

ChIP-on-chip Experimental Design  
QuickStart Guide



# This guide is intended for use in conjunction with the Affymetrix Chromatin Immunoprecipitation (ChIP) Protocol, available for download at [www.affymetrix.com/support](http://www.affymetrix.com/support).

The guide is a collation of all protocol information, hints and tips shared by highly experienced ChIP-on-chip researchers, at key Affymetrix ChIP-on-chip workshops and scientific meetings. For more information and online webforum practical help, go to [www.chiponchip.net](http://www.chiponchip.net). This webforum is hosted by Affymetrix to enable sharing of protocol information and other useful information between labs. On this website, you will also find details of regional workshops, practical courses and larger scientific meetings.

## Sections:

1. Antibody Selection & Validation
2. Fixation
3. Sonication
4. IP
5. Amplification
6. Control Options: Input/Mock IP/Non-specific Ab?
7. Sample Preparation & Hybridization
8. Data Analysis
9. Validation of Data

# ChIP-on-chip experimental design considerations

Genome-wide analysis of transcription regulation using ChIP-on-chip identifies disease-related modifications at the epigenetic and transcription machinery levels, including histone modification (blue), transcription factor binding (purple) and DNA methylation (green).

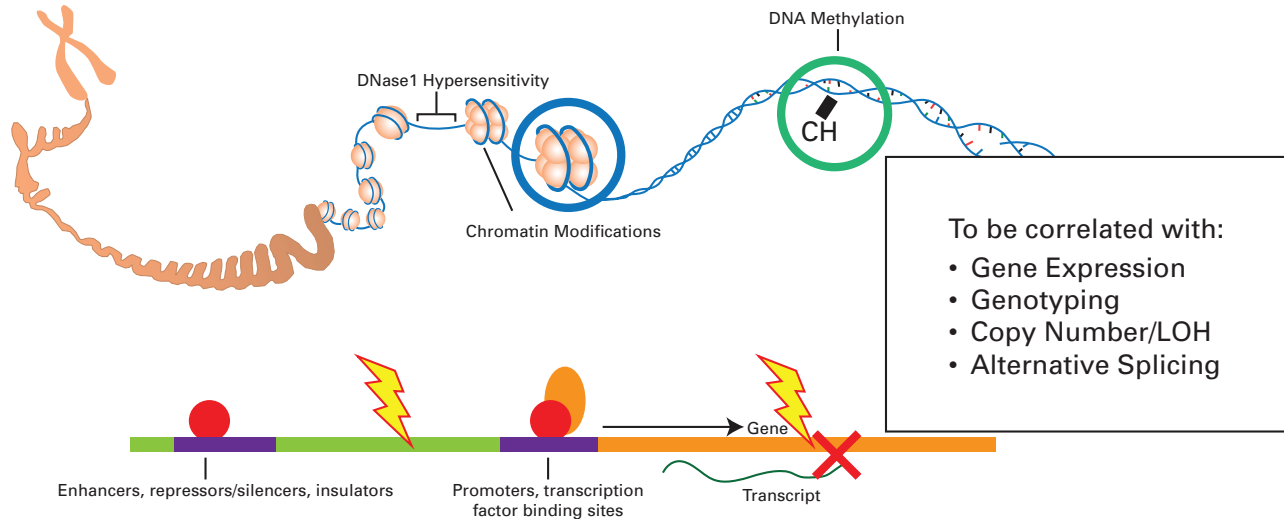


Figure 1. ChIP-on-chip is a powerful technique that can be used to identify sites of transcription factor binding (purple), as well as sites of DNA-methylation (green) or histone modification (blue).



## **There is no single protocol that will work for every factor that you wish to analyze.**

Some consideration should be given to genome size and complexity, and the expected frequency of the binding events you are analyzing. The Affymetrix ChIP-on-chip protocol was co-developed by Mark Biggin and Xiao-Yong Li of the Lawrence Berkeley National Laboratory, who were specifically looking for transcription factor binding in the *Drosophila* genome.

If you are looking for more frequent events, such as histone modification or DNA methylation, you will need to optimize the protocol accordingly. In general, the more complex the genome and the more frequent the binding event, the higher stringency you will need to apply to antibody selection, etc.

We have tried to summarize the main points in the workflow, which may require some degree of optimization.

## Hints and Tips...

---

- The antibody is probably the single most critical point of your experimental design. It is highly recommended to spend time validating the specificity and affinity of your antibody, as the rest of the workflow will be dependent on its success.
- Polyclonal antibodies have higher “pass” rate for ChIP applications (recognize multiple epitopes), but suffer more batch-to-batch variation.
- Pass rate for ChIP is generally very low (20% for TF’s, 50% for histone modifications).
- Abcam produces a very wide range of ChIP-grade antibodies, and performs batch-to-batch testing on a subset of these. Visit [www.abcam.com](http://www.abcam.com) for more information.
- Success in ChIP-qPCR is not a guarantee of success in ChIP-on-chip, because the enrichment factor required for array detection tends to be higher (>10X).
- If you are looking for more frequent binding events (e.g. histone modifications), and your antibody is highly specific, you may be able to get away with lower enrichment factors (e.g. >3X).
- If there are no known positive sequences, which can be used as a control for your target protein, you could try to select probable targets from expression and/or knockdown data.
- In the worst case, where you cannot predict ANY potential sites of binding, you may want to run one single hybridization, and identify putative sites of enrichment from this; however, this should be done as a last resort.

# Antibody Validation and Selection

---

- If possible, use a commercially available ChIP-grade antibody that has been tested, not only for performance in ChIP, but also for batch-to-batch variation.
- If no commercial ChIP-grade antibody is available, or if you want to use a homemade antibody, this should be tested for specificity and affinity/enrichment factor. As a minimum, we recommend:
  - Testing specificity with different peptides in an ELISA assay, peptide blocking in western blot, or using an RNAi knockdown of your target protein.
  - Testing specificity in your ChIP assay by selecting known positive sequences and negative sequences by RT-PCR post-IP.
- Western blotting is a good indicator of success. Antibodies that work in immunoprecipitation and immunohistochemistry are likely to be successful.
- You always need to test enrichment factor in your ChIP assay (see below). In general, we recommend a minimum enrichment factor of 10-fold for success in ChIP-on-chip.
- Use 1-10  $\mu\text{g}$  antibody per ChIP (25  $\mu\text{g}$  chromatin), and check that signal:background is good. You can pre-couple the antibody to your beads, and then can use up to 15  $\mu\text{g}$  antibody.
- Test different wash conditions to make as stringent as possible (150 – 500 mM NaCl).

## Hints and Tips...

---

- Make sure your formaldehyde contains MeOH – this prevents dimerization of the formaldehyde and subsequent insoluble complex formation.
- If you are examining co-factors and other protein:protein interactions, as well as your protein:DNA, you may want to use a second cross-linking agent in addition to formaldehyde. Formaldehyde is inefficient at cross-linking protein to protein. You may want to try agents such as DDMP.
- Formaldehyde does destroy proteases, but many labs use protease inhibitor as well – cocktails are available from Sigma and Roche. Roche's contains PMSF, which can block epitopes, but has a very short half-life in aqueous solution (35 mins pH8.0), so should not be a problem.

# Fixation

---

- Start with  $0.5 - 2 \times 10^8$  cells.
- You can use small independent IPs and then pool.
- We recommend fixation in 1% formaldehyde for 10 minutes.
- Some have reported success with shorter times (2 mins) and / or lower concentrations of formaldehyde (i.e. final concentration of 0.37%).
- Ideally, aim to minimize the fixation, to avoid formation of insoluble fractions that are then impossible to fragment by sonication.
- If working with tissue samples, grind the tissue as much as possible, freeze/thaw and then add paraformaldehyde solution. Freezing in liquid nitrogen is recommended.

## Hints and Tips...

---

- If sonication efficiency is low, check that cells are not over-fixed.
- If foaming is a problem, use a lower concentration of SDS, or increase the volume.
- Foaming is also an indicator that the sample is over-heating – keep the sample on ice and perform shorter pulses of sonication.
- Visit [www.diagenode.com/pages/bioruptor.html](http://www.diagenode.com/pages/bioruptor.html) for more details on BioRuptor, which uses lower frequency, so you will need to sonicate for longer. But you can use a wide range of volumes (0.2mL – 50mL).
- Treat samples with RNase before agarose gel analysis – otherwise you may find two populations of fragments post-sonication, only one of which is actually RNA.
- Storing chromatin post-sonication: if kept on ice, it can be stored for several days. It can keep at -70C for up to three months.
- If using nuclease digestion, you will end up with relatively short fragments (<250bp), which may affect the choice of amplification method (sometimes RP-PCR is not so efficient with short fragments).

# Sonication

---

- Key factors affecting sonication efficiency:
  - Cellular background
  - Sonicator model
  - Concentration of cells
- Aim for average fragment length of approximately 300-500 bp.
- Some researchers increase sonication (e.g. to 300 bp) helping to make peaks more defined; however, this can have a potential effect on downstream amplification choices, and overall sensitivity of the detection due to the number of fragments able to hybridize to the probe.
- Keep chromatin concentration constant between samples.
- BioRuptor seems to be the sonicator of choice for many labs doing ChIP-on-chip – waterbath sonicators have been reported to result in greater tube-to-tube consistency.
- For cells that are highly resistant to sonication – try micrococcal nuclease digestion.



## Hints and Tips...

---

- Some researchers report higher variation using magnetic beads, although other labs report good success. The advantage is no pre-clear step and easy cleanup.

# IP

---

- You can pre-couple Ab to beads or first mix with sample, then add beads. There doesn't appear to be a strong bias in favor of either.
- Protein A, G beads have been used successfully, as well as magnetic beads (Dyna). Be aware of species cross-reactivity of the protein A/G you are using. The quality of beads is very important.
- Use up to 10  $\mu\text{g}$  antibody per IP.
- You can pre-block with single-stranded herring sperm DNA or alternative.
- 1 – 1.5 hours is usually sufficient for IP, but you can leave overnight for convenience.



## Hints and Tips...

---

- If using dUTP incorporation, make sure your Taq polymerase doesn't have proofreading ability.
- dUTP tends to give more reproducible fragmentation than DNase1.
- However, some groups are successfully using DNase1 – you just need to titrate the enzyme and control the length of digestion to ensure consistency between samples.
- WGA kits are available from SIGMA and GE Healthcare.
- It is possible to incorporate dUTP into the WGA kits, for subsequent fragmentation.
- Typically you should expect 4 – 30ng DNA from a single ChIP, from which you can take an aliquot for amplification.

# Amplification

---

- Just two factors are important in amplification:
  - You need to generate sufficient DNA for hybridization to array(s).
  - You need to ensure that enrichment is maintained after amplification.
- Affymetrix has recommended Random Primed PCR because it works well for us, and is an easy and efficient way to incorporate dUTP for subsequent fragmentation.
- LM-PCR has also been used successfully by some labs, but has been reported to give more sample-to-sample variability and noisier data.
- Many labs now focusing on whole-genome amplification, report very consistent data, low bias effects, and large quantities of DNA produced, but you may need a relatively high amount of starting DNA (>25ng).



## Hints and Tips...

---

- Choice of control seems to be specific to each lab.
- Complexity of the sample may be relevant in the choice of control.
- Input DNA is reported to give more variability between control samples themselves – this may be due to complexity of input sample.
- Non-specific IP works well in many hands, but you may need to consider trying to match the expected binding frequency of non-specific and specific Ab's to match the final complexity.

# Control Options

---

- What is the “best” control to use in a ChIP experiment?
- You may need to initially run >1 control, and compare your sample to each, to see if one is better than the other.
- The ideal is to compare the same sample under two different conditions, to see the effect of the change in condition.
- Alternatively, there are three commonly used controls:
  - Input DNA
  - Mock IP – where the sample is split into two, one-half is IP’d and the other half is treated exactly the same, but without any Ab.
  - Non-specific IP – using another Ab, one that is highly unlikely to show any cross-reactivity with your target Ab. Many labs use IgG for this.
- The advantage of Affymetrix arrays is that you can run a single control per study as long as other conditions remain unchanged.



# Sample Preparation and Hybridization

---

- You will need 2 – 20 µg DNA to hybridize to your array. A good starting point would be 7 µg, but it may be worth trying a couple of different amounts.
  - The key consideration is likely to be the frequency of binding: if you are looking at a rare binding event, such as a transcription factor, you may be able to use a small quantity of DNA. If looking at medium frequency events (e.g. histone modification) or high frequency events (e.g. DNA methylation), you may need higher quantities (10 – 20 µg respectively).
- Use Affymetrix Double-stranded DNA Terminal Labeling Kit to fragment and label your DNA with biotin (P/N 900812).
- Use GeneChip® Sample Cleanup Module to purify your samples prior to hybridization (P/N 900371).
- Use GeneChip® Hybridization, Wash and Stain Kit to complete the assay (P/N 900720).

## Hints and Tips...

---

- TAS and IGB are both free to download at [www.affymetrix.com](http://www.affymetrix.com).
- You can also find a very easy-to-follow online tutorial of TAS and IGB at [www.affymetrix.com/support/learning](http://www.affymetrix.com/support/learning).
- For other commercial ChIP-on-chip data analysis solutions, go to:

[www.Partek.com](http://www.Partek.com)  
[www.Genomatix.de](http://www.Genomatix.de)

- For other freeware solutions, go to:

<http://genome.dfci.harvard.edu/xsliu/>



# Data Analysis

---

- Tiling Analysis Software (TAS) allows you to perform basic peak identification in your CEL file and link to the genome library (.bpmmap) file.
- The Integrated Genome Browser (or alternatives such as the UCSC Browser) allows you to upload annotations and visualize your data against these and across the genome.
- For more advanced data analysis and to link to biological pathway data, more sophisticated tools are becoming available from commercial vendors such as Genomatix and Partek.
- The MAT algorithm is another freeware data analysis tool, developed by X. Shirley Liu's team at the Dana-Farber Cancer Institute, which has been very powerful for ChIP-on-chip.



# Validation of Data

---

- ENCODE guidelines suggest taking 48 IP'd regions and six negative regions. These should be validated by qPCR.
- This should be seen as a starting point: you will need to get a feel for the appropriate threshold by validating higher numbers of putative peaks.
- Try generating a ROC curve using known ChIP events vs. negative regions of no IP – you will find a point at which the sensitivity vs. false discovery rate is acceptable. For example, take the top 100 candidate regions, then the top 200, top 300 etc. And verify the data from each strata.
- Using two cut-offs, one with <1% false positives, one with <5% false positives, for example.
- A minimum of three biological replicates should be run for each ChIP (or at least three technical replicates at level of antibody IP).
- Biological replicates should have  $R^2 > 0.5$  (using only significant values).

# Affymetrix Arrays and Reagents for ChIP-on-chip

Reagents for ChIP-on-chip Applications	Part Number
GeneChip WT Double Stranded DNA Terminal Labeling Kit – 30 reactions	900812
GeneChip Sample Cleanup Module – 30 reactions	900371
GeneChip Hybridization Wash and Stain Kit – 30 reactions	900720
GeneChip Control Oligo B2, 3nM – 30 reactions	900301

Whole-genome Arrays for ChIP-on-chip	Part Number	Resolution
GeneChip Human Tiling 2.0R Array Set (7 array set)*	900772	35 bp
GeneChip Mouse Tiling 2.0R Array Set (7 array set)*	900852	35 bp
GeneChip Arabidopsis Tiling 1.0R Array	900594	35 bp
GeneChip <i>S. cerevisiae</i> Tiling 1.0R Array	900645	5 bp
GeneChip <i>S. pombe</i> Tiling 1.0FR Array	900647	20 bp
GeneChip <i>C. elegans</i> Tiling 1.0R Array	900935	25 bp
GeneChip Drosophila Tiling 1.0R Array	900588	35 bp

\*Single arrays are available from these sets

# Affymetrix Whole-genome Arrays for ChIP-on-chip

Single Arrays from Human Whole-genome Set	
Array	Part Number
GeneChip Human Tiling 2.0R Array A (chromosomes 1, 6)	900779
GeneChip Human Tiling 2.0R Array B (chromosomes 2, 9, 19)	900780
GeneChip Human Tiling 2.0R Array C (chromosomes 23, 22, 21, X, Y and mitochondria)	900781
GeneChip Human Tiling 2.0R Array D (chromosomes 4, 5, 18, 20)	900782
GeneChip Human Tiling 2.0R Array E (chromosomes 5, 7, 16)	900783
GeneChip Human Tiling 2.0R Array F (chromosomes 8, 11, 12)	900784
GeneChip Human Tiling 2.0R Array G (Chromosomes 10, 13, 14, 17)	900785

Single Arrays from Mouse Whole-genome Set	
Array	Part Number
GeneChip Mouse Tiling 2.0R Array A (chromosomes 1, 9, 19)	900894
GeneChip Mouse Tiling 2.0R Array B (chromosomes 2, mitochondria, X, Y and unassigned)	900896
GeneChip Mouse Tiling 2.0R Array C (chromosomes 3, 7, 18)	900897
GeneChip Mouse Tiling 2.0R Array D (chromosomes 4, 11, 17)	900895
GeneChip Mouse Tiling 2.0R Array E (chromosomes 2, 12, 15)	900898
GeneChip Mouse Tiling 2.0R Array F (chromosomes 6, 8, 16)	900899
GeneChip Mouse Tiling 2.0R Array G (Chromosomes 10, 13, 14)	900900

## Focused Arrays

Array	Part Number	Resolution
<b>Human</b>		
GeneChip Human Promoter 1.0R Array	900775 (2)	35 bp
	900776 (6)	
	900777 (30)	
GeneChip Human Chromosome 21/22 2.0R Array	900936 (6)	16 bp
GeneChip ENCODE 2.0R Array	900937 (6)	7 bp
<b>Mouse</b>		
GeneChip Mouse Promoter 1.0R Array	900889 (2)	35 bp
	900890 (6)	
	900891(30)	

## Custom Array Options

Custom Arrays are often an attractive approach to ChIP-on-chip experiments. You can define any content, at any resolution, any strand detection and any application with a range of array formats accommodating 38,000 – 6.5 million unique oligonucleotides per array.

#### **AFFYMETRIX, INC.**

3420 Central Expressway  
Santa Clara, CA 95051 U.S.A.  
Tel: 1-888-DNA CHIP (1-888-362-2447)  
Fax: 1-408-731-5441  
sales@affymetrix.com  
support@affymetrix.com

#### **AFFYMETRIX UK Ltd**

Voyager, Mercury Park,  
Wycombe Lane, Wooburn Green,  
High Wycombe HP10 0HH  
United Kingdom  
Tel: +44 (0) 1628 552550  
Fax: +44 (0) 1628 552585  
French Tel: 0800 919505  
German Tel: 0180 3001334  
saleseurope@affymetrix.com  
supporteurope@affymetrix.com

#### **AFFYMETRIX JAPAN K.K.**

Mita NN Bldg., 16 F  
4-1-23 Shiba, Minato-ku,  
Tokyo 108-0014 Japan  
Tel: +81-(0)3-5730-8200  
Fax: +81-(0)3-5730-8201  
salesjapan@affymetrix.com  
supportjapan@affymetrix.com

[www.affymetrix.com](http://www.affymetrix.com)

**For research use only.**

**Not for use in diagnostic procedures.**



© 2007 Affymetrix, Inc. All rights reserved.  
Affymetrix, the Affymetrix logo, and GeneChip are registered  
trademarks owned or used by Affymetrix, Inc.