

# Whitepaper

## *Using Chromatin Immunoprecipitation (ChIP) to Determine Protein Binding Sites on DNA*

This whitepaper discusses the use of chromatin immunoprecipitation arrays to identify binding sites for transcription factors, determine areas of active transcription, or assess modifications of genome structure by histone-binding in a region of the genome.

---

### Introduction

Recently, researchers have begun using microarrays to identify the DNA products from chromatin immunoprecipitations (ChIP) as a way to determine protein binding sites on DNA (1, 2, 3). The basic concept is simple: the ChIP precipitates the DNA sequences bound by the protein of interest. The researcher then takes the precipitated DNA, labels it, and hybridizes it to a microarray. The resulting microarray data indicate the DNA sequences that were bound by the protein of interest.

ChIP reactions can use antibodies to precipitate DNA bound by transcription factors, polymerases, histones, or any other protein. Therefore, researchers are able to identify binding sites for transcription factors, determine areas of active transcription, or assess the modifications of the histones in a region of the genome using this method.

### Identifying Target Genes of Human Transcription Factors

One of the most common uses of ChIP array experiments is the determination of binding sites for transcription factors. Transcription factors are regulatory proteins that bind to DNA and are involved in the control of gene expression. It is well understood that deregulation of certain transcription factors plays a significant role in human cancer development. For example, overexpression of the transcription factor c-Myc is found in 80% of breast cancers, 70% of colon cancers, 90% of gynecological cancers, and 50% of hepatocellular carcinomas (4). Therefore, understanding the role of these transcription factors and their targets is of central interest to cancer researchers.

Sequencing of the human genome has provided a wealth of information that is now allowing rapid advances in the understanding of gene regulation. For example, 5' exon prediction software has provided estimates of the number of promoters in the human genome. Davuluri et al. have developed software called FirstEF and used it to predict 68,645 first exons (5). Of these, 30,000 are considered highly probable promoters. Such information allows the analysis of thousands of promoter sequences and provides a comprehensive dataset for a bioinformatics approach to finding target promoters of human transcription factors.

Using a bioinformatics approach, software programs have been developed to identify E2F sites in sets of mammalian promoters (6, 7). Similarly, scanning the human genome for Myc binding sites has identified a large set of promoters that contain Myc consensus sites (8). Follow-up analyses of the predicted E2F and Myc target promoters using chromatin immunoprecipitation assays have indicated that many of the putative targets

**Bioinformatics  
Combined with  
Experimental  
Methods**

are bound by E2F or Myc family members in living cells (7, 8). Therefore such bioinformatics approaches are very instructive and have provided researchers with a large number of potential target genes.

Two major concerns arise when one relies solely upon a bioinformatics approach. First, many consensus sites may reside within inactive chromatin and may be inaccessible for interaction with a transcription factor. For example, computational identification of binding sites for the PIF/GMEB transcription factor identified sites which were bound *in vitro* by the factor, but which were extremely sensitive to the methylation status of the DNA (9). Thus these sites would not be bound if the consensus site resided within hypermethylated heterochromatin. Similarly a binding site may reside within active euchromatin but be inaccessible due to specific nucleosomal positioning. Also the overlap of a consensus binding site with a site for a more abundant and/or higher affinity DNA binding protein can prevent *in vivo* binding. For example the consensus Ets site in the c-Myc promoter is occupied *in vivo* only if the overlapping E2F site is mutated (10). Thus methylation, nucleosomal positioning, and overlapping binding sites can lead to the identification of a set of false-positive target promoters using *in silico* methods.

The second major concern is that many true positives may be overlooked if the transcription factor is recruited to the DNA via a site that does not exactly match a consensus site. Recruitment via a non-consensus site could be accomplished via protein-protein interaction. For example, the interaction between YY1 and E2F2 is believed to allow stable recruitment of E2F2 to the cdc6 promoter (11). Such promoters would be overlooked by identification schema that rely solely upon the presence of a high affinity consensus site in a promoter region. Also transcription factors such as steroid hormone receptors and POU domain proteins can recognize a variety of similar, but not identical, sites (12, 13). Thus, the lack of a strict consensus can make it difficult to identify true target promoters using only sequence analysis software.

Clearly, experimental methods that can allow global analyses and identification of binding sites for human transcription factors would complement bioinformatics approaches.

### ChIP Assay Considerations

Over the last several years, experimental methods have been developed to identify genomic targets of DNA-binding factors. All of these approaches utilize as a first step the chromatin immunoprecipitation (ChIP) assay to selectively enrich for chromatin fragments bound by transcription factors. Researchers then use the chromatin fragments to probe microarrays containing intergenic sequences. The list of factors that have been analyzed on a genome-wide basis using yeast as a model system is quite large (1, 2, 14, 15, 16, 17, 18). In contrast, fewer studies have been performed using the human genome, mostly due to the lack of availability of appropriate microarrays.

Several groups have begun creating their own specific human intergenic microarrays. A recent study utilized a DNA microarray comprising nearly all of the nonrepetitive sequences of human chromosome 22 (19). However, the published use of this new array has been limited to analysis of mRNA expression profiles. In a separate set of studies, two different approaches—both using spotted PCR fragments—have successfully been used to identify target genes of human transcription factors. The microarrays used in these studies differ in whether they contain researcher-selected promoter regions or randomly chosen clones. A PCR fragment microarray including the sequence spanning from -700 to +200 of 1,444 human genes has been utilized to find novel promoters bound by E2F (20). Such an array by necessity is biased in that the particular 1,444 promoters analyzed were researcher selected, or regions based on a set of criteria. Using this approach, it would be feasible to prepare PCR fragments representing the sequences upstream of all predicted first exons in the human genome. However, this Herculean project has not been attempted to date. As an intermediate step, researchers have utilized a microarray that contains 7,776 CpG island clones (1, 21). CpG islands tend to be found in intergenic regions and at the 5' ends of genes. Thus, use of CpG arrays allows a relatively unbiased analysis of many thousands of human intergenic regions. However a distinct disadvantage to these arrays is that not all human promoters contain CpG islands.

Clearly an alternative solution is necessary, and NimbleGen offers a flexible and powerful solution.

### The NimbleGen Solution

#### Custom ChIP Microarrays

Roche NimbleGen offers researchers the ability to look for protein binding sites in any region of a genome at any resolution. This service uses custom tiling arrays that contain approximately 380,000 60mer oligonucleotide probes per array. These probes can be spaced at user-defined intervals anywhere in the genome and can also be designed to hybridize to customer-selected control sequences.

NimbleGen DNA microarray flexibility is based on the Maskless Array Synthesizer (MAS) technology. This technology, developed by Roche NimbleGen, retains the advantages of *in situ* light-mediated DNA array synthesis but avoids the flexibility limitations associated with existing DNA array technologies. NimbleGen has adapted a digital processor, initially designed for computer projection and digital cinema applications, to the field of DNA microarray technology (22, 23). The heart of the MAS technology is the DMD (Digital Micromirror Device), manufactured by Texas Instruments. The DMD has individual, addressable mirrors

that control spatial light projection during DNA chip synthesis. This development opens up dramatic new possibilities for DNA array synthesis, such as:

- limited production run DNA chips
- multiple genome chips
- complete design flexibility
- iterative design refinement
- probe set optimization

Enormous chip volumes are no longer required to justify the high up-front cost associated with traditional chip production.

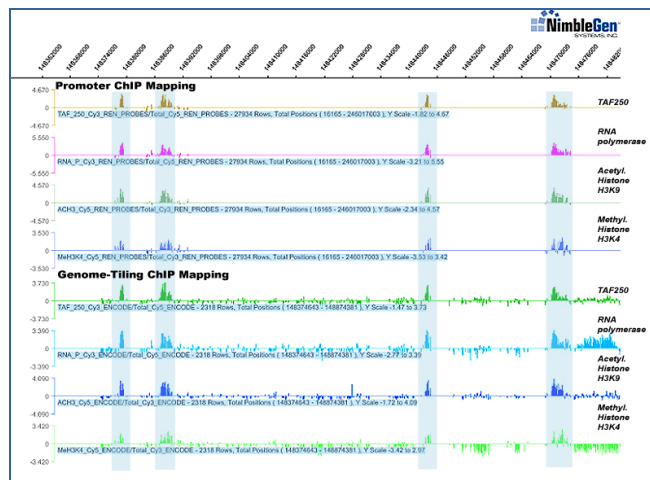
### Service Overview

Customers send Roche NimbleGen their ChIP samples, define the region of the genome in which they are interested, and specify the probe spacing. Roche NimbleGen then performs the following services:

- array design, based on consultation with customer
- sample labeling (post immunoprecipitation and PCR)
- array manufacture
- sample hybridization
- array scanning
- image gridding and data extraction

### Preliminary Data Analysis and Graphic Presentation

The graphical output displays the ratio of enrichment of the sample of interest versus the control sample mapped to the chromosomal location. Depending on the state of the annotation of the genome selected, known genes, exons, and repetitive elements will also be mapped to facilitate interpretation of the results. Figure 1 illustrates an example of the data output from this service.



**Figure 1.** NimbleGen ChIP graphical data output allows for rapid scanning of results that otherwise can be overwhelming as a result of the sheer number of data points generated during an experiment. It also provides a very useful way of archiving results and comparing results from different experiments. In this figure, results from “Promoter ChIP Mapping” of human chromosome 1 and “Genome-Tiling ChIP Mapping” through this same chromosome reveal identical hits (highlighted columns).

► For details on the ChIP array specifications, please see the *NimbleGen Chromatin Immunoprecipitation (ChIP) Microarrays and Services datasheet*.

### ChIP Protocol for Human Cells

The chromatin immunoprecipitation assay has been used for many years to study protein/DNA interactions in living cells. However, much of the initial work used this assay to study yeast or *Drosophila* proteins. More recently, several labs have adapted this assay for the analysis of mammalian site-specific DNA binding factors, including the Farnham lab (24) at the University of California, Davis, and the Ren lab at the University California, San Diego Branch of the Ludwig Institute for Cancer Research (LICR) (25). Roche NimbleGen recommends the protocols developed at these two labs for researchers working with human cells.

A typical experiment consists of the following samples and controls:

- Total DNA. The "total" sample is an aliquot of chromatin that has not been immunoprecipitated but has been crosslinked, sonicated, and processed in a manner identical to the experimental samples.
- Protein of interest. Antibody immunoprecipitated DNA.
- A control sample in which primary antibody was omitted (no Ab control).

All labeling reactions of the ChIP fragments are performed by Roche NimbleGen as part of the ChIP array service. This requirement ensures reproducible results and allows us to spike controls into the labeling reaction to assess the quality of the labeling reaction and the array hybridization.

### Array Design

Experiments performed by Roche NimbleGen indicate that a tiling array with a probe interval of 40 - 100 bases is optimum for ChIP studies designed to map transcription factor binding sites in the human genome. The exact interval chosen depends on the ChIP fragment size, the level of resolution desired, and the amount of genomic DNA to be covered. At this recommended coverage, the array will cover about 40 megabases of genomic DNA with 60mer probes.

For ChIP studies looking at histone modifications, the probe interval can be larger, and we recommend a spacing of 500 bases. This spacing allows the researcher to cover nearly 200 megabases of genomic DNA per array. The researcher also has the option of having replicates on the array or tiling both strands of DNA to increase the statistical significance of the experiment.

## ChIP Results

Results from the experiments are provided in three forms:

- original array images from the experiments.
- extracted data from the images in a format that can be immediately imported into a commercial analysis package, such as Spotfire or GeneSpring.
- graphical output such as that shown in Figure 1.

This graphical format allows for rapid scanning of results that otherwise can be overwhelming as a result of the sheer number of data points generated during an experiment. It also provides a very useful way of archiving results and comparing results from different experiments.

## Conclusion

For researchers interested in studying transcriptional regulation, the histone code, polymerase activity, or any protein/DNA binding, the NimbleGen ChIP array service offers a unique opportunity to study these processes in fine detail across any region of a genome.

## References

1. Ng, H. H., et al. Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. *Genes Dev.*: 16: 806-819, 2002.
2. Ren, B., et al. Genome-wide location and function of DNA binding proteins. *Science*: 290: 2306-2309, 2000.
3. Weinmann, A. S., et al. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.*: 16: 235-244, 2002.
4. Gardner, L., et al. The c-Myc Oncogenic Transcription Factor. *The Encyclopedia of Cancer*, Second Ed. 2002.
5. Davuluri, R. V., et al. Computational identification of promoters and first exons in the human genome. *Nat. Genet.*: 29: 412-417, 2001.
6. Aerts, S., et al. Toucan: deciphering the cis-regulatory logic of co-regulated genes. *Nuc. Acids Res.*: 31: 1753-1764, 2003.
7. Kel, A. E., et al. Computer-assisted identification of cell cycle-related genes - new targets for E2F transcription factors. *J. Mol. Biol.*: 309: 99-120, 2001.
8. Fernandez, P. C., et al. Genomic targets of the human c-Myc protein. *Genes Dev.*: 17: 1115-1129, 2003.
9. Burnett, E., Christensen, J., and Tattersall, P. A consensus DNA recognition motif for two KDWK transcription factors identifies flexible-length, CpG-methylation sensitive cognate binding sites in the majority of human promoters. *J. Mol. Biol.*: 314: 1029-1039, 2001.
10. Albert, T., et al. The chromatin structure of the dual c-Myc promoter P1/P2 is regulated by separate elements. *J. Biol. Chem.*: 276: 20482-20490, 2001.
11. Schlisio, S., et al. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J.*: 21: 5775-5786, 2002.
12. Loven, M. A., et al. Estrogen response elements alter coactivator recruitment through allosteric modulation of estrogen receptor beta conformation. *J. Biol. Chem.*: 276: 45282-45288, 2001.
13. Millevoi, S., et al. Atypical binding of the neuronal POU protein N-Oct3 to noncanonical DNA targets. Implications for heterodimerization with HNF-3 beta. *Eur. J. Biochem.*: 268: 781-791, 2001.
14. Damelin, M., et al. The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress. *Mol. Cell.*: 9: 563-573, 2002.
15. Iyer, V. R., et al. Genomic binding sites of the yeast cell-cycle transcription factor SBF and MBF. *Nature*: 409: 533-538, 2001.
16. Lee, T. I., et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*: 298: 799-804, 2002.
17. Lieb, J. D., et al. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.*: 28: 327-334, 2001.
18. Wyrick, J. J., et al. Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science*: 294: 2357-2360, 2001.
19. Rinn, J. L., et al. The transcriptional activity of human chromosome 22. *Genes Dev.*: 17: 529-540, 2003.
20. Ren, B., et al. E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. *Genes Dev.*: 16: 245-256, 2002.
21. Wells, J., et al. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene*: 22: 1445-1460, 2003.
22. Nuwaysir, E. F., et al. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.*: 12: 1749-1755, 2002.
23. Singh-Gasson, S., et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotech.*: 17: 974-978, 1999.
24. Kirmizis, A., et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.*: 18:1592-1605, 2004.
25. Kim, T.H., et al. A high-resolution map of active promoters in the human genome. *Nature* advance online publication; published online 29 June 2005.

\*\*This research was funded by NIH Grant 1R01 HG003129-01.

## For More Information

Toll-free in US: (877) NimbleGen / (877) 646-2534  
(608) 218-7600

ngsales@nimblegen.com

www.nimblegen.com



HIGH - DEFINITION GENOMICS™

© June 2008 Roche NimbleGen, Inc. All Rights Reserved.  
05227917001 • Reprinted 06/08 • Original Publication 09/05



Roche NimbleGen, Inc.  
500 S. Rosa Road  
Madison, WI 53719 USA