A Blueprint for a Human Epigenome Project: The AACR Human Epigenome Workshop

Peter A. Jones¹ and Robert Martienssen²

¹University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California and ²Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Abstract

Epigenetic processes control the packaging and function of the human genome and contribute to normal and pathologic states, including cancer. The time is ripe to undertake an international effort to identify all the chemical changes and relationships between chromatin constituents that provide function to the genetic code. A timely workshop of leading experts, convened by the American Association for Cancer Research (AACR), confirmed that the technology is at hand to begin defining human epigenomes at high resolution. (Cancer Res 2005; 65(24): 11241-6)

Introduction

The Human Genome Project, which defined the exact sequence of almost all of the 10⁹ bases in DNA, is one of the greatest scientific achievements of our time. The availability of the complete sequence has provided scientific opportunities and opened new fields of genomics research that simply did not exist a short while ago. As powerful as it is, however, the sequence itself does not predict how the genome is packaged in chromosomes and chromatin to provide for the differential expression of genes, which is essential for organismal development and differentiation. Epigenetic processes that initiate and maintain heritable patterns of gene expression without changing the sequence of the genome provide this key function.

Recent advances in epigenetic research have provided new high-throughput technologies for the analyses of DNA cytosine methylation patterns and histone modification marks that give structure and function to the human genome. A Human Epigenome Project (HEP) would "genomicize" the epigenetic research that is currently being conducted on a piecemeal basis and would pave the way for unforeseen breakthroughs in understanding normal and disease states. A HEP would offer much information of direct relevance to public health because of the acknowledged role of epigenetics in normal and pathologic processes, such as aging, mental health, and cancer, among many others. The epigenome interacts with the environment and changes with age. Several mental health syndromes, such as Rett syndrome, involve proteins that interpret cytosine methylation signals. Cancer is increasingly being recognized as having an epigenetic component that may be contributing equally, if not more so, to mutations in driving tumor initiation and progression. These are compelling reasons for applying existing technologies to deciphering our epigenome in its entirety.

There is, therefore, ample justification to begin this project; however, several issues need to be resolved before a coordinated effort is mounted. Among these issues are questions, such as the following: How many epigenomes should be sequenced and at what level of resolution? Which of the bewildering array of histone modifications should be examined? How will the project be coordinated? Is the technology ready? To begin to answer these questions, the AACR organized a Human Epigenome Workshop in June 2005 that was focused on planning a HEP. This workshop followed a 2004 National Cancer Institute-sponsored workshop, "Epigenetic Mechanisms in Cancer," chaired by Andrew P. Feinberg and Terumi Kohwi-Shigematsu, which had preliminary discussions of these issues.

The AACR-sponsored workshop developed key strategies to answer these questions and concluded that the time is ripe to begin a HEP. It developed a mission statement to serve as a guiding principle: "The goal of the Human Epigenome Project is to identify all the chemical changes and relationships among chromatin constituents that provide function to the DNA code, which will allow a fuller understanding of normal development, aging, abnormal gene control in cancer, and other diseases as well as the role of the environment in human health."

The following is a summary of the findings and the recommendations of the workshop.

State of the Science

The AACR Human Epigenome Workshop was organized into five scientific sessions, each addressing the current status of a major theme in epigenetic research.

Genomic Microarrays and Their Analysis, chaired by Bradley E. Bernstein, discussed several successful applications of genomic-scale technologies for analyzing epigenetic data and emphasized that the technologies are maturing rapidly.

- Bradley E. Bernstein gave an overview of high-throughput technologies for mapping chromatin modifications focusing on available chromatin immunoprecipitation (ChIP)/chip (microarray) and ChIP/serial analysis of gene expression (SAGE) platforms. These approaches can identify novel and functional elements in the human genome, including a T-cell specific enhancer that would not have been detected by DNA sequence analysis alone.

- Thomas R. Gingeras described how the genome can be interrogated in an unbiased manner, using tiling array studies, and showed that particularly for transcription analyses, an unbiased approach can uncover regulatory sequences that would otherwise be missed.

- The need for analytic approaches to handle epigenetic data was addressed by Rebecca W. Doerge, who discussed methods of accounting for technical variation in DNA methylation experiments.
DNA Methylation Analysis was chaired by Stephan Beck.

- Stephan Beck and Florian Eckhardt presented updates on the European Union– and Wellcome Trust–funded HEP, which aims to “identify, catalog, and interpret the DNA methylation profiles of all human genes in major tissues.”
- One of the more surprising themes to emerge concerning DNA methylation was that, in contrast to conventionally accepted dogma, CpG islands and gene promoters are frequently methylated in normal tissue.
- Kunio Shiota presented a genome-wide analysis of tissue-specific differentially methylated regions in the mouse, which should provide a valuable basis for comparison in a HEP.
- Hiroyuki Sasaki showed how DNA methyltransferases contribute to genomic imprinting.
- Manel Esteller showed surprising epigenetic variability in identical twins.
- Andrew P. Feinberg described rapid recent advances in epigenomic technology, including array-based methylation analysis, ChIP/chip, and SAGE-linked methylation and chromatin analysis. He also emphasized the importance of epigenetic analysis of normal tissue and the idea that epigenetic changes may precede genetic changes in cancer. In addition, the work emphasizes the importance of understanding the epigenetic basis of stem cell alterations for increasing cancer risk.

Chromatin Modification, chaired by Thomas Jenuwein, described a joint effort of German and Austrian groups that has comprehensively mapped histone modifications on mouse chromosome 10 and the plans for a project to map mouse chromosomes 6, 8, 18, 19, and X.

- Several speakers presented data showing the essential role of histone modification in imprinting. Robert Feil showed specifically that methylation of histones in an imprinting control region is responsible for regulation of a cluster of imprinted genes on chromosome 7.
- Histone methylation and acetylation have been most frequently studied to date, but studies presented by Shelley L. Berger showed that additional marks, including phosphorylation, ubiquitylation, and sumolation also influence gene expression in Saccharomyces cerevisiae.

Model Organisms/CpG Islands, chaired by Philip Avner, pointed out the use of model organisms to a HEP. Model organisms provide a means of getting from correlation to causality by intervening experimentally in ways not possible in humans.

- Anne C. Ferguson-Smith described how mammalian-imprinted domains provide a powerful paradigm for functional investigations of the epigenetic control of normal and abnormal gene expression. In particular, she emphasized the role of small and large noncoding RNAs in epigenetic control, and that epigenetically repressed states can occur in the absence of DNA methylation at promoters. She also presented a method of manually preparing tiling arrays, allowing 94% coverage through a 150-kb region consisting of 45% repeats.
- William A. Held presented data indicating there are many tissue-specific differentially methylated regions (TDM) in the genome. Identification of 150 TDMs indicates that almost 60% are in CpG island gene promoter regions, and that they are likely to have an important role in tissue-specific gene expression.

Disease States was chaired by Stephen B. Baylin; he pointed out that although epigenetics plays a prominent role in cancer, it undoubtedly is important in many other diseases as well. He also emphasized the importance of examining cancer as a disease not only of altered methylation, but also of altered chromatin structure. Aberrations in methylation lead to abnormalities in tightening or loosening of chromatin that directly affect gene expression.

- Analysis of the colon cancer cell line HCT 116, presented by Stephen B. Baylin, showed that these cells contain four genetic mutations in β-catenin, transforming growth factor β receptor, MLH1, and p16. Fourteen genes, however, were silenced by hypermethylation, including the wild-type allele of p16, an important tumor suppressor.
- Methylation status was also shown to have predictive value. Data from Peter W. Laird showed that in 148 human breast carcinomas, methylation of the estrogen receptor gene correlated with increased survival in response to tamoxifen and was a better predictor of response than hormone receptor status itself. These results were obtained using the high-throughput method, MethylLight.
- Toshikazu Ushijima presented a study of DNA methylation in 140 neuroblastoma cases, which showed that poor prognosis correlated almost exclusively with high methylation of a particular CpG island. He also presented data showing that Helicobacter pylori, believed to be a causative agent of gastric cancer, induces aberrant methylation.
- Joseph F. Costello presented a large-scale integrated genomic and epigenomic analysis of 25 brain tumors that showed the putative tumor suppressor gene Wnk2 was silenced by DNA methylation much more frequently than it was lost to deletion.

![Loci affected by epigenetic mechanisms are often independent of those affected by genetic mechanisms](image-url)

**Figure 1.** Comparative contributions of genetic and epigenetic mechanisms to genetic alterations in analysis of 25 brain tumors by array comparative genomic hybridization and restriction landmark genomic scanning. The proportion of loci affected by methylation (light blue, partial methylation; dark blue, homozygous methylation) and copy number change (gray, deletion; yellow, gain; orange, gain but methylated) and convergence of deletion and methylation (red). Modified and reproduced with permission of Nature Genetics (1).
the loci tested for genetic (copy number) change and epigenetic (aberrant methylation) change, 67% were not affected by either type of mechanism. Of those that showed alteration (Fig. 1), 66% were due to methylation, 18% were due to deletion, and 7% showed deletion of one allele and methylation of the other (1).

Loss of imprinting (LOI) at the *IGF2* locus is correlated with a 5-fold increase in incidence for colon cancer. Andrew P. Feinberg presented data showing that in a mouse LOI model, tumors are more numerous and larger, and increased numbers of undifferentiated cells are found in the intestine, suggesting that LOI results in a prolonged undifferentiated state, providing a larger pool for transformation.

In addition to DNA methylation, histones and the proteins that modify them carry an enormous amount of epigenetic information. Manel Esteller presented data showing that histone methylation and acetylation marks correlate with progression in 90% of the patients with colon cancer whom his group studied.

In summary, it is becoming evident that cancer cannot be comprehensively understood or addressed in the clinic unless the scope of the epigenetic effect on cancer is revealed. It is worth noting that in addition to its role in contributing to a more complete understanding of cancer, a HEP would also have relevance to many other scientific areas, including stem cell therapies, developmental biology, neurologic disorders, and a host of other diseases.

**Developments in Epigenomic Technology**

Technologies designed to examine epigenetic phenomena at the genomic level have shown remarkable progress over the last few years. Not only is the technology capable of sustaining genomic efforts, but several groups are already engaged in genome-scale projects that are delivering new insights into epigenetic mechanisms. In addition, there is a concordance of results in that different labs using different approaches are obtaining similar results, which is encouraging evidence of the validity of these new techniques. Furthermore, these results are leading to shifting paradigms concerning the epigenetic code, indicating that genome-scale efforts are not only possible, but are also necessary if the field of epigenomics is to advance.

**Chromatin immunoprecipitation/chip methodology.** Probably the most broadly used technology to examine histone modifications is the ChIP/chip methodology, in which the products of ChIP are applied to microarrays. Histones tails are subject to a number of chemical modifications believed to influence higher order chromatin structure and function. In ChIP/chip experiments, cells are typically fixed, cross-linking DNA and histones, and then sonicated to fragment the chromatin into pieces several nucleosomes in length. The resulting cell lysate is treated with antibodies targeting particular modifications, and the chromatin is collected and amplified, most often using PCR-based methods or in vitro transcription. This material is hybridized to genomic microarrays, available from several commercial sources, and the results are analyzed with regard to any number of questions to determine where histone modifications line up in the genome or how they differ when cell types are compared.

Alternative systems employing constructs in which biotin tags were attached to histones, which were pulled down with streptavidin and analyzed on arrays, have also yielded spectacular insights into epigenomic organization. This was described by Steven Henikoff and Christian Beisel.

**High-throughput DNA methylation analysis.** DNA methylation is currently being analyzed using genomic microarrays and bisulfite sequencing. In a pilot phase of the European HEP, published in November 2004 (2), the effort successfully analyzed 253 loci covering 90 genes within the human MHC by bisulfite sequencing, identifying methylation variable positions in six tissue types. Data were verified by mass spectrometry, and agreement was close to 90%. Phase I of the project is currently underway, and >5,000 loci on four chromosomes (6, 13, 20, and 22), covering ~10% of human genes, are expected to be completed by Fall 2006. Moreover, the effort identified methylation markers with clinical significance for colon and breast cancer.

Jeffrey A. Jeddeloh, of Orion Genomics, LLC, described an alternative to bisulfite sequencing based on genome partitioning technology. The company had initially developed this technology to sequence plant genomes but was now applying it to the human genome. These technologies, such as MethylScope, sort a randomly fragmented genome into methylated and unmethylated populations employing the restriction enzyme *Mc*<sub>RC</sub>B. Sorted material is hybridized to genomic DNA tiling arrays developed by NimbleGen Systems, Inc. Blood DNA from four age- and race-matched individuals is being analyzed, with one twelfth of the first genome currently completed. Calculation suggests that the final analysis, expected soon, will deliver methylation occupancy data for >80% of the entire genome, as well as the majority of the unique/single-copy portion. These data will be made publicly available when completed.

The technologies that have been developed to date are not without challenges, but none seem intractable. Probably the most significant difficulty concerns the specificity and consistency of antibodies available for ChIP. Commercially available antibodies are polyclonal, and several groups reported variations in products from the same company over time. Some groups have seen success with monoclonals, whereas others have experienced difficulties. Frequent occurrence of coincident markers (e.g., acetylation and trimethylation of Lys4 on active chromatin) provides a measure of internal control and ameliorates the problem somewhat. The difficulty of antibody specificity should be resolvable, but the scale of the issue renders it unlikely to be tackled by one group alone. Standardized reagents would constitute a likely benefit of a HEP that would help to push the field forward.

Also mentioned were the computational challenges presented by the amounts of data being collected. As was pointed out, the Human Genome Project gathered data that were more or less linear and could be analyzed using digital methods. A HEP effort, however, would require the analysis of data that are orthogonal in nature, owing to the fact that there is not just one epigenome: epigenetic states vary between tissues, among individuals, and in healthy versus disease states. This variation is in fact the essence of the epigenetic code, which makes it so vital to collect and analyze epigenetic data. Several computational groups are already engaged in developing tools to handle these data sets. And as was the case with the Human Genome Project, new tools will no doubt continue to evolve as the technology and biology progress. Most likely, the establishment of a formal HEP would drive the further development of technology.

Bioinformatic strategies for the epigenome was the subject of a previous workshop in December 2004, organized by Denise Barlow and Robert Martienssen at the Banbury Center, Cold Spring Harbor, NY. Many of the issues raised then, including the scope and rigor of analysis required for data collection and interpretation, were further explored at that workshop.

Perhaps the most encouraging aspect about the new technologies is that they are already informing biological questions and leading
the field to question some of its paradigms. Among these is the
dogma that CpG islands and gene promoters are not methylated
in normal tissue. John M. Greally showed that a surprisingly high
proportion of CpG islands are in fact methylated, as many as 40%,
depending on the tissues studied. Several ChIP/chip studies showed
that some modifications to histones are actually occurring slightly
downstream of transcription start sites, not directly over the sites
as previously believed; these results have implications for under-
standing the role of histones in transcription initiation. These
shifting paradigms are the direct result of studies taking a global
view of epigenetic marks in different cell types and show that a
comprehensive effort is now needed to advance our understanding
of the epigenetic code and its function in normal and disease states.

The Way Forward: Potential Structure of a Human
Epigenome Project

It is important to emphasize that there is not one "epigenome," but
rather many different "epigenomes" that define normal and disease
states. Deciphering these epigenomes will tell us not only how the
genome is packaged, but, more importantly, comparative assess-
ment will also indicate how it is functionally organized into states
that allow for differential output in specific tissues. Furthermore,
knowledge of the baseline in differentiated tissues is also essential
to understanding genomic function. In view of the fact that several
disease states, particularly cancer, have been directly linked to
epigeneric changes, it is now timely to mount such an international
undertaking.

Reference epigenomes. There are in fact many epigenomes,
necessitating the need to define a subset of epigenomes to be
determined at a high level of resolution. The premise is that a
limited number of normal cell types can be chosen and char-
acterized at every level: everything that can now be measured (all
methylation sites on the DNA and all chromatin marks) would be
cataloged.

Which cell types to choose and whether cell lines or primary
tissues should be used would have to be agreed upon. The amounts
of material necessary for certain experiments (ChIP/chip typically
requires on the order of 5 mg of chromatin) might necessitate the
use of cell lines. Cell lines also offer the benefits of ready availability
and some degree of standardization; however, there is some
question as to their relevance to real biological situations.
Particularly, with regard to epigenetic states, a relatively high level
of variability might be introduced in different handling situations.

One potential source of primary tissue is peripheral blood cells,
which could be collected and sorted to obtain fairly homogeneous
populations. An additional advantage is that stem cell and
developmental biology has been studied fairly extensively in the
hematopoietic system, and lymphocytes are banked in many
locations. Therefore, data emanating from a HEP could be readily
extended upon by investigators outside the field, which would
leverage results.

An alternative source of primary tissue might be fibroblasts
cultured from foreskins, which are readily available and, being at an
early developmental stage, could be expected to display less
individual epigenetic variation than tissues from older individuals.
Cultured human fibroblasts are relatively stable in culture, are
relatively homogeneous, and never undergo spontaneous transfor-
mation. A limitation would be gender restriction; placental tissue
could offer similar benefits and limitations.

Second-tier resolution. In addition to reference epigenomes, a
second set of samples would be analyzed at lower resolution. Single
nucleotide resolution is not necessary to obtain meaningful
information about comparative epigenetic states. Techniques, such
as ChIP/chip and MethylScope, can point out and direct
investigation to regions of the genome to which "drill-down"
techniques can then be applied. Bisulfite sequencing remains the
gold standard for single nucleotide resolution, but resolution at
that level may not be necessary or even appropriate to obtain
meaningful information about regions of chromatin. Roland Green
suggested that a multitiered approach could begin immediately on
a variety of tissues, such as several cancers, young and aging brain
tissue, and samples from a metabolic disease, among others. This
approach would result in a small number of epigenomes being
analyzed at single nucleotide resolution; then a progressively larger
number of epigenomes would be examined using methods of
decreasing resolution (Fig. 2).

Such a HEP would begin to answer some of the most
fundamental biological questions that epigenetics can address:
What are the marks that distinguish a proliferating cell from a
derivating cell? Which marks might account for aging and for
disease?

Wider functional/mechanistic implications. In addition to
settling the issue of tissues and cell types to examine, the regions
of the genome to be surveyed must be determined. One possibility
is to include all known promoters and start sites that have been
associated with differential methylation. Many additional start
sites exist both within genes and in intergenic regions, however,
and the functions of these are currently unknown. They need to be
determined if a complete picture of gene expression is to be
assembled. Investigators working on issues of imprinting are not
necessarily concerned with promoter regions. Eric D. Green
described how the National Human Genome Research Institute
has recently chosen regions for analysis in a "cross-section" of the
genome for the ENCODE project (Encyclopedia of DNA Elements).
ENCODE is annotating 30 Mb constituting 1% of the human
genome across 44 regions. Fourteen of these, ranging from 0.5 to
2 Mb, were manually selected based on the existence of current
data sets; the remainder, in 0.5 Mb fragments, were computationally
selected to represent areas of high, medium, and low gene
content and human-mouse sequence conservation.

![Figure 2. Proposed multitiered organization for a HEP.](http://example.com/figure2.png)
In addition to cataloging modifications, it will also be necessary to engage in mechanistic studies. The presence of active marks on chromatin correlates fairly well with transcription, but so-called repressive marks do not necessarily correlate, at least at the resolution of the studies undertaken thus far. Information is not yet readily available for issues, such as the importance of timing, the sequential addition of marks to histones and DNA, and how these marks might interact in synergetic fashions in "chromatin landscapes." Speaking metaphorically, Hiroyuki Sasaki said, "We need to understand which genes are on the desktop, which are on the bookshelf, and which are in a locked filing cabinet." Also critical is defining the mechanism of memory: how marks are retained through mitotic and meiotic cell division.

Many of the questions that a comprehensive study of the Human Epigenome would explore are relevant but not limited to concerns of cancer research. They include issues of high interest to the public at large, such as the use of stem cell therapies. Particularly regarding the safety of such therapies, epigenomic profiling before cell replacement could confirm that the differentiated state is the correct one and could also be used to detect any potentially malignant epigenetic changes. One could make the argument that it will not be safe to implement therapies resulting from therapeutic cloning until these issues are addressed. In describing the importance of epigenomics to the general public, the intersection of the field with philosophical issues of nature versus nurture should be emphasized. Studies of identical twins have shown that 3-year-old twins share very similar profiles of global gene expression, but that by age 50, their profiles differ as widely from one another as they do from unrelated individuals in the general population (Manel Esteller).

Therefore, epigenetic marks can be described as marks of experience; they contribute to differentiating us from our younger selves.

**Action Plan to Advance the Human Epigenome Project**

It will be essential to inform and educate the public and decision makers about the overriding importance of this field for the prevention, detection, diagnosis, prognosis, and treatment of cancer and other diseases. It is clear that an international HEP is needed, and encouraging collaborations and partnerships with scientists around the world who have a shared interest in solving the problems of cancer and other intractable diseases, to join in a coordinated effort in support of a HEP. Our growing knowledge of epigenetic mechanisms suggests that this complex network of molecular signals that regulate gene activity has vast implications for understanding the initiation, growth, and spread of cancer. Furthermore, the technologies needed to examine epigenetic changes at the genomic level have advanced sufficiently to facilitate the success of this project. Such an effort should be launched as soon as it is feasible so that thus, we may begin to unravel the role played by epigenomic forces in cancer and other diseases.

**AACR Workshop on the Human Epigenome Speaker and Participant List**

**Peter A. Jones and Robert Martienssen, Co-Chairpersons**

- Grace S. Ault
  - National Cancer Institute
  - Bethesda, MD
- Joseph F. Costello
  - University of California
  - San Francisco, CA
- Philip Avner
  - Institut Pasteur
  - Paris, France
- Jennifer Couch
  - National Cancer Institute
  - Rockville, MD
- Stephen B. Baylin
  - Johns Hopkins University
  - Baltimore, MD
- Rebecca W. Doerge
  - Purdue University West Lafayette, IN
- Stephan Beck
  - Wellcome Trust Sanger Institute
  - Cambridge, England
- Florian Eckhardt
  - Epigenomics AG
  - Berlin, Germany
- Christian Beisel
  - University of Heidelberg
  - Heidelberg, Germany
- Manel Esteller
  - Centro Nacional de Investigaciones Oncológicas
  - Madrid, Spain
- Shelley Berger
  - Wistar Institute
  - Philadelphia, PA
- Robert Feil
  - Institute of Molecular Genetics
  - Montpellier, France
- Bradley E. Bernstein
  - Harvard University
  - Cambridge, MA
- Andrew P. Feinberg
  - Johns Hopkins University
  - School of Medicine
  - Baltimore, MD
- Joseph F. Costello
  - University of California
  - San Francisco, CA
- Philip Avner
  - Institut Pasteur
  - Paris, France
- Rebecca W. Doerge
  - Purdue University West Lafayette, IN
- Stephan Beck
  - Wellcome Trust Sanger Institute
  - Cambridge, England
- Christian Beisel
  - University of Heidelberg
  - Heidelberg, Germany
- Shelley Berger
  - Wistar Institute
  - Philadelphia, PA
- Bradley E. Bernstein
  - Harvard University
  - Cambridge, MA
- Andrew P. Feinberg
  - Johns Hopkins University
  - School of Medicine
  - Baltimore, MD
Acknowledgments

Sincere thanks are due to Suzanne Clancey for her editorial assistance with this report.

References

A Blueprint for a Human Epigenome Project: The AACR Human Epigenome Workshop

Peter A. Jones and Robert Martienssen


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/65/24/11241

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.