Establishment of a Human RNA Pol II Pausing System and the Identification of O-GlcNAc Cycling Regulating Pol II Pausing and Elongation

Brian A. Lewis* and David Levens

Gene Regulation Section, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, NIH, 9000 Rockville Pike, Bethesda, MD 20892. *Corresponding author: brian.lewis@nih.gov

Abstract

Paused RNA polymerase II is the major regulated step of transcription in metazoans. We describe here a unique human cell-free transcription system that recapitulates RNA pol II pausing and assemble paused pol IIs on the human CMV IE, SV40, and heat shock promoters, all is the case in vivo. We then use the system to show that PARP-1 and CDK12/13 inhibitors directly affect pausing and elongation. We then show that O-GlcNAcylation is required for the establishment of a paused pol II: inhibition of OGT allowed pol II to bypass pausing and begin to elongate. Addback of rOGT or the pausing factor NELF re-established the pausing. In vivo nascent RNA measurements showed that OGA inhibition blocks elongation. These data show that cell-free systems can recapitulate RNA pol II pausing, that PARP-1 and CDK12/13 directly regulate RNA pol II elongation, and identify O-GlcNAc cycling regulating both pausing and pause release.

Introduction

Our understanding of transcriptional regulation by RNA polymerase II has undergone major revisions over the past several years. The predominant view of pol II transcriptional regulation assigned the primary regulatory point to the promoter region of a gene (consisting of various core promoter

26 elements and activator binding sites) and the congregation of general transcription factors that were then
27 recruited by an activator. The ordered assembly of factors at the promoter forms the preinitiation
28 complex (PIC). This view in part arose from the identification of the general factors and the eventual
29 reconstitution of pol II-dependent transcription. Consequently, the formation of the PIC was thought to
30 be the major regulatory step at promoters: upon initiation of transcription, pol II immediately enters
31 productive elongation.

32 Over the past 10 years genome-wide studies have shown that many, if not the majority of
33 promoters, from *Drosophila* to humans, are regulated at the paused pol II step, where an engaged pol II
34 sits between +30 and +75 downstream of the transcriptional start site (Adelman and Lis 2012). Thus, it
35 is important to understand the nature and regulation of the paused pol II both as a fundamental problem
36 and because many diseases are now considered "diseases of uncontrolled transcriptional elongation"
37 (Chen et al. 2018).

38 A paused pol II is thought to be established by two factors, DSIF and NELF (Wada et al. 1998b,
39 1998a; Yamaguchi et al. 1999), that together bind to pol II and effectively misalign the active site of pol
40 II from the DNA template preventing nucleotide incorporation and elongation (Bernecky et al. 2017;
41 Vos et al. 2018). Pol II release occurs via the phosphorylation of DSIF and NELF by P-TEFb (Yamada
42 et al. 2006; Fujinaga et al. 2004), after which NELF is lost, and the polymerase moves into a productive
43 elongation state (for review see Guo and Price 2013).

44 However, this model is incomplete. The addition of purified DSIF and NELF does not establish
45 a paused pol II, but only slows elongation (Cheng and Price 2007; Renner et al. 2001). These data
46 imply that additional factors are required to establish a paused pol II. Secondly, another catalytic activity,
47 poly-ADP ribose polymerase 1 (PARP-1), also regulated pol II release in cells (Gibson et al. 2016), and
48 so elongation is no longer simply due to P-TEFb activity.

49 More recently, an O-GlcNAcylation requirement in transcription has emerged. O-GlcNAc is
50 found on serine and threonine resides, often in a mutually exclusive relationship with phosphorylation.

There is only one transferase, the O-GlcNAc transferase (OGT) and one removal enzyme, O-GlcNAc aminidase (OGA) (Bond and Hanover 2013). High densities of OGA, OGT and O-GlcNAc co-localize with paused promoters in humans, *M. musculus*, *Drosophila*, and *C. elegans* (Resto et al. 2016; Love et al. 2010; Deplus et al. 2013; Krause et al. 2018). Previous experiments suggested that OGA activity is necessary for pol II elongation in a human cell-free system (Resto et al. 2016). For OGA to have a substrate, OGT must have acted prior to OGA, but it was not clear whether OGT activity occurred before or after PIC formation.

A cell-free transcription system that recapitulates pol II pausing would be highly desirable, enabling the dissection of the pausing machinery and the regulation of pause establishment and release. We describe here a human cell-free transcription system (CFS) that contains an actively engaged paused pol II which can be released via P-TEFb or SEC (which contains P-TEFb), Sarkosyl, or excess SPT5. This system was used to identify other, previously unknown components of the pausing machinery. Specifically, inhibition of the catalytic activity of OGT results in the loss of paused pol II, which instead immediately elongates, effectively bypassing the pausing step. This pausing defect can be rescued by the addition of rOGT or purified human NELF complex, suggesting that NELF-dependent pausing requires OGT. The CFS shows that pol II pausing can be recapitulated in human derived nuclear extracts. Additionally, these data show that the decision to pause is an intrinsic part of the initiation process and is predicated on the action of OGT, identifying O-GlcNAc cycling as a necessary regulatory event in RNA pol II pausing and release.

Results

The cell-free system recapitulates in vivo elongation events

Recently, one of us developed a nuclear extract from HeLa cells and the elongation properties of this extract were used to identify OGA as an elongation factor (Resto et al. 2016). One interesting

76 observation from those experiments was that the RNA products below the tRNA band at +75 were not
77 affected by the inhibition of either OGA or P-TEFb, whereas full-length elongation was completely
78 blocked. Lastly, the specific localization of these RNAs between +20 and +75 is exactly where one
79 would expect to find a paused pol II population (for review see Kwak et al. 2013). The following
80 experiments set out to determine the extent to which the CFS recapitulated in vivo behaviors and whether
81 these polymerases between +20 and +75 were in fact paused.

82 **Elongation inhibitors that block elongation in vivo show elongation defects in the CFS**

83 In order to establish the legitimacy of the CFS, it was necessary to determine whether the system
84 recapitulated the effects of known in vivo elongation inhibitors. Several groups have shown in vitro and
85 in vivo that THZ1 inhibitor blocks pol II elongation (Nilson et al. 2015; Kwiatkowski et al. 2014). THZ1
86 shows the highest specificity for the CDK7 kinase (IC50 = 27nM) while also inhibiting CDK12/13 at
87 higher concentrations (IC50 = 250nM) (Kwiatkowski et al. 2014). Addition of THZ1 (4 μM final) to
88 the CFS showed a clear elongation block of full-length 548 nt RNA product (lane 4, Fig. 1B).
89 Additionally, the RNA population between +20 and +75 decreased while there is an increase in RNAs
90 chased beyond the tRNA band at +75 [Note that tRNAs in the extract are factitiously labeled by CCA
91 tRNA nucleotidyl transferase and the ^{32}P-CTP (Adamson et al. 2003)]. The expected elongation block
92 with the CDK9 inhibitor flavopiridol was also observed (lane 3, Fig. 1B) (Cheng et al. 2012).

93 The CFS was assessed for the effects with the CDK12/13 inhibitor THZ531 (IC_{50}=158 nM/69
94 nM for CDK12/13 and IC_{50}=8.5 μM/10.5 μM for CDK7 and CDK9, respectively), that also blocked pol
95 II elongation in vivo (Zhang et al. 2016). Titration of increasing concentrations of THZ531 (in the same
96 volume of DMSO) into the CFS showed effects similar to those of THZ1: a reduction of short RNAs
97 between +20 and +75, an increase in RNAs above +75, and a block of full-length RNA synthesis
98 (compare lane 2 to lanes 3-5, Fig. 1C).

99 Kraus and colleagues have described a positive role for PARP-1 in pol II elongation in vivo
100 (Gibson et al. 2016). To examine the CFS for possible dependence on PARP-1, we added the PARP-1

101 inhibitor PJ34 to the CFS post-PIC formation, which resulted in a clear block in elongation with loss of
102 the 548 run-off product (lane 3, Fig. 1E), identical to flavopiridol (Fig. 1B), while having no effect on
103 the +20/+75 RNA population; DMSO alone had no effect (lane 2, Fig. 1D).

104 These experiments show that the CFS can recapitulate known elongation defects previously
105 documented in vivo with flavopiridol, THZ1, THZ531, and PJ34 (Nilson et al. 2015; Zhang et al. 2016;
106 Kwiatkowski et al. 2014; Gibson et al. 2016).

107 Lastly, we compared our extracts to an extract made with an ammonium sulfate-based protocol
108 and which has previously been used to study RNA pol II elongation (Adamson et al. 2003; Cheng and
109 Price 2007, 2009; Renner et al. 2001). We assayed both types of extracts side-by-side and confirmed
110 that the behaviors of these ammonium sulfate-derived extract ($AmSO_4$ NE) is entirely different:
111 elongation is extremely efficient with little to no accumulation of RNAs below +75, as compared to our
112 extracts (Figure 1E, compare lanes 1 and 2 to lane 3). This experiment further suggested that we had
113 developed an extract that was functionally and biochemically distinct.

114 **Defining a paused pol II population**

115 To further address whether the RNAs between +20 and +75 represent a paused pol II population
116 (delineated by the red brackets in Figure 2), we undertook several litmus tests. If these are paused
117 polymerases, then one would expect that they satisfy several criteria: the paused pol II population should
118 not chase into longer RNAs unless appropriately stimulated, should be sensitive to the detergent
119 Sarkosyl and the addition of DSIF (a known pausing factor), and P-TEFb should stimulate release of the
120 paused pol II and the synthesis of longer RNAs. Additionally, promoters known to establish a paused
121 pol II in vivo should do so in the CFS.

122 **Sarkosyl releases the paused pol II**

123 The release of paused pol II with the detergent Sarkosyl has been documented previously, in
124 both in vivo nuclear run-on assays and in Drosophila nuclear extracts (Rougvie and Lis 1988; Li et al.
125 2013). The proposed mechanism of Sarkosyl release is that treating the paused pol II with a mild

126 detergent removes the pausing factors thereby permitting pol II release and elongation (Wissink et al.
127 2019).

128 We set out to examine the stability of the putative paused pol II population and the effects of
129 Sarkosyl on that population. A pulse-chase time course was done to determine the stability of the paused
130 pol II. A chase with unlabeled CTP was allowed for 1, 3, 5, or 10 minutes (Fig. 2A, lanes 1-4). The
131 +20/+75 RNAs are remarkably stable, with minimal incorporation into longer products even after a 10'
132 chase. We next added Sarkosyl to the CFS after a 5' chase followed by an additional 5'chase, which
133 released a fraction of +20/+75 RNAs into longer species, in contrast to a 10' chase without Sarkosyl
134 (compare lanes 4 and 5, Fig. 2A). Note that because of the excess of cold NTPs in the chase, the longer
135 labeled RNAs here must have originated from the short, pulsed RNA population that elongated after the
136 Sarkosyl addition.

137 This experiment shows that the +20/+75 RNAs are nascent transcripts engaged with paused
138 polymerases that can elongate after Sarkosyl addition, as described in vivo, but otherwise do not chase
139 even after 10'. This is in contrast to other published elongation systems where the entire pol II
140 population will chase within 3 minutes to full-length RNA products (Cheng and Price 2007).

141 **GST-SPT5 stimulates elongation in the cell-free system**

142 The DSIF complex, a heterodimer of SPT4 and SPT5, is one of two known pausing factors
143 (Yamaguchi et al. 2013), but can also stimulate elongation (Chen et al. 2009; Yamada et al. 2006). As
144 such, one might expect that the addition of rSPT5 to the CFS would stimulate elongation, either by a
145 squelching mechanism where it titrated out other pausing factors, or through its positive elongation
146 activity. To that end, equivalent amounts of either rGST or rGST-SPT5 were titrated into the CFS after
147 PIC formation. After a 10' incubation, the standard pulse-chase assay was performed (Fig. 2B). The
148 addition of GST-SPT5 stimulated release of the +20/+75 RNA/paused pol II population, as there is a
149 diminution of RNA around +20 and a concomitant increase of RNAs longer than +75 (Fig. 2B, lanes 6-
150 8), while GST protein alone had little effect (Fig. 2B, lanes 2-5). Although the stimulation of elongation

151 was not efficient as no increase in full-length products was seen, this experiment indicates that the CFS
152 contains a pol II population that elongates in the presence to excess SPT5 as we predicted.

153 **Pausing occurs in the CFS using a human heat shock promoter and SV40 promoter**

154 We next asked whether other promoters known to harbor a paused pol II would show this
155 behavior in the CFS. The SV40 promoter was one of the first promoters suggested to have a paused pol
156 II (Skolnik-David and Aloni 1983). Titration of the SV40 promoter DNA into the CFS clearly showed
157 the establishment of a paused pol II population between +20 and +75 (lanes 1-5, Fig. 2C).

158 We extended these observations to the well-known pausing on heat shock promoters. We used
159 the human HSPA1B promoter (Bunch et al. 2014) in our standard pulse/chase assay and found that a
160 discrete paused pol II population existed between +20 and +75 nt, (Fig. 2C, lanes 6-7). These data show
161 that pausing occurs as a default event on three different promoters (CMV, SV40, and HSPA1B)
162 harboring a paused pol II in vivo.

163 **P-TEFb and Super Elongation Complex stimulate pause release in the CFS**

164 We next determined whether the CFS and the paused pol II population would respond to the
165 addition of recombinant P-TEFb, which is required for pause release in vivo. In order to assay P-TEFb
166 directly, the endogenous activity must first be removed from the CFS. Using a biotinylated-CMV
167 promoter template conjugated to magnetic beads, one can isolate various steps in the transcription
168 processes, wash off soluble factors, and then add back the activity in question to assess its function
169 (Resto et al. 2016; Adamson et al. 2003; Cheng and Price 2009). As indicated in Fig. 3A, a 30" pulse
170 was used to label RNA and transcription was stopped with EDTA. The beads were isolated and washed
171 with a buffer containing 60 mM KCl, and then resuspended in transcription buffer. As shown in lane 1,
172 Fig. 3C, there is no full-length product synthesized after the addition of NTPs for the chase step. The
173 addition of a concentrated 0.3M KCl elongation fraction eluted from a P11 column (P.3c) (Wada et al.
174 1998a), stimulated synthesis of full-length RNA (lane 2, Fig. 3C). The addition of recombinant P-TEFb,
175 purified from Sf9 cells (Fig. 3B), further stimulated elongation (lane 3, Figure 3C), showing that the

176 paused polymerases isolated on the immobilized templates are active and capable of being released into
177 productive elongation and synthesizing a full-length RNA.

178 The Super Elongation Complex (SEC) physically connects the ELL elongation factor to the P-
179 TEFb kinase (Luo et al. 2012; Lin et al. 2010). Both ELL and P-TEFb can stimulate elongation in vitro
180 by themselves and the SEC can stimulate elongation as a discrete complex (Biswas et al. 2011). The
181 question remains however, whether the SEC can stimulate pause release. Flag-tagged SEC (Lin et al.
182 2010) (F-SEC) was purified from 293 cells and western blot analysis showed that it contained several
183 SEC subunits (E1-E3, Fig. 3D). The F-SEC phosphorylated rGST-CTD, as did rP-TEFb, as expected
184 (Fig. 3E) (Biswas et al. 2011; Lin et al. 2010). Next, F-SEC was assayed using immobilized templates
185 (Fig. 3F). Titration of F-SEC relative to the elution buffer (containing FLAG peptide) showed
186 stimulation of isolated paused pol II complexes into intermediate and full-length RNA products
187 (compare lanes 2-5 with 6-9, Fig. 3G). Thus, these stably isolated, engaged, paused polymerases can
188 enter productive elongation with either P-TEFb or SEC.

189 These experiments all show that the CFS described here behaves as one would demand of a
190 system to study RNA pol II pausing and elongation and to address the necessary factors and mechanisms
191 of regulation. We now use this system to discover and explore previously unidentified factors in
192 pausing and pause release.

193 **The OGT substrate UDP-GlcNAc blocks elongation in the CFS**

194 Previously it was shown that OGT is required for PIC formation (Ranuncolo et al. 2012; Lewis
195 et al. 2016) and that OGA is required for proper elongation in the CFS (Resto et al. 2016). Since OGA
196 activity requires an O-GlcNAcylated substrate, prior OGT catalytic activity is also implicated in the
197 transition to elongation.

198 To establish a direct requirement for O-GlcNAcylation in the CFS the substrate of OGT, UDP-
199 GlcNAc, was titrated into the CFS (Fig. 4A). Titration of UDP-GlcNAc (after PIC formation) resulted
200 in a concentration-dependent decrease in full-length RNA (lanes 1-4, Fig. 4A). A similar result was

201 obtained with immobilized templates and the 0.3M P11 elongation fraction P.3c: addition of rOGT and
202 UDP-GlcNAc decreased synthesis of full-length RNA (compare lanes 2 and 3 in Fig. 4B). That
203 elongation was blocked suggested that pausing had increased via OGT activity and the O-
204 GlcNAcylation of pausing factors.

205 **Pausing and elongation factors are O-GlcNAcylated**

206 To test if pausing factors are O-GlcNAcylated, nuclear extract was incubated with wheat germ
207 agglutinin (WGA)-agarose beads to affinity purify O-GlcNAcylated proteins and the bound proteins
208 were assayed by western blot. NELF-A and -E bound the WGA-beads (left-most lane, Fig. 4C) but did
209 not bind beads that had been preincubated with the competitor GlcNAc, indicating that the WGA
210 interaction was specific to O-GlcNAcylation (+GlcNAc, Fig. 4C). As a control for the possible loss of
211 O-GlcNAcylation by endogenous OGA, the OGA inhibitor PUGNAc (+PUG; Fig. 4C) was added but
212 did not alter the presence of O-GlcNAcylated NELF subunits. The literature for mass spectroscopy data
213 of O-GlcNAcylated proteins further shows that at least 30 pausing and elongation factors are O-
214 GlcNAcylated (Fig. 4D), including DSIF and NELF (Hahne et al. 2012, 2013).

215 **OGT inhibitors reduce pausing and allow RNA pol II to elongate**

216 Previously, addition of the OGT catalytic inhibitor ST045849 (Gross et al. 2005) affected
217 transcription (Ranuncolo et al. 2012), but at what step was not determined. Thus, we set out to examine
218 the effects of OGT inhibition in a pulse-chase assay and to determine where in the transcription process
219 OGT acted.

220 The addition of ST04 along with CMV promoter DNA at the start of PIC formation completely
221 abrogated all RNA synthesis (compare lanes 1 and 2, Fig. 5A). However, ST04 added after PIC
222 formation showed a release of pol II from the paused population into elongation (post-PIC, compare
223 lanes 1 and 4, Fig. 5A), while PUGNAc showed the expected elongation defect (Resto et al. 2016) (lane
224 5, Fig. 5A). Titration of ST04 and a second OGT inhibitor ST06 showed similar effects: 0. 1mM ST04

225 and to a lesser extent 1mM ST06 showed that the paused pol II population at +20/+75 was released into
226 longer RNAs (compare lane 2 to lanes 3 and 6, Fig. 5B).

227 More recently, the Walker lab synthesized several additional OGT inhibitors (Ortiz-Meoz et al.
228 2015). Titration of these inhibitors (OSMI-2a and OSMI-4a) post-PIC formation showed an increase in
229 full-length RNA relative to the untreated and DMSO-treated nuclear extract (compare lanes 1 and 2 to
230 lanes 3 and 4, Fig. 5C). Further refinement of the ST04 titration showed the same effects: escape of the
231 paused pol II population, especially the population just above +20, and a significant increase in the full-
232 length RNA product (lanes 5-7, Fig. 5C). As the ST04 concentration increased further, a second effect
233 appeared closer to the full-length RNA at 548 nt (as seen in lane 4, Fig. 5A and lanes 3 and 6, Fig. 5B),
234 suggesting a second OGT requirement for elongation into a full-length RNA at 548nt. Thus, four
235 different OGT inhibitors show that OGT activity is necessary to establish a paused pol II, and in the
236 absence of OGT, the pol II does not pause, but immediately moves into an elongation phase.

237 **OGT reverses the pausing defect of OGT inhibitors**

238 If OGT is required for pausing, then the addition of rOGT and UDP-GlcNAc should reverse the
239 effects of ST04. This experiment is readily done with immobilized templates. PICs were formed for
240 30', then, where indicated, ST04 was added for 10'. Transcription was initiated with a 30" pulse
241 followed by inhibition with 20 mM EDTA. The beads were then isolated using a magnet, washed to
242 remove STO4, NTPs, and nuclear extract, and then resuspended in transcription buffer. To the indicated
243 reactions either rOGT or rOGT plus UDP-GlcNAc were added for 15', followed by a 5' chase with cold
244 NTPs (Fig. 5D). As expected, the untreated templates did not appreciably elongate after washing and
245 the ST04 reaction showed much higher levels of RNAs longer than +75 (lanes 1 and 2, Fig. 5E). The
246 addition of rOGT has a slight effect on the levels of these RNAs (lane 3, Fig. 5E) but the addition of
247 both rOGT and its nucleotide-sugar substrate UDP-GlcNAc dramatically suppressed the levels of RNAs
248 longer than +75 (lane 4, Fig. 5E), indicating that OGT catalytic activity reversed the pause-neutralizing

249 effect of ST04. This experiment shows that OGT catalytic activity is directly necessary for
250 establishment of the paused pol II.

251 **NELF rescues the pausing defect of OGT inhibition**

252 The release of the paused pol II after treatment with ST04 suggested that the NELF pausing factor
253 was specifically lost with OGT inactivation. To determine whether the addition of NELF back to a
254 ST04-treated EEC could prevent the release of the paused polymerases, we again isolated paused pol II
255 on immobilized templates (as in Fig. 5D), but after washing, added the P.3c elongation fraction with or
256 without ST04 (lanes 1-3, Fig. 5F). With STO4 added, there was a clear release of paused pol II as
257 indicated by the chasing of label from RNAs that are slightly larger than +20 (red arrow #1) into RNAs
258 throughout the paused region (red arrows #2-4) and then into and beyond +75 (red bracket, #5; compare
259 lanes 2 and 3, Fig. 5F). The titration of purified Flag-tagged NELF completely abrogated the release of
260 these paused polymerases as indicated by the loss of all labeled RNA longer than +75 (red bracket, #5),
261 as well as the reduction in RNA levels at and just below the +75 region (red arrows #2-4), and finally
262 the retention of the short RNA population just above the +20 mark (red arrow #1) that had released into
263 longer RNAs with STO4 (compare the RNA intensity indicated by arrow #1 between lanes 3 and 4-6).
264 Mechanistically, these data suggest that O-GlcNAcylation is required for pausing by maintaining the
265 paused pol II in a state receptive to NELF. Loss of O-GlcNAcylation causes the loss of NELF and
266 release of the paused pol II into an elongation-competent species.

267 **The OGA inhibitor PUGNAc blocks elongation in vivo by suppressing RNA pol II pause release**

268 To analyze the perturbation of O-GlcNAcylation on pausing and elongation in vivo, we used a
269 variation on the SLAM-seq/TT-seq (Muhar et al. 2018; Schwalb et al. 2016). The incorporation of 4-
270 thiouridine in vivo offered a rapid and efficacious approach to measuring changes in nascent RNA
271 expression. The 4sU approach (Fig. 6A) requires only the isolation of the RNA after treatment of the
272 cells with 4sU, followed by a brief treatment with iodoacetamide to alkylate the 4sU. During library
273 PCR amplification, the alkylated 4sU is recognized as a C and thus a G is incorporated opposite the

274 alkyl-4sU. During bioinformatic analysis the T > C conversions can be measured and isolated,
275 identifying the pulsed, actively transcribed RNA population (Muhar et al. 2018). A library made with
276 random primers permits the specific identification of nascent RNAs.

277 We treated HeLa cells with either the CDK9 inhibitor flavopiridol (FP) or the OGA inhibitor
278 PUGNAc using concentrations known to have effects in vivo. As a negative control, HeLa cells were
279 treated with an equivalent amount of DMSO. We used FP as a positive control as others have noted a
280 significant reduction in elongation (Zhou et al. 2012) and an increase in the nascent RNA-defined paused
281 pol II population at the 5' end of genes after FP treatment (Laitem et al. 2015). As expected, metagene
282 analysis of replicate FP treatments showed an increase in the paused pol II population just downstream
283 of the transcriptional start site (TSS) followed by decreasing incorporation of 4sU, indicative of an
284 elongation block (Figure 6B). Importantly, we found that the PUGNAc treatment resulted in a similar
285 effect where pol II had accumulated above DMSO levels at the 5' end of genes and a decreasing
286 incorporation of 4sU. In contrast, the control DMSO-treated cells showed less accumulation of a paused
287 pol II population and more importantly an upward sloping/increase in 4sU incorporation. These results
288 show that OGA inhibition in vivo blocks elongation and promotes the accumulation of paused pol II at
289 the 5' end of genes, much like the behavior of flavopiridol, a known elongation inhibitor. Lastly, the
290 requirement for OGA both in vivo and in the CFS requires, by definition, prior O-GlcNAcylation by
291 OGT, and which we have shown to occur prior to OGA activity in the CFS (Fig. 5).

292

293

294

295 **Discussion**

296 **Summary of the evidence of the recapitulation of RNA pol II pausing and elongation**

297 We present here evidence that a human cell-free system (CFS) recapitulates the expected
298 behaviors of pol II pausing and elongation. The CFS establishes a bona fide paused pol II, based on 12

299 different criteria (Figures 1-3). 1) The CFS synthesizes a population of paused RNAs found between
300 +20 and +75 downstream of the TSS. 2) These RNAs are stably engaged to the DNA template and
301 remain associated with the pol II. 3) The +20/+75 RNA population is intrinsically stable and alone will
302 not chase into longer RNAs. 4) Sarkosyl treatment of this population on washed immobilized templates
303 results in the release of the pol II and the synthesis of longer RNAs. 5) The treatment of the CFS with
304 the CDK9 inhibitor flavopiridol inhibits elongation after +75 but does not perturb the +20/+75 RNA
305 population. 6-7) The CFS is also dependent on CDK12/13 and PARP-1 catalytic activities, whose
306 requirements were shown in vivo (Gibson et al. 2016; Zhang et al. 2016), but which remained to be seen
307 in a direct, biochemical transcription assay. 8) In contrast, treatment with the CDK7 inhibitor, THZ1,
308 eliminates the +20/+75 RNA population, as observed in other experiments (Nilson et al. 2015). 9)
309 Excess SPT5 results in the synthesis of RNAs longer than +75, as expected if titrating out associated
310 pausing factors. 10) The +20/+75 RNA population is released upon treatment with either rP-TEFb or
311 purified SEC. 11) Using the CFS, we established paused pol II on the human heat shock promoter of
312 the HSPA1B gene. 12) Price and colleagues recently showed that in vivo the CMV IE promoter used
313 here and all other CMV promoters are in fact paused (Parida et al. 2019). These data show that the CFS
314 establishes an engaged, paused RNA pol II population between +20 and +75 and which can enter
315 productive elongation. Furthermore, we show that the human HSPA1B promoter, which harbors a
316 paused pol II in vivo, also assembled a paused pol II in the CFS (Fig. 2).

317 These data make clear that chromatin and nucleosomes are not required for pausing per se. No
318 chromatin has been added to the CFS nor have steps been taken to assemble any of these templates with
319 nucleosomes. Our data also reflect CMV transcription in vivo, where pausing occurs at all viral
320 promoters despite not being packaged into chromatin (Parida et al., 2019). Pausing is a direct, default
321 consequence of transcription initiation.

322 **O-GlcNAc cycling is necessary for RNA pol II pausing and elongation**

323 O-GlcNAc and OGA ChIP-seq peaks overlap paused pol II positions in human BJAB cells
324 (Resto et al. 2016). OGT ChIP-seq peaks map to promoters in mice (Deplus et al. 2013) and OGT
325 shRNA reduced pol II and O-GlcNAc levels at several human B-cell promoters (Ranuncolo et al. 2012).
326 Mass spectroscopy data indicated that over 20 pausing and elongation-associated factors, including
327 DSIF and NELF (Fig. 4), are O-GlcNAcylated (Hahne et al. 2012), as are the bovine and human RNA
328 pol II C-terminal domains (Kelly et al. 1993; Ranuncolo et al. 2012). The modification of these factors
329 accounts for the co-localization of O-GlcNAc and paused pol II at promoters in vivo. Previous
330 experiments showed that inhibition of OGA in the CFS blocked elongation as effectively as the CDK9
331 inhibitor flavopiridol (Resto et al. 2016). Those experiments indicated that OGA catalytic activity was
332 required for elongation and secondly, that OGT must have acted prior to that point in order to create a
333 substrate for the OGA. The co-localization of OGT, OGA, and the O-GlcNAc modification at promoters
334 in vivo, the reduction in pol II promoter occupancy after OGT knockdown, and the OGA requirement
335 for elongation all point to a model where O-GlcNAc cycling regulates RNA pol II recruitment and
336 elongation.

337 However, it was unclear at what steps in transcription OGT acted. The experiments in Fig. 5
338 precisely delineated OGT activity in both PIC formation and pause establishment. The rescue of the
339 OGT inhibition with rOGT showed that pausing was a direct consequence of OGT catalytic activity.
340 OGT inhibition was also rescued by the addition of purified NELF, strongly suggesting that the
341 mechanism by which OGT established a paused pol II is via recruitment of NELF (Fig. 5). Furthermore,
342 the titration of UDP-GlcNAc into the CFS blocked synthesis of a full-length RNA product, showing that
343 increased O-GlcNAcylation decreased the elongation capacity of the CFS (Fig. 4). Likewise, the
344 addition of rOGT and UDP-GlcNAc to an immobilized template assay also blocked elongation (Fig. 4).
345 Lastly, we have confirmed the OGA, and by extension, the OGT requirement for elongation by showing
346 a significant accumulation of paused pol II in vivo after OGA inhibition (Fig. 6).

347 Our results suggest a model where O-GlcNAcylation addition and removal respectively regulate
348 the establishment of a paused pol II and the subsequent release of the paused pol II into productive
349 elongation. O-GlcNAcylation may regulate phosphorylation by blocking access of kinases to the
350 necessary serine and threonine residues. OGA activity removes this block and permits phosphorylation
351 to occur, leading to pol II escape into productive elongation. Alternatively, O-GlcNAcylation
352 addition/removal is directly regulating the activities of the pausing and elongation protein machinery.

353 **A multi-factor model of RNA pol II pausing and release**

354 These data can be further assembled into a multifactor model of pol II pause establishment and
355 release. Pol II pausing is established by DSIF and NELF in conjunction with the catalytic action of
356 CDK7 (Larochelle et al. 2012) and OGT. Release of the paused pol II into productive elongation is also
357 a multifactor release step requiring P-TEFb (Guo and Price 2013), OGA (Resto et al. 2016), and PARP-
358 1 (Gibson et al. 2016) (Fig. 7). Lastly, the substrates ATP, UDP-GlcNAc, and NAD+ are all high energy
359 compounds. Hence, we hypothesize that these high energy donors may satisfy specific energy
360 requirements for pausing and pause release in addition to the effect of post-translational modification
361 itself.

362 Furthermore, these high energy donor compounds are synthesized by various metabolic
363 pathways. The OGT substrate UDP-GlcNAc is made from fructose-6-phosphate, glutamine, and acetyl
364 CoA, derived from the glycolytic pathway, amino acid uptake, and lipid metabolism, respectively. The
365 PARP substrate NAD+ is derived from tryptophan and connects oxidative phosphorylation and the
366 Krebs cycle. Finally, ATP is the output from oxidative phosphorylation. If these enzymes are
367 responsive to the metabolic flux of their substrates, then the pol II machinery at the 5' proximal end of
368 genes may be a sensor of the nutrient/metabolic state of the cell, modulating transcriptional output in
369 response to changing metabolic flux.

370 **The cell-free transcription system**

371 Pol II pausing is of paramount importance, as most genes now are known to be regulated at that
372 stage and not at the earlier PIC formation/initiation stage. The lack of a human nuclear extract system
373 that recapitulates pausing has hindered understanding of pausing. The cell-free system described here
374 will allow further exploration of the depths of pol II pausing, the factors involved in this step, and how
375 pausing and pause release might be regulated. Such a system allows one to directly probe these events,
376 in contrast to the indirect genetic experiments to which one must resort for in vivo studies. Through the
377 combination of the bottom-up functional biochemistry and in vivo genetic approaches, we should be
378 able to further understand RNA pol II pausing and its regulation.
379

380 **Acknowledgments**

381 The authors wish to thank Ali Shilatifard for the F-SEC cell line, Nathaniel Gray for THZ1 and
382 THZ531, David Price for the $AmSO_4$ extract, Yuki Yamaguchi for the GST-SPT5 expression vector and
383 F-NELF HeLa cell line, Jane Jones of the NCI Frederick Protein Synthesis Core Facility for the
384 production and isolation of rP-TEFb, Suzanne Walker for supplying the OSMI inhibitors, and the NCI
385 Sequencing Facility for library construction and Illumina sequencing. B.A.L. wishes to especially thank
386 Tobius Neuman for assistance in implementing the SLAM-DUNK bioinformatic analysis.
387

388 **Funding:** This work was supported by the Intramural Research Program, Center for Cancer Research,
389 National Cancer Institute.
390

391 **Author contributions:** B.A.L. conceived of the work, designed and conducted the experiments,
392 analyzed the data, and wrote the manuscript. D.L. contributed to the data analysis and manuscript
393 writing.
394

395 **Competing interests:** The authors declare no competing interests.

396

397 **Data and materials availability:** Raw and processed data are available in the GEO Repository,

398 accession number GSE158746: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158746

399

400 **Materials and Methods**

401 **In vitro transcription assays**

402 Assays were all done (NE-based and EEC-based) as in the appropriate figure schematics and as

403 described previously (Resto et al. 2016), except that a 60 mM KCl salt wash buffer (Cheng and Price

404 2007) was used to wash the EECs.

405

406 **Proteins, HeLa nuclear extract, and HeLa P.3 Fraction**

407 rP-TEFb was synthesized in Tni-FNL cells by co-expression of CDK9-his6 and cyclinT1 subunits and

408 then purified via IMAC. rGST and rGST-SPT5 proteins were purified as described previously for other

409 GST-tagged proteins (Ranuncolo et al. 2012). HeLa nuclear extract was prepared as described, using a

410 0.4M KCl step for elution from nuclei to make the nuclear extract(Abmayr et al. 2006). F-SEC was

411 purified from 293 Flp-in-TRex cells as indicated (Lin et al. 2010). rOGT and GST-CTD were purified

412 as described previously (Ranuncolo et al. 2012). HeLa P.3 fraction was isolated by fractionation of a

413 HeLa nuclear extract with a P11 column. The column was eluted with 0.1, 0.3, 0.5, and 1M KCl BC

414 buffer (containing 20 mM Tris pH7.9 @ 4ºC, 10% glycerol, 0.2 mM DTT, 0.2 mM EDTA). The 0.3M

415 peak was pooled and precipitated with 50% ammonium sulfate. The resulting precipitate (P.3c) was

416 suspended in 1/10th volume (relative to the starting material volume) BC buffer without KCl and

417 dialyzed against 0.1M KCl BC buffer. F-NELF complex was affinity-purified from stably transfected

418 HeLa cells.

419

420 **WGA assay**

421 20 μL wheat germ agglutinin (WGA) agarose beads was washed 1x with 500 μL BC100, spun to remove
422 buffer with 26 gauge needle/syringe. 50 μL (10 mg/mL) nuclear extract was mixed with 50 μL H_2O
423 and incubated with the WGA beads for 60 minutes with rotation at room temperature. For the addition
424 of PUGNAc, 10 μL of 20 mM PUGNAc was added to the 100 μL mixture of extract and H_2O. To use
425 GlcNAc as competitor, 50 μL 1M GlcNAc was added to the 50 μL extract and incubated with the WGA
426 beads. After incubating for one hour, beads were washed 3x with BC100, eluted with Laemmli sample
427 buffer, separated by denaturing PAGE, transferred to nitrocellulose, and analyzed for NELF by western
428 blot.
429
430 **Kinase assay**
431 0.2 μL GST CTD 1-26, 0.5 μL F-AFF1 or 0.5 μL rPTEFb, and 8.5 μL H_2O, 1 μL 10x P-TEFb kinase
432 buffer (50mM Tris pH7.5, 5mM DTT, 5mM $MnCl_2$, 4mM $MgCl_2$ (Zhou et al. 2000)) were mixed and
433 incubated for 30 minutes at 30^0C. Sample buffer was added to each sample and separated by denaturing
434 PAGE followed by transfer to nitrocellulose and western blotting with anti-phosphoserine 2 antibody.
435
436 **4sU experiment**
437 Approximately, $1x10^7$ HeLa cells were grown on 10 cm plates and treated at 70% confluency with either
438 DMSO or either flavopiridol (1 mM final) or PUGNAc (100 mM final) for 60 minutes. Cells were then
439 treated with 1 mM final concentration of 4-thiouridine for 10 minutes. RNA was isolated using Trizol
440 (Invitrogen) and treated with iodoacetamide as indicated (Herzog et al. 2017; Muhar et al. 2018).
441 Libraries were constructed with indexed random primers and subjected to paired-end sequencing by the
442 NCI Core Sequencing Facility. Two separate experiments were done using the same conditions.
443 Sequences were analyzed by the SLAM-DUNK software (https://t-neumann.github.io/slamdunk/)
444 (Neumann et al. 2019) and deeptools (https://deeptools.readthedocs.io/en/latest/index.html) (Ramírez et

445 al. 2014). Raw and processed data are available in the GEO Repository, accession number GSE158746:

446 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158746

447

448 **Antibodies**

449 Anti-pol II CTD phosphoserine S2 (Abcam ab5095)

450 Anti-ELL2 (Abcam ab194445)

451 Anti-CDK9 (Bethyl A303-493A)

452 Anti-Flag (Sigma)

453 Anti-NELF-A (Santa Cruz sc-32911)

454 Anti-NELF-E (Santa Cruz sc-32912)

455

456 **Reagents**

457 ST045849 and ST060266 (TimTec), Flavopiridol (Sigma F3055), PJ34, THZ531 (APExBIO, A8736),

458 PUGNAc (Sigma A7229), ^{32}P-CTP (3000 Ci/mmol, Perkin Elmer), NTPs (Roche), Sarkosyl (Sigma),

459 WGA-agarose, GlcNAc (Sigma), UDP-GlcNAc (Sigma).

460

461 **References**

462

463 Abmayr SM, Yao T, Parmely T, Workman JL. 2006. Preparation of Nuclear and Cytoplasmic Extracts

464 from Mammalian Cells. *Curr Protoc Mol Biol* **75**: 12.1.1-12.1.10.

465 Adamson TE, Shore SM, Price DH. 2003. Analysis of RNA polymerase II elongation in vitro.

466 *Methods Enzymol* **371**: 264–75. http://www.ncbi.nlm.nih.gov/pubmed/14712706.

467 Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles

468 in metazoans. *Nat Rev Genet* **13**: 720–31.

469 http://www.ncbi.nlm.nih.gov/pubmed/22986266 (Accessed November 5, 2012).

470 Bentley DL, Groudine M. 1986. Novel promoter upstream of the human c-myc gene and
471 regulation of c-myc expression in B-cell lymphomas. *Mol Cell Biol* **6**: 3481–3489.
472 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citat
473 ion&list_uids=3540591.

474 Bernecky C, Plitzko JM, Cramer P. 2017. Structure of a transcribing RNA polymerase II–
475 DSIF complex reveals a multidentate DNA–RNA clamp. *Nat Struct Mol Biol* **24**.
476 http://www.nature.com/doifinder/10.1038/nsmb.3465.

477 Biswas D, Milne TA, Basrur V, Kim J, Elenitoba-Johnson KSJ, Allis CD, Roeder RG. 2011.
478 Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through
479 distinct partner protein complexes. *Proc Natl Acad Sci* **108**: 15751–15756.
480 http://www.pnas.org/cgi/doi/10.1073/pnas.1111498108.

481 Bond MR, Hanover JA. 2013. *O*- GlcNAc Cycling: A Link Between Metabolism and Chronic
482 Disease. *Annu Rev Nutr* **33**: 205–229.
483 http://www.annualreviews.org/doi/10.1146/annurev-nutr-071812-161240.

484 Bunch H, Zheng X, Burkholder A, Dillon ST, Motola S, Birrane G, Ebmeier CC, Levine S,
485 Fargo D, Hu G, et al. 2014. TRIM28 regulates RNA polymerase II promoter-proximal
486 pausing and pause release. *Nat Struct Mol Biol* **21**: 876–83.
487 http://dx.doi.org/10.1038/nsmb.2878.

488 Chen FX, Smith ER, Shilatifard A. 2018. Born to run: control of transcription elongation by
489 RNA polymerase II. *Nat Rev Mol Cell Biol* **70**. https://www.nature.com/articles/s41580-
490 018-0010-5.

491 Chen Y, Yamaguchi Y, Tsugeno Y, Yamamoto J, Yamada T, Nakamura M, Hisatake K,
492 Handa H. 2009. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative
493 roles in RNA polymerase II elongation. *Genes Dev* **23**: 2765–2777.

494 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2788331&tool=pmcentrez&r
495 endertype=abstract (Accessed January 7, 2013).

496 Cheng B, Li T, Rahl PB, Adamson TE, Loudas NB, Guo J, Varzavand K, Cooper JJ, Hu X,
497 Gnatt A, et al. 2012. Functional association of Gdown1 with RNA polymerase II poised
498 on human genes. *Mol Cell* **45**: 38–50.
499 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3259526&tool=pmcentrez&r
500 endertype=abstract (Accessed December 19, 2012).

501 Cheng B, Price DH. 2009. Isolation and functional analysis of RNA polymerase II elongation
502 complexes. *Methods* **48**: 346–52.
503 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2754188&tool=pmcentrez&r
504 endertype=abstract (Accessed January 7, 2013).

505 Cheng B, Price DH. 2007. Properties of RNA polymerase II elongation complexes before and
506 after the P-TEFb-mediated transition into productive elongation. *J Biol Chem* **282**:
507 21901–21912. http://www.ncbi.nlm.nih.gov/pubmed/17548348 (Accessed November 5,
508 2012).

509 Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, Dawson M a,
510 Volkmar M, Putmans P, Calonne E, et al. 2013. TET2 and TET3 regulate GlcNAcylation
511 and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* **32**: 645–655.
512 http://www.ncbi.nlm.nih.gov/pubmed/23353889.

513 Eick D, Bornkamm GW. 1986. Transcriptional arrest within the first exon is a fast control
514 mechanism in c-myc gene expression. *Nucleic Acids Res* **14**: 8331–8346.
515 https://doi.org/10.1093/nar/14.21.8331.

516 Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. 2004. Dynamics of human
517 immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates

518 negative effectors from the transactivation response element. *Mol Cell Biol* **24**: 787–95.
519 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=343783&tool=pmcentrez&re
520 ndertype=abstract.

521 Gibson BA, Zhang Y, Jiang H, Hussey KM, Shrimp JH, Lin H, Schwede F, Yu Y, Kraus WL.
522 2016. Chemical genetic discovery of PARP targets reveals a role for PARP-1 in
523 transcription elongation. *Science* **353**: 45–50.
524 http://www.ncbi.nlm.nih.gov/pubmed/27256882 (Accessed July 22, 2016).

525 Gross BJ, Kraybill BC, Walker S. 2005. Discovery of O-GlcNAc Transferase Inhibitors. *J Am*
526 *Chem Soc* **127**: 14588–14589. http://www.ncbi.nlm.nih.gov/pubmed/16231908.

527 Guo J, Price DH. 2013. RNA polymerase II transcription elongation control. *Chem Rev* **113**:
528 8583–603. http://www.ncbi.nlm.nih.gov/pubmed/23919563 (Accessed October 4, 2016).

529 Hahne H, Gholami AM, Kuster B. 2012. Discovery of O -GlcNAc-modified Proteins in
530 Published Large-scale Proteome Data. *Mol Cell Proteomics* **11**: 843–850.
531 http://www.mcponline.org/lookup/doi/10.1074/mcp.M112.019463.

532 Hahne H, Sobotzki N, Nyberg T, Helm D, Borodkin VS, van Aalten DMF, Agnew B, Kuster
533 B. 2013. Proteome wide purification and identification of O-GlcNAc-modified proteins
534 using click chemistry and mass spectrometry. *J Proteome Res* **12**: 927–36.
535 http://www.ncbi.nlm.nih.gov/pubmed/23301498%5Cnhttp://dx.doi.org/10.1021/pr30096
536 7y.

537 Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, Von
538 Haeseler A, Zuber J, Ameres SL. 2017. Thiol-linked alkylation of RNA to assess
539 expression dynamics. *Nat Methods* **14**: 1198–1204.

540 Kelly WG, Dahmus ME, Hart GW. 1993. RNA polymerase II is a glycoprotein. Modification

541 of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* **268**: 10416–24.
542 http://www.ncbi.nlm.nih.gov/pubmed/8486697.

543 Krause MW, Love DC, Ghosh SK, Wang P, Yun S, Fukushige T, Hanover JA. 2018.
544 Nutrient-driven O-GlcNAcylation at promoters impacts genome-wide RNA Pol II
545 distribution. *Front Endocrinol (Lausanne)* **9**.

546 Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how
547 promoters direct initiation and pausing. *Science* **339**: 950–3.
548 http://www.sciencemag.org/cgi/doi/10.1126/science.1229386.

549 Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, Dastur A, Amzallag
550 A, Ramaswamy S, Tesar B, et al. 2014. Targeting transcription regulation in cancer with
551 a covalent CDK7 inhibitor. *Nature* **511**: 616–620.
552 http://www.nature.com/doifinder/10.1038/nature13393.

553 Laitem C, Zaborowska J, Isa NF, Kufs J, Dienstbier M, Murphy S. 2015. CDK9 inhibitors
554 define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat*
555 *Struct Mol Biol* **22**: 396–403. http://www.nature.com/doifinder/10.1038/nsmb.3000.

556 Larochelle S, Amat R, Glover-Cutter K, Sansó M, Zhang C, Allen JJ, Shokat KM, Bentley
557 DL, Fisher RP. 2012. Cyclin-dependent kinase control of the initiation-to-elongation
558 switch of RNA polymerase II. *Nat Struct Mol Biol* **19**: 3–8.
559 http://www.ncbi.nlm.nih.gov/pubmed/23064645 (Accessed October 30, 2012).

560 Lewis BA, Burlingame AL, Myers SA. 2016. Human RNA Polymerase II Promoter
561 Recruitment *in Vitro* Is Regulated by *O* -Linked *N* -Acetylglucosaminyltransferase
562 (OGT). *J Biol Chem* **291**: 14056–14061.
563 http://www.jbc.org/lookup/doi/10.1074/jbc.M115.684365 (Accessed July 11, 2016).

564 Li J, Liu Y, Rhee HS, Ghosh SKB, Bai L, Pugh BF, Gilmour DS. 2013. Kinetic competition
565 between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol*
566 *Cell* **50**: 711–722. http://dx.doi.org/10.1016/j.molcel.2013.05.016.

567 Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, Washburn MP, Conaway
568 JW, Conaway RC, Shilatifard A. 2010. AFF4, a Component of the ELL/P-TEFb
569 Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription
570 Elongation to Leukemia. *Mol Cell* **37**: 429–437.
571 http://dx.doi.org/10.1016/j.molcel.2010.01.026.

572 Love DC, Ghosh S, Mondoux MA, Fukushige T, Wang P, Wilson MA, Iser WB, Wolkow
573 CA, Krause MW, Hanover JA. 2010. Dynamic O-GlcNAc cycling at promoters of
574 Caenorhabditis elegans genes regulating longevity, stress, and immunity. *Proc Natl Acad*
575 *Sci* **107**: 7413–7418.
576 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citat
577 ion&list_uids=20368426 (Accessed November 5, 2012).

578 Luo Z, Lin C, Guest E, Garrett AS, Mohaghegh N, Swanson S, Marshall S, Florens L,
579 Washburn MP, Shilatifard A. 2012. The Super Elongation Complex Family of RNA
580 Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output.
581 *Mol Cell Biol* **32**: 2608–2617. http://mcb.asm.org/cgi/doi/10.1128/MCB.00182-12
582 (Accessed December 19, 2012).

583 Muhar M, Ebert A, Neumann T, Umkehrer C, Jude J, Wieshofer C, Rescheneder P, Lipp JJ,
584 Herzog VA, Reichholf B, et al. 2018. SLAM-seq defines direct gene-regulatory
585 functions of the BRD4-MYC axis. *Science (80-)* **360**: 800–805.
586 http://www.sciencemag.org/lookup/doi/10.1126/science.aao2793.

587 Neumann T, Herzog VA, Muhar M, Von Haeseler A, Zuber J, Ameres SL, Rescheneder P.

588 2019. Quantification of experimentally induced nucleotide conversions in high-
589 throughput sequencing datasets. *BMC Bioinformatics* **20**: 1–16.

590 Nilson KAA, Guo J, Turek MEE, Brogie JEE, Delaney E, Luse DSS, Price DHH. 2015.
591 THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol Cell* **59**:
592 1–13. http://linkinghub.elsevier.com/retrieve/pii/S1097276515005262.

593 Ortiz-Meoz RF, Jiang J, Lazarus MB, Orman M, Janetzko J, Fan C, Duveau DY, Tan ZW,
594 Thomas CJ, Walker S. 2015. A Small Molecule That Inhibits OGT Activity in Cells.
595 *ACS Chem Biol* **10**: 150309142052000.
596 http://pubs.acs.org/doi/abs/10.1021/acschembio.5b00004.

597 Parida M, Nilson KA, Li M, Ball CB, Fuchs HA, Lawson CK, Luse DS, Meier JL, Price DH.
598 2019. Nucleotide Resolution Comparison of Transcription of Human Cytomegalovirus
599 and Host Genomes Reveals Universal Use of RNA Polymerase II Elongation Control
600 Driven by Dissimilar Core Promoter Elements. *MBio* **10**: 1–20.

601 Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb.
602 *Mol Cell* **23**: 297–305. http://www.ncbi.nlm.nih.gov/pubmed/16885020.

603 Ping YH, Rana TM. 2001. DSIF and NELF Interact with RNA Polymerase II Elongation
604 Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA
605 Polymerase II and DSIF during Transcription Elongation. *J Biol Chem* **276**: 12951–
606 12958.

607 Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. DeepTools: A flexible platform
608 for exploring deep-sequencing data. *Nucleic Acids Res* **42**: 187–191.

609 Ranuncolo SM, Ghosh S, Hanover J a., Hart GW, Lewis B a. 2012. Evidence of the
610 involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in

611 vitro and in vivo. *J Biol Chem* **287**: 23549–23561.

612 Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. 2001. A highly purified RNA
613 polymerase II elongation control system. *J Biol Chem* **276**: 42601–9.
614 http://www.ncbi.nlm.nih.gov/pubmed/11553615 (Accessed January 22, 2013).

615 Resto M, Kim BH, Fernandez AG, Abraham BJ, Zhao K, Lewis BA. 2016. O-GlcNAcase is
616 an RNA polymerase II elongation factor coupled to pausing factors SPT5 and TIF 1β. *J*
617 *Biol Chem* **291**: 22703–22713. http://www.ncbi.nlm.nih.gov/pubmed/27601472
618 (Accessed October 4, 2016).

619 Rougvie AE, Lis JT. 1988. The RNA polymerase II molecule at the 5??? end of the uninduced
620 hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54**: 795–804.
621 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citat
622 ion&list_uids=3136931.

623 Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016.
624 TT-seq maps the human transient transcriptome. *Science (80-)* **352**: 1225–1228.
625 http://www.ncbi.nlm.nih.gov/pubmed/27257258%5Cnhttp://science.sciencemag.org/cont
626 ent/sci/352/6290/1225.full.pdf.

627 Skolnik-David H, Aloni Y. 1983. Pausing of RNA polymerase molecules during in vivo
628 transcription of the SV40 leader region. *EMBO J* **2**: 179–184.

629 Vos SM, Farnung L, Boehning M, Wigge C, Linden A, Urlaub H, Cramer P. 2018. Structure
630 of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*.
631 http://www.nature.com/articles/s41586-018-0440-4.

632 Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog
633 GA, Winston F, et al. 1998a. DSIF, a novel transcription elongation factor that regulates

634 RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes*
635 *Dev* **12**: 343–356.

636 Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. 1998b. Evidence that P-TEFb
637 alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in
638 vitro. *EMBO J* **17**: 7395–7403.
639 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1171084&tool=pmcentrez&r
640 endertype=abstract.

641 Wissink EM, Vihervaara A, Tippens ND, Lis JT. 2019. Nascent RNA analyses: tracking
642 transcription and its regulation. *Nat Rev Genet* **20**: 705–723.

643 Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. 2006. P-TEFb-mediated
644 phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription
645 elongation. *Mol Cell* **21**: 227–37. http://www.ncbi.nlm.nih.gov/pubmed/16427012
646 (Accessed January 7, 2013).

647 Yamaguchi Y, Shibata H, Handa H. 2013. Transcription elongation factors DSIF and NELF:
648 Promoter-proximal pausing and beyond. *Biochim Biophys Acta - Gene Regul Mech* **1829**:
649 98–104. http://dx.doi.org/10.1016/j.bbagrm.2012.11.007.

650 Yamaguchi YY, Takagi TT, Wada TT, Yano KK, Furuya AA, Sugimoto SS, Hasegawa JJ,
651 Handa HH. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF
652 to repress RNA polymerase II elongation. *Cell* **97**: 41–51.
653 http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10199401&re
654 tmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.1016/S0092-8674(00)80713-
655 8.

656 Zhang T, Kwiatkowski N, Olson CM, Dixon-Clarke SE, Abraham BJ, Greifenberg AK,
657 Ficarro SB, Elkins JM, Liang Y, Hannett NM, et al. 2016. Covalent targeting of remote

658 cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat Chem Biol* **12**: 876–884.
659 http://www.nature.com/doifinder/10.1038/nchembio.2166.

660 Zhou M, Halanski MA, Radonovich MF, Kashanchi F, Peng J, Price DH, Brady JN. 2000. Tat
661 Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II
662 Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1
663 Transcription. *Mol Cell Biol* **20**: 5077–5086.
664 http://mcb.asm.org/cgi/content/abstract/20/14/5077.

665 Zhou Q, Li T, Price DH. 2012. RNA polymerase II elongation control. *Annu Rev Biochem* **81**:
666 119–43. http://www.ncbi.nlm.nih.gov/pubmed/22404626 (Accessed November 6, 2012).

667
668
669
670
671
672
673
674
675
676
677
678
679
680
681

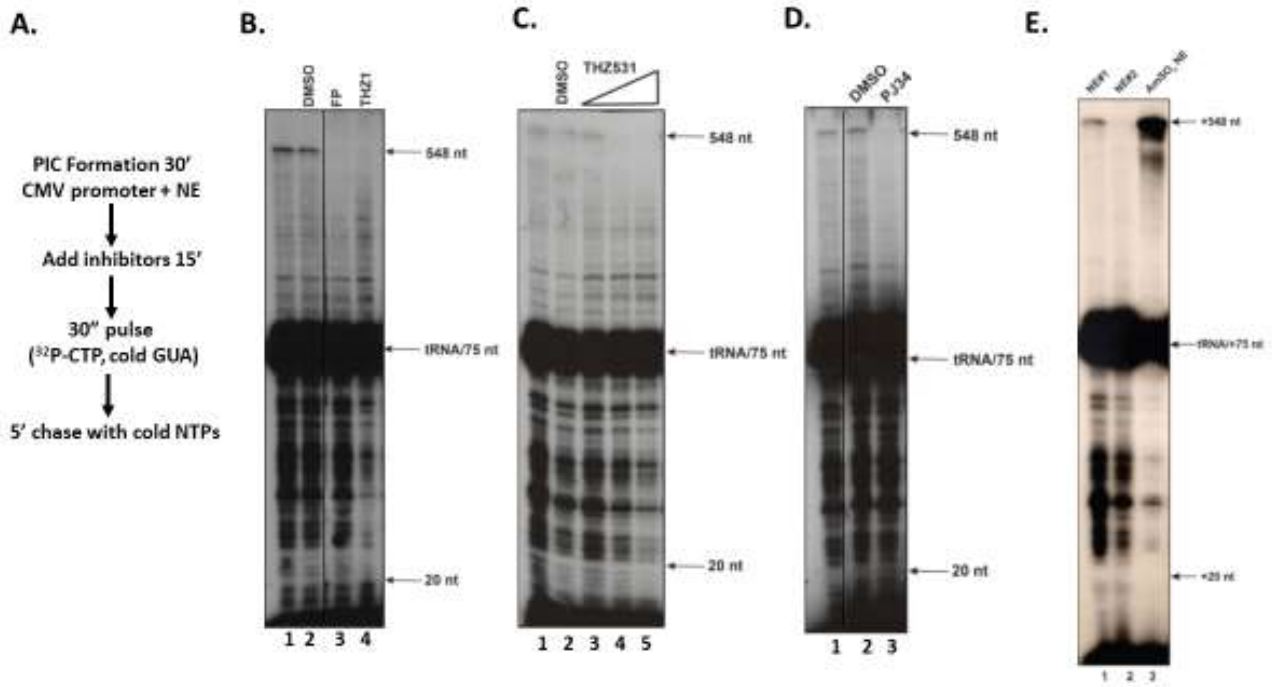

Figure 1

Figure 1. Cell-free system is dependent on CDK7, CDK9, CDK12/13, and PARP-1.

A. Schematic of steps in the experiments in panels B-D.

B. Shown is a pulse-chase assay using the DMSO solvent alone, 1 μM CDK9 inhibitor flavopiridol (FP), or 4 μM CDK7 inhibitor THZ1 in DMSO. Transcription products were analyzed on an 8% urea/TBE gel followed by autoradiography. Vertical black line between lanes 2 and 3 indicates a deletion of intervening lanes relative to the original autoradiograph.

C. Shown is a pulse-chase assay using the DMSO solvent alone and a titration of the CDK12 inhibitor THZ531 (0.4 μM, 4 μM, 40 μM final concentrations). Each THZ531 addition was in the same volume of DMSO as in lane 2.

D. Shown is a pulse-chase assay using the DMSO solvent alone or the PARP-1 inhibitor PJ34 (80 μM final concentration). Vertical black line between lanes 1 and 2 indicates a deletion of intervening lanes relative to the original autoradiograph.

E. Shown is a pulse-chase assay of two nuclear extracts made as indicated in Materials and Methods (lanes 1 and 2) and compared to an extract made with an ammonium sulfate based protocol (lane 3) (Adamson et al. 2003).

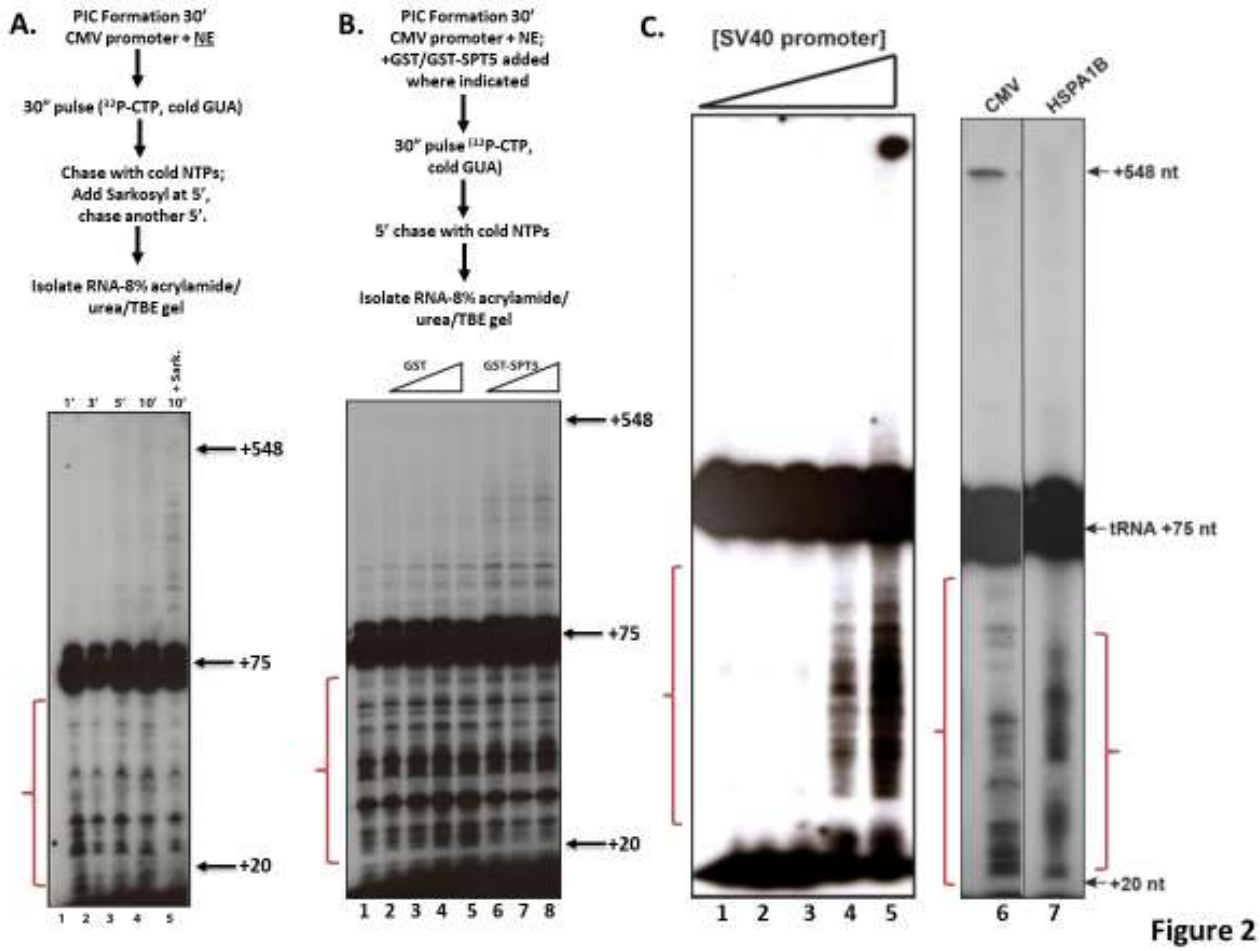

Figure 2. Sarkosyl and excess SPT5 elicit release of paused pol II.

A. Pausing system is sensitive to Sarkosyl and paused pol II does not chase. On the left is a schematic of steps in the experiment in this panel. Shown are pulse-chase assays allowed to chase for either 1, 3, 5, or 10 minutes (lane 1-4) or for 5', followed by the addition of 0.6% final concentration of Sarkosyl, followed by an additional 5' chase (lane 5).

B. Pol II is released by excess SPT5. Shown on the left is a schematic of steps in the pulse-chase assay on the right. Equivalent amounts of rGST or rGST-SPT5 were titrated into PICs that had been previously formed for 30'. After a 15' incubation, the pulse-chase assay was performed on each titration.

C. SV40 and heat shock promoters establish a paused pol II. SV40 promoter DNA (10, 25, 50, 100, 200 ng; lanes 1-5, respectively) was incubated with nuclear extracts for 30' followed by a standard pulse-chase assay. Lane 6 is a CMV control pulse/chase assay and lane 7 is a pulse/chase assay using the human HSPA1B promoter. These two lanes were merged from the same autoradiograph. Transcription products were analyzed on an 8% urea/TBE gel followed by autoradiography.

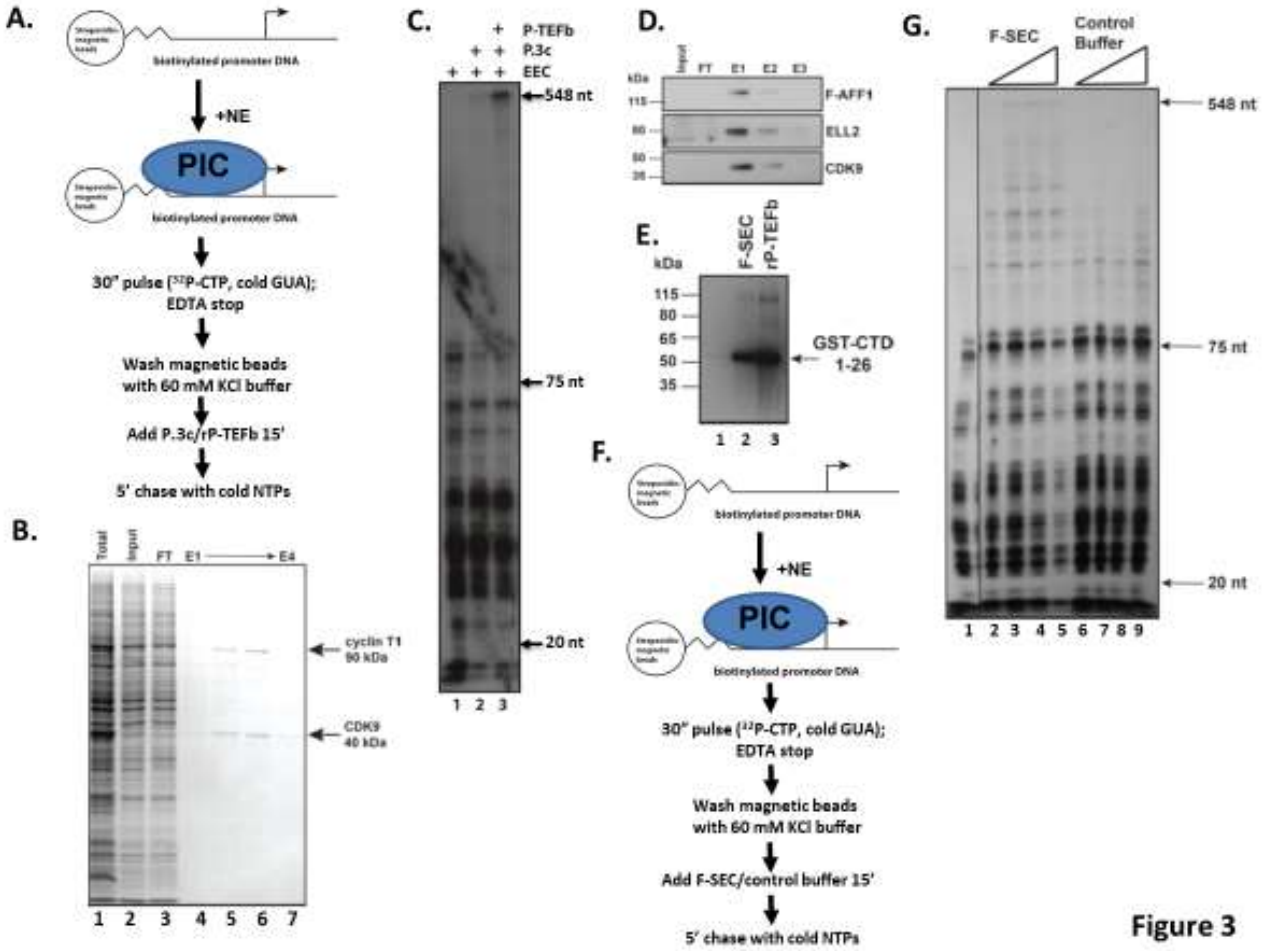

Figure 3

714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730

731 **Figure 3. Super Elongation Complex (SEC) and P-TEFb stimulate release of paused RNA pol II.**
732 A. Shown is a schematic of steps in the immobilized template assays in panels C and F.
733 B. Purification of rPTEFb. Baculovirus expression vectors containing human P-TEFb and cyclin
734 T1 were expressed in Sf9 cells and purified by Ni-Sepharose affinity chromatography.
735 C. Pol II elongation is stimulated by rP-TEFb. Pulsed immobilized templates were isolated after
736 they had been isolated and washed to remove ^{32}P-CTP/GUA and nuclear extract. The indicated factors
737 (P.3c is a concentrated P11 0.3M elution that contains some elongation activity-see lane 2) were added
738 to immobilized templates and incubated for 15', followed by a chase with unlabeled rNTPs. rP-TEFb
739 plus a P11 elongation fraction (P.3c) was added to immobilized templates and compared to either no
740 additional factors (lane 1) or the P.3c fraction alone (lane 2) as indicated in panel A. Transcription
741 products were analyzed on an 8% urea/TBE gel followed by autoradiography.
742 D. Purification of F-SEC. Flag-tagged SEC subunit AFF1 was purified from 293 cells using M2-
743 agarose and eluted by excess Flag peptide. Input extract contained undetectable levels of SEC subunits,
744 but which were concentrated by affinity purification (E1-E3), as indicated by western blot analysis for
745 SEC subunits AFF1, ELL2, and CDK9.
746 E. Kinase activity of F-SEC was assessed by incubating purified F-SEC (lane 2) or rP-TEFb
747 (positive control, lane 3) with the known substrate GST-CTD, containing the first 26 repeats of human
748 pol II CTD. Phosphorylation was detected by western blot using an anti-phosphoserine 2 CTD antibody.
749 Lane 1 contains only the GST-CTD and the kinase reaction buffer.
750 F. Schematic shows the steps in the immobilized template assay in panel G.
751 G. F-SEC stimulates release of paused pol II. Either purified F-SEC or the control elution buffer
752 were titrated into pulsed immobilized templates after they had been isolated and washed to remove ^{32}P-
753 CTP /GUA and nuclear extract. After a 15' incubation, unlabeled NTPs were added for 5' to chase any
754 elongation competent pol II into making full-length RNAs. Transcription products were analyzed on an
755 8% urea/TBE gel followed by autoradiography. Vertical black line between lanes 1 and 2 indicates a
756 deletion of intervening lanes relative to the original autoradiograph.

757
758
759
760
761
762
763

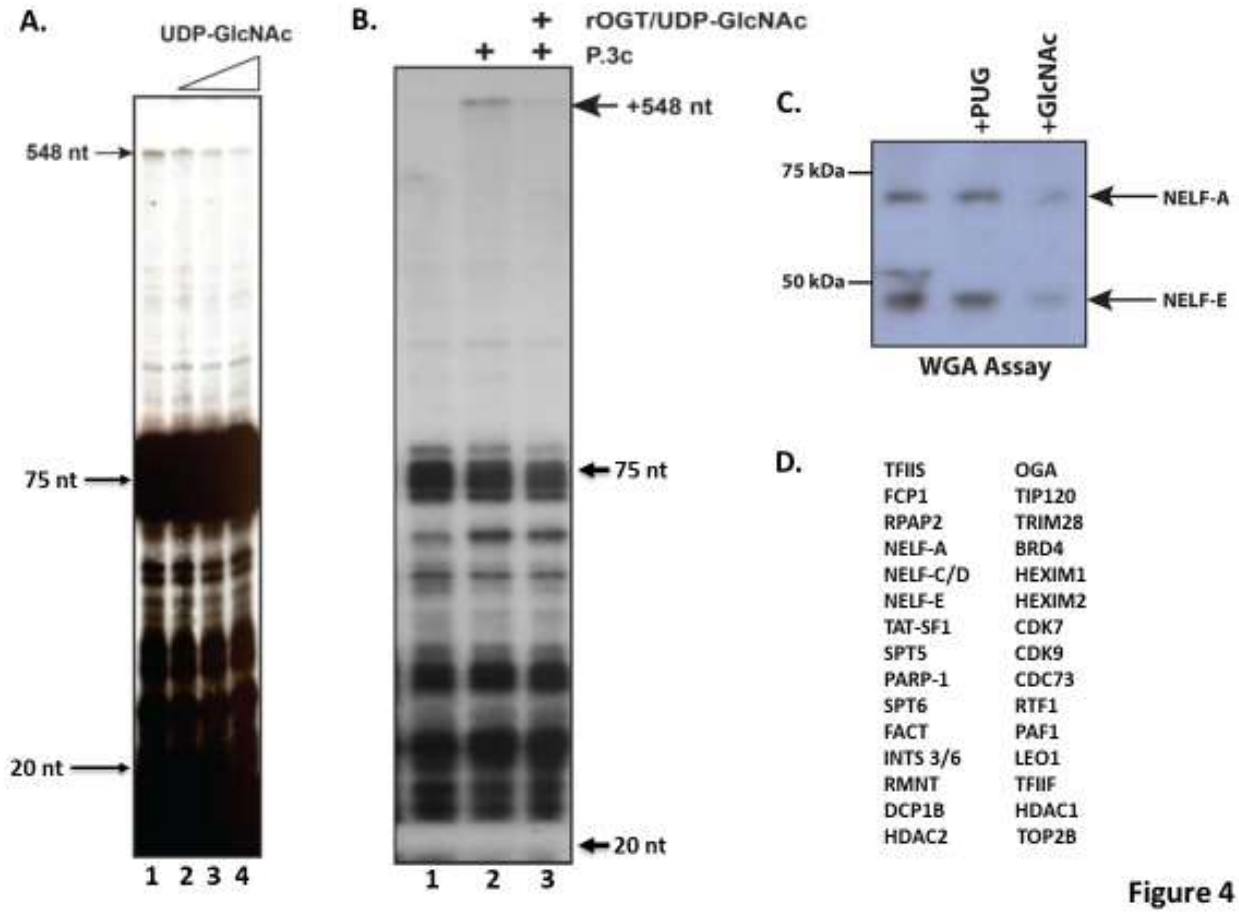

Figure 4. O-GlcNAcylation activity blocks RNA pol II elongation.

A. UDP-GlcNAc causes a decrease in elongation in nuclear extracts. Standard pulse-chase assays were performed after incubating preformed PICs with increasing amounts of the OGT substrate UDP-GlcNAc (0.1, 0.2, 0.5, and 1 mM final concentrations).

B. UDP-GlcNAc decreases elongation of EEC templates. Immobilized templates and pulsed EECs were isolated and washed in low salt transcription buffer. To the indicated reactions, either P.3c or P.3c plus 0.25 μg rOGT and 0.4 mM UDP-GlcNAc were added for 15' prior to chasing labeled RNAs for 5' with unlabeled NTPs.

C. Pausing factor NELF is O-GlcNAcylated. Nuclear extracts (incubated as indicated with either 3 mM PUGNAc or 0.5M GlcNAc) were incubated with wheat germ agglutinin-agarose for 1 hour and washed several times. The recovered agarose beads were eluted in sample buffer, separated by PAGE, and transferred to nitrocellulose. NELF subunits were detected by western blot.

D. Shown is a list of O-GlcNAcylated pausing and elongation factors that have been identified by mass spectroscopy(Hahne et al. 2012, 2013).

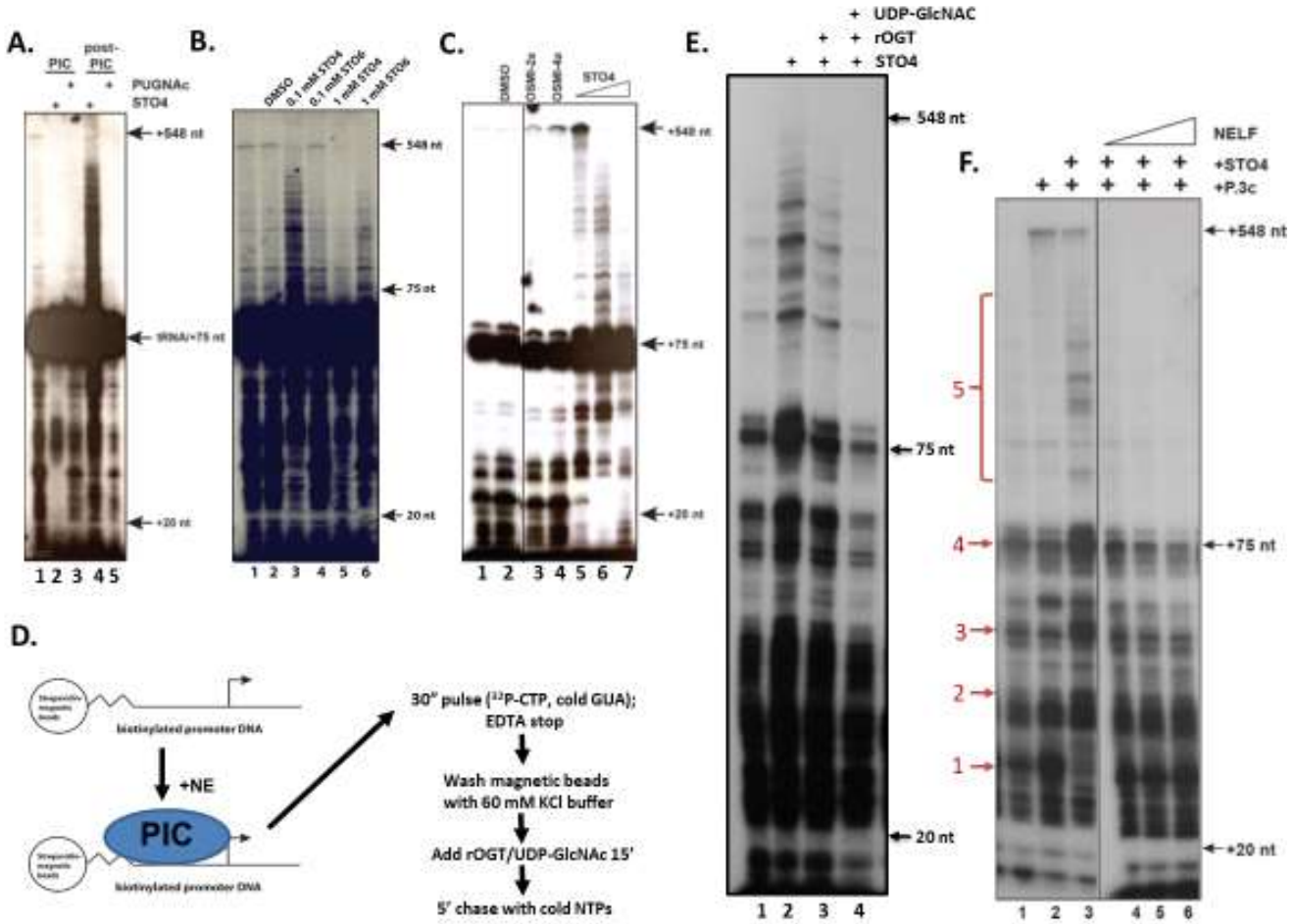
A.
PIC
post-PIC
PUGNAc
STO4
+548 nt
tRNA/+75 nt
+20 nt
1 2 3 4 5
B.
DMSO
0.1 mM STO4
0.1 mM STO6
1 mM STO4
1 mM STO6
548 nt
75 nt
20 nt
1 2 3 4 5 6
C.
DMSO
OSM-2a
OSM-4a
STO4
+548 nt
+75 nt
+20 nt
1 2 3 4 5 6 7
D.
biotinylated promoter DNA
+NE
PIC
biotinylated promoter DNA
30" pulse (32P-CTP, cold GUA); EDTA stop
Wash magnetic beads with 60 mM KCl buffer
Add rOGT/UDP-GlcNAc 15'
5' chase with cold NTPs
E.
+ UDP-GlcNAC
+ rOGT
+ STO4
548 nt
75 nt
20 nt
1 2 3 4
F.
NELF
+STO4
+P.3c
+548 nt
+75 nt
+20 nt
5
4
3
2
1
1 2 3 4 5 6

Figure 5

796 **Figure 5. OGT is required to establish a paused pol II.**

797 A. OGT inhibitor addition after PIC formation causes release of paused pol II. 40 μM ST04 or
798 3mM PUGNAc were added simultaneously with CMV promoter DNA or 30' afterwards (to allow PICs
799 to form). These reactions were then assayed by standard pulse-chase assay. Transcription products
800 were analyzed on an 8% urea/TBE gel followed by autoradiography.

801 B. OGT inhibition causes release of paused pol II. The indicated final concentrations of either ST04
802 or ST06 were added to preformed PICs which were then assayed by pulse-chase. Transcription products
803 were analyzed on an 8% urea/TBE gel followed by autoradiography.

804 C. Three OGT inhibitors affect pause release. Either DMSO, 80 μM OSMI-2a or 160 μM OSMI-
805 4a were added after PIC formation followed by a standard pulse-chase assay (lanes 1-4). Lanes 5-7
806 show a titration of STO4 (40, 80, 200 μM final concentrations). Transcription products were analyzed
807 on an 8% urea/TBE gel followed by autoradiography. Vertical black line between lanes 2 and 3 indicates
808 a deletion of intervening lanes relative to the original autoradiograph.

809 D. Schematic shows the steps in the immobilized template assay in panel E.

810 E. rOGT rescues the OGT inhibitor ST04 aberrant elongation effect. Immobilized templates were
811 created by allowing PICs for form for 30', followed by a 15' incubation with 0.1 mM ST04 and then a
812 30'' pulse with ^{32}P-CTP /GUA followed by addition of EDTA to stop transcription. Beads were washed
813 as indicated to remove all rNTPs, ST04, and nuclear extract. Individual reactions were suspended in
814 transcription buffer and either rOGT or rOGT and 0.4 mM UDP-GlcNAc were added as indicated, for
815 15', followed by a 5' chase with unlabeled NTPs. Transcription products were analyzed on an 8%
816 urea/TBE gel followed by autoradiography.

817 F. NELF complex reverses the ST04-induced pause release. Immobilized template assays were
818 set up similarly to panels D and E, but with the addition of the P.3c elongation fraction ± ST04 after the
819 isolation of the pulsed EEC template, followed by the titration of F-NELF (lanes 4-6). After a 15'
820 incubation period, cold NTP were added as the chase step. Transcription products were analyzed on an
821 8% urea/TBE gel followed by autoradiography. Vertical black line between lanes 3 and 4 indicates a
822 deletion of intervening lanes relative to the original autoradiograph.

823

824

825

826

827

828

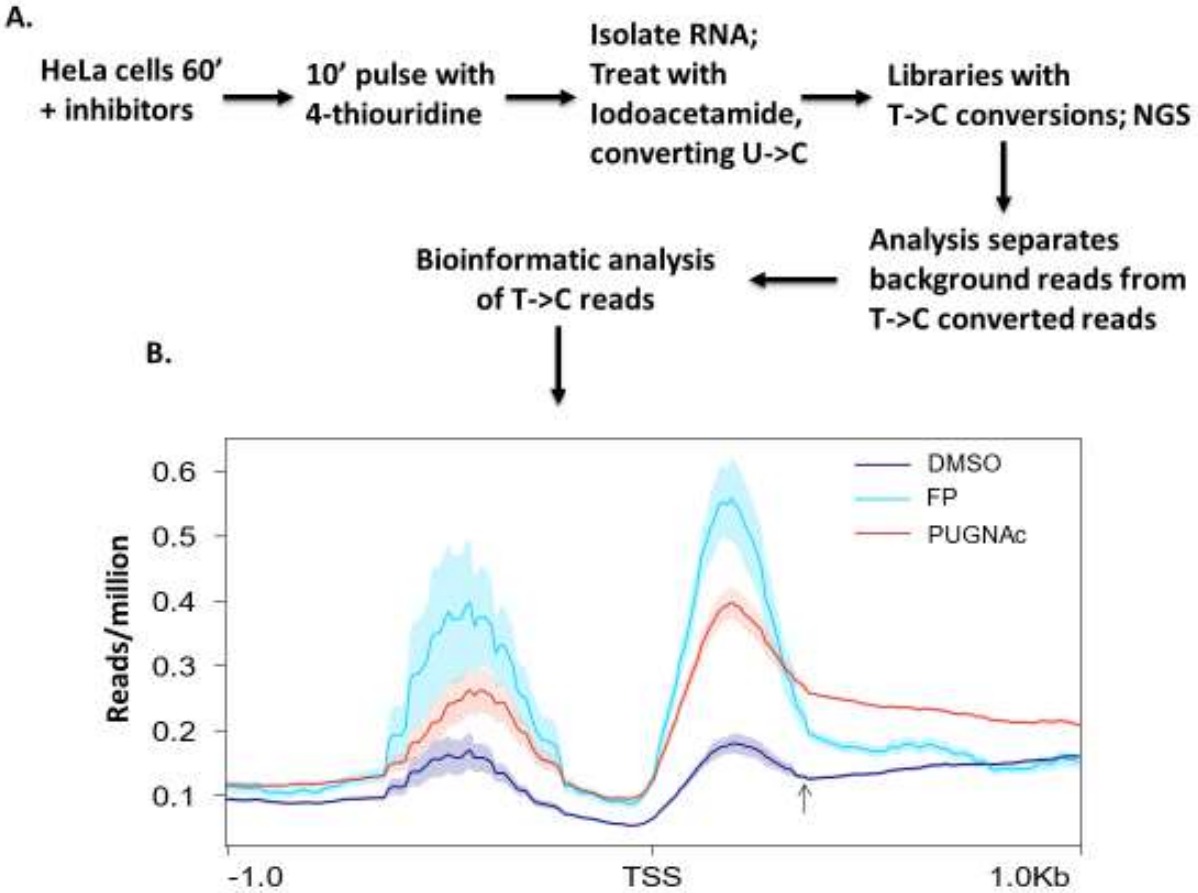

Figure 6

Figure 6. Inhibition of the O-GlcNAc aminidase blocks elongation in vivo.

A. Schematic of the 4-thiouridine experiment.

B. Metagene analysis of 4sU-incorporated RNA, mapping 4sU-RNAs based on distributions relative to the transcriptional start site (TSS) of genes, normalized to reads per million.

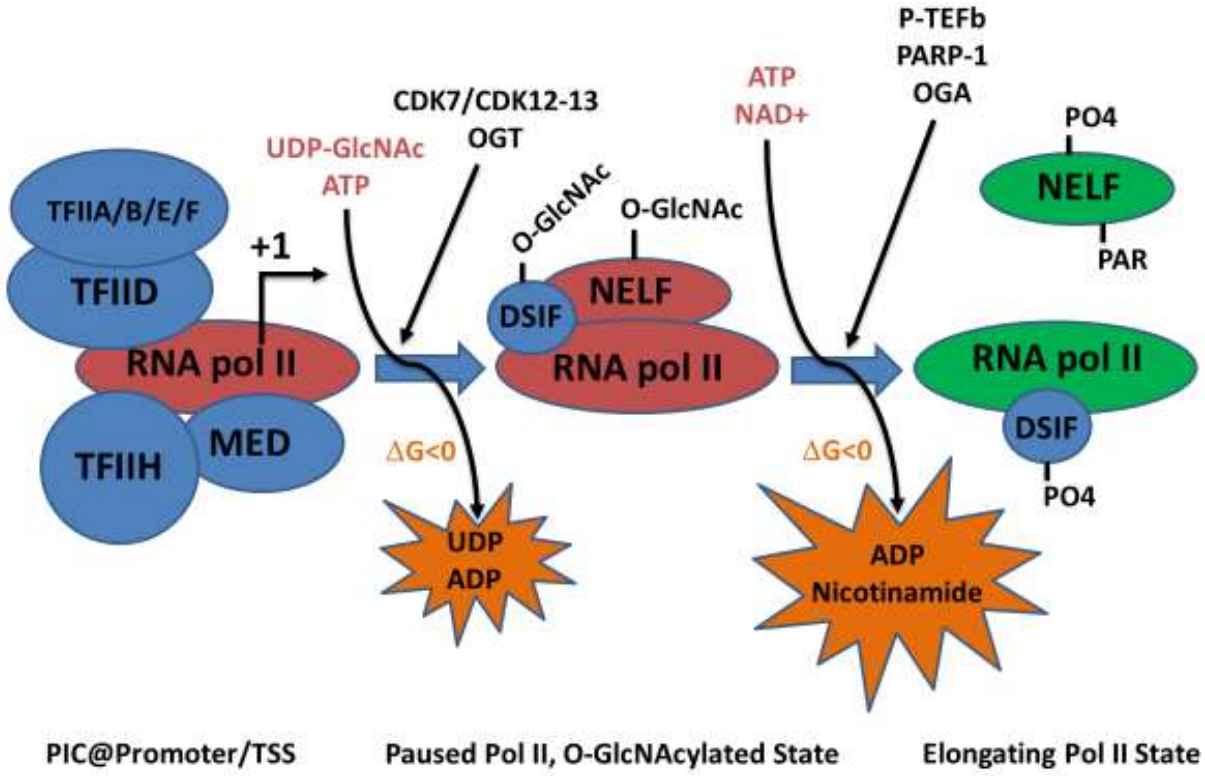

Figure 7

Figure 7. A multi-factor release model of RNA polymerase II pausing and elongation. The model proposes that there is a paused pol II complex in an O-GlcNAcylated state, partially defined by O-GlcNAcylated pausing factors DSIF and NELF and established by the concomitant catalytic activities of CDK7 (Larochelle et al. 2012), CDK12/13, and OGT. Release into productive elongation occurs by the action of P-TEFb (Guo and Price 2013), PARP-1 (Gibson et al. 2016), and OGA (Resto et al. 2016). This removes GlcNAc from the targeted proteins, poly-ADP ribosylates NELF (Gibson et al. 2016), possibly inducing NELF release and/or inactivation, and phosphorylates DSIF (Fujinaga et al. 2004; Ping and Rana 2001; Yamada et al. 2006; Peterlin and Price 2006), which converts DSIF into a positive elongation factor. The orange "bursting star" surrounding the various metabolites indicates that these hydrolysis products are associated with the release of free energy.