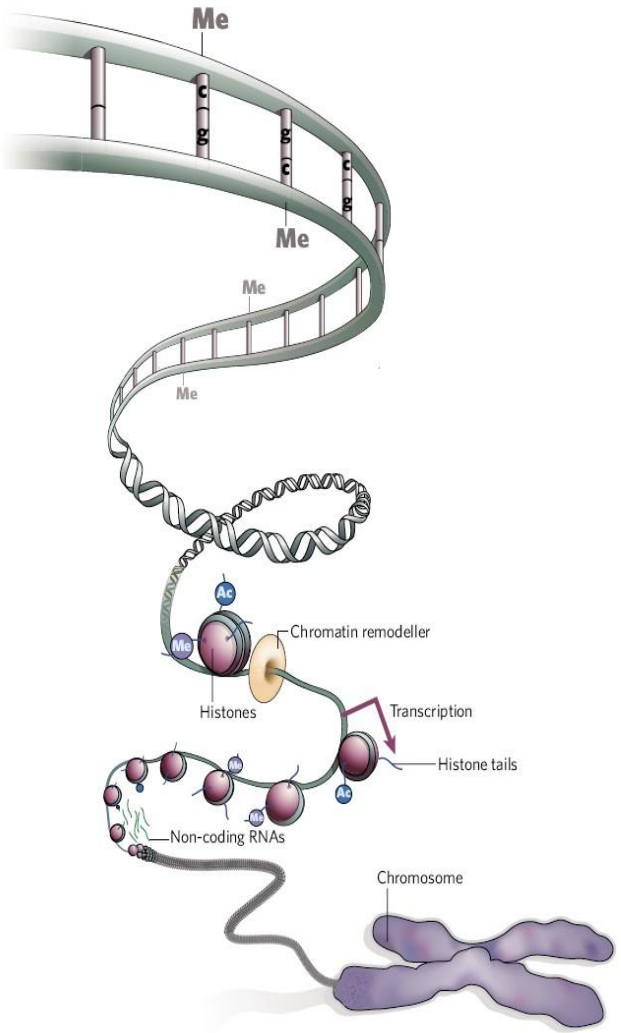
GOALS, STRUCTURE, POLICIES & GUIDELINES

January 10, 2013



Note: This document supersedes all previous versions of the IHEC Policy document. Moving forward, the IHEC Policy Document will be revised as needed by the IHEC Executive Committee.

INTERNATIONAL HUMAN EPIGENOME CONSORTIUM

GOALS, STRUCTURE, POLICIES & GUIDELINES

TABLE OF CONTENTS

Section

A. Introduction
B. Consortium Goals
C. Background to the Consortium
D. Structure of the Consortium
E. Consortium Policies and Guidelines

INTERNATIONAL HUMAN EPIGENOME CONSORTIUM

A. Introduction

The deciphering of the human genome sequence has helped our understanding of biological processes in health and diseases. However, the way in which the genomic information is organized within the cell, through epigenetic processes, is known to play a major role in regulating gene expression and in controlling specific cellular functions. Epigenetic processes go beyond DNA-stored information and are essential for packaging and interpreting the genome, are fundamental to normal development and cell differentiation, and are increasingly recognized as being involved in human disease.

Epigenetic mechanisms include histone modification, positioning of histone variants, nucleosome remodelling, DNA methylation, and small and non-coding RNAs, among others. In concert with transcription factors and other DNA-binding proteins, these epigenetic mechanisms, which may be inherited from cell to cell, regulate gene expression patterns to govern the development of the > 250 cell types in the human body. While the DNA sequence is identical in almost all of these diverse cell types, their epigenetic profiles are very distinct. The modulation of these epigenetic profiles significantly contributes to embryonic development, differentiation, and cell identity, transitions from a stem cell to a lineage-committed cell, and underlies responses to environmental signals (e.g., hormones, nutrients, stress, and damage). In many respects, the epigenetic interpretation of the genome (i.e. epigenomic information) represents a "second code" that programs and stabilises the DNA-based information in diverse biological contexts.

Mis-steps in epigenomic programming have been directly implicated in common human diseases including but not limited to diabetes, cardiopulmonary diseases, neuropsychiatric disorders, imprinting disorders, inflammation, autoimmune diseases, and cancer as well as in ageing. Importantly, epigenomic changes are potentially reversible by drug treatments. This has significant implications for the prevention and treatment of these major human diseases. Indeed, several inhibitors of chromatin-modifying enzymes, including histone deacetylase (HDAC) inhibitors and DNA methyltransferase (DNMT) inhibitors have now been FDA and EU approved and are being used in clinical practice with good prognosis for tumor regression. Therefore, epigenetic-based therapy is now a reality in the clinic. However, to maximize the potential of such therapeutic approaches, it is critically important that there be a more comprehensive characterization of the epigenetic changes that occur during normal development, adult cell renewal, and disease, and of the relationships between genetic and epigenetic variation and their impact on health.

Regenerative medicine is a very promising approach for many diseases. Recently, major progress has been achieved in cellular reprogramming to generate pluripotent cells from human somatic cells. These new sources of pluripotent cells are potentially useful for the production of genetically compatible material for cellular therapy. Reprogramming involves

changes in epigenetic profiles and it will be important to have reference epigenome maps of all relevant human cell types to evaluate the importance and the consequences of these epigenetic changes.

Environment and nutrition have strong and durable influences on our health. Differences in epigenetic profiles are known to be induced by environmental and nutrition changes, so that maps for reference epigenomes will greatly broaden our understanding of how the environment and nutrition will modulate epigenetic alterations. This new, non DNA-based, knowledge will have a major impact for novel avenues in preventing and diagnosing disease.

Recent technological improvements allow high throughput mapping of epigenome in a very reproducible and standardized way. It is now possible, with these new technologies, to map the entire epigenome of a human cell. Scientists and representatives of major funding agencies have decided to launch the International Human Epigenome Consortium (IHEC). Just as the Human Genome Project provided a reference 'normal' sequence for studying human disease, IHEC will provide high-resolution reference epigenome maps to the research community. These maps will integrate the various epigenetic layers of detailed DNA methylation, histone modification, nucleosome occupancy and corresponding coding and non-coding RNA expression in different normal and disease cell types. The epigenome reference maps will be of great utility in basic and applied research, have an immediate impact on understanding many diseases, and will hopefully lead to the discovery of new means to control them. Although the project should have a human focus, it will be essential to involve model organisms to obtain mechanistic insights as to the functionality of epigenomic parameters or "codes."

Studies in model organisms such as yeast, fly and mouse have yielded fundamental discoveries across many fields of biology, including notable advances in our understanding of epigenetic mechanisms of gene control. Moreover, disease models established in the mouse have furthered understanding of mechanisms of cancer, aging and other disorders. Such models have led to the identification of promising drug targets and enabled initial evaluation of candidate therapeutic agents. Epigenomic maps for model organisms can thus provide essential information to further these studies and to benchmark the underlying organismal, cellular and disease models against human counterparts. We therefore recommend that up to 10% of IHEC funding be made available for epigenomic mapping studies in model organisms.

IHEC will coordinate epigenome mapping and characterisation worldwide to avoid redundant research effort, to implement high data quality standards, to coordinate data storage, management and analysis and to provide free access to the epigenomes produced. The expectations are that the outcome of the research carried out by the members of IHEC will be extensive. First and foremost will be the availability of reference human epigenomes to the world-wide research community. Second, will be valuable information on the methods utilized by IHEC members to produce, analyze, and integrate large epigenomic datasets related to health and diseases, in human and in model organisms. Third, it will become possible to compare different human populations thereby evaluating the impact of environment and nutrition on the epigenome. IHEC will facilitate communication among the members and provide a forum for coordination, with the objective of maximizing efficiency among the scientists working to understand, treat, and prevent diseases.

B) CONSORTIUM GOALS

Primary Goals

1. Coordinate the production of reference maps of human epigenomes for key cellular states relevant to health and diseases. To have a substantial coverage of the human epigenome, the IHEC sets the ambitious goal to decipher at least 1000 epigenomes within the next 7-10 years.

   To reach this goal, the consortium will use robust and validated technologies to generate:

   - very high resolution maps of informative histone modifications
   - high resolution DNA methylation maps
   - landmark maps for transcription start sites of all protein coding genes
   - entire catalogue of and expression patterns of non-coding and small RNAs
   - comparative analysis of epigenome maps of model organisms relevant to human health and diseases

2. IHEC will focus on key cellular states such as stemness, immortality, proliferation, differentiation, senescence, and stress, thereby generating new knowledge that will catalyse progress in health research and regenerative medicine.

   Surveys of individuals, pedigrees and genetically identical twins will be used to determine the relationship between genetic and epigenetic variation worldwide. A long term IHEC goal is to determine the extent to which the epigenome has shaped human populations over generations and in response to the environment.

   IHEC would differ from and complement other ongoing projects such as ENCODE (ENCyclopedia Of DNA Elements). The ENCODE project is focused on defining the functional DNA sequences in the genome, whereas IHEC would define the patterns of epigenetic regulation occurring at those sequences in different primary cells.

3. Coordinate rapid distribution of the data to the entire research community with minimal restrictions, to accelerate translation of this new knowledge into health and diseases. IHEC will coordinate the development of common bioinformatics standards, data models and analytical tools to organize, integrate and display whole epigenomic data generated from this important international effort.

4. IHEC will set up the efficient structure that will coordinate this international effort so that the interest and priorities of individual participants, self-organising consortia, funding agencies and nations are addressed. IHEC will encourage the minimal amount of redundancy between the different epigenetics efforts around the world. IHEC will also interact and coordinate its efforts with other international projects, such as the International Cancer Genomic Consortium (ICGC) and ENCODE.

Secondary Goals

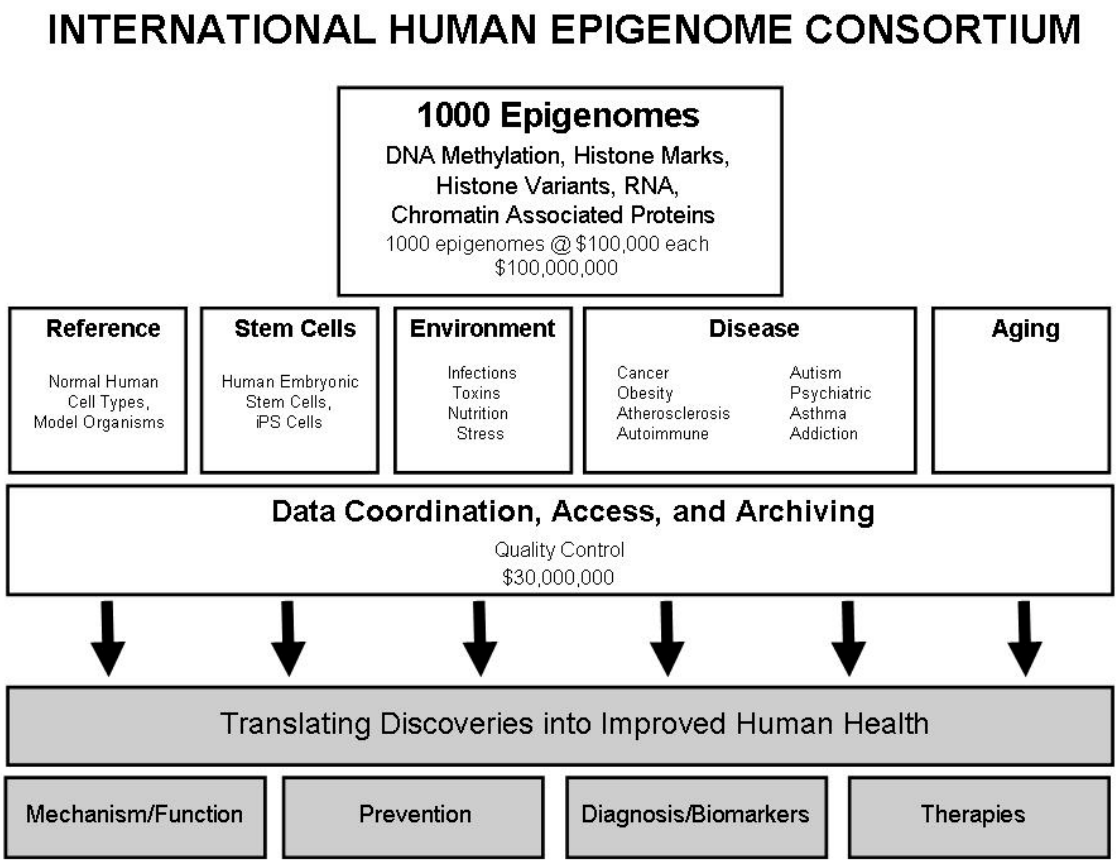
5. Catalyse the development of new and robust technologies that will facilitate the characterisation and functional analysis of the epigenome in health and diseases thereby driving substantially down the costs of epigenome mapping.

6. Support the dissemination of knowledge and standards related to new technologies, software, and methods to facilitate data integration and sharing between epigenetic researchers around the globe.

Timeline

The IHEC consortium has the ambitious goal of producing 1000 epigenome maps (see Figure). Although high-throughput technologies are evolving very rapidly, a 7-10 year program needs to be secured for the generation and coordination of epigenomic reference maps for the major primary human cell types in health and disease.

A rate limiting step might be the availability, in sufficient amount, of specific human cells or tissues for carrying out full epigenome mapping analysis. Special attention will be given in the IHEC steering committee to address this issue and to facilitate sample collection and exchange. IHEC might also interact with other international efforts such as the International Cancer Genomic Consortium that could provide access to large collections of tumour samples for epigenome mapping analysis.



C. Background to the Consortium:

Individual investigators have been studying epigenetics for several decades; however, concerted efforts to organize the epigenomics research community are quite recent, and none has sought to engage the community on a broad-based, international level.

Europe has a strong tradition for epigenetics research. Epigenetics research programs have been funded by different European countries and by the EU Framework Programme (FP6/FP7). More than €50 million (US$79 million) has been allocated to networks and consortia that focus on central epigenetic questions such as DNA methylation (HEP, Human Epigenome Project), chromatin profiling (HEROIC, High-Throughput Epigenetic Regulatory Organization In Chromatin), and treatment of neoplastic disease (EPITRON, EPIgenetic Treatment Of Neoplastic Disease). Special attention was given to structuring the epigenetic research landscape in Europe via the successful "The Epigenome" Network of Excellence (www.epigenome-noe.net).

In the past few years, there have been several efforts to organize the epigenetics research community in the **United States** and develop the support and structure for a Human Epigenome Project. Several workshops sponsored by the National Cancer Institute (NCI) (2004, Epigenetic Mechanisms in Cancer Think Tank, http://www.cancer. gov/think-tanks-cancer-biology/page7; by National Institute of Environmental Health Sciences (NIEHS), 2005 (Environmental Epigenomics, Imprinting and Disease Susceptibility, http://www.landesbioscience.com/journals/epigenetics/article/2642/) and by the American Association for Cancer Research (AACR) (June 2005, (http://www.aacr.org/page9673.aspx) pointed towards the need of initiating a Human Epigenome Project that could take full advantage of advances in several existing US and European initiatives. On the heels of these workshops, the AACR Human Epigenome Task Force, a cross-disciplinary group of international investigators, was formed to design a strategy and develop a timetable for the implementation of an International Human Epigenome Project.

The US National Institute of Health selected Epigenomics as a NIH Roadmap Program taking into account growing awareness that the epigenome may have widespread and profound implications for human health and disease. The overarching goal of the NIH Roadmap Epigenomics Program is to target scientific gaps that must be overcome in order to translate the promise of epigenetic science into applications that maximally affect human health and a wide range of common complex human diseases. The NIH Roadmap Epigenomics Program consists of six distinct components. 1. The Reference Epigenome Mapping Centers (REMCs) will map reference epigenomes of a variety of normal human cells and tissues. The reference epigenomes will be mapped with respect to DNA methylation, histone modifications, and complement of non-coding RNAs. 2. An Epigenomics Data Analysis and Coordination Center (EDACC) provides an informatics and analysis resource to assist components of the program by coordinating and facilitating common data format structure and integrative analyses of the epigenomic data. 3. The National Center for Biotechnology Information (NCBI) is generating a repository and long-term data archive for the Roadmap Epigenomics Program as well as a user friendly Epigenetics Public Interface. 4. The Technology Development in Epigenetics initiative supports the development of innovative technologies that have the potential to significantly change the way that epigenomics research can be performed in the future. 5. The Discovery of Novel Epigenetic Marks in Mammalian Cells initiative was developed in recognition that our basic understanding of epigenetic modifications may be incomplete with respect to the universe of epigenetic regulatory marks. 6. The Epigenomics of Human Health and Disease initiative supports research on fundamental epigenomic changes or mechanisms underlying specific diseases; conditions of development or aging; or response to exposures (physical, chemical, behavioral, and social factors).

Asia is also active in fostering epigenomics research, with a major emphasis placed on disease epigenomes, especially those in liver and gastric cancers. An international meeting, Genome-wide Epigenetics 2005, was held in Tokyo. Scientists from Yonsei University (South Korea), the National Cancer Center (Japan), the Shanghai Cancer Institute (China) and the Genome Institute (Singapore) also organized several meetings to facilitate the exchange of information in epigenomics (Seoul, 2006; Osaka 2007). In December 2006, a Japanese Society for Epigenetics was formed. Clearly, Asia is now poised to contribute strongly to global epigenomics research.

Interest is building in **Canada** for a broad-ranging "Epigenetics, Environment and Health" (EEH) initiative, led by the Canadian Institutes of Health Research (CIHR). Canada has existing strength in several areas of epigenetics research - including cancer, stem cells, early human development and neuroscience - and is looking to more broadly examine the interplay between environmental signals and the genome that underlie individual differences in health.

National support for the Human Epigenome Project is also mounting in **Australia** with the formation of the Australian Alliance for Epigenetics in 2008 (http://www.epialliance.org.au). Australian meetings devoted to epigenetics were initiated in 1996 in Heron Island with the first workshop on bisulphite sequencing and this has been followed by biannual meetings hosted by different States in Sydney, Canberra and Perth, with the most recent being held in Melbourne in December 2009 to showcase Australia's strengths in the global epigenetic research arena.

In March 2009, the NIH Roadmap Epigenomics Program Working Group, organized a workshop in Bethesda on: "Exploring International Epigenomic Coordination". The workshop brought together high level policy makers from major funding agencies worldwide and top scientists from all continents. The main purpose of this workshop was to explore the coordination of an international epigenomics project that would integrate the NIH Roadmap Epigenomics Program with other international efforts. The concept of an International Human Epigenome Consortium (IHEC) was strongly supported by all participants.

As a result of this meeting, an IHEC Interim Executive Committee was established with Phil Avner, Bradley Bernstein, Susan Clark, Amanda Fisher, Thomas Jenuwein, Peter Jones, En Li, Robert Martienssen, Jacques Remacle (observer), Bing Ren, John Satterlee (observer), and Kazu Ushijima. In 2009 the Interim Executive Committee prepared a draft policy document describing the proposed scope and policies for IHEC. The committee was assisted by working groups/committees dealing with the scientific planning, the data release policy, the data standard, the funding strategies, etc. The Interim Executive Committee also prepared the program of the IHEC launch conference which took place in Paris on Jan 25-26, 2010. Over 90 scientists and funding agency representatives attended this meeting to discuss and revise the proposed scope and policies of IHEC (Nature (2010) **463**:587). An Epigenome Network of Excellence web link was used to provide all Paris meeting invitees and the global scientific community the opportunity to comment on the proposed scope and policies of IHEC.

D. Structure of the Consortium

The IHEC is a confederation of members that share the common goals and principles described in this document and have agreed to work in a coordinated and collaborative manner within a consortium.

Members consist of Funding Members and Research Members, each of which is an individual or allied group that will provide a level of funding or scientific expertise to undertake a significant part of the research tasks foreseen in the International Human Epigenome Consortium. Each member will have the responsibility for financially or scientifically supporting a significant research program.

It is recognised that Funding Members interested in joining the consortium may not have funds in place to support projects aligned with the goals of IHEC and thus may be unable to immediately commit the requisite funds. In the absence of a qualifying research project, these funding agencies will be invited to join IHEC as a funding member with observer status until December 2011. This should allow sufficient time for observers to secure funds, to plan initiatives of large magnitude, and to make a firm commitment.

Funding agencies planning to join IHEC are encouraged to submit a letter indicating their intent (see sample letter of intent at the end of this document).

Categories of membership are defined as follow:

IHEC Funding Members:

1) Single funding agency or
2) Alliance of organisations, with a representative from a single organisation with the coalition appointed to the IHEC Executive Committee (EXEC)

Funding agencies are encouraged to become Funding Members as they become ready to contribute to the IHEC and adopt the Consortium's policies and guidelines.

To be considered as an IHEC Funding Member, the funding agency should provide substantial support of a minimum of $10 million US in total distributed over 5 years to a project/programme in line with the IHEC objectives. These contributions should not include overhead/indirect costs and equipment. IHEC Funding Members will be invited to nominate a representative to the IHEC Executive Committee (described below) and will have an active role in the governance of the IHEC initiative. Nominations must be received before June 30, 2010. Additional funding agencies are encouraged to become IHEC Funding Members in the future, as they become ready to contribute to the IHEC and adopt the Consortium's policies and guidelines.

IHEC Funding Associated Members:

Funding agencies committing less than $10 million US/5 years total contribution to this international effort are also very much welcome to join IHEC. These agencies would be considered "IHEC Funding Associated Members". These IHEC Funding Associated Members will collectively elect a maximum of two representatives to the IHEC Executive Committee. These additional seats will rotate between IHEC Funding Associated Members.

IHEC Research Members:

To join the IHEC as a Research Member, nominations must originate from an IHEC Funding Member that will provide support to the research organisation. Nominations are reviewed and approved by the Executive Committee. Research Members will have the

demonstrated capability and capacity to support major research effort within IHEC and will accept the set of commonly agreed-upon policies and guidelines described in this document. Such organizations will need to have existing or committed funds from an IHEC Funding Member.

Research Members can be:

a) A research center or network of national or international research groups organized to perform major research activities in line with the IHEC goals;

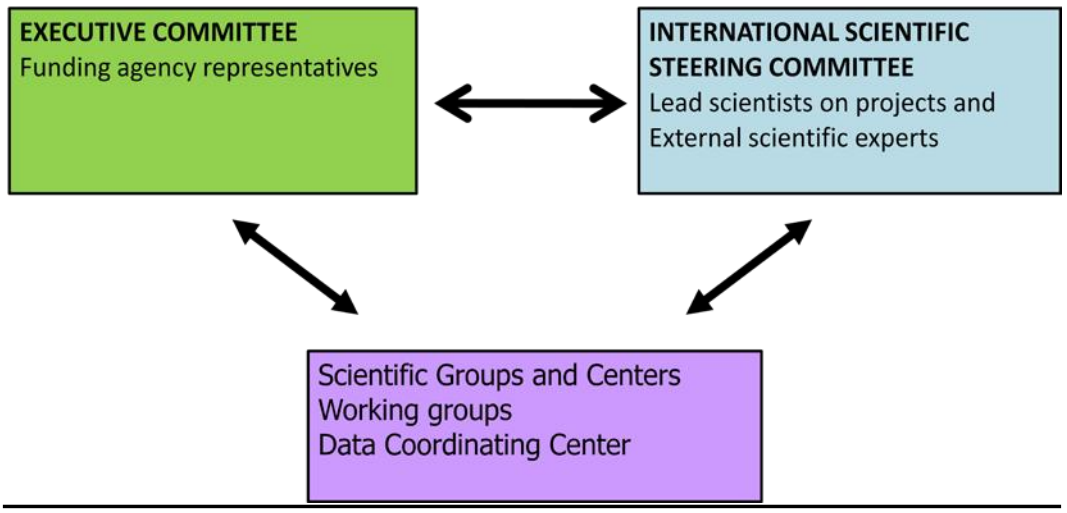
b) Other centers (Data management, ethics, etc..) which contribute significantly to IHEC

Given that these organizations will likely have different structures, and include many investigators, scientific managers and technical staff, each organization will be asked to nominate representatives to participate in IHEC coordination activities, such as the International Scientific Steering Committee, working groups, workshops, and IHEC meetings.

Note that IHEC itself will not have funds to support scientific research. Rather, it is the responsibility of each IHEC funding agency member or associated member to support research projects aligned with the goals of IHEC.

Structure:

IHEC will utilize a distributed organizational model. This model has been successfully used in other international genome projects, where high standards and policies have been determined at the outset, and acceptance and adherence were prerequisite for joining. This model relies on the interaction among funders (providing oversight), an international scientific steering committee (setting guidelines) and scientific groups and centers (Data Production Centers and regional- or national-level Data Analysis and Coordination Centers involved in data production, quality assessment and data management). The strength of the Consortium's structure relies not only with its component parts but also in the bilateral flow of information between the groups.



Governance

Missions for the different IHEC committees will be clearly defined and the decision-making process will be streamlined and agreed upon by all participants.

- **IHEC Executive Committee (EXEC)**

Oversight of the IHEC will be provided by an **IHEC Executive Committee (EXEC)**, constituted of individuals nominated by the IHEC Funding Members. The Executive Committee will:

- Review and accept nominations of new Members
- Work closely with the International Scientific Steering Committee
- Revise and adopt new recommendations related to IHEC policies
- Track data deposition, data quality, and data accessibility across projects
- Periodically provide data updates to funding agencies
- Provide a forum for resolution of any conflicts, should they arise
- Provide a forum to resolve conflicting issues
- Make recommendations concerning the recruitment of consultants or establish expert committees on issues related to science, law, IPR, ethics, funding, communications, etc.
- Develop a communication strategy, with special focus on communication with the public.

A key responsibility of the Executive Committee will be to discuss IHEC policies and guidelines and revise these as issues such as technological improvements arise.

Transition: Following the Bethesda workshop in March 2009 (see section C), an **Interim IHEC Executive Committee** was constituted: Phil Avner, Bradley Bernstein, Susan Clark, Amanda Fisher, Thomas Jenuwein, Peter Jones, En Li, Robert Martienssen, Jacques Remacle (observer), Bing Ren, John Satterlee (observer), and Kazu Ushijima. The main mission of this Interim Executive Committee has been to prepare the draft policy document for the IHEC. Once this phase is completed, the management of IHEC will transition to the IHEC executive committee which will be formed through nominations by participating funding agency as described earlier.

- **International Scientific Steering Committee (ISSC)**

An **International Scientific Steering Committee (ISSC)** will be comprised of lead scientists on projects and principal scientific leaders in the field of Epigenetics. This group will interact frequently, through phone conferences, e-mail and regular meetings, to:

- Act as scientific coordinating body
- Assess progress
- Address arising issues of scientific nature
- Encourage exchange of the best protocols and practices
- Establish temporary or permanent subcommittees to complete focused tasks
- Establish quality standards
- Facilitate data dissemination to the scientific community

- **Data Coordination**

A **Data Coordination Center** will manage data flow from projects and centers to the IHEC database and public repositories in coordination with national or regional Data Analysis and Coordination Centers (see details in the Data Management section found later in this

document). In cases where patient protection requires an access controlled database, the DCC will provide a summary of the data available and indicate how access to the data can be obtained. The Data Coordination Center will provide regular progress reports to the EXEC and ISSC.

The Data Coordination Center will also coordinate data quality standards which will be required for each contributing Data Production Center. These quality assessments will focus on evaluation of a standardized cell model which will be distributed to all centers for evaluation by epigenomic assays. The resulting data will be evaluated by the Data Coordination Center and collaborating national or regional Data Analysis and Coordination Centers to ensure accuracy and uniformity for each production center and across the consortium.

E. Consortium Policies and Guidelines

In planning the IHEC, the Scientific Planning Committee recognised the importance of generating a document that would be communicated widely, and contain sufficient information to allow funding agencies and scientists in many countries to make decisions on future participation. This includes both a limited number of principles that are central to the project and recommendations to the readers based on what currently considered "best practices".

What is a consortium policy?

A consortium policy is a principle which consortium members agree to follow, during the course of the project. Although policies will likely be long-lasting, the IHEC will periodically review its policies.

Policies are highlighted in grey

What is a consortium guideline?

Consortium guidelines refer to recommendations made by the IHEC working groups that offer advice as to what are believed to constitute current "best practices" at a given time. Given the rapid changes in technologies and new knowledge that be gleaned from the data generated by IHEC or other groups, it is expected that these guidelines will evolve. The IHEC may choose to make most of its recommendations as guidelines rather than policies to allow flexibility in approaches and promote innovation.

In this document, the guidelines are written in blue-shaded

In this overall IHEC policy paper, the following issues must be addressed:

E.1. Informed Consent, Access and Ethical Oversight

1. Informed Consent

IHEC proposes that certain Core Ethical Elements be respected by all members as a precondition of membership. These elements apply to the collection of samples to be analysed and to consent surrounding the use of these samples (e.g. samples from disease patients) in the context of the IHEC project. Following these policies are guidelines that IHEC-member projects should consider in matters related to consent. IHEC-member projects will be responsible for carrying out these policies and guidelines, taking into account the differences between local, socio-cultural and legal requirements.

1.1 Sample donation

Core Bioethical Elements:

IHEC members should convey to sample donors that:

- The IHEC is a coordinated effort among related scientific projects being carried out around the world
- Participation in IHEC projects is voluntary
- Samples and data collected will be used to decipher human epigenomes in health and disease and will include genome sequencing
- Patient care will not be affected by their decision regarding participation
- The sample collected will be of limited quantity, sample access will be tightly controlled and will depend on the policy and practices of the IHEC-member project. A small percentage of the sample may be shared with international laboratories for the purposes of performing quality control studies.
- Data derived from the samples collected and data generated by the IHEC members will be made accessible to the IHEC members and other international researchers through either an open or a controlled access database under terms and conditions that will maximize participant confidentiality
- Those accessing data and samples will be required to affirm that they will not attempt to re-identify samples donors.
- There is a risk of being identified from data available on the databases
- Once the data is placed in open databases, it cannot later be withdrawn
- In controlled access databases, the links to (local) data that can identify an individual will be destroyed upon withdrawal.
- IHEC members agree not to make claims to possible intellectual properties derived from primary data
- No profit from eventual commercial products will be returned to subjects donating samples

Box 1. IHEC guidelines for information that should be provided to sample donors (IHEC acknowledges that the informed consent process used by IHEC members will necessarily differ according to local, socio-cultural and legal requirements)

1. IHEC administration, oversight, funding, duration, ethics, scientific approval and contact persons;
2. Who will be recruited and the approach;
3. Procedures involved in participation including physical and psychological risks;
4. Information on the kinds of samples and data that will be collected;
5. Protections in place 'locally' to ensure the confidentiality of samples and data;
6. Research uses of data (IHEC members are encouraged to seek the broadest level of consent that is appropriate at the local level)
7. Whether access to samples will be available for purposes such as validation, quality control, research, etc.;
8. Whether access to medical/administrative health records will be sought;
9. Provided it is agreed at recruitment, if clinically important and validated findings emerge during the initial recruitment and screening phase, or in the early research, attempts will be made to pass this information back via the clinician, by whatever mechanism may be agreed at the local level;
10. Information on whether or not compensation/reimbursement is available;
11. Withdrawal procedures, such as sample retrieval and/or destruction and data coding and anonymization procedures;
12. Ownership of samples;
13. Prospects for third-party commercialization and intellectual property procedures;
14. Purposes for which the uses of data and samples will not be allowed (if required to be named by country);
15. How information on the general results of the research will be disseminated;
16. Contact persons, should participants have concerns.

1.2 Human Embryonic Stem Cells

Epigenetics plays a major in cellular differentiation and cell commitment. To characterise epigenetic profiles associated with different differentiation states, some IHEC projects may use human Embryonic Stem Cells (hESC).

IHEC participants will scrupulously verify that all IHEC projects involving hESCs have the necessary legal approvals from the regional or local ethical committees prior to the start of these projects. If necessary, an ad-hoc ethical committee will be constituted to monitor ethical issues (e.g. the best way to handle incidental findings) within the IHEC projects should they arise.

1.3 Data Access and Patient Protection

The nature of the data produced by IHEC members, including clinical annotation or epigenetic profiles raises important issues concerning data protection on individuals. The patient/individual protection policies developed for IHEC are designed to balance two important goals: to facilitate investigation of epigenomic changes related to diseases and, at the same time, to respect and protect the donors/individuals whose materials and data have been or will contribute to IHEC-member projects. It is technically possible that epigenomic information generated by the projects comprising the IHEC could lead to identification of an individual if similar specimen data from that person (or a blood relative) were obtained from a third-party database and correlated. There is also a risk of individual identification by computer-based analysis of the clinical data in conjunction with, for example, third-party demographic and healthcare management databases. This potential identification could then publicly link the individual to his/her clinical information collected by the participating projects, and could lead to social risks such as discrimination or loss of privacy.

The first category, Open Access Datasets, will be publicly accessible and contain only data that cannot be aggregated to generate a dataset unique to an individual. The second category, Controlled Access Datasets, will contain composite genomic and clinical data that are associated to a unique, but not directly identified, person.

IHEC Open Access Datasets	IHEC Controlled Access Datasets

- Disease Pathology
  o Histological type or subtype
- General Patient/Person Information
  o Gender
  o Age range
- Gene expression (normalised)
- Epigenetic mapping data
  o Methylation marks
  o Histone modification positions
  o Small RNA levels
- Genotype frequencies
	- Detailed Phenotype and Clinical Data
	- Gene expression (probe-level data)
	- Raw genotype data
	- Gene-sample identifier links
	- Genome sequence files
	-

Management of controlled access data will likely require dedicated hardware for storage and computing. A credentialing server is required to manage data access and the DCC will work closely with the organizations that grant access to researchers within and beyond the consortium. Archival data will be stored at dbGAP for data generated by US producers and at the EGA (European Genome-phenome Archive) for non-US producers as is done by the ICGC. Procedures enabling researchers to obtain controlled access data are already in place.

E.2. Data Quality and Standards

E2.1 Experimental approaches for generating large-scale epigenomic datasets

POLICY: The reference epigenome map of a cell may include the collection of DNA methylation, chromatin modifications, positions of nucleosomes and variants, and abundance of each RNA species. A number of technologies have been developed for each kind of epigenomic dataset. IHEC membership requires that the experimental approaches used by each member to generate epigenome dataset be in accordance with the following criteria:

- The accuracy of the method must have been extensively validated, by more than one research group, and have been published in peer-reviewed journals;
- The method should interrogate epigenetic events across the entire genome;
- The method should examine the epigenetic event at a high resolution;
- The detailed assay methods should be readily available to the public;
- Instrumentation required for the method should be generally accessible.

Guidelines

1. To facilitate distribution, interpretation and comparison of epigenomic data from different research centers, the IHEC will, through working groups, make recommendations and guidelines on the experimental approaches for generating the various epigenome maps.

E2.2 Quality control for biological materials and reagents -

POLICY:

- To ensure general utility of the reference genome maps generated by each IHEC member, it is necessary that the source of the biological materials is properly documented, and a set of standard quality measurements are implemented.
- Biological or chemical reagents used for the generation of epigenome maps should be properly characterized and the documentation should be publicly available.
- Antibodies are key reagents for mapping the epigenome, either for ChIP based analysis of chromatin modifications, or mapping methylated cytosines. There are a growing number of well-characterized antibodies. However, great variation in quality exists between antibodies from different vendors and between different lots, therefore all antibodies used for these experiments should be validated. The data generated to characterize the antibody should be made publically available.

Guidelines:

IHEC will, through working groups, make recommendations on the tissue collection, cell culturing, reagents documentation and antibody validation procedures.

E2.3 Data quality verification and validation -

POLICY:

- To ensure the accuracy and reproducibility of the reference epigenome maps, it is important that at least two biological replicates are carried out for each type of epigenome map.

- The accuracy of the data should be validated using a gold standard method to determine the specificity and sensitivity of the large-scale analysis.

Guidelines:

- The IHEC will make recommendations on the criteria for quality control process on each different experimental approach used for obtaining epigenomic information.

E2.4 Data reporting -

POLICY

- IHEC membership requires immediate and pre-publication release of epigenome map dataset, in accordance with data release policy below (E4);

Guidelines

- IHEC will make recommendations on the data format and quality control procedure.
- Data should be submitted to public databases, and should be flagged as being part of the IHEC project upon submission
- Processed data (metadata) should also be submitted to the relevant DCC.

E.3. IHEC Data Coordination Center and Data Management

POLICY: The IHEC Data Coordination Center (DCC) will be established to ensure consistency and efficiency in data formats

1. Initial processing and availability of both freely available non-patient epigenomic data (e.g. cell lines, model organism data) and data subject to usage limitations based on the consent obtained from the patient. Such an effort requires large scale bioinformatics support.
2. In coordination with national- or regional-level Data Analysis and Coordination Centers, the DCC will implement a modular and flexible data flow pipeline capable of accepting, processing and tracking a range of mixed data types. The pipeline codes provides coordinated systems for supporting a variety of analysis programs that are all capable of reading and writing standard input and output file formats.
3. The NIH Roadmap Epigenomics Program and ENCODE projects have established and emerging standards for handling with epigenomic data obtained via sequencing platforms. These standards can be adopted almost immediately.

4. For other data types, such as DNA methylation data, the concepts in this paragraph apply, but the standard formats and analysis methods may be under development. Briefly, sequencing data is submitted by the production/ mapping centers in SRF/FASTQ/BAM and initial consistent processing (such as alignment for sequence data) is done either at the production center or the DCC based on agreed procedures. The completion of these steps are recorded in a DCC- created tracking database that is widely visible so that the pipeline code (and the people working in the project) can know what has been done, what needs to be done and allows identification of files that have apparent problems either due to bad data or software/hardware failure. Raw and processed data files must be subject to quality assessment to ensure that the proper samples were assayed and file integrity assessment to prevent data corruption and the results of these assessments stored.
5. Sample data storage and a sample identification method will be appropriately managed so that resources created by multiple members of the IHEC and other projects can be integrated. For example, other projects may generate whole-genome sequences, phenotype resources or other complementary data that will enhance the value of IHEC data.
6. Management of controlled access data must be done at scale and will likely require dedicated hardware for storage and computing. A credentialing server is required to manage data access and the DCC will work closely with the organizations that grant access to researchers within and beyond the consortium. Archival data will be stored at dbGAP for data generated by US producers and at the EGA (European Genome-phenome Archive) for non-US producers as is done by the ICGC. Procedures enabling researchers to obtain controlled access data are already in place.
7. The last step in the DCC pipeline is to make the data available for analysis groups within and beyond the project. This includes loading summary and raw data into visualization systems such as a genome browser; making data available through commercial or consortium compute clouds; and/ or creating queryable database systems such as Biomarts.
8. Transfer of large data sets requires appropriate infrastructure including adequate bandwidth and robust hardware configurations on the sending and receiving end. In coordination with national or regional Data Analysis and Coordination Centers, the DCC will work with major Internet backbone providers to ensure consistent support.

E.4. IHEC Data Release Policies

A guiding principle for research funded by public (governmental and charitable) organizations is to maximize benefit to the public while, at the same time protecting the interests of sample donors and their relatives. The data release policies of the IHEC are intended to achieve these dual objectives.

Responsibility of data producers and users

POLICY: The members of the International Human Epigenome Consortium (IHEC) are committed to the principles of rapid data release to the scientific community

The IHEC consortium will deliver large data sets on human epigenomes in health and diseases, thereby establishing a unique data resource that will be freely accessible to the scientific community. As for other large data resource projects (including genome-sequencing projects, (the Roadmap Epigenomics Program, the ENCODE Project, the international HapMap Project, the 1000 Genomes Project, the International Cancer Genome Consortium, the Human Microbiome Project, and the MetaHIT project), IHEC will commit to the principle of rapid data release to the scientific community. To implement this policy, the IHEC members (Funding agencies, Data producers, Data analysts/users) will follow the recommendations made during the Toronto (2009) conference that was sponsored by the NIH, Wellcome Trust, EU Commission and Genome Canada. These Toronto principles are listed in the table below.

Funding agencies should facilitate the specification of data-release policies for relevant projects by:

1. Explicitly informing applicants of data-release requirements, especially mandatory
2. prepublication data release
3. Ensuring that evaluation of data release plans is part of the peer review process
4. Proactively establishing analysis plans and timelines for projects releasing data prepublication
5. Fostering investigator-initiated prepublication data release
6. Helping to develop appropriate consent, security, access and governance mechanisms that protect research participants while encouraging prepublication data release
7. Providing long-term support of databases

Data producers should state their intentions and enable analyses of their data in coordination with central or regional Data Coordination Centers by:

1. Informing data users about the data being generated, data standards and quality, planned analyses, timelines, and relevant contact information, ideally through publication of a citable marker paper near the start of the project or by provision of a citable URL at the project or funding agency website
2. Providing relevant metadata (e.g., questionnaires, phenotypes, environmental conditions, and laboratory methods) that will assist other researchers in reproducing and/or independently analysing the data, while protecting interests of individuals enrolled in studies focusing on humans
3. Ensuring that research participants are informed that their data will be shared with other scientists in the research community
4. Publishing their initial global analyses, as stated in the marker paper or citable URL, in a timely fashion
5. Creating databases designed to archive all data (including underlying raw data) in an easily retrievable form and facilitate usage of both pre-processed and

Data analysts/users should freely analyse released prepublication data and act responsibly in publishing analyses of those data by:

1. Respecting the scientific etiquette that allows data producers to publish the first global analyses of their data set
2. Reading the citeable document associated with the project
3. Accurately and completely citing the source of prepublication data, including the version of the data set (if appropriate)
4. Being aware that released prepublication data may be associated with quality issues that will be later rectified by the data producers
5. Contacting the data producers to discuss publication plans in the case of overlap between planned analyses
6. Ensuring that use of data does not harm research participants and is in conformity with ethical approvals

Scientific journal editors
should engage the research community about issues related to prepublication data release and provide guidance to authors and reviewers on the third-party use of prepublication data in manuscript

For data sets obtained from patient samples, IHEC will take all necessary precautions to adhere to patient wishes and avoid potential patient identification. As a result of this policy, patient sample data will also be released rapidly but will only be accessible through a controlled access procedure (see section E.2.).

E5. IHEC Publication Policy

POLICY:

- Individual IHEC investigators may publish the results of their own work. PIs within the consortium may wish to notify other consortium members of their intent to publish or to coordinate back to back publications to maximize scientific impact.
- To balance the interests of all stakeholders, resource users are asked to respect the ability of the IHEC members to publish an initial analysis of their own data in a timely manner so as not to slow the progress of science. For the purpose of this data release policy, "timely" refers to an initial period of **nine months** after the **release** of the data into public databases.

The period of nine months after the release of data into public databases will provide time for the resource producers to have a protected opportunity to **publish** initial analyses of the data they have generated. The nine-month period will be established for each submitted dataset by the creation of a timestamp at the time the data are posted for public access and will apply to all data types, including primary, interpreted (as defined above), validation, and biological characterization data. The timestamp will be maintained by the public database where the data will be publicly accessed and archived.

During this time period, resource users may work without restriction to analyze, and otherwise use the data in their own work, but they are requested not to submit their analyses or conclusions for publication. The publication moratorium by resources users ends either at the expiration of the nine-month protected period or when the data have been published, whichever is shorter. During the protected period, resource users and resource producers are encouraged to communicate their activities for purposes of either establishing collaborations or organizing simultaneous publications.

To facilitate comparison of data between different groups involved in the IHEC, all publications by IHEC members should, to the extent possible, include data on the reagents used including common cell lines or common antibodies to encourage members of the scientific community to generate data using these reagents to enable comparison to IHEC data.

Investigators outside of the IHEC are encouraged to use the data generated from this endeavor, and are asked to follow the guidelines elaborated in this document. Any user of the data is responsible for being aware of the publication status of the data they use and treat them accordingly. IHEC will monitor the use of this policy; evidence of significant misuse will result in modifications to these guidelines.

Use of unpublished data within the initial 9-month period:
Investigators accessing data generated by the IHEC and performing analyses on unpublished data are urged to proceed with submission for publication only after the 9-month period has elapsed. If there is a strong desire to publish prior to the 9-month period, data users are expected to use appropriate scientific etiquette and 1. discuss their plans to use the pre-publication data with the IHEC resource producer(s) and 2. should obtain their consent prior to using the unpublished data in their individual publications or grant submissions. The time stamp-related moratorium on publication is expected to apply to submission of manuscripts for publication by resource users. Resource users are expected to acknowledge the IHEC data producers, as mentioned previously, as well as to

acknowledge to the funding organization(s) that supported the work in all resulting oral or written presentations, disclosures, or publications of the analyses. All investigators, through their roles as journal and grant reviewers, and journal editors, are asked to help maintain a high standard of respect for the scientific contribution of the IHEC members.

Use of published or unpublished data for which the 9-month time stamp has expired:
Following expiration of the projected publication period, any investigator may submit manuscripts without restriction, including integrated analyses using multiple unrestricted datasets. For unpublished and unrestricted datasets, users are expected to provide proper acknowledgement as suggested above.

E6. IHEC Intellectual Property Policy

Long-standing scientific policies have encouraged the rapid release and ready accessibility of genomic data to the broad research community. A related issue of availability pertains to any intellectual property rights that might be sought by data generators, and the effects that the exercise of such rights have on access to the data.

In some respects, the cases of genomic sequence data and haplotype data were relatively easy to deal with because the data themselves do not have "utility" (in the patent law sense of the term). However, for epigenomic data the applicability of this argument is not as obvious. One purpose of the IHEC is to generate data that identify or define epigenomic marks and chromatin states that have biological function, and therefore might be considered to have utility and be able to be patented. Therefore, the use of patents in ways that might restrict access to large amounts or broad categories of data, e.g., all DNAse hypersensitivity sites, is an issue that needs to be addressed.

POLICY: The IHEC's primary interest is to ensure the widespread availability of all information and any inventions that are generated and, therefore, encourages all epigenomics data resource producers to consider placing all information generated from their project-related efforts in the public.

In the cases in which the consortium members elect to exercise their intellectual property rights, IHEC encourages consideration of maximal use of non-exclusive licensing of patents to allow for broad access and stimulate the development of multiple products.

The IHEC also encourages users of IHEC data to act responsibly and share the effort involved in maintaining unrestricted access to the data.

Thus, for example, if a resource user were to incorporate IHEC generated epigenomic data into an invention, the subsequent license should not restrict the access of others to the IHEC epigenomic data. For this purpose, the term "resource users" is meant to include both researchers who are members of the IHECs and researchers who are not.

E7. Software Sharing Policy

POLICY: IHEC participants will fully disclose algorithms, software source code, and experimental methods to the other members of the consortium for purposes of scientific evaluation. IHEC strongly encourages consortium participants to make these methods available to the broad research community as well.

1. The software should be freely available to biomedical researchers and educators in the non-profit sector, such as institutions of education, research institutions, and government laboratories.
2. The terms of software availability should permit the commercialization of enhanced or customized versions of the software, or incorporation of the software or pieces of it into other software packages.
3. To preserve utility to the community, the software should be transferable such that another individual or team can continue development in the event that the original investigators are unwilling or unable to do so.
4. The terms of software availability should include the ability of researchers to modify the source code and to share modifications with other colleagues. The software engineer is responsible for creating the original and subsequent "official" versions of a piece of software, and should provide a plan to manage the dissemination or adoption of improvements or customizations of that software by others. This plan should include a method to distribute other user's contributions such as extensions, compatible modules, or plug-ins.

E8. Tissues Selection and Model Organisms: IHEC coordination

POLICY: The IHEC International Scientific Steering Committee will provide recommendations on the prioritization of tissues and epigenomes to be analyzed.

A human epigenome project is inherently more complex than other genomic projects because there are many human epigenomes. Work has already begun in earnest to map embryonic stem cell epigenomes, their differentiated derivatives, and selected primary tissues as part of the NIH Roadmap Epigenomics Program. There are many potential normal and disease states of interest making tissue and disease type selection important. Additionally, model organisms can provide profound insights into the functional relevance of the epigenetic marks being mapped by IHEC members.

1. There is therefore a need to coordinate the selection of tissues to be studied, epigenomes to be analyzed, what detail is acceptable, what quality standards will be applied, etc.

2. The consortium acknowledges that it is important to limit duplication of activities and minimize redundancy at the same time as it is necessary to have a structure which allows for data to be cross-checked in different locations.

3. The recommendations of priorities for which tissues and epigenomes will be studied will be the responsibility of the International Scientific Steering Committee.

4. The Committee will provide real time feedback to members as to which specific projects are being undertaken and/or planned by the different members.

5. The members agree to provide timely information to the steering committee as to which specific projects they are focused on in their epigenome studies. For example, the Roadmap Epigenomics Program has already begun to map embryonic stem cell epigenomes, their differentiated derivatives, and selected primary tissues. This knowledge will be useful in helping other consortium members undertake epigenetic analysis of additional tissue types and disease states. Another example might be the use of model organisms such as yeast, Arabidopsis or knockout mice which will be used to probe the functional interrelationships between different epigenetic marks.

6. IHEC members therefore agree to provide information concerning epigenomic samples to be analyzed to the steering committee which can be displayed on the website before starting a project to avoid duplication and redundancy of activities.

LETTER OF INTENT TO JOIN THE INTERNATIONAL HUMAN EPIGENOME CONSORTIUM (IHEC)

To: Chair, IHEC Executive Committee
c/o Eric Marcotte, PhD
CIHR - Institute of Genetics
McGill University
3649 Promenade Sir William Osler
Room 279
Montreal, QC
H3G 0B1
Canada

LETTER OF INTENT TO JOIN THE INTERNATIONAL HUMAN EPIGENOME CONSORTIUM (IHEC)

The International Human Epigenome Consortium (IHEC) coordinates an international collaborative effort for the production of reference maps of epigenomes for key cellular states relevant to health and disease. IHEC sets the ambitious goal to decipher at least 1000 epigenomes within the next 7-10 years to have a substantial coverage of the human epigenome (Nature, Vol 463, 4 February 2010, 596-597).

This letter of intent is to inform the IHEC Executive Committee that [funding agency name] from [country] is willing to join the IHEC collaborative effort and planning to finance research with an anticipated budget of [foreseen budget] aligned with the IHEC scientific objectives, funding level, policies and governance structure.

Our agency and the scientists involved in the funded projects will follow the principles listed below to join this international collaboration. The principles are as follows:

- Free and open release of data and resources generated;
- Promotion of maximum efficiency in the generation of epigenome mapping technologies through the sharing of information regarding new approaches and technologies developed during the programs;
- Coordination of the efforts through the sharing of production plans to minimize unnecessary redundancy;
- Coordination of public communications regarding individual efforts or the international effort as a whole, including the establishment of a common Web portal; and
- Coordination on issues such as archiving and distribution to ensure the data and resources generated are readily accessible to the scientific community.

We therefore nominate [Mr, Ms, Dr, Prof] [name of the representative of the funding agency] to represent our funding agency on the IHEC Executive Committee with observer status until the above funding is in place. We also nominate [Dr. Prof., name of the scientific principal investigator] to be an observer member of the International Scientific Steering Committee.

[If you have the information available, you are encouraged, but not required, to provide a brief description of your planned IHEC aligned program, which may include a listing of proposed cell/tissue types to be mapped.]

Signature
Name of [Minister]/[Director]/[Funding Agency Legal Representative]
Date