

# Applying whole-genome studies of epigenetic regulation to study human disease

J.D. Lieb<sup>a</sup> S. Beck<sup>b</sup> M.L. Bulyk<sup>c</sup> P. Farnham<sup>d</sup> N. Hattori<sup>e</sup> S. Henikoff<sup>f</sup>  
X.S. Liu<sup>g</sup> K. Okumura<sup>h</sup> K. Shiota<sup>e</sup> T. Ushijima<sup>i</sup> J.M. Greally<sup>j</sup>

<sup>a</sup>Department of Biology and Carolina Center for Genome Sciences, The University of North Carolina at Chapel Hill, NC (USA)

<sup>b</sup>Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge (UK)

<sup>c</sup>Division of Genetics, Department of Medicine, Department of Pathology, and Harvard/MIT Division of Health Sciences and Technology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA (USA)

<sup>d</sup>Department of Pharmacology and the Genome Center, University of California-Davis, Davis, CA (USA)

<sup>e</sup>Laboratory of Cellular Biochemistry, Animal Resource Sciences/Veterinary Medical Sciences, The University of Tokyo, Tokyo (Japan)

<sup>f</sup>Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA (USA)

<sup>g</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA (USA)

<sup>h</sup>Department of Life Sciences, Faculty of Bioresources, Mie University, Mie (Japan)

<sup>i</sup>Carcinogenesis Division, National Cancer Center Research Institute, Tokyo (Japan)

<sup>j</sup>Departments of Medicine (Hematology) and Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY (USA)

Manuscript received 26 September 2005; accepted in revised form for publication by M. Schmid, 6 October 2005.

Epigenetics may be broadly defined as the study of processes that produce a heritable phenotype that is not strictly dependent on DNA sequence. The definition has traditionally been restricted to processes that occur in the cell's nucleus, with the term 'heritable' having a loose meaning that can be applied to either the entire organism or single cells. For example, a process that produces a phenotype only in a specific cell type (for instance, chromatin-mediated maintenance of a differentiated state) is usually considered epigenetic even if it is not directly inherited, but instead must be re-established or actively maintained at each cell division. Given this definition, the field of epigenetics has long focused on proteins that affect DNA packaging, and thereby affect the utilization of the genetic information encoded in the DNA template. This focus extends to the enzymatic modification of those proteins, and to the enzymatic modification of the DNA template itself, primarily DNA methylation.

This review is written in conjunction with the international symposium on *Genome-wide Epigenetics 2005*, held at the University of Tokyo, Japan on November 8–10, 2005. Over the past decade, the field of epigenetics has undergone an exciting expansion in the number of researchers, techniques available and our understanding of epigenetic phenomena. The purpose of this short review is not to summarize all of these advances, but rather to guide the reader to more detailed sources of information by sketching an outline of the major thrusts in the field, emphasizing mammalian epigenetics in particular.

## Mediators of epigenetic regulation of the genome

Epigenetic information manifests itself by regulating gene expression, often reflected by biological phenotypes. The mechanistic mediators of heritable gene expression appear to be similar whether one is considering epigenetic inheritance on the organismal level (from parent to offspring) or the cellular level (from parent to daughter cells). For example, a mediator of epigenetic regulation is cytosine methylation. The action of maintenance methyltransferases ensures that the methylation of pairs of cytosines on complementary DNA strands at CpG dinucleotides is propagated to both pairs of daughter DNA molecules following DNA replication (Bestor,

Request reprints from John M. Greally  
Departments of Medicine (Hematology) and Molecular Genetics  
Albert Einstein College of Medicine  
1300 Morris Park Avenue, Bronx, NY 10461 (USA)  
telephone: +1 718 430 2875; fax: +1 718 824 3153  
e-mail: [jgreally@aecom.yu.edu](mailto:jgreally@aecom.yu.edu)

2000). The inheritance of gene regulatory information from parental gametes defines the phenomenon of genomic imprinting, for which cytosine methylation is a central mediator (Li et al., 1993).

Cytosine methylation is the best-studied mediator of epigenetic regulation. In mammalian genomes, the vast majority of methylation of cytosines occurs when the cytosine is followed by a guanine, a so-called CpG dinucleotide (Clark et al., 1995). The methyl group protrudes into the major groove of the DNA double helix, preventing the binding of transcription factors that would otherwise bind locally in a sequence-specific manner (Bell and Felsenfeld, 2000; Hark et al., 2000), and facilitating the binding of methyl-binding proteins that generally attach to DNA in a sequence non-specific manner (Jorgensen and Bird, 2002). The methylation of cytosines may not be the initiating event in a change from activity to silencing of a locus, but it is certainly associated with a number of molecular changes. These include the alteration of covalent modifications of histone tails (El-Osta and Wolffe, 2000), the recruitment of methyl-binding proteins (Jorgensen and Bird, 2002) and proteins normally associated with distinctive chromatin states (such as SWI/SNF components; Hari Krishnan et al., 2005), and the compaction of the chromatin with silencing of gene expression locally (Nguyen et al., 2001).

It is possible that cytosine methylation is induced by primary events involving the covalent modification of histone tails. The histone octamers that form the nucleosome protrude N-terminal tails, where it was thought that various combinations of methylation, acetylation, phosphorylation and other post-translational modifications could be established to create what has been referred to as the histone code (Jenuwein and Allis, 2001). However, this combinatorial concept is being challenged by data suggesting that the pattern of modifications may be more simple than the potential complexity, establishing what is more like a binary on or off pattern (Schubeler et al., 2004; Dion et al., 2005; Liu CL et al., 2005). As described for methylcytosine, these modified histone molecules are capable of recruiting further molecules and complexes that influence chromatin structure and local gene activity. Whether histone tail modifications induce or are induced by cytosine methylation, it seems that the disruption of one component of this system can induce changes in other components also (Ben-Porath and Cedar, 2001).

The recent finding in fission yeast that small RNAs are targeted to heterochromatin (Verdel et al., 2004) suggests that epigenetic gene regulation in mammalian cells may also involve the actions of such short RNAs. In addition, it has been observed that a subset of DNA methyltransferases and methyl-binding proteins have RNA-binding properties (Jeffery and Nakielnny, 2004), supporting the idea that RNA molecules may influence the regulation of expression of their own or other loci. However, a point worth stressing is that the major consequence of all these biochemical processes, which ultimately alter gene expression, is the regulation of transcription factor binding. Transcription factors with activating effects bind in modules at specific loci to facilitate the loading of the RNA polymerase complex locally, a prerequisite for gene ex-

pression. The relationship between epigenetic regulation and transcription factor binding tends to be overlooked in reviews of either field. However, as we emphasize below, techniques to study chromatin components and transcription factors *in vivo* are equally advanced, allowing integration of data and major advances in the understanding of the relationship of epigenetic alterations with transcription factor binding. An example of such a study shows a relationship between binding of the CTCF nuclear protein and cytosine methylation in a large-scale analysis (Mukhopadhyay et al., 2004). The use of high-throughput techniques to study epigenetic regulatory processes in large genomic regions or genome-wide has been described as an 'epigenomic' approach (Fazzari and Grealley, 2004). Our insights into epigenetic regulatory processes have to date been founded mostly on detailed studies of individual loci; now that epigenomic approaches have been developed, we are beginning to accrue knowledge based on less detailed studies of large numbers of loci.

### **Déjà vu: Why study the whole epigenome?**

Although many of the details remain unclear, few people today doubt the importance of epigenetic processes (in addition to the genetic code) for biological and medical research. A full understanding of epigenetic inheritance and how epigenetic regulatory information is distributed in the genome therefore ought to be the ultimate aim of any genome-wide epigenetics or Epigenome Project. Of course, there is currently much debate on whether the time and technology is ready for such an enormous undertaking and those who remember the early discussions about the Human Genome Project will find some of the arguments reminiscent of that time. For the Genome Project, the turning point came in 1995, when the more pragmatic protagonists at the time pointed out that a (draft) genome sequence could be generated with existing technology (Marshall, 1995). Anticipating, of course, that the effort would be self-catalyzing once started which, with the luxury of hindsight, turned out to be true when the Human Genome was announced finished in 2003, under budget and ahead of time. Exactly when the discussion about the Epigenome Project started is arguable, most likely it started independently in the minds of multiple scientists. In Pubmed, the first indexed mentioning of 'Epigenome' appeared in 1987 (Kieser, 1987) and 'Epigenome Project' was mentioned the first time only in 2002 (Novik et al., 2002). Therefore, a 'spoo' editorial in a leading science magazine (*Nature Biotechnology*) was quite timely when it suggested in August 2000 that a trilogy be required to complete the Human Genome Project: 'The Draft', 'The Closure' and 'Epigenetics Strikes Back'. With the 'Draft' and the 'Closure' delivered, the time has come to explore the next frontier, the 'Epigenome'.

The rationale for studying the epigenome and how this might be accomplished has been reviewed recently (Murrell et al., 2005). Basically, different cell types require different epigenomes to control – in time and space – the correct execution of their individual transcriptional programs and (revers-

ible) epigenetic modulation has the flexibility to achieve the required fine-tuning of the underlying nuclear genome. A recent study further suggests that the transcriptional program is much larger and more complex than previously thought (Cheng et al., 2005). High-resolution transcript mapping of ten human chromosomes revealed that the majority of (mostly non-polyadenylated) transcripts map to loci for which there is currently no annotation available. Yet, these transcripts of unknown function may have important implications for gene expression and (epi) genotype-phenotype associations. The power to detect any alterations anywhere in the epigenome will therefore be of great value to understand what goes wrong if a program is executed incorrectly, particularly in the context of disease. Obviously, our ability to analyze and interpret entire epigenomes is very much dependent upon the availability of adequate technologies which will be discussed in the next section.

### Techniques for genome-wide epigenetic studies

Much of the recent progress in understanding epigenetic phenomena is directly attributable to technologies that allow researchers to pinpoint the genomic location of proteins that package and regulate access to the DNA. The advent of DNA microarrays and inexpensive DNA sequencing has allowed many of those technologies to be applied to the whole genome. The focus of the Tokyo conference is the study of epigenetics on a genome-wide scale, so here we will focus on technologies that enable a comprehensive investigation of the genome. We note that Bas van Steensel's group has recently published a series of very informative reviews relevant to this subject (Loden and van Steensel, 2005; van Leeuwen and van Steensel, 2005; van Steensel, 2005).

#### *Genome-wide cytosine methylation studies*

Cytosine methylation has been the best-studied epigenetic mechanism, largely because the experimental techniques for its study have been available for some time. The mainstays of experimental approaches have been the sensitivity of certain restriction enzymes to cytosine methylation and the relative resistance of methylcytosine to bisulfite-induced deamination. Experiments based on these approaches have revealed great detail about individual loci. The expansion of these studies from single-locus to genome-wide approaches is still in development. A platform that has proven to be very robust is the use of restriction landmark genome scanning (RLGS), which uses a rare-cutting methylation-sensitive restriction enzyme (such as *NotI*) in combination with a second enzyme to create a profile of *NotI* digestion products. The digested *NotI* overhanging ends are radiolabeled and separated by 2-dimensional gel electrophoresis. Approximately 2,000 sites can be resolved using RLGS. Using virtual image RLGS (vi-RLGS), the actual pattern of *NotI* fragments can be compared with the computationally predicted pattern for that genome (Matsuyama et al., 2003). This allows differences observed by gel electrophoresis to be linked to specific genomic locations for subsequent validation studies. The technique

suffers from being technically challenging, difficult to validate even with vi-RLGS, and limited in terms of the number of loci that are testable (Fazzari and Greally, 2004). Nonetheless, it has been the single most informative technique describing differences in cytosine methylation between tissues, stages of differentiation and in disease (Shiota et al., 2002; Kremensky et al., 2003; Song et al., 2005).

Other restriction enzyme-based techniques have also been developed, and high-throughput bisulfite sequencing using MALDI-TOF looking at numerous loci in parallel has been the foundation of the European Human Epigenome Project (Bradbury, 2003; Rakyan et al., 2004). Most of the techniques used to date have been reviewed comprehensively (Laird, 2003; Ushijima, 2005), although several recent publications update this list (Ching et al., 2005; Hu et al., 2005; Weber et al., 2005). The techniques differ in terms of how they select a fraction of the genome with distinctive cytosine methylation. Most use the sensitivity of restriction enzymes to cytosine methylation as the means of discrimination, while a new technique uses the affinity of an anti-methylcytosine antibody to enrich the methylated fraction of the genome (methylated DNA immunoprecipitation, MeDIP; Weber et al., 2005). The techniques also differ in terms of how they detect the identity of the sequences that constitute these distinctive fractions of the genome, with gel electrophoresis now largely superseded by microarray-based approaches or high-throughput sequencing.

At present, there is no technique that can test the methylation status of every CpG dinucleotide in the genome. Furthermore, little attention has been paid to the effects of the heterogeneity of CpG dinucleotide clustering in the genome, apart from using this property to define CpG islands (Gardiner-Garden and Frommer, 1987; Takai and Jones, 2002). Clustering of CpGs increases the number of informative cytosines or restriction sites in some areas compared with others, so that a comparison of cytosine methylation genome-wide will show much stronger signals of methylation or the absence of methylation at sites of clustering than sites where CpGs are relatively more sparse. Possibly because of this, our attention is focused on CpG-enriched regions as particularly informative. Our appreciation of the role of less CpG-rich regions in dynamic methylation patterns remains limited as a consequence.

#### *Mapping the genomic location of chromatin proteins and enzymes by ChIP-chip*

Chromatin components and determinants of chromatin dynamics are often the mediators of epigenetic effects. The genome-wide mapping of these chromatin marks by ChIP-chip has led to important insights regarding the mechanism of transcriptional and epigenetic memory, and how different chromatin states are propagated through the genome. In yeast, ChIP-chip has been used to determine the distribution of non-histone chromatin components (Smith et al., 2003; Glynn et al., 2004; Lengronne et al., 2004; Weber et al., 2004), enzymes involved in histone modification (Kurdistani et al., 2002; Ng et al., 2003; Humphrey et al., 2004), and post-translational histone modifications themselves (Bernstein et al.,

2002; Robyr et al., 2002; Kurdistani et al., 2004; Loden and van Steensel, 2005; van Leeuwen and van Steensel, 2005).

The 'ChIP-chip' method uses microarrays to expand the number of loci testable by Chip and is currently the most popular technique for high-resolution genome-wide mapping of chromatin proteins and enzymes. ChIP-chip was developed in the late 1990s (Editorial, *Nature Genetics*, 1998) and first published shortly thereafter (Ren et al., 2000; Iyer et al., 2001). Generally, cells are fixed, chromatin is fragmented, and the protein of interest is purified, along with any associated DNA, using an antibody or an affinity tag. The DNA that was co-purified with the protein is then detected using a DNA microarray and mapped back to the genome, allowing the binding position of the protein to be inferred. The details of this technique have been reviewed extensively (Kuo and Allis, 1999; Orlando, 2000; Nal et al., 2001; Farnham, 2002; Johnson and Bresnick, 2002; Weinmann and Farnham, 2002; Kurdistani and Grunstein, 2003; Lieb, 2003; Oberley and Farnham, 2003; Spencer et al., 2003; Bernstein et al., 2004a; Buck and Lieb, 2004; Hanlon and Lieb, 2004; Im et al., 2004; Oberley et al., 2004; Robyr et al., 2004). Initial ChIP-chip studies focused on the genome-wide location of well-studied yeast transcription factors. These studies empirically verified the effectiveness of the method by identification of previously known gene targets, and discovery of new and physiologically relevant targets (Ren et al., 2000; Iyer et al., 2001; Lieb et al., 2001; Simon et al., 2001; Horak and Snyder, 2002). In addition, DNA-binding motifs corresponding to the known affinities of the proteins were derived from the ChIP-chip data. In the short time since those original studies, ChIP-chip has been applied to hundreds of transcription factors in organisms ranging from bacteria to humans (van Steensel, 2005).

More recently, the approach of assessing global chromatin structure has been extended to more complex genomes. In *Drosophila*, the global acetylation pattern of histones H3 and H4 and the global methylation pattern of histone H3 lysine 4 (H3K4) and lysine 79 (H3K79) have been monitored using ChIP-chip (Schubeler et al., 2004). H3K4 di- and trimethylation, H3K79 dimethylation and H3 and H4 acetylation were all found to be present in the same genes. These results are in agreement with the genome-wide pattern of H3K4 methylation (Bernstein et al., 2002) and Set1p binding (Ng et al., 2003) in yeast and suggest that, in *Drosophila*, as in yeast, a pattern of nucleosome modifications distinguishes actively transcribed genes from repressed genes throughout the genome. Detailed genome-scale studies have also been carried out in *Arabidopsis* (Lippman et al., 2004).

Although the antibodies used for ChIP-chip studies of chromatin modifications can often be used across species, as of the printer's deadline there is only one published report of a genome-wide, high-resolution ChIP-chip study in mammalian cells (Kim TH et al., 2005). This is mainly due to difficulty in producing inexpensive and easy-to-use full-genome microarrays. The initial mammalian ChIP-chip experiments identified binding sites for the c-Myc, Max, Gata1, E2F and Rb transcription factors in cultured human cells (Horak et al., 2002; Weinmann et al., 2002; Li et al., 2003; Wells et al.,

2003). By practical necessity, the DNA microarrays used in these pioneering studies represented only a fraction of the genome. For the c-Myc and Max studies, DNA microarrays were constructed with PCR products spanning the proximal promoters of 4,839 of the approximately 30,000 human genes (Li et al., 2003). Later studies that examined the genomic distribution of the mammalian Set1 homolog MLL used an expanded version of this array that included 19,000 proximal promoters (Guenther et al., 2005). To map the mammalian transcription factors E2F and Rb, DNA microarrays were created with 7776 CpG island clones (Weinmann et al., 2002; Wells et al., 2003). CpG islands are short stretches of DNA containing a high density of non-methylated CpG dinucleotides, and are frequently associated with the promoters and the first exon of a gene (Antequera and Bird, 1993). Studies of transcription factor binding across human chromosomes 21 and 22 have been carried out (Martone et al., 2003; Cawley et al., 2004; Euskirchen et al., 2004; Carroll et al., 2005). These chromosome 21 and 22 arrays were also used for the first genome-scale study of histone modifications (H3K4me2, H3K4me3, and H3K9/14Ac) in human cells (Bernstein et al., 2005). This year a group of researchers published the first truly genomic human ChIP-chip experiment, using an array of 14.5 million 50 mers that covered the unique regions of the human genome at 100 bp resolution to analyze the distribution of TFIID on the human genome (Kim TH et al., 2005). Microarrays similar to the one used in that study are in commercial development, and many chromatin modifications are being mapped as part of the NIH's ENCODE project (2004) (Boguski, 2004; van Steensel, 2005).

There are some challenges inherent to any ChIP-chip experiment, and experiments that aim to determine the distribution of chromatin components and histone modifications are especially subject to these concerns. Several factors could create non-biological variation in results, including the effects of fixation, epitope accessibility, antibody specificity, microarray content, and underlying bulk nucleosome occupancy. These challenges, and suggestions for overcoming them, have been discussed at length in recent reviews (Buck and Lieb, 2004; Hanlon and Lieb, 2004; Loden and van Steensel, 2005; van Leeuwen and van Steensel, 2005).

#### *ChIP-chip alternatives: Dam-ID and in vivo biotin-tagging*

Current alternatives to ChIP-chip include the use of DNA adenine methyltransferase-Identification (Dam-ID) (van Steensel and Henikoff, 2000, 2003; van Steensel et al., 2001; Greil et al., 2003). In this method, the eukaryotic chromatin protein of interest is fused with the bacterial enzyme DNA adenine methyltransferase. Any DNA with which the protein is associated will then be methylated at adenine, a mark normally restricted to prokaryotes. The methylated DNA can then be purified with antibodies or probed with methyl-sensitive restriction enzymes to deduce the location of the protein's interaction with genomic DNA. The Dam-ID is very powerful because it does not require antibodies to be raised against each factor of interest and obviates the need for crosslinking. However, it requires expression of a recombinant protein,

careful control of the level and duration of expression, and cannot be used to map post-translational histone modifications. A second technique requires that the chromatin component be tagged with the biotin ligase recognition peptide, and that this fusion protein and biotin ligase be co-expressed in the same cell type. Cellular biotin is conjugated to the chromatin protein of interest, allowing streptavidin to be used on sheared chromatin from these cells to isolate the DNA at which this chromatin component is located. This approach has been used to track histone H3.3 localization in cultured *Drosophila* cells (Mito et al., 2005).

*Genome-wide in vitro methods to determine DNA binding specificity and quantify the influence of epigenetic processes*

Many epigenetic effects are mediated, ultimately, through their effect on transcription factors and transcriptional regulation. To understand how epigenetic processes influence transcriptional programs, accurate binding-site descriptions of hundreds of transcription factors will likely be needed. DNA-binding specificity can be determined by many well-established methods, including binding site selection (SELEX) (Oliphant et al., 1989; Blackwell et al., 1990; Tuerk and Gold, 1990) and electrophoretic mobility shift assays (EMSA) (Fried and Crothers, 1981). However, these methods are labor-intensive, not amenable to high-throughput analysis, and do not sample the full range of a protein's natural *in vivo* DNA substrates. As described below, one can use ChIP-chip experiments to infer binding motifs from computational analysis of the ChIP-enriched sequences (Liu et al., 2002). While ChIP-chip is a powerful method, the ability to infer relevant and accurate binding sites is dependent on a number of factors, including adequate expression of the protein of interest, nuclear localization and DNA binding by that protein, and availability of a specific antibody that is capable of recognizing the protein in the context of a specific DNA-protein complex. Furthermore, the discovery of binding sites from ChIP-chip data is complicated by the effects of protein-protein interactions, and the cooperative and competitive DNA binding of other proteins *in vivo*. Two high-throughput *in vitro* methods for the determination of DNA-binding specificity have emerged recently: Protein Binding Microarrays (PBMs) (Bulyk et al., 2001; Mukherjee et al., 2004) and DIP-chip (DNA Immunoprecipitation followed by microarray analysis) (Liu X et al., 2005).

In the PBM approach, a purified, epitope-tagged DNA binding protein is incubated with a microarray spotted with double-stranded DNA corresponding either to short, synthetic double-stranded oligonucleotides (Bulyk et al., 2001; Mukherjee et al., 2004) or longer PCR products (Mukherjee et al., 2004). The protein-bound microarray is then washed gently and stained with a fluorophore-conjugated anti-tag antibody (for example, an anti-GST antibody would be used with GST-tagged protein); alternatively, a fluorophore-conjugated anti-transcription factor primary antibody could be used. The stained array is then washed, spun dry, and scanned with a standard microarray scanner. The PBM experimental protocol takes less than a day, and multiple PBM microarray

experiments can be performed in parallel. The PBM data, normalized by DNA concentration, are analyzed with a motif finding algorithm in order to identify the protein's DNA binding sites. Although PBM experiments can be performed using some 'standard' binding conditions, one could examine the effects of altered binding conditions, such as alterations in buffer composition, transcription factor concentration, or even the effects of small molecule or protein cofactors. Importantly, the PBM experiments do not require prior knowledge of the culture conditions in which the transcription factor is expressed, do not require potentially limiting tissue sources, and do not require protein-specific antibodies. Comparison of PBM and ChIP-chip data has revealed that the respective DNA binding site motifs can correspond quite well (Mukherjee et al., 2004). Moreover, comparison of the *in vivo* and *in vitro* binding data will permit an analysis of what local sequence context features may contribute to binding site usage. PBM data will be useful for generating binding site motifs for known and predicted transcription factors with poorly or uncharacterized DNA binding specificities. Such binding site data, coupled with further computational analyses, will permit the prediction of the regulatory roles of transcription factors, as well as the prediction of candidate *cis* regulatory modules (i.e., transcriptional enhancers) (Bulyk, 2003; Philippakis et al., 2005 and submitted).

In DIP-chip, a purified DNA binding protein is incubated with purified, sheared yeast genomic DNA. Protein-DNA complexes are separated from unbound DNA using immunoprecipitation or affinity purification. Purified DNA fragments are amplified, labeled fluorescently, and identified by hybridization to a DNA microarray. Computational methods are then used to define a binding site based on enriched sequences. DIP-chip, while similar in concept to ChIP-chip, can overcome some of the limitations of ChIP-chip by inferring accurate DNA-binding specificities under well-defined and easily varied *in vitro* conditions. The experimental protocol for DIP-chip can also be used for a rather different purpose, which is comparing the sites of binding *in vitro* with the sites of binding *in vivo*, as defined by ChIP-chip. Comparisons of DIP-chip and ChIP-chip experiments will be useful in determining how much of the specificity of *in vivo* interactions depends on chromatin and other epigenetic factors, and how much is inherent to the protein and DNA itself. Both PBMs and DIP-chip are powerful adjuncts to ChIP-chip and Dam-ID experiments, and are efficient and accurate methods for determining *in vitro* DNA binding specificity.

*Sequencing-based alternatives to microarray detection*

The DNA purified from a ChIP experiment can be cloned and sequenced, providing an alternative to microarray-based detection (Impey et al., 2004; Roh et al., 2004, 2005; Chen and Sadowski, 2005; Kim J et al., 2005; Ng et al., 2005). Currently, the cost of performing DNA microarray hybridization on commercial arrays that cover the whole human genome is prohibitive to all but the wealthiest labs. Even with these 'whole genome' arrays, the whole genome is not represented, only the non-repetitive portions. Therefore, sequencing-based methods could prove to be very attractive, particularly for

larger genomes. Rather than sequencing the entire cloned DNA fragments, most of the referenced methods are similar to SAGE, and create small (~17–21 bp) tags from each enriched fragment. The main question surrounding these methods is how many sequencing reads must be performed before an adequate sampling of all enriched sequences is achieved. Consider the case in which a 20-fold enrichment of targets is achieved by IP, and targets represent 1% of all genomic fragments. If a sequencing approach is chosen, only ~17% of all sequenced tags would be IP targets at all, and for each experiment, a very large number of clones would have to be sequenced to sample the entire IP result with sufficient coverage to identify targets with confidence. This method may become feasible by devising high-throughput schemes to increase the practical enrichment and decrease background prior to sequencing. These may include pre-screening clones for repetitive elements, including a subtractive hybridization step, modification of the standard ChIP experiment to include a second IP, or size selection to limit nonspecific clones and repetitive elements (Weinmann and Farnham, 2002). Another challenge of sequence-based methods includes assigning the sequence tag back to a unique position in the genome (Kim J et al., 2005). What may ultimately drive these alternatives to the forefront are future reductions in sequencing costs. New commercially available sequencing technologies, such as massively parallel signature sequencing (MPSS) (Brenner et al., 2000) and more recently the emulsion-based method coupled to pyrosequencing (Margulies et al., 2005) can sequence tens of thousands of DNA fragments simultaneously, and could be used to sample an entire ChIP. An advantage of a sequencing-based approach is that the results obtained should be more directly correlated with *in vivo* events when compared with microarray-based approaches.

#### *New genomic techniques for understanding chromatin*

New technologies and methods for understanding chromatin and epigenetics continue to emerge. Many are based on long-trusted biochemical assays coupled to new detection techniques. DNase I hypersensitive sites have been a defining characteristic of ‘active’ chromatin for decades, and several new methods have arisen to map hypersensitive sites quantitatively and genome-wide (Dorschner et al., 2004; Sabo et al., 2004). Researchers have also directly measured nucleosome occupancy with genomic methods. A clever alternative to ChIP-chip was performed by digesting yeast chromatin with micrococcal nuclease, purifying mononucleosomes, and using high-resolution tiling arrays to look for missing linker DNA (Yuan et al., 2005). Using this method, the authors were able to obtain a very high-resolution map of nucleosome occupancy on yeast chromosome III, which was consistent with earlier, lower-resolution ChIP-chip studies (Bernstein et al., 2004b; Lee et al., 2004). Another group separated mammalian chromatin with a sucrose gradient, and using microarrays with ~1 Mb resolution showed that the ‘open’, chromatin is composed of underlying DNA with a high gene density (Gilbert et al., 2004). In a reciprocal approach, another group isolated chromatin enriched for Histone H1 and relatively resistant to DNase I digestion, and found an inverse correla-

tion between the isolated chromatin and gene expression (Weil et al., 2004).

The least-understood aspects of chromatin, long-range interactions and nuclear domain organization, remain enigmatic largely because few quantitative techniques exist for their study. Fluorescence *in situ* hybridization studies of epigenetically-distinctive regions have revealed differences in the resolution of sister chromatids believed to be a marker of DNA replication timing (Greally et al., 1998; Kagotani et al., 2002; Takebayashi et al., 2005). Recently, several new techniques have emerged that are providing insights into the higher-order organization of chromatin (Carter et al., 2002; Dekker et al., 2002; Lebrun et al., 2003). Casolari et al. mapped the genome-wide localization of several components of the *S. cerevisiae* nuclear transport machinery, and their data suggests that the most highly transcribed genes are held near nuclear pores at the periphery of the nucleus to facilitate rapid transport of mRNA to the cytoplasm (Casolari et al., 2004).

#### **Analysis and interpretation of epigenomic data**

An obvious prediction at this stage of development of the field of epigenomics is that our ability to generate experimental results will shortly overwhelm our ability to manage and analyze these complex, highly-dimensional data. Some of these problems will require standardization of vocabularies and the definition of common data elements to allow the kind of syntactic and semantic interoperability that allows us to integrate diverse sources of data. However, for the purpose of this review we will focus on the more immediate need to analyze the results of epigenomic studies.

#### *Nucleosomal occupancy experiments and related computational analysis*

While genomic DNA is generally packaged into chromatin as nucleosomes, local nucleosomal organization is often disrupted around functional regulatory sequences, making the regions accessible for transcription factors. These regions, typically ~250 bp in length, are hypersensitive (HS) to DNase I, thus can be detected throughout the genome by comparing the DNase treated to non-treated genomic DNA using high-throughput quantitative PCR (Dorschner et al., 2004). Given a collection of experimentally detected HS sequences and non-HS control sequences, a support vector machine (SVM) can be trained to differentiate HS from non-HS sequences and predict new HS sequences in the genome with 85% accuracy. In the SVM, each sequence is embedded into a vector that contains the counts of all k-mers (k = 1 to 6) in the sequence. The SVM found the CG dinucleotide to be the most important sequence feature in the determination of HS, and predicted 65% of HS sequences to be > 5 kb from the nearest transcription start site. The HS sequences were also found to be enriched with CTCF recognition sequences, which are signature insulator elements (Noble et al., 2005).

In another nucleosome occupancy study, differentially labeled micrococcal nuclease (MNase)-treated and non-treated yeast genomic DNA were hybridized to microarrays with

overlapping 50-mer probes tiled every 20 bp across the yeast chromosome (Yuan et al., 2005). Nucleosome-associated DNA is resistant to MNase digestion, whereas the nucleosome-free DNA regions or linkers are hypersensitive. The authors constructed a hidden Markov model to look for well-positioned nucleosomes as a run of 6–8 high-ratio probes (~146 bp) and delocalized nucleosomes as a run of >9 high-ratio probes (Yuan et al., 2005). The remaining regions with runs of low-ratio probes were considered nucleosome-depleted DNA regions. The hidden Markov model identified nucleosome-depleted regions of ~150 bp at ~200 bp upstream of many ORFs. These regions are more conserved than the background genomic sequence, include many transcription factor binding sites, and are enriched for poly-A and poly-T.

### *ChIP-chip statistical analysis*

Noise is inevitable in ChIP-chip experiments due to the random nature of chromatin shearing, nonspecific binding during the immunoprecipitation step, infidelities in DNA amplification and labeling, and the noise intrinsic to microarray synthesis and hybridization. The means of dealing with these problems to date have been based on performing independent experimental replicates, although advances in molecular techniques and microarray resources are improving our ability to define signals with confidence based on single experiments. To combine experimental replicates, an error model first proposed by Hughes et al. (2000) was adapted to ChIP-chip experiments (Simon et al., 2001; Lee et al., 2002; Ren et al., 2002; Harbison et al., 2004). For ChIP-chip conducted on a transcription factor, a (ratio, *P* value) pair is reported for each intergenic probe. The ratio is a weighted average of ChIP over control fold-change in the replicates, where weight is based on the fold-change and systematic error in each replicate. The product of ratio and weight sum follows a normal distribution, so its *P* value can be calculated. A *P* value of 0.001 has been often used as the threshold to call a binding target, at which level false positive and false negative rates are estimated as 6–10% and 24%, respectively (Lee and Young, 2000). Each intergenic sequence represented on the microarray is assigned to the gene(s) downstream of it. An intergenic region is assigned to two genes if it lies in between two divergently transcribed genes ( $\leftarrow\rightarrow$ ), and not assigned if it lies in between two convergently transcribed genes ( $\rightarrow\leftarrow$ ).

### *Transcription factor motif finding*

ChIP-chip experiments commonly identify protein-DNA binding targets at a resolution of hundreds of base pairs, but precise protein-DNA binding sites are often only 5–20 bp in length. Although many chromatin regulators do not have sequence-specific DNA recognition themselves, they can be recruited to specific sequence locations in the genome by interacting with other proteins (Cosma et al., 1999; Hampsey and Reinberg, 2003). After targets of ChIP-chip against a transcription factor or chromatin regulator are identified, biologists often proceed to use computational tools to find enriched sequence motifs from the ChIP-targets to characterize the precise protein-DNA binding sites. Earlier computational methods such as MEME (Bailey and Elkan, 1994; Grundy et

al., 1996), AlignACE (Roth et al., 1998; Kurdistani et al., 2002), and BioProspector (Liu et al., 2001) look for sequence motifs enriched in all the ChIP-target sequences compared to the genome background. The DNA fragments most highly enriched by ChIP often contain multiple binding sites for the same protein, so MDscan (Liu et al., 2002) was developed to begin motif searches from these sequences. The insight that sequences with more and stronger motif occurrences often have higher ChIP-enrichment values motivated REDUCE (Bussemaker et al., 2001; Wang et al., 2002) and Motif Regressor (Conlon et al., 2003; Ben-Yehuda et al., 2005) to find motifs whose occurrence in sequences correlate with ChIP-enrichment ratio. Perhaps the most powerful tools are meta-motif finders such as TAMO (Gordon et al., 2005) that combine different motif-finding algorithms to incorporate sequence pattern enrichment, evolutionary conservation, microarray correlation, and known motif databases to derive the most likely binding motif.

While formaldehyde is used with the belief that it stabilizes protein-DNA interactions, evidence from 20 years ago indicates that it works in a ChIP experiment by trapping DNA when protein constituents of chromatin cross-link (Solomon and Varshavsky, 1985). However, the effect to cause protein-protein cross-linking allows the interactions between two transcription factors or chromatin regulators to be studied. An antibody used to immunoprecipitate one of an interacting complex of proteins will pull down the DNA targets of both proteins, although the indirect target might have weaker enrichment (Iyer et al., 2001). Therefore, if the ChIP targets of two factors overlap significantly, yet the two factors are not expected to have the same DNA recognition properties, it is reasonable to hypothesize that the two proteins may be part of the same complex. This has led to the discovery of interactions between Irf1 and Rap1 (Wade et al., 2004), and between Sum1 and Hst1 (Robert et al., 2004). Computational motif finding can even predict protein-interaction partners using ChIP-chip data from just a single protein. For example, the canonical binding motifs of both the estrogen receptor (ER) and a forkhead protein were discovered from ChIP targets of the ER (Carroll et al., 2005). Based on the motif discoveries and the fact that forkhead protein FoxA1 is highly co-expressed with ER, the authors predicted the interaction between ER and FoxA1, and later proved that FoxA1 is required for ER binding (Carroll et al., 2005).

### *Analysis of ChIP-chip data on high-resolution tiling arrays*

As reviewed above, many microarray platforms have been used to perform ChIP-chip experiments in organisms with large genomes. For tiled arrays, two platforms are currently in use, long oligonucleotide arrays from NimbleGen (Kim TH et al., 2005), and Affymetrix 25-mer oligonucleotide microarrays. Due to the significant probe variability and noise associated with the Affymetrix arrays, robust analysis algorithms are required. The first method to analyze the ChIP-chip experiments on Affymetrix tiled arrays was the non-parametric Wilcoxon rank sum test (Cawley et al., 2004; Bernstein et al., 2005). This method looks at probes in a sliding window of

800–1000 bp, ranks all the ChIP and control probes together in each window by their perfect match-mismatch (PM-MM) values, and checks whether the sum of probe ranks in the ChIPs are significantly higher than that in the controls. To adjust for probe variability, another approach is to estimate the baseline hybridization behavior of a probe across ChIP and control experiments from many different conditions and laboratories (Carroll et al., 2005; Li et al., 2005). A hidden Markov model was then applied to find regions of probes whose (PM-MM) values are much higher than the estimated baseline behavior. Since the full-genome Affymetrix tiled arrays are still a developing platform that is not currently commercially available, we expect many new analysis algorithms developed in the near future.

Different statistical analysis programs are more appropriate for data derived from the NimbleGen oligonucleotide arrays. These arrays are composed of long oligomers (50-mers) and only perfect matches to the sequence are built on the arrays. One analysis program that has been developed to identify binding sites from ChIP-chip data obtained from NimbleGen arrays is based on percentile-derived thresholds and *P* value determined peak widths which are calculated according to the Waterman extension of the Erdos-Renyi Law. Due to the inherent noise in the ChIP-chip data generated, this analysis program was used to predict peaks from three independent array experiments and then the peaks that are present in at least two of the three arrays are categorized as binding sites. Using very conservative thresholds and *P* values, E2F1 binding sites in HeLa cells were identified from ChIP-chip experiments. Interestingly, between 30 and 40% of all promoters within the 30 Mb encompassing the ENCODE regions were categorized as bound by E2F1 (Bieda et al., manuscript submitted).

The ChIP-chip technique is a very young technology at present, and there are differing opinions on a number of experimental variables, including the appropriate experimental controls, the optimal microarray design and hybridization conditions, whether experimental replicates will continue to be necessary as the technology improves, and whether the same experimental controls and data analytical techniques are appropriate for transcription factor binding and histone modification studies. A lot of progress is expected in this field in the immediate future.

## Epigenetics and disease

While the focus in this review has been on the basic scientific questions and techniques in whole-genome epigenetic studies, the goal towards which researchers in this field are aiming is the application of these techniques to the study of human disease. The two areas where most insights have been gained are in mammalian development and in cancer biology. We focus on these areas here, while recognizing that many other applications are being developed, including interesting new insights into the role of epigenetics in aging (Fraga et al., 2005) and sex-specific regulation of autosomal genes (Sarter et al., 2005).

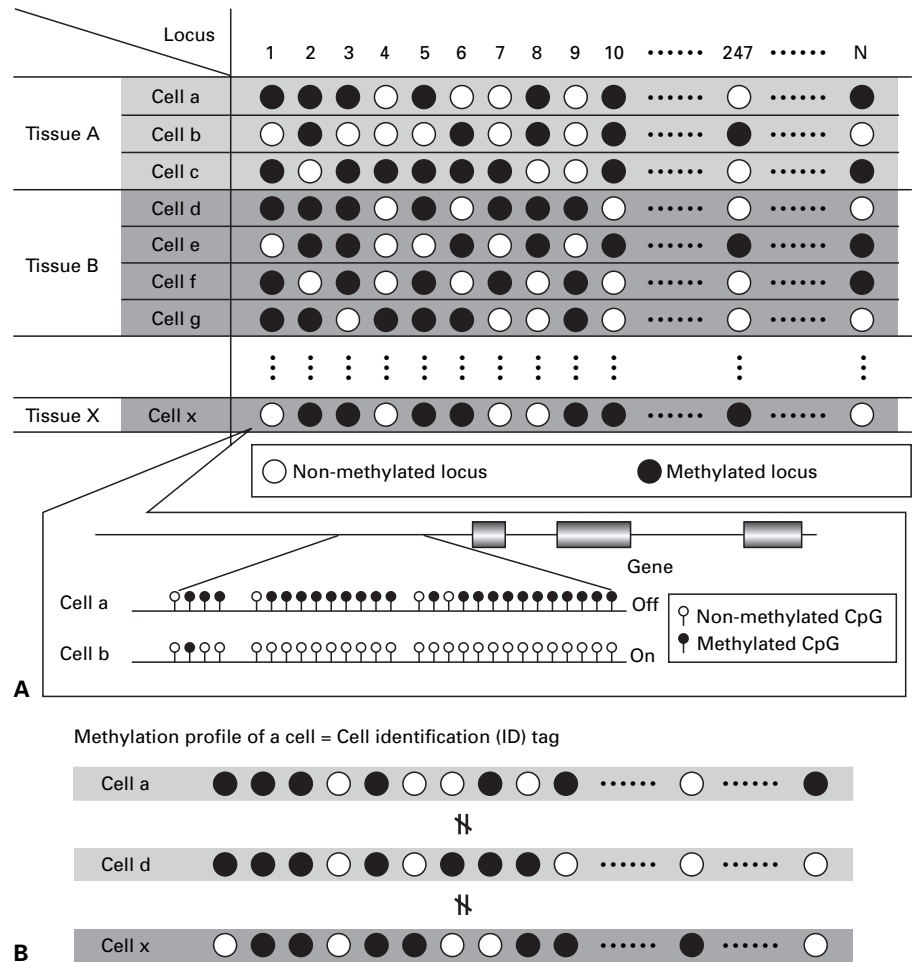
## DNA methylation profiles in normal and abnormal development

Activation of a particular set of genes and inactivation of others underlies the differentiation of cells and mammalian development. Many genes have what are described as tissue-dependent and differentially-methylated regions, or T-DMRs (Shiota, 2004). DNA methylation at T-DMRs is involved in tissue-specific and developmentally-regulated gene expression. For example, *Pou5f1* (*Oct4*) is a member of the POU family of transcription factors, essential for normal mammalian development (Okamoto et al., 1990; Rosner et al., 1990; Scholer et al., 1990; Nichols et al., 1998; Niwa et al., 2000). DNA methylation of the *Oct4* T-DMR was identified by RLGS and found to play a critical role in silencing of the locus (Hattori et al., 2004). The *Oct4* gene has no CpG island at the transcription start site but has a relatively CG-rich and TATA-less promoter (Okazawa et al., 1991; Sylvester and Scholer, 1994). In the placenta of DNA methyltransferase 1 (*Dnmt1*)-deficient mice, most CpGs were unmethylated and *Oct4* was expressed ectopically, supporting a causal role for cytosine methylation in the regulation of this gene. Expression of *Sry* is also another example of DNA methylation of a T-DMR mediating gene silencing (Nishino et al., 2004). *Sry* is a master gene for testis differentiation in mammals (Koopman et al., 1991) and has a promoter region with rare CpGs. Certain prolactin/growth hormone gene family members also have promoters with CpGs but not enough to constitute a CpG island that are regulated by DNA methylation (Ngo et al., 1996; Cho et al., 2001). DNA methylation can therefore be involved in the gene regulation even when (or perhaps especially when) CpG dinucleotide density is insufficient to encode a CpG island.

Rat sphingosine kinase 1 (*Sphk1*) is an example of a gene for which tissue-specific expression is regulated by DNA methylation at a T-DMR (Imamura et al., 2001). The T-DMR is located approximately 800 bp upstream of the first exon of *Sphk1* in a 200-bp region at the 5' edge of a CpG island. The T-DMR of *Sphk1* is conserved in the human and mouse genomes and these T-DMRs are also targets of DNA methylation (Imamura et al., 2004). Other genes with T-DMR in CpG islands have also been described, including the *Ednrb* (endothelin receptor type B) (Pao et al., 2001), *Pomc1* (proopiomelanocortin) (Newell-Price et al., 2001) and *Serpinb5* (*Maspin*) (Futscher et al., 2002) genes. In addition, genome-wide screening identified CpG islands that are methylated in a manner reflecting their gametic origin, leading to their being termed germline differentially-methylated regions (gDMRs) (Strichman-Almashanu et al., 2002). These loci can be considered a further type of T-DMR occurring within CpG islands.

These loci represent the characterized subset of a much larger group of loci for which tissue-specific methylation has been identified using RLGS. A genome-wide analysis focusing on 1,500 CpG islands and CG-rich regions identified 247 T-DMRs, which were methylated or unmethylated depending on cell- or tissue-type (Okazaki et al., 1995; Ohgane et al., 1998, 2005; Shiota et al., 2002). The methylation profile of one cell or tissue type creates a profile for a cell or tissue type that distinguishes it from other cells or tissues (Fig. 1A). Con-





**Fig. 1.** DNA methylation profiles define cell types. **(A)** A given cell type has a unique pattern of DNA methylation at genes. This number is increasing as we study more and more loci to identify all genes regulated by DNA methylation. The methylation status of a gene is correlated with the transcriptional state of the gene. **(B)** The DNA methylation profile of one cell type is different from that of others and unique to the cell type, akin to an identification (ID) tag for the cell.

Considering that there are 16,100 CpG islands in the mouse haploid genome (March 2005/mm6 mouse genome assembly, UCSC Genome Browser, <http://genome.ucsc.edu>), of which RLGS can only sample a subset, the number of T-DMRs is likely to expand with further studies, allowing even more complex specific DNA methylation profiles. A further consequence of these studies is to replace the old idea of universal unmethylation of CpG islands with a model in which CpG methylation differences, even at CpG islands, mediate differences in cell states.

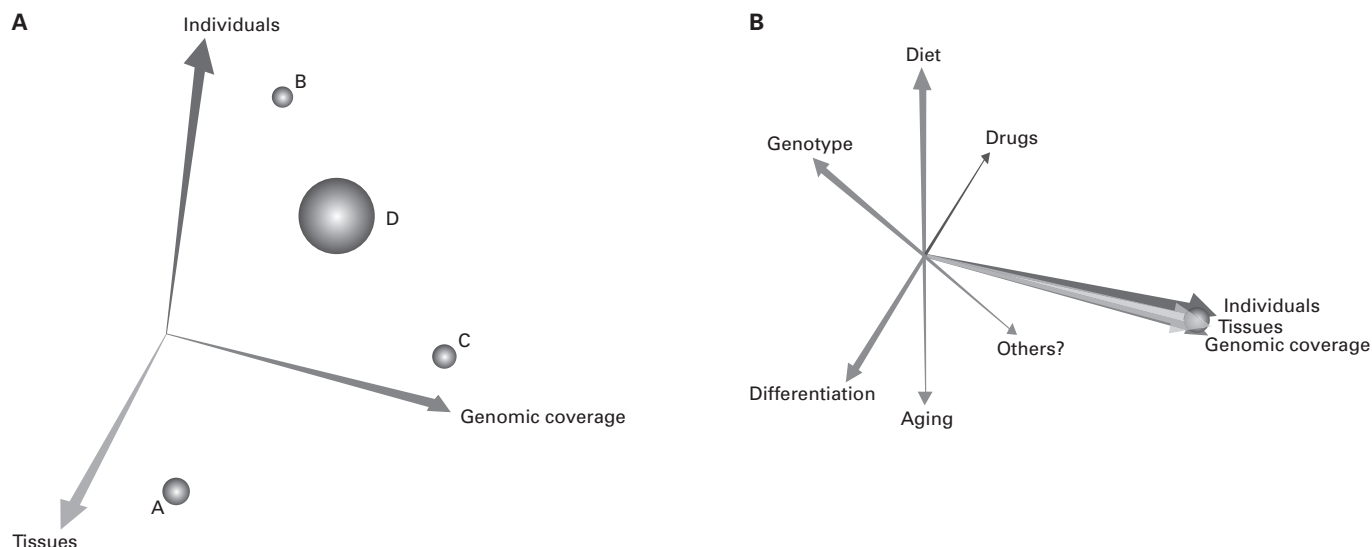
If the DNA methylation profile is a unique identification code for a cell (Fig. 1) and is involved in the regulation of gene expression, a change of the DNA methylation profile will cause alteration of the properties of the cell. Cloned animals created by somatic cell nuclear transfer have been found to have aberrant DNA methylation profiles in every tested tissue (Ohgane et al., 2001). As placentomegaly (placental overgrowth) is a frequently observed phenotype of cloned animals regardless of the origin or gender of donor cells (Ogura et al., 2000; Tanaka et al., 2001; Wakayama and Yanagimachi, 2001), RLGS was performed to determine whether this specific problem reflected abnormal epigenetic regulation. A specific pattern of aberrant DNA methylation at the *Sal13* locus

was found in the cloned mice exhibiting the placentomegaly (Ohgane et al., 2004).

With advances in genome-wide techniques to study epigenetic regulation, extending beyond cytosine methylation studies and testing more loci than allowed by RLGS, it is probable that our ability to detect variability associated with normal development, cloning, assisted reproductive technology and fetal exposures will be enhanced. A critical resource that will allow multiple investigators to compare their observations will be a common database structure allowing the DNA methylation and other epigenetic regulatory characteristics of genomic loci to be assembled.

#### *Epigenetics and cancer*

Although epigenetic alterations are being recognized to occur in disorders of development, the major current example of a disease state involving disordered epigenetic regulation is carcinogenesis (Jones and Baylin, 2002). Aberrant methylation of CpG islands in promoter regions is involved in inactivation of various tumor-suppressor genes, such as *RBI*, *p16 (CDKN2A)*, *VHL*, *MLH1*, E-cadherin (*CDH1*) and *BRCA1*, processes that occur in a number of cancers, including colon, bladder, stomach, liver, breast, uterine, and renal



**Fig. 2.** A Human Epigenome Project is an inherently complex undertaking. In this review, we discuss the multiple means by which epigenetic organization of the genome can be studied. Our emphasis on a whole-genome approach is represented in panel **A** by the position C on the graph – a focus on studying the whole genome as opposed to fewer loci in multiple tissues (A) or in multiple individuals (B). The ideal situation is represented by position D – capturing information genome-wide from many tissues and individuals. However, even this ambitious undertaking neglects to take into account the further dimensions shown in panel **B**, each representing currently-recognized influences on epigenetic organization, and also fails to take into account the multitude of genome-wide assays that can now be performed in each situation. This simple diagram illustrates the complexity inherent to a comprehensive Human Epigenome Project.

malignancies. In some cancers, such as gastric cancers, tumor suppressor genes are inactivated more frequently by their promoter methylation than by mutations (Ushijima and Sasako, 2004). At the same time, global genomic hypomethylation is occurring in most cancer cells (Feinberg and Tycko, 2004). Hypomethylation can lead to genomic instability and contribute to tumor formation (Gaudet et al., 2003). Hypomethylation of specific CpG islands causes aberrant expression of some cancer testis antigen genes, such as *MAGE* genes (De Smet et al., 1999) and can lead to loss of imprinting (LOI) (Feinberg and Tycko, 2004).

Epigenetic inactivation of *MLH1* induces microsatellite instability (Kane et al., 1997), while inactivation of *CHFR* induces loss of cell cycle checkpoint function (Mizuno et al., 2002). On the other hand, some cancer cells harbor intrinsic abnormalities that induce increased rates of de novo methylation (Ushijima et al., 2005). The increased rates lead to what has been described as the CpG island methylator phenotype (CIMP), which was originally reported in colorectal cancers (Toyota et al., 1999). However, there still remains a controversy over the presence of CIMP (Yamashita et al., 2003), and a genome-wide analysis of well-characterized CpG islands is awaited.

Factors that induce aberrant DNA methylation include aging and chronic inflammation and possibly viral infection (Issa et al., 1994, 2001). Cell division is considered to be a major factor in the induction of abnormal methylation (Velicescu et al., 2002). It has been proposed that decreased gene expression and methylation of scattered CpG sites ('seeds of

methylation') are involved in the induction of dense methylation of CpG islands (Song et al., 2002; De Smet et al., 2004; Ushijima and Okochi-Takada, 2005). The development of convenient assay systems is necessary to identify individual factors that induce DNA methylation, and the detailed molecular mechanisms for induction of aberrant methylation need to be understood.

The epigenetic alterations in cancer are now being used in cancer diagnosis and treatment (Miyamoto and Ushijima, 2005). For diagnostic purposes, aberrant DNA methylation can be used first to detect cancer cells in biopsy or laboratory specimens and cancer-derived free DNA in serum/plasma (Laird, 2003; Belinsky, 2004). Aberrant DNA methylation has an advantage over mutations because aberrant methylation can be detected sensitively, while targets for diagnosis can be readily identified by a genome-wide search for differences in DNA methylation (Ushijima, 2005). Secondly, aberrant DNA methylation can be used to predict a disease phenotype, such as prognosis, responses to chemotherapies, or occurrence of adverse effects. Methylation of *MGMT*, a gene involved in repair of O<sup>6</sup>-methylguanine, is a useful predictor of the responsiveness of tumors to alkylating agents in gliomas and diffuse large B-cell lymphomas (Esteller and Herman, 2004). The presence of CIMP in neuroblastomas is a strong predictor of poor clinical outcome (Abe et al., 2005). Thirdly, aberrant methylation in non-cancerous tissue has potential as a cancer risk marker. The presence of the LOI in normal colonic mucosa and peripheral leukocytes is associated with risk of colorectal cancers (Cui et al., 2003). An advantage of work-

ing with cytosine methylation as a diagnostic marker is that DNA is a more stable nucleic acid than RNA and is consequently less sensitive to specimen handling than RNA-based assays.

The plasticity of epigenetic information offers a good target for cancer therapeutics (Egger et al., 2004). Administration of the DNA demethylating agent 5-aza-deoxycytidine (5-aza-dC) at low doses for a prolonged period turned out to be more effective than higher doses in hematological malignancies (Issa et al., 2004). To overcome the short half-life of 5-aza-dC, the drug Zebularine, which can be administered orally, has been developed (Cheng et al., 2004). Histone deacetylases (HDACs), are also well studied as targets of therapeutics (Yoshida et al., 2001; Marks et al., 2004). Various HDAC inhibitors have been developed for therapeutic purposes, and tumor cells are known to show higher sensitivity to these agents than normal cells for reasons that remain unclear. Phase I/II trials are now under way for solid tumors (Marks et al., 2004).

### Genome-wide epigenetics 2005: towards a Human Epigenome Project

In this review, we have placed an emphasis on the technical advances in the analysis of the epigenetic organization of the whole genome. We also highlight the need to develop means of organization and analysis of epigenetic data. While we focused on the role of epigenetic dysregulation in developmental abnormalities and cancer, the list of disorders to which epigenetic abnormalities could be contributing is very

large. If resources are to be used with maximum efficiency to test how epigenetic regulation contributes to human disease, it has become clear that investigator-initiated research will need to be complemented by the kind of common resources that were used for the human genome project. These resources first and foremost should address the difficulties of data management, integration and analysis. In addition, just as technology innovation drove the human genome project, attention needs to be paid to the development of the technologies that will allow genome-wide epigenetics studies, especially applied to the limited numbers of cells that can be isolated to a high degree of purity by techniques such as laser capture microscopy. The goal is simple – to define the role of epigenetics in human disease, allowing new insights, diagnostic tests and drug targets. Current efforts to create a coherent Human Epigenome Project include a recent workshop sponsored by the American Association for Cancer Research (Jones and Martienssen, 2005). While such an undertaking is daunting in its complexity (Fig. 2), the goal is unquestionably worthwhile.

### Acknowledgements

The contributions of the following speakers to the Whole-Genome Epigenetics 2005 conference are gratefully acknowledged: Drs. Kuniya Abe, Hiroyuki Aburatani, Izuhito Hatada, Tim Hui-Ming Huang, Fumio Ishino, Peter A. Jones, Tetsuji Kakutani, Hiroshi Kataoka, Takeo Kubota, Hiroki Kurihara, Robert Martienssen, Mitsuyoshi Nakao, Mikio Nakazono, Masaki Okano, Mitsuo Oshimura, Hiroyuki Sasaki, Toru Shimada, Yoichi Shinkai, Masakazu Sugiyama, Kazunari Taira, Shoji Tajima, Satoshi Tanaka, Nobuhiro Tsutsumi, Shintaro Yagi and Minoru Yoshida.

### References

- Abe M, Ohira M, Kaneda A, Yagi Y, Yamamoto S, Kitano Y, Takato T, Nakagawara A, Ushijima T: CpG island methylator phenotype is a strong determinant of poor prognosis in neuroblastomas. *Cancer Res* 65:828–834 (2005).
- Antequera F, Bird A: Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA* 90:11995–11999 (1993).
- Bailey TL, Elkan CP: Fitting a mixture model by expectation-maximization to discover motifs in biopolymers. *ISMB*:28–36 (1994).
- Belinsky SA: Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer* 4:707–717 (2004).
- Bell AC, Felsenfeld G: Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405:482–485 (2000).
- Ben-Porath I, Cedar H: Epigenetic crosstalk. *Mol Cell* 8:933–935 (2001).
- Ben-Yehuda S, Fujita M, Liu XS, Gorbatyuk B, Skoko D, Yan J, Marko JF, Liu JS, Eichenberger P, Rudner DZ, et al: Defining a centromere-like element in *Bacillus subtilis* by identifying the binding sites for the chromosome-anchoring protein RacA. *Mol Cell* 17:773–782 (2005).
- Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL: Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci USA* 99:8695–8700 (2002).
- Bernstein BE, Humphrey EL, Liu CL, Schreiber SL: The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Meth Enzymol* 376:349–360 (2004a).
- Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL: Global nucleosome occupancy in yeast. *Genome Biol* 5:R62 (2004b).
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, et al: Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120:169–181 (2005).
- Bestor TH: The DNA methyltransferases of mammals. *Hum Mol Genet* 9:2395–2402 (2000).
- Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN, Weintraub H: Sequence-specific DNA binding by the c-Myc protein. *Science* 250:1149–1151 (1990).
- Boguski MS: ENCODE and ChIP-chip in the genome era. *Genomics* 83:347–348 (2004).
- Bradbury J: Human epigenome project-up and running. *PLoS Biol* 1:E82 (2003).
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, et al: Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630–634 (2000).
- Buck MJ, Lieb JD: ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83:349–360 (2004).
- Bulyk ML: Computational prediction of transcription-factor binding site locations. *Genome Biol* 5:201 (2003).
- Bulyk ML, Huang X, Choo Y, Church GM: Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* 98:7158–7163 (2001).
- Bussemaker HJ, Li H, Siggia ED: Regulatory element detection using correlation with expression. *Nat Genet* 27:167–171 (2001).
- Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, et al: Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122:33–43 (2005).
- Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P: Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32:623–626 (2002).
- Casolari JM, Brown CR, Komili S, West J, Hieronymus H, Silver PA: Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* 117:427–439 (2004).

- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al: Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116:499–509 (2004).
- Chen J, Sadowski I: Identification of the mismatch repair genes *PMS2* and *MLH1* as p53 target genes by using serial analysis of binding elements. *Proc Natl Acad Sci USA* 102:4813–4818 (2005).
- Cheng JC, Yoo CB, Weisenberger DJ, Chuang J, Wozniak C, Liang G, Marquez VE, Greer S, Orntoft TF, Thykjaer T, et al: Preferential response of cancer cells to zebularine. *Cancer Cell* 6:151–158 (2004).
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al: Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308:1149–1154 (2005).
- Ching TT, Maunakea AK, Jun P, Hong C, Zardo G, Pinkel D, Albertson DG, Fridlyand J, Mao JH, Shchors K, et al: Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of *SHANK3*. *Nat Genetics* 37:645–651 (2005).
- Cho JH, Kimura H, Minami T, Ohgane J, Hattori N, Tanaka S, Shiota K: DNA methylation regulates placental lactogen I gene expression. *Endocrinology* 142:3389–3396 (2001).
- Clark SJ, Harrison J, Frommer M: CpNpG methylation in mammalian cells. *Nat Genet* 10:20–27 (1995).
- Conlon EM, Liu XS, Lieb JD, Liu JS: Integrating regulatory motif discovery and genome-wide expression analysis. *Proc Natl Acad Sci USA* 100:3339–3344 (2003).
- Cosma MP, Tanaka T, Nasmyth K: Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* 97:299–311 (1999).
- Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, Feinberg AP: Loss of *IGF2* imprinting: a potential marker of colorectal cancer risk. *Science* 299:1753–1755 (2003).
- De Smet C, Lurquin C, Lethe B, Martelange V, Boon T: DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* 19:7327–7335 (1999).
- De Smet C, Loriot A, Boon T: Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene *MAGE-A1* in tumor cells. *Mol Cell Biol* 24:4781–4790 (2004).
- Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 295:1306–1311 (2002).
- Dion MF, Altschuler SJ, Wu LF, Rando OJ: Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci USA* 102:5501–5506 (2005).
- Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, Mack J, Hall R, Goldy J, Sabo PJ, et al: High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* 1:219–225 (2004).
- Editorial: Getting hip to the chip. *Nat Genet* 18:195–197 (1998).
- Egger G, Liang G, Aparicio A, Jones PA: Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457–463 (2004).
- El-Osta A, Wolffe AP: DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease. *Gene Expr* 9:63–75 (2000).
- ENCODE Project Consortium: The ENCODE (Encyclopedia Of DNA Elements) Project. *Science* 306:636–640 (2004).
- Esteller M, Herman JG: Generating mutations but providing chemosensitivity: the role of O<sup>6</sup>-methylguanine DNA methyltransferase in human cancer. *Oncogene* 23:1–8 (2004).
- Euskirchen G, Royce TE, Bertone P, Martone R, Rinn JL, Nelson FK, Sayward F, Luscombe NM, Miller P, Gerstein M, et al: CREB binds to multiple loci on human chromosome 22. *Mol Cell Biol* 24:3804–3814 (2004).
- Farnham PJ: In vivo assays to examine transcription factor localization and target gene specificity. *Methods* 26:1–2 (2002).
- Fazzari MJ, Gready JM: Epigenomics: beyond CpG islands. *Nat Rev Genet* 5:446–455 (2004).
- Feinberg AP, Tycko B: The history of cancer epigenetics. *Nat Rev Cancer* 4:143–153 (2004).
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al: Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102:10604–10609 (2005).
- Fried M, Crothers DM: Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 9:6505–6525 (1981).
- Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, Domann FE: Role for DNA methylation in the control of cell type specific maspin expression. *Nat Genet* 31:175–179 (2002).
- Gardiner-Garden M, Frommer M: CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282 (1987).
- Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R: Induction of tumors in mice by genomic hypomethylation. *Science* 300:489–492 (2003).
- Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA: Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118:555–566 (2004).
- Glynn EF, Megee PC, Yu HG, Mistrot C, Unal E, Koshland DE, DeRisi JL, Gerton JL: Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS Biol* 2:E259 (2004).
- Gordon DB, Nekudova L, McCallum S, Fraenkel E: TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. *Bioinformatics* 21:3164–3165 (2005).
- Gready JM, Starr DJ, Hwang S, Song L, Jaarola M, Zemel S: The mouse *H19* locus mediates a transition between imprinted and non-imprinted DNA replication patterns. *Hum Mol Genet* 7:91–96 (1998).
- Greif F, van der Kraan I, Delrow J, Smothers JF, de Wit E, Bussemaker HJ, van Driel R, Henikoff S, van Steensel B: Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes Dev* 17:2825–2838 (2003).
- Grundy WN, Bailey TL, Elkan CP: ParaMEME: a parallel implementation and a web interface for a DNA and protein motif discovery tool. *Comput Appl Biosci* 12:303–310 (1996).
- Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E, Young RA: Global and Hox-specific roles for the MLL1 methyltransferase. *Proc Natl Acad Sci USA* 102:8603–8608 (2005).
- Hampsey M, Reinberg D: Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* 113:429–432 (2003).
- Hanlon SE, Lieb JD: Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr Opin Genet Dev* 14:697–705 (2004).
- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, et al: Transcriptional regulatory code of a eukaryotic genome. *Nature* 431:99–104 (2004).
- Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L, Craig JM, Jones PL, Sif S, et al: Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* 37:254–264 (2005).
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM: CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* 405:486–489 (2000).
- Hattori N, Nishino K, Ko YG, Hattori N, Ohgane J, Tanaka S, Shiota K: Epigenetic control of mouse *Oct-4* gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem* 279:17063–17069 (2004).
- Horak CE, Snyder M: ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol* 350:469–483 (2002).
- Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SC, Snyder M: GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIP-chip analysis. *Proc Natl Acad Sci USA* 99:2924–2929 (2002).
- Hu M, Yao J, Cai L, Bachman KE, van den Brule F, Velculescu V, Polyak K: Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet* 37:899–905 (2005).
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, et al: Functional discovery via a compendium of expression profiles. *Cell* 102:109–126 (2000).
- Humphrey EL, Shamji AF, Bernstein BE, Schreiber SL: Rpd3p relocation mediates a transcriptional response to rapamycin in yeast. *Chem Biol* 11:295–299 (2004).
- Im H, Grass JA, Johnson KD, Boyer ME, Wu J, Bresnick EH: Measurement of protein-DNA interactions in vivo by chromatin immunoprecipitation. *Methods Mol Biol* 284:129–146 (2004).
- Imamura T, Ohgane J, Ito S, Ogawa T, Hattori N, Tanaka S, Shiota K: CpG island of rat sphingosine kinase-1 gene: tissue-dependent DNA methylation status and multiple alternative first exons. *Genomics* 76:117–125 (2001).
- Imamura T, Miyauchi-Senda N, Tanaka S, Shiota K: Identification of genetic and epigenetic similarities of *SPHK1/Sphk1* in mammals. *J Vet Med Sci* 66:1387–1393 (2004).
- Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH: Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* 119:1041–1054 (2004).
- Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB: Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* 7:536–540 (1994).
- Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA: Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res* 61:3573–3577 (2001).

- Issa JP, Garcia-Manero G, Giles FJ, Mannari R, Thomas D, Faderl S, Bayar E, Lyons J, Rosenfeld CS, Cortes J, et al: Phase I study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood* 103: 1635–1640 (2004).
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409:533–538 (2001).
- Jeffery L, Nakielny S: Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem* 279:49479–49487 (2004).
- Jenuwein T, Allis CD: Translating the histone code. *Science* 293:1074–1080 (2001).
- Johnson KD, Bresnick EH: Dissecting long-range transcriptional mechanisms by chromatin immunoprecipitation. *Methods* 26:27–36 (2002).
- Jones PA, Baylin SB: The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3:415–428 (2002).
- Jones PA, Martienssen RA: Blueprint for a human epigenome project (HEP). *Cancer Res* 65:11241–11246 (2005).
- Jorgensen HF, Bird A: MeCP2 and other methyl-CpG binding proteins. *Ment Retard Dev Disabil Res Rev* 8:87–93 (2002).
- Kagotani K, Takebayashi S, Kohda A, Taguchi H, Paulsen M, Walter J, Reik W, Okumura K: Replication timing properties within the mouse distal chromosome 7 imprinting cluster. *Biosci Biotechnol Biochem* 66:1046–1051 (2002).
- Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R: Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* 57:808–811 (1997).
- Kieser JA: Epigenetic canalization and phenotypic change: a minimax model. *Med Hypotheses* 22: 105–110 (1987).
- Kim J, Bhinge AA, Morgan XC, Iyer VR: Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat Methods* 2:47–53 (2005).
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: A high-resolution map of active promoters in the human genome. *Nature* 436:876–880 (2005).
- Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R: Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351:117–121 (1991).
- Kremenskoy M, Kremenska Y, Ohgane J, Hattori N, Tanaka S, Hashizume K, Shiota K: Genome-wide analysis of DNA methylation status of CpG islands in embryoid bodies, teratomas, and fetuses. *Biochem Biophys Res Commun* 311:884–890 (2003).
- Kuo MH, Allis CD: In vivo cross-linking and immunoprecipitation for studying dynamic protein: DNA associations in a chromatin environment. *Methods* 19:425–433 (1999).
- Kurdistani SK, Grunstein M: In vivo protein-protein and protein-DNA crosslinking for genomewide binding microarray. *Methods* 31:90–95 (2003).
- Kurdistani SK, Robyr D, Tavazoie S, Grunstein M: Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet* 31:248–254 (2002).
- Kurdistani SK, Tavazoie S, Grunstein M: Mapping global histone acetylation patterns to gene expression. *Cell* 117:721–733 (2004).
- Laird PW: The power and the promise of DNA methylation markers. *Nat Rev Cancer* 3:253–266 (2003).
- Lebrun E, Fourel G, Defossez PA, Gilson E: A methyltransferase targeting assay reveals silencer-telomere interactions in budding yeast. *Mol Cell Biol* 23:1498–1508 (2003).
- Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD: Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36:900–905 (2004).
- Lee TI, Young RA: Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* 34:77–137 (2000).
- Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al: Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799–804 (2002).
- Lengronne A, Katou Y, Mori S, Yokobayashi S, Kelly GP, Itoh T, Watanabe Y, Shirahige K, Uhlmann F: Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 430:573–578 (2004).
- Li E, Beard C, Jaenisch R: Role for DNA methylation in genomic imprinting. *Nature* 366:362–365 (1993).
- Li W, Meyer CA, Liu XS: A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics* 21 Suppl 1:i274–i282 (2005).
- Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B: A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci USA* 100:8164–8169 (2003).
- Lieb JD: Genome-wide mapping of protein-DNA interactions by chromatin immunoprecipitation and DNA microarray hybridization. *Methods Mol Biol* 224:99–109 (2003).
- Lieb JD, Liu X, Botstein D, Brown PO: Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* 28:327–334 (2001).
- Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al: Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476 (2004).
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ: Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3:e328 (2005).
- Liu X, Brutlag DL, Liu JS: BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*:127–138 (2001).
- Liu XS, Brutlag DL, Liu JS: An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* 20:835–839 (2002).
- Liu X, Noll DM, Lieb JD, Clarke ND: DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res* 15:421–427 (2005).
- Loden M, van Steensel B: Whole-genome views of chromatin structure. *Chromosome Res* 13:289–298 (2005).
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380 (2005).
- Marks PA, Richon VM, Miller T, Kelly WK: Histone deacetylase inhibitors. *Adv Cancer Res* 91:137–168 (2004).
- Marshall E: A strategy for sequencing the genome 5 years early. *Science* 267:783–784 (1995).
- Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, et al: Distribution of NF-kappaB binding sites across human chromosome 22. *Proc Natl Acad Sci USA* 100:12247–12252 (2003).
- Matsuyama T, Kimura MT, Koike K, Abe T, Nakano T, Asami T, Ebisuzaki T, Held WA, Yoshida S, Nagase H: Global methylation screening in the *Arabidopsis thaliana* and *Mus musculus* genome: applications of virtual image restriction landmark genomic scanning (Vi-RLGS). *Nucleic Acids Res* 31:4490–4496 (2003).
- Mito Y, Henikoff JG, Henikoff S: Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37:1090–1097 (2005).
- Miyamoto K, Ushijima T: Diagnostic and therapeutic applications of epigenetics. *Jpn J Clin Oncol* 35:293–301 (2005).
- Mizuno K, Osada H, Konishi H, Tatematsu Y, Yatabe Y, Mitsudomi T, Fujii Y, Takahashi T: Aberrant hypermethylation of the CHFR prophase checkpoint gene in human lung cancers. *Oncogene* 21:2328–2333 (2002).
- Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA, Bulyk ML: Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat Genet* 36:1331–1339 (2004).
- Mukhopadhyay R, Yu W, Whitehead J, Xu J, Lezcano M, Pack S, Kanduri C, Kanduri M, Ginjala V, Vostrov A, et al: The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res* 14:1594–1602 (2004).
- Murrell A, Rakan VK, Beck S: From genome to epigenome. *Hum Mol Genet* 14 Spec No 1:R3-R10 (2005).
- Nal B, Mohr E, Ferrier P: Location analysis of DNA-bound proteins at the whole-genome level: untangling transcriptional regulatory networks. *Bioessays* 23:473–476 (2001).
- Newell-Price J, King P, Clark AJ: The CpG island promoter of the human proopiomelanocortin gene is methylated in nonexpressing normal tissue and tumors and represses expression. *Mol Endocrinol* 15:338–348 (2001).
- Ng HH, Robert F, Young RA, Struhl K: Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11: 709–719 (2003).
- Ng P, Wei CL, Sung WK, Chiu KP, Lipovich L, Ang CC, Gupta S, Shahab A, Ridwan A, Wong CH, et al: Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat Methods* 2:105–111 (2005).
- Ngo V, Gourdji D, Laverriere JN: Site-specific methylation of the rat prolactin and growth hormone promoters correlates with gene expression. *Mol Cell Biol* 16:3245–3254 (1996).
- Nguyen CT, Gonzales FA, Jones PA: Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. *Nucleic Acids Res* 29:4598–4606 (2001).
- Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95:379–391 (1998).
- Nishino K, Hattori N, Tanaka S, Shiota K: DNA methylation-mediated control of *Sry* gene expression in mouse gonadal development. *J Biol Chem* 279:22306–22313 (2004).
- Niwa H, Miyazaki J, Smith AG: Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 24:372–376 (2000).
- Noble WS, Kuehn S, Thurman R, Yu M, Stamatoyanopoulos J: Predicting the in vivo signature of human gene regulatory sequences. *Bioinformatics* 21 Suppl 1:i338–i343 (2005).

- Novik KL, Nimmrich I, Genc B, Maier S, Piepenbrock C, Olek A, Beck S: Epigenomics: genome-wide study of methylation phenomena. *Curr Issues Mol Biol* 4:111–128 (2002).
- Oberley MJ, Farnham PJ: Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins. *Methods Enzymol* 371:577–596 (2003).
- Oberley MJ, Tsao J, Yau P, Farnham PJ: High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. *Methods Enzymol* 376:315–334 (2004).
- Ogura A, Inoue K, Ogonuki N, Noguchi A, Takano K, Nagano R, Suzuki O, Lee J, Ishino F, Matsuda J: Production of male cloned mice from fresh, cultured, and cryopreserved immature Sertoli cells. *Biol Reprod* 62:1579–1584 (2000).
- Ohgane J, Aikawa J, Ogura A, Hattori N, Ogawa T, Shiota K: Analysis of CpG islands of trophoblast giant cells by restriction landmark genomic scanning. *Dev Genet* 22:132–140 (1998).
- Ohgane J, Wakayama T, Kogo Y, Senda S, Hattori N, Tanaka S, Yanagimachi R, Shiota K: DNA methylation variation in cloned mice. *Genesis* 30:45–50 (2001).
- Ohgane J, Wakayama T, Senda S, Yamazaki Y, Inoue K, Ogura A, Marh J, Tanaka S, Yanagimachi R, Shiota K: The *Sal3* locus is an epigenetic hotspot of aberrant DNA methylation associated with placentomegaly of cloned mice. *Genes Cells* 9:253–260 (2004).
- Ohgane J, Hattori N, Shiota K: Analysis of tissue-specific DNA methylation during development. *Methods Mol Biol* 289:371–382 (2005).
- Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H: A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60:461–472 (1990).
- Okazaki Y, Okuizumi H, Sasaki N, Ohsumi T, Kuromitsu J, Hirota N, Muramatsu M, Hayashizaki Y: An expanded system of restriction landmark genomic scanning (RLGS Ver. 1.8). *Electrophoresis* 16:197–202 (1995).
- Okazawa H, Okamoto K, Ishino F, Ishino-Kaneko T, Takeda S, Toyoda Y, Muramatsu M, Hamada H: The *oct3* gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *EMBO J* 10:2997–3005 (1991).
- Oliphant AR, Brandl CJ, Struhl K: Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol Cell Biol* 9:2944–2949 (1989).
- Orlando V: Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* 25:99–104 (2000).
- Pao MM, Tsutsumi M, Liang G, Uzvolgyi E, Gonzales FA, Jones PA: The endothelin receptor B (EDNRB) promoter displays heterogeneous, site specific methylation patterns in normal and tumor cells. *Hum Mol Genet* 10:903–910 (2001).
- Philippakis AA, He FS, Bulyk ML: Modulefinder: a tool for computational discovery of cis regulatory modules. *Pac Symp Biocomput*:519–530 (2005).
- Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A, et al: DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2:e405 (2004).
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, et al: Genome-wide location and function of DNA binding proteins. *Science* 290:2306–2309 (2000).
- Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD: E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* 16:245–256 (2002).
- Robert F, Pokholok DK, Hannett NM, Rinaldi NJ, Chandy M, Rolfe A, Workman JL, Gifford DK, Young RA: Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell* 16:199–209 (2004).
- Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, Suka N, Grunstein M: Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* 109:437–446 (2002).
- Robyr D, Kurdistani SK, Grunstein M: Analysis of genome-wide histone acetylation state and enzyme binding using DNA microarrays. *Methods Enzymol* 376:289–304 (2004).
- Roh TY, Ngau WC, Cui K, Landsman D, Zhao K: High-resolution genome-wide mapping of histone modifications. *Nat Biotechnol* 22:1013–1016 (2004).
- Roh TY, Cuddapah S, Zhao K: Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 19:542–552 (2005).
- Rosner MH, Vigano MA, Ozato K, Timmons PM, Poirier F, Rigby PW, Staudt LM: A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* 345:686–692 (1990).
- Roth FP, Hughes JD, Estep PW, Church GM: Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol* 16:939–945 (1998).
- Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, McArthur M, Stamatoyannopoulos JA: Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc Natl Acad Sci USA* 101:4537–4542 (2004).
- Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW: Sex differential in methylation patterns of selected genes in Singapore Chinese. *Hum Genet* 117:402–403 (2005).
- Scholer HR, Ruppert S, Suzuki N, Chowdhury K, Gruss P: New type of POU domain in germ line-specific protein Oct-4. *Nature* 344:435–439 (1990).
- Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, et al: The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 18:1263–1271 (2004).
- Shiota K: DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. *Cytogenet Genome Res* 105:325–334 (2004).
- Shiota K, Kogo Y, Ohgane J, Imamura T, Urano A, Nishino K, Tanaka S, Hattori N: Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells* 7:961–969 (2002).
- Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, et al: Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* 106:697–708 (2001).
- Smith CD, Smith DL, DeRisi JL, Blackburn EH: Telomeric protein distributions and remodeling through the cell cycle in *Saccharomyces cerevisiae*. *Mol Biol Cell* 14:556–570 (2003).
- Solomon MJ, Varshavsky A: Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci USA* 82:6470–6474 (1985).
- Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA: Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc Natl Acad Sci USA* 102:3336–3341 (2005).
- Song JZ, Stirzaker C, Harrison J, Melki JR, Clark SJ: Hypermethylation trigger of the glutathione-S-transferase gene (*GSTP1*) in prostate cancer cells. *Oncogene* 21:1048–1061 (2002).
- Spencer VA, Sun JM, Li L, Davie JR: Chromatin immunoprecipitation: a tool for studying histone acetylation and transcription factor binding. *Methods* 31:67–75 (2003).
- Strichman-Almashanu LZ, Lee RS, Onyango PO, Perlman E, Flam F, Frieman MB, Feinberg AP: A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res* 12:543–554 (2002).
- Sylvester I, Scholer HR: Regulation of the Oct-4 gene by nuclear receptors. *Nucleic Acids Res* 22:901–911 (1994).
- Takai D, Jones PA: Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 99:3740–3745 (2002).
- Takebayashi S, Sugimura K, Saito T, Sato C, Fukushima Y, Taguchi H, Okumura K: Regulation of replication at the R/G chromosomal band boundary and pericentromeric heterochromatin of mammalian cells. *Exp Cell Res* 304:162–174 (2005).
- Tanaka S, Oda M, Toyoshima Y, Wakayama T, Tanaka M, Yoshida N, Hattori N, Ohgane J, Yanagimachi R, Shiota K: Placentomegaly in cloned mouse concepti caused by expansion of the spongiotrophoblast layer. *Biol Reprod* 65:1813–1821 (2001).
- Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP: CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 96:8681–8686 (1999).
- Tuerk C, Gold L: Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510 (1990).
- Ushijima T: Detection and interpretation of altered methylation patterns in cancer cells. *Nat Rev Cancer* 5:223–231 (2005).
- Ushijima T, Okochi-Takada E: Aberrant methylations in cancer cells: Where do they come from? *Cancer Sci* 96:206–211 (2005).
- Ushijima T, Sasako M: Focus on gastric cancer. *Cancer Cell* 5:121–125 (2004).
- Ushijima T, Watanabe N, Shimizu K, Miyamoto K, Sugimura T, Kaneda A: Decreased fidelity in replicating CpG methylation patterns in cancer cells. *Cancer Res* 65:11–17 (2005).
- van Leeuwen F, van Steensel B: Histone modifications: from genome-wide maps to functional insights. *Genome Biol* 6:113 (2005).
- van Steensel B: Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat Genet* 37 Suppl:S18–24 (2005).
- van Steensel B, Henikoff S: Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* 18:424–428 (2000).
- van Steensel B, Henikoff S: Epigenomic profiling using microarrays. *Biotechniques* 35:346–350, 352–354, 356–357 (2003).
- van Steensel B, Delrow J, Henikoff S: Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* 27:304–308 (2001).

- Velicescu M, Weisenberger DJ, Gonzales FA, Tsai YC, Nguyen CT, Jones PA: Cell division is required for de novo methylation of CpG islands in bladder cancer cells. *Cancer Res* 62:2378–2384 (2002).
- Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D: RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303:672–676 (2004).
- Wade JT, Hall DB, Struhl K: The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature* 432:1054–1058 (2004).
- Wakayama T, Yanagimachi R: Mouse cloning with nucleus donor cells of different age and type. *Mol Reprod Dev* 58:376–383 (2001).
- Wang W, Cherry JM, Botstein D, Li H: A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99:16893–16898 (2002).
- Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D: Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37:853–862 (2005).
- Weber SA, Gerton JL, Polancic JE, DeRisi JL, Koshland D, Megee PC: The kinetochore is an enhancer of pericentric cohesin binding. *PLoS Biol* 2:E260 (2004).
- Weil MR, Widlak P, Minna JD, Garner HR: Global survey of chromatin accessibility using DNA microarrays. *Genome Res* 14:1374–1381 (2004).
- Weinmann AS, Farnham PJ: Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 26:37–47 (2002).
- Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ: Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 16:235–244 (2002).
- Wells J, Yan PS, Cechvala M, Huang T, Farnham PJ: Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* 22:1445–1460 (2003).
- Yamashita K, Dai T, Dai Y, Yamamoto F, Perucho M: Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* 4:121–131 (2003).
- Yoshida M, Furumai R, Nishiyama M, Komatsu Y, Nishino N, Horinouchi S: Histone deacetylase as a new target for cancer chemotherapy. *Cancer Chemother Pharmacol* 48 Suppl 1:S20–S26 (2001).
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ: Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309:626–630 (2005).