Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W., Mitsakakis, N., Ryan, O. W., Datta, N., Jojic, V., Pal, C., Canadien, V., Richards, D., Beattie, B., Wu, L. F., Altschuler, S. J., Roweis, S., Frey, B. J., Emili, A., Greenblatt, J. F., and Hughes, T. R. (2003). A panoramic view of yeast noncoding RNA processing. *Cell* **113,** 919–933.

Shai, O., Morris, Q. D., Blencowe, B. J., and Frey, B. J. (2006). Inferring global levels of alternative splicing isoforms using a generative model of microarray date. *Bioinformatics* **22,** 606–613.

Shoemaker, D. D., Schadt, E. E., Armour, C. D., He, Y. D., Garrett-Engele, P., McDonagh, P. D., Loerch, P. M., Leonardson, A., Lum, P. Y., Cavet, G., Wu, L. F., Altschuler, S. J., Edwards, S., King, J., Tsang, J. S., Schimmack, G., Schelter, J. M., Koch, J., Ziman, M., Marton, M. J., Li, B., Cundiff, P., Ward, T., Castle, J., Krolewski, M., Meyer, M. R., Mao, M., Burchard, J., Kidd, M. J., Dai, H., Phillips, J. W., Linsley, P. S., Stoughton, R., Scherer, S., and Boguski, M. S. (2001). Experimental annotation of the human genome using microarray technology. *Nature* **409,** 922–927.

Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.* **24,** 4501–4505.

Wang, H., Hubbell, E., Hu, J. S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M. A., Ares, M., Kulp, D. C., and Haussler, D. (2003). Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* **19**(Suppl. 1), i315–i322.

Wiegant, J. C., van Gijlswijk, R. P., Heetebrij, R. J., Bezrookove, V., Raap, A. K., and Tanke, H. J. (1999). ULS: A versatile method of labeling nucleic acids for FISH based on a monofunctional reaction of cisplatin derivatives with guanine moieties. *Cytogenet. Cell Genet.* **87,** 47–52.

Wolber, P. K., Collins, P. J., Lucas, A. B., De Witte, A., and Shannon, K. W. (2006). The Agilent *in situ*-synthesized microarray platform. *Methods Enzymol.* **410** (this volume), 28–57.

Xing, F., Hiley, S. L., Hughes, T. R., and Phizicky, E. M. (2004). The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.* **279,** 17850–17860.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30,** e15.

Yeakley, J. M., Fan, J. B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S., and Fu, X. D. (2002). Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20,** 353–358.

[15] Mapping the Distribution of Chromatin Proteins by ChIP on Chip

By NICOLAS NÈGRE, SERGEY LAVROV, JÉRÔME HENNETIN, MICHEL BELLIS, and GIACOMO CAVALLI

Abstract

The ChIP on chip method combines chromatin immunoprecipitation (ChIP) with hybridization on DNA microarrays (chip). The ChIP technique allows one to obtain a DNA sample enriched in sequences

METHODS IN ENZYMOLOGY, VOL. 410 0076-6879/06 $35.00
Copyright 2006, Elsevier Inc. All rights reserved.
DOI: 10.1016/S0076-6879(06)10015-4

bound by transcription factors or chromatin-associated proteins. Usually, ChIP is used to test whether specific candidate sequences are bound by a transcription factor, but microarrays are a powerful tool that allows testing large pools of sequences at once. This chapter presents the pipeline of a ChIP on chip method that can be applied to map the binding sites of chromatin-associated proteins along *Drosophila* chromosomes at different developmental stages. This chapter provides protocols for ChIP, for quality control tests of ChIP samples, for microarray design, for hybridization of the ChIP samples onto microarrays, and for initial analysis of the data. In addition, this chapter discusses the most important steps in each of the protocols as well as the importance of bioinformatic analysis in order to extract valuable biological information from the data sets.

Introduction

The genomes of many model organisms have been completely sequenced since the beginning of the year 2000, including the human genome (Lander *et al.*, 2001) and the genome of the fly *Drosophila melanogaster* (Adams *et al.,* 2000). The postgenomic era coincides with an extensive use of microarray techniques that were initially described in 1995 for the study of gene expression in *Arabidopsis thaliana* (Schena *et al.,* 1995). There are three major types of microarray platforms available for most of the organisms studied, based on the depositing of cDNA amplicons, of oligonucleotides, and on polymerase chain reaction (PCR) amplicons (Gupta and Oliver 2003). In *Drosophila*, most of these microarrays have been used to study whole genome expression profiles by hybridizing mRNA isolated from different developmental stages (Arbeitman *et al.,* 2002; Stolc *et al.,* 2004) or mutant versus wild-type flies.

Complementary information that is needed to understand how genes are regulated at a genomic scale comes from the identification of transcription factor-binding sites. Several regulatory pathways involving transcription factors are well described genetically and, in the best cases, the *trans*-acting regulatory factors and their binding sites on the regulatory sequences of individual target genes have been identified. What is not known are the global features of gene networks involving these factors. The development of microarray technologies allows one to extend the analysis from individual genes to the whole genome and to reach an integrated understanding of gene regulatory phenomena. To this aim, one useful approach, originally developed in yeast, is the so-called ChIP

on chip technique (Ren *et al.,* 2000), which combines ChIP (Orlando *et al.,* 1997) with hybridization of the ChIP samples onto microarrays.

This chapter describes a pipeline for application of the ChIP on chip method to the study of the distribution of chromatin-associated proteins along *Drosophila* chromosomes. Because this pipeline has no steps that are specific to the *Drosophila* system, except for the genomic sequence used for the design of tiling-path microarrays, it should be adaptable to ChIP samples coming from virtually any kind of biological material and organism. Our laboratory is particularly interested in understanding the role of the Polycomb (PcG) and trithorax-group (trxG) proteins in the regulation of gene expression during development (Ringrose and Paro, 2004). These proteins are well-known regulators of the patterned expression of homeotic genes. PcG proteins maintain the repressed state of homeotic genes in the segments where they were initially repressed by early acting transcription factors, whereas trxG proteins maintain active states in the appropriate regions of the body. PcG and trxG proteins act as multimeric complexes that associate to chromatin through regulatory sequences named PREs and TREs for Polycomb and trithorax response elements (Simon *et al.,* 1993), but most of these factors do not bind directly to DNA. Instead, they are recruited at target genes by other proteins and by specific histone modification marks. This is the case of the proteins forming the Polycomb Repressive Complex 1 (PRC1) (Francis *et al.,* 2001), which are recruited to PREs through the recognition of trimethylation of lysine 27 of histone H3 via the chromodomain that is present in the Polycomb (PC) protein (Min *et al.,* 2003). Although PcG and trxG proteins bind more than 100 loci on polytene chromosomes, based on immunostaining assays (Franke *et al.,* 1992), only 11 PREs have been described molecularly to date, and most of them are located in the homeotic gene complexes Bithorax and Antennapedia. Based on their sequence, no extensive homology has been identified between these PREs, and moreover their size varies, as well as their distance to target genes. This is a typical example of the type of problems encountered during the study of gene regulatory proteins. Few target genes are known, and thus their mutual relations as well as their organization along chromosomes are a mystery. Moreover, because the *in vivo* target sequences for the factors of interest cannot be predicted reliably, bioinformatic studies cannot be used in order to identify the chromosomal distribution of these factors. Therefore, ChIP on chip is the most powerful approach toward reaching this goal. For instance, this approach has been used successfully for mapping PcG protein binding to a large number of genomic sites in mouse tumor cell lines (Kirmizis *et al.,* 2004).

Chromatin Immunoprecipitation

In most ChIP applications, chemical reagents such as formaldehyde are used to cross-link proteins to DNA in a covalent manner. A specific antibody is then used to immunoprecipitate the protein of interest. The cross-linking step allows one to coimmunoprecipitate the DNA fragments bound to this protein. The final product of this assay is a pool of genomic DNA sequences, usually ranging from 200 bp to 1 kb. Ideally, only sequences specifically cross-linked to the protein of interest should be recovered after immunoprecipitation (IP), but in practice any genomic DNA sequence binds with low affinity to the beads used to recover the immunoprecipitated material. The background DNA sticking to the beads in ChIP experiments is controlled for by performing a control sample (named ''mock'') where no antibody is added during the IP. The specific protein targets are enriched above this background in a good ChIP sample, while the mock IP shows the same background without specific enrichments. For this reason, detection of the enriched fragments is done by comparing the IP sample with the mock sample.

Multiple agents can be used for cross-linking, such as UV (Zhang *et al.,* 2004) or methylene blue (Liu *et al.,* 2000). The most commonly used is formaldehyde (HCHO), a chemical that induces protein–protein and protein–DNA cross-links, and is particularly convenient as it allows studying not only DNA-binding proteins, but also proteins that do not bind DNA directly but associate to chromatin via other proteins.

A critical issue concerning the use of formaldehyde is the accessibility of the biological material to this agent. *Drosophila* ChIP protocols have been originally developed for cultured cells or dechorionated *Drosophila* embryos that are permeable to the formaldehyde solution. Because the other developmental stages of *Drosophila* are characterized by the presence of an impermeable cuticle, we developed a ChIP protocol that can use any tissue by directly crushing the material in the presence of formaldehyde.

Solutions and Materials

- Formaldehyde 37%
- Glycine solution 2.5 *M*
- 10% *N*-lauroylsarcosine
- Buffer A1: 60 m*M* KCl, 15 m*M* NaCl, 4 m*M* $MgCl_2$, 15 m*M* HEPES (pH 7.6), 0.5% Triton X-100, 0.5 m*M* dithiothreitol (DTT), protease inhibitors cocktail (complete, EDTA free, Roche, use following manufacturer's instructions)
- Lysis buffer: 140 m*M* NaCl, 15 m*M* HEPES (pH 7.6), 1 m*M* EDTA, 0.5 m*M* EGTA, 1% Triton X-100, 0.5 m*M* DTT, 0.1% sodium deoxycholate, 0.05% SDS, protease inhibitor cocktail

- Centricon columns (YM-100, cutoff 100 kDa from Amicon): Should be blocked by bovine serum albumin (BSA) (1 mg/ml) in phosphate-buffered saline (PBS) and then washed with PBS only before use
- Branson sonifier 250, equipped with a microtip of 5 mm diameter
- Sodium azide
- 5 mg/ml RNase A (DNase free)
- Protein A-Sepharose (PAS) suspension: 100 mg of CL-4B (Amersham, 17–0780–01) PAS should be resuspended in 1 ml of lysis buffer + 0.1 mg/ml BSA for 50% suspension. Wash in lysis buffer two to three times and equilibrate in lysis buffer for 1 h at 4° on a rotating wheel. Store up to 1 week at 4°.
- TE: 10 m*M* Tris–HCl (pH 8.0), 1 m*M* EDTA
- Elution buffer 1: 10 m*M* EDTA, 1% SDS, 50 m*M* Tris–HCl (pH 8)
- Elution buffer 2: 0.67% SDS in TE
- Proteinase K solution (250 μl): 0.5 μl of 20 mg/ml glycogen solution, 5 μl of 20 mg/ml proteinase K stock, 244.5 μl TE
- 4 *M* lithium chloride
- Polynucleotide kinase (Promega) 10,000 units/ml
- Klenow fragment polymerase (Promega) 5000 units/ml
- T4 DNA ligase (Promega) 400,000 units/ml with its supplied buffer
- 10 μM linker DNA: two oligonucleotides must be annealed: (i) a 24-mer of sequence 5′-AGA AGC TTG AAT TCG AGC AGT CAG (phosphorylated at 5′ end); (ii) a 20-mer of sequence 5′-CTG CTC GAA TTC AAG CTT CT. Store oligonucleotides in small aliquots at –20°. To produce the linker, mix 20 μl of 100 μM 24-mer phosphorylated primer and 20 μl of 100 μM 20-mer primer in 160 μl of TE. Incubate in a PCR machine using the following program: 5 min at 70° (remove secondary structures), 5 min at 55° (annealing). Let cool down slowly (0.01°/s) to 25°, incubate 2 h at this temperature and then cool down (0.01°/s) to 4°. Aliquot the 10 μM linker and store at –20°.
- 2 m*M* dNTP mix: prepare a mix with 20 μl of each dNTP (Promega, PCR grade, 100 m*M*) in 920 μl H_2O. Make 10× 100-μl aliquots and store at –20°.
- *Taq* polymerase (Promega) 5000 units/ml
- QIAquick PCR purification kit (Qiagen)

Formaldehyde Cross-Linking of Chromatin from Any Drosophila *Tissue*

1. Homogenize the material (about 150 to 200 mg of dried biological material is sufficient for four to five independent immunoprecipitations) in 5 ml of buffer A1 + formaldehyde at 1.8% final concentration (290 μl of 37% solution) at room temperature using first a Potter homogenizer and then a Dounce with type A pestle (three strokes). Wait 15 min (total time

starting from beginning of homogenization). Add glycine solution to 225 m*M* (540 μl of 2.5 *M* solution for 6 ml of cross-linked mixture), mix, and incubate 5 min. Put on ice.

2. Transfer the homogenate into a 15-ml tube. Centrifuge 5 min at 4000*g* at 4°. Discard the supernatant. Add 3 ml of buffer A1, resuspend the pellet, and spin down the same way. Repeat the washing step three times.

3. Wash once in 3 ml of lysis buffer without SDS. Spin down 5 min at 4000*g*.

4. Resuspend the cross-linked material in 0.5 ml of lysis buffer + SDS to 0.1% and *N*-lauroylsarcosine to 0.5%. Incubate 10 min at 4° in a rotating wheel.

5. Sonicate the chromatin to make it soluble. Parameters for the Branson sonifier are as follow: power 2, duty cycle 100%, four times 30 s with 2-min intervals. Sonication must be made in a conical-shaped tube with the tip of the sonifier going just at the limit between the cylindrical part and the conical bottom. Place the tube on melting ice during sonication to avoid excess heating of the chromatin.

6. Rotate 10 min on a rotating wheel at 4°. Transfer into Eppendorf tubes and centrifuge 5 min at room temperature at maximum speed. Transfer supernatant to a new tube. Add another 0.5 ml of lysis buffer to the pellet and rotate for 10 min. Repeat centrifugation and combine the supernatants. Centrifuge the combined supernatants 2× 10 min at maximum speed. The supernatant from this stage is the chromatin extract. Cross-linked chromatin can be stored at –80° for several months at this stage. Add sodium azide to 0.02% for storage.

7. Put the chromatin extract in a Centricon column, centrifuge 3× 40 min (or more) at 1000 *g* while adding lysis buffer. At least 3 volumes of lysis buffer should pass through the column. Bring the final volume of chromatin extract to 1 ml with lysis buffer.

Chromatin Immunoprecipitation Procedure

1. To an amount of chromatin corresponding to 150 mg of biological material, suspended in a final volume of 1 ml (in lysis buffer), add 100 μl of PAS suspension for preincubation. Incubate several hours or overnight at 4° and then remove PAS. Cross-linked chromatin at this stage can be stored several days at 4° or frozen at –70°.

2. Check DNA recovery as follows. From the 1-ml solution described earlier, take a 100-μl aliquot, add proteinase K up to 100 μg/ml and SDS to 1%, incubate 6 h at 60°, then 20 min at 70°, add RNase A to 50 μg/ml, and incubate for an additional 2 h at 37°. Extract the DNA with phenol-chloroform and precipitate with ethanol. After resuspension, run the sample on an agarose gel to check amount and size of DNA.

3. Separate the chromatin sample into 4× 250-μl aliquots (one aliquot is sufficient for one IP). Immunoprecipitate the chromatin by adding the antibody (Ab) of interest. The amount of Ab should be determined empirically. In the case of affipure Abs, it might be in the order of 2–5 μg. If the concentration is not known, one should use the same concentration as used in regular IP experiments using soluble protein extracts. Do not forget to perform a control IP without Ab, called ''mock.'' For microarray hybridization, one mock should be done for each different antibody, as the hybridization requires a large amount of material.

4. Incubate 4 h at 4° on a rotating wheel, add 50 μl of PAS suspension, and incubate 4 h or overnight. Spin down PAS and proceed to washes.

5. Wash PAS 4× with lysis buffer, followed by 2× with TE (without protease inhibitors). Each wash is for 5 min at 4°, using 1 ml of solution.

6. For elution of precipitated material, spin down PAS. Add 100 μl of elution buffer 1, mix, and incubate 10 min at 65°. Spin down PAS and transfer supernatant to a new tube. Add 150 μl of elution buffer 2 to PAS, mix, centrifuge at full speed, and transfer the eluate to a tube together with the eluate from the first centrifugation. The combined material is the ''chromatin precipitate'' (approximately 250 μl).

7. Incubate the chromatin precipitate 6 h (or overnight) at 65° to reverse cross-links. Add 250 μl of proteinase K solution and incubate at 50° for 2 to 3 h.

8. Add 55 μl of 4 *M* LiCl and 500 μl of phenol-chloroform. Mix and centrifuge at full speed at room temperature. Transfer the aqueous phase to a new tube and precipitate with 1 ml of cold 100% ethanol. Wash with 750 μl of 70% ethanol. Spin down and dry the precipitate.

9. Dissolve in 25 μl of water. This is the ''native'' ChIP sample. At this stage, one can quantify the amount of DNA on a slot blot hybridized with genomic DNA, including a standard curve with known amounts of genomic DNA in the blot. One should normally obtain around several nanograms in total from these samples (see Fig. 1A). Microarray hybridizations require 1 μg of DNA per slide; therefore an amplification step by blunt-end linker-mediated PCR (LM-PCR) is necessary.

10. Add 3 μl of T4 ligase buffer and 1 μl (10 U) of polynucleotide kinase (PNK). Incubate at 37° for 30 min.

11. Inactivate PNK at 68° for 20 min. Cool down to 37°. Repair staggered DNA ends by adding 1 μl of 2 m*M* dNTP mix and 1 μl of Klenow fragment (5 U) and incubating for 30 min at 37°. Inactivate at 75° for 10 min and cool to 4°.

12. Take 9 μl of the reaction mixture (store the remaining amount at –20° for further experiments) and add 1 μl of 10 m*M* ATP, 1 μl of 1 μ*M* linker, and 4 U of T4 DNA ligase. Incubate overnight at 4°.

13. Use the ligation mixture directly for PCR amplification [add to tube with ligation reaction 44 μl of bidistilled H_2O (PCR grade), 8 μl of *Taq*

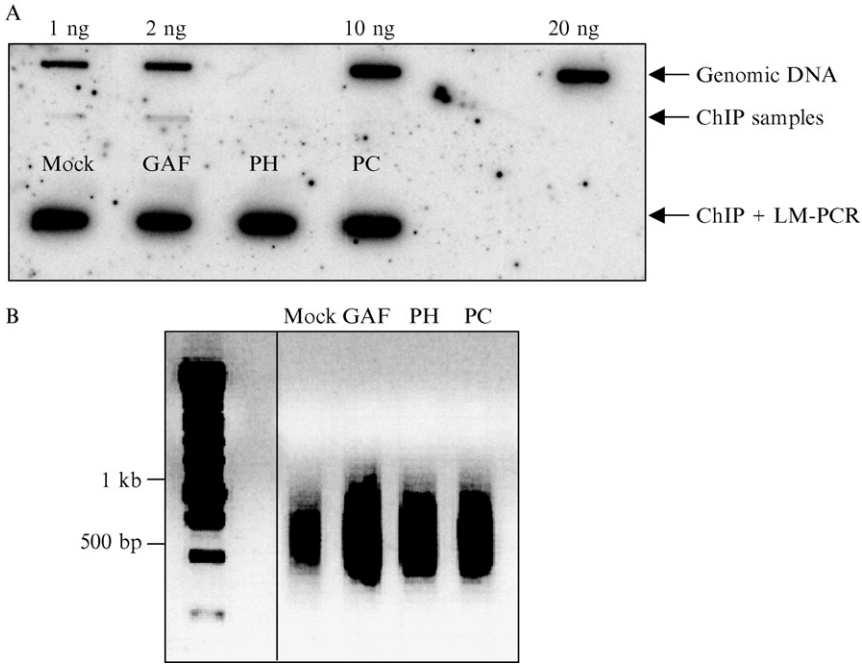

FIG. 1. Quality of ChIP samples. (A) Quantification of ChIP samples by slot blot experiment. The upper lane corresponds to a standard curve of Drosophila genomic DNA. One microliter of the ChIP samples before and after PCR amplification was deposited below. The amount of DNA present in ChIP samples before PCR amplification is 0.7 ng/μl for mock, 1.1 ng/μl for GAF, and 0.3 ng/μl for both PH and PC (the total volume of the samples being 25 μl). After PCR amplification, the amount of DNA is homogeneous in all ChIP samples and is approximately 50 ng/μl (total volume 100 μl). (B) An example of the distribution of DNA fragments obtained after PCR amplification of ChIP samples on a 1% agarose gel. The size of the fragments ranges from 200 bp to 1 kb and is centered on 500 bp.

polymerase buffer, 10 μl of 2 m*M* dNTP mix, 5 μl of 25 m*M* solution of $MgCl_2$, 0.5 μl *Taq* polymerase, 1.5 μl of 20-mer primer]. PCR amplification is as follows: 2 min at 94° (1 cycle), 34 cycles of 1 min at 94°/1 min at 55°/3 min at 72°, and 1 min at 94°/1 min at 55°/10 min at 72° (1 cycle).

14. Purify PCR products on Qiagen QIAquick columns (following manufacturer's instructions). The final elution of the DNA is in 100 μl of Qiagen "EB buffer" (10 m*M* Tris–HCl, pH 8.5).

Quality Control of ChIP Samples

After obtaining the ChIP sample and before using it for microarray hybridizations, several quality controls should be done.

1. Determine the concentration of the eluted DNA by a spectrophotometer (OD_{260}).
2. Check the size and yield of DNA further by agarose gel electrophoresis (see also Fig. 1B).
3. Finally, measure the DNA amount carefully by slot blot as follows: load a standard concentration curve of *Drosophila* genomic DNA of 1, 2, 10, and 20 ng onto a slot blot apparatus using a GeneScreen Plus nylon membrane (NEN). In addition, load approximately 1 to 4 ng of the DNA samples onto the slot blot. Estimate these amounts from the OD measure and agarose gel. Note that, in general, samples out of the QIAquick columns are approximately 50 ng/μl, so they should be diluted appropriately. Hybridize these samples using 50 ng of radiolabeled *Drosophila* genomic DNA and measure the signals by a phosphorimager in order to evaluate DNA yields precisely (Fig. 1A).

Evaluation of ChIP Specificity by Southern Blot

Before applying the ChIP sample on microarrays, a good evaluation of ChIP quality is needed. In our case, we utilize well-characterized PREs that represent positive controls. These DNA fragments are digested by appropriate restriction enzymes and are loaded onto agarose gels that are hybridized with the DNA from the ChIP. Briefly, the procedure is as follows.

1. Run 1 μg of the different restriction digested plasmids (containing the fragments of interest) on 1% agarose gels in duplicate.
2. Transfer the gels to two nylon membranes (NEN membranes, GeneScreen +) by Southern blot overnight using 10× SSC. Then, cross-link the membranes with UV or by incubation for 1 h at 80° in a hybridization oven.
3. Label 50 ng of both mock and ChIP samples separately by incorporating ^{32}P-labeled dCTP. We typically perform random priming reactions with a PrimeaGene kit (Promega).
4. Hybridize the two membranes with the mock or the ChIP sample overnight at 65°.
5. Wash the membranes 4× 10 min in 0.1% SDS, 2× SSC at 65°.
6. Finally, expose the hybridized membranes 1 to 5 days in a Storm Phosphorimager cassette (Molecular Dynamics). Scan the cassette with the Storm Phosphorimager and quantify the signals with the ImageQuant software. For each fragment of the digested plasmid, the ''fold change'' is calculated as a ratio between the ChIP and the Mock samples. One example of such a test is shown in Fig. 2.

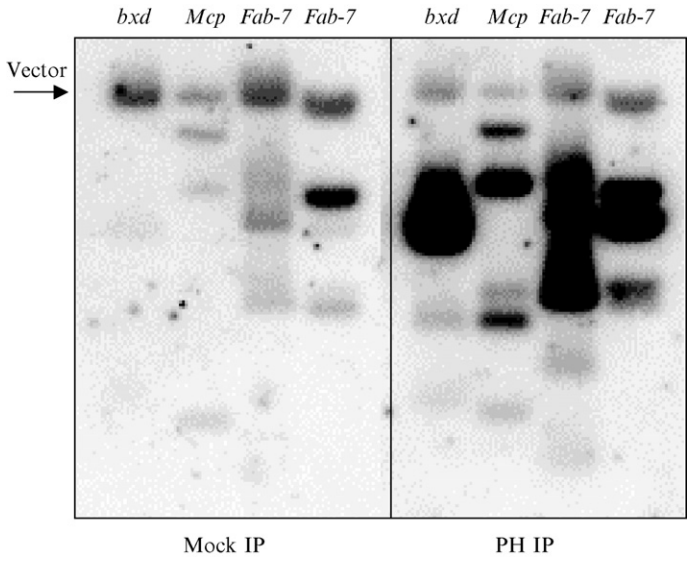

FIG. 2. Specificity of ChIP samples by Southern blot. Known PREs from the *Bithorax-Complex* (BX-C) region (*Fab-7, Mcp,* and *bxd*) were used as positive controls to test the quality of ChIP. Digested plasmids were migrated on a 1% agarose gel and then transferred onto a nylon membrane by Southern blot. One membrane is hybridized with the mock sample (left). Several fragments are enriched in the PH IP sample (right), as expected from previous analysis. One fragment also hybridizes strongly in the mock sample, illustrating that some of the fragments are overrepresented upon blunt-end ligation-mediated PCR.

Chips for ChIPs

When mapping binding sites of gene regulatory factors, cDNA arrays or arrays containing oligonucleotides spanning the coding regions of genes (typically used for transcriptome studies) are not convenient, as the binding sites of these factors are in regulatory regions that are often located far away from the coding part of the gene. Ideally, one should put the whole euchromatic genome on microarrays. This requires building tiling path arrays, where the features printed on the glass slides can be oligonucleotides or PCR fragments continuously covering large genomic regions. We assembled such a tiling path to cover 7 Mb of *Drosophila* X-chromosome euchromatin by PCR products.

Microarray Design and Production of PCR Amplicons

PCR amplicons were designed based on release 2 of the *Drosophila* genome sequence. We chose a 1.9-kb average amplicon size in order

to cover a large genomic area and to keep the cost at a reasonable level. For the design, we segmented the genome sequence in fragments within which the oligonucleotide design was made. For segmentation, we applied the following set of rules. Adjacent sequences overlap on a window of 100 bp, and the oligonucleotide design for the reverse oligonucleotide of the first amplicon and the forward oligonucleotide of the second fragment is made in this overlapping window. Adjacent fragments have been designed in such a way to have different sizes of 1.7, 1.9, and 2.1 kb, respectively. This allows checking the specificity of the PCR products by visual inspection of their migration in agarose gels (Fig. 3). Finally, repeat sequences were not masked, and retrospective data mining revealed that they did not perturb the obtention of profiles. However, it must be noted

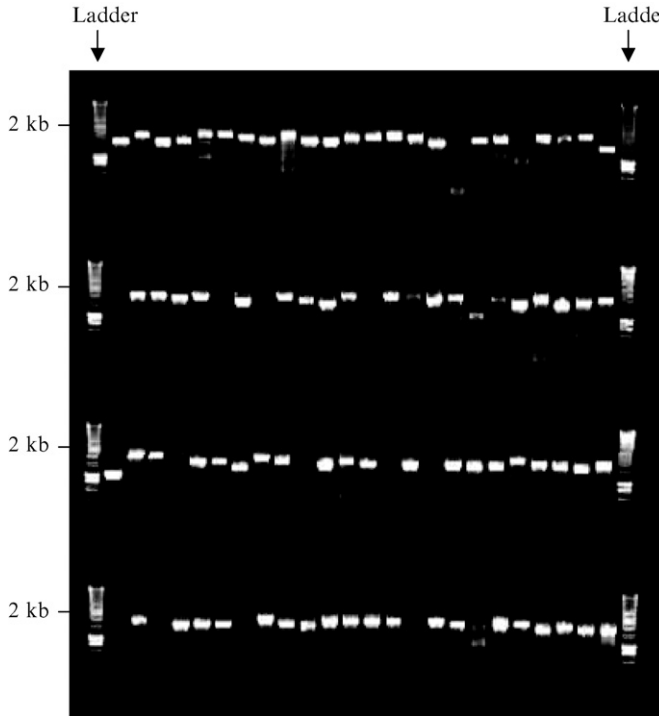

FIG. 3. Quality of PCR products deposited on tiling path arrays. An example of a 1% agarose gel loaded with 96 PCR products obtained from *Drosophila* genomic DNA is shown. In most of the cases, the size and yield of the amplicons are correct. Amplification of the fragments giving a low amount of DNA, a wrong size, or a band that is not unique is repeated under modified PCR conditions in order to recover them. If the rework fails, these amplicons are excluded from the microarrays.

here that the proteins studied do not generally bind to repeat sequences. It might be wise to separate repeats from unique sequences in the case of proteins potentially binding to DNA repeats. The quality of each PCR product was checked systematically on agarose gels for yield, size, and specificity (see Fig. 3). The percentage of success was about 84%. A total of 4153 fragments were thus obtained. In addition, the specificity of the amplicons was verified by sequencing several randomly chosen PCR fragments.

This PCR-based approach allows tiling across large genomic regions without missing any sequences and, as such, allows one to describe the distribution profile of the protein of interest. However, it involves a large amount of work, and the size of the amplicons determines the resolution of data. Although we designed amplicons of roughly 2 kb, the maximal resolution of the ChIP technique is 200 to 500 bp. Therefore, reducing the amplicon size would improve data quality, but the trade-off is a greater amount of work. Furthermore, this approach requires one to have access to a facility that can array the amplicons and print custom slides. Finally, one disadvantage regarding this PCR-based design is the heterogeneity between the fragments in terms of size, melting temperature, and CG content. This results in some degree of variability in the efficiency of hybridization of different amplicons and requires the use of good normalization and calibration tools (see later). Thus, this approach is very well suited for routine systematic analysis of large regions of interest, but extending it to the whole genome is a challenging endeavor that is better suited for a community effort than for a single laboratory.

An alternative to this approach is to use oligonucleotide tiling path arrays. The current options are short oligonucleotides (from Affymetrix) in very high-density microarrays or longer oligonucleotides that can be spotted at relatively high densities (>100,000 per slide) by maskless photolithography (Kirmizis *et al.,* 2004; Stolc *et al.,* 2004). The first option allows for whole-genome tiling, but the use of short oligonucleotides might be a source of specificity problems in some genomic regions. The second option warrants better specificity, but requires the design of multiple microarrays for whole-genome tiling, multiplying the amount of ChIP material required for one experiment. In addition, tiling path oligonucleotide arrays are expensive. In summary, both alternatives have advantages and drawbacks, and one should evaluate carefully which approach is best suited for the project.

Printing and Processing Microarrays Prior to Hybridization

This step strongly depends on the microarray equipment available. In our case, printing is done with two different setups, both giving satisfactory results. PCR products are deposited into 384-well plates and dried. Then,

the PCR products are resuspended into 23 μl of 3$\times$ SSC or in 50% dimethyl sulfoxide using a robot. The plates are allowed to resuspend during 1 week in order to obtain a homogeneous solution before printing. The printing of microarrays is done by using (i) an Omnigrid arrayer (GeneMachines) equipped with 16 steel pins that fit the 384-well plates used to print poly-L-lysine-coated glass slides or (ii) a Lucidea array spotter (Amersham) with a 24-pen print head used to print amino-silane-coated microarray glass slides. Because most experiments are performed with slides printed with the first setup, the following protocols correspond to this type of slides. After the printing step, slides are stored inside a slide box and put into a dry chamber filled with drierite. Before using the slides, they should be processed as follows.

Solutions and Materials

- Diamond pen
- Humid chambers suitable for incubating regular glass slides (Sigma)
- Metal slide racks (VWR) and glass chambers where the metal tray can fit (VWR)
- Succinic anhydride (Sigma)
- 1,2-Dichloroethane (DCE; Acros Organics)
- 1-Methylimidazole (Fluka)
- 95% ethanol
- Drierite (Sigma)
- Slide boxes and dry chamber

Procedure

1. Mark the boundaries of each array on the back (using a diamond pen) to mark the area for deposition of the coverslip during hybridization (see later). Also, label the date and ID of each array at the side of the slide.
2. Fill the bottom of a humid chamber with the maximum volume of 4$\times$ SSC in such a way that slides do not touch the liquid once deposited in the chamber.
3. Put the humid chamber under a binocular and place a lamp on it.
4. Place arrays face down over 4$\times$ SSC and cover the chamber with its lid. This procedure will rehydrate the arrays.
5. Allow the slides to rehydrate for 5 to 15 min. With the help of the binocular, check that each spot has grown to its maximum size and is homogeneous in shape. Check also that the spots do not touch.
6. Once rehydrated, snap dry each array (face up) on an 80° inverted heat block for 3 s. Stop immediately when you see a heat wave crossing the slide.

7. Place the arrays in a metal slide rack.

8. Prepare the blocking solution. Dissolve 1.9 g of succinic anhydride in 380 ml of DCE. The solution should appear turbid.

9. Add 4.75 μl of methylimidazole to the solution, which will become immediately clear.

Transfer the clear solution to a glass chamber immediately and plunge the slide rack containing the processed arrays in this solution.

10. Shake for 1 h at room temperature on an orbital shaker. During this time, prepare 2 liters of boiling distilled water in a 4-liter beaker.

11. Discard the blocking solution appropriately (chemical waste), and fill the chamber with DCE. Wash slides for 1 min. Again, discard the DCE in the chemical waste.

12. Gently place the metallic slide rack into the boiling water and shake slowly for 2 min.

13. Remove the rack and place it into a glass chamber containing 95% ethanol. Repeat five times for 30 s each and then place the rack into a chamber containing distilled water.

14. Bring the chamber to the centrifuge and spin the rack for 5 min at 1000 rpm.

15. The slides are ready to use and can be stored in a slide box placed in a dry chamber.

Labeling and Hybridization

One microgram of the mock IP and 1 μg of the IP samples should be labeled for hybridization. Labeling consists of random priming by insertion of fluorophore-coupled nucleotides, such as Cy3 and Cy5 dCTPs. It is essential to perform labeling and double hybridization in the two color channels at the same time, as labeling and hybridization are a major source of variability (see later). Labeling and hybridization are performed as follows.

Solutions and Materials

- Bioprime DNA labeling kit: for random-priming labeling of ChIP samples do not use the dNTP mix provided in the kit
- 10× dNTP mix: prepare using PCR grade dNTPs: 1.2 m*M* each dATP, dGTP, and dTTP, 0.8 m*M* dCTP, diluted in TE
- Cy5-dCTP and Cy3-dCTP (Amersham, 1 m*M* stocks)
- Microcon YM-30 filter (Amicon/Millipore)
- Yeast tRNA (Invitrogen; make a 5-mg/ml stock)
- Poly(dA-dT) (Sigma; make a 1-mg/ml stock)
- TE: 10 m*M* Tris–HCl (pH 8.0), 1 m*M* EDTA

- 20× SSC
- SDS 20%
- 24 × 30-mm coverslips
- Hybridization chambers (Proteomic Solutions)
- Glass chambers and their adapted glass slide racks for the SSC washes
- Wash IA: 1.14× SSC; 0.0285% SDS
- Wash IB: 1.14× SSC
- Wash II: 0.228× SSC
- Wash III: 0.057× SSC
- Centrifuge adapted for slide racks (e.g., Jouan CR412 tabletop centrifuge with swinging rotor and adapters for 96-well plates)
- Genepix scanner 4000B (Axon Instruments)

Procedure

1. Add 1 μg DNA of the sample to be labeled into a single PCR tube (no strips).

2. Add bidistilled H_2O to bring the total volume to 21 μl. Add 20 μl of 2.5× random primer/reaction buffer mix (from the Bioprime kit). Boil 5 min (in a PCR machine) and then place on ice.

3. On ice, add 5 μl of the10× dNTP mix.

4. Add 2 μl Cy5-dCTP in the experimental IP and 2 μl Cy3-dCTP into the mock IP. As microarrays require multiple replicates, interchange the fluorophore for further replicates of the same ChIP sample. This ''dye swap'' step allows one to eliminate variations of signals due to fluorophores.

5. Add 1 μl of the Klenow fragment from the Bioprime kit and incubate at 37° for 1 to 2 h. Then stop the reaction by adding 5 μl 0.5 *M* EDTA, pH 8.0.

6. Purify the DNA probe using a Microcon column as follows. Add 450 μl TE, pH 7.4, to the stopped labeling reaction. Lay onto Microcon filter. Spin 10 min at 8000*g*. Invert and spin 1 min at 8000*g* to recover the purified probe into a new tube (~20–40 μl volume). Combine the experimental and the mock-purified probes (Cy3 and Cy5 labeled) in a new Eppendorf tube. Then add 20 μg yeast tRNA and 20 μg poly(dA-dT) [this blocks hybridization to poly(A) tails of cDNA array elements] in 400 μl TE, pH 7.4.

7. Concentrate with a Microcon column as described earlier (8000*g*, ~18 min, then check volume every 1 min until volume is 26.9 μl or less). Then adjust the volume of the probe mixture to exactly 26.9 μl with bidistilled H_2O.

At this step, it is possible to bring the labeled probe immediately to the DNA array facility (or one can keep it frozen for later use). For immediate hybridization, proceed as follows.

8. For a total volume of 33 μl, covering a 24 $\times$ 30-mm coverslips, add 4.95 μl of 20$\times$ SSC (final concentration of about 3$\times$) to the 26.9 μl of the probe and mix by finger tapping. Then add 1.21 μl of 5% SDS. Boil the sample for 2 min and spin down in a small tabletop centrifuge to recover condensed droplets. Allow the sample to cool down to room temperature before applying to arrays.

9. Pipette two droplets (15 μl) of 3$\times$ SSC in the bottom of the hybridization chamber. Pipette 33 μl of the hybridization mixture onto a coverslip and then place the microarray glass slide (with the DNA side facing down) on the drop of hybridization mixture. Flip the slide quickly. Try to avoid air bubbles. If any bubbles form, remove them by tapping gently with forceps. Stick two thin Parafilm strips at the left and right sides of the coverslip and put the slide in the hybridization chamber (DNA face up). Add some small droplets of 3$\times$ SSC at the external side of each Parafilm strip in order to humidify the chamber. Take particular care to avoid these droplets from mixing with the hybridization mixture below the coverslip (the Parafilm strips should effectively isolate the coverslip). Close the hybridization chamber and incubate in a water bath at 42° for 15 min to adjust the temperature. Then place the chambers into a water bath at 64° for 16 to 18 h for hybridization.

10. After hybridization, wash the arrays several times in SSC. Dry the hybridization chamber with a towel. Unscrew the chamber and remove all traces of water.

11. Place the arrays, singly, in a rack inside the wash IA bath. Without any movement, wait for the coverslip to slip down by itself. If necessary, use forceps to carefully remove the coverslip. Avoid scrapping the slide. Agitate vigorously for 2 min.

12. Remove arrays from the rack and rinse in wash IB bath without a rack to remove traces of SDS. Then place the arrays in a rack in the wash II bath. Agitate vigorously for 2 min and transfer the rack into wash III bath. Again, agitate vigorously for 2 min. Then transfer the rack in an appropriate centrifuge already balanced and centrifuge for 2 min at 1000*g* in order to dry the slides (proceed rapidly to avoid uneven drying of the slides).

13. Store the slides in a dark chamber to avoid decay of the signal. Preferably scan the arrays immediately using a Genepix scanner.

Data Acquisition

1. Before scanning, do a prescan and examine the histogram of the intensity distribution of signals for each channel. The two channels should show overlapping curves. If this is not the case, change the voltage settings

of the scanner for each channel in order to obtain comparable signals. Once the settings are good, scan the array in the selected area of the slide at maximum resolution and save the picture.

2. Open the .gal file generated during the spotting of the slides that gives the ID of each feature on the slide with its coordinates. It generates a .gps file that corresponds to the virtual grid of the slides. Save each .gps file generated for each microarray.

3. Adjust the grid of the .gps file to the picture of the array. Each feature of the grid should correlate to the physical feature on the slide. The Genepix software calculates the signal intensity of each feature inside the virtual feature of the grid. It also calculates the median values for each feature and several normalization factors. We use the ratio between the average median intensities in the two channels as the normalization factor.

4. Once the grid is well set up, press the ''analyze'' button. It will generate a text tab separated file that can be saved as a .gpr file. This file gives all the necessary information of single microarray experiments, for example, date and name of the array, and, for each feature, its ID, coordinates, intensity in each channel, and ratio of medians. The normalization factor is given in a separate line in this file.

5. Open the .gpr file with Microsoft Excel software.

6. In order to obtain the normalized ratio for each feature ID, also known as ''fold change,'' create a new column, paste in it the value of the normalization factor, and then multiply it by the ratio of medians.

7. In the Excel file, reorder the features by their position in the genome. In our case, the ID number of the PCR fragments corresponds to the absolute coordinate of the sequence in the *Drosophila* genome. Use the graphic assistant to generate the distribution profile of the protein along the chromosomes. One example of such a distribution is seen in Fig. 4, which presents distribution of the chromatin proteins PC, Polyhomeotic (PH), and GAGA factor (GAF), over a genomic region containing the gene *polyhomeotic*, a well-known target of all these factors (Bloyer *et al.,* 2003). A further example is shown in Fig. 5 for the PH protein in the whole 7 Mb tiling path of the X chromosome.

One important point should be noted here. One might expect that all signals of individual spots on the chips hybridized with the mock IP sample are equal, as they should represent background hybridization to spotted fragments of roughly the same size (2 kb in our case). In fact, the background sequences are not totally represented randomly. The PCR efficiency, the spotting procedure, and the sequence complexity of each DNA fragment affect the efficiency of hybridization of each spot. However, the level of nonspecific attachment of each fragment to the IP beads

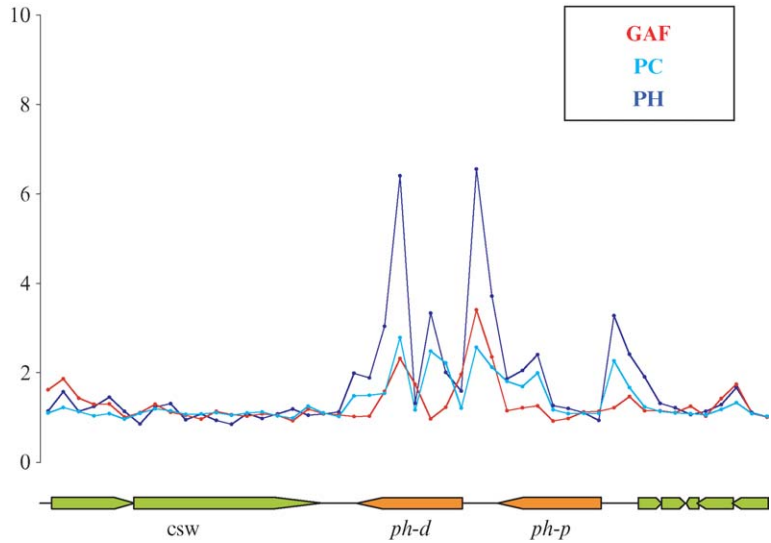

FIG. 4. Profile of chromatin proteins in the *ph-d ph-p* region. This graph represents the distribution of three chromatin components (GAF in red, PH in blue, and PC in light blue) on the *ph-d ph-p* region. The *ph* gene is duplicated in this region and encodes for the PH protein itself. Strong binding sites for the three proteins occur upstream and downstream of the *ph-d* gene. A binding site for PC and PH but not for GAF is observed upstream of the *ph-p* gene.

and minor skews in the ligation-mediated PCR amplification procedure may affect the amount of fragment present in the ChIP sample. All these sources of variation lead to signal intensities varying 5- to 10-fold for different fragments, even in control IP samples without antibody. An example of this variation is seen in Fig. 5, where the signal levels for the mock IP sample do not form a flat curve, but have peaks and valleys throughout the tiling path. Therefore, one should not be surprised to find signal variation in the mock IP channel. Fortunately, however, the ChIP technique is reproducible, and the dynamic range of detection in microarray experiments is large. Therefore, this variation in signal intensity does not prevent specific enrichments from being detected reproducibly, as can be seen in Fig. 5 (bottom).

Statistical Analysis of Data

At this point, basic data, as well as a graphic display of the protein distribution profile, are available. However, this is not the end of the analysis. It is crucial to determine how significant the profiles that have

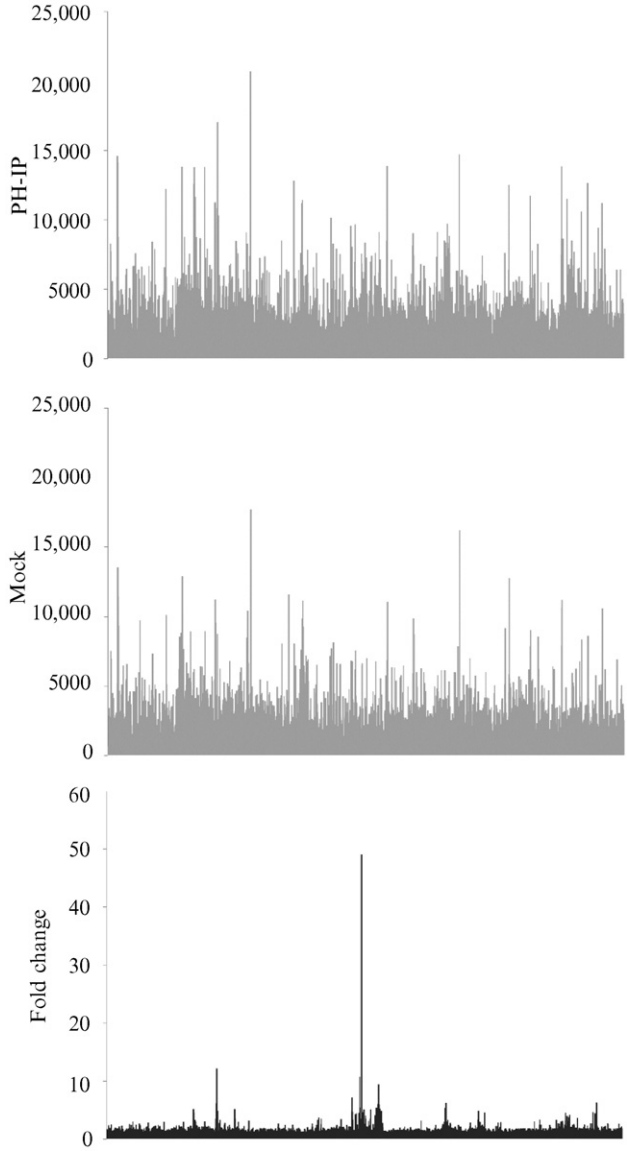

FIG. 5. Distribution of the hybridization signal on the microarrays. These graphs present the distribution of signals on the X chromosome. The normalized signal intensity for both ChIP (top) and mock (center) samples is shown in gray. The two graphs show a very well-correlated profile, consistent with most of the PH/mock ratios being close to 1. Only a minor fraction of the signals is enriched in the PH-ChIP sample, giving rise to the fold change graph (bottom).

been obtained are and what chromosomal features correspond to the peaks and valleys seen in the profiles. A large body of knowledge exists concerning statistical analysis of microarray data in the field of transcriptional studies. Even if no particular method, algorithm, or software prevails upon others for detecting significant variations, trivial methods have been found to be generally inappropriate and should be dismissed. The ''fold change'' is one of these simple methods. In this method, ratios greater than an arbitrary threshold (often set at twofold) are considered significant. Unfortunately, the fold change technique is inefficient in detecting small variations in the upper part of the signal range (generating false negatives); however, it selects a number of nonsignificant variations in the lower part of the signal range (generating false positives). This method is still used because it is simple to apply and because results are expressed in an intuitive way. Several alternative approaches have been proposed to correct for its obvious defects, but they are generally not applied to ChIP on chip analysis.

In order to improve robustness, specificity, and sensitivity of the analysis, the rank difference analysis of microarray (RDAM) method was adapted to ChIP on chip analysis (Nègre *et al.,* 2006). RDAM (Martin *et al.,* 2004) replaces raw signals by their rank, expressed on a 0–100 scale, which is a powerful normalizing procedure. Also, RDAM does not reduce replicated signals to their means, but instead considers variations, expressed as rank differences, between individual experimental points. Finally, RDAM estimates the total number of truly varying signals, assigns a p value to each signal variation, characterizes the selection of a signal using the false discovery rate in order to estimate the expected amount of false positive signals that may be present in the selected sample, and estimates the percentage of truly varying signals included in the selection (sensitivity). A detailed description of the RDAM method is found in Martin *et al*., 2004. Its application to ChIP on chip data has shown that this method is superior to the fold change, that is, setting an arbitrary threshold, for example, 2 in the fold change method would incur both in false positive and in false negative estimations that are not statistically significant when analyzed by RDAM. For this reason, we recommend analyzing ChIP on chip data using methods that are able to estimate the statistical significance of the enrichment of each feature, such as RDAM.

The statistical comparison of microarray samples also shows an interesting feature of ChIP on chip samples, namely that sample labeling and hybridization vary in independent experiments. For technical reasons, one set of the slides has been hybridized in autumn 2003 and the end of the replicates during winter 2004 in a different laboratory. We applied a variant of the method (Hennetin and Bellis, 2006) to display the relationships between

all experimental points by tracing a dendrogram. Instead of doing comparisons between an experimental point and the ''median'' point and selecting points by the *p* value, we made internal comparisons inside each condition (IP channel vs mock channel) and retained the 200 most varying probes in each comparison. Then we calculated the distance between any two comparisons as follows: d = 1 – #com/200, where #com is the number of common probes between the two 200 top lists. As observed on the corresponding dendrogram (Fig. 6), the pool of 2004 experiments is generally apart from the 2003 set. This type of diagram can help select a subset of points used and to conduct further statistical analysis and to discard

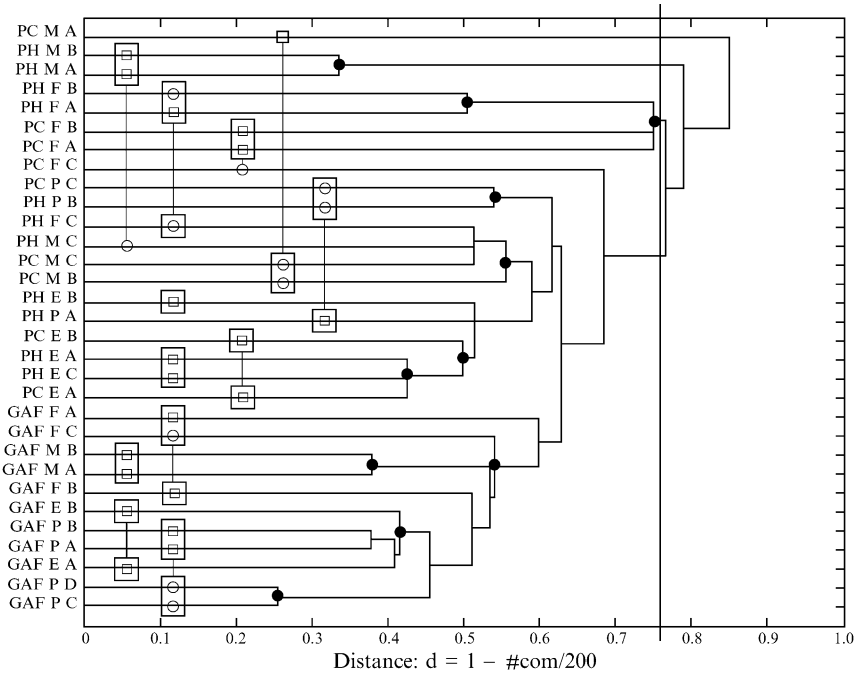

FIG. 6. Dendrogram displaying the correlation between different microarray replicates. Each experimental point displayed on the ordinate axis refers to a comparison between the IP channel and the mock channel. Names are abbreviated by designating the protein used in ChIPs (PH, PC, or GAF), the developmental stage analyzed (E, embryos; P, pupae; F, adult females; M, adult males), and the replicate experiment (A to D). Open circles and squares indicate, respectively, 2004 and 2003 samples. All experimental points belonging to the same condition are linked and aligned vertically. Boxes indicates those points selected for further statistical analysis by applying a threshold distance of $d = 0.76$, indicated by the vertical line. Black circles indicate the distance between the two best replicates in each condition.

outlier experiments. The retained samples can be subsequently analyzed by the RDAM method to estimate the significance of enrichments for each sample in each microarray feature. As an example we applied a threshold distance of 0.76, which is the minimal value allowing one to select at least two replicates for each condition. Other possibilities would have been to select in each condition the two replicates separated by the minimal distance, as indicated by closed circles in Fig. 6, or to take the whole set of replicates for subsequent analysis.

Graphic Comparison of ChIP on Chip Data with Genome Annotations

After applying a statistical analysis on microarray data, a list of binding sites for the protein of interest is obtained. Once the absolute coordinates of the binding sites along the genome are known, they can be visualized and compared to the genomic annotations using a variety of graphic interfaces. One of these tools, which is of simple use, is the GBrowse interface. In the case of the *Drosophila* genome, this is available at http://flybase.bio.indiana.edu/cgi-bin/gbrowse_fb/dmel. This interface allows external users to upload and display their own genomic annotations in the field: ''Upload/Remote Annotations.'' A set of simple rules, described in the ''Help'' section of the web site, can be applied to create a text file from microarrays data. This file must be formatted as in the following example:

reference = chrX
PHBindingSites BindingSite1 1979305-1981305
PHBindingSites BindingSite2 2001696-2003696
etc.

The first line indicates the genomic entity (BACs, scaffolds, chromosomes) to which the absolute coordinates of the annotations apply. Here, chrX indicates the X chromosome. Then, each of the following lines refers to an annotation corresponding to the binding sites detected by ChIP on chip. For each annotation, the generic description of the data set (PHBindingSites), the name of a particular feature (BindingSite1), and the absolute start and end coordinates of the feature (1979305–1981305) should be given. Each field must be space separated. When the text file is created and uploaded into GBrowse, a category line named ''PHBindingSites'' appears at the bottom of the graphic interface and displays squares for each described features at their appropriate position on the X chromosome.

Perspectives

To date, major limitations for ChIP on chip studies are the biological material, the efficiency of ChIP, and the microarray substrates. Although the ChIP method described here can be used with any *Drosophila* tissue, we are still unable to perform ChIP in specific cell types, and a major future challenge will be to isolate sufficient amounts of pure individual cell types for ChIP. One could imagine tagging specific cell types with GFP, dissociate cells from tissues, and then sort them by FACS. An alternative approach might rely on *in vivo* protein biotinylation (de Boer *et al.,* 2003). In this approach, tissue-specific expression of the BirA bacterial biotin ligase (by appropriate transgenes) coupled to expression of a fusion protein between the protein of interest and a target peptide tag can lead to *in vivo* biotinylation of the target protein. In conjunction with formaldehyde cross-linking, this might allow one to recover the biotinylated protein of interest in the cell type to be studied by using the strong and highly specific biotin–streptavidin interaction.

Another technical point that needs to be improved is the amount of biological material needed for ChIP. For the moment, no less than tens of milligram amounts of cells are needed for one IP, which is equivalent to more than 10^7 cells. Perhaps this might be scaled down by improving the efficiency of immunoprecipitation and/or by developing a microarray fluidic technology allowing much smaller volumes (i.e., amounts) of labeled DNA to be hybridized.

A further aspect that is likely to improve with time concerns the microarrays themselves. It is becoming possible to analyze large genomic regions or even whole genomes, but at the moment this requires large amounts of work, and in most cases one genome can only be contained in several chips. Future improvements in the array density should allow one to obtain whole genome data from a single array, to improve reproducibility to a point where data from different labs can be compared directly, and to reduce the cost of the experiments. In summary, disposing of a large number of directly comparable data on the whole-genome transcription profiling as well as protein location for multiple, evolutionarily related organisms will be of invaluable importance in order to understand how gene networks established by transcription factors modulate gene expression throughout development and evolution.

Acknowledgments

We are very grateful to K.P. White (Yale University) for giving N.N. access to his microarray platform and to L.V. Sun for training him in microarray production and

processing. We thank the Montpellier L-R Genopole for arraying, spotting, and processing custom microarrays. N.N is supported by the Ministère de l'Enseignement Supérieur, the European Molecular Biology Organisation (EMBO), and the Association pour la Recherche sur le Cancer. J.H. was supported by a postdoctoral fellowship from the CNRS. G.C. was supported by grants from the CNRS (Programme ''Puces à ADN''), the Association pour la Recherche sur le Cancer, the Human Frontier Science Program Organization, the European Union FP6 (Network of Excellence The Epigenome and STREP 3D Genome), the Indo-French Centre for Promotion of Advanced Research, and the Ministère de l'Enseignement Supérieur, ACI BCMS.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., and Venter, J. C. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**(5461), 2185–2195.

Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W., and White, K. P. (2002). Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**(5590), 2270–2275.

Bloyer, S., Cavalli, G., Brock, H. W., and Dura, J. M. (2003). Identification and characterization of polyhomeotic PREs and TREs. *Dev. Biol.* **261**(2), 426–442.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **100**(13), 7480–7485.

Francis, N. J., Saurin, A. J., Shao, Z., and Kingston, R. E. (2001). Reconstitution of a functional core polycomb repressive complex. *Mol. Cell* **8**(3), 545–556.

Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H. W., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**(8), 2941–2950.

Gupta, V., and Oliver, B. (2003). Drosophila microarray platforms. *Brief Funct. Genom. Proteom.* **2**(2), 97–105.

Hennetin, J., and Bellis, M. (2006). Clustering methods for analyzing large data sets: Gonad development, a study case. *Methods Enzymol.* **411,** 387–407.

Kirmizis, A., Bartley, S. M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., and Farnham, P. J. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**(13), 1592–1605.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J.,

Wagner, F., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822), 860–921.

Liu, Z. R., Sargueil, B., and Smith, C. W. (2000). Methylene blue-mediated cross-linking of proteins to double-stranded RNA. *Methods Enzymol.* **318,** 22–33.

Martin, D. E., Demougin, P., Hall, M. N., and Bellis, M. (2004). Rank difference analysis of microarrays (RDAM), a novel approach to statistical analysis of microarray expression profiling data. *BMC Bioinformat.* **5**(1), 148.

Min, J., Zhang, Y., and Xu, R. M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**(15), 1823–1828.

Negre, N., Hennetin, J., Sun, L. V., Lavrov, S., Bellis, M., White, K. P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during *Drosophila* development. *PLOS Biol.* **4,** e170.

Orlando, V., Strutt, H., and Paro, R. (1997). Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**(2), 205–214.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290** (5500), 2306–2309.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38,** 413–443.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235), 467–470.

Simon, J., Chiang, A., Bender, W., Shimell, M. J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev. Biol.* **158**(1), 131–144.

Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P. E., Bussemaker, H. J., and White, K. P. (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**(5696), 655–660.

Zhang, L., Zhang, K., Prandl, R., and Schoffl, F. (2004). Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation. *Biochem. Biophys. Res. Commun.* **322** (3), 705–711.