bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

RNA Synthesis is Associated with Multiple TBP-Chromatin Binding Events

Hussain A. Zaidi[1], David T. Auble[1], Stefan Bekiranov[1]

[1]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA 22908

Cellular processes including transcription are inherently dynamic. Currently, the dynamics of transcription and other molecular processes in the cell are poorly understood[1] because of a lack of methods that measure fundamental kinetic parameters in vivo. Precise estimation of the chromatin-binding on- and off-rates of general transcription factors (TFs) would allow stochastic modeling of pre-initiation complex formation[2,3], RNA polymerase recruitment and elongation, and transcription[4,5]. Live-cell imaging at specific multi-copy genes is capable of yielding the residence time of TF-chromatin interactions at high temporal resolution (i.e., second timescale)[6] but does not allow these measurements at single-copy genes. Cross-linking kinetic (CLK) analysis is a high spatial and temporal resolution method that enables estimation of the in-vivo TF-chromatin on- and off-rates at single-copy loci[7]. However, alternative approaches to estimating these kinetic parameters are needed to independently verify CLK as well as live-cell imaging approaches[8]. We developed and applied a physical modeling approach using chemical kinetic theory that directly estimates the residence time of TATA-binding protein (TBP) across the yeast genome from TBP competition ChIP data[9]—another high-spatial resolution method[10,11], which was generally believed to be low temporal resolution (20 min or greater). Using this approach, we are capable of estimating TBP-chromatin residence times on the minute time-scale across the yeast genome, demonstrating that competition ChIP is actually a relatively high temporal resolution method. Comparing TBP-chromatin residence times with nascent RNA transcription rates[12], we find that ~5 TBP binding events are associated with productive RNA synthesis at the typical gene. Our results paint, for the first time, a highly dynamic, stochastic picture of pre-initiation complex formation with multiple rounds of partial assembly and disassembly before productive RNA polymerase elongation.

We use kinetic theory to model the in vivo competitive dynamics of the induced competitor and the endogenous TBP in competition ChIP experiment[9] with TF-chromatin binding on-rate (k_a) and off-rate (k_d) (Supplemental Text Sec 1). In order to fit TBP competition ChIP two-color Agilent tiling microarray data[9] to our kinetic model, we first normalize, background subtract, and scale the data for TBP peaks within gene promoters to theoretically expected values at the start of induction ($t = 0$) and steady state or equilibrium ($t \rightarrow \infty$) (Supplemental Text and Extended Data Fig. 1). The kinetic theory explicitly accounts for induction of the competitor. Consequently, we fit the ratio of the induced (denoted by B) over endogenous (denoted by A) TBP concentration ($C_B(t)/C_A$) determined from Western blots as a function of induction time[9] to a Hill-like sigmoid function with Hill coefficient $n = 4$ (Fig. 1a). The fit yields a characteristic time-scale

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

for TBP induction $t_0^{ind} = 22$ min and the steady state ratio of induced over endogenous TBP concentration ($C_B(t \to \infty)/C_A \to 2.233$). The normalized competition ChIP data is also well approximated by an $n = 4$ Hill-like equation with a time-scale parameter t_0 (Supplemental Eqn. 3), which quantifies the overall turnover response including induction and turnover dynamics at every TBP peak. We use this Hill-like equation to background subtract and scale the data to the theoretical in-vivo (denoted by superscript i) ratio of fractional occupancy of the competitor $\theta_B^i(t)$ to the endogenous $\theta_A^i(t)$ TBP which must satisfy the boundary conditions at the start of induction ($\theta_B^i(0)/\theta_A^i(0) = 0$) and steady state ($\theta_B^i(t)/\theta_A^i(t) \to 2.233$ as $t \to \infty$) (Supplemental Eqn. 24 and Supplemental Text).

We simultaneously numerically solve and fit the in vivo chemical kinetic equations of competitive binding between species A and B (Supplemental Eqns. 1 and 2) to normalized, scaled competition ChIP data (see Extended Data Figs. 2 and Supplemental Text). We (and others[9-11]) ignore the impact of cross-linking theoretically as competition ChIP data was gathered at one cross-linking time (20 min of formaldehyde cross-linking) and show that the resulting off-rate, k_d, could be modestly biased (Extended Data Fig. 3 a-d) using a generalization of the CLK framework to competition ChIP (Supplemental Eqns. 4-8). This framework could be used to correct the bias if data is gathered at various crosslinking times[7]. As noted by Lickwar et al.[11], we also find that the in vivo ratio of induced over endogenous TF as a function of induction time is insensitive to the on-rate, k_a, and is very sensitive to the off-rate or residence time, $t_{1/2} = \ln(2)/k_d$ (Extended Data Fig. 3 e-f). Consequently, we only arrive at relatively precise values of the residence time ($t_{1/2}$) by fitting the kinetic equations to competition ChIP ratio data.

Stratifying TBP-containing promoters in 2-minute bands of t_0, we show that the average normalized, scaled ratio of competitor over endogenous signals as a function of induction time progressively shows slower rise as t_0 increases (i.e., moves to the right) (Fig. 1b) with corresponding residence times increasing from 1.4 to 70 minutes (Fig. 1c), showing that residence times could be estimated from the ratio. Indeed, given that fitting the Hill-like equation and chemical kinetic equations should yield highly correlated results, we find a smooth relationship between $t_{1/2}$ and t_0 up to a point where numerically fitting the chemical kinetic equations becomes unstable at $t_0 < 24.5$ min (Fig. 1d and Extended Data Fig. 4). This numerical instability is due to the fact that for promoters with $t_0 < 24.5$ min, the separation between the normalized, scaled data and the induction curve are well within the noise of the competition ChIP data. For $t_0 > 24.5$ min, the normalized, scaled data yield excellent fits to the chemical kinetic equations, move progressively to the right with increasing residence time and, remarkably, allow residence times as short as 1.3 minutes to be estimated (Figs. 1 e-g, Extended Data Fig. 4 e-h, and Extended Data Table 1). Importantly, the lower bound on the time resolution is determined by the noise in the induction and competition ChIP data and not the induction time of the competitor.

Because a physical model of competitive binding rooted in reaction-rate theory naturally follows the profiles of the normalized, scaled data as a function of induction time, we are able to apply stringent noise criteria on the residuals of each fit (Supplemental Text) and reliably estimate TBP residence times for 794 Pol II and 205 Pol III genes (Extended

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

Data Table 1). This is in comparison to earlier estimates of relative turnover, r, for 602 Pol II and 264 Pol III genes that were obtained using linear regression to a subset of the data (i.e., 10, 20, 25 and 30 min time points)[9]. While r and our estimates of k_d are correlated (Extended Data Fig. 5a), r is also strongly correlated with the $t = 0$ ratio of induced over competitor ChIP signals (Extended Data Fig. 5), which suggests insufficient background subtraction influencing the estimates of r. Nevertheless, in agreement with estimates of r made by van Werven et al.[9], we find TBP residence times are notably shorter for Pol II compared to Pol III genes (Fig. 2a) and to a lesser extent for TATA compared to TATA-less genes[13] (Fig. 2b). In contrast to van Werven et al.[9], we find no significant differences between TBP residence times comparing SAGA containing and SAGA free genes (Extended Data Fig. 6d) or TFIID containing and TFIID free genes (Extended Data Fig. 6g). Given that Pol III genes tend to be higher expressed[14] and have longer TBP residence times than Pol II genes, we were surprised to find marginally shorter TBP residence times at highly expressed ribosomal protein (RP) genes (Extended Data Fig. 6j) compared to other genes (Fig. 2c). This finding was consistent with higher nascent RNA transcription rates (TRs)[12] for shorter TBP residence times at Pol II genes (Fig. 2d). Shorter residence times were also associated with higher levels of transcriptional extrinsic noise[15] (Fig. 2e) consistent with recent findings[16]. With estimates of TR and TBP $t_{1/2}$, we defined *transcriptional efficiency*, which is the product of the transcription rate and TBP residence time ($TRt_{1/2}$) and represents the number of TBP residence times or binding events associated with productive elongation of Pol II and transcription. Strikingly, we find low transcriptional efficiencies for Pol II genes (Fig. 2f). The median $TRt_{1/2}$ across Pol II promoters is ~5 TBP binding events for productive RNA synthesis to proceed (Fig. 2f). These findings are consistent with a stable scaffold model of PIC formation[17] and function for Pol III genes, and rapid, highly stochastic TPB/PIC dynamics at higher expressed Pol II genes.

To gain further insights into the upstream regulation and/or downstream impact of TPB-chromatin binding dynamics especially on regulation of gene expression, we compared TBP residence times to previously derived Rap1[11] and nucleosome[10] relative turnover rates (λ) and their relative residence times (λ^{-1}). Notably, we show that λ equals k_d plus a time-dependent function (Supplemental Eqn. 28, Extended Data Fig. 7) and can be excessively biased if normalized ChIP ratios are not properly scaled to ratios of fractional occupancies before model fitting[11] (Supplemental text and Extended Data Fig. 8). Nevertheless, we find TBP residence time is correlated with Rap1 λ^{-1} at non-RP Pol II genes but not at RP Pol II genes (Fig. 3a). TBP $t_{1/2}$ (corr $= -0.11$; Extended Data Fig. 9a) and Rap1 λ^{-1} (Fig. 3b) show weak negative and positive correlation with transcription rate, respectively, while transcriptional efficiency is modestly correlated with Rap1 relative residence time at non-RP Pol II genes (Fig. 3c). Interestingly, the majority of sites, for which Rap1 relative residence times were determined (ranging from 30-150 min), contain highly dynamic TBP ($t_{1/2} < 1.4$ min or $t_0 < 24.5$ min; Extended Data Fig. 9b), further supporting our findings that Rap1 relative residence times[11] are 20 to 30 fold higher than, but likely correlated with, Rap1 $t_{1/2}$. While +1 nucleosome dynamics are poorly correlated with TBP residence time (Figs. 3d, Extended Data Figs. 9 c,d), they are positively correlated with transcription rate (Figs. 3e, Extended Data Fig.

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

9e) and efficiency (Figs. 3f, Extended Data Fig. 9f). These results suggest that while the dynamics and not merely the presence (Extended Data Fig. 6m) of transcription factors like Rap1 potentially regulate TBP/PIC dynamics, TBP and Rap1 recruitment and dynamics are not the rate-limiting step in transcription at Pol II genes. Conversely, the dynamics of factor that regulate elongation including +1 nucleosome turnover may play a more critical role in determining transcription rate and efficiency.

To further assess the hypothesis that transcription factors associated with elongation as opposed to PIC and Pol II recruitment or initiation are the rate limiting step in transcription, we tested the effect that the presence or absence of 202 transcription factors mapped to the yeast genome[18] had on TBP residence time, transcription rate and transcription efficiency. We subdivided loci for which we had estimates of TBP residence time into quartiles of the number transcription, initiation and elongation factors, respectively. As expected, the presence of greater numbers of transcription, initiation and elongation factors at promoters had no significant impact on TBP residence times (Figs. 4a-c) but yielded higher transcription rates (Figs. 4d-f) and efficiencies (Fig. 4g-i). Strikingly, the presence of more elongation factors had a much greater impact on both transcription rate (Fig. 4f) and efficiency (Fig. 4i) compared to that of initiation factors (Figs. 4e,h), consistent with our hypothesis. For each of the 202 factors, we also assessed the significance of differences of TBP residence times, transcription rate and efficiency at sites with the factor present compared to sites with that factor absent. We only found one factor, Tfa2 (a TFIIE subunit), whose presence or absence yielded significantly different residence times (Fig. 5a). However, Tfa2 had a more modest effect on transcription rate (Fig. 5b) and no significant effect on efficiency (Fig. 5c). In contrast, we find that 46% and 50% of all the initiation and elongation factors mapped, respectively, significantly modulate transcription rate and efficiency (Fig. 5d,e). Not surprisingly, many of these factors are members of initiation and elongation complexes whose enrichment at promoters leads to both increased transcription rate and efficiency (Fig. 5f).

We find that the dynamics as opposed to the presence of factors like Rap1 affect TBP residence time, suggesting a number of unknown dynamic relationships and regulatory roles for transcription factors that await characterization and testing using methods that measure dynamic parameters including residence time and transcriptional efficiency. These approaches and measurements should ultimately allow the stochastic processes of pre-initiation complex formation, Pol II recruitment and elongation and transcription to be characterized quantitatively.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank members of the Bekiranov lab, particularly Brian Capaldo, and the Auble lab, especially Elizabeth Hoffman, for helpful suggestions. This research was supported by NIH grant R21 GM110380 (awarded to S.B and D.T.A) and NIH grant R01 GM55763 (awarded to D.T.A).

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

Author Contributions S.B. conceived the approach and guided method development and data analysis. H.A.Z. developed and applied the methods and performed the data analysis. D.T.A. made critical analysis suggestions. H.A.Z., D.T.A. and S.B. interpreted the results and wrote the paper.

Author Information The authors declare no competing financial interests. Correspondence and requests for materials, software and information should be addressed to S.B. (sb3de@virginia.edu) or H.A.Z (haz4z@virginia.edu).

Methods

Background subtraction, normalization and scaling of competition ChIP data. The raw data generated by van Werven et al.[9] (ArrayExpress E-M-TAB-58) reported the optical signal intensity for induced (S_{HA}) and endogenous (S_{Avi}) protein concentrations hybridized on an Agilent whole-genome microarray. S_{HA} and S_{Avi} were replicated by swapping Cy_3 and Cy_5 dyes to take into account dye-specific variations in the intensity of the optical signal. In our analysis, we used the geometric average of the two dye-swapped ratios, R_m, as described in Supplemental Text. Non-specific background probes were identified by fitting a normal curve to the right edge of the $t = 0$ minute $\log_2(S_{HA}/S_{Avi})$ data as shown in Extended Data Figure 1. For our analysis, we selected signal probes in the tail of the normal fit to the non-specific background with a false discovery rate (FDR) of 0.05 or less in the $t = 0$ minute data. R_m values were normalized across time points, t, by dividing them by the background mean obtained from the normal fit to the background probes (Extended Data Figure 1). To quantify the induction of HA over time, we fit a Hill-like sigmoid curve with $n = 4$ to the ratio of the concentration of HA over Avi $\left(\frac{C_B^i(t)}{C_A^i}\right)$, where A and B denote Avi and HA, respectively. The fit gives an induction time (t_0^{ind}) of 22 minutes and the saturation value of HA/Avi concentration ratio of 2.23 (Supplemental Eqn. 3, and Fig. 1a). We relate the empirical values of normalized R_m for the signal probes in our data to the ratio of the in-vivo fractional occupancy of HA (θ_B^i) and Avi (θ_A^i) as $\alpha R_m - B = \frac{\theta_B^i}{\theta_A^i}$, where B is the locus-specific differential background between HA and Avi at $t = 0$ minutes, and α denotes a scale factor which effectively quantifies the ratio of the antibody affinities for HA and Avi (Supplemental Text Section 2). To determine α and B at every TBP peak, a Hill-like sigmoid curve (with $n = 4$) with the added term B was fit to the normalized R_m (Supplemental Eqn. 24). B was subtracted from R_m and α was determined as the ratio of the asymptotic in-vivo concentration ratio of HA/Avi (2.23) over the asymptotic R_m value. Hence, after scaling and background subtraction, $\alpha R_m - B$ satisfied the two boundary conditions: $\frac{\theta_B^i}{\theta_A^i} = 0$ for $t = 0$, and $\frac{\theta_B^i}{\theta_A^i} \to 2.23$ as $t \to \infty$, as required by the kinetic model of in-vivo competitive binding.

Estimation of residence time by fitting chemical kinetic theory based model of competitive binding to normalized, scaled competition ChIP data. The model for in-vivo competitive binding dynamics between endogenous Avi (subscript A) and

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

competitor HA (subscript B) TBP is described by mass-action differential equations linear in the TBP-chromatin association rate (k_a^i) and dissociation rate (k_d^i):

$$\frac{d\theta_B^i}{dt} = -k_d^i \theta_B^i + k_a^i \frac{C_B^i(t)}{C_A^i}\left(1 - \theta_A^i - \theta_B^i\right) \quad (1)$$

$$\frac{d\theta_A^i}{dt} = -k_d^i \theta_A^i + k_a^i \left(1 - \theta_A^i - \theta_B^i\right) \quad (2)$$

We have assumed that the association and dissociation rates for endogenous and competitor TBP are the same, superscript *i* denotes "in-vivo", and we have absorbed the experimentally undetermined endogenous concentration (C_A^i) into k_a^i, such that k_a^i and k_d^i have units of inverse minutes (Supplemental Text Section 2). Equations (1) and (2) cannot be solved analytically due to the time dependence of C_B^i, but a solution can be derived for the case when the induction of HA is instantaneous: $\frac{C_B^i(t)}{C_A^i} = 0$ for $t < 0$ and $\frac{C_B^i(t)}{C_A^i} = constant$ for $t \geq 0$. Inserting the actual time dependent $\frac{C_B^i(t)}{C_A^i}$ in the ideal induction solution gives an approximate solution to the Equations (1) and (2) (Supplemental Eqns. 19 and 20, Extended Data Fig. 2).

We fit the approximate solution of ideal induction to the normalized, scaled ratio data using NonlinearModelFit in Mathematica (Supplemental Text Section 5.1). The ratio data is almost insensitive to k_a^i, and hence, we can reliably extract only k_d^i. The approximate solution introduces a bias in our estimate of k_d^i, which we fix using a pre-generated look-up table (Supplemental Text Section 5.2). Finally, we use our bias-corrected estimates as the starting point for a numerical one-dimensional Newton's method fit of Equations (1) and (2) to find the minimum of the fit residual and extract k_d^i (Supplemental Text Section 5.3). To calculate the derivative of the fit residual required at each iteration of Newton's method, we numerically solve the in-vivo differential equations using NDSolve in Mathematica. Exceptions to the fitting procedure where we had to change the starting estimate of k_d^i or the step size for Newton's method are noted in Supplemental Text Section 5.

Statistical analyses of residence time, transcription rate and transcription efficiency data. Throughout the main text and the supplement, quoted correlations are Spearman correlation coefficients unless otherwise stated. Kolmogorov-Smirnoff (KS) test was conducted in R using the ks.test function to determine the p-values reported in Figs. 2, 3, and 4 of the main text. For Fig. 5a of the main text, permutation test (which is useful in particular when the test statistic does not follow a normal distribution) was used to calculate the false discovery rate (FDR) for $t_{1/2}$, TR, and $TRt_{1/2}$. In other words, loci across the genome were partitioned into two sets for each transcription factor: those that showed a significant enrichment of the transcription factor above the background as determined by Venters et al.[18], and those that did not. These two sets were used to conduct permutation test for $t_{1/2}$, TR, and $TRt_{1/2}$ using permTS in the perm library in R, which gives the mean difference between the two sets along with the p-value for the mean difference. The p-value was adjusted using the Benjamini-Hochberg correction[19] using the p.adjust function in R to derive FDR estimates. In Fig. 5d the FDR for $TRt_{1/2}$

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

was plotted against the FDR for TR, and transcription factors were listed in descending order of $TRt_{1/2}$ mean differences. The blue dots (representing TFs that affect TR more significantly than $TRt_{1/2}$) were chosen with a TR FDR < 0.06 and $TRt_{1/2}$ FDR > 0.1. Red dots (representing TFs that are significant in permutation tests for both TR and $TRt_{1/2}$) were chosen with TR FDR < 0.1 and $TRt_{1/2}$ FDR < 0.1. Finally, black dots (representing TFs that potentially affect $TRt_{1/2}$ more than TR) were chosen with TR FDR > 0.1 and $TRt_{1/2}$ FDR < 0.1, or TR FDR > 0.45 and $TRt_{1/2}$ FDR < 0.3.

Figure Legends

Figure 1 | Estimation of TBP residence time from model fit of normalized, scaled competition ChIP data. a, Ratio of concentration of competitor TBP (denoted by B) to the concentration of endogenous TBP (denoted by A) as a function of induction time measured by Western blotting[9] along with a sigmoid fit to the data. The fit gives an induction time t_0^{ind} of 22 minutes and a saturation concentration ratio of 2.233 as the induction time goes to infinity. **b,** Plot of normalized, scaled competition ChIP ratio data (competitor/endogenous) versus induction time, stratified and averaged in bands of 2 minutes for t_0, an estimate of the overall turnover response time, ranging from 24.5 minutes to greater than 40 minutes, along with induction data (dashed). The average data shows a progressively slower rise as t_0 increases. **c,** Normalized density of residence times, $t_{1/2}$, obtained from data within each t_0 band (same color scheme as panel (**b**)) showing that larger t_0 leads to longer residence times. **d,** $\log_2$-$\log_2$ plot of $t_{1/2}$ versus t_0 showing a monotonic relationship between $t_{1/2}$ and t_0 for $t_0 > 24.5$ min. For loci with $t_0 < 24.5$ min, the noise in the data and the induction curve makes $t_{1/2}$ estimates imprecise. As a consequence, estimates of residence times shorter than ~ 90 seconds are unreliable. **e – g,** Representative fits of our chemical kinetic theory based model to the normalized, scaled competition ChIP ratio data and estimates of $t_{1/2}$, along with the Hill-like sigmoid function fit to the induction data (dashed) with $t_0^{ind} = 22$ min. The colors of the data and the fits correspond to the appropriate t_0 band shown in (**b**).

Figure 2 | Multiple, minute-scale TBP-chromatin binding events required for transcription at high expressed Pol II genes. a, Normalized density of $t_{1/2}$ for Pol II and Pol II genes showing that median Pol II residence time is 3 minutes, while the median for Pol III genes is 9 minutes. The difference between the two distributions is significant with a Kolmogorov-Smirnoff (KS) p-value = 2.2e-16. **b,** Normalized $t_{1/2}$ density for TATA-containing versus TATA-less promoters. TATA-containing promoters have over all shorter residence times than TATA-less promoters (KS p-value = 0.0075). **c,** RP genes have marginally shorter residence times compared to non-RP genes (median RP $t_{1/2} = 1.4$ min and median non-RP $t_{1/2} = 1.6$ min; KS p-value=0.25). **d,** Promoters in the highest quartile of TR tend to have shorter $t_{1/2}$ than promoters in the lowest quartile (KS p-value = 0.005). **e,** Promoters with higher extrinsic transcriptional noise have lower residence time (KS p-value = 0.048). **f,** Normalized density of transcription efficiency (defined as the transcription rate multiplied by residence time, $TRt_{1/2}$) on the

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

$\log_2$ scale showing that for a typical Pol II promoter, ~5 TBP turnovers are required before a single molecule of RNA is successfully transcribed.

Figure 3 | TBP dynamics are correlated with Rap1 but not +1 nucleosome dynamics. a-c, $\log_2$-$\log_2$ scatterplot of Rap1 relative residence time (λ^{-1}) versus $t_{1/2}$ (**a**), TR (**b**), and $TRt_{1/2}$ (**c**) for non-RP genes (red) and RP genes (blue). Rap1 λ^{-1} is correlated well with $t_{1/2}$ and $TRt_{1/2}$ at non-RP genes, but not at RP genes. λ^{-1} is mildly correlated with TR, while $t_{1/2}$ is mildly anti-correlated with TR. **d-f,** Normalized $t_{1/2}$ (**d**), TR (**e**) and $TRt_{1/2}$ (**f**) densities for genes containing hot and cold +1 nucleosomes. Hot nucleosomes were in the top quartile of nucleosome turnover and cold were in the bottom quartile. There was no difference in the residence times between hot and cold nucleosomes (KS p-value=0.50) (**d**), but hot nucleosomes tended to have higher TR (KS p-value=0.007) (**e**) and higher $TRt_{1/2}$ (KS p-value=1.3e-7) (**f**).

Figure 4 | High numbers of elongation factors at Pol II promoters are associated with higher transcription rates and efficiencies. a-c, Normalized density of $t_{1/2}$ on $\log_2$ scale for genes with the upper quartile numbers of bound TFs and genes with the lower quartile numbers of bound TFs (out of 202 mapped TFs[18]) showing that $t_{1/2}$ is not modulated by the number of overall TFs (**a**), initiation TFs (**b**) or elongation TFs (**c**). **d-f,** Normalized density of TR on the $\log_2$ scale for genes with the upper quartile numbers of bound TFs and genes with the lower quartile numbers of bound TFs showing that TR is modulated by the number of overall TFs (KS p-value=8.6e-5) (**d**), initiation TFs (KS p-value=9.8e-4) (**e**) and elongation TFs (KS p-value=3.14e-11) (**f**). **g-i,** Normalized density of $TRt_{1/2}$ on $\log_2$ scale for genes with the upper quartile numbers of bound TFs and genes with the lower quartile numbers of bound TFs showing that $TRt_{1/2}$ is significantly modulated by the number of overall TFs (KS p-value=2.4e-4) (**g**), initiation TFs (KS p-value=0.05) (**h**) and elongation TFs (KS p-value=8.5e-8) (**i**).

Figure 5 | Presence of TFIIE results in lower TBP residence times. a-c, The presence of Tfa2 yielded lower $t_{1/2}$ (KS p-value=5.4e-3) (**a**) and TR (KS p-value=0.028) (**b**), but $TRt_{1/2}$ was not affected by Tfa2 (**c**). **d,** Significance of the increase or decrease that the presence versus absence of each TF (each dot) of 202 mapped TFs[18] had on $TRt_{1/2}$ (y-axis) and/or TR (x-axis) quantified by performing a permutation t-test and applying a false discovery rate (FDR) correction. Low FDRs were used to identify TFs that are likely to affect the transcription rate only (blue dots), the transcription efficiency only (black dots) or both (red dots). **e,** Ranked list of TFs (categorized according to access, orchestration, initiation and elongation[18]) whose presence compared to absence at genes yielded the most significant higher ("Fastest" in table) or lower ("Slowest" in table) TR (blue), $TRt_{1/2}$ (black), or both TR and $TRt_{1/2}$ (red). **f,** Select multi-protein complexes from the list in (**e**) highlighting the role of each complex in increasing ("Fast Subunits" in table) or decreasing ("Slow Subunits" in table) TR and/or $TRt_{1/2}$ using the same color scheme as (**d**) and (**e**).

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

References

1 Hager, G. L., McNally, J. G. & Mistelli, T. Transcription Dynamics. *Mol. Cell* **35**, 741-753 (2009).

2 Boettiger, A. N., Ralph, P. L. & Evans, S. N. Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS Comput Biol* **7**, e1001136, doi:10.1371/journal.pcbi.1001136 (2011).

3 Boettiger, A. N. Analytic approaches to stochastic gene expression in multicellular systems. *Biophys J* **105**, 2629-2640, doi:10.1016/j.bpj.2013.10.033 (2013).

4 Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science* **332**, 475-478 (2011).

5 Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* **332**, 472-474 (2011).

6 Karpova, T. S. *et al.* Concurrent Fast and Slow Cycling of a Transcriptional Activator at an Endogenous Promoter. *Science* **319**, 466-469 (2008).

7 Poorey, K. *et al.* Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342**, 369-372, doi:10.1126/science.1242369 (2013).

8 Sprouse, R. O. *et al.* Regulation of TATA binding protein dynamics in living yeast cells. *PNAS* **105**, 13304-13308 (2008).

9 van Werven, F. J., van Teeffelen, H. A., Holstege, F. C. & Timmers, H. T. Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. *Nat Struct Mol Biol* **16**, 1043-1048, doi:10.1038/nsmb.1674 (2009).

10 Dion, M. F. *et al.* Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**, 1405-1408, doi:10.1126/science.1134053 (2007).

11 Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251-255, doi:10.1038/nature10985 (2012).

12 Pelechano, V., Chavez, S. & Perez-Ortin, J. E. A complete set of nascent transcription rates for yeast genes. *PLoS One* **5**, e15442, doi:10.1371/journal.pone.0015442 (2010).

13 Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699-709 (2004).

14 Moqtaderi, Z. & Struhl, K. Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. *Mol Cell Biol* **24**, 4118-4127 (2004).

15 Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. *Mol Cell* **45**, 483-493, doi:10.1016/j.molcel.2011.11.035 (2012).

16 Ravarani, C. N., Chalancon, G., Breker, M., de Groot, N. S. & Babu, M. M. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat Commun* **7**, 10417, doi:10.1038/ncomms10417 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/045161; this version posted March 22, 2016. 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 4.0 International license.

17 Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225-229, doi:10.1038/35041603 (2000).

18 Venters, B. J. *et al.* A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Mol Cell* **41**, 480-492, doi:10.1016/j.molcel.2011.01.015 (2011).

19 Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375 (2003).

Fig. 1

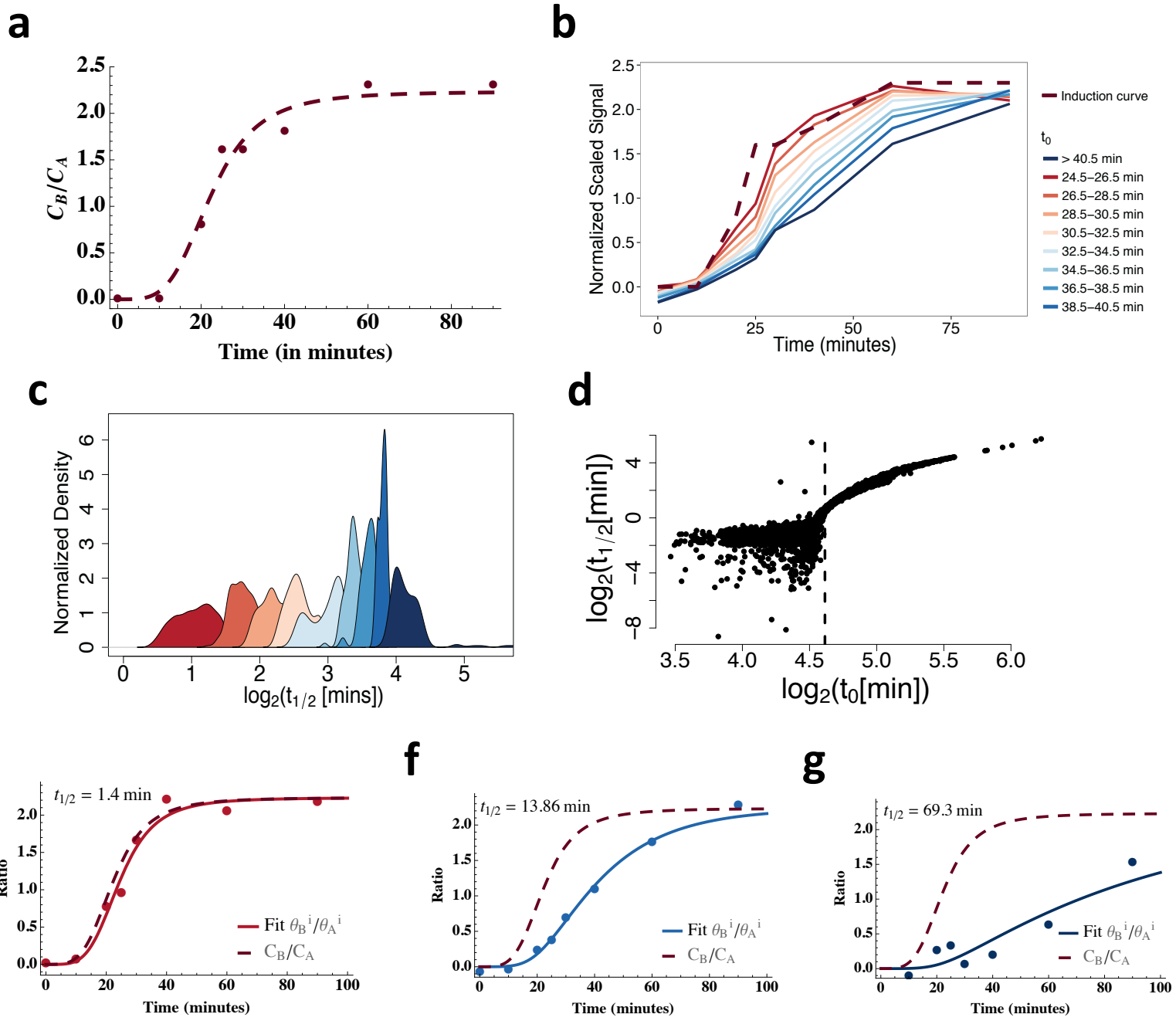

a
C_B/C_A
Time (in minutes)
b
Normalized Scaled Signal
Time (minutes)
Induction curve
t_0
> 40.5 min
24.5–26.5 min
26.5–28.5 min
28.5–30.5 min
30.5–32.5 min
32.5–34.5 min
34.5–36.5 min
36.5–38.5 min
38.5–40.5 min
c
Normalized Density
$\log_2(t_{1/2}$ [mins])
d
$\log_2(t_{1/2}$[min])
$\log_2(t_0$[min])
e
$t_{1/2}$ = 1.4 min
Ratio
Fit $\theta_B{}^i/\theta_A{}^i$
C_B/C_A
Time (minutes)
f
$t_{1/2}$ = 13.86 min
Ratio
Fit $\theta_B{}^i/\theta_A{}^i$
C_B/C_A
Time (minutes)
g
$t_{1/2}$ = 69.3 min
Ratio
Fit $\theta_B{}^i/\theta_A{}^i$
C_B/C_A
Time (minutes)

Fig. 2

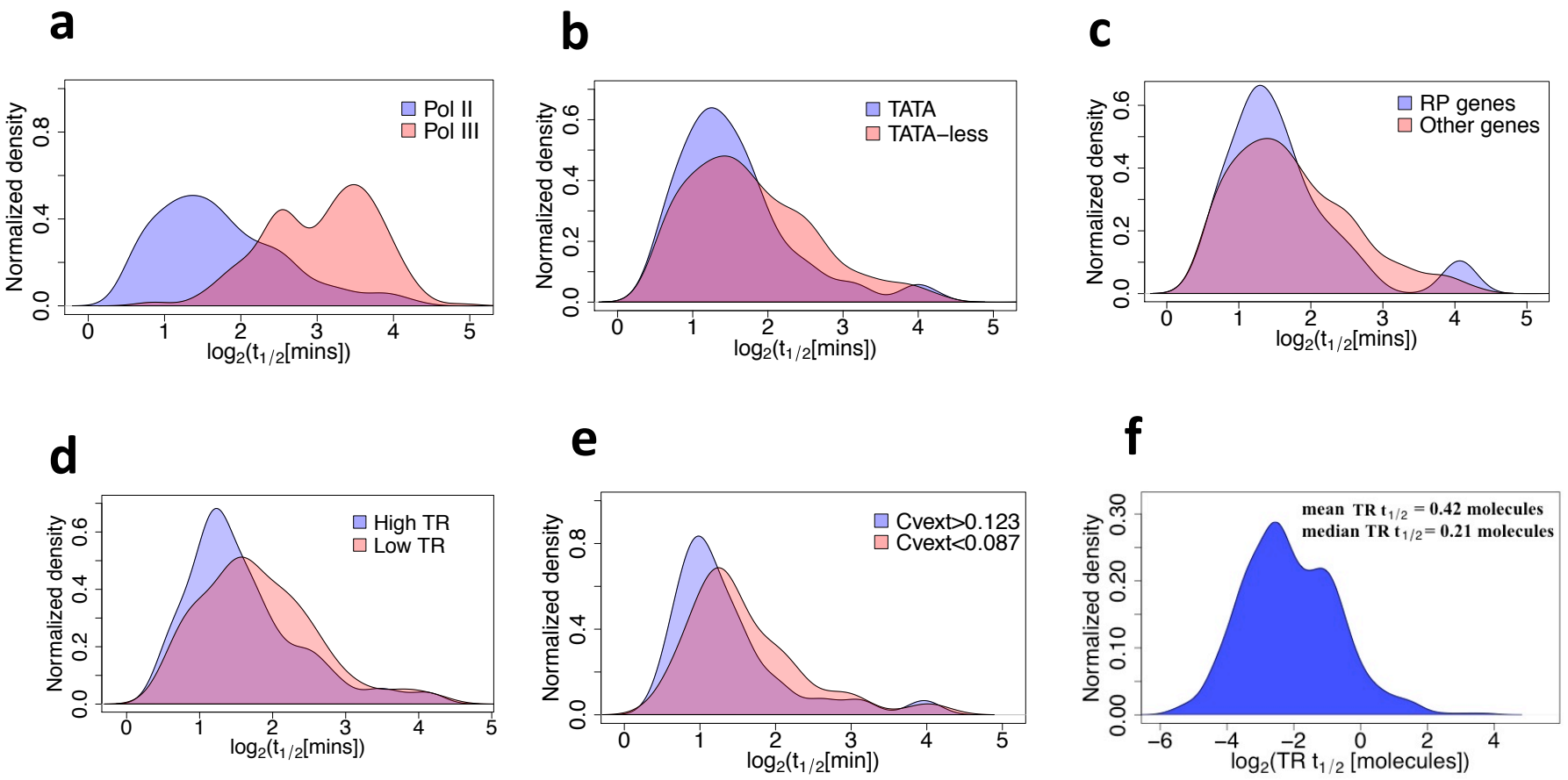

a
Normalized density
Pol II
Pol III
log2(t1/2[mins])
b
Normalized density
TATA
TATA–less
log2(t1/2[mins])
c
Normalized density
RP genes
Other genes
log2(t1/2[mins])
d
Normalized density
High TR
Low TR
log2(t1/2[mins])
e
Normalized density
Cvext>0.123
Cvext<0.087
log2(t1/2[min])
f
Normalized density
mean TR t1/2 = 0.42 molecules
median TR t1/2 = 0.21 molecules
log2(TR t1/2 [molecules])

Fig. 3

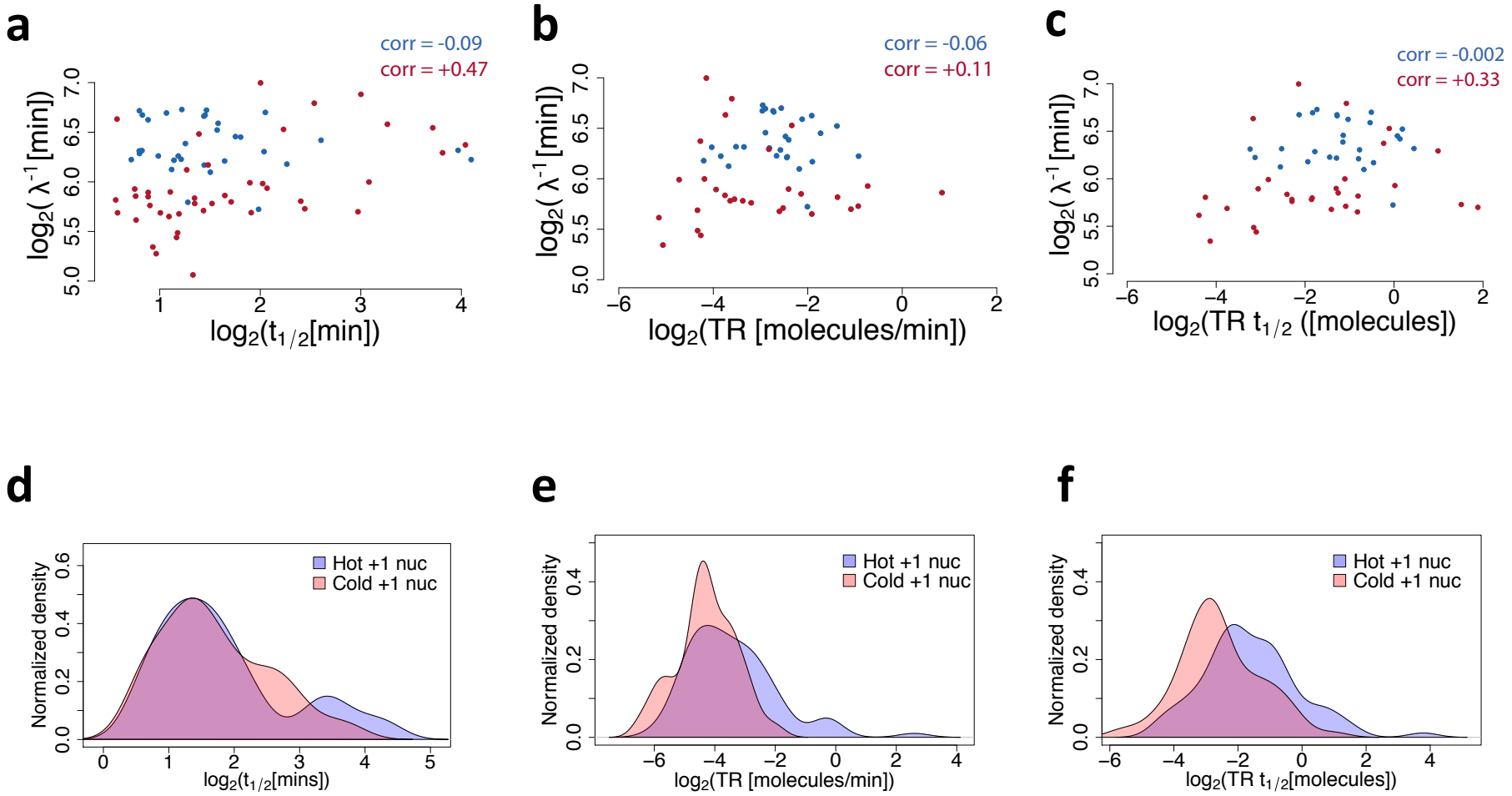

a
corr = -0.09
corr = +0.47
log2(λ-1 [min])
log2(t1/2[min])
b
corr = -0.06
corr = +0.11
log2(λ-1 [min])
log2(TR [molecules/min])
c
corr = -0.002
corr = +0.33
log2(λ-1 [min])
log2(TR t1/2 ([molecules])
d
Hot +1 nuc
Cold +1 nuc
Normalized density
log2(t1/2[mins])
e
Hot +1 nuc
Cold +1 nuc
Normalized density
log2(TR [molecules/min])
f
Hot +1 nuc
Cold +1 nuc
Normalized density
log2(TR t1/2[molecules])

Fig. 4

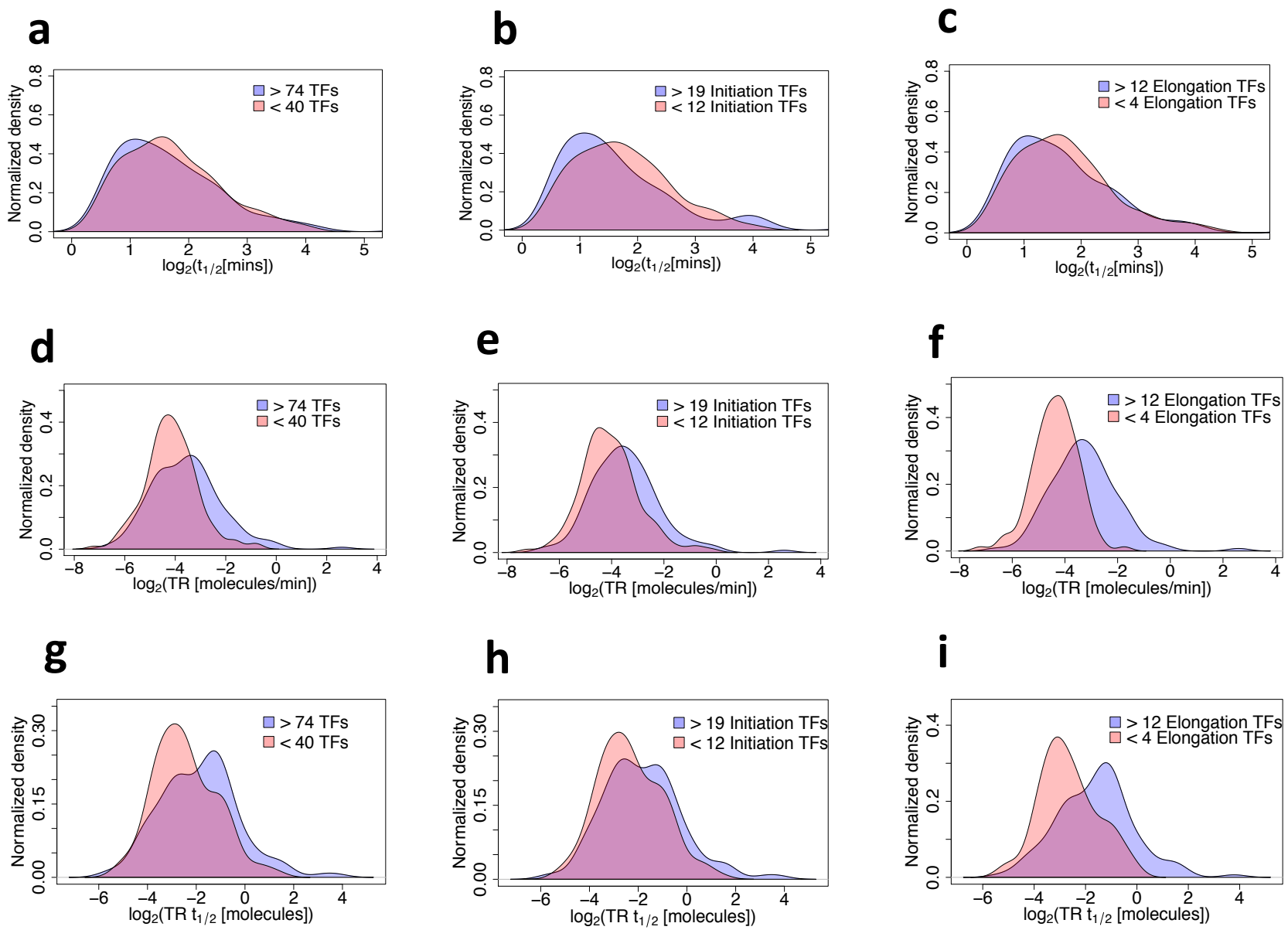
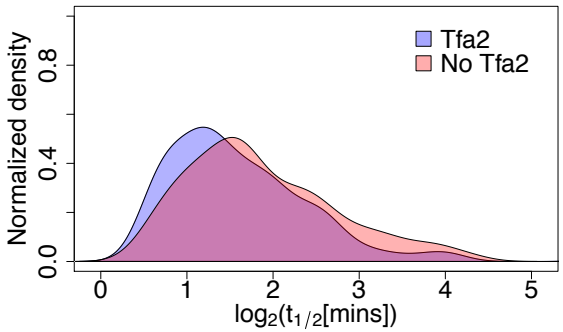
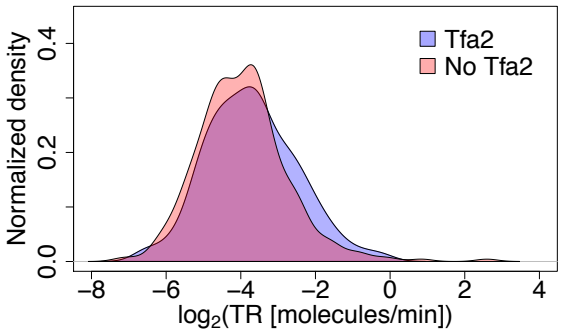

a
b
c
d
e
f
g
h
i
> 74 TFs
< 40 TFs
> 19 Initiation TFs
< 12 Initiation TFs
> 12 Elongation TFs
< 4 Elongation TFs
Normalized density
log2(t1/2[mins])
log2(TR [molecules/min])
log2(TR t1/2 [molecules])

Fig. 5

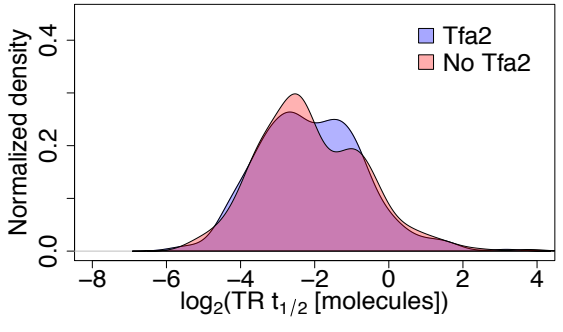
a



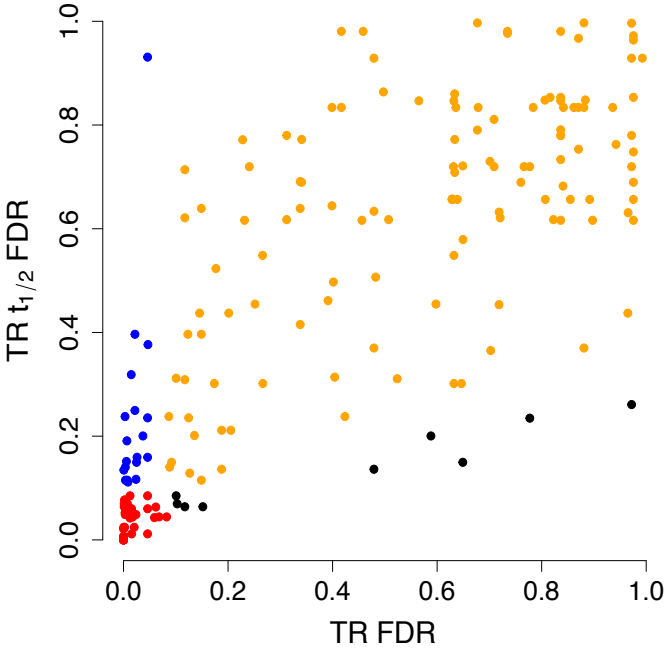
b



c



d



e

Access	Orchestration	Initiation	Elongation
Fastest: Chd1 Isw1 Ycs4 Esa1 Nhp6a Epl1 Sir2 Spt2 Bre1 Set1 Tup1 Ahc1 Bre2 Eaf3 Htb1	Fastest: Cin5 Pho2 Ume6 Sfp1 Ifh1 Rap1 Dig1 Spt23 Msn2 Skn7	Fastest: Spt7, Bur6 Tfa1, Ssl2 Kin28, Taf14 Med4, Tfa2 Tbp, Ncb2 Hfi1, Tfb1 Taf3, Tfb3 Taf13, Tfg1 Spt3, Tfc6 Taf4 Srb4	Fastest: Dst1 Paf1 Ctr9 Spt16, Rpo21 Spt6, Rpb3 Rpb2, Iws1 Rpb7 Ess1 Pob3 Rtt103 Pcf11 Rlr1 Rna14 Thp1
Slowest: Swc1 Htz1 Vps72 Ioc3 Irc20	Slowest: Xpb1 Fkh2	Slowest: Rgr1	

f

Complex	Fast Subunits	Slow subunits
nuA4	Esa1, Epl1, Eaf3	
Tho	Rlr1, Thp1	
SAGA/SLIK	Spt7, Chd1, Hfi1, Spt3	
NC2	Bur6, Ncb2	
Nucleosome	Htb1	Htz1
TFIIE	Tfa1, Tfa2	
TFIIH	Ssl2, Kin28, Tfb1, Tfb3	
Mediator	Med4, Srb4	Rgr1
FACT	Spt16, Pob3	
SwrC		Swc1, Vps72
Spt6	Spt6, Iws1	
PolII	Rpb2, Rpo21, Rpb3	
TFIID	Taf14, Taf3, Taf13, Taf4	
Compass	Set1, Bre2	