BLUEPRINT: mapping human blood cell epigenomes

Joost H.A. Martens, and Hendrik G. Stunnenberg

Radboud University, Department of Molecular Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences, Nijmegen, The Netherlands

E-mail: *j.martens@ncmls.ru.nl* and *h.stunnenberg@ncmls.ru.nl* doi:10.3324/haematol.2013.094243

The sequencing of the human genome, now well over a decade ago,[1] provided important new insights into the variation of sequence elements present in the genome. An immediate question that surfaced was how one single genome can give rise to all the different cell types of the human body, and how all these different cell types know how and when to interpret these genomic elements. Since then, it has become widely appreciated that this regulation of DNA elements is closely connected to how the DNA is packaged within the nucleus of each individual cell. This packaging is orchestrated via chromatin, the complex of DNA, RNA and proteins that provides functionality to the genome. Chromatin is, in contrast to DNA, highly dynamic, and this plasticity is achieved through placing chemical tags on cytosines at the DNA level and on histones at the nucleosome level providing an index or blueprint on top of the genomic information in different cellular environments (Figure 1).[2,3] The field of epigenetics aims to define the combination of chemical tags that index the genome, how these marks are recognized and how this translates into specific gene regulation and interpretation of the functionality of DNA elements. Initially, epigenetic studies focused on single loci or genomic regions of several kilobases, but technological advances in recent years have for the first time allowed the establishment of genome wide epigenomic maps of individual cell types. To face the challenge of the plasticity and the shear number and diversity of epigenetic marks, worldwide consortia have been founded which aim to establish epigenomic maps of different kinds for a variety of species. Specific for human, these efforts are coordinated by the International Human Epigenome Consortium (IHEC; *http://www.ihec-epigenomes.org/*) that aims to deliver 1,000 reference epigenomes for human cell types. Within IHEC, quality standards are set for every aspect of epigenomic mapping, from experimental setup and metadata collection to data storage and primary analysis. Moreover, it provides communication platforms that allow data sharing and minimizes redundancy between different individual research projects. Currently, seven epigenomic projects have been initiated and more are expected to join IHEC in the coming years.

BLUEPRINT (*www.blueprint-epigenome.eu*) represents the European initiative to establish epigenomic maps. BLUEPRINT[4] has chosen to focus on generating epigenomic maps of a wide variety of cell types from the blood, and to provide at least 100 reference epigenomes as a resource to the scientific community. These reference epigenomes will include primary human cells from healthy individuals but also blood-based diseases. The reference epigenomes from healthy individuals will provide an important framework for understanding normal hematopoiesis as well as providing a reference point for epigenomes generated for hematologic malignancies such as autoimmune diseases and common leukemias of myeloid and lymphoid origin. Of the latter, BLUEPRINT will include several of the most prevalent and clinical B-cell malignancies in adults and children, i.e. chronic lymphocytic leukemia, follicular lymphoma, diffuse large B-cell lymphoma and Burkitt's lymphoma that are already subjected to complete genomic and RNA sequencing in two European ICGC (*http://icgc.org*) projects,[5] thereby further facilitating genome *versus* epigenome comparisons.

The BLUEPRINT project, which was initiated in the fall of 2011, celebrated its first data release last April when 12 full epigenomes of neutrophils and monocytes from both adult blood and cord blood became available through different data portals including the ENSEMBL[6] and UCSC[7] browsers, BIO-MART[8] and a visual interface developed by GENOMATIX.[9] For each sample, the release included information on 6 IHEC recommended informative histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K9me3 and H3K27me3) (Figure 2), on DNA methylation at base pair resolution, on RNA expression, and on genome-wide accessibility through DNAseI-seq analysis. Future releases will also include epigenomic data on differentiation pathways, such as monocyte to macrophage and B-cell differentiation, and from more rare cell types from healthy donors as well as on diseased cell types.

As an epigenomic project specifically focusing on hematopoiesis, BLUEPRINT is expected to make a major contribution to the field of blood epigenetics. The epigenomic maps generated within BLUEPRINT will provide comprehensive indexes of chromatin organization and associated functionality that will serve as an entry point for further investigations into the key transcription factors and regulatory networks that establish, regulate or maintain epigenomic features.[10] Within one cell type, the epigenomic maps will allow the identification of gene classes with similar patterns of epigenetic fea-

Figure 1. Schematic overview of epigenetic modifications and effects on chromatine structure and accessability of genes.

A

phosphorylation

methylation
repressive lysine

H2B | H2A

H4 | H3

K36me3

methylation
active lysine

K27me3

K4me1
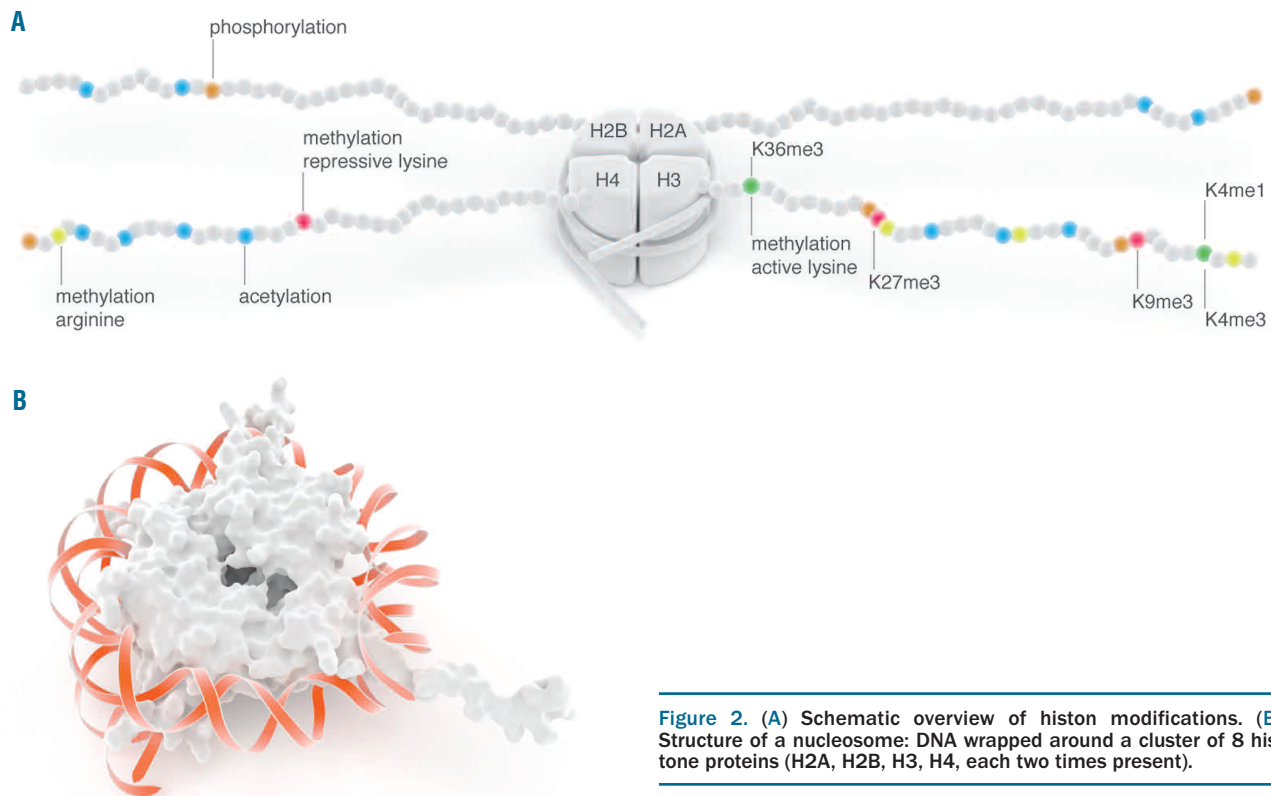
K4me3

K9me3

methylation
arginine

acetylation

B

Figure 2. (A) Schematic overview of histon modifications. (B) Structure of a nucleosome: DNA wrapped around a cluster of 8 histone proteins (H2A, H2B, H3, H4, each two times present).

tures that likely represent clusters with coordinated regulation of gene expression. Upon systematic comparisons between cell types, clusters of genes may prove to be coordinately regulated by epigenetic mechanisms throughout the hematopoietic differentiation program.[11,12] In addition, chromatin state maps identification will allow assignment of the functional states (such as active, inactive or poised) based on epigenetic profiles. The analysis will divide the genome into epigenetic segments comprising combinations of different epigenetic features such as DNA methylation, histone modifications and accessibility[13] and relate these to function. The identification of such segments and comparison with perturbed epigenetic landscapes in disease will trigger investigations into the restoration or repair of the epigenetic code of these elements.

Another goal of BLUEPRINT will be to investigate epigenetic variation between individuals by studying two cell types from at least 200 healthy donors, while a third cell type from the same donors will be analyzed by a Canadian IHEC project (*http://ihec-epigenomes.org/research/projects/epigenomic-platform-program/*). These donors will be subjected to complete genome sequencing by the Wellcome Trust Sanger Institute, as well as part of the UK10K project (*http://www.uk10k.org/*) allowing epigenetic and genetic variation to be correlated. As the epigenome is expected to be more plastic and influenced by many environmental factors, including diet, age and environmental exposure,[14] a likely finding will be that the epigenomes are more variable between individuals then their genomes. Moreover, it will reveal the natural epigenetic variation between cell types from different individuals and to what extent this variation is influenced by variations in the genome sequence. In addi-

tion, it will be extremely interesting to investigate the extent to which epigenetic variation is translated into differential gene expression.

BLUEPRINT will also set out to predict consequences of sequence variants such as those linked to disease and thereby create a better understanding of the relationship between genetics and epigenetics. For example, since most SNPs identified in genome-wide association studies are located outside of protein coding genes,[15] epigenomic mapping allows the functional characterization of a regulatory region, for example by identifying new or alternative promoters, poised or active enhancers. Moreover, the DNAseI-seq experiment can be extended to footprint analysis[16] that will allow the putative identification of transcription factor binding to a particular genomic location and can, in relation with GWAS studies, also predict whether binding of transcription factor(s) will be disturbed or gained. Apart from the sequence variation in regulatory regions, BLUEPRINT will also investigate the genome-wide effects of epigenetic gene mutation. Many of the disease epigenomes will be determined in leukemias that, as in many other cancers, contain defects in main epigenetic regulators such as MLL (H3K4me3), DNMT3A (DNAme) and TET2 (DNA demethylation) in subtypes of acute myeloid leukemia.[17] Comparisons between these and reference epigenomes form healthy donors is expected to further unravel how disturbance of epigenetic mechanisms facilitate the acquisition of the hallmarks of cancer.[18,19] As, in contrast to genetic changes, these epigenetic modifications are reversible, BLUEPRINT has also initiated studies that focus on epigenetic drug treatment. These include efforts to identify compounds that modulate epigenetic regulators as well as to

identify novel epigenetic targets through RNA interference approaches.

As far as clinical impact is concerned, BLUEPRINT will nurture the identification of biomarkers, specifically focusing on the discovery of novel epigenetic biomarkers. Epigenomes hold great promise as discovery engines for identification of a disease state or environmental influence (e.g. exposure to toxins, infections, drugs of abuse or psychosocial stress) and thus will be useful for diagnosis, prognosis or monitoring disease progression and remission. Within BLUEPRINT, biomarker development will focus on analyzing DNA methylation in childhood acute lymphoblastic leukemia both for prognosis and personalized therapy, and on myeloid diseases, specifically acute myeloid leukemia and myelodysplastic syndrome, for determining the efficacy of epigenetic drug treatment.

Finally, since BLUEPRINT represents one of the founding human epigenomic projects within IHEC, it has developed together with other IHEC consortia and ENCODE (http://genome.ucsc.edu/ENCODE/) standards for other epigenetic projects to produce, analyze, and integrate large datasets and will maintain this function. A main challenge within BLUEPRINT, as also in other epigenomic projects, is that the full spectrum of hematopoietic cell types from the stem cell to fully differentiated cells entails analysis of abundant cell types such as monocytes and neutrophils, as well as more rare or ultra rare cell types such as hematopoietic stem cells. Primary hematopoietic stem cells and progenitors cannot be obtained in sufficient quantities to perform full epigenomic mapping using existing methods. Therefore, BLUEPRINT is investing in developing new miniaturization technologies to allow high throughput mapping of the more rare blood cell types. As such, based on ongoing efforts and future projects, BLUEPRINT will continue to provide important guidance and guidelines for further epigenomic mapping. These will not be restricted to experimental issues, but will also be extended to data analysis, especially focusing on large-scale epigenomic data integration that remains one of the major challenges within these multidimensional projects.

BLUEPRINT, as the first so-called high impact EU research initiative project, can be expected to make significant contributions to our understanding of the hematopoietic epigenome, both in health and disease. We anticipate that these will ultimately facilitate the development of new predictors of disease, as well as providing new entry points for diagnosis and drug therapies.

Dr Joost Martens is a research group leader at the Radboud University, Nijmegen, the Netherlands. He is mainly interested in perturbed transcription factors and epigenomics in AML. Dr. Hendrik Stunnenberg is Head of the Molecular Biology Department at the Radboud University, Nijmegen, the Netherlands. His main interest is epigenetic and transcriptional regulation in ES and blood cells and is coordinating the BLUEPRINT project (BLUEPRINT – A BLUEPRINT of Haematopoietic Epigenomes).

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
2. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell. 2007;128(4):669-81.
3. Maunakea AK, Chepelev I, Zhao K. Epigenome mapping in normal and disease States. Circ Res. 2010;107(3):327-39.
4. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012;30(3):224-6.
5. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012;44(11):1236-42.
6. http://www.ensembl.org/Homo_sapiens/Location/View?g=ENSG00000130544;r=19:7069455-7087979;time=1372059055
7. http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=ftp://ftp.ebi.ac.uk/pub/Databases/blueprint/releases/20130301/homo_sapiens/hub/hub.txt
8. http://blueprint.bsc.es/
9. http://blueprint.genomatix.de/
10. Martens JH, Stunnenberg HG, Logie C. The decade of the epigenomes? Genes Cancer. 2011;2(6):680-7.
11. Satterlee JS, Schubeler D, Ng HH. Tackling the epigenome: challenges and opportunities for collaboration. Nat Biotechnol. 2010;28(10):1039-44.
12. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell. 2010;6(5):479-91.
13. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol. 2010;28(8):817-25.
14. Suter MA, Aagaard-Tillery KM. Environmental influences on epigenetic profiles. Semin Reprod Med. 2009;27(5):380-90.
15. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
16. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature. 2012;489(7414):83-90.
17. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer. 2012;12(9):599-612.
18. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
19. Shen H, Laird PW. Interplay between the Cancer Genome and Epigenome. Cell. 2013;153(1):38-55.