FUZZINESS AND NOISE IN NUCLEOSOMAL ARCHITECTURE

Oscar Flores, Özgen Deniz, Montserrat Soler-López and Modesto Orozco

SUPPLEMENTARY TABLES

default	**score_w**	**0.6**		**score_h**	**0.4**		
	FcF	FoF	WcW	WoW	M-F	M--W	+1_missing
R1 (2x)	226	89	1808	1674	106	605	42
R2 (2x)	183	59	1942	1606	86	464	20
As (2x)	277	167	1624	1419	171	658	34
Ov (2x)	204	205	1405	1285	314	840	56
Un (2x)	478	358	763	623	211	361	15
var1	**score_w**	**0.5**		**score_h**	**0.5**		
	FcF	FoF	WcW	WoW	M-F	M--W	+1_missing
R1 (2x)	149	63	1785	1618	103	608	42
R2 (2x)	109	46	1916	1574	74	476	20
As (2x)	136	113	1722	1476	133	696	34
Ov (2x)	211	210	1261	1147	345	809	56
Un (2x)	360	273	853	677	174	398	15
var2	**score_w**	**0.4**		**score_h**	**0.6**		
	FcF	FoF	WcW	WoW	M-F	M--W	+1_missing
R1 (2x)	287	152	1433	1362	149	562	42
R2 (2x)	187	79	1643	1415	104	446	20
As (2x)	238	185	1410	1259	194	635	34
Ov (2x)	341	354	1016	935	424	730	56
Un (2x)	431	335	726	596	198	374	15

Supplementary Table S1. Nucleosomes call evaluation based on nucleR TSS clustering by applying different thresholds. The number of genes clustered in the main nucleosome pattern annotations are displayed according to the default parameters score_w = 0.6 and score_h=0.4; score_w=0.5 and score_h=0.5 (*var1*); and score_w=0.4 and score_h=0.6 (*var2*). Key: **R1 (2x)** – replica 1 paired-end dataset; **R2** – replica 2; **As** – asynchronized; **Ov** – MNase-overdigested; **Un** – underdigested.

	-1 Nucleosome			NFR		+1 Nucleosome	
	M	**F**	**W**	**Closed**	**Open**	**F**	**W**
R1 (1x)	648 9.68%	2544 38.02%	2992 44.71%	2858 42.71%	2207 32.98%	2405 35.94%	3779 56.47%
R2 (1x)	239 3.57%	2094 31.29%	395 59.06%	3726 55.68%	1860 27.79%	1390 20.77%	4895 73.15%
R1 (2x)	711 10.62%	1508 22.53%	4021 60.09%	2939 43.92%	2300 34.37%	1029 15.38%	5211 77.87%
R2 (2x)	550 8.22%	1070 15.99%	3852 57.56%	2742 40.97%	2040 30.48%	684 10.22%	4788 71.55%
As. (2x)	829 12.39%	1733 25.9%	3656 54.63%	2794 41.75%	2225 33.25%	1302 19.46%	4916 73.46%
Ov. (2x)	1154 17.24%	1886 28.18%	3235 48.34%	2559 38.24%	2328 34.79%	1321 19.74%	4954 74.03%
Un. (2x)	572 8.55%	2530 37.81%	2051 30.65%	2302 34.4%	1752 26.18%	1846 27.59%	3307 49.42%

Supplementary Table S2. Distribution of -1/+1 nucleosomes and NFRs classification for each sample. -1 nucleosomes are classified as missing (M), fuzzy (F) or well-positioned (W) and +1 nucleosomes are either F or W. NFRs can have open or closed configuration depending on the NFR width. Key: **1x** – Single End, **2x** – Paired End, **R1** – Replica 1, **R2** – Replica2, **As** – Asynchronous, **Ov** – Overdigested, **Un** – Underdigested.

	Same classification					Variable classification				
vs.	**Coverage**	**Cluster**	**-1 Nuc.**	**+1 Nuc.**	**NFR**	**Coverage**	**Cluster**	**-1 Nuc.**	**+1 Nuc.**	**NFR**
R1 (1x) R1 (2x)	6311 (94.31%)	2552 (38.14%)	4109 (61.4%)	4261 (63.67%)	4599 (68.72%)	23 (0.34%)	1513 (22.61%)	489 (7.31%)	448 (6.69%)	390 (5.83%)
R2 (1x) R2 (2x)	5969 (89.2%)	2399 (35.85%)	3480 (52%)	4301 (64.27%)	3751 (56.05%)	63 (0.94%)	1514 (22.62%)	791 (11.82%)	335 (5.01%)	591 (8.83%)
R1 (1x) R2 (1x)	5418 (80.96%)	2014 (30.1%)	3606 (53.89%)	4205 (62.84%)	4091 (61.13%)	222 (3.32%)	1994 (29.8%)	1208 (18.05%)	716 (10.7%)	710 (10.61%)
R1 (2x) R2 (2x)	5980 (89.36%)	3181 (47.53%)	4068 (60.79%)	4529 (67.68%)	4260 (63.66%)	93 (1.39%)	992 (14.82%)	433 (6.47%)	470 (7.02%)	231 (3.45%)
R1 (2x) As (2x)	5123 (76.55%)	2964 (44.29%)	4131 (61.73%)	4758 (71.1%)	4482 (66.98%)	228 (3.41%)	1563 (23.36%)	858 (12.82%)	819 (12.24%)	322 (4.81%)
R2 (2x) As (2x)	5781 (86.39%)	2927 (43.74%)	3874 (57.89%)	4393 (65.65%)	4157 (62.12%)	120 (1.79%)	1144 (17.1%)	616 (9.21%)	611 (9.13%)	212 (3.17%)
R1 (2x) Ov (2x)	5616 (83.92%)	2881 (43.05%)	3966 (59.26%)	4808 (71.85%)	4452 (66.53%)	91 (1.36%)	1649 (24.64%)	840 (12.55%)	857 (12.81%)	322 (4.81%)
R2 (2x) Ov (2x)	5874 (87.78%)	2754 (41.15%)	3612 (53.97%)	4443 (66.39%)	4155 (62.09%)	106 (1.58%)	1181 (17.65%)	680 (10.16%)	670 (10.01%)	217 (3.24%)
R1 (2x) Un (2x)	4177 (62.42%)	1377 (20.58%)	2597 (38.81%)	3426 (51.2%)	2917 (43.59%)	405 (6.05%)	2111 (31.55%)	1179 (17.62%)	1090 (16.29%)	705 (10.53%)
R2 (2x) Un (2x)	4971 (74.28%)	1355 (20.25%)	2376 (35.51%)	3202 (47.85%)	2767 (41.35%)	221 (3.3%)	1820 (27.2%)	1184 (17.69%)	1047 (15.65%)	586 (8.76%)
Ov (2x) Un (2x)	6342 (94.77%)	1915 (28.62%)	3105 (46.4%)	3737 (55.84%)	3279 (49%)	19 (0.28%)	1661 (24.82%)	721 (10.77%)	731 (10.92%)	547 (8.17%)
As (2x) Ov (2x)	6421 (95.95%)	3315 (49.54%)	4381 (65.47%)	5033 (75.21%)	4776 (71.37%)	14 (0.21%)	1229 (18.37%)	414 (6.19%)	484 (7.23%)	259 (3.87%)
As (2x) Un (2x)	5890 (88.02%)	1703 (25.45%)	2876 (42.98%)	3648 (54.51%)	3065 (45.8%)	79 (1.18%)	1908 (28.51%)	982 (14.67%)	882 (13.18%)	620 (9.26%)

Supplementary Table S3. Different pair-wise metrics of nucleosome similarity/dissimilarity. In order to obtain robust estimations of the similarity/dissimilarity of gene architectures between samples we defined the following metrics. *Coverage*: A gene is considered as stable if Pearson's correlation between two samples in the window -300:300 from the TSS is greater than 0.7; is considered as variable if correlation is smaller than 0.5. *Cluster*: We consider a gene stable if the cluster stays the same; we considered a significant variable architecture when 2 of the clustering dimensions (-1/NFR/+1) vary between samples. *+1/-1 Nucleosome*: We consider a nucleosome in the same classification if nucleR's classification is the same for two samples; we considered a gene variable if the absolute difference in nucleR's score is bigger than 0.25 points. *NFR*: we consider a gene stable if the classification of the NFR is the same (open/close/overlap/missing); we considered a change as significant if the change in distance between -1/+1 nucleosomes is more than 100bp. Genes which do not satisfy any of the two criteria are in considered out of the stability/variability threshold. Percentages are relative to the total number of genes in the SacCer3 genome. Key: **1x** – Single End, **2x** – Paired End, **R1** – Replica 1, **R2** – Replica2, **As** – Asynchronous, **Ov** – Overdigested, **Un** – Underdigested.

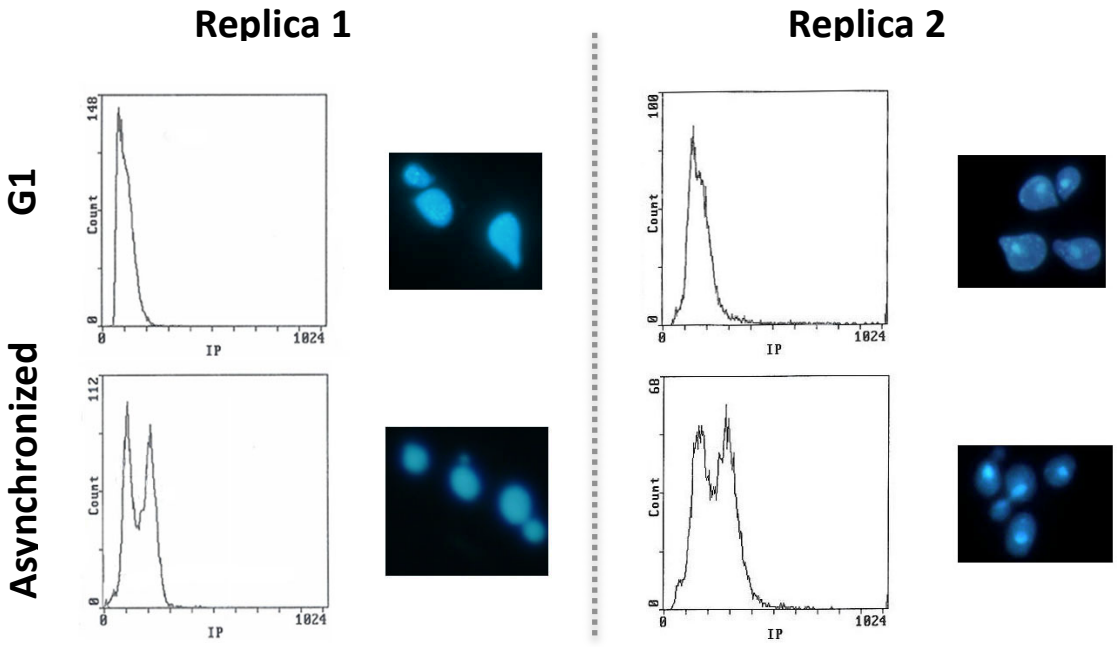
	Stable classification			Variable classification		
vs.	**Coverage**	**Δcalls**	**Δscore**	**Coverage**	**Δcalls**	**Δscore**
R1 (1x) **R1 (2x)**	5688 (85%)	3077 (45.98%)	6010 (89.81%)	169 (2.53%)	446 (6.66%)	33 (0.49%)
R2 (1x) **R2 (2x)**	5436 (81.23%)	3142 (46.95%)	5862 (87.6%)	194 (2.9%)	679 (10.15%)	39 (0.58%)
R1 (1x) **R2 (1x)**	3776 (56.43%)	2976 (44.47%)	5634 (84.19%)	486 (7.26%)	386 (5.77%)	59 (0.88%)
R1 (2x) **R2 (2x)**	5212 (77.88%)	3261 (48.73%)	5759 (86.06%)	341 (5.1%)	577 (8.62%)	53 (0.79%)
R1 (2x) **As (2x)**	3190 (47.67%)	2901 (43.35%)	4958 (74.09%)	869 (12.99%)	461 (6.89%)	96 (1.43%)
R2 (2x) **As (2x)**	4578 (68.41%)	2971 (44.4%)	5461 (81.6%)	459 (6.86%)	537 (8.02%)	72 (1.08%)
R1 (2x) **Ov (2x)**	4297 (64.21%)	3096 (46.26%)	5129 (76.64%)	463 (6.92%)	417 (6.23%)	114 (1.7%)
R2 (2x) **Ov (2x)**	5047 (75.42%)	3448 (51.52%)	5462 (81.62%)	427 (6.38%)	321 (4.8%)	91 (1.36%)
R1 (2x) **Un (2x)**	2550 (38.11%)	2493 (37.25%)	3938 (58.85%)	1176 (17.57%)	769 (11.49%)	262 (3.92%)
R2 (2x) **Un (2x)**	3678 (54.96%)	2647 (39.55%)	4279 (63.94%)	767 (11.46%)	508 (7.59%)	224 (3.35%)
Ov (2x) **Un (2x)**	6058 (90.53%)	2786 (41.63%)	5631 (84.15%)	121 (1.81%)	459 (6.86%)	48 (0.72%)
As (2x) **Ov (2x)**	6101 (91.17%)	3014 (45.04%)	6094 (91.06%)	91 (1.36%)	554 (8.28%)	17 (0.25%)
As (2x) **Un (2x)**	5280 (78.9%)	2462 (36.79%)	5294 (79.11%)	237 (3.54%)	828 (12.37%)	101 (1.51%)

Supplementary Table S4. Different pair-wise metrics of nucleosome similarity/dissimilarity in gene body regions. *Coverage*: a particular gene is considered 'stable' when Pearson's correlation between the whole gene body of two samples is greater than 0.7; otherwise, it is considered 'variable' when correlation is smaller than 0.5. Δ*calls*: a gene is considered 'stable' when it has the same number of nucleosome calls inside the gene body; otherwise, it is considered 'variable' when the difference of nucleosome calls between samples is more than 2 nucleosomes. Δ*score*: a particular gene is considered to be the same between two samples when the equivalent nucleosome calls have a mean difference of nucleR score lower than 0.15; otherwise, the gene is assigned as variable when this difference is higher than 0.25. Genes which do not satisfy any of the two criteria are in considered out of the stability/variability threshold. Percentages are relative to the total number of genes in the SacCer3 genome. *Note: the present metrics differ from those presented in the previous Supplementary Table 3, in order to account for an arbitrary number of nucleosome calls in gene body regions.* Key: **1x** – Single End, **2x** – Paired End, **R1** – Replica 1, **R2** – Replica2, **As** – Asynchronous, **Ov** – Overdigested, **Un** – Underdigested.

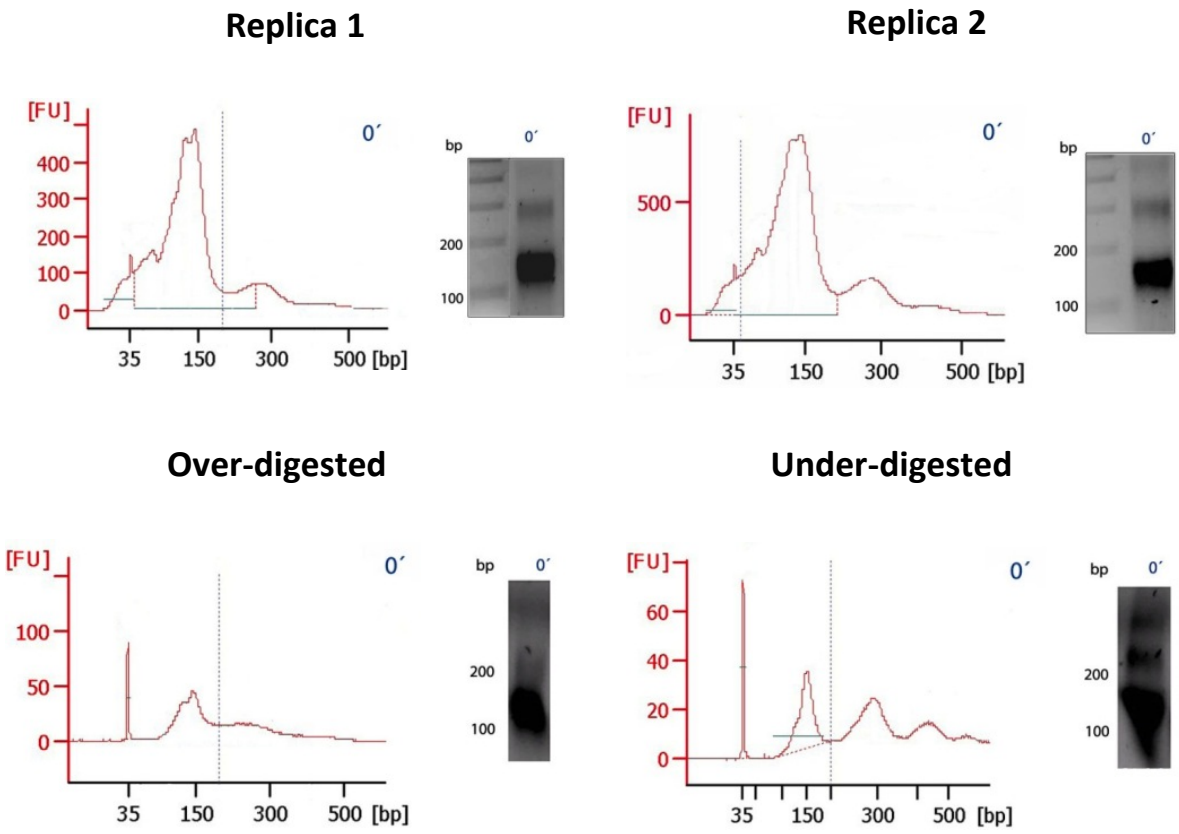
	Stable classification			Variable classification		
vs.	**Coverage**	**Δcalls**	**Δscore**	**Coverage**	**Δcalls**	**Δscore**
R1 (1x) R1 (2x)	6210 (92.80%)	3965 (59.25%)	5868 (87.69%)	81 (1.21%)	227 (3.39%)	27 (0.4%)
R2 (1x) R2 (2x)	5944 (88.82%)	2893 (43.23%)	5729 (85.61%)	118 (1.76%)	830 (12.4%)	50 (0.75%)
R1 (1x) R2 (1x)	5455 (81.52%)	3411 (50.97%)	5220 (78%)	273 (4.08%)	82 (1.23%)	92 (1.37%)
R1 (2x) R2 (2x)	5929 (88.60%)	4397 (65.71%)	5614 (83.89%)	174 (2.6%)	281 (4.2%)	76 (1.14%)
R1 (2x) As (2x)	5136 (76.75%)	3907 (58.38%)	4825 (72.10%)	364 (5.44%)	54 (0.81%)	176 (2.63%)
R2 (2x) As (2x)	5738 (85.74%)	4303 (64.30%)	5366 (80.19%)	242 (3.62%)	92 (1.37%)	108 (1.61%)
R1 (2x) Ov (2x)	5562 (83.11%)	3885 (58.05%)	4787 (71.53%)	209 (3.12%)	252 (3.77%)	179 (2.67%)
R2 (2x) Ov (2x)	5820 (86.97%)	4477 (66.90%)	5109 (76.34%)	223 (3.33%)	87 (1.3%)	118 (1.76%)
R1 (2x) Un (2x)	4344 (64.91%)	2890 (43.19%)	3928 (58.70%)	649 (9.7%)	179 (2.67%)	447 (6.68%)
R2 (2x) Un (2x)	4975 (74.34%)	3000 (44.83%)	3880 (57.98%)	445 (6.65%)	160 (2.39%)	453 (6.77%)
Ov (2x) Un (2x)	6248 (93.37%)	3248 (48.54%)	5254 (78.51%)	67 (1%)	118 (1.76%)	87 (1.3%)
As (2x) Ov (2x)	6353 (94.93%)	4059 (60.65%)	5982 (89.39%)	43 (0.64%)	290 (4.33%)	33 (0.49%)
As (2x) Un (2x)	5837 (87.22%)	3009 (44.96%)	4858 (72.59%)	163 (2.44%)	149 (2.23%)	170 (2.54%)

Supplementary Table S5. Different pair-wise metrics of nucleosome similarity/dissimilarity around TSSs (window -300:+300). Metrics used here are the same than in Supplementary Table S4.

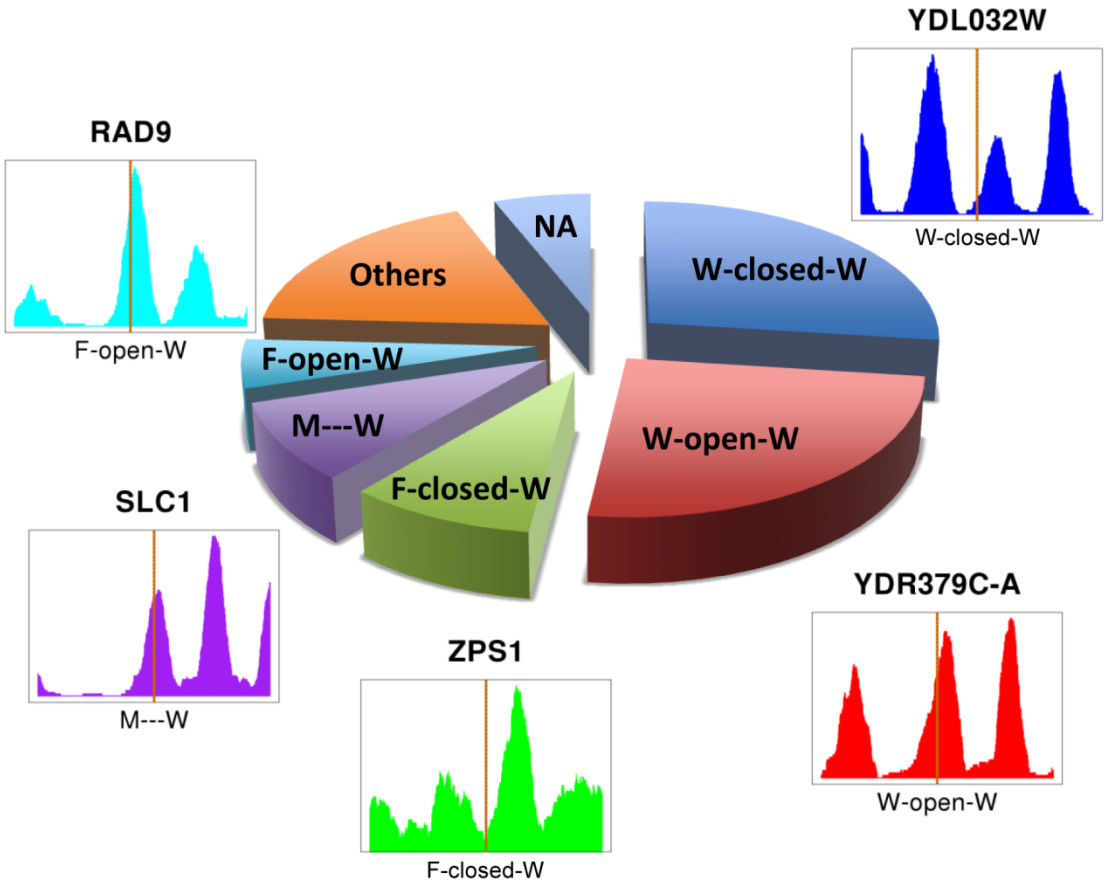
SUPPLEMENTARY FIGURES



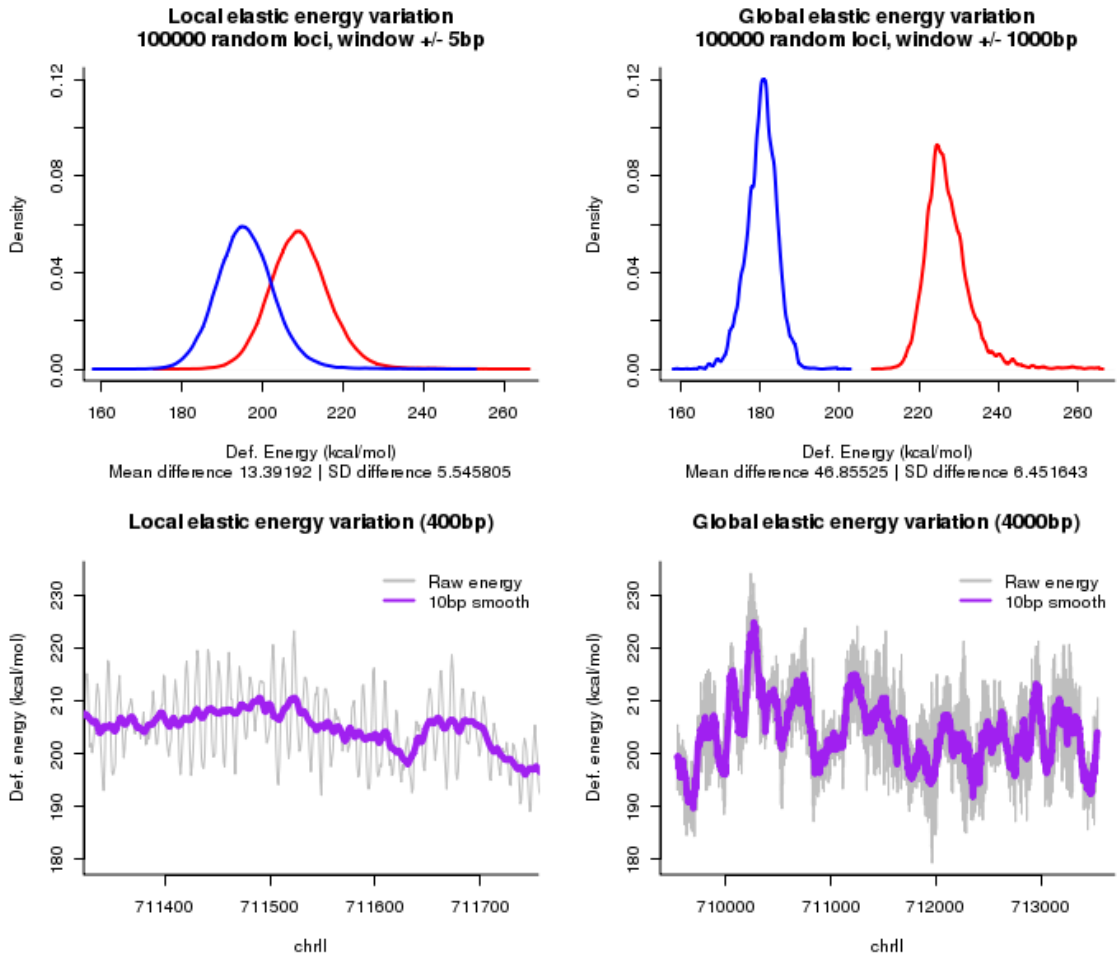
Supplementary Figure S1. Flow cytometry analysis and fluorescence microscope images of late G1 synchronized cells (upper panel) and asynchronous cells (lower panel) for *replicas* 1 (left) and 2 (right).



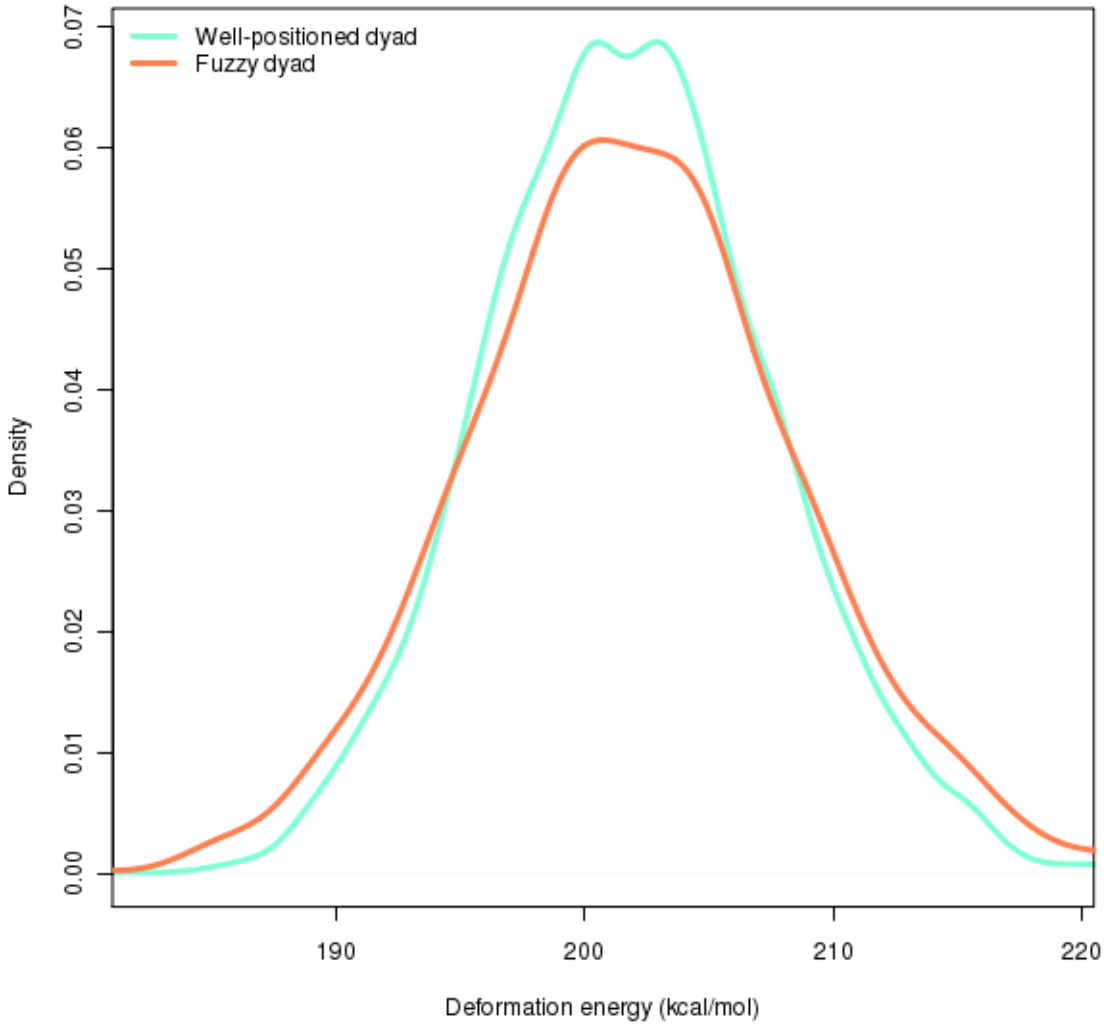
Supplementary Figure S2. MNase digestion profiles of replica 1 (top-left), replica 2 (top-rigth), over-digested sample (bottom-left) and under-digested sample (bottom-right). The left panels show the size distribution of digested DNA molecules as measured by Bioanalyzer and the right panels show the agarose gel analysis of digestion products.



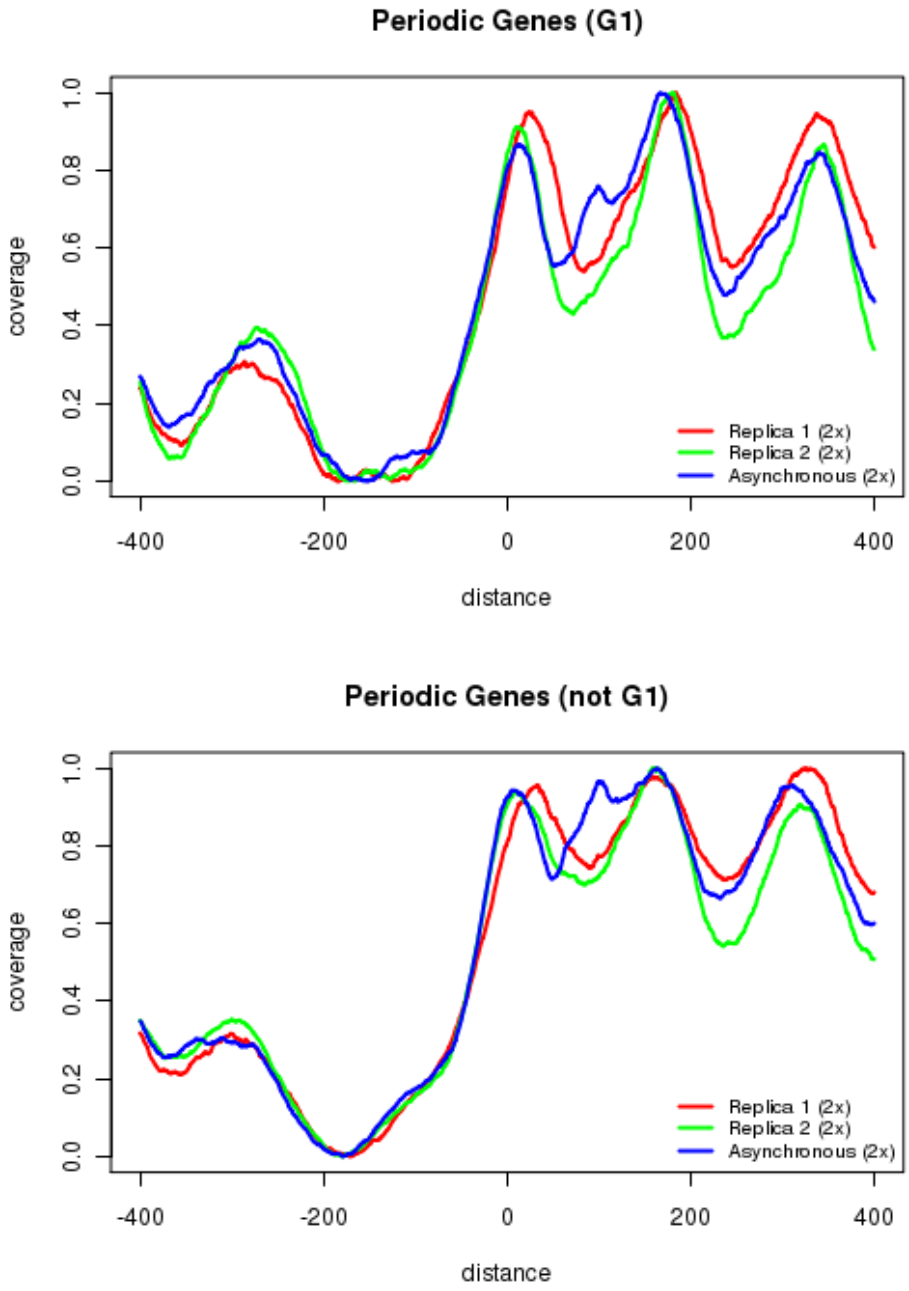
Supplementary Figure S3. **Gene clustering according to nucleosomal architecture at transcription start sites.** Pie-chart shows the gene distribution for the most populated classes in the sample Replica 2 (2x). For every class, an example of the nucleosome coverage around the TSS of a representative gene is illustrated (window -300:300 from the TSS, marked in red). All plots show the coverage in 5'->3' direction, representing the +1 nucleosome as the peak overlapping or immediately downstream TSS and the -1 nucleosome as the peak right upstream of +1. In the case of SLC1, -1 nucleosome peak is not detected in the -300:300 window.



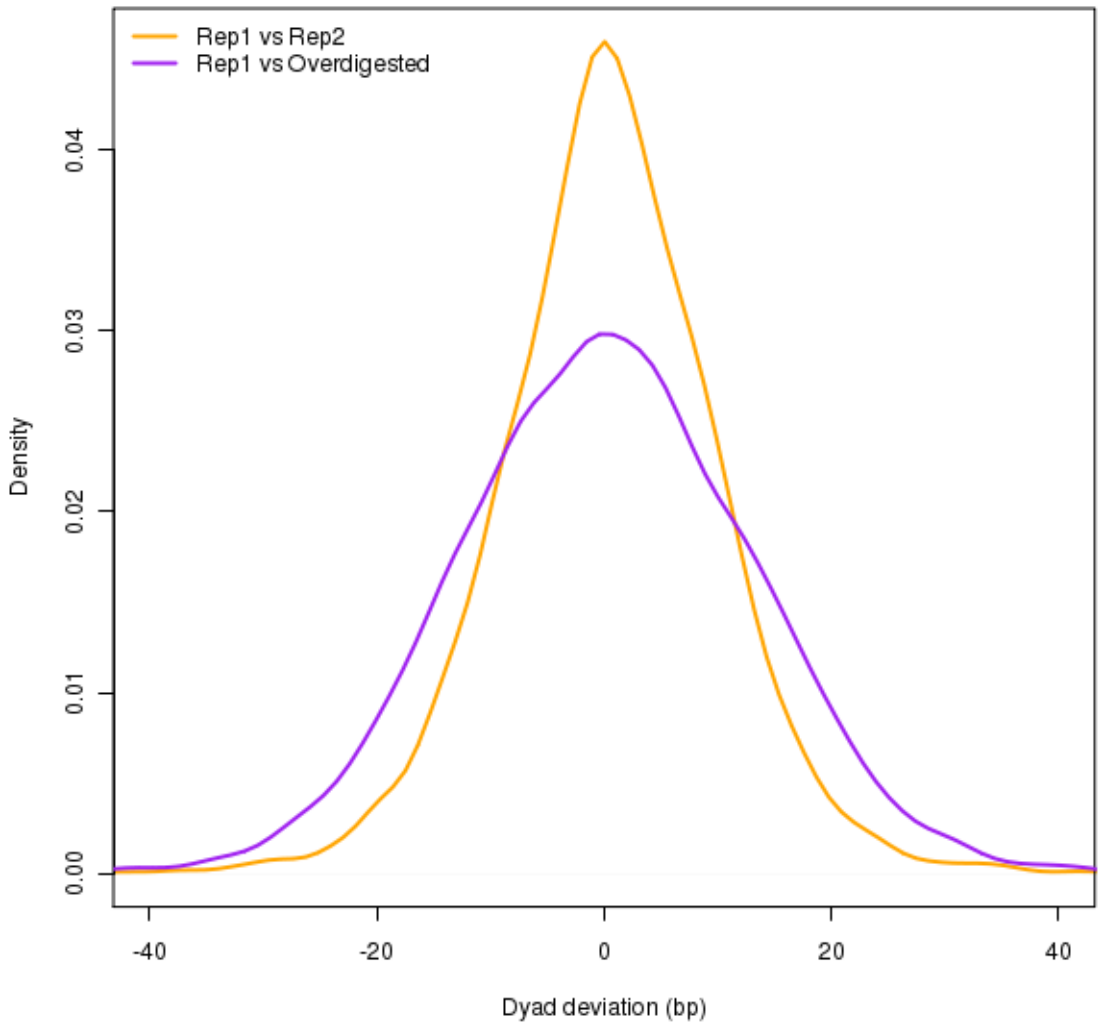
Supplementary Figure S4. Local and global energy variation. Despite local energy variation involves a strong periodicity of 10bp with small energy fluctuations (around 13.39 kcal/mol) (left), these don't act as a strong regulator of the nucleosome fuzziness. Global energy barriers with a larger mesoscopic effect (around 46.86 kcal/mol) could act as intrinsic regulator of the nucleosome phasing along different cells. On top, minimum (blue) and maximum (red) values in a window of +/- 5bp (left) and +/- 1000bp right of 100000 random loci. On the bottom, we show the raw energy (grey) and the 10bp average (purple) of single random region of the chromosome II (left: 400bp window, right:4000bp window).



Supplementary Figure S5. Deformation energy of well-positioned and fuzzy nucleosomes. Deformation energy around +/-5bp around the peak summit has been calculated for annotated -1/+1 nucleosomes. Mean value of every 10 possible combinations was used to account for local periodicity. Fuzzy and well-positioned nucleosomes are taken from the common ones in *replicas* 1 and 2.



Supplementary Figure S6. Effect of cell-cycle periodic genes in nucleosome map. Coverage of cell-cycle periodic genes is shown for G1 related genes (top, 211 genes) and in other stages (bottom, 365 genes). Asynchronous sample (blue) shows a larger perturbation between the +1/-1 nucleosome peaks in both cases.



Supplementary Figure S7. Comparison of dyad deviations due to different digestion. Dyad distances of annotated -1/+1 nucleosomes (coverage peak summits) have been calculated between biological replicates and over-digested sample. Absolute mean deviation between Rep1 and Rep2 is 14.25 bp, compared with 18.75bp (+4.5bp) in the case of Rep1 with over-digested sample.