Supplementary material

Additional figures for "The yeast exoribonuclease Xrn1 and associated factors modulate RNA polymerase II processivity in 5' and 3' gene regions" by Jonathan Fischer, Yun S. Song, Nir Yosef, Julia di Iulio, L. Stirling Churchman, and Mordechai Choder.

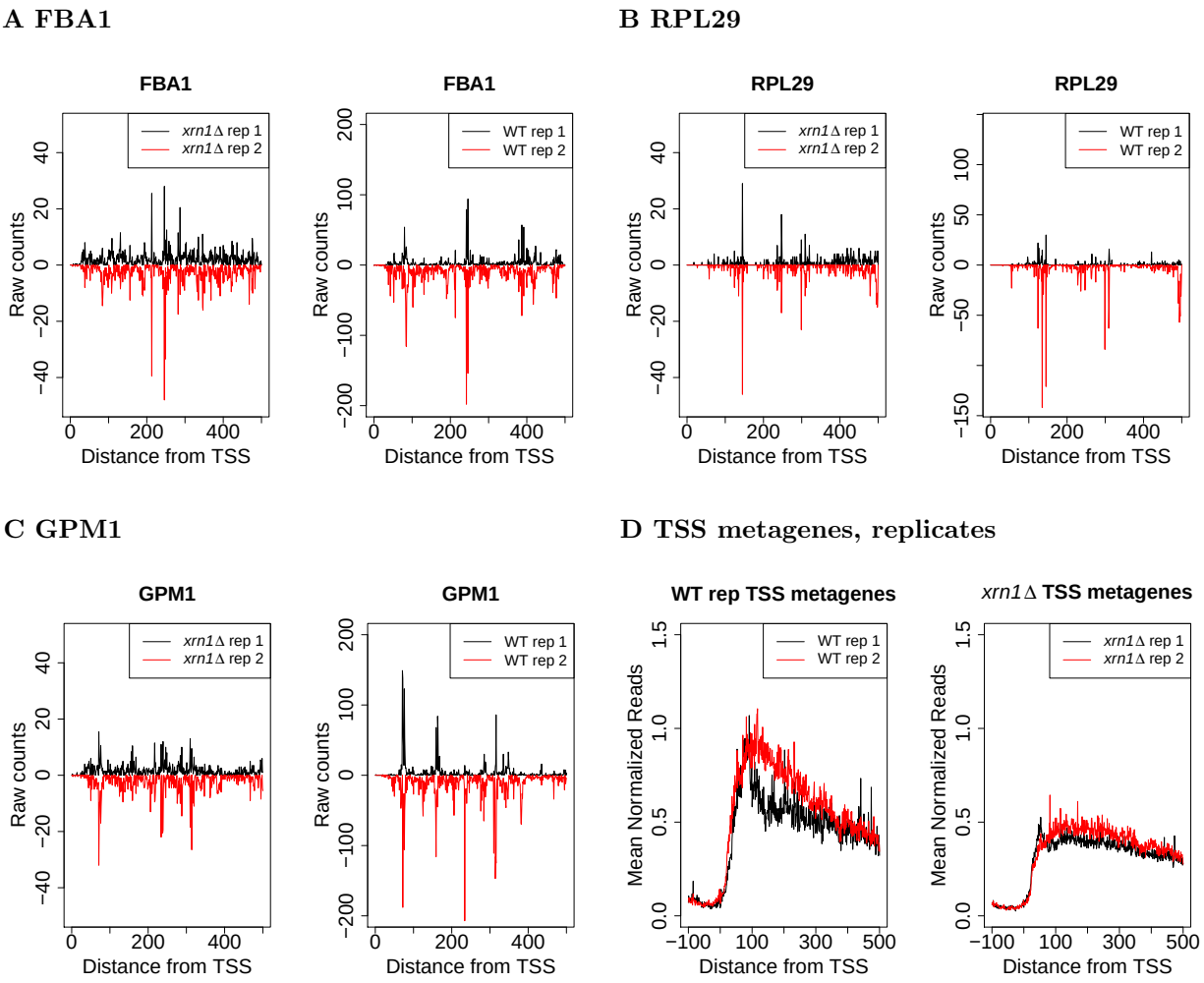
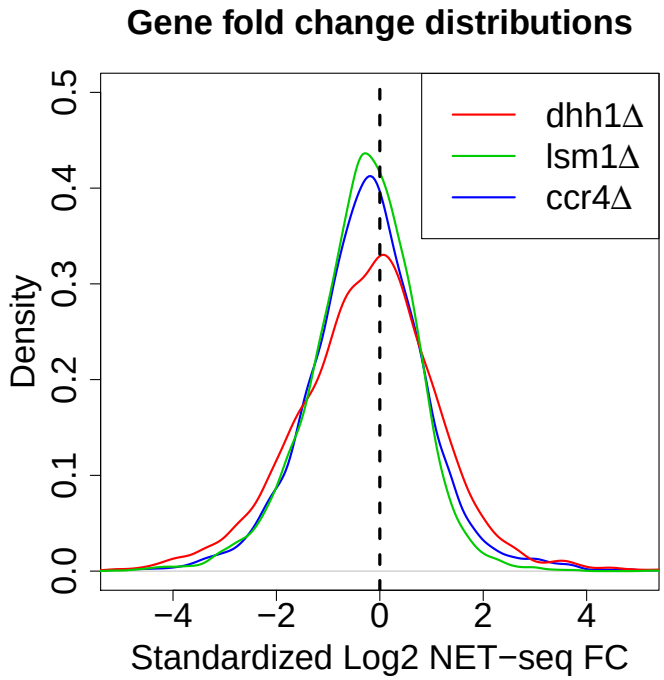


Fig. S1 Comparison of WT and $xrn1\Delta$ replicates. (**A**), (**B**), (**C**): Plots of raw NET-seq reads for the first 500 bp of 3 highly expressed genes in replicate pairs; (**A**) - $FBA1$, (**B**) - $RPL29$, (**C**) - $GPM1$. (**D**): Plots of metagenes for the region beginning 100 bp before TSS and ending 500 bp after. Reads within replicate strains were aggregated at each location across all genes and averaged. They were then normalized using the procedure described in the Methods section of the primary manuscript.

A NET-seq FCs, *dhh*1Δ, *lsm*1Δ, *ccr*4Δ



B Elongation efficiency FCs, *xrn*1Δ **and** *rpb*4Δ

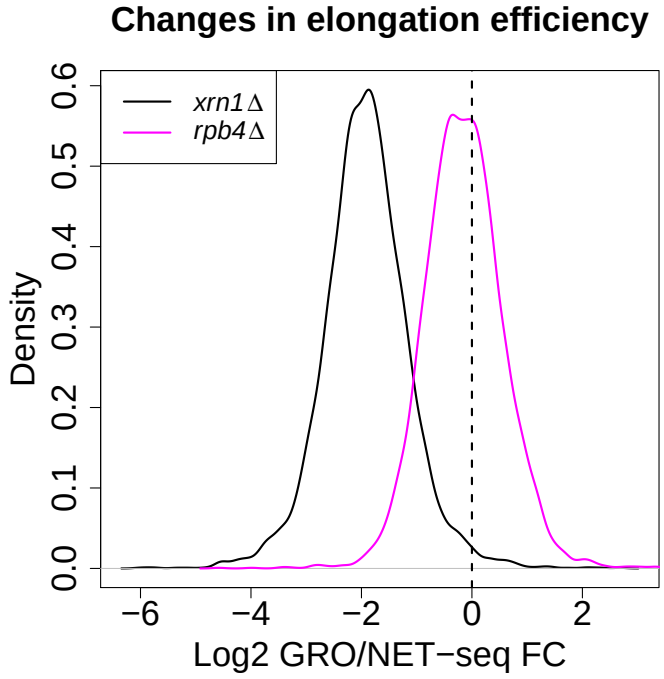
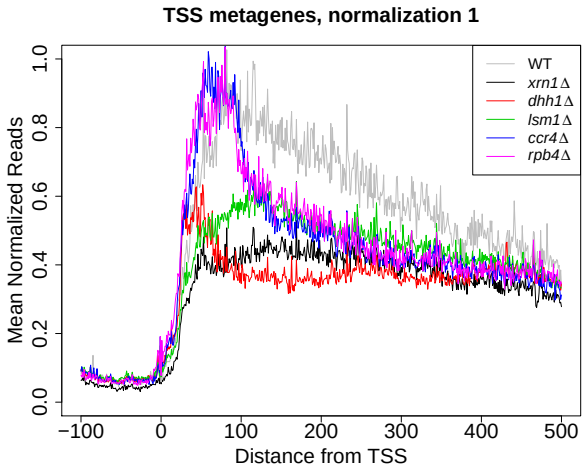
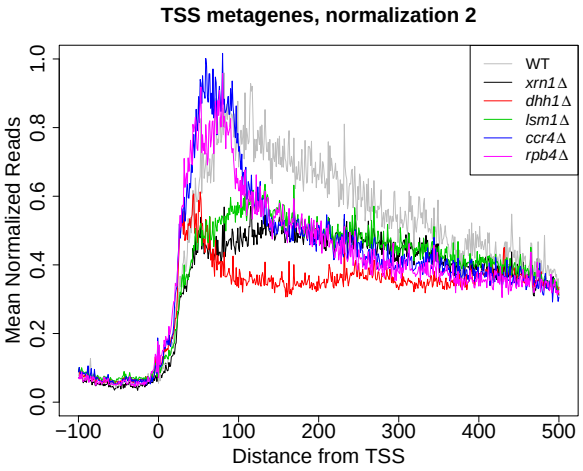


Fig. S2 Fold changes in NET-seq Pol II occupancy and GRO/NET-seq ratios. (**A**) Same procedure as Fig. 1A but for the displayed mutants. (**B**) Elongation efficiencies were estimated by taking the ratio of GRO to NET-seq values for all genes in WT, *xrn*1Δ, and *rpb*4Δ strains and then applying a log2 transformation. Fold changes were then computed by taking the difference of mutant elongation efficiencies to those in the WT for each of the two mutants in question. GRO data were only available for *xrn*1Δ and *rpb*4Δ.

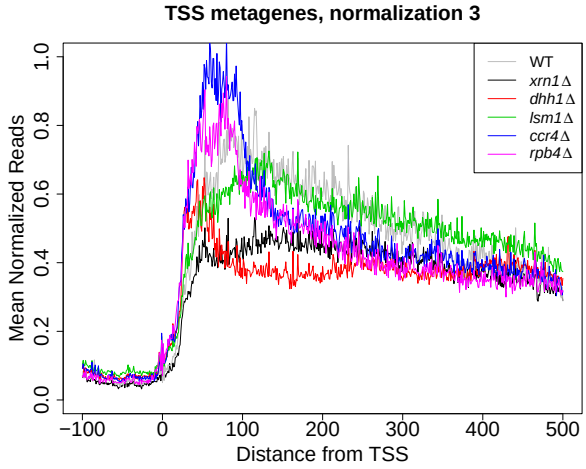
A Normalization 1



B Normalization 2



C Normalization 3



D Normalization 4

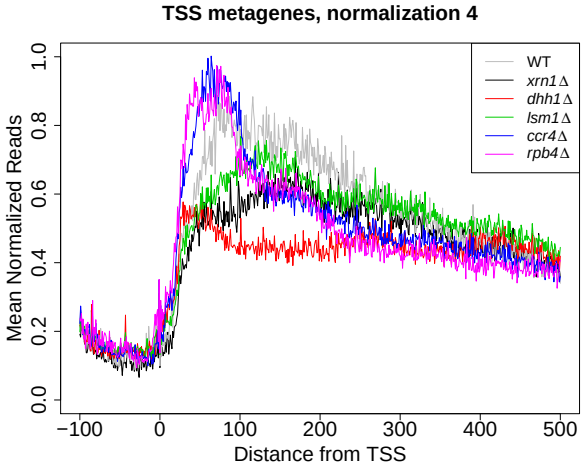


Fig. S3 Effect of different normalizations on TSS metagenes. Plots of TSS metagenes for each strain using different normalization procedures. The different normalization procedures, as well as method of metagene construction, are described fully in the Methods section. In brief, the normalizations were as follows: 1 - housekeeping genes were identified using GRO or cDTA and DESeq2's normalization procedure employed; 2 - DESeq2's procedure was applied to all genes; 3 - samples were divided by the sum of all reads in considered ORFs; 4 - the reads in each ORF were divided by the total reads in that ORF. These profiles were then averaged across ORFs. For 3 and 4, a common global rescaling was applied to the final profiles for all strains to faciliate visual comparison with the results of the other normalizations.

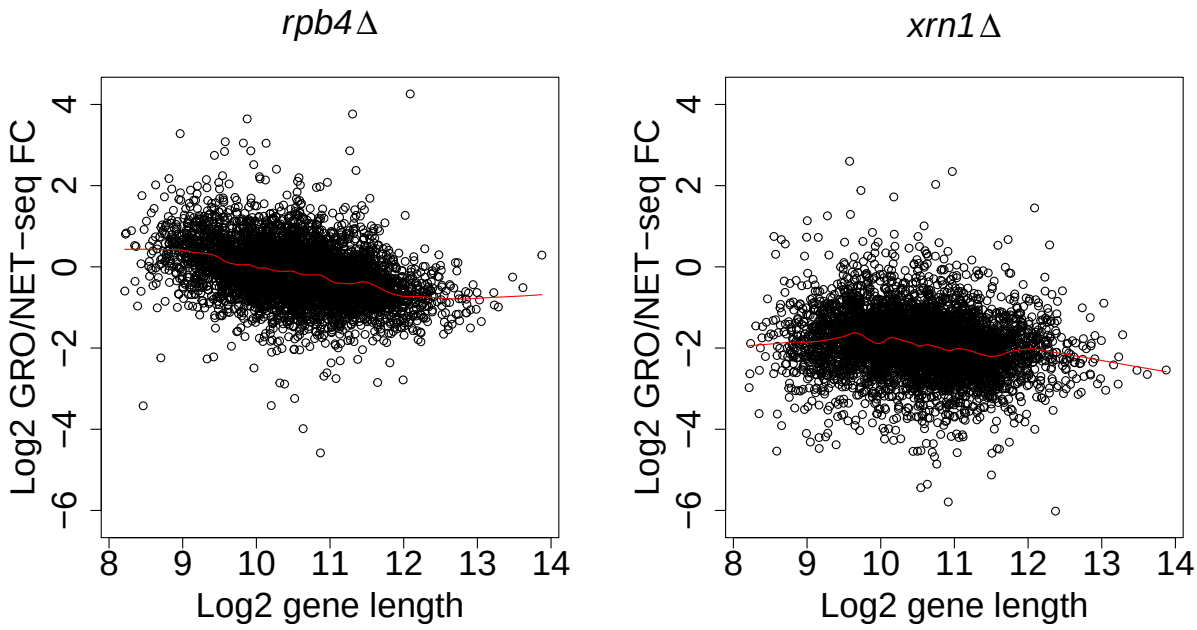


Fig. S4 Log2 fold changes in transcriptional efficiency as a function of gene length for *rpb4*Δ and *xrn1*Δ. The transcriptional efficiency for each gene was estimated as the ratio of GRO (1, 28) to NET-seq signal over whole genes. Red lines are LOWESS fits. GRO data are only available for *xrn1*Δ and *rpb4*Δ.

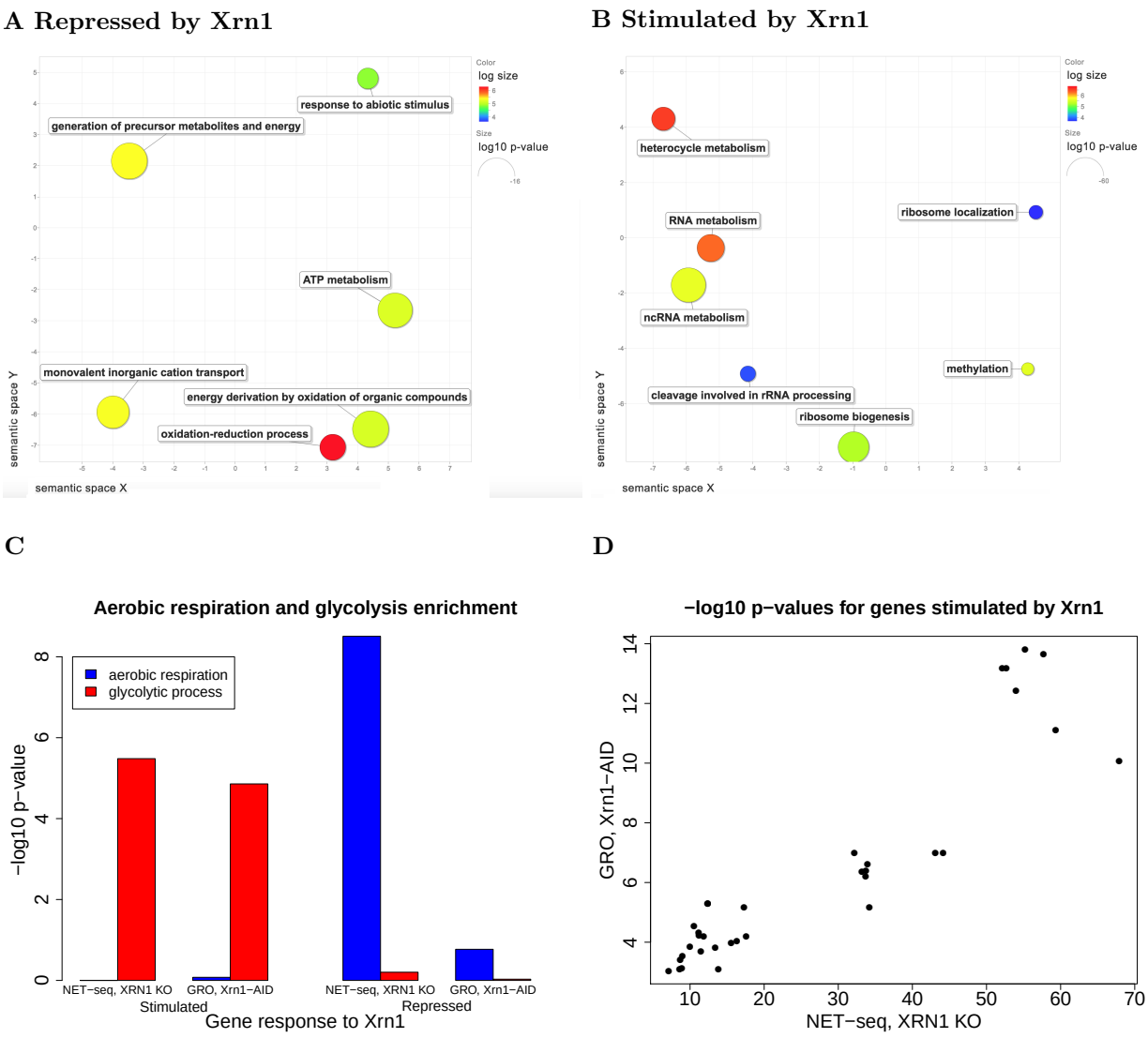
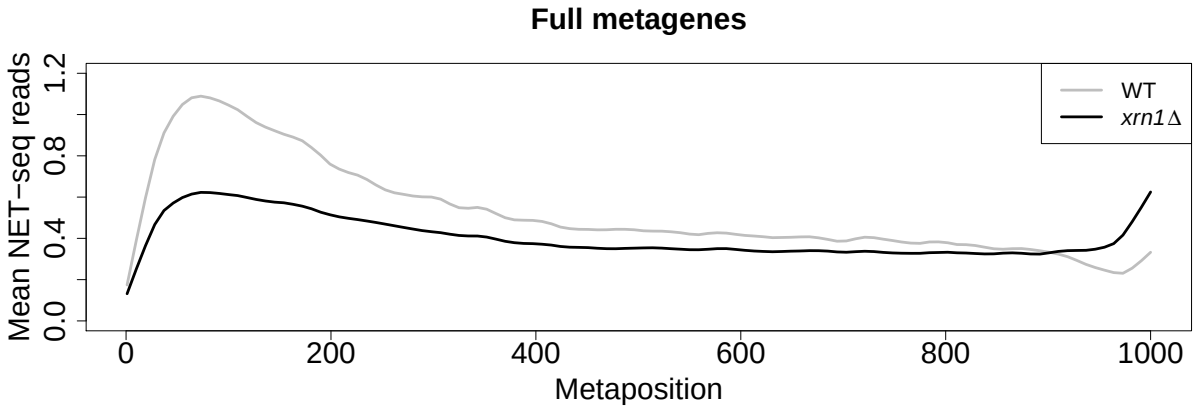
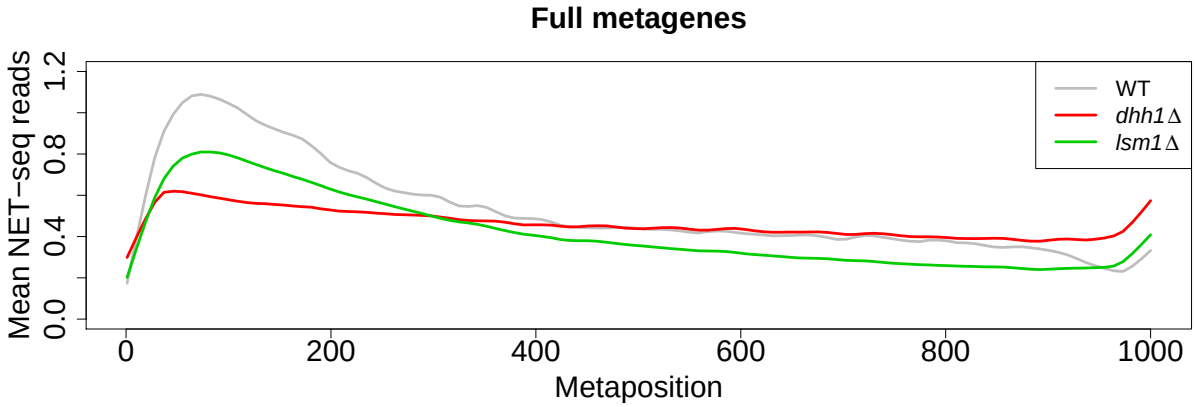


Fig. S5 GO enrichment analysis for *xrn*1Δ**.** (**A**), (**B**): Genes were ranked based on their fold changes and GO enrichment analysis was performed using GOrilla (84). Plotting was then done using REVIGO (98). Each circle represents an enriched GO term, the size of which indicates the p-value of enrichment. Circle colors represent the number of genes annotated to that GO term, and the distances between circles reflect the similarity between given GO terms (terms closer to one another in the plot have more genes in common). (**C**): Genes annotated to the aerobic respiration and glycolytic process GO terms were obtained from PANTHER (99) and enrichments p-values were computed using the minimum hypergeometric test as implemented in the mHG R package (83, 100). For the "NET-seq, XRN1 KO" bars, genes were sorted by their FCs in the *xrn*1Δ strain as in (**A**) and (**B**). For the "GRO, Xrn1-AID", genes were ranked by their FCs in the Xrn1 degradation experiment of (37). (**D**): Scatterplot of *p*-values which were enriched among Xrn1-stimulated genes in both experiments. Only enriched GO terms are returned by (84).

A *xrn*1Δ



B *dhh*1Δ **and** *lsm*1Δ



C *ccr*4Δ **and** *rpb*4Δ

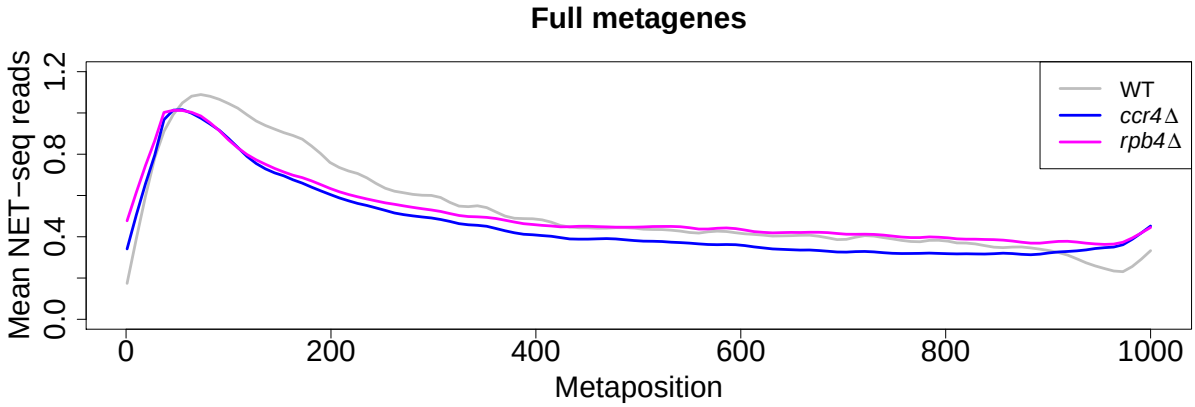


Fig. S6 Comparison of normalized full-body metagenes using all genes. Normalized reads were aggregated from TSS to PAS and were then re-scaled to lengths of 1000 bp. The read counts corresponding to the new "metapositions" were averaged to yield a picture of Pol II occupancy along whole gene bodies using all genes regardless of length. Different panels show comparisons between WT and the indicated deletion strains.

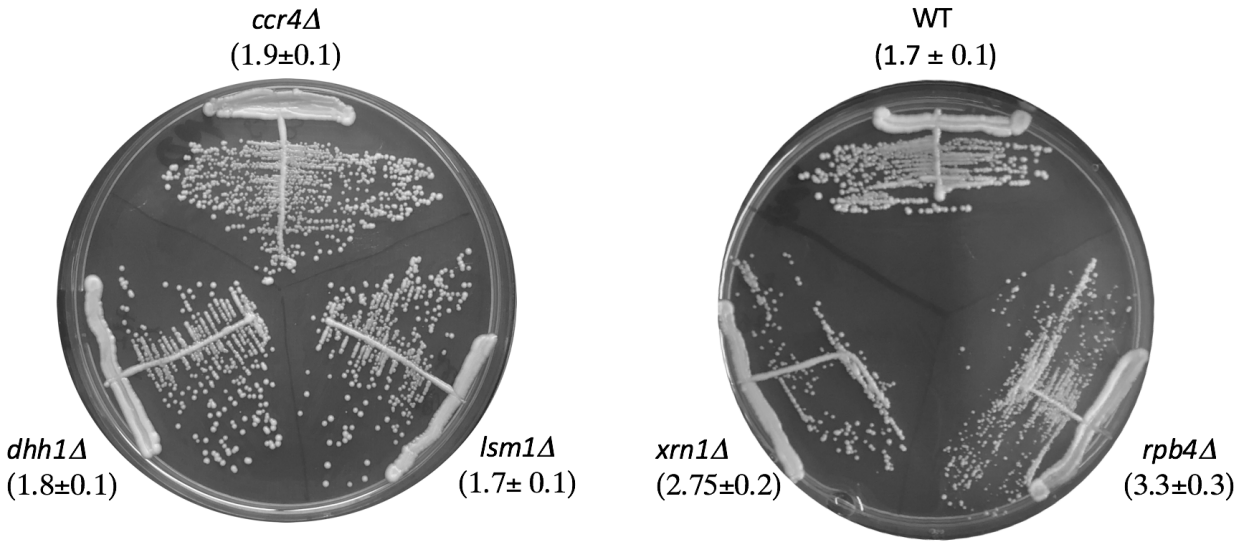
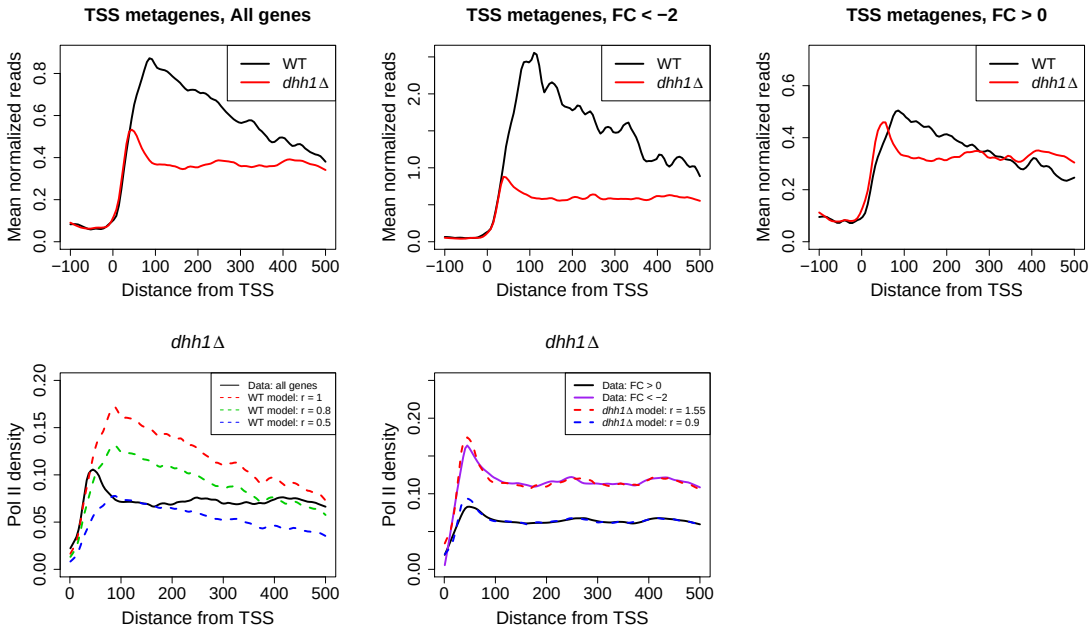
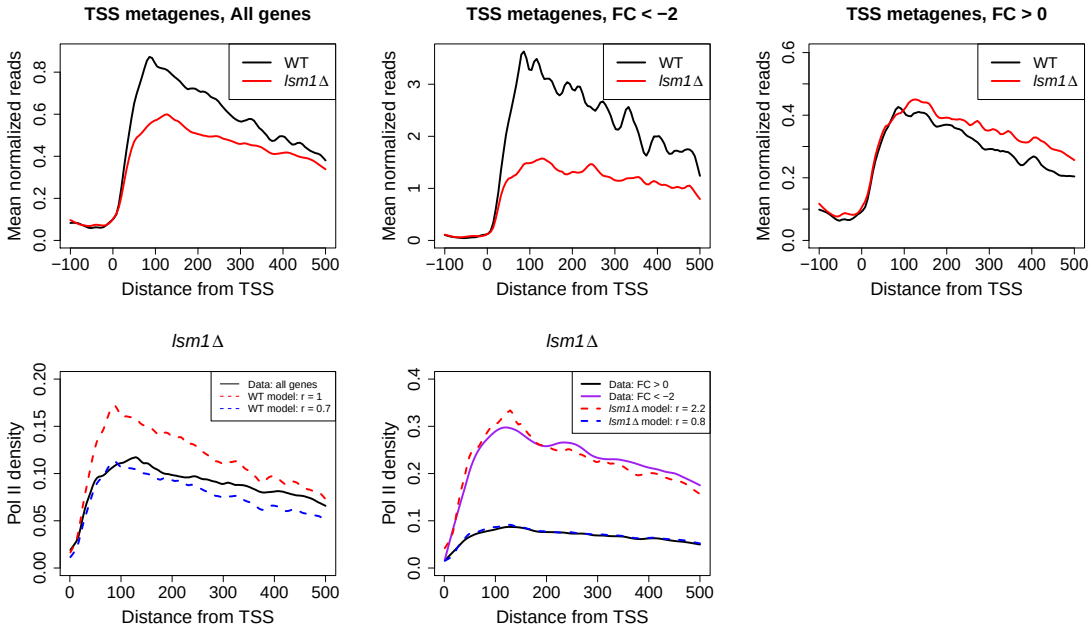


Fig. S7 Growth of the studied yeast strains. Optimally proliferating cells were streaked on plates with rich medium (YPD). The plates were incubated at 30 ^{o}C for 2 days before photos were taken. In parallel, the doubling time was determined in liquid YPD medium. The results are indicated in brackets and expressed as hours.

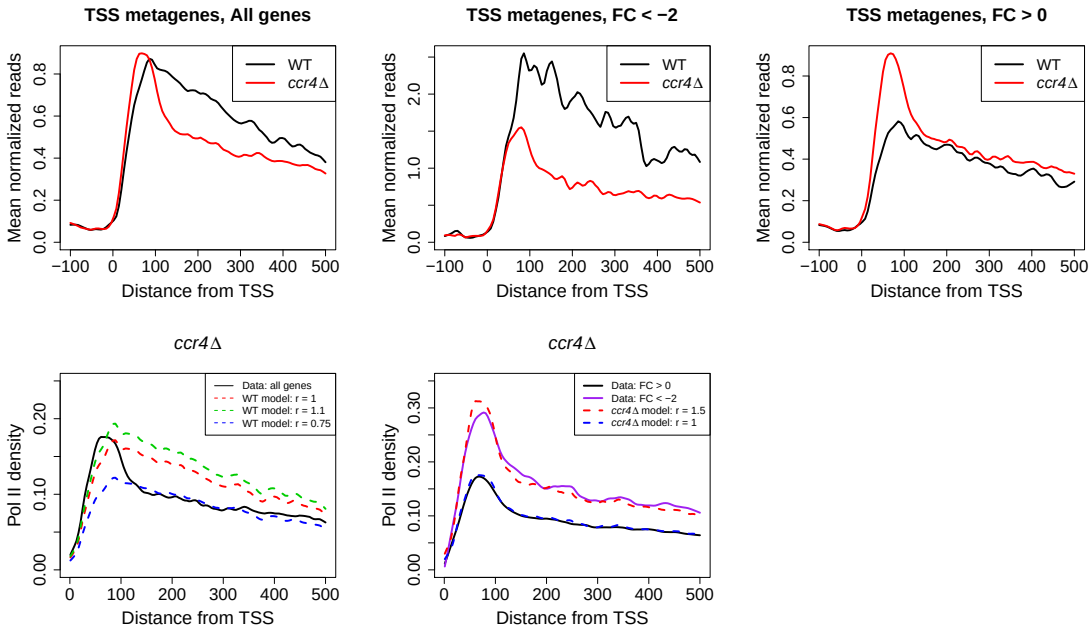
A $dhh1\Delta$



B $lsm1\Delta$



C *ccr4*Δ



D *rpb4*Δ

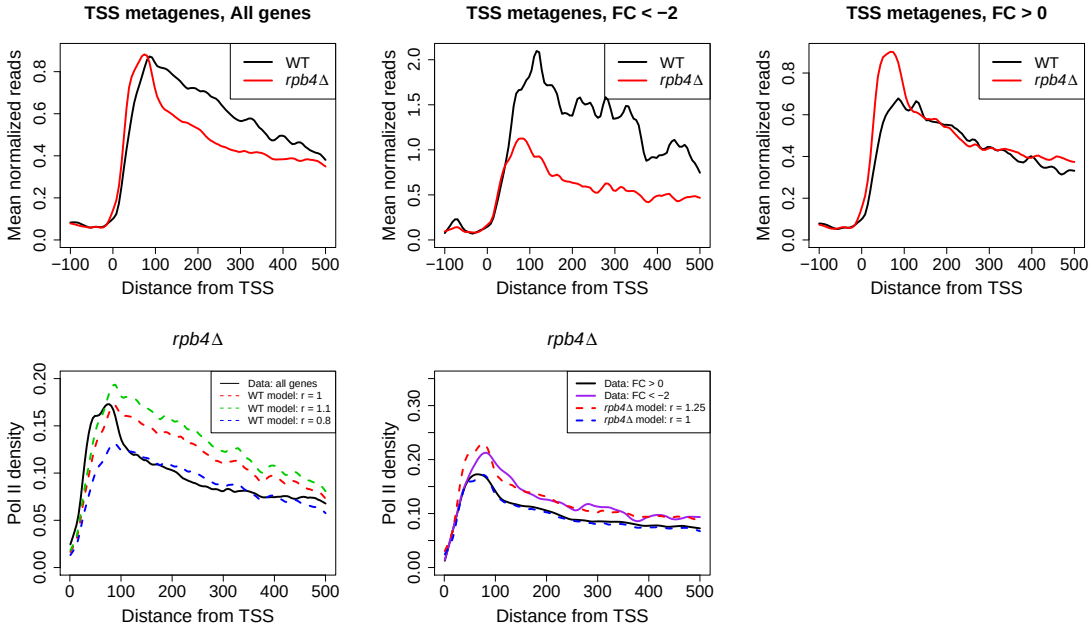
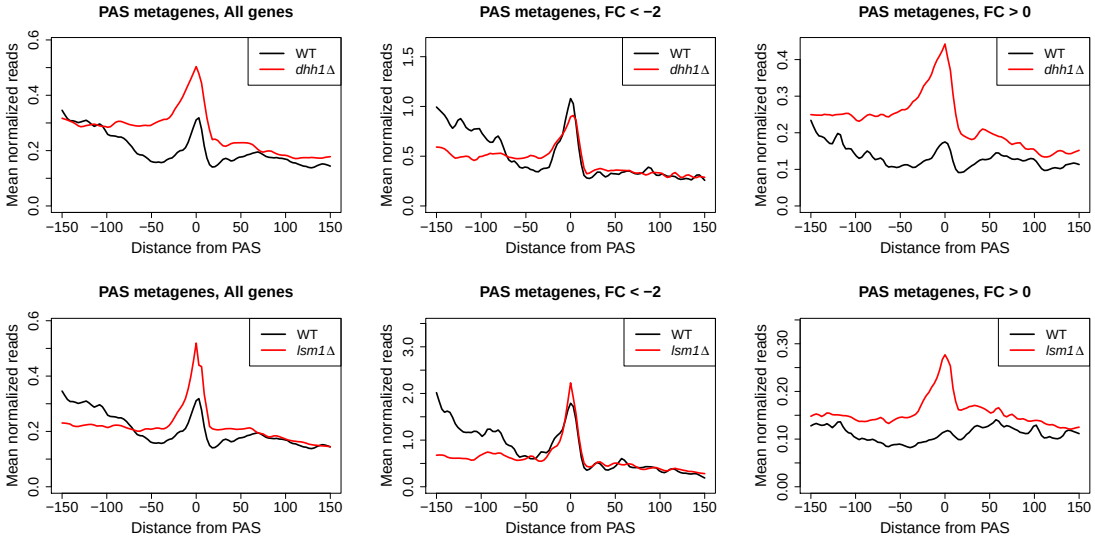


Fig. S8 Normalized metagene profiles near TSS for other DF KOs. Same procedure as Fig. 3A but for (**A**) *dhh*1Δ, (**B**) *lsm*1Δ, (**C**) *ccr*4Δ, and (**D**) *rpb*4Δ. Note that the axes are on different scales to facilitate comparison of profile shapes within each panel.

A *dhh*1Δ **and** *lsm*1Δ



B *ccr*4Δ **and** *rpb*4Δ

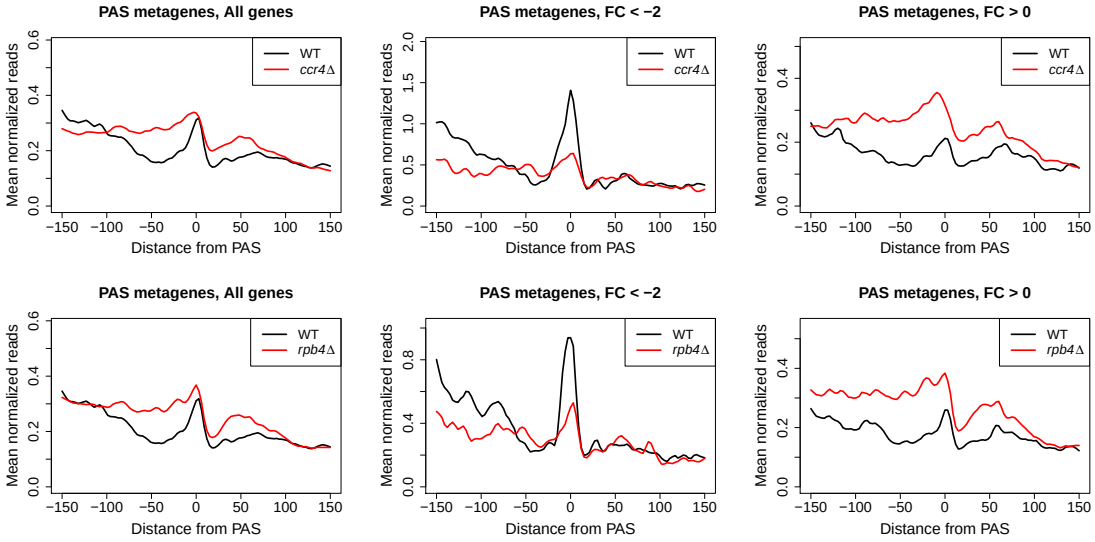


Fig. S9 Normalized metagene profiles near PAS for other DF KOs. Same procedure as Fig. 5A but for (**A**) *dhh*1Δ, *lsm*1Δ, and (**B**) *ccr*4Δ, *rpb*4Δ. Note that the axes are on different scales to facilitate comparison of profile shapes within each panel.

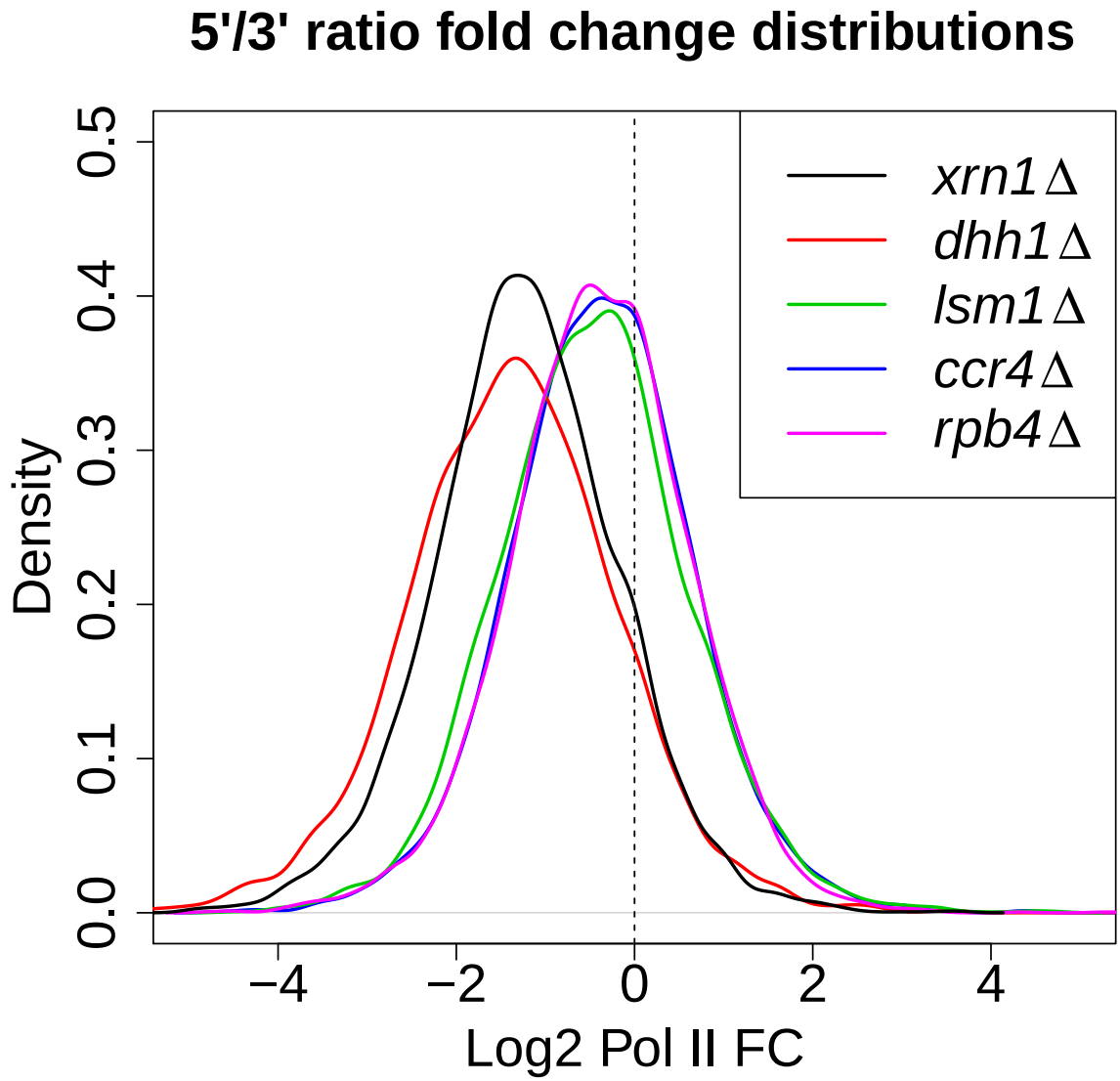
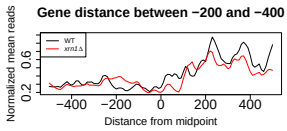
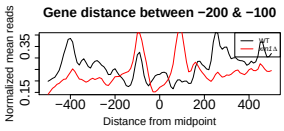
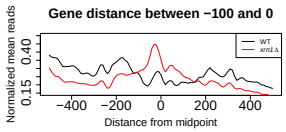


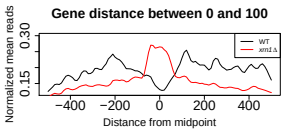
Fig. S10 Fold changes in $5'/3'$ occupancy ratios. Reads were aggregated in the first 300 bp downstream of TSS and upstream of PAS. The ratio of $5'$ and $3'$ reads was then taken for each gene in each strain, and log2 fold changes were computed with respect to WT samples.

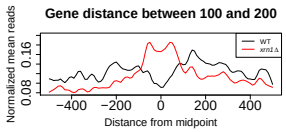
A $xrn1\Delta$

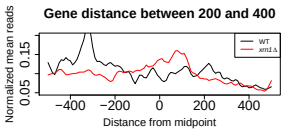




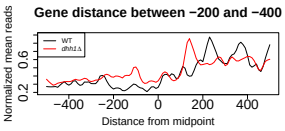


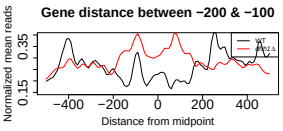


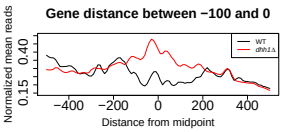


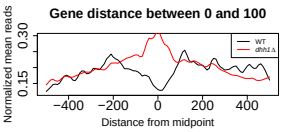


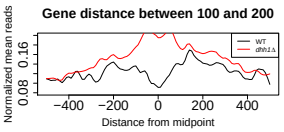
B $dhh1\Delta$

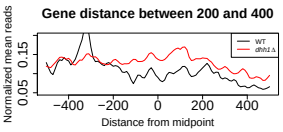




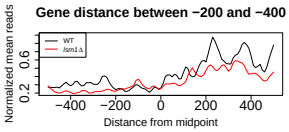


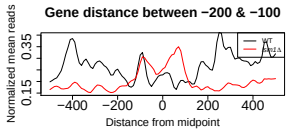


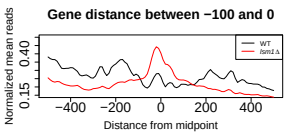


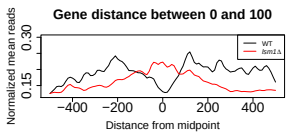


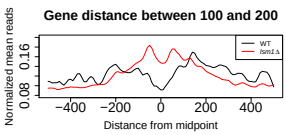
C $lsm1\Delta$

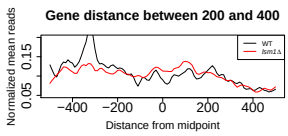




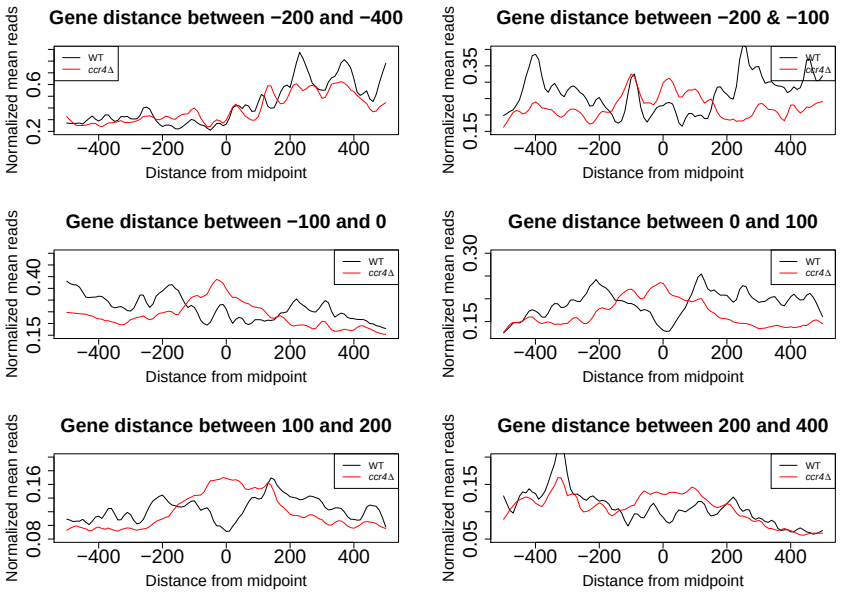








D *ccr4*Δ



E *rpb4*Δ

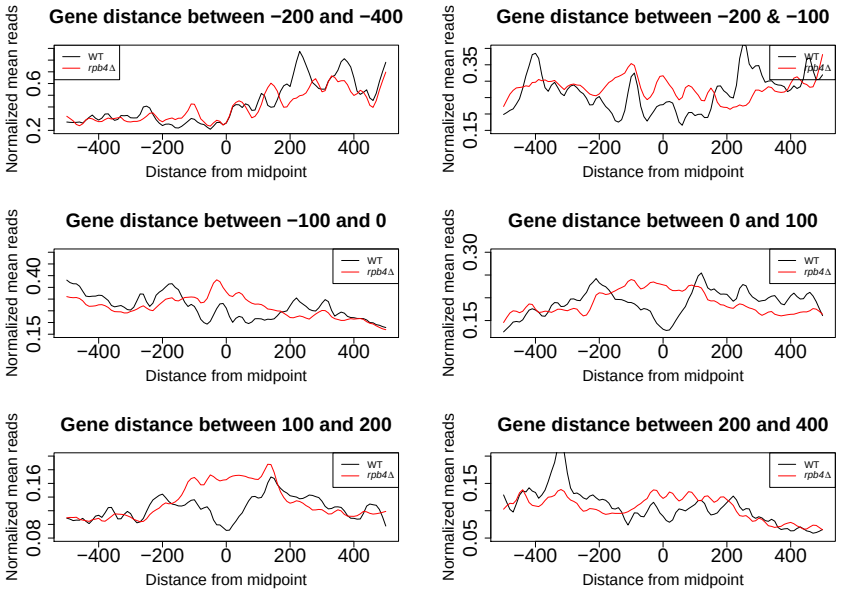


Fig. S11 Metagenes for convergent gene pairs after binning by distance between genes. The distance between genes was computed by taking the difference between the locations of the PAS on the plus and minus strands, respectively. Convergent gene pairs were then stratified into the six windows shown above based on these distances, where negative values indicate overlapping genes. Within each group, reads on both strands falling within 500 bp of a midpoint between genes were averaged at each position and plotted.

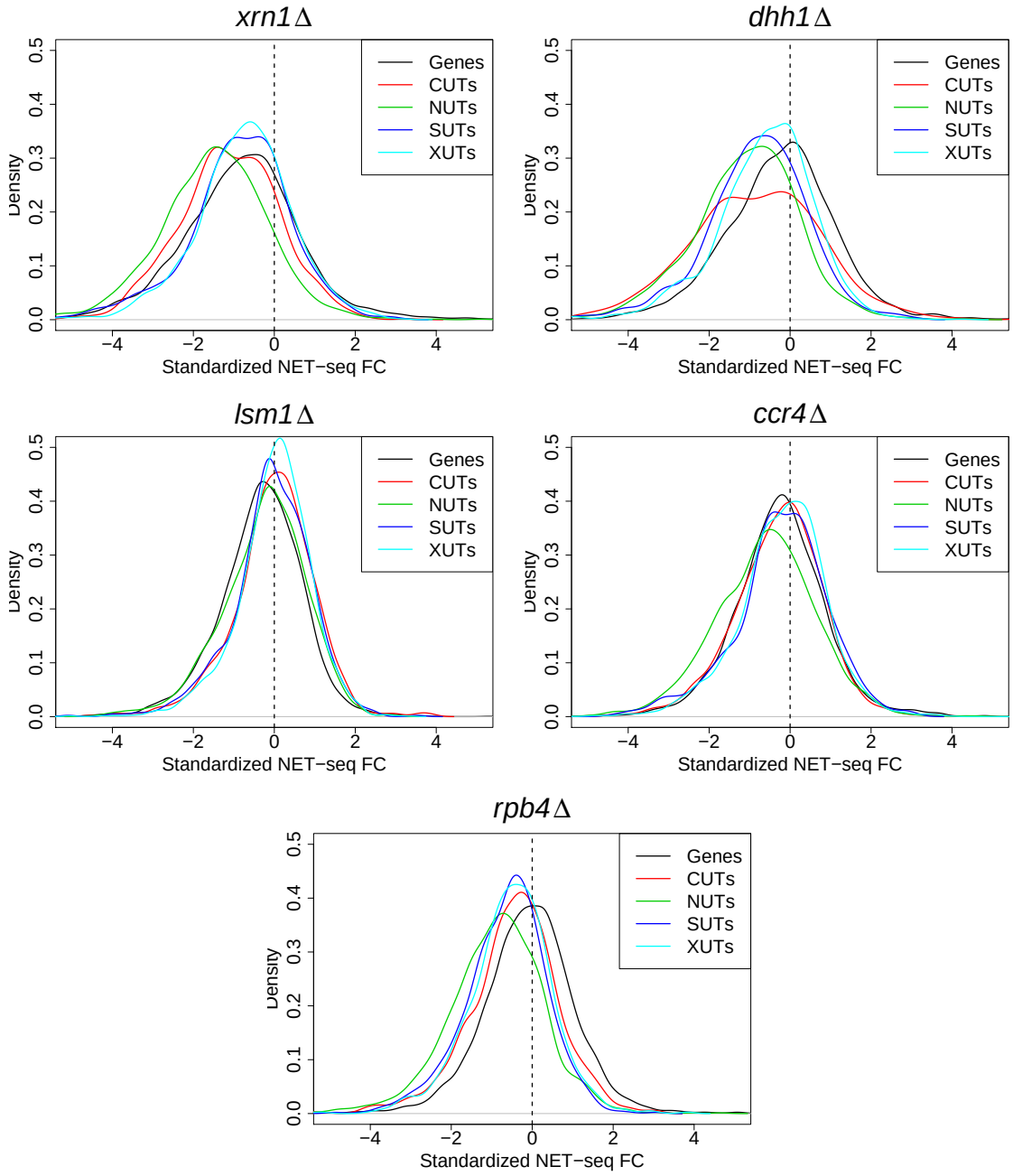


Fig. S12 Pol II occupancy fold changes across different transcript types. Fold changes were computed by the same procedure described in Fig. 1A.

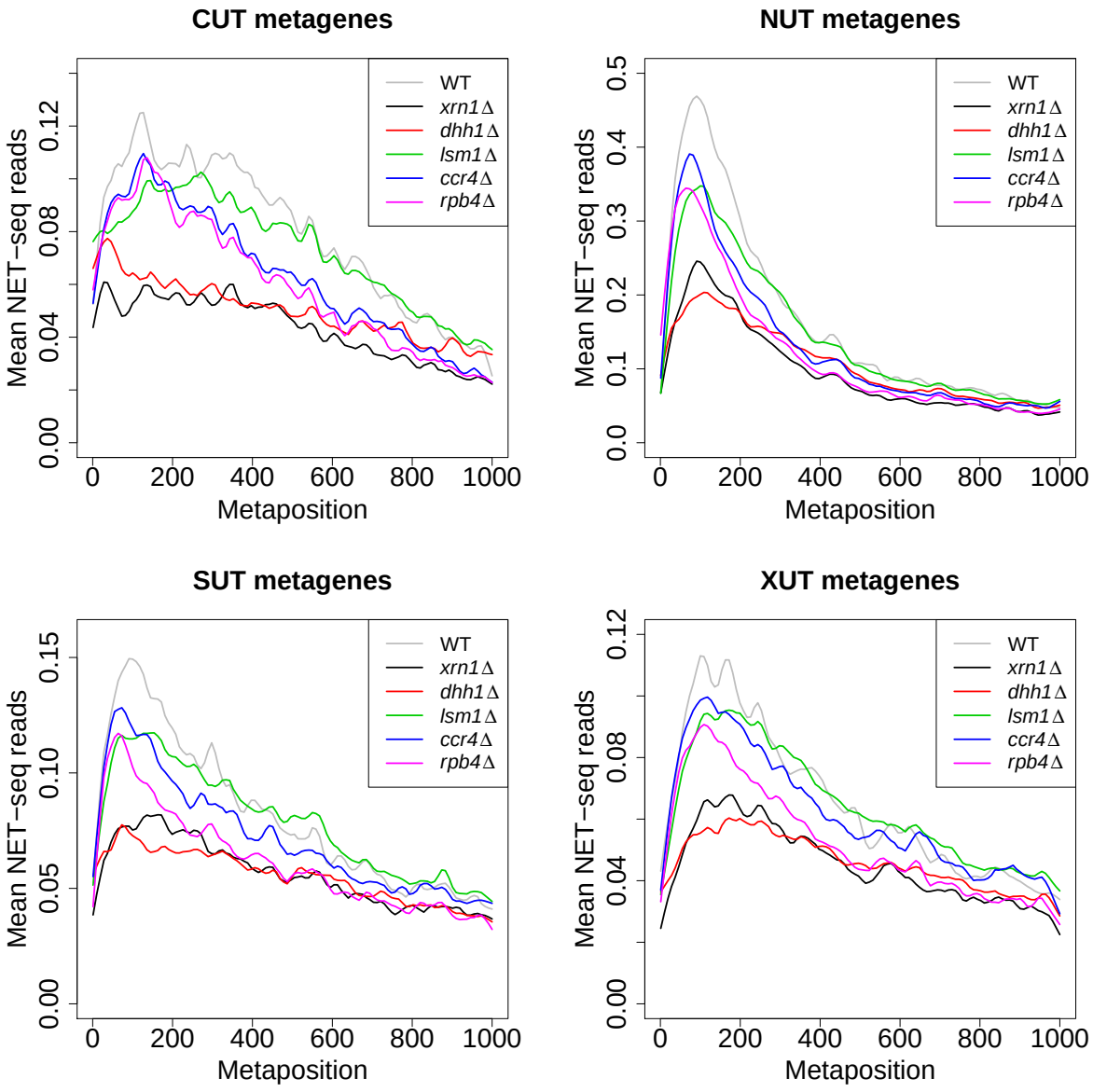
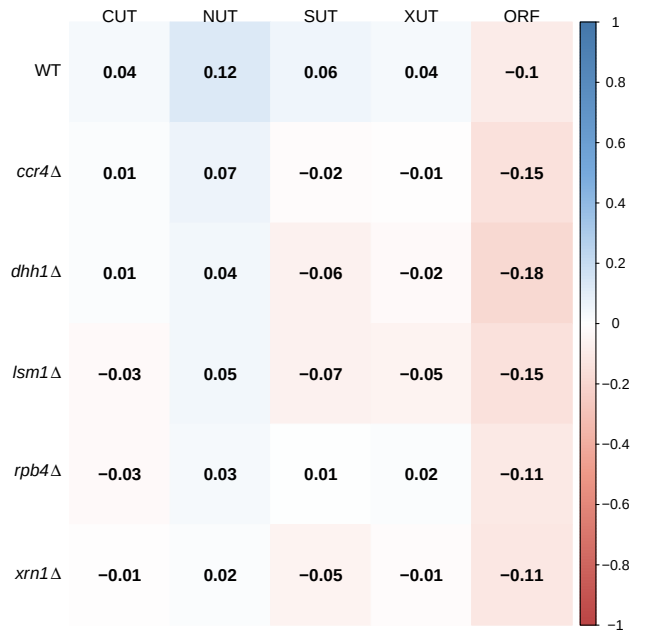


Fig. S13 Smoothed metagene profiles for all annotated ncRNAs. The procedure used to generate Fig. S6 was applied to CUTs, NUTs, SUTs, and XUTs with LOWESS smoothing due to the high noise levels in these transcripts.

A Correlations among NET-seq raw values in convergent pairs



B Correlations among NET-seq FCs in convergent pairs

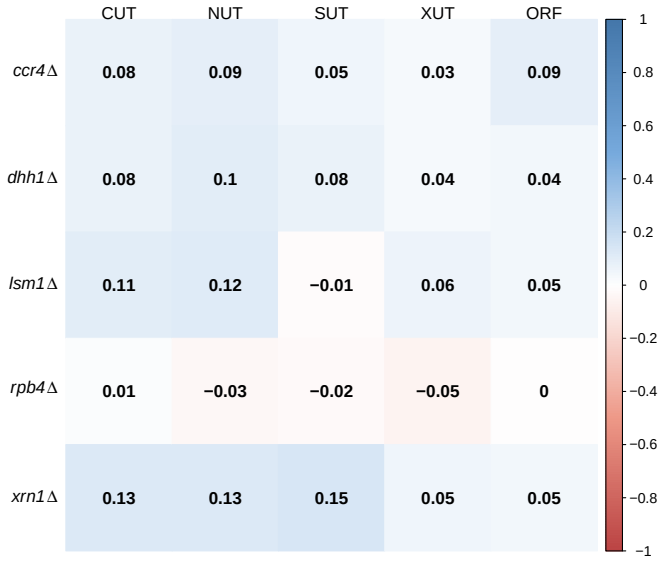
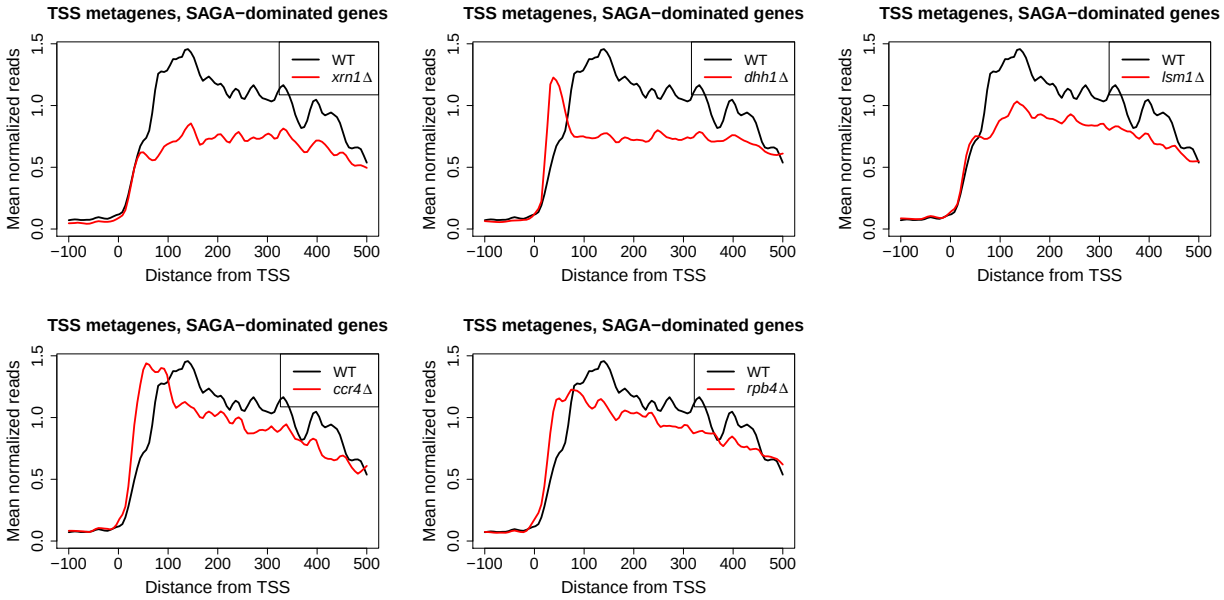


Fig. S14 Correlations of raw and standardized fold changes in sense and antisense Pol II levels for genes and convergent antisense transcripts as measured by NET-seq. Each entry gives the within strain correlation between Pol II (A) levels or (B) fold changes in coding regions and transcripts of the given type which are in convergent antisense orientations with coding regions. Fold changes were computed using DESeq2 (33).

A Metagenes near TSS (SAGA-dominated genes)



B Metagenes near PAS (SAGA-dominated genes)

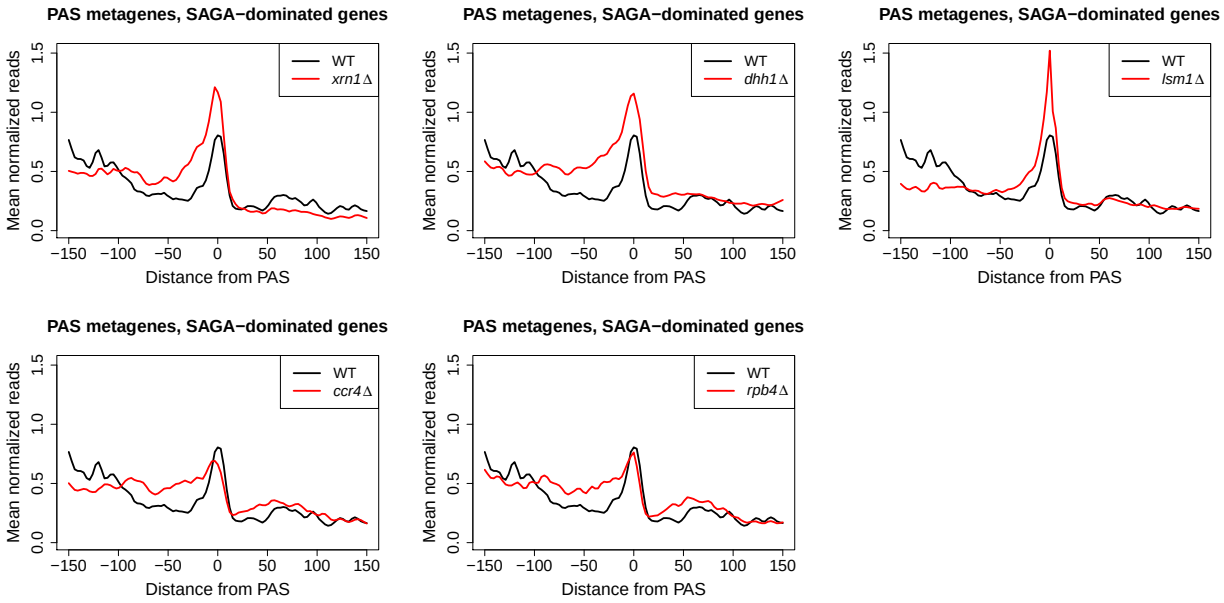
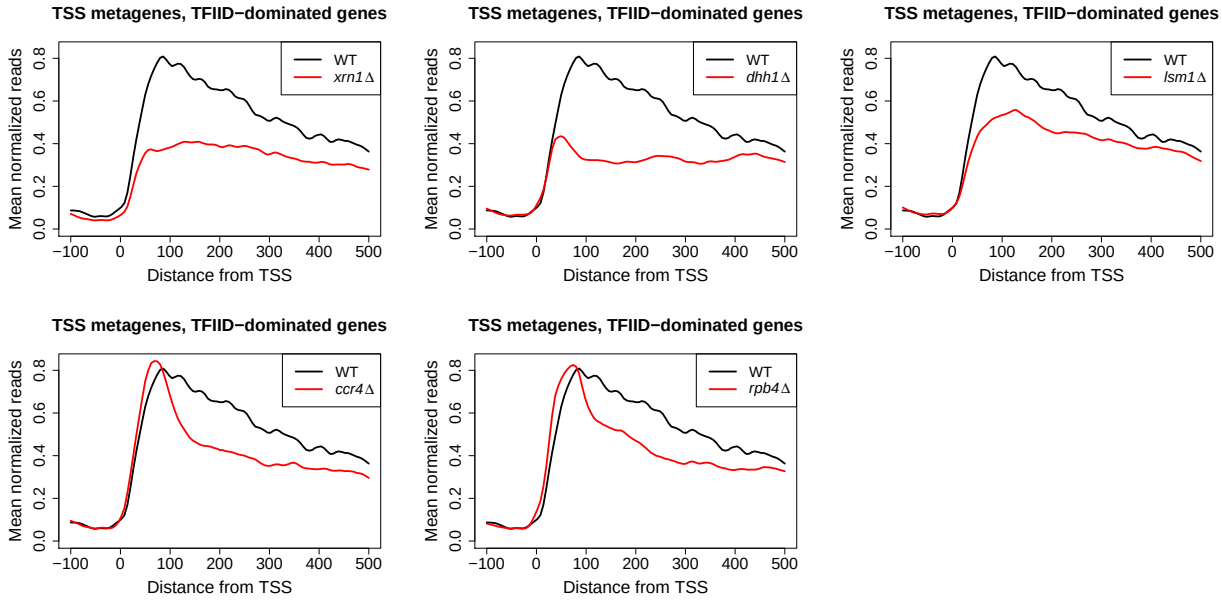


Fig. S15 Normalized metagene profiles near TSS and PAS for SAGA-dominated genes. Reads were aggregated as described in Figs. 3A (TSS) and 5A (PAS) but restricted to genes previously considered to be SAGA-dominated.

A Metagenes near TSS (TFIID-dominated genes)



B Metagenes near PAS (TFIID-dominated genes)

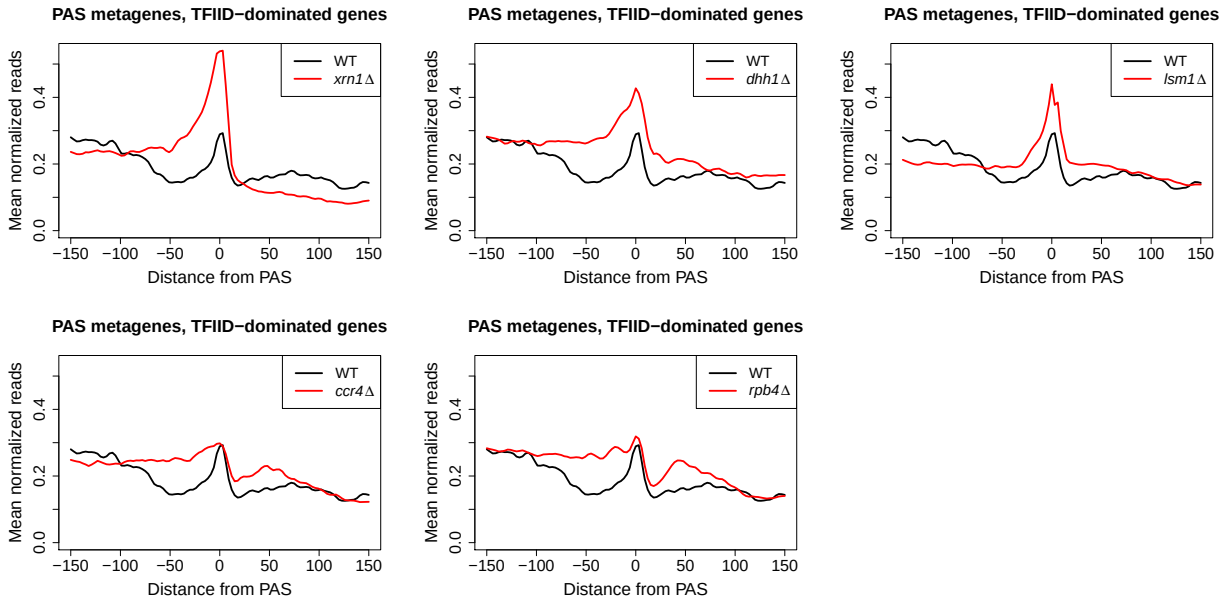


Fig. S16 Normalized metagene profiles near TSS and PAS for TFIID-dominated genes. Reads were aggregated as described in Figs. 3A (TSS) and 5A (PAS) but restricted to genes previously considered to be TFIID-dominated.