Genome-scale profiling of histone H3.3 replacement patterns

Yoshiko Mito[1], Jorja G Henikoff[1] & Steven Henikoff [1,2]

Histones of multicellular organisms are assembled into chromatin primarily during DNA replication. When chromatin assembly occurs at other times, the histone H3.3 variant replaces canonical H3. Here we introduce a new strategy for profiling epigenetic patterns on the basis of H3.3 replacement, using microarrays covering roughly one-third of the *Drosophila melanogaster* genome at 100-bp resolution. We identified patterns of H3.3 replacement over active genes and transposons. H3.3 replacement occurred prominently at sites of abundant RNA polymerase II and methylated H3 Lys4 throughout the genome and was enhanced on the dosage-compensated male X chromosome. Active genes were depleted of histones at promoters and were enriched in H3.3 from upstream to downstream of transcription units. We propose that deposition and inheritance of actively modified H3.3 in regulatory regions maintains transcriptionally active chromatin.

Eukaryotic DNA is tightly wrapped around core histones to form nucleosome particles for compact packaging in both interphase nuclei and mitotic chromosomes[1]. Histones also have roles in gene regulation and are implicated in epigenetic inheritance. During development, the active state of a gene is inherited, and this 'cellular memory' of gene activity is thought to be propagated during DNA replication by the distribution of nucleosome particles to daughter chromatids[2]. For propagation to occur, heritable chromatin features, such as covalent modifications and variant histones, must distinguish active genes from inactive genes. An understanding of cellular memory will have broad implications for studies of epigenetic processes in general, including development.

Most studies of chromatin inheritance focused on covalent histone modifications. Modifications that correspond to transcriptionally active states are strongly associated with one another[3–5], suggestive of a single underlying process. Histone replacement is a good candidate for this mechanism. Whereas the bulk of nucleosome assembly occurs behind the replication fork, some also occurs independently of replication. In multicellular organisms, replication-independent assembly can be distinguished from replication-coupled assembly by the incorporation of alternative histone H3 variants[6]. Replication-coupled assembly deposits primarily canonical histone H3, which is expressed at high levels in S phase under tight cell cycle control, whereas replication-independent assembly deposits the constitutive variant H3.3 but not H3. There are probably important mechanistic differences between replication-coupled and replication-independent assembly, because replication-coupled deposition occurs in gaps created by the synthesis of newly replicated DNA, whereas replication-independent assembly occurs by replacement

of nucleosomal histones. Consistent with this hypothesis, distinct chaperone complexes associated with H3 and H3.3 catalyze replication-coupled and replication-independent assembly, respectively[7]. Replication-independent replacement is a potential mechanism for the propagation of active chromatin, because it takes place preferentially at actively transcribed loci where many histone modifications are enriched[6,8–10].

If replication-independent nucleosome assembly is generally responsible for maintaining active chromatin, then the distribution of H3.3 should reflect patterns of active transcription and associated histone modifications. Consistent with this expectation, biochemical fractionation of H3 variants shows that H3.3 is relatively enriched in modifications that are characteristic of active chromatin in both plants and animals[11–13]. Furthermore, cytogenetic analysis showed that H3.3 is deposited rapidly at highly transcribed loci in *D. melanogaster*, suggesting that replication-independent assembly is coupled to transcriptional elongation[9]. H3.3 is also enriched at actively transcribed human genes[10]. But the low resolution of these studies and the relatively limited number of gene targets examined did not allow for general conclusions about chromatin patterns. High-resolution analysis is needed to determine whether patterns of replication-independent assembly correspond to transcribed regions of active genes.

Here we used H3.3 and H3 as markers for alternative assembly pathways. We introduced these variants into cultured *D. melanogaster* cells with a biotin epitope tag, allowing for high-affinity purification of nucleosomes containing specific histone variants. We then profiled localization patterns using high-resolution microarray analysis and identified patterns of H3.3 replacement at active genes and families of transposons.

[1]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. [2]Howard Hughes Medical Institute, USA. Correspondence should be addressed to S.H. (steveh@fhcrc.org).
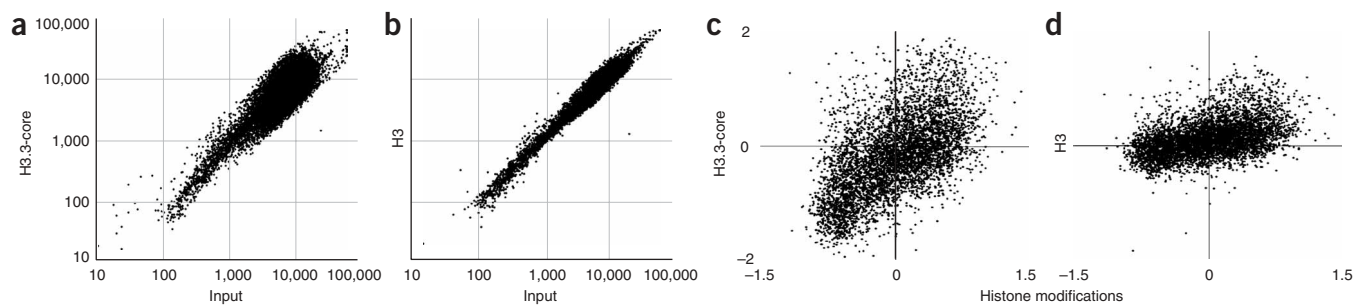
Figure 1 Replication-independent deposition corresponds to active modifications at genes. (**a,b**) Representative scatter plots of raw signal intensities measured on cDNA microarrays for input versus pull-down samples for (**a**) H3.3-core and (**b**) H3. (**c,d**) Representative scatter plots of relative enrichment for the average of three strongly correlated histone modifications (acetylation of H3, H3K79me2 and acetylation of H4)[3] versus (**c**) H3.3-core and (**d**) H3 measured on cDNA microarrays. Western-blot analysis showed that essentially all biotin–H3.3-core and biotin-H3 was pulled down (**Supplementary Fig. 6** online). Results obtained using full-length H3.3, results of MNase digestion, pull-down and cDNA microarray analysis were similar to those obtained using H3.3-core (**Supplementary Fig. 7** online).

RESULTS

H3.3 is preferentially deposited in genic regions

To profile deposition patterns, we tagged histones at their N termini with biotin ligase recognition peptide (BLRP)[14] and expressed them together with *Escherichia coli* biotin ligase in *D. melanogaster* S2 cells. We treated nuclei with micrococcal nuclease (MNase) to digest chromatin to yield mostly mononucleosomes and then extracted chromatin. We measured replication-coupled assembly by pull-down of biotin-H3 and replication-independent assembly by pull-down of biotin–H3.3-core, in which deletion of the N-terminal tail prevents replication-coupled deposition[6]. We purified DNA from input and pull-down fractions and assayed the matched DNAs by hybridization to microarrays representing ∼12,000 annotated *D. melanogaster* cDNAs. If a histone variant is uniformly distributed at loci represented on the array, then a scatter plot of pull-down versus input signal intensities will show a tight distribution along the diagonal. But H3.3-core had a broad distribution around the diagonal, indicating that a large fraction of the gene loci on the array was enriched or depleted in pull-down versus input fractions, whereas the H3 control had a tight distribution (**Fig. 1a,b**). This indicates that H3.3 levels vary from gene to gene, whereas H3 levels are relatively uniform.

To determine whether the gene-to-gene variation in H3.3 levels was related to variations in histone modifications, we compared H3.3-core and H3 levels to published data sets of active histone modifications[3]. We found that, relative to H3, H3.3-core was strongly associated with several active modifications (**Fig. 1c,d** and **Supplementary Fig. 1** online). Furthermore, we observed strong associations despite using different pull-down methods (MNase-treated noncrosslinked nuclei and native DNA versus formaldehyde-treated nuclei, sonication and PCR-amplified DNA), suggesting that these relationships have a biological basis.

A limitation of cDNA-based profiling is that it does not uncover inter- and intragenic patterns. Therefore, we used a tiling array consisting of 60-mer oligonucleotides spaced 40 bp apart covering all annotated sequences of *D. melanogaster* chromosome arm 2L, chromosome 4 and selected regions from proximal chromosome 2R and distal chromosome X. As we did for the cDNA array, we used input and pull-down fractions from S2 cells stably transformed with biotin-tagged H3.3-core, H3.3 and H3. In all cases, we obtained sufficiently high yields of pull-down material to allow for microarray hybridization without PCR amplification. **Figure 2a** shows a map of the H3.3 density (log ratio) for a typical gene-rich region comprising

∼0.5% of loci on the array, where the 10-kb bar spans a distance equivalent to 100 60-mers. The appearance of continuous peaks and valleys shows prominent patterns of H3.3 deposition. The reproducibility of these patterns is confirmed by the very similar appearance of a biological replicate (**Fig. 2b**). These patterns were not obviously affected by removal of the N-terminal tail, because H3.3-core had very similar patterns (**Fig. 2c**). In contrast, the distribution of H3 was relatively featureless (**Fig. 2d**). We conclude that replication-independent deposition of H3.3 can generate patterns along the chromosome.

We next compared H3.3 patterns with the distribution of a histone modification and of active transcription on tiling arrays. For chromosome arm 2L, data sets are available from chromatin immunoprecipitation experiments assayed on tiling microarrays at 1.5-kb resolution[3,15]. Both dimethylation at Lys4 of H3 (H3K4me2) and RNA polymerase II (Pol II; using an antibody to the C-terminal domain) had patterns that corresponded closely with those of H3.3 (**Fig. 2e,f**). We also observed this similarity between H3.3 and markers of active chromatin and transcription for other gene-rich regions. Therefore, replication-independent deposition of H3.3 marks actively transcribed genes genome-wide and corresponds to patterns of active chromatin modification.

Transposons have distinctive chromatin profiles

In addition to genes, 684 annotated transposons are represented on each tiling array. These fall into 88 families, 12 of which are represented by at least five elements. Most elements in a family are sufficiently similar in DNA sequence that the 'hybridizable' DNA from these elements throughout the genome will efficiently anneal to most copies represented on the array. To determine H3.3 levels in transposon families, we aligned all elements of approximately equal lengths from each family and plotted H3.3 profiles, using the signal ratio of H3.3 to H3 hybridized on the same array. Each element family had a characteristic replication-independent replacement signature (**Fig. 3** and **Supplementary Fig. 2** online). Some element families, such as copia and 297, had strongly positive H3.3/H3 log ratios, indicative of H3.3 enrichment, whereas others, such as HMS Beagle and opus, had strongly negative log ratios, indicative of H3.3 depletion relative to the genome as a whole. We observed this wide range not only for LTR retrotransposons, which are the most abundant type in the *D. melanogaster* genome, but also for non-LTR retrotransposons and DNA transposons (**Fig. 3** and data not shown). We interpret these differences between transposon families and variations in a transposon
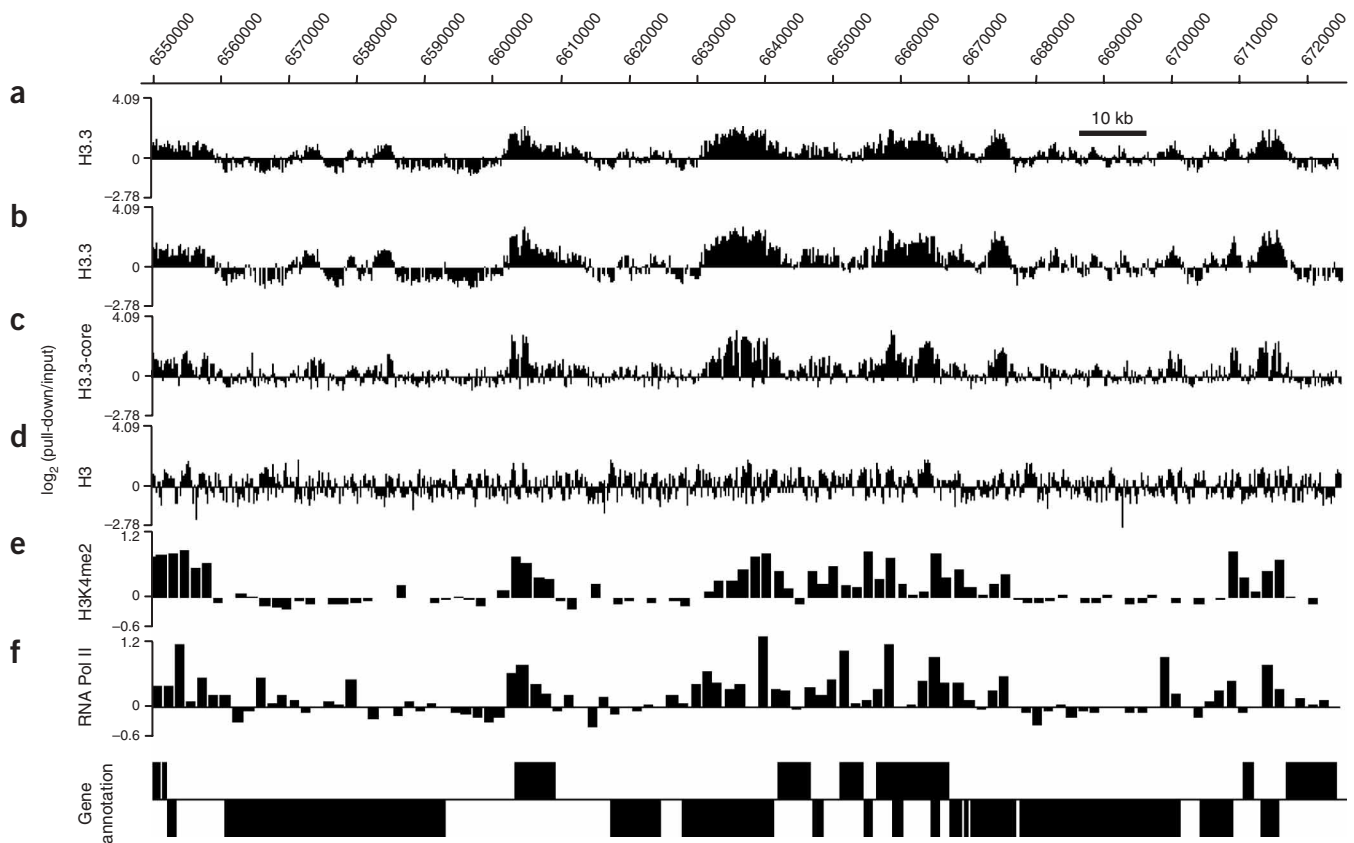
1091

Figure 2 H3.3 enrichment corresponds to H3K4me2 and RNA Pol II at genes. An example of ~0.5% of the tiling array showing single log₂ signal ratios (pull-down/input) represented by bars. (**a**,**b**) H3.3 profiles for two biological replicates, (**c**) a profile for H3.3-core and (**d**) a profile for H3. (**e**,**f**) $\log_2$ signal ratio profiles for (**e**) H3K4me2 (ref. 3) and (**f**) RNA Pol II (ref. 15). Annotated genes from FlyBase release 3.2.2 are shown for comparison; the extent of each gene oriented 5′ to 3′ from left to right is indicated by a box above the line and that of each gene oriented 5′ to 3′ from right to left by a box below the line. The scale at the top indicates FlyBase chromosome 2L coordinates.

(*e.g.*, Tirant) to represent different fractions of transcriptionally active copies genome-wide. A similar interpretation was proposed to account for quantitative differences in histone modifications at mouse transposable elements[16]. H3.3 is rich in euchromatin and poor in heterochromatin[6]; therefore, transposon families with most members in euchromatin may have high H3.3/H3 ratios, and those that are common in (mostly unannotated) heterochromatin may have low H3.3/H3 ratios.

Histone densities change with RNA Pol II levels

Transcription rates are inversely proportional to core histone density in yeast, suggesting that nucleosomes become depleted as a result of the transcriptional process[17,18]. We asked whether a similar relationship occurs for *D. melanogaster* genes. We used MNase digestion to release mostly mononucleosome particles for extraction and DNA hybridization analysis, allowing us to estimate relative nucleosome density on the basis of signal intensity in the input channel for each tiling array. We plotted a moving average of input signal intensities for genic transcription units (that were experimentally determined; FlyBase 3.2.2) as a function of RNA Pol II density averaged over the transcription unit, comparable to an analysis of yeast transcription rate versus histone density data[18]. For all input data sets, we observed similar changes in nucleosome density: genes with low RNA Pol II density showed little change in nucleosome density, whereas genes with increasing RNA Pol II densities showed gradual depletion (**Fig. 4**).

We applied the same procedure to measure the densities of tagged histones in the pull-down channels as a function of RNA Pol II densities. H3.3-core and H3.3 densities begin to increase at low to moderate RNA Pol II densities, eventually leveling off at high RNA Pol II densities (**Fig. 4**). Relative to input, these profiles resembled those seen for active histone modifications in *D. melanogaster* cells[3]. We conclude that replication-independent assembly of H3.3 is coupled to transcription genome-wide.

We also observed that the H3 signal intensity increased slightly with increasing RNA Pol II density (**Fig. 4**). This was unexpected, because uniform replication-coupled deposition of H3 should give a profile similar to that for input, and previous studies did not detect replication-independent deposition of H3 (refs. 6,9,19). It is possible that a small amount of replication-independent deposition of H3 takes place at active genes as a result of its constitutive production from a Cu⁺⁺ inducible promoter rather than from its native S-phase limited promoter. A low level of constitutive replication-independent incorporation of H3 seems possible, as only three amino acids on the core that differ between H3 and H3.3 target the histone for one or the other pathway[6]. Replication-independent deposition of H3 would mask its expected depletion after replacement by H3.3 (**Fig. 4**). Such masking could account for the lack of visible depletion of H3 in individual regions that are rich in H3.3 (**Fig. 2**). Minor replication-independent deposition of H3 could also account for the slight positive slope in the comparison between H3 and active
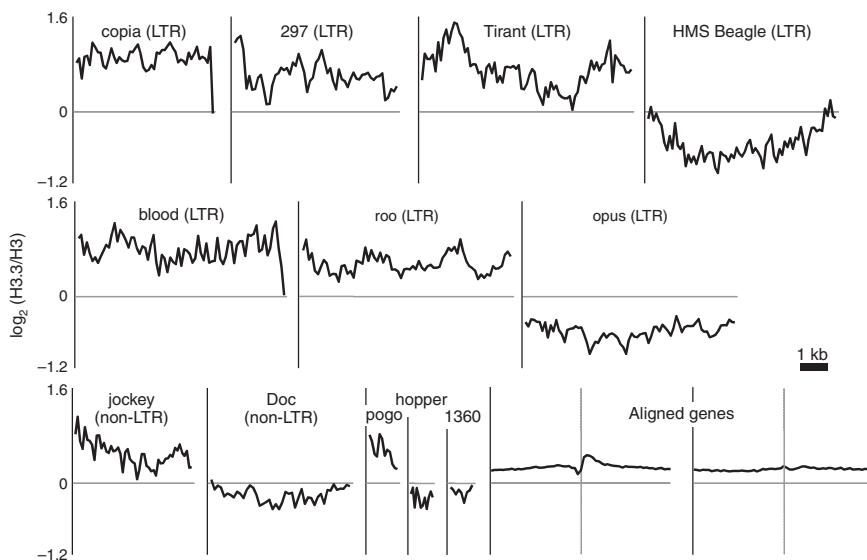
Figure 3 Transposons have distinct chromatin patterns. Twelve transposon families have distinct H3.3 enrichment or depletion patterns. $\log_2$ ratios of H3.3/H3 are shown. Cy3-labeled H3.3 and Cy5-labeled H3 were hybridized on the same array to allow a direct comparison, although very similar profiles were obtained for ratios of H3.3/input to H3/input and H3.3-core/input to H3/input from multiple biological replicates (**Supplementary Fig. 2** online). For comparison, average gene ratio profiles from the same data set were aligned at 5′ and 3′ ends and plotted as in **Figure 5a** on the same y axis scale (lower right panels).

modifications (**Fig. 1d** and **Supplementary Fig. 1** online). In any case, identically tagged H3.3-core and H3.3 show much greater increases in signal intensity than H3 controls (**Fig. 4**).

Active genes show pronounced H3.3 replacement patterns

To determine whether there are replication-independent assembly patterns common to genes, we calculated average chromatin profiles using GenBank annotations from release 3.2.2 of the *D. melanogaster* genome. We found that 2,865 transcription units included annotated 5′ and 3′ ends of genes represented on each tiling array, indicating that these annotations were based on full-length cDNAs. We aligned all 2,865 genes at their 5′ and 3′ ends and averaged the pull-down to input signal ratios across the genes from 60-mers on either side out to 3 kb (half the average transcription unit length; **Supplementary Fig. 3** online). As expected from the enrichment of H3.3 at sites of H3K4me2 and RNA Pol II enrichment (**Fig. 2**), H3.3 had a higher signal ratio than H3 throughout transcribed regions (**Fig. 5a**). We obtained similar results for H3.3-core, indicating that the difference between H3.3-core and H3 is attributable to replication-independent assembly and not to some inadvertent consequence of removal of the N-terminal tail to prevent replication-coupled assembly. Unexpectedly, however, we found that enrichment extended beyond genes on either side. Enrichment was specific for genes, because the 684 transposons

on the array showed only background levels of H3.3 when similarly aligned.

To some extent, gene enrichment patterns represent contributions from adjacent, over-lapping and nested genes. Therefore, we excluded 60-mers from the averages of aligned genes where either the 5′ or 3′ end overlapped the annotated portion of an adjacent gene (**Supplementary Fig. 3** online) and replotted the gene profiles. This exclusion greatly reduced the 5′ upstream peak (**Fig. 5b**), suggesting that clustering of divergently transcribed genes was responsible for much of the 5′ upstream enrichment of H3.3. Nevertheless, enrichment of H3.3 relative to H3 persisted both upstream and downstream.

Enrichment of H3.3 was prominent in the coding region, peaking at approximately position +300 bp. We considered the possibility that this peak primarily represents an enrichment of H3.3 in small genes that do not extend beyond ∼1 kb. To minimize this and other complications arising from averaging genes of different lengths, we determined signal ratio profiles for all genes ≥6 kb in length, because each of them contributes once and only once to every plotted interval. Although the small peak just upstream of the 3′ end disappeared, the large 5′ peak was essentially unchanged for genes ≥6 kb (**Fig. 5c**). Notably, the H3.3 signal ratio profile for large genes showed a discontinuity between position +3 kb of aligned 5′ ends and position −3 kb of aligned 3′ ends (**Fig. 5c**); there was no discontinuity for H3 or for signal ratios for all genes (**Fig. 5a,b**). We attribute this feature to the gradual reduction of replication-independent replacement over the entire body of genes, and longer-than-average genes would show

Figure 4 Changes in nucleosome and histone density profiles with increasing RNA Pol II density. A moving average of raw signal intensities with window size of 80 genes and step size of 1 gene was calculated[18] for 1,643 genes ordered by increasing RNA Pol II density averaged over each transcription unit. Gray curves represent the indicated input $\log_2$ signal intensities; curves of other colors represent pull-down $\log_2$ signal intensities. To facilitate comparison on the same log scale, experimental differences between data sets were minimized by equalizing the lowest RNA Pol II value in each data set (adding or subtracting a constant intensity value to each point in the data set), which resulted in an arbitrary $\log_2$ signal intensities baseline. $\log_2$ signal intensities of H3.3 (orange curve), H3.3-core (red curve) and H3 (blue curve), represent averages from two biological replicates. Arrows indicate pairing of each pull-down with its respective input measured on the same array, which differ from one another because of experimental variation. Similar profiles were obtained for a dye swap of one of the H3.3-core data sets (data not shown).
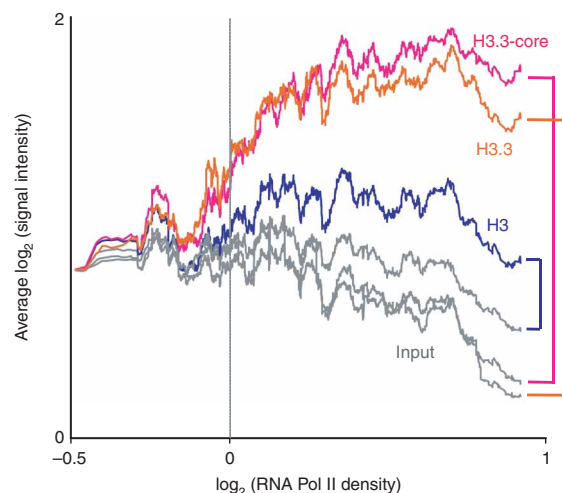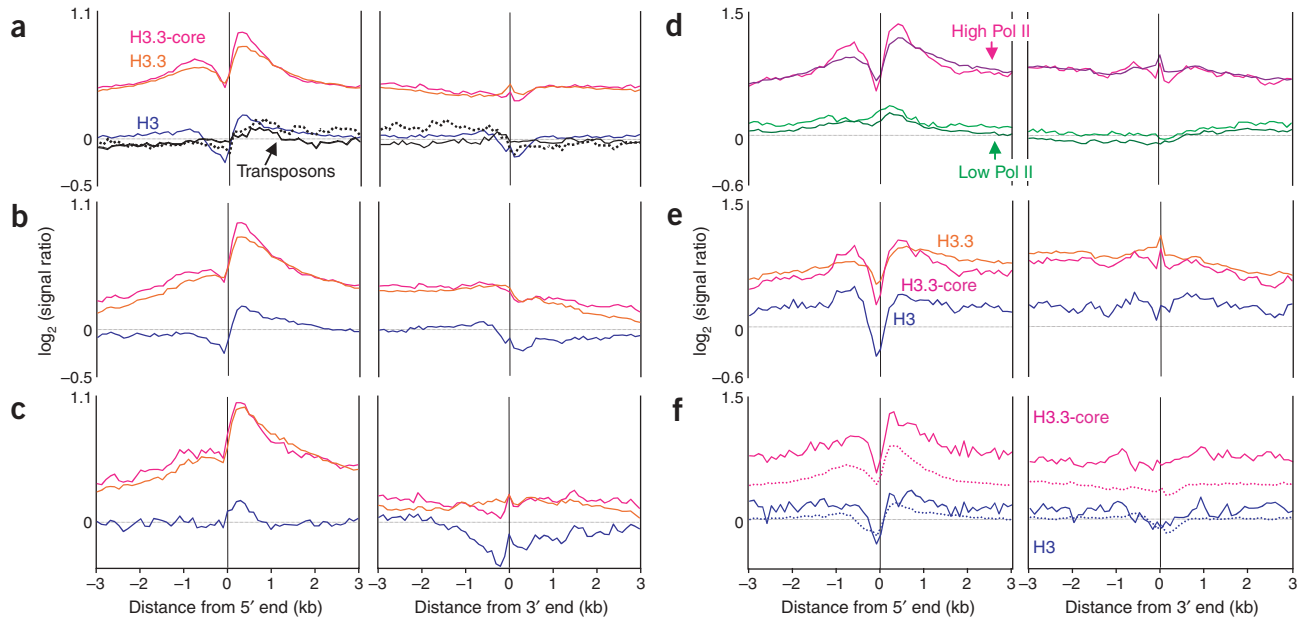
Figure 5 H3.3 shows enrichment patterns over genes and flanking sequences. (**a**) H3.3-core (red curves), H3.3 (orange) and H3 (blue) signal ratio profiles (solid lines) for all 2,865 genes with annotated 5′ (left) and 3′ (right) ends. We used 684 similarly aligned transposons as controls (solid black for H3.3-core and dotted for H3). (**b**) H3.3-core (red curves), H3.3 (orange) and H3 (blue) signal ratio profiles after removal of the contributions from neighboring and overlapping genes. (**c**) Same data as shown in **b** for genes ≥ 6 kb. (**d**) H3.3 and H3.3-core profiles calculated separately for the subsets of genes (one-third each) with the highest (purple curves for H3.3 and magenta for H3.3-core) and lowest (dark green curves for H3.3 and light green for H3.3-core) integrated RNA Pol II levels for all 1,643 genes on chromosome 2L with annotated 5′ and 3′ ends. (**e**) $\log_2$ ratios of the subsets of genes (one-third each) with the highest versus lowest RNA Pol II densities (calculated as high RNA Pol II $\log_2$ (pull-down/input) – low RNA Pol II $\log_2$ (pull-down/input); solid lines). (**f**) H3.3-core (red) and H3 (blue) log ratios for 168 X-chromosome genes on the array (solid lines) compared with 2,697 autosomal genes (dotted lines).

continuing reduction through the middle interval. We conclude that the gradual reduction in H3.3 signal throughout transcribed regions reflects transcription-coupled replacement of nucleosomes, consistent with inferences drawn from studies of long yeast genes[20,21]. H3.3 levels near the 3′ ends of long genes were substantially lower than levels upstream of promoters, in agreement with the comparison of H3.3 levels at promoters and downstream regions of several human genes[10]. Upstream enrichment of H3.3 suggests that replication-independent assembly is not just a consequence of transcriptional elongation through genes but is also a process with potential for gene regulation.

We expected that specific H3.3 features would be more prominent in active versus inactive genes because of transcription-coupled replication-independent assembly[6,8,9]. Therefore, we rank-ordered genes on the basis of RNA Pol II density, divided them into three subsets and compared signal ratio profiles for the genes in the subsets with the highest (presumed mostly active) versus the lowest (presumed mostly inactive) RNA Pol II density. We observed a marked enrichment in the H3.3 signal ratio profile for active genes from far upstream to far downstream of aligned genes (**Fig. 5d**). In contrast, the H3.3 signal ratio profile for the subset with low RNA Pol II density was close to background levels. We obtained similar results when we compared the quarters with the highest and lowest RNA Pol II levels or rank-ordered genes on the basis of H3K4me2 level (data not shown). We conclude that active genes and their flanking regions are sites of pronounced replication-independent replacement.

The detection of high levels of replication-independent replacement both upstream and downstream of active gene promoters (**Fig. 5d**) led us to wonder whether nucleosomes are removed from active *D. melanogaster* promoters, as has been documented for yeast promoters[22–24]. We confirmed this possibility by plotting the log ratios

of subsets with high versus low RNA Pol II density for all H3 forms (**Fig. 5e**), which should cancel out gene-specific differences that are not related to transcriptional activity. We found that the sharp dip at 0 to –100 bp was a consistent feature of genes. We conclude that promoters of active genes are relatively depleted of both H3- and H3.3-containing nucleosomes.

H3.3 is enriched on the male X chromosome

Cytological studies identified replication-independent replacement at highly transcribed genes[6,8,9]. In these cases, high levels of transcription might cause almost complete nucleosome depletion followed by replacement with H3.3-containing nucleosomes[9]. Alternatively, a more gradual replacement process might be occurring at most transcribed genes. For example, a small fraction of H3/H4 subunits might be lost as transcription proceeds, and this damage would be repaired during reassembly of nucleosomes in the wake of RNA polymerase by insertion of new H3.3/H4 subunits[26]. Gradual replacement is consistent with the graded increase in H3.3 enrichment that we observed with increasing RNA Pol II abundance (**Fig. 4**). Such a gradual replacement process would mean that H3.3 levels can provide a sensitive measure of differences in the number of rounds of transcriptional elongation.

Twofold differences in levels of transcriptional elongation have been proposed to underlie *D. melanogaster* X-chromosome dosage compensation[27–29]. This leads to the expectation that H3.3 will be enriched on the male X chromosome relative to autosomes. We can test this expectation because our study used a male *D. melanogaster* cell line, S2 cells[30], and our tiling arrays represent ∼15% of the annotated X chromosome. We found an enrichment of H3.3 over the 168 annotated X-chromosome genes relative to 2,697 autosomal genes
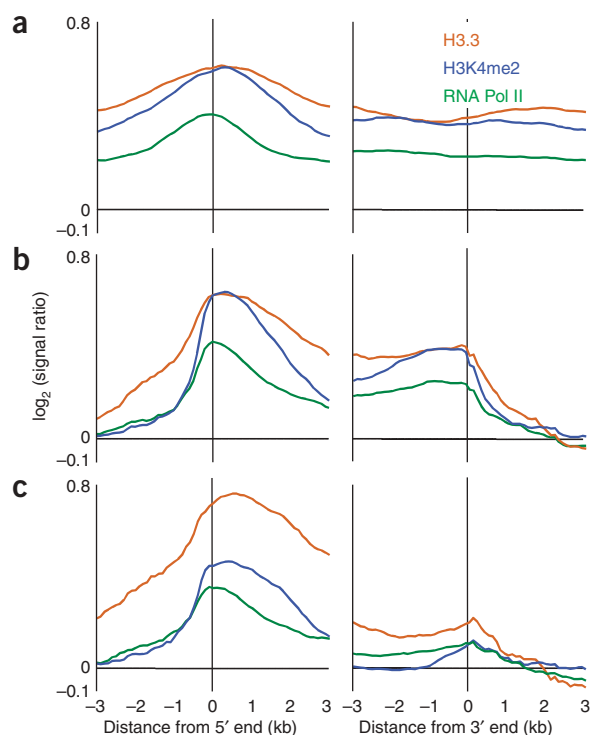
a

b

c

Figure 6 H3.3, RNA Pol II and H3K4me2 show similar enrichment patterns from upstream to downstream of transcription start sites. (**a**) $\log_2$ ratio for all 1,643 chromosome 2L genes with both 5′ and 3′ annotated ends that are represented in all data sets. (**b**) Same data as shown in **a** after removal of the contributions from neighboring and overlapping genes. (**c**) Same data as shown in **b** for 397 genes $\geq$ 6 kb in length.

(**Fig. 5f**) or relative to 155 genes in the corresponding distal stretch of chromosome 2L (**Supplementary Fig. 4** online). This relative enrichment of H3.3 on the male X chromosome genes occurred throughout the transcription unit and both upstream and downstream, with a profile similar to but elevated above that for autosomal H3.3. The correspondence of H3.3 levels to twofold transcriptional upregulation of X-chromosome genes further indicates that replication-independent assembly is a gradual rather than an all-or-none replacement process.

H3.3 deposition corresponds to active histone modification

The similarities between H3.3, RNA Pol II and H3K4me2 densities along the chromosome (**Fig. 2**) led us to ask whether they were also similarly distributed at genic resolution. Therefore, we aligned the 5′ and 3′ ends of the 1,665 genes from chromosome arm 2L for which RNA Pol II and H3K4me2 tiling data were available[3,15] and averaged their densities. For each ∼1.5 kb interval for which these data were available, we also calculated the average H3.3 density by integrating over tiled 60-mers encompassed by the interval. In this way, we could compare profiles for replication-independent assembly, RNA Pol II and an active modification on the same scale at ∼1.5 kb resolution. These profiles were generally similar to one another (**Fig. 6**). Like H3.3, RNA Pol II and H3K4me2 showed broad enrichment over the 5′ end of genes (**Fig. 6a**). When overlaps were removed to reduce the contribution of neighboring genes, the three profiles changed in concert (**Fig. 6b**). We observed gradual decreases in density downstream of the 5′ end in all three cases. These decreases seemed to occur throughout the gene region, because when we plotted profiles for large ($\geq$6 kb) genes, we observed discontinuities for RNA Pol II and

H3K4me2 that were similar to those for H3.3 (**Fig. 6c**). This suggests that transcriptional elongation, replication-independent assembly and enrichment of H3K4me2 are coupled[25].

DISCUSSION

We used high-resolution mapping of nucleosomes and tagged histone variants to explore the relationship between nucleosome assembly pathways and genomic features. Both cDNA and genomic tiling arrays showed a close correspondence between histone H3.3 and active histone modifications at the level of single genes. Our high resolution analysis also showed that genes have distinct chromatin features that might be no larger than single nucleosomes, including a sharp dip at the promoters of active genes (**Fig. 5e**). Although nucleosomes have previously been reported to be depleted at particular active promoters and genome-wide for yeast genes in general[17,18,22,23,31], our survey of roughly one-third of the *D. melanogaster* genome indicates that this is a consistent feature of active, but not inactive, genes in *D. melanogaster* and suggests that it is general for eukaryotes.

H3.3 replacement profiling is a new tool for identifying epigenetic patterns[32]. Previous large-scale studies used antibodies to histone modifications for this purpose[3,4,33]. But the complexity of histone modifications and the variable affinity of different antibodies can lead to ambiguities that may limit their utility in developing useful public resources analogous to genomic sequences. In contrast, H3.3-versus-H3 patterns are simple marks of active transcription that capture much of what is detected using antibodies (**Figs. 2** and **6**). These replication-independent and replication-coupled variants are universal for plants and animals, they are extraordinarily conserved (H3.3 is 100% identical in flies and humans) and, as we show here, biotin-tagged nucleosomes can be purified in a single step. Streptavidin binds biotin with an affinity that is orders of magnitude higher than that of binding of typical antibodies used for chromatin affinity purification to modified histones[14]. Furthermore, regulated expression of biotin ligase and H3.3 should allow for tissue-specific and time-restricted epigenetic profiling, which is not feasible when profiling histone modifications by chromatin immunoprecipitation.

We show the value of our profiling strategy by identifying the enrichment of H3.3 on male X-chromosome genes. X-chromosome enrichment occurs uniformly from upstream to downstream relative to the H3.3 enrichment over autosomes (**Fig. 5f** and **Supplementary Fig. 4** online), consistent with the idea that replication-independent replacement is a gradual process that is sensitive to twofold differences in transcription levels. This shows that H3.3 abundance can be used to measure levels of dosage compensation, which thus far has been assayed primarily by measuring gene products[27–29]. Unequivocal H3.3 enrichment was shown using only ∼5% of the data on our array, which was not specifically designed to test models for dosage compensation. Therefore, our profiling system and our analysis strategy should be broadly applicable to investigations of this and other developmental processes, providing a general epigenomic strategy for plants and animals.

The genome-wide patterns of H3.3 incorporation that we detected for genes and transposons imply a possible mechanism for the generation of histone modification patterns. We propose that replication-independent assembly gradually deposits H3.3 with each round of transcription and that this replacement process leads to enrichment of active modifications throughout the cell cycle. After replication, H3.3 and its associated modifications would be diluted by H3, and then the enrichment process would begin again. Whether enrichment of active modifications results from the action of modifying enzymes on pre-assembled H3.3/H4 subunits or on H3.3-containing nucleosomes after

they are assembled is unknown, but in some cases, both H3 and H4 are modified in assembly complexes[34,35], consistent with the idea that histone modification patterns are generated when histones are deposited.

We propose that epigenetic patterns reflect the balance between replication-coupled assembly, which dilutes H3.3 nucleosomes by a factor of two every cell cycle, and replication-independent assembly, which replaces nucleosomes at different rates at active genes. In this way, different rates of replication-independent replacement would lead to different relative abundances of H3.3, which is enriched in active modifications, and of H3, which is enriched in silent modifications. The sharp increase and gradual decrease of H3.3 beginning at the 5′ ends of active genes would represent different rates of turnover during transcriptional elongation, as observed in yeast genes[20,21]. At the most highly transcribed genes, such as at puff sites in *D. melanogaster*[9] and rDNA genes[6,36], transcription-coupled replacement would be at a higher level than what we see here for average genes that are much more weakly expressed. The prominent dip at active promoters would represent depletion by eviction or sliding of nucleosomes, as observed in yeast[17,18,22–24]. Our results are also consistent with the detection of higher levels of H3.3 upstream than downstream for a selected set of human genes[10]: with both a probe density (∼1 probe per 10 kb versus 1 probe per 100 bp) and a number of genes (2,685 versus 18) roughly two orders of magnitude higher than in the previous study[10], our analysis identified general features of chromatin that were not detected in that study.

We were surprised to find that H3.3 was enriched far upstream and downstream of transcribed regions. We speculate that intergenic transcription is the process responsible for replication-independent assembly over such long distances from genes[37,38]. Recent genome-wide studies indicate that intergenic transcription is widespread, but the transcripts themselves might have no function[39,40]. For example, at the *D. melanogaster* bithorax complex, intergenic transcription derepresses developmental regulatory genes by transcribing through *cis*-acting elements necessary for silencing by Polycomb protein complexes[41–43]. Derepression seems to be caused by the process of transcription itself and not by the presence of the resulting transcript[43]. Notably, transcription through several predicted *cis*-acting regulatory elements shows the same developmental restrictions as the closely adjacent genes that they are likely to regulate. Thus, a new paradigm for developmental regulation is that key regulatory genes are kept silent by default, and transcription factors activate them by promoting upstream transcription. For these genes to remain permanently active, however, there must be a process that remembers the active state through the cell cycle. H3.3/H4 deposition in the wake of the activating polymerase can potentially provide a general mechanism for developmental memory: transcription both displaces histones that are deposited during replication bearing default silencing modifications and inserts new histones that are enriched in active modifications.

METHODS

Plasmid constructs, cell culture and transformation. We used a plasmid vector (pRMHA3; ref. 44) with a cloning site downstream of the mouse metallothionein promoter to insert both the histone gene constructs and the *E. coli* BirA biotin ligase gene. We fused coding sequences for histone H3, H3.3 and H3.3-core (with amino acids 3–35 inclusive deleted) described previously[6] at their 5′ ends in-frame with a sequence encoding the 23–amino acid BLRP[45] followed by a TEV cleavage site (T. Furuyama, personal communication). We inserted this chimeric sequence into the plasmid vector downstream of the metallothionein promoter. We similarly inserted the BirA gene into a vector with a puromycin-resistance gene for selection in *D. melanogaster* S2 cells[46]. We grew cells at 25 °C in serum-free medium (Invitrogen) supplemented with L-glutamine[47] and cotransfected them with a 19:1 mixture of plasmids encoding

histone and plasmids encoding BirA by electroporation[48]. We selected cells that were stably transformed with plasmids and maintained them in serum-free medium supplemented with L-glutamine, 5% fetal bovine serum and 4 µg ml⁻¹ puromycin. We simultaneously induced the metallothionein promoters on both plasmids by adding 0.5 mM CuSO₄ at early log phase and grew cells for 2–3 d before collecting them (**Supplementary Fig. 5** online). To confirm the correct behavior of constructs, we checked expression levels by western-blot analysis (**Supplementary Fig. 6** online). We also confirmed correct localization by staining of cytological preparations using streptavidin–Alexa 488 (Molecular Probes); biotin–H3.3-core localized to euchromatin and biotin-H3 localized to DAPI-rich regions[6] (data not shown).

Chromatin and DNA preparation. We prepared nuclei as described[11] and extracted chromatin (input) using a modification of a previously described procedure[49]. We purified biotinylated histone-containing nucleosome particles (pull-down) and obtained essentially complete recovery in all cases (**Supplementary Fig. 6** online and data not shown). For microarray analysis, we extracted DNA from both input and pull-down samples (**Supplementary Fig. 7** online).

To ascertain the specificity of biotinylation and streptavidin pull-down, we transformed cells using a mutant version of BLRP in which the lysine target of biotinylation on the H3.3-core chimeric protein was changed to an arginine. We treated this mutant BLRP control cell line in parallel with wild-type BLRP lines. In the mutant control pull-downs, we detected less than 2% of the amount of DNA detected in wild-type BLRP pull-downs (≤0.1 µg of mutant BLRP–H3.3-core versus 7–14 µg of wild-type BLRP–H3 or –H3.3-core pulled down in parallel).

Microarray analyses. For cDNA analysis, we used random priming[50] and hybridization[3] to spotted arrays containing 11,858 annotated *D. melanogaster* genes (**Fig. 1** and **Supplementary Fig. 1** online). Three biological replicates yielded similar results. The cDNA products of most of these genes, and the array fabrications and hybridizations, are essentially the same as previously used[3], allowing for rigorous comparisons between variant and modification abundances. Genomic profiling was done by NimbleGen Systems as part of a Chromatin Immunoprecipitation Array Service[33]. Arrays were designed to provide 100-bp resolution by tiling 60-mers with 40-bp spacing through all annotated sequence for chromosome arm 2L (22.2 Mb), chromosome 4 (1.2 Mb), the proximal half of 2R (11.0 Mb), the distal tip of X (2.0 Mb) and assorted repeats and single-copy regions from elsewhere in the genome (NimbleGen Systems, Inc.). A total of 378,912 tiling 60-mers and 10,000 random 60-mers were present on each array. We provided DNA from input (3 µg) and pull-down (4 µg) samples to NimbleGen Systems for differential labeling by random priming with Cy3 or Cy5 and hybridization to oligonucleotide arrays. For each experiment, input and pull-down channel signal intensities and scaled log₂ ratios were provided by NimbleGen Systems. Input and pull-down signals were compared to signals for the 10,000 random 60-mers on the array, and, on the basis of this comparison, the signals were judged to be adequate in the context of continuous tiling for all arrays.

Computational analyses. We analyzed cDNA arrays as described[3]. We identified common loci between our data sets and those previously reported[3] and excluded loci with weak or missing signals in either data set from the analysis.

For tiling arrays, we visually examined log₂ ratio profiles from each array hybridization and compared them to other data sets and annotations (converted to gff format) using NimbleGen SignalMap (*e.g.*, **Fig. 2**). We downloaded RNA Pol II and H3K4me2 log₂ ratios from a ∼1.5 kb tiling of chromosome arm 2L (ref. 15) and averaged RNA Pol II values for each gene with experimentally determined 5′ and 3′ ends (based on annotated transcription units from FlyBase release 3.2.2) in common between the two tiling designs. We aligned array data sets by chromosome position and, for each gene, weighted the log₂ ratios by the amount of overlap with the gene region for averaging. For each of the 1,643 genes, we rank-ordered these averaged RNA Pol II values and used them to divide genes into classes with high and low values.

For analysis of the 2,865 gene and 684 transposon 5′ and 3′ ends, we identified the 100-bp tiling interval and overlapping 60-mers containing each end and aligned the 30 60-mers on either side. This allowed calculation of

average input, pull-down and log$_2$ ratios for the 3 kb on either side of the 5′ and 3′ ends at 100-bp resolution from values assigned to the 60-mers. To exclude contributions from overlapping genes in some analyses, we truncated 60-mer alignments at the 5′ or 3′ end of the adjacent gene on either side.

To avoid effects of G+C content on hybridization efficiency, we equalized the 60-mer G+C content in each 100-bp interval when averaging aligned genes (**Figs. 5 and 6** and **Supplementary Fig. 4** online). We excluded the most AT-rich 60-mers from each AT-rich interval to increase GC richness or did the opposite to decrease GC richness. In this way, each interval would be represented by a set of 60-mers with precisely the same average GC content as on the array as a whole (42.17%).

For analysis of transposon families, we manually selected all annotated members represented on the tiling array that seemed to be intact. For each family, we aligned 60-mers starting at the 60-mer nearest to the annotated 5′ end and averaged the scores assigned to 60-mers over the body of the transposon, as we did for end analysis.

URLs. RNA Pol II and H3K4me2 log$_2$ ratios from a ∼1.5 kb tiling of chromosome arm 2L were downloaded from http://chromosome2l.mit.edu/. FlyBase release 3.2.2 was downloaded from http://www.flybase.net/.

Accession codes. GEO: GSE3031.

Note: Supplementary information is available on the Nature Genetics website.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Wolffe, A.P. *Chromatin: Structure and Function* (Academic, San Diego, 1992).
2. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
3. Schubeler, D. *et al.* The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**, 1263–1271 (2004).
4. Bernstein, B.E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
5. Kurdistani, S.K., Tavazoie, S. & Grunstein, M. Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721–733 (2004).
6. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).
7. Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116**, 51–61 (2004).
8. Janicki, S.M. *et al.* From silencing to gene expression: Real-time analysis in single cells. *Cell* **116**, 683–698 (2004).
9. Schwartz, B.E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev.* **19**, 804–814 (2005).
10. Chow, C.M. *et al.* Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep.* **6**, 354–360 (2005).
11. McKittrick, E., Gafken, P.R., Ahmad, K. & Henikoff, S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl. Acad. Sci. USA* **101**, 1525–1530 (2004).
12. Waterborg, J.H. Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. *J. Biol. Chem.* **265**, 17157–17161 (1990).
13. Johnson, L. *et al.* Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res.* **32**, 6511–6518 (2004).
14. de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **100**, 7480–7485 (2003).
15. MacAlpine, D.M., Rodriguez, H.K. & Bell, S.P. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* **18**, 3094–3105 (2004).
16. Martens, J.H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
17. Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O. & Schreiber, S.L. Global nucleosome occupancy in yeast. *Genome Biol.* **5**, R62 (2004).
18. Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D. & Lieb, J.D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
19. Ahmad, K. & Henikoff, S. Centromeres are specialized replication domains in heterochromatin. *J. Cell Biol.* **153**, 101–110 (2001).
20. Schwabish, M.A. & Struhl, K. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **24**, 10111–10117 (2004).
21. Kristjuhan, A. & Svejstrup, J.Q. Evidence for distinct mechanisms facilitating transcript elongation through chromatin *in vivo*. *EMBO J.* **23**, 4243–4252 (2004).
22. Boeger, H., Griesenbeck, J., Strattan, J.S. & Kornberg, R.D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11**, 1587–1598 (2003).
23. Reinke, H. & Horz, W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11**, 1599–1607 (2003).
24. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
25. Workman, J.L. & Abmayr, S.M. Histone H3 variants and modifications on transcribed genes. *Proc. Natl. Acad. Sci. USA* **101**, 1429–1430 (2004).
26. Henikoff, S., Furuyama, T. & Ahmad, A. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet.* **20**, 320–326 (2004).
27. Henikoff, S. & Meneely, P. Unwinding dosage compensation. *Cell* **72**, 1–2 (1993).
28. Smith, E.R., Allis, C.D. & Lucchesi, J.C. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J. Biol. Chem.* **276**, 31483–31486 (2001).
29. Straub, T., Dahlsveen, I.K. & Becker, P.B. Dosage compensation in flies: Mechanism, models, mystery. *FEBS Lett.* **579**, 3258–3263 (2005).
30. Copps, K. *et al.* Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. *EMBO J.* **17**, 5409–5417 (1998).
31. Roh, T.Y., Ngau, W.C., Cui, K., Landsman, D. & Zhao, K. High-resolution genome-wide mapping of histone modifications. *Nat. Biotechnol.* **22**, 1013–1016 (2004).
32. van Steensel, B. Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat. Genet.* **37** Suppl, S18–S24 (2005).
33. Kirmizis, A. *et al.* Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**, 1592–1605 (2004).
34. Sarraf, S.A. & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15**, 595–605 (2004).
35. Tyler, J.K. *et al.* The RCAF complex mediates chromatin assembly during DNA replication and repair. *Nature* **402**, 555–560 (1999).
36. Thiriet, C. & Hayes, J.J. Replication-independent core histone dynamics at transcriptionally active loci *in vivo*. *Genes Dev.* **19**, 677–682 (2005).
37. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus. *Mol. Cell* **5**, 377–386 (2000).
38. Lipshitz, H.D., Peattie, D.A. & Hogness, D.S. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev.* **1**, 307–322 (1987).
39. Wyers, F. *et al.* Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
40. Johnson, J.M., Edwards, S., Shoemaker, D. & Schadt, E.E. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* **21**, 93–102 (2005).
41. Hogga, I. & Karch, F. Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of Polycomb-mediated silencing. *Development* **129**, 4915–4922 (2002).
42. Rank, G., Prestel, M. & Paro, R. Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol. Cell. Biol.* **22**, 8026–8034 (2002).
43. Schmitt, S., Prestel, M. & Paro, R. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.* **19**, 697–708 (2005).
44. Bunch, T.A., Grinblat, Y. & Goldstein, L.S. Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. *Nucleic Acids Res.* **16**, 1043–1061 (1988).
45. Beckett, D., Kovaleva, E. & Schatz, P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999).
46. Iwaki, T., Figuera, M., Ploplis, V.A. & Castellino, F.J. Rapid selection of *Drosophila* S2 cells with the puromycin resistance gene. *Biotechniques* **35**, 482–484, 486 (2003).
47. Echalier, G. *Drosophila Cells in Culture* (Academic, New York, 1997).
48. 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).
49. Blower, M.D., Sullivan, B.A. & Karpen, G.H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002).
50. Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46 (1999).