

# 6

## Genome-wide Mapping of Histone Modifications by GMAT

SURESH CUDDAPAH AND KEJI ZHAO\*

*Laboratory of Molecular Immunology, NHLBI, NIH, 9000 Rockville Pike,  
Bethesda, MD 20892, USA*

### Introduction

Recent studies have estimated the total number of genes in the human genome to be approximately 35 000 (Baltimore, 2001). This is surprisingly small with it being only twice the number found in the fruit fly, *Drosophila melanogaster* (Misra *et al.*, 2002). It is clear that the higher organisms, during evolution, have attained their enormous complexity not only by increasing the size of their genome but also by more complex mechanisms of gene regulation.

The eukaryotic genome is organized into chromatin. The nucleosome, the basic unit of chromatin, is composed of DNA wrapped around an octamer of the four core histone proteins, H2A, H2B, H3, and H4. The maintenance and control of epigenetic information in the nucleus are conferred by post-translational modification of histones and also by DNA methylation. While DNA methylation is generally believed to be involved in gene silencing, gene regulation by histone modification is a more dynamic process. The amino terminal histone tails, which extend outwards of the nucleosome, are subject to a variety of post-translational modifications, such as acetylation, methylation, phosphorylation, ubiquitination, and ADP ribosylation (Sims *et al.*, 2003). The accessibility of nucleosomal DNA to transcription factors is affected by the type and position of the modifications (Turner, 1993; Strahl and Allis, 2000). Acetylation of histones, for example, is generally associated with transcriptional activation (Wu and Grunstein, 2000). However, methylation of histones can have both activating and repressing effects. While lysine 4 tri-methylation of histone H3 correlates with transcriptional activation, lysine 9 and lysine 27

---

\*To whom correspondence may be addressed (zhaok@nhlbi.nih.gov)

---

Abbreviations: ChIP, chromatin immunoprecipitation; CNS, conserved non-coding sequence; GMAT, genome-wide mapping technique; H3-K9/K14 acetylation, histone H3 lysine 9 and 14 diacetylation; HS sites, hypersensitive sites; K, lysine; ORF, open reading frame; SAGE, serial analysis of gene expression; TCR, T cell receptor.

tri-methylation correlates with the repression of the associated gene (Kouzarides, 2002). These histone modifications are believed to constitute the 'histone code', which conceptualizes the type and combination of histone modifications leading to varied transcriptional output (Strahl and Allis, 2000; Turner, 2002).

A clearer understanding of epigenetic information and its regulation is essential in the light of the importance attributed to it with respect to cancer and many other diseases (Wolffe and Matzke, 1999; Feinberg and Tycko, 2004). Several large-scale analyses of histone modifications in various species have been undertaken recently (Robyr *et al.*, 2002; Schubeler *et al.*, 2004; Bernstein *et al.*, 2005; Xu *et al.*, 2005). In this review, we outline some of the methods used to identify histone modifications, most of which involve the chromatin immunoprecipitation-based techniques. We discuss in detail the genome-wide mapping technique (GMAT) that we have developed (Roh *et al.*, 2004). GMAT is a quantitative and unbiased global mapping technique that does not depend on pre-selected sequences. Here, the advantages of this method in genome-wide mapping and its applications and significant findings using this method have been discussed.

### **Chromatin immunoprecipitation (ChIP)-based assays**

Chromatin immunoprecipitation (ChIP) is the technique where DNA and protein interaction is stabilized by a covalent bond generated by crosslinking with formaldehyde. The crosslinked chromatin is fragmented, usually by sonication. Antibodies are used to pull down the proteins of interest and, along with them, the bound DNA. The crosslinking is then reversed and quantitative PCR carried out with specific primers to amplify pre-selected regions, thus giving the information regarding the binding of protein to the chromatin regions of interest. ChIP studies have shown that histones in the telomeric regions are hypoacetylated in relation to the active gene regions (Braunstein *et al.*, 1993). Such a strategy, though very powerful for analysing a single gene or a particular genomic region, has a severely restricted scope on a larger scale because of the fact that only a small subset of regions can realistically be analysed. The technique becomes much more powerful when it is combined with methods with which large-scale analysis of the chromatin immunoprecipitated DNA can be carried out.

### **High-throughput analysis of chromatin immunoprecipitated DNA**

Several methods that can be used for high-throughput analysis of ChIP DNA have been reported recently (Horak and Snyder, 2002; Liang *et al.*, 2004; Roh *et al.*, 2004, 2005; Kim *et al.*, 2005). Chromatin immunoprecipitation combined with DNA microarray detection (ChIP-on-chip) has been used successfully in several genome-wide studies to determine protein binding sites (Ren *et al.*, 2000; Iyer *et al.*, 2001). Using this method, histone modifications have been studied extensively in *Saccharomyces cerevisiae* (Bernstein *et al.*, 2002; Robyr *et al.*, 2002). The application of the ChIP-on-chip technique to analyse the histone modifications in yeast genome has been very successful due to the availability of DNA microarrays that cover most of the genome (Bernstein *et al.*, 2002; Kurdistani *et al.*, 2004). However, it is still difficult to analyse the whole human genome using the ChIP-on-chip

method because the available DNA microarrays cover only a small portion of the human genome. To analyse the true global histone modifications in the human genome, an unbiased genome-wide analysis method is needed. To this end, we have developed the genome-wide mapping technique (GMAT) (Roh *et al.*, 2004), which is a combination of chromatin immunoprecipitation and serial analysis of gene expression (SAGE).

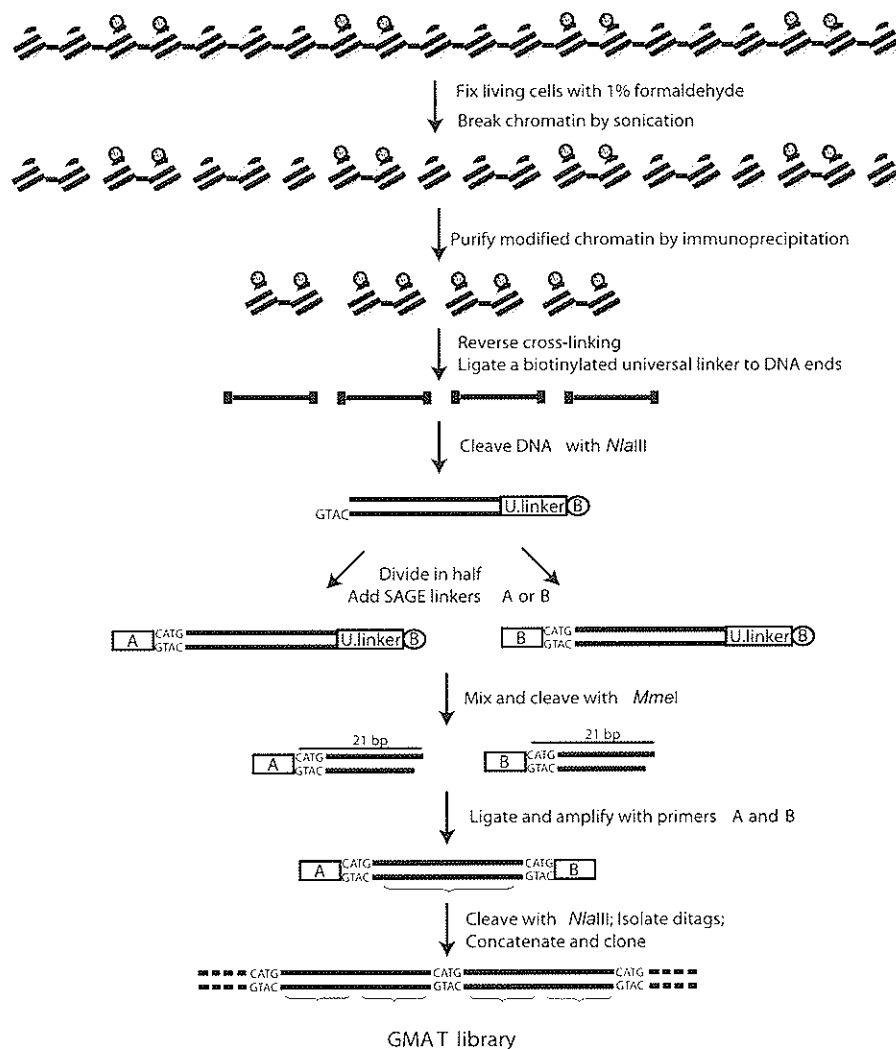
### Serial analysis of gene expression (SAGE)

Serial analysis of gene expression (SAGE) is a powerful technique that can be used to obtain a quantitative and comprehensive measure of mRNA levels in a given sample and that also can identify previously unknown transcripts (Velculescu *et al.*, 1995). SAGE is based on the principle that short sequence tags contain sufficient information for the identification of a transcript, and the number of tags from a particular transcript reflects the absolute abundance of the transcript in the genome. In SAGE, cDNA is synthesized from the mRNA isolated from cells. The cDNA is cleaved with an anchoring enzyme (*Nla*III) and a linker is ligated. A class II restriction enzyme that cleaves about 14 bp away from the recognition site is used to cleave the DNA to obtain short cDNA fragments, called SAGE tags. The tags, which are 9 to 10 bp long, are concatenated to form a long DNA molecule consisting of 25–50 tags, which are then sequenced. The sequence obtained is then analysed to identify the genes expressed and their level of expression. The original SAGE procedure used the class II restriction enzyme *Bsm*FI, which cleaves the cDNA 14 bp away from its recognition site, thus resulting in a 9–10 bp SAGE tag (Velculescu *et al.*, 1995). The SAGE procedure was later modified by replacing the enzyme *Bsm*FI with another class II restriction enzyme, *Mme*I, which cuts 21 bp away from the recognition site to generate 21 bp tags. This modified procedure, called the long SAGE, has improved the ability to unequivocally identify the gene (Saha *et al.*, 2002).

SAGE has been used widely to generate gene expression profiles in yeast (Velculescu *et al.*, 1997) and humans (Patino *et al.*, 2002). The application of SAGE has thus resulted in a better understanding of the molecular mechanisms involved in many different types of cancers, such as gastrointestinal (Zhang *et al.*, 1997), lung (Hibi *et al.*, 1998), pancreatic (Argani *et al.*, 2001), and thyroid (Takano *et al.*, 2000). A combination of SAGE and DNA microarray has been used to identify genes that are differentially expressed in breast cancer (Nacht *et al.*, 1999). Thus, SAGE is a valuable tool with diagnostic and therapeutic potential.

### Genome-wide mapping technique (GMAT)

The genome-wide mapping technique (GMAT) identifies the DNA immunoprecipitated with specific antibodies by long SAGE analysis. In the GMAT procedure (Figure 6.1), cells are crosslinked with formaldehyde to covalently stabilize histone binding to DNA, followed by sonication to fragment the DNA to 300–500 bp in size. The chromatin is then immunoprecipitated with specific antibodies to isolate the protein and the bound DNA. The DNA is purified after reversing the crosslinking. Following this, a biotinylated universal linker is ligated to the ChIP DNA ends. The DNA is then digested with the 4 bp cutter *Nla*III, which recognizes CATG sequence.



**Figure 6.1.** Schematic representation of genome-wide mapping technique (Roh *et al.*, 2004).

The linker-ligated DNA is isolated using streptavidin beads. The fragments thus obtained are divided in half, and SAGE linkers containing the restriction site for class II restriction enzyme *Mme*I is ligated to the *Nla*III digested ends. *Mme*I digestion of the DNA releases fragments consisting of the linker and a 21 bp tag from the ChIP DNA. The two sets of tags with linkers are mixed, ligated, and amplified by PCR using linker primers. Following amplification, the ditags are released by *Nla*III digestion. The GMAT library is generated by concatenating the ditags and cloning in a sequencing vector. SAGE 2000 and in-house software were used to extract the tag sequences from the sequencing data and to map the tags onto the chromosomes. The frequency of occurrence of a tag in the GMAT library is the direct and quantitative representation of the occurrence of the epitope recognized by the ChIP

antibody in that particular locus. To avoid mapping of a particular tag onto different regions of the genome due to the occurrence of repeat sequences, the observed tags have been normalized by dividing their number with the number of tag sites that occur in the genome.

### **5' ends of open reading frames (ORF) are highly acetylated in *Saccharomyces cerevisiae***

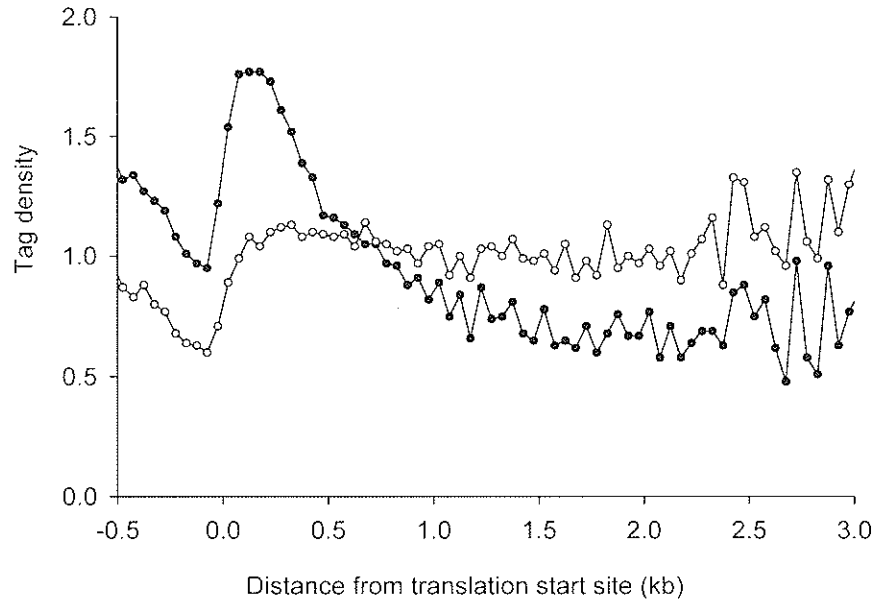
GMAT has been used to map the histone H3, and H4 acetylation of *S. cerevisiae* (Roh *et al.*, 2004). The analysis of 6000 yeast genes has shown the first 500 bases of the ORF to be the most hyperacetylated region in the genome, while the acetylation was much lower in the promoter. This can be explained in part by the studies on *PHO5* promoter of *S. cerevisiae*, where a dramatic loss of acetylation was observed upon induction of the promoter (Boeger *et al.*, 2003; Reinke and Horz, 2003). During chromatin remodelling, besides the changes in chromatin remodelling complexes, covalent modification of histones, such as acetylation, is an intermediate step. Removal of nucleosome marks the end point of the process of chromatin remodelling (Struhl, 1998). Nucleosome loss occurs broadly at many yeast promoters (Boeger *et al.*, 2003; Lee *et al.*, 2004), thus leading to a loss of histones, which accounts for the reduced acetylation seen in the promoter regions in *S. cerevisiae*.

### **GCN5 regulates H3 acetylation in the promoter and 5' coding region**

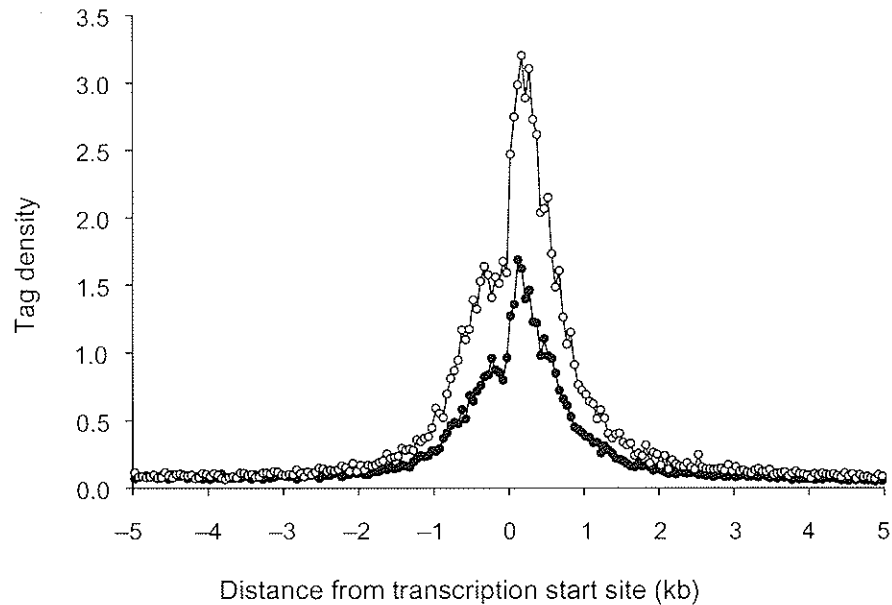
GCN5, a histone acetyl transferase, is a part of the yeast multisubunit complex, SAGA (Timmers and Tora, 2005). GCN5 plays a global role in regulating histone H3 acetylation, but not H4 acetylation (Vogelauer *et al.*, 2000). Deletion of *GCN5* has a profound effect on the acetylation status of the genome. Comparison of GMAT libraries created from the wild-type and *GCN5*-deleted *S. cerevisiae* revealed that the deletion of *GCN5* targets histone H3 acetylation in the promoter and the first 500 bp of the coding region (Figure 6.2). The acetylation peak observed in this region in the wild-type yeast was lost, and the acetylation was more uniformly distributed throughout the genome. Acetylation in the promoter region was also reduced by a large extent. Such a loss of acetylation was observed only in histone H3 and not in H4, which is not a substrate of the GCN5 enzyme (Roh *et al.*, 2004).

### **Promoter regions in the human genome are marked by histone hyperacetylation**

Most of the current knowledge about histone modifications and its regulation of gene expression is from the studies on lower eukaryotes like yeast. Using GMAT, we have analysed the H3 acetylation patterns of 21 355 annotated genes of the human genome by aligning their transcription start sites (Roh *et al.*, 2005). The study clearly showed histone hyperacetylation in the promoter region, which is defined as the region, 1 kb on either side of the transcription start site. The promoter region, which makes up only 0.8% of the genome, contains 23.5% of the acetylation tags in the GMAT library. The levels of acetylation dropped significantly and rapidly on



**Figure 6.2.** Deletion of *GCN5* led to reduced H3 acetylation in the promoter and 5' coding regions. 6040 yeast genes were aligned relative to their translation start sites and the normalized tag densities plotted. The acetylation peak observed in the 5' end of the coding region in the wild-type yeast (filled circle) is lost in the *GCN5* mutant (open circle).



**Figure 6.3.** Promoters of active genes have higher levels of H3 acetylation. 21 355 human genes were aligned relative to their transcription start sites. Active genes are depicted as open circles and inactive genes as filled circles.

either side of the promoter. Studies on 57 human genes by Liang and colleagues showed a similar enrichment of acetylation in the promoter region and a reduction away from the promoter (Liang *et al.*, 2004). Acetylation of active genes in the promoter region was much higher than those of the inactive genes (Figure 6.3), which is consistent with the results obtained from *Drosophila* cells (Schubeler *et al.*, 2004). Contrary to earlier studies that may have suggested the chromatin domain is uniformly acetylated when activated (Litt *et al.*, 2001), the high-resolution nature of the GMAT data indicates that the acetylation is very much localized to a few regions, even in active chromatin domains.

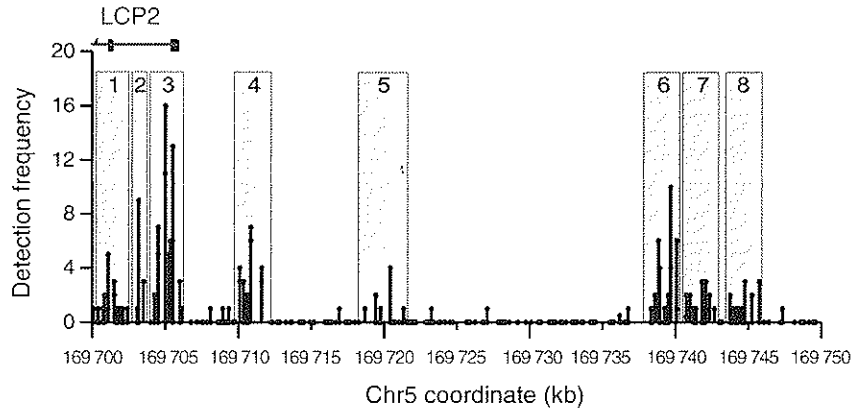
### **Histone acetylation could protect CpG islands from methylation**

There are 27 058 CpG islands in the human genome, which account for about 1% of the genomic DNA (Karolchik *et al.*, 2003). They are normally found in the promoter region of housekeeping genes and are unmethylated. When methylated, they are inhibitory to the genes associated with them. DNA methylation in CpG islands is a common feature within heterochromatin (Jaenisch and Bird, 2003). Histone modifications are believed to be involved in keeping the CpG islands unmethylated and the associated genes active (Fahrner *et al.*, 2002). GMAT analysis of the human genome has shown that most of the CpG islands are associated with acetylated histone H3. There appears a clear boundary between the CpG islands and their neighbouring sequences. High levels of histone acetylation are observed within the CpG islands. However, there is a marked reduction in the acetylation outside the CpG islands, which rapidly reaches the basal levels within a 200 bp region (Roh *et al.*, 2005).

Histone modification has been strongly implicated in influencing DNA methylation (Sims *et al.*, 2003). Histone H3–K9 methylation has been shown to provide a signal for DNA methylation and subsequent gene repression (Tamaru and Selker, 2001). Studies in mouse embryonic stem cells have shown that H3–K9 methylation is important for DNA methylation in pericentromeric satellite repeats and in imprinted genes (Lehnertz *et al.*, 2003; Xin *et al.*, 2003). Since H3–K9 acetylation is mutually exclusive with methylation, the H3–K9 acetylation in the CpG islands inhibits the K9 methylation event, and therefore the subsequent DNA methylation. Knowledge of distribution of other euchromatic modifications, such as H3–K4 methylation on the CpG islands, will provide further information on the mechanism involved in CpG methylation, and also whether there is any cooperativity of the various histone modifications in this phenomenon.

### **Acetylation islands mark regulatory elements in the genome**

While recent studies have clearly shown that the promoter regions are associated with high levels of histone acetylation, there has been no information as to the status of acetylation in the rest of the euchromatic region in the human genome. Does the distribution of acetylation follow any pattern in these regions? The high-resolution nature of the GMAT analysis provides interesting insights to the question. The analysis shows that many active loci in the human genome are decorated with discrete regions of acetylation, besides high levels of acetylation in the promoter



**Figure 6.4.** Chromatin domains have discrete regions of acetylation. Shown are the acetylation islands in resting human T cells. X-axis shows the chromosome coordinate and Y-axis shows the tag detection frequency. Acetylation islands are highlighted and numbered. The arrow indicates the position of the gene and its orientation.

region. These isolated pockets of acetylation, called ‘acetylation islands’, have high levels of acetylation compared to the surrounding regions (*Figure 6.4*). They totalled 21 481 in the intergenic regions and 25 332 in the transcribed regions of the human genome (Roh *et al.*, 2005). The specific distribution of the acetylation islands raises the possibility of these being involved in some kind of regulation of the gene. Comparative genomic analysis has identified approximately 240 000 conserved non-coding sequences (CNSs) in the human and mouse genomes (Hardison, 2000), which have been suggested to be regulatory elements. Comparison between the acetylation islands and CNSs indicates that most of the acetylation islands are associated with CNSs (Roh *et al.*, 2005). This extensive co-localization of acetylation islands with the CNSs indicates that the histone acetylation may mark functional transcriptional and/or chromatin regulatory elements. Indeed, most of the known transcriptional regulatory elements in T cells are co-localized with the acetylation islands. For example, all the well-characterized regulatory elements in the CD4, CD8, and IL2RA loci are correlated with histone acetylation, suggesting that histone acetylation is an epigenetic marker for functional regulatory elements.

#### **Induced acetylation islands could narrow down search for regulatory elements**

CNSs may constitute some of the potential regulatory elements in a genome. However, since a specific tissue/cell type expresses only a subset of the genes, the particular cell may use only a subset of the CNSs. A big challenge facing any investigator is to identify experimentally the functional CNSs in a particular cell type. Histone acetylation, with its well-established role in gene activation, is a potential index of events occurring in a cell. GMAT analysis of human T cells shows a genome-wide change in the histone H3 acetylation levels upon induction of the cells by TCR signalling. Increase in acetylation in ~4000 loci and decrease in acetylation in approximately the same number of loci have been observed. These



loci may play critical roles in mediating the TCR-signalling-induced gene activation and chromatin decondensation. For example, TCR signalling induces two acetylation islands in the Th2 locus, one of which is co-localized with a CNS required for the coordinated expression of the cytokine genes. Further studies have shown that both these induced islands function as transcriptional enhancers in reporter assays (Roh *et al.*, 2005). This shows that induced acetylation islands are valid epigenetic marks and could reveal potential regulatory regions.

### Concluding remarks

GMAT is an unbiased and high-resolution method of mapping genome-wide chromatin modifications. As we have shown, it is a powerful technique to determine true global patterns of epigenetic modifications. A similar method, SACO, has been used for the identification of CREB target sites in the rat genome (Impey *et al.*, 2004). Since the cost of sequencing is reducing rapidly because of the breakthroughs in sequencing technologies such as massive parallel signature sequencing, the potential for application of GMAT to study human epigenomes and their regulation will be unsurpassed.

### References

- ARGANI, P., ROSTY, C., REITER, R.E. *ET AL.* (2001). Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. *Cancer Research* **61**, 4320–4324.
- BALTIMORE, D. (2001). Our genome unveiled. *Nature* **409**, 814–816.
- BERNSTEIN, B.E., HUMPHREY, E.L., ERLICH, R.L. *ET AL.* (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8695–8700.
- BERNSTEIN, B.E., KAMAL, M., LINDBLAD-TOH, K. *ET AL.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181.
- BOEGER, H., GRIESENBECK, J., STRATTAN, J.S. AND KORNBURG, R.D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular Cell* **11**, 1587–1598.
- BRAUNSTEIN, M., ROSE, A.B., HOLMES, S.G., ALLIS, C.D. AND BROACH, J.R. (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes and Development* **7**, 592–604.
- FAHRNER, J.A., EGUCHI, S., HERMAN, J.G. AND BAYLIN, S.B. (2002). Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Research* **62**, 7213–7218.
- FEINBERG, A.P. AND TYCKO, B. (2004). The history of cancer epigenetics. *Nature Reviews Cancer* **4**, 143–153.
- HARDISON, R.C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends in Genetics* **16**, 369–372.
- HIBI, K., LIU, Q., BEAUDRY, G.A. *ET AL.* (1998). Serial analysis of gene expression in non-small cell lung cancer. *Cancer Research* **58**, 5690–5694.
- HORAK, C.E. AND SNYDER, M. (2002). Global analysis of gene expression in yeast. *Functional and Integrative Genomics* **2**, 171–180.
- IMPEY, S., MCCORKLE, S.R., CHA-MOLSTAD, H. *ET AL.* (2004). Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**, 1041–1054.
- IYER, V.R., HORAK, C.E., SSAFE, C.S., BOTSTEIN, D., SNYDER, M. AND BROWN, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538.
- JAENISCH, R. AND BIRD, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* **33** Suppl., 245–254.

- KAROLCHIK, D., BAERTSCH, R., DIEKHANS, M. *ET AL.* (2003). The UCSC Genome Browser Database. *Nucleic Acids Research* **31**, 51–54.
- KIM, T.H., BARRERA, L.O., ZHENG, M. *ET AL.* (2005). A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880.
- KOUZARIDES, T. (2002). Histone methylation in transcriptional control. *Current Opinion in Genetics and Development* **12**, 198–209.
- KURDISTANI, S.K., TAVAZOIE, S. AND GRUNSTEIN, M. (2004). Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721–733.
- LEE, C.K., SHIBATA, Y., RAO, B., STRAHL, B.D. AND LIEB, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics* **36**, 900–905.
- LEHNERTZ, B., UEDA, Y., DERJUCK, A.A. *ET AL.* (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biology* **13**, 1192–1200.
- LIANG, G., LIN, J.C., WEI, V. *ET AL.* (2004). Distinct localization of histone H3 acetylation and H3–K4 methylation to the transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7357–7362.
- LITT, M.D., SIMPSON, M., RECILLAS-TARGA, F., PRIOLEAU, M.N. AND FELSENFELD, G. (2001). Transitions in histone acetylation reveal boundaries of three separately regulated neighbouring loci. *EMBO Journal* **20**, 2224–2235.
- MISRA, S., CROSBY, M.A., MUNGALL, C.J. *ET AL.* (2002). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biology* **3**, research 0083.1–0083.22.
- NACHT, M., FERGUSON, A.T., ZHANG, W. *ET AL.* (1999). Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Research* **59**, 5464–5470.
- PATINO, W.D., MIAN, O.Y. AND HWANG, P.M. (2002). Serial analysis of gene expression: technical considerations and applications to cardiovascular biology. *Circulation Research* **91**, 565–569.
- REINKE, H. AND HORZ, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Molecular Cell* **11**, 1599–1607.
- REN, B., ROBERT, F., WYRICK, J.J. *ET AL.* (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309.
- ROBYR, D., SUKA, Y., XENARIOS, I. *ET AL.* (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* **109**, 437–446.
- ROH, T.Y., NGAU, W.C., CUI, K., LANDSMAN, D. AND ZHAO, K. (2004). High-resolution genome-wide mapping of histone modifications. *Nature Biotechnology* **22**, 1013–1016.
- ROH, T.Y., CUDDAPAH, S. AND ZHAO, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes and Development* **19**, 542–552.
- SAHA, S., SPARKS, A.B., RAGO, C. *ET AL.* (2002). Using the transcriptome to annotate the genome. *Nature Biotechnology* **20**, 508–512.
- SCHUBELER, D., MACALPINE, D.M., SCALZO, D. *ET AL.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes and Development* **18**, 1263–1271.
- SIMS, R.J., 3RD, NISHIOKA, K. AND REINBERG, D. (2003). Histone lysine methylation: a signature for chromatin function. *Trends in Genetics* **19**, 629–639.
- STRAHL, B.D. AND ALLIS, C.D. (2000). The language of covalent histone modifications. *Nature* **403**, 41–45.
- STRUHL, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes and Development* **12**, 599–606.
- TAKANO, T., HASEGAWA, Y., MATSUZUKA, F. *ET AL.* (2000). Gene expression profiles in thyroid carcinomas. *British Journal of Cancer* **83**, 1495–1502.
- TAMARU, H. AND SELKER, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283.
- TIMMERS, H.T. AND TORA, L. (2005). SAGA unveiled. *Trends in Biochemical Sciences* **30**, 7–10.

- TURNER, B.M. (1993). Decoding the nucleosome. *Cell* **75**, 5–8.
- TURNER, B.M. (2002). Cellular memory and the histone code. *Cell* **111**, 285–291.
- VELCULESCU, V.E., ZHANG, L., VOGELSTEIN, B. AND KINZLER, K.W. (1995). Serial analysis of gene expression. *Science* **270**, 484–487.
- VELCULESCU, V.E., ZHANG, L., ZHOU, W. ET AL. (1997). Characterization of the yeast transcriptome. *Cell* **88**, 243–251.
- VOGELAUER, M., WU, J., SUKA, N. AND GRUNSTEIN, M. (2000). Global histone acetylation and deacetylation in yeast. *Nature* **408**, 495–498.
- WOLFFE, A.P. AND MATZKE, M.A. (1999). Epigenetics: regulation through repression. *Science* **286**, 481–486.
- WU, J. AND GRUNSTEIN, M. (2000). 25 years after the nucleosome model: chromatin modifications. *Trends in Biochemical Sciences* **25**, 619–623.
- XIN, Z., TACHIBANA, M., GUGGIARI, M., HEARD, E., SHINKAI, Y. AND WAGSTAFF, J. (2003). Role of histone methyltransferase G9a in CpG methylation of the Prader–Willi syndrome imprinting centre. *Journal of Biological Chemistry* **278**, 14996–15000.
- XU, F., ZHANG, K. AND GRUNSTEIN, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **121**, 375–385.
- ZHANG, L., ZHOU, W., VELCULESCU, V.E. ET AL. (1997). Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272.