Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Elizabeth Anne Bowman Date

Regulation of Transcription by RNA Polymerase II in *C. elegans*

By

Elizabeth Anne Bowman
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology

_____

William G. Kelly, Ph.D.
Advisor

_____

Jeremy M. Boss, Ph.D.
Committee Member

_____

Xiaodong Cheng, Ph.D.
Committee Member

_____

John Lucchesi, Ph.D.
Committee Member

_____

Daniel Reines, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

Regulation of Transcription by RNA Polymerase II in *C. elegans*

By

Elizabeth Anne Bowman
B. S., Belmont University, 2008

Advisor: William G. Kelly, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2013

Abstract

Regulation of Transcription by RNA Polymerase II in *C. elegans*
By Elizabeth Anne Bowman

Virtually all cells within a multicellular organism contain the same DNA. Despite this, multicellular organisms are made of different cell types, and the development of different cell types from the same DNA is due to differential expression of protein coding genes. The first step of gene expression is transcription in which RNA Polymerase II (Pol II) copies a gene into an RNA transcript. Transcription involves three steps, 1) initiation, or selection of a transcriptional site, 2) elongation, or the productive generation of RNA Pol II, and 3) termination of transcription. While these steps of transcription have been well studied in single cell systems, the regulation of transcription during the development of multicellular organisms is not well characterized. Therefore, I have studied transcriptional regulation during development in the multicellular organism, *C. elegans*.

In this thesis, I have analyzed mutations in the large catalytic subunit of *C. elegans* Pol II, called AMA-1, and have mapped these onto the structure of the yeast protein to identify structural components required for normal function and regulation of this protein. In pursuing further characterization of the role of Pol II during development, I discovered a form of tissue-specific regulation of transcriptional elongation in *C. elegans*. The transition from Pol II initiation to elongation is accompanied by phosphorylation of serine 2 (Ser2) of the Pol II carboxy terminal domain (CTD), which was thought to be carried out by the canonical Ser2 kinase, CDK-9. However, I have uncovered CDK-9 independent Ser2 phosphorylation in the *C. elegans* germline. Instead, this modification is performed by the recently discovered Ser2 kinase, CDK-12. This is in contrast to somatic tissues where both CDK-9 and CDK-12 contribute to Ser2 phosphorylation. I observed that neither CDK-12 nor Ser2 phosphorylation are required in the germline for fertility. These studies are the first to show tissue-specific differences in the regulation of Ser2 phosphorylation, a marker of Pol II elongation. Overall, I have generated a novel model of tissue-specific transcriptional regulation during *C. elegans* development, further defining the regulatory mechanisms of this “general” process in multicellular organisms.

.

Regulation of Transcription by RNA Polymerase II in *C. elegans*

By

Elizabeth Anne Bowman
B. S., Belmont University, 2008

Advisor: William G. Kelly, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2013

Table of Contents

CHAPTER 1: INTRODUCTION

Eukaryotes have three DNA-dependent RNA polymerases, designated RNA polymerase I, II, and III (*1*). Genes transcribed by RNA polymerase I and III are limited to the non-coding ribosomal and transfer genes, respectively (*2-4*). RNA polymerase II (Pol II[1]) is a twelve subunit complex that transcribes protein coding genes and many non-coding RNA (ncRNA) genes. Although differential expression of some ncRNAs is increasingly recognized for its importance in biological processes (*5*), regulation of protein expression is still largely considered to lay at the heart of development, and thus gene transcription by Pol II is tightly controlled in the cell. In addition to the "where" and "when" of activity, Pol II transcription also requires efficient navigation through a chromatin template as well as coordination with RNA processing machinery. While some of these challenges also extend to the other RNA polymerases, proper Pol II transcription faces many unique challenges, as described below.

The process of transcription is commonly described as having three phases: 1) initiation, which includes transcription start site selection and the formation of the first few phosphodiester bonds in the RNA transcript, 2) elongation, or navigation of RNA polymerase through a locus as it lengthens the RNA transcript, and 3) termination, or the release of polymerase when it reaches the end of the gene being transcribed. For appropriate gene expression, recruitment of Pol II to a locus is perhaps the most conceptually straightforward mode of regulation. Thus, Pol II initiation was thought to be the major mode of transcription regulation for many years. Early studies

[1] Commonly used abbreviations- Pol II: RNA Polymerase II; ncRNA: non-coding RNA; CTD: carboxy terminal domain, referring to the Pol II CTD; CTR: carboxy terminal repeat, referring to the Spt5 CTR; AA: amino acid; Ser: serine; Ser2-P: Pol II CTD serine 2 phosphorylation; Ctk: carboxy terminal kinase, Ctk1/2/3 is a Ser2 kinase complex; Bur: bypasses upstream requirement, Bur1/2 is a Ser2 kinase complex; Lsk: latrunculin sensitive kinase, Lsk1 is a Ser2 kinase; Cdk: cyclin dependent kinase, Cdk9, 12, and 13 are Ser2 kinases; Spt: suppressor of Ty's, Spt4/5 makes up the DSIF transcription elongation complex; DSIF: DRB sensitive inhibitory factor; DRB, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole, transcription elongation inhibitor; NELF: negative elongation factor; P-TEF: positive transcription elongation factor, P-TEFb is composed of Cdk9 and Cyclin T; PPP: Pol II promoter proximal pausing; TBP: TATA binding protein; TAF: TBP-associated factor; PGC: primordial germ cell; ORF: open reading frame; UTR: untranslated region; MOS-SCI: Mos transposon-mediated single copy insertion; INTACT: isolation of nuclei tagged in specific cell types; BLRP: biotin ligase recognition peptide

showed that purified Pol II cannot accurately initiate transcription at gene start sites *in vitro*. Thus, it became clear that accurate initiation requires additional factors to recruit Pol II to transcription start sites. [reviewed in (*6*)]. Many very important and historic studies in the 1980s and early 1990s set out to identify the so called "general transcription factors" (GTFs) important for Pol II initiation [this has been extensively reviewed, but for contemporary reviews see (*7*)]. The identification of GTFs started with the establishment of a system by the Roeder group that had faithful initiation on a template (*7*), followed by purification of the factors required for this activity by this group and the Sharp group (*8*). Today, we understand the regulation of Pol II initiation by the GTFs at a molecular level (*9*).

During the past twenty years, as we have developed a clear picture of Pol II initiation regulation, attention has shifted to studying the regulation of the elongation phase of Pol II transcription. Because there are many barriers to Pol II elongation, Pol II recruited to genes is not initially competent for efficient transcription [reviewed (*10*)]. Instead, Pol II cycles through a few rounds of abortive transcription, creating short (2-14 nt) RNA products, until it becomes competent for elongation. Following this, Pol II faces additional challenges to processivity and is further regulated by positive and negative elongation factors. These factors both directly regulate Pol II activity or regulate association of other transcription factors (discussed below). Today, we know a great deal about regulation of Pol II elongation (specific reviews are cited through the text), but I am going to focus the remainder of my discussion on a few, highly conserved general elongation factors along with Pol II itself.

The Pol II CTD

Early purifications of Pol II identified three major migrating forms of Pol II (IIa, IIb, IIo) [a contemporary review (*11*)]. While the difference in these forms was not initially clear, biochemical studies in the 1980's showed that Pol II elongation is carried out by the highest migrating form, IIo, which is a phosphorylated form of a lower migrating form, IIa (*12*). At about

the same time, cloning studies of the largest subunit of Pol II (*13*), identified a unique, repeated motif on the C terminal end, now called the "C terminal domain", or CTD, composed of dozens of repeats of the heptapeptide sequence Y_1-S_2-P_3-T_4-S_5-P_6-S_7. The overwhelming majority of phosphorylation present in the Pol II IIo form is found within the CTD (*12*). Conversion from initiated Pol II to Pol II that is actively elongating is accompanied by a conversion of the unphosphorylated IIa form to the phosphorylated IIo form *in vitro* (*14*) and *in vivo* (*15*). These observations established a model that CTD phosphorylation is important for Pol II elongation *in vivo*, although this is not necessarily required for transcription *in vitro* (*16*).

It should be noted that this CTD is commonly referred to as the "Pol II CTD". This name is technically a misnomer. As mentioned above, Pol II is a multisubunit complex and the repetitive CTD is found within the large subunit of Pol II; however, because this subunit is only found within the Pol II complex, and because the other Pol II subunits do not have similar C terminal repeated extensions, I will continue to use this term here.

Since the early studies of the Pol II CTD, which were initially performed with radioactive labeling of Pol II purified by chromatography, many studies over the past 25 years that have utilized antibodies against CTD modifications, CTD mutants, or mass spectrometry analysis, have established a model of CTD phosphorylation through the transcription cycle. Briefly, and in general, Pol II recruited to a gene is hypophosphorylated on the CTD and becomes phosphorylated on serine 5 (Ser5) of the repeat during the transition from initiation to early elongation and phosphorylated on the CTD's serine 2 (Ser2) during productive elongation [reviewed (*17, 18*)]. While the CTD can be modified in many other ways, such as proline isomerization (*19*), threonine phosphorylation (*20*), arginine methylation (*21*), glycosylation (*22*), and serine 7 phosphorylation (*23*), Ser5 and Ser2 phosphorylation are the best studied and appear to be the most conserved general marks of transcription. Differential phosphorylation of Ser2 and Ser5 through the transcription cycle plays an important role in regulation of gene expression by

recruiting transcription associated proteins at the appropriate phase of the transcription cycle [discussed below and (*17*)].

The remainder of this introduction focuses on the transition of Pol II from initiation to productive elongation and the factors that positively and negatively regulate that step, thus much of my discussion will be focused on phosphorylation of Ser2 of the Pol II CTD. Other useful reviews on different phases of Pol II elongation can be found (*18*). The importance of the factors that regulate this transition is highlighted by the fact that mutations of many of these genes have been associated with human diseases including, but not limited to, HIV-1 (*24*), leukemia (*25*), NUT midline carcinoma (*26*), other cancers (*27*), cardiac hypertrophy (*28, 29*), and other diseases (*30*).

Positive elongation factors: Pol II CTD Serine 2 kinases

All eukaryotes studied thus far contain two different Ser2 kinases, Bur1/CDK9 and Ctk1/Lsk1/CDK12 (Table 1.1). Until the last 4 years, most studies of Ser2 phosphorylation in eukaryotic systems only focused on one of these kinases. In budding yeast, Ctk1 was considered to be the sole Ser2 kinase and Bur1 was not thought to directly phosphorylate the CTD. In higher organisms, the Bur1 homolog, CDK9, was considered the major CTD kinase while the role of the Ctk1 homolog, CDK12, was not explored. Retrospectively, it is interesting to consider that the proteins thought to be the major Ser2 kinases, Ctk1 in yeast and CDK9 in higher eukaryotes, are not the closest sequence homologs in these two organisms. The discrepancy between the roles of these kinases across eukaryotic systems was not explored until recently. Studies in the past 4 years have shifted the paradigm of Ser2 phosphorylation, converging on a model in which the role of the Ser2 kinases is quite conserved from yeast to humans (Figure 1.1B,C). Here, I am going to review the early studies carried out in different systems as well as the current model of Ser2 phosphorylation regulation by both Ser2 kinases.

The yeast Ser2 kinases: Ctk1/Lsk1 and Bur1/Cdk9

Many different CTD kinases were discovered in the 1990s [reviewed (*31*)]. The first CTD Ser2 kinase was identified in biochemical studies in budding yeast (*32*) and later cloned as a cyclin-dependent kinase, Ctk1 (*33*). Early studies by the Greenleaf lab determined that that this kinase is found in a complex containing Ctk1, its cyclin partner, Ctk2, and an additional complex component, Ctk3, which is important for stability of the complex (*34, 35*). While null mutants of these components are viable, they all show similar cold sensitive phenotypes (*34*). Although initially shown to directly affect transcription elongation rate *in vitro* (*36*), studies through the mid-2000s focused on the role of the Ctk1 target, Ser2 phosphorylation, in recruitment of RNA processing factors and chromatin-associated factors (discussed in detail below). The fission yeast homolog of Ctk1 is called Lsk1 as it was first identified as a mutant that causes increased sensitivity to a cell cycle drug, latrunculin A [Lsk1: latrunculin sensitive kinase (*37*)]. Similar to budding yeast *ctk1* mutants, fission yeast *lsk1* mutants are conditionally lethal (*38*) (Figure 1.1A).

The other budding yeast Ser2 kinase complex components, Bur1 and Bur2, were first identified as transcriptional regulators in 1993 using a screen to identify factors that could bypass the need for upstream activating sequences (UAS) of a reporter gene [BUR, bypass UAS requirement (*39*)]. These proteins were found to comprise a cyclin dependent kinase complex in which Bur1 is the kinase and Bur2 is the cyclin (*40*). While genetic studies linked Bur1 activity with phosphorylation of the CTD (*41*), it wasn't shown to be a direct CTD kinase *in vivo* until several years later (*42, 43*).

bur1 null mutants were initially characterized as lethal (*44*), but were later found to be extremely slow growing (*45*). Similarly, *bur2* null mutants are also viable but slow growing (*40*). Interestingly, an analog sensitive mutant of *bur1* (discussed below) is viable and healthy (*42*). This could be due to incomplete inactivation of kinase activity *in vivo* as the authors of this study suggest, although no kinase activity from this mutant was detected *in vitro*. The fission yeast Bur1

homolog, Cdk9, named due to its similarity to Cdk9 in higher eukaryotes, is an essential protein as mutants are inviable (*46*) (Figure 1.1A).

Bur1/Cdk9 and Ctk1/Lsk1 regulate yeast Ser2 phosphorylation together

Until recently, Ctk1 was thought to be the major Ser2 kinase in budding yeast (*47*); however early analysis of the roles of Pol II Ser2 kinases were complicated by the specific mutants used for each. Specifically, Ctk1 is not an essential protein, and null deletion mutants clearly showed that Ctk1 regulates Ser2 phosphorylation levels *in vivo* (*47*). In contrast, Bur1 is essential and studies suggesting that Bur1 does not regulate Ser2 phosphorylation levels used point mutants that may only partially affected kinase activity (*41*).

Recent studies have moved away from classical genetics to chemical genetics approaches. This approach has improved the understanding of the direct targets of these kinases and has even reversed our understanding of the essential nature of these kinases. Kinase chemical genetics was pioneered by the Shokat group [reviewed (*48*)]. These studies utilize a mutation in the kinase catalytic domain that allows normal function under standard conditions but renders the kinase inactive when exposed to a bulky ATP analog that does not fit into the smaller mutant ATP binding pocket. These mutants are thus called analog sensitive, or "*as*" mutants. In contrast to using kinase inhibitors that may have off target effects, introducing this ATP analog into cell culture specifically inhibits the mutant kinase. Finally, these mutants can be used to refine the identification of endogenous kinase targets (*49*).

Ctk1 was initially shown to be responsible for detectible Ser2-P in yeast during the diauxic shift (*47*), and although this was the focus for 15 years, this same study also showed that only about 50% of Ser2-P was decreased in *ctk1* mutants during logarithmic growth. A recent study using a chemical genetics approach showed that Bur1 is the kinase responsible for the remaining Ser2 phosphorylation during logarithmic growth (*42*). While previous studies did not detect a role for Bur1 in Ser2 phosphorylation (*41*), this recent study used a viable *as-bur1* mutant to

show that inhibition of activity leads to ~60% reduction in Ser2-P. Furthermore, they showed that the very trace amounts (~10%) of Ser2-P they detected in a *ctk1* deletion strain during diauxic shift are lost with Bur1 inhibition. In summary, while Ctk1 is responsible for the majority of Ser2-P, loss of Bur1 activity shows more dramatic decreases in Ser2-P than the level it is directly responsible for; i.e. there is a 60% reduction in a *bur1* mutant, but it is only responsible for the trace amounts of Ser2-P in a *ctk1* mutant. These authors suggested an epistatic relationship between Bur1 and Ctk1 that places Bur1 upstream of Ctk1 during transcription of a gene. This conclusion is supported by the fact that in a *ctk1* mutant, Ser2-P at the 3' end of a gene is more affected than 5' Ser2-P. These results overall suggest a model in which both Bur1 and Ctk1 contribute to Ser2-P, but that Bur1 activity on Pol II is upstream of, or more 5' than, Ctk1 and that Bur1 activity stimulates Ctk1 activity (Figure 1.1B).

Fission yeast regulate Ser2-P in a similar fashion. While an overwhelming majority of detectible Ser2-P is absent with loss of activity of the Ctk1 homolog, Lsk1 (*38, 50*), recent data using analog sensitive mutants has also shown that fission yeast Cdk9 is responsible for some level of Ser2-P (*51, 52*). This effect of Cdk9 loss on Ser2-P levels was also shown in a different study which interpreted this result differently (*50*). The authors concluded that the loss of Ser2-P in *cdk9* mutants may be due to decreased elongation rather than through a direct role for Cdk9 in Ser2 phosphorylation (Figure 1.1B).

It should be noted that Ctk1 and Bur1 have other functions besides phosphorylation of the Pol II CTD. Bur1 has other well characterized targets including the transcriptional elongation factor, Spt5 (discussed below) and DNA repair protein, Rad6 (*53*). Bur1 has also been shown to regulate genome integrity through regulation of the replication protein A (*54*). Finally, there have been several reports of roles for Ctk1 in rRNA transcription (*55-57*).

Yeast Ser2 kinases regulate transcription-associated factor recruitment

Many of the studies of the yeast Ser2 kinases have focused on their role in the regulation of RNA processing and chromatin modifying complexes. While most relationships between these kinases and transcription associated processes are likely to be indirect through interactions with the kinase product, Ser2-P, in fission yeast the C terminus of Cdk9 directly binds to the mRNA capping factor, Pct1 (*52, 58*). This domain is conserved in the homologous budding yeast protein, Bur1, and may also help coordinate mRNA capping; however, it is not conserved in higher organisms where capping is instead mediated through Ser5 phosphorylation (*59, 60*). Capping may also be redundantly regulated through Ser5-P in fission yeast as the Cdk9 C-terminus domain is not essential, and its removal only causes conditional growth defects (*58*). The similar domain is also not essential in the budding yeast Bur1 (*41*).

Ctk1 genetically interacts with 3' processing factors (*61*), and this interaction is likely mediated through a direct physical interaction between these RNA processing factors and Ser2-P for appropriate timing of Pol II termination. While these factors can be recruited to genes in the absence of Ctk1, changes in the polyA site choices are observed (*62*). As opposed to Cdk9 in other systems (described below), Ctk1 and Bur1 are not required for cotranscriptional recruitment of splicing factors (*63*). However, efficient splicing may not be as tightly regulated in yeast since only 5% of genes are spliced in budding yeast and only a handful of those are alternatively spliced (*64*). Thus, as is the case in higher eukaryotes (discussed below), Ser2 kinases regulate interaction of at least a subset of RNA processing factors with the transcription machinery.

Ctk1 has been shown to regulate recruitment of chromatin factors to sites of transcription, supporting a role for Pol II CTD phosphorylation in recruitment of chromatin modifying factors. The most well studied interaction is with the histone 3 lysine 36 (H3K36) histone methyl transferase, Set2. Set2 is recruited to the hyper-phosphorylated CTD (*65*) through its C terminus (*66*) to regulate H3K36 trimethylation (H3K36me3) in the body of genes (*67*). Ctk1 is also required for COMPASS-mediated H3K4me1 levels (*68*) and Ctk1 activity prevents the spread of

transcription start site-associated H3K4me2/3 into gene bodies (*69*). Thus, Ctk1-mediated Ser2-P aids in the recruitment of histone modifying complexes to coordinate the Pol II elongation stage with the appropriate histone marks.

Surprisingly, although Ctk1-mediated Ser2-P appears to coordinate transcription with RNA processing factors and histone modifiers, *ctk1* is not an essential gene (*33*) and has a minimal effect on gene expression by microarray analysis (*61*). The only major transcriptional defect observed in a Ctk1 mutant is an abnormal accumulation of Pol II at the 3' end of genes (*70*). As Ctk1-mediated Ser2-P was shown to be important for 3' processing (see above), the authors suggest this accumulation compensates for slower 3' processing during Pol II termination in a *ctk1* mutant. Thus, many of the processes in which Ctk1-mediated Ser2-P has been implicated are not essential or can function fairly normally under optimal growth conditions in budding yeast.

While Ctk1 is not essential in budding yeast, the second position serine of the CTD heptad is required for viability (*71*). Perhaps the second position serine is essential because the minimal level of Bur1-dependent Ser2-P that remains in *ctk1* mutant strains is required for Ser2-P function (*42*). Importantly, Ser2 phosphorylation is not essential in fission yeast as both *lsk1* mutants and mutation of all serines in the second position of the CTD to alanine do not significantly affect yeast viability during mitotic growth (*50*).

Bur1/2 also regulates an overlapping but not identical subset of chromatin modifiers as Ctk1. Bur1/2 is involved in the H3K36 methylation pathway as Set2 mutants, and overexpression of the K36 demethylase Rph1, bypass the requirement for Bur1 (*72, 73*). H3K36me3 is reduced in *bur1* mutants (*45*), although it is unclear if Bur1 regulates H3K36me3 through its Ser2 kinase activity.

In addition to regulation of H3K36me3, Bur1/2 also regulates several other histone modifications including H2BK123 ubiquitination, H3K4 methylation, and possibly H3K79 methylation (*53, 74*). While the specific effects on some of these modifications are contested, the role of Bur1/2 in these marks is likely through regulation of recruitment of the polymerase associated factor complex, Paf1c, which binds elongating Pol II to mediate interaction of other

elongation and chromatin remodeling factors (*74*). In budding yeast, Paf1c is recruited through Bur1-mediated Pol II CTD and Spt5 phosphorylation (discussed below) (*75*) through three separate phospho-CTD interaction domains on the Paf1c components Cdc73, Rrf1 and Ctr9 (*76*). While the Bur1 homolog, Cdk9, is important for Paf1c recruitment in fission yeast, this may be specifically through Spt5 (discussed below) rather than the Pol II CTD, as the viable CTD S2A mutation does not affect levels of a downstream modification, H2BK123 ubiquitination (*77*).

In summary, both Ctk1 and Bur1 are important for recruitment of histone modifying complexes, and phosphorylation of both the Pol II CTD and Spt5 may play overlapping roles in the appropriate recruitment of these factors.

Regulation of yeast Ser2 kinases

Mirroring their roles in Ser2 phosphorylation, both Ctk1 and Bur1 are associated with elongating Pol II by ChIP (*78, 79*) (*41*), and this has been shown genome wide for Ctk1 (*80*). Thus, these kinases appear to play general roles in transcription, and unlike Cdk9 of higher eukaryotes (discussed below), their recruitment to loci may be through general mechanisms, rather than gene-specific factors. The C terminus of Bur1 contains a phospho-CTD interaction domain (amino acids #442-523), which interacts with Ser5-P to enhance Bur1 recruitment to genes (*42*). While Ctk1 has been suggested to interact with Pol II through a Bur1-phosphorylated CTD (*42*), Ctk1 is only capable of recruitment to loci following deubiquitylation of H2BK123 by Ubp8, which occurs in the wake of Pol II transcription (*81*). Because the factors involved in kinase activity identified thus far are not gene-specific regulators, recruitment of Ser2 kinases may not play a role in gene-specific regulation of Pol II elongation, unlike in higher eukaryotes (discussed below). Instead, these kinases seem to act largely as general factors that mark elongating Pol II for the recruitment of other factors at the proper stage of transcription.

The metazoan Ser2 kinases: CDK9 and CDK12

Until recent studies by the Greenleaf lab (*82*), CDK9 was thought to be the sole Ser2 kinase in metazoans and it was thought to carry out all functions attributed to the two yeast Ser2 kinases (*83*). When in complex with its cyclin partner, Cyclin T, CDK9 is also referred to as positive transcription elongation factor b, or P-TEFb, as it was first identified as having an ability to enhance Pol II elongation *in vitro* (*84*). CDK9 is a very well studied protein, in part because of early association of CDK9 function during HIV infection and more recently because of the role of CDK9 in the regulation of Pol II promoter proximal pausing (both discussed below). Because of this, CDK9 has been extensively studied including characterization of gene-specific recruitment of CDK9, the interaction of CDK9 with its binding partners, and regulation by post-translational modifications. While the role of CDK12 in Ser2 phosphorylation is more recently appreciated, its importance in proper transcription and development is already well recognized (described below).

Identification of CDK9

CDK9 was identified as a metazoan CTD kinase in 1990s by the Price and Rice labs as a factor that positively regulates transcriptional activity. The Price lab identified this P-TEF activity in *Drosophila* Kc cell nuclear extract using an *in vitro* transcription system sensitive to 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole, or DRB (*85*), a purine nucleoside analog previously shown to inhibit Pol II elongation (*86*). This positive activity was later identified as two complexes, called P-TEFa and P-TEFb (*87*), and P-TEFb was shown to be a Pol II CTD kinase (*84*). At the same time, studies of HIV-1 transcription in human cells showed that efficient transcription of HIV-1 is DRB-sensitive (*88, 89*). Shortly thereafter, the Rice lab identified a DRB-sensitive kinase required for HIV-1 transcription, and this kinase was shown to be capable of phosphorylating the Pol II CTD (*90*). Although found in different systems, these two activities were soon identified as homologous activities and the P-TEFb complex from both *Drosophila* and

human was shown to contain a cyclin dependent kinase, CDK9, (*91*) and a cyclin, Cyclin T (*92-94*).

While CDK9 has been suggested to associate with Cyclin K (discussed below), most of the studies of CDK9 have been with its association with Cyclin T homologs. *Drosophila* has one Cyclin T homolog, but humans have two, Cyclin T1 and T2, with Cyclin T2 alternatively spliced to form T2a and T2b (*92, 94*). Among the human cyclins, the N terminal cyclin boxes that bind CDK9 are highly similar (*95*), while the C terminal halves of these cyclins are fairly divergent, pointing to non-overlapping functions of these cyclins. In fact, Cyclin T1 and Cyclin T2 have been shown to be non-redundant as Cyclin T2 is essential in mouse (*96*) and there is little overlap between genes affected after knock down of either (*97*).

Early studies established the Pol II CTD as the major target of CDK9 within Pol II (*98, 99*); however, there have been conflicting reports about the specific residues targeted by CDK9. *In vitro* transcription assays initially suggested that CDK9 phosphorylates Ser5 (*100, 101*), but later studies using CTD constructs with more heptad repeats showed that this was likely an artifact of short CTD substrates in the early studies (*98, 102*). However, a more recent report also shows that CDK9 can in fact phosphorylate both Ser5 and Ser7 on long CTD substrates *in vitro* (*103*). The role of Ser2 as the CDK9 target is corroborated by *in vivo* studies that have shown that Ser2 phosphorylation but not Ser5 phosphorylation is affected when CDK9 is inactivated [*C. elegans* (*104*); *Drosophila* heat shock genes (*105*) and other active loci (*106*) on polytene chromosomes].While *Drosophila* CDK9 mutants have reduced levels of Ser2 and Ser5 phosphorylation (*107*), these authors suggest that the decrease in Ser5 might be indirect due to a negative effect on transcription. Finally, although not a general requirement for proper Pol II transcription, recent studies have also shown that CDK9-mediated Thr4 phosphorylation of the Pol II CTD is essential for proper histone expression (*20*).

CDK9/P-TEFb activity is regulated by its association with large complexes

CDK9 activity is tightly regulated in the cell and much of it is found within large regulatory complexes. These complexes have been extensively characterized in *Drosophila*, mouse, and human, and while these organisms appear to regulate CDK9 activity in a similar fashion, this is not the case for other eukaryotes, including yeast and *C. elegans*, which lack obvious homologs of the genes for most of these proteins (Table 1.1).

In *Drosophila* and mammals, about half of cellular P-TEFb is found within an inhibitory complex composed of an abundant non-coding RNA named 7SK, the HEXIM1 and HEXIM2 proteins, the LARP7 protein, and the 7SK capping protein, MEPCE [reviewed in (*108*)]. The 7SK RNA activity was identified by two groups who detected it as an RNase sensitive CDK9 cofactor that inhibited its activity (*109*), and as a component of the CDK9 complex detectible by silver stain but not coomassie stain (*110*). The HEXIM1/2 proteins were later found to be important components of this inhibition (*111, 112*), and they are only capable of inhibiting CDK9 activity in the presence of 7SK (*113, 114*). HEXIM1 and 2 both contain coiled coil domains that form dimers in the cell (*115, 116*). The inhibitory complex contains one 7SK molecule, two P-TEFbs, and a HEXIM1/2 heterodimer (*117*). LARP7, and the 7SK capping protein, MEPCE were later found to be an important component of this complex (*27, 118, 119*), which binds and methylates the 5’ end of 7SK RNA for its stability.

For many years, P-TEFb activity was thought to exist either in a “small” active complex composed of CDK9 and Cyclin T, or the “large” inhibitory complex with 7SK/HEXIM/LARP7/MEPCE. However, many studies have suggested that P-TEFb is most active when associated with another large complex, composed of many previously identified elongation factors, now deemed the “super elongation complex”, or SEC. This complex was identified through purification of aberrant fusion proteins that cause mixed-lineage leukemia in humans (*120*). These mutants include fusions between the human H3K4 methyltransferase, MLL, and several partners, often including ELL1, ENL, AFF1, and AF9 (*120*). In this study, it was

found that a set of common MLL fusion partners form a complex with P-TEFb to regulate transcriptional elongation. A similar complex was identified by others around the same time using a similar approach (*121, 122*), and through purification of CDK9 and HIV-1 TAT (*123, 124*). Earlier studies also hinted at this complex, as many SEC members were identified in P-TEFb complexes, including AFF4 (*125*) AFF4, AF9/ENL (*126*) and AFF4, AF9 (*127*). Although different studies vary slightly on the composition and contacts made within this complex, it is agreed that AFF4 serves as a scaffold that binds other components (*122, 128-130*). More details regarding this complex can be found in any of several excellent reviews (*25, 131*).

CDK9 recruitment is regulated by gene-specific factors

For P-TEFb to activate transcription it must be recruited to sites of transcription. This is evident from biochemical and molecular studies which must use large excesses of this complex for biochemical activity *in vitro* (*132*) or which artificially tether P-TEFb to a gene through either DNA or RNA for activity [commented on in (*133*)]. Although both Cyclin T1 and T2 can bind to Pol II directly through regions in their C terminal end (*134, 135*), they must be recruited by transcriptional co-factors (*136*), deemed "P-TEFb facilitators" (*137*). In the best studied P-TEFb system, HIV-1 transcription, P-TEFb is recruited by a HIV-encoded protein, TAT (*91*).

At endogenous genes, P-TEFb has been shown to be directly recruited to genes through well-studied transcription factors including but not limited to HSF (*138*), c-Myc (*139-142*), MyoD (*143, 144*), SKIP (*145*), ERα (*146*), and general factors of the mediator complex [CDK8 (*147, 148*), MED26 (*149*), and MED23 (*150*)], the mRNA cap binding complex (*151*), Paf1c (*128*), and Gdown (*152*). It is clear that P-TEFb recruitment to specific genes following a transcriptional activating signal is a general step for efficient transcription (Figure 1.1C). The only known examples of Pol II transcripts that are exceptions to this rule include the short U2 snRNA (*153*) and the p53 target gene, p21 (*154*), which do not require either CDK9 or Ser2-P for efficient transcription.

Many studies have also identified the double bromodomain containing protein, BRD4, as a general factor for recruiting P-TEFb (*155*). BRD4 was shown to bind P-TEFb free of the inhibitory 7SK/HEXIM complex (*155, 156*) and recruit P-TEFb to sites of transcription by binding acetylated chromatin (*157, 158*). While it is unclear if BRD4 is a general P-TEFb recruiter during interphase, a major role of BRD4 may be to mark genes for expression just after mitosis (*159-161*). BRD4 has also been shown to directly phosphorylate Ser2 of the Pol II CTD, but it is unclear how this activity is regulated with CDK9 and CDK12 activity (*162*).

Additional roles for CDK9

While many studies have linked Pol II CTD phosphorylation with regulation of transcription-associated processes such as RNA processing and chromatin remodeling [reviewed (*17*)], several studies have linked CDK9 with these processes directly. First, as in yeast, CDK9 is required for bulk H2BUb, H3K4me3, and H3K36me3 though CTD phosphorylation (*163*) as well as through another target, SPT5 (*164*). Interestingly, CDK9 may more directly regulate H2BUb through phosphorylation of the H2B ubiquitin ligase, UBE2A (*165*).

Although the bulk of P-TEFb is found within speckles in the nucleus where it overlaps with transcription and RNA processing machinery (*138, 166, 167*), there have been few studies on the role of CDK9 in regulation of mRNA processing directly. While CDK9 and Ser2-P are not required for transcription of the small U2 snRNA gene, they were previously suggested to be important for 3’ processing (*153*), although later studies showed that 3’ processing is regulated through NELF rather than P-TEFb directly (*168*). In addition CDK9 was shown to be important for 3’ processing of histone RNA (*163*), although this was later determined to be through CDK9-mediated Thr4 phosphorylation (*20*). Finally, components of the 7SK/HEXIM/LARP7 complex, rather than P-TEFb itself was shown to be important for proper alternative splicing in zebrafish (*169*).

CDK12/Cyclin K: the elusive CTD kinase complex

As described above, the role of both Bur1 and Ctk1 in Ser2 phosphorylation in yeast has only been recently appreciated. Surprisingly, this was recognized only a year before CDK12 was shown to be a Ser2 kinase in metazoans. Thus, within a year, the different models for eukaryotic Ser2-P converged on a highly conserved model where Bur1/CDK9 acts more 5' to Ctk1/CDK12 to regulate Ser2 phosphorylation.

Although only recently characterized as a Ser2 kinase, homologs of CDK12 and its cyclin partner, Cyclin K, had been identified in other studies. Humans contain two CDK12 paralogs, CDK12 and CDK13. CDK12 was first identified in 2000 as a cyclin-related kinase, rich in RS domains, that localizes to nuclear speckles, and was first named CrkRS (*170*). CDK13 was identified a year before this, first identified as a Cdc2-like protein, and initially called CDC2L5 (CDC2 like 5) (*171*). Sequence similarity between CDK13 and CDK9 was noted in this study and this lead to the suggestion that CDK13 may be involved in gene regulation. It was later shown to interact with the splicing factor p32, disrupt splicing when overexpressed (*172*), and regulate splicing of the HIV-1 transcript (*173*).

Cyclin K was initially identified as a suppressor of *S. cereviase* cell cycle-regulating cyclin mutants (*174*). However, even in this early study, Cyclin K was shown to be associated with a kinase capable of phosphorylating the Pol II CTD. Until recently, this kinase was thought to be CDK9. The CDK9-Cyclin K interaction was first detected by yeast two hybrid analysis (*175*) and this complex was shown to be able to at least partly substitute for P-TEFb activity *in vitro* (*175*), and *in vivo* when tethered to a transcriptional unit through RNA (*176*). Interestingly, later studies showed that Cyclin K-bound CDK9 inhibits HIV transcription (*177, 178*). Additional studies on the CDK9/Cyclin K complex suggested that it regulates DNA damage response (*179*), discussed in more detail below.

The interaction between CDK12 and Cyclin K has been nebulous. The first characterization of CDK12 did not find an interaction with any cyclin tested (K, T, or M) (*170*). The first CDK12/13 cyclin partner identified was cyclin L (*180, 181*). In fact, the CDK12-Cyclin K interaction was not directly shown until 2011 (*182*).

The role of CDK12 in CTD phosphorylation was first shown in 2010 when Bartowiak and colleagues showed that it phosphorylates Ser2 *in vitro* and *in vivo* (*82*). ChIP studies revealed that CDK12 is located at and regulates Ser2-P towards the 3' end of genes. In addition, they also showed that a *S. cerevisiae* Bur1 chimera with the CDK9 kinase domain rescues a *bur1* deletion while a Ctk1/CDK12 kinase domain chimera rescues a *ctk1* mutant (*82*). Based on these observations, as well as the fact that inhibition of CDK9 activity frequently reduces all detectible Ser2-P in other studies, Bartowiak and colleagues proposed a model that more closely matches Ser2-P regulation in yeast: CDK9 phosphorylates Ser2-P at the 5' end of genes and this activity is upstream of and required for CDK12-mediated Ser2-P towards the 3' end of genes (*183*) (Figure 1.1A,C).

The interaction between CDK12 and Cyclin K was directly shown the following year (*182*) and this group showed that this interaction is specific and could not detect either a CDK9/Cyclin K or CDK12/Cyclin L complex. More interestingly, this group also showed that the genes most affected by loss of CDK12/Cyclin K are long genes with a large number of exons. DNA damage response genes were down regulated and cells lacking CDK12/Cyclin K accumulate spontaneous DNA damage. Finally, while they acknowledge that previous studies suggested a role for CDK12 in splicing regulation, they did not detect dramatic defects in splicing (*182*). Taking this study into account, it is possible that at least some of the previous results suggesting a role for the CDK9/Cyclin K complex in DNA damage (*179*) may have been detecting roles for the CDK12/Cyclin K complex instead.

A few other recent studies have hinted at the importance of the CDK12/Cyclin K complex in proper development. Mutations in CDK12 have been identified in gastric cancer (*184*) ovarian

carcinoma (*185, 186*), and chronic kidney disease (*187*). In addition, Cyclin K levels are important for differentiation. While Cyclin K levels are high in pluripotent embryonic stem cells (ESCs), they decrease through differentiation and experimentally decreasing CDK12 levels in ESCs leads to differentiation (*188*). Finally, CDK12 levels are undetectable in artemia (brine shrimp) during diapause, a dormant phase of eggs when little transcription is occurring (*189*). Thus, CDK12/Cyclin K is highly regulated during differentiation and development, and the mechanism of this regulation needs to be more extensively characterized.

Spt4/Spt5: The evolution of dual functions of a conserved transcription elongation factor

Counteracting positive regulators of transcription elongation such as the Pol II CTD kinases are negative elongation factors. These include the DRB sensitivity inducing factor complex (DSIF) composed of SPT5 and SPT4, and the negative elongation factor complex (NELF) composed of NELF-A, NELF-B, NELF-C/D, and NELF-E (*190*). The DSIF components were initially referred to as p160 and p14 based on their apparent molecular weight, but given the fact that they were initially cloned from human DNA based on their homology to yeast Spt5 and Spt4 (*191, 192*) and their sequence and functional homology to yeast Spt5 and Spt4, I will refer to them as SPT5 and SPT4. In addition, because the DSIF complex received its name from a biochemical property detected in human cells, I will refer to this complex as SPT4/SPT5 across organisms for simplicity. Because NELF is only conserved in higher eukaryotes such as *Drosophila* and mammals, I am focusing the majority of my discussion on the SPT4/SPT5 complex. SPT4/SPT5 has been recently considered a negative elongation complex; however, it may not play this role outside of the context of NELF (discussed in detail below).

Identification of SPT4/SPT5

Spt4 and Spt5 were first identified in a screen as factors that regulate transcription in yeast (*193*). Many studies since then have defined these proteins as important, conserved factors in

transcription elongation. Spt4 and Spt5 form a heterodimer (*194*) and although SPT4 is highly conserved from yeast to human it appears to be dispensable for SPT4/SPT5 activity as loss of SPT4 is not lethal in the systems tested (*104, 195*) and is only conditionally lethal in yeast (*196, 197*). In contrast, SPT5 is essential in all eukaryotes tested (*104, 198-201*). SPT4 (102 amino acids in *S. cerevisiae*) is a much smaller protein than SPT5 (1063 amino acids) and although it contains a Zn finger domain that has been shown to be important for function yeast (*196, 197, 202*), this is not the case for *in vitro* studies using human factors (*194*). Furthermore, there are conflicting reports about the ability of this domain to bind nucleic acids (*203, 204*). Finally, all of the known functional contacts made by the SPT4/SPT5 complex with other proteins are mediated through SPT5 (discussed in detail below). Thus, much of the function of the SPT4/SPT5 complex is attributed to SPT5 directly and it has been proposed that SPT4 plays mostly a structural role in the complex (*194*).

SPT4/SPT5 mediates PPP in higher eukaryotes by bridging NELF and Pol II binding

As the members of the DSIF complex, SPT4/SPT5 has gained recent attention as a regulator of Pol II promoter proximal pausing (PPP), although this mechanism likely only functions in higher eukaryotes where the NELF complex is present. Given the attention that recent reviews have given to this function of DSIF in negative elongation regulation (*18, 205*), I am only going to summarize the current model of PPP. In the late 1990s to early 2000s, Handa and colleagues performed beautiful biochemical work that established Pol II elongation as a point of transcriptional regulation by SPT4/SPT5 and NELF. Specifically, SPT4/SPT5 interacts directly with Pol II (*206*), bridging contact between NELF and Pol II (*207*). Although the mechanism of NELF-mediated pausing is poorly understood, several models have been proposed. NELF may affect the intrinsic elongation properties of Pol II itself by increasing the time Pol II spends at pause sites (*132*), or by inhibiting positive elongation factors such as TFIIS (*208*), and/or it may affect the ability of Pol II to navigate through chromatin [discussed in (*168, 209*)]. Although

discussed in detail elsewhere, it seems likely that at least part of NELF's mechanism of action is to regulate Pol II directly as the initial studies that identified NELF were *in vitro* Pol II elongation assays on a naked DNA template (*210*). Elongation repression is reversed by P-TEFb (*132, 211*). P-TEFb phosphorylates both SPT5 [discussed in detail below, (*212*)] and NELF (*213*) components, and following this phosphorylation, NELF leaves the elongating Pol II complex while SPT4/SPT5 remains associated (*214, 215*).

After dissociation of NELF, SPT4/SPT5 acts to enhance transcription elongation [details below, (*212*)]. Thus, SPT4/SPT5 appears to repress elongation simply by acting as a facilitator of NELF-mediated repression. One study showed that the NELF-A can bind Pol II directly (*190*); however, neither SPT4/SPT5 nor NELF can repress Pol II elongation on their own in a highly purified system (*190, 210*), and in *Drosophila*, NELF requires SPT4/SPT5 to bind Pol II (*216*). The only report of SPT4/SPT5-mediated pausing *in vivo* includes the NF-kB target gene, A20, which has a "paused" polymerase profile in the absence of any detectable NELF binding (*217*).

If SPT4/SPT5 plays mostly a positive role in elongation in the absence of NELF, one would expect that in organisms that lack NELF, SPT4/SPT5 would act as a positive enhancer of transcription elongation. While NELF is only conserved in higher eukaryotic species (*190*), both SPT4 and SPT5 are highly conserved proteins among eukaryotes and SPT5 is the only known transcription elongation factor with a domain conserved from bacteria [reviewed in (*218*)]. Studies performed in yeast, which lack NELF, suggest that Spt4/Spt5 acts only as a positive elongation factor (*219, 220*). While most of the studies in yeast have focused on Spt5 in recruiting transcription associated factors (discussed below), no negative role in transcription elongation has been described thus far. It would be particularly interesting to determine if SPT4/SPT5 has a positive and/or negative role in transcription elongation in other common model systems that lack NELF such as *Arabidopsis* or *C. elegans* using an *in vitro* transcription elongation assay. If SPT4/SPT5 acts as a positive elongation factor in these systems, it would seem that SPT4/SPT5 acquired its negative role in Pol II elongation with the evolution of NELF-mediated transcription

pausing. This would mean that analogies with the PPP check point in higher eukaryotes may be limited in lower systems.

SPT4/SPT5 is a conserved transcription elongation factor that enhances Pol II processivity

While many studies have linked SPT4/SPT5 function to enhancing RNA production by regulating either RNA maturation or navigation through chromatin (discussed below), it seems likely that SPT4/SPT5 must be able to stimulate elongation outside of this role as a "platform" because the initial characterization of SPT4/SPT5 as an elongation factor used *in vitro* transcription assays on naked templates in the absence of RNA processing (*206*). Furthermore, SPT5 is related to a bacterial elongation factor, NusG, which enhances bacterial RNA polymerase elongation directly. The striking conservation of SPT5 from bacteria was first noticed by Handa and colleagues when they isolated it as an elongation repressor (*206*). While much attention has been given to the negative role of SPT4/SPT5, several biochemical structural studies have characterized the positive role SPT5 plays in elongation and many recent reviews have highlighted these studies (*218, 221-223*).

SPT5 is a multi-domain protein, containing an N terminal acidic domain which likely facilitates the nuclear localization of the protein (*224*), a NusG homology domain, called NGN, that directly binds to RNA polymerase complexes (*225, 226*), KOW domains that likely mediate other protein-protein or protein-nucleic acid interactions (*226-228*), and a C terminal repetitive "CTR" domain that can be phosphorylated by P-TEFb/Bur1 (discussed in detail below). While this general domain structure is conserved among eukaryotes, there are variations in the number of KOW domains and the repetitive nature of the CTR across eukaryotes (discussed below). Both the bacterial NusG and archaeal Spt5 proteins contain only the NGN domain and a single KOW domain [reviewed in (*223*)] and the relative simplicity of these proteins have allowed structural analysis.

The best understanding of how SPT5 homologs facilitate elongation have come from structural studies. Co-crystallization of archaeal Spt4/Spt5 with RNA polymerase shows that the NGN domain of Spt5 binds to the coiled-coil domain of the "clamp" of the large subunit of RNA polymerase (*225, 228*). The RNA polymerase "clamp" reaches out to downstream DNA about to be transcribed and Spt5 binding creates a complete enclosure around this DNA, preventing RNA polymerase from disengaging (*226, 228*). The DNA enclosure enhances RNA polymerase processivity similar to the role of the DNA clamp in the DNA polymerase complex. This structural interaction is conserved in bacterial NusG binding to RNA polymerase (*226*). The conservation of both the NGN domain of SPT5 homologs and the coiled coil domain in RNA polymerases from bacteria to eukaryotes support the model that eukaryotic SPT5 enhances RNA polymerase processivity directly.

Interplay of Pol II elongation factors

As discussed in detail above, the Pol II elongation kinases regulate transcription-associated processes through phosphorylation of the Pol II CTD. However, many studies have shown that phosphorylation of the SPT5 CTR is just as important as the Pol II CTD in the recruitment of transcription associated factors.

Phosphorylation of the SPT5 CTR

As mentioned above, SPT5 contains a repetitive C terminal domain that plays an important role in transcriptional regulation. While SPT5 is conserved from bacteria to eukaryotes (*206*), the CTR is only present in eukaryotic SPT5. The repetitive nature of the CTR and similarity to the repetitive Pol II CTD were first described in the cloning of *SPT5* in yeast (*198*), although the first studies showing SPT5 phosphorylation were performed in human (*229*). The repetitive C terminus of SPT5 has been called a C terminal region, CTR, in budding yeast and C terminal

domain, CTD, in fission yeast. I will refer to it as the CTR to prevent confusion with the Pol II CTD.

Although the existence of a somewhat repetitive, extended CTR on SPT5 is conserved across eukaryotes, there has been controversy over the consensus repeat sequence. Initial studies identified 15 copies of the 6 amino acid repeat S-A/T-W-G-G-A/Q in budding yeast (*198*), a highly regular 18 copies of 9 amino acid repeat T-P-A-W-N-S-G-S-K in fission yeast (*220*), and two different repeats, G-S-R/Q-T-P-X-Y and P-T/S-P-S-P-Q/A-S/G-Y, in human (*229*). A fairly recent analysis of the SPT5 CTR across species suggested that the conserved repeat is G-S-R/Q-T-P conserved from yeast to humans (*214*); however, it is not clear if this is the functional repeat across eukaryotes.

Although early biochemical studies of human SPT5 showed conflicting reports about the requirement of the CTR on DSIF activity *in vitro* (*207, 212*), Handa and colleagues have shown that SPT5 phosphorylation is important for the activating but not repressive function of SPT5. They thus suggested a "switch model" whereby P-TEFb converts the repressive SPT4/SPT5/NELF complex to a positive SPT4/SPT5 complex by phosphorylation the SPT5 CTR (*214*). While yeast Bur1 also phosphorylates SPT5, it is unclear how this affects its role in directly regulating Pol II elongation (*75*).

As described above, the SPT5 CTR is a conserved target for Bur1 and CDK9, despite the lack of sequence conservation within the CTR. In budding yeast, Bur1 has been shown to phosphorylate Spt5 on the CTR while Ctk1 cannot (*43, 75*). In fission yeast, Cdk9 specifically phosphorylates T1 of the CTR 9 amino acid repeat (*46, 77*). Studies in human cells have shown that the CTR is phosphorylated by P-TEFb (*212*) on both Thr and Ser residues of the CTR (*230*). Interestingly, studies have shown that CDK7 may also play a role in phosphorylation of the CTR, although its kinase activity on the CTR is not as robust as CDK9's activity (*231*).

Spt4/Spt5 recruitment of transcription associated factors

The Bur1/CDK9 targets, the Pol II CTD and SPT5 CTR seem to act in concert to recruit RNA processing and chromatin remodeling factors. Thus, in addition to directly stimulating Pol II processivity (discussed above), SPT4/SPT5 also indirectly helps Pol II navigate through a repressive chromatin environment (*232, 233*). Specifically, SPT4/SPT5 has been shown to directly interact with, and in some cases, recruit chromatin modifying complexes. Spt4 and Spt5 were initially identified in yeast in a screen with the chromatin remodeler Spt6 (*195*). Since then, Spt4/Spt5 have been shown to directly interact with this factor (*219, 234*) as well as FACT (*235*), Rpd3S (*236*), and the MSL complex in *Drosophila* (*200*). Spt5 also regulates the activity of the Paf1c complex: Spt5 directly interacts with Paf1c (*237*), and phosphorylation of the Spt5 CTR enhances Paf1c recruitment to transcription loci (*43, 238*). Interestingly, CTR phosphorylation is also important for Paf1c-mediated elongation stimulation on a naked template *in vitro* (*238*).

As with the Pol II CTD, SPT5 also regulates mRNA processing factors including capping factors in human (*239, 240*), splicing and 3' processing factors in budding yeast (*235, 241, 242*), and the exosome in *Drosophila* (*243*). In fission yeast, Spt5 recruits capping enzymes (*244*), but this does not require phosphorylation of the CTR (*220*). It is interesting to note that capping enzyme recruitment in fission yeast may be shared with Cdk9 as Cdk9 is found as a heterotrimer with the capping factor Pct1 (*58*).

Due to the fact that SPT5 can recruit many of the same factors to transcriptional units as the Pol II CTD, Handa and colleagues have suggested that the SPT5 CTR functions as a "mini CTD" that can recruit transcription associated factors in collaboration with the Pol II CTD (*214*). This shared function of the SPT5 is highlighted by the fact that deletion of the Spt5 CTR is synthetically lethal with Pol II CTD truncations in fission yeast (*220*).

Genome-wide model or gene-specific function?

Almost all of the factors described above were first characterized using biochemical studies on naked DNA templates. This, along with the fact that these factors appear to have fairly uniform effects on the genes tested *in vitro* and *in vivo*, led to the model that regulation of transcription elongation occurs through a general mechanism (Figure 1.1), and that elongation factors are general transcription factors. Furthermore, genome wide ChIP studies of the location of these factors also suggest they were general factors as they are located on a surprisingly high number of genes (*224*, *245*) and have a large overlap with Pol II's location (*246*, *247*).

Because these factors appear to be general based on ChIP analysis, does that mean that transcriptional regulation of all these genes is uniform? Recent reviews have highlighted the evidence in support of both general and gene-specific role of Ser2 phosphorylation and transcriptional elongation (*205*, *248*). I, however, favor a model (Figure 1.1) whereby these elongation factors play largely general roles in transcription. However, in metazoans, where transcription elongation is a regulated step of gene expression, the activity of and regulation by these factors is modified by signaling pathways and gene specific transcription factors. Thus, gene-specific transcription factors create a gene-specific balance between positive and negative regulation of Pol II elongation. While these factors likely act genome-wide to regulate elongation in the systems tested, no study thus far has studied tissue-specific differences in the regulation of transcriptional elongation.

Germline-specific gene regulation

The germline is arguably the most important cell lineage in multicellular organisms because it is the only one that contributes to the perpetuation of a species. To perform this function, the germline differentiation capacity cycles during development, differentiating into highly specialized, haploid cells, and transforming back into the totipotent one cell zygote upon

fertilization. Because germ cells alone contribute to the next generation, they must be protected from: 1) improper inheritance of aberrant genomic information, including genetic and epigenetic components, and 2) terminal somatic differentiation of germ cells. In order to protect the genome, germline-specific small RNAs, called piRNAs prevent transposon-mediated mutagenesis [reviewed in (*249-251*)]. In addition, inheritance of aberrant epigenetic information is combated through epigenetic reprogramming, or erasure of many epigenetic marks at germline specification (*252*) and fertilization (*253*). Although these are very important aspects of germline development, I will focus on transcriptional regulation in the germline.

In order to protect itself from terminal somatic transformation, gene expression is uniquely controlled in germ cells (Figure 1.2A). As discussed below, transcriptional activity is regulated through both expression of germline-specific transcription initiation factors as well as through a phase of transcriptional repression during germ cell specification. In addition to this, germ cells prevent somatic differentiation through regulation of translation [(*254, 255*) reviewed (*256*)], although I will limit my discussion to regulation of gene expression through transcription.

Expression of variant basal transcription factors is a conserved hallmark of the germline

Transcriptional regulators are generally classified as basal transcription factors or gene-specific transcription factors (simply TFs). Basal transcription factors are considered to be universal machinery of transcription, including preinitiation complex (PIC) factors that perform the mechanics of preparing DNA for transcription by an RNA polymerase, elongation factors that facilitate polymerase movement through a locus, and termination factors. Classical models of transcription suggest that this machinery is regulated in a spatiotemporal fashion by gene-specific transcription factors that bind both specific DNA elements of a locus and the basal transcription factors to enhance recruitment of transcriptional machinery to a gene. As expected, gene-specific transcription factors are highly regulated to allow proper spatiotemporal gene activation [reviewed in (*257*)]. However, over the past decade, it has become increasingly obvious that basal

transcription factors, including PIC components, are not always “general” and variability of transcriptional machinery composition can play regulatory roles in transcription [reviewed in (*258, 259*)]. It should be noted that, the existence of variants contradicts the previous name “general transcription factor” given to these factors. Thus, these factors are more recently being referred to as “basal” transcription factors, and I will refer to them with this name.

The Pol II basal transcription factor PIC components include TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and the mediator complex [reviewed extensively in (*260*)]. Many of the studies on basal transcription variants have been with TFIID, which is a large, multisubunit complex, composed of the TATA binding protein, TBP, and TBP-associated factors, or TAFs. Many eukaroyotes contain two TBP paralogs, TBP and TBP-like factor, or TLF/TRF2. In addition, mammals contain a third TBP paralog, called TRF3/TBP2. Many eukaryotes also contain paralogs of several TAFs (*259*), some of which are discussed below. Where many of these variants likely act as basal transcription factors, it should also be noted that TAFs are also present in chromatin remodeling complexes in some systems [discussed in (*261*)].

In addition to TFIID, *Drosophila* and mammals also have variants of TFIIA subunits [reviewed in (*262*)]. TFIIA is encoded by two genes, TFIIAαβ and TFIIAγ, and TFIIAαβ can be cleaved into two polypeptides to form a trimeric TFIIA complex. While many organisms contain homologs of TFIIAαβ, called TFII*A-L*ike *F*actor or ALF, there is only one TFIIAγ paralog in all organisms analyzed.

Expression of variant basal transcription factors is a common phenomenon in germ cell lineages (Figure 1.2A). While the variants expressed are largely specific to a species, the use of variant basal transcription factors in germline regulation is common [reviewed in (*261*)]. These germline specific TFIID factors include 5 TAF variants in *Drosophila* sperm [reviewed in (*263*), no hitter (TAF4 paralog), cannonball (TAF5 paralog), meiosis I arrest (TAF6 paralog), spermatocyte arrest (TAF8 paralog) and ryan express (dTAF12 paralog)]; a TAF7 paralog, TAF7L, in mammalian sperm (*264*); a TAF4 paralog, TAF4b, in mammalian oocytes

(*265*); and TBP paralogs, TBP2/TRF3 in xenopus (*266*) and mice oocytes (*267*). Interestingly, the difference between the TAF4/TAF4b paralogs were also shown to be important in embryonic stem cells, where TAF4b is required for transcription of self renewal genes and TAF4 is more important for differentiation (*268*). Finally, the TFIIA subunit variant, ALF, was also shown to function specifically in human (*269, 270*) and mouse (*271*) testis. Thus, the expression of variant basal transcription factors in the germline suggests a unique transcriptional initiation regulation that may protect germ cell identity.

A phase of transcriptional repression is a conserved hallmark of germline specification

The germline also protects itself from somatic differentiation during specification. In sexually reproducing organisms, germ cells are commonly specified in the embryo, when many differentiation signals are highly abundant in order to initiate the somatic identities of most cell lineages (*272*). The germline must protect itself from these fates. Although there are very distinctive differences in specification of the germline across species [reviewed extensively (*272, 273*)], one conserved component of specification includes repression of zygotic transcriptional elongation during specification of the first germline cells, or primordial germ cells, PGCs (*274*) (Figure 1.2A). Interestingly, despite striking similarities of repression mechanisms, the proteins that carry out this repression appear to be unique to a species and are not conserved.

The earliest evidence of PGC transcriptional repression was found in *Drosophila* embryos in the 1970s. Here, it was shown that PGCs fail to incorporate radioactive nucleotides at levels comparable to somatic neighbors (*275, 276*). Almost thirty years later, the factor responsible for this repression, called polar granule component, or *pgc*, was identified (*277*). *pgc* was initially identified through a screen for factors enriched in PGCs and important for PGC specification (*278*), and it was suggested that *pgc* encoded a ncRNA. Later studies identified a sequencing error in the initial analysis of this locus, and correction of this error showed that *pgc* contains two possible ORFs (*279, 280*), and that ORF1, encoding a 71 amino acid polypeptide, is the

functional unit (*280*). These recent studies also uncovered the molecular details of transcriptional repression, showing that the PGC protein directly binds and sequesters P-TEFb, blocking its ability to release paused Pol II in a mechanism independent of the other P-TEFb inhibitory complex, 7SK/HEXIM (*279*). PGC protein activity is autonomous as overexpression in somatic cells also blocks transcription (*281*).

Transcriptional repression occurs by two separate mechanisms in *C. elegans* germline blastomeres (Figure 1.2B). Repression is carried out in somatic and germline blastomers of 1-4 cell embryos through the OMA-1/2 proteins, which sequester the basal transcription factor, TAF-4, in the cytoplasm to prevent transcription initiation (*282*). In addition, transcription is further repressed specifically in the germline precursors by the protein PIE-1. PIE-1 was initially identified as a factor important for early embryo cell fate determination (*283, 284*), and later shown to be important for germline specification (*285*) through repression of Pol II elongation (*286, 287*). Repression of elongation by PIE-1 was later shown to require a Pol II CTD-like heptad motif found in PIE-1 (*288*). This domain is thought to prevent phosphorylation of Ser2 of the Pol II CTD in a manner analogous to the *Drosophila* PGC protein by directly binding Cyclin T (*289, 290*). Interestingly, PIE-1 has other functions in germline blastomeres including limiting Ser5-P (*290*) and regulating expression of a germline factor, NOS-2 (*291*).

Most recently, repression of transcription elongation in PGCs has been shown in the deuterostome ascidians (*292*). Here, a protein called Posterior End Mark, or PEM, blocks transcriptional elongation by binding CDK9. Finally, both *Xenopus* (*293*) and mouse (*294*) PGCs have a period of low Ser2-P early in specification, although the mechanism behind transcriptional repression in these organisms has yet to be identified. In summary, although the specific factors that mediate repression are not conserved, regulation of transcriptional elongation is a hallmark of germline specification.

Introduction to the C. elegans *germline*

Because of the short time span between generations and genetic tractability, *C. elegans* is an excellent model for understanding germline biology. As is well highlighted in a review series (*295*), the *C. elegans* germline is a well studied tissue and much of the regulation of germ cell specification and development is well understood in this organism.

Tissue specification begins at the first cell division in the *C. elegans* embryo, where the zygote, P_0, divides asymmetrically to produce a somatic blastomere and a germline blastomere, P_1. Similarly, P_1 again asymmetrically divides to produce somatic and germline blastomeres. This asymmetric division continues until P_4 divides to produce the *C. elegans* primordial germ cells (PGCs), Z2 and Z3, which remain quiescent through embryogenesis (Figure 1.2B).

After newly hatched larvae begin feeding, the two germ cells, Z2/Z3 begin to proliferate, producing germ cells until the third larval stage of development. At this time, the oldest germ cells begin spermatogenesis, producing haploid sperm until the end of the fourth larval stage, when the gonad switches irreversibly to oogenesis [reviewed (*296*)]. The oocytes produced are then fertilized by the larval-produced sperm. A pool of germline stem cells continuously produces more germ cells that will progress through gametogenesis through the life of the adult animal. Thus, the adult gonad contains germ cells at all stages of development. Fertilized embryos are stored within the uterus until the 30-50 cell stage, when they are then expelled and continue embryogenesis-- and the cycle begins again [described in more detail in (*295*)].

Considering both the heritability of germ cell chromatin and the ability of these cells to regain totipotency following fertilization, there has been a great deal of work toward understanding the epigenetics of these cells [reviewed in (*297*)] and the mechanism of gene expression through translational repression (*256*), and small RNAs (*298*). Despite the identification of germline-specific basal transcription factor variants in other organisms described above, until now there has been little work on germline-specific regulation of transcriptional machinery in *C. elegans*

Transcriptional regulation in *C. elegans*

C. elegans has been an attractive model system for almost 4 decades (*299*); however there has been somewhat limited experimental analysis of the transcriptional machinery, especially compared to other common eukaryotic systems such as yeast, *Drosophila*, and mammalian systems. Much of the analysis of transcriptional regulation is limited because: 1) mutants of transcriptional machinery components are lethal, and 2) there are no immortalized cell culture systems for high resolution analysis of transcriptional activity in a homogeneous cell population. Thus, for my thesis work I set out to address these points and to further understand the mechanism of transcriptional regulation in *C. elegans*.

Characterization of transcription mutants in C. elegans

Because of the essential role of transcription in development, null mutants of transcription components generally cause *C. elegans* to arrest early in development. In addition, many early developmental processes are driven by factors loaded into the embryo (*300*), complicating the analysis of transcriptional activity in young animals. Thus, dissection of transcription-associated processes is challenging because of the limited number of conditional transcription mutants in *C. elegans*. To address this problem, the next chapters detail an analysis of temperature sensitive mutants of the general transcription factors, *ama-1* and *taf-6.2*. The *C. elegans ama-1* gene, which encodes the large subunit of Pol II, was first cloned in 1988 (*301*). At this time, several hypomorphic mutations of *ama-1* were isolated (*302*, *303*); however, because these studies were done when sequence analysis was challenging, the specific DNA changes in these mutants were largely undetermined. In order to better characterize these mutants, which could be potential tools for studying transcription in *C. elegans*, I identified the *ama-1* mutations by sequencing this gene in these mutants. I then utilized the structure of the yeast Pol II complex (*304*) to predict the resulting defects that these mutants [Chapter 2, (*305*)].

In addition to this, I also characterized two other conditional mutants thought to affect transcriptional activity. These were first isolated in a screen for mutations that result in temperature sensitive embryonic arrest (*306*). At the restrictive temperature, embryos arrest at a point similar to the arrest of embryos lacking *ama-1* function (*307*). I mapped these mutations to the *taf-6.2* gene, which is a component of the general transcription factor, TFIID [Chapter 3, (*308*)]. In these two studies, I characterized temperature sensitive mutants in transcriptional machinery that provide useful tools for studying transcription in *C. elegans*.

***Tissue-specific transcriptional regulation in* C. elegans**

In exploring transcriptional activity in *C. elegans*, I also identified tissue-specific regulation of transcriptional elongation. Initial studies by our lab identified CDK-9 independent phosphorylation of serine 2 (Ser2-P) of the Pol II CTD in the *C. elegans* embryonic germ cells (*309*). Expanding on this work, I characterized a unique pattern of Ser2 phosphorylation (Ser2-P) specific to the germline throughout development. My results are the first demonstration of tissue-specific differences in the regulation of Ser2-P in eukaryotes and suggest that germline-specific elongation regulation may comprise a novel facet of germline identity (Chapter 4).

Finally, as described above, genomic or biochemical analysis of tissue-specific phenomena is challenging in *C. elegans* because of the heterogeneity of different cell types in an animal. In order to address this problem, I initiated a project that would allow the purification of germ cell nuclei from *C. elegans* embryos. This approach, called isolation of nuclei tagged in specific cell types, or INTACT, utilizes the high affinity biotin-avidin interaction to purify specific nuclei that have been biotinylated *in vivo* (*310*). I have developed the tools required to optimize the purification, which will be very useful in moving towards genomic and biochemical analysis of transcriptional regulation in *C. elegans* (Chapter 5).

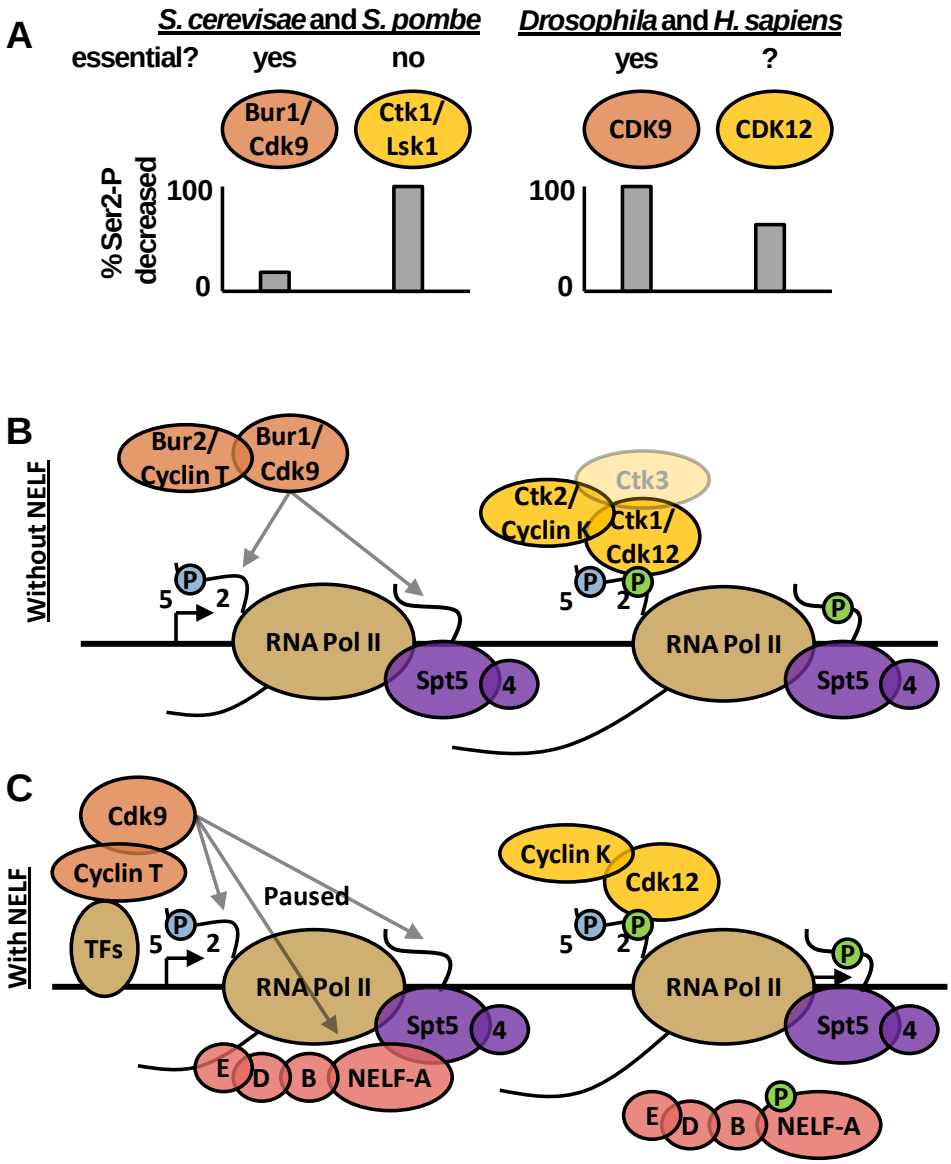
Table 1.1: RNA Polymerase II elongation factor homologs

		S. cerevisae	*S. pombe*	*C. elegans*	*Drosophila*	Mammalians
5' Ser2	kinase	Bur1	Cdk9	CDK-9	CDK9	CDK9/PITARE
	cyclin(s)	Bur2	Pch1	CIT-1.1	CYCT	CCNT1
				CIT-1.2		CCNT2a/b
3' Ser2	kinase(s)	Ctk1	Lsk1	CDK-12	CDK12	CDK12/ CrkRS
						CDK13/CDC2L5
	cyclin(s)	Ctk2	Lsc1	CCNK-1	CYCK	CCNK
		Ctk3	Lsg1	*	*	*
DSIF		Spt4	Spt4	SPT-4	SPT4/p160	SPT4/p160
		Spt5	Spt5	SPT-5	SPT5/p14	SPT5/p14
NELF		*	*	*	NELF-A	NELFA/WHSC2
		*	*	*	NELF-B	NELFB/COBRA1
		*	*	*	NELF-D/TH1	NELFC/NELFD/ TH1/HSPC130
		*	*	*	NELF-E	NELFE/RD†

*no significant homolog detected

† homologs detected in lower eukaryotes contain an RRM similar to that found in NELFE but do not appear to be functional homologs

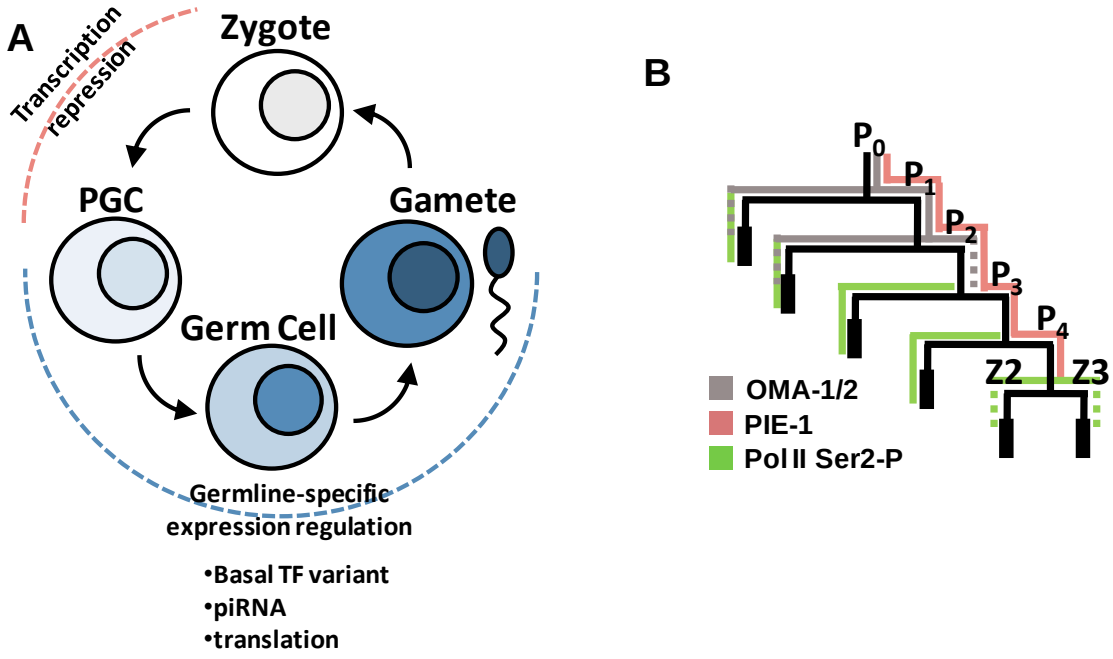
Figure 1.1: Schematic of the role of Ser2 kinases in transcriptional elongation regulation



A. Loss of Bur1/Cdk9 or Ctk1/Lsk1/Cdk12 activity results in different effects on Ser2-P levels in yeast versus metazoans. While Ctk1/Lsk1 are responsible for the overwhelming bulk of Ser2-P in yeast *S. cerevisae* and *S. pombe*, both Cdk9 and Cdk12 regulate Ser2-P levels in the metazoans, *Drosophila* and huamns. Importantly, loss of Cdk9 in metazoans results in complete loss of Ser2-P in metazons suggesting it is required upstream of Cdk12 activity. Finally, while Bur1 and Cdk9 are largely essential genes (*S. cerevisae bur1* mutants are extrememly slow growing), Ctk1/Lsk1

is not essential in yeast. B. Transcription elongation has not been demonstrated to be a regulated step of gene expression in organisms without NELF homologs, including yeast and *C. elegans*. Here, the kinases Bur1/CDK9 and Ctk1/CDK12 are recruited through general mechanisms to phosphorylate Ser2 of the Pol II CTD and the SPT5 CTR. These phosphorylation marks act to enhance recruitment of other Pol II elongation associated factors such as RNA processing factors and chromatin remodelers. C. In organisms that contain the NELF complex, such as *Drosophila* and mammals, the transition from Pol II initiation to elongation is a regulated step of gene expression. Here, SPT4/SPT5 mediates NELF binding to the Pol II complex, "pausing" Pol II and preventing productive elongation. Regulated recruitment of CDK9 and Cyclin T through transcription factors and phosphorylation of NELF releases this pause. At the same time, CDK9 phosphorylates the Pol II CTD on Ser2 and the SPT5 CTR. Downstream of this regulated step, the CDK12 complex further phosphorylates the Pol II CTD.

Figure 2.1: Schematic of germline transcriptional regulation



A. The germline cycle starts with specification of primordial germ cells (PGCs) in the embryo, followed by proliferation to form germ cells, and specialization during meiosis to form gametes. Fertilization starts the cycle over again. Transcriptional repression is a hallmark of PGC specialization. Gene expression is regulated in germ cells through expression of variant basal transcription factors, and post-transcriptional regulation by piRNAs and unique translation regualtion . B. Schematic of the C. elegans PGC (Z2/Z3) specification. Z2/Z3 are born from the P lineage (described in the text). The OMA-1/2 and PIE-1 proteins prevent transcription in the early embryo and P lineage respectively. Phosphorylation of Ser2 of the Pol II CTD appears in the embryo when these proteins are degraded.

Citations

1. R. G. Roeder, W. J. Rutter, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234 (1969).
2. R. H. Reeder, R. G. Roeder, Ribosomal RNA synthesis in isolated nuclei. *J Mol Biol* **67**, 433 (1972).
3. R. Weinmann, R. G. Roeder, Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. *Proc Natl Acad Sci U S A* **71**, 1790 (1974).
4. R. Weinmann, H. J. Raskas, R. G. Roeder, Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. *Proc Natl Acad Sci U S A* **71**, 3426 (1974).
5. J. F. Kugel, J. A. Goodrich, Non-coding RNAs: key regulators of mammalian transcription. *Trends Biochem Sci* **37**, 144 (2012).
6. R. A. Young, RNA polymerase II. *Annual review of biochemistry* **60**, 689 (1991).
7. P. A. Weil, D. S. Luse, J. Segall, R. G. Roeder, Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* **18**, 469 (1979).
8. M. Samuels, A. Fire, P. A. Sharp, Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J Biol Chem* **257**, 14419 (1982).
9. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature*, (2013).
10. A. Dvir, Promoter escape by RNA polymerase II. *Biochim Biophys Acta* **1577**, 208 (2002).
11. J. L. Corden, Tails of RNA polymerase II. *Trends Biochem Sci* **15**, 383 (1990).
12. D. L. Cadena, M. E. Dahmus, Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *The Journal of biological chemistry* **262**, 12468 (1987).
13. L. A. Allison, M. Moyle, M. Shales, C. J. Ingles, Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599 (1985).
14. H. Lu, O. Flores, R. Weinmann, D. Reinberg, The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10004 (1991).
15. T. O'Brien, S. Hardin, A. Greenleaf, J. T. Lis, Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**, 75 (1994).
16. H. Serizawa, J. W. Conaway, R. C. Conaway, Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. *Nature* **363**, 371 (1993).
17. S. Egloff, S. Murphy, Cracking the RNA polymerase II CTD code. *Trends Genet* **24**, 280 (2008).
18. Q. Zhou, T. Li, D. H. Price, RNA polymerase II elongation control. *Annual review of biochemistry* **81**, 119 (2012).
19. C. G. Noble *et al.*, Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12**, 144 (2005).
20. J. P. Hsin, A. Sheth, J. L. Manley, RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science (New York, N.Y.)* **334**, 683 (2011).
21. R. J. Sims, 3rd *et al.*, The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**, 99 (2011).
22. S. M. Ranuncolo, S. Ghosh, J. A. Hanover, G. W. Hart, B. A. Lewis, Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem* **287**, 23549 (2012).
23. R. D. Chapman *et al.*, Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780 (2007).
24. M. Ott, M. Geyer, Q. Zhou, The control of HIV transcription: keeping RNA polymerase II on track. *Cell host & microbe* **10**, 426 (2011).
25. E. Smith, C. Lin, A. Shilatifard, The super elongation complex (SEC) and MLL in development and disease. *Genes & development* **25**, 661 (2011).
26. J. Yan, J. Diaz, J. Jiao, R. Wang, J. You, Perturbation of BRD4 protein function by BRD4-NUT protein abrogates cellular differentiation in NUT midline carcinoma. *The Journal of biological chemistry* **286**, 27663 (2011).

27. N. He *et al.*, A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. *Molecular cell* **29**, 588 (2008).
28. A. P. C. O. N. C. R. J. Rice, Pmid, Dysregulation of positive transcription elongation factor B and myocardial hypertrophy. *Circulation research* **104**, 1327 (2009).
29. V. Krystof, I. Chamrad, R. Jorda, J. Kohoutek, Pharmacological targeting of CDK9 in cardiac hypertrophy. *Medicinal research reviews* **30**, 646 (2010).
30. A. M. Alazami *et al.*, Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. *Human mutation* **33**, 1429 (2012).
31. G. Prelich, RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell* **1**, 153 (2002).
32. J. M. Lee, A. L. Greenleaf, A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II. *Proc Natl Acad Sci U S A* **86**, 3624 (1989).
33. J. M. Lee, A. L. Greenleaf, CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. *Gene expression* **1**, 149 (1991).
34. D. E. Sterner, J. M. Lee, S. E. Hardin, A. L. Greenleaf, The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. *Molecular and cellular biology* **15**, 5716 (1995).
35. G. Hautbergue, V. Goguel, Activation of the cyclin-dependent kinase CTDK-I requires the heterodimerization of two unstable subunits. *The Journal of biological chemistry* **276**, 8005 (2001).
36. J. M. Lee, A. L. Greenleaf, Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. *The Journal of biological chemistry* **272**, 10990 (1997).
37. J. Karagiannis, A. Bimbo, S. Rajagopalan, J. Liu, M. K. Balasubramanian, The nuclear kinase Lsk1p positively regulates the septation initiation network and promotes the successful completion of cytokinesis in response to perturbation of the actomyosin ring in Schizosaccharomyces pombe. *Mol Biol Cell* **16**, 358 (2005).
38. J. Karagiannis, M. K. Balasubramanian, A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. *PLoS One* **2**, e433 (2007).
39. G. Prelich, F. Winston, Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription in vivo. *Genetics* **135**, 665 (1993).
40. S. Yao, A. Neiman, G. Prelich, BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo. *Molecular and cellular biology* **20**, 7080 (2000).
41. M. C. Keogh, V. Podolny, S. Buratowski, Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Molecular and cellular biology* **23**, 7005 (2003).
42. H. Qiu, C. Hu, A. G. Hinnebusch, Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Molecular cell* **33**, 752 (2009).
43. Y. Liu *et al.*, Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and cellular biology* **29**, 4852 (2009).
44. S. Murray, R. Udupa, S. Yao, G. Hartzog, G. Prelich, Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase. *Molecular and cellular biology* **21**, 4089 (2001).
45. Y. Chu, A. Sutton, R. Sternglanz, G. Prelich, The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *Molecular and cellular biology* **26**, 3029 (2006).
46. Y. Pei, S. Shuman, Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis. *The Journal of biological chemistry* **278**, 43346 (2003).
47. M. Patturajan, N. K. Conrad, D. B. Bregman, J. L. Corden, Yeast carboxyl-terminal domain kinase I positively and negatively regulates RNA polymerase II carboxyl-terminal domain phosphorylation. *The Journal of biological chemistry* **274**, 27823 (1999).

48. A. C. Bishop, O. Buzko, K. M. Shokat, Magic bullets for protein kinases. *Trends Cell Biol* **11**, 167 (2001).
49. J. R. Tietjen *et al.*, Chemical-genomic dissection of the CTD code. *Nature structural & molecular biology* **17**, 1154 (2010).
50. D. Coudreuse *et al.*, A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Curr Biol* **20**, 1053 (2010).
51. L. Viladevall *et al.*, TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. *Molecular cell* **33**, 738 (2009).
52. A. Guiguen *et al.*, Recruitment of P-TEFb (Cdk9-Pch1) to chromatin by the cap-methyl transferase Pcm1 in fission yeast. *The EMBO journal* **26**, 1552 (2007).
53. A. Wood, J. Schneider, J. Dover, M. Johnston, A. Shilatifard, The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. *Molecular cell* **20**, 589 (2005).
54. E. Clausing *et al.*, The transcription elongation factor Bur1-Bur2 interacts with replication protein A and maintains genome stability during replication stress. *The Journal of biological chemistry* **285**, 41665 (2010).
55. C. Bouchoux *et al.*, CTD kinase I is involved in RNA polymerase I transcription. *Nucleic acids research* **32**, 5851 (2004).
56. S. Grenetier, C. Bouchoux, V. Goguel, CTD kinase I is required for the integrity of the rDNA tandem array. *Nucleic acids research* **34**, 4996 (2006).
57. S. Rother, K. C. I. N. G. D. J. Strasser, Pmid, The RNA polymerase II CTD kinase Ctk1 functions in translation elongation. *Genes & development* **21**, 1409 (2007).
58. Y. Pei, B. Schwer, S. Shuman, Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *The Journal of biological chemistry* **278**, 7180 (2003).
59. C. K. Ho, S. Shuman, Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**, 405 (1999).
60. C. R. Rodriguez *et al.*, Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol Cell Biol* **20**, 104 (2000).
61. D. A. Skaar, A. L. Greenleaf, The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. *Molecular cell* **10**, 1429 (2002).
62. S. H. Ahn, M. Kim, S. Buratowski, Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell* **13**, 67 (2004).
63. S. L. McKay, T. L. Johnson, An investigation of a role for U2 snRNP spliceosomal components in regulating transcription. *PloS one* **6**, e16077 PHST (2011).
64. C. Neuveglise, C. Marck, C. Gaillardin, The intronome of budding yeasts. *C R Biol* **334**, 662 (2011).
65. T. Xiao *et al.*, Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & development* **17**, 654 (2003).
66. N. J. Krogan *et al.*, Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and cellular biology* **23**, 4207 (2003).
67. M. L. Youdell *et al.*, Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology* **28**, 4915 (2008).
68. A. Wood *et al.*, Ctk complex-mediated regulation of histone methylation by COMPASS. *Molecular and cellular biology* **27**, 709 (2007).
69. T. Xiao *et al.*, The RNA polymerase II kinase Ctk1 regulates positioning of a 5' histone methylation boundary along genes. *Molecular and cellular biology* **27**, 721 (2007).
70. H. Kim *et al.*, Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology* **17**, 1279 (2010).
71. M. L. West, J. L. Corden, Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223 (1995).
72. M. C. Keogh *et al.*, Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593 (2005).

73. T. Kim, S. Buratowski, Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. *The Journal of biological chemistry* **282**, 20827 (2007).
74. R. N. Laribee *et al.*, BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Current biology : CB* **15**, 1487 (2005).
75. K. Zhou, W. H. Kuo, J. Fillingham, J. F. Greenblatt, Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6956 (2009).
76. H. Qiu, C. Hu, N. A. Gaur, A. G. Hinnebusch, Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *The EMBO journal* **31**, 3494 (2012).
77. M. Sanso *et al.*, A positive feedback loop links opposing functions of P-TEFb/Cdk9 and histone H2B ubiquitylation to regulate transcript elongation in fission yeast. *PLoS genetics* **8**, e1002822 PHST (2012).
78. E. J. Cho, M. S. Kobor, M. Kim, J. Greenblatt, S. Buratowski, Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & development* **15**, 3319 (2001).
79. M. Kim, S. H. Ahn, N. J. Krogan, J. F. Greenblatt, S. Buratowski, Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The EMBO journal* **23**, 354 (2004).
80. A. Mayer *et al.*, Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology* **17**, 1272 (2010).
81. A. Wyce *et al.*, H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. *Molecular cell* **27**, 275 (2007).
82. B. Bartkowiak *et al.*, CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & development* **24**, 2303 (2010).
83. A. Wood, A. Shilatifard, Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. *Cell cycle (Georgetown, Tex.)* **5**, 1066 (2006).
84. N. F. Marshall, J. Peng, Z. Xie, D. H. Price, Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of biological chemistry* **271**, 27176 (1996).
85. N. F. Marshall, D. H. Price, Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and cellular biology* **12**, 2078 (1992).
86. L. A. Chodosh, A. Fire, M. Samuels, P. A. Sharp, 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J Biol Chem* **264**, 2250 (1989).
87. N. F. Marshall, D. H. Price, Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of biological chemistry* **270**, 12335 (1995).
88. M. Braddock, A. M. Thorburn, A. J. Kingsman, S. M. Kingsman, Blocking of Tat-dependent HIV-1 RNA modification by an inhibitor of RNA polymerase II processivity. *Nature* **350**, 439 (1991).
89. R. A. Marciniak, P. A. Sharp, HIV-1 Tat protein promotes formation of more-processive elongation complexes. *The EMBO journal* **10**, 4189 (1991).
90. X. Yang, C. H. Herrmann, A. P. Rice, The human immunodeficiency virus Tat proteins specifically associate with TAK in vivo and require the carboxyl-terminal domain of RNA polymerase II for function. *Journal of virology* **70**, 4576 (1996).
91. Y. Zhu *et al.*, Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes & development* **11**, 2622 (1997).
92. J. Peng, Y. Zhu, J. T. Milton, D. H. Price, Identification of multiple cyclin subunits of human P-TEFb. *Genes & development* **12**, 755 (1998).
93. P. Wei, M. E. Garber, S. M. Fang, W. H. Fischer, K. A. Jones, A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451 (1998).
94. J. Peng, N. F. Marshall, D. H. Price, Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *The Journal of biological chemistry* **273**, 13855 (1998).
95. D. Ivanov *et al.*, Cyclin T1 domains involved in complex formation with Tat and TAR RNA are critical for tat-activation. *Journal of molecular biology* **288**, 41 (1999).
96. J. Kohoutek *et al.*, Cyclin T2 is essential for mouse embryogenesis. *Molecular and cellular biology* **29**, 3280 (2009).

97. R. Ramakrishnan, W. Yu, A. P. Rice, Limited redundancy in genes regulated by Cyclin T2 and Cyclin T1. *BMC research notes* **4**, 260 PHST (2011).
98. M. Zhou *et al.*, Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Molecular and cellular biology* **20**, 5077 (2000).
99. G. Napolitano, B. Majello, P. Licciardo, A. Giordano, L. Lania, Transcriptional activity of positive transcription elongation factor b kinase in vivo requires the C-terminal domain of RNA polymerase II. *Gene* **254**, 139 (2000).
100. Y. Ramanathan *et al.*, Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *The Journal of biological chemistry* **276**, 10913 (2001).
101. Y. Ramanathan, S. M. Reza, T. M. Young, M. B. Mathews, T. Pe'ery, Human and rodent transcription elongation factor P-TEFb: interactions with human immunodeficiency virus type 1 tat and carboxy-terminal domain substrate. *Journal of virology* **73**, 5448 (1999).
102. R. Pinhero, P. Liaw, K. Bertens, K. Yankulov, Three cyclin-dependent kinases preferentially phosphorylate different parts of the C-terminal domain of the large subunit of RNA polymerase II. *European journal of biochemistry / FEBS* **271**, 1004 (2004).
103. K. Glover-Cutter *et al.*, TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology* **29**, 5455 (2009).
104. E. Y. Shim, A. K. Walker, Y. Shi, T. K. Blackwell, CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. *Genes & development* **16**, 2135 (2002).
105. A. K. Boehm, A. Saunders, J. Werner, J. T. Lis, Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Molecular and cellular biology* **23**, 7628 (2003).
106. Z. Ni, B. E. Schwartz, J. Werner, J. R. Suarez, J. T. Lis, Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Molecular cell* **13**, 55 (2004).
107. J. C. Eissenberg, A. Shilatifard, N. Dorokhov, D. E. Michener, Cdk9 is an essential kinase in Drosophila that is required for heat shock gene expression, histone methylation and elongation factor recruitment. *Molecular genetics and genomics : MGG* **277**, 101 (2007).
108. B. M. Peterlin, J. E. Brogie, D. H. Price, 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley interdisciplinary reviews. RNA* **3**, 92 (2012).
109. V. T. Nguyen, T. Kiss, A. A. Michels, O. Bensaude, 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322 (2001).
110. Z. Yang, Q. Zhu, K. Luo, Q. Zhou, The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317 (2001).
111. A. A. Michels *et al.*, MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Molecular and cellular biology* **23**, 4859 (2003).
112. J. H. Yik *et al.*, Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Molecular cell* **12**, 971 (2003).
113. A. A. Michels *et al.*, Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *The EMBO journal* **23**, 2608 (2004).
114. R. Chen, Z. Yang, Q. Zhou, Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. *The Journal of biological chemistry* **279**, 4153 (2004).
115. D. Blazek, M. Barboric, J. Kohoutek, I. Oven, B. M. Peterlin, Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. *Nucleic acids research* **33**, 7000 (2005).
116. J. H. Yik, R. Chen, A. C. Pezda, Q. Zhou, Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *The Journal of biological chemistry* **280**, 16368 (2005).
117. Q. Li *et al.*, Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *The Journal of biological chemistry* **280**, 28819 (2005).

118. B. J. Krueger *et al.*, LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic acids research* **36**, 2219 (2008).
119. C. Jeronimo *et al.*, Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Molecular cell* **27**, 262 (2007).
120. C. Lin *et al.*, AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular cell* **37**, 429 (2010).
121. A. Yokoyama, M. Lin, A. Naresh, I. Kitabayashi, M. L. Cleary, A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer cell* **17**, 198 (2010).
122. D. Biswas *et al.*, Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15751 (2011).
123. N. He *et al.*, HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Molecular cell* **38**, 428 (2010).
124. B. Sobhian *et al.*, HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Molecular cell* **38**, 439 (2010).
125. M. C. Estable *et al.*, MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein. *Journal of biomedical science* **9**, 234 (2002).
126. E. Bitoun, P. L. Oliver, K. E. Davies, The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Human molecular genetics* **16**, 92 (2007).
127. D. Mueller *et al.*, Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS biology* **7**, e1000249 PHST (2009).
128. N. He *et al.*, Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E636 (2011).
129. Z. Luo *et al.*, The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. *Molecular and cellular biology* **32**, 2608 (2012).
130. S. Chou *et al.*, HIV-1 Tat recruits transcription elongation factors dispersed along a flexible AFF4 scaffold. *Proceedings of the National Academy of Sciences of the United States of America*, (2012).
131. Z. Luo, C. Lin, A. Shilatifard, The super elongation complex (SEC) family in transcriptional control. *Nature reviews. Molecular cell biology* **13**, 543 (2012).
132. D. B. Renner, Y. Yamaguchi, T. Wada, H. Handa, D. H. Price, A highly purified RNA polymerase II elongation control system. *The Journal of biological chemistry* **276**, 42601 (2001).
133. A. De Luca, M. De Falco, A. Baldi, M. G. Paggi, Cyclin T: three forms for different roles in physiological and pathological functions. *Journal of cellular physiology* **194**, 101 (2003).
134. R. Taube, X. Lin, D. Irwin, K. Fujinaga, B. M. Peterlin, Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes. *Molecular and cellular biology* **22**, 321 (2002).
135. T. Kurosu, F. Zhang, B. M. Peterlin, Transcriptional activity and substrate recognition of cyclin T2 from P-TEFb. *Gene* **343**, 173 (2004).
136. B. M. Peterlin, D. H. Price, Controlling the elongation phase of transcription with P-TEFb. *Molecular cell* **23**, 297 (2006).
137. B. Cheng, D. H. Price, Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *The Journal of biological chemistry* **282**, 21901 (2007).
138. J. T. Lis, P. Mason, J. Peng, D. H. Price, J. Werner, P-TEFb kinase recruitment and function at heat shock loci. *Genes & development* **14**, 792 (2000).
139. S. R. Eberhardy, P. J. Farnham, c-Myc mediates activation of the cad promoter via a post-RNA polymerase II recruitment mechanism. *The Journal of biological chemistry* **276**, 48562 (2001).
140. S. R. Eberhardy, P. J. Farnham, Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *The Journal of biological chemistry* **277**, 40156 (2002).

141. S. Kanazawa, L. Soucek, G. Evan, T. Okamoto, B. M. Peterlin, c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* **22**, 5707 (2003).
142. V. H. Cowling, M. D. Cole, The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding. *Molecular and cellular biology* **27**, 2059 (2007).
143. C. Simone *et al.*, Activation of MyoD-dependent transcription by cdk9/cyclin T2. *Oncogene* **21**, 4137 (2002).
144. C. Giacinti, L. Bagella, P. L. Puri, A. Giordano, C. Simone, MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. *Journal of cellular physiology* **206**, 807 (2006).
145. V. Bres, N. Gomes, L. Pickle, K. A. Jones, A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes & development* **19**, 1211 (2005).
146. B. M. Wittmann, K. Fujinaga, H. Deng, N. Ogba, M. M. Montano, The breast cell growth inhibitor, estrogen down regulated gene 1, modulates a novel functional interaction between estrogen receptor alpha and transcriptional elongation factor cyclin T1. *Oncogene* **24**, 5576 (2005).
147. A. J. Donner, C. C. Ebmeier, D. J. Taatjes, J. M. Espinosa, CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature structural & molecular biology* **17**, 194 (2010).
148. M. Belakavadi, J. D. Fondell, Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation. *Molecular and cellular biology* **30**, 2437 (2010).
149. H. Takahashi *et al.*, Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92 (2011).
150. W. Wang *et al.*, Mediator MED23 regulates basal transcription in vivo via an interaction with P-TEFb. *Transcription* **4**, 39 (2013).
151. T. Lenasi, B. M. Peterlin, M. Barboric, Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *The Journal of biological chemistry* **286**, 22758 (2011).
152. B. Cheng *et al.*, Functional association of Gdown1 with RNA polymerase II poised on human genes. *Molecular cell* **45**, 38 (2012).
153. J. Medlin *et al.*, P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *The EMBO journal* **24**, 4154 (2005).
154. N. P. Gomes *et al.*, Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes & development* **20**, 601 (2006).
155. Z. Yang *et al.*, Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular cell* **19**, 535 (2005).
156. M. K. Jang *et al.*, The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular cell* **19**, 523 (2005).
157. D. C. Hargreaves, T. Horng, R. C. I. N. C. J. Medzhitov, Pmid, Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**, 129 (2009).
158. A. Zippo *et al.*, Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* **138**, 1122 (2009).
159. Z. Yang, N. He, Q. Zhou, Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. *Molecular and cellular biology* **28**, 967 (2008).
160. A. Dey, A. Nishiyama, T. Karpova, J. McNally, K. Ozato, Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Molecular biology of the cell* **20**, 4899 (2009).
161. K. Mochizuki *et al.*, The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. *The Journal of biological chemistry* **283**, 9040 (2008).
162. B. N. Devaiah *et al.*, BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* **109**, 6927 (2012).
163. J. Pirngruber *et al.*, CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO reports* **10**, 894 (2009).
164. J. Pirngruber, A. Shchebet, S. A. Johnsen, Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell cycle (Georgetown, Tex.)* **8**, 3636 (2009).

165. A. Shchebet, O. Karpiuk, E. Kremmer, D. Eick, S. A. Johnsen, Phosphorylation by cyclin-dependent kinase-9 controls ubiquitin-conjugating enzyme-2A function. *Cell cycle (Georgetown, Tex.)* **11**, 2122 (2012).
166. A. Marcello *et al.*, Visualization of in vivo direct interaction between HIV-1 TAT and human cyclin T1 in specific subcellular compartments by fluorescence resonance energy transfer. *The Journal of biological chemistry* **276**, 39220 (2001).
167. C. H. Herrmann, M. A. Mancini, The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions. *Journal of cell science* **114**, 1491 (2001).
168. S. Egloff, H. Al-Rawaf, D. O'Reilly, S. Murphy, Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. *Molecular and cellular biology* **29**, 4002 (2009).
169. M. Barboric *et al.*, 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7798 (2009).
170. T. K. Ko, E. Kelly, J. Pines, CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *Journal of cell science* **114**, 2591 (2001).
171. F. Marques *et al.*, A new subfamily of high molecular mass CDC2-related kinases with PITAI/VRE motifs. *Biochemical and biophysical research communications* **279**, 832 (2000).
172. Y. Even *et al.*, CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *Journal of cellular biochemistry* **99**, 890 (2006).
173. R. Berro *et al.*, CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. *Journal of virology* **82**, 7155 (2008).
174. M. C. Edwards, C. Wong, S. J. Elledge, Human cyclin K, a novel RNA polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity. *Molecular and cellular biology* **18**, 4291 (1998).
175. T. J. Fu, J. Peng, G. Lee, D. H. Price, O. Flores, Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *The Journal of biological chemistry* **274**, 34527 (1999).
176. X. Lin, R. Taube, K. Fujinaga, B. M. Peterlin, P-TEFb containing cyclin K and Cdk9 can activate transcription via RNA. *The Journal of biological chemistry* **277**, 16873 (2002).
177. E. Urano *et al.*, Cyclin K/CPR4 inhibits primate lentiviral replication by inactivating Tat/positive transcription elongation factor b-dependent long terminal repeat transcription. *AIDS (London, England)* **22**, 1081 (2008).
178. S. Z. Khan, D. Mitra, Cyclin K inhibits HIV-1 gene expression and replication by interfering with cyclin-dependent kinase 9 (CDK9)-cyclin T1 interaction in Nef-dependent manner. *The Journal of biological chemistry* **286**, 22943 (2011).
179. D. S. Yu *et al.*, Cyclin-dependent kinase 9-cyclin K functions in the replication stress response. *EMBO reports* **11**, 876 (2010).
180. H. H. Chen, Y. C. Wang, M. J. Fann, Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Molecular and cellular biology* **26**, 2736 (2006).
181. H. H. Chen, Y. H. Wong, A. M. Geneviere, M. J. Fann, CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochemical and biophysical research communications* **354**, 735 (2007).
182. D. Blazek *et al.*, The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes & development* **25**, 2158 (2011).
183. B. Bartkowiak, A. L. Greenleaf, Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? *Transcription* **2**, 115 (2011).
184. Z. J. Zang *et al.*, Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing. *Cancer research* **71**, 29 (2011).
185. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609 (2011).
186. S. L. Carter *et al.*, Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology* **30**, 413 (2012).
187. C. Pattaro *et al.*, Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS genetics* **8**, e1002584 PHST (2012).

188. Q. Dai *et al.*, Cyclin K-containing kinase complexes maintain self-renewal in murine embryonic stem cells. *The Journal of biological chemistry* **287**, 25344 (2012).
189. Y. Zhao *et al.*, Involvement of cyclin K posttranscriptional regulation in the formation of Artemia diapause cysts. *PloS one* **7**, e32129 PHST (2012).
190. T. Narita *et al.*, Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Molecular and cellular biology* **23**, 1863 (2003).
191. P. W. Chiang *et al.*, Isolation, sequencing, and mapping of the human homologue of the yeast transcription factor, SPT5. *Genomics* **38**, 421 (1996).
192. P. W. Chiang *et al.*, Isolation and characterization of the human and mouse homologues (SUPT4H and Supt4h) of the yeast SPT4 gene. *Genomics* **34**, 368 (1996).
193. F. Winston, D. T. Chaleff, B. Valent, G. R. Fink, Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics* **107**, 179 (1984).
194. D. K. Kim *et al.*, Structure-function analysis of human Spt4: evidence that hSpt4 and hSpt5 exert their roles in transcriptional elongation as parts of the DSIF complex. *Genes to cells : devoted to molecular & cellular mechanisms* **8**, 371 (2003).
195. M. S. Swanson, F. Winston, SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics* **132**, 325 (1992).
196. E. A. Malone, J. S. Fassler, F. Winston, Molecular and genetic characterization of SPT4, a gene important for transcription initiation in Saccharomyces cerevisiae. *Molecular & general genetics : MGG* **237**, 449 (1993).
197. B. Schwer, S. Schneider, Y. Pei, A. Aronova, S. Shuman, Characterization of the Schizosaccharomyces pombe Spt5-Spt4 complex. *RNA (New York, N.Y.)* **15**, 1241 (2009).
198. M. S. Swanson, E. A. Malone, F. Winston, SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Molecular and cellular biology* **11**, 4286 (1991).
199. S. Guo *et al.*, A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366 (2000).
200. M. Prabhakaran, R. L. Kelley, Mutations in the Transcription Elongation Factor SPT5 Disrupt a Reporter for Dosage Compensation in Drosophila. *PLoS genetics* **8**, e1003073 PHST (2012).
201. S. A. Patel, M. C. Simon, Functional analysis of the Cdk7.cyclin H.Mat1 complex in mouse embryonic stem cells and embryos. *The Journal of biological chemistry* **285**, 15587 (2010).
202. M. A. Basrai, J. Kingsbury, D. Koshland, F. Spencer, P. Hieter, Faithful chromosome transmission requires Spt4p, a putative regulator of chromatin structure in Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 2838 (1996).
203. S. Wenzel, K. Schweimer, P. Rosch, B. M. Wohrl, The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with alpha/beta type topology. *Biochemical and biophysical research communications* **370**, 414 (2008).
204. M. Guo *et al.*, Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. *Structure (London, England : 1993)* **16**, 1649 (2008).
205. Y. Yamaguchi, H. Shibata, H. Handa, Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. *Biochimica et biophysica acta* **2012 Nov 29**, doi 10 (2012).
206. T. Wada *et al.*, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development* **12**, 343 (1998).
207. Y. Yamaguchi *et al.*, Structure and function of the human transcription elongation factor DSIF. *The Journal of biological chemistry* **274**, 8085 (1999).
208. M. Palangat, D. B. Renner, D. H. Price, R. Landick, A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15036 (2005).
209. G. S. Chang *et al.*, Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in Dictyostelium. *Genome research* **22**, 1098 (2012).
210. Y. Yamaguchi *et al.*, NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41 (1999).
211. T. Wada, T. Takagi, Y. Yamaguchi, D. Watanabe, H. Handa, Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *The EMBO journal* **17**, 7395 (1998).

212. D. Ivanov, Y. T. Kwak, J. Guo, R. B. Gaynor, Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Molecular and cellular biology* **20**, 2970 (2000).
213. K. Fujinaga *et al.*, Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and cellular biology* **24**, 787 (2004).
214. T. Yamada *et al.*, P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular cell* **21**, 227 (2006).
215. Y. H. Ping, T. M. Rana, DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *The Journal of biological chemistry* **276**, 12951 (2001).
216. A. Missra, D. S. Gilmour, Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11301 (2010).
217. L. Amir-Zilberstein *et al.*, Differential regulation of NF-kappaB by elongation factors is determined by core promoter type. *Molecular and cellular biology* **27**, 5246 (2007).
218. F. Werner, A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of molecular biology* **417**, 13 (2012).
219. G. A. Hartzog, T. Wada, H. Handa, F. Winston, Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes & development* **12**, 357 (1998).
220. S. Schneider, Y. Pei, S. Shuman, B. Schwer, Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Molecular and cellular biology* **30**, 2353 (2010).
221. V. Svetlov, E. Nudler, Clamping the clamp of RNA polymerase. *The EMBO journal* **30**, 1190 (2011).
222. G. A. Hartzog, C. D. Kaplan, Competing for the clamp: promoting RNA polymerase processivity and managing the transition from initiation to elongation. *Molecular cell* **43**, 161 (2011).
223. G. A. Hartzog, J. Fu, The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et biophysica acta* **2012 Sep 6**, doi 10 (2012).
224. T. Komori, N. Inukai, T. Yamada, Y. Yamaguchi, H. Handa, Role of human transcription elongation factor DSIF in the suppression of senescence and apoptosis. *Genes to cells : devoted to molecular & cellular mechanisms* **14**, 343 (2009).
225. A. Hirtreiter *et al.*, Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic acids research* **38**, 4040 (2010).
226. F. W. Martinez-Rucobo, S. Sainsbury, A. C. Cheung, P. Cramer, Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal* **30**, 1302 (2011).
227. F. Beckouet, S. Mariotte-Labarre, G. Peyroche, Y. Nogi, P. Thuriaux, Rpa43 and its partners in the yeast RNA polymerase I transcription complex. *FEBS letters* **585**, 3355 (2011).
228. B. J. Klein *et al.*, RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 546 (2011).
229. A. A. Stachora, R. E. Schafer, M. Pohlmeier, G. Maier, H. Ponstingl, Human Supt5h protein, a putative modulator of chromatin structure, is reversibly phosphorylated in mitosis. *FEBS letters* **409**, 74 (1997).
230. J. B. Kim, P. A. Sharp, Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *The Journal of biological chemistry* **276**, 12317 (2001).
231. S. Larochelle *et al.*, Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nature structural & molecular biology* **13**, 55 (2006).
232. T. K. Quan, G. A. Hartzog, Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. *Genetics* **184**, 321 (2010).
233. J. R. Stevens *et al.*, FACT, the Bur kinase pathway, and the histone co-repressor HirC have overlapping nucleosome-related roles in yeast transcription elongation. *PloS one* **6**, e25644 PHST (2011).

234. N. J. Krogan *et al.*, RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Molecular and cellular biology* **22**, 6979 (2002).
235. D. L. Lindstrom *et al.*, Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Molecular and cellular biology* **23**, 1368 (2003).
236. S. Drouin *et al.*, DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS genetics* **6**, e1001173 PHST (2010).
237. S. L. Squazzo *et al.*, The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO journal* **21**, 1764 (2002).
238. H. Chen *et al.*, Repression of RNA polymerase II elongation in vivo is critically dependent on the C-terminus of Spt5. *PloS one* **4**, e6918 PHST (2009).
239. Y. Wen, A. J. Shatkin, Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & development* **13**, 1774 (1999).
240. K. Glover-Cutter, S. Kim, J. Espinosa, D. L. Bentley, RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology* **15**, 71 (2008).
241. Y. Xiao *et al.*, Analysis of a splice array experiment elucidates roles of chromatin elongation factor Spt4-5 in splicing. *PLoS computational biology* **1**, e39 PHST (2005).
242. A. Mayer *et al.*, The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Molecular and cellular biology* **32**, 1321 (2012).
243. E. D. Andrulis, E. Guzman, P. Doring, J. Werner, J. T. Lis, High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes & development* **14**, 2635 (2000).
244. Y. Pei, S. Shuman, Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *The Journal of biological chemistry* **277**, 19639 (2002).
245. D. A. Gilchrist *et al.*, NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & development* **22**, 1921 (2008).
246. R. Pavri *et al.*, Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122 (2010).
247. P. B. Rahl *et al.*, c-Myc regulates transcriptional pause release. *Cell* **141**, 432 (2010).
248. J. Drogat, D. Hermand, Gene-specific requirement of RNA polymerase II CTD phosphorylation. *Molecular microbiology* **84**, 995 (2012).
249. T. U. Banisch, M. Goudarzi, E. Raz, Small RNAs in germ cell development. *Current topics in developmental biology* **99**, 79 (2012).
250. H. Ishizu, H. Siomi, M. C. Siomi, Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes & development* **26**, 2361 (2012).
251. M. S. Cook, R. Blelloch, Small RNAs in germline development. *Current topics in developmental biology* **102**, 159 (2013).
252. J. A. Hackett, J. J. Zylicz, M. A. Surani, Parallel mechanisms of epigenetic reprogramming in the germline. *Trends in genetics : TIG* **28**, 164 (2012).
253. M. E. Gill, S. Erkek, A. H. Peters, Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Current opinion in cell biology* **24**, 387 (2012).
254. P. Rangan *et al.*, Temporal and spatial control of germ-plasm RNAs. *Current biology : CB* **19**, 72 (2009).
255. C. Merritt, D. Rasoloson, D. Ko, G. Seydoux, 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Curr Biol* **18**, 1476 (2008).
256. F. Lai, M. L. King, Repressive translational control in germ cells. *Molecular reproduction and development*, doi 10 (2013).
257. F. Spitz, E. E. Furlong, Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics* **13**, 613 (2012).
258. J. A. D'Alessio, K. J. Wright, R. Tjian, Shifting players and paradigms in cell-specific transcription. *Molecular cell* **36**, 924 (2009).
259. F. Muller, A. Zaucker, L. Tora, Developmental regulation of transcription initiation: more than just changing the actors. *Current opinion in genetics & development* **20**, 533 (2010).
260. M. C. Thomas, C. M. Chiang, The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology* **41**, 105 (2006).

261. U. Kolthur-Seetharam, I. Martianov, I. Davidson, Specialization of the general transcriptional machinery in male germ cells. *Cell cycle (Georgetown, Tex.)* **7**, 3493 (2008).
262. T. Hoiby, H. Zhou, D. J. Mitsiou, H. G. Stunnenberg, A facelift for the general transcription factor TFIIA. *Biochimica et biophysica acta* **1769**, 429 (2007).
263. V. C. Li *et al.*, Molecular evolution of the testis TAFs of Drosophila. *Molecular biology and evolution* **26**, 1103 (2009).
264. Y. Cheng *et al.*, Abnormal sperm in mice lacking the Taf7l gene. *Molecular and cellular biology* **27**, 2582 (2007).
265. A. E. Falender, M. Shimada, Y. K. Lo, J. S. Richards, TAF4b, a TBP associated factor, is required for oocyte development and function. *Dev Biol* **288**, 405 (2005).
266. W. Akhtar, G. J. Veenstra, TBP2 is a substitute for TBP in Xenopus oocyte transcription. *BMC biology* **7**, 45 PHST (2009).
267. E. Gazdag *et al.*, TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes & development* **23**, 2210 (2009).
268. A. Bahat *et al.*, TAF4b and TAF4 differentially regulate mouse embryonic stem cells maintenance and proliferation. *Genes to cells : devoted to molecular & cellular mechanisms* **18**, 225 (2013).
269. A. B. Upadhyaya, S. H. Lee, J. DeJong, Identification of a general transcription factor TFIIAalpha/beta homolog selectively expressed in testis. *The Journal of biological chemistry* **274**, 18040 (1999).
270. J. Ozer, P. A. Moore, P. M. Lieberman, A testis-specific transcription factor IIA (TFIIAtau) stimulates TATA-binding protein-DNA binding and transcription activation. *The Journal of biological chemistry* **275**, 122 (2000).
271. S. Y. Han *et al.*, TFIIAalpha/beta-like factor is encoded by a germ cell-specific gene whose expression is up-regulated with other general transcription factors during spermatogenesis in the mouse. *Biology of reproduction* **64**, 507 (2001).
272. R. N. Seervai, G. M. Wessel, Lessons for inductive germline determination. *Molecular reproduction and development*, doi 10 (2013).
273. J. Solana, Closing the circle of germline and stem cells: the Primordial Stem Cell hypothesis. *EvoDevo* **4**, 2 PHST (2013).
274. A. Nakamura, M. Shirae-Kurabayashi, K. Hanyu-Nakamura, Repression of early zygotic transcription in the germline. *Current opinion in cell biology* **22**, 709 (2010).
275. M. M. Lamb, C. D. Laird, Increase in nuclear poly(A)-containing RNA at syncytial blastoderm in Drosophila melanogaster embryos. *Developmental biology* **52**, 31 (1976).
276. M. Zalokar, Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. *Developmental biology* **49**, 425 (1976).
277. R. G. Martinho, P. S. Kunwar, J. Casanova, R. C. I. N. C. B. M. Lehmann, P. R, A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. *Current biology : CB* **14**, 159 (2004).
278. A. Nakamura, R. Amikura, M. Mukai, S. Kobayashi, P. F. Lasko, Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science (New York, N.Y.)* **274**, 2075 (1996).
279. K. Hanyu-Nakamura, H. Sonobe-Nojima, A. Tanigawa, P. Lasko, A. Nakamura, Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* **451**, 730 (2008).
280. G. Timinszky, M. Bortfeld, A. G. Ladurner, Repression of RNA polymerase II transcription by a Drosophila oligopeptide. *PloS one* **3**, e2506 PHST (2008).
281. J. M. de Las Heras, R. G. Martinho, R. Lehmann, J. Casanova, A functional antagonism between the pgc germline repressor and torso in the development of somatic cells. *EMBO reports* **10**, 1059 (2009).
282. T. Guven-Ozkan, Y. Nishi, S. M. Robertson, R. C. I. N. C. O. Lin, Pmid, Global transcriptional repression in C. elegans germline precursors by regulated sequestration of TAF-4. *Cell* **135**, 149 (2008).
283. S. E. Mango, C. J. Thorpe, P. R. Martin, S. H. Chamberlain, B. Bowerman, Two maternal genes, apx-1 and pie-1, are required to distinguish the fates of equivalent blastomeres in the early Caenorhabditis elegans embryo. *Development (Cambridge, England)* **120**, 2305 (1994).

284. C. C. Mello, B. W. Draper, M. Krause, H. Weintraub, J. R. Priess, The pie-1 and mex-1 genes and maternal control of blastomere identity in early C. elegans embryos. *Cell* **70**, 163 (1992).
285. C. C. Mello *et al.*, The PIE-1 protein and germline specification in C. elegans embryos. *Nature* **382**, 710 (1996).
286. G. Seydoux, M. A. Dunn, Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development (Cambridge, England)* **124**, 2191 (1997).
287. G. Seydoux *et al.*, Repression of gene expression in the embryonic germ lineage of C. elegans. *Nature* **382**, 713 (1996).
288. C. Batchelder *et al.*, Transcriptional repression by the Caenorhabditis elegans germ-line protein PIE-1. *Genes & development* **13**, 202 (1999).
289. F. Zhang, M. Barboric, T. K. Blackwell, B. M. Peterlin, A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. *Genes & development* **17**, 748 (2003).
290. D. Ghosh, G. Seydoux, Inhibition of transcription by the Caenorhabditis elegans germline protein PIE-1: genetic evidence for distinct mechanisms targeting initiation and elongation. *Genetics* **178**, 235 (2008).
291. C. Tenenhaus, K. Subramaniam, M. A. Dunn, G. Seydoux, PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of Caenorhabditis elegans. *Genes & development* **15**, 1031 (2001).
292. G. Kumano, N. Takatori, T. Negishi, T. Takada, H. Nishida, A maternal factor unique to ascidians silences the germline via binding to P-TEFb and RNAP II regulation. *Current biology : CB* **21**, 1308 (2011).
293. T. Venkatarama *et al.*, Repression of zygotic gene expression in the Xenopus germline. *Development (Cambridge, England)* **137**, 651 (2010).
294. Y. Seki *et al.*, Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development (Cambridge, England)* **134**, 2627 (2007).
295. N. Pazdernik, T. Schedl, Introduction to germ cell development in Caenorhabditis elegans. *Advances in experimental medicine and biology* **757**, 1 (2013).
296. S. Zanetti, A. Puoti, Sex determination in the Caenorhabditis elegans germline. *Advances in experimental medicine and biology* **757**, 41 (2013).
297. P. M. Van Wynsberghe, E. M. Maine, Epigenetic control of germline development. *Advances in experimental medicine and biology* **757**, 373 (2013).
298. S. E. Castel, R. A. Martienssen, RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100 (2013).
299. S. Brenner, The genetics of Caenorhabditis elegans. *Genetics* **77**, 71 (1974).
300. L. R. Baugh, A. A. Hill, D. K. Slonim, E. L. Brown, C. P. Hunter, Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development* **130**, 889 (2003).
301. T. M. Rogalski, D. L. Riddle, A Caenorhabditis elegans RNA polymerase II gene, ama-1 IV, and nearby essential genes. *Genetics* **118**, 61 (1988).
302. T. M. Rogalski, A. M. Bullerjahn, D. L. Riddle, Lethal and amanitin-resistance mutations in the Caenorhabditis elegans ama-1 and ama-2 genes. *Genetics* **120**, 409 (1988).
303. T. M. Rogalski, M. Golomb, D. L. Riddle, Mutant Caenorhabditis elegans RNA polymerase II with a 20,000-fold reduced sensitivity to alpha-amanitin. *Genetics* **126**, 889 (1990).
304. C. D. Kaplan, K. M. Larsson, R. D. Kornberg, The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell* **30**, 547 (2008).
305. E. A. Bowman, D. L. Riddle, W. Kelly, Amino Acid Substitutions in the Caenorhabditis elegans RNA Polymerase II Large Subunit AMA-1/RPB-1 that Result in alpha-Amanitin Resistance and/or Reduced Function. *G3* **1**, 411 (2011).
306. A. Golden *et al.*, Metaphase to anaphase (mat) transition-defective mutants in Caenorhabditis elegans. *The Journal of cell biology* **151**, 1469 (2000).
307. J. A. Powell-Coffman, J. Knight, W. B. Wood, Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev Biol* **178**, 472 (1996).
308. E. A. Bowman, G. Seydoux, W. G. Kelly, Temperature sensitive mutants of the RNA Polymerase II TFIID initiation factor, taf-6.2. . *Worm Breeder's Gazette* **18**, (2011).

309. H. Furuhashi *et al.*, Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics & chromatin* **3**, 15 PHST (2010).
310. R. B. Deal, S. Henikoff, The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. *Nature protocols* **6**, 56 (2011).

CHAPTER 2: AMINO ACID SUBSTITUTIONS IN THE C. ELEGANS RNA POLYMERASE II LARGE SUBUNIT, AMA-1/RPB-1, THAT RESULT IN A-AMANITIN RESISTANCE AND/OR REDUCED FUNCTION

Bowman, E.A.; Riddle, D.L.; Kelly, W.G.[2]

In order to try to characterize temperature sensitive mutants of *ama-1*, the large subunit of RNA Polymerase II, I identified the specific mutations in a large collection of *ama-1* mutants previously isolated (*1-3*). Here, I described the specific point mutations in this collection, and further speculate about the defects caused by these mutants by mapping them onto the structure of the *S. cerevisiae* Pol II complex. I also further characterize the phenotype a temperature sensitive *ama-1* mutation that efficiently reduces transcriptional activity at restrictive temperatures. This work has been previously published (*4*).

[2] E.A.B. wrote the manuscript and conducted all experiments. D.L.R. provided the worm strains and helped conceive experiments. W.G.K. conceived, supervised and provided funding for the study.

Abstract

Mutations in the *C. elegans* RNA Polymerase II AMA-1/RPB-1 subunit that cause α-amanitin resistance and/or developmental defects were isolated previously. I identified 12 of these mutations and mapped them onto the *S. cerevisiae* RPB-1 structure to provide insight into AMA-1 regions that are essential for development in a multi-cellular organism.

Introduction

The DNA-directed RNA polymerase II (Pol II) holoenzyme is a ~500 kDa complex responsible for transcribing protein-coding and other genes in eukaryotes. The Pol II complex is composed of 12 subunits that are highly conserved from yeast to human (Figure 2.1). The largest subunit, AMA-1/RPB1, is a ~200kDa multi-domain protein which makes up much of the functional core of Pol II (*5*). During transcription, dsDNA enters between the two lobes of the Pol II jaw-like structure, largely contacting the cleft domain of the AMA-1/RPB1 protein. At the base of the jaws lies the active site domain of AMA-1/RPB1. While this is the domain primarily responsible for RNA polymerization, components of both the cleft and funnel domains are required for transcription as they allow translocation of the template DNA within the Pol II complex. Specifically, the AMA-1/RPB1 trigger loop, located outside of the active site, is essential for proper catalysis by monitoring substrate selectivity as ribonucleotides pass through the AMA-1/RPB1 pore domain to the active site (*6*). In addition, interactions between the trigger loop, an α-helix that spans between the Pol II jaws, and the bridge helix are thought to mediate the flexibility of the bridge helix which is important for translocation of the duplex DNA (*7, 8*).

Understanding the mechanism of Pol II transcription has been aided by specific inhibitors of this complex. The “death cap” mushroom toxin, α-amanitin, prevents transcriptional elongation in most eukaryotes by sterically blocking the intramolecular interactions between the trigger loop and bridge helix required for translocation (*6-8*). Mutations that confer α-amanitin-resistance are

mostly located in the "funnel" domain of AMA-1/RPB1, a region close to the active site of the enzyme which contains the trigger loop (*9*).

Studies in *C. elegans* were among the first to characterize mutations in AMA-1/RPB1 and their effects on both α-amanitin binding and developmental processes (*1-3*). These studies mapped point mutations within the *ama-1* gene but did not identify the mutations (*10*). The original collection of mutations represents a variety of *ama-1* alleles, including α-amanitin resistant, hypomorphic, and putative null alleles. In order to further define the nature of these alleles, I sequenced a number of the original mutations and mapped them onto the highly homologous *S. cerevisiae* Pol II structure to provide potential structure-function information for these regions of the protein. I show that mutations that disrupt susceptibility to α-amanitin lie within the toxin binding site and that one of these also disrupts Pol II function at elevated temperatures. Furthermore, I identify mutations in conserved regions of the protein that cause significant alterations in RNA polymerase function and provide further insight into its mechanism.

Results

A large, valuable collection of mutant *C. elegans* worms with changes in the RNA polymerase II large subunit gene, *ama-1/rpb-1* was generated by EMS mutagenesis in the 1980s. This represents the second largest collection of metazoan mutants in *ama-1/rpb-1*, yet the identification of the corresponding changes in the *ama-1* gene that result in these defects have remained unexplored. I have thus revisited this resource to further characterize these *ama-1* mutations and to investigate their possible structural effects.

α-Amanitin Resistant Mutations

α-Amanitin binds AMA-1/RPB1 and blocks transcriptional elongation by preventing AMA-1/RPB1 trigger loop-mediated substrate selection and bridge helix flexibility during translocation

(*6-9*). I sequenced two previously isolated mutants in the *C. elegans ama-1* gene that demonstrated α-amanitin-resistance (*1-3*). For one allele, *ama-1(m118),* I verified a mutation previously identified as a C777Y substitution (D. M. Bird and D. L. Riddle, unpublished). I also identified a novel *C. elegans* α-amanitin-resistant mutation, *ama-1(m322)*, as an R739H substitution (Table 2.1). Mapping the *ama-1* sequence onto the *S. cerevesiae* structure of RPB1 showed that both mutations were in the trigger loop of AMA-1 "funnel" domain (Figure 2.2). The arginine residue in yeast that is homologous to R739 in AMA-1, R726, provides a hydrogen bond with α-amanitin, suggesting that R739 performs a similar function in *C. elegans* (Figure 2.2b, Table 2.1) (*9*). An R to H change in the *C. elegans* protein would alter the distance critical for this hydrogen bond in the α-amanitin binding pocket and weaken this interaction. An identical substitution in the corresponding amino acid in *Drosophila* and in mouse cells has also been shown to inhibit α-amanitin binding [RpbII215-4; R741H (*11-13*); RpII215-A21, R749P, (*14*)]. Interestingly, the corresponding mutation in yeast suppresses a transcription start site defect [sit1-290, (*15*)].

I also identified the mutations in an α-amanitin "super-resistant" strain [DR1099: *ama-1(m118m526)*] that was isolated after further mutagenesis of the *ama-1(m118)* strain (*1*). The *ama-1* gene in this strain carries a second mutation, G785E, which I predict to sterically block the α-amanitin binding pocket. Alteration of the corresponding amino acid also inhibits α-amanitin binding in the mouse protein (*14*).

In addition to the high resistance to α-amanitin, DR1099 displays temperature-sensitive defects consistent with defective RNA Pol II function (*1*). *ama-1(RNAi)* embryos exhibit gastrulation defects and arrest at ~120 cells (*16*). Similarly, when DR1099 animals are shifted to a restrictive temperature (25° C), their embryos fail to gastrulate and arrest at ~120 cells (Table 2.2). Importantly, however, immunofluorescence analyses of these embryos detected significant levels of an epitope that correlates with the elongating form of RNA Pol II, phosphorylation of Ser2 on the CTD repeat peptide [(*17*); data not shown]. Thus, while RNA Pol II function is not

completely compromised at the restrictive temperature, elongation processivity may be significantly affected, perhaps by a temperature-dependent mechanism that mimics inhibition of elongation by α-amanitin. This seems likely when considering the potential impact of G785E on the predicted structure. The bulky and charged E side-chain could insert within and disrupt the interface between the trigger loop and bridge helix of AMA-1 (Figure 2.2b). It is unclear, however, whether the effects of the G785E substitution are autonomous, or exist only in the context of the second *m118* C777Y substitution.

Interestingly, residues that when altered confer α-amanitin resistance in *C. elegans* are also changed in the *Giardia lamblia* rpb1 gene, which is naturally α-amanitin resistant. The positions in *Girardia rbp1* that correspond to those in *C. elegans ama-1*, and the amino acid differences between the species, are as follows: *Ce* R739=*Gl* S851, *Ce* C777=*Gl* S889, and *Ce* G785=*Gl* S897 (*18*).

Hypomorphic and Null Mutations

Recessive-lethal alleles of *ama-1* were also isolated through further mutagenesis of the *ama-1(m118)* strain. (*2, 3*). The phenotypes of these mutants range from temperature sensitive (*ts*) sterile (presumed hypomorphic allele) to L1 arrest (presumed null allele; i.e., phenocopies *ama-1* deletion alleles). In order to better correlate Pol II structural alteration with phenotype, I sequenced ten of these mutants and mapped them onto the *S. cerevisiae* Pol II structure (Table 2.1, Figure 2.3a, b). The following mutations represent alterations in residues and domains that are highly conserved among yeast, worms, flies, and humans (see Figure 2.1 for AMA-1 sequence alignment).

The *ama-1(m370)* mutation, G636R, which yields a null function phenotype, affects a residue predicted to lie at the cap of a beta sheet in the “pore” domain of AMA-1. This amino acid is in a tightly packed region of the Pol II protein and the dramatic G to R residue change likely disrupts this packing and thus indirectly disrupts catalysis (Figure 2.3d).

The *ama-1(m370)* strain, DR880 [*ama-1(m118m370*)], was further mutagenized to identify suppressors of the *ama-1(m370)* lethal phenotype (*2*). I identified the presumed suppressing mutation in one of these strains, DR976: *ama-1(m118m370m417)*, as an A746V conversion. Surprisingly, upon placement onto the yeast structure, the positions of G636 (*ama-1(m370)* mutation) and A746 [*ama-1(m417)* mutation] are predicted to lie at least 27Å apart within two different domains of AMA-1/RPB1 (Figure 2.3d). G636 is predicted to lie within the pore domain, and A746 within the distant funnel domain. Further outcross experiments confirmed the very tight linkage of *ama-1(m370)* and *ama-1(m417)*, supporting the conclusion that the suppression phenotype is indeed caused by the A746V conversion.

A similar long-distance effect has also been observed in *Drosophila*. The RpII215^{K1} mutation (*D.m.* S678N, *S.c.* S663) is found within the pore domain of *Drosophila* RPB1 and causes a *ts* phenotype. This phenotype can be rescued by two different intragenic mutations in the funnel domain (RpII215^{R4}: *D.m.* H713L, *S.c.* 698-Q; RpII215^{R10}: *D.m.* S747L, *S.c.* 732-L), which are 49Å and 24Å away from RpII215^{K1}, respectively (*11, 19, 20*). Sequences of more internal revertants of *ama-1* hypomorphs might reveal how common it is for intragenic revertants to exert their effects over long distances.

The remaining five mutations identified in this study that are highly conserved residues all lie within the "cleft" domain of AMA-1/RPB1 (Figure 2.3e) (*5*). This domain makes up a large portion of the DNA binding region of the lower "jaw" of Pol II. The *ama-1(m332)* mutation, V869M, which results in a null phenotype, corresponds to a residue in the *S. cerevisiae* structure that lies within a densely packed region of Pol II. While this particular amino acid is not conserved in yeast, it is in a very well conserved domain and changing the small hydrophobic valine to a larger, more hydrophilic methionine may cause clashing with nearby side chains and disrupt this packing. The *ama-1(m235)* and *ama-1(m367)* mutants (G1086E and G1110E respectively) correspond to yeast positions that are both found in the trigger loop of the cleft domain. This domain is thought to couple nucleotide recognition and catalysis during Pol II

translocation, and substitutions at either glycine could decrease flexibility of the catalytically-important bridge helix (*6, 7*). The *ama-1(m235)* mutation is also in close proximity to Q838 in the bridge helix (Figure 2.3f). This contact is likely important since the sequence of the entire bridge helix is highly conserved, so the glycine to glutamate is presumably poorly tolerated. The G1110 residue does not contact the bridge helix directly, but it is also in a well-packed region of AMA-1. Conversion of G1110 to E in *ama-1(m367)* presumably disrupts packing in this domain. An EMS-induced revertant of *ama-1(m367)* (DR877), contained only the *ama-1(m118)* α-amanitin-resistant mutation; the E1110 substitution had been converted back to glycine.

The *ama-1(m238)* mutation, G1406R, results in *ts* maternal effect embryonic lethal phenotype. G1406 is predicted to be in the "switch 1 domain" of the cleft domain which is thought to be important for the Pol II conformational changes that allow template binding (*21*). Mutation of this amino acid in yeast causes cold sensitivity and slow growth by affecting transcription start site selection (*22, 23*). The *ts* phenotype in *C. elegans* may involve a similar mechanism.

The amino acid substitution in *ama-1(m396)*, S1336F, is also found in the cleft domain and is specifically in the binding interface between RPB1 and RPB5 (Figure 2.3g). In yeast, the corresponding amino acid, T1318, is predicted to have ionic interactions with R 11 and R14 of the RPB5 subunit. This binding is likely conserved in *C. elegans* as this side chain hydroxyl and the RPB-5 arginine residues are conserved (RPB-5 R10 and R13 in *C. elegans*), and mutation of this amino acid could disrupt RPB1-5 interactions.

Two mutations identified by sequencing resulted in amino acid substitutions in positions where there was little conservation. The *ama-1(m251)* mutation, A363V, has a temperature sensitive maternal effect embryonic lethal phenotype. A363 corresponds to a glycine amino acid in yeast and a proline in flies and humans. Although the residue itself is not conserved, the surrounding residues are highly conserved among all four species, and is located within the active site domain of yeast RPB1 (Figure 2.3c). While this mutation results in a conservative amino acid change, A363 is also predicted to lie within the RPB1-RPB2 interface, and disruption of this

interaction may explain its *ts* phenotype. However, how a conservative change in *C. elegans* can affect AMA-1 function, yet the normal protein in flies and humans is maintained as a proline, is not understood.

The *ama-1(m236)* mutation, N1051I, results in sterility and *ts* larval arrest and is predicted to lie within the AMA-1/RPB1 "foot" domain. The asparagine in *C. elegans* is a significant alteration relative to the other species, which have either a serine (yeast) or threonine (fly and human). The surrounding amino acids are also significantly diverged, although this domain serves as a binding site for the yeast mRNA capping enzyme, CE (*24*). The yeast amino acid position corresponding to N1051 [T1038] lies on the edge of the CE electron density in the yeast Pol II-CE co-crystal (*24*). It is likely that much of the structure of this region is devoted to the proper presentation of important/conserved residues for this interaction. Thus the N to I substitution might result in temperature-dependent instability in the interaction between the *C. elegans* capping enzyme and AMA-1.

Discussion

Sequencing these previously identified mutants of C. elegans *ama-1* helps to complete mutagenesis studies done over twenty years ago. Overall, the positions of the mutations identified by sequencing closely match their positions originally determined by fine-structure genetic mapping (Figure 2.3a). The genetic results thus provide strong supporting data to conclude that the nucleotide changes identified are causative for the functional phenotypes observed in these mutants.

I identified two novel *C. elegans* α-amanitin-resistant mutations. One of these mutations, *ama-1(m526)*, confers a tight *ts* maternal effect lethal phenotype that may result from reduced efficiency of elongation, and could be useful in further studies of Pol II function. In addition, assigning the structural positions of null and hypomorphic mutations may provide important structure-function clues for understanding how these mutations lead to different functional

consequences, which have biological read-outs as different developmental phenotypes. Further biochemical analyses on the structure-function relationship of the mutations isolated from these early genetic screens will provide new information about the Pol II structure, and this emphasizes the mutual benefits that combined genetic and biochemical/structural approaches can provide. This report should serve as a valuable community resource for those seeking to understand such structure/function relationships, as well as those seeking to employ defined *ama-1* mutations in their studies.

Acknowledgements

Work is funded by NIH GM 63102 (W.K.), and an NSF GRFP (E.A.B.). I would like to thank Danny Reines, Sujata Bhattacharyya, and Rebecca Adams for reading and helpful comments on the manuscript.

Tables

Table 2.1: Summary of α-amanitin-resistance and hypomorphic AMA-1 mutations

Strain [a]	Allele	DNA mutation	AA change	Corresponding S. cerevisiae AA	Location in structure [b]	Terminal phenotype [c] 20°	25°
N2		N. M. [d]					
DR680 [e]	m118	3891 g→a	777 C→Y	764 C	Funnel, near α-amanitin binding site	Adult (F)	Adult (F)
DR786 [e]	m322	3479 g→a	739 R→H	726 R	Funnel, near α-amanitin binding site	Adult (F)	Adult (F)
DR1099 [e]	m118	3891g→a	777 C→Y	764 C	Funnel, near α-amanitin binding site	Adult (F)	Adult (ME)
	m526	3915 g→a	785 G→E	772 G	Funnel, near α-amanitin binding site		
DR731 [f]	m118	3891g→a	777 C→Y	764 C		Adult (F)	Adult (ME)
	m251	1717 c→t	363 A→V	355 G	Active site, α-helix 8, between rpb1-2		
DR730 [f]	m118	3891g→a	777 C→Y	764 C		Adult (F)	Adult (ME)
	m238	6621 g→a	1406 G→R	1388 G	Cleft, loop		
DR892 [f]	m118	3891 g→a	777 C→Y	764 C		Adult (ME)	Mid larval arrest
	m396	6412 c→t	1336 S→F	1318 T	Cleft, β-sheet 44, between rpb1-5		
DR682 [f]	m118	3891g→a	777 C→Y	764 C		Mid larval to adult (ME) arrest	Mid larval arrest
	m235	5208 g→a	1086 G→E	1073 G	Cleft (trigger loop), β-sheet 36		
DR683 [f]	m118	3891g→a	777 C→Y	764 C		Adult (ST)	L1 larval arrest
	m236	5103 a→t	1051 N→I	1038 T	Foot, β-sheet 35		
DR811 [g]	m118	3891g→a	777 C→Y	764 C		L1 larval arrest	L1 larval arrest
	m332	4166 g→a	869 V→M	856 T	Cleft, α-helix 24		
DR880 [g]	m118	3891g→a	777 C→Y	764 C		L1 larval arrest	L1 larval arrest
	m370	3169 g→a	636 G→R	623 G	Pore 1, β-sheet 17		
DR976 [h]	m118	3891g→a	777 C→Y	764 C		Adult (F)	Adult (F)
	m370	3169 g→a	636 G→R	623 G			
	m417	3500 c→t	746 A →V	733 A	Funnel, β-sheet 21		
DR877 [g]	m118	3891g→a	777 C→Y	764 C		L1 larval arrest	L1 larval arrest
	m367	5280 g→a	1110 G→E	1097 G	Cleft (trigger loop), β-sheet 37		
DR966 [i]	m118	3891g→a	777 C→Y	764 C		Adult (F)	Adult (F)
	m367	N. M. [d]					
	m414	N. M. [d]					

2-4 Kb overlapping fragments of *ama-1* were amplified from 10 worms of each genotype using high-fidelity Phusion polymerase (Finnzymes) and sequenced (Macrogen USA Sequencing). Sequencing covered the entire coding region, including introns and exons. The sequence was compared to the F36A4.7 (*ama-1*) sequence on wormbase.org. If a mutation was identified, the corresponding *ama-1* fragment was independently amplified and sequence to confirm that the mutation was not due to the amplification step. For strains with early arrest phenotypes (DR811, DR880, DR877), heterozygous balanced worms were used for amplification and sequencing, and a mutation was identified as a heterozygous (double) peak in the sequencing chromatogram. If a mutation was identified in a balanced strain, the corresponding *ama-1* fragment from homozygous *ama-1* mutant worms [marked by *dpy-13(e184)*] were amplified and

sequenced. Numbering of the DNA sequence is from the genomic sequence, starting from the translation start site. All DNA mutations that were identified were found within exons; no mutations within introns were observed. The presumed amino acid change that would result from the DNA mutation is indicated. The *C. elegans* and *S. cereviase* AMA-1/RPB-1 amino acid sequences were aligned using Clustal W (*25*) and the homologous *S. cerevisiae* amino acid corresponding to the mutated residues in each strain are indicated.

[a] Isolated previously (DR680, DR786, DR683, DR682, DR730, DR731, (*3*); DR1099, (*1*); DR892, DR811, DR880, DR976 DR877, DR966, (*2*)

[b] Domain and secondary structure specified as in (*5*)

[c] Phenotypes characterized previously (*2*). ME, maternal effect embryonic lethal; ST, does not lay eggs; F, fertile, producing 70-90 progeny.; L1 arrest is a null phenotype: terminal phenotype of AMA-1 deletion allele is L1 arrest, indicating there is enough wild type maternally inherited AMA-1 activity to complete embryogenesis (*3*)

[d] No Mutation

[e] α-amanitin resistant

[f] Hypmorphic ts- mutant

[g] Null mutant

[h] Rescue of DR880 phenotype

[i] Rescue of DR877 phenotype

Table 2.2: DR1099 phenotype characterization

	16° (n=8)	Shifted as L1, 25° (n=7)	Shifted as L4, 25° (n=7)
Brood size	174.5±44.4	31.0±16.1	80.6±34.3
Embryonic lethality	0%, n=1396	95.4%, n=217	98.8%, n=564
Gastrulation defective	No	Yes	Yes

Worms were either maintained at 16°, or shifted from 16° to 25° at either the L1 or L4 larval stage and phenotypes of progeny were recorded. For brood size, the total number of embryos produced were counted. Embryonic lethality was recorded as the percent of unhatched embryos. To assay gastrulation, embryos were probed using rabbit anti-PGL-1, which labels the primordial germ cells, Z2/Z3, and DAPI to stain DNA as previously described (*26*). Embryos were scored as gastrulation defective if Z2/Z3 were found to remain among the external layer of embryonic nuclei (*16*).

Figures

Figure 2.1: Sequence alignments of Saccharomyces cerevisiae (y), Caenorhabditis elegans (c), Drosophila melanogaster (d), and Homo sapiens (h) AMA-1/RPB-1 amino acid sequences.

C. elegans mutations identified in this study are highlighted in green. Sequence conservation indicated by red (identical), yellow (conserved), and grey (semi-conserved) highlights. Alignment done with Clustal W (*25*).

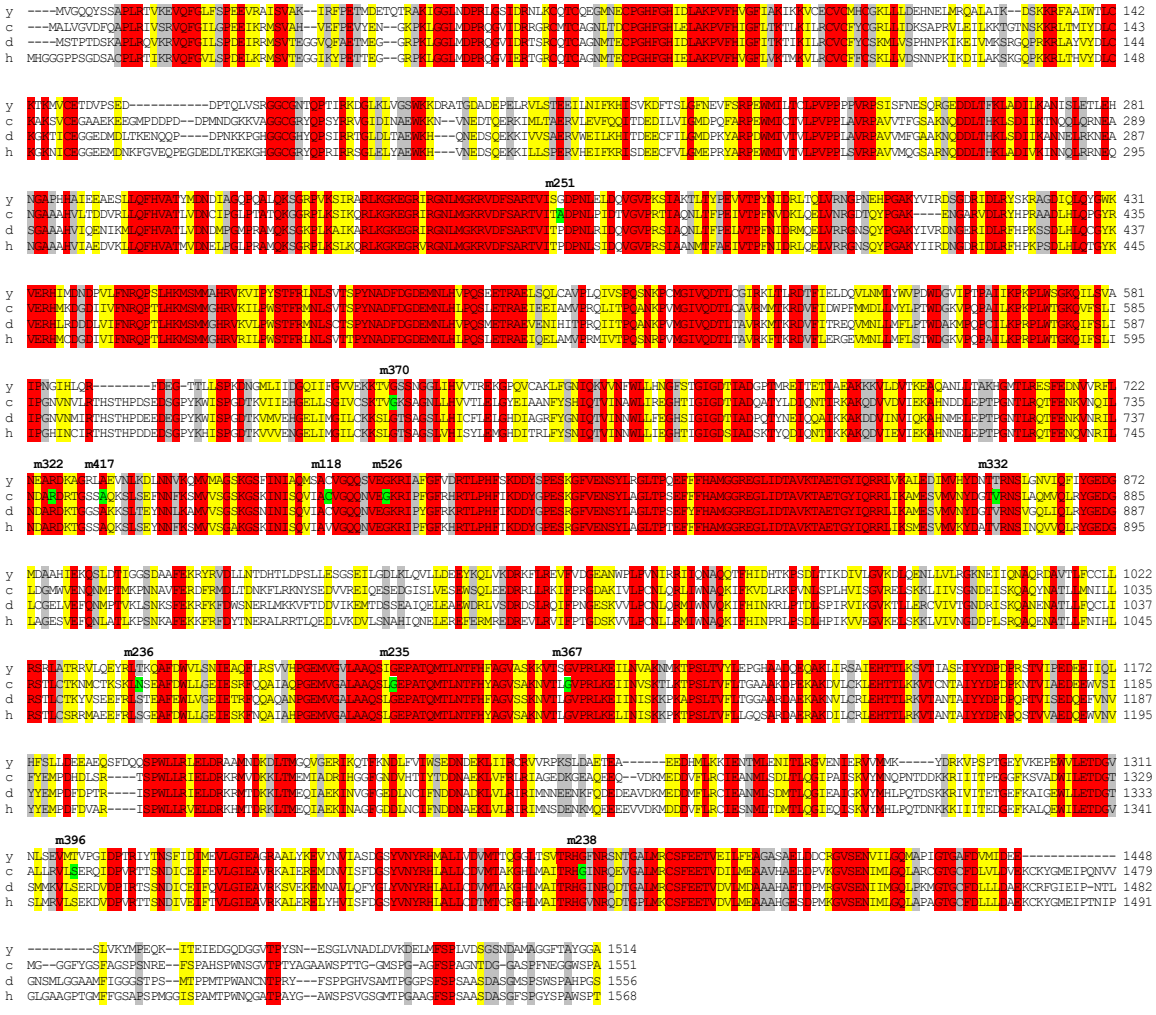


Figure 2.2: Structural position of* C. elegans *α-amanitin resistant mutations in the homologous* S. cerevisae *structure.

A) Location of α-amanitin binding site in the *S. cerevisae* Pol II structure (PDB ID: 3cqz) (*6*). α-amanitin (arrow) binds to the AMA-1/RPB1 subunit between the "funnel" and "cleft" domains (light olive and bright green domains, respectively). **B**) Stereo image of the location of amino acids corresponding to *C. elegans* α-amanitin-resistance inducing mutations. All amino acids changes are within the α-amanitin binding site and are conserved from yeast to *C. elegans*. The amino acid corresponding to m322 makes a hydrogen bond with α-amanitin. Figure made in Pymol.

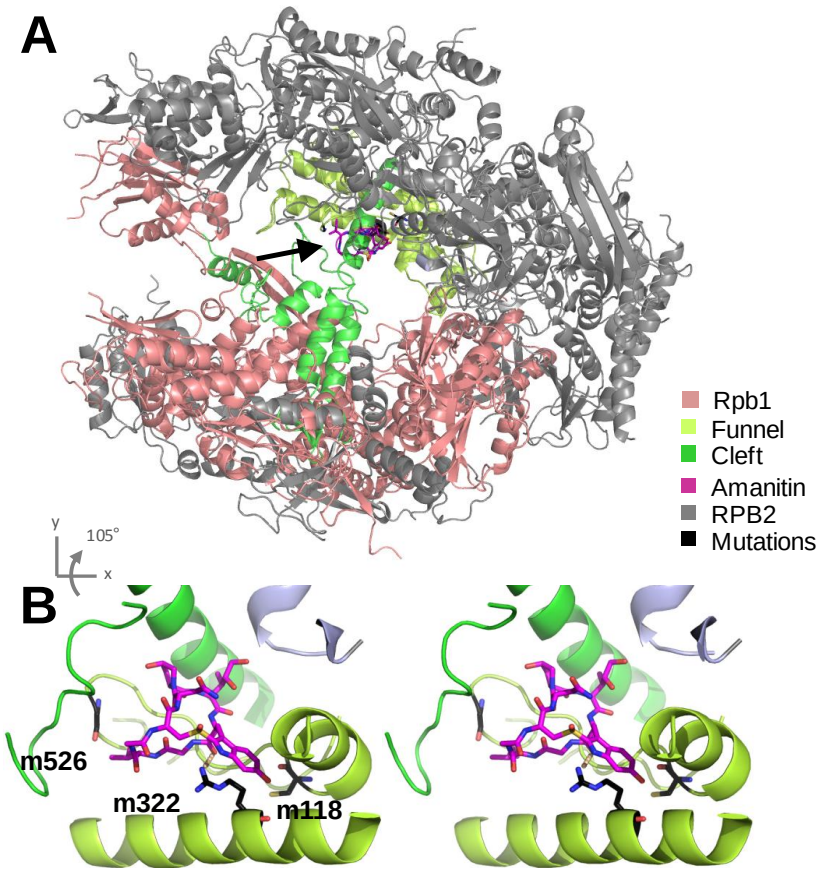
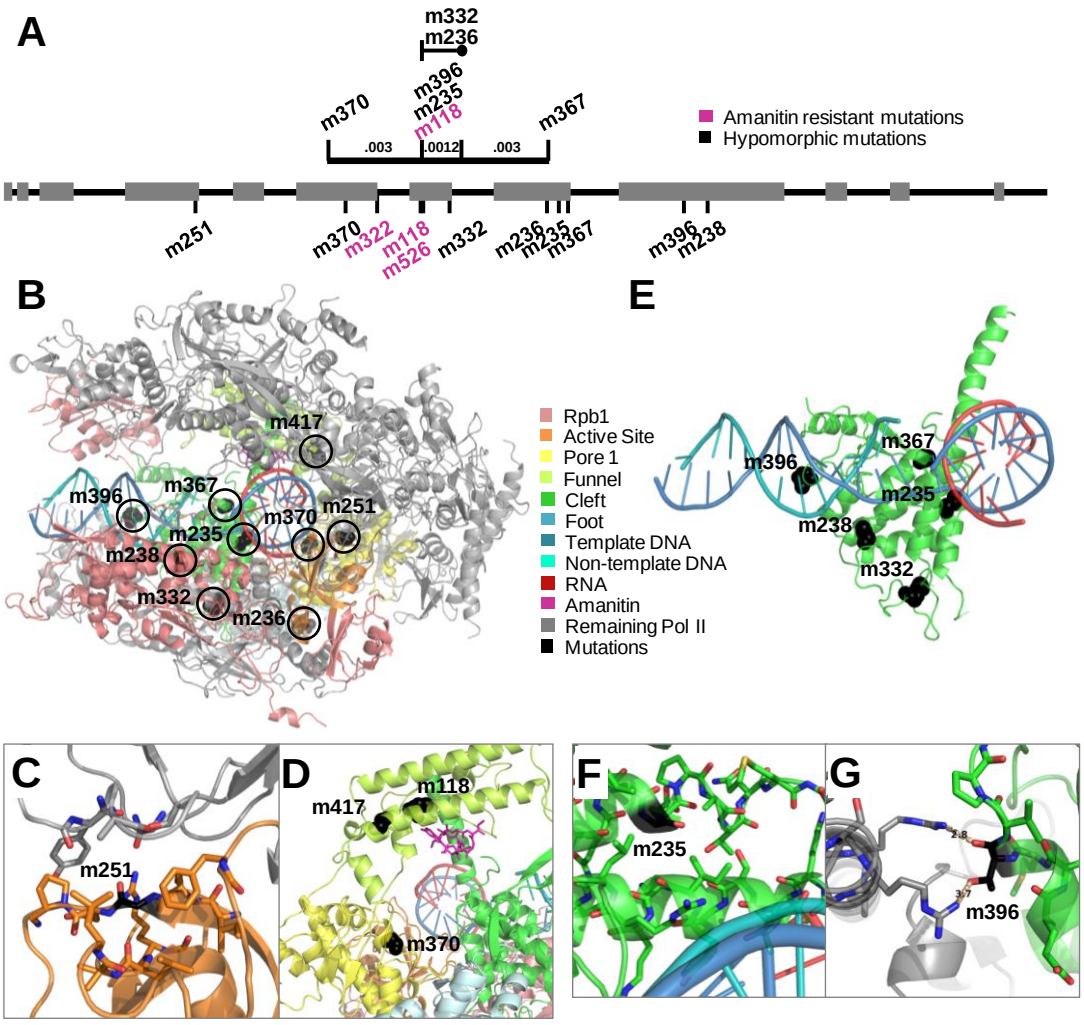


Figure 2.3: Positions of the C. elegans AMA-1/RPB1 mutations mapped to corresponding residues in the homologous S. cerevisae structure.

A) Comparison of location of mutations along the *ama-1* sequence (exons in grey boxes) with previous fine-structure genetic map position (*10*). **B**) Structural location of hypomorphic and null mutations in the Pol II structure (PDB ID: 2vum) (*27*). Domains and domain-like regions are identified according to (*5*); mutations are identified by their allele names. See Table 2.1 for allele descriptions.. **C**) Structural location of m251 in RPB1-RPB2 binding face. **D**) Structural location of mutations found in DR976: m118, α-amanitin resistance mutation; m370, null mutation; and m417, rescue mutation. Mutations m370 and m417 are approximately 27Å apart. **E**) Structural location of mutations within cleft domain and DNA binding domain. **F**) Structural location of m235 in the cleft "trigger loop" and proximity to bridge helix (bottom α-helix). **G**) Structural location of m396 and hydrogen bonds to RPB5 R11 and R14 in 3cqz structure (*6*). Bond distance is indicated. Figures made in Pymol.


A
m332
m236
m396
m235
m118
m370
m367
.003
.0012
.003
Amanitin resistant mutations
Hypomorphic mutations
m251
m370
m322
m118
m526
m332
m236
m235
m367
m396
m238
B
m417
m367
m396
m370
m251
m235
m238
m332
m236
Rpb1
Active Site
Pore 1
Funnel
Cleft
Foot
Template DNA
Non-template DNA
RNA
Amanitin
Remaining Pol II
Mutations
E
m367
m396
m235
m238
m332
C
m251
D
m417
m118
m370
F
m235
G
m396

Citations

1. T. M. Rogalski, M. Golomb, D. L. Riddle, Mutant Caenorhabditis elegans RNA polymerase II with a 20,000-fold reduced sensitivity to alpha-amanitin. *Genetics* **126**, 889 (1990).
2. T. M. Rogalski, A. M. Bullerjahn, D. L. Riddle, Lethal and amanitin-resistance mutations in the Caenorhabditis elegans ama-1 and ama-2 genes. *Genetics* **120**, 409 (1988).
3. T. M. Rogalski, D. L. Riddle, A Caenorhabditis elegans RNA polymerase II gene, ama-1 IV, and nearby essential genes. *Genetics* **118**, 61 (1988).
4. E. A. Bowman, D. L. Riddle, W. Kelly, Amino Acid Substitutions in the Caenorhabditis elegans RNA Polymerase II Large Subunit AMA-1/RPB-1 that Result in alpha-Amanitin Resistance and/or Reduced Function. *G3* **1**, 411 (2011).
5. P. Cramer, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, N.Y.)* **292**, 1863 (2001).
6. C. D. Kaplan, K. M. Larsson, R. D. Kornberg, The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Molecular cell* **30**, 547 (2008).
7. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941 (2006).
8. F. Brueckner, P. Cramer, Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol* **15**, 811 (2008).
9. D. A. Bushnell, P. Cramer, R. D. Kornberg, Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1218 (2002).
10. A. M. Bullerjahn, D. L. Riddle, Fine-structure genetics of ama-1, an essential gene encoding the amanitin-binding subunit of RNA polymerase II in Caenorhabditis elegans. *Genetics* **120**, 423 (1988).
11. Y. Chen, J. Weeks, M. A. Mortin, A. L. Greenleaf, Mapping mutations in genes encoding the two large subunits of Drosophila RNA polymerase II defines domains essential for basic transcription functions and for proper expression of developmental genes. *Molecular and cellular biology* **13**, 4214 (1993).
12. D. E. Coulter, A. L. Greenleaf, Properties of mutationally altered RNA polymerases II of Drosophila. *The Journal of biological chemistry* **257**, 1945 (1982).
13. A. L. Greenleaf, L. M. Borsett, P. F. Jiamachello, D. E. Coulter, Alpha-amanitin-resistant D. melanogaster with an altered RNA polymerase II. *Cell* **18**, 613 (1979).
14. M. S. Bartolomei, J. L. Corden, Clustered alpha-amanitin resistance mutations in mouse. *Mol Gen Genet* **246**, 778 (1995).
15. J. Archambault *et al.*, Stimulation of transcription by mutations affecting conserved regions of RNA polymerase II. *Journal of bacteriology* **180**, 2590 (1998).
16. J. A. Powell-Coffman, J. Knight, W. B. Wood, Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev Biol* **178**, 472 (1996).
17. A. Saunders, L. J. Core, J. T. Lis, Breaking barriers to transcription elongation. *Nature reviews. Molecular cell biology* **7**, 557 (2006).
18. V. Seshadri, A. G. McArthur, M. L. Sogin, R. D. Adam, Giardia lamblia RNA polymerase II: amanitin-resistant transcription. *The Journal of biological chemistry* **278**, 27804 (2003).
19. M. A. Mortin, W. J. Kim, J. Huang, Antagonistic interactions between alleles of the RpII215 locus in Drosophila melanogaster. *Genetics* **119**, 863 (1988).
20. I. Krasnoselskaya, J. Huang, T. Jones, C. Dezan, M. A. Mortin, Selection and analysis of rare second-site suppressors of Drosophila RNA polymerase II mutations. *Molecular & general genetics : MGG* **258**, 457 (1998).
21. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**, 1876 (2001).
22. M. Hampsey, J. G. Na, I. Pinto, D. E. Ware, R. W. Berroteran, Extragenic suppressors of a translation initiation defect in the cyc1 gene of Saccharomyces cerevisiae. *Biochimie* **73**, 1445 (1991).

23. R. W. Berroteran, D. E. Ware, M. Hampsey, The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. *Molecular and cellular biology* **14**, 226 (1994).
24. M. H. Suh *et al.*, A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. *The Journal of biological chemistry* **285**, 34027 (2010).
25. J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673 (1994).
26. C. E. Schaner, G. Deshpande, P. D. Schedl, W. G. Kelly, A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Dev Cell* **5**, 747 (2003).
27. F. Brueckner, P. Cramer, Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature structural & molecular biology* **15**, 811 (2008).

CHAPTER 3: TEMPERATURE SENSITIVE MUTANTS OF THE RNA POLYMERASE II TFIID INITIATION FACTOR, *TAF-6.2*

Bowman, E.A.; Kelly, W.G.[3]

To identify a conditional transcriptional mutant, I identified and characterized two previously isolated temperature sensitive embryonic lethal mutations as being alleles of *taf-6.2*, a gene encoding a conserved component of the basal transcriptional factor TFIID. Herein I describe how I mapped the mutations to a small region on chromosome IV, sequenced the candidate *taf-6.2* gene to identify the mutation, and complemented the phenotype with a wild type copy of the *taf-6.2* gene. This work has been previously published in a non-peer reviewed journal, *The Worm Breeder's Gazette* (*1*).

[3] E.A.B. wrote the manuscript and conducted all experiments. W.G.K. conceived, supervised and provided funding for the study.

Abstract

Many of the factors regulating early embryo transcription are contributed maternally; thus, transcription regulation studies in *C. elegans* have been challenging. In an attempt to identify temperature sensitive (ts) alleles that prevent transcription at the restrictive temperature, I characterized existing mutant strains that displayed temperature sensitive phenotypes that correlated with defects in embryonic transcription. In the course of these analyses, I identified two ts alleles of the TFIID initiation factor subunit, *taf-6.2*. These alleles were originally isolated by M. Wallenfang in the Seydoux lab in a screen for temperature sensitive early embryonic lethals (Emb) (*2*). The strains JH873 (*ax701*) and JH686 (*ax514*) were thought to have ts defective transcription based on the fact that they: 1) did not express an early zygotic reporter, and 2) when adults were shifted to 25°C the embryos arrested at a point similar to *ama-1(RNAi)* (*3, 4*). Indeed, the information I initially received regarding these mutations was that they were assumed to be ts mutations in *ama-1*.

I first mapped the *ax701* mutation in JH873 to chromosome IV, and positioned it between *dpy-13* and *unc-17* (8/9 DpyNonUnc; 9/53 UncNonDpy). A candidate gene in this region that matched the phenotypes observed for *ax701* was *taf-6.2* (Y37E11AL.8). Sequencing of *taf-6.2* in JH873 (*ax701*) identified 2 missense mutations (R144H, G459A) in the TAF6 domain (AA105-475) of the *taf-6.2* coding region. During the course of mapping and sequencing *ax701*, it was found that *ax514* did not complement *ax701*, suggesting that *ax514* was also located in *taf-6.2*. Sequencing of *taf-6.2* in JH686 also identified a missense mutation (G421E), which was also in the TAF6 domain of *taf-6.2*. These mutations were not found in either an N2 strain or the parent strain mutagenized in the original screen, JH150. In addition, no mutations were found in non-coding regions of the *taf-6.2* ORF of either strain. The strains were outcrossed six times each, and the lesions I identified tracked with the ts Emb phenotype in the outcross.

Finally, the Emb phenotypes can be rescued with a WT copy of *taf-6.2*. A construct containing the *taf-6.2* ORF plus 200 bp upstream and downstream of the ORF (pBAB1) was injected into the outcrossed *ax701* strain with PD118.20 (myo-3:GFP). Nine independent GFP^+ rescued lines that could grow at the restrictive temperature (25°C) were obtained. The *taf-6.2* transgene also rescued JH686 (*ax514*) to viability at 25°C.

In addition to the phenotypes mentioned above, further lines of evidence suggest that *ax701* prevents transcription at the restrictive temperature. Embryos raised at the restrictive temperature failed to gastrulate, as is observed in *ama-1(RNAi)* (*4*). In addition, shifted *ax701 unc-17* embryos showed decreases in the transcriptional elongation marker, phosphorylation of Ser2 on the CTD repeat peptide (H5 antibody, data not shown).

In summary, I have identified two temperature sensitive alleles of the transcription initiation factor, *taf-6.2*. While I have confirmed that mutations in *taf-6.2* are responsible for the embryonic arrest phenotypes that suggest that RNA Polymerase II transcription is shut down or severely compromised in these strains, I have not yet verified the transcription defect using high resolution or biochemical assays. Overall, I think these alleles will be important to the worm community as a way to modulate transcription in a time- and temperature-sensitive manner. For this reason I published a brief description of the identification and characterization of these mutations in the *Worm Breeders Gazette*, a non-peer reviewed publication for the *C. elegans* research community that provides otherwise unpublished data and techniques for researchers(*1*). I have made the *unc-17 taf-6.2* outcrossed strains available to the community through the CGC [KW1973: *taf-6.2(ax701) unc-17(e113)* IV and KW1975: *taf-6.2(ax514) unc-17(e113)* IV].

Citations

1. E. A. Bowman, G. Seydoux, W. G. Kelly, Temperature sensitive mutants of the RNA Polymerase II TFIID initiation factor, taf-6.2. . *Worm Breeder's Gazette* **18**, (2011).
2. A. Golden *et al.*, Metaphase to anaphase (mat) transition-defective mutants in Caenorhabditis elegans. *The Journal of cell biology* **151**, 1469 (2000).
3. M. R. Wallenfang, G. Seydoux, cdk-7 Is required for mRNA transcription and cell cycle progression in Caenorhabditis elegans embryos. *Proc Natl Acad Sci U S A* **99**, 5527 (2002).
4. J. A. Powell-Coffman, J. Knight, W. B. Wood, Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev Biol* **178**, 472 (1996).

CHAPTER 4: P-TEFB-INDEPENDENT RNA POLYMERASE II PHOSPHORYLATION AND GERMLINE-SOMA DISTINCTION IN C. ELEGANS

Bowman, E.A.; Bowman, C.R.; Kelly, W.G.[4]

Following the recent description of *Drosophila* and mammalian CDK12 as a kinase capable of phosphorylating Ser2 of the Pol II CTD, I characterized the two kinases responsible for Ser2 phosphorylation (Ser2-P) in *C. elegans*, CDK-9 and CDK-12. Previous studies in our laboratory identified a CDK-9 independent form of Ser2-P in *C. elegans* embryonic germ cells (*1*). I have expanded that work by confirming that embryonic germ cell Ser2-P is independent of both P-TEFb components, CDK-9 and Cyclin T. I also identified the kinase complex responsible for embryonic germ cell Ser2-P as CDK-12 and Cyclin K. Finally, I showed that CDK-12 is the kinase responsible for Ser2-P throughout germline development. This is the first demonstration of tissue-specific regulation of Ser2-P and transcriptional elongation. This work has been submitted for publication and is currently under revision (*2*).

[4] E.A.B. conceived the study, wrote the manuscript and conducted most experiments. C.R.B. provided technical assistance for quantification of immunofluorescence data and other experiments. W.G.K. supervised and provided funding for the study.

Abstract

RNA Polymerase II (Pol II) elongation in metazoans is thought to require phosphorylation of Serine 2 (Ser2) of the Pol II C-terminal domain (CTD) by P-TEFb (CDK-9-Cyclin T). I show that while Ser2 phosphorylation (Ser2-P) in *C. elegans* embryonic somatic tissues agrees with this model, Ser2-P in germ cells is independent of P-TEFb and instead requires the recently characterized CDK-12-Cyclin K complex. Surprisingly, loss of CDK-12 and detectable Ser2-P from germ cells has little impact on fertility. In contrast, loss of CDK-9 from the germ line has little effect on Ser2P but results in sterility. These results uncouple Ser2-P from germline development and suggest that Ser2-P regulation is not the essential role of CDK-9 in these cells. I further show that CDK-12's role in the germline includes co-transcriptional H3K36me3 regulation. Overall, transcriptional elongation in *C. elegans* appears to be uniquely regulated in the germline which may comprise a novel facet of germline identity.

Introduction

The large subunit of RNA Polymerase II (Pol II) contains a C terminal domain (CTD) that includes many repeats of the heptapeptide sequence with consensus $Y_1S_2P_3T_4S_5P_6S_7$ (42 in *C. elegans*). The phosphorylation states of specific residues in the repeats are thought to be important for proper mRNA synthesis. Specifically, Pol II recruited to genes is hypophosphorylated on the CTD. Serine 5 (Ser5) becomes phosphorylated during transcription initiation and serine 2 (Ser2) is phosphorylated on elongating Pol II (*3, 4*). In addition, dephosphorylation of both Ser2 and Ser5 is important for recycling of Pol II complexes for reinitiation (*5*). The different phosphorylation marks help recruit various Pol II transcription factors, RNA processing machinery, and histone/nucleosome modifiers at the proper time during the transcription cycle (*3, 6*).

The transition between Pol II initiation and elongation is a regulated process that includes positive and negative regulatory factors that interact with Pol II. Pol II elongation is inhibited by two complexes, DSIF (DRB-sensitivity inducing factor, composed of Spt4 and Spt5) and NELF (Negative Elongation Factor). Phosphorylation of these factors releases their inhibition of Pol II elongation and this appears to be tightly linked to Pol II CTD Ser2 phosphorylation (*7*). Whereas DSIF is highly conserved among eukaryotes, NELF is absent in many eukaryotes including *S. cerevisiae* or *C. elegans* (*8*).

In budding yeast Ser2 and DSIF phosphorylation requires two kinases, Bur1 and Ctk1. Bur1 can phosphorylate both DSIF and Ser2, but Ctk1 is thought to carry out the majority of Ser2 phosphorylation (Ser2-P) (*9*). In metazoans, the positive transcription elongation factor b complex, or P-TEFb, composed of CDK9 and Cyclin T, has historically been considered the sole DSIF-NELF and Pol II Ser2 kinase in metazoans, because loss of CDK9 leads to complete loss of Ser2-P (*10*). Recently a homolog of yeast Ctk1, CDK12, was shown to be required for a large fraction of Ser2-P in *Drosophila* and humans (*11*). It was thus suggested that Ser2-P in metazoans may more closely match the elongation process in budding yeast: CDK9 phosphorylates the Pol II CTD and DSIF (and NELF when present), and these events precede CDK12-mediated phosphorylation of Ser2 on elongating Pol II (*12*). This model could explain the requirement for CDK9 for all Ser2 observed in these systems.

Here I show that Ser2-P is differentially regulated in the somatic and germ cell lineages in *C. elegans*. Ser2-P regulation in the soma requires CDK-9 acting upstream of CDK-12 for Ser2-P; however Ser2-P in the germ line is independent of CDK-9 and instead only requires CDK-12. Surprisingly, CDK-12-dependent Ser2-P is not required for germline development. These results show that Ser2-P is not an obligate component of Pol II-mediated gene expression in all tissues, and they further suggest that differences in basic modes of transcriptional regulation may be important features of soma and germ line distinction.

Materials and Methods

Strains Used

The strains used were: wild-type N2 (Bristol); KW2090 *cdk-9(tm2884) I/hT2* qIS48; KW2088 *cdk-12(tm3846) IV/qC1* qIS26; KW2212 *cdk-12(ok3664) IV/qC1* qIS26; VC456 *cit-1.2(gk241)* III; SS803 *mes-4(bn85) V/DnT1-GFP*; JH1382 *dpy-18(e364) pie-1(zu127) III/qC1*; and other transgenic lines created for this study (Table 4.2, 4.3). Some strains were provided by the CGC (Minneapolis, MN, USA), funded by NIH Office of Research Infrastructure Programs (P40 OD010440), and the National Bioresource Project for the Experimental Animal "Nematode *C. elegans*" (Kawada-cho, Shinjuku-ku, Tokyo).

RNA interference

RNAi knock down was performed by soaking and feeding as previously described (*13*). RNAi library clones were verified by sequencing and used for all experiments (*14*). RNAi was considered efficient if 100% embryonic lethality (for *cdk-9*, *cyclin T*, *or fcp-1*) or larval arrest (for *cdk-12* and *ccnk-1*) was observed.

Immunofluorescence

Immunofluorescence was performed using methanol/formaldehyde fixation (*15*) with the following antibodies: rat anti-Ser2-P (3E10, 1/10 dilution), rat anti-Ser5-P (3E8, 1/10 dilution), rat anti-Ser7-P (4E12, 1/10 dilution) (*16*), goat anti-PIE-1 (cN-19, 1/100 dilution, Santa Cruz Biotech, Dallas, TX, USA #sc9245), rabbit anti-PGL-1 [1/10,000 dilution (*17*)], and mouse anti-H3K36me3 [CMA333, 0.25 μg/ml (*13*)]. Methanol/acetone fixation (*18*) was used for anti-GFP (Milipore MAB3580, 1/200 dilution), anti-AMA-1 (AMA-1 N terminal, 1/10,000 dilution; Novus Biological, Littleton, CO, USA, #38520002), anti-H3K4me2 [CMA303, 1/20 dilution (*19*)], and anti-P-granules [OIC1D4, 1/4 dilution (*20*)]. Fixed animals were incubated with primary

antibodies 12-16 hours at 4°C. Secondary antibodies included Alexa Fluor 594-and Alexa Fluor 488-conjugated donkey antibodies (1:500) (Molecular Probes, Eugene, OR, USA).

Images of immunofluorescence data were taken on a Leica DMRXA (Hamamatsu Photonics, Hamamatsu, Japan) microscope using Simple PCI software (Hamamatsu Photonics). Fiji (*21*) was used for quantification of raw immunofluorescence data. Either the two Z2/Z3 nuclei or 3 representative nuclei were identified by DAPI staining and average signal intensity was measured. Background signal was identified as non-nuclear staining and subtracted from the nuclear staining. Standard error of measurements was determined according to (*22*). Embryos at ~100-300 cells were analyzed. For embryo staging, the number of nuclei in an embryo was counted using Fiji (*21*). Non-overlapping Z-stacks were taken of each embryo. The DAPI channel of each layer transformed into a binary image, the Watershed protocol was applied to separate individual nuclei, and the number of particles (nuclei, 200-2000 pixel units) was counted.

Quantification of Ser2-P was performed using the 3E10 antibody rather than H5 antibody because the H5 antibody recognizes an AMA-1-independent signal in mitotic nuclei that interferes with quantitative approaches (*23*).

Protein Isolation and Western

75 adult worms were transferred to 5 µL 1X PBS in a 1.5 mL microfuge tube. 75 µL lysis buffer (100 mM Tris HCl pH 8.0, 3 mM $MgCl_2$, 300 uM KCl, 0.1% NP40, 20% glycercol, 40 mM DTT) was added before snap freezing in liquid nitrogen. Samples were thawed at room temperature, and refrozen in liquid nitrogen twice before sonication for 10 minutes (Diagenode Bioruptor, high setting) in a 4°C water bath. Samples were centrifuged at 4°C for 1 minute at 13K rpm to remove debris and supernatant removed to a new tube. Protein concentration was analyzed by Bradford (Bio-Rad Laboratoires, Hercules, CA, USA).

Equal amounts of protein were run on a 4-20% SDS-PAGE gel (Bio-Rad Laboratoires) and transferred for Western analyses. Membranes were probed with anti-Ser2-P (H5, 1/5,000,

Covance, Princeton, NJ, USA), anti-AMA-1 N-terminal antibody (1/10,000), or anti-actin (1/10,000, Milipore, Billerica, MA, USA, MAB1501) and detected with HRP-conjugated secondaries using SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology, Rockford, IL, USA).

Transgenic Lines

Fluorescence-tagged constructs were generated by conventional cloning methods (Table 4.4, 4.5). Construct promoters, ORFs, and 3' UTRs were amplified from *C. elegans* N2 genomic DNA and construct tags were amplified from the indicated plasmids (Table 4.5) using Phusion Taq (Finnzymes, Waltham, MA, USA). Transgene components were cloned into pCR-Blunt II-TOPO Zero Blunt Cloning Plasmid (Invitrogen, Grand Island, NY, USA). Finished constructs were subcloned into MOS-SCI plasmids (*cit-1.1*, *cit-1.2*, *ccnk-1*, *spt-5* in to pCFJ151; *cdk-9* into pCFJ356; *cdk-12* into pCFJ352, from AddGene, Cambridge, MA, USA). Kinase-dead point mutants were created using QuikChange XL (Aglient, Santa Clara, CA, USA).

Single copy insertion transgenic lines were made by standard MOS-SCI protocol (*24*). Insertion of the construct into the respective MOS transposon site was confirmed by PCR and at least two independent expressing lines were assayed for expression patterns. Representative transgenic lines used for this study are indicated (Table 4.2).

Fluorescence and DNA visualization in intact worms

5-10 adult worms were picked and quickly moved between the following solutions in a multi-well depression slide: ice cold 80% MeOH 45 seconds; room temperature 2 ug/mL 4',6-diamidino-2-phenylindole (DAPI) 30 seconds; room temperature dH_2O 2-10 minutes. Worms were mounted under a glass cover slip in 7 µL 5% glycerol on a slide for visualization using Leica DMRXA (Hamamatsu Photonics, Hamamatsu, Japan) microscope and Simple PCI software

(Hamamatsu Photonics). This maintained GFP and mCherry signal and allowed DAPI to stain all nuclei except in embryos.

Results

Tissue-specific requirements for CDK-9 and CDK-12 in Ser2 phosphorylation

Previous studies have indicated that CDK-9 activity is required for all detectable Ser2-P in the *C. elegans* embryo (*25*), but Ser2-P appearance is independent of CDK-9 in the two primordial germ cells (PGCs), Z2 and Z3 (*13*). Although our initial studies implicated the Tousled-like kinase, TLK-1, as a Ser2 kinase in the germ cells, the cell-cycle defects in the *tlk-1* depleted embryos made it difficult to conclude that this was a direct effect (*13*). To further investigate embryonic Ser2 kinase requirements, I depleted other components of the CDK-9-Cyclin T complex as well as homologs of the newly described CDK-12-Cyclin K Ser2 kinase complex.

As expected, *cdk-9(RNAi)* resulted in loss of detectable Ser2-P in all embryonic somatic cells, but significant levels of Ser2-P remained unaffected by this treatment in Z2/Z3 (Figure 4.1A,B, 4.2A). RNAi of the two Cyclin T paralogs yielded identical results to *cdk-9(RNAi)* (Figure 4.1A, 4.2E). Knock down of either CDK-9 or Cyclin T caused defective gastrulation and early embryonic arrest at ~120 cells (Figure 4.3). These phenotypes correlate with defective somatic zygotic transcription and occur with depletion of the large subunit of Pol II, AMA-1 (*26*). Thus, while CDK-9 and Cyclin T appear to be important for both zygotic transcription and Ser2-P in somatic cells, neither are required for Ser2-P in Z2/Z3.

CDK12 is responsible for significant Ser2-P in both *Drosophila* and human cells (*11, 27, 28*). An uncharacterized gene, *B0285.1*, is the closest homolog to *Drosophila* CDK12 in *C. elegans* (previously *cdtl-7*, hereafter named *cdk-12*). In contrast to *cdk-9(RNAi)*, *cdk-12(RNAi)* resulted in undetectable Ser2-P levels in the PGCs, and ~60% reduction of Ser2-P in embryonic somatic nuclei (Figure 4.1A,B, 4.2A). The RNAi conditions used were specific for each individual gene,

as knock down of either kinase did not affect expression of a transgene expressing the other (Figure 4.4).

Cyclin K is the cyclin partner of CDK12 in other systems (*11, 27, 28*). I performed RNAi against several homologs of *Drosophila* Cyclin K, and found that knock down of the closest sequence homolog, *F43D2.1* (hereafter named *ccnk-1*), phenocopies *cdk-12(RNAi)*, and is likely the cyclin partner of CDK-12 (Figure 4.1A,B, 4.2F). Knock down of either CDK-12 or Cyclin K results in early larval arrest, but no specific developmental defects were observed other than what appeared to be arrested growth (Figure 4.3).

The above results demonstrate that CDK-12, rather than CDK-9, is the predominant Ser2 kinase in embryonic germ cells. The Ser2-P antibody used in this study [3E10 antibody (*16*)] is specific for Pol II in *C. elegans* since *ama-1(RNAi)* depletes all antibody binding (*13*). Another antibody against the Ser2-P epitope shows similar patterns as the 3E10 antibody in RNAi experiments (H5 antibody; data not shown). CDK-12 appears to be specific for Ser2 in the embryo, since another phosphoepitope, Ser5-P, is unaffected by *cdk-12(RNAi)*. I did observe decreases in Ser5-P with *cdk-9(RNAi)* (Figure 4.2B,D), however this may be a direct or indirect effect as CDK-9 is reported to phosphorylate Ser5 in human cells (*29*).

I also considered that decreased Ser2-P after *cdk-12* knockdown could be an indirect result of increased Pol II CTD phosphatase activity. Knock down of the best characterized Pol II CTD Ser2 phosphatase, FCP-1, shows increased Ser2-P in *C. elegans* oocytes (*30*). *fcp-1(RNAi)* also results in increases in Ser2-P levels in embryonic somatic and germline nuclei (Figure 4.5A,B, 4.2G). RNAi against *fcp-1* in combination with *cdk-12* caused a Ser2-P reduction in Z2/Z3 similar to *cdk-12(RNAi)* alone, showing that the reduction of Ser2-P in these cells is not due to an ectopic increase in FCP-1 activity (Figure 4.5A,B, 4.2G).

Embryonic Germ line-Soma conversion yields switch in kinase dependency

A maternal protein called PIE-1inhibits Ser2 phosphorylation and Pol II transcription in the *C. elegans* embryonic germline precursor cells, called P-cells (*31*). PIE-1 has been proposed to inhibit Ser2 phosphorylation by sequestering P-TEFb until PIE-1 is degraded at the birth of Z2/Z3; indeed, degradation of PIE-1 in Z2/Z3 correlates with Ser2-P appearance in these cells (Zhang et al., 2003). However, our assays above suggested that the germline Ser2-P kinase inhibited by PIE-1 is likely to be CDK-12-Cyclin K. Indeed delaying PIE-1 degradation in Z2/Z3 by knocking down ZIF-1, a SOCS-box protein involved in PIE-1 degradation, caused a delay in appearance of CDK-12-dependent Ser2-P Z2/Z3 (Figure 4.6A,B,C). The loss of PIE-1 in the P-cells causes a transformation of germ line into soma, so I next tested whether the ectopic Ser2-P that appears in P-cells in *pie-1(zu127)* mutants is CDK-12 dependent. Surprisingly, the Ser2-P in the *pie-1(zu127)* transformed P-cells and their descendants became sensitive to *ckd-9(RNAi)* (Figure 4.6D), indicating that transformation from germ line to soma is accompanied by a switch from CDK-12 to CDK-9 in kinase requirements for Ser2-P regulation. This further illustrates the tight correlation between germline identity and CDK-9-independent Ser2-P regulation.

CDK-12—dependent Ser2-P is not essential for early germ cell proliferation

I next examined the requirement for CDK-12 in post-embryonic germ cells. Z2/Z3 are born in the embryo after about 2 hours of development (~100 cells), and remain mitotically quiescent through the remaining 10 hours of embryogenesis and re-enter the cell cycle after hatching (*32, 33*). Embryos with CDK-12 knock down complete embryogenesis but arrest shortly after hatching, so I performed immunofluorescence on newly-hatched L1 larval offspring of RNAi-treated animals. *cdk-12(RNAi)* larvae had decreased Ser2-P in somatic nuclei, but lost detectable Ser2-P in the dividing germline cells (Figure 4.7A,B). Thus, CDK-12 is also required for Ser2-P in post-embryonic germ cells. The somatic Ser2-P observed in *cdk-1(RNAi)* larvae is presumably

due to CDK-9 activity, although this could not be assessed because of early embryonic lethality of *cdk-9(RNAi)* animals.

Importantly, the *cdk-12(RNAi)* larvae examined in this experiment arise from embryos with no history of detectable Ser2-P in their germ cells. Yet despite Ser2-P absence, the L1 germ cells initially proliferate after the embryos hatch, but arrest with the rest of the soma in the larvae (Figure 4.3). The larval arrest appears to be related to a lack of growth, since the animals move and feed normally but eventually die.

CDK-12 is the predominant germline Ser2 kinase

I next examined the requirements for CDK-9 and CDK-12 in adult germ cells. The adult gonad is a syncytium of germ cell nuclei that are produced from a pool of proliferating stem cells at the distal end of the gonad arm, and these cells progress sequentially through meiosis and gametogenesis as they migrate toward the proximal end. Thus, the adult gonad contains germ cells linearly arranged in discrete, sequential stages of development that are highly transcriptionally active. I individually targeted each kinase by RNAi and used CDK-9:mCherry and CDK-12:GFP transgenes to assess RNAi knockdown. By initiating RNAi treatment at the last larval (L4) stage, I obtained adults with loss of detectable kinase protein in younger cells in the distal end of the gonad, and near complete loss in older nuclei at the proximal end (Figure 4.8A).

Surprisingly, although *cdk-9(RNAi*) causes loss of CDK-9:mCherry below detection, this had little effect on Ser2-P (Figure 4.8B, 4.9G). In contrast, *cdk-12(RNAi)* causes near complete loss of Ser2-P in all germ cells (Figure 4.8B). Some Ser2-P remains detectable in older proximal nuclei in *cdk-12(RNAi)* ovaries, presumably due to CDK-12 protein produced before RNAi treatment was maximal. The dependence of Ser2-P on CDK-12, with little requirement for CDK-9, is not specific to female germ cells as I observed similar results with RNAi in male gonads (data not shown). The significant decrease of Ser2-P in *cdk-12(RNAi)* germ cells, which represents roughly half of the total nuclei in an adult worm, is also readily apparent by western blot analyses (Figure

4.9A). Thus, Ser2-P in germ cells at all stages of the germ cell cycle is largely independent of CDK-9 and dependent on CDK-12.

Neither *cdk-9(RNAi)* nor *cdk-12(RNAi)* significantly affect Ser5-P levels in adult germ cells (Figure 4.9C). In contrast, *cdk-12(RNAi)* results in a significant reduction of Ser7-P levels (Figure 4.9D), but it is unclear if this is a direct effect.

CDK-9, CDK-12, and cyclin partners are ubiquitously expressed

Tissue-specific differences in Ser2 kinase requirements could be due to differences in expression of the kinases and/or their cyclin partners. In the absence of available antibodies for these components in *C. elegans*, I analyzed their expression using low copy, integrated fluorescently-tagged transgenes under control of their endogenous promoters and 3'UTRs (*24*).

All components tested were ubiquitously expressed (Figure 4.10A). These expression patterns were observed for at least two independent transgenic lines of each. The expression of the CDK-9, CDK-12, and Cyclin T (CIT-1.2) transgenes likely recapitulate endogenous protein localization because the transgenes rescue their respective mutant strains (a Cyclin K mutation is unavailable; Table 4.1). Furthermore, the kinase activity of both CDK-9 and CDK-12 is essential, as transgenes expressing kinase-inactivated point mutants did not rescue mutant strains (data not shown).

I further analyzed the subnuclear distribution of the fusion proteins in adult germ cells. The X chromosome is transcriptionally suppressed in meiotic germ cells in *C. elegans* (*34*). Both CDK-9 and CDK-12 and their component cyclins, are detectably associated with autosomal chromatin but excluded from the X chromosome until late oogenesis, when X-linked transcription is activated (Figure 4.10B, Cyclin T/K data not shown). Thus, while only CDK-12 activity is required for Ser2-P in the germ line, both the CDK-9-Cyclin T and CDK-12-Cyclin K complexes are distributed in patterns that correlate with chromatin undergoing robust transcription.

Both Cyclin T paralogs are ubiquitously expressed; however, the transgenic expression of each appeared reduced in the adult gonad and embryonic germ cells relative to somatic expression (Figure 4.10A, 4.11A,B). To test if this reduced expression correlates with a decreased requirement in germ cells for CDK-9-Cyclin T-dependent Ser2-P, I increased Cyclin T expression in the germ line. Replacing the *cit-1.2 3'UTR* with the *ccnk-1 3'UTR* in the CIT-1.2 construct results in substantially increased expression of CIT-1.2 in both embryonic and adult germ cells, but this did not change the Ser2 kinase requirements (Figure 4.11). The reduced requirement for CDK-9 in germline Ser2-P does not appear to be due to reduced Cyclin T expression in germ cells.

Tissue-specific differences in Ser2 kinase requirements could also be due to the presence or absence of other factors involved in transcription elongation. I examined the expression of SPT-5, the major functional component of DSIF in *C. elegans* and other systems (*25*, *35*, *36*). As observed for the CTD kinases, SPT-5:GFP is ubiquitously expressed and excluded from the X chromosome in adult meiotic germ cells (Figure 4.12, data not shown). Thus, tissue-specific differences in Pol II Ser2 phosphorylation are not likely due to differential expression of the kinases, cyclins, or the DSIF complex.

CDK-9, but not CDK-12, is required for proper germline development

Because of the different requirements for CDK-9 and CDK-12 in germline Ser2-P, I wanted to determine if either of these kinases is essential for germline development. Because ubiquitous loss of CDK-9 or CDK-12 function causes early developmental arrest, I developed a method to remove protein activity specifically from the germ line. I replaced the endogenous 3'UTR with the 3'UTR of the *pal-1* gene which prevents translation in the germ line until late oogenesis in the adult (*37*). CDK-9 and CDK-12 transgenes with the *pal-1 3' UTR* show ubiquitous somatic expression, but no expression in post-embryonic germ cells until late in adult oogenesis (Figure

4.13A). Expressing the *kinase:pal-1 3'UTR* transgenes in the respective mutants therefore allows rescue of kinase function in all somatic lineages, but no rescue in post-embryonic germ cells.

I created balanced mutant strains expressing the *kinase::pal-1 3'UTR* transgenes and analyzed their homozygous mutant offspring. Both *kinase:pal-1 3'UTR* transgenes fully rescue somatic development as the transgenic homozygotes grow to adulthood in parallel with their heterozygous siblings and show no obvious somatic phenotypes. In contrast, the absence of either CDK-9 or CDK-12 kinase activities in germ cells has quite different consequences.

The absence of detectable CDK-9 expression in germ cells causes dramatic sterility, with animals producing only ~50-100 germ cells per gonad (~1/10 of the normal size, Figure 4.13B). The surviving germ cells exhibited robust Ser2-P in their chromatin, further supporting the conclusion that Ser2-P is independent of CDK-9 in germ cells (Figure 4.13C). No CDK-9:mCherry was detected in these germ cells or those in earlier stages, which also suggests that germ cell proliferation can still occur in the absence of CDK-9 activity. While persistence of trace amounts of maternally derived CDK-9 mRNA or protein cannot be ruled out, it would have to have been at very low levels distributed among a significant number of nuclei.

Surprisingly, and in contrast to the CDK-9 results, the *cdk-12(ok3664)* animals expressing the *cdk-12::pal-1 3'UTR* transgene show few if any germline defects. Although neither CDK-12:GFP nor Ser2-P was detected at any post-embryonic germ cell stage until late oogenesis (Figure 4.13D,E), these animals produce large numbers of functional germ cells and are fertile (Figure 4.13D; note embryos in these animals). Therefore neither CDK-12 nor Ser2-P phosphorylation are essential for normal germ cell proliferation and development at normal temperatures. Interestingly, 100% of homozygotes expressing the *cdk-12:pal-1 3' UTR* transgene grow up sterile when raised at an elevated temperature (25°C; data not shown).

CDK-12 is required for normal H3K36me3levels in germ cells.

Ser2-P is proposed to play multiple roles in transcription regulation, including the recruitment of RNA processing factors and histone modifiers. Previous studies in yeast have shown that Ser2-P regulates histone H3 lysine 36 trimethylation (H3K36me3) levels by recruiting the K36 methyltransferase, Set2, during Pol II transcription elongation (*38*). *C. elegans* has two H3K36 methyltransferases: MET-1, a transcription-dependent K36 HMT and MES-4, an H3K36 methyltransferase that can function independently of transcription in both embryos and adults (*13, 39, 40*). I tested if transcription-dependent H3K36me3 in the germ line is dependent on CDK-12.

Previous immunofluorescence studies have shown that detectable H3K36 dimethylation (H3K36me2) in the adult germ line requires MES-4 (*39*). In contrast, I found that *mes-4(RNAi)* or *mes-4(bn85)* mutants still exhibit substantial levels of H3K36 trimethylation (H3K36me3) in adult germ cells (Figure 4.14A,C, 4.15A,C). This suggests that a large fraction of this modification is due to transcription-dependent activities of MET-1, which allowed us to assess the requirement for CDK-12 in transcription dependent H3K36me3.

I targeted *mes-4* by RNAi in the *cdk-12::pal-1 3'UTR* transgenic *cdk-12(ok3664)* mutants and examined H3K36me3 levels in germ cell chromatin. Decreased Ser2-P in germ cells correlated with a significant decrease in detectable H3K36me3 (Figure 4.14A,B—compare white/grey bars with pink/red bars). Both total H3K36me3 levels (compare white and pink bars) and MES-4—independent H3K36me3 levels showed significant decreases (Figure 4.14A,C—compare grey and red bars). Similar results were obtained with *cdk-12(RNAi)* in *mes-4(bn85)* mutants (Figure 4.15A,C). H3K36me3 in the *C. elegans* germ line thus requires Ser2-P produced by CDK-12.

I also observed decreases in H3K36me3 levels after *cdk-9(RNAi)* in *mes-4(bn85)* mutant germ cells (Figure 4.15A,C). This was surprising, since loss of CDK-9 has little effect on Ser2-P in germ cells. Given the more severe germ cell phenotypes caused by decreased CDK-9, this may

be an indirect effect on H3K36me3 levels; although a similar decrease in H3K36me3 levels has been observed in yeast Bur1 mutants (*41*).

Discussion

Embryonic Ser2-P regulation highlights tissue-specific regulation

Although phosphorylation of Ser2 of the Pol II CTD has been studied for over 20 years, the role of CDK12 in this process has only recently been appreciated (*11*). This is understandable considering that the loss of P-TEFb (CDK-9-Cyclin T) activity in most analyses caused loss of all Ser2-P, thereby obviating an apparent requirement for additional kinases. The more severe effect on Ser2-P observed with loss of P-TEFb has been reinterpreted as the P-TEFb complex playing a role upstream of CDK12 in Ser2 phosphorylation (*12*). Embryonic somatic tissues in *C. elegans* are consistent with this model as knock down of either P-TEFb component results in complete loss of Ser2-P. In contrast CDK-12-Cyclin K knock down only results in a 60% decrease in Ser2-P, a decrease remarkably consistent with that observed in other systems (*11*, *27*).

The regulation of Ser2-P in somatic lineages also agrees with current models during the transcription cycle. CDK-9-Cyclin T and FCP-1 knock down causes defective gastrulation, a phenotype associated with defective zygotic transcription in *C. elegans* (*26*). While this phenotype could be due to other functions of these proteins, it seems most likely that these effects are due to defective Pol II transcription, i.e., CDK-9 knock down prevents proper Pol II elongation whereas FCP-1 prevents proper Pol II recycling, in agreement with the current models (*5*, *42*).

The novelty of the current study is the discovery that transcription regulation in the *C. elegans* germ line dramatically differs with the current model of metazoan transcription elongation. Not only does CDK-9 not play the major role in phosphorylation of Ser2, but phosphorylation of Ser2 by CDK-12 does not require P-TEFb activity in this lineage. This is the first demonstration of Pol

II Ser2-P not requiring, directly or indirectly, CDK-9/P-TEFb activity in a multicellular organism. Furthermore, the *pie-1* mutant results show that this is a basic difference between somatic and germ line Pol II regulation, as transformation of germ line to soma is accompanied by a switch from CDK-12- to CDK-9-dependent regulation of Ser2-P.

Novel regulation of Pol II is a common feature of germ cell specification in many organisms (*43*), yet I was initially surprised to identify P-TEFb-independent Ser2-P in the *C. elegans* PGCs, Z2/Z3. Previous studies have suggested that P-TEFb inhibition is important for transcriptional repression of the germline precursors that give rise to the PGCs (*31, 44*). The germline precursor cells, named P cells, are transcriptionally repressed by a general transcriptional repressor, PIE-1 (*31*). These studies suggested that after PIE-1 degradation at the birth of the PGCs, Z2/Z3, P-TEFb becomes free to phosphorylate Pol II Ser2. I find instead that PIE-1 activity correlates with CDK-12 repression in the P-cells. However, PIE-1 was also shown to inhibit Ser5 phosphorylation by CDK-7, which may be the more critical role for PIE-1's suppression of Pol II in the P-cells (Ghosh and Seydoux, 2008). It is interesting to note that some PIE-1 mutations appeared to uncouple Ser2-P regulation from transcriptional repression and embryonic lethality in these studies.

Kinase requirements in development

Although animals lacking *cdk-9* and *cdk-12* function arrest at different points, both kinases are essential for normal somatic development and growth. Furthermore, the kinase activity of these proteins is essential for somatic development as kinase-dead point mutants do not rescue lethality of deletion mutants. However, while both of these kinases play a role in somatic Ser2-P levels, it is unclear if Ser2 is the sole, or even essential, target for either CDK-9 or CDK-12 in somatic tissues.

In contrast to the soma, CDK-9, but not CDK-12 is required for normal germ line development and growth, yet in this lineage CDK-9 is not required for its most-studied activity,

phosphorylation of Ser2 in the Pol II CTD. This may not be surprising as CDK-9 has many other, well-characterized roles in transcriptional regulation (*45*). Although our data suggest that CDK-9 may be involved in some aspect of germline transcriptional regulation (it is associated with DNA and is reduced on the transcriptionally-suppressed X chromosome in the adult gonad), it is not yet clear which of the multiple proposed roles for CDK-9 comprise its essential function in the germ line.

Because phosphorylation of Ser2 in germ cells does not require CDK-9, I had the unique ability to assess the requirement for Ser2 phosphorylation in transcription and development apart from any other roles of the essential CDK-9 protein. Given its essential function in larval growth, I was surprised to find that CDK-12 was not required for germline development or fertility. I assume that the near-normal development of germ line stem cells, their normal entry and progression through the meiotic program, and the successful production of functional gametes are unlikely to occur with significant alterations in transcriptional regulation. Thus, neither CDK-12 nor Ser2-P is required for transcription in general at any stage of germ cell development in *C. elegans*. Furthermore, our results indicate that CTD phosphoepitopes may not always be accurate indicators of different stages of transcriptional regulation; e.g., absence of Ser2-P may not always reflect an absence of elongating Pol II.

What role does Ser2 phosphorylation play in germline transcription? One role of CTD phosphorylation is to provide binding sites for complexes involved in transcript processing and nucleosome dynamics with the elongating polymerase. Given the essential and predominant role of co- and post-transcriptional processing in *C. elegans* germ cells, it is unlikely these mechanisms are grossly affected by the loss of Ser2-P in the *cdk-12* mutant at optimal growth conditions. However, I did observe a dramatic decrease in co-transcriptional H3K36me3 caused by loss of CDK-12. The role of H3K36 methylation, which is largely limited to gene bodies, is thought to decrease chromatin remodeling activity in the wake of Pol II elongation and restrict initiation to the proper region (*46, 47*). The loss of this integrity could most dramatically impact

the transcriptome during post-embryonic growth or during stress. Furthermore, binding of transcription-associated RNA processing factors or other chromatin-modifying complexes may be more severely disrupted during stress in the absence of Ser2-P.

Thus, while Ser2-P is not essential for germline transcription during optimal conditions, it may be important for proper gene expression regulation during larval growth or under stress. Indeed, germ cell development is impacted by defective CDK-12 activity at higher temperature, since the *cdk-12(ok3664)* rescued with the *pal-1:3'UTR* construct is sterile at 25°C. This requirement under stress is similar to what is observed in budding yeast, where deletion of Ctk1, the major Ser2 kinase, causes cold sensitive growth (*48*).

The tissue-specific difference in kinase requirements in bulk Ser2-P is readily apparent, but it is possible that these differences actually reflect a difference in gene-specific Ser2-P regulation. Specifically, CDK-9 and CDK-12 could regulate Ser2-P in distinct sets of genes and the presence or absence of detectable Ser2-P in different tissues may reflect different sets of genes predominantly expressed in these tissues. It is also possible that the very small but quantifiable amounts of Ser2-P that I see in *cdk-12(RNAi)* or mutant germ cells represents specific CDK-9 target loci that are required for germ cell development and viability. Finally, a decreased correlation between transcription and Ser2-P may be a general property of pluripotent cells, as Ser2-P is undetectable in melanocyte and other stem cells (*49*). In any case, regulation of Ser2-P in *C. elegans* germ cells is clearly unique from other described systems: the majority of Ser2-P is not downstream of CDK-9 activity and is not required for development.

Mechanism behind tissue-specific kinase requirements on Ser2-P

It is unclear why Ser2-P is regulated differently in *C. elegans* soma and germ line. I was unable to identify a molecular basis for this phenomenon as all kinase/cyclin components are all ubiquitously expressed, including another known target of CDK-9, DSIF. It is possible that the mechanism behind tissue-specific Ser2-P kinase requirements may lie in post-translational

modifications to these proteins or expression of other factors not explored in this study. However, this difference in requirements may represent a basic difference in the way transcription is regulated between these lineages.

As described above, regulation of Ser2-P in embryonic somatic tissues is consistent with the current model of metazoan Ser2-P regulation: CDK9 is required upstream of CDK12. Previous studies have suggested that this major "upstream" role of CDK9 in metazoans is to alleviate negative regulation of "promoter proximal paused" polymerase, or Pol II that has initiated transcription but cannot progress to elongation. In the absence of CDK9, Pol II remains paused and is thus not accessible to CDK12-mediated Ser2 phosphorylation. In systems with "paused" polymerase, this pausing is largely mediated by the negative regulator, NELF, (*7*). In contrast to these systems, NELF-mediated Pol II pausing is not a general mechanism of gene regulation in *C. elegans* as it does not contain homologs to NELF components. Furthermore, "paused" polymerase signatures are also largely absent from *C. elegans* Pol II profiles, excluding a few hundred genes in starved *C. elegans* larvae [(*50*); Kruesi, W., Core, L., Waters, C., Lis, J.T. and Meyer, B.J, personal communication]. Thus, while the upstream requirement of CDK-9 for somatic Ser2-P in embryos is unclear, it is intriguing to speculate that there may be a mechanism of Pol II elongation regulation analogous to NELF-mediated "pausing" in embryonic soma and in starved larvae. This may be regulated by DSIF in combination with soma-specific factors.

In contrast to the soma, Ser2-P regulation in the germline more closely matches yeast Ser2-P regulation. Here, the CDK-9 homolog, Bur1, is not required upstream of Ser2 phosphorylation by the CDK-12 homolog, Ctk1 (*9*). The similarity between this tissue and yeast is particularly interesting as both organisms lack NELF-mediated pausing, and thus don't appear to require CDK-9/Bur1 for the progression of initiated Pol II into elongation and CDK-12/Ctk1-mediated Ser2-P. However, CDK-9 is essential for germline development in worms, as Bur1 is for viability in yeast, and the essential target is still unclear in both organisms.

The striking differences between these two modes of Ser2 regulation likely reflect inherent differences in the regulation of the transition between Pol II initiation and elongation in soma versus germline. Control of this transition may be more important in somatic development, in which the spatial and temporal regulation of tissue-specific gene expression is tightly regulated. In contrast, this transition to elongation may not be as tightly regulated in *C. elegans* germ cells, which predominantly relies on post-transcriptional regulation for spatial and temporal protein patterns (*37*). In addition, there is a significant epigenetic contribution to the guidance of germ cell transcription in *C. elegans* that might decrease reliance on dynamic promoter regulation by kinases (*13, 40, 51*).

The transition from initiated to elongating Pol II has also been proposed to provide a checkpoint for proper pre-mRNA capping (*52*). It is important to note that the majority of *C. elegans* genes are trans-spliced at their 5' end to produce a capped mRNA message (*53*). It is possible that differences in trans-splicing may play a role in tissue-specific Ser2 kinase requirements, but there doesn't appear to be an enrichment or depletion of germline function in genes that are trans-spliced versus those that are not (*54*). However, many genes in *C. elegans* are in operons, with trans-splicing providing the processing that separates the individual mRNAs in polycistronic messages (*53*). A strong bias has been reported for genes expressed in germ cells to exist in operons, which may obviate a need for coordinating transitions during elongation with mRNA capping in genes expressed in these cells (*55*).

In summary, I have identified novel and tissue-specific regulation of Ser2-P in *C. elegans*. I believe that this demonstration of P-TEFb-independent Ser2-P in the *C. elegans* germline, the first to our knowledge in a multicellular organism, will provide a unique model for understanding the role of Ser2-P in transcription regulation during development. It is also intriguing to speculate that the different kinase requirements for Ser2-P and germ line-specific modes of Pol II elongation regulation may lie at the heart of germline immortality.

Acknowledgements

I thank Rebecca Adams, Dr. Sujata Bhattacharyya, Dr. Roger Deal, and Dr. Daniel Reines for reading and commenting on the manuscript. I also thank Dr. Christian Frøkjær-Jensen for reagents and advice on the MOS-SCI method and Dr. Sujata Bhattacharyya, Dr. Jason Lieb, and members of the Kelly lab for helpful discussion of the project. This work was supported by National Institutes of Health grant GM 077600 (W.G.K.) and the National Science Foundation Graduate Research Fellowship Program (E.A.B.).

Figures

Figure 4.1: Kinase and cyclin dependency of Ser2-P in embryonic soma and germ cells.

A. Anti-Ser2-P immunofluorescence (green) analyses of embryos exposed to RNAi targeting CDKs and cyclins as indicated and counter-stained with DAPI (red). Arrows indicates the two primordial germ cells, Z2/Z3, marked by anti-Pgl-1 (blue). B. Quantification of anti-Ser2-P immunofluorescence relative to that of anti-AMA-1 (total Pol II) in either Z2/Z3 or representative somatic nuclei, normalized to vector controls. "±" indicates standard error (see experimental procedures).

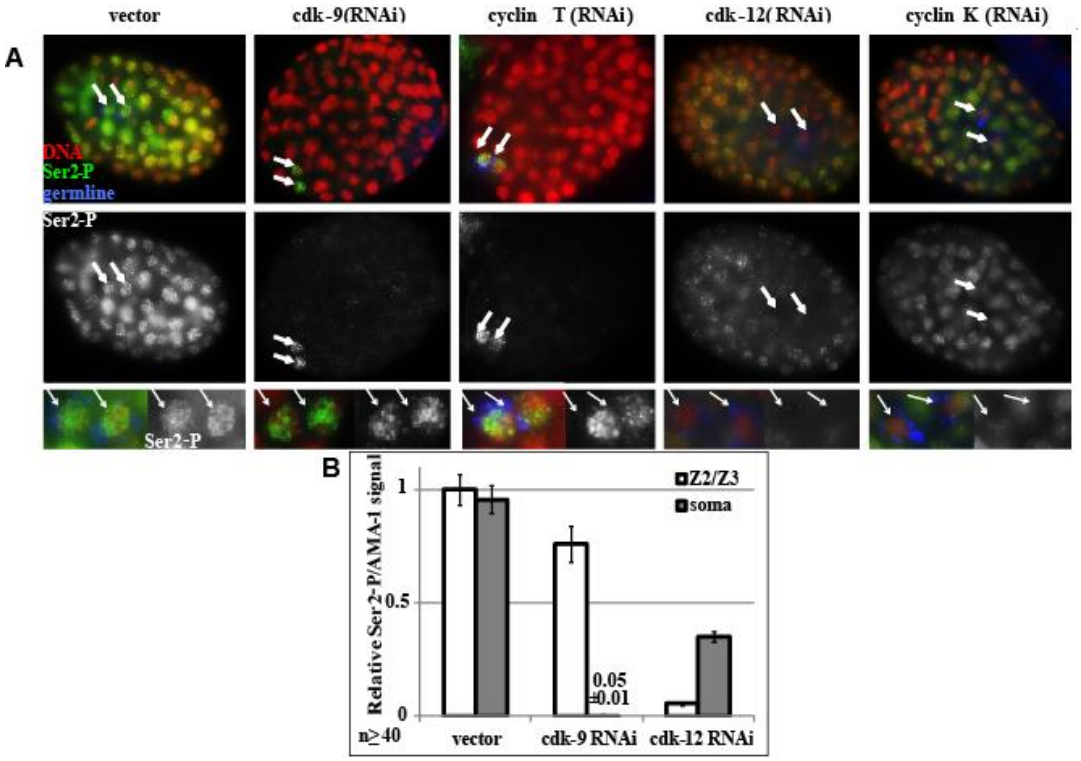


Figure 4.2: Quantification of Pol II and CTD modifications in the embryo.

A, E, F, G. Quantification of 3E10:Ser2-P immunofluorescence. B. Quantification of anti-Ser5-P (3E8) immunofluorescence. C, H. Quantification of total anti-AMA-1 immunofluorescence. D. Quantification of anti-Ser5-P signal versus total anti-AMA-1 immunofluorescence. Signal intensities measured in Z2/Z3 or representative somatic nuclei ±s.e.m. (see materials and methods for statistical analysis) in the indicated RNAi conditions. Signal is normalized to vector-treated Z2/Z3 signal.

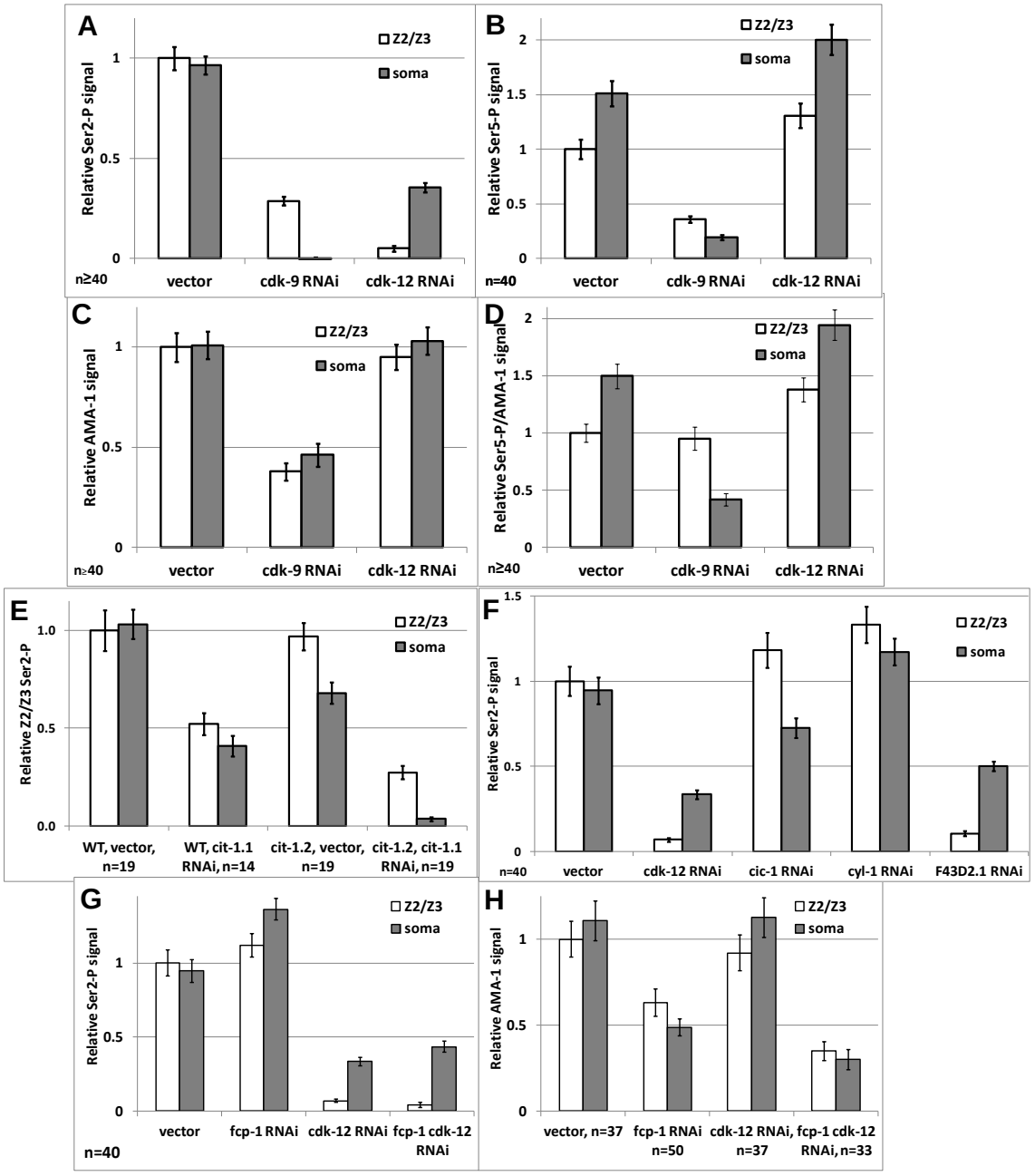

A
Relative Ser2-P signal
Z2/Z3
soma
n≥40
vector
cdk-9 RNAi
cdk-12 RNAi
B
Relative Ser5-P signal
n=40
C
Relative AMA-1 signal
D
Relative Ser5-P/AMA-1 signal
E
Relative Z2/Z3 Ser2-P
WT, vector, n=19
WT, cit-1.1 RNAi, n=14
cit-1.2, vector, n=19
cit-1.2, cit-1.1 RNAi, n=19
F
Relative Ser2-P signal
n=40
cdk-12 RNAi
cic-1 RNAi
cyl-1 RNAi
F43D2.1 RNAi
G
Relative Ser2-P signal
fcp-1 RNAi
fcp-1 cdk-12 RNAi
H
Relative AMA-1 signal
vector, n=37
fcp-1 RNAi n=50
cdk-12 RNAi, n=37
fcp-1 cdk-12 RNAi, n=33

Figure 4.3: Arrest point of RNAi treated worms.

DIC images of progeny ~2 days following indicated RNAi treatments. Vector treated animals develop normally while progeny following cdk-12 RNAi arrest shortly after hatching and progeny following cdk-9 RNAi arrest as embryos. All images are on the same scale. Solid lines outline the germ cells and dotted box shows larval gonad region expanded illustrating proliferation of the germ line.

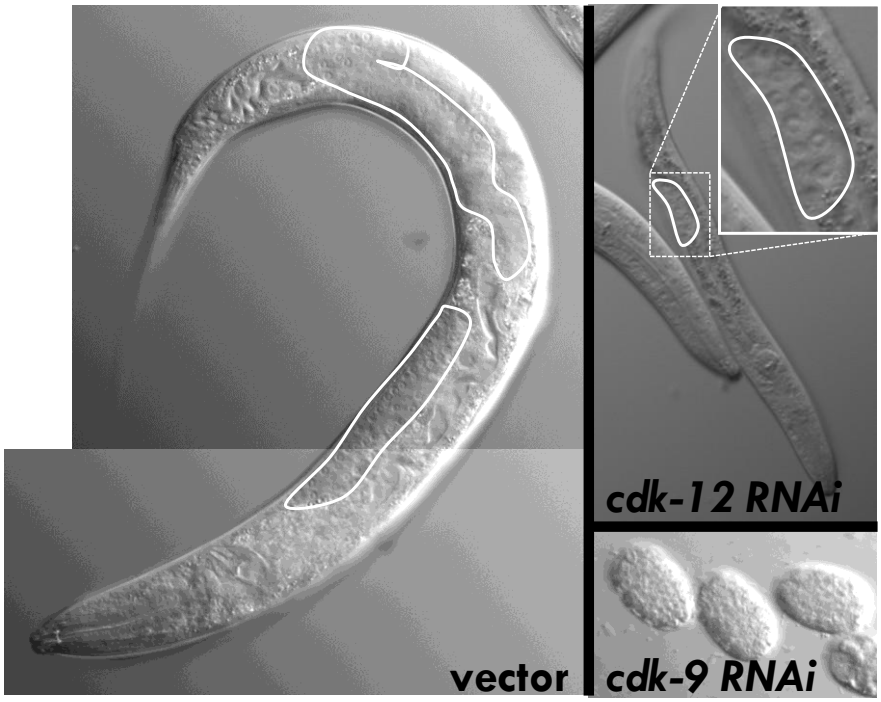


Figure 4.4: Analysis of kinase RNAi knockdown specificity and efficiency in CDK-9; CDK-12 double transgenic animals.

Fluorescently tagged CDK-9:mCherry and CDK-12:GFP transgenes show that the appropriate kinase was specifically and robustly knocked down in all embryonic nuclei under the RNAi conditions used.

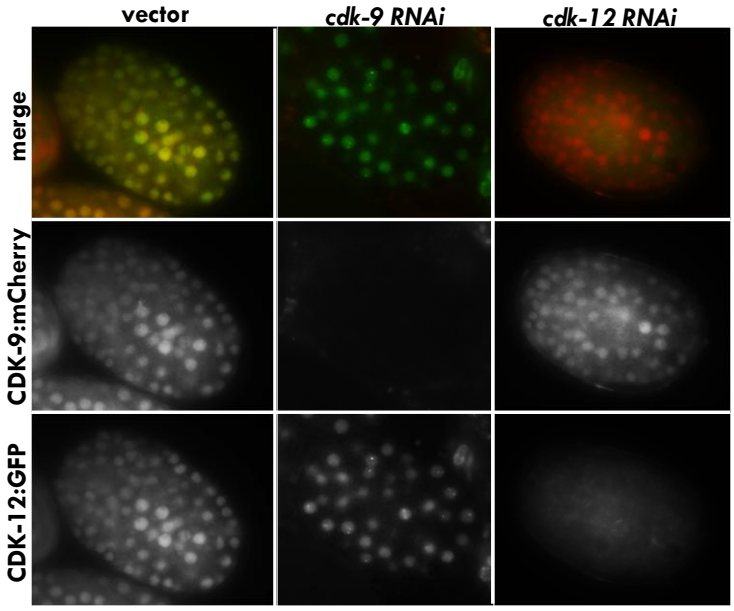


Figure 4.5: FCP-1 phosphatase activity does not contribute to CDK-12-dependent Ser2-P.

A. Anti-Ser2-P immunofluorescence (green) analyses of embryos exposed to RNAi targeting *cdk-12* and/or *fcp-1* and counterstained with DAPI (red). Arrows indicate the germ cells, Z2/Z3, marked by anti-Pgl-1 (blue). B. Quantification of Ser2-P immunofluorescence, as in Figure 4.1.

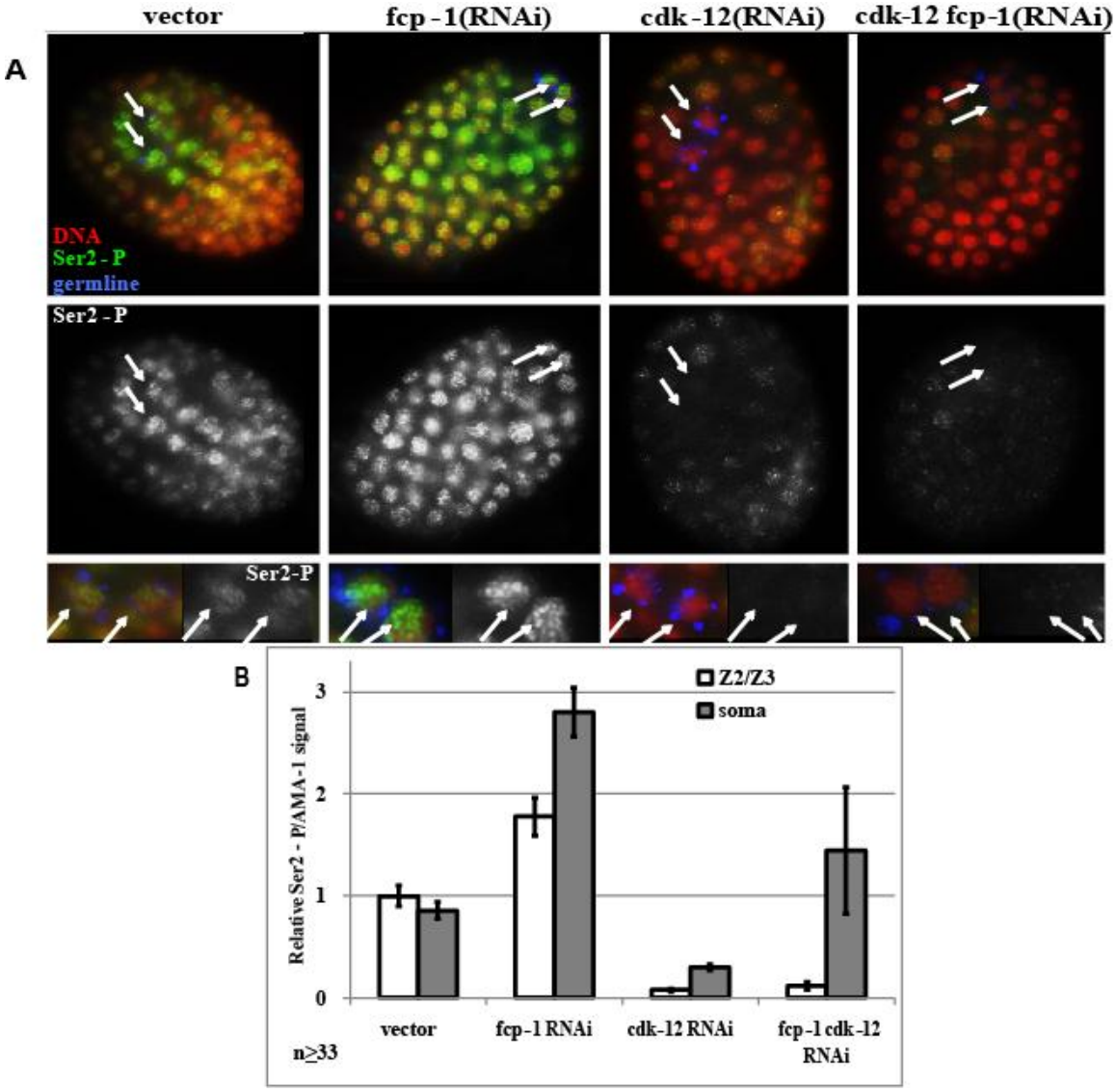


Figure 4.6: PIE-1 regulates CDK-12-dependent Ser2-P.

A. Anti-Ser2-P (green) and PIE-1 (blue) immunofluorescence analyses of embryos exposed to RNAi targeting *zif-1* and counterstained with DAPI (red). Box indicates region expanded in inset with germ cells indicated by arrows. B. Quantification of PIE-1 immunofluorescence in embryos with the Z2/Z3 precursor cell, P4, or embryos in which Z2/Z3 have been specified, staged by cell number and normalized to levels in P4 vector control. C. Quantification of Ser2-P immunofluorescence in PGCs relative to surrounding somatic nuclei in embryos staged by cell number. Bars indicate ±s.e.m. D. Anti-Ser2-P immunofluorescence (green) analyses of *pie-1(zu127)* embryos as in Figure 4.1.

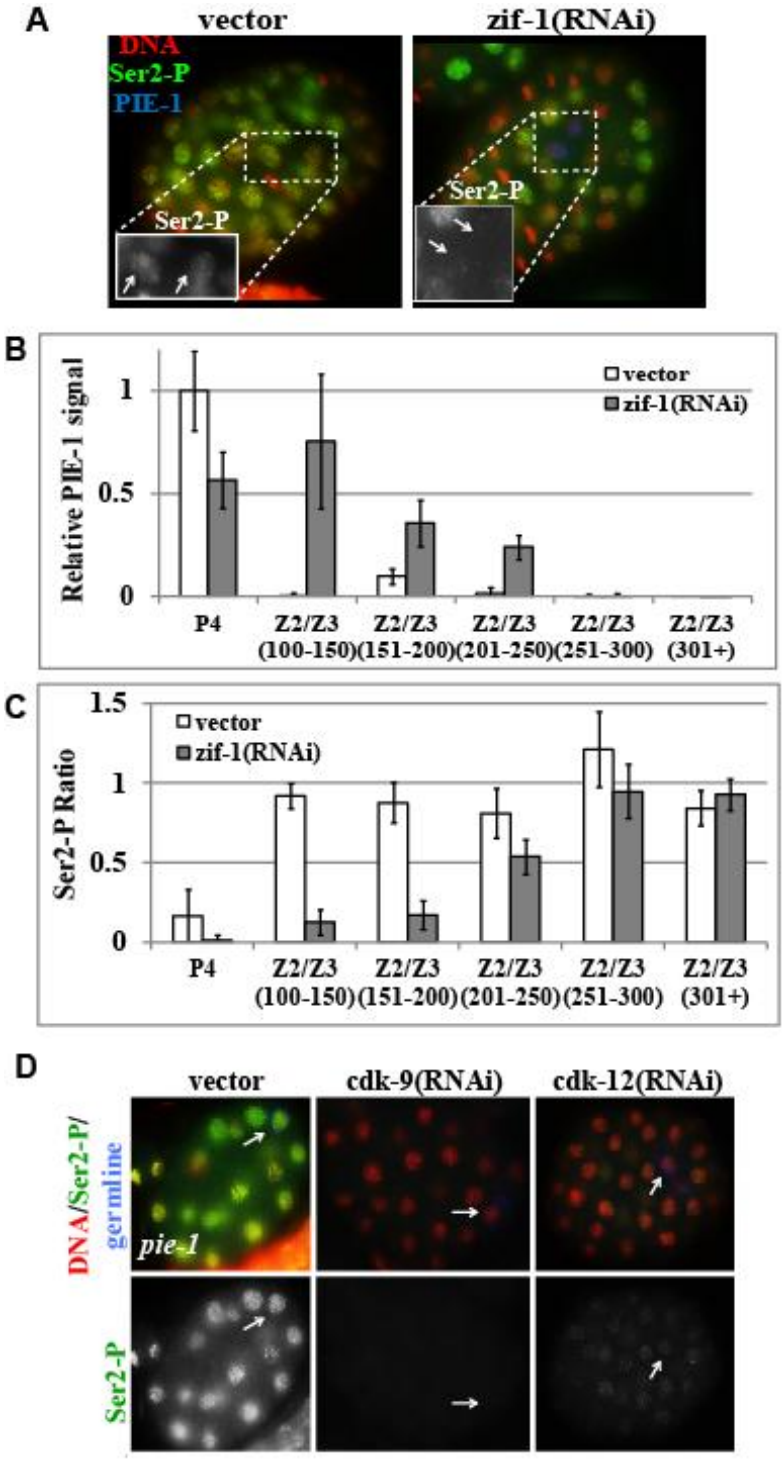

A
vector
zif-1(RNAi)
DNA
Ser2-P
PIE-1
Ser2-P
Ser2-P
B
Relative PIE-1 signal
1
0.5
0
vector
zif-1(RNAi)
P4
Z2/Z3 (100-150)
Z2/Z3 (151-200)
Z2/Z3 (201-250)
Z2/Z3 (251-300)
Z2/Z3 (301+)
C
Ser2-P Ratio
1.5
1
0.5
0
vector
zif-1(RNAi)
P4
Z2/Z3 (100-150)
Z2/Z3 (151-200)
Z2/Z3 (201-250)
Z2/Z3 (251-300)
Z2/Z3 (301+)
D
vector
cdk-9(RNAi)
cdk-12(RNAi)
DNA/Ser2-P/ germline
pie-1
Ser2-P

Figure 4.7: Ser2-P, but not proliferation, requires CDK-12 in early larvae.

A. Anti-Ser2-P immunofluorescence analyses of hatched larvae that developed from embryos exposed to *cdk-12(RNAi)* as in Figure 4.1. Box indicates region expanded in inset with germ cells outlined. B. Quantification of anti-Ser2-P immunofluorescence signal in germ cells relative to surrounding gut nuclei and normalized to RNAi vector controls. Bars indicate ±s.e.m.

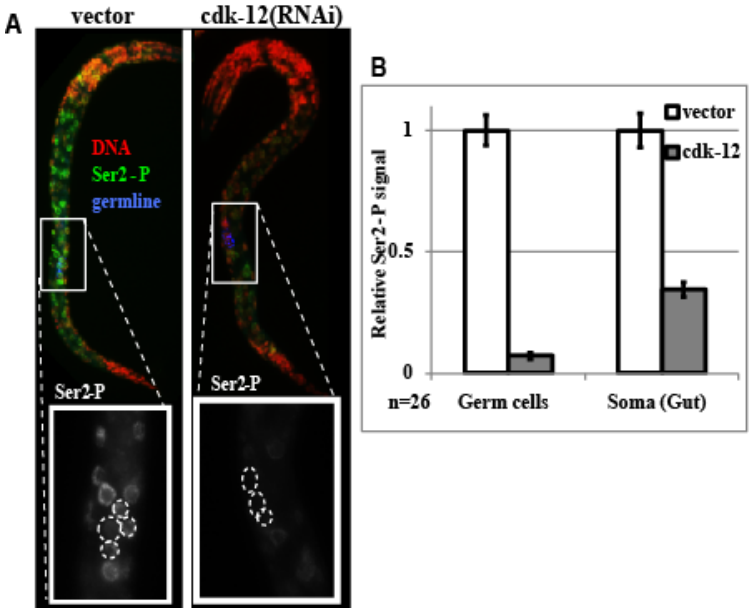


Figure 4.8: Ser2-P in adult germline cells is independent of P-TEFb.

A. *kinase(RNAi)* targeting in animals expressing CDK-9:mCherry and CDK-12:GFP showing that the appropriate kinase was specifically and robustly knocked down in the germline (outlined regions) under the RNAi conditions used. B. Anti-Ser2-P immunofluorescence analyses of dissected ovaries exposed to the indicated RNAi conditions. Mitotic (stem cell pool) and meiotic regions of the gonad are indicated. Distal and proximal ends of the gonad are marked by "d" and "p" respectively. C. Quantification of anti-Ser2-P immunofluorescence signal versus total AMA-1 protein in nuclei at the extreme distal or proximal end of the gonad ±standard error (see experimental procedures). Signal is normalized to vector-treated signal.

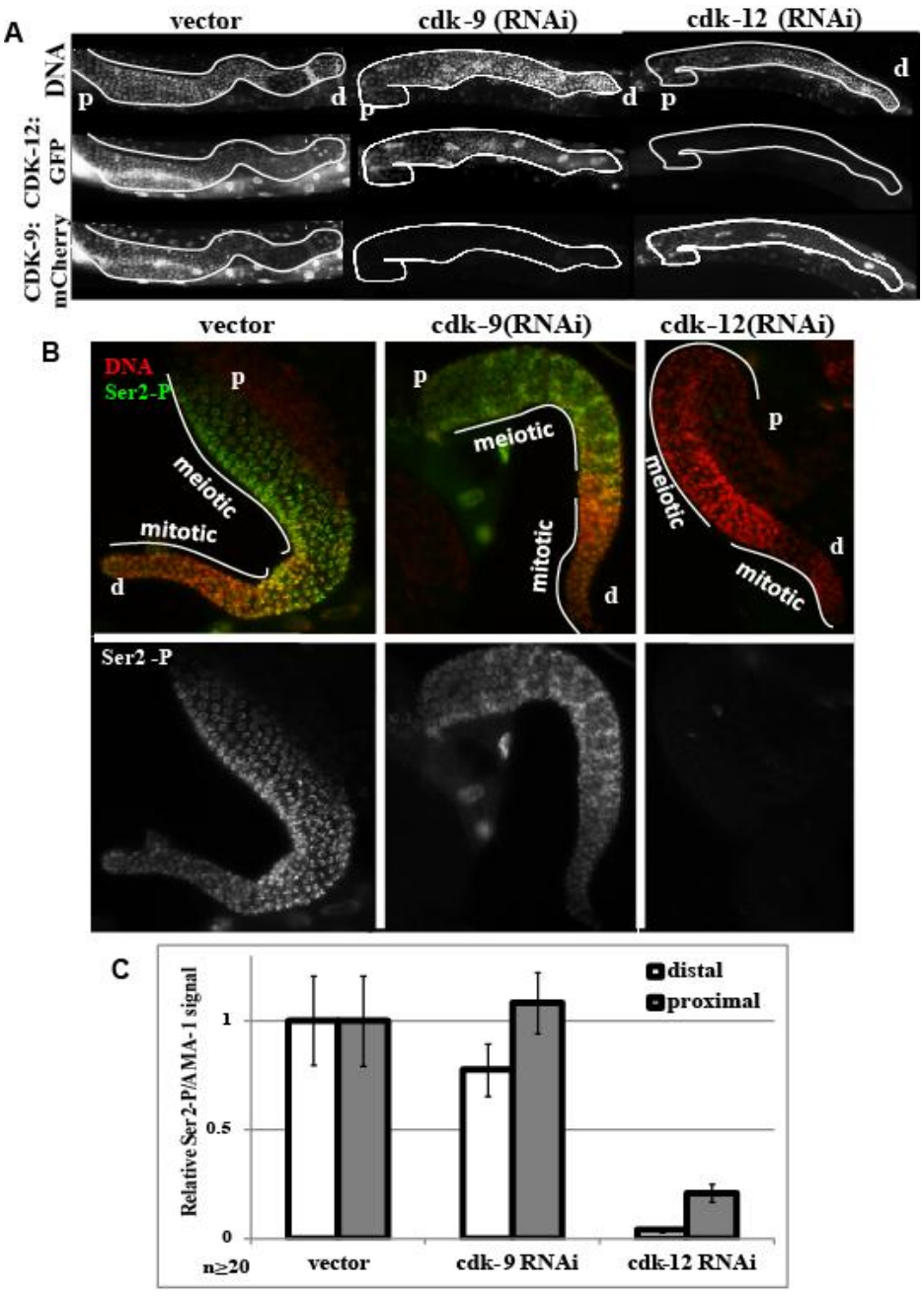

A
vector
cdk-9 (RNAi)
cdk-12 (RNAi)
DNA
CDK-12: GFP
CDK-9: mCherry
p
d
B
vector
cdk-9(RNAi)
cdk-12(RNAi)
DNA
Ser2-P
meiotic
mitotic
Ser2 -P
C
distal
proximal
Relative Ser2-P/AMA-1 signal
1
0.5
0
n≥20
vector
cdk-9 RNAi
cdk-12 RNAi

Figure 4.9: Quantification of Pol II and CTD modifications in the adult and gonad.

A. Western blot of protein extracted from worms following RNAi conditions. The same amount of protein was loaded in each lane. Total AMA-1 was detected using an antibody against the N terminus of AMA-1 and Ser2-P was detected using the H5 antibody. Decreases in total AMA-1 in cdk-12(RNAi) were not consistently observed. B. Quantification of Ser2-P relative to total AMA-1 levels normalized to vector-treated sample. ±s.e.m., n=3. C-H. Quantification of immunofluorescence signals in the adult gonad: C. Quantification of relative Ser5-P signal versus total AMA-1. D. Quantification of relative Ser5-P signal versus total AMA-1. E. Quantification of anti-Ser5-P (3E8) immunofluorescence. F. Quantification of anti-Ser7-P (4E12) immunofluorescence. G. Quantification of anti-Ser2-P (3E10) immunofluorescence. E. Quantification of total anti-AMA-1 immunofluorescence. Signal intensities measured at the extreme distal or proximal end of the gonad ±s.e.m. (see materials and methods for statistical analysis). Signal is normalized to vector-treated signal.

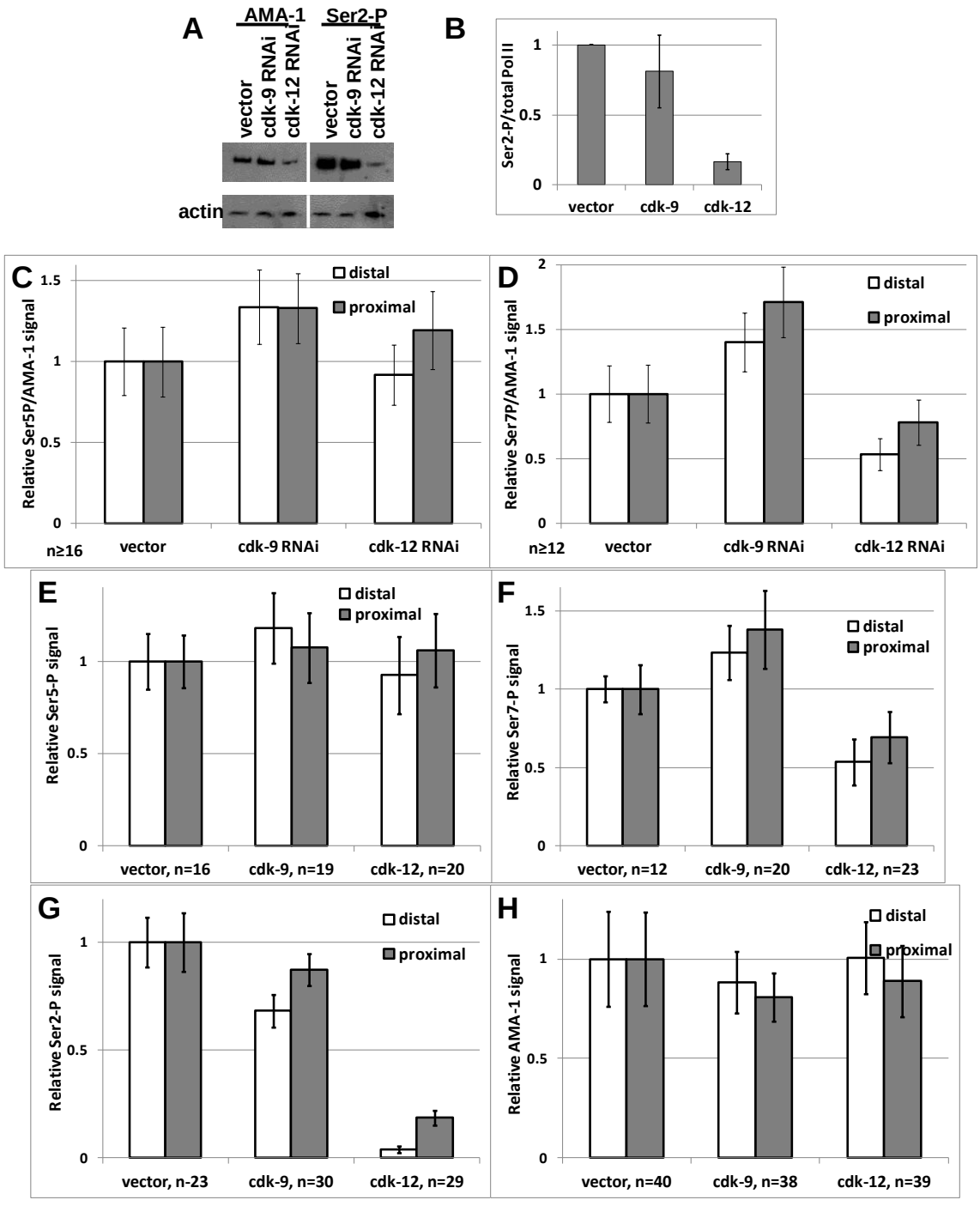

A
AMA-1
Ser2-P
vector
cdk-9 RNAi
cdk-12 RNAi
actin
B
Ser2-P/total Pol II
vector
cdk-9
cdk-12
C
Relative Ser5P/AMA-1 signal
distal
proximal
n≥16
vector
cdk-9 RNAi
cdk-12 RNAi
D
Relative Ser7P/AMA-1 signal
n≥12
E
Relative Ser5-P signal
vector, n=16
cdk-9, n=19
cdk-12, n=20
F
Relative Ser7-P signal
vector, n=12
cdk-9, n=20
cdk-12, n=23
G
Relative Ser2-P signal
vector, n-23
cdk-9, n=30
cdk-12, n=29
H
Relative AMA-1 signal
vector, n=40
cdk-9, n=38
cdk-12, n=39

Figure 4.10: CDK-9, CDK-12, and their cyclin partners are ubiquitously expressed.

A. Expression in adult hermaphrodites of transgenic, single-copy insertions of fluorescently tagged CDK-9 (mCherry and GFP tagged versions show the same pattern, data not shown), Cyclin T (CIT-1.2 shown, CIT-1.1 has the same pattern; patterns identical with GFP or FLAG tags, not shown) CDK-12, and Cyclin K. Outlines around the adult gonad (white), oocytes (blue), and embryos (orange) show these factors are expressed in the germline and maternally loaded into embryos. Dotted boxes indicate regions expanded. B-C. Co-staining gonads with GFP and H3K4me2 show chromosomes that lack H3K4me2 (X chromosome) also lack CDK-9 and CDK-12. Representative nulcei (dotted box) shown below.

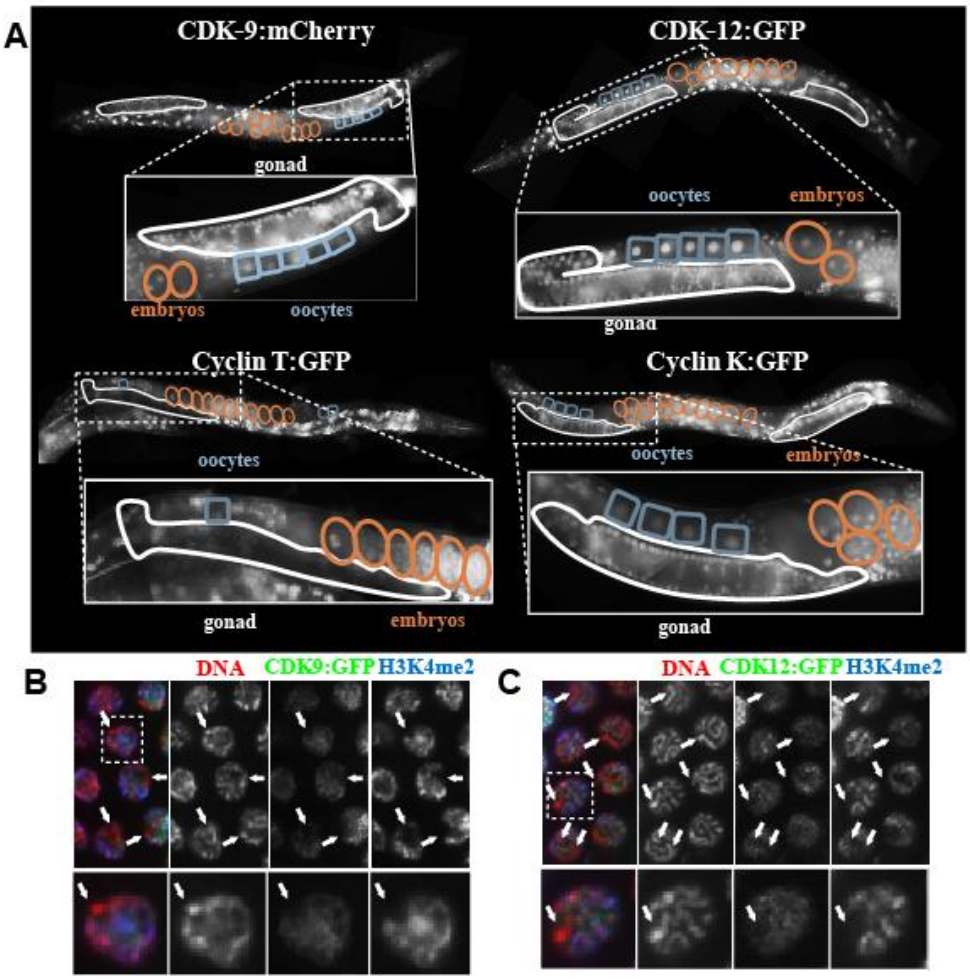


Figure 4.11: Increasing germline expression of Cyclin T (CIT-1.2) does not change germline Ser2-P kinase requirements.

A. GFP expression in an adult hermaphrodite transgenic animals expressing CIT-1.2:GFP with either its endogenous 3'UTR or ccnk-1 3'UTR. White outline is around gonad. B. Anti-GFP immunofluorescence of MOS-SCI transgenic lines expressing CIT-1.2 with either its endogenous 3'UTR or ccnk-1 K 3'UTR. Blue, pgl staining and arrows indicate the primordial germ cells, Z2/Z3. C. Anti-Ser2-P immunofluorescence images of dissected CIT-1.2:GFP:ccnk-1 3'UTR hermaphrodite gonads exposed to the indicated RNAi conditions. D. Anti-Ser2-P immunofluorescence images of CIT-1.2:GFP:ccnk-1 3' UTR embryos exposed to the indicated RNAi conditions. Blue, pgl staining and arrows indicate the germcells, Z2/Z3, expanded in inset below.

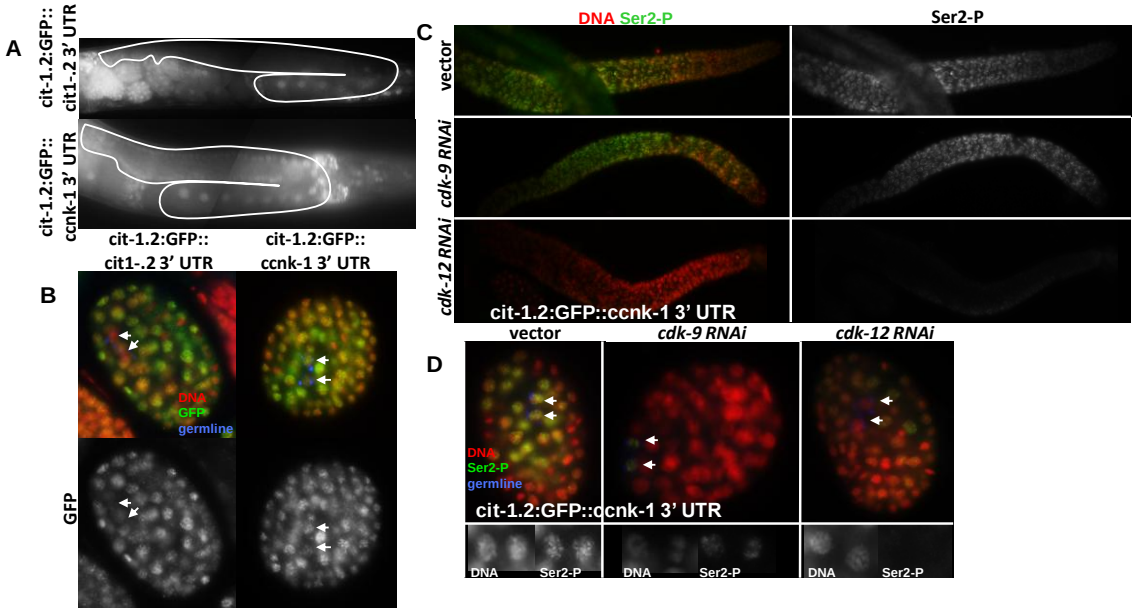


Figure 4.12: SPT-5 is ubiquitously expressed.

Expression in adult hermaphrodite of fluorescently tagged DSIF component, SPT-5. Outlines around the adult gonad (white), oocytes (blue), and embryos (orange) show these factors are expressed in the germline and maternally loaded into embryos. Dotted boxes indicate regions expanded.

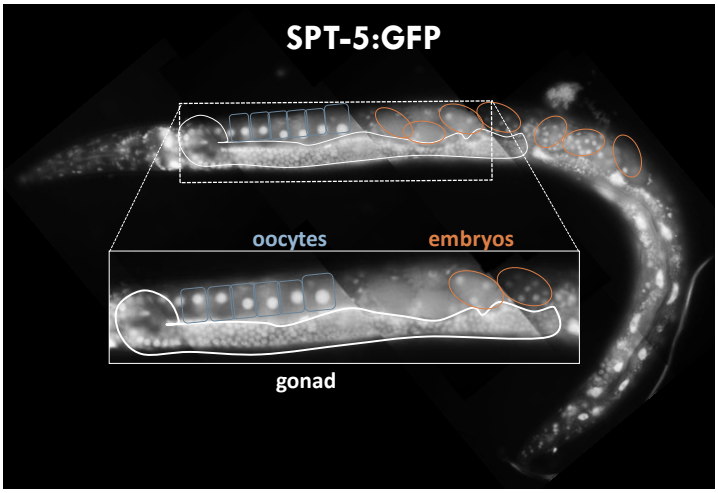


Figure 4.13: CDK-9, but not CDK-12 or Ser2P, is essential for germline development.

A. Expression of both *pal-1 3'UTR* tagged kinases (only *cdk-9::pal-13'UTR* is shown) is observed in all somatic cells, but not in germ cells expression until late oogenesis. B. Expression of *cdk-9:pal-1 3'UTR* in homozygous *cdk-9(tm2884)* adult hermaphrodites worms. Dotted box region expanded in inset. White circle surrounds germ cells. Arrows point to somatic gonad nuclei C. Anti-Ser2-P immunofluorescence analyses of *cdk-9(tm2884), cdk-9:pal-1 3'UTR* hermaphrodite ovaries (outlined in white). D. Expression of *cdk-12:pal-1 3'UTR* in adult hermaphrodite worms in homozygous *cdk-12(ok3664)* background. White outline surrounds gonad. E. Ser2-P immunofluorescence images of *cdk-12(ok3664), cdk-12:pal-1 3'UTR* hermaphrodite gonads.

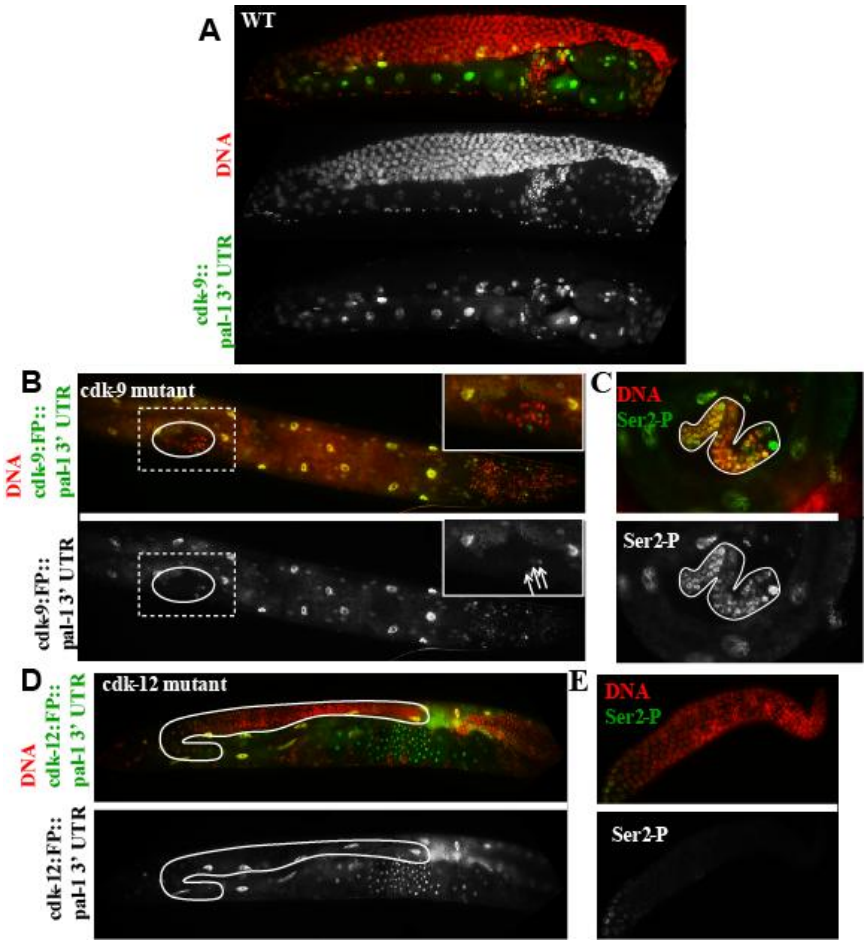


Figure 4.14: CDK-12 regulates H3K36me3 levels.

A. Ser2-P and H3K36me3 immunofluorescence images of hermaphrodite gonads in either *cdk-12(ok3664)/+*; *cdk-12:pal-1 3'UTR* (WT) or *cdk-12(ok3664)/cdk-12(ok3664)*; *cdk-12:pal-1 3'UTR* (cdk-12 mutant) animals exposed to the indicated RNAi conditions. B. Quantification of Ser2-P immunofluorescence signal ±s.e.m. C. Quantification of H3K36me3 immunofluorescence signal ±s.e.m . Nuclei selected for analysis were from the most distal mitotic and meiotic regions of the gonad.

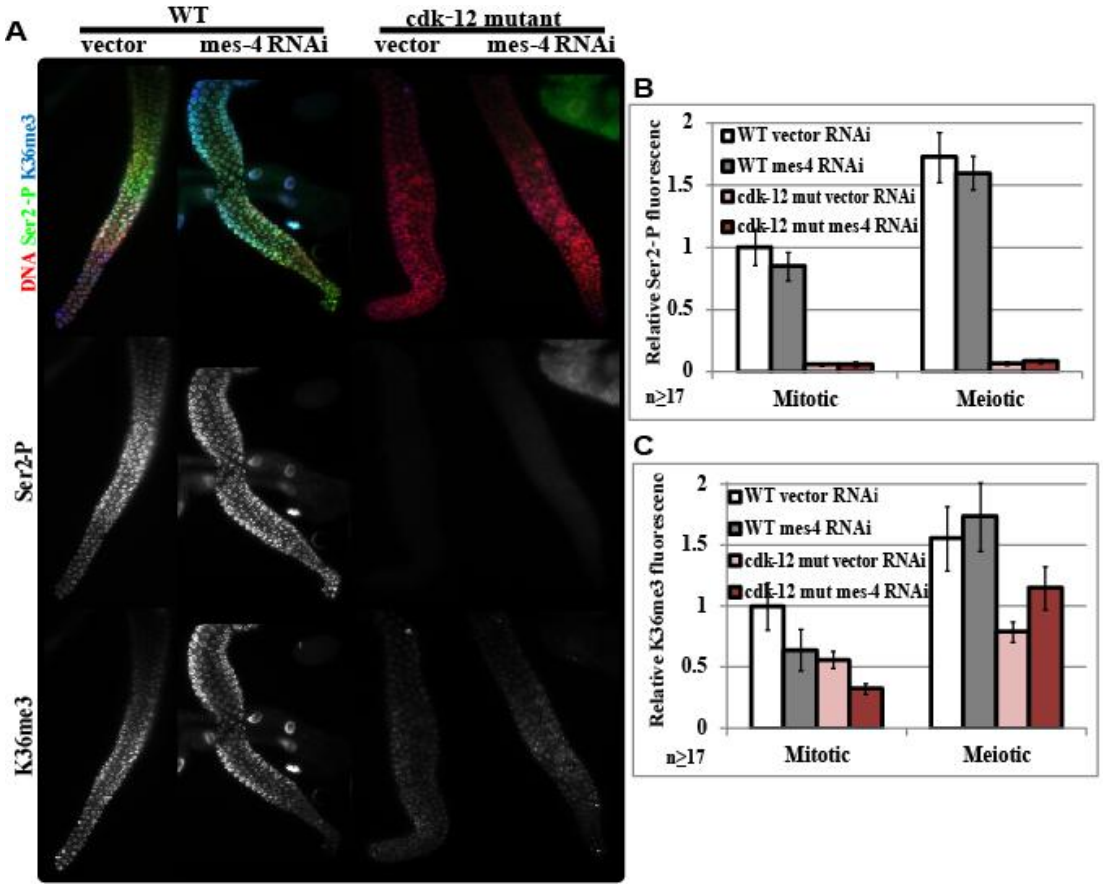
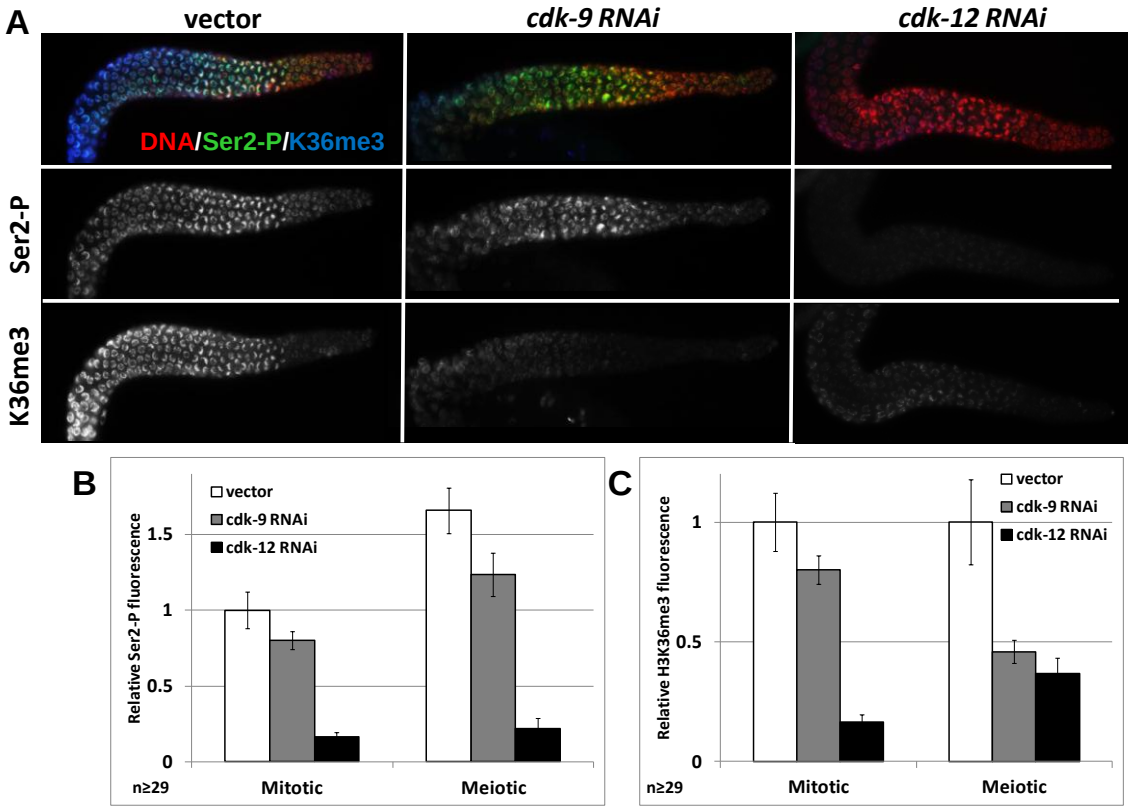


Figure 4.15: CDK-9 and CDK-12 regulate transcription-dependent H3K36me3.

A. Anti-Ser2-P and anti-H3K36me3 immunofluorescence images of dissected hermaphrodite gonads from *mes-4(bn85)* worms exposed to the indicated RNAi conditions. B. Quantification of anti-Ser2-P immunofluorescence signal ±s.e.m. C. Quantification of anti-H3K36me3 immunofluorescence signal ±s.e.m. Nuclei selected for analysis were from the most distal mitotic and meiotic regions of the gonad.



Tables

Table 4.1: Cyclin T RNAi Embryonic Lethality

	cit-1.2:GFP		cit-1.2(gk214)		cit-1.2(gk214); cit-1.2:GFP	
RNAi treatment	vector	*cit-1.1(RNAi)*	vector	*cit-1.1(RNAi)*	vector	*cit-1.1(RNAi)*
Average # emb laid	30±3	25±3	13±5	9±6	24±4	24±5
Emb lethality	0.19%	0.23%	30.34%	100%	0.24%	0.00%

Adult worms (n=17) were transferred to individual plates following vector or *cit-1.1* RNAi treatment for 6 hours. The number of embryos laid and unhatched after 24 hours were counted.

Table 4.2: New Transgenic Lines

Name	Description	Genotype	Plasmid
KW2112	CDK-12:GFP	ckSi6 (unc-119, cdk-12:GFP) I; unc-119(ed3) III	pBAB66
KW2157	CDK-12 (D462N):GFP	ckSi9 (unc-119, cdk-12:D462N:GFP) I; unc-119(ed3) III	pBAB75
KW2194	CDL-12:GFP:mex-5 3' UTR	ckSi17 (unc-119, cdk-12:GFP:mex-5 3' UTR) I; unc-119(ed3) III	pBAB96
KW2206	CDK-12:GFP:pal-1 3' UTR	ckSi26 (unc-119, cdk-12:GFP:pal-1 3' UTR) I; unc-119(ed3) III	pBAB97
KW2117	CCNK-1;FLAG	ckSi10 (unc-119, ccnk-1:FLAG) II; unc-119(ed3) III	pBAB76
KW2147	CCNK-1:GFP	ckSi15(unc-119, ccnk-1:GFP)II; unc-119(ed3) III	pBAB88
KW2115	CDK-9:mCherry	ckSi4 (unc-119, cdk-9:mCherry) II; unc-119(ed3) III	pBAB61
KW2167	CDK-9:GFP	ckSi13 (unc-119, cdk-9:GFP) II; unc-119(ed3) III	pBAB86
KW2159	CDK-9(D235N):GFP	ckSi12 (unc-119, cdk-9D235N:mCherry) II; unc-119(ed3) III	pBAB85
KW2195	CDK-9:mCherry:mex-5 3' UTR	ckSi20 (unc-119, cdk-9:GFP:mex-5 3' UTR) II; unc-119(ed3) III	pBAB102
KW2196	CDK-9:mCherry:pal-1 3' UTR	ckSi21 (unc-119, cdk-9:GFP:pal-1 3' UTR) II; unc-119(ed3) III	pBAB103
KW2096	CIT-1.1:FLAG	ckSi2 (unc-119, cit-1.1:FLAG) II; unc-119(ed3) III	pBAB58
KW2098	CIT-1.2:FLAG	ckSi3 (unc-119, cit-1.2:FLAG) II; unc-119(ed3) III	pBAB59
KW2140	CIT-1.2:GFP	ckSi14(unc-119, cit-1.2:GFP)II; unc-119(ed3) III	pBAB87
KW2104	SPT-5:GFP	ckSi5 (unc-119, spt-5:GFP) II; unc-119(ed3) III	pBAB62
KW2237	CIT-1.2:FLAG:CCNK-1 3' UTR	ckSi25 (unc-119, cit-1.2:FLAG:ccnk-1 3' UTR) II; unc-119(ed3) III	pBAB110

Table 4.3: Additional Strains Made

Name	Description	Genotype
KW2126	CDK-12:GFP rescue of cdk-12 deletion	cdk-12(tm3846)III; ckSi6 (unc-119, cdk-12:GFP) I
KW2214	CDK-12:GFP rescue of cdk-12 deletion	cdk-12(ok3664)III; ckSi6 (unc-119, cdk-12:GFP) I
KW2209	CDK-12(D462N):GFP in balanced cdk-12 deletion	cdk-12(ok3664)/qC1 qIs26 (lag-2:GFP; rol-6) III; ckSi9 (unc-119, cdk-12 D462N:GFP) I; unc-119(ed3) III
KW2210	CDK-12:GFP:mex-5 3' UTR in balanced cdk-12 deletion	cdk-12(ok3664)/qC1 qIs26 (lag-2:GFP; rol-6) III; ckSi17 (unc-119, cdk-12:GFP:mex-5 3' UTR) I
KW2211	CDK-12:GFP:pal-1 3' UTR in balanced cdk-12 deletion	cdk-12(ok3664)/qC1 qIs26 (lag-2:GFP; rol-6) III; ckSi18 (unc-119, cdk-12:GFP:pal-1 3' UTR) I
KW2183	CDK-9:mCherry rescue of cdk-9 deletion	cdk-9(tm2884) I; ckSi4 (unc-119, cdk-9:mCherry) II
KW2181	CDK-9(D235N):mCherry in balanced cdk-9 deletion mutant	cdk-9(tm2884)/ht2 qIs48 (myo-2:GFP) I, III; ckSi12 (unc-119, cdk-9 D235N:mCherry) II
KW2204	CDK-9:mCherry:mex-5 3' UTR in balanced cdk-9 deletion	cdk-9(tm2884)/ht2 qIs48 (myo-2:GFP) I, III; ckSi20 (unc-119, cdk-9:mCherry:mex-5 3' UTR) II
KW2205	CDK-9:mCherry:pal-1 3' UTR in balanced cdk-9 deletion	cdk-9(tm2884)/ht2 qIs48 (myo-2:GFP) I, III; ckSi21 (unc-119, cdk-9:mCherry:pal-1 3' UTR) II
KW2185	CDK-12:GFP, CDK-9:mCherry	ckSi6 (unc-119, cdk-12:GFP) I; ckSi4 (unc-119, cdk-9:mCherry) II; unc-119(ed3) III
KW2268	CIT-1.2:GFP rescue of cit-1.2 deletion	cit-1.2(gk241) III; ckSi14(unc-119, cit-1.2:GFP) II

Table 4.4: Cloning Methods

Name	Parent vector	Description	Made from pBAB plasmid	Cloning method	Primers
pBAB30	pCR blunt	cit-1.1 ORF	-	blunt ligation with PCR prod: cit-1.1 ORF	BB336, BB240
pBAB31	pCR blunt	cit-1.2 ORF	-	blunt ligation with PCR prod: cit-1.2 ORF	BB316, BB317
pBAB33	pCR blunt	cdk-12 ORF	-	blunt ligation with PCR prod: cdk-12 ORF	BB334, BB226
pBAB34	pCR blunt	cdk-9 5' ORF	-	blunt ligation with PCR prod: cdk-9 5' ORF	BB333, BB237
pBAB35	pCR blunt	spt-5 ORF	-	blunt ligation with PCR prod: spt-5 ORF	BB298, BB299
pBAB36	pCR blunt	Pcit-1.1 cit-1.1 ORF	pBAB30	KpnI, NheI ligation with PCR prod: Pcit-1.1	BB322, BB323
pBAB37	pCR blunt	Pcit-1.2 cit-1.2 ORF	pBAB31	BamHI, NheI ligation with PCR prod: Pcit-1.2	BB312, BB313
pBAB39	pCR blunt	Pcdk-12 cdk-12 ORF	pBAB33	KpnI, NheI ligation with PCR prod: Pcdk-12	BB320, BB321
pBAB40	pCR blunt	Pcdk-9 cdk-9 5' ORF	pBAB34	KpnI, NheI ligation with PCR prod: Pcdk-9	BB318, BB319
pBAB41	pCR blunt	Pcdk-9 cdk-9 whole ORF	pBAB40	AvrII, SbfI ligation with PCR prod: Pcit-1.5	BB238, BB332
pBAB42	pCR blunt	Pspt-5 spt-5 ORF	pBAB35	KpnI, NheI ligation with PCR prod: Pspt-5	BB300, BB301
pBAB43	pCR blunt	Pcit-1.1 cit-1.1 ORF FLAG	pBAB36	SbfI, NotI ligation with PCR product: FLAG from pFS26	BB330, BB331
pBAB44	pCR blunt	Pcit-1.2 cit-1.2 ORF FLAG	pBAB37	SbfI, NotI ligation with PCR product: FLAG from pFS26	BB330, BB331
pBAB46	pCR blunt	Pcdk-12 cdk-12 ORF GFP	pBAB39	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB47	pCR blunt	Pcdk-9 cdk-9 whole ORF mCherry	pBAB41	SbfI, NotI ligation with PCR product: mCherry pFS26	BB328, BB329
pBAB48	pCR blunt	Pspt-5 spt-5 ORF GFP	pBAB42	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB49	pCR blunt	Pcit-1.1 cit-1.1 ORF FLAG cit-1.1 3' UTR	pBAB43	NotI, ApaI ligation with PCR prod: cit1.1 3' UTR	BB310, BB311
pBAB50	pCR blunt	Pcit-1.2 cit-1.2 ORF FLAG cit-1.2 3' UTR	pBAB44	NotI, ApaI ligation with PCR prod: cit1.2 3' UTR	BB314, BB315
pBAB51	pCR blunt	Pccnk-1 ccnk-1 ORF FLAG ccnk-1 3' UTR	pBAB57+ pBAB65	KpnI, SbfI ligation	
pBAB52	pCR blunt	Pcdk-12 cdk-12 ORF GFP cdk-12 3' UTR	pBAB46	NotI, ApaI ligation with PCR prod: cdk-12 3' UTR	BB306, BB307
pBAB53	pCR blunt	Pcdk-9 cdk-9 ORF mCherry cdk-9 3' UTR	pBAB47	NotI, ApaI ligation with PCR prod: cdk-9 3' UTR	BB304, BB392
pBAB54	pCR blunt	Pspt-5 spt-5 ORF GFP spt-5 3' UTR	pBAB48	NotI, ApaI ligation with PCR prod: spt-5 3' UTR	BB302, BB303
pBAB57	pCR blunt	Pccnk-1 ccnk-1 ORF FLAG cit-1.1 3' UTR	pBAB49	KpnI, SbfI ligation with PCR prod: Pccnk-1 ccnk-1 ORF	BB324, BB228
pBAB58	pCFJ151	cit-1.1 FLAG	pBAB49	blunt ligation after KpnI, ApaI digestion	
pBAB59	pCFJ151	cit-1.2 FLAG	pBAB50	blunt ligation after BamHI, ApaI digestion	
pBAB60	pCFJ151	cdk-12 GFP	pBAB52	blunt ligation after KpnI, ApaI digestion	
pBAB61	pCFJ151	cdk-9 mCherry	pBAB53	blunt ligation after KpnI, ApaI digestion	
pBAB62	pCFJ151	spt-5 GFP	pBAB54	blunt ligation after KpnI, ApaI digestion	
pBAB65	pCR blunt	Pcit-1.2 cit-1.2 ORF FLAG ccnk-1 3' UTR	pBAB50	NotI, ApaI ligation with PCR product: ccnk-1 3' UTR	BB308, BB309
pBAB66	pCFJ352	cdk-12 GFP	pBAB52	blunt ligation after KpnI, ApaI digestion	
pBAB68	pCR blunt	cdk-9 5' ORF D235N	pBAB34	Qiagen XL QC	BB385, BB386
pBAB71	pCR blunt	cdk-12 ORF D462N	pBAB33	Qiagen XL QC	BB383, BB384
pBAB75	pCFJ352	cdk-12 D462N GFP	pBAB66+ pBAB71	SphI, NheI ligation	
pBAB76	pCFJ151	ccnk-1 FLAG	pBAB51	blunt ligation after KpnI, ApaI digestion	
pBAB79	pCR blunt	Pspt-5 spt-5 ORF FLAG spt-5 3' UTR	pBAB54	SbfI, NotI ligation with PCR product: FLAG from pFS26	BB330, BB331
pBAB80	pCR blunt	Pcdk-12 cdk-12 ORF FLAG cdk-12 3' UTR	pBAB52	SbfI, NotI ligation with PCR product: FLAG from pFS26	BB330, BB331

pBAB81	pCR blunt	Pcit-1.1 cit-1.1 ORF GFP cit-1.1 3' UTR	pBAB49	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB82	pCR blunt	Pcit-1.2 cit-1.2 ORF GFP cit-1.2 3' UTR	pBAB50	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB83	pCR blunt	Pcdk-9 cdk-9 ORF GFP cdk-9 3' UTR	pBAB53	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB84	pCR blunt	Pccnk-1 ccnk-1 ORF GFP ccnk-1 3' UTR	pBAB51	SbfI, NotI ligation with PCR product: GFP from pFS19	BB326, BB327
pBAB85	pCFJ151	cdk-9 D235N mCherry	pBAB61+ pBAB68	EcoNI, SphI ligation	
pBAB86	pCFJ151	cdk-9 GFP	pBAB83+ pBAB61	NotI, SphI ligation	
pBAB87	pCFJ151	cit-1.2 GFP	pBAB82+ pBAB59	NotI, NheI ligation	
pBAB88	pCFJ151	ccnk-1 GFP	pBAB84+ pBAB76	NotI, SphI ligation	
pBAB96	pCFJ352	cdk-12:GFP:mex-5 3' UTR	pBAB66	NotI, SnaBI ligation with PCR product: mex-5 3' UTR as specified in (*37*)	BB408, BB424
pBAB97	pCFJ352	cdk-12:GFP:pal-1 3' UTR	pBAB66	NotI, SnaBI ligation with PCR product: pal-1 3' UTR as specified in (*37*)	BB410, BB425
pBAB102	pCFJ151	cdk-9:mCherry:mex-5 3' UTR	pBAB61	NotI, XhoI ligation with PCR product: mex-5 3' UTR as specified in (*37*)	BB408, BB414
pBAB103	pCFJ151	cdk-9:mCherry:pal-1 3' UTR	pBAB61	NotI, XhoI ligation with PCR product: pal-1 3' UTR as specified in (*37*)	BB410, BB415

Table 4.5: Primer sequences

Name	Sequence	Description
BB226	gactCCTGCAGGcttgaaaaatactgactg	cdk-12 ORF rev SbfI
BB228	gactCCTGCAGGaaaagttgtgagcttctttttctcc	ccnk-1 ORF rev SbfI
BB237	gactCCTGCAGGgtcacctaggccacgctttgaaaatcgatg	cdk-9 inside ORF first half rev SbfI AvrII use with BB235
BB238	gactGGCCGGCCcctaggaaaccccctgtggaaaag	cdk-9 inside ORF second half fwd FseI AvrII use with BB236
BB240	gactCCTGCAGGttctagttcaccatcttccaaatc	cit-1.1 ORF rev SbfI
BB245	TCAGAGCTCTAATCGGCGGT	cdk-9(tm2884) deletion confirmation fwd
BB246	TTGCGGTGGCCGAGGTATAC	cdk-9(tm2884) deletion confirmation rev
BB247	ACTCGGCCTGTGTAAGTTAT	cdk-12(tm3846) deletion confirmation fwd
BB248	AGCTCGCCTCTGCAAACAAT	cdk-12(tm3846) deletion confirmation rev
BB249	taattttccgggtccttgtg	cit-1.2(gk241) deletion confirmation fwd
BB250	atggcctcaacttcttcacg	cit-1.2(gk241) deletion confirmation rev
BB298	gactGCTAGCatgtcctctgacgaaagtgatgc	spt-5 ORF fwd NheI
BB299	gatcCCTGCAGGagtttcgctatgcattttgcagc	spt-5 ORF rev minus stop SbfI
BB300	GATCggtaccGCCCGGGCgcaagttgtgggttttggac	spt-5 promoter fwd KpnI SrfI
BB301	gactGCTAGCtgctaactgaaacatttaagtaaat	spt-5 promotoer rev NheI
BB302	gactGCGGCCGCaaagttgttcactttactatttattc	spt-5 3' UTR fwd NotI
BB303	gactGGGCCCGCCCGGGCctcttcattcttgatctcac	spt-5 3' UTR rev SrfI ApaI
BB304	gactGCGGCCGCgctctttttccctatttttttcc	cdk-9 3' UTR fwd NotI
BB306	gactGCGGCCGCaaattctgatttttttgttgatta	cdk-12 3' UTR fwd NotI
BB307	gactGGGCCCGCCCGGGCttttgatccactgctgcttg	cdk-12 3' UTR rev SrfI ApaI
BB308	gactGCGGCCGCttttcaaaaatctaatatttctatat	ccnk-1 3' UTR fwd NotI
BB309	gactGGGCCCGCCCGGGCaaccacacccactttcaagc	ccnk-1 3' UTR rev SrfI ApaI
BB310	gactGCGGCCGCttatttttagttcgtatttttattag	cit-1.1 3' UTR fwd NotI
BB311	gactGGGCCCGCCCGGGCttttatccccaaatcttgatgag	cit-1.1 3' UTR rev SrfI ApaI
BB312	GATCggatccGCCCGGGCtgaaacctggacgacacaag	cit-1.2 promoter fwd BamHI SrfI
BB313	gactGCTAGCactgatcaatgctgaaaaaaatatat	cit-1.2 promoter rev NheI
BB314	gactGCGGCCGCgagcttccctcactgttatttccg	cit-1.2 3' UTR fwd NotI
BB315	gactGGGCCCGCCCGGGCcatcatgccttgtcatttcc	cit-1.2 3' UTR rev SrfI ApaI
BB316	gactGCTAGCatgtcgaattcgaacaaattgatcg	cit-1.2 ORF fwd NheI
BB317	gatcCCTGCAGGaacgagctccccttcctccatctc	cit-1.2 ORF rev minus stop SbfI
BB318	GATCggtaccGCCCGGGCaataataaaaaccacgggtttcagg	cdk-9 promoter fwd KpnI SrfI
BB319	acgtGCTAGCttgctctgaaaattgttaa	cdk-9 promoter rev NheI
BB320	GATCggtaccGCCCGGGCacgcattatcattgcgtttg	cdk-12 promoter fwd KpnI SrfI
BB321	acgtGCTAGCggctgaaaatgataagaatattaaag	cdk-12 promoter rev NheI
BB322	GATCggtaccGCCCGGGCaatgttcacgacgaaacacg	cit-1.1 promoter fwd KpnI SrfI
BB323	acgtGCTAGCcgcatttgagtttaattctc	cit-1.1 promoter rev NheI
BB324	GATCggtaccGCCCGGGCcgcggaacgtttataattca	ccnk-1 promotoer fwd KpnI SrfI
BB326	AAAGCCTGCAGGgATGAGTAAAGGAGAAG	GFP fwd +1 SbfI
BB327	GCAGGCGGCCGCTTATTTGTATAGTTC	GFP rev STOP NotI
BB328	agctcCCTGCAGGgATGGTGAGCAAGGGCGAGGAG	mCherry fwd +1 SbfI
BB329	tctaGCGGCCGCTTACTTGTACAGCTCGTCCATGC	mCherry rev STOP NotI
BB330	agctCCTGCAGGgGCCGCAGATTAC	FLAG fwd +1 SbfI
BB331	tctaGCGGCCGCTTACTTATCATCATC	FLAG rev STOP NotI
BB332	CATcCCTGCAGGaaaaatagtatcgcgatattgtc	cdk-9 ORF rev SbfI
BB333	gactGCTAGCatgagtgctcaaaactatcacgcc	cdk-9 ORF fwd NheI
BB334	gactGCTAGCatggaaatatcgccagg	cdk-12 ORF fwd NheI
BB336	gactGCTAGCatgtcggtgtcgagtcgaggcg	cit-1.1 ORF fwd NheI
BB383	gagctgaagattgctaatctcggactggcac	cdk-12 kinase-dead QC D462N fwd
BB384	gtgccagtccgagattagcaatcttcagctc	cdk-12 kinase-dead QC D462N rev
BB385	gaatactcaaacttgccaattttggactagctcgg	cdk-9 kinase-dead QC D235N fwd
BB386	ccgagctagtccaaaattggcaagtttgagtattc	cdk-9 kinase-dead QC D235N rev
BB392	gactCGATCGGCCCGGGCgtcaccacgtgtggctattg	cdk-9 3' UTR rev SrfI PvuI
BB408	gactGCGGCCGCtaggttgtatgttaccacac	mex-5 3' UTR fwd NotI
BB410	gactGCGGCCGCataagtactcatctacttacaaag	pal-1 3' UTR fwd NotI
BB414	gactCTCGAGattccataaaaaaccatccg	mex-5 3' UTR rev XhoI
BB415	gactCTCGAGtggatagttaatctcatc	pal-1 3' UTR rev XhoI
BB424	gacTACGTAattccataaaaaaccatccg	mex-5 3' UTR rev SnaBI
BB425	gactTACGTAtggatagttaatctcatc	pal-1 3' UTR rev SnaBI
BB451	tccttctcccatagtcgtgc	cdk-12(ok3664) deletion confirmation fwd
BB452	tgacattcattttgccgcta	cdk-12(ok3664) deletion confirmation rev

Citations

1. H. Furuhashi *et al.*, Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics & chromatin* **3**, 15 PHST (2010).
2. E. A. Bowman, C. R. Bowman, W. G. Kelly, P-TEFb-independent RNA polymerase II phosphorylation and germline-soma distinction in C. elegans. *Development*, (2013).
3. S. Egloff, S. Murphy, Cracking the RNA polymerase II CTD code. *Trends Genet* **24**, 280 (2008).
4. Q. Zhou, T. Li, D. H. Price, RNA polymerase II elongation control. *Annu Rev Biochem* **81**, 119 (2012).
5. H. Cho *et al.*, A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev* **13**, 1540 (1999).
6. A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, P. Cramer, A structural perspective of CTD function. *Genes Dev* **19**, 1401 (2005).
7. Y. Yamaguchi, H. Shibata, H. Handa, Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. *Biochim Biophys Acta*, (2012).
8. T. Narita *et al.*, Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol* **23**, 1863 (2003).
9. H. Qiu, C. Hu, A. G. Hinnebusch, Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* **33**, 752 (2009).
10. A. Wood, A. Shilatifard, Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. *Cell Cycle* **5**, 1066 (2006).
11. B. Bartkowiak *et al.*, CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**, 2303 (2010).
12. B. Bartkowiak, A. L. Greenleaf, Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? *Transcription* **2**, 115 (2011).
13. H. Furuhashi *et al.*, Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics Chromatin* **3**, 15 (2010).
14. R. S. Kamath, J. Ahringer, Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**, 313 (2003).
15. G. Seydoux, M. A. Dunn, Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development* **124**, 2191 (1997).
16. R. D. Chapman *et al.*, Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780 (2007).
17. I. Kawasaki *et al.*, PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in C. elegans. *Cell* **94**, 635 (1998).
18. C. E. Schaner, G. Deshpande, P. D. Schedl, W. G. Kelly, A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Dev Cell* **5**, 747 (2003).
19. H. Kimura, Y. Hayashi-Takanaka, Y. Goto, N. Takizawa, N. Nozaki, The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct Funct* **33**, 61 (2008).
20. S. Strome, W. B. Wood, Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. *Cell* **35**, 15 (1983).
21. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676 (2012).
22. S. J. H. Ashcroft, C. Pereira, *Practical Statistics for the Biological Sciences: Simple Pathways to Statistical Analyses*. (Palgrave, 2003).
23. A. K. Walker, Y. Shi, T. K. Blackwell, An extensive requirement for transcription factor IID-specific TAF-1 in Caenorhabditis elegans embryonic transcription. *J Biol Chem* **279**, 15339 (2004).
24. C. Frokjaer-Jensen *et al.*, Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet* **40**, 1375 (2008).

25. E. Y. Shim, A. K. Walker, Y. Shi, T. K. Blackwell, CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. *Genes Dev* **16**, 2135 (2002).
26. J. A. Powell-Coffman, J. Knight, W. B. Wood, Onset of C. elegans gastrulation is blocked by inhibition of embryonic transcription with an RNA polymerase antisense RNA. *Dev Biol* **178**, 472 (1996).
27. D. Blazek *et al.*, The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**, 2158 (2011).
28. J. Kohoutek, D. Blazek, Cyclin K goes with Cdk12 and Cdk13. *Cell Div* **7**, 12 (2012).
29. K. Glover-Cutter *et al.*, TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol* **29**, 5455 (2009).
30. A. K. Walker, P. R. Boag, T. K. Blackwell, Transcription reactivation steps stimulated by oocyte maturation in C. elegans. *Dev Biol* **304**, 382 (2007).
31. D. Ghosh, G. Seydoux, Inhibition of transcription by the Caenorhabditis elegans germline protein PIE-1: genetic evidence for distinct mechanisms targeting initiation and elongation. *Genetics* **178**, 235 (2008).
32. C. E. Schaner, W. G. Kelly, Germline chromatin. *WormBook*, 1 (2006).
33. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100**, 64 (1983).
34. W. G. Kelly *et al.*, X-chromosome silencing in the germline of C. elegans. *Development* **129**, 479 (2002).
35. M. S. Swanson, F. Winston, SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics* **132**, 325 (1992).
36. Y. Yamaguchi *et al.*, Structure and function of the human transcription elongation factor DSIF. *J Biol Chem* **274**, 8085 (1999).
37. C. Merritt, D. Rasoloson, D. Ko, G. Seydoux, 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Curr Biol* **18**, 1476 (2008).
38. K. O. Kizer *et al.*, A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol* **25**, 3305 (2005).
39. L. B. Bender *et al.*, MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. *Development* **133**, 3907 (2006).
40. A. Rechtsteiner *et al.*, The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genet* **6**, (2010).
41. Y. Chu, A. Sutton, R. Sternglanz, G. Prelich, The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *Mol Cell Biol* **26**, 3029 (2006).
42. N. J. Fuda *et al.*, Fcp1 dephosphorylation of the RNA polymerase II C-terminal domain is required for efficient transcription of heat shock genes. *Mol Cell Biol* **32**, 3428 (2012).
43. A. Nakamura, G. Seydoux, Less is more: specification of the germline by transcriptional repression. *Development* **135**, 3817 (2008).
44. F. Zhang, M. Barboric, T. K. Blackwell, B. M. Peterlin, A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. *Genes Dev* **17**, 748 (2003).
45. J. Pirngruber, A. Shchebet, S. A. Johnsen, Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell Cycle* **8**, 3636 (2009).
46. S. Venkatesh *et al.*, Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452 (2012).
47. S. Buratowski, T. Kim, The role of cotranscriptional histone methylations. *Cold Spring Harb Symp Quant Biol* **75**, 95 (2011).
48. J. M. Lee, A. L. Greenleaf, CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. *Gene Expr* **1**, 149 (1991).
49. R. Freter, M. Osawa, S. Nishikawa, Adult stem cells exhibit global suppression of RNA polymerase II serine-2 phosphorylation. *Stem Cells* **28**, 1571 (2010).
50. L. R. Baugh, J. Demodena, P. W. Sternberg, RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324**, 92 (2009).

51. L. J. Gaydos, A. Rechtsteiner, T. A. Egelhofer, C. R. Carroll, S. Strome, Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell Rep* **2**, 1169 (2012).
52. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720 (2012).
53. T. Blumenthal, Trans-splicing and operons in C. elegans. *WormBook*, 1 (2012).
54. M. A. Allen, L. W. Hillier, R. H. Waterston, T. Blumenthal, A global analysis of C. elegans trans-splicing. *Genome Res* **21**, 255 (2011).
55. V. Reinke, A. D. Cutter, Germline expression influences operon organization in the Caenorhabditis elegans genome. *Genetics* **181**, 1219 (2009).

CHAPTER 5: TOWARDS THE PURIFICATION OF C. ELEGANS PRIMORDIAL GERM CELL NUCLEI BY INTACT FOR GENOMIC ANALYSIS

Bowman, E.A.; Kelly, W.G.[5]

In order to move towards genomic and biochemical characterization of tissue-specific transcriptional regulation in *C. elegans*, I initiated the application of a previously developed method to purify embryonic germ cell nuclei. This method, called isolation of nuclei tagged in specific cell types, or INTACT, was previously developed in *Arabidopsis* for the purification of specific cells from root tissue (*1*). I have developed the tools required for this project, but optimization of nuclei purification is still in process. While this work is unpublished, the reagents have been distributed to several labs in order to allow others to use the technology developed thus far.

[5] E.A.B. wrote the manuscript and conducted all experiments. E.A.B. and W.G.K. conceived the study. W.G.K. supervised and provided funding for the study.

Introduction

One of the challenges of studying the development of metazoan model systems such as *Arabidopsis*, *C. elegans*, and *Drosophila*, is their multicellularity. While many studies have exploited the apparent homogeneity of cell culture systems to understand molecular details of biology through biochemical and genomic techniques, it is desirable to move toward utilizing these methods in tissues isolated from whole animal models. Isolation of these tissues is challenging for multicellular organisms where the tissue is either too small to reproducibly isolate or a tissue is made of many cell types that are impossible to separate by dissection. Methods such as fluorescence activated cell sorting of cell types or even microdissection have proven to be useful; however, these assays require special expertise and equipment and still may not have the sensitivity to isolate cell types existing at a low frequency in a population.

To combat some of the problems, a new technique was developed in *Arabidopsis* to isolate cell-specific nuclei for genome-wide assays (*1*). This method is based on selectively tagging the outer nuclear membrane in specific cells with biotin in a specific cell type. The tagged nuclei are then separated from others by exploiting the high affinity biotin-avidin interaction. Specifically, a biotin ligase recognition peptide (BLRP) tagged nuclear envelope protein is expressed in the cell of interest using a cell-specific promoter. The BLRP is then biotinylated by an exogenously expressed *E. coli* biotin ligase, BirA. Nuclei are isolated and biotinylated nuclei are separated using avidin-coated magnetic beads. The beauty of this method, called isolation of nuclei tagged in specific cell types, or INTACT, is that it uses methods well established across different model organisms, transgenics and affinity-isolation, with no specialized equipment needed.

Perhaps the most important cell lineage in any multicellular organism is the germline. While germ cells become highly differentiated into haploid gametes, they are the only ones that contribute to the next generation and also uniquely retain the ability to transform into the

totipotent zygote. Because of the ease of genetic manipulation and short generation time, *C. elegans* is an attractive model for understanding germline development.

Specification of the germline in *C. elegans* begins at the first embryonic cell division. The one-cell embryo, P_0, divides asymmetrically to produce a somatic blastomere, AB, and germline blastomere, P_1. This asymmetric division of P cells into somatic blastomeres and germline-destined P cells continues until the last germ cell precursor, P_4, divides symmetrically to produce the first germline-restricted cells, or primoridial germ cells (PGCs), Z2 and Z3 (*2*). In the P lineage, the pluripotent differentiation potential is maintained through genome-wide transcriptional inhibition through the protein PIE-1, which prevents Pol II elongation (*3*, *4*). Once P_4 divides to produce Z2/Z3, PIE-1 is marked for degradation by the ubiquitin ligase, ZIF-1 (*5*), and a canonical marker of transcriptional elongation, phosphorylation of serine 2 of the RNA polymerase II CTD (Ser2-P) appears (*6*). The specific mechanism of PIE-1-mediated transcriptional repression is unique to *C. elegans*; however transcription inhibition is a general phenomenon of germ line specification (*7*).

While Z2 and Z3 have high levels of Pol II CTD Ser2 phosphorylation, changes in chromatin modifications suggest that Z2/Z3 may not be highly transcriptionally active. The active marks H3K4me3, H3K4me2, and H3K18Ac are either absent prior to, or lost just after, the birth of Z2/Z3 (*8*). Another transcription-associated modification, H3K36me3, is dependent on a unique, transcription-independent methyltransferase (*9*). The repressive mark H3K27me3 also increases in Z2/Z3 (*10*). In addition, while the phosphorylated Pol II CTD Ser2 epitope is present at the birth of Z2/Z3, this signal is transient and decreases in Z2/Z3 throughout the rest of embryogenesis, suggesting that broad transcription may not be ongoing (*9*). Finally, Z2/Z3 arrest in prophase and do not divide again until after hatching, suggesting that there might not be a need for global transcription in these cells during embryonic development. While this data suggests that there may not be active transcription in Z2/Z3, a global analysis of transcription is required for these cells.

Conventional assays for global transcription such as oligo(T) in situ hybridization or BrUTP incorporation assays are not feasible in Z2/Z3 due to the presence of maternally loaded transcripts (*11*) and transcription of rRNA molecules in these cells (*6*). Thus, transcriptional activity must be assayed by an alternative method. Analyses of Pol II distribution in the genome can answer two very important questions that are critical to understanding the role of Pol II in PGCs: it can suggest the transcriptional competency of Z2/Z3, and it can act as a "transcription screen", providing evidence for the genes that may be important for PGC pluripotency and development.

In order to obtain a high resolution, genome-wide understanding of transcription in the *C. elegans* embryonic germline precursor cells, P_4/Z2/Z3, I attempted to develop the INTACT method for use in the *C. elegans* germline cells. Specifically, I optimized expression of a BLRP-tagged nuclei envelop protein specifically in P_4/Z2/Z3 within embryos and BirA within the germline and ubiquitously in the early embryo. I hope this will provide the beginnings of a method to isolate Z2/Z3 in high enough quantities for genome-wide analysis of RNA polymerase II localization.

Methods

Strains used

Strains used in this study included Bristol N2; EG6699 *ttTi5605 II, unc-119(ed3) III*; EG6701 *ttTi4348 I; unc-119(ed3) III*; and other transgenic lines generated in this study (Table 5.1).

Generation of transgenic lines

Standard cloning methods were used to generate transgenic constructs (Table 5.2, 5.3). For the *npp-9*:BLRP construct, the *pie-1* promoter, *npp-9* ORF, and *nos-2* 3' UTR were amplified from N2 genomic DNA with the indicated primers (Table 5.3) and ligated into a plasmid containing the *unc-119* 5'UTR, cDNA and 3'UTR. A tag consisting of mCherry, BLRP, and FLAG, called CBF,

was amplified by the following PCR reactions [1) mCherry_F_C9 and mChBio_C_Ri on mCherry containing plasmid; 2) mCherry_F_C9 and mChBio_C_Ro9 on 1st PCR; 3) mCherry_F_C9 and FLAG_Ri on 2nd PCR; 4) mCherry_F_C9 and FLAG_Ro on 3rd PCR] and ligated to create pFS26.

In order to add a PEST sequence to the end of the *npp-9:BLRP* transgene sequence using the SbfI site, an internal SbfI site was removed by subcloning a region into the pCR-Blunt II-TOPO Zero Blunt Cloning Plasmid (Invitrogen, Grand Island, NY, USA) and using the Quik Change site directed mutagenesis kit (Aglient, Santa Clara, CA, USA) with indicated primers (Table 5.2, 5.3) to create a silent mutation. The NPP-9:BLRP was subcloned into the MOS-SCI expression vector, pCFJ151. A PEST sequence previously shown to be effective in *C. elegans* (*12*) and point mutants previously shown to reduce destabilizing activity in other systems (*13, 14*) were created from long primers and then cloned onto the end of the *npp-9* construct.

The BirA construct was made by replacing the promoter and 3' UTR of pSO221 (*15*) with *pie-1* and *nos-2* sequences respectively (Tables 5.2, 5.3) and was subcloned into the MOS-SCI expression vector, pCFJ352. A synthetic, *C. elegans* codon optimized BirA, called ceBirA, with introns was synthetically created (Clontech) and subcloned into this vector. The remaining cloning was performed using conventional cloning methods as indicated (Tables 5.2, 5.3).

Transgenic lines were made via standard MOS-SCI methods (*16*). Insertion was confirmed by PCR and representative lines are indicated (Table 5.1).

Visualization of transgenic animals and immunofluorescence

GFP immunofluorescence was performed with methanol/acetone fixation (*8*): anti-GFP (Milipore MAB3580, 1/200 dilution) and anti-P-granules [OIC1D4, 1/4 dilution (*17*)]. Fixed animals were incubated with primary antibodies 12-16 hours at 4°C. Secondary antibodies included Alexa Fluor 594-and Alexa Fluor 488-conjugated donkey antibodies (1:500) (Molecular Probes, Eugene, OR, USA).

Images of immunofluorescence data or native GFP and mCherry were taken on a Leica DMRXA (Hamamatsu Photonics, Hamamatsu, Japan) microscope using Simple PCI software (Hamamatsu Photonics).

Results

Expression of NPP-9:BLRP in embryonic germline precursors

The INTACT method was initially adapted to *C. elegans* to isolate muscle nuclei (*18*). In order to expand this method to isolate embryonic germline precursor nuclei, I need to express the nuclei tagging construct, NPP-9:BLRP, in these nuclei. To do this, I constructed this transgene with the maternal germline-expressing *pie-1* promoter and the *nos-2 3' UTR*, which has been reported to prevent translation of the protein until the Z2/Z3 mother, P_4, is born. As with other studies (*19, 20*), I detected transgene expression in both the maternal germline and P_4/Z2/Z3. Unfortunately, I also detected high levels of the transgenic protein maternally loaded into embryos which can be seen ubiquitously in somatic embryonic nuclei until the ~200 cell stage, long after P_4/Z2/Z3 are born (Figure 5.1). I thought that this contaminating expression, not previously reported with transgenic lines with the *nos-2 3' UTR*, was likely due to the stability of nuclear envelope-associated proteins, and the loss of the signal in mid embryogenesis was likely due to dilution of the maternal fusion protein rather than turnover.

In order to try to eliminate this contaminating somatic signal, I attempted to reduce the stability of the fusion protein by adding a destabilizing PEST signal to the 3' end of the construct that had been optimized to reduce the half life of a fusion protein in *C. elegans* (*12*). Because of the strength of this PEST sequence in destabilizing a protein, and because I am utilizing a low copy expression system, I also tested several mutants of the PEST sequence that have been shown to reduce the effectiveness of this destabilizing sequence (*13, 14*). After testing several mutants, I identified one that was effective at reducing the expression of this fusion protein within the

maternal germline. With this PEST sequence (mPEST), NPP-9:BLRP is expressed in the distal mitotic germ cells but protein levels decrease toward the proximal end of the gonad where the oocytes are produced(Figure 5.1). NPP-9:BLRP:mPEST is undetectable in 1-2 cell embryos but is highly expressed in P_4/Z2/Z3 (Figure 5.1). I was therefore able to produce what appears to be the ideal expression of the nuclear tagging protein for isolation of P_4/Z2/Z3.

Expression of BirA in the germline

As mentioned above, the INTACT method developed in *C. elegans* was aimed at isolating muscle nuclei. In order use this method for isolation of P_4/Z2/Z3, I exchanged the muscle-specific BirA promoter for the germline-specific *pie-1* promoter, and added the *nos-2 3' UTR*. Unfortunately, this construct did not have any detectible expression. The lack of germline BirA expression could be due to RNAi-mediated silencing, as often happens with multi-copy transgenes (*21*). Typically when this occurs, germline expression can be detected in the first few generations after construction of the transgenic line but then reduces shortly after. RNAi-mediated silencing of BirA seemed unlikely for two reasons: first, I was never able to detect BirA:GFP expression, even in early generations, and second, I could not detect expression even with a single copy transgene. However, I wanted to make sure that lack of expression was not due to germline silencing by knocking down the RNAi machinery. Previously, RNAi directed against the RNAi gene *rde-3* was shown to reverse transgene silencing (C. Mello, personal communication). While I could restore expression of a different transgenic line that had been silenced with *rde-3 (RNAi)*, I could never detect BirA:GFP expression with *rde-3 (RNAi)* (data not shown). Thus, I determined that lack of expression was likely due to the some sequence-specific aspect of the BirA transgene.

I determined that the lack of expression was specifically due to the BirA coding region because a construct lacking this region of the transgene had germline GFP expression (data not shown). This BirA protein was taken directly from *E. coli*, which does not have introns or codons

optimized for *C. elegans* expression, which may prevent expression in *C. elegans*. I reasoned that this was not a problem in the gut because there is tighter regulation of expression of exogenous genes in the germline (*21*).

To optimize germline expression of BirA, I synthetically constructed a BirA construct with artificial introns and codons optimized for *C. elegans* expression, called ceBirA (*22*). This transgene successfully expresses in the maternal germline, is maternally loaded into embryos and detectible until the ~300 cell stage, and is highly expressed in P_4/Z2/Z3 (Figure 5.2). Combined with the NPP-9:BLRP, this provided the proper expression of the nuclear tagging system for nuclei isolation

Discussion

As genomics analysis is becoming more straightforward and possible on smaller sample sizes, the *in vivo* binding of factors in specific cells is becoming more feasible. The current challenge for this analysis is the purification of specific cells from organisms or tissues with multiple cell types. This indeed is a challenge in tissues that are refractory to previous methods of cell purification such as FACS or microdissection. The *C. elegans* embryo is a unique challenge because it has a chitin outer membrane that makes it challenging to obtain large numbers of dissociated cells that are not damaged when the egg shell is broken. Thus, the INTACT method of cell-specific nuclei purification is an attractive approach as isolation of nuclei from *C. elegans* embryos is likely to be more straightforward than isolation of intact cells. In addition, utilization of the avidin-biotin system for purification would likely increase purity and yield compared to traditional FACS sorting.

Thus far, I have constructed both the ceBirA and NPP-9:BLRP transgenic lines that have proper expression for this method. At this point, I are optimizing the nuclei purification step; however, once this is established, it is likely that high numbers of *C. elegans* PGCs will be purified for high resolution genome-wide analysis of transcription-associated factors.

As an alternative approach to identifying the genome-wide location of Pol II in *C. elegans* PGCs, I have also begun an alternative approach to pull down Pol II specifically from P_4/Z2/Z3. This approach includes expressing a BLRP-tagged Pol II subunit, RPB-3, specifically in P_4/Z2/Z3, and ceBirA in embryonic nuclei. Thus far, I have successfully prepared the constructs for this approach (Table 5.2, 5.3), and are in the process of creating the transgenic lines.

Acknowledgments

I would like to thank Dr. Sujata Bhattacharyya for pointing out the INTACT method and identifying it as a potential method to isolate *C. elegans* embryonic PGCs. I would also like to thank Dr. Florian Steiner and Dr. Steve Henikoff for reagents used in this study and Dr. Steve L'Hernault and Dr. Andy Fire for their advice in improving the expression of BirA in the germline and help in designing this construct, respectively.

Tables

Table 5.1: Transgenic lines

Name	Description	Genotype	Plasmid
KW2152	npp-9:CBF:PEST	ckSi7 (unc-119, Ppie-1 npp-9 CBF PEST nos-2 3' UTR) II; unc-119(ed3) III	pBAB67
KW2127	npp-9:CBF	ckSi8 (unc-119, Ppie-1 npp-9 CBF nos-2 3' UTR) II; unc-119(ed3) III	pBAB70
KW2123	ceBirA:GFP	ckSi11 (unc-119, Ppie-1 ceBirA GFP npp-9 3' UTR) I; unc-119(ed3) III	pBAB78
KW2173	npp-9:CBF:mPEST	ckSi19 (unc-119, Ppie-1 npp-9 CBF mPEST nos-2 3' UTR) II; unc-119(ed3) III	pBAB100
KW2186	npp-9:CBF:msPEST short	ckSi22 (unc-119, Ppie-1 npp-9 CBF msPEST nos-2 3' UTR) II; unc-119(ed3) III	pBAB105
KW2193	npp-9:CBF:mPEST; ceBirA:GFP	ckSi11 (unc-119, Ppie-1 ceBirA GFP npp-9 3' UTR) I; ckSi19 (unc-119, Ppie-1 npp-9 CBF mPEST nos-2 3' UTR) II; unc-119(ed3) III	pBAB78, pBAB100

Table 5.2: Cloning Methods

Name	Parent vector	Description	Made from plasmid	Cloning method	Primers
pFS19		Ppie-1 BirA GFP npp-9 3' UTR	pSO221	AscI, FseI ligation with PCR prods: Ppie-1, npp-9 3' UTR	Ppie-1_F, Ppie-1_R, npp-9UTR_F, npp-9UTR_R
pFS26		Ppie-1 npp-9 mCherry FLAG BLRP nos-2 3' UTR		ligation with PCR prods: Ppie-1, npp-9 ORF, BLRP, FLAG, BLRP, nos-2 3' UTR	Ppie-1_F, Ppie-1_R, npp-9_F, npp-9_R, mChBio_C_Ri, mChBio_C_Ro9, mCherry_F_C9, FLAG_RI, FLAG_RO, nos-2_3'_F, nos-2_3'_R
pBAB20	pCR blunt	mCherry BLRP FLAG nos-2 3' UTR		blunt ligation	BB297, BB298
pBAB24	pCR blunt	mCherry BLRP FLAG nos-2 3' UTR QC SbfI	pBAB20	QC to remove SbfI from mCherry	BB 233, BB234
pBAB42	pFS26	Ppie-1 npp-9 CBF no stop nos-2 3' UTR	pBAB24, pFS26	MscI, XmaI ligation	
pBAB63	pFS26	Ppie-1 npp-9 CBF PEST nos-2 3' UTR	pBAB42	SbfI, XmaI ligation using PCR prod: PEST	BB297, BB244
pBAB67	pCFJ151	Ppie-1 npp-9 CBF PEST nos-2 3' UTR	pBAB63, pBAB70	NheI, NotI ligation	
pBAB70	pCFJ151	Ppie-1 npp-9 CBF nos-2 3' UTR	pFS26	SpeI, ApaI cut and blunt ligation to pCFJ151	
pBAB73	pUC57	ceBirA w/ introns		BirA synthetically created (Genescript) with introns and codon optimized, named ceBirA	
pBAB74	pFS19	Ppie-1 ceBirA GFP npp-9 3' UTR	pFS19, pBAB73	FseI, SbfI ligation	
pBAB78	pCFJ352	Ppie-1 ceBirA GFP npp-9 3' UTR	pBAB74	SpeI, ApaI cut and blunt ligation to pCFJ151	
pBAB94	pFS26	Ppie-1 npp-9 CBF msPEST nos-2 3' UTR	pBAB63	XmaI, SbfI ligation using PCR prod: PEST mutant A428E A430E A431E T436A and minus last 5 AA	BB407, BB406
pBAB95	pFS26	Ppie-1 npp-9 CBF mPEST nos-2 3' UTR	pBAB63	XmaI, SbfI ligation using PCR prod: PEST mutant A428E A430E A431E T436A	BB407, BB405
pBAB100	pCFJ151	Ppie-1 npp-9 CBF mPEST nos-2 3' UTR	pBAB70+ pBAB95	NheI, NotI ligation	
pBAB105	pCFJ151	Ppie-1 npp-9 CBF msPEST nos-2 3' UTR	pBAB70+ pBAB94	NheI, NotI ligation	
pBAB107	pFS26	Ppie-1 npp-9 CBF m2PEST nos-2 3' UTR	pBAB63	XmaI, SbfI ligation using PCR prod: PEST mutant2 A428E A430E A431E	BB438, BB405
pBAB113	pCFJ151	Ppie-1 npp-9 CBF m2PEST nos-2 3' UTR	pBAB70+ pBAB107	NheI, NotI ligation	
pBAB119	pFS19	Ppie-1 ceBirA NLS GFP	pFS19	SbfI, XmaI ligation with PCR prod: NLS GFP	BB481, BB482
pBAB120	pCFJ352	Ppie-1 NLS ceBirA	pBAB119 +pBAB78	SbfI ApaI ligation	
pBAB121	pCFJ151	Prpb3 rpb3 FLAG BLRP	pBAB70	AvrII, BsrGI ligation using PCR prod: Prpb3 rpb3 ORF	BB483, BB485
pBAB122	pCFJ152	Prpb3 rpb3 FLAG BLRP mPEST	pBAB100	AvrII, BsrGI ligation using PCR prod: Prpb3 rpb3 ORF	BB483, BB485
pBAB123	pCFJ153	Ppie-1 rpb3 FLAG BLRP mPEST	pBAB100	FseI, BsrGI ligation using PCR prod: rpb3 ORF	BB484, BB485
pBAB124	pCFJ154	Ppie-1 rpb3 FLAG BLRP	pBAB70	FseI, BsrGI ligation using PCR prod: rpb3 ORF	BB484, BB485
pBAB125	pFS26	Ppie-1 npp-9 CBF rpb3 3' UTR	pFS26	XmaI, ApaI ligation using PCR prod: rpb3 3' UTR	BB486, BB487
pBAB126	pCFJ156	Prpb3 rpb3 FLAG BLRP rpb3 3' UTR	pBAB121	NotI, NheI ligation using PCR prod: rpb3 3' utr+ addional vector components	BB488, BB489
pBAB127	pCFJ157	Prpb3 rpb3 FLAG BLRP mPEST rpb3 3' UTR	pBAB122	NotI, NheI ligation using PCR prod: rpb3 3' utr+ addional vector components	BB488, BB489

Table 5.3: Primers

Name	Sequence	Description
Ppie-1_F	ATAggcgcgccAGATCTCTAAAAGTTACATAAAATTG	Ppie-1 5' AscI
Ppie-1_R	TATggccggccCTGGAAAAGAAAATTTGATTTTTAATTG	Ppie-1 3' FseI
npp-9UTR_F	ATAcccgggctgcagactactcttaagTCCTCTGCTGATTATTTAAATTATTTTATTA	npp-9 3' UTR 5' XmaI
npp-9UTR_R	ATAgggcccTAAATCTCATTCTTCAATGCATTCAC	npp-9 3' UTR 3' ApaI
mChBio_C_Ri	TCTTCTGAGAGTCGAGGATCTGACGAAGAGAAGAAGCCATcttgtacagc tcgtccatgc	CBF construction
mChBio_C_Ro9	taCCCGGGttaAGATCCTCCAGCGTTAGAACGCCACTCCATCTTCTGAG AGTCGAGGATC	CBF construction
mCherry_F_C9	ataATGCATaatggtgagcaagggcgagg	CBF construction
FLAG_RI	TGTAATCGCCATCGTGATCCTTGTAATCTGCGGCCGCAGATCCTCCA GCGTTAGAAC	CBF construction
FLAG_RO	TATcccgggTTACTTATCATCATCATCCTTGTAATCGATATCGTGATCCTT GTAATCGCCATCGTGATCC	CBF construction
npp-9_F	ATAggccggccATGAGCGATCAGAAGCCGGTAAG	npp-9 ORF 5' FseI
npp-9_R	ATAcctgcaggCATCTTCGATTGAGCATCCTGAAC	npp-9 ORF 3' SbfI
nos-2UTR_F	ATAcccgggTAGAAGATCCAATTTCTCAATAC	nos-2 3' UTR 5' XmaI
nos-2UTR_R	ATAgggcccTCATCCTATCCTGGACTGG	nos-2 3' UTR 3' ApaI
BB233	CCAGGACTCCTCACTGCAGGACGGC	mCherry SbfI QC F
BB234	GCCGTCCTGCAGTGAGGAGTCCTGG	mCherry SbfI QC R
BB244	gatccccgggCTACACATTGATCCTAGCAGAAGC	GFP PEST 3' STOP and XmaI
BB297	gatcCCTGCAGGgATGGATGAACTATACAAAC	PEST 5' +1 SbfI
BB298	gactGCTAGCatgtcctctgacgaaagtgatgc	spt-5 ORF 5' NheI
BB405	gatcCCCGGGCTACACATTGATCCTAGCAGAAGCACAGGCTGCAGGG TGACGGTCCATCCCGCTCTCCTGGGCACAAGACATGGGCAGCG	PEST 3' with XmaI
BB406	gatcCCCGGGCTAAGAAGCACAGGCTGCAGGGTGACGGTCCATCCCG CTCTCCTGGGCACAAGACATGGGCAGCG	PEST 3' -last 5 aa with XmaI
BB407	gatcCCTGCAGGgATGGATGAACTATACAAACTTAGCCATGGCTTCCC GCCGGAGGTGGAGGAGCAGGATGATGGCGCGCTGCCCATGTC	PEST 5' A428E A430E A431E T436A with SbfI
BB438	gatcCCTGCAGGgATGGATGAACTATACAAACTTAGCCATGGCTTCCC GCCGGAGGTGGAGGAGCAGGATGATGGCACGCTGCCCATGTC	PEST 5' A428E A430E A431E with SbfI
BB481	gatcCCTGCAGGgaaaaaatgactgctccaaag	gfp+NLS 5' SbfI
BB482	gatcCCCGGGTTAgccatgtgtaatcccagcagc	gfp+stop 3' XmaI
BB483	gatcCCTAGGctgaaaaattttaaaatttgaaataac	rpb3 promoter AvrII
BB484	gatcGGCCGGCCatgccgtacgcaaatcaacc	rpb3 ORF FseI
BB485	gatcTGTACAcctgctgctgggcatgagcttg	rpb3 ORF3' -STOP BsrGI +1
BB486	gatcCCCGGGactatttggatctctaatttatttacg	rpb3 3' UTR XmaI
BB487	gatcGGGCCCattctgcgaaaaaaaagccg	rpb3 3' UTR ApaI
BB488	gatcGCGGCCGCAGATTACAAGGATCACGATGG	rpb3 3' utr section for subclone NotI
BB489	gatcGCTAGCGGGCCCattctgcgaaaaaaagc	rpb3 3' utr section for subclone NheI

Figures

Figure 5.1: Addition of a destabilizing PEST sequences causes selective expression of npp-9:BLRP in P_4/Z2/Z3.

A) mCherry expression of NPP-9:BLRP transgenic line with gonad (white), oocytes (orange), and embryos (blue) outlined. Dotted box represents region expanded with germ cells indicated with arrow. B) Anti-FLAG immunofluorescence (green) analysis of dissected gonad and embryos from the NPP-9:BLRP:mPEST transgenic line counter-stained with DAPI (red). Dotted grey line represents the path germ cells make as they progress through the gonad. White arrows indicates the two primordial germ cells, Z2/Z3, marked by anti-PGL-1 (blue).

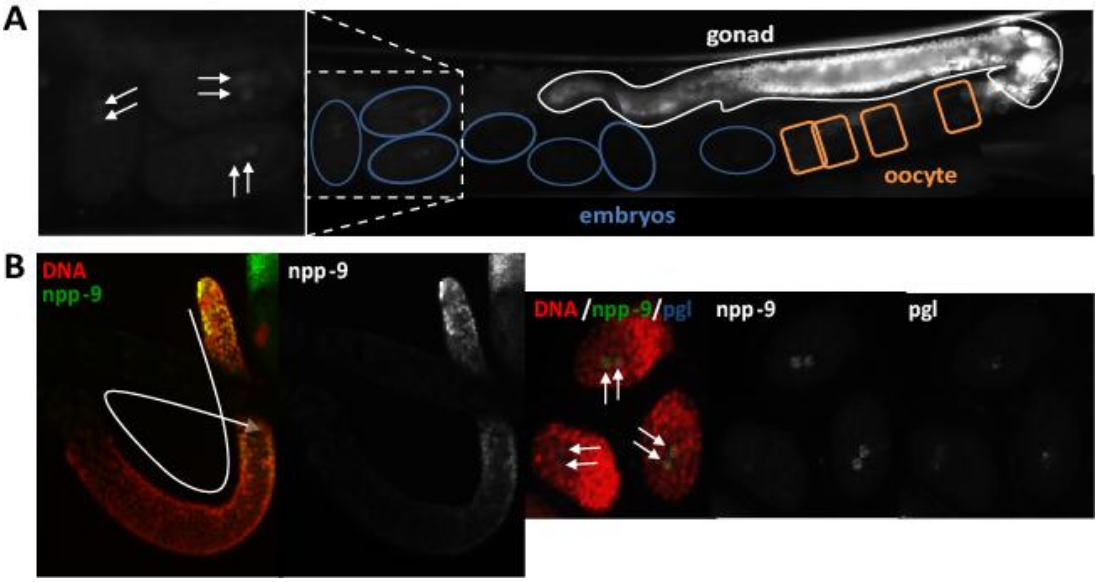
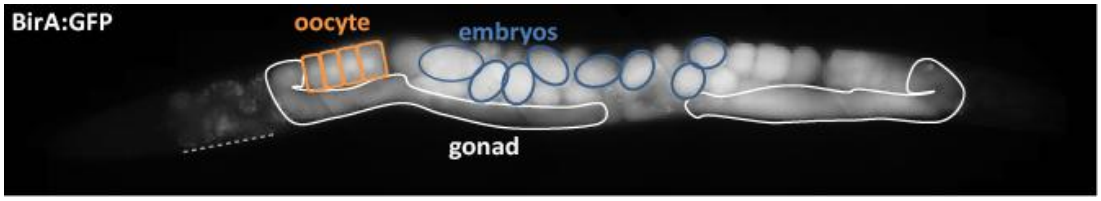


Figure 5.2: ceBirA expresses in the adult germline and is maternally loaded into embryos.

GFP expression of ceBirA:GFP transgenic line with gonad (white), oocytes (orange), and embryos (blue) outlined. Grey dotted line marks background gut autofluorescence.



Citations

1. R. B. Deal, S. Henikoff, The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. *Nature protocols* **6**, 56 (2011).
2. J. T. Wang, G. Seydoux, Germ cell specification. *Advances in experimental medicine and biology* **757**, 17 (2013).
3. F. Zhang, M. Barboric, T. K. Blackwell, B. M. Peterlin, A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. *Genes & development* **17**, 748 (2003).
4. D. Ghosh, G. Seydoux, Inhibition of transcription by the Caenorhabditis elegans germline protein PIE-1: genetic evidence for distinct mechanisms targeting initiation and elongation. *Genetics* **178**, 235 (2008).
5. P. M. Checchi, W. G. Kelly, emb-4 is a conserved gene required for efficient germline-specific chromatin remodeling during Caenorhabditis elegans embryogenesis. *Genetics* **174**, 1895 (2006).
6. G. Seydoux, M. A. Dunn, Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development (Cambridge, England)* **124**, 2191 (1997).
7. A. Nakamura, M. Shirae-Kurabayashi, K. Hanyu-Nakamura, Repression of early zygotic transcription in the germline. *Current opinion in cell biology* **22**, 709 (2010).
8. C. E. Schaner, G. Deshpande, P. D. Schedl, W. G. Kelly, A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Developmental cell* **5**, 747 (2003).
9. H. Furuhashi *et al.*, Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics & chromatin* **3**, 15 PHST (2010).
10. L. B. Bender *et al.*, MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. *Development* **133**, 3907 (2006).
11. G. Seydoux, A. Fire, Soma-germline asymmetry in the distributions of embryonic RNAs in Caenorhabditis elegans. *Development* **120**, 2823 (1994).
12. A. R. Frand, S. Russel, G. C. I. N. P. B. O. Ruvkun, P. e, Functional genomic analysis of C. elegans molting. *PLoS biology* **3**, e312 PHST (2005).
13. X. Li *et al.*, Generation of destabilized green fluorescent protein as a transcription reporter. *The Journal of biological chemistry* **273**, 34970 (1998).
14. J. Takeuchi, H. Chen, M. A. Hoyt, P. Coffino, Structural elements of the ubiquitin-independent proteasome degron of ornithine decarboxylase. *The Biochemical journal* **410**, 401 (2008).
15. S. L. Ooi, J. G. Henikoff, S. Henikoff, A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Res* **38**, e26 (2010).
16. C. Frokjaer-Jensen *et al.*, Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet* **40**, 1375 (2008).
17. S. Strome, W. B. Wood, Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. *Cell* **35**, 15 (1983).
18. F. A. Steiner, P. B. Talbert, S. Kasinathan, R. B. Deal, S. Henikoff, Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res* **22**, 766 (2012).
19. S. Jadhav, M. Rana, K. Subramaniam, Multiple maternal proteins coordinate to restrict the translation of C. elegans nanos-2 to primordial germ cells. *Development (Cambridge, England)* **135**, 1803 (2008).
20. I. D'Agostino, C. Merritt, P. L. Chen, G. Seydoux, K. Subramaniam, Translational repression restricts expression of the C. elegans Nanos homolog NOS-2 to the embryonic germline. *Developmental biology* **292**, 244 (2006).
21. W. G. Kelly, S. Xu, M. K. Montgomery, A. Fire, Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. *Genetics* **146**, 227 (1997).
22. M. Stenico, A. T. Lloyd, P. M. Sharp, Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. *Nucleic acids research* **22**, 2437 (1994).

CHAPTER 6: DISCUSSION

While most work from the Kelly lab focuses on epigenetic regulation in the germline, the goal of my project was to better understand the regulation of transcription in *C. elegans*. During my thesis work, I characterized mutants of the general transcriptional machinery (Chapters 2,3), identified tissue-specific regulation of transcriptional elongation (Chapter 4), and adapted a system for purifying germ cell-specific nuclei in *C. elegans* (Chapter 5).

New tools for the analysis of gene expression in *C elegans*

Because general transcription factors are essential for development, null mutants of basal transcription machinery are lethal. Thus, utilization of conditional mutants is important for the analysis of these factors in development. Towards this goal, I identified and characterized point mutants of the two transcription factors, *ama-1* and *taf-6.2*, in hopes that a subset of these mutants could be utilized as conditional transcription mutants. While these factors did not prove useful for the remainder of my work, two of the *ama-1* mutants have already been successfully used to analyze the role of transcription in other biological processes (*1*). These mutants should prove to be useful to others for similar analysis.

Identification of germline-specific Ser2 phosphorylation in *C. elegans*

Regulation of transcriptional elongation is a well established mode of gene expression control in metazoans. Canonically, this step of transcription is accompanied by phosphorylation of Ser2 of the Pol II CTD by the P-TEFb kinase complex CDK9-Cyclin T. Early studies of P-TEFb activity in *C. elegans* suggested that phosphorylation of Ser2 occurs by a mechanism analogous to other metazoan systems, with all Ser2-P requiring P-TEFb activity (*2*).

Because of this previous result, I was surprised to find that the two PGCs, Z2/Z3, in *C. elegans* embryos retain Ser2-P following CDK-9 knock down (*3*). There could be several explanations for this. First, Z2/Z3 could be less sensitive to RNAi treatment. There have been reports that some tissues in *C. elegans* are resistant to RNAi, including late embryonic tissues,

sperm, and neurons (*4*, *5*). However, I suspected that this was not the case for Z2/Z3 as I have observed other phenotypes in these cells following RNAi (*3*, *6*, *7*). Initially I suggested that another kinase, TLK-1, was responsible for phosphorylation of Ser2 in Z2/Z3 along with CDK-9 (*3*). I now believe that loss of Ser2-P in *tlk-1 cdk-9(RNAi)* embryos is likely to be an indirect effect as I could not see an effect on Ser2-P levels after TLK-1 knock down alone (E.A. Bowman, W. Kelly, unpublished), and *tlk-1 cdk-9(RNAi)* embryos show severe cell division defects that might indirectly contribute to the loss of Ser2-P in Z2/Z3.

Thus, after the discovery that *Drosophila* and human CDK12 could phosphorylate Ser2 of the CTD, I analyzed the role of *C. elegans* CDK-12 in the PGCs. I was surprised to find cell-specific regulation of Ser2-P in the *C. elegans* embryo (Chapter 4). In embryonic somatic blastomeres, CDK-9 acts upstream of and is required for CDK-12 activity to phosphorylate Ser2. In contrast to this, Ser2-P is largely independent of CDK-9 in the embryonic germ cells, Z2/Z3, and instead is fully dependent on CDK-12 in this tissue (Figure 6.1A).

I initially suspected that CDK-9 independent Ser2-P might be specific to Z2/Z3 because of unique transcriptional activity markers in these cells. While Z2 and Z3 have high levels Ser2-P, a marker of transcriptional elongation, they have an epigenetic pattern characteristic of silent chromatin (described in the introduction). Because of these conflicting markers of transcription and unclear transcriptional activity in Z2/Z3, I reasoned that the CDK-9-independent Ser2-P in Z2/Z3 may not actually reflect active transcription. I was surprised, however, to find that this unique mode of Ser2-P regulation is not unique to these embryonic germ cells but instead occurs in transcriptionally active germ cells throughout development. Thus, somatic cells and germ cells regulate Ser2-P differently even during stages of high levels of transcription in germ cell development (Figure 6.1A).

Comparison of Ser2 phosphorylation models

While CDK12 has previously been identified as a Ser2 kinase (*8*), this work is the first example CDK9-independent Pol II Ser2-P in metazoans. CDK9 Ser2 kinase activity occurs upstream of CDK12 in metazoan systems tested. Loss of CDK9 activity generally results in loss of detectible Ser2-P while loss of CDK12 reduces Ser2-P levels by about 60% (*8, 9*). I see very similar kinase requirements in the *C. elegans* soma (compare Figure 1.1A and Figure 6.1A). Thus, I have concluded that CDK-9 acts upstream of, and is required for CDK-12 activity in the *C. elegans* soma (Figure 6.1B).

In contrast, Ser2-P appears to be fully dependent on CDK-12 in the *C. elegans* germline. This is similar to the regulation of Ser2-P in yeast. In both budding and fission yeast, the CDK12 homolog Ctk1/Lsk1 is responsible for the bulk (>90%) of Ser2-P. In addition, this phosphorylation largely does not require upstream activity of the CDK9 homolog, Bur1/Cdk9. In contrast to the *C. elegans* germline, however, the yeast Bur1/Cdk9 proteins are responsible for a small fraction (~10%) of Ser2-P upstream of Ctk1 activity. In the *C. elegans* germline, knock down of CDK-9 does not appear to significantly reduce Ser2-P levels while CDK-12 knock down results in near complete loss of Ser2-P. Thus, I have concluded that CDK-12 alone is required for detectible Ser2-P in the germline. There is a chance that CDK-9 does act upstream of CDK-12 to phosphorylate Ser2; however, it would only be responsible for the <5% of Ser2-P signal remaining following loss of CDK-12. Furthermore, unlike the soma, this activity is not required for CDK-12-mediated Ser2-P (compare Figure 1.1A and Figure 6.1A; Figure 6.1B).

While regulation of Ser2-P in the *C. elegans* soma by CDK-9 matches that seen in other metazoans, there are likely distinct differences in the mechanism of CDK-9 activity. For example, in higher eukaryotes CDK9 activity is tightly regulated by large protein complexes (described in Chapter 1) that include the inhibitory 7SK/HEXIM and stimulatory SEC complex. *C. elegans* does not contain homologs for the 7SK/HEXIM components, and thus likely does not sequester

CDK-9 in an analogous inhibitory complex. While *C. elegans* does contain homologs for a subset of the SEC components, specifically ELL1 (Y24D9A.1) and AFF9 (Y55B1BR.1 and Y55B1BR.2), they don't appear to play a major in the regulation of CDK-9 activity in *C. elegans* as knock down does not affect Ser2-P levels or viability of the embryo (E.A. Bowman, W. Kelly, unpublished). Finally, CDK9 activity is also regulated by the protein BRD4 in higher eukaryotes and there is not a BRD4 homolog in *C. elegans*.

Ser2-P is not required for germline development

One of the most interesting findings is that while CDK-12 is required for Ser2-P in the *C. elegans* germline, it is not essential for germline development under optimal growth conditions (Figure 6.1A). This was initially very surprising because 1) CDK-12 is essential for larval somatic development and 2) Ser2-P is thought to be important for transcription associated processes in metazoans. As described in the introduction, Ser2-P is thought to be important for recruiting RNA processing and chromatin modification in all eukaryotes studied to date, including yeast. However, as another parallel between yeast Ser2-P regulation in the *C. elegans* germline and yeast, the bulk Ser2 kinase in yeast, Ctk1, is not essential under optimal growth conditions. Thus, perhaps Ser2-P enhances the efficiency of transcription associated processes but is not absolutely required for them.

In contrast to this, CDK-9 is essential for *C. elegans* fertility (Figure 6.1A). This was also initially surprising as CDK-9 is not required for the process it is best known for, phosphorylation of Ser2, in the germline. While the essential CDK-9 target in the germline is unclear, the fact that it is an essential protein is again similar to the severe growth defects of Bur1/Cdk9 mutants in budding and fission yeast. The essential target of Bur1/Cdk9 in yeast is also not known; however, I suspect that the targets of CDK-9 in *C. elegans* mirror the targets of Bur1/Cdk9 in yeast. It also seems likely that CDK-9 also phosphorylates the CTR of SPT-5 in *C. elegans*, although my studies did not address this possibility.

A possible role for a 5' transcriptional checkpoint in *C. elegans*

Why is there tissue-specific Ser2-P regulation in *C. elegans*? As Ser2-P is most tightly associated with transcription elongation, the most likely reason for this difference is that transcription elongation is regulated differentially between the soma and germline. As described in the introduction, CDK9 has been implicated in the release of paused polymerase in higher metazoans. The fact that Ser2-P in *C. elegans* somatic tissues is regulated similarly to higher metazoans may suggest an analogous mechanism in these tissues. Additionally, the fact that somatic cells lacking CDK-9 retain a marker of initiated Pol II, Ser5-P, in the absence of the elongation marker, Ser2-P, supports a role for CDK-9 in mediating the transition between Pol II initiation and elongation. Finally, a previous study of CDK-9 activity in *C. elegans* embryos showed that transgenes expressing somatic developmental genes fail to be expressed in embryos lacking CDK-9 (*2*). Thus, in the soma, CDK-9 activity likely regulates an essential step of gene expression following Pol II initiation.

Is there a pausing step in transcriptional elongation in the soma but not the germline? In higher eukaryotes, pausing is largely mediated by the NELF complex; however, *C. elegans* does not have homologs of any NELF component (see Chapter 1, Table 5.1). Furthermore, genome-wide analysis has not revealed an accumulation of Pol II at the 5' end of the majority of genes, which is a well characterized marker of Pol II pausing (Kruesi, W., Core, L., Waters, C., Lis, J.T. and Meyer, B.J, personal communication). While the lack of an accumulation of Pol II at the 5' end of most genes suggests that pausing is not a common phenomenon in *C. elegans*, it is possible that a regulated step between transcription initiation and elongation could exist, but that under normal conditions, Pol II stalling here is very short lived.

Furthermore, while promoter proximal Pol II accumulation is not commonly observed in *C. elegans* tissues under normal growth conditions, an accumulation of Pol II at the 5' end of a small fraction (~200) of genes has been detected in *C. elegans* larvae that have been starved for food

(*10*). While the factors involved in this accumulation is not known, this Pol II accumulation clearly demonstrates that post-promoter proximal events can be a regulated step of transcription in *C. elegans*. In absence of a clear mechanism and because 5' accumulation is not readily detected on most genes by ChIP as it is in other systems (*11*), I will call this hypothetical regulated 5' elongation step a "5' check point" (Figure 6.1C).

Why would somatic cells have a regulated 5' check point but germ cells do not? Perhaps this is due to the unique property of germ cells to maintain the ability to regain totipotency following fertilization. While germ cells do become highly specialized, perhaps transcription is regulated in such a way to allow these cells to quickly regain totipotency. Studies in human embryonic stem cells have suggested that wide-spread transcription in these cells allow them to maintain their pluripotent state (*12*). This model suggests that cell differentiation is not necessarily mediated solely by activation of differentiation factors, but also by repression of pluripotency factors. While initially counterintuitive, perhaps the *C. elegans* soma has tight regulation of transcriptional elongation to facilitate differentiation while the *C. elegans* germline allows more permissive transcription, preventing commitment down a differentiation path. If transcription is more permissive in the *C. elegans* germline, expression of improper genes may be repressed by posttranscriptional mechanisms such as RNAi and translation repression.

If there is a 5' check point from which CDK-9 must release Pol II in *C. elegans*, the mechanism for this is not clear. The SPT-4/SPT-5 complex, DSIF, is commonly referred to as a negative regulator of elongation; however, it is not clear if this complex can mediate pausing in the absence of NELF. In support of SPT-4/SPT-5-mediated pausing in *C. elegans*, the previous study of CDK-9 in the embryo showed that a heat shock transgene, which requires CDK-9 for expression, is expressed when both CDK-9 and SPT-4/SPT-5 are knocked down in the embryo (*2*). The authors of this study suggest that this is because CDK-9 is required for release of negative elongation regulation by SPT-4/SPT-5. This however, is not a general phenomenon in the soma because they do not see expression of any developmental transgenes when both CDK-9

and SPT-4/SPT-5 are knocked down. Furthermore, if Pol II could undergo efficient elongation in the absence of both CDK-9 and SPT-4/SPT-5, it seems likely that CDK-12-mediated Ser2-P would be detected; however, this is not the case [(*2*) and E.A. Bowman, W. Kelly, unpublished]. Finally, these embryos have a very similar arrest point to embryos with CDK-9 knock down, suggesting that bulk transcription elongation is likely still absent. Expression of SPT-5 in the *C. elegans* germline also argues against a role for SPT-4/SPT-5 in a 5' checkpoint. SPT-5 is ubiquitously expressed and is also found on the transcriptionally engaged autosomes in the germline. It seems highly unlikely that SPT-4/SPT-5 alone is mediating a 5' check point in the *C. elegans* soma, but there is no such role in the germline. Thus, if CDK-9 does act to release Pol II from a 5' check point in the soma, there is likely to be a soma-specific pausing factor (SSPF) that mediates this step, which may or may not work in combination with SPT-4/SPT-5 in a mechanism analogous to NELF in higher eukaryotes (Figure 6.1C).

Conservation of germline-specific transcriptional elongation control

As described in the introduction, there are many examples of germline-specific basal transcription factors. Thus far, the factors that have been identified are components of the preinitiation complex (PIC). I have characterized germline-specific regulation of a transcription elongation associated process, phosphorylation of Ser2 of the Pol II CTD. Unlike the germline-specific expression of PIC components described in other systems, the two kinases that phosphorylate Ser2 are ubiquitously expressed. Thus, I speculate that there is expression of a novel, tissue-specific factor which mediates this differential regulation.

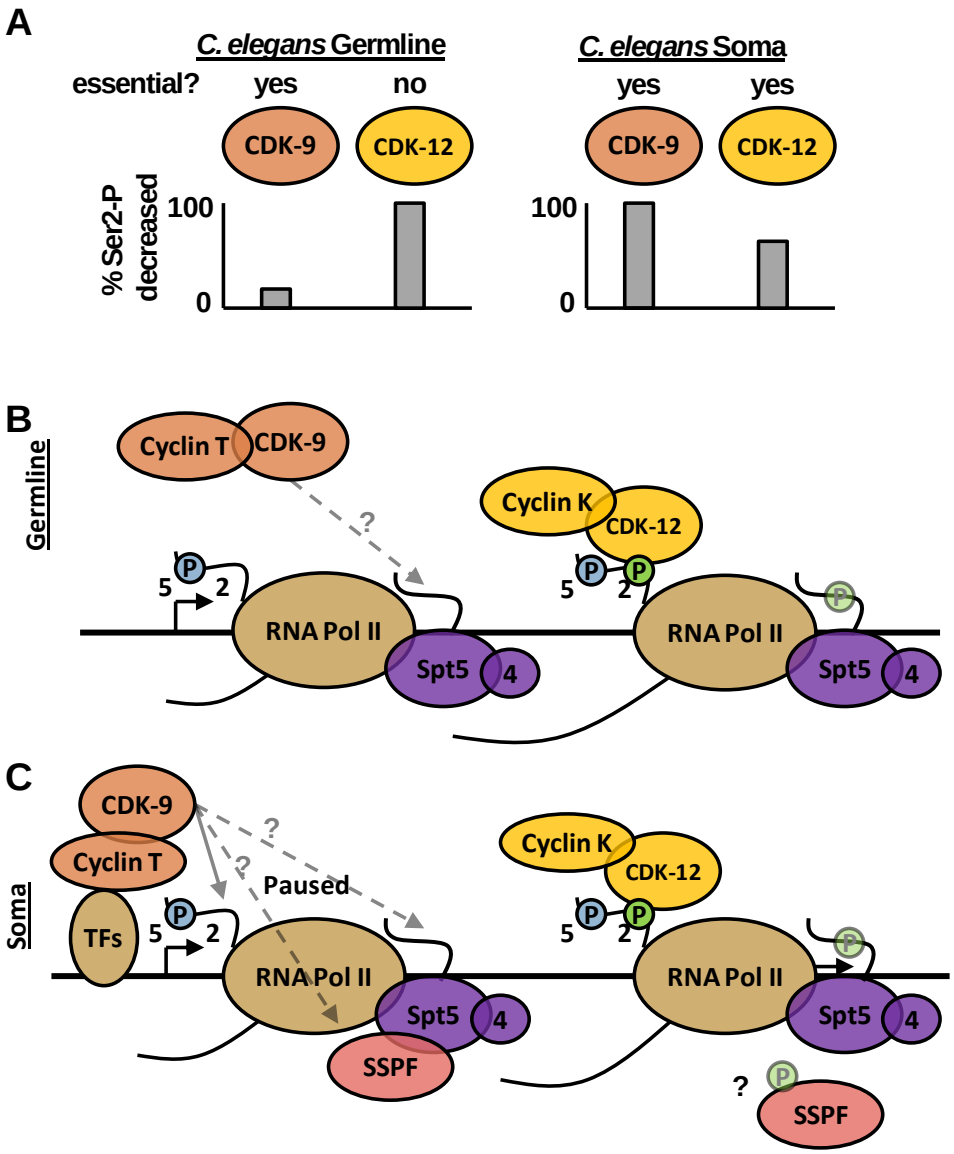
Is germline-specific regulation of transcription elongation a conserved process in eukaryotes? While the two Ser2 kinases, CDK9 and CDK12, have been shown to be capable of phosphorylating Ser2 in both *Drosophila* and mammalian systems, these analyses were performed in cell culture studies rather than in an organism (*8*). Furthermore, because the role of CDK12 as a Ser2 kinase was only fairly recently described, there has been no characterization of

tissue-specific differences in roles of these kinases *in vivo*. Thus, it is not clear whether tissue-specific kinase requirements for Ser2 phosphorylation could exist in other systems.

While the role of the Ser2 kinases has not been explored in the germline of other systems, regulation of transcription elongation is highly regulated during germline specification. As described above, there is a period of transcription elongation repression for all organisms explored to date. In *Drosophila*, the transcription elongation repressor, PGC, sequesters P-TEFb from Pol II, preventing Ser2 phosphorylation (*13*). While PGC was shown to bind CDK9, these studies were done before CDK12 was described as a Ser2 kinase and it is unclear if PGC also regulates CDK12 activity. In addition, the kinase responsible for the appearance of Ser2-P following PGC protein degradation has not been determined, although these studies did show that overexpression of CDK9 in a *pgc* heterozygous animal results in premature Ser2-P in PGCs (*13*). Thus, while CDK9 is capable of phosphorylation Ser2 under these conditions, it is not clear which is the endogenous kinase responsible for phosphorylation of Ser2 in the *Drosophila* germline. Similarly, in ascidians, the transcriptional repressor in germ cell blastomeres has been shown to bind CDK9, but any role for CDK12 in this process or the kinase responsible for the appearance of Ser2-P have not been characterized. Finally, the kinase responsible for the reappearance of Ser2-P in mammal PGCs following specification has not been analyzed.

As in *C. elegans*, early expression analysis of *cdk9* and *cdk12* transcripts suggested these factors are ubiquitously expressed in mammals (*14-19*). Interestingly, and in support of a role for CDK12-Cyclin K in Ser2 phosphorylation in the germline, both of these factors are highly abundant in mammalian sperm cells (*14*, *15*). Given the recent appreciation for the importance of transcription elongation in gene expression, it seems likely that tissue-specific regulation of transcription elongation may be a conserved mode of germline gene expression is not confined to *C. elegans*.

Figure 6.1: Summary of thesis and schematic of model



A. Loss of CDK-9 or CDK-12 activity results in different effects on Ser2-P levels in the C. elegans germline versus soma (compare to Figure 1.1A). While CDK-12 is responsible for the overwhelming bulk of Ser2-P in the germline, both CDK-9 and CDK-12 regulate Ser2-P levels in the soma. Importantly, loss of CDK-9 in the soma results in complete loss of Ser2-P, suggesting it is required upstream of CDK-12 activity. Finally, while CDK-9 and CDK-12 are essential for somatic development and CDK-9 is essential for germline development, CDK-12 is not essential

in the germline. B. As an analogy to eukaryotic organisms that do not have a NELF homolog, transcription elongation may not a regulated step of gene expression in the germline. While CDK-9 is essential in the germline, it may not directly phosphorylate Ser2 of the Pol II CTD but it may target the SPT-5 CTR. C. In the soma, the transition from Pol II initiation to elongation may be a regulated step of gene expression. Here, SPT4/SPT5 in combination with a soma-specific pausing factor (SSPF) may regulate a 5' check point prior to productive Pol II elongation. Regulated recruitment of CDK9 and Cyclin T through transcription factors and possibly phosphorylation of the SSPF may release Pol II from this check point. At the same time, CDK9 phosphorylates the Pol II CTD on Ser2 and may phosphorylate the SPT5 CTR. Downstream of this regulated step, the CDK12 complex further phosphorylates the Pol II CTD.

Citations

1. S. Rohner *et al.*, Promoter- and RNA polymerase II-dependent hsp-16 gene association with nuclear pores in Caenorhabditis elegans. *The Journal of cell biology* **200**, 589 (2013).
2. E. Y. Shim, A. K. Walker, Y. Shi, T. K. Blackwell, CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. *Genes & development* **16**, 2135 (2002).
3. H. Furuhashi *et al.*, Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics & chromatin* **3**, 15 PHST (2010).
4. R. S. Kamath, M. Martinez-Campos, P. Zipperlen, A. G. Fraser, J. Ahringer, Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome Biol* **2**, RESEARCH0002 (2001).
5. A. G. Fraser *et al.*, Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* **408**, 325 (2000).
6. P. M. Checchi, W. G. Kelly, emb-4 is a conserved gene required for efficient germline-specific chromatin remodeling during Caenorhabditis elegans embryogenesis. *Genetics* **174**, 1895 (2006).
7. C. E. Schaner, G. Deshpande, P. D. Schedl, W. G. Kelly, A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Developmental cell* **5**, 747 (2003).
8. B. Bartkowiak *et al.*, CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & development* **24**, 2303 (2010).
9. D. Blazek *et al.*, The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes & development* **25**, 2158 (2011).
10. L. R. Baugh, J. Demodena, P. W. Sternberg, RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324**, 92 (2009).
11. S. Nechaev, K. Adelman, Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et biophysica acta* **1809**, 34 (2011).
12. S. Efroni *et al.*, Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437 (2008).
13. K. Hanyu-Nakamura, H. Sonobe-Nojima, A. Tanigawa, P. Lasko, A. Nakamura, Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* **451**, 730 (2008).
14. M. C. Edwards, C. Wong, S. J. Elledge, Human cyclin K, a novel RNA polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity. *Molecular and cellular biology* **18**, 4291 (1998).
15. C. Simone *et al.*, Activation of MyoD-dependent transcription by cdk9/cyclin T2. *Oncogene* **21**, 4137 (2002).
16. J. Kohoutek *et al.*, Cyclin T2 is essential for mouse embryogenesis. *Molecular and cellular biology* **29**, 3280 (2009).
17. P. Wei, M. E. Garber, S. M. Fang, W. H. Fischer, K. A. Jones, A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451 (1998).
18. J. Peng, Y. Zhu, J. T. Milton, D. H. Price, Identification of multiple cyclin subunits of human P-TEFb. *Genes & development* **12**, 755 (1998).
19. A. De Luca *et al.*, Pattern of expression of cyclin T1 in human tissues. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **49**, 685 (2001).