Molecular Cell, Volume *67*

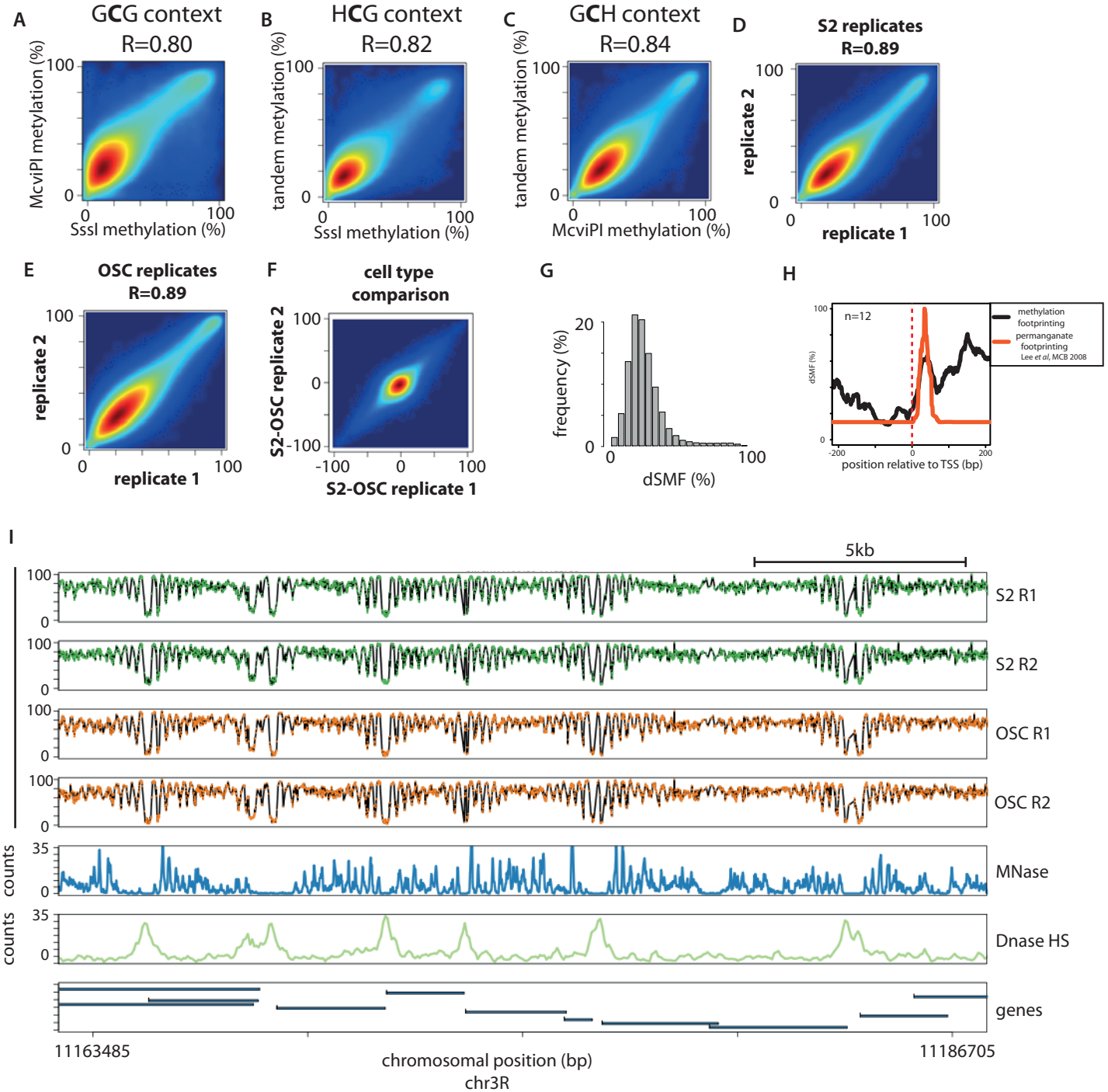
Supplemental Information

Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters

Arnaud R. Krebs, Dilek Imanci, Leslie Hoerner, Dimos Gaidatzis, Lukas Burger, and Dirk Schübeler

Supplementary Figure 1: Highly reproducible DNA footprinting using two recombinant methyl-transferases. Related to Figure 1.

A, M.CviPI - GpC methyltransferase and M.SssI - CpG methyltransferase create similar footprinting patterns. Smoothed scatter plot representing average methylation observed at GCG context genome-wide in S2 cells that were incubated with either M.CviPI or M.SssI. B, CpG methylation in tandem treated samples is comparable to CpG methylation in cells treated with M.SssI only. Smoothed scatter plot representing the average genome wide methylation in CpG context for the two S2 cells samples. C, GpC methylation in tandem treated samples is comparable to GpC methylation in cells treated with M.CviPI only. Smoothed scatter plot representing the average genome wide methylation in GpC context for the two S2 cells samples. D, Tandem treated biological replicates are highly correlated. Smoothed scatter plot representing the average genome wide methylation in GpC or CpG context for the two biological replicates in S2 cells. E, Tandem treated biological replicates are highly correlated. Smoothed scatter plot representing the average genome wide methylation in GpC or CpG context for the two biological replicates in OSC cells. F, Methylation footprinting detects variation of accessibility between cell lines. Smoothed scatter plot representing the average methylation difference for S2 versus OSC cells compared between biological replicates. G, Most of the genome is protected from methylation. Histogram depicting the distribution of average methylation levels genome-wide in S2 cells. H, The non-nucleosomal footprint observed downstream of TSSs aligns with the position of the transcriptional bubble. Composite profile for a subset of promoters profiled by permanganate-footprinting (Lee et al., 2008). Displayed is the overlay of methylation footprints with permanganate footprinting signal as smoothed average (lines). I, Dual Enzyme Single Molecule Footprinting (dSMF) over a 25 kb chromosomal region illustrating periodicity in the accessibility signal and the increased accessibility observed at regulatory regions. Shown is average methylation of single cytosine from two biological replicates for each cell type (S2 - green dots; OSC- orange dots) and smoothed signal (black line). Read counts are displayed across the region for additional datasets as smoothed signal (colored lines – dataset as indicated).



Krebs *et al*, Figure S1

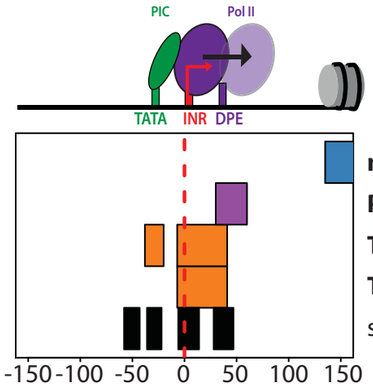
Supplementary Figure 2: A strategy for single molecule quantification of binding at promoters. Related to Figure 2.

A, Schematic representation summarizing the positions observed to be occupied by the PIC, Pol II and nucleosomes at TSSs. B, Schematic representation of the relative position of the four bins used for collecting footprint data at the single molecule level. Color bars representing the position of each collection bin were overlaid on the average footprint of the example gene used in Figure 2A. Bins were positioned to capture the footprints of a promoter bound nucleosome (grey), the PIC (green), the initiating Pol II (red), and the engaged Pol II (purple). C, Schematic representation of the footprint patterns used to define each state (methylated Cs – accessible – light grey, unmethylated Cs - protected - black). States were named according to the complexes expected to bind at these positions. D, E, Comparison of an example footprint pattern organized using hierarchical clustering (D) or the single molecule classification method (E). Data are compared here for the gene *CG45486* (as in Figure 2A) to illustrate that patterns observed with an unsupervised clustering method are captured by the single molecule classification method. F, Boxplots depicting the frequency of each footprint pattern across promoters genome-wide. Boxplots were organized and colored according to their usage in the simplified 5 states classification (as in panel C). Sub-states from the unassigned category (grey) tend to have low frequency. G, Global relation between the frequencies of the 16 states at promoters and independent bulk measurements of Pol II and nucleosomes depicted as a heatmap (Pearson correlations). 6 states showed no or low correlation with the tested biological features and were classified as 'unassigned'. The patterns showing low accessibility cluster together with MNase-seq and were grouped in a single 'nucleosome' state. H, Relative abundance of states across promoters genome wide. Histogram depicting the distribution of frequencies observed for each state across all promoters. The nucleosome occupied state shows a bimodal distribution with promoters having either low (~10%) or high (~80%) nucleosome occupancy. The abundance of accessible molecules varies broadly, PIC footprints occur generally at low frequency, with the exception of a defined small subset of promoters. The Pol II state shows a broad distribution at a frequency up to ~65% of occupied

molecules at some TSSs. The unassigned state shows a normal distribution centered around ~15%, indicating that most (> 85%) of the molecules are classified into one of the 5 studied states. I, Abundance of states measured by single molecule footprinting is highly correlated between biological replicates. Heatmap depicting the pearson correlation coefficient calculated by comparing promoter state abundances genome wide within replicates and between cell types. Correlation in state abundance is R>0.85 for all states between biological replicates. J, Difference in state abundance between S2 and OSC cells is correlated to changes measured by independent bulk methods. Heatmap representing pearson correlation coefficients of the differences observed across promoters between cell types. Changes in Pol II or PIC abundance as measured by single molecule quantification correlate with accessibility (DHS) and expression changes (RNA-seq). In contrast, changes in nucleosome occupancy as determined by single molecule quantification are anti-correlated to gene activity changes. K, Validation of the TBP dsRNA knock down. Relative TBP mRNA abundance after 72h incubation with dsRNA as measured by RT-qPCR (fold change calculated using the comparative Ct method). Error bars represent the variation across two biological replicates. L, Validation of α-amanitin mediated Pol II degradation. Western blot against Pol II and TBP after incubation of S2 cells with 20µg/mL α-amanitin for 24h. M, TBP knock-down leads to a reduction of the frequency of PIC states (states 3 and 4). Boxplots depicting the relative state frequencies after TBP knock down for a focused set of TATA-containing promoters (See Figure 5 for details). Shown are average frequencies over two biological replicates. N, Pol II degradation leads to a reduction in the frequency of Pol II states (states 4 and 5). Boxplots depicting the relative state frequencies after Pol II α-amanitin mediated Pol II degradation. The TBP specific sate frequency remains unaffected (state 3). Shown are average frequencies over two biological replicates. O, P, The simplified 5 state classification accurately describe variations in levels of PIC and Pol II. Same boxplot representation as in M and N using the 16 states classification. Q-S, Effects of TBP and Pol II depletion on the frequencies of footprints at the Fur1 gene. Stack of individual molecule reads illustrating the discrete distribution of accessible (methylated Cs – light grey) and protected regions (unmethylated Cs - black). TBP knock down leads to a

reduction of the frequency of footprints upstream of the TSS from 42% (Q) to 9% (R). Pol II degradation leads to a reduction of the frequency of the footprint downstream of the TSS from 48% to 21% without affecting frequency of the upstream footprint (S). T, State redistribution at an ecdysone response gene (*Hr46*) upon hormonal induction. Barplot depicting the cumulative distribution of state frequencies before and after ecdysone induction (time as indicated). Also shown is expression of *Hr46* during the timecourse (RNAseq - RPKM)(upper panel – white to black scale). Induction causes increase of Pol II occupied molecules at the cost of unbound molecules.

A



nucleosome (Marvitch *et al* 2008)

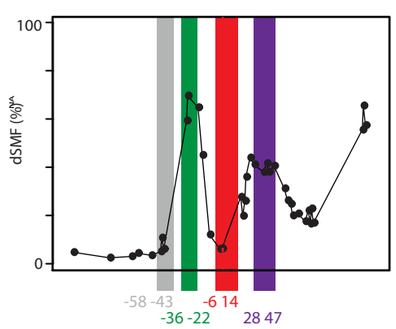
Pol II (Nechaev, *et al* 2010, Kwak *et al*, 2013)

TFIID+TFIIA (Cianfrocco *et al*, 2013, Louder *et al* 2016)

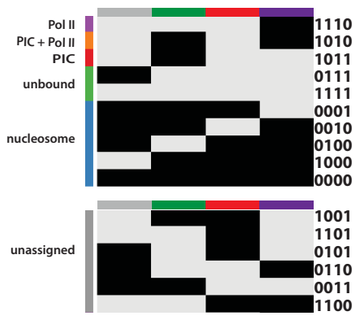
TFIID (Cianfrocco *et al*, 2013, Louder *et al* 2016)

single molecule classification bins

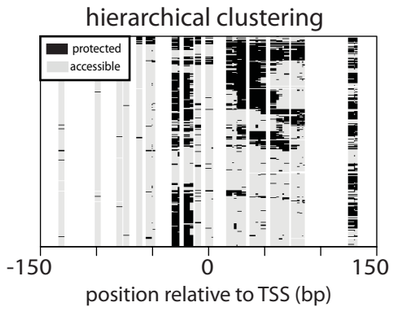
B



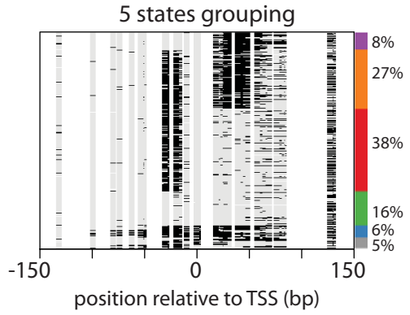
C



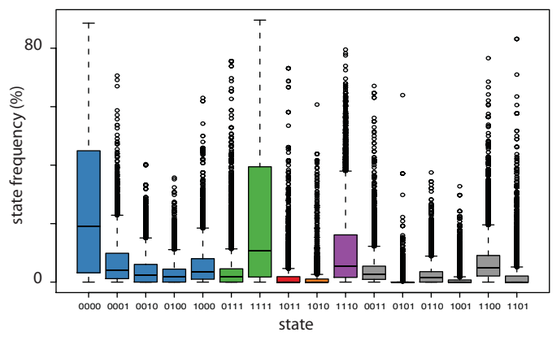
D



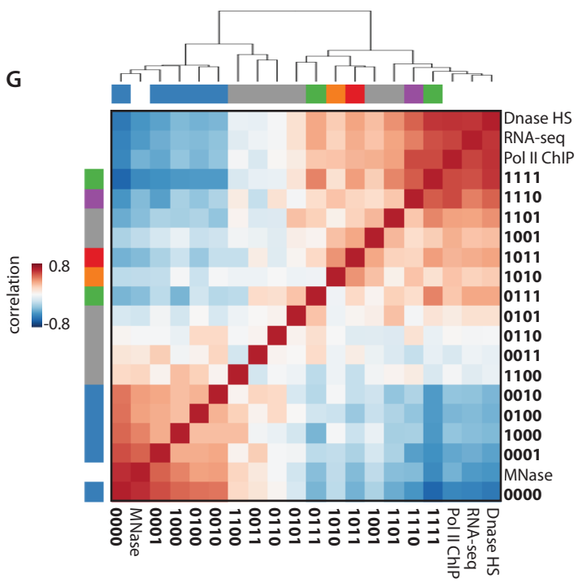
E



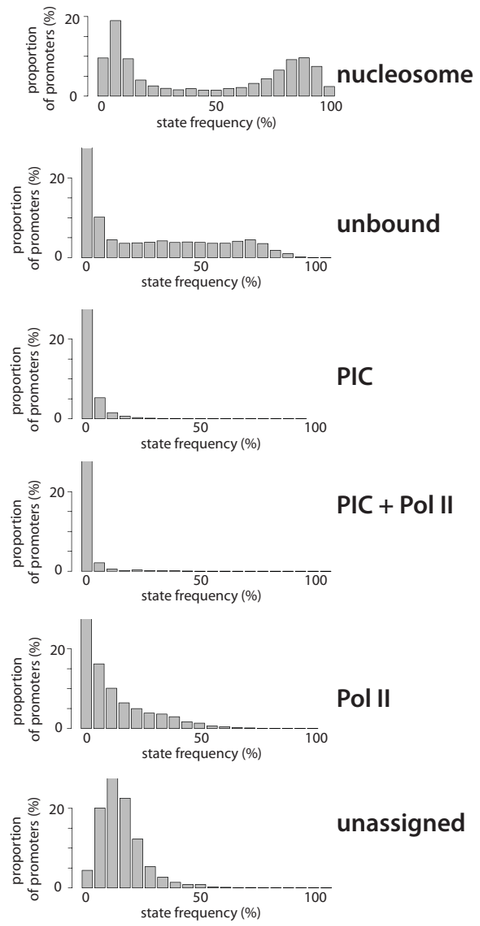
F



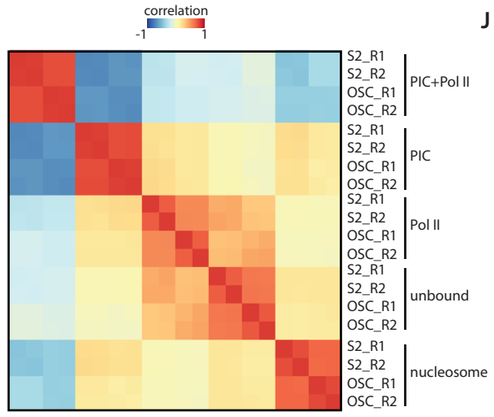
G



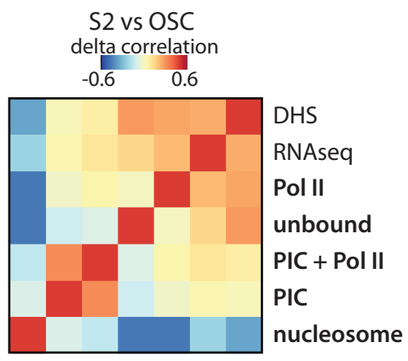
H



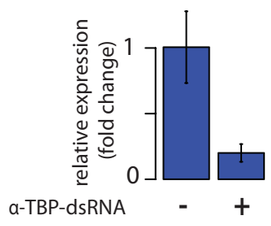
I



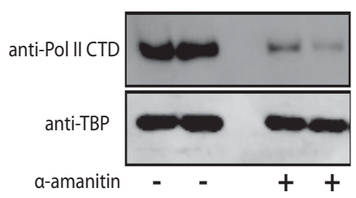
J



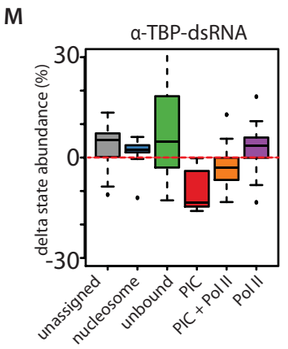
K



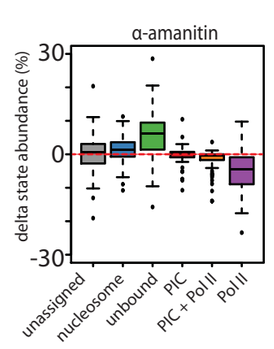
L



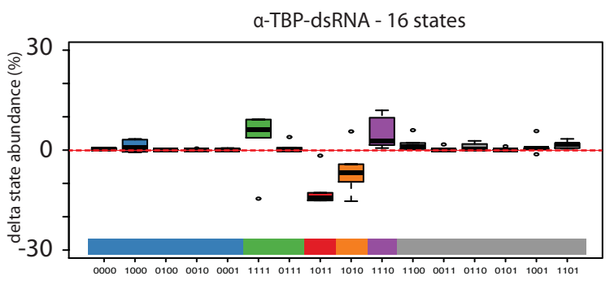
M



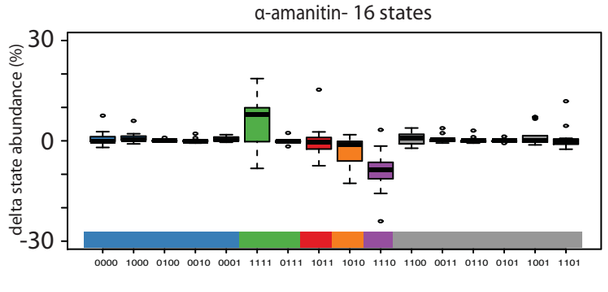
N



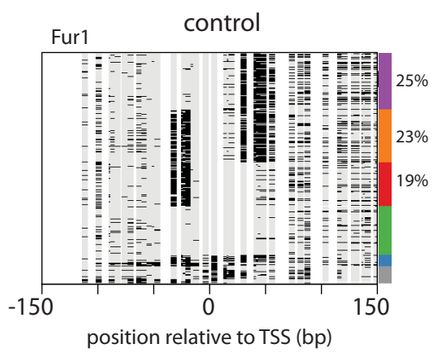
O



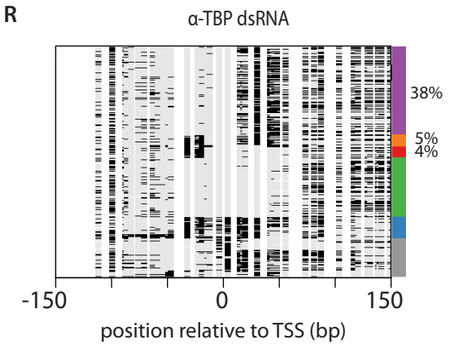
P



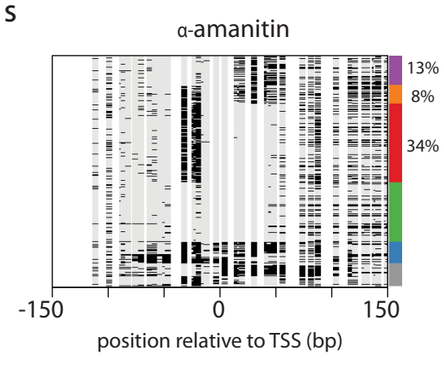
Q



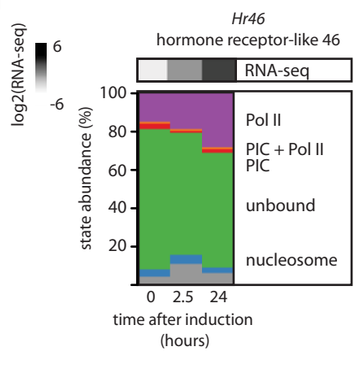
R



S



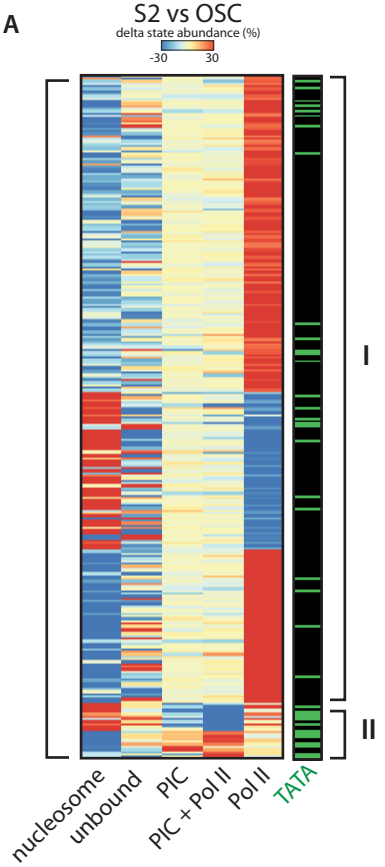
T

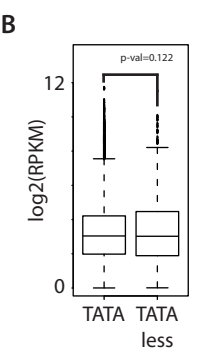


Krebs *et al*, Figure S2

Supplementary Figure 3: Effects of promoter structure on footprinting patterns. Related to Figure 3.

A, Core promoter sequence elements determine changes of state distribution between cell types. Heatmap showing differences in state abundances between S2 and OSC cells for the 10% most changing promoters. Side-bar displays presence (present: color line, absent: black line) of core promoter elements at the promoters. Dynamics of Pol II is observed at all promoters while the PIC dynamics is most evident at TATA containing promoters. Promoters were organized by hierarchical clustering and arbitrarily grouped in two categories (I-II). B, Distribution of expression levels is comparable for TATA-containing and TATA-less promoters. Comparative boxplot showing RNA-seq levels of genes with or without a TATA box. p-value results from a bi-directional t-test illustrating the similarity of the distributions.

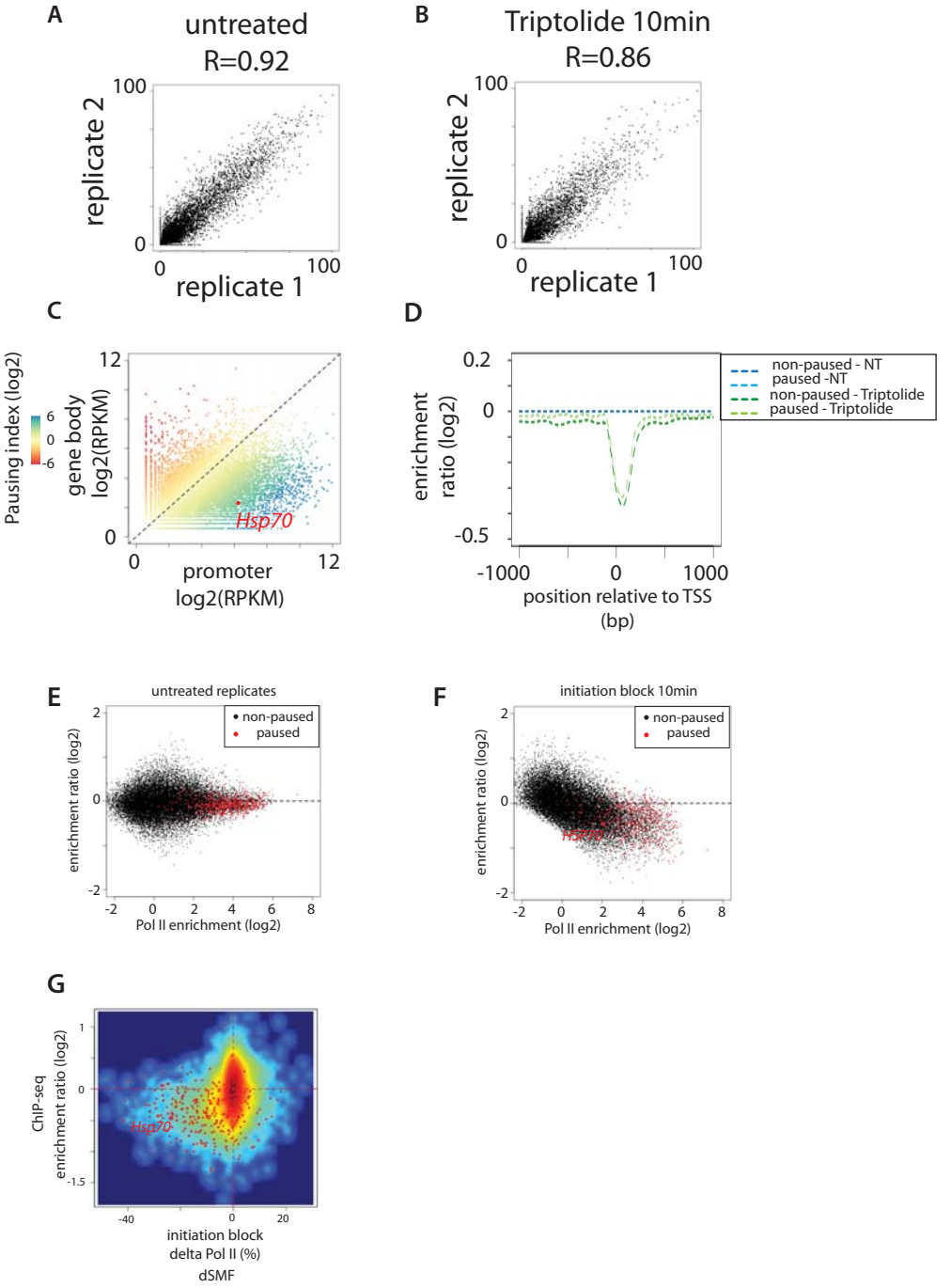




Krebs *et al*, Figure S3

Supplementary Figure 4: Polymerase turnover at promoters genome-wide. Related to Figure 4.

A, B, Frequency of Pol II footprints is highly correlated between biological replicates. Scatter plot representing the frequency of Pol II footprints (state4+5) at promoters genome-wide in untreated S2 cells (A), or cells incubated 10 minutes with Triptolide (B). C, Calculation of the pausing index using GRO-seq data. Scatter plot comparing GRO-seq signal at the promoter and in the body of drosophila genes. The color code depicts the 'pausing index' defined as the GRO-seq signal ratio between promoters and gene body. The red dot depicts the position of the *Hsp70* gene in the plot. Paused genes were defined as active genes having 8 times more signal at promoter compared to gene body (log2 (pausing index) >3) and low read counts in their gene body (log2(RPKM)<3). D, Pol II turnover is high at paused genes. Composite profiles of Pol II enrichment around TSSs in untreated cells and cells incubated with Triptolide for 10 minutes as measured by ChIP-seq. Enrichment fold change (log2) upon Triptolide treatment for paused (green dotted line) and non-paused genes (blue dotted line). E, F, Inhibition of transcription initiation leads to global changes in Pol II occupancy genome-wide. (E) Scatterplot displaying the fold change between two untreated biological replicates as a function of Pol II enrichment. Gene categories are color-coded (paused-red, non-paused-black). (F) Global changes in Pol II occupancy are observed upon inhibition of initiation. Scatterplot displaying the fold change in Pol II occupancy upon Triptolide treatment as a function of Pol II enrichment in the untreated conditions. Gene categories are color-coded (paused-red, non-paused-black). G, ChIP-seq and dSMF identify comparable changes in Pol II occupancy upon inhibition of transcription initiation for 10 minutes. Smoothed scatter plot showing the differences in frequency of Pol II occupancy (state 4+5) measured by dSMF compared to the fold change in occupancy measured by ChIP-seq. Paused genes are highlighted using red dots. The position of the *Hsp70* gene is depicted in the plot.

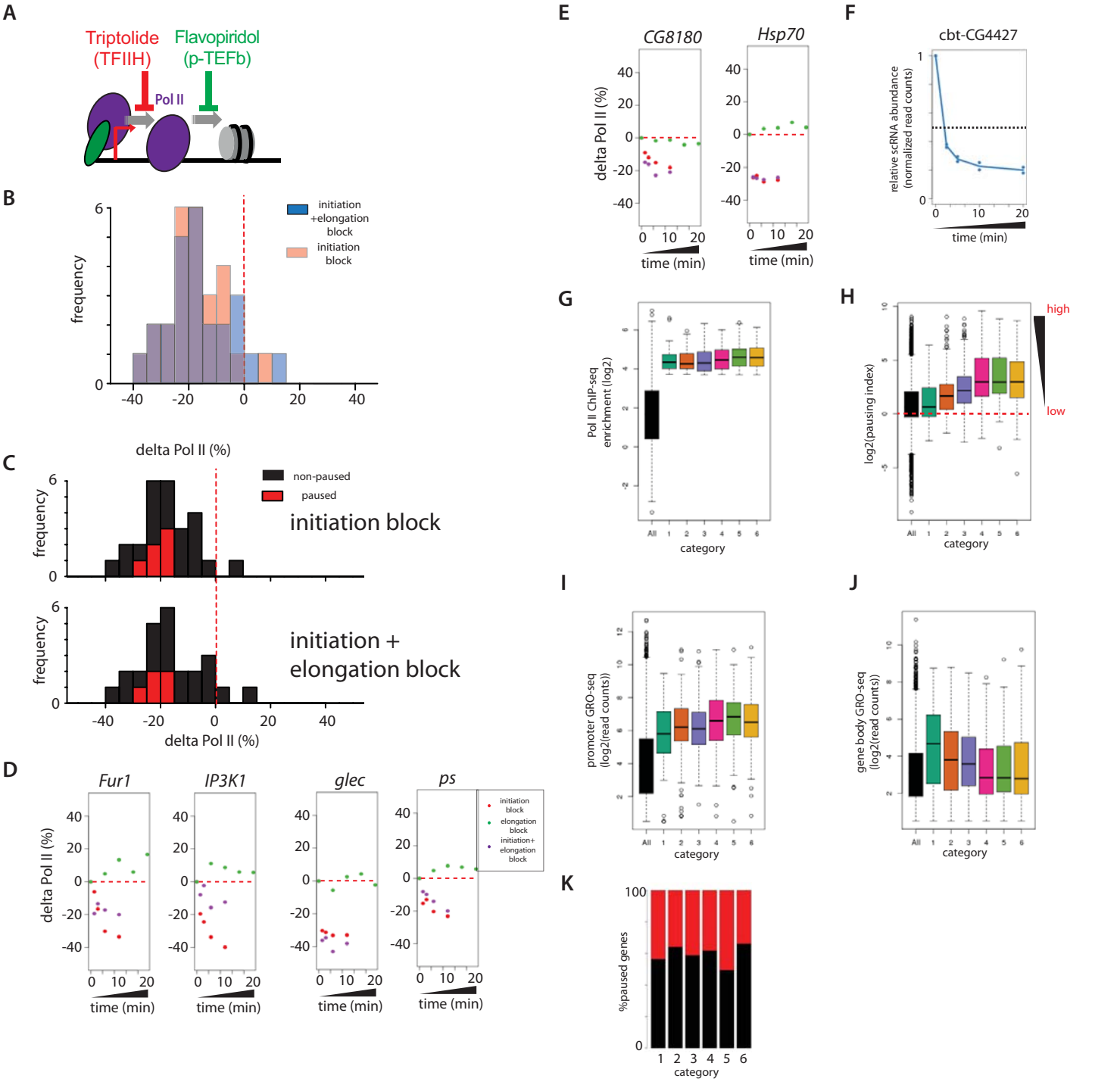


Krebs *et al*, Figure S4

Supplementary Figure 5: Rapid Pol II turnover at the promoter of paused genes. Related to Figure 5.

A, Schematic representation of discrete steps in the initiation process and their inhibition by small molecules. Changes in frequency of Pol II bound molecules are measured after blocking either entry to initiation (Triptolide - red bar) or elongation (Flavopiridol - green bar) or both simultaneously. B, Elongation block partially rescues the loss of engaged Pol II observed upon inhibition of initiation. Overlaid histograms representing the distribution of relative changes in frequency of Pol II occupancy in cells treated either with the inhibitor of initiation alone (light red) or in combination with the inhibitor of elongation (light blue). C, Loss of Pol II at paused genes is independent of elongation. Histogram describing the distribution of relative changes in frequency of Pol II occupancy at different promoter categories (non-paused – black; paused - red) upon incubation with different combinations of inhibitors (10 min). D, E, Time course of the changes in Pol II binding frequency upon inhibition of initiation or elongation. Changes are shown for example genes categorized as non-paused (D), or paused (E). Colored dots represent the amount of engaged Pol II for each time point after various treatment (red – initiation; green – elongation; purple – initiation and elongation). Average over two independent biological measurements is plotted for each time point. F, Comparison of the scRNA decay kinetics at the cbt gene shows similar kinetics when measured genome-wide (F - this study) or on single promoter example (See Figure 4A from (Henriques et al., 2013)).G, Differences in scRNA stability cannot be explained by differential Pol II occupancy. Distribution of Pol II occupancy as measured by ChIP-seq for each group of scRNA kinetics . H, A majority of paused genes show rapid scRNA decay upon inhibition of initiation. Boxplot depicting the distribution of Pol II pausing indexes in each group of scRNA kinetics. Groups with fast scRNA decay (<5min; groups 4-6), contain the majority of genes with high pausing index. I-J, Boxplot depicting the distribution of GRO-seq signal in promoters (I), or gene bodies (J) in each group of scRNA kinetics. Groups with fast scRNA decay (<5min; groups 4-6), contain the majority of genes with low GRO-seq signal in the gene bodies. K, None of the kinetics categories show enrichment for paused genes. Barplot depicting the proportion of paused genes in each group of scRNA kinetics (black:

unpaused, red: paused). P-value of a squared-chi test was used to test the distribution within categories (p-value= 0.0134).



Krebs *et al*, Figure S5