International Human Epigenome Consortium Phase II Roadmap Document

(November 2018)

A. Introduction

The sequence of the human genome provides a foundation for understanding cellular processes in health and disease. However, the way in which this primary genetic information is organized within the cell also plays a critical role in its interpretation and execution. Epigenetic processes are essential for packaging and higher level organization of the genome and thus are fundamental to normal development and cell differentiation, and are increasingly recognized as being involved in human disease.

Motivated by a desire to provide a foundation from which to build an understanding of normal epigenome states leading scientists and research funding agency representatives founded the International Human Epigenome Consortium (IHEC) in the fall of 2010. Primary goals for this first phase of IHEC were to define and coordinate the production of reference maps of human epigenomes for key cellular states relevant to health and diseases, to facilitate rapid distribution of the data to the research community, and to accelerate translation of this new knowledge to improve human health. A critical component of IHEC in reaching these goals was to coordinate the development of common bioinformatics standards, data models and analytical tools to organize, integrate and display the large quantities of epigenomic data generated.

Since its inception IHEC members have generated epigenomic datasets from more than 3,400 samples including close to 380 complete reference epigenomes from a wide variety of primary human tissues and cells. These data have been made available in a processed form to the research community through interactive data portals (e.g. www.epigenomesportal.ca/ihec/) and in an unprocessed form through protected and open access public repositories. In depth analyses of these data and experimental and computational methodologies for epigenomic analysis have been reported in over 2,700 primary publications including in the context of two coordinated packages in Nature and Cell family journals.

Ongoing technological advances, including those developed by IHEC members have driven increasing efficiencies and resolution to existing epigenomic measurements while enabling new dimensions of epigenomic analysis. As these technologies mature, IHEC members have begun to apply them in the context of epigenomic mapping and new initiatives provide opportunities for IHEC to add additional dimensions (e.g. time,

variation at the cellular and organism levels, higher order structures) to existing epigenome maps while continuing to expand the catalog of cell types profiled (**Fig. 1**).

B. Background

Epigenetic mechanisms include histone and DNA modification, positioning of nucleosomes and histone variants, organization of higher order chromatin structures, and small and non-coding RNAs, among others. In concert with transcription factors and other DNA-binding proteins, the resulting epigenetic states, which may be inherited through cell division, regulate RNA 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 distinct. The modulation of these epigenetic profiles is essential for normal embryonic development, differentiation, and cell identity, and thus transitions from a stem cell to a lineage-committed cell, and underlies responses to environmental signals (e.g., metabolites, hormones, nutrients, stress, and damage). In many respects, the epigenetic interpretation of the genome (i.e. epigenomic information) represents a "second code" that programs DNA-based information in diverse biological contexts.

Errors 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 possibility of therapeutic targets has significant implications for the prevention and treatment of these major human diseases. Indeed, several inhibitors of chromatin-modifying enzymes, including histone deacetylase and histone and DNA methyltransferase inhibitors are being used effectively in clinical practice. 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, 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. 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 states 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 induced by environmental and nutritional changes,

so that maps for reference epigenomes will greatly broaden our understanding of how the environment and nutrition will modulate epigenetic alterations. This new knowledge will have a major impact for novel avenues in preventing and diagnosing disease.

While epigenetic mechanisms have been implicated in cancer as well as rare and common diseases that impact growth, aging and neurodevelopment, there has not been a concerted effort to coordinate and translate emerging findings into society. Such an effort becomes more demanding when one considers how research has made incredible strides in the understanding of epigenetic mechanisms while increasingly public perception and action has become decoupled from the science. IHEC will continue to address this challenge through the development of policies to advance translation of epigenetic findings into society.

Technological advancements now allow for the reproducible and standardized mapping of epigenomic information. 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 and histone modification, nucleosome position and corresponding coding and non-coding RNA expression in different normal and disease cell types. The epigenome reference maps have been 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[1]. Although IHEC maintains its' human focus, it will be essential to involve model organisms to obtain mechanistic insights as to the functionality of epigenomic parameters.

Studies in model organisms such as yeast, fly, and mouse have yielded fundamental discoveries across many fields of biology, including notable advances in our knowledge 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. In support of these efforts IHEC will continue to 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 (**Fig. 1**). The outcome of the research carried out by the members of IHEC will continue to be extensive[2, 3]. First and foremost is availability of an increasing catalog 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.

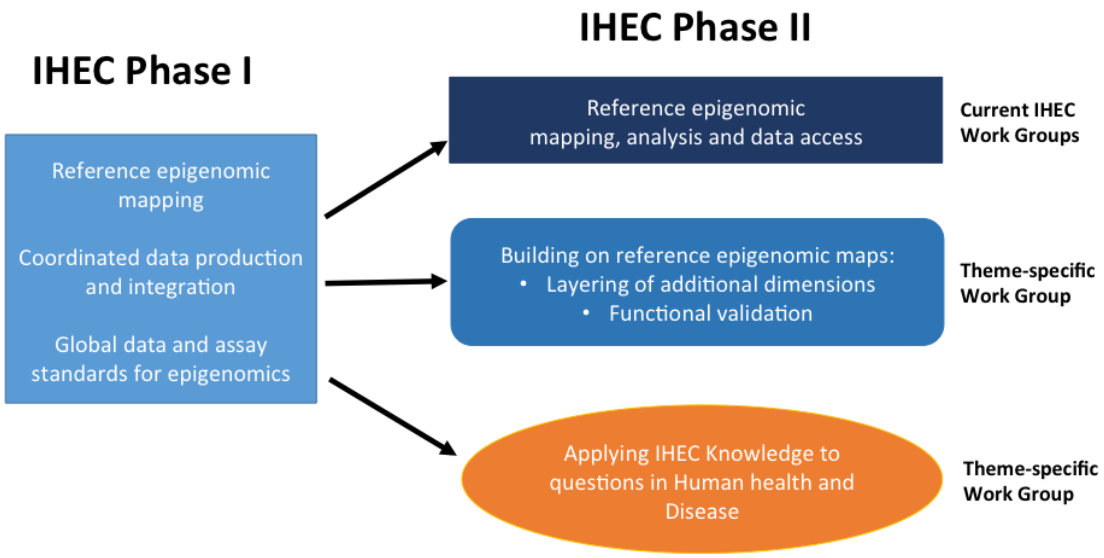

Fig. 1: IHEC phase II themes

C. Consortium Goals

1. DATA GENERATION

Coordinate the generation and annotation of sequencing based human epigenomic datasets from preferentially primary human cells and tissues [2,3] including quantitative measurements of histone and DNA modifications and RNA expression. Adherence to standardized experimental and bioinformatics procedures and standards. Coordination of reference cell production to avoid or minimize redundancy and maximize the diversity to generate a human epigenome cell atlas.

2. DATA INTERPRETATION

Develop methods for integrated epigenome interpretation. Use epigenomes to connect expression and regulation on a cell specific level. Generate methods and standards for quantitative chromatin state measurements, chromatin dynamics, open[4] and higher order chromatin contacts[5] and single cell measurements[6, 7]. Link this information to existing two dimensional reference epigenome maps. Generate tools to decompose/deconvolute complex multicellular/multistate datasets using these high-resolution single cell data.

3. DATA DISTRIBUTION

Coordinate rapid distribution of unprocessed and uniformly processed data to the entire research community with minimal restrictions, to accelerate translation of this new knowledge into health and diseases. IHEC will develop and distribute common bioinformatics standards, data models and analytical tools to organize, integrate and display epigenomic data generated by members and the broader epigenetics research community.

4. BIOMEDICAL APPLICATION

Coordinate the production and public availability of reduced representative epigenomic datasets (primarily NGS) and their interpretation based on IHEC reference epigenomes. Use IHEC resources to generate comparable datasets in medically relevant cohorts comparing normal and diseased cell types from. Develop IHEC standards to quantitate and interpret epigenetic variation within and across normal and diseased human populations.

5. KNOWLEDGE MOBILIZATION

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

References

1 Stricker, S. H., Koferle, A. & Beck, S. From profiles to function in epigenomics. *Nat Rev Genet* **18**, 51-66, doi:10.1038/nrg.2016.138 (2017).

2 Stunnenberg, H. G., International Human Epigenome, C. & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1145-1149, doi:10.1016/j.cell.2016.11.007 (2016).

3 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

4 Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322, doi:10.1016/j.cell.2007.12.014 (2008).

5 Davies, J. O., Oudelaar, A. M., Higgs, D. R. & Hughes, J. R. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods* **14**, 125-134, doi:10.1038/nmeth.4146 (2017).

6 Angermueller, C., Lee, H. J., Reik, W. & Stegle, O. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol* **18**, 67, doi:10.1186/s13059-017-1189-z (2017).

7 Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* **11**, 817-820, doi:10.1038/nmeth.3035 (2014).