Fig. S1

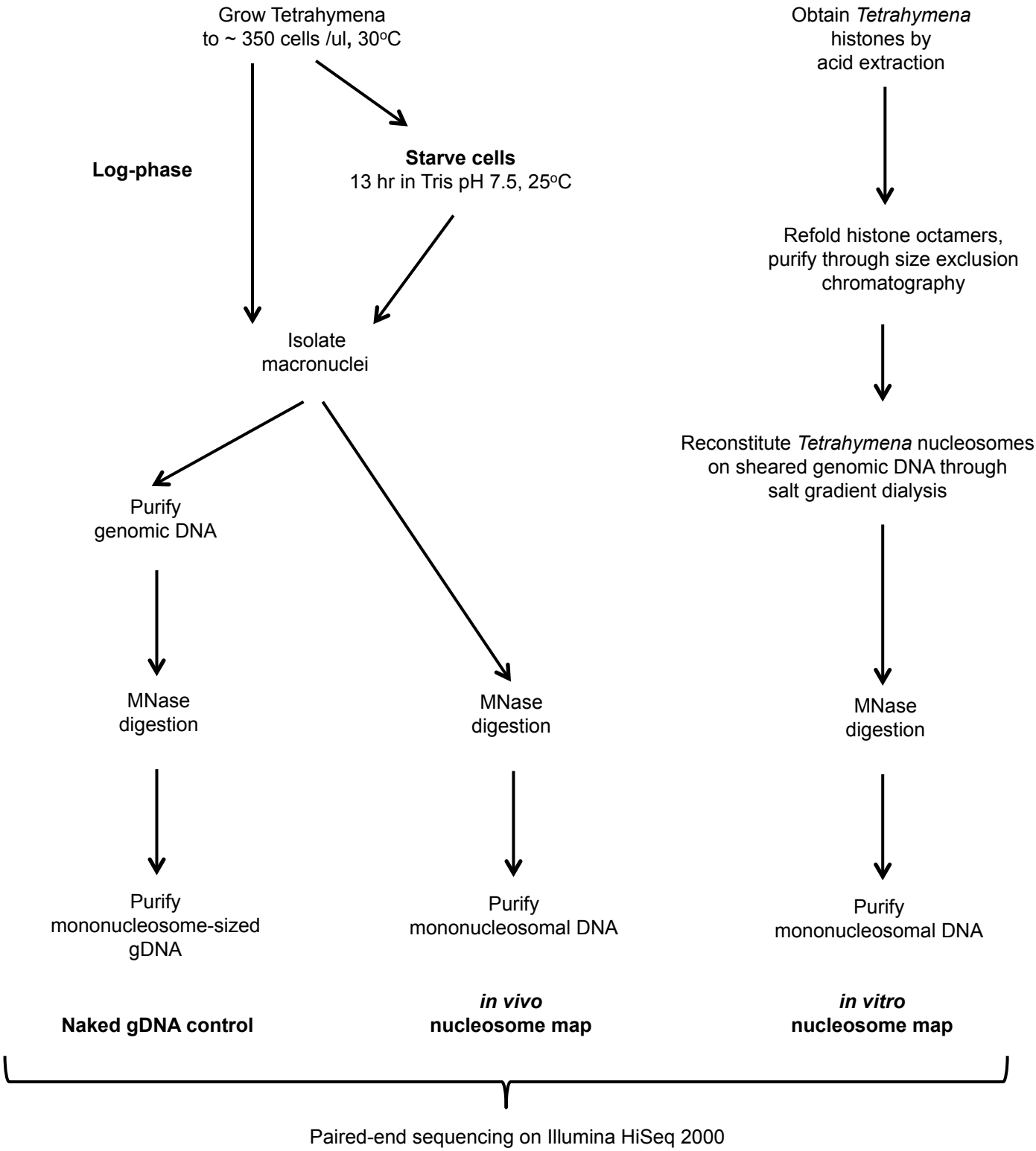


Figure S1. Workflow of *Tetrahymena* nucleosome mapping experiments. Macronuclei were isolated from starved or log-phase *Tetrahymena* and digested with MNase. Separately, *Tetrahymena* histones were acid-extracted, refolded into octamers, assembled on genomic DNA through salt gradient dialysis, and subsequently treated with MNase. No *trans*-acting factors were added during chromatin assembly. The mononucleosomal DNA from *in vivo* and *in vitro* MNase digests was gel-purified for subsequent Illumina sequencing.

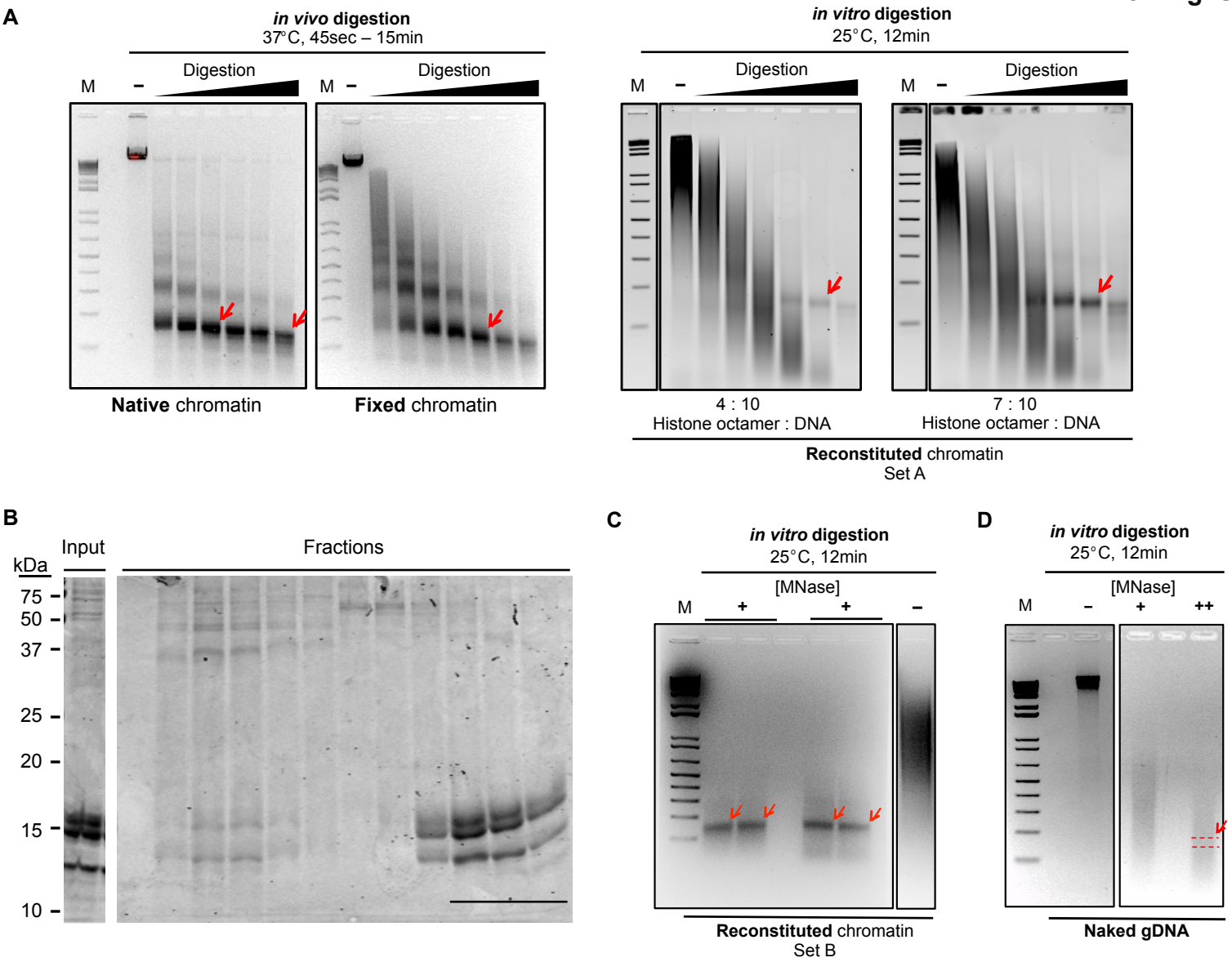


Figure S2. Gel analysis of *Tetrahymena* chromatin samples. (A) Macronuclei from log-phase or starved cells yielded nucleosome ladders upon MNase digestion *in vivo*, similar to other eukaryotes. A protected mononucleosome-sized fragment was observed after *in vitro* reconstituted chromatin after MNase treatment, with no evidence of laddering. Mononucleosomal DNA samples marked with a red arrow were gel-purified for subsequent paired-end sequencing. (B) Size exclusion chromatography of refolded *Tetrahymena* histone octamers. The fractions highlighted with a horizontal black bar were pooled and concentrated for subsequent *in vitro* reconstitution experiments with *Tetrahymena* genomic DNA. (C) Light and heavy digestion of *in vitro* reconstituted chromatin. Samples are from reconstitution set "B" (see Supplemental Methods). They were digested with either (33 Kunitz Units) or 21.96 µl (66 Kunitz Units) MNase, and labeled as light ("+") and heavy ("++") digest, respectively. Mononucleosome-sized fragments denoted by the red arrow were excised and gel-purified, for subsequent paired-end sequencing. (D) MNase digestion of naked *Tetrahymena* macronuclear gDNA. The DNA was digested with either 2.37 or 4.74 Kunitz Units of MNase (labeled as "+" and "++" respectively). A ~150bp mononucleosome-sized fragment was excised, as denoted by the dotted red lines and arrow. This sample was gel-purified and subjected to paired-end sequencing. 'M' denotes kb(+) DNA ladder (Life Technologies); 'kDa' denotes Precision Plus Protein Dual Color protein ladder (Bio-Rad).

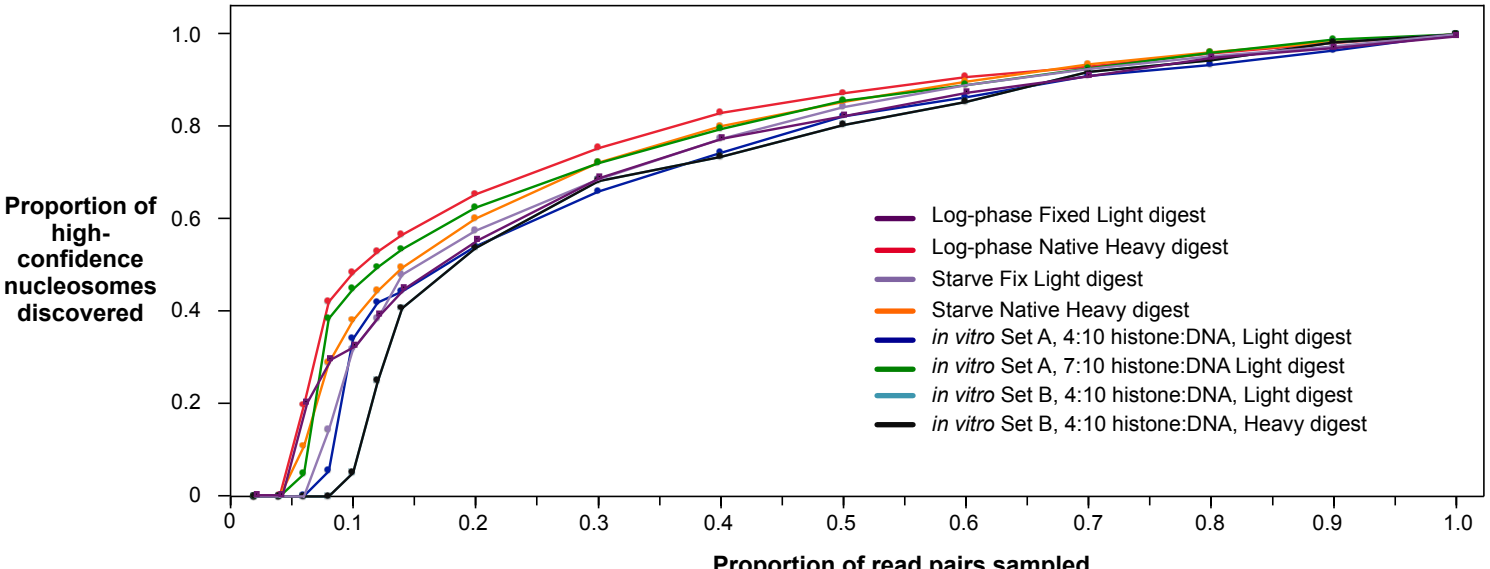


Figure S3. Subsampling of MNase-seq data. Varying fractions of mapped read pairs from each dataset were randomly subsampled, and used for nucleosome calling through DANPOS (Chen et al. 2013a). Reads pairs were mapped to all chromosomes in the October 2008 build of the *Tetrahymena* SB210 reference genome (Eisen et al. 2006). Each point represents subsampling of a particular fraction of read pairs. The number of called high confidence nucleosomes ($p < 1e^{-8}$) approached saturation before full sampling of *in vivo* and *in vitro* data, indicating that nucleosomes are well-sampled in all datasets.

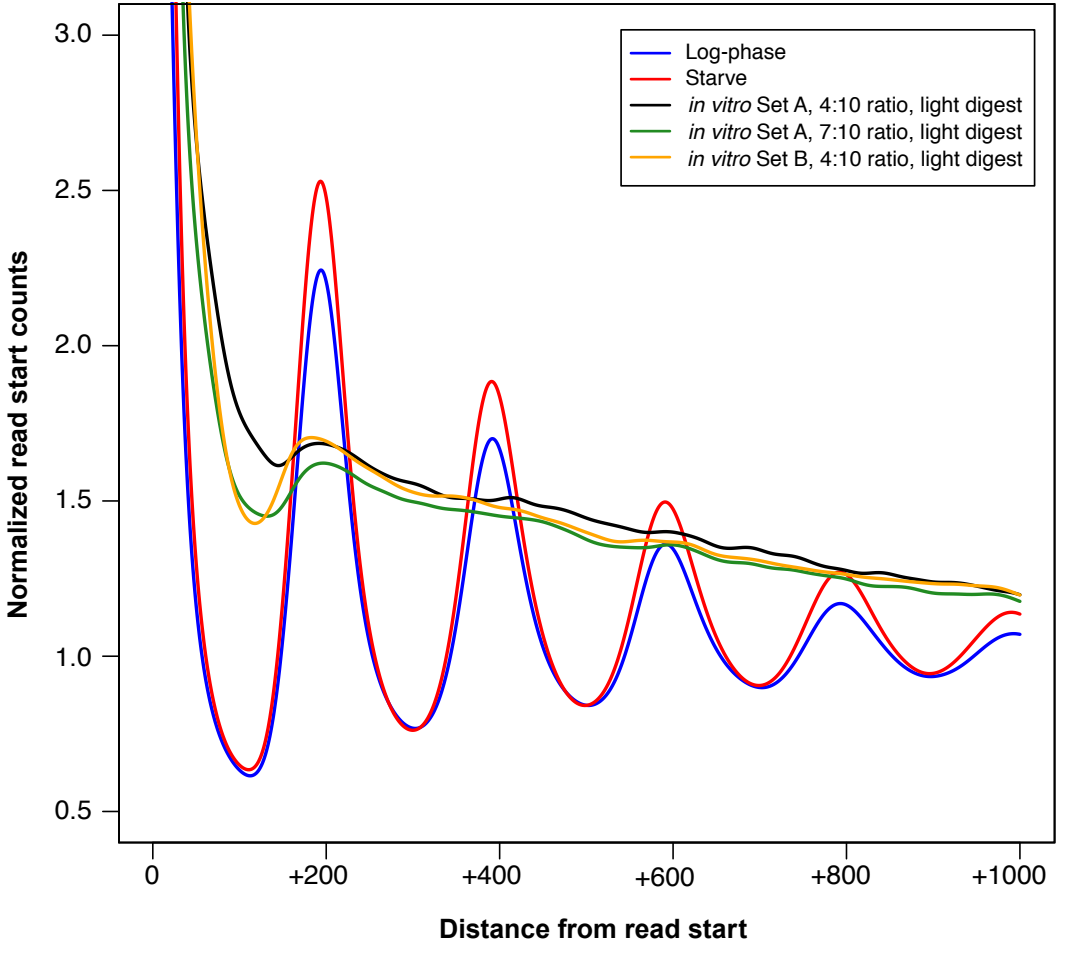


Figure S4. Phasograms of *in vitro* and *in vivo* MNase-seq datasets. A distinct 200bp periodicity is specifically observed within *in vivo* datasets (log-phase and starve), suggesting the presence of regular nucleosome arrays. This is consistent with our gel analysis (Supplemental Fig. S2A) and other independent studies (Gorovsky et al. 1978).

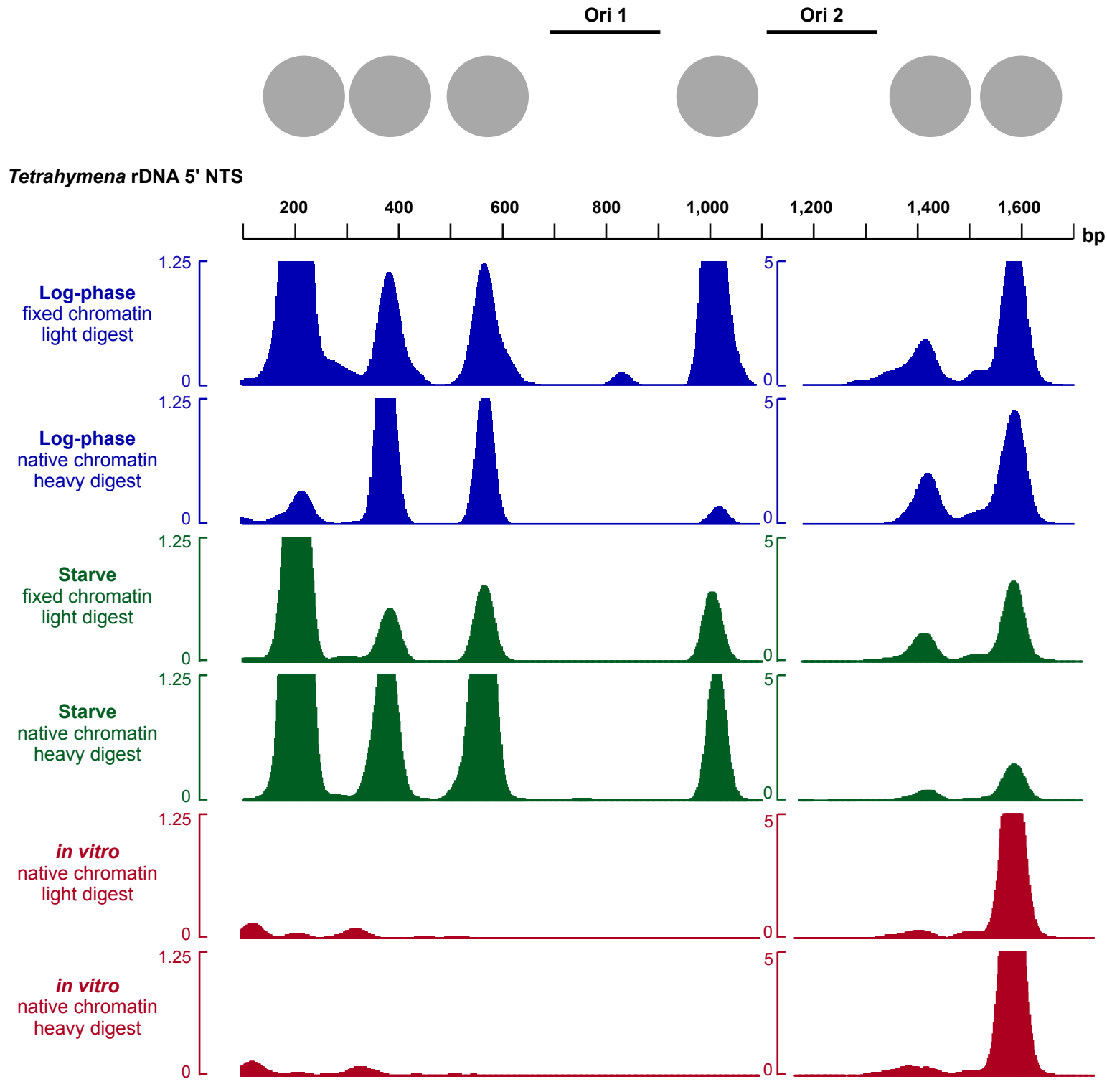


Figure S5. Nucleosome dyad counts along the 5' NTS of the *Tetrahymena* ribosomal DNA locus. Only uniquely mapping reads were considered when tabulating nucleosome dyads from MNase-seq reads, at this locus. Blue and green tracks represent *in vivo* data from chromatin digested to different extents with MNase. Well-positioned nucleosomes *in vivo* flank both origins of replication *in vivo*, corroborating independent studies that mapped nucleosome positions through Southern analysis. We excluded the 5'-most end (0-200 bp) of the rDNA locus from this analysis, because of potentially ambiguous mapping of the 5' end of read pairs to the adjacent palindromic arm of this locus.

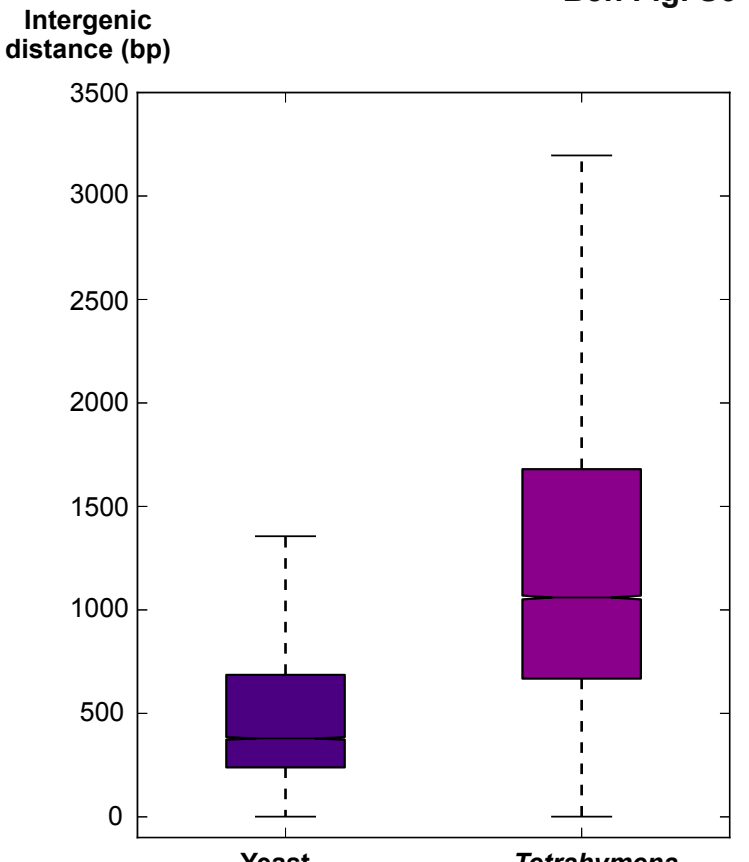


Figure S6. Distribution of intergenic distances in *Tetrahymena* and yeast. An intergenic distance is designated as the length between 3' end of the open reading frame of a gene, and the 5' end of an adjacent open reading frame. *Tetrahymena* intergenic regions tend to be longer than those in yeast.

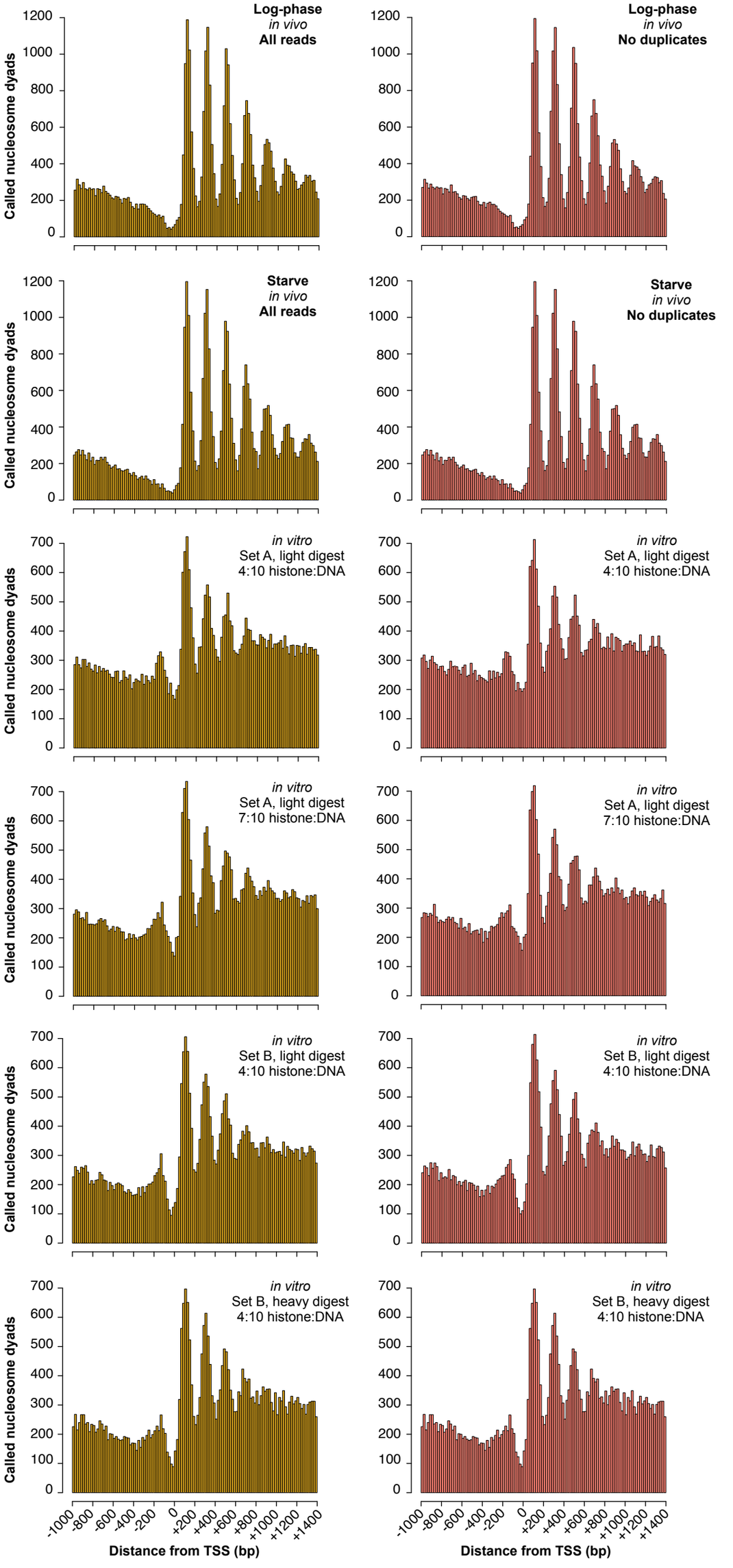


Figure S7. Removing duplicate reads does not affect the *in vivo*-like nucleosome organization near TSSs *in vitro*. Histograms of called nucleosome dyads within each *in vitro* and log-phase *in vivo* datasets, around TSSs. The various *in vitro* datasets analyzed here are described in Methods. Stringent filters for absolute and conditional nucleosome positioning were applied, such that ~35% of nucleosomes were discarded. The nucleosome organization remains similar *in vivo* and *in vitro* even when duplicate reads are removed, suggesting that it is not an artifact arising from over-amplification of Illumina libraries.

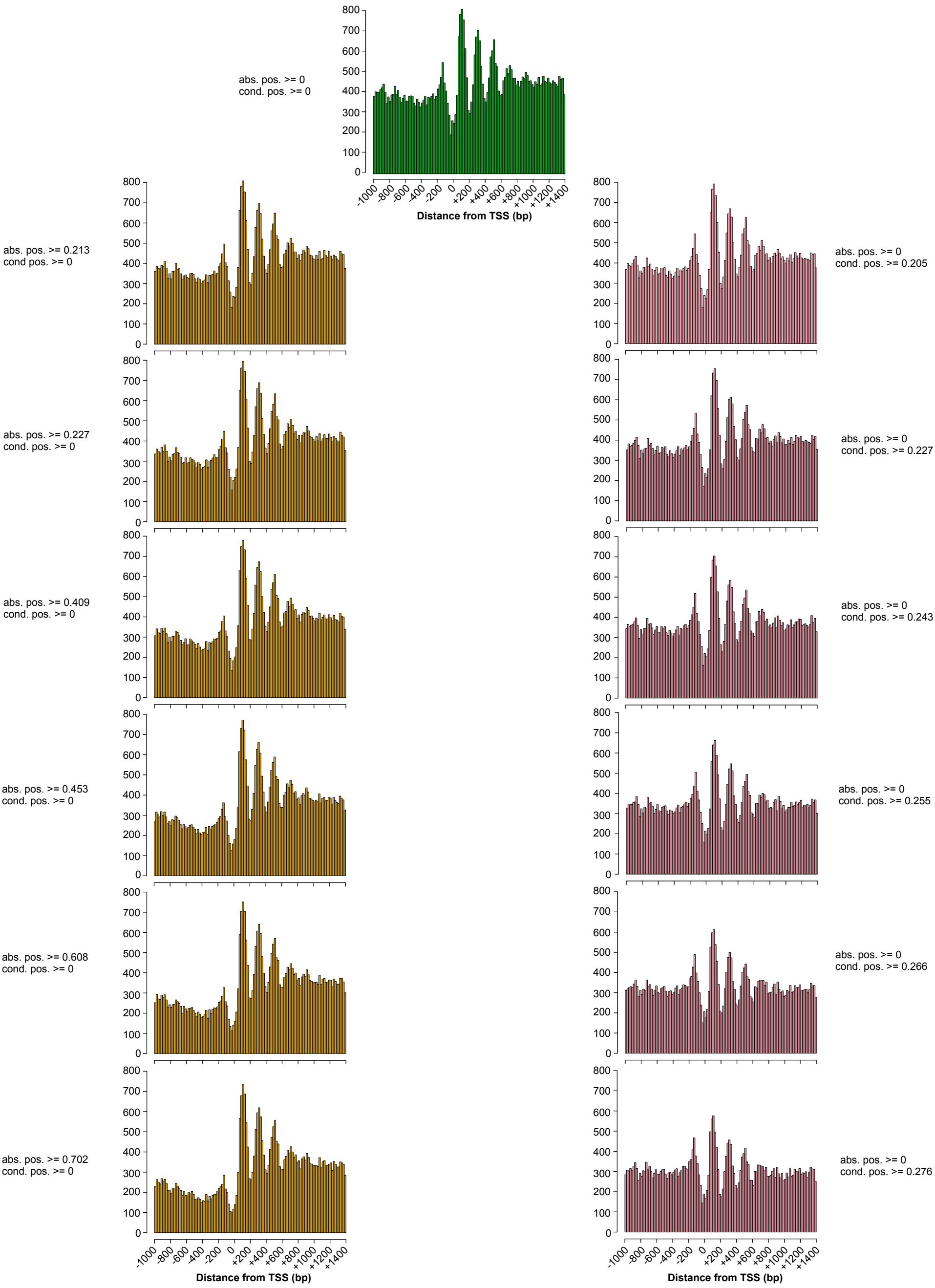


Figure S8. The phased pattern of *in vitro* nucleosome positions is robust to variation in nucleosome calling criteria. Cutoffs for absolute positioning (abs. pos.) and conditional positioning (cond. pos.) were separately varied, such that up to 30% of called nucleosomes were respectively removed. The filtered data were then used to plot histograms of called nucleosome positions, relative to the TSS. The regular organization of nucleosomes downstream of TSSs is readily observed, even when using the most stringent filtering criteria.

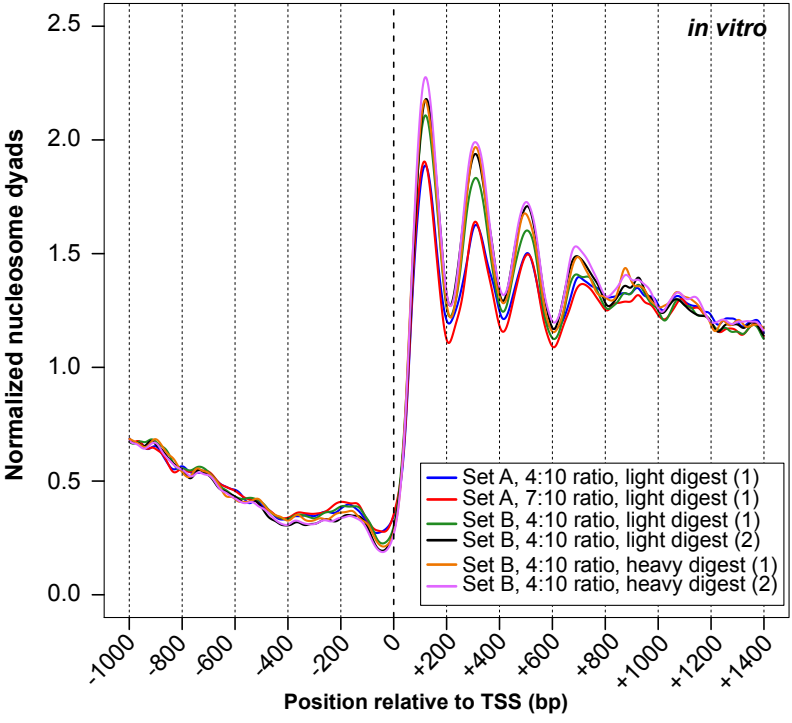


Figure S9. The *in vivo*-like pattern in reconstituted chromatin is robust to experimental variation. Averaged nucleosome dyad counts around the TSS are plotted for each *in vitro* reconstitution experiment performed in this study (see Methods for explanation of each sample). Bracketed numbers indicate individual replicates. A regular pattern is clearly observed in all samples, encompassing variation in MNase digestion, histone:DNA ratio, and reconstitution reaction volume.

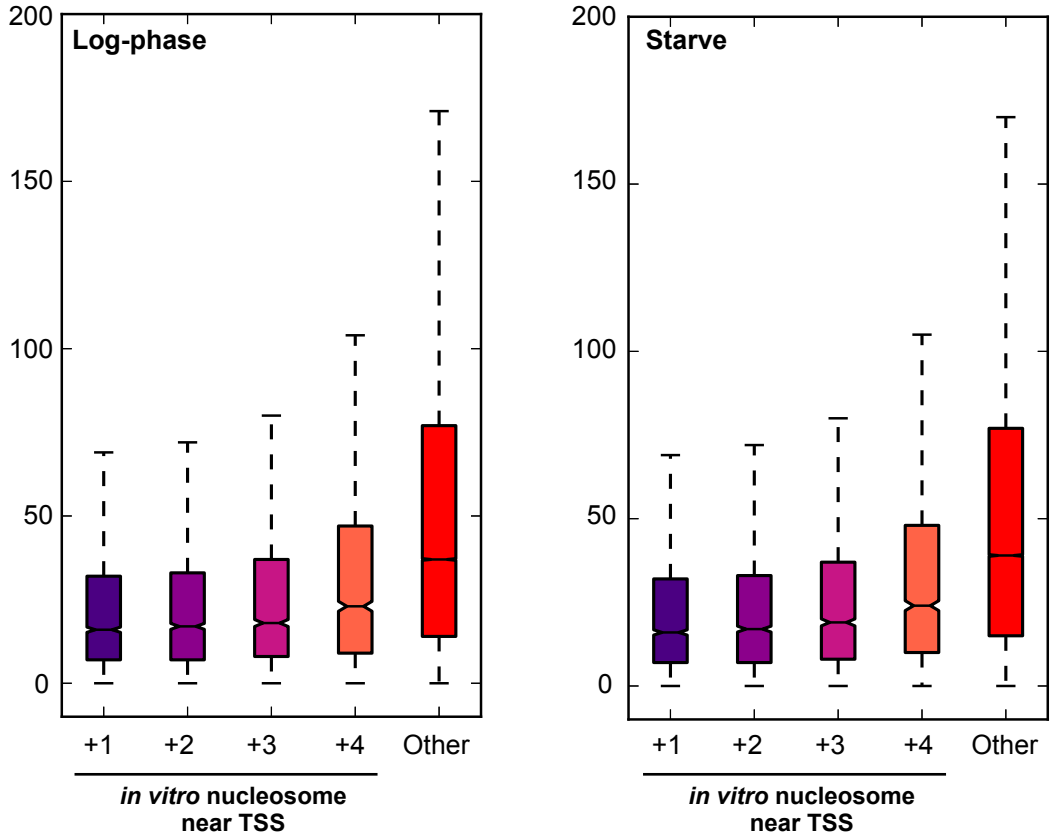


Figure S10. Sites closest to the TSS show greatest correspondence between *in vitro* and *in vivo* nucleosome positions. For *in vitro* nucleosomes in the +1, +2, +3, and +4 positions downstream of the TSS, the distance to the nearest *in vivo* nucleosome is calculated. "Other" represents *in vitro* nucleosomes in the genome that not located at +1 to +4 positions. Nucleosomes at the +1 position *in vitro* most closely overlap with a nucleosome *in vivo*, suggesting that the +1 nucleosome is intrinsically favorable for nucleosome formation.

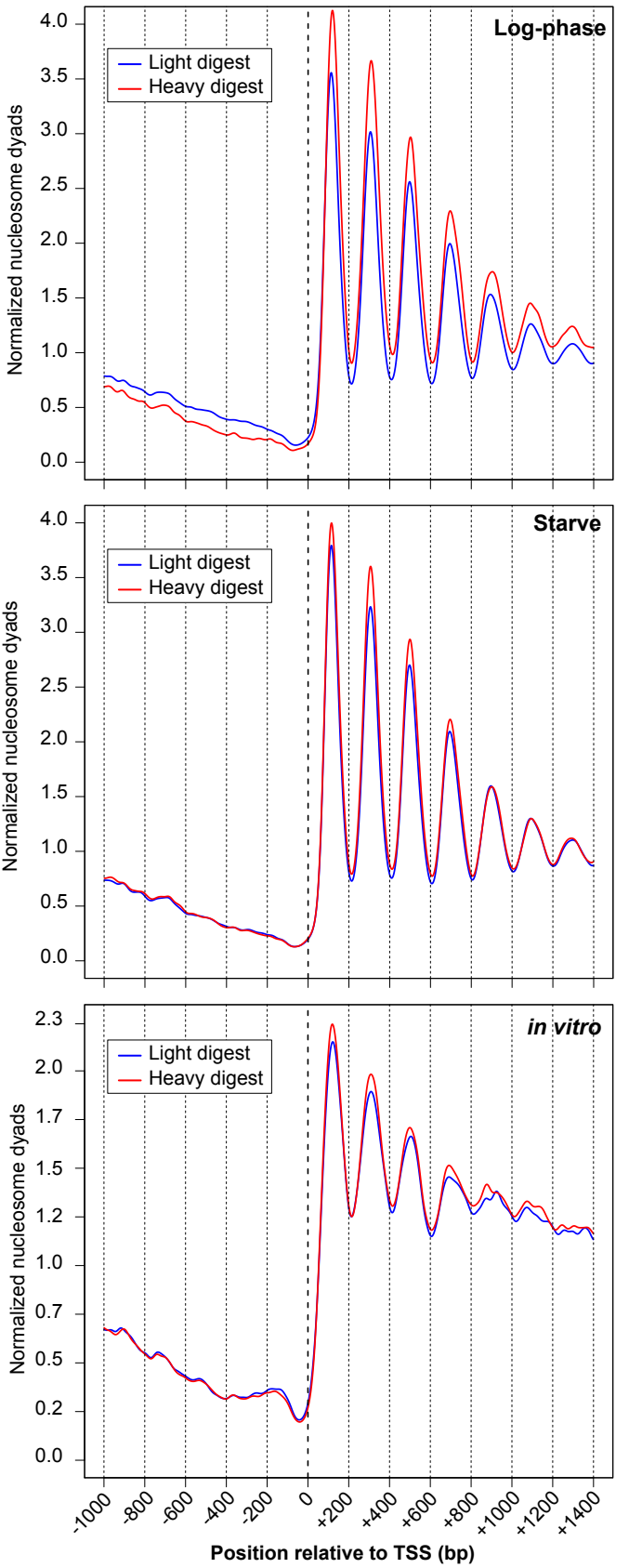


Figure S11. The stereotypical nucleosome pattern downstream of TSSs is robust to extended MNase digestion. Averaged nucleosome dyad counts around the TSS are plotted for (fixed) lightly digested and (native) heavily digested chromatin. The *in vitro* sample corresponds to reconstituted chromatin from experimental set B (see Methods).

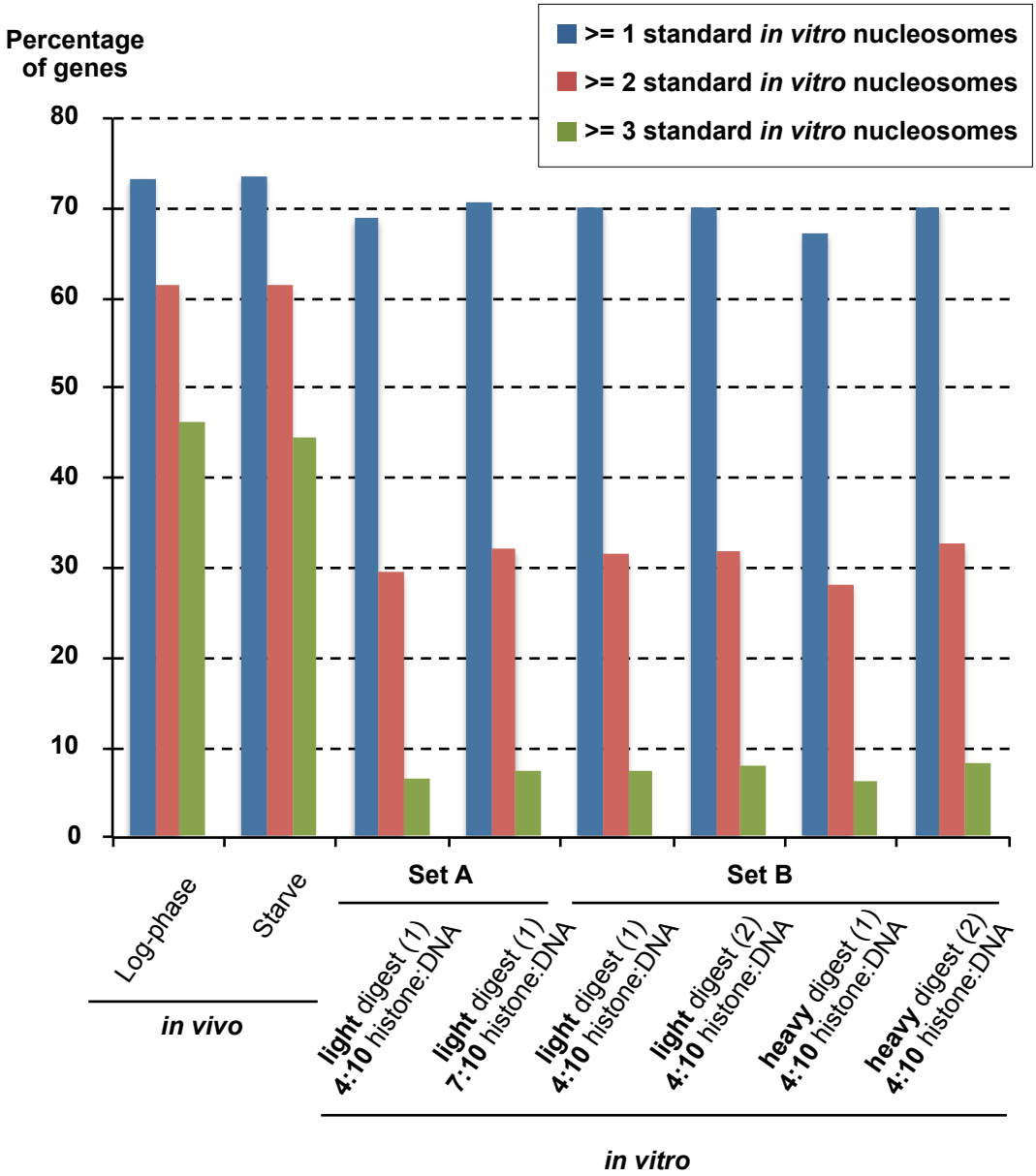


Figure S12. The distinctive *in vitro* nucleosome organization in *Tetrahymena* genes is robust to experimental variation. For each dataset, the number of genes with at least 1, 2, or 3 standard nucleosomes is counted, as described in Fig. 3. The consensus positions of the +1, +2, and +3 positions are respectively called from each dataset, by examining aggregate plots of nucleosome dyads around TSSs. Bracketed numbers denote individual replicates.

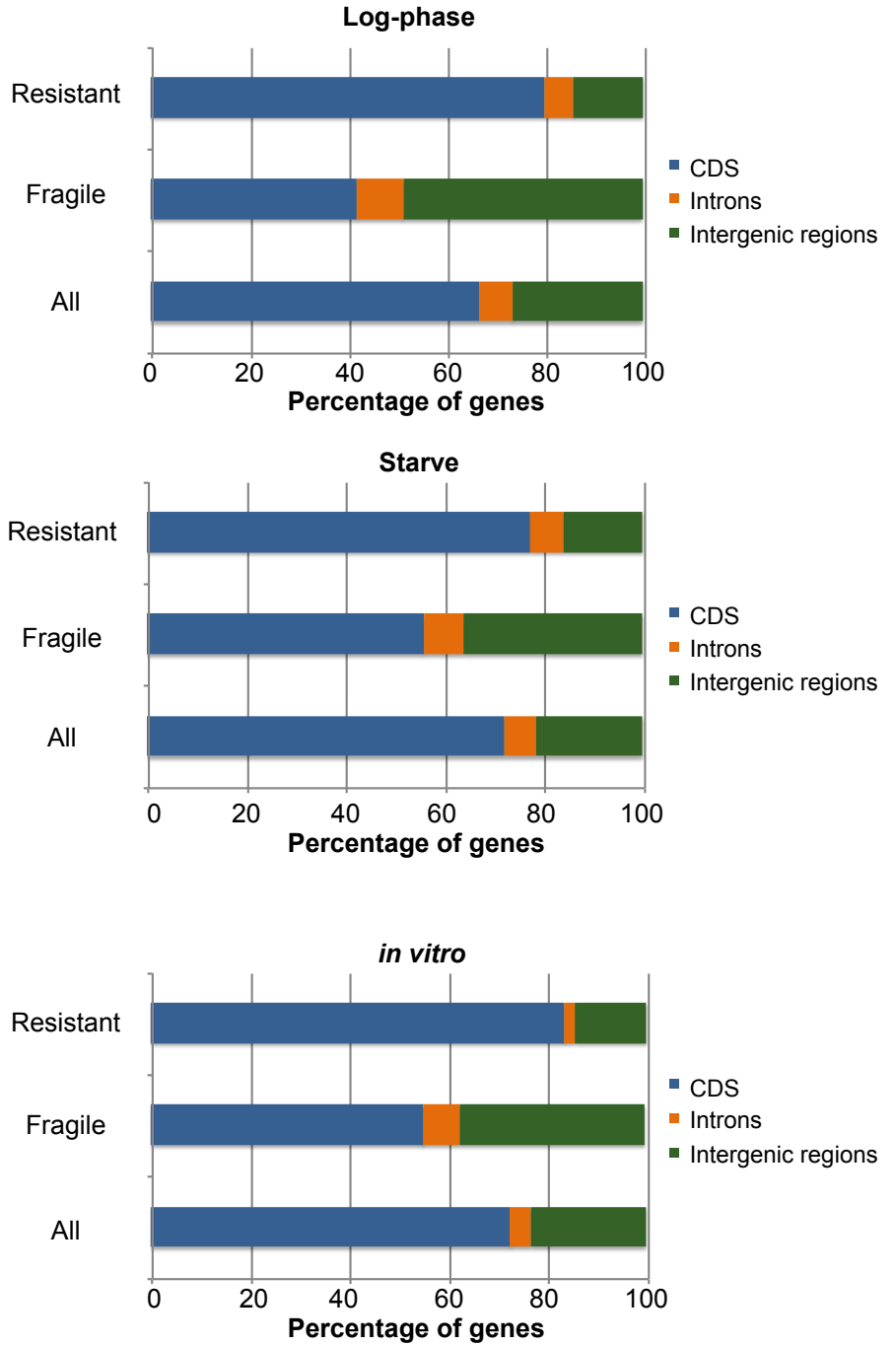


Figure S13. Coding regions are enriched in resistant nucleosomes and depleted in fragile nucleosomes. Resistant and fragile nucleosomes are annotated as nucleosomes that are either invariant or labile between the light and heavily digested chromatin samples (see Methods). These nucleosomes are then matched to respective genomic locations.

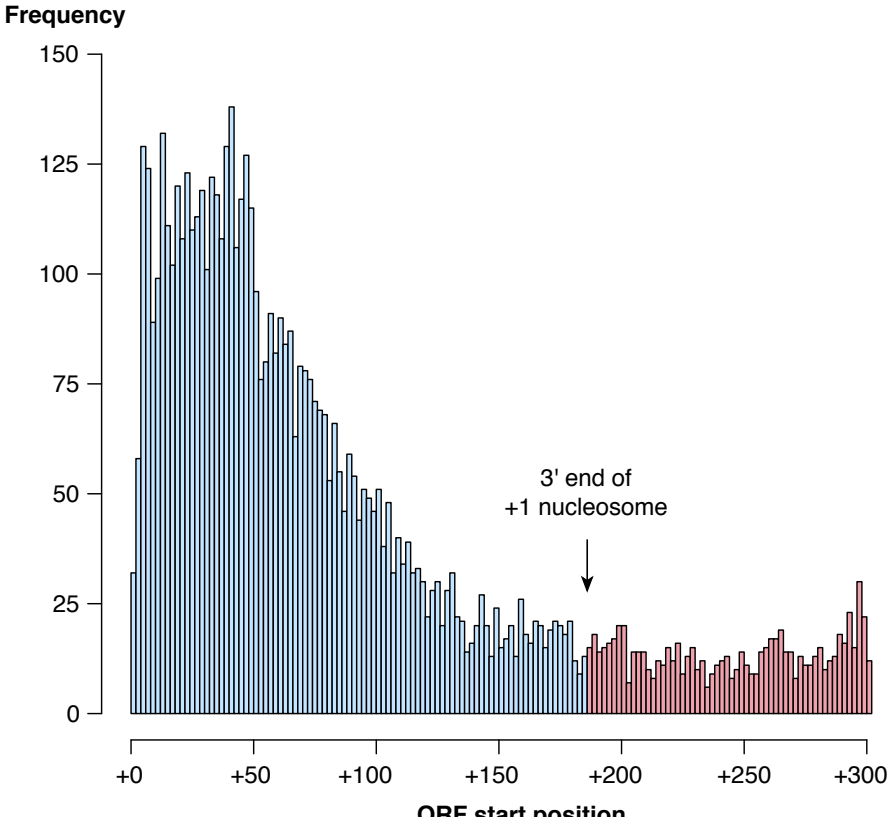


Figure S14. Distribution of *Tetrahymena* ORF start positions. Bars shaded in blue represent ORF start positions that lie upstream of the 3' end of the +1 nucleosome, while red bars represent ORF start positions downstream of it. Most 5' UTRs (given by the distance between the TSS and the ORF start position) are short, with a median length of 60 bp. Thus, most nucleosomes downstream of the TSS (such as those at the +1, +2, and +3 positions) lie within *Tetrahymena* open reading frames.

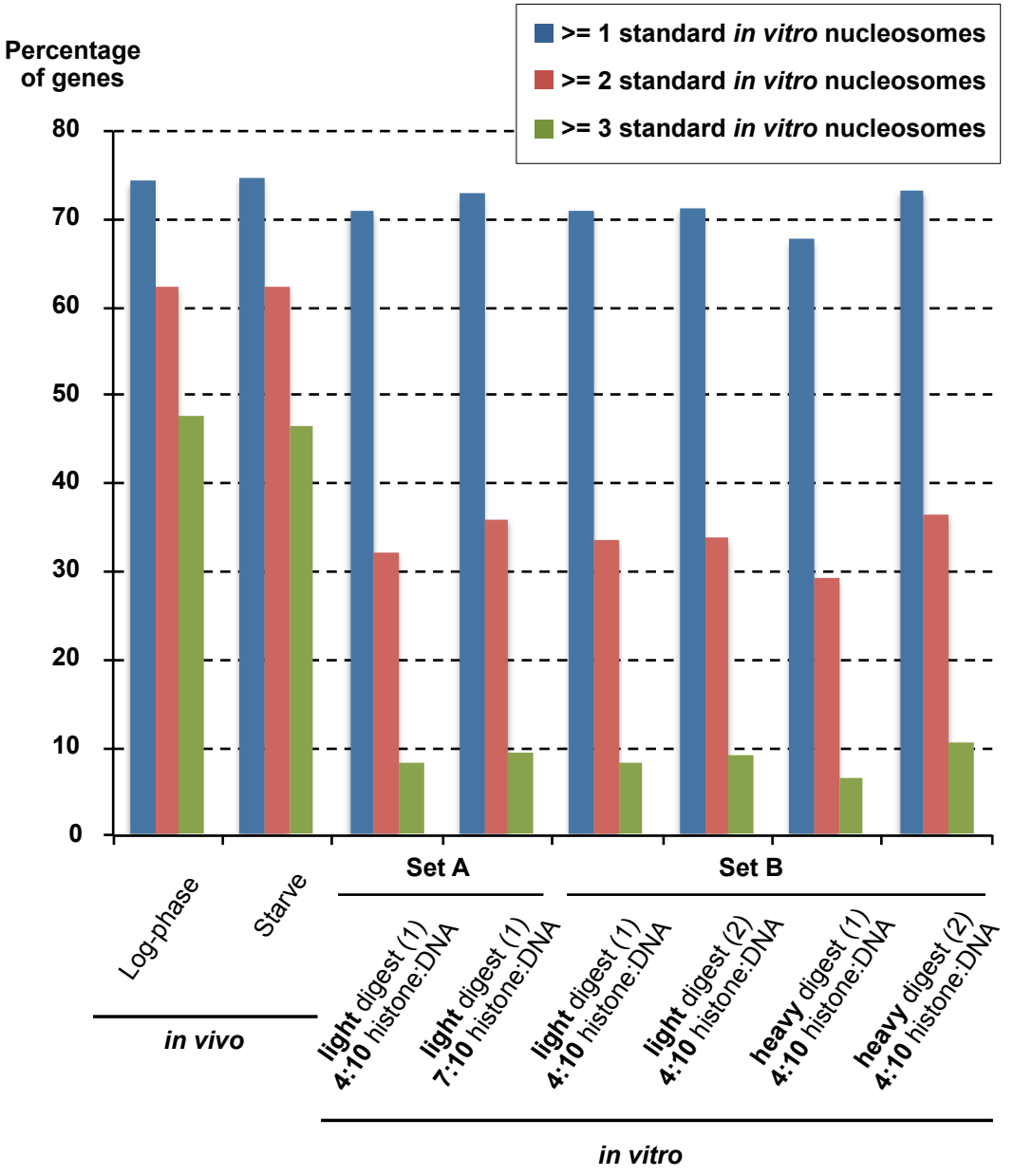


Figure S15. Nucleosome filtering criteria does not explain the observation of subsets of standard nucleosomes *in vitro*. Nucleosomes called from each MNase-seq dataset are normally filtered according to absolute and conditional positioning thresholds, to remove poorly positioned nucleosomes. Here, the filtering step is omitted, and the number of genes with at least 1, 2 or 3 standard nucleosomes is counted for each dataset. A similar number of genes with standard nucleosomes is observed for each dataset (as compared to Supplemental Fig. S12). This indicates that the nucleosome filtering criteria do not lead to the removal of standard nucleosomes from the dataset. Bracketed numbers denote individual replicates.

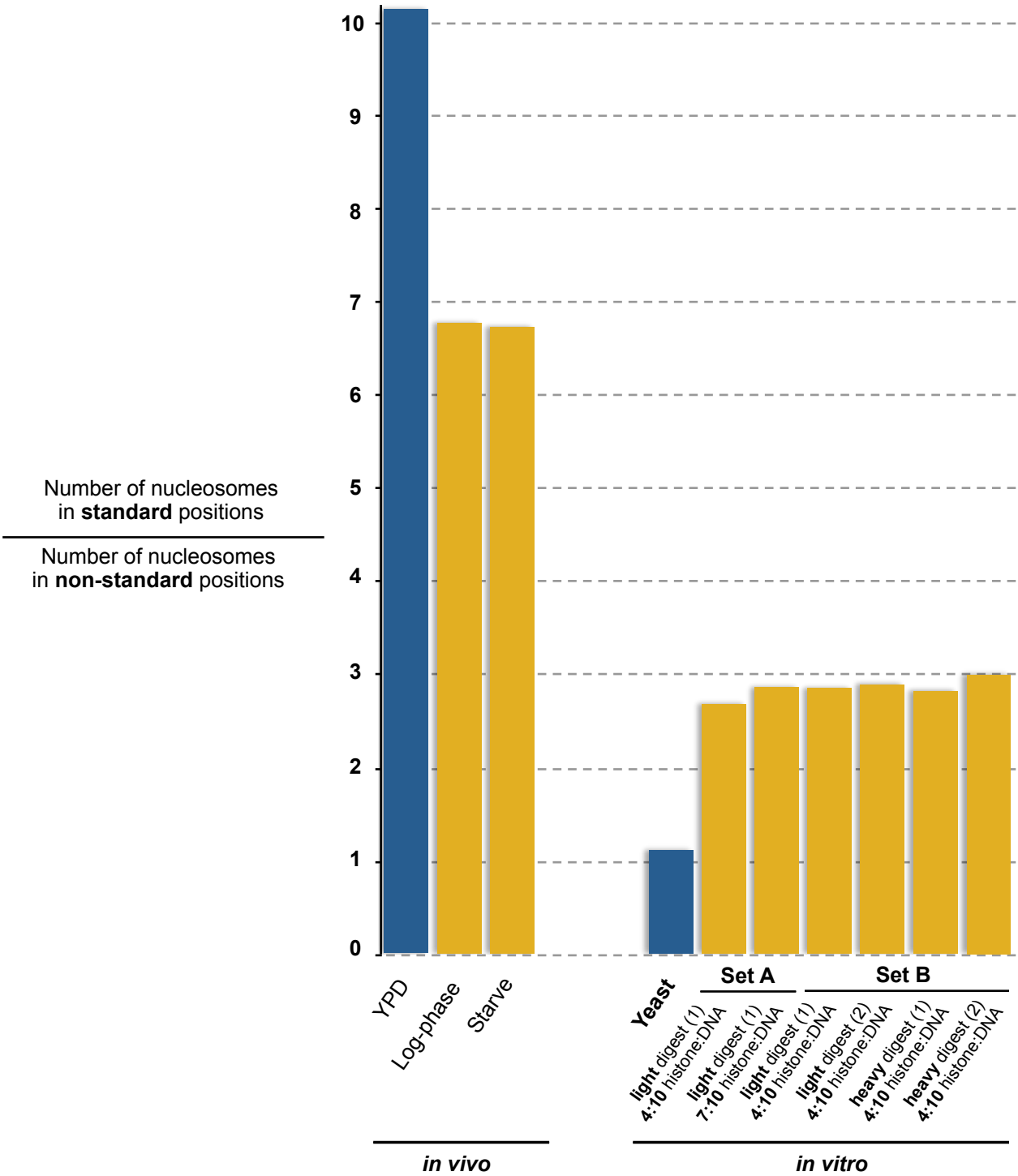


Figure S16. The high ratio of standard to non-standard nucleosomes within *in vitro Tetrahymena* chromatin is robust to variation in experimental conditions. Similar ratios were obtained for all individual *in vitro* experiments, despite changes in reaction volume, histone:DNA ratio, and the extent of MNase digestion. Bracketed numbers denote individual replicates. Standard and non-standard positions were defined as in Fig. 3.

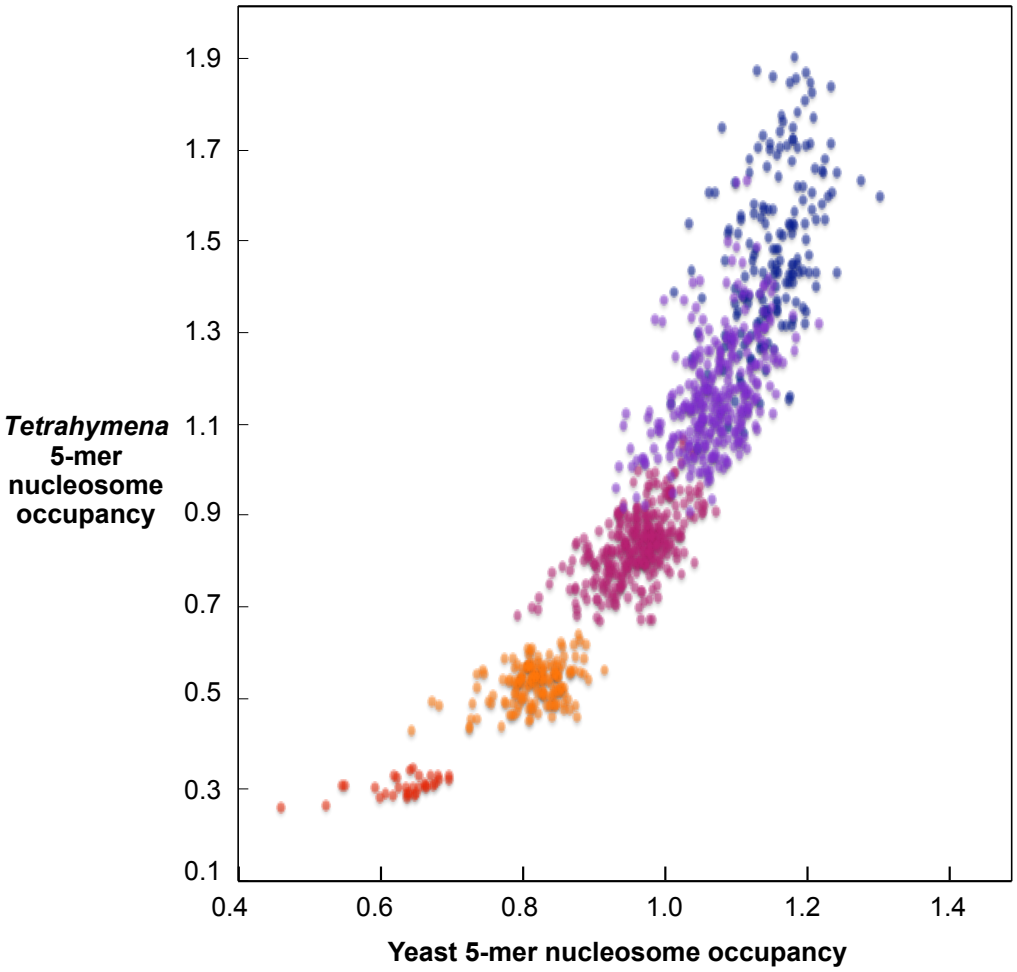


Figure S17. Comparison of normalized nucleosome occupancy of 5-mers in the yeast and *Tetrahymena* genomes. Occupancy data were calculated from the number of extended *in vitro* MNase-seq reads that span each unique 5-mer, normalized by the average 5-mer read count within each genome. This represents the relative intrinsic affinities of histone octamers for various unique DNA sequences. A strong correlation between *Tetrahymena* and yeast nucleosome occupancies is observed, indicating that histone octamers from both species share similar nucleosome sequence preferences. Colored data points progressing from dark blue to red denote increasing AT content.

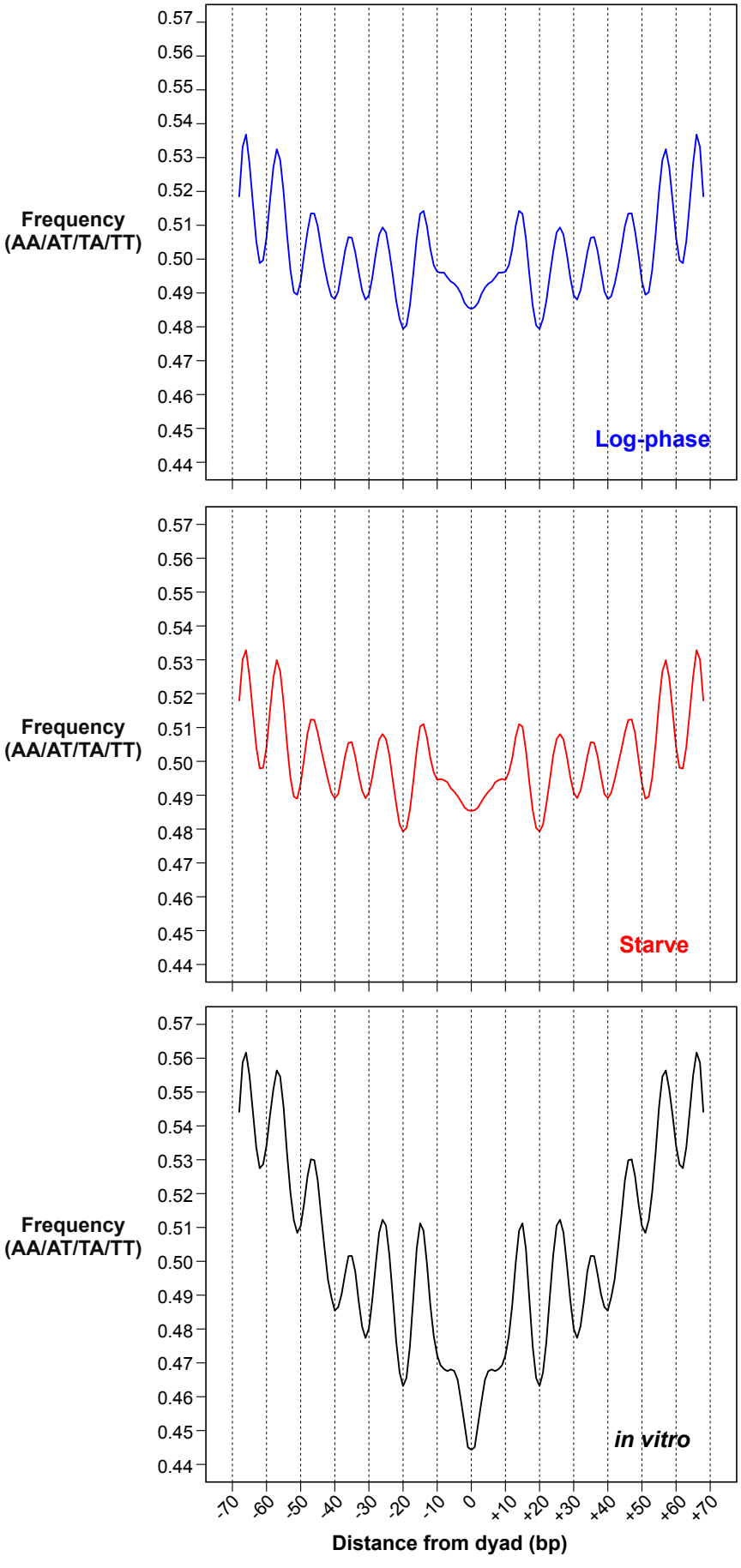


Figure S18. Rotational positioning of *Tetrahymena* nucleosomes. AA/TT/AT/TA dinucleotide frequencies were calculated as a 3bp sliding window average across nucleosomal DNA. A clear 10bp periodicity is observed, and is more distinct *in vitro* than *in vivo*. This is consistent with the larger role that nucleosome sequence preferences play in guiding nucleosome positions *in vitro*.

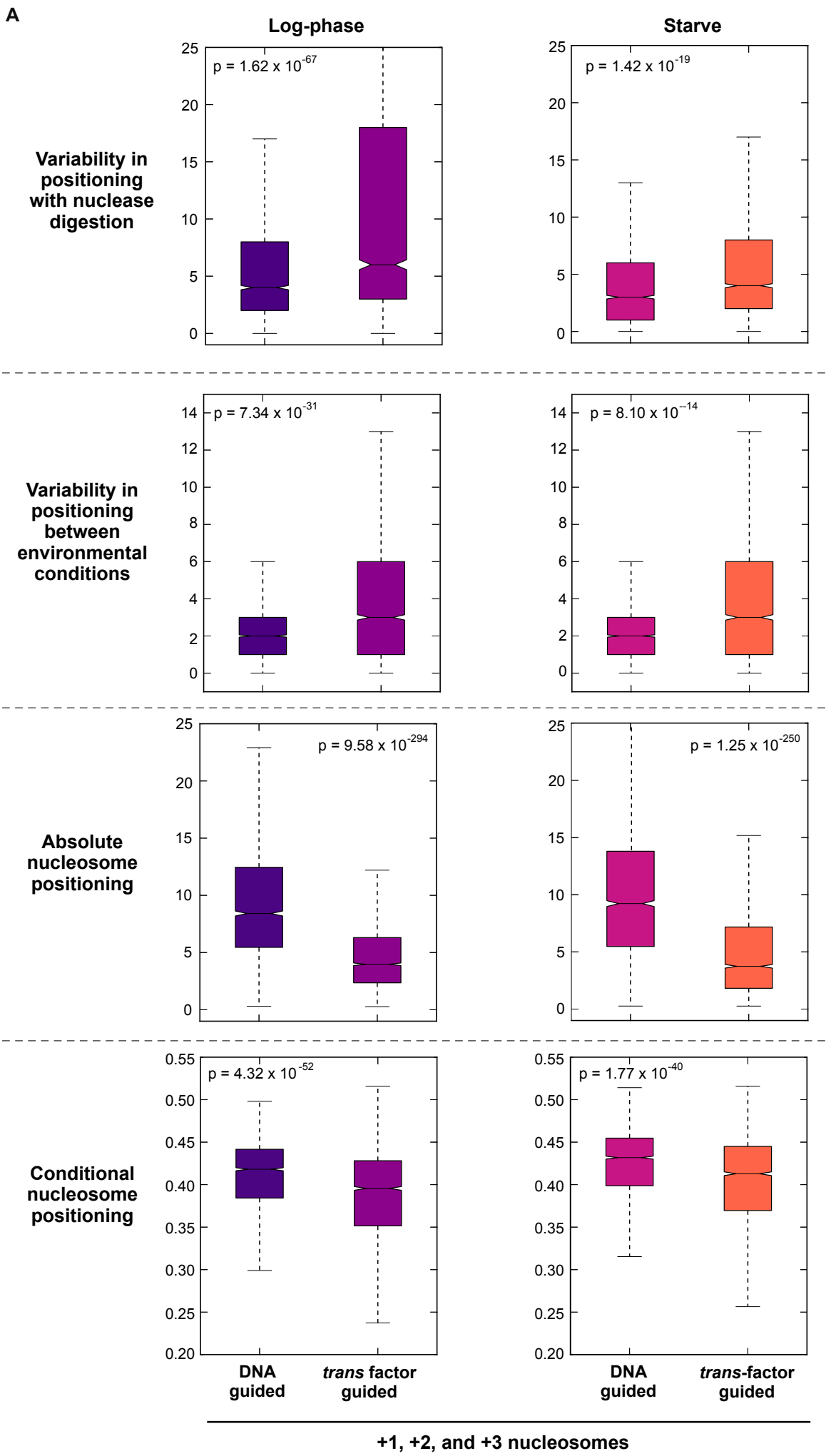
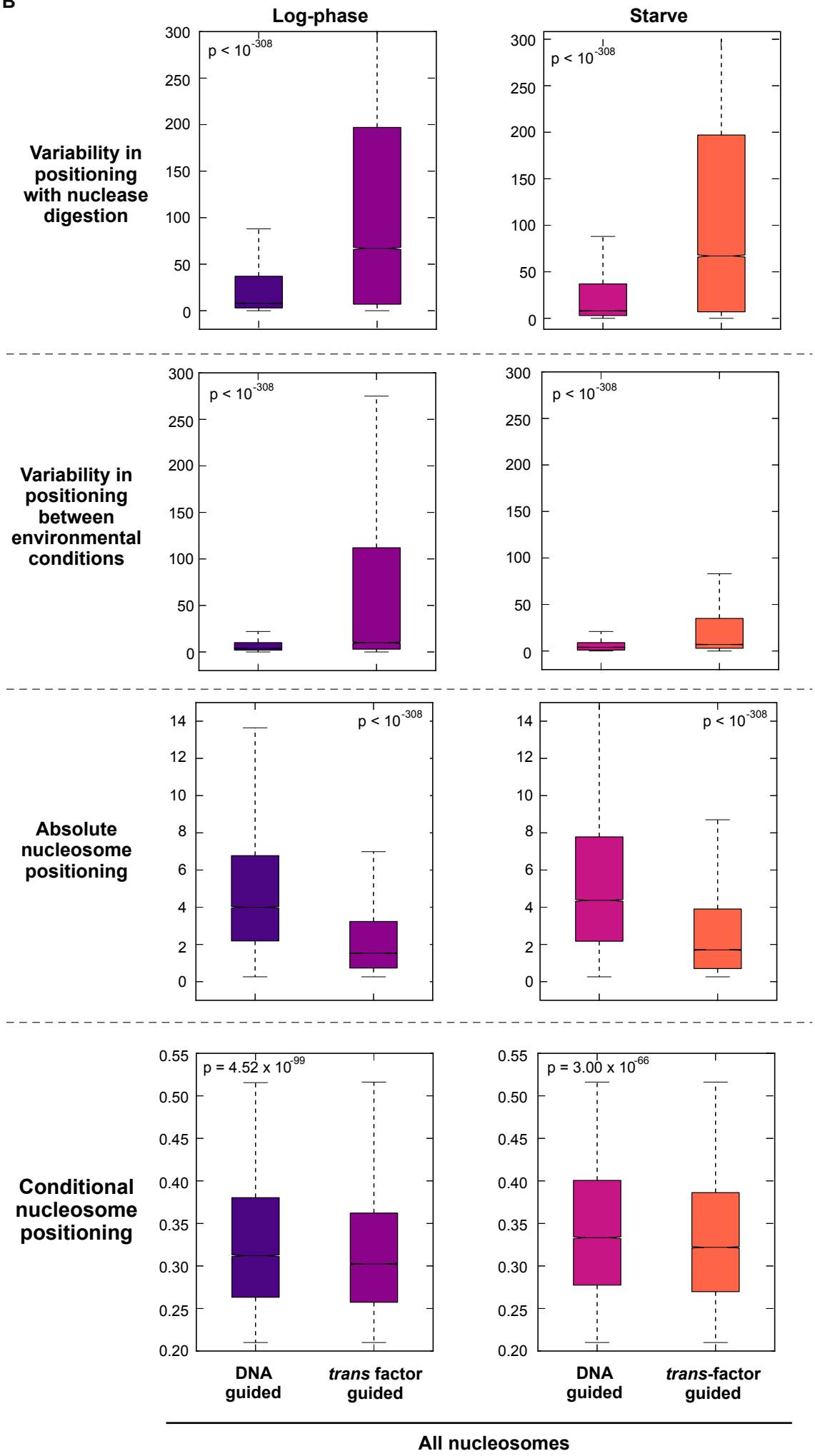
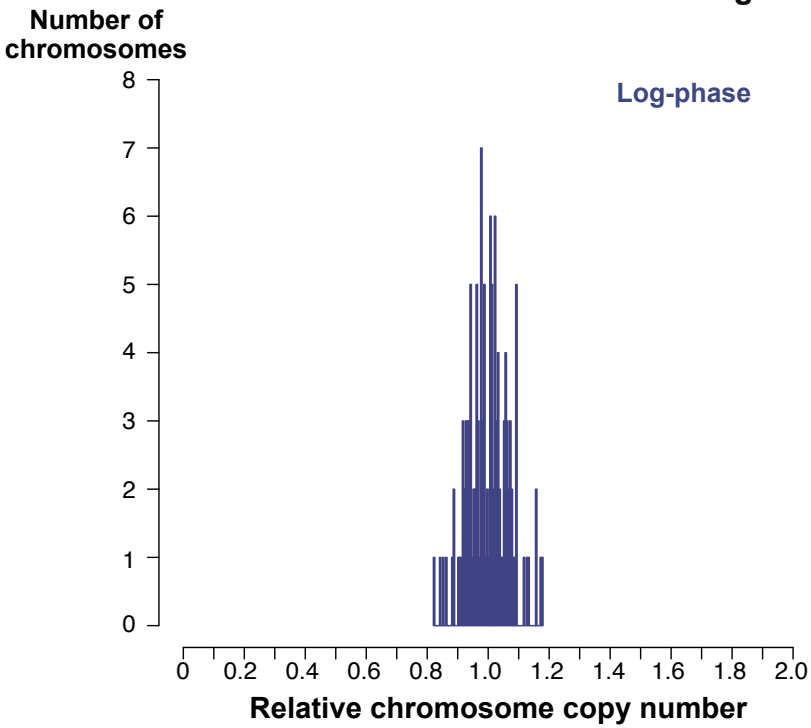


Figure S19. DNA-guided nucleosomes are more strongly positioned *in vivo*, and exhibit less variability in translational positions between different nutritional conditions and with extended MNase digestion. A nucleosome *in vivo* is classified as "DNA-guided" if it lies within 10 bp from a nucleosome *in vitro*. On the other hand, a nucleosome *in vivo* that lies greater than 73 bp from a nucleosome *in vitro* is classified as "*trans* factor-guided". For every *in vivo* nucleosome in a particular environmental condition (eg. log-phase), its distance to the nearest nucleosome in another environmental condition (eg. starve) is calculated. These distances are tabulated for all DNA-guided and *trans* factor-guided nucleosomes, respectively, and are denoted as the "variability in positioning between different environmental conditions". An analogous analysis is performed between lightly digested versus heavily digested chromatin, to obtain the "variability in positioning with nuclease digestion". See Methods for description of 'light' and 'heavy' chromatin digests. Absolute nucleosome positioning and conditional nucleosome positioning are calculated as described in Methods. (A) Analysis specifically of standard +1, +2, and +3 nucleosomes downstream of the TSS. (B) Analysis of all nucleosomes across the genome in log-phase and starve conditions, respectively.

B


Log-phase
Starve
Variability in positioning with nuclease digestion
Variability in positioning between environmental conditions
Absolute nucleosome positioning
Conditional nucleosome positioning
p < 10-308
p = 4.52 x 10-99
p = 3.00 x 10-66
DNA guided
trans factor guided
trans-factor guided
All nucleosomes



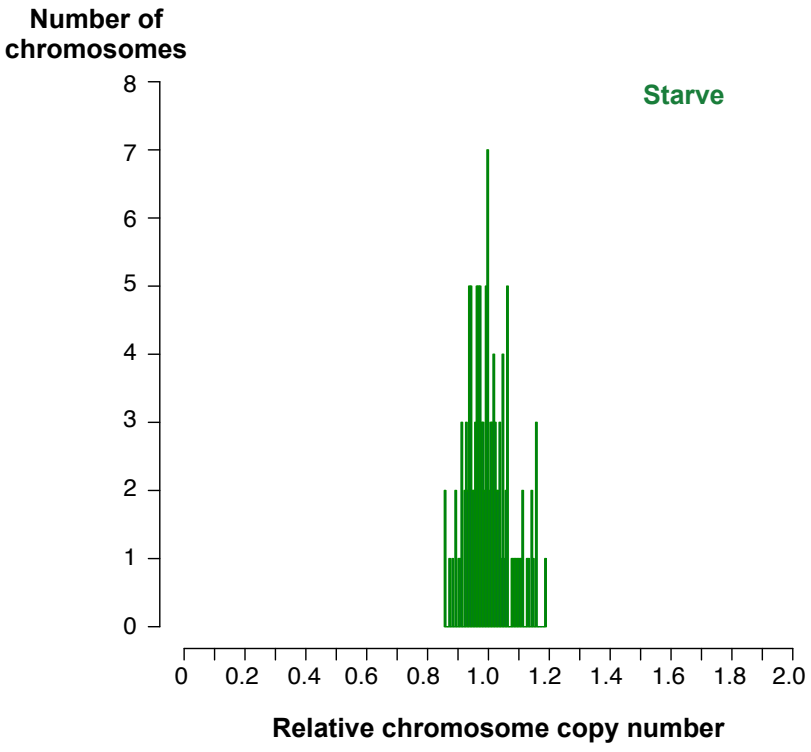


Figure S20. *Tetrahymena* chromosomes exhibit little variation in chromosome copy number. Sheared *Tetrahymena* genomic DNA from either log-phase or starved cells was sequenced, and the number of mapped reads per kilobase per million mapped reads was calculated for each chromosome. The resulting values were normalized by the genome-wide average, to obtain the relative chromosome copy number.

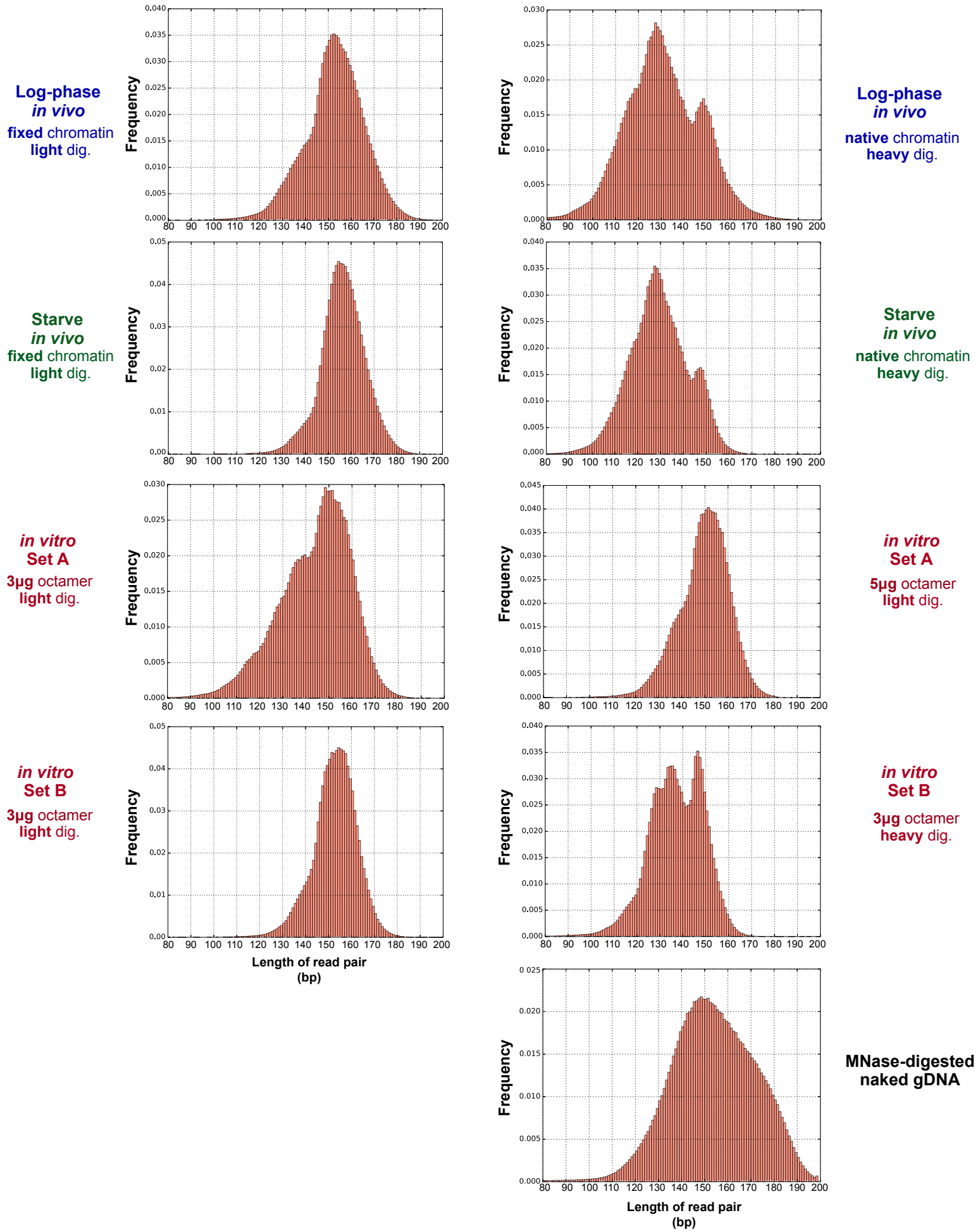


Figure S21. Distribution of fragment lengths for various MNase-seq datasets. Individual fragment lengths were calculated as the distance between the 5' ends of each read in a mapped read pair.

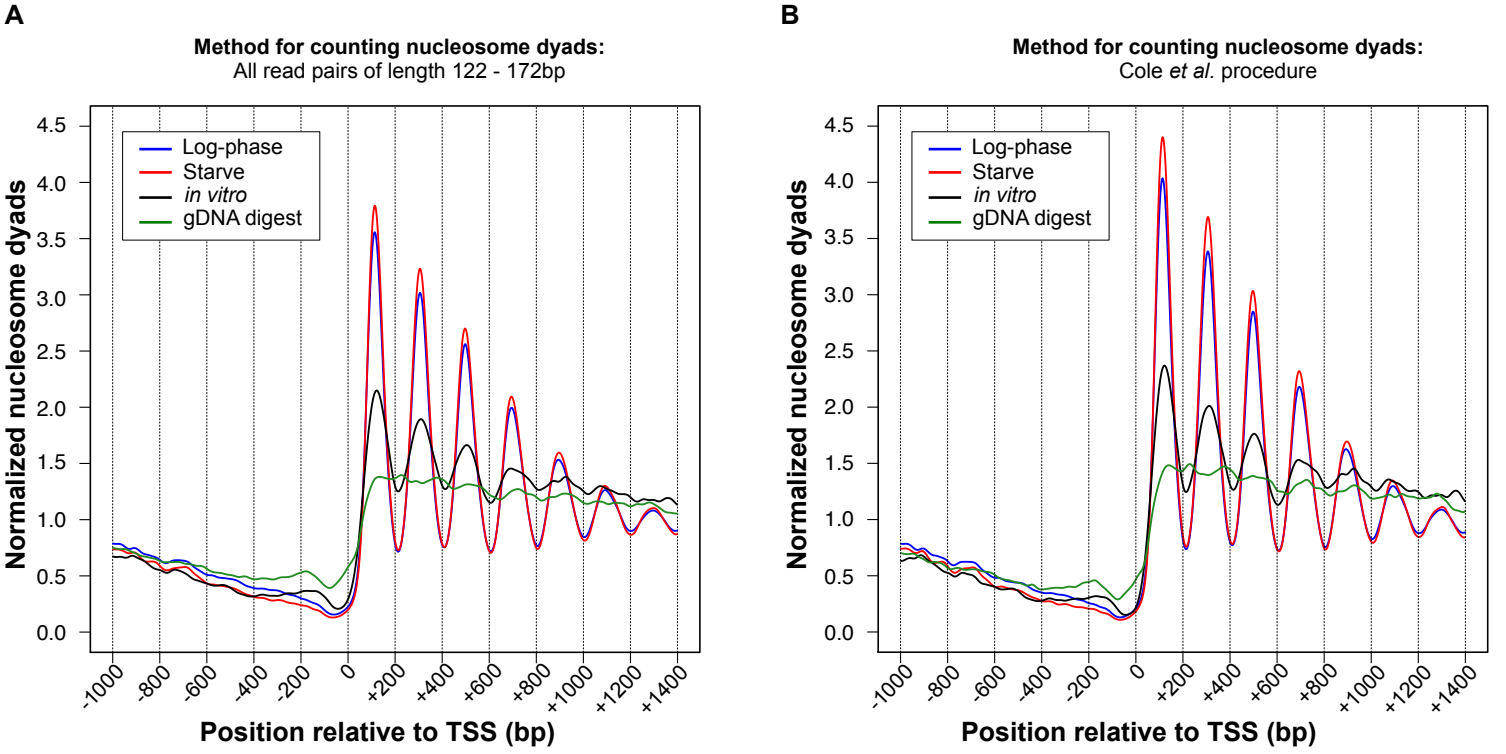


Figure S22. The distribution of nucleosome dyads around TSSs is consistent between different methods for calculating nucleosome positioning. (A) The mid-point of each 122 – 172 bp read pair is designated as a nucleosome dyad. (B) The two-step procedure by Cole et al. is used. Here, "high accuracy" nucleosome dyad positions are assigned from read pairs of length 144 – 150 bp. The mid-points of the remaining read pairs in the dataset are then assigned to these "high accuracy" positions.

Table S1

Sample information								Fragment length		
Sequencing Mode	**Number of mapped reads***	**Type of DNA**	**Source**	**Nutritional condition**	**Formaldehyde fixation**	**Extent of MNase digestion**	**Description**	****Mean**	**Median**	**Std Dev**
Paired-end	7581138	Mononucleosomal DNA	Macronucleus (*in vivo*)	Log-phase	Yes	Light	Biological replicate A	148.9	150.0	13.7
Paired-end	9417895	Mononucleosomal DNA	Macronucleus (*in vivo*)	Log-phase	Yes	Light	Biological replicate B	154.8	155.0	11.7
Paired-end	8375184	Mononucleosomal DNA	Macronucleus (*in vivo*)	Log-phase	No	Heavy	Biological replicate A	133.3	132.0	17.1
Paired-end	6435260	Mononucleosomal DNA	Macronucleus (*in vivo*)	Log-phase	No	Heavy	Biological replicate B	127.0	127.0	14.2
Paired-end	6133025	Mononucleosomal DNA	Macronucleus (*in vivo*)	Starve	Yes	Light	Biological replicate A	154.2	155.0	9.8
Paired-end	6177167	Mononucleosomal DNA	Macronucleus (*in vivo*)	Starve	Yes	Light	Biological replicate B	157.3	157.0	10.0
Paired-end	6866269	Mononucleosomal DNA	Macronucleus (*in vivo*)	Starve	No	Heavy	Biological replicate A	129.2	129.0	11.7
Paired-end	8424740	Mononucleosomal DNA	Macronucleus (*in vivo*)	Starve	No	Heavy	Biological replicate B	128.9	129.0	13.7
Paired-end	10017296	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Light	50ul in vitro reconstitution reaction (Set A), 4:10 histone:DNA ratio	143.5	146.0	15.4
Paired-end	14571585	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Light	50ul in vitro reconstitution reaction (Set A), 7:10 histone:DNA ratio	149.0	150.0	11.0
Paired-end	3679464	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Light	150ul in vitro reconstitution reaction (Set B), 4:10 histone:DNA ratio, Replicate 1	134.5	135.0	11.3
Paired-end	6492854	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Light	150ul in vitro reconstitution reaction (Set B), 4:10 histone:DNA ratio, Replicate 2	138.1	139.0	11.9
Paired-end	300531	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Heavy	150ul in vitro reconstitution reaction (Set B), 4:10 histone:DNA ratio, Replicate 1	150.2	151.0	9.9
Paired-end	4121508	Mononucleosomal DNA	Reconstituted chromatin (*in vitro*)	*in vitro*	No	Heavy	150ul in vitro reconstitution reaction (Set B), 4:10 histone:DNA ratio, Replicate 2	153.6	154.0	9.5
Paired-end	10329843	MNase-digested genomic DNA from starved cells	Naked genomic DNA	NA	No	Light	Naked gDNA, digested by MNase and size-selected for mononucleosome-sized DNA fragments	153.4	153.0	18.0

Table S1 (contd.)

Sample information								Fragment length		
Sequencing Mode	**Number of mapped reads***	**Type of DNA**	**Source**	**Nutritional condition**	**Formaldehyde fixation**	**Extent of MNase digestion**	**Description**	****Mean**	**Median**	**Std Dev**
Single-end	17887355	Covaris-sheared genomic DNA from log-phase cells	Naked genomic DNA	NA	No	NA	*Tetrahymena* genomic DNA, sheared in a sonicator	112.0		
Single-end	7522145	Covaris-sheared genomic DNA from starved cells	Naked genomic DNA	NA	No	NA	*Tetrahymena* genomic DNA, sheared in a sonicator	119.0		

* For paired-end datasets, the number of properly mapped read pairs is listed.

* For single-reads datasets, the number of mapped individual reads is listed.

**Mean fragment length values were directly calculated for paired-end datasets, while they were estimated using cross-correlation analysis for single-read datasets

Table S1. Summary of Illumina datasets generated in this study. Fragment lengths for paired-end datasets were calculated directly, as the distance between the 5' ends of the reads in a mapped read pair. For single-end datasets, the average fragment length was estimated using cross-correlation analysis (Kharchenko et al. 2008). Reads counts denote the total number of reads mapped to two-telomere (complete) chromosomes in the *Tetrahymena* SB210 genome assembly.

Table S2

Standard nucleosome pattern				*Tetrahymena in vitro*						Yeast *in vitro*
+1	+2	+3	Experimental Set	A	A	B	B	B	B	
			Histone:DNA ratio	4:10	7:10	4:10	4:10	4:10	4:10	4:10
			MNase digestion	Light (1)	Light (1)	Light (1)	Light (2)	Heavy (1)	Heavy (2)	
■	■	■	% genes with specific pattern	11.6	10.9	10.5	10.4	11.2	11.5	5.9
■	■	■		11.1	11.0	10.9	11.3	12.9	11.9	8.6
■	■	■		7.1	7.3	7.7	7.9	7.7	6.2	5.4
■	■	■		10.7	11.0	11.8	11.4	11.2	11.6	9.0
■	■	■		6.5	6.6	6.1	6.0	6.2	6.2	4.0
■	■	■		7.1	7.4	7.4	7.8	7.2	7.4	7.3
■	■	■		7.1	8.2	8.0	8.6	6.7	9.0	6.9

Table S2. Frequency of genes with standard nucleosome patterns. Each nucleosome pattern is depicted by a row of 3 shaded boxes on the left of the table. A green box denotes the presence of a standard nucleosome, while a grey box denotes its absence. The corresponding frequency of the pattern is listed for each dataset, on the right of the table. A distribution of frequencies is observed for each *Tetrahymena in vitro* dataset, with no clearly dominant pattern. This indicates that multiple nucleosome patterns – each mainly exhibiting subsets of standard nucleosomes – explain the regular aggregate pattern *in vitro*. All of these individual patterns occur with higher frequency in *Tetrahymena* than yeast.

Table S3

	Codon position 1	
	Number of Gs and Cs	Number of As and Ts
Within DNA sequence guided nucleosome	103934	172616
Within *trans*-factor guided nucleosome	54536	101145

	Codon position 2	
	Number of Gs and Cs	Number of As and Ts
Within DNA sequence guided nucleosome	83960	192590
Within *trans*-factor guided nucleosome	41558	114123

	Codon position 3	
	Number of Gs and Cs	Number of As and Ts
Within DNA sequence guided nucleosome	73281	203269
Within *trans*-factor guided nucleosome	37942	117739

Table S3. Codons within DNA-guided nucleosomes exhibit higher GC content than those within *trans* factor-guided nucleosomes. Both types of nucleosomes are defined in Supplemental Fig. S19. Codons that lie no greater than 73 bp from a called nucleosome are considered as lying within the corresponding DNA-guided or *trans* factor-guided nucleosome.

Table S4

			Codon Frequency			
Amino Acid	**Codon**	**Codon GC content**	**Within DNA-guided nucleosomes, relative to within *trans* factor-guided nucleosomes**	**Within DNA-guided nucleosomes**	**Within trans factor-guided nucleosomes**	**Within all regions**
A	GCG	1.000	0.361	0.019	0.052	0.019
A	GCT	0.667	0.857	0.537	0.626	0.537
A	GCA	0.667	1.309	0.348	0.266	0.340
A	GCC	1.000	1.705	0.097	0.057	0.104
C	TGT	0.333	0.672	0.465	0.693	0.535
C	TGC	0.667	1.742	0.535	0.307	0.465
D	GAT	0.333	0.923	0.801	0.868	0.805
D	GAC	0.667	1.510	0.199	0.132	0.195
E	GAA	0.333	0.951	0.801	0.842	0.811
E	GAG	0.667	1.261	0.199	0.158	0.189
F	TTC	0.333	0.942	0.313	0.332	0.311
F	TTT	0.000	1.029	0.687	0.668	0.689
G	GGA	0.667	0.737	0.434	0.588	0.438
G	GGT	0.667	1.082	0.391	0.361	0.398
G	GGC	1.000	3.255	0.117	0.036	0.112
G	GGG	1.000	3.997	0.059	0.015	0.053
H	CAT	0.333	0.929	0.667	0.718	0.695
H	CAC	0.667	1.181	0.333	0.282	0.305
I	ATA	0.000	0.725	0.333	0.460	0.346
I	ATT	0.000	1.224	0.509	0.416	0.509
I	ATC	0.333	1.268	0.158	0.124	0.146
K	AAA	0.000	0.957	0.696	0.727	0.707
K	AAG	0.333	1.114	0.304	0.273	0.293
L	CTC	0.667	0.835	0.095	0.114	0.072
L	TTA	0.000	0.864	0.388	0.449	0.414
L	CTT	0.333	0.878	0.212	0.241	0.217
L	CTA	0.333	1.396	0.100	0.072	0.107
L	TTG	0.333	1.600	0.174	0.108	0.159
L	CTG	0.667	1.991	0.031	0.016	0.031

Table S4. Biases in synonymous codon usage are encoded within distinct nucleosomal regions. DNA-guided and *trans*-factor guided nucleosomes are defined as in Supplemental Fig. S19. We examined nucleosomes from the log-phase dataset (light MNase digest), and from the set "B" *in vitro* experiment (performed with 4:10 histone:DNA, light MNase digest). Codons were considered as lying within a nucleosome, according to criteria described in Supplemental Table S3. Each group of synonymous codons was analyzed separately. The most GC-rich codon(s) within each group of synonymous codon is highlighted in red. For each codon, we calculated the ratio of its frequency in sequences that lie within DNA-guided nucleosomes, to its corresponding frequency within sequences that lie within *trans* factor-guided nucleosomes. Codons enriched in DNA-guided nucleosomes have the corresponding value > 1, and are highlighted in red. It quantifies the impact of accommodating DNA-guided nucleosomes on synonymous codon usage. The underlying codon usage for 13 out of 18 amino acids was biased towards GC-rich codons within coding regions that overlap with DNA-guided

Table S4 (contd.)

			Codon Frequency			
Amino Acid	**Codon**	**Codon GC content**	**Within DNA-guided nucleosomes, relative to within *trans* factor-guided nucleosomes**	**Within DNA-guided nucleosomes**	**Within trans factor-guided nucleosomes**	**Within all regions**
N	AAT	0.000	0.917	0.722	0.787	0.739
N	AAC	0.333	1.308	0.278	0.213	0.261
P	CCC	1.000	0.520	0.096	0.184	0.090
P	CCA	0.667	0.738	0.320	0.433	0.335
P	CCG	1.000	1.014	0.012	0.012	0.015
P	CCT	0.667	1.543	0.572	0.371	0.560
Q	CAA	0.333	0.894	0.234	0.261	0.225
Q	TAA	0.000	0.950	0.522	0.550	0.556
Q	TAG	0.333	1.120	0.192	0.172	0.172
Q	CAG	0.667	3.055	0.052	0.017	0.047
R	AGA	0.333	0.902	0.719	0.797	0.729
R	CGT	0.667	0.982	0.093	0.094	0.108
R	CGA	0.667	0.985	0.015	0.016	0.017
R	CGC	1.000	1.798	0.036	0.020	0.036
R	AGG	0.667	1.864	0.135	0.072	0.108
R	CGG	1.000	4.550	0.003	0.001	0.002
S	TCG	0.667	0.452	0.022	0.049	0.024
S	AGT	0.333	0.808	0.194	0.240	0.210
S	TCA	0.333	0.992	0.267	0.269	0.250
S	TCT	0.333	1.075	0.319	0.297	0.313
S	AGC	0.667	1.362	0.146	0.107	0.149
S	TCC	0.667	1.363	0.053	0.039	0.054
T	ACA	0.333	0.745	0.354	0.475	0.376
T	ACT	0.333	1.120	0.488	0.436	0.496
T	ACC	0.667	1.711	0.136	0.079	0.105
T	ACG	0.667	2.320	0.022	0.010	0.022
V	GTG	0.667	0.843	0.087	0.103	0.084
V	GTT	0.333	0.909	0.500	0.550	0.495
V	GTA	0.333	1.130	0.281	0.249	0.297
V	GTC	0.667	1.348	0.132	0.098	0.124
Y	TAT	0.000	0.845	0.667	0.789	0.695
Y	TAC	0.333	1.579	0.333	0.211	0.305
M	ATG	0.333	1.000	1.000	1.000	1.000
W	TGG	0.667	1.000	1.000	1.000	1.000

Table S5

		Amino Acid Frequency			
Amino acid	**Weighted codon GC content**	**Within DNA-guided nucleosomes, relative to within *trans* factor-guided nucleosomes**	**Within DNA-guided nucleosomes**	**Within *trans* factor-guided nucleosomes**	**Within all regions**
I	0.049	0.845	0.072	0.085	0.079
N	0.087	0.868	0.083	0.096	0.086
K	0.098	0.899	0.086	0.096	0.091
Y	0.102	1.048	0.039	0.038	0.041
F	0.104	0.759	0.049	0.065	0.049
Q	0.164	0.818	0.083	0.101	0.092
L	0.230	0.955	0.089	0.093	0.088
M	0.333	1.413	0.022	0.016	0.019
T	0.376	1.270	0.044	0.035	0.044
E	0.396	0.978	0.064	0.065	0.068
D	0.398	0.987	0.054	0.055	0.051
V	0.403	1.268	0.045	0.035	0.043
S	0.409	1.033	0.080	0.078	0.079
H	0.435	1.346	0.017	0.013	0.016
R	0.437	1.711	0.030	0.017	0.030
C	0.488	1.046	0.015	0.015	0.014
W	0.667	2.003	0.008	0.004	0.006
P	0.702	1.188	0.033	0.028	0.029
A	0.708	1.543	0.044	0.028	0.037
G	0.721	1.113	0.043	0.038	0.037

Table S5. Biases in amino acid composition are encoded within distinct nucleosomal regions. DNA-guided and trans-factor guided nucleosomes are defined as in Supplemental Fig. S19. We examined nucleosomes from the log-phase dataset (light MNase digest), and from the set “B” *in vitro* experiment (performed with 4:10 histone:DNA, light MNase digest). Amino acids whose corresponding codons lie no greater than 73 bp from a called nucleosome dyad are considered as lying within the nucleosome. Weighted codon GC content values were calculated as the sum of GC contents of synonymous codons specifying an amino acid, respectively multiplied by their respective codon frequencies. Amino acids were ranked according to their weighted codon GC content, as shaded from low (blue) to high (red). For each amino acid, we calculated the total frequency of its codons that lie within 73 bp of DNA-guided nucleosome dyads, divided by the total frequency of those that lie within 73 bp of trans factor-guided nucleosome dyads. Amino acids whose codons are enriched in DNA-guided nucleosomes have the corresponding value > 1. Amino acids specified by GC-rich codons tend to be enriched in DNA-guided nucleosomes.