bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Promoter scanning during transcription initiation in *Saccharomyces cerevisiae*: Pol II in the "shooting gallery"

Chenxi Qiu[1,a,§], Huiyan Jin[1,b,§], Irina Vvedenskaya[2,3], Jordi Abante Llenas[4,c], Tingting Zhao[5], Indranil Malik[1,d], Scott L. Schwartz[6,e], Ping Cui[1], Pavel Čabart[1,f], Kang Hoo Han[7], Richard P. Metz[6], Charles D. Johnson[6], Sing-Hoi Sze[1,8], B. Franklin Pugh[7], Bryce E. Nickels[2,3], Craig D. Kaplan[5]*

[1]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843-2128

[2]Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, 08854

[3]Department of Genetics, Rutgers University, Piscataway, NJ, 08854

[4]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128

[5]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, 15260

[6]Genomics and Bioinformatics Service, Texas A&M AgriLife, College Station, TX 77845

[7]Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA 16802

[8]Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843-3127

*To whom correspondence should be addressed.

§Equal contributions

[a]Current Address: Department of Medicine, Division of Translational Therapeutics, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215

[b]Current Address: Roche Nimblegen, Madison, WI 53719

[c]Current Address: Whitaker Biomedical Engineering Institute, Johns Hopkins University, Baltimore, MD 21218

[d]Current Address: Department of Neurology, University of Michigan, Ann Arbor, MI, 48109, USA

[e]Current Address: Covera Health, New York City, NY 10017

[f]Current Address: First Faculty of Medicine, Charles University, BIOCEV, 252 42 Vestec, Czech Republic

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Background

The majority of eukaryotic promoters utilize multiple transcription start sites (TSSs). How multiple TSSs are specified at individual promoters across eukaryotes is not understood for most species. In *S. cerevisiae*, a preinitiation complex comprised of Pol II and conserved general transcription factors (GTFs) assembles and opens DNA upstream of TSSs. Evidence from model promoters indicates that the PIC scans from upstream to downstream to identify TSSs. Prior results suggest that TSS distributions at promoters where scanning occurs shift in a polar fashion upon alteration in Pol II catalytic activity or GTF function.

Results

To determine extent of promoter scanning across promoter classes in *S. cerevisiae*, we perturbed Pol II catalytic activity and GTF function and analyzed their effects on TSS usage genome-wide. We find that alterations to Pol II, TFIIB, or TFIIF function widely alter the initiation landscape consistent with promoter scanning operating at all yeast promoters, regardless of promoter class. Promoter architecture, however, can determine extent of promoter sensitivity to altered Pol II activity in ways that are predicted by a scanning model.

Conclusions

Our observations coupled with previous data validate this scanning model for Pol II initiation in yeast – which we term the “shooting gallery”. In this model, Pol II catalytic activity, and the rate and processivity of Pol II scanning together with promoter sequence determine the distribution of TSSs and their usage. Comparison of TSS distributions and their relationship to promoter sequence among other eukaryotes suggest some, but not all, share characteristics of *S. cerevisiae*.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BACKGROUND

Gene expression can be regulated at all levels, and its proper control is critical for cellular function. Transcription regulation has been of intense interest for decades as it determines how much RNA is synthesized for a given gene or locus. Much regulation occurs at the first step in transcription, initiation. A multitude of signals can be integrated with the activities of transcriptional regulators that converge on individual gene promoters. Subsequent to the integration of regulatory information, RNA Polymerase II (Pol II) and general transcription factors (GTFs) must recognize core promoters to together initiate transcription at specific sequences, transcription start sites (TSSs). As with any biochemical process, the efficiency of individual steps will shape the overall output. Thus, determinants of core promoter output during initiation, both overall expression level and the exact position of transcription start sites (TSSs), will be affected by the efficiency of biochemical events during initiation. How different core promoters modulate biochemical steps in initiation, and the nature of their functional interactions with the initiation machinery, are not fully understood.

Classes of eukaryotic core promoters can be distinguished by DNA sequence motifs and chromatin structure (reviews of the core promoter over time [1-6]). These features together comprise a promoter's architecture, which may also correlate with differential recruitment or requirement for particular GTF complexes. A theme across eukaryotes is that core promoters can be broadly separated into two main classes by examination of architectural features and factor requirements. A number of studies indicate that the most common eukaryotic promoters are nucleosome-depleted regions (NDRs) flanked by positioned nucleosomes, which can support divergent transcription through assembly of pre-initiation complexes (PICs) proximal to flanking nucleosomes (with exceptions)[7-18]. We will adhere to the definition of "core promoter" as representing the DNA elements and chromatin structure that facilitate transcription in one direction, to avoid definitional confusion that a "promoter" inherently drives divergent transcription [19-21]. In yeast, promoter classes have been distinguished in many ways with the end result generally being two main classes of promoter are recognized [9-11, 22]. These classes are distinguished by the presence or absence of a consensus TATA element [23, 24], presence or absence of stereotypical nucleosome organization [11], enrichment for specific transcription factor binding [7, 25, 26], enrichment for non-TATA sequence motifs [27, 28], and differential sensitivity to mutations in GTFs or transcription coactivators [23, 25, 26]. Core promoters attached to defined NDRs tend to lack canonical TATA-elements. Conversely, in yeast and other eukaryotes, core promoters with TATA elements can lack stereotypical nucleosome organization and may have nucleosomes positioned over the TATA box in the absence of gene activation. While there have been a number of additional core promoter elements identified in other organisms, especially *Drosophila melanogaster* [29], we will focus on the distinction provided by presence or absence of TATA-elements.

The TATA element serves as a platform for core promoter binding of the TATA-Binding Protein (TBP). TBP recognition of promoter DNA is assumed to be critical for PIC formation and Pol II promoter specificity. Functional distinction in promoter classes is supported by studies showing differential factor recruitment and requirements between them, with TATA promoters showing higher SAGA dependence and reduced Taf1 (a TFIID subunit) recruitment [23-26], though more recent data have been interpreted as both SAGA and TFIID functioning at all yeast promoters [30, 31]. Conversely, TATA-less promoters show higher Taf1 recruitment and greater requirement for TBP-Associated Factor (TAF) function. Given differences in reported factor requirements and promoter architectures, it is important to understand the mechanistic differences between promoters and how these relate to gene regulation.

TSS selection in *Saccharomyces cerevisiae* has been used as a model to understand how initiation factors collaborate to promote initiation [32, 33]. The vast majority of yeast core

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

promoters specify multiple TSSs [34-36], and multiple TSS usage is now known to be common to the majority of core promoters in other eukaryotes [37-41]. Biochemical properties of RNA polymerase initiation lead to TSSs selectively occurring at a purine (R=A or G) just downstream from a pyrimidine (Y=C or T) – the $Y_{-1}R_{+1}$ motif [42]. $Y_{-1}R_{+1}$ motifs may be additionally embedded in longer sequence motifs (the Inr element)[43, 44]. In yeast, the initiation factor TFIIB has been proposed to "read" TSS sequences to promote recognition of appropriate TSSs, with structural evidence supporting positioning of TFIIB to read DNA sequences upstream of a TSS [45, 46].

Yeast differs from other model eukaryotes in that TSSs for TATA-containing core promoters are generally dispersed, and are found ~40-120 nt downstream from the TATA [47]. Conversely, TSSs at TATA promoters in other organisms are tightly associated ~31 nt downstream of the TATA (with the first T in "TATA" being +1)[48]. As TATA promoters represent ~10% of promoters across well-studied organisms, they are the minority. Classic experiments using permanganate footprinting of melted DNA showed that promoter melting at two TATA promoters in yeast, *GAL1* and *GAL10*, occurs far upstream of TSSs, at a distance downstream from TATA, where melting would occur in other eukaryotes that have TSSs closer to the TATA element [49]. This discovery led Giardina and Lis to propose that yeast Pol II scans downstream from TATA boxes to find TSSs. A large number of mutants have been found in yeast that perturb TSS selection, allowing the genetic architecture of Pol II initiation to be dissected, from those in Pol II subunit encoding genes *RPB1, RPB2, RPB7,* and *RPB9*, to GTF encoding genes *SUA7* (TFIIB), *TFG1* and *TFG2* (TFIIF), and *SSL2* (TFIIH), and the conserved transcription cofactor *SUB1* [50-70]. Mutants in GTFs or Pol II subunits have been consistently found at model promoters to alter TSS usage distributions in a polar fashion by shifting TSS distributions upstream or downstream relative to WT. These observations coupled with analysis of TSS mutations strongly support the directional scanning model for Pol II initiation (elegantly formulated in the work of Kuehner and Brow)[53].

Previous models for how initiation might be affected by Pol II mutants suggested that Pol II surfaces important for initiation functioned through interactions with GTFs within the PIC. We have previously found that altering residues deep in the Pol II active site, unlikely to be directly interacting with GTFs, but instead altering Pol II catalytic activity had strong, allele-specific effects on TSS selection for model promoters [71-73]. Observed effects were polar in nature, and consistent with the Pol II active site acting downstream of a scanning process but during TSS selection and not afterwards. In other words, Pol II catalytic efficiency appears to directly impact TSS selection. For example, it appeared that increased Pol II catalytic activity increased initiation probability, leading to an upstream shift in TSS usage at candidate promoters because less DNA needs to be scanned on average prior to initiation. Conversely, lowering Pol II catalytic activity results in downstream shifts to TSS usage at candidate promoters, because more promoter DNA has to be scanned prior to initiation. In general, candidate promoters examined for TSS selection have mostly been TATA containing (for example *ADH1*, *HIS4*), thus it is not known how universal Pol II initiation behavior or mechanisms are across all yeast core promoters, which likely comprise different classes with distinct architectures. To examine initiation by promoter scanning on a global scale in yeast, we perturbed Pol II or GTF activity genetically to examine changes to TSS usage across a comprehensive set of yeast promoters that likely represent all promoter classes in yeast. We have found that promoter scanning appears to be universal across yeast core promoters. Furthermore, we find that core promoter architecture correlates with sensitivity of core promoters to TSS perturbation in Pol II and initiation factor mutants. Our results have enabled formulation a model where Pol II and GTF function together in initiation to promote Pol II initiation efficiency at favorable DNA sequences. Finally, initiation by core promoter scanning makes predictions about the relationship between usable TSSs in a core promoter and the distribution of TSS usage. We compare yeast TSS distributions to a number of other eukaryotes and find that some, but not all, examined

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

eukaryotic model organisms have TSS distributions consistent with predictions of the "shooting gallery" scanning model.

RESULTS

Initiation mutants affect TSS selection globally in *Saccharomyces cerevisiae*

We previously found that yeast strains mutant for Pol II key active site residues important for normal catalysis showed polar effects on TSS selection at the model *ADH1* promoter in addition to some other promoters [72, 73]. *ADH1* is a TATA-containing promoter with major TSSs positioned at 90 and 100 nucleotides downstream of its TATA box. A number of mutants in Pol II and initiation factors also show TSS selection effects at *ADH1*. TSS selection effects have been hypothesized to relate to alterations in initiation sequence specificity, while the stereotypical polar effects of TSS-altering mutants are consistent with effects on scanning and not necessarily sequence specificity. These are not mutually exclusive models, and to understand better how Pol II activity and GTFs cooperate to identify TSSs, we mapped capped RNA 5' ends genome-wide in *S. cerevisiae* using TSS-seq for WT, a series of Pol II catalytic mutants, a TFIIB mutant (*sua7-58A5*)[71], and a TFIIF mutant (*tfg2*Δ*146-180*)[74]. Positions of capped RNA 5' ends are taken to represent positions of TSSs as Pol II-initiated RNA 5' ends are capped shortly after emerging from the enzyme after initiation. We first determined how reproducible our pipeline (**Figure 1A**) was across the yeast genome, examining correlation of read positions corresponding to 5' ends across all genome positions containing at least three mapped reads in each library being compared (**Figure 1B, Supplemental Figures 1,2**). Examples of correlations between biological replicates are shown in **Figure 1B** for WT, one catalytically fast Pol II allele (*rpb1* E1103G)[75-77], and one catalytically slow Pol II allele (*rpb1* H1085Y)[73]. We refer to fast Pol II alleles as "gain of function" (GOF) alleles and slow Pol II alleles as "loss of function" (LOF) alleles [78]. Correlation plots for all other strains are shown in **Supplemental Figure 1**. Clustering analysis of Pearson correlation coefficients among libraries aggregated from biological replicates for each strain indicates that Pol II and initiation mutant classes can be distinguished based on RNA 5' end mapping alone (**Figure 1C**). **Supplemental Figure 2** shows clustering of Pearson correlation coefficients of individual biological replicates for reads in promoter regions.

We focused our analyses first on promoter windows predicted from the localization of PIC components by Rhee and Pugh [7] and anchored on TATA or "TATA-like" elements (core promoter elements or CPE underlying PIC assembly points) at the +1 position of the promoter window (**Figure 1D**). RNA 5' ends mapping to the top genome strand of these putative promoter windows indicates that these windows are associated with putative TSSs as expected. The majority of observed TSSs are downstream of predicted CPE/PIC locations from Rhee and Pugh, with TSSs originating from a range of distances from predicted CPE/PIC positions. We note that a fraction of promoter windows has TSSs arising from positions suggesting that the responsible PICs for those TSSs assemble at additional positions, either upstream or downstream.

Given the distinct and polar alterations of TSS distribution at model genes by Pol II GOF and LOF mutants, we asked if attributes of RNA 5' end distributions within promoter windows could also distinguish mutant classes. To do this, we examined two attributes of TSS usage: the change in position of the median TSS usage in the promoter window from WT (TSS "shift"), and the change in the width between positions encompassing 80% of the TSS usage distribution (from 10% to 90%, the change (Δ) in TSS "spread", **Figure 1A**). Promoter regions were two-dimensionally hierarchically clustered for both attributes across all TSS libraries and TSS libraries were subsequently clustered for each attribute individually (left of **Figure 1E** shows

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TSS shift and right shows ΔTSS spread). First, we observed that profiles of shift in TSS position or alteration in TSS spread were sufficient to distinguish GOF *rpb1* mutants from LOF *rpb1* mutants. Second, Pol II and GTF mutants showed widespread directional shifting of TSSs across nearly all promoters. Pol II GOF and *tfg2Δ146-180* strains exhibited primarily upstream shifts in TSS distributions within promoter windows, while Pol II LOF and *sua7-58A5* exhibited primarily downstream shifts. Directional shifts are consistent with previously observed shifts at individual promoters, such as *ADH1*, suggesting that promoter scanning is operating across all yeast promoter classes.

We examined changes in TSS distribution relative to promoter class and Pol II mutant strength to determine how each relates to magnitude of TSS changes. To visualize changes, we separated promoters using classification by Taf1-enrichment or depletion as done previously. While recent work indicates that TFIID (containing Taf1) functions at all yeast promoters[30], differential recruitment of Taf1 correlates with promoter nucleosome organization, underlying DNA sequence composition, and DNA element enrichment (TATA *etc*) [7, 11, 23, 24, 27], suggesting this metric is a useful proxy for promoter class. **Figure 2A** shows example heat maps of the difference of normalized TSS distributions between WT and a Pol II GOF or a Pol II LOF mutant. The stereotypical patterns of polar changes to TSS distributions where distribution of TSSs shift upstream (increases upstream and decreases downstream, such as *rpb1* E1103G, or shift downstream (increases downstream and decreases upstream, such as *rpb1* H1085Y, are observed across essentially all promoters, and for all mutants examined including GTF mutants (**Supplemental Figure 3**). By determining the shift in median TSS position in promoter windows, we can see that mutants exhibit different strengths of effects on TSS distributions (**Figure 2B**). A double mutant between *tfg2Δ146-180* and *rpb1* E1103G shows enhancement of TSS defects across promoter classes (**Figure 2B, 2C**), similarly to what has been observed at for defects *ADH1* [71]. Examination of average TSS shift and measured in vitro elongation rate for Pol II mutants shows a correlation between the strength of in vivo TSS selection defect and in vitro Pol II elongation rate [72, 73] (**Figure 2D**). These results are consistent with TSS selection being directly sensitive to Pol II catalytic activity.

Altered TSS motif usage in TSS-shifting mutants

To understand the basis of directional TSS shifting in Pol II mutants, we asked how changes to TSS selection relate to potential sequence specificity of initiation (**Figure 3**). Earlier studies of TSS selection defects in yeast suggested that mutants might have altered sequence preferences in the PIC[32]. Our identified TSSs reflect what has been observed previously for Pol II initiation preferences, *i.e.* the simplest TSS motif is $Y_{-1}R_{+1}$ as in most eukaryotes, with the previously observed budding yeast specific preference for A_{-8} at strongest TSSs [34](**Figure 3A**). Preference for $Y_{-1}R_{+1}$ is common across RNA polymerases and likely reflects the stacking of an initiating purine (R, A/G) triphosphate onto a purine at the -1 position on the template strand (reflected as pyrimidine (Y, C/T) on the transcribed strand)[42]. Within the most strongly expressed promoters, preference for A_{-8} is greatest for the primary TSS, and is reduced from secondarily to tertiarily preferred TSSs, even though these sites also support substantial amounts of initiation. Examination of the most focused, expressed promoters – promoters that contain the majority of their TSSs in a narrow window – reveals potential preferences at additional positions. We analyzed TSS usage within promoter windows by dividing all TSSs into 64 motifs based on identities of the -8, -1, and +1 positions (**Figure 3B**). We asked if Pol II or GTF mutants altered apparent preferences among these 64 motifs. Based on aggregate usage of sequences across our promoter set, we found that the top 4 used motifs were $A_{-8}Y_{-1}R_{+1}$, with the next preferred motifs found among B_{-8}(not A)$Y_{-1}R_{+1}$. Pol II and GTF mutants have clear effects on motif usage distribution concerning the -8A position. Upstream TSS shifting mutants (Pol II GOF and *tfg2Δ146-180*) show a decreased preference for $A_{-8}Y_{-1}R_{+1}$ motifs concomitant

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with a gain in relative usage of $B_{-8}Y_{-1}R_{+1}$ motifs, while downstream TSS shifting mutants (Pol II LOF and *sua7-58A5*) have the converse effect, though primarily increases in $A_{-8}C_{-1}A_{+1}$ and $A_{-8}C_{-1}G_{+1}$. Total TSS usage might be affected by strong effects at a subset of highly expressed promoters, therefore we also examined motif preference on a promoter by promoter basis (**Supplemental Figure 4A,B**). *rpb1* E1103G TSS preferences illustrate that the reduction in preference for $A_{-8}Y_{-1}R_{+1}$ motifs is observed across yeast promoters (**Supplemental Figure 4A**) while H1085Y shows the converse (**Supplemental Figure 4B**).

Different models might explain why initiation mutants alter apparent TSS sequence selectivity, and in doing so lead to polar changes to TSS distribution or vice versa (**Figure 3C**). First, relaxation of a reliance on A_{-8} would allow, on average, earlier initiation in a scanning window because non-A_{-8} sites would be more accessible to the PIC, whereas increased reliance on A_{-8} would have the opposite effect. Alternatively, altered Pol II catalytic activity or GTF function may broadly affect initiation efficiency across all sites. In this case, for there to be an apparent change to TSS selectivity, there would need to be a corresponding polar distribution in TSS motifs within promoter regions. It has already been observed that yeast promoter classes sequence distributions deviate from random across promoters. Here, we examined sequence distributions for individual nucleotides and for select $A_{-8}Y_{-1}R_{+1}$ motifs relative to median TSS position for yeast promoters (**Figure 3D**, **Supplemental Figure 4C**). As noted previously, yeast promoter classes differ based on their distributions of A/T [27, 79]. In Wu and Li, promoters were classified based on their nucleosome structure. Our classification based on Taf1-enrichment similarly divides yeast promoters with Taf1-depleted promoters highly enriched for T and depleted for A on the top DNA strand (**Supplemental Figure 4C**). Furthermore, the extent of depletion or enrichment correlates with promoter expression level in vivo, fitting with prediction based on reporter promoter analyses [80]. Enrichment or depletion of individual nucleotides would also be expected to potentially alter distributions of $N_{-8}Y_{-1}R_{+1}$ TSS motifs. Therefore, we extended our analyses to $N_{-8}Y_{-1}R_{+1}$ motifs (**Figure 3D**). We find that $A_{-8}C_{-1}A_{+1}$, the apparent preferred TSS motif for Pol II in yeast, is markedly enriched at median TSS and downstream positions with a sharp drop off upstream, with enrichment also showing correlation with apparent promoter expression level. A less preferred motif, $T_{-8}T_{-1}A_{+1}$, shows the opposite enrichment pattern (enriched upstream of median TSS, depleted downstream). This biased distribution in promoter sequence for TSS sequence motifs makes it difficult to determine whether apparent altered sequence specificity is a cause or consequence of altered TSS distributions.

TSS motif efficiency and usage altered across a number of TSS motifs

To examine further, we looked at TSS distributions by a method that allows us to determine if the average shapes of distributions are changed or merely shifted. We examined the efficiency of TSS usage by individual TSS motifs, with efficiency determined as the ratio of observed reads for a particular TSS to the sum of those reads and all downstream reads, as defined by Kuehner and Brow [53]. This calculation allows the probability of usage of TSSs to be compared within the framework of the polar scanning process. Scanning from upstream to downstream will create greater apparent usage for upstream TSSs relative to a downstream TSS, even if they are identical in initiation efficiency. If Pol II mutants primarily affect initiation efficiency across TSSs we have specific expectations for how efficiency will be affected. For example, if Pol II LOF alleles decrease efficiency across sequences we predict that median observed efficiency of TSS usage will be lower on average over all promoter positions relative to WT, except at the most downstream positions. This would reflect a spreading out of the usage distribution to downstream positions as more Pol II would continue to scan to downstream relative to WT. Conversely, if Pol II GOF alleles increase efficiency across sequences, we might expect the

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

median efficiency to increase for upstream promoter positions but return to baseline efficiency sooner than WT.

To partially account for innate sequence differences among TSS motifs, we examined TSS usage and efficiency across promoters for specific $N_{-8}Y_{-1}R_{+1}$ motifs (**Figure 4**). Usage is defined as the reads found in particular TSS relative to the total reads for that promoter, whereas efficiency is an estimate of the strength of a TSS, assuming a polar scanning process. As an example, **Figure 4A** focuses on H1085Y and E1103G effects on TSSs containing the $A_{-8}C_{-1}A_{+1}$ motif. Median % usage (median % reads for $A_{-8}C_{-1}A_{+1}$ motifs found at each promoter position for promoter class) shows that E1103G increases $A_{-8}C_{-1}A_{+1}$ motif usage at upstream positions relative to WT usage for each promoter class, even though aggregate $A_{-8}C_{-1}A_{+1}$ motif usage is lower for this mutant. Conversely, H1085Y decreases $A_{-8}C_{-1}R_{+1}$ motif usage at almost all promoter positions except for the most downstream motifs for each promoter class. Examining $A_{-8}C_{-1}A_{+1}$ motif efficiency, which is another way to consider the distribution of reads, we find that E1103G essentially shifts the efficiency curve upstream relative to a WT strain, or to H1085Y, which shifts TSS usage distribution downstream, but appears to do so by reducing $A_{-8}C_{-1}R_{+1}$ motif efficiency at almost all promoter positions. This reflects a "flattening" of the usage distribution and would be expected from an overall reduction in initiation efficiency across TSS motifs. Extending this motif analysis to a range of $N_{-8}Y_{-1}R_{+1}$ motifs used at different levels (**Figure 4B, 4C**) we observe that upstream shifting mutants shift usage upstream for all examined motifs (**Figure 4B**) while downstream shifting mutants have the opposite effects on motif usage. In contrast, when examining $N_{-8}Y_{-1}R_{+1}$ motif efficiencies across promoter positions, downstream shifting mutants tend to reduce efficiencies across promoter positions while upstream shifting mutants shift TSS efficiencies upstream (**Figure 4C**). These analyses are consistent with upstream shifting mutants exhibiting increased efficiency across TSS motifs, which shifts both usage and observed efficiency curve to upstream positions, while downstream mutants reduce the efficiency curve and essentially flatten the usage distributions, as would be expected from reduced initiation efficiency across promoter positions.

Non-TATA promoter sequence motifs do not appear to function like TATA-elements

High-resolution TSS data allow us to evaluate promoter features relative to observed median TSS positions instead of using annotated TSS (one per gene and not necessarily accurate) from the Saccharomyces Genome Database. As has previously been determined, a minority of yeast promoters contain consensus TATA elements (TATAWAWR) and these are enriched in Taf1-depleted promoters (illustrated in **Figure 5A**) within ~50-100 basepairs upstream of TSS clusters but not in Taf1-enriched promoters. Furthermore, TATA enrichment tracks with apparent expression level determined by total RNA 5' reads within promoter windows. On the basis of finding TATA-like elements within ChiP-exo signal for GTFs along with a stereotypical pattern to the ChiP-exo signal, it has been proposed by Rhee and Pugh that promoters lacking consensus TATA elements can use TATA-like elements (TATAWAWR with one or two mismatches) for function analogous to a TATA element [7]. Evidence for the function of such TATA-like elements is sparse. In vitro experiments suggested that a TBP footprint is positioned over potential TATA-like element in *RPS5* promoter, but the element itself is not required for this footprint [81]. In contrast, more recent results have suggested modest requirement for TATA-like elements at three promoters (~2-fold) in an in vitro transcription system [82]. Examination of the prevalence of elements with two mismatches from TATA consensus TATAWAWR within relatively AT-rich yeast promoter regions suggests that there is a high probability of finding a TATA-like element for any promoter (**Figure 5A**). Taf1-enriched promoters show enrichment for an alternate sequence motif, a G-capped A tract (sequence GAAAAA), also called the GA-element (GAE) [27, 28]. This positioning of GAEs approximately 50-100 bp upstream of TSSs is reminiscent of TATA positioning (**Figure 5A**), and the GAE has been proposed to function as a

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

core promoter element at non-TATA promoters [28]. Other studies describe the relationship of this element to nucleosome positioning and suggest that these elements may function directionally in nucleosome remodeling at NDR promoters as asymmetrically distributed poly dA/dT elements [83, 84]. To understand how these potential elements function in gene expression, we cloned a number of candidate promoters upstream of a *HIS3* reporter and deleted or mutated identified TATA, TATA-like, or GAE elements and examined effects on expression by Northern blotting (**Figure 5B**, **Supplemental Figure 5**). As expected, in general, identified consensus TATAs positioned upstream of TSSs were important for normal expression of the *HIS3* reporter. In contrast, neither TATA-like or GAE elements in general had strong effects on expression, though some individual mutations affected expression to the same extent as mutation of TATA elements in the control promoter set. We conclude that GAE or TATA-like elements do not generally function similarly to consensus TATAs for promoter expression.

TSS-shifting initiation mutants alter PIC-component positioning consistent with promoter scanning model

Given results above suggesting that TATA-like or GAE elements may not generally function as core promoter elements and therefore may lack value as potential PIC landmarks, we performed ChiP-exo for GTFs TFIIB (Sua7) and TFIIH (Ssl2) to directly examine PIC component localization in WT, *rpb1* H1085Y, and *rpb1* E1103G cells. ChiP-exo [85] was performed in duplicate for all strains, examined for reads per promoter window correlation (**Supplemental Figure 6**) and reads from replicate libraries aggregated. We reasoned that ChiP-exo would allow us to determine where the PIC localizes for all promoter classes and, moreover, how PIC localization may be altered by Pol II mutants that alter TSS utilization. Previous work anchored ChiP-exo signal for PIC components over TATA or TATA-like sequences and identified a stereotypical overall pattern for crosslinks relative to these anchor positions, which were interpreted as relating to potential structure of the PIC open complex [7]. Subsequent work has identified that crosslinking in ChiP-exo can have some sequence bias [86] and this sequence bias may reflect partially the stereotypical crosslinking patterns observed around TATA/TATA-like sequences. Because the PIC must access TSSs downstream from the site of assembly, it is likely that observed ChiP-exo signal reflects the occupancies of PIC components across promoters and not just the site(s) of assembly. Using TATA-like sequences as anchors, Taf1-enriched promoters were found to have PIC components on average closer to TSSs than they were for Taf1-depleted promoters [7]. Here, we used our high resolution TSS mapping data coupled with determination of median position of ChiP-exo signal for Ssl2 or Sua7 within promoter windows to examine distance between putative PIC position and initiation zone as reflected by observed median TSSs (**Figure 5C**). We confirm that on average, ChiP-exo signal for PIC components is closer to median TSS position for Taf1-enriched promoters versus Taf1-depleted promoters.

We reasoned that if ChiP-exo signal for PIC components at least partially reflects promoter scanning, *i.e.* the interaction of PIC components with downstream DNA between PIC assembly position and zone of initiation, then Pol II mutants that alter TSS usage distribution should also alter PIC component distribution across promoters. As illustrated in **Figure 5C**, ChiP-exo signal for Sua7 (TFIIB) appears furthest upstream on the top DNA strand while signal for Ssl2 appears furthest downstream on the bottom DNA strand as expected for factors positioned at the upstream and downstream edges of the PIC. We observed modest changes to the aggregate distribution of ChiP-exo signal for both Taf1-enriched and Taf1-depleted promoter classes, with effects most obvious on the downstream edge of the PIC as detected by Ssl2 signal on the bottom strand of promoter DNA (**Figure 5D**). In single molecule experiments examining putative promoter scrunching in the Pol II PIC, scrunching behavior was similar regardless of whether all NTPs (to allow initiation) were present [87]. This observation suggested the possibility that

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

putative promoter scanning driven by TFIIH-mediated scrunching might be uncoupled to initiation, meaning that TFIIH translocation might continue independently of whether Pol II had initiated or not. However, we observed altered PIC component localization in Pol II mutants predicted to directly alter initiation efficiency but not necessarily other aspects of PIC function such as scanning (directly). Thus, there may in fact be coupling of initiation and scanning in vivo. Apparent coupling has been observed in magnetic tweezers experiments where a short unwinding event that is strictly TFIIH-dependent can be extended to a larger unwinding event by addition of NTPs, presumably reflecting Pol II transcription [88].

Relationships of TSS-selection altering initiation mutants with promoter architectural features

TSSs evolve at certain distances from the site of PIC assembly. This means that TSSs will be found at a range of distances from sites of initial assembly and will theoretically require scanning of different distances. We asked whether presumed scanning distance correlated with sensitivity to Pol II mutants for either TSS shifting or apparent promoter expression (**Figure 6**). We observed a modest correlation for TSS shifting extent based on where TSSs are relative to PIC location for Taf1-enriched promoters (**Figure 6A**), but a stronger correlation for Taf1-depleted promoters (**Figure 6B**) and the subset of Taf1-depleted promoters with consensus TATA boxes (**Figure 6C**). These latter promoters have putative PIC assembly points at greater distances from TSSs on average. Within the range of distances where most of these promoters have their TSSs, promoters with TSSs evolved at downstream positions show the greatest effects of upstream-shifting mutants on the TSS distribution (the TSS shift). Conversely, promoters with TSSs evolved at upstream positions show the greatest effects of downstream shifting mutants. These results are consistent with a facet of promoter architecture correlating with altered initiation activity but with potential upstream and downstream limiters on this sensitivity (see Discussion for more). We also asked if initiation mutants' effects on apparent expression, as measured by differential expression analysis for total TSS-seq reads within promoter windows could be related to promoter architecture. We used DEseq2 [89] to examine initiation mutant effects on putative expression relative to TSS-PIC distance (**Figure 6D, 6E**). That said, we observe divergent trends depending on class of initiation mutant, independent of promoter class, where Pol II LOF mutants exhibit a negative relationship on expression relative to WT as TSSs get closer to the position of the PIC while GOF initiation mutants (Pol II and *tfg2*) have the opposite trend (**Figure 6D, 6E**). Interestingly, *sua7-58A5* does not behave for gene expression similarly to Pol II LOF mutants, even though they are similar for other aspects of TSS defects. This correlation only explains a fraction of the changes in apparent RNA levels, but this is not unexpected. There are many reasons why gene expression will be altered in transcription factor mutants in addition to consequences directly from initiation defects (elongation or termination defects or secondarily from changes in cellular signaling and RNA stability). Our results suggest that either PIC-TSS distance or a correlated variable with promoter architecture may determine sensitivity of promoters to how initiation defects may contribute to gene expression defects. To examine if a strongly co-regulated gene class such as ribosomal protein genes might have an unequal distribution across TSS-PIC distances and thus drive some of the observed trend, we examined effects on gene expression for RP genes (**Supplemental Figure 8**). RP genes show the same trends as promoters overall for apparent expression vs. PIC-TSS distance even though they would be expected *a priori* to be co-regulated [90].

The majority of yeast promoters, especially the Taf1-enriched class, are found within a nucleosome depleted region (NDR) and flanked by an upstream -1 and a downstream +1 nucleosome. Previous work showed association between ChiP-exo for GTFs and +1 nucleosomes [7], and our data illustrate this as well (**Figure 7A**). We find that ChiP-exo for PIC

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

450 components track with nucleosome position with some flexibility. How the PIC recognizes these
451 promoters in the absence of a TATA-box is an open question. Our results are consistent with
452 the fact that TFIID has been found to interact with nucleosomes [91] and with the possibility that
453 the +1 nucleosome may be instructive for, or responsive to, PIC positioning. Nucleosomes have
454 previously been proposed as barriers to Pol II promoter scanning to explain the shorter distance
455 between PIC-component ChiP-exo footprints and TSSs at Taf1-enriched promoters [7].
456 Nucleosomes can be remodeled or be moved by transcription in yeast [8, 92], likely during
457 initiation as even for promoters with NDRs, TSSs can be found within the footprints of the +1
458 nucleosome. We do not observe a differential barrier for downstream shifting in Pol II or GTF
459 mutants at Taf1-enriched promoters, which have positioned nucleosomes (**Figure 2B**), thus it
460 remains unclear whether the +1 nucleosome can act as a barrier for Pol II scanning or TSS
461 selection from the existing data.

462 To determine if altered initiation and PIC positioning of Pol II mutants, especially downstream
463 shifting *rpb1* H1085Y, occurs in conjunction with altered +1 nucleosome positioning, we
464 performed MNase-seq in *rpb1* H1085Y and E1103G mutants along with a WT control strain
465 (**Figure 7B-I**). Determination of nucleosome positioning by MNase-seq can be sensitive to a
466 number of variables (discussed in [93]), therefore we isolated mononucleosomal DNA from a
467 range of digestion conditions and examined fragment length distributions in MNase-seq libraries
468 from a number of replicates (**Supplemental Figure 9A**) to ensure we had matched digestion
469 ranges for WT and mutant samples. We asked if +1 nucleosome midpoints were affected in
470 aggregate, if array spacing over genes was altered, or if individual +1 nucleosomes shifted on
471 average in Pol II mutants vs. WT. For H1085Y, we observed a slight but clear shift for the
472 aggregate +1 position (**Figure 7B**). Aligning genes of Taf1-enriched promoters by the +1
473 nucleosome position in WT suggests that H1085Y nucleosomes show increased spacing at the
474 +3, +4, and +5 positions relative to WT (**Figure 7B**). The downstream shift in aggregate +1
475 position also is reflected at the individual nucleosome level across H1085Y replicates (box plots,
476 **Figure 7C**). To ask if this effect on nucleosomes reflected a global defect across genes or
477 instead correlated with transcription (whether it be initiation or elongation), we performed the
478 same analyses on the top (**Figure 7D,E**) and bottom expression decile Taf1-enriched promoters
479 (**Figure 7F,G**). The downstream shift was apparent in top expression decile promoters but not in
480 bottom expression decile promoters, as would be predicted if the alteration were coupled to
481 transcription. For *rpb1* E1103G, we did not observe the same trend and effects were either
482 reduced or not present among all replicates (**Figure 7H,I**). To potentially identify subpopulations
483 of nucleosomes, we employed a more sophisticated analysis of nucleosomes using approach of
484 Zhou *et al* [93] (**Supplemental Figure 9B**). This approach recapitulated a similarly slight effect
485 of H1085Y on shifting the +1 nucleosome downstream across most H1085Y datasets relative to
486 WT.

487 **Comparisons of TSS distributions and properties in eukaryotes**

488 Properties of TSS distributions have been examined in a number of species and linked to
489 aspects of promoter architecture or expression behavior of underlying genes. Promoters are
490 termed “broad” or “dispersed” when many TSSs are utilized and “narrow” or “focused” when
491 TSSs are tightly clustered. The evolution of promoter sequence coupled with the mechanism of
492 TSS selection will be expected to shape TSS distributions at promoters, with potential outcomes
493 on promoter properties or transcript diversity. The broad preference for $Y_{-1}R_{+1}$ initiator elements
494 in eukaryotes is one defining factor for TSS selection. The presence and quality (consensus Inr
495 [1, 44], for example) or absence of these sites will contribute to TSS distributions at promoters.
496 The density of TSS usage will also be dependent on the mechanism of initiation. Promoter
497 scanning, which has only been strongly posited as an initiation mechanism in *S. cerevisiae*,
498 makes predictions about the density of observed TSSs relative to the density of potential $Y_{-1}R_{+1}$

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

initiator elements. Scanning predicts that YpR dinucleotides within a scanning window will be "seen" as potential initiators and potentially used by a scanning PIC. Conversely, if TSSs are individually specified by distinct, non-scanning PICs, YpR dinucleotides unlinked to PIC assembly will not be used, therefore $Y_{-1}R_{+1}$ usage relative to YpR dinucleotide density should be distinct for species utilizing different initiation mechanisms. To state differently, the density of TSS used to potential TSS motifs observed will be predictably different between a scanning mechanism a direct specification of an individual TSS or a small group of TSSs by an individual PIC. To determine if there were any commonalities between TSS distributions of *S. cerevisiae* and other eukaryotes, we compared them by a set of simple metrics (**Figure 8**).

First we examined the relationship between TSS spread and expression level in yeast (**Figure 8A**). Recent analyses in mouse and human have indicated that highly expressed promoters have more focused TSS distributions [94], beyond association with TATA elements, which can be associated with a single TSS. This relationship was suggested previously in yeast from analysis of a subset of native yeast promoters [80]. We find a similar correlation for both yeast promoter classes (**Figure 8A**) as found in mouse and human. We next asked whether promoter TSS spread distributions were similar between *S. cerevisiae*, *S. pombe* [95], *D. melanogaster* [39, 96], *D. rerio* (Zebrafish)[97], or Human (K562 cells)[98] (**Figure 8B**). With the caveat that detection method (steady state RNA analysis methods like TSS-seq or CAGE vs. nascent methods such as PRO-cap) might bias distributions, we find similar distributions for 101 nt promoter windows in all species with *S. pombe* being an outlier, where promoters are on average much narrower. The difference in distributions between CAGE [39] and PRO-cap [96] datasets in Drosophila could reflect source of RNA, but more likely reflect enrichment for lowly expressed promoters and enhancers in the PRO-cap dataset relative to highly expressed, focused promoters in the CAGE dataset. Within analyzed promoters across species, we wished to determine the relationship between YpR dinucleotide density on the top promoter strand and $Y_{-1}R_{+1}$ TSS usage (**Figure 8C-E**). We examined each dataset for fraction of TSSs that were $Y_{-1}R_{+1}$ on a promoter by promoter basis as well as fraction of sequencing reads that were $Y_{-1}R_{+1}$ (**Supplemental Figure 10A**). While our dataset showed the strongest $Y_{-1}R_{+1}$ preference, median $Y_{-1}R_{+1}$ site fraction was ~50% while median $Y_{-1}R_{+1}$ sequencing read fraction was ≥ 75% across species. We find that YpR dinucleotide density relative to promoter spread is similar across species (**Figure 8C**). The TSS usage relative to this density did differ among species indicating differences in TSS distributions within spreads (**Figure 8D**). For this analysis, we considered a $Y_{-1}R_{+1}$ site "used" if its usage was ≥2% of reads for that promoter. Zebrafish zygotic promoters were at least superficially similar to *S. cerevisiae* while Zebrafish maternal promoters and other species showed a lower rate of $Y_{-1}R_{+1}$ usage relative to YpR density. This is especially interesting because Zebrafish maternal and zygotic promoters can be closely positioned in the genome but maternal promoters are distinguishable by WW motif enrichment ~30 nt upstream of individual TSSs [97]. This positioning for individual TSSs is reminiscent of TATA-box positioning to TSSs in metazoans [1] and is consistent with individual TSS specification for maternal promoters. In contrast, Zebrafish zygotic promoters lack these WW elements. Promoters in a number of metazoans, especially mammals, can be CpG rich. While CpG is a YpR dinucleotide, it does not appear to be preferred – for most species the fraction of reads deriving from $C_{-1}G_{+1}$ sites was lower than fraction of TSSs deriving from CpGs (**Supplemental Figure 10B**). Therefore, we examined fraction non-CpG site used/present relative to promoter spread (**Figure 8E**). In this analysis, *S. cerevisiae*, human, and Zebrafish zygotic promoters were distinguished from Drosophila, *S. pombe*, and Zebrafish maternal promoters. These analyses suggest that TSS distributions in some eukaryotes beyond *S. cerevisiae* share attributes consistent with being derived from a scanning mechanism.

DISCUSSION

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Budding yeast has been a powerful model for understanding key mechanisms for transcription by Pol II. An early identified difference in promoter behavior for yeast TATA-containing promoters from classically studied TATA-containing human viral promoters such as adenovirus major late led to proposals that initiation mechanisms were fundamentally different between these species [47, 99]. TSSs for yeast TATA promoters were found downstream and spread among multiple positions while TSSs for viral and cellular TATA promoters were found to be tightly positioned ~31 nt downstream of the beginning of the element [48]. This positioning for TSSs at TATA promoters holds for many species including *S. pombe* [95] but not budding yeast. This being said, genome-wide studies of initiation indicate that the vast majority of promoters use multiple TSSs, though evolution appears to restrict TSS usage at highly expressed promoters in multiple species, including budding yeast (our work, [79, 94]. How these TSSs are generated and if by conserved or disparate mechanisms is a critical unanswered question in gene expression.

We have shown here that Pol II catalytic activity, as determined by mutations deep in the active in the essential "trigger loop", confer widespread changes in TSS distributions across the genome regardless of promoter type. Mutants in core Pol II GTFs TFIIB (*sua7* mutant) or TFIIF (*tfg2* mutant) confer defects of similar character to downstream shifting or upstream shifting Pol II alleles, respectively. The changes observed are consistent with a model wherein TSSs are displayed to the Pol II active site directionally from upstream to downstream, with the probability of initiation controlled by the display or scanning rate, and by Pol II catalytic rate. This system is analogous to a "shooting gallery" where targets (TSSs) move relative to a fixed firing position (the Pol II active site)[100]. In this model, Pol II catalytic activity, the rate of target movement, *i.e.* scanning rate, and the length of DNA that can be scanned *i.e.* scanning processivity, should all contribute to initiation probability at any particular sequence. Biochemical potential of any individual sequence will additionally contribute to initiation efficiency. Our results suggest that Pol II and tested GTF mutants affect initiation efficiency across sequence motifs and that differential effects in apparent motif usage genome-wide likely result from skewed distributions of bases within yeast promoters. Our in vivo results are consistent with elegant in vitro transcription experiments showing reduction of ATP levels (substrate for initiating base or for bases called for in very early elongation) confers downstream shifts in start site usage [101]. Reduction in substrate levels in vitro, therefore, is mimicked by reduction of catalytic activity in vivo.

How template sequence contributes to initiation beyond positions close to the template pyrimidine specifying the initial purine, and how they interact with scanning, is an open question. For models employing a scanning mechanism such as the "shooting gallery", it can be imagined that bases adjacent to the TSS affect TSS positioning to allow successful interaction with the first two NTPs, while distal bases such as the -8T on the template strand (-8A on the non-template strand) stabilize or are caught by interaction with the yeast TFIIB "reader" to hold TSSs in the active site longer during scanning [45]. Critical to this model are the structural studies just cited of Sainsbury *et al* on an artificial initial transcribing complex showing direct interaction of Sua7 D69 and R64 and -8T and -7T on the template strand. There are a number of ways TFIIB may alter initiation efficiency beyond recognition of upstream DNA. TFIIB has also been proposed by Sainsbury et al to allosterically affect Pol II active site Mg^{2+} binding and RNA-DNA hybrid positioning [45, 46]. Direct analysis of Kuehner and Brow [53] found evidence for lack of effect of *sua7* R64A on efficiency of one non--8A site, while -8A sites were affected, consistent with this residue functioning as proposed. We isolated individual motifs to examine efficiency (**Figure 4C**), and our tested *sua7-58A5* allele reduced efficiencies of both -8A and non--8A motifs alike. This allele contains a five-alanine insertion at position 58 in Sua7, likely reducing efficiency of the B-reader but possibly leaving some R64 interactions intact. Specific tests of Sua7 R64 mutants under controlled promoter conditions will directly address whether this

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

599 contact confers TSS selectivity. Additionally, altered selectivity alleles of Sua7 would be
600 predicted if interactions with the template strand were altered.

601 Core transcriptional machinery for Pol II initiation is highly conserved in eukaryotes leading to
602 the general expectation that key mechanisms for initiation will be conserved. While it has long
603 been believed that budding yeast represents a special case for initiation, this has not
604 systematically been addressed in eukaryotes. The question of how broadly conserved are
605 initiation mechanisms in eukaryotic gene expression is open for a number of reasons. There are
606 examples of diverse transcription mechanisms within organisms across development, for
607 example tissues, cells, or gene sets using TBP-related factors to replace TBP in initiation roles.
608 For example, in zebrafish, distinct core promoter “codes” have been described for genes that
609 are transcribed in oocytes (maternal transcription) versus those transcribed during zygotic
610 development (zygotic transcription) [97]. The maternal code is proposed to utilize an alternate
611 TBP for initiation, while zygotic promoters utilize TBP. Distinct core promoters are used to drive
612 maternal and zygotic expression. For genes transcribed both maternally and zygotically, distinct
613 TSS clusters specific to each phase of development can be quite close to one another in the
614 genome and may have superficially similar distribution characteristics, for example promoter
615 widths or spreads. However, individual maternal TSSs are each stereotypically positioned
616 relative to an upstream sequence motif, while individual zygotic TSSs are not. This difference
617 can be detected in the densities of TSSs used in a cluster. Our analysis indicates that
618 differential density between maternal and zygotic does not relate to differential YpR dinucleotide
619 density, and that zygotic TSS usage density in zebrafish appears similar to TSS usage density
620 in budding yeast, while maternal TSS usage density is distinct. The conservation of scanning
621 between organisms or for subsets of promoter classes within other eukaryotes is an open
622 question.

623 Another major question is how promoters without TATA-elements are specified. Organization of
624 PIC components is relatively stereotypical within a number of species, as detected by ChIP
625 methods for Pol II and GTFs [7, 102, 103], with the caveat that these are population-based
626 approaches. The most common organization for promoters across examined eukaryotes is a
627 NDR flanked by positioned nucleosomes. Such NDRs can support transcription bidirectionally,
628 reflecting a pair of core promoters with TSSs proximal to the flanking nucleosomes [13, 14, 17-
629 19, 104-106]. While sequence elements have been sought for these promoters, an alternate
630 attractive possibility is that NDR promoters use nucleosome positioning to instruct PIC
631 assembly. The association of TSSs with the edges of nucleosomes is striking across species,
632 though in species with high levels of promoter proximal pausing, nucleosomes may be
633 positioned downstream of the pause. Transcription itself has been linked to promoter
634 nucleosome positioning, turnover, or exchange in yeast. Given that MNase analyses reflects
635 bulk nucleosome populations, and depending on kinetics of initiation and the duration of
636 chromatin states supporting initiation (expected to be relatively infrequent), the nature of
637 initiating chromatin is unclear.

638 Finally, how does initiation interact with nucleosomes? In a scanning model, Pol II activity will
639 not be expected to control the interactions with the downstream nucleosome. Instead, TFIIH
640 bound to downstream DNA and translocating further downstream to power scanning, will be
641 expected to be the major interaction point of the PIC and the +1 nucleosome. This model
642 explains why downstream nucleosomes may not limit changes to scanning incurred by
643 alterations to Pol II activity, because Pol II will be acting downstream of the TFIIH-nucleosome
644 interaction. DNA translocation by TFIIH is expected to be competitive with the +1 nucleosome
645 for DNA as scanning proceeds into the territory of the nucleosome. Indeed, transcription and
646 TFIIH activity are proposed to drive H2A.Z exchange in the +1 nucleosome [92]. How TFIIH
647 activity is controlled to either allow scanning in addition to promoter opening or be restricted to

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

promoter opening is a major question in eukaryotic initiation. The *S. cerevisiae* CDK module of TFIIH has been implicated in restricting initiation close to the core promoter in vitro, but no in vivo evidence has emerged in vivo for this mechanism [107]. TFIIH components have long been implicated in controlling activities of the two ATPases – Ssl2 and Rad3 in yeast, XPB and XPD in humans – to enable or promote transcription or nucleotide excision repair [108-110]. These inputs may regulate activity of ATPases and their ability to be coupled to translocation activity analogous to paradigms for DNA translocase control in chromatin remodeling complexes [111].

METHODS

Yeast strains, plasmids, and oligonucleotides

Yeast strains used in this study were constructed as described previously [71-73]. Briefly, plasmids containing *rpo21/rpb1* mutants were introduced by transformation into a yeast strain containing a chromosomal deletion of *rpo21/rpb1* but with a wild type *RPO21/RPB1 URA3* plasmid. GTF mutant parental strains used for GTF single or GTF/Pol II double mutant analyses were constructed by chromosomal integration of GTF mutants into their respective native locus by way of two-step integrations [71]. Strains used in ChiP-exo were TAP-tagged at target genes (*SSL2*, *SUA7*) using homologous recombination of TAP tag amplicons (Puig et al., 2001) obtained from the yeast TAP-tag collection [112] (Open Biosystems) and transferred into our lab strain background [113]. All strains with mutations at chromosomal loci were verified by selectable marker, PCR genotyping, and sequencing. *rpo21/rpb1* mutants were introduced to parental strains with or without chromosomal locus mutation by plasmid shuffling [114], selecting for cells containing *rpo21/rpb1* mutant plasmids (Leu^+) in absence of the *RPB1* WT plasmid (Ura^-), thus generating single *rpo21/rpb1* mutation strain or double mutant strains combining mutations in GTF and *rpo21/rpb1* alleles. Yeast strains in all experiments were grown on YPD (1% yeast extract, 2% peptone, 2% dextrose) medium unless otherwise noted. Mutant plasmids for yeast promoter analyses were constructed by Quikchange mutagenesis (Stratagene) following adaptation for use of Phusion DNA polymerase (NEB) [115]. All oligonucleotides were obtained from IDT. Yeast strains, plasmids, and oligonucleotide sequences are described in **Additional File 1**.

Sample preparation for 5′-RNA sequencing

Yeast strains were diluted from a saturated overnight YPD culture and grown to mid-log phase (~$1.5x10^7$/ml) in YPD and harvested. Total RNA was extracted by a hot phenol-chloroform method [116], followed by on-column incubation with DNase I to remove DNA (RNeasy Mini kit, Qiagen), and processing with a RiboZero rRNA removal kit (Epicentre/Illumina) to deplete rRNA. To construct the cDNA library, samples were treated with Terminator 5′ phosphate-dependent exonuclease (Epicentre) to remove RNAs with 5′ monophosphate (5′ P) ends, and remaining RNAs were purified using acid phenol/chloroform pH 4.5 (Ambion) and precipitated. Tobacco acid pyrophosphatase (TAP, Epicentre) was added to convert 5′ PPP or capped RNAs to 5′ P RNAs. RNAs were purified using acid phenol/chloroform and a SOLiD 5′ adaptor was ligated to RNAs with 5′ P (this step excludes 5′ OH RNAs), followed by gel size selection of 5′ adaptor ligated RNAs and reverse transcription (SuperScript III RT, Invitrogen) with 3′ random priming. RNase H (Ambion) was added to remove the RNA strand of DNA-RNA duplexes, cDNA was size selected for 90-500 nt lengths. For SOLiD sequencing, these cDNA libraries were amplified using SOLiD total RNA-seq kit (Applied Biosystems) and SOLiD Barcoding kit (Applied Biosystems), final DNA was gel size selected for 160-300 nt length, and sequenced by SOLiD (Applied Biosystems) as described previously [117, 118].

5′-RNA sequencing data analyses

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SOLiD TSS raw data for libraries 446-465 was based on 35 nt short reads. The data were delivered in XSQ format and subsequently converted into Color Space csfasta format. Raw data for libraries VV497-520 was in FASTQ format. Multiple read files from each library were concatenated and aligned to *S. cerevisiae* R64-1-1 (SacCer3) reference genome from Saccharomyces Genome Database. We explored the possibility that alignments might be affected by miscalling of 5′ end base of the SOLiD reads. We trimmed one base at the 5′ end of the reads of the TSS libraries VV497-520, and aligned the trimmed reads independently from the raw reads for direct comparison. The alignment rates did not differ significantly, indicating 5′ end of our SOLiD libraries reads were not enriched for sequencing errors more than the rest of the reads. In addition, we asked if the alignment was affected by 5′ micro-exons in some *S. cerevisiae* genes by using TopHat to allow for potential splicing in alignment [119]. As we did not observe such splicing, we proceeded with Bowtie [120] allowing 2 mismatches but only retained uniquely mapped alignments. The aligned BAM files were converted to bedgraphs, and 5′ base (start tag) in each aligned read was extracted using Bedtools (v2.25.0) for downstream analyses [121]. Mapping statistics for TSS-seq, MNase-seq, and ChIP-exo libraries are described in **Additional File 2**.

To assess the correlation between biological replicates and different mutants, base-by-base coverage correlation between libraries was calculated for all bases genome-wide and for bases up and downstream of the promoter windows identified by [7](408 nt total width, described below). Given that Pearson correlation is sensitive to variability at lower coverage levels, correlations for positions above arbitrarily chosen thresholds (3 reads or 10 reads per position) were calculated and 3 read threshold correlations are shown. Heat scatter plots were generated the LSD R package (4.0-0) (cite Schwalb, Bjoern, Achim Tresch, and Romain Francois. "LSD Lots of Superior Depictions." *The Comprehensive R Archive Network* (2011)) and compiled in Adobe Photoshop. Heatmaps were generated using Morpheus (https://software.broadinstitute.org/morpheus/) or JavaTreeView [122] and Cluster [123].

To create base-by-base coverage in selected windows of interest, computeMatrix reference-point() function from the deepTools package (2.1.0) was used [124]. There were two types of windows of interest. First, the promoter windows were established by expanding 200 nt up and downstream from the TATA/TATA-like elements identified by [7] (here we term them TATA/TATA-like centered windows) (408 nt total width). Most of these windows (5945/6044) were centered on TATA/TATA-like element annotated in [7], while 99 promoters did not have annotated TATA/TATA-like element and were centered on the TFIIB ChiP-exo peak. Second, we established windows centered on transcription start sites (TSSs) to investigate TSSs at promoters in a core promoter element-independent manner (here we term them TSS-anchored windows). For the TSS anchored windows, we first determined the 50th percentile (median) TSS (see next paragraph for details) in the TATA/TATA-like centered promoter windows with WT TSS reads derived from *RPB1* WT libraries 446, 456, 497, and 499 (see below) and expanded 200 nt upstream and 200 nt downstream from this “median” TSS position (401 nt total width), adjusting this window one time based on new TSSs potentially present after shifting the window, and then displaying 250 nt upstream and 150 nt downstream from the median TSS position.

Several characteristics of TSS utilization were calculated as following: (1) The position of the TSS containing the 50th percentile of reads in the window and was termed the “median” TSS. (2) Distance between 10th percentile and 90th percentile TSS position in each promoter was used to measure the width of the TSS distribution, termed the “TSS Spread”. Specifically, TSS positions with 10th and 90th percentile reads were determined in a directional fashion (from upstream to downstream), the absolute value of the difference between two positions by

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

subtraction was calculated as "TSS Spread". (3) Total reads in windows of interest were summed as a measurement of apparent expression. (4) Normalized densities in windows were calculated as fraction of reads at each TSS position relative to the total number of reads in the window. The normalized densities were subsequently used for examination of TSS usage distribution at each promoter independent of expression level, comparison among different libraries, and start site usage pattern changes in mutants, and visualization. (5) Differential expression analyses of promoter expression determined from (3) were computed by DESeq2 [89]. The log2 fold changes (log2FC) of the expression between mutant and wild type were taken from the output of DESeq2 analysis and used to examine the dependency on TSS distance to PIC. Lowly expressed promoters with fewer than 100 reads from the sum of four WT replicates were filtered out given their high coefficient of variations. We observed that replicates of each strain (WT or mutant) were highly correlated at the base coverage level as well as primary characteristics of TSS usage (distance to core promoter element, apparent expression). We therefore aggregated the counts from replicate strains for downstream analyses (*i.e.*, aligned reads for all replicates of each strain were combined and treated as single "merged library"). Mutant vs WT relative changes of median TSS (**Figure 1E**), TSS spread and normalized TSS densities (**Figure 2**) in the indicated windows are calculated in R and visualized in Morpheus or Graphpad Prism 8.

In the TSS motif analyses, two major characteristics were computed. First was TSS usage defined by the number of reads at each TSS divided by the total number of reads in the promoter window. Second, we calculated TSS efficiency by dividing TSS reads at an individual position by the reads at or downstream of the TSS, as a proxy to estimate how well each TSS gets utilized with regard to the available Pol II (TSS efficiency)[53]. TSS positions with ≥20% efficiency calculated with ≤ 5 reads were excluded (which definitionally are only found at the downstream edges of windows). The corresponding −8, −1, +1 position underlying each TSS ($N_{-8}N_{-1}N_{+1}$ motif) was extracted by Bedtools getfasta (v2.25.0). Start site motif compilation was done by WebLogo for indicated groups of TSSs. Reads for each $N_{-8}N_{-1}N_{+1}$ motif of interest were summed, and fraction of the corresponding motif usage in total TSS reads was calculated for each library. Differences of fraction of start site motif usage in WT and mutants were calculated by subtracting the WT usage fraction from that in each mutant.

ChiP-exo sequencing

ChiP-exo experiments were performed as described previously[85, 125]. Briefly, yeast strains were amplified to mid-log phase (~$1.5X10^7$/ml) in rich medium (YPD) from a saturated overnight culture, crosslinked with formaldehyde (1% final concentration from 37% formaldehyde solution, (Mallinckrodt)) for 20 min and then quenched by 0.25M glycine (from 2.5M stock, pH 7). Cells were washed, lysed with glass beads by beat beating (30s on 60s off, 7-8 rounds), visually inspected under microscope, followed by shearing of chromatin to ~200-500 nt fragments by sonication at 4 °C (Diagenode Bioruptor, 30s on 30s off, high power, 18 cycles or until desired size reached) in FA-lysis buffer without detergents. Solubilized chromatin was Immunoprecipitated using IgG-bound sepharose resin and washed. Immunoprecipitated chromatin on resin was end polished by T4 DNA polymerase (NEB). Adaptor sequences were subsequently ligated on both ends of linker DNA, followed by nick repair using phi29 polymerase (NEB). λ exonuclease (NEB) was used for 5′ → 3′ digestion of sonicated protein-dsDNA and RecJ exonuclease for 5′ → 3′ digestion of ssDNA to minimize background. Resin was washed, DNA was eluted with TEV protease (Invitrogen), and crosslinks reversed by incubation at 65 °C with Proteinase K (Roche). DNA was extracted by phenol:chloroform:isoamyl alcohol (each overhang of dsDNA was different and corresponded to one border of protein binding), denatured to ssDNA, amplified to dsDNA with oligos with -OH on both ends to amplify exonuclease treated strand only. Adaptors were ligated only to λ

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exonuclease digested ends of dsDNA, ligated DNA was then amplified, size selected for 120−160 nt and were sequenced by SOLiD sequencing (Applied Biosystems)

First replicate of ChiP-exo reads were aligned to V56 reference genome by Corona Lite software provided by SOLiD allowing up to 3 mismatches. Only uniquely aligned reads were kept. ChiP-exo reads alignments were converted to SacCer3 (V64) genome alignment using the LiftOver tool (http://genome.ucsc.edu/) for downstream analyses. Second replicate of ChiP-exo reads were aligned to V64 genome. 5′ base of each aligned reads were extracted to a tabular coverage file in a strand conscious way, and then converted to bed format. We separated the ChiP-exo reads on two strands relative to the promoters in study: TOP strand (reads on the same strand the direction of promoter transcription, thus TOP may map to either Watson (forward) strand or Crick (reverse) strand on genome) and BOTTOM strand (opposite strand of TOP strand). Replicates of ChiP-exo reads for each strain were compared by base-by-base (after 10-read thresholding) correlation. Characteristics of ChiP-exo reads distribution (median position, levels) in promoter windows were also compared, and we concluded that replicates showed reasonable correlations, considering variability in ChiP-exo reads base-by-base between replicates. Therefore, we merged aligned reads from replicates as a single library for downstream analyses. For each promoter, ChiP-exo 5′ tags were assigned to TOP or BOTTOM strand at each promoter window and analyzed separately. Base-by-base read density in selected windows of interest was computed by computeMatrix reference-point() function from deepTools package (2.1.0)[124], similar to our TSS analyses above. The read density at each position was subsequently normalized to total reads in the window (we termed normalized density). ChiP-exo “median” position and total reads in each window were calculated similarly with those of TSS reads, as described above. Median ChiP-exo position was used as a proxy for core promoter element position in Taf1-enriched promoter classes where no proven functional promoter motif has been identified and tested.

Nucleosome MNase sequencing

Nucleosomal DNAs were prepared by a method described elsewhere [126] with the following modifications. Yeast strains were grown in rich medium (YPD) to mid-log phase (~1.5X10^7/ml) and cross-linked with methanol-free formaldehyde (1% final concentration, Polysciences Inc) for 30 min and quenched with 0.25M final concentration of glycine (from 2.5M stock, pH 7). Cells were washed and digested with zymolyase-20T (Sunrise International) (6mg for 500ml culture) for ~17 min or until ~90% cells appeared as spheroplasts, followed by MNase (Thermo Fisher Scientific) digestion with different amount of MNase to generate “less” and “more” digested nucleosomes (in general, digests were limited such that at least mono, di, and trinucleosomes were still apparent after agarose gel electrophoresis). Crosslinks on nucleosomes were reversed at 65 °C in the presence of Proteinase K (G-Biosciences) overnight. DNA was extracted by phenol/chloroform, and digested with RNase A (Thermo Fisher Scientific) to remove RNAs. Nucleosomal DNA was separated on 1.5% agarose gels containing SYBR gold dye (Thermo Fisher Scientific) and mono-nucleosome bands were identified and selected under blue light and gel purified (Omega Biotek). Mononucleosomal DNA fragments were sequenced on an Illumina HiSeq 2500 instrument (2x125 paired-end sequencing). Paired-end nucleosome reads were aligned to V64 (SacCer3) reference genome using Bowtie2 [127] allowing 1 mismatch, with only uniquely mapped alignments are kept. We used Samtools [128] to extract the alignments to build genome coverage for visualization and start and end position of sequenced DNA fragments. Using the start and end positions of each fragments, fragment length and midpoint position of each fragment were calculated.

Midpoints were analyzed in two main windows of interest. First was median TSS centered window (−250 upstream and +150 downstream based on median TSS position as above).

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Second, windows were identified based on determined WT +1 nucleosome peak position, as described below using custom scripts (NucSeq v1.0)[129]. Midpoints were assigned to relative coordinates of the window and smoothed using a triweight kernel (75 nt up/downstream total width with a uniform kernel with 5 nt up/downstream width) to get a "smoothed" midpoints profile. The nucleosome peak was called by identifying the local maximum using the smoothed profile. This method enabled us to call a single peak position in ranges of 150 nt windows using the smoothed nucleosome midpoints profiles, thus determining one peak per nucleosome. Average chromosomal coverage (sum of raw midpoints divided by current chromosome length) was calculated for each chromosome as a read threshold per position. The first peak downstream of the median TSS position that had larger than or equal to 20% of chromosomal average coverage and was also within a reasonable position range for a +1 nucleosome was annotated as the +1 nucleosome peak at each promoter (if present). +1 nucleosome peaks were separately identified in 4 WT libraries (replicates for "less" and "more" digested chromatin), The replicates for "less" digested WT +1 nucleosome peaks showed greater correlation. Therefore, we took the average of +1 nucleosome peaks between two "less" digested WT libraries and used as the center for +1 nucleosome-based window. 500 nt up/downstream of these base positions led to 5660 +1 nucleosome centered 1001 nt wide windows, allowing observation of up to 8 nucleosomes surrounding +1 nucleosomes. Nucleosome midpoints were subsequently assigned to this window using the same method as above. Aggregated nucleosome midpoints analysis was done by sorting the promoters by promoter class, expression level (TSS reads in window) followed by summing the nucleosome midpoint counts at each position in the window.

Analysis of TSS usage in different species

Genome-wide TSS sequencing datasets analyzed for different eukaryotes were: CAGE sequencing in Drosophila by the Celniker lab [39], PROcap sequencing in Drosophila by Lis lab [96]; GROcap sequencing in human by the Lis lab [98]; CAGE sequencing in zebrafish by the Lenhard lab [97]; and DeepCAGE sequencing in *S. pombe* by the Shao lab [95]. Aligned start site reads data were retrieved from read archives or authors in various formats (tab, bigwig, wig *etc*.). Replicate libraries were merged and treated as one library after determining they were well-correlated with each other. Start site tags files were formatted to Bed6 format (UCSC) and major start sites determined by authors were used to create 101 nt TSS centered windows for each promoter in a strand conscious manner. When a major TSS cluster was identified instead of single start site at each promoter by the authors' respective analyses, the midpoint of the TSS cluster was used as the center of the window. Start site tags were then mapped into TSS-centered promoter windows using a custom R script. Separate files containing underlying sequence in the same window of each promoter from appropriate reference genome were created for each library.

Because $Y_{-1}R_{+1}$ start site motifs are strongly preferred by Pol II, we specifically analyzed each YR motif separately and in aggregate. $Y_{-1}R_{+1}$ usage and "spread" calculation were computed as described above. Positions that had ≥ 2% of the total reads in the window were considered as meaningful TSSs, which were further classified into $C_{-1}G_{+1}$, non-CG $Y_{-1}R_{+1}$ and non- $Y_{-1}R_{+1}$ based on TSS base and the preceding deoxynucleotide. Only TSSs within the "spread" (10^{th} to 90^{th} percentile reads in the promoter windows) were used for subsequent analyses and plotting. XY scatter plots and the linear regression fits were generated using GraphPad Prism 7.

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consent for publication

Not applicable

Availability of data and materials

Genomics datasets generated in the current study are available in the NCBI BioProject, under the accession number PRJNA522619. Promoters analyzed in yeast, genomic positions, and attributes (ChIP-exo median positions and +1 nucleosome positions) are described in **Additional File 3**. Published TSS datasets re-analyzed here may be accessed at: ArrayExpress accession E-MTAB-3188 (*S. pombe*, [95]), Sequence Read Archive accession SRX015329 (*D. melanogaster*, [39]), Gene Expression Omnibus (GEO) under accession GSE42117 (*D. melanogaster*, [96]), Sequence Read Archive SRA097279 and SRA104816 (*Danio rerio*, [97]), GEO under accession GSE60456 (Human, [98]). Coordinates for promoter windows in other species are described in **Additional Files 4-7.**

Competing interests

B.F.P. has a financial interest in Peconic, LLC, which utilizes the ChIP-exo technology implemented in this study and could potentially benefit from the outcomes of this research. All other authors declare that they have no competing interests.

Funding

Initial funding for this project was provided by grants from the National Institutes of Health R01GM097260 and Welch Foundation A-1763 to C.D.K. We acknowledge funding from NIH R01GM120450 to C.D.K. and R35GM118059 to B.E.N.

Authors' contributions

C.Q. analyzed data, made figures, contributed to writing the manuscript. H.J. initiated project, generated strains, prepared material for TSS-seq, generated material and libraries for MNase-seq, analyzed data, piloted most informatics approaches, and generated outline of the manuscript. I.V. generated libraries for TSS-seq. P.Č. generated strains for ChIP-exo analyses. J.A.L. collaborated with H.J. on nucleosome positioning analyses and generated scripts and code for the analyses. T.Z. provided informatics analysis of TSS-seq data. I.M. constructed strains and performed Northern blotting for promoter variant studies. S.S. initiated informatics analyses for TSS-seq in yeast. P.Č. constructed strains and performed Northern blotting for promoter variant studies. K.H.H. prepared ChIP-exo samples for sequencing. R.P.M. and C.D.J consulted on Illumina sequencing strategies and library preparation. S-H.Z. implemented MNase analyses as described in [93]. B.F.P. consulted on ChIP-exo and enabled sequencing of ChIP-exo samples. B.E.N provided funding and consulted on development of TSS-seq for yeast Pol II RNAs. C.D.K conceived the project, guided analyses, made figures and wrote the manuscript. All others read and approved the final manuscript except for S.S. who was unable to be contacted.

Acknowledgements

Mahmoud Bassal and Kaplan lab members are acknowledged for discussions and comments on the manuscript.

REFERENCES

1. Vo Ngoc L, Wang YL, Kassavetis GA, Kadonaga JT: **The punctilious RNA polymerase II core promoter.** *Genes Dev* 2017, **31:**1289-1301.

2. Kadonaga JT: **Perspectives on the RNA polymerase II core promoter.** *Wiley Interdiscip Rev Dev Biol* 2012, **1:**40-51.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Juven-Gershon T, Kadonaga JT: **Regulation of gene expression via the core promoter and the basal transcriptional machinery.** *Dev Biol* 2010, **339:**225-229.

4. Juven-Gershon T, Hsu JY, Kadonaga JT: **Perspectives on the RNA polymerase II core promoter.** *Biochem Soc Trans* 2006, **34:**1047-1050.

5. Smale ST, Kadonaga JT: **The RNA polymerase II core promoter.** *Annu Rev Biochem* 2003, **72:**449-479.

6. Butler JE, Kadonaga JT: **The RNA polymerase II core promoter: a key component in the regulation of gene expression.** *Genes Dev* 2002, **16:**2583-2592.

7. Rhee HS, Pugh BF: **Genome-wide structure and organization of eukaryotic pre-initiation complexes.** *Nature* 2012, **483:**295-301.

8. Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N: **High-resolution nucleosome mapping reveals transcription-dependent promoter packaging.** *Genome Res* 2010, **20:**90-100.

9. Tirosh I, Barkai N, Verstrepen KJ: **Promoter architecture and the evolvability of gene expression.** *J Biol* 2009, **8:**95.

10. Jiang C, Pugh BF: **A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome.** *Genome Biol* 2009, **10:**R109.

11. Tirosh I, Barkai N: **Two strategies for gene regulation by promoter nucleosomes.** *Genome Res* 2008, **18:**1084-1091.

12. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, et al: **Nucleosome organization in the Drosophila genome.** *Nature* 2008, **453:**358-362.

13. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM: **Bidirectional promoters generate pervasive transcription in yeast.** *Nature* 2009, **457:**1033-1037.

14. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A: **Widespread bidirectional promoters are the major source of cryptic transcripts in yeast.** *Nature* 2009, **457:**1038-1042.

15. Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH: **RNA exosome depletion reveals transcription upstream of active human promoters.** *Science* 2008, **322:**1851-1854.

16. Core LJ, Lis JT: **Transcription regulation through promoter-proximal pausing of RNA polymerase II.** *Science* 2008, **319:**1791-1792.

17. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA: **Divergent transcription from active promoters.** *Science* 2008, **322:**1849-1851.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Jin Y, Eser U, Struhl K, Churchman LS: **The Ground State and Evolution of Promoter Region Directionality.** *Cell* 2017, **170:**889-898 e810.

19. Andersson R, Chen Y, Core L, Lis JT, Sandelin A, Jensen TH: **Human Gene Promoters Are Intrinsically Bidirectional.** *Mol Cell* 2015, **60:**346-347.

20. Duttke SH, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U: **Human promoters are intrinsically directional.** *Mol Cell* 2015, **57:**674-684.

21. Duttke SH, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U: **Perspectives on Unidirectional versus Divergent Transcription.** *Mol Cell* 2015, **60:**348-349.

22. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C: **A high-resolution atlas of nucleosome occupancy in yeast.** *Nat Genet* 2007, **39:**1235-1244.

23. Huisinga KL, Pugh BF: **A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae.** *Mol Cell* 2004, **13:**573-585.

24. Basehoar AD, Zanton SJ, Pugh BF: **Identification and distinct regulation of yeast TATA box-containing genes.** *Cell* 2004, **116:**699-709.

25. Kuras L, Kosa P, Mencia M, Struhl K: **TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo.** *Science* 2000, **288:**1244-1248.

26. Li XY, Bhaumik SR, Green MR: **Distinct classes of yeast promoters revealed by differential TAF recruitment.** *Science* 2000, **288:**1242-1244.

27. Wu R, Li H: **Positioned and G/C-capped poly(dA:dT) tracts associate with the centers of nucleosome-free regions in yeast promoters.** *Genome Res* 2010, **20:**473-484.

28. Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, Soding J, Cramer P: **A conserved GA element in TATA-less RNA polymerase II promoters.** *PLoS One* 2011, **6:**e27595.

29. Vo Ngoc L, Kassavetis GA, Kadonaga JT: **The RNA Polymerase II Core Promoter in Drosophila.** *Genetics* 2019, **212:**13-24.

30. Warfield L, Ramachandran S, Baptista T, Devys D, Tora L, Hahn S: **Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID.** *Mol Cell* 2017, **68:**118-129 e115.

31. Baptista T, Grunberg S, Minoungou N, Koster MJE, Timmers HTM, Hahn S, Devys D, Tora L: **SAGA Is a General Cofactor for RNA Polymerase II Transcription.** *Mol Cell* 2017, **68:**130-143 e135.

32. Hampsey M: **The Pol II initiation complex: finding a place to start.** *Nat Struct Mol Biol* 2006, **13:**564-566.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Corden JL: **Yeast Pol II start-site selection: the long and the short of it.** *EMBO reports* 2008, **9:**1084-1086.

34. Zhang Z, Dietrich FS: **Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE.** *Nucleic acids research* 2005, **33:**2838-2851.

35. Park D, Morris AR, Battenhouse A, Iyer VR: **Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements.** *Nucleic Acids Res* 2014, **42:**3736-3749.

36. Pelechano V, Wei W, Steinmetz LM: **Extensive transcriptional heterogeneity revealed by isoform profiling.** *Nature* 2013, **497:**127-131.

37. Chen RA, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LW, Jeffers TE, Ahringer J: **The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures.** *Genome Res* 2013, **23:**1339-1347.

38. Yamashita R, Sathira NP, Kanai A, Tanimoto K, Arauchi T, Tanaka Y, Hashimoto S, Sugano S, Nakai K, Suzuki Y: **Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis.** *Genome Res* 2011, **21:**775-789.

39. Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, Yu C, Booth BW, Zhang D, Wan KH, et al: **Genome-wide analysis of promoter architecture in Drosophila melanogaster.** *Genome Res* 2011, **21:**182-192.

40. Consortium F, the RP, Clst, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, et al: **A promoter-level mammalian expression atlas.** *Nature* 2014, **507:**462-470.

41. Nepal C, Hadzhiev Y, Previti C, Haberle V, Li N, Takahashi H, Suzuki AM, Sheng Y, Abdelhamid RF, Anand S, et al: **Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis.** *Genome Res* 2013, **23:**1938-1950.

42. Gleghorn ML, Davydova EK, Basu R, Rothman-Denes LB, Murakami KS: **X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides.** *Proc Natl Acad Sci U S A* 2011, **108:**3566-3571.

43. Smale ST, Baltimore D: **The "initiator" as a transcription control element.** *Cell* 1989, **57:**103-113.

44. Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SH, Kadonaga JT: **The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters.** *Genes Dev* 2017, **31:**6-11.

45. Sainsbury S, Niesser J, Cramer P: **Structure and function of the initially transcribing RNA polymerase II-TFIIB complex.** *Nature* 2013, **493:**437-440.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

46. Sainsbury S, Bernecky C, Cramer P: **Structural basis of transcription initiation by RNA polymerase II.** *Nat Rev Mol Cell Biol* 2015, **16:**129-143.

47. Struhl K: **Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast.** *Cell* 1987, **49:**295-297.

48. Breathnach R, Chambon P: **Organization and expression of eucaryotic split genes coding for proteins.** *Annu Rev Biochem* 1981, **50:**349-383.

49. Giardina C, Lis JT: **DNA melting on yeast RNA polymerase II promoters.** *Science* 1993, **261:**759-762.

50. Yang C, Ponticelli AS: **Evidence that RNA polymerase II and not TFIIB is responsible for the difference in transcription initiation patterns between Saccharomyces cerevisiae and Schizosaccharomyces pombe.** *Nucleic Acids Res* 2012, **40:**6495-6507.

51. Goel S, Krishnamurthy S, Hampsey M: **Mechanism of start site selection by RNA polymerase II: interplay between TFIIB and Ssl2/XPB helicase subunit of TFIIH.** *J Biol Chem* 2012, **287:**557-567.

52. Khaperskyy DA, Ammerman ML, Majovski RC, Ponticelli AS: **Functions of Saccharomyces cerevisiae TFIIF during transcription start site utilization.** *Mol Cell Biol* 2008, **28:**3757-3766.

53. Kuehner JN, Brow DA: **Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model.** *J Biol Chem* 2006, **281:**14119-14128.

54. Pal M, Ponticelli AS, Luse DS: **The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II.** *Mol Cell* 2005, **19:**101-110.

55. Majovski RC, Khaperskyy DA, Ghazy MA, Ponticelli AS: **A functional role for the switch 2 region of yeast RNA polymerase II in transcription start site utilization and abortive initiation.** *J Biol Chem* 2005, **280:**34917-34923.

56. Freire-Picos MA, Krishnamurthy S, Sun ZW, Hampsey M: **Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex.** *Nucleic acids research* 2005, **33:**5045-5052.

57. Ghazy MA, Brodie SA, Ammerman ML, Ziegler LM, Ponticelli AS: **Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II.** *Mol Cell Biol* 2004, **24:**10975-10985.

58. Chen BS, Hampsey M: **Functional interaction between TFIIB and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation.** *Mol Cell Biol* 2004, **24:**3983-3991.

59. Faitar SL, Brodie SA, Ponticelli AS: **Promoter-specific shifts in transcription initiation conferred by yeast TFIIB mutations are determined by the sequence in the immediate vicinity of the start sites.** *Mol Cell Biol* 2001, **21:**4427-4440.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

60. Pappas DL, Jr., Hampsey M: **Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in Saccharomyces cerevisiae.** *Mol Cell Biol* 2000, **20:**8343-8351.

61. Wu WH, Pinto I, Chen BS, Hampsey M: **Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB.** *Genetics* 1999, **153:**643-652.

62. Bangur CS, Faitar SL, Folster JP, Ponticelli AS: **An interaction between the N-terminal region and the core domain of yeast TFIIB promotes the formation of TATA-binding protein-TFIIB-DNA complexes.** *The Journal of biological chemistry* 1999, **274:**23203-23209.

63. Pardee TS, Bangur CS, Ponticelli AS: **The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection.** *The Journal of biological chemistry* 1998, **273:**17859-17864.

64. Sun ZW, Tessmer A, Hampsey M: **Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in Saccharomyces cerevisiae.** *Nucleic acids research* 1996, **24:**2560-2566.

65. Sun ZW, Hampsey M: **Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in Saccharomyces cerevisiae.** *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92:**3127-3131.

66. Pinto I, Wu WH, Na JG, Hampsey M: **Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast.** *J Biol Chem* 1994, **269:**30569-30573.

67. Berroteran RW, Ware DE, Hampsey M: **The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations.** *Mol Cell Biol* 1994, **14:**226-237.

68. Pinto I, Ware DE, Hampsey M: **The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo.** *Cell* 1992, **68:**977-988.

69. Hampsey M, Na JG, Pinto I, Ware DE, Berroteran RW: **Extragenic suppressors of a translation initiation defect in the cyc1 gene of Saccharomyces cerevisiae.** *Biochimie* 1991, **73:**1445-1455.

70. Knaus R, Pollock R, Guarente L: **Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4.** *Embo J* 1996, **15:**1933-1940.

71. Jin H, Kaplan CD: **Relationships of RNA polymerase II genetic interactors to transcription start site usage defects and growth in Saccharomyces cerevisiae.** *G3 (Bethesda)* 2014, **5:**21-33.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

72. Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, et al: **From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II.** *Cell* 2013, **154:**775-788.

73. Kaplan CD, Jin H, Zhang IL, Belyanin A: **Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo.** *PLoS Genet* 2012, **8:**e1002627.

74. Eichner J, Chen HT, Warfield L, Hahn S: **Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex.** *EMBO J* 2010, **29:**706-716.

75. Kaplan CD, Larsson KM, Kornberg RD: **The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin.** *Mol Cell* 2008, **30:**547-556.

76. Kireeva ML, Nedialkov YA, Cremona GH, Purtov YA, Lubkowska L, Malagon F, Burton ZF, Strathern JN, Kashlev M: **Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation.** *Mol Cell* 2008, **30:**557-566.

77. Malagon F, Kireeva ML, Shafer BK, Lubkowska L, Kashlev M, Strathern JN: **Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil.** *Genetics* 2006, **172:**2201-2209.

78. Qiu C, Erinne OC, Dave JM, Cui P, Jin H, Muthukrishnan N, Tang LK, Babu SG, Lam KC, Vandeventer PJ, et al: **High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop.** *PLoS Genet* 2016, **12:**e1006321.

79. Lubliner S, Keren L, Segal E: **Sequence features of yeast and human core promoters that are predictive of maximal promoter activity.** *Nucleic Acids Res* 2013, **41:**5569-5581.

80. Lubliner S, Regev I, Lotan-Pompan M, Edelheit S, Weinberger A, Segal E: **Core promoter sequence in yeast is a major determinant of expression level.** *Genome Res* 2015, **25:**1008-1017.

81. Kamenova I, Warfield L, Hahn S: **Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription.** *Mol Cell Biol* 2014, **34:**2929-2943.

82. Donczew R, Hahn S: **Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters.** *Mol Cell Biol* 2018, **38**.

83. Lorch Y, Maier-Davis B, Kornberg RD: **Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions.** *Genes Dev* 2014, **28:**2492-2497.

84. Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF, Korber P: **Genomic Nucleosome Organization Reconstituted with Pure Proteins.** *Cell* 2016, **167:**709-721 e712.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

85. Rhee HS, Pugh BF: **ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy.** *Curr Protoc Mol Biol* 2012, **Chapter 21:**Unit 21 24.

86. Rossi MJ, Lai WKM, Pugh BF: **Genome-wide determinants of sequence-specific DNA binding of general regulatory factors.** *Genome Res* 2018, **28:**497-508.

87. Fazal FM, Meng CA, Murakami K, Kornberg RD, Block SM: **Real-time observation of the initiation of RNA polymerase II transcription.** *Nature* 2015, **525:**274-277.

88. Tomko EJ, Fishburn J, Hahn S, Galburt EA: **TFIIH generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning.** *Nat Struct Mol Biol* 2017, **24:**1139-1145.

89. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15:**550.

90. Warner JR: **The economics of ribosome biosynthesis in yeast.** *Trends Biochem Sci* 1999, **24:**437-440.

91. Bhuiyan T, Timmers HTM: **Promoter Recognition: Putting TFIID on the Spot.** *Trends Cell Biol* 2019.

92. Tramantano M, Sun L, Au C, Labuz D, Liu Z, Chou M, Shen C, Luk E: **Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex.** *Elife* 2016, **5**.

93. Zhou X, Blocker AW, Airoldi EM, O'Shea EK: **A computational approach to map nucleosome positions and alternative chromatin states with base pair resolution.** *Elife* 2016, **5**.

94. Xu C, Park JK, Zhang J: **Evidence that alternative transcriptional initiation is largely nonadaptive.** *PLoS Biol* 2019, **17:**e3000197.

95. Li H, Hou J, Bai L, Hu C, Tong P, Kang Y, Zhao X, Shao Z: **Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE.** *RNA Biol* 2015, **12:**525-537.

96. Kwak H, Fuda NJ, Core LJ, Lis JT: **Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.** *Science* 2013, **339:**950-953.

97. Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, Gehrig J, Dong X, Akalin A, Suzuki AM, et al: **Two independent transcription initiation codes overlap on vertebrate core promoters.** *Nature* 2014, **507:**381-385.

98. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT: **Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers.** *Nat Genet* 2014, **46:**1311-1320.

99. Struhl K: **Molecular mechanisms of transcriptional regulation in yeast.** *Annual review of biochemistry* 1989, **58:**1051-1077.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

100. Kaplan CD: **Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae.** *Biochim Biophys Acta* 2013, **1829:**39-54.

101. Fishburn J, Galburt E, Hahn S: **Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II.** *J Biol Chem* 2016, **291:**13040-13047.

102. Lai WK, Pugh BF: **Genome-wide uniformity of human 'open' pre-initiation complexes.** *Genome Res* 2017, **27:**15-26.

103. Shao W, Zeitlinger J: **Paused RNA polymerase II inhibits new transcriptional initiation.** *Nat Genet* 2017, **49:**1045-1051.

104. Scruggs BS, Gilchrist DA, Nechaev S, Muse GW, Burkholder A, Fargo DC, Adelman K: **Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin.** *Mol Cell* 2015, **58:**1101-1112.

105. Kaplan CD: **Pairs of promoter pairs in a web of transcription.** *Nat Genet* 2016, **48:**975-976.

106. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322:**1845-1848.

107. Murakami K, Mattei PJ, Davis RE, Jin H, Kaplan CD, Kornberg RD: **Uncoupling Promoter Opening from Start-Site Scanning.** *Mol Cell* 2015, **59:**133-138.

108. Singh A, Compe E, Le May N, Egly JM: **TFIIH subunit alterations causing xeroderma pigmentosum and trichothiodystrophy specifically disturb several steps during transcription.** *Am J Hum Genet* 2015, **96:**194-207.

109. Egly JM, Coin F: **A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor.** *DNA Repair (Amst)* 2011, **10:**714-721.

110. Compe E, Egly JM: **TFIIH: when transcription met DNA repair.** *Nat Rev Mol Cell Biol* 2012, **13:**343-354.

111. Clapier CR, Iwasa J, Cairns BR, Peterson CL: **Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes.** *Nat Rev Mol Cell Biol* 2017, **18:**407-422.

112. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS: **Global analysis of protein expression in yeast.** *Nature* 2003, **425:**737-741.

113. Winston F, Dollard C, Ricupero-Hovasse SL: **Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C.** *Yeast* 1995, **11:**53-55.

114. Boeke JD, Trueheart J, Natsoulis G, Fink GR: **5-Fluoroorotic acid as a selective agent in yeast molecular genetics.** *Methods Enzymol* 1987, **154:**164-175.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

115. Xia Y, Chu W, Qi Q, Xun L: **New insights into the QuikChange process guide the use of Phusion DNA polymerase for site-directed mutagenesis.** *Nucleic Acids Res* 2015, **43:**e12.

116. Schmitt ME, Brown TA, Trumpower BL: **A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae.** *Nucleic acids research* 1990, **18:**3091-3092.

117. Goldman SR, Sharp JS, Vvedenskaya IO, Livny J, Dove SL, Nickels BE: **NanoRNAs prime transcription initiation in vivo.** *Mol Cell* 2011, **42:**817-825.

118. Vvedenskaya IO, Goldman SR, Nickels BE: **Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends.** *Methods Mol Biol* 2015, **1276:**211-228.

119. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25:**1105-1111.

120. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.

121. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.

122. Saldanha AJ: **Java Treeview--extensible visualization of microarray data.** *Bioinformatics* 2004, **20:**3246-3248.

123. de Hoon MJ, Imoto S, Nolan J, Miyano S: **Open source clustering software.** *Bioinformatics* 2004, **20:**1453-1454.

124. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T: **deepTools: a flexible platform for exploring deep-sequencing data.** *Nucleic Acids Res* 2014, **42:**W187-191.

125. Rhee HS, Pugh BF: **Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution.** *Cell* 2011, **147:**1408-1419.

126. van Bakel H, Tsui K, Gebbia M, Mnaimneh S, Hughes TR, Nislow C: **A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription.** *PLoS Genet* 2013, **9:**e1003479.

127. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-359.

128. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.

129. Abante J: **NucSeq v1.0.** v1.0 edition: Zenodo; 2016.

FIGURE LEGENDS

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. Genome-wide analysis of TSS selection in *S. cerevisiae*. A. Overview of method and description of metrics used in analyzing TSS distributions at yeast promoters. **B.** Reproducibility of TSS-seq analysis demonstrated by correlation plots determine RNA 5' ends across all genome positions with ≥3 reads in each library for biological replicates of WT, *rpb1* E1103G, and *rpb1* H1085Y libraries. **C.** Heat map illustrating hierarchical clustering of Pearson correlation coefficients between aggregate (combined biological replicate) libraries for all strains. Clustering distinguishes known reduced function *rpb1* alleles ("slow" or LOF) and increased activity *rpb1* alleles ("fast" or GOF). **D.** TSSs generally map downstream of core promoters predicted by Rhee and Pugh from GTF ChiP-exo data. **E.** Basic metrics of TSS distribution changes distinguish classes of TSS-usage affecting alleles. Determination of change in median TSS position (upstream shift in median position is negative, downstream shift in median position is positive, see Methods) or change in width of TSS distribution (see A, "Δ TSS spread") are sufficient to differentiate two main classes of *rpb1* mutants and separate them from GTF mutants. Heat maps show individual yeast promoter regions on the y-axis and the measured TSS shift or ΔTSS Spread from TSS-seq data for TSS-usage affecting mutants on the x-axis hierarchically clustered in both dimensions.

Figure 2. Pol II and GTF mutants confer polar shifts in TSS-usage across all promoter classes in *S. cerevisiae*. A. Heat maps show TSS distribution changes in a fast (E1103G) or a slow (H1085Y) Pol II mutant relative to WT. 401 nt promoter windows were anchored on measured median TSS position in our WT strain and TSS distributions in WT or mutant strains were normalized to 100%. Differences in distribution between WT and mutant TSS usage were determined by subtracting the normalized WT distribution from normalized mutant distributions. Promoters are separated into those classified as Taf1-enriched, Taf1-depleted, or neither and rank-ordered on the y-axis based on total reads in WT (from high to low). Gain in relative mutant TSS usage is positive while loss in relative mutant usage is negative. **B.** Significant polar shifts in TSS usage are apparent for examined *rpb1* mutants (except *rpb1* F1084I) across promoter classes. All box plots are Tukey plots unless otherwise noted (see Methods). Promoters examined are n=3494 (>200 reads total expression in WT). **C.** Significant polar shifts in TSS usage are apparent for examined GTF mutants and an *rpb1 tfg2* double mutant shows exacerbated TSS shifts relative to the single mutants (compare C to B). Promoters examined are as in (B). **D.** Average TSS shifts in Pol II *rpb1* mutants correlate with their measured in vitro elongation rates. Error bars on TSS shifts and elongation rates are bounds of the 95% confidence intervals of the means. Elongation rates are from [73, 75]. Mutants slower than WT in vitro exhibit downstream shifts in TSS distributions while mutants faster than WT in vitro exhibit upstream shifts in TSS distributions correlating with the strengths of their in vitro elongation rate defects and their in vivo growth rate defects. Promoters examined are as in (B,C).

Figure 3. TSS motif usage and alterations in TSS-usage affecting mutants. A. Preferred Y-1R+1 motif usage observed in our data as expected. *S. cerevisiae* selective enrichment of A at -8 is apparent at the most highly used starts in promoters with higher expression (compare primary/top (1°) TSSs with secondary (2°) or tertiary TSSs from promoters within the top decile of expression). Promoters exhibiting very narrow TSS spreads (focused) show additional minor enrichments for bases near the TSS. **B.** Overall TSS motif usage in WT and TSS-usage affecting mutants. Motifs were separated by -8 -1 +1 identities (64 motifs) as the vast majority of TSS reads derive from $N_{-8}Y_{-1}R_{+1}$ sequences. (Top) Percent motif usage determined for individual strains and displayed in heat map hierarchically clustered on y-axis to group strains with similar motif usage distribution. (Bottom) Difference heat map illustrating relative changes in $N_{-8}Y_{-1}R_{+1}$ motif usage in heat map hierarchically clustered on y-axis to group strains with similar motif usage difference distribution. **C.** Alteration in motif usage and apparent changes to reliance on an A_{-8} could arise from a number of possibilities. Alterations in TSS efficiencies in

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mutants could result in upstream or downstream shifts in TSS distribution if mutants have decreased or increased reliance, respectively, on a particular motif. Conversely, alteration in initiation efficiency in general (increase or decrease) could alter TSS motif usage if TSS motifs are unevenly distributed across yeast promoters (example distribution for hypothetical motif *N*). **D.** TSS motifs are unevenly distributed across yeast promoters and differentially enriched correlating with steady state promoter expression levels. (Top) the apparent highest used $A_{-8}Y_{-1}R_{+1}$ motif ($A_{-8}C_{-1}A_{+1}$) and (bottom) the less preferred $T_{-8}T_{-1}A_{+1}$ motif are compared for Taf1-enriched or Taf1-depleted promoters.

Figure 4. TSS-usage mutants alter TSS usage efficiencies across TSS motifs consistent with promoter scanning initiation at all promoters. A. Median usage (left) or "efficiency" (right) for $A_{-8}C_{-1}A_{+1}$ sites across promoters in WT, *rpb1* E1103G, or *rpb1* H1085Y strains for Taf1-enriched (left pair of graphs) or Taf1-depleted promoters (right pair of graphs). Usage is defined as median percent reads found at any $A_{-8}C_{-1}A_{+1}$ sites by promoter position relative to a baseline position (the median TSS in WT). Efficiency is calculated from a model that assumes promoter scanning from upstream to downstream positions and is defined as number of TSS-seq reads mapping to a genome position divided by the sum of those reads and any downstream reads within a defined promoter window. *rpb1* E1103G and *rpb1* H1085Y both shift usage but alter efficiency differently. **B.** Altered usage across TSS motifs in TSS-usage affecting mutants. Heat maps show difference in aggregate usage normalized to promoter number for different $N_{-8}Y_{-1}R_{+1}$ TSS motifs. Strains are ordered on the x-axis from left-to-right from strongest downstream shifter to strongest upstream shifter. Promoter positions from -100 (upstream) to +100 (downstream) flanking the median TSS position in WT Regardless or promoter class, TSS-usage affecting mutants cause polar effects on distribution of TSS usage when examining motifs separately. **C.** Motif efficiency calculated as in (A) for a subset of $N_{-8}Y_{-1}R_{+1}$ TSS motifs for all mutants. Heat maps are ordered as in (B). Downstream shifting mutants in (B) generally reduce TSS usage efficiencies across promoter positions. Upstream shifting mutants in (B) generally shift TSS efficiencies upstream.

Figure 5. Attributes of core promoter classes and PIC positioning in TSS-usage affecting mutants. A. Enrichment by expression decile in WT of putative core promoter elements in Taf1-enriched and Taf1-depleted promoters. TATA consensus (TATAWAWR, W=A/T, R=A/G) is enriched in Taf1-depleted promoters while the GA-rich element (GAAAAA) is enriched in Taf1-enriched promoters. Yeast promoters are relatively AT-rich so there is a high probability of "TATA-like" elements differing from the TATA consensus by two mismatches. **B.** Tested GAE or TATA-like elements do not greatly contribute to expression from promoters where tested. Expression by Northern blotting for promoters or classes of promoter mutant fused to a reporter gene. Promoter mutants are normalized to the respective WT for each promoter. "Delete" mutants represent deletions of particular element types. "Mutant" elements represent elements where base composition has been altered. **C.** GTF positioning by promoter classes determined by ChiP-exo for Sua7 (TFIIB) or Ssl2 (TFIIH). For each promoter, the median position of ChiP-exo reads on the top (TOP) or bottom (BOT) DNA strand was used to estimate GTF positioning. Left graph shows histogram of estimated GTF positions for Taf1-enriched promoters while right graph shows histogram of estimated GTF positions for Taf1-depleted promoters. **D.** Pol II mutant effects on GTF positioning as detected by ChiP-exo for Sua7 (TFIIB) or Ssl2 (TFIIH). Aggregate ChiP-exo signal for Taf1-enriched or depleted promoters on top (TOP) or bottom (BOT) DNA strands in WT, *rpb1* H1085Y, or *rpb1* E1103G. Curves on graph indicate LOWESS smoothing of aggregate ChiP-exo reads for the top 50% of promoters determine by ChiP-exo reads in WT cells.

Figure 6. Promoter architecture influences sensitivity to TSS-usage affecting mutants. A-C. Distance of TSS to GTF position or core promoter position can correlate with extent of TSS

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

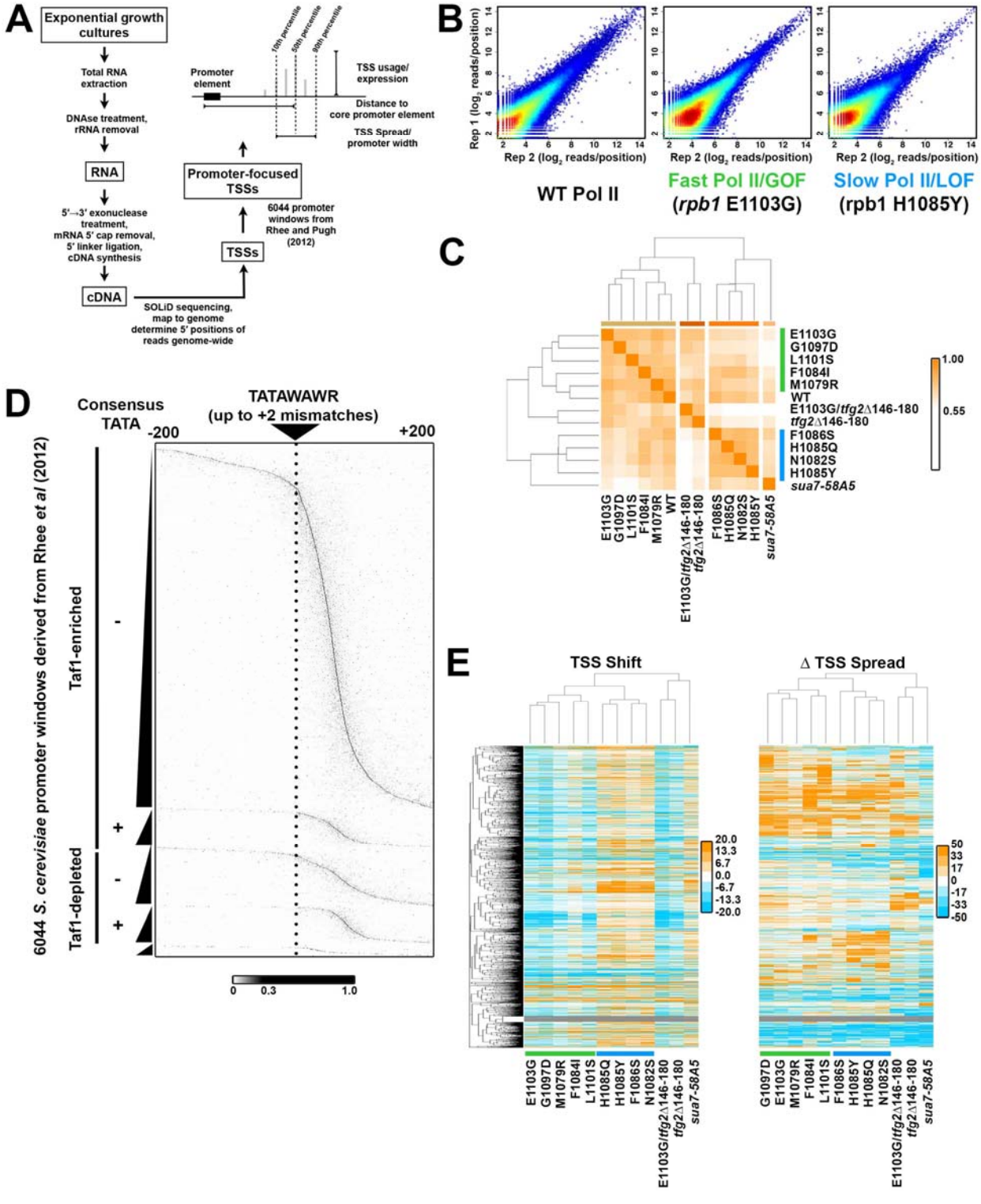
shift in TSS mutants. Dashed lines are linear regression plots for TSS shift vs. GTF to TSS distance. GTF position determined by average of median ChiP-exo signal from top and bottom DNA strands for Sua7 and Ssl for **A.** Taf1-enriched promoters or **B.** Taf1-depleted promoters. **C.** Core promoter-TSS distance for Taf1-depleted TATA-element containing promoters correlates with extent of TSS shifts in TSS-usage affecting mutants. Dashed lines are linear regression plots for TSS shift vs. TATA-element to TSS distance for Taf1-depleted promoters with TATA elements. **D.** Correlation of differential expression (log_2(mutant/WT)) with TSS to PIC distance for Taf1-enriched promoters. **E.** Correlation of differential expression (log_2(mutant/WT)) with TSS to PIC distance for Taf1-depleted promoters.

Figure 7. Relationship of promoter chromatin architecture to PIC position and effects of TSS-usage affecting mutants on nucleosome positioning. A. Nucleosome midpoints as determined by MNase-seq (dashed lines) and GTF ChiP-exo signals for Taf1-enriched promoters (solid line smoothes of scatter plots) are aggregated by promoter quintiles determined by TSS-+1 nucleosome midpoint position. Nucleosome midpoints are from WT strain and the same data are shown as reference for each ChiP-exo plot. First to fifth quintiles are promoters with the closest +1 nucleosome to furthest, respectively. Fifth quintile promoters likely have a weak +1 nucleosome and thus the determined +1 nucleosome is in some cases like the +2. ChiP-exo aggregate data shows correlation with +1 nucleosome-TSS distance. **B.** Nucleosome positioning in WT and H1085Y for Taf1-enriched promoters aligned by +1 nucleosome in WT (left), over genes (-200 to +800 from +1 nucleosome position, right). **C.** Determined +1 nucleosome position for WT and H1085Y Taf1-enriched promoters for individual MNase-seq libraries relative to position determined by averaging the four WT libraries. Box plots are Tukey plots (see Methods). **D. and E.** Nucleosome positioning analyses as in B, C for top expression decile Taf1-enriched promoters for WT and H1085Y. **F. and G.** Nucleosome positioning analyses as in B, C for bottom expression decile Taf1-enriched promoters for WT and *rpb1* H1085Y. **H. and I.** Nucleosome positioning analyses as in B, C for Taf1-enriched promoters for *rpb1* E1103G. WT data from B, C shown as reference.

Figure 8. TSS distribution characteristics for select eukaryotes. A. TSS spread (distance defining positions of 10%-90% percentile of the TSS distribution) plotted for Taf1-enriched (left) or depleted (right) promoters separated by expression decile for WT yeast. Spread determined here for 401 nt promoter windows. **B.** Distribution of promoter “widths” (TSS spread) for TSS-seq for 101 nt promoter windows across a number of eukaryotic TSS-seq or related methodologies, including *S. cerevisiae* (WT data from this work), *S. pombe* (Li *et al*), *D. melanogaster* (Hoskins *et al* CAGE or Kwak *et al* PRO-cap), *Danio rerio* Zebrafish Maternal promoters (512 stage) or Zygotic (prim20) (Haberle *et al*), or Human PRO-cap data (Core *et al*). **C.** Number of available YR dinucleotides within spread regions for promoters in (B) with the following alterations. *S. pombe* promoter class limited to “notSP” (not single TSS promoters) as defined by Li et al. M1 and M2 classes for human data as defined by Core et al are separated. Hoskins et al CAGE data were not analyzed due to CAGE artifact potential for adding an extra untemplated C during reverse transcription of RNA 5’ ends. **D.** Number of $Y_{-1}R_{+1}$ dinucleotides used at ≥2% of total reads for promoter region vs. spread width for promoters in (C). Fraction of non-CpG $Y_{-1}R_{+1}$ dinucleotides used at ≥2% of total reads for a promoter region relative to available non-CpG $Y_{-1}R_{+1}$ dinucleotides vs spread width for promoters in (C).

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1427 **FIGURES AND FIGURE LEGENDS**



1428

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. Genome-wide analysis of TSS selection in *S. cerevisiae*. A. Overview of method and description of metrics used in analyzing TSS distributions at yeast promoters. **B.** Reproducibility of TSS-seq analysis demonstrated by correlation plots determine RNA 5' ends across all genome positions with ≥3 reads in each library for biological replicates of WT, *rpb1* E1103G, and *rpb1* H1085Y libraries. **C.** Heat map illustrating hierarchical clustering of Pearson correlation coefficients between aggregate (combined biological replicate) libraries for all strains. Clustering distinguishes known reduced function *rpb1* alleles ("slow" or LOF) and increased activity *rpb1* alleles ("fast" or GOF). **D.** TSSs generally map downstream of core promoters predicted by Rhee and Pugh from GTF ChiP-exo data. **E.** Basic metrics of TSS distribution changes distinguish classes of TSS-usage affecting alleles. Determination of change in median TSS position (upstream shift in median position is negative, downstream shift in median position is positive, see Methods) or change in width of TSS distribution (see A, "Δ TSS spread") are sufficient to differentiate two main classes of *rpb1* mutants and separate them from GTF mutants. Heat maps show individual yeast promoter regions on the y-axis and the measured TSS shift or ΔTSS Spread from TSS-seq data for TSS-usage affecting mutants on the x-axis hierarchically clustered in both dimensions.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

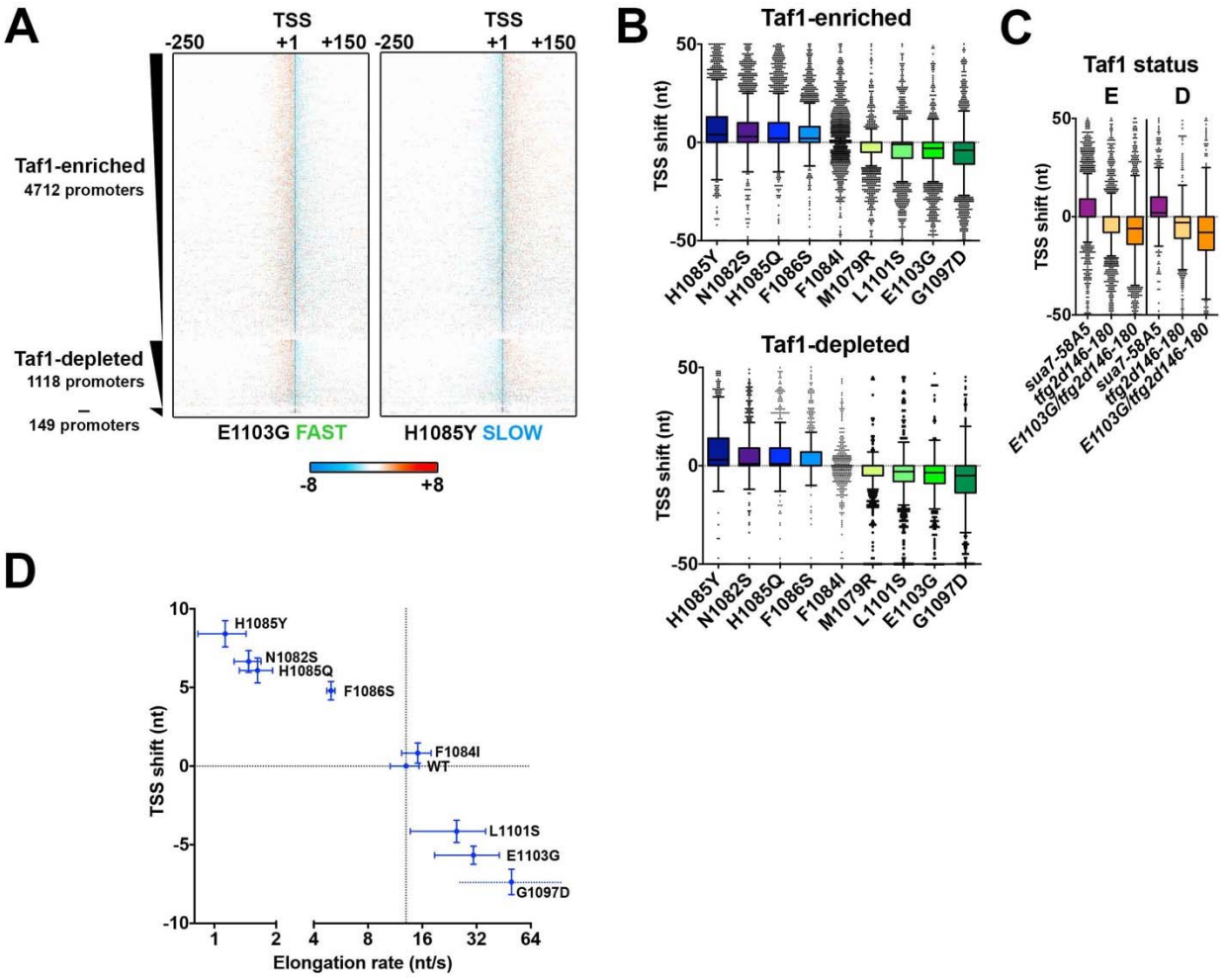


Figure 2. Pol II and GTF mutants confer polar shifts in TSS-usage across all promoter classes in *S. cerevisiae*. A. Heat maps show TSS distribution changes in a fast (E1103G) or a slow (H1085Y) Pol II mutant relative to WT. 401 nt promoter windows were anchored on measured median TSS position in our WT strain and TSS distributions in WT or mutant strains were normalized to 100%. Differences in distribution between WT and mutant TSS usage were determined by subtracting the normalized WT distribution from normalized mutant distributions. Promoters are separated into those classified as Taf1-enriched, Taf1-depleted, or neither and rank-ordered on the y-axis based on total reads in WT (from high to low). Gain in relative mutant TSS usage is positive while loss in relative mutant usage is negative. **B.** Significant polar shifts in TSS usage are apparent for examined *rpb1* mutants (except *rpb1* F1084I) across promoter classes. All box plots are Tukey plots unless otherwise noted (see Methods). Promoters examined are n=3494 (>200 reads total expression in WT). **C.** Significant polar shifts in TSS usage are apparent for examined GTF mutants and an *rpb1 tfg2* double mutant shows exacerbated TSS shifts relative to the single mutants (compare C to B). Promoters examined are as in (B). **D.** Average TSS shifts in Pol II *rpb1* mutants correlate with their measured in vitro elongation rates. Error bars on TSS shifts and elongation rates are bounds of the 95% confidence intervals of the means. Elongation rates are from (**XXX**). Mutants slower than WT in vitro exhibit downstream shifts in TSS distributions while mutants faster than WT in vitro exhibit upstream shifts in TSS distributions correlating with the strengths of their in vitro elongation rate defects and their in vivo growth rate defects. Promoters examined are as in (B,C).

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

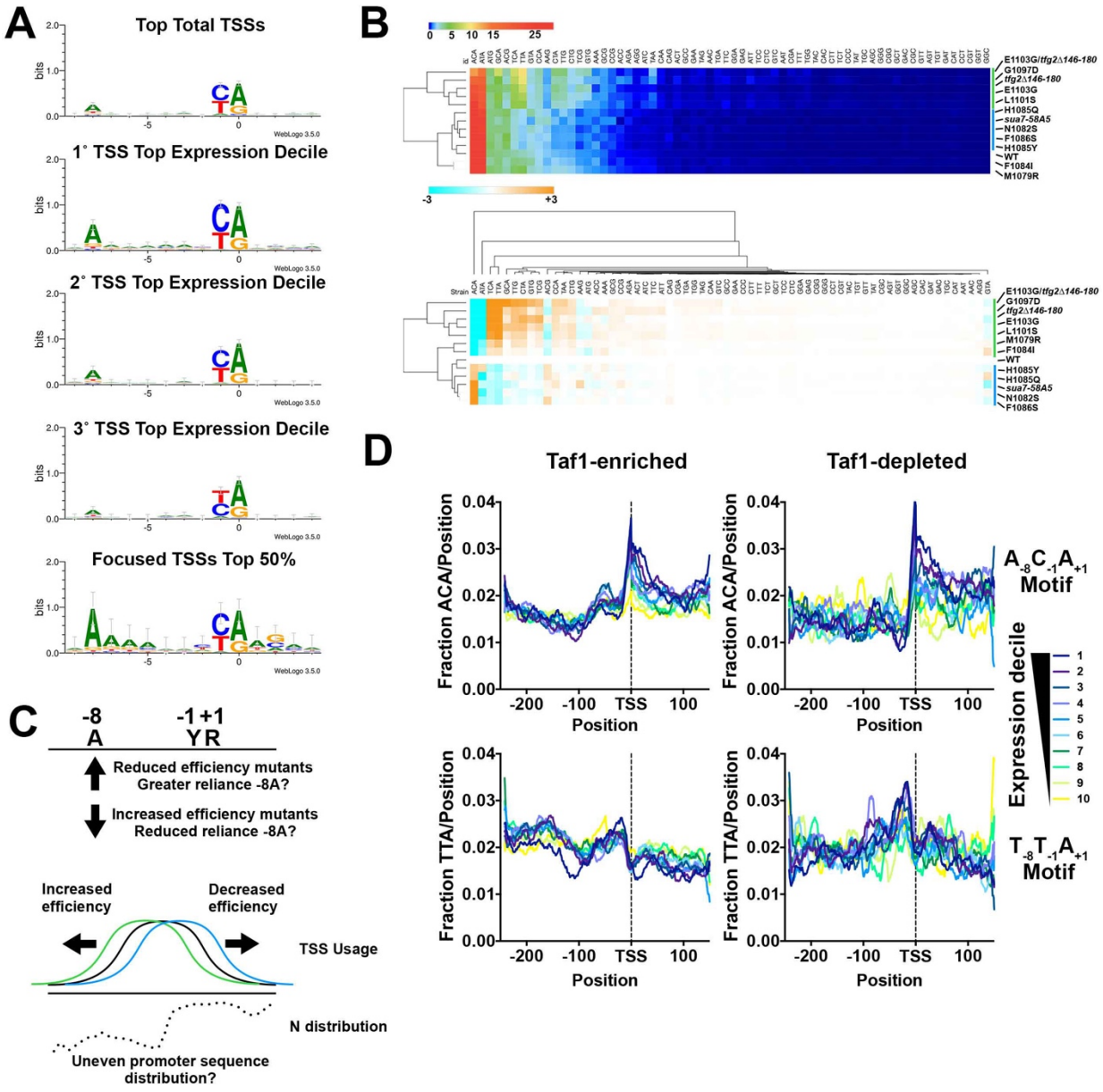


Figure 3. TSS motif usage and alterations in TSS-usage affecting mutants. A. Preferred Y-1R+1 motif usage observed in our data as expected. *S. cerevisiae* selective enrichment of A at -8 is apparent at the most highly used starts in promoters with higher expression (compare primary/top (1˚) TSSs with secondary (2˚) or tertiary TSSs from promoters within the top decile of expression). Promoters exhibiting very narrow TSS spreads (focused) show additional minor enrichments for bases near the TSS. **B.** Overall TSS motif usage in WT and TSS-usage affecting mutants. Motifs were separated by -8 -1 +1 identities (64 motifs) as the vast majority of TSS reads derive from $N_{-8}Y_{-1}R_{+1}$ sequences. (Top) Percent motif usage determined for individual strains and displayed in heat map hierarchically clustered on y-axis to group strains with similar motif usage distribution. (Bottom) Difference heat map illustrating relative changes in $N_{-8}Y_{-1}R_{+1}$ motif usage in heat map hierarchically clustered on y-axis to group strains with similar motif usage difference distribution. **C.** Alteration in motif usage and apparent changes to reliance on an A_{-8} could arise from a number of possibilities. Alterations in TSS efficiencies in mutants could result in upstream or downstream shifts in TSS distribution if mutants have

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

decreased or increased reliance, respectively, on a particular motif. Conversely, alteration in initiation efficiency in general (increase or decrease) could alter TSS motif usage if TSS motifs are unevenly distributed across yeast promoters (example distribution for hypothetical motif *N*). **D.** TSS motifs are unevenly distributed across yeast promoters and differentially enriched correlating with steady state promoter expression levels. (Top) the apparent highest used $A_{-8}Y_{-1}R_{+1}$ motif ($A_{-8}C_{-1}A_{+1}$) and (bottom) the less preferred $T_{-8}T_{-1}A_{+1}$ motif are compared for Taf1-enriched or Taf1-depleted promoters.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

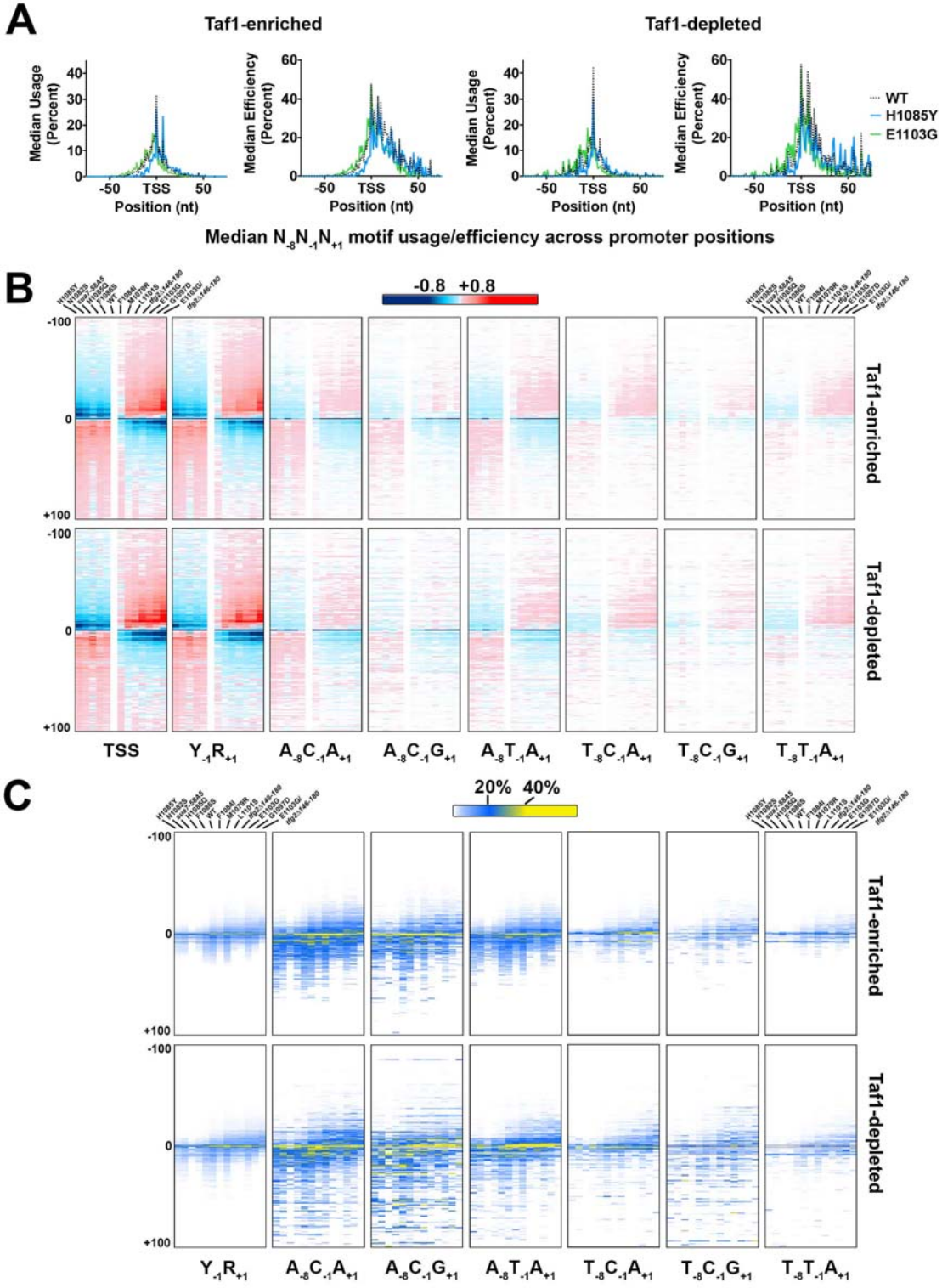


Figure 4. TSS-usage mutants alter TSS usage efficiencies across TSS motifs consistent with promoter scanning initiation at all promoters. A. Median usage (left) or "efficiency"

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(right) for $A_{-8}C_{-1}A_{+1}$ sites across promoters in WT, *rpb1* E1103G, or *rpb1* H1085Y strains for Taf1-enriched (left pair of graphs) or Taf1-depleted promoters (right pair of graphs). Usage is defined as median percent reads found at any $A_{-8}C_{-1}A_{+1}$ sites by promoter position relative to a baseline position (the median TSS in WT). Efficiency is calculated from a model that assumes promoter scanning from upstream to downstream positions and is defined as number of TSS-seq reads mapping to a genome position divided by the sum of those reads and any downstream reads within a defined promoter window. *rpb1* E1103G and *rpb1* H1085Y both shift usage but alter efficiency differently. **B.** Altered usage across TSS motifs in TSS-usage affecting mutants. Heat maps show difference in aggregate usage normalized to promoter number for different $N_{-8}Y_{-1}R_{+1}$ TSS motifs. Strains are ordered on the x-axis from left-to-right from strongest downstream shifter to strongest upstream shifter. Promoter positions from -100 (upstream) to +100 (downstream) flanking the median TSS position in WT regardless or promoter class, TSS-usage affecting mutants cause polar effects on distribution of TSS usage when examining motifs separately. **C.** Motif efficiency calculated as in (A) for a subset of $N_{-8}Y_{-1}R_{+1}$ TSS motifs for all mutants. Heat maps are ordered as in (B). Downstream shifting mutants in (B) generally reduce TSS usage efficiencies across promoter positions. Upstream shifting mutants in (B) generally shift TSS efficiencies upstream.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

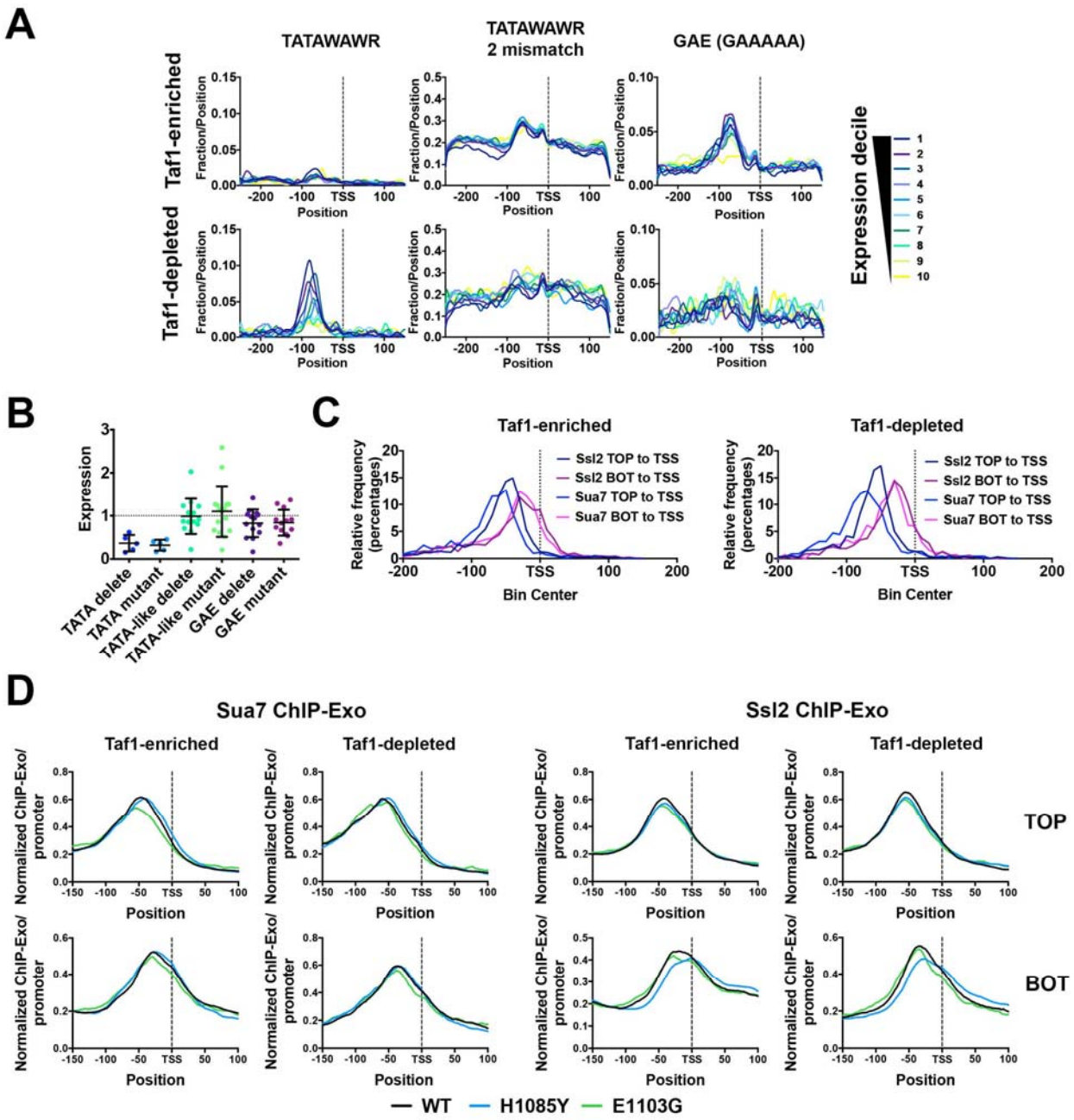


Figure 5. Attributes of core promoter classes and PIC positioning in TSS-usage affecting mutants. A. Enrichment by expression decile in WT of putative core promoter elements in Taf1-enriched and Taf1-depleted promoters. TATA consensus (TATAWAWR, W=A/T, R=A/G) is enriched in Taf1-depleted promoters while the GA-rich element (GAAAAA) is enriched in Taf1-enriched promoters. Yeast promoters are relatively AT-rich so there is a high probability of "TATA-like" elements differing from the TATA consensus by two mismatches. **B.** Tested GAE or TATA-like elements do not greatly contribute to expression from promoters where tested. Expression by Northern blotting for promoters or classes of promoter mutant fused to a reporter gene. Promoter mutants are normalized to the respective WT for each promoter. "Delete" mutants represent deletions of particular element types. "Mutant" elements represent elements where base composition has been altered. **C.** GTF positioning by promoter classes determined

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by ChiP-exo for Sua7 (TFIIB) or Ssl2 (TFIIH). For each promoter, the median position of ChiP-exo reads on the top (TOP) or bottom (BOT) DNA strand was used to estimate GTF positioning. Left graph shows histogram of estimated GTF positions for Taf1-enriched promoters while right graph shows histogram of estimated GTF positions for Taf1-depleted promoters. **D.** Pol II mutant effects on GTF positioning as detected by ChiP-exo for Sua7 (TFIIB) or Ssl2 (TFIIH). Aggregate ChiP-exo signal for Taf1-enriched or depleted promoters on top (TOP) or bottom (BOT) DNA strands in WT, *rpb1* H1085Y, or *rpb1* E1103G. Curves on graph indicate LOWESS smoothing of aggregate ChiP-exo reads for the top 50% of promoters determine by ChiP-exo reads in WT cells.

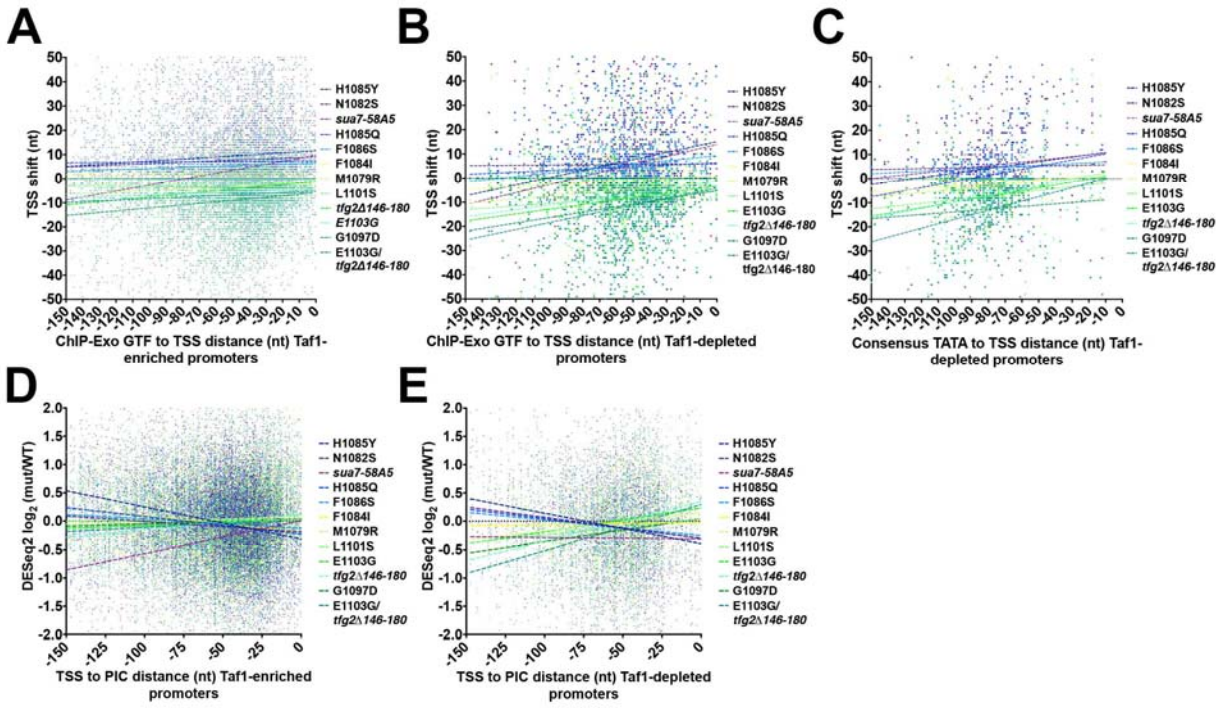
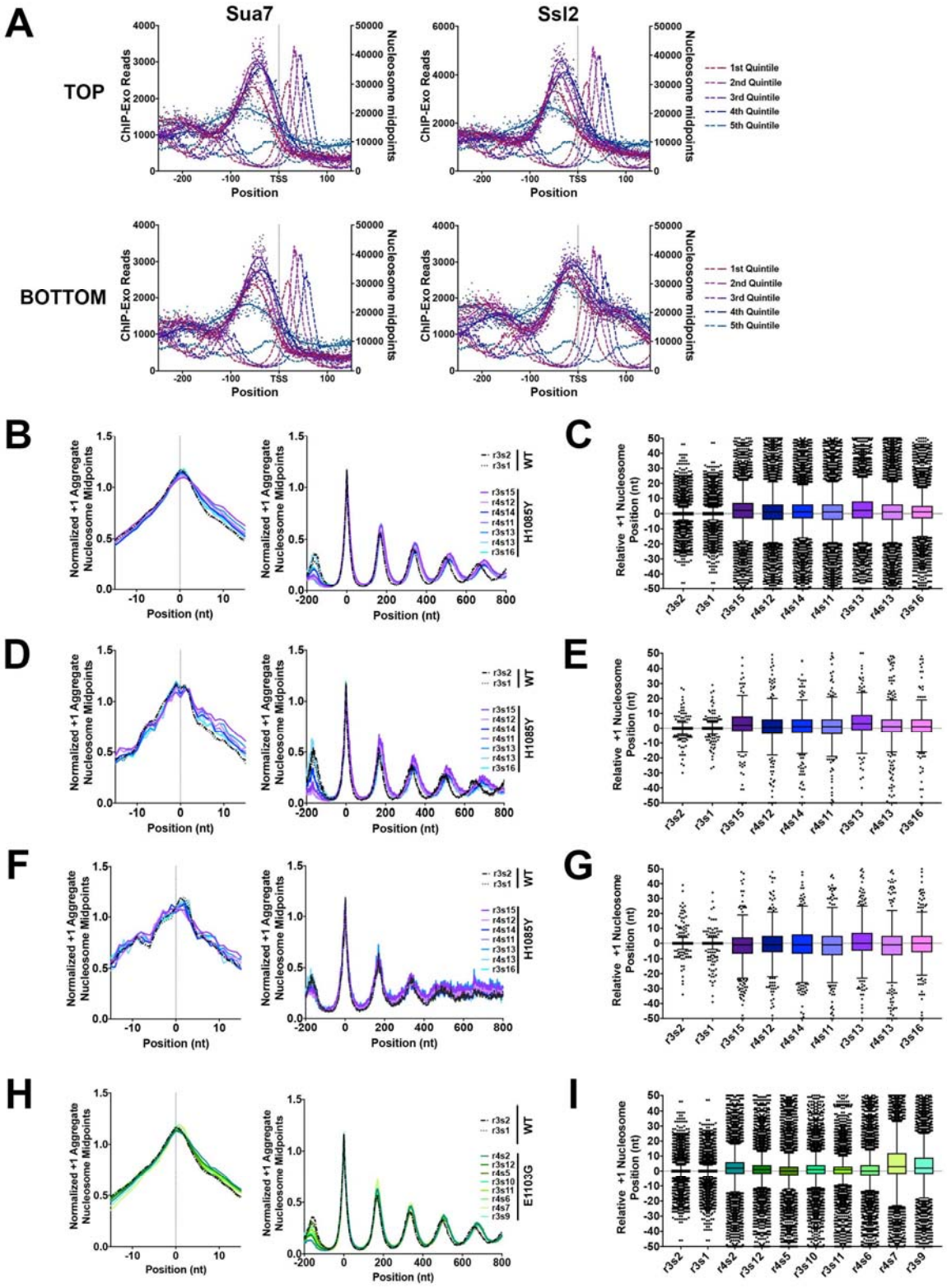


Figure 6. Promoter architecture influences sensitivity to TSS-usage affecting mutants. A-C. Distance of TSS to GTF position or core promoter position can correlate with extent of TSS shift in TSS mutants. Dashed lines are linear regression plots for TSS shift vs. GTF to TSS distance. GTF position determined by average of median ChiP-exo signal from top and bottom DNA strands for Sua7 and Ssl for **A.** Taf1-enriched promoters or **B.** Taf1-depleted promoters. **C.** Core promoter-TSS distance for Taf1-depleted TATA-element containing promoters correlates with extent of TSS shifts in TSS-usage affecting mutants. Dashed lines are linear regression plots for TSS shift vs. TATA-element to TSS distance for Taf1-depleted promoters with TATA elements. **D.** Correlation of differential expression (log_2(mutant/WT)) with TSS to PIC distance for Taf1-enriched promoters. **E.** Correlation of differential expression (log_2(mutant/WT)) with TSS to PIC distance for Taf1-depleted promoters.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1542

1543 **Figure 7. Relationship of promoter chromatin architecture to PIC position and effects of**
1544 **TSS-usage affecting mutants on nucleosome positioning. A.** Nucleosome midpoints as

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determined by MNase-seq (dashed lines) and GTF ChiP-exo signals for Taf1-enriched promoters (solid line smoothes of scatter plots) are aggregated by promoter quintiles determined by TSS-+1 nucleosome midpoint position. Nucleosome midpoints are from WT strain and the same data are shown as reference for each ChiP-exo plot. First to fifth quintiles are promoters with the closest +1 nucleosome to furthest, respectively. Fifth quintile promoters likely have a weak +1 nucleosome and thus the determined +1 nucleosome is in some cases like the +2. ChiP-exo aggregate data shows correlation with +1 nucleosome-TSS distance. **B.** Nucleosome positioning in WT and H1085Y for Taf1-enriched promoters aligned by +1 nucleosome in WT (left), over genes (-200 to +800 from +1 nucleosome position, right). **C.** Determined +1 nucleosome position for WT and H1085Y Taf1-enriched promoters for individual MNase-seq libraries relative to position determined by averaging the four WT libraries. Box plots are Tukey plots (see Methods). **D. and E.** Nucleosome positioning analyses as in B, C for top expression decile Taf1-enriched promoters for WT and H1085Y. **F. and G.** Nucleosome positioning analyses as in B, C for bottom expression decile Taf1-enriched promoters for WT and *rpb1* H1085Y. **H. and I.** Nucleosome positioning analyses as in B, C for Taf1-enriched promoters for *rpb1* E1103G. WT data from B, C shown as reference.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

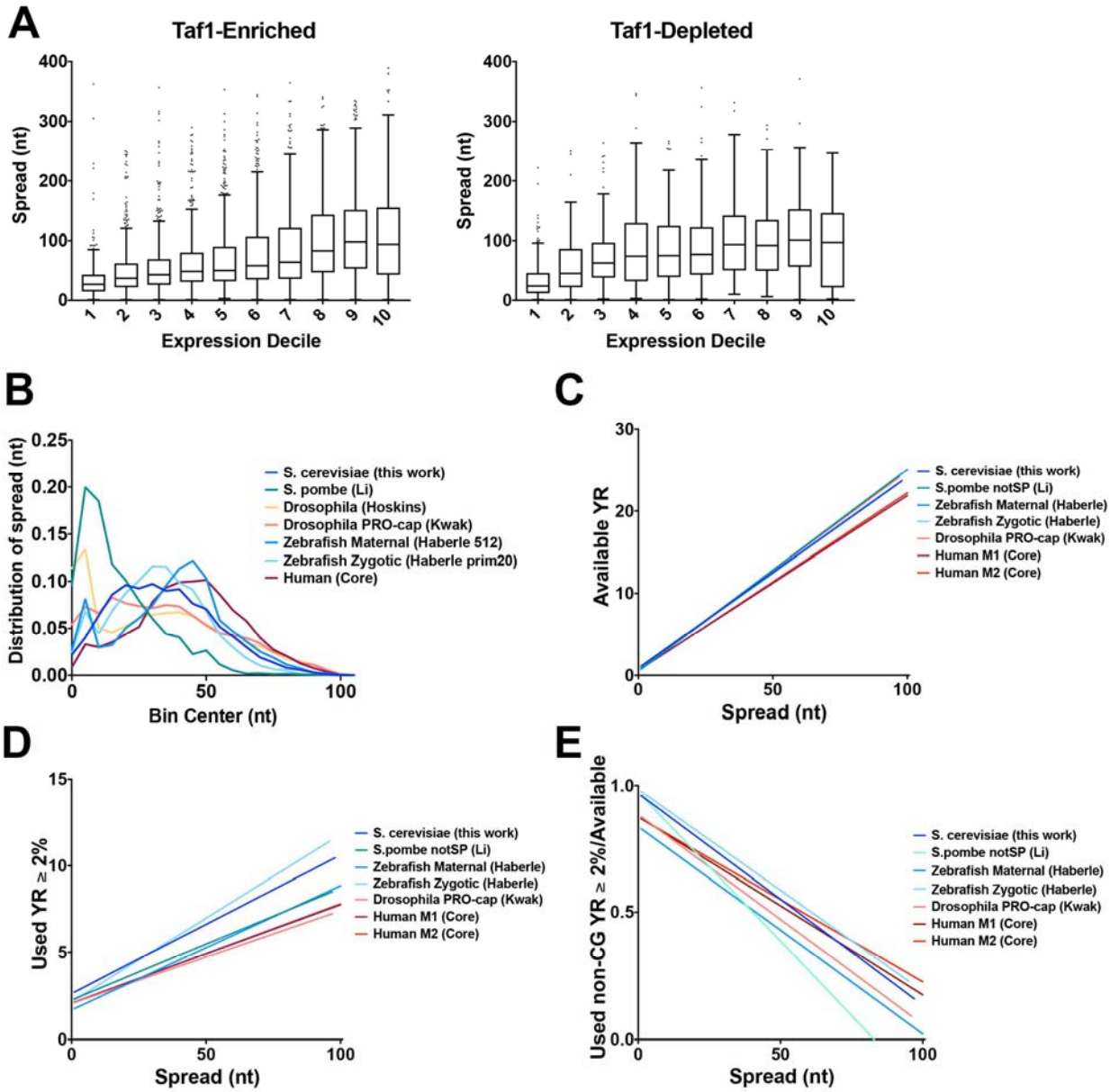
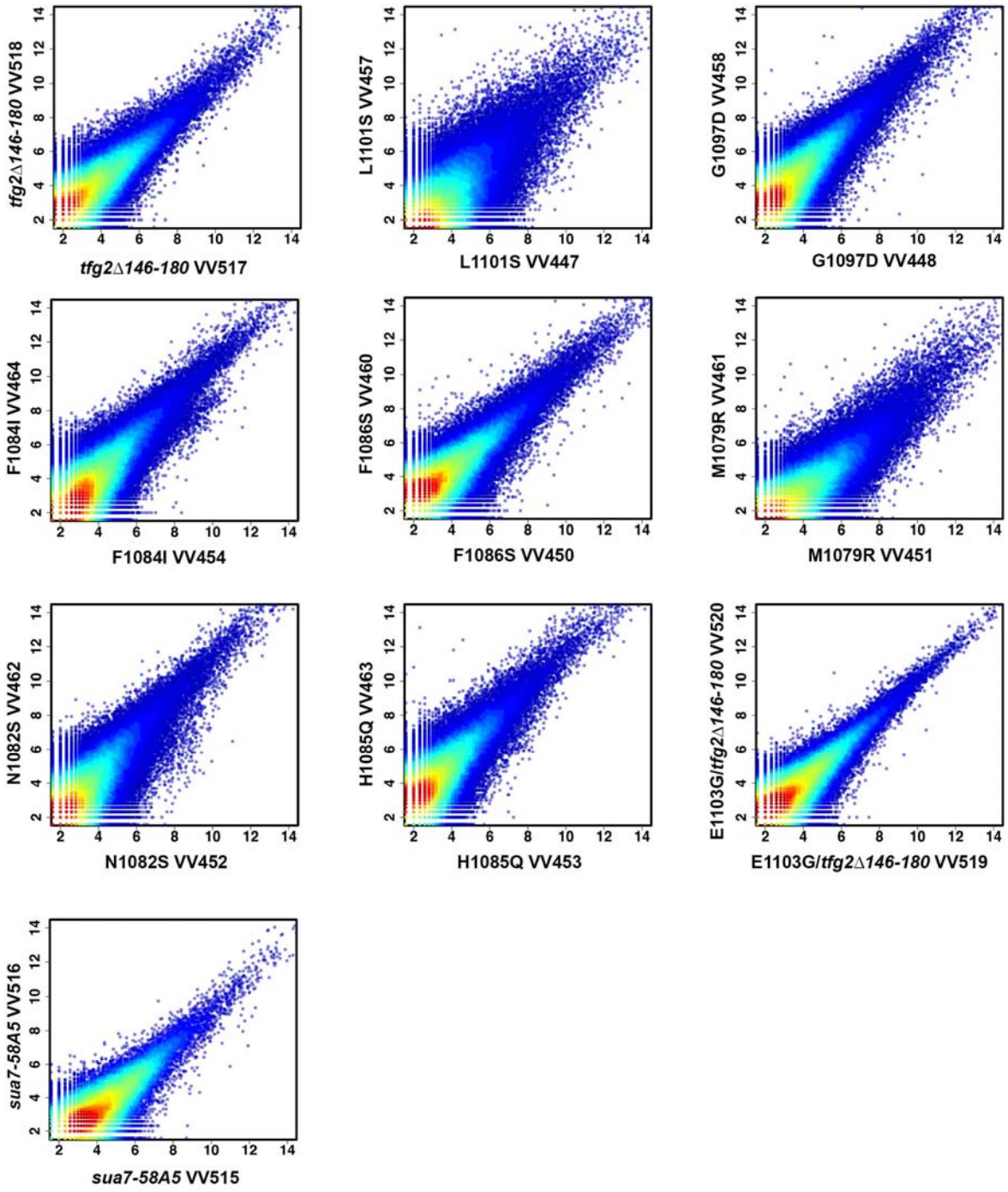


Figure 8. TSS distribution characteristics for select eukaryotes. A. TSS spread (distance defining positions of 10%-90% percentile of the TSS distribution) plotted for Taf1-enriched (left) or depleted (right) promoters separated by expression decile for WT yeast. Spread determined here for 401 nt promoter windows. **B.** Distribution of promoter "widths" (TSS spread) for TSS-seq for 101 nt promoter windows across a number of eukaryotic TSS-seq or related methodologies, including *S. cerevisiae* (WT data from this work), *S. pombe* (Li *et al*), *D. melanogaster* (Hoskins *et al* CAGE or Kwak *et al* PRO-cap), *Danio rerio* Zebrafish Maternal promoters (512 stage) or Zygotic (prim20) (Haberle *et al*), or Human PRO-cap data (Core *et al*). **C.** Number of available YR dinucleotides within spread regions for promoters in (B) with the following alterations. *S. pombe* promoter class limited to "notSP" (not single TSS promoters) as defined by Li et al. M1 and M2 classes for human data as defined by Core et al are separated. Hoskins et al CAGE data were not analyzed due to CAGE artifact potential for adding an extra untemplated C during reverse transcription of RNA 5' ends. **D.** Number of $Y_{-1}R_{+1}$ dinucleotides

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

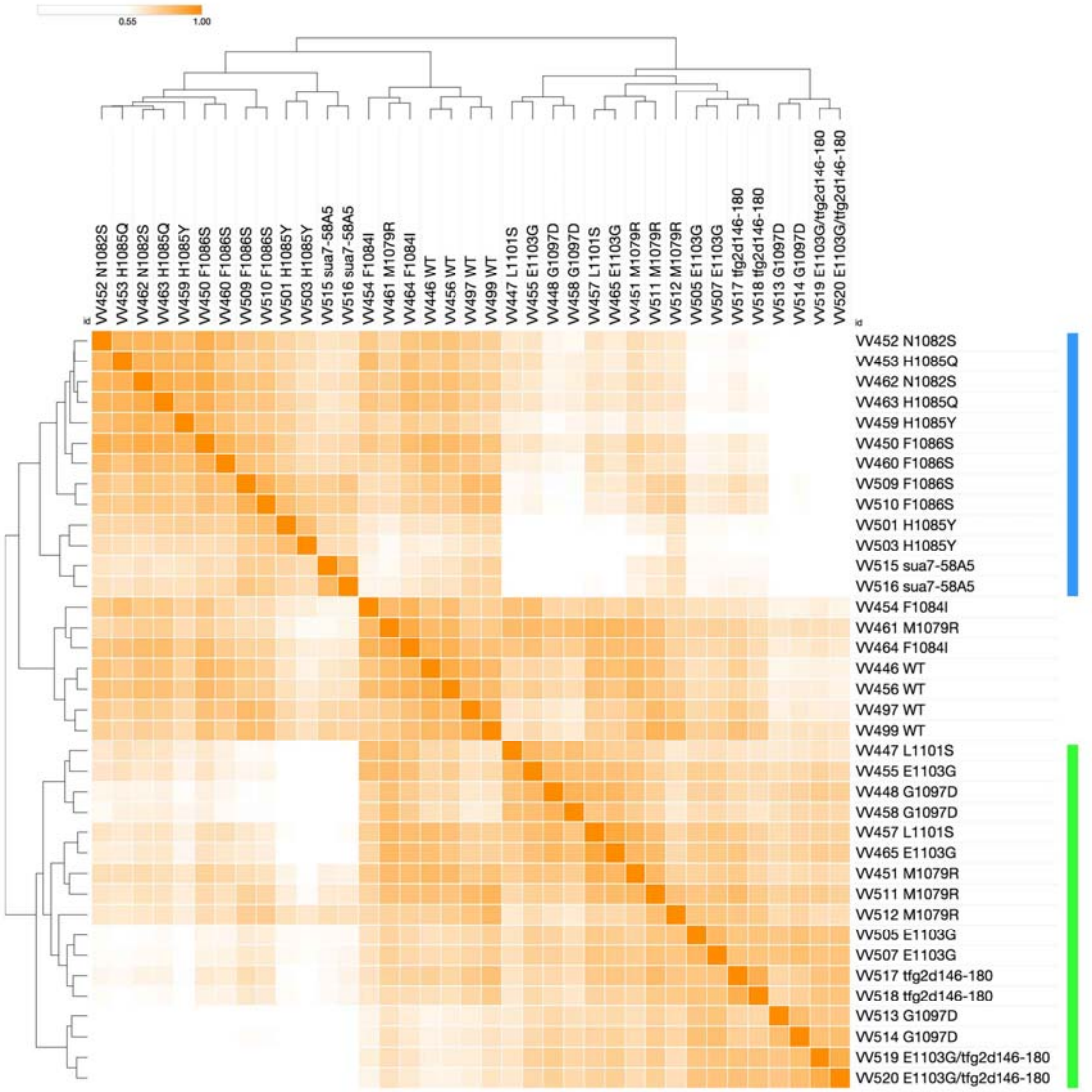
used at ≥2% of total reads for promoter region vs. spread width for promoters in (C). Fraction of non-CpG $Y_{-1}R_{+1}$ dinucleotides used at ≥2% of total reads for a promoter region relative to available non-CpG $Y_{-1}R_{+1}$ dinucleotides vs spread width for promoters in (C).

SUPPLEMENTAL FIGURES AND LEGENDS



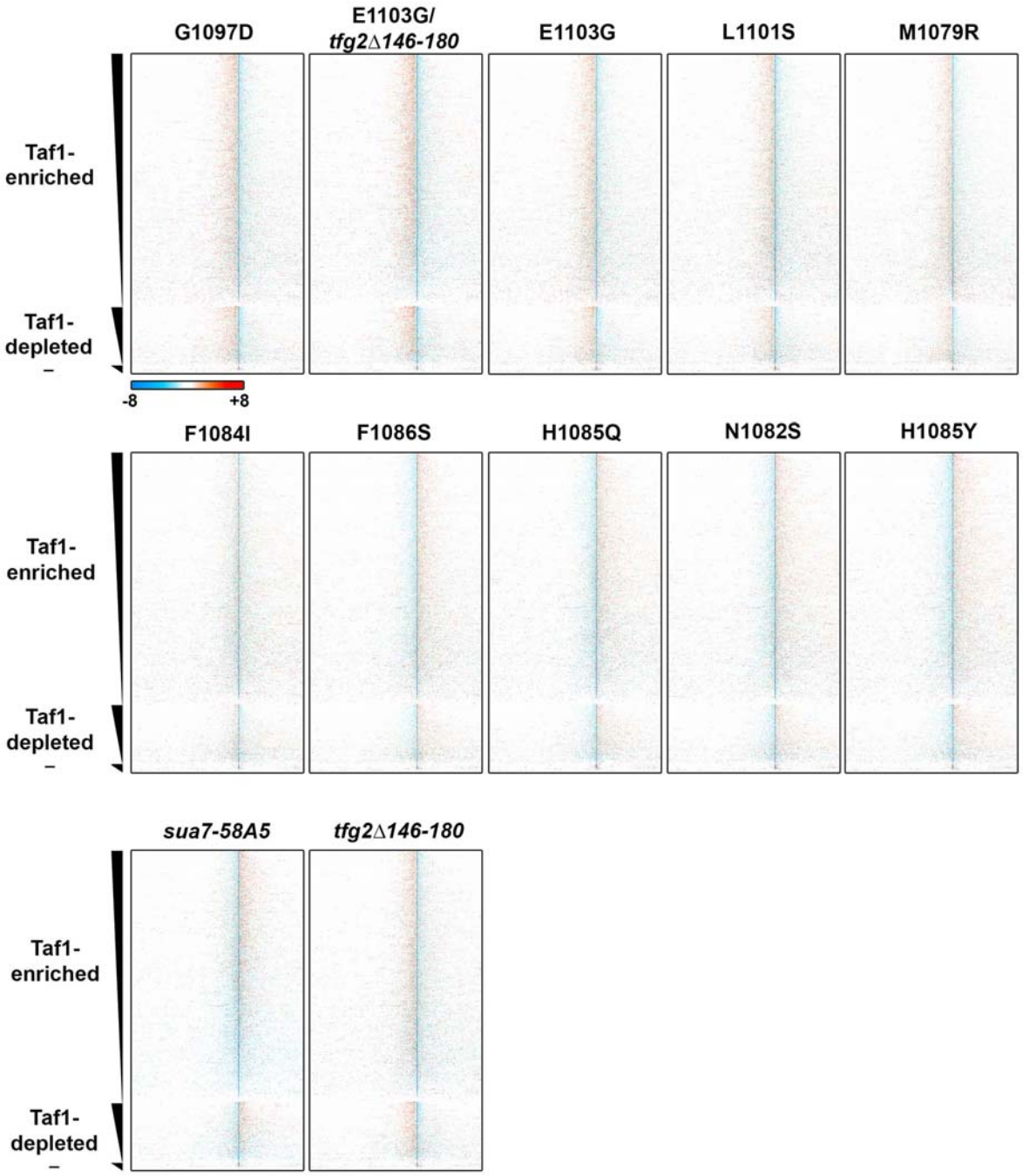
Supplemental Figure 1. Example correlation plots for biological replicate TSS-seq libraries. Plots show all genome positions with ≥ 3 reads in each library.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



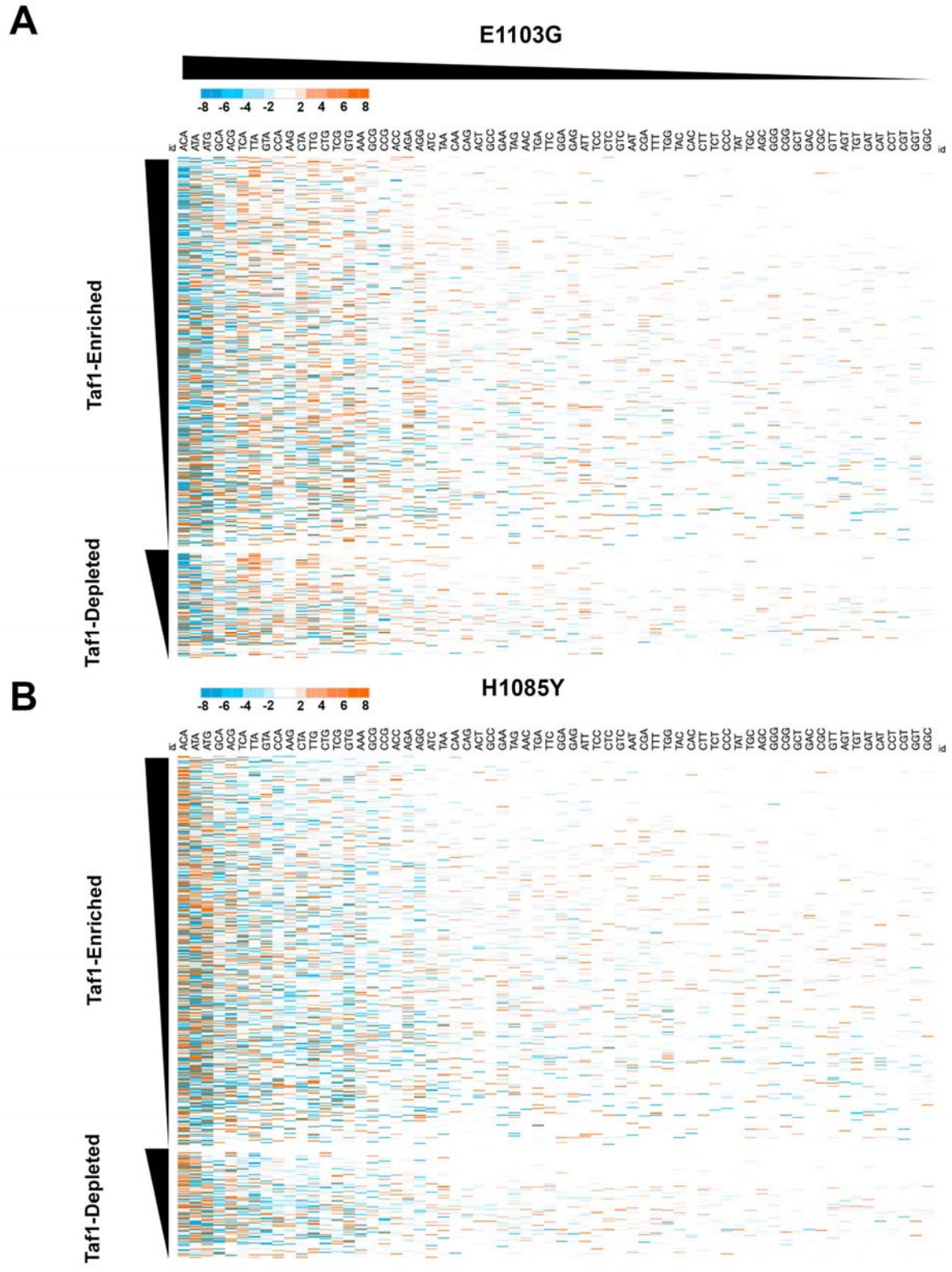
Supplemental Figure 2. Correlation matrix for individual TSS-seq libraries for genome positions within promoter windows. Pearson r correlation coefficients for all TSS-seq library comparisons displayed in a hierarchically clustered heat map. Promoter windows in this analysis were defined by Rhee and Pugh predicted 8-mer TATA or TATA-like core promoter element position +/- 200 nucleotides upstream and downstream. Libraries VV446-465 represent one batch of libraries and VV497-520 represent a separate batch. Clustering distinguishes two major classes of TSS-seq libraries correlating with upstream TSS-shifting and downstream TSS-shifting.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



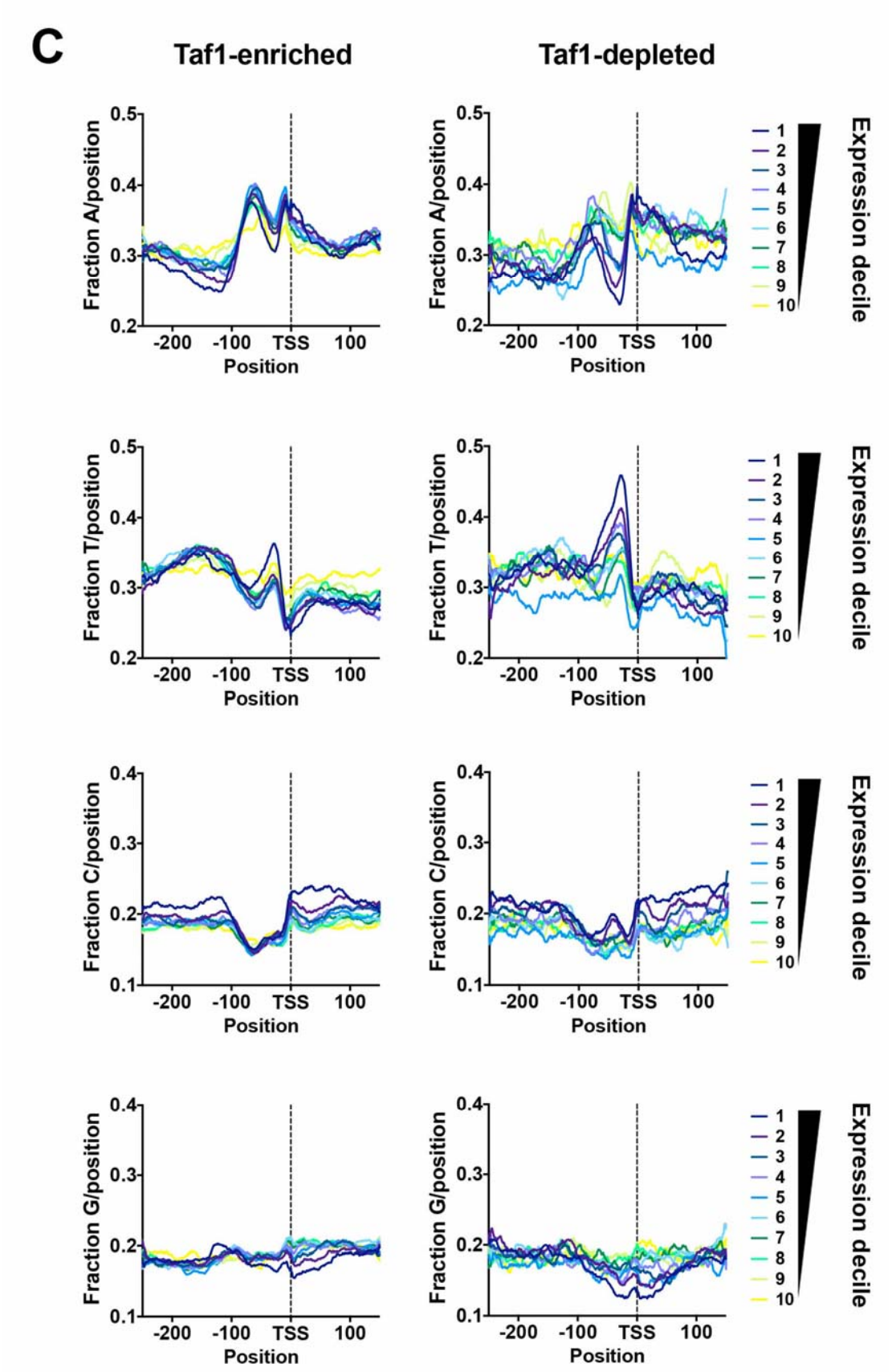
Supplemental Figure 3. Polar effects on TSS distributions observed for majority of TSS-usage affecting mutants genome wide. Heat maps as in Figure 2A. H1085Y and E1103G maps from Figure 2A shown here for comparison with all other heat maps. Maps are arranged from strongly upstream shifting to strongly downstream shifting (top left to middle right). Downstream shifting *sua7-58A5* and upstream shifting *tfg2Δ146-180* GTF mutants are shown in bottom row.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1599

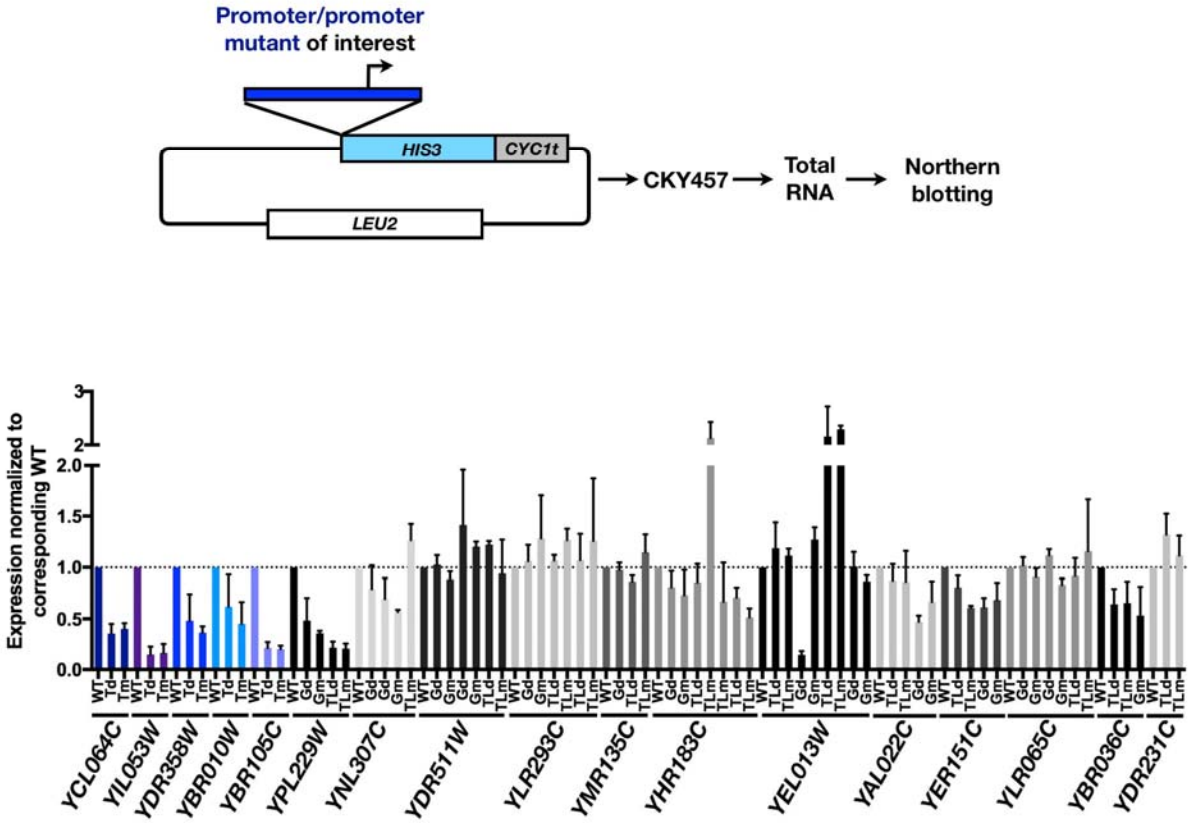
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1600

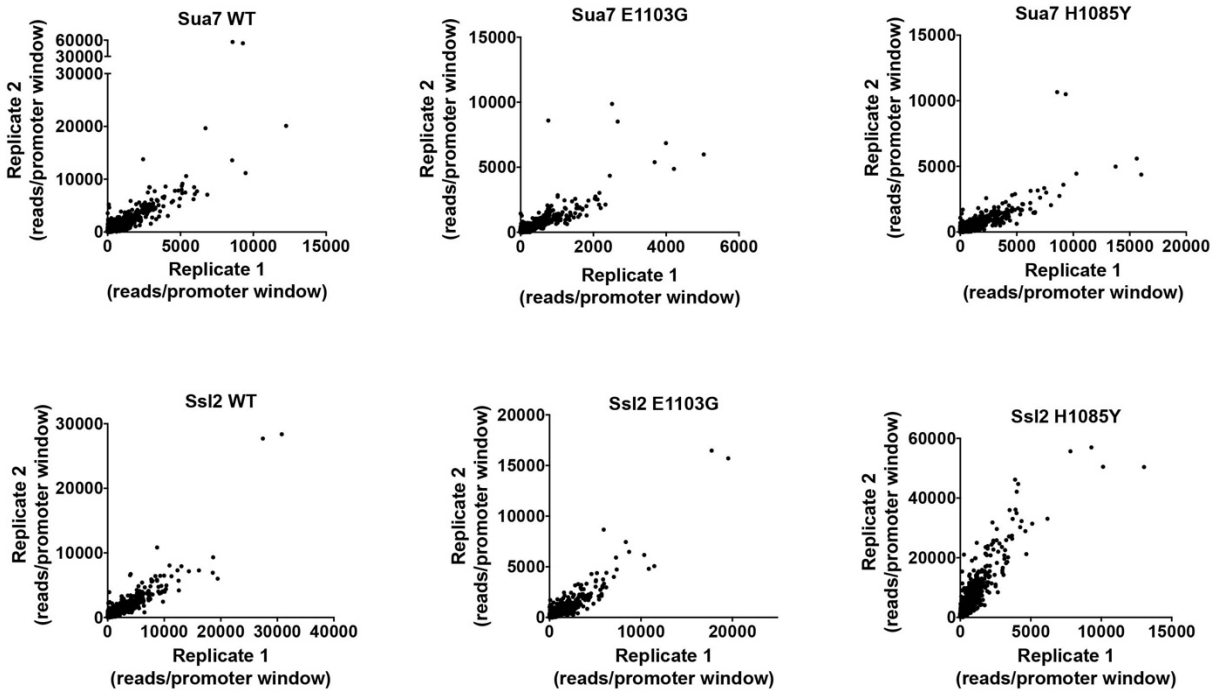
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Figure 4. Effects of *rpb1* H1085Y and *rpb1* E1103G mutants on TSS motif usage for $N_{-8}Y_{-1}R_{+1}$ motifs at the individual promoter level. A and B. Heat maps illustrating differences in percent motif usage for individual promoters (y-axis) for the 64 $N_{-8}Y_{-1}R_{+1}$ motifs (x-axis) in *rpb1* E1103G (**A**) or *rpb1* H1085Y (**B**) are shown. Motifs are rank ordered based on overall usage across genome in WT yeast (high to low from left to right) and promoters are separated into Taf1-enriched and Taf1-depleted classes and rank ordered within class by expression (high to low from top to bottom). **C.** Distribution of bases on the top promoter strand for Taf1-enriched or depleted promoters, separated by expression decile in WT cells.



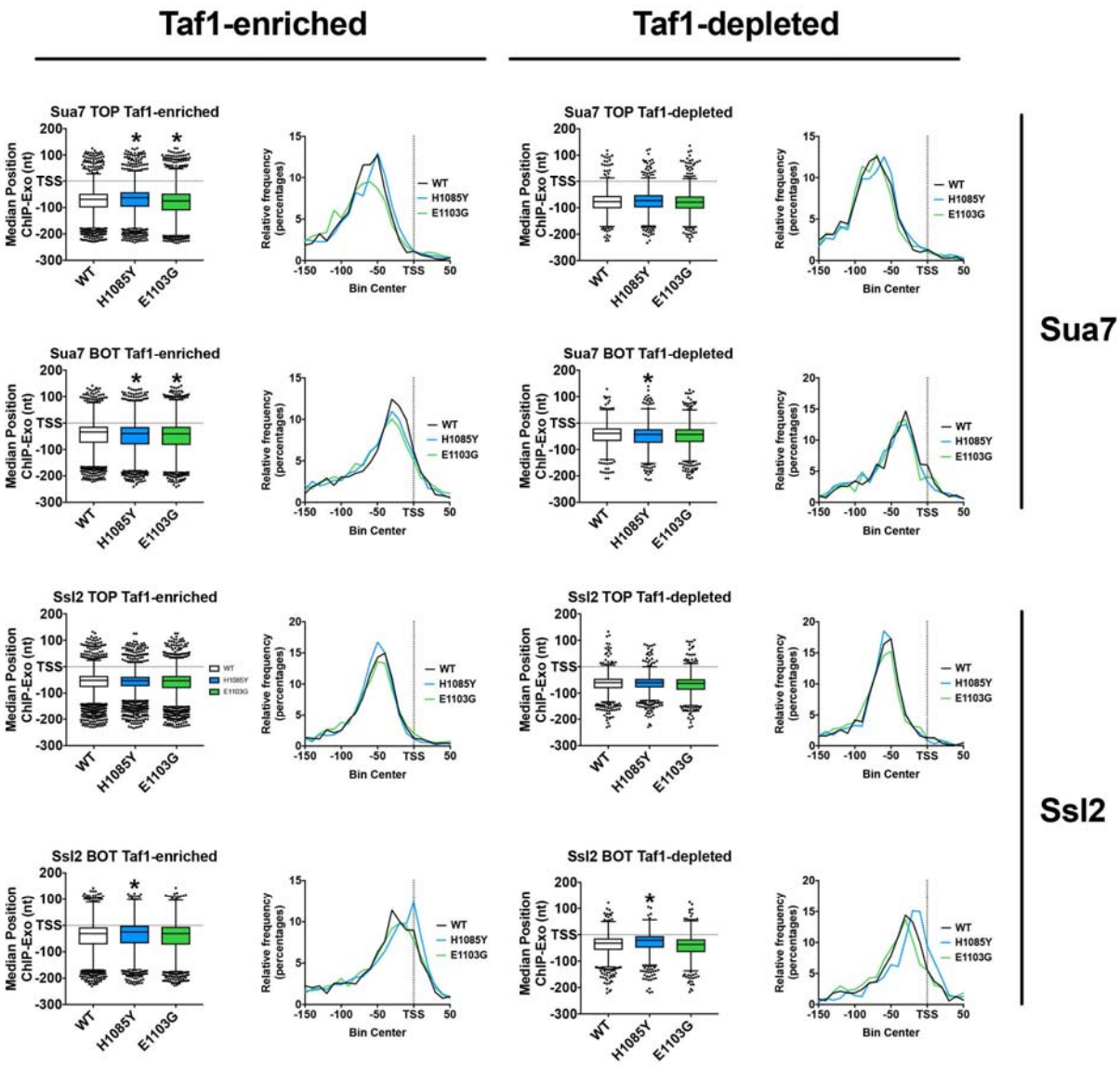
Supplemental Figure 5. Effects on expression level of putative core promoter element mutations. (Top) Schematic of reporter plasmids fusing promoters of interest (up to ATG) to a *HIS3* ORF/*CYC1* terminator reporter. (Bottom) Quantification of Northern blotting for control WT or promoters mutated (Tm) or deleted (Td) for consensus TATA elements (promoters shaded in blue), mutated or deleted for GAE (Gm or Gd, respectively) or mutated or deleted for TATA-like elements identified by Rhee and Pugh or our own analyses (TLm, TLd, respectively). Bars are mean +/- standard deviation of the mean (n=≥3).

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



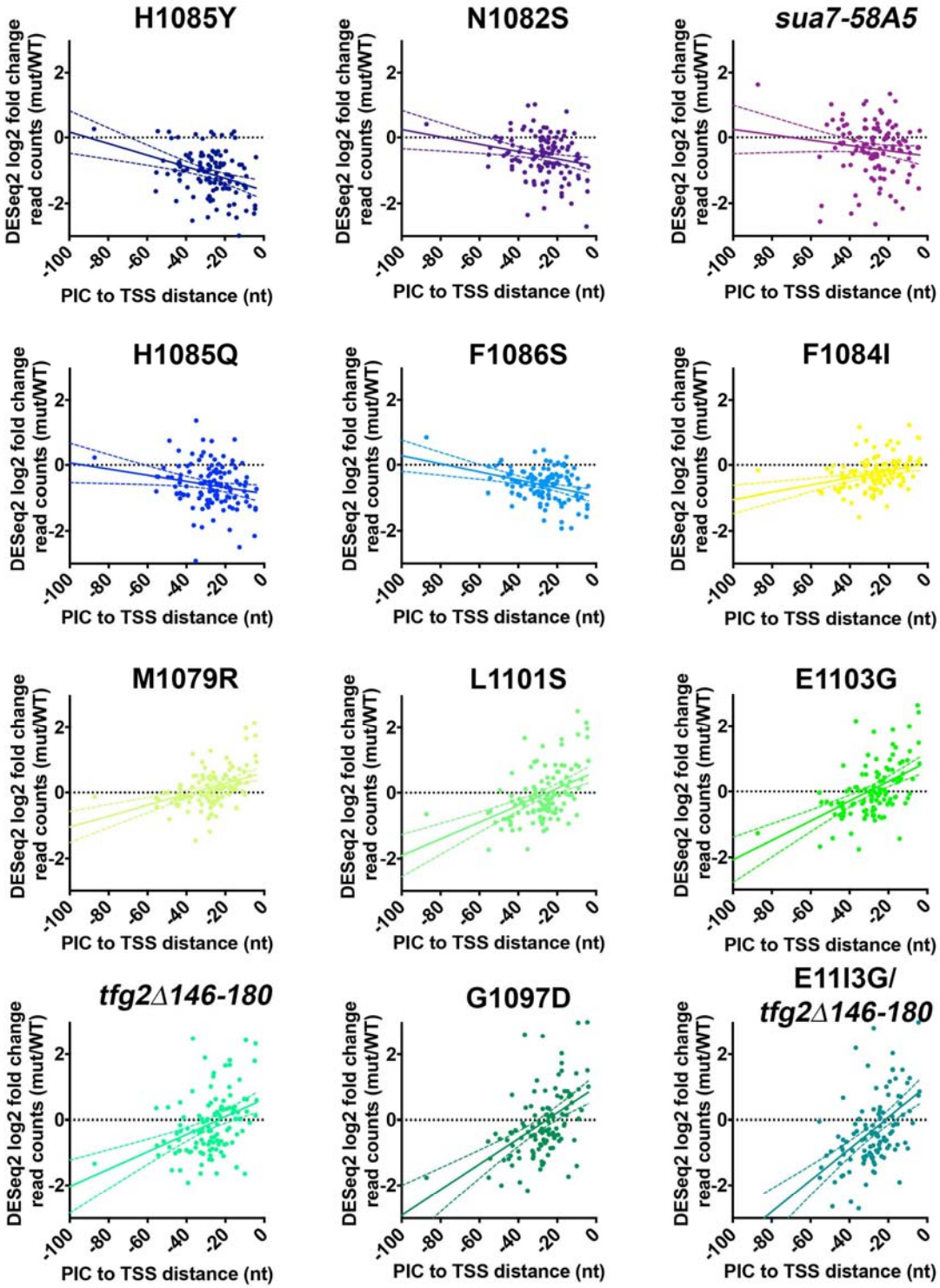
Supplemental Figure 6. Correlation between ChIP-exo replicates. Promoter-mapped ChIP-exo tags compared for two biological replicates for WT, *rpb1* E1103G, and *rpb1* H1085Y in Ssl2-TAP and Sua7-TAP strains.

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplemental Figure 7. GTF positioning in *rpb1* H1085Y and *rpb1* E1103G mutants as determined by ChiP-exo. Graphs show determined median position of ChiP-exo sequencing reads for top (TOP) or bottom (BOT) DNA strands for Sua7 (TFIIB) or Ssl2 (TFIIH) in Taf1-enriched promoters (left two columns) or Taf1-depleted promoters (right two columns) for WT, *rpb1* H1085Y, or *rpb1* E1103G strains. Promoters analyzed represent the top 50% of promoters as determined by ChiP-exo reads for **XXX** in WT cells. Box plots indicate distribution of determined positions relative to median TSS position (negative values indicate upstream positioning). Box plots are Tukey plots (see Methods). Asterisks indicate p≤0.05 as determined by Kruskal-Wallis test with Dunn's correction for multiple comparisons in Graphpad Prism 7.0e. Line graphs indicate histogram of ChiP-exo position determinations in WT, *rpb1* H1085Y, or *rpb1* E1103G strains (derived from data illustrated in box plot to left of each histogram).

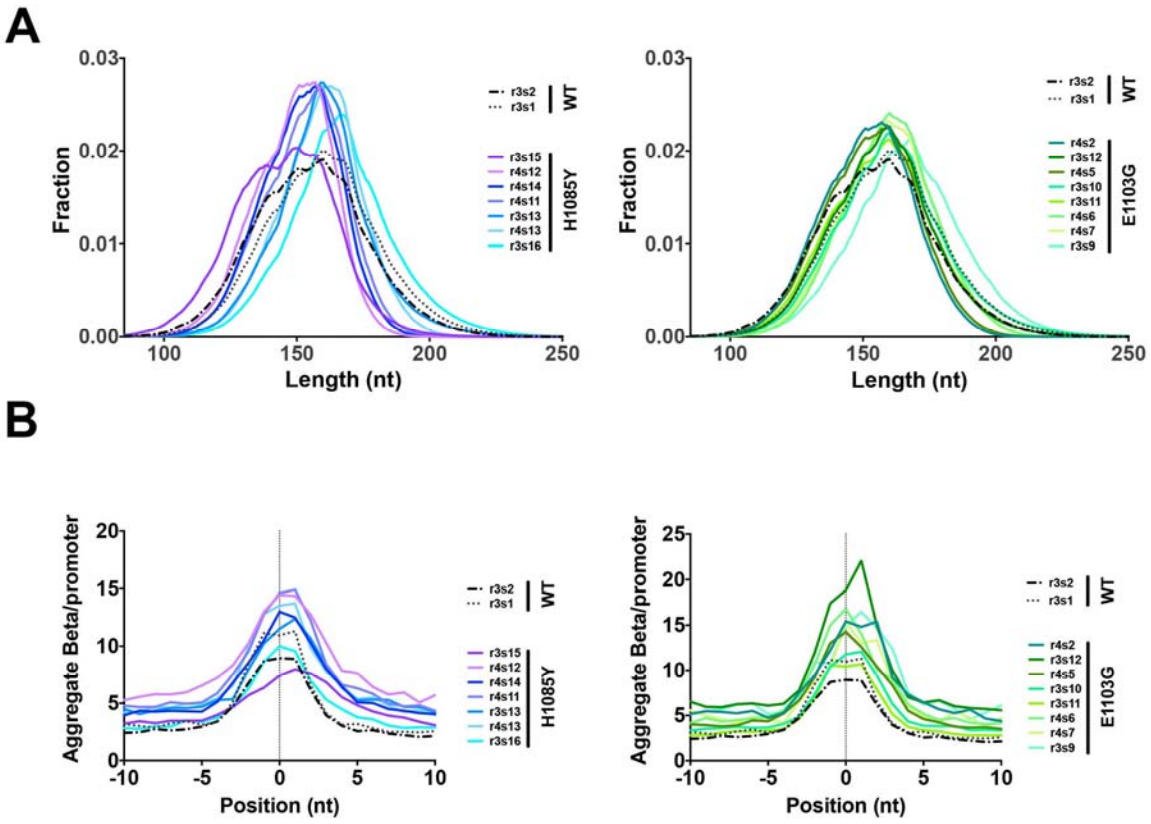
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplemental Figure 8. Correlation of apparent expression changes for ribosomal protein (RP) genes with PIC to TSS distance. PIC to TSS distance determined as in Figure 6.

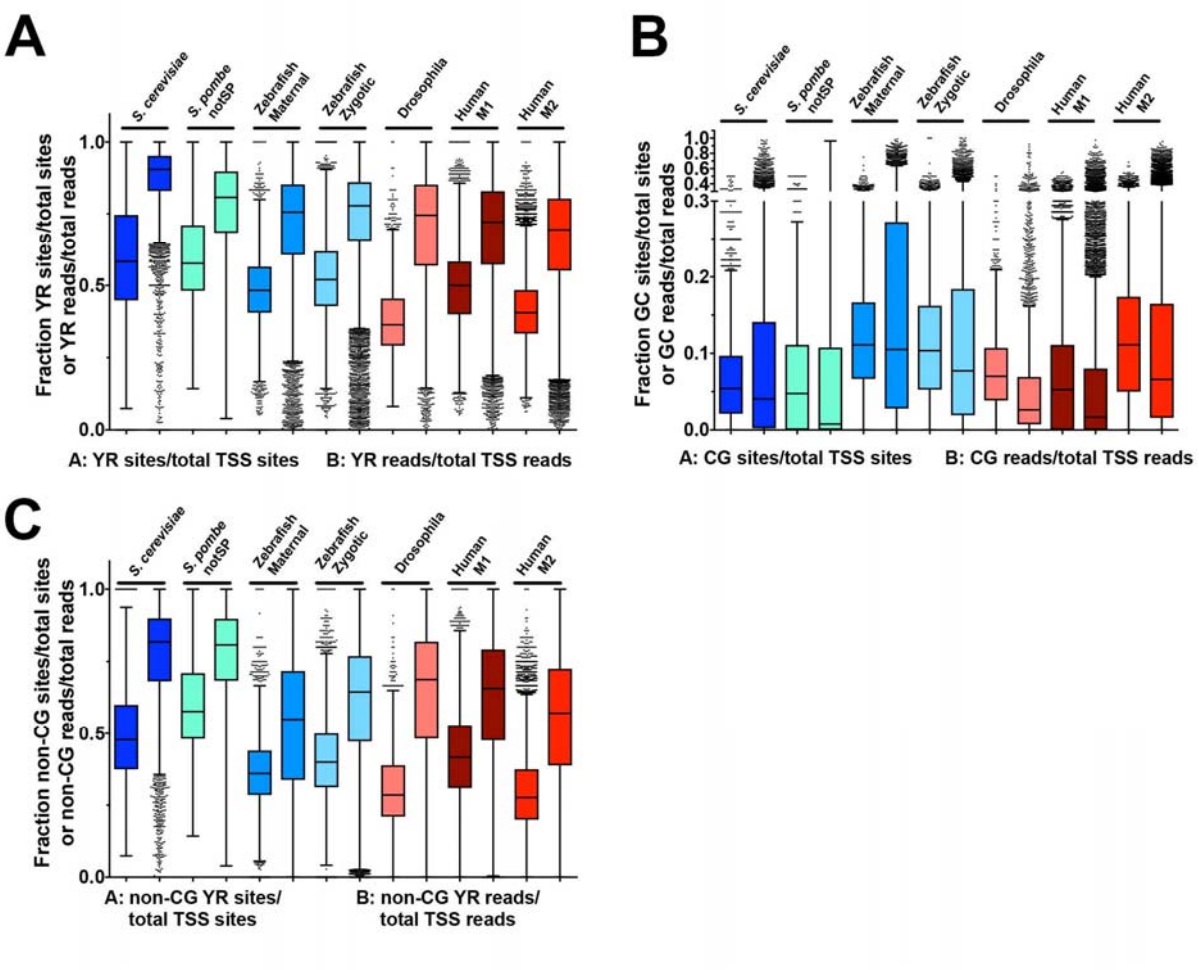
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Differential apparent expression determined by total read counts for promoter windows attributed to adjacent RP genes using DEseq2 analysis of individual TSS-seq libraries for WT and mutant yeast strains. Lines indicate linear regression lines and dashed lines indicate 95% confidence interval of the regression line.



Supplemental Figure 9. MNase-seq analyses of nucleosome positions in WT, *rpb1* H1085Y, and *rpb1* E1103G mutants. A. Paired-end sequencing fragment length distributions in WT and H1085Y MNase-seq libraries (left) and in WT (as left, shown for reference) and E1103G MNase-seq libraries (right). Libraries arranged within groups from most digested (top) to least digested (bottom). **B.** Probability of nucleosome positioning ("Beta") values determined by method of Zhou *et al* for MNase-seq libraries arranged as in (A).

bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

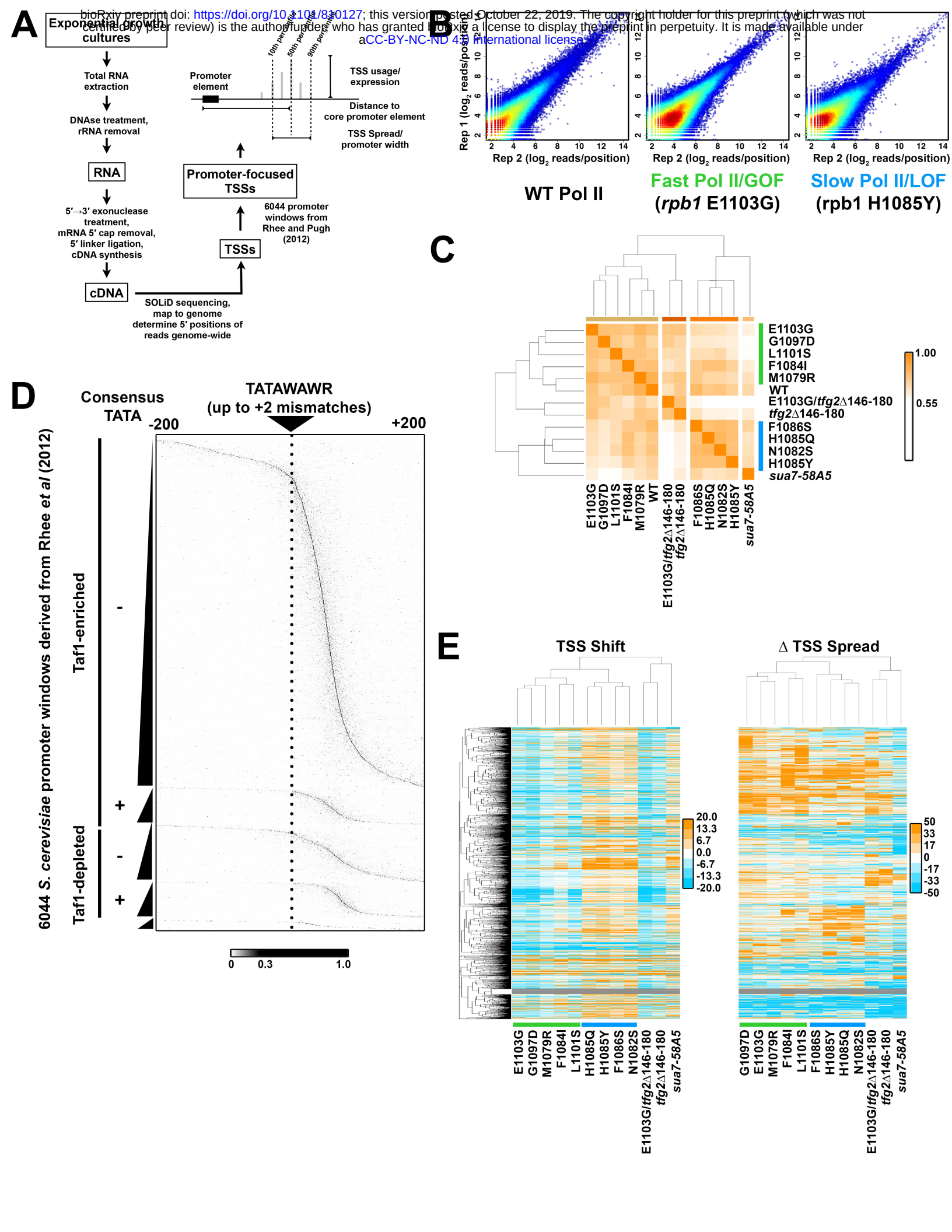


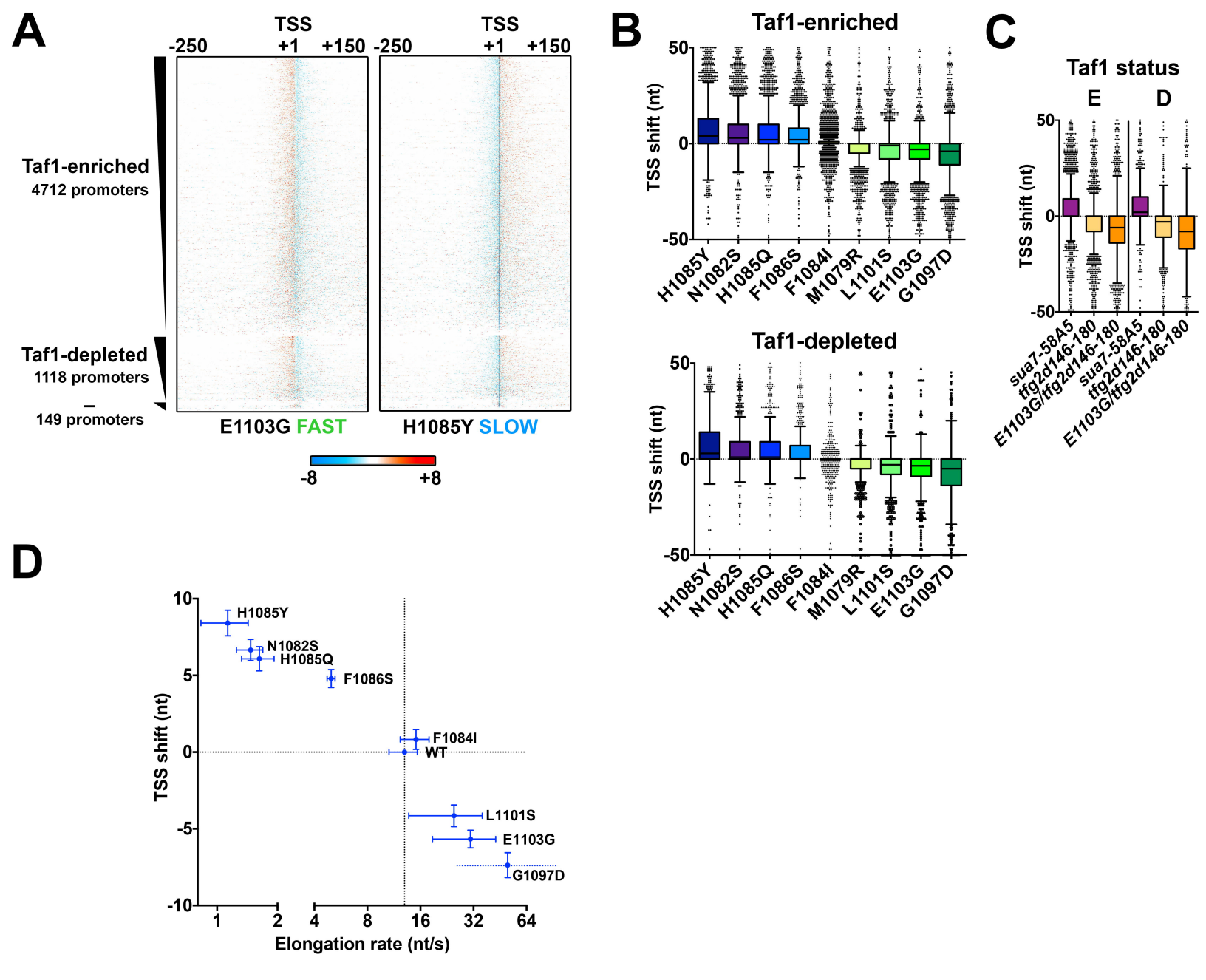
Supplemental Figure 10. $Y_{-1}R_{+1}$ usage preferences for select eukaryotes. A. Box plots indicate fraction of $Y_{-1}R_{+1}$ TSS sites (left boxes of paired data) or TSS reads (right boxes of paired data) relative to total TSS sites or reads observed for promoters with greater than 50 reads in data sets. Data sets are described in Figure 8. Box plots are Tukey plots (see Methods). **B.** Box plots indicate fraction of $C_{-1}G_{+1}$ TSSs (left boxes of paired data) or TSS reads (right boxes of paired data) relative to total TSS sites or reads observed for promoters with greater than 50 reads in data sets. Data sets are as in (A). **C.** Box plots indicate fraction of non-$C_{-1}G_{+1}$ $Y_{-1}R_{+1}$TSS sites (left boxes of paired data) or TSS reads (right boxes of paired data) relative to total TSS sites or reads observed for promoters with greater than 50 reads in data sets. Data sets are as in (A).

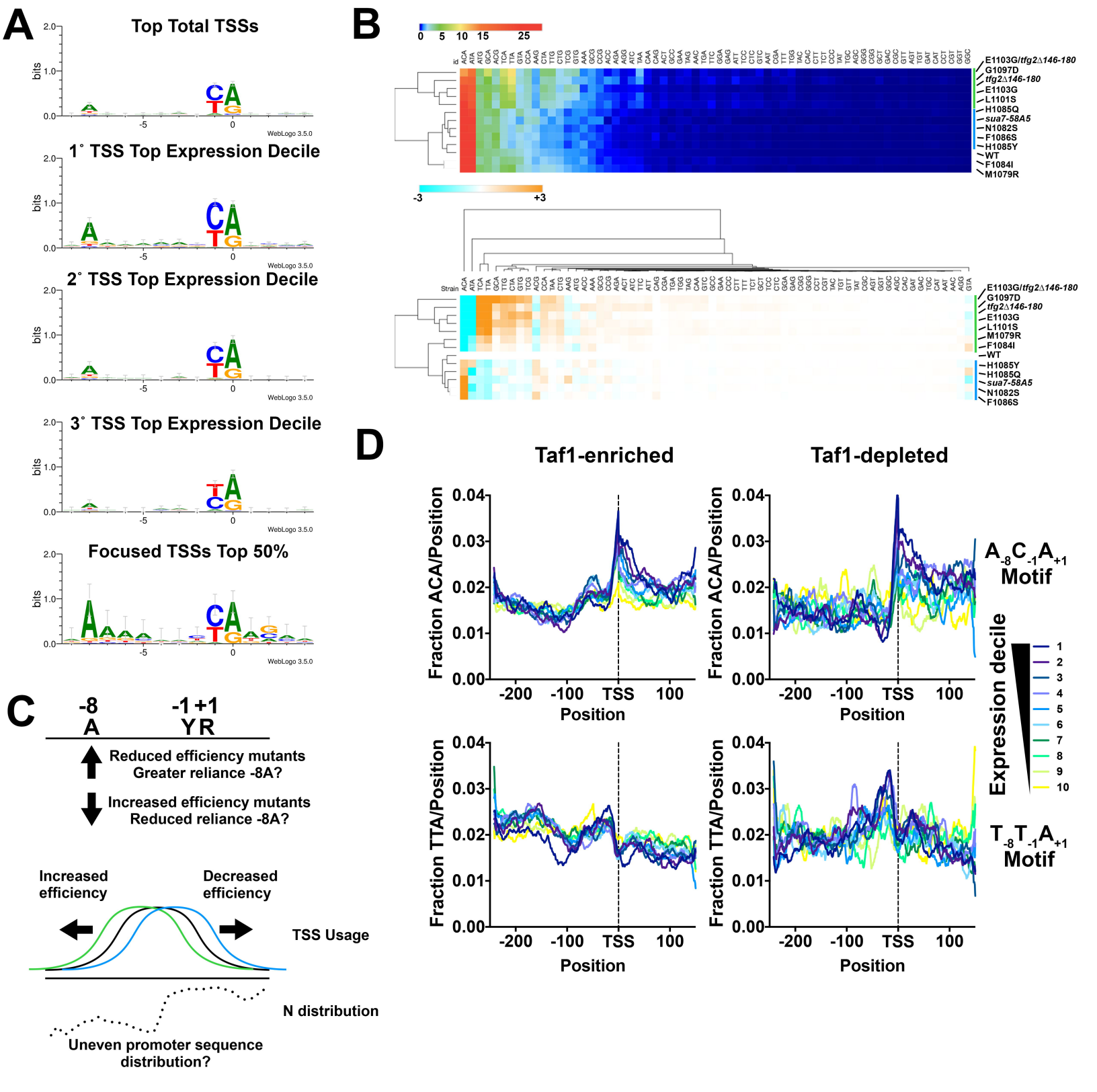
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1662

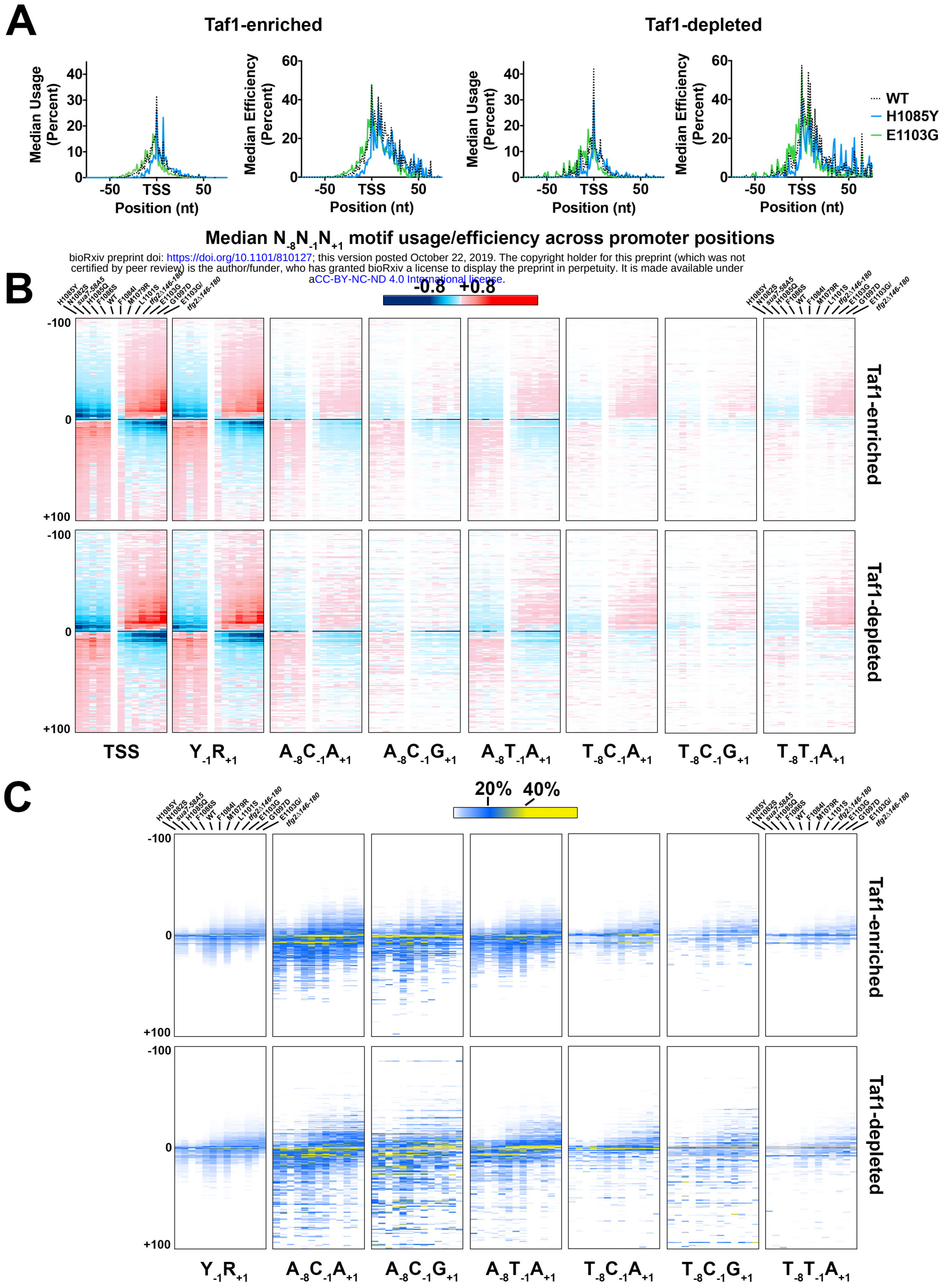
bioRxiv preprint doi: https://doi.org/10.1101/830127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

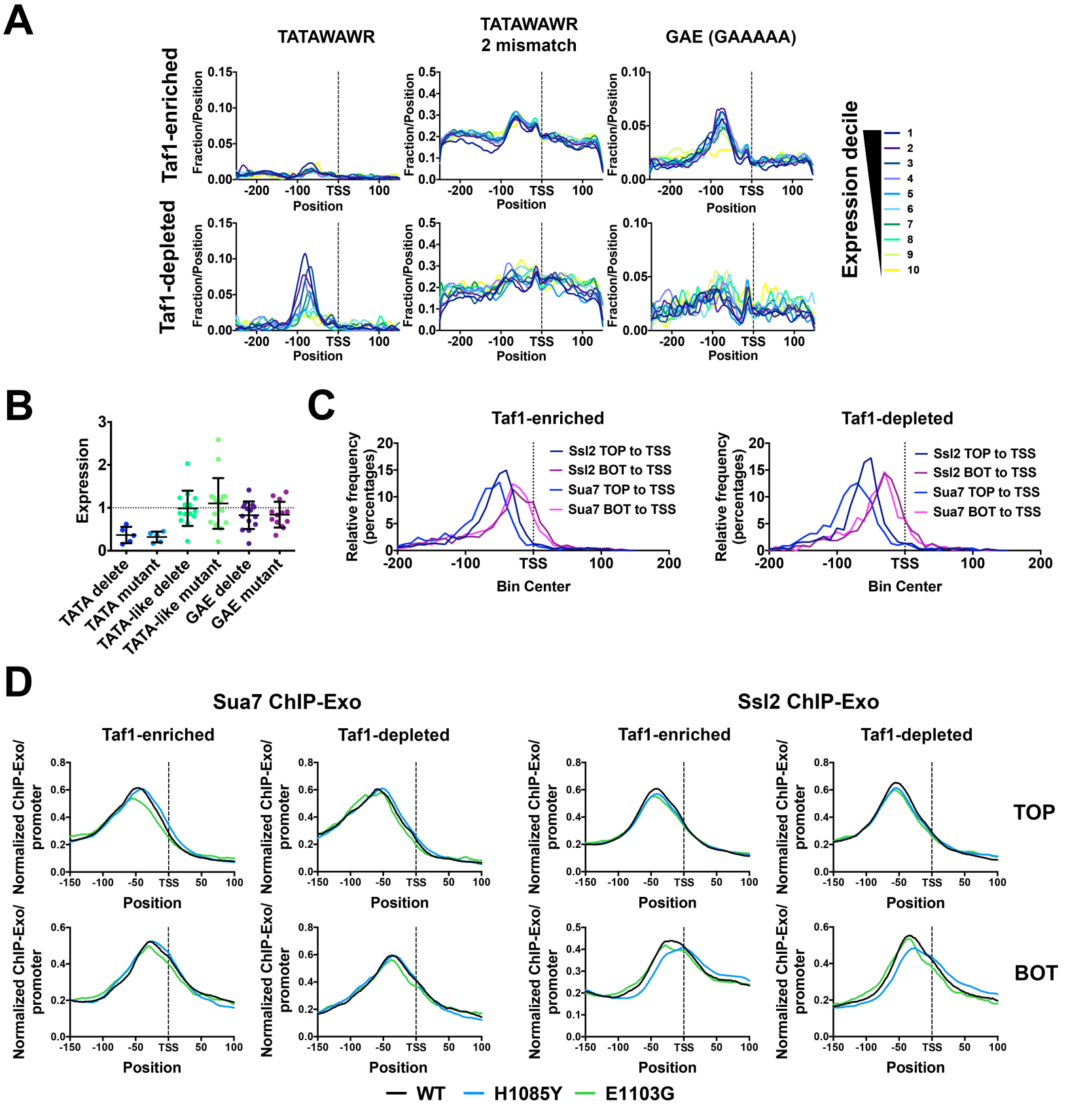


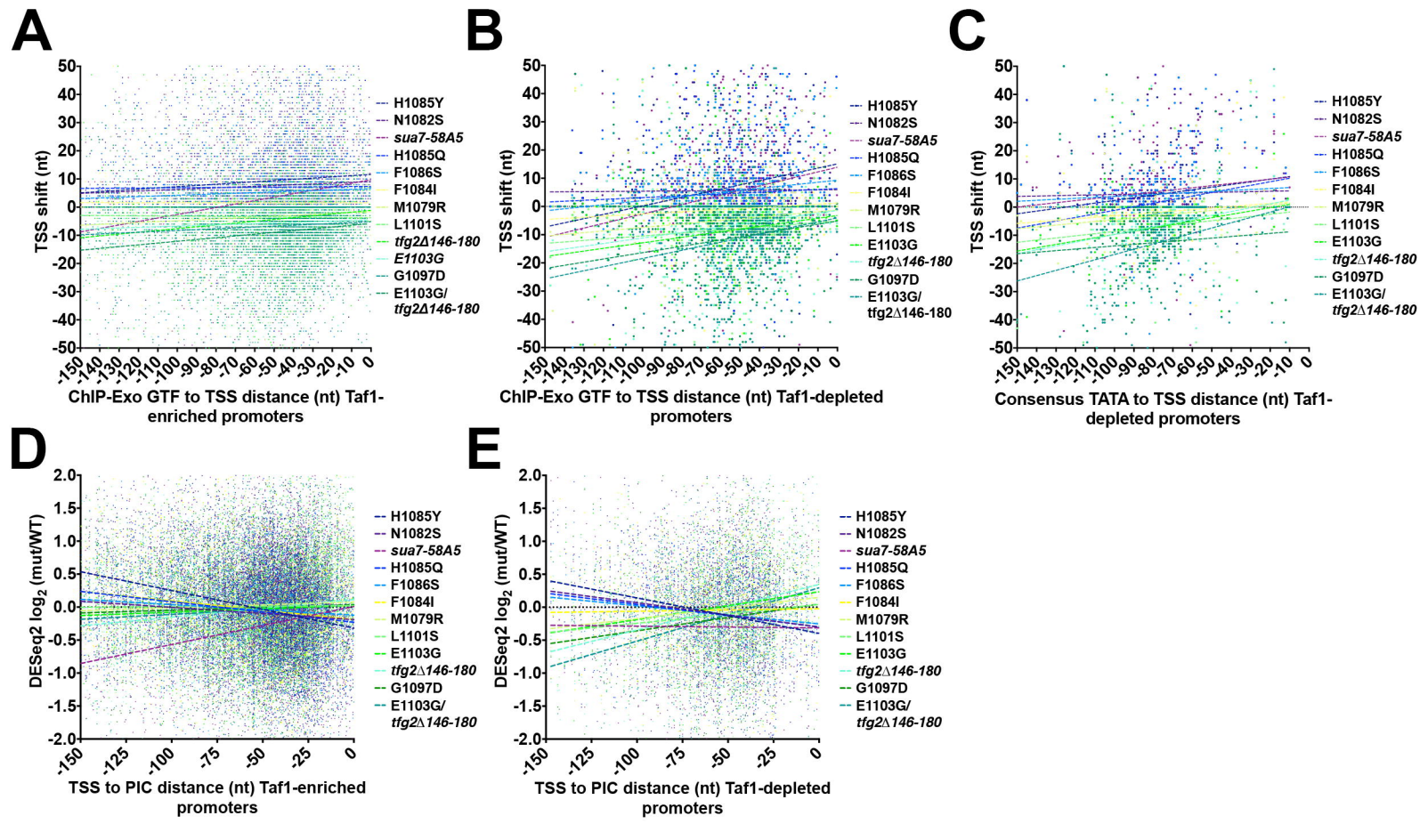

A
TSS
-250
+1
+150
TSS
-250
+1
+150
Taf1-enriched
4712 promoters
Taf1-depleted
1118 promoters
149 promoters
E1103G FAST
H1085Y SLOW
-8
+8
B
Taf1-enriched
TSS shift (nt)
50
0
-50
H1085Y
N1082S
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
G1097D
Taf1-depleted
TSS shift (nt)
50
0
-50
H1085Y
N1082S
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
G1097D
C
Taf1 status
E
D
TSS shift (nt)
50
0
-50
sua7-58A5
tfg2d146-180
E1103G/tfg2d146-180
sua7-58A5
tfg2d146-180
E1103G/tfg2d146-180
D
TSS shift (nt)
10
5
0
-5
-10
H1085Y
N1082S
H1085Q
F1086S
F1084I
WT
L1101S
E1103G
G1097D
1
2
4
8
16
32
64
Elongation rate (nt/s)


A
Top Total TSSs
1˚ TSS Top Expression Decile
2˚ TSS Top Expression Decile
3˚ TSS Top Expression Decile
Focused TSSs Top 50%
bits
WebLogo 3.5.0
B
E1103G/tfg2Δ146-180
G1097D
tfg2Δ146-180
E1103G
L1101S
H1085Q
sua7-58A5
N1082S
F1086S
H1085Y
WT
F1084I
M1079R
Strain
C
-8 A
-1 +1 YR
Reduced efficiency mutants
Greater reliance -8A?
Increased efficiency mutants
Reduced reliance -8A?
Increased efficiency
Decreased efficiency
TSS Usage
N distribution
Uneven promoter sequence distribution?
D
Taf1-enriched
Taf1-depleted
Fraction ACA/Position
Fraction TTA/Position
Position
TSS
A-8C-1A+1 Motif
T-8T-1A+1 Motif
Expression decile

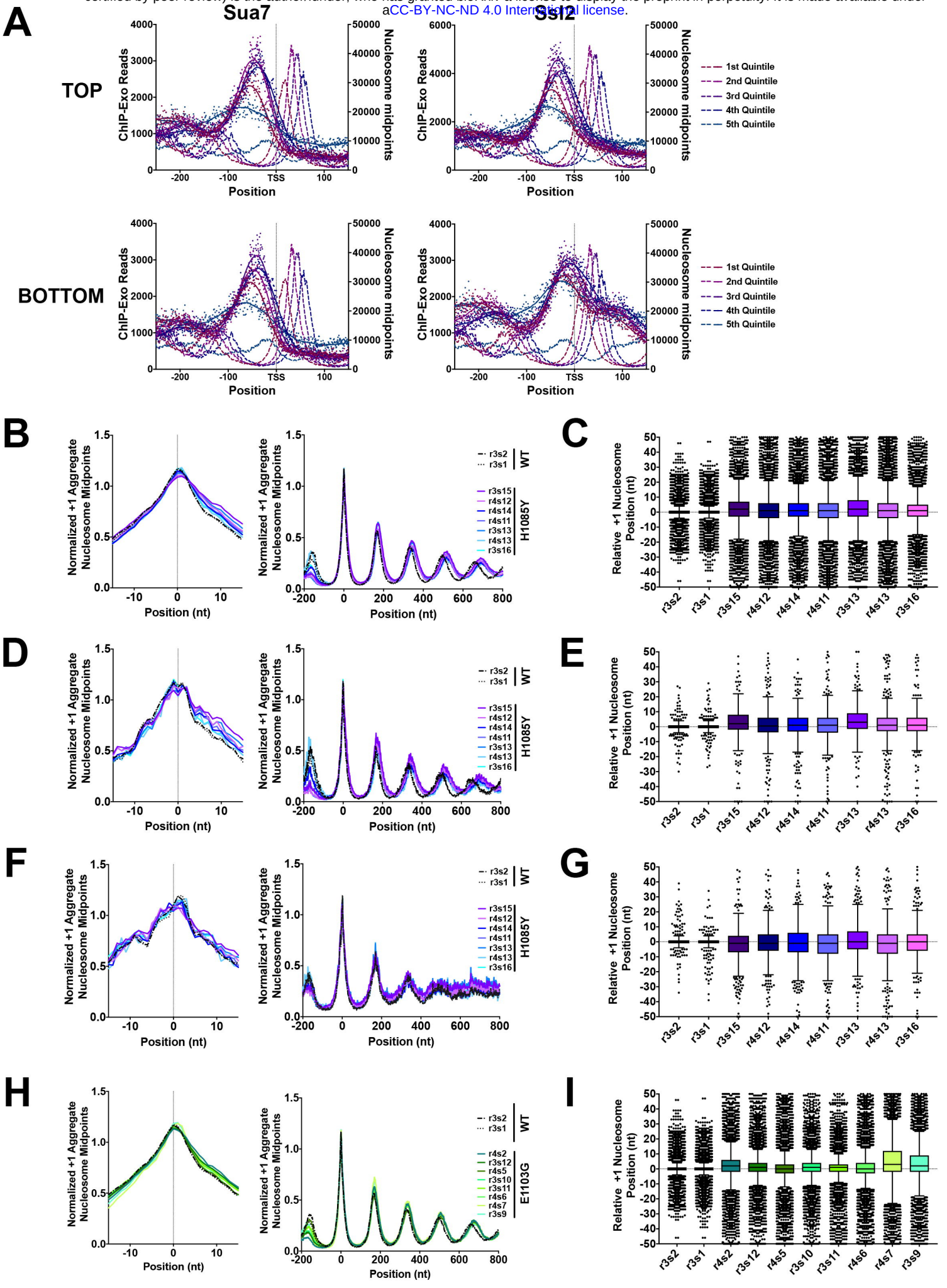
bioRxiv preprint doi: https://doi.org/10.1101/810127; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

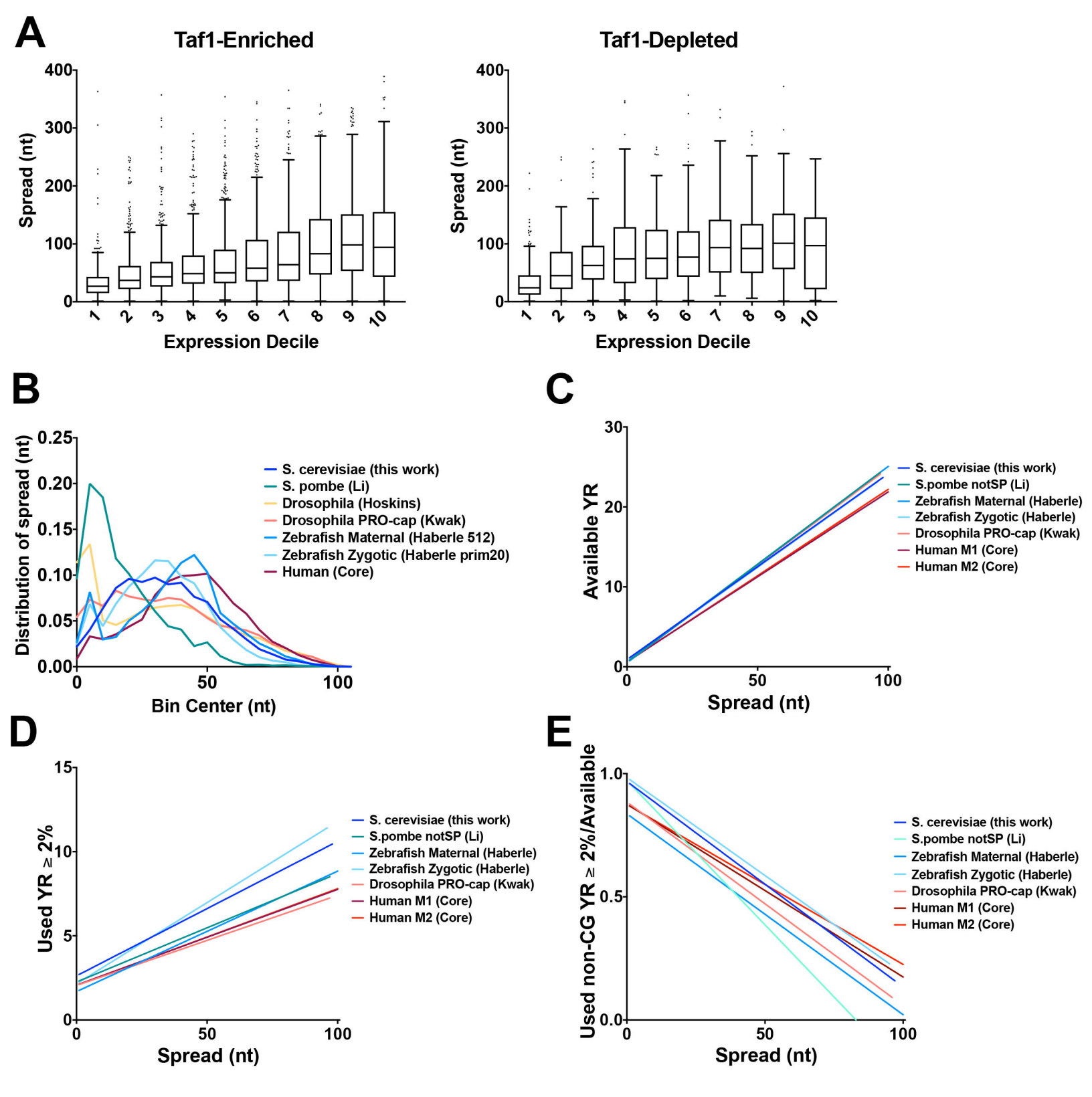



A
TATAWAWR
TATAWAWR
2 mismatch
GAE (GAAAAA)
Taf1-enriched
Taf1-depleted
Fraction/Position
Position
-200
-100
TSS
100
Expression decile
1
2
3
4
5
6
7
8
9
10
B
Expression
TATA delete
TATA mutant
TATA-like delete
TATA-like mutant
GAE delete
GAE mutant
C
Taf1-enriched
Taf1-depleted
Relative frequency
(percentages)
Ssl2 TOP to TSS
Ssl2 BOT to TSS
Sua7 TOP to TSS
Sua7 BOT to TSS
Bin Center
D
Sua7 ChIP-Exo
Ssl2 ChIP-Exo
Taf1-enriched
Taf1-depleted
Normalized ChIP-Exo/
promoter
Position
TOP
BOT
WT
H1085Y
E1103G


A
TSS shift (nt)
ChIP-Exo GTF to TSS distance (nt) Taf1-enriched promoters
H1085Y
N1082S
sua7-58A5
H1085Q
F1086S
F1084I
M1079R
L1101S
tfg2Δ146-180
E1103G
G1097D
E1103G/ tfg2Δ146-180
B
TSS shift (nt)
ChIP-Exo GTF to TSS distance (nt) Taf1-depleted promoters
H1085Y
N1082S
sua7-58A5
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
tfg2Δ146-180
G1097D
E1103G/ tfg2Δ146-180
C
TSS shift (nt)
Consensus TATA to TSS distance (nt) Taf1-depleted promoters
H1085Y
N1082S
sua7-58A5
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
tfg2Δ146-180
G1097D
E1103G/ tfg2Δ146-180
D
DESeq2 log2 (mut/WT)
TSS to PIC distance (nt) Taf1-enriched promoters
H1085Y
N1082S
sua7-58A5
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
tfg2Δ146-180
G1097D
E1103G/ tfg2Δ146-180
E
DESeq2 log2 (mut/WT)
TSS to PIC distance (nt) Taf1-depleted promoters
H1085Y
N1082S
sua7-58A5
H1085Q
F1086S
F1084I
M1079R
L1101S
E1103G
tfg2Δ146-180
G1097D
E1103G/ tfg2Δ146-180

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




A
Taf1-Enriched
Taf1-Depleted
Spread (nt)
Expression Decile
B
Distribution of spread (nt)
Bin Center (nt)
S. cerevisiae (this work)
S. pombe (Li)
Drosophila (Hoskins)
Drosophila PRO-cap (Kwak)
Zebrafish Maternal (Haberle 512)
Zebrafish Zygotic (Haberle prim20)
Human (Core)
C
Available YR
Spread (nt)
S. cerevisiae (this work)
S.pombe notSP (Li)
Zebrafish Maternal (Haberle)
Zebrafish Zygotic (Haberle)
Drosophila PRO-cap (Kwak)
Human M1 (Core)
Human M2 (Core)
D
Used YR ≥ 2%
Spread (nt)
S. cerevisiae (this work)
S.pombe notSP (Li)
Zebrafish Maternal (Haberle)
Zebrafish Zygotic (Haberle)
Drosophila PRO-cap (Kwak)
Human M1 (Core)
Human M2 (Core)
E
Used non-CG YR ≥ 2%/Available
Spread (nt)
S. cerevisiae (this work)
S.pombe notSP (Li)
Zebrafish Maternal (Haberle)
Zebrafish Zygotic (Haberle)
Drosophila PRO-cap (Kwak)
Human M1 (Core)
Human M2 (Core)