

# Interpreting Chromatin Immunoprecipitation Experiments

Kevin Struhl<sup>1,\*</sup>

<sup>1</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

\*Correspondence: kevin@hms.harvard.edu

## Introduction

Chromatin immunoprecipitation (ChIP) is a widely applied technique for measuring the association of proteins with specific genomic regions in living cells. Formaldehyde is used to generate protein-protein and protein-DNA cross-links between molecules located in close proximity *in vivo*, and the resulting material is fragmented (typically by sonication, but sometimes with micrococcal nuclease) to an average DNA length of 300–500 bp. This “input” sample is immunoprecipitated with an antibody against a desired protein, modified peptide (e.g., to detect acetylated, phosphorylated, or methylated versions of a protein), or epitope (when the protein of interest is epitope tagged), thereby selectively enriching DNA sequences that directly or indirectly (via another protein) crosslink with the desired protein (or modified variant). The basic measurement of any ChIP experiment is “immunoprecipitation efficiency” or “IP:input ratio” for a given genomic region, which is defined as the amount of PCR product in the IP sample divided by the amount of PCR product in the input sample. ChIP can be combined with microarray technology (Bernstein et al., 2004; Buck and Lieb, 2004; Ren and Dynlacht, 2004) or large-scale DNA sequencing (Impey et al., 2004; Loh et al., 2006; Wei et al., 2006) to identify protein-binding sites on a large-scale or unbiased genome-wide basis. Variants of ChIP can be used to measure RNA-protein interactions *in vivo* (Gilbert et al., 2004) or DNA-protein interactions *in vitro* (Liu et al., 2005).

ChIP experiments often involve a comparison of multiple samples representing different conditions. Protein occupancy *in vivo* can be analyzed under different environmental conditions, at different stages of the cell cycle, throughout a developmental time course, in different cell types, and in wild-type versus genetically altered cells. ChIP snapshots are useful for rapid kinetic analysis (time points every 30–60 s) of genomic processes (Katan-Khaykovich and Struhl, 2002; Bryant and Ptashne, 2003), because cellular enzymes are immediately inactivated upon formaldehyde addition. ChIP can also specifically follow the behavior of epitope-tagged mutant proteins unable to support cell growth, by analyzing them in cells containing an untagged version of the wild-type protein (Mencia and Struhl, 2001).

## Basic Controls for the ChIP Procedure

Although the ChIP procedure is straightforward and relatively standard (Aparicio et al., 2004), the quality of a ChIP experiment depends on (1) reproducible cell breakage, chromatin solubilization, and DNA fragmentation,

(2) quality and specificity of the antibody, (3) optimal and reproducible immunoprecipitation that maximizes recovery of the desired protein-DNA complexes while minimizing background DNA not crosslinked to the protein, and (4) high-quality quantitative PCR analysis, which is critically dependent on PCR primer design. Basic controls for these parameters are often not presented in publications, and hence not easily assessed. When quantitative PCR reactions are assessed by gel electrophoresis after a fixed number of cycles, it is desirable to show a standard curve indicating that the analysis is being performed in the linear range. Ideally, genomic regions of interest are analyzed in the same reaction mixture with a constant “reference region” that serves as an internal normalization control. Real-time PCR analysis largely eliminates these issues, although the reactions must be robust and reproducible (i.e., superimposable curves that show ~1.9-fold amplification per cycle).

ChIP experiments are quantitative, and they should be presented as such. Quantitative analysis requires an assessment of experimental variability at the level of individual samples, biological replicates of the samples performed under a single condition, and, in many cases, multiple samples representing different conditions. ChIP experiments typically involve at least three biological replicates (i.e., samples prepared from three independent populations of cells). Two biological replicates may be sufficient in cases where the effects are quantitatively robust, but more than three may be required in cases where the effects are subtle. Quantitative values from ChIP experiments should include error bars (typically standard deviations).

## Specificity Controls and Experimental Background

ChIP experiments typically involve “specificity controls” at the level of both DNA and protein. Specificity at the DNA level is accomplished by examining multiple genomic regions for a given pair of input and immunoprecipitated samples. This permits quantitative measurements of the relative levels of protein association with different genomic regions in an internally controlled manner. To assess specificity at the protein level, parallel immunoprecipitations of a given input sample are performed with the antibody of interest and with an unrelated (or no) antibody or preimmune serum. When appropriate, a comparable comparison can be made with parallel immunoprecipitations of input samples from strains expressing epitope-tagged or untagged versions of the protein of interest.

There are two types of experimental background that are linked to the specificity controls. True experimental background arises from DNA fragments in the IP sample that do not represent the desired DNA-protein complexes, and it can be assessed with the protein specificity controls. A different type of background arises from actual crosslinking of the protein that occurs all over the genome at a low level in a nonspecific manner. At “negative control” regions, immunoprecipitations with the antibody of interest often give slightly higher IP efficiencies than control immunoprecipitations. However, it is unclear whether this effect is physiologically meaningful (true nonspecific association) or an experimental artifact (random collisions between the protein and DNA that are captured by crosslinking).

Experimental design and interpretation of a ChIP experiment depends on the type of protein being analyzed. Many ChIP experiments involve proteins that are highly localized to specific genomic regions, either by direct DNA binding or by recruitment via interactions with DNA-bound proteins. In such cases, it is presumed that the protein of interest does not bind (or is weakly and nonspecifically associated) with the vast majority of genomic regions. Some proteins (e.g., RNA polymerase II and associated elongation factors) can associate with extended genomic regions, although association is presumed to be specific to these regions. On the other hand, some proteins (e.g., modified histones, HMG proteins, chromatin-modifying activities) associate to varying extents with many, and perhaps all, genomic regions. In the extreme case, histones associate with virtually the entire genome, although some genomic regions are relatively depleted.

### Basic Principles for Interpretation of ChIP Experiments

A key principle for interpreting ChIP experiments is that, for a given sample, IP efficiencies at different genomic regions reflect relative levels of protein association. This principle depends on the assumption that the physical crosslinking event for a given protein to the chromatin template is equally efficient at all genomic locations. Although difficult to prove rigorously, there are many lines of evidence consistent with this assumption, and clear cases of differential crosslinking efficiency due to locus-specific differences in DNA and/or protein conformation have yet to be described. Another key principle is that the level of protein association at a given locus represents a cell-averaged and time-averaged snapshot taken at the time of formaldehyde addition. For example, a value of X could reflect all cells having an occupancy level of X, or half the cells in a population having a value of 2X and the remaining half having no occupancy at all. In addition, ChIP cannot distinguish between subpopulations of a given protein that have differential kinetic or stability properties when associating with genomic sequences.

Absolute IP efficiencies and fold enrichments cannot be used to compare binding characteristics of different proteins, to provide absolute measurements of protein occupancy on specific genomic regions, or to determine relative stoichiometry of factors on a given sequence. Crosslinking per se is very inefficient, and the number and physical location of amino acid and nucleotide residues within the interacting protein surfaces that react with formaldehyde vary considerably among protein-protein and protein-DNA interactions. Proteins directly interacting with DNA can be crosslinked by a single event, usually but not always resulting in higher IP efficiencies than proteins that indirectly associate with DNA and require multiple crosslinking events. Proteins also vary in whether they stably or transiently associate with genomic sequences. Lastly, the absolute IP efficiency depends on the quality of the antibody-antigen interaction as well as the antibody concentration, and the fold enrichment depends on both the absolute IP efficiency and on the background. These considerations can also apply to a comparison of wild-type and mutant (or variant) forms of the same protein. A mutant protein might lack sequences that affect interactions with the antibody and/or crosslinking to DNA.

In certain situations, ChIP can be used to estimate absolute levels of protein occupancy. The principle is that there is a maximal ChIP signal that corresponds to 100% occupancy, that is, all the time in all cells. It is reasonable to suppose that a consistent, maximal ChIP signal among multiple loci might represent complete occupancy (e.g., very highly transcribed genes in the case of initiation factors or totally repressed genes in the case of repressors). Given this assumption, absolute levels can be estimated for any genomic region analyzed in the same samples.

### Antibody Specificity and Accessibility

Ideally, antibody specificity is assessed in the context of a ChIP experiment by using mutant cells lacking the protein of interest or lacking the potential to make the modification of interest. Similarly, the use of epitope-tagged proteins permits the control of an untagged strain. In organisms where such genetic approaches are difficult, RNAi can be used to deplete the protein of interest and reduce the ChIP signal. However, RNAi rarely eliminates the protein, and the relationship between protein level and ChIP signal is not straightforward. An alternative approach, often not possible, is to use two different antibodies to a given protein and show that the results are similar.

Unfortunately, a rigorous antibody control is often not available. Western blotting is often used to assess antibody quality, but there is only a modest relationship between the apparent specificity of a protein on a western blot and specificity in a ChIP experiment. In general, artifactual ChIP results due to crossreactivity of antibodies are rare, because crossreacting proteins are unlikely to associate with specific genomic regions

and to be immunoprecipitated with high efficiency. However, when analyzing a member of a multiprotein family, it is important to assess crossreactivity of the antibody with other family members. Lastly, antibody specificity is often inferred from ChIP results that meet expectations from other lines of experiments. For example, cells with a mutation in a putative binding site represent a useful control for DNA-binding proteins, as do mutant cells or environmental conditions suspected to be important for the protein of interest. Of course, such “control” experiments represent circular reasoning.

A hypothetical consideration is that the relevant epitope might be inaccessible to the antibody in the context of crosslinked chromatin, thereby generating a false negative result. This concern is extremely unlikely with epitope-tagged proteins, because the epitope is unlikely to have a specific interaction with other proteins or DNA sequences. It is minimized when using polyclonal antibodies that recognize multiple determinants within a protein. More generally, a protein or DNA interaction that masks an epitope is unlikely to be efficiently crosslinked and unlikely to persist after the denaturation step that occurs prior to the immunoprecipitation. Overall, epitope masking is not a significant concern, although it might be relevant when a protein of interest does not appear to interact with any genomic sequence.

A more significant, but unappreciated, problem arises when using antibodies against specific peptides corresponding to a modified (e.g., acetylated, phosphorylated, or methylated) histone or other protein. Formaldehyde modifies lysine residues at high efficiency, even if this does not result in crosslinking. If a lysine is near the modified residue of interest and part of the epitope recognized by the antibody, the IP efficiency will appear low even though the protein associates with the genomic region. In such cases, it might be useful to decrease the crosslinking time or formaldehyde concentration.

### **Analysis of Proteins that Associate with Specific Genomic Regions**

For proteins that associate specifically with certain genomic regions, the best way to interpret the data is to compare the IP efficiencies for different genomic regions from the same input and IP samples. This approach utilizes the concept of a “background” level for all “negative control” fragments that do not associate specifically with the protein of interest. Background defined in this manner includes true experimental background as well as any nonspecific binding (meaningful or artifactual). A typical background level is ~0.03%, although this can vary depending on the antibody and the elution method. The choice of suitable negative control regions is often based on expectation; for example, the central portions of protein-coding regions are unlikely to associate with transcriptional initiation factors or specific DNA-binding proteins. In the absence of any expectation, the background level can only be based on multiple regions having similar IP efficiencies at the level of a typical back-

ground. It is highly desirable that ChIP experiments analyze multiple negative control regions, both to confirm that these regions are truly negative controls and to permit an accurate assessment of the background level and experimental error.

ChIP measurements of this type represent specific binding, with the fold enrichment of a genomic region over the background being related to the relative level of protein association *in vivo*. It is useful to define “relative protein occupancies” for different regions by subtracting the background from the observed IP efficiencies. For example, if the background level is defined as one unit, a genomic region showing 6-fold enrichment over background will have five occupancy units. As absolute IP efficiencies and fold enrichments can vary among replicate experiments, it is useful to arbitrarily define a “positive control” genomic region(s) as having a fixed number of occupancy units in all biological replicates. In this way, occupancy units for other genomic regions will be defined relative to that of the “positive control” in the same pair of input and IP samples. The advantage of this approach is that relative levels of protein association with different genomic regions are largely unaffected by differences in absolute IP efficiencies and fold enrichments.

### **Determining Relative Occupancy Levels of Different Proteins at Specific Genomic Regions**

Relative occupancy levels of different proteins at genomic regions can be determined by performing parallel immunoprecipitations with different antibodies, ideally on the same crosslinked chromatin sample. Occupancy units for individual factors are determined independently as described above, and occupancy ratios are defined in arbitrary units. The relative occupancy ratios for the different genomic regions are valid, but they do not represent an absolute stoichiometric relationship. To account for potential sample-to-sample variations among replicate experiments, a given occupancy ratio should be defined for a specific genomic region and ratios at all other genomic regions calculated in relative terms. Using this rationale, the relative associations of TBP and the general transcription factors TFIIA and TFIIB were shown to be essentially constant at all promoters, whereas the TAF:TBP occupancy ratios vary considerably (Kuras et al., 2000). Occupancy ratios determined from such experiments do not address whether two proteins co-occupy a given genomic region or mutually compete for the same genomic region.

### **Sequential ChIP for Determining Co-Occupancy of Two Proteins to Specific Genomic Regions**

Sequential chromatin immunoprecipitation (SeqChIP; also referred to as Re-ChIP, ChDIP, and double ChIP) can determine whether two proteins can simultaneously associate with the same genomic region *in vivo*. In SeqChIP, protein-DNA complexes from the first immunoprecipitation are subjected to an additional

immunoprecipitation with an antibody of a different specificity. The key concept for quantitating SeqChIP experiments is that, if two proteins completely co-occupy DNA, the fold enrichment of the sequential ChIP should be equal to the product of the fold enrichments of the individual ChIPs (Geisberg and Struhl, 2004). By performing SeqChIP experiments in both directions, this approach can be used to distinguish between complete, partial, and no co-occupancy of two factors on specific genomic regions.

### Mapping Protein-Binding Sites by ChIP

Due to the size of DNA in the crosslinked chromatin samples, ChIP detection of protein association to a genomic region defined by a single PCR product only roughly localizes where the protein actually binds *in vivo*. However, a protein bound to a specific DNA site generates a predicted ChIP profile that depends on the size of the fragmented DNA in the crosslinked chromatin sample and the length of the PCR products used for the analysis (Kadosh and Struhl, 1998). The predicted profile is trapezoidal, with peak ChIP signals centered at the actual binding site and extending the length of the PCR product. ChIP signals gradually and symmetrically decrease at regions flanking the peak in a manner that depends on DNA fragment length. As a consequence, protein-binding sites can be mapped to ~50 bp resolution with a sufficient number of PCR primer pairs that define closely located or overlapping regions spanning the genomic locus (Hall et al., 2006). The binding site is defined by the center of the peak, much in the same manner that peaks are defined in biochemical fractionations.

### Analysis of Proteins that Associate with Many or All Genomic Regions

For some proteins, it is inappropriate to interpret the data in terms of occupancy units and specific versus nonspecific binding sites. For example, histones associate with essentially all genomic regions, and the level of a particular chromatin modification typically occurs in a continuum. Similar issues are likely to pertain to HMG and other nonhistone proteins as well as chromatin-modifying activities that modify histones on a genome-wide basis. In these cases, the concept of "negative control" regions all giving the same IP efficiency is incorrect.

Instead, control immunoprecipitations with preimmune serum or with an antibody to a protein not found in the cell are used as a negative control. Quantitative analysis of the relative level of protein association is then presented as simple IP efficiencies in which the background contribution is subtracted. When IP efficiencies for the protein of interest are relatively high, background subtraction has little effect, and the requirement for the negative control is less important. It is difficult to determine whether a given region is devoid of a particular modification, because this involves subtraction of two small numbers that are

prone to error. As discussed above, variations among replicate samples are best addressed by giving a specific genomic region an arbitrarily defined value, which is used to determine the relative levels of all other genomic regions.

### Comparing Protein Occupancy in Different Biological Samples

ChIP experiments often involve a comparison of multiple samples representing different physiological conditions, cell types, or genetically modified variants. Variability in IP efficiency and experimental background among the different samples is a key issue in such comparisons, and it is difficult to control internally for these parameters. This is unlike ChIP analyses on an individual sample, where protein association at different genomic regions is expressed in relative terms that are internally controlled within each sample. In theory, positive and negative control regions that are predicted to be unaffected by the physiological condition or genetic constitution can be used for normalization and background subtraction of different biological samples. Although useful, this approach relies on assumptions that are difficult to prove, and in any event represents circular reasoning. Moreover, this approach cannot be used in situations where the physiological or genetic condition has a general effect on protein association (e.g., reducing protein binding at most or all target sites). In such cases, one must rely solely on sample-to-sample reproducibility from independent trials of the same experiment. As a consequence, comparing protein occupancy among different biological samples is inherently more error prone than analyses of individual ChIP samples.

### Assessing Experimental Error and Significance

An individual ChIP measurement is typically presented as the average value of replicate samples (usually at least three). Experimental error is often expressed as the standard deviation for each ChIP measurement and shown as error bars in figures. Alternatively, an average standard deviation for the entire ChIP experiment can be applied to individual values (except in cases where the variance seems particularly large). Assessment of experimental error is more accurate with more individual measurements, so it is particularly useful to assay multiple positive and negative controls in addition to the genomic regions of interest.

In the many experiments where measurements are relative within the same pair of input:IP samples and normalized to a positive control(s), sample-sample variability in immunoprecipitation efficiency is largely factored out. In such cases, a typical standard deviation is ~±25% of the mean, and with three biological replicates, a 2-fold difference between two regions has a *p* value of ~0.05, which represents the conventional limit for a believable effect. With more biological replicates or better reproducibility, effects that are less than 2-fold can be measured reliably. In addition, one can obtain much more

accurate data (standard deviations ~5%) in experiments where a single PCR reaction is used to simultaneously compare protein association at two alleles that generate different-sized PCR products (Katan-Khaykovich and Struhl, 2002). However, in experiments involving multiple biological samples, there is no easy way to control for variability in immunoprecipitation efficiency, and it is difficult to define a positive control that should be quantitatively unchanged among the different samples. As a consequence, comparisons among different biological samples are more error prone, making it necessary to perform more biological replicates to get information of comparable accuracy.

### Biological Information from ChIP Experiments

Given highly likely assumptions, ChIP provides a quantitative measure of the relative levels of protein association at genomic regions *in vivo*. Thus, when protein association is detected at one or more regions, the failure to observe protein association at other regions indicates a lack of detectable binding and cannot be dismissed as an uninterpretable negative result. The detection limit varies greatly among ChIP experiments, and assay sensitivity increases in accord with the fold enrichment observed at the genomic region with the highest level of protein occupancy. On the other hand, the failure to observe protein association at any genomic region tested cannot be interpreted, unless such association is observed in other biological samples that are analyzed in parallel.

By definition, ChIP represents a snapshot of a true physiological condition, and the observed protein association does not need to be validated by any independent line of experimentation. However, ChIP does not provide any information about how the protein associates with genomic regions, with the exception that analysis of mutant cells can identify which part(s) of the genomic region is required for the association. Moreover, ChIP cannot distinguish whether a protein binds DNA directly or associates with the genomic region indirectly via protein-protein interactions. Biochemical analysis of the protein-DNA interaction is valuable to determine the basis of protein-DNA association *in vivo*, but it is inappropriate to use it as validation for the ChIP result *per se*.

Because ChIP can only measure protein association, it does not provide any information about the function of protein bound at the genomic region(s) of interest. Whereas ChIP results are very useful for generating or invalidating hypotheses, knowledge of the function of the bound protein must come from other lines of experimentation such as transcriptional or genetic analysis. Thus, ChIP provides crucial information about biological phenomena that is extremely difficult to obtain by any other method, but it needs to be integrated with other experimental information to elucidate the underlying molecular mechanisms that occur in living cells.

### REFERENCES

- Aparicio, O.M., Geisberg, J.V., and Struhl, K. (2004). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences *in vivo*. In *Current Protocols in Molecular Biology*, F.A. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (New York: John Wiley & Sons), pp. 21.3.1–21.3.17.
- Bernstein, B.E., Humphrey, E.L., Liu, C.L., and Schreiber, S.L. (2004). The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol.* **376**, 349–360.
- Bryant, G.O., and Ptashne, M. (2003). Independent recruitment *in vivo* by Gal4 of two complexes required for transcription. *Mol. Cell* **11**, 1301–1309.
- Buck, M.J., and Lieb, J.D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360.
- Geisberg, J.V., and Struhl, K. (2004). Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions *in vivo*. *Nucleic Acids Res.* **32**, e151.
- Gilbert, C., Kristjuhan, A., Winkler, G.S., and Svejstrup, J.Q. (2004). Elongator interactions with nascent mRNA revealed by RNA immunoprecipitation. *Mol. Cell* **14**, 457–464.
- Hall, D.B., Wade, J.T., and Struhl, K. (2006). An HMG protein, Hmo1, associates with promoters of many ribosomal protein genes and throughout the rRNA gene locus in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **26**, 3672–3679.
- 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.
- Kadosh, D., and Struhl, K. (1998). Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin *in vivo*. *Mol. Cell. Biol.* **18**, 5121–5127.
- Katan-Khaykovich, Y., and Struhl, K. (2002). Dynamics of global histone acetylation and deacetylation *in vivo*: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev.* **16**, 743–752.
- Kuras, L., Kosa, P., Mencia, M., and Struhl, K. (2000). TAF-containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science* **288**, 1244–1248.
- Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. (2005). DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res.* **15**, 421–427.
- Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440.
- Mencia, M., and Struhl, K. (2001). A region of TAF130 required for the TFIID complex to associate with promoters. *Mol. Cell. Biol.* **21**, 1145–1154.
- Ren, B., and Dynlacht, B.D. (2004). Use of chromatin immunoprecipitation assays in genome-wide location analysis of mammalian transcription factors. *Methods Enzymol.* **376**, 304–315.
- Wei, C.L., Wu, Q., Vega, V., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y.T., Weng, Z., et al. (2006). A global map of p53 transcription factor binding sites in the human genome. *Cell* **124**, 207–219.

### Please cite this article as:

Struhl, K. (2007). Interpreting Chromatin Immunoprecipitation Experiments. In *Evaluating Techniques in Biochemical Research*, D. Zuk, ed. (Cambridge, MA: Cell Press), <http://www.cellpress.com/misc/page?page=ETBR>.