Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells

Taeil Kim[a], Shwetha Pazhoor[a], Musheng Bao[a], Zhiqiang Zhang[a], Shino Hanabuchi[a], Valeria Facchinetti[a], Laura Bover[a], Joel Plumas[b], Laurence Chaperot[b], Jun Qin[c], and Yong-Jun Liu[a,1]

[a]Department of Immunology, Center for Cancer Immunology Research, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030; [b]Department of Research and Development, Etablissement Français du Sang Rhône-Alpes Grenoble, 38701 La Tronche, France; and [c]Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030

Edited by Ralph M. Steinman, The Rockefeller University, New York, NY, and approved July 14, 2010 (received for review May 10, 2010)

Toll-like receptor 9 (TLR9) senses microbial DNA and triggers type I IFN responses in plasmacytoid dendritic cells (pDCs). Previous studies suggest the presence of myeloid differentiation primary response gene 88 (MyD88)-dependent DNA sensors other than TLR9 in pDCs. Using MS, we investigated C-phosphate-G (CpG)-binding proteins from human pDCs, pDC-cell lines, and interferon regulatory factor 7 (IRF7)-expressing B-cell lines. CpG-A selectively bound the aspartate-glutamate-any amino acid-aspartate/histidine (DExD/H)-box helicase 36 (DHX36), whereas CpG-B selectively bound DExD/H-box helicase 9 (DHX9). Although the aspartate-glutamate-alanine-histidine box motif (DEAH) domain of DHX36 was essential for CpG-A binding, the domain of unknown function 1605 (DUF1605 domain) of DHX9 was required for CpG-B binding. DHX36 is associated with IFN-α production and IRF7 nuclear translocation in response to CpG-A, but DHX9 is important for TNF-α and IL-6 production and NF-κB activation in response to CpG-B. Knocking down DHX9 or DHX36 significantly reduced the cytokine responses of pDCs to a DNA virus but had no effect on the cytokine responses to an RNA virus. We further showed that both DHX9 and DHX36 are localized within the cytosol and are directly bound to the Toll-interleukin receptor domain of MyD88 via their helicase-associated domain 2 and DUF domains. This study demonstrates that DHX9/DHX36 represent the MyD88-dependent DNA sensors in the cytosol of pDCs and suggests a much broader role for DHX helicases in viral sensing.

cytosolic sensor | innate immunity

The innate immune response is the first line of the host defense system in response to microbial infections. Pattern recognition receptors (PRRs) are sentinels to detect pathogen-associated molecular patterns (PAMPs) and to initiate a downstream signaling cascade leading to the activation of type I IFN and inflammatory cytokines (1). PRRs have been categorized into several families, including Toll-like receptors (TLR) (2), retinoic acid inducible gene I (RIG-I)–like receptors (RLR) (3), and Nod-like receptors (4). In addition, absent in melanoma 2 (AIM2) (5) and RNA polymerase III (6, 7) recently have been identified as cytosolic DNA sensors.Plasmacytoid dendritic cells (pDCs), also known as "professional type I IFN-producing cells," are a specialized cell type for mounting antiviral innate immune responses and are characterized by their selective expression of TLR7 and TLR9 for sensing viral RNA and DNA and constitutive expression of interferon regulatory factor 7 (IRF-7) for rapid IFN responses (2, 8, 9). Although TLR9 was shown to be critical for endowing pDCs with the ability to sense microbial DNA within the endosome compartment, there is a major gap in understanding how TLR9 binds DNA and whether TLR9 represents the only DNA sensor in pDCs. Recent studies suggested the presence of a myeloid differentiation primary response gene 88 (MyD88)-dependent viral sensor other than TLR9 in pDCs (10–12). However, the nature of this sensor is unknown.

Microbial nucleic acids, including their genomic DNA/RNA and replicating intermediates, work as strong PAMPs (13), so finding PRR-sensing pathogenic nucleic acids and investigating their signaling pathway is of general interest. Cytosolic RNA is recognized by RLRs, including RIG-I, melanoma differentiation-associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2). RIG-I senses 5′-triphosphate dsRNA and ssRNA or short dsRNA with blunt ends. MDA5 mainly senses long dsRNA. RIG-I and MDA5 have tandem caspase activation and recruitment domains (CARD) at the N-terminal region that interact with other CARD-containing proteins such as the mitochondrial antiviral signaling (MAVS) protein (14–17). MAVS transduces the signaling cascade through IκB kinase (IKK)-related kinases such as TNF receptor-associated factor family member-associated NF-κB activator (TANK)-binding kinase-1 (TBK1) and inducible IKK, culminating in the activation of IRF3 and inducing the transcription of type I IFNs such as IFN-β.

Cytosolic DNA, whether self or nonself, is a potent pathogenic stimulus of the innate immune system. When bacteria or virus infect the cells and their DNA is introduced into cytosol, the innate immune response is triggered to produce type I IFNs. DNase II-deficient mice, which can not digest DNA from engulfed apoptotic cells, produced robust amounts of type I IFN mediating IRF3 activation (18, 19). Several studies have suggested the existence of cytosolic DNA sensors within the innate immune system (20–22). Intracellular administration of the double-stranded B-form of DNA into mouse embryonic fibroblasts triggered TBK1/IRF3-dependent, TLR9/MyD88-independent, and RIG-I–independent type I IFN responses (21). An earlier study suggested that Z-DNA binding protein 1, also known as "DNA-dependent activator of IRF" (DAI), is a cytosolic DNA sensor (23). However, one study using DAI-knockout mice failed to confirm DAI as the cytosolic DNA sensor (24). AIM2 is another cytosolic DNA sensor that activates inflammasome but is not involved in the type I IFN response (5). Recently, RNA polymerase III was found to sense microbial DNA in cytosol, triggering an RNA intermediate-dependent type I IFN response (6, 7). Whether there are cytosolic sensors that bind DNA directly is not known.

Here we show that biochemical purification of C-phosphate-G (CpG)-binding proteins led to the identification of aspartate-glutamate-any amino acid-aspartate/histidine (DExD/H)-box helicase 36 (DHX36) and DExD/H-box helicase 9 (DHX9) as specific

Author contributions: T.K. and Y.-J.L. designed research; T.K., S.P., M.B., and Z.Z. performed research; S.H., V.F., L.B., J.P., L.C., and J.Q. contributed new reagents/analytic tools; T.K. and Y.-J.L. analyzed data; and T.K. and Y.-J.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: yjliu@mdanderson.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1006539107/-/DCSupplemental.

sensors for CpG-A and CpG-B, respectively, in pDC cells. DHX36 could sense CpG-A by direct binding via the aspartate-glutamate-alanine-histidine box motif (DEAH) domain, whereas DHX9 could sense CpG-B via its domain of unknown function (DUF). Both DHXs are critical for sensing viral DNA pathogens to trigger differential cytokine responses. Under CpG treatment, DHX36 and DHX9 are localized in cytosol but not in endosomal structures and bind to the Toll-IL receptor (TIR) domain of MyD88 via their helicase-associated 2 (HA2) and DUF domains, leading to the activation of IRF7 and p50 (NF-κB).

Results

DHX36 and DHX9 Associate with CpG-A and CpG-B, Respectively. We used biochemical approaches to determine CpG-binding proteins in human primary pDCs, in a human pDC cell line (Gen2.2), and in an IRF7-transfected B-cell line (Namalwa). The Gen2.2 human pDC cell line is a human leukemia cell line that shares identical features with human primary pDCs (25). The Namalwa cell line is a TLR9-expressing human B-cell line transfected with IRF7 that has the ability to produce large amounts of type I IFN in response to CpG-oligodeoxynucleotide (ODN). We first generated biotinylated CpG-A (ODN 2216) and CpG-B (ODN 2006) and confirmed that both were as potent as nonconjugated CpG in activating human primary pDCs and Gen2.2 and Namalwa cells. Then we isolated ≈1.0×10^8 human primary pDCs with >95% purity by anti-blood dendritic cell antigen 2 (BDCA2) magnetic bead sorting from ≈100 human peripheral blood buffy coat samples. Human primary pDCs from each donor were incubated for 4 h with CpG-A+CpG-B (control) or with biotin-CpG-A or biotin-CpG-B. We chose this incubation period because CpG induced significant IFN gene transcription in pDCs at 4 h. Parallel experiments were performed using Gen2.2 and Namalwa cells.

To purify the CpG-bound protein complexes, whole-cell lysates from pDCs treated with biotin-CpG-A, biotin-CpG-B, or CpG-A+CpG-B were purified with NeutrAvidin (NA)-conjugated beads that specifically bind the biotin moiety. The polypeptides bound to biotin-CpGs were separated by gradient polyacrylamide gel electrophoresis and visualized by Coomassie staining. As shown in Fig. 1*A*, several polypeptide bands were observed that were unique to the biotin-CpG-A and biotin-CpG-B pulldowns and were absent from the control CpG-A+CpG-B pulldown. Moreover, these polypeptide bands, specific for either the biotin-CpG-A or biotin-CpG-B pulldown, also were observed in the pulldown experiments using Gen2.2 and Namalwa cells.

The polypeptide bands specific for biotin-CpG-A and biotin-CpG-B pulldowns were excised from the gel and analyzed by liquid chromatography (LC)-MS. We found DHX36 at 110 kDa in biotin-CpG-A pulldowns and DHX9 at 140 kDa in biotin-CpG-B pulldowns. The DExD/H box family includes a large number of

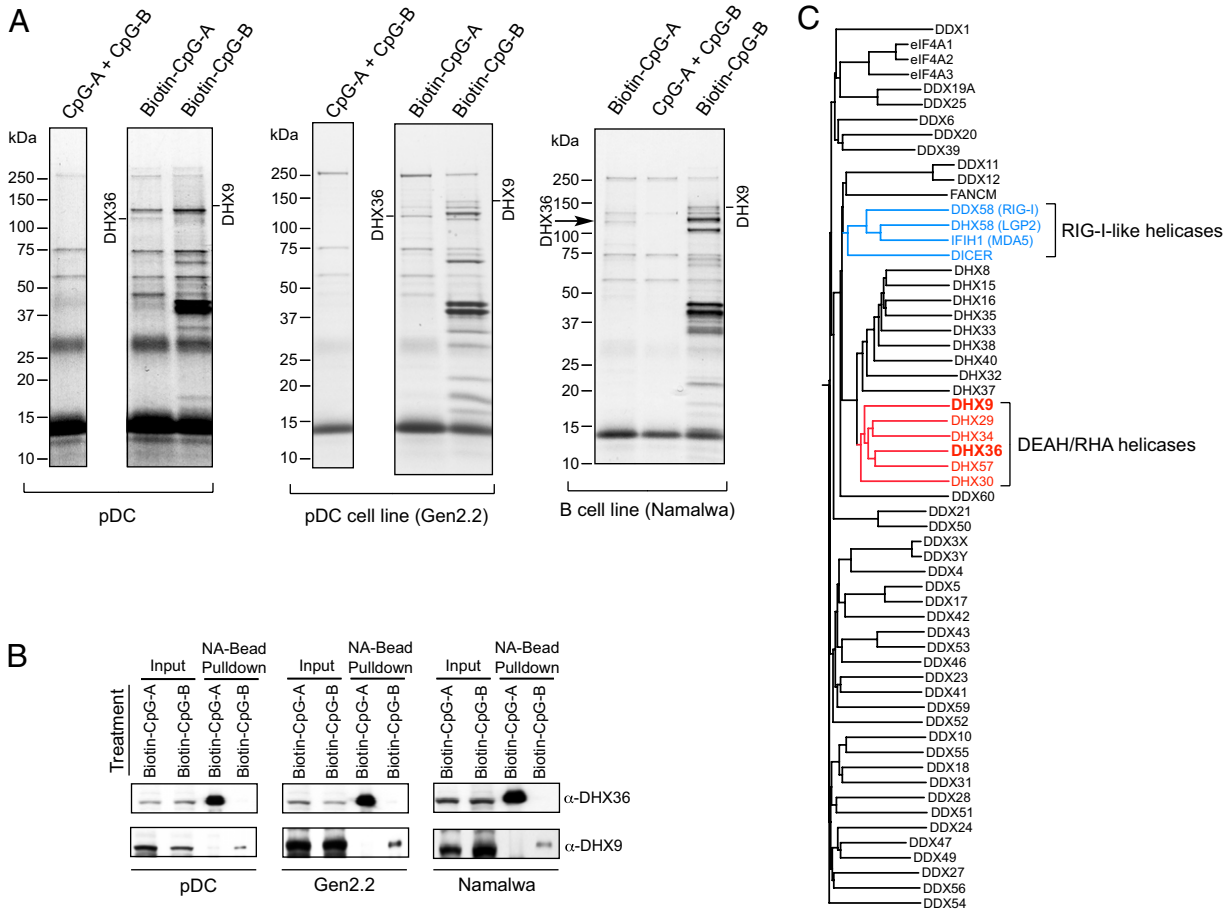


Fig. 1. DHX36 and DHX9 associate with CpG-A and CpG-B, respectively. (*A*) Coomassie staining of CpG-A–associated proteins and CpG-B–associated proteins purified with NA beads from primary pDCs (*Left*), from a pDC cell line (Gen2.2) (*Center*), and from a B-cell line (Namalwa) (*Right*) treated with both CpG-A and CpG-B (CpG-A+CpG-B), with biotin-CpG-A, or with biotin-CpG-B. Proteins identified by LC-MS are as indicated. The sample treated with CpG-A+CpG-B served as control. (*B*) Purified CpG-A–bound proteins and CpG-B—bound proteins from pDCs (*Left*), from Gen2.2 cells (*Center*), and from Namalwa cells (*Right*) were analyzed by immunoblotting with anti-DHX36 and anti-DHX9 antibodies, as indicated. (*C*) Phylogeny of 59 human DExD/H helicases was analyzed by ClustalW analysis.

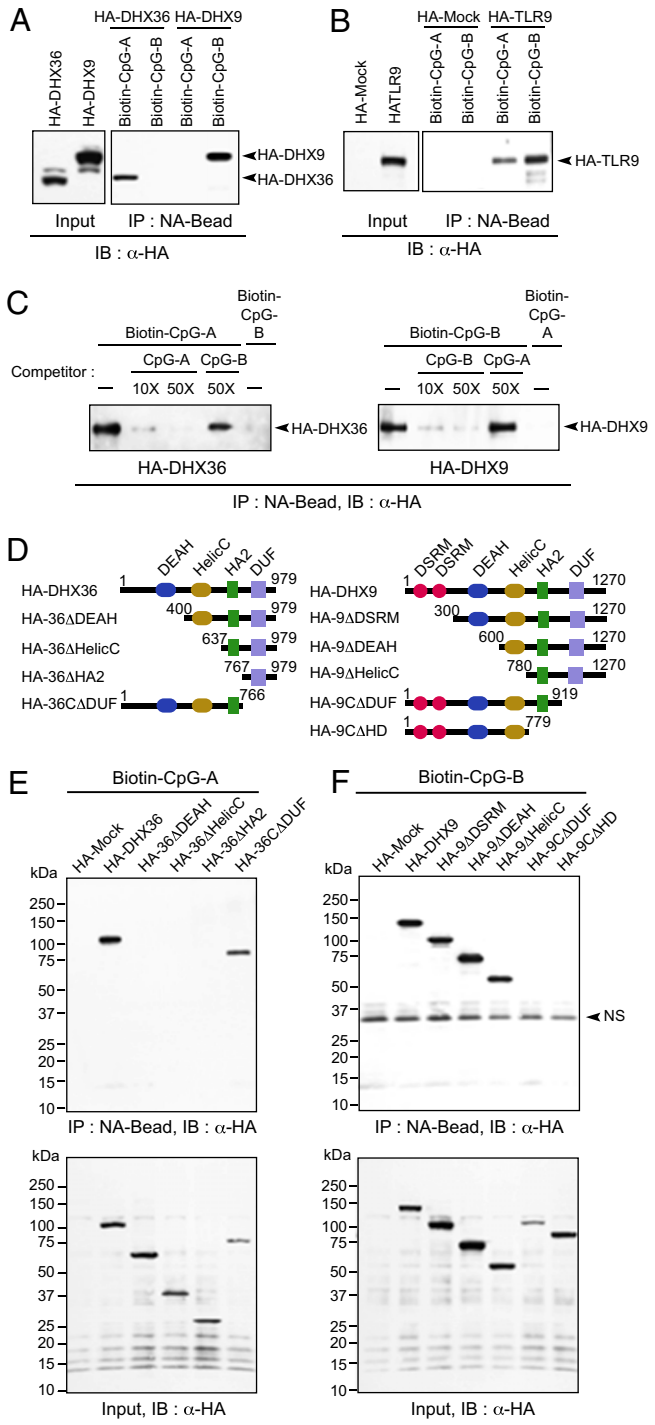


Fig. 2. Domains responsible for specific interactions of DHX36 with CpG-A and of DHX9 with CpG-B. (*A*) Pulldown assays were performed by incubating whole-cell lysates from HA-DHX36– and HA-DHX9–expressing 293T cells with biotin-CpG-A or biotin-CpG-B with NA beads. Bound proteins were analyzed by immunoblotting with anti-HA antibody. (*B*) Pulldown and immunoblotting assays were performed as described in *A* but with whole-cell lysates from HA-Mock or HA-TLR9–expressing 293T cells. (*C*) Competitive pulldown assays. (*Left*) Whole-cell lysates from HA-DHX36–expressing 293T cells were preincubated in the absence or presence of nonconjugated CpG-A (10× and 50×; 1×=1 μM) or CpG-B (50×) and then were incubated with 1× of biotin-CpG-A or biotin-CpG-B along with NA beads. (*Right*) Whole-cell lysates from HA-DHX9–expressing 293T cells were preincubated in the absence or presence of nonconjugated CpG-B (10× and 50×) or CpG-A (50×) and then were incubated with 1× of biotin-CpG-B or biotin-CpG-A along with NA beads. Bound proteins were analyzed by immunoblotting with anti-HA antibody. (*D*) Schematic representations of DHX36 and DHX9 and their serial-deletion mutants. Numbers denote amino acid residues. DEAH, aspartate-glutamate-alanine-histidine box motif; DSRM, double-stranded RNA binding motif; DUF 1605, domain of unknown function 1605; HA2, helicase-associated domain 2; HelicC, helicase C-terminal domain. HA2, helicase-associated domain 2. (*E* and *F*) Pulldown assays were performed by incubating lysates from 293T cells expressing HA-tagged full-length and deletion mutants of DHX36 with biotin-CpG-A (*E*) or by incubating HA-tagged full-length and deletion mutants of DHX9 with biotin-CpG-B (*F*) along with NA beads. Bound proteins were analyzed by immunoblotting with anti-HA. NS, nonspecific band.

proteins that play important roles in RNA metabolism. Members of this family act as RNA helicases or unwindases, using the energy from ATP hydrolysis to unwind RNA structures or to dissociate RNA–protein complexes in cellular processes that require modulation of RNA structures, and are distinguished by the presence of several conserved motifs, including the characteristic DExD/H sequence (in which "x" can be any amino acid). These proteins are highly conserved from viruses and bacteria to humans (26). Although initially classified as RNA helicases, many members of the DExD/H helicase family, including DHX36 and DHX9, were found to display DNA helicase activity (27, 28). Sequence analyses revealed that DHX36 and DHX9 are in the same DEAH/RHA helicase subfamily, which is close to but is clearly separated from the RIG-I–like helicase subfamily (Fig. 1*C*). Because a subgroup of the DExD/H helicase family, including RIG-I, MDA5, and LGP2, was found to function as cytosolic sensors of viral RNA for mounting innate antiviral immune responses (29), we hypothesized that DHX36 and its close relative DHX9 may be the sensors for viral DNA in pDCs.

To confirm whether DHX36 specifically binds CpG-A and DHX9 specifically binds CpG-B, we performed immunoblotting analyses using antibodies specific to DHX36 or DHX9. As shown in Fig. 1*B*, anti-DHX36 antibody detected a 110-kDa band in the biotin-CpG-A pulldown, and anti-DHX9 antibody detected a 140-kDa band in the biotin-CpG-B pulldown.

DHX36 Binds CpG-A via the DEAH Domain, and DHX9 Binds CpG-B via the DUF Domain. To confirm further that the recognition of DHX36 is specific to CpG-A and that the recognition of DHX9 is specific to CpG-B, competition experiments were performed. HA-tagged DHX36 and DHX9 were expressed in 293T cells. The cell lysates were incubated with biotin-CpG-A or biotin-CpG-B in the presence or absence of nonconjugated CpG-A or CpG-B as competitors. Pulldown and immunoblotting analyses showed that DHX36 bound to CpG-A but not to CpG-B (Fig. 2*A*). The binding of DHX36 to biotin-CpG-A was blocked by nonconjugated CpG-A in a dose-dependent fashion, but it was not blocked by nonconjugated CpG-B (Fig. 2*C*). Similarly, DHX9 bound to biotin-CpG-B but not to biotin-CpG-A (Fig. 2*A*). The binding of DHX9 to biotin-CpG-B was blocked by nonconjugated CpG-B in a dose-dependent fashion, but it was not blocked by nonconjugated CpG-A (Fig. 2*C*). Both biotin-CpG-A and biotin-CpG-B were shown to bind to TLR9 (Fig. 2*B*).

DHX36 and DHX9 belong to the RHA subfamily of DExD/H helicase and share the conserved domain structures DEAH, helicase C terminal domain (HelicC), HA2, and DUF1605 (Fig. 2*D*). To define the CpG-A–binding domain of DHX36 and the CpG-B–binding domain of DHX9, serial deletion mutants of DHX36 and DHX9 were expressed in 293T cells, and their ability to bind CpG-A or CpG-B was analyzed. We found that the DEAH domain of DHX36 was essential for binding CpG-A, and the DUF domain of DHX9 was essential for binding CpG-B (Fig. 2 *E* and *F*).

IMMUNOLOGY

DHX36 and DHX9 Are Critical for Microbial DNA-Mediated Cytokine Responses. We next used siRNA technology to investigate the function of DHX36 and DHX9 in pDC responses to CpG-A and CpG-B. Because the manipulation of gene expression has been unsuccessful in human primary pDCs, we performed the experiments in Gen2.2 cells, a pDC cell line derived from human pDC leukemia that expresses TLR7/9, transcription factor 4, and IRF7 and which has all the phenotypic and functional features associated with human primary pDCs (25). Although the CpG-A response by pDCs is characterized by high IFN, the CpG-B response by pDCs is characterized by low IFN but high TNF and IL-6 synthesis (30, 31). Knocking down endogenous DHX36 by two different siRNAs (siDHX36-1 and siDHX36-2) led to >50% reduction in IFN-α production by the Gen2.2 cells in response to CpG-A, whereas knocking down DHX9 by two different siRNAs (siDHX9-1 and siDHX9-2) had no effect on CpG-A–induced IFN-α secretion (Fig. 3 *A* and *B*, *Left*). In contrast, DHX9 knockdown led to an approximate 50% reduction in TNF and IL-6 responses as well as a residual IFN-α response to CpG-B by Gen2.2 cells (Fig. 3 *B*, *Right* and *C*). To determine the function of DHX9/36 in antiviral innate immune responses, pDC cell lines were challenged with DNA virus (HSV) or RNA virus (influenza A virus, Flu A). We found that DHX9/36 knockdown inhibited pDC responses to HSV but not to Flu A virus (Fig. 3*D*). These results suggest that DHX36 and DHX9 are critical for sensing CpG-ODN and viral pathogens and that DHX36/DHX9-mediated sensing is dependent on MyD88.

Both DHX36 and DHX9 Bind MyD88 via the TIR Domain in the Cytosol of pDC. We next conducted experiments to investigate the cellular localization of DHX36 and DHX9 and whether and how they interact with MyD88. Using specific antibodies to DHX9 or DHX36, we found that both were localized within the cytosol of pDCs, and neither was localized within the early endosome marked by the transferrin receptor (TfR) or the late endosome marked by lysosomal-associated membrane protein 1 (LAMP-1) (Fig. 4*A*). We also fractionated endosomal structures by ultracentrifugation based on step-gradient sucrose cushion and found that DHX9 and DHX36 were not fractionated with endosomal structure (Fig. 4*B*). These results indicate that DHX9 and DHX36 function in the cytosol but not in endosomal structures.

A recent study showed that the tobacco plant TLR-like "R" antiviral protein, encoded by the *N* gene, bound directly via its TIR domain to the Tobacco mosaic virus pathogen elicitor protein, p50 helicase (32). This report prompted us to determine whether DHX9/36 binds to MyD88 directly. To do so, we expressed the DHX9 and DHX36 HA-tagged deletion mutants depicted in Fig. 2*D* with Myc-tagged deletion mutants of MyD88 in 293T cells (Fig. 4*C*) and performed coimmunoprecipitation assays. We observed that MyD88 indeed could interact with both DHX36 and DHX9 through the TIR domain of MyD88 (Fig. 4*D*). Moreover, the HA2 and DUF domains of DHX36 and DHX9 were critical for interaction with MyD88, and the HelicC domain of DHX9 was involved also (Fig. 4*E*). This result suggests that DHX9 and DHX36 indeed bind to MyD88 via interactions between the HA2 and DUF domains of DHX and the TIR domain of MyD88.

DHX36 and DHX9 Trigger Downstream Signaling to Activate IRF7 and p50 (NF-κB). To investigate the requirement of DHX36 and DHX9 for CpG-triggered signaling culminating in the activation of IRF7 and p50 (Fig. 5 *A* and *B*), we knocked down MyD88, TLR9, DHX36, and DHX9 expression in Gen2.2 cells and monitored nuclear localization of IRF7 and p50 (Fig. 5*D*). As shown in Fig. 5*C*, when MyD88 or TLR9 expression was dampened by siRNA, nuclear localization of IRF7 by CpG-A and nuclear localization of p50 by CpG-B was diminished. Interestingly, whereas knockdown of DHX36 diminished the nuclear localization of IRF7 by CpG-A but not the nuclear localization of p50 by CpG-B, knockdown of DHX9 inhibited nuclear localization of p50 by CpG-B but not the nuclear localization of IRF7 by CpG-A. These results suggest that DHX36-mediated sensing of CpG-A and DHX9-mediated sensing of CpG-B trigger different signal pathways in pDCs.

Discussion

In this study we have demonstrated that DHX36 and DHX9, members of the RHA subfamily of DExD/H helicases, represent

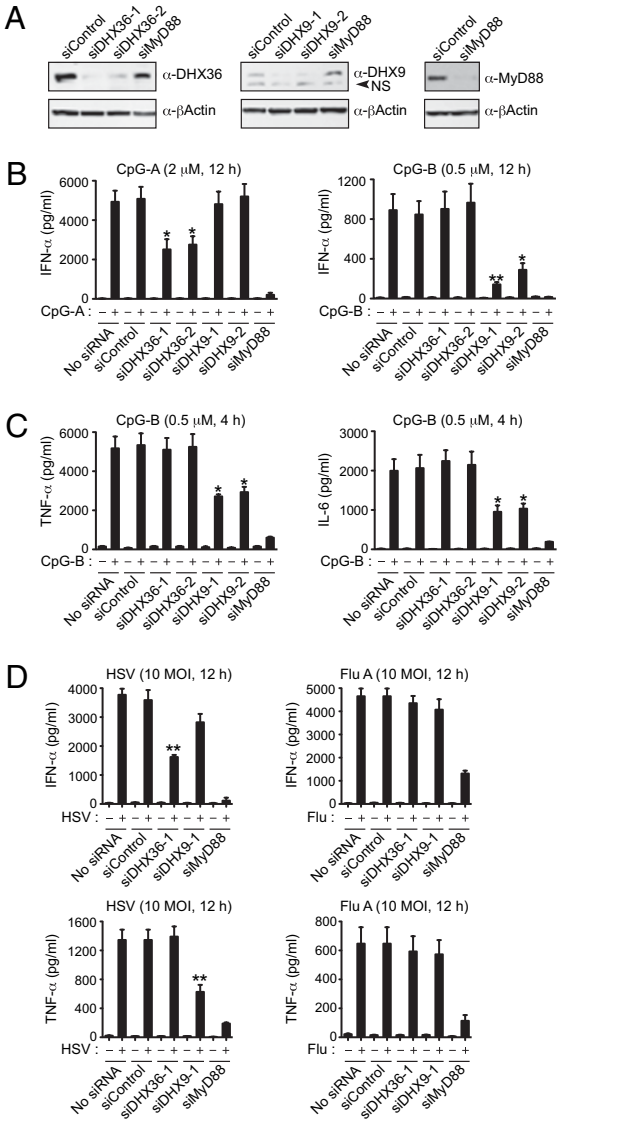
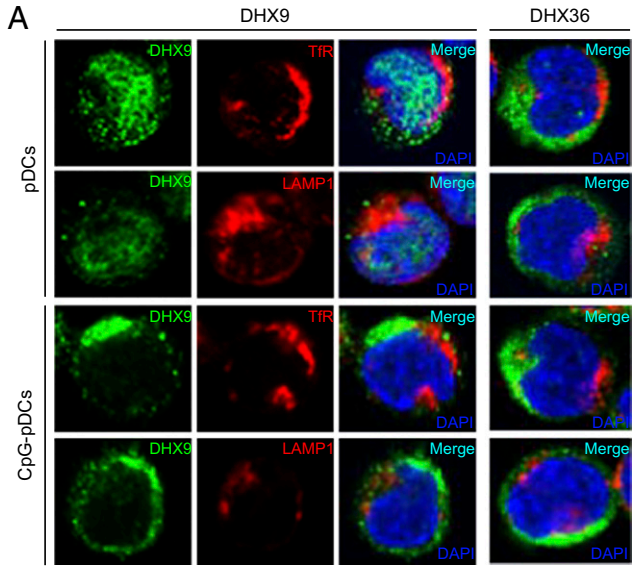
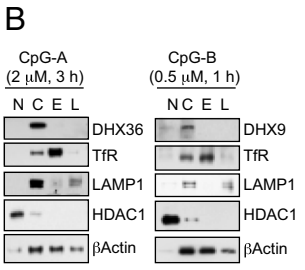
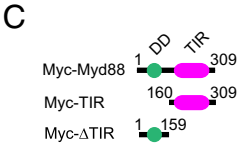
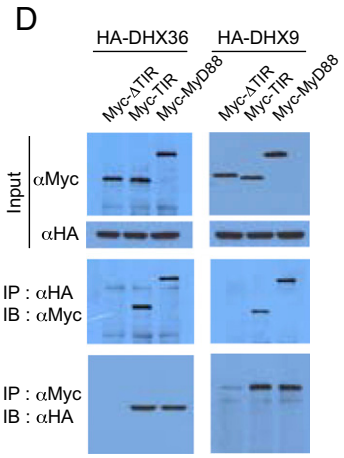


Fig. 3. DHX36 and DHX9 are critical for microbial DNA-mediated cytokine responses. (*A*) Gen2.2 cells were transfected with nonspecific siRNA (siControl), two siRNAs targeting DHX36 (siDHX36-1 and siDHX36-2), two siRNAs targeting DHX9 (siDHX9-1 and siDHX9-2), or a MyD88-targeting siRNA (siMyD88). Endogenous DHX36, DHX9, and MyD88 were monitored by immunoblotting with anti-DHX36, anti-DHX9, and anti-MyD88 antibodies, as indicated at the right. NS, nonspecific bands. (*B* and *C*) ELISAs to monitor IFN-α production from Gen2.2 cells transfected with siRNA, as indicated, upon treatment with 2 μM of CpG-A for 12 h or 0.5 μM of CpG-B for 12 h (*B*) or to monitor TNF-α and IL-6 production upon treatment with 0.5 μM of CpG-B for 4 h (*C*). (*D*) ELISAs to monitor IFN-α and TNF-α production from Gen2.2 cells transfected with siRNA, as indicated, upon treatment with HSV or Flu A (multiplicity of infection, 10) for 12 h. Data are mean ± SD from three independent experiments. $*P < 0.05$ and $**P < 0.01$ versus sample transfected with siControl and treated with CpG or virus.









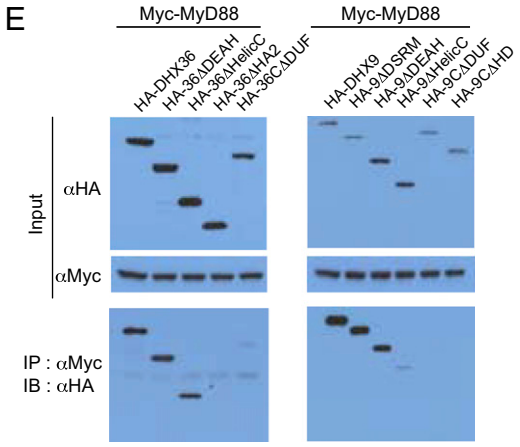
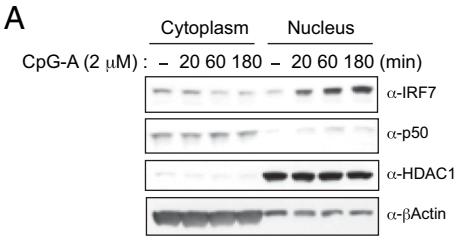
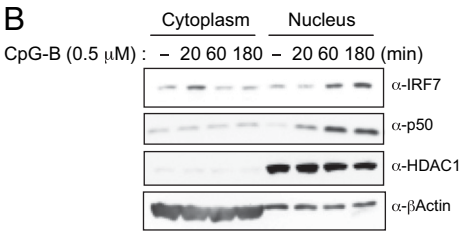
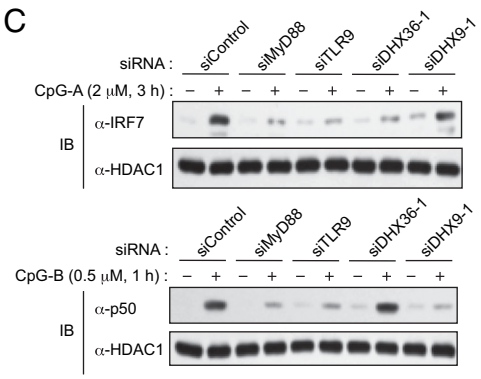


Fig. 4. DHX36 and DHX9 interact with MyD88 in the cytosol of pDCs. (*A*) Confocal images of pDCs untreated or treated with CpG-A or CpG-B. The three columns on the left show staining of DHX9 and TfR (an early endosome marker) or LAMP1 (a late endosome marker). The column on the right shows staining of DHX36 and TfR or LAMP1. (*B*) Endosomal fractionation. Fractionated early endosome and late endosome by ultracentrifugation based on step-gradient sucrose was examined by immunoblotting with anti-DHX9, anti-DHX36, anti-TfR, anti-LAMP1, anti-histone deacetylase 1 (anti-HDAC1), and anti-β-Actin antibodies. (*C*) Schematic representations of MyD88 and two deletion mutants. Numbers denote amino acid residues. DD, death domain; TIR, Toll-IL-1R homologous domain. (*D*) Lysates from 293T cells expressing Myc-tagged full-length or deletion mutants of MyD88 with HA-DHX36 or HA-DHX9 were immunoprecipitated with anti-HA antibody, followed by immunoblotting with anti-Myc antibody, or vice versa. (*E*) Lysates from 293T cells expressing HA-tagged full-length or deletion mutants of DHX36 or DHX9 with Myc-MyD88 were immunoprecipitated with anti-Myc antibody, followed by immunoblotting with anti-HA antibody.







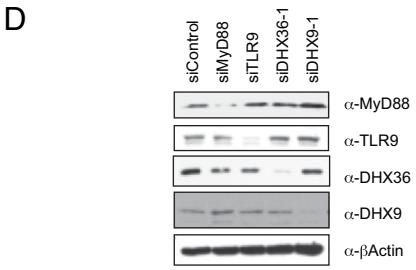


Fig. 5. DHX36 and DHX9 trigger downstream signaling cascades. (*A* and *B*) Nuclear fractions from Gen2.2 cells stimulated with 2 μM of CpG-A (*A*) or 0.5 μM of CpG-B (*B*) in a time-dependent manner were immunoblotted with anti-IRF7 or anti-p50 antibodies. HDAC1 and β-Actin were used as nuclear and cytosolic markers, respectively. (*C*) Nuclear fractions from Gen2.2 cells transfected with siRNA, as indicated, upon treatment with 2 μM of CpG-A for 3 h (*Upper*) or 0.5 μM of CpG-B for 1 h (*Lower*) were immunoblotted with anti-IRF7 or anti-p50 antibodies. HDAC1 was used as a loading control. (*D*) Endogenous DHX36, DHX9, TLR9, and MyD88 were monitored by immunoblotting with anti-DHX36, anti-DHX9, anti-TLR9, and anti-MyD88 antibodies, as indicated at the right.

the CpG-A sensor and the CpG-B sensor, respectively, in human pDCs. Therefore, the DExD/H helicase family has a dedicated RIG-I–like subfamily for sensing viral RNA, as shown by pre-

vious studies (29), and a dedicated RHA subfamily for sensing viral DNA, as shown here. Interestingly, each member of the RIG-I subfamily of DExD/H helicases senses viral RNA differently. For example, RIG-I senses 5′ triphosphate ds/ssRNA, MDA5 senses cytoplasmic dsRNA (33), and LGP2 senses dsRNA independent of 5′ triphosphates (34). In parallel, we found that two members of the RHA subfamily of DExD/H helicases sense microbial DNA differently: DHX36 senses CpG-A using the DEAH domain, whereas DHX9 senses CpG-B using the DUF domain. The DEAH domain has ATP-dependent helicase activity and is known to bind RNA or DNA. As indicated by its name, the function of the DUF domain had not been identified; here we report a putative function for the DUF domain as a binding domain to unmethylated CpG ODN. Because cell lysates were used for in vitro binding assay, the possibility that indirect binding of helicases to CpGs was mediated by other proteins cannot be excluded completely. In eukaryotes, a total of 59 DExD/H helicases have been grouped into four subfamilies: RIG-1-like, DEAH/RHA, aspartate-glutamate-alanine-aspartate (DEAD)-Box, and Snf1-related kinase interacting protein 2 (Ski2)-like (26). Our study suggests that DExD/H helicases may play a much broader role in antiviral innate immune responses than previously thought. Indeed, a recent study demonstrated that Dicer-2, another RIG-I–like DExD/H helicase, senses viral nucleic acids in the *Drosophila* innate immune system (35). The potential roles of the other DExD/H helicase family members in sensing microbial nucleic acids remain to be explored.

Studies using *Myd88*-knockout mice have shown that the innate immune responses of pDCs to viral DNA, CpG-A, and CpG-B are totally dependent on MyD88 (36). We confirm these results in our current study by showing that MyD88 knock down by siRNA completely abolishes the innate immune responses of pDCs to both CpG-A and CpG-B. Previous studies have indicated the presence of TLR9-independent, MyD88-dependent viral DNA sensors in pDCs (10–12). Our study identified DHX36 and DHX9 as MyD88-dependent viral DNA sensors in pDCs that are localized within the cytosol and bind MyD88 directly via its TIR domain. These findings suggest that MyD88 serves as a key adaptor molecule for both endosomal TLRs and cytosolic helicases. Interestingly, the interaction between the TIR domain and helicases already has been documented in the plant world (32). This study suggests that the MyD88-dependent endosomal TLR9 sensor and the cytosolic DHX9/36 sensor may play complementary roles in viral DNA sensing, with TLR9 sensing viral entry and DHX9/36 sensing viral replication.

Materials and Methods

Statistical Analysis. Statistically significant differences were determined by unpaired, two-tailed, Student's *t* test. *P* values <0.05 were considered statistically significant. Statistical analysis of data was done using GraphPad Prism version 5 for Macintosh (GraphPad Software).

Phylogenetic Analysis. All protein sequences of listed DExD/H helicases were retrieved from the UniProt browser with the Swiss-Prot Knowledgebase, which is manually annotated and reviewed. These sequences were aligned, and all residues that contained gaps were removed from the alignment. The ClustalW multiple sequence alignment program with neighbor joining estimation was used to generate a phylogenetic tree.

Other detailed methods are provided in *SI Text*.

ACKNOWLEDGMENTS. We thank Stephanie Watowich and Shao-Cong Sun for discussions and Melissa Wentz for critical reading. We thank all our colleagues in our laboratory.

1. Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. *Cell* 124:783–801.
2. Blasius AL, Beutler B (2010) Intracellular toll-like receptors. *Immunity* 32:305–315.
3. Yoneyama M, Fujita T (2007) RIG-I family RNA helicases: Cytoplasmic sensor for antiviral innate immunity. *Cytokine Growth Factor Rev* 18:545–551.
4. Meylan E, Tschopp J, Karin M (2006) Intracellular pattern recognition receptors in the host response. *Nature* 442:39–44.
5. Schroder K, Muruve DA, Tschopp J (2009) Innate immunity: Cytoplasmic DNA sensing by the AIM2 inflammasome. *Curr Biol* 19:R262–R265.
6. Ablasser A, et al. (2009) RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol* 10: 1065–1072.
7. Chiu YH, Macmillan JB, Chen ZJ (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* 138:576–591.
8. Hemmi H, et al. (2000) A Toll-like receptor recognizes bacterial DNA. *Nature* 408: 740–745.
9. Colonna M, Trinchieri G, Liu YJ (2004) Plasmacytoid dendritic cells in immunity. *Nat Immunol* 5:1219–1226.
10. Hochrein H, et al. (2004) Herpes simplex virus type-1 induces IFN-alpha production via Toll-like receptor 9-dependent and -independent pathways. *Proc Natl Acad Sci USA* 101:11416–11421.
11. Jung A, et al. (2008) Lymphocytoid choriomeningitis virus activates plasmacytoid dendritic cells and induces a cytotoxic T-cell response via MyD88. *J Virol* 82:196–206.
12. Hokeness-Antonelli KL, Crane MJ, Dragoi AM, Chu WM, Salazar-Mather TP (2007) IFN-alphabeta-mediated inflammatory responses and antiviral defense in liver is TLR9-independent but MyD88-dependent during murine cytomegalovirus infection. *J Immunol* 179:6176–6183.
13. Takeuchi O, Akira S (2009) Innate immunity to virus infection. *Immunol Rev* 227: 75–86.
14. Kawai T, et al. (2005) IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol* 6:981–988.
15. Seth RB, Sun L, Ea CK, Chen ZJ (2005) Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* 122:669–682.
16. Xu LG, et al. (2005) VISA is an adapter protein required for virus-triggered IFN-beta signaling. *Mol Cell* 19:727–740.
17. Meylan E, et al. (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* 437:1167–1172.
18. Okabe Y, Kawane K, Akira S, Taniguchi T, Nagata S (2005) Toll-like receptor-independent gene induction program activated by mammalian DNA escaped from apoptotic DNA degradation. *J Exp Med* 202:1333–1339.
19. Yoshida H, Okabe Y, Kawane K, Fukuyama H, Nagata S (2005) Lethal anemia caused by interferon-beta produced in mouse embryos carrying undigested DNA. *Nat Immunol* 6:49–56.
20. O'Riordan M, Yi CH, Gonzales R, Lee KD, Portnoy DA (2002) Innate recognition of bacteria by a macrophage cytosolic surveillance pathway. *Proc Natl Acad Sci USA* 99: 13861–13866.
21. Ishii KJ, et al. (2006) A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nat Immunol* 7:40–48.
22. Stetson DB, Medzhitov R (2006) Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* 24:93–103.
23. Takaoka A, et al. (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* 448:501–505.
24. Ishii KJ, et al. (2008) TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* 451:725–729.
25. Chaperot L, et al. (2006) Virus or TLR agonists induce TRAIL-mediated cytotoxic activity of plasmacytoid dendritic cells. *J Immunol* 176:248–255.
26. Linder P (2006) Dead-box proteins: A family affair—active and passive players in RNP-remodeling. *Nucleic Acids Res* 34:4168–4180.
27. Vaughn JP, et al. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem* 280:38117–38120.
28. Zhou K, et al. (2003) RNA helicase A interacts with dsDNA and topoisomerase IIalpha. *Nucleic Acids Res* 31:2253–2260.
29. Beutler B, et al. (2007) Genetic analysis of resistance to viral infection. *Nat Rev Immunol* 7:753–766.
30. Vollmer J, et al. (2004) Characterization of three CpG oligodeoxynucleotide classes with distinct immunostimulatory activities. *Eur J Immunol* 34:251–262.
31. Verthelyi D, Ishii KJ, Gursel M, Takeshita F, Klinman DM (2001) Human peripheral blood cells differentially recognize and respond to two distinct CPG motifs. *J Immunol* 166:2372–2377.
32. Burch-Smith TM, et al. (2007) A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biol* 5:e68.
33. Kato H, et al. (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441:101–105.
34. Pippig DA, et al. (2009) The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nucleic Acids Res* 37:2014–2025.
35. Deddouche S, et al. (2008) The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in drosophila. *Nat Immunol* 9:1425–1432.
36. Adachi O, et al. (1998) Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* 9:143–150.

Cell

Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation

Daniel A. Gilchrist,[1] Gilberto Dos Santos,[1] David C. Fargo,[2] Bin Xie,[4] Yuan Gao,[4,5] Leping Li,[3] and Karen Adelman[1,*]
[1]Laboratory of Molecular Carcinogenesis
[2]Library and Information Services
[3]Biostatistics Branch
National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA
[4]Division of Genomics, Epigenomics and Bioinformatics, Lieber Institute for Brain Development
[5]Neurogeneration and Stem Cell Biology Program, Institute of Cell Engineering
Johns Hopkins University, Baltimore, MD 21205, USA
*Correspondence: adelmank@niehs.nih.gov
DOI 10.1016/j.cell.2010.10.004

SUMMARY

Metazoan transcription is controlled through either coordinated recruitment of transcription machinery to the gene promoter or regulated pausing of RNA polymerase II (Pol II) in early elongation. We report that a striking difference between genes that use these distinct regulatory strategies lies in the "default" chromatin architecture specified by their DNA sequences. Pol II pausing is prominent at highly regulated genes whose sequences inherently disfavor nucleosome formation within the gene but favor occlusion of the promoter by nucleosomes. In contrast, housekeeping genes that lack pronounced Pol II pausing show higher nucleosome occupancy downstream, but their promoters are deprived of nucleosomes regardless of polymerase binding. Our results indicate that a key role of paused Pol II is to compete with nucleosomes for occupancy of highly regulated promoters, thereby preventing the formation of repressive chromatin architecture to facilitate further or future gene activation.

INTRODUCTION

Eukaryotic gene expression begins with recruitment of the transcription machinery to a gene promoter and formation of a preinitiation complex composed of RNA polymerase II (Pol II) and general transcription factors (Roeder, 2005). This step is highly regulated and is enhanced by DNA sequence motifs within the promoter region, which are recognized by general transcription factors to stabilize transcription complex assembly (Juven-Gershon et al., 2008). Interestingly, these core promoter motifs are more prevalent at highly regulated genes than at constitutively active housekeeping genes, suggesting that these two classes of promoters might use different mechanisms to attract the transcription machinery (Basehoar et al., 2004; Hendrix et al., 2008).

Chromatin structure also impacts polymerase recruitment by modulating promoter accessibility, and activation of some genes requires disassembly of promoter nucleosomes by ATP-dependent chromatin-remodeling complexes (Cairns, 2009). In the yeast *Saccharomyces cerevisiae*, highly regulated promoters are particularly likely to be occluded by nucleosomes before activation, making these genes reliant on nucleosome remodeling for transcription (Tirosh and Barkai, 2008). However, global mapping of nucleosomes in yeast has revealed that most promoter regions display low nucleosome occupancy even when the gene is inactive (Yuan et al., 2005; Albert et al., 2007), suggesting that assembly of promoter nucleosomes is inherently disfavored. Indeed, yeast promoter DNA sequences often contain rigid poly (dA:dT) tracts that deter nucleosome assembly (Iyer and Struhl, 1995). Accordingly, intrinsic sequence preferences for nucleosome formation contribute significantly to accessibility of yeast promoters in vivo (Sekinger et al., 2005; Kaplan et al., 2009; Zhang et al., 2009).

Human and *Drosophila* promoters are also generally nucleosome deprived in a manner that is not dependent on gene expression (Mito et al., 2005; Ozsolak et al., 2007; Mavrich et al., 2008; Schones et al., 2008). However, the mechanisms for this nucleosome depletion appear to be different than in yeast. Metazoan genes are much more G+C-rich than their yeast counterparts and, in contrast to yeast, are reported to intrinsically favor nucleosome formation around their promoters (Kaplan et al., 2009; Tillo et al., 2010). Thus, active mechanisms must contribute to the broad nucleosome depletion observed in metazoans, such as recruitment of chromatin-remodeling complexes or association of the transcription machinery (Kim et al., 2005; Ozsolak et al., 2007). Indeed, pausing of Pol II near promoters can affect both the positioning (Mavrich et al., 2008; Schones et al., 2008) and occupancy of nucleosomes (Gilchrist et al., 2008).

Polymerase pausing was first described at the *Drosophila* heat shock genes, where Pol II synthesizes 25–50 nucleotides (nt) of RNA prior to heat shock and then halts to "wait" for an activating signal (Rougvie and Lis, 1988; Lis, 1998). Heat shock immediately triggers the release of paused polymerase into the gene,

allowing an extremely rapid and robust transcriptional response (Lis, 1998). Rapid activation of heat shock genes is also favored by the lack of nucleosomes within the initially transcribed region (Wu, 1980), which would otherwise present barriers to efficient elongation (Izban and Luse, 1992). Although promoter-proximal pausing was once considered a rare phenomenon, recent work has demonstrated that it is a common regulatory strategy in higher eukaryotes (Muse et al., 2007; Zeitlinger et al., 2007; Core et al., 2008; Nechaev et al., 2010; Rahl et al., 2010). However, despite the growing appreciation for the widespread nature of pausing, the functions of paused Pol II remain to be elucidated.

We investigated the relationships among pausing, gene activity, and chromatin structure by performing high-resolution mapping of Pol II, pause-inducing factors, and nucleosomes across the *Drosophila* genome. Our data reveal that Pol II pausing occurs globally and plays a decisive role in determining promoter nucleosome occupancy. Moreover, we find that genes regulated by pausing rather than Pol II recruitment have distinct "default" chromatin architectures specified by their DNA sequences. Although recruitment-limited genes have intrinsically nucleosome-deprived promoters, genes with paused Pol II require polymerase occupancy to prevent promoter nucleosome assembly. These findings indicate that a gene's intrinsic nucleosome occupancy in the naive, or default, state is instructive for gene regulation and suggest that the interplay between static information within promoter DNA sequences and the dynamics of polymerase pausing facilitates precise control of gene expression.

RESULTS

Pausing of Pol II Is Widespread and Occurs at Highly Active Genes

Regulation of Pol II pausing involves the coordinated action of both negative and positive elongation factors (Marshall and Price, 1992). Shortly after transcription initiation, the pause-inducing factors *n*egative *el*ongation *f*actor (NELF) and *D*RB-*s*ensitivity *i*nducing *f*actor (DSIF) associate with the polymerase and decrease elongation efficiency (Yamaguchi et al., 2002; Wu et al., 2003; Cheng and Price, 2007; Lee et al., 2008). To examine the prevalence of pausing during early elongation, we used genome-wide ChIP-chip on high-density tiling arrays to compare NELF and DSIF distribution in *Drosophila* S2 cells with that of Pol II (see Figure S1A and Table S1 available online). Heat maps representing fold enrichment over input DNA (Figure 1A) reveal a broad colocalization of NELF, total Pol II, and DSIF near promoters. In fact, the average promoter signals for these factors correspond extremely well (Figure 1B; Figure S1B), indicating that NELF and DSIF generally associate with Pol II in the promoter-proximal region. Additionally, in agreement with recent reports (Rahl et al., 2010), most genes show enrichment in Pol II signal near promoters relative to downstream regions, suggesting that recruited polymerases are generally released inefficiently into genes.

Release of paused polymerase into productive elongation is triggered by the kinase activity of the positive transcription elongation factor b (P-TEFb) (Marshall and Price, 1995; Peterlin and Price, 2006). P-TEFb phosphorylates the Serine-2 residues on the Pol II C-terminal domain, DSIF and NELF, leading to dissociation of NELF and recruitment of factors that facilitate transcription elongation and RNA processing. The tight correlation between NELF and Pol II signals near promoters suggests that each round of transcription involves NELF-mediated pausing, such that active genes should be enriched in NELF. To confirm this, we identified active genes by performing ChIP-chip with an antibody that recognizes the Serine-2 phosphorylated (Ser2-P) form of Pol II. All heat maps shown in Figure 1A have genes rank-ordered from highest Ser2-P Pol II enrichment within the gene to lowest, clustering active genes at the top. Expression analysis confirms that genes with elevated Ser2-P Pol II signal produced significant levels of mRNA (Figure 1A, mRNA). Notably, the most active promoters were highly enriched in NELF (e.g., Ef2b; Figure 1C; Figure S1C), suggesting that NELF is universally present during early elongation, even at the most highly expressed genes.

To determine whether NELF-bound polymerases were engaged in transcription, we evaluated RNA production from each transcription start site (TSS). We found that >85% of Pol II-bound promoters generate significant short (<100 nt) transcripts (Nechaev et al., 2010), strongly supporting the idea that Pol II pauses promoter proximally at these genes (Figure 1A; Figure S1D). Thus, the majority of *Drosophila* genes occupied by Pol II display the key hallmarks of polymerase pausing: occupancy by NELF and DSIF, promoter-proximal enrichment of Pol II signal, and the synthesis of short RNA transcripts. Notably, these findings suggest that it is not the initiation of NELF-mediated pausing, but rather the rate of pause release that is regulatory for transcription.

NELF Broadly Affects Promoter-Proximal Pol II Occupancy

To evaluate the impact of pausing on Pol II promoter occupancy, we investigated the changes in polymerase distribution upon depletion of NELF (Figure S1E). These experiments demonstrated that NELF depletion using RNA interference (RNAi) globally reduced promoter-proximal polymerase levels (Figure 1A, right panel; Figure S1F) ($p < 0.0001$). Composite Pol II profiles demonstrate that the average promoter signal is substantially reduced by NELF RNAi (Figure 1D), at both highly active and less active genes (Figure 1E). These results are consistent with widespread NELF-mediated pausing during early elongation and provide further evidence that pausing is a general step in the transcription cycle. However, although NELF RNAi widely impacts Pol II promoter occupancy, polymerase loss at individual genes varies in magnitude, suggesting that some promoters are more reliant upon NELF to achieve maximal Pol II occupancy.

We investigated why genes showed differential responses to NELF RNAi, focusing on genes bound by Pol II in untreated S2 cells. Heat maps depicting ChIP-chip signal around these promoters are shown in Figure 2A, with genes rank-ordered from most to least Pol II loss upon NELF depletion. Consistent with NELF RNAi releasing paused polymerases (Muse et al., 2007), promoters with the highest levels of promoter-proximal Pol II and NELF enrichment in untreated cells experienced the largest losses in polymerase signal upon NELF depletion

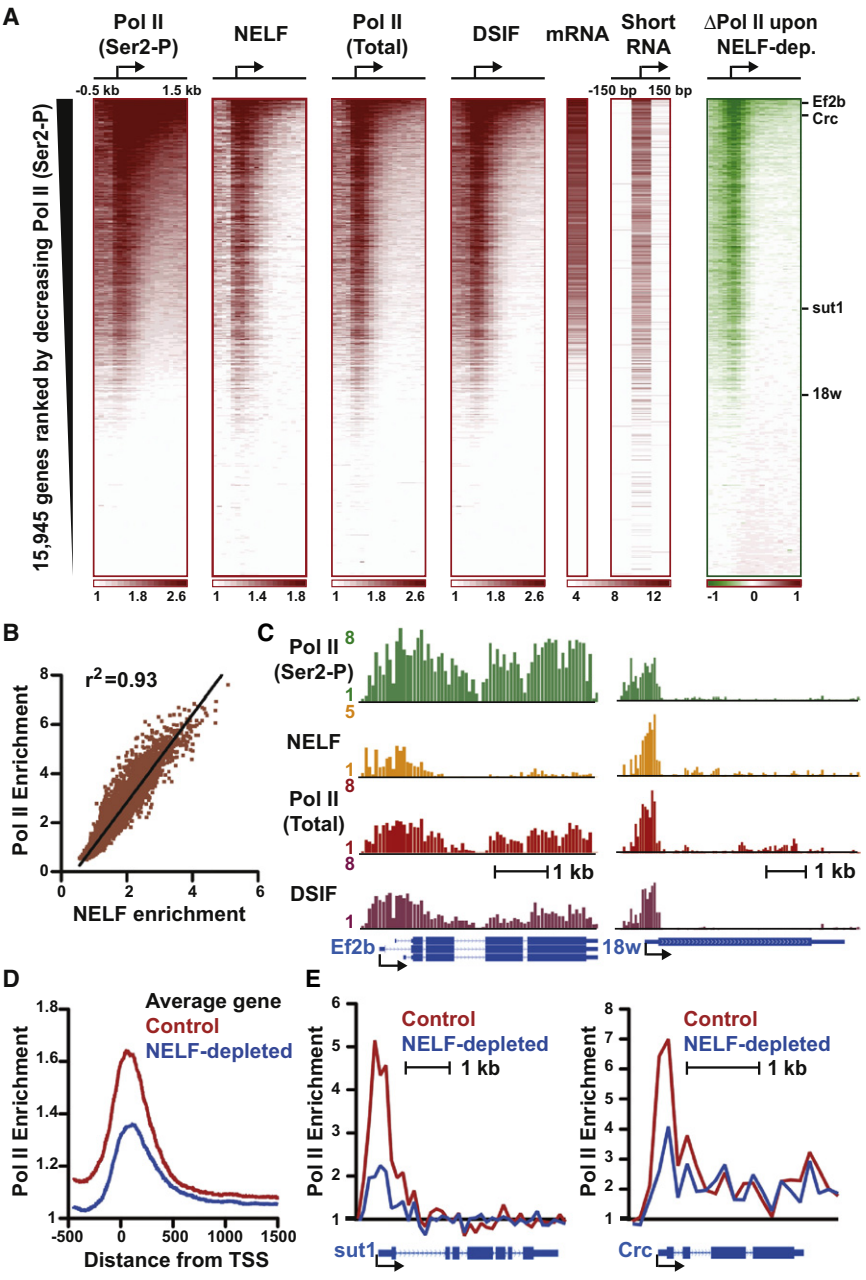


Figure 1. Pol II, NELF, and DSIF Globally Co-localize Near Promoters

(A) Average fold enrichment over genomic DNA from ChIP-chip experiments is shown in 100 bp windows surrounding *Drosophila* TSSs (shown as arrows) for actively elongating Pol II (Ser2-P), NELF (α-NELF-B), total Pol II (α-Rpb3), and DSIF (α-Spt5), with color bars at bottom indicating range. Expression levels determined by microarray (mRNA) and short RNAs derived from paused Pol II (Nechaev et al., 2010) are shown in Log_2 units. The change in Pol II signal following NELF RNAi is shown at right, as compared to control samples. Range is depicted in color bar, where red signifies gain and green indicates loss in signal.

(B) The average enrichment for total Pol II and NELF around promoters (± 250 bp) are strongly correlated.

(C) ChIP-chip data for indicated factors displayed as fold enrichment at Ef2b (CG2238), a gene with considerable elongating Pol II (left), and 18w (CG8896), a gene with little evidence of productive elongation (right). Gene models below depict exons as boxes and introns as lines.

(D) Composite Pol II distribution profiles surrounding all promoters in control and NELF-depleted cells reveal a general decrease in promoter occupancy upon NELF RNAi.

(E) NELF depletion affects Pol II promoter occupancy at genes with very little Pol II enrichment within the gene (sut1, CG8714), and with polymerase signal throughout the transcription unit (Crc, CG9429).

See also Figure S1 and Table S1.

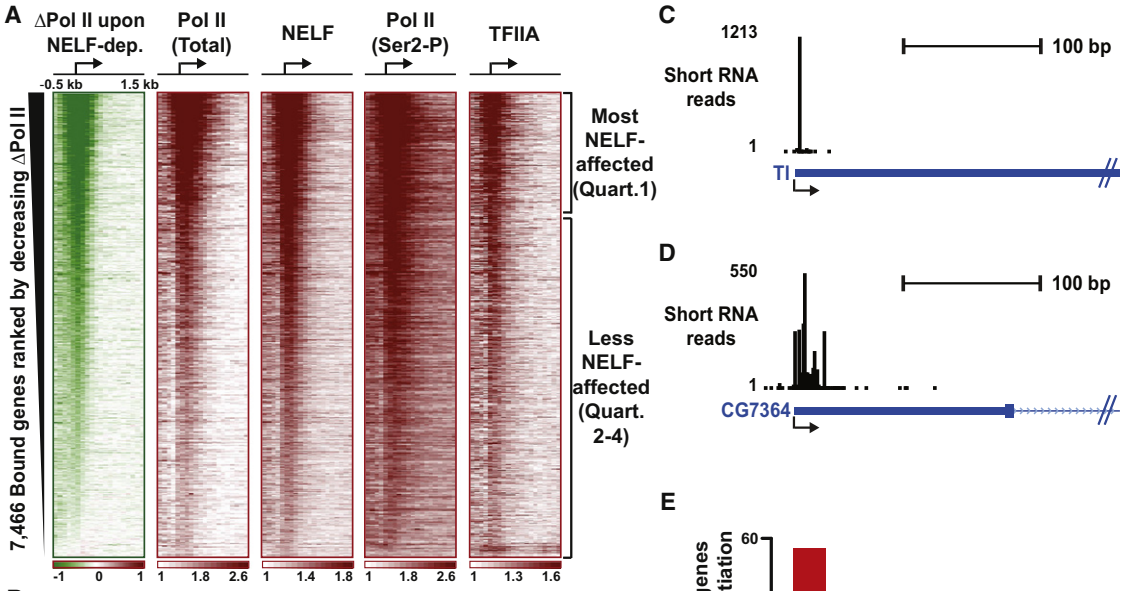
less affected by NELF (Quartiles 2–4) include housekeeping genes involved in basic cellular processes (Figure S2C).

Genes with Paused Pol II Show High Levels of Preinitiation Complex Formation and Focused Initiation

The most NELF-affected genes also displayed a distinct sequence composition near their promoters. In agreement with recent reports (Hendrix et al., 2008; Lee et al., 2008), these genes were enriched in binding sites for GAGA factor, a protein important for pausing at the heat shock genes (Shopland et al., 1995), as well as a number of well-defined promoter motifs, such as the TATA box, Initiator (Inr), and Downstream Promoter Element (DPE) (Figure 2B; Table S2). Interestingly, we found that two G+C-rich motifs that were overrepresented at the most NELF-affected genes, the DPE and Pause Button, were both located between positions +26 and +33 at these genes (Figure S3A) (Juven-Gershon et al., 2008). The precise coincidence of these sequence motifs with the peak of paused Pol II supports the idea that G+C-richness within the initially transcribed region influences elongation efficiency (Hendrix et al., 2008; Nechaev et al., 2010).

(Figure S2A). Notably, the most NELF-affected genes (Quartile 1, Figure 2A, upper bracket) were among the most active (Figure S2B), confirming that NELF-mediated pausing plays a role at active genes.

Gene ontology analysis of the most NELF-affected genes (Quartile 1) supports the idea that pausing is a favored regulatory mechanism at genes that require synchronous, precise control of expression (Muse et al., 2007; Gilchrist et al., 2008; Hendrix et al., 2008; Boettiger and Levine, 2009): these genes tend to encode highly regulated components of developmental and stimulus-responsive pathways (Figure S2C). In contrast, genes



B

Sequence element	Most NELF-affected	Less NELF-affected	P-value
Inr	719 (44.5%)	931 (19.2%)	2.4×10^{-84}
DPE	324 (20.1%)	301 (6.2%)	1.8×10^{-52}
GAGA	438 (27.1%)	594 (12.3%)	2.4×10^{-41}
Pause button	502 (31.1%)	906 (18.7%)	3.4×10^{-24}
TATA	163 (10.1%)	235 (4.8%)	5.9×10^{-13}
E-box	315 (19.5%)	1437 (29.7%)	5.3×10^{-16}
HOX RE	158 (9.8%)	866 (17.9%)	1.0×10^{-15}
DREF	395 (24.5%)	1536 (31.7%)	2.7×10^{-08}

Figure 2. Genes with Prominent NELF-Mediated Pausing Have Strong Promoters and More Focused Transcription Initiation

(A) Heat maps depict loss of Pol II signal upon NELF depletion or fold enrichment for the factors indicated at genes bound by Pol II in untreated cells. The rank order places promoter regions (± 250 bp) that lose the most Pol II signal upon NELF depletion at the top, and those least affected at bottom.

(B) Promoter motifs are enriched among the most NELF-affected genes (Quartile 1), whereas less NELF-affected genes (Quartiles 2–4) are more likely to possess activator binding sites, such as the E-box, homeo domain response element (Hox RE), or DNA-replication-related element binding factor (DREF). Pol II-bound genes with high confidence TSS annotation were analyzed (n = 6461; including 1615 of the most and 4846 of the less NELF-affected genes), and the number and percentage of genes that possess each motif are shown, along with p value (Fisher's exact test).

(C and D) Examples of genes that display highly focused transcription initiation (TI, CG5490) or more dispersed initiation patterns (CG7364). Shown are the number of short RNA 5′ ends, at single-nucleotide resolution, that map near each TSS.

(E) The most NELF-affected genes (Quartile 1) have more focused initiation than genes less affected by NELF RNAi (Quartiles 2–4). Initiation was considered focused when ≥50% of total promoter-proximal reads (± 50 bp from TSS) mapped to a single location.

See also Figures S2 and S3, and Table S2.

We found that 60% of the most NELF-affected genes possess at least one of the three core promoter motifs (TATA, Inr, and DPE), compared with only 27% of the less NELF-affected genes. Strong core promoters are thought to direct transcription initiation that is focused around a single nucleotide position (e.g., Figure 2C), whereas the absence of such motifs leads to more dispersed initiation (e.g., Figure 2D) (Juven-Gershon et al., 2008). Thus, we probed whether the observed enrichment in core promoter sequences at the most NELF-affected genes impacted the mode of transcription initiation at these genes. Mapping the 5′ ends of short capped RNAs (Nechaev et al., 2010) around the promoters of the most NELF-affected genes (Quartile 1) revealed that they experienced much more focused initiation than did less NELF-affected genes (Quartiles 2-4, Figure 2E; Figure S3B).

In agreement with the idea that highly NELF-affected genes contain intrinsically stronger promoters, we find that the general transcription factor TFIIA is significantly enriched at the most NELF-affected genes (Figure 2A and Figure S3C). Moreover, the general correspondence between occupancy by TFIIA and paused Pol II suggests that pausing may stabilize binding of general transcription factors, facilitating subsequent rounds of reinitiation at these promoters.

Conversely, weaker promoters with fewer core motifs were observed at genes that were less affected by NELF-mediated pausing, consistent with polymerase recruitment being inefficient and likely rate-limiting at these genes. Moreover, the less NELF-affected genes were enriched in binding sites for transcription activators (Figure 2B), suggesting a greater reliance on extrinsic factors for recruitment of the transcription machinery. Thus, these findings point to a relationship between promoter strength and the rate-limiting step of transcription: genes where pause release is rate-limiting have strong promoters that drive efficient recruitment of Pol II, whereas

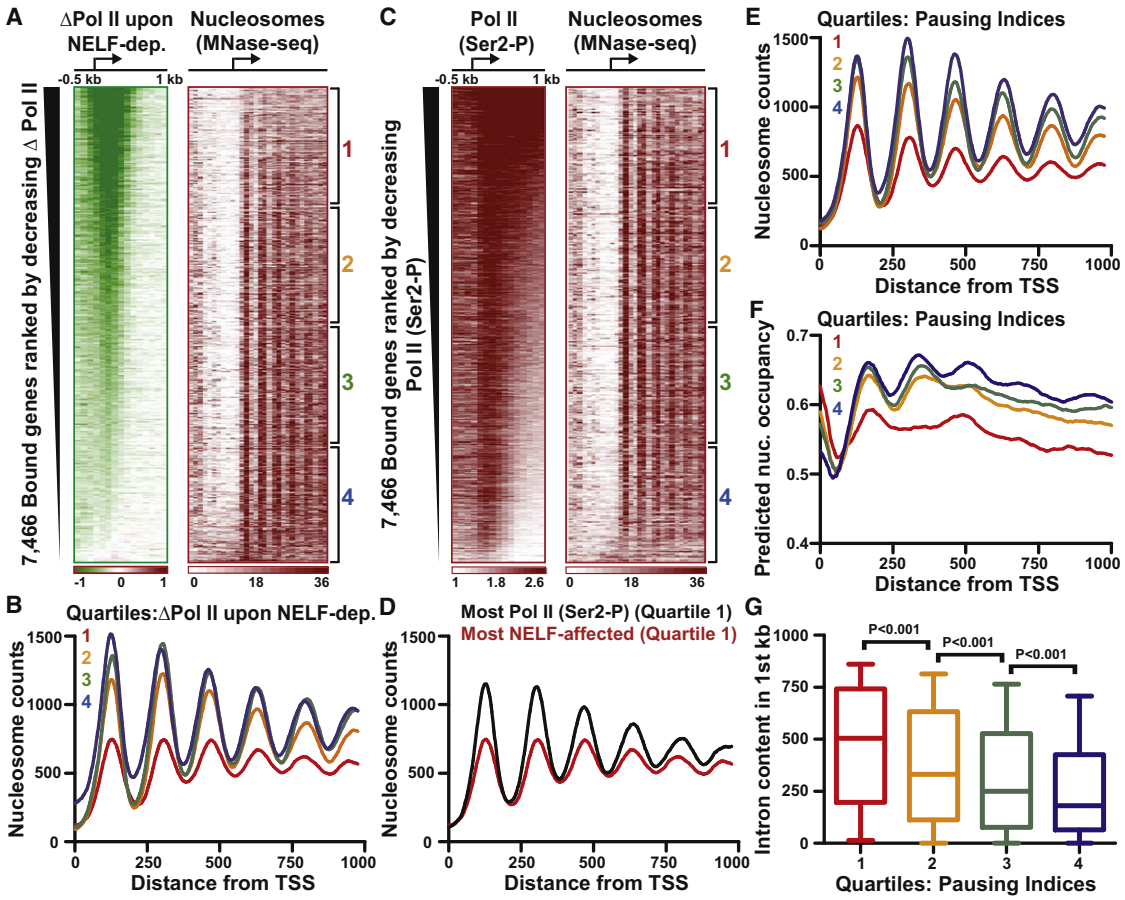


Figure 3. Nucleosomes Are Depleted Downstream of Promoters with Highly Paused Pol II

(A) The most NELF-affected genes are preferentially depleted of downstream nucleosomes. Heat maps show the change in Pol II signal upon NELF depletion for Pol II–bound genes (as in Figure 1A) and nucleosome occupancy determined by paired-end MNase-seq (color intensity indicates the number of read centers that lie in each 50 bp bin).

(B) Pol II–bound genes were divided into quartiles according to the effect of NELF depletion on Pol II promoter occupancy, from most affected (Quartile 1) to least affected (Quartile 4). Nucleosome distribution at genes in each quartile was determined by summing the number of nucleosome centers mapping to each position from the TSS to +1 kb.

(C) Transcription elongation modestly disrupts chromatin architecture. Heat maps show Ser2-P Pol II signal and nucleosome distribution at genes rank ordered by levels of Ser2-P enrichment within the gene.

(D) Nucleosome occupancy is lower downstream of the most NELF-affected genes than at genes with the most active elongation (panel C, Quartile 1).

(E) Nucleosome occupancy at genes separated into quartiles by pausing indices, where Quartile 1 represents genes with the most pausing.

(F) Predicted nucleosome occupancy at genes in each quartile of pausing indices, based on intrinsic DNA sequence preferences of nucleosome formation (Kaplan et al., 2009).

(G) Intron content is shown for genes in each quartile of pausing indices, revealing significantly elevated intron levels at genes that are highly affected by NELF depletion (Kruskal-Wallis test, boxes depict 25th through 75th percentiles, whiskers show 10th through 90th percentiles).

See also Figure S4.

recruitment-limited genes have weaker promoters and depend on additional factors for their activation.

Pol II Pausing Is Linked to Nucleosome Deprivation Downstream of the TSS

We next mapped nucleosomes across the *Drosophila* genome using micrococall nuclease (MNase) digestion of chromatin followed by high-throughput paired-end sequencing. We achieved >30-fold coverage of the genome (assuming one nucleosome every 200 bp), with 32.5 million reads that mapped uniquely to the 170 megabase genome. The distribution of these reads around TSSs of Pol II-bound genes is shown in Figure 3A as the number of read centers that mapped to each 50-bp bin. These data confirm that Pol II-occupied *Drosophila* promoters display a nucleosome-deprived region around the TSS (Mavrich et al., 2008). However, the most and least NELF-affected genes differed substantially in their nucleosome distributions downstream of the TSS.

Genes that were less affected by NELF depletion (Quartiles 2–4) exhibit a canonical, well-organized nucleosome

architecture with a clear periodicity (Figure 3A, ~170 bp internucleosomal spacing). In contrast, the most NELF-affected genes (Quartile 1) show lower nucleosome occupancy and less organized chromatin structure. Composite metagene analysis of nucleosome distribution showed that the most NELF-affected genes contain far fewer nucleosomes within the initially transcribed region than genes less impacted by NELF RNAi (Figure 3B).

Pol II disrupts nucleosomes as it transcribes, and the considerable levels of Ser2-P Pol II detected at the most NELF-affected genes raised the possibility that the observed nucleosome deprivation could result from polymerase elongation. To address this issue, we analyzed nucleosome occupancy at Pol II–bound genes when ordered by descending levels of active elongation (Ser2-P Pol II signal; Figure 3C; Figure S4A). If polymerase elongation were largely responsible for low nucleosome occupancy, then the most actively transcribed genes should be particularly depleted of nucleosomes. In contrast, despite having much higher levels of Ser2-P Pol II signal (Figure S4B), genes with the most active elongation exhibit higher nucleosome density than the most NELF-affected genes (Figure 3D), indicating that Pol II elongation is not the dominant cause of nucleosome disruption within NELF-affected genes.

Nucleosome Depletion at Paused Genes Argues against a Role for Nucleosomes in Establishing Paused Pol II

To further probe the link between paused Pol II and promoter-proximal nucleosome organization, we investigated nucleosome distributions at genes with varying levels of pausing, as judged by their "pausing index," which was calculated as the ratio of the Pol II signal near promoters (TSS ± 250 bp) to the downstream region (+500 bp to the end of the gene), as described in Muse et al. (2007). Higher ratios reflect greater promoter-proximal enrichment of polymerase, and thus genes with the highest pausing indices (Quartile 1) display the most paused Pol II. Consistent with our analysis of NELF-affected genes, the most paused genes show the lowest nucleosome occupancy within the initially transcribed region (Figure 3E; Figure S4C).

Pol II-bound genes containing TATA, Inr, or PB/DPE motifs also show reduced nucleosome density downstream of the TSSs relative to the average bound gene (Figure S4D), consistent with recent reports suggesting that these motifs are associated with diminished nucleosome organization (Albert et al., 2007; Mavrich et al., 2008). However, genes with GAGA elements showed the lowest average nucleosome occupancy, similar to that at the most NELF-affected genes (Figure S4D). These data are consistent with the known role of GAGA factor in recruiting chromatin remodeling complexes (Tsukiyama et al., 1994) and suggest that GAGA binding broadly leads to histone eviction. Notably, the presence of GAGA-binding sites at the most NELF-affected genes corresponded to a dramatic depletion of promoter-proximal nucleosomes (Figure S4E), indicating that many of these genes, like the *Drosophila* heat shock genes (Wu, 1980), are effectively nucleosome-free within the initially transcribed region. This finding argues strongly against recent suggestions that nucleosomes cause pausing by imposing a stable barrier to elongation (Schones et al., 2008; Mavrich et al., 2008). In contrast, we find higher promoter-proximal nucleosome occupancy at genes that display less pausing, implying that the presence of nucleosomes is unlikely to establish paused Pol II.

Nucleosome Occupancy Is Intrinsically Disfavored Downstream of Paused Promoters

To evaluate the role of DNA sequence in establishing different chromatin structures, we determined the favored positions for nucleosome occupancy around *Drosophila* promoters using algorithms based on inherent sequence preferences for nucleosome formation (Kaplan et al., 2009). Surprisingly, these analyses revealed that sequences downstream of the most highly paused promoters intrinsically disfavor nucleosome occupancy (Figure 3F, Quartile 1), suggesting that the nucleosome depletion observed at these genes is specified by their DNA sequence (compare Figure 3F; Figure S4C).

Notably, introns are enriched in nucleosome-disfavoring sequences and have lower nucleosome occupancy than exons in vivo (Schwartz et al., 2009). Highly regulated *Drosophila* genes, and in particular those involved in development, are known to possess long introns, leading us to investigate whether an elevated intron content downstream of highly paused genes might contribute to their nucleosome depletion. Indeed, genes with the highest pausing indices had significantly higher intron content within the first 1 kb than did less paused genes (Figure 3G). These intriguing results suggest that introns may serve a role in deterring nucleosome formation at highly paused genes, helping to establish distinct downstream chromatin architectures.

Promoters of Highly Paused Genes Favor Nucleosome Assembly

Consistent with prior work (Mito et al., 2005; Ozsolak et al., 2007; Mavrich et al., 2008; Schones et al., 2008), we found that Pol II–bound promoters are generally depleted of nucleosomes (Figure 4A). However, it remained unclear whether this depletion is entirely caused by the presence of Pol II, or whether sequence preferences for nucleosome formation contribute as well (Tillo et al., 2010). To address this question, we determined the predicted nucleosome occupancy around promoters in each pausing index quartile. Strikingly, genes with the highest pausing indices (Figure 4B, Quartile 1) contain promoters that intrinsically favor assembly of a nucleosome over the TSS, whereas this tendency is diminished at genes with less pausing (Quartiles 2–4). This result suggests that highly paused promoters encode an inherently repressive chromatin structure that is counteracted by pausing of Pol II. In contrast, less paused promoters may not require Pol II pausing to deter promoter nucleosome formation, instead possessing sequences that disfavor nucleosome assembly.

We showed previously that loss of paused Pol II upon NELF depletion leads to increased nucleosome occupancy and down-regulation of gene expression at several highly paused promoters (Gilchrist et al., 2008). The data shown in Figure 4B suggest that increased nucleosome occupancy at these genes is driven by sequences that favor nucleosome assembly. To test this model on a global scale, we mapped nucleosomes in NELF-depleted and mock-RNAi treated cells using MNase-seq. Figure 4C shows one example of a highly paused gene

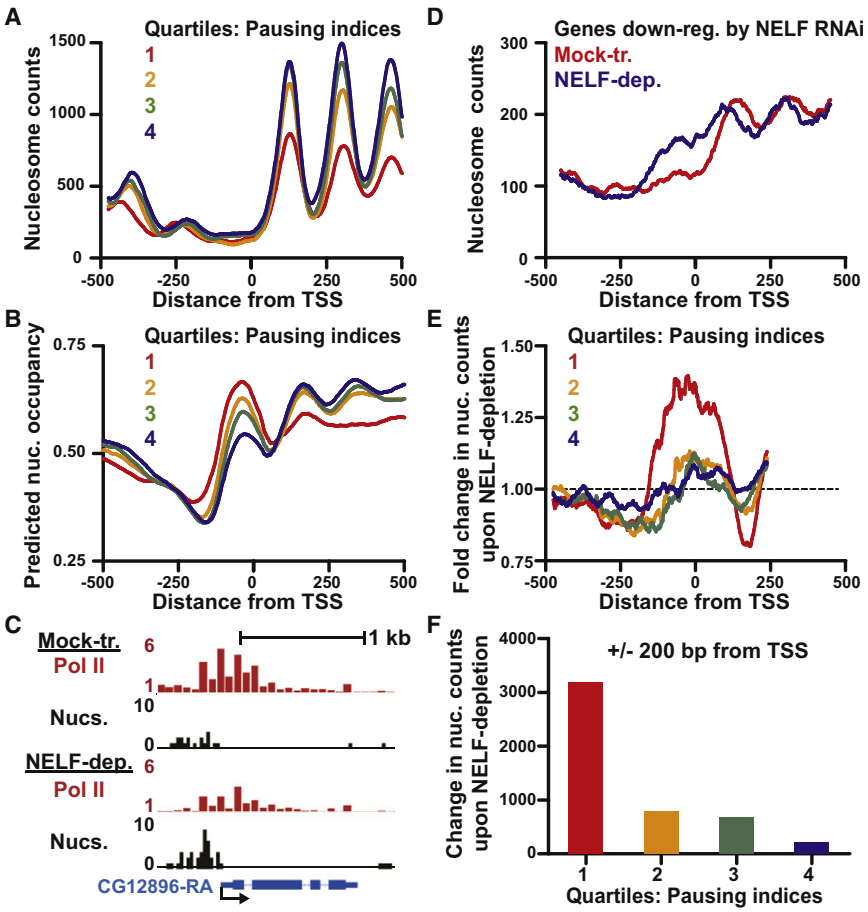


Figure 4. DNA Sequences at Paused Genes Favor High Promoter Nucleosome Occupancy

(A) Nucleosome occupancy at Pol II–bound genes (n = 7466) separated into quartiles by pausing indices, where Quartile 1 represents the most paused genes. Nucleosome occupancy at genes in each quartile was determined by summing the number of nucleosome centers mapping to each position.

(B) Predicted nucleosome occupancy at genes in each quartile of pausing indices, based on intrinsic DNA sequence preferences of nucleosome formation, as in Kaplan et al. (2009).

(C) Loss of Pol II upon NELF depletion is accompanied by increased nucleosome occupancy. Pol II ChIP-chip fold enrichment (red) and MNase-seq read distribution (black, depicts read centers in 25-bp bins) around a highly NELF-affected gene (CG12896) in mock-treated and NELF-depleted samples.

(D) Genes downregulated by NELF depletion show increased promoter nucleosome occupancy. Nucleosome occupancy (calculated as in A) at genes whose expression decreased >2-fold following NELF depletion.

(E and F) NELF depletion leads to increased nucleosome occupancy over highly paused promoters. The change in nucleosome counts upon NELF depletion (MNase-seq reads in NELF-depleted/mock-treated samples) is shown for genes in each quartile of pausing indices as fold change in read number at each position (E) or the raw increase in the number of nucleosome reads ± 200 bp from the TSS (F).

See also Table S3.

with low promoter nucleosome occupancy in mock-treated cells. Depletion of NELF results in a reduction in Pol II promoter signal, and an accompanying increase in promoter-proximal nucleosome levels. This finding can be extended broadly to genes whose expression is downregulated following NELF RNAi (>2-fold change; see microarray expression data in Table S3), which show increased promoter nucleosome occupancy in NELF-depleted cells (Figure 4D). Interestingly, these genes also show a shift in nucleosome position, with downstream nucleosomes moving toward the promoter following NELF depletion, implying a dynamic relationship between Pol II and nucleosome binding at these promoters.

Furthermore, the increase in nucleosome occupancy over the TSS following NELF RNAi is a general feature of highly paused genes. Comparing the nucleosome levels in NELF-depleted versus mock-treated cells revealed a considerable increase in nucleosome occupancy surrounding promoters of the most paused genes (Figures 4E and 4F, Quartile 1). In contrast, NELF RNAi resulted in much smaller changes in nucleosome occupancy at genes with less paused Pol II (Figure 4F). These results demonstrate that NELF-mediated pausing inhibits nucleosome occupancy of the most highly paused promoters and that loss of pausing allows these genes to assume the default nucleosome organization specified by the underlying DNA sequence.

Pol II Binding Inhibits Nucleosome Occupancy at the Most Paused Genes

We further investigated nucleosome architecture around highly paused versus less paused promoters by comparing Pol II and nucleosome occupancy at individual genes in their repressed and activated states. To accomplish this, we took advantage of the fact that a 24-hr treatment of *Drosophila* cells with the steroid hormone ecdysone causes marked changes in gene expression (Dimarcq et al., 1997). Pol II ChIP-chip was performed with and without ecdysone treatment to identify genes that transitioned between Pol II–bound and unbound states (or vice versa) during this treatment. We then used quantitative PCR on MNase-digested chromatin to investigate changes in nucleosome occupancy that accompanied these Pol II transitions, focusing on genes that were highly paused (Quartile 1) or less paused (Quartiles 3 or 4) in the active state.

We found that highly paused genes had nucleosome-occluded promoters in the absence of Pol II and that polymerase binding substantially reduced nucleosome levels at these genes (Figures 5A and 5B; Figure S5). In contrast, genes lacking paused Pol II (Figures 5C and 5D) were generally depleted of nucleosomes, even in the unbound state. These data further support the notion that sequences around highly paused promoters specifically favor nucleosome assembly and that paused Pol II prevents nucleosome formation around these promoters.

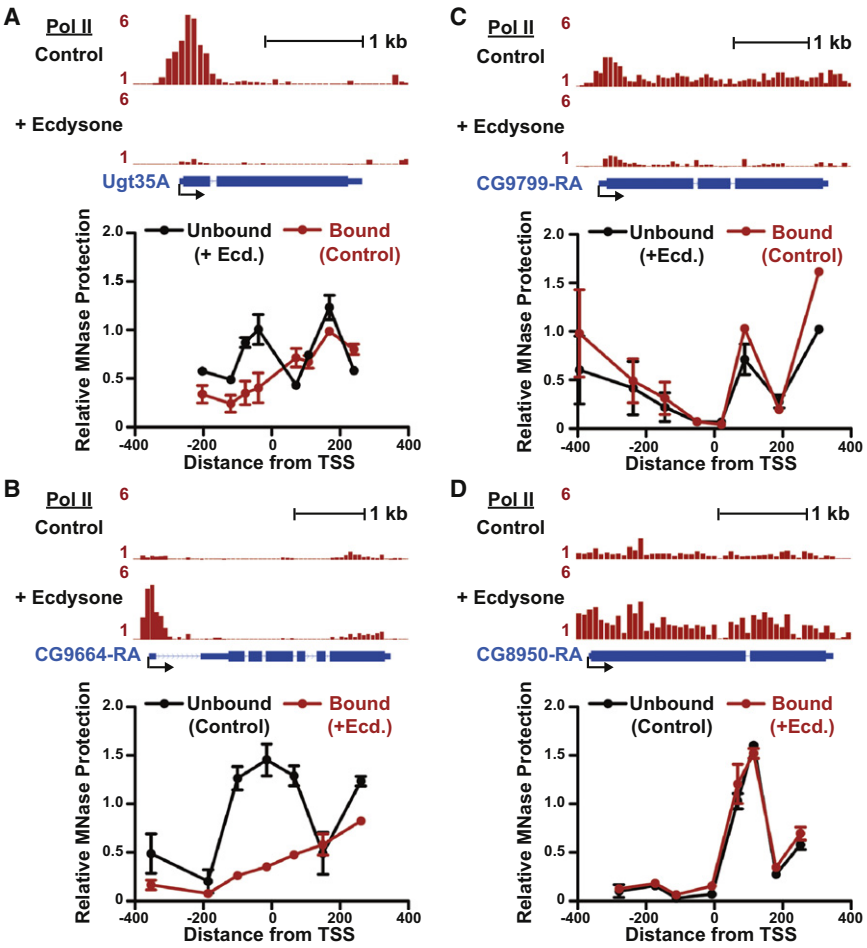


Figure 5. The Relationship between Pol II and Nucleosome Occupancy at Highly Paused and Less Paused Genes

Pol II distribution in S2 cells ± 24 hr treatment with ecdysone is depicted as fold enrichment from ChIP-chip experiments. MNase protection assays are shown below to compare nucleosome occupancy at each gene in the Pol II–bound versus unbound state. 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.

(A) Ugt35A (CG6644), a gene with a high pausing index in control cells that becomes unbound by Pol II following treatment with ecdysone.

(B) CG9664, a gene unbound by Pol II in control cells that becomes highly paused in ecdysone-treated cells.

(C) Ecdysone causes CG9799, a gene with a low pausing index in control cells, to become unbound by Pol II.

(D) CG8950, an unbound gene in control cells, has uniform Pol II distribution upon ecdysone treatment.

See also Figure S5.

Genes Adopt Their Predicted Nucleosome Organization in the Absence of Pol II Binding

The above data suggest that sequences around the most highly paused genes in *Drosophila* S2 cells inherently favor nucleosome occupancy that is high over promoters and lower downstream. To ascertain whether these characteristics would be conserved in a different context, we determined Pol II distribution in *Drosophila* 0–16-hr-old embryos and compared this to embryo nucleosome occupancy reported previously (Mavrich et al., 2008). Figure 6A displays these data, with genes rank-ordered by descending pausing index in embryos. Importantly, we found that nucleosome depletion downstream of the most highly paused promoters is not limited to S2 cells but is maintained in developing embryos (Figure 6B). Likewise, calculation of predicted nucleosome occupancies for genes in each pausing index quartile in embryos corroborated data from S2 cells (Figure 6C; Figure S6A): the most highly paused genes in embryos exhibited higher predicted nucleosome occupancy over the promoter than within the gene (Figure 6C, Quartile 1), whereas genes with the least Pol II pausing (Quartile 4) favored higher nucleosome occupancies downstream.

These experiments also identified many promoters that showed high pausing indices in embryos but were unoccupied by Pol II in S2 cells (e.g., Figure 6D; Figure S6B). If our model is correct, then these promoters should be depleted of nucleosomes in the presence of paused Pol II but occluded by nucleosomes in the absence of polymerase binding. To test this idea, we analyzed nucleosome distribution at genes that were highly paused in embryos (Quartile 1) but were unbound by Pol II in S2 cells. In support of our model, these genes contain a nucleosome positioned directly over the promoter in the Pol II–unbound state (S2 cells, Figure 6E), and this nucleosome is displaced in the presence of paused Pol II (embryos, Figure S6C). Notably, although Pol II binding substantially decreases promoter-proximal nucleosome occupancy at these genes, it only modestly affects downstream nucleosome levels (Figure S6C). In contrast, promoters with the lowest pausing indices in embryos were generally nucleosome deprived, regardless of Pol II occupancy (Figure 6E; Figure S6D), and had higher nucleosome density within the gene, consistent with sequence-based predictions.

Taken together, our data demonstrate that genes generally assume their sequence-predicted nucleosome architecture in the absence of the transcription machinery. That genes with different levels of pausing possess such distinct default states suggests that there is a fundamental relationship between intrinsic chromatin structure and gene regulatory strategies.

DISCUSSION

Our data support a general model for gene regulation wherein the underlying DNA sequence around promoters directly

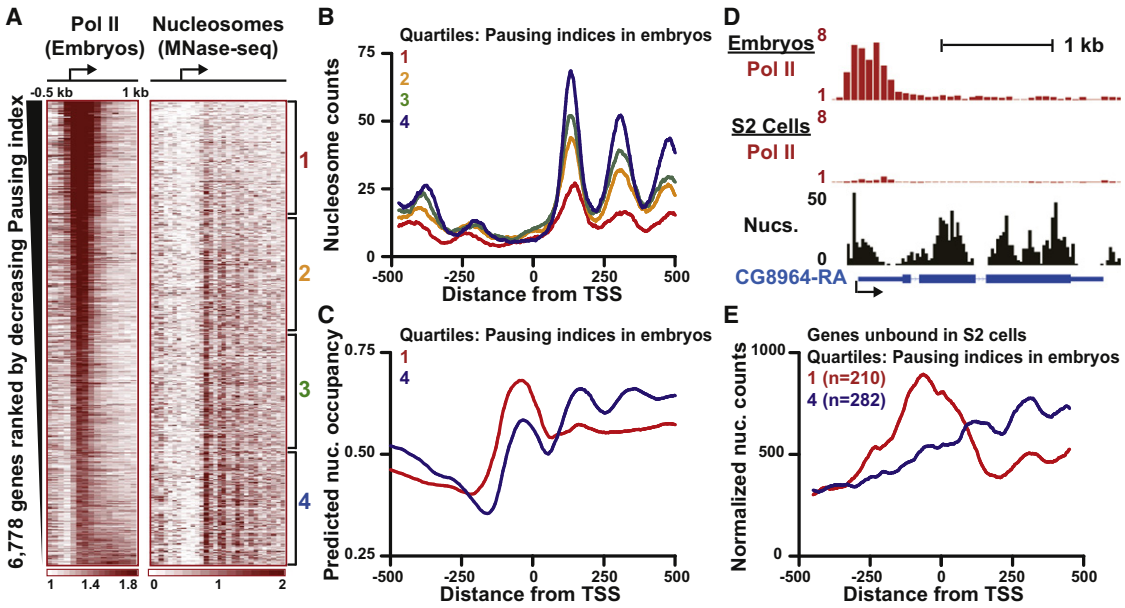


Figure 6. Promoters Adopt Their Predicted Nucleosome Configuration in the Absence of Paused Pol II

(A) Pol II distribution and nucleosome occupancy (H2A.Z nucleosomes from Mavrich et al., 2008) in *Drosophila* embryos. Genes are rank ordered by descending pausing indices in embryos.

(B) The most highly paused genes in embryos are depleted of downstream nucleosomes. Pol II–bound genes in embryos were divided into quartiles on the basis of pausing indices, and composite metagene analyses of nucleosome reads around their promoters were generated from the data in Mavrich et al. (2008).

(C) Predicted nucleosome occupancy at genes in each quartile of pausing indices as determined in *Drosophila* embryos.

(D) A promoter with a high pausing index in embryos (top panel) that is not occupied by Pol II in S2 cells (middle panel) becomes occluded by nucleosomes in the unbound state. Bottom panel shows nucleosome occupancy at the unbound gene in S2 cells as determined by MNase-seq, with read centers displayed in 25-bp bins.

(E) In the absence of Pol II, in vivo nucleosome occupancy closely resembles predictions. Nucleosome occupancies were determined in S2 cells, where these genes are not bound by Pol II. Shown are genes that are highly paused (Quartile 1) or lacking paused Pol II (Quartile 4) in embryos.

See also Figure S6.

influences both chromatin architecture and the step in the transcription cycle that is rate limiting for gene expression. We find that genes with high levels of Pol II pausing (Figure 7A) inherently favor the formation of nucleosomes over the promoter, establishing an active competition between Pol II and nucleosomes for promoter occupancy. We propose that this intrinsically repressive chromatin structure prevents aberrant expression of paused genes, which are often components of highly regulated pathways. Nucleosome remodeling, which likely is initiated by proteins such as GAGA factor, would be required to disassemble nucleosomes at these promoters and allow for gene activity (Figure 7A, small red arrow). Nucleosome removal would uncover strong promoter motifs that facilitate efficient, stable recruitment of the transcription machinery (Figure 7A, large green arrow). Extended NELF-mediated pausing of polymerase at these promoters makes the transition to productive elongation slow (Figure 7A, small red arrow). However, upon pause release, low levels of downstream nucleosomes would minimize barriers to transcription elongation, and additional Pol II molecules would be rapidly recruited to maintain high Pol II occupancy and prevent nucleosome formation.

In contrast, genes that lack extended pausing (Figure 7B) appear to disfavor promoter nucleosome assembly and instead harbor nucleosomes flanking the nucleosome-deprived promoter region. Localized DNA accessibility near TSSs could both help target the transcription machinery to the promoter region and diminish the requirement for nucleosome remodeling to allow gene activity. The dearth of core promoter elements could make these genes more reliant on activator binding for recruitment of the transcription machinery, and Pol II recruitment would be the rate-limiting step for expression of these genes (Figure 7B, small red arrow). Pausing would be short-lived at these genes, and despite higher downstream nucleosome occupancy, polymerase escapes efficiently into productive synthesis.

Importantly, these two strategies present different opportunities for gene regulation. Highly paused genes present two distinct steps at which they can be regulated: promoter accessibility and release of Pol II from pausing. We propose that this two-step mechanism facilitates precise control of gene expression. We envision that the first step, nucleosome remodeling, functions as a molecular switch that relieves repression by chromatin to permit expression. This step can be temporally uncoupled from gene activation and could potentiate genes for future activation rather than prompting their immediate expression. The second step, release of paused Pol II, might be analogous to a volume dial, which permits fine-tuning of expression levels in response to changing conditions. Transcription levels could be rapidly regulated solely by manipulating the efficiency

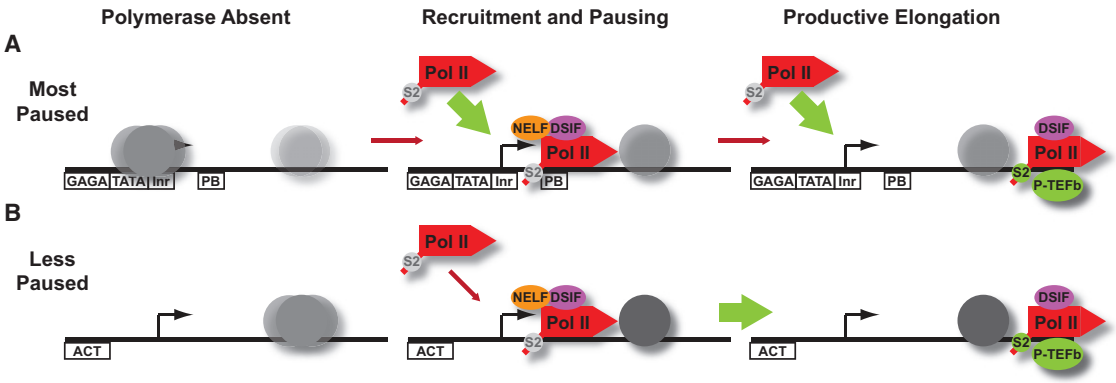


Figure 7. Different Default Chromatin Architectures Specify Distinct Gene Regulatory Strategies

(A) The most paused promoters are inherently occluded by nucleosomes (shown as gray ovals, where color intensity denotes occupancy levels) prior to Pol II binding. Regulated chromatin remodeling (red arrow) can expose strong promoter motifs (shown as boxes below DNA) that allow for efficient Pol II recruitment (green arrow). Subsequent recruitment of P-TEFb and pause release are also regulated at these genes (second red arrow), providing an additional opportunity for gene regulation.

(B) Less paused genes display weaker, nucleosome-deprived promoter regions. Polymerase recruitment is rate-limiting at these genes (red arrow), and perhaps more dependent on activators (ACT, binding site shown as box). Pol II is bound by DSIF and NELF at these genes, but pausing is transient and the polymerase moves efficiently into the gene (green arrow).

of P-TEFb recruitment through its interactions with DNA-binding transcription activators and histone modifications (Peterlin and Price, 2006; Rahl et al., 2010). This idea is supported by observations that activation of highly paused genes is both fast and synchronous (Lis, 1998; Boettiger and Levine, 2009). In contrast, genes that lack promoter-proximal pausing and nucleosome occupancy rely chiefly on a single-step mechanism to alter gene expression: regulated, step-wise recruitment of the transcription machinery. This mode of regulation has been suggested to be inherently more stochastic and prone to transcriptional noise (Boettiger and Levine, 2009), which may explain why many genes regulated by recruitment are constitutively active housekeeping genes.

We provide evidence that NELF-mediated pausing during early elongation is a general feature of the transcription cycle that is exploited at some genes to regulate transcription output. We propose that each round of transcription entails pausing, perhaps serving as an early "checkpoint" to ensure proper maturation of the elongation complex before release into productive elongation. At some genes, this halt in elongation may be transient, whereas at others it may involve a long-lived paused complex that becomes rate-limiting for gene expression. Importantly, these results imply that the release from pausing through P-TEFb recruitment is an important, regulated step that broadly impacts gene expression, in agreement with recent work (Peterlin and Price, 2006; Rahl et al., 2010). We note that general recruitment of NELF during early elongation likely explains the seemingly paradoxical observation made in several systems that NELF levels increase at activated genes that experience robust recruitment of additional Pol II.

Our data also reveal that the inherent preference toward repression of highly regulated promoters by nucleosome occlusion is an evolutionarily conserved phenomenon (Tirosh and Barkai, 2008). Moreover, our results are in agreement with recent work in yeast revealing that Pol II plays a role in displacing nucleosomes from promoter regions (Weiner et al., 2009). However, in yeast, nucleosome disassembly is coupled directly to gene activation, whereas in *Drosophila* nucleosome disassembly is coupled to Pol II pausing. Perhaps *Drosophila* and other metazoans have evolved promoter-proximal pausing as an additional layer of regulation to accommodate increased demands for precise and rapid gene regulation during organism development and responses to stress. In addition, it might be beneficial to maintain highly regulated promoters poised in an open chromatin state, to prevent their incorporation into the more inaccessible, condensed heterochromatin that exists in metazoans.

In summary, we report that a primary function of paused Pol II is to prevent promoter-proximal nucleosome formation. This represents a fundamental shift in our thinking about the role of Pol II pausing, which has long been thought to simply repress gene expression. Instead, we argue that pausing should be viewed as a mechanism to fine-tune gene expression, and to potentiate genes for further or future activation. In addition, we have shown that sequence-specified default nucleosome architecture instructs the regulatory properties of *Drosophila* promoters. We propose that metazoans have evolved a gene regulatory strategy in which nucleosomes and paused Pol II compete for promoter occupancy, affording multiple opportunities for regulation of gene expression.

EXPERIMENTAL PROCEDURES

ChIP-Chip Experiments

Untreated and RNAi-treated *Drosophila* S2 cells were cross-linked, and DNA was immunoprecipitated, amplified, and labeled for ChIP-chip, as described elsewhere (Gilchrist et al., 2009). NimbleGen tiling arrays that span the *Drosophila* genome (2.1 million probes) were probed according to manufacturer's instructions. Data shown represent average probe signals from at least two biological replicates. Antibodies and detailed methods are described in Extended Experimental Procedures.

Defining Nucleosome Positions by MNase-seq

MNase-digested chromatin from untreated, mock-treated, and NELF-depleted S2 cells was prepared as described in Gilchrist et al. (2008), except that 200 μl of chromatin was digested with 20 units of MNase (Worthington) for 45 min at 25°C. Following gel purification, mono-nucleosome sized fragments (100–200 bp) were subjected to sequencing using the Illumina paired-end protocol. The resulting data set from untreated samples included 32.5 million unique read pairs identifying both ends of fragments $\geq$ 120 bp and $\leq$ 180 bp in length, whereas 11.2 million read pairs were obtained from each of the RNAi-treated samples.

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 as described elsewhere (Kaplan et al., 2009). Genomic position average occupancy values were placed in gene context relative to TSSs using custom scripts. The resulting predictions of nucleosome occupancy with respect to individual TSSs were used to generate metagene analyses of predicted nucleosome occupancy for select groups of genes as noted in the text.

ACCESSION NUMBERS

Genomic data described in this work have been deposited in GEO under accession number GSE20472.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and three tables and can be found with this article online at doi: 10.1016/j.cell.2010.10.004.

ACKNOWLEDGMENTS

We thank S. Nechaev, T. Kunkel, and G. Hu for critical reading of the manuscript. We acknowledge L. Pederson for production of the NELF-B protein, J. Tucker and the NIEHS microarray core for help with arrays, and S. Dai and J. Grovenstein for computational support. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (grant Z01 ES101987 to K.A. and grant ES101765 to L.L.). D.A.G. and K.A. designed experiments, D.A.G., G.D.S., B.X., and Y.G. performed experiments, D.A.G., D.C.F., L.L., and K.A. performed data analysis, and D.A.G. and K.A. prepared the manuscript.

Received: May 21, 2010
Revised: August 18, 2010
Accepted: September 27, 2010
Published: November 11, 2010

REFERENCES

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature *446*, 572–576.

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

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. Science *325*, 471–473.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193–198.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J. Biol. Chem. *282*, 21901–21912.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Dimarcq, J.L., Imler, J.L., Lanot, R., Ezekowitz, R.A., Hoffmann, J.A., Janeway, C.A., and Lagueux, M. (1997). Treatment of l(2)mbn *Drosophila* tumorous blood cells with the steroid hormone ecdysone amplifies the inducibility of antimicrobial peptide gene expression. Insect Biochem. Mol. Biol. *27*, 877–886.

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 *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 Dev. *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. Proc. Natl. Acad. Sci. USA *105*, 7762–7767.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. *14*, 2570–2579.

Izban, M.G., and Luse, D.S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J. Biol. Chem. *267*, 13647–13655.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter—the gateway to transcription. Curr. Opin. Cell Biol. *20*, 253–259.

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.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876–880.

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*. Mol. Cell. Biol. *28*, 3290–3300.

Lis, J. (1998). Promoter-associated pausing in promoter architecture and post-initiation transcriptional regulation. Cold Spring Harb. Symp. Quant. Biol. *63*, 347–356.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol. Cell. Biol. *12*, 2078–2090.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. *270*, 12335–12338.

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.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090–1097.

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. Nat. Genet. *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 *327*, 335–338.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nat. Biotechnol. *25*, 244–248.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. *579*, 909–915.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. Cell *54*, 795–804.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. *16*, 990–995.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol. Cell *18*, 735–748.

Shopland, L.S., Hirayoshi, K., Fernandes, M., and Lis, J.T. (1995). HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev. *9*, 2756–2769.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High nucleosome occupancy is encoded at human regulatory sequences. PLoS ONE *5*, e9129.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res. *18*, 1084–1091.

Tsukiyama, T., Becker, P.B., and Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature *367*, 525–532.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2009). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. *20*, 90–100.

Wu, C. (1980). The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. Genes Dev. *17*, 1402–1414.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol. Cell. Biol. *22*, 2918–2927.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science *309*, 626–630.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. Nat. Genet. *39*, 1512–1516.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847–852.

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 hr 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 hr 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 the Tjian 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 heat maps 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 min 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 $\geq$ 120 bp and $\leq$ 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 $\geq$ 120 bp and $\leq$ 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 hr, and 200 μl chromatin was digested with 20 units MNase for 10 min 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 heat maps. 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 ($\geq$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 versus 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 Heat Maps

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. Heat maps 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 90th percentile enrichment value for each data set around promoters (± 250 bp).

The heat maps 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 heat map 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 heat map 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 heat maps 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 heat map where red represents gain and green loss of Pol II ChIP enrichment.

The MNase-seq heat maps 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 heat map 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, Figure 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 versus 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 (1171 genes) were excluded from analysis of Pausing index, including 337 bound genes (leaving 7466 bound genes of sufficient length for analysis). After calculation of Pausing indices, the gene list was separated into quartiles, with the top quartile (1866 genes, or ~11% of all 17,116 genes) displaying Pausing indices > 2.155.

The heat maps in Figures 2 and 3 display these 7466 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 hr 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 7110 promoters being called bound by polymerase, 6778 of which had probes downstream of +500 and so were used for calculation of Pausing indices. The heat map in Figure 5 displays these 6778 genes rank ordered by Pausing index.

In the absence of ecdysone 8055 genes were considered bound and 7661 considered for Pausing index calculation, and in the presence of ecdysone, 5641 genes were observed to be bound, and 5443 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 = 1866 genes), and this is referred to as "Normalized nucleosome counts."

Data involving small numbers of genes (Figure 4E and Figures S4D and S4E) were smoothed with 50 bp moving averages, and Figures 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 hr 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 7466 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 (number of nucleosome centers)$_{\text{NELF-depleted}}$ − (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 7466 genes used to calculate Pausing indices.

For identification of sequence motifs we used a list of 6461 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 $\leq$ 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 overrepresented 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 7422 of the 7466 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 Dev. *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 Biol. *4*, 3.

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 Dev. *22*, 1921–1933.

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 *48*, 398–408.

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. Proc. Natl. Acad. Sci. USA *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 Res. *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. Nat. Protoc. *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 Biol. *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. Mol. Cell. Biol. *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. Nat. Genet. *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 *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. Nat. Protoc. *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 Dev. *7*, 2235–2245.

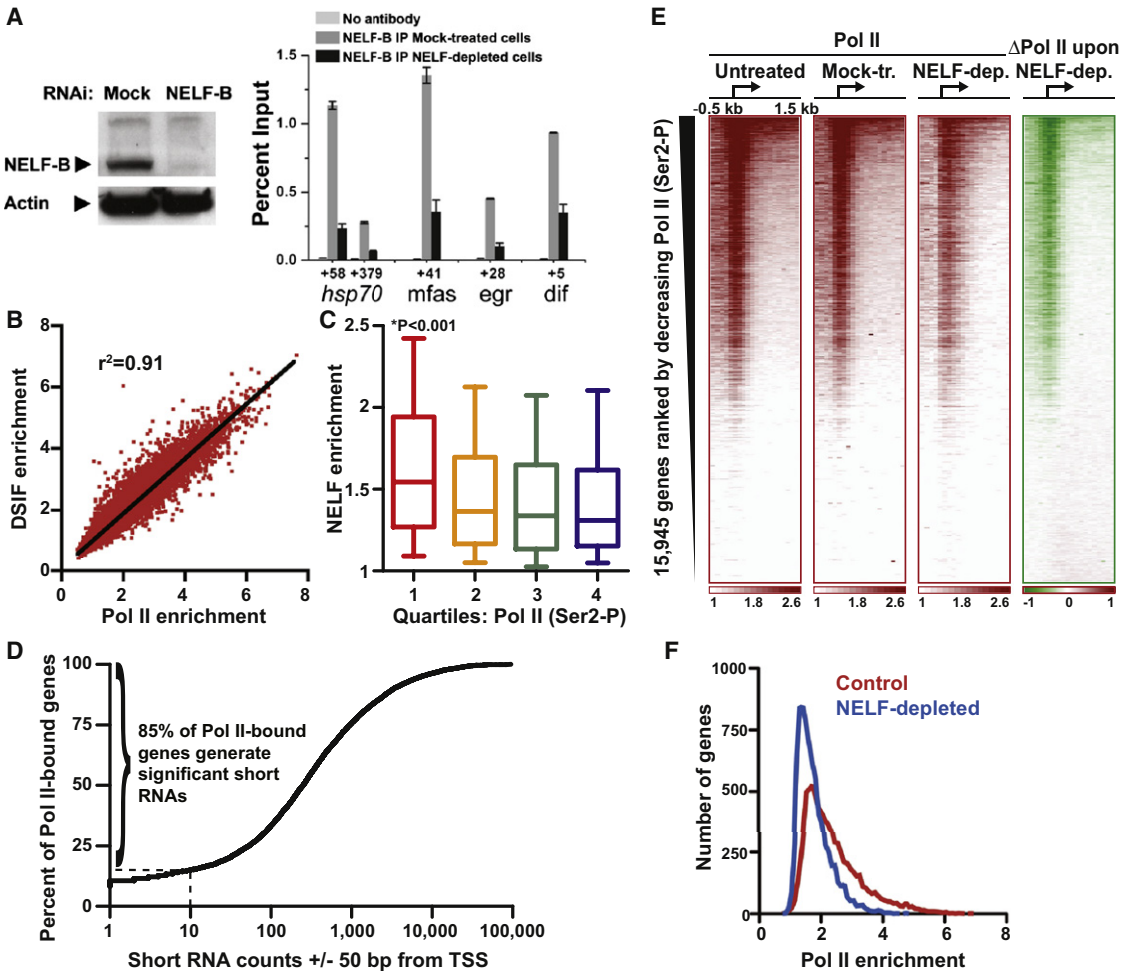


Figure S1. NELF-Mediated Pausing Is a Widespread Phenomenon, Related to Figure 1

(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.

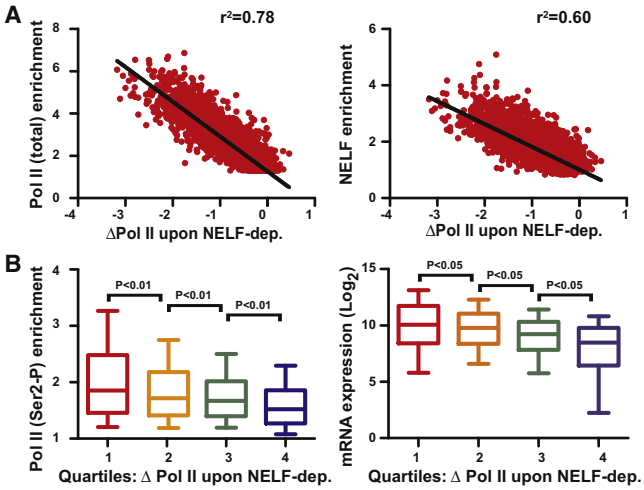
(B) The average total Pol II signal around each of the 17,116 unique promoters (± 250 bp) corresponds very well with the promoter-proximal DSIF signal at these promoters.

(C) NELF is enriched at genes with the highest levels of active transcription. Pol II-bound genes (n = 7466) 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 7466 bound genes, 6357 (85%) had ≥ 10 short RNAs mapping near their TSS, consistent with the presence of paused Pol II at these genes.

(E) Heat maps 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 7466 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) toward 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).



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}

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 = 7466 Pol II-bound genes). Right panel shows a good correlation between NELF promoter enrichment versus 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 (1268 of 1615 Most NELF-affected; 4056 of 4846 Less NELF-affected).

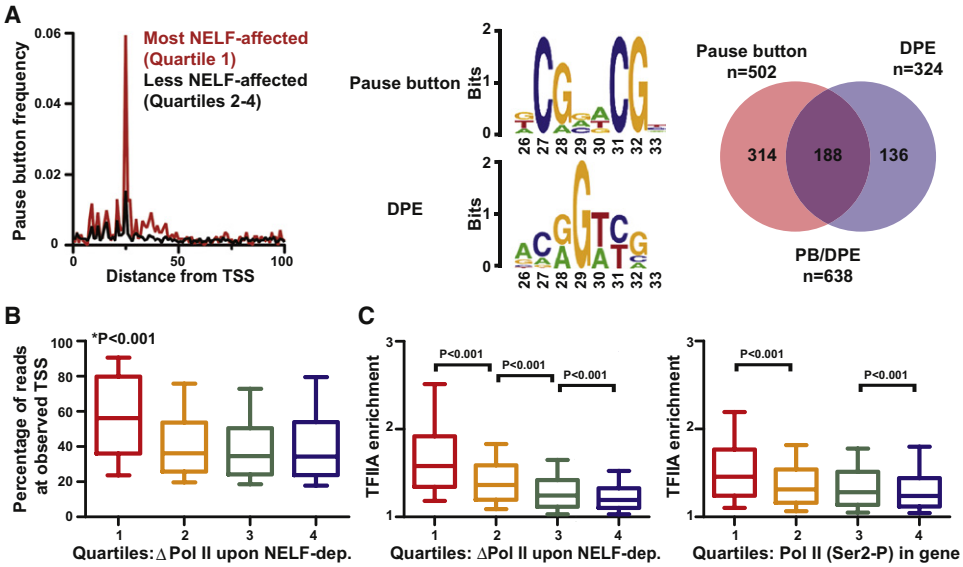


Figure S3. The Sequences of the Most NELF-Affected Genes May Induce Pausing and Facilitate Focused Transcription Initiation from Stable Preinitiation Complexes, Related to Figure 2

(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 1615 genes possess a consensus Pause button) and genes that are less NELF-affected (907 of 4846 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.

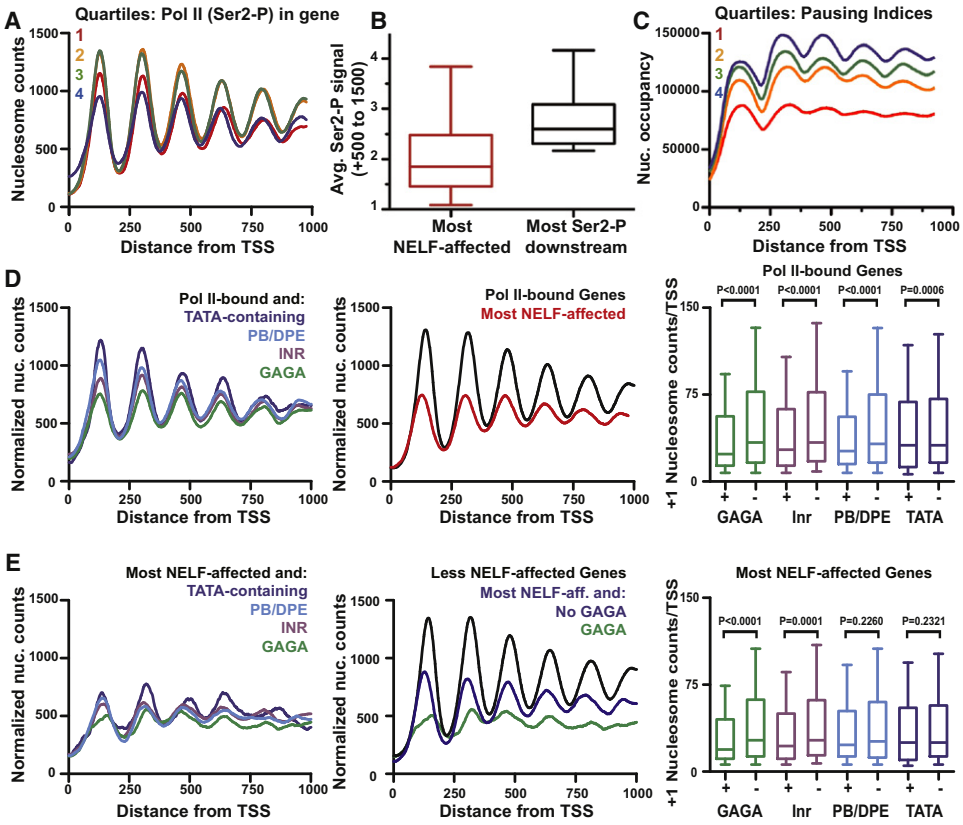


Figure S4. Elongating Pol II Only Modestly Disrupts Promoter-Proximal Nucleosomes, and Nucleosome Occupancy Is Lowest at Genes with GAGA-Binding Sites, Related to Figure 3

(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 = 1866) and the most highly active (Quartile 1 ranked by Ser2-P Pol II signal in the region +500 to +1500; n = 1866).

(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 Figure 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 = 6461) with: TATA (n = 398), Pause button or DPE (n = 1689), Inr (n = 1650) or GAGA (n = 1032) 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.

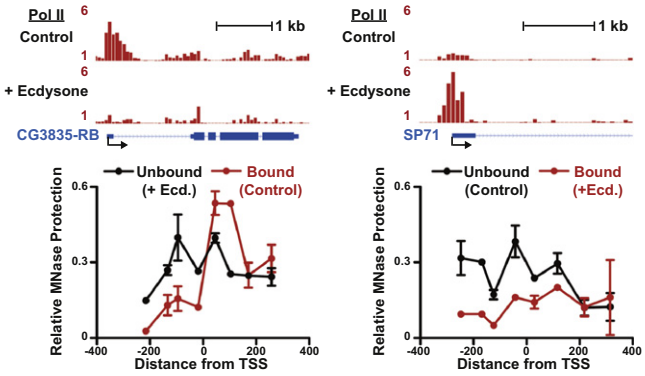


Figure S5. Antagonistic Relationship between Pol II and Nucleosome Binding at Highly Paused Genes Revealed by Ecdysone Treatment, Related to Figure 5

Left panel: CG3835-RA, a promoter with a high pausing index in control cells that becomes unbound by Pol II following 24 hr 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).

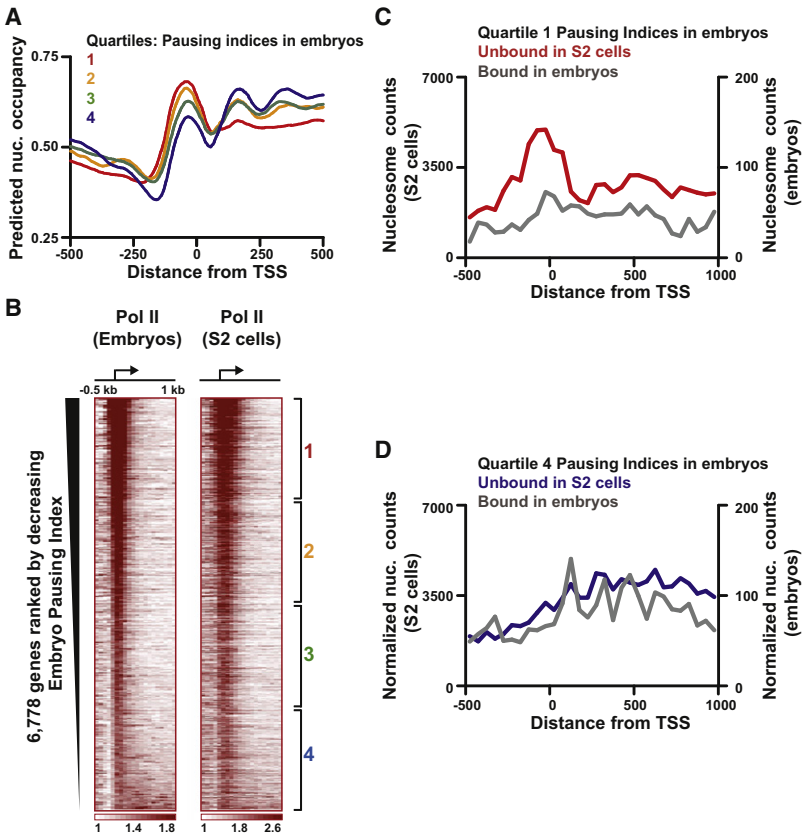


Figure S6. Paused Polymerase Prevents Promoters from Adopting Their Default Chromatin States, Related to Figure 6

(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 = 6778), divided into quartiles based on their Pausing indices (Quartile 1 has the highest Pausing Indices).

(B) Pol II distribution in *Drosophila* embryos versus S2 cells. Genes in both heat maps 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) versus (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).