Supplementary Information

Interaction of Transcriptional Regulators with Specific Nucleosomes Across the *Saccharomyces* Genome

R. Thomas Koerber, Ho Sung Rhee, Cizhong Jiang, and B. Franklin Pugh
Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802

Supplementary Tables (*Examples only – see associated spreadsheets*)

Supplementary Table 1 – Excel workbook containing the nucleosome calls for all datasets used here. There are six worksheets (Bdf1, Vps72, Rap1, Reb1, Rpo21,and Srm1). The top 150 nucleosomes used for this analysis and all significant nucleosomes (P <0.05) are marked.

Bdf1 worksheet example

number	chrom	strand	start	end	peak_height	comment
1	chr12	A	459832	459978	812.2328535	Top 150
2	chr4	A	1135911	1136057	766.2404961	Top 150
3	chr1	A	120044	120190	434.679088	Top 150
4	chr4	A	392028	392174	347.5648211	Top 150

Supplementary Table 2 – Excel workbook containing a list of all genes, and the clusters they belong to. "0" indicates that the gene was not part of the main clusters.

Gene list example

Feature_ID	Bdf1/Vps72	Rap1	Reb1	Rpo21	Srm1
YAL016W					
YAL017W	2	1			
YAL019W					
YAL026C					
YAL032C					
YAL033W		0			
YAL034W-A					
YAL035W	2			1	
YAL036C	0			0	

Supplementary Table 3 – Excel workbook containing frequency distribution information used to generate composite graphs for individual clusters. All graphs were binned every 25 bp, and smoothed every 3 bins with the midpoint of the bin being reported. Each cluster of graphs is for the same gene list to which the factor-bound top 150 nucleosomes were mapped.

Bdf1 worksheet information

Cluster 1 (C1) – Bdf1 located at +1 nucleosome. Traces included:

A	B	C	D	E	F	G
Bin	H3/H4 (Bdf1 C1)	H2AZ (Bdf1 C1)	Bdf1 C1	Rpo21 Nuc (Bdf1 C1)	Rpo21 Std (Bdf1 C1)	Vps72 (Bdf1 C1)

Figure 2C (top) – B (All), C (H2A.Z), D (Bdf1-bound), G (Vps72-bound)
Figure 3E – D (Bdf1-bound), E (Pol II-bound nucleosomes), F (Pol II)

Cluster 2 – Bdf1 located at the +2 nucleosome. Traces included:

I	J	K	L	M	N	O
Bin	H3/H4 (Bdf1 C2)	H2AZ (Bdf1 C2)	Bdf1 C2	Rpo21 Nuc (Bdf1 C2)	Rpo21 Std (Bdf1 C2)	Vps72 (Bdf1 C2)

Figure 2C (bottom) – J (All), K (H2A.Z), L (Bdf1-bound), O (Vps72-bound)
Figure 3F – L (Bdf1-bound), M (Pol II-bound nucleosomes), N (Pol II)

Bdf1 CoIP worksheet information

A	B	C
Bin	H3/H4 (Bdf1 CoIP)	Bdf1 CoIP

Figure 3B – B (All nucleosomes), C (Bdf1-bound di-nucleosome)

Rap1 worksheet information

A	B	C
Bin	H3/H4 (Rap1)	Rap1

Figure 4C – B (All nucleosomes), C (Rap1-bound nucleosomes)

Reb1 worksheet information

A	B	C	D	E
Bin	H3/H4 (Reb1 C1)	Reb1 C1	H3/H4 (Reb1 C2)	Reb1 C2

Figure 5C (top) – For Cluster 1: B (All nucleosomes), C (Reb1-bound nucleosomes)
Figure 5C (bottom) – For Cluster 2: D (All nucleosomes), E (Reb1-bound nucleosomes)

Rpo21 worksheet information

A	B	C	D
Bin	H3/H4 (Rpo21)	Rpo21 Nuc	Rpo21 Std (Rpo21 Nuc)

Figure 6F – B (All nucleosomes), C (Pol II-bound nucleosomes), D (Pol II)

Srm1 worksheet information

A	B	C
Bin	H3/H4 (Srm1)	Srm1

Figure 6C – B (All nucleosomes), C (Srm1-bound nucleosomes)

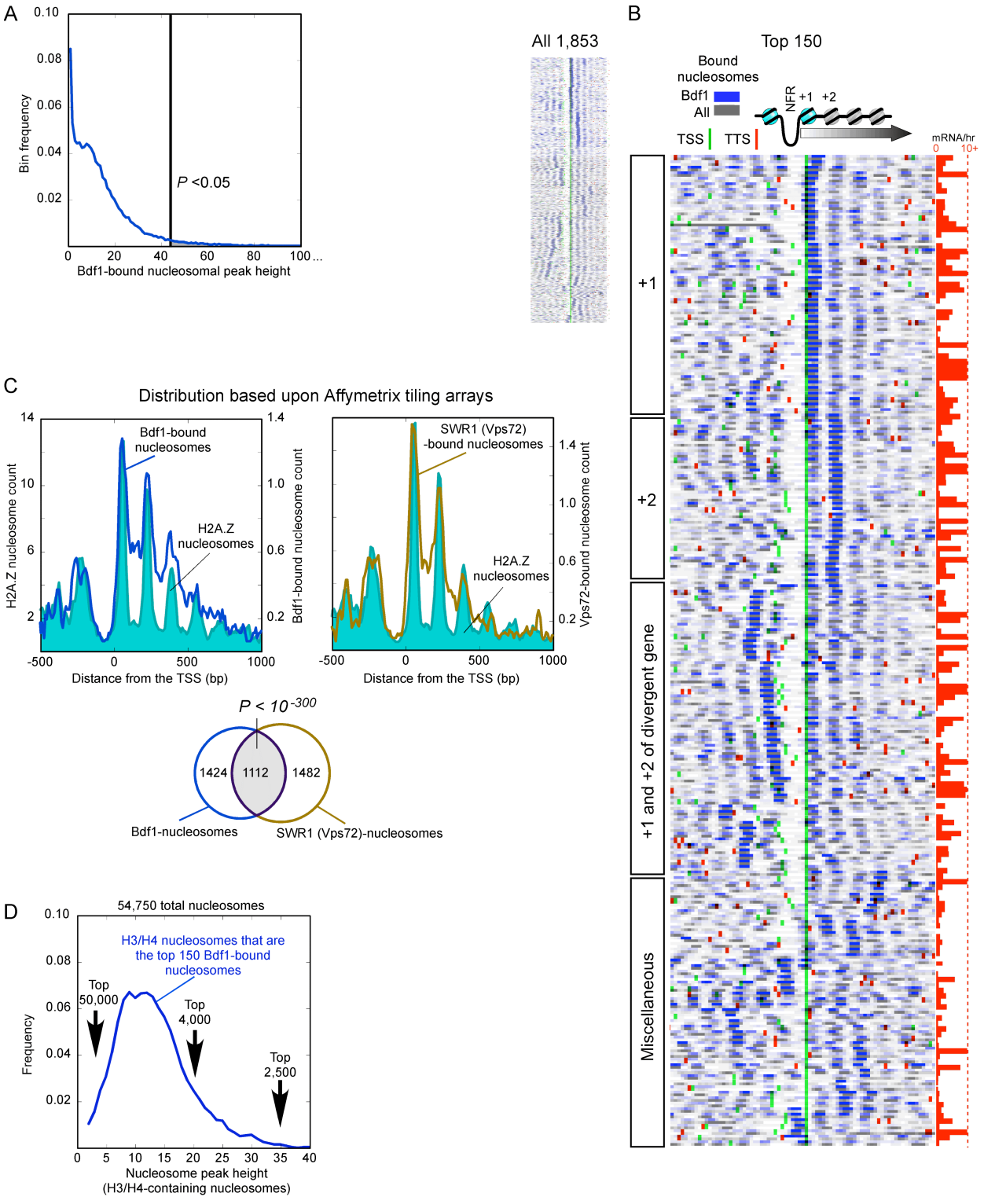
Supplementary Fig. 1 | Distribution of Bdf1-bound nucleosomes around the TSS.

A, Frequency distribution of tag counts for all nucleosomal calls in the Bdf1-bound nucleosomal data set. Bins sizes are in 1 tag increments, with the total bin frequency adding up to 1. Bin count >100 are not shown. All nucleosomal calls to the right of the vertical line are considered to be statistically significant.

B, Distribution of tag counts at genes that are located within 1 kb of the top 150 Bdf1-bound nucleosomes (or all 1853 significant bound nucleosomes). Each pair of rows represents a gene in which the tag counts for Bdf1-bound nucleosomes (blue) or all (H3 and H4-containing) (Mavrich et al., 2008) nucleosomes (gray) are binned in 10 bp increments as to their distance to the TSS. Bins counts were smoothed using a 3-bin moving average. Deeper color indicates higher bin counts. Bdf1-bound nucleosomal patterns were arranged using K-means. Transcript start (TSS) and stop (TTS) sites are shown on the plot using green and red bars, respectively. Transcription frequencies (Holstege et al., 1998) are shown as mRNA/hr (red horizontal bars), with the red dashed line representing 10 mRNA/hr.

C, Hybridization-based assays verify that Bdf1-nucleosome interactions and Vps-72 nucleosome interactions are enriched at the +1 and +2 nucleosomes. The raw ChIP material used for SOLiD sequencing was hybridized to Affymetrix high-density (5 bp spacing) whole genome tiling arrays. Bdf1-bound (blue trace) or Vps72-bound (gold trace) nucleosome midpoint locations were then determined using MAT (Johnson et al., 2006), and plotted as a smoothed distribution (10 bp bin, 3-bin moving average) relative to the TSS (David et al., 2006) at all genes that contain a Bdf1-interacting nucleosome (P <0.05, left panel) or Vps-72-interacting nucleosome (P <0.05, right panel). The cyan filled plot reports the distribution of H2A.Z-containing nucleosomes for the same collection of genes represented in each panel (Albert et al., 2007), in which nucleosomes were called in a similar manner but using GeneTrack software (Albert et al., 2008). Note that in both cases the analysis calls nucleosomes via detection of local hybridization or tag maximum, regardless of overall signal intensity relative to the rest of the genome. The analysis works well for identifying nucleosomes, but does not consider occupancy level compared to the rest of the genome. As a result there tends to be an over-emphasis of nucleosomes that represent contaminating background nucleosomes or low factor occupancy, since for the factor-nucleosome interaction assay they represent the vast majority of all detectable nucleosomes. Bin counting was stopped 300 bp before another TSS or TTS was reached, and then normalized to the number of genes counted in that bin. The Venn diagram indicates the intersection of Bdf1-bound and Vps72-bound nucleosomes (as opposed to the intersection of genes as in Fig. 2D).

D, Frequency distribution of peak heights that were determined byGeneTrack (Albert et al., 2008). Peak height provides a measure of positioning strength. If the nucleosomes that were selected as the top 150 Bdf1-bound nucleosomes were amongst the most highly position in the genome, then their distribution of peak heights (in the H3/H4 dataset of all nucleosomes) are expected to be far to the right of the top 2,500 positioned nucleosomes in genome (indicated by the arrow). This is not observed. Instead their distribution of peak heights (or positioning strength) is about average.



Supplementary Figure 1

Supplementary Fig. 2 | Distribution of Rap1-bound nucleosomes around the TSS

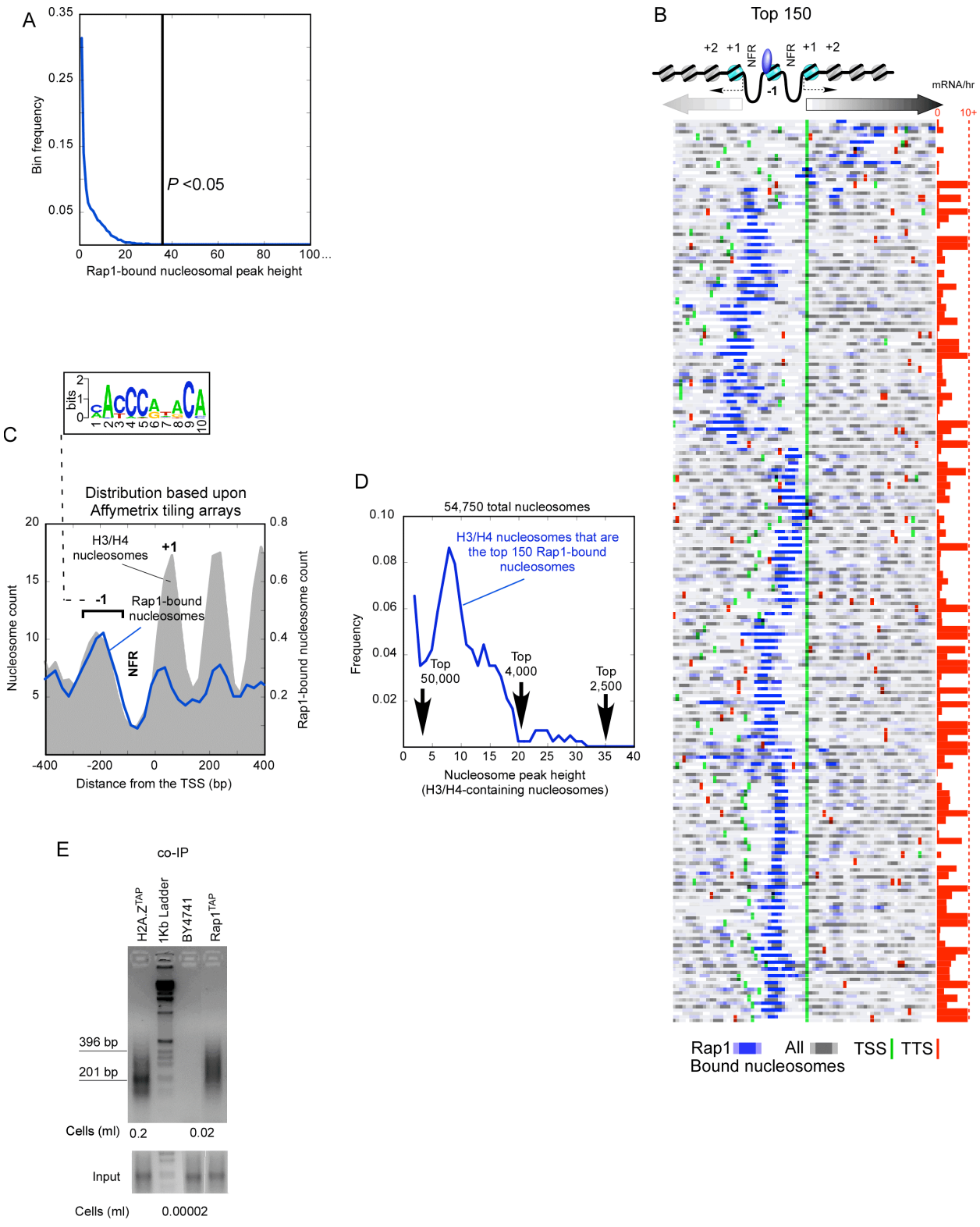
A, Frequency distribution of tag counts for all nucleosomal calls in the Rap1-bound nucleosomal data set. Bin sizes are in 1-tag increments, with the total adding up to 1. Bin counts >100 are not shown. All nucleosomal calls to the right of the vertical line are considered to be statistically significant.

B, Distribution of tag counts at genes that are located within 1 kb of the top 150 Rap1-bound nucleosomes. Each pair of rows represents a gene in which the tag counts for Rap1-bound nucleosomes (blue) or all (H3 and H4-containing) (Mavrich et al., 2008) nucleosomes (gray) are binned in 10 bp increments as to their distance to the TSS. Bins counts were smoothed using a 3-bin moving average. Deeper color indicates higher bin counts. Rap1-bound nucleosomal patterns were arranged using K-means. Transcript start (TSS) and stop (TTS) sites are shown on the plot using green and red bars, respectively. Transcription frequencies (Holstege et al., 1998) are shown as mRNA/hr (red horizontal bars), with the red dashed line representing 10 mRNA/hr.

C, Hybridization-based assays verify that Rap1-nucleosome interactions are enriched at the -1 nucleosomes. The raw ChIP material used for SOLiD sequencing was hybridized to Affymetrix high-density (5 bp spacing) whole genome tiling arrays. Rap1-bound nucleosome midpoint locations (blue trace) were then determined using MAT (Johnson et al., 2006), and plotted as a smoothed distribution (10 bp bin, 3-bin moving average) relative to the TSS (David et al., 2006) at all genes that contain a Rap1-interacting nucleosome (P <0.05). The gray filled plot reports the distribution of all (H3/H4-containing) nucleosomes for the same collection of genes represented in each panel (Albert et al., 2007), in which nucleosomes were called in a similar manner but using GeneTrack software (Albert et al., 2008). See note in Supplementary Fig. 1C regarding nucleosome calls vs. tag counts. Bin counting was stopped 300 bp before another TSS or TTS was reached, and then normalized to the number of genes counted in that bin. MEME analysis of Rap1-bound nucleosomal DNA produced the indicated Rap1 motif.

D, Frequency distribution of peak heights that were determined by GeneTrack (Albert et al., 2008). Peak height provides a measure of positioning strength. If the nucleosomes that were selected as the top 150 Rap1-bound nucleosomes were amongst the most highly position in the genome, then their distribution of peak heights (in the H3/H4 dataset of all nucleosomes) is expected to be far to the right of the nucleosome peak height that is marked #2,500 in rank order. Peak heights of other ranks are also shown with arrows. This was not observed. Instead their distribution of peak heights (or positioning strength) is about average.

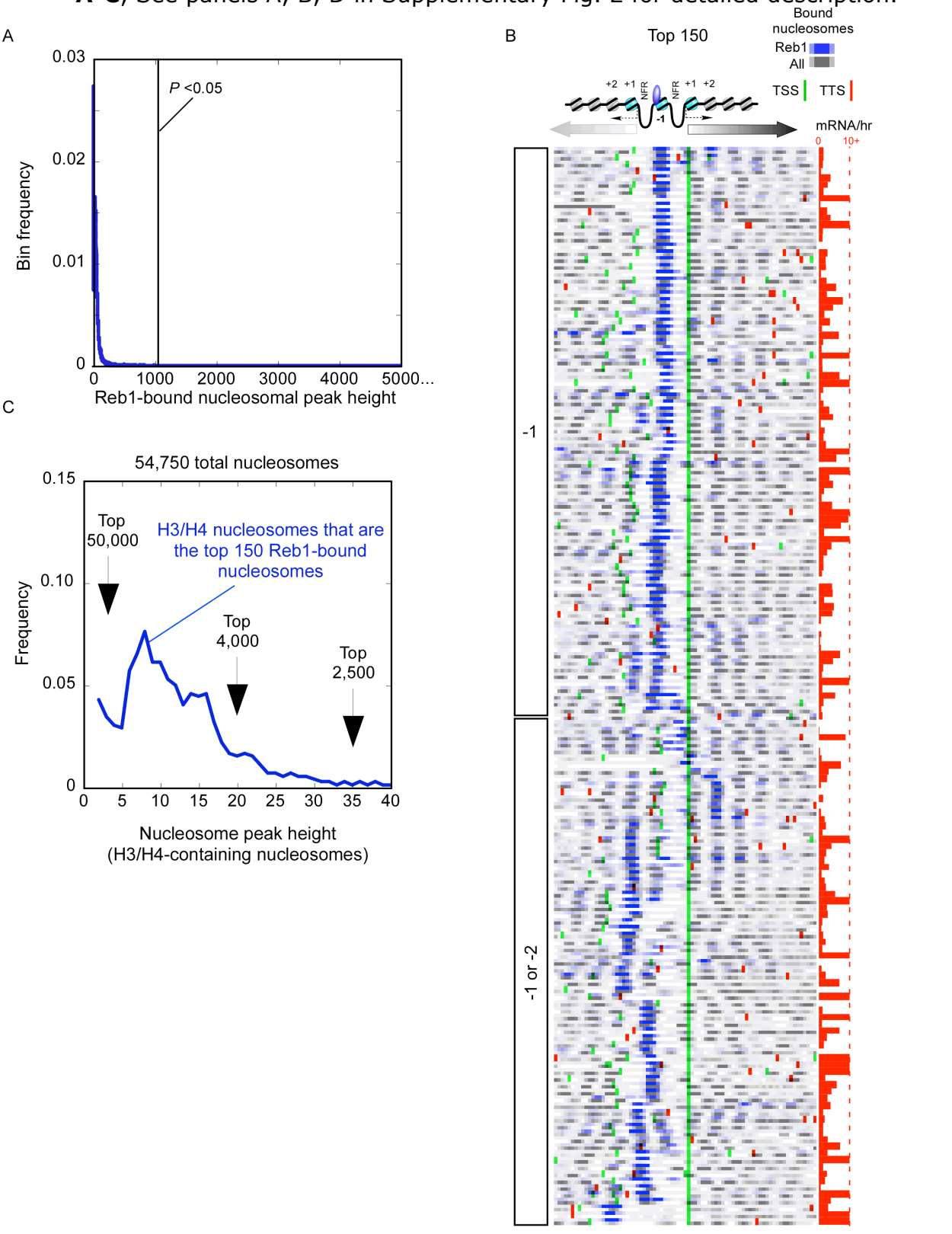
E, Co-immunoprecipitation of Rap1-bound nucleosomes. Reactions were performed as described for Bdf1 (Fig. 3A). Although some di-nucleosomes were detected, the ratio of di- to mono-nucleosomes was the same as input, indicating that there was not selective enrichment of Rap1-bound di-nucleosomal DNA.



Supplementary Figure 2

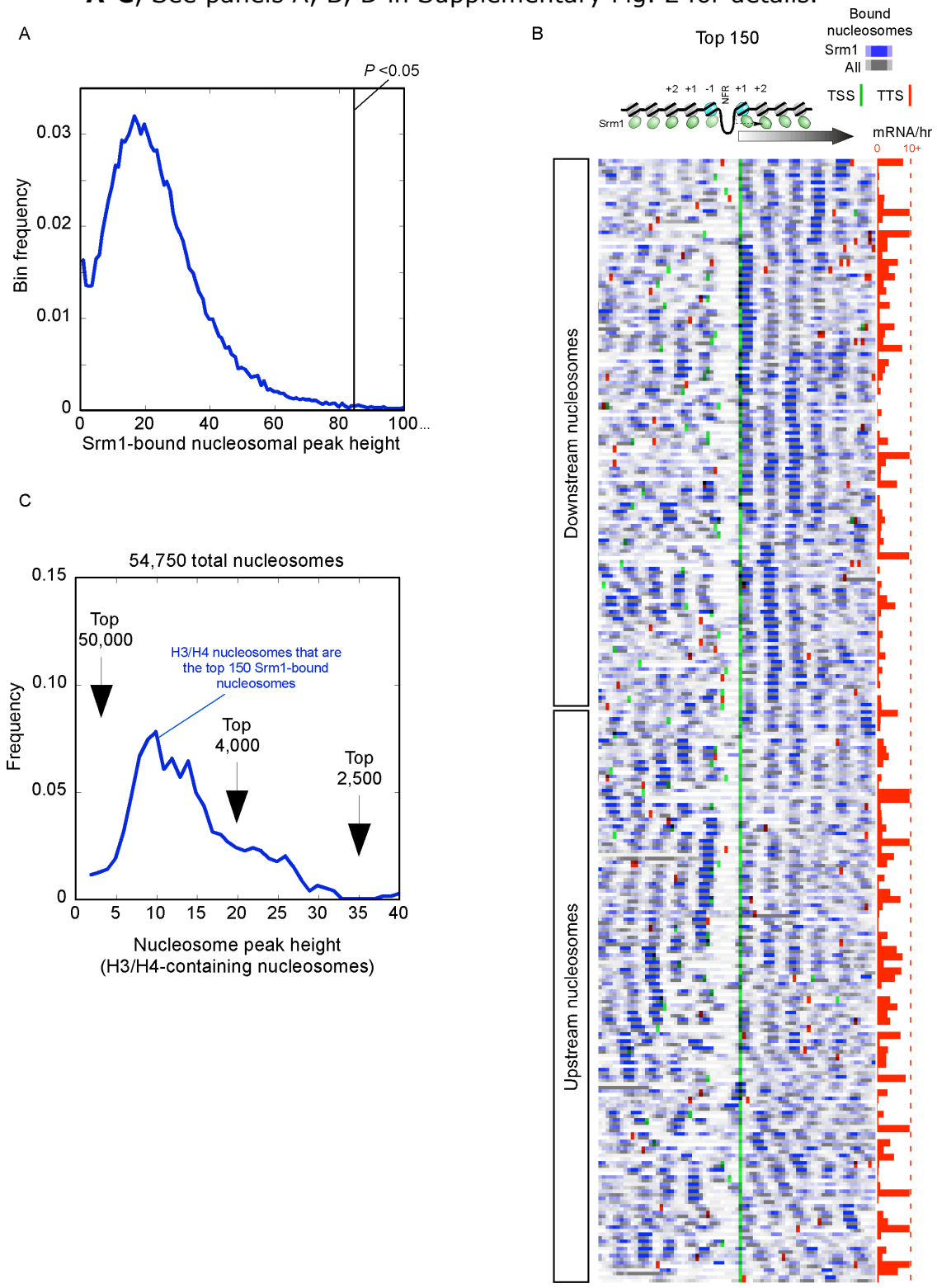
Supplementary Fig. 3 | Distribution of Reb1-bound nucleosomes around the TSS

A-C, See panels A, B, D in Supplementary Fig. 2 for detailed description.



Supplementary Fig. 4 | Distribution of Srm1-bound nucleosomes around the TSS

A-C, See panels A, B, D in Supplementary Fig. 2 for details.



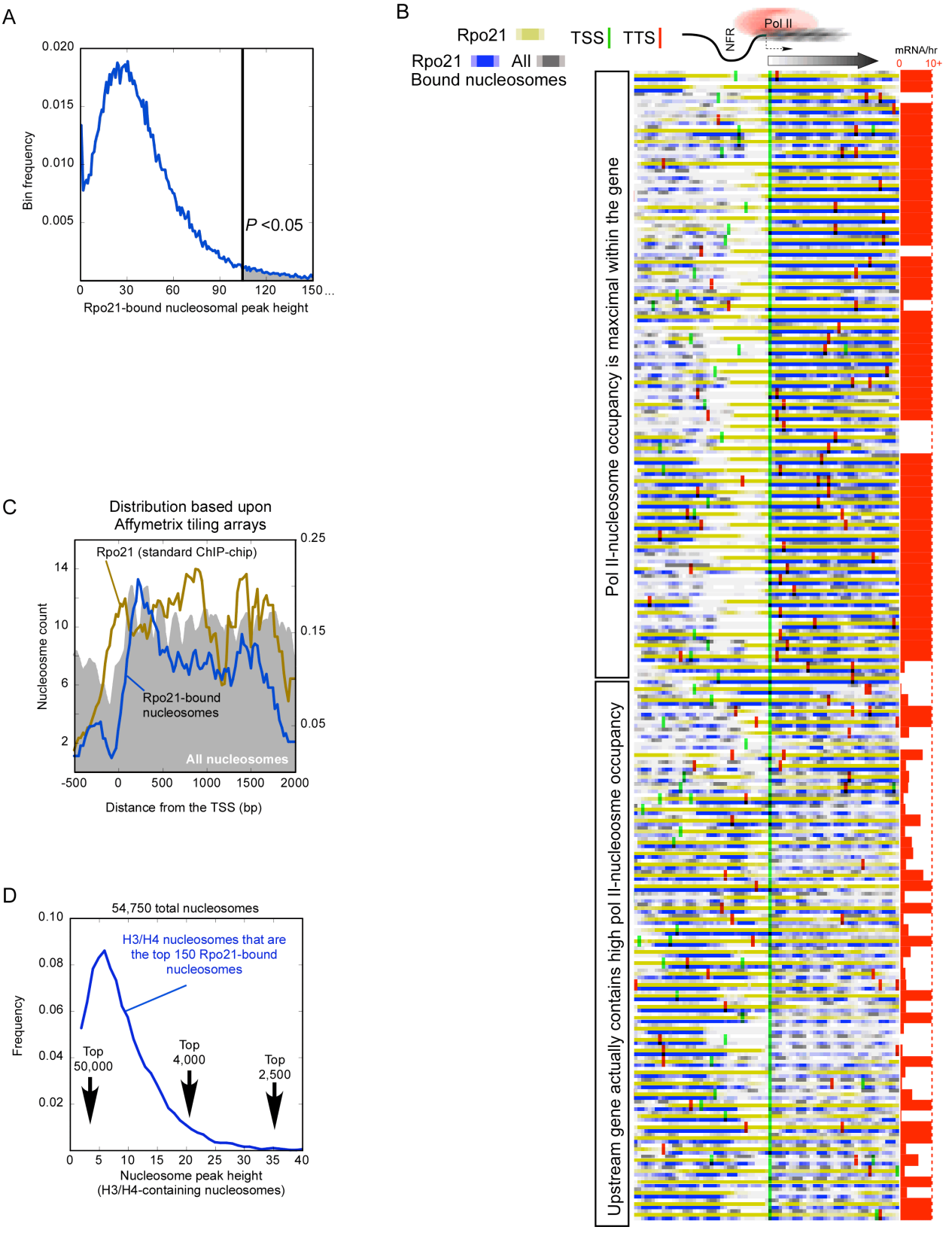
Supplementary Fig. 5 | Distribution of Pol II-crosslinked nucleosomes around the TSS

A, Frequency distribution of tag counts for all nucleosomal calls in the Rpo21-bound nucleosomal data set. Bin sizes are in 1-tag increments, with the total adding up to 1. Bin counts >150 are not shown. All nucleosomal calls to the right of the vertical line are considered to be statistically significant, although they tend to be highly delocalized.

B, Distribution of tag counts at genes that are located within 1 kb of the top 150 Rpo21-bound nucleosomes. Each set of three rows represents a gene in which the tag counts for Rpo21-bound nucleosomes (blue), all (H3 and H4-containing) (Mavrich et al., 2008) nucleosomes (gray), or standard sonication based distribution of total Pol II based upon ChIP-chip (gold), are binned in 10 bp increments as to their distance to the TSS. Bins counts were smoothed using a 3-bin moving average. Deeper color indicates higher bin counts. Rpo21-bound nucleosomal patterns were arranged using K-means. Transcript start (TSS) and stop (TTS) sites are shown on the plot using green and red bars, respectively. Transcription frequencies (Holstege et al., 1998) are shown as mRNA/hr (red horizontal bars), with the red dashed line representing 10 mRNA/hr. The lower half of the dataset reflects genes whose upstream neighbor actually possesses at least one of the top scoring Rpo21-nucleosome peaks (and thus can be ignored for the purposes of this study).

C, Hybridization-based assays verify that Pol II (Rpo21)-nucleosome interactions result in displacement from canonical positions. The raw ChIP material used for SOLiD sequencing was hybridized to Affymetrix high-density (5 bp spacing) whole genome tiling arrays. Rpo21-bound nucleosome midpoint locations (blue trace) were then determined using MAT (Johnson et al., 2006), and plotted as a smoothed distribution (10 bp bin, 3-bin moving average) relative to the TSS (David et al., 2006) at all genes that are in the top 25 percentile of transcription frequency (Holstege et al., 1998). The gold trace reports the distribution of all Rpo21 at the same genes (sonication based ChIP-chip). The gray filled plot reports the distribution of H3/H4-containing nucleosomes for the same collection of genes represented in each panel (Albert et al., 2007), in which nucleosomes were called in a similar manner but using GeneTrack software (Albert et al., 2008). See note in Supplementary Fig. 1C. Bin counting was stopped 300 bp before another TSS or TTS was reached, and then normalized to the number of genes counted in that bin.

D, Frequency distribution of H3/H4 peak heights that were determined by GeneTrack (Albert et al., 2008). Peak height provides a measure of positioning strength. If the nucleosomes that were selected as the top 150 Rpo21-bound nucleosomes were amongst the most highly position in the genome, then their distribution of peak heights (in the H3/H4 dataset of all nucleosomes) is expected to be far to the right of the nucleosome peak height that is marked #2,500 in rank order. Peak heights of other ranks are also shown with arrows. This was not observed. Instead their distribution of peak heights (or positioning strength) is about average.



Supplementary Figure 5

Supplementary References

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Albert, I., Wachi, S., Jiang, C., and Pugh, B.F. (2008). GeneTrack--a genomic data processing and visualization framework. Bioinformatics *24*, 1305-1306.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA *103*, 5320-5325.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717-728.

Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci U S A *103*, 12457-12462.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res *18*, 1073-1083.