ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos

Thomas Sandmann[1], Janus S Jakobsen[1] & Eileen E M Furlong[1]

EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany. Correspondence should be addressed to E.E.M.F. (furlong@embl.de).

Published online 25 January 2007; doi:10.1038/nprot.2006.383

This protocol describes a method to detect *in vivo* associations between proteins and DNA in developing *Drosophila* embryos. It combines formaldehyde crosslinking and immunoprecipitation of protein-bound sequences with genome-wide analysis using microarrays. After crosslinking, nuclei are enriched using differential centrifugation and the chromatin is sheared by sonication. Antibodies specifically recognizing wild-type protein or, alternatively, a genetically encoded epitope tag are used to enrich for specifically bound DNA sequences. After purification and polymerase chain reaction-based amplification, the samples are fluorescently labeled and hybridized to genomic tiling microarrays. This protocol has been successfully used to study different tissue-specific transcription factors, and is generally applicable to *in vivo* analysis of any DNA-binding proteins in *Drosophila* embryos. The full protocol, including the collection of embryos and the collection of raw microarray data, can be completed within 10 days.

INTRODUCTION

A large number of DNA-binding and chromatin-associated proteins are recruited to organize, package, transcribe or repair the genetic information stored in the nucleus of eukaryotic cells. An important step toward understanding any of these (and other) processes at a molecular level is to determine the sites of direct protein–DNA interactions.

ChIP-on-chip allows genome-wide protein–DNA interactions to be assayed *in vivo*

Approaches such as DNAseI footprinting[1] and electrophoretic mobility shift assays[2] allow monitoring of binding events *in vitro*, whereas chromatin immunoprecipitation studies are a valuable tool to probe these essential interactions *in vivo*. By co-precipitating the protein of interest with its associated DNA sequences from cellular or embryonic chromatin extracts, a snapshot of binding site occupancy—reflecting complex variables such as chromatin accessibility, combinatorial binding and/or competition with other DNA-associated factors—can be obtained (**Fig. 1**). Typically, chemical reagents or ultraviolet irradiation are used to stabilize transient interactions by introducing covalent crosslinks. Formaldehyde, a small molecule that readily penetrates biological samples, induces (partially) reversible crosslinks between ε-lysine groups of proteins and (i) neighboring peptide bonds (protein–protein crosslinking) or (ii) amino groups of partially denatured DNA bases (protein–DNA crosslinks)[3,4]. Ultraviolet irradiation, on the other hand, induces irreversible bonds exclusively between nucleic acids and directly bound proteins, but cannot penetrate more than a few cell layers, requiring dissociation of cells from larger tissues and embryos[5–7].

After crosslinking, cells are lysed and the chromatin is sheared, to fragments typically between 0.2 and 1 kb in length, and sequences specifically bound by the protein of interest are purified by immunoprecipitation. Traditionally, Southern blot[8] or quantitative real-time polymerase chain reaction (qPCR)[9] have been used to

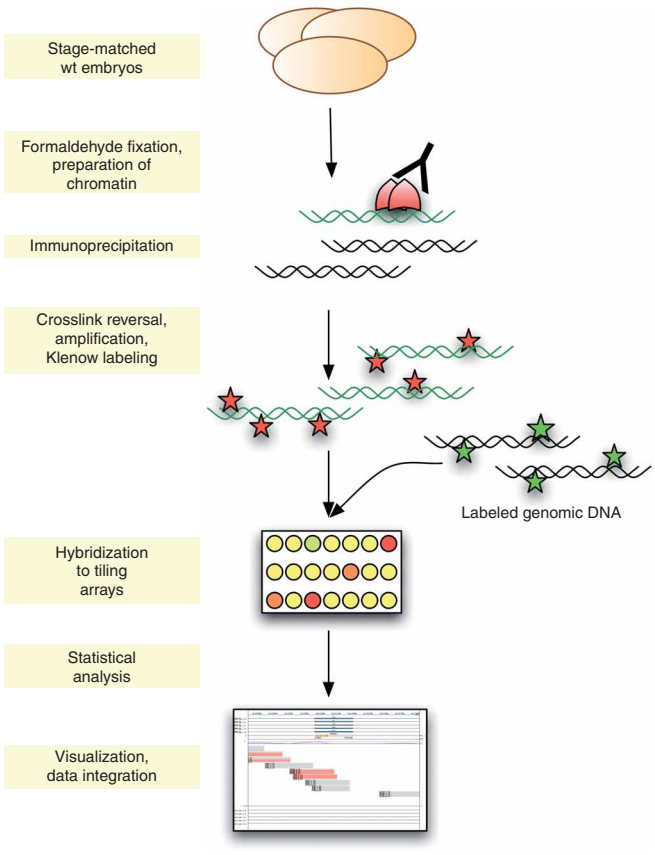

Figure 1 | Schematical overview of ChIP-on-chip experiments. A population of wild-type (or transgenic) *D. melanogaster* embryos is dechorionated and covalent bonds between proteins as well as proteins and nucleic acids are introduced by formaldehyde crosslinking. Shearing the DNA allows immunoprecipitation of short sequences associated with the protein of interest. After partial reversal of the crosslinks, these can be amplified, labeled and hybridized against a genomic reference to genomic tiling arrays. Raw data processing followed by statistical analysis allows identification of significantly enriched sequences bound by the protein of interest.

monitor the differential enrichment of specific sequences, requiring prior knowledge of candidate regions to design suitable probes or oligonucleotide primers[3]. Analyzing chromatin immunoprecipitates using genomic tiling arrays (ChIP-on-chip), on the other hand, allows an unbiased detection of protein–DNA interactions. This powerful combination has highlighted, for example, the role of Polycomb and other transcriptional regulators in human embryonic stem cells[10,11], the occupancy of human promoters by polymerase II preinitiation complexes[12], the recruitment of histone-modifying enzymes to genomic loci in yeast[13], the recruitment of the *Drosophila* dosage–compensation complex to the X-chromosome[14–16], heat–shock factor binding to the second chromosome[17] and the dynamic binding of the tissue-specific transcription factor Mef 2 during *Drosophila* myogenesis[18].

Scope of this protocol

This protocol describes the steps required for ChIP-on-chip analysis of DNA–protein interactions in *Drosophila melanogaster* embryos, including formaldehyde fixation, precipitation of the protein of interest, purification, ligation-mediated PCR amplification and labeling of the enriched sequences, as well as detection of significantly enriched sequences using two-channel genomic tiling microarrays (**Fig. 1**). It has been used successfully to study the genome-wide location of different transcriptional regulators (falling into several structural superfamilies) at multiple stages of development using two different microarray platforms (ref. 18 and T.S., J.S.J. & E.M.F., unpublished results). Earlier protocols designed for chromatin immunoprecipitations from whole *Drosophila embryos* included purification of crosslinked chromatin by isopycnic centrifugation through a CsCl gradient[3,5], a time-consuming procedure difficult to apply to large number of samples (e.g., for time-course studies). Following Birch-Machin *et al.*[17], we have replaced this step by differential enrichment of a crude nuclear subcellular fraction. Using this approach, we have successfully assayed the binding of transcription factors expressed in different subsets of the *Drosophila* mesoderm.

The chromatin immunoprecipitates obtained using this protocol can be analyzed by different means, for example, hybridization to single-channel microarrays[19] or by high-throughput sequencing[20] without requiring changes in the enrichment procedure.

An alternative technique, DamID, is based on genetically fusing the gene of interest with DNA adenine methyltransferase (Dam) and driving the expression of the fusion protein in the tissue of interest[21]. Dam methylates adenine residues within its recognition sites (GA^{m}TC) in the vicinity of the DNA-bound fusion protein. These sequences can therefore be released from isolated genomic DNA with the restriction enzyme *Dpn*1, which specifically cleaves these methylated sequence motifs. This technique provides an independent, sensitive means of obtaining genome-wide localization data similar to that from ChIP-on-chip approaches[22]. To ensure very low expression levels of the fusion protein, it is typically placed under the control of a "leaky" promoter. Although this avoids non-endogenously high expression levels, it leads to ubiquitous expression in all cells of the transgenic organism and may confound the detection of tissue-specific protein–DNA associations. Also, DamID cannot be used for high-resolution mapping of protein binding sites, owing to methylation of Dam recognition sites within several kilobases around the direct binding site of the fusion protein (see refs. 23,24 for recent reviews).

Experimental controls

A number of parameters strongly influence the quality of the data obtained in ChIP-on-chip experiments[25,26]. To achieve sufficiently high levels of specific enrichment, high-quality antibodies—capable of recognizing their antigen after formaldehyde fixation—must be available. To identify *in vivo* binding sites of any DNA-associated protein, the effect of unspecific enrichment of unrelated DNA sequences must be tightly controlled. At least three different effects contribute to the detection of false-positive sequences:

(1) *Unexpected "off-target effects" of the antibodies/antisera used in the experiment*: Even high-quality antibodies can show cross-reactivity in formaldehyde-fixed material and yield seemingly "specific" but biologically misleading results in ChIP-on-chip studies. We, therefore, strongly recommend using two or more antibodies/antisera raised independently against the same protein, whenever available.

Control precipitations from sufficiently large populations of homozygous mutant embryos lacking the protein of interest (if obtainable, e.g., in the case of non-essential gene loci) represent the best control to exclude unspecific results. Similarly, if the gene of interest has been fused to an epitope tag in genetically modified animals, the corresponding non-tagged wild-type strain provides the best choice to control "off-target effects" as well as other sources of experimental noise described below.

(2) *Unspecific enrichment of chromatin through binding to unrelated IgGs*: In the absence of a genetically different control strain, a suitable "mock reaction" needs to be performed on wild-type embryos alongside the 'specific' ChIP-on-chip experiment (**Fig. 2a**). Polyclonal sera contain many different IgGs, only a subset of which will recognize the protein of interest with high affinity. We have successfully performed ChIP-on-chip experiments by comparing the results obtained with unpurified serum (raised against the protein of interest) with those obtained using preimmune serum from the same animal.

(3) *Unspecific enrichment of chromatin as a consequence of adhesion to protein A Sepharose, reaction tubes, etc.*: Finally, even control experiments conducted without any antibody/antiserum yield significant enrichment of specific DNA sequences when compared to similarly amplified genomic reference DNA, most likely owing to sequence-dependent adhesion to the materials and reagents (e.g., reaction tubes, protein A (ProtA) beads, etc.). In the absence of any other control reaction, a minimum mock reaction can be performed with a general IgG control[27] or by simply proceeding through the procedure without adding any antibodies/antisera. In our experience, the amplification of mock reactions yields DNA yields comparable to those from real experimental samples.

Experimental design: replication

Biological replication is of paramount importance for the application of statistical tools to identify significantly enriched sequences in ChIP-on-chip experiments. To assay a single stage of development, several independent embryo collections must be performed and individual chromatin preparations prepared. Assaying at least four independent biological samples per developmental time point typically recovers the vast majority of known binding sites (an even number of experiments is required to perform the experiment with two different antibodies, while an experiment limited to two repeats would lack statistical power). More repeats may be required,

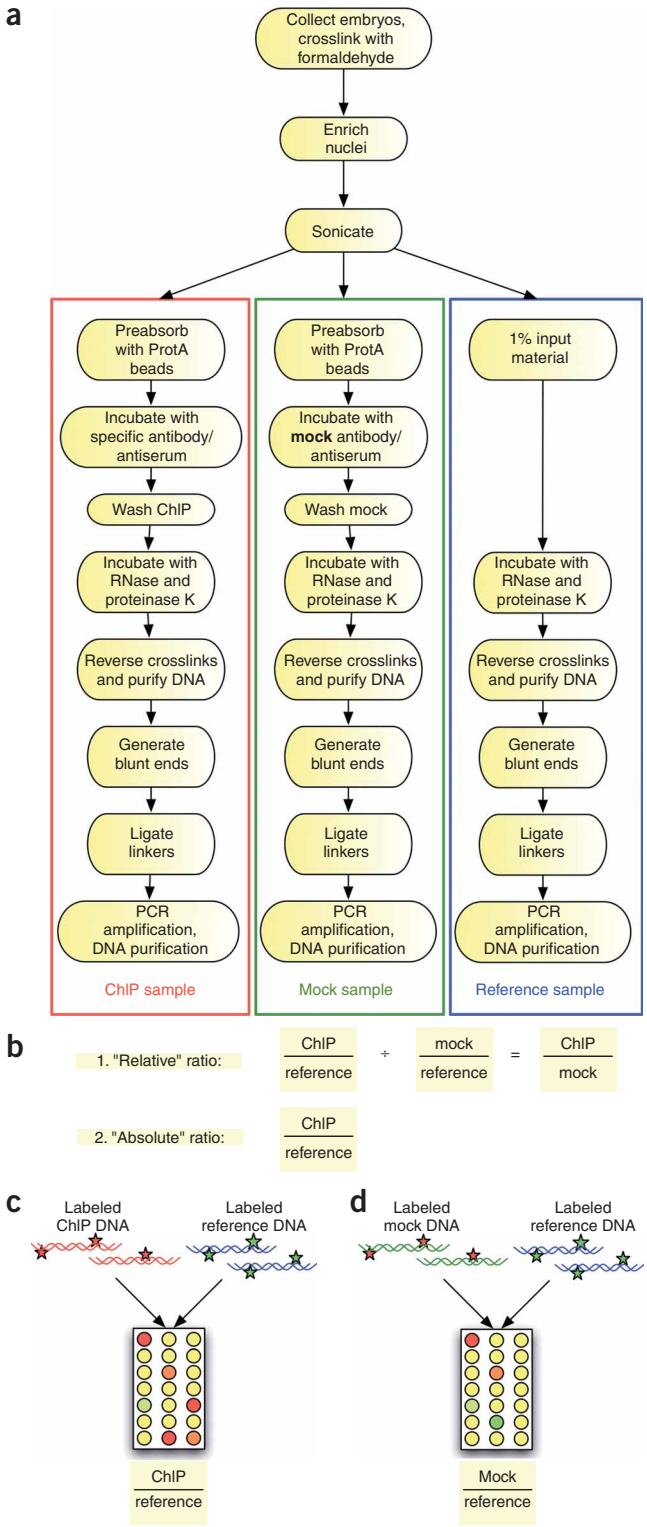

Figure 2 | Overview of the steps outlined in this protocol. (**a**) Schematic overview of the chromatin immunoprecipitations and amplification procedure. Three samples are obtained: (i) amplified ChIP DNA (red box); (ii) corresponding amplified mock control DNA (green box); and (iii) reference DNA obtained by amplifying 1% of the input material. (**b**) Two enrichment ratios can be obtained for each arrayed feature: a "relative" ratio, reflecting the difference between the ChIP and mock reactions, and an "absolute" ratio, indicating the ChIP enrichment when compared to (amplified) genomic DNA. (**c**) ChIP and mock DNA can be compared in a "reference design" by hybridizing both samples individually against the genomic reference.

if very small enrichment ratios need to be evaluated for statistical significance or if large fluctuations in enrichment ratios are observed (see ref. 28 for a systematic review of levels of replication, sample size or analysis strategies).

In our experience, there is considerable variation between different immunoprecipitations from the same starting material. We therefore perform several precipitations from the same chromatin (technical replicates), assay the enrichment by quantitative real-time PCR[9] and choose only successfully enriched samples for amplification and microarray analysis.

Experimental design: microarray analysis

Both experimental and mock precipitations give rise to DNA sequences enriched over a genomic reference. We consider the "relative" enrichment of the experimental sample compared to its mock counterpart (**Fig. 2b**) as well as the "absolute" enrichment of the experimental sample compared to a genomic reference sample (**Fig. 2b**) of interest. These ratios can be obtained by adopting a "reference design", in which both experimental "ChIP" and "mock" samples are separately hybridized to the same (amplified) genomic reference DNA sample (**Fig. 2c,d**). Typically, we first identify significant differences between ChIP and mock samples ("relative ratio", e.g., by applying SAM; see below) and then consider only those features also yielding a minimum "absolute" fold enrichment between "ChIP" and reference samples.

As opposed to a direct comparison of ChIP and mock samples on the same array, this design does not require dye-swaps[29] and allows different statistical approaches to be evaluated at a later stage. Alternatively, if ChIP and mock DNA are compared directly to each other on the same array, one must ensure the correct pairwise relationship between the samples before hybridization—for example, precipitation from the same starting material, serum/preimmune serum pairs originating from the same animal. Also, dye-swaps need to be included in the experimental design to avoid dye-bias effects.

Post-processing and statistical analysis of the microarray data

Similar to raw data obtained from other microarray-based assays, ChIP-on-chip data need to undergo basic processing steps to normalize the independently acquired intensities, alleviate spatial non-homogeneities, etc.[30]. As opposed to expression profiling data, which most often reveals gene expression changes in both positive and negative directions, chromatin immunoprecipitations generate more asymmetric data: specifically bound sequences will exclusively give rise to positive ratio changes (compare **Fig. 3a** and **b**). A number of analysis techniques specifically designed for ChIP-on-chip data sets have been published[31–34]. Users with computational experience can take advantage of the growing suite of useful R-modules available from the Bioconductor project[35] (http://www.bioconductor.org/). We encourage the user to test different approaches and compare the results, especially when evaluating positive control regions.

As a first approximation, although often techniques and tools used for the analysis of expression profiling data can be used to identify specifically enriched microarray features. Below, both post-processing and statistical analysis using the freely available,

platform independent TM4 analysis package[36] (Midas version 2.19 and MeV version 4.0) are described.

Quality control: known binding sites can be assayed efficiently by quantitative real-time PCR

A prerequisite for quality control and optimization of this protocol for use with any DNA-binding protein is knowledge about a previously identified direct binding site. The successful enrichment of this genomic region compared to a bona fide negative control region can be monitored very efficiently using quantitative real-time PCR[9]. In our experience, the high sensitivity of this method compared to end-point analysis of traditional PCRs using agarose gel electrophoresis allows monitoring of optimal antiserum/antibody amounts, determination of the required minimal amounts of chromatin or a comparison between precipitations performed with different antibodies raised against the same protein.

We recommend designing oligonucleotide primers closely (<100 bp) flanking known bindings sites ("positive controls"), as well as within sequences not expected to be bound by the protein of interest, for example, within the coding sequence of unrelated genes ("negative controls"). As the absence of binding is difficult to predict, we typically test a number of negative control regions to identify outliers unexpectedly enriched in the experiment.

To evaluate an experiment, we assay each sample (including mock reactions) individually with both positive and negative primer pairs. To account for differences in amplification fidelity of different oligonucleotide primer pairs, the results are normalized to a dilution series of an unrelated calibration control (e.g., genomic DNA). Each sample's enrichment ratio is calculated as (relative amounts obtained with the positive control primers)/(relative amounts obtained with the negative control primers). Typically, mock samples yield ratios around 1, whereas the enrichment of different positive controls range between 3- and 150-fold, depending on the nature of the DNA-binding protein, the quality of the antibody and the *in vivo* occupancy of the DNA sequence. Typical enrichment ratios obtained with two antisera raised independently against the muscle-specific transcription factor Mef2 are shown in **Figure 4**.

Comparing the relative amounts of product obtained from ChIP samples with those from reference DNA (retained in Step 27 and processed alongside the precipitated material) allows calculation of the fraction of specific sequences precipitated in the experiment. Depending on the nature of the protein of interest and its occupancy of a specific binding site, this number can vary greatly, but is expected to remain relatively consistent throughout the experiment.

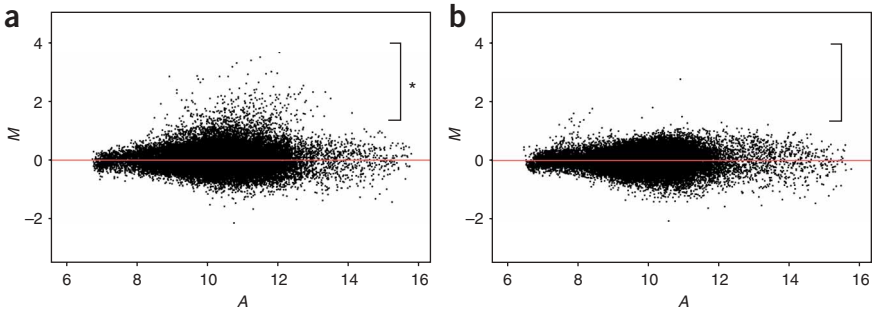

Figure 3 | Enriched sequences can be identified by comparing ChIPs performed with specific anti-Mef2 or preimmune antisera. Amplified DNA obtained from chromatin immunoprecipitations using anti-Mef2 antiserum (**a**) or mock reactions with preimmune serum (**b**) was hybridized against genomic DNA on two different microarrays. *M* versus *A* plots are generated by plotting the $\log_2$ enrichment ratio (*M* axis) of each feature against the logarithm of the product of the intensities in both channels recorded for the same arrayed feature (*A* axis). Several hundred spots are enriched more than twofold ($M>1$) with specific antiserum (**a**, asterisk), whereas only few features are found in this range in the mock experiment (**b**), representing nonspecifically enriched sequences.

MATERIALS

REAGENTS

Staged collections of wild-type *D. melanogaster* embryos (see REAGENT SETUP)
- Formaldehyde (Polysciences, cat. no. 18814); opened vials should be resealed and stored at 4 °C for no longer than a few days ! CAUTION Formaldehyde is a potent mutagen and an expected human carcinogen
- Crosslinking solution: 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 1.8% formaldehyde (v/v), 50 mM HEPES, pH 8.0 ▲ CRITICAL Prepare immediately before use to avoid inactivation of formaldehyde during storage
- Stop solution: 125 mM glycine, 0.1% Triton X-100 (v/v) in PBS
- *n*-Heptane (Sigma, cat. no. H9629) ! CAUTION Highly flammable, toxic to aquatic organisms
- Methanol (Merck, cat. no. 106009) ! CAUTION Highly flammable, toxic
- Sodium hypochlorite (Merck, cat. no. 105614) ! CAUTION Corrosive, contact with acids liberates toxic gas
- PBT: 0.1% Triton X-100 (v/v) in PBS
- Glycerol (Merck, cat. no. 104091)
- Sodium deoxycholate (Sigma, cat. no. D6750)
- *N*-Lauroylsarcosine (Sigma, cat. no. L9150)
- Cell lysis buffer: 85 mM KCl, 0.5% IGEPAL CA-630 (v/v), 5 mM HEPES, pH 8.0
- Nuclear lysis buffer: 10 mM EDTA, 0.5% *N*-lauroylsarcosine (w/v), 50 mM HEPES, pH 8.0
- Protease inhibitors (1,000× stock): Leupeptin: 10 mg ml^{-1} in DMSO; pepstatin: 10 mg ml^{-1} in DMSO; aprotinin: 10 mg ml^{-1} in water
- Phenylmethylsulfonyl fluoride (PMSF) ! CAUTION PMSF is toxic and severely corrosive to the eyes
- RIPA buffer: 140 mM NaCl, 1 mM EDTA, 1% Triton X-100 (v/v), 0.1% SDS (w/v), 0.1% sodium deoxycholate (w/v), 10 mM Tris-HCl, pH 8.0
- RIPA500 buffer: 500 mM NaCl, 1 mM EDTA, 1% Triton X-100 (v/v), 0.1% SDS (w/v), 0.1% sodium deoxycholate (w/v), 10 mM Tris-HCl, pH 8.0
- Lithium chloride buffer: 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL CA-630 (v/v), 0.5% sodium deoxycholate (w/v), 10 mM Tris-HCl, pH 8.0
- TE buffer: 1 mM EDTA, 10 mM Tris-HCl, pH 8.0
- ProtA Sepharose CL4B (Sigma, cat. no. P9424) (use as 50% slurry)
- Protein G Sepharose CL4B (Sigma, cat. no. P3296) (use as 50% slurry)
- Bovine serum albumin (BSA) (Sigma, cat. no. A7906)
- RNase A (Qiagen, cat. no. 1006693)
- Proteinase K (Roche, cat. no. 745723)
- Phenol:chloroform:isoamylalcohol (Ambion, cat. no. 9732) ! CAUTION Phenol is highly corrosive and can cause severe burns
- Chloroform (Sigma, cat. no. C2432) ! CAUTION Chloroform is toxic and is a suspected carcinogen
- Glycogen (Roche, cat. no. 901393)
- T4 DNA polymerase (Roche, cat. no. 1004786)
- T4 DNA ligase (Roche, cat. no. 799099)
- Cy3-dUTP fluorophore (GE-Healthcare, cat. no. PA53022)

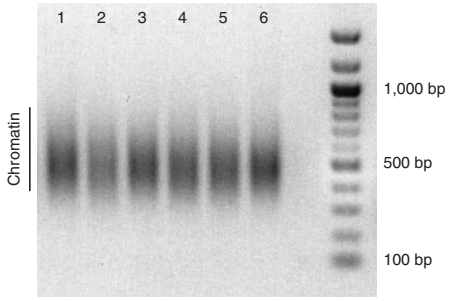

Figure 4 | Crosslinked chromatin can be sheared reproducibly by sonication. To verify the average sequence length within different batches of chromatin, 50 µl of chromatin (obtained in Step 21) was treated with RNase A and proteinase K and the covalent crosslinks were partially reversed at 65 °C. The DNA was purified by phenol–chloroform extraction and ethanol precipitation and analyzed by agarose gel electrophoresis on a 1% gel. Six independent chromatin preparations (lanes 1–6) were processed. All show a similar unimodal sequence length distribution (visualized by ethidium bromide staining) with a peak around 500 bp.

- Cy5-dUTP fluorophore (GE-Healthcare, cat. no. PA55022)
- 100× Denhardt's reagent: 1 g Ficoll 400 (Sigma, cat. no. F4375), 1 g polyvinylpyrrolidone (Sigma, cat. no. P5288), 1 g BSA, H_2O to 50 ml
- 20× SSC (Ambion, cat. no. 9765)
- Hybridization solution: 1% SDS (w/v), 5× Denhardt's reagent, 0.8 mg ml^{-1} polyA (Sigma, cat. no. P9403), 50% formamide (v/v) (Sigma, cat. no. F5786), 3× SSC
- Hybridization wash solution 1: 1× SSC/0.03% SDS (w/v), filtered
- Hybridization wash solution 2: 0.2× SSC, filtered
- Hybridization wash solution 3: 0.05× SSC, filtered
- Qiaquick min-elute PCR purification kit (Qiagen, cat. no. 28004)
- Bioprime CGH Genomic DNA labeling system (Invitrogen, cat. no. 18095-011)
- Annealing buffer: 1 mM EDTA, 50 mM NaCl, 10 mM Tris, pH 8.0
- Annealed oligonucleotide linkers (see REAGENT SETUP) ▲ CRITICAL Avoid contamination with foreign DNA or crosscontamination between samples. We recommend the use of filter-tips

EQUIPMENT

- Nitex membranes, 120 µm (Sefar)
- Sieves (Fisher Scientific, cat. no. 14-306A)
- 15 ml Dounce homogenizer (Wheaton)
- Bioruptor sonicator water bath (Diagenode)
- Siliconized reaction tubes (Biozym, cat. no. 710176)
- 2 ml phase-lock heavy gel tubes (Eppendorf, cat. no. 0032-005-152)
- Speed-vac concentrator
- Nanodrop spectrophotometer ND-1000 (Nanodrop Technologies) or spectrophotometer with a sufficiently narrow quartz cuvette
- Hybridization chambers (Corning, cat. no. 2551)
- Thermocycler
- Microarray scanner, for example, Genepix 4000B (Molecular Devices)

REAGENT SETUP

Staged collections of wild-type *D. embryos* Large numbers of wild-type embryos can be obtained from population cages[37]. Three cages seeded with 20 g of adult wild-type flies typically allow collection of ≥1.5 g from a 2-h lay. We recommend not to use the first 2–3 collections of the day, as egg laying is often delayed during night owing to overcrowding of the food containers. Consequently, embryogenesis starts within the uterus and the early collections usually contain a broad range of developmental stages.

Preparation of annealed oligonucleotide linkers ● TIMING **2.5 h** To provide priming sites for PCR amplification, double-stranded oligonucleotide linkers are ligated to both ends of the precipitated DNA fragments. The linker is obtained by annealing the following two complementary oligonuclotides (the 20-mer oligomer is also used later to prime the PCR amplification reaction)[38]: 24-mer oligonucleotide: AGA AGC TTG AAT TCG AGC AGT CAG, 5′ phosphorylated; 20-mer oligonucleotide: CTG CTC GAA TTC AAG CTT CT. Resuspend each oligonucleotide in annealing buffer to a final concentration of 100 µM. Mix equal amounts of 24- and 20-mer oligonucleotides in a 1.5 ml reaction tube. Incubate the tube in a beaker with approximately 400 ml boiling water for 5 min, move the beaker from the heat plate and let the water slowly cool to room temperature (20–25 °C). Annealed linkers can be stored at −20 °C for years. Dilute an aliquot to 1 µM before use in a ligation reaction (Step 48). Alternatively, a slow temperature ramp in a thermocycler can be used to generate double-stranded linkers.

PROCEDURE

Embryo collection and crosslinking ● TIMING 1 h

▲ CRITICAL This protocol produces 4 ml of crosslinked chromatin from 1.5 g of embryos, representing sufficient material for at least ten immunoprecipitations.

▲ CRITICAL Unless indicated otherwise, centrifugation and incubation steps are performed at room temperature (20–25 °C).

1| Collect large amounts (≥1.5 g) of staged wild-type or transgenic embryos from population cages and dechorionate them in a stirred 3% sodium hypochlorite (v/v) solution (or 50% commercial bleach) at room temperature for 2.5 min.

! CAUTION Sodium hypochlorite is corrosive and contact with acids liberates toxic gas.

2| Transfer the embryos into pre-weighed sieves and wash them well with tap water. Remove as much liquid as possible by blotting the sieve with a paper towel, then weigh the sieve and calculate the net weight of the embryos. Resuspend them in 10 ml PBT per 1.5 g of embryos in a glass beaker.

3| Place a sufficient number of Nitex membranes (approximately 5 × 5 cm each) onto a pile of tissue paper and pipette 10 ml of embryo suspension (corresponding to approximately 1.5 g of embryos) onto each membrane. Fold the membrane over to cover the embryos and blot them dry with a paper towel.

4| Transfer each membrane into a separate 50 ml Falcon tube with 10 ml crosslinking solution + 30 ml heptane, shake off the embryos, recover the membrane and shake the tube vigorously at room temperature (20–25 °C) for 15 min.

! CAUTION *n*-Heptane is highly flammable and toxic to aquatic organisms. Formaldehyde is a potent mutagen and an expected human carcinogen.

▲ CRITICAL This step crosslinks proteins to chromatin as well as to other proteins. The time required for this step should be kept constant between repeated collections. Importantly, as some proteins are more easily crosslinked to chromatin

than others, the formaldehyde concentration/length of crosslinking reaction might require optimization for different proteins of interest.

5| Pellet the embryos in each tube by centrifugation at 500*g* for 1 min, replace the supernatant with 30 ml stop solution and shake vigorously at room temperature for at least 1 min to stop the crosslinking reaction.
! CAUTION The supernatant contains *n*-heptane and formaldehyde.

6| Pellet the embryos by centrifugation at 500*g* for 1 min, carefully decant the supernatant and wash the pellet with 50 ml PBT.

7| Pellet the embryos by centrifugation at 500*g* for 1 min, decant the supernatant and resuspend the embryos in approximately 10 ml PBT per tube. Transfer the embryos onto separate Nitex membranes as in Step 3, fold the membrane over to cover the embryos and blot them dry with a paper towel.

8| Transfer a small number of embryos (100–200) from any of the membranes into a microfuge tube containing 0.5 ml heptane and 0.5 ml methanol. Shake vigorously to devitellinize the embryos, let them settle and then remove as much liquid as possible. Wash the embryos with methanol twice and store at −20 °C. This sample from the collection is set aside to evaluate if the collected embryos are at the correct developmental stage (see **Box 1**).
! CAUTION Methanol is toxic and highly flammable.

9| Transfer the remaining aliquots of dry crosslinked embryos from the Nitex membranes (Step 7) into separate cryotubes and freeze them in liquid nitrogen.
■ **PAUSE POINT** Crosslinked embryo collections can be stored at −80 °C for at least 1 year.
Preparation of chromatin extracts
● **TIMING** Approximately 3 h, depending on the number of chromatin extracts prepared.

10| Each 1.5 g aliquot of crosslinked embryos (Step 9) provides sufficient material for at least ten immunoprecipitations. We recommend processing no more than four samples at the same time. Thaw each sample at room temperature, resuspend them in 15 ml cold PBT supplemented with protease inhibitors and 1 mM PMSF and transfer the suspension into a 15 ml Dounce homogenizer.
! CAUTION PMSF is toxic and severely corrosive to the eyes.

11| Homogenize each 1.5 g aliquot of embryos in a 15 ml Dounce homogenizer on ice by applying 20 strokes with the loose-fitting pestle.

12| Transfer the lysate into a suitable centrifuge tube and centrifuge at 400*g* at 4 °C for 1 min to precipitate the vitelline membranes and large debris.

13| Decant the supernatant into a fresh centrifuge tube and centrifuge at 1,100*g* at 4 °C for 10 min. Decant the supernatant and discard it.

14| Resuspend the cell pellet in 15 ml cold cell lysis buffer supplemented with protease inhibitors and 1 mM PMSF.
! CAUTION PMSF is toxic and severely corrosive to the eyes.

15| Homogenize the cells in a 15 ml Dounce homogenizer on ice by applying 20 strokes with the tight-fitting pestle. Split the sample into two approximately 8 ml aliquots in two separate 15 ml Falcon tubes.

16| Centrifuge the samples at 2,000*g* at 4 °C for 4 min to pellet the nuclei. Discard the supernatant.
■ **PAUSE POINT** Nuclei can be frozen in liquid nitrogen and can be stored for at least 1 year at −80 °C.

BOX 1 | EVALUATING THE DISTRIBUTION OF DEVELOPMENTAL STAGES WITHIN A COLLECTION

For staging purposes, rehydrate the sample set aside in Step 8 by rinsing the embryos once in 1 ml of each of the following solutions (v/v): (1) 70% methanol/30% PBT, (2) 50% methanol/50% PBT, (3) 30% methanol/70% PBT. Finally, wash the embryos once in 1 ml PBT for 5 min. Replace the supernatant with 85% glycerol and mount the embryos on a microscope slide. The developmental stages present in the collection can now be examined microscopically using the morphological features defined by Campos-Ortega and Hartenstein[39]. Compare the distribution of developmental stages between repeated collections and exclude those containing inappropriate stages from further analysis, if necessary.

17| Resuspend each pellet in 1 ml of cold nuclear lysis buffer supplemented with protease inhibitors and 1 mM PMSF and incubate at room temperature for 20 min.
! CAUTION PMSF is toxic and severely corrosive to the eyes.

18| Add 1 ml of cold nuclear lysis buffer supplemented with protease inhibitors and 1 mM PMSF to each sample and sonicate 10× using a precooled Bioruptor sonicator water bath (15 s on/15 s off cycles, low-energy settings). Ensure that the water bath remains cold by adding small amounts of ice after five sonication cycles.
! CAUTION PMSF is toxic and severely corrosive to the eyes.

19| Transfer the chromatin into 1.5 ml Eppendorf tubes and centrifuge at 20,000*g* at 4 °C for 10 min.

20| Pool the supernatants to ensure a homogenous sample, remove 50 µl for quality assessment and freeze the remaining chromatin in 400 µl aliquots in liquid nitrogen.
■ PAUSE POINT The samples can be stored at −80 °C for at least 1 year.

21| To determine the yield and average fragment length of the chromatin preparations, dilute the 50 µl of chromatin set aside in Step 20 with 50 µl buffer TE. Add RNase to 50 µg ml^{-1}, treat the sample as described in Steps 34–40 and resuspend the purified DNA in 50 µl TE. Determine the concentration of sheared DNA spectrometrically and inspect its size distribution by gel electrophoresis using a 1% agarose gel. The Bioruptor settings used in this protocol give rise to an average fragment length of approximately 500 bp (**Fig. 4**). Check if the samples have been sheared to comparable fragments lengths within the desired range.
? TROUBLESHOOTING

Preabsorption + antibody addition ● TIMING 1.5 h sample preparation, 1 h preincubation, overnight incubation
22| Pool 25 µl of 50% ProtA Sepharose suspension per precipitation (including mock reactions) and wash the beads once with 1 ml RIPA buffer.
▲ CRITICAL STEP Antibodies raised in different species differ in their affinity toward ProtA. Alternatively, protein G or ProtA/G Sepharose beads can be used.

23| Pellet the beads by centrifugation at 1,000*g* for 2 min, discard the supernatant and resuspend in 100 µl RIPA buffer per 25 µl ProtA beads.

24| Thaw an aliquot of chromatin and adjust the final volume to 500 µl with cold TE buffer. Depending on the abundance of the protein of interest, typically 200–400 µl of chromatin is used for a single immunoprecipitation reaction. Optimal immunoprecipitation depends on the correct ratio between accessible antigen and antibody. We recommend starting with 200 µl of chromatin and titrating the amount of antibody yielding highest enrichment ratios. Afterwards, possible improvement of these enrichment ratios by proportional up-scaling of the amounts of chromatin and antibody can be evaluated.

25| To each sample, add 100 µl 10% Triton X-100, 100 µl 1% deoxycholate, 100 µl 1% SDS, 100 µl 1.4 M NaCl, 10 µl 100 mM PMSF, in the order given. Mix the sample after addition of each separate component by inverting the reaction tube several times.

Component	Amount	Final
10% Triton-X 100	100 µl	1%
12% eoxycholate	100 µl	0.1%
1% SDS	100 µl	0.1%
1.4 M NaCl	100 µl	0.14 M
100 mM PMSF	10 µl	1 mM

! CAUTION PMSF is toxic and severely corrosive to the eyes.

26| Resuspend the beads prepared in Step 23 and add 100 µl of the bead suspension to each sample. Incubate on a rotating wheel at 4 °C for 1 h and pellet the beads by centrifugation at 1,000*g* at 4 °C for 2 min.
This acts as a preclearing step to remove chromatin nonspecifically binding to ProtA Sepharose.

27| Transfer the supernatant to a new, siliconized tube and avoid any bead carryover. Retain 10 µl of the sample (1% input) in a separate tube and store at 4 °C until Step 34.
▲ CRITICAL STEP Always use siliconized tubes for immunoprecipitations to reduce nonspecific adherence to the reaction tubes.

28| Add 3–20 µl of serum or a suitable amount of purified antibody to each tube. Incubate at 4 °C overnight on a rotating wheel.
▲ CRITICAL STEP The optimal amount of antiserum/antibody is difficult to predict, as it depends on the abundance of the protein of interest, the concentration of specific IgGs in the serum, etc. We recommend to test a range of different

antiserum/antibody amounts (e.g., 5, 10 and 20 µl) and to monitor the enrichment of a known binding site using quantitative real-time PCR (see below).
▲ CRITICAL STEP Always include a mock control. Incubate separate mock reactions with the same amount of preimmune serum or use an unrelated antibody raised in the same species to control for nonspecifically enriched sequences.

29| For each precipitation (including mock reactions), wash 25 µl of 50% ProtA suspension with 1 ml RIPA buffer supplemented with 1 mg ml^{-1} BSA, pellet the beads by centrifugation at 1,000*g* for 2 min and incubate in 1 ml RIPA buffer supplemented with 1 mg ml^{-1} BSA on a rotating wheel at 4 °C overnight. Preblocking the ProtA Sepharose beads with BSA reduces nonspecific binding of chromatin.

Purification of immunocomplexes ● TIMING 3 h incubation, 2 h wash steps, 30 min RNase digest
30| Centrifuge the preblocked ProtA beads at 1,000*g* for 2 min, discard the supernatant and resuspend the beads in 100 µl RIPA buffer per reaction.

31| Add 100 µl of ProtA suspension to each (approximately 1 ml) chromatin sample from Step 27 and incubate at 4 °C on a rotating wheel for 3 h.

32| To purify the antigen–antibody complexes, pellet the beads by centrifugation at 1,000*g* for 2 min, discard the supernatant and rinse the beads once with 1 ml cold RIPA buffer.

33| Pellet the antigen–antibody complexes again by centrifugation at 1,000*g* for 2 min and wash them with 1 ml of each of the following buffers at 4 °C on a rotating wheel for 10 min: 1× with RIPA buffer, 4× with RIPA500, once with LiCl buffer, 2× with TE.

34| Resuspend the beads in 100 µl TE buffer supplemented with 50 µg ml^{-1} RNase A and incubate at 37 °C for 30 min. From this point on, include also the 1% input sample retained in Step 27. Add 90 µl TE buffer to yield a final volume of 100 µl before RNase addition.

35| Add SDS to a final concentration of 0.5% (w/v) from a 10% stock and incubate with 0.5 mg ml^{-1} proteinase K at 37 °C overnight.

Crosslink reversal and DNA purification ● TIMING 6 h incubation, 30 min extraction, 1.5 h precipitation
36| Incubate the samples at 65 °C for 6 h to partially reverse the crosslinks.

37| Extract the DNA by combining the sample with 300 µl phenol:chloroform:isoamylalcohol in a pre-spun phase-lock tube, vortex briefly and centrifuge at >15,000*g* at room temperature for 5 min. Add 300 µl chloroform, vortex briefly and centrifuge again at >15,000*g* at room temperature for 5 min. Transfer the aqueous sample to a fresh 1.5 ml reaction tube. (The sepharose beads will precipitate below the phase lock.)
! CAUTION Phenol is highly corrosive and can cause severe burns. Chloroform is toxic and is a suspected carcinogen.
▲ CRITICAL STEP In our experience, replacing the phenol–chloroform extraction step with column-based purification methods (e.g., Qiaquick PCR purification kit, Qiagen) can interfere with subsequent real-time PCR analysis.

38| Supplement the samples with 0.25 mg ml^{-1} glycogen, add 25 µl 3 M sodium acetate solution, pH 5.2, and 550 µl ethanol, vortex briefly and incubate the sample at −80 °C for ≥30 min.

39| Centrifuge the sample at 4 °C at >15,000*g* for 30 min to precipitate the DNA, wash the pellet once with 1 ml of 70% ethanol and centrifuge again at >15,000*g* at 4 °C for 10 min.

40| Resuspend the purified DNA in 30 µl TE.
■ PAUSE POINT Samples can be stored at −80 °C for at least 1 year.

41| If sequence-specific binding sites for the protein of interest are known, evaluate the efficiency of enrichment by quantitative real-time PCR (assay 1–2 µl of sample per PCR; **Fig. 5**). This quality control avoids costly hybridizations of unsuccessful immunoprecipitations. Alas, if previously identified binding sites are lacking, microarray analysis can still be performed and used to identify enriched sequences. We strongly recommend to evaluate these "novel" binding sites by quantitative PCR and to optimize the protocol at this step, instead of assessing the results only after microarray analysis. As the full protocol contains many steps, quality control after each major part helps identifying problems early and identifies troubleshooting needs.
? TROUBLESHOOTING

Generation of blunt ends on sample DNA ● TIMING 45 min

42| Set up the following reaction for each sample and incubate at 37 °C for 30 min.

▲ **CRITICAL STEP** Also include a "no DNA" control containing water instead of sample DNA to monitor possible contamination problems.

Component	Amount	Final
Sample DNA	7 μl	—
5× T4 polymerase buffer	10 μl	1×
2 mM dNTPs	2.5 μl	0.1 mM
T4 DNA polymerase	1 μl (5 U μl−1)	0.1 U μl^{-1}
Water	29.5 μl	—

43| Purify the DNA using the Qiagen min-elute kit following the manufacturer's instructions.

44| Reduce the volume to less than 7 μl using a speed-vac centrifuge.

■ **PAUSE POINT** The purified DNA can be stored in TE buffer at −20 °C for years.

Linker ligation ● TIMING Overnight incubation, 2 × 2 h PCR

45| In a thin-walled PCR tube, set up the following ligation reaction for each sample and incubate the reaction at 16 °C overnight.

Component	Amount	Final
Blunt-ended, purified DNA sample from Step 44	Up to 7 μl	—
10× ligation buffer	1 μl	1×
Annealed linkers (1 μM)	1 μl	0.1 μM
T4 DNA ligase (5 U μl^{-1})	1 μl	0.5 U μl^{-1}
Water	To 10 μl	—

▲ **CRITICAL STEP** The linkers will be ligated to any blunt-ended DNA. Avoid contamination with foreign DNA and also process the "no DNA" control reaction without sample DNA to monitor possible contamination problems.

46| To each ligation reaction, add the following reagents to set up a PCR amplification reaction:

Component	Amount	Final
Ligation reaction from Step 45	10 μl	—
10× amplitag buffer (15 mM Mg^{2+})	10 μl	1× (1.5 mM Mg^{2+})
2 mM dNTPs	10 μl	0.2 μM
Amplitag polymerase (5 U μl^{-1})	1 μl	0.05 U μl^{-1}
100 μM 20-mer oligonucleotide	1 μl	1 μM
Water	68 μl	—

47| Perform a PCR using the following cycling conditions:

Step	Time	Temperature (°C)
1	2	55
2	5	72
3	5	94
4	1	94
5	1	55
6	1	72
19 cycles of 4–6		
7	5	72
8	Hold	4

▲ **CRITICAL STEP** Depending on the length of the sheared DNA fragments and thermocycler performance, the PCR program might need to be adjusted. This program was used successfully in a PTC-200 cycler (MJ-Research).
Limiting the number of PCR cycles to 20 reduces the risk of nonlinear amplification, which can skew the distribution of DNA molecules in the sample.

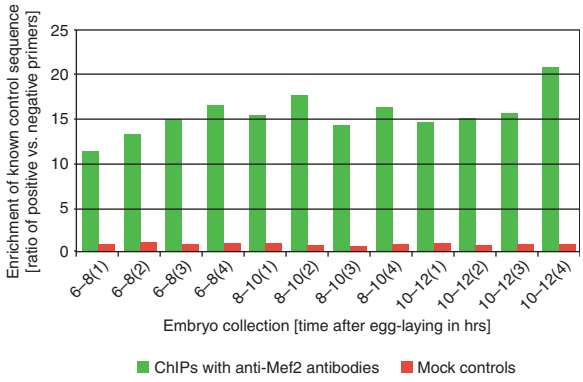

Figure 5 | Enrichment of a known Mef2-binding site within the *Act57B* locus. Four independent embryo collections were assayed at each developmental time point and assayed with one of two different anti-Mef2 antisera (at each time point, chromatin preparations 1 + 2 were precipitated with serum A and 3 + 4 with serum B). Specific enrichment of a known binding site within the Act57B enhancer[43] was assayed by quantitative real-time PCR and compared to a negative control region. All ChIPs performed with anti-Mef2 antisera (green) yield at least an 11.5-fold enrichment over the negative control, whereas mock precipitations performed with the respective preimmune sera (red) yield no enrichment (ratios around 1). There is noticeable variation in the enrichment ratios obtained with the same antiserum (compare, e.g., samples 10–12(3) with 10–12(4)), reflecting variation between different chromatin preparations as well as biological variation in the starting material.

■ **PAUSE POINT** The amplicons generated in this step provide a stock for an additional round of PCRs and can be stored at −20 °C.

48| To generate sufficient amounts of DNA for microarray analysis, combine the following reagents with 10 µl from each unpurified PCR from Step 47:

Unpurified PCR reaction from Step 47	10 µl	—
10× amplitag buffer (15 mM Mg^{2+})	10 µl	1× (1.5 mM Mg^{2+})
2 mM dNTPs	10 µl	0.2 µM
Amplitag polymerase (5 U μl^{-1})	1 µl	0.05 U μl^{-1}
100 µM 20-mer oligonucleotide	1 µl	1 µM
Water	68 µl	—

Amplify the sample using the thermocycler program shown in Step 47. Common yields range between 2 and 6 µg of DNA, depending on the protein and the developmental stages of interest.
▲ **CRITICAL STEP** To assess the reproducible amplification of enriched sequences, quantitative real-time PCR analysis should be used. Avoid column-based purification systems before the analysis; instead, simply dilute the unpurified PCRs in water (e.g., 1:25 after Step 47 or 1:200 after Step 48) and assay 2 µl per reaction.
? TROUBLESHOOTING

49| Analyze 5 µl of PCR product by 1% agarose gel electrophoresis. The fragment length distribution should be similar to that of the initial chromatin preparation. No product should be detectable in the "no DNA" control sample.

50| Purify the amplification products using the min-elute PCR purification kit according to the manufacturer's conditions.
■ **PAUSE POINT** The purified DNA can be stored in TE buffer at −20 °C for years.

Direct Klenow labeling of amplified DNA ● TIMING 3.5 h

51| Bring 0.5 µg of PCR-amplified DNA to a volume of 10.5 µl with H_2O and mix with 10 µl of 2.5× random prime reaction buffer.

52| To denature the DNA, incubate the sample at 100 °C for 5 min and immediately place on ice afterwards.

53| Add 2.5 µl of 10× UTP mix, 1.5 µl Cy3 or Cy5-dUTP and 0.5 µl Klenow polymerase mix and incubate at 37 °C in the dark for ≥3 h (the incubation can be extended to run, e.g., overnight).
▲ **CRITICAL STEP** Protect the Cy dyes and all labeled products from light as much as possible.

54| Stop the reaction by adding 5 µl of 0.5 M EDTA and purify the probes with the Qiagen min-elute PCR purification kit according to the manufacturer's instructions. Use dark-tinted Eppendorf tubes to reduce exposure to light.

55| Quantify the spectrometric absorption at 260, 550 and 650 nm, for example, using a Nanodrop spectrometer, and calculate the specific activity according to the following equations (valid for double-stranded DNA):
Frequency of Cy3 incorporation (dye/100 bases) = $4.33 \times A_{550}/A_{260}$.
Frequency of Cy5 incorporation (dye/100 bases) = $2.6 \times A_{650}/A_{260}$.
Expected DNA yield: approximately 1 µg of double-stranded DNA per labeling reaction.
Expected specific activity: 1.5–2.5% (Cy5) and 2.0–2.5% (Cy3).
? TROUBLESHOOTING

56| Combine 1 µg of the respective Cy3- and Cy5-labeled samples and lyophilize the DNA using a speed-vac centrifuge protected from light at room temperature.
▲ **CRITICAL STEP** Depending on the microarray platform used, the amounts of labeled DNA probes may need to be adjusted.

Hybridization of microarrays in Corning hybridization chambers ● TIMING 15 min hybridization setup, 12–16 h hybridization, 15 min wash steps

▲ CRITICAL As for microarray assays in general, it is difficult to give general guidelines for the use of genomic tiling arrays based on different technological platforms. In particular, the amounts of probe required for hybridization, the nature and the volume of the hybridization solution as well as the wash conditions vary greatly depending on the system used. The following conditions have been used successfully with long (3 kb) PCR probes spotted onto Codelink glass slides (GE Healthcare) in sodium phosphate buffer or high-density 60-mer oligonucleotide microarrays. We advise the user to obtain information about appropriate parameters and conditions from the supplier of the specific microarrays used for the experiment.

57| Resuspend the labeled probe in 35 µl of hybridization solution.

58| Place a suitable coverslip on a clean, even surface.

59| Add 15 µl 3× SSC buffer into each humidifier well of the hybridization chamber.

60| Denature the probe at 100 °C in a heat block for 2 min, centrifuge for 10 s in a tabletop centrifuge to remove condensation from the lid and wait for an additional 10 s before applying the probe to the center of the coverslip.

▲ CRITICAL STEP Avoid creating bubbles or remove them before the next step. Proceed swiftly to avoid cooling and reannealing of the probe DNA molecules.

61| Carefully lower the microarray, DNA side facing down, onto the coverslip, until the probe spreads evenly between the two glass surfaces.

62| Quickly place the slide into the hybridization chamber, attach and seal the lid and submerge the chamber in a level 42 °C water bath. Hybridize for 12–16 h in the dark.

Washing procedure and data acquisition ● TIMING 20 min

63| Fill three separate slide wash chambers fitted with slide racks with hybridization buffers 1, 2 and 3.

64| Disassemble the hybridization chamber and quickly submerge each array into hybridization solution 1.

▲ CRITICAL STEP Avoid removing the coverslip while the array is exposed to air to prevent washing artifacts. Instead, let the coverslip slide off in the wash solution by itself.

65| Gently move the slide rack up and down for 2 min. Avoid scratching the array with the loose coverslip.

66| Using two forceps, carefully hold each array by the ends and quickly transfer it into hybridization wash solution 2. Wash the arrays again while gently moving the rack for 2 min.

67| Move the arrays into hybridization wash solution 3, as described in Step 66, and wash them while gently moving the rack for 2 min.

68| Move the last wash chamber next to a centrifuge fitted with microtiter plate holders. Lift the slide rack out of the washing solution and blot it very briefly on a lint-free paper towel before centrifuging at 500*g* for 5 min to drain the remaining liquid from the slides.

▲ CRITICAL STEP Move the slide holder from the wash solution into the centrifuge swiftly and immediately start the centrifuge, as liquid slowly evaporating from the slides can lead to wash artifacts.

69| Proceed to scan the microarray as soon as possible, as the fluorescence will decay over time.

? TROUBLESHOOTING

The required parameters for data acquisition and image processing depend strongly on the microarray platform and the scanner technology used. In general, similar settings as for acquiring expression profiling data can be used.

Basic data analysis using the TIGR TM4 microarray analysis suite

70| Download and install the following tools from the TIGR TM4 webpage (http://www.tm4.org/): TIGR Midas, TIGR MeV and the ExpressConverter tool available from the utilities menu on the TM4 webpage.

71| First, convert the raw microarray data obtained to TM4's .mev format before the analysis: the ExpressConverter tool can be used to convert Genepix, Imagene, ScanArray, ArrayVision as well as Agilent data files. Specify the input files by choosing File/select input files. If all files contain data generated with the same microarray platform, choosing "many to one" from the "Ann file" menu option can generate a single annotation file. We recommend using median intensity values (the default choice). Press File/Start converting to start the conversion process.

72| After conversion, start the MIDAS module of the TM4 suite to post-process the data. First, specify the input .mev file(s) by choosing a suitable option from the "read data" menu.

73| At each step, the analysis parameters can be set in the main Midas window. To exclude data from those spots rejected by the image analysis software due to bad signal-to-noise ratios or low intensities, tick the four boxes specifying the use of flags as well as background checking and acknowledge the use of the appropriate data column from the .mev file (the default choice).

74| Next, add a "Locfit Normalization (LOESS)" step from the "Operations" menu and leave the default parameters unchanged. This includes a LOESS normalization step in the work-flow, which will adjust the $\log_2$ ratios to zero in an intensity-dependent manner. Setting the "Mode" parameter to "block" (default) instructs Midas to analyze each quadrant of spots deposited by the same print-tip separately, correcting for physical differences introduced during the printing process (for details on choosing suitable normalization techniques for microarray analysis, please see ref. 30).

75| Finally, choose "Write" from the "Write data" menu to select an output directory. We encourage the user to explore the considerable functionality of the Midas program available in the "Operations" menu. Nevertheless, for a basic analysis, these steps (file conversion, specification of input files, flag removal, LOESS normalization, writing of output files) are sufficient.

76| Press the "Execution" button and follow the on-screen prompts to specify the output file and directory.

77| Within the output directory, the normalized data can be identified by the _MDS.mev suffix. To access the graphical output generated by MIDAS at each step, switch from the "Design" to the "Investigation" panel within the main window and browse to the "lowess" folder within the output directory. To plot different graphs, right-click the respective normalized data files (indicated with "G" icons). Choose, for example, a file with a .prc suffix to obtain an "MA-plot" similar to the ones shown in **Figure 3**. Analogous plots for the original, raw data can be found in the "raw" folder.

78| To identify significantly enriched sequences, the MeV module provides a large variety of statistical, analytical and visualization tools. Open the program and start the analysis with the "Load data" option found in the "File" menu. Browse to the folder containing the normalized data files and choose the appropriate files (indicated by the _MDS.mev suffix) by pressing the "add" or "add all" button. Similarly, specify the annotation file, which was automatically generated by ExpressConverter, by pressing the "add" button on the lower half of the screen. This file can be found in the same folder as the raw data files. Finally, start the analysis with the "Load" button at the bottom.

79| The main MeV window is divided into two parts, the vertical analysis tree on the left and the result window on the right. The latter will now display a color-coded display of your data (to change the display, select a suitable color scheme from the "Display/Color Scheme" menu). To simply export the data, including annotations present in the original data file, choose "Save Matrix" from the file menu.

80| To exclude microarray features from the analysis that have not yielded data in a minimum number of hybridizations (e.g., owing to low-intensity signals or mechanical defects of the arrays), select "Adjust Data/Data Filters/Percentage Cutoff Filter" and specify an appropriate percentage. In case of four different samples, for example, entering 50% will ensure that at least two data points are present for the analysis.

81| After this operation, a new "Data filter—Percentage cutoff" entry will be visible in the overview panel on the left. Expand this entry and choose the filtered data set as a new "Data Source" by right-clicking the "Expression Image" icon. To illustrate the statistical analysis, we will now describe the "Significance of Microarrays" approach, available from the "Analysis/Statistics/SAM" menu.

82| The pop-up window allows you to choose a suitable comparison. Choose, for example, "Two-class unpaired" analysis to identify specifically enriched regions from groups of real and mock ChIP-on-chip experiments hybridized against genomic DNA on different microarrays (**Fig. 6a**). (Choose "Two-class paired" when a clear pairwise relationship between a mock and its corresponding ChIP sample exists, for example, if both samples were precipitated from the same starting material. If ChIP and mock experiments were hybridized to a single microarray, choose "One class" instead.)

83| Indicate which samples represent a mock experiment, by placing it into group B.

84| To obtain a *q*-value for each sequence present on the microarray, activate the "Yes (slow!)" button in the center of the window. *q*-Values represent an estimate of the false-discovery rate within the significant population. If, for example, 100 features receive a *q*-value $\leq 5\%$, five false-positive sequences are expected within this set[39].

85| Start the analysis by pressing "OK" and accept the number of possible permutations suggested in the following pop-up window.

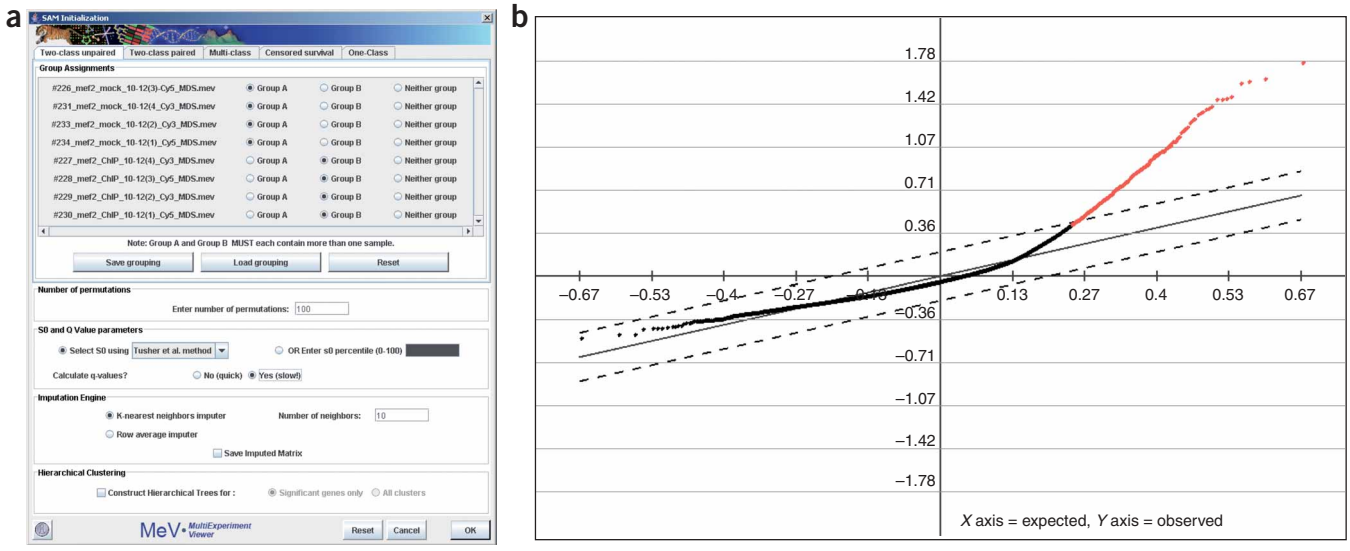

Figure 6 | SAM analysis using TM4's MeV module reveals sequences significantly enriched in anti-Mef2 but not in mock ChIPs. For an unpaired SAM analysis, data from four microarrays hybridized with biologically independent anti-Mef2 ChIP samples are compared with four data sets from mock ChIPs (screenshot from MeV, **a**). The qq-plot obtained from the SAM analysis (**b**) shows a striking asymmetric distribution of features strongly deviating from random values (red spots deviating from the diagonal). As expected for biochemical enrichment, at the chosen cutoff (dashed lines), several hundred features show significant positive deviation ($\log_2 > 0$, red), whereas none are significantly depleted. Final cutoffs for this experiment were chosen on the base of *q*-values combined with a minimum enrichment ratio cutoff.

86| A graphical overview of the results is displayed in an interactive quantile-quantile-plot (qq-plot). Each analyzed sequence feature is represented by a dot. Feature coordinates deviating from the central gray diagonal represent sequences significantly overrepresented in one of the two experimental groups. An example of a qq-plot is shown in **Figure 6b**. Changing the delta value by using the slider below the graph allows the significance cutoff of the experiment to be changed. The number of significant genes as well as the expected median number of false positives at this cutoff is indicated below the graph. Choose a suitable cutoff and press "OK". If *q*-values have been calculated, other cutoffs can be evaluated later. As with any statistical test, there are no general rules how to select a suitable cutoff. In our experience, it is often useful to combine a stringent *q*-value cutoff (e.g., 5%) with a minimum median enrichment ratio to exclude significantly, but very lowly enriched features. We recommend using known positive controls to select suitable cutoffs.

87| Expand the new "SAM" icon in the overview panel on the left. Different graphical representation of genes judged "significant" or "not significant" are available, as are numerical data tables. Right-click and save the data tables for further analysis by choosing "save all clusters."

? TROUBLESHOOTING

● TIMING

The whole procedure can be performed within 10 days.
Day 1: Steps 1–9 (population cages must be seeded at least 1–2 days before. As crosslinked embryos can be stored, several consecutive days of collections can be performed, for example, to acquire repeats or different developmental stages)
Day 2: Steps 10–20 (approximately 3 h)
Day 2: Step 21 (RNase treatment, approximately 1 h; proteinase K treatment, overnight)
Day 3: Step 21 (8 h)
Day 4: Steps 22–29 (approximately 1.5 h, overnight incubation)
Day 5: Steps 30–35 (approximately 6 h, overnight incubation)
Day 6: Steps 36–40 (approximately 8 h)
Day 7: Optional Step 41 (quantitative real-time PCR, approximately 4 h)
Day 7: Steps 42–45 (approximately 1 h, overnight incubation)
Day 8: Steps 46–53 (approximately 5.5 h, overnight incubation)
Day 9: Steps 54–62 (approximately 3 h, overnight incubation)
Day 10: Steps 63–69 (approximately 1 h, depending on the number of microarrays processed)
Day 10: Steps 70–87 (initial data analysis, $\geq$1 h)

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
21	Low DNA yields in chromatin preparations	Incomplete lysis by Dounce homogenization	Try a different Dounce homogenizer and/or increase the number of strokes
		Settings for centrifugation steps incorrect	Check the conversion of r.p.m. to *g*-force for the centrifuge used
21	Sheared chromatin size distribution is too large	Incomplete shearing of the DNA during sonication	Increase the number and/or length of sonication cycles
21	Sheared chromatin size distribution is too small	Excessive shearing of the DNA during sonication	Reduce the number and/or length of sonication cycles
41	No DNA detected in any sample assayed by quantitative real-time PCR Low enrichment ratios for positive controls	Carry over of phenol during the purification of immunoprecipitated DNA	Use phase-lock tubes to reduce phenol carry over
		Suboptimal amount of antiserum/antibody used	Titrate the optimal amount of antiserum/antibody required, while keeping the amount of chromatin constant
		Low affinity of antiserum/antibodies toward protein A	Consult the literature to determine if either protein A (e.g., rabbit, guinea-pig) or protein G (different mouse IgG subtypes, etc.) Sepharose should be used
		The antiserum/antibody used might recognize the antigen with low affinity	Reduce the stringency of the washing steps (Step 32) by reducing the salt concentration of RIPA500. Omit LiCl wash step
		The antiserum/antibody used might not recognize formaldehyde crosslinked proteins or the recognized epitope is not accessible *in vivo*	Choose a different antibody or fuse the epitope tag to the other end of the protein of interest
		The chosen "negative" control region might be nonspecifically enriched in the ChIP	Increase the stringency and/or the number of washing steps (Step 32). Optimize the amount of antibody/antiserum used Always use siliconized tubes for the chromatin immunoprecipitations
		The chosen "negative" control region might be specifically enriched in the ChIP	Choose additional "negative" control regions Repeat the real-time PCR with new negative controls and compare with the previous results
		The chosen positive control may be of low affinity or only be occupied with low frequency	Choose a different positive control region
41/48	Variable enrichment of positive controls	The protein of interest might be present in low amounts or in a small subset of cells	Increase the amount of chromatin used for each precipitation
48	Low enrichment ratios for positive controls only after amplification	Sub-optimal blunting of precipitated DNA, only sub-population was amplified	Repeat Steps 42–44 with fresh reagents
48	"No DNA" amplification yields >10% of the amount of DNA obtained with the experimental samples	Contamination problem during generation of blunt ends, linker ligation and/or amplification reactions	First repeat the second amplification reaction (Step 48) and avoid any crosscontamination. If the problem persists, start over with a fresh blunting reaction (Step 42)
48	Low yields of DNA after amplification of all samples	PCR conditions in your thermocycler might have to be adjusted	Try another thermocycler. Optimize thermocycler program
		Sub-optimal blunting of precipitated DNA, only sub-population was amplified	Repeat Steps 42–44 with fresh reagents
48	Low yields of DNA after amplification of specific samples	Low amount of sample DNA precipitated	Increase the cycle number of the second amplification reaction (Step 48)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
55	Low specific activity after Klenow labeling	Batch to batch variation of Cy dyes	Try a different batch of Cy-dUTPs
			Increase the amount of Cy-dUTP in the reaction
55	Low amount of DNA generated during Klenow labeling	Low quality of template DNA	Increase the amount of template DNA
69	Low intensity of probe signal obtained from the microarray, especially in the Cy5 channel	Bleaching of the Cy dyes	Repeat and protect the dyes from ambient ozone and reduce exposure to light. Scan after the end of hybridization as soon as possible
		Amount of probe used too low for the specific microarray platform	Increase the total amount of probe DNA in the hybridization
87	Few specifically enriched sequences discovered after SAM analysis, but positive controls are recovered	Antiserum/antibodies might precipitate only high-affinity sites	Optimize the precipitation conditions
87	Few specifically enriched sequences discovered after SAM analysis and positive controls are not significant or barely significant	The protein of interest might be difficult to crosslink with chromatin	Evaluate different crosslinking conditions, for example, vary formaldehyde concentration and/or time of fixation
		If more than one antiserum/antibody was used, the quality of the reagents might be too different for a merged SAM analysis	Analyze the data obtained with each antiserum/antibody separately
			Optimize the precipitation conditions for each antiserum/antibody separately Evaluate other statistical tests, especially nonparametrical, rank-based methods[31]

ANTICIPATED RESULTS

Chromatin preparation

The final experimental resolution of any ChIP-on-chip approach depends on the size of sheared DNA fragments as well as the spacing and length of the arrayed probes. The procedure outlined in Step 18 typically yields sequences with a modal distribution centered around 500 bp (**Fig. 4**). Owing to ongoing mitosis during *Drosophila* embryogenesis, the amount of chromatin extracted from the collected embryos using this protocol depends on the developmental stages of interest; generally, later stages yield larger amounts of DNA. After crosslink reversal, we typically recover approximately 160 ng μl^{-1} pure DNA from embryos collected between 6 and 8 h after egg laying (predominantly stages 11 + 12), 220 ng μl^{-1} at 8–10 h (stages 12 + 13) and 250 ng μl^{-1} at 10–12 h (stages 14 + 15). As the complexity of the embryo increases with developmental time, formaldehyde fixation conditions might have to be adjusted; especially at very late stages, crosslinking efficiency might be reduced as a consequence of cuticle formation.

Microarray analysis

Once suitable samples have been identified by quantitative real-time PCR, they are amplified to obtain sufficient amounts of DNA for labeling and hybridization to a genomic tiling array. In this protocol, double-stranded linkers are ligated to both ends of the precipitated samples and subsequently used as priming sites for PCR amplification. Alternatively, priming sites can be introduced by random priming[19] or PCR amplification can be circumvented using an RNA intermediate[40].

The data shown in **Figure 3** are obtained using a single spotted microarray featuring PCR fragments with an average length of 3 kb, together covering approximately 50% of the *Drosophila* genome (additional data sets are available from ArrayExpress under accession code E-TABM-56). Immunoprecipitated ChIP or mock samples were hybridized against (amplified) genomic reference DNA. At 10–12 h of development, several hundred features are enriched by the chromatin immunoprecipitation procedure performed with specific antiserum (cloud of points with high $\log_2$ ratio values M (e.g., >1; **Fig. 3a**, asterisk), as opposed to a control ChIP using the corresponding preimmune serum (**Fig. 3b**).

Depending on the array platform of choice, different statistical analysis methods are available (reviewed in ref. 25). High-density microarrays can tile a genome at a higher resolution but are not available for many model organisms. Specialized analysis strategies take advantage of simultaneous detection of a ChIP-enriched region by multiple probes and draw information

from the positioning of neighboring probes[34]. If the arrayed sequences are relatively long, as in the example presented in this protocol, and enriched sequences are expected to be detected by a single probe, each spot on the array can be analyzed separately, as in traditional expression profiling experiments using cDNA arrays.

SAM analysis, as implemented in the free, platform-independent TM4 (ref. 36) package, has been used to identify features differentially enriched in anti-Mef2 or mock experiments. Four independent ChIP and mock reactions were performed and hybridized against genomic DNA. After normalization, they are grouped together accordingly for the analysis (**Fig. 6a**) for a two-class unpaired test: mock reactions in group A and specific anti-Mef2 data in group B. The resulting qq-plot (**Fig. 6b**) illustrates a large number of specifically enriched sequences (red).

As opposed to an unpaired analysis, a paired test can account, for example, for differences between different chromatin preparations. On the other hand, this approach reduces the number of permutations available for the SAM analysis, limiting statistical power. We therefore encourage the experimenter to analyze the data using both paired and unpaired designs and to evaluate the results using known controls. In the case of Mef2, both paired and unpaired approaches yield similar results.

The identified significantly enriched sequences represent regions in the *Drosophila* genome associated with the factor of interest. Care should be taken, although, to eliminate repetitive regions as well as known transposon-derived sequences from the analysis, before investigating these sequences in their genomic contexts (e.g., using the UCSC genome browser[41], MAMMOT[42] or other visualization tools).

ACKNOWLEDGMENTS We are grateful to the laboratories of Robert White, Steve Russell (Cambridge) and Renato Paro (Basel) for sharing their ChIP (on chip) protocols with us.

COMPETING INTERESTS STATEMENT

Published online at http://www.natureprotocols.com

1. Galas, D.J. & Schmitz, A. DNAse footprinting: a simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Res.* **5**, 3157–3170 (1978).
2. Fried, M. & Crothers, D.M. Equilibria and kinetics of lac repressor–operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505–6525 (1981).
3. Strutt, H. & Paro, R. Mapping DNA target sites of chromatin proteins *in vivo* by formaldehyde crosslinking. *Methods Mol. Biol.* **119**, 455–467 (1999).
4. Toth, J. & Biggin, M.D. The specificity of protein–DNA crosslinking by formaldehyde: *in vitro* and in *Drosophila* embryos. *Nucleic Acids Res.* **28**, e4 (2000).
5. Orlando, V. & Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell* **75**, 1187–1198 (1993).
6. Walter, J. & Biggin, M.D. Measurement of *in vivo* DNA binding by sequence-specific transcription factors using UV cross-linking. *Methods* **11**, 215–224 (1997).
7. Solano, P.J. *et al.* Genome-wide identification of *in vivo Drosophila* Engrailed-binding DNA fragments and related target genes. *Development* **130**, 1243–1254 (2003).
8. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517 (1975).
9. Dorak, M. *Real Time PCR* (Taylor & Francis, Oxford, 2006).
10. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
11. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
12. Kim, T.H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
13. Robert, F. *et al.* Global position and recruitment of HATs and HDACs in the yeast genome. *Mol. Cell* **16**, 199–209 (2004).
14. Alekseyenko, A.A., Larschan, E., Lai, W.R., Park, P.J. & Kuroda, M.I. High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* **20**, 848–857 (2006).
15. Gilfillan, G.D. *et al.* Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.* **20**, 858–870 (2006).
16. Legube, G., McWeeney, S.K., Lercher, M.J. & Akhtar, A. X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in *Drosophila*. *Genes Dev.* **20**, 871–883 (2006).
17. Birch-Machin, I. *et al.* Genomic analysis of heat-shock factor targets in *Drosophila*. *Genome Biol.* **6**, R63 (2005).
18. Sandmann, T. *et al.* A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell* **10**, 797–807 (2006).
19. Schwartz, Y.B. *et al.* Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* **38**, 700–705 (2006).
20. Wei, C.L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
21. Greil, F., Moorman, C. & van Steensel, B. [16] DamID: mapping of *in vivo* protein–genome interactions using tethered DNA adenine methyltransferase. *Methods Enzymol.* **410**, 342–359 (2006).
22. Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **103**, 12027–12032 (2006).
23. Bulyk, M.L. DNA microarray technologies for measuring protein–DNA interactions. *Curr. Opin. Biotechnol.* **17**, 422–430 (2006).
24. Orian, A. Chromatin profiling, DamID and the emerging landscape of gene expression. *Curr. Opin. Genet. Dev.* **16**, 157 (2006).
25. Buck, M.J. & Lieb, J.D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
26. Lee, T.I., Johnstone, S.E. & Young, R.A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protocols* **1**, 729 (2006).
27. Hanlon, S.E. & Lieb, J.D. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr. Opin. Genet. Dev.* **14**, 697–705 (2004).
28. Simon, R., Radmacher, M.D. & Dobbin, K. Design of studies using DNA microarrays. *Genet. Epidemiol.* **23**, 21–36 (2002).
29. Dobbin, K., Shih, J.H. & Simon, R. Statistical design of reverse dye microarrays. *Bioinformatics* **19**, 803–810 (2003).
30. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **32** (suppl.): 496–501 (2002).
31. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **573**, 83–92 (2004).
32. Buck, M.J., Nobel, A.B. & Lieb, J.D. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Genome Biol.* **6**, R97 (2005).
33. Gibbons, F.D., Proft, M., Struhl, K. & Roth, F.P. Chipper: discovering transcription-factor targets from chromatin immunoprecipitation microarrays using variance stabilization. *Genome Biol.* **6**, R96 (2005).
34. Ji, H. & Wong, W.H. TileMap create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629–3636 (2005).
35. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
36. Saeed, A.I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
37. Sullivan, M., Ashburner, R. & Hawley, S. *Drosophila Protocols* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000).

38. Strutt, H., Cavalli, G. & Paro, R. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* **16**, 3621–3632 (1997).
39. Campos-Ortega, J.A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, Heidelberg, 1985).
40. Liu, C.L., Schreiber, S.L. & Bernstein, B.E. Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics* **4**, 19 (2003).
41. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res.* **31**, 51–54 (2003).
42. Ryder, E., Jackson, R., Ferguson-Smith, A. & Russell, S. MAMMOT—a set of tools for the design, management and visualization of genomic tiling arrays. *Bioinformatics* **22**, 883–884 (2006).
43. Kelly, K.K., Meadows, S.M. & Cripps, R.M. *Drosophila* MEF2 is a direct regulator of Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. *Mech. Dev.* **110**, 39–50 (2002).