

## **UC Merced**

### **UC Merced Previously Published Works**

#### **Title**

Genome-Wide Chromatin Immunoprecipitation in *Candida albicans* and Other Yeasts

#### **Permalink**

<https://escholarship.org/uc/item/3xs0q6bn>

#### **Authors**

Lohse, Matthew B  
Kongsomboonvech, Pisiwat  
Madrigal, Maria  
et al.

#### **Publication Date**

2016

#### **DOI**

10.1007/978-1-4939-3079-1\_10

Peer reviewed



Published in final edited form as:

Methods Mol Biol. 2016 ; 1361: 161–184. doi:10.1007/978-1-4939-3079-1\_10.

## Genome-Wide Chromatin Immunoprecipitation in *Candida albicans* and Other Yeasts

Matthew B. Lohse, Pisiwat Kongsomboonvech, Maria Madrigal, Aaron D. Hernday, and Clarissa J. Nobile

### Abstract

Chromatin immunoprecipitation experiments are critical to investigating the interactions between DNA and a wide range of nuclear proteins within a cell or biological sample. In this chapter we outline an optimized protocol for genome-wide chromatin immunoprecipitation that has been used successfully for several distinct morphological forms of numerous yeast species, and include an optimized method for amplification of chromatin immunoprecipitated DNA samples and hybridization to a high-density oligonucleotide tiling microarray. We also provide detailed suggestions on how to analyze the complex data obtained from these experiments.

### Keywords

Chromatin immunoprecipitation; *Candida albicans*; Yeast; ChIP-chip; ChIP-seq

## 1 Introduction

Chromatin immunoprecipitation (ChIP) is a method that allows for the investigation of the interaction between a protein of interest and DNA in a cell or biological sample. In order to perform ChIP, live cells are cross-linked then lysed, the chromatin is sheared, specific DNA fragments associated with the protein of interest are immunoprecipitated from the lysate using antibodies, and the bound DNA fragments are purified away from the protein upon reversal of the cross-links (Fig. 1). The overall goal of the ChIP procedure is to determine the specific genomic binding sites of the DNA-associated protein of interest. In general, genome-wide ChIP procedures with *Candida albicans* are similar to those performed in the model yeast *Saccharomyces cerevisiae* [1]. There are, however, several *Candida*-specific modifications in terms of cell lysis and DNA shearing that we highlight in this chapter that are critical for successful *Candida* genome-wide ChIP experiments. The protocol described below has been used successfully for several distinct morphological forms of numerous yeast species, including *C. albicans*, *S. cerevisiae*, *Kluyveromyces lactis*, and *Histoplasma capsulatum* [2 – 9]. The detailed methods described in this chapter include an optimized method for amplification of ChIP DNA samples and hybridization to a high-density

<sup>17</sup>If Round B yields less than 1 µg total DNA, follow this alternate Round C approach:

Set up two 100 µL Round C reactions for each sample, using 200–400 ng of Round B DNA per tube. Perform amplification and cleanup as described in the standard Round C approach, but pool the two independent reactions prior to Zymo<sup>25</sup> column purification.

oligonucleotide tiling microarray (ChIP-chip) (*see* also ref.[10]). We also include a section on how to analyze the data obtained from genome-wide ChIP experiments. Although the protocols described here are focused on ChIP-chip, much of what we outline also applies to genome-wide ChIP-seq methods, which combine ChIP with high-resolution next-generation sequencing.

## 2 Materials

### 2.1 Chromatin Immunoprecipitation Buffers (See Note 1)

- 1 TBS: 20 mM Tris-HCl (pH 7.5), 150 mM NaCl.
- 2 Lysis buffer: 50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate.
- 3 Lysis buffer with 500 mM NaCl: 50 mM HEPES/KOH (pH 7.5), 500 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate.
- 4 Wash buffer: 10 mM Tris-HCl (pH 8.0), 250 mM LiCl, 0.5 % NP-40, 0.5 % Na-deoxycholate, 1 mM EDTA.
- 5 Elution buffer: 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1 % SDS.
- 6 TE/0.67 % SDS: 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.67 % SDS.
- 7 TE/1 % SDS: 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 % SDS.
- 8 4 M LiCl.
- 9 2.5 M glycine (prepared fresh) in ddH<sub>2</sub>O.
- 10 10 mg/mL proteinase K in TE (prepared fresh).
- 11 10 mg/mL glycogen (in TE).

### 2.2 Culture Growth and Cross-Linking

- 1 37 % formaldehyde solution (use freshly opened bottles).
- 2 2.5 M glycine (make fresh in ddH<sub>2</sub>O).
- 3 Ice-cold TBS.
- 4 Liquid nitrogen.

### 2.3 Cell Lysis and Immunoprecipitation

- 1 Ice-cold lysis buffer.
- 2 Complete protease Inhibitor cocktail EDTA-free.
- 3 0.5 mm glass beads.
- 4 Clamped horizontal shaking vortex adaptor.

---

<sup>1</sup>Use autoclaved ddH<sub>2</sub>O and baked glassware when making buffers to avoid DNA contamination. This is especially important for the final wash buffers and post-elution steps.

- 5 70 % ethanol.
- 6 18-G needles.
- 7 26-G needles.
- 8 Diagenode Bioruptor™ (preferred) or Microtip sonicator (alternative).
- 9 TE/1 % SDS.
- 10 5 µg of affinity-purified polyclonal antibody or 2–10 µg of monoclonal antibody.
- 11 50 % slurry of protein A or protein G Sepharose beads.
- 12 TBS.

#### 2.4 Recovery of Immunoprecipitated DNA

- 1 18-G needles.
- 2 Lysis buffer.
- 3 Lysis buffer with 500 mM NaCl.
- 4 Wash buffer.
- 5 TE.
- 6 Elution buffer.
- 7 TE/0.67 % SDS.

#### 2.5 Cross-Link Reversal and DNA Cleanup

- 1 Proteinase K mix: 238 µL TE, 1 µL 10 mg/mL glycogen, 10 µL 10 mg/mL proteinase K (per sample).
- 2 TE.
- 3 5 mg/mL glycogen.
- 4 10 mg/mL proteinase K.
- 5 4 M LiCl.
- 6 Phenol–chloroform–isoamyl alcohol (25:24:1), pH 8.0.
- 7 Ice-cold 100 % ethanol.
- 8 Ice-cold 70 % ethanol.
- 9 TE with 100 µg/mL RNaseA.

#### 2.6 Strand Displacement Amplification

- 1 ddH<sub>2</sub>O.
- 2 2.5× SDA buffer: 125 mM Tris–HCl (pH 7.0), 12.5 mM MgCl<sub>2</sub>, 25 mM βME, 750 µg/mL random DNA nonamers (dN9) (make fresh or store aliquots without βME at –20 °C and add βME immediately prior to use).

- 3 dNTP mix (1.25 mM each nucleotide).
- 4 50 U/ $\mu$ L *exo*-Klenow.
- 5 0.5 M EDTA.
- 6 DNA Clean and Concentrator™ Columns (Zymo Research).
- 7 DNA binding buffer (Zymo Research).
- 8 DNA wash buffer (Zymo Research).
- 9 10× aminoallyl-dNTP stock solution (12.5 mM dATP, 12.5 dCTP, 12.5 mM dGTP, 5 mM dTTP, 7.5 mM aa-dUTP).

## 2.7 Dye Coupling

- 1 ddH<sub>2</sub>O.
- 2 Fresh 1 M sodium bicarbonate, pH 9.0.
- 3 Cy3 and Cy5 monoreactive dye (Amersham).
- 4 DMSO.
- 5 DNA binding buffer (Zymo Research).
- 6 DNA wash buffer (Zymo Research).

## 2.8 ChIP -Chip Hybridization

- 1 ddH<sub>2</sub>O.
- 2 1 mg/mL Human Cot-1 DNA (Invitrogen).
- 3 10× CGH/CoC blocking agent (Agilent).
- 4 2× Hi-RPM hybridization buffer (Agilent).
- 5 Oligo aCGH/ ChIP wash buffer 1 (Agilent).
- 6 Oligo aCGH/ ChIP wash buffer 2 (Agilent).
- 7 Acetonitrile.
- 8 Drying and stabilization solution (Agilent).

## 3 Methods

### 3.1 Culture Growth and Cross-Linking

- 1 Grow 200–400 mL of planktonic cells to an OD<sub>600</sub> of 0.4 (*see* Note 2).
- 2 Add a final concentration of 1 % fresh formaldehyde (stock is at 37 %) and cross-link for 15 min at room temperature on a platform shaker.

---

<sup>2</sup>200 mL of planktonic cells at an OD<sub>600</sub> of 0.4 is sufficient for a batch of lysate which equates to about 10 individual ChIPs. For biofilm cells, use 1 × 6-well plate/strain (4 mL/well; 24 mL total volume).

- 3 Quench cross-linking with freshly made 2.5 M glycine to a final concentration of 125 mM, and incubate for 5 min at room temperature on a platform shaker.
- 4 Collect cells by centrifugation for 10 min at 1,000×g in a fixed angle centrifuge rotor.
- 5 Decant and resuspend pellets in 10 mL ice-cold TBS.
- 6 Transfer cell suspension to 15 mL Falcon tubes, pellet, decant and repeat the wash once more.
- 7 Resuspend pellet in 2 mL ice-cold TBS, and separate cell suspension to two 2 mL Sarstaedt tubes (for 400 mL cell volume) (*see* Note 3).
- 8 Pellet and decant before proceeding to lysis step, or freeze the decanted pellets in liquid nitrogen and store at −80 °C.

### 3.2 Cell Lysis and Immunoprecipitation

- 1 Thaw cell pellets on ice, and resuspend in 700 µL ice-cold lysis buffer containing complete protease inhibitor cocktail (EDTA-free) (Roche).
- 2 Transfer cell suspension to a clean 1.75 mL microfuge tube preloaded to the 500 µL mark with 0.5 mm glass beads.
- 3 Place on a clamped horizontal tube adaptor on a vortex mixer at 4 °C, and lyse cells for 30 min to 2 h (*see* Note 4).
- 4 Observe cell lysis under a microscope; if more than 90 % of the cells are lysed continue to next step (*see* Note 5).
- 5 Recover the lysate by inverting the microfuge tubes containing the lysate/bead mixture, wipe the bottom of the tube with 70 % ethanol, allow the tube to dry, and then pierce the bottom of the tube with a 26-G needle. Open the microfuge tube, place it into a 5 mL falcon tube (right side up), and pierce the falcon tube (above the level of the bottom of the microfuge tube) using an 18-G needle attached to a vacuum line. This will cause the lysate to flow through to the bottom of the falcon tube (*see* Note 6). Recover the lysate, and transfer 300 µL to each of two new 1.75 mL microfuge tubes for Bioruptor shearing (*see* Note 7).
- 6 Shear chromatin by sonication in a Diagenode Bioruptor™ with the following settings: 15 min, high setting, 30 s on, 1 min off (*see* Note 8).

<sup>3</sup>For 200 mL planktonic cell volume and for 24 mL biofilm cell volume, resuspend in 1 mL ice-cold TBS and store in one 2 mL Sarstaedt tube.

<sup>4</sup>Lysis times may vary depending upon the cell type and growth conditions.

<sup>5</sup>Cells should appear as a mixture of dead cell “ghosts” and fragmented cell debris by phase contrast microscopy.

<sup>6</sup>As an alternate method, recover the lysate by centrifugation into a larger tube.

<sup>7</sup>Alternatively, if using a microtip sonicator for shearing, transfer the entire lysate to one new 1.75 mL microfuge tube.

<sup>8</sup>If using a microtip sonicator, use the following settings: 5 × 20 s at level 2, 100 % duty cycle, with 1 min on ice between each pulse. In our experience, using the Bioruptor to shear results in yields of smaller fragment sizes, tighter shear distribution, and greater consistency than the microtip sonication method.

- 7 Pellet the cell debris for 5 min at full speed at 4 °C and transfer the supernatant containing the lysate to a new tube.
- 8 Remove 50 µL of the lysate and transfer to a new tube containing 200 µL TE/1 % SDS. This is the “input DNA” sample; store at –20 °C until it will be processed along with the immunoprecipitated DNA.
- 9 Aliquot and dilute the sheared lysate according to the number of IPs that will be performed. Use 50–500 µL of crude lysate in 500 µL (final volume) lysis buffer (with fresh protease inhibitors) for each IP.
- 10 Add appropriate antibody (e.g., 5 µg of custom affinity-purified polyclonal antibody, 2 µg of monoclonal anti-c-myc antibody, or 10 µg of anti-FLAG M2 monoclonal antibody), and incubate at 4 °C overnight on a nutator agitator.
- 11 The following day, wash bead slurry as follows. For 20 reactions, you will need ~1 mL of a 50 % slurry of appropriate type of Sepharose beads (protein A or protein G; ~500 µL bed volume). Use wide-bore P1000 tips for bead dispensing. Spin 4000×g, 10 s (double bed volume for washes since you lose ~30–40 % beads during washes). Wash two times with TBS and three times with lysis buffer. Resuspend in lysis buffer to 50 % slurry.
- 12 Add 50 µL of washed 50 % slurry of protein A or protein G Sepharose beads to each immunoprecipitation sample, using wide-bore P200 tips to dispense the bead slurry. Incubate at 4 °C for 2 h on a nutator agitator.

### 3.3 Recovery of Immunoprecipitated DNA (See Note 9)

- 1 Wash beads as follows (*see* Note 10): Pellet for 1 min at 1000×g at room temperature.
- 2 Draw off the supernatant with an 18-G needle on a vacuum line.
- 3 Wash with buffers indicated below for 5 min each while mixing on a nutator:
  - Two times with 1 mL lysis buffer.
  - Two times with 1 mL lysis buffer with 500 mM final NaCl.
  - Two times with 1 mL wash buffer.
  - One time with 1 mL TE.
- 4 After final wash, draw off TE and add 110 µL elution buffer, vortex on gentle setting, and incubate for 10 min at 65 °C, mixing every 2 min by gentle vortex.
- 5 Pellet for 30 s at full speed at room temperature, and remove 100 µL of supernatant to a new tube.

<sup>9</sup>For optimal recovery of immunoprecipitated DNA during the washes, keep samples at room temperature and buffers ice cold.

<sup>10</sup>Wash buffer temperatures, incubation temperatures, and incubation times can be optimized for each antibody; however, we have found that ice-cold buffers and 5 min incubations at room temperature work best for most antibodies.

- 6 Add 150  $\mu$ L of TE with 0.67 % SDS to remaining bead bed and vortex vigorously. Pellet by spinning full speed, 1 min, remove 150  $\mu$ L of supernatant, and pool together with previous eluate (250  $\mu$ L total).

### 3.4 Cross-Link Reversal and DNA Cleanup

- 1 Incubate the IP samples and “input DNA” samples (from previous steps) at 65  $^{\circ}$ C for 16 h.
- 2 Add 250  $\mu$ L of proteinase K mix (for each sample: 238  $\mu$ L TE, 1  $\mu$ L 10 mg/mL glycogen, 10  $\mu$ L 10 mg/mL proteinase K). Incubate at 37  $^{\circ}$ C for 2 h (*see Note 11*).
- 3 Add 55  $\mu$ L of 4 M LiCl and 555  $\mu$ L of cold phenol–chloroform– isoamyl alcohol (25:24:1), pH 8.0. Mix by vortexing briefly, spin at 10,000 $\times$ g for 2 min, and remove 500  $\mu$ L of the top aqueous layer to a fresh tube.
- 4 Add 1 mL of ice-cold 100 % ethanol to the collected aqueous layer and incubate at  $-20^{\circ}$  C overnight or at  $-80^{\circ}$  C for at least 1 h.
- 5 Centrifuge at 10,000 $\times$ g at 4  $^{\circ}$ C for 30 min. Decant carefully using a 1 mL pipette.
- 6 Wash the pellet with 950  $\mu$ L ice-cold 70 % ethanol (wash by gentle inversion), spin for 10 min at 10,000 $\times$ g at 4  $^{\circ}$ C, decant, spin briefly, and remove any residual ethanol.
- 7 Air-dry the pellets and resuspend as follows: use 25  $\mu$ L of TE for the IP samples and 100  $\mu$ L of TE with 100  $\mu$ g/mL RNaseA for the “input DNA” samples (*see Note 12*).
- 8 Incubate the input samples at 37  $^{\circ}$ C for 1 h to degrade RNA, and the IPs at 37  $^{\circ}$ C for 10 min to ensure pellet is dissolved, and store all at  $-20^{\circ}$  C (*see Note 13*).

### 3.5 Strand Displacement Amplification of Chromatin IP Samples (See Note 14)

The following amplification protocol uses high concentration exo-Klenow with random DNA nonamers (dN9s) in order to perform strand displacement amplifications of the IP and input DNA samples for ChIP experiments. Prior to amplification, input and IP DNA concentrations are normalized by dilution of the input DNA for each corresponding IP based on qPCR values for a nonenriched locus (e.g., *ADE2* locus). Input and IP samples are amplified separately, in parallel, and should yield similar amounts of product after each

<sup>11</sup>Make a fresh proteinase K solution from lyophilized powder.

<sup>12</sup>RNase A stock is prepared at 10 mg/mL in ddH<sub>2</sub>O. The RNase A stock solution should be boiled for 10 min before adding to TE to remove DNase activity.

<sup>13</sup>An optional DNA clean up step could be performed on the “input DNA” using a commercial DNA clean up kit, however this adds an additional variable (relative to the IP DNA), and could produce spikes in the ChIP-chip data. Therefore, it is recommended to leave the RNase A in the “input DNA” sample and avoid the cleanup steps prior to amplification. We also recommend monitoring the sheer distribution of the input DNA sample prior to proceeding with subsequent analysis of ChIP samples. Test the sheer distribution by running ~200–500 ng of purified input DNA on a 2 % agarose gel at ~5 V/cm. Average sheer size from the Bioruptor is typically ~200 bp, with most fragments distributed between 100 and 400 bp.

<sup>14</sup>This is a nonspecific amplification, and any contaminating DNA will be amplified. Therefore, perform all amplification rounds with gloves, filter tips, autoclaved ddH<sub>2</sub>O, and reagents free of any potential DNA contamination.



round of amplification (typically three rounds). Round B amplification can be omitted if the IP DNA concentration is sufficient.

### 3.5.1 Round A Primary Amplification

- 1 Mix 12  $\mu\text{L}$  of IP sample or 12  $\mu\text{L}$  of diluted input (diluted in TE) (*see* Note 15), 12  $\mu\text{L}$  of ddH<sub>2</sub>O, and 20  $\mu\text{L}$  of 2.5 $\times$  SDA buffer.
- 2 Incubate at 95  $^{\circ}\text{C}$  for 5 min and then immediately transfer samples to an ice water bath for 5 min.
- 3 Add 5  $\mu\text{L}$  of dNTP mix (1.25 mM each nucleotide).
- 4 Add 1  $\mu\text{L}$  of 50 U/ $\mu\text{L}$  *exo*-Klenow, and mix by pipetting.
- 5 Incubate at 37  $^{\circ}\text{C}$  for 2 h with a heated lid thermal cycler (*see* Note 16).
- 6 Purify product using Zymo<sup>25</sup> columns by adding 10 volumes of binding buffer (450–50  $\mu\text{L}$  sample), bind to column, wash 1 $\times$  with 200  $\mu\text{L}$  binding buffer, 2 $\times$  with 200  $\mu\text{L}$  wash buffer, spin for 1 min at 10,000 $\times g$  to dry, and elute with 30  $\mu\text{L}$  H<sub>2</sub>O into a new tube.
- 7 Check 1.5  $\mu\text{L}$  of the sample on a NanoDrop spectrophotometer. If the total yield is  $\geq$  400 ng, then skip to Round C. If not, continue to Round B.

### 3.5.2 Round B Secondary Amplification

- 1 Mix 24  $\mu\text{L}$  of Round A DNA, 20  $\mu\text{L}$  of 2.5 $\times$  SDA buffer.
- 2 Repeat **steps 2–7** of Round A Primary Amplification, but elute with 50  $\mu\text{L}$  of H<sub>2</sub>O after purifying product.

### 3.5 Strand Displacement Amplification of Chromatin IP Samples (See Note 14)

For this final amplification, perform 100  $\mu\text{L}$  reactions with 1–2  $\mu\text{g}$  total Round B DNA for each sample. This will yield ~2.5- to 3-fold amplification.

- 1 Mix 1–2  $\mu\text{g}$  of Round B DNA in H<sub>2</sub>O to 48  $\mu\text{L}$  total volume, then add 40  $\mu\text{L}$  of 2.5 $\times$  SDA.
- 2 Incubate for 5 min at 95  $^{\circ}\text{C}$  and immediately transfer the samples to an ice water bath for 5 min.
- 3 Add 10  $\mu\text{L}$  of 1.25 mM aminoallyl-dNTP mix (1:10 dilution of stock solution).
- 4 Add 2  $\mu\text{L}$  of 50 U/ $\mu\text{L}$  *exo*-Klenow.
- 5 Incubate at 37  $^{\circ}\text{C}$  for 2 h with heated lid in a thermal cycler (*see* Note 16).
- 6 Purify the Round C product using Zymo<sup>25</sup> columns by adding 10 volumes of binding buffer (900–100  $\mu\text{L}$  sample), bind to column, wash 1 $\times$  with 200  $\mu\text{L}$

<sup>15</sup>Equalize the input and IP samples based on the qPCR values for a non-enriched locus (e.g., the *ADE2* locus).

<sup>16</sup>The reactions may sit up to ~2 h at 10  $^{\circ}\text{C}$  following amplification or add 5  $\mu\text{L}$  0.5 M EDTA at pH 8.0 and store at  $-20^{\circ}\text{C}$ .

binding buffer, 2× with 200 µL wash buffer, spin for 1 min at 10,000×g to dry, and elute with 50 µL of H<sub>2</sub>O into a new tube.

- 7 Check 1.5 µL of the sample on a NanoDrop; the yield should be about 5 µg of total DNA per reaction.

### 3.6 Dye Coupling

- 1 Speed-vac the amplified input and IP reactions from Round C to 9 µL volume, or until dry.
- 2 Resuspend with H<sub>2</sub>O to 9 µL final volume and add 1 µL of fresh 1 M sodium bicarbonate, pH of 9.0 (*see* Note 18).
- 3 Add 1.25 µL of Cy3 for the input sample or Cy5 for the IP sample (*see* Note 19).
- 4 Incubate the labeling reactions at room temperature for 1 h in the dark.
- 5 Purify the dye-coupled DNA with Zymo<sup>25</sup> columns by adding 800 µL of Zymo DNA binding buffer to each of the samples and load onto a Zymo column. Wash once with 200 µL binding buffer, wash twice with 200 µL wash buffer, and spin at 10,000×g for 1 min to dry. Elute using 50 µL of H<sub>2</sub>O into a new tube. Check concentration on a NanoDrop spectrophotometer using the “microarray” setting to quantitate the total yield and dye-coupling efficiency. A minimum of 2–4 pmol/µL Cy3 or Cy5 is ideal.
- 6 Equalize the input and IP samples to 5 µg each in order to hybridize to a 1× 244 K format Agilent microarray (*see* Note 20).

### 3.7 ChIP-Chip Hybridization (Adapted from the Agilent Oligo aCGH/ChIP-On-Chip Hybridization Kit)

- 1 Mix 5 µg each of the input and IP samples, and bring volume to 150 µL in ddH<sub>2</sub>O (*see* Note 21).
- 2 Add 50 µL of 1 mg/mL Human Cot-1 DNA.
- 3 Add 50 µL of 10× CGH/CoC blocking agent.
- 4 Add 250 µL of 2× Hi-RPM hybridization buffer.
- 5 Mix and quick spin to collect sample.
- 6 Incubate at 95 °C for 3 min and transfer to 37 °C for 30 min.
- 7 Spin at full speed in a microcentrifuge for 1 min, carefully remove 490 µL, and load the sample onto a gasket slide. Cover with the array slide and assemble the hybridization chamber.

<sup>18</sup>Prepare the sodium bicarbonate fresh in ddH<sub>2</sub>O on the day of labeling and pH using a pH meter.

<sup>19</sup>We use Amersham monoreactive dye packs (Cat. # PA 23001 and PA25001). Each tube contains enough dye for eight labeling reactions. Resuspend the dye in 10 µL DMSO and use 1.25 µL of dye per labeling reaction. If fewer than eight labeling reactions are to be performed, any unused dye can be desiccated and stored at 4 °C in the dark.

<sup>20</sup>Agilent custom oligonucleotide arrays, hybridization buffers, and wash buffers consistently yield high-quality data.

<sup>21</sup>Although 5 µg is optimal, a minimum of 1 µg each of input and IP samples is sufficient for hybridization without any significant decrease in data quality. Be sure that equal amounts of the input and IP sample are used for hybridization.

- See demo video at: <http://agilent.cnpg.com/video/flatfiles/189/>
- 8 Hybridize at 65 °C for 40 h in an Agilent microarray hybridization oven at a rotation speed of 20 rpm.
- 9 Disassemble the array and wash using Agilent wash buffers with mixing (using a magnetic stir bar) as follows (*see* Note 22):
  - Agilent oligo aCGH/C hIP-on-Chip wash buffer 1 for 5 min at 25 °C.
  - Agilent oligo aCGH/C hIP-on-Chip wash buffer 2 for 5 min at 32 °C.
  - Acetonitrile for 1 min at 25 °C.
  - Agilent drying and stabilization solution (*see* Note 23) for 30 s at 25 °C (*see* Note 24).
- 10 Scan slides in GenePix 4000B scanner, and grid array.

### 3.8 Preliminary Data Analysis (See Note 25)

The data analysis sections described below provide an overview of many of the analyses we have commonly performed for sequence-specific DNA binding proteins. Not all of the analyses described are relevant to every situation and our analyses may not include analyses relevant to your situation. These analyses may also be performed with data generated from ChIP-seq datasets.

- 1 For all of the downstream ChIP-chip data analysis described below, we use MochiView [11], however there are alternative programs available. MochiView is freely available at <http://johnsonlab.ucsf.edu/mochi.html> (*see* Note 26). An example plot of ChIP-chip data visualized using MochiView is shown in Fig. 2.
- 2 Normalize the enrichment values for every probe on the array by LOWESS normalization using Agilent Chip Analytics software or other normalization software, and import the data into MochiView.
- 3 Raw ChIP-chip data should be uploaded to a publicly available database, such as the Gene Expression Omnibus ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)), and an accession number should be obtained (*see* Note 27).

### 3.9 Peak Calling

- 1 Identify binding events by smoothing the data using MochiView. This utility applies a smoothing function to the Chip Analytics log<sub>2</sub> enrichment values,

<sup>22</sup>For disassembly, hold the microarray/gasket slide submerged in wash buffer 1 while gently gripping sides of the microarray slide. Gently pry the gasket slide off of the array by inserting the tip of a plastic forceps between the outer edge of the two slides and lightly twist the forceps. The gasket slide will fall away, while the array should remain in your hands. Be sure to avoid any contact with the printed array surface.

<sup>23</sup>If Agilent drying and stabilization solution contains precipitate, place bottle in a 37 °C water bath prior to use. Depending on the amount of precipitate, it may be necessary to dissolve the precipitate overnight in the water bath; cool to RT before use.

<sup>24</sup>In order to ensure even drying of the array, remove the slide holder from the Agilent drying and stabilization slowly; minimize water droplets on the arrays.

<sup>25</sup>At least two independent biological replicates for each strain should be used for data analysis.

<sup>26</sup>Some alternatives to MochiView include UCSC's Genome Browser, freely available at <http://genome.ucsc.edu/> [17] and CisGenome, freely available at <http://www.biostat.jhsph.edu/~hji/cisgenome/> [18].

followed by the application of a peak detection algorithm, where all binding peaks are assigned a  $P$  value using permutation testing.

- 2 Use peak-finding significance thresholds at the default settings,  $P$  0.001 for the experimental IPs, and  $P$  0.05 for the control IPs (*see* Note 28).
- 3 User-defined cut-offs for the minimum value for peak inclusion post-smoothing should be determined using the distribution of log-ratios for each experimental IP, and should be set at two standard deviations from the mean of  $\log_2$ -transformed fold enrichments. User-defined cut-offs for the minimum value for peak inclusion post-smoothing ranging from 0.27 to 0.36 (1.5 standard deviations from the mean of  $\log_2$ -transformed fold enrichments) should be used for the untagged or delete IP control data sets (*see* Note 29).

### 3.10 Peak Curation and Assignment (See Note 30)

- 1 Eliminate any peaks that overlap with a “red-flagged” location set (*see* Note 27).
- 2 Map the cleaned list of peaks to intergenic regions and to specific genes using defined criteria, such as those described in the following steps (*see* Note 31).
- 3 If a peak falls entirely over an intergenic region map it to that intergenic region.

<sup>27</sup>In addition to submitting datasets to a curated database, several additional steps will help readers to make use of your data. Downloading and interpreting entire datasets is a barrier to many readers, so we suggest including certain supplemental files with any manuscript containing genome-wide ChIP data. If you have a list of “red flagged” locations used to remove spurious peaks, a sheet with said locations should be provided. To allow for quick examination of the data, we include an Excel file with a list of genes and the maximum enrichment value for any peak upstream of that gene. Ideally, this list should include all genes but at a minimum it should contain all genes with upstream binding sites. There should be columns for each genome-wide ChIP dataset as well as other datasets being used. *See* supplementary files in Hernday et al. [9] and Nobile et al. [2] for examples of such files. Also important is an Excel or text file with peak locations, enrichment values, and the adjacent genes for each peak can also be of use. *See* supplementary files from Cain et al. [4], Nobile et al. [2], and Hernday et al. [9] for examples. In addition, a file with plots centered on each peak overlaid with the experimental and control data traces are useful. We generate these plots in MochiView, *see* supplementary files from Hernday et al. [9] and Nobile et al. [2] for examples. Include also tables for any higher order analysis such as binding site overlap comparisons. If any DNA binding motifs are reported, include an Excel or text file with the data used to make the Position Specific Weight Matrix as such information cannot readily be determined based on figures alone. *See* supplementary files from Hernday et al. [9] for an example of such a file. Describe the predictive value of all motifs, ideally by the relevant ROC plots. *See* Supplemental Fig. 1 from Cain et al. [4] for an example of such a figure. The methods section of the paper reporting the ChIP data should include information for any antibodies used, such as sequences polyclonal antibodies were raised against or suppliers of commercially available monoclonal antibodies. Methods sections should also explain the criteria used for peak calling as well as the details of any analyses performed.

<sup>28</sup>Peak-finding significance thresholds may be adjusted accordingly to assess the quality of the data. For greater confidence, the amount of sampling can be increased tenfold from the default setting to 100,000 (number of random samples to compare against each peak), and 100 (maximum number of random samples passing for inclusion of peak).

<sup>29</sup>Adjustments to peak inclusion cutoffs should not alter the majority of called peaks in a good dataset.

<sup>30</sup>The peak curation steps described stem in part from our observation of a number of likely artifactual peaks near tRNA and over ribosomal genes, based on the fact that these loci showed variable but substantial enrichment in the majority of deletion control ChIP-chip experiments that were performed with antibodies against a deleted target [9, 19]. Recent reports have indicated that ChIP-chip data often results in spurious areas of enrichment at or near highly transcribed regions, such as tRNA [20]. Such spurious peaks often escape culling during the peak calling process depending on the nature of a given control; as such we take the precaution of automatically eliminating anything overlapping our “red flagged” dataset. That said, entirely different sets of assumptions may apply with other types of proteins, such as histone deacetylases, where certain categories of peaks, including peaks covering ORFs, may be physiologically relevant [21]. Peaks near tRNA or over ribosomal genes are not spurious a priori, so it may not be possible to eliminate them unless they have appeared in multiple control datasets. Care should be taken when dealing with such peaks and in making any conclusions based on their presence or absence.

<sup>31</sup>We perform the peak curation and assignment analysis using MochiView but other software packages could be used. We use the criteria described in the methods section for assigning peaks to intergenic regions and genes, but other criteria may be more appropriate in a given situation. Normally we ignore small features such as tRNAs or spurious ORFs when making gene proximity assignments. Depending on the specific circumstances, different assumptions may be appropriate. Whatever criteria are chosen, it is important to apply them consistently across the dataset and to list them when reporting the dataset.

- 4 If a peak partially overlaps with both an intergenic region and an open reading frame (ORF), assign it to that intergenic region.
- 5 If a peak is positioned over a short open reading frame such that it also overlaps with the intergenic regions on each side of the ORF, assign it to both flanking intergenic regions.
- 6 If a peak is positioned entirely over an open reading frame such that it does not overlap any intergenic region, omit it from further analysis.
- 7 If multiple peaks fall within a single intergenic region, assign the maximum enrichment value of the individual peaks to that intergenic region (and any associated genes) for the purpose of further analysis.
- 8 Consider any gene whose start codon is located immediately downstream of a given intergenic region (that is whose start codon is on the same side of the ORF as a given intergenic region rather than on the opposite end of the ORF) as potentially regulated if that intergenic region is bound by the factor in question. As such, a given intergenic region may be associated with two (in the case of divergent ORFs), one (in the typical case), or zero (in the case of convergent ORFs) genes.
- 9 In the case of divergent ORFs (two genes controlled by a given intergenic region), a distinction is not made between the regulation of the two genes based on the distance from the binding site/s. Both genes are considered to be potential targets of the binding site.

### 3.11 Binding Site and Regulon Analysis

- 1 Once lists of the intergenic regions and genes that are potential targets for the regulator of interest are obtained, begin the following analysis (*see* Note 32).
- 2 If transcription data is available for deletion or overexpression of the regulator in question or for conditions related to the regulator, determine if the target genes are preferentially regulated in a regulator-dependent manner relative to the genome as a whole.
- 3 Perform Gene Ontology (GO term) mapping or other similar analysis on the target set of genes in order to determine if there are enriched classes of genes in the regulon.
- 4 Determine whether the binding events occur at a constant distance from a common feature, such as a start codon, transcription start site, or other known regulator target sites.
- 5 Use multiple motif search algorithms, such as SCOPE [12], Bioprosppector [13], and MEME [14], to identify DNA sequences whose occurrence correlates with the binding sites (*see* Note 33).

---

<sup>32</sup>In general, the results and significance of a given analysis will improve as the number of locations (or genes) being used increases and the average size of the locations being used decreases. Although searches can be performed using entire intergenic regions, it is preferable to use discrete peaks and to have these peaks be as narrowly defined as possible.

- 6 Having identified a list of potential DNA-binding motifs, perform the following analyses to evaluate their predictive power (*see* Note 34).
- 7 Create a control location set randomly selected from unbound intergenic regions. The average size of members of this control location set should be the same as the average size of the experimental peaks. If whole intergenic regions are used, the control set should contain unbound intergenic regions whose average size is equivalent to the experimental set (*see* Note 35).
- 8 Once the control set has been created, determine how many sites in the experimental and control location sets have one or more instances of the motif for a series of stringency criteria. These can then be plotted on a graph, such as an ROC (Receiver Operating Characteristic) plot, with true positives (percent of experimental locations passing) and false positives (percent of control locations passing) on the two axes.
- 9 Search for DNA-binding motifs using only a fraction of the binding sites (i.e., the top 50 % of peaks or a randomly selected subset of the peaks) and subject the resulting motifs to **steps 7 and 8** of Subheading 3.11 (*see* Note 36).
- 10 Once you have identified one or more motifs with good predictive value, ideally for full as well as partial datasets, proceed to the following analyses.
- 11 If the factor in question has a published DNA-binding motif or one or more homologs with a published motif, compare your motif to the existing motifs. Comparisons between two or more motifs can be performed using programs like MochiView and motifs can be compared against libraries of motifs using online databases, such as YeastTract (<http://www.yeasttract.com>) [15], MOTIF (<http://www.genome.jp/tools/motif/>), and TOMTOM (<http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi>) [16]. Using these same tools, determine whether part or your entire motif matches that of a known regulator (*see* Note 37).

---

<sup>33</sup>When dealing with online servers for DNA-binding motif identification, it is important to remember that the different servers have their own strengths and weaknesses (for review, *see* [22]) and that the server used as well as the settings chosen for a given server may limit the type of motifs that can be found. The use of multiple tools or servers as well as different settings within a server may provide a better motif, while also allowing for the identification of a broader range of motif types.

<sup>34</sup>It is important to remember that the top motif hit (or even top several hits) may not necessarily be the best at explaining the location set. Rather, the output of these tools should be seen as a series of candidates that need to be further examined in detail. Likewise, the statistical value (i.e., *P* or *E* value) for a given putative motif does not in and of itself indicate that the motif is valid. Depending on the overall makeup of the genome (i.e., AT or GC rich), it is possible to get spurious but statistically significant motifs that consist of repeats of a given nucleotide or a repeated sequence of 2–3 nucleotides that occur throughout the genome. For example, spurious motif hits that are rich in “A” and “T” nucleotides are not uncommon given that promoters generally tend to be “AT” rich. Thus, as a general rule, the motifs produced should be considered as a set of candidates for further analyses.

<sup>35</sup>There are several important considerations to take into account in order to avoid unintentionally skewing of the results or potentially misinterpreting data. For almost every analysis or calculation, it is important to consider how the specific intergenic length in question compares to the average intergenic length of the genome as a whole. It is critical to make sure that any control location sets are equivalently sized to the relevant experimental set as most analyses will tend to favor the longer location set if the two sets are not equivalently sized.

<sup>36</sup>Using only the full location set for motif analysis contains a level of bias, even when coupled with the follow-up tests we describe, as it is evaluating a motif against the dataset used to create the motif. To reduce this bias, we recommend performing parallel analyses on motifs developed using only portions of the binding sets (i.e., the full target list, the top half of the target list based on enrichment values, and the top third or quarter of the list based on enrichment values). An example of the multiple rounds of motif analysis described in the methods section can be found in Cain et al. [4]. We consider this analysis to be critical in the motif identification process given the frequency with which our searches returned spurious motifs as the best hit for a given location set.

- 12 Determine whether the motif/s occurs at a constant distance from a common feature, such as a start codon, a transcription start site, or the motif for another transcription factor (*see* Note 38).
- 13 Assess the individual occurrences of your motif at your binding sites. Determine whether your motif/s can be subdivided into distinct sub-motifs or whether any pairs of your motifs consistently occur with fixed spacing between them.

### 3.12 Analysis of Multiple Datasets (See Note 39)

- 1 If possible, produce equivalently sized and formatted location sets for each of the datasets to be compared (*see* Note 40). There are three general approaches for comparing the overlap of binding datasets.
- 2 Compare the lists of genes downstream of the binding sites for each factor (*see* Note 41).
- 3 Compare the overlap of the intergenic regions bound by the various factors rather than the genes (*see* Note 42).
- 4 Compare the overlap of the binding peaks (*see* Note 43).
- 5 For whichever comparison/s is made, subdivide the overlap into all the various possible combinations (i.e., “A and B” and “A and B and C”), as well as broader

<sup>37</sup>Although changes in preferred DNA-binding sites for homologs of a given regulator have been reported [23], it is more common that they remain similar or even identical [3]. A match between your motif and a previously reported one for a homolog of your target is another piece of evidence in support of the motif you have developed. If there is a large difference between your motif and a previously reported motif, it may be useful to verify the new motif through *in vivo* or *in vitro* experiments.

<sup>38</sup>If your target has one or more cofactors, motif searches may return the motifs for the cofactors in addition to or in lieu of your target's motif. Depending on the motif search settings and the spacing of the DNA-binding sequences, these may occur together in one longer motif or as two independent motifs. As such, you should look to see whether your motif/s can be subdivided into multiple motifs (one of which might match a previously known motif) or whether any pair of your motifs occurs with a constant spacing between them. If the analysis suggests that your target does have a cofactor, it can be informative to rerun motif finding analysis on the subsets of sites with or without the cofactor in order to improve the relevant motifs in each case and to determine if the DNA-binding preferences of your target differ between the two cases [9, 24].

<sup>39</sup>Depending on what datasets are available in your case, a wide range of analyses can potentially be performed. When dealing with more complex regulons, you may have produced or have access to previously reported binding data for a number of other factors. If such datasets exist, it can be informative to examine the degree of overlap between the different sets of binding sites. Although we have traditionally performed this analysis in MochiView, it can be performed with a wide variety of available software.

<sup>40</sup>As with motif identification, it is important to avoid unintentionally skewing of the results or potentially misinterpreting data. As a rule of thumb, more importance should be attached to the presence of a peak rather than the absence of a peak. For almost every analysis or calculation, it is important to consider how the specific intergenic length in question compares to the average intergenic length of the genome as a whole. It is critical to make sure that any control location sets are equivalently sized to the relevant experimental set as most analyses will tend to favor the longer location set if the two sets are not equivalently sized. It is also often fruitful to perform each analysis in parallel with different target lists (i.e., the full target list, the top half of the target list based on enrichment values, and the top third or quarter of the list based on enrichment values). Likewise, it may be worthwhile to evaluate whether various metrics or analyses are linked in some way to binding enrichment values.

<sup>41</sup>The target gene based overlap analysis, which can be quickly conducted in Microsoft Excel, will provide a basic understanding of the overlap (if any) between the factors. Although quick, it is of limited use compared to some of the other analyses we describe.

<sup>42</sup>The intergenic regions overlap analysis is conceptually similar to the gene based analysis, but remains useful for many of the follow-up analyses. The most involved version of this analysis looks for overlap in binding peaks themselves. Since intergenic regions will sometimes contain multiple binding peaks for a given factor and since different factors will often be positioned slightly askew to each other, this analysis requires careful curation of the various types of binding events. Some events may require manual judgment calls as overlapping patterns may not necessarily fit into a simple category.

<sup>43</sup>The overlap between binding sites or intergenic regions can also be examined using a visualization program such as Cytoscape (<http://www.cytoscape.org/>) [25]. Regardless of the approach taken, when dealing with high degrees of overlap between two or more factors in ChIP datasets, it is important to remember that the absence of evidence for a peak for one factor at a location should not be taken as evidence of certain absence as there are a number of experimental reasons, such as epitope masking, as to why binding could be occluded or missed.

categories (i.e., “bound by  $n$  or more factors”). Perform the following analysis on the different subgroups (*see* Note 44).

- 6 Perform Gene Ontology (GO) term analysis for the genes regulated by specific categories of regulator binding events.
- 7 If transcriptional profiling is available for deletions or overexpressions of various regulators or for conditions related to the regulators, determine if the genes regulated by specific categories of regulators are preferentially regulated (*see* Note 45).
- 8 If available, determine whether the DNA-binding motifs for the factors in question are better at predicting certain types of binding events.
- 9 Determine whether certain types of binding events correspond with higher or lower levels of binding enrichment for the factors in question.
- 10 Determine whether the overlap of binding events for any combination of factors occurs more frequently than would be expected by chance. There are several possible metrics (described below) for making such a calculation and we recommend using all of them to account for possible biases. These metrics all work on a similar principle, comparing the number of observed binding events to the number of predicted events for a given combination (*see* Note 46).
- 11 Make a comparison based on the fraction of intergenic regions bound by each regulator, treating all intergenic regions as equally probable targets for regulatory binding. For each regulator determine the fraction of possible intergenic regions bound; divide the number of intergenic regions the regulator binds by the total number of intergenic regions in the genome. Proceed to **step 14** of Subheading 3.12.
- 12 Make a comparison based on the fraction of intergenic regions bound by each regulator, correcting for the difference in intergenic length. To do this, determine a length correction factor for each regulator by dividing the mean length of the intergenic regions bound by each regulator by the mean length of all intergenic regions in the genome. Multiply the fraction of possible intergenic regions bound by this length correction factor, before proceeding with the calculations in step 14 of Subheading 3.12.

---

<sup>44</sup>There are a variety of approaches to take in regard to this analysis, each with their own benefits and drawbacks. In all cases, it is worth remembering that the number of possible combinations for “ $n$ ” factors equals  $(2n) - 1$  so the effort involved will greatly increase as more factors are included.

<sup>45</sup>When looking for overrepresentation or underrepresentation of some condition in regard to an experimental set relative to the genome as a whole, the hypergeometric distribution test is often the most relevant statistical test. In such cases, the chi-squared test may also be valid if there are a large number of targets. It is worth remembering that a lack of regulation of targets for a given regulator in a specific condition does not necessarily mean that binding events have no function. The Gal4 binding sites upstream of the galactose metabolic genes in *S. cerevisiae* are a good example of this; Gal4 binds under almost all conditions, but is only functional in response to a specific metabolic cue ([26], for review *see* ref. [27]). At the same time, remember that not all binding sites are functional.

<sup>46</sup>We have discussed three possible metrics for determining whether binding site overlap occurs more frequently than would be expected by chance. We recommend using all three of them to account for possible biases in each method. These are described in detail in the Supplemental Methods from Hernday et al. [9].



- 13** Make a comparison that accounts for the fact that short intergenic regions provide a relatively smaller evolutionary target, and thus are less likely by chance than longer intergenic regions to acquire active binding motifs. For each intergenic region, calculate the region's weighted length as the product of the region's length and the inverse of the mean length of all intergenic regions. For each regulator, calculate a length-corrected fraction of bound intergenic regions by taking the sum of weighted lengths for all intergenic regions bound by that regulator and dividing by the sum of weighted lengths for all intergenic regions in the genome. Use this corrected value for the calculations in step 14 of Subheading 3.12.
- 14** Determine the fraction of regulators predicted to be bound by any given combination of regulators by multiplying the fraction of possible intergenic regions for the relevant regulators. Substitute in the relevant length corrected factors from **step 12** or **13** of Subheading 3.12 if performing either of those analyses. Multiply this predicted fraction of binding sites with a particular combination by the total number of intergenic regions in the genome in order to get the predicted number of regulators bound by a given combination of regulators. Compare the observed number of binding events to the predicted number of events.
- 15** If data is available for the binding of a homolog of your target/s in another species, compare the targets of a given factor/s between those species. Mapping gene orthologs between species is beyond the scope of this chapter, however we recommend mapping orthologs in both directions and using multiple orthology lists if possible (*see* Note 47).

## Acknowledgments

This work was supported by National Institutes of Health (NIH) grants R00AI100896 and R01AI049187.

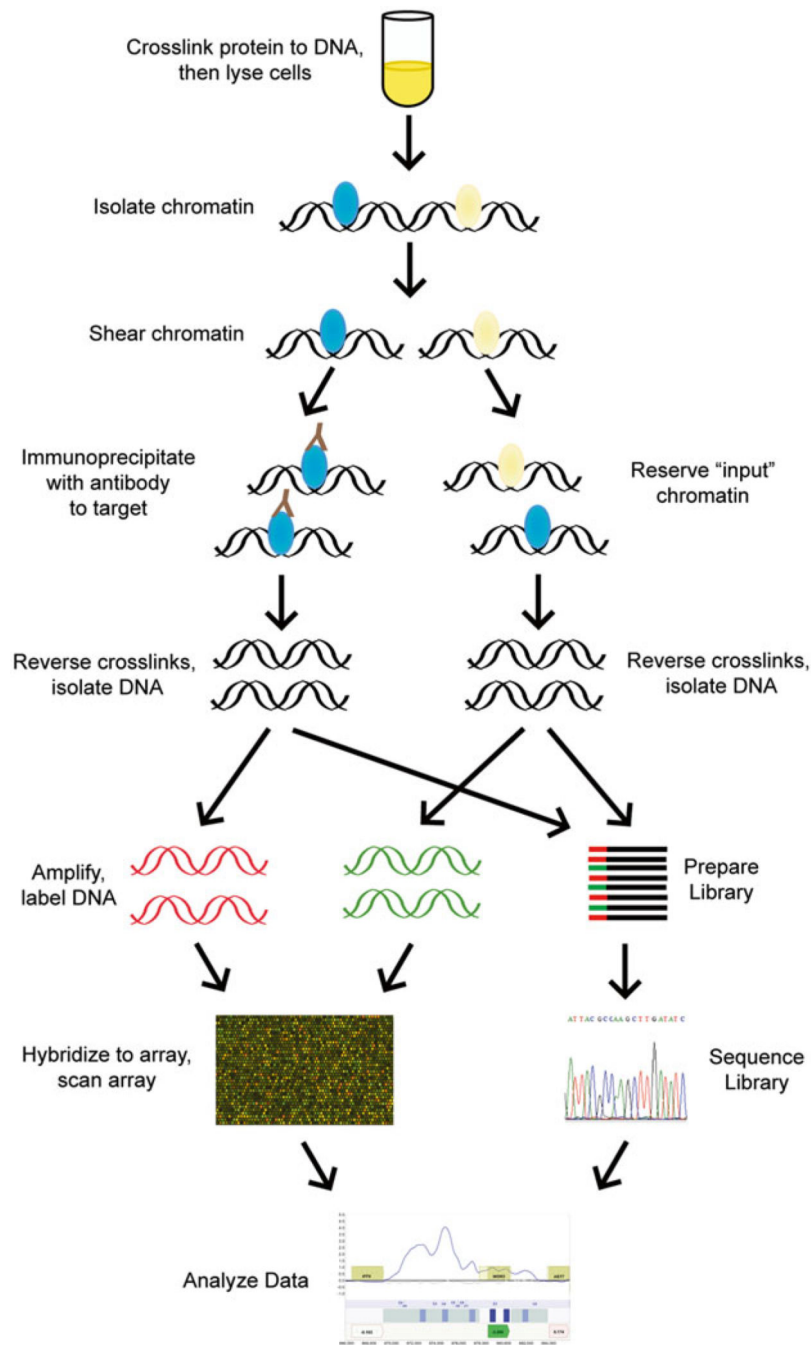
## References

1. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc.* 2006; 1:729–748. [PubMed: 17406303]
2. Nobile CJ, et al. A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell.* 2012; 148:126–138. [PubMed: 22265407]
3. Tuch BB, et al. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* 2008; 6:e38. [PubMed: 18303948]

---

<sup>47</sup>Depending on the data available, it may be informative to compare the targets of a given factor/s across multiple species. Although many of the analyses described above can be performed in this case, additional considerations come into play. The method described assumes a number of genome rearrangements between the species in question, making it necessary to map orthologs of targets between the species. When comparing between closely related species with few to no genome rearrangements, it may be possible to directly compare equivalent intergenic regions and even specific instances of DNA-binding motifs. We have not provided a detailed method for mapping gene orthologs between species as it is beyond the scope of this chapter. The exact details of the mapping process will depend on the resources available for the species in question, our methods are general recommendations based on our experience. If multiple independent orthology calls are available, the different mappings will often have at least subtle differences. Therefore, we recommend performing this analysis with at least two sets of orthology calls if multiple sets are available. Map orthology in both directions (i.e., from species A to B as well as for from B to A) and perform the analysis for both gene lists. Limiting mapping to clear cases of one to one homology is simplest, but often results in little useable information. As such, we recommend considering more ambiguous cases like one to two, two to two, or even many to many. An example of this approach can be found in Cain et al. [4].

4. Cain CW, et al. A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics*. 2012; 190:511–521. [PubMed: 22095082]
5. Baker CR, et al. Protein modularity, cooperative binding, and hybrid regulatory States underlie transcriptional network diversification. *Cell*. 2012; 151:80–95. [PubMed: 23021217]
6. Nguyen VQ, Sil A. Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. *Proc Natl Acad Sci U S A*. 2008; 105:4880–4885. [PubMed: 18339808]
7. Beyhan S, et al. A temperature-responsive network links cell shape and virulence traits in a primary fungal pathogen. *PLoS Biol*. 2013; 11:e1001614. [PubMed: 23935449]
8. Pérez JC, Johnson AD. Regulatory circuits that enable proliferation of the fungus *Candida albicans* in a mammalian host. *PLoS Pathog*. 2013; 9:e1003780. [PubMed: 24385896]
9. Hernday AD, et al. Structure of the transcriptional network controlling white-opaque switching in *Candida albicans*. *Mol Microbiol*. 2013; 90:22–35. [PubMed: 23855748]
10. Hernday AD, et al. Genetics and molecular biology in *Candida albicans*. *Methods Enzymol*. 2010; 470:737–758. [PubMed: 20946834]
11. Homann OR, Johnson AD. MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol*. 2010; 8:49. [PubMed: 20409324]
12. Chakravarty A, et al. A novel ensemble learning method for de novo computational identification of DNA binding sites. *BMC Bioinformatics*. 2007; 8:249. [PubMed: 17626633]
13. Liu X, Brutlag DL, Liu JS. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*. 2001:127–138. [PubMed: 11262934]
14. Bailey TL, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009; 37:W202–W208. [PubMed: 19458158]
15. Teixeira MC, et al. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2014; 42:D161–D166. [PubMed: 24170807]
16. Gupta S, et al. Quantifying similarity between motifs. *Genome Biol*. 2007; 8:R24. [PubMed: 17324271]
17. Kent WJ, et al. The human genome browser at UCSC. *Genome Res*. 2002; 12:996–1006. [PubMed: 12045153]
18. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol*. 2008; 26:1293–1300. [PubMed: 18978777]
19. Lohse MB, et al. Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. *Proc Natl Acad Sci U S A*. 2013; 110:7660–7665. [PubMed: 23610392]
20. Teytelman L, et al. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A*. 2013; 110:18602–18607. [PubMed: 24173036]
21. Hnisz D, et al. A histone deacetylase adjusts transcription kinetics at coding sequences during *Candida albicans* morphogenesis. *PLoS Genet*. 2013; 8:e1003118. [PubMed: 23236295]
22. Tompa M, et al. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005; 23:137–144. [PubMed: 15637633]
23. Baker CR, Tuch BB, Johnson AD. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc Natl Acad Sci U S A*. 2011; 108:7493–7498. [PubMed: 21498688]
24. Askew C, et al. The zinc cluster transcription factor Ahr1p directs Mcm1p regulation of *Candida albicans* adhesion. *Mol Microbiol*. 2011; 79:940–953. [PubMed: 21299649]
25. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–2504. [PubMed: 14597658]
26. Ren B, et al. Genome-wide location and function of DNA binding proteins. *Science*. 2000; 290:2306–2309. [PubMed: 11125145]
27. Traven A, Jelicic B, Sopta M. Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep*. 2006; 7:496–499. [PubMed: 16670683]



**Fig. 1.** Overview of the ChIP-chip and ChIP-seq experimental workflows. In brief, DNA is cross-linked to proteins, isolated from lysed cells, and then sheared into fragments. At this point, a fraction of the sample is separated to process independently as the “input” sample. The protein of interest is then immunoprecipitated from the experimental sample with an antibody against that protein. The cross-links are then reversed for both samples and the DNA isolated. For ChIP-chip, the DNA is amplified and labeled in preparation for hybridization to a high-density oligonucleotide tiling microarray. If performing ChIP-seq,

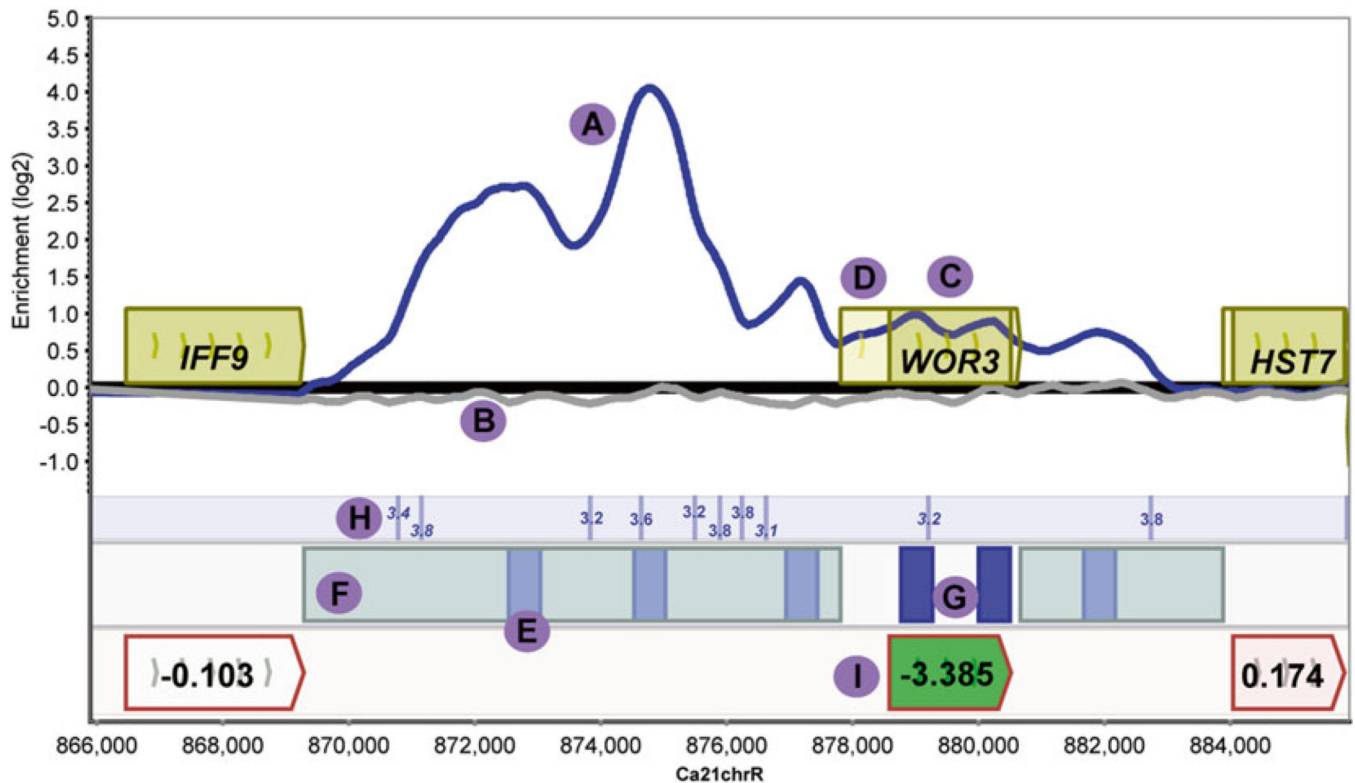
the DNA would be used for library generation in preparation for sequencing. Data generated from both approaches are the starting points for further analysis

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 2.**

Sample of a MochiView screenshot illustrating concepts relating to the analysis of ChIP-chip data. Data in this figure is from a ChIP-chip experiment of *C. albicans* Efg1, a regulator of white-opaque switching, in the opaque cell type. ChIP-chip data for Efg1 (blue, A) and an Efg1 delete control (grey, B) are shown. Open reading frames are represented by yellow boxes (C), lighter yellow represents the untranslated region (D), genes above the bold line are transcribed in the sense direction and genes below the bold line are transcribed in the antisense direction. The x-axis represents ORF chromosomal locations and the y-axis represents the ChIP-chip enrichment value ( $\log_2$ ). A lower track illustrates 500 bp peaks of Efg1 binding (blue boxes, E) and intergenic regions bound by Efg1 (light blue boxes, F) as well as 500 bp Efg1 peaks that do not fall in an intergenic region and would normally be excluded from further analysis (dark blue boxes, G). High-scoring instances of the Efg1 DNA-binding motif (maximum possible score 4.17) is indicated (H). The plot also contains data from a microarray analysis of an opaque *efg1* deletion strain versus wildtype opaque cells; values are on a log 2 scale (I) with downregulated genes in green and upregulated genes in red, color intensity represents differences from wild type (with darker colors having greater differences). The data for Chip-chip, Efg1 motif, and Efg1 delete microarray analysis was taken from Hernday et al. [9]. Plots were constructed using MochiView 1.46, see Homman et al. [11]