ARTICLES

Nucleosome organization in the *Drosophila* genome

Travis N. Mavrich[1,2]*, Cizhong Jiang[1,2]*, Ilya P. Ioshikhes[3], Xiaoyong Li[4], Bryan J. Venters[1,2], Sara J. Zanton[1,2], Lynn P. Tomsho[2], Ji Qi[2], Robert L. Glaser[5], Stephan C. Schuster[2], David S. Gilmour[1], Istvan Albert[2] & B. Franklin Pugh[1,2]

Comparative genomics of nucleosome positions provides a powerful means for understanding how the organization of chromatin and the transcription machinery co-evolve. Here we produce a high-resolution reference map of H2A.Z and bulk nucleosome locations across the genome of the fly *Drosophila melanogaster* and compare it to that from the yeast *Saccharomyces cerevisiae*. Like *Saccharomyces*, *Drosophila* nucleosomes are organized around active transcription start sites in a canonical −1, nucleosome-free region, +1 arrangement. However, *Drosophila* does not incorporate H2A.Z into the −1 nucleosome and does not bury its transcriptional start site in the +1 nucleosome. At thousands of genes, RNA polymerase II engages the +1 nucleosome and pauses. How the transcription initiation machinery contends with the +1 nucleosome seems to be fundamentally different across major eukaryotic lines.

Knowledge of the precise location of nucleosomes in a genome is essential to understand the context in which chromosomal processes such as transcription and DNA replication operate. A common theme to emerge from recent genome-wide maps of nucleosome locations is a general deficiency of nucleosomes in promoter regions and an enrichment of certain histone modifications towards the 5′ end of genes[1–7]. A high resolution genomic map of nucleosome locations in the budding yeast *S. cerevisiae* has further revealed the nucleosomal context of *cis*-regulatory elements and transcriptional start sites[1–7]. However, such context has not been established in multicellular eukaryotes, and so fundamental questions remain: Is there a common theme by which genes of multicellular eukaryotes position their nucleosomes with respect to functional chromosomal elements? Are such themes and their underlying rules evolutionarily conserved across eukaryotes? What are the functional implications for those themes that differ across the major eukaryotic lines? To address these questions, we have produced a genome-wide high-resolution map of H2A.Z (also known as H2Av) and bulk nucleosome locations in the embryo of the fruitfly *D. melanogaster*. H2A.Z is widely distributed in *Drosophila*[8], but some evidence points to specialized roles[9,10]. In *Saccharomyces*, H2A.Z replaces H2A at the 5′ end of active genes[11–14], and thus provides a focused representation of promoter chromatin architecture.

Drosophila embryos are composed of a wide variety of cell types in which subsets of genes may elicit distinct gene expression programmes[15,16]. Global gene expression profiles during all stages of *Drosophila* development from 8–12 h post fertilization to a young adult fly are correlated (Supplementary Fig. 1), which possibly reflects the broad expression pattern of the large repertoire of housekeeping genes in most cell types during development[15,16]. This general spatial and temporal independence of gene expression provides impetus to use whole embryos to develop a reference nucleosome map. Indeed, our map reveals that nucleosomes are generally well organized, despite cell type heterogeneity.

'Closed' and 'open' chromatin organization

Embryos were treated with formaldehyde, and H2A.Z nucleosome core particles were immunopurified (Supplementary Figs 2 and 3). H2A.Z-containing nucleosomes (652,738) were sequenced (Supplementary Fig. 4) and mapped to 207,025 consensus locations in the *Drosophila* r5.2 reference genome (Fig. 1a and Supplementary Fig. 2b, see browser at http://atlas.bx.psu.edu/), thereby providing >3-fold depth of coverage (Supplementary Fig. 5). Correction for micrococcal nuclease (MNase) digestion bias was imposed (Supplementary Fig. 6). Those 112,750 nucleosomes detected three or more times were further analysed, although patterns were identical when all nucleosomes were analysed. The internal median error of the data was 4 bp (Supplementary Fig. 7).

Figure 1b displays the predominant embryonic distribution of H2A.Z nucleosomes relative to the transcription start site (TSS) of all coding genes, and is compared to the pattern previously derived from *Saccharomyces*[1]. Patterns around noncoding genes are shown in Supplementary Fig. 8. Eighty five per cent of *Drosophila* coding genes (11,994 of the 14,143) contained at least one H2A.Z nucleosome (detected three or more times) within 1 kb of the TSS. H2A.Z levels correlated with gene expression (Fig. 1c and Supplementary Fig. 9), as has been seen on individual genes and in *Saccharomyces*[12,13,17].

H2A.Z nucleosomes were predominantly distributed at 175-bp intervals from the TSS (compared to 165-bp in *Saccharomyces*[1], Fig. 1b), demonstrating that a predominant organizational pattern exists for H2A.Z nucleosomes in *Drosophila* embryos that transcends a spatial and temporal context. The H2A.Z pattern was compared to the distribution of bulk nucleosomes (that is, those containing any combination of H2A.Z and H2A), determined using high-density tiling arrays (36-bp probe spacing). Within genic regions, the same organizational pattern was found (Supplementary Fig. 10). For both data sets, a nucleosome-depleted region was evident immediately upstream of the +1 nucleosome, which probably reflects a nucleosome-free core promoter region (NFR), as first detected in

[1]Center for Gene Regulation, and [2]Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [3]Department of Biomedical Informatics and Department of Molecular and Cellular Biochemistry, Davis Heart and Lung Research Institute, The Ohio State University, Columbus, Ohio 43210, USA. [4]Berkeley Drosophila Transcription Network Project, Genomics Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA. [5]Wadsworth Center, New York State Department of Health and Department of Biomedical Sciences, State University of New York, Albany, New York 12201-2002, USA.
*These authors contributed equally to this work.

Saccharomcyces[7]. Like *Saccharomyces*, a −1 nucleosome was detected ~180 bp upstream of the TSS. However, in contrast, it lacked H2A.Z.

Surprisingly, the genic array of *Drosophila* nucleosomes started ~75 bp further downstream from the equivalent position in *Saccharomyces*, placing the +1 nucleosome at +135 (Fig. 1b and Supplementary Fig. 10). This shift has important implications in how the TSS is presented to RNA polymerase II (Pol II). In *Saccharomyces*, the TSS resides within the nucleosome border, potentially allowing the nucleosome to regulate start-site selection and efficiency[1]. In *Drosophila*, the predominant arrangement of nucleosomes might allow unimpeded access to the TSS, with potential blockage occurring downstream after initiation.

Drosophila have well-defined core promoter elements, such as TATA, initiator (Inr), downstream promoter element (DPE) and motif ten element (MTE), which bind to the general transcription machinery[18–22], although these elements are not found in most genes. For genes lacking these core promoter elements or having a DPE, the canonical nucleosome organization was observed (black pattern in Supplementary Fig. 11), which was more robust when only H2A.Z-containing nucleosomes were examined (blue pattern). In contrast, genes containing TATA, Inr or MTE had a diminished canonical nucleosome organization and a diminished NFR, indicating that these classes of genes may have a more compact and gene-specific chromatin architecture, including a positioned nucleosome over the TSS. Consequently, they might be more dependent on chromatin remodelling for expression. When genes become transcriptionally competent, resident nucleosomes could adopt a more open and canonical organization, which includes replacing H2A with H2A.Z. Three observations support this hypothesis. First, H2A.Z and bulk nucleosomes at highly expressed genes were more uniformly organized than those at genes with a lower expression (Supplementary Fig. 9). Second, bulk nucleosomes for genes that contained H2A.Z at their 5′ end displayed the canonical pattern, whereas those lacking H2A.Z did not (Supplementary Fig. 10, black plot versus red trace). Third, within any class of genes except those having an Inr, H2A.Z nucleosomes adopted a more canonical organization than the bulk set of nucleosomes (Supplementary Fig. 11). These results suggest that transcription and the presence of H2A.Z are linked to an open and uniform chromatin architecture at promoter regions.

DNA motif organization around nucleosomes

Recent genome sequencing of 12 *Drosophila* species of differing evolutionary distance has provided an unprecedented opportunity to identify conserved DNA sequence motifs[23]. By comparing the distribution of motifs around the TSS[23], we found four recurring patterns: 27 motifs were classified as 'nucleosomal', 57 as 'anti-nucleosomal', 12 as 'fixed', and 98 as 'random' (left panels in Fig. 2a and Supplementary Fig. 12). Nucleosomal and anti-nucleosomal patterns matched the general distribution of where nucleosomes were relatively enriched or depleted, respectively, relative to the TSS (see Fig. 1b). Fixed elements were at a defined distance from the TSS, and random elements lacked patterning. The nucleosomal and anti-nucleosomal patterns suggest that certain motifs are organized to be downstream of the TSS in the midst of nucleosomal arrays, whereas others are organized to be upstream of the TSS, where nucleosomes are relatively depleted.

We examined the organizational relationship of these DNA motifs to individual H2A.Z nucleosomes within the whole genome (right panels of Fig. 2a and Supplementary Fig. 12, and all motifs in Fig. 2b). Notably, nucleosomal motifs were consistently enriched on the H2A.Z nucleosome surface, whereas anti-nucleosomal motifs were consistently depleted. Individual fixed motifs were mostly depleted of H2A.Z nucleosomes. These findings along with several controls (Supplementary Fig. 13) suggest that motifs and nucleosomes adopt a preferred organization around each other, regardless of their genomic location. This organization could be linked to co-evolution of base sequence composition bias in and around nucleosomes. The functional importance of such context remains to be determined.

Nucleosome-positioning sequences

We examined whether the positions of *Drosophila* H2A.Z nucleosomes are at least partly defined by the underlying DNA sequence pattern, and investigated whether such a pattern might be evolutionarily conserved. We determined the frequency of dinucleotides across *Drosophila* H2A.Z nucleosomal DNA because 10-bp periodic patterns of certain dinucleotides enhance the wrapping and positioning of DNA around the histone core (Fig. 3a and Supplementary Fig. 14). As seen in *Saccharomyces*, 10-bp periodic patterns of A and/or T (A/T) dinucleotides running counter-phase to G/C dinucleotides was observed. The modest amplitudes of the pattern suggest that such periodicities are infrequent, and are thus used selectively (that is, most nucleosomes lack underlying positioning signals).

We further investigated the rules of nucleosome positioning by scanning promoter regions for correlations to nucleosome positioning sequences previously identified for a relatively small number of eukaryotic nucleosomes[24], in which AA/TT (yeast[25] and worms[26]) or CC/GG (human)[27] dinucleotides occur in a biased and/or periodic

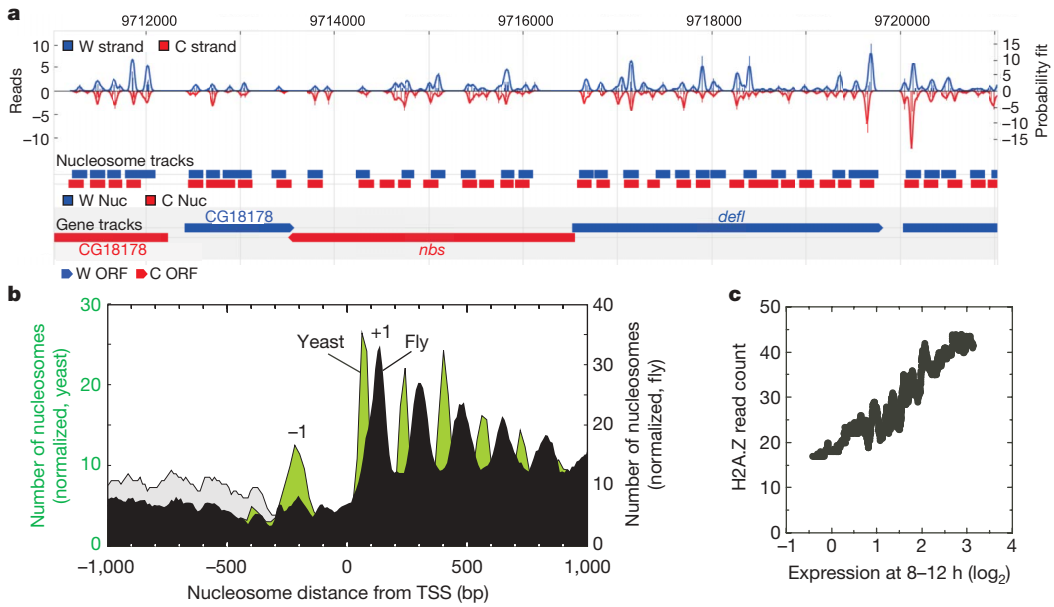

Figure 1 | H2A.Z nucleosomal organization around the 5′ end of *Drosophila* genes. a, Browser shot of an arbitrary locus (*nbs–defl*). The bar graph represents the number of 'W' or 'C' (top and bottom traces, respectively) strand reads mapped to each coordinate. **b**, Composite distribution of H2A.Z nucleosomes relative to the TSS. Nearby genes were either included (grey) or eliminated (black) from the analysis, and normalized accordingly. The equivalent *Saccharomyces* profile is shown in green[1]. **c**, Correlation of the number of H2A.Z nucleosomal sequencing reads (per gene) to messenger RNA levels in embryos (8–12 h).

arrangement across nucleosomal DNA. Unlike in yeast, the AA/TT positioning pattern failed to identify nucleosome locations (Fig. 3b, black trace). However, the CC/GG pattern (Supplementary Fig. 15) reproduced the exact position of the +1 nucleosomes (Fig. 3b, red trace), indicating that the *Drosophila* +1 nucleosome may be positioned in part by CC/GG-based positioning sequences that are used preferentially in metazoans. Consistent with this, +1 nucleosomes are highly positioned around the 5′ end of genes (Supplementary Fig. 16).

The ends of genes are nucleosome-free

Despite H2A.Z being enriched at the 5′ end of genes, substantial levels were detected throughout the genome, which allowed us to examine nucleosome organization at the 3′ end of genes (Fig. 4a and Supplementary Fig. 17a). Notably, H2A.Z nucleosome levels spiked near the open reading frame (ORF) end points and then dropped precipitously further downstream into the intergenic regions, where transcripts terminate. The spike occurred ~30 bp upstream from the stop codon and ~160 bp upstream of the transcript poly(A) site. A similar nucleosome drop-off was seen when bulk nucleosomes were examined (Supplementary Fig. 17b), but was not evident at genes that lacked H2A.Z. Thus, like the 5′ end, the presence of H2A.Z may be linked to a more open chromatin architecture at the 3′ end of genes. The change in nucleosome density coincided with alterations in nucleosome positioning sequences (Fig. 4b). Thus, such '3′ NFRs' might be defined in part by the underlying DNA sequence. Conceivably, 3′ NFRs might function in transcription termination.

RNA polymerase contacts the +1 nucleosome

The location of the +1 nucleosome at the 5′ end of genes is notable because its upstream border resides at approximately +62 (relative to the TSS), which is near where Pol II pauses during the transcription cycle[3,28–32]. To examine the potential linkage between Pol II pausing and nucleosome positions, we first determined the genome-wide location of Pol II in embryos at 1,956 putatively paused genes (Fig. 5a). Pol II was concentrated in a ~300 bp region that peaked around +90, which overlaps the region bound by the +1 nucleosome; this is consistent with other recent placements[30–32]. Indeed, the distribution of paused Pol II, as directly measured by permanganate reactivity of thymines on a statistically robust subset of ~50 genes (yellow trace in Fig. 5a and Supplementary Fig. 18a), indicates that pausing occurs between +20 and +50, with the centre at +35 (ref. 30). This high-resolution permanganate footprinting data, which represents the most definitive means of assessing Pol II pausing, places the front edge of Pol II (~16 bp downstream of the bubble[33]) within ~10 bp of the +1 nucleosome border.

The location of the +1 H2A.Z nucleosome was similar (but not identical) whether or not paused Pol II was present (Fig. 5b), indicating that Pol II was not likely to be the cause of the nucleosome shift compared to *Saccharomyces*. Instead, the positioned +1 nucleosome might be contributing to Pol II pausing, which is consistent with other studies[34–37]. Other factors including negative elongation factor (NELF) are likely to make significant contributions to pausing as well[30,38,39].

Intriguingly, genes that contained a paused Pol II showed a ~10 bp downstream shift of H2A.Z nucleosomes (P-value = 10^{-9}; Fig. 5b). The same shift was observed if H2A.Z sequencing reads (rather than

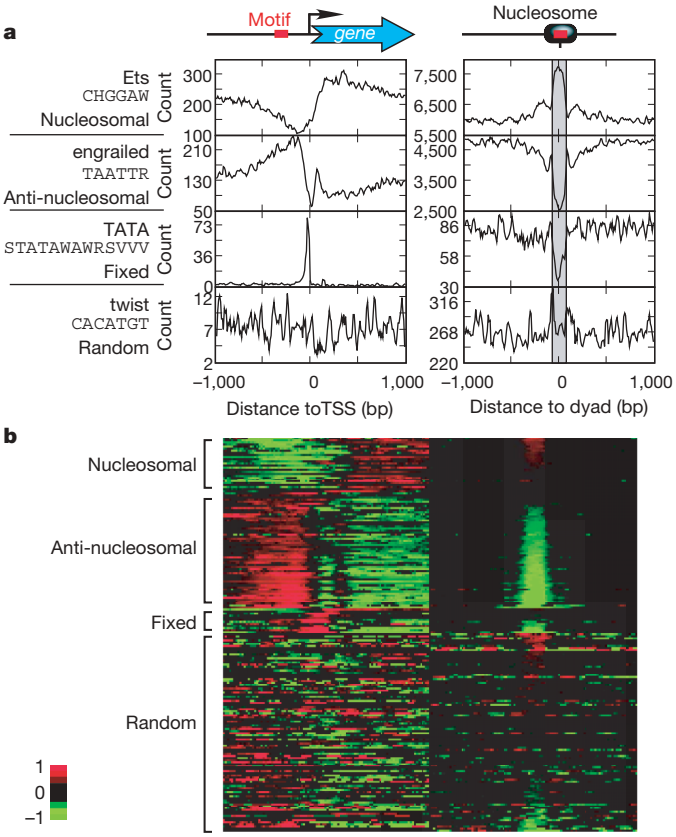

Figure 2 | Organization of conserved DNA motifs around TSSs (left) and nucleosomes (right). **a**, Distribution of representative conserved DNA motifs from four distribution classes. The motif name (TATA) or the protein that binds to the motif (E-26-specific or Ets, engrailed and twist) is indicated above the motif DNA sequence **b**, Distribution of all motifs plotted with Treeview[46]. Bin counts from all motifs in Supplementary Fig. 12 were converted to fold deviations from the regional average (±1 kb), converted to a $\log_2$ scale, and plotted. Red, black and green denote above, near or below average deviations, respectively.

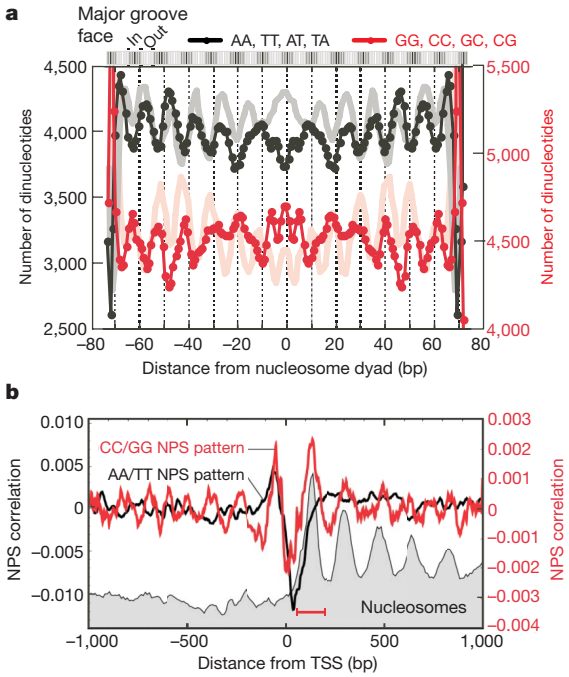

Figure 3 | Positioning properties of *Drosophila* nucleosomes and DNA. **a**, Composite distribution of WW (AA, TT, AT and TA represented by a black trace) and SS (GG, CC, GC and CG represented by the red trace) dinucleotides along the 147-bp axis of nucleosomal DNA (P-value, 0). The equivalent yeast profile[1] is shown in light shading in the background. **b**, Average correlation of all *Drosophila* promoter regions to nucleosome positioning sequence (NPS) patterns, comparing an AA/TT and a CC/GG pattern. The distribution of H2A.Z nucleosomes from Fig. 1b is shown as a grey backdrop. The span of the +1 nucleosome is indicated by the red horizontal bar.

nucleosomes) or bulk nucleosomes are plotted (Supplementary Fig. 19a, b). The shift suggests that, as part of the pausing process, Pol II collides with the +1 nucleosome, possibly displacing it downstream by one turn of the DNA helix. If the downstream nucleosomes are positioned mainly by the principles of statistical positioning[40,41], rather than the underlying DNA sequence, then a shift of the +1 nucleosome is expected to have a ripple effect on downstream nucleosomes.

To test the prediction that Pol II is engaging the +1 nucleosome, bulk mononucleosomes were prepared from formaldehyde-crosslinked embryos and immunoprecipitated with antibodies directed against Pol II. DNA corresponding to mononucleosomes (~150 bp) was gel-purified and mapped to the entire *Drosophila* genome with high-resolution tiling arrays. Figure 5c (black trace) shows that the distribution of nucleosome–Pol II crosslinking at Pol II-paused genes peaked at the +1 nucleosome. This was not seen at genes lacking a paused Pol II or H2A.Z. The selective enrichment at +1 demonstrates that Pol II is predominantly engaged with the +1 nucleosome, and therefore the +1 nucleosome may be instrumental in establishing the paused state.

Conclusions

The high resolution map of *Drosophila* nucleosomes reveals evolutionarily conserved and divergent principles of nucleosome organization. Genes that possess H2A.Z nucleosomes are likely to have experienced a transcription event. They tend to have nucleosome-free promoter and termination regions and intervening arrays of uniformly positioned nucleosomes that become less uniform towards the 3′ end of the gene. H2A.Z nucleosomes in general might not block assembly of the transcription machinery at transcriptionally 'experienced' promoters. However, repressed promoters or those containing Inr elements do seem to have an H2A nucleosome over the TSS.

Conserved DNA sequence motifs (and thus any proteins that bind to them) tend to have an organizational relationship with nucleosomes. 'Anti-nucleosomal' motifs including those for proteins such as engrailed, even skipped, fushi tarazu, giant, hunchback and knirps tend to be located upstream of the TSS and might contribute to the exclusion of nucleosomes over the core promoter. Indeed some have anti-nucleosomal activity[42,43]. 'Nucleosomal' motifs include sites for achaete, antennapedia, dorsal, tramtrack and others. Their preference for locations downstream of the TSS where nucleosomes are well organized raises the possibility that they contribute to nucleosome organization.

In *Saccharomyces*, the location of the TSS just inside the +1 nucleosome border allows the nucleosome potentially to exert control over initiation; however, in *Drosophila*, most genes might position the +1 nucleosome to interact with a transcriptionally engaged paused polymerase. It is not known whether the +1 nucleosome is causative or just participatory in the pausing. It is now becoming clear that metazoans also regulate transcription through Pol II pausing rather than solely through transcription complex assembly[3,31,32,44]. The nucleosome map and its context to DNA regulatory elements, presented here, provides a framework for designing

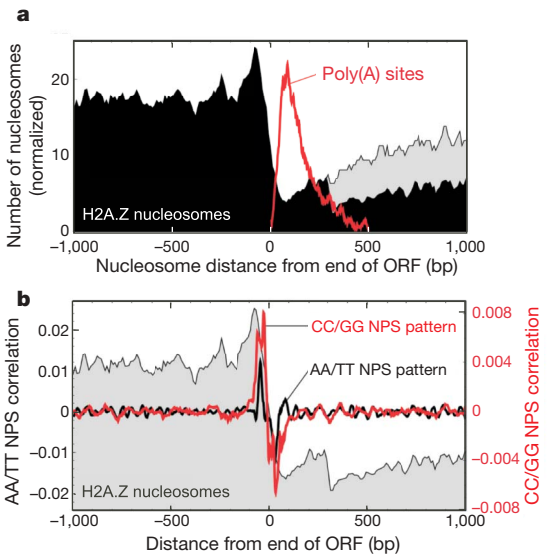

Figure 4 | H2A.Z nucleosomal organization around the 3′ end of *Drosophila* genes. **a**, Composite distribution of nucleosomes relative to ORF end points. Also shown is the distribution of transcript termination sites (poly(A) sites) in red. Nearby genes were either included (grey) or eliminated (black) from the analysis. **b**, Average correlation of all *Drosophila* gene terminal regions to AA/TT or CC/GG NPS patterns. The nucleosome profile from **a** is shown as a grey backdrop.

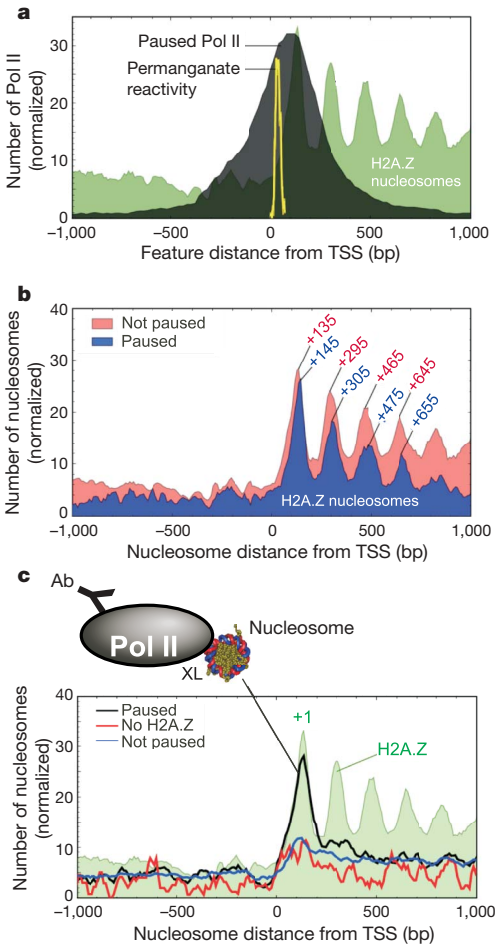

Figure 5 | Distribution of Pol II and Pol II-engaged nucleosomes around the 5′ end of genes. **a**, Genome-wide location of 'paused' Pol II relative to the TSS. Chromatin immunoprecipitation (ChIP)–chip genomic profiling of Pol II was conducted on *Drosophila* embryos. The black filled plot shows the distribution of Pol II at 1,956 'paused' Pol II genes (see Methods). The yellow trace shows the distribution of permanganate-reactive thymines (an indicator of pausing) in 50 genes that undergo pausing[30]. The distribution of Pol II at these 50 genes is similar to the bulk profile (Supplementary Fig. 18a). **b**, H2A.Z nucleosomal distribution at 1,956 genes that contained 'paused' Pol II. The 'not paused' class represents those for which Pol II is either absent or not paused. **c**, Distribution of Pol II-bound nucleosomes. Pol II ChIP was performed on bulk MNase-digested mono-nucleosomal DNA and hybridized to genome-wide tiling arrays. The three traces include distributions at 1,956 Pol II 'paused' genes and control distributions either at 788 genes that have 'no H2A.Z' within 1 kb of the TSS or at 8,736 genes of the 'not paused' class.

experiments and analysing existing data to understand how metazoans regulate transcription.

METHODS SUMMARY

D. melanogaster embryos (0–12 h) were collected and crosslinked with formaldehyde. H2A.Z was immunoprecipitated from chromatin digested with MNase. Mononucleosomal DNA was gel-purified and sequenced using Roche GS20/FLX pyrosequencing technology[1,45]. Chromatin from crosslinked embryos was also solubilized by sonication and/or MNase digestion, where indicated, and Pol II immunoprecipitated. Bulk nucleosomes were not immunoprecipitated. MNase-treated samples were gel-purified in the 75-200 bp range. DNA samples were then hybridized to Affymetrix *Drosophila* tiling microarrays (36 bp average probe spacing).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 February; accepted 26 March 2008.
Published online 13 April 2008.

1. Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446,** 572–576 (2007).
2. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
3. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77–88 (2007).
4. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39,** 1235–1244 (2007).
5. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genet.* **37,** 1090–1097 (2005).
6. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122,** 517–527 (2005).
7. Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309,** 626–630 (2005).
8. Leach, T. J. *et al.* Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster*. *J. Biol. Chem.* **275,** 23267–23272 (2000).
9. Updike, D. L. & Mango, S. E. Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS Genet.* **2,** e161 (2006).
10. Swaminathan, J., Baxter, E. M. & Corces, V. G. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of *Drosophila* heterochromatin. *Genes Dev.* **19,** 65–76 (2005).
11. Lieb, J. D. & Clarke, N. D. Control of transcription through intragenic patterns of nucleosome composition. *Cell* **123,** 1187–1190 (2005).
12. Raisner, R. M. *et al.* Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123,** 233–248 (2005).
13. Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123,** 219–231 (2005).
14. Li, B. *et al.* Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl Acad. Sci. USA* **102,** 18385–18390 (2005).
15. Hild, M. *et al.* An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5,** R3 (2003).
16. Tomancak, P. *et al.* Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8,** R145 (2007).
17. Bruce, K. *et al.* The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res.* **33,** 5633–5639 (2005).
18. Purnell, B. A., Emanuel, P. A. & Gilmour, D. S. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.* **8,** 830–842 (1994).
19. Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20,** 4754–4764 (2000).
20. Lim, C. Y. *et al.* The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* **18,** 1606–1617 (2004).
21. Biggin, M. D. & Tjian, R. Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* **53,** 699–711 (1988).
22. 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).
23. Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450,** 219–232 (2007).
24. Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* **262,** 129–139 (1996).
25. Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nature Genet.* **38,** 1210–1215 (2006).
26. Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **16,** 1505–1516 (2006).
27. Kogan, S. B., Kato, M., Kiyama, R. & Trifonov, E. N. Sequence structure of human nucleosome DNA. *J. Biomol. Struct. Dyn.* **24,** 43–48 (2006).
28. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol. Cell. Biol.* **6,** 3984–3989 (1986).
29. Law, A., Hirayoshi, K., O'Brien, T. & Lis, J. T. Direct cloning of DNA that interacts *in vivo* with a specific protein: application to RNA polymerase II and sites of pausing in *Drosophila*. *Nucleic Acids Res.* **26,** 919–924 (1998).
30. Lee, C. *et al.* NELF and GAGA factor are linked to promoter proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* (in the press).
31. Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nature Genet.* **39,** 1507–1511 (2007).
32. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39,** 1512–1516 (2007).
33. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292,** 1876–1882 (2001).
34. Brown, S. A., Imbalzano, A. N. & Kingston, R. E. Activator-dependent regulation of transcriptional pausing on nucleosomal templates. *Genes Dev.* **10,** 1479–1490 (1996).
35. Brown, S. A. & Kingston, R. E. Disruption of downstream chromatin directed by a transcriptional activator. *Genes Dev.* **11,** 3116–3121 (1997).
36. Carey, M., Li, B. & Workman, J. L. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol. Cell* **24,** 481–487 (2006).
37. Bondarenko, V. A. *et al.* Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* **24,** 469–479 (2006).
38. Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276,** 42601–42609 (2001).
39. Wu, C. H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17,** 1402–1414 (2003).
40. Kornberg, R. The location of nucleosomes in chromatin: specific or statistical. *Nature* **292,** 579–580 (1981).
41. Kornberg, R. D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16,** 6677–6690 (1988).
42. Lehmann, M. Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet.* **20,** 15–22 (2004).
43. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of *cis*-regulatory domains. *Science* **315,** 1408–1411 (2007).
44. Lis, J. T. Imaging *Drosophila* gene activation and polymerase pausing *in vivo*. *Nature* **450,** 198–202 (2007).
45. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
46. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants HG004160 (B.F.P.) and GM47477 (D.S.G.). We thank M. Biggin for early access to the Pol II chromatin immunoprecipitation (ChIP)–chip data, R. Fan for supplying the antibody raised against the *Drosophila* Pol II subunit Rpb3, and C. Lee for help in identifying paused Pol II.

Author Contributions T.N.M. prepared and purified the nucleosomes including Pol II-bound nucleosomes; C.J. analysed the nucleosome-mapping data and its relationship to other genomic features; I.P.I. performed computational analyses related to nucleosome-positioning sequences; X.L. conducted ChIP–chip on Pol II; B.J.V. conducted ChIP–chip and analysis on GAF; S.J.Z. provided bioinformatics support; L.P.T. constructed libraries and sequenced nucleosomal DNA; J.Q. mapped sequencing reads to the yeast genome; R.L.G. provided H2A.Z antibodies; S.C.S. directed the DNA-sequencing phase; D.S.G. directed embryo preparations and helped to interpret the data; I.A. developed computational approaches to derive nucleosome maps from the read locations and developed the associated browser; and B.F.P. directed the project, interpreted the data and wrote the paper.

Author Information Sequence data are deposited in the NCBI Trace Archives TI SRA000283 under the Sequencing Center designation 'CCGB', and microarray data are deposited in the ArrayExpress under accession numbers E-MEXP-1515, E-MEXP-1519 and E-MEXP-1520. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.F.P. (bfp2@psu.edu).

METHODS

Egg collection, dechorionation, crosslink and nuclei preparation. Eight grams of Oregon R embryos 0–12 h old were collected and crosslinked at a time, similarly to what has been described previously[47,48]. The embryos were dechorionated for 90 s, washed, and equally divided into eight 50 ml tubes containing 10 ml ChIP-FIX (50 mM HEPES, pH 7.6, 100 mM NaCl, 0.1 mM EDTA, 0.5 mM EGTA, 2% formaldehyde) and 30 ml heptane. The tubes were vigorously shaken for 15 min at 25 °C and then centrifuged at 1,500*g* for 1 min at 25 °C, followed by the removal of the aqueous layer. The embryos were washed as follows: once with PBS + 0.01% TritonX-100 + 0.125 M glycine, and twice with PBS + 0.01% TritonX-100. For each wash, 10 ml of the indicated buffer was added, the tubes were shaken for 1 min and centrifuged as before, and the aqueous layer was removed (except after the last wash, in which the heptane was also removed). The embryos were washed once more with 20 ml PBS + 0.01% TritonX-100.

The crosslinked embryos were resuspended in 40 ml homogenization buffer (10 mM HEPES, pH 7.6, 0.3 M sucrose, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA, 1 mM DTT, 1 mM sodium bisulphite, 0.2 mM PMSF) in a dounce with 10 strokes of the loose pestle and 15 strokes of the tight pestle, while on ice. The sample was centrifuged at 2,000*g* for 10 min at 4 °C, the supernatant was removed, and the pellet was washed once in 10 ml NPS buffer. After centrifuging at 16,000*g* for 5 min at 4 °C, the supernatant was removed and the pellet was frozen in liquid nitrogen and stored at −80 °C until ready for chromatin digestion.

H2A.Z nucleosome preparations. See Supplementary Fig. 2 for a schematic of the procedure. MNase digestion of chromatin was carried out similarly to what has been described previously[1]. Fifteen grams of embryos were thawed and resuspended, using a dounce, in a total volume of 36 ml in NPS buffer + 1 mM β-mercaptoethanol. MNase (40 kU) was added to the sample and incubated for 2 h at 25 °C on a rototorque. The sample was chilled on ice for 10 min and EDTA was added to 10 mM final concentration to quench the digestion. The sample was then centrifuged at 15,000*g* for 10 min at 4 °C and the supernatant was discarded. The pellet was washed twice with 36 ml formaldehyde (FA) lysis buffer + SDS and centrifuged as before, discarding the supernatant each time.

To solubilize digested chromatin, the sample was briefly sonicated. The sample was resuspended in 18 ml FA lysis buffer + SDS, and equally divided into fifteen 15-ml tubes. Two samples were sonicated simultaneously in a Diagenode Bioruptor on medium power for five sessions (each session consisting of 30 s ON and 30 s OFF). Samples were then transferred to 1.7-ml tubes and centrifuged at 16,000*g* for 10 min at 4 °C. The pellets were discarded and the supernatants were frozen at −20 °C until ready for ChIP.

The supernatants were thawed on ice and centrifuged at 15,000*g* for 10 min at 4 °C to remove any debris. Solubilized digested chromatin from 15 g of embryos was combined and the volume was increased fourfold with FA lysis buffer to dilute out the SDS. After this, the chromatin was divided equally into five 15-ml tubes. To each tube, 170 µl anti-H2A.Z antibody[8] was added and samples were incubated for 14 h at 4 °C on a rototorque. Chromatin was pre-cleared by incubating with 115 µl bed volume Sepharose 4B (Amersham, 17-0120-01) for 15 min at 4 °C on a rototorque. The resin was centrifuged at 2,000*g* for 3 min at 25 °C, and the supernatants were transferred to a new tube containing 115 µl bed volume protein A-Sepharose (Amersham, 17-0780-01). Samples were incubated for 1.5 h at 4 °C on a rototorque.

The resin was centrifuged at 1,000*g* for 2 min at 25 °C. The supernatants were removed and the resins were combined into two tubes and washed in series with the following buffers: twice with FA lysis buffer, twice with high-salt wash buffer, twice with FA wash 2 buffer, twice with FA wash 3 buffer, and once with TE (10 mM Tris-Cl, pH 8, 1 mM EDTA). For each wash, 15 ml of the indicated buffer was added, the sample was incubated for 5 min at 25 °C on a rototorque, the resin was centrifuged at 1,000*g* for 2 min at 25 °C, and the supernatant was discarded. Additionally, the resin was transferred to a new tube during the second FA lysis and FA wash 3 washes to reduce background. Finally, the resin in each tube was resuspended in 5.85 ml ChIP elution buffer (25 mM Trizma, 2 mM EDTA, 0.2 M NaCl, 0.5% SDS) and incubated for 15 min at 65 °C. The resin was centrifuged and all eluates were combined in a single tube.

The crosslinks were reversed and the ChIP DNA was isolated and gel-purified as described previously[1] (see Supplementary Fig. 3), with the exception that during gel-purification DNA fragments ranging in size from 100 to 200 bp were excised. Gel-purified DNA was subject to pyrosequencing using the Roche GS20/FLX in accordance with the manufacturer's instructions. Raw sequencing reads can be accessed through NCBI Trace Archives TI SRA000283 under the Sequencing Center designation 'CCGB'. Bulk downloads or specific queries of nucleosome positions can be accessed from http://atlas.bx.psu.edu/ or Supplementary Table 1.

H2A.Z nucleosome mapping. Sequencing reads were aligned to the FlyBase *D. melanogaster* reference genome (release 5.2) using BLAST. Sequences were aligned to all regions with >90% identity. Aligned regions were denoted on the *Drosophila* genome browser (http://atlas.bx.psu.edu/) by the coordinate of each read midpoint. Because the entire nucleosomal DNA was sequenced, both nucleosome borders were identified, thereby allowing nucleosome positions to be defined relative to each border: 73 bp interior to each end. The clustered distribution of reads (bar graph) were smoothed at a coarse-grain level, as described previously[1], using a value of 20 as the 'smoothing parameter' (see Supplementary Fig. 2). The smoothing included adjustments of bar heights (read locations) in proportion to empirical determinations of MNase bias at each cut site (Supplementary Fig. 4). Nucleosome positions were defined as the closest genomic coordinate to the peak of the smoothed distribution. Three coarse-grain determinations were made using: only the Watson (W) strand, only the Crick (C) strand, and both strands. The error between the W and C calls is presented in Supplementary Fig. 7.

Analysis was performed on only those 112,750 nucleosomes that were defined by three or more reads, although virtually identical patterns (and conclusions) were achieved when all H2A.Z nucleosomes (defined by one or more reads) were included (data not shown). Nucleosome fuzziness (Supplementary Fig. 16) was calculated as the standard deviation for the set of reads for which midpoint locations (defined above) resided within 73 bp of the coarse-grain nucleosomal midpoint.

Nucleosome distribution profiles. The annotation of all *Drosophila* features (TSS included) were downloaded from FlyBase release 5.2 (ftp://ftp.flybase.net/genomes/dmel/dmel_r5.2_FB2007_01/fasta). The TSSs are annotated in the feature transcripts, and represent both experimental and computationally derived determinations. Only 3,419 (~24% of total) messenger RNA genes have alternative transcripts, many of which have the same TSS but with variant exon and intron structure. To remove the redundancy, only the 5′-most TSS was used in this study for genes with alternative transcripts. This eliminates TSSs located internal to genes. However, indistinguishable results were obtained when we used the 3′-most TSS for each gene (data not shown). The facts that *Saccharomyces*[49,50] and *Drosophila* TSS annotations were each derived from multiple laboratory sources, include both experimental and computational predictions, are internally consistent (yielding robust nucleosome landscape patterns), and are widely used, would indicate that there is unlikely to be a level of random or systematic error in the annotations that would lead to the offset of the nucleosome landscape reported in Fig. 1b.

In Fig. 1b, only those genes with a +1 H2A.Z nucleosome were examined, although essentially the same result was obtained when all genes were examined. In Fig. 4a, only those genes with an H2A.Z nucleosome within 1 kb of the ORF end were examined, although essentially the same result was obtained when all genes were examined.

Regions located internal to or within 300 bp of a nearby gene were removed from the analysis, except where indicated, to minimize potential influence from nearby genes. A minimum of 300 bp in either direction of the reference feature was analysed. None of these filters substantially affected the distribution of the data or conclusions. Equivalent filters were applied to the other data sets (described below). Nucleosome distances from the reference feature were binned in 10-bp intervals. Bin data were normalized to number of regions represented in each bin, and smoothed using a moving average. The size of the moving average (typically between three and eight bins) was set to provide the optimal balance between signal and noise for purposes of visual display. The size of moving average within the chosen range had no effect on the conclusions drawn.

For Fig. 5b, nucleosome positions for 1,956 'paused' genes were binned (10-bp bins) and plotted as a three-bin moving average. The number of genes used in the 'not-paused' class was 8,736. This number includes genes that either lack Pol II or lack a 'paused' Pol II. A paired *t*-test confirmed that the 'paused' pattern is not a subset of the 'not–paused' pattern (*P*-value, 10^{-43}), and that it is shifted from the 'not paused' pattern (*P*-value, 10^{-9}).

Preparation of bulk nucleosomes. Nucleosomal DNA from 3 g of Oregon R embryos (0–12 h old) was prepared similarly to H2A.Z nucleosomal DNA as described above, with the following change. After sonication, solubilized DNA was treated with an equal volume of 6 M urea, and was then dialyzed for 20–24 h against 100-fold volume of FA lysis buffer + 0.05% SDS at 4 °C (SpectraPor membrane, 6,000–8,000 molecular weight cut-off, 632650). This treatment enhanced the recovery of Pol II during immunoprecipitation (see below). A sample of this material was gel-purified (the remainder was subjected to Pol II immunoprecipitation), and was then amplified by ligation-mediated polymerase chain reaction (LM-PCR). The bulk nucleosomal DNA was amplified by LM-PCR and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays, each of which contains about 3 million oligonucleotide 25-bp probes that cover the euchromatic portion of the genome at an average resolution of about

one per 36 bp. Bulk nucleosome mapping data were analysed using the 'model-based analysis of tiling arrays' (MAT) program[51], using the following settings: bandwidth 70 bp, minimum probe count 3, MaxGap 147 bp (which reflects the nucleosomal DNA length), 25-mer probe and 36 bp average probe separation. Interval cutoff was set to a *P*-value of 1×10^{-2}. We developed a peak detection algorithm that iteratively selects the maximal, non-overlapping peaks based on the MAT score assigned to the probe. The script is available on request. A total genome-wide count of 415,119 nucleosomes was obtained by these criteria (Supplementary Table 5).

H2A.Z correlation to messenger RNA expression. Data in Fig. 1c were previously normalized to 0–4 h expression to control for the presence of maternal RNA[15]. Shown is a moving median of 300 gene windows from 7,657 genes. Similar profiles were obtained when only the H2A.Z read count in the +1 nucleosome was considered, or if the genic read count was normalized to gene length (not shown).

DNA motif distribution. The 194 motifs were obtained by combining the novel motifs and known transcription factor (TF) motifs from a previous report[23]. Three motifs were excluded owing to rare occurrences (twin of eyeless recognition motif, TGGAGGDGGWAHTMATBVRTGWDD DRKKMW; glass recognition motif, CAATGCACTTCTGGGGCTTCCAC; and abnormal chemosensory jump 6 recognition motif, TGCATAATTA ATTAC). Four short or highly degenerate motifs (prospero recognition motif, CWYBDCY; *apterous* recognition motif, TAAT; mitochondrial transcription factor A recognition motif, TTATS; and bric a brac 1 recognition motif, WHWWWWWW-WWKK) caused computational failure owing to over-abundance. Their locations were therefore restricted to within ±1 kb of the TSS.

On the basis of the motif consensus sequences, all occurrences of each motif were identified on both strands across the entire fly genome. Palindromic motifs were only searched on one strand. Motif midpoint distances to the TSS were calculated and binned (size = 10 bp). Bin counts were smoothed using a seven-bin moving average and plotted in Fig. 2b and Supplementary Fig. 12. Each motif was classified into one of four classes on the basis of visual inspection of the profiles (Supplementary Table 4), as shown in Fig. 2. To derive the values in Fig. 2b, bin counts were normalized to the average bin count from ±1 kb of the TSS, and then $\log_2$ transformed. The frequency of each motif distance from an H2A.Z nucleosome midpoint located throughout the entire genome was measured in the same manner. Several negative controls were performed as described in Supplementary Fig. 13.

GAGAG site identification. The *D. melanogaster* genome (r5.2) was scanned for GAGAG motifs that resided within 20 bp of another GAGAG motif (Supplementary Table 3), because they tend to co-occur[43,52]. Such clusters were binned in 20-bp intervals based on the distance from the TSS. Binned data were smoothed using a five-bin moving average. Clusters that were within 300 bp of an adjacent gene were removed from the analysis (which had little effect on the distribution).

GAGA-associated factor binding. ChIP–chip data for genome-wide GAF were obtained from ref. 30, and represent GAGA-associated factor (GAF) binding in Schneider 2 cells, as detected by Affymetrix *Drosophila* tiling 1.0R arrays. Signal analysis, interval analysis and peak calling were performed using MAT software[51]. The bandwidth value was 70 and the MaxGap value was 150 bp, and peaks were called using a 1% false-discovery rate threshold. The distance from the TSS was calculated for the 2,903 GAF peaks above a 1% false-discovery rate threshold. The GAF composite distribution profile was generated by binning these distance values in 20-bp intervals for all genes that contain a GAF peak within 1 kb of the transcription start site. The data were smoothed using an eight-bin sliding window average in one-bin steps.

Dinucleotide frequencies. To survey for all 16 dinucleotides across nucleosomal DNA, we used all 7,600 nucleosomal DNA sequences that were exactly 147 bp long. Dinucleotide counts were performed in the 5′–3′ direction on both strands and were summed. Data were smoothed using a 3 bp moving average, except at position ±73 and ±72 at which 1 bp and 2 bp moving averages, respectively, were used.

Nucleosome-positioning sequences. Correlation profiles were calculated as described elsewhere[25], using the AA/TT dinucleotide pattern obtained previously[24]. We also used a *Drosophila*-specific pattern in which the AA/TT dinucleotide pattern was derived from the top 1,000 most well positioned H2A.Z nucleosomes, representing <1% of all nucleosomes, and obtained essentially the same result (data not shown). The CC/GG pattern was derived from the same nucleosome set used to derive the AA/TT pattern. This pattern is shown in Supplementary Fig. 15. The *Drosophila* pattern represents dinucleotide positional frequency profiles dyad-symmetrized as described previously[24]. In all cases, the number of nucleosomes used to define the search pattern represents <1% of all nucleosome locations, and therefore the nucleosomes used in the search pattern are expected to contribute <1% to resulting genome-wide correlation profiles.

Genome-wide mapping of RNA polymerase II. The genome-wide distribution of Pol II was determined as described elsewhere[53]. In brief, stage-14 embryos were crosslinked with formaldehyde; chromatin was isolated by $CsCl_2$ gradient ultracentrifugation and sonicated; and Pol II was immunoprecipitated using the 8WG16 antibody (Covance) against the C-terminal domain of the Pol II large subunit. The Pol II ChIP and control ChIP DNA samples, along with DNA purified from input chromatin samples, were amplified and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays. The ChIP–chip data were analysed using TiMAT—software developed by the Berkeley *Drosophila* Transcription Network Project[53]. The number of Pol II-crosslinked regions was 4,286, and were identified by genome-wide analysis at a 1% false discovery rate based on the symmetric null distribution method of TiMAT (Supplementary Table 6). These regions were ranked on the basis of the peak window score of each region. Within these regions, peaks of Pol II were located using the Mpeak algorithm[54]. All parameters for Mpeak were set at the default, except for the 'Largest search range', which was reduced to 1.0 kb. 1,956 genes were found to contain at least one Pol II Mpeak within 1 kb of a TSS, and no Pol II Mpeak between +1 kb and the end of the gene. These genes are defined as putative Pol II-paused genes. Pol II locations in Fig. 5a were binned in 20-bp intervals. Eight-bin moving averages were generated and plotted as a filled line graph. Essentially the same result was obtained when all genes were examined (not shown). 8,736 genes lacked a Pol II Mpeak anywhere in the transcription unit, of which 3,742 contained Pol II crosslinked regions somewhere in the body of the gene, and 4,994 lacked significant Pol II binding.

Genome-wide mapping of nucleosomes bound to RNA polymerase II. Material that was accumulated before gel purification of bulk nucleosomes (see above) was incubated for 16–18 h at 4 °C with 54 µl anti-Rpb3 antibodies (raised in rabbit), and immunoprecipitated as described for H2A.Z. Immunoprecipitated DNA was not detected in the absence of antibody or in the absence of crosslinking (not shown). Eluate DNA was gel-purified in the size range of 75–200 bp, and 30 ng DNA was amplified by LM-PCR (primer sequence, GCGGTGACCCGGGAGATCTGAATTC). Two biological replicates were prepared for hybridization with the GeneChip WT Double-Stranded DNA Terminal Labelling Kit (Affymetrix 900812) using the manufacturer's recommended protocol, and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays. Nucleosome–Pol II interaction data were analysed using MAT, as described above for the bulk nucleosome analysis. A total genome-wide count of 82,969 Pol II-crosslinked nucleosomes were obtained using this criteria (Supplementary Table 7). The distribution of these nucleosomes relative to the TSS is displayed in Fig. 5c, and represents the average of two biological replicates. The nucleosome count data were binned in 10 bp and further smoothed using a three-bin moving average.

47. Fyrberg, E. & Goldstein, L. *Drosophila melanogaster*: Practical Uses in Cell and Molecular Biology. *Methods Cell Biol.* **44,** 1–732 (1994).
48. Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J. & Paro, R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17,** 5141–5150 (1998).
49. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103,** 5320–5325 (2006).
50. Zhang, Z. & Dietrich, F. S. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5′ SAGE. *Nucleic Acids Res.* **33,** 2838–2851 (2005).
51. Johnson, W. E. *et al.* Model-based analysis of tiling-arrays for ChIP–chip. *Proc. Natl Acad. Sci. USA* **103,** 12457–12462 (2006).
52. van Steensel, B., Delrow, J. & Bussemaker, H. J. Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl Acad. Sci. USA* **100,** 2580–2585 (2003).
53. Li, X. Y. *et al.* Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6,** e27 (2008).
54. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436,** 876–880 (2005).