

BLUEPRINT TO DECODE THE EPIGENETIC SIGNATURE WRITTEN IN BLOOD

David Adams¹, Lucia Altucci², Stylianos E Antonarakis³, Juan Ballesteros⁴, Stephan Beck⁵, Adrian Bird⁶, Christoph Bock⁷, Bernhard Boehm⁸, Elias Campo⁹, Andrea Caricasole¹⁰, Fredrik Dahl¹¹, Emmanouil T Dermitzakis³, Tariq Enver⁵, Manel Esteller¹², Xavier Estivill¹³, Anne Ferguson-Smith¹⁴, Jude Fitzgibbon¹⁵, Paul Flicek¹⁶, Claudia Giehl¹⁷, Thomas Graf¹³, Frank Grosveld¹⁸, Roderic Guigo¹³, Ivo Gut¹⁹, Kristian Helin²⁰, Jonas Jarvius²¹, Ralf Küppers²², Hans Lehrach²³, Thomas Lengauer⁷, Åke Lernmark²⁴, David Leslie¹⁵, Markus Loeffler²⁵, Elizabeth Macintyre²⁶, Antonello Mai²⁷, Joost HA Martens²⁸, Saverio Minucci²⁹, Willem H Ouwehand¹⁴, Pier Giuseppe Pelicci²⁹, Hélène Pendeville³⁰, Bo Porse²⁰, Vardhman Rakan¹⁵, Wolf Reik³¹, Martin Schrappe³², Dirk Schübeler³³, Martin Seifert³², Reiner Siebert³⁴, David Simmons³⁵, Nicole Soranzo¹, Salvatore Spicuglia³⁶, Michael Stratton¹, Hendrik G Stunnenberg^{28*}, Amos Tanay³⁷, David Torrents³⁸, Alfonso Valencia³⁹, Edo Vellenga⁴⁰, Martin Vingron²³, Jörn Walter⁴¹, Spike Willcocks⁴²

¹Wellcome Trust Sanger Institute, Hinxton, UK.

²Second University of Naples, Naples, Italy.

³Universite De Geneve, Geneva, Switzerland.

⁴Vivia Biotech, Madrid, Spain.

⁵University College London, London, UK.

⁶University of Edinburgh, Edinburgh, UK.

⁷Max Planck Institute for Informatics, Saarbrücken, Germany.

⁸University of Ulm, Ulm, Germany.

⁹Institut D'investigacions Biomediques August Pi I Sunyer, Barcelona, Spain.

¹⁰Siena Biotech Spa, Siena, Italy.

¹¹Halo Genomics AB, Uppsala, Sweden.

¹²Institut D'investigacio Biomedica De Bellvitge, Barcelona, Spain.

¹³Centre for Genomic Regulation, Barcelona, Spain.

¹⁴University of Cambridge, Cambridge, UK.

¹⁵Queen Mary University of London, London, UK.

¹⁶European Bioinformatics Institute, Hinxton, UK.

¹⁷European Research and Project Office GmbH, Saarbrücken, Germany.

¹⁸Erasmus University Medical Centre Rotterdam, Rotterdam, The Netherlands.

¹⁹Centro Nacional de Analisis Genomico, Barcelona, Spain.

²⁰University of Copenhagen, Copenhagen, Denmark.

²¹Sigolis AB, Uppsala, Sweden.

²²Universitaetsklinikum Essen, Essen, Germany.

²³Max Planck Institute for Molecular Genetics, Berlin, Germany.

²⁴Lund University, Malmö, Sweden.

²⁵University of Leipzig, Leipzig, Germany.

²⁶Centre National de la Recherche Scientifique, Paris Descartes University, Paris, France.

²⁷Sapienza University of Rome, Rome, Italy.

²⁸Radboud University, Nijmegen, The Netherlands.

²⁹European Institute of Oncology, Milan, Italy.

³⁰Diagenode SA, Liège, Belgium.

³¹The Babraham Institute, Cambridge, UK.

³²Genomatix Software GmbH, Munich, Germany.

³³Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.

³⁴Christian-Albrechts-Universität zu Kiel, Kiel, Germany.

³⁵Cellzome AG, Heidelberg, Germany.

³⁶Institut National de la Santé et de la Recherche Médicale, Marseille, France.

³⁷Weizmann Institute of Science, Rehovot, Israel.

³⁸Barcelona Supercomputing Center, Barcelona, Spain.

³⁹Centro Nacional de Investigaciones Oncológicas, Madrid, Spain.

⁴⁰University Medical Centre Groningen, Groningen, The Netherlands.

⁴¹University of Saarland, Saarbrücken, Germany.

⁴²Oxford Nanopore Technologies Ltd., Oxford, UK.

*e-mail: h.stunnenberg@ncmls.ru.nl

To the Editor:

Last October, scientists gathered in Amsterdam to celebrate the start of BLUEPRINT (<http://www.blueprintepigenome.eu/>), an EU-funded consortium that will generate epigenomic maps of at least 100 different blood cell types. With this initiative, Europe has pledged a substantial contribution to the ultimate goal of the International Human Epigenome Consortium (IHEC) to map 1,000 human epigenomes. Here, we provide a brief background to the scientific questions that prompted the formation of BLUEPRINT, summarize the overall goals of BLUEPRINT and detail the specific areas in which the consortium will focus its initial efforts and resources.

In mammals, nucleated cells share the same genome but have different epigenomes depending on the cell type and many other factors, resulting in an astounding diversity in phenotypic plasticity with respect to morphology and function. This diversity is defined by cell-specific patterns of gene expression, which are controlled through regulatory sites in the genome to which transcription factors bind. In eukaryotes, access to these sites is orchestrated via chromatin, the complex of DNA, RNA and proteins that constitutes the functional platform of the genome. In contrast with DNA, chromatin is not static but highly dynamic, particularly through modifications of histones at nucleosomes and cytosines at the DNA level that together define the epigenome, the epigenetic state of the cell. Advances in new genomics technologies, particularly next-generation sequencing, allow the epigenome to be studied in a holistic fashion, leading to a better understanding of chromatin function and functional annotation of the genome. Yet little is known about how epigenetic characteristics vary between different cell types, in health and disease or among individuals. This lack of a quantitative framework for the dynamics of the epigenome and its determinants is a major hurdle for the translation of epigenetic observations into regulatory models, the identification of associations between epigenotypes and diseases, and the subsequent development of new classes of compounds for disease prevention and treatment. The task, however, is daunting as each of the several hundred cell types in the human body is expected to show specific epigenomic features that are further expected to respond to environmental inputs in time and space. The research community has realized these limitations and the need for concerted action.

The IHEC was founded to coordinate large-scale international efforts toward the goal of a comprehensive human epigenome reference atlas (<http://www.ihec-epigenomes.org/>). The IHEC will coordinate epigenomic mapping and characterization worldwide to avoid redundant research efforts, implement high data quality standards, coordinate data storage, management and analysis, and

provide free access to the epigenomes produced. The maps generated under the umbrella of the IHEC contain detailed information on DNA methylation, histone modification, nucleosome occupancy, and corresponding coding and noncoding RNA expression in different normal and diseased cell types. This will allow integration of different layers of epigenetic information for a wide variety of distinct cell types and thus provide a resource for both basic and applied research.

BLUEPRINT aims to bridge the gap in our current knowledge between individual components of the epigenome and their functional dynamics through state-of-the-art analysis in a defined set of primarily human hematopoietic cells from healthy and diseased individuals. Mammalian blood formation or hematopoiesis is one of the best-studied systems of stem cell biology. Blood formation can be viewed as a hierarchical process, and classically, differentiation is defined to occur along the myeloid and lymphoid lineages. The identity of cellular intermediates and the geometry of branch points are still under intense investigation and therefore provide a paradigm for delineation of fundamental principles of cell fate determination and regulation of proliferation and lifespan, which differ considerably between different types of blood cells.

BLUEPRINT will generate reference epigenomes of at least 50 specific blood cell types and their malignant counterparts and aim to provide high-quality reference epigenomes of primary cells from >60 individuals with detailed genetic and, where appropriate, medical records. To account for and quantify the impact of DNA sequence variation on epigenome differences, BLUEPRINT will work whenever possible on samples of known genetic variation, including samples from the Cambridge BioResource (Cambridge, UK), the International Cancer Genome Consortium and the British Diabetic Twin Study for disease-discordant monozygotic twin samples. The Wellcome Trust Sanger Institute (Hinxton, UK) will also provide full genomic sequencing for up to 100 samples. BLUEPRINT will harness existing proven technologies to generate reference epigenomes, including RNA-Seq for transcriptome analysis, bisulfite sequencing for methylome analysis, DNase-Seq for analysis of hypersensitive sites and ChIP-Seq for analysis of at least six histone marks. Moreover, BLUEPRINT aims to develop new technologies to enhance high-throughput epigenome mapping, particularly when using few cells.

BLUEPRINT is initially focusing on four main areas. One main goal of the project is to comprehensively analyze diverse epigenomic maps and make them available as an integrated BLUEPRINT-IHEC resource to the scientific community. Integration is envisioned for related projects within species (e.g., the 1000 Genomes Project) and between species (e.g., modENCODE) to better understand functional aspects (e.g., shared pathways) and the evolution of cell lineage development. Analysis of the BLUEPRINT data is expected to catalyze a better understanding of the relationship between epigenetic and genomic information and will form the basis for generation of new methods (e.g., epigenetic imputation) for prediction of epigenetic states from epigenomic profiles. Such prediction methods will facilitate a move toward a more quantitative knowledge and modeling of epigenetic mechanisms. As a result, such models could in the future assist in 'reverse engineering' of regulatory networks to repair or restore epigenetic codes that have been perturbed by disease.

A second goal of BLUEPRINT is to systematically link epigenetic variation with phenotypic plasticity in health and disease. This will be attempted in three ways. First, genetic and epigenetic variation in two blood cell types from 100 healthy individuals will be analyzed. These measurements will be combined with whole-genome and transcriptome sequencing to dissect the interplay between common DNA

sequence variation and the epigenome. This will allow estimation, assessed by changes in transcription, of the degree to which epigenetic variation is driven by genetic variation and how this variation affects function. Second, epigenetic profiles will be generated from comparable cell types in three different mouse strains for which high-quality sequence is now available. Determining genotype-epigenotype variation in mouse will allow detailed comparison with human data sets, contribute to generation of experimentally testable hypotheses in a tractable model organism and provide a framework for future comparative epigenomic analyses of other purified mouse cell populations. Third, the possible role of epigenetic variation (that is, DNA methylation) in the etiology of common diseases will be determined by the first comprehensive epigenome-wide association study of any human disease. This is designed to discriminate between epigenetic variation that is a cause or a consequence of disease, a currently unresolved issue in epigenome-wide association studies. Type 1 diabetes has been chosen as an exemplar because certain blood cells (e.g., CD14⁺ monocytes) are relevant to type 1 diabetes pathogenesis and because BLUEPRINT partners have already conducted a successful pilot study demonstrating the involvement of methylation-variable positions, the epigenetic equivalent of single-nucleotide polymorphisms, in type 1 diabetes. Of the >100 methylation-variable positions identified, many have been associated with immune genes and have been found present before disease diagnosis, as well as in patients positive for diabetes-associated autoantibodies but disease-free after 12 years¹. These findings suggest methylation-variable positions to be involved in type 1 diabetes etiology as they cannot be explained by genetic heterogeneity or twinning, metabolic dysfunction, insulin or other pharmacological treatment.

Another goal of BLUEPRINT is to foster the clinical relevance of epigenetic analysis by including a major effort in the biomarker area. The focus will be identifying biomarkers for more accurate prognosis and personalized therapy of childhood acute lymphoblastic leukemia and for determining the efficacy of epigenetic drug treatment in acute myeloid leukemia and myelodysplastic syndrome. Biomarker development will focus specifically on DNA methylation to maximize compatibility with clinical diagnostics and capitalize on the clinical sequencing infrastructure becoming available in European countries².

BLUEPRINT's last goal is to identify new compounds that interact with epigenetic regulators. Epigenetic modifications are reversible and thus have the potential to be modified by small molecule drugs. The first epigenetic targeting drugs (DNA methyltransferase and histone deacetylase (HDAC) inhibitors) have efficacy in some types of cancer but are currently limited by lack of specificity and high toxicity. The full potential of epigenetic-based therapies has not yet been realized owing to limitations of current knowledge on specific targets and imperfect strategies of current drug discovery approaches based on recombinant, single target assays. Although deregulated expression of a few epitargets has been found in cancer, functional results are required to determine which targets qualify as candidates warranting follow-up studies. BLUEPRINT will use RNA interference screens to validate epigenetic targets in ex vivo and in vivo settings. Particularly relevant is the concept of direct in vivo validation of candidate targets, a relatively new approach, on a selected set of candidate drug targets. In vivo RNA interference-based screens in murine leukemia models and patient-derived human samples will be carried out to achieve the most relevant level of preclinical validation possible. Cell assays and screens will be complemented by in vivo screens, used for further validation and, most importantly, allow further characterization of biological phenotypes and assays to be developed for drug discovery. Thus,

the conventional process of sequential validation (from in vitro to in vivo studies) will be improved to provide a more innovative strategy.

BLUEPRINT will also pioneer approaches to drug discovery based on isolating epigenetic complexes directly from human cells or tissues, thus preserving their native multicomponent structure. This is an essential requirement for focused target identification and compound optimization to produce clinically relevant therapeutics. For instance, by using beads modified with HDAC inhibitors to capture interacting proteins from cell lysates and quantitative mass spectrometry to determine drug affinities for those complexes, Cellzome (Heidelberg, Germany) has already identified novel histone deacetylase–protein complexes specific for defined cell states and several new targets for HDAC inhibitors³. By screening existing, non-HDAC inhibitors for binding to HDAC complexes, potential new uses for these drugs as epigenetic modulators can be uncovered, allowing either direct application as epigenetic therapy or as a new starting point for lead optimization on HDAC targets. Additionally, the Cellzome platform has revealed unexpected selectivity of known HDAC clinical-stage inhibitors and lack of activity against specific HDAC protein complexes. We think many of these basic principles can be extended to other epigenetic target classes.

In summary, the commencement of BLUEPRINT signals a new era for European biomedical research that we hope will harness the power of epigenomic dynamics in health and disease. We expect the results will contribute to our knowledge of genome biology and facilitate the definition of new avenues for pharmaceutical intervention, prediction and diagnosis of disease.

Competing financial interests

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturebiotechnology>.

References

1. Rakyan, V.K. et al. PLoS Genet. 9, e1002300 (2011).
2. Callaway, E. Nature 467, 766–767 (2010).
3. Bantscheff, M. et al. Nat. Biotechnol. 29, 255–265 (2011).