## Research

# High nucleosome occupancy is encoded at X-linked gene promoters in *C. elegans*

Sevinç Ercan,[1,2,5,6] Yaniv Lubling,[3,5] Eran Segal,[3,4,6] and Jason D. Lieb[1,2]

[1]Department of Biology, Carolina Center for the Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA; [2]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA; [3]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel; [4]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

We mapped nucleosome occupancy by paired-end Illumina sequencing in *C. elegans* embryonic cells, adult somatic cells, and a mix of adult somatic and germ cells. In all three samples, the nucleosome occupancy of gene promoters on the X chromosome differed from autosomal promoters. While both X and autosomal promoters exhibit a typical nucleosome-depleted region upstream of transcript start sites and a well-positioned +1 nucleosome, X-linked gene promoters on average exhibit higher nucleosome occupancy relative to autosomal promoters. We show that the difference between X and autosomes does not depend on the somatic dosage compensation machinery. Instead, the chromatin difference at promoters is partly encoded by DNA sequence, because a model trained on nucleosome sequence preferences from *S. cerevisiae* in vitro data recapitulate nearly completely the experimentally observed difference between X and autosomal promoters. The model predictions also correlate very well with experimentally determined occupancy values genome-wide. The nucleosome occupancy differences observed on X promoters may bear on mechanisms of X chromosome dosage compensation in the soma, and chromosome-wide repression of X in the germline.

[Supplemental material is available for this article. The sequencing and microarray data from this study have been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession no. GSE20136.]

Eukaryotic genomes are packaged by nucleosomes, which consist of ~147 bp DNA wrapped around an octamer of histone proteins H2A, H2B, H3, and H4 (Luger et al. 1997). DNA sequence, chromatin remodelers, transcription factors, and structural proteins such as linker histones affect nucleosome locations, and likewise nucleosome positions can affect the binding of regulatory proteins such as transcription factors. Determining nucleosome occupancy across the genome is required to understand the relation between nucleosomes and underlying cellular processes such as transcription.

Many of the techniques that map nucleosomes rely on preferential digestion of linker DNA by the micrococcal nuclease (MNase) (Clark 2010). Microarray analysis or direct sequencing of MNase-generated mononucleosomal DNA (MNase-seq) provides a high-resolution measure of nucleosome occupancy across the genome. Using MNase-seq, a previous study mapped nucleosomes in *C. elegans* using a mixture of all stages of the life cycle (Valouev et al. 2008). The *C. elegans* life cycle is highly dynamic, consisting of the embryo, four larval stages, and adult, which contains a rapidly proliferating germline that constitutes approximately half of all adult nuclei. Another study in *C. elegans* used embryonic extracts and a biotinylation tagging system to examine nucleosome occupancy in different chromatin fractions by microarray analysis (Ooi et al. 2010). These data reflect enrichment or depletion of nucleosome occupancy in different chromatin fractions relative to whole-genome chromatin.

Our study builds on these previous nucleosome mapping efforts by separating two stages of the life cycle (embryos and adults), and by further separating the somatic and germ cells of the adult stage by using two mutant strains. The first mutant (*glp-1*) produces germlineless adults, in which almost all cells are fully differentiated somatic cells. The second mutant (*fem-2*) produces animals with a feminized germline that cannot produce sperm, and are therefore incapable of producing embryos. These mutant animals provide a mixture of somatic cells and a fully developed germline.

In *C. elegans*, the X chromosome is subjected to two different chromosome-wide transcriptional repression mechanisms. The first involves transcriptional silencing of the X chromosome in the germline during meiosis (Kelly and Aramayo 2007). The second involves somatic dosage compensation, in which the transcription from both of the X chromosomes in XX hermaphrodites is halved to match the transcriptional output from the single X in males (Ercan and Lieb 2009). This reduction is mediated by the dosage compensation complex, which binds to active gene promoters on the X chromosome (Ercan et al. 2007, 2009; Jans et al. 2009). Here, we focus on nucleosome occupancy differences between the X chromosome and autosomes.

We found that in both embryos and adults, *C. elegans* X chromosome promoters exhibit higher measured nucleosome occupancy than the promoters of autosomal genes. This is partly dependent on DNA sequence, because GC content, which correlates with higher nucleosome occupancy, is also higher at the X promoters. A model trained on in vitro nucleosome preferences of yeast DNA correlates well with in vivo nucleosome occupancy in *C. elegans*. The higher nucleosome occupancy on the promoters of X-linked genes may relate to mechanisms that regulate X chromosome transcription in soma and germline.

## Results

### Data generation and processing

We prepared chromatin from three distinct *C. elegans* populations. First, to sample nucleosome occupancy during development, we used wild-type embryos. Second, to examine nucleosome occupancy in mature somatic cells, we prepared extract from *glp-1(q224)* adults, which lack a germline. We refer to these animals as "germlineless adults" throughout the manuscript. Finally, to examine chromatin organization in a mixture of purely adult somatic and germ cells, we used *fem-2(b245)* adults, which contain a germline but cannot produce embryos due to defects in sperm production. These animals are referred to as "germline-containing adults" from this point on. We digested chromatin with increasing concentrations of MNase, and purified mononucleosomal DNA by gel electrophoresis (Supplemental Fig. 1A).

Three lanes of paired-end Illumina GAIIx sequencing were performed for each of two independent replicates. We obtained ~40 million pairs of reads per replicate, and aligned the reads to *C. elegans* WS170 (ce4) genome (see Methods). For each sample, the majority (50%–80%) of the sequenced fragments were between 130 and 160 bp in length (Supplemental Fig. 1B), consistent with the nucleosome core particle length of ~147 bp. The presence of fragments shorter and longer than 147 bp may result from both differential accessibility of the genome to MNase and the presence of variant nucleosomes that stably wrap differing lengths of DNA. The distributions of autosomal and X chromosome fragment lengths were nearly identical (Supplemental Fig. 1C), and using only those fragments that are longer than 146 bp recapitulates our main result (Supplemental Fig. 1D).

To prevent inaccuracies in measured nucleosome occupancy caused by differences in mappability, we excluded any region of the genome within 147 bp of a read that maps to multiple locations. This filtering step resulted in a final count of 17–27 million unique read pairs per replicate (Supplemental Table 1). The paired reads were joined, and the number of reads that cover each coordinate was normalized by $\log_2$ transforming the ratio of reads at a given base to the average genome coverage. The two replicates of each data set correlated well with each other (correlation coefficients for embryo 0.91, germlineless adults 0.90, and germline-containing adults 0.96). Therefore, we averaged the base coverage ratios from the replicates and used this track as a measure of nucleosome occupancy (data from each replicate and the average are shown for two loci in Supplemental Fig. 2).

### The nucleosome occupancy of X chromosome promoters is higher than that of autosomes, as measured by micrococcal nuclease digestion

For each of the six *C. elegans* chromosomes, we aligned all genes by their annotated transcript start sites, and plotted the average nucleosome occupancy. The most striking feature of these plots was increased nucleosome occupancy in X-linked gene promoters relative to the autosomal gene promoters (highlighted with an arrow in Fig. 1A). This difference was specific to the region near the transcript start site. Increased nucleosome occupancy was not a general feature of the X chromosome, because no such difference was apparent when the data were aligned using other genomic features, including the annotated transcript stop sites (Fig. 1B).

Focusing on the 5′ region and collapsing all of the results from autosomes into a single plot reveals that the higher nucleosome occupancy occurs mostly at the promoter region, and to smaller extent at the nucleosome just downstream from the transcript start site (Fig. 1C). The difference in nucleosome occupancy between X and autosomes was also observed in adult somatic cells, and in animals containing both somatic and germ cells (Supplemental Fig. 3A,B). We confirmed the increased nucleosome occupancy that we observed on X by reanalysis of data from a previous study that used Applied Biosystems (Life Technologies) SOLiD System sequencing to profile nucleosome occupancy in mixed stage animals (Supplemental Fig. 3C; Valouev et al. 2008).

Genes on the X chromosome are separated by longer intervals than autosomal genes (Fig. 1D), and neighboring genes can influence the nucleosome occupancy profiles of genes that are closely positioned (Supplemental Fig. 3D). Therefore, we tested whether the shorter intergenic regions on autosomes might somehow be driving the difference in nucleosome occupancy we observed near transcript start sites. We did this by considering only genes on X and autosomes with an upstream intergenic region larger than 2 kb. Using this set of promoters, the nucleosome occupancy difference between X and autosomal promoters persisted (Fig. 1E). Importantly, the apparent nucleosome occupancy differences at regions greater than 500 bp upstream of the transcript start sites disappeared, suggesting that these were caused by neighboring genes.

### The differences in measured nucleosome occupancy between X and autosomes occur chromosome-wide

We next examined whether the difference between X and autosomes was due to just a few genes, or was a more widespread property of X-linked genes. A histogram of nucleosome occupancy (Fig. 2A) shows that the entire distribution of X-linked promoter nucleosome occupancy values is significantly shifted to the right (*t*-test, $P < 10^{-12}$). This is also illustrated by a heat map representation of nucleosome occupancy of all genes on the X versus the autosomes (Fig. 2B). Both the heat map and the histogram show that although the X harbors many gene promoters that have a typical nucleosome-depleted region, nucleosome occupancy is higher at the 5′ of genes compared to autosomal genes.

It is well known that MNase cuts naked DNA with sequence preferences that match the sequences characteristic to linker DNA (Keene and Elgin 1981; Yuan et al. 2005; Albert et al. 2007; Valouev et al. 2008; Clark 2010). Despite this correspondence between the sequence preferences of MNase and linker DNA, it has been firmly established that MNase digestion of chromatin results in 147 bp particles that are nucleosomes (Lutter 1979, 1981; Cartwright and Elgin 1982; Jessee et al. 1982; Prunell and Kornberg 1982; Cartwright et al. 1983; Richmond et al. 1984; Hayes et al. 1990; Flaus et al. 1996; Richmond and Davey 2003; Widlak and Garrard 2006; Visnapuu and Greene 2009). As expected, digestion of naked *C. elegans* DNA with MNase also resulted in a higher signal on X promoters than on autosomes (Fig. 2C).

Independent of MNase digestion and sequencing, others measured nucleosome occupancy by histone H3 ChIP microarray analysis in L3 larvae (Liu et al. 2011) and early embryos (Rechtsteiner et al. 2009). Using data from these studies, we found higher H3 enrichment around transcript start sites on the X compared to autosomes (−0.02 vs. −0.13 for L3 [Fig. 2D], and 0.04 vs. −0.02 for early embryo [Fig. 2E]; *t*-test, $P < 2.4 \times 10^{-18}$ for L3 and $P < 4.4 \times 10^{-12}$ for early embryo). These H3 ChIP data sets were generated using chromatin extracts prepared by sonication instead of MNase treatment, and H3 enrichment was measured by
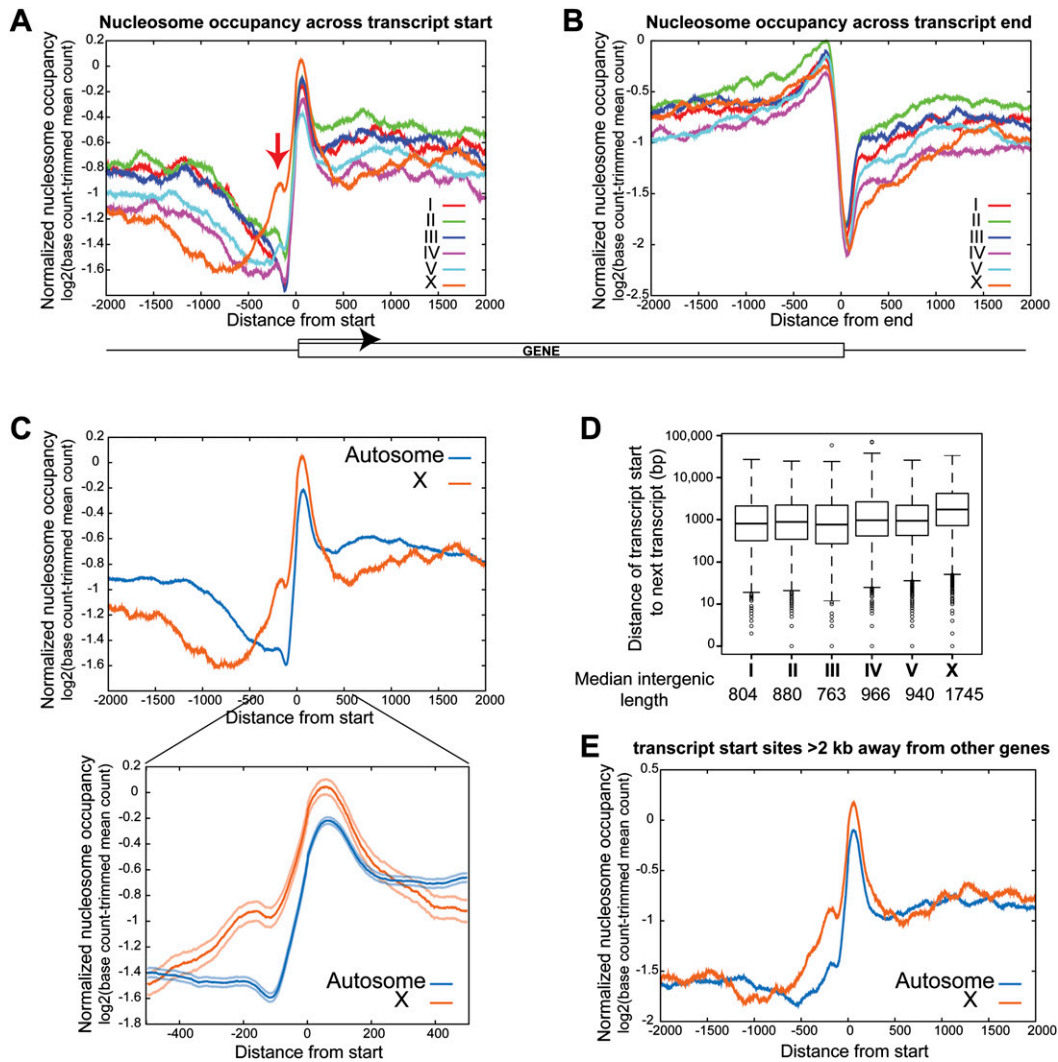
**Figure 1.** Higher nucleosome occupancy on X. Normalized embryo nucleosome occupancy per base pair was averaged across all transcripts, and plotted across the transcript start (*A*) and end (*B*) for each chromosome. A schematic representation of a gene is shown *below* the panels, and the red arrow in panel *A* indicates the higher nucleosome occupancy on the X promoters. (*C*) The same as *A*, but the data from autosomes are collapsed. The *lower* panel zooms to the 1-kb region around the transcript start. The lighter shades of each line represent two standard errors from the mean. (*D*) The distance between each transcript start to the nearest upstream transcript is plotted for each chromosome. Median length is reported *below*. (*E*) Same as *A*, but for only genes whose transcript starts are more than 2 kb away from other genes are used. This consisted of 5084 autosomal and 1480 X chromosome transcripts.

microarrays instead of high-throughput sequencing, and corroborate higher nucleosome occupancy at X promoters.

### Differences in GC content between X and autosomal promoters may underlie the differences in experimentally measured nucleosome occupancy

It has long been known that GC content correlates positively with nucleosome occupancy (Widom 2001; Lee et al. 2007; Peckham et al. 2007; Chung and Vingron 2009; Schwartz et al. 2009), and recent studies indicate that it may indeed be the dominating factor in DNA-encoded nucleosome occupancy (Tillo and Hughes 2009). Therefore, we examined whether the DNA sequence composition of X-linked promoters could account for the difference in the MNase-generated occupancy profiles. Indeed, at individual promoters across the genome ($-300$ to $+200$ relative to the transcript

start site), GC content correlated positively with nucleosome occupancy (Supplemental Fig. 3E, Pearson correlation coefficient 0.79). The difference in GC content between X and autosome promoters mirrored the measured difference in nucleosome occupancy, with the largest difference in both measures occurring ~150 bp upstream of the transcript start site (see arrow, Fig. 3A). Similar to the measured nucleosome occupancy values (Fig. 2A), the distribution of GC content values for individual promoters on X was shifted positively relative to autosomal values (Fig. 3B). In the region surrounding the transcript start sites, GC content of the X (36.7%) was higher than that of each individual autosome (35.8% for I, 36.0% for II, 35.6% for III, 35.2% for IV, and 35.1% for V). On average, GC content differed by 1.2% (35.5% on autosomes vs. 36.7% GC on X). At the same region, measured nucleosome occupancy on X-linked promoters is 1.6-fold higher relative to autosomes.
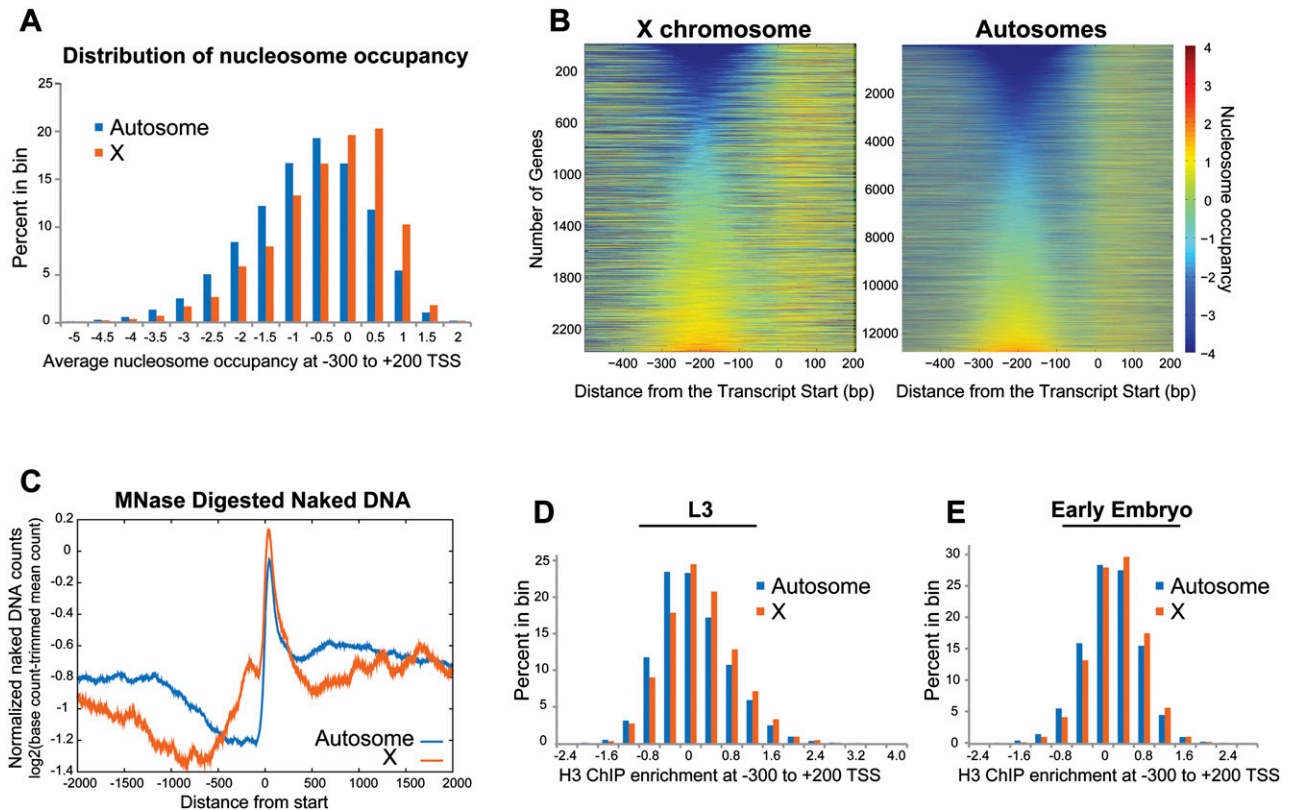
**Figure 2.** Nucleosome occupancy values at the beginning of X and autosomal genes. (*A*) Distribution of average embryo nucleosome occupancy at −300 to +200 bp of the transcript start sites among X and autosomes. (*B*) Heat map representation of nucleosome occupancy at all genes on the X and autosomes. Genes were sorted by their average occupancy between −300 to −100 bp from the transcript start site. The scale for the color gradient for occupancy values is shown on the *right*. (*C*) Naked DNA was extensively digested by MNase and sequenced. The read coverage per base pair is normalized and is plotted across the X and autosomal transcript start sites. (*D,E*) Histograms illustrate the distribution of average H3 ChIP enrichment between X and autosomes in L3 (*D*) and early embryos (*E*). Average ChIP enrichment was calculated from microarray probes within the −300 to +200 of the transcript start sites.

## A DNA-sequence based model correlates with nucleosome occupancy throughout the genome and recapitulates X and autosome differences

A nucleosome occupancy prediction model that is based on in vitro assembly of yeast nucleosomes (Kaplan et al. 2009) correlated well with in vivo *C. elegans* nucleosome occupancy. The Pearson correlation coefficient of 0.84 for embryos and 0.75 for germlineless adults (Supplemental Fig. 2F) is similar to the correlation observed between the model and in vivo nucleosome occupancy in yeast (0.74) (Kaplan et al. 2009). The in vitro model also predicted the higher nucleosome occupancy of X promoters (Fig. 3C). Two major components of this model are GC content and dinucleotide frequency. To test whether dinucleotide frequency is also different between X and autosomes, we aligned 147 bp sequenced fragments and counted the frequency of dinucleotides. Dinucleotide periodicity was apparent in the MNase digestion of chromatin (Supplemental Fig. 4A) as previously observed (Kaplan et al. 2009), but not in the digestion of naked DNA (Supplemental Fig. 4B). We next examined only nucleosomes that reside near transcript start sites (−300 to +200). There were no striking differences in dinucleotide frequencies between X and autosomes, although a disturbance in the pattern was observed 20 to 30 bp from the dyad among nucleosomes in X-linked promoters, the significance of which is not clear (Supplemental Fig. 4C).

## Differences in operon usage or repeat frequency do not account for nucleosome occupancy differences between X and autosome promoters

The *C. elegans* genome contains over 1000 operons, which account for about 15% of all protein-coding genes (Blumenthal et al. 2002). However, fewer than 5% of the operons are on the X (Blumenthal et al. 2002; Reinke and Cutter 2009). To ensure that this difference in gene content was not driving our result, we selected only genes that were either not part of operons, or were the first gene transcribed within operons. *C. elegans* transcripts undergo *trans*-splicing, which removes the beginning of RNA transcripts and adds one of two RNA splice leaders to each transcript, either "SL1" or "SL2" (Blumenthal 1995). Genes that are not in operons or are the first gene in operons typically receive the SL1 leader, with genes internal to operons receiving SL2. When we compared X and autosomal SL1-spliced genes (Gerstein et al. 2010), we found the same increased nucleosome occupancy at −300 to +200 bp from transcript starts (Fig. 4A) as reported for all genes (Fig. 1C). In contrast, SL2-spliced genes derived from the X and autosomes, which are internal to operons, have much more similar nucleosome occupancy profiles surrounding their transcript start sites (Fig. 4B).

The X chromosome sequence differs from autosomes in other ways. For example, the X chromosome contains fewer repeat
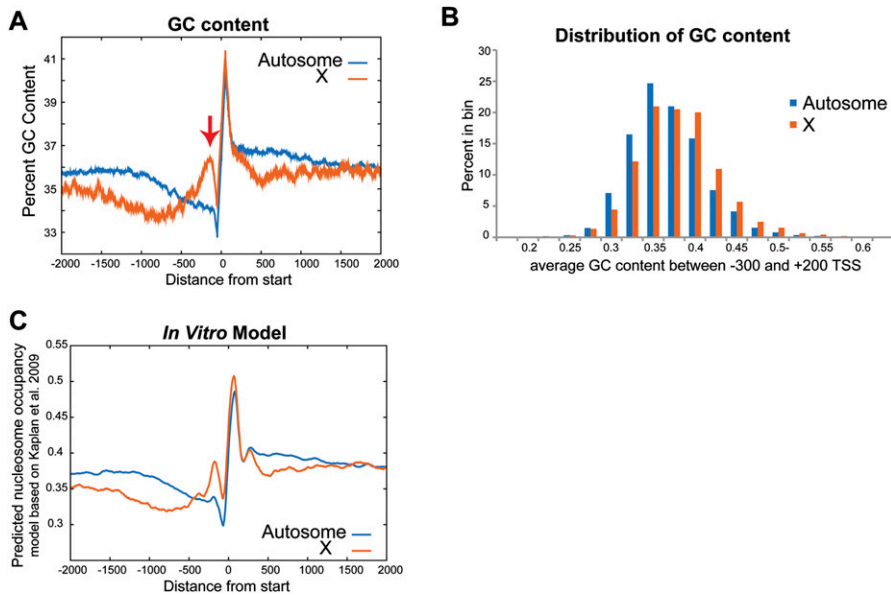
**Figure 3.** DNA sequence factors that influence nucleosome occupancy on X and autosomes. (*A*) Percentage of GC bases in a 50 base-pair window surrounding each base pair is averaged and plotted separately across X and autosomal transcript start coordinates. (*B*) Distribution of average GC content (−300 to +200 of the transcript start sites) among X and autosomes. (*C*) Normalized nucleosome occupancy per base pair was predicted by a model derived from in vitro reconstitution of nucleosomes on yeast DNA. The values from this model are averaged and plotted across the X and autosomal transcript start sites.

pattern that is similar to XX animals, with higher nucleosome occupancy at the beginning of X genes relative to autosomal genes (Fig. 5B).

## Nucleosome occupancy differences among X and autosomal promoters are predicted in a related nematode species, but not in *Drosophila*

We examined whether the differences in nucleosome occupancy extended to other species. *C. briggsae* is a nematode separated from *C. elegans* by ~100 million years of evolution, but which may employ a dosage compensation mechanism similar to that in *C. elegans*. In *C. briggsae*, nucleosome occupancy on X promoters is also predicted to be higher than autosomes (Fig. 5C). In contrast, predicted nucleosome occupancy in *D. melanogaster* lacks differences between X and autosomal promoters (Fig. 5D). *Drosophila* is estimated to have shared a last common ancestor with *C. elegans* more than 800 million years ago and has a completely different dosage compensation system (Straub and Becker 2007).

sequences, and therefore has more unique sequences, than autosomes. Consequently, the short sequence reads could be assigned to the X chromosome 1.13-fold more frequently per kilobase than on autosomes (Supplemental Table 2). Differences in the ability to map short sequence reads also occur across the transcript start and end sites (Supplemental Fig. 5A,B), as has been observed in the human genome (Rozowsky et al. 2009). However, the increased nucleosome occupancy we see at X-linked gene promoters is not due to such differences, because we only consider uniquely mappable regions on autosomes and the X. In addition, the mappability of the genome does not affect analysis of GC content or the in vitro model of nucleosome occupancy.

### Higher nucleosome occupancy at X-linked promoters does not depend on dosage compensation

To test whether the binding of the dosage compensation complex to X affects nucleosome occupancy at X-linked promoters, we mapped nucleosomes in a mutant strain that is karyotypically male (XO), but is transformed into a hermaphrodite for ease of growth and culture [*her-1(e1520) sdc-3(y126) V; xol-1(y9) X*] (Davis and Meyer 1997). In these animals, the dosage compensation complex does not bind to the X chromosome. We found that nucleosome occupancy in these XO animals was similar to that in XX animals (Fig. 5A). As expected, half the number of reads map to X. Correcting for the karyotype by doubling the number of reads assigned to X results in a nucleosome occupancy

## Discussion

There are several challenges associated with profiling nucleosome occupancy in *C. elegans* by MNase-seq. First, the true site of transcriptional initiation in *C. elegans* is not well-annotated due to *trans*-splicing (Krause and Hirsh 1987; Hastings 2005). Thus, throughout this manuscript we refer to the "transcript start site" rather than the "transcription start site," because in almost all cases the exact site of transcription initiation at a given gene is not known. For most of our analyses, we used the longest transcript coordinates annotated in the WS170 assembly of the genome. If we instead use transcript start coordinates from recent RNA-seq experiments, which are independently derived but still suffer from the ambiguity caused by *trans*-splicing, the results agree with our main finding that X promoters harbor higher nucleosome occupancy (Fig. 4A; data not shown). Second, nucleosome occupancy measurements recorded at each genomic region reflect occupancy
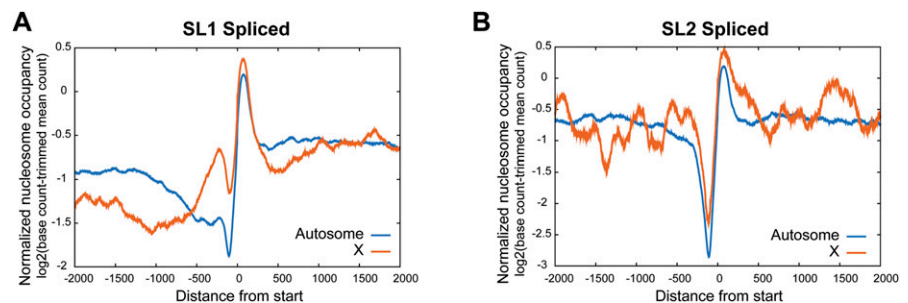


**Figure 4.** Nucleosome occupancy across the SL1 and SL2 spliced transcript starts. Normalized embryo nucleosome occupancy per base pair was averaged and plotted across X and autosomal transcript starts that have an SL1 (*A*) or SL2 (*B*) splice leader.
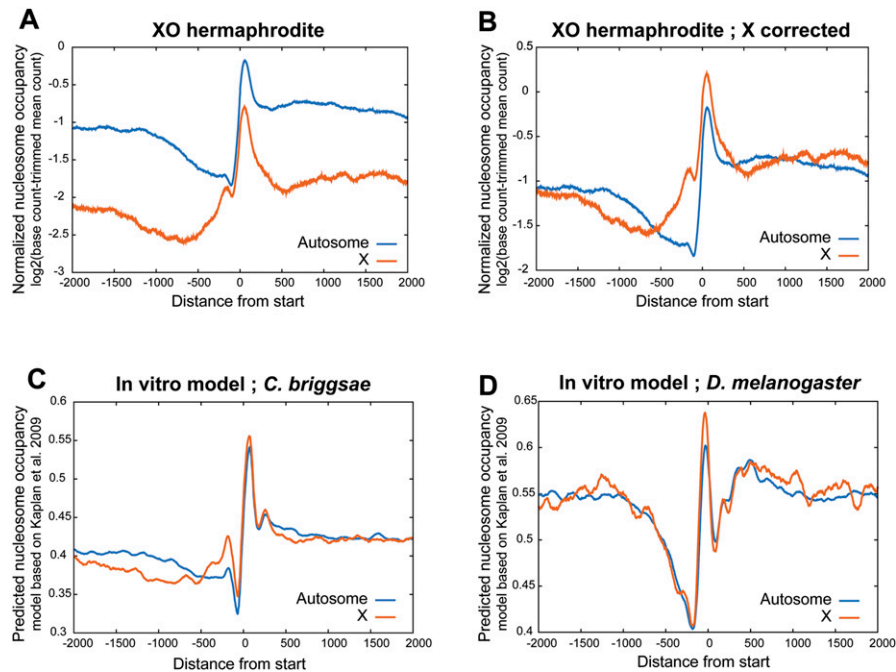
**Figure 5.** Higher nucleosome occupancy at the X promoters does not depend on somatic dosage compensation. (*A*) Nucleosome occupancy was measured in L3 XO hermaphrodites, which do not undergo dosage compensation. Nucleosome occupancy is averaged and plotted across X and autosomal transcript start sites. (*B*) X chromosome reads were doubled to correct for a single X in XO animals. (*C,D*) Nucleosome occupancy prediction from the in vitro model is plotted for X and autosomes of *C. briggsae* (*C*) and *D. melanogaster* (*D*).

in mixed tissues and multiple cells. This could result in a more "blurred" nucleosome map, especially in a highly dynamic sample such as embryos. Consistent with this possibility, overall nucleosome coverage of the genome appears to be more even in embryos relative to adults (Supplemental Fig. 5C). Third, high-throughput sequencing of the *C. elegans* genome contains an inherent positive bias toward GC content (Hillier et al. 2008). However, this bias would apply to the entire genome, not just X promoters, and our finding was corroborated by independent data using the SOLiD sequencing platform, a model for nucleosome occupancy derived from in vitro nucleosome assemblies, straightforward measures of GC content, and histone ChIP experiments detected on DNA microarrays (Fig. 2D,E).

Nucleosome occupancy has been studied in many organisms, including yeast (Yuan et al. 2005; Lee et al. 2007; Kaplan et al. 2009; Zhang et al. 2009), *Aspergillus* (Nishida et al. 2010), the malaria parasite (Ponts et al. 2010), *C. elegans* (Valouev et al. 2008; Ooi et al. 2010), *Drosophila* (Mavrich et al. 2008), and human tissue culture cells (Schones et al. 2008). The results of our experiments are similar to previous studies with respect to the relationship between overall nucleosome occupancy and gene structure. Both the 5′ and 3′ ends of *C. elegans* genes harbor lower nucleosome occupancy, with a well-positioned nucleosome just downstream from the transcript start. *C. elegans* exons also contained higher nucleosome occupancy compared to introns (Supplemental Fig. 5D). Similar to other studies (Schones et al. 2008), we did not find a simple correlation between nucleosome occupancy and transcription. On average, in genes that change in expression between embryos and adults, nucleosome occupancy tended to increase at promoters when RNA levels decreased, and vice versa (correlation coefficient −0.12).

Unlike the autosomes, the X chromosome is subject to chromosome-wide regulation in both the soma and germline. During meiosis, the X chromosome is devoid of chromatin marks that correlate with active transcription (Kelly et al. 2002; Bender et al. 2006). In the soma, dosage compensation complex binds to promoters of expressed genes on the X (Ercan et al. 2007; Jans et al. 2009), and represses transcription by a factor of two (Meyer and Casson 1986). Here, we report higher nucleosome occupancy in promoters on the X chromosome relative to autosomal promoters in both embryonic and adult cells in *C. elegans*. The higher nucleosome occupancies on X did not depend on dosage compensation, and may be due to the higher GC content on X chromosome promoters relative to autosomes.

Both in vivo and in vitro evidence suggest that nucleosomes are generally repressive to transcription (for review, see Workman 2006). Thus, higher nucleosome occupancy on X may confer resistance to transcription. In humans, the DNA at tissue-specific promoters tends to specify higher nucleosome occupancy (Tillo et al. 2010), perhaps because these promoters should be activated in only a subset of the tissues. Consistent with this hypothesis, genes whose expression depends strongly on chromatin remodelers have higher nucleosome occupancy in their promoters (Ioshikhes et al. 2006; Field et al. 2008; Tirosh and Barkai 2008). Perhaps higher nucleosome occupancy at *C. elegans* X chromosome promoters facilitates the activity of chromatin remodelers that differentially regulate X-linked transcription in multiple tissues and life stages.

Nucleosome occupancy at gene promoters evolves according to the expression program of that gene (Field et al. 2009; Tirosh et al. 2009, 2010; Tsankov et al. 2010). It is possible that the high nucleosome occupancy at X promoters coevolved with the somatic and germline dosage compensation mechanisms that regulate X transcription. In a related nematode, *C. briggsae*, X is also predicted to have higher nucleosome occupancy (Fig. 5C). In contrast, no such difference is observed in *D. melanogaster* (Fig. 5D), where transcription from the single X in males is increased twofold, probably through an elongation-mediated mechanism (Straub and Becker 2007). *C. elegans* dosage compensation is mediated by a condensin-like protein complex, with four out of five members also involved in chromosome assembly and condensation during mitosis and meiosis (Csankovszki et al. 2009). Whether nucleosome occupancy and positioning helps to provide specificity to condensin targeting will be exciting to explore.

## Methods

### Worm strains and growth

Mixed-stage embryos (wild type N2) were isolated and treated with 2% formaldehyde for 15 min. Germlineless adults were obtained by growing JK1107 [*glp-1(q224)*] at permissive temperature (15°C), then shifting L1s to restrictive temperature (25°C) until they reach

adulthood. Germline containing adults DH245 [*fem-2(b245)*] were grown similar to JK1107 strain, except that the L1s were shifted to 23°C. TY2205 [*her-1(e1520) sdc-3(y126) V; xol-1(y9) X*] was grown at 20°C and L1s were filtered to separate from dead embryos, and grown until they reached L3 stage. The worms were collected in one volume of PBS and frozen by dripping into liquid nitrogen. A cryo-mortar and pestle were used to grind worms to a fine powder. The powder was incubated in 1% formaldehyde for 1 min before chromatin preparation.

## MNase digestion and preparation of mononucleosomal DNA for sequencing

Samples were washed and dounced in dounce buffer (0.35 M sucrose, 15 mM HEPES-KOH at pH 7.5, 0.5 mM EGTA, 0.5 mM $MgCl_2$, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% Triton X-100, 0.25% NP-40). Samples were pelleted and washed with MNase digestion buffer (110 mM NaCl, 40 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM HEPES/KOH at pH 7.5) then resuspended in six volumes of MNase digestion buffer, and digested with increasing amounts of MNase at 25°C for 10 min. DNA was purified from each sample and mononucleosomal DNA was gel purified. Sequencing libraries were prepared from 50 ng of DNA. End repair was done with NEB Klenow, T4 DNA polymerase, and T4 PNK for 30 min at 20°C. Exo(–) Klenow fragment and dATP was used to add adenosine at the 3′ ends, 1 h at 37°C. DNA were ligated to pair-end adaptors (Illumina) and amplified by PCR. DNA between 200 and 400 bp in size was gel purified. Paired-end sequencing was performed by GAIIx at the UNC high-throughput sequencing facility.

## RNA preparation and expression array analysis

Embryos in 10 volumes of TRIzol (Invitrogen) were freeze-cracked. Adult RNA was prepared from frozen powder in TRIzol. Purification was done according to the manufacturer's protocol. RNA was further cleaned up using the Qiagen RNeasy kit. Labeling and hybridization to expression arrays were performed at Roche NimbleGen.

## Data processing

Raw data was processed through Illumina pipeline. Each pair of reads was mapped independently to the WS170 genome, using a proprietary software (SeqHit) allowing a single mismatch and not allowing any gaps. If a read mapped more than one region of the genome, that read was extended (on both sides) up to 147 bp, and removed from the analysis, as being part of a non-unique region. We used only those pairs of uniquely mapped reads that were less than 200 bp apart. The two read edges were extended to cover the coordinates in between. The raw coverage of a coordinate is calculated by adding up the number of read pairs that cover it. Zero value is assigned to base pairs that were in a unique region but not covered by any read pair. Raw data files and wiggle tracks of nucleosome coverage per base pair are at the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) and the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) under accession number GSE20136. Any region of base coverage greater than 10 times the genome median, were trimmed to equal 10 times the median. The number of reads that cover each coordinate was normalized by $log_2$-transforming the ratio of reads to the average genome coverage. The two replicates of each data set were averaged and used as a measure of nucleosome occupancy. Wiggle tracks of normalized files are available at the modENCODE Data Coordination Center (DCC) (http://www.modencode.org/). DCC accession numbers for nucleosome occupancy in embryos, germlineless, and germline-containing adults are 2763, 2764, and 3276,

respectively. The DCC accession numbers for H3 ChIP experiments in L3 and early embryos are 2407 and 2312, respectively. H3 ChIP enrichment was calculated by standardizing $log_2$ ratios of the intensity of H3 ChIP channel to the input DNA channel obtained from hybridization to NimbleGen microarrays (080922_modENCODE_CE_chip_HX1, GEO platform number GPL8647) that tile the whole genome at 50-bp resolution.

## Data analysis

Transcript coordinates were composed from WormBase WS170. For genes with multiple transcripts, only the longest transcript was selected. Transcripts shorter than 200 bp were discarded (71 transcripts). In vitro model predictions were done on *C. briggsae* (UCSC genome version Cb3), and on *D. melanogaster* (UCSC genome version dm3).

More detailed descriptions of our methods are provided in the Supplemental material.

## References

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446:** 572–576.

Bender LB, Suh J, Carroll CR, Fong Y, Fingerman IM, Briggs SD, Cao R, Zhang Y, Reinke V, Strome S. 2006. MES-4: An autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the *C. elegans* germ line. *Development* **133:** 3907–3917.

Blumenthal T. 1995. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet* **11:** 132–136.

Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, et al. 2002. A global analysis of Caenorhabditis elegans operons. *Nature* **417:** 851–854.

Cartwright IL, Elgin SC. 1982. Analysis of chromatin structure and DNA sequence organization: Use of the 1,10-phenanthroline-cuprous complex. *Nucleic Acids Res* **10:** 5835–5852.

Cartwright IL, Hertzberg RP, Dervan PB, Elgin SC. 1983. Cleavage of chromatin with methidiumpropyl-EDTA·iron(II). *Proc Natl Acad Sci* **80:** 3213–3217.

Chung HR, Vingron M. 2009. Sequence-dependent nucleosome positioning. *J Mol Biol* **386:** 1411–1422.

Clark DJ. 2010. Nucleosome positioning, nucleosome spacing and the nucleosome code. *J Biomol Struct Dyn* **27:** 781–793.

Csankovszki G, Collette K, Spahl K, Carey J, Snyder M, Petty E, Patel U, Tabuchi T, Liu H, McLeod I, et al. 2009. Three distinct condensin complexes control *C. elegans* chromosome dynamics. *Curr Biol* **19:** 9–19.

Davis TL, Meyer BJ. 1997. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development* **124:** 1019–1031.

Ercan S, Lieb JD. 2009. C. elegans dosage compensation: A window into mechanisms of domain-scale gene regulation. *Chromosome Res* **17:** 215–227.

Ercan S, Giresi PG, Whittle CM, Zhang X, Green RD, Lieb JD. 2007. X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nat Genet* **39:** 403–408.

Ercan S, Dick LL, Lieb JD. 2009. The *C. elegans* dosage compensation complex propagates dynamically and independently of X chromosome sequence. *Curr Biol* **19:** 1777–1787.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216. doi: 10.1371/journal.pcbi.1000216.

Field Y, Fondufe-Mittendorf Y, Moore IK, Mieczkowski P, Kaplan N, Lubling Y, Lieb JD, Widom J, Segal E. 2009. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat Genet* **41**: 438–445.

Flaus A, Luger K, Tan S, Richmond TJ. 1996. Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. *Proc Natl Acad Sci* **93**: 1370–1375.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechsteiner A, Ikegami K, et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE Project. *Science* **330**: 1775–1787.

Hastings KE. 2005. SL *trans*-splicing: easy come or easy go? *Trends Genet* **21**: 240–247.

Hayes JJ, Tullius TD, Wolffe AP. 1990. The structure of DNA in a nucleosome. *Proc Natl Acad Sci* **87**: 7405–7409.

Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, et al. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods* **5**: 183–188.

Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. *Nat Genet* **38**: 1210–1215.

Jans J, Gladden JM, Ralston EJ, Pickle CS, Michel AH, Pferdehirt RR, Eisen MB, Meyer BJ. 2009. A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes Dev* **23**: 602–618.

Jessee B, Gargiulo G, Razvi F, Worcel A. 1982. Analogous cleavage of DNA by micrococcal nuclease and a 1-10-phenanthroline-cuprous complex. *Nucleic Acids Res* **10**: 5823–5834.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362–366.

Keene MA, Elgin SC. 1981. Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure. *Cell* **27**: 57–64.

Kelly WG, Aramayo R. 2007. Meiotic silencing and the epigenetics of sex. *Chromosome Res* **15**: 633–651.

Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villeneuve AM, Reinke V. 2002. X-chromosome silencing in the germline of *C. elegans*. *Development* **129**: 479–492.

Krause M, Hirsh D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**: 753–761.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235–1244.

Liu T, Rechsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung M-S, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P, et al. 2011. Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* **21**: doi: 10.1101/gr.115519.110.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.

Lutter LC. 1979. Precise location of DNase I cutting sites in the nucleosome core determined by high resolution gel electrophoresis. *Nucleic Acids Res* **6**: 41–56.

Lutter LC. 1981. DNase II digestion of the nucleosome core: Precise locations and relative exposures of sites. *Nucleic Acids Res* **9**: 4251–4265.

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, et al. 2008. Nucleosome organization in the *Drosophila* genome. *Nature* **453**: 358–362.

Meyer BJ, Casson LP. 1986. *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* **47**: 871–881.

Nishida H, Motoyama T, Suzuki Y, Yamamoto S, Aburatani H, Osada H. 2010. Genome-wide maps of mononucleosomes and dinucleosomes containing hyperacetylated histones of *Aspergillus fumigatus*. *PLoS ONE* **5**: e9916. doi: 10.1371/journal.pone.0009916.

Ooi SL, Henikoff JG, Henikoff S. 2010. A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. *Nucleic Acids Res* **38**: e26. doi: 10.1093/nar/gkp1090.

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z. 2007. Nucleosome positioning signals in genomic DNA. *Genome Res* **17**: 1170–1177.

Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG. 2010. Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res* **20**: 228–238.

Prunell A, Kornberg RD. 1982. Variable center to center distance of nucleosomes in chromatin. *J Mol Biol* **154**: 515–523.

Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, Wang W, Kimura H, Lieb JD, Strome S. 2009. The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genet* **6**: e1001091. doi: 10.1371/journal.pgen.1001091.

Reinke V, Cutter AD. 2009. Germline expression influences operon organization in the *Caenorhabditis elegans* genome. *Genetics* **181**: 1219–1228.

Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**: 145–150.

Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. 1984. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**: 532–537.

Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol* **27**: 66–75.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887–898.

Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**: 990–995.

Straub T, Becker PB. 2007. Dosage compensation: The beginning and end of generalization. *Nat Rev Genet* **8**: 47–57.

Tillo D, Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* **10**: 442. doi: 10.1186/1471-2105-10-442.

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. 2010. High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5**: e9129. doi: 10.1371/journal.pone.0009129.

Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* **18**: 1084–1091.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**: 659–662.

Tirosh I, Sigal N, Barkai N. 2010. Divergence of nucleosome positioning between two closely related yeast species: Genetic basis and functional consequences. *Mol Syst Biol* **6**: 365. doi: 10.1038/msb.2010.20.

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol* **8**: e1000414. doi: 10.1371/journal.pbio.1000414.

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, et al. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* **18**: 1051–1063.

Visnapuu ML, Greene EC. 2009. Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat Struct Mol Biol* **16**: 1056–1062.

Widlak P, Garrard WT. 2006. Unique features of the apoptotic endonuclease DFF40/CAD relative to micrococcal nuclease as a structural probe for chromatin. *Biochem Cell Biol* **84**: 405–410.

Widom J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys* **34**: 269–324.

Workman JL. 2006. Nucleosome displacement in transcription. *Genes Dev* **20**: 2009–2017.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**: 626–630.

Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K. 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* **16**: 847–852.