

Supplemental Methods and Figures

Poorey et al. “RNA Synthesis Precision is Regulated by Preinitiation Complex Turnover”

ChIP Tiling Array Data

Raw array data (CEL files) or the intensity data of the treatment (ChIP sample) and control (mock sample) were quantile normalized separately within replicate groups and the median intensity was scaled to a common target of 100. To determine the size and the significance of the difference between ChIP and mock sample data (signal estimates and the p values) we applied a Wilcoxon Rank-Sum test to log transformed PM-MM values, $\log_2(\max(\text{PM}_i - \text{MM}_i, 1))$, whose genomic coordinate i (i.e., central position of 25mer PM/MM probe sequence) fell within a 500 bp (i.e., the ChIP fragmentation size) sliding window. The Hodges-Leman estimator, which is the estimator associated with the Wilcoxon Rank Sum test, was calculated along with $-10\log_{10}(\text{p-value})$. This analysis was applied to the following datasets:

1. WT TBP occupancy: 3 replicates of myc-tagged TBP ChIP as treatment and mock ChIP as control in WT cells.
2. *mot1-42* TBP occupancy: 3 replicates of myc-tagged TBP ChIP as treatment and mock ChIP as control in *mot1-42* cells.
3. WT TFIIB occupancy: 2 replicates of TFIIB ChIP as treatment and mock ChIP as control in WT cells.
4. *mot1-42* TFIIB occupancy: 2 replicates of TFIIB ChIP as treatment and mock ChIP as control in *mot1-42* cells.
5. Differential TBP occupancy: 3 replicates of myc-tagged TBP ChIP in *mot1-42* cells as treatment and 3 replicates of myc-tagged TBP ChIP in WT cells as control.
6. Differential TFIIB occupancy: 2 replicates of TFIIB ChIP in *mot1-42* cells as treatment and 3 replicates of TFIIB ChIP in WT cells as control.

The above analysis was performed using TAS (Tiling Analysis Software version 1.1), which generates graph files (.bar and .txt) files which contain signal and log transformed p-values as a function of genomic coordinates.

RNA Tiling Array Data

Estimates of RNA levels were made from the raw array data (CEL files) by first quantile normalizing all replicate arrays and scaling the data to a target median intensity of 100. We applied the Wilcoxon Signed-Rank test to the normalized $\log_2(\max(\text{PM}_i - \text{MM}_i, 1))$ values whose genomic coordinate i fell within a 100 bp (i.e., a length much smaller than the typical ORF) sliding window to calculate the log transformed probability ($-10\log_{10}(\text{p-value})$) that the RNA was detected above noise levels. The associated Hodges-Leman estimator was used to estimate RNA levels. This analysis was applied to the following samples:

1. Total RNA in WT cells (*MOT1* isogenic control, 2 replicates).
2. Total RNA in *mot1-42* cells (2 replicates).
3. Total RNA in WT cells (*SET2* isogenic control, 2 replicates)

4. Total RNA in *set2Δ* cells (2 replicates)

Differential RNA levels were estimated using the same procedure described for the ChIP data including a window size of 500 bp with *mot1-42* RNA samples as treatment and WT RNA samples as control.

ChIP-chip Peak Identification

In order to identify the location and height of TBP and TFIIB ChIP-chip peaks, we started by applying a series of cutoffs to the signal data, from x to y in intervals of z (as illustrated in Figure S3B), which result in a series of segments or intervals for each cutoff. To avoid characterizing spurious, noisy local peaks, we further joined intervals for every cutoff that are ≤ 50 bp apart and eliminated those that are < 100 bp long. This yielded a series of intervals which encompass TBP and TFIIB peaks which were found by searching for the maximum signal in each interval.

Associating ChIP-chip Peaks to Genes

Using the TBP and TFIIB peaks, we associated a peak with a gene if it was within -300 to +50 bp of the annotated transcription start site (TSS). We allowed multiple genes which satisfy the distance criteria, to be associated with a single peak. Conversely, a single gene can be associated with multiple peaks, each of which satisfies the distance cutoff separately.

Average Plots

We generated average RNA and ChIP enrichment plots for selected genes in Figures 6, 8, and S5 and S7A, C, D by setting the TSS of the genes to a common value of 0 (effectively aligning the TSSs) and averaging the signal over genes as a function of genomic coordinate. For Figure S7B, the gene end was set to 0 and the signal was averaged over the gene upstream and downstream. We applied modest smoothing to the data by calculating the averages in a sliding window ranging from 1 to 50bp in length depending on the plot.

Heat Maps

The Heat maps shown in Figures 8C- and D were generated by discretizing absolute and differential *mot1-42* versus WT TBP and TFIIB signals (i.e., maximum signal value within -300 to +50 bp of the TSS) into a 30x30 grid. The position of the box on the grid represents the median absolute or differential TBP (x-axis) and TFIIB (y-axis) value of the genes in the box. Similarly, absolute or differential RNA levels of genes falling within a box were summarized by their median value and assigned a color based on percentile (e.g. dark brown for upper 20% absolute or differential RNA levels).

RNA Length Change Algorithm

We applied a 0.3 cutoff to differential RNA signal data generated using a 500 bp sliding window. By comparing these segments to annotations, we identified 5' and 3' length changes in *mot1-42* relative to WT transcription. These length changes were either positive or negative and represented a different RNA defect as described in Figure S2. Cases where the significant differential expression segment mapped to two or more

annotated genes were separately characterized. Putative length changes that did not satisfy all the criteria below were filtered out:

- The overlap between annotations and the significant differential signal segments was at least 100 bp long.
- The length of the defect computed was ≥ 150 bp long.
- Median signal in the defect was significantly different from the baseline differential expression value as described in Figure 2A. In the case of extensions where the segment fell beyond the boundaries of the annotation, the median signal in the extended region, S_{ext} , was greater than 0.44 to make a length change call. In the case of truncations, we defined two quantities: (1) S_{overlap} as the median of the differential signal in the region where the differential segment and the annotation overlapped and (2) S_{int} as the median signal in the region internal to the gene that did not overlap the differential expression segment. For a truncation to be called, we required $S_{\text{overlap}} - S_{\text{int}} \geq 0.44$.
- Spliced genes (283 total) were excluded because differential expression of spliced genes spans more than one CDS segment and were therefore erroneously called as length changes.

SUT/CUT Analysis

We assessed the association between SUTs/CUTs (Neil et al. 2009) and significant changes in Mot1-mediated changes in RNA expression (i.e., differential level and/or location of expression) from the tiling array analysis. Genes were classified according to the location of SUTs/CUTs: (1) overlapping the gene; (2) located in the promoter of the gene and (3) located at the 3' end of the gene. We used BEDTools to perform the overlap analysis (Quinlan et al, 2010). Based on an analytical formula, we found the overlap of SUTs/CUTs with RNA length changes was the expected number due to random chance. However the association with RNAs that were differentially up or down regulated were found to be potentially significant from the same analysis. To further test this association, we randomly associated SUTs/CUTs with genes while maintaining the same number of overlapping, 5' and 3' proximal cases. We generated 10,000 random associations of SUTs/CUTs with genes and calculated both p-values and the enrichment of observed over random.

Comparison of Shifts in TBP Localization and RNA Length Changes

A number of analyses were performed to assess the extent to which aberrant TBP localization correlated with changes in RNA lengths. Similar to the RNA length change calculation, we first calculated the differential TBP signal between *mot1-42* and WT cells. We then calculated average plots of the differential signal near the TSS for all genes. We found a general tendency for the differential TBP peak to be localized closer to the TSS or even within the ORF when compared to the WT peak. This is consistent with the average plots shown in Figure 8A where the differences between *mot1-42* and WT tend to be largest near the TSS. To assess if this differential shift in TBP signal correlated with RNA length changes, we identified the differential peak positions and calculated the difference between the differential peak and the WT peak, Δ . For each of the four RNA length change classes as well as those that display no detectable length changes (i.e., nulls), we computed the distribution of Δ values. We observed little to no

significant differences between each of the RNA length change classes and the null class distributions. This was reflected in insignificant p-values derived from a Kolmogorov-Smirnov test between the null and each of the RNA length change classes separately. We then assessed the extent to which the shifts between the WT and the *mot1-42* TBP peaks correlated with the LI and EI length change classes. We first calculated the difference between *mot1-42* and WT peak positions as well as the difference between the apparent site of initiation of the new RNA and the annotated TSS. We calculated the Pearson correlation coefficient for these two sets of differences and found a correlation coefficient of -0.51 for the downstream initiation events.

S1

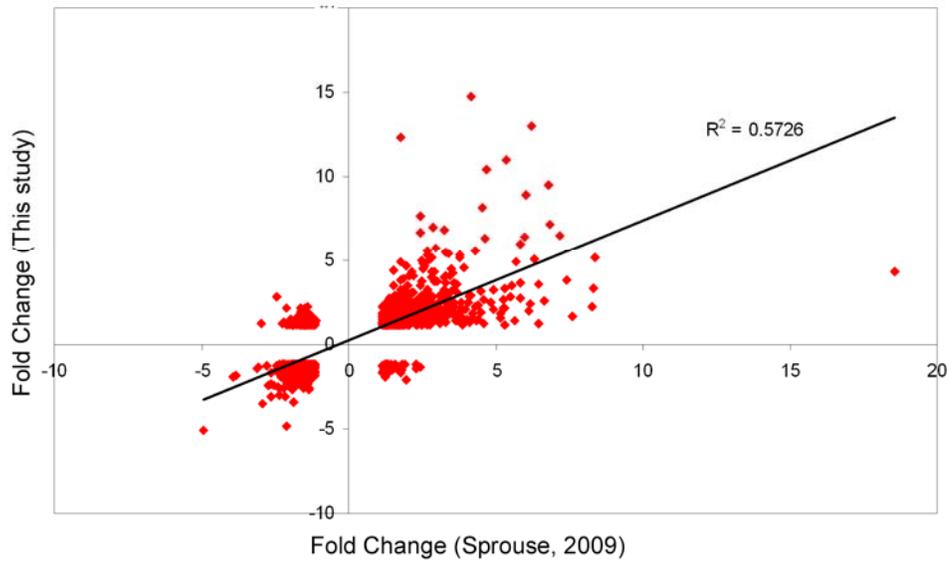


Figure S1: Comparison of *mot1* microarray fold changes from Sprouse et al. (2009) and the present study. Note the linear correlation between Mot1-dependent fold changes derived from Agilent Yeast Oligo arrays (conventional microarrays; Sprouse et al. 2009) and those derived from this study using Affymetrix *S. cerevisiae* 1.0R tiling arrays. The slope of the best-fit line is 0.73. The technical differences in experimental design, array sensitivity, and data analysis account for why the datasets are correlated but the slope of the line is not equal to 1.

S2

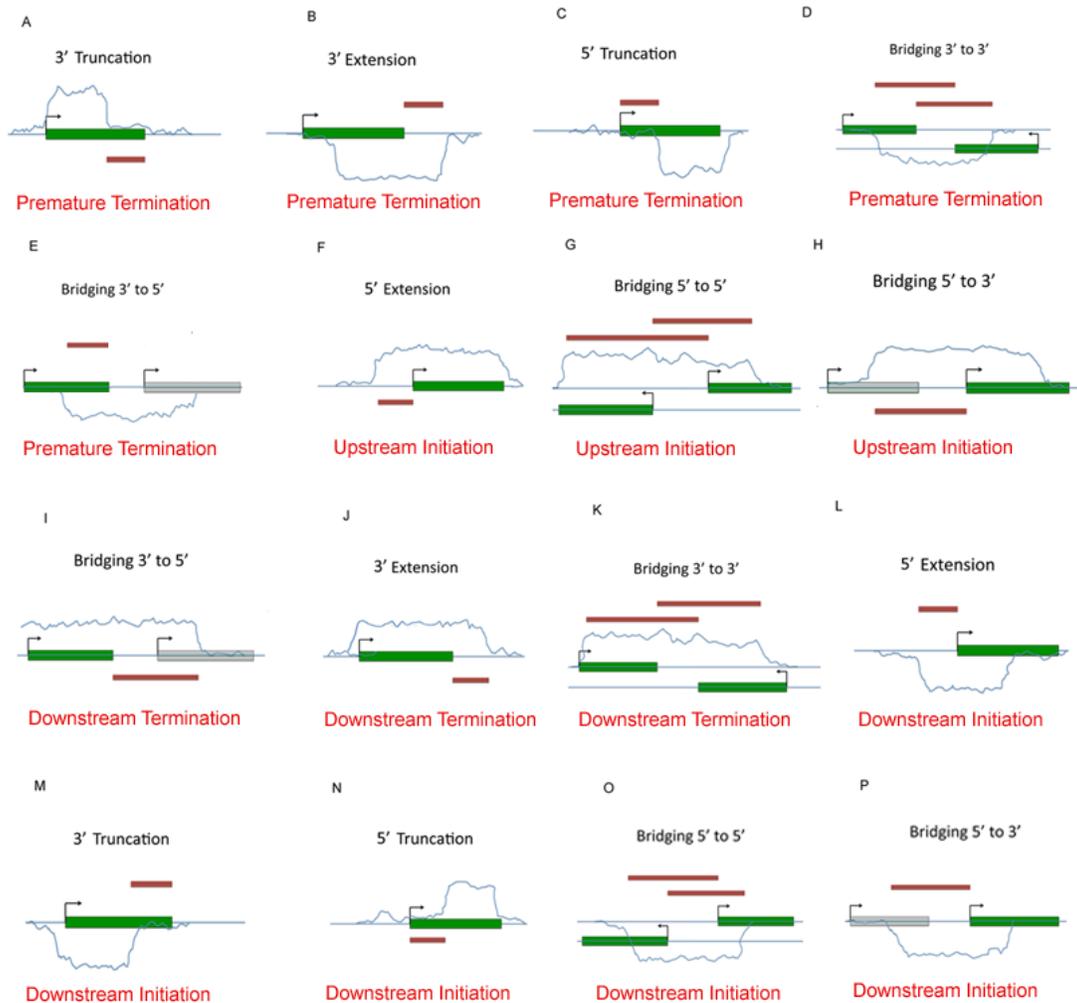


Figure S2: Biological complexity in the observed RNA length changes. (A-P) The gene-dense nature of the *S. cerevisiae* genome, which gives rise to transcription of ~90% of the sequence, leads to complexity in the interpretation of observed RNA length changes, particularly when a differential RNA signal bridges two adjacent annotations. Schematic representations of RNA length changes are shown, with differential RNA shown in the blue curves and the brown segments showing the RNA length change segments called by the algorithm. The length changes were called with respect to a given gene or genes as shown in green.

S3

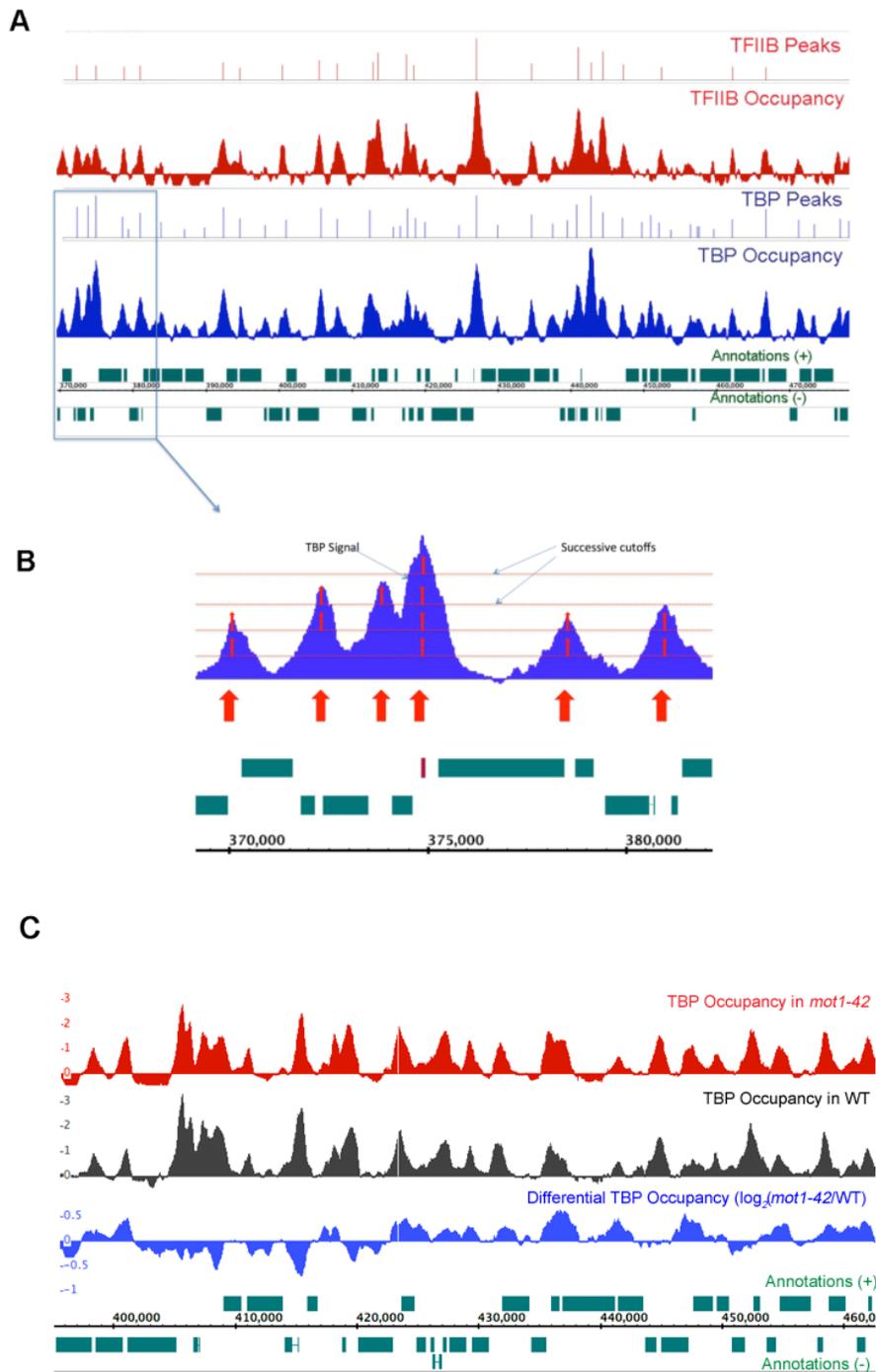


Figure S3: Overview of TBP and TFIIB peak-finding method. (A) Screenshot of a region of chromosome 14 showing TBP occupancies in WT cells (black) and TFIIB (red) ChIP-on-chip signals. The vertical lines shown above each signal track indicate the positions of TBP and TFIIB identified by the algorithm and the annotated genes by green bars (Pol II transcribed) and a red bar (Pol III transcribed gene). (B) Enlargement of a portion of panel (A) shows the principle of the method. In each successive segment

(horizontal line), the position of maximum signal was computed, thereby estimating the location of the bound factor (red arrows). Annotated Pol II transcribed genes are indicated by green bars, and the red bar denotes a tRNA gene. (C) Screenshot of a region of chromosome 2 showing TBP occupancy in *mot1-42* (red), TBP occupancy in WT (black), and differential TBP occupancy (*mot1-42*/WT; blue). Annotated genes are shown as green bars.

S4

A

	TBP	TFIIB
Total Peaks - WT	5226	1932
Total Peaks - <i>mot1-42</i>	4810	2461

B

Number of TBP Peaks	WT	<i>mot1-42</i>
0	2055	1997
1	4697	4772
2	215	194
3	1	5

C

Number of TFIIB Peaks	WT	<i>mot1-42</i>
0	4649	4122
1	2279	2756
2	40	87
3	0	3

D

	WT	<i>mot1-42</i>
TBP Peaks associated with TATA box	2474	2275
TBP Peaks not associated with TATA Box	2752	2535

Figure S4: Summary of counts obtained from the peak finding algorithm. (A) The total number of TBP and TFIIB peaks computed by the peak finding algorithm in WT and *mot1-42* cells. (B) The table shows the distribution in the number of TBP peaks associated with a given promoter. Note that most promoters (4697) were associated with a single TBP peak in WT cells. (C) The table shows the distribution of the number of TFIIB peaks associated with a given promoter. Note that most promoters (4649) did not have an associated TFIIB peak in WT cells. (D) The table shows the distribution of TBP

peaks with and without TATA boxes. TATA boxes were defined as in Basehoar et al. 2004. Note that in contrast to the results shown in Table 1, these assignments of TBP peaks to TATA sequences were made without regard for whether the detected TATA box was located within a promoter. Overall, about 20% of Pol II promoters possess a TATA box; the large number of TATA-associated TBP peaks in this analysis indicates that a substantial number of peaks are not in promoters.

S5

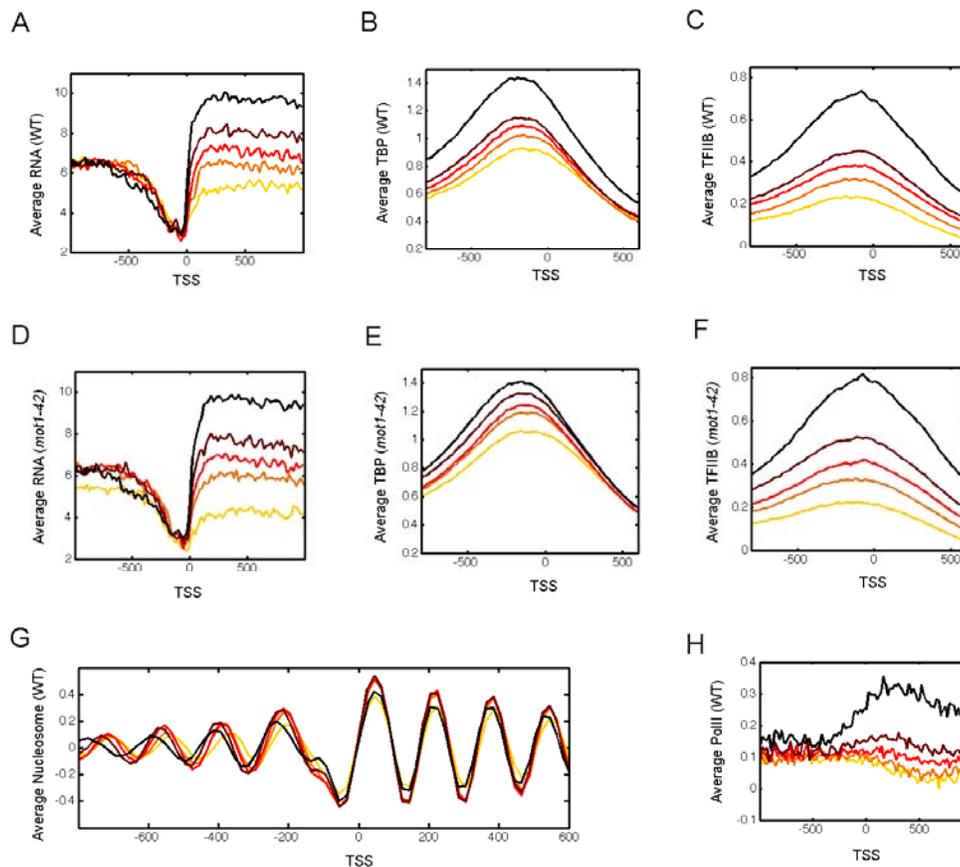


Figure S5: Average profiles of TBP, TFIIB, nucleosome positioning, and Pol II stratified by expression level. (A) Average RNA plots for genes in WT cells stratified by expression level (i.e., quintiles). (B, C) Average plots for TBP and TFIIB (as indicated) in WT cells for genes grouped by RNA level depicted in (A). The curves were generated using a 20 bp sliding window. Note that on average, TBP and TFIIB levels increased with increasing RNA level. (D-F) Average profiles for RNA, TBP and TFIIB as in (A-C) but for genes in *mot1-42* cells. (G) Average nucleosome profile for genes segregated by expression level. Nucleosome data were obtained from Whitehouse et al. 2007. (H) Average Pol II occupancies in WT cells (Steinmetz et al. 2006) stratified by relative RNA level. The curves were made with 50 bp window smoothing of the signal.

S6

A

Type of interaction	SUTs			CUTs		
	Sense/ Antisense	Mot1- repressed	Mot1- activated	Sense/ Antisense	Mot1- repressed	Mot1- activated
Overlapping annotations	3/551	130 ¹	118 ²	0/626	95	234 ¹
Promoter sharing	59/563	103	221	142/622	123	261
3' Proximal	33/170	49	39	88/93	54	28

¹ enriched² depleted

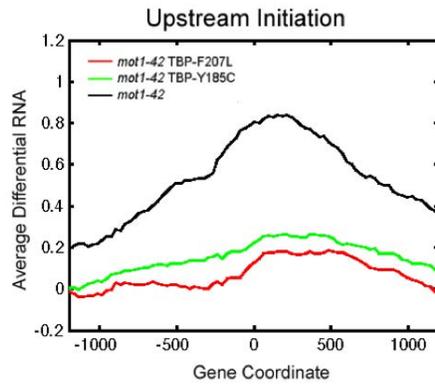
B

Length Change Class	SUTs			CUTs		
	Overlapping annotations	Promoter sharing	3' Proximal	Overlapping annotations	Promoter sharing	3' Proximal
Premature Termination	35	39	3	35	37	3
Premature Termination Mot1-repressed	3	10	2	3	5	1
Premature Termination Mot1-activated	32	30	1	32	32	2
Upstream Initiation	16	6	2	7	7	3
Downstream Termination	17	14	6	16	17	4
Downstream Initiation	4	7	4	13	9	2

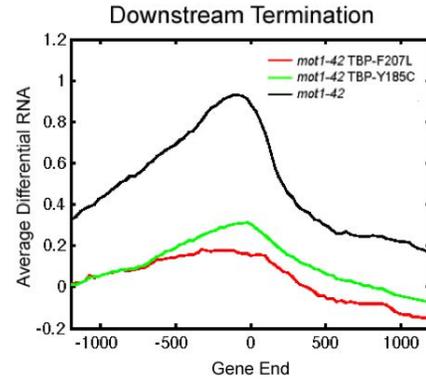
Figure S6: Summary of counts obtained from the SUT/CUT analysis. (A) The table shows the characterization of the genes according to position with respect to the CUT/SUT and Mot1 regulation (Mot1-activated or Mot1-repressed) for the 3 classes of association: (1) overlapping the gene; (2) located in the promoter of the gene and (3) located at the 3' end of the gene. The significantly enriched cases over random are represented in red and are schematized in Figure 6, and counts in blue represent significant depletion from the random counts. (B) The table shows the number of genes associated with a SUT/CUT class (3 categories listed above).

S7

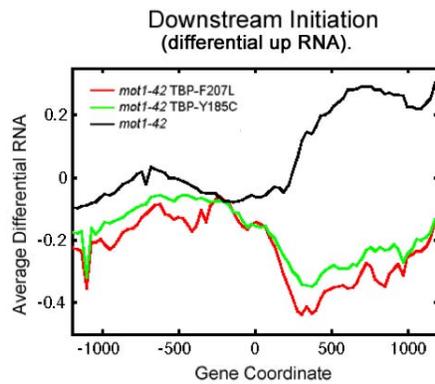
A



B



C



D

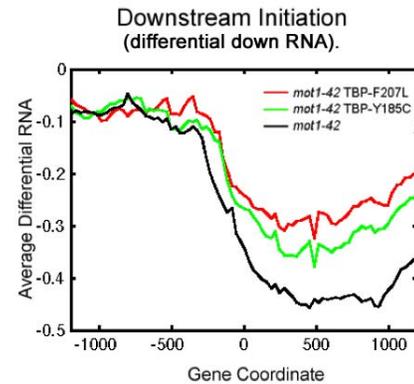


Figure S7: Suppression of transcription length changes in *mot1-42* cells by TBP bypass alleles: (A-D) Average profiles of differential RNA for genes showing upstream initiation (A), downstream termination (B), and downstream initiation (C, D) length changes in *mot1-42* versus WT cells (black lines). The x-axis indicates the position along the chromosome in base pairs relative to the start of the gene annotation (zero in A, C, and D) or gene end (zero in B). The average differential RNA profiles for these same genes in *mot1-42* cells harboring TBP-F207L versus WT (red lines) and *mot1-42* TBP-Y185C versus WT (green lines) show that the bypass alleles at least partially suppress each of the RNA length change categories.

S8

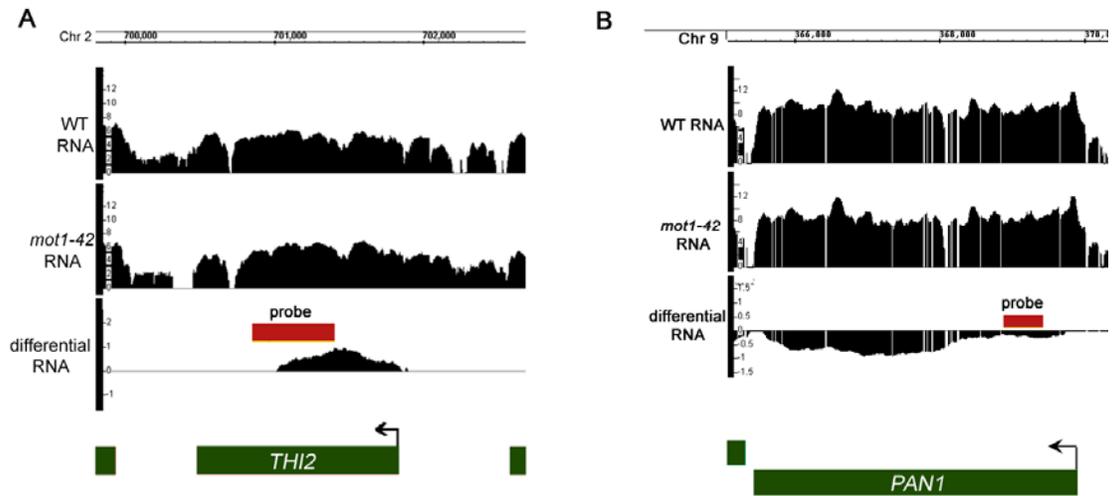


Figure S8. Schematic showing *THI2* and *PAN1* RNA signals and probes used for Northern blotting. Integrated Genome Browser screenshots show \log_2 WT, *mot1-42*, and differential RNA levels (*mot1-42*/WT). A) The *THI2* probe (red box) spanned +240 to +930 with respect to the start of the open reading frame. B) The *PAN1* probe (red box) spanned +420 to +864 with respect to the start of the open reading frame.

References

- Basehoar, A.D., S.J. Zanton, and B.F. Pugh. 2004. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell* **116**: 699-709.
- Neil, H., C. Malabat, Y. d'Aubenton-Carafa, Z. Xu, L.M. Steinmetz, and A. Jacquier. 2009. Widespread Bidirectional Promoters are the Major Source of Cryptic Transcripts in Yeast. *Nature* **457**: 1038-1042.
- Quinlan, AR and Hall, IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**:841–842.
- Sprouse, R.O., M.N. Wells, and D.T. Auble. 2009. TATA-Binding Protein Variants that Bypass the Requirement for Mot1 In Vivo. *J. Biol. Chem.* **284**: 4525-4535.
- Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. 2006. Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. *Mol. Cell* **24**: 735-746.
- Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**: 1031-1035.

Poorey et al. "RNA Synthesis Precision Is Regulated by Preinitiation Complex Turnover".

Supplemental Table 1. Primer sequences used in this study

primers for
RT-PCR

Gene	Forward	Reverse
PDR11 3'	5'-GCCAGGGAAGCGAGATCTAA-3'	5'-AGCAATCAAAGAACGGTCCA-3'
PDR11 5'	5'-ACCCAACATCAGCGCTCTTC-3'	5'-TCAGATGCTTGTGACCATGGA-3'
PDR11 promoter	5'-TCACACACTTAAACCCTTTTCTCA-3'	5'-TTTTTACCAACCCGTTTACGA-3'
EMP47 promoter	5'-AGAAGCCTTAATTGGTATTGAATAAT-3'	5'-TCCCTGATGCTAACTATTTGTTAATTT-3'
ACT1 promoter	5'-CAGCTTTTAGATTTTTCACGCTTA-3'	5'-TTTTCGATCTTGGGAAGAAAA-3'
THI2 5'	5'-ATCAGTTCATTACAGGCTCGCA-3'	5'-TTTCTGGCACAATGGTTTCG-3'
THI2 3'	5'-CATCAATATGGCCTCGTGCC-3'	5'-CAACTCCAACGAGCGATGC-3'
PDC5 5'	5'-CTTTGAATGGTATTGCCGTTTC-3'	5'-CGTGCAAAACACCGACATGT-3'
PDC5 3'	5'-TCACAGTCGGCGCTCTATTG-3'	5'-TCACAGTCGGCGCTCTATTG-3'
EMP47 5'	5'-ACATGGGCTGGATTGCTAGAA-3'	5'-TGCATCTGAAGTGCACCCAA-3'
EMP47 3'	5'-AGGGAAGCTACCAAACGTCATG-3'	5'-AATTCGTCGACCTGAGGACC-3'
TAT1 3'	5'-AAATCCAAACAGCTCGCGTT-3'	5'-AAATTGGTCATCCTCTTAAAAAATTTTC-3'
TAT1 3' ext	5'-CAAAGCTTCCACCTTGCACC-3'	5'-CAAACATCCATCAGTCTGCAAC-3'
ARN1 3'	5'-GGGATCCTAAATTGACTGATACTGTTG-3'	5'-CGCCATCCTCGATGTATTCC-3'
ARN1 3' ext	5'-TTTCTCTCTACAAGACTCTACCATTCTAA-3'	5'-AAAGGAATATGCAAAGAACTAGACATTTTC-3'
ACT1 5'	5'-TCGTCGGTAGACCAAGACACC-3'	5'-CTTTTTCACCCATACCGACCAT-3'
ACT1 3'	5'-CATGTTCCAGGTATTGCCG-3'	5'-GCCAAAGCGGTGATTCCT-3'
PCA1 5'	5'-GCTGCAATACCAATTCGAACC-3'	5'-CGATGCAAAGCCTACCCTTC-3'
PCA1 3'	5'-CGCAGTCAAAAAGACCGGTT-3'	5'-CGTCGTTTGTCCGTCACC-3'
STE11 5'	5'-CCACTGCTACGCTATCGATGAAT-3'	5'-CACAGTGCTTTTCAGGAATCAATT-3'
STE11 3'	5'-ATAGGCACAAACACGACCCC-3'	5'-TCCTTCTGACGTAGCCCAGG-3'

primers for
northern
probes

THI2	5'-ATCAGTTCATTACAGGCTCGCA-3'	5'-CAACTCCAACGAGCGATGC-3'
PAN1	5'-CAATCCATACAACCCCAAGG-3'	5'-ACTGTGCTTGTGAGGAGGT-3'
ACT1	5'-GTTCTAGCGCTTGCACCATC-3'	5'-TTCCAAATTTCAAGCCCCTA-3'