

 Contenido archivado el 2024-06-18



A BLUEPRINT of Haematopoietic Epigenomes

Informe

Información del proyecto

BLUEPRINT

Identificador del acuerdo de subvención:
282510

[Sitio web del proyecto](#) 

Proyecto cerrado


Fecha de inicio
1 Octubre 2011

Fecha de finalización
30 Septiembre 2016

Financiado con arreglo a
Specific Programme "Cooperation": Health

Coste total
€ 39 664 716,88

Aportación de la UE
€ 29 996 664,00

Coordinado por
STICHTING RADBOUD
UNIVERSITEIT
 Netherlands

Final Report Summary - BLUEPRINT (A BLUEPRINT of Haematopoietic Epigenomes)

Executive Summary:




BLUEPRINT was a large-scale research project receiving close to 30-million-euro funding from the EU. 42 leading European universities, research institutes and industry entrepreneurs participated in what is one of the two first so-called high impact research initiatives to receive funding from the EU.

The BLUEPRINT consortium was formed with the aim to further our understanding of how genes are activated or repressed in both healthy and diseased human cells. The project focused on distinct types of haematopoietic cells from healthy individuals and diseased counterparts to advance and exploit


knowledge of the underlying biological processes and mechanisms in health to ultimately treat disease.


Reference epigenomes have been generated by state-of-the-art technologies in accordance with quality standards set by the International Human Epigenome Consortium (IHEC) of which BLUEPRINT was a cornerstone consortium. BLUEPRINT has developed protocols for the analysis of single or very low number of cells, which enables the analysis of rare to ultra-rare cell types and, importantly, of clinically relevant material such as biopsies. Thus, BLUEPRINT has generated around 3500 data sets of healthy and diseased cell types.

BLUEPRINT also focused on identifying genetic variants responsible for changes in gene expression levels and epigenetic changes at the level of the entire human genome of three blood cell types from 200 healthy donors, in collaboration with its Canadian IHEC partners.

The very considerable amount of data generated required a coordinated approach for data storage, handling and distribution as well as the implementation of a common data analysis strategy. All quality controlled data was made publicly accessible through links to the BLUEPRINT (<http://dcc.blueprint-epigenome.eu/#/home> ) DeepBlue (<http://deepblue.mpi-inf.mpg.de> ) and the IHEC (<http://epigenomesportal.ca/ihec/> ) data portals and can be visualized in several genome browsers. The data are frequently used by non-BLUEPRINT-members.

The resource-generating activities have been complemented by hypothesis-driven research into blood-based diseases such as autoimmune disease Type 1 Diabetes Mellitus (T1DM). BLUEPRINT's research led to several game changing studies on common leukaemias and to the discovery and validation of epigenetic markers for diagnostic use and for epigenetic target identification. Since epigenetic changes are reversible, they can be targets for the development of novel and more individualised medical treatments. BLUEPRINT uncovered epigenetic memory and metabolic programs of innate immune cells which led to a paradigm shift in immunology. The tolerant and trained macrophage phenotypes are laid down in their epigenomes; the exploration of this invaluable resource opens new doors for the treatment of immune diseases.

The involvement of innovative companies has energized epigenomic research in the private sector by the development of smart technologies which led to the establishment of a spin-off SME Cambridge Epigenetix (<http://www.cambridge-epigenetix.com> ) and three patent applications, one of which is to predict the clinical evolution of patients suffering of chronic lymphocytic leukemia (CLL).

During the 5-year project duration, the BLUEPRINT consortium has published over 280 publications, with data still to be analysed and to be published. Importantly, BLUEPRINT initiated and coordinated the submission of a package of publications collected from the IHEC consortia, which appeared on November 17th, 2016 in Cell and in other journals (<http://www.cell.com/consortium/IHEC> ). In this issue, the importance of and opportunities provided by epigenome research were discussed in an IHEC essay and aspects of data sharing and privacy protection was discussed in a commentary. The coordinated release of this set of papers significantly raised the attention for epigenetic research in the scientific literature, in the general media, in biotech and in the clinic.

In summary, the outcome of the BLUEPRINT project has exceeded all the set goals and the insights from the BLUEPRINT project provide hitherto unexplored opportunities for targeted (epigenetic) diagnostics, new treatments and preventive measures for specific diseases in individual patients, known as 'personalised medicine'.

Project Context and Objectives:

The research activities of the BLUEPRINT project were divided into 4 Research Areas and 15 Work Packages (WPs). Research Area 5, covering 4 Work Packages, was set up for training, networking and outreach activities.

Research Area 1 – Human reference epigenomes and Data coordination and analysis

WP1 (Standardization and Quality Control of Reagents) aimed at defining Standard Operating Procedures (SOP) for reference epigenome generation using Chromatin Immunoprecipitation coupled to Next Generation Sequencing (ChIP-seq) to ensure consistent data of the highest quality and to define performance benchmarks. In short, in ChIP-Seq, cross-linked DNA-protein complexes are precipitated with an antibody that is specific to the DNA-binding protein and the DNA is sequenced. The quality of the antibody is critical and determines the level of enrichment over background. BLUEPRINT partner DIAGENODE ensured that partner labs within BLUEPRINT use only validated antibodies and established standard protocols to ensure that each participant will generate ChIP-seq data of the highest quality and reproducibility.

WP2 (Standardization and Quality Control of Sequencing) aimed at the production and standardization of high quality sequencing data for whole-genome sequencing in particular for bisulfite treated genomic DNA (BS-Seq). BLUEPRINT, together with its IHEC partners, set out to establish standard operating procedures (SOP), protocols for data exchange, data access, shared analysis methods and pipelines, and development of a common ground for analysis of results and standards for the biological and interpretation of sequence data.

WP3 (Epigenome of haematopoietic cells) aimed at generating a resource to study the regulation of the differentiation of haematopoietic stem cell (HSC), via committed precursor cells and mature blood cells from cord and adult blood provided by Cambridge BioResource. To achieve the necessary high purity and great homogeneity, standardized cell selection protocols had to be developed. Another major objective was to generate (partial) epigenomes from monocyte and neutrophilic granulocyte of 200 subjects (for WP10). At later stages, new objectives were taken up in the portfolio. First, the development/establishment of novel single and low cell approaches (μ WGBS for DNA methylation and iChIP for histone marks) to facilitate the analysis of rare and ultra-rare cell types obtained from adult blood and bone marrow of healthy individuals. Second, the transcriptome of the entire B-cell differentiation program by MARS-seq.

WP4 (Epigenome of normal and neoplastic B- and T-cells) aimed at generating epigenomes of B-cell malignancies that account for approximately two thirds of all hematologic neoplasms and around 80% of the lymphoid tumors with increasing incidence. B-cell malignancies are biologically heterogeneous and the outcome strongly depends on the biological subtype defined by the underlying genetic aberrations and the maturation stage. BLUEPRINT teamed up with two European ICGC projects (Spain, Germany) to

complement ongoing whole genome and transcriptome sequencing activities by generating epigenome data on the same cancer samples and on normal B-cells. The second focus of WP4 was on T-lineage acute lymphoblastic leukaemia (T-ALL) that consists of a heterogeneous group of acute leukaemias that are arrested at various stages of T cell development. Notably, oncogenic markers for therapeutic stratification of T-ALLs are not yet available, stressing the needs for new biomarker development. The second aim in leukemic T-cell malignancies was to generate and exploit epigenomes of T-cell prolymphocytic leukaemia (T-PLL) that consists of more mature T-cell leukaemia.

WP5 (Epigenome of acute myeloid leukaemias) aimed at generating and exploiting reference epigenomes of acute myeloid leukaemias (AML) which is the most common acute leukaemia affecting adults, accounting for approximately 41% of leukemic deaths in Europe. Its incidence is expected to increase as the population ages. Although several risk factors for AML have been identified, the specific cause of AML remains unclear. As an acute leukaemia, AML progresses rapidly and is typically fatal within weeks or months if untreated. The epigenome resource was expected to reveal the AML specific epigenomic features and establish the variety within this subclass of leukaemia's: adult AMLs associated with specific chromosomal translocations, with normal karyotype, with monosomal karyotype and patients with MyeloDysplastic Syndrome (MDS).

WP6 (Data Coordination Centre – DCC) focused on development and implementation of data storage, data handling and data distribution system. This includes the primary (low-level) data analysis methods including the alignment of sequence reads to the appropriate reference genome and an initial calling of positive regions. The resulting uniformly processed data will be made available to the BLUEPRINT Data Analysis Group (DAG) and the wider scientific community. The aim was to set up the proposed framework in line with the developments in other large scale genome-based consortiums (especially ICGC and IHEC) and take into account the needs created by large data sets, intensive computation and distributed analysis and provide the wider community with tools to analyse these large and highly dimensional data sets.

WP7 (Data Analysis Group - DAG) aimed at the design and implementation of common “data analysis strategies”, by analyzing and evaluating current methodologies, by developing new methods and by proposing and coordinating the data analysis strategy. This WP will coordinate the participation of biologists and bioinformaticians of all the project partners in the DAG, thus avoiding unnecessary duplications and fostering collaboration. The DAG is ultimately responsible of generating the final project integrated analysis.

Research Area 2 – Causes and consequences of epigenetic variation and utilize this knowledge for improved diagnostics of human diseases

WP8 (DNA methylation variation in T1DM) aimed at providing a resource to study the epigenetic deregulation in a non-cancer common disease, Type 1 Diabetes Mellitus (T1DM). This chronic autoimmune disease results from an interaction between genes, many involved with the immune response, and environmental factors. Such interaction can induce a destructive inflammatory response which leads to clinical disease. BLUEPRINT hypothesized that the gene-environment interaction will be mediated, in whole or in part, by epigenetic effects including DNA methylation. BLUEPRINT focused on generating DNA methylation profiles of Monozygotic (MZ) twins discordant for the disease (thereby eliminating

genetic variability) with the major aim to identify disease-associated methylation variable positions and to determine their causality in the etiology of T1DM.

WP9 (Biomarker development) aimed to determine the clinical value of epigenetic biomarkers for hematopoietic malignancies. Cancers of the blood system are promising targets for epigenetic biomarker development. To establish epigenetic biomarkers as a key component in the toolkit of personalized medicine, WP9 aimed to define targets drawing from hundreds of reference epigenomes that were to be generated in BLUEPRINT and other consortia like TCGA and ICGC. Partners in WP9 aimed at performing benchmarking of new and established assays for DNA methylation biomarkers to ensure robustness and reproducibility.

WP10 (The effect of common sequence variation on the epigenomic landscape) aimed at identifying and validating quantitative differences in the epigenome marks and the RNA landscape that are associated with common sequence variation. The relationship between common sequence variation and the risk of common diseases on the one hand and disease-relevant quantitative traits (QTs) on the other by Genome-Wide Association Studies (GWAS) has been a resounding success. BLUEPRINT aimed at bridging the large gap between the levels of observed and explained heritability. The primary motivation to test the association between common sequence variation and variation in epigenetic marks and the RNA landscape is to determine what fraction of the latter is due to underlying DNA sequence variation and what fraction is epigenetically determined. To minimize “experimental noise” and therefore enhance the power to discover modest effects, BLUEPRINT set out to generate the epigenomes of three peripheral blood cell types, the monocyte (an important central orchestrator of adaptive immunity and a bridge between innate and adaptive immunity), the neutrophilic granulocyte (the frontline cell for innate immunity) and CD4 T naïve cells from 100 females and 100 males in collaboration with the IHEC partner from McGill.

WP11 (Mouse models to quantify variation in reference epigenomes). It is well known that the epigenetic state is dynamic with the potential to change in differentiation, in response to external cues (such as pathogens), in aging or as the cell acquires a diseased state. The comparative differences in epigenetic state in different biological contexts are key to understanding the regulatory functions of epigenetic modifications in health and disease. BLUEPRINT aimed at quantifying genotype-epigenotype variation in two haematopoietic cell types and in inbred mouse strains (thereby eliminating genetic differences) to determine to what extent variation between epigenomes is a reflection of random non-functional epigenetic events or of the underlying sequence, and if so, to determine the functional and non-functional consequences.

Research Area 3 – Development and validation of novel technology for high-throughput epigenome mapping

WP12 (Technology development for profiling of cytosine (hydroxy)methylation) aimed at developing and applying novel strategies and methods to study DNA methylation, a key component of vertebrate epigenomes. The methods available at the start of the project had specific limitations in resolution and throughput. Furthermore, the discovery of hydroxymethylation (5hmC) added a further layer of epigenetic information, and posed new challenges for its specific detection. Since bisulfite cannot distinguish methylation from hydroxymethylation, one important aim was to develop new chemistry or sequencing

approaches to sequence 5mC and 5hmC at single nucleotide resolution. A second major goal was to develop approaches to obtain single nucleotide methylation maps of a fraction of the genome applicable in high throughput projects (such as large cohorts). The third aim was to generate methylomes for DNA that is bound by specific factors by combining Chromatin-IP with bisulfite sequencing to relate specific methylation of 5mC and/or 5hmC to protein binding and histone modification states, a method needed to functionally understand, amongst others, allelic heterogeneity in DNA methylation. Lastly, the development of bisulfite-based approaches to determine DNA methylation at single base resolution using very low DNA input of DNA from single cells was added to the portfolio to facilitate analysis of (ultra-)rare cell types and of biopsies.

WP13 (Technology optimization for microscale application) aimed at developing pipelines for ChIP-sequencing starting from low amounts of chromatin to facilitate the study of rare and ultra-rare populations of cells. Following an open call for partners, BLUEPRINT decided to focus on two aspects. Firstly, the use of microfluidics technology and the development of a BLUEPRINT custom designed microfluidic chip with the company Fluidigm, the world-leader in the field. Secondly, to develop a prototype microsonicator (by partner Diagenode) for fragmentation of chromatin from very low cell numbers, as the input for the microfluidic chip.

Research Area 4 – Identification of new compounds interfering with the regulators of epigenetics profiles

WP14 (Identification and validation of Epi-targets) aimed at to characterize known and identify novel chromatin-associated proteins with a role in the establishment and/or the maintenance of the cancer phenotype and to elucidate their function in relevant model systems and to push the enzymes and factors forward for drug development in WP15. Epigenetic changes are causally linked to oncogenesis and tumor progression. Epigenetic changes are prevalent in all types of cancer, regardless of their histological origin, are observed at all stages of tumor development and induction of epigenetic changes in animal models is sufficient to initiate tumorigenesis. The finding that epigenetic proteins are often aberrantly expressed in cancer makes them attractive targets for drug treatment and hence the prospect of an epigenetic treatment of cancer is high. The best-studied functional consequences of epigenomic changes are those resulting from aberrant DNA-methylation and histone deacetylases which boosted the development of epigenetic drugs for these enzymes. However, they constitute a very small fraction of the 600-700 proteins involved in chromatin organization. New targets and new drugs were and are needed to treat cancer.

WP15 (Compound development and screening) aimed at applying a ‘targeted’ drug design approach using both in vitro and in cell-based assays to exploit the structure-based compound deconvolution and to perform biomedical validation in mice and preclinical approaches. In WP15, the chemical proteomics approach called Episphere, developed by the SME Cellzome, was to be exploited to characterize existing and novel compounds and to characterize uncovered epigenetic complexes. Episphere is a unique approach to measure potency and selectivity across a wide range of epigenetic targets in their native state.

Research Area 5 – Training, networking, communication and outreach

WP16 (Training, Networking and Communication) aimed at organizing internal and external training

programs. This included setting up and funding a lab exchange program for young investigators as well as the organization of topical workshops, training courses as well as symposia and satellite workshops: these activities were often organized jointly with other consortia and most of the time they were open to non-members. A second important objective was the establishment of an infrastructure (website) for internal communication, for exchange of information on technology, materials, data, protocols, and antibodies.

WP17 (Dissemination and Outreach) aimed at coordinating the contacts with other consortia active in the area of epigenome research, like EU project EpiGeneSys and IDEAL as well as the International Human Epigenome Consortium (IHEC). In collaboration with these consortia, WP17 aimed at joint efforts to impact on (the perception of) epigenomics on health research and health programs. Moreover, WP17 had the goal of producing newsletters and short films. Also, the public website was a task of this WP. And lastly, WP17 aimed at developing a visual interface for public data dissemination (by Genomatix), to enable scientists that are not experts in working with epigenome data, can search the data generated by BLUEPRINT.

WP18 (Project Management) aimed at setting up an effective management structure to ensure that the overall coordination of this unique High Impact project would run smoothly. This was absolutely imperative given the 42 partner organizations and some 55 Principal Investigators involved in the BLUEPRINT project and the ambitious multidisciplinary objectives.

WP19 (Key Performance Indicators) aimed at providing an annual overview of the major outputs of the consortium by means of key performance indicators.

Project Results:

WP1 – Standardization and quality control of reagents for ChIP

WP1 aimed at defining the Standard Operating Procedures (SOP) for reference epigenome generation using the ChIP-seq technique to ensure consistent data of highest quality and to define and control performance benchmarks for critical steps of data generation. In the ChIP process, the quality of the antibody is one of the critical components for an effective ChIP experiments as it determines the level of enrichment over background. The SME-partner Diagenode has produced and validated different sets of ChIP-seq grade high quality antibodies. Extensive Quality Control (QC) metrics have been developed for 'BLUEPRINT' grade antibodies, which includes Elisa, Western blot, peptide array, chromatin immunoprecipitation (ChIP) on selected genomic locations and ChIP-sequencing. As decided by the International Human Epigenome Consortium (IHEC), the core histone modifications to be profiled for a full epigenome are H3K4me3, H3K4me1, H3K9me3, H3K27me3, H3K27ac and H3K36me3. In addition, antibodies and protocols were developed for H3K9/14ac, H2AZac, H2AZ, H4K20me3 and H3K79me3. The performance of the antibodies in ChIP-seq was validated by several BLUEPRINT groups and in part by Cellzome using Mass Spectrometry. All antibodies were successfully generated, validated and provided to BLUEPRINT partners to guarantee high quality ChIP-seq data for primary hematopoietic cell types. The validated antibodies were made commercially available to the wider community. Protocols and QC metrics have been published on the BLUEPRINT and IHEC websites.

Another step determining reproducibility, consistency and efficiency is the sonication. This aspect became critical as BLUEPRINT moved from abundant cell types for which 'bulk' amount of chromatin can be obtained to rare or ultra-rare cell types. Existing equipment and procedures could not be used and major efforts were undertaken to miniaturize the sonicator and the sonication volume to volumes below 10

microliters. By introducing ultrasound waves generated by a piezo element, and a proof-of-principle of fragmentation by the newly integrated EMBL team (see WP13), Diagenode developed a substantially upgraded prototype that was distributed and tested by several BLUEPRINT partners. The SME-partner Diagenode filed an US patent for this prototype (Method and apparatus for fragmenting DNA sequences, US 20120264228 A1).

WP2 – Standardization and Quality control of sequencing

The central promise of the Blueprint consortium was to provide at least 100 high-quality full epigenomes from the different cell types of the hematopoietic cell system to the International Human Epigenome Consortium Project and the research community. Within the BLUEPRINT project, the role of WP2 was to generate these high-quality epigenomes. BLUEPRINT members made a major contribution to the IHEC Assay Standards group that defined the qualities of the data that needed to be provided. According to the consensus of IHEC, a full epigenome consists of the six core histone marks (H3K4me3, H3K36me3, H3K27ac, H3K4me1, H3K27me3, H3K9me3), RNA sequencing and Whole Genome Bisulphite Sequencing with a minimum of 30x coverage. At the outset of the project, all laboratory procedures were optimized and turned into standard operating procedures. Dedicated computational pipelines were developed in WP6/7 for the processing of the data and all data transferred to EpiRR through an efficient data transfer protocol that was established in BLUEPRINT.

BLUEPRINT successfully established an efficient process chain for handling cell types that started with the chromatin immunoprecipitation, the most variable step that initially required large amount of input material. Once the ChIP-seq for the six core histone marks was successfully completed and passed the QC, RNA sequencing (strand-specific, paired-end 100 cycles per end and between 80-100 million reads) and the WGBS (minimally 30x coverage) were launched so that a maximum number of full epigenomes could be completed. Purification of homogeneous cell types, ChIP-seq, RNA-seq and WGBS-seq were generated by different partners in different countries. Logistically this required the movement of cell materials from the University of Cambridge and clinical partners to Radboud University; upon its successful completion, RNA was shipped to the MPIMG for RNA sequencing and DNA was shipped to the CNAG for WGBS. For the needs of BLUEPRINT additional histone marks and DNA accessibility assays, initially DNaseI hypersensitivity and as it became available during the course of the project the Assay for Transposase Accessible Chromatin (ATAC)-seq, were produced for a more comprehensive description of certain of the cell types.

WGBS standard reporting format: WGBS data can provide information on sequence variants as well as methylation status, but there was no standard way of reporting this information. BLUEPRINT partner CNAG therefore developed an extension to the VCF4.2 format, developed for reporting sequence variants, that allows methylation specific information and quality measures to be combined with sequence variant calls in a standardized manner. This format has arisen from discussions in the IHEC Assay Standards group and has been refined in consultation with members of the TCGA and Roadmap projects, and in particular with the developers of the Bis-SNP. An important feature of the new format is that it is still a valid VCF file, and so existing packages for sorting, filtering, indexing, compressing and viewing VCF files (i.e. vcftools and bcftools) will work with these files. The extended VCF format for whole genome bisulfite sequencing has been adopted by IHEC and has received strong support from TCGA and ICGC projects to adopt it as the standard WGBS reporting form.

WP3 – Epigenome of haematopoietic cells

This WP has produced the most exhaustive to date reference atlas of epigenomic and expression data for essentially 'all' blood cell types. At the start of BLUEPRINT, the technology of transcriptome, methylome and ChIP-seq using very low amounts of cells (very rare cell types like stem cells and progenitor cells) was in its infancy.

Hence in the first 2-3 years of the project, the focus was on generating and exploiting epigenomes of the more abundant, mature cell types. This culminated in the joint publication of the first fruits from the integrated epigenome profiling efforts that included deep bioinformatics mining of the epigenomes (together with partners in WP6 and WP7). Three major publications appeared in Science (September 2014) describing epigenetic and transcriptome profiling. Two manuscripts were based on studies performed in collaboration with Associate Member prof. Mihai Netea (Nijmegen, The Netherlands) assessing tolerance and training in innate immunity. Epigenetic programming during monocyte to macrophage differentiation represents one of the cornerstone processes in innate host defense. Immunological imprinting - either 'tolerance' or 'trained immunity' - after an infection or vaccination determines the functional fate of monocytes and macrophages, and the susceptibility of the host to secondary infections. Saeed et al. described the transcriptional and epigenetic changes taking place when monocytes were exposed to LPS and beta-Glucan. This highly cited study showed that tolerance and training are written in the epigenome and maintained during the differentiation of the monocytes to macrophages providing evidence for epigenetic memory. Cheng et al. described that the metabolism changes entirely upon differentiation of monocytes to macrophages. This highly cited study indicated that induction of aerobic glycolysis through an Akt-mTOR-HIF-1 α pathway represents the metabolic basis of trained immunity. These two connected studies caused a paradigm shift. Up to this point, innate immune cells, in contrast to the lymphoid cells, were thought not to have a memory. That the finding has important implications for modifying the immune system to fight disease was highlighted in a Science editorial. In the third Science publication, Chen et al. analysed paired-end RNA-seq analysis of six progenitor cell types showing extensive alternative splicing of factors involved in lineage determination. This study revealed amongst others the extensive usage of alternative transcriptional start sites, exons or 3' untranslated regions that was unknown before. This will allow updating the reference databases and together with the characterization of the regulatory space of each cell type will allow scientists studying blood and immune diseases, both common and rare, to have the most complete reference set to identify the cause of diseases. Nature Reviews Genetics and Nature Reviews Immunology published news and views about these publications.

In the course of the project, low cell and single cell technology were developed by BLUEPRINT partners and others and could be integrated into BLUEPRINT's main goal of generating epigenome as a resource to the community. Methods for DNA methylation of low and single cells were developed within BLUEPRINT (see WP12) as well as single cell RNA-seq approaches. Bottle-necks in ChIP-seq in tackling rare and ultra-rare cell types could be overcome by integration of ChIPmentation (Schmidl et al. 2015), a low cell protocol developed by BLUEPRINT partner Bock and integration of Ido Amit, who published a low cells protocols using chromatin indexing (Lara-Astiaso et al. 2014) and co-ChIP approaches (Weiner et al. 2016), as a partner in BLUEPRINT to study rare cell types.

Exploitation of ultralow and single cell bisulfite and RNA sequencing, epigenomes from the rare cell types, Hematopoietic Stem Cell (HSC), Multipotent Progenitor (MPP), Lymphoid-primed Multipotential Progenitors (LMPP), Common Lymphoid Progenitor (CLP), Early T-cell precursor (ETP), Progenitor for B and NK (BNKP), Common Myeloid Progenitor (CMP), Megakaryocytic and Erythroid Progenitor (MEP),

Granulocyte-Macrophage Progenitors (GMP) and several stages of megakaryocytes (MK) cells has been published by Farlik et al. (2016) in a close collaboration between the Cambridge (Frontini) and Vienna (Bock) laboratories. Furthermore, three biological replicates of 30 cell subpopulations (including ultra-rare subpopulations) covering the entire B-cell differentiation program from hematopoietic stem cells to terminally-differentiated plasma cells have been analysed at the level of the epigenome (iChIP-seq) and RNA sequencing (MARS-seq). This study is due to be submitted for publication beginning of 2017.

The Nijmegen laboratory, together with many BLUEPRINT partners, published in Novakovic et al. (Cell 2016) a highly detailed analysis of the epigenomic and transcriptomic analysis as a follow-up of the Science publications of 2014. This study showed that tolerance of macrophages induced by LPS, a constituent of bacterial cell wall, can be reverted by treatment of tolerized macrophages with the 'training' compound beta-Glucan. The treatment was shown to be effective in vitro cultured macrophages as well as tolerance induced in vivo in volunteers in the endotoxemia model. These results may have important implications for treatment of sepsis patients for which there is currently no effective treatment.

Javierre et al. (Cell 2016) generated high-resolution maps of promoter interactions in 17 human primary blood cell types. They showed that the interaction patterns are cell type specific and segregate with the hematopoietic tree. Importantly, promoter-interacting regions are enriched for regulatory chromatin features and eQTLs and link non-coding GWAS variants with putative target genes. Arts et al. (Cell Metabolism 2016) extended on their previous study that the cellular metabolism undergoes major shifts in beta-glucan-trained monocytes. Glucose, glutamine, and cholesterol metabolism were shown to be crucial in trained immunity and accumulation of fumarate is essential for epigenetic changes in trained immunity. A comprehensive analysis of 112 whole genome bisulfite sequencing data from many isolated cell types and blood cell stages by Schuyler et al. (Cell Reports 2016) provides a broad resource and a reference for comparison with diseased blood cell types. Galindo-Albarran et al. (Cell Reports 2016) showed that human neonatal CD8⁺ T cells have a distinctive transcription and chromatin landscape, and are biased towards an innate immune response as compared to adult CD8⁺ T cells. This bias could explain the sensitivity of neonates to infections and inflammation

In summary, WP3 has delivered beyond expectation and the data generated will be one of IHEC cornerstones for years to come. BLUEPRINT has successfully generated and explored the epigenetic mechanisms underlying all blood cell types formation from the hematopoietic stem cell, progenitors to the mature cell types. WP3 has characterized all stages of B cell development and explored immune tolerance's molecular mechanisms. The potential of these data has been showcased in a series of high impact articles published in leading journals in 2016. Because all data were made freely available for the scientific community to analyse and use as reference early on after generation, we expect that many more discoveries will follow in the near future. The data will undoubtedly have a significant impact in precision medicine, haematologist working on blood cancers will find the data on progenitor cells and B cells development extremely useful to place different leukaemia cells and lymphomas (see WP4 and WP5) in a hierarchical model and to identify their origins.

WP4 – Epigenome of normal and neoplastic B- and T-cells

Lymphoid malignancies account for approximately two thirds of all hematologic neoplasms with increasing incidence. Around 80-90% of the lymphoid tumors are derived from B-lineage cells, the reminder from T-lineage cells. Both B- and T-cell neoplasms are biologically heterogeneous and outcome strongly depends

on the biological subtype defined by the underlying genetic aberrations and the maturation stage at which the malignant cells are 'frozen'. In order to better understand the biological and clinical heterogeneity of lymphatic tumors, WP4 of BLUEPRINT characterised the epigenome of the B- and T-cell tumors most common in Europe across all age groups, namely precursor B- and T-lineage acute lymphoblastic leukaemia (preB- and T-ALL), chronic lymphocytic leukaemia (CLL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL), Burkitt lymphoma (BL), Mantle cell lymphoma (MCL) and T-prolymphocytic leukemia (T-PLL) including various molecular subtypes of these diseases. In order to better understand epigenetic aberrations in the tumor cells, in parallel the epigenome of different normal B- and T-cell populations were analysed as a joint effort between teams in WP3 and WP4.

Despite considerable technical challenges posed by the rareness of some cell populations, full reference epigenomes of each three samples of naive B cells from peripheral blood, naive B cells from tonsils, germinal center B cells, memory B cells and plasma cells have been generated. Similarly, at least one full reference epigenome for each of five major T cell subpopulations was generated. By these analyses, we have made several key contributions to the field of normal lymphopoiesis, such as the characterization of B- and T-cells at different maturations stages sorted from healthy donors (Kulis et al., Nature Genetics 2015, Galindo et al., Cell Reports in press) or from an in vitro differentiation system (Caron et al., Cell Reports 2015). Among others, we characterized for the first time, DNA methylome during an entire cellular differentiation process, i.e. B-cell differentiation. We observed that non-CpG methylation disappeared upon B-cell commitment whereas CpG methylation changed extensively during B-cell maturation. Modulation of CpG methylation showed an accumulative pattern mostly targeting enhancers, heterochromatin and polycomb-repressed regions. The methylation loss in enhancers was associated with upregulation of key B-cell transcription factors and affected genes relevant for the immune system showing high and variable expression levels. On the contrary, demethylation of heterochromatin and methylation gain of polycomb-repressed areas did not have an apparent functional impact in B cells. Instead, we observed that this epigenetic signature was prevalent in long-lived B cells, and also in neoplastic B cells, suggesting the presence of an epigenetic drift associated with cellular longevity, both under physiological and pathological conditions. A summary figure is shown in the attached document.

Based on the findings in the normal lymphocyte populations, we have characterised the genome of neoplastic lymphocyte populations. In this context, we have completed the reference epigenomes of B and T cell neoplasms, leading to a total of 31 complete reference epigenomes and 24 partial epigenomes. In addition, several "light epigenomes" have been generated, analysing single informative marks. The epigenomic characterization has been further extended to some representative cell lines serving as models for the respective lymphomas. Based on these analyses, we could publish the first complete DNA methylomes including lists of differential epigenetic marks and disease-specific biomarkers of several B-cell tumors, including chronic lymphocytic leukaemia (Kulis et al., Nat Genet 2012), mantle cell lymphoma (Queiros et al., Cancer Cell accepted for publication), multiple myeloma (Agirre et al., Genome Res 2015), follicular lymphoma and Burkitt lymphoma (Kretzmer et al., Nat Genet 2015). Moreover, integrative (epi)genomic analyses in the context of a collaboration between BLUEPRINT and the Spanish and German ICGC projects has led to the identification of non-coding mutations with functional impact (Puente et al., Nature 2015), new oncogenic long non-coding RNAs (Doose et al., PNAS 2015) and a link among somatic mutation, DNA methylation and transcriptional control (Kretzmer et al., Nat Genet 2015). Finally, the clinical relevance of some epigenomic biomarkers has been validated in independent series (Queiros et al., Leukemia, 2015).

In summary, WP4 has delivered far beyond expectation and the data generated will be one of IHEC and ICGC cornerstones for years to come. Not only did the teams provide novel and valuable insight into the epigenetic contribution to pathobiology, they also determined and validated a signature profile for classification and prognosis.

WP5 – The epigenome of acute myeloid leukaemia

The main objective of WP5 was to establish and analyse reference epigenomes of myeloid diseases, in particular acute myeloid leukaemia (AML). These analyses intended to highlight the AML specific epigenomic features but also to establish the epigenomic variety within this subclass of leukaemia's. For this, BLUEPRINT focussed on different AML subtypes that from a clinical point of view represent leukaemia's with good, intermediate and adverse prognosis. Subtypes analysed in BLUEPRINT included the good prognosis translocation t(15;17), t(8;21) and inv(16) AMLs, the adverse prognosis complex karyotype AMLs as well as intermediate prognosis AMLs harbouring mutations in the Nucleophosmine (NPM1) gene and the FLT3 gene. In addition, AML samples were screened for further genetic mutations including those in epigenetic enzymes such as DNMT3A, TET2, IDH1 and EZH2, transcription factor such as RUNX1 as well as splicing factors such as SRSF2 and SF3B1 to obtain a complete overview on genetic alterations underlying these samples. For each AML cell type RNA for RNA-seq analysis, DNA for bisulfite-seq analysis (WGBS), chromatin for ChIP-seq analysis and DNaseI treated nuclei for DNaseI-seq analysis, were isolated allowing the establishment of full BLUEPRINT epigenomes.

In total 23 full AML epigenomes were established while for an additional 18 full ChIP-seq and RNA-seq profiles were made. All data was transferred to EGA and are available for the community either as processed data via the BLUEPRINT-DCC site or as raw data via application to the Data Access Committee (DAC). In addition to these samples, several ex vivo treated samples were mined in the context of drug screenings performed in WP14/WP15 or the DNA methylation analysis of WP9. Finally, epigenomes of AMLs were examined before and after xenotransplantation. For this, several primary AMLs (p0) were transplanted in mice (p1 and p2) and analysed for chromatin structure alterations in collaborations with activities in WP14/15. Downstream analysis of epigenomic data focussed on identification of the commodities and differences within the adult AML epigenomes, but also in relation to normal myeloid differentiation. A first publication appeared as part of the IHEC package. Mandoli et al. (Cell Reports 2016), investigated the effects of the t(8;21) acute myeloid leukaemia (AML)-associated oncoprotein AML1-ETO on the transcriptome and epigenome in t(8;21) patient cells. They show that the interplay of the epigenome and transcription factors prevents apoptosis in t(8;21) AML cells which may have ramification for treatment. Several other publications are in preparation and will be submitted in the first half of 2017.

WP6 – Data Coordination Centre

BLUEPRINT produced a large volume of data: 2,558 experiments were run on 1,040 samples taken from 498 donors and 7 cell lines. Their analysis led to the further creation of 11,278 files. The mission of WP 6 was to keep track of it all and make the data available to the wider scientific community, with a focus on security, availability, discoverability, visualisation and reusability.

All data and results produced by BLUEPRINT are available to the scientific community for further analysis and to support new research. For the raw data types that include potentially identifiable human genetic variation and which was collected under an appropriate consent agreement, data is shared with 'bone fide'

researchers via the European Genome-phenome Archive (EGA) (<https://www.ebi.ac.uk/ega/dacs/EGAC00001000135>) where it enjoys the appropriate level of security. These data can be obtained by applying to the BLUEPRINT Data Access Committee (blueprint-dac@ebi.ac.uk) with details of an appropriate scientific project in the context of a recognized research organisation. The data can then be downloaded in an encrypted form from the EGA. Every other dataset produced by the consortium is available openly via a number of means. Raw data that could be released openly was deposited at the European Nucleotide Archive. Downstream results are available on our FTP site (<ftp://ftp.ebi.ac.uk/pub/databases/blueprint>). During the course of the project, updates to the public data were published in half yearly releases. This system ensured in particular that the analysis teams were consistently working from the same files, at the same time as the data was being produced. This synchronicity not only sped the project, it allowed the downstream analysis to provide feedback to the data production teams.





To facilitate discovery of a specific data file of interest within the collection, we developed user-friendly tools to search through the BLUEPRINT collection. In particular, the Data Portal (<http://dcc.blueprint-epigenome.eu/>) allows any visitor to quickly browse through the different BLUEPRINT products. Filtering by a number of criteria, they can obtain identifiers to private datasets and direct links to public ones.

Genome specialists will likely prefer to visualise BLUEPRINT data in the context of their other genomic annotations. We set up the data so that it can be automatically loaded onto popular genomics tools, in particular:

- the Ensembl Genome Browser:
(http://ensembl.org/Homo_sapiens/Location/View?g=ENSG00000130544;contigviewbottom=url:http://ftp.ebi.ac.uk/pub/databases/blueprint/releases/current_release/homo_sapiens/hub/hub.txt;format=DATAHUB;menu=Blueprint%20data)
- the UCSC Genome Browser:
(http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hubUrl=http://ftp.ebi.ac.uk/pub/databases/blueprint/releases/current_release/homo_sapiens/hub/hub.txt)
- the Genomatix browser (<https://blueprint.genomatix.de/>).

BLUEPRINT is one of the largest available genomic data collections, and as such presents challenges to some commonly used tools. In particular, our data is highly dimensional, and genomic browsers are not designed to view many datasets simultaneously. We, therefore, developed new data visualisation and exploration tools to explore our results. Since this is a new problem, and designing an effective user interface is a matter of trial and error, we developed three separate prototypes. The Data Analysis Portal (http://blueprint-dev.bioinfo.cnio.es/release_2016-08/) guides the visitor through a number of preliminary questions, leading them eventually to a collection of colourful and interactive plots. The DeepBlue portal (<http://deepblue.mpi-inf.mpg.de/dashboard.php#ajax/dashboard.php>) lets the visitor choose a number of filters and criteria to find the data of their choice, then extract what they need from those files. Thirdly, the GenomeStats browser (<http://genomestats.blueprint-epigenome.eu>) lets the visitor summarise any subset of BLUEPRINT experiments into a single file, which can easily be




downloaded or viewed in the traditional genomic browsers mentioned above.

Finally, to ensure the longevity and reusability of BLUEPRINT's work, we took a leadership role in the International Human Epigenome Consortium (IHEC - <http://ihec-epigenomes.org/> ). BLUEPRINT set up and maintained the EpiRR registry (<http://www.ebi.ac.uk/vg/epirr> ) where a user can find references to all available epigenomic datasets, including BLUEPRINT's. These references are then fed into IHEC's data portal (<http://epigenomesportal.ca/ihec/> ). This massive integrative work required us to coordinate with all the major epigenome consortia, and together define and implement common standards for data and metadata quality (<http://ihec-epigenomes.org/research/reference-epigenome-standards/> ).

Through this array of automated web services, we expect BLUEPRINT datasets to be accessible to the research community and the public and serve as a basis for integrative analyses long after the official end of the project.

WP7 – Data Analysis Group

Work Package 7 had the ambitious objective to organize, distribute and analyse the large epigenome datasets produced from around 100 different hematopoietic cell types. From the very beginning, an effort was made to establish a common data processing and primary analysis framework. This served the purpose of making the data trustable, robust and comparable among different experiments and cell types. These efforts have produced more than 2.500 analysis products from around 2.558 experiments.

A major challenge has been to do make this large dataset and the raw data accessible to the scientific community in a fast, easy and intuitive way. Usually, the process to identify and download the data is tough and cumbersome, especially for complex and large datasets. Here, in coordination with WP6, BLUEPRINT has made major efforts to simplify, organize and improve the access and visibility of the BLUEPRINT data, including the DCC portal (<http://dcc.blueprint-epigenome.eu/#/home> ) which is the main entry point for the raw and primary processed data. This effort has been combined with other more specific strategies like DeepBlue (<http://deepblue.mpi-inf.mpg.de/> ) with programmatic data access for advanced computational users, or BDAP (http://blueprint-dev.bioinfo.cnio.es/release_2016-08/#/ ) a comparative epigenome web portal accessible for a broader (non-expert) audience. These resources provide access to downstream analyses like gene expression values, hypo or hyper-methylated regions or histone binding profiles, among others for the different cell types in BLUEPRINT. In addition, trackhubs are available for ENSEMBL and UCSC browsers to visualize and explore the data in a genomic context. BLUEPRINT is convinced that all these efforts will help scientists to address more complex questions, providing fast and intuitive access to this impressive dataset avoiding experimental and analytical duplicities.

The huge amount and complexity of data generated by BLUEPRINT provided opportunities to explore new methodologies and tools that could take advantage of this rich dataset to further our understanding of hematopoiesis and related diseases. A first type of methods provides large-scale analysis and data interpretation, including RnBeads (Assenov et al., Nature Methods 2014) or metilene (Jühling et al., Genome Research 2016) for DNA methylation, NucHunter to infer nucleosome positions from ChIP-Seq data (Mammana et al., Bioinformatics 2013), CoSi to characterize splicing in RNA-seq experiments (Pervouchine et al., Bioinformatics 2012) or methods to analyze single cell RNA-seq (Jaitin et al., Science

2014).

The second type are those that perform multidimensional epigenomic data annotation, like the Ensembl Regulatory Build (Zerbino et al., Genome Biology 2015) and EpiExplorer (Halachev et al., Genome Biology 2012). Finally, the third types of methods developed by BLUEPRINT are those related with the integration of data from different sources. An example of BLUEPRINT results in this respect are methods integrating ChIP-seq data with protein evolution, network analysis and literature curation to elucidate the role of different epigenomic components in a molecular communication network (Juan et al., Cell Reports 2016). Furthermore, Pancaldi et al. (Genome Biology 2016) developed the chromatin assortativity measure to contextualize epigenomic data fusing 3D chromatin interaction networks produced by different chromosome capture methods, identifying factors that play a structural role in shaping chromatin. In addition, Carrillo de Santa Pau et al. (2016) and Paul et al. (2015) have developed different algorithmic frameworks to study hematopoietic differentiation from an epigenome and transcriptome view.

All these approaches and others produced within BLUEPRINT have been applied to several projects carried out in collaboration with biologists and physicians to address their hypotheses from a computational point of view. This collaboration between different sides has been very fruitful providing new insights in the relation between cell fate determination and disease (such as Kulis et al., 2015; Galindo-Albarrán et al., 2016) or non-coding variants and variability (Aistle et al., Cell 2016; Chen et al., Cell 2016), among others.

In conclusion, WP7 has made a very impressive dataset publicly available to a wide audience that can now take advantage of the full downstream analysis and the different methods generated. These efforts will open a new era to further the study of the complex relationship between the epigenome and the genome, advancing our basic science knowledge and improving clinical practices.

WP8 – DNA methylation variation in T1DM

Diabetes is a devastating disease responsible for substantial mortality and morbidity in the developed world. The incidence of diabetes has substantially increased in recent years. Of the two major types of diabetes, type 1 diabetes (T1D) can be the most severe, being the second commonest chronic childhood disease after asthma. BLUEPRINT postulated that the gene-environment interaction could be mediated, in whole or in part, by epigenetic effects including DNA methylation in selected peripheral blood cells critical to the disease and exposed to the metabolic changes responsible for diabetes complications. BLUEPRINT performed an epigenome-wide association study across 406,365 DNA methylation sites (CpGs) in 52 monozygotic (identical) twin pairs discordant for T1D in three immune effector cell types: CD4⁺ T cells, CD19⁺ B cells, and CD14⁺CD16⁻ monocytes. By using disease-discordant monozygotic twins, our strategy reduced major confounding effects, such as inter-individual genetic variability and in utero effects.

We observed a substantial enrichment of differentially variable CpG positions (DVPs) in T1D twins compared to their healthy co-twins across all cell types. Compared to the healthy, unrelated individuals, patients with T1D showed cell type-specific enrichment of these changes and the same changes were found to be temporally stable and enriched at gene regulatory elements. Integration with cell type-specific gene regulatory circuits highlighted pathways involved in immune cell metabolism and the cell cycle, including mTOR signalling. Evidence from cord blood of new-borns who progressed to overt T1D, suggests that the DVPs likely emerged after birth. Our findings, based on the largest study of its type to

date involving 772 methylomes, implicate epigenetic changes that could contribute to disease pathogenesis in T1D. This study (Paul et al., Nature Communications 2016) has illustrated, for the first time, the potential effect of epigenetic changes in T1D, which could potentially be translated into diagnostic and prognostic epigenetic biomarkers.

WP9 – Biomarker development

Epigenetic biomarkers have huge potential for precision medicine. The objective WP9 was to establish a solid foundation for widespread use of epigenetic biomarkers in the personalized therapy of blood cancers and other diseases. Toward this goal, BLUEPRINT established and validated both experimental and computational methods for epigenetic biomarker development, and successfully conducted biomarker development projects for selected hematopoietic malignancies. Moreover, BLUEPRINT contributed to the broader dissemination of good practices in epigenetic biomarker development through review articles, workshops, and presentations.

Among all results and milestones that have emerged from research in the context of biomarker development in BLUEPRINT, we highlight six key achievements – selected on the basis of their broad relevance and their illustrative character for the global impact of BLUEPRINT on the fields of epigenetic biomarker development and precision medicine.

1. DNA methylation assay validation and benchmarking: BLUEPRINT initiated and coordinated a community-wide benchmarking study that comprehensively established the accuracy and robustness of DNA methylation biomarkers for clinical research and diagnostics (Bock et al., Nature Biotechnology 2016). The paper summarizes a huge body of work (including ~500 pages of supplementary reports and documentations) contributed by 18 different laboratories in seven countries. This type of work was feasible only in the context of a project of the scope and scale of BLUEPRINT, and it has already had substantial impact by boosting global interest in epigenetic biomarkers for a broad range of diseases.

2. Open-source software for DNA methylation analysis: To facilitate and standardize the bioinformatics of epigenetic biomarker development, a software package for DNA methylation analysis based on all widely used assays has been developed (Assenov et al., Nature Methods 2014). The RnBeads software has established a powerful pipeline for analysing DNA methylation profiles and for identifying biomarker candidates, thereby making epigenetic biomarker development much more accessible for clinical researchers.

3. A validated DNA methylation biomarker for disease subtype prediction in paediatric ALL: Building upon an initial large-scale discovery analysis of DNA methylation in 764 children newly diagnosed with paediatric acute lymphoblastic leukaemia (ALL) (Nordlund et al., Genome Biology 2013), a DNA methylation biomarker for disease subtype prediction in paediatric ALL has been developed and validated (Nordlund et al., Clin. Epigenetics 2015). This biomarker uses DNA methylation to accurately classify samples according to established, karyotypically defined subgroups of paediatric ALL. It is particularly useful when karyotypic information is unavailable or of low quality/sensitivity, and it will contribute to better diagnosis for patients that cannot be categorized based on karyotypic information alone.

4. A validated DNA methylation biomarker for patient stratification in CLL: Utilizing large DNA methylation

maps for chronic lymphocytic leukemia (CLL) established by BLUEPRINT (Kulis et al., 2012), a DNA methylation biomarker for patient stratification in CLL has been developed and validated (Queiros et al., 2015). This biomarker accurately distinguishes between three major subgroups of CLL that correspond to different stages in normal B cell development. By identifying a robust minimal set of CpGs whose DNA methylation levels are sufficient to predict disease subtype and testing the resulting biomarker in two independent cohorts, a strong case has been made for the diagnostic potential of epigenetic biomarkers for patient stratification in CLL.

5. Proof-of-concept for open chromatin analysis for patient stratification in CLL: Looking beyond DNA methylation, ATAC-seq profiles for 88 samples obtained from 55 CLL patients have been established to provide proof-of-concept for using chromatin accessibility as a biomarker (Rendeiro et al., Nature Communications 2016). Bioinformatic analysis showed that characteristic chromatin signatures reflect and predict the three major subtypes of CLL that were also identified by the DNA methylation biomarker described above. Furthermore, disease subtype-specific regulatory networks were derived from the chromatin accessibility data, suggesting that the ATAC-seq assay has future potential as a biomarker that reflects precise regulatory differences between disease subtypes.

6. Education & dissemination: Good practices of epigenetic biomarker development have been discussed and disseminated in numerous review articles, workshops, and presentations. Most notably, BLUEPRINT researchers summarized the state-of-the-art of DNA methylation profiling in the clinic (Heyn and Esteller, 2012), provided detailed instructions on how to analyse and interpret DNA methylation data (Bock, 2012), and contributed to recommendations for the design and analysis of epigenome-wide association studies (Michels et al., 2013).

The six key achievements highlight above showcase the global impact of BLUEPRINT on the fields of epigenetic biomarker development and precision medicine.

WP10 – The effect of common sequence variation on the epigenome landscape

Many human complex diseases are characterized by dysregulation of immune and inflammatory activity. However, the repertoire of immune genes and cell subsets implicated in the pathogenesis of individual disease can vary dramatically. Genome-wide association studies (GWAS) have contributed to expanding catalogues of implicated genes and pathways for many complex human diseases, and are beginning to shed light on shared and unique etiological and pathological components of disease. A key challenge is that these disease variants map predominantly to noncoding regions of the human genome, where they are predicted to alter regulatory function. Linking susceptibility variants to their respective causative genes and cell-specific regulatory elements thus remains a main priority in order to realize the potential of association studies to advance understanding of disease biology and etiology, leading to therapeutic advances.

Molecular quantitative trait locus (QTL) studies, testing for associations between genetic variants and intermediate phenotypes, in particular gene expression levels, provide powerful approaches to annotate the putative consequence of disease associations. Particularly when applied to primary human cells, these studies allow to unravel the cell- and context-specific regulatory effects of complex disease variants. Here, we report an integrated analysis of genetic, epigenetic and transcriptomic datasets in the three major cells of the human immune system, namely CD14⁺ monocytes, CD16⁺ neutrophils and CD4⁺ naïve T cells.

Monocytes contribute to maintenance of the resident macrophage pool under steady state conditions, and migrate to sites of infection in the tissues and divide/differentiate into macrophages and dendritic cells to elicit an immune response. Neutrophil granulocytes (neutrophils) are primary blood cells of the innate immune and inflammatory response system that form a first line of organismal response to bacterial and fungal infection, migrating within minutes to sites of infection, attracted by local tissue factors and resident macrophages during the acute phase of inflammation. Finally, CD4+ naïve T cells are part of the adaptive immune response system, representing mature helper T cells that have not yet encountered their cognate antigen.

BLUEPRINT, with a strong contribution of its Canadian IHEC partner from McGill, generated high-resolution whole-genome sequence, transcriptome, DNA methylation and histone modification datasets in up to 197 individuals selected from a population-based sample, and applied different genomic analyses to investigate genetic and epigenetic influences to transcription and RNA splicing in the three cell types. We demonstrate colocalization of molecular trait QTLs with 345 unique genetic variants predisposing to seven human autoimmune diseases, involving all molecular data types. Overall, the data and results deepen our understanding of genetic and epigenetic regulation of the transcriptional machinery in three primary cells of the immune system, and inform the formulation and testing of functional hypotheses for human complex disease. The first fruits from the efforts of BLUEPRINT to describe the genetic and epigenetic variation have appeared in the IHEC package (Chen et al, Cell 2016; Astle et al., Cell 2016)

WP11 – Mouse models to quantify variation in reference epigenomes

WP11 has delivered integrated epigenomes and transcriptomes from purified populations of naïve B and naïve T cells isolated from two strains of mouse (C57BL6 and CAST and reciprocal hybrids), from males and females. Because these cells are not actively dividing and are pure, they are very homogenous and the data we have generated is very high quality. Datasets of this nature allow us to make hypotheses to address questions about how the epigenome regulates the function of the genome and importantly, to test these experimentally. Unlike humans who have highly variable DNA sequences, the genome of a mouse strain is identical from one mouse to another and this allows us to separate epigenetic functions from genetic ones.

In this Work Package, BLUEPRINT has identified high levels of hydroxymethylcytosine in naïve T cells and addressed the relationship of this mark with DNA methylation and gene expression. We have also identified and characterised very striking levels of unmodified DNA that are conserved in human, and addressed the function and epigenetic states associated with this feature. Our datasets allow the quantification of parental origin specific gene expression, the analysis of sex-specific epigenetic and transcriptional differences and the relationship between genetic and epigenetic variation through the analysis of two genetically different strains of mouse.

Many transcription factors regulate gene expression by binding to specific DNA sequences consisting of the four-base alphabet. However, epigenetics provides additional potential specificity for factor binding. In collaboration with Viner and Hoffman, we have used our BLUEPRINT methylomes and hydroxymethylomes from C57BL6 to model transcription factor binding to this expanded alphabet that includes epigenetic modifications providing an approach to determine novel insights into the rules by which gene regulators interact with specific gene sequences - Viner et al. (bioRxiv 2016).

WP12 – Technology development for profiling of cytosine (hydroxyl)methylation

The purpose of this Work Package was the development of novel epigenomic technologies which could be used by the BLUEPRINT consortium and the European and international research communities. In general, technology development could be expected to accelerate the progress with obtaining epigenomes of interest for BLUEPRINT, especially when dealing with cell types of low abundance or that are inherently more difficult to profile. WP12 developed six new epigenomics technologies, contributed significantly to the establishment of the company Cambridge Epigenetix, and we expect will also have a lasting impact on young epigeneticists careers, and on ongoing and future flagship programmes.

Epigenomics technologies developed

Most of the technologies developed centre around the methylome (and other DNA modifications) and its context-dependent regulation. DNA methyltransferases interact with chromatin in order to create methylomes, but profiling of the methyltransferases themselves has been notoriously difficult. However, through the advent of unique tagging of the endogenous enzymes, high quality profiles have been obtained, allowing initial insights into chromatin contexts that are permissive or non-permissive, respectively, for methylation targeting. Complementing these insights, a method was developed with which the methylome is sequenced after enriching for certain chromatin environments. For example, chromatin bound to particular transcription factors or modified in particular ways is pulled down (enriched) with subsequent sequencing of the methylome, showing the precise methylation patterns that exist in this context.

New methods were developed to sequence DNA modifications other than methylcytosine, such as hydroxymethylcytosine, formylcytosine, or carboxycytosine. This includes base-resolution techniques as well as pulldown-ones in order to maximize either resolution or throughput and cost-effectiveness. (Booth et al. Science 2012; Iurlaro et al. Genome Biology 2013).

A major achievement was the development of methods for single cell epigenomics. Cell identity is established and changed in single cells, and understanding these dynamics needs single cell profiling technologies. Notably, a technique was established that sequences the methylome in single cells (50% genome coverage at this point), revealing profound epigenetic heterogeneity between cells (Smallwood et al., Nature Methods 2014; Clark et al., Genome Biology 2016; Angermueller et al., Nature Methods 2016; Farlik et al., Cell Reports 2015). In a further development, it was possible to combine methylome and transcriptome sequencing from the same single cell, revealing for the first time the intricate connections between methylome and transcription variation from cell to cell in development, disease, and ageing

WP13 – Technology optimization for microscale application

Chromatin-IP is the key experimental approach to locate histone modifications and transcription factors in the genome, yet for robust and automatized detection this key assay requires large numbers of cells. The aim of WP13 was to develop prototypes for ChIP miniaturization and automation. Priority was given to the development of a microsonicator and the development of a microfluidic system. In the course of the project, a call was launched to integrate new partners to develop new technologies which led to the integration of two EMBL teams (Furlong and Mertens) in the BLUEPRINT consortium and a close collaboration with Fluidigm, a leading company in microfluidics technology.

As part of the project, SME partner Diagenode managed to develop a prototype of a micro-sonication system for chromatin shearing in microliter scale volumes. This piezo-electric transducer (PZT) based

microsonicator desktop device was shared with partner RU and performance was validated. In addition, custom microfluidic chips were developed by Fluidigm and tested by RU for sequence library construction and for low cell input ChIP. Proof-of-concept using these microfluidic chips for low cell input ChIP-seq was provided by profiling four canonical histone modifications. Further optimization is ongoing and it is anticipated that the BLUEPRINT custom microfluidic chip will become commercially available in the course of 2017. The integrated workflow is compatible with Fluidigm hardware that is on the market and will allow parallel analysis of clinical samples (biopsies).

WP14 - Identification and validation of Epi-targets.

In this WP, mouse models of leukaemia based on expression of AML1-ETO or MLL-AF9 have been used. In addition, a mouse model of AML with mutated NPM carrying the FLT3-ITD mutation has been established. Collectively, these three models present the major subtypes of human AML. The mouse models have been used to screen for epigenetic targets causal for initiation and/or maintenance of leukemogenesis by in vivo shRNA-based approaches. The libraries have targeted chromatin-associated proteins with a putative role in epigenetic control (histone methyltransferases, histone demethylases, HDACs, HATs, Chromatin remodelers etc.). Next Generation DNA sequencing has been used to determine selection of genes encoding for potential Epi-Targets. In mouse MLL-AF9 cells, 5 potential targets have been identified. Depletion of every single one of these 5 genes impairs growth and colony formation of mouse MLL-AF9 cells as well as of human MLL-rearranged AML cells. In addition, the putative histone demethylase Jmjd1c has been discovered as potential therapeutic target in leukaemia since its depletion increases apoptosis in both mouse and human leukemic cells (Sroczyńska et al., Blood 2014).

In addition the screening of a custom-made lentiviral library containing 1200 shRNAs targeting 110 genes coding for epigenetic modifiers involved in cancer development has been completed in 4 different AML samples: AML5 (t(1;2); NPM wt; FLT3 mutated); AML9 (complex karyotype; NPMWT; FLT3 WT); AML-IEO20 (t(9;11); NPMWT; FLT3 WT); AML-IEO23 (normal karyotype, mutant NPM, mutant FLT3).

Attention was focussed on the depleted shRNAs since they should target putative oncogenes that, in turn, could identify new therapeutic targets.

CBX2 and BRD9, belonging to the reader class, have been identified as potential drug targets. Several analyses revealed that both CBX2 and BRD9 are strongly expressed in different human cancer cell lines compared to normal cells. In addition, RNA-seq analysis performed on 200 different AML samples revealed that both BRD9 and CBX2 are overexpressed compared to normal progenitor cells, and are strongly reduced in more differentiated cells such as monocytes and macrophages (Di Costanzo et al., submitted for publication).

Furthermore, HDAC2 has been identified as having a specific role in leukemogenesis, connected to the expression of genes involved in activation of immune responses, such as genes in the human leukocyte antigen (HLA) family, required in the effector stages of antitumor immunity (Conte et al., Oncotarget 2015). Finally, a novel role for HDAC1, oncosuppressor in tumorigenesis, oncogene in tumor maintenance, has been described (Santoro et al., Blood 2013).

Hits characterization and epigenome profiling of xenotransplanted human AMLs

To identify chromatin remodeling factors (CRFs) playing a role in transdifferentiation (TD) and reprogramming into pluripotency, two systems of TD, namely the C/EBPa induced conversion of mouse and human B cell lines into macrophages have been established. The first is the murine C11 cell line that

contains an estradiol inducible C/EBPα ER transgene. The second is the Burkitt lymphoma line Seraphina containing C/EBP ER (BLaER1 cells). Several shRNA candidates inhibiting or accelerating TD have been identified (such as HDAC1, UBA1, Sin3B, USP27X and CHD6, CHD8, L2MBTL3, PRKAA1, respectively). In silico screen of CRFs down or upregulated during the C/EBPα induced TD of BLaER1 cells identified as a top candidate WHSC1 as this gene becomes not only downregulated during human but also mouse TD into macrophages, as well as during Yamanaka factor induced reprogramming of mouse pre-B cells into iPS cells (Di Stefano et al., Nature 2014).

By LC-MS/MS analysis, BRD9 (a candidate target identified in human AML ex vivo blasts) was identified interacting with proteins belonging to SWI/SNF complex: SMARCA4, SMARCA2, SS18, SMARCD1. BRD9 was also identified to interact with two proteins: GLTSCR1 and GLTSCR1L, the function of which is still unknown, and with a subset of proteins involved in the DNA repair machinery: EMSY, CHD9, BRCA2, UBR5, ATAD5. This data highlighted that BRD9 is part of SWI/SNF complex (Del Gaudio et al., submitted for publication).

59 human samples (54 primary and 5 relapsed AMLs) have been transplanted into NOD SCID IL2RG null (NSG) immunocompromised mice of which 24 successfully engrafted. The aim was to develop mouse models for screening of novel compounds against leukaemias. For this approach to be successful and informative, it is essential that the epigenome of the xenotransplanted AML reflects the original human AML makeup. For this purpose, the epigenome from the original disease has been compared to the pattern obtained in the xenotransplanted mice (from two different sites: bone marrow and spleen) and genome-wide maps have been generated for 3 modifications (H3K4me3, H3K27me3 and H3K27Ac) associated with gene activation and gene repression. The hierarchical clustering approach used to assess the degree of similarity confirmed that the chromatin structure of the original human samples is maintained in xenotransplanted AMLs which strongly suggest that the mice can be used for compounds screening.

WP15 – Compound development and screening

The goal of Work Package 15 was the discovery and development of chemical compounds with the ability to modulate epigenetic processes. The motivation for the use of compounds targeting the epigenome was both for research – e.g. as chemical tools to study processes in cells – as well as for starting points for the development/optimization of therapeutic agents for the treatment of cancers.

The targets of “epigenetic drugs” are typically proteins that associate with, and modify the function of chromatin. These protein targets frequently constitute enzymes, but also non-enzymatic “readers” of chromatin states such as the Bromodomain-containing proteins. The enzyme targets consist mostly of “writers” or remodelers of chromatin modifications, such as DNA methyltransferases, histone methyltransferases, and histone acetyltransferases, as well as the corresponding “erasers”, the demethylases and deacetylases. The “reader” targets comprise proteins which bind to the chromatin modifications induced by the “writers”, including the bromodomain “acetyl readers”, and several classes of “methyl readers”. WP 15 focussed on the design, synthesis and biochemical/biological evaluation of compounds acting on all of these target classes:

- Modulation of DNA methylation: Two series of DNMT inhibitors have been reported, the first with analogues of SGI-1027 more selective for DNMT with respect to other methyltransferases, and characterized by the same potency in cancer cells along with lower toxicity. The latter is a series of properly substituted quinazolines selective against DNMT3A and active in leukaemia.

- Modulation of histone acetylation: BLUEPRINT reported the discovery of MC2392, a context-selective ATRA/HDAC hybrid molecule that selectively inhibits the PML-RAR α -HDAC complex in APL without affecting any other HDAC-containing complexes. (De Bellis et al., Cancer research 2014) We also described other types of HDAC inhibitors (1,3,4-oxadiazoles, cinnamyl and pyrrolacrylic derivatives) and demonstrated activity in cancer cells including sarcoma CSCs, and activity as tools in the modulation of globin gene expression (β -thalassemia) or in activation of VEGFR (Valente et al., J Med Chem 2014). We also reported on compounds targeting the Sirtuin-type histone deacetylases, including different series of Sirt1/2 inhibitors active in CSCs, and selective Sirt2 or Sirt5 inhibitors, the first arresting leukaemia cell growth at low micromolar level, the latter inducing autophagy and mitophagy through increase of ammonia levels (Polletta et al., Autophagy 2015). In addition, we performed a screen for modulators of SirT1 starting from a library of 10000 molecules. More than 10 potential modulators have been selected for further characterization which is ongoing.

- Modulation of histone methylation: We described bis-bromophenyl compounds endowed with dual HAT/EZH2(PRC2) inhibitory activity, which proved highly active as apoptosis inducers in cancer cell lines and in vivo. These dual agents were more potent than the related single-target inhibitors alone or in combination (manuscript submitted). We also developed derivatives of tranylcypromine, a weak KDM1/LSD1 inhibitor, with increased potency with respect to the prototype and potent in leukaemia at low micromolar concentration. Moreover, we developed novel pan-KDM inhibitors by linking pharmacophores active for LSD1 and JMJ family demethylases which are under investigation for cancer treatment. We described the in silico, in vitro, and cell-based characterization of the compound PKF118-310, an antagonist of transcription factor 4 (TCF4)/ β -catenin signalling, as inhibitor of KDM4A. PKF118-310 potential inhibitor activity was discovered via virtual screening on the crystal structure of KDM4A. Peptide-based histone tri-methylation assay developed in-house confirmed the potent KDM4A inhibitor activity. The protein target was validated by cellular thermal shift assay experiments. PKF118-310 anticancer activity was observed in both liquid and solid tumor cells, and shown to have a dose- and time-dependent effect. We demonstrated the previously unreported inhibitory action of PKF118-310 on KDM4A. Our findings open up the possibility of developing the first KDM4A-specific inhibitors and derivatives. We also performed a screen for modulators of KDM4A and identified 23 potential modulators for further characterization.

- Acetyl readers: The BET tandem-bromodomain proteins are showing great promise in early stage clinical trials. We established ChemSeq and Click-Seq technology to enable mechanistic studies on small molecule BET inhibitors and the interaction of their primary targets with chromatin. These techniques exploit drugs modified with biotinylated and/or click chemistry functions, to pull down their targets and the associated genomic sequences that are then mapped by NGS. We applied ChemSeq in combination with ChIP-seq to quantify drug sensitivity at individual genomic sites by treating cells with BET inhibitor in a dose-response fashion. The concentration-dependent displacement of BET proteins from Transcriptional Start Sites as well as Enhancer sites was monitored and quantified. Differences were observed in the extent of BET proteins displacement at the same concentration of inhibitor depending on the isoform (BRD2, BRD3 and BRD4) and on cell type. These genomic signatures of different inhibitors are expected to provide biomarkers and to guide the design of more selective second generation compounds. The first-generation clinical compounds are, however, neither selective for the different members of the target family, nor for the N-terminal versus C-terminal domains. Based on chemical proteomics screening

techniques, new pharmacophores were identified acting selectively on the first bromodomain of the BET target class. These inhibitors bear potential for second-generation compounds with an improved therapeutic index. By the same chemical proteomics screening, we identified novel small molecule inhibitors of Brd9 which possess nanomolar affinity. Functionalized analogues of these compounds tethered to beads identified different Brd9-containing complexes associated with different novel chemotypes. Brd9-selective inhibitors were selective for the BAF complex, whereas less selective compounds showed cross-reactivity with Brd7 and also bind the PBAF complex.

- **Methyl readers:** In contrast to the promising inhibitors of the acetyl-lysine-binding bromodomain proteins, there are no drug-like inhibitors for the even larger target classes of methyl-lysine binding domains. We performed a chemoproteomics-based screen of 20,000 compounds (a subset of the GSK library selected by a chemoinformatics approach) mimicking dimethyllysine/ dimethylarginine. The assay was formatted as a competition binding assay in nuclear lysates of HuT-78 cells to identify hits to endogenous methyl-readers (SPIN1, MPHOSPH8, UHRF1, CBX5, PHF23 and PHF8). The low number of hits obtained indicated limited tractability of methyl-binding domains. We identified fragment-like hits with low potency for SPIN1 binding to biotinylated H3K4Me3 histone tails with reasonable ligand efficiency, which may represent useful scaffolds for further optimization. Novel chemotypes were also identified for the Tudor-PHD domain proteins PHF19 and PHF1.

Work Package 16 – Training, networking and communication

Within the scope of WP16, BLUEPRINT has successfully developed an infrastructure for training, networking and communication activities.

Several workshops, training courses and symposia were organized, often jointly with other FP7 consortia, like the Network of Excellence EpiGeneSys, Marie Curie ITN network EpiTrain, collaborative project IDEAL, the German DEEP consortium and the Italian EpiGen consortium. Topics of these meetings were amongst others: Computational epigenomics, 450k analysis, Modelling approaches in Epigenetics, Bioethics of epigenomics data, Epigenomic data – a public resource, Epigenetic mechanisms in Health and Disease. Attendance varied from small groups (30-40 participants) to larger meetings with an audience of >100 participants.

In addition to the self-organized meetings, BLUEPRINT also sponsored a number of activities, like a symposium at the Human Genome Meeting (HGM) in Kuala Lumpur (Malaysia) in 2015, a satellite symposium at the NetSci meeting in Zaragoza in 2015, as well as a meeting of the London Epigenetics club in 2014. This also offered the opportunity to present BLUEPRINT and its outcomes at these meetings. For additional networking and training activities, a lab exchange program was set up. This program allowed BLUEPRINT participants, but later also other scientists, to visit BLUEPRINT laboratories to gain experience with techniques and to exchange information. The program covered the travel and housing costs for the visitors. This program was used by 9 BLUEPRINT members and 4 external scientists.


WP17 – Dissemination and outreach

For dissemination and outreach a website was set-up at the start of the project: www.blueprint-epigenome.eu. This website provides information about the participants, the goals and the outcomes of the project. This includes access to the epigenomics data generated, but also protocols, an overview of all the publications, films that were produced by the consortium etc. Furthermore, the BLUEPRINT website has an internal password-protected domain which is used to exchange information which is for consortium

members only.

In the final year, the website was used to announce and provide all relevant information about the BLUEPRINT/IHEC meeting which took place in Brussels (September 2016). This meeting was a joint activity with the IHEC community but it was also the final dissemination event of the BLUEPRINT consortium: during 2 'Days of Science', the major outcomes of the BLUEPRINT project were presented. Moreover, state-of-the art research in the field of epigenomics and closely-related areas was presented by renowned invited scientists.

In the final month, the website was again redone to communicate the release of a package of publications, by members of the International Human Epigenome Consortium (IHEC) in Cell and Cell Press journals. Over half of those publications had BLUEPRINT members in the lead. Besides the scientific publications, also an interview and a short film about a paper were released by Cell, as well as an editorial by Cell and an Essay by the IHEC members.

As part of the dissemination activities, also a user-friendly visual Epigenome Interface was developed (see <https://blueprint.genomatix.de/> ) to allow the non-expert scientist to visualize the epigenome data generated by the BLUEPRINT consortium.

To inform the wider scientific community as well as the more laymen public, BLUEPRINT has generated a short introductory film at the start of the project and a second short film and 3 thematic clips in the end phase of the project. The short films give an overview of the plans and the achievements of the project. The videos can be found on the BLUEPRINT website. Moreover, newsletters were generated about the project aims as well as about a number of publications from BLUEPRINT partners, presenting important outcomes of the project. These newsletters can also be found on the website.

Last but not least, BLUEPRINT has been the European cornerstone of the International Human Epigenome Consortium (IHEC) and as such has greatly contributed to this international initiative. Not only were BLUEPRINT members well represented in many of the committees and working groups of IHEC, BLUEPRINT was responsible for a major part of the data currently (end 2016) in the IHEC data portal as well as to the IHEC paper package that was published in Cell and Cell Press journals in November 2016. IHEC offered an important forum for networking and dissemination of new technological development and scientific insights. BLUEPRINT co-organized 2 of its meetings (kick-off meeting – October 2011 and final dissemination event – September 2016) together with IHEC.

Potential Impact:

BLUEPRINT reference epigenomes – impact on Health

The overriding aim of the BLUEPRINT project was to generate >100 epigenome profiles for the wider community to mine. BLUEPRINT decided to concentrate its efforts on blood cells and blood cell derived disease, in particular cancer. This has generated a comprehensive epigenetic description of this single 'organ', i.e. blood, that is second to none in terms of quality and range and has been produced to the highest standards to date.

The availability of high quality reference data is crucial for the advancement of knowledge and research and from the very first data releases, access to the data has been requested. BLUEPRINT has provided full epigenomes in excess of 140 according to IHEC standards from high purity cells derived from the hematopoietic tree in healthy individuals. Additional epigenetic marks have been measured on these cell

types to increase the amplitude of the BLUEPRINT epigenomes dataset which will turn into a world reference for many applications. BLUEPRINT has advanced the characterisation to date of the regulatory mechanisms in megakaryopoiesis and erythropoiesis that are highly relevant as platelets and red cells are involved in diseases such as stroke, myocardial infarction and anaemia.

BLUEPRINTs potential impact on immune disease treatment

The BLUEPRINT efforts on innate immune cells and their response to pathogens has been a game changer and has important ramification for the way we think about immune disease and its treatment, in particular of sepsis. Sepsis is a complex syndrome, triggered by infection, and the leading cause of death in intensive care units worldwide. Sepsis pathology can be separated into two states, an early inflammatory state and a late immune-tolerance state. The tolerized state is associated with high susceptibility to secondary hospital infections, due to a suppressed immune response. Despite positive results in sepsis research primarily in animal models, mortality rates have remained relatively unchanged (20-40%) for sepsis patients.

The epigenetic characterization of the monocyte-to-macrophage differentiation and the effect of exposure of monocyte to pathogens, led to the novel concept of trained macrophages (Saeed et al, Science 2014; reviewed in Netea et al, Science 2016). It was well-known that resting innate immune cells of healthy individuals respond to stimulation with pathogen-associated molecular patterns (PAMPs) by releasing a large array of proinflammatory mediators that activate short term responses like phagocytosis and killing of the invading microorganisms, and initiate long-term adaptive immune responses such as T-cell activation and antibody production. BLUEPRINTs early studies showed that macrophages remember exposure to pathogens when they were still monocytes. Our studies showed that the memory was written in the epigenome. In subsequent studies (Novakovic et al., Cell 2016), using an in vitro model of macrophage tolerance, BLUEPRINT identified several key pathways and transcription factors that show early differential expression in LPS exposed monocytes leading to an altered epigenome in the resulting tolerized macrophages. This altered epigenome is associated with pathways involved in phagocytosis and cytokine release. Future investigations should focus on how to (epigenetically) reprogram and re-activate immunologically-defective, or decommission aberrant functions of tolerized human monocytes and macrophages in sepsis patients.

BLUEPRINT and genetic variants predisposing to disease

This study sought to understand the complex molecular events associated with predisposition to severe human complex diseases. Here we focused on six common human disease such as Type 1 diabetes, rheumatoid arthritis, inflammatory bowel disease and celiac disease that affect a large proportion of the human populations. These diseases are characterized by dysregulation of immune and inflammatory pathways. BLUEPRINT studied three major human cells (namely monocytes, neutrophils and T-cells) known to play an important role in adaptive immune, and inflammatory response systems in humans and previously implicated in the pathogenesis of certain immune diseases. For instance, in inflammatory bowel diseases including Crohn's and ulcerative colitis, monocytes are involved in the recruitment of neutrophils during intestinal inflammation. An example of an autoimmune disease is Type 1 diabetes, which is caused by a T cell mediated destruction of insulin producing pancreatic beta cells.

BLUEPRINT generated the most expansive resource of molecular variations for these cells, and used it to demonstrate that hundreds of genetic variants that predispose to these devastating human diseases can be ascribed with high certainty to specific molecular variations, including changes in levels of expression of

genes. These results provide invaluable new information to drug development efforts, by indicating specific genes and molecular pathways that could be targeted to treat these diseases, using new and existing drugs.

BLUEPRINT blood cancer epigenomes and their impact on diagnosis and treatment

BLUEPRINT has generated full epigenomes of blood related cancer cells using the same standards. Lymphoid malignancies account for approximately two thirds of all hematologic neoplasms with increasing incidence. They affect all age groups and pose a considerable socio-economic burden as to expensive treatment, recurrent treatment failures and relapsing diseases requiring additional treatment. The BLUEPRINT datasets have significantly advanced our understanding of haematological cancers and allowed us to determine the cell of origin of many of these diseases. The concepts and biomarkers developed aid to unravel novel pathogenetic mechanisms probably targetable by innovative treatments. Finally, prognostic epigenomic biomarkers like those identified in CLL might help to save patients with good prognosis from extensive and expensive therapies and related side effects, whereas it can also identify high-risk patients in which the disease course might benefit from innovative approaches. These validated biomarkers are likely to find their way into clinical practice.

BLUEPRINTs contributed to the field of biomarker development also by performing a community-wide benchmarking study that established the accuracy and robustness of DNA methylation assays and provides confidence for their use as epigenetic biomarkers. Integrated wet-lab (Infinium/RRBS) and computational (CHAMP/RnBeads) pipeline for DNA methylation analysis makes epigenetic biomarker development much more accessible for clinical researchers.

BLUEPRINT and public data access




Next to generating highly valuable resources, BLUEPRINT has made significant efforts to make the data publicly and as freely as possible available within the boundaries defined by the informed consent. BLUEPRINT was the driving force in the International Human Epigenome Consortium (IHEC) to set standards not only at the level of data quality control and meta data standardisation but also in harmonizing data formats and data availability through the portals and web browsers. Thus, BLUEPRINT provided services to consortium members and researchers and the general public to timely make it a very impressive dataset publicly available to a wide audience that can now take advantage of the full downstream analysis and the different methods generated. These efforts will open a new era to further the study of the complex relationship between the epigenome and the genome, advancing our basic science knowledge and improving clinical practices.

In addition to the efforts to make the data available to the experts scientific community, Genomatix has developed an integrated graphical user interface that allows access to all data generated within the BLUEPRINT consortium combined with other publicly available data that might provide additional relevant background knowledge. The online web-based interface provides intuitive and straightforward access even for users not familiar with the complexities of querying biological databases or repositories. This enables scientists but also people working in the medical environment – e.g in the area of blood cancers - to query and mine the results of BLUEPRINT for their research questions, integrate the data in treatment strategies or even develop approaches for new therapies. They can hence create tremendous added value to the research efforts within BLUEPRINT.

Socio-economic impact of BLUEPRINT on SME

One of the SMEs participating in BLUEPRINT was Diagenode who played a central role in generation and validation of antibodies in the frame of BLUEPRINT. Parallel antibody batches, QC'ed and validated in the same way as for BLUEPRINT, were commercialized. Second, Diagenode launched the ONE (Bioruptor) early 2016 and although it is still early phase, several demos to users are ongoing. It may not be finished yet as Diagenode may get involved in automated ChIP-seq with a BLUEPRINT-custom-designed microfluidic-device for the Fluidigm C1 machine as the producer/seller of kits. BLUEPRINT not only gave Diagenode new products in particular 'BLUEPRINT-grade' Abs and a micro sonication device, but also broader brand recognition within the field and valuable contacts with experts in the epigenetic field. Diagenode sales have been increasing annually by more than 25% the last years, reaching 18 million Euro consolidated this year. This is not only but also thanks to BLUEPRINT. Diagenode has hired about 40 employees during the last 4 years of the BLUEPRINT project. This may look small at the scale of the European or national economy but small progress, one day may make strong companies.

Another SME, Genomatix, has undertaken significant efforts to increase the awareness about the existence of its data portal and the other BLUEPRINT resources. This includes a series of webinars (that have also been recorded and made available via youtube), publication of data releases in our social media channels like facebook and linkedin. Genomatix is now expanding its product portfolio by adding data and analysis strategies for epigenomics derived from results and methods of the BLUEPRINT consortium. The data generated within the consortium is being integrated in existing meta-analysis-solutions (e.g. Genomatix Software Suite) creating new analysis products. In addition, Genomatix has integrated technological/computational methodologies implemented within BLUEPRINT into further product lines (GeneGrid) and offers analysis strategies for commercial services.

Notably, the invention of oxidative bisulfite sequencing (oxBS-seq) led to the establishment of the company Cambridge Epigenetix (<https://www.cambridge-epigenetix.com> ) who recently attracted substantial funding from Google Ventures. The development of single cell epigenomics techniques helped with the establishment of the Sanger/EBI Single Cell Genomics Centre (<http://www.sanger.ac.uk/science/collaboration/sanger-institute-ebi-single-cell-genomics-centre> ) . These exciting developments have recently been harnessed into an international initiative by which it is hoped to map all cells (and eventually their epigenomes) in the human body by single cell approaches (<http://www.sanger.ac.uk/news/view/international-human-cell-atlas-initiative> ) .

BLUEPRINT and compounds interfering with the regulators of epigenetics profiles

BLUEPRINT has identified new pathogenetic mechanisms for leukemogenesis in vivo (mice) and ex vivo. The novel causal epi-targets discovered might represent tools of relevance for further diagnostic and prognostic developments. In addition, innovative therapeutic strategies might be applied exploiting the knowledge developed. BLUEPRINT focused on the design, synthesis and biochemical/biological evaluation of chemical compounds with the ability to modulate epigenetic processes yielding a plethora of starting points for the development/optimization of therapeutic agents for the treatment of cancers. The development of a compound to a lead and finally an approved drug is a long process. BLUEPRINT's efforts are a very first step. Further developing of the compounds requires collaboration with SMEs such as Epi-C founded by a BLUEPRINT partner (www.epi-c.com) and Pharma companies. Many of the compounds have been patented and the inventors are in the process of raising interest to further develop the compounds.

BLUEPRINT and training, networking, communication and outreach

WP16 + WP17 have successfully contributed to create an efficient training & dissemination platform and ensure training of young scientists throughout the duration of the project. The lab exchange program as well as the workshops and the annual consortium meetings offered ample opportunity to PhD students, postdoctoral researchers as well as technicians to gain further insights in the research and technological development in epigenomics and related fields. Not only could they obtain technical experimental expertise, but they also gained networking experience thanks to the highly collaborative nature of the project. Especially, also the annual consortium meetings provided great opportunity to interact with the different disciplines represented by the 42 partners in the BLUEPRINT project. Also, the poster sessions that were organized contributed to the communication between junior and senior scientists. The final dissemination event in Brussels (September 2016) offered even a wider network with many representatives from the International Human Epigenome Consortium and high-level speakers that could be easily approached.

The establishment of a public website where a lot of information about the consortium can be found, the newsletters, videos and the many papers published in leading scientific journals and presented at national and international conferences have greatly enhanced the impact of the project: thus a large community has been able to learn about the achievements of the project and its contribution to more basic science as well as to insights into important biological pathways may in the end have impact on personalized medicine as it is currently developing.

The communication via newsletters and videos that have been produced by BLUEPRINT also provide the more general public access to the project results, illustrating the importance of what can be achieved when large consortia receive substantial funding from the European Commission.

List of Websites:

www.blueprint-epigenome.eu

Contact information BLUEPRINT coordinator

Prof. dr. Hendrik Stunnenberg

Radboud University

Department of Molecular Biology

PO Box 9101, NL-6500 HB Nijmegen, the Netherlands

Email: h.stunnenberg@ncmls.ru.nl

Phone: +31 24 3610524

BLUEPRINT project manager

Dr. Marion J.G. Bussemakers

Radboud University

Department of Molecular Biology

PO Box 9101, NL-6500 HB Nijmegen, the Netherlands

Email: m.bussemakers@ncmls.ru.nl

Phone: +31 24 3615157

Última actualización: 9 Marzo 2017

Permalink: <https://cordis.europa.eu/project/id/282510/reporting/es>

European Union, 2025

