Molecular Cell, Volume *48*

Supplemental Information

A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning: A Unifying Model for Establishing the Genome-wide Pattern

Amanda L. Hughes, Yi Jin, Oliver J. Rando, and Kevin Struhl

Supplemental Inventory

Figure S1, related to Figures 1,2: NDRs are generally better-maintained over sequence from *K. lactis* than over *D. hansenii* sequence.

Figure S2, related to Figures 1-5: Bulk chromatin is not affected in YAC-bearing strains.

Figure S3, related to Figure 4: +1 nucleosome shifts associated with transcription.

Figure S4, related to Figures 4, 5: Comparison of RNA-Seq and MNase-Seq datasets.

Figure S5, related to Figure 4: Nucleosome positioning shifts are associated with shifts in TSS.

Figure S6, related to Figure 4, 5: Example of fortuitous NDR in *D. hansenii* YAC.

Figure S7, related to Figure 6: Yeast whole cell extracts poorly position nucleosomes.

SUPPLEMENTAL FIGURES

Figure S1

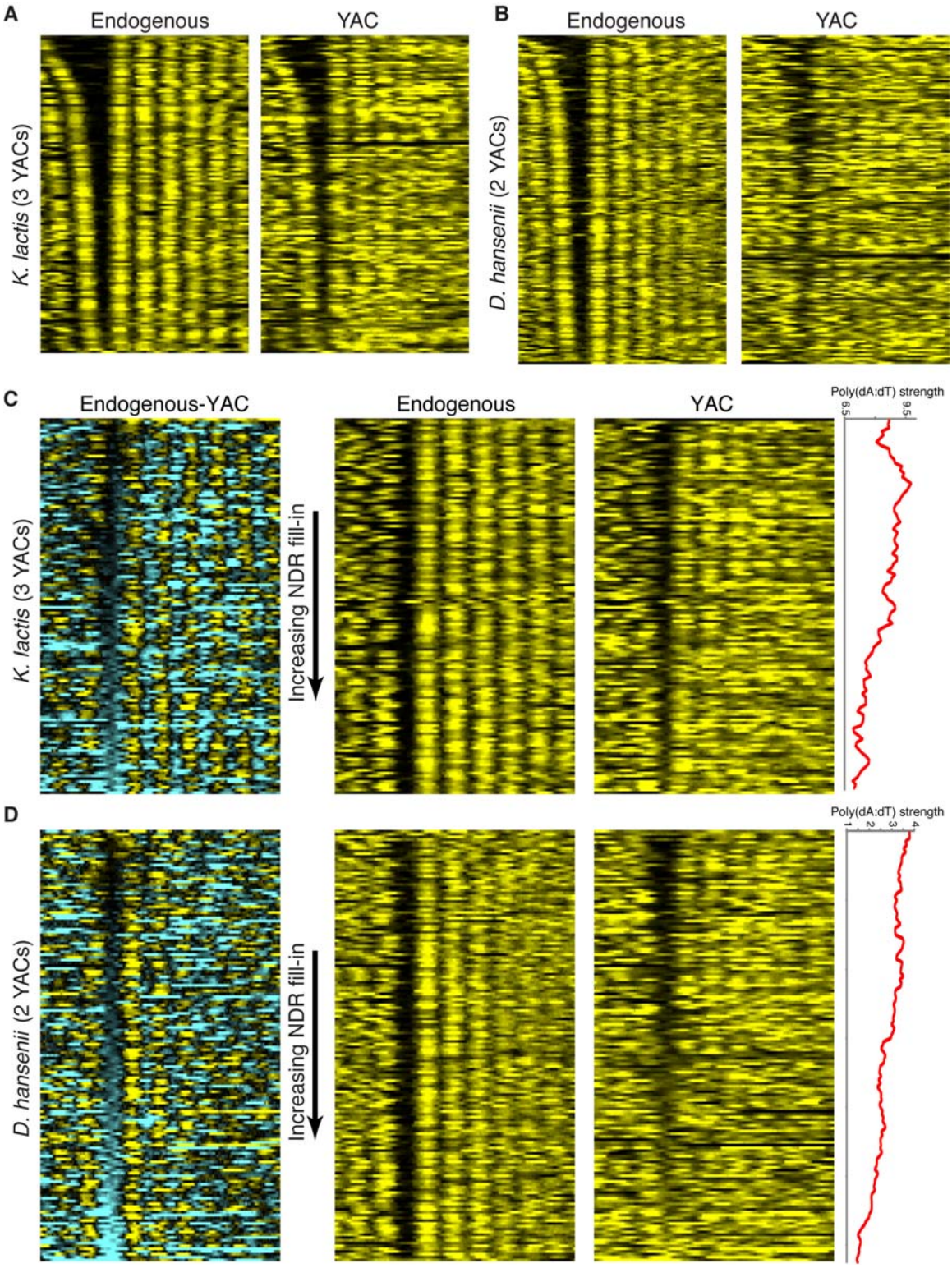


Figure S2

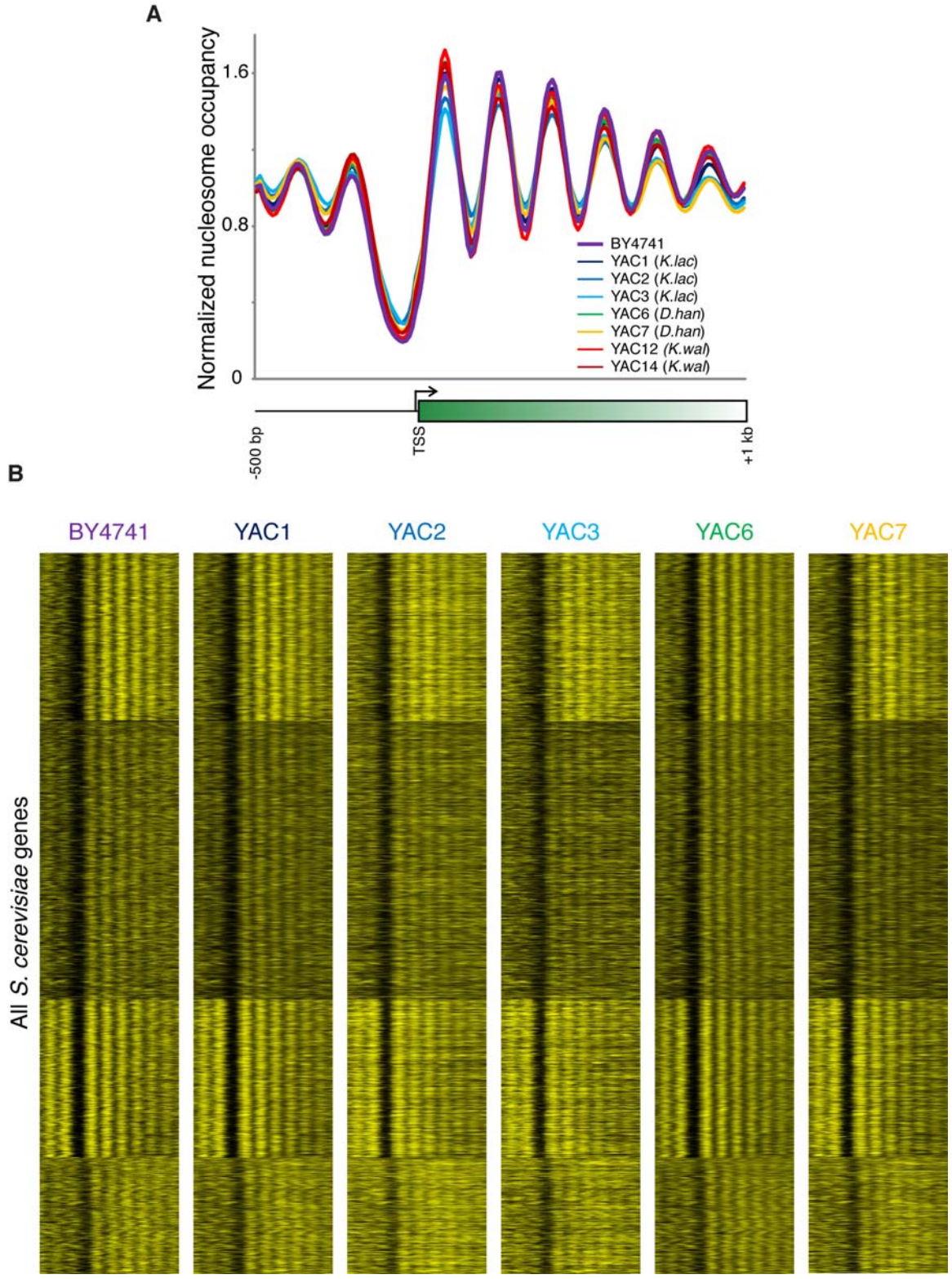

A
Normalized nucleosome occupancy
1.6
0.8
0
BY4741
YAC1 (K.lac)
YAC2 (K.lac)
YAC3 (K.lac)
YAC6 (D.han)
YAC7 (D.han)
YAC12 (K.wal)
YAC14 (K.wal)
-500 bp
TSS
+1 kb
B
BY4741
YAC1
YAC2
YAC3
YAC6
YAC7
All S. cerevisiae genes

Figure S3

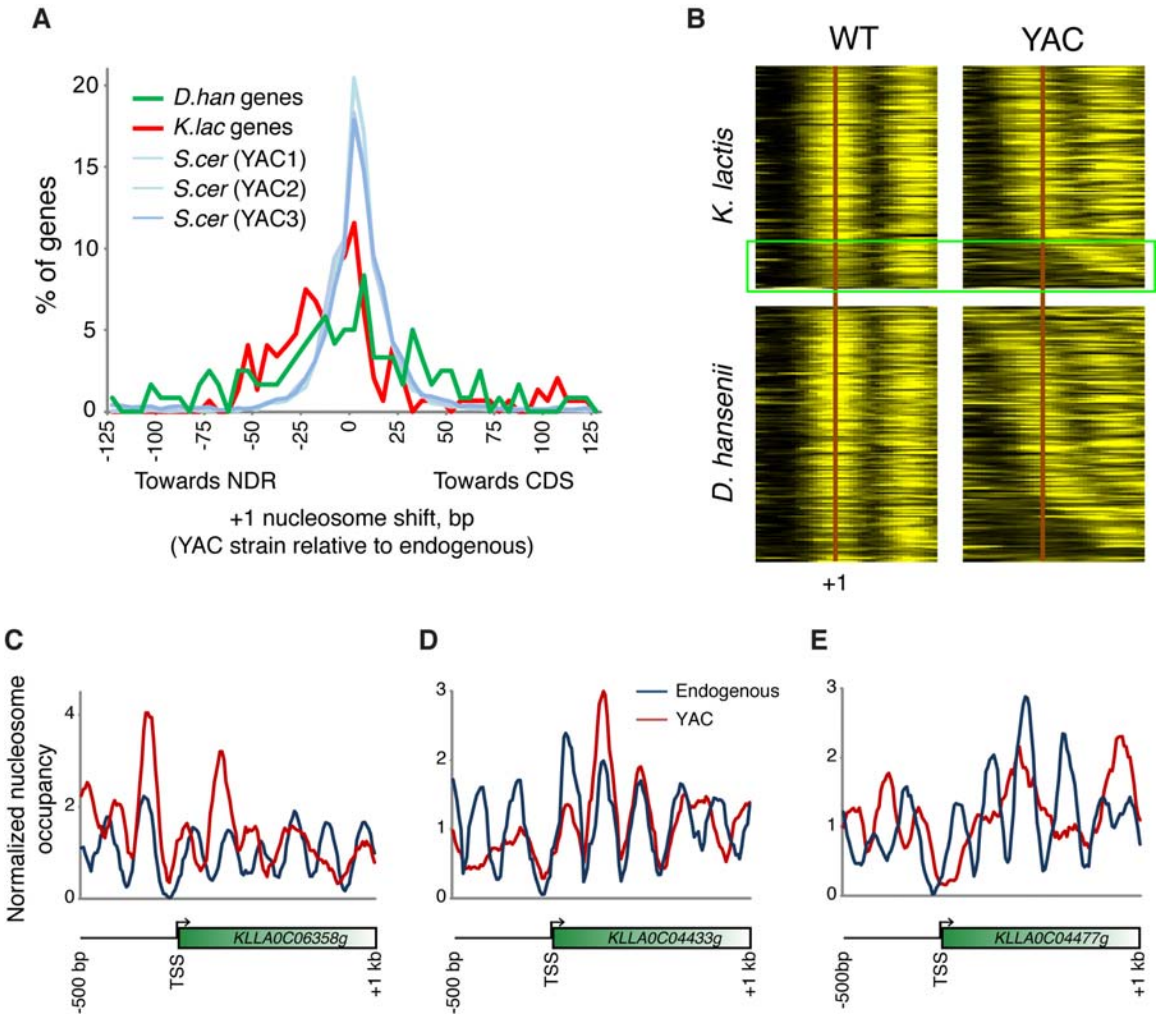

A
% of genes
D.han genes
K.lac genes
S.cer (YAC1)
S.cer (YAC2)
S.cer (YAC3)
Towards NDR
Towards CDS
+1 nucleosome shift, bp
(YAC strain relative to endogenous)
B
WT
YAC
K. lactis
D. hansenii
+1
C
D
E
Normalized nucleosome occupancy
Endogenous
YAC
KLLA0C06358g
KLLA0C04433g
KLLA0C04477g
-500 bp
TSS
+1 kb
-500bp

Figure S4

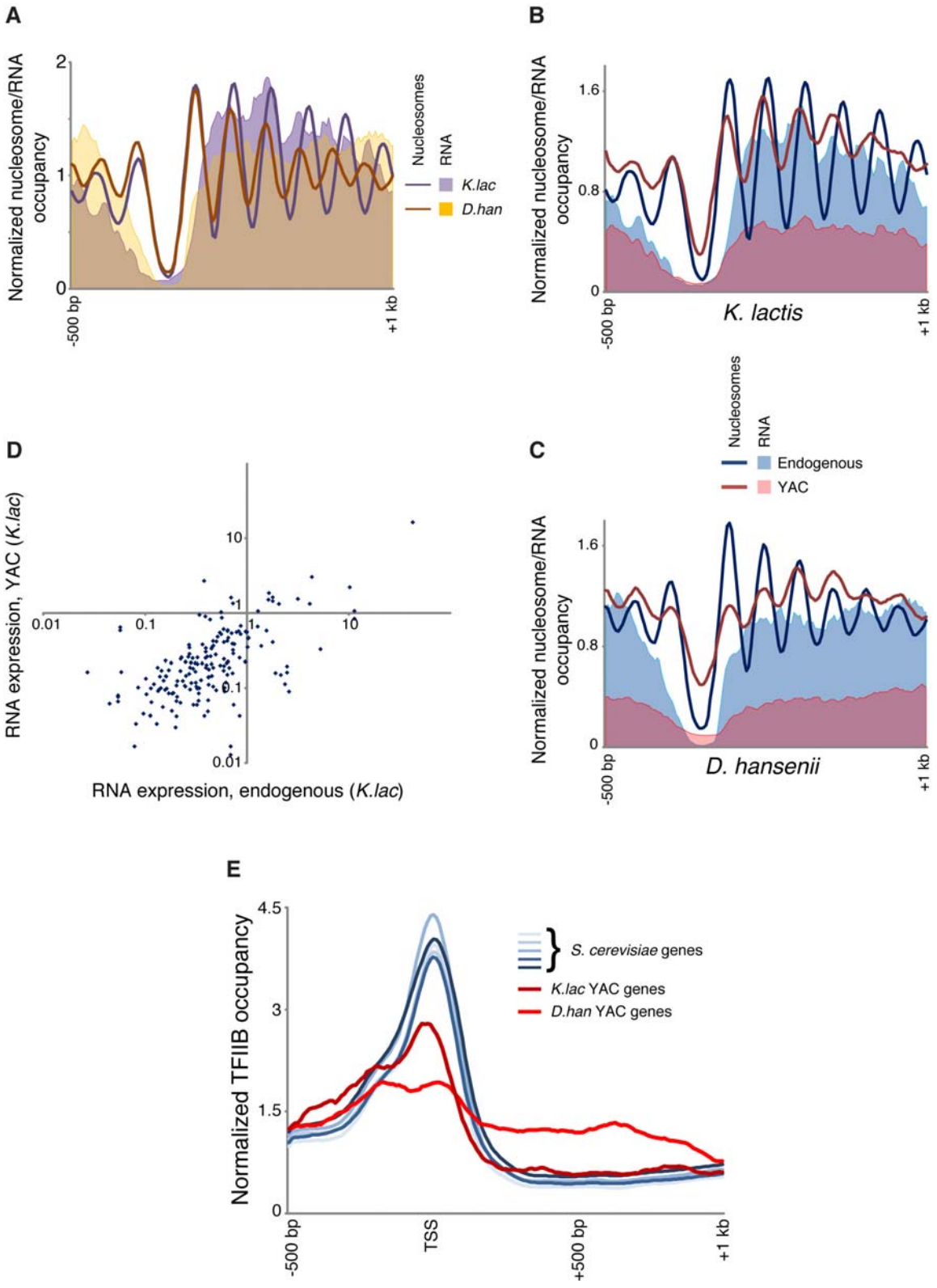

A
Normalized nucleosome/RNA occupancy
2
1
0
-500 bp
+1 kb
Nucleosomes
RNA
K.lac
D.han
B
Normalized nucleosome/RNA occupancy
1.6
0.8
0
-500 bp
+1 kb
K. lactis
C
Nucleosomes
RNA
Endogenous
YAC
Normalized nucleosome/RNA occupancy
1.6
0.8
0
-500 bp
+1 kb
D. hansenii
D
RNA expression, YAC (K.lac)
10
1
0.1
0.01
0.01
0.1
1
10
RNA expression, endogenous (K.lac)
E
Normalized TFIIB occupancy
4.5
3
1.5
0
-500 bp
TSS
+500 bp
+1 kb
S. cerevisiae genes
K.lac YAC genes
D.han YAC genes

Figure S5

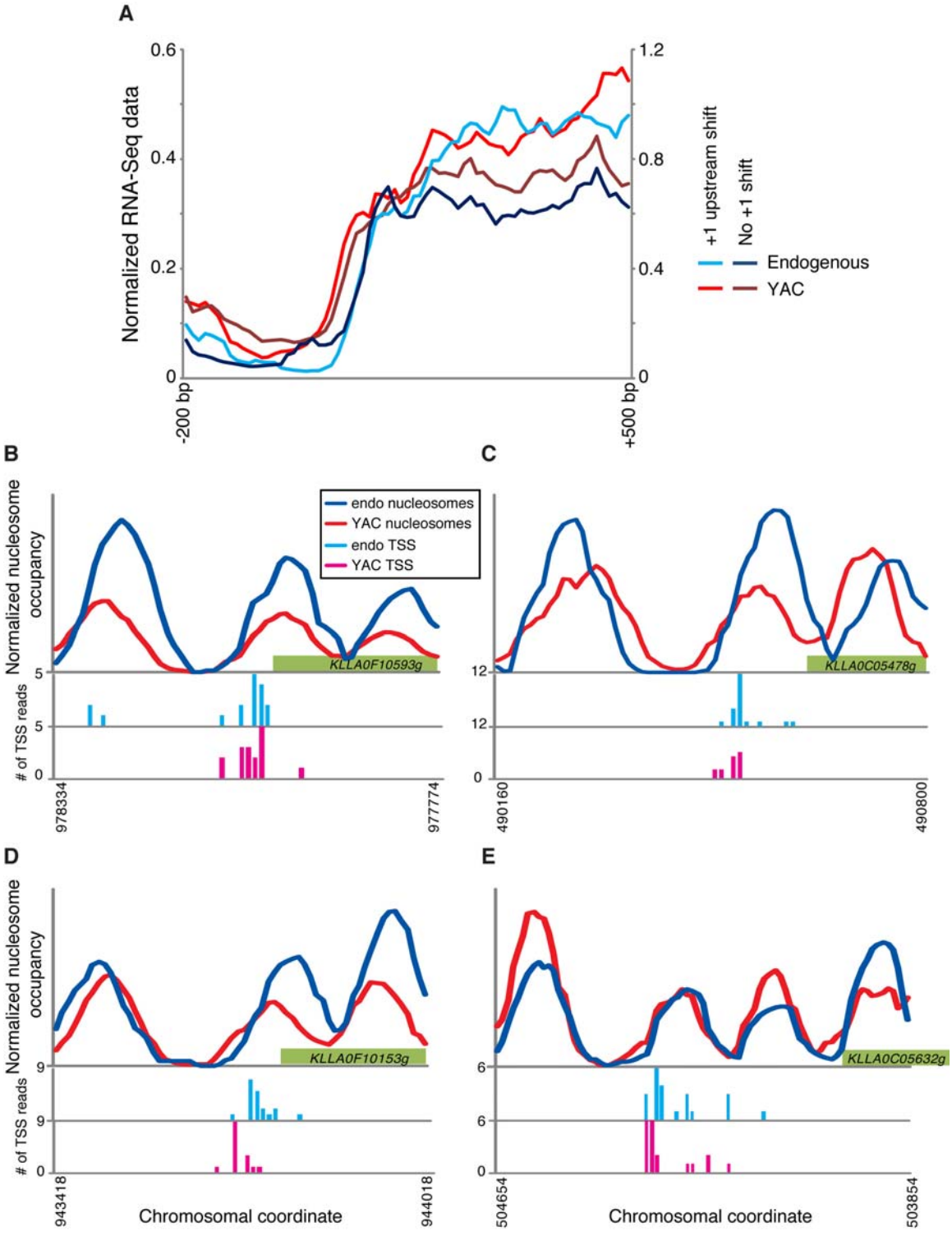

A
Normalized RNA-Seq data
0.6
0.4
0.2
0
1.2
0.8
0.4
0
-200 bp
+500 bp
+1 upstream shift
No +1 shift
Endogenous
YAC
B
C
Normalized nucleosome occupancy
endo nucleosomes
YAC nucleosomes
endo TSS
YAC TSS
KLLA0F10593g
KLLA0C05478g
# of TSS reads
978334
977774
490160
490800
D
E
KLLA0F10153g
KLLA0C05632g
943418
944018
504654
503854
Chromosomal coordinate
Chromosomal coordinate

Figure S6

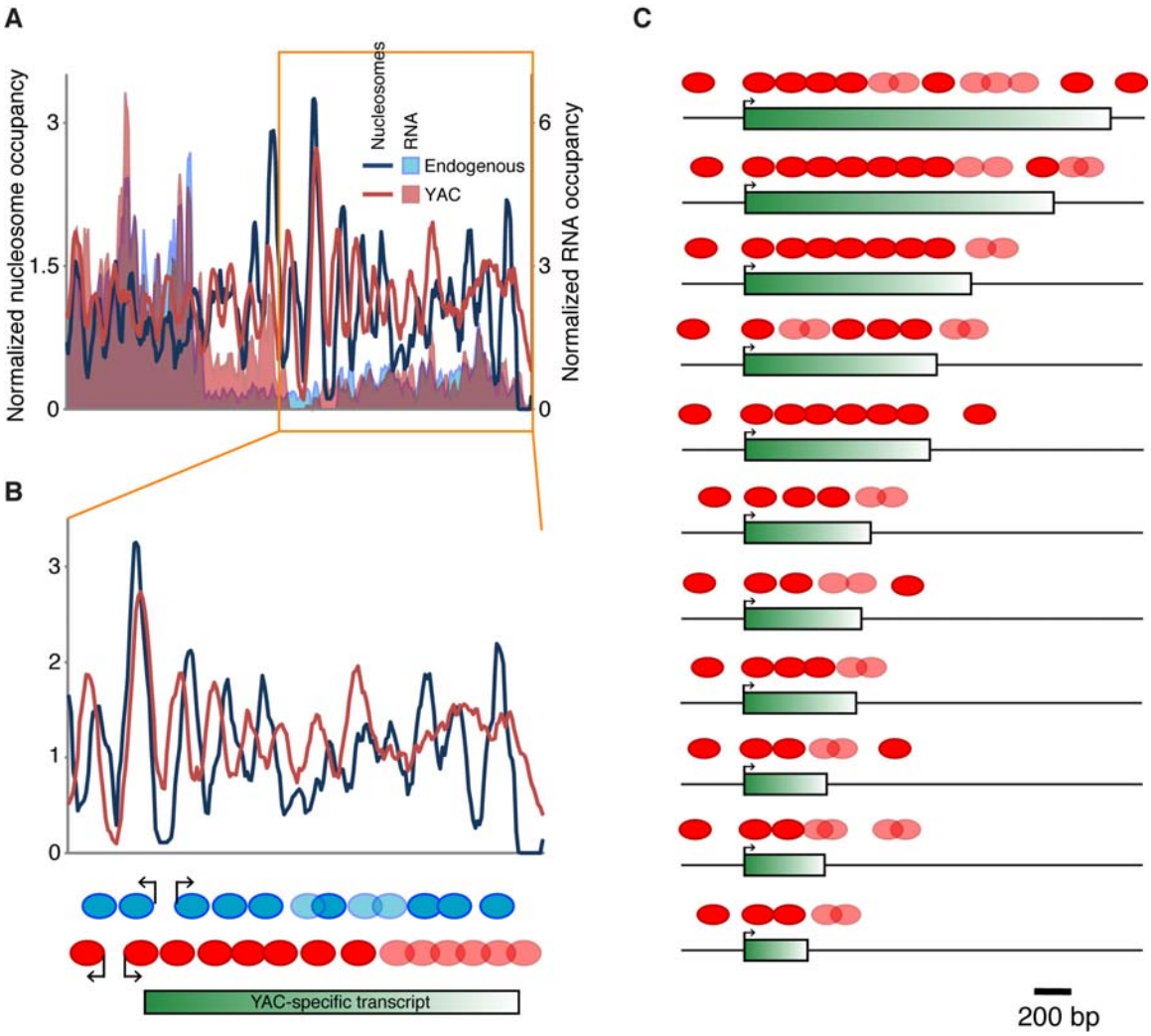
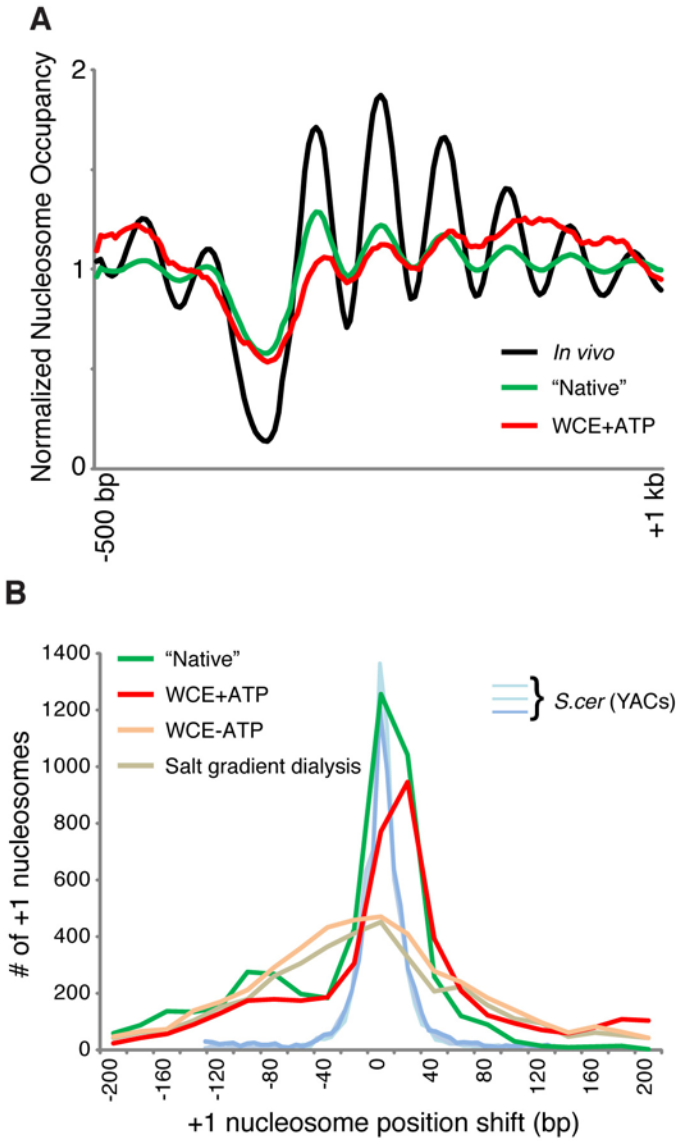

A
Normalized nucleosome occupancy
3
1.5
0
Nucleosomes
RNA
Endogenous
YAC
6
3
0
Normalized RNA occupancy
B
3
2
1
0
YAC-specific transcript
C
200 bp

Figure S7


A
Normalized Nucleosome Occupancy
2
1
0
-500 bp
+1 kb
In vivo
"Native"
WCE+ATP
B
of +1 nucleosomes
1400
1200
1000
800
600
400
200
0
"Native"
WCE+ATP
WCE-ATP
Salt gradient dialysis
S.cer (YACs)
-200
-160
-120
-80
-40
0
40
80
120
160
200
+1 nucleosome position shift (bp)

SUPPLEMENTAL FIGURE LEGENDS

Figure S1, related to Figures 1-2

NDRs are generally better-maintained over sequence from *K. lactis* than over *D. hansenii* sequence

(**A-B**) Nucleosome mapping data for all genes from *K. lactis* (**A**) or *D. hansenii* (**B**) are shown for wild-type and YACs, as indicated. Genes are sorted by wild-type NDR width.

(**C**) NDR maintenance correlates with poly(dA:dT) elements. As in **Figures 2A-B**, but for *K. lactis*.

(**D**) Identical to **Figures 2A-B**, reproduced here for comparison between *D. hansenii* and *K. lactis*.

Figure S2, related to Figures 1-5

Bulk chromatin is not affected in YAC-bearing strains

(**A**) Nucleosome sequencing data for all strains in this study was mapped to the *S. cerevisiae* genome, and data are averaged for all genes aligned by the +1 nucleosome.

(**B**) Data for all *S. cerevisiae* genes are shown for wild-type (BY4741) and 5 YAC-bearing strains. Genes are sorted by K means clustering (K=4) of BY4741 dataset.

Figure S3, related to Figure 4

+1 nucleosome shifts associated with transcription.

(**A**) Distribution of shifts for +1 nucleosomes. Distributions are shown for the shifts between +1 positions in wild-type and YAC-bearing strains. The three *S. cerevisiae* distributions show changes in +1 positioning for *S. cerevisiae* genes in the indicated YACs, showing that technical variability or analytical variability do not account for nucleosome position changes for YAC-associated genes.

(**B**) Data for all *K. lactis* and *D. hansenii* genes, from wild-type and YACs. Data are shown from –100 to +300 bp relative to the +1 nucleosome upstream border (+1

nucleosome center indicated as a red line). Green box for *K. lactis* genes indicates a set of genes with low +1 nucleosome occupancy in wild-type, where the downstream shift in the YAC is likely due to failure to correctly call the low occupancy +1 nucleosome in the YAC.

(**C-E**) Examples of *K. lactis* genes exhibiting different +1 nucleosome shifts in the YAC context, including an upstream shift (**C**), an unchanged +1 (**D**), and a downstream shift (**E**).

Figure S4, related to Figures 4-5

Comparison of RNA-Seq and MNase-Seq datasets.

(**A**) TSS positioning relative to +1 nucleosome. Averaged nucleosome data (solid lines) or RNA-Seq data (shaded area) for *K. lactis* and *D. hansenii* wild type cells, as indicated. All genes are aligned by +1 nucleosome position.

(**B-C**) Lower expression of genes on YACs relative to endogenous expression. For *K. lactis* (**B**) or *D. hansenii* (**C**), nucleosome data and RNA-Seq data are plotted as indicated for YAC-associated genes. For YAC-associated genes, RNA-Seq data are normalized to whole-genome RNA data (e.g. including *S. cerevisiae* genes, in contrast to the normalization in **Figure 4**), with lower normalized abundance indicating that YAC-associated genes are expressed at lower levels than are endogenous genes, assuming similar RNA content of the various species.

(**D**) RNA abundance for *K. lactis* genes. Normalized RNA-Seq data from *K. lactis* (x axis) or YAC-bearing strains (y axis) is scatter-plotted with each point representing a single gene. mRNA abundance data are shown as reads per kb per million reads. Note good correlation between endogenous and YAC-based expression, indicating that differences between poorly and highly-expressed genes are maintained in a foreign nuclear environment.

(**E**) ChIP-Seq was carried out for TFIIB in the YAC-bearing strains. TSS-aligned data are shown for all *S. cerevisiae* genes in each strain, or for all YAC-based genes for either *K. lactis* or *D. hansenii*. TFIIB ChIP could not be carried out for the other organisms due to

the limited cross-reactivity of our antibody. Note that foreign promoters continue to recruit TFIIB in *S. cerevisiae*, but that for both organisms TFIIB exhibits worse localization in the YAC than expected, due to divergence of regulatory information between species. This is especially true for *D. hansenii*.

Figure S5, related to Figure 4

Nucleosome positioning shifts are associated with shifts in TSS.

(**A**) Averaged nucleosome mapping and RNA-Seq data are shown for all *K. lactis* genes exhibiting either no +1 nucleosome shift (less than 20 bp in either direction) in the YAC, or a 20 bp or more upstream shift in the YAC. Note that RNA-Seq data shift 5' in both cases when *K. lactis* genes are expressed in *S. cerevisiae*, but that genes with 5' nucleosome shifts exhibit a greater 5' shift, consistent with a constant distance being maintained between TSS and +1 nucleosome positioning. Note that genes with 3' shifts in the +1 nucleosome are not included, as these largely represent genes where the +1 nucleosome decreases occupancy and hence is miscalled (**Figure S3B**).

(**B-E**) TSS mapping for 4 individual genes. TSSs were mapped by 5' RACE (**Methods**), and individual clone locations are shown as indicated below the nucleosome mapping data. Note that for two genes with little +1 nucleosome position shift (**B**, **E**) there is little change in TSS, whereas the two genes with 5' shifts in the +1 nucleosome (**C**, **D**) also show upstream shifts in the TSS in the YAC.

Figure S6, related to Figures 4-5

Examples of fortuitous NDRs in *D. hansenii* YACs.

(**A-B**) As in **Figure 5**, another example of a fortuitous NDR that occurs only in YACs, and is associated with transcription. (**A**) shows nucleosome and RNA data for 4 kb surrounding a YAC-specific NDR, with (**B**) showing nucleosome data only for the indicated region. Note increasing nucleosome fuzziness in the YAC nucleosome data at the 3' end of the transcript.

(**C**) Extent of positioned nucleosome array is linked to RNA transcript length. Schematic interpretation of the nucleosome positioning for RNA transcripts with different lengths, derived from fortuitous coding region NDRs. RNA transcript is shown in green rectangle and nucleosome positioning is shown in solid red (well positioned nucleosome) and transparent light red (less positioned nucleosome) above the rectangle. Black arrows indicate inferred TSSs.

Figure S7, related to Figure 6

Yeast whole cell extracts poorly position nucleosomes

(**A**) Nucleosome mapping data from intact yeast ("*in vivo*"), from lysed yeast equilibrated prior to crosslinking ("native"), or from yeast genomic DNA incubated with yeast whole cell extract and ATP (WCE+ATP) *in vitro* (Zhang et al., 2011). Yeast whole cell extract performs significantly better than salt gradient dialysis, as reported, but the nucleosome positions recovered nonetheless do not precisely match nucleosome positions recovered from intact yeast – compare "native" and *in vivo* positioning.

(**B**) +1 nucleosome positioning in "native" yeast extracts exhibits systematic deviation from *in vivo* positioning. +1 nucleosome positions were called, and distance from +1 nucleosome positions *in vivo* to the positions recovered in various datasets is shown as a histogram. As a comparison for technical variation, we show data for +1 positioning variability in our YAC strains (to keep datasets on the same y axis we used a 5 bp bin size for YACs rather than the 10 bp used for other datasets).

SUPPLEMENTAL TABLE

Table S1, related to the RESULTS

Strains used in this study

Strains	Description
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
AB1380	MATa ura3-52 trp1-289 lys2-1 ade2-1 can1-100 his5 ρ+ ψ+
K. lactis	CLIB 209
D. hansenii	NCYC 2572
K. waltii	NCYC 2644
YAC1	AB1380 + 128 kb YAC (*K. lactis* Chromosome F 872404~1000550)
YAC2	AB1380 + 143 kb YAC (*K. lactis* Chromosome C 339713~482935)
YAC3	AB1380 + 136 kb YAC (*K. lactis* Chromosome C 443175~578764)
YAC6	AB1380 + 115 kb YAC (*D. hansenii* Chromosome C 1165392~1280355)
YAC7	AB1380 + 216 kb YAC (*D. hansenii* Chromosome D 1148162~1364529)
YAC12	AB1380 + 120 kb YAC (*K. waltii* contig S0 133276~N/A*)
YAC14	AB1380 + 131 kb YAC (*K. waltii* contig S33 510773~641751)

*N/A: The coordinate is unable to be identified, due to incomplete *K. waltii* genome sequences.