PROSPECT
Journal of Cellular Biochemistry 107:19–29 (2009)

Journal of Cellular Biochemistry

Putting Chromatin Immunoprecipitation Into Context

Vincent Zecchini* and Ian G. Mills*

Uro-Oncology Research Group, CRUK Cambridge Research Institute, Robinson Way, Cambridge CB2 0RE, United Kingdom

ABSTRACT

Chromatin immunoprecipitation (ChIP), when paired with sequencing or arrays, has become a method of choice for the unbiased identification of genomic-binding sites for transcription factors and epigenetic marks in various model systems. The data generated is often then interpreted by groups seeking to link these binding sites to the expression of adjacent or distal genes, and more broadly to the evolution of species, cell fate/differentiation or even cancer development. Against this backdrop is an ongoing debate over the relative importance DNA sequence versus chromatin structure and modification in the regulation of gene expression (Anon. [2008a] Nature 454: 795; Anon. [2008b] Nature 454: 711–715; Henikoff et al. [2008] Science 322: 853; Madhani et al. [2008] Science 322: 43–44). Rationally there is a synergy between the two and the goal of a biologist is to characterise both comprehensively enough to explain a cellular phenotype or a developmental process. If this is truly our goal then the critical factor in good science is an awareness of the constraints and potential of the biological models used. The reality however is often that this discussion is polarised by funding imperatives and the need to align to a transcription factor or epigenetic camp. This article will discuss the extrapolations involved in using ChIP data to draw conclusions about these themes and the discoveries that have resulted. J. Cell. Biochem. 107: 19–29, 2009.

KEY WORDS: GENE EXPRESSION; TRANSCRIPTION FACTOR; CHROMATIN IMMUNOPRECIPITATION

Of the three billion base pairs within the human genome only around 1.5% encode proteins [Anon., 2004]. The remaining non-coding sequences in part regulate gene expression and understanding transcriptional regulation is of fundamental importance in making sense of human development and disease processes. With this in mind, it is necessary to address a number of critical questions: 1—what are the transcriptional regulatory sequences for every gene? 2—what is the *cis*-regulatory code underpinning tissue-specific and developmental gene expression? 3—how are epigenetic information and imprinting propagated throughout development and what is their role in regulating gene expression? Addressing these questions requires a comprehensive description of the DNA sequences that interact with transcriptional regulatory proteins, the temporal regulation of these associations and a comprehensive catalogue of the factors involved. Early efforts to tackle these questions did not take into account the status of the chromatin structure in compacting DNA and regulating the access of proteins to regulatory elements [Klug and Famulok, 1994]. Therefore, a series of approaches have been developed to study protein–DNA interactions in the nucleus under physiological conditions, which has evolved into chromatin immunoprecipitation (ChIP) [Das et al., 2004] (Fig. 1). Living cells are treated with a chemical cross-linking reagent that fixes protein to their DNA targets. Chromosomes are then extracted and fragmented using DNA shearing or enzymic digestion and specific DNA sequences associated with specific proteins are enriched using immuno-affinity purification with a specific antibody against a target protein. Purified DNA fragments can then be analysed using Southern blotting or PCR to determine whether a specific sequence is present. Conventional ChIP is normally limited to known proteins and known or suspected target sequences but does not readily allow for the identification of novel protein binding sites or target sequences on a genome-wide scale. This limitation can be overcome by the use of microarrays or direct sequencing as a readout [Collas and Dahl, 2008].

By combining ChIP with DNA microarrays, the ChIP-on-chip method in principle allows the unbiased detection of DNA binding sites for proteins throughout the genome. ChIP-on-chip involves the amplification and fluorescent labelling of ChIP-purified DNA followed by hybridisation to DNA microarrays along with a control DNA sample corresponding to total genomic DNA. In practice array probes are designed to cover non-repetitive sequences principally and so coverage is less complete than that achievable by direct sequencing using next generation platforms such as Solexa. Array elements that correspond to genomic-binding sites have significantly higher fluorescent signal intensity than the control DNA. The availability of high-density oligonucleotide arrays with whole

*Correspondence to: Vincent Zecchini or Ian G. Mills, Uro-Oncology Research Group, CRUK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.
E-mail: vinny.zecchini@cancer.org.uk; ian.mills@cancer.org.uk

Received 2 January 2009; Accepted 6 January 2009 • DOI 10.1002/jcb.22080 • 2009 Wiley-Liss, Inc.
Published online 9 February 2009 in Wiley InterScience (www.interscience.wiley.com).

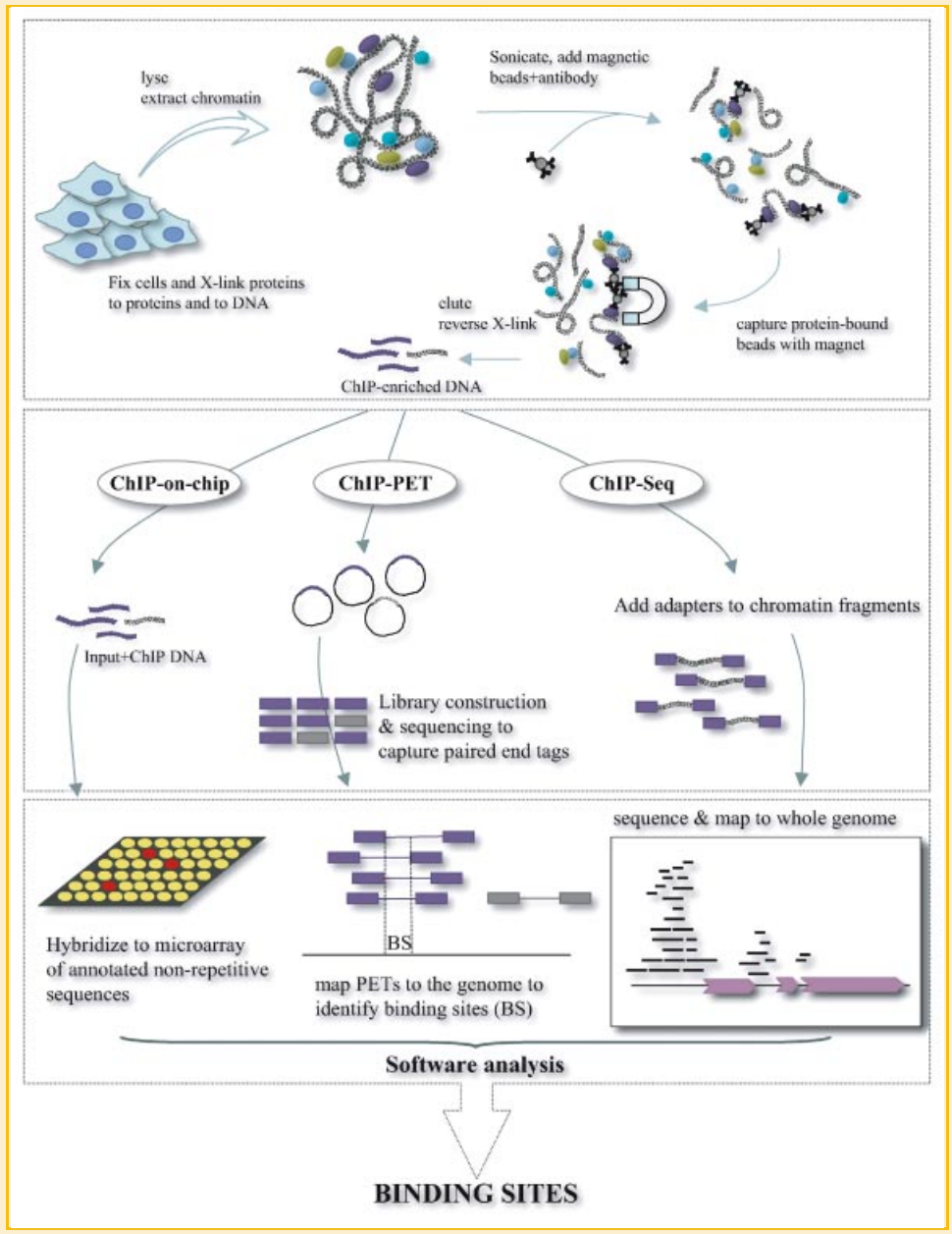


Fig. 1. Existing ChIP platforms: variations on the same theme. The ChIP technology allows the identification of specific genomic sequences that are in direct physical interaction with transcription factors and other nuclear proteins on a genome-wide basis. It produces a library of DNA sites that a particular factor was bound to in vivo. The technique can be divided into three phases. In the first one, the 'wet' phase, cells or tissue are treated with a chemical cross-linker resulting in protein–protein and protein–DNA binding. The cells are then lysed, chromatin is extracted and sheared by sonication resulting in double-stranded chunks of DNA fragments less than 1 kb in length. A complex of magnetic beads and antibody specifically directed against the protein of interest is added to the fragmented chromatin. The antibody-bound fraction is magnetically separated from the unbound fraction, the chromatin eluted, the cross-link reversed and the DNA purified. In the next phase, the 'semi-wet' phase, processing of the samples can be carried out by the means of different technological platforms. In the Chip-on-chip method, the ChIP-enriched chromatin is amplified and denatured to produce the single-stranded DNA fragments. These are labelled with a fluorescent tag and are incubated with the DNA microarray (tiled with short, single-stranded sequences covering the genomic region of interest). Complementary fragments will hybridise to the array, forming a double-stranded DNA fragment. The fluorescent signals from the array are then captured. The analysis of the raw data constitutes the 'dry' phase of ChIP-on-chip experiments and is also the trickiest part of the technique. Typical problems encountered during the analysis include the chip read-out, inadequate methods to subtract background noise, and suitable algorithms that normalise the data and make it available for subsequent statistical analysis. In ChIP-PET, which stands for *P*aired-*E*nd *T*ags, the ChIP-enriched DNA is cloned into a plasmid-based library. The plasmids are digested by restriction enzymes to yield a library of concatenated paired-end ditags sequences where each ditag represent the 5′-most and 3′-most termini of the ChIPed chromatin fragments initially cloned into the original library. These concatenated PETs are sequenced and their locations are mapped to the genome to delineate the boundaries of protein ChIP-enriched chromatin. ChIP-Seq combines chromatin immunoprecipitation with high throughput parallel whole-genome sequencing to identify binding sites of chromatin-associated proteins. After purification, adapters are added to the DNA fragments and the tagged DNA fragments are then amplified. They are then sequenced simultaneously using a genome sequencer. The analysis software aligns sample sequences to a known genomic sequence to identify the ChIP-enriched fragments. The depth of sequencing (i.e., the number of mapped sequence tags), the size of the genome and the distribution of the target factor all determine the sensitivity of the ChIP-Seq technology. Unlike the ChIP-on-chip technique, the accuracy of the ChIP-Seq is not limited by the spacing of predetermined probes. By integrating a large number of short reads, it is possible to achieve highly precise binding site localisation. Compared to ChIP-on-chip, ChIP-Seq can locate a protein binding site within tens of base pairs of the actual protein binding site.

genome coverage has improved the sensitivity and specificity in the detection of protein binding sites.

The alternative to this is tag-based high throughput sequencing, an approach originally applied to transcription profiling as a method known as Serial Analysis of Gene Expression (SAGE) [Chen, 2006]. In the guise of SAGE, it involves the isolation of a unique short DNA tag at the 3′-end of each cDNA, concatenation of multiple sequence tags to create a library of tag clones, followed by large scale sequencing to obtain tens of thousands of tag sequences resulting in a gene expression profile. Owing to the complexity of the human genome, the number of tag sequencing runs required to obtain a reliable and comprehensive map of protein–DNA interactions for a given transcription factor is considerably greater than the number of runs required for a SAGE-based gene expression profile. Consequently, it has only been a viable alternative to ChIP-on-chip with the recent development of high-throughput picoliter DNA sequencers. This provides the potential for a genuinely unbiased examination of DNA binding sites for proteins since no region of the genome is excluded prior to experimentation based on sequence characteristics (satellite/transposon density, etc.). This however will inevitably create issues at the data analysis and validation stages.

ChIP-based methods provide a direct means of examining protein–DNA interactions in cells with the consequence that the results are likely to have some physiological relevance. They do not rely on a prior knowledge of transcriptional regulatory sequences based on more in vitro approaches and have in many cases redefined the target sites for transcription factors previously investigated using other strategies. There are, however, potential drawbacks that require appropriate controls. The main limitation is the dependency on the quality and specificity of antibodies of available antibodies for proteins of interest. Given that the complex that is immunoprecipitated is chemically cross-linked, antibody specificity as determined using other approaches, such as Western blotting or confocal microscopy, do not extrapolate to ChIP. This may reflect the altered accessibility of epitopes within the cross-linked complex in a ChIP reaction and it is therefore good practice to screen panels of antibodies against a target protein. Alternatively the protein of interest can be ectopically overexpressed with an epitope or the epitope can be inserted into the genome in a targeted manner using homologous recombination [Zhang et al., 2008].

A significant factor in interpreting the results of ChIP studies is the degree to which the binding site and epigenetic information is context dependent. Context can be defined as a dependency on the predominant phase of the cell cycle for a given cell population, the differentiation or disease status of the cells, tissues or organism, the environment in which those cells, tissues or organisms find themselves and the evolutionary divergence between species. If binding sites or epigenetic marks are highly dynamic and influenced by these factors, then defining general principles becomes much more difficult. Some of these general principles precede the sequencing of the human genome and include, for example, the idea that transcription is regulated primarily through regulatory sequence elements situated predominantly upstream of target genes and proximal (within a few kilobases) of these genes. This has influenced the development of some platforms, promoter arrays for ChIP-on-chip experiments for example, and has also influenced subsequent data analysis and specifically correlations between transcription factor binding sites and changes in gene expression.

Increasingly researchers are seeking to derive ever greater 'relevancy' from their work by extrapolating from data gleaned in one system to others. It is therefore timely to reconsider what is known and unknown and what we as researchers are actually therefore basing our interpretations on. The traditional approach to characterising promoters has been to clone regions predicted to occur 1–5 kb upstream of a putative transcriptional start site, as defined by TATA boxes or equivalent motifs, into reporter constructs. The outcome has been significantly influenced by chance in that much of this work was undertaken before DNA binding sites for transcriptional regulators had been defined. The recent work of the The Encyclopedia of DNA Elements (ENCODE) Project focussing on 30 Mb—~1% of the human genome—provided some striking insights into the uses to which this DNA is put both in a single species and by making cross-species comparisons [Birney et al., 2007]. ENCODE has focussed on a collection of 44 genomic loci ranging in size from 500 kb to 2 Mb with the aim of comprehensively identifying functional elements in the human genome. This identified 118 promoters of which 96 were promoters of previously known transcripts and 22 were novel. By taking an unbiased, multi-group approach ENCODE has challenged some of the dogma that has previously informed our interpretations of ChIP data [Birney et al., 2007]. They have shown that: 1—whilst only a small fraction of DNA sequence encodes proteins, the human genome is pervasively transcribed. The majority of bases in the human genome are associated with at least one primary transcript and many link distal regions to establish protein-coding loci. 2—Regulatory sequences that surround transcription start sites are symmetrically distributed with no bias towards upstream regions. 3—Chromatin accessibility and histone modifications are highly predictive of the presence and activity of transcription start sites. 4—The majority of functional elements within DNA sequences are not actively constrained across evolution. These elements are in effect neutral despite being biologically active and are in other words of no discernible/specific benefit to the organism. These elements may provide the raw material for further rounds of natural selection and the development of new lineages.

These observations post-date the majority of recent ChIP studies and the work that we go on to discuss in this review will benefit from being viewed, or perhaps reviewed, in the context of ENCODE and similar studies.

APPLYING ChIP-on-chip TO EPIGENETICS

A fundamental aspect of genome regulation is chromatin organisation and DNA methylation. Histone modifications and DNA methylation states constitute the epigenetic information that controls animal development and cell function. Understanding the exact roles of these epigenetic marks and the mechanisms of function is an important component in delineating developmental programs and gene regulation [Anon, 2008a,b; Henikoff et al., 2008; Madhani et al., 2008]. ChIP-on-chip has been used to directly reveal histone modifications at specific loci on a genomic scale in a bid to

facilitate the understanding of the relationship between histone modifications and gene expression for a large number of genes in parallel. To this effect, a wide array of antibodies have been developed and characterised to recognise either the core histones or peptides with the specific modifications, such as acetylation, methylation, ubiquitination or phosphorylation. Several groups have applied ChIP-on-chip to examination of the nucleosome distribution in vivo. Lieb and colleagues performed ChIP-on-chip to identify the location of core histone H3 and H4 in the yeast genome and found that the nucleosomes are not evenly distributed along the yeast chromosomes [Lee et al., 2004]. Instead, the coding sequences have lower density of histones on average than the intergenic regions. Actively transcribed genes have the lowest density of nucleosomes, indicating that the nucleosomes are displaced during transcription. This study was extended by Rando and colleagues, who analysed the precise location of mono-nucleosomes along the yeast chromosomes using high-resolution oligo arrays [Yuan et al., 2005]. The analysis revealed that the promoters of actively transcribed genes are nucleosome free. Young and colleagues also observed a similar depletion of nucleosomes at the coding and transcriptional starts of the genes [Pokholok et al., 2005]. Taken together, these studies have established that nucleosomes are dynamically distributed along the genome, and showed that changes of nucleosome organisation accompany transcriptional activities. Similar chromatin dynamics have also been observed in higher eukaryotes. A ChIP-on-chip investigation into the chromatin structures in the *Drosophila* genome showed that promoters of actively transcribed genes are generally devoid of normal histone H3 [Mito et al., 2005]. However, these promoters are associated with a variant form of H3, H3.3, which is deposited to the transcribed sequences by a replication-independent mechanism. Other histone variants have been analysed by ChIP-on-chip. The position of histone variant H2A.Z along the yeast genome was also determined and found to flank the silent heterochromatin regions to prevent their spread [Raisner et al., 2005]. Nucleosomes with this histone H2A variant are preferentially located at promoters, again suggesting distinct chromatin organisation at transcription start sites [Guillemette et al., 2005].

Besides chromatin organisation and distribution of histone variants, chromatin modifications constitute another important aspect of epigenetic information. An extensive array of modifications was found, and many have been functionally linked to transcription. Several groups have systematically examined the histone modifications throughout the yeast genome and correlated them with loading of transcription factors and gene expression levels [Roh et al., 2004; Pokholok et al., 2005; Rando, 2007; Shivaswamy and Iyer, 2007]. These studies have revealed a surprisingly simple pattern of correlation between histone modifications and gene expression. The various acetylation and methylation of histone H3 and H4 are tightly correlated with each other and with gene expression. These marks are found at nearly all the active promoters, and the dynamic levels of modification appear to generally correlate with gene transcription [Pokholok et al., 2005]. Tri-methylation of lysine 36 of histone H3 (H3K36me3) appears to be correlated with transcription elongation or termination, as they occur mainly in the transcribed regions [Pokholok et al., 2005]. In contrast, methylation of lysine 9 of histone H3 is located within heterochromatin and centromeric and telomeric regions [Pokholok et al., 2005; Rao et al., 2005].

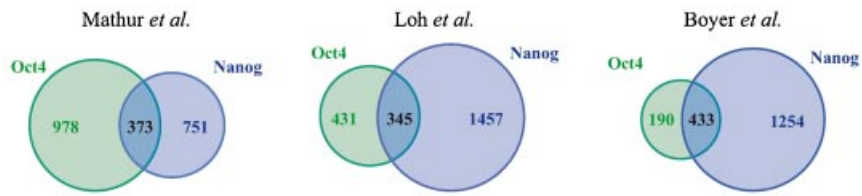
A similar distribution of histone modification patterns along the genome has also been observed in higher eukaryotes. A number of groups have examined the histone H3 acetylation and methylation in mammalian cells and in *Drosophila*. Similar to yeast, tri-methylation of lysine 4 of histone H3 (H3K4me3) is predominantly located at active promoters whereas tri-methylation of lysine 27 of histone H3 (H3K27me3) is preferentially present at inactive promoters [Kirmizis et al., 2004]. In addition, H3 acetylation occurs in promoter regions and at other genomic sequences that correspond to enhancers [Bernstein et al., 2005]. Recent reports demonstrated that DNA methylation can be monitored with a modified ChIP-on-chip method [Mohn et al., 2009]. Schubeler and colleagues used an antibody that specifically recognises the methyl-cytosine to isolate methylated DNA from cancer cells and identified the sites of DNA methylation using BAC or CpG island arrays [Weber et al., 2005]. The results revealed many differences between the DNA methylation profiles in cancer and normal cells, confirming that alteration of epigenetic programs contributes to tumorigenesis [Weber and Schubeler, 2007].

ChIP-on-chip IN STEM CELL RESEARCH AND CELLULAR DIFFERENTIATION

Embryonic stem (ES) cells are pluripotent cells derived from the inner cell mass (ICM) of the developing blastocyst. These cells possess self-renewal capacity and can generate virtually every cell type in the body. Sorting out the mechanisms underlying pluripotency and self-renewal of ES cells holds the key to understanding animal development and realising the therapeutic potential of ES cells in regenerative medicine. In an effort to dissect the transcriptional regulatory networks involved in maintaining a stem cell state, recent genomic studies using ChIP combined with genome-wide technologies have identified target genes regulated by three key transcription factors, Oct4, Nanog and Sox2 [Loh et al., 2008]. To date, target binding sites for these transcription factors have been identified by a number of groups for both human and mouse ES cells using different ChIP platforms [Boyer et al., 2005; Loh et al., 2006; Mathur et al., 2008] (Fig. 2).

In a study using ChIP-PET, Loh et al. [2006] identified 1,083 and 3,006 binding sites for Oct4 and Nanog, respectively, in mouse ES cells. The authors validated the functionality of these binding sites by complementing their ChIP dataset with RNAi microarray expression profiling. In a slightly more recent study employing a different platform (ChIP-on-chip), Mathur et al. [2008] also sought to identify genomic targets for the same transcription factors in mouse ES cells. Their approach identified 1,351 and 1,124 binding sites for Oct4 and Nanog, respectively. Both the Loh and Mathur datasets describe an extensive number of targets that are enriched for genes that play a role in development and cell fate specification. In addition, both sets of results suggest that Oct4 and Nanog can co-occupy some of their targets, that both Oct4 and Nanog can

a



b

Authors	System studied	Antibodies used	Platform used	Overlap between platforms*	Common genes bound by Oct4 and Nanog
Mathur *et al.*, 2008	mouse ES cells	Oct4: Santa Cruz sc-8628 Nanog: Bethyl	**ChIP-on-chip** ~20,000 genes -4kb to +4kb relative to TSS	Oct4: 9.3-24% Nanog: 19.5-28.1%	Atbf1, Cdyl, Eomes, Foxc1, Gbx2, Gsh2, Irx2, Jarid2, Nanog, REST**, Rfx4, Rif1, Sall4, Smarcad1, Sox2, Tcf7l1, Tdgf1, Tif1, Zic3
Loh *et al.*, 2006	mouse ES cells	Oct4: Santa Cruz Nanog: Cosmo Bio	**ChIP-PET** whole genome	Oct4: 9.1% Nanog: 18%	Cdx2, Jarid2, Sox2
Boyer *et al.*, 2005	human ES cells	Oct4: Santa Cruz sc-8628, sc9081. Nanog: R&D Systems	**ChIP-on-chip** ~18,000 genes -8kb to +2kb relative to TSS		

* The first number represents the % of targets identified by ChIP-PET and recovered in the ChIP-on-chip dataset, the second one represents the % of targets identified in by ChIP-on-chip and recovered in the ChIP-PET dataset.
**REST identified as target in both ChIP-PET and ChIP-on-chip but binding not within the same region.

Fig. 2. Identification of Oct4 and Nanog chromatin binding sites using ChIP technology. a: Number of Oct4 and Nanog targets identified using ChIP technology in the studies discussed here. b: Overlaps and discrepancies between these studies.

potentially activate or repress their targets and that Oct4 and Nanog, in addition to binding other promoters, can also bind their own and each other's promoters. However, as the authors report, there are also substantial differences in the data obtained through these different platforms that illustrate not only the need to apply caution in using these data in a complementary manner but, perhaps more importantly, also the need for a standardised data analysis methodology to compare different experiments. In order to determine how the analysis method and threshold criteria influence the agreement between the ChIP-PET and ChIP-on-chip experiments, Mathur et al. generated recovery curves. They show that 24% of the Oct4 targets identified by ChIP-PET were recovered in the ChIP-on-chip dataset within a distance of 1 kb. Conversely, 9.3% of the Oct4 targets identified by ChIP-on-chip were recovered in the ChIP-PET dataset. For Nanog, these numbers are 28.1% and 19.5%, respectively.

A previous study by Boyer et al. [2005] also used a ChIP-on-chip approach to identify Oct4 and Nanog binding sites in human ES cells and it is interesting to compare these results to those obtained in mouse. This comparison reveals that in both human and mouse, Oct4 and Nanog occupy a large number of transcriptionally active and silent genes, many of which have been shown to regulate lineage specification and cell fate determination. Still, only 9.1% of Oct4-bound genes and 13% of Nanog-bound genes overlapped between the studies. This limited overlap may suggest that differences may exist in the networks controlled by Oct4 and Nanog between species. Once more the different technology platforms and reagents used in the two studies may contribute to the discrepancies observed. Boyer et al. [2005] screened regions spanning 10-kb upstream transcription start sites of approximately 18,000 annotated genes, representing roughly 6% of the human genome. However, a significant number of binding sites may be located outside promoter regions. Indeed, previous work on mapping transcription factors binding sites using unbiased whole-genome approaches showed that some mammalian transcription factors bind sites outside proximal promoter elements. Unbiased mapping of binding sites in ES cells with ChIP-PET is therefore particularly important in the context of mammalian systems since regulatory elements are not always comprised within the 5′ proximal region of the first exon [Cawley et al., 2004].

The discrepancies highlighted between these three different studies could arise from the limitations inherent to the methods used. In ChIP-PET experiments, the cloning, sequencing and mapping all leave margin for errors whereas in ChIP-on-chip, observations are restricted to regions tiled on the array and the resolution is limited by the size of the probes, their spatial

distribution and the average chromatin fragment size. On top of these limitations, it is important to consider other sources of variability: binding sites may be differentially occupied at different times during the cell cycle, different antibodies may be used and the processing of samples can vary between laboratories. ChIP-Seq, the most recent addition to the ChIP family, aims to address many of the issues such as genome coverage, sequencing depth and binding resolution that are encountered by other techniques.

Separately, each study can be taken as a partial representation of the overall ES cell regulatory network but it is the integration of the data from multiple platforms that provides a more detailed overview of the factors involved in the ES cell transcriptional network. However, caution must be applied when extrapolating data in this way as the discrepancies observed in the results may reflect true biological differences between the samples due to a different cell status. Indeed, the binding of transcription factors to their DNA target must not be seen as a two dimension geographical chart but as a three dimension succession of maps that are temporally altered in response to various stimuli. Therefore, the representation of the overall ES cell regulatory network obtained by merging several different experiments as described in this review may not represent an accurate snapshot of a particular population of ES cells at a given stage but the superimposition of temporally different and incomplete stages. Only once the complete epigenetic and transcription factor binding map is obtained for the various time points, can we complete the picture by trying to place the events sequentially as they occur to achieve the phenotype we are interested in.

In parallel to the transcriptional profiling studies described above, studies on epigenetic markers have suggested that epigenetic profiles may be indicators of stem cell identity [Azuara et al., 2006]. They show that the epigenetic profile of pluripotent ES cells is different from that of embryonic carcinoma cells, haematopoietic stem cells and their differentiated progeny [Azuara et al., 2006]. Silent, lineage-specific genes replicated earlier in ES cells had high levels of acetylated H3K9 and methylated H3K4, usually considered as markers of open chromatin. These were combined with H3K27me3 at some silent genes. This suggests that pluripotency is characterised by a specific epigenetic profile where lineage-specific genes are primed for expression but that their expression is repressed if they carry the repressive H3K27me3 histone modification.

At present, we, as researchers, still have little evidence as to which of the two charts, that of transcription factor binding sites or that of epigenetic chromatin modifications, contains the true blueprint for gene expression. Is the epigenetic chromatin context the main modulator of transcription? Or is it simply a secondary outcome of the transcription factor's associated chromatin-modifying activities that facilitates transcriptional regulation? Or is the information locked in the genomic sequence motifs and recognised by the transcription factors the only drive behind controlling transcription? It is likely that the two mechanisms act in concert as a double level of security to ensure a tight regulation of transcription. It will be interesting to see if the superimposition of these maps leads to a more accurate picture of activation or silencing of gene transcription.

EVOLUTIONARY CONSERVATION OF TRANSCRIPTION FACTORS AND BINDING SITES

It is acknowledged that evolutionarily conserved sets of tissue-specific transcription factors determine a cell's transcription during development and do so by recognising short DNA sequence motifs. How transcription factors discriminate between those motifs is believed to be dependent on a range of influences including chromatin structure and cellular signalling/environment. Sequence comparisons alone across species are poor predictors and even when both the sequence motifs and the transcription factors are highly conserved between, for example, mouse and human, the precise target genes and binding site locations diverge. Similar observations have been made in cross-species comparisons of *Drosophila*, yeast and mammals. Mechanisms that determine tissue-specific transcriptional development may be significantly more complex than simply loss or gain of local sequence motifs. In a recent study, the contribution of genetic sequence to transcription was isolated using a mouse model of Down's syndrome containing part of human chromosome 21 [Wilson et al., 2008]. This allowed the comparison of orthologous mouse and human sequences in the same nuclei when isolated from other environmental and experimental variables. Liver was chosen as a representative tissue because the bulk of the cellular content is hepatocytes that can be easily isolated and are highly conserved in structure and function. In this unique context, isolated from many trans-regulatory influences, the authors showed that transcription factors encoded by the mouse genome could bind to human sequences identically to transcription factors encoded by the human genome in a native tissue setting [Wilson et al., 2008]. Exploring epigenetic marks, the authors assessed H3K4me3, a mark that mostly associates with transcription start sites and correlates with gene expression. Overall they found that around 85% of these marks were conserved between the human and mouse setting for chromosome 21 [Wilson et al., 2008]. The authors therefore concluded that regions of differential H3K4me3 between divergent species are dictated principally by cis-acting genetic sequence. Neither the cellular environment nor differences between human and mouse chromatin remodelling complexes were reported to be significant.

This study is an elegant tour de force. The development of a mouse containing a mosaic genome has undoubtedly allowed the contribution of sequence to be separated from other cellular and environmental influences. However, in applying this model in this manner, a self-fulfilling study arises. If there are protein differences (divergent signalling pathways or environmental responses) that affect gene expression then, presently, whilst we can transpose parts of a genome from species-to-species, we cannot transpose these trans-acting factors. Indeed in an earlier study, 41–89% of binding events at orthologous promoters were found to be species-specific depending on the transcription factor (FOXA2, HNF1A, HNF4A or HNF6) and comparing mouse to human. However such divergence should not be ignored in all cases. Many transcription factors, rather than acting as master regulators with constitutive activity once expressed, are activated dynamically in response to cytokines, growth factors and hormones. Hopefully the model will be employed to explore the targeting of other transcription factors in other tissues to address this point.

CELL-CYCLE DEPENDENT TRANSCRIPTION FACTOR ACTIVITY AND BINDING

Transcription factor activity depends on chromatin remodelling and the composition of transcriptional complexes. Most ChIP experiments published so far have been undertaken in unsynchronised cell lines. Although binding sites and epigenetic marks can be catalogued with relative precision across the genome they potentially reflect the predominant sites of occupancy and chromatin modifications in a population in G1/S, G2/M and G0. Any cell cycle specificity is therefore potentially masked. This becomes highly relevant for transcription factors with potent but divergent effects on phenotypes spanning proliferation through to terminal differentiation [Vias et al., 2008]. Differentiation is often associated with cell cycle arrest whereas proliferation is associated with cell cycle progression. Transcription factors such as the androgen receptor and oestrogen can contribute to both differentiation and proliferation in a context-dependent manner. In cell culture it is possible to synchronise populations of cells by chemically inducing reversible arrest at checkpoints and releasing these blocks.

Immunoprecipitation of oestrogen receptor alpha from synchronised cells was used to compare proteins differentially associated in G1/S and G2/M fractions [Okada et al., 2008]. Principal classes of chromatin-modifying complex include those that act through the modification of histones and those that act in an ATP-dependent manner to rearrange nucleosomal arrays. The histone deacetylase NuRD was detectable only in the G2/M fraction whilst components of the SWI/SNF complex were detectable in asynchronous and G1/S populations [Okada et al., 2008]. The authors went on to demonstrate that the NuRD complex inhibits oestrogen receptor transcriptional activity in G2/M. By implication the effects of these distinct associations on chromatin structure will be different and affect the transcriptional network activated by oestrogens or anti-oestrogens. This possibility has yet to be tested by the many groups tackling nuclear hormone receptor biology by characterising genomic targets for these proteins in synchronised cells.

APPLYING ChIP-on-chip TO CANCER

Cancer is a complex set of diseases characterised by accumulation of mutations in the genome and aberrant expression of multiple genes. A significant number of cancer-associated mutations occur in genes encoding transcription factors. Identifying the genomic-binding sites for these transcription factors is critical to understanding the molecular basis of cancer. Studies using ChIP-on-chip or ChIP-SAGE have identified direct target genes regulated by a growing number of transcription factors implicated in cancers, including the androgen receptor (AR), p53 and the oestrogen receptor (ER) [Massie and Mills, 2008]. In addition, these experiments have also revealed unexpected modes of action by these factors. In particular, binding site recognition appears to be dependent on the recruitment of complexes of transcription factors to clusters of binding motifs that tend to be of the order of 6-mer consensus nucleotide sequences. In the case of the androgen receptor, there is discernible co-clustering/co-enrichment of AR binding sites with sites for the oncogenic ETS family of transcription factors at around 75–80% of proximal promoter binding sites in the LNCaP cell line identified using a Nimblegen promoter array covering 25,000 gene targets [Massie et al., 2007]. Interestingly, in the same cell line using a tiling array with coverage of chromosomes 21 and 22, the co-enrichment is for other families of transcription factors and in particular GATA-3, Oct1 and FoxA1 principally at distal or enhancer sites located up to 100 kb away from identifiable transcription start sites [Wang et al., 2007]. The latter observation very much follows a pattern established for the oestrogen receptor in the MCF-7 breast cancer cell line [Carroll et al., 2006]. The implication of these studies is that whilst the AR and ER are transcription factors that have long been targeted therapeutically in prostate and breast cancers, other families of transcription factors may be equally or more significant in tumours in redirecting the AR and ER to drive expression of gene targets associated with disease.

What is presently missing however is a direct link between these carefully controlled ChIP studies in cancer cell lines and comparable assessments of transcription factor binding and function in cells extracted from clinical material. What we therefore base our understanding of transcription factor function in cancers on is therefore largely inference and correlation. In prostate tumours we know that ETS transcription factors are overexpressed often due to chromosomal rearrangements and gene fusions affecting ETV1, ERG, ETV4 and other family members [Alipov et al., 2005; Kumar-Sinha et al., 2008]. By combining these observations with a functional association between the androgen receptor and ETS1 in a cell line Massie et al., concluded that these overexpressed ETS transcription factors may affect AR signalling in prostate tumours [Alipov et al., 2005; Massie et al., 2007]. This is clearly difficult to prove categorically; however it is possible to co-stain tumour sections for the AR and interacting proteins with a high degree of precision owing to the development of fluorescent quantum dots [Shi et al., 2008]. An obvious preliminary question is whether the same cells actually express these associating proteins in tumours.

A whole-genome approach was taken to map oestrogen receptor binding sites in the MCF-7 cell line and motif co-enrichment revealed a highly significant co-enrichment of the PAX transcription factor motif with oestrogen receptor binding sites [Hurtado et al., 2008]. Based on previous reports of the overexpression of one member of this family, PAX2, in a subset of breast cancers, this protein became the subject of follow-up work [Silberstein et al., 2002]. Hurtado et al. [2008] identified for the first time an oestrogen receptor binding site within *ERBB2* and found that, unlike most other PAX sites, this binding site was occupied by PAX2 after both oestrogen and tamoxifen treatment and *ERBB2* transcription was repressed. Knockdown of PAX2 expression relieved repression. The hypothesis that PAX2 is a key determinant of ERBB2-mediated tamoxifen-resistance was supported by immunohistochemistry on ER-positive tumours that showed the PAX2 positive tumours were associated with significantly improved recurrence-free survival. The paper concludes by proposing that the *alter ego* of PAX2 is an ER transcriptional co-activator called AIB1 and that, consequently, the best prognosis for patients undergoing tamoxifen treatment

is potentially for those that are PAX2-positive, ER-positive and AIB1-negative [Hurtado et al., 2008]. Tumours from these patients indeed had the lowest levels of ERBB2 expression. The proposed model is elegant and the correlations are significant. There are however some pertinent lessons to be drawn from this work. Firstly, the original research to identify oestrogen receptor binding sites throughout the genome in the same cell line yielded 3,665 sites and a co-enrichment of FoxA1 and oestrogen receptor binding sites [Carroll et al., 2006]. The present whole-genome dataset comprises some 8,525 sites and now captures PAX site co-enrichment [Hurtado et al., 2008]. This illustrates how crucial it can be to select thresholds in interpreting ChIP data. It also indicates the importance of being able to refine this complex dataset down to a single credible target recognisable as significant. Given that there are other established routes to ERBB2 overexpression in breast tumours including genomic amplification [Mano et al., 2007], the clinically significant facet of the work is not the mechanism, which is not directly testable in the clinical material, but the correlation. What the field lacks is an association between the site and the proteins in ChIP from clinical material and this is universally true for research that seeks to employ ChIP to shed light on transcriptional networks in cancer.

CONCLUSIONS AND FUTURE PROSPECTS

ChIP approaches have rapidly been adopted by researchers working in almost all fields of biology. As a stand-alone, ChIP data are most impressive in providing whole-genome snapshots of transcription factor binding sites or epigenetic modifications in well defined model systems in which biological diversity is limited or tightly controlled. Assuming that there is significant evolutionary conservation of the proteins and modifications, these snapshots can also allow phylogenetic comparisons to be made. Consequently the most impressive data is generated by the simplest systems, be these liver specimens from multiple species or simple and abundant organisms such as yeast, and focussing on constitutive/core transcriptional machinery and epigenetic events. Applying ChIP to proteins that are highly regulated by extracellular stimuli/cell signalling and are expressed in small sub-populations of cells in tissues, such as the androgen receptor or oestrogen receptor, presents far greater challenges. Datasets generated with defined treatments in well characterised cell lines for such proteins are often robust and, when paired with profiles of histone modifications, can indeed highlight target genes for these transcription factors [Jia et al., 2008].

Issues arise however when trying to explain what the relationship is between the androgen or oestrogen receptor and other transcription factor families, trying to identify distal gene targets in a high-throughput manner or attempting to extrapolate from these datasets to tissues or even other cell lines. These issues can largely be summarised by our inability to refine in a truly unbiased manner the bewildering arrays of co-enriched binding motifs for families of transcription factors and of distal targets once DNA looping is invoked down to those that are most relevant. The challenge is exacerbated by our present inability to apply ChIP directly to the material that we profess the greatest interest in, tumour and normal tissues. Consequently much of this work incorporates hunches, extrapolation and an element of supposition. What do we need to do to improve this?

We need to be able to ChIP from much smaller quantities of material, equivalent to merely a few thousand cells rather than several million. Clearly this will happen first for the most abundant DNA-bound proteins, the histones. However significant efforts need to be put in to achieving results from sub-populations of cells in tissues for transcription factors. There is thankfully some progress in this regard with protocols now available for ChIP on 10,000 cells and fewer [Acevedo et al., 2007; Collas and Dahl, 2008]. Additionally, as it becomes ever clearer that transcription factors can bind at sites that are distal from at least known transcription start sites—note that ENCODE has revealed there may be large numbers that are 'unknown' or presumed not to exist—we need chromosome conformation capture (3C) technology to evolve to become a truly robust, genome-wide approach for enriching associations between distal sequences [Simonis et al., 2007].

Finally, we need to be able to reconstitute or enrich protein complexes on binding sites of particular interest, be this ER/PAX2 site in intron 1 of the ERBB2 gene or elsewhere, and identify the full complement of proteins within such complexes in an unbiased and comprehensive manner. There is really little point in advertising a co-dependent transcription factor or pioneer factor as a target for cancer treatment if there is sufficient redundancy in a multi-transcription factor complex for the gene expression and tumour growth is maintained even when such a factor is effectively targeted. This may prove possible if we can amplify and biotin-label immunoprecipitated DNA sequences and use them as scaffolds for the enrichment of proteins in a sequence-dependent manner, rather than merely sequence them or hybridise them to arrays. Nihilists may then argue that in the absence of chromatin structure such protein assemblies are meaningless but, interestingly, in vitro chromatin reconstitution has been undertaken for many years by researchers studying nucleosomal packing [Lusser and Kadonaga, 2004]. Pairing such an approach with ChIP-based isolation of DNA, biotin-tagging to allow enrichment of protein–DNA complexes on an avidin matrix, and more sensitive semi-quantitative mass spectroscopy based on stable isotope labelling by amino acids in cell culture (SILAC) or an equivalent strategy would be a step forward. SILAC-based proteomic screening was recently used to show that the basal transcription factor TFIID directly binds to the H3K4me3 mark via the plant homeodomain (PHD) finger of TAF3 [Vermeulen et al., 2007].

Such a multi-disciplinary proteomic strategy would remove our current reliance on DNA sequence and motifs to predict classes of bound proteins and would cast a comprehensive light for the first time on proteins critical for the regulation of transcription but with no intrinsic DNA binding capacity. Doubtless the research community will worry about artefacts in attempting to achieve such a goal. Indeed there will be artefacts but provided experiments are controlled and associations are validated on genomic material, this is a challenge that needs to be embraced. Otherwise we will remain with DNA motifs, sequencing technologies and supposition in attempting to describe the true complexity of protein–DNA complexes. Presently we can schematise them as core machinery

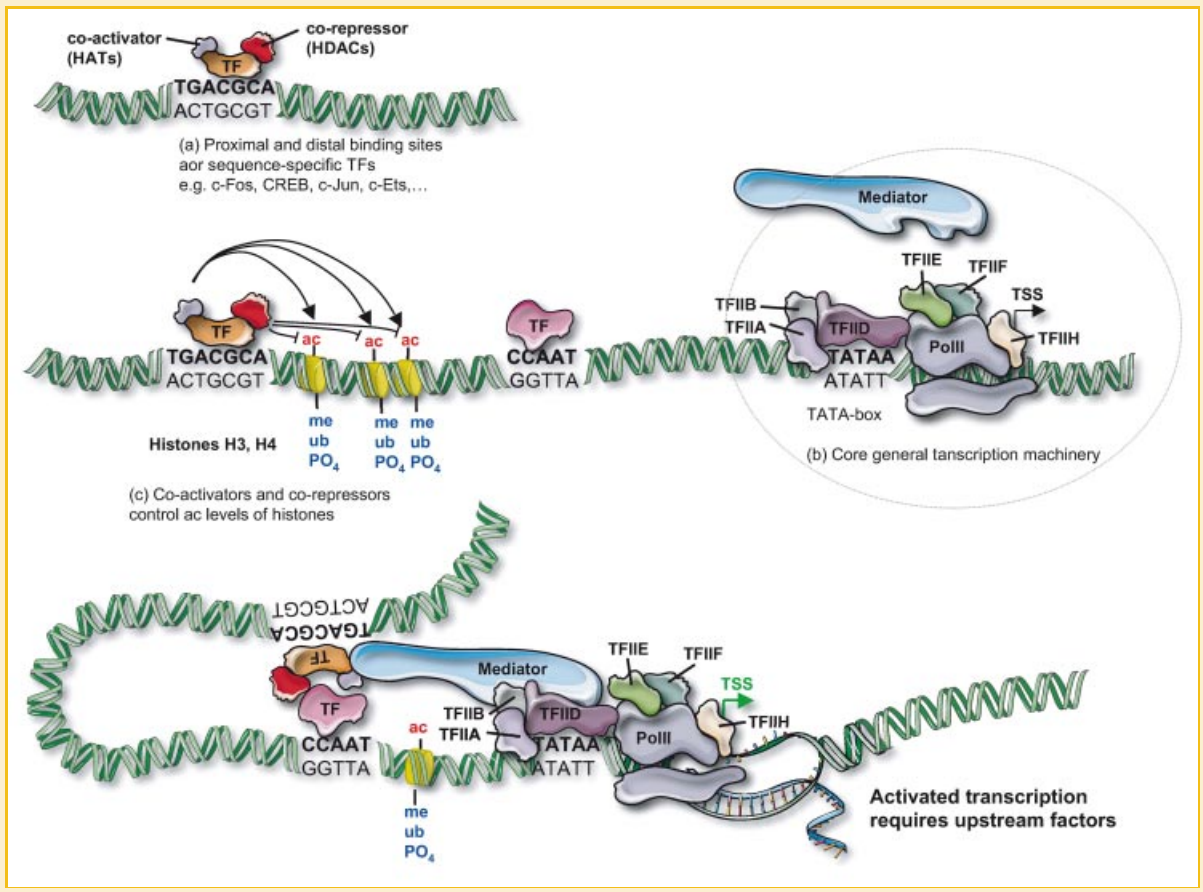


Fig. 3. The regulation of transcription in eukaryotes, whose complexity is succinctly summarised in the cartoon above, is a coordinated process involving the combined function of multiple protein complexes. a: Sequence-specific transcription factors (TFs) which modulate the transcription of specific target genes bind to proximal promoter elements and/or more distal regulatory sequences such as enhancers and silencers which can be situated upstream and/or downstream of the transcriptional start site (TSS). These TFs recruit histone acetyltransferases (HATs) (see c) resulting in the remodelling of chromatin and localised histone acetylation. This allows the access of additional TFs to *cis*-regulatory sequences that organise gene transcription via multiple interactions with other co-regulators and the core general transcription machinery, resulting in the recruitment of the Pol II complex to the transcription start site (see b). b: Ubiquitous general TFs that bind to core promoter DNA elements such as TATA-box and allow the specific recruitment of the Pol II complex to the core promoter. The assembly of general TFs and Pol II complex at core promoters is initiated by the binding of TFIID to the core promoter. The TFIID–DNA binding is stabilised by binding of TFIIB to TFIID. This then allows the recruitment of TFIIF, Pol II, TFIIE and TFIIH and the subsequent recruitment of TFIIF in association with Pol II to the complex. The mediator complex can be recruited to most genes via its interaction with components of the general transcription machinery. c: Co-regulators, either co-activators or co-repressors, which play essential roles in mediating the effects of TFs on the regulation of transcription. They are recruited to promoter regions of specific genes via interaction with sequence-specific TFs. General co-regulators often have basal functions and are recruited to most genes via association with the general transcription machinery, whereas TF-associated co-regulators that are primarily recruited to promoters by gene-specific TFs often have chromatin structure modifying activities. Chromatin-modifying co-regulators include histone acetyltransferases (HATs) and histone deacetylases (HDACs), and control the acetylation (ac) levels of specific lysine residues within the core histone tails. Histone acetylation by HATs generally correlates with gene activation, while gene silencing is often associated with histone deacetylation and HDACs are generally part of co-repressor complexes. In addition, histones are also modified by phosphorylation (PO_4), methylation (me) and ubiquitination (ub). These modifications form the epigenetic information of the genome. At present, it is still unclear what part of this information is inherited by the cell and what is not. The resulting code of histone modifications is recognised by specific protein domains present in TFs.

and transcription factors, whose associations are regulated by DNA loops and chromatin structure (Fig. 3), but little more.

ACKNOWLEDGMENTS

We would like to acknowledge the support of the National Institute for Health Research (NIHR) which funds the Cambridge Biomedical Research Centre. V.Z. is a postdoctoral researcher funded by a Cancer Research UK Programme Grant and I.G.M is a CRUK core-funded Associate Scientist.

REFERENCES

Anon. 2004. Finishing the euchromatic sequence of the human genome. Nature 431:931–945.

Anon. 2008a. Epigenomics: Tackling the epigenome. Nature 454:795.

Anon. 2008b. Moving AHEAD with an international human epigenome project. Nature 454:711–715.

Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, Farnham PJ. 2007. Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques 43:791–797.

Alipov G, Nakayama T, Ito M, Kawai K, Naito S, Nakashima M, Niino D, Sekine I. 2005. Overexpression of Ets-1 proto-oncogene in latent and clinical prostatic carcinomas. Histopathology 46:202–208.

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. 2006. Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8:532–538.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ III, Gingeras TR, Schreiber SL, Lander ES. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120:169–181.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447:799–816.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122:947–956.

Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M. 2006. Genome-wide analysis of estrogen receptor binding sites. Nat Genet 38:1289–1297.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR. 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.

Chen J. 2006. Serial analysis of binding elements for human transcription factors. Nat Protoc 1:1481–1493.

Collas P, Dahl JA. 2008. Chop it, ChIP it, check it: The current status of chromatin immunoprecipitation. Front Biosci 13:929–943.

Das PM, Ramachandran K, vanWert J, Singal R. 2004. Chromatin immunoprecipitation assay. Biotechniques 37:961–969.

Guillemette B, Bataille AR, Gevry N, Adam M, Blanchette M, Robert F, Gaudreau L. 2005. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol 3:e384.

Henikoff S, Strahl BD, Warburton PE. 2008. Epigenomics: A roadmap to chromatin. Science 322:853.

Hurtado A, Holmes KA, Geistlinger TR, Hutcheson IR, Nicholson RI, Brown M, Jiang J, Howat WJ, Ali S, Carroll JS. 2008. Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. Nature 456(7222):663–666.

Jia L, Berman BP, Jariwala U, Yan X, Cogan JP, Walters A, Chen T, Buchanan G, Frenkel B, Coetzee GA. 2008. Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity. PLoS ONE 3:e3645.

Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ. 2004. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev 18:1592–1605.

Klug SJ, Famulok M. 1994. All you wanted to know about SELEX. Mol Biol Rep 20:97–107.

Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. 2008. Recurrent gene fusions in prostate cancer. Nat Rev Cancer 8:497–511.

Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36:900–905.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38:431–440.

Loh YH, Ng JH, Ng HH. 2008. Molecular framework underlying pluripotency. Cell Cycle 7:885–891.

Lusser A, Kadonaga JT. 2004. Strategies for the reconstitution of chromatin. Nat Methods 1:19–26.

Madhani HD, Francis NJ, Kingston RE, Kornberg RD, Moazed D, Narlikar GJ, Panning B, Struhl K. 2008. Epigenomics: A roadmap, but to where? Science 322:43–44.

Mano MS, Rosa DD, De Azambuja E, Ismael GF, Durbecq V. 2007. The 17q12-q21 amplicon: Her2 and topoisomerase-IIalpha and their importance to the biology of solid tumours. Cancer Treat Rev 33:64–77.

Massie CE, Mills IG. 2008. ChIPping away at gene regulation. EMBO Rep 9:337–343.

Massie CE, Adryan B, Barbosa-Morais NL, Lynch AG, Tran MG, Neal DE, Mills IG. 2007. New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. EMBO Rep 8:871–878.

Mathur D, Danford TW, Boyer LA, Young RA, Gifford DK, Jaenisch R. 2008. Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. Genome Biol 9:R126.

Mito Y, Henikoff JG, Henikoff S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet 37:1090–1097.

Mohn F, Weber M, Schubeler D, Roloff TC. 2009. Methylated DNA Immunoprecipitation (MeDIP). Methods Mol Biol 507:55–64.

Okada M, Takezawa S, Mezaki Y, Yamaoka I, Takada I, Kitagawa H, Kato S. 2008. Switching of chromatin-remodelling complexes for oestrogen receptor-alpha. EMBO Rep 9:563–568.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122:517–527.

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. 2005. Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. Cell 123:233–248.

Rando OJ. 2007. Global patterns of histone modifications. Curr Opin Genet Dev 17:94–99.

Rao B, Shibata Y, Strahl BD, Lieb JD. 2005. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. Mol Cell Biol 25:9447–9459.

Roh TY, Ngau WC, Cui K, Landsman D, Zhao K. 2004. High-resolution genome-wide mapping of histone modifications. Nat Biotechnol 22:1013–1016.

Shi C, Zhou G, Zhu Y, Su Y, Cheng T, Zhau HE, Chung LW. 2008. Quantum dots-based multiplexed immunohistochemistry of protein expression in human prostate cancer cells. Eur J Histochem 52:127–134.

Shivaswamy S, Iyer VR. 2007. Genome-wide analysis of chromatin status using tiling microarrays. Methods 41:304–311.

Silberstein GB, Dressler GR, Van Horn K. 2002. Expression of the PAX2 oncogene in human breast cancer and its role in progesterone-dependent mammary growth. Oncogene 21:1009–1016.

Simonis M, Kooren J, de Laat W. 2007. An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 4:895–901.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131:58–69.

Vias M, Ramos-Montoya A, Mills IG. 2008. Terminal and progenitor lineage-survival oncogenes as cancer markers. Trends Mol Med 14:486–494.

Wang Q, Li W, Liu XS, Carroll JS, Janne OA, Keeton EK, Chinnaiyan AM, Pienta KJ, Brown M. 2007. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell 27: 380–392.

Weber M, Schubeler D. 2007. Genomic patterns of DNA methylation: Targets and function of an epigenetic mark. Curr Opin Cell Biol 19:273–280.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, 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.

Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT. 2008. Species-specific transcription in mice carrying human chromosome 21. Science 322:434–438.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309:626–630.

Zhang X, Guo C, Chen Y, Shulha HP, Schnetz MP, LaFramboise T, Bartels CF, Markowitz S, Weng Z, Scacheri PC, Wang Z. 2008. Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies. Nat Methods 5:163–165.