

Genome-wide location analysis: insights on transcriptional regulation

R. David Hawkins and Bing Ren*

Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine,
UCSD School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA

Received January 31, 2006; Accepted February 22, 2006

Gene expression analysis of microarray data can provide a global view of the transcriptome of a cell or specific tissue type, revealing important information about the kinds of signaling pathways, genes and protein classifications that are active. However, transcript profiles alone do not reveal how expression levels are controlled or which transcription factors (TFs) are responsible. Establishing transcriptional regulatory networks requires knowledge of TFs bound to promoter, enhancer and repressor elements. Accessibility of these sites and an additional level of control are mediated by chromatin and DNA modifications. Genome-wide location analysis is a tool for identifying protein–DNA interaction sites on a genomic scale. Applications of this tool are proving invaluable in determining *in vivo* target genes of TFs, epigenetic marks and *cis*-regulatory elements. Here, we will discuss how advances have been made in each of these categories and how this has helped to elucidate regulatory networks and control mechanisms.

INTRODUCTION

The advent of genome-wide location analysis (GWLA) has provided a direct means of determining regulatory interactions on a global scale (1,2). GWLA is the pairing of chromatin immunoprecipitation (ChIP) and DNA microarray (chip) technologies, commonly referred to as ChIP-on-chip or ChIP-chip (Fig. 1). This technology allows for the interrogation of protein–DNA interactions *in vivo* and the mapping of those interactions to precise locations in a genome. Initial studies identified the locations of specific transcription factor binding sites (TFBSs) within the yeast genome. GWLA has been subsequently utilized to ascertain protein–DNA interactions in the genomes of various organisms.

In this review, we will focus on recent applications of GWLA and how those studies have contributed to three primary categories of transcriptional regulation. A significant proportion of work has focused on epigenetics and, more specifically, histone modifications. Additionally, just as in the original studies, individual transcription factors (TFs) have been assessed to establish regulatory networks. The third category covers advances made in establishing global maps of common regulatory elements. The studies covered here should illustrate how this approach can be utilized to construct more informative genome maps, as well as provide a systems biology approach to understanding regulatory networks.

EPIGENETICS: THE ROLE OF HISTONES AND THEIR MODIFICATIONS

In recent years, a large body of data has been generated for histone marks on the genomes of various organisms, with the primary focus in yeast. Histones are important in chromatin structure, in DNA replication and repair and in transcription through post-translational modifications made to nucleosomes at the N-terminal tail of histones. These modifications include acetylation, methylation, ubiquitination, phosphorylation, poly-ADP-ribosylation and, most recently, sumoylation (3,4). The addition and removal of these modifications are an important part of transcription regulation and transcript elongation. It is generally accepted that upon histone acetylation, genes become actively transcribed, whereas deacetylation leads to transcriptional repression (5). A number of yeast studies have utilized GWLA to ascertain the effects of acetylation. Initially, this was done by identifying the genome-wide location of histone acetylases (HATs) and histone deacetylases (HDACs) (6–9). These studies illustrated the feasibility of mapping histone modifications and provided global-scale correlations of the presence of HAT or HDAC with gene expression or repression, respectively.

Much of the focus is currently on understanding how post-translational modifications correlate or regulate transcriptional

*To whom correspondence should be addressed. Tel: +1 8588225767; Fax: +1 8585347816; Email: biren@ucsd.edu

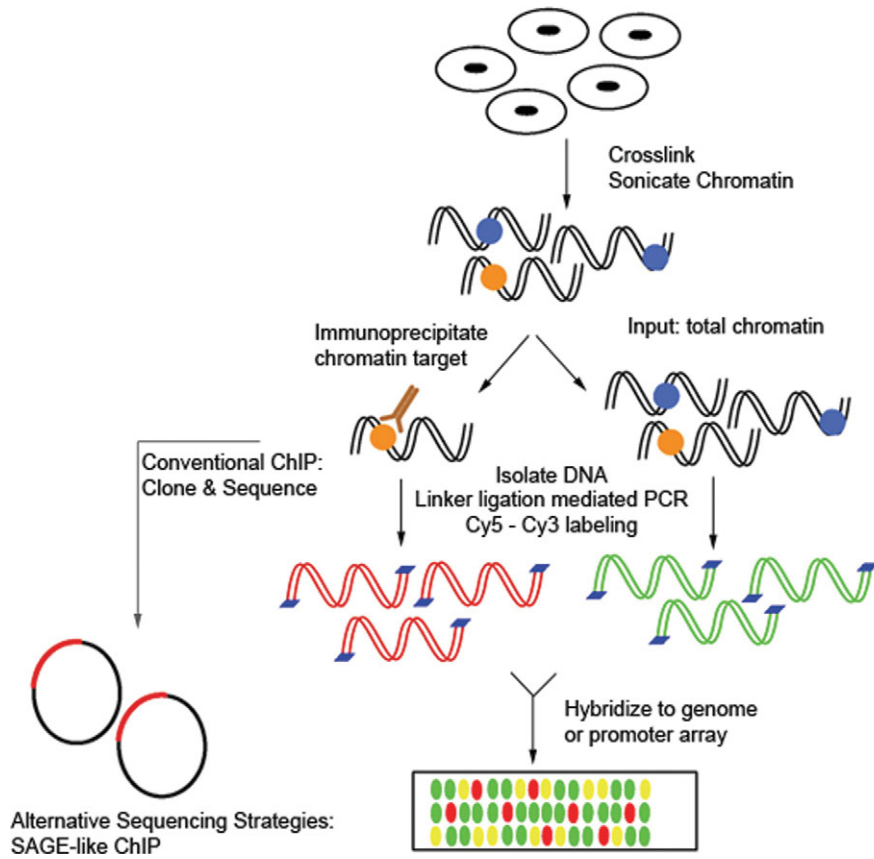


Figure 1. ChIP and microarray diagram. Cells are treated with formaldehyde to crosslink proteins bound to DNA. Cells are lysed and the chromatin is sonicated. The DNA-binding protein is immunoprecipitated, typically via antibody-linked beads. Isolated DNA, from IP and input, is amplified via linkers and ligation-mediated PCR. Samples are labeled and hybridized to genomic microarrays. Total chromatin input serves as the reference. A more detailed overview can be found elsewhere (33). Conventional ChIP requires sequencing of IP DNA. SAGE-like modifications have been implemented to cut sequencing cost.

activity. The acetylation of histones H3 and H4 marks the yeast genome in intergenic segments corresponding to the promoter regions primarily at the 5' end of the gene (10). These marks show enrichment at transcriptionally active genes. A remarkable pattern has emerged for histone H3 lysine 4 (H3K4) methylation at transcribed genes. Trimethylation was enriched at the 5' end, whereas dimethylation was enriched in the middle of gene and monomethylation was enriched at the 3' end of these genes (Fig. 2). However, other H3 trimethylated lysines had a more uniform distribution across genic regions. Higher resolution maps in yeast validated these findings, along with assessing an increased number of modifications (11). Assessment of modified histone marks on human chromosomes 21 and 22 demonstrated that trimethylation at H3K4 was also correlated with the 5' end of actively transcribed genes (12). Additionally, work in *Drosophila* has shown some differences from yeast in modification patterns. For instance, dimethylation of H3K79 in flies is correlated with transcription (13), whereas in yeast, this mark does not increase with transcription activity (10). Additional studies of these modifications on a more global scale could provide insight into unannotated genes in higher eukaryotes, which still lack detailed epigenetics maps like those under construction in yeast.

Histone occupancy studies in yeast revealed that there is a depletion of both H3 and H4 histones at regulatory regions of active genes. There is also a correlation with the loss of nucleosomes at promoters and the rate of gene transcription; however, this depletion does not extend into coding regions. These findings were reproducible under conditions that induce specific gene expression. Yeast grown under varying conditions induced environmentally specific gene activation, and these newly expressed genes showed nucleosome depletions at their promoters (14). Similarly, a complete genome-scale map of nucleosome deposition revealed nucleosome-free regions (NFRs) at transcription start sites (TSSs). By crosslinking nucleosomes and digesting DNA with micrococcal nuclease, bound regions were then hybridized to tiled, yeast genome microarrays. Tiling arrays were critical for providing the resolution necessary to identify exact nucleosome locations and for distinguishing linker regions. NFRs of ~150 bp were typically 200 bp upstream of start codons (15). Boundaries of the NFRs are defined by well-positioned nucleosomes flanking each side (Fig. 2). Histone H2A.Z is known to replace H2A and block the spread of euchromatic silencing (16). GWLA revealed that the nucleosomes that flank the NFRs comprised the histone variant H2A.Z (17). Histone acetylation is required for

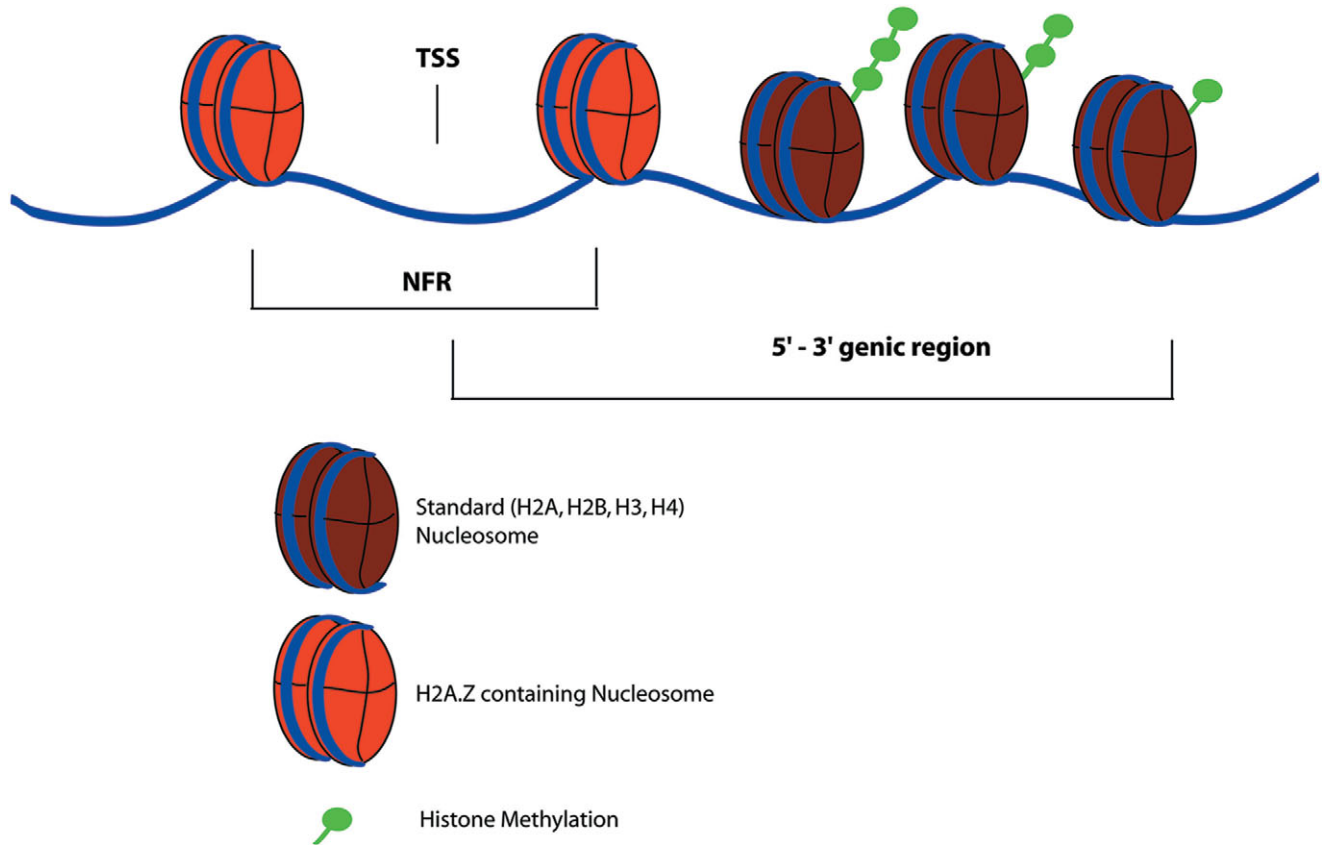


Figure 2. Nucleosome location and histone modifications. Nucleosomes comprise two copies of histones H2A, H2B, H3 and H4. Histone variant H2A.Z replaces H2A at the boundaries of NFRs that encompass the TSS. Methylation of histone H3 decreases (from tri- to di- to monomethylation) as nucleosomes occupy positions from the 5' to the 3' genic regions of DNA.

H2A.Z enrichment and deposition. Although there is no precedent for DNA elements to direct histone deposition, further examination identified a binding site for Reb1, a general regulatory factor, and an adjacent sequence of (dT:dA)₇. By comparing NFRs in seven species of *Saccharomyces*, regions of conservation were found outside TFBS. These NFRs contained tracks of poly-dA or poly-dT important for establishing the regions (14). These studies suggest that a sequence motif may play a role in determining which regions are resistant to chromatin silencing.

Histone H3 can also be replaced by its variant histone H3.3. This occurs in a replication-independent manner of nucleosome assembly (18). This variant differs by only a few amino acids, making it very hard to ascertain via a specific antibody. Utilizing a clever *in vivo* biotin tagging method (19,20), H3 was distinguished from H3.3 following induced expression in *Drosophila* S2 cells (21). GWLA revealed H3.3 deposition varied from gene to gene, whereas H3 remained uniform. The replication-independent deposition of H3.3 was typically in genic regions, marking actively transcribed genes in the genome, and correlates with the increase of RNA *PoIII* at active genes, as H3.3 is gradually enriched with each round of transcription. The level of H3.3 deposition provides a measurement that is sensitive enough to indicate levels of transcription. This could have important implications for how transcript levels are measured on a global scale.

The epigenome is a dynamic physical genome. Cancer epigenetics will greatly benefit from this approach by mapping the epigenomes of diseased versus normal cells. Many cancerous cell types show a loss of histone modifications when compared with normal counterparts (22). The CpG island hyper- and hypomethylation states also play important roles in cancer (23). It is now possible to map DNA methylation marks via immunoprecipitation of 5-methyl-cytosine and microarray (24,25). With the Human Epigenome Project (www.epigenome.org) underway in Europe (26), these high-throughput approaches are bound to be key technologies for such an endeavor.

REGULATORY NETWORKS

GWLA is an ideal methodology for identifying the regulatory sites bound by specific TFs within a genome. The information gained is environmentally specific, therefore, yeast is a preferable model organism for testing various growth or stress conditions. An example of this is illustrated by the mapping of regulatory sites for the MAPK-activated TF Ste12 (27). Haploid yeast cells were examined under conditions that induced either mating behavior or induction of filamentous growth. GWLA revealed Ste12 bound promoters for non-treatment conditions, mating-induced stimulation and filamentous growth conditions. Localization in MAPK kinase deletion

strains expanded the regulatory network by identifying that a specific MAPK kinase, Fus3, inhibited the binding of Ste12 to filamentation genes under mating conditions. By determining the genome-wide locations of Ste12 under various conditions, a detailed regulatory network was constructed, illustrating how upstream signaling events play a role in target gene activation under specific conditions. It also revealed that Ste12 has both condition-specific and non-specific binding partners, by examining common binding partners, and placed target genes into functional categories that corresponded to the environment. This added level of information is important in understanding gene regulation, as it is never as simple as a single regulator for an individual gene.

Ultimately, it will be important to know the regulatory binding sites of all known TFs within a genome and how the occupation of those sites varies under different environmental conditions, including various cell types in multicellular organisms. The occupancy sites of 203 TFs in yeast grown in rich media were recently determined (28). Furthermore, testing 84 of these TFs in at least one of 12 different conditions led to the identification of 11 000 unique TF-promoter interactions. The protein–DNA interaction data set allowed prediction of DNA binding consensus sites for the TFs using a combination of motif predicting algorithms [e.g. AlignAce (29), MDScan (30) and MEME (31)]. The authors identified highly significant ($P < 0.001$) motifs for 116 TFs and less significant motifs for 147 TFs. Such a detailed analysis of the complete regulatory code allowed the authors to define promoter architecture and distinct patterns of regulatory actions based on different environmental conditions (Table 1). Mapping the TFBS for virtually all yeast TFs provides an encyclopedic level of regulatory information and complex networks of interactions, some of which may be translational to higher eukaryotes.

Major strides are being made to study human transcriptional regulation, especially in relation to disease, using human cell lines and promoter microarrays. Initial forays have identified the targets of important cancer-related TFs. The *c-Myc* oncogene is involved in cell growth, proliferation and apoptosis, making it a critical component of many cancer cell types (32). GWLA was employed to understand the role of *c-Myc* in Burkitt's lymphoma cells (33). *c-Myc* bound ~15% of the promoters interrogated, indicating a broader role in global-gene expression along with its binding partner, Max. The ~85% overlap of bound promoters by the binding partners illustrates not only the validity of the technique but their roles as common global regulators. *Myc* binding sites did not always contain consensus DNA binding elements, identifying additional targets missed by only assessing E-boxes (34). CpG island arrays also demonstrated that a number of *Myc*:Max bound genes are repressed (35). The *Myc*–Max network was expanded in *Drosophila* by examining another interacting partner, Mnt—a Mad protein, and utilizing DamID assays and microarrays [(36) method described therein]. Likewise, elucidating the targets of the mixed-lineage leukemia histone methyltransferase (MLL1) illustrated that it too plays a global role in transcription, localizing to the 5' end of active genes (37). As there is an ever growing interest in understanding the role of oncogenic TFs and how their regulatory network leads to cancer growth and progression,

Table 1. Promoter architecture and TF binding behavior

	Explanation
Promoter architecture	
Single regulator	Promoter contains a binding site for a single transcriptional regulator
Repetitive motifs	Multiple-binding motifs for the same regulator
Multiple regulators	More than one binding site corresponding to multiple regulators
Co-occurring regulators	Contains binding sites for presumably interacting pairs of regulators
Binding	
Condition invariant	Regulators that bind the same set of promoters under two different conditions
Condition enabled	Regulators are activated following environmental change
Condition expanded	Regulators actively bind a subset of promoters under one condition and bind additional promoters following environmental change
Condition altered	Regulators that switch promoter preference following environmental change

additional factors have been analyzed via GWLA and human promoter arrays, including TCF4 in colorectal cancer cells (38). This TF is prominently bound to the endothelin-1 promoter, a gene commonly secreted from many solid tumors (39). TCF4 is known to interact with beta-catenin, which is frequently activated by oncogenic mutations in colorectal cancer cells (40). This is just one of many examples that illustrates how developmentally important TFs are commandeered in cancer cells. The AP-1 complex is made up of Jun, Fos and ATF proteins and is known to be involved in both tumor suppression and tumorigenesis (41). Promoters targeted by *c-Jun* and ATF2, two components of the AP-1 complex, were identified, and these TFBSs differ depending on the phosphorylation state of the protein and in response to cisplatin treatment, a common anti-cancer therapeutic (42). Again, the upstream signaling events played a key role in deducing gene regulation and target promoters. However, in some instances, more may be at play than just the state of phosphorylation, as the presence of important co-factors may be critical for regulation. In a study of cAMP-response element binding (CREB) protein, DNA binding by CREB was assessed in a phosphorylated or unphosphorylated state (43). Under these differing states, the protein showed no significant change in bound target sites. What did appear to be critical was the recruitment of a co-activator to inducible sites, specifically CBP and potentially p300. With a large number of TFs implicated in cancer biology, identifying regulatory cascades could provide novel targets for therapeutics.

Establishing regulatory networks for tissue-specific TFs or tissue-specific targets will provide invaluable information on how a cell maintains or establishes its cellular identity. Several recent GWLAs have provided such insight. Odom *et al.* (44) established regulatory networks for HNF1 α , HNF4 α and HNF6, which are required for normal function of the liver and pancreas (45–48). Complex regulatory maps revealed both tissue-specific and commonly regulated genes. These factors interact in various combinations to regulate respective subsets of genes. The architecture of the regulatory maps was elucidated into mechanistic models including

multicomponent loops, feedforward loops and autoregulation. Mutations in *HNF1 α* and *HNF4 α* cause maturity-onset diabetes of the young (49), and the regulatory networks established in this study should provide valuable information about this disease. The authors also suggested that deciphering the regulatory role of HNF genes could provide a mechanistic understanding of how polymorphisms near the *HNF4 α* gene increase the risk of type II diabetes. This model will likely prove applicable to a number of diseases, as genome-wide association studies have identified single nucleotide polymorphisms correlated with changes in gene expression levels (50,51).

Before establishing the transcriptional hierarchy to determine a cell fate, progenitor cells must relinquish their transcriptional stronghold on the genes that define their phenotype. Stem cells have the potential to be therapeutics for many diseases, and a core set of transcriptional regulators in human embryonic stem cells was recently investigated to understand their role in maintaining pluripotency. OCT4, NANOG and SOX2 are early markers in development thought to be near the top of the transcriptional hierarchy of ES cells and critical for maintaining pluripotency (52–55). Their genomic targets revealed that a surprising number of promoters are co-occupied. However, not all of these genes are actively transcribed. One or more of these TFs occupied promoters of 1303 actively transcribed genes, whereas 957 occupied promoters matched inactive genes. Many inactive genes appear to be developmentally important, indicating a role for repressor functions by the TFs in order to help maintain pluripotency. Again, a more global approach with several factors allowed the authors to identify mechanistic models of regulation among these factors. The regulatory models not only provide information about maintaining pluripotency but also suggest which programs need to be activated (de-repressed) to differentiate the cells into a defined lineage. However, just as the data illustrates the co-acting TFs function both as activators and repressors, it also demonstrates the need for a more complete understanding of other factors, implying that co-factors or epigenetic marks likely distinguish between these two roles.

***cis*-REGULATORY ELEMENTS**

Multicellular organisms, especially mammals, have evolved complex gene regulatory mechanisms that control tissue and developmentally specific gene expression. Many types of transcriptional regulatory elements have been discovered. These elements include promoters, enhancers, repressors and insulators that are positioned throughout the genome, and their identification has been particularly challenging. A significant step in mapping active promoters within a cell was accomplished using a component of the pre-initiation complex, TAF1 (56,57). Assembly of this complex at promoters is required for the initiation of transcription of protein encoding genes in eukaryotes (58). This was first demonstrated by assessing bound sites in ~1% of the human genome corresponding to 44 different loci known as ENCODE regions (56,59). A follow-up study utilized the first full, non-repetitive, human genome tiling array set consisting of ~14.5 million 50mers (57). In this study, 12 150 TAF1 binding sites were

mapped, 87% of which were within 2.5 kb of known TSSs, determined by 5' end sequencing-based methods. The more distant promoters suggest the presence of unidentified genes. Many of these promoters for putative genes were bound by other marks or located within 2.5 kb of the 5' end of EST-based predicted genes. Several were confirmed with corresponding transcripts by gene expression analysis. On the basis of a conservative number of promoters that matched known genes and unannotated genes in this study, the authors estimate that ~13% of human genes still remain unannotated. Another remarkable feature is that a large number of genes appear to have multiple promoters active in a single cell, corresponding with either different 5'-UTRs or first exon, distinct differences in protein sequences or variations in splicing and polyadenylation. Typically, tandem promoters are thought to play a role in indicating cell-specific transcripts, however, this study suggests that there are additional roles.

Several other studies have successfully identified *cis*-regulatory elements, many of which are enhancers. The NF- κ B protein, p65—known to bind enhancers (60), bound a large number of sites on human chromosome 22, and 90% were greater than 1 kb from the TSS (61). Assessment of Sp1, p53 and c-Myc binding on human chromosomes 21 and 22 revealed that nearly 50% of the binding sites were outside promoters or intragenic regions (62). Similarly, estrogen receptor (ER) binds a number of sites on chromosomes 21 and 22, the majority of which are outside of proximal promoters. Approximately half the sites contained the ER binding motif flanked by a forkhead-binding motif. Functional validation verified that these distal sites act as enhancer elements (63). Although these studies were limited in genome scope, they provide an unbiased assessment of where TFs bind. A more global scope of *cis*-regulatory elements has been accomplished by coupling sequencing strategies with ChIP, primarily SAGE-like ChIP [SACO (64), SABE (65), STAGE (66) and PET-ChIP (67)].

CONCLUSION

Identification of regulatory elements provides a more complete and detailed map of a genome. However, most regulatory elements, i.e. enhancers, repressors and insulators, still lack global identification. Of the ~2000 predicted human TFs (68,69), genomic binding sites for only a handful have been mapped and this was within a few cell-types and under fewer conditions. Future endeavors would benefit from the development of high-throughput methods to screen hundreds of TFs and assess their loss of function (e.g. RNAi followed by expression analysis), along with better computational prediction tools. One such step was recently made for predicting enhancer elements (70). Establishing a network of interactions that incorporates TF binding, co-factors, various regulatory elements and epigenetics will be paramount toward our understanding of transcriptional regulation and deciphering the transcriptional regulatory code.

Conflict of Interest statement. None declared.

REFERENCES

- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
- Iyer, V.R., Horak, C.E., Saffe, 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.
- Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.
- Santos-Rosa, H. and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur. J. Cancer*, **41**, 2381–2402.
- Gibbons, R.J. (2005) Histone modifying and chromatin remodelling enzymes in cancer and dysplastic syndromes. *Hum. Mol. Genet.*, **14** (Spec no. 1), R85–R92.
- Lieb, J.D., Liu, X., Botstein, D. and Brown, P.O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.*, **28**, 327–334.
- Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S.K., Wang, A., Suka, N. and Grunstein, M. (2002) Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell*, **109**, 437–446.
- Kurdistani, S.K., Robyr, D., Tavazoie, S. and Grunstein, M. (2002) Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat. Genet.*, **31**, 248–254.
- Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K. and Young, R.A. (2004) Global position and recruitment of HATs and HDACs in the yeast genome. *Mol. Cell*, **16**, 199–209.
- Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E. *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, **122**, 517–527.
- Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N. and Rando, O.J. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.*, **3**, e328.
- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., III, Gingeras, T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
- Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., Van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J. *et al.* (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.*, **18**, 1263–1271.
- 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. *Nat. Genet.*, **36**, 900–905.
- Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J. and Rando, O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**, 626–630.
- Meneghini, M.D., Wu, M. and Madhani, H.D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell*, **112**, 725–736.
- Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J. and Madhani, H.D. (2005) Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*, **123**, 233–248.
- Ahmad, K. and Henikoff, S. (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell*, **9**, 1191–1200.
- Beckett, D., Kovaleva, E. and Schatz, P.J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.*, **8**, 921–929.
- de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F. and Strouboulis, J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl Acad. Sci. USA*, **100**, 7480–7485.
- Mito, Y., Henikoff, J.G. and Henikoff, S. (2005) Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.*, **37**, 1090–1097.
- Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Roperio, S., Petrie, K. *et al.* (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.*, **37**, 391–400.
- Laird, P.W. (2005) Cancer epigenetics. *Hum. Mol. Genet.*, **14** (Spec no. 1), R65–R76.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. and Schubeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*, **37**, 853–862.
- Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H. *et al.* (2006) Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nat. Genet.*, **38**, 149–153.
- 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) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.*, **2**, e405.
- Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R. and Young, R.A. (2003) Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell*, **113**, 395–404.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
- Roth, F.P., Hughes, J.D., Estep, P.W. and Church, G.M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.*, **16**, 939–945.
- Liu, X.S., Brutlag, D.L. and Liu, J.S. (2002) An algorithm for finding protein–DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.*, **20**, 835–839.
- Bailey, T.L. and Elkan, C. (1995) The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **3**, 21–29.
- Pelengaris, S., Khan, M. and Evan, G. (2002) c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer*, **2**, 764–776.
- Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q. and Ren, B. (2003) A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc. Natl Acad. Sci. USA*, **100**, 8164–8169.
- Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A. and Amati, B. (2003) Genomic targets of the human c-Myc protein. *Genes Dev.*, **17**, 1115–1129.
- Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H. and Penn, L.Z. (2003) Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. *Curr. Biol.*, **13**, 882–886.
- Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C. *et al.* (2003) Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. *Genes Dev.*, **17**, 1101–1114.
- Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E. and Young, R.A. (2005) Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl Acad. Sci. USA*, **102**, 8603–8608.
- Kim, T.H., Xiong, H., Zhang, Z. and Ren, B. (2005) beta-Catenin activates the growth factor endothelin-1 in colon cancer cells. *Oncogene*, **24**, 597–604.
- Kusuhara, M., Yamaguchi, K., Nagasaki, K., Hayashi, C., Suzaki, A., Hori, S., Handa, S., Nakamura, Y. and Abe, K. (1990) Production of endothelin in human cancer cell lines. *Cancer Res.*, **50**, 3257–3261.
- Kinzler, K.W. and Vogelstein, B. (1996) Lessons from hereditary colorectal cancer. *Cell*, **87**, 159–170.
- Eferl, R. and Wagner, E.F. (2003) AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer*, **3**, 859–868.
- Hayakawa, J., Mittal, S., Wang, Y., Korkmaz, K.S., Adamson, E., English, C., Ohmichi, M., McClelland, M. and Mercola, D. (2004) Identification of promoters bound by c-Jun/ATF2 during rapid large-scale gene activation following genotoxic stress. *Mol. Cell*, **16**, 521–535.
- Zhang, X., Odom, D.T., Koo, S.H., Conkright, M.D., Canetti, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E. *et al.* (2005) Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl Acad. Sci. USA*, **102**, 4459–4464.

44. Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K. *et al.* (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science*, **303**, 1378–1381.
45. Costa, R.H., Kalinichenko, V.V., Holterman, A.X. and Wang, X. (2003) Transcription factors in liver development, differentiation, and regeneration. *Hepatology*, **38**, 1331–1347.
46. Bell, G.I. and Polonsky, K.S. (2001) Diabetes mellitus and genetically programmed defects in beta-cell function. *Nature*, **414**, 788–791.
47. Shih, D.Q., Screenan, S., Munoz, K.N., Philipson, L., Pontoglio, M., Yaniv, M., Polonsky, K.S. and Stoffel, M. (2001) Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes*, **50**, 2472–2480.
48. Jacquemin, P., Lemaigre, F.P. and Rousseau, G.G. (2003) The Onecut transcription factor HNF-6 (OC-1) is required for timely specification of the pancreas and acts upstream of Pdx-1 in the specification cascade. *Dev. Biol.*, **258**, 105–116.
49. Fajans, S.S., Bell, G.I. and Polonsky, K.S. (2001) Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.*, **345**, 971–980.
50. Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S. and Cheung, V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
51. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
52. Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.*, **17**, 126–140.
53. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, **113**, 643–655.
54. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, **113**, 631–642.
55. Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, **95**, 379–391.
56. Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M. *et al.* (2005) Direct isolation and identification of promoters in the human genome. *Genome Res.*, **15**, 830–839.
57. Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876–880.
58. Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.*, **72**, 449–479.
59. Consortium, T.E.P. (2004) The ENCODE (encyclopedia of DNA elements) Project. *Science*, **306**, 636–640.
60. Schjerve, H., Brandtzaeg, P. and Johansen, F.E. (2001) A novel NF-kappa B/Rel site in intron 1 cooperates with proximal promoter elements to mediate TNF-alpha-induced transcription of the human polymeric Ig receptor. *J. Immunol.*, **167**, 6412–6420.
61. Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M. *et al.* (2003) Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc. Natl Acad. Sci. USA*, **100**, 12247–12252.
62. Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J. *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.
63. Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoutte, J., Shao, W., Hestermann, E.V., Geistlinger, T.R. *et al.* (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, **122**, 33–43.
64. Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G. and Goodman, R.H. (2004) Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell*, **119**, 1041–1054.
65. Chen, J. and Sadowski, I. (2005) 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.
66. Kim, J., Bhing, A.A., Morgan, X.C. and Iyer, V.R. (2005) Mapping DNA–protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat. Methods*, **2**, 47–53.
67. Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z. *et al.* (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell*, **124**, 207–219.
68. Tupler, R., Perini, G. and Green, M.R. (2001) Expressing the human genome. *Nature*, **409**, 832–833.
69. Messina, D.N., Glasscock, J., Gish, W. and Lovett, M. (2004) An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res.*, **14**, 2041–2047.
70. Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E. and Taipale, J. (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*, **124**, 47–59.