Epigenetic Contributions to the Cancer Transcriptome

DJ Smiraglia and MJ Campbell, Roswell Park Cancer Institute, Buffalo, NY, USA

© 2014 Elsevier Inc. All rights reserved.

Glossary
Acetylation Enzymatic addition of an acetyl group (COCH3) to a protein (histones).
Chromatin structure Higher-level organization of the DNA wrapped around histones.
Chromosomal translocation Rearrangement of chromosomal parts between two nonhomologous chromosomes.
CpG dinucleotide Deoxycytidine triphosphate five prime to deoxyguanosine triphosphate.
CpG islands Stretches of DNA 200 bp or larger having a GC content greater than 50% and an observed CpG/expected CpG greater or equal to 0.6.
Epigenetics Above the genetics. The heritable chemical marks that do not change the coding of the genome, but may change the organization and transcription of the genome.
Hemimethylated DNA methylation state where one strand has methylated CpG, but the other strand is not methylated. Typically found during DNA replication.
Heterodimer Interaction of two different proteins to create a dimer.
Hypermutable A base having a higher tendency for mutation than other bases.
Ligand Small molecule that binds to a receptor molecule such as dihydrotestosterone binding to the androgen receptor.
Methylation Enzymatic addition of a methyl group (CH3) to DNA or protein (histones).
Nascent strand The newly produced strand of DNA during DNA replication created by DNA polymerase as a copy of the template strand.
Pericentromeric region Chromosomal region on either side of the centromere.
Pluripotent Cells with the ability to differentiate into multiple cell types.
Redox status The balance in the cell between oxidants and antioxidants. A measure of oxidative stress.
Template strand DNA strand being copied by DNA polymerase to create the nascent strand of DNA during replication.
Transcriptome The set of all RNA molecules transcribed in a cell, including mRNA, miRNA, rRNA, tRNA, and other noncoding RNAs.

Abbreviations
2-HG 2-Hydroxyglutarate
5mC 5-Methylcytosine
AIB1 Amplified in breast cancer-1
AML Acute myeloid leukemia
AR Androgen receptor
ARID AT-rich interaction domain
BRAF v-Raf murine sarcoma viral oncogene homolog B1
CDKN1A Cyclin-dependent kinase inhibitor 1A
CT Cancer-testis
CTBP C-terminal-binding proteins
DNA Deoxyribonucleic acid
DNMT DNA methyltransferases
ES cell Embryonic stem cell
ETO Eight twenty one protein
EZH2 Enhancer of zeste homolog
FOX P Forkhead box P family
G9a Euchromatic histone lysine N-methyltransferase
H2A H2A histone family
H2B H2B histone family
H3 H3 histone family
H4 H4 histone family
HATs Histone acetyltransferases
HDACs Histone deacetylases
HP1 Heterochromatin protein 1
IAP Intracisternal A-particle
IDH1 Isocitrate dehydrogenase 1
JARID Jumonji/ARID domain-containing protein
KDM1A/LSD1 Lysine (K)-specific demethylase 1A
KDMs Lysine demethylases
KMTs Lysine methyltransferases
LINEs Long interspersed elements
LOH Loss of heterozygosity
MBD Methyl binding domain
MeCP2 Methyl-CpG-binding protein 2
MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
MLL Mixed-lineage leukemia
mTOR Mechanistic target of rapamycin
MYOD1 Myogenic differentiation 1
NAD Nicotinamide adenine dinucleotide
NANOG Homeobox protein NANOG
NCOA Nuclear receptor coactivator
NCOR Nuclear receptor corepressor
NRs Nuclear receptors
p16 Cyclin-dependent kinase inhibitor 2A
PCNA Proliferating cell nuclear antigen
PHD Plant homeodomain
PLZF Promyelocytic leukemia zinc finger
PML Promyelocytic leukemia
Introduction

Every normal human haploid genome has over three billion base pairs of deoxyribonucleic acid (DNA) that encode over 20,000 protein coding genes and a large complement of non-coding genes whose functions are often less well understood. The average length of a base pair of DNA is 0.34 nm long. That means that if the human genome were arranged as one long string, there would be approximately 2 m of DNA in each diploid nucleus. Of course, the genome is not arranged as one long string. It is broken into chromosomes, but more importantly, it is wrapped around nucleosomes and those nucleosomes take on higher-ordered structures, ultimately creating an orderly packaging of the genome allowing it to fit into the nucleus of a cell. This packaging structure is referred to as the chromatin structure. Each time a cell divides, those three billion base pairs need to be unpacked, replicated, and then repackaged into two new nuclei. There are approximately ten trillion nucleated cells in an adult human body – all coming from a single fusion of two haploid genomes in the zygote. This requires tremendous efficiency and accuracy in unpackaging and repackaging of chromatin and coordination with the processes of DNA replication.

Beyond the need to simply protect and fit all those base pairs of DNA in each nucleus, the systems regulating chromatin structure of each cell must be able to package the DNA correctly in order to allow the transcriptional machinery to have access to the required gene coding regions. Each individual cell type has the exact same set of genes encoded by the DNA, but unique combinations of these genes are expressed in a highly regulated cell type-dependent fashion. To a great extent, that is determined by cis-acting elements and directed by specific transcription factors. However, this is also governed by how the DNA is packaged in each cell type, which in part determines which genomic regions are accessible to transcription factors. The regulation of chromatin structure, and therefore the regions of DNA accessible to transcriptional machinery, is accomplished through covalent modifications of both the histone proteins making up the nucleosomes and the DNA itself, namely, DNA methylation. This level of regulation of the genome, which is above (and does not change) the coding potential of the genome, is referred to as epigenetics. By altering the accessibility of DNA regions to the transcriptional machinery, epigenetics contributes significantly to the transcriptome of each cell. This article will explore epigenetic regulation of transcription and how these processes become distorted in cancer to result in abnormal cancer transcriptomes.

Epigenetics: DNA Methylation and the Histone Code

How can an epithelial cell in the duct of a prostate be so different from an endothelial cell of a blood vessel when, in a single person, they share identical genetic codes? The answer, of course, is that a different combination of genes is expressed in the two cell types that create the unique phenotypes of the cells; but how? To an extent, this is determined by the combination of transcription factors, both general and tissue-specific (or cell type-specific), that are available to regulate the correct combination of genes. Epigenetic mechanisms provide a second layer of regulation that in some cases supersedes, or at least juxtaposes with, the cis regulatory elements recognized by transcription factors. Epigenetic mechanisms control what cis elements are accessible to the general transcription machinery and tissue-specific transcription factors. There are many aspects to epigenetics, some more well understood than others. For this article, we will focus on DNA methylation and the most well-studied and well-understood histone modifications.

DNA Methylation

DNA methylation is a chemical modification of the fifth carbon of cytosine adding a methyl (CH3) group in place of hydrogen. The vast majority of this occurs in the context of cytosine (C) five prime (5’) to guanine (G) – in other words, in a CpG context where the 5’-3’ orientation is indicated by the ‘p’ representing the phosphate backbone of the DNA. There are two critical features of this chemical modification of the DNA that define it as the prototypical epigenetic mark. (1) It does not affect the coding of the genome. 5-Methylcytosine (5mC) as part of a codon still encodes for the same amino acid as unmethylated cytosine. 5mC still base pairs with G. (2) This chemical modification is heritable to daughter cells. The 5mC mark is copied to the nascent strand of DNA during DNA replication, thereby perpetuating the mark in daughter cells. These are two defining characteristics of epigenetics – heritable changes that do not affect the genetic code.

The majority (~85%) of CpG sites in the genome are methylated and this is mainly found in repetitive elements such as short interspersed elements and long interspersed elements and satellite DNA repeats in pericentromeric regions. The distribution of CpG in mammalian genomes is unusual. There are 16 possible dinucleotides and so each should be expected to make up approximately 1/16th of the genome. However, CpG is only 1/80th of the genome. This is partially explained by the fact that 5mC is hypermutable to thymine through spontaneous deamination. DNA repair systems can recognize and repair the ensuing G:T
mismatch and correct it to G:C. However, because the mutation event (deamination of 5mC to T) results in a base normally found in DNA, these repair systems are relatively inefficient compared to those that repair some other types of mutations. If not repaired, the result is a transition mutation from a 5mC:T base pair to a T:A base pair after replication. It is thought that over the course of evolution, the mammalian genome has become partially depleted of 5mC due to this process. Two implications come from this: (1) there is probably less evolutionary pressure on the regions of the genome that remain highly methylated since they have withstood the increased mutation rate without apparent harm to fitness, and (2) the genomic regions that remain unmethylated might be under higher evolutionary pressure and therefore must remain unmethylated to avoid the potentially higher mutation rate that in turn would harm fitness. Consistent with these ideas, ∼15% of the CpG sites in the genome that are unmethylated are mostly found in clusters called CpG islands. These CpG islands are defined in part by having nearly the expected frequency of CpG dinucleotides (1/16th). Furthermore, consistent with the idea that these might be evolutionarily important regions, many CpG islands are found in the promoter regions of genes, suggesting that the high mutation rate associated with 5mC might be intolerable. The evolutionary conservation of CpG islands and of the ∼1/16th frequency of CpG within these regions argues for their functional significance.

While most CpG islands in the mammalian genome are unmethylated, the majority of the CpG islands found on the inactive copy of the X chromosome in a female cell are methylated. In addition, imprinted genes, expressed from either the paternal or the maternal allele, are associated with CpG island regions methylated only on one allele, with these regions referred to as differentially methylated regions. In addition, the cancer–testis (CT) antigen genes (such as those of the melanoma antigen, L antigen, and G antigen families) often have methylated CpG island promoters in all normal tissues except the testes. Methylation of CpG islands in the promoter regions of genes is associated with silencing of transcription. Accordingly, CT antigen genes are silent in normal tissues, where their CpG island-containing promoters are methylated, but become expressed in many cancers if those CpG islands lose methylation. The unmethylated state of the majority of CpG island promoters puts them into a transcriptionally permissive structure. This permissive state means that if positive regulatory signals are present, those promoters are available to the basal transcriptional machinery and specific transcription factors. It is critical to also consider that these DNA methylation states are in the context of chromatin and that this chromatin structure is in part being defined by modifications of histones making up the nucleosomes.

**Histone Modifications**

Nucleosomes are made up of four dimers of the histone proteins H2A, H2B, H3, and H4. All of these histone proteins can be posttranslationally modified at specific residues at the amino tail. The amino tails of the histone proteins protrude outside the core of the nucleosome and so are accessible to enzymatic activity and accessible to other proteins that might be either recognized or influenced by the modifications. There are many different types of posttranslational modification of histone tails on multiple amino acid residues, but the majority of epigenetic research has focused on methylation and acetylation of specific lysine residues. Broadly speaking, acetylation of lysine residues in histone tails is associated with a more open chromatin structure where adjacent nucleosomes tend not to cluster together. On the other hand, deacetylated histone tails are associated with a closed chromatin structure where nearby nucleosomes can be tightly packed together making the DNA inaccessible to transcription factors and the basal transcriptional machinery.

The lysine (K) residues at positions 4, 9, 27, and 36 of histone H3 appear to play key regulatory roles in transcriptional control. For example, histone H3 lysine 4 trimethylation (H3K4me3) is found in the promoter regions of actively transcribed genes. This mark is mutually exclusive with H3K9me, which instead is associated with transcriptionally silent promoter regions. Acetylation of H3K9 is found along with the methylation of H3K4 at active promoter regions. An individual lysine in a histone tail can be either methylated or acetylated at any one moment, but not both. Therefore, methylation and acetylation of H3K9 are mutually exclusive. Similarly, H3K27 can be either acetylated or methylated, with acetylation associated with active gene transcription and methylation associated with gene silencing. H3K36 is typically found as acetylated in promoters of active genes but methylated in the bodies of those genes, while K36 is unmodified at silent genes regardless of position in the promoter or gene body. At enhancer regions of active genes, H3K4 monomethylation is found along with acetylation of K9 and K27, while enhancers of silent genes retain H3K4me, but do not have acetylation at K9 and 27.

DNA methylation patterns coexist with these and other histone modifications. In the promoter regions of active genes, a lack of DNA methylation is found and tightly associated with trimethylation of H3K4. This is a typical signature of an active promoter. On the other hand, a silenced promoter will often show high levels of DNA methylation, a complete lack of H3K4 methylation, but a gain of H3K9 or H3K27 trimethylation. Together, DNA methylation and histone modifications collaborate to create the chromatin structure context for gene-containing and noncoding regions of the genome. In broad terms, the genome is broken into heterochromatin and euchromatin. Euchromatic regions are characterized by harboring genes that may be expressed and generally have low levels of DNA methylation and high levels of histone acetylation. Euchromatic regions also have a more open structure, meaning that the adjacent nucleosomes do not pack tightly together, in part due to the high degree of negative charge repelling adjacent nucleosomes from the acetyl groups added to lysines. Conversely, heterochromatin is characterized by higher levels of DNA methylation and lower levels of histone acetylation. Heterochromatin regions do not contain expressed genes and have nucleosomes that are tightly packed, restricting access to the DNA for DNA binding proteins. Other proteins that can recognize some of the histone modifications associated with heterochromatin bind these regions and contribute to maintaining the closed nucleosome structure.

**Writers, Readers, and Erasers of Epigenetic Marks**

Epigenetic marks such as the described DNA methylation, histone methylation, and histone acetylation are chemical
modifications of DNA or histone tails that are created or removed enzymatically. Furthermore, in order for these chemical marks to have a biological function, they must be recognized or read to impart a biological signal. There are large families of enzymes and other proteins that work together to create, interpret, and remove epigenetic signals.

**DNA Methylation**

DNA methylation is written by a family of DNA methyltransferase enzymes (DNMTs) that all have the same carboxy terminal catalytic domain, but different amino terminal domains that affect protein interactions. DNMTs catalyze the addition of a methyl group donated by s-adenosylmethionine (SAM) on the number 5 carbon of cytosine in the context of CpG. DNMT1 is the major DNMT expressed at high levels in all tissues. During DNA replication, DNMT1 is recruited to the replication fork through interaction with proliferating cell nuclear antigen (PCNA). The interaction with PCNA increases the processivity of DNMT1, helping it to follow along the replication fork where it is able to copy the DNA methylation pattern from the template strand, which is already methylated, to the nascent strand that is not methylated. This hemimethylated state is a key signal and binding site for DNMT1. The protein ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1), which also binds DNMT1 at the replication fork, plays an important role in recognizing the hemimethylated state and directing the activity of DNMT1 to these positions. In this way, DNMT1 maintains the DNA methylation pattern from the parent cell to the daughter cell (Figure 1). This heritability of DNA methylation is a key feature defining DNA methylation as an epigenetic mark.

If DNMT1 is responsible for maintaining DNA methylation patterns, what is responsible for establishing new methylation patterns de novo? Two other DNMT family members, DNMT3a and DNMT3b, are often referred to as the de novo methyltransferases. Unlike DNMT1, these enzymes normally methylate DNA that is unmethylated on both strands and do not have binding preference to the hemimethylated state – a feature central to DNMT1’s maintenance function. The de novo methyltransferases are most highly expressed in germ cells and early embryonic development where they are critical to reestablishing the normal DNA methylation pattern after the demethylation of the genome that occurs in these unique settings. The roles of DNMT3a and DNMT3b are not completely redundant. DNMT3a is a distributive enzyme, while DNMT3b is a processive enzyme. These characteristics fit very well with the types of regions they tend to methylate (Figure 2, Animated). DNMT3a is important for focal methylation of either single-copy genes or regions where there are no long stretches of CpG to methylate. On the other hand, the high processivity of DNMT3b is conducive to its role in methylating the highly repetitive pericentromeric regions where there are long stretches with many CpG positions to be methylated. Another DNMT3 family member is DNMT3L, which has no catalytic activity because of having a nonconsensus catalytic domain. Nevertheless, DNMT3L plays an important role in DNA methylation because it interacts with DNMT3a and DNMT3b. The interaction with DNMT3a increases the activity of DNMT3a and has been shown to be essential for maternal imprinting.

DNA methylation can be read by a number of proteins to help translate the chemical signal of the methyl group into the biological function of suppressing transcriptional activity when found in a promoter region. A family of proteins that can uniquely recognize methylated CpG (as opposed to unmethylated CpG) contain a similar domain called the methylation binding domain (MBD). These proteins (MeCP2, MBD1, MBD2, MBD3, and MBD4), along with ZBTB33/KAI3O (zinc finger and BTB domain-containing protein 33), which has a different domain to recognize methylated DNA, are found in complexes that contain other chromatin-modifying enzymes such as histone deacetylases (HDACs). Methyl-CpG-binding protein 2 (MeCP2) binds to methylated DNA and recruits SIN3A (SIN3 transcription regulator homolog A), which in turn recruits HDACs leading to a situation where the regions of DNA methylation coexist with the regions...
of deacetylated histones that can form a compact, closed chromatin structure. The consequence of such a structure is to exclude binding of transcription factors and to prevent interaction with the basal transcriptional machinery. Similarly, MBD2 also binds methylated DNA and is part of the nucleosome-remodeling and histone deacetylase complex that includes an HDAC and Mi-2-like proteins, which are ATP-dependent chromatin-remodeling factors. Therefore, in order to accomplish transcriptional silencing, the chemical signal of DNA methylation is read by specific proteins that can influence the acetylation state of the nearby histones.

Proper control and distribution of DNA methylation are critical to cell and organism development and survival. Genetic knockout of DNMT1, DNMT3a, or DNMT3b was shown to be lethal in mice. The importance of normal regulation of DNA methylation is further illustrated in two examples affecting methylation of repetitive elements. Repetitive elements are typically highly methylated in mammalian genomes. One purpose is to limit the ability of mobile elements to insert randomly in the genome, thereby reducing chances of insertional mutations, and another is to block any potential promoter activity these elements may have. In the mouse genome, the Agouti gene affects hair color, and in some strains of mice, an intracisternal A-particle (IAP), an endogenous retroviral sequence, has inserted upstream of the normal promoter for the Agouti gene. IAPs generally harbor very strong promoter elements and DNA methylation of IAPs silences these promoters. In mice that have the IAP inserted upstream of the promoter, the Agouti gene is properly regulated, expressed only in the skin, and coat color is normal when DNA methylation of the IAP is maintained. However, if the DNA methylation of the IAP is lost, this causes unregulated expression. The loss of methylation of the very strong promoter in the IAP allows it to become active and it takes over the normal control of the Agouti gene. This affects not only expression levels but also in what tissues the Agouti gene is expressed, resulting in ubiquitous expression of the Agouti gene leading to yellow coat color, obesity, and, perhaps surprisingly, tumor development.

A second example of the consequences of abnormal methylation patterns comes in the human condition called immunodeficiency–centromeric instability–facial anomalies syndrome. This autosomal recessive disorder is caused by a point mutation in the DNMT3b gene. This mutation results in hypomethylation (insufficient methylation) of long stretches of satellite DNA repeats high in CpG content found in the pericentromeric regions, affecting their structural stability. The pericentromeric regions are generally highly methylated and associated with deacetylated histones contributing to a compact, heterochromatin structure. Insufficient DNA methylation opens this structure making it less stable and results in radial chromosomes where the arms of chromosomes 1, 9, and 16 fuse together.

H3K4 and H3K9 Modifications

Modifications of the histone tails are epigenetic marks that are more highly dynamic than DNA methylation. Most DNA methylation is very stable once established and contributes to transcriptional silencing and chromosomal stability. The lack of DNA methylation at most CpG island promoters is stably maintained to keep those promoters available to transcriptional regulation. The histone modifications associated with heterochromatin are also highly stable. However, in the euchromatin, where gene expression can be dynamically regulated, for example, through differentiation and in response to environmental cues, histone tail modifications are highly dynamic and contribute significantly to the changing transcriptional state of such genes. Until recently, DNA methylation in these regions was thought to be stable, but new evidence suggests that DNA methylation may also be somewhat dynamic in genes that respond to specific signals.

Many positive-acting transcription factors recruit histone acetyltransferases (HATs) that acetylate H3K9 in the promoter regions to help promote the assembly of the basal transcriptional machinery. The HAT superfamily includes at least 20 different and diverse proteins including CLOCK (circadian locomotor output cycles kaput) and NCOA1. Several sub-families exist, including the p300/CBP (EIA binding protein p300) family (e.g., p300), the GCN5 (K (lysine) acetyltransferase 2A) family (e.g., KAT2A), the MYST family (e.g., MYST1), and SRC/p160 nuclear receptor coactivator family (e.g., NCOA1). The acetylation and positive association with

Figure 2  (Animated) DNMT3a and DNMT3b functions. DNMT3a and DNMT3b are referred to as the de novo methyltransferases because they methylate DNA that is unmethylated on both strands, unlike DNMT1 that methylate hemimethylated DNA. DNMT3a and DNMT3b have differences in their catalytic domains that make DNMT3a a distributive enzyme and DNMT3b a highly processive enzyme with higher intrinsic activity. DNMT3a, therefore, has to load onto the DNA strand multiple times to methylate a stretch of CpGs, while DNMT3b can load once and then move along the strand methylating many CpGs before falling off the DNA strand. These differences in form affect the function of the enzymes with DNMT3a mainly used by the cell to methylate small targeted regions, while DNMT3b is used to methylate long stretches of DNA with high CpG content such as those found in pericentromeric regions.
transcriptional activation are opposed not only by deacetylation through recruitment of HDACs to the regions but also by writers of the methylation marks at H3K9 and H3K27. Eighteen HDACs are known in humans that are classified into four classes based on homology that include the HDAC1–11 and seven sirtuin (SIRT) members. Twenty-eight different lysine methyltransferases (KMTs) are known to act on histones reflecting the importance of the methylated state of key residues for the control of evolutionarily conserved transcriptional programs. Some of the more well-studied H3K9 methyltransferases include C9a (euchromatic histone lysine N-methyltransferase), SUVR9H1/2, and SETDB1/2. These enzymes add one, two, or three methyl groups to H3K9, but can only do so if the lysine has been first deacetylated. Therefore, the recruitment of HDACs to the locus is a prerequisite for addition of the methyl group associated with transcriptional silencing. Conversely, to reactivate the gene, the acetyl groups must be added to H3K9, but first, the methyl groups must be erased. This is done through the action of lysine demethylases (KDMs). At least 20 KDMs are divided into two major groups that include the LSD family members (e.g., KDM1A/LSD1) and the jumonji family (e.g., JmC domain-containing histone demethylation protein 3A and jumonji/ARID domain-containing protein, JARID), proteins containing AT-rich interaction domains (ARIDs). In order to switch a gene back on, after being turned off through deacetylation and methylation of H3K9, the KDMs must be recruited to the locus to remove the methyl groups to allow for the activity of the HATs recruited by the positive-acting transcription factors.

Regions of heterochromatin are much more stable in their retention of transcriptionally repressive chromatin marks. In part, this is through the cointeraction with DNA methylation and histone modifications and proteins that recognize both types of marks. This cooperation helps to cement the transcriptionally repressive state in a more stable way than transient transcriptional repression described in the text earlier. A key player in this cooperation is heterochromatin-associated chromdomain protein 1 (HP1) that is able to read the H3K9 methylation mark and bind to chromatin. HP1 also binds to both an H3K9 methylation writer and a DNA methylation writer, SUV39H1 and DNMT1, respectively, and therefore brings these enzymes to the targeted region. In mouse embryonic fibroblasts that are double null for Suv39H1 and Suv39H2, both writers of the H3K9 methylation mark, there is a significant loss of DNA methylation at major satellite repeats indicating an important link between H3K9 methylation and DNA methylation. Moreover, at least two readers of the DNA methylation mark, MeCP2 and MBD1, have been shown to interact with the writers of the H3K9 methylation mark, SUV39H1 and SETDB1. So, once the DNA methylation mark and/or the H3K9 methylation is set up in a region of heterochromatin, the structure becomes self-reinforcing because both marks are read by readers that recruit writers of the two marks. The presence of DNA methylation leads to binding of MeCP2 and/or MBD1, which recruits H3K9 methylases. On the other hand, the presence of H3K9 methylation leads to binding of HP1, which recruits DNMT1s and H3K9 methylases.

The structure of some of the enzymes writing and erasing epigenetic marks plays a crucial role in the mutual exclusivity of many marks. For example, CpG methylation and H3K4me3 are mutually exclusive marks. Consistent with this observation, the mixed-lineage leukemia (MLL) family of histone methylases, which is responsible for writing the H3K4me3 mark, binds DNA through a CXXC domain that strongly prefers binding to unmethylated CpG. Conversely, DNMT3L, which interacts with and increases the activity of DNMT3a, binds the unmodified version of H3K4 and thereby acts as a sensor of the demethylated state of H3K4. Thereafter, it can recruit DNMT3a and allow for methylation of the CpGs. Similarly, H3K4me3 and H3K9me3 are mutually exclusive. The histone KDM4 plant homeodomain (PHD) finger protein 8 demethylates H3K9me2 and H3K9me1. This enzyme has a PHD-type zinc finger that binds to H3K4me3, which gives the jumonji domain region access to the H3K9me2 allowing for enzymatic removal of the methyl groups at H3K9. Conversely, JARID domain-containing histone KDMs such as JARID1A (KDM5a) have PHD-type zinc finger domains that bind to H3K9me3, allowing access for the demethylase activity of the jumonji domain at H3K4. These are examples of situations where the writer or eraser of an epigenetic mark recognizes the presence (PHD domains recognizing histone methylation states) or absence (CXXC domain of MLL recognizing the lack of CpG methylation or DNMT3L recognizing the lack of H3K4 methylation) of the mutually exclusive mark, thereby ensuring that these marks are read correctly and ensuring their mutual exclusivity.

**H3K27 Methylation**

H3K27 methylation plays an important role in pluripotency and differentiation. As pluripotent stem cells begin to differentiate down a lineage committed path, new sets of genes must be turned on and turned off. Furthermore, maintaining the pluripotent state, which is a unique status for a cell, is dependent on the correct set of genes being activated or repressed. The pluripotent state is significantly different than that of a terminally differentiated cell because this cell must maintain the possibility of turning on multiple unique programs of genes to allow for differentiation down different pathways. This means that the packaging of the genome in pluripotent stem cells has special considerations, namely, maintaining the potential to activate many different genes, a consideration that a terminally differentiated cell does not have. Epigenetic regulation of the genome plays a major role in properly setting up and maintaining this packaging through the action of the polycomb group proteins. The polycomb group proteins were originally identified in *Drosophila* and were found to be critical in exerting repressive control though chromatin modification of homeotic genes responsible for body patterning and differentiation. The mammalian homologs have been identified and shown to play a similar role in regulating the homeobox A, B, C, and D clusters. The mammalian polycomb genes create four polycomb repressive complexes, PRC1–4. One of the core components of PRC2 is the histone H3K27 histone methylase EZH2 (enhancer of zeste homolog). When EZH2 methylates H3K27, this mark creates a docking site for PRC1, which is able to enforce transcriptional repression.

In human embryonic stem cells (ES cells), the components of the PRC2 complex were found near the transcription start
sites of a set of genes that were nearly mutually exclusive with the set of genes that have RNA polymerase II (Pol II), which is part of the basal transcriptional machinery. This is consistent with PRC2 playing an important role in transcriptional repression. Furthermore, H3K27 methylation was found at these same gene promoter regions. A significant fraction of genes harboring PRC2 and H3K27 methylation in ES cells are genes that remain silent in ES cells, but become expressed when ES cells differentiate. Most of these genes encode for proteins and transcription factors that play a critical role in differentiation and development. These are exactly the type of genes that need to be silent in a pluripotent stem cell, but not permanently silenced since they are required to be expressed when the pluripotent cell is driven to differentiate toward lineage commitment. However, the presence of the repressive chromatin mark and structure at these genes in embryonic stem cells is only half of the story. Three key transcription factors, NANOG (homeobox protein NANOG), OCT4 (POU class 5 homeobox 1), and SRY (sex-determining region Y)-box, are essential for stem cells to maintain a pluripotent state. These transcription factors were found at promoters of expressed genes in ES cells as would be expected, but they were also found at the promoters of silenced genes that had PRC2 binding and the H3K27 methylation repressive mark. Furthermore, these genes also had the active epigenetic mark, H3K4 methylation. Therefore, these genes are in a unique bivalent chromatin structure harboring both activating epigenetic marks (H3K4 methylation) and silencing epigenetic marks (H3K27 methylation). This situation puts the genes in a silent but poised state. Notably, these genes do not have H3K9 methylation like that associated with permanently silenced heterochromatin regions. Though there is some DNA methylation found at such genes, it is low-level methylation. Once the pluripotent stem cell receives a signal to differentiate down a specific pathway – muscle cell differentiation, for example – genes critical to that pathway, such as MYOD1 (myogenic differentiation 1), lose binding of PRC2 and lose the repressive H3K27 methylation mark while retaining the active H3K4 methylation mark. This opens the chromatin structure to facilitate the assembly of basal transcriptional machinery to turn the gene on. While the status of MYOD1 changes as the cells commit to muscle differentiation, genes involved in differentiation down other cell lineages, such as the IKAROS gene for the hematopoietic lineage, retain PRC2 binding and H3K27 methylation and remain silenced.

In addition to activating appropriate gene programs of expression needed for differentiation, genes such as Nanog, which are essential to a pluripotent phenotype, need to be repressed in differentiating cells. Once again, chromatin structure and epigenetic regulation play a central role. In mouse ES cells, this occurs in a stepwise manner. First, repressive transcription factor complexes directly bind the Nanog promoter to enforce transient repression. This is similar in process to the transient repression of environmentally responsive genes that may be temporarily silenced until required, and expression is initiated upon receiving a specific signal. Second, the H3K9 methylase G9a is recruited to the Nanog locus along with a complex containing HDACs and H3K4 demethylases. In the second stage, the surrounding nucleosomes become depleted of the acetyl groups that are typically associated with active transcription and methylation at H3K4, also associated with transcriptionally active promoters. This facilitates the addition of the methylation to H3K9, which in turn allows for docking of HP1 and subsequent heterochromatin formation. Thus, in the second stage, the process moves from a highly regulated transient gene repression to a much more stably repressive heterochromatin structure. Lastly, the presence of G9a, as well as HP1, leads to the recruitment of DNMT3b and DNMT1 to allow for DNA methylation of the region and further stabilization of the repressive heterochromatin structure. It is important to note that while it is clear from this stepwise process that DNA methylation is not needed to initiate transcriptional silencing, it plays an important role in stabilizing and maintaining that silencing. The cell is unable to revert back to the pluripotent stem cell-like state after differentiation and silencing of genes like Nanog, in part because of the DNA methylation stabilizing the silencing of genes like Nanog. Experimentally blocking the acquisition of DNA methylation does not prevent the silencing of these genes. However, without the presence of DNA methylation, these genes are able to be reactivated and the pluripotent phenotype can be reattained.

**Distortion of Epigenetic Enzymes in Cancer**

Given that each histone modification is governed by antagonistic groups of enzymes, it is not surprising that altered gene expression patterns in disease such as cancer are frequently associated with altered expression and activity of these enzymes. Similarly, the expression of DNMTs and other components that govern DNA methylation is frequently altered. The list of enzymes identified to be involved in setting and removing histone modifications is increasingly numerous. Many of these enzymes are cofactors or binding partners for transcription factors. Alternatively, transcription regulatory factors can contain intrinsic histone-modifying capacity. It is also apparent that some histone modifications are governed either on a larger chromosomal scale or even globally, whereas other modifications have a much more restricted pattern.

Two points are particularly important in considering the extent of redistribution and altered patterns of histone modifications in cancer. The first is that the steady-state level of each modification represents a dynamic balance between the effects of the modifying and demodifying enzymes, with turnover likely to vary from one part of the genome to another and between cell types. Furthermore, such turnover is intimately associated with cell cycle status, cell-cell interactions, and cell lineage commitment. Secondly, many, if not all, of the enzymes are dependent upon, or influenced by, metabolites and components present in the intra- or extracellular environment. At the simplest level, many of these enzymes depend on cofactors such as acetyl-CoA, NAD (nicotinamide adenine dinucleotide), and s-adenosylmethionine for their activity, and in turn, these levels will depend on the metabolic or redox state of the cell. More subtle effects can be derived from metabolism. For example, naturally occurring short-chain fatty acids are inhibitors of class I HDACs and nicotinamide is an inhibitor of the NAD-dependent deacetylase SIRT1 and can be derived intrinsically within a cell or tissue.
Epigenetic processes are central to sustaining and regulating transcriptional plasticity – that is, the choice, amplitude, and period of target gene expression. The interactions between histones and DNA are frequent, intimate, and varied due to the high number of posttranslational modifications governed by multiple enzyme families. Steady-state levels of histone modifications represent a dynamic balance between antagonistic enzymes. Furthermore, key histone modifications form a template for attracting and directing the DNA methylation machinery to CpG regions. Nuclear receptors (NRs) form a bridge linking extrinsic or intrinsic signals to chromatin modifications resulting in cyclical transcriptional outputs at target loci. The dynamic regulation of histone modifications in response to NR ligand binding has been well documented. However, the potential of DNA methylation changes in this dynamic setting has received relatively little attention.

Dynamic control of transcription of NR-inducible genes is an important concept that illustrates the finely tuned dynamic equilibrium between epigenetic-modifying enzymes that write and remove marks associated with transcriptionally active or inactive states. This is illustrated in (Figure 3, Animated). The addition of ligand for the NR causes a shift in the dynamic equilibrium by altering the local recruitment of enzymatic activities. NRs cycle between large complexes to either activate or repress target gene transcription causing the locus to rapidly switch between a transcriptionally silent but poised state to a state conducive to high-efficiency transcription. The switching between complexes is highly dynamic, in part, leading to oscillating target gene expression. Repression arises due to interaction with corepressors, which are associated with HDACs and histone demethylases (KDMs) that condense chromatin structure at target genes. Activating complexes, on the other hand, include histone acetylase activity (HAT) and KDMs that remove repressive methyl marks on H3K9 and H3K27. This demethylation allows for acetylation of H3K9 and for methylation of H3K4. This chromatin structure is permissive to transcriptional machinery binding and elongation. DNA methylation is also altered in these dynamic processes with increased DNA methylation correlating with loss of H3K4 methylation and gain of H3K9 methylation. Conversely, the removal of the methyl group from CpG dinucleotides coincides with gain of H3K4 methylation and loss of H3K9 methylation. These dynamic changes in response to ligand-bound NRs represent active epigenetic regulation of transcriptional response. Distortions to these dynamic processes have the potential to push the dynamic equilibrium toward one extreme or the other, potentially leading to a locus acquiring a permanently open structure facilitating high-level transcription or a permanently closed, heterochromatic structure leading to gene silencing.

Figure 3 (Animated) Epigenetic dynamics and regulation of transcription. For genes that respond to stimuli such as ligand-bound nuclear receptor transcription factors, the rate of transcription oscillates between high- and low-level rates of firing from the promoter. The epigenetic state of the promoter including acetylation and methylation of histones and methylation of DNA also oscillates and contributes to the oscillating transcription rate. As the transcription rate increases, there is a concomitant increase in histone acetylation and H3K4 methylation (active epigenetic marks) and a decrease in H3K9 methylation and DNA methylation (inactive epigenetic marks). The dynamic equilibrium of these epigenetic marks is determined by the relative activity of readers and writes of the marks, which is in part determined by the recruitment of transcription factors. When ligand is added to the system, this shifts the dynamic equilibrium toward active transcription. In the cancer setting, these processes can be distorted at multiple levels, ultimately leading toward pushing strongly the dynamic equilibrium to one extreme or the other. This can result in an epigenetically regulated over expression or silencing of a gene.
The effects of metabolic changes on gene expression are a strongly reemergent area in cancer biology. In 1927, Otto Warburg suggested that tumors derive energy primarily from the conversion of glucose to lactic acid and only partially through cellular respiration. Subsequently, in the 1950s, he went on to propose that all causes of cancer were manifestations of mechanisms that disabled cellular respiration in favor of fermentation (now termed aerobic glycolysis). This counterintuitive shift in dependence away from respiration (oxidative phosphorylation) toward a metabolically expensive and wasteful utilization of glucose exemplifies tumor biology. As a result, the role of aberrant glucose metabolism in cancer has become a central tenant of the cancer paradigm. The inefficient production of ATP via glucose conversion to lactic acid requires vastly increased glycolytic flux; this distortion is exploited with positron emission tomography scans to visualize tumors. Furthermore, mitochondrial metabolic pathways are central to creating and maintaining the acetyl-CoA pools through the conversion of pyruvate to acetyl-CoA during oxidation of glucose or β-oxidation of fatty acids. The availability of the acetyl-CoA pools affects the ability of HATs to acetylate histones. SAM pools, essential for both DNA and histone methylation, are also affected by mitochondrial metabolism through the generation of serine, which is the primary source of one-carbon units (through conversion to glycine) fueling one-carbon metabolism. Thus, the nucleosome, through the array of histone modifications it carries and the enzymes that put them in place, is a finely tuned sensor of the metabolic state of the cell and the composition of its environment. In this manner, nucleosome structure provides a platform through which external environmental and internal variables can influence genomic function.

Finally, the enzymes that govern histone methylation are also distorted in cancer with both loss of function and gain of function. Expression patterns of histone-modifying enzymes are even able to discriminate between tumor samples and their normal counterparts and cluster the tumor samples according to cell type. Candidate studies of specific enzymes have suggested that expression changes are significant in different cancers. For example, overexpression of G9a, a H3K9 KMT, occurs in lung and breast cancers and associates with aggressiveness. However, enzymes that deacetylate H3K9 are also overexpressed in cancers, including breast cancer, and suggest that the targeting to specific sets of genes is critical to the carcinogenic role played by these enzymes. More recently, The Cancer Genome Atlas has comprehensively profiled the cancer genomes in multiple tumors and identified that multiple components of key complexes in various chromatin-remodeling complexes are altered by mutation and expression changes. Therefore, the precise specificity of a distorted epigenetic complex in cancer is most likely dependent on the combination of both the enzyme complex and recruitment to target gene(s). For example, the mutation of KDM6A/UTX results in the inability to relieve H3K37me3 repression. Gain of function also occurs, for example, increased targeting of methyltransferases KMT1A/SUZ12 to CDKN1A (cyclin-dependent kinase inhibitor 1A) leads to sustained H3K9me2 and transcriptional silencing. Similarly, the KMTs/MLLs are overexpressed in prostate cancer and sustain levels of H3K27me3 at key targets such as DAP21B, a RAS regulatory molecule, thereby leading to metastasis. These observations illustrate deregulation of the enzymes that control histone lysine methylation is common but most likely highly targeted. This contextual nature is typified by KDM1A/LSD1, which can target the demethylation of either H3K9me2 or H3K4me2 and thereby drive both gene activation and repression by the androgen receptor (AR).

### Interplay Between Altered Transcriptional Signals and Epigenetic States in Cancer

Epigenetic regulatory mechanisms operate in response to signals from the cellular microenvironment of the tumor, including signals from associated stromal or immune cells. Thus, the niche in which cells find themselves is an important determinant of their epigenetic properties and raises the possibility that histone marks can be modified by environmental conditions that alter metabolic and redox status, leading to a heritable alteration in cell phenotype. Such epigenetic disruption is not restricted to single nucleotides but rather can be spread over larger regions and therefore comparable to deletions and amplifications in terms of scale of the effect. They can act alongside conventional genetic and cytogenetic alterations, either inherited or de novo, to cause the biallelic silencing of tumor suppressor genes that can be the first step in the development of a cancer.

The NR superfamily also illustrates the key concepts of distorted and selected transcription in cancer due to altered regulation of histone modifications. NRs are the largest superfamily of transcription factors in humans and generally form active heterodimers to control networks that regulate homeostasis, energy metabolism, and xenobiotic handling. These receptors are intimately associated with the control of self-renewal in a number of epithelial systems, notably the prostate and mammary glands. For example, studies in the prostate have established that the AR cooperates with WNT (wingless-type MMTV integration site family) and mechanistic target of rapamycin pathways to induce proliferation. Equally, other receptors, such as VDRs (vitamin D receptors), PPARs (peroxisome proliferator-activated receptors), and RARs (retinoic acid receptors), exert mitotic restraint at least in part by antagonizing WNT signaling and activation of cell cycle arrest through the regulation of gene targets such as CDKN1A (encodes p21<sup>WAF1/CIP1</sup>) and IGBP3 (insulin-like growth factor-binding protein 3).

Cancer is typified by the actions of individual receptors becoming selective and the NR network collectively displaying a loss of transcriptional plasticity. The AR transcriptional program evolves toward increased targeting of proliferative gene promoters and decreased targeting of prodifferentiation genes. Similarly, within breast cancer, the transcriptional actions of the ERα appear to become increasingly selective for gene targets associated with proliferation and survival, such as cyclin D1, and away from targets associated with differentiation, such as the progesterone receptor. Equally, in a range of solid tumors and myeloid leukemia, NRs that normally exert mitotic restraint, such as the VDR, RARs, and PPARs, become skewed, with selective silencing of antiproliferative target genes. Combined, oncogenic transcriptional rigidity reflects the simultaneous distorted regulation of target loci such that proliferative and survival signals are enhanced and antimitotic inputs are either limited or...
lost. This filtering of transcriptional choices during cancer progression has significant therapeutic implications.

More recently, genome-wide approaches have revealed considerable variability in the networks of interactions capable of bringing about varied transcriptional responses. For example, in prostate cancer, as the disease progresses, there are altered levels of H3K4me1 and H3K4me2 on gene enhancer regions in the so-called AR-independent state, where cells have evolved resistance to antiandrogen therapies. In this new state, the targeted increase of H3K4me1 and H3K4me2 at different enhancer regions allows the cells to initiate a different AR transcriptional program.

Taken together, these findings support the concept that the actions of major transcription factor families are selective at several levels to govern the expression of subtranscriptomes that are phenotypically related. The flexibility of transcriptional actions includes the exact choice of target sequence, the timing, and the amplitude and magnitude of transcription and integration with other transcriptional programs and signal transduction events. In malignancy, the dexterity of targeting and regulation is blunted and instead transcription factors become addicted to specific subtranscriptomes, for example, those associated with blockade of programmed cell death and progression through the cell cycle.

**Loss of Function and Gain of Function of Transcriptional Coactivators and Corepressors**

One manner by which transcriptional actions are distorted is through the altered expression of associated cofactors that have either an intrinsic or associated capacity to regulate histone modifications. The diversity of coactivator and corepressors is extreme. Several examples are strongly illustrative of underlying mechanisms of how altering the capacity of transcriptional regulation occurs. In essence, the altered expression and function of these key proteins alters the equilibrium of key histone modifications and therefore allows the gene regulatory actions of a given transcription factor to be more or less pronounced.

Both coactivators and corepressors display loss of function and gain of function and can result in similar phenotypes. The loss of a coactivator can lead to suppressed ability of a transcription factor to transactivate a given target. Similarly, the gain of function of corepressors can also limit transactivation ability and enhance transrepression. The opposite patterns will in turn enhance the transactivation function. For example, NCOA3/SRC3 is situated within a common area of chromosomal amplification in breast cancer on chromosome 20q. Initially, a cDNA was isolated from this region that contained a putative target gene that was termed AIB1 (amplified in breast cancer-1). Subsequently, this gene was found to be a member of the SRC coactivator family and was amplified and overexpressed in breast and ovarian cancer cell lines and in breast cancer biopsies. NCOA3/SRC3/AIB1 interacts with ERs in a ligand-dependent fashion and enhances the regulation of target genes. Specifically, the protein has intrinsic HAT activity and also acts to recruit other CBP/p300 in an allosteric manner. Therefore, elevated expression increases the ability of the ERs to transactivate a given gene target. Additionally, NCOA3 has been found to be a potent histone acetyltransferase able to enhance the function of multiple NRs.

Compared to their coactivator cousins, the corepressors are somewhat underexplored. Again, key proteins, originally identified for their repressive interactions with NRs, illustrate how their deregulated function alters chromatin and attenuates gene regulation. NCOR1 and NCOR2/SMRT were cloned in 1995 using NR as bait. Both proteins exist in large multimeric complexes (~2.0 MDa) with histone deacetylases and other histone-modifying enzymes. These complexes are recruited by many different transcription factors to repress gene activity including NRs, MAD/MXI (MAX interactor 1, dimerization protein), MYOD, ETO (eight twenty one protein), CBP (core-binding factor, runt domain), FOXP (forkhead box P family), AP-1 (activator protein 1), and NF-xB (nuclear factor kappa-B DNA binding subunit) factors. The importance of targeted basal repression by corepressors is evident in the lethality of the Ncor1−/− and Ncor2/Smrt−/− mice. These models reveal enhanced function of transcription factors, notably Pparα in adipocytes and FoxP1 in cardiomyocytes.

Well-established oncogenic roles for NCOR1 and NCOR2/SMRT have been elucidated in acute promyelocytic leukemia (APL) that results from a fusion between the NR, RARs, and either the PML or promyelocytic leukemia zinc finger (PLZF) genes. Both chimeric proteins sustain NCOR1 interactions and consequently RARs-mediated cell differentiation is blocked, in part, as a result of maintaining a condensed chromatin structure around the promoters of RAR target genes that govern normal hematopoietic differentiation. In the PML–RAR fusion, this can be overcome by pharmacological dosing with retinoic acid. The PLZF–RAR fusion is resistant to retinoic acid alone and treatment with a combination of retinoic acid and HDAC inhibitors has shown promising results. Similarly, in acute myeloid leukemia (AML), the AML1/ETO fusion protein promotes leukemogenesis by recruiting NCOR1 and again impeding transcriptional regulation. The importance of NCOR1 binding in the treatment of these disease states exemplifies the relevance of the corepressors in firstly driving critical oncogenic events but secondly providing a rational targeted strategy toward HDACs.

Expression profiling in solid tumors has revealed altered NCOR1 and NCOR2/SMRT expression and localization, for example, in breast, bladder, and prostate cancers. However, to date, uncertainty remains over their precise role in solid tumors, especially in the case of breast and prostate cancers where the etiology of disease is intimately driven by the actions of steroid hormone NRs. Indeed, the ability of the ligand-free NR conformation to bind NCOR1 and NCOR2/SMRT is central to therapeutic exploitation with receptor antagonists such as tamoxifen in the case of breast cancer. Therefore, ambiguity exists over the extent and timing of NCOR1 and NCOR2/SMRT expression changes, as they relate to initiation and progression of disease. Secondly, it remains unclear how changes in NCOR1 and NCOR2/SMRT expression relate to different NRs and other transcription factors that exert either promitotic or antimitotic and survival effects.

Similar contextual aspects to function have emerged for other corepressors. For example, C-terminal-binding proteins (CTBP1 and CTBP2) colocalize with polycomb group complexes, thereby leading to gene silencing. CTBP1 also displays dehydrogenase activity and appears to serve as a cellular redox sensor with repression being acutely sensitive to NAD+ /NADH levels. CTBP1 is implicated in the transrepression...
events mediated by ERα and, again, its deregulation suggests a role that distorts and restricts the normal flexibility of receptor transcription. For example, CTBP1 was found to interact in vitro and in vivo with the zinc finger protein ZNF366 and mediate selective estrogen-dependent and HDAC-mediated repression of ERα target genes in breast cancer cells. Similarly, the REST corepressor (RCOR/CoREST) can block the expression of key target genes and, in turn, is part of a complex with HDAC1, HDAC2, BHC80 (BRAF35–HDAC complex protein), BRAF35, and the histone demethylase KDM1/LSD1.

**Higher-Order Chromatin Interactions Associated with Transcription**

Another theme that has emerged concerning epigenetic regulation of transcription is higher-order chromatin interactions. It seems that large-scale chromatin rearrangement, through looping, is frequent and widespread, can be inter- or intrachromosomal, and is guided by transcription factors, key pioneer factors, and chromatin-modifying enzymes. Improved microscopic techniques have recently shown nascent RNA on the surface of protein-dense transcription factories (gene hubs) that seem to correspond to structures previously termed nuclear speckles.

A clear example of these interactions has been illustrated in the transcriptional responses of B cells where translocation of genes occurs from separate chromosomes and nuclear regions to common sites referred to as transcription factories. These sites contain significant levels of RNA Pol II and other proteins, including factors required for elongation, chromatin remodeling, capping, splicing, and nonsense-mediated decay. Recruitment of genes to transcription factories is highly selective, with certain genes and chromosome regions colocalizing far more frequently than expected by chance. Intriguingly, the sites of chromosome translocation associated with various cancers often colocalize. For example, Myc (v-Myc myelocytomatosis viral oncogene homolog (avian)) and Igh (immunoglobulin heavy locus) tend to colocallize and their fusion, in human lymphoid cells, is a common cause of Burkitt’s lymphoma. These rapid movements are associated with movements of the nuclear architecture and involve ATP-dependent mechanisms that involve a chromosome locus usually located at the nuclear periphery being rapidly translocated to the interior in a direction perpendicular to the nuclear membrane.

Defining the distributions of histone modifications, DNA methylation patterns, gene expression, and even higher-order chromatin structures in cell line models is being organized by research consortia, for example, ENCODE (Encyclopedia of DNA Elements). These integrated genome-wide data sets support the idea that the distribution of histone marks is strongly associated with genomic architecture, such as gene and enhancer regions, and is interwoven with DNA methylation to control gene expression.

**Genetic Contribution to Epigenetic Distortion of Transcription in Cancer**

By the mid to late 1990s, we knew a great deal about how genetics contributes to cancer with the discovery of many proto-oncogenes that become amplified or mutated into an oncogenic form and of tumor suppressor genes that become mutated or deleted. With the discovery that tumor suppressor genes sometimes have CpG island hypermethylation in cancers and that this correlated with silencing of the gene came the idea that epigenetics might play an important role in carcinogenesis, perhaps as large a role as genetic abnormalities. The fact that pharmacological inhibition of the DNMTs could sometimes reverse CpG island hypermethylation and lead to reexpression of the gene not only opened up new therapeutic potential but also supported the idea of a causative relationship between hypermethylation and gene silencing. By the early 2000s, the field had discovered that CpG island methylation was widespread throughout cancer genomes. This has been borne out with newer and more comprehensive techniques to study DNA methylation in cancers including next-generation sequencing-based methodologies. We now know that many CpG islands gain hypermethylation in cancer – some at very high frequency, some in many types of cancer, and some in very specific types of cancer. We also know that many repetitive elements that are normally methylated become hypomethylated in cancer. These basic observations of the shifting DNA methylation pattern in cancer are illustrated in (Figure 4, Animated). A number of basic questions remain about the significance of epigenetic changes in cancer. How do we define which epigenetic events can really drive the disease in the same sense that expression of oncogenic RAS can drive it? How much of the milieu of epigenomic abnormality is important to the phenotype and how much is just along for the ride because it occurred, but caused no selective disadvantage to the cancer? Are epigenetic events positively selected for and to what degree do genetic and epigenetic events collaborate for the malignant phenotype?

Some specific examples can begin to shed light on some of these questions. In a study of 50 pancreatic ductal carcinomas, researchers found that 24 had homozygous deletion of the p16 (cyclin-dependent kinase inhibitor 2A) tumor suppressor gene and 18 more showed loss of heterozygosity (LOH) with loss-of-function mutations in the remaining allele. While seven others had LOH at the locus, but were wild-type for the remaining allele, and one sample showed retention of heterozygosity and was wild-type at both alleles of p16. When CpG island methylation of the p16 promoter was investigated, 7 of the 50 cases showed aberrant CpG island methylation. All seven were cases where a wild-type allele of p16 was retained (out of eight). Strikingly, DNA methylation was not found at any of the 18 cases with a mutant p16 allele. Not only do these findings demonstrate that CpG island methylation of a promoter can both substitute for, and collaborate with, loss-of-function genetic mutations by resulting in silencing of the gene, but they also strongly support the idea that positive selection is a driving force for aberrant CpG island methylation since no mutated alleles became methylated. Methylating a tumor suppressor gene allele that harbors a loss-of-function mutation would be redundant and would therefore not have any selective pressure to drive its occurrence. However, methylating the one remaining wild-type allele would be strongly selected for as this would lead to a complete loss of function of the tumor suppressor. The HCT116 human colon cancer cell line provides a similar example where both the p16 gene and the DNA repair gene MLH1 (mutL homolog 1, colon cancer, nonpolyposis type 2...
(E. coli) have one mutated allele and methylation of the non-mutated allele. In this case, within the same cell, there are two genes where both genetic and epigenetic mechanisms are collaborating to enforce tumor suppressor gene loss of function.

That genetic and epigenetic mechanisms in carcinogenesis are inextricably intertwined is an unavoidable conclusion when one considers that the readers, writers, and erasers of epigenetic marks are genetically encoded and therefore subject to mutation. Conversely, the stewards of genomic integrity are also genetically encoded and are subject to epigenetic regulation. As new sequencing technologies make it feasible to sequence all genes in cancers from individual patients, we are beginning to build a broad picture of what genes become mutated in cancer and at what frequency. Indeed, summary data from The Cancer Genome Atlas project have now revealed that genes that encode for proteins that govern chromatin modifications are mutated in multiple cancers. Furthermore, the frequency and distribution of these mutations and other genetic aberrations strongly support roles as tumor suppressors and oncogenes. Similarly, these same technologies are making it possible to assess DNA methylation patterns genome-wide and patterns of other epigenetic marks such as H3K9 methylation and H3K27 methylation. Through such efforts, it has become apparent that many genes known to play a role in epigenetic regulation can become mutated, thereby proving a genetic route to affect epigenetics.

The MLH1 gene is a mismatch repair gene that plays an important role in genome stability by helping to maintain the integrity of microsatellite repeats (short two to six nucleotide tandem repeats also known as simple sequence repeats). Inheriting a mutant form of MLH1 results in greatly increased risk for hereditary nonpolyposis colorectal cancer. The sporadic form of colorectal cancer is also associated with loss of function of MLH1, most often through CpG island methylation in the promoter of MLH1. Loss of function of MLH1 causes microsatellite instability and, in the case of some colorectal cancers, results in the mutation of the transforming growth factor beta receptor II (70/80 kDa) gene, which contains a microsatellite repeat. This mutation is a driver event in colorectal cancer that is caused by loss of function of MLH1 through either genetic or epigenetic (or both) mechanisms. Furthermore, colorectal cancers with methylation of MLH1 and with the V600E mutation of the BRAF (v-raf murine sarcoma viral oncogene homolog B1) oncogene are very likely to have a strong CpG island methylator phenotype (CIMP). CIMP-positive colorectal cancers have exceptionally high rates of CpG island hypermethylation compared to CIMP-negative tumors.

An example of how genetic mutation can impact epigenetics can be seen in gliomas where there is a high frequency of a gain-of-function mutation of the isocitrate dehydrogenase (IDH1) gene. IDH1 catalyzes the oxidative decarboxylation of isocitrate to α-ketoglutarate. The R132H mutation of IDH1 results in
the enzyme producing 2-hydroxyglutarate (2-HG) instead of α-ketoglutarate. This innocuous sounding metabolic effect from a mutant enzymatic activity can have profound effects on the DNA methylation pattern because 2-HG inhibits the activity of the ten eleven translocation enzymes (TEL1–3). The TET enzymes induce demethylation of CpG by oxidizing the 5mC to 5-hydroxymethylcytosine. This form is further oxidized to 5-formylcytosine and to 5-carboxylcytosine, both of which can be recognized by the base excision repair pathway enzymes. Thymine–DNA glycosylase, or other glycosylases, excises this damaged base and then an unmethylated cytosine is replaced, resulting in demethylation at that position. The TET enzymes not only are inhibited by 2-HG, which is the new metabolite produced by mutant IDH1, but also require α-ketoglutarate for activity, which is the metabolite no longer being produced by the mutant IDH1. Therefore, IDH1 mutation in glioma results in reduced ability to demethylate CpG via the oxidation of 5mC. This translates into a greatly increased degree of CpG island hypermethylation in gliomas that have IDH1 mutations compared to those that do not. Furthermore, this difference in the genetic and epigenetic basis of the disease has biological significance because patients with IDH1 mutant glioma present at a younger age and have significantly improved outcomes. Therefore, this provides an example of how a single genetic mutation can have profound effects on epigenetics and that this translates into different biology of the disease, thereby demonstrating the functional significance of the epigenetic effects. This also highlights again the intimate link between epigenetics and metabolism. Epigenetic modifications such as DNA methylation and histone methylation and acetylation are dependent upon metabolite pools for methyl groups (SAM) and acetyl groups (acetyl Co-A). Therefore, perturbations in metabolism during carcinogenesis can impact epigenetics.

Lastly, mutations in the histones themselves can result in globally altered histone modification patterns. Histone H3 has a variant form of the protein (histone H3.3) encoded by the H3F3A (H3 histone, family 3A) gene. This gene was recently found to be mutated at high frequency in a pediatric brain cancer called diffuse intrinsic pontine glioma, with the mutation resulting in the lysine (K) residue at position 27 being replaced with a methionine (M) (H3.3K27M). This mutant histone variant protein was found to account for between 4% and 18% of the total histone H3 in these cancers. As discussed in the preceding text, H3K27 is a target of methylation by the enzyme EZH2, and this plays an important role in epigenetic regulation of transcription. The mutant H3.3K27M cannot be methylated at M27. Moreover, the presence of this mutant protein, even when accounting for as little as 4% of the total histone H3 in the cell, dramatically reduces the global levels of H3K27 methylation. This mutant histone variant acts as a dominant negative mutation by blocking the enzymatic activity of EZH2. Thus, once again, there is a situation where mutation of a single gene can lead to global changes in the epigenetic patterns of the genome.

Conclusions

Epigenetic distortion of transcription in cancer occurs at many levels. The scope of epigenetic changes in cancer cells is vast and a major research challenge is to understand which groups of changes have functional significance and alter the biology of the disease. Furthermore, because epigenetic changes can be combinatorial in their effect, the complexity of epigenetic distortion in cancer far exceeds that of genetic distortion. This adds greatly to the potential heterogeneity within a single tumor. For example, while an early deletion of a tumor suppressor gene may be a driving genetic factor in the early development of the disease, it is highly unlikely to revert later in disease progression or within a small subpopulation of cells, when its loss might no longer be a driving factor in the survival and proliferation of the advanced disease. Epigenetic silencing of the same tumor suppressor gene, on the other hand, is evidently reversible. Therefore, various subpopulations may reacquire expression of the gene and this reexpression might in fact be beneficial as the disease progresses and the cancer cells are faced with evolving challenges, such as immune challenge or growth in a new environment. This potential reversibility of epigenetic changes poses incredible challenges to researchers to understand the full scope and biological significance of these changes. However, it also offers a unique therapeutic opportunity. If the silencing of a tumor suppressor gene is epigenetic, rather than genetic, then therapeutic options might exist to inhibit the accumulation and maintenance of the epigenetic marks responsible for gene silencing, which might then lead to reactivation of the tumor suppressor gene. In fact, such epigenetic therapies currently are in use in some specific situations where inhibitors of DNMTs and HDACs are used with the idea of leading to reactivation of epigenetically silenced genes. Increasing the efficiency and, perhaps more importantly, the specificity of epigenetic-based therapy holds considerable promise for improved treatments for cancer.

Further Reading

Epigenetics: DNA Methylation and the Histone Code

Writers, Readers, and Erasers of Epigenetic Marks

Distortion of Epigenetic Enzymes in Cancer
Interplay Between Altered Transcriptional Signals and Epigenetic States in Cancer

Loss-of-Function and Gain-of-Function of Transcriptional Co-activators and Co-repressors

Higher-Order Chromatin Interactions Associated with Transcription

Genetic Contribution to Epigenetic Distortion of Transcription in Cancer