Epigenetic Mechanisms in Human Disease

Andrew P. Feinberg, Mitsuo Oshimura, and J. Carl Barrett

If one accesses the genome websites, either the publicly or privately funded, and examines the information, one will find only the four letters A, G, T, and C in various combinations. However, one will not find a fifth letter we know to exist, namely C for 5-methylcytosine. For cancer research, this is a significant issue as alterations in DNA methylation in human cancer were first discovered in 1983 (1), and since that time, hundreds of labs have examined epigenetic alterations in human tumors and their role both in the activation of tumor promoter genes as well as silencing of tumor suppressor genes. Although epigenetic modifications are not shown in the databases, a clue within those databases to the importance of epigenetics comes from the comparison of the published human and mouse sequences. This analysis reveals that ~1.5% of the conserved sequences in both genomes represent the coding sequences of genes, but another 1.5% of the genome represents conserved noncoding elements that may be within introns or at a considerable distance from the genes themselves. For the most part, these sequences are GC-rich and therefore potential targets of methylation. Furthermore, we are only beginning to understand sequences involved in epigenetic regulation of gene expression. Indeed, most of our efforts that have identified the molecular basis of disease have been focused upon the genes themselves rather than upon these regulatory sequences.

It should therefore be no surprise that epigenetic mechanisms are found to play a role in many human diseases, including cancer, but also that epigenetic mechanisms lie at the very heart of our understanding of stem cell therapy, animal cloning, complex traits, and aging. After all, what is the difference between a somatic cell and an early embryonic cell other than epigenetic, given that their genomes are the same, but their function and developmental potential are quite different and yet stably inherited as the cells divide. An intensive two-day conference sponsored by the Center for Cancer Research of the National Cancer Institute in Bethesda, Maryland, brought together an outstanding group of investigators from eight countries studying epigenetic mechanisms in development and disease. To our knowledge, it was the first such conference that addressed epigenetic mechanisms of human disease generally, not only cancer, and the proceedings are available in their entirety online.

Disease Mechanisms Involving Chromatin-modifying Genes

Mutations in several chromatin-modifying genes are directly involved in human diseases. One of these is Rett syndrome, which Zoghbi and colleagues showed is caused by mutations in the methylcytosine-binding protein-encoding gene \textit{MECP2} (2). Rett syndrome is an X-linked disorder limited to females whose etiology had eluded investigators for decades. Zoghbi and colleagues using linkage analysis localized and screened a number of genes and were surprised to find \textit{MECP2} mutations. Rett syndrome is limited to females because the null phenotype is lethal. One would expect a chromatin-modifying gene to cause developmental malformations or be an embryonic lethal. However, in the case of Rett syndrome, girls are born completely normally but gradually lose their ability to speak and walk, and they develop stereotypic movement disorders followed by progressive dementia and death. Zoghbi now reports that the protein is highly expressed in the brain, and its appearance correlates with neuronal maturation. A mouse model also appears normal until 6 weeks of age. One of the most interesting questions in Rett syndrome is whether the silencing of genes associated with chromatin may also be involved in the silencing of neuronal activity that is essential for normal brain maturation.

Another fascinating disorder involving chromatin-modifying genes is ICF syndrome, which stands for immunodeficiency, chromosome instability, and facial anomalies. Like the other syndromes, the phenotype is surprising. Evani Viegas-Pequignot and colleagues found that ICF is caused by mutations in the de novo DNA methyltransferase \textit{DNMT3B} (3). The patients show decondensation of centromeric heterochromatin similar to cells treated with 5-aza-deoxycytidine and similar to cytogenetic abnormalities in cancer cells, although the patients do not appear to be at increased risk of cancer. An important lesson regarding the effect of methylation comes from studies of the inactive X chromosome, which while variably demethylated in ICF, does not show full reactivation of the unmethylated genes. However, imprinted loci are also demethylated, and thus this disorder may involve imprinting disturbances as well. A second group of ICF patients do not show \textit{DNMT3B} mutations, and thus the hunt is on for additional causative genes.

A third chromatin-modifying gene that causes human disease is \textit{ATRX}, which encodes a chromatin-remodeling protein in the \textit{SNF2} family and contains a characteristic ATPase/helicase domain (4). \textit{ATRX} contains a cysteine-rich region similar to domains in the \textit{DNMT3B} family, and mutations in this X-linked gene cause mental retardation, urogenital abnormalities, and \textit{x}-thalassemia caused by down-regulation of the \textit{x}-globin genes. Douglas Higgs discovered the basis of this disorder and now reports both hypomethylation in the transcribed CpG rich regions of rDNA arrays and subtelomeric sequences, and hypermethylation in Y-specific repeats. In this case, the mutation is an embryonic lethal.

Mechanisms for Cancer and Birth Defects Involving Imprinted Genes

Bernard Horstemke, through painstaking cytogenetic and molecular analysis of patients with Prader-Willi syndrome and Angelman syndromes, has identified two nearby but distinct ICRs\(^2\) that appear to regulate normal imprinting and, when disrupted, lead to epigenetic inactivation of genes up to hundreds of kilobases distant (5). Although...
the ubiquitin E3A ligase gene has been identified as the Angelman syndrome gene (6), the Prader-Willi syndrome gene is still unknown because no definitive gene with germ-line mutations has been identified. There is more uncertainty than previously thought about the regulation of imprinting of the domain because several families with imprinting mutations affecting methylation over a large region show no microdeletions. However, Horsthemke and colleagues have identified a family of snoRNAs encoded within introns of UBE3A. These snoRNAs are unusual because they lack telltale RNA complementarity, but they are the only sequences within this complex of imprinted genes conserved between human and mouse and have now become his chief candidates for a function in this disease.

The first epigenetic disorder linked to cancer is Beckwith-Wiedemann syndrome, a disorder of overgrowth and birth defects with a 1000-fold increased frequency of childhood solid tumors such as Wilms’ tumor. Andrew Feinberg and Michael DeBaun found that LOI of IGFB2 is specific for cancer risk in this disorder, and LOI of the antisense transcript LIT1 is specific for birth defects (7). Feinberg and colleagues previously showed that LOI of IGFB2 is a common epigenetic alteration in cancer and that the mechanism in embryonal tumors is hypermethylation of an ICR upstream of H19, impeding access of the insulator protein CTCF to the maternal allele, thereby activating IGFB2 (8, 9). However, they now show that hypomethylation, rather than hypermethylation, is the mechanism of LOI in colorectal cancer and that the methylation changes directly affect IGFB2. These results suggest a represor recruitment model for IGFB2 silencing in adult tissues (see also the article in this issue; Ref. 10). The results should also not be entirely surprising because hypomethylation was the first epigenetic alteration observed in cancers.

Continuing on that theme, Oshimura and his coworker, Makiko Meguro, described LOI of IGFB2 and PEG1/MEST in lung cancer as well as epigenetic down-regulation of the imprinted gene, PEG3, in glioma cells. Their group has taken a genome-wide approach to identifying antisense and sense transcripts expressed either from the maternal or paternal allele, the assay performed by first isolating specific parental chromosomes in monochromosome hybrids. They have also taken a somatic genetic approach using the DT40 avian shuttle system, which has a high frequency of homologous recombination, and then transferring chromosomes into a mammalian background to do rapid knockouts. Targeted deletion of the LIT1 CpG island abolished LIT1 expression of the paternal chromosome, accompanied by activation of the normally silent paternal alleles of multiple imprinted loci at the centromeric domain, the first demonstration of a role of LIT1 as an imprinting control center (11). This technology is valuable not only for studying imprinted gene but for genetic studies in general. Rolf Ohlsson has made the provocative observation that nucleosome phasing may be responsible for silencing of H19 and IGF2 because it appears to regulate accessibility of CTCF to its binding sites.

An exciting new development in the epigenetics of cancer is the discovery of a paralogue of CTCF, which has been previously shown to be a highly conserved, multifunctional, multivalent nuclear factor with the properties of a tumor suppressor at 16q22 (12). CTCF is ubiquitously expressed in all somatic cells and involved in promoter activation and repression, hormone-inducible gene silencing, and creation of constitutive or methylation-sensitive chromatin boundaries (12). Both CTCF and its new mammalian paralogue, somewhat whimsically termed BORIS (13), were discovered by Victor Lobanenkov of the National Institute of Allergy and Infectious Disease. CTCF and BORIS share an identical 11 Zn-finger domain to interact with the same spectrum of CTCF/BORIS-binding sites but diverge at the NH2- and COOH-termini (13).

BORIS is a candidate gene for modulating epigenetic reprogram-ming and development, as Rolf Ohlsson (University of Uppsala, Uppsala, Sweden) has found that it is expressed in a restricted cell subset within the testis and precedes CTCF expression in development. Ohlsson showed that BORIS can replace CTCF on the H19 ICR, potentially marking these sequences for methylation acquisition in the male germ line. This is a promising result, as Lobanenkov showed that BORIS is located within the amplicon on 20q13 commonly involved in breast cancer and, in fact, is abnormally activated in those tumors. He suggested that CTCF and BORIS could serve as counteracting genes by manifesting and reprogramming epigenetic states, respectively. The number of CTCF/BORIS target sites occupied in vivo may be limited, as Ohlsson described an average one such site per 200 kb on chromosome 22.

ARHI is another imprinted tumor suppressor gene, described by Yinhua Yu at M. D. Anderson Cancer Center (14), who now reports that its expression is modulated both by DNA methylation and histone acetylation. Histone acetylation was also the focus of studies by Wataru Yasui of Hiroshima University in Japan, who is identifying genes differentially acetylated in tumors and has found that trichostatin activates p21 and phosphorylated Rb protein. A genome-wide approach has also been taken by Toshikazu Ushijima of the National Cancer Research Institute (Tokyo, Japan) who has been screening for hypermethylated sequences using methylation-sensitive representa-tional difference analysis. We hope that he will also perform the reciprocal experiment looking for loss of methylation and gene activation, given the studies described earlier from other groups.

Genomics Approaches to Epigenetic Mechanisms

A number of researchers are attempting to apply modern genomic tools and information to epigenetic studies. One such approach is a major joint university/corporate/government effort by the Riken Genomic Science Center, led by Yoshihide Hayashizaki, to obtain a complete functional annotation of mouse cDNAs, which is available on line.4 Readers would also be interested to know that Minoru Ko at the National Institute of Aging has made publicly available 31,000 mouse EST cDNAs on line.5

Bruce Howard has been performing chromatin immunoprecipitation using affinity-purified polyclonal antiacetyl lysine antibodies by differential display to identify differentially acetylated sequences. A total of 100,000 to 200,000 sequences has been identified to date, and they plan on a higher throughput approach. The Feinberg lab has taken a related approach to identifying normally methylated CpG islands, which may account for ~1% of CpG islands in the genome. About half of these are differentially methylated on maternal and paternal chromosomes and mark the location of novel imprinted genes. The other half appear to be methylated in most tissues and may be involved in tissue-restricted gene expression (15). In this case, loss of methylation of these normally methylated islands may lead to aberrant gene activation of, for example, members of the cancer-testis gene family such as MAGE and BORIS.

Lessons for Disease Mechanisms from Model Organisms

Several advances in model systems suggest novel mechanism for human disease. For example, Daphne Preuss and colleagues at University of Chicago have identified 47 single copy genes within centromeres that are highly methylated and yet expressed, suggesting novel mechanisms for their regulation (16). Victor Corces of Johns Hopkins University, whose group had identified the insulator protein Gypsy and the interacting protein Mod4 in Drosophila (17), showed

4 Internet address: genome.gsc.riken.go.jp.
5 Internet address: lgsun.grc.nia.nih.gov/cDNA/.

6785
EPIGENETIC MECHANISMS IN HUMAN DISEASE

provocative confocal photomicrographs of Gypsy in the polytene chromosome. This implies that the molecules interact with each other forming loops within the interphase chromosome. This is an idea that can trace its origins to the description of the nuclear matrix by Donald Coffey (18), a former AACR president. In fact, Corces’s loops are sensitive to RNase treatment unlike the original matrix preps of Coffey, which raises the provocative idea that an integral part of chromatin remodeling may be RNA itself. Steven Henikoff described a fascinating model that might explain the lack of an enzymatic machinery for removing lysine methylation of histones, an important source of chromatin programming and gene regulation. Replication-independent deposition of a variant histone H3, termed H3.3 that differs from H3 by only four amino acids, might displace modified histone and lead to gene reactivation.

The role of epigenetics in reprogramming the donor nucleus was discussed by Kevin Sinclair of the Scottish Agricultural College, where earlier abnormalities of the IGF2R gene had been found in cloned sheep (19). They have now found that several ICRs are inappropriately methylated in the newly formed zygote. In addition, Yong-Mahn Han of the Korean Research Institute of Bioscience and Biotechnology described aberrant methylation of many loci in cloned embryos compared with normal embryos produced either in vitro or in vivo.

Biochemical Insights

Several biochemical advances also provided insights into epigenetic mechanisms of human disease. Masahiro Sirakawa performed a crystal structure analysis of MECP2. The crystal structure indicates how the protein may access nucleosomal DNA without encountering steric interference from core histones. Interestingly, some residues of MECP2 that are mutated in Rett syndrome are located at the protein-DNA interface. Timothy Bestor (Columbia University) described the interesting alternative transcript of DNMT1 that is ovary specific and is required for maintenance of imprints in preimplantation development (20). It raises the important idea that alternative splicing or alternative promoter usage may be important in modulating epigenetic chromatin structure. Another DNA helicase, LSH, was recently found by Catherine Muegge (National Cancer Institute) to be required for normal DNA methylation (21). If one takes these data together, it suggests that chromatin-remodeling complexes contain methyltransferases, helicases, transcription factors, DNA binding proteins, and RNA, and that they interact with sequences both outside of genes as well as nucleosomes within genes.

Key Questions and Their Relationship to Cancer Research

We suggest that three key questions facing epigeneticists are: (a) how is epigenetic memory established and maintained? More precisely, what are the critical determinants in chromatin remodeling that establish and maintain an epigenetic state? Hints to the answer to this question come from this meeting, e.g., BORIS, alternative transcripts of methyltransferases, a role for Lsh and other chromatin-remodeling factors, special histones and histone-modifying enzymes, and perhaps a regulatory role for RNA. Experiments like those of Victor Corces suggest that subnuclear compartmentalization is a key ingredient of chromatin states. Most studies of chromatin to date have been focused on the nucleosome, the promoter, and immediately apposite modifiers. However, we suggest that a full understanding will require elucidation of the role and function of these elements within specific topological domains, for example, on the nuclear periphery. This question is critical to understanding cancer biology because cancer is as much a disorder of disturbed epigenetic memory as it is of mutational errors. The additional hope for cancer epigenetics is that it may lend itself to creative therapeutic or chemoprevention because by definition epigenetic changes are more malleable than conventional mutations.

(b) What is the role of epigenetic modification in speciation? This is an oft-heard but promising area. Just as one can clearly establish hierarchical trees of evolution for gene variation among species, a similar organization must exist for targets of epigenetic modification. The high degree of conservation of GC-rich nonexonic sequence suggests that is so. In a provocative discussion, Carmen Sapienza (Temple University) suggested at the meeting that epigenetic modifications could underlie meiotic drive. We suggest that variations in cis-acting regulatory sequences may have profound effects on the pattern of regulation, not only of genes but also of groups of genes within domains such as imprinted gene clusters. A hint that this might be so comes from a recent study showing that although the exonic sequences vary little among primates, the pattern of gene expression in the brain varies greatly (22). It will be interesting to make comparisons among such expressed genes. Perhaps some of the same genes are involved that show altered expression in Rett syndrome. The implication for cancer biology is that we should examine these abundant, but until now little understood, regulatory elements for alterations in cancer. Perhaps mutations in such sequences might provide a genetic basis for epigenetic alterations in cancer, either in the tumors themselves or as predisposing factors in families.

(c) How does epigenetic disruption cause human disease? This was the rationale for the conference. We already have part of the answer to this question, particularly for some monogenic disorders and cancer, as described above. In addition, several common complex diseases such as bipolar affective disorder appear to involve, in part, genes that show parent of origin-specific transmission, as was described at the meeting by Raymond DePaolo (Johns Hopkins University). In cancer, also, epigenetic alterations may precede tumor development (23). If so, our current tools for analysis of the genetic basis of complex traits may not be adequate. For example, linkage analysis or transmission disequilibrium tests may require consideration of parent of origin. Also, one may need to examine epigenetic associations such as methylation of individual genes in addition to conventional haplotype association. Our current genomic tools are not well-equipped to address such issues. As mentioned in the introduction to this report, methylation is not systematically archived in the genome databases. In addition, despite the intense interest in identifying and cataloguing human haplotypes for the benefit of disease researchers, there is almost no consideration at present of epigenetic modifications. It was therefore gratifying and encouraging that David Lipman (Director, National Center for Biotechnology Information) showed intense interest in epigenetics at the conference. We hope that investigators in our field will contribute to the open source and extremely rich and dynamic repository of information that the National Center for Biotechnology Information has pioneered.

Acknowledgments

We thank Richard Lee for helpful comments.

References


Epigenetic Mechanisms in Human Disease

Andrew P. Feinberg, Mitsuo Oshimura and J. Carl Barrett


Updated version Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/62/22/6784

Cited articles This article cites 21 articles, 7 of which you can access for free at: http://cancerres.aacrjournals.org/content/62/22/6784.full#ref-list-1

Citing articles This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/62/22/6784.full#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.