ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 6, pp. 589-594. © Pleiades Publishing, Ltd., 2007.
Original Russian Text © S. B. Akopov, I. P. Chernov, S. S. Bulanenkova, Yu. V. Skvortsova, A. S. Vetchinova, L. G. Nikolaev, 2007, published in Biokhimiya, 2007, Vol. 72, No. 6, pp. 725-732.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM06-338, April 1, 2007.

REVIEW

Methods for Identification of Epigenetic Elements in Mammalian Long Multigenic Genome Sequences

S. B. Akopov, I. P. Chernov, S. S. Bulanenkova, Yu. V. Skvortsova, A. S. Vetchinova, and L. G. Nikolaev*

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (495) 330-6538; E-mail: lev@humgen.siobc.ras.ru

Received December 28, 2006
Revision received February 8, 2007

Abstract—Epigenetic elements of the genome, i.e. elements that determine stably inherited changes in gene expression without changes in the genomic DNA sequence, are essential tools of genetic regulation in higher eukaryotes. The complete sequencing of the human and other genomes allowed studies to be started on positioning of these elements within long multigenic regions of the genome, which is a prerequisite for a comprehensive functional annotation of genomes. This mini-review considers some recent experimental approaches to the high-throughput identification and mapping of epigenetic elements of mammalian genomes, including the mapping of methylated CpG sites, open and closed chromatin regions, and DNase I hypersensitivity sites.

DOI: 10.1134/S0006297907060016

Key words: functional genomics, epigenome, CpG methylation, chromatin structure, hypersensitivity to DNase I, gene regulation

Upon the complete sequencing of a number of mammalian genomes [1-3], the study on regulatory systems determining the phenotypic variety of living organisms becomes a current problem of the coming post-genomic era. Epigenetic elements, i.e. elements that determine stably inherited changes in gene expression without changes in the genomic DNA sequence, are the most important components of the genome regulatory machinery [4]. These elements include DNA methylation sites, open and condensed chromatin regions, etc. Obviously, a comprehensive annotation of the genome has to contain data on the position of these elements in the genome of all types of the organism's cells.

The number of "classic" *cis*-regulatory elements (promoters, enhancers, silencers, insulators) in the eukaryotic genome reaches tens of thousands [5], whereas the amount of potential epigenetic elements is orders of magnitude higher than this number and equals some millions. Therefore, at present the problem of a genome-wide identification and mapping of epigenetic elements is almost insuperable and needs to be simplified. The problem may be simplified differently: first, all these elements may be mapped inside a relatively small stretch of the genome; second, a functional part of the genome may be chosen, such as the protein encoding sequences or CpG islands; third, approaches of comparative genomics may be used, such as a differential display or subtractive hybridization, which can give information about the difference in location of epigenetic markers in various tissues or organisms. We think that approaches using the most possibly comprehensive mapping of epigenetic elements in the limited but sufficiently long (several Mbp) multigenic regions of the genome are the most adequate. In such regions, mechanisms of high-throughput regulation of the genome can be studied and new approaches for mapping the epigenetic elements can be tested. Afterwards, these completely characterized regions can be combined into maps of whole chromosomes and genomes.

Approaches for mapping epigenetic elements are being developed in the framework of the international consortium ENCODE (the ENCyclopedia Of DNA Elements, http://www.genome.gov/10005107) [6]. The strategy of our studies is similar to that of the ENCODE project. We try to design a complete map of the *cis*-regu-

Abbreviations: bp) base pair; kb) thousand base pairs.

* To whom correspondence should be addressed.

latory and epigenetic elements located in a human genome multigenic segment 1 Mbp in length.

In the present review, some recent experimental approaches to the high-throughput mapping of mammalian genome epigenetic elements and results of them are considered. All these approaches include the same crucial stage—obtaining libraries of relatively short DNA fragments highly enriched with the corresponding elements. Upon designing such libraries, the mapping of regulatory epigenetic elements becomes a purely technical problem and can be realized using various technologies, such as hybridization with genomic microarrays, high-throughput sequencing, SAGE-similar (serial analysis of gene expression) technologies, and others shortly described further.

MAPPING OF METHYLATED CpG SITES

Methylation of cytosines at the 5-position inside the dinucleotide CpG plays a key role in the regulation of gene expression in higher eukaryotes [7]. To reveal and analyze methylated cytosines numerous approaches have been proposed and described in recent reviews [8, 9]. Note that the number of potential CpG methylation sites coincides with the number of CpG dinucleotides in the genome; therefore, they are mapped by simplified methods. We shall consider in brief some recently published methods for the high-throughput identification of methylated CpG.

Methyl-sensitive PCR. This approach uses as a template for the PCR amplification the genomic DNA modified by treatment with bisulfite, which converts into thymines only unmethylated cytosines [8, 9]. Thus, the amplified product will contain DNA fragments with different nucleotide sequences corresponding to the original methylation state of the fragment.

A fluorescently labeled product of this PCR was used as a probe for hybridization with an oligonucleotide chip capable of distinguishing methylated and unmethylated sites. This approach was used for identification and mapping of methylated CpG sites from the regulatory regions of some genes located both inside and outside CpG islands in some normal and tumor tissues [10-12].

The bisulfite approach combined with bisulfite sequencing and mass spectrometry is used in the framework of the Human Epigenome project [13]. The purpose of this project is the genome-wide mapping and analysis of the genome methylation state. In the preliminary stage of the project, methylation of the main human histocompatibility complex genes was analyzed [13]. The intron and exon regions of 90 genes were analyzed in various cell types, and the majority of these sequences were characterized either by hyper- or hypomethylated state, which indicates their involvement in genomic regulation.

An identification of bisulfite-modified bases using methyl-specific primers was proposed recently [14]. The methylation profiles of 1536 CpG sites were analyzed from the 5′-regulatory region of 371 genes for some cell lines and in normal and tumor lung tissue. A number of markers were identified, which allowed the discrimination of lung adenocarcinoma and normal lung tissue by methylation degree.

Methylation-sensitive representational difference analysis (MS-RDA). This approach is used for comparison of the methylation degree in the genomes of two cell or tissue types. To prepare PCR-amplified fragments of genomic DNA (amplicons) from two different sources, the methyl-sensitive endonuclease *Hpa*II was used. The resulting amplicons were compared using subtractive hybridization [15, 16]. This approach identifies differentially methylated genome regions in two different cell lines or tissues. In addition to the simplification of subtractive hybridization, the number of mapped methylated CpG is lowered because only CpG in the *Hpa*II recognition sites are under consideration and because of imperfection of the original RDA method [17].

Methylation-sensitive restriction landscape genomic scanning (RLGS-M). The RLGS-M [18] method is based on using methyl-sensitive restriction endonucleases with long recognition sites. Upon the genomic DNA cleavage with such an enzyme and radiolabeling, the DNA fragments are separated by two-dimensional gel electrophoresis combined with treatment with an additional restriction enzyme after the first dimension.

On using the methyl-sensitive restriction enzyme *Not*I with the restriction sites (GCGGCCGC) mainly located inside CpG islands, the RLGS profile allows the determination of both the methylation state and number of methylated copies of a sequence under analysis. This method can be used for comparison of the methylation degree of CpG islands of various (including normal and tumor) tissues [19].

Immunoprecipitation of methylated DNA. Immunoprecipitation with antibodies to methylated cytosines was used combined with microarrays for determination of the methylation profile of all human chromosomes at the resolution of 80 kb and for defining the methylation profile of large CpG islands [20].

Methylated-CpG island recovery assay (MIRA). The MIRA approach is based on the ability of proteins containing the methyl-CpG-binding domain, such as MBD2, to specifically bind with methylated sequences of DNA. Ultrasonically fragmented DNA from cells or tissues was incubated with the immobilized MBD2 in the presence of its partner in binding, the protein MBD3L1, which increased the affinity of MBD2 for methylated DNA [21]. The prepared DNA was used as a template for PCR. This procedure mainly reveals the highly methylated regions of CpG islands. Later, MIRA was combined with hybridization with microarrays that allowed identification of many genes containing the methylated regulatory regions in primary lung cancer and melanoma [22].

Non-methylated genomic sites coincidence cloning (NGSCC). This approach [23] consists of two stages. In the first stage, genomic DNA is treated with a methyl-sensitive restriction endonuclease, which cleaves only unmethylated sites, and the fragments containing unmethylated CpG are selectively amplified by a PCR-suppression technique [24]. In the second stage, using an original procedure of cloning the coincidental sequences, from this pool only those fragments are selected and cloned that belong to the genome regions under study. This approach was used to analyze the distribution profile of hypomethylated CpG in a region 1 Mbp in length in human chromosome 19 in DNA specimens from a normal testis and seminoma. A high level of expression of some genes inside the hypomethylated clusters correlated with hypomethylation of the genome regions including them; other genes were expressed independently of the methylation status. Combined with the technique of rapid identification of genomic splits (RIDGES) [25], the NGSCC approach allows the determination of methylation profiles for long regions of DNA.

Analysis of methylation profile using endonuclease McrBC. To analyze the methylation profile, the fragmented genomic DNA was treated with the methyl-sensitive endonuclease McrBC, and the fraction depleted in methylated CpG was purified by electrophoresis in agarose gel [26, 27]. This fraction was used for hybridization with genomic microarrays in parallel with the control fraction (untreated with McrBC). The methylation profile of the human genomic DNA was determined for more than 21,000 sites, including 79% of annotated sites of transcription start. The analysis revealed a complicated profile of DNA methylation along every autosome, with a tendency for increase in the methylation density with approaching telomeres. About 2% of CpG islands were methylated with a high density, 17% contained a considerable quantity of 5-methylcytosines, and the methylation level did not correlate with the position of CpG islands relative to sites of transcription start.

Identification of methylated CpG sites by mass spectrometry. A highly efficient approach has been recently proposed for identification of bases transformed during the reaction with bisulfite, which includes hybridization with a library of peptidyl-nucleic acids and time-of-flight mass spectrometry with laser desorption-ionization (MALDI-TOF) [28]. This approach is specified by high sensitivity and reliability in quantitative determination of methylation level.

Reduced representation bisulfite sequencing (RRBS). Genomic DNAs of mouse cells and of the same cells deprived of methyltransferases Dnmt1, Dnmt3a, and Dnmt3b were cleaved with *Bgl*II, and for further analysis fragments 500-600 bp in length were chosen. The ends of these fragments were supplemented with synthetic primers, and the fragments were treated with bisulfite, PCR-amplified, cloned, and sequenced [29]. Analysis of sequences of about 1000 clones (~400,000 bp) has shown that the method can be used for genome-wide studies. It should be noted that this approach allows the identification of methylated cytosines in sequences other than CpG.

IDENTIFICATION OF OPEN AND CLOSED CHROMATIN REGIONS

Studies of specific features of chromatin structure of transcribed regions resulted in a concept of "open" or "decondensed" (active) and "closed" or "condensed" (inactive) chromatin and its involvement in the regulation of gene activity [30]. Regions of the open and closed chromatin play an important role in the regulation of gene transcription, and this is the reason for their identification and mapping inside elongated regions of the genome. Some approaches have been proposed for high-throughput identification and mapping of open and closed chromatin regions.

***In vitro* detection of specific chromatin regions.** The chromatin structure of the human genome was analyzed by cleavage of DNA inside isolated nuclei using micrococcal nuclease and separation of regions with the compact and condensed chromatin by centrifugation in a sucrose density gradient [31]. The distribution of the closed and open chromatin regions inside the genome was studied using hybridization with metaphase chromosomes and genomic microarrays. Positions of the open chromatin regions correlated with positions of the genome regions with higher gene density but did not correlate with the level of their expression. The authors think that this is caused by location of inactive genes in the open chromatin or active genes inside the compact chromatin regions.

In another scheme, chromatin was fractionated using differential solubility of the compact (enriched with histone H1) and open chromatin. The total chromatin was treated with micrococcal nuclease and divided into three fractions: soluble in 5 mM $MgCl_2$, in EDTA, and insoluble. It was also fractionated by its accessibility for cleavage by DNase I. The fractions were hybridized with microarrays carrying the coding sequences of about 22,000 genes, and data on the chromatin state were compared with data on gene expression obtained with the same microarrays. The level of gene expression was inversely proportional to the level of chromatin openness [32].

Mapping of regions accessible for Dam-methylation. The open and closed chromatin regions were identified *in vivo* by testing the chromatin accessibility for the methylase Dam of *E. coli* [33]. The method is based on the ability of Dam methylase to methylate adenine residues in the GATC sequence [34] and also on an approach similar to the above-described NGSCC method [23]. The distri-

bution of the open chromatin regions (accessible for the Dam methylase) was established inside the human genome region of 140 kb in the cell line HEK-293. The Dam-methylated regions were found in the gene introns and exons and also in the intergene space. The distribution of these regions along DNA was uneven and correlated with the transcriptional activity of genes in the locus. Moreover, the DNA regions accessible for the Dam methylase in many cases were overlapped with the regions hypersensitive to DNase I.

Mapping of chromatin regions containing modified histones. The posttranslational modification of histones is one of the fundamental mechanisms of regulation of gene activity and forms an epigenetic profile of the genome, or the "histone code" [35]. Modified side chains of amino acid residues in histones often act as binding sites for specific proteins that, in turn, induce changes in the chromatin structure and gene expression. Acetylation of lysine residues in histones is usually associated with the open state of chromatin and active transcription, whereas methylation of lysine and arginine residues can correlate with both active and inactive chromatin [36].

A genome-wide mapping of regions containing acetylated H3 histone was described in resting and activated human T-cells. Chromatin was immunoprecipitated with antibodies to diacetylated (K9/K14) H3 histone in combination with a SAGE-like approach [37]. Only 1.2% of the chromatin was found to contain the diacetylated H3 histone. In total, the H3 histone was highly acetylated in the genome regions rich with genes. A high acetylation level was characteristic for promoter regions. It was concluded that the increased openness of chromatin and a high level of gene expression in the given region of the genome were more likely not due to the total increase in the histone H3 acetylation, but rather to the hyperacetylation of proteins associated with promoters and other regulatory elements, such as enhancers, locus control regions (LCR), and insulators.

The level of histone H3 methylation at lysine-9 (K9) was analyzed in the cell line and tissue specimens of human colon cancer using immunoprecipitation of chromatin with antibodies to the histone H3 methylated or acetylated at K9 and a microarray carrying CpG islands of the human genome [38]. With CpG-microarrays, a strong correlation was shown between a high ratio of K9 methylation to the K9 acetylation and methylation of CpG islands of DNA.

The technique of chromatin immunoprecipitation with analysis on microarrays (ChIP-on-chip) was also used for studies on the methylation profile of the H3 and H4 histones in the protein-coding regions of the human genome. Positions of the histone H3 methylated at K4, K36, and K79 (associated with gene activation) and at K9, K27, and K20 (associated with gene repression) were mapped in human monocytic THP-1 cells. The correlation was studied between the histone H3 methylation at different positions and its acetylation at the K9/K14 position. The modifications H3-K4Me2, H3-K4Me3, H3-K36Me2, and H3-K79Me2 were shown to be associated with H3 hyperacetylation and elevated gene activity, whereas the modifications H3-K9Me2, H3-K9Me3, H3-K27Me2, and H4-K20Me2 were associated with the H3 hypoacetylation [39].

Histone H4 methylation at arginine-3 was studied in HL-60 line cells using chromatin immunoprecipitation and direct cloning and sequencing of the precipitated DNA fragments. All identified sites were located either in introns, or no farther than 5 kb from the 5′-ends of the corresponding genes. A considerable part of the genes adjacent to the sites containing the methylated H4 was involved in signal transduction and development, including the immune response of myeloma cells [40].

Mapping of DNase I-hypersensitive regions of DNA. Positions of DNase I-hypersensitive regions of DNA indicate the location of essential regulatory elements of the genome, such as promoters, enhancers, insulators, etc. [41]. Highly efficient approaches for identification and mapping of these sites have been recently developed.

One of these approaches includes limited hydrolysis of isolated nuclei with DNase I, isolation of soluble DNA, cleavage with restriction endonuclease forming sticky ends, and directional cloning of the resulting fragments. The library of fragments containing the sites of DNase I cleavage from human $CD4^+$ cells was sequenced, and positions of the cleavage sites were mapped [42]. The sites of DNase I cleavage were more frequently located in the genome regions containing regulatory elements, such as 5′-adjacent regions to genes, CpG islands, and sequences similar in the human and mouse genome, which suggested their functional importance [43]. In the next work of the same authors, 230,000 sites cleavable by DNase I were mapped. By quantitative real-time PCR, about 80% of these sites were shown to be mapped in the genome regions containing regulatory elements, including regions 2 kb upstream of the genes, CpG islands, and highly conservative sequences.

Another group of researchers [44] used a subtractive hybridization to obtain a library enriched with DNA fragments containing sites hypersensitive to DNase I. The resulting fragments were mapped in the human chromosome 21 region ~30 Mbp in length. The hypersensitive sites were mainly located near regions of transcription start and in CpG islands.

Methods for identification of sites preferentially cleaved by DNase I were recently adapted for hybridization with genomic microarrays. The two described approaches include slightly different methods for preparation of a hybridization probe. The first approach [45] includes biotinylation of DNA fragments obtained upon treatment with different concentrations of DNase and purification of these fragments on a column with immobilized streptavidin. The other approach [46] uses a selec-

tion of short fragments containing neighboring sites of cleavage with DNase I. Both types of probes were hybridized with microarrays carrying 44 regions of the human genome with total length of ~30 Mbp chosen for the first stage of the ENCODE project [6].

Various epigenetic elements essentially determine mechanisms of mammalian genome functioning. Identification and mapping of these elements in the genome is a large-scale problem, which is being solved by a number of highly efficient approaches. Nevertheless, the current methods are unable to provide for a genome-wide mapping of millions of epigenetic elements, especially on taking into account their tissue specificity. To solve this problem, new methods suitable for genome-wide analysis must be elaborated.

The authors are grateful to E. D. Sverdlov for valuable remarks and discussion.

This work was supported by the Russian Foundation for Basic Research (project No. 05-04-48814), the Program of Scientific School Support (NSh project 2006.2003.4), the Russian Academy of Sciences Presidium Program on Molecular and Cellular Biology, and the INTAS Program (project No. 01-0279).

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) *Nature*, **409**, 860-921.
2. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001) *Science*, **291**, 1304-1351.
3. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002) *Nature*, **420**, 520-562.
4. Jaenisch, R., and Bird, A. (2003) *Nat. Genet.*, **33** (Suppl.), 245-254.
5. Maston, G. A., Evans, S. K., and Green, M. R. (2006) *Annu. Rev. Genomics Hum. Genet.*, **7**, 29-59.
6. ENCODE consortium (2004) *Science*, **306**, 636-640.
7. Bird, A. (2002) *Genes Dev.*, **16**, 6-21.
8. Fraga, M. F., and Esteller, M. (2002) *Biotechniques*, **33**, 632, 634, 636-649.
9. Dahl, C., and Guldberg, P. (2003) *Biogerontology*, **4**, 233-250.
10. Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A. R., Gutig, D., Grabs, G., et al. (2002) *Nucleic Acids Res.*, **30**, e21.
11. Gitan, R. S., Shi, H., Chen, C. M., Yan, P. S., and Huang, T. H. (2002) *Genome Res.*, **12**, 158-164.
12. Shi, H., Maier, S., Nimmrich, I., Yan, P. S., Caldwell, C. W., Olek, A., and Huang, T. H. (2003) *J. Cell. Biochem.*, **88**, 138-143.
13. Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., et al. (2004) *PLoS Biol.*, **2**, e405.
14. Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E. W., Wu, B., Doucet, D., Thomas, N. J., Wang, Y., Vollmer, E., et al. (2006) *Genome Res.*, **16**, 383-393.
15. Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T., and Nagao, M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 2284-2289.
16. Kaneda, A., Takai, D., Kaminishi, M., Okochi, E., and Ushijima, T. (2003) *Ann. N. Y. Acad. Sci.*, **983**, 131-141.
17. Lisitsyn, N., and Wigler, M. (1993) *Science*, **259**, 946-951.
18. Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., et al. (1993) *Electrophoresis*, **14**, 251-258.
19. Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., et al. (2000) *Nat. Genet.*, **24**, 132-138.
20. Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., and Schubeler, D. (2005) *Nat. Genet.*, **37**, 853-862.
21. Rauch, T., and Pfeifer, G. P. (2005) *Lab. Invest.*, **85**, 1172-1180.
22. Rauch, T., Li, H., Wu, X., and Pfeifer, G. P. (2006) *Cancer Res.*, **66**, 7939-7947.
23. Azhikina, T., Gainetdinov, I., Skvortsova, Y., Batrak, A., Dmitrieva, N., and Sverdlov, E. (2004) *Mol. Genet. Genomics*, **271**, 22-32.
24. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., et al. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6025-6030.
25. Azhikina, T., Gainetdinov, I., Skvortsova, Y., and Sverdlov, E. (2006) *Mol. Genet. Genomics*, **275**, 615-622.
26. Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D., et al. (2004) *Nature*, **430**, 471-476.
27. Lippman, Z., Gendrel, A. V., Colot, V., and Martienssen, R. (2005) *Nat. Meth.*, **2**, 219-224.
28. Schatz, P., Distler, J., Berlin, K., and Schuster, M. (2006) *Nucleic Acids Res.*, **34**, e59.
29. Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005) *Nucleic Acids Res.*, **33**, 5868-5877.
30. Woodcock, C. L. (2006) *Curr. Opin. Struct. Biol.*, **16**, 213-220.
31. Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., and Bickmore, W. A. (2004) *Cell*, **118**, 555-566.
32. Weil, M. R., Widlak, P., Minna, J. D., and Garner, H. R. (2004) *Genome Res.*, **14**, 1374-1381.
33. Bulanenkova, S., Snezhkov, E., Nikolaev, L., and Sverdlov, E. (2007) *Genetica*, **130**, 83-92.
34. Singh, J., and Klar, A. J. (1992) *Genes Dev.*, **6**, 186-196.
35. Jenuwein, T., and Allis, C. D. (2001) *Science*, **293**, 1074-1080.
36. Lam, A. L., Pazin, D. E., and Sullivan, B. A. (2005) *Chromosoma*, **114**, 242-251.
37. Roh, T. Y., Cuddapah, S., and Zhao, K. (2005) *Genes Dev.*, **19**, 542-552.
38. Kondo, Y., Shen, L., Yan, P. S., Huang, T. H., and Issa, J. P. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 7398-7403.
39. Miao, F., and Natarajan, R. (2005) *Mol. Cell. Biol.*, **25**, 4650-4661.

40. Balint, B. L., Gabor, P., and Nagy, L. (2005) *Immunobiology*, **210**, 141-152.
41. Felsenfeld, G., and Groudine, M. (2003) *Nature*, **421**, 448-453.
42. Crawford, G. E., Holt, I. E., Whittle, J., Webb, B. D., Tai, D., Davis, S., Margulies, E. H., Chen, Y., Bernat, J. A., Ginsburg, D., et al. (2006) *Genome Res.*, **16**, 123-131.
43. Crawford, G. E., Holt, I. E., Mullikin, J. C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E. D., Wolfsberg, T. G., et al. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 992-997.
44. Sabo, P. J., Humbert, R., Hawrylycz, M., Wallace, J. C., Dorschner, M. O., McArthur, M., and Stamatoyannopoulos, J. A. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 4537-4542.
45. Crawford, G. E., Davis, S., Scacheri, P. C., Renaud, G., Halawi, M. J., Erdos, M. R., Green, R., Meltzer, P. S., Wolfsberg, T. G., and Collins, F. S. (2006) *Nat. Meth.*, **3**, 503-509.
46. Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006) *Nat. Meth.*, **3**, 511-518.