An Introduction to Epigenetics

Roger Hoover

First Edition, 2012

ISBN 978-81-323-4071-3

Published by:
White Word Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1 - Epigenetics

Chapter 2 - Transgenerational Epigenetics

Chapter 3 - Genomic Imprinting

Chapter 4 - Methylated DNA Immunoprecipitation

Chapter 5 - Bisulfite Sequencing

Chapter 6 - DNA Methylation

Chapter 7 - Nutriepigenomics

Chapter 8 - Paramutation and Sex-Determination System

Chapter 9 - Soft Inheritance, Structural Inheritance and Testis Determining Factor

Chapter 10 - X-Inactivation

Chapter 11 - Sex Determination and Differentiation (Human)

Chapter 12 - Histone

Chapter 13 - XY Sex-Determination System

Chapter 14 - Haplodiploid Sex-Determination System and Temperature-Dependent Sex Determination

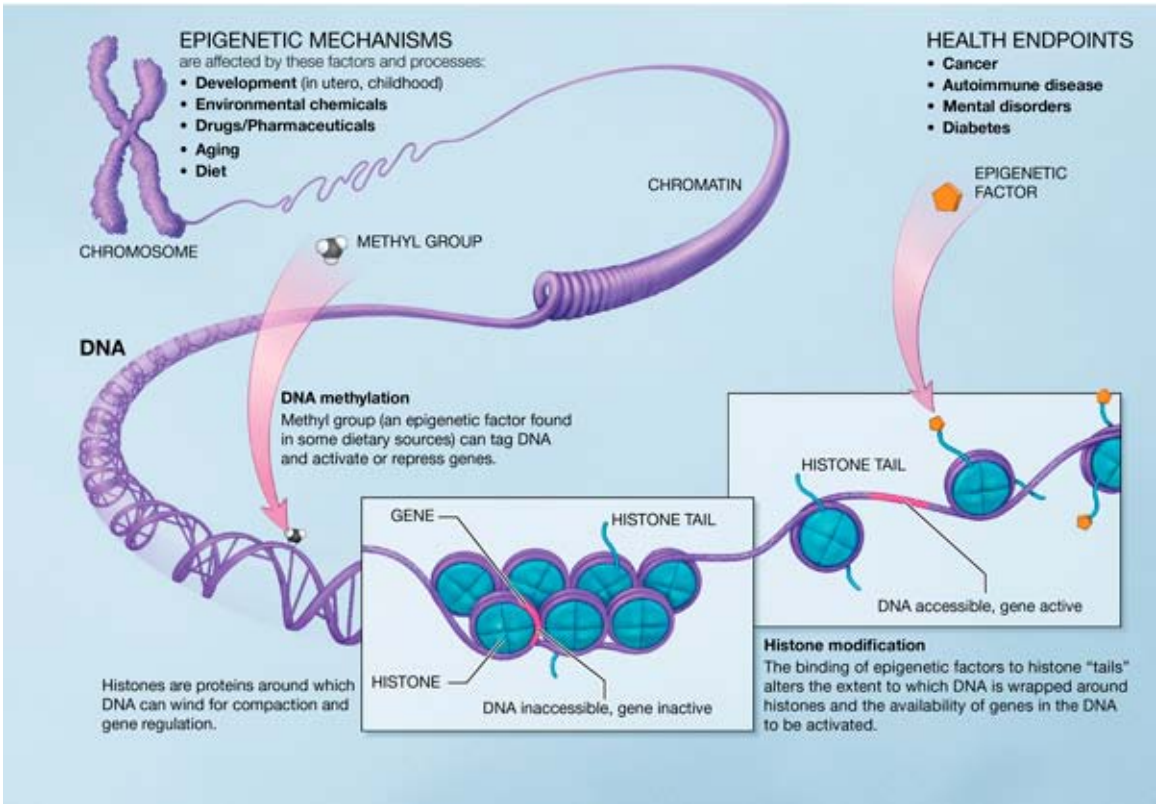
Chapter 1

Epigenetics

In biology, and specifically genetics, **epigenetics** is the study of heritable changes in phenotype (appearance) or gene expression caused by mechanisms other than changes in the underlying DNA sequence, hence the name *epi-* (Greek: *επί-* over, above) *-genetics*. These changes may remain through cell divisions for the remainder of the cell's life and may also last for multiple generations. However, there is no change in the underlying DNA sequence of the organism; instead, non-genetic factors cause the organism's genes to behave (or "express themselves") differently.

One example of epigenetic changes in eukaryotic biology is the process of cellular differentiation. During morphogenesis, totipotent stem cells become the various pluripotent cell lines of the embryo which in turn become fully differentiated cells. In other words, a single fertilized egg cell – the zygote – changes into the many cell types including neurons, muscle cells, epithelium, blood vessels etc. as it continues to divide. It does so by activating some genes while inhibiting others.

Etymology and definitions



Epigenetic mechanisms

Epigenetics (as in "epigenetic landscape") was coined by C. H. Waddington in 1942 as a portmanteau of the words *genetics* and *epigenesis*. *Epigenesis* is an old word which has more recently been used to describe the differentiation of cells from their initial totipotent state in embryonic development. When Waddington coined the term the physical nature of genes and their role in heredity was not known; he used it as a conceptual model of how genes might interact with their surroundings to produce a phenotype.

Robin Holliday defined epigenetics as "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms." Thus *epigenetic* can be used to describe anything other than DNA sequence that influences the development of an organism.

The modern usage of the word in scientific discourse is more narrow, referring to heritable traits (over rounds of cell division and sometimes transgenerationally) that do not involve changes to the underlying DNA sequence. The Greek prefix *epi-* in *epigenetics* implies features that are "on top of" or "in addition to" genetics; thus *epigenetic* traits exist on top of or in addition to the traditional molecular basis for inheritance.

The similarity of the word to "genetics" has generated many parallel usages. The "epigenome" is a parallel to the word "genome", and refers to the overall epigenetic state of a cell. The phrase "genetic code" has also been adapted—the "epigenetic code" has been used to describe the set of epigenetic features that create different phenotypes in different cells. Taken to its extreme, the "epigenetic code" could represent the total state of the cell, with the position of each molecule accounted for in an *epigenomic map*, a diagrammatic representation of the gene expression, DNA methylation and histone modification status of a particular genomic region. More typically, the term is used in reference to systematic efforts to measure specific, relevant forms of epigenetic information such as the histone code or DNA methylation patterns.

The psychologist Erik Erikson used the term *epigenetic* in his theory of psychosocial development. That usage, however, is of primarily historical interest.

Molecular basis of epigenetics

The molecular basis of epigenetics is complex. It involves modifications of the activation of certain genes, but not the basic structure of DNA. Additionally, the chromatin proteins associated with DNA may be activated or silenced. This accounts for why the differentiated cells in a multi-cellular organism express only the genes that are necessary for their own activity. Epigenetic changes are preserved when cells divide. Most epigenetic changes only occur within the course of one individual organism's lifetime, but, if a mutation in the DNA has been caused in sperm or egg cell that results in fertilization, then some epigenetic changes are inherited from one generation to the next. This raises the question of whether or not epigenetic changes in an organism can alter the basic structure of its DNA, a form of Lamarckism.

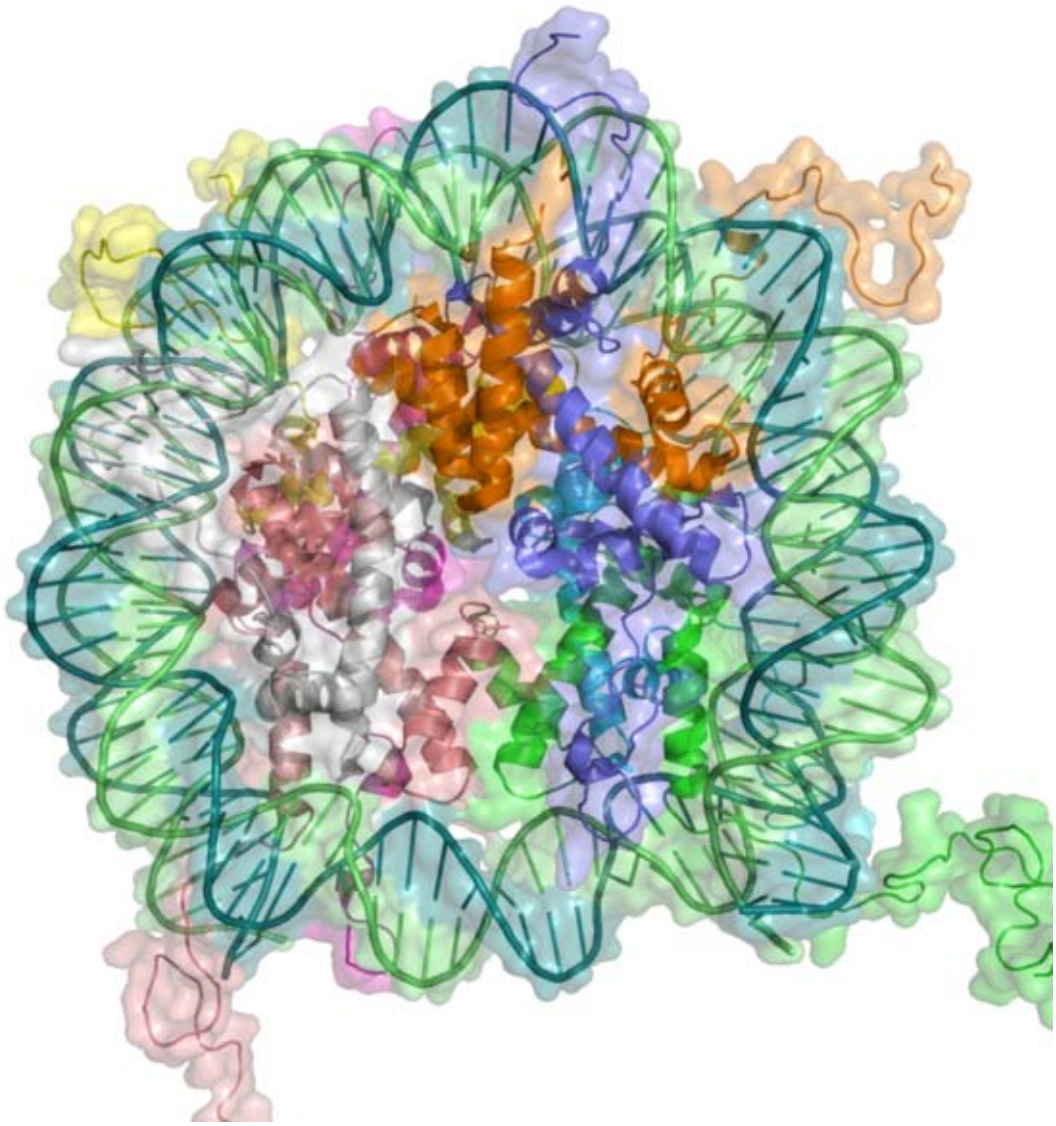
Specific epigenetic processes include paramutation, bookmarking, imprinting, gene silencing, X chromosome inactivation, position effect, reprogramming, transvection, maternal effects, the progress of carcinogenesis, many effects of teratogens, regulation of histone modifications and heterochromatin, and technical limitations affecting parthenogenesis and cloning.

Epigenetic research uses a wide range of molecular biologic techniques to further our understanding of epigenetic phenomena, including chromatin immunoprecipitation (together with its large-scale variants ChIP-on-chip and ChIP-seq), fluorescent in situ hybridization, methylation-sensitive restriction enzymes, DNA adenine methyltransferase identification (DamID) and bisulfite sequencing. Furthermore, the use of bioinformatic methods is playing an increasing role (computational epigenetics).

Mechanisms

Several types of epigenetic inheritance systems may play a role in what has become known as cell memory:

DNA methylation and chromatin remodeling



DNA associates with histone proteins to form chromatin.

Because the phenotype of a cell or individual is affected by which of its genes are transcribed, heritable transcription states can give rise to epigenetic effects. There are several layers of regulation of gene expression. One way that genes are regulated is through the remodeling of chromatin. Chromatin is the complex of DNA and the histone proteins with which it associates. Histone proteins are little spheres that DNA wraps around. If the way that DNA is wrapped around the histones changes, gene expression can change as well. Chromatin remodeling is accomplished through two main mechanisms:

1. The first way is post translational modification of the amino acids that make up histone proteins. Histone proteins are made up of long chains of amino acids. If

the amino acids that are in the chain are changed, the shape of the histone sphere might be modified. DNA is not completely unwound during replication. It is possible, then, that the modified histones may be carried into each new copy of the DNA. Once there, these histones may act as templates, initiating the surrounding new histones to be shaped in the new manner. By altering the shape of the histones around it, these modified histones would ensure that a differentiated cell would stay differentiated, and not convert back into being a stem cell.

2. The second way is the addition of methyl groups to the DNA, mostly at CpG sites, to convert cytosine to 5-methylcytosine. 5-Methylcytosine performs much like a regular cytosine, pairing up with a guanine. However, some areas of genome are methylated more heavily than others and highly methylated areas tend to be less transcriptionally active, through a mechanism not fully understood. Methylation of cytosines can also persist from the germ line of one of the parents into the zygote, marking the chromosome as being inherited from this parent (genetic imprinting).

The way that the cells stay differentiated in the case of DNA methylation is clearer to us than it is in the case of histone shape. Basically, certain enzymes (such as DNMT1) have a higher affinity for the methylated cytosine. If this enzyme reaches a "hemimethylated" portion of DNA (where methylcytosine is in only one of the two DNA strands) the enzyme will methylate the other half.

Although histone modifications occur throughout the entire sequence, the unstructured N-termini of histones (called histone tails) are particularly highly modified. These modifications include acetylation, methylation, ubiquitylation, phosphorylation and sumoylation. Acetylation is the most highly studied of these modifications. For example, acetylation of the K14 and K9 lysines of the tail of histone H3 by histone acetyltransferase enzymes (HATs) is generally correlated with transcriptional competence.

One mode of thinking is that this tendency of acetylation to be associated with "active" transcription is biophysical in nature. Because it normally has a positively charged nitrogen at its end, lysine can bind the negatively charged phosphates of the DNA backbone. The acetylation event converts the positively charged amine group on the side chain into a neutral amide linkage. This removes the positive charge, thus loosening the DNA from the histone. When this occurs, complexes like SWI/SNF and other transcriptional factors can bind to the DNA and allow transcription to occur. This is the "cis" model of epigenetic function. In other words, changes to the histone tails have a direct affect on the DNA itself.

Another model of epigenetic function is the "trans" model. In this model changes to the histone tails act indirectly on the DNA. For example, lysine acetylation may create a binding site for chromatin modifying enzymes (and basal transcription machinery as well). This Chromatin Remodeler can then cause changes to the state of the chromatin. Indeed, the bromodomain — a protein segment (domain) that specifically binds acetyl-

lysine — is found in many enzymes that help activate transcription, including the SWI/SNF complex (on the protein polybromo). It may be that acetylation acts in this and the previous way to aid in transcriptional activation.

The idea that modifications act as docking modules for related factors is borne out by histone methylation as well. Methylation of lysine 9 of histone H3 has long been associated with constitutively transcriptionally silent chromatin (constitutive heterochromatin). It has been determined that a chromodomain (a domain that specifically binds methyl-lysine) in the transcriptionally repressive protein HP1 recruits HP1 to K9 methylated regions. One example that seems to refute this biophysical model for acetylation is that tri-methylation of histone H3 at lysine 4 is strongly associated with (and required for full) transcriptional activation. Tri-methylation in this case would introduce a fixed positive charge on the tail.

It has been shown that the histone lysine methyltransferase (KMT) is responsible for this methylation activity in the pattern of histones H3 & H4. This enzyme utilizes a catalytically active site called the SET domain (Supressor of variegation, Enhancer of zeste, Trithorax). The SET domain is a 130-amino acid sequence involved in modulating gene activities. This domain has been demonstrated to bind to the histone tail and causes the methylation of the histone.

Differing histone modifications are likely to function in differing ways; acetylation at one position is likely to function differently than acetylation at another position. Also, multiple modifications may occur at the same time, and these modifications may work together to change the behavior of the nucleosome. The idea that multiple dynamic modifications regulate gene transcription in a systematic and reproducible way is called the histone code.

DNA methylation frequently occurs in repeated sequences, and helps to suppress the expression and mobility of 'transposable elements': Because 5-methylcytosine is chemically very similar to thymidine, CpG sites are frequently mutated and become rare in the genome, except at CpG islands where they remain unmethylated. Epigenetic changes of this type thus have the potential to direct increased frequencies of permanent genetic mutation. DNA methylation patterns are known to be established and modified in response to environmental factors by a complex interplay of at least three independent DNA methyltransferases, DNMT1, DNMT3A and DNMT3B, the loss of any of which is lethal in mice. DNMT1 is the most abundant methyltransferase in somatic cells, localizes to replication foci, has a 10–40-fold preference for hemimethylated DNA and interacts with the proliferating cell nuclear antigen (PCNA). By preferentially modifying hemimethylated DNA, DNMT1 transfers patterns of methylation to a newly synthesized strand after DNA replication, and therefore is often referred to as the ‘maintenance' methyltransferase. DNMT1 is essential for proper embryonic development, imprinting and X-inactivation.

Histones H3 and H4 can also be manipulated through demethylation using histone lysine demethylase (KDM). This recently identifited enzyme has a catalytically active site

called the Jumonji domain (JmjC). The demethylation occurs when JmjC utilizes multiple cofactors to hydroxylate the methyl group, thereby removing it. JmjC is capable of demethylating mono-, di-, and tri-methylated substrates. .

Chromosomal regions can adopt stable and heritable alternative states resulting in bistable gene expression without changes to the DNA sequence. Epigenetic control is often associated with alternative covalent modifications of histones. The stability and heritability of states of larger chromosomal regions are often thought to involve positive feedback where modified nucleosomes recruit enzymes that similarly modify nearby nucleosomes. A simplified stochastic model for this type of epigenetics is found here .

Because DNA methylation and chromatin remodeling play such a central role in many types of epigenic inheritance, the word "epigenetics" is sometimes used as a synonym for these processes. However, this can be misleading. Chromatin remodeling is not always inherited, and not all epigenetic inheritance involves chromatin remodeling.

It has been suggested that the histone code could be mediated by the effect of small RNAs. The recent discovery and characterization of a vast array of small (21- to 26-nt), non-coding RNAs suggests that there is an RNA component, possibly involved in epigenetic gene regulation. Small interfering RNAs can modulate transcriptional gene expression via epigenetic modulation of targeted promoters.

RNA transcripts and their encoded proteins

Sometimes a gene, after being turned on, transcribes a product that (either directly or indirectly) maintains the activity of that gene. For example, Hnf4 and MyoD enhance the transcription of many liver- and muscle-specific genes, respectively, including their own, through the transcription factor activity of the proteins they encode. RNA signalling includes differential recruitment of a hierarchy of generic chromatin modifying complexes and DNA methyltransferases to specific loci by RNAs during differentiation and development. Other epigenetic changes are mediated by the production of different splice forms of RNA, or by formation of double-stranded RNA (RNAi). Descendants of the cell in which the gene was turned on will inherit this activity, even if the original stimulus for gene-activation is no longer present. These genes are most often turned on or off by signal transduction, although in some systems where syncytia or gap junctions are important, RNA may spread directly to other cells or nuclei by diffusion. A large amount of RNA and protein is contributed to the zygote by the mother during oogenesis or via nurse cells, resulting in maternal effect phenotypes. A smaller quantity of sperm RNA is transmitted from the father, but there is recent evidence that this epigenetic information can lead to visible changes in several generations of offspring.

Prions

Prions are infectious forms of proteins. Proteins generally fold into discrete units which perform distinct cellular functions, but some proteins are also capable of forming an infectious conformational state known as a prion. Although often viewed in the context of

infectious disease, prions are more loosely defined by their ability to catalytically convert other native state versions of the same protein to an infectious conformational state. It is in this latter sense that they can be viewed as epigenetic agents capable of inducing a phenotypic change without a modification of the genome.

Fungal prions are considered epigenetic because the infectious phenotype caused by the prion can be inherited without modification of the genome. PSI+ and URE3, discovered in yeast in 1965 and 1971, are the two best studied of this type of prion. Prions can have a phenotypic effect through the sequestration of protein in aggregates, thereby reducing that protein's activity. In PSI+ cells, the loss of the Sup35 protein (which is involved in termination of translation) causes ribosomes to have a higher rate of read-through of stop codons, an effect which results in suppression of nonsense mutations in other genes. The ability of Sup35 to form prions may be a conserved trait. It could confer an adaptive advantage by giving cells the ability to switch into a PSI+ state and express dormant genetic features normally terminated by premature stop codon mutations.

Structural inheritance systems

In ciliates such as *Tetrahymena* and *Paramecium*, genetically identical cells show heritable differences in the patterns of ciliary rows on their cell surface. Experimentally altered patterns can be transmitted to daughter cells. It seems existing structures act as templates for new structures. The mechanisms of such inheritance are unclear, but reasons exist to assume that multicellular organisms also use existing cell structures to assemble new ones.

Functions and consequences

Development

Somatic epigenetic inheritance, particularly through DNA methylation and chromatin remodeling, is very important in the development of multicellular eukaryotic organisms. The genome sequence is static (with some notable exceptions), but cells differentiate into many different types, which perform different functions, and respond differently to the environment and intercellular signalling. Thus, as individuals develop, morphogens activate or silence genes in an epigenetically heritable fashion, giving cells a "memory". In mammals, most cells terminally differentiate, with only stem cells retaining the ability to differentiate into several cell types ("totipotency" and "multipotency"). In mammals, some stem cells continue producing new differentiated cells throughout life, but mammals are not able to respond to loss of some tissues, for example, the inability to regenerate limbs, which some other animals are capable of. Unlike animals, plant cells do not terminally differentiate, remaining totipotent with the ability to give rise to a new individual plant. While plants do utilise many of the same epigenetic mechanisms as animals, such as chromatin remodeling, it has been hypothesised that plant cells do not have "memories", resetting their gene expression patterns at each cell division using positional information from the environment and surrounding cells to determine their fate.

Medicine

Epigenetics has many and varied potential medical applications. Congenital genetic disease is well understood, and it is also clear that epigenetics can play a role, for example, in the case of Angelman syndrome and Prader-Willi syndrome. These are normal genetic diseases caused by gene deletions or inactivation of the genes, but are unusually common because individuals are essentially hemizygous because of genomic imprinting, and therefore a single gene knock out is sufficient to cause the disease, where most cases would require both copies to be knocked out.

Evolution

Although epigenetics in multicellular organisms is generally thought to be a mechanism involved in differentiation, with epigenetic patterns "reset" when organisms reproduce, there have been some observations of transgenerational epigenetic inheritance (e.g., the phenomenon of paramutation observed in maize). Although most of these multigenerational epigenetic traits are gradually lost over several generations, the possibility remains that multigenerational epigenetics could be another aspect to evolution and adaptation. A sequestered germ line or Weismann barrier is specific to animals, and epigenetic inheritance is expected to be far more common in plants and microbes. These effects may require enhancements to the standard conceptual framework of the modern evolutionary synthesis.

Epigenetic features may play a role in short-term adaptation of species by allowing for reversible phenotype variability. The modification of epigenetic features associated with a region of DNA allows organisms, on a multigenerational time scale, to switch between phenotypes that express and repress that particular gene. When the DNA sequence of the region is not mutated, this change is reversible. It has also been speculated that organisms may take advantage of differential mutation rates associated with epigenetic features to control the mutation rates of particular genes. Interestingly, recent analysis have suggested that members of the APOBEC family of cytosine deaminases are capable of simultaneously mediating genetic and epigenetic inheritance using similar molecular mechanisms.

Epigenetic changes have also been observed to occur in response to environmental exposure—for example, mice given some dietary supplements have epigenetic changes affecting expression of the agouti gene, which affects their fur color, weight, and propensity to develop cancer.

More than 100 cases of transgenerational epigenetic inheritance phenomena have been reported in a wide range of organisms, including prokaryotes, plants, and animals.

Epigenetic effects in humans

Genomic imprinting and related disorders

Some human disorders are associated with genomic imprinting, a phenomenon in mammals where the father and mother contribute different epigenetic patterns for specific genomic loci in their germ cells. The best-known case of imprinting in human disorders is that of Angelman syndrome and Prader-Willi syndrome—both can be produced by the same genetic mutation, chromosome 15q partial deletion, and the particular syndrome that will develop depends on whether the mutation is inherited from the child's mother or from their father. This is due to the presence of genomic imprinting in the region. Beckwith-Wiedemann syndrome is also associated with genomic imprinting, often caused by abnormalities in maternal genomic imprinting of a region on chromosome 11.

Transgenerational epigenetic observations

Marcus Pembrey and colleagues also observed in the Överkalix study that the paternal (but not maternal) grandsons of Swedish boys who were exposed during preadolescence to famine in the 19th century were less likely to die of cardiovascular disease; if food was plentiful then diabetes mortality in the grandchildren increased, suggesting that this was a transgenerational epigenetic inheritance. The opposite effect was observed for females—the paternal (but not maternal) granddaughters of women who experienced famine while in the womb (and therefore while their eggs were being formed) lived shorter lives on average.

Cancer and developmental abnormalities

A variety of compounds are considered as epigenetic carcinogens—they result in an increased incidence of tumors, but they do not show mutagen activity (toxic compounds or pathogens that cause tumors incident to increased regeneration should also be excluded). Examples include diethylstilbestrol, arsenite, hexachlorobenzene, and nickel compounds.

Many teratogens exert specific effects on the fetus by epigenetic mechanisms. While epigenetic effects may preserve the effect of a teratogen such as diethylstilbestrol throughout the life of an affected child, the possibility of birth defects resulting from exposure of fathers or in second and succeeding generations of offspring has generally been rejected on theoretical grounds and for lack of evidence. However, a range of male-mediated abnormalities have been demonstrated, and more are likely to exist. FDA label information for Vidaza(tm), a formulation of 5-azacitidine (an unmethylatable analog of cytidine that causes hypomethylation when incorporated into DNA) states that "men should be advised not to father a child" while using the drug, citing evidence in treated male mice of reduced fertility, increased embryo loss, and abnormal embryo development. In rats, endocrine differences were observed in offspring of males exposed to morphine. In mice, second generation effects of diethylstilbesterol have been described occurring by epigenetic mechanisms.

Recent studies have shown that the Mixed Lineage Leukemia (MLL) gene causes leukemia by rearranging and fusing with other genes in different chromosomes, which is a process under epigenetic control.

Other investigations have concluded that alterations in histone acetylation and DNA methylation occur in various genes influencing prostate cancer.

In 2008, the National Institutes of Health announced that $190 million had been earmarked for epigenetics research over the next five years. In announcing the funding, government officials noted that epigenetics has the potential to explain mechanisms of aging, human development, and the origins of cancer, heart disease, mental illness, as well as several other conditions. Some investigators, like Randy Jirtle, PhD, of Duke University Medical Center, think epigenetics may ultimately turn out to have a greater role in disease than genetics.

DNA methylation in cancer

DNA methylation is an important regulator of gene transcription and a large body of evidence has demonstrated that aberrant DNA methylation is associated with unscheduled gene silencing, and the genes with high levels of 5-methylcytosine in their promoter region are transcriptionally silent. DNA methylation is essential during embryonic development, and in somatic cells, patterns of DNA methylation are generally transmitted to daughter cells with a high fidelity. Aberrant DNA methylation patterns have been associated with a large number of human malignancies and found in two distinct forms: hypermethylation and hypomethylation compared to normal tissue. Hypermethylation is one of the major epigenetic modifications that repress transcription via promoter region of tumour suppressor genes. Hypermethylation typically occurs at CpG islands in the promoter region and is associated with gene inactivation. Global hypomethylation has also been implicated in the development and progression of cancer through different mechanisms.

Variant histones H2A in cancer

The histone variants of the H2A family are highly conserved in mammals, playing critical roles in regulating many nuclear processes by altering chromatin structure. One of the key H2A variants, H2A.X, marks DNA damage, facilitating the recruitment of DNA repair proteins to restore genomic integrity. Another variant, H2A.Z, plays an important role in both gene activation and repression. A high level of H2A.Z expression is ubiquitously detected in many cancers and is significantly associated with cellular proliferation and genomic instability.

Cancer Treatment

Current research has shown that epigenetic pharmaceuticals could be a putative replacement or adjuvant therapy for currently accepted treatment methods such as radiation and chemotherapy, or could enhance the effects of these current treatments. It

has been shown that the epigenetic control of the proto-onco regions and the tumor suppressor sequences by conformational changes in histones directly affects the formation and progression of cancer Epigenetics also has the factor of reversibility, a characteristic that other cancer treatments do not offer.

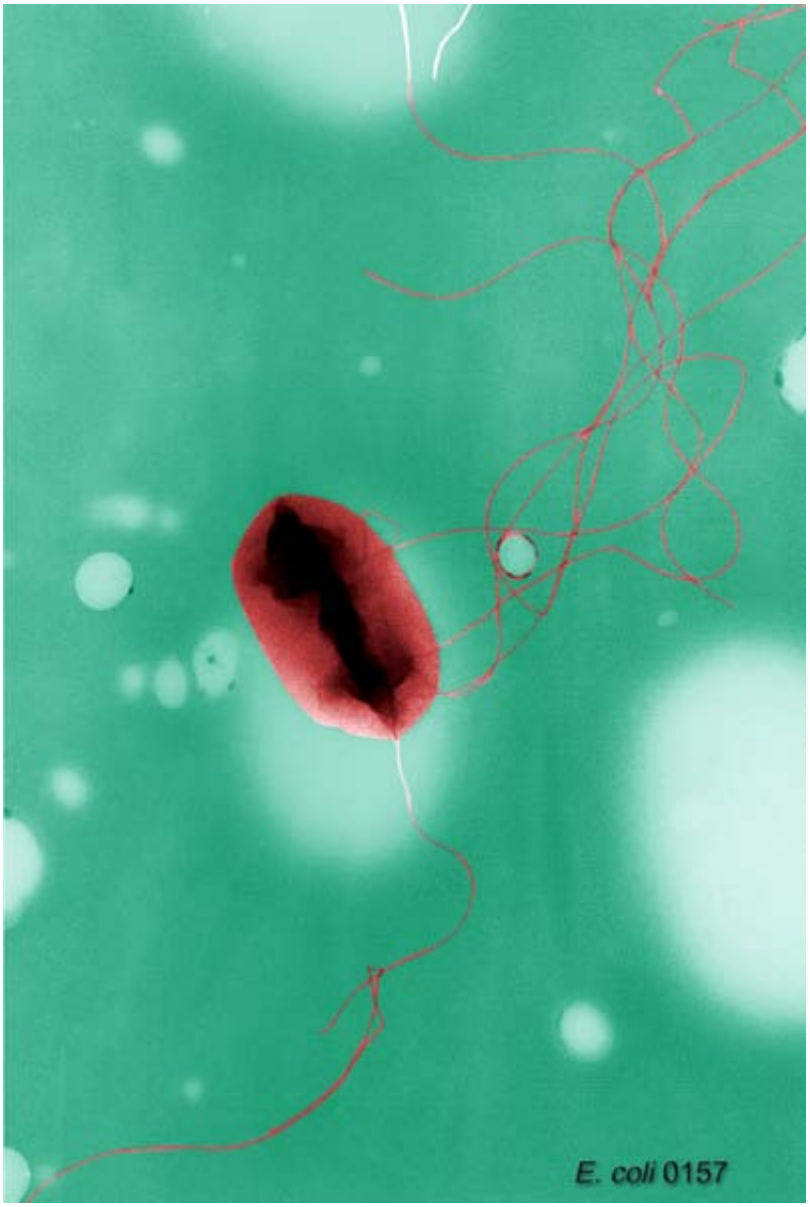
Drug development has mainly focused on Histone Acetyltransferase (HAT) and Histone Deactylase (HDAC), including the introduction of the new pharmaceutical Vorinostat, a HDAC inhibitor, to the market. HDAC specifically has been shown to play an integral role in the progression of oral squamous cancer

Current front-runner candidates for new drug targets are Histone Lysine Methyltransferases (KMT) and Protein Arginine Methyltransferases (PRMT).

Twin studies

Recent studies involving both dizygotic and monozygotic twins have produced some evidence of epigenetic influence in humans.

Epigenetics in microorganisms



Escherichia coli bacteria

Bacteria make widespread use of postreplicative DNA methylation for the epigenetic control of DNA-protein interactions. Bacteria make use of DNA adenine methylation (rather than DNA cytosine methylation) as an epigenetic signal. DNA adenine methylation is important in bacteria virulence in organisms such as *Escherichia coli*, *Salmonella, Vibrio, Yersinia, Haemophilus*, and *Brucella*. In *Alphaproteobacteria*, methylation of adenine regulates the cell cycle and couples gene transcription to DNA replication. In *Gammaproteobacteria*, adenine methylation provides signals for DNA replication, chromosome segregation, mismatch repair, packaging of bacteriophage, transposase activity and regulation of gene expression.

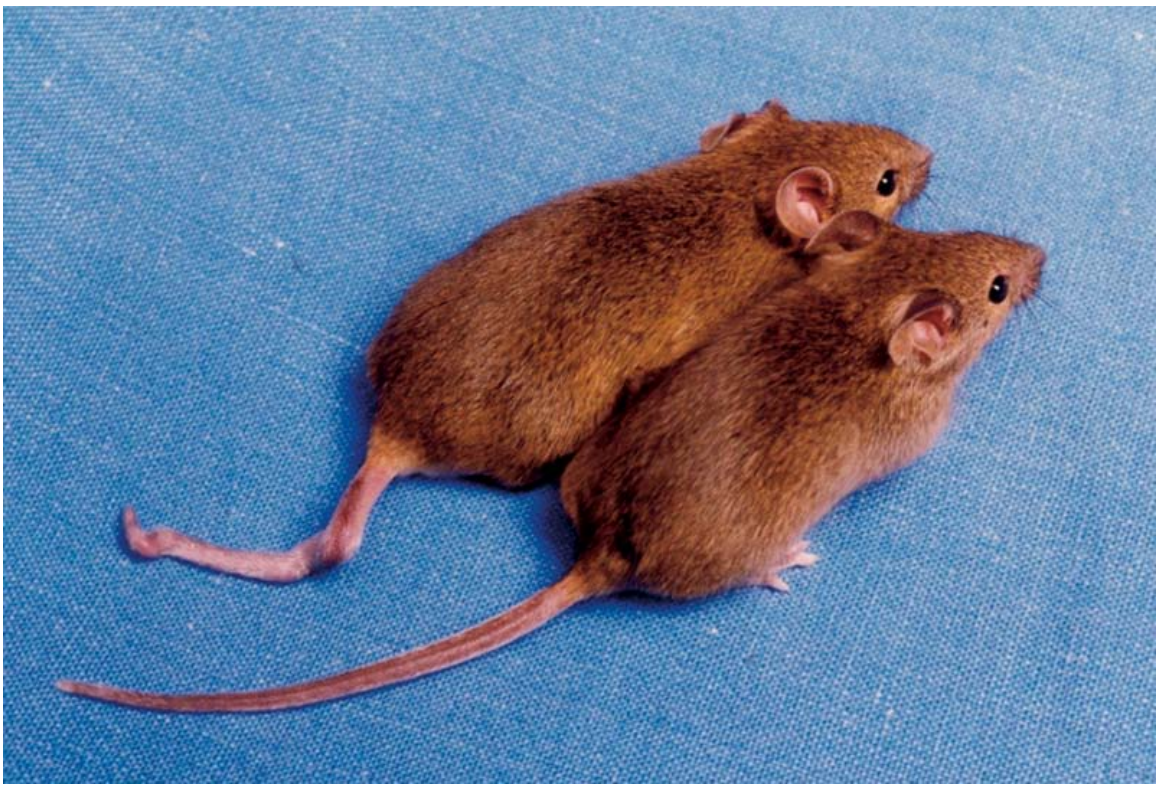
The filamentous fungus *Neurospora crassa* is a prominent model system for understanding the control and function of cytosine methylation. In this organisms, DNA methylation is associated with relics of a genome defense system called RIP (repeat-induced point mutation) and silences gene expression by inhibiting transcription elongation.

The yeast prion PSI is generated by a conformational change of a translation termination factor, which is then inherited by daughter cells. This can provide a survival advantage under adverse conditions. This is an example of epigenetic regulation enabling unicellular organisms to respond rapidly to environmental stress. Prions can be viewed as epigenetic agents capable of inducing a phenotypic change without modification of the genome.

Chapter 2

Transgenerational Epigenetics

Epigenetic inheritance is the transmittance of information from one generation to the next that affects the traits of offspring without alteration of the primary structure of DNA (i.e. the sequence of nucleotides) or from environmental cues. The term “epigenetic inheritance” is used to describe both cell-cell and organism-organism information transfer, while **transgenerational epigenetics** typically refers only to the latter. Although these two levels of epigenetic inheritance are equivalent in unicellular organisms, they may have distinct mechanisms and evolutionary distinctions in multicellular organisms.



Cloned mice with different DNA methylation patterns causing kinks in the tail of one but not the other.

Four general categories of epigenetic modification are known: 1) self-sustaining metabolic loops, in which a mRNA or protein product of a gene stimulates transcription of the gene (e.g. *Wor1* gene in *Candida albicans*), 2) structural templating in which structures are replicated using a template or scaffold structure on the parent (e.g. prions, proteins that replicate by changing the structure of normal proteins to match their own), 3) chromatin marks, in which methyl or acetyl groups bind to DNA nucleotides or histones thereby altering gene expression patterns (e.g. *Lcyc* gene in *Linaria vulgaris* described below), and 4) RNA silencing, in which small RNA strands interfere (RNAi) with the transcription of DNA or translation of mRNA (known only from a few studies, mostly in *Caenorhabditis elegans*).

For some epigenetically influenced traits, the epigenetic marks can be induced by the environment and some marks are heritable, leading some to view epigenetics as a relaxation of the rejection of Lamarckian evolution of acquired characters.

Major Controversies in the History of the Inheritance

Humans have recognized that traits of the parents are often seen in offspring. This insight led to the practical application of selective breeding of plants and animals, eventually leading to domestication, but did not address the central question of inheritance: how are these traits conserved between generations, and what causes variation?

Blending vs. Particulate Inheritance

Addressing these related questions, scientists during the time of the Enlightenment largely argued for the blending hypothesis, in which parental traits were homogenized in the offspring much like buckets of different colored paint being mixed together. Critics of Charles Darwin's *On the Origin of Species*, pointed out that under this scheme of inheritance, variation would quickly be swamped by the majority phenotype. In the paint bucket analogy, this would be seen by mixing two colors together and then mixing the resulting color with only one of the parent colors 20 times; the rare variant color would quickly fade.

Unknown to most of the European scientific community, a monk by the name of Gregor Mendel had resolved the question of how traits are conserved between generations through breeding experiments with pea plants. Charles Darwin thus did not know of Mendel's proposed "particulate inheritance" in which traits were not blended but passed to offspring in discrete units that we now call genes. Darwin came to reject the blending hypothesis even though his ideas and Mendel's were not unified until the 1930s, a period referred to as the Modern Synthesis.

Inheritance of Innate vs. Acquired Characteristics

In his 1809 book, *Philosophie Zoologique*, Jean-Baptiste Lamarck recognized that each species experiences a unique set of challenges due to its form and environment. Thus, he proposed that the characters used most often would accumulate a “nervous fluid.” Such

acquired accumulations would then be transmitted to the individual's offspring. In modern terms, a nervous fluid transmitted to offspring would be a form of epigenetic inheritance.

Lamarckism, as this body of thought became known, was the standard explanation for change in species over time when Charles Darwin and Alfred Russel Wallace co-proposed a theory of evolution by natural selection in 1859. Responding to Darwin and Wallace's theory, a revised neo-Lamarckism attracted a small following of biologists, though the Lamarckian zeal was quenched in large part due to Weismann's famous experiment in which he cut off the tails of mice over several successive generations without having any effect on tail length. Thus the emergent consensus that acquired characteristics could not be inherited became canon.

Origin of Epigenetics and Revision of the Modern Synthesis

Non-genetic variation and inheritance, however, proved to be quite common. Concurrent to the Modern Synthesis (unifying Mendelian genetics and natural selection), C. H. Waddington was working to unify developmental biology and genetics. In so doing, he coined the word "epigenetic" to represent the ordered differentiation of embryonic cells into functionally distinct cell types despite having identical primary structure of their DNA. Waddington's epigenetics was sporadically discussed, becoming more of a catch-all for puzzling non-genetic heritable characters rather than advancing the body of inquiry. Consequently, the definition of Waddington's word has itself evolved, broadening beyond the subset of developmentally signaled, inherited cell specialization.

Does epigenetic inheritance compromise the foundation of the Modern Synthesis? Outlining the Central Dogma of Molecular Biology, Francis Crick succinctly stated, "DNA is held in a configuration by histone[s] so that it can act as a passive template for the simultaneous synthesis of RNA and protein[s]. *None* of the detailed "information" is in the histone (italic added for emphasis). However, he closes the article stating, "this scheme *explains the majority* of the present experimental results!" (italic added for emphasis). Indeed the emergence of epigenetic inheritance (in addition to advances in the study of evolutionary-development, phenotypic plasticity, evolvability, and systems biology) has strained the current framework of the Modern Synthesis and prompted the re-examination of previously dismissed evolutionary mechanisms.

Origin and Inheritance of Epigenes

Epigenetic variation may take one of four general forms. Others may yet be elucidated, but currently self-sustaining feedback loops, spatial templating, chromatin marking, and RNA-mediated pathways modify epigenes at the level of individual cells. Epigenetic variation within multicellular organisms may be endogenous, generated by cell-cell signaling (e.g. during cell differentiation early in development), or exogenous, a cellular response to environmental cues.

Removal vs. retention of epigenetic marks

In sexually reproducing organisms, much of the epigenetic modification within cells is reset during meiosis (e.g. marks at the FLC locus controlling plant vernalization), though some epigenetic responses have been shown to be conserved (e.g. transposon methylation in plants). Differential inheritance of epigenetic marks due to underlying maternal or paternal biases in removal or retention mechanisms may lead to the assignment of epigenetic causation to some parent of origin effects in animals and plants.

Removal of epigenetic marks

In mammals, male and female gametes join during fertilization in different cell cycle states and with different configuration of the genome. The epigenetic marks of the male are rapidly diluted. First, the protamines associated with male DNA are replaced with histones from the female's cytoplasm, most of which are acetylated due to either higher abundance of acetylated histones in the female's cytoplasm or through preferential binding of the male DNA to acetylated histones. Second, male DNA is systematically demethylated in many organisms, however the mechanism and functional outcome of this process has yet to be elucidated.

Recognition of the importance of epigenetic programming to the establishment and fixation of cell line identity during early embryogenesis has recently stimulated interest in artificial removal of epigenetic programming. Epigenetic manipulations may allow for restoration of totipotency in stem cells or cells more generally, thus generalizing regenerative medicine.

Retention of epigenetic marks

Cellular mechanisms may allow for co-transmission of some epigenetic marks. During replication, DNA polymerases working on the leading and lagging strands are coupled by the DNA processivity factor proliferating cell nuclear antigen (PCNA), which has also been implicated in patterning and strand crosstalk that allows for copy fidelity of epigenetic marks. Work on histone modification copy fidelity has remained in the model phase, but early efforts suggest that modifications of new histones are patterned on those of the old histones and that new and old histones randomly assort between the two daughter DNA strands. With respect to transfer to the next generation, many marks are removed as described above. Emerging studies are finding patterns of epigenetic conservation across generations. For instance, centromeric satellites resist demethylation. The mechanism responsible for this conservation is not known, though some evidence suggests that methylation of histones may contribute.

Decay of epigenetic marks

Whereas the mutation rate in a given 100 base gene may be 10^{-7} per generation, epigenes may “mutate” several times per generation or may be fixed for many generations. This raises the question: are changes in epigene frequencies is evolution? Rapidly decaying

epigenetic effects on phenotypes (i.e. lasting less than three generations) may explain some of the residual variation in phenotypes after genotype and environment are accounted for. However, distinguishing these short-term effects from the effects of the maternal environment on early onotgeny remains a challenge.

Contribution to Phenotypes

The relative importance of genetic and epigenetic inheritance is subject to debate. Though hundreds of examples of epigenetic modification of phenotypes have been published, few studies have been conducted outside of the laboratory setting. Therefore, the interactions of genes and epigenes with the environment cannot be inferred despite the central role of environment in natural selection. Experimental methodologies for manipulating epigenetic mechanisms are nascient (e.g.) and will need rigorous demonstration before studies explicitly testing the relative contributions of genotype, environment, and epigenotype are feasible.

Effects on Fitness

Epigenetic inheritance may only affect fitness if it predictably alters a trait under selection. Evidence has been forwarded that environmental stimuli are important agents in the alteration of epigenes. Ironically, Darwinian evolution may act on these neo-Lamarckian acquired characteristics as well as the cellular mechanisms producing them (e.g. methyltransferase genes). Epigenetic inheritance may confer a fitness benefit to organisms that deal with environmental changes at intermediate timescales. Short-cycling changes are likely to have DNA-encoded regulatory processes, as the probability of the offspring needing to respond to changes multiple times during their lifespans is high. On the other end, natural selection will act on populations experiencing changes on longer-cycling environmental changes. In these cases, if epigenetic priming of the next generation is deleterious to fitness over most of the interval (e.g. misinformation about the environment), these genotypes and epigenotypes will be lost. For intermediate time cycles, the probability of the offspring encountering a similar environment is sufficiently high without substantial selective pressure on individuals lacking a genetic architecture capable of responding to the environment. Naturally, the absolute lengths of short, intermediate, and long environmental cycles will depend on the trait, the length of epigenetic memory, and the generation time of the organism. Much of the interpretation of epigenetic fitness effects centers on the hypothesis that epigenes are important contributors to phenotypes, which remains to be resolved.

Deleterious effects

Inherited epigenetic marks may be important for regulating important components of fitness. In plants, for instance, the *Lcyc* gene in *Linaria vulgaris* controls the symmetry of the flower. Linnaeus first described radially symmetric mutants, which arises when *Lcyc* is heavily methylated. Given the importance of floral shape to pollinators, methylation of *Lcyc* homologues (e.g. *CYCLOIDEA*) may have deleterious effects on plant fitness. In animals, numerous studies have shown that inherited epigenetic marks can increase

susceptibility to disease. Transgenerational epigenetic influences are also suggested to contribute to disease, especially cancer, in humans. Tumor methylation patterns in gene promotors have been shown to correlate positively with familial history of cancer. Furthermore, methylation of the *MSH2* gene is correlated with early-onset colorectal and endometrial cancers.

Putatively adaptive effects

Experimentally demethylated seeds of the model organism *Arabidopsis thaliana* have significantly higher mortality, stunted growth, delayed flowing, and lower fruit set, indicating that epigenes may increase fitness. Furthermore, environmentally induced epigenetic responses to stree have been shown to be inherited and positively correlated with fitness. In animals, communal nesting changes mouse behavior increasing parental care regimes and social abilities that are hypothesized to increase offspring survival and access to resources (such as food and mates), respectively.

Macroevolutionary Patterns

Inherited epigenetic effects on phenotypes have been documented in bacteria, protists, fungi, plants, and animals. Though no systematic study of epigenetic inheritance has been conducted (most focus on model organisms), there is preliminary evidence that this mode of inheritance is more important in plants than in animals. The early differentiation of animal germlines is likely to preclude epigenetic marking occurring later in development, while in plants and fungi somatic cells may be incorporated into the germ line.

Life history patterns may also contribute to the occurrence of epigenetic inheritance. Sessile organisms, those with low dispersal capability, and those with simple behavior may benefit most from conveying information to their offspring via epigenetic pathways. Geographic patterns may also emerge, where highly variable and highly conserved environments might host fewer species with important epigenetic inheritance.

Chapter 3

Genomic Imprinting

Genomic imprinting is a genetic phenomenon by which certain genes are expressed in a parent-of-origin-specific manner. It is an inheritance process independent of the classical Mendelian inheritance. Imprinted genes are either expressed **only** from the allele inherited from the mother (e.g. *H19* or *CDKN1C*), or in other instances from the allele inherited from the father (e.g. *IGF-2*). Forms of genomic imprinting have been demonstrated in insects, mammals and flowering plants.

Genomic imprinting is an epigenetic process that involves methylation and histone modifications in order to achieve monoallelic gene expression without altering the genetic sequence. These epigenetic marks are established in the germline and are maintained throughout all somatic cells of an organism.

Appropriate expression of imprinted genes is important for normal development, with numerous genetic diseases associated with imprinting defects including Beckwith-Wiedemann syndrome, Silver-Russell Syndrome, Angelman Syndrome and Prader-Willi Syndrome.

Overview

In diploid organisms, somatic cells possess two copies of the genome. Each autosomal gene is therefore represented by two copies, or alleles, with one copy inherited from each parent at fertilisation. For the vast majority of autosomal genes, expression occurs from both alleles simultaneously. In mammals, however, a small proportion (<1%) of genes are imprinted, meaning that gene expression occurs from only one allele. The expressed allele is dependent upon its parental origin. For example, the gene encoding Insulin-like growth factor 2 (IGF2/Igf2) is only expressed from the allele inherited from the father.

The phrase "imprinting" was first used to describe events in the insect *Pseudococcus nipae*. In Pseudococcids or mealybugs (Homoptera, Coccoidea) both the male and female develop from a fertilised egg. In females, all chromosomes remain euchromatic and functional. In embryos destined to become males, one haploid set of chromosomes becomes heterochromatinised after the sixth cleavage division and remains so in most

tissues; males are thus functionally haploid. In insects, imprinting describes the silencing of the paternal genome in males, and thus is involved in sex determination. In mammals, genomic imprinting describes the processes involved in introducing functional inequality between two parental alleles of a gene.

Imprinted genes in mammals

That imprinting might be a feature of mammalian development was suggested in breeding experiments in mice carrying reciprocal translocations. Nucleus transplantation experiments in mouse zygotes in the early 1980s confirmed that normal development requires the contribution of both the maternal and paternal genomes. The vast majority of mouse parthenogenones/gynogenones (with two maternal or egg genomes) and androgenones (with two paternal or sperm genomes) die at, or before, the blastocyst/implantation stage. In the rare instances that they develop to postimplantation stages, gynogenetic embryos show better embryonic development relative to placental development, while for androgenones, the reverse is true. Nevertheless, for the latter, only a few have been described.

Parthenogenetic/gynogenetic embryos have twice the normal expression level of maternally derived genes, and lack expression of paternally expressed genes, while the reverse is true for androgenetic embryos. It is now known that there are at least 80 imprinted genes in humans and mice, many of which are involved in embryonic and placental growth and development. Various methods have been used to identify imprinted genes. In swine, Bischoff et al 2009 compared transcriptional profiles using short-oligonucleotide microarrays (Affymetrix Porcine GeneChip) to survey differentially expressed genes between parthenotes (2 maternal genomes) and control fetuses (1 maternal, 1 paternal genome) An intriguing study surveying the transcriptome of murine brain tissues revealed over 1300 imprinted gene loci (approximately 10-fold more than previously reported) by Illumina RNA-sequencing (RNA-Seq) technology from F1 hybrids resulting from reciprocal crosses.

No naturally occurring cases of parthenogenesis exist in mammals because of imprinted genes. Experimental manipulation of a paternal methylation imprint controlling the Igf2 gene has, however, recently allowed the creation of rare individual mice with two maternal sets of chromosomes - but this is not a true parthenogenone. Hybrid offspring of two species may exhibit unusual growth due to the novel combination of imprinted genes.

Genetic mapping of imprinted genes

At the same time as the generation of the gynogenetic and androgenetic embryos discussed above, mouse embryos were also being generated that contained only small regions that were derived from either a paternal or maternal source. The generation of a series of such uniparental disomies, which together span the entire genome, allowed the creation of an imprinting map. Those regions which when inherited from a single parent result in a discernible phenotype contain imprinted gene(s). Further research showed that within these regions there were often numerous imprinted genes. Around 80% of

imprinted genes are found in clusters such as these, called imprinted domains, suggesting a level of co-ordinated control. More recently, genome-wide screens to identify imprinted genes have used differential expression of mRNAs from control fetuses and parthenogenetic or androgenetic fetuses hybridized to expression arrays, allele-specific gene expression using SNP genotyping arrays, transcriptome sequencing, in silico prediction pipelines...to name a few.

Imprinting mechanisms

Imprinting is a dynamic process. It must be possible to erase and re-establish the imprint through each generation. The nature of the imprint must therefore be epigenetic (modifications to the structure of the DNA rather than the sequence). In germline cells the imprint is erased, and then re-established according to the sex of the individual; i.e. in the developing sperm (during spermatogenesis), a paternal imprint is established, whereas in developing oocytes (oogenesis), a maternal imprint is established. This process of erasure and reprogramming is necessary such that the current imprinting status is relevant to the sex of the individual. In both plants and mammals there are two major mechanisms that are involved in establishing the imprint; these are DNA methylation and histone modifications.

Regulation

The grouping of imprinted genes within clusters allows them to share common regulatory elements, such as non-coding RNAs and differentially methylated regions (DMRs). When these regulatory elements control the imprinting of one or more genes, they are known as imprinting control regions (ICR). The expression of non-coding RNAs, such as *Air* on mouse chromosome 17 and KCNQ1OT1 on human chromosome 11p15.5, have been shown to be essential for the imprinting of genes in their corresponding regions.

Differentially methylated regions are generally segments of DNA rich in cytosine and guanine nucleotides, with the cytosine nucleotides methylated on one copy but not on the other. Contrary to expectation, methylation does not necessarily mean silencing; instead, the effect of methylation depends upon the default state of the region.

Functions of imprinted genes

The control of expression of specific genes by genomic imprinting is unique to therian mammals (placental mammals and marsupials) and flowering plants. Imprinting of whole chromosomes has been reported in mealybugs. and a fungus gnat (*Sciara*). It has also been established that X-chromosome inactivation occurs in an imprinted manner in the extra-embryonic tissues of mice, where it is always the paternal X-chromosome which is silenced.

The majority of imprinted genes in mammals have been found to have roles in the control of embryonic growth and development, including development of the placenta. Other

imprinted genes are involved in post-natal development, with roles affecting suckling and metabolism.

Theories on the origins of imprinting

A widely accepted hypothesis for the evolution of genomic imprinting is the "parental conflict hypothesis." Also known as the kinship theory of genomic imprinting, this hypothesis states that the inequality between parental genomes due to imprinting is a result of the differing interests of each parent in terms of the evolutionary fitness of their genes. The father's genes that encode for imprinting gain greater fitness through the success of the offspring, at the expense of the mother. The mother's evolutionary imperative is often to conserve resources for her own survival while providing sufficient nourishment to current and subsequent litters. Accordingly, paternally expressed genes tend to be growth promoting whereas maternally expressed genes tend to be growth limiting. In support of this hypothesis, genomic imprinting has been found in all placental mammals, where post-fertilisation offspring resource consumption at the expense of the mother is high; it has not been found in oviparous birds or monotremes (a class of oviparous mammals) where there is relatively little post-fertilisation resource transfer and therefore less parental conflict.

However, our understanding of the molecular mechanisms behind genomic imprinting show that it is the maternal genome that controls much of the imprinting of both its own and the paternally-derived genes in the zygote, making it difficult to explain why the maternal genes would willingly relinquish their dominance to that of the paternally-derived genes in light of the conflict hypothesis. Several other hypotheses that propose a coadaptive reason for the evolution of genomic imprinting have been proposed.

Others have approached their study of the origins of genomic imprinting from a different side, arguing that natural selection is operating on the role of epigenetic marks as machinery for homologous chromosome recognition during meiosis, rather than on their role in differential expression. This argument centers on the existence of epigenetic effects on chromosomes that do not directly affect gene expression, but do depend on which parent the chromosome originated from. This group of epigenetic changes that depend on the chromosome's parent of origin (including both those that affect gene expression and those that do not) are called parental origin effects, and include phenomena such as paternal X inactivation in the marsupials, nonrandom parental chromatid distribution in the ferns, and even mating type switching in yeast. This diversity in organisms that show parental origin effects has prompted theorists to place the evolutionary origin of genomic imprinting before the last common ancestor of plants and animals, over a billion years ago.

Natural selection for genomic imprinting requires genetic variation in a population. A hypothesis for the origin of this genetic variation states that the host-defense system responsible for silencing foreign DNA elements, such as genes of viral origin, mistakenly silenced genes whose silencing turned out to be beneficial for the organism. There appears to be an over-representation of retrotransposed genes, that is to say genes that are

inserted into the genome by viruses, among imprinted genes. It has also been postulated that if the retrotransposed gene is inserted close to another imprinted gene, it may just acquire this imprint.

Problems associated with imprinting

Imprinting may cause problems in cloning, with clones having DNA that is not methylated in the correct position. It is possible that this is due to a lack of time for reprogramming to be completely achieved. When a nucleus is added to an egg during somatic cell nuclear transfer, the egg starts dividing in minutes, as compared to the days or months it takes for reprogramming during embryonic development. If time is the responsible factor, it may be possible to delay cell division in clones, giving time for proper reprogramming to occur.

An allele of the "callipyge" (from the Greek for "beautiful buttocks"), or CLPG, gene in sheep produces large buttocks consisting of muscle with very little fat. The large-buttocked phenotype only occurs when the allele is present on the copy of chromosome 18 inherited from a sheep's father and is *not* on the copy of chromosome 18 inherited from that sheep's mother.

Examples

Prader-Willi/Angelman

The first imprinted genetic disorders to be described in humans were the reciprocally inherited Prader-Willi syndrome and Angelman syndrome. Both syndromes are associated with loss of the chromosomal region 15q11-13 (band 11 of the long arm of chromosome 15). This region contains the paternally expressed genes (SNRPN and NDN) and the maternally expressed gene (UBE3A).

- Paternal inheritance of a deletion of this region is associated with Prader-Willi syndrome (characterised by hypotonia, obesity, and hypogonadism).
- Maternal inheritance of the same deletion is associated with Angelman syndrome (characterised by epilepsy, tremors, and a perpetually smiling facial expression).

NOEY2

NOEY2 is a paternally expressed imprinted gene located on chromosome 1 in humans. Loss of NOEY2 expression is linked to an increased risk of ovarian and breast cancers; in 41% of breast and ovarian cancers the protein transcribed by NOEY2 is not expressed, suggesting that it functions as a tumor suppressor gene Therefore, if a person inherits both chromosomes from the mother, the gene will not be expressed and the individual is put at a greater risk for breast and ovarian cancer.

Other

Other conditions involving imprinting include Beckwith-Wiedemann syndrome, Silver-Russell syndrome, and pseudohypoparathyroidism.

Transient neonatal diabetes mellitus can also involve imprinting.

Imprinted genes in plants

A similar imprinting phenomenon has also been described in flowering plants (angiosperms). During fertilisation of the egg cell, a second, separate fertilization event gives rise to the endosperm, an extraembryonic structure that nourishes the embryo in a manner analogous to the mammalian placenta. Unlike the embryo, the endosperm is often formed from the fusion of two maternal cells with a male gamete. This results in a triploid genome. The uneven ratio of maternal to paternal genomes appears to be critical for seed development. Some genes are found to be expressed from both maternal genomes while others are expressed exclusively from the lone paternal copy.

Chapter 4

Methylated DNA Immunoprecipitation

Methylated DNA immunoprecipitation (MeDIP or mDIP) is a large-scale (chromosome- or genome-wide) technique that is used to enrich for methylated DNA sequences. It consists of isolating methylated DNA fragments via an antibody raised against 5-methylcytosine (5mC). This technique was first described by Weber M. *et al.* and has helped pave the way for viable methylome-level assessment efforts, as the purified fraction of methylated DNA can be input to high-throughput DNA detection methods such as high-resolution DNA microarrays (MeDIP-chip) or next-generation sequencing (MeDIP-seq). Nonetheless, understanding of the methylome remains rudimentary; its study is complicated by the fact that, like other epigenetic properties, patterns vary from cell-type to cell-type.

Background

DNA methylation, referring to the reversible methylation of the 5' position of cytosine by methyltransferases, is a major epigenetic modification in multicellular organisms. In mammals, this modification primarily occurs at CpG sites, which in turn tend to cluster in regions called CpG islands. There is a small fraction of CpG islands that can overlap or be in close proximity to promoter regions of transcription start sites. The modification may also occur at other sites, but methylation at either of these sites can repress gene expression by either interfering with the binding of transcription factors or modifying chromatin structure to a repressive state.

Disease condition studies have largely fueled the effort in understanding the role of DNA methylation. Currently, the major research interest lies in investigating disease conditions such as cancer to identify regions of the DNA that has undergone extensive methylation changes. The genes contained in these regions are of functional interest as they may offer a mechanistic explanation to the underlying genetic causes of a disease. For instance, the abnormal methylation pattern of cancer cells was initially shown to be mechanism through which tumor suppressor-like genes are silenced, although it was later observed that a much broader range of gene types are affected.

Other technologies

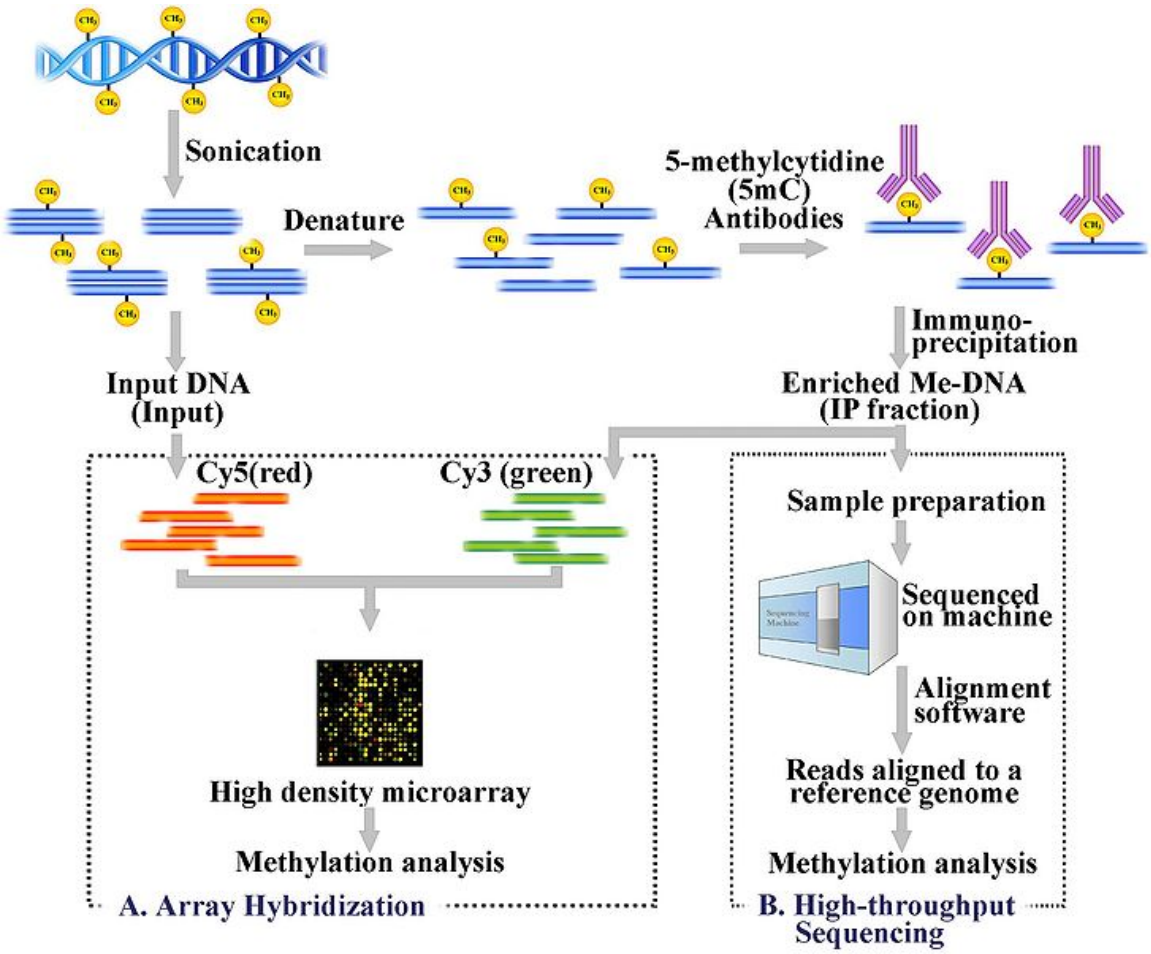
There are two approaches to methylation analysis: typing and profiling technologies. Typing technologies are targeted towards a small number of loci across many samples, and involve the use of techniques such as PCR, restriction enzymes, and mass spectrometry. Profiling technologies such as MeDIP are targeted towards a genome- or methylome-wide level assessment of methylation; this includes restriction landmark genomic scanning (RLGS), and bisulfite conversion-based methods, which rely on the treatment of DNA with bisulfite to convert unmethylated cytosine residues to uracil.

Limitations of other technologies

Other methods mapping and profiling the methylome have been effective but are not without their limitations that can affect resolution, level of throughput, or experimental variations. For instance, RLGS is limited by the number of restriction sites in genome that can be targets for the restriction enzyme; typically, a maximum of ~4100 landmarks can be assessed. Bisulfite sequencing-based methods, despite possible single-nucleotide resolution, have a drawback: the conversion of unmethylated cytosine to uracil can be unstable. In addition, when bisulfite conversion is coupled with DNA microarrays to detect bisulfite converted sites, the reduced sequence complexity of DNA is a problem. Microarrays capable of comprehensively profiling the whole-genome become difficult to design as fewer unique probes are available.

Methods

The following sections outline the method of MeDIP coupled with either high-resolution array hybridization or high-throughput sequencing. Each DNA detection method will also briefly describe post-laboratory processing and analysis. Different post-processing of the raw data is required depending on the technology used to identify the methylated sequences. This is analogous to data generated using ChIP-chip and ChIP-seq.

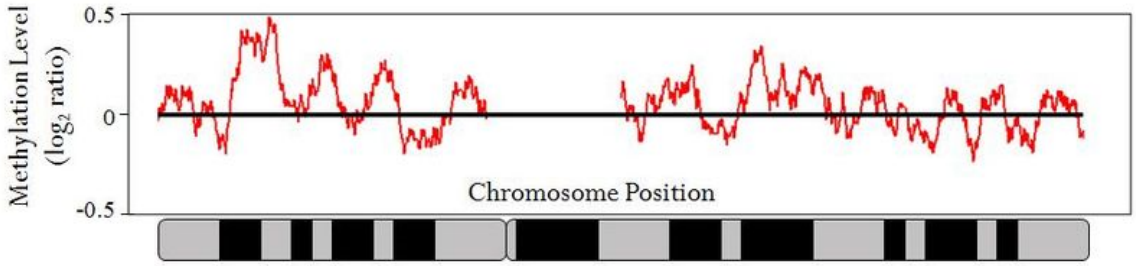


Workflow overview of the MeDIP procedure. MeDIP procedure is followed by array-hybridization (A) or high-throughput/next generation sequencing (B)

Methylated DNA immunoprecipitation (MeDIP)

Genomic DNA is extracted (DNA extraction) from the cells and purified. The purified DNA is then subjected to sonication to shear it into random fragments. This sonication process is quick, simple, and avoids restriction enzyme biases. The resulting fragments range from 300 to 1000 base pairs (bp) in length, although they are typically between 400 and 600 bp. The short length of these fragments are important in obtaining adequate resolution, improving the efficiency of the downstream step in immunoprecipitation, and reducing fragment-length effects or biases. Also, the size of the fragment affects the binding of 5-methyl-cytidine (5mC) antibody because the antibody needs more than just a single 5mC for efficient binding. To further improve binding affinity of the antibodies, the DNA fragments are denatured to produce single-stranded DNA. Following denaturation, the DNA is incubated with monoclonal 5mC antibodies. The classical immunoprecipitation technique is then applied: magnetic beads conjugated to anti-mouse-IgG are used to bind the anti-5mC antibodies, and unbound DNA is removed in the supernatant. To purify the DNA, proteinase K is added to digest the antibodies and release the DNA, which can be collected and prepared for DNA detection.

MeDIP and array-based hybridization (MeDIP-chip)



Simulated data to visualize typical analysis using MeDIP-chip.

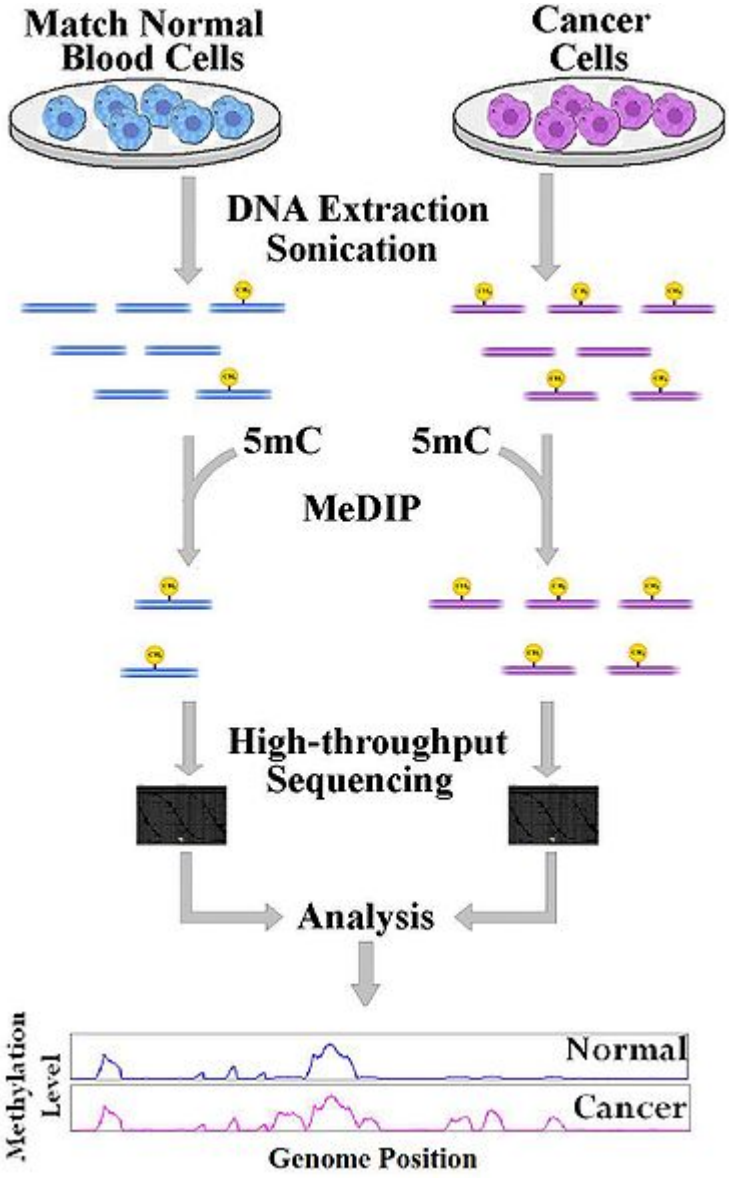
A fraction of the input DNA obtained after the sonication step above is labeled with cyanine-5 (Cy5; red) deoxy-cytosine-triphosphate while the methylated DNA, enriched after the immunoprecipitation step, is labeled with cyanine-3 (Cy3; green). The labeled DNA samples are cohybridized on a 2-channel, high-density genomic microarray to probe for presence and relative quantities. The purpose of this comparison is identify sequences that show significant differential expression, thereby confirming the sequence of interest is enriched. Array-based identification of MeDIP sequences are limited to the array design. As a result, the resolution is restricted to the probes in the array design. There are additional standard steps required in signal processing to correct for hybridization issues such as noise, as is the case with most array technologies.

MeDIP and high-throughput sequencing (MeDIP-seq)

The MeDIP-seq approach, i.e. the coupling of MeDIP with next generation, short-read sequencing technologies such as 454, Illumina (company) (Solexa), and SoLiD (Applied Biosystems), was first described by Down *et al.* in 2008. The high-throughput sequencing of the methylated DNA fragments produces a large number of short reads (36-50bp or 400 bp, depending on the technology). The short reads are aligned to a reference genome using alignment software such as Mapping and Assembly with Quality (Maq), which uses a Bayesian approach, along with base and mapping qualities to model error probabilities for the alignments. The reads can then be extended to represent the ~400 to 700 bp fragments from the sonication step. The coverage of these extended reads can be used to estimate the methylation level of the region. A genome browser such as Ensembl can also be used to visualize the data.

Validation of the approach to assess quality and accuracy of the data can be done with quantitative PCR. This is done by comparing a sequence from the MeDIP sample against an unmethylated control sequence. The samples are then run on a gel and the band intensities are compared. The relative intensity serves as the guide for finding enrichment. The results can also be compared with MeDIP-chip results to help determine coverage needed.

Downstream bioinformatics analysis



Simple workflow demonstrating a typical experiment using MeDIP-seq.

The DNA methylation level estimations can be confounded by varying densities of methylated CpG sites across the genome when observing data generated by MeDIP. This can be problematic for analyzing CpG-poor (lower density) regions. One reason for this density issue is its effect on the efficiency of immunoprecipitation. In their study, Down *et al.* developed a tool to estimate absolute methylation levels from data generated by MeDIP by modeling the density of methylated CpG sites. This tool is called Bayesian tool for methylation analysis (Batman). The study reports the coverage of ~90% of all CpG sites in promoters, gene-coding regions, islands, and regulatory elements where methylation levels can be estimated; this is almost 20 times better coverage than any previous methods.

Studies using MeDIP-seq or MeDIP-chip are both genome-wide approaches that have the common aim of obtaining the functional mapping of the methylome. Once regions of DNA methylation are identified, a number of bioinformatics analyses can be applied to answer certain biological questions. One obvious step is to investigate genes contained in these regions and investigate the functional significance of their repression. For example, silencing of tumour-suppressor genes in cancer can be attributed to DNA methylation. By identifying mutational events leading to hypermethylation and subsequent repression of known tumour-suppressor genes, one can more specifically characterize the contributing factors to the cause of the disease. Alternatively, one can identify genes that are known to be normally methylated but, as a result of some mutation event, is no longer silenced.

Also, one can try and investigate and identify whether some epigenetic regulator has been affected such as DNA methyltransferase (DNMT); in these cases, enrichment may be more limited.

Limitations of MeDIP

Limitations to take note when using MeDIP are typical experimental factors. This includes the quality and cross-reactivity of 5mC antibodies used in the procedure. Furthermore, DNA detection methods (i.e. array hybridization and high-throughput sequencing) typically involve well established limitations. Particularly for array-based procedures, as mentioned above, sequences being analyzed are limited to the specific array design used.

Most typical limitations to high-throughput, next generation sequencing apply. The problem of alignment accuracy to repetitive regions in the genome will result in less accurate analysis of methylation in those regions. Also, as was mentioned above, short reads (e.g. 36-50bp from an Illumina Genome Analyzer) represent a part of a sheared fragment when aligned to the genome; therefore, the exact methylation site can fall anywhere within a window that is a function of the fragment size. In this respect, bisulfite sequencing has much higher resolution (down to a single CpG site; single nucleotide level). However, this level of resolution may not be required for most applications, as the methylation status of CpG sites within < 1000 bp has been shown to be significantly correlated.

Applications of MeDIP

- Weber *et al.* 2005 determined that the inactive X-chromosome in females is hypermethylated on a chromosome wide level using MeDIP coupled with microarray.

- Keshet *et al.* 2006 performed a study on colon and prostate cancer cells using MeDIP-chip. The result is a genome-wide analysis of genes lying in hypermethylated regions as well as conclude that there is an instructive mechanism of de novo methylation in cancer cells.

- Zhang *et al.* 2006 obtained a high resolution methylome mapping in Arabidopsis using MeDIP-chip.
- Novak *et al.* 2006 used the MeDIP-chip approach to investigate human breast cancer for methylation associated silencing and observed the inactivation of the HOXA gene cluster

Chapter 5

Bisulfite Sequencing

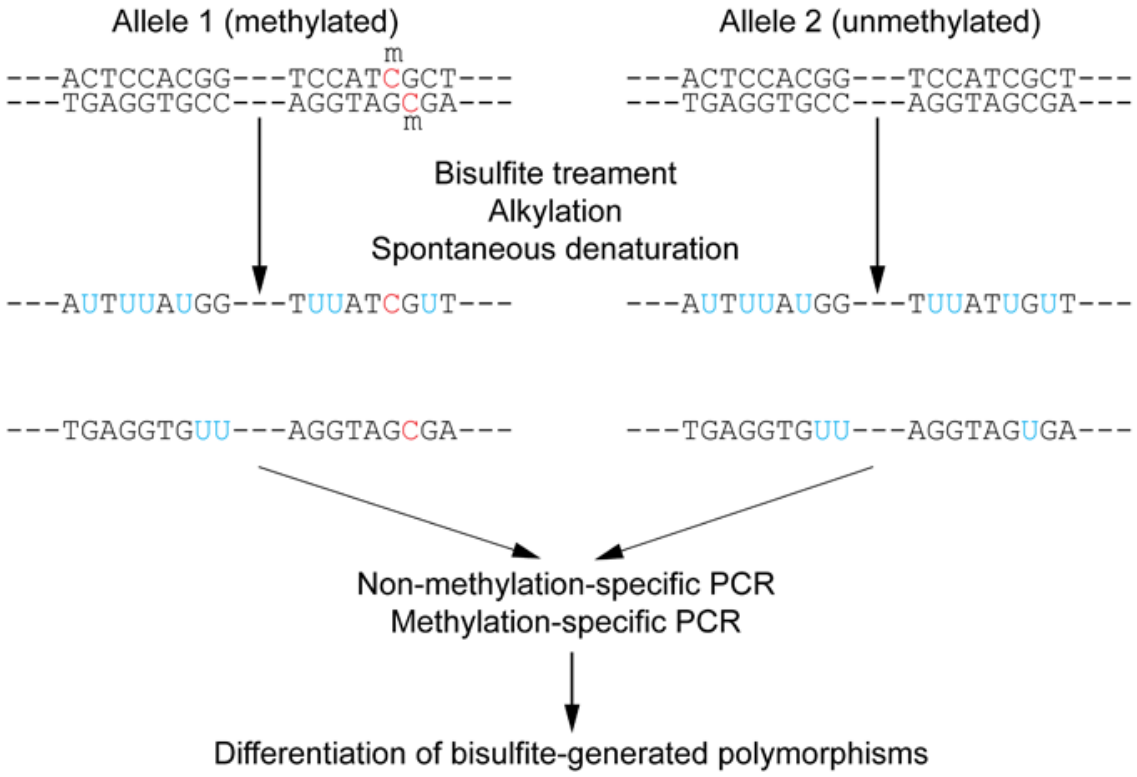


Figure 1: Outline of bisulfite conversion of sample sequence of genomic DNA. Nucleotides in blue are unmethylated cytosines converted to uracils by bisulfite, while red nucleotides are 5-methylcytosines resistant to conversion.

Bisulfite sequencing is the use of bisulfite treatment of DNA to determine its pattern of methylation. DNA methylation was the first discovered epigenetic mark, and remains the most studied. In animals it predominantly involves the addition of a methyl group to the carbon-5 position of cytosine residues of the dinucleotide CpG, and is implicated in repression of transcriptional activity.

Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Thus, bisulfite treatment introduces specific changes in the DNA sequence that depend on the methylation status of individual cytosine residues, yielding single- nucleotide resolution information about the methylation status of a segment of DNA. Various analyses can be performed on the altered sequence to retrieve this information. The objective of this analysis is therefore reduced to differentiating between single nucleotide polymorphisms (cytosines and thymidine) resulting from bisulfite conversion (Figure 1).

Methods

Bisulfite sequencing applies routine sequencing methods on bisulfite-treated genomic DNA to determine methylation status at CpG dinucleotides. Other non-sequencing strategies are also employed to interrogate the methylation at specific loci or at a genome-wide level. All strategies assume that bisulfite-induced conversion of unmethylated cytosines to uracil is complete, and this serves as the basis of all subsequent techniques. Ideally, the method used would determine the methylation status separately for each allele. Alternative methods to bisulfite sequencing include Combined Bisulfite Restriction Analysis and methylated DNA immunoprecipitation (MeDIP).

Methodologies to analyze bisulfite-treated DNA are continuously being developed. To summarize these rapidly evolving methologies, numerous review articles have been written.

The methodologies can be generally divided into strategies based on methylation-specific PCR (MSP) (Figure 3), and strategies employing polymerase chain reaction (PCR) performed under non-methylation-specific conditions (Figure 2). Microarray-based methods use PCR based on non-methylation-specific conditions also.

Non-methylation-specific PCR based methods

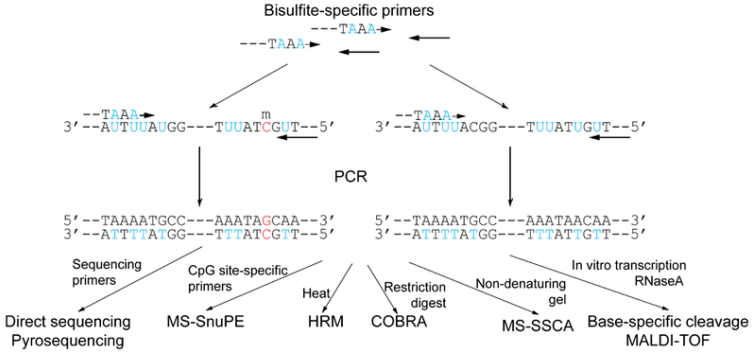


Figure 2: DNA methylation analysis methods not based on methylation-specific PCR. Following bisulfite conversion, the genomic DNA is amplified with PCR that does not discriminate between methylated and non-methylated sequences. The numerous methods available are then used to make the discrimination based on the changes within the amplicon as a result of bisulfite conversion.

Direct sequencing

The first reported method of methylation analysis using bisulfite-treated DNA utilized PCR and standard dideoxynucleotide DNA sequencing to directly determine the nucleotides resistant to bisulfite conversion. Primers are designed to be strand-specific as well as bisulfite-specific (i.e., primers containing non-CpG cytosines such that they are not complementary to non-bisulfite-treated DNA), flanking (but not involving) the methylation site of interest. Therefore, it will amplify both methylated and unmethylated sequences, in contrast to methylation-specific PCR. All sites of unmethylated cytosines are displayed as thymines in the resulting amplified sequence of the sense strand, and as adenines in the amplified antisense strand. This technique required cloning of the PCR product prior to sequencing for adequate sensitivity, and therefore was a very labour-intensive method unsuitable for higher throughput. Alternatively, nested PCR methods can be used to enhance the product for sequencing.

All subsequent DNA methylation analysis techniques using bisulfite-treated DNA is based on this report by Frommer et al. (Figure 2). Although most other modalities are not true sequencing-based techniques, the term "bisulfite sequencing" is often used to describe bisulfite-conversion DNA methylation analysis techniques in general.

Pyrosequencing

Pyrosequencing has also been used to analyze bisulfite-treated DNA without using methylation-specific PCR. Following PCR amplification of the region of interest, Pyrosequencing is used to determine the bisulfite-converted sequence of specific CpG sites in the region. The ratio of C-to-T at individual sites can be determined quantitatively based on the amount of C and T incorporation during the sequence extension. The main limitation of this method is the cost of the technology. However, Pyrosequencing does well allow for extension to high-throughput screening methods.

A further improvement to this technique was recently described by Wong et al., which uses allele-specific primers that incorporate single-nucleotide polymorphisms into the sequence of the sequencing primer, thus allowing for separate analysis of maternal and paternal alleles. This technique is of particular usefulness for genomic imprinting analysis.

Methylation-sensitive single-strand conformation analysis (MS-SSCA)

This method is based on the single-strand conformation polymorphism analysis (SSCA) method developed for single-nucleotide polymorphism (SNP) analysis. SSCA differentiates between single-stranded DNA fragments of identical size but distinct sequence based on differential migration in non-denaturating electrophoresis. In MS-SSCA, this is used to distinguish between bisulfite-treated, PCR-amplified regions containing the CpG sites of interest. Although SSCA lacks sensitivity when only a single nucleotide difference is present, bisulfite treatment frequently makes a number of C-to-T conversions in most regions of interest, and the resulting sensitivity approaches 100%.

MS-SSCA also provides semi-quantitative analysis of the degree of DNA methylation based on the ratio of band intensities. However, this method is designed to assess all CpG sites as a whole in the region of interest rather than individual methylation sites.

High resolution melting analysis (HRM)

A further method to differentiate converted from unconverted bisulfite-treated DNA is using high-resolution melting analysis (HRM), a real-time PCR-based technique initially designed to distinguish SNPs. The PCR amplicons are analyzed directly by temperature ramping and resulting liberation of an intercalating fluorescent dye during melting. The degree of methylation, as represented by the C-to-T content in the amplicon, determines the rapidity of melting and consequent release of the dye. This method allows direct quantitation in a single-tube assay, but assesses methylation in the amplified region as a whole rather than at specific CpG sites.

Methylation-sensitive single-nucleotide primer extension (MS-SnuPE)

MS-SnuPE employs the primer extension method initially designed for analyzing single-nucleotide polymorphisms. DNA is bisulfite-converted, and bisulfite-specific primers are annealed to the sequence up to the base pair immediately before the CpG of interest. The primer is allowed to extend one base pair into the C (or T) using DNA polymerase terminating dideoxynucleotides, and the ratio of C to T is determined quantitatively.

A number of methods can be used to determine this C:T ratio. At the beginning, MS-SnuPE relied on radioactive ddNTPs as the reporter of the primer extension. Fluorescence-based methods or Pyrosequencing can also be used. However, matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF) mass spectrometry analysis to differentiate between the two polymorphic primer extension products can be used, in essence, based on the GOOD assay designed for SNP genotyping. Ion pair reverse-phase high-performance liquid chromatography (IP-RP-HPLC) has also been used to distinguish primer extension products.

Base-specific cleavage/MALDI-TOF

A recently described method by Ehrich et al. further takes advantage of bisulfite-conversions by adding a base-specific cleavage step to enhance the information gained from the nucleotide changes. By first using in vitro transcription of the region of interest into RNA (by adding an RNA polymerase promoter site to the PCR primer in the initial amplification), RNase A can be used to cleave the RNA transcript at base-specific sites. As RNase A cleaves RNA specifically at cytosine and uracil ribonucleotides, base-specificity is achieved by adding incorporating cleavage-resistant dTTP when cytosine-specific (C-specific) cleavage is desired, and incorporating dCTP when uracil-specific (U-specific) cleavage is desired. The cleaved fragments can then be analyzed by MALDI-TOF. Bisulfite treatment results in either introduction/removal of cleavage sites by C-to-U conversions or shift in fragment mass by G-to-A conversions in the amplified reverse strand. C-specific cleavage will cut specifically at all methylated CpG sites. By analyzing

the sizes of the resulting fragments, it is possible to determine the specific pattern of DNA methylation of CpG sites within the region, rather than determining the extent of methylation of the region as a whole. This method demonstrated efficacy for high-throughput screening, allowing for interrogation of numerous CpG sites in multiple tissues in a cost-efficient manner.

Methylation-specific PCR (MSP)

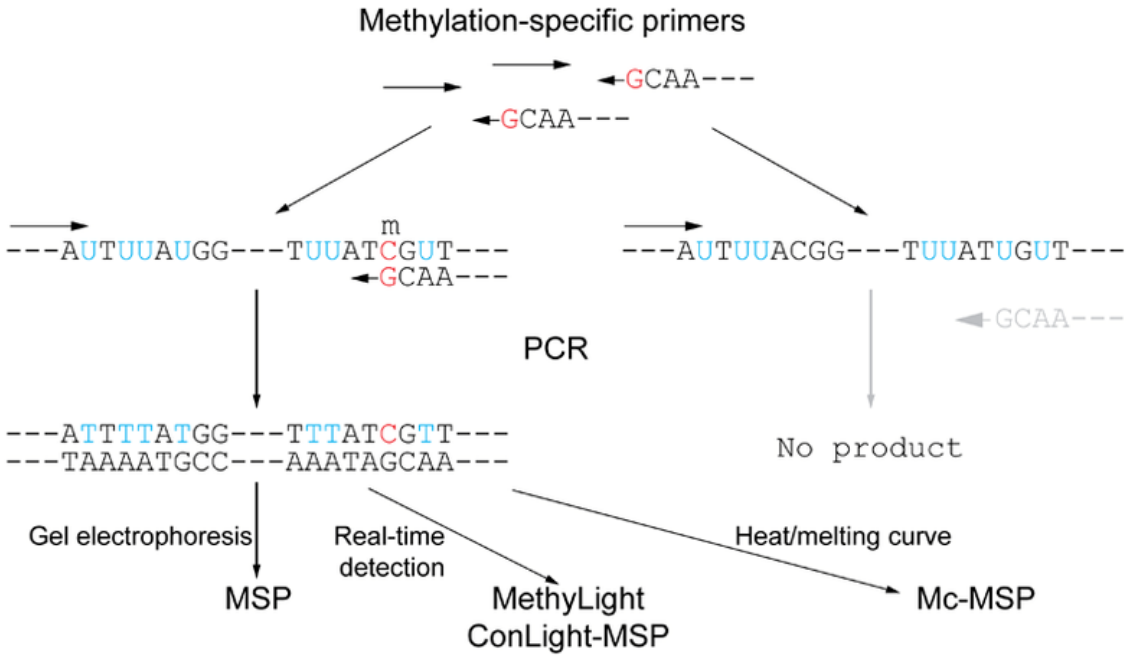


Figure 3: Methylation-specific PCR is a sensitive method to discriminately amplify and detect a methylated region of interest using methylated-specific primers on bisulfite-converted genomic DNA. Such primers will anneal only to sequences that are methylated, and thus containing 5-methylcytosines that are resistant to conversion by bisulfite. In alternative fashion, unmethylated-specific primers can be used.

This alternative method of methylation analysis also uses bisulfite-treated DNA but avoids the need to sequence the area of interest. Instead, primer pairs are designed themselves to be "methylated-specific" by including sequences complementing only unconverted 5-methylcytosines, or, on the converse, "unmethylated-specific", complementing thymines converted from unmethylated cytosines. Methylation is determined by the ability of the specific primer to achieve amplification. This method is particularly useful to interrogate CpG islands with possibly high methylation density, as increased numbers of CpG pairs in the primer increase the specificity of the assay. Placing the CpG pair at the 3'-end of the primer also improves the sensitivity. The initial report using MSP described sufficient sensitivity to detect methylation of 0.1% of alleles. In general, MSP and its related protocols are considered to be the most sensitive when interrogating the methylation status at a specific locus.

The MethyLight method is based on MSP, but provides a quantitative analysis using real-time PCR. Methylated-specific primers are used, and a methylated-specific fluorescence reporter probe is also used that anneals to the amplified region. In alternative fashion, the primers or probe can be designed without methylation specificity if discrimination is needed between the CpG pairs within the involved sequences. Quantitation is made in reference to a methylated reference DNA. A modification to this protocol to increase the specificity of the PCR for successfully bisulfite-converted DNA (ConLight-MSP) uses an additional probe to bisulfite-unconverted DNA to quantify this non-specific amplification.

Further methodology using MSP-amplified DNA analyzes the products using melting-curve analysis (Mc-MSP). This method amplifies bisulfite-converted DNA with both methylated-specific and unmethylated-specific primers, and determines the quantitative ratio of the two products by comparing the differential peaks generated in a melting-curve analysis. A high-resolution melting analysis method that uses both real-time quantification and melting analysis has been introduced, in particular, for sensitive detection of low-level methylation

Microarray-based methods

Microarray-based methods are a logical extension of the technologies available to analyze bisulfite-treated DNA to allow for genome-wide analysis of methylation. Oligonucleotide microarrays are designed using oligonucleotide pairs targeting CpG sites of interest, with one complementary to the unaltered methylated sequence, and the other to the C-to-U-converted unmethylated sequence. The oligonucleotides are also bisulfite-specific to prevent binding to DNA incompletely converted by bisulfite. The Illumina Methylation Assay is one such assay that applies the bisulfite sequencing technology on a microarray level to generate genome-wide methylation data.

Limitations

Incomplete conversion

Bisulfite sequencing relies on the conversion of every single unmethylated cytosine residue to uracil. If conversion is incomplete, the subsequent analysis will incorrectly interpret the unconverted unmethylated cytosines as methylated cytosines, resulting in false positive results for methylation. Only cytosines in single-stranded DNA are susceptible to attack by bisulfite, therefore denaturation of the DNA undergoing analysis is critical. It is important to ensure that reaction parameters such as temperature and salt concentration are suitable to maintain the DNA in a single-stranded conformation and allow for complete conversion. Embedding the DNA in agarose gel has been reported to improve the rate of conversion by keeping strands of DNA physically separate.

Degradation of DNA during bisulfite treatment

A major challenge in bisulfite sequencing is the degradation of DNA that takes place concurrently with the conversion. The conditions necessary for complete conversion, such as long incubation times, elevated temperature, and high bisulfite concentration, can lead to the degradation of about 90% of the incubated DNA. Given that the starting amount of DNA is often limited, such extensive degradation can be problematic. The degradation occurs as depurinations resulting in random strand breaks. Therefore the longer the desired PCR amplicon, the more limited the number of intact template molecules will likely be. This could lead to the failure of the PCR amplification, or the loss of quantitatively accurate information on methylation levels resulting from the limited sampling of template molecules. Thus, it is important to assess the amount of DNA degradation resulting from the reaction conditions employed, and consider how this will affect the desired amplicon. Techniques can also be used to minimize DNA degradation, such as cycling the incubation temperature.

Other concerns

A potentially significant problem following bisulfite treatment is incomplete desulfonation of pyrimidine residues due to inadequate alkalization of the solution. This may inhibit some DNA polymerases, rendering subsequent PCR difficult. However, this situation can be avoided by monitoring the pH of the solution to ensure that desulphonation will be complete.

A final concern is that bisulfite treatment greatly reduces the level of complexity in the sample, which can be problematic if multiple PCR reactions are to be performed (2006). Primer design is more difficult, and inappropriate cross-hybridization is more frequent.

Applications: genome-wide methylation analysis

The advances in bisulfite sequencing have led to the possibility of applying them at a genome-wide scale, where, previously, global measure of DNA methylation was feasible only using other techniques, such as Restriction landmark genomic scanning. The mapping of the human epigenome is seen by many scientists as the logical follow-up to the completion of the Human Genome Project. This epigenomic information will be important in understanding how the function of the genetic sequence is implemented and regulated. Since the epigenome is less stable than the genome, it is thought to be important in gene-environment interactions.

Epigenomic mapping is inherently more complex than genome sequencing, however, since the epigenome is much more variable than the genome. While an individual has only one genome, one's epigenome varies with age, differs between tissues, is altered by environmental factors, and shows aberrations in diseases. Such rich epigenomic mapping, however, representing different ages, tissue types, and disease states, would yield valuable information on the normal function of epigenetic marks as well as the mechanisms leading to aging and disease.

Direct benefits of epigenomic mapping include probable advances in cloning technology. It is believed that failures to produce cloned animals with normal viability and lifespan result from inappropriate patterns of epigenetic marks. Also, aberrant methylation patterns are well characterized in many cancers. Global hypomethylation results in decreased genomic stability, while local hypermethylation of tumour suppressor gene promoters often accounts for their loss of function. Specific patterns of methylation are indicative of specific cancer types, have prognostic value, and can help to guide the best course of treatment.

Large-scale epigenome mapping efforts are under way around the world and have been organized under the Human Epigenome Project. This is based on a multi-tiered strategy, whereby bisulfite sequencing is used to obtain high-resolution methylation profiles for a limited number of reference epigenomes, while less thorough analysis is performed on a wider spectrum of samples. This approach is intended to maximize the insight gained from a given amount of resources, as high-resolution genome-wide mapping remains a costly undertaking.

Chapter 6

DNA Methylation

DNA methylation involves the addition of a methyl group to the 5 position of the cytosine pyrimidine ring or the number 6 nitrogen of the adenine purine ring (cytosine and adenine are two of the four bases of DNA). This modification can be inherited through cell division. DNA methylation is typically removed during zygote formation and re-established through successive cell divisions during development although the latest research shows that hydroxylation of methyl group occurs rather than complete removal of methyl groups in zygote . DNA methylation is a crucial part of normal organismal development and cellular differentiation in higher organisms. DNA methylation stably alters the gene expression pattern in cells such that cells can "remember where they have been" or decrease gene expression; for example, cells programmed to be pancreatic islets during embryonic development remain pancreatic islets throughout the life of the organism without continuing signals telling them that they need to remain islets. In addition, DNA methylation suppresses the expression of viral genes and other deleterious elements that have been incorporated into the genome of the host over time. DNA methylation also forms the basis of chromatin structure, which enables cells to form the myriad characteristics necessary for multicellular life from a single immutable sequence of DNA. DNA methylation also plays a crucial role in the development of nearly all types of cancer.

DNA methylation involves the addition of a methyl group to DNA — for example, to the number 5 carbon of the cytosine pyrimidine ring — in this case with the specific effect of reducing gene expression. DNA methylation at the 5 position of cytosine has been found in every vertebrate examined. In adult somatic tissues, DNA methylation typically occurs in a CpG dinucleotide context; non-CpG methylation is prevalent in embryonic stem cells.

In mammals

DNA methylation is essential for normal development and is associated with a number of key processes including genomic imprinting, X-chromosome inactivation, suppression of repetitive elements, and carcinogenesis.

Between 60% and 90% of all CpGs are methylated in mammals. Methylated C residues spontaneously deaminate to form T residues; hence CpG dinucleotides steadily mutate to TpG dinucleotides, which is evidenced by the under-representation of CpG dinucleotides in the human genome (they occur at only 21% of the expected frequency). (On the other hand, spontaneous deamination of unmethylated C residues gives rise to U residues, a mutation that is quickly recognized and repaired by the cell.)

Unmethylated CpGs are often grouped in clusters called *CpG islands*, which are present in the 5' regulatory regions of many genes. In many disease processes, such as cancer, gene promoter CpG islands acquire abnormal hypermethylation, which results in transcriptional silencing that can be inherited by daughter cells following cell division. Alterations of DNA methylation have been recognized as an important component of cancer development. Hypomethylation, in general, arises earlier and is linked to chromosomal instability and loss of imprinting, whereas hypermethylation is associated with promoters and can arise secondary to gene (oncogene suppressor) silencing, but might be a target for epigenetic therapy.

DNA methylation may affect the transcription of genes in two ways. First, the methylation of DNA itself may physically impede the binding of transcriptional proteins to the gene, and second, and likely more important, methylated DNA may be bound by proteins known as methyl-CpG-binding domain proteins (MBDs). MBD proteins then recruit additional proteins to the locus, such as histone deacetylases and other chromatin remodeling proteins that can modify histones, thereby forming compact, inactive chromatin, termed silent chromatin. This link between DNA methylation and chromatin structure is very important. In particular, loss of methyl-CpG-binding protein 2 (MeCP2) has been implicated in Rett syndrome; and methyl-CpG-binding domain protein 2 (MBD2) mediates the transcriptional silencing of hypermethylated genes in cancer.

Research has suggested that long-term memory storage in humans may be regulated by DNA methylation.

DNA methylation in cancer

DNA methylation is an important regulator of gene transcription and a large body of evidence has demonstrated that aberrant DNA methylation is associated with unscheduled gene silencing, and the genes with high levels of 5-methylcytosine in their promoter region are transcriptionally silent. DNA methylation is essential during embryonic development, and in somatic cells, patterns of DNA methylation are generally transmitted to daughter cells with a high fidelity. Aberrant DNA methylation patterns have been associated with a large number of human malignancies and found in two distinct forms: hypermethylation and hypomethylation compared to normal tissue. Hypermethylation is one of the major epigenetic modifications that repress transcription via promoter region of tumour suppressor genes. Hypermethylation typically occurs at CpG islands in the promoter region and is associated with gene inactivation. Global hypomethylation has also been implicated in the development and progression of cancer through different mechanisms.

DNA methyltransferases

In mammalian cells, DNA methylation occurs mainly at the C5 position of CpG dinucleotides and is carried out by two general classes of enzymatic activities – maintenance methylation and *de novo* methylation.

Maintenance methylation activity is necessary to preserve DNA methylation after every cellular DNA replication cycle. Without the DNA methyltransferase (DNMT), the replication machinery itself would produce daughter strands that are unmethylated and, over time, would lead to passive demethylation. DNMT1 is the proposed maintenance methyltransferase that is responsible for copying DNA methylation patterns to the daughter strands during DNA replication. Mouse models with both copies of DNMT1 deleted are embryonic lethal at approximately day 9, due to the requirement of DNMT1 activity for development in mammalian cells.

It is thought that DNMT3a and DNMT3b are the *de novo* methyltransferases that set up DNA methylation patterns early in development. DNMT3L is a protein that is homologous to the other DNMT3s but has no catalytic activity. Instead, DNMT3L assists the *de novo* methyltransferases by increasing their ability to bind to DNA and stimulating their activity. Finally, DNMT2 (TRDMT1) has been identified as a DNA methyltransferase homolog, containing all 10 sequence motifs common to all DNA methyltransferases; however, DNMT2 (TRDMT1) does not methylate DNA but instead methylates cytosine-38 in the anticodon loop of aspartic acid transfer RNA.

Since many tumor suppressor genes are silenced by DNA methylation during carcinogenesis, there have been attempts to re-express these genes by inhibiting the DNMTs. 5-Aza-2'-deoxycytidine (decitabine) is a nucleoside analog that inhibits DNMTs by trapping them in a covalent complex on DNA by preventing the β-elimination step of catalysis, thus resulting in the enzymes' degradation. However, for decitabine to be active, it must be incorporated into the genome of the cell, which can cause mutations in the daughter cells if the cell does not die. In addition, decitabine is toxic to the bone marrow, which limits the size of its therapeutic window. These pitfalls have led to the development of antisense RNA therapies that target the DNMTs by degrading their mRNAs and preventing their translation. However, it is currently unclear whether targeting DNMT1 alone is sufficient to reactivate tumor suppressor genes silenced by DNA methylation.

In plants

Significant progress has been made in understanding DNA methylation in the model plant *Arabidopsis thaliana*. DNA methylation in plants differs from that of mammals: while DNA methylation in mammals mainly occurs on the cytosine nucleotide in a CpG site, in plants the cytosine can be methylated at CpG, CpHpG, and CpHpH sites, where H represents any nucleotide but guanine.

The principal *Arabidopsis* DNA methyltransferase enzymes, which transfer and covalently attach methyl groups onto DNA, are DRM2, MET1, and CMT3. Both the DRM2 and MET1 proteins share significant homology to the mammalian methyltransferases DNMT3 and DNMT1, respectively, whereas the CMT3 protein is unique to the plant kingdom. There are currently two classes of DNA methyltransferases: 1) the *de novo* class, or enzymes that create new methylation marks on the DNA; and 2) a maintenance class that recognizes the methylation marks on the parental strand of DNA and transfers new methylation to the daughters strands after DNA replication. DRM2 is the only enzyme that has been implicated as a *de novo* DNA methyltransferase. DRM2 has also been shown, along with MET1 and CMT3 to be involved in maintaining methylation marks through DNA replication. Other DNA methyltransferases are expressed in plants but have no known function.

It is not clear how the cell determines the locations of *de novo* DNA methylation, but evidence suggests that, for many (though not all) locations, RNA-directed DNA methylation (RdDM) is involved. In RdDM, specific RNA transcripts are produced from a genomic DNA template, and this RNA forms secondary structures called double-stranded RNA molecules. The double-stranded RNAs, through either the small interfering RNA (siRNA) or microRNA (miRNA) pathways direct de-novo DNA methylation of the original genomic location that produced the RNA. This sort of mechanism is thought to be important in cellular defense against RNA viruses and/or transposons, both of which often form a double-stranded RNA that can be mutagenic to the host genome. By methylating their genomic locations, through an as yet poorly-understood mechanism, they are shut off and are no longer active in the cell, protecting the genome from their mutagenic effect.

In fungi

It can be seen that many fungi have low levels (0.1 to 0.5%) of cytosine methylation, whereas other fungi have as much as 5% of the genome methylated.

This value seems to vary both among species and among isolates of the same species. There is also evidence that DNA methylation may be involved in state-specific control of gene expression in fungi.

Although brewers' yeast (*Saccharomyces*) and fission yeast (*Schizosaccharomyces*) have very little DNA methylation, the model filamentous fungus *Neurospora crassa* has a well-characterized methylation system. Several genes control methylation in *Neurospora* and mutation of the DNA methyl transferase, *dim-2*, eliminates all DNA methylation but does not affect growth or sexual reproduction. While the *Neurospora* genome has very little repeated DNA, half of the methylation occurs in repeated DNA including transposon relics and centromeric DNA. The ability to evaluate other important phenomena in a DNA methylase-deficient genetic background makes *Neurospora* an important system in which to study DNA methylation.

In bacteria

Adenine or cytosine methylation is part of the restriction modification system of many bacteria, in which specific DNA sequences are methylated periodically throughout the genome. A methylase is the enzyme that recognizes a specific sequence and methylates one of the bases in or near that sequence. Foreign DNAs (which are not methylated in this manner) that are introduced into the cell are degraded by sequence-specific restriction enzymes and cleaved. Bacterial genomic DNA is not recognized by these restriction enzymes. The methylation of native DNA acts as a sort of primitive immune system, allowing the bacteria to protect themselves from infection by bacteriophage.

E. coli DNA adenine methyltransferase (Dam) is an enzyme of ~32 kDa that does not belong to a restriction/modification system. The target recognition sequence for *E. coli* Dam is GATC, as the methylation occurs at the N6 position of the adenine in this sequence (G meATC). The three base pairs flanking each side of this site also influence DNA–Dam binding. Dam plays several key roles in bacterial processes, including mismatch repair, the timing of DNA replication, and gene expression. As a result of DNA replication, the status of GATC sites in the *E. coli* genome changes from fully methylated to hemimethylated. This is because adenine introduced into the new DNA strand is unmethylated. Re-methylation occurs within two to four seconds, during which time replication errors in the new strand are repaired. Methylation, or its absence, is the marker that allows the repair apparatus of the cell to differentiate between the template and nascent strands. It has been shown that altering Dam activity in bacteria results in increased spontaneous mutation rate. Bacterial viability is compromised in dam mutants that also lack certain other DNA repair enzymes, providing further evidence for the role of Dam in DNA repair.

One region of the DNA that keeps its hemimethylated status for longer is the origin of replication, which has an abundance of GATC sites. This is central to the bacterial mechanism for timing DNA replication. SeqA binds to the origin of replication, sequestering it and thus preventing methylation. Because hemimethylated origins of replication are inactive, this mechanism limits DNA replication to once per cell cycle.

Expression of certain genes, for example those coding for pilus expression in *E. coli*, is regulated by the methylation of GATC sites in the promoter region of the gene operon. The cells' environmental conditions just after DNA replication determine whether Dam is blocked from methylating a region proximal to or distal from the promoter region. Once the pattern of methylation has been created, the pilus gene transcription is locked in the on or off position until the DNA is again replicated. In *E. coli*, these pilus operons have important roles in virulence in urinary tract infections. It has been proposed that inhibitors of Dam may function as antibiotics.

On the other hand DNA cytosine methylase targets CCAGG and CCTGG sites to methylate cytosine at the C5 position (C meC(A/T)GG). The other methylase enzyme, EcoKI, causes mehtylation of adenine in the sequences AAC(N6A)GTGC and GCAC(N6A)GTT.

Most strains used by molecular biologists are derivatives of K-12, and possess both Dam and Dcm, but there are commercially available strains which possess dam-/dcm- activity. In fact, it is possible to unmethylate the DNA extracted from dam+/dcm+ strains by transforming into dam-/dcm- strains. This would help digest sequences that are not being recognized by methylation-sensitive restriction enzymes.

Detection

DNA methylation can be detected by the following assays currently used in scientific research:

- Methylation-Specific PCR (MSP), which is based on a chemical reaction of sodium bisulfite with DNA that converts unmethylated cytosines of CpG dinucleotides to uracil or UpG, followed by traditional PCR. However, methylated cytosines will not be converted in this process, and primers are designed to overlap the CpG site of interest, which allows one to determine methylation status as methylated or unmethylated.
- The HELP assay, which is based on restriction enzymes' differential ability to recognize and cleave methylated and unmethylated CpG DNA sites.
- ChIP-on-chip assays, which is based on the ability of commercially prepared antibodies to bind to DNA methylation-associated proteins like MCP2.
- Restriction landmark genomic scanning, a complicated and now rarely-used assay based upon restriction enzymes' differential recognition of methylated and unmethylated CpG sites; the assay is similar in concept to the HELP assay.
- Methylated DNA immunoprecipitation (MeDIP), analogous to chromatin immunoprecipitation, immunoprecipitation is used to isolate methylated DNA fragments for input into DNA detection methods such as DNA microarrays (MeDIP-chip) or DNA sequencing (MeDIP-seq).
- Molecular break light assay for DNA adenine methyltransferase activity – an assay that relies on the specificity of the restriction enzyme DpnI for fully methylated (adenine methylation) GATC sites in an oligonucleotide labeled with a fluorophore and quencher. The adenine methyltransferase methylates the oligonucleotide making it a substrate for DpnI. Cutting of the oligonucleotide by DpnI gives rise to a fluorescence increase.

Chapter 7

Nutriepigenomics

Nutriepigenomics is the study of food nutrients and their effects on human health through epigenetic modifications. There is now considerable evidence that nutritional imbalances during gestation and lactation are linked to non-communicable diseases, such as obesity, cardiovascular disease, diabetes, hypertension, and cancer. If metabolic disturbances occur during critical time windows of development, the resulting epigenetic alterations can lead to permanent changes in tissue and organ structure or function and predispose individuals to disease

Overview

Epigenetics relates to heritable changes in gene function that occur independently of alterations in primary DNA sequence. Two major epigenetic mechanisms implicated in nutriepigenomics are DNA methylation and histone modification. DNA methylation in gene promoter regions usually results in gene silencing and influences gene expression. While this form of gene silencing is extremely important in development and cellular differentiation, aberrant DNA methylation can be detrimental and has been linked to various disease processes, such as cancer . The methyl groups used in DNA methylation are often derived from dietary sources, such as folate and choline, and explains why diet can have a significant impact on methylation patterns and gene expression . Gene silencing can also be reinforced through the recruitment of histone deacetylases to decrease transcriptional activation. Conversely, histone acetylation induces transcriptional activation to increase gene expression. Dietary components can influence these epigenetic events, thereby altering gene expression and disturbing functions such as appetite control, metabolic balance and fuel utilization .

Various genetic sequences can be targeted for epigenetic modification. A transcriptome-wide analysis in mice found that a protein-restricted (PR) diet during gestation resulted in differential gene expression in approximately 1% of the fetal genes analyzed (235/22,690). Specifically, increased expression was seen in genes involved in the p53 pathway, apoptosis, negative regulators of cell metabolism, and genes related to epigenetic control . Additional studies have investigated the effect of a PR-diet in rats and

found changes in promoter methylation of both the glucocorticoid receptor and peroxisome proliferator-activated receptor (PPAR) . Altered expression of these receptors can result in elevated blood glucose levels and affect lipid and carbohydrate metabolism . Feeding a PR-diet to pregnant and/or lactating mice also increased expression of glucokinase, acetyl-CoA carboxylase, PPARα, and acyl-CoA oxidase . Changes in expression were reportedly due to epigenetic regulation of either the gene promoter itself, or promoters of transcription factors that regulate gene expression. Additional genes that have been shown, either by in vitro or in vivo studies, to be regulated by epigenetic mechanisms include leptin, SOCS3, glucose transporter (GLUT)-4, POMC, 11-β-hydroxysteroid dehydrogenase type 2 and corticotrophin releasing hormone. Epigenetic modification of these genes may lead to "metabolic programming" of the fetus and result in long-term changes in metabolism and energy homeostasis .

Nutriepigenomics and Development

The period of development in which the nutritional imbalance occurs is very important in determining which disease-related genes will be affected. Different organs have critical developmental stages, and the time point at which they are compromised will predispose individuals to specific diseases . Epigenetic modifications that occur during development may not be expressed until later in life depending on the function of the gene . While the majority of studies implicate prenatal and perinatal periods as critical time windows, some research has shown that nutritional intake during adulthood can also affect the epigenome.

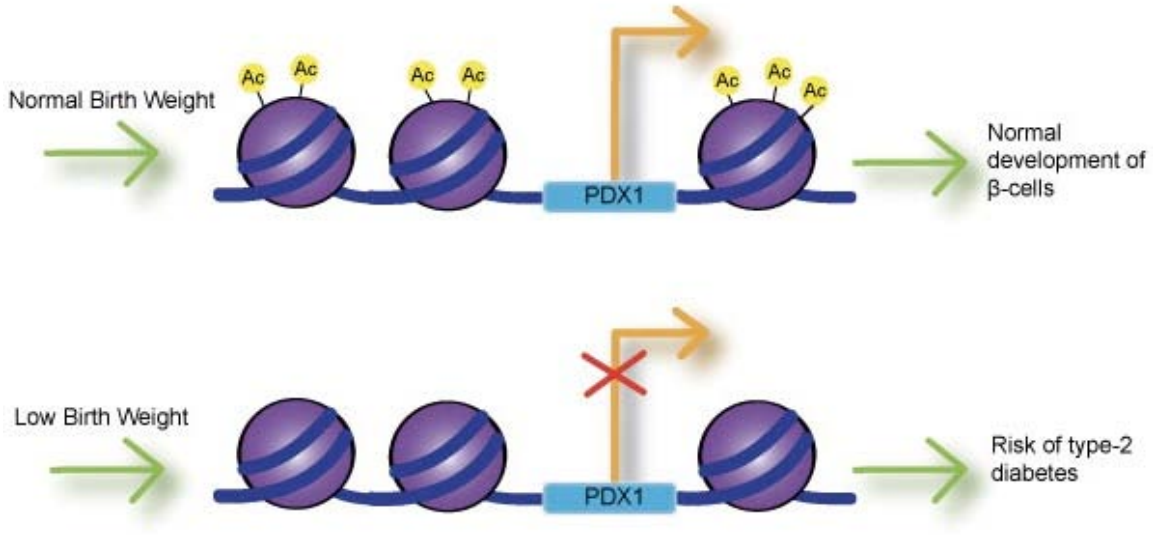
Prenatal

Developmental plasticity is a term used to describe the process in which fetuses adapt to their environment. Environmental cues, including dietary components, present in the in utero environment can induce significant changes in the expression of the genome through epigenetic modifications . Fetal developmental plastic responses can cause changes in lean body mass, endocrinology, blood flow and vascular loading, and lead to increased risk of various diseases in adulthood.

Low Birth Weight

Fetal exposure to calcium, folate, magnesium, high or low protein, and zinc have all been associated with birth weight. Numerous studies have investigated the link between birth weight and risk of disease and have found that low birth weight is significantly associated with coronary heart disease, stroke and type-2 diabetes. Most importantly, these associations occurred after adjusting for lifestyle factors, implying a genetic basis for onset of disease . Impaired insulin secretion is associated with low birth weight and can lead to insulin resistance as babies accumulate body fat . Studies using intrauterine growth retarded (IUGR) rats have found that growth inhibition can lead to decreased expression of Pdx1 transcription factor, which is essential for differentiation and function of pancreatic beta cells . Decreased histone acetylation at the proximal promoter of Pdx1

is responsible for reduced Pdx1 expression and subsequently results in a cascade of histone deacetylation and methylation events that can result in type-2 diabetes.



Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1

Obesity

Obesity during pregnancy and high-fat maternal diets both show strong associations with obesity in offspring. As the number of overweight reproductive-age women increases, the number of overweight children and infants also increases . It has been postulated that maternal obesity causes an accumulation of fat in fetal adipose tissue (adiposity) and predisposes babies for obesity in childhood and adulthood . Animal studies have shown that maternal overnutrition may impact brain development and cause disruptions to programming of the hypothalamus. Offspring that were exposed to a high-fat or high-caloric maternal diet had increased levels of insulin, glucose and leptin. It is hypothesized that these elevations are due to disturbances in the complex neuronal network that includes the neuropeptide Y (NPY) and proopiomelanocortin (POMC) pathways . This altered neuronal signaling can consequently impact food-intake behavior and lead to diet-induced obesity in adulthood. While epigenetic modifications are most likely involved in the development of obesity, the specific target genes have yet to be identified. Genes involved in adipogenesis, such as fibroblast growth-factor-2, phosphatase and tensin homologue, cyclin-dependent kinase inhibitor 1A and oestrogen receptor-alpha, possess multiple CpG islands in their promoter sites and may act as epigenetic targets . Furthermore, it has been shown that prenatal exposure to a hypomethylating agent, such as bisphenol A (BPA), is associated with increased body weight and suggests modified DNA methylation as a mechanism for increasing susceptibility to obesity .

Folate

It has long been realized that maternal folate intake during pregnancy is linked to fetal development and growth, and can reduce the risk of serious birth defects. Folate is a source of S-adenosyl methionine (SAM), which is used to supply DNA methyltransferases with methyl groups. Therefore, changes in folate supply have a substantial effect on DNA methylation patterns. Low levels of folate are associated with an increased risk of preterm delivery, poor growth of the placenta and uterus, and intrauterine growth retardation . Several complex diseases, including cancer, cardiovascular diseases and autism have also been linked to maternal folate status. Based on animal studies it has been hypothesized that reduced folate intake could increase the risk of neural tube defects by reducing the amount of methylayed DNA during cranial neural tube closure . Recently it was discovered that folate protection from congenital heart defects is linked to epigenetics and Wnt signaling. Multiple environmental factors target the Wnt signaling pathway during embryogenesis and can cause misregulation of the pathway. Folic acid metabolism generates SAM, thereby altering the methylation states of histones H3K9, H3K4, and H3K27 and genetically altering Wnt signaling .

Perinatal

Another critical developmental time window is the perinatal period, which refers to the time period immediately before and after birth. It has been shown that maternal diet in late pregnancy and an infant's diet in the beginning weeks can all have significant impacts on gene expression. Therefore, perinatal nutrition refers to both late-stage in utero nutrition and lactation.

Bone Health

Bone mass and the development of osteoporosis have been studied in relation to perinatal nutrition. An important factor to consider when investigating perinatal nutrition is whether the baby was breast-fed or formula-fed. Studies have shown that breast-fed babies have increased bone mass compared to those were not breast-fed, and that this small increase in bone mass during a period of critical development could potentially program the skeleton to continue along a "healthy" growth trajectory . It has also been shown that maternal vitamin D insufficiency during late pregnancy is associated with reduced bone size and mineral mass in late childhood. Peak bone mass has shown to be a good predictor of risk of fracture and osteoporosis, with even a small increase in peak bone mass resulting in a much lower risk of bone fracture . Research shows that genetic markers explain only a small proportion of variation in bone mass and risk of fracture. Therefore, healthy bone programming is most likely influenced by various epigenetic mechanisms, such as imprinting of the growth promoting genes IGF-2, or changes to the hypothalamic-pituitary-adrenal axis (HPA).

Neurodevelopment

Imbalances in maternal nutrition can also have a significant effect on fetal neurodevelopment. Brain development occurs most rapidly during fetal development and infancy, and research has shown that exposure to certain environmental conditions can have long lasting effects on cognition. Specifically, n-3 fatty acids, iodine, iron and choline have been shown to influence brain development and impact cognitive ability and behavior. The greatest evidence for a link between nutrition and neurodevelopment comes from studies that show low birth weight associated with low IQ and increased risk of schizophrenia. Several studies suggest that breast-feeding promotes long-term neurodevelopment by providing the nutrients necessary for proper brain development. A study in mice showed that choline-deficient diets during the late gestation period impaired fetal brain development, including decreased cell proliferation and reduced visual-spatial and auditory memory. These cognitive changes appeared to be due to altered histone and DNA methylation patterns in the fetal hippocampus, thus providing a link between maternal nutrition, epigenetics, and early brain development.

Type-1 Diabetes

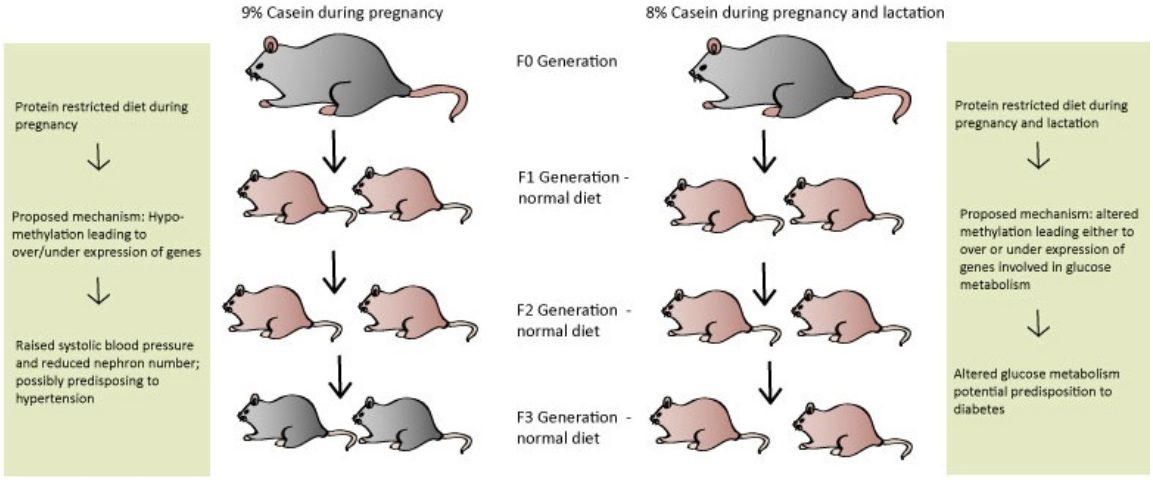
It has been postulated that breast-feeding may also protect against type-1 diabetes, with research showing that formula-fed infants are at an increased risk of developing islet autoantibodies. Individuals with type-1 diabetes experience a pre-clinical diabetes phase characterized by autoimmunity against pancreatic islets. The introduction of certain foods in the first few months of life, such as berries and cereal, is significantly associated with increased risk of islet autoantibody development compared to babies who are exposed to solid foods later in life. While the pathogenesis behind development of autoantibodies remains largely unknown, it is very probable that an epigenetic link exists between perinatal diet and risk of type-1 diabetes.

Adulthood

The majority of research in nutriepigenomics has focused on nutritional imbalances during gestation and lactation periods. However, foods that are consumed during adulthood can also impact gene expression and disease pathogenesis. Cancer is the disease most commonly associated with adult nutrition and epigenetic modifications. DNA hypomethelation promotes cancer progression by allowing increased gene transcription, while hypermethylation can silence tumor suppressor genes and further promote uncontrolled cell division and tumor formation. Compounds found in foods, such as genistein and tea polyphenols, are able to regulate DNA methyltransferases and histone acetylation in cultured cancer cells and may provide protection against certain types of cancer . Other dietary compounds, such as diallyl disulfide present in garlic and sulforaphane present in cruciferous vegetables, have been associated with cancer prevention in clinical trials . This is due to their ability to inhibit histone deacetylase (HDAC) enzymes and prevent silencing of important regulatory genes.

Transgenerational effects

Many believe epigenetic regulation is cleared during the fertilization process, yet more evidence for transgenerational effects (TGEs) are being revealed. These TGEs take place when the epigenetic regulatory patterns are not sufficiently erased during fertilization, possibly due to nutrition levels in previous generations. Later generations may be affected from caloric and protein restriction, high-fat interventions and endocrine disruption in earlier generations. Differences within the nutritional behavior of the maternal rat are believed to cause malprogramming in the F1 generation and may then be passed to subsequent generations. Maternal rats fed a PR-diet during the entire length of pregnancy led to metabolic-related problems in the F1 and F2 generations, even with normal nutrition during the F1 pregnancy. These effects have also been seen in the F3 generation depending on the length of protein restriction. If protein restriction occurred solely during pregnancy, the F1 and F2 offspring had higher systolic blood pressure and lower nephron numbers, possibly predisposing them to hypertension. Altered glucose utilization was detected in the grand-offspring of maternal rats fed a PR-diet during pregnancy and lactation, potentially resulting in diabetes later on in life



Transgenerational effects of maternal protein restriction

Protein-restriction in the F0 generation led to hypomethylation of promoters involved in metabolism in the F1 and F2 generations, even though the F1 pregnant rat was given a normal diet . The exact mechanism of this situation has yet to be elucidated; however, direct transmission is a distinct possibility, meaning the epigenetic marks were preserved during spermatogenesis and oogenesis, when they are normally erased.

Models used in nutriepigenomic studies

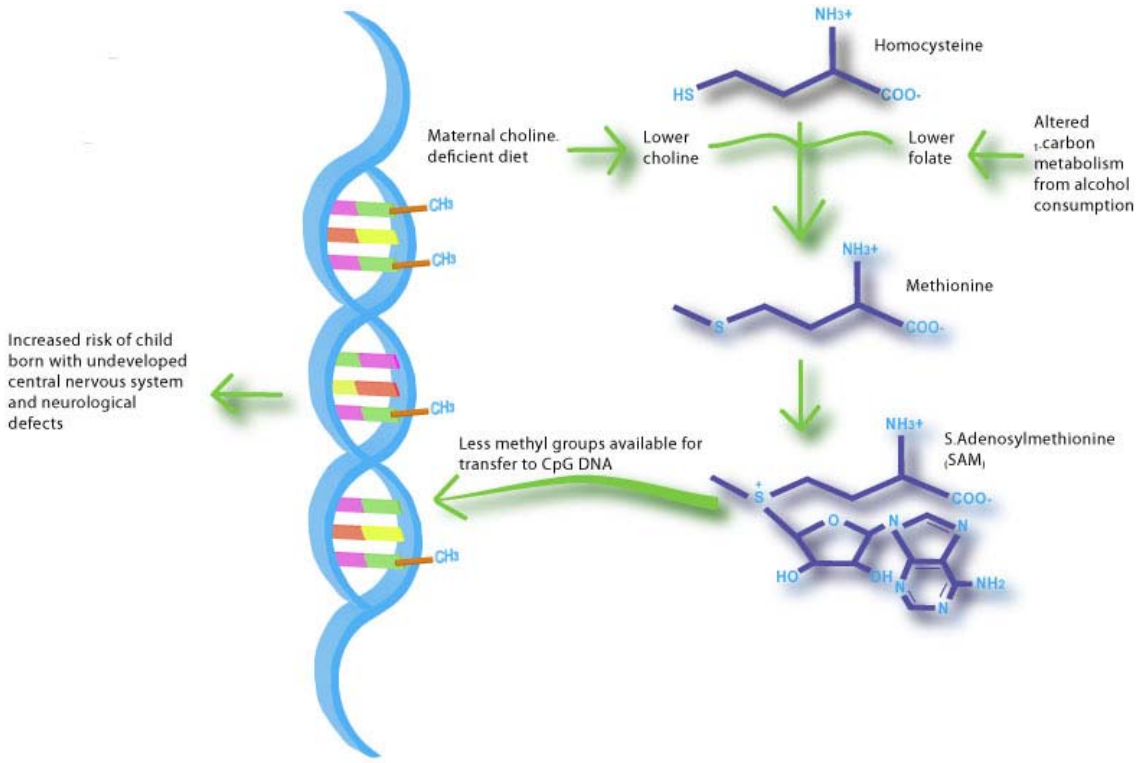
Most research to date use common rodent models to investigate the role of nutrition on phenotype . Popular areas to investigate include IUGR studies, whereby rodents, and sometimes sheep, are subjected to a variety of nutritional conditions. A model for studying IUGR in rodent was developed by Simmons et al (2010) and is used to

investigate type II diabetes . The maternal rats have their uterine arteries ligated, causing altered use of glucose and insulin in the fetus and can therefore serve as a model for diabetes. These growth-retarded rats were found to be highly similar to human fetuses, as they both display symptoms such as lowered glucose and insulin levels. Gestational diabetes may also be studied through chemical induction using streptozotocin treatment of the pregnant rats. Streptozotocin can cause destruction of the beta cells within the pancreas depending on the concentration given.

The predominant means of investigating nutriepigenetics involves varying the nutritional conditions to which a subject is exposed to and monitoring the effects thereafter. Restricting caloric and protein intake are the two most common methods . A pregnant rodent may have their caloric intake reduced up to 30-50% of normal intake. Protein restricted rodents are given 8-9% casein, as opposed to control rats that are fed 20% casein. Micronutrients, such as zinc and iron, may also be restricted to investigate the effects on offspring. Additionally, rats fed diets lacking or including methyl donors are often used to study the effects of diet on epigenomics, as variations within the methylation of DNA are common means of silencing or expressing genes . Supplementing maternal rats with folic acid, vitamin B12, choline and betaine leads to increased levels of DNA methylation at CpG sites and causes a coat color change . This is an example of epigenetically modifiable loci called a “metastable epiallele”, of which only a few have been identified. The above is an example of the “agouti” gene locus, whereby the insertion of a transposable element upstream to the Agouti gene is hypermethylated from the supplementation and causes a change in the rat’s coat color. Diets containing higher carbohydrate and fat content attempt to mimic typical Western-style diets may also be used in nutriepigenetic studies . Another method used is “catch-up”, where offspring of rats born to mothers subjected to various diets are subsequently cross-fostered to mothers fed normal diets .

Future directions

The possibilities of utilizing nutriepigenomics for intervention are quite expansive. This can include preventative therapies, such as providing an optimal regime for nutrition during pregnancy and lactation . It is already common place for pregnant mothers to supplement their diets with choline and folate to prevent the development of neurological disabilities in the fetus.



The nutriepigenetic pathway of maternal choline-deficient diets helps to elucidate the development of fetal alcohol syndrome

A highly specific diet, termed an “EpiG diet”, may be employed for an individual believed to be at higher risk of developing a metabolic disorder . These diets may include supplementation with methyl donors, such as folate. There are also many other natural compounds, such as resveratrol, curcumin and green tea that have been termed “epigenetic modifiers”, as they have anti-cancer capabilities in addition to being used as treatments for metabolic diseases . However, the functions of these compounds still require long-term studies to evaluate their effect over time.

There also exists potential for therapeutic treatments that may correct metabolic disorders, such as type II diabetes . Components of garlic and cruciferous vegetables are known to possess HDAC inhibitors that modify the acetylation of histone proteins and may contain a protection against cancer . These same compounds have also been implicated in irritable bowel syndrome (IBS) and colon cancer, as they may modify the histones normally implicated in these diseases .

Elucidation of disease pathways is another future direction for nutriepigenomic studies. For example, choline-deficient diets and alcohol metabolism during pregnancy may have very similar metabolic pathways . Therefore, animal studies using choline-restricted diets may assist in investigations of fetal alcohol spectrum disorders.

When compared to studies of maternal transmission, investigations into the role of paternal diets are lacking. A review demonstrated the nutrition of both parents do in fact play a role in determining the health of their offspring . A germ-line study reported paternal rats fed a high-fat diet led to insulin dysfunction in the F1 offspring . While this likely occurs via epigenetic modifications similar to those postulated in the maternal diets, the exact mechanism remains to be defined. Assessing the role of epigenetic mechanisms may be easier using paternal inheritance, as sperm transmits epigenetic and genetic information, whereas the female cells also transmit mitochondrial DNA .

Chapter 8

Paramutation and Sex-Determination System

Paramutation

In epigenetics, **paramutation** is an interaction between two alleles of a single locus, resulting in a heritable change of one allele that is induced by the other allele. Paramutation violates Mendel's first law, which states that in the process of the formation of the gametes (egg or sperm) the allelic pairs separate, one going to each gamete, and that each allele remains completely uninfluenced by the other. In paramutation an allele in one generation heritably affects the other allele in future generations, even if the allele causing the change is itself not transmitted. What may be transmitted are patterns of DNA methylation or RNAs such as piRNAs, siRNAs, miRNAs or other regulatory RNAs. Through proper breeding, paramutation can result in isogenic sibling plants with drastically different phenotypes.

Paramutation was first discovered and studied in maize (*Zea mays*) by R.A. Brink at the University of Wisconsin–Madison in the 1950s. Brink noticed that specific weakly expressed alleles of the *red1 (r1)* locus in maize, which encodes a transcription factor that confers red pigment to corn kernels, can heritably change specific strongly expressed alleles to a weaker expression state. The weaker expression state adopted by the changed allele is heritable and can, in turn, change the expression state of other active alleles in a process termed **secondary paramutation**. Brink showed that the influence of the paramutagenic allele could persist for many generations.

Interestingly, paramutation can result in a single allele of a gene controlling a spectrum of phenotypes. At *r1* in maize, for example, the weaker expression state adopted by an allele following paramutation can range from completely colorless to nearly fully-colored kernels. This is an exception to the general observation that continuous variation is controlled by many genes.

Allelic interactions similar to paramutation have since been reported in other organisms, including tomato, pea, and mice.

The molecular basis of paramutation is being unraveled, almost exclusively in maize. Paramutation may share common mechanisms with other epigenetic phenomena, such as gene silencing and genomic imprinting. In maize, paramutation seems to share many traits with the well understood RNA-directed DNA-methylation pathway in *Arabidopsis thaliana*, even though it has never been observed in the famous model plant. Alleman (2006) reported that, in maize, "paramutation is RNA-directed. Stability of the chromatin states associated with paramutation and transposon silencing requires the *mop1* gene, which encodes an RNA-dependent RNA polymerase." Exactly how the RNA produced by this polymerase causes paramutation in maize is not yet understood, but like other epigenetic changes, it involves the covalent modification of DNA and/or the DNA-bound histone proteins without changing the DNA sequence itself.

Sex-determination system

A **sex-determination system** is a biological system that determines the development of sexual characteristics in an organism. Most sexual organisms have two sexes. In many cases, sex determination is genetic: males and females have different alleles or even different genes that specify their sexual morphology. In animals, this is often accompanied by chromosomal differences. In other cases, sex is determined by environmental variables (such as temperature) or social variables (the size of an organism relative to other members of its population). The details of some sex-determination systems are not yet fully understood.

Chromosomal determination

XX/XY sex chromosomes

The **XX/XY sex-determination system** is the most familiar sex-determination systems, as it is found in human beings, most other mammals, as well as some insects. However, at least one monotreme, the platypus, presents a particular sex determination scheme that in some ways resembles that of the ZW sex chromosomes of birds, and also lacks the SRY gene, whereas some rodents, such as several Arvicolinae (voles and lemmings), are also noted for their unusual sex determination systems. The platypus has ten sex chromosomes; males have an XYXYXYXYXY pattern while females have ten X chromosomes. Although it is an XY system, the platypus' sex chromosomes share no homologues with eutherian sex chromosomes. Instead, homologues with eutherian sex chromosomes lie on the platypus chromosome 6, which means that the eutherian sex chromosomes were autosomes at the time that the monotremes diverged from the therian mammals (marsupials and eutherian mammals). However, homologues to the avian DMRT1 gene on platypus sex chromosomes X3 and X5 suggest that it is possible the sex-determining gene for the platypus is the same one that is involved in bird sex-

determination. However, more research must be conducted in order to determine the exact sex determining gene of the platypus.

In the XY sex-determination system, females have two of the same kind of sex chromosome (XX), while males have two distinct sex chromosomes (XY). Some species (including humans) have a gene SRY on the Y chromosome that determines maleness; others (such as the fruit fly) use the presence of two X chromosomes to determine femaleness. The XY sex chromosomes are different in shape and size from each other unlike the autosomes, and are termed allosomes.

XX/X0 sex determination

In this variant of the XY system, females have two copies of the sex chromosome (XX) but males have only one (X0). The *0* denotes the absence of a second sex chromosome. This system is observed in a number of insects, including the grasshoppers and crickets of order Orthoptera and in cockroaches (order Blattodea).

The nematode *C. elegans* is male with one sex chromosome (X0); with a pair of chromosomes (XX) it is a hermaphrodite.

ZW sex chromosomes

The **ZW sex-determination system** is found in birds and some insects and other organisms. The ZW sex-determination system is reversed compared to the XY system: females have two different kinds of chromosomes (ZW), and males have two of the same kind of chromosomes (ZZ). In the chicken, this was found to be dependent on the expression of DMRT1.

Haplodiploidy

Haplodiploidy is found in insects belonging to Hymenoptera, such as ants and bees. Unfertilized eggs develop into haploid individuals, which are the males. Diploid individuals are generally female but may be sterile males. Thus, if a queen bee mates with one drone, her daughters share ¾ of their genes with each other, not ½ as in the XY and ZW systems. This is believed to be significant for the development of eusociality, as it increases the significance of kin selection. This is common also in wasps that are parasitic and in the male greenflies.

Non-genetic sex-determination systems

Many other sex-determination systems exist. In some species of reptiles, including alligators, some turtles, and the tuatara, sex is determined by the temperature at which the egg is incubated. Some species, such as some snails, practice sex change: adults start out male, then become female. In tropical clown fish, the dominant individual in a group becomes female while the other ones are male, and blue wrasse fish are the reverse. In the

marine worm *Bonellia viridis*, larvae become males if they make physical contact with the female, and females if they end up on the bare sea floor.

Some species, however, have no sex-determination system. Hermaphrodites include the common earthworm and certain species of snails. A few species of fish, reptiles, and insects reproduce by parthenogenesis and are female altogether.

In some arthropods, sex is determined by infection, as when Bacteria of the genus *Wolbachia* alter their sexuality; some species consist entirely of ZZ individuals, with sex determined by the presence of *Wolbachia*.

Other unusual systems:

- Swordtail fish
- The Chironomus midge species
- The Platypus has 10 sex chromosomes but lacks the mammalian sex-determining gene SRY, meaning that the process of sex determination in the Platypus remains unknown.

Chapter 9

Soft Inheritance, Structural Inheritance and Testis Determining Factor

Soft inheritance

Soft inheritance is the term coined by Ernst Mayr to include such ideas as Lamarckism, that an organism can pass on characteristics that it acquired during its lifetime to its offspring. It contrasts with modern ideas of inheritance, which Mayr called hard inheritance. Since Mendel, modern genetics has held that the hereditary material is impervious to environmental influences (except, of course, mutagenic effects). In soft inheritance "the genetic basis of characters could be modified either by direct induction by the environment, or by use and disuse, or by an intrinsic failure of constancy, and that this modified genotype was then transmitted to the next generation." Concepts of soft inheritance are usually associated with the ideas of Lamarck and Geoffroy. The concept of hard inheritance holds sway today.

One of the first statements in favour of hard inheritance was made by the English surgeon William Lawrence in 1819. His ideas on heredity were many years ahead of their time, as this extract shows: "The offspring inherit only [their parents'] connate peculiarities and not any of the acquired qualities". This is as clear a rejection of soft inheritance as one can find. However, Lawrence qualified it by including the origin of birth defects owing to influences on the mother (an old folk superstition). So Mayr places Wilhelm His, Sr. in 1874 as the first unqualified rejection of soft inheritance. August Weismann, in 1883, gave a comprehensive denial of Lamarkism (soft inheritance) and with his distinction between germ and soma provided a general ideology of hard inheritance which survives to the present day.

Recent work in plants and mammals on the role of the environment on epigenetic modifications of DNA have led to the argument that inherited epigenetic variation is a kind of soft inheritance.

Structural inheritance

Structural inheritance or **cortical inheritance** is the transmission of a trait in a living organism by a self-perpetuating spatial structures. This is in contrast to the transmission of digital information such as is found in DNA sequences, which accounts for the vast majority of known genetic variation.

Examples of structural inheritance include the propagation of prions, the infectious proteins of diseases such as scrapie (in sheep and goats), bovine spongiform encephalopathy ('mad cow disease') and Creutzfeld-Jakob disease (although the protein-only hypothesis of prion transmission has been considered contentious until recently.) Prions based on heritable protein structure also exist in yeast. Structural inheritance has also been seen in the orientation of cilia in protozoans such as *Paramecium* and *Tetrahymena*, and 'handedness' of the spiral of the cell in *Tetrahymena*, and shells of snails. Some organelles also have structural inheritance, such as the centriole, and the cell itself (defined by the plasma membrane) may also be an example of structural inheritance.

Testis determining factor

Testis-determining factor (**TDF**) is a general term for the gene (or product thereof) that results in maleness in humans and some other species.

Certain genes cause chemical reactions that result in the development of testes. Embryos are gonadally identical, regardless of genetic sex, until a certain point in development; then the testis-determining factor causes male sex organs to develop, whereas lack of this factor will cause the embryo to develop as physically female.

The TDF factor is encoded by the SRY gene located in the Y chromosome. It is a DNA-binding protein that enhances other transcription factors, or is a transcription factor itself. Its expression directly or indirectly causes the development of primary sex cords, which will later develop to seminiferous tubules. These cords form in the central part of the yet-undifferentiated gonad, turning it into a testis. The testis then starts secreting testosterone and the Mullerian Inhibiting Substance.

Older texts discuss the role of the HY antigen in the control of testicular development, which was later disproven.

Role in disease

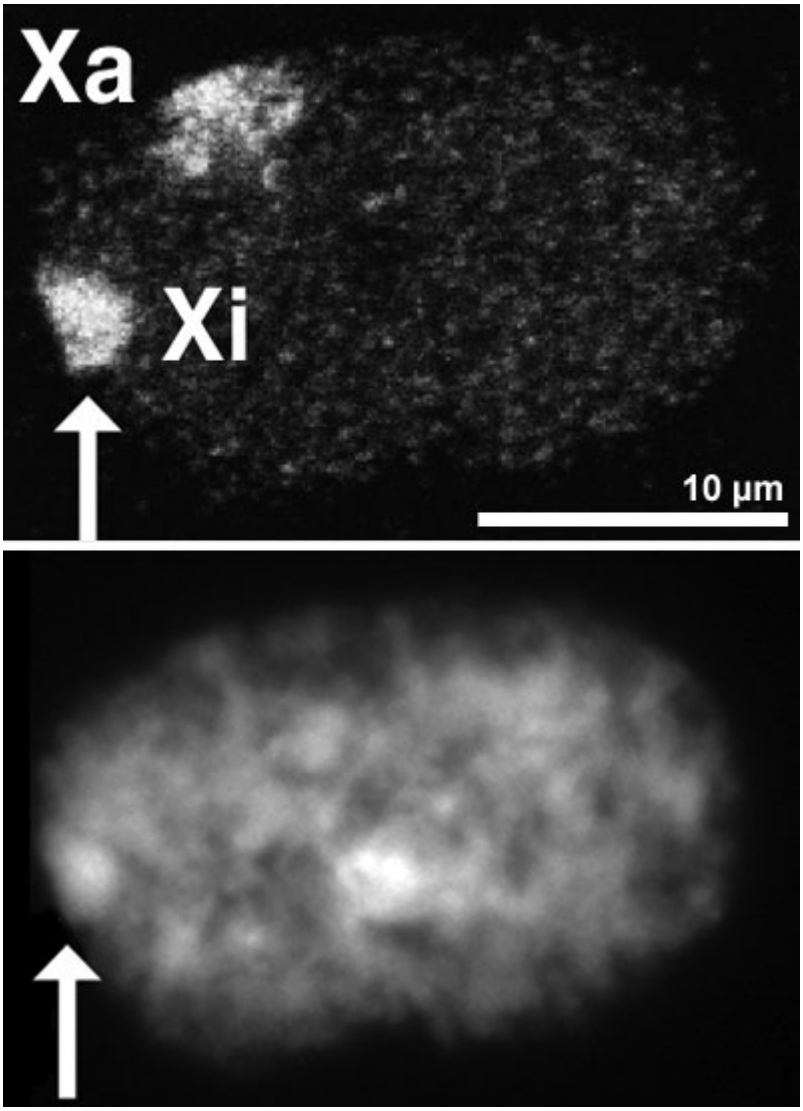
The TDF gene has some interesting implications. The genetic recombination of Crossing over can cause the gene to be transferred on to the X chromosome. In this case, the X chromosome will initiate testis development; so, regardless of whether the person has a Y chromosome, the person will turn into a male. Though everything else will be developed as if it were a female (other sex-related alleles), the apparent sex will be male (a syndrome known as XX male syndrome).

On the converse, such a cross-over event also can result in a Y chromosome that is missing the Sex-Determining Region (SRY), which contains the TDF, replaced with the corresponding sequence from the end of the X chromosome. Individuals that inherit this Y chromosome will develop as females, despite having the normal male chromosomal set of one X and one Y. This is called Swyer syndrome (46XY, genotypic male but phenotypic female).

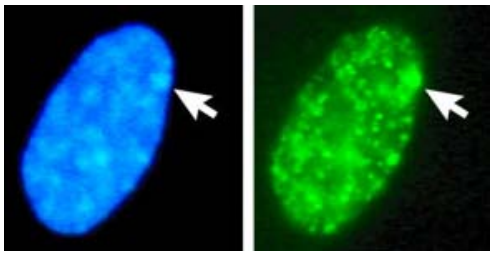
Chapter 10

X-Inactivation

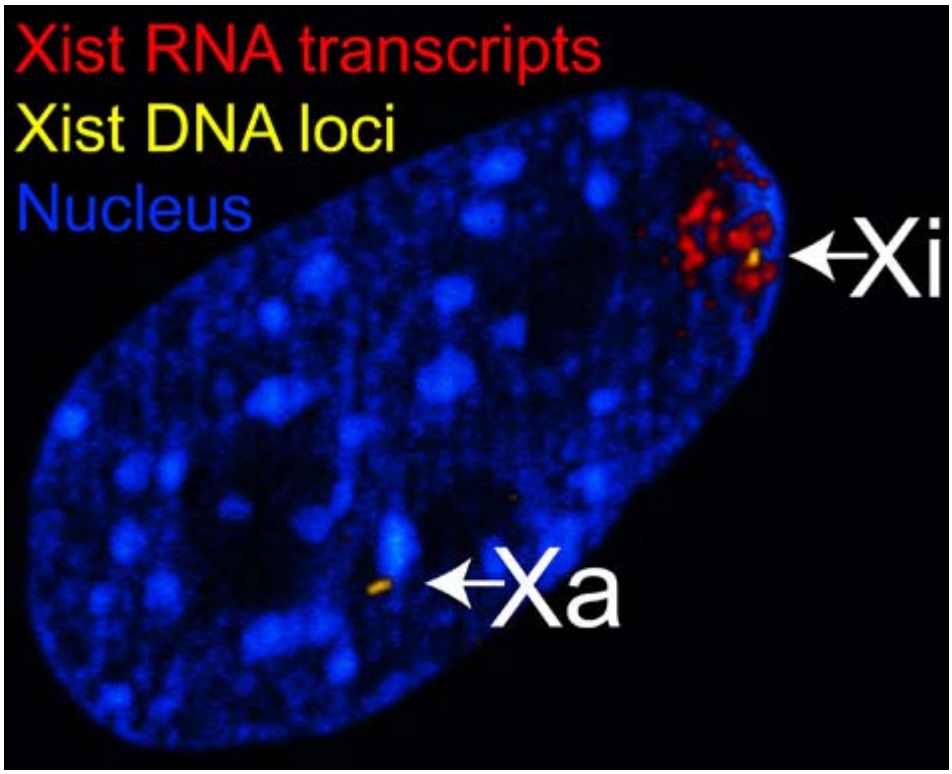
The coloration of tortoiseshell cats is a visible manifestation of X-inactivation. The black and orange alleles of a fur coloration gene reside on the X chromosome. For any given patch of fur, the inactivation of an X chromosome that carries one gene results in the fur color of the other, active gene.



Nucleus of a female cell. Top: Both X-chromosomes are detected, by FISH. Bottom: The same nucleus stained with a DNA stain (DAPI). The Barr body is indicated by the arrow, it identifies the inactive X (Xi).



An interphase female human fibroblast cell. Arrows point to sex chromatin on DNA (DAPI) in cell nucleus(left), and to the corresponding X chromatin (right).
Left: DNA (DAPI)-stained nucleus. Arrow indicates the location of Barr body(Xi). Right: DNA associated histones protein detected



The figure shows confocal microscopy images from a combined RNA-DNA FISH experiment for Xist in fibroblast cells from adult female mouse, demonstrating that Xist RNA is coating only one of the X-chromosomes. RNA FISH signals from Xist RNA are shown in red color, marking the inactive X-chromosome (Xi). DNA FISH signals from Xist loci are shown in yellow color, marking both active and inactive X-chromosomes (Xa, Xi). The nucleus (DAPI-stained) is shown in blue color. The figure is adapted from:

X-inactivation (also called **lyonization**) is a process by which one of the two copies of the X chromosome present in female mammals is inactivated. The inactive X chromosome is silenced by packaging into transcriptionally inactive heterochromatin. X-inactivation occurs so that the female, with two X chromosomes, does not have twice as many X chromosome gene products as the male, which only possess a single copy of the X chromosome. The choice of which X chromosome will be inactivated is random in placental mammals such as mice and humans, but once an X chromosome is inactivated it will remain inactive throughout the lifetime of the cell and its descendants in the organism. Unlike the random X-inactivation in placental mammals, inactivation in marsupials applies exclusively to the paternally derived X chromosome.

History

In 1959 Susumu Ohno showed that the two X-chromosomes of mammals were different: one appeared like the autosomes; the other was condensed and heterochromatic. This finding suggested, independently to two groups of investigators, that one of the X-

chromosomes underwent inactivation. In 1961, Mary Lyon proposed the random inactivation of one female X chromosome to explain the mottled phenotype of female mice heterozygous for coat color genes. The Lyon hypothesis also accounted for the findings that one copy of the X chromosome in female cells was highly condensed, and that mice with only one copy of the X chromosome developed as infertile females. Ernest Beutler, studying heterozygous females for Glucose-6-phosphate dehydrogenase (G6PD) deficiency, independently proposed that there were two red cell populations of erythrocytes in such heterozygotes: deficient cells and normal cells, depending on whether the inactivated X chromosome contains the normal or defective G6PD allele.

Mechanism

Timing

All mouse cells undergo an early, imprinted inactivation of the paternally-derived X chromosome in two-cell or four-cell stage embryos. The extraembryonic tissues (which give rise to the placenta and other tissues supporting the embryo) retain this early imprinted inactivation, and thus only the maternal X chromosome is active in these tissues.

In the early blastocyst, this initial, imprinted X-inactivation is reversed in the cells of the inner cell mass (which give rise to the embryo), and in these cells both X chromosomes become active again. Each of these cells then independently and randomly inactivates one copy of the X chromosome. This inactivation event is irreversible during the lifetime of the cell, so all the descendants of a cell which inactivated a particular X chromosome will also inactivate that same chromosome. This phenomenon, which can be observed in the coloration of calico cats when females are heterozygous for the X-linked gene, should not be confused with mosaicism, which is a term that specifically refers to differences in the genotype of various cell populations in the same individual; X-inactivation, which is an epigenetic change that results in a different phenotype, is *not* a change at the genotypic level. For an individual cell or lineage the inactivation is therefore skewed or 'non-random' this can give rise to mild symptoms in female 'carriers' of X-linked genetic disorders.

X-inactivation is reversed in the female germline, so that all oocytes contain an active X chromosome.

Selection of one active X chromosome

Normal females possess two X chromosomes, and in any given cell one chromosome will be active (designated as Xa) and one will be inactive (Xi). However, studies of individuals with extra copies of the X chromosome show that in cells with more than two X chromosomes there is still only one Xa, and all the remaining X chromosomes are inactivated. This indicates that the default state of the X chromosome in females is inactivation, but one X chromosome is always selected to remain active.

It is hypothesized that there is an autosomally-encoded 'blocking factor' which binds to the X chromosome and prevents its inactivation. The model postulates that there is a limiting blocking factor, so once the available blocking factor molecule binds to one X chromosome the remaining X chromosome(s) are not protected from inactivation. This model is supported by the existence of a single Xa in cells with many X chromosomes and by the existence of two active X chromosomes in cell lines with twice the normal number of autosomes.

Sequences at the **X inactivation center** (**XIC**), present on the X chromosome, control the silencing of the X chromosome. The hypothetical blocking factor is predicted to bind to sequences within the XIC.

Chromosomal component

The X-inactivation center (XIC) on the X chromosome is necessary and sufficient to cause X-inactivation. Chromosomal translocations which place the XIC on an autosome lead to inactivation of the autosome, and X chromosomes lacking the XIC are not inactivated.

The XIC contains two non-translated RNA genes, Xist and Tsix, which are involved in X-inactivation. The XIC also contains binding sites for both known and unknown regulatory proteins.

Xist and Tsix RNAs

The X-inactive specific transcript (Xist) gene encodes a large non-coding RNA that is responsible for mediating the specific silencing of the X chromosome from which it is transcribed. The inactive X chromosome is coated by Xist RNA, whereas the Xa is not. The Xist gene is the only gene which is expressed from the Xi but not from the Xa. X chromosomes which lack the Xist gene cannot be inactivated. Artificially placing and expressing the Xist gene on another chromosome leads to silencing of that chromosome.

Prior to inactivation, both X chromosomes weakly express Xist RNA from the Xist gene. During the inactivation process, the future Xa ceases to express Xist, whereas the future Xi dramatically increases Xist RNA production. On the future Xi, the Xist RNA progressively coats the chromosome, spreading out from the XIC; the Xist RNA does not localize to the Xa. The silencing of genes along the Xi occurs soon after coating by Xist RNA.

Like Xist, the Tsix gene encodes a large RNA which is not believed to encode a protein. The Tsix RNA is transcribed antisense to Xist, meaning that the Tsix gene overlaps the Xist gene and is transcribed on the opposite strand of DNA from the Xist gene. Tsix is a negative regulator of Xist; X chromosomes lacking Tsix expression (and thus having high levels of Xist transcription) are inactivated much more frequently than normal chromosomes.

Like Xist, prior to inactivation, both X chromosomes weakly express Tsix RNA from the Tsix gene. Upon the onset of X-inactivation, the future Xi ceases to express Tsix RNA (and increases Xist expression), whereas Xa continues to express Tsix for several days.

Silencing

The inactive X chromosome does not express the majority of its genes, unlike the active X chromosome. This is due to the silencing of the Xi by repressive heterochromatin, which coats the Xi DNA and prevents the expression of most genes.

Compared to the Xa, the Xi has high levels of DNA methylation, low levels of histone acetylation, low levels of histone H3 lysine-4 methylation, and high levels of histone H3 lysine-9 methylation, all of which are associated with gene silencing. Additionally, a histone variant called macroH2A is exclusively found on nucleosomes along the Xi.

Barr bodies

DNA packaged in heterochromatin, such as the Xi, is more condensed than DNA packaged in euchromatin, such as the Xa. The inactive X forms a discrete body within the nucleus called a Barr body. The Barr body is generally located on the periphery of the nucleus, is late replicating within the cell cycle, and, as it contains the Xi, contains heterochromatin modifications and the Xist RNA.

Expressed genes on the inactive X chromosome

A fraction of the genes along the X chromosome escape inactivation on the Xi. The Xist gene is expressed at high levels on the Xi and is not expressed on the Xa. Other genes are expressed equally from the Xa and Xi; mice contain few genes which escape silencing whereas up to a quarter of human X chromosome genes are expressed from the Xi. Many of these genes occur in clusters.

Many of the genes which escape inactivation are present along regions of the X chromosome which, unlike the majority of the X chromosome, contain genes also present on the Y chromosome. These regions are termed pseudoautosomal regions, as individuals of either sex will receive two copies of every gene in these regions (like an autosome), unlike the majority of genes along the sex chromosomes. Since individuals of either sex will receive two copies of every gene in a pseudoautosomal region, no dosage compensation is needed for females, so it is postulated that these regions of DNA have evolved mechanisms to escape X-inactivation. The genes of pseudoautosomal regions of the Xi do not have the typical modifications of the Xi and have little Xist RNA bound.

The existence of genes along the inactive X which are not silenced explains the defects in humans with abnormal numbers of the X chromosome, such as Turner syndrome (X0) or Klinefelter syndrome (XXY). Theoretically, X-inactivation should eliminate the differences in gene dosage between affected individuals and individuals with a normal

chromosome complement, but in affected individuals the dosage of these non-silenced genes will differ as they escape X-inactivation.

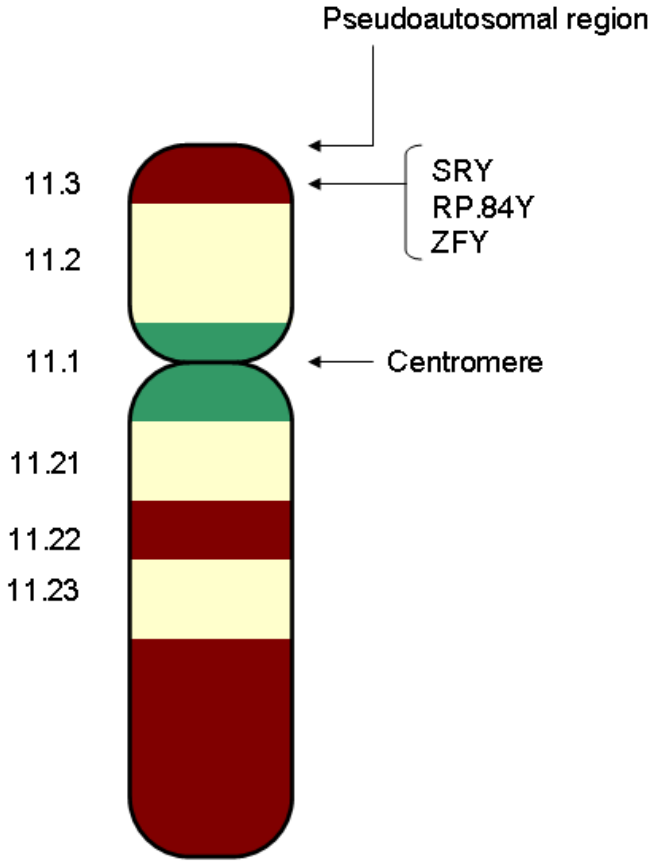
The precise mechanisms that control escape from X-inactivation are not known, but silenced and escape regions have been shown to have distinct chromatin marks. It has been suggested that escape from X-inactivation might be mediated by expression of long non-coding RNA (lncRNA) within the escaping chromosomal domains.

Uses in experimental biology

Stanley Michael Gartler used X chromosome inactivation to demonstrate the clonal origin of cancers. Examining normal tissues and tumors from females heterozygous for isoenzymes of the sex-linked G6PD gene demonstrated that tumor cells from such individuals express only one form of G6PD, whereas normal tissues are composed of a nearly equal mixture of cells expressing the two different phenotypes. This pattern suggests that a single cell, and not a population, grows into a cancer.

Chapter 11

Sex Determination and Differentiation (Human)



The Human Y Chromosome showing the SRY gene which codes for a protein regulating sexual differentiation.

Human sex refers to the processes by which an individual becomes either a male or female during development.

The Jost Paradigm

Under typical circumstances, the sex of an individual will be determined and expressed through the following mechanisms:

- Chromosomal Sex (genetic): Presence or absence of Y chromosome
- Gonadal Sex (Primary Sex Determination): Controlled by presence or absence of testis determining factor (TDF)
- Phenotypic Sex (Secondary Sex Differentiation): Determined by the hormonal products produced by the gonads.

Sex determination

Sex determination at the chromosome level

For the majority of individuals, sex determination is as simple as the presence or absence of a Y chromosome. Those individuals with a Y chromosome (including XXY, XXXY, etc.) will develop into males, and those without one will become female. Some individuals, however, will undergo what is referred to as primary sex reversal, whereby the X and Y chromosomes [cross over] and exchange genetic material. This relatively rare occurrence (approximately 1 in 20000 births) can lead to males with two X chromosomes and females with a Y chromosome.

Testis determining gene

During the late 1980s and early 1990s, coinciding with the mapping of the human genome, researchers began to look for the specific gene on the Y chromosome that, up until then, had been known as the testis determining factor (TDF). Through the study of individuals that underwent primary sex reversal (that is, XX males and XY females), researchers determined that the TDF must lie on the Y chromosome in a location that would permit its exchange to the X chromosome during cross over. In 1985, Dr. David C. Page published an article in Nature boldly stating that the TDF was the ZFY gene on the Y chromosome. However, Dr. MS Palmer later discovered a ZFY analogue on the X chromosome, providing evidence that ZFY was in fact not the TDF. Eventually, Dr. Peter Koopman was able to prove that the SRY gene is the TDF from studies on XX males.. The following evidence further supports this claim:

- SRY is Y specific, and there is no analogue on the X chromosome.
- SRY is deleted or mutated in XY females
- It undergoes expression within the testis at the time of testis differentiation.
- Its sequence suggests that its protein has a DNA binding motif because it has high homology to an 80 amino acid long DNA binding region (HMG box).

SRY a repressor?

Recently, it has been suggested by some that the SRY gene acts as a repressor or inhibitor of another gene, "Z", that is involved in female development. Previously, it was stated that the SRY sequence suggests the presence of a DNA binding motif. Also, the idea that SRY is a repressor is further supported by the fact that a small percentage of sex reversal cases cannot be explained by the absence of SRY and could be due to a mutation in some gene "Z" that prevents the binding of SRY and its subsequent antagonist action.

Other sex determination genes

- DAX1: Exerts its effects early on in development. There is some debate over what its role is in the development of testis. It is a candidate for gene "Z".
- SOX9: mutations in this gene cause severe dwarfism, and a bone disorder called campomelic dysplasia, which occurs in many sex reversed males.

Sex differentiation

Sex differentiation refers to the expression of phenotypic attributes specific to the sex of an individual. While gonad development is a result of the presence or absence of the sex determination gene SRY on the Y chromosome, sex differentiation is determined by the hormonal products produced by the gonads.

Testosterone

In the 1930s, Alfred Jost determined that the presence of testosterone was required for Wolffian duct development in the female rabbit.

Müllerian inhibiting substance

Jost also observed that while testosterone was required for Wolffian duct development, the regression of the Müllerian duct was due to another substance. This was later determined to be Müllerian inhibiting substance (MIS), a 140 kD dimeric glycoprotein that is produced by sertoli cells. MIS blocks the development of Müllerian ducts, promoting their regression.

5-alpha dihydrotestosterone (DHT)

Testosterone is converted to the more potent DHT by 5-alpha reductase. DHT is necessary to exert androgenic effects farther from the site of testosterone production, where the concentrations of testosterone are too low to have any potency. A 5-alpha reductase deficiency results in an androgen disorder characterized by female phenotype or severely undervirilized male phenotype with development of the epididymis, vas deferens, seminal vesicle, and ejaculatory duct, but also a pseudovagina.

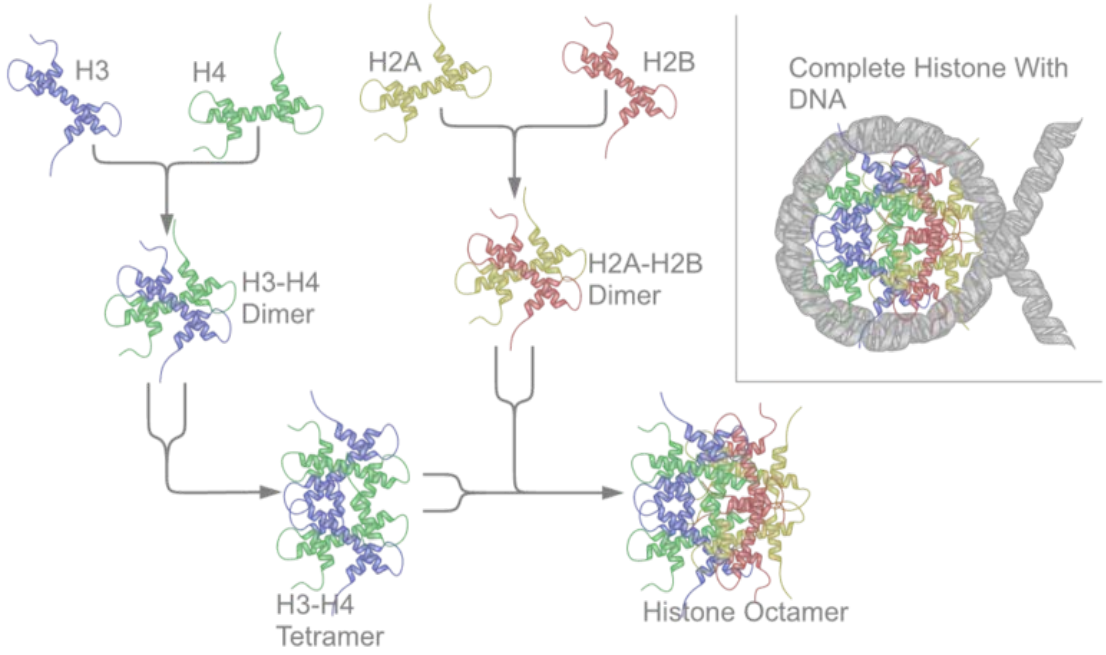
Pathologies

The following disorders are caused by a malfunction in the sex determination and differentiation process:

- Congenital Adrenal Hyperplasia - Inability of adrenal to produce sufficient cortisol, leading to increased production of testosterone resulting in severe masculinization of 46 XX females.
- Persistent Müllerian Duct Syndrome - A rare type of pseudohermaphroditism that occurs in 46 XY males, caused by either a mutation in the Müllerian inhibiting substance (MIS) gene, on 19p13, or its type II receptor, 12q13. Results in a retention of Müllerian ducts (persistance of rudimentary uterus and fallopian tubes in otherwise normally virilized males), unilateral or bilateral undescended testes and sometimes causes infertility.
- Male Pseudohermaphroditism - Failure of androgen production or inadequate androgen response, which can cause incomplete masculinization in XY males. Varies from mild failure of masculinization with undescended testes to complete sex reversal and female phenotype (Androgen insensitivity syndrome)
- Swyer syndrome. A form of complete gonadal dysgenesis, mostly due to mutations in the first step of sex determination; the SRY genes.

Chapter 12

Histone

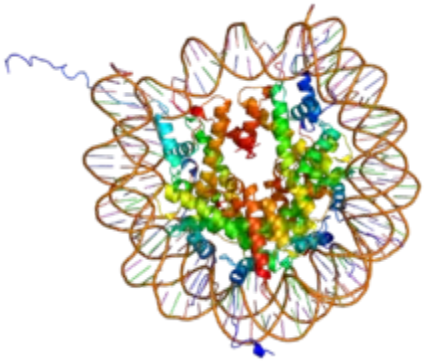


Schematic representation of the assembly of the core histones into the nucleosome.

In biology, **histones** are highly alkaline proteins found in eukaryotic cell nuclei, which package and order the DNA into structural units called nucleosomes. They are the chief protein components of chromatin, acting as spools around which DNA winds, and play a role in gene regulation. Without histones, the unwound DNA in chromosomes would be very long (a length to width ratio of more than 10 million to one in human DNA). For example, each human cell has about 1.8 meters of DNA, but wound on the histones it has about 90 millimeters (?) of chromatin, which, when duplicated and condensed during mitosis, result in about 120 micrometers of chromosomes.

Classes

Core histone H2A/H2B/H3/H4



PDB rendering of H2AFJ based on 1aoi.

Identifiers	
Symbol	Histone
Pfam	PF00125
Pfam clan	CL0012
InterPro	IPR007125
SCOP	1hio

linker histone H1 and H5 family

PDB rendering of HIST1H1B based on 1ghc.

Identifiers	
Symbol	Linker_histone
Pfam	PF00538
InterPro	IPR005818

SMART	SM00526
SCOP	1hst

Histones "are highly conserved and can be grouped into five major classes: H1/H5, H2A, H2B, H3, and H4". These are organised into two super-classes as follows:

- core histones – H2A, H2B, H3 and H4
- linker histones – H1 and H5

Two of each of the core histones assemble to form one octameric nucleosome core particle by wrapping 147 base pairs of DNA around the protein spool in 1.65 left-handed super-helical turn. The linker histone H1 binds the nucleosome and the entry and exit sites of the DNA, thus locking the DNA into place and allowing the formation of higher order structure. The most basic such formation is the 10 nm fiber or beads on a string conformation. This involves the wrapping of DNA around nucleosomes with approximately 50 base pairs of DNA separating each pair of nucleosomes (also referred to as linker DNA). The assembled histones and DNA is called chromatin. Higher-order structures include the 30 nm fiber (forming an irregular zigzag) and 100 nm fiber, these being the structures found in normal cells. During mitosis and meiosis, the condensed chromosomes are assembled through interactions between nucleosomes and other regulatory proteins.

The following is a list of human histone proteins:

Super family	Family	Subfamily	Members
Linker	H1	H1F	H1F0, H1FNT, H1FOO, H1FX
		H1H1	HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T
Core	H2A	H2AF	H2AFB1, H2AFB2, H2AFB3, H2AFJ, H2AFV, H2AFX, H2AFY, H2AFY2, H2AFZ
		H2A1	HIST1H2AA, HIST1H2AB, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2AG, HIST1H2AI, HIST1H2AJ, HIST1H2AK, HIST1H2AL, HIST1H2AM
		H2A2	HIST2H2AA3, HIST2H2AC
	H2B	H2BF	H2BFM, H2BFO, H2BFS, H2BFWT

		H2B1	HIST1H2BA, HIST1H2BB, HIST1H2BC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H2BI, HIST1H2BJ, HIST1H2BK, HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BO
		H2B2	HIST2H2BE
	H3		
		H3A1	HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J
		H3A2	HIST2H3C
		H3A3	HIST3H3
	H4		
		H41	HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L
		H44	HIST4H4

Structure

The nucleosome core is formed of two H2A-H2B dimers and a H3-H4 tetramer, forming two nearly symmetrical halves by tertiary structure (C2 symmetry; one macromolecule is the mirror image of the other). The H2A-H2B dimers and H3-H4 tetramer also show pseudodyad symmetry. The 4 'core' histones (H2A, H2B, H3 and H4) are relatively similar in structure and are highly conserved through evolution, all featuring a 'helix turn helix turn helix' motif (which allows the easy dimerisation). They also share the feature of long 'tails' on one end of the amino acid structure - this being the location of post-translational modification (see below).

It has been proposed that histone proteins are evolutionarily related to the helical part of the extended AAA+ ATPase domain, the C-domain, and to the N-terminal substrate recognition domain of Clp/Hsp100 proteins. Despite the differences in their topology, these three folds share a homologous helix-strand-helix (HSH) motif.

Using an electron paramagnetic resonance spin-labeling technique, British researchers measured the distances between the spools around which eukaryotic cells wind their DNA. They determined the spacings range from 59 to 70 Å.

In all, histones make five types of interactions with DNA:

- Helix-dipoles from alpha-helices in H2B, H3, and H4 cause a net positive charge to accumulate at the point of interaction with negatively charged phosphate groups on DNA
- Hydrogen bonds between the DNA backbone and the amide group on the main chain of histone proteins
- Nonpolar interactions between the histone and deoxyribose sugars on DNA
- Salt bridges and hydrogen bonds between side chains of basic amino acids (especially lysine and arginine) and phosphate oxygens on DNA
- Non-specific minor groove insertions of the H3 and H2B N-terminal tails into two minor grooves each on the DNA molecule

The highly basic nature of histones, aside from facilitating DNA-histone interactions, contributes to the water solubility of histones.

Histones are subject to post translational modification by enzymes primarily on their N-terminal tails, but also in their globular domains. Such modifications include methylation, citrullination, acetylation, phosphorylation, SUMOylation, ubiquitination, and ADP-ribosylation. This affects their function of gene regulation.

In general, genes that are active have less bound histone, while inactive genes are highly associated with histones during interphase. It also appears that the structure of histones has been evolutionarily conserved, as any deleterious mutations would be severely maladaptive.

Function

Compacting DNA strands

Histones act as spools around which DNA winds. This enables the compaction necessary to fit the large genomes of eukaryotes inside cell nuclei: the compacted molecule is 40,000 times shorter than an unpacked molecule.

Chromatin regulation

Histones undergo posttranslational modifications which alter their interaction with DNA and nuclear proteins. The H3 and H4 histones have long tails protruding from the nucleosome which can be covalently modified at several places. Modifications of the tail include methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, citrullination and ADP-ribosylation. The core of the histones H2A, H2B and H3 can also be modified. Combinations of modifications are thought to constitute a code, the so-called "histone code". Histone modifications act in diverse biological processes such as gene regulation, DNA repair and chromosome condensation (mitosis).

The common nomenclature of histone modifications is:

- The name of the histone (*e.g.* H3)

- The single letter amino acid abbreviation (*e.g.* K for Lysine) and the amino acid position in the protein
- The type of modification (Me: methyl, P: phosphate, Ac: acetyl, Ub: ubiquitin)

So H3K4me1 denotes the monomethylation of the 4th residue (a lysine) from the start (i.e., the N-terminal) of the H3 protein.

Examples of histone modifications in transcription regulation include:

Type of modification	Histone						
	H3K4	**H3K9**	**H3K14**	**H3K27**	**H3K79**	**H4K20**	**H2BK5**
mono-methylation	activation	activation		activation	activation	activation	activation
di-methylation		repression		repression	activation		
tri-methylation	activation	repression		repression	activation, repression		repression
acetylation		activation	activation				

History

Histones were discovered in 1884 by Albrecht Kossel. The word "histone" dates from the late 19th century and is from the German "Histon", of uncertain origin: perhaps from Greek *histanai* or from *histos*. Until the early 1990s, histones were dismissed by most as inert packing material for eukaryotic nuclear DNA, based in part on the "ball and stick" models of Mark Ptashne and others who believed transcription was activated by protein-DNA and protein-protein interactions on largely naked DNA templates, as is the case in bacteria. During the 1980s, work by Michael Grunstein demonstrated that eukaryotic histones repress gene transcription, and that the function of transcriptional activators is to overcome this repression. We now know that histones play both positive and negative roles in gene expression, forming the basis of the histone code.

The discovery of the H5 histone appears to date back to 1970's, and in classification it has been grouped with H1.

Conservation across species

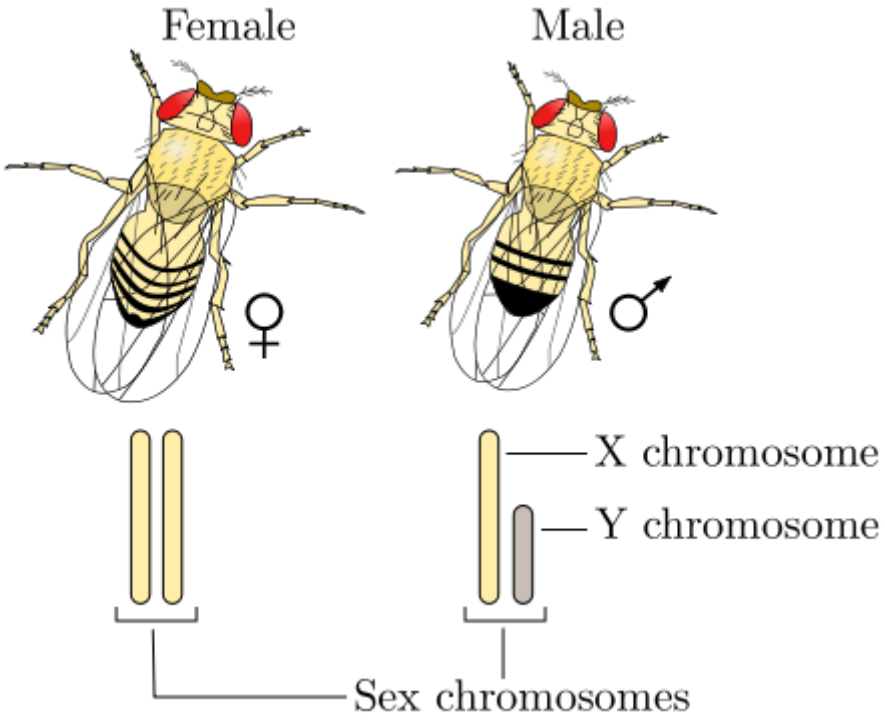
Histones are found in the nuclei of eukaryotic cells, and in certain Archaea, namely Euryarchaea, but not in bacteria. Archaeal histones may well resemble the evolutionary precursors to eukaryotic histones. Histone proteins are among the most highly conserved proteins in eukaryotes, emphasizing their important role in the biology of the nucleus.[:939] In contrast mature sperm cells largely use protamines to package their genomic DNA, most likely because this allows them to achieve an even higher packaging ratio.

Core histones are highly conserved proteins, that is, there are very few differences among the amino acid sequences of the histone proteins of different species. Linker histone usually has more than one form within a species and is also less conserved than the core histones.

There are some *variant* forms in some of the major classes. They share amino acid sequence homology and core structural similarity to a specific class of major histones but also have their own feature that is distinct from the major histones. These *minor histones* usually carry out specific functions of the chromatin metabolism. For example, histone H3-like CenpA is a histone only associated with the centromere region of the chromosome. Histone H2A variant H2A.Z is associated with the promoters of actively transcribed genes and also involved in the prevention of the spread of silent heterochromatin. Another H2A variant H2A.X binds to the DNA with double strand breaks and marks the region undergoing DNA repair. Histone H3.3 is associated with the body of actively transcribed genes.

Chapter 13

XY Sex-Determination System



Drosophila sex-chromosomes

The **XY sex-determination system** is the sex-determination system found in humans, most other mammals, some insects (*Drosophila*) and some plants (*Ginkgo*). In this system, females have two of the same kind of sex chromosome (XX), and are called the homogametic sex. Males have two distinct sex chromosomes (XY), and are called the heterogametic sex. However, an opposite scheme is found in birds.

The XY sex determination system was first described independently by Nettie Stevens and Edmund Beecher Wilson in 1905.

History

Ancient ideas on gender determination

Since ancient times, people believed that the gender of an infant is determined by how much heat a man's sperm had during insemination. Aristotle wrote that:

> “ ...the semen of the male differs from the corresponding secretion of the female in that it contains a principle within itself of such a kind as to set up movements also in the embryo and to concoct thoroughly the ultimate nourishment, whereas the secretion of the female contains material alone. If, then, the male element prevails it draws the female element into itself, but if it is prevailed over it changes into the opposite or is destroyed. ”

Aristotle claimed that the male principle was the driver behind gender determination, such that if the male principle was insufficiently expressed during reproduction, the fetus would develop as a female. In contrast, modern genetics has developed a view on sex determination in which no one single factor is responsible with determining gender, a number of pro-male, anti-male and pro-female genes being responsible.

Beginnings of genetics of sex

Edmund Beecher Wilson and Nettie Stevens are credited with discovering in 1905 the chromosomal XY sex-determination system, the idea that males have XY sex chromosomes and females have XX sex chromosomes.

The first clues to the existence of a factor which determines the development of testis in mammalians came from experiments carried out by Alfred Jost, who castrated embryonic rabbits in utero and noticed that they all developed as female. In the wake of Jost's experiments, C. E. Ford and his team discovered that the Y gene was needed for a foetus to develop as male when they examined patients with Turner's syndrome, who grew up as phenotypic females, and found them to be X0 (hemizygous for X and no Y). All these observations lead to a consensus that a dominant gene which determines testis development (TDF) must exist on the mammalian chromosome Y.

The search for a testis-determining factor (TDF) lead a team of scientists in 1990 to discover a region of the Y chromosome which is necessary for the male sex determination and which was named SRY (**S**ex-determining **R**egion of the **Y** chromosome).

Mechanisms

Some species (including most mammals) have a gene or genes on the Y-chromosome that determine maleness. In the case of humans, a single gene (*SRY*) on the Y-chromosome

acts as a signal to set the developmental pathway towards maleness. Other mammals use several genes on the Y-chromosome for that same purpose. Not all male-specific genes are located on the Y-chromosome.

Other species (including most *Drosophila* species) use the presence of two X chromosomes to determine femaleness. One X chromosome gives putative maleness. The presence of Y-chromosome genes is required for normal male development.

Humans, as well as some other organisms, can have a chromosomal arrangement that is contrary to their phenotypic sex, that is, XX males or XY females. See, for example, XX male syndrome and androgen insensitivity syndrome.

Birds have a similar system of sex determination (ZW sex-determination system), in which it is the females which are heterogametic (ZW), while males are homogametic (ZZ).

Recent studies on the genetic factors which influence gender traits

For a long time, biologists had believed that the female form was the default template for the mammalian foetuses of both sexes. After the discovery of the testis-determining gene SRY, many scientists still believed that the genetic mechanism which determines a foetus to develop into a male form was initiated by the SRY gene, which was thought to be responsible for the production of testosterone and its overall effects on body and brain development. This perspective still shared the classical way of thinking, that in order to produce two sexes, nature has developed a default, female pathway, and an active pathway, by which male genes would initiate the process of determining a male sex, as something which is developed in addition to and based on the default female form. This view is no longer considered accurate by most scientists who study the genetics of sex. In an interview for the *Rediscovering Biology* website, researcher Eric Vilain described how the paradigm changed since the discovery of the SRY gene:

> “ For a long time we thought that SRY would activate a cascade of male genes. It turns out that the sex determination pathway is probably more complicated and SRY may in fact inhibit some anti-male genes.
>
> The idea is instead of having a simplistic mechanism by which you have pro-male genes going all the way to make a male, in fact there is a solid balance between pro-male genes and anti-male genes and if there is a little too much of anti-male genes, there may be a female born and if there is a little too much of pro-male genes then there will be a male born.
>
> We [are] entering this new era in molecular biology of sex determination where it's a more subtle dosage of genes, some pro-males, some pro-females, some anti-males, some anti-females that all interplay with each other rather than a simple linear pathway of genes going one after the other which makes it very fascinating but very complicated to study. ”

In mammals, including humans, the SRY gene is responsible with triggering the development of non-differentiated gonads into testes, rather than ovaries. However, there are cases in which testes can develop in the absence of an SRY gene. In these cases, the SOX9 gene, involved in the development of testes, can induce their development without the aid of SRY. In the absence of SRY and SOX9, no testes can develop and the path is clear for the development of ovaries. Even so, the absence of the SRY gene or the silencing of the SOX9 gene are not enough to trigger sexual differentiation of a foetus in the female direction. A recent finding indicates that ovary development and maintenance is an active process, regulated by the expression of a "pro-female" gene, FOXL2. In an interview for the *TimesOnline* edition, study co-author Robin Lovell-Badge explained the significance of the discovery:

> “ We take it for granted that we maintain the sex we are born with, including whether we have testes or ovaries. But this work shows that the activity of a single gene, FOXL2, is all that prevents adult ovary cells turning into cells found in testes. ”

Implications for human health and social policy

Looking into the genetic determinants of human sex can have wide-ranging consequences. Scientists have been studying different sex determination systems in fruit flies and animal models to be able to understand how the genetics of sexual differentiation can influence biological processes like reproduction, ageing and disease. Since many genetic mechanisms involved in determining sexually dimorphic traits have been preserved from the fruitflies to mice and humans, understanding how these genetic

mechanisms work can lead to improved health care which takes into account sex differences and also to changes in how people understand and perceive sex differences, which can influence how they treat each other in social settings.

Chapter 14

Haplodiploid Sex-Determination System and Temperature-Dependent Sex Determination

Haplodiploid sex-determination system

The haplodiploid sex-determination system is typical of bees and wasps.

The **haplodiploid sex-determination system** determines the sex of the offspring of many hymenopterans (bees, ants, and wasps), spider mites, coleopterans (bark beetles) and rotifers. In this system, sex is determined by the number of sets of chromosomes an individual receives. An offspring formed from the union of a sperm and an egg develops as a **female**, and an unfertilized egg develops as a **male**. This means that the males have half the number of chromosomes that a female has, and are haploid. This haplodiploid sex-determination system produces a number of peculiarities; chief among these is that a male has no father and cannot have sons, but he has a grandfather and can have grandsons. Haplodiploidy is postulated as having paved the way for the evolution of eusociality in the Hymenoptera and a few other taxa although this is a matter of considerable debate.

Mechanisms

Several models have been proposed for the genetic mechanisms of haplodiploid sex-determination. The model most commonly referred to is the *complementary allele model*. According to this model, if an individual is heterozygous for a certain locus, it develops into a female, whereas hemizygous and homozygous individuals develop into males. In other words, diploid offspring develop from fertilized eggs, and are normally female, while haploid offspring develop into males from unfertilized eggs. Diploid males would be infertile, as their cells would not undergo meiosis to form sperm. Therefore the sperm would be diploid, which means that their offspring would be triploid. Since hymenopteran mother and sons share the same genes they may be especially sensitive to inbreeding: Inbreeding reduces the number of different sex alleles present in a population, hence increasing the occurrence of diploid males.

After mating, fertile Hymenopteran females store the sperm in an internal sac called the spermatheca. The mated female controls the release of stored sperm from within the organ: If she releases sperm as an egg passes down the oviduct, the egg is fertilized. Social bees, wasps, and ants can modify sex ratios within colonies to maximize relatedness among members, and to generate a workforce appropriate to surrounding conditions.

Sex-determination in honey bees

In honeybees the drones (males) are entirely derived from the queen, their mother. The diploid queen has 32 chromosomes and the haploid drones have 16 chromosomes. Drones produce sperm cells that contain their entire genome, so the sperm are all genetically identical except for mutations. The genetic makeup of the female worker bees is half derived from the mother, and half from the father, but the male bees' genetic makeup is entirely derived from the mother. Thus, if a queen bee mates with only one drone, any two of her daughters will share, on average, 3/4 of their genes. The diploid queen's genome is recombined for her daughters, but the haploid father's genome is inherited by his daughters "as is".

While workers can lay unfertilized eggs that become their sons, haplodiploid sex-determination system is beneficial to the individual due to indirect selection. Since the worker is more related to the queen's daughters (her sisters) than to her own offspring, helping the queen's offspring to survive aids the spread of the same genes that the worker possesses more efficiently than direct reproduction. Batches of worker bees are short lived and are constantly being replaced by the next batch, so this kin selection is possibly a strategy to ensure the proper working of the hive. However, since queens usually mate with a dozen drones or more, not all workers are full sisters. Due to the separate storage of drone sperm, a specific batch of brood may be more closely related than a specific batch of brood laid at a later date.

Shared gene proportions in haplo-diploid sex-determination system relationships

Sex	Daughter	Son	Mother	Father	Full Sister	Full Brother
Female	1/2	1/2	1/2	1/2	3/4	1/4
Male	1	*N/A*	1	*N/A*	1/2	1/2

Temperature-dependent sex determination

Temperature-dependent sex determination (TSD) is type of environmental sex determination in which the temperature the eggs experience determines the sexes of the organisms that hatch. It is most prevalent and common among amniote vertebrates that are classified under the reptile class, but is also used among some birds, such as the Australian Brush-turkey. It differs from the chromosomal sex-determination systems common among vertebrates. It is a type of environmental sex determination (ESD); in other ESD systems, some factors such as population determine the sex of organisms.

The eggs are affected by the temperature at which they are incubated during the middle one-third of embryonic development. This critical period of incubation is known as the thermosensitive period (TSP). The specific time of sex-commitment is known due to several authors resolving histological chronology of sex differentiation in the gonads of turtles with TSD.

Types

Within the mechanism, two distinct patterns have been discovered and named Pattern I and Pattern II, with Pattern I further divided into IA and IB. Pattern IA has a single transition zone, where eggs predominantly hatch males if incubated below this temperature zone, and predominantly hatch females if incubated above it. Pattern IB also has a single transition zone, but females are produced below it and males above it. Pattern II has two transition zones, with males dominating at intermediate temperatures and

females dominating at both extremes. Very near or at the pivotal temperature of sex determination, mixed sex ratios and, more rarely, intersex individuals are produced.

In turtles with TSD, males are generally produced at lower incubation temperatures than females (TSD IA), with this change occurring over a range of temperatures as little as 1-2 °C. At cooler temperatures ranging between 22.5 and 27 degrees Celsius mostly male turtles arise, and at warmer temperatures around 30 degrees Celsius only female turtles arise. In lizards and crocodilians, this pattern is reversed (TSD IB).

It has been proposed that essentially all modes of TSD are actually type II and those that deviate from the expected female-male-female pattern are simply never exposed to extreme temperature ranges on either end.

Hormones In TSD Systems

Synergism between temperature and hormones has also been identified in these systems. Administering estradiol at male-producing temperatures generates females that are physiologically identical to temperature-produced females. The reverse experiment, males produced at female temperatures, only occurs when a nonaromatizable testosterone or an aromatase inhibitor is administered, indicating that the enzyme responsible for conversion of testosterone to estradiol, aromatase, plays a role in female development.

Interestingly, hormones and temperature show signs of acting in the same pathway, in that less hormone is required to produce a sexual shift as the incubation conditions near the pivotal temperature. It has been proposed that temperature acts on genes coding for such steroidogenic enzymes, and testing of homologous GSD pathways has provided a genic starting point. Yet, the genetic sexual determination pathway in TSD turtles is poorly understood and the controlling mechanism for male or female commitment has not been identified.

While sex hormones have been observed to be influenced by temperature, thus potentially altering sexual phenotypes, specific genes in the gonadal differentiation pathway display temperature influenced expression . In some species such important sex-determining genes as *DMRT1* and those involved in the Wnt signalling pathway could potentially be implicated as genes which provide a mechanism (opening the door for selective forces) for the evolutionary development of TSD.

Adaptive significance

The adaptive significance of TSD is currently not well understood. One possible explanation that TSD is common in amniotes is phylogenetic inertia – TSD is the ancestral condition in this clade and is simply maintained in extant lineages because it is currently adaptively neutral or nearly so. . Indeed, recent phylogenetic comparative analyses imply a single origin for TSD in most amniotes around 300 million years, with several more recent independent origins of TSD in squamates.. Consequently, the adaptive significance of TSD in all but the most recent origins of TSD may have been

obscured by the passage of deep time, with TSD potentially being maintained in many amniote clades simply because it ‘works’ (i.e. has no overall fitness costs along the lines of the phylogenetic inertia explanation).

Other work centers on a 1977 theoretical model (the Charnov–Bull model), predicted that selection should favour TSD over chromosome-based systems when "the developmental environment differentially influences male versus female fitness"; this theoretical model was empirically validated thirty years later but the generality of this hypothesis in reptiles is questioned.
"Temperature sex determination could allow the mother to determine the sex of her offspring by varying the temperature of the nest in which her eggs are incubated. However there is no evidence thus far that sex ratio is manipulated by parental care"