Epigenomics

Anne C. Ferguson-Smith · John M. Greally ·
Robert A. Martienssen
Editors

Epigenomics

Springer

Editors
Dr. Anne C. Ferguson-Smith
Department of Physiology
Development and Neuroscience
University of Cambridge
Downing Street
Cambridge CB2 3EG
UK

Prof. John M. Greally
Albert Einstein College of
Medicine
1300 Marris Park Ave.
Bronx NY 10461
Ullman 921
USA

Dr. Robert A. Martienssen
Cold Spring Harbor Laboratory
I Bungtown Rd.
Cold Spring Harbor NY
1 1724-221 2
USA

ISBN: 978-1-4020-9186-5 e-ISBN: 978-1-4020-9187-2

DOI 10.1007/978-1-4020-9187-2

Library of Congress Control Number: 2008936212

Cover design: Boekhorst Design b.v.

Cover picture: From the article by Steven Henikoff, used with permission.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Now a popular and widely used term, "Epigenetics" has changed its meaning several times since "epigenesis" was debated by Greek philosophers as an alternative to existence of the homunculus (Aristotle, *On the Generation of Animals*). Waddington famously proposed that an "epigenetic landscape" underlies alternative cell fates, incorporating the idea, still novel at the time, that genes might contribute to this landscape. But it was the merging of embryology with non-Mendelian inheritance that ultimately led to the commonly accepted meaning of "Epigenetics" today. In this molecular age, we define epigenetics by what it is not – epigenetic changes are alterations in the hereditary material (usually chromosomes) that are not accompanied by changes in the DNA sequence. Certainly, epigenetics profoundly influences embryonic development, but equally it impacts chromosome organization, genome defense, heredity, gene expression and evolution.

Just as genetics allowed us the first glimpse of the genome in the form of linkage maps of genes and cytogenetic landmarks (Creighton and McClintock, 1931), so epigenetics allowed the first glimpse of the epigenome, with the discovery of widespread modifications of chromosomal material, including DNA and histones, associated with epigenetic regulation. Immunocytochemistry revealed striking correlations between, for example, histone H4 acetylation and dosage compensation, while molecular biology revealed correlations between, for example, DNA methylation and transposon inactivation. It was quickly realized that many of these modifications were conserved among most if not all eukaryotes, and were associated with a wide diversity of "epigenetic" phenomena. But it was only when the first whole genome sequences became available that the concept of the "epigenome" took hold. Just as DNA sequence can be aligned with the chromosome, so epigenetic modifications can be aligned with the DNA sequence. The epigenome has emerged at nucleosome and nucleotide resolution from genome profiling using high density programmable microarrays, chromatin immunoprecipitation and next generation DNA sequencing, as well as analytical procedures to detect and display significant associations. This volume commences with a section describing the current technologies employed in mapping epigenomes and the challenges associated with the analysis and visualization of these large datasets within their genomic context. The impact of the technology cannot be underestimated and as such, is recognised in many of the subsequent contributions.

In subsequent chapters, the current understanding of these epigenetic maps is reviewed within the context of epigenetic phenomena in eukaryotes. The role that model organisms have played in unveiling key conserved mechanisms is a general theme, as is the idea of an epigenetic "code", comprising histone and DNA modifications, guided by modifying enzymes, and in some cases RNA interference, and interpreted by histone and DNA binding proteins that recognize these modifications and signal their downstream effects. The relationship between epigenetic states, non-protein coding transcripts and sub-nuclear localization is also explored. The impact of these mechanisms on gene activity and repression, developmental memory, imprinting, X inactivation, genome defense and chromosome organization is reviewed in animals, plants and fungi. We also review the first glimpses of the human epigenome in differentiating cells and relate these mechanisms to human disease and development.

In the conclusion of his chapter describing the relevance of variant histones to epigenome function, Steven Henikoff reminds us that, since the completion of the draft human genome sequence, the 21st century is often referred to as the "post-genomics era". The combination of major advances in our understanding of model epigenetic processes and remarkable technological advances, as illustrated in this volume, have ushered in what he suggests might rather be considered an "epigenomics era". Whatever it is called, the integration of the epigenetic components required to consolidate DNA into its highly regulated chromatin context, is now recognized as profoundly important for understanding the functions of normal and compromised genomes. The development of effective new tools having the ability to modulate epigenetic states and influence genome function in clinical contexts, seems a realistic goal. We have witnessed the birth of a major new discipline in genetics that has taken us inside chromosomes to unfold multiple highly-regulated dynamic epigenetic landscapes within which a genome is parceled and to which it responds. We are grateful to all the authors of this volume for sharing their research, views and ideas, and for revealing these landscapes.

United Kingdom — Anne C. Ferguson-Smith
USA — John Greally
USA — Rob Martienssen

Contents

Contributors

R.M. Amasino Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

E. Ballestar Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Catalan Institute of Oncology (ICO-IDIBELL), 08907 L'Hospitalet de Llobregat, Barcelona, Spain, eballestar@iconcologia.net

B.E. Bernstein Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142, USA; Molecular Pathology Unit and Center for Cancer Research, MGH, Charlestown, MA 02129, USA; Department of Pathology, Harvard Medical School, Boston, MA 02115, USA, Bernstein.Bradley@mgh.harvard.edu

J.A. Birchler Division of Biological Sciences, Tucker Hall, University of Missouri, Columbia, MO 65211, USA, BirchlerJ@Missouri.edu

E. Birney European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, birney@ebi.ac.uk

R.M. Blumenthal Department of Medical Microbiology & Immunology, and Program in Bioinformatics & Proteomics/Genomics, University of Toledo Health Science Campus, Toledo, OH 43614-2598, USA

A.B. Brinkman Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, The Netherlands, a.brinkman@ncmls.ru.nl

X. Cheng Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA, xcheng@emory.edu

A. Dean Laboratory of Cellular and Developmental Biology, NIDDK, NIH, Building 50, Room 3154, 50 South Drive, MSC 8028, Bethesda, MD 20892, USA, anndean@helix.nih.gov

R.W. Doerge Department of Statistics, Purdue University, 150 North University Street, West Lafayette, IN 47907 USA

M.R. Doyle Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

M. Esteller Cancer Epigenetics Group, Cancer Epigenetics and Biology Programme (PEBC), Catalan Institute of Oncology (ICO-IDIBELL), 08907 L'Hospitalet de Llobregat, Barcelona, Spain, mesteller@iconcologia.net

C.R. Farthing Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, UK

A.P. Feinberg Epigenetics Center, Institute for Basic Biomedical Sciences, and the Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA, afeinberg@jhu.edu

A.C. Ferguson-Smith Professor of Developmental Genetics, Department of Physiology Development and Neuroscience, University of Cambridge, Physiology Building, Downing Street, Cambridge CB2 3EG, UK, afsmith@mole.bio.cam.ac.uk

G. Ficz Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, UK

S. Fiering Department of Microbiology/Immunology and Genetics, Dartmouth Medical School, Rubin 622, DHMC, Lebanon, NH 03756, USA, fiering@dartmouth.edu

P. Flicek European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, flicek@ebi.ac.uk

J.M. Greally Einstein Center for Epigenomics, Department of Genetics, Albert Einstein College of Medicine, Bronx, NY USA

I.R. Henderson Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK, irh25@cam.ac.uk

J.B. Hollick University of California, Department of Plant and Microbial Biology, 111 Koshland Hall #3102, Berkeley CA, USA, hollick@nature.berkeley.edu

M. Jonathan Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK

C. Ladd-Acosta Epigenetics Center Institute for Basic Biomedical Sciences, and the Department of Medicine, Johns Hopkins University School of Medicine, Battimore, ND 21205, USA, claddac1@jhmi.edu

J.T. Lee Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Simches 6.624, 185 Cambridge St., Boston, MA 02114, USA, lee@molbio.mgh.harvard.edu

S. Locke Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor NY 11724, USA, locke@cshl.edu

R. Martienssen Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor NY 11724, USA. martiens@cshl.edu

K.R. McEwen Department of Physiology Development and Neuroscience, University of Cambridge, Physiology Building, Downing Street, Cambridge CB2 3EG, UK, krm29@cam.ac.uk

A. Meissner Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142, USA; Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA, alex@broad.mit.edu

A. Petronis Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, 250 College street, Toronto ON M5T 1R8, Canada, arturas_petronis@camh.net

V. Pirrotta Department of Molecular Biology and Biochemistry, Rutgers University, 604 Allison Road, Piscataway, NJ 08854, USA, pirrotta@biology.rutgers.edu

W. Reik Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, UK; Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK

M.R. Rountree Institute of Molecular Biology, University of Oregon, Eugene, OR 97403; USA, mrountree@molbio.uoregon.edu

Y.B. Schwartz Department of Molecular Biology and Biochemistry, Rutgers University, 604 Allison Road, Piscataway, NJ 08854, USA

E.U. Selker Institute of Molecular Biology, University of Oregon, Eugene, OR 97403; USA, selker@molbio.uoregon.edu

S. Shibata Department of Stem Cell Biology, Graduate School of Medical Sciences, Kanazawa University, 13-1 Takara-machi, Kanazawa, Ishikawa 920-8640, Japan, shinwa@med.kanazawa-u.ac.jp

N. Springer University of Minnesota, Department of Plant Biology, 250 Biosciences Center, 1445 Gortner Ave., Saint Paul MN 55105, springer@umn.edu

S. Henikoff Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, Washington 98109, USA, steveh@fhcrc.org

H.G. Stunnenberg Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, The Netherlands, h.stunnenberg@ncmls.ru.nl

J.M.A. Turner Division of Stem Cell Research and Developmental Genetics, MRC National Institute for Medical Research, London NW7 1AA, UK

C. Wadelius Department of genetics and pathology, Rudbeck Laboratory, Uppsala University, SE-75185 Uppsala, Sweden, Claes.Wadelius@genpat.uu.se

S.-Y. Yoo Department of Statistics, Purdue University, 150 North University Street, West Lafayette, IN 47907 USA

Part I
Epigenomic Technologies and Analytical Approaches

Strategies for Epigenome Analysis

A.B. Brinkman and H.G. Stunnenberg

Abstract The eukaryotic genome is packaged into nucleosomes, which form the basal unit of chromatin, the physiological form of DNA within the nucleus. Apart from its function in compacting the immense nuclear DNA molecules, chromatin also serves as a platform onto which multiple signalling pathways converge to cooperate in determining the expression status of mRNAs and other (non-coding) RNA molecules. Epigenetic profile analysis aims to determine what changes on the nucleosomes cooperate to establish and maintain DNA sequence-independent heritable traits such as those determining cell identity.

Keywords Profiling · Histone modifications · DNA methylation · ChIP-on-chip · ChIP-seq

1 Introduction

Epigenetic changes can occur on all building blocks of the nucleosome to ultimately constitute the epigenome. They include (i) post-translational histone modifications; (ii) incorporation of histone variants; (iii) remodelling of the DNA-histone interaction; (iv) methylation of DNA; (v) association with transcription (co)factors; (vi) local changes in nucleosome density; (vii) changes in long-range chromatin interactions and compaction.

In this chapter we will focus on the different methods that are available to profile epigenetic changes. In particular, we will focus on methods that are suitable for high-throughput –or whole-genome– profiling, and we will discuss logical strategies to approach such profiling and consider aspects that we have found to be of importance.

The availability of highly specific antibodies for chromatin-modifying proteins and their resulting modifications, in combination with chromatin immunoprecipitation and immunofluorescence-based techniques has enabled mapping of epigenetic

A.B. Brinkman (✉)
Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, The Netherlands

A.C. Ferguson-Smith et al. (eds.), *Epigenomics,* DOI 10.1007/978-1-4020-9187-2_1,

changes. This has generated a wealth of information that drastically changed the concept of expression regulation from a transcription factor-based view into a system in which transcription factors and chromatin interact and cooperate. The advent of genomic tiling microarrays –and recently, massive parallel sequencing– have enabled global and whole-genome epigenetic profiling studies. Such large-scale analyses will allow the generation of new insights into how chromatin is utilized and shaped according to the cell's necessities, ultimately defining its epigenome.

The concepts outlined above are of particular importance for the development of novel treatment strategies for cancer. It has become clear that epigenetic alterations act as "surrogates" for genetic changes in cancer, and as epigenetic alterations are mitotically heritable, they play the same roles and undergo the same selective processes as genetic alterations in the development of cancer. However, epigenetic alterations are more feasible to be reversed than genetic aberrations, thus providing opportunities for therapeutic intervention.

2 ChIP

In 1988 a key publication from the Varshavsky laboratory (Solomon et al. 1988) demonstrated the use of chromatin immunoprecipitation (or ChIP) to study regulation of the Hsp70 genes in Drosophila. Modifications of the original protocol have been developed since, but the basic protocol includes the *in vivo* cross-linking of chromatin-bound proteins using formaldehyde, to 'freeze' the *in vivo* situation, and the generation of short random fragments of this chromatin using sonication. The obtained cross-linked chromatin fragments are subsequently used in an immunoprecipitation step using antibodies directed against the protein or histone modification of interest. After immunoprecipitation the isolated antibody-chromatin-complexes are decrosslinked and the DNA is purified. In parallel, the input material –or non-immunoprecipitated chromatin– is decrosslinked and DNA is purified likewise. Both fractions are subjected to quantitative PCR using primers specific for the genomic region of interest. In this way the magnitude of enrichment by ChIP is determined at a specific genomic position. Depending on the antibody, ChIP allows for profiling chromatin-associated factors, histone modifications, histone variants as well as local nucleosome density.

3 ChIP-on-Chip

Combining ChIP with genomic tiling array hybridization (CHIP-on-chip) or massive-parallel sequencing (ChIP-seq) instead of gene-specific PCR allows one to analyze larger genomic regions or even the whole genome. Several steps within such a procedure are of critical importance for a successful profiling experiment. In Fig. 1 we present a generalized strategy that can be followed to establish a ChIP-based profiling approach – the Roman numerals depict strategic points shown in Fig. 1. The strategy starts with a factor of interest, which is assumed to be a chromatin-associated protein (I).

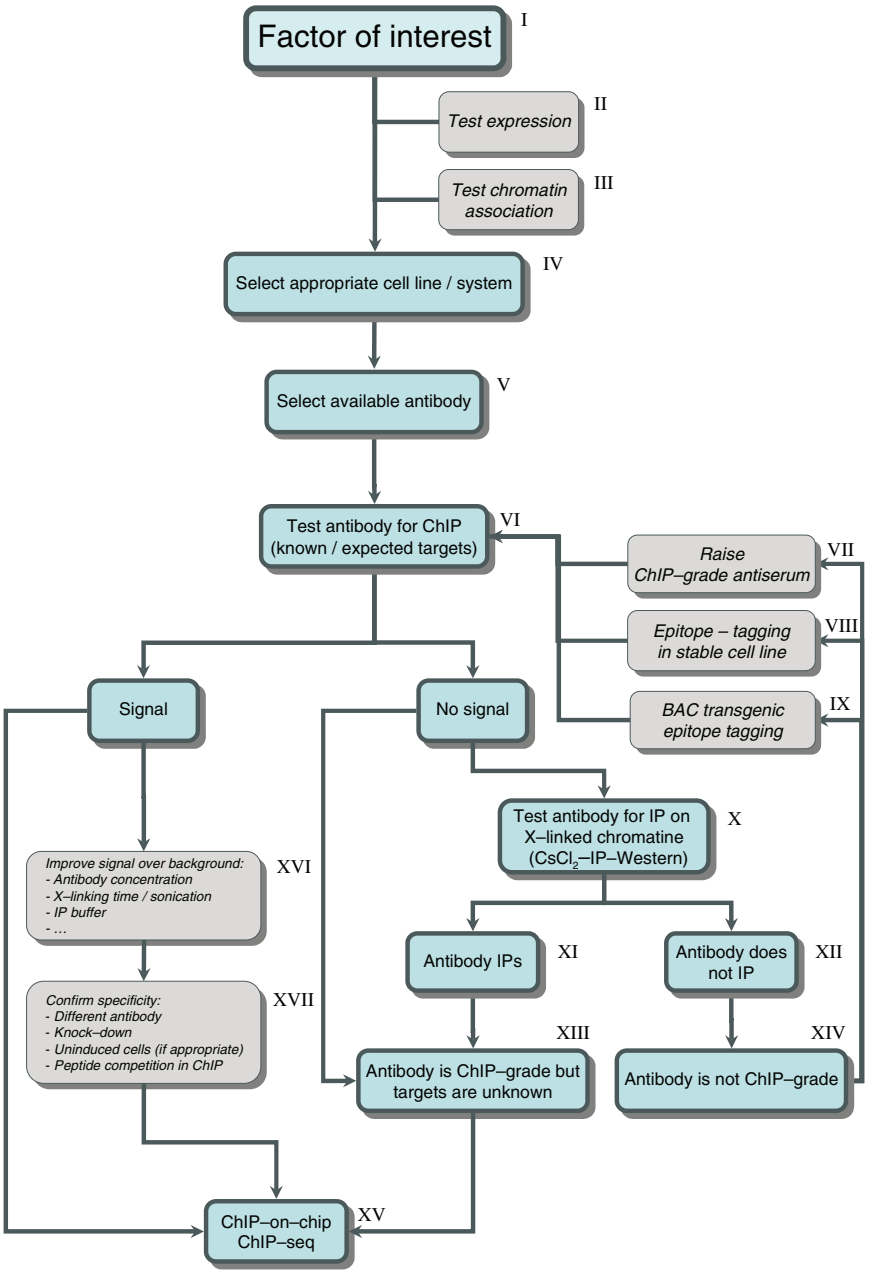

Fig. 1 Generalized strategy for a ChIP-based profiling. See text for further explanation

Since a ChIP experiment requires the availability of a relatively large amount of cells, the initial use of cultured cells is inevitable, but whenever conditions are optimized it should be possible to use primary cells or tissue-derived cells. The choice for a particular cell line (IV) most often depends on whether the protein is expressed or not (II). In addition, it may be relevant in this stage to confirm its chromatin-association (III). If an antibody is available, this can be done by cesium-chloride ($CsCl_2$)-density gradient centrifugation of formaldehyde-crosslinked material. Western blot analysis of the fraction containing crosslinked DNA-protein complexes (after decrosslinking) will reveal whether the protein of interest is chromatin-associated.

The most critical issue of ChIP is the availability of a high-quality ChIP-grade antibody (V). Commercial antibodies are available for many chromatin-associated proteins, and even though they may be specific in Western analysis or efficient in a normal immunoprecipitation, most of them are not ChIP-(profiling)-grade, even when marketed as such. A large-scale ChIP-profiling experiment requires substantial enrichment over background (a control region that is not enriched), even more than for a single ChIP-qPCR. Clearly, only a few ChIP-grade antibodies are also ChIP-profiling grade. In our experience, enrichment above 20-fold over background will give profiles of reasonable to good quality.

Whenever an antibody needs to be produced against the protein of interest (VII), generating a polyclonal antiserum against several different peptides (or a protein domain) will be preferred. A polyclonal antibody is directed against a range of different epitopes, which increases the chance for obtaining one that is ChIP-grade. In addition, the performance of a ChIP-grade polyclonal antibody will usually improve after affinity purification. An important constraint on immunogen selection is to avoid lysine residues within the amino acid sequence. Lysine residues are the primary targets for formaldehyde cross-linking, and epitopes containing lysines will be (at least partly) functionally destroyed by formaldehyde cross-linking. Most commonly used monoclonal antibodies against epitope-tags are not ChIP-grade for this reason (see Table 1). Although they have been successfully used by others, we have never reached substantial enrichment using MYC, and inconsistent results were obtained using FLAG. HA and HSV provide better alternatives because they lack lysines. We found that they are ChIP-grade, although

Table 1 Epitope tag sequences

Epitope tag	Amino acid sequence	ChIP-grade
MYC	EQKLISEEDL	-
FLAG	DYKDDDDK	+/–
VSV-G	YTDIEMNRLGK	*Not tested*
V5	GKPIPNPLLGLDST	*Not tested*
HA	YPYDVPDYA	+
HSV	QPELAPEDPED	+
TY-1	EVHTNQDPLD	+
2xTY-1	EVHTNQDPLDAEVHTNQDPLD	++
ERα	SLQKYYITGEAEGFPATV	++

the level of enrichment is not outstanding. We have recently exploited the use of two different universal epitope tags: TY-1 (and 2xTY-1) and ERα. For both epitopes monoclonal antibodies are available. Although a single lysine is present within the periphery of the ERα-tag, both tags show excellent performance in ChIP and also in ChIP-on-chip and ChIP-seq experiments. The use of the ERα-tag requires that ERα is not expressed endogenously or that its ligand is depleted from the culture medium, since we found only very low chromatin association of endogenous ERα in the latter case. Alternatively, a ChIP-reChIP approach using first TY-1 followed by ERα could be used, or vice versa.

Whenever epitope tagging is used one has to make sure that the tagged protein is expressed at (near-)endogenous levels. In our hands, ChIP profiles of overexpressed proteins contain many sites of enrichment that are otherwise absent in profiles of the endogenously expressed protein. For generating a stable cell line expressing a tagged protein (VIII), multiple independent clones should be screened for ones with near-endogenous expression levels.

An exploratory ChIP-qPCR experiment can be performed if one or more genomic target sites are known or can be predicted (VI). The absence of any signal over background (a genomic control region) could either mean that the antibody is not ChIP-grade (XIV) or that the wrong genomic targets were selected for qPCR (XIII). This can be analyzed by performing an immunoprecipitation on crosslinked chromatin that has been purified following cesium-chloride ($CsCl_2$)-density gradient centrifugation (X), which is used to remove non-crosslinked proteins. If the antibody precipitates the protein of interest from the crosslinked fraction –that is, a signal is obtained in Western blot– the antibody is most likely ChIP-grade, and one could proceed with a pilot ChIP-on-chip or ChIP-seq experiment to detect targets (XV). Whenever the exploratory ChIP experiment reveals sufficient enrichment over background, the ChIP protocol may need optimization (XVI). This may include antibody concentration, crosslinking and/or sonication time, and composition of the immunoprecipitation buffer, such as SDS concentration. Whenever the protein of interest is expected to be chromatin-associated through secondary interactions, additional protein-protein crosslinking may improve the level of enrichment (Zeng et al. 2006). In some cases it may be necessary to determine the specificity of the antibody in ChIP (XVII). This could involve the use of different antibodies against the same protein, RNAi-mediated knock-down of the protein, or performing a ChIP experiment using untreated/uninduced cells (e.g. in the case of many nuclear hormone receptors).

For ChIP-on-chip profiling, a number of additional (post-ChIP) aspects are of importance. First, ChIP DNA must be amplified to obtain sufficiently large amounts for labelling and array hybridization (typically several micrograms per microarray hybridization). Depending on the antibody the amount of ChIP DNA is within the nanogram range, and at least a 1000-fold amplification is necessary. Commonly used DNA amplification methods include linker-mediated PCR (LM-PCR), T7 polymerase-based linear amplification, and whole-genome amplification (WGA). In LM-PCR, linkers are ligated to both ends of the ChIP DNA fragments, which allows PCR amplification using universal primers (Ren 2000). In T7-based linear amplification, oligo-dT tails of defined length are added to the 3' termini of the DNA

fragments, to which a T7-promoter sequence can be annealed. After conversion into dsDNA, the resulting DNA is used as a template for *in vitro* transcription using the T7 RNA polymerase. The transcribed RNA molecules are subsequently converted to dsDNA (Liu et al. 2003). WGA involves the proprietary GenomePlex® amplification technique (Sigma-Aldrich, St. Louis, MO, USA) in which DNA fragments are primed to generate a library of DNA fragments with defined 3' and 5' termini. This library is then replicated using linear amplification in the initial stages, followed by a limited round of geometric amplifications. We have used LM-PCR with some success for ChIPs against transcription factors or locally enriched histone marks like H3K4 tri-methylation and H3/H4 acetylation. However, the same procedure was not successful for histone marks that are distributed over larger regions, like H3K9 and H3K27 tri-methylation. In addition, we found that on complex samples (like human genomic DNA) LM-PCR amplification introduces a considerable bias. This is shown in Fig. 2. Upon hybridization of unamplified DNA, LM-PCR-amplified DNA and T7-amplified DNA, the obtained probe intensities were compared using scatter plots. Comparison of two unamplified genomic DNA samples shows that their probe intensities correlate well (Fig. 2A, R^2=0.85). When unamplified DNA is compared with the same DNA after LM-PCR amplification the correlation decreases dramatically (Fig. 2B, R^2= 0.52). In contrast, the correlation is maintained to a much greater extent when unamplified DNA is compared with T7 amplified DNA (Fig. 2C, R^2=0.75). This shows that T7 amplification is superior to LM-PCR, even upon multiple rounds of amplification (not shown). WGA has also been shown to eliminate bias problems (O'Geen et al. 2006), although we have not applied this technique in our laboratory.

Second, the choice of microarray platform is an important determinant for obtaining confident signal to noise ratios. Three different platforms have been mostly used for ChIP-on-chip analysis: Affymetrix, Nimblegen and Agilent. The main differences between these platforms are probe length, probe spacing and probe density. Whereas probe length ranges from 25 to 75 bp, probe densities currently vary from 2.4×10^5 to 6.5×10^6 probes per array. We have experienced that Nimblegen arrays with 50-bp probes provide good probe density and signal to noise ratios.

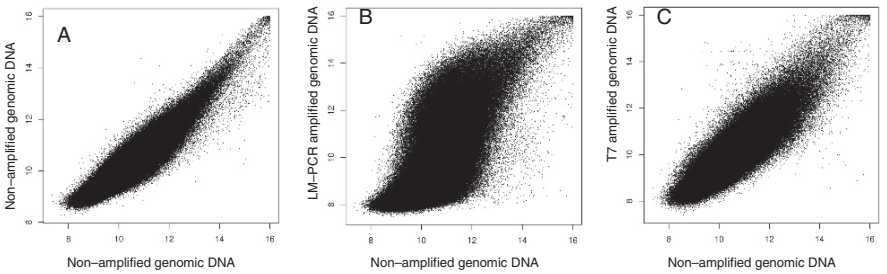

Fig. 2 Comparison of two random amplification methods used in ChIP-on-chip. Human non-amplified genomic DNA was hybridized against (**A**), human non-amplified DNA as a control; (**B**), human genomic DNA amplified using LM-PCR; (**C**), human genomic DNA amplified using T7-based linear amplification. Log^2 probe intensities are plotted

The 60-mer probe design of Agilent arrays shows better signal to noise ratios, but the number of probes per array is lower. Although it is possible to interrogate the whole human or mouse genome using tiling arrays, more directed analyses may be performed using promoter-tiling arrays or dedicated custom arrays containing selected genomic regions.

Third, data interpretation mostly starts with peak-calling: the assignment of enriched regions. Several independently developed peak recognition algorithms are available, e.g. MPeak (Zheng et al. 2007), TileMap (Ji and Wong 2005). In addition, array manufacturers provide similar software tools, e.g. Nimblegen's SignalMap. MPeak is a model-based method that assumes a triangular shape for enriched regions, while Tilemap uses test-statistics to define regions of enrichment and calculates statistically significant enriched regions between a sample and control dataset. Because of the different methodologies, the performance of these algorithms will vary depending on the factor to be profiled. For instance, MPeak may not be the method of choice in profiling H3K27 tri-methylation because of the presence of many non-triangular enrichment patterns. In contrast, for transcription factor profiling all methods mentioned have been used successfully in our laboratory. However, all algorithms will detect distinct but largely overlapping peak sets. Selecting peaks that are commonly found by all methods will significantly decrease the number of false-positives. The false-positive rate has to be determined empirically using ChIP-qPCR analysis of a number of randomly selected peaks (>50 for a genome-wide analysis). Such validation is of critical importance, because in our experience the number of detected peaks is subject to considerable variation just because of the selected method and its peak-detection settings.

4 ChIP-Seq

ChIP-seq represents a recent advance in epigenetic profiling. It involves direct massive parallel sequencing of individual DNA molecules obtained in a ChIP experiment. Counting of the number of sequence reads within a specific genomic region is proportional to the local level of enrichment. To date, a number of studies have described ChIP-seq approaches to profile histone-modifications (Mikkelsen et al. 2007, Barski et al. 2007) or transcription factors (Robertson et al. 2007, Johnson et al. 2007), using Solexa sequencing technology (Illumina Inc.). The technology starts with the generation of clusters of about 1000 identical DNA fragments, each originating from a single DNA molecule. Cluster generation is achieved by solid-phase DNA amplification onto a glass-surface. Up to 50 million clusters can be generated and sequenced in parallel, each producing a sequence read of 36 bases. Other platforms for massive-parallel sequencing are 454 Life Sciences and SOLiD. For ChIP, the advantages of massive-parallel sequencing over microarray hybridization are the increased throughput and dramatically increased signal to noise ratio. In Fig. 3, a comparison between a ChIP-on-chip profile and a ChIP-seq profile is shown. The screenshot shows the enrichment of a transcription factor at one of its target genes. While the enrichment of the ChIP-on-chip profile is displayed as the

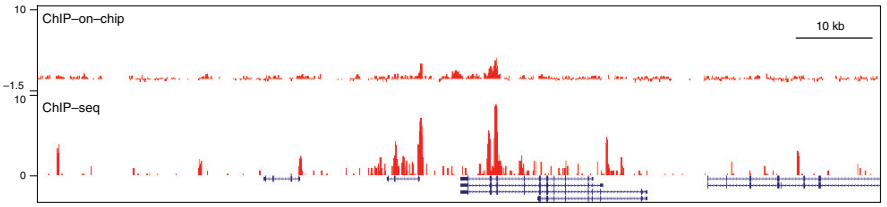

Fig. 3 Screenshot showing a comparison between a ChIP-on-chip profile (top panel, $\log^2$ of ChIP/input ratio) and a ChIP-seq profile (bottom panel, $\log^2$ of the number of sequence reads counted per 100 bp window) generated from the same cell line. The profile shows enrichment of a transcription factor at one of its target genes. Annotated genes are indicated at the bottom

ratio of ChIP DNA over non-immunoprecipitated (input) DNA, enrichment in the ChIP-seq profile is displayed as the number of sequence reads counted per 100-bp window within the ChIP DNA sample. The difference between enrichments represented in this way is over 40-fold: while ChIP-on-chip enrichment is maximally 2.9 (log2), ChIP-seq enrichment is 8.3 (log2). In addition, the average background enrichment within the regions flanking these peaks is clearly lower in the CHIP-seq profile. The increased signal to noise ratio obtained using ChIP-seq allows assignment of peaks with high confidence, and permits identification of binding sites that are not evident from the ChIP-on-chip profile. A single analysis-run (one out of eight lanes of a flowcell) using Solexa sequencing is able to provide genome-wide coverage for a ChIP experiment, although this is of course dependent on the factor to be profiled and the specificity of the used antibody. More sequence reads may be necessary to optimize the signal to noise ratio. Profiling of a widespread histone mark like H3K27 tri-methylation requires more sequence reads than profiling a transcription factor that has a limited number of binding sites throughout the genome, because of the difference in complexity of the ChIP DNA sample.

5 DNA Methylation

Within the last decades it has become clear that gene-silencing through DNA methylation involves multiple steps that cooperate to establish a gene-repressive state. Whereas the role of DNA methylation in gene-body and intergenic sequence is much less clear, promoter methylation has been strongly linked to gene-silencing. Roughly half of the human promoters are located within a CpG-island, a region in which the CpG dinucleotide motif is overrepresented relative to the average genome (Gardiner-Garden and Frommer 1987). While virtually all cytosine bases of a CpG are methylated, those within a CpG island are generally maintained methylation-free. However, within cancer cells a number of genes become inactivated through hypermethylation of their promoter CpG-islands. This phenomenon offers great potential towards the identification of cancer-specific epigenetic aberrations termed differentially methylated regions (DMRs). A model for DNA methylation-induced repression is shown in Fig. 4. DNA methyltransferases (DNMTs) are recruited

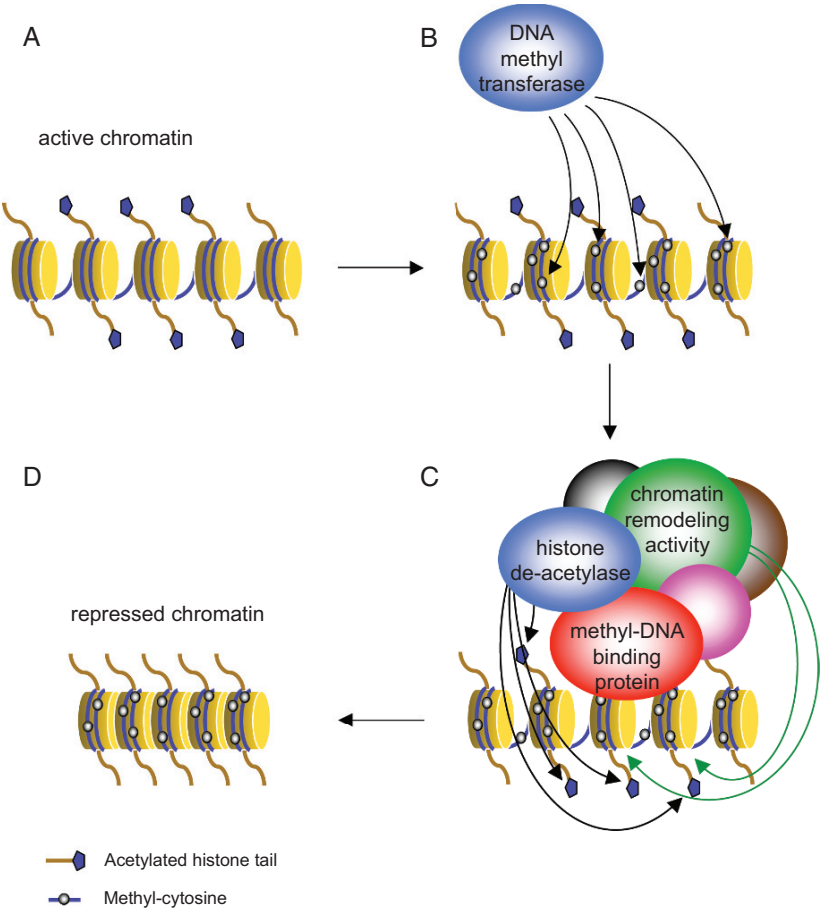

Fig. 4 General model for gene repression through DNA methylation. (**A**), Active chromatin, schematically displayed as open and acetylated. (**B**), DNA methyl transferases (DNMTs) are recruited to locally increase DNA methylation levels. (**C**), The methyl-DNA mark acts as a target for methyl-DNA binding proteins that specifically recognize and bind this mark. Methyl-DNA binding proteins are mostly part of multi-protein complexes that contain chromatin-modifying activity represented by histone-deacetylases (HDACs) and chromatin remodelers (e.g. MBD2-NuRD, (Le Guezennec et al. 2006, Feng and Zhang 2001, Ng et al. 1999)). (**D**), The recruited chromatin-modifying activities establish a repressive chromatin configuration, schematically shown as deacetylated and compacted chromatin

by currently poorly defined mechanism(s), and increase local DNA methylation. The DNA methylation marks act as targets for methyl-DNA binding proteins (MBPs) that specifically recognize and bind such sites. Some of these MBPs act as direct repressors through the presence of transcriptional repression domains, but their major repressive activity is exerted via associated proteins, including histone deacetylases (HDACs), chromatin remodelers, and histone methyltransferases (HMTs), which cooperate to establish a repressed chromatin state. HDACs remove

acetyl groups from histone tails, which is closely associated with repressive chromatin. The various proteins with chromatin-modifying activities represent attractive targets for pharmacological inhibition of repression to reactivate repressed target genes. Indeed, a growing number of HDAC inhibitors are in development or are already in phase I-II clinical trials and effectively induce differentiation, growth arrest and/or apoptosis in tumorigenic cells, whereas normal cells appear to be less sensitive.

6 Profiling of DNA Methylation

In contrast to profiling of histone modifications, transcription factors and histone variants, for which ChIP is the most widely used method, profiling of DNA methylation is possible through a wide variety of techniques (at least 20 have been described). They differ in the resolution of methylation mapping, the ability to give qualitative rather than quantitative measurements, and in their potential to be used in global rather than gene-specific analysis. In this overview we will only discuss a selection of the available methods, focusing on the most widely used methods and those that are amendable for large-scale analysis (see also Chapter "Sequencing the Epigenome" by Meissner and Bernstein).

Methods to profile DNA methylation patterns can be roughly divided in three categories based on the principle of distinguishing methylated from unmethylated DNA. These include (i) bisulphite conversion, (ii) digestion with methylation sensitive restriction enzymes, and (iii) capture of methylated DNA fragments using a recombinant methyl-DNA binding protein domain (MBD) or a monoclonal anti-5-methyl-cytosine antibody.

6.1 Bisulphite Conversion Based Methods

Bisulphite sequencing (Frommer et al. 1992) provides quantitative information on methylation within the analyzed cell population at the basepair level. It involves chemical conversion of only the unmethylated cytosines to thymidines, followed by PCR amplification, cloning and sequencing of the obtained fragments. The latter steps however make it a relatively labour-intensive technique that can only be scaled up using a pipeline-infrastructure (Eckhardt et al. 2006). The design of primers that specifically amplify bisulphite-converted sequences puts constraints on the choice of genomic regions to be targeted, and only up to 250 bp-fragments can be analyzed at the time. This makes whole-genome coverage almost impossible. However, for validation purposes it is regarded as the 'golden standard'.

Bisulphite pyrosequencing (Colella et al. 2003) is a sequencing-by-synthesis technology in which nucleotide incorporation is measured using a luminescence signal. This is generated by pyrophosphate release through a cascade consisting of four enzymes. Bisulphite-converted DNA is subject to PCR amplification in which one of the two primers is biotinylated. The amplified DNA is immobilized on streptavidin beads, converted to ssDNA and sequenced. This method allows for

quantitative analysis, but is limited to 25–30 bases in length, although modifications of the protocol have been shown to generate sequence reads of up to 75 bp (Dupont et al. 2004).

In Methylation-Specific PCR (MSP) (Herman et al. 1996) bisulphite conversion is combined with amplification by primer pairs that specifically amplify regions in which CpGs are methylated or unmethylated. MSP is a very fast and sensitive technique, although it should be regarded as a qualitative rather than a quantitative technique. It can only be scaled up to medium-throughput because it requires pre-selection of regions of interest for PCR amplification.

In COmbined Bisulphite Restriction Analysis (COBRA) (Xiong and Laird 1997) DNA is converted using bisulphite and amplified using gene-specific PCR. The amplified fragments are digested with *BstUI*, which recognition site is destroyed whenever overlapping CpGs are unmethylated. The digested DNA is subsequently analyzed using Southern blotting to determine the cut/uncut ratio. COBRA provides quantitative information on methylation and can be used for validation purposes, but like MSP it can only be scaled up to medium-throughput.

MethyLight (Eads et al. 2000) is a sensitive technology in which genomic DNA is converted using bisulphite. Subsequently, selected genomic regions are amplified using TaqMan amplification technology that requires two amplification primers flanking a fluorescent probe. MethyLight can be used in different ways: methylation-dependent sequence discrimination at the *probe* level (quantitative mode), discrimination at the *amplification primer* level (qualitative, similar to MSP), discrimination at both the *primer and probe* levels (semi-quantitative mode), or no methylation-dependent discrimination (control mode). Like MSP, MethyLight requires only low amounts of DNA of modest quality, which allows the analysis of material from various sources (e.g. fixed/fragmented patient material). MethyLight is a medium-throughput technology that requires target pre-selection, but it is more flexible than MSP.

The GoldenGate assay for methylation in combination with BeadArray technology (Ilumina Inc.) concomitantly interrogates over 1,500 individual CpGs in 96 samples (Laird, Illumina Application Note, www.illumina.com). The method is based on two specific probe pairs for a specific CpG site: one pair for the methylated state and one pair for the unmethylated state. The pairs are annealed to bisulphite-converted DNA, and the gap that remains between the oligo pair is extended and ligated, creating a PCR template that can be amplified using fluorescent primers. The resulting fragments are subsequently hybridized to beads that contain a known, probe-pair specific sequence tag that allows the fluorescent signal to be assigned to the specific CpG site in question. For every CpG site a two-channel fluorescent signal reports the methylated to unmethylated ratio. The GoldenGate assay is quantitative and amendable to high-throughput analysis.

6.2 Restriction Enzyme Based Methods

In Differential Methylation Hybridization (DMH) (Huang et al. 1999) adapters are ligated to MseI-fragmented genomic DNA, which are then cut with a methylation-

sensitive restriction enzyme like *BstUI* or *McrBC*, although other enzyme combinations have been used as well (Nouzova et al. 2004). The uncut DNA is then PCR-amplified and hybridized onto microarrays. DMH compares samples with a reference sample so as to identify differentially methylated regions. The use of microarrays allows for high-throughput analysis. DMH is semi-quantitative.

HpaII tiny fragment Enrichment by Ligation-mediated PCR (HELP) (Khulan et al. 2006) involves digestion of genomic DNA by HpaII and in parallel its methylation-insensitive isoschizomer *MspI*. The two different populations of fragments are amplified and size-selected using LM-PCR and hybridized on genomic tiling microarrays. HELP allows both intragenomic profiling and intergenomic comparisons of DNA methylation and is quantitative.

6.3 Capture of Methylated DNA Fragments

Methylated-CpG Island Recovery Assay (MIRA) (Rauch et al. 2006, Rauch and Pfeifer 2005) is an affinity capture assay that is similar to those described in earlier studies in which the MBD domain of MeCP2 was used to capture methylated DNA fragments (Cross et al. 1994, Shiraishi et al. 1999). MIRA is a modification of this initial capturing method, and utilizes full-length GST-MBD2b complexed with His-MBD3L1. The latter stimulates the methyl-DNA binding affinity of MBD2. Genomic DNA is fragmented using *MseI*, linkers are ligated to the generated ends and methylated DNA is captured by the MBD2b/MBD3L1 protein complex. After washing, methylated DNA fragments are purified and analyzed using microarrays. Both CpG-islands and genomic tiling arrays have been used (Rauch et al. 2008). MIRA is a high-throughput method that could potentially be used to profile methylation along the complete genome. It is presumably unbiased, since no DNA sequence bias for MBD2's methyl-DNA binding activity has been reported, as opposed to that of MeCP2 (Klose et al. 2005). It is currently unknown whether the method is semi-quantitative or quantitative.

Methylated DNA ImmunoPrecipitation (MeDIP) is a capture assay that utilizes a monoclonal antibody that recognizes 5-methyl-cytosine. Despite the potential of various DNA methylation analysis methods for large-scale analysis, generation of whole-genome DNA methylation profiles had not been performed until the MeDIP technique became available. Studies using human cancer cells (Weber et al. 2005, 2007) and *Arabidopsis thaliana* (Zilberman et al. 2007, Zhang et al. 2006) demonstrated that MeDIP, when combined with whole-genome tiling arrays, provides a powerful solution for whole-genome methylation profiling. The application of MeDIP was first published by the Schübeler lab in 2005 (Weber et al. 2005). It includes the shearing of high-molecular-weight genomic DNA into 300–500 bp fragments using sonication, and subsequent immunoprecipitation using an anti-5-methyl-cytosine antibody. The captured DNA fragments are then analyzed using genomic tiling arrays. The basic principle of MeDIP is very similar to MIRA, but where MIRA uses recombinant proteins that recognize dsDNA, MeDIP uses an antibody that requires ssDNA. Both methods are ideally suited for whole-genome

methylation profiling. We have extensively evaluated and optimized the MeDIP-on-chip approach, and we have no evidence for any sequence bias in MeDIP. A common problem with methyl-DNA (immuno)capturing approaches is that the efficiency with which DNA is captured is dependent on the number of methylated CpGs within the fragment. This "CpG-density factor" complicates quantification of methylation, assignment of an average "no-methylation" value, and cross-comparison between profiles of different samples. Our results indicate that MeDIP can be used in a quantitative manner by addition of exogenous control DNAs to the capturing step that serves normalization purposes. An additional advantage of capture over restriction-based methods is that it can be used to profile patient DNA samples even if these are fragmented. We have successfully used DNA isolated from paraffin-embedded fixed biopsies.

7 MeDIP-seq

When combined with massively parallel sequencing, MeDIP (or MIRA) would be a very powerful method to generate whole-genome DNA methylation profiles. The main problem with such approaches, however, is the presence of bulk quantities of highly methylated repetitive DNA among captured DNA fragments. Analyses in our laboratory have shown that in a typical MeDIP experiment several classes of repeats (e.g. ALUs, satellites) are recovered with high efficiency (up to 100%) due to the high methylation content of these elements. A genomic DNA sample that is to be randomly sequenced will normally yield mostly unique sequences that can be assigned to a single genomic locus. In contrast, a DNA sample that has been enriched using MeDIP will primarily yield sequence reads that correspond to (highly methylated) repetitive sequence, which cannot be assigned unequivocally to specific genomic regions. In order to detect sufficient numbers of unique sequences within such samples, the total number of sequence reads has to be increased dramatically. This represents a major challenge towards the analysis of MeDIP samples by massive parallel sequencing. In contrast to MeDIP-seq, MeDIP-on-chip does not have this problem since genomic tiling array probes are repeat-masked.

8 Concluding Remarks

The past decades have seen an enormous increase in the number of studies focused on epigenetic changes, both from a molecular and from a clinical point of view. It has become clear that epigenetic aberrations play an important role in disease, and rational 'epidrug' design will require an intimate knowledge of the epigenetic alterations that underlie disease pathways. Numerous examples of gene-by-gene studies have focused on epigenetic aberrations at 'usual suspect' genes, and have combined DNA methylation and ChIP analyses to provide mechanistic details on repression mechanisms. Nevertheless, it is currently poorly understood how (at a global scale) epigenetic marks such as DNA methylation are interpreted and

translated into changes in gene expression. Several lines of evidence indicate that methyl-DNA binding proteins are functionally specialized (Hendrich et al. 2001), and their recruitment to genomic sites may not be determined only by recognition of methylated-DNA (Klose et al. 2005). Interestingly, a recent profiling analysis of MeCP2 –a prototype methyl-DNA binding protein with repressive activity– suggested that the majority of MeCP2-associated promoters are unmethylated and active (Yasui et al. 2007). In addition, the activities of various methyl-DNA binding proteins appears to be influenced by post-translational modifications of these proteins themselves (Chen et al. 2003, Le Guezennec et al. 2006, Tan and Nakielny 2006, Lyst et al. 2006). Together, this suggests the translation of the methyl-DNA mark into a repression signal is a highly regulated process involving many proteins and signalling pathways. Recently developed global profiling techniques for DNA methylation as well as ChIP provide opportunities for unbiased analyses to unravel molecular repression pathways at a global scale. Such approaches will allow comparative analyses of multiple different epigenomes. This kind of 'comparative epigenomics' will greatly increase our understanding of how epigenomes are shaped, and will provide opportunities to identify novel diagnostic/prognostic targets, and avenues for pharmacological interference with gene silencing.

References

Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823–37.

Chen, W.G., et al., *Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2.* Science, 2003. **302**(5646): 885–9.

Colella, S., et al., *Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites.* Biotechniques, 2003. **35**(1): 146–50.

Cross, S.H., et al., *Purification of CpG islands using a methylated DNA binding column.* Nat Genet, 1994. **6**(3): 236–44.

Dupont, J.M., et al., *De novo quantitative bisulfite sequencing using the pyrosequencing technology.* Anal Biochem, 2004. **333**(1): 119–27.

Eads, C.A., et al., *MethyLight: a high-throughput assay to measure DNA methylation.* Nucleic Acids Res, 2000. **28**(8): E32.

Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nat Genet, 2006. **38**(12): 1378–85.

Feng, Q. and Y. Zhang, *The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes.* Genes Dev, 2001. **15**(7): 827–32.

Frommer, M., et al., *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.* Proc Natl Acad Sci U S A, 1992. **89**(5): 1827–31.

Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): 261–82.

Hendrich, B., et al., *Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.* Genes Dev, 2001. **15**(6): 710–23.

Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.* Proc Natl Acad Sci U S A, 1996. **93**(18): 9821–6.

Huang, T.H., M.R. Perry, and D.E. Laux, *Methylation profiling of CpG islands in human breast cancer cells.* Hum Mol Genet, 1999. **8**(3): 459–70.

Ji, H. and W.H. Wong, *TileMap: create chromosomal map of tiling array hybridizations.* Bioinformatics, 2005. **21**(18): 3629–36.

Johnson, D.S., et al., *Genome-wide mapping of in vivo protein-DNA interactions.* Science, 2007. **316**(5830): 1497–502.

Khulan, B., et al., *Comparative isoschizomer profiling of cytosine methylation: the HELP assay.* Genome Res, 2006. **16**(8): 1046–55.

Klose, R.J., et al., *DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG.* Mol Cell, 2005. **19**(5): 667–78.

Le Guezennec, X., et al., *MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties.* Mol Cell Biol, 2006. **26**(3): 843–51.

Liu, C.L., S.L. Schreiber, and B.E. Bernstein, *Development and validation of a T7 based linear amplification for genomic DNA.* BMC Genomics, 2003. **4**(1): 19.

Lyst, M.J., X. Nan, and I. Stancheva, *Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins.* Embo J, 2006. **25**(22): 5317–28.

Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): 553–60.

Ng, H.H., et al., *MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex.* Nat Genet, 1999. **23**(1): 58–61.

Nouzova, M., et al., *Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays.* J Pharmacol Exp Ther, 2004. **311**(3): 968–81.

O'Geen, H., et al., *Comparison of sample preparation methods for ChIP-chip assays.* Biotechniques, 2006. **41**(5): 577–80.

Rauch, T. and G.P. Pfeifer, *Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer.* Lab Invest, 2005. **85**(9): 1172–80.

Rauch, T., et al., *MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells.* Cancer Res, 2006. **66**(16): 7939–47.

Rauch, T.A., et al., *High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer.* Proc Natl Acad Sci U S A, 2008. **105**(1): 252–7.

Ren, B., et al., *Genome-wide location and function of DNA binding proteins.* Science, 2000. **290**(5500): 2306–9.

Robertson, G., et al., *Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.* Nat Methods, 2007. **4**(8): 651–7.

Shiraishi, M., Y.H. Chuu, and T. Sekiya, *Isolation of DNA fragments associated with methylated CpG islands in human adenocarcinomas of the lung using a methylated DNA binding column and denaturing gradient gel electrophoresis.* Proc Natl Acad Sci U S A, 1999. **96**(6): 2913–8.

Solomon, M.J., P.L. Larsen, and A. Varshavsky, *Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene.* Cell, 1988. **53**(6): 937–47.

Tan, C.P. and S. Nakielny, *Control of the DNA methylation system component MBD2 by protein arginine methylation.* Mol Cell Biol, 2006. **26**(19): 7224–35.

Weber, M., et al., *Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.* Nat Genet, 2005. **37**(8): 853–62.

Weber, M., et al., *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.* Nat Genet, 2007. **39**(4): 457–66.

Xiong, Z. and P.W. Laird, *COBRA: a sensitive and quantitative DNA methylation assay.* Nucleic Acids Res, 1997. **25**(12): 2532–4.

Yasui, D.H., et al., *Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes.* Proc Natl Acad Sci U S A, 2007. **104**(49): 19416–21.

Zeng, P.Y., et al., *In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation.* Biotechniques, 2006. **41**(6): 694, 696, 698.

Zhang, X., et al., *Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis.* Cell, 2006. **126**(6): 1189–201.
Zheng, M., et al., *ChIP-chip: data, model, and analysis.* Biometrics, 2007. **63**(3): 787–96.
Zilberman, D., et al., *Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription.* Nat Genet, 2007. **39**(1): 61–9.

Sequencing the Epigenome

Alexander Meissner and Bradley E. Bernstein

Abstract The term 'epigenome' refers to the complete description of chemical changes to DNA and histones as they map onto the genome in a given cell type. A comprehensive genomewide catalog of epigenetic control elements and how these vary across cell states could offer critical insight into the relationships between genotype, phenotype and environment, and serve as a catalyst for future studies of the epigenetic mechanisms that regulate normal physiology and human disease. Our ability to characterize mammalian epigenomes has been markedly enhanced by technological developments in recent years. In particular, the introduction of ultra high-throughput sequencing has improved the precision, comprehensiveness and throughput of techniques for mapping chromatin and DNA methylation. This chapter will largely focus on these new applications and their use for high resolution interrogation of mammalian epigenomes.

Keywords DNA methylation · Histone modifications · Epigenome · Bisulfite sequencing · ChIP-Seq

1 Introduction

1.1 Epigenetics

Epigenetic modifications provide essential regulatory information that does not alter the primary nucleotide sequence (Epi: on top or in addition to). DNA methylation is generally associated with repressive contexts and stably propagated through cell division by DNA methyltransferases. Despite being the most extensively studied epigenetic modification in mammals experimental data for its genome-wide distribution, it's dynamic role during differentiation and its relationship with histone modifications remain limited (Bernstein et al. 2007; Bird, 2002). Post-translational histone modifications are implicated in epigenetic regulatory pathways such as

A. Meissner(✉)
Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA 02142, USA
e-mail: alex@broad.mit.edu

A.C. Ferguson-Smith et al. (eds.), *Epigenomics,* DOI 10.1007/978-1-4020-9187-2_2,