

Chromatin profiling using targeted DNA adenine methyltransferase

Bas van Steensel^{1,3}, Jeffrey Delrow² & Steven Henikoff¹

Chromatin is the highly complex structure consisting of DNA and hundreds of associated proteins. Most chromatin proteins exert their regulatory and structural functions by binding to specific chromosomal loci. Knowledge of the identity of these *in vivo* target loci is essential for the understanding of the functions and mechanisms of action of chromatin proteins. We report here large-scale mapping of *in vivo* binding sites of chromatin proteins, using a novel approach based on a combination of targeted DNA methylation and microarray technology. We show that three distinct chromatin proteins in *Drosophila melanogaster* cells each associate with specific sets of genes. HP1 binds predominantly to pericentric genes and transposable elements. GAGA factor associates with euchromatic genes that are enriched in (GA)_n motifs. A *Drosophila* homolog of *Saccharomyces cerevisiae* Sir2p is associated with several active genes and is excluded from heterochromatin. High-resolution, genome-wide maps of target loci of chromatin proteins ('chromatin profiles') provide new insights into chromatin structure and gene regulation.

Our approach is based on the observation that *Escherichia coli* DNA adenine methyltransferase (Dam) can be targeted *in vivo* to binding sequences of a chromatin protein of interest by tethering Dam to

the chromatin protein¹. This leads to specific methylation of DNA near the binding sites of the chromatin protein. We developed a protocol for measuring adenine methylation levels of a large number of genomic loci in parallel using DNA microarray technology, allowing the large-scale identification of *in vivo* target loci of a chromatin protein of interest (Fig. 1). To test this protocol, we transfected *Drosophila* Kc cells either with a Dam fusion protein or with Dam only. The latter served as a reference to normalize for local differences in chromatin accessibility to methylation by Dam (ref. 1). Twenty-four hours after transfection, we isolated genomic DNA and purified methylated regions. We labeled purified methylated regions from experimental and control cells with the fluorochromes Cy3 and Cy5, respectively, and mixed and co-hybridized these to DNA microarrays containing approximately 300 *Drosophila* cDNAs and a few genomic DNA fragments. Target sequences of the chromatin proteins of interest were identified based on the Cy3: Cy5 fluorescence ratio, which indicates the relative targeted methylation level of each probed sequence.

We first tested the mapping technique for *Drosophila* heterochromatin protein-1 (HP1), which is predominantly associated with pericentric heterochromatin^{1,2}. We mapped HP1 target loci using a fusion protein (Dam-HP1) consisting of Dam linked to the amino terminus of full-length HP1. A scatter diagram of the microarray hybridization signals (Fig. 2a) and a chromosomal map of Cy3: Cy5 ratios of all probed loci (Fig. 2b) showed that most loci display an almost identical Cy3: Cy5 ratio, indicating only background levels of methylation and hence no detectable association of these loci with HP1. A distinct set of loci, however, show higher Cy3: Cy5 ratios. These loci represent targets of HP1. In support of this interpretation, we found that loci with elevated Cy3: Cy5 ratios included *Bari-1*, *cta*, *28S* and the histone gene cluster (*HisC*), all of which we had previously identified as HP1 targets¹. The euchromatic loci *5S*, *Prat*, *GART* and *tosca*, which served as negative controls, did not have elevated Cy3: Cy5 ratios.

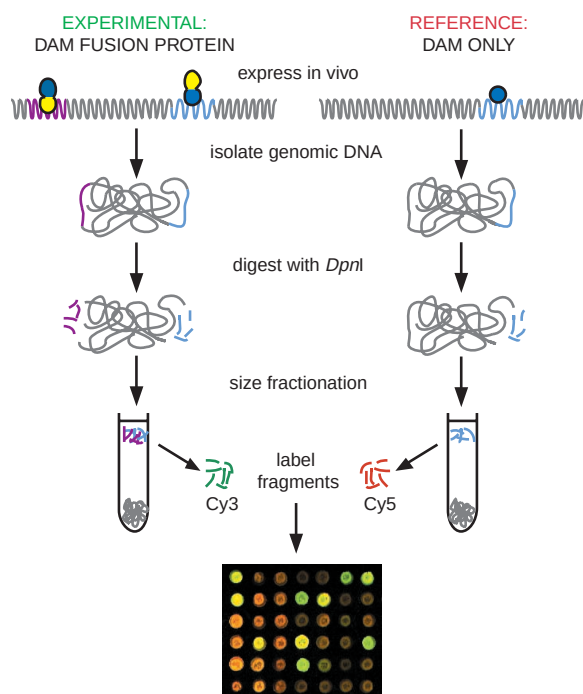
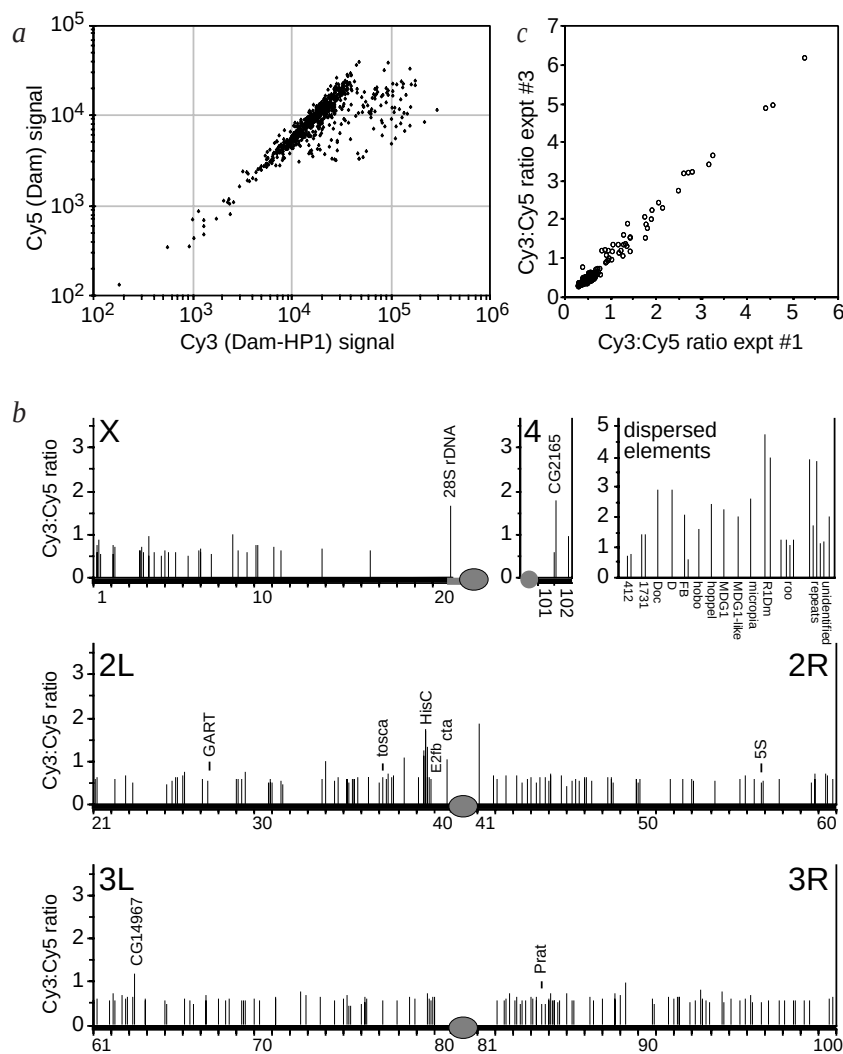


Fig. 1 Experimental procedure for identifying *in vivo* targets of chromatin proteins. In parallel experiments, cells are transfected with a Dam fusion protein (experimental, left) and with Dam only (reference, right). Unfused Dam methylates 'open' chromatin regions (blue); the Dam-fusion protein methylates both 'open' chromatin regions and target loci of the Dam fusion partner (purple). Methylated regions are purified as small fragments by sucrose gradient centrifugation after digestion with *DpnI* (which only cuts methylated GATC sequences), fluorescently labeled and used to probe microarrays. Target loci are identified based on the ratios of the Cy3 (pseudo-colored green) and Cy5 (pseudo-colored red) hybridization signals. These Cy3: Cy5 ratios reflect the relative abundances of the probed sequences in purified (that is, methylated) experimental and reference DNA samples, and thus indicate the level of binding of the chromatin protein of interest to the probed sequences. Note that in practice 'open' regions are rarely completely methylated by unfused Dam (ref. 1 and Web Fig. A), hence targeted methylation in such a region can still be detected as an increase in the Cy3: Cy5 ratio. We have also not found any regions in the genome that are completely inaccessible to Dam.

¹Howard Hughes Medical Institute and ²DNA Array Facility, Fred Hutchinson Cancer Research Center, Seattle, Washington. ³Present address: Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands. Correspondence should be addressed to B.v.S. (e-mail: steensel@chem.uva.nl).



Fig. 2 Mapping of HP1 target loci. **a**, Scatter diagram of a representative experiment showing the Cy3 and Cy5 signals of all cDNAs on the array. **b**, Chromosomal map of Cy3:cy5 ratios (representative experiment). Probed loci are indicated by their approximate position on the standard polytene chromosome map. Centromeres are indicated by gray ovals; the large heterochromatic proximal region of the X chromosome is depicted as a gray rectangle (not to scale). Some genes with relatively high levels of HP1 binding are labeled. Dispersed repetitive elements (mostly transposons) are shown in a separate graph. **c**, Comparison of Cy3:cy5 ratios between two experiments ($r=0.99$, $P<10^{-4}$).



Although the cut-off between 'target' and 'non-target' Cy3:cy5 ratios is arbitrary, the differences in Cy3:cy5 ratios between probed loci are highly reproducible. Pair-wise comparisons of 3 independent experiments showed correlation coefficients between 0.95 and 0.99 (Fig. 2c), and on average the standard deviation of the Cy3:cy5 ratios was less than 10% of the observed mean values. Hence, loci that show only a modest increase in Cy3:cy5 ratio over background levels (for example, gene *CG14967*; Fig. 2b) are likely to be associated with HP1 *in vivo*, although perhaps relatively weakly. Moreover, differences in Cy3:cy5 ratios between genes in our assay may underestimate differences in protein binding. By Southern-blot analysis, we found that the *Bari-1* locus displays approximately eightfold higher HP1-targeted methylation than the 5S locus (data not shown and ref. 1), yet our microarray experiments indicate only an approximately fourfold difference. Such a microarray-specific compression effect has been observed previously³.

Both the *HisC* locus and the gene *cta*, which are located near the centromere on the left arm of chromosome 2, are associated with HP1. In contrast, *E2f2*, which lies between these two loci, shows no detectable HP1 binding, indicating a discontinuous distribution of HP1 in this region. This is consistent with a banded pattern of HP1 staining in this region in polytene chromosomes (data not shown). At least one gene located on a euchromatic chromosome arm (*CG14967*) reproducibly associates with HP1 (Fig. 2b), albeit weakly.

We found that HP1 binds to a wide variety of transposable elements. Of 13 different transposons present on the microarray, 12 showed moderate to strong association with HP1 (Fig. 2b). The average Cy3:cy5 ratio of these elements taken as a group was significantly higher than that of all other loci taken together (1.86 ± 1.06 versus 0.62 ± 0.22 , average of three experiments; $P < 10^{-4}$, Mann-Whitney U-test). This is consistent with the enrichment of most transposable elements in pericentric heterochromatin⁴⁻⁶, the primary region of HP1 staining in polytene chromosomes². Given their high abundance, transposons must form a major class of HP1

targets in the *Drosophila* genome. It may be that HP1 is involved in genome defense against parasitic elements⁷.

As a second test of the mapping technique we chose *Drosophila* GAGA factor (GAF), which binds hundreds of euchromatic sites in polytene chromosomes^{8,9}. We mapped GAF target loci using a fusion protein (GAF-Dam) consisting of Dam linked to the carboxy terminus of full-length GAF. A chromosomal map of GAF-Dam binding (Fig. 3a) indicates that GAF interacts with many genes, although to varying degrees. Correlation analysis of independent experiments showed that the results were highly reproducible ($r=0.90$, data not shown). Moreover, C- and N-terminal Dam fusions of GAF gave similar results (Fig. 3b), indicating that Dam, when fused to either end of GAF, does not interfere with correct targeting of GAF.

Genes that bind GAF strongly have no common function or expression pattern. Because GAF binds GA-rich regulatory elements¹⁰⁻¹³, we investigated whether the GAF target loci that we identified are enriched in such elements. Indeed, loci that display moderate-to-strong GAF binding have significantly higher average densities of GAGAG and GAGAGAG sequences than loci with low GAF binding (Fig. 3c). These results further confirm that mapping of binding sites using targeted methylation is sensitive and specific.



letter

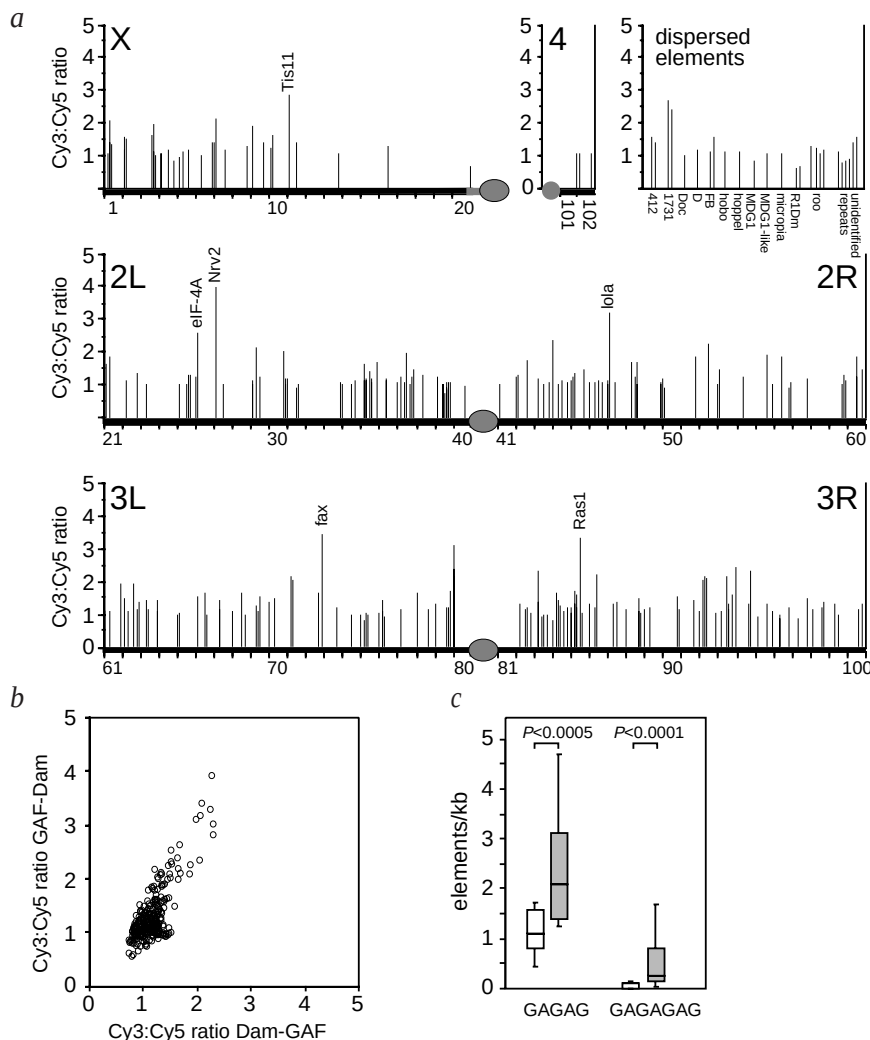


Fig. 3 Mapping of GAF target loci. **a**, Chromosomal map of Cy3:cy5 ratios (average of two experiments) using a GAF-Dam fusion protein. Some genes with relatively high levels of GAF binding are labeled. **b**, Scatter graph showing the correlation between Cy3:cy5 ratios obtained with GAF-Dam and Dam-GAF fusion proteins ($r=0.80$ in 2 independent comparisons, $P<10^{-4}$). **c**, Box plot showing the relative abundances of GAGAG and GAGAGAG sequence elements in probed regions with low (open boxes) and high (filled boxes) levels of GAF binding. Horizontal lines represent the 10th, 25th, 50th (median), 75th and 90th percentiles. P values are according to the Mann-Whitney U-test.

of HP1 binding contain low levels of Sir2, indicating that Sir2 is not part of HP1-containing heterochromatin. Immunofluorescence microscopy of Kc cell nuclei confirmed that the Sir2-Dam fusion protein is largely excluded from HP1-rich regions (Web Fig. B). Although we cannot rule out that Sir2 binds to a small subset of heterochromatic sequences not represented on our microarray, our results indicate that Sir2 does not have a major role in gene silencing, in contrast to Sir2p in *S. cerevisiae*. Consistent with this, some of the protein domains in Sir2p that are essential for silencing are not conserved in Sir2 in *Drosophila* and other higher eukaryotes^{17,18}. Possibly, one or more of the other four Sir2-like proteins in *Drosophila*¹⁹ are associated with

Finally, we identified target loci of Sir2, a still-uncharacterized *Drosophila* homolog of *S. cerevisiae* Sir2p. In budding yeast, Sir2p has a role in silencing of genes in the silent mating-type loci, telomeric regions and rDNA locus^{14,15}. The homology to yeast Sir2p suggested that *Drosophila* Sir2 might be associated with heterochromatin.

Mapping results for Sir2 (Fig. 4) indicated that Sir2 is associated with numerous genes. Among the strongest Sir2-binding loci are several euchromatic, constitutively expressed genes encoding, for example, translation factors, putative ribosomal proteins, α -tubulin, hsc4 and EIP40. This indicates that *Drosophila* Sir2 binds to a set of active genes, unlike yeast Sir2p. The binding to active genes is not due to the 'open' chromatin conformation of these genes, because the binding data are corrected for local differences in accessibility. Independent mapping of gene accessibility (Web Fig. A) confirmed that Sir2 does not generally bind to 'open' chromatin regions.

The results also indicate that Sir2 does not detectably interact with the 28S genes in the rDNA gene cluster (Fig. 4). Again, this is in contrast to yeast Sir2p, which binds to the rDNA repeat¹⁶. This lack of binding is not due to a general inaccessibility of the nucleolar compartment, because HP1 associates with the 28S genes¹ (Fig. 2b).

Comparison of the distributions of HP1 and Sir2 by means of a bivariate scatter diagram (Fig. 5) revealed that loci with high levels

heterochromatin. Pair-wise comparisons of the distributions of GAF versus HP1 and Sir2 are available (Web Fig. C).

We have reported a simple and efficient large-scale technique for mapping of protein-binding sites in the genome of *Drosophila*. This technique may prove to be particularly useful in higher eukaryotes, in which other global mapping approaches based on chromatin immunoprecipitation methods²⁰⁻²² may fail due to the high complexity of the genome and insufficient specificity of antibodies. Expression of Dam has no toxic side effects in flies^{1,23}. Because methylation by tethered Dam spreads over 2–5 kb from a discrete protein-binding sequence¹, target loci may be mapped with a resolution of a few kilobases. In certain cases, sequence comparisons of all identified target loci may reveal common sequence elements that potentially mediate the recruitment of the chromatin protein, thus effectively increasing the mapping resolution. Because our microarrays contained mostly cDNA probes, our analysis focused primarily on transcribed regions. Arrays of genomic fragments systematically covering promoter and enhancer regions should provide detailed insight into the associations of chromatin proteins with *cis*-acting elements. With larger microarrays it will be possible to obtain extremely detailed genome-wide maps of the binding patterns of chromatin proteins. We expect that such 'chromatin profiles' will yield unprecedented insights into the functions and mechanisms of action of chromatin proteins.

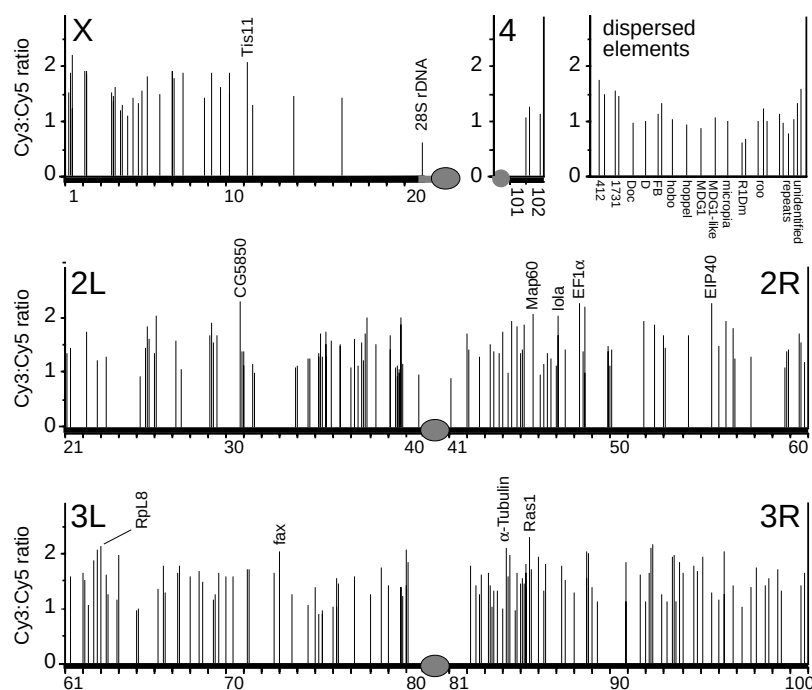


Fig. 4 Mapping of Sir2 target loci. A chromosomal map of Cy3:cy5 ratios (average of two experiments) is shown. Some genes of particular interest or with high levels of Sir2 binding are labeled. Results were highly reproducible ($r=0.81$ between 2 independent experiments; data not shown).

trophoresis) and concentrated them by isopropanol precipitation. Typically, this procedure yielded 20–50 μ g methylated DNA. We processed genomic DNA from control cells transfected without plasmid in parallel, which generally gave an 80–90% lower yield of <2.5-kb fragments, indicating that the methylated DNA is ~80–90% pure. An estimated 20–50% of the methylated DNA consists of plasmid DNA that was taken up by the cells during transfection (data not shown). This plasmid DNA does not interfere with the subsequent labeling and hybridization procedure.

Microarray hybridizations. We labeled purified methylated DNA (1 μ g) with Cy3- or Cy5-dCTP (Amersham Pharmacia) by random priming³. We mixed and co-hybridized

the labeled experimental and reference DNA samples to a microarray in 3 $\times$ SSC (450 mM sodium chloride and 45 mM sodium citrate, pH 7.0) in the presence of 0.22% SDS, poly [dA-dT] (20 μ g), yeast tRNA (100 μ g) and unlabeled *DpnI*-digested plasmid (25 μ g) encoding the fusion protein that was used for transfection. Hybridization was performed at 63 $^{\circ}$ C for 16 h followed by sequential washings at RT in 1 $\times$ SSC/0.03% SDS, 1 $\times$ SSC, 0.2 $\times$ SSC and 0.05 $\times$ SSC. We spun washed arrays dry in a centrifuge and immediately scanned them using a GenePix 4000 fluorescent scanner (Axon Instruments).

Data processing and analysis. We used GenePix 3.0 image analysis software for image processing, and StatView software (Abacus Concepts) for statistical analysis.

We normalized the measured Cy3:cy5 ratios to allow comparison between experiments. For this, 16 spots of total *Drosophila* genomic DNA, present on each microarray, served as an internal normalization standard. We performed normalization by dividing the measured Cy3:cy5 ratios by the average Cy3:cy5 ratio measured in the 16 genomic DNA spots. Thus, for a given probed locus, a Cy3:cy5 value >1 indicates more binding of the chromatin protein than the average binding level in the entire genome; a value <1 means that the binding is less than the average binding level in the entire genome.

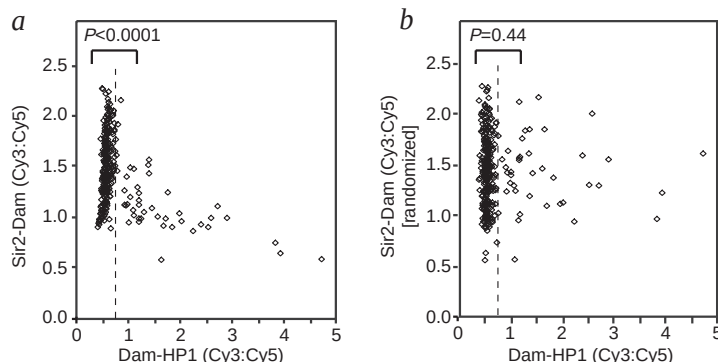
Note: supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Methods

Descriptions of plasmids, microarray construction, immunocytochemistry procedures and data normalization are available (see Web Methods). Data and plasmid sequences are available on our web site (<http://blocks.fhcr.org/DamID>).

Transfections and DNA purification. We used *Drosophila* Kc cells because they are easily transfected and provide a homogeneous cell population. We cultured Kc cells and performed transfections as described²⁴. Expression of the Dam proteins was driven by the low constitutive activity of the uninduced hsp70 promoter, ensuring very low expression levels. After 24 h, we isolated genomic DNA as described²⁵, except we included an RNase incubation before the second proteinase K treatment. We digested ~0.5–1.5 mg DNA (from ~10⁹ cells) for 16 h with *DpnI* (500–1,000 units; New England Biolabs) and size-fractionated the digested DNA by ultracentrifugation in a Beckman SW-40TI swing-out rotor for 16 h at 79,000g on 11 ml sucrose gradients (5–30%) containing Tris-HCl (10 mM, pH 7.4), NaCl (150 mM) and EDTA (10 mM). We pooled gradient fractions containing DNA fragments smaller than approximately 2.5 kb (as judged by agarose gel elec-

Fig. 5 Pair-wise comparison of binding patterns of HP1 and Sir2. **a**, Bivariate scatter graph showing the Cy3:cy5 ratios of all probed loci. Values are the average of three (Dam-HP1) or two (Sir2-Dam) independent experiments. To test whether Sir2 is excluded from HP1 targets, loci were divided into two categories based on their level of association with HP1, taking the average HP1 level of all loci (Cy3:cy5 ratio 0.74, dashed line) as threshold. The distributions of Sir2 levels in the 'high HP1' and 'low HP1' categories were then compared using the Mann-Whitney U-test (P value is indicated). This test was highly robust: for all threshold levels >0.62, we found $P<0.05$. **b**, Same graph as in (a), but the values for Sir2-Dam were randomly reassigned to a different locus, simulating how a scatter graph would look if the distribution of Sir2 were not correlated with HP1. The P value in (b) is the average of 20 simulations.





letter

For each probed locus, the standard deviation (s.d.), expressed as a percentage of the mean Cy3: Cy5 ratio, was calculated based on replicate experiments. Averaged over all loci, the s.d. was 6.8% (Dam–HP1), 11.7% (GAF–Dam), 10.5% (Dam–GAF) and 10.1% (Sir2–Dam). Moreover, only 0.3%, 0.3%, 3.4% and 6.8% of all probed loci had a s.d. >30%, respectively. Thus, the data are highly reproducible. For this analysis, we omitted one experiment with Dam–HP1 because of its relatively high baseline level and poor dynamic range, which systematically skewed the s.d. values. Nevertheless, this experiment showed good correlation ($r=0.95$) with the other experiments, and visual inspection showed that the binding patterns were qualitatively identical.

We identified five ESTs that were initially thought to represent unique genes along the euchromatic arms (based on the available 5' sequence) as HP1 targets. Sequencing of both 5' and 3' ends indicated that these cDNA clones contain repetitive sequences, which may be a cloning artifact that occurred during the construction of the CK library²⁶. (These clones are represented in the dispersed elements section of Figs. 2–4.) The fact that these clones were identified as HP1 targets underscores the sensitivity of our assay.

For analysis of the density of (GA)_n elements, we selected 15 loci showing low GAF–Dam binding (Cy3: Cy5 ratios 0.97 ± 0.09) and 15 loci showing high GAF–Dam binding (Cy3: Cy5 ratios 2.77 ± 0.65) based on availability of the complete probe sequence. We obtained corresponding genomic sequences from the BDGP/Celera genomic database and included in our analysis introns smaller than 5 kb, as well as 3 kb of sequence upstream and downstream of the probed region, because methylation by tethered Dam extends *in cis* over a few kilobases¹. On average, approximately 7.5 kb was analyzed per probed region. We counted GAGAG and GAGAGAG sequences in both orientations; we counted partially overlapping elements separately (for example, the sequence TGAGAGAGC contains one GAGAGAG and two GAGAG elements).

Acknowledgments

We thank E. Giniger for coordinating the assembly of the Northwest Fly Consortium microarray; C. Neal for help with microarray construction and analysis; P. Ng for help with the EST annotation; J. Smothers for the anti-HP1 antibody; J. O'Brien for technical assistance; and members of the Henikoff lab for suggestions.

Received 20 September 2000; accepted 6 February 2001.

- van Steensel, B. & Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature Biotechnol.* **18**, 424–428 (2000).
- James, T.C. *et al.* Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of *Drosophila*. *Eur. J. Cell Biol.* **50**, 170–180 (1989).

- Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
- Charlesworth, B., Jarne, P. & Assimacopoulos, S. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III Element abundances in heterochromatin. *Genet. Res.* **64**, 183–197 (1994).
- Pimpinelli, S. *et al.* Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc. Natl. Acad. Sci. USA* **92**, 3804–3808 (1995).
- Carmena, M. & Gonzalez, C. Transposable elements map in a conserved pattern of distribution extending from β -heterochromatin to centromeres in *Drosophila melanogaster*. *Chromosoma* **103**, 676–684 (1995).
- Henikoff, S. Heterochromatin function in complex genomes. *Biochim. Biophys. Acta* **1470**, O1–O8 (1999).
- Tsukiyama, T., Becker, P.B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
- Benyajati, C. *et al.* Multiple isoforms of GAGA factor, a critical component of chromatin structure. *Nucleic Acids Res.* **25**, 3345–3353 (1997).
- Biggin, M.D. & Tjian, R. Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* **53**, 699–711 (1988).
- Soeller, W.C., Oh, C.E. & Kornberg, T.B. Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* **13**, 7961–7970 (1993).
- 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).
- O'Brien, T., Wilkins, R.C., Giardina, C. & Lis, J.T. Distribution of GAGA protein on *Drosophila* genes *in vivo*. *Genes Dev.* **9**, 1098–1110 (1995).
- Guarente, L. Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* **14**, 1021–1026 (2000).
- Gartenberg, M.R. The Sir proteins of *Saccharomyces cerevisiae*: mediators of transcriptional silencing and much more. *Curr. Opin. Microbiol.* **3**, 132–137 (2000).
- Gotta, M. *et al.* Localization of Sir2p: the nucleolus as a compartment for silent information regulators. *EMBO J.* **16**, 3243–3255 (1997).
- Cuperus, G., Shafaatian, R. & Shore, D. Locus specificity determinants in the multifunctional yeast silencing protein Sir2. *EMBO J.* **19**, 2641–2651 (2000).
- Cockell, M.M., Perrod, S. & Gasser, S.M. Analysis of Sir2p domains required for rDNA and telomeric silencing in *Saccharomyces cerevisiae*. *Genetics* **154**, 1069–1083 (2000).
- Frye, R.A. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochem. Biophys. Res. Commun.* **273**, 793–798 (2000).
- Blat, Y. & Kleckner, N. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* **98**, 249–259 (1999).
- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2001).
- Iyer, V.R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
- Wines, D.R., Talbert, P.B., Clark, D.V. & Henikoff, S. Introduction of a DNA methyltransferase into *Drosophila* to probe chromatin structure *in vivo*. *Chromosoma* **104**, 332–340 (1996).
- Henikoff, S., Ahmad, K., Platero, J.S. & van Steensel, B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl. Acad. Sci. USA* **97**, 716–721 (2000).
- de Lange, T. *et al.* Structure and variability of human chromosome ends. *Mol. Cell. Biol.* **10**, 518–527 (1990).
- Kopczynski, C.C. *et al.* A high throughput screen to identify secreted and transmembrane proteins involved in *Drosophila* embryogenesis. *Proc. Natl. Acad. Sci. USA* **95**, 9973–9978 (1998).