A Blueprint for an International Cancer Epigenome Consortium.
A Report from the AACR Cancer Epigenome Task Force

Stephan Beck1, Bradley E. Bernstein2, Robert M. Campbell4, Joseph F. Costello5, Dashyant Dhanak9, Joseph R. Ecker6, John M. Greally11, Jean-Pierre Issa10, Peter W. Laird7, Kornelia Polyak3, Benjamin Tycko12, and Peter A. Jones8, for the AACR Cancer Epigenome Task Force

Recent whole-genome sequencing of thousands of human cancers has uncovered an unexpectedly large number of mutations in genes, which control the epigenome. Many of these recurrent mutations are in epigenetic writers (e.g., EZH2, DNMT3a), readers (BRD4-NUT fusions), erasers (UTX mutations), and chromatin remodeling enzymes (BAF subunits of SWI/SNF). They occur at high frequencies and may therefore be drivers of carcinogenesis, which is supported by functional data in some cases. The data have reemphasized the role of epigenetics in human cancer and pointed to a much closer collaboration between genetic and epigenetic events in carcinogenesis (1). The realization of the commonality of epigenetic pathways in neoplasia has opened the door to the development of drugs that will target these defects and likely be therapeutic for specific types of cancers. There is therefore a need to map both the genetic and epigenetic landscapes in cancers to provide a backdrop for the rapid development and deployment of new epigenetically targeted therapies.

It is now possible to map human epigenomes in great detail, and several international and national efforts are underway to achieve this task in normal tissues. Mapping of cancer epigenomes is a more formidable task, but one that is clearly attainable with current technologies and is likely to become increasingly achievable over the next few years. The AACR Cancer Epigenome Task Force recommends the establishment of a public–private partnership, the International Cancer Epigenome Project (ICEP), to map a defined number of cancer epigenomes that may be relevant for the development of future drugs. Mapping of cancer epigenomes will also show the inter- and intracase variability between and within tumors and have important implications for our understanding of the epidemiology, detection, prevention, and treatment of cancer.

Cell Purity/Isolation

The human epigenome, in contrast to the genome, is highly cell type–specific in normal tissues because cellular differentiation states are largely regulated by the interaction of transcription factors with the epigenome. Thus, purified cells relatively homogeneous for a particular differentiation state are necessary for a precise definition of the normal epigenome. The human epigenome is modified by multiple additional factors including physiologic life events, such as aging, parity, menopause, and environmental factors, such as nutrition (including methyl donors such as folic acid) and exposure to infectious and toxic agents. The genome and the epigenome interact, so the epigenome has to be evaluated in terms of specific genetic contexts. For example, DNA methylation and histone modification patterns are known to be influenced by germline genetic polymorphism including single-nucleotide polymorphisms (SNP) and DNA copy number variants. Thus, purified cells from individuals with previous or concurrent genotypic data and detailed epidemiologic information are the best choice to use for the definition of human epigenomes. Cell purification poses a challenge and is best conducted from fresh tissue samples, but normal fresh tissues such as hematopoietic cells, breast, and gastrointestinal tract are fairly easily available in sufficient quantities from individuals without cancer.

The mapping of the cancer epigenome is also challenging, as most tumors display remarkable intratumor genetic and epigenetic diversity. Detailed knowledge of the intratumor diversity is particularly important for the interpretation of epigenetic changes as it can serve as surrogate for longitudinal sampling, which is an effective but slow and expensive way to distinguish driver from passenger epimutations. Tumor-specific somatic genetic changes may also directly affect epigenetic profiles (e.g., by affecting the enzymes or chromatin modifiers/histones themselves) or indirectly (e.g., by changing intracellular levels of substrates and/or cofactors that modulate switching the cells to a more stem cell-like state). The identification of analogous cell types in normal and neoplastic tissue is important as cancer cells usually have more plastic and heterogeneous phenotypes than normal cells.
The need for fresh tissue samples, matched normal tissue, and well-characterized cell differentiation hierarchies currently limits the tumor types for which complete epigenetic studies can be conducted. Hematopoietic malignancies, breast, gastrointestinal (especially colon), brain, and lung cancers are some of the most likely candidates. On the other hand, pediatric tumors are potentially the most likely to be driven by aberrant epigenetic programs as shown by the paucity of somatic mutations in some of these tumor types (e.g., retinoblastoma) and mutations in chromatin-related genes (e.g., histone H3 in pediatric glioblastoma multiforme). However, technologies are continuously improving, allowing for the use of fewer and fewer cells as starting material. Methods for analysis of archival frozen tissues (e.g., isolated nuclei) and even formalin-fixed samples (for DNA methylation) are also becoming available. Funding technology development and optimization now, rather than conducting studies on poorly fractionated cells might avoid the development of uninterpretable misleading results that can hamper progress.

Setting the goals and scope of the proposed ICEP therefore requires detailed attention to the tissue requirements unique to epigenome assays and proactive approaches to reduce sample-related failures. DNA methylation analysis is feasible on many unfixed and archival tumor samples but hydroxymethylation mapping has not yet been deployed within The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) or International Cancer Genome Consortium (ICGC) projects. Other essential epigenome assays such as chromatin immunoprecipitation sequencing (ChIP-Seq) are reproducible on primary cultures or fresh frozen tissue but not on archival tissue stored in fixative. To best accommodate the tissue quality, quantity, and cellularity requirements, the cancer epigenome project will take a prospective approach to sample collection. This approach will raise the success rate for each sample and assay while simultaneously enabling more uniform collection of clinical information. We also expect to incorporate a limited set of retrospectively collected samples for practical purposes and for pilot profiling, concurrent with prospective tissue collection. An inclusive approach will be taken to best represent diversity of the patient population.

Intratumoral heterogeneity is a major complicating factor for therapy, particularly with single agents. Therefore, addressing regional and cellular heterogeneity in the tumor epigenome will be a priority in the cancer epigenome project. This stands apart from other cancer projects that prioritize sample size but ignore intratumoral heterogeneity, which has recently been shown to be highly informative. This design feature will allow novel insights into epigenomic patterns in the tumor cells themselves, in non-neoplastic cells of the supportive tumor microenvironment, and between phenotypically distinct malignant cell types within the same tumor. There are practical considerations when enriching for select cell types that will allow cellular heterogeneity to be assessed in some but not all cancer types.

Elucidating the evolution of the tumor epigenome over time is an important goal of the ICEP that also impacts sample selection. Distinguishing the epimutations that drive tumorigenesis and progression from the much larger number of alterations that are functionally inconsequential is a major challenge. While the characteristics of true driver epimutations are still being defined, we expect drivers to show evidence of natural and/or therapy-induced selection in vitro and to exhibit quantifiable effects on the target gene transcript levels and downstream gene networks. These characteristics can be assessed in the epigenome and transcriptome trajectories obtained from profiling multiple tumor samples collected over time from individual patients.

The ICEP aims to impact clinical oncology in part by advancing the application of promising epigenetic therapeutics. The project will therefore incorporate tumor types for which matching short-term cultures can be obtained. This will allow direct comparisons of epigenome patterns and in vitro drug response on an individual patient-specific basis. For example, mapping the distribution of a given mark in the tumor will allow initial predictions of the transcriptional consequence of inhibition of the enzyme that adds or removes the marks, which can then be tested directly on the matching culture. Ultimately, this parallel approach could couple epigenetic or transcription biomarkers to drug response for these novel agents.

In summary, to define true malignancy-associated epigenetic alterations, we need to profile paired analogous cell types from normal and cancer tissues purified to homogeneity as much as possible. Although this is a challenging task, several recent studies from multiple laboratories have shown that it is feasible at least for some organ/cancer types. In contrast, epigenetic biomarkers could be identified even from bulk tissue samples, and even these can have some immediate use, but their ability to inform therapies aimed at cancer-specific epigenetic alterations is more limited. Most importantly, epigenetic profiles are genetic context dependent, thus, the genotype of the same samples (both normal and tumor) would also have to be characterized.

Methodology/Miniaturization/Single-Molecule Analyses

The cost of whole-genome sequencing continues to rapidly decrease so that single-base resolution DNA methylation maps of genomes created using bisulfite sequencing have also become less expensive. Histone modification analysis has made great strides with the introduction of ChIP-Seq. While many new insights have been gained using the technology, it does not yet offer high resolution. Also the reproducibility of peak calls in duplicate samples is approximately 90%, a precision that somewhat compromises comparisons of different samples. Moreover, new marks on DNA (e.g., 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine) and histones (e.g., crotonylation) have recently been identified, so our knowledge of the epigenome parts list is still incomplete. To understand the roles of these novel epigenomic marks, cost-effective technologies need to be developed for accurately quantifying and localizing these modifications in cancer cell genomes.

In evaluating technology for a cancer epigenome project, one should first consider current knowledge, limitations, and
the potential uses of the information. It is clear that cancers have very different epigenomes than adjacent normal tissues. What is less clear is how much variation there is when comparing normal epigenomes between different individuals or how much intratumoral heterogeneity there is in epigenetic patterns. Whole-genome studies also suggest that differences between cancer and normal are often quantitative rather than all or none. For these reasons, quantitative single-base resolution analysis using the smallest number of cells possible is suggested to be the preferred technology for a comprehensive approach to the problem.

For DNA methylation analysis of cancer genomes, we can envision 2 different approaches to wisely manage resources—whole-genome studies at significant sequencing depth on a limited number of samples and targeted sequencing using custom DNA capture technology in a large number of cases. Currently, bisulfite treatment coupled with deep sequencing is the preferred approach. If accurate technologies to detect 5-methylcytosine (and other DNA modifications) at a single-molecule level were available, they would be strong candidates to supplement (or replace) bisulfite-based approaches. For histone modifications, limitations of current technologies preclude large-scale analyses in a substantial number of tumors. Technology development is needed to refine or to develop alternatives to enrichment-based approaches. Ideally, future technologies would allow the capture of base resolution information for all DNA modifications, protein DNA interactions including transition factor locations and nucleosome positioning, and higher-order genome structural information (e.g., chromatin conformation capture). Development of methods that allow simultaneous capture of multiple types of epigenomic information (e.g., NOMe-Seq and ChIP-BS-Seq/ BisChIP-Seq) in a dynamic fashion from long (>10–100 kb) DNA molecules in single cancer cells should be the long-term objective of the project.

In summary, the epigenome is dynamic, unlike the primary sequence of DNA and remains incompletely understood. Advances in genomic technology are likely to continue over the next few years, and a strategic ICEP would adopt a flexible approach that allows rapid incorporation of new technologies and new understanding from these analyses.

**Integrating Genomics and Epigenomics**

A full understanding of cancer epigenomics requires the incorporation of associated genomic information. The interplay between cancer genomics and epigenomics covers many facets and can flow in both directions.

At the most basic level, the probability that genomic loci undergo cancer-related somatic epigenetic change is affected by local sequence context, including G:G content and CpG density, proximity to repetitive elements, nuclear architecture, and the presence of DNA-binding protein recognition sequences. In addition to these fixed germline sequence influences, the many somatic genetic changes that arise in cancer are likely to affect the epigenome. These include copy number alterations such as amplifications and deletions, as well as copy number neutral LOH, translocations, and mutations. Not only are these structural genomic changes likely to influence local epigenetic marks, but just as importantly, a correct molecular interpretation of the epigenomic data requires accurate underlying copy number, structural, and sequence information.

There has also been an impressive surge in the discovery of somatic mutations of epigenetic writers, erasers, remodelers, and insulators in cancer, highlighting the functional importance of epigenetics to oncogenesis. A comprehensive epigenomic characterization of cancers with and without such mutations can help us understand how these mutations contribute to cellular transformation, but a full picture will require further experimental investigations. Most epigenetic therapies to date have targeted wild-type chromatin regulators. These recently identified mutations offer exciting opportunities for novel therapeutic strategies, including drugs targeted to mutant forms although relatively few of these are gain of function. However, EZH2 mutations are clearly a gain of function, whereas DNMT3A mutations, though recurrent, do not seem to be a gain of function.

Another genetic–epigenetic link is the strong association observed between some epigenetic subtypes of cancer and the occurrence of specific somatic mutations. Examples include the association between CpG island methylator phenotype (CIMP) in colorectal cancer and mutation of the BRAF oncogene, and of G-CIMP in glioblastoma with mutation of the IDH1 gene. This is also an area that will benefit greatly from additional mechanistic studies.

As mentioned earlier, the discrimination between epigenetic driver and passenger events is one of the main challenges for the functional interpretation of the results of a cancer epigenome project. Integration of epigenomic data with genomic and expression data and extending this integrated approach to a full functional pathway analysis can help to identify the most functionally relevant epigenomic changes.

Ideally, a pilot cancer epigenome project would be staggered with first a discovery phase, followed by a clinical phase. The discovery phase should take advantage of existing genomic data contained within TCGA or ICGC, ideally by analyzing some of the same samples. If this proves to be logistically difficult, then single-cell studies or available xenografts propagated from these samples may serve as surrogates. Alternatively, prospective collections of sufficient amounts of tumor material could be considered. The clinical phase would use the integrated genomic and epigenomic information to propose novel therapeutic strategies.

**Computational and Analytic Approaches**

A major challenge in an international cancer epigenome project is that of managing, analyzing, and making available the data and information generated to the community. Significant problems include the sizes of the datasets involved making it difficult even to move data from one location to another. It is also likely that sample collection and performance of assays will take place in multiple locations, imposing the need to collect and organize metadata from the outset to help downstream analyses to account for any variability at this early stage.
We believe that the central goal should be to create an interface for data users that allows intuitive exploration of the epigenomics data in ways that are useful to basic science researchers and clinicians, even if they lack expertise in computation. This is a very significant challenge, but it is critical if the data are to have maximal impact and value to the community. As outlined later, this will require substantial planning efforts and infrastructure, including hardware, software, computational scientists, and active community participation.

One key to the development of the computational infrastructure for this kind of project will be that it should exist in a reasonably functional form before the generation of the data that it will have to accommodate. If the samples for analysis take some time to acquire, characterize, and process, this offers a window of opportunity to develop the core system and database for the project. It will be a challenge to design a system in advance that anticipates all of the eventual needs of the project amidst rapidly evolving technologies and analysis tools, and to get this development from design to execution in a compressed timeframe.

To help address this challenge, there are some existing paradigms from which this project could learn. First, TCGA has been working on the characterization and analysis of different cancers for some time, and has developed database and analytic processes that could be adopted. Second, the Roadmap Epigenomics Consortium has developed a Human Epigenome Browser (http://epigenomegateway.wustl.edu/) that allows integrated viewing of multidimensional epigenomic data similar to those that will be generated by ICEP. Beyond biology, the same sort of challenges have been addressed by the astrophysics community in establishing the Virtual Observatory, which involved large datasets of different types that needed to be preprocessed, integrated, and made accessible through web interfaces (http://www.us-vo.org/). This is an example of an international initiative that had to deal with complexities comparable with those faced by an international cancer epigenome project and could serve as a valuable paradigm.

The analysis of data can be divided into 2 major areas. One is the conversion of raw data such as DNA sequences into biologic information, such as DNA methylation levels or sequence variant calls. This could be referred to as primary data analysis. The second is interpretation and effect prediction, which requires integration with further sources of information, the secondary or integrative step. These sources of information include genomic position and relationship to genomic annotations or a comparison of this information with that of other samples with different characteristics using the same molecular assay (e.g., differentially methylated regions distinguishing cancer subtypes) or using a different assay (e.g., comparing DNA methylation and histone modification patterns). A key to the success of the Virtual Observatory was that the workflows to process data were defined by consensus, allowing subsequent exploration of the data to be based on processed and not primary data. A significant benefit is that this reduced the sizes of datasets required. Similar consensus and transparency about how data are processed, at least at the primary stage, will be critical to the feasibility of providing a useful cancer epigenomics resource to the community.

This need to impose standards should be balanced by facilitating the exploration of the data by the community, in particular allowing analytic algorithms developed by community users to be used to explore the data. This raises the question of how hardware resources will be found to support the project. Our favored model is of dedicated hardware to support the project, with a central data warehouse complemented by geographically distributed hardware resources, all using common formats and database structures and containing all associated metadata mirrored across multiple sites for security and optimized for fast access. Hardware used for production (data preprocessing and automated analysis) should be separated from that used for data exploration by user algorithms. Cloud computing solutions such as Biomim- bus could also be part of the solution, and we recognize that this is a fast-moving field that will expand in terms of potential solutions over time.

**Epigenetic-Targeted Drug Discovery and Development**

Numerous reports suggest that aberrant epigenetic modification and subsequent gene transcription are critical in a variety of disease states including cancer, neurodegeneration, inflammation, and autoimmunity. Such modifications are controlled by epigenetic "writers" (e.g., histone transferases, acetyltransferases, ubiquitin ligases, etc.), "readers" (e.g., bromo, chromo, TUDOR, and PHD domain proteins) or "erasers" (demethylases, deacetylases, etc.). In a cancer context, drugs that modulate the activities of these epigenetic modifiers could lead to either silencing of oncogenes and/or reexpression of tumor suppressor genes. In addition, the profile of these epigenetic target genes could be used as a prognostic for patient selection and/or pharmacodynamic marker of response to epigenetically targeted therapeutics.

Despite a recent surge of research activity in academia and industry, the biochemical and biologic activities of most epigenetic proteins have yet to be fully characterized. Similarly, the relevance of specific epigenetic "marks" to phenotypic changes in normal and/or diseased cells has yet to be understood. Nonetheless, there have been some exciting breakthroughs that convey optimism about epigenetic therapies. DNA methyltransferase (DNMT) inhibitors, such as Vidaza and Dacogen, have been approved for use in myelodysplastic syndrome, and broad-spectrum histone deacetylase (HDAC) inhibitors (Zolinza, Istodax, etc.) are used clinically for the treatment of cutaneous T-cell lymphoma. It was recently reported in a small phase I/II clinical trial that patients with non-small cell lung carcinoma responded favorably to a combination of an HDAC inhibitor (Entinostat) and Vidaza (2). In addition, a 4-gene signature was found to predict sensitivity to the drug combination.

Successful drug discovery relies on targeting key proteins and/or pathways important in driving the disease phenotype. Identification of such targets requires an understanding of the underlying biology and increasingly, restricting trials to
patients most likely to respond to therapy. In contemporary cancer drug discovery programs, targets are selected on the basis of human genetics (mutations, translocations, amplifications, etc.), clinical outcome data, supporting experimental biology, unmet medical need, and tractability for small-molecule intervention. Experience with kinase inhibitors suggests activating mutations/translocations may also represent the "low hanging fruit" of epigenetic targets. However, only a few such mutations have been reported thus far (3) and many occur with very low frequency. Several inactivating mutations of epigenetic proteins are known, but it is often difficult to drug such targets (i.e., with an activator). Many chromatin modifiers are known to exist in heterogeneous complexes whose composition can change and alter the biologic response to this complex. Pharmacologic intervention to modulate the epigenetic response may be accomplished by drugs that affect these protein–protein interactions and/or by inhibition of any inherent enzymatic activity.

There are many unanswered questions in discovering and developing epigenetic modifiers for cancer therapy. These include:

1. Whether existing chemical libraries provide tractable starting points.
2. The availability of structural information and if such crystal structures are relevant in the absence of a multicomponent complex.
3. Whether biochemical or binding assays reflect the activity of the physiologic complex.
4. The role of histone versus nonhistone substrates.
5. The acute and/or long-term safety profiles of novel, selective epigenetic modifiers.

In addition, target validation and subsequent assay development for drug screening is complicated by a paucity of high-quality biologic and chemical tools: active proteins, antibodies (total protein, Ac/Me-specific), mutant constructs (e.g., catalytically inactive), crystal structures, and reference compounds. To address the daunting task of building a comprehensive "epigenetics toolbox," public–private partnerships have emerged (e.g., the Structural Genomics Consortium), whereby all partners share precompetitive information to accelerate early drug discovery. It is proposed that similar private–public partnerships may also be required to better understand the biology of these diverse epigenetic proteins and develop appropriate animal models, as they are likely to differ from conventional models. In summary, epigenetically targeted drug discovery is still in its infancy, but the promise of epigenetic modifiers as therapeutics is encouraging.

Concluding Remarks

The advances discussed here provide clear evidence that any perceived gap between desired versus attainable goal of ICEP is closing rapidly and the time has come to move our attention toward implementation of ICEP. Mapping of cancer epigenomes with the aim of developing an integrated understanding of the cross-talk between genetic and epigenetic changes will facilitate patient benefit through the selection of appropriate targets for drug discovery and development. Importantly, ICEP will provide clinically relevant biomarkers (e.g., gene signatures) and relevant animal models that better predict clinical response in a stratified patient population. From the perspective of compound screening, there are still many biologic and chemical tools required to enable epigenetic assays. It is clear that close collaboration between academia and industry will help facilitate the development of these tools to speed up the delivery of epigenetic therapies to patients.

Steering Committee

Peter A. Jones, Chairperson
University of Southern California, Los Angeles, CA
Stephen B. Baylin
Johns Hopkins University, Baltimore, MD
Bradley E. Bernstein
Massachusetts General Hospital, Boston, MA
Susan Clark
Garvan Institute of Medical Research, The University of New South Wales, Sydney, Australia
John M. Greally
Albert Einstein College of Medicine, Bronx, NY
Peter W. Laird
University of Southern California, Los Angeles, CA
Dirk Schubeler
Friedrich-Miescher Institute for Biomedical Research, Basel, Switzerland
Toshikazu Ushijima
National Cancer Research Institute, Tokyo, Japan

A Blueprint for an International Cancer Epigenome Consortium
The Cancer Epigenomics Task Force

Bruno Amati
European Inst. of Oncology, Milan, Italy
Stephan Beck
University College London Cancer Institute, London, UK
Robert M. Campbell
Eli Lilly and Company, Indianapolis, IN
John D. Carpenter
TGen, Phoenix, AZ
Arul M. Chinnaiyan
University of Michigan, Ann Arbor, MI
Joseph F. Costello
University of California, San Francisco, CA
Dashyant Dhanak
GlaxoSmithKline, Collegeville, PA
Joseph R. Ecker
Salk Institute for Biological Sciences, La Jolla, CA
Manel Esteller
Institut d'Investigació Biomèdica de Bellvitge, Barcelona, Spain
Kristian Helin
University of Copenhagen, Copenhagen, Denmark
James G. Herman
Johns Hopkins University, Baltimore, MD
Jean-Pierre Issa
Temple University, Philadelphia, PA
Laurie Jackson-Grusby
Children’s Hospital Boston, Boston, MA
Young-Joon Kim
Yonsei University, Seoul, South Korea

Jonathan D. Licht
Lurie Comprehensive Cancer Center, Northwestern University, Chicago, IL
Xiaole Shirley Liu
Dana-Farber Cancer Institute, Boston, MA
Huck Hui Ng
National University of Singapore, Singapore
Kornelia Polyak
Dana-Farber Cancer Institute, Boston, MA
Daniel Reinberg
NYU Langone Medical Center, New York, NY
Victoria M. Richon
Epizyme, Inc., Cambridge, MA
Henk Stunnenberg
Radboud University, Nijmegen, The Netherlands
Benjamin Tycko
Columbia University, New York, NY
Jun Wang
Beijing Genomics Institute, Shenzhen, Guandong, China

Disclosure of Potential Conflicts of Interest

S. Beck has ownership interest (including patents) in Alacris Theranostics and is a consultant/advisory board member of GATC Biotech. R.M. Campbell has commercial research grant from Eli Lilly and Company. K. Polyak has commercial research grant from Novartis and is a consultant/advisory board member of Novartis and Millennium. P.A. Jones is a consultant/advisory Board member of Eli Lilly and Company. No potential conflicts of interest were disclosed by the other authors.

Received September 18, 2012; accepted October 8, 2012; published OnlineFirst November 27, 2012.

References

A Blueprint for an International Cancer Epigenome Consortium. A Report from the AACR Cancer Epigenome Task Force


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-3658

Cited articles
This article cites 3 articles, 1 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/24/6319.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/72/24/6319.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.