INTERNATIONAL HUMAN EPIGENOME CONSORTIUM

The International Human Epigenome Consortium is enabling free and effective distribution of epigenomic data, all while maintaining best practice guidelines. Returning with similar vigour and enthusiasm, Chair of the Executive Board **Dr Eric Marcotte** speaks openly about the aims of the Consortium, the progress of its international efforts and why epigenomics research is so essential to further understand, and ultimately prevent, both common and complex diseases

In your opinion, why is epigenomics research particularly important?

Epigenetic mechanisms provide a compelling model for how environmental influences can modify the expression of the genome in long-lasting ways. The traditional life science approach of identifying candidate genes – and testing their role in model systems – has proven incredibly valuable in helping to understand a range of disorders. But many of the variants detected through the Genome-Wide Association Studies (GWAS) map to noncoding regions of the genome. A comprehensive understanding of epigenomics could help unlock the origins of complex diseases, by shining a light on the underlying regulatory processes that have proven resistant to analysis by conventional genomic approaches. By their very nature, many common diseases are multifactorial, and likely the result of complex processes involving the interplay of both genetic and environmental contributions.

What are the aims of the International Human Epigenome Consortium (IHEC)?

The overarching objective of IHEC is to help address fundamental questions on how the environment interacts with us, during development and ageing, to influence health and disease. Ultimately, our goal is to translate these discoveries into improved human health, with a better understanding of gene-environment mechanisms. At present, we are focused on reference epigenome mapping in 1000 cells and tissues.

Can you elaborate on the reference epigenome standards set by the IHEC? Why are these important?

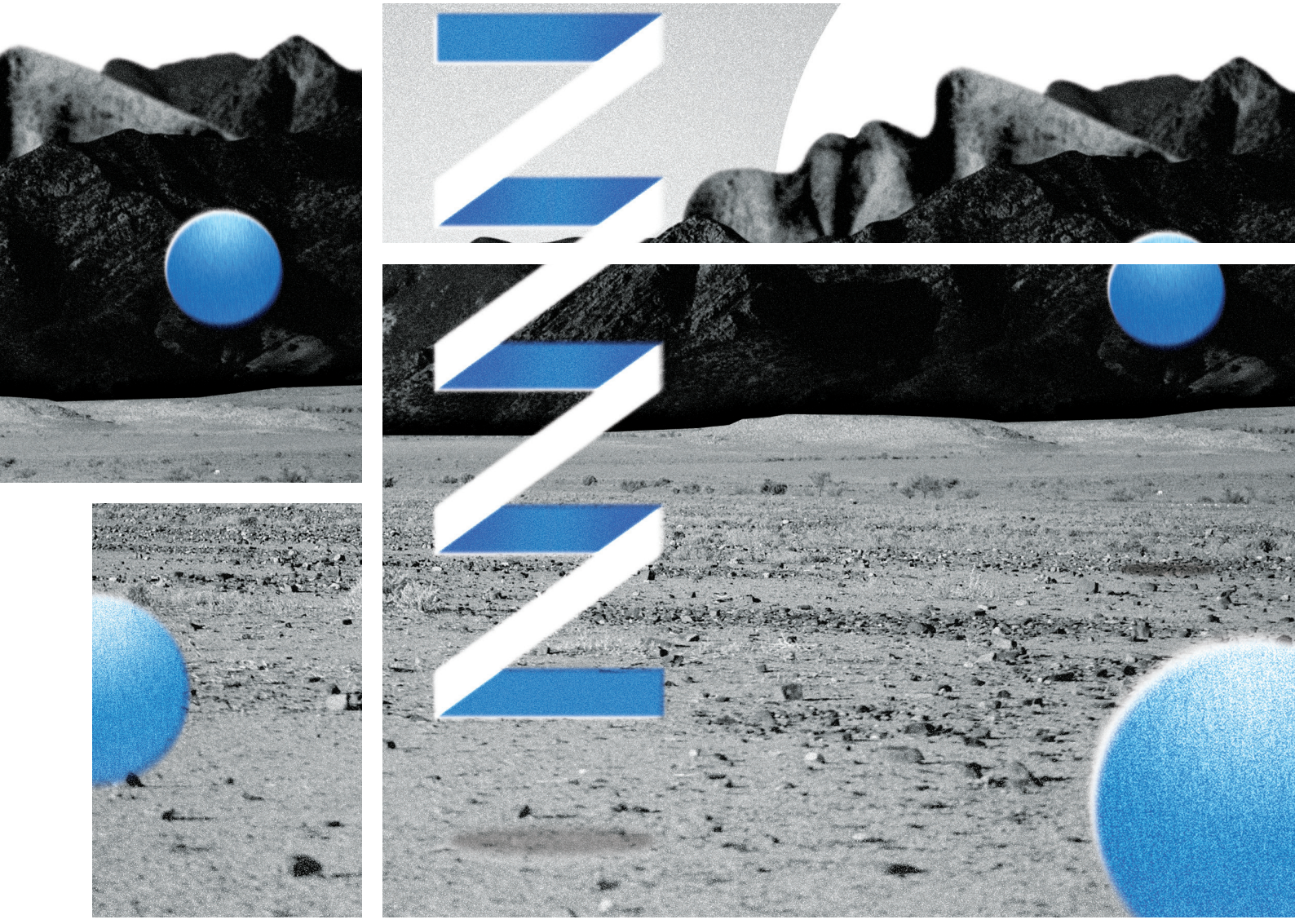
IHEC reference epigenome guidelines are minimal standards based on current knowledge of the elements that contribute to epigenomic regulation in humans, and the current state of epigenomic mapping technologies. As technology and research continue to advance, these standards are regularly reviewed and revised by the core IHEC workgroups and committees. In a sense, you could consider the initial reference standards developed and approved by IHEC as a second-generation set, building upon earlier national programmes (eg. the National Institutes of Health (NIH) Roadmap Epigenomics programme). We have continually updated these guidelines with incremental revisions over the last few years, and are currently looking at a more significant versioning change with an expanded set of standards.

Is it difficult to perform epigenomic analyses?

While recent technological advancements have facilitated the detailed analysis of epigenomic marks across the whole genome, there certainly remain challenges to performing accurate, comprehensive and reproducible epigenomic studies. These require not only access to state-of-the-art genomic sequencing technologies, but also familiarity and experience with rapidly developing approaches and methodologies. IHEC has established classes of high-resolution mapping approaches that we believe will accelerate the scientific exploration of human epigenomic information. In that process, we have defined the assays required for these classes, and set standardised protocols and quality control metrics for each assay. As these standards are made freely available to the larger community, they provide a framework for any group seeking to perform similar high-quality epigenomic mapping.

By what other means is the IHEC enabling the effective distribution of epigenomic data?

As international research leaders, many of the members of the IHEC workgroups and committees also serve on other national and international initiatives involving large data sets of genomic information. Just as this brings in to IHEC the experience of approaches taken (and lessons learned) from other consortia, it also allows for a broader consideration of epigenomic information into the activities of those consortia. The dissemination of best practices within aligned fields is a



key indicator of initiative success, and we know that our members have influenced other consortia on what sort of epigenomic data they consider, and how to best interpret and represent it.

This high-level coordination between international consortia is key to ensuring the maximum benefit of our individual efforts. One example is how IHEC adopted a common set of core ontologies in organising and describing our research output. Ontologies are the formal naming and definition of terms, properties and interrelationships of entities within a given system. Ontologies are critical to organise and relate information in genomic analyses, to allow accurate interpretation and sharing of information. While each consortium will have unique needs (depending on the specific assays, technologies, tissues, etc. that are used), it is critical for data interoperability that a consistent set of ontologies be applied in areas where consortia overlap. The ENCODE project (a consortium that is building a comprehensive list of all the functional elements in the human genome), the NIH Roadmap Epigenomics programme, and IHEC all approved a common core set of ontologies in 2013.

The Data Portal is an example of how IHEC was able to move quickly to integrate the perspectives and needs of all consortium members

What is the IHEC Data Portal and why is this a key initiative?

One of the main goals of IHEC is to make publicly available comprehensive sets of reference epigenomes relevant to human health and disease in a timely way. The IHEC Data Portal is a new community resource that can be used to view, search and download all the open access data released by the different IHEC-associated projects. It was custom designed to meet the needs of the epigenomic research community, with a sophisticated set of tools to easily isolate and identify specific tracks of relative interest among the extensive (and growing) collection of IHEC epigenomic data sets. All the member projects work closely with the Portal designers to keep the content updated, and to advise on new features and functionality. The current ability to overlay specific epigenomic data on a popular genome browser – and download all of the full IHEC data sets right from the common portal – make it a particularly valuable resource to the larger community.

Given the distributed nature of IHEC, can you explain how the resources were found to create the Portal?

While the Data Portal's content is truly a collaborative effort across all of IHEC, the main credit for developing and maintaining the Portal itself

rests largely with one of the Canadian-funded IHEC data analysis and coordination centres. This centre was able to access significant additional national resources in Canada, including programming support and a high-performance computing cluster to manage the large volume of international data associated with the generation of IHEC reference epigenome maps. The successful leveraging of new resources into the Consortium speaks to the strength of the IHEC's coordinated approach, and the perceived value by other stakeholders.

THE DATA PORTAL – AT A GLANCE

IHEC makes available comprehensive sets of reference epigenomes relevant to health and disease. The IHEC Data Portal can be used to view, search and download the data already released by the different IHEC-associated projects.

Developed and maintained by:

The McGill Epigenomics Data Coordination Centre (EDCC)

The McGill Epigeonmics Mapping Centre (MEMC)

Funded by:

Canadian Institutes of Health and Research (CIHR)'s Canadian Epigenetics Environment and Health Research Consortium (CEEHRC)

Genome Quebec

Genome Canada

Compute Canada

CANARIE

How challenging is the task of coordinating the IHEC's international members and stakeholders?

The Data Portal is an example of how IHEC was able to move quickly to integrate the perspectives and needs of all consortium members. Within six months, our pan-IHEC working group on data integration succeeded in designing and launching the Portal with a preliminary data set from several member countries. Within another three months, the first full release from all the data-producing members was made available to the larger community. That is a remarkable timeframe and achievement, given the complexity of creating a novel infrastructure and integrating thousands of data sets dispersed across the globe. It is a testament to the level of engagement and dedication of the members of the IHEC workgroups and committees.

What is the current status of epigenomics research?

This is a question best answered by the research leaders on our International Scientific Steering Committee. But from my perspective, I note that there have recently been many improvements on the key epigenomic technologies used by IHEC, as highlighted at our most recent Science Day. The ability to perform many of our core assays – to our current quality standards and read depths – greatly reduced amounts of starting material and opens up numerous new possibilities.

Neuroscience is a good example of a field where it is difficult to obtain sufficient primary tissue for analysis from human subjects. Any advances that reduce the amount of tissue required, or that can allow a wider range of previously isolated and stored material to be used, can thus have huge practical benefits in facilitating new research. It also opens up the possibility of expanding our reference epigenome standards set, potentially allowing interesting new assays to be included for the same amount of initial tissue going forward.

Where do you envisage this field advancing over the next few years?

The increasing number of epigenomic analyses per cell type that we will be able to perform should help us better understand the impact of genetic variation and environmental factors on health. I expect that epigenomic marks will increasingly form part of longitudinal studies in healthy ageing, leading to the potential development of new and robust biomarkers of disease.

More immediately, a common theme that I have noticed among many genomic consortia is the importance of performing integrative analyses of the comprehensive datasets we generate. While the individual datasets themselves can be of great significance and open up new avenues of research in specific fields, the potential to glean fundamental insights from integrative analyses can be revolutionary. This is especially true when combining genomic and epigenomic data, in both healthy and diseased cells and tissues. These could broadly redefine our understanding of the genome and its regulation, up-ending previous dogmas and establishing whole new fields or approaches. The computational and bioinformatics resources needed for this are significant, and I expect this to be a major focus for many national and international initiatives in the coming years. While hard to predict where this will take us, none of this would be possible without the coordinated multilateral approach to generating comprehensive reference datasets through consortia like IHEC in the first place.