Supplemental Material for:

**Nucleosome fragility is associated with future transcriptional response to developmental cues and stress in *C. elegans***

Tess E. Jeffers[1]* and Jason D. Lieb[1]

**TABLE OF CONTENTS**



## SUPPLEMENTAL RESULTS

### 10 bp periodicity of MNase digestion

Without MNase, chromatin remained intact and undigested. After addition of the enzyme, a stereotypic chromatin ladder rapidly formed, and a small proportion of total chromatin became mononucleosomal. As digestion proceeded, the mononucleosomal fraction increased while polynucleosomal fractions were depleted (**Supplemental Fig. 1A**). Mononucleosomal DNA fragments released earliest during the digestion were larger (median size of 2-minute nucleosomal fragments: 180 bp) than fragments released later in the timecourse (median of 30-minute nucleosomes: 155 bp) (**Supplemental Fig. 1B,C**). Among the digestion timepoints, nucleosome sizes decreased in 10 bp increments. This is consistent with the well-established observation that with increasing lengths of digestion time, MNase will trim long linkers down to the core nucleosome particle.

MNase preferentially cleaves linker DNA between adjacent nucleosomes. The initial fragment released from two MNase cut sites can range in size from 147 bp (the nucleosome core particle) to 147 bp + the sum of the lengths of the flanking linker DNA. In mammalian chromatin, this

initial fragment is approximately 180 to 200 bp. Fragment length is progressively shortened with additional digestion time (Axel 1975). MNase digestion within the linker region is semi-quantized, occurring in ~10.5 bp periodicity, which most likely reflects the helical turn of the DNA and the frequency of WW dinucleotides that MNase preferentially digests (McGhee and Felsenfeld 1983; Deniz et al. 2011; Trifonov and Sussman 1980; Ioshikhes et al. 2011). The longer nucleosome fragments we observe likely represent nucleosome core particles + flanking linkers that are not associated with the nucleosome. It is possible that these additional fragments are associated with other proteins. For example, the typical fragment size observed from standard MNase digestion in yeast is either 160 bp, if histone H1 is associated with the nucleosome, or 147 bp if H1 has been digested away. Thus, 147 bp represents the DNA associated with the 'core' nucleosome particle, and fragments smaller than 147 bp may indicate partially unwrapped or crowded nucleosomes (Chereji and Morozov, 2015), or extensively digested nucleosome fragments. We confirmed that mononucleosomal DNA fragments released earliest during the digestion were larger (median size of 2-minute nucleosomal fragments: 180 bp) than fragments released later in the timecourse (median of 30-minute nucleosomes: 155 bp) (**Supplemental Fig. 1B,C**).

**Comparison between MNase-seq and ChIP-seq nucleosome occupancy**

Nucleosome occupancy should be lower at more highly transcribed genes. Indeed, we confirmed that more highly transcribed genes had lower occupancy in our assay as measured by MNase-seq (**Supplemental Figure S5C**). We suggest that although fewer nucleosomes occupy highly transcribed genes, these nucleosomes are more resistant to MNase digestion. This interpretation is consistent with our data and with many recent descriptions of MNase sensitive and resistant nucleosomes: Chereji et al. observed MNase resistant nucleosomes in the gene body (Ooi et al. 2010; Chereji et al. 2015), and Mieczkowski et al. found that "*changes in accessibility at enhancers, promoters and other regulatory regions do not correlate with*

*changes in nucleosome occupancy. Moreover, high nucleosome occupancy does not necessarily preclude high accessibility, which reveals novel principles of chromatin regulation"* (Mieczkowski et al. 2016). The data presented in our manuscript, together with the evidence in these papers, provide strong support for the notion that nucleosome fragility measurements are not a simple consequence of changes in nucleosome occupancy.

As mentioned above, nucleosome occupancy should be lower at more highly transcribed genes, and we confirmed that more highly transcribed genes had lower occupancy in our assay as measured by MNase-seq (**Supplemental Figure S5C).** In contrast, when occupancy was measured by modENCODE histone H3 ChIP-seq in embryo chromatin, higher nucleosome occupancy was observed at more highly transcribed genes (**Supplemental Figure S5B**). This discrepancy is not unique to the modENCODE dataset. We searched the literature to find additional comparisons between gene expression and histone occupancy as measured by ChIP, and found that an incongruent relationship is not unique to this dataset. Perez-Lluch et al. showed that silent genes have lower H3 occupancy than ubiquitously transcribed or tissue-specifically expressed genes. (Pérez-Lluch et al. 2015). Similarly, Flensburg et al. also found higher H3 occupancy at expressed genes and promoters than at silent genes or promoters in mouse fetal liver (Flensburg et al. 2014). These results, in part, likely reflect the difficulties of achieving a "true" measure of nucleosome occupancy. ChIP coupled with sonication is advantageous because it does not use MNase, but it is subject to the potential biases and lower resolution inherent to sonication (Auerbach et al. 2009). Measuring occupancy by MNase digestion (or ChIP coupled with MNase digestion) affords higher resolution, but may be influenced by the accessibility of the adjacent linkers to MNase rather than the occupancy itself.

Several pieces of evidence suggest that nucleosome fragility is not the simple consequence of nucleosome depletion. (1) First, we found that nucleosome fragility on average was highest at

the promoters and gene bodies of lowly transcribed genes (see main text **Figure 4F**). As shown in **Figure 4H**, lowly transcribed genes have high nucleosome occupancy. (2) After heat shock, the –1 nucleosomes of the 14 heat shock-responsive genes showed no difference in nucleosome occupancy (as measured by MNaseTC) and also showed the greatest increase in nucleosome fragility (see **Figure 3D**). Thus, changes in occupancy and fragility are wholly separable. Taking into account our results together with those reported in the literature, the general conclusion is that regions with high fragility are generally also highly nucleosome occupied (using either MNase or H3 ChIP). Only during heat shock do we observe lower nucleosome occupancy and higher fragility. The reason for this behavior at heat shock genes remains to be elucidated by future experiments.

**Effects of DNA sequence content at nucleosome core and nucleosome linkers**

The sequence preference of MNase digestion bias has been established for decades (Hörz and Altenburger 1981; Dingwall et al. 1981; Cockell et al. 1983; Keene and Elgin 1981; McGhee and Felsenfeld 1983) and has been extensively reviewed. The consensus view is that while sequence preferences exist when tested on naked DNA, those effects are negligible on a chromatin substrate because the effect of nucleosomes dominates the cutting site preferences (Kaplan et al. 2010). To ensure that MNase sequence preferences are not responsible for the observed patterns in nucleosome fragility, we performed additional analyses to investigate how well nucleotide content in the linker, where most MNase cuts occur, can predict nucleosome fragility scores. We found that variance in fragility scores is better explained by the AT content of the nucleosome itself than it is by the sequence composition of the adjacent linkers. We calculated the average AT content of the two linkers flanking each nucleosome. Then, we asked whether the linker AT content was correlated with nucleosome fragility or resistance. Fragility scores are weakly correlated with AT content in the linker (R = 0.198, **Supplemental Figure S6B**. However, this correlation is not as strong as the correlation between AT content at the

nucleosome core and fragility scores observed at those nucleosomes (R = 0.378). We observed no correlation between resistant nucleosomes and linker GC content.

We performed an additional analysis to explore this issue further. We asked whether the weak correlation between linker AT content and nucleosome fragility scores could be explained by the correlation between AT content in the nucleosome core and AT content in the adjacent linkers (R=0.342, data not shown). The model without linker nucleotide content explained 23.54% of the variance in fragility scores. We asked how much additional variance in nucleosome fragility scores could be explained by adding the linker AT content to the model. After the addition of linker sequence content, the model increased by 0.73% to now explain 24.27% of the total variance (**Supplemental Figure S7F**). The model regarding variance in resistance scores did not change after addition of linker GC content.

In addition, we used a multiple linear regression model built on the variables above to predict fragility scores at all nucleosomes. In the version of the model where only GC content at the nucleosome core was taken into account, the correlation between observed and predicted fragility was 0.480 (data not shown). After addition of the average GC content of the flanking linkers, the correlation between observed and predicted fragility scores increased slightly to 0.487 (**Supplemental Figure S7G**).

To further test the differential influence of linker and core DNA, we asked how well these variables could predict fragility scores. The correlation between observed fragility scores and predicted fragility scores was 0.199 when linker GC content alone was used (**Supplemental Figure S7H)**, and 0.378 when nucleosome core GC content alone was used (**Supplemental Figure S7I)**. Addition of both core GC and linker GC content increased the correlation to 0.385 (**Supplemental Figure S7J)**. Thus, while flanking linker scores can provide small increases

helping to explain the overall fragility score of a nucleosome, the largest single component in explaining the variance in fragility scores is the GC content of the nucleosome itself, where the MNase enzyme is not acting. The results of this analysis, in conjunction with the consensus in the literature regarding the negligible influence of MNase sequence preferences on a chromatin template, indicate that our conclusions are not driven by MNase cleavage preferences.

## SUPPLEMENTAL METHODS

### MNase Digestion Time Course

MNase digestion was performed as previously described (Ercan et al. 2011), with slight alterations. Micrococcal nuclease (Worthington LS004798) was resuspended in water at 50 U/uL and frozen in individual aliquots at -80°C. To control for variability in enzyme activity, individual aliquots were removed from the freezer and thawed on ice, and never reused. Mixed-stage embryos were incubated with chitinase (Sigma Cat #C6137), washed, and dounced in dounce buffer (0.35 M sucrose, 15 mM HEPES-KOH pH 7.5, 0.5 mM EGTA, 5 mM $MgCl_2$, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% TritonX-100, 0.25% NP-40) to extract nuclei. Nuclei 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 pH 7.5). MNase was added, and at each timepoint (0, 2, 4, 8, 15, or 30 minutes after enzyme addition) a fraction of the reaction was removed, quenched with EDTA, and stored on ice. Samples were treated with Proteinase K for 2 hours at 55°C, then incubated overnight at 65°C to reverse crosslinks. DNA was isolated from RNA and proteins using phenol:chloroform extraction and RNase A treatment for 1 hour at 37°C. Mononucleosome-sized fragments (100 to 200 bp) were extracted from a 2% agarose gel and purified using a Qiagen gel extraction kit.

### Heat shock

Mixed-stage embryos were isolated as described and split into two pools. One pool was incubated at 34°C for 20 minutes with intermittent brief mixing, while the other pool nutated at room temperature. After 20 minutes, an aliquot from each pool was saved for RNA-seq, while the remaining embryos were fixed for 30 minutes in 2% formaldehyde at room temperature.

**RNA isolation**

Embryos were dropped into TRIzol (Life Technologies) and flash frozen in liquid nitrogen after incubation for 20 minutes at room temperature or 34°C heat shock. Embryos were homogenized by thawing at 37°C and refreezing in liquid nitrogen 3x. Total RNA was isolated using a TRIzol/chloroform extraction followed by RNeasy Mini (Qiagen) preparation with On Column DNaseI Digestion (Qiagen).

**Illumina Sequencing and Post-Processing**

Paired-end sequencing was performed by the Princeton University Sequencing Core Facility according to Illumina protocols. Paired end reads were mapped to the UCSC Oct. 2010 (WS220/ce10) genome release using Bowtie (v1.1.2) with stringent multimapping parameters:

bowtie -q -X 2000 --fr -p 1 -S -n 2 -e 70 -l 28 --pairtries 100 --maxbts 125 -k 1 -m 1 --un /Unmapped_Reads.fastq --phred33-quals /ce10 -1 /read1.fastq -2 /read2.fastq

**Nucleosome analysis**

Reads with insert sizes between 100 and 250 bp were kept for downstream analysis.

Replicates were first processed individually, then pooled after confirming a high degree of correlation between replicates. Nucleosome analysis was performed as described previously (Kaplan et al. 2010; Gossett and Lieb 2012).

*Coverage:* Nucleosome coverage was calculated by extending the filtered mapped reads to their fragment length and measuring the sum of reads covering each bp. To normalize for variation between samples, nucleosome coverage was scaled by 1/(mean coverage), yielding a mean nucleosome coverage of 1.0.

*Dyad Occupancy:* Dyads are approximated as the center of a paired-end fragment. The number of dyads at each base pair was scaled by 1/(mean dyad density), then Gaussian smoothed with a standard deviation of 20 bp.

*Nucleosome calls:* Nucleosome positions were identified from dyad density maps using a previously reported greedy algorithm (Albert et al. 2007; Gossett and Lieb 2012). Using the local maxima of the dyad density as the nucleosome center *p*, the size of the nucleosome (the nucleosome-protected region) was determined by measuring the average length of all reads that covered the nucleosome center. The standard deviation of the nucleosome center (the nucleosome "fuzziness") was calculated for each called nucleosome as the standard deviation of dyads around the mean. Nucleosome occupancy was defined as the number of dyads that fell within 50 bp of the nucleosome center.

*Boundary nucleosomes:* Using these called nucleosome positions 5′ and 3′ boundary nucleosomes were identified for the 20,578 RefSeq annotated genes. 5′ +1 nucleosomes were identified as the first nucleosome call with a dyad coordinate downstream of the 1st coding exon. Similarly, the 3′ boundary nucleosome was identified as the first nucleosome call with a dyad coordinate upstream of the TTS. Because *C. elegans* utilizes trans-splicing, the 5′ end of mature polyadenylated mRNAs, and the RefSeq annotations used in this analysis, do not reflect the exact base pair position of transcription initiation. Although recent studies have used novel methods to identify the true transcription initiation sites (Chen et al. 2013; Kruesi et al. 2013; Saito et al. 2013), the TSS annotations are only known for a subset of expressed genes in a

small number of stages (Chen et al. only tested embryos, only 31.7% of genes had a TSS in at least 1 of 3 stages tested in Kruesi et al., Saito et al., only tested embryo and adult). For completeness, we chose to instead investigate the full set of known genes using their first coding exon as an alignment point.

**Nucleosome Fragility and Resistance scores**

The pooled "intermediate" nucleosome profile was generated by pooling the reads from the 4, 8, and 15 minute time points from each replicate. The pooled reads were used to generate average nucleosome positions, fuzziness scores, and occupancies. To identify regions of the genome that were liberated earlier or later than average, we subtracted the occupancy of the pooled sample from either the 2 minute (2m – pool = Fragility score) or the 30 minute samples (30m – pool = Resistance score). To highlight regions significantly enriched with this signal, we considered the 10% of nucleosomes with the highest fragility or resistance scores as Fragile or Resistant Nucleosomes.

**Nucleosome occupancy scores**

Reads from the 4, 8, and 15 minute digestion timepoints were pooled and used to calculate nucleosome coverage and nucleosome dyad occupancy scores (see above). Nucleosome dyad occupancy is referred to as "Nucleosome Occupancy".

**Genes associated with fragile or resistant nucleosomes**

We found fragile nucleosomes enriched at the -1 nucleosome and resistant nucleosomes were enriched at the +1 nucleosome (**Fig. 4**). Given the strong association with the gene start, we identified two sets of genes that contain fragile and resistant nucleosomes, respectively, +/- 500 bp from their transcript start site.

**Gene ontology analysis**

Gene lists were uploaded to the FatiGO web server (babelomics.bioinfo.cipf.es) and compared against the background set of all *C. elegans* genes (Al-Shahrour et al. 2004). P-values were calculated using the Fisher's exact test, and corrected for multiple testing using the FDR procedure of Benjamini and Hochberg (Benjamini and Hochberg 1995). Corrected p-values and GO terms were then input in to REVIGO to reduce and visualize significantly enriched GO clusters (Supek et al. 2011).

**Stable and developmentally regulated genes**

Pre-normalized transcriptome sequencing data was downloaded from: https://www.encodeproject.org/comparative/transcriptome/ (Gerstein et al. 2014; Spencer et al. 2011). For each gene, we calculated the coefficient of variation (CV): $c_v = \frac{\sigma}{\mu}$. We took the 1000 genes with the highest CVs as the set of developmentally regulated genes, and the set of 1000 genes with the lowest CVs as the set of stably expressed genes.

**RNA-seq analysis**

Unstranded mRNA libraries were prepared from total RNA for RNA-seq using the Illumina TruSeq RNA Library Preparation Kit v2 (RS-122-2001). RNA-seq reads were mapped to the *C. elegans* WS220 Gene Annotation Model using Tophat2 (v0.7) (Trapnell et al. 2012). The resulting alignment files were quantified using HT-Seq (v0.4.1) and the RefSeq gene annotations for WS220 (Anders et al. 2015). Total read counts per gene were normalized for differential expression using DESeq2 (v1.0.19) in R (v3.0.1) (Love et al. 2014).

**SUPPLEMENTAL FIGURE LEGENDS**

**Figure S1:** Accompanying Figure 1.

**(A)** Representative image of an N2 embryo MNase digestion timecourse after gel electrophoresis. For each timepoint, mononucleosome-sized fragments were excised from the gel (white box) and used for paired-end Illumina DNA sequencing. Size markers ("M") are indicated.

**(B,C)** Mononucleosome fragments are shorter with increasing MNase digestion time, in 10 bp increments. Comparison of nucleosome sizes between native (**B**) and formaldehyde-fixed (**C**) timecourse**.**

**(D→G)** Heatmaps showing distribution of (**D**) fragile and (**E**) resistant nucleosomes in native mixed stage embryos, as well as (**F**) fragility and (**G**) resistance in fixed mixed stage embryos. Genes are sorted by gene length, maximum gene length shown is 3500 bp. Genes are aligned at the first nucleosome downstream from the transcript start site (+1 nucleosome), named the 5′ boundary nucleosome. Yellow line cartoons the center of the +1 and terminal nucleosomes. (**H**) Line plot comparing fragility and resistance in native (dotted lines) and fixed (solid lines) embryos around the 5′ (**left**) and 3′ (**right**) boundary nucleosomes.

**Figure S2:** Accompanying Figure 1.

**(A,B)** Cross – timecourse and cross – replicate Spearman correlation in 150 bp windows between two replicates of mixed-stage embryo MNaseTC (Jee et al. 2011). RT = replicate 1, 2RT = replicate 2. (**A**) Unclustered timepoints. (**B**) Hierarchical clustering of Spearman correlation between timepoints.

**(C→F)** Heatmaps from (**C**) 2 minute, (**D**) average of 4, 8, and 15 minute nucleosomes, and (**E**) 30 minute nucleosomes, sorted by length. Genes are aligned at "0", the center of the 1$^{st}$ nucleosome downstream from the transcript start site. Nucleosome dyad occupancy is shown in blue – yellow. (**F**) Line plot average of dyad density around the 5′ and 3′ boundary nucleosomes

for the early (2 minute), intermediate (average of 4, 8, and 15 minute), and late (30 minute) timepoints.

**(G)** Scatterplot comparing fragility and resistance scores at 448,600 nucleosomes. Fragility Score = 41.2 is the threshold above which a nucleosome is considered a “fragile nucleosome”. Resistance Score = 37.9 is the threshold above which a nucleosome is considered a “resistant nucleosome”. Plot is colored by density of points. Purple = least dense, Red = most dense.

**Figure S3:** Accompanying Figure 2. Ubiquitously HOT regions are fragile *in vivo.* Average fragility, resistance, and intermediate nucleosome occupancy plotted around the center of 847 ubiquitously HOT regions identified in Araya et al (Araya et al. 2014).

**Figure S4**: Accompanying Figure 3.

**(A)** RNA-seq cross-correlation plot for two replicates of room temperature and two replicates of heat shock embryos.

**(B)** Spearman correlation between two replicates of heat shock and two replicates of room temperature control mixed stage embryo MNaseTC experiments, unclustered (**B**) or hierarchically clustered (**C**) (Jee et al. 2011).

**Figure S5:** Accompanying Figure 4.

(**A**) Scatterplots comparing intermediate nucleosome occupancy (average of 4, 8, and 15 minute nucleosomes) and fragility scores, 2 minute nucleosome occupancy and fragility scores, intermediate nucleosome occupancy and resistance scores, 30 minute nucleosome occupancy and resistance scores, 2 minute nucleosome fuzziness and fragility scores, intermediate nucleosome fuzziness and fragility scores, intermediate nucleosome fuzziness and resistance scores, and 30 minute nucleosome fuzziness and resistance scores.

(**B**) Nucleosome occupancy scores from histone H3 ChIP-seq around the 5′ and 3′ boundary nucleosomes averaged over expression quintiles (highest expressed 20% in black, lowest expressed 20% in lightest gray) (Ho et al. 2014a). Quintile 1: 0 to 4.5 normalized counts. Quintile 2: 4.5 to 65. Quintile 3: 65 to 619. Quintile 4: 619 to 2209. Quintile 5: > 2209.

(**C**) Same as in (**B**), except nucleosome occupancy scores from the intermediate (average of 4, 8, and 15 minute timepoints) as measured in this study by MNase-seq are plotted.

(**D**) $Log_2$ DESeq2-normalized number of reads measured by mRNA-seq at 20,785 genes, ordered by their relative expression.

(**E**) Heatmap of H3 ChIP-seq scores at genes ordered as in (**D**). Genes were aligned at the center of the first nucleosome downstream from the transcript start site, known as the +1 or 5′ Boundary Nucleosome. ChIP-seq data from Ho et al., 2014.

(**F**) Same as in (**E**), except intermediate nucleosome occupancy scores measured by MNase-seq are plotted.

(**G**) Scatterplots comparing nucleosome occupancy at 448,600 nucleosomes as measured by histone H3 ChIP-seq or by MNase-seq performed in this study.

(**H**) Scatterplots comparing nucleosome occupancy as measured by histone H3 ChIP-seq and nucleosome fragility or resistance scores measured in this study.

**Figure S6:** Accompanying Figure 4.

**(A)** Transcription factor binding sites in the embryo are biased towards expressed genes. Heatmaps sorted by gene expression (highest expressed genes at top). modENCODE ChIP-seq of transcription factors from embryo stages are shown for comparison. EM: embryo. LE: late embryo. EE: early embryo. mxdE: mixed-stage embryo. ChIP data from Araya et al. 2014 (Araya et al. 2014), expression data from mixed-stage embryos, this study.

**(B)** Scatterplots displaying the correlation between nucleosome fragility (**top**) or nucleosome resistance (**bottom**) and expression (**left**), GC content at nucleosome core (**middle-left**),

average of GC content at two flanking linkers (**middle-right**), and HTZ-1 occupancy (**right**). Expression is measured as the number of normalized RNA-seq counts that mapped to a nucleosome. HTZ-1 occupancy is from (Ho et al. 2014).

**(C)** The gene body of highly expressed genes is not fragile. Fragility scores averaged at the highest 20% (green), highest 5% (pink), or highest 1% (yellow) of expressed genes in the embryo. Genes are aligned at the +1 nucleosome.

**(D)** H2A.Z containing nucleosomes are not enriched at fragile or resistant nucleosomes. Of the total 5,042 H2A.Z nucleosomes, 1,486 were also identified to be fragile, while 2614 were identified as resistant (Ho et al. 2014a).

**Figure S7:** Accompanying Figure 5.

**(A)** Fragile nucleosomes are enriched for TATA box motifs; both fragile and resistant nucleosomes are underrepresented with T-block motifs. 146 bp sequences underneath fragile, resistant, and random nucleosomes were searched for perfect matches to TATAAA, TTTTTT (T6), or TTTTTTTTT (T9) motifs using FIMO (Grant et al. 2011). Numbers of perfect matches were normalized to the number of nucleosomes used: fragile nucleosomes, 48,110; resistant nucleosomes, 48,089; randomly chosen nucleosomes, 48,000. All comparisons were significant by chi-square test with a p-value < 2.2E-16.

**(B)** Spearman correlation between histone modifications, histone variants, and other ChIP-seq experiments in 150 bp sliding windows. 80 mM and 350 mM extracted nucleosomes from (Ooi et al. 2010). LEM-2 from (Ikegami et al. 2010). All other data from (Ho et al. 2014a).

**(C)** Fragile nucleosomes are associated with 80 mM salt-extracted nucleosomes. Resistant nucleosomes are associated with H3.3, H3K36me1/2/3, H3K79me1. Histone post-translational modifications score at all nucleosomes (gray), fragile nucleosomes (green), or resistant nucleosomes (orange). Histone modifications from (Ho et al. 2014a). H3.3 and 80 mM extracted chromatin from (Ooi et al. 2010).

(**D**) Enrichment of fragile and resistant nucleosomes at chromatin states. Chromatin states from (Ho et al. 2014b). Blank box: no enrichment of histone post-translational modification. +, ++, +++, ++++: relative enrichment of histone post-translational modification in the chromatin state. **(E)** Longer linkers are weakly correlated with (**left**) higher nucleosome fragility and (**right**) decreased nucleosome resistance. Flanking liner length: total length of linker surrounding a nucleosome (upstream + downstream). Fragility quartiles: all nucleosomes are assigned a fragility score. Nucleosomes are broken into equally sized quartiles. Q1 = 120,205 nucleosomes with lowest fragility scores, Q4 = 120,205 nucleosomes with highest fragility scores. Same process applied for resistance quartiles. Statistical significance evaluated with chi square test. * p-value < 0.05, ** p-value < 1e-08, *** p-value < 1e-12. All other comparisons not significant. **(F)** GC content of the nucleosome core is the strongest predictor of nucleosome fragility and resistance scores. Relative importance of each feature for predicting nucleosome fragility and resistance scores was calculated using the 'lmg' metric from the R package "relimpo": the $R^2$ contribution averaged over orderings among regressors (Gromping 2006). (**Top**) The model composed of %gc + genic + expr + fStd + iStd + linker + HelT + MGW + OCR2 + ProT + Roll + WW + SS + PhastC + LinkGC explains 24.27% of the variance in nucleosome fragility scores. Of the explained variance, GC content is the most explanatory. (**Bottom**) The model composed of %gc + genic + expr + fStd + iStd + linker + HelT + MGW + OCR2 + ProT + Roll + WW + SS + PhastC + LinkGC explains 8.13% of the variance in nucleosome resistance scores. Of the explained variance, GC content is the most explanatory. %gc = average GC content of nucleosome core, gene = genic or non genic sequence, expr = expression level, fStd = nucleosome fuzziness at 2m timepoint, iStd = nucleosome fuzziness at average of intermediate timepoints (4, 8, 15m), linker = length of flanking linker, fOcc = nucleosome occupancy at 2m timepoint, iOcc = nucleosome occupancy at average of intermediate timepoints (4, 8, 15m), HelT = helical twist, MGW = major groove width, ORC2 = hydroxyl radical cleavage ORChID2) , ProT = propeller twist, Roll = DNA roll, WW = frequency of (AA, AT, TA, or TT) dinucleotides,

SS = frequency of (GG, GC, CG, CC) dinucleotides, PhastC = PhastCons concervation score of nucleosome core sequence, LinkGC = average % GC content of two flanking linker sequences. DNA shape properties are from (Chiu et al. 2015).

(**G**) Correlation between observed and predicted fragility (**top**) and resistance (**bottom**) scores using the model: score = %gc + genic + expr + fStd + iStd + linker + HelT + MGW + OCR2 + ProT + Roll + WW + SS + LinkGC. Model was trained on 90% of the data and tested on the remaining 10%.

(**H→J**) Correlation between observed and predicted fragility scores using either (**H**) only average linker GC content, (**I**) only GC content of nucleosome core, or (**J**) both average linker GC content and GC content of nucleosome core.

(**K,L**) Normalized signal from (**K**) 80 mM salt-extracted nucleosomes and **(L)** H3.3 ChIP'd nucleosomes from (Ooi et al. 2010). Genes are sorted by expression level as measured in the current study.

(**M,N**) Correlation between (**M**) 80 mM salt extracted nucleosome occupancy and fragility scores measured in this study. (**N**) Correlation between H3.3 occupancy and resistance scores measured in this study.

**SUPPLEMENTAL TABLE LEGENDS:**

**Table S1: Additional datasets used in this study.** Table provides details on which datasets were used, which figures in this paper used the data, the original reference for the dataset, and from where the dataset was downloaded.

**SUPPLEMENTAL REFERENCES**


Al-Shahrour F, Diaz-Uriarte R, Dopazo J. 2004. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20**: 578–580.

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.

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Araya CL, Kawli T, Kundaje A, Jiang L, Wu B, Vafeados D, Terrell R, Weissdepp P, Gevirtzman L, Mace D, et al. 2014. Regulatory analysis of the C. elegans genome with spatiotemporal resolution. *Nature* **512**: 400–405.

Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrancois P, Struhl K, Gerstein M, Snyder M. 2009. Mapping accessible chromatin regions using Sono-Seq. *Proceedings of the National Academy of Sciences* **106**: 14926–14931.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**: 289–300.

Chen RAJ, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LW, Jeffers TE, Ahringer J. 2013. The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. *Genome Research*.

Chereji RV, Kan T-W, Grudniewska MK, Romashchenko AV, Berezikov E, Zhimulev IF, Guryev V, Morozov AV, Moshkin YM. 2015. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. *Nucleic Acids Research* **44**: 1036–1051.

Chiu TP, Yang L, Zhou T, Main BJ, Parker SCJ, Nuzhdin SV, Tullius TD, Rohs R. 2015. GBshape: a genome browser database for DNA shape annotations. *Nucleic Acids Research* **43**: D103–D109.

Cockell M, Rhodes D, Klug A. 1983. Location of the Primary Sites of Micrococcal Nuclease Cleavage on the Nucleosome Core. *Journal of Molecular Biology* **170**: 423–446.

Dingwall C, Lomonossoff GP, Laskey RA. 1981. High sequence specificity of micrococcal nuclease. *Nucleic Acids Research* **9**: 2659–2673.

Ercan S, Lubling Y, Segal E, Lieb JD. 2011. High nucleosome occupancy is encoded at X-linked gene promoters in C. elegans. *Genome Research* **21**: 237–244.

Flensburg C, Kinkel SA, Keniry A, Blewitt ME, Oshlack A. 2014. A comparison of control samples for ChIP-seq of histone modifications. *Frontiers In Genetics* **5**: 1–8.

Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, et al. 2014. Comparative analysis of the transcriptome across distant species. *Nature*

**512**: 445–448.

Gossett AJ, Lieb JD. 2012. In Vivo Effects of Histone H3 Depletion on Nucleosome Occupancy and Position in Saccharomyces cerevisiae ed. O.J. Rando. **8**: e1002771. http://dx.plos.org/10.1371/journal.pgen.1002771.s016.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Gromping U. 2006. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software* **17**: 1–27.

Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A, et al. 2014a. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449–452.

Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A, et al. 2014b. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449–452.

Hörz W, Altenburger W. 1981. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Research* **9**: 2643–2658.

Ikegami K, Egelhofer TA, Strome S, Lieb JD. 2010. Caenorhabditis elegans chromosome arms areanchored to the nuclear membrane viadiscontinuous association with LEM-2. *Genome biology* **11**: R120.

Jee J, Rozowsky J, Yip KY, Lochovsky L, Bjornson R, Zhong G, Zhang Z, Fu Y, Wang J, Weng Z, et al. 2011. ACT: aggregation and correlation toolbox for analyses of genome tracks. *Bioinformatics* **27**: 1152–1154.

Kaplan N, Hughes TR, Lieb JD, Widom J, Segal E. 2010. Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. *Genome biology* **11**: 1–12.

Keene MA, Elgin SCR. 1981. Micrococcal Nuclease as a Probe of DNA Sequence Organization and Chromatin Structure. *Cell* **27**: 57–64.

Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. 2013. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife* **2**: e00808–e00808.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.

McGhee JD, Felsenfeld G. 1983. Another Potential Artifact in the Study of Nucleosome Phasing by Chromatin Digestion with Micrococcal Nuclease. *Cell* **32**: 1205–1215.

Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ, et al. 2016. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nature Communications* **7**: 1–11.

Ooi SL, Henikoff JG, Henikoff S. 2010. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Research* **38**: e26–e26.

Pérez-Lluch S, Blanco E, Tilgner H, Curado J, Ruiz-Romero M, Corominas M, Guigó R. 2015. Absence of canonical marks of active chromatin in developmentally regulated genes. *Nature Publishing Group* **47**: 1158–1167.

Saito TL, Hashimoto SI, Gu SG, Morton JJ, Stadler M, Blumenthal T, Fire A, Morishita S. 2013. The transcription start site landscape of C. elegans. *Genome Research* **23**: 1348–1361.

Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, et al. 2011. A spatial and temporal map of C. elegans gene expression. *Genome Research* **21**: 325–341.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms ed. C. Gibas. *PlOS ONE* **6**: e21800.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**: 562–578.