Supplemental Information

Table of Contents: **Page number**

Figure S1. NELF-mediated Pausing is a Widespread Phenomenon. Related to Figure 1.

(A) Specificity of the NELF-B antibody is shown by the decrease in signal observed in western (left panel) and ChIP experiments (right panel) when cells are partially depleted of NELF-B by RNAi. The graph depicts percent input obtained from ChIP experiments with mock-treated or NELF-depleted cells; x-axis coordinates designate the center position of primer pairs with respect to the TSS for each gene.

(B) The average total Pol II signal around each of the 17,116 unique promoters (+/- 250 bp) corresponds very well with the promoter-proximal DSIF signal at these promoters.

(C) NELF is enriched at genes with the highest levels of active transcription. Pol II-bound genes (n=7,466) were divided into quartiles based on the Ser2-P Pol II enrichment within the gene (+500 to +1500) and the levels of NELF enrichment (average signal +/- 250 bp) in each quartile are shown. The top quartile shows significantly higher NELF levels than all other quartiles ($P<0.001$, Kruskal-Wallis test), but quartiles 2-4 are not significantly different. Boxes show 25-75th percentiles, whiskers denote 10-90th.

(D) The short RNAs derived from paused Pol II (Nechaev et al., 2010) were mapped around each Pol II-bound TSS, from +/- 50 bp. A total of 10 reads within this 101 bp window was statistically significant (described in Nechaev et al., 2010), shown as dotted line. Of 7,466 bound genes, 6,357 (85%) had ≥10 short RNAs mapping near their TSS, consistent with the presence of paused Pol II at these genes.

(E) Heatmaps show fold enrichment of total Pol II for control cells that were untreated or treated with dsRNA targeting β-galactosidase (Mock-tr.) compared to cells depleted of NELF (NELF-dep.). Genes are ordered and displayed as in Figure 1A. The change in Pol II signal following NELF RNAi is shown at right as compared to control samples (untreated + mock-RNAi treated). Range is depicted in color bar, where red signifies gain and green indicates loss in signal.

(F) Pol II promoter enrichment is broadly decreased in NELF-depleted cells. Shown are the numbers of genes with indicated levels of Pol II enrichment around promoters (+/- 250 bp) for all 7,466 Pol II-bound genes in Control and NELF-depleted cells. Depletion of NELF shifts the median Pol II enrichment (represented by the peak of the curve) towards the left, indicating generally lower Pol II levels. This shift is significant ($P<0.0001$, Mann-Whitney test), and is in good agreement with our previous work using partial *Drosophila* genome arrays (Muse et al., 2007).

Figure S2. NELF-depletion Affects Highly-Regulated Genes That Can Be Highly Transcribed. Related to Figure 2.

(A) Left panel shows that the average fold enrichment for total Pol II in the promoter-proximal region (TSS +/- 250 bp) in untreated S2 cells corresponds well with the change in Pol II promoter enrichment in cells depleted of NELF (n=7,466 Pol II-bound genes). Right panel shows a good correlation between NELF promoter enrichment vs. the change in Pol II enrichment following NELF RNAi for these same genes.

(B) Left panel shows the Ser2-P Pol II levels within genes (+500 to +1500, levels in untreated S2 cells) for genes in each quartile of Pol II loss upon NELF-depletion (where Quartile 1 are the Most NELF-affected genes, and Quartile 4 are the least), demonstrating that the most NELF-affected genes are more actively transcribed than

genes that are less affected by NELF RNAi. Comparisons between each quartile are statistically significant (P<0.01, Kruskal-Wallis test). Right panel displays the mRNA expression levels for genes in each quartile (Muse et al., 2007), confirming that the Most NELF-affected genes are highly transcribed. P-values for statistical comparisons between quartiles are shown and calculated as in left panel. Boxes show 25-75th percentiles, whiskers show 10-90th percentiles.
(C) GO terms significantly enriched among the most NELF-Affected genes indicate they are frequently regulated through developmental or other signaling pathways. Less NELF-affected genes are more likely to be involved in cellular housekeeping functions. Data represent genes in each group for which a functional annotation existed (1,268 of 1,615 Most NELF-affected; 4,056 of 4,846 Less NELF-affected).

Figure S3. The Sequences of the Most NELF-affected Genes May Induce Pausing and Facilitate Focused Transcription Initiation from Stable Pre-initiation Complexes. Related to Figure 2.
(A) Pause button and DPE are enriched at the most NELF-affected genes. Left panel: frequency of occurrence of the Pause button motif (Hendrix et al., 2008) at each position relative to the observed TSSs for the most NELF-affected genes (502 of 1,615 genes possess a consensus Pause button) and genes that are less NELF-affected (907 of 4,846 genes with Pause button). In both groups of genes, the Pause button motif is most likely to begin at position +26 with respect to the TSS. Center panel: the Pause button shares considerable sequence similarity with the DPE. Shown are the logos for information content for the DPE (Lee et al., 2008) and for the Pause button, derived from motif searches of the most NELF-affected genes as described in Extended experimental procedures. We note that the consensus we arrived at for the Pause button is nearly identical to that reported in (Hendrix et al., 2008). Right panel: overlap of the most NELF-affected genes (n=1615) that possess a Pause button and/or DPE motif using the search parameters described in Extended Experimental Procedures. Although there was considerable overlap between the genes that matched each consensus, there were also many genes whose sequences represented much better matches to one motif than the other.
(B) The most NELF-affected genes display significantly more focused transcription initiation than less NELF-affected genes. Short, capped RNAs (described in Nechaev et al., 2010) were mapped around each Pol II-bound TSS (+/- 50) and the percentage of reads generated from each position in this interval was determined. The location from which the most reads arose was considered as the “observed” TSS and the percentage of reads mapping to this location was reported. This percentage is shown, for genes divided into quartiles based on the level of Pol II loss upon NELF-depletion (where Quartile 1 represents the most NELF-affected genes, and Quartile 4 the least). The most NELF-affected genes showed significantly more reads from the observed TSSs (P<0.001, Kruskal Wallis test), but there were not significant differences between quartiles 2-4. Boxes show 25-75th percentiles, whiskers show 10-90th.
(C) The general transcription factor TFIIA is enriched at the most NELF-affected genes, suggesting that these genes possess stable Pre-initiation complexes. Left panel: genes divided into quartiles based on Pol II loss upon NELF-depletion, were evaluated for their promoter-proximal enrichment of TFIIA signal (+/- 250 bp from TSS). The most NELF-

affected genes showed significantly higher levels of TFIIA occupancy (median enrichment= 1.58), and TFIIA levels decreased with each quartile ($P<0.001$, Kruskal-Wallis test). Boxes show 25-75th percentiles, whiskers show 10-90th percentiles. Right panel: similar analyses performed on genes separated into quartiles based on levels of active elongation (Ser2-P levels from +500 to +1500) also show the expected enrichment of TFIIA at the most highly active genes, although the median enrichment (median=1.46) was lower than at the most NELF-affected genes.

Figure S4. Elongating Pol II Only Modestly Disrupts Promoter-proximal Nucleosomes, and Nucleosome Occupancy is Lowest at Genes with GAGA Binding Sites. Related to Figure 3.

(A) Composite nucleosome profiles are shown for genes bound by Pol II, separated into quartiles based on the amount of Pol II (Ser2-P) within the region +500 to +1500 bp relative to the TSS. The most actively elongated genes (Quartile 1) show little reduction in nucleosome occupancy near the promoter, but some depletion within the gene compared to Quartiles 2 and 3. Interestingly, Quartile 4, which has very low levels of Pol II elongation, displays less well positioned nucleosomes, suggesting that Pol II helps to maintain promoter-proximal nucleosome occupancy and organization.

(B) Comparison of Ser2-P Pol II signal between the most NELF-affected genes (Quartile 1; n=1,866) and the most highly active (Quartile 1 ranked by Ser2-P Pol II signal in the region +500 to +1500; n=1,866).

(C) Nucleosome occupancy profiles are shown for genes bound by Pol II, separated into quartiles based on Pausing Indices. To better compare data with predictions (which represent occupancy over the whole nucleosome), reads were extended from nucleosome centers (shown in Fig. 3E) 73-bp up- and downstream to cover 147-bp regions. The graph depicts, for each quartile, the number of these 147-bp sequences that overlap with the indicated locations.

(D) The presence of core promoter motifs is correlated with lower nucleosome occupancy. Normalized nucleosome occupancy (normalization to allow for comparison between groups with varying gene numbers is described in Extended experimental procedures) is shown for Pol II-occupied genes that generate short RNAs (n=6,461) with: TATA (n=398), Pause button or DPE (n=1,689), Inr (n=1,650) or GAGA (n= 1,032) elements. For comparison, nucleosome distribution is also shown at the average Pol II-bound gene, and at the most NELF-affected genes. Right panel: occupancy of the first (+1) downstream nucleosome (from +50 to +220 bp with respect to TSS) is shown for genes with each motif, relative to genes lacking that motif. P-values refer to comparisons between each indicated pair of gene groups (Mann-Whitney test). Boxes depict 25-75th percentiles and whiskers from 10-90th percent.

(E) Composite nucleosome profiles at the Most NELF-affected genes with TATA, PB/DPE, Inr or GAGA motifs (as in Figure 2B). Greater nucleosome-depletion is observed at NELF-affected genes with GAGA binding sites (n=438), than at genes that lack GAGA motifs (n=1177), but these genes still show far lower nucleosome occupancy than genes with little NELF-mediated pausing. Nucleosome density at the +1 location is shown for genes that possess each motif relative to genes lacking each motif, as in part D.

Figure S5. Antagonistic Relationship Between Pol II and Nucleosome Binding at Highly Paused Genes Revealed by Ecdysone Treatment. Related to Figure 5.
Left panel: CG3835-RA, a promoter with a high pausing index in control cells that becomes unbound by Pol II following 24 hour treatment with ecdysone (top panel). Increased MNase protection following ecdysone treatment indicates increased nucleosome occupancy following Pol II loss (bottom panel). Data points represent average qPCR signal of DNA protected against MNase digestion from two biological replicates at primer pairs centered at the indicated distance from the TSS; error bars depict range.
Right panel: The SP71 promoter is unbound by Pol II in control cells and becomes highly paused following ecdysone treatment (top panel). Promoter nucleosome occupancy decreases in response to ecdysone treatment and Pol II binding (bottom panel).

Figure S6. Paused Polymerase Prevents Promoters from Adopting their Default Chromatin States. Related to Figure 6.
(A) Genes with the most paused Pol II in embryos are predicted to have high promoter nucleosome occupancy, but lower levels within the gene. Predicted nucleosome occupancy (data from models described in Kaplan et al., 2009) is shown for genes that are bound by Pol II in embryos (n=6,778), divided into quartiles based on their Pausing indices (Quartile 1 has the highest Pausing Indices).
(B) Pol II distribution in *Drosophila* embryos vs. S2 cells. Genes in both heatmaps are rank ordered by descending Pausing index in embryos. Genes are divided into quartiles based on their Pausing indices in embryos, as depicted by brackets on right.
(C) Pol II binding displaces promoter-proximal nucleosomes at highly paused genes. Comparison of nucleosome occupancy around genes with the highest Pausing indices in embryos (Quartile 1) that become unbound by Pol II in S2 cells (n=210). Nucleosomes surrounding these genes are shown when they are: unbound by Pol II in S2 cells (red, left y-axis), or bound by paused Pol II in embryos (gray, right y-axis). Due to the low coverage of H2A.Z nucleosome data (Mavrich et al., 2008), nucleosome counts for this figure were summed over each 50 bp bin from -500 to +1000, rather than for each nucleotide position (as in Figure 6E). The y-axis values reflect the 34-fold difference in nucleosome coverage in the region (-500 to +1000) obtained from S2 cells (n=7,662,427 reads) relative to embryos (n=224,817 reads).
(D) Comparison of nucleosome occupancy around genes with the lowest Pausing indices in embryos that are unbound by Pol II in S2 cells (n=282). Shown are total nucleosome counts in 50 bp bins for these genes when they are: unbound by Pol II in S2 cells (blue, left y-axis), or bound by Pol II in embryos (gray, right y-axis). Nucleosome counts in this figure were normalized to account for the larger number of genes in (D, n=282) vs. (C, n=210), so that nucleosome levels in these graphs could be directly comparable (i.e. raw counts in D were multiplied by 210/282=0.74).

ChIP-chip Samples Compared	r^2
Pol II (total) 1 vs. Pol II (total) 2	0.94
NELF-B 1 vs. NELF-B 2	0.86
DSIF (Spt5) 1 vs. DSIF (Spt5) 2	0.95
Pol II (Ser2-P) 1 vs. Pol II (Ser2-P) 2	0.94
Total Pol II: Untreated 1 vs. Untreated 2	0.96
Total Pol II: Untreated Avg. vs. Mock-treated	0.95
TFIIA 1 vs. TFIIA 2	0.92
Pol II (embryo) 1 vs. Pol II (embryo) 2	0.91
Pol II (no ecdysone) 1 vs. Pol II (no ecdysone) 2	0.95
Pol II (+ ecdysone) 1 vs. Pol II (+ ecdysone) 2	0.92

Table S1. ChIP-chip Signals from Independent Biological Replicates Are In Good Agreement. Shown are correlation coefficients (r^2) for linear regressions of average fold enrichment in the promoter region (+/- 250 bp from 17,116 annotated TSSs) between two biological replicates for each of the indicated comparisons. Also shown is a comparison between the average Pol II signal from two Untreated samples compared to that of one Mock-treated sample.

Motif	Search space	k-mer	**Most NELF-affected** (n=1615)	**Less NELF-affected** (n=4846)
GAGA	-109 to +109	GAGA	438	594
Inr	-4 to 0	TCAGTT	203	194
		TCAGTC	102	86
		TCATTC	61	61
		TTAGTT	63	84
		TCACAC	50	153
		TCAGTG	40	65
		TTCTTT	43	32
		TCAGTA	34	35
		TTAGTC	32	39
		TTAGTA	22	35
		GCAGTT	23	51
		CCAGTT	21	56
		TCATTT	29	41
DPE	+23 to +30	CGGACG	68	31
		CGGTCG	30	27
		CGCACG	28	29
		AAGACG	21	23
		TGGTCG	15	19
		AAGTCG	21	27
		CAGACG	18	36
		CGGTTG	50	36
		CGGTTC	41	26
		AGGTTG	12	11
		AGGTTC	10	12
		CAGTTG	9	20
		CGGTTT	13	13
E-box	-109 to +109	CAGCTG	209	1001
		CACGTG	51	206
		CACCTG	41	197
		CAGGTG	38	175
HOX RE	-109 to +109	TTAATT	202	872
		TTAATG	49	282
		ATAATT	122	671

Table S2. Sequences Employed In Motif Searches. Shown for the indicated motifs (GAGA, Inr, DPE, E-box, and homeo domain Response Element) are the number of occurrences of 4-mer or 6-mer sequences within the designated search spaces around the TSSs of genes that were the most NELF-affected (Quartile 1) or less NELF-affected (Quartiles 2-4). Since some genes may display several occurrences of an individual motif, these numbers are slightly higher than those shown in Figure 2B (which gives the number of genes that possess one or more occurrence of each motif).

Table S3. Genes Bound by Pol II Whose RNA Abundance is Significantly Changed Following NELF-RNAi. In attached excel spreadsheet.

List of genes that were significantly changed by 96 hour RNAi-mediated depletion of NELF in S2 cells, and could be assigned uniquely to one of the 7,466 genes considered for calculation of Pausing Indices. Significant differences were those with fold change>2.0 and P<0.001 in comparisons vs. untreated cells or mock-treated cells; 98 down-regulated and 17 up-regulated genes were thus identified. The fold change and P-value for each RNA species (designated by its Affymetrix probe ID, annotation symbol and gene symbol) is given for NELF-depleted cells vs. both untreated and mock-treated samples.

Extended Experimental Procedures

S2 Cell Growth

Drosophila S2 cells used in most experiments were obtained from the Drosophila Genome Resource Center and grown in M3 media (Sigma) supplemented with bacto-peptone and yeast extract + 10% FBS (GIBCO) as recommended (https://dgrc.cgb.indiana.edu/cells/support/protocols.html). RNAi was performed as described previously (Gilchrist et al., 2009). Experiments on NELF-depleted cells were performed 96 hours after addition of dsRNA targeting NELF-B and NELF-E.

For experiments involving treatment of cells with ecdysone to induce terminal differentiation and Pol II distribution changes, we used *Drosophila* S2 cells purchased from Invitrogen (this line of S2 cells is much more ecdysone-responsive) that were grown in Schneider's media +10% FBS without additional supplementation. For these experiments, S2 cells were treated with 1 µM 20-hydroxyecdysone (Sigma) for 24 hours prior to making ChIP material or isolating chromatin for MNase digestion.

ChIP-chip Localization of Pol II and Elongation Factors

Total Pol II (Rpb3 antibody) and Serine2-phophorylated Pol II (Abcam ab5095) were immunoprecipitated as described (Muse et al., 2007). The NELF-B antibody was raised in rabbit against a soluble NELF-B fragment consisting of amino acid residues 150-561 (Figure S1A). The NELF fragment was produced in *E. coli* as a GST-fusion protein and purified on a glutathione based affinity resin using standard protocols. GST was removed after purification using a TEV cleavage site, leaving only three N-terminal amino acids that were not native to NELF-B. The Spt5 antibody recognizes the C-terminus and was a gift from the Lis laboratory (Andrulis et al., 2000). The TFIIA antibody was a gift from theTjian laboratory (Yokomori et al., 1993). Immunoprecipitations from Drosophila embryo chromatin were performed as described (Sandmann et al., 2006).

Immunoprecipitated material and input DNA were amplified and NimbleGen two-color arrays were probed according to manufacturers suggestions as described (Gilchrist et al., 2009). The arrays contain 2.1 million isothermal probes (50-75 bp) that tile the entire annotated Drosophila genome at 65-bp resolution (Henikoff et al., 2009). Two independent biological replicates for each antibody or condition were in good agreement (Table S1); therefore the data were averaged such that heatmaps and metagene analyses reflect the averaged (n=2) values. The exception to this is the mock-RNAi treated sample, for which there was only one replicate; since data from this sample were in excellent agreement with the two untreated samples all three data sets were combined for comparison with NELF-depleted samples.

MNase-seq and Analysis of Nucleosome Positions

MNase-digested chromatin was prepared for two biological replicates as described in (Gilchrist et al., 2008) except that 200 µl chromatin was digested with 20 units MNase (Worthington) for 45 minutes at 25° C. Following gel purification, mono-nucleosome sized fragments (100-200 bp) were subjected to paired-end sequencing using the

Illumina paired-end protocol. Paired-end reads from two lanes of one biological replicate (total of 35,523,197 reads) and one lane of the second biological replicate (21,319,682 reads) were mapped with Bowtie 0.12.3 to the *D. melanogaster*, Flybase, r5.22 index (Langmead et al., 2009). Paired-end reads were aligned allowing for 2 mismatches and a maximum insert size for valid alignment of 1000 (Bowtie options -v 2 -x 1000). Biological replicates were in good agreement and were combined, resulting in a data set of 32,544,596 unique read pairs identifying both ends of fragments ≥120 bp and ≤180 bp in length (presumed mono-nucleosomes).

MNase-seq was also performed with cells depleted of NELF using RNAi or mock-treated cells, resulting in data sets of 11,218,731 and 11,188,903 unique read pairs identifying both ends of fragments ≥120 bp and ≤180 bp in length, respectively.

Determination of MNase protection in ecdysone-treated and control S2 cells (Invitrogen) was performed as described in (Gilchrist et al., 2008), except that cells were treated with 1 µM 20-hydroxyecdysone or vehicle for 24 hours, and 200 µl chromatin was digested with 20 units MNase for 10 minutes at 25° C.

Data Analysis

Gene Lists Used and Designation of Observed TSSs

A set of non-redundant Drosophila promoters was employed for the generation of heatmaps. This list was constructed from the genome annotations for *D. melanogaster* from Flybase (build r5.17, April 2009; gff genome file downloaded from ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r5.17_FB2009_04/gff/). This list included all Pol II-derived transcripts (mRNA, snoRNA, snRNA, ncRNA) except for miRNAs, as the location of miRNA annotations in the genome build denotes the location of the mature miRNA product (which does not include actual start site of transcription for the precursor RNA), resulting in a total of 22,202 elements. To avoid use of redundant TSSs, we compressed multiple isoforms of the same gene that share a TSS, or have start sites within 25 bp of each other, into a single element maintaining the most upstream TSS (longest gene) for further analyses. Additionally, when more than one gene (unique CG identifiers) had annotated TSSs on the same strand within +/- 25 bp of each other, only one of these genes was retained to generate a set of 17,109 unique TSSs within this window.

Using this gene list, we mapped RNA-seq reads derived from our previous isolation of short, capped nuclear RNAs (Nechaev et al., 2010) around each annotated TSS in the region +/- 150 bp (or +/- 50 bp for genes that were <300 bp from the nearest gene), and the location to which the most reads mapped was called the “observed TSS” if the number of reads at this location was statistically significant (≥3). For each promoter, the total number of short RNA reads mapping within +/-50 bp of the observed TSS was then determined, and the percentage of total reads arising from the observed TSS was calculated to evaluate how focused vs. dispersed initiation was at each TSS. We defined focused transcription initiation as ≥50% of all reads mapping within 50 bp of the observed TSS arising from the observed TSS.

Manual inspection of the ChIP and RNA-seq data surrounding several promoters of interest revealed that the primary peaks of Pol II, DSIF and short transcripts were significantly (>150 bp) offset from the annotated TSSs, but were in agreement with either a GenBank *D. melanogaster* mRNA or spliced EST. We therefore manually incorporated the following seven additional TSSs to our list of 17,109 non-redundant TSSs, for a total of 17,116.

Annotation symbol_ mRNA or spliced EST	Chr	strand	TSS	end
CG11709_AF207541	chrX	plus	11,455,967	11,456,813
CG7571_GH947758	chr3L	plus	17,480,736	17,504,793
CG5903_BT021237	chr3R	minus	11,972,444	11,971,397
CG10520_M59501	chr3R	plus	213,459	215,535
CG5576_AY051558	chr2R	minus	14,298,988	14,297,296
CG4183_AY069419	chr3L	minus	9,370,460	9,369,518
CG4859_C0332215	chr2R	minus	20,572,613	20,562,188

Generation of Heatmaps

We created a search space around each of these 17,116 TSSs from -500 to +1500 bp and divided this into twenty 100-mer bins (from -500 to -401, -400 to -301, etc. excepting 101-mer bin 6, from 0 to 100). The center of each ChIP-chip probe was used to designate its location. The average probe signal, expressed as fold enrichment over input DNA was then determined for each bin, with most bins containing 1-2 probe centers. We note that 274 of 342,320 bins lacked probe centers and thus fold enrichment data. Heatmaps were generated from this fold enrichment data using Partek, and genes were ordered as described in figure legends. Scale bars designate fold-enrichment values, which range from 1 (representing the normalized genome-wide average) to an upper boundary representing the 90^{th} percentile enrichment value for each data set around promoters (+/- 250 bp).

The heatmaps in Figure 1, which show genes rank ordered by Pol II (Ser2-P) signal within the gene (from +500 to +1500), include only 15,945 of the 17,116 Drosophila genes, because genes that were too short to possess probes downstream of +500 were excluded.

The mRNA heatmap depicts GCRMA normalized expression levels (Log_2) derived from our prior analysis (Muse et al., 2007). The scale-bar depicts the range in expression values, from Log_2 4-12. The heatmap showing the location of short RNAs depicts the number of 5'-end sequencing reads obtained from capped, nuclear RNAs of length 25-120 bp (Nechaev et al., 2010). These data are shown in 100-nt, strand specific bins (only sense strand reads are shown) that are centered on the TSSs (-150 to -51, -50 to +50, and +51 to +150). Like mRNA data, the number of short RNA reads at each TSS is depicted using a scale from Log_2 4-12.

To generate heatmaps showing the change in Pol II enrichment upon depletion of NELF, we averaged enrichment values of the untreated and mock-treated samples (n=3) for each 100-mer bin, and from this value subtracted the averaged (n=2) NELF-depleted Pol II ChIP enrichment within the corresponding bin. The change in fold enrichment for each bin is shown as a heatmap where red represents gain and green loss of Pol II ChIP enrichment.

The MNase-seq heatmaps in Figure 3 depict paired-end reads of mono-nucleosomal DNA derived from S2 cells, in regions surrounding TSSs from -500 to +1000 bp divided into thirty 50-mer bins (from -500 to -451, -450 to -401, etc. excepting 51-mer bin 11, from 0 to 50). The center of each fragment (size restricted for fragments ≥120 bp and ≤180 bp) recovered by unique read pairs was used to designate its location; scale bars indicate the number of read centers in each bin from 0 to 36 (80% of all TSSs had fewer than 36 nucleosome centers positioned in the region +101 to +150, where the +1 nucleosome is typically centered).

The MNase-seq heatmap in Figure 5 depicts long sequencing reads (454 technology) of mono-nucleosomal DNA derived from Drosophila embryos and described previously (Mavrich et al., 2008). Sequences were downloaded from http://atlas.bx.psu.edu/data/dmel/ and aligned with blastn to the D. melanogaster genome database using a local NCBI-BLAST2 resource. For each sequence that aligned uniquely with >= 90% identity, the alignment generating the highest bit score was considered a nucleosome location. This yielded a total of 643,929 nucleosomes that mapped uniquely to the genome. The centers of nucleosome locations were mapped in regions surrounding TSSs from -500 to +1000 bp divided into thirty 50-mer bins.

Calculation of Average ChIP-chip Enrichments and Metagene Analysis of Pol II Distribution

To calculate average ChIP-chip signal enrichment within a given region (e.g. TSS +/-250 bp, Fig. 1B), for each gene, fold enrichment values for all probes whose centers were contained within the designated region were averaged.

Metagene analyses to compare Pol II distribution in control vs. NELF-depleted cells were performed on all 17,116 unique TSSs. Probes with centers located from -500 to +1500 with respect to TSSs were selected, and the average probe signal at each bp was calculated from all probes centered at that position. Data were smoothed with a 100-nucleotide sliding window. Data from the untreated and mock-treated samples were averaged to generate the Control metagene (n=3) and compared to averaged data from NELF-depleted samples (n=2).

Determination of Bound TSSs and Calculation of Pausing Index

Unless otherwise indicated, data used to determine which promoters were bound by Pol II and to calculate Pausing indices were derived from average probe signals of 4 ChIP-chip experiments performed on untreated S2 cells (total Pol II antibody, α-Rpb3). TSSs were designated as bound by Pol II if the average fold enrichment in the region +/-250

bp from the annotated TSS was >1.3; 7,803 of 17,116 TSSs (45.6%) were classified as Pol II-bound using this cutoff. NELF depletion significantly decreased Pol II enrichment near bound promoters (+/- 250 bp from the TSS): the median Pol II signal dropped from 2.164 to 1.654 (P-value<0.0001, Mann Whitney test).

The Pausing index was calculated as (average probe enrichment TSS+/- 250bp)/(average probe enrichment from +500 bp to end of gene). Bound genes that lacked probes downstream of +500 (1,171 genes) were excluded from analysis of Pausing index, including 337 bound genes (leaving 7,466 bound genes of sufficient length for analysis). After calculation of Pausing indices, the gene list was separated into quartiles, with the top quartile (1,866 genes, or ~11% of all 17,116 genes) displaying Pausing indices >2.155.

The heatmaps in Figures 2 and 3 display these 7,466 bound genes that were long enough for determination of Pol II signal within the downstream region and calculation of Pausing index.

Analyses of Pol II occupancy and Pausing index were also performed on the ChIP-chip data from *Drosophila* embryos and S2 cells +/- 24 hour treatment with ecdysone, using the same parameters described above (bound=average fold enrichment +/- 250 bp from TSS >1.3). In embryos, this resulted in a total of 7,110 promoters being called bound by polymerase, 6,778 of which had probes downstream of +500 and so were used for calculation of Pausing indices. The heatmap in Figure 5 displays these 6,778 genes rank ordered by Pausing index.

In the absence of ecdysone 8,055 genes were considered bound and 7,661 considered for Pausing index calculation, and in the presence of ecdysone, 5,641 genes were observed to be bound, and 5,443 of these were long enough for analysis of Pausing index.

Metagene Analysis of Nucleosome Occupancy Patterns

The centers of DNA fragments (size restricted for fragments $\geq$120 bp and $\leq$180 bp) identified by unique read pairs were used to designate their locations. Nucleosome occupancy profiles for each group of genes were generated by summing these read centers at each position from -500 to +1000 with respect to the TSS. These raw sums of read counts are referred to throughout simply as “Nucleosome counts”.

When groups within a comparison contained different numbers of TSSs, data for each group were normalized to the number of genes within one quartile of bound genes (n=1,866 genes), and this is referred to as “Normalized nucleosome counts”.

Data involving small numbers of genes (Figure 4E, S4D, S4E) were smoothed with 50 bp moving averages, and Figure 4D and 6E used 100 bp moving averages. Given the low coverage of the H2A.Z nucleosome data (Mavrich et al., 2008), Figure 6A was also smoothed using a 50 bp moving average.

Determination of Changes in Nucleosome Occupancy in Response to NELF RNAi
To calculate fold change in nucleosome occupancy upon NELF-depletion the number of paired-end read centers at each position (-500 to +250) from NELF-depleted cells was divided by the number of paired-end read centers from mock-treated cells. Data were normalized to the total number of reads from the NELF-depleted sample, and were smoothed with 50 bp moving averages.

To identify changes in nucleosome occupancy at genes downregulated by NELF-RNAi, microarray analysis of three biological replicates for untreated, mock-treated, and NELF-depleted S2 cells were performed as described (Gilchrist et al., 2008) using Drosophila Genome 2.0 Genechip® arrays (Affymetrix, Santa Clara, CA). The Rosetta Resolver system was used to calculate P-values and fold changes for transcripts in each experiment. NELF-depleted samples yielded 98 transcripts that were significantly changed after 96 hours of RNAi relative to either untreated and mock-treated cells (P-value <0.001 and fold-change >2, see Table S3), and could be assigned uniquely to one of the 7,466 genes considered for calculation of Pausing Indices. Nucleosome occupancy profiles for these genes were generated by summing read centers at each position from -500 to +500 with respect to the TSS, as determined in either mock-treated or NELF-depleted cells. Data were smoothed with 100 bp moving averages.

The change in nucleosome counts for each quartile of Pausing Indices in the region TSS +/- 200 bp was calculated as $(\text{number of nucleosome centers})_{\text{NELF-depleted}}$ - $(\text{number of nucleosome centers})_{\text{mock-treated}}$.

Calculation of Intron Content, Analysis of Consensus Motifs and Gene Ontology

The number of bp of intron within the region (TSS to +1000 bp) was counted for each of the 7,466 genes used to calculate Pausing indices.

For identification of sequence motifs we used a list of 6,461 genes that were bound by Pol II in S2 cells, and for which we could accurately map the TSS used in S2 cells using short RNA species (Nechaev et al., 2010). This allowed us to precisely determine the "observed" TSS for these genes, allowing for better sequence alignment and detection of promoter motifs. The genes within this group that lost the most Pol II following NELF-depletion (Quartile 1) were considered Most NELF-affected (n=1615). The remaining genes were considered Less NELF-affected (n=4846, Quartiles 2-4).

To identify genes with a TATA box or DREF motif, we scanned each promoter (TSS -109 bp to +1 bp) with the TRANSFAC position weight matrix (PWM) for the TATA-box (ID: M00252) or DREF (ID: M00488). Promoters with a PWM score *p*-value ≤0.0005 were considered to contain a TATA-box or DREF element (Li, 2009). To identify genes with initiator (Inr), GAGA, DPE, E-box, and HOX RE (homeo domain response element) motifs, we searched for exact matches to corresponding motif *k*-mers at specific locations within each promoter, as indicated in Table S2. To identify genes containing the Pause button motif, we constructed a PWM from the consensus motif reported in (Hendrix et al., 2008). This was used as the starting PWM for the EM algorithm in

GADEM (Li, 2009) to identify the Pause button in sequences in the Most NELF-affected group using a *p*-value cutoff of 0.001. The resulting optimized Pause button PWM was used to scan for the Pause button in the Less NELF-affected group.

The program DAVID (http://david.abcc.ncifcrf.gov/) was used to determine which Gene Ontology Biological Processes were over-represented by the Most NELF-affected and Less NELF-affected genes (Dennis et al., 2003; Huang da et al., 2009).

Predictions of Nucleosome Occupancy

D. melanogaster (Fly dm3) genome-wide nucleosome positioning prediction data for average occupancy (predicted probability for each position in the genome to be covered by any nucleosome) were downloaded from http://genie.weizmann.ac.il/software/nucleo_genomes.html (Kaplan et al., 2009). These data include values for the six euchromatic arms chr2L, chr2R, chr3L, chr3R, chr4 and chrX but exclude heterochromatic, scaffold and mitochondrial regions. These regions retain 7,422 of the 7,466 genes considered as bound by Pol II in S2 cells. Genomic position average occupancy values were placed in gene context relative to TSS using custom scripts. The resulting predictions of nucleosome occupancy with respect to individual TSSs were averaged to generate metagene analyses of predicted nucleosome occupancy for select groups of genes as noted in the text.

Supplemental References

Andrulis, E.D., Guzman, E., Doring, P., Werner, J., and Lis, J.T. (2000). High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes & development *14*, 2635-2649.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology *4*, P3.

Gilchrist, D.A., Fargo, D.C., and Adelman, K. (2009). Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. Methods (San Diego, Calif *48*, 398-408.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes & development *22*, 1921-1933.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proceedings of the National Academy of Sciences of the United States of America *105*, 7762-7767.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome research *19*, 460-469.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Molecular and cellular biology *28*, 3290-3300.

Li, L. (2009). GADEM: a genetic algorithm guided formation of spaced dyads coupled with an EM algorithm for motif discovery. J Comput Biol *16*, 317-329.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

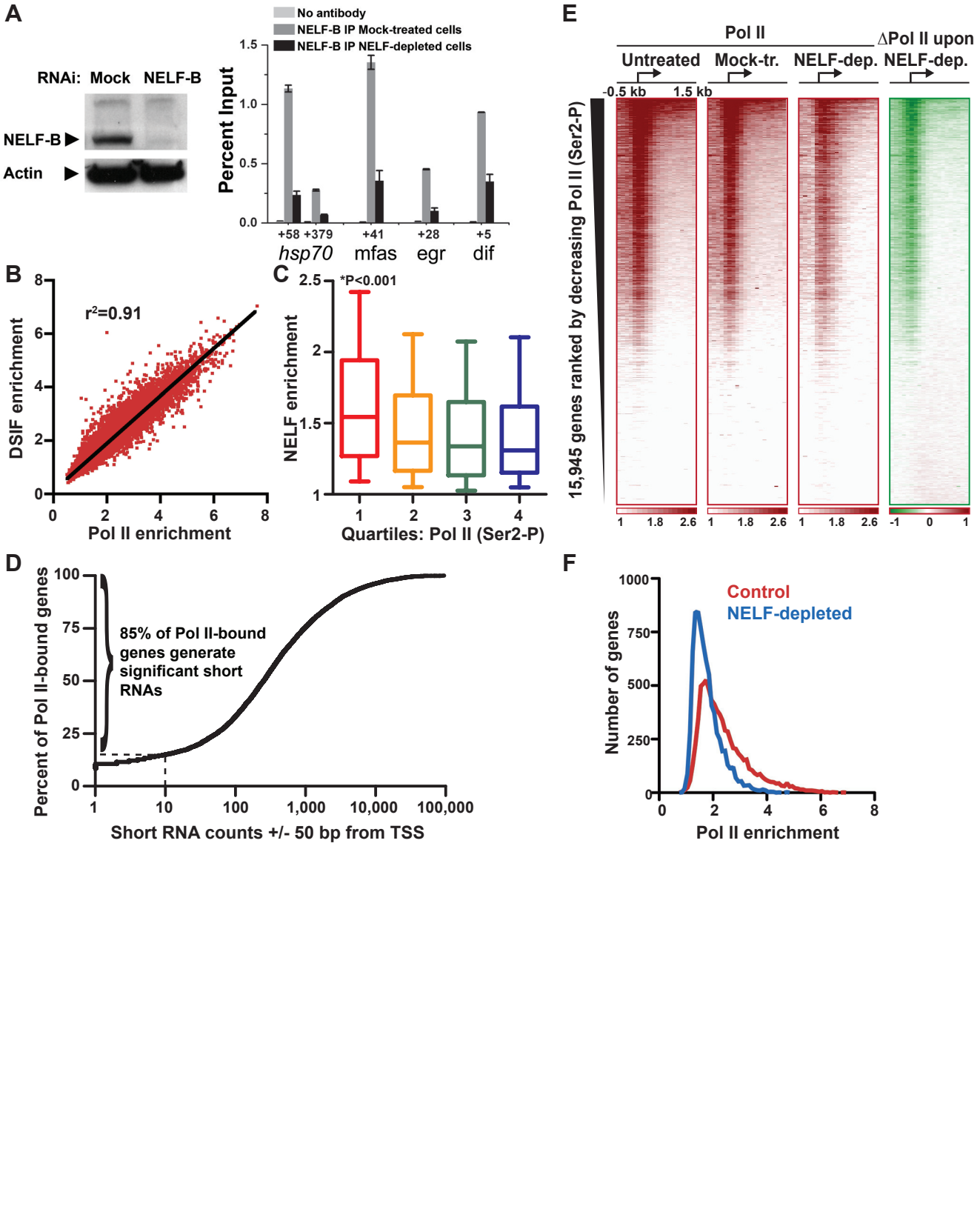
Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nature genetics *39*, 1507-1511.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science (New York, NY *327*, 335-338.

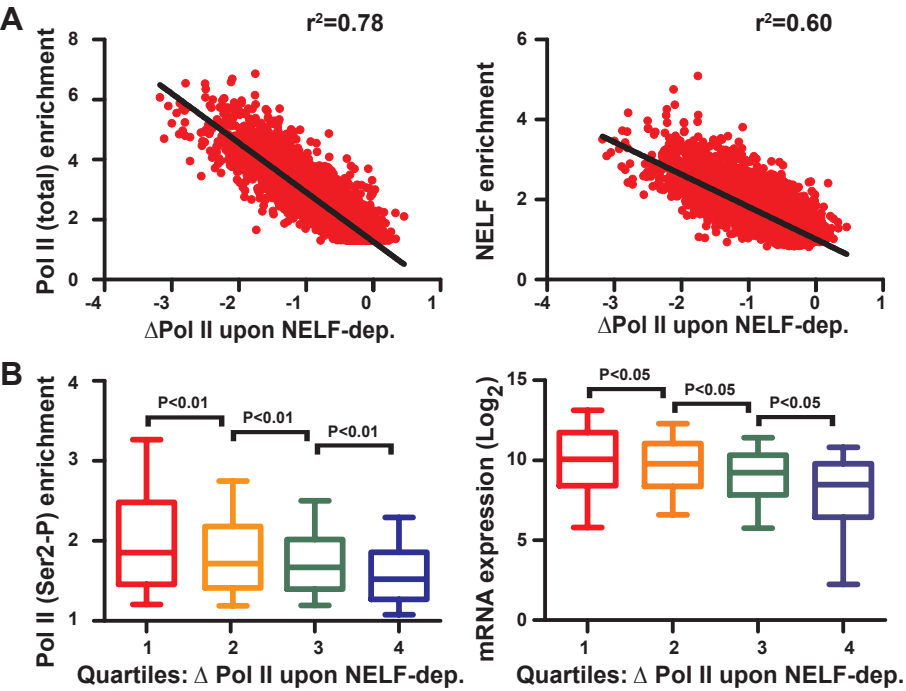
Sandmann, T., Jakobsen, J.S., and Furlong, E.E. (2006). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nature protocols *1*, 2839-2855.

Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.L., and Tjian, R. (1993). Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. Genes & development *7*, 2235-2245.

Supplemental Figure S1



Supplemental Figure S2



C

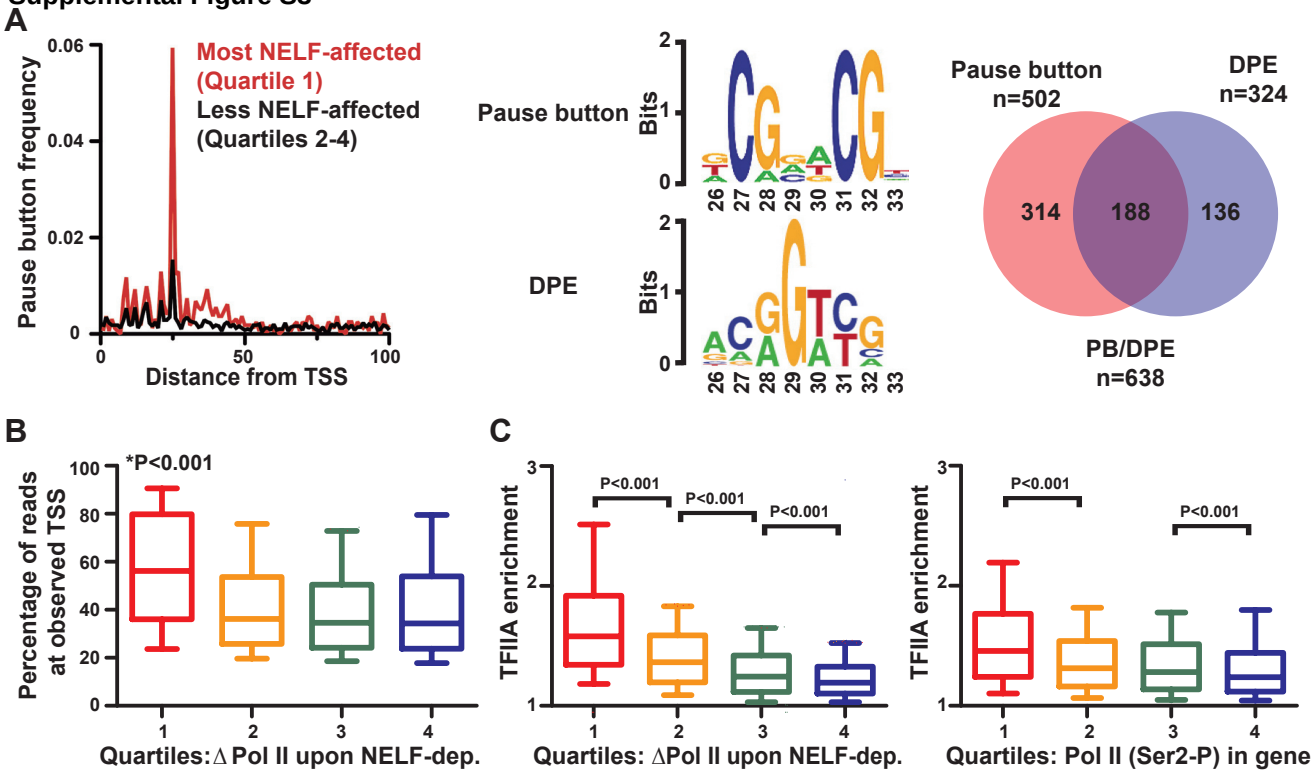
Most NELF-affected (Quartile 1)

GO Term	Count (n=1268)	P-value	Benjamini
developmental process	389 (30.7%)	6.3×10^{-23}	1.4×10^{-21}
multicellular organismal process	396 (31.2%)	1.3×10^{-14}	9.2×10^{-14}
death	60 (4.7%)	1.3×10^{-10}	4.7×10^{-10}
locomotion	55 (4.3%)	1.2×10^{-06}	3.3×10^{-06}
response to stimulus	178 (14.0%)	1.5×10^{-05}	3.7×10^{-05}
biological adhesion	43 (3.4%)	5.6×10^{-05}	1.2×10^{-04}
growth	38 (3.0%)	6.0×10^{-05}	1.2×10^{-04}
reproduction	127 (10.0%)	3.9×10^{-04}	6.6×10^{-04}

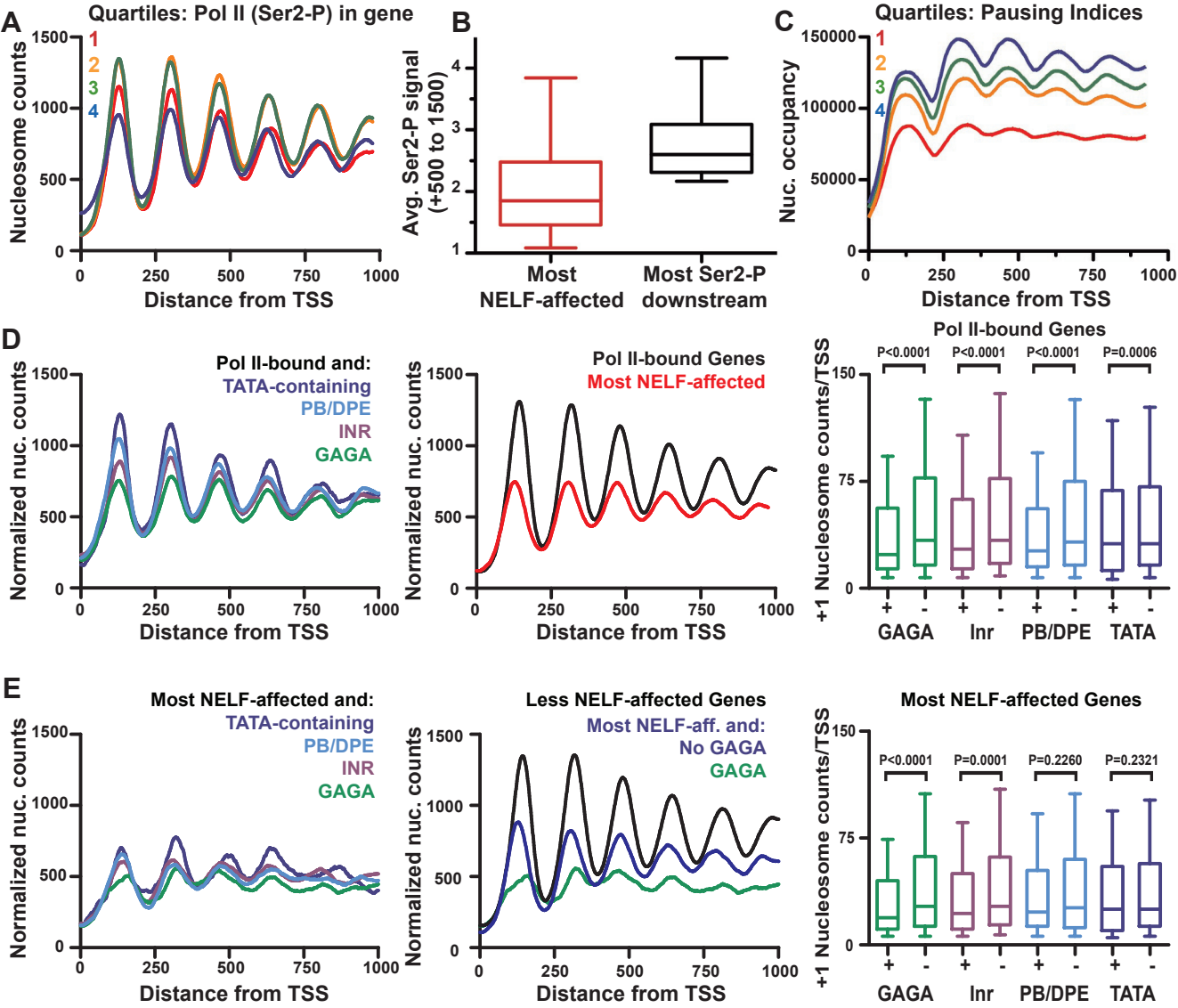
Less NELF-affected (Quartiles 2-4)

GO Term	Count (n=4056)	P-value	Benjamini
cellular process	2009 (49.5%)	7.1×10^{-111}	1.4×10^{-109}
cellular component organization	793 (19.6%)	3.7×10^{-53}	3.7×10^{-52}
metabolic process	1610 (39.7%)	3.8×10^{-34}	2.5×10^{-33}
cellular component biogenesis	301 (7.4%)	5.1×10^{-28}	2.5×10^{-27}
biological regulation	942 (23.2%)	1.7×10^{-16}	8.9×10^{-16}
localization	594 (14.6%)	4.6×10^{-07}	1.5×10^{-06}
establishment of localization	506 (12.5%)	1.3×10^{-06}	3.6×10^{-06}

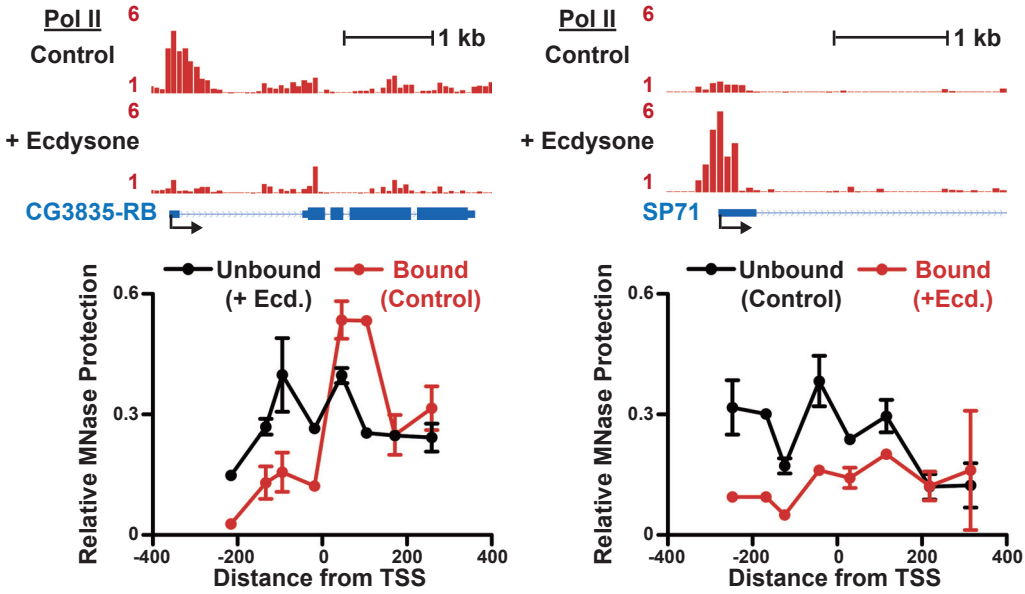
Supplemental Figure S3



Supplemental Figure S4



Supplemental Figure S5



Supplemental Figure S6