DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES

DER FAKULTÄT FÜR BIOLOGIE

DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Analysis of RNA Pol II CTD phosphorylation in CDK9 analog-sensitive cells

GUOLI MA

June 2021

Completed at the Helmholtz Center Munich

German Research Center for Environment and Health (GmbH)

Institute of Functional Epigenetics

Department of Molecular Epigenetics

Date of Submission:	17.06.2021
First Examiner:	Prof. Dr. Dirk Eick
Second Examiner:	Prof. Dr. Angelika Böttger

Date of oral examination: 04.10.2021

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorliegende Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist .

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

Ich erkläre weiter, dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht**

unterzogen habe.

München, am 17. Juni 2021

Guoli Ma

Summary

The C-terminal domain (CTD) of the large subunit Rpb1 of RNA polymerase II comprises multiple tandemly repeated heptapeptides with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, with up to 52 repeats in the mammalian protein. This domain plays an essential role in the control of mRNA synthesis, as well as RNA processing, and its phosphorylation is a key feature of its function. The formation of initiation complexes on promoter DNA involves RNA polymerase II molecules with unphosphorylated CTD, whereas the CTD becomes phosphorylated during or after the transition to elongation.

The CTD is the site of multiple phosphorylations, with up to five potential phosphorylation sites in a consensus heptapeptide repeat. However, in vivo phosphorylation occurs mainly on serine residues. The hyper-phosphorylated form of the mammalian CTD has on average one phosphate/repeat, although the number of phosphates on the CTD at any point of the transcription cycle has not yet been determined.

CTD kinases and phosphatases create distinct phosphorylation patterns that change during the transcription cycle, and antibodies generated against each phosphorylated amino acid within a consensus heptad have been used to show that five different residues in the CTD heptad repeat are indeed phosphorylated in vivo. However, antibody reagents are limited in specificity and exhibit variable avidity, and it has been challenging to establish the in vivo stoichiometry of CTD phosphorylation.

I used a previously described combined genetic and mass spectrometry approach to map phosphorylation sites. Introducing several tryptic cutting sites into CTD makes it accessible to mass spectrometric analysis. By additionally replacing several amino acids, each tryptic peptide becomes unique and identifiable. This approach has shown that not all heptads are equally phosphorylated. Phosphorylation on Ser5 and Ser2 is most abundant.

In order to study the impact on CTD phosphorylation by a specific kinase, like cyclin-dependent kinase 9 (CDK9), I utilized a CDK9 analog-sensitive kinase generated by CRISPR/CAS9 and introduced it into this system.

Coupling analog-sensitive kinase mutant with stable isotope labeling and mass spectrometry, I can map and quantify heptad phosphorylation regulated by a specific kinase in sub-regions within whole CTD.

Specifically, phosphorylation of Ser5 in heptad repeat 2 and 3 was strikingly reduced 15 and 60 minutes after inhibition of CDK9. Mutation of these Ser5 residues caused prominent upregulation of a large set of interferon inducible transcripts. This suggests that specific heptads in CTD are involved in the regulation of specific gene classes.

Table of contents

1 Introduction

All cellular activities are instructed by the genomic DNA. DNA is the blueprint for the production of RNAs and proteins. To get access to this genetic information, transcription of RNA is the first step (Li & Xie, 2011).

Genomic DNA is packed into chromatin in the nucleus along with numerous regulatory macromolecules (van Steensel & Furlong, 2019). Gene expression starts with the remodeling of chromatin 3D architecture. First, transcription activators bind to specific DNA sequences, called promoter and enhancer. Next, they recruit chromatin remodelers or modifiers to open chromatin to make it accessible for RNA polymerase around other transcription factors (Andersson & Sandelin, 2020). RNA polymerases, together with transcription factors, form a complex. This complex initiates the synthesis of an RNA molecule that matches one of the two DNA strands. Finally, the newly synthesized RNA molecule is processed (Cramer, 2019c).

1.1 Discovery of eukaryotic RNA polymerases

Fifty years ago, several pioneering studies via classical biochemistry approach purified the three eukaryotic RNA polymerases: Pol I, Pol II, and Pol III. A massive number of further studies on eukaryotic transcription flourished subsequently (Lindell et al., 1970; Roeder & Rutter, 1969).

After identifying each RNA polymerase, the next step was to investigate the functions of each polymerase. The three polymerases were found to be differentially sensitive to α-amanitin, a mushroom toxin. Pol I is insensitive, Pol II is sensitive at low concentrations and can be completely inhibited, and Pol III is only partially sensitive (Kedinger et al., 1970). This information allowed several key experiments with Pol I, II, and III. RNA Pol I synthesizes the RNAs of the ribosomal subunits (rRNAs). Due to the fundamental requirement of ribosomes for protein synthesis, Pol I produces up to 60% of total RNA in eukaryotic cells (Grummt, 1998). Pol III synthesizes short, highly abundant non-coding RNAs, including all transfer RNA (tRNA), 5S rRNA, and other essential RNAs, like U6 small nuclear RNA and 7SL RNA. Genes transcribed by Pol III usually are housekeeping genes expressed in all cell types and most conditions (Lata et al., 2021). Due to its crucial role in protein synthesis, Pol III transcription is

coupled with the demand for protein synthesis (Willis, 1993). Pol II synthesizes precursors of protein-encoding mRNAs and snRNAs, microRNAs, and long non-coding RNAs (Young, 1991).

1.2 Compositions of eukaryotic RNA polymerases

RNA polymerases in prokaryotes and eukaryotes are evolutionary conserved. However, prokaryotes contain only a single type of RNA polymerase, while eukaryotes contain three different types.

Pol I is a 590kDa enzyme composed of 14 protein subunits. Pol III is the largest enzyme with 17 subunits. Pol II contains 12 subunits, Rpb1–Rpb12. The two largest subunits, Rpb1 and Rpb2, are conserved in all three RNA polymerases. The structure reveals that these two large subunits form a cleft as a backbone of the polymerase, and the smaller subunits surround the periphery. This architecture serves as an active center and allows the loading of the DNA template. The critical elements of the Pol II are the clamp, which is mobile and can close the active center, and the wall and dock, which bind initiation factors, such as TFIIB (Cramer, 2019b) (Figure 1|).

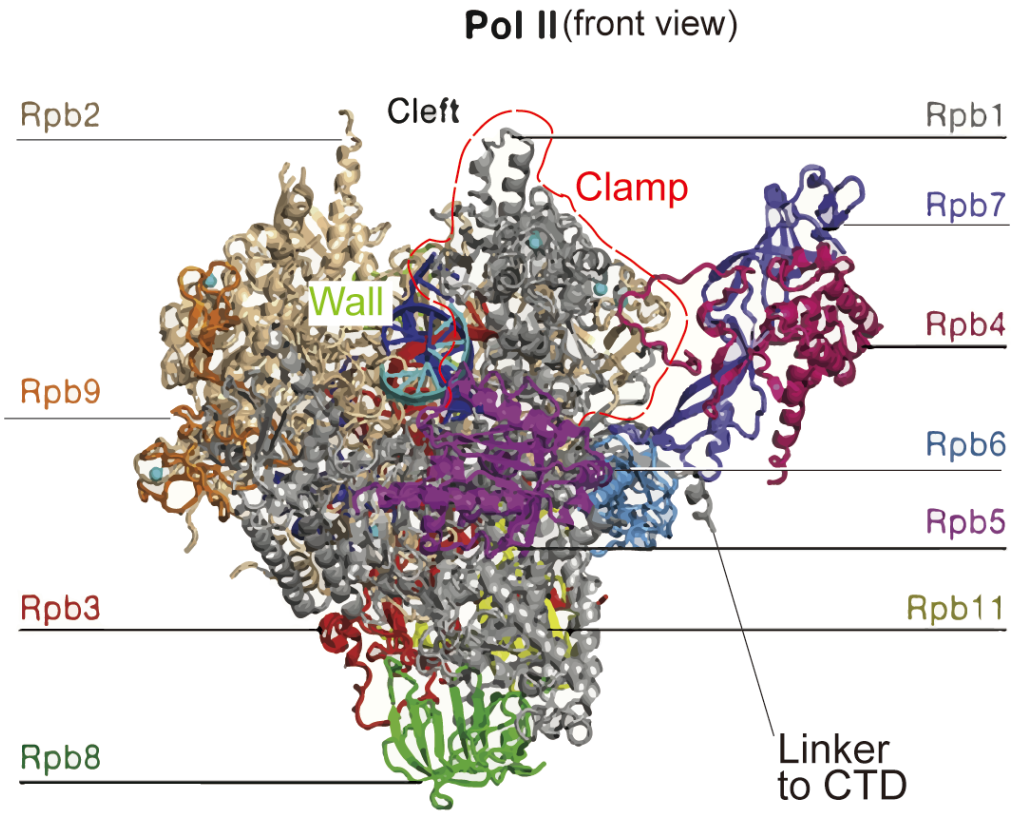


Figure 1| **Composition of RNA Pol II.** Pol II contains 12 subunits. The largest two subunits Rbp1 and Rbp2, form a cleft, the active center. The unstructured carboxyl-terminal domain (CTD) of Rpb1 is connected to Rpb1 by the linker region. Modified from (Cramer, 2019a).

1.3 RNA Pol II carboxy-terminal domain (CTD)

Rpb1, the largest subunit of Pol II, has a unique carboxy-terminal domain (CTD), a long, repetitive, and low complexity polypeptide region. Such unique CTD is only found in Pol II. CTD undergoes post-translational modifications and dynamically couples with the transcription stages and other transcriptional processes. Thus, it regulates gene expression.

1.3.1 Composition of CTD

CTD consists of tandem heptad repeats with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. This heptad sequence is highly conserved throughout all eukaryotes, with the number of repeats: 52x in mammals, 26x in S. *cerevisiae*, 29x in S. *pombe,* and 42x in Drosophila (Figure 2|) (Chapman et al., 2008). Most heptads follow the consensus sequence in yeast, while mammals show more deviations from the consensus sequence, with 21 consensus and 31 non-consensus repeats. The first 26 repeats are highly conserved, whereas the distal repeats are less conserved. The size of the CTD varies dramatically between species, with the lowest number being zero in *Trypanosoma brucei* and only five heptad repeats in *Plasmodium yoelii*. Between the CTD domain and Rpb1 main body is a linker region.

The conservation of heptad repeats in the CTD region differs through species. For example, the CTD of yeast is highly conserved. In budding yeast, there are 7 out of 26 non-consensus repeats. In contrast, CTD of Drosophila contains only two consensus repeats. In humans, there are 21 consensus repeats of 52 (Figure 2|).

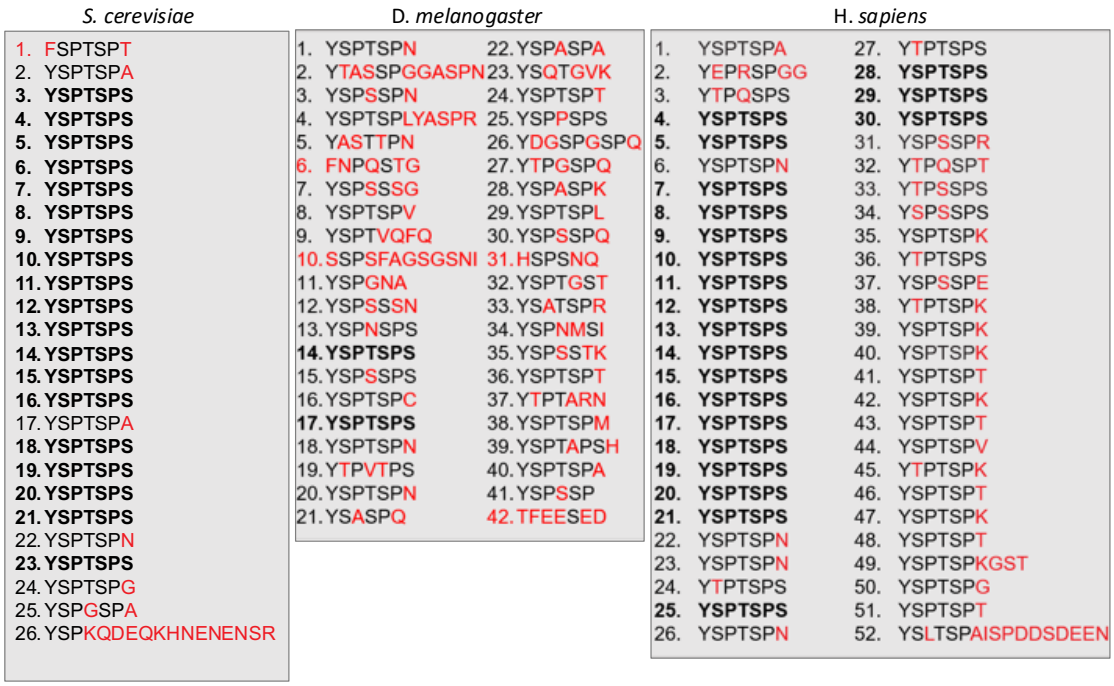


Figure 2| **The conservation of the CTD.** The carboxy-terminal domain (CTD) of S*accharomyces cerevisiae* (26 repeats), Drosophila melanogaster (42 repeats), and human (52 repeats) RNA Pol II is depicted. The repeat-sequence consensus is highlighted in black, and variants from the consensus sequence are labeled in red.

A full-length CTD is not required for viability or even for maintaining the growth rates in budding yeast. A growth assay study in yeast showed that eight consensus repeats are sufficient for viability. Yeast with ten consensus repeats grows similar fast as wild type (West & Corden, 1995). However, some other work in yeast indicated that a repeat number reduced CTD has instability of DNA (Morrill et al., 2016). Thus, although a minimum length of CTD maintains viability, only a full-length CTD is fully functional. In addition, replacing some residues within CTD in yeast is tolerated and gives a mild phenotype (West & Corden, 1995).

1.3.2 Modifications of CTD

The sequence of the CTD consists of amino acid residues that allow post-translational modifications (PTMs), which include phosphorylation, glycosylation, proline isomerization (Figure 3|) (Custódio & Carmo-Fonseca, 2016; Eick & Geyer, 2013; Harlen & Churchman, 2017; Yurko & Manley, 2018; Zaborowska, Egloff, et al., 2016). In addition, arginine and lysine residues in non-consensus repeats of CTD become modified by acetylation, mono-, di- and tri-methylation, and ubiquitylation Argine1810 (R1810), located in repeat 31, has been shown to be asymmetrically or symmetrically methylated (Voss et al., 2015). Recently, citrullination of R1810 was also added to the

list of modifications (Sharma et al., 2019).

Direct analysis of CTD modifications from yeast to humans by mass spectrometry has been established in recent years (Burriss & Mosley, 2019). In addition, an endogenously modified cleavable CTD is accessible for MS analysis, enabling modification mapping to individual residues within the CTD sequence (Schüller et al., 2016; Suh et al., 2016).

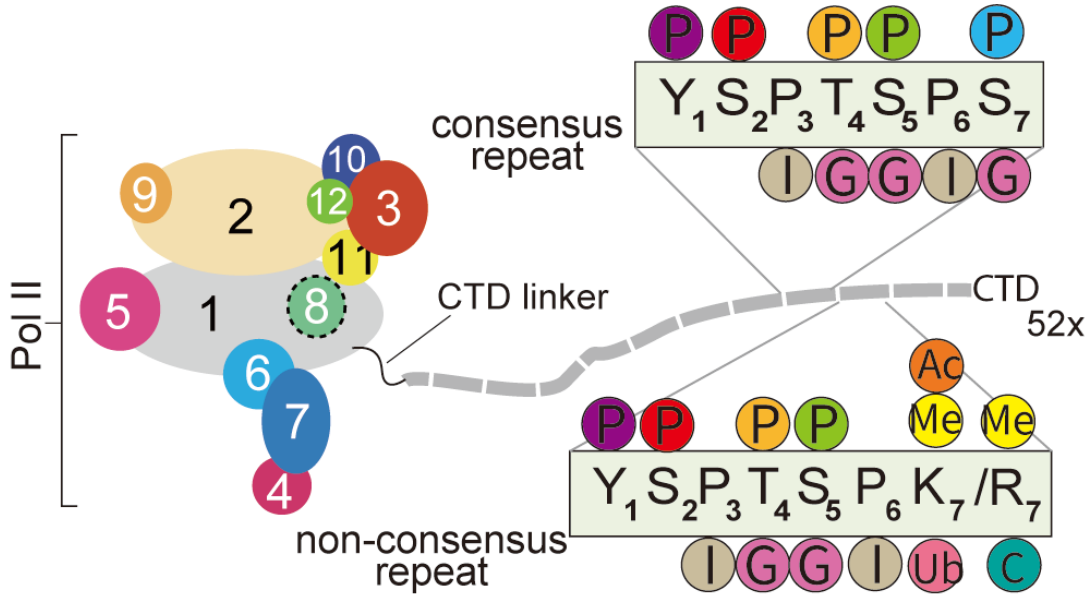


Figure 3| **Modifications of CTD repeats.** Left: Pol II with subunits Rpb1–Rpb12 numbered 1–12, the CTD of Rpb1 with 52 repeats is indicated as a dashed line. Right: modifications of each residue on consensus and non-consensus repeats. (P) dots, phosphorylation of Tyr1, Ser2, Thr4, Ser5 and Ser7; (G) dots, glycosylation; (I) dots, isomerization; (Ac) dot, acetylation; (Me) dots, methylation; (Ub) dot, ubiquitination; (C) dot, citrullination of R1810.

By now, the most well-studied CTD modification is phosphorylation. Within a single heptad, 5 out of 7 amino acid residues can be phosphorylated (Tyr1, Ser2, Thr4, Ser5, and Ser7). Phosphorylation of these residues is performed by specific enzymes, called kinases, that catalyze the transfer of phosphate residues. The removal of these phosphates is facilitated by another enzyme family: phosphatases. Chromatin immunoprecipitation sequencing (ChIP-seq) analysis, using specific CTD modification antibodies, indicated that the relative abundance of these modifications changes during the transcription cycle along the gene body (Figure 4|) (Hintermair et al., 2012; Nojima et al., 2015). For example, Ser2P and Ser5P are the most abundant and well-studied, but their distributions are different throughout the gene body. Ser5P reaches a maximum level around the transcription start site (TSS). After initiating transcription, Ser5P signals are thought to be reduced rapidly and maintained at lower levels until transcription termination. Differently, Ser2P signals increase along the gene body,

suddenly decreasing where transcription terminates (Corden, 2013). ChIP-seq also indicated that Ser7P, Tyr1P, acetylated Lys7, and mono- or di-methylated Lys7 are also highest at the TSS in mammals, Thr4P is highest at 3′ends. Tyr1P is also associated with antisense TSS and active enhancers. The ubiquitination and glycosylation have not yet been mapped genome-wide. The phosphorylation patterns of each amino acid appear and have been characterized throughout the genome and are mostly conserved across the species, from yeast to mammals. One exception is Tyr1P in S. *cerevisiae*, which shows a similar phosphorylation pattern to that of Ser2P. In metazoan, Tyr1P level is highest near TSS, and it is associated with Pol II peak in both direction, sense, and antisense (Eick & Geyer, 2013). Moreover, Thr4P shows a more uniform distribution along with the genes in yeast.

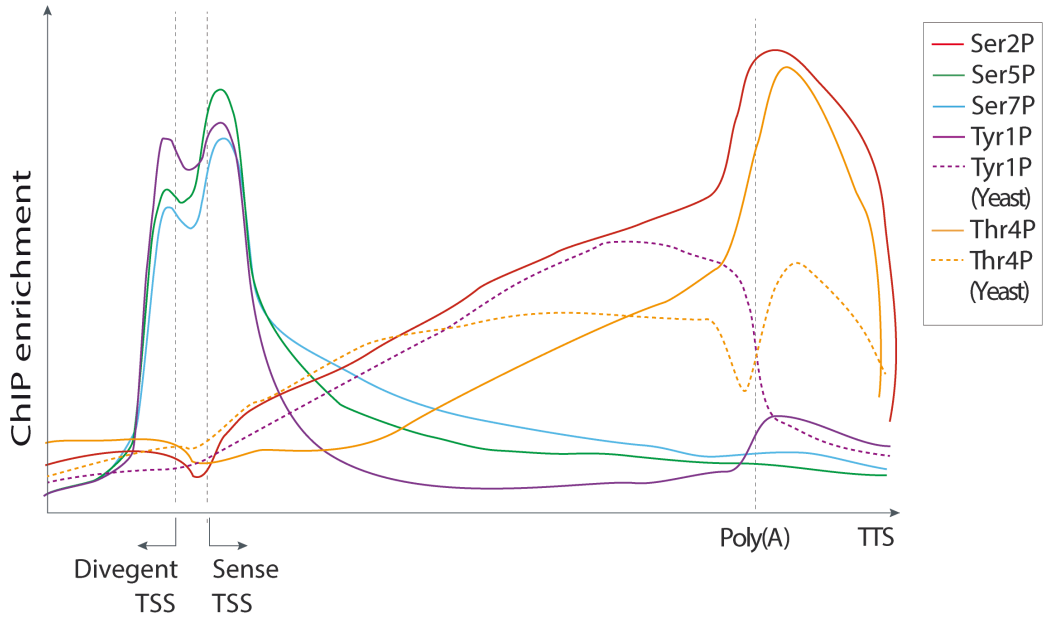


Figure 4| **CTD phosphorylation and transcription cycle.** Average chromatin immunoprecipitation (ChIP) profiles of phosphorylation of the Pol II-CTD across protein-coding genes in Homo sapiens and *Saccharomyces cerevisiae*. The profiles of budding yeast and humans are highly similar, except for Tyr1P and Thr1P (dash lines). Modified from (Harlen & Churchman, 2017).

1.4 Enzymes in CTD modification

CTD serves as an excellent binding platform of cellular factors in the proximity of the nascent transcript. During transcription and RNA processing, various enzymes dynamically modify CTD, and this modified CTD allows the recruitment of regulatory proteins. Specific modification of CTD is coupled with particular transcriptional events. Critical components of the CTD code are CTD modifications generated by CTD modifying enzymes– the writers, CTD associated factors read these modifications– the

readers, and CTD code modification are removed by enzymes, like phosphatases– the erasers. Combining these regulatory elements of CTD leads to precise control of transcription and RNA processing in spatial and temporal manners.

Identification of factors involved in transcription interacting with the CTD is continually expanding (Corden, 2019; Genes Robles & Coin, 2019; Harlen et al., 2016; Ramani et al., 2020). Although the mechanism of how the kinases recognize and modify the correct residue is not well understood, both CDK7 and CDK9 likely affect the Ser5P level in the gene body (Zaborowska et al., 2016). CDK7 is considered a Ser5-specific kinase (Ramani et al., 2020). CDK9 is already known as a Ser2 kinase but can also phosphorylate Ser5 in vitro (Czudnochowski et al., 2012). CDK8, a Mediator complex subunit, has Ser2 and Ser5 phosphorylation activity in vitro (Galbraith et al., 2010). An in vitro study showed that P-TEFb associated with BRD4 is specific for Ser5P when Ser7 is pre-phosphorylated (Glover-Cutter et al., 2009). DYRK1A is a known gene-specific kinase that phosphorylates both Ser2 and Ser5; however, the extent and generality of its role in Ser5 phosphorylation remain unclear (Vona et al., 2015). A recent study shows DYRK1A is a Ser2-specific kinase (Ramani et al., 2020). In DNA damage response, a study showed that glycogen synthase kinase 3 (GSK-3) phosphorylates the CTD on Ser2 and Ser5 in vitro, but preferentially when the CTD is pre- phosphorylated (Nieto Moreno et al., 2020). Erk2 shows dual specificity for Ser2 and Ser5 (Ramani et al., 2020). Inhibition of CDK12 significantly affects Ser2 CTD phosphorylation in DNA-damage-response genes, whereas a minor impact on global Ser2P was detected (Liang et al., 2015). In addition to Ser2, CDK12 can phosphorylate Ser5 and Ser7 in vitro (Bösken et al., 2014). CDK13 also shows phosphorylation ability on Ser2 and Ser5 in vitro (Greifenberg et al., 2016). CDK11, which associates with the Pol II elongation complex near the 3' end, promotes Ser2 phosphorylation and contributes to the recruitment of cleavage- and- polyadenylation factors (Pak et al., 2015). Some other factors can also bind the CTD, recruiting or associating with other CTD kinases or factors, thus regulating transcription. A member of the JmjC protein family, JMJD5, binds to CTD with a unique phosphorylation code generated by CDK9 to release Pol II from the pause site (Liu et al., 2020). Therefore, a wide range of kinases is responsible for Ser2P and Ser5P.

Besides Ser2P and Ser5P other modifications of the CTD have been described. Polo-

like kinase 3 (PLK3) shows a Thr4-specific kinase activity in human HEK293 cells (Hintermair et al., 2012). In human cells, the non-consensus Arg1810 undergoes symmetric and asymmetric di-methylation by protein arginine methyltransferase (PRMT5) (Zhao et al., 2016) and co-activator-associated arginine methyltransferase 1 (CARM1) (Sims et al., 2011), respectively. Eight lysines in mammalian CTD are acetylated by p300/KAT3B acetyltransferase (Schröder et al., 2013).

Although some CTD kinases are well studied and have a clear function, some other kinases remain controversial. For example, in Drosophila, CDK9 is the most critical kinase for Ser2 at the 5' region, while during elongation, CDK12 takes charge of Ser2 phosphorylation (Bartkowiak et al., 2010). In humans, both CDK12 and CDK13 can phosphorylate the CTD during elongation. CDK12 phosphorylates Ser5, Ser7, and Ser2 in vitro, whereas CDK13 phosphorylates Ser5 and Ser2 in vitro (Bösken et al., 2014). Nevertheless, inhibition of CDK12 or CDK13 has shown a minimal effect on the global level of CTD phosphorylation in vivo and minimal impact on gene expression (Greenleaf, 2019; Liang et al., 2015). Thus, the relation or interaction between CDK9, CDK12, and CDK13 is still unclear.

Elimination of the modifications is as crucial as the production of the modifications, which allows Pol II to initiate a new cycle of transcription and mediates transitions in the ongoing transcription cycle (Eick & Geyer, 2013). For example, RPAP2 (Rtr1 in yeast) de-phosphorylates Ser5P (Egloff et al., 2012). Ssu72, as a phosphatase, de-phosphorylates Ser5 and Ser7, which is associated with transcription termination (Zhang et al., 2012). The evolutionarily conserved TFIIF-associated CTD phosphatase 1 (Fcp1) de-phosphorylates Ser2P mainly at the end of protein-coding genes (Schreieck et al., 2014).

1.5 CTD modifications in transcription regulation

The CTD is not required for the Pol II catalyzing. However, it is crucial for regulating multiple steps during transcription and the coupling of transcription from several co-transcriptional-related events. Such transcription regulatory character is often achieved via the post-transcriptional modification of specific peptides.

The transcription process in eukaryotes consists of three major stages: initiation,

elongation, and termination (Eick & Geyer, 2013). In general, dynamic changes in CTD phosphorylation are strongly coupled with transcription stages, with a dramatic shift appearing at transitions between transcription stages. Furthermore, these phosphorylation changes promote or inhibit the transcription factor's binding ability to the CTD, thus control these transitions.

1.5.1 Initiation

All RNA polymerases need to break similar barriers to initiate transcription, mainly involving DNA template opening, specific promoter region binding, transcription starting site recognition, and transition to productive elongation (Figure 5|). But unfortunately, the polymerase cannot do it alone.

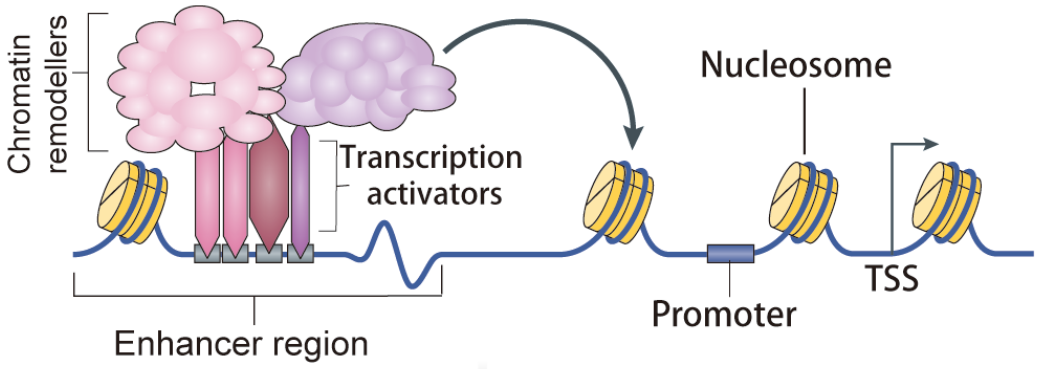


Figure 5| **A simplified model for transcription activation.** Transcription starts with an opening chromatin structure. Briefly, the transcription activators bind on enhancer regions at a far distance from the promoter region. Transcription activators then recruit chromatin remodelers to modulate chromatin structure for docking other activators such as the Mediator complex. Modified from (Soutourina, 2018).

In prokaryotes, the singular RNA polymerase showed no transcription activity in a situation with an isolated polymerase and a target DNA sequence. However, when adding RNA polymerase-free cellular extract, transcription happens (Magee et al., 1975). Further studies focused on identifying which components are required for transcription conduction, known as General Transcription Factors (GTFs). These components can assemble and bind onto the promoter domain, serving as a platform for RNA polymerase locating and binding. Transcription in prokaryotes is served by only one GTF, while eukaryotes employ multiple GTFs. These factors needed for transcription activation of a complex called the preinitiation complex (PIC). The components of the PIC vary when compared with different RNA polymerases in the eukaryotic system. For transcription initiation, Pol II is recruited by the PIC complex, which is composed of 6 essential subunits: TFIIB, TFIID, TFIIH, TFIIE, TFIIF, and

TFIIA (not essential for PIC function, but necessary to stabilize TFIID binding to DNA) (Figure 6|).

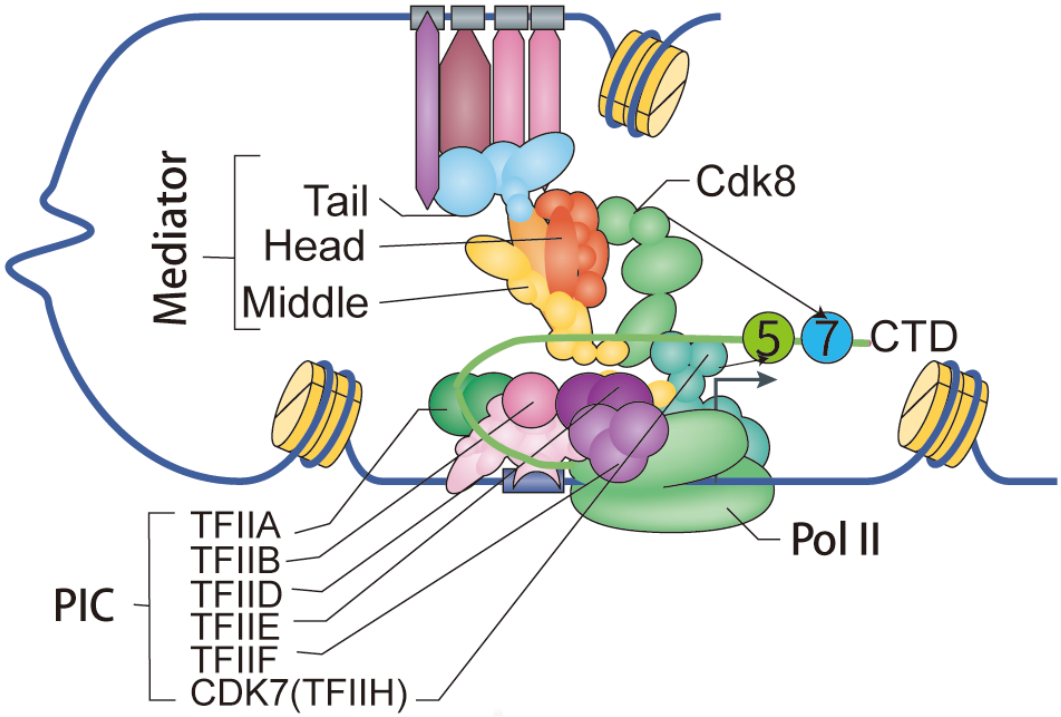


Figure 6| **A simplified model of transcription initiation.** Transcriptional activators are recruited and assemble the basal transcriptional machinery called the preinitiation complex (PIC) at the promoter region. The Pol II is part of this PIC with an unphosphorylated CTD that interacts with the Mediator complex. Mediator has a high affinity for unphosphorylated CTD. Upon phosphorylation of Ser5 by cyclin-dependent kinase 7 (CDK7; subunit of TFIIH), this affinity is lost, and Pol II escapes from the promoter. Additionally, phosphorylation of Ser7 by CDK8 assists promoter escape. Modified from (Soutourina, 2018).

The Mediator is a multi-subunit complex and transcriptional co-regulator conserved in eukaryotes (30 subunits in humans and 25 subunits in S. *cerevisiae*). It is required for transcription initiation and is often considered a bridge to connect transcription factors to basal factors, including Pol II and general transcription factors (GTF), promoting preinitiation complex assembling at the promoter region. During transcription initiation, Mediator is recruited to the enhancer region via interacting with specific transcription factors. Then, the Mediator can contact the PIC positioned at the promoter region over a far distance by looping chromatin, contributing to PIC assembling and stabilization.

In the presence of the PIC and general transcription factors, an un-modified CTD of Pol II is recruited to promoters. CTD is essential for PIC formation and stability because the Mediator complex is essential for promotor-enhancer looping formation. Notably, the un-modified CTD has a high affinity for Mediator by forming several hydrophobic interactions with Mediator (Harlen & Churchman, 2017).

CTD modification is also regulated by Pol II escaping from the promoter and the transition from initiation to elongation. Cyclin-dependent kinase 7 (CDK7; Kin28 in S. *cerevisiae*), a sub-component of general transcription factor TFIIH, phosphorylates CTD Ser5 and Ser7. The role of Ser7 phosphorylation is not clear. In vitro experiments indicate that Ser7P can promote CTD for further phosphorylation (Czudnochowski et al., 2012). However, phosphorylation on Ser5 is more apparent. It can reduce the hydrogen bonding affinity between Pol II and Mediator, boosting promoter escape (Wong et al., 2014). Ser5P also recruits the capping enzyme at early nascent RNA 5', which protects the transcript from degradation (Fabrega et al., 2003). TFIIH recruitment requires CDK8, which is the kinase subunit of Mediator. CDK7 mediated phosphorylation of Ser5 of CTD dissociates the Mediator and releases Pol II (Allen & Taatjes, 2015). Thus, a modified CTD is dynamically phosphorylated and de-phosphorylated by kinases and phosphatases throughout the transcription process, whereas the un-modified CTD is predominantly implicated in transcription initiation.

1.5.2 Promoter-proximal pausing and elongation

In metazoans, there is a critical regulatory process during elongation called promoter-proximal pausing. When Pol II escapes from the initiation site and synthesizes a short, nascent RNA, it then pauses promoter-proximally, waiting for other signals to enter the downstream gene body. Only when paused Pol II can transit into the productive elongation phase, a full-length functional mRNA be produced (Figure 7|).

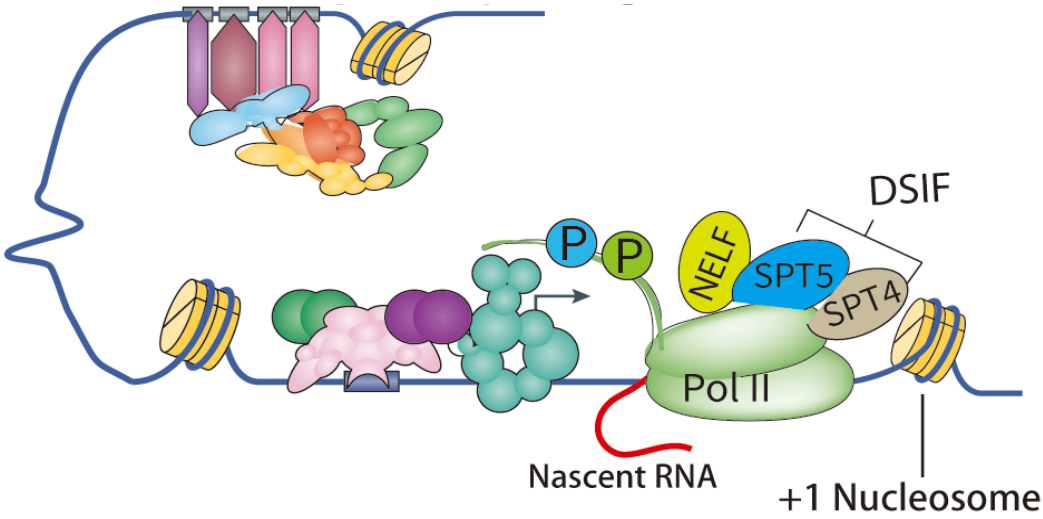


Figure 7| **A simplified model of promoter escape and pausing.** Pol II is highly phosphorylated at Ser5 and Ser7 of CTD when escaping from TSS. Soon after the TSS, before the +1 nucleosome, Pol II is paused and stabilized by negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF), waiting for further signals. Modified from (Soutourina, 2018).

Studies in the early 1990s showed that several essential regulatory genes in mammalian cells, such as c-MYC (Strobl & Eick, 1992) and c-FOS (Plet et al., 1995), have an enrichment of Pol II just downstream of transcription starting site (TSS). Additionally, Pol II accumulation at promoters was found on the HIV long terminal repeat (LTR) (Kao et al., 1987). Following studies using chromatin immunoprecipitation (ChIP) assays provide evidence that proximal pausing of Pol II is widespread in metazoans (Core & Adelman, 2019).

In vitro studies found that two complexes, NELF and DSIF, are required for transcription repression by DRB (Yamaguchi et al., 1999). DSIF is composed of transcription elongation factors SPT5 and SPT4, which together form a heterodimer called dichloro-1-β-D-ribofuranosyl- benzimidazole [DRB] sensitivity-inducing factor (DSIF) that regulates transcription elongation. When mutating certain positions in SPT5, P-TEFb failed to phosphorylate SPT5, preventing Pol II release, even though P-TEFb was fully functional in Pol II phosphorylation (Yamada et al., 2006). Like Pol II CTD, SPT5 also has a pentapeptide repeat region, phosphorylated by both the kinase subunit of TFIIH and P-TEFb. SPT5 is also involved in the recruitment and activation of the capping complex (Doamekpor et al., 2014). After transcribing 20-100 nuclear bp downstream of the TSS, the arising nascent RNA from the Pol II RNA exit channel is recognized by SPT5. This critical contact with nascent RNA allows DSIF to contribute to the block on further transcription by Poll II (Bernecky et al., 2017).

The negative elongation factor (NELF) is composed of four subunits. NELF- A and NELF- C (or its isoform NELF- D) form the core subunit, associated with NELF- B and NELF- E (Conaway & Conaway, 2019). NELF recognizes the Pol II–SPT5 interface and associates with the elongation complex at the promoter-proximal region (Vos et al., 2018). All subunits of NELF can associate with RNA, and it offers a possibility that NELF interactions with early nascent RNA could prevent premature termination (Vos et al., 2018). Several studies from bacteria to humans have shown that early elongation complexes transcribe slower and not as stable as complexes downstream within gene bodies, rendering the RNA polymerase quite sensitive to gene sequence (Core & Adelman, 2019; Luse, 2013). The paused Pol II complex is stabilized by DSIF and NELF. NELF can also prevent the enzymatic reaction of Pol II at the catalytic site (Adelman & Henriques, 2018).

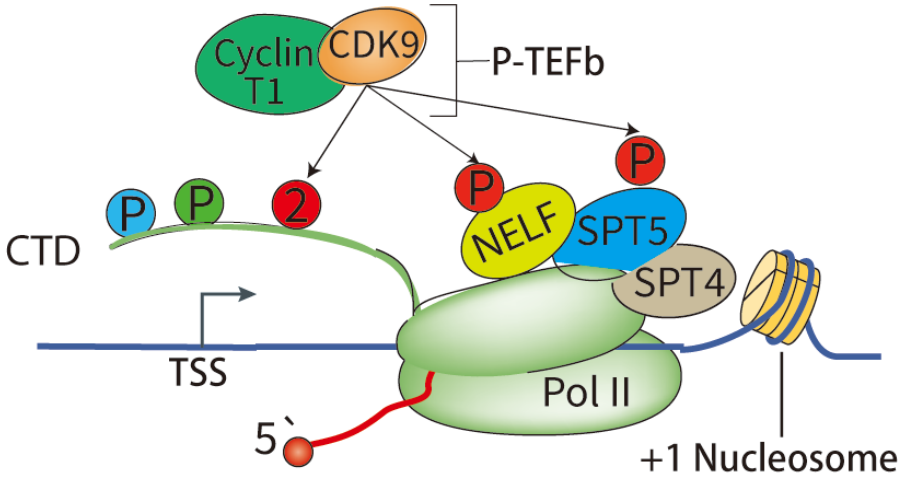


Figure 8| **A simplified model for pause release.** CDK9, together with Cyclin T1, arrives at paused Pol II and phosphorylates NELF, DSIF, and Ser2 of CTD. This phosphorylation causes the release of NELF and switches DSIF from the pausing factor into a positive elongation factor, thus releasing Pol II from the pausing site. Modified from (Soutourina, 2018).

The transition from proximal pausing into productive elongation is triggered by positive transcription elongation factor b (P-TEFb) (Figure 8|). P-TEFb is a cyclin-dependent kinase-containing complex including the catalytic subunit, CDK9, and a regulatory subunit cyclin T (Peng et al., 1998). P-TEFb associates with other factors, including the acetylated histone-binding, protein bromodomain-containing 4 (BRD4), which is often associated with a large complex called the super elongation complex (SEC) (Yang et al., 2005), transcription activators such as MYC (Baluapuri et al., 2019), nuclear factor-κ B (NFκB), and the MED26 subunit of the Mediator complex. In addition, P-TEFb is also associated with other elongation factors and chromatin remodelers in the SEC. P-TEFb was identified in vitro by purifying Drosophila cell extracts as a factor sensitive to the transcription elongation inhibitor 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB). Treatment of DRB leads to a global defect of Pol II release and reduces nascent transcripts' stability (Marshall & Price, 1995). P-TEFb kinase activity is tightly regulated when being associated with other larger complexes: the super elongation complex (SEC), bromodomain-containing protein 4 (BRD4), and 7SK small nuclear ribonucleic protein (snRNP) (Chen et al., 2018). P-TEFb-7SK snRNP complex is considered an inactive form of P-TEFb and represses transcription activity by trapping a large portion of cellular P-TEFb through a dimer of HEXIM1 and HEXIM2. The majority of cellular P-TEFb, up to 90%, is arrested in the 7SK- P-TEFb complex (Zhou et al., 2012). In contrast, P-TEFb associated with SEC and BRD4 comprises the active form of P-TEFb to release Pol II from the pause site. In particular, P-TEFb with SEC is the most active form.

Once a gene is active to initiate transcription, a transition of P-TEFb from repressive form to active form is conducted. This process depends on multiple factors, including KRAB-associated protein 1 (KAP1; also known as TRIM28), the serine/ arginine-rich splicing factor 1 (SRSF1), and SRSF2, which associate with 7SK snRNP on chromatin (Ji et al., 2013). P-TEFb is well known to phosphorylate DSIF, NELF, and Pol II CTD (Marshall & Price, 1995). Phosphorylation of NELF by P-TEFb causes its release from chromatin. Meanwhile, phosphorylation of the SPT5 CTR by P-TEFb triggers DSIF switching from a pausing factor to be a positive elongation factor. Furthermore, phosphorylation of the Pol II CTD at Ser2 by P-TEFb is required to transit Pol II from initiation into elongation, progress the transcription into the elongation phase, and is considered a mark of active, productive elongation (Egloff et al., 2012) (Figure 9|).

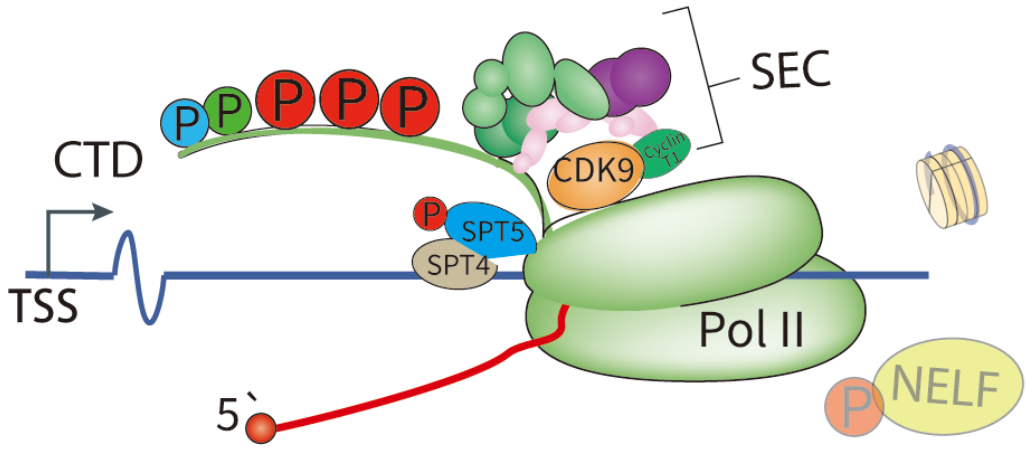


Figure 9| **A simplified model for productive elongation.** During elongation, the Ser2P of the CTD continuously increases. In contrast, the levels of Ser5P and Ser7P are getting low, which promotes the recruitment of elongation factors, chromatin-modifying, and RNA-processing factors that regulate co-transcriptional processes. Modified from (Soutourina, 2018).

Ser2 phosphorylation by P-TEFb is promoted when the CTD is pre-phosphorylated at Ser7 (Czudnochowski et al., 2012). Since Ser7 phosphorylation is conducted by the TFIIH kinase subunit Cdk7 during initiation, this suggests that the Ser2-P by CDK9 can only occur after successful transcription initiation. Thus, P-TEFb is not responsible for the total of Ser2-P throughout CTD during transcription. Instead, the kinases Cdk12 and Cdk13 play a significant role in this phosphorylation. Particularly, Cdk12, often associated with Pol II during elongation, and Ser2-P levels continually increase along the downstream gene, peaking at the 3′ ends (Mayer et al., 2010). In addition, the phosphorylated CTD recruits the PAF1 complex and other transcription elongation factors, such as SPT6 and the human transcription elongation regulator 1 like (TCERG1) (Liu et al., 2013).

P-TEFb together activated and recruited by BRD4 and SEC, phosphorylates Ser2 of the Pol II CTD. The BET proteins, like BRD4, have two tandem bromodomains for recognizing acetylated lysine residues and an extra terminal domain for interacting with regulatory factors such as JMJD6, thus can direct these proteins towards transcriptionally active chromatin regions, such as promoters and enhancers (Konuma et al., 2017). BRD4 interacts with P-TEFb, causing the disassociation of P-TEFb from the 7SK complex, increasing free P-TEFb, which can further assemble into its active complex (Schröder et al., 2012). Rapid degradation of BRD4 leads to reduced Ser2 phosphorylation without a noticeable change in the chromatin occupancy of CDK9 (Winter et al., 2017). Moreover, BRD4 was reported to possess kinase activity and that it phosphorylates Ser2 residues of the CTD in vitro, independent from P-TEFb (Devaiah et al., 2012). However, an in vitro experiment showed that the vital function of BRD4 on P-TEFb is specific for Ser5-P when Ser7 is pre-phosphorylated (Larochelle et al., 2012). How the SEC is recruited to chromatin is less clear. Transcription factor HES1 binds and recruits the SEC to activate gene transcription (Liu et al., 2017). For genes like FOS and JUN, the SEC is recruited by the Integrator (Gardini et al., 2014). Other studies show that the SEC recognizes specific chromatin profiles (Wan et al., 2017). For example, Mediator subunit 26 (MED26) promotes SEC recruitment by interacting with the SEC subunits ELL- associated factor 1 (EAF1) and EAF2 (Lens et al., 2017). Nevertheless, SEC recruitment is present at most gene-active regions by Pol II.

With more and more factors demonstrated in the pausing-elongation transition, such as BRD4, CDK9, PAF1, a substantial part of this regulation is mediated by CTD kinases, emphasizing the importance of CTD phosphorylation in this process.

The continuous development of increasingly specific P-TEFb kinase inhibitors, beginning with DRB, flavopiridol, and precise inhibition strategies such as analog-sensitive kinase technology, enable more detailed studies of P-TEFb during transcription in living cells (Garriga & Graña, 2014; Gressel et al., 2017; Yamaguchi et al., 1999) At the same time, the list of phosphorylation targets of P-TEFb keeps expanding. For example, a recent study revealed that a CTD linker region of Pol II is phosphorylated by P-TEFb, followed by SPT6 binding to this phosphorylated linker (Vos et al., 2018).

1.5.3 Termination

Hundreds of thousands of transcription reactions are always going to end by their termination. Transcription termination occurs when RNA polymerase and nascent RNA are detached from the gene body. Termination occurs in a variety of ways when polymerase responding to different gene inputs. So far, three termination pathways have been described in metazoans, poly (A) dependent pathway for mRNAs, Sen1 dependent pathway for snRNAs, and a stem-loop structure-dependent way for replication-dependent histone genes.

In most protein-coding genes, the termination process is poly-adenylation signal-dependent (PAS), which defines the transcription boundary of mRNA 3' end. PAS is encoded within the DNA template and includes an AAUAAA motif often coupled with upstream U-rich and downstream U/ GU-rich sequences (Figure 10|). These sequences are recognized by a cleavage and polyadenylation (CPA) complex, assembled from cleavage stimulatory factor, cleavage and polyadenylation specificity factor (CPSF) I and IIm sub-complexes. The cleavage starts with the recognition of AAUAAA, performed by CPSF30 and WDR33 (components of CPSF). Next, the endonuclease CPSF73 cleaves nascent RNA at the position between the AAUAAA and downstream elements (Eaton & West, 2020). Notably, a PAS mutation can cause termination failure (Whitelaw & Proudfoot, 1986).

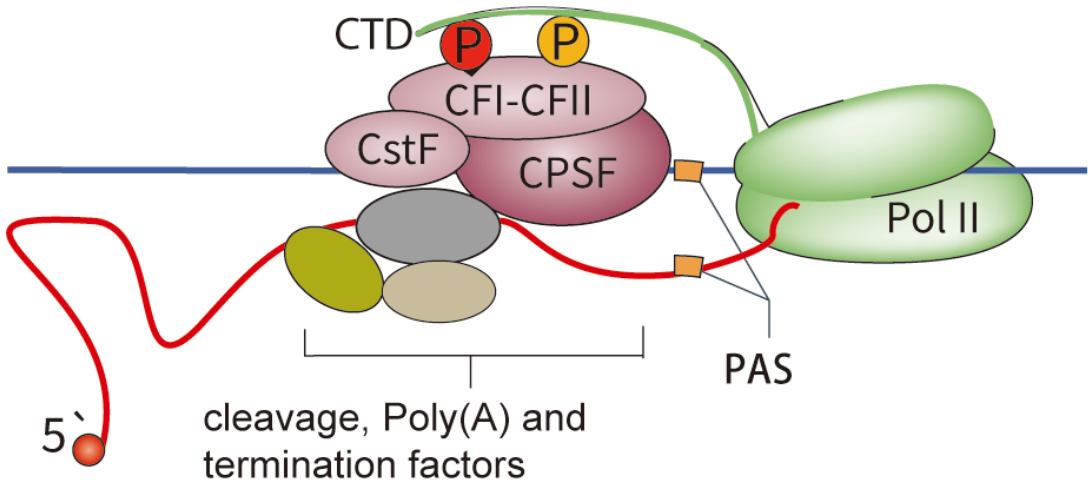


Figure 10| **A simplified model for termination.** Pol II with a high level of Ser2P and Thr4P recruits cleavage, polyadenylation factors, and termination factors that release Pol II from the DNA. PAS, polyadenylation site. Modified from (Soutourina, 2018).

Two different models were developed to explain PAS-dependent termination,

allosteric/anti-terminator, and torpedo models. The allosteric/anti-terminator model proposes that transcription of the PAS dissociates the factors from elongating complexes, leading to a conformational change within Pol II, resulting in a loss of elongation anti-termination factors ultimately leads to termination. Anti-terminator factors are aimed to suppress termination until passing the PAS element. When depleting these factors, premature cleavage and polyadenylation are observed, including the depletion of SCAF4, SCAF8, U1 snRNA, CDK12, or nuclear poly(A) binding protein (PABPN1) (Eaton & West, 2020; Gregersen et al., 2019). In vitro assays have shown that PAS transcription induces conformational changes within Pol II that promote termination even without cleavage (Zhang et al., 2015). In the torpedo model, the cleavage of the pre-mRNA by the CPF–CF complex provides a position for a 5′ to 3′ exonuclease (human XRN2 and budding yeast Rat1), which degrades the nascent RNA tail up to the transcribing Pol II, causing the elongating Pol II to disassociate from the DNA template (Eaton & West, 2020). This model would explain the coupling between cleavage and termination. XRN2 depletion inhibits PAS-dependent termination, and mutation of the XRN2 active site delays the termination process (Eaton et al., 2018).

Recent studies show evidence for the unified termination mechanism of these two original models. PAS cleavage promotes the deceleration of Pol II caused by the de-phosphorylation of SPT5 by PNUTS/PP1 and constitutes an allosteric switch (Cortazar et al., 2019). In addition, PAS cleavage is also associated with increased Thr4 phosphorylation of Pol II CTD. This change allows XRN2, the exonuclease, to quickly terminate Pol II.

Pol II CTD phosphorylation is associated with transcription termination in both yeast and humans. In yeast, Ser2P is bound directly by subunits of the cleavage, polyadenylation, and termination complexes, such as polyadenylation factor I (Pcf11) and regulator of Ty1 transposition protein 103 (Rtt103) (Lunde et al., 2010). Other CTD phosphorylation such as Tyr1P, Thr4P, and Ser7P are also implicated in recruiting termination factors. Tyr1P prevents the binding of Rtt103 and Pcf11 to the CTD in vitro (Mayer et al., 2012). Tyr1P levels decrease before PAS in yeast, and the overall Tyr1P level is much lower than the Ser2P signal. A Thr4P peak signal is also shown downstream of PAS, and it binds Rtt103 (Harlen et al., 2016).

In non-coding RNA transcription in S. *cerevisiae*, Ser5P is associated with the NNS complex, which is responsible for the termination of the transcription of small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and cryptic unstable transcripts (CUTs). Mainly, Nrd1, a subunit of the NNS complex, binds to Ser5P, thus, brings NNS to Pol II (Vasiljeva et al., 2008). Other CTD residues are also found in non-coding transcript termination. For example, Ser7P was shown to be involved in 3' processing of snRNA gene transcription in mammals (Egloff et al., 2007), and Thr4P signal was found in 3′ processing at histone genes (Hsin et al., 2011). Thus, these CTD modifications collaborate to regulate termination, cleavage, and polyadenylation of both protein-coding and non-coding RNAs from yeast to mammals.

1.6 CTD in co-transcriptional regulation

The CTD also plays a role in several co-transcriptional regulatory events involving pre-mRNA capping, splicing, histone modification, and chromatin remodeling (Harlen & Churchman, 2017).

1.6.1 5' capping

When a nascent transcript emerges at the Pol II RNA exit channel, the CTD Ser5P promotes the recruitment of the capping complex (Komarnitsky et al., 2000a), which conducts the first step of mRNA processing. Early capping of the nascent RNA is essential, not only because 5' capping can protect fragile nascent transcript from degradation, but it is also required for mRNA processing, including splicing, cleavage, and poly-adenylation (Topisirovic et al., 2011). This mechanism is highlighted by a CTD Ser5 to Ala mutation in S. pombe, which causes a lethal phenotype and can be rescued by tethering the capping complex to polymerase (Morrill et al., 2016).

1.6.2 Splicing

When transcription elongation goes on, often associating with another mRNA processing event, splicing can be regulated by CTD (Bentley, 2014). In yeast, splicing occurs co-transcriptionally and immediately after the 3' splice site is transcribed. The spliceosome interacts with Ser5P CTD. With ChIP-seq, a high level of Ser5P and Ser7P signal and a low level of Thr4P signal are associated with the 3' splice site (Milligan et al., 2016). Moreover, a signal of U1 small nuclear ribonucleoprotein A, a component of

the U1 spliceosomal subcomplex, is also found at 3′ splice sites (Herzel et al., 2017). In mammals, a phosphorylated CTD promotes splicing. Less phosphorylated CTD decreases recruitment of spliceosome when mutating Ser2. The mNET-seq approach reveals that high levels of Ser5P CTD are associated with spliceosome from the cell (Nojima et al., 2015; Nojima et al., 2018). Thus, an increased Ser5P signal promotes spliceosome recruitment at 3' splice sites in both yeast and mammals. However, the detailed interaction mechanisms of phosphorylated CTD can be different because the variations in average numbers and lengths of introns differ between species. Besides, Pol II has been found paused at 3' splice sites (Mayer et al., 2015). Together, this suggests that splicing is tightly coupled with elongation and CTD phosphorylation. These observations provide further evidence that CTD phosphorylation plays a role in co-transcriptional splicing.

1.6.3 Chromatin remodelling

Chromatin modification and 3D genome organization are known to regulate gene expression. Interestingly, transcription activity can also alter chromatin (van Steensel & Furlong, 2019). Furthermore, Pol II CTD contributes to the regulation of histone chaperones. Ser5P can recruit complex proteins associated with Set1 (COMPASS) at gene 5' region, conduct H3K4 methylation, and activate gene transcription (Terzi et al., 2011). In addition, Ser5P and Ser2P are essential for the recruitment of SET2, a methyltransferase on H3K36. H3K36me is involved in transcription elongation and chromatin stability and can recruit HDAC sub-component RPD3S (Li et al., 2003). The RPD3S is directly recruited by the CTD, so are other factors, including the SET3 complex, the histone acetyl-transferase (HAT) complexes NUA4 and SAGA (Ginsburg et al., 2009; Govind et al., 2007). Furthermore, during the elongation, Ser2P recruits and actives topoisomerase I (Baranello et al., 2016). These studies suggest a significant role of CTD phosphorylation in histone modification. However, the presiding details of which form of CTD modification regulates histone modification across gene bodies remain to be discovered. Nevertheless, CTD is also implicated in the regulation of the 3D genome.

1.7 Pol II CTD condensates: an additional layer of gene regulation

The dynamic phosphorylation of Pol II CTD has essential roles in transcription regulation. However, due to the repetitive nature of the CTD, this region represents low-complexity domain (LCD) properties. It was unclear how phosphorylation-dependent molecular interactions were able to provide sufficient specificity to a spatiotemporal variety of co-transcriptional events (Harlen & Churchman, 2017). During the transcription of a specific gene from the whole genome, hundreds of specific proteins and RNAs must work together at the targeting gene locus to ensure reliable transcription activation. Indeed, transcription components of gene regulation machinery are often found in a dynamic cluster where those factors are at a high concentration. Such clustering of those molecules is often required to conduct their interactions (Banani et al., 2017; Chong et al., 2018; Hnisz et al., 2017). Transcription factors exist in those membrane-less clusters, which appear in a liquid-like character and often contain low-complexity domains (LCDs) or intrinsically disordered regions (IDRs). The LCD-LCD has numerous weak, non-covalent interactions between numerous factors. Such multivalent interactions promote liquid-liquid phase separation (LLPS), thereby forming membrane-less compartments or bio-molecular condensates, such as granules or nuclear speckles (Sabari, 2020). These condensates have functional consequences for gene expression. The RNA Pol II molecule concentration correlates with the number of transcribed RNA molecules (Cho et al., 2016), and preventing clustering of the transcription machinery causes reduction of gene expression. A optogenetic engineered transcription factor formed condensates and amplified gene expression in a living cell (Wei et al., 2020), indicating that formation of transcription factor condensates promote transcription of associated genes. Modifying the CTD also alters the physical properties of condensates. Pol II condensation in vitro and clustering in cell nuclei is dependent on the length of the CTD (Boehning et al., 2018), indicating that the clustering ability of Pol II depends on multivalent interactions. The molecular properties of LCDs in condensate formation can be altered by post-translational modification, enabling cells to control the valency and composition of droplets (Hnisz et al., 2017). A phosphorylated Pol II CTD can drive the polymerase switch from initiation condensate to splicing condensate (Guo et al., 2019), which shows that CTD phosphorylation can

regulate condensate preference suggesting another mechanism of gene regulation by the CTD. CTD length contributes to CTD-CTD interaction, enables the recruitment of multiple polymerases, and enhances initiation droplet formation (Quintero-Cadena et al., 2020).

1.8 Pre establishment of the work

CTD phosphorylations are placed by kinases and are removed by phosphatases. These modifiers create distinct phosphorylation patterns that direct the recruitment of specific factors. Antibodies generated against each specific phosphorylated residue in the CTD heptad have been used to explore CTD modifications of each residue. Nevertheless, the use of antibodies does not allow the precise quantification of CTD phosphorylation and the exact mapping of a phosphorylation site to a specific heptad. A mass spectrometry (mass spec) approach is ideal for addressing this issue. However, the wild-type CTD sequence is not accessible by mass spec due to very few trypsin digestions sites. Besides, the nature of the tandem heptad repeats makes it challenging to identify specific peptides. A genetically modified CTD sequence, created by introducing multiple trypsin cleavage sites and replacing several amino acids in consensus repeat, can be used to generate unique tryptic peptides (Figure 11|) (Schüller et al., 2016). Although such genetically engineered CTD is no longer a wild-type CTD, CTD function is not dramatically affected by multiple point mutations. Furthermore, sequence-modified CTDs support proliferation rates similar to wild-type.

The previous mass spec results provide 100% CTD sequence coverage. On average, the CTD contains 1/3 phosphorylated heptads. However, most phosphorylated heptads contain one single phosphorylation. Moreover, not all heptads are equally phosphorylated, suggesting that the placed phosphorylations diverge along the CTD (Schüller et al., 2016). In other words, kinases and phosphatases might have preferential targeting positions within the CTD.

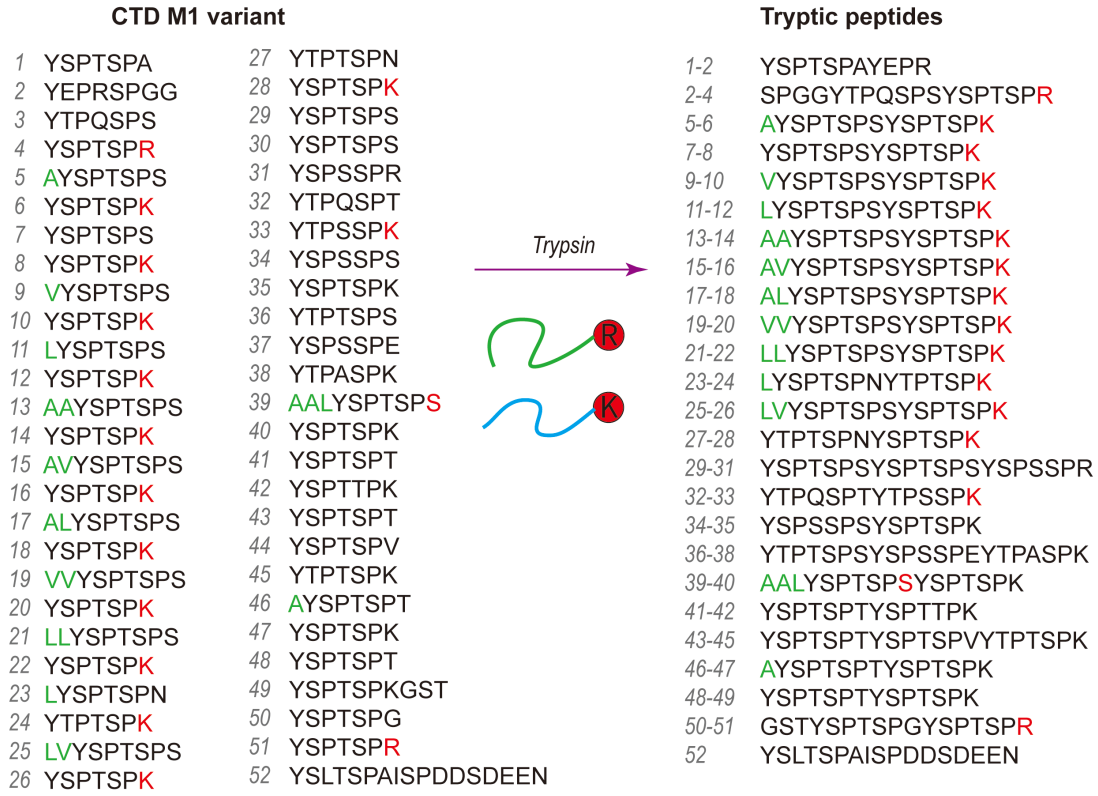


Figure 11| **Schematic sequence of a mass spec accessible CTD variant.** (Left) Pol II CTD sequence of variant M1. Red letters: amino acid substitution. Green letters: added residues. (Right) List of tryptic CTD peptides of variant M1. Numbers indicate CTD repeats covered by the corresponding peptide**.** Modified from (Schüller et al., 2016).

Since the mass spec optimized CTD variants provide deeper insights into CTD phosphorylation, I aim to answer how the CTD phosphorylation changes after kinases are specifically inhibited. To study a specific kinase function, researchers often used chemical compounds for kinase inhibition, such as flavopiridol (Chao & Price, 2001) and DRB, which target CDK9 (Matija & Price, 2006). However, although these chemical inhibitors have a dominant target enzyme, these earlier studies are compromised by the significant non-specific targeting of other CDKs, such as CDK1, CDK2, and CDK7 (Albert et al., 2014).

In recent times, an analog-sensitive (AS) kinase approach was created to provide precise, rapid inhibition of a particular kinase (Lopez et al., 2014). A CDK9 analog-sensitive cell line was generated via CRISPR/Cas for Raji B-cells by our collaborator (Gressel et al., 2017). This engineered CDK9 kinase contains a mutated ATP binding pocket that allows an ATP analog-molecule to bind the pocket (Figure 12|). Both wild-type and analog kinases can freely bind and conduct kinase activity with endogenous

ATP. However, The CDK9 analog-sensitive kinase also allows the ATP analog to bind. Once the ATP analog was applied to the cellular environment, it competes with endogenous ATP for the ATP binding pocket of the analog kinase, thus achieving rapid and specific CDK9 inhibition that does not affect other kinases. Our prior studies utilizing this cell line found that promoter-proximal pausing increased after 15 minutes of CDK9 inhibition, culminating in global transcription downregulation (Gressel et al., 2017).

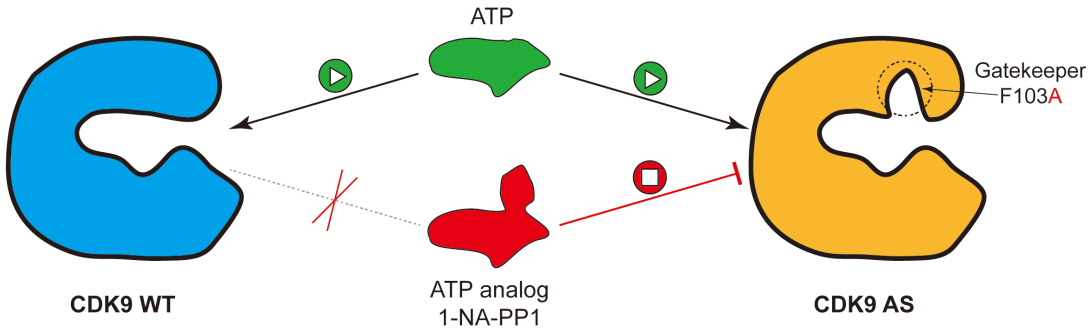


Figure 12| **Targeting strategy of analog-sensitive CDK9.** Only in F103A mutant cells, the 1-NA-PP1 inhibitor can compete with endogenous ATP docking at an ATP-binding pocket. Whereas, in WT cells, an ATP analog could not bind CDK9.

Combining mass spec accessible CTD variants and CDK9 analog-sensitive cells allows us to address the phosphorylation patterns changes after CDK9 inhibition.

1.9 Aim and scope of present work

CTD kinases and phosphatases generate discrete phosphorylation patterns that change throughout the transcription cycle, and antibodies raised against each CTD-specific phosphorylated amino acid were utilized to demonstrate that five different residues in the CTD heptad repeat are phosphorylated in vivo. However, due to the limited specificity and varying avidity of antibody reagents, determining the in vivo stoichiometry of CTD phosphorylation has proven difficult. Indeed, studies exhibited a heptad (or location) specific phosphorylation and function of Pol II CTD between the tandem repeats (Babokhov et al., 2018; Pinhero et al., 2004). Our previous study combined a genetic and mass spectrometry approach to map phosphorylation sites on each heptad (Schüller et al., 2016). Results show that each heptad has a unique phosphorylation pattern, indicating that the CTD heptads may all have different phosphorylation patterns in a sub-region and that these patterns are placed by different enzymes.

A specific kinase inhibition strategy should be adopted to evaluate the phosphorylation changes on each heptad by a single kinase. To achieve specific inhibition on CDK9 in cells, our collaborator Weihua Qin (Leonhard lab at LMU) generated a CDK9 analog-sensitive cell with CRISPR/CAS9 (Gressel et al., 2017). CDK9 was engineered to accept analogies of ATP that are not efficiently used by wild-type kinase. This engineered kinase, referred to as "analog-sensitive" or "-as," is sensitive to protein kinase inhibitor variants that do not block the activity of non-mutant kinases.

My goal is to combine the mass spectrometry cleavable CTD constructs with CDK9 analog-sensitive kinase, thus exploring the phosphorylation changes that occur upon selective inhibition of CDK9. Ultimately, this will allow me to identify hotspots within the CTD dramatically affected by CDK9 and study their biological function in follow-up experiments.

2 Results

2.1 The quantitative analysis of Pol II CTD phosphorylation at each single phosphosite in response to CDK9 inhibition

This study builds on a previously described genetically engineered mass spec cleavable CTD construct and CDK9 analog-sensitive kinase. Furthermore, stable isotope labeling with amino acids in cell culture (SILAC) was combined with high-accuracy mass spectrometry for precise CTD phosphorylated peptide quantification. Thus, establishing a recombinant Pol II CTD construct in a CDK9 analog-sensitive cell line.

To combine MS cleavable CTD construct with CDK9 analog-sensitive cells, I introduced five different mass spec cleavable CTD constructs (CDK9as M1 to M5) into CDK9 analog-sensitive cells. Expression plasmids containing recombinant, α-amanitin resistant Pol II Rpb1 with variant CTDs were transfected into CDK9as cells by electroporation. Cells were cultured in RPMI media containing 1 µg/ml tetracycline and selected by adding 1mg/ml neomycin; only successfully transfected cells will survive in the medium. After that, only cells expressing recombinant Pol II adequately grow in the presence of 2 µg/ml α-amanitin in a second selection.

Apart from a CTD truncated Pol II (IIb) type, all five variants expressed hyperphosphorylated Pol II (IIo) and hypo-phosphorylated Pol II (IIa) (Figure 13| A). The transcriptionally active Pol II is represented by the IIo form, while the transcriptionally inactive Pol II is represented by the IIa form. IIb form exists in all recombinant polymerase constructs. In comparison to recombinant WT, all five CTD variants had similar Pol II expression levels and a similar IIo/IIa ratio. CTD variants M1-M5 showed altered migration behaviors in gel electrophoresis, possibly due to genetic manipulation. The IIo forms of all variants migrated slower than wild-type forms, while the IIa form migrated faster (Figure 13| A). Charge variations in the highly mutated CTD result in different electrophoresis positions for the IIo and IIa forms. Overall, the electrophoresis profile of the CTD variants in CDK9 analog-sensitive cells is similar to that of these CTD constructs in WT Raji cells (Schüller et al., 2016).

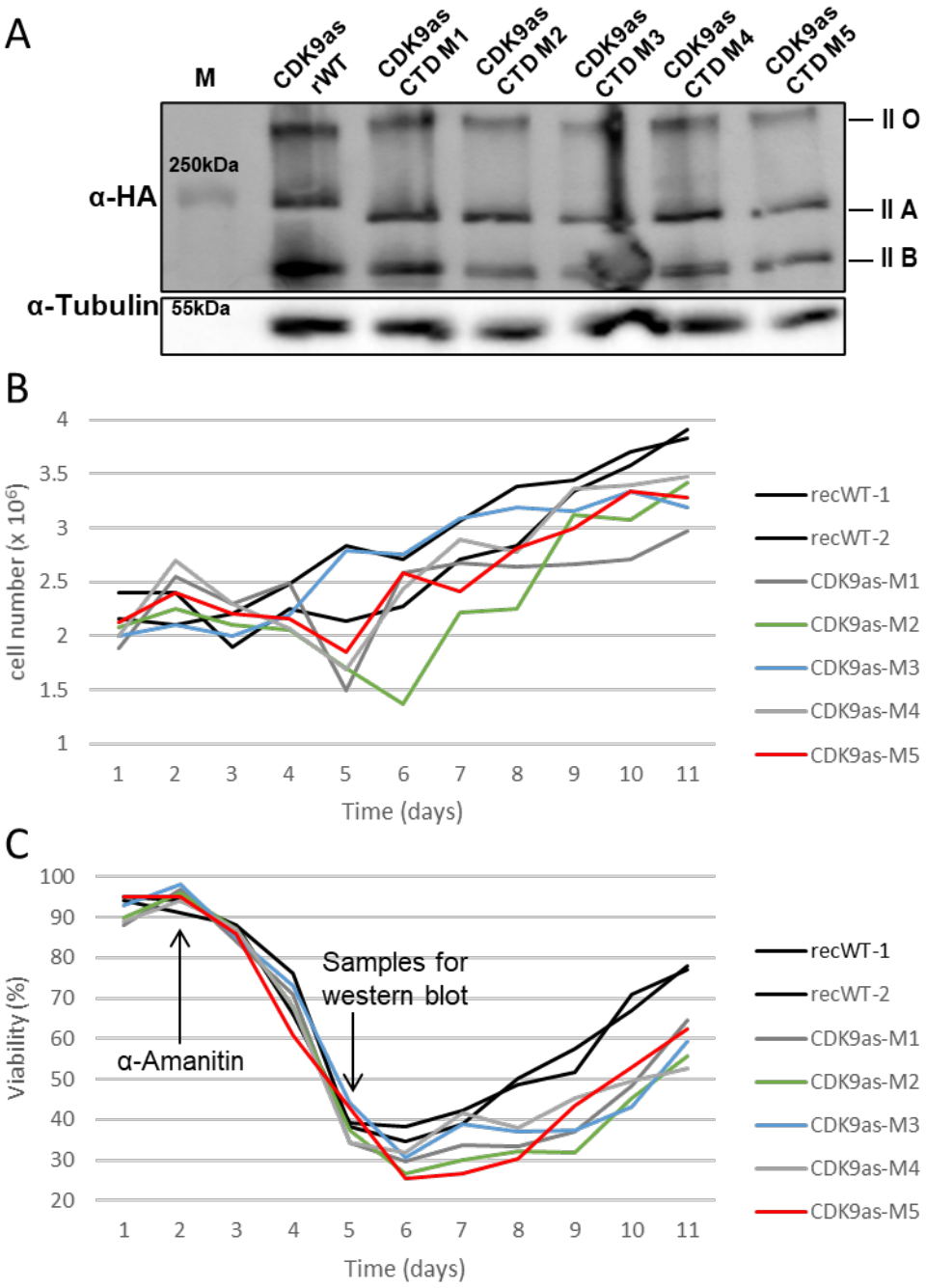


Figure 13| **CDK9 analog-sensitive CTD mass spec variants. (A)** Rpb1 expression of CDK9 analog-sensitive- CTD mass spec variants. Whole-cell lysates were collected after three days under α-amanitin selection. Recombinant WT CTD in CDK9 analog-sensitive cells served as a positive control. Pol II CTD was detected by HA-specific antibodies, and α-Tubulin served as a loading control. Cell proliferation **(B)** and viability **(C)** of CDK9 analog-sensitive CTD mass spec variants. The number of living and dead cells was calculated by using trypan blue staining.

The growth character of CTD variants was monitored daily by counting the numbers of living and dead cells during the 11 days selection period (Figure 13| B-C). Measurement was started when each transfected sample reached 90% viability during neomycin selection. Tetracycline was removed on day 1 to express recombinant Pol II Rpb1. α-

amanitin was added on day 2 to select fully expressed recombinant Pol II (Figure 13| B). Since only those cells that appropriately express recombinant Pol II can survive, all the transfectants showed a decreased cell population and significant drop of cell viability during the following 3-4 days after adding α-amanitin, and slowly increase the population and viability at a later time. Recombinant WT cells showed a faster recovery compared to the mutants. After ten days of α-amanitin selection, all mutants are viable but with slightly lower viability and slower growth rate than recombinant WT. All the transfectants continually recovered up to 90% viability after three weeks of α-amanitin selection (cell counting only covers the first 11 days).

Thus, the mass spec cleavable CTD variants were successfully introduced and expressed in CDK9 analog-sensitive cells. The next step is mapping phosphorylation patterns of CTD in CDK9 analog-sensitive cells after inhibition of CDK9.

2.2 Establishment of SILAC mass spectrometry approach on Raji cell for phospho- CTD quantification.

2.2.1 Isolation and identification of recombinant Pol II CTD by SILAC-MS approach

Quantitative data of an absolute number of individual phosphorylated heptads are required to investigate the phosphorylation changes of CTD heptads when CDK9 is inhibited.

Mass spectrometry approaches that aim to quantitate proteins come with several challenges. The peptide signal strength of mass spec measurement varies from time to time. Furthermore, when parallel samples were processed separately before MS analysis, methodological variability is introduced. Such variability affects the overall consistency among replicates and the background noise between different samples. Only when significant log-fold changes are measured, proteome analyses that describe differences between samples can produce meaningful results. Metabolic labeling using stable isotopes is frequently used for more accurate quantification (Geiger et al., 2011). Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) is a metabolic labeling technique for mass spectrometric (MS)-based quantitative proteomics (Ong & Mann, 2007). With the isotope labeling of heavy nitrogen or carbon, the heavy amino

acids create a distinct and known mass difference between the samples and affect the isotope distribution of the peptides predictably, thereby making data interpretation and quantification more accurate and robust (Hoedt et al., 2019). Trypsin is the most widely used proteinase for mass spec sample preparation and cleaves peptides at arginine (R) and lysine (K) residues. Lys is an essential amino acid. Arg is essential for mammalian cell culture.

Both cannot be converted from other amino acids by cellular metabolism, which can influence labeling efficiency (Eagle, 1955; Zhang & Neubert, 2009). Here, we choose the isotopes of C^{13} and N^{15} on R and K for labeling. Such double labeling ensures every single digested peptide created by trypsin (except the last heptad of the CTD) contains a heavy labeled amino acid distinguished by the mass peak shift. Phosphopeptides present in low abundance when compared to unphosphorylated peptides (Steen et al., 2006). In order to analyze CTD phosphopeptides, an enriched phosphopeptide of CTD is ideal to obtain a successful detection and identification. To achieve this, firstly, the engineered CTD was isolated by immunoprecipitating with HA antibody. Secondly, the CTD phosphopeptides were enriched by TiO_2 beads.

I applied SILAC to quantify the impact of CDK9as inhibition, following a stringent protocol: In our SILAC experiments, two CTD variants M1 and M3, were used (see Materials chapter for sequences), and different times of CDK9 inhibition treatment were performed on each CTD variant. Briefly, CTD variant CDK9as Raji cells were cultured in a growth medium supplemented with amino acids with a stable isotope of C^{13} + N^{15} label for K and R (Heavy-medium). The appropriate concentration was tested in a previous study (Decker et al., 2019). Next, unlabeled cells were grown in a medium with standard amino acids (Light-medium) (Figure 14|). Cells were maintained in SILAC media for five doubling times to ensure that the whole cell population has incorporated the labeled amino acids. Then, the heavy labeled cells were treated with CDK9 inhibitor, and the control cells were treated with DMSO for 15 minutes or 1 hour, right after an equal amount of labeled and control cells were mixed. In parallel, another set of heavy labels was swapped, treated with CDK9 inhibitor and DMSO, then mixed up. The mixed sample containing 300 million cells was lysed. Recombinant Pol II was purified by α-HA, Rpb1 peptides including Pol IIa and IIo bands were excised from SDS-PAGE Coomassie-stained gel. The isolated gel pieces with Rpb1 were trypsin

digested. Each fraction was enriched for phospho-peptides on TiO_2 beads and analyzed by LC-MS/MS (described in the Methods section). Serine, threonine, and tyrosine phosphorylation were searched for in the mapping process. Phosphate groups were identified and localized within the heptad sequences according to the PTM score. Data were normalized to correct for uneven sample mixing. Four sets of mass spec data from two different mutants with two time-points inhibition were collected. Data from six replicates (3+3 swapped labeling) of each experiment were combined into a single experimental analysis.

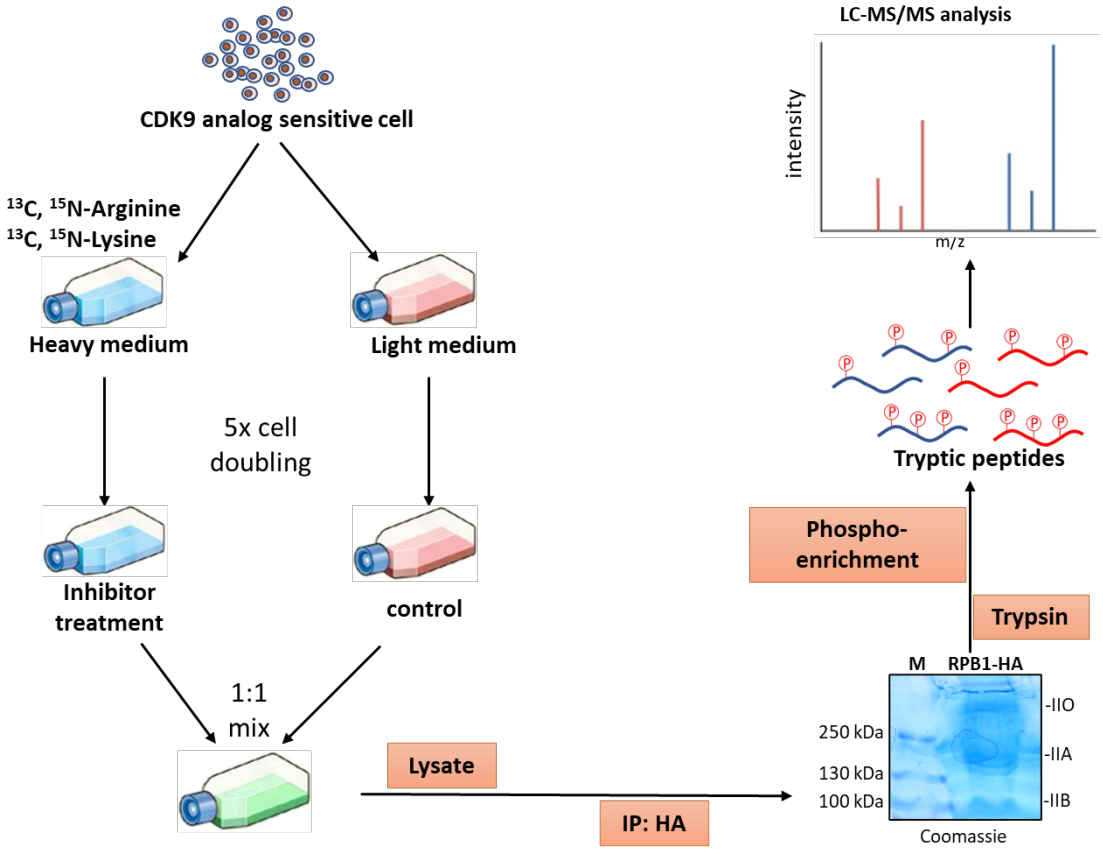


Figure 14| **Workflow of the SILAC (Stable isotope labeling with amino acids in cell culture) labeling.** Schematic outline of the SILAC mass spectrometry approach. CTD variant CDK9as cells are metabolically labeled by growing them in a medium containing either non (light) labeled or $C^{13} + N^{15}$ K and R (Heavy) labeled amino acids. Heavy and light cells are treated with 10µM 1NA-PP1(inhibitor) or DMSO (Ctrl), respectively. Cells from the two experimental conditions are harvested, mixed in a 1:1 ratio, and lysed to obtain a protein extract. Ha-tagged Pol II was precipitated, trypsin-digested and phospho-enriched. A swapped treatment for heavy and light cells is performed in parallel. The heavy and light peptides are distinguishable by mass change shifts.

The data analysis shifted MS absorption peaks using $C^{13} + N^{15}$ labeled K and R, allowing us to distinguish CDK9 inhibition from non-inhibition samples. We can then compare phospho-changes on each heptad with precisely quantified peptides. This

method produces two samples with different isotopic compositions but no other variations in cell growth, sample preparation, or MS analysis, thereby avoiding methodological bias in those steps.

2.2.2 SILAC labeling efficiency

With the SILAC approach, different cell types show different levels of isotope incorporation (Ong et al., 2002). Therefore, before SILAC labeling, it is necessary to evaluate the peptide labeling efficiency with isotope-labeled amino acids under different conditions. Therefore, heavy and nonlabeled arginine and lysine concentrations were first applied as the manufacturer recommended in this optimization assay (seen in methods). Significantly, standard fetal calf serum (FCS) was compared to dialyzed FCS, which was used to prevent contamination with non-labeled counterparts of the SILAC amino acids.

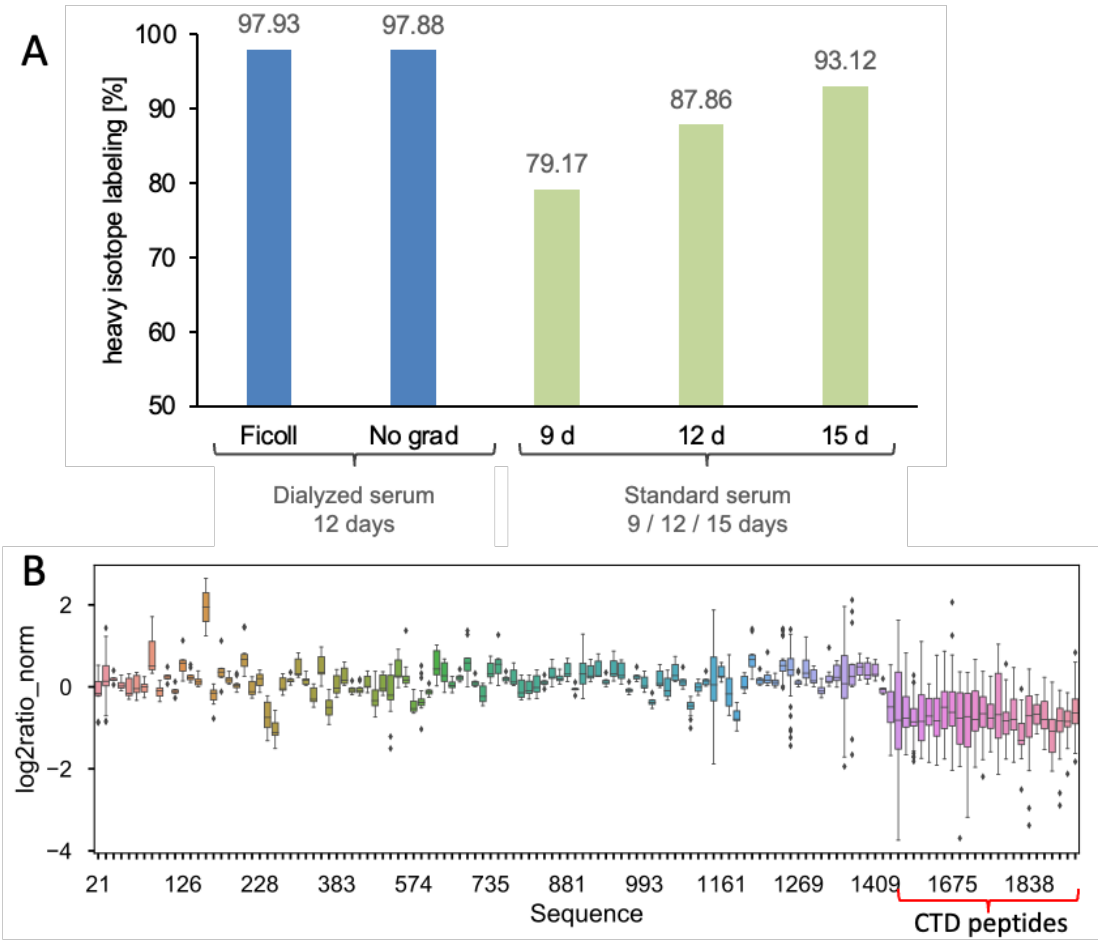


Figure 15| **SILAC labeling efficiency of Rpb1. (A)** General incorporation efficiency of heavy isotope labeling. Cells were labeled and supplied with different sera. **(B)** An overall incorporation efficiency was evaluated by assessment of SILAC ratios between unlabeled and labeled Rpb1. The resulting peptides are analyzed by LC-MS/MS using the Max-Quant software. The proteins are identified based on the MS/MS spectra, while the ratios of the according to SILAC pairs are used for relative quantification.

Our results showed that cultured cells with dialyzed serum show 98% of heavy isotope incorporation within all detected peptides (Figure 15| A), indicating sufficient incorporation of a heavy amino acid after five times cell doubling. Next, we performed SILAC-mass spec analysis of CTD peptides under the same condition, which obtains 98% labeling efficiency. Mass spec data shows complete coverage detection of Rpb1 for peptides from amino acid 1 to 1937. Due to the lack of R or K residues, only the peptide containing the 52nd heptad repeat was not included in the heavy/light ratio (Figure 15| B). Thus, an overall labeling ratio of the heavy and light peptides is close to 1:1. However, the heavy/light labeling ratio of the 3' region that contains the CTD peptides was dropped down. This drop-in heavy/light labeling ratio indicates that the labeling efficiency of the CTD region is affected. Therefore, the amount of heavy labeled peptides is reduced for all CTD peptides.

What causes the Rpb1's heavy/light ratio shift? An altered abundance of Rpb1 due to synthesis and degradation processes might be the underlying cause. So, we addressed the labeling efficiency of two parallel swapped pipelines, the "heavy treated+ light untreated" sample and the "light treated+ heavy untreated" sample (Figure 16| A-B). A similar peptide frequency was observed when comparing samples between heavy and light labeling and samples between inhibited and non- inhibited. Thus, the abundance of Pol II remained unchanged. However, the heavy labeled peptides in both heavy treated and light treated samples showed decreased labeling intensity, whereas control peptides exhibited similar higher labeling intensity in both populations. The previous result showed that the heavy/light ratio dropped down at the C- terminal end of Rpb1. Therefore, we evaluated the labeling efficiency of the CTD part and non-CTD region of Rpb1. The result showed that heavy peptides were missing only for the CTD part of Rpb1 but not for the non- CTD part (Figure 16| C).

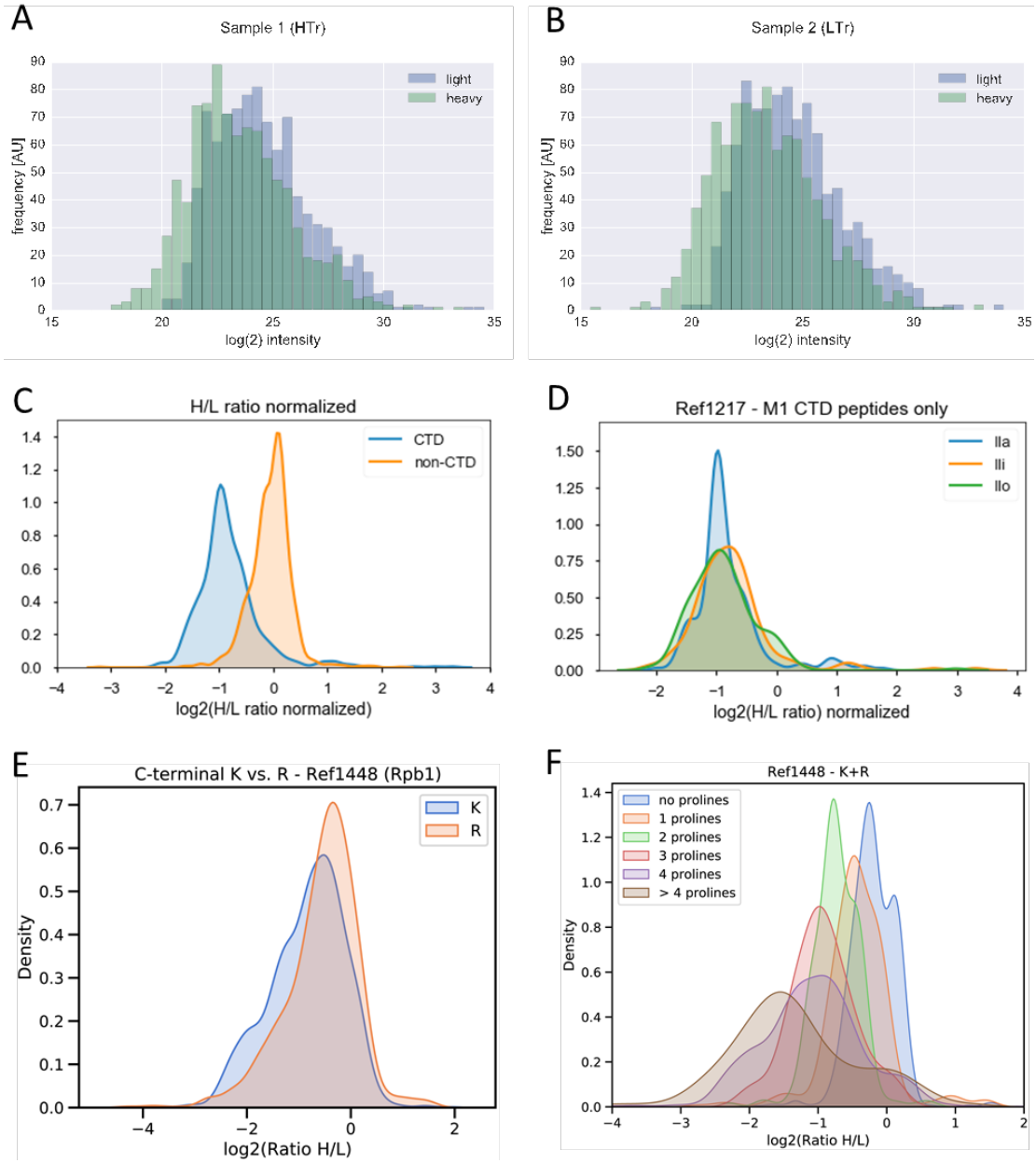


Figure 16| **Labeling efficiency of Pol II CTD.** Overview of Pol II labeling efficiency **(A)** in heavy labeled CDK9 inhibition sample and **(B)** in nonlabelled CDK9 inhibition sample. **(C)** labeling efficiency comparison between CTD peptides and non-CTD peptides. **(D)** labeling efficiency comparison between three Pol II forms of CTD peptides. Hyper- phosphorylated Pol II (IIo), hypo- phosphorylated Pol II (IIa), and intermediate phosphorylated Pol II (IIi). **(E)** labeling efficiency comparison between arginine(R) containing peptides and lysine(K) containing peptides. **(F)** Degree of arginine to proline conversion. Labeling efficiency was evaluated by the amount of arginine to proline conversion.

The labeling efficiency at the 5' part of Rpb1 is relatively standard, while the genetically modified CTD region incorporated less heavy amino acids. One explanation and a common issue when losing labeled isoforms is the turnover of individual proteins (Pratt et al., 2002). Thus, we looked at arginine and lysine individually. A significant shift of lysine labeling was detected, as well as a detectable arginine labeling shift (Figure 16| E). Studies show that arginine can be converted to proline in some mammalian cells

during cell culture, consequently altering quantification accuracy (Ong & Mann, 2007). A set of proline conversions was observed. We found that an increasing amount of incorporation of isotope into proline could contribute to the labeling shift (Figure 16| F).

Nevertheless, we have a set of samples with the same inhibition by swapping isotope labeling, and the heavy/light labeling ratio can be normalized to the labeling-swapped samples. Furthermore, only successfully labeled peptides were taken into account. Therefore, several SILAC labeling experiments were conducted under the same conditions as described above.

2.3 Analysis of SILAC-MS CTD phosphorylation patterns in response to CDK9 inhibition

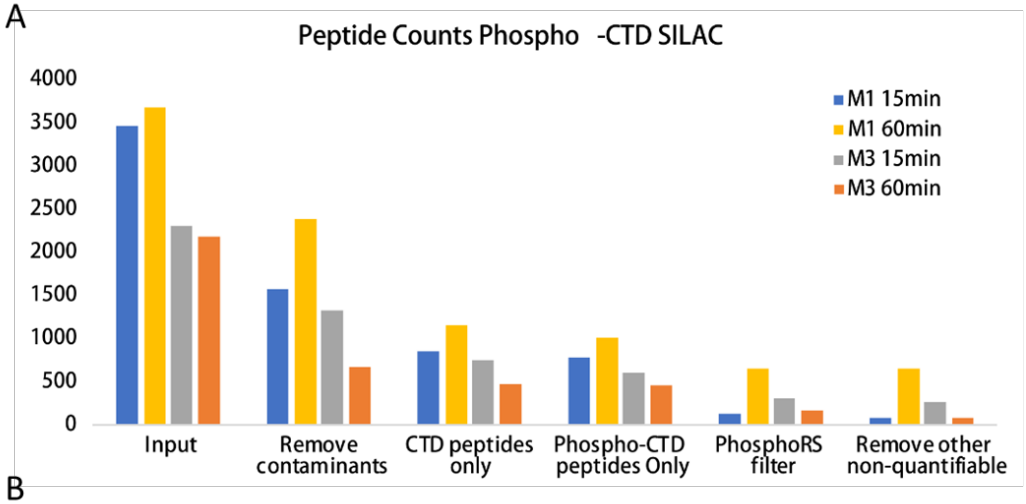
To explore CTD phosphorylation by inhibiting CDK9, the period of inhibition is critical. Long-term inhibition resulted in a reduction of all Pol II transcription (Lam et al., 2001). Nevertheless, flashing inhibition might not be enough to identify the effect on peptide phosphorylation changes. Previous work in our lab found that 1-hour inhibition of CDK9 resulted in a substantial difference in the cellular phospho-proteome and a detectable difference of Ser2P phosphorylation after 2-hour CDK9 inhibition (Decker et al., 2019). With a 15-minute inhibition, a measurable increase in transcription pausing was observed for nascent transcripts (Gressel et al., 2017). Thus, to achieve a consistent result, inhibition with 15 minutes and 1 hour was applied. In addition, a consistent concentration of ATP analog inhibitor (10 M) was utilized. Two CTD variants, M1 and M3, were investigated to minimize the impact of the sequence variation of genetically manipulated CTD. Four experiments were finally conducted, CTD M1 variant with 15 minutes inhibition (M1 15min), CTD M1 variant with 60 minutes inhibition (M1 60min), CTD M3 variant with 15 minutes inhibition (M3 15min), and CTD M3 variant with 60 minutes inhibition (M3 60min).

2.3.1 CTD peptide identification and analysis

MaxQuant software (1.6.17.0) was used to process mass spec raw data (M1 15min, M1 60min, M3 15min, M3 60min) with default settings for peak detection, peptide ranking, mass calibration, and database searches for protein recognition, peptide quantification,

and summary statistics (Tyanova et al., 2015). CTD peptides were first searched in a given line (CTD M1 or M3 variants) and then scored using a probability-based algorithm named peptide score using a target-decoy-based false discovery rate (FDR) approach. Oxidation (M), Acetyl, Phospho (STY), Deamidation (NQ) were set as variable modifications. After each step of filtration of four data sets, accumulation of peptide count was listed (Figure 17|).

A



B

Experiment	Input	Remove contaminants	CTD peptides only	Phospho -CTD peptides Only	PhosphoRS filter	Remove other non-quantifiable
M1 15min	3468	1570	855	784	128	82
M1 60min	3678	2388	1160	1009	654	654
M3 15min	2303	1330	747	606	308	263
M3 60mn	2181	670	469	459	168	84

Figure 17| **Data analysis clean-up pipeline.** 4 experiments were conducted on CDK9 analog-sensitive mass spec M1 and M3 variant, with CDK9 inhibition for 15 minutes and 60 minutes. **(A)** Bar plots showing enrichment of peptide count with different filtration, **(B)** the number of peptides passes through each filtration.

A significant amount of peptide contamination was found in all samples, especially in M1 15min and M3 60min (Figure 17| A). Calculation of phosphorylation site probabilities in CTD peptides was applied using phosphoRS (Tyanova et al., 2015). A large drop of phospho- CTD was observed in three samples except for M1 60min, indicating a significant amount of phosphosites could not be identified. A phosphoRS site probability cutoff, 654 phosphorylated CTD peptides, was identified, accounting for 64.8 % of the peptides identified without any false localization rate (FLR)-based filtering. The other three samples had a much lower identification rate, 16.3% (M1 15min), 50.8% (M3 15min), and 36.6% (M3 60min) (Figure 17| B). In the end, phosphosites identified CTD peptides were counted of each sample with a number of 82 (M1 15 min), 654 (M1 60 min), 263 (M3 15 min), and 84 (M3 60 min). For specific

phospho- site quantification analysis, only the M1 60 minutes inhibition sample, which includes 654 phosphoRS identified CTD peptides, was performed to obtain solid statistical results.

2.3.2 Impact of CDK9 inhibition on CTD phosphorylation

To address the impact of CDK9 inhibition on CTD peptides, an overall CTD phosphorylation profile of CDK9as cells was evaluated. CTD tryptic phosphopeptides were counted in three categories, peptide contains mono- phosphorylation site (1P), peptide contains di- phosphorylation sites (2P), and peptide contains tri- phosphorylation sites (3P).

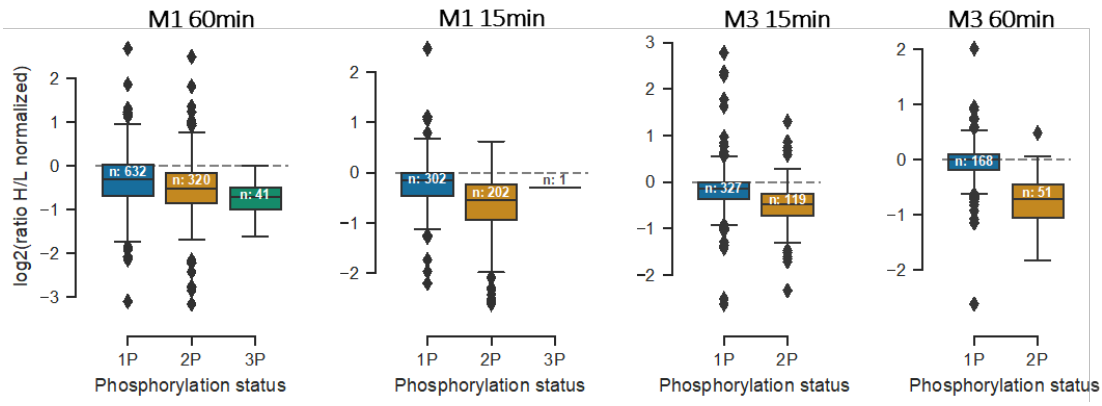


Figure 18| **CTD phosphopeptides distribution.** Overview of phosphorylation changes in mono-, di-, tri- phosphorylation. CDK9 inhibition was performed on CTD M1 and M3 variants for 15 and 60 minutes.

A global decrease of phosphorylated tryptic CTD peptides was observed after CDK9 inhibition in both variants and two different inhibition duration. Peptides that contained tri- and di-phosphorylation sites were primarily decreased. Peptides that carried mono-phosphorylation sites were slightly decreased, in one condition (M3 60min) even unchanged. The phosphorylation state turnover from multiple phosphorylation sites to single phosphorylation may contribute to the minimal change of the mono-phosphorylation population. Interestingly, the abundance ratio of 1P/2P phosphorylation differed between M1 and M3 variants. In CTD M1, 1P: 2P is approximately 1.5-2, which is significantly lower than the CTD M3 sample, which has a 3:1 ratio (Figure 18|). Thus, a transient CDK9 inhibition with 15 minutes and 1 hour on CDK9 analog-sensitive cells is sufficient to observe an accumulative phosphopeptides change of CTD. Further analysis of individual phosphosites can be conducted.

2.3.3 CTD phosphosite distribution in response to CDK9 inhibition

Raji cells were treated with a multiple kinases target inhibitor flavopiridol in Schueller's work for an extended period of treatment (4 hours), demonstrating the phosphorylation changes of CTD are detectable by mass spectrometry, with a significant reduction of Ser2P and a minimum effect of Ser5P and Thr4P. However, due to non-selective targeting of CDK9 by flavopiridol, and indirect effects due to prolonged inhibition, targeting specificity by a particular kinase could not be addressed. With an analog-sensitive kinase and shorter inhibition, the direct impact of a certain kinase can be achieved. To achieve a solid statistical analysis, the M1 60 min sample with the highest number of 654 phosRS identified CTD peptides was taken into phosphosites analysis.

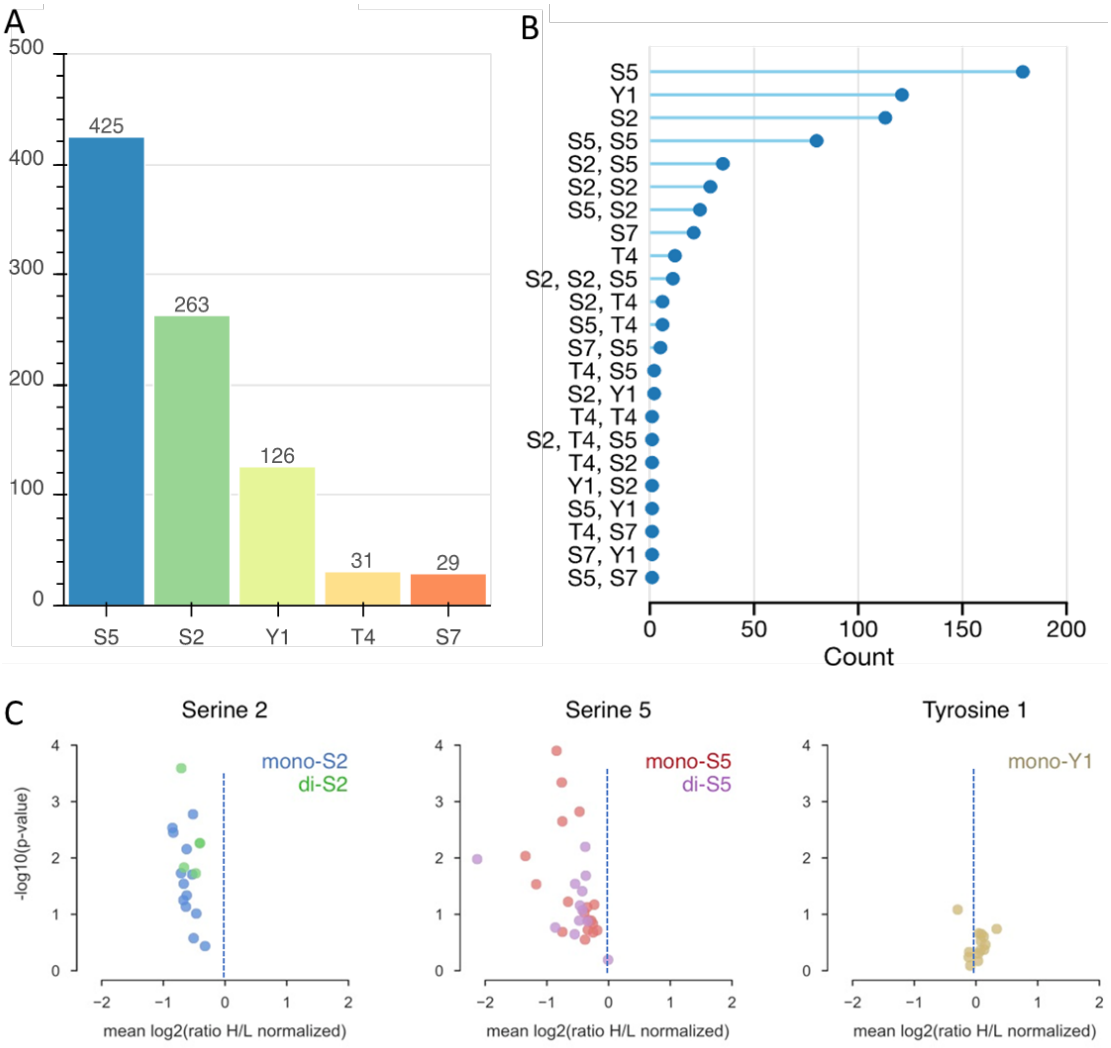


Figure 19| **CTD phosphorylation profile of CDK9 inhibition with specified phosphosites. (A)** Frequency observation counts by a different type of phosphorylation. **(B)** Tryptic peptide counts by different phosphorylation patterns. **(C)** Phosphorylation of Ser2 and Ser5 is decreased by CDK9 inhibition. 1NA-PP1-analog treatment for 1 hour.

Phosphorylation signals on all five potential residues were observed. Phosphorylated Ser5 (48.6%), Tyr1 (30.1%), and Ser2 (14.4%) were the most abundant phosphosites, whereas the phosphorylation on Thr4 (3.5%) and Ser7 (3.3%) took the lowest abundance (Figure 19| A). As previously found (Schüller et al., 2016), most tryptic peptides contained mono- phosphorylation. The most abundant peptides were mono-phosphorylated on Ser5, Tyr1, or Ser2. Whereas only a few mono-phosphorylation on Ser7 and Thr4 were detected. Peptides with di- phosphorylation on Ser5 in two consecutive heptads were also observed at high frequency (Figure 19| B).

Phosphorylation changes on single residues were evaluated. Phosphorylation of Ser2 as a well-known CDK9 target was significantly decreased when CDK9 was inhibited (Bowman & Kelly, 2014; Parua et al., 2018). Interestingly, a decrease of Ser5P is also detected (Figure 19| C). On the other hand, Tyr1 stayed unchanged as a non-CDK9 target.

2.3.4 Phosphorylation patterns on each heptad repeat

The previous study indicated CTD heptad repeats displayed a diverged phospho-pattern across 52 repeats (Schüller et al., 2016). Particular enzymes could modify CTD at different heptads. Thus, it is necessary to evaluate phosphorylation changes on each heptad. The peptide frequency of each tryptic peptide containing mono-, di-, tri-phosphorylation sites were evaluated to analyse phosphorylation changes on each heptad. The M3 CTD variant contains peptides with more tri-heptads, which led to less fragmentation generated by trypsin digestion, which turned out less abundance. M1 15 minutes inhibition and M3 60 minutes inhibition samples were processed in a non-mass spec environmental clean lab, with a large portion of contamination, which resulted in low CTD peptide detection. M3 60 minutes sample was therefore excluded due to non 100% CTD sequence coverage. All other three samples showed 100% coverage of the entire CTD (except the last 52nd peptide). Different amounts of phosphorylated CTD peptides were observed: 446 peptides of M1 15 min sample, 993 peptides of M1 60 min sample, and 505 peptides of M3 15 min (Figure 20|).

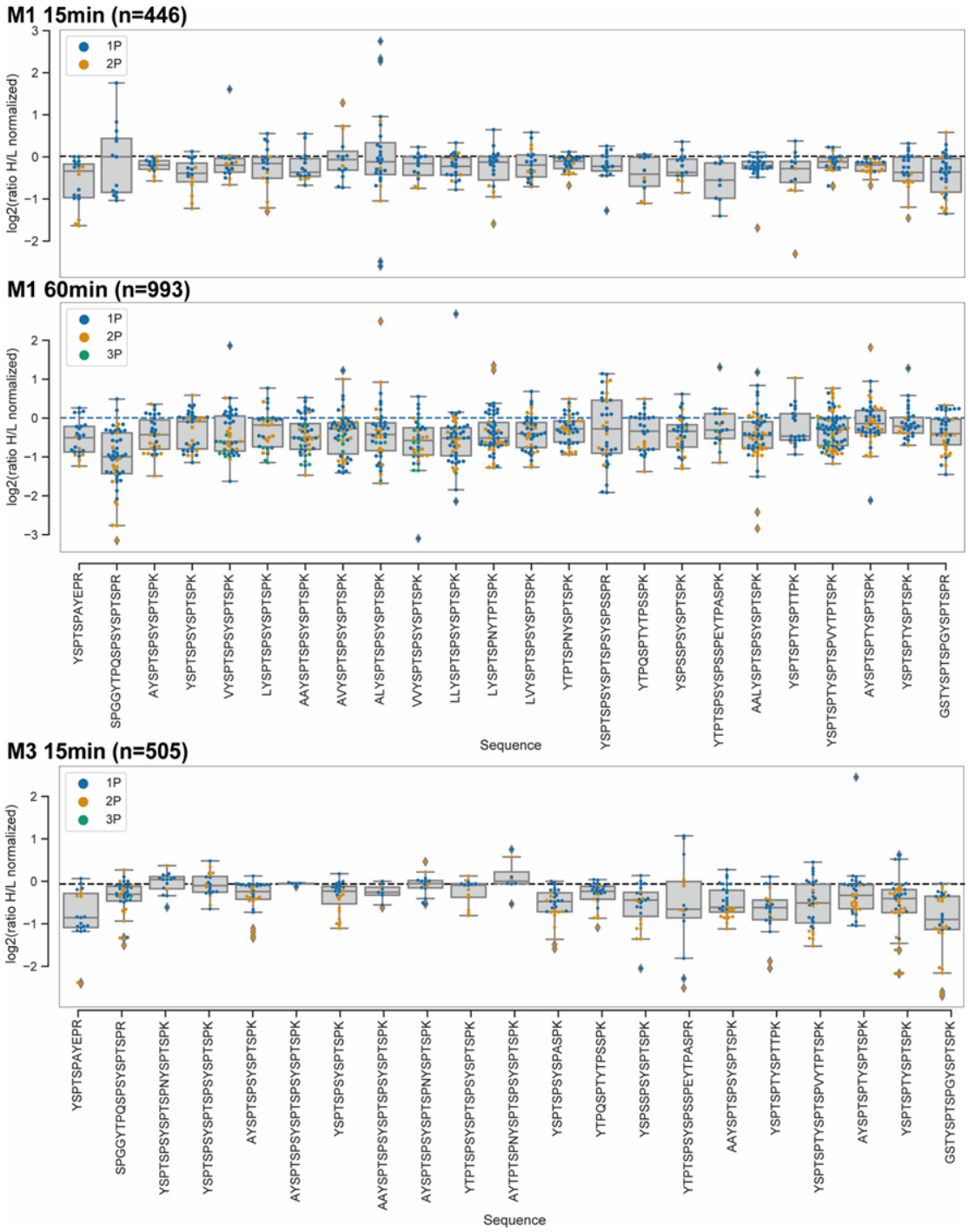


Figure 20| **Differential response of CDK9 inhibition among CTD peptides.** Phosphorylation changes on each tryptic CTD peptide. X-axis: digested CTD peptide from heptad 1 to 52, and each peptide contains 2-3 heptads. Y-axis: normalized log2 ratio of heavy/light phospho- peptides observations (based on ms intensity). Legend dots represent mono-, di-, tri- phosphorylation.

First, the abundance of peptides observed varies in the three experiments. Thus, the peptide chemical property and peptide length may influence the detection rates during the mass spec measurement. Additionally, peptides with lower abundance may stay un-phosphorylated in physiological conditions since our result showed most peptides preferred low phosphorylation status with mono- phosphorylation. Secondly,

phosphorylation in all peptides displayed a reduction in response to CDK9 inhibition in all three experiments. The reduction level diverged at different peptides, indicating that CDK9 prefers specific CTD heptad repeats.

2.3.5 Analysis of CDK9 targeting heptad

Peptides in the M1 60 min experiment with the highest peptide count were taken into more profound analysis. The phosphorylation of the second peptide containing heptad 2, 3 and 4 showed the most significant response to CDK9 inhibition, with around 50% reduction on average (Figure 20|), suggesting a possible CDK9 targeting site. In contrast, the phosphorylation of the second last peptide was almost unaffected.

For the second peptide, 36 phosphorylated peptides were observed. 34 out of 36 peptides contain phosphorylation of Ser5 on heptad 2, 3, or 4 (Figure 21| A). More specifically, phosphorylation forms of Ser5P on heptad 2 and 3 are most abundant and most significantly affected. Heptad 2 and 3 can nearly reach a 90% phosphorylation decrease. In comparison, the phosphorylation of the second last peptide remained almost stable after CDK9 inhibition (Figure 21| B).

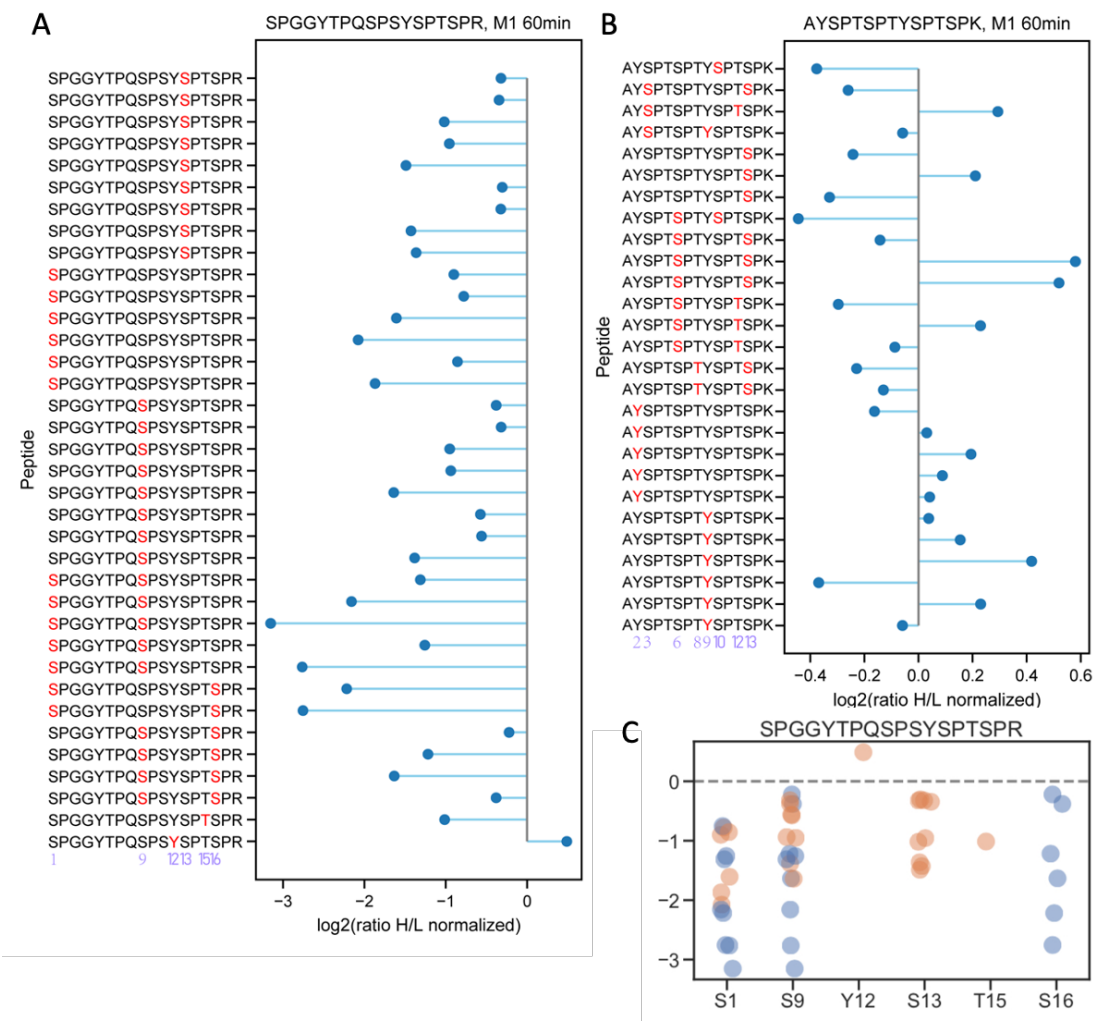


Figure 21| **CDK9 targets CTD.** Phosphorylation changes on each residue of **(A)** the second tryptic peptide and **(B)** the second last peptide. Phosphopeptides were identified by using PhosR. Phosphorylation sites were labeled in red. Normalized log2 ratio of heavy/light phospho- peptides was evaluated for each type of phosphopeptides. **(C)** Accumulation plot of the phospho- residue of the second peptide. The peptide contains mono- phosphorylation labeled in orange, peptide contains di-phosphorylation labeled in blue.

As previous data have shown, CDK9 mainly targets Ser2 and Ser5. The second tryptic peptide contains nine potential phosphorylation sites. First, the phosphorylation log2 change on a single phospho- residue was evaluated (Figure 21| A and C). The most dramatic log2 fold change of phosphorylation is seen for Ser5 of heptad2 (position 1) and heptad3 (position 9), with mono- phosphorylation either on heptad2 or heptad3 and with di- phosphorylation on both heptads, which hits most significant phosphorylation decrease with a number up to 90%. Also, di- phosphorylation of Ser5 on heptad1 and 3 (position 1 and position 16) displayed a dramatic decrease, and it did not take dominant abundance. Thus, the Ser5 of heptad2 and heptad3 are the most affected residues on CTD of CDK9 inhibition. Next, a possible function of these CDK9 potential targets was determined.

2.3.6 Designing serine 5 mutant of CDK9 target

According to our observation of mass spec data, serine 5 on CTD heptad 2,3 and 4 are affected mainly by CDK9. Therefore, these three serine sites could be the target of CDK9 kinase.

To further validate this hypothesis, I generated several CTD mutants with serine 5 to alanine mutations on heptad 2 or 3 (Figure 22|). Three different CTD constructs with Ser5 to alanine mutation on heptad2 or heptad3 were generated. CTD S5A-1 carried a single mutation of S5A on heptad2, S5A-2 carried a single mutation of S5A on heptad3, S5A-3 carried a double mutation of S5A on heptad2 and3. All three CTD mutants were synthesized (GeneArt, Regensburg) and ultimately cloned into the episomal expression vector pRX4-267 (Meininghaus et al., 2000) using a two-step cloning strategy (4.2.1CTD plasmids and cloning strategy).

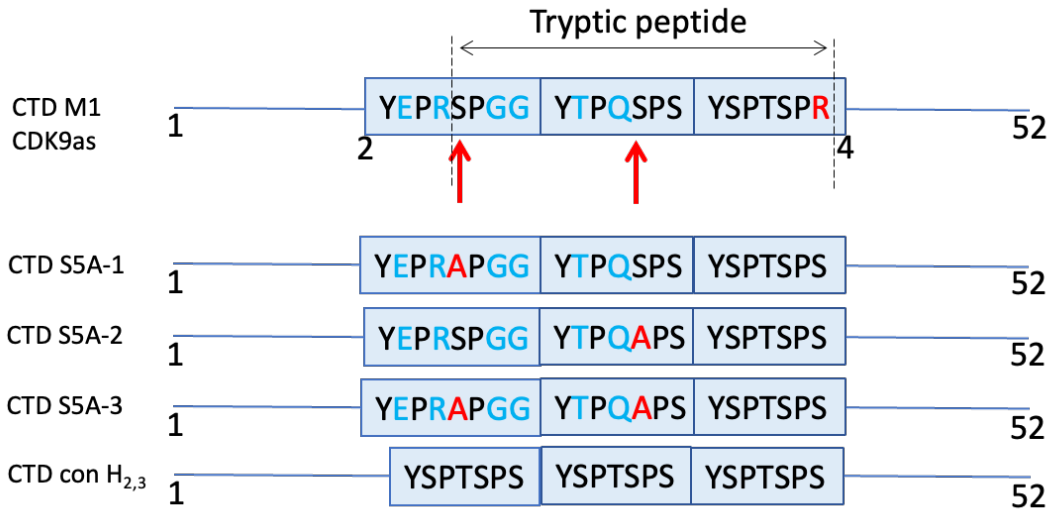


Figure 22| **Schematic representation of serine 5 mutations.** The sequence of heptad 2-4 is shown in the box. The second tryptic peptide of CTD M1 variant is located within the dashed line. Heptads 1, 2, 4, and 52 of the CTD are marked. Ser5 of heptad 2 and 3 is indicated by a red arrow. Three serine to alanine mutants are designed as CTD S5A-1 for single mutation of serine to alanine on heptad2; CTD S5A-2 for single mutation of serine to alanine on heptad3; CTD S5A-3 for double mutation of serine to alanine on heptad2 and 3; CTD con $H_{2,3}$ for consensus heptad replacement of heptad2 and 3.

The pRX4-267 vector includes a full-length, HA-tagged mouse Rpb1 gene with 28 exons (Figure 31|). Exon 28 encodes CTD, which can be substituted by reconstituted CTD. A point mutation in the Rpb1 gene (N793D) provides resistance to α-amanitin (Bartolomei & Corden, 1987). Endogenous Pol II is inhibited in the presence of α-amanitin, enabling an examination of genetically modified Pol II in the cell. The vector uses the Epstein-Barr virus (EBV) as its replication origin and can be held episomally

in human cells that express EBV-nuclear antigen 1 (EBNA1). As a positive control, an expression vector with the WT-CTD sequence was used and was referred to as rWT (recombinant wild-type).

Analysis of these mutants may help us validate the CDK9 targeting sites on Pol II CTD and further study the biological function of these targeting sites.

2.4 Functional study of serine 5 on heptad 2 and 3

To discover more about the function of CDK9 in the phosphorylation of Ser5 in heptads 2 and 3, we examined Ser5 to alanine mutants.

2.4.1 Establishing S5A mutant CTD construct in CDK9as cell

CDK9 analog-sensitive cells were electroporated with the expression vector containing the mutant CTD sequence. To select positively transfected cells, cells were cultured for 2-3 weeks in RPMI growth media containing 1 mg/ml neomycin and 1 µg/ml tetracycline. Expression of the recombinant Pol II was induced by eliminating tetracycline from the medium at cell viability of 90-95 %. 24 hours after induction, cells were cultured in the presence of 2 µg/ml of α-amanitin. 96 hours after the addition of α-amanitin, whole-cell lysates were prepared for western blot analysis. The hypo-phosphorylated (IIa) and hyper-phosphorylated (IIo) forms of Pol II were expressed in all S5A mutants, as well as rWT cell, which was detected using the -HA antibody (3F10) (Figure 23| A). The presence of the Pol IIO and IIa forms in the mutants indicated that these polymerases are transcriptionally functional. Depletion of heptad 1-3 led to Pol II degradation (Chapman et al., 2005), generated a CTD truncated Rpb1 IIb form. To monitor pol II degradation, a mutant containing heptads 2 and 3 were substituted by consensus sequences was involved, referred to as CTD con $H_{2,3.}$ A moderate degraded pol II was detected in S5A-1, S5A-2, compared to complete substitution of CTD con $H_{2,3}$.

Interestingly, the Pol IIa and Pol IIo bands migrated faster than rWT in all mutants, especially in S5A-3. Such point mutation on single and double amino acids may cause an altered CTD modification, further affecting peptide migration behavior in SDS PAGE gel. Some mutants like S5A-3 exhibited a greater Rpb1 expression level. It could

be due to a faster recovery after α-amanitin treatment. All strains showed a similar Rpb1 expression level with full recovery of cell growth after three weeks of α-amanitin selection. The un-transfected Raji cells served as a negative control and did not show any signal, while rWT cells served as a positive control and expressed all types of Pol II. The Rpb1 antibody can detect endogenous and recombinant Pol II expression (Pol 3.3). Only recombinant Pol II was observed in transfected cells, indicating endogenous polymerases were largely eliminated. The phosphorylation of Ser2 and Ser5 on the CTD was probed with an S2P antibody (3E10) and S5P antibody (3E8). All the mutants exhibited a similar phosphorylation level of Ser2 and Ser5, indicating a fully functional polymerase during the transcription process. Thus, a set of S5A mutant cell lines was successfully generated.

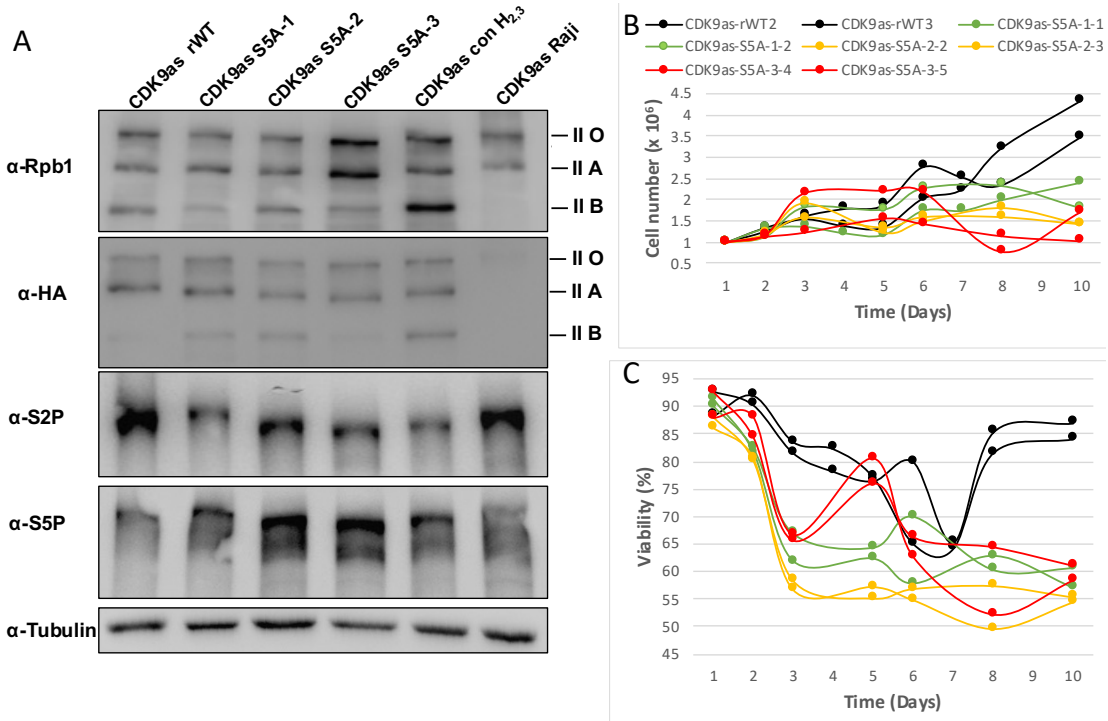


Figure 23| **RNA Pol II expression and cell growth of CTD S5A CDK9 analog-sensitive mutants. (A)** Rpb1 expression of CTD S5A CDK9as mutants. Whole-cell lysates were collected after four days under α-amanitin selection. rWT and un-transfected CDK9as Raji cells served as a positive and negative control, respectively. Heptads 2 and 3 were replaced by consensus sequences in CTD con $H_{2,3}$. Pol II CTD was detected by antibodies using Rpb-1, HA, Ser2-P, Ser5-P, and α-Tubulin served as a loading control. Cell proliferation **(B)** and viability **(C)** of CTD S5A CDK9as mutants. The number of living and dead cells was calculated by using trypan blue staining.

The CTD S5A mutant's growth profile was assessed using cell number and viability (Figure 23| B-C). First, tetracycline was washed out to induce the expression of the recombinant polymerase. Then, 24 hours after induction, α-amanitin was applied. Next, the number of living and dead cells was measured daily by Trypan-blue staining.

Finally, the total number of living cells was multiplied by the splitting factor by which the culturing cells were divided over the period of measurement.

Graphs showed that all the strains display a decrease in cell number and cell viability after adding α-amanitin. α-amanitin 100% inhibits transcription by endogenous Pol II and only allows transcription conducted by recombinant Pol II. Such a transition from endogenous to recombinant polymerases dramatically impacts basal cellular activities and takes time for cells to recover from this transition. After one week of α-amanitin selection, recWT shows a recovery in cell viability (Figure 23| C), whereas the mutants cell remained at a low viability level. Similar to total cell number, all mutants displayed a slower growth rate compared to recWT. After two weeks of α-amanitin selection, all mutants were viable and could continuously proliferate. However, the growth velocity of all three mutants was slower than recWT.

The S5A mutants were viable and recombinant polymerase was stable and transcription active. Next, a functional analysis was conducted to analyze these CDK9 targeting sites.

2.4.2 Sensitivity of S5A mutant cell to CDK9 inhibition.

To have maximum targeting effects, we only took S5A-3, which has a double mutation on both heptads.

Here, we treated S5A-3 mutant and recWT cells with two different concentrations of CDK9 inhibitor 1-NA-PP1 for a time series of up to two hours. Cells were then released from inhibition by wash-out inhibitor, and samples were collected at different time points as indicated (Figure 24| A-B). S5A-3 mutant and recWT both displayed a phosphorylation decrease of Pol IIo form by HA and S5P antibody when CDK9 inhibition was implied in two concentrations. A recovery signal of increased IIo form was shown after releasing from CDK9 inhibition. However, the Pol IIo form did not show a significant difference between recWT and S5A mutant from the western blot, although both displayed response to CDK9 inhibition.

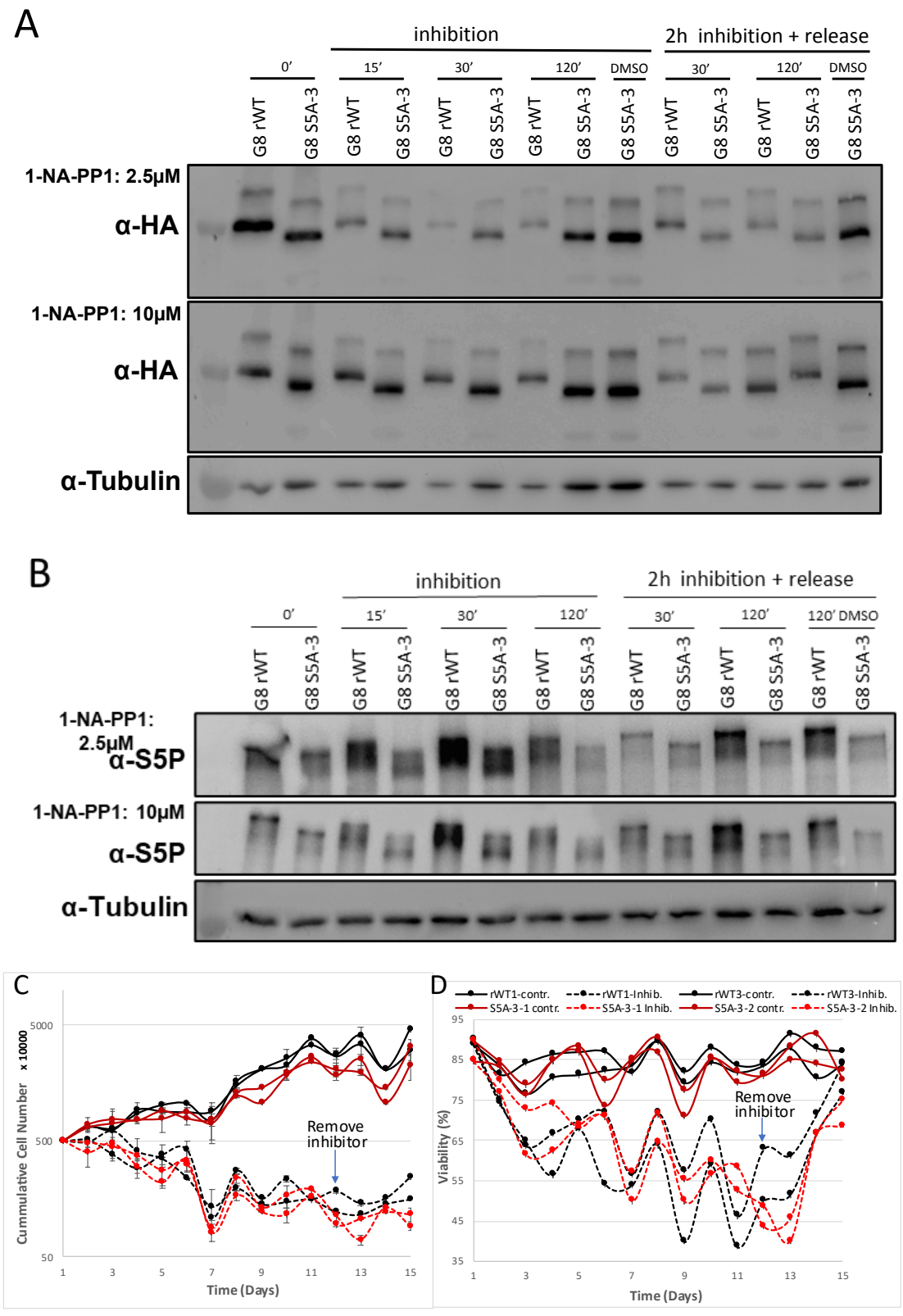


Figure 24| **CDK9 inhibition of CTD S5A CDK9 analog-sensitive mutants.** Rpb1 expression of CTD S5A mutants with CDK9 inhibition. rWT and mutant cells were treated with two concentrations of 1-NA-PP1 and DMSO. Cells were treated in 15, 30, 120 mins, and 30, 120 mins release of washing out 1-NA-PP1 after 120 mins treatment. Whole-cell lysates were collected right after each particular time point. Pol II CTD was detected by antibodies using HA **(A)**, Ser5P **(B),** and α-Tubulin served as a loading control. Cell proliferation **(C)** and viability **(D)** of CTD S5A CDK9as mutants. The number of living and dead cells was calculated by using trypan blue staining.

Next, we asked whether a difference appeared in the cell growth of the S5A mutant in long-term CDK9 inhibition. A proliferation assay in the presence of ATP analog 1-NA-PP1 was performed. The stably transfected cells were treated with 2.5 μM of 1-NA-PP1 for 11 days, 1-NA-PP1 was then washed out, and cells were cultured in a standard growth medium. The cell number was measured every second day to detect the long-term impact of CDK9 inhibition. The cell numbers continually decreased while applying CDK9 inhibition, both in S5A mutant and recWT cells. However, there is no significant difference between the two cell lines (Figure 24| C-D). Cells were then released from CDK9 inhibition, showing an increasing cell number and a recovery in cell viability, but again, no significant difference between mutant and recWT cells.

Double mutations of S5A on heptad2 and 3 did not modulate the function of CDK9 on CTD phosphorylation. And no significant difference in sensitivity to CDK9 long-term inhibition. S5A mutant did not show any difference in the sensitivity in western blot and growth behavior. Considering the redundancy of CTD with 52 repeats, altering 2 of 156 serines may not give a distinct phenotype. To access the direct output of Pol II CTD, a transcriptome analysis was carried out.

2.4.3 Total RNA-seq analysis of CTD S5A mutant

Next, the effect of S5A mutation on transcription was analyzed. Looking at nascent transcripts may not be ideal for studying the mutant polymerase, as processed RNAs are excluded from such an analysis To study a possible phenotype of the S5A mutant, a whole transcriptome analysis was performed.

Total RNA-seq was performed on CDK9 analog-sensitive (CDK9as) Raji cell with double serine 5 mutation on CTD heptad 2 and 3. CDK9as cells with recombinant WT CTD served as control. Cells were collected and lysed in Trizol. In cooperation with Stefan Krebs (group of Helmut Blum, Gene Center), a total RNA-seq was performed on CDK9 analog-sensitive (CDK9as) Raji cell with double serine 5 mutation on CTD heptad 2 and 3. Five biological replicates were prepared for each sample. Cells were collected and lysed in Trizol.

RNA-seq data was analysed by Michael Kluge (group of Caroline Friedel, LMU). A standard principal component analysis (PCA) sample quality check was used to

determine sample similarity among the five biological replicates and two principal cell groups. A distinct separation was shown between two groups of samples (Figure 25| A). Furthermore, five replicates of the recWT sample were separated from other samples and mapped together in a group at the PC1 dimension. By contrast, five S5A-3 mutant samples exhibited more deviation in a group. MT3 displayed a significant variation compared to other replicates of S5A-3 mutant with negative relevance on PC2 dimension. So MT3 was considered as an outlier and was eliminated for further analysis.

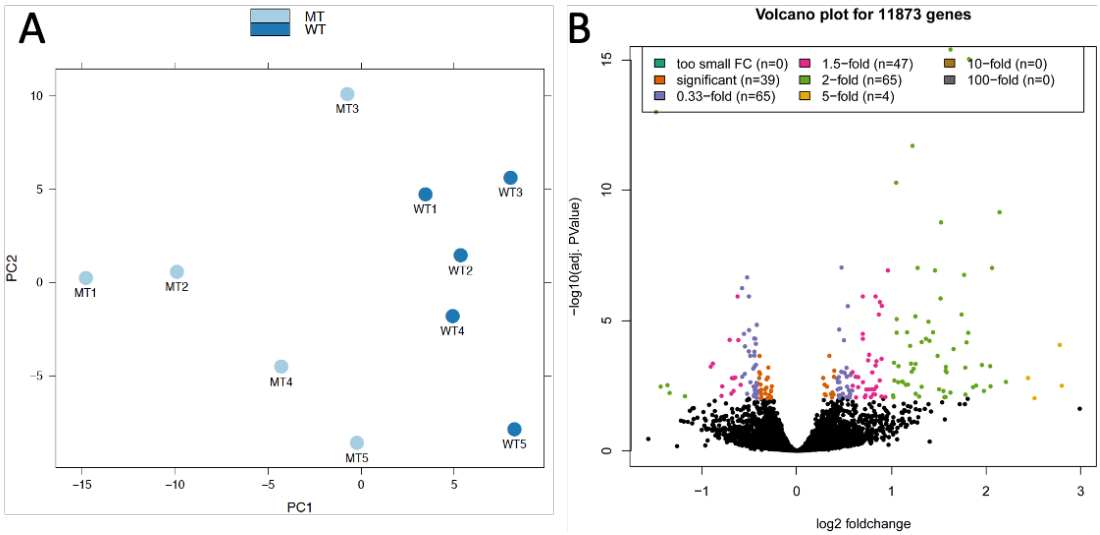


Figure 25| **Principal component analysis and differential gene expression analysis of S5A mutant. (A)** Principal component plot analysis of RNA-seq data. Gene expression changes were investigated in rWT and S5A-3 mutants. Each group has five replicates, WT1-5 and MT1-5. **(B)** Volcano plots of differential gene expression levels with log2 fold change (*x*-axis) against -log10 (adj. *P*-value, *y*-axis) in rWT and S5A -3 mutant. DEGs (fold change with significant, 0.33 fold, 1.5 fold, 2 fold, and 5 fold, $P < 0.01$) are orange, purpure, pink, green, and yellow.

Differential gene expression (DGE) analysis was conducted for recWT and S5A-3 mutant. Out of 11873 genes, applying logFC and sorting for up and down-regulated genes were sorted, total of 220 significantly differentially expressed genes were obtained, of which 142 genes were increased in abundance. In comparison, 78 genes were decreased in abundance (Figure 25| B). All the decreased genes had a log2 fold change value, not excess -1. The most significantly changed genes showed an up-regulation in the S5A-3 mutant cell line with log2 fold change value up to 2.79 (Table 1).

Thus, the S5A-3 mutant cell exhibited a difference in transcriptome outcome. 220 differential expressed genes were observed in the S5A-3 mutant cell. In addition, the

most significant DE genes were upregulated.

2.4.4 Functional enrichment analysis of differentially expressed genes

Next, to better understand these DE genes' biological function, the significant pathways of DE genes were analyzed. Thus, the Gene set enrichment analysis (GSEA) on differentially expressed genes in S5A-3 mutant versus recWT transcriptome was applied (Subramanian et al., 2005). The Molecular signatures database (MSigDB) was applied, and the genes were filtered by hallmark gene sets or using Gene Ontology (GO) categories to identify significantly enriched or depleted groups of genes (Liberzon et al., 2015).

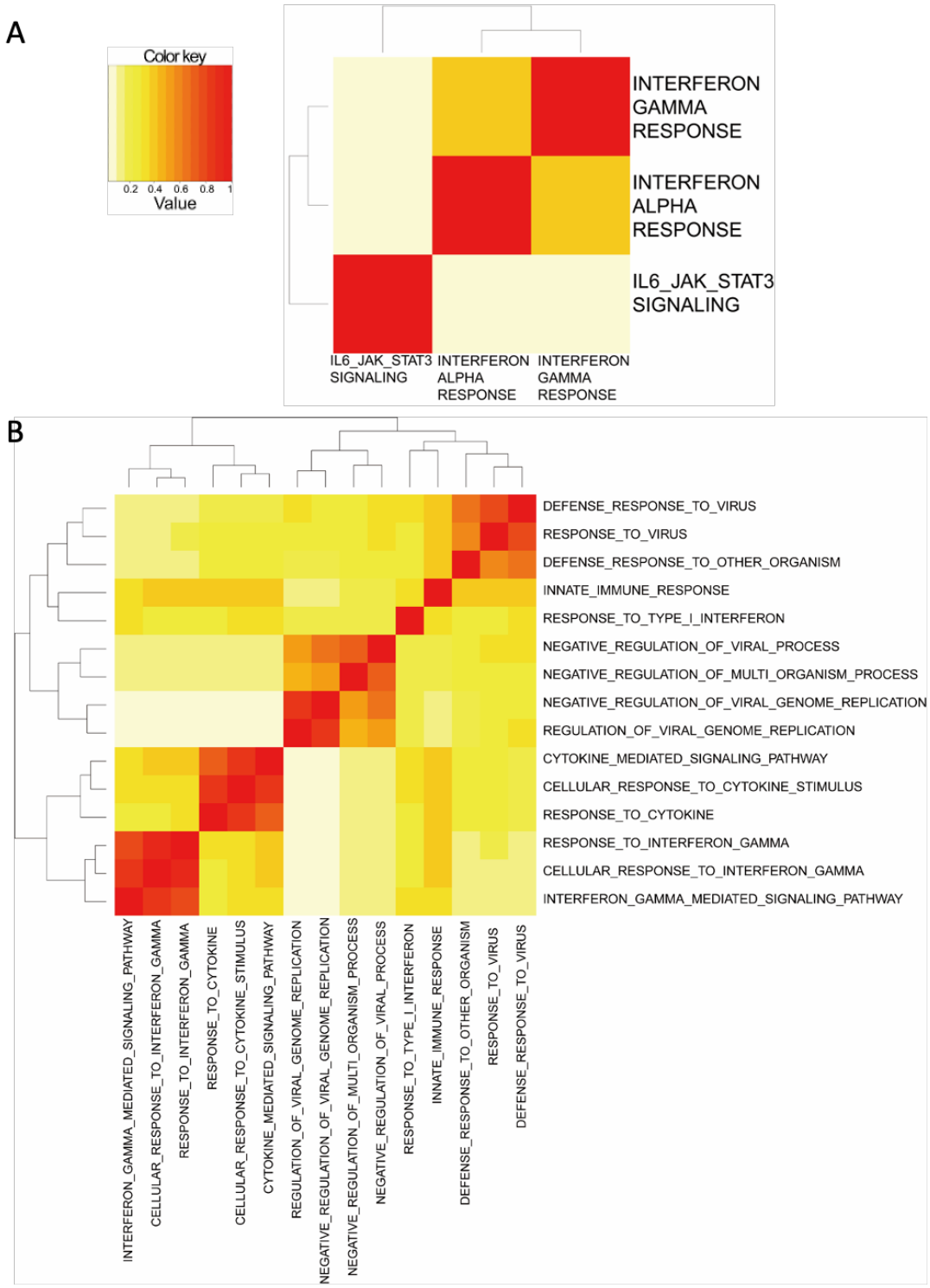


Figure 26| **Gene set enrichment analysis (GSEA) of CTD S5A-3 mutant differential expression genes. (A)** GO- GSEA analysis. **(B)** Hallmark- GSEA analysis.

Three upregulated signaling pathways were identified in the S5A-3 mutant sample by GO classes, including interferon-gamma response (78 genes), interferon-alpha response (49 genes), and IL6-JAK-STAT3 (19 genes) signaling (Figure 26| A) (data of gene numbers is not shown). A comparison of interferon alpha response and gamma response genes indicated that a high number of shared genes was enriched in both

pathways. Thus, an overall interferon response stimulation was found in the S5A-3 mutant. Furthermore, the hallmark analysis provided a more precise picture of the significant genes related to functional groups. The top gene groups in differential gene expression related to interferon-stimulated genes, respond to cytokine and to the virus infection (Figure 26| B).

Interestingly, most DE genes were functionally related to immune response, displaying a distinct group of differentially expressed genes in the S5A-3 mutant. The gene list demonstrated the occurrence of upregulated expression of genes involved in IFN signaling (Table 1).

2.4.5 Interferon stimulated gene (ISG) - the most dysregulated genes

Overview of the most significant DE genes, genes with log2 fold change > ±1.5 were listed (Table 1).

name	Gene description	log2FC	aveLog2CPM	PValue
IFI27	interferon alpha inducible protein 27	2.79	407.01	3.53E-05
MX2	MX dynamin like GTPase 2	2.77	165.45	3.59E-07
IFIT1	Interferon Induced Protein With Tetratricopeptide Repeats 1	2.51	192.91	1.76E-04
IFI6	interferon alpha inducible protein 6	2.44	212.63	1.42E-05
DDX60	DExD/H-box helicase 60	2.21	108.49	2.18E-05
DHRS7	Dehydrogenase/Reductase 7	2.14	47.77	3.63E-13
DHRS2	Dehydrogenase/Reductase 2	2.06	56.92	8.21E-11
SAMD9L	sterile alpha motif domain containing 9 like	2.04	439.75	3.77E-05
PI16	Peptidase Inhibitor 16	2.04	26.17	3.45E-06
IFI44L	interferon induced protein 44 like	1.97	3303.13	7.06E-05
OASL	2'-5'-oligoadenylate synthetase like	1.95	46.72	2.99E-06
XAF1	XIAP Associated Factor 1	1.89	29.99	3.42E-05
SMIM38	Small Integral Membrane Protein 38	1.86	28.54	4.39E-05
PROM1	Prominin 1	1.82	109.46	1.63E-19
CXCL9	C-X-C motif chemokine ligand 9	1.81	550.75	8.86E-08
LAMA5	Laminin Subunit Alpha 5	1.79	31.86	2.76E-07
STRA6	Signaling Receptor And Transporter Of Retinol	1.78	23.69	1.11E-04
AL158136.	lncRNA	1.78	202.68	4.20E-06
SSXP10	processed_pseudogene	1.76	83.99	1.98E-10
GRID1	Glutamate Ionotropic Receptor Delta Type Subunit 1	1.74	80.55	1.23E-08
AC104825.	lncRNA	1.65	47.02	5.51E-07
IRF4	interferon regulatory factor 4	1.62	145.77	3.49E-20
STAT1	signal transducer and activator of transcription 1	1.62	1785.73	8.37E-05
TCTE1	T-Complex-Associated-Testis-Expressed 1	1.58	40.06	7.12E-06
AC027801.	lncRNA	1.57	34.82	3.86E-06
CSF1R	Colony Stimulating Factor 1 Receptor	1.57	59.69	5.93E-06
LINC02842	lncRNA	1.57	32.34	1.51E-04
AC104823.	lncRNA	1.56	42.30	1.55E-04
KLKB1	Kallikrein B1	1.55	27.36	7.13E-05
GABRA5	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha5	1.52	157.29	1.03E-12
CCL17	C-C Motif Chemokine Ligand 17	1.51	118.52	2.43E-09

Table 1| **List of differential expression genes.** Genes with log2 fold change values over ±1.5 were shown in the list. Interferon stimulated genes were highlighted in yellow. The estimation of expression levels of a transcript is transformed to average log number of reads Counts Per Million mapped reads (aveLogCPM). $P < 0.01$.

15 out of 31 most dysregulated genes (lncRNA, pseudogene excluded) were interferon-stimulated genes (highlighted in yellow). Surprisingly, the top five dysregulated genes were all interferon-stimulated genes: Interferon-alpha inducible protein 27 (IFI27), MX dynamin-like GTPase 2 (MX2), a typical ISGs- Interferon Induced Protein With Tetratricopeptide Repeats 1 (IFIT1), Interferon-alpha inducible protein 6 (IFI6) and DExD/H-box helicase 60 (DDX60).

2.4.6 Interferon stimulated gene expression

Since a significant portion of differentially expressed genes can be assigned to the interferon response, I aimed to study further the connection between the S5A-3 CTD mutation and increased levels of interferon-stimulated gene (ISG) transcripts. One possibility could be that mutant CTD displays an increased affinity to the ISG promoter region. ISGs have conserved elements of ISRE or IRF-responsive element at the promoter (Michalska et al., 2018; Mostafavi et al., 2016). Thus, besides other co-factors, the Pol II could specifically initiate ISGs transcription. Another possibility is that S5A-3 mutant polymerase produces aberrant double-stranded RNA (dsRNA). The formation of endogenous dsRNA may trigger a cellular antiviral response and interferon signaling (Chung et al., 2018).

First, we checked secreted IFNs to determine whether IFN signaling is functionally active or not. Secreted IFNs of CTD S5A mutant and WT cells were measured by ELISA (Figure 27| A-B). IFN secretion was not detected in mutant or WT cells (Figure 27| A). It was only detected after dsRNA stimulation, indicating that IFN signals were not produced in S5A-3 mutant cells, implying that the downstream regulation of ISGs was involved in regulating the IFN signal.

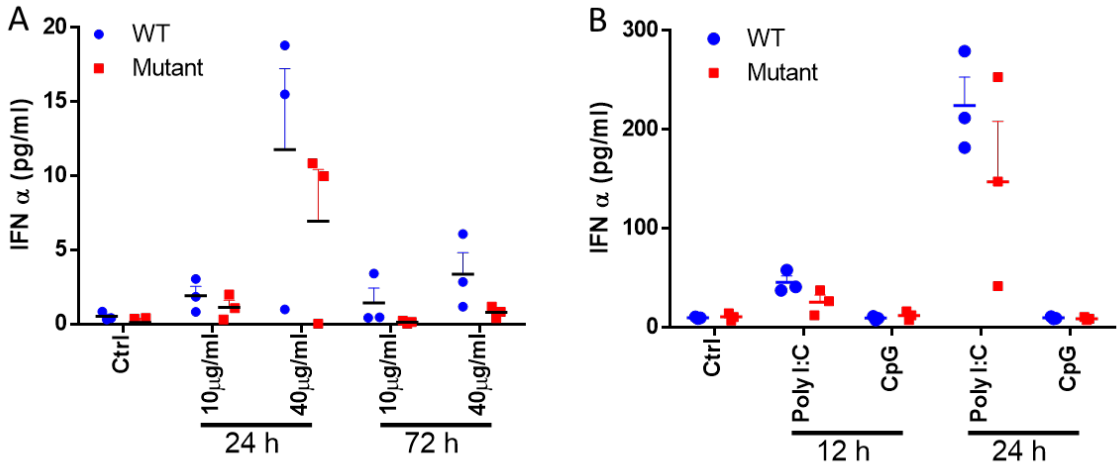


Figure 27| **Interferon and aberrant RNA processing. (A)** ELISA of interferon-alpha with Poly I: C induction and **(B)**. ELISA of interferon-alpha with Poly I: C and CpG.

3 Discussion

This study combined a trypsin cleavable CTD construct with an analog-sensitive cell line to map CTD phosphorylation patterns in response to selective CTD kinase inhibition (here with CDK9). We found that phosphorylation of the second tryptic peptide of CTD was mainly affected by CDK9 inhibition. Furthermore, mutations of the second peptide revealed upregulation of interferon-stimulated genes, indicating a link between CTD heptad2 and 3 and interferon response genes.

3.1 Establishment of CTD phosphopeptides quantification system

3.1.1 Combination of trypsin cleavable CTD and analog-sensitive kinase

What is the advantage of using mass spec for evaluating CTD phosphorylation? Phosphorylation-specific antibodies for each phospho-site (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) have been used to measure the dynamics of CTD phosphorylation across the transcription cycle by western blots and ChIP seq (Chapman et al., 2007; Hintermair et al., 2012). However, these antibodies do not differentiate between which heptad is phosphorylated on CTD. The mass spec approach allows for unequivocal detection of modified residue(s) in each heptad and allows for a high-throughput experiment. In addition, it allows to evaluate CTD phosphorylation on every single heptad using a trypsin cleavable CTD (Schüller et al., 2016; Suh et al., 2016). Each CTD heptad displays a divergent and distinct phosphorylation profile, according to prior work in our lab. When applying CTD kinases inhibitor, such as flavopiridol, the phosphorylation changes on each heptad were also different, indicating that a given kinase may have a specific targeting region on CTD (Schüller & Eick, 2016).

Why did we choose analog-sensitive kinase approach to study the function of a certain kinase? Exploring the function of a particular kinase by chemical inhibitor, however, has a limitation. Many kinase inhibitors, including those used as targeted cancer therapeutic agents, are polypharmacological and inhibit multiple targets simultaneously at applied concentrations, such as flavopiridol, a CDK9 inhibitor, inhibiting other

CDKs (Bösken et al., 2014; Greifenberg et al., 2016). Instead, analog-sensitive kinase technology allows for a rapid, reversible, and precise inhibition of specific kinase in cells and living organisms (Bishop et al., 2000). The AS kinase has a space-creating mutation in the kinase's ATP-binding pocket, as well as a bulky ATP-competitive molecule that matches the mutant ATP pocket's shape (Lopez et al., 2014). As a result, the cellular role of individual kinases can be determined. Previous studies found that the ATP-analog molecule inhibited CDK9 very fast after 15 minutes of inhibition (Gressel et al., 2017). Due to the interaction network between the CTD kinases (Devaiah & Singer, 2012), prolonged inhibition of CDK9 may impact other kinases. With CDK9as cells, a short rapid inhibition is ideal for revealing the direct effect of CDK9 on CTD.

There has not been any research into mapping the CTD heptad phosphorylation pattern by inhibiting specific CTD kinases. Therefore, I combined trypsin cleavable CTD and CDK9 analog-sensitive cells. Transfection of engineered CTD constructs into CDK9 analog-sensitive cells was successful. Cells were viable and proliferated a bit slower than wild-type cells (Figure 13|). It enabled us to map phosphorylation patterns on each heptad with a more direct and specific CDK9 inhibition.

3.1.2 SILAC labeling for CTD phosphopeptides quantification

Stable Isotope Labeling by/with Amino acids (SILAC) approach is often utilized to quantitatively analyze phosphopeptides. A significant advantage of this method is that cells from different groups can be combined and analyzed together by mass spec (Mann, 2006). However, it also faces challenges statistically and technically (Solari et al., 2015), especially when applied to CTD phosphopeptides. CTD phosphorylation is a dynamic reversible post-translational modification (PTMs) by CTD kinases and phosphatases (Introduction 1.4). During the transcription, molecules interact with chromatin and transcription complex within a time scale of milliseconds (Ranjan et al., 2020). Within this time frame, the CTD code is read, written, and erased. It turns out that phosphopeptides have transient and limited stoichiometry. Therefore, to enrich, detect, and identify these phosphopeptides, many shortfalls must be solved.

Several amino acids can be used for SILAC labeling. In addition to arginine and lysine,

such as alanine, glycine, leucine, serine, and tyrosine are applied for proteomics studies (Beynon & Pratt, 2005). Apart from that, an isotope with a different mass also needs to be considered. The mass difference between the light and heavy forms of peptide pairs needs to be at least 4 Da to ensure proper separation of heavy and light peptides. Otherwise, isotope peak clusters from light and heavy peptides will overlap (especially for large peptides). We choose L-lysine (8): $^{13}C^{15}N$ and L-arginine (10): $^{13}C^{15}N$, which provide a 6 Da difference for each tryptic peptide. With a minimum requirement of 5 cell doubling times, 98% labeling efficiency of peptides can be reached. Cells reached 98% labeling efficiency after 12 days of labeling (Figure 15| A). TiO_2 was used for phosphopeptides enrichment to increase the peptide detection by mass spec. Phosphopeptides display an impaired ionization efficiency compared to non-phosphopeptides (Beausoleil et al., 2006). Thus, enrichment of phosphopeptides was introduced. Only a few non-phosphopeptides were detected (Figure 17|), indicating that the phospho-enrichment by metal oxide affinity chromatography (MOAC) using TiO_2 was sufficient. Thus, this SILAC-mass spec system for quantifying CTD phosphopeptides was able to be used.

3.1.3 Loss of labeling efficiency of the carboxy-terminus of Rpb1

The labeling efficiency of Rpb1 displayed a significant drop at the CTD region, and the intensity of this drop differed between peptides (Figure 15| B). In contrast, the rest part of Rpb1 gave an ideally normal labeling efficiency. Losing part of labeled peptides may not reflect a comprehensive CTD phosphorylation profile. Therefore, it is necessary to find out the cause of the peptide loss. What causes the labeling loss of the CTD region? Several factors may contribute to this issue.

Insufficient trypsin digestion might cause the poor detection of the peptide. Ideally, trypsin digestion results in peptides with homogeneous characteristics and an average length of 14 amino acids (Burkhart et al., 2012). Enzymatic cleavage may be drastically suppressed in the presence of phospho- amino acids, resulting in missed cleavage, which may directly impact phosphorylation stoichiometry determination (Dickhut et al., 2014). This limitation is often omitted in phosphopeptides quantification rather than phosphorylation sites. From the results (Figure 20|), the number of peptides detected of each tryptic peptide was different, and seldom peptides containing multiply

phosphorylated sites (>3) were seen. It raises the concern of the missed cleavage of the hyperphosphorylated peptides. In addition, acetylation or di- and trimethylation of lysine and arginine residues can make peptides resistant to trypsin digestion (Lee et al., 2011). Moreover, the missed cleavages were not equally distributed between lysine and arginine residues. Missed lysine cleavage sites were present at disproportionally high levels compared to missed arginine sites (Saveliev et al., 2013). Thus, we evaluated miss cleavage frequency. No significant miss cleavage was found (Figure 28|).

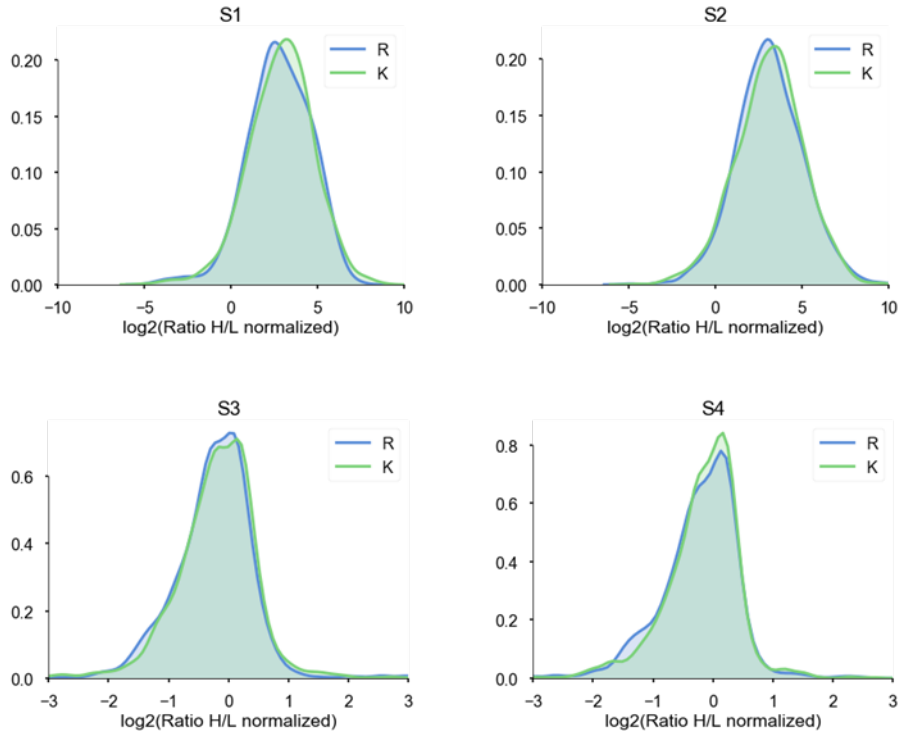


Figure 28| **Distribution of H/L ratios of peptides with missed cleavage sites > 0.** S1 = heavy K and R only, 12 days, trypsin + LysC; S2 = heavy K and R only, 15 days, trypsin + LysC; S3 = heavy/light K and R mixed 1:1, 12 days, Trypsin + LysC; S4 = heavy/light K and R mixed 1:1, 12 days, Trypsin only.

A larger dose of trypsin was utilized, with a ratio of 1:20 (trypsin: protein) instead of the typical 1:100, to guarantee thorough digestion. Another option would be a combined proteases usage, such as trypsin + Lys C, which would contribute to complete digestion. However, it was not improved a lot. Trypsin only and trypsin+LysC gave a similar labeling efficiency (data not shown).

Some cell lines can metabolically convert arginine to proline, which is a recognized concern when utilizing arginine labeling (Ong & Mann, 2007; Van Hoof et al., 2007). In MS, this will cause the signal of a proline-containing peptide to split into multiple signals, depending on the number of prolines in the sequence. Unfortunately, the Rpb1 CTD is an extreme proline-rich region, and each heptad contains two prolines. In a

heavy labeling medium, such proline conversion introduces a heavy proline from heavy arginine, which leads to additional mass changes of the peptide. Decreasing the concentration of arginine in the medium may prevent this conversion from happening (Van Hoof et al., 2007).

Nevertheless, this will also cause the growth condition to change when compared to the standard growth medium. Reduction of the arginine concentration may lead to cell growth defects. Thus, we use the same concentration of arginine and lysine as a standard medium for Raji cell culturing. The options left is either to exclude all proline-containing peptides from quantitation; or for each heavy proline-containing peptide, add up all its split signals for quantitation, or even use specially designed labeling strategies to cancel out the error. In theory, identifying the miss counted heavy proline containing peptides and bring them back into account could increase the representant of the CTD peptides pool and improve the statistics value. In fact, this option did not improve the labeling efficiency and brought a huge duty of algorisms to the machine. Thus, we decided to eliminate these heavy proline-containing peptides for further analysis.

3.2 Analysis of CTD phosphopeptides after CDK9 inhibition

3.2.1 CTD phospho- patterns after CDK9 inhibition

CDK9 is a critical component of P-TEFb, regulates transcriptional pausing release. At the pausing site, CDK9 phosphorylates Ser2 of CTD, which is pre phosphorylated at Ser5 and Ser7. Inhibition of CDK9 caused Pol II stuck at the pausing site. Prolonged inhibition led to premature termination (Introduction 1.5.2). How is the profile of CTD phosphorylation when CDK9 is inhibited? Are Ser5P and Ser7P increased, indicating a paused pol II?

With a timeframe of one-hour inhibition, the polymerases can transcribe the whole length of most genes. Thus, most of pol II been analyzed were not past pausing site before the inhibition implied to make sure all polymerases were affected by CDK9 inhibition. The results turned out the highest phosphorylation at Ser5, followed by Ser2 and Tyr1 (Figure 19| A-B), indicating a significant paused Pol II was observed. On the other hand, a significant amount of Ser7P was not found primarily, possibly due to the

heavy K and R substitution on Ser7 of the mass spec CTD construct, which could reduce CDK9 kinetic efficiency, as a pre-phosphorylated Ser7 exhibited 4-fold higher CDK9 activity on CTD substrate compared to a substrate with K substitution on Ser7 (Czudnochowski et al., 2012).

With the keenness for specific phosphorylation of CTD during the transcription process, certain CDK must identify the particular residue in a short-given time once they are recruited to pol II CTD. Such a puzzle is especially faced by CDKs when dealing with Ser2/Ser5 recognition, both of which are Ser-Pro motifs (Komarnitsky et al., 2000b).

CDK9 phosphorylation of CTD was mainly considered happening on Ser2. Evidence exhibited Ser5P is not a target of CDK9. Cells treated with CDK9 inhibitor caused phosphorylation changes of Ser2, but not Ser5, on elongating Pol II (Albert et al., 2016). Meanwhile, Ser5 exhibited as a target of CDK9 in several cases. P-TEFb phosphorylated both Ser2 and Ser5 in vitro (Ebmeier et al., 2017) and gave a preference on Ser5 (Czudnochowski et al., 2012). A structural motif study in vitro of CTD kinases characterized the CDK7 and ERK2 as Ser5 specific kinases (Ramani et al., 2020). Results showed both Ser2P and Ser5P were affected after CDK9 inhibition (Figure 19| C), implying that they are both CDK9 substrates. Even Lys was heavily introduced in the seventh spot, which will reduce CDK9's recognition on Ser5 (Ramani et al., 2020).

CDK9 is crucial for the cell in transcriptional processes and cell differentiation, cell fate responses, development, and other biological processes. SILAC proteomics study of CDK9 in our lab indicates CDK9 inhibition can increase pausing and delay polymerase recycling. Thus, long-term inhibition can decrease transcription activity. In our S5A-3 mutant, only 250 genes showed different expression, and most of them were upregulated. Here are some possibilities to explain. First, it is a matter of how much serine 2 residues are phosphorylated other than which heptad is phosphorylated. Second, the phosphorylation of NELF and DSIF is not affected in the mutant cell. It is sufficient to allow the polymerase to pass through the pausing site into the elongation phase. Pause release is governed by CDK9, multi- targets of CDK9 including Ser2 CTD, NELF, DSIF, SPT6 (Tellier et al., 2020) contribute to pausing control.

Although CDK9 is a primary regulator of Pol II pause release, several studies indicate

that other factors are also involved in pause checkpoint regulation. A phosphatase-PP2A, recruited by Integrator subunit INTS6, antagonized CDK9 at the phosphorylation level on DSIF and CTD Ser2. Disassociation of PP2A and INTS6 turned out to be resistant to CDK9 inhibition (Vervoort et al., 2021). Without disturbing NELF or DSIF, paused Pol II still transit into the elongation phase (Kwak & Lis, 2013; Vos et al., 2018; Yamaguchi et al., 2013), indicating CTD in pause control is not crucial. Interestingly, an entire CTD truncated Pol II displayed functional pause and elongation (Gerber & Roeder, 2020).

3.2.2 The target of CDK9 on CTD

How do CTD kinases phosphorylate the 52 repeats of CTD? Do kinases equally phosphorylate each heptad or display specificity to a certain heptad at a specific location of the entire CTD? Where do kinases start to phosphorylate a CTD? Unfortunately, there is no clear answer up to date.

Previous work from our lab exhibited diverge phosphorylation changes between heptads after four hours of flavopiridol treatment. Our SILAC results here displayed a different level of response to CDK9 inhibition from all CTD peptides (Figure 20|), indicating that each heptad has different sensitivity of response to CDKs inhibition. A dramatic phosphopeptide decrease of the second peptide from variant M1 1h after CDK9 was observed sample (Figure 21|), indicating a major target site of CDK9. However, the CTD phosphorylation patterns vary in different inhibition times and different CTD constructs (Figure 20|). We could not see a solid phosphopeptides distribution pattern under a different period of CDK9 inhibition. More extended inhibition could have affected more polymerases at the early stage of transcription. Moreover, extended inhibition could impact other CDKs, thus indirectly affecting CTD phosphorylation (Devaiah & Singer, 2012; Nekhai et al., 2002). Different manipulation between variants M1 and M3 CTD construct may also change the motif of kinase recognition.

As discussed in the previous section, the recognition motif of CDKs may explain why there is no solid phospho-pattern after CDK9 inhibition. The recognition of Ser2/ Ser5 can be affected by residues character at the nearby position. A significant amount of

non-consensus heptads residues provides characteristic deviation to affect CTD phosphorylation. Additionally, these non-consensuses also bring new modifications, potentially recruit other factors to regulate the transcription process.

3.3 Functional analysis of CTD heptad 2 and 3

3.3.1 Interferon genes and CTD S5A mutant

In the present study, S5A-3 mutants were made to evaluate the function of Ser5 on heptad2 and 3 and the connection with CDK9. We observed differentially expressed genes and pathways for the S5A-3 mutant. Only 250 genes were shown to be differently expressed, and no significant growth defect in the mutant cell. It may be due to the functional redundancy of CTD heptad repeats. We found that most DE genes were upregulated in the S5A-3 mutant rather than downregulated. Gene Set Enrichment Analysis (GSEA) analysis displayed most upregulated genes were interferon response related. These genes contained a conserved regulatory element at the promoter region(Chow & Gale Jr, 2015).

IFNs, which are best known for their antiviral properties, are grouped into type-I and type-II IFNs. Type-I IFNs are mainly produced directly after viral infection, whereas type-II IFNs are secondarily produced by activated T and NK cells (Chow & Gale, 2015). The interferon genes in our list contain both type-I and type-II IFNs. Environment changes and mechanical stimulation can induce interferon (Ivashkiv & Donlin, 2014; Stifter et al., 2019). However, it may not be the case in our experiment since the wild-type and mutant cells were cultured in the same condition.

Why are the interferon-stimulated genes active? One possibility is that mutant CTD may produce aberrant dsRNA that stimulate an interferon response (Figure 29|). The dsRNA is an essential threat to the cells. Many viral infections are associated with cytoplasmic dsRNA. Various sensor proteins can detect dsRNA properties, such as the 5′ phosphate group of viral RNAs or hybrids of a given length, but not particular nucleotide sequences. Interferon signaling, which stimulates cellular stress pathways and innate immunity, as well as caspase activation, which triggers apoptosis, are all activated by the sensors. The principal sources of endogenous dsRNA include mitochondrial transcripts, repetitive nuclear sequences, including short and long

interspersed elements (SINEs, LINEs), and endogenous retroviruses (ERVs), as well as natural sense–antisense transcript pairs (Sadeq et al., 2021). Endogenous retroviruses produce a considerable amount of dsRNA, which is also essential for the synthesis of lncRNAs (H. Lee et al., 2019; Palazzo & Koonin, 2020). A large proportion of lncRNA transcripts are rapidly degraded because they lack protective modifications such as splicing, polyadenylation, and capping, which would allow them to be exported from the nucleus. In other circumstances, however, contact with local cellular components such as chromatin remodeling complexes may boost the stability and prevent degradation (Fabbri et al., 2019).

In our top upregulated gene list, 5 out of 25 genes expressed lncRNA (Table 1), indicating an increasing amount of dsRNA. We also assumed aberrant mRNA processing might generate RNA secondary structure, provides a source of dsRNAs to trigger interferon signaling.

Another possibility is that mutant CTD, together with chromatin remodelers and transcription activators, have a greater affinity to bind and stabiles at the promoter and enhancer region of interferon-stimulated genes, which contains a conserved region of ISRE or IRF-responsive element (Schoggins, 2019), and actively transcribes interferon genes (Figure 29|). IRF3, CDK9, and unphosphorylated Pol II were recruited to the ISRE element of ISG promoters to active transcription of ISGs (Wang et al., 2017). RSV infection increases the active form of P-TEFb by encouraging the switch of CDK9 binding with the inhibitory 7SK small nuclear RNA to BRD4 (Au-Yeung & Horvath, 2018). BRD4 coordinates the recruitment of elongation factors to regulates ISGs transcription (Patel et al., 2013). IFI27 is upregulated by IFNs, and trap transcription factors NR4A1 and Crm1 decreased Pol II transcription activation (Papac et al., 2012). BRD4 was found to bind to the TSS region of IFIT1 through acetylated histones, recruits P-TEFb and NELF/DSIF, and altered the binding of Pol II, thus regulating ISGs transcription (Patel et al., 2013). Some other ISGs, such as STAT1, OASL were also found to process the same regulation path (Bisgrove et al., 2007).

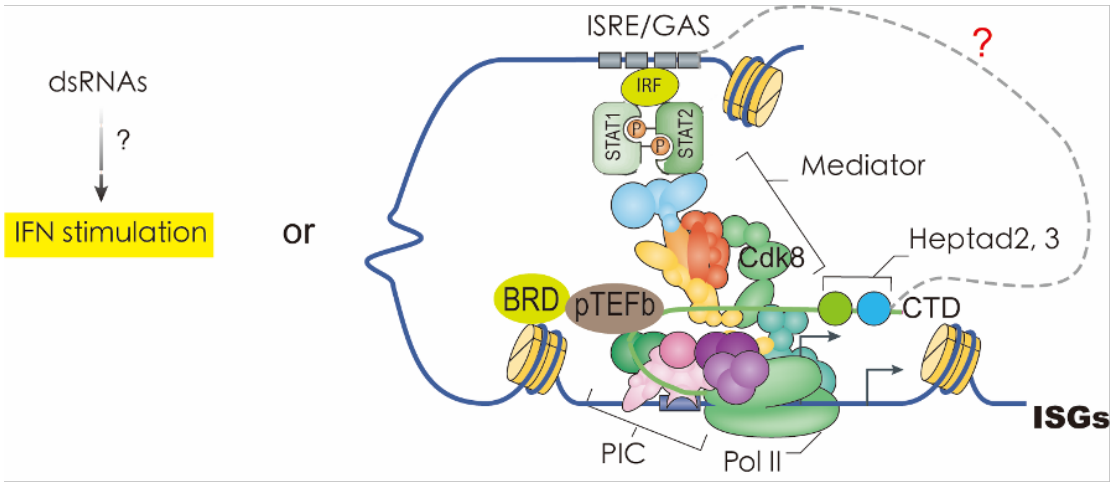


Figure 29| **Induction of interferon-stimulated genes of CTD S5A mutant.** CTD S5A mutant cells may produce abnormal endogenous dsRNAs, which triggers interferon (IFN) stimulation. The heterodimer of STAT1 and STAT2, associates with the IFN-regulatory factor, binds to the ISRE or GAS elements of ISG promoters and enhancers, recruiting chromatin remodelers and transcriptional activators. Thus, S5A mutants may have the advantage of regulating the transcription complex at the promoter and enhancer region of ISGs.

Similar to CDK9 inhibition, the viral proteins interact with CDK9 (Palermo et al., 2008; Zaborowska et al., 2014) and alter CTD phosphorylation (Rice et al., 1994) after infection, thus repress host gene transcription (Zaborowska, Isa, et al., 2016). For example, HSV protein ICP22 triggers significant loss of Ser2P of CTD (Fraser & Rice, 2007). Alternatively, virus protein hijacks CDK9 in a host cell to promote the transcription of virus genomes. For example, HIV Tat competes with HEXIM1 for interacting with cyclin T1 to active CDK9 to further transcribe viral genes (Barboric et al., 2007). In addition, interferon signaling was involved after the virus infection. RNA-seq analysis reveals upregulation of 324 interferons stimulated genes after CDK9 inhibition (Zhang et al., 2018). An additional hint connection might trigger interferon signal when an active form of P-TEFb does not sufficiently recognize and phosphorylate Ser2 CTD to stimulate host genes transcription.

Five lncRNAs were significantly upregulated in our list. A depletion of the ubiquitous Pol-II-associated transcription elongation factor SPT6 leads to the formation of extended, polyadenylated lncRNAs that are both chromatins restricted, and form increased levels of RNA: DNA hybrid (R-loops) that are associated with DNA damage (Nojima, Tellier, et al., 2018). LncRNAs transcription exhibited different Pol II phospho-CTD isoforms compared to protein-coding genes transcription (Schlackow et al., 2017). The S5A mutant may cause different CTD phosphorylation, preferentially actives the lncRNAs transcription.

Endogenous retrovirus (ERV) genes also cause the stimulation of interferon (Goel et al., 2017; Wienerroither et al., 2015). Several ERV genes appeared in our list of upregulated genes. The mitochondrial antiviral signaling protein (MAVS)-interferon regulatory factors (IRFs) pathway is activated by ERV induction via MDA5 and, to a lesser extent, a retinoic acid-inducible gene I (RIG1). LINEs associate with various dsRNA binding proteins, mostly PKR and MDA5, and their expression has been linked to the activation of an interferon 1 response (Kim et al., 2018). An interferon (IFN) response was only observed after high-level expression of the hairpin construct in a transfected human cell line (HEK293) (Nejepinska et al., 2012). BRD4 and P-TEFb were rarely detected on immune response genes prior to interferon stimulation. Likewise, NELF and DSIF are also seldom recruited to immune genes in the absence of stimulation (Patel et al., 2013).

3.3.2 Co-evolution of interferon response and CTD heptad

A fully functional IFN system, including IFN-stimulated genes, is present in cartilaginous fish and bony fish, as well as in tetrapod, suggesting that IFN earliest emerged in jawed vertebrates. Along with the interleukin (IL)-10 cytokine family, they are believed to have originated from a class II helical cytokine ancestor (Secombes & Zou, 2017). The whole-genome multiplication that occurred between invertebrates and vertebrates (Nakatani et al., 2007) may have provided capacity for new genes that initially included IL-10 and IFN ancestor, leading to the formation of the IL-10 family genes, as well as the IFN type I and IFN type III systems.

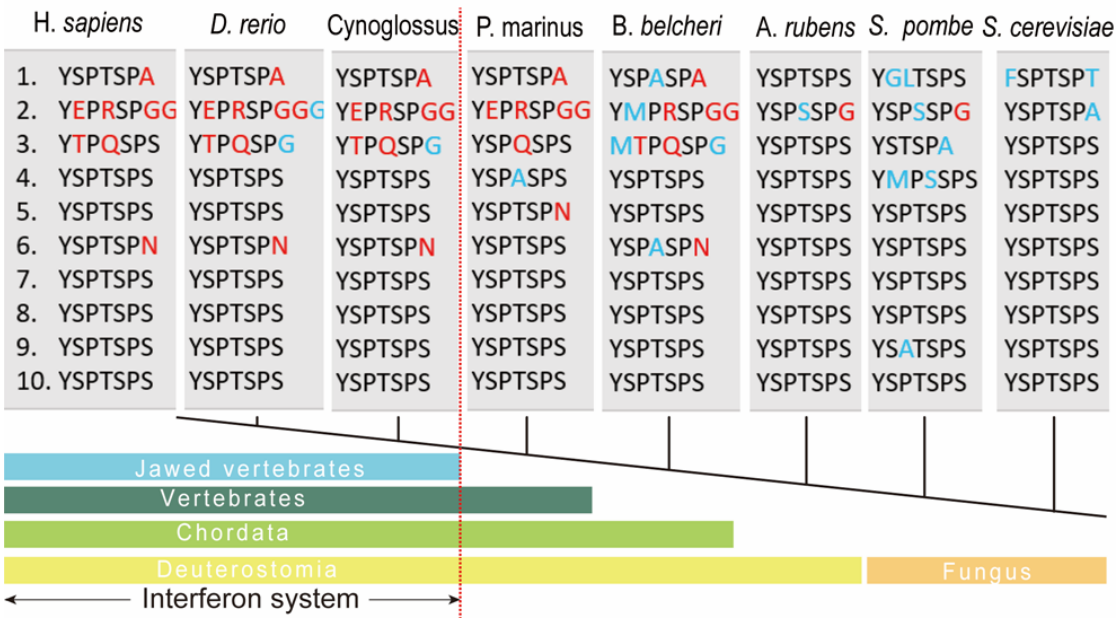


Figure 30| **The conservation of the CTD non-consensus.** Alignment of heptad 1-10 of eukaryotic RNA pol II CTD. Conserved non-consensus residues in mammals are highlighted in red, and those non-consensus residues in other organisms are in blue. Taxonomy was marked with a color bar. Interferon existing classes were distinguished by the red dash line.

Redundancy (length of heptad repeats) and diversity (non-consensus residues) of CTD keep increasing during the evolution (Chapman et al., 2008; Stiller & Hall, 2002). Unique and conserved non-consensus at heptad 1-3 were seen in vertebrates. Depletion of heptad 1-3 in Raji cells heavily affects CTD stability (Chapman et al., 2005), indicating the non-consensus may carry additional modifications, thus, stabilizing CTD. Interestingly, such non-consensus of heptad 1-3 in mammals were emerged and gradually established during vertebrate's evolution. Such a co- evolutional process might give a hint for a connection between interferon genes and non-censuses of heptad 2 and 3. These non-consensus residues may provide specific recruitment of factors to ISRE and IRF, the conserved promoter element of interferon genes.

4 Materials and Methods

4.1 Material

4.1.1 Chemicals

1 kb DNA ladder	Invitrogen, Karlsruhe
1,4-Dithiothreitol (DTT)	Carl Roth GmbH&CoKG, Karlsruhe
1-NA-PP1	Calbiochem, Merck Millipore, Darmstadt
4-thiouracil (4sU)	Sigma-Aldrich Chemie GmbH, München
Agarose	Invitrogen, Karlsruhe
Bromophenol blue (BPB)	Sigma-Aldrich Chemie GmbH, München
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich Chemie GmbH, München
DPBS	(Gibco) Life Technologies, Eggenstein
Ethanol (EtOH), absolute	Merck, Darmstadt
SYBR™ Safe DNA Gel Stain	Invitrogen, Karlsruhe
Ethylendiaminetetraacetic acid (EDTA)	Carl Roth GmbH&CoKG, Karlsruhe
Fetal bovine serum (FBS)	Sigma-Aldrich Chemie GmbH, München
Glycerol 86%	Carl Roth GmbH&CoKG, Karlsruhe
Glycine	Carl Roth GmbH&CoKG, Karlsruhe
Isopropanol, absolute	Carl Roth, Karlsruhe
L-glutamine 200mM	(Gibco) Life Technologies, Eggenstein
Methanol (MeOH), absolute	Merck, Darmstadt
Neomycin (G148)	Promega Corp., Wisconsin, USA
Penicillin/Streptomycin 10k U/ml	(Gibco) Life Technologies, Eggenstein
Phosphatase inhibitor cocktail	Roche Diagnostics, Penzberg
Polyacrylamide 30% (PAA)	Carl Roth GmbH+Co.KG, Karlsruhe
Powdered milk, blotting grade	Carl Roth GmbH+Co.KG, Karlsruhe
Prestained Protein Ladder Plus	Fermentas, St. Leon-Rot
Protease inhibitor cocktail	Roche Diagnostics, Penzberg
Protein A-Sepharose beads	GE Healthcare, Munich
Protein G-Sepharose beads	GE Healthcare, Munich

QIAzol	Qiagen GmbH, Hilden
RPMI medium 1640	(Gibco) Life Technologies, Eggenstein
Sodium dodecyl sulfate (SDS)	Carl Roth GmbH+Co.KG, Karlsruhe
Tetracycline	Promega Corp., Wisconsin, USA
Tetramethylethylenediamine (TEMED)	Carl Roth GmbH+Co.KG, Karlsruhe
Tris(hydroxymethyl)aminomethane (TRIS)	Merck, Darmstadt
TRIzol	Thermo Fisher Scientific, Waltham
Trypan blue	Invitrogen, Karlsruhe
Tween-20	Sigma-Aldrich Chemie GmbH, München
Poly(I:C) LMW	InvivoGen, Toulouse,France
SILAC RPMI (w/o L-Gln, L- Arg, L-Lys)	Silantes GmbH, Munich, Germany
SILAC FBS (dialyzed)	Silantes GmbH, Munich, Germany
L-Lysine(8) HCl, isotopic labeling: $^{13}C^{15}N$	Silantes GmbH, Munich, Germany
L-Arginine(10) HCl, isotopic labeling: $^{13}C^{15}N$	Silantes GmbH, Munich, Germany
L-Lysine(0), isotopic labeling: unlabelled	Silantes GmbH, Munich, Germany
L-Arginine(0), isotopic labeling: unlabelled	Silantes GmbH, Munich, Germany
L-Glutamine, isotopic labeling: unlabelled	Silantes GmbH, Munich, Germany
TiO_2 beads	GL Sciences, Tokyo, Japan

4.1.2 Consumables

Agar plates	Greiner Bio-One GmbH, Frickenhausen
SUSPENSION CULTURE FLASK, 50, 250 , 650 ML	Greiner Bio-One GmbH, Frickenhausen
Cover slides	Menzel, Braunschweig
Cryovials 1.5 ml	1.5 ml Nunc GmbH, Wiesbaden
ECL detection reagent	GE healthcare, München

Gel blotting Paper GB003	Schleicher & Schuell, Deutschland
Gene Pulser cuvettes	Bio-Rad Laboratories GmbH, München
Hybond N+ nylon membrane	GE Healthcare, München
Laboratory glassware S	Duran Productions GmbH & Co. KG, Mainz
Nitrile gloves duoSHILD S	SHIELD Scientific, Bennekom, NL
Parafilm	Carl GmbH+Co.KG, Karlsruhe
Pasteur pipettes	Hirschmann Laborgeräte, Eberstadt
Pipette tips ART 10, 20, 200, 1000µl	MolecularBio-Products, San Diego
Falcon® 15 mL Polystyrene Centrifuge Tube	Corning GmbH, Kaiserslautern
Falcon® 50 mL Polystyrene Centrifuge Tube	Corning GmbH, Kaiserslautern
RNA nanochip kit	Technologies, Waldbronn
Scalpel Braun	Scalpel Braun, Tuttlingen
Sterile cup filters	Merck Millipore, Darmstadt
Streptavidin beads	Miltenyi Biotec, Bergisch Gladbach
T4 DNA Ligase	New England Biolabs, Frankfurt a.M.
Whatmann gel blotting paper	Schleicher & Schuell, Germany

4.1.3 Consumable kits

Agilent RNA 6000 pico kit	Agilent Technologies, USA
DNA Mini/Maxi kits	Qiagen GmbH, Hilden
NucleoSpin gel extraction kit	MACHEREY-NAGEL GmbH, Düren
RNA Library Preparation kit	Epicenter, USA
RNA library Preparation Kit	Illumina, USA
QIAEX II Gel extraction	Qiagen GmbH, Hilden
Ribo-Zero-rRNA removal kit	Epicenter, USA
SG Cell Line 4D-NucleofectorTM X Kit L	Lonza, Köln
NucleoSpin Plasmid EasyPure, Mini kit	MACHEREY-NAGEL GmbH, Düren

4.1.4 Technical instruments

-20°C freezer Eco Plus	Siemens, München
-80°C freezer Hera Freeze	Thermo Fisher Scientific, Waltham
AQUAline AL 12 waterbath	LAUDA, Lauda-Königshofen
Bacteria shaker (Series 25)	New Brunswick ScientificCo., NJ, USA
Bio-Rad PowerPac 300	Bio-Rad Laboratories GmbH, München
Branson Sonifier 250	Heinemann Ultraschall- und Labortechnik
Countess automated cell counter	(Invitrogen) Thermo Fisher Scientific, Waltham
Electrophoresis equipment	Bio-Rad Laboratories GmbH, München
Electroporator (eukaryotic cells)	Bio-Rad Laboratories GmbH, München
Eppendorf Centrifuge 5417R	Eppendorf Deutschland, Wesseling-Berzdorf
Eppendorf Thermomixer C	Eppendorf Deutschland, Wesseling-Berzdorf
Focused ultrasonicator S220	Covaris Ltd, Brighton, UK
Fridge KU 1710 Vario	Liebherr, Biberach
Fuchs-Rosenthal chamber	GLW Gesellschaft für Laborbedarf GmbH
Hypercassette	Amersham Biosciences, Freiburg
Illumina HiAeq 1500	Illumina Inc., San Diego, CA, USA
iMac 27-inch, Late 2013	Apple Inc., Cupertino, CA, USA
Incubator HERA cell 150	Heraeus Sepatech GmbH, Osterode
Inkubator Heraeus 6000	Heraeus Sepatech GmbH, Osterode
Laminar Flow Hood	BDK Luft-und Reinraumtechnik GmbH
Magnet stirrer M23	GLW, Würzburg
Microwave NNV 689W	Panasonic, Hamburg
Mighty Small II SE260 Electrophoresis Unit	Amersham Biosciences, Freiburg
Mighty Small Transphor Unit TE 22	GE Healthcare, München
MiniSpin Plus centrifuge	Eppendorf Deutschland, Wesseling-Berzdorf

Multi-calimatic pH-meter	Knick GmbH+Co.KG, Berlin
Nanodrop 1000	Thermo Fisher Scientific, Waltham
Odyssey imaging system	LI-COR Biosciences GmbH, Bad Homburg
PipetMan P	Gilson, Bad Camberg
Power supply Peqlab EV202	VWR Life Science, Erlangen Primo
Rollermixer SRT 6	Stuart Equipment, Staffordshire, UK
Rotanta 460 R centrifuge	Hettich, Tuttlingen
Rotina 380 centrifuge	Hettich, Tuttlingen
Scanmaker i800 plus	Microtek, Taiwan
Sunrise Photometer	Tecan Group Ltd., Männedorf, CH
UV lamp Peqlab VL-4. LC	VWR Life Science, Erlangen
Vert light microscope	Carl Zeiss Jena GmbH, Göttingen
Vi-CELL XR cell counter	Beckman Coulter, München
Vortexer Reax 2000	Heidolph Instruments GmbH, Schwabach
xCELLigence RTCA DP	Roche Diagnostics, Mannheim
4D-NucleofectorTM X Unit	Lonza, Köln
4D-NucleofectorTM Core Unit	Lonza, Köln

4.1.5 Buffer and solutions

0.7% Agarose-Gel	2.1 g agarose powder 300 ml 1X TAE buffer Boil in microwave cool down to 65°C Add syber safe (1:1000 use)
2x Tris/SDS pH 8.8	90.72 gms Tris base 10 ml of 20% SDS Adjust volume to 1000 ml with H2O Adjust pH to 8.8

2xTris/SDS pH 6.8	30.24 gms Tris base 10 ml of 20% SDS Adjust volume to 1000 ml with H2O Adjust pH to 6.8
Blocking buffer for western blott	5% milk in 1X TBS-T
IP Lysis buffer	50 mM Tris/HCl, pH 8.0 1% NP40 150 mM NaCl Phosphatase inhibitor (1 pill/10ml) Protease inhibitor (1 pill/10ml)
IP Wash buffer	50 mM Tris/HCl, pH 8.0 150 mM NaCl Phosphatase inhibitor (1 pill/10ml) Protease inhibitor (1 pill/10ml)
Laemmli-Buffer (2x)	2% SDS 100 mM DTT 10 mM EDTA 20% Glycerol 60 mM Tris/HCl; pH 6.8 0.01% Bromophenolbue
Laemmli-Buffer (6x)	9% SDS 375 mM Tris/HCl; pH 6.8 9% β-mercaptoethanol 50% Glycerol 0.06% Bromophenolblue
PBS	137 mM NaCl 2,7 mM KCl

	4,3 mM Na2HPO4*6H2O 1,4 mM KH
SDS-PAGE separating gel (6.5%)	4.3 mL PAA 30% 10 mL 2xTris/SDS pH 8.8 5.5 mL H2O 1167 µL APS 17 µL TEMED
SDS-PAGE Stacking gel (4%)	1.5 ml 30% PAA 7.5 ml 2xTris/SDS pH 6.8 5.9 ml H2O 90 µl APS 20 µl TEMED
SDS-PAGE-running buffer (10x)	60.4 g Tris/Base 288 g Glycin 5 ml SDS 20% add 2 L H2O
SDS-PAGE-transfer buffer (10x)	60.4 g Tris/Base 288 g Glycin 5 mL SDS 20% 200 mL methanol add 2 L H2O
TiO2-Elution Buffer	40 µl Ammonium solution (28%) in 960 µl H2O, pH 11,3
TiO2-Loading Buffer	80% Acetonitrile 5% TFA 1 M Glycolic acid

TiO2-Washing Buffer 2	10% Acetonitrile 0,2% TFA
Tris acetate EDTA (TAE) 1x	40 mM Tris Acetate 1 mM EDTA adjust pH to pH 8.0
Tris buffered saline (TBS) 1x	10 mM NaCl 1 mM Tris/HCl pH 7.5

4.1.6 Antibodies

	source	clone	species	usage
HA	Roche Diagnostics	3F10	rat	WB 1:100
HA	Roche Diagnostics	12CA5	rat	WB 1:100
Rpb1	Dr. E Kremmer	Pol3.3	mouse	WB 1:5
CTD Ser2-P	Dr. E Kremmer	3E10	rat	WB 1:10
CTD Ser5-P	Dr. E Kremmer	3E8	rat	WB 1:10
CTD Ser7-P	Dr. E Kremmer	4E12	rat	WB 1:10
CTD Try1-P	Dr. E Kremmer	3E8	rat	WB 1:10
CTD Thr4-P	Dr. E Kremmer	3E10	rat	WB 1:10
α-Tubulin	Sigma (T9026)	T6557	mouse	WB 1:20,000
GAPDH	Dr. E Kremmer	5C4	rat	WB 1:5,000

Secondary antibodies	Source	Usage
Alexa Fluor 680 anti rat IgG	Invitrogen	WB 1:20,000
HRP anti mouse IgG	Promega	WB 1:5,000
HRP anti rabbit IgG	Promega	WB 1:5,000
HRP anti rat IgG+IgM	Jackson Laboratories	WB 1:5,000
IRDye 800 anti mouse IgG	Rockland Inc.	WB 1:10,000

4.1.7 Cloning materials

Enzymes	Source
Age1	New England Biolabds GmbH, Frankfurt am Main
Not1	New England Biolabds GmbH, Frankfurt am Main
AvrII	New England Biolabds GmbH, Frankfurt am Main
BspE1	New England Biolabds GmbH, Frankfurt am Main
Not1-HF	New England Biolabds GmbH, Frankfurt am Main
Age1-HF	New England Biolabds GmbH, Frankfurt am Main
Not1	Thermo Fisher Scientific, Waltham
Age1	Thermo Fisher Scientific, Waltham

Plasmids and bacteria	Source/ reference
RX2-287	subcloning vector): Vector containing last exon (CTD) of the α-amanitin resistant Pol II Rpb1 gene
RX4-267	expression vector): A tetracycline-regulated expression vector containing the ⟨-amanitin resistant and HA-tagged mouse Rpb1 gene
DH10B	E.coli strain purchased from Invitrogen GmbH, Karlsruhe was used for the cloning of all plasmid DNA

Primers for sequencing	Source/reference
WT fwd	5'CTCCTGCTGACGCACCTGTTCT3'
CTD fwd	5'CCTTTGTCTTTTCCTATAGGTGGTGC3'
CTD rev	5'GTCAGACAACCTCGGTGGCCTGTGTG3'

Cells	Expression vector	Cell line	Resistance	Source/reference
WT	\	Raji	\	(JV, 1964).
recWT	RX4-267	Raji	Neomycin, α-amanitin	(Meininghaus et al., 2000)

S5A H_2	RX4-267	Raji	Neomycin, α-amanitin	From this study
S5A H_3	RX4-267	Raji	Neomycin, α-amanitin	From this study
S5A $H_{2,3}$	RX4-267	Raji	Neomycin, α-amanitin	From this study
CDK9as		Raji	\	(Gressel et al., 2017)
CDK9as-recWT	RX4-267	Raji	Neomycin, α-amanitin	From this study
CDK9as-S5A H_2	RX4-267	Raji	Neomycin, α-amanitin	From this study
CDK9as-S5A H_3	RX4-267	Raji	Neomycin, α-amanitin	From this study
CDK9as-S5A $H_{2,3}$	RX4-267	Raji	Neomycin, α-amanitin	From this study

4.2 Methods

4.2.1 CTD plasmids and cloning strategy

GeneArt synthesized all of the Pol II CTD variant sequences, which were then incorporated into the pOK-RQ vector (KanR). Inside the synthetic CTD chain, there are no restriction sites for AvrII, NotI, AgeI, BspEI, NgoMIV, NheI, SpeI, or ClaI. AvrII and NotI restriction sites flanked the CTD frame on both sides. A two-step cloning technique was used to clone the CTD sequences into the expression vector RX4-267 (Fig.32 A-B).

Using restriction enzymes AvrII and NotI, the synthesized CTD sequence was first excised from the pOK-RQ (KanR) vector and incorporated into the subcloning vector RX2-287. The CTD construct in the RX2-287 vector was then fragmentary with BspEI and NotI restriction enzymes before being embedded into the AgeI/NotI site of the RX4-267 expression vector. As a result, the recombined Rpb1 in the expression vector has a HA tag at the N-terminus, an α-amanitin tolerant mutation at Rpb1's N793D, and a reconstitute CTD. Using restriction enzymes AvrII and NotI, the synthesized CTD sequence was first excised from the pOK-RQ (KanR) vector and incorporated into the subcloning vector RX2-287. The CTD construct in the RX2-287 vector was then fragmentary with BspEI and NotI restriction enzymes before being embedded into the AgeI/NotI site of the RX4-267 expression vector. As a result, the recombined Rpb1 in the expression vector has a HA tag at the N-terminus, an α-amanitin tolerant mutation at Rpb1's N793D, and the reconstitute CTD.

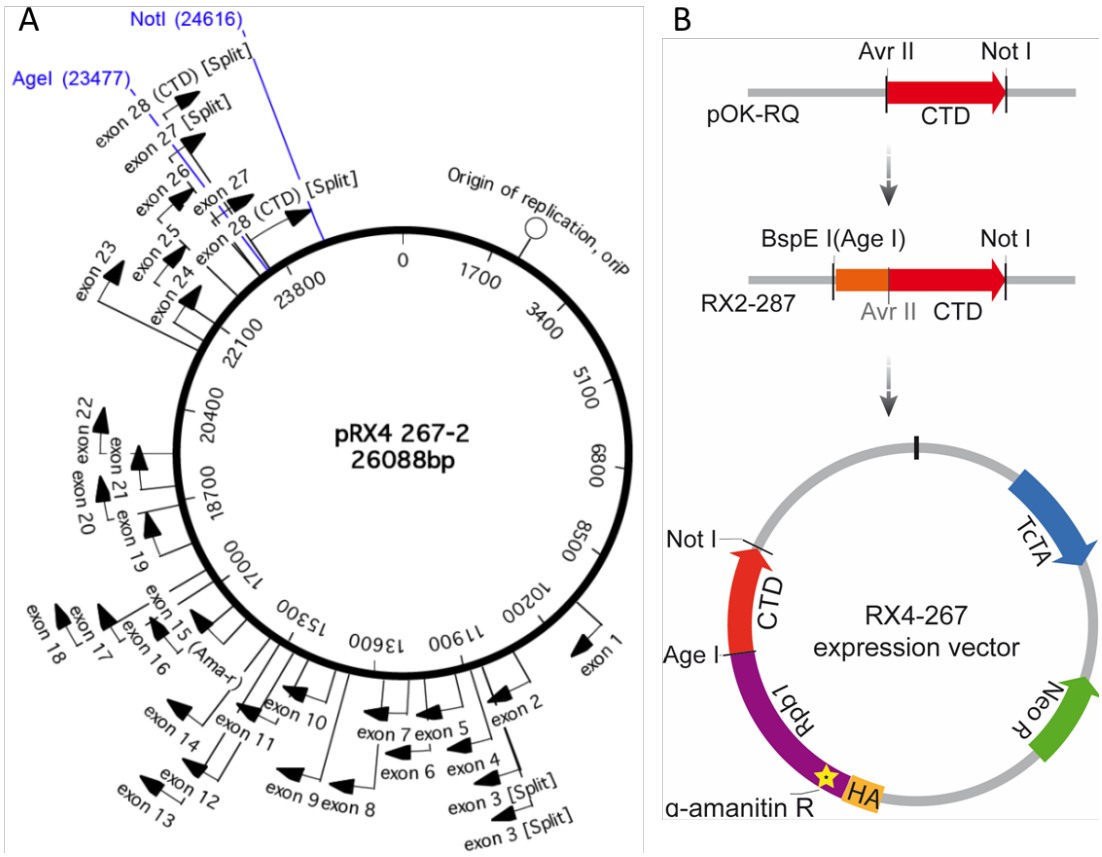


Figure 31| **A schematic representation of the two-step cloning procedure. (A)** pRX4-267 end vector scheme The expression vector pRX4-267 contains all 28 exons of the mouse Rpb1 gene. Using restriction enzymes AgeI and NotI, the sequence encoding CTD can be easily replaced by the given artificial CTD sequences. (B) AvrII and NotI restriction enzymes carry out GeneArt's CTD sequence from the pOK-RQ vector. The flanked sequence is inserted into the RX2- 287 subcloning vector, and the CTD sequence from RX2-287 is flanked by BspEI and NotI restriction enzymes and inserted into the AgeI/NotI sites of the RX4-267 expression vector.

4.2.2 DNA digestion using restriction endonucleases

DNA fragments and vectors of cloning were digested with appropriate restriction enzymes. Restriction enzymes from New England Biolabs (NEB) and Thermo Fisher Scientifics are being used. The digesting procedure was followed exactly as directed by the manufacturer. The digested products were electroporated on an agarose gel, which enabled the desired fragment to be recognized and extracted.

4.2.3 Ligation of DNA constructs

The excited restriction enzymic digest products from an agarose gel were purified using the NucleoSpin gel extraction kit, as directed by the manufacturer. The purified CTD insert and vector was mixed in a 3:1 molar ratio, and the mixture was incubated for 10 minutes at room temperature with T4 DNA ligase and buffer. The ligase was

then heated inactive for 10 minutes at 65°C.

4.2.4 Transformation of competent cells

50 µL of competent DH10B E. Coli cells were mixed either with 5 µL of the ligation product or with 3 µg of plasmid DNA, followed by incubation for 15 min on ice. Heat-shocked cells were exposed to 42°C for 60 seconds before being placed on ice for 10 minutes. 400 µL of non-antibiotic LB medium was applied, and the cells were precultured for 40 minutes at 37°C on a thermo block. The cells were centrifuged and resuspended in 50 µL of LB medium before plating the suspension cells on antibiotic-selective agar plates with a stiral stick. The plates were incubated for 16 hours at 37°C.

4.2.5 Miniprep of bacterial plasmid DNA

On an agar plate, a single colony was picked from a transformed cell. In an incubation shaker at 37°C for 16 hours, a selected colony was cultured in 2 mL of LB medium with suitable antibiotics. Typically, the culturing cells attain a concentration of 3–6 OD. 1.5 mL of culturing cells were centrifuged at > 12,000 g for 30 seconds. Remove the supernatant and discard it. The NucleoSpin® Plasmid EasyPure kit was then used to purify the plasmids. The cell pellet was fully resuspended by vortexing 150 µL of Buffer A1. Then 250 µL of Buffer A2 was added and gently combined by inverting the tube several times. 350 µL Buffer A3 was applied and thoroughly mixed by inverting the tube until the mixture turned colorless. The tube was centrifuged at maximum speed for 3 minutes (> 12,000 x g). The supernatant was loaded onto the column and centrifuged at 1,000 x g for 30 seconds. After discarding the flow-through, 450 µL of Buffer AQ was applied and centrifuged at full speed for 1 minute (> 12,000 x g). The column was put in a 1.5 mL Ep tube and washed for 1 minute at room temperature (18–25 °C) with 50 mL Buffer AE before being centrifuged for 1 minute at full speed (> 12,000 x g). Enzyme digestion and electroporation can be used to validate purified plasmids.

4.2.6 Maxiprep of bacterial plasmid DNA

A stored aliquot from a verified plasmid obtained from Miniprep was added into 400ml LB medium containing appropriate antibiotics and incubated at 37°C for 16-18h in an

incubation shaker. QIAGEN's maxi prep kit was used to purify the samples. An overnight bacterial culture was harvest by centrifuging at 6000 x g for 15 min at 4°C. The bacterial pellet was resuspended in 10 ml Buffer P1. After supplying 10ml Buffer P2, the suspensions were thoroughly mixed by vigorously inverting 4–6 times and incubated at room temperature (15–25°C) for 5 minutes. After applying 10 mL prechilled Buffer P3, the suspensions were thoroughly blended by vigorously inverting 4–6 times. After that, the fluid was incubated for 20 minutes on ice. The column was washed with 2 x 30 ml Buffer QC by gravity flow. 15 mL Buffer QF was used to elute DNA into a new 50 mL tube. After that, DNA was precipitated by applying 10.5 mL isopropanol to the supernatant and centrifuging for 30 minutes at 4°C at 15,000 x g. The DNA pellet was washed in 5 mL 70 % ethanol at room temperature and centrifuged for 10 minutes at 15,000 x g. After discarding the supernatant and air-drying the pellet for 5–10 minutes, DNA was removed using TE buffer (pH 8.0).

4.2.7 Cell culturing

The Raji cell stocks were thawed in a 37°C water bath after being extracted from liquid nitrogen. The cells were suspended in 50 mL of growth medium and centrifuged for 3 minutes at 1200 rpm. The cell pellet was suspended in 10ml growth media containing RPMI supplemented with 10% fetal calf serum, 2 mM L-glutamine, and 1% penicillin/streptomycin and transferred into a T-75 culture flask at 37°C and 5% CO_2. The cells are kept at 37°C with 5% CO_2 in an incubator. For the next few days, the medium was replaced every 2 to 3 days, and cultures were split when they reached $1x10^6$ viable cells/ml.

4.2.8 Stable transfection of Raji cell

Raji cells were stably transfected by electrophoresis using Nucleofection™ Process. Cells were divided 1:2 one day before transfection. A portion of the culture medium was heated to 37°C in advance. $2x\ 10^6$ cells were used per transfection cell. The cells were centrifuged for 10 minutes at 90x g at room temperature. The cell pellet was resuspended with 100 µL of 4D-Nucleofector™ solution after the supernatant was depleted. To each aliquot, 3 g of plasmid was added. The master mixes were loaded

into Nucleocuvette™ Vessels and then placed in the retainer of the 4D-Nucleofector X Unit. After the Nucleofection process was completed, the cells were resuspended in 1ml of the pre-warmed medium.

The cells were incubated at 37°C with 5% CO_2 in a 6-wells plate. When the medium turns yellow after 1-2 days, start selection with Neo (1 mg/mL G418) and 1 µg/ml tetracycline. Positive selection by splitting (1:2, medium yellow) for 2-4 weeks until viability of 90-95 % was achieved, after which the cells were split several times until 70ml. Wash-out tet was used to induce the expression of recombinant Rpb1. Cells were incubated with a complete medium for 24h. An aliquot can be taken for testing recombinant Rpb1 expression. Endogenous polymerase was removed by selecting recombinant Rpb1 with α-amanitin (2 µg/mL). Cells were kept in α-amanitin selection medium for 2-4 weeks until 90% viability is achieved.

4.2.9 Cell proliferation assays

A Fuchs-Rosenthal counting chamber was used to count the cells. Take 100-200 µl of cell suspension under sterile conditions. Mix in an equal volume of 0.04 % Trypan Blue with a gentle pipetting operation. Using an inverted phase-contrast microscope at 20x magnification, fill both sides of the chamber with cell suspension (approximately 5-10 µl). Using a hand tally counter, count the number of viable (bright) and non-viable (blue) cells. Calculate the number of viable and non-viable cells, as well as the cell viability of each sample.

4.2.10 Western analysis of the whole-cell extract

Whole-cell lysis was prepared from cultured cells. $1x10^6$ cells were collected in an Ep tube, centrifuged, and washed once with 1x PBS at 1200rpm for 5 minutes, and the pellet was resuspended and lysed in 100ml of 2x Laemmli buffer. The lysate was boiled and sonicated. Samples were centrifuged at full speed for 10 min. Alternatively, the supernatant was loaded onto the gel or held at -20°C.

An SDS-PAGE gel with appropriate polyacrylamide concentration was cast and assembled in a gel-running tank filled with running buffer. Protein samples were heated

at 95°C for 5min. Samples and a protein marker were loaded onto the gel using a gel-loading tip after a brief spin. As the sample moved through the stacking gel, the gels were continuously run at a lower current (30mA/per gel). When samples were drafted in the separating gel, the gels were continuously run at a higher current (45mA/per gel). When the blue front reaches the desired position, electrophoresis is terminated. Transfer buffer was used to soak a 0.2 μm nitrocellulose membrane and Whatman papers. Gel and nitrocellulose membrane was placed in between two Whatman paper. Any bubbles or dead spaces at the interface between any layers should be removed. Apply 450mA to the wet transfer system continuously for 90 minutes. Ponceau S reversible staining can be used to display transferred protein bands. A successfully transferred membrane was blocked with 5% milk in TBST for 1 hour at room temperature or overnight at 4°C with persistent rocking. The primary antibodies were diluted according to the dilution ratio. Primary antibodies were incubated on membranes for 1 hour at room temperature or overnight at 4°C. 3x TBST was used to wash the membranes for 5 minutes each. HRP conjugated secondary antibody or secondary fluorescence antibody is used to incubate the membranes.

4.2.11 Coomassie blue and Ponceau S staining

The SDS-PAGE gel was fixed in fix solution (50% methanol,10% glacial acetic acid) for 2 hours with gentle agitation. The gel was then stained for 2–3 hours with gentle agitation in a staining solution (0.1% Coomassie Brilliant Blue, 50 % methanol, and 10% glacial acetic acid). The gel was deposited in 5% glacial acetic acid after destaining in solution (40% methanol and 10% glacial acetic acid) several times until there was no unspecific background.

The blotted membrane was stained for 5 minutes in Ponceau S staining solution (0.1 % Ponceau S in 1% acetic acid). The membrane was then soaked in a 5 % acetic acid (v/v) aqueous solution for 5 minutes to evaluate protein transfer efficiency from the SDS gel to the nitrocellulose membrane. Ponceau S staining can be reversibly washed out in TBST and processed continuously for antibody incubation.

4.2.12 SILAC labeling

Raji cells were cultured in SILAC specific RPMI 1640 media without Lysine, arginine,

and glutamine, then supplemented with 10% dialyzed fetal bovine serum (Silantes), heavy isotopes of $^{13}C^{15}N$-labeled L-arginine (200 mg/L), and L-lysine (40 mg/L) (Silantes), referred as heavy labeling. Meanwhile, control cells were cultured in SILAC RPMI 1640 media with unlabeled L-arginine (40 mg/L) and L-lysine (200 mg/L) to respectively replace heavy Lys or Arg. All cell lines were cultivated and kept at 37 ° C in an incubator containing 5% CO_2. The culture media were replaced every 3 days. To obtain >% labeling, cells were maintained in a 'labeled' medium for at least five cell passages.

4.2.13 Inhibition of CDK9 analog-sensitive cell

SILAC-labeled were collected and washed in phosphate-buffered saline once (PBS). Cell pellets were resuspended and seeded at a concentration of 1 x 10^6/ml with 1.5x 10^8 cells per flask. Cells were treated with 10 μM ATP analog by adding 150 μl of 1-NA-PP1 stock solution or 150 μl of DMSO for an appropriate span of time (1h or 15 minutes). Following CDK9 inhibition, heavy labeled 1-NA-PP1-treated cells were mixed in a 1:1 ratio with unlabeled DMSO-treated cells. There were three biological replicates of the paired mixed sample prepared. In parallel, three corresponding label-swapped replicates were prepared, in which unlabeled cells were treated with 1-NA-PP1, and DMSO control cells were heavily labeled with $^{13}C^{15}N$ arginine and lysine isotope.

4.2.14 Cell lysis and immunoprecipitation

Primary antibodies were combined with protein A/G sepharose beads and rolled for 6 hours at 4°C. On the following day, the requisite number of Raji cells were collected and washed twice in cold PBS. A sufficient amount of pellets is resuspended in lysis buffer (100ul lysis buffer for 1x 10^6 cells). Cells were lysed for 40 minutes at 4°C on a rocker and then sonicated to shear chromatin. For 15 minutes, cell lysates were centrifuged at 10,000g. The supernatants were then combined with antibody-conjugated protein A/G beads and incubated on a foam roller overnight at 4°C. The next day, samples were collected by centrifugation at 1200rpm for 2 minutes, followed by four washes with wash buffer and resuspension in 2x Laemmli buffer. The samples were boiling at 95°C for 5 minutes. IP quality can be assessed using Western analysis. After western blot validation, the IP samples were loaded onto an 8 % acrylamide gel,

allowed to run into separating gel for a brief period of time, and the electrophoresis was stopped until the 300kDa ladder was clearly visible, and the gel was stained with Coomassie to visualize gel bands.

4.2.15 Proteins in-gel digestion

The IIo and IIa bands of Rpb1 were excised from each sample and sliced into 1mm 3 slices, which were then washed twice with 150 µL of 20 mM NH_4HCO_3 (ABC) at room temperature to eliminate unpolymerized acrylamide. After that, the gels were destained by incubating 2-3 times with 50% 20 mM ABC /50% ultrapure acetonitrile (ACN) for 20 minutes at 37°C with 650 rpm shaking until the Coomassie blue color was almost out. From there, the gel pieces were washed three times with 20 mM NH_4HCO_3 for ten minutes each time and dehydrated by incubating with ACN three times for ten minutes each time. After being dehydrated and dried, gel slices were rehydrated for 1 hour at room temperature using a rehydration solution (10 mM DTT/20 mM ABC). Next, gels were incubated in IAA solution (55 mM Iodoacetamide, 20 mM ABC) for 30 minutes in darkness to reduce disulfide bonds and alkylate-free cysteines, allowing optimal trypsin access to cleavage sites within the peptide.

After being dehydrated and dried with ACN, gel slices were digested overnight at 37 °C with 100 µl MS-grade trypsin (Promega, USA). Peptides in solution were transferred to a new tube, and peptides that remained in the gel were extracted twice at room temperature with 100 µl extraction solution (50% ACN/ 0.25% trifluoroacetic acid). A total of 50 µl of 100% ACN was used twice for maximum extraction. The extracted supernatant was dried at room temperature in a vacuum centrifuge. Phospho-peptide enrichment was then performed on the samples.

4.2.16 Phospho-peptide enrichment

Samples were incubated with 100µl loading buffer (1M Glycolic acid/ 80% ACN/ 5% TFA) for 30 minutes on ice. Titansphere Phos-TiO_2 Beads were added into samples. After 10 min incubation, phosphopeptides were bonded onto TiO_2 beads. After 2x wash, phosphopeptides were eluted with 50 µl Elution buffer (28% Ammonia solution) and concentrated by speed-vac and acidified in 30 µl of 3% TFA for desalting.

4.2.17 Peptide clean-up and LC-MS/MS analysis

Styrenedivinylbenzene-reversed phase-sulfonated StageTips (SDB-RPS; 3M Empore) columns were assembled in two layers with SDB-RPS chromatography paper and 200 µl pipette tip. The acidified samples were directly applied to SDB-RPS StageTips. The column was rinsed Two times with 100 µl 1% (vol/vol) TFA in isopropanol and 0.2% (vol/vol) TFA. Peptides are eluted with SDB-RPS elution buffer (60 µL) (80% ACN, 1.25% NH_4OH). The samples were concentrated to 2 µl using a SpeedVac centrifuge at 45°C for 30 minutes. Reconstitute peptides in 10 µl MS loading buffer (2% acetonitrile, 0.3% TFA). The peptides can now be stored at -20°C, and the samples can be analyzed by mass spectrometry.

Proteins were identified using a Thermo Ultimate 3000 RSLCnano system and isolated in a 15-cm analytical column (75 µm ID with ReproSil-Pur C18-AQ 2.4 m from Dr. Maisch) using a 50-minute mobile phase A gradient from 3 to 16% acetonitrile in 0.1% formic acid, followed by a 10-minute mobile phase B gradient from 16 to 26% acetonitrile in 0.1% formic acid.

The HPLC effluent was electrosprayed directly into a QexactiveHF(Thermo) in data-dependent mode, which consisted of one full MS scan in profile mode followed by seven MS/MS scans in centroid mode for each scan cycle. At m/z 400, full survey scan MS spectra (from m/z 375–1600) were acquired with R=60,000 resolution (AGC target of $3x10^6$). The ten most intense peptide ions with charge states ranging from 2 to 5 were isolated sequentially to a target value of $1x10^5$ and fragmented at a normalized collision energy of 27%. Spray voltage of 1.5 kV; no sheath and auxiliary gas flow; a heated capillary temperature of 250°C; ion selection threshold of 33.000 counts were typical mass spectrometric conditions.

Maxquant software 1.5.2.8 was used as the database search engine to handle the mass spec data. For database searching, the following settings were configured: Initial peptide mass tolerance was set to 10ppm; fragment mass tolerance was 0.5 Da; Peptide FDR, 0.1; Protein FDR, 0.01 Min. peptide Length, 5; Variable modifications, Oxidation (M); Fixed modifications, Carbamidomethyl (C); Peptides for protein quantitation, razor and unique; MaxQuant's normalized heavy to light ratios are log2 transformed for statistical analysis.

4.2.18 RNA-seq library preparation

RNA seq analysis was performed on CTD S5A mutant cell and CDK9 analog-sensitive cell with 1-hour inhibition. The cells were washed twice in PBS, resuspended in TRIzol (Qiagen), and snap-freezing in liquid nitrogen before being stored at -80 °C until library preparation. The library preparation and RNA sequencing were conducted in cooperation with Stefan Krebs of the Gene Center's Helmut Blum Lab (LMU Munich).

After thawing, Total RNA was extracted from cells using the Directzol kit (Zymo Research, Irvine CA, USA) according to the manufacturer's instructions. Purified RNA was measured using a Nanodrop 1000 instrument and shown to have a purity with OD260/280 = 1.8-2.2 and OD260/230 >1.8. The Bioanalyzer 2100 was used to verify the purity of the RNA using the RNA Nano Assay kit (Agilent Technologies, USA). Only RNA samples with a RIN greater than 8 were sequenced.

Following that, the RNAs proceeded through strand-specific cDNA library preparation. mRNA molecules containing poly-A were purified using poly-T oligo-attached magnetic beads. The mRNAs were then fragmented into small pieces. Using reverse transcriptase and random primers, the cleaved RNA fragments were reverse transcribed into the first-strand cDNA. The RNA template was removed and double-stranded cDNA generated by synthesizing a replacement strand. The End Repair was performed to converts the overhangs into blunt ends using T4 DNA polymerase and Klenow DNA polymerase. To prepare the DNA fragments for ligation to the adapters, which have a single 'T' base overhang at their 3' end, the 3' end of the blunt phosphorylated DNA fragments is Adenylated. The adapters were ligated to the ends of the DNA fragments to prepare hybridization to a single-read flow cell. The ligation products were then evaluated for appropriate insert size using a DNA 1000 chip on the Agilent Bioanalyser 2100. The samples were sequenced in single-read mode on an Illumina HiSeq1500 with a read length of 100 bp. Each sample had a sequencing depth of 50 million reads.

4.2.19 Bioinformatics analysis of RNA-seq data

Bioinformatics assessment was conducted by Michael Kluge from the research group of\sCaroline Friedel at the Institute for Informatics (LMU Munich). Sequencing reads were mapped to the human genome (hg38) and human rRNA sequences using ContextMap. Using featureCounts on the GENCODE gene annotations, the number of

reading fragments per gene was determined. First, the principal component analysis (PCA) was performed using limma. Differential gene expression analysis was performed using limma, edgeR, and DESeq2. The DAVID Web er was used to perform functional enrichment analysis for Gene Ontology terms. Using Benjamini and Hochberg's process, significantly enriched terms were calculated with a multiple testing correction of P-value 0.05.

V Bibliography

Adelman, K., & Henriques, T. (2018). Transcriptional speed bumps revealed in high resolution. *Nature*. https://doi.org/10.1038/d41586-018-05971-8

Albert, T. K., Rigault, C., Eickhoff, J., Baumgart, K., Antrecht, C., Klebl, B., Mittler, G., & Meisterernst, M. (2014). Characterization of molecular and cellular functions of the cyclin-dependent kinase CDK9 using a novel specific inhibitor. *British Journal of Pharmacology*, *171*(1), 55–68. https://doi.org/10.1111/bph.12408

Albert, Thomas K., Antrecht, C., Kremmer, E., & Meisterernst, M. (2016). The establishment of a hyperactive structure allows the tumour suppressor protein p53 to function through P-TEFb during limited CDK9 kinase inhibition. *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0146648

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: A central integrator of transcription. *Nature Reviews Molecular Cell Biology*, *16*(3), 155–166. https://doi.org/10.1038/nrm3951

Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, *21*(2), 71–87. https://doi.org/10.1038/s41576-019-0173-8

Au-Yeung, N., & Horvath, C. M. (2018). Transcriptional and chromatin regulation in interferon and innate antiviral gene expression. In *Cytokine and Growth Factor Reviews* (Vol. 44, pp. 11–17). Elsevier Ltd. https://doi.org/10.1016/j.cytogfr.2018.10.003

Babokhov, M., Mosaheb, M. M., Baker, R. W., & Fuchs, S. M. (2018). Repeat-specific functions for the C-Terminal domain of RNA Polymerase II in budding yeast. *G3: Genes, Genomes, Genetics*, *8*(5), 1593–1601. https://doi.org/10.1534/g3.118.200086

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S. M., Adhikari, B., Schwarz, J. D., Narain, A., Vogt, M., Wang, S. Y., Düster, R., Jung, L. A., Vanselow, J. T., Wiegering, A., Geyer, M., Maric, H. M., Gallant, P., Walz, S., … Wolf, E. (2019). MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Molecular Cell*, *74*(4), 674-687.e11. https://doi.org/10.1016/j.molcel.2019.02.031

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.7

Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B. N., Chung, H. J., Chan-Salis, K. Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., Piotrowski, J., Thomas, C. J., Singer, D. S., Pugh, B. F., Pommier, Y., Przytycka, T. M., Kouzine, F., Lewis, B. A., Zhao, K., & Levens, D. (2016). RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell*, *165*(2), 357–371. https://doi.org/10.1016/j.cell.2016.02.036

Barboric, M., Yik, J. H. N., Czudnochowski, N., Yang, Z., Chen, R., Contreras, X., Geyer, M., Peterlin, B. M., & Zhou, Q. (2007). Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Research*, *35*(6), 2003–2012. https://doi.org/10.1093/nar/gkm063

Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., Adelman, K., Lis, J. T., & Greenleaf, A. L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes and Development*. https://doi.org/10.1101/gad.1968210

Bartolomei, M. S., & Corden, J. L. (1987). Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. *Molecular and Cellular Biology*, *7*(2), 586–594. https://doi.org/10.1128/mcb.7.2.586

Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J., & Gygi, S. P. (2006). A probability-

based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology*, *24*(10), 1285–1292. https://doi.org/10.1038/nbt1240

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, *15*(3), 163–175. https://doi.org/10.1038/nrg3662

Bernecky, C., Plitzko, J. M., & Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.3465

Beynon, R. J., & Pratt, J. M. (2005). Metabolic labeling of proteins for proteomics. In *Molecular and Cellular Proteomics* (Vol. 4, Issue 7, pp. 857–872). Elsevier. https://doi.org/10.1074/mcp.R400010-MCP200

Bisgrove, D. A., Mahmoudi, T., Henklein, P., & Verdin, E. (2007). Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(34), 13690–13695. https://doi.org/10.1073/pnas.0705053104

Bishop, A. C., Ubersax, J. A., Pøtsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., Shimizu, E., Tsien, J. Z., Schultz, P. G., Rose, M. D., Wood, J. L., Morgan, D. O., & Shokat, K. M. (2000). A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, *407*(6802), 395–401. https://doi.org/10.1038/35030148

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., McSwiggen, D. T., Kokic, G., Dailey, G. M., Cramer, P., Darzacq, X., & Zweckstetter, M. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural and Molecular Biology*, *25*(9), 833–840. https://doi.org/10.1038/s41594-018-0112-y

Bösken, C. A., Farnung, L., Hintermair, C., Schachter, M. M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R. P., Eick, D., & Geyer, M. (2014). The structure and substrate specificity of human Cdk12/Cyclin K. *Nature Communications*, *5*.

https://doi.org/10.1038/ncomms4505

Bowman, E. A., & Kelly, W. G. (2014). RNA Polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation A tail of two kinases. *Nucleus*, *5*(3), 224–236.

Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., & Zahedi, R. P. (2012). Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *Journal of Proteomics*, *75*(4), 1454–1462. https://doi.org/10.1016/j.jprot.2011.11.016

Burriss, K. H., & Mosley, A. L. (2019). Methods review: Mass spectrometry analysis of RNAPII complexes. In *Methods* (Vols. 159–160, pp. 105–114). Academic Press Inc. https://doi.org/10.1016/j.ymeth.2019.03.013

Chao, S. H., & Price, D. H. (2001). Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo. *Journal of Biological Chemistry*, *276*(34), 31793–31799. https://doi.org/10.1074/jbc.M102306200

Chapman, R. D., Conrad, M., & Eick, D. (2005). Role of the Mammalian RNA Polymerase II C-Terminal Domain (CTD) Nonconsensus Repeats in CTD Stability and Cell Proliferation. *Molecular and Cellular Biology*, *25*(17), 7665–7674. https://doi.org/10.1128/mcb.25.17.7665-7674.2005

Chapman, R. D., Heidemann, M., Albert, T. K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., & Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science*, *318*(5857), 1780–1782. https://doi.org/10.1126/science.1145977

Chapman, R. D., Heidemann, M., Hintermair, C., & Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. In *Trends in Genetics* (Vol. 24, Issue 6, pp. 289–296). https://doi.org/10.1016/j.tig.2008.03.010

Chen, F. X., Smith, E. R., & Shilatifard, A. (2018). Born to run: Control of transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *19*(7),

464–478. https://doi.org/10.1038/s41580-018-0010-5

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, *361*(6400). https://doi.org/10.1126/science.aar2555

Chow, K. T., & Gale Jr, M. (2015). SnapShot: Interferon Signaling. *CELL*, *163*, 1808-1808.e1. https://doi.org/10.1016/j.cell.2015.12.008

Chow, K. T., & Gale, M. (2015). SnapShot: Interferon Signaling. In *Cell* (Vol. 163, Issue 7, pp. 1808-1808.e1). Cell Press. https://doi.org/10.1016/j.cell.2015.12.008

Chung, H., Calis, J. J. A., Wu, X., Sun, T., Yu, Y., Sarbanes, S. L., Dao Thi, V. L., Shilvock, A. R., Hoffmann, H. H., Rosenberg, B. R., & Rice, C. M. (2018). Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. *Cell*, *172*(4), 811-824.e14. https://doi.org/10.1016/j.cell.2017.12.038

Conaway, R. C., & Conaway, J. W. (2019). The hunt for RNA polymerase II elongation factors: a historical perspective. *Nature Structural and Molecular Biology*, *26*(9), 771–776. https://doi.org/10.1038/s41594-019-0283-1

Corden, J. L. (2013). The RNA Polymerase II C-terminal Domain: Tethering Transcription to Transcript and Template. *Chemical Reviews*, *113*(11), 8423–8455. https://doi.org/10.1021/cr400158h

Corden, J. L. (2019). An Arginine Nexus in the RNA Polymerase II CTD. *Molecular Cell*, *73*(1), 3–4. https://doi.org/10.1016/j.molcel.2018.12.013

Core, L., & Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes and Development*, *33*(15–16), 960–982. https://doi.org/10.1101/gad.325142.119

Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., & Bentley, D. L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and

Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Molecular Cell*, *76*(6), 896-908.e4. https://doi.org/10.1016/j.molcel.2019.09.031

Cramer, P. (2019a). Eukaryotic Transcription Turns 50. *Cell*, *179*(4), 808–812. https://doi.org/10.1016/j.cell.2019.09.018

Cramer, P. (2019b). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

Cramer, P. (2019c). Five decades of eukaryotic transcription. *Nature Structural & Molecular Biology*, *26*(9), 757. https://doi.org/10.1038/s41594-019-0303-1

Custódio, N., & Carmo-Fonseca, M. (2016). Co-transcriptional splicing and the CTD code. In *Critical Reviews in Biochemistry and Molecular Biology* (Vol. 51, Issue 5, pp. 395–411). Taylor and Francis Ltd. https://doi.org/10.1080/10409238.2016.1230086

Czudnochowski, N., Bösken, C. A., & Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms1846

Decker, T. M., Forné, I., Straub, T., Elsaman, H., Ma, G., Shah, N., Imhof, A., & Eick, D. (2019). Analog-sensitive cell line identifies cellular substrates of CDK9. *Oncotarget*, *10*(65), 6934–6943. https://doi.org/10.18632/oncotarget.27334

Devaiah, B. N., Lewis, B. A., Cherman, N., Hewitt, M. C., Albrecht, B. K., Robey, P. G., Ozato, K., Sims, R. J., & Singer, D. S. (2012). BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1120422109

Devaiah, B. N., & Singer, D. S. (2012). Cross-talk among RNA polymerase II kinases modulates C-terminal domain phosphorylation. *Journal of Biological Chemistry*, *287*(46), 38755–38766. https://doi.org/10.1074/jbc.M112.412015

Dickhut, C., Feldmann, I., Lambert, J., & Zahedi, R. P. (2014). Impact of digestion conditions on phosphoproteomics. *Journal of Proteome Research*, *13*(6), 2761–2770. https://doi.org/10.1021/pr401181y

Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S., & Lima, C. D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes and Development*. https://doi.org/10.1101/gad.242768.114

Eagle, H. (1955). Nutrition needs of mammalian cells in tissue culture. *Science*, *122*(3168), 501–504. https://doi.org/10.1126/science.122.3168.501

Eaton, J. D., Davidson, L., Bauer, D. L. V., Natsume, T., Kanemaki, M. T., & West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes and Development*, *32*(2), 127–139. https://doi.org/10.1101/gad.308528.117

Eaton, J. D., & West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics*, *36*(9), 664–675. https://doi.org/10.1016/j.tig.2020.05.008

Ebmeier, C. C., Erickson, B., Allen, B. L., Allen, M. A., Kim, H., Fong, N., Jacobsen, J. R., Liang, K., Shilatifard, A., Dowell, R. D., Old, W. M., Bentley, D. L., & Taatjes, D. J. (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. *Cell Reports*, *20*(5), 1173–1186. https://doi.org/10.1016/j.celrep.2017.07.021

Egloff, S., Dienstbier, M., & Murphy, S. (2012). Updating the RNA polymerase CTD code: Adding gene-specific layers. *Trends in Genetics*, *28*(7), 333–341. https://doi.org/10.1016/j.tig.2012.03.007

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., & Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*.

https://doi.org/10.1126/science.1145989

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., & Murphy, S. (2012). Ser7 phosphorylation of the CTD recruits the RPAP2 ser5 phosphatase to snRNA genes. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2011.11.006

Eick, D., & Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical Reviews*, *113*(11), 8456–8490. https://doi.org/10.1021/cr400071f

Fabbri, M., Girnita, L., Varani, G., & Calin, G. A. (2019). Decrypting noncoding RNA interactions, structures, and functional networks. In *Genome Research* (Vol. 29, Issue 9, pp. 1377–1388). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gr.247239.118

Fabrega, C., Shen, V., Shuman, S., & Lima, C. D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Molecular Cell*, *11*(6), 1549–1561. https://doi.org/10.1016/S1097-2765(03)00187-4

Fraser, K. A., & Rice, S. A. (2007). Herpes Simplex Virus Immediate-Early Protein ICP22 Triggers Loss of Serine 2-Phosphorylated RNA Polymerase II. *Journal of Virology*, *81*(10), 5091–5101. https://doi.org/10.1128/jvi.00184-07

Galbraith, M. D., Donner, A. J., & Espinosa, J. M. (2010). CDK8: A positive regulator of transcription. In *Transcription*. https://doi.org/10.4161/trns.1.1.12373

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J. M., Wagner, E. J., Lazar, M. A., Shilatifard, A., & Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2014.08.004

Garriga, J., & Graña, X. (2014). CDK9 inhibition strategy defines distinct sets of target genes. *BMC Research Notes*, *7*(1), 1–10. https://doi.org/10.1186/1756-0500-7-301

Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., & Mann, M. (2011). Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature Protocols*, *6*(2), 147–157. https://doi.org/10.1038/nprot.2010.192

Genes Robles, C. M., & Coin, F. (2019). Conducting the CTD orchestra. *Nature Chemical Biology*, *15*(2), 97–98. https://doi.org/10.1038/s41589-018-0201-6

Gerber, A., & Roeder, R. G. (2020). The CTD Is Not Essential for the Post-Initiation Control of RNA Polymerase II Activity. *Journal of Molecular Biology*, *432*(19), 5489–5498. https://doi.org/10.1016/j.jmb.2020.07.010

Ginsburg, D. S., Govind, C. K., & Hinnebusch, A. G. (2009). NuA4 Lysine Acetyltransferase Esa1 Is Targeted to Coding Regions and Stimulates Transcription Elongation with Gcn5. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.01033-09

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R. P., & Bentley, D. L. (2009). 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*. https://doi.org/10.1128/mcb.00637-09

Goel, S., Decristo, M. J., Watt, A. C., Brinjones, H., Sceneay, J., Li, B. B., Khan, N., Ubellacker, J. M., Xie, S., Metzger-Filho, O., Hoog, J., Ellis, M. J., Ma, C. X., Ramm, S., Krop, I. E., Winer, E. P., Roberts, T. M., Kim, H. J., McAllister, S. S., & Zhao, J. J. (2017). CDK4/6 inhibition triggers anti-tumour immunity. *Nature*, *548*(7668), 471–475. https://doi.org/10.1038/nature23465

Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K., & Hinnebusch, A. G. (2007). Gcn5 Promotes Acetylation, Eviction, and Methylation of Nucleosomes in Transcribed Coding Regions. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2006.11.020

Greenleaf, A. L. (2019). Human CDK12 and CDK13, multi-tasking CTD kinases for

the new millenium. In *Transcription*. https://doi.org/10.1080/21541264.2018.1535211

Gregersen, L. H., Mitter, R., Ugalde, A. P., Nojima, T., Proudfoot, N. J., Agami, R., Stewart, A., & Svejstrup, J. Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. *Cell*, *177*(7), 1797-1813.e18. https://doi.org/10.1016/j.cell.2019.04.038

Greifenberg, A. K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C. A., Anand, K., Blazek, D., & Geyer, M. (2016). Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Reports*, *14*(2), 320–331. https://doi.org/10.1016/j.celrep.2015.12.025

Gressel, S., Schwalb, B., Decker, T. M., Qin, W., Leonhardt, H., Eick, D., & Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *ELife*, *6*, 1–24. https://doi.org/10.7554/eLife.29736

Grummt, I. (1998). Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I. *Progress in Nucleic Acid Research and Molecular Biology*. https://doi.org/10.1016/S0079-6603(08)60506-1

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., Spille, J. H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., Abraham, B. J., Boija, A., Decker, T. M., Rimel, J. K., Fant, C. B., Lee, T. I., Cisse, I. I., Sharp, P. A., Taatjes, D. J., & Young, R. A. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, *572*(7770), 543–548. https://doi.org/10.1038/s41586-019-1464-0

Harlen, K. M., & Churchman, L. S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, *18*(4), 263–273. https://doi.org/10.1038/nrm.2017.10

Harlen, K. M., Trotta, K. L., Smith, E. E., Mosaheb, M. M., Fuchs, S. M., & Churchman, L. S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct

and Varied Roles for Each Phospho-CTD Residue. *Cell Reports*, *15*(10), 2147–2158. https://doi.org/10.1016/j.celrep.2016.05.010

Herzel, L., Ottoz, D. S. M., Alpert, T., & Neugebauer, K. M. (2017). Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology*, *18*(10), 637–650. https://doi.org/10.1038/nrm.2017.63

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., Chapman, R. D., Andrau, J. C., & Eick, D. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO Journal*, *31*(12), 2784–2797. https://doi.org/10.1038/emboj.2012.123

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, *169*(1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007

Hoedt, E., Zhang, G., & Neubert, T. A. (2019). Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) for Quantitative Proteomics. *Advances in Experimental Medicine and Biology*, *1140*, 531–539. https://doi.org/10.1007/978-3-030-15950-4_31

Hsin, J. P., Sheth, A., & Manley, J. L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. *Science*. https://doi.org/10.1126/science.1206034

Ivashkiv, L. B., & Donlin, L. T. (2014). Regulation of type i interferon responses. In *Nature Reviews Immunology* (Vol. 14, Issue 1, pp. 36–49). NIH Public Access. https://doi.org/10.1038/nri3581

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C. Y., Xiao, R., Burge, C. B., & Fu, X. D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell*.

https://doi.org/10.1016/j.cell.2013.04.028

Kao, S. Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*. https://doi.org/10.1038/330489a0

Kedinger, C., Gniazdowski, M., Mandel, J. L., Gissinger, F., & Chambon, P. (1970). α-Amanitin: A specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/0006-291X(70)91099-5

Kim, Y., Park, J., Kim, S., Kim, M. A., Kang, M. G., Kwak, C., Kang, M., Kim, B., Rhee, H. W., & Kim, V. N. (2018). PKR Senses Nuclear and Mitochondrial Signals by Interacting with Endogenous Double-Stranded RNAs. *Molecular Cell*, *71*(6), 1051-1063.e6. https://doi.org/10.1016/j.molcel.2018.07.029

Komarnitsky, P., Cho, E. J., & Buratowski, S. (2000a). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes and Development*. https://doi.org/10.1101/gad.824700

Komarnitsky, P., Cho, E. J., & Buratowski, S. (2000b). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes and Development*, *14*(19), 2452–2460. https://doi.org/10.1101/gad.824700

Konuma, T., Yu, D., Zhao, C., Ju, Y., Sharma, R., Ren, C., Zhang, Q., Zhou, M. M., & Zeng, L. (2017). Structural Mechanism of the Oxygenase JMJD6 Recognition by the Extraterminal (ET) Domain of BRD4. *Scientific Reports*. https://doi.org/10.1038/s41598-017-16588-8

Kwak, H., & Lis, J. T. (2013). Control of transcriptional elongation. In *Annual Review of Genetics* (Vol. 47, pp. 483–508). Annual Reviews Inc. https://doi.org/10.1146/annurev-genet-110711-155440

Lam, L. T., Pickeral, O. K., Peng, A. C., Rosenwald, A., Hurt, E. M., Giltnane, J. M.,

Averett, L. M., Zhao, H., Davis, R. E., Sathyamoorthy, M., Wahl, L. M., Harris, E. D., Mikovits, J. A., Monks, A. P., Hollingshead, M. G., Sausville, E. A., & Staudt, L. M. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, *2*(10), research0041.1. https://doi.org/10.1186/gb-2001-2-10-research0041

Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J. J., Shokat, K. M., Bentley, D. L., & Fisher, R. P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2399

Lee, H., Zhang, Z., & Krause, H. M. (2019). Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? In *Trends in Genetics* (Vol. 35, Issue 12, pp. 892–902). Elsevier Ltd. https://doi.org/10.1016/j.tig.2019.09.006

Lee, J. Y., Kim, J. Y., Park, G. W., Cheon, M. H., Kwon, K. H., Ahn, Y. H., Moon, M. H., Lee, H. J., Paik, Y. K., & Yoo, J. S. (2011). Targeted mass spectrometric approach for biomarker discovery and validation with nonglycosylated tryptic peptides from N-linked glycoproteins in human plasma. *Molecular and Cellular Proteomics*, *10*(12). https://doi.org/10.1074/mcp.M111.009290

Lens, Z., Cantrelle, F. X., Peruzzini, R., Hanoulle, X., Dewitte, F., Ferreira, E., Baert, J. L., Monté, D., Aumercier, M., Villeret, V., Verger, A., & Landrieu, I. (2017). Solution Structure of the N-Terminal Domain of Mediator Subunit MED26 and Molecular Characterization of Its Interaction with EAF1 and TAF7. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2017.09.001

Li, B., Howe, L. A., Anderson, S., Yates, J. R., & Workman, J. L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M212134200

Li, G. W., & Xie, X. S. (2011). Central dogma at the single-molecule level in living cells. In *Nature*. https://doi.org/10.1038/nature10315

Liang, K., Gao, X., Gilmore, J. M., Florens, L., Washburn, M. P., Smith, E., & Shilatifard, A. (2015). Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.01426-14

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., & Rutter, W. J. (1970). Specific inhibition of nuclear RNA polymerase II by α-amanitin. *Science*. https://doi.org/10.1126/science.170.3956.447

Liu, H., Ramachandran, S., Fong, N., Phang, T., Lee, S., Parsa, P., Liu, X., Harmacek, L., Danhorn, T., Song, T., Oh, S., Zhang, Q., Chen, Z., Zhang, Q., Tu, T. H., Happoldt, C., O'Conner, B., Janknecht, R., Li, C. Y., … Zhang, G. (2020). JMJD5 couples with CDK9 to release the paused RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(33), 19888–19895. https://doi.org/10.1073/PNAS.2005745117

Liu, J., Fan, S., Lee, C. J., Greenleaf, A. L., & Zhou, P. (2013). Specific interaction of the transcription elongation regulator TCERG1 with RNA polymerase II Requires simultaneous phosphorylation at Ser2, Ser5, and Ser7 within the carboxyl-terminal domain repeat. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M113.460238

Liu, Kun., Shen, Dan., Shen, J., Gao, S. M., Li, B., Wong, C., Feng, W., & Song, Y. (2017). The Super Elongation Complex Drives Neural Stem Cell Fate Commitment. *Developmental Cell*. https://doi.org/10.1016/j.devcel.2017.02.022

Lopez, M. S., Kliegman, J. I., & Shokat, K. M. (2014). The logic and design of analog-sensitive kinases and their small molecule inhibitors. In *Methods in Enzymology* (Vol. 548, Issue C, pp. 189–213). Academic Press Inc.

https://doi.org/10.1016/B978-0-12-397918-6.00008-2

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., & Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology*, *17*(10), 1195–1201. https://doi.org/10.1038/nsmb.1893

Luse, D. S. (2013). Promoter clearance by RNA polymerase II. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2012.08.010

Magee, B. B., Paoletti, J., & Magee, P. T. (1975). Separation of lymphocyte chromatin into template active fractions with specificity for eukaryotic RNA polymerase II or prokaryotic RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(12), 4830–4834. https://doi.org/10.1073/pnas.72.12.4830

Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology*, *7*(12), 952–958. https://doi.org/10.1038/nrm2067

Marshall, N. F., & Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.270.21.12335

Matija, B., & Price, D. H. (2006). Review Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, *23*, 297–305. https://doi.org/10.1016/j.molcel.2006.06.014

Mayer, A., Di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A., & Churchman, L. S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*. https://doi.org/10.1016/j.cell.2015.03.010

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., & Cramer, P. (2012). CTD Tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*, *336*(6089), 1723–1725. https://doi.org/10.1126/science.1219651

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.1903

Meininghaus, M., Chapman, R. D., Horndasch, M., & Eick, D. (2000). Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. *Journal of Biological Chemistry*, *275*(32), 24375–24382. https://doi.org/10.1074/jbc.M001883200

Michalska, A., Blaszczyk, K., Wesoly, J., & Bluyssen, H. A. R. (2018). A positive feedback amplifier circuit that regulates interferon (IFN)-stimulated gene expression and controls type I and type II IFN responses. *Frontiers in Immunology*, *9*(MAY), 1–17. https://doi.org/10.3389/fimmu.2018.01135

Milligan, L., Huynh-Thu, V. A., Delan-Forino, C., Tuck, A., Petfalski, E., Lombraña, R., Sanguinetti, G., Kudla, G., & Tollervey, D. (2016). Strand-specific, high-resolution mapping of modified RNA polymerase II. *Molecular Systems Biology*. https://doi.org/10.15252/msb.20166869

Morrill, S. A., Exner, A. E., Babokhov, M., Reinfeld, B. I., & Fuchs, S. M. (2016). DNA instability maintains the repeat length of the yeast RNA polymerase II C-terminal domain. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M115.696252

Mostafavi, S., Yoshida, H., Moodley, D., Leboité, H., Rothamel, K., Raj, T., Ye, C. J., Chevrier, N., Zhang, S. Y., Feng, T., Lee, M., Casanova, J. L., Clark, J. D., Hegen, M., Telliez, J. B., Hacohen, N., De Jager, P. L., Regev, A., Mathis, D., & Benoist, C. (2016). Parsing the Interferon Transcriptional Network and Its Disease

Associations. *Cell*, *164*(3), 564–578. https://doi.org/10.1016/j.cell.2015.12.032

Nakatani, Y., Takeda, H., Kohara, Y., & Morishita, S. (2007). *Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates*. 1254–1265. https://doi.org/10.1101/gr.6316407.

Nejepinska, J., Malik, R., Filkowski, J., Flemr, M., Filipowicz, W., & Svoboda, P. (2012). DsRNA expression in the mouse elicits RNAi in oocytes and low adenosine deamination in somatic cells. *Nucleic Acids Research*, *40*(1), 399–413. https://doi.org/10.1093/nar/gkr702

Nekhai, S., Zhou, M., Fernandez, A., Lane, W. S., Lamb, N. J. C., Bradyr, J., & Kumar, A. (2002). HIV-1 Tat-associated RNA polymerase C-terminal domain kinase, CDK2, phosphorylates CDK7 and stimulates Tat-mediated transcription. In *Biochem. J* (Vol. 364).

Nieto Moreno, N., Villafañez, F., Giono, L. E., Cuenca, C., Soria, G., Muñoz, M. J., & Kornblihtt, A. R. (2020). GSK-3 is an RNA polymerase II phospho-CTD kinase. *Nucleic Acids Research*, *48*(11), 6068–6080. https://doi.org/10.1093/nar/gkaa322

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M., & Proudfoot, N. J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*. https://doi.org/10.1016/j.cell.2015.03.027

Nojima, T., Rebelo, K., Gomes, T., Grosso, A. R., Proudfoot, N. J., & Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell*, *72*(2), 369-379.e4. https://doi.org/10.1016/j.molcel.2018.09.004

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S. M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., & Proudfoot, N. J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Molecular Cell*, *72*(6),

970-984.e7. https://doi.org/10.1016/j.molcel.2018.10.011

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics*. *Molecular & Cellular Proteomics*, *1*, 376–386. https://doi.org/10.1074/mcp.M200025-MCP200

Ong, S.-E. E., & Mann, M. (2007). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, *1*(6), 2650–2660. https://doi.org/10.1038/nprot.2006.427

Pak, V., Eifler, T. T., Jäger, S., Krogan, N. J., Fujinaga, K., & Peterlin, B. M. (2015). CDK11 in TREX/THOC regulates HIV mRNA 3′ end processing. *Cell Host and Microbe*. https://doi.org/10.1016/j.chom.2015.10.012

Palazzo, A. F., & Koonin, E. V. (2020). Functional Long Non-coding RNAs Evolve from Junk Transcripts. In *Cell* (Vol. 183, Issue 5, pp. 1151–1161). Cell Press. https://doi.org/10.1016/j.cell.2020.09.047

Palermo, R. D., Webb, H. M., Gunnell, A., & West, M. J. (2008). Regulation of transcription by the Epstein-Barr virus nuclear antigen EBNA 2. *Biochemical Society Transactions*, *36*(4), 625–628. https://doi.org/10.1042/BST0360625

Papac-Milicevic, N., Breuss, J. M., Zaujec, J., Ryban, L., Plyushch, T., Wagner, G. A., Fenzl, S., Dremsek, P., Cabaravdic, M., Steiner, M., Glass, C. K., Binder, C. J., Uhrin, P., & Binder, B. R. (2012). The interferon stimulated gene 12 inactivates vasculoprotective functions of NR4A nuclear receptors. *Circulation Research*, *110*(8). https://doi.org/10.1161/CIRCRESAHA.111.258814

Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., & Fisher, R. P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*, *558*(7710), 460–464. https://doi.org/10.1038/s41586-018-0214-z

Patel, M. C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., Heightman, T. D., Tamura, T., & Ozato, K. (2013). BRD4 Coordinates Recruitment of Pause Release Factor P-TEFb and the Pausing Complex NELF/DSIF To Regulate Transcription Elongation of Interferon-Stimulated Genes. *Molecular and Cellular Biology*, *33*(12), 2497–2507. https://doi.org/10.1128/mcb.01180-12

Peng, J., Marshall, N. F., & Price, D. H. (1998). Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.273.22.13855

Pinhero, R., Liaw, P., Bertens, K., & Yankulov, K. (2004). Three cyclin-dependent kinases preferentially phosphorylate different parts of the C-terminal domain of the large subunit of RNA polymerase II. *Eur. J. Biochem*. https://doi.org/10.1111/j.1432-1033.2004.04002.x

Plet, A., Eick, D., & Blanchard, J. M. (1995). Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene*.

Pratt, J. M., Petty, J., Riba-Garcia, I., Robertson, D. H. L., Gaskell, S. J., Oliver, S. G., & Beynon, R. J. (2002). Dynamics of protein turnover, a missing dimension in proteomics. *Molecular & Cellular Proteomics : MCP*, *1*(8), 579–591. https://doi.org/10.1074/mcp.M200046-MCP200

Quintero-Cadena, P., Lenstra, T. L., & Sternberg, P. W. (2020). RNA Pol II Length and Disorder Enable Cooperative Scaling of Transcriptional Bursting. *Molecular Cell*, *79*(2), 207-220.e8. https://doi.org/10.1016/j.molcel.2020.05.030

Ramani, M. K. V., Escobar, E. E., Irani, S., Mayfield, J. E., Moreno, R. Y., Butalewicz, J. P., Cotham, V. C., Wu, H., Tadros, M., Brodbelt, J. S., Zhang, Y. J., Kumar, M., Ramani, V., Escobar, E. E., Irani, S., Mayfield, J. E., Moreno, R. Y., Butalewicz, J. P., Cotham, V. C., … Zhang, Y. J. (2020). Structural Motifs for CTD Kinase Specificity on RNA Polymerase II during Eukaryotic Transcription. *ACS Chemical Biology*, *15*(8), 2259–2272.

https://doi.org/10.1021/acschembio.0c00474

Ranjan, A., Nguyen, V. Q., Liu, S., Wisniewski, J., Kim, J. M., Tang, X., Mizuguchi, G., Elalaoui, E., Nickels, T. J., Jou, V., English, B. P., Zheng, Q., Luk, E., Lavis, L. D., Lionnet, T., & Wu, C. (2020). Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. *ELife*, *9*. https://doi.org/10.7554/eLife.55667

Rice, S. A., Long, M. C., Lam, V., & Spencer, C. A. (1994). RNA polymerase II is aberrantly phosphorylated and localized to viral replication compartments following herpes simplex virus infection. *Journal of Virology*, *68*(2), 988–1001. https://doi.org/10.1128/jvi.68.2.988-1001.1994

Roeder, R. G., & Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*. https://doi.org/10.1038/224234a0

Sabari, B. R. (2020). Biomolecular Condensates and Gene Activation in Development and Disease. *Developmental Cell*, *55*(1), 84–96. https://doi.org/10.1016/j.devcel.2020.09.005

Sadeq, S., Al-Hashimi, S., Cusack, C. M., & Werner, A. (2021). Endogenous double-stranded RNA. *Non-Coding RNA*, *7*(1), 1–18. https://doi.org/10.3390/ncrna7010015

Saveliev, S., Bratz, M., Zubarev, R., Szapacs, M., Budamgunta, H., & Urh, M. (2013). Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nature Methods*, *10*(11), i–ii. https://doi.org/10.1038/nmeth.f.371

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. https://doi.org/10.1016/j.molcel.2016.11.029

Schoggins, J. W. (2019). Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology*, *6*(1), 567–584. https://doi.org/10.1146/annurev-virology-

092818-015756

Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., & Passmore, L. A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2753

Schröder, S., Cho, S., Zeng, L., Zhang, Q., Kaehlcke, K., Mak, L., Lau, J., Bisgrove, D., Schnölzer, M., Verdin, E., Zhou, M. M., & Ott, M. (2012). Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111.282855

Schröder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D. A., Kaehlcke, K., Cho, S., Pollard, K. S., Capra, J. A., Schnölzer, M., Cole, P. A., Geyer, M., Bruneau, B. G., Adelman, K., & Ott, M. (2013). Acetylation of RNA Polymerase II Regulates Growth-Factor-Induced Gene Transcription in Mammalian Cells. *Molecular Cell*, *52*(3), 314–324. https://doi.org/10.1016/j.molcel.2013.10.009

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T. M., Cramer, P., Imhof, A., & Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular Cell*, *61*(2), 305–314. https://doi.org/10.1016/j.molcel.2015.12.003

Secombes, C. J., & Zou, J. (2017). Evolution of interferons and interferon receptors. In *Frontiers in Immunology* (Vol. 8, Issue MAR, p. 209). Frontiers Research Foundation. https://doi.org/10.3389/fimmu.2017.00209

Sharma, P., Lioutas, A., Fernandez-Fuentes, N., Quilez, J., Carbonell-Caballero, J., Wright, R. H. G., Di Vona, C., Le Dily, F., Schüller, R., Eick, D., Oliva, B., & Beato, M. (2019). Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription. *Molecular Cell*, *73*(1), 84-96.e7. https://doi.org/10.1016/j.molcel.2018.10.016

Sims, R. J., Rojas, L. A., Beck, D., Bonasio, R., Schüller, R., Drury, W. J., Eick, D., & Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science*. https://doi.org/10.1126/science.1202663

Solari, F. A., Dell'Aica, M., Sickmann, A., & Zahedi, R. P. (2015). Why phosphoproteomics is still a challenge. *Molecular BioSystems*, *11*(6), 1487–1493. https://doi.org/10.1039/c5mb00024f

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology*, *19*(4), 262–274. https://doi.org/10.1038/nrm.2017.115

Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., & Kirschner, M. W. (2006). Phosphorylation analysis by mass spectrometry: Myths, facts, and the consequences for qualitative and quantitative measurements. *Molecular and Cellular Proteomics*, *5*(1), 172–181. https://doi.org/10.1074/mcp.M500135-MCP200

Stifter, S. A., Bhattacharyya, N., Sawyer, A. J., Cootes, T. A., Stambas, J., Doyle, S. E., Feigenbaum, L., Paul, W. E., Britton, W. J., Sher, A., & Feng, C. G. (2019). Visualizing the Selectivity and Dynamics of Interferon Signaling In Vivo. *Cell Reports*, *29*(11), 3539-3550.e4. https://doi.org/10.1016/j.celrep.2019.11.021

Stiller, J. W., & Hall, B. D. (2002). Evolution of the RNA polymerase II C-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(9), 6091–6096. https://doi.org/10.1073/pnas.082646199

Strobl, L. J., & Eick, D. (1992). Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1992.tb05409.x

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for

interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Suh, H., Ficarro, S. B., Kang, U. B., Chun, Y., Marto, J. A., & Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular Cell*, *61*(2), 297–304. https://doi.org/10.1016/j.molcel.2015.12.021

Tellier, M., Zaborowska, J., Caizzi, L., Mohammad, E., Velychko, T., Schwalb, B., Ferrer-Vicens, I., Blears, D., Nojima, T., Cramer, P., & Murphy, S. (2020). CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. *Nucleic Acids Research*, *48*(14), 7712–7727. https://doi.org/10.1093/nar/gkaa514

Terzi, N., Churchman, L. S., Vasiljeva, L., Weissman, J., & Buratowski, S. (2011). H3K4 Trimethylation by Set1 Promotes Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.05590-11

Topisirovic, I., Svitkin, Y. V., Sonenberg, N., & Shatkin, A. J. (2011). Cap and cap-binding proteins in the control of gene expression. In *Wiley Interdisciplinary Reviews: RNA*. https://doi.org/10.1002/wrna.52

Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., & Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. *PROTEOMICS*, *15*(8), 1453–1456. https://doi.org/10.1002/pmic.201400449

Van Hoof, D., Pinkse, M. W. H., Oostwaard, D. W. Van, Mummery, C. L., Heck, A. J. R., & Krijgsveld, J. (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics [1]. In *Nature Methods* (Vol. 4, Issue 9, pp. 677–678). Nat Methods. https://doi.org/10.1038/nmeth0907-677

van Steensel, B., & Furlong, E. E. M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*, *20*(6), 327–337. https://doi.org/10.1038/s41580-019-0114-6

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., & Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.1468

Vervoort, S. J., Welsh, S. A., Devlin, J. R., Barbieri, E., Knight, D. A., Offley, S., Bjelosevic, S., Costacurta, M., Todorovski, I., Kearney, C. J., Sandow, J. J., Fan, Z., Blyth, B., McLeod, V., Vissers, J. H. A., Pavic, K., Martin, B. P., Gregory, G., Demosthenous, E., … Johnstone, R. W. (2021). The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell*, *0*(0). https://doi.org/10.1016/j.cell.2021.04.022

Vona, C. D., Bezdan, D., Islam, A. B. M. M. K., Salichs, E., López-Bigas, N., Ossowski, S., & Luna, S. D. L. (2015). Chromatin-wide profiling of DYRK1A reveals a role as a gene-specific RNA polymerase II CTD kinase. *Molecular Cell*, *57*(3), 506–520. https://doi.org/10.1016/j.molcel.2014.12.026

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., & Cramer, P. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*, *560*(7720), 607–612. https://doi.org/10.1038/s41586-018-0440-4

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M. A., Heidemann, M., Flatley, A., Imhof, A., Gut, M., Gut, I., Kremmer, E., Andrau, J. C., & Eick, D. (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. *Transcription*, *6*(5), 91–101. https://doi.org/10.1080/21541264.2015.1114983

Wan, L., Wen, H., Li, Y., Lyu, J., Xi, Y., Hoshii, T., Joseph, J. K., Wang, X., Loh, Y. H. E., Erb, M. A., Souza, A. L., Bradner, J. E., Shen, L., Li, W., Li, H., Allis, C. D., Armstrong, S. A., & Shi, X. (2017). ENL links histone acetylation to oncogenic

gene expression in acute myeloid leukaemia. *Nature*. https://doi.org/10.1038/nature21687

Wang, W., Xu, L., Su, J., Peppelenbosch, M. P., & Pan, Q. (2017). Transcriptional Regulation of Antiviral Interferon-Stimulated Genes. *Trends in Microbiology*, *25*(7), 573–584. https://doi.org/10.1016/j.tim.2017.01.001

Wei, M. T., Chang, Y. C., Shimobayashi, S. F., Shin, Y., Strom, A. R., & Brangwynne, C. P. (2020). Nucleated transcriptional condensates amplify gene expression. *Nature Cell Biology*, *22*(10), 1187–1196. https://doi.org/10.1038/s41556-020-00578-6

West, M. L., & Corden, J. L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics*, *140*(4), 1223–1233.

Whitelaw, E., & Proudfoot, N. (1986). Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO Journal*, *5*(11), 2915–2922. https://doi.org/10.1002/j.1460-2075.1986.tb04587.x

Wienerroither, S., Shukla, P., Farlik, M., Majoros, A., Stych, B., Vogl, C., Cheon, H. J., Stark, G. R., Strobl, B., Müller, M., & Decker, T. (2015). Cooperative Transcriptional Activation of Antimicrobial Genes by STAT and NF-κB Pathways by Concerted Recruitment of the Mediator Complex. *Cell Reports*, *12*(2), 300–312. https://doi.org/10.1016/j.celrep.2015.06.021

Willis, I. M. (1993). RNA polymerase III: Genes, factors and transcriptional specificity. In *European Journal of Biochemistry*. https://doi.org/10.1111/j.1432-1033.1993.tb17626.x

Winter, G. E., Mayer, A., Buckley, D. L., Erb, M. A., Roderick, J. E., Vittori, S., Reyes, J. M., di Iulio, J., Souza, A., Ott, C. J., Roberts, J. M., Zeid, R., Scott, T. G., Paulk, J., Lachance, K., Olson, C. M., Dastjerdi, S., Bauer, S., Lin, C. Y., … Bradner, J.

E. (2017). BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.06.004

Wong, K. H., Jin, Y., & Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2014.03.024

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., & Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2005.11.024

Yamaguchi, Y., Shibata, H., & Handa, H. (2013). Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol. 1829, Issue 1, pp. 98–104). Biochim Biophys Acta. https://doi.org/10.1016/j.bbagrm.2012.11.007

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., & Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*. https://doi.org/10.1016/S0092-8674(00)80713-8

Yang, Z., Yik, J. H. N., Chen, R., He, N., Moon, K. J., Ozato, K., & Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2005.06.029

Young, R. A. (1991). RNA polymerase II. In *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.bi.60.070191.003353

Yurko, N. M., & Manley, J. L. (2018). The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. *Transcription*, *9*(1), 30–40. https://doi.org/10.1080/21541264.2017.1338176

Zaborowska, J., Baumli, S., Laitem, C., O'Reilly, D., Thomas, P. H., O'Hare, P., & Murphy, S. (2014). Herpes simplex virus 1 (HSV-1) ICP22 protein directly interacts with cyclin-Dependent kinase (CDK)9 to Inhibit RNA polymerase II transcription elongation. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0107654

Zaborowska, J., Egloff, S., & Murphy, S. (2016). The pol II CTD: New twists in the tail. *Nature Structural and Molecular Biology*, *23*(9), 771–777. https://doi.org/10.1038/nsmb.3285

Zaborowska, J., Isa, N. F., & Murphy, S. (2016). P-TEFb goes viral. In *BioEssays* (Vol. 38, pp. S75–S85). John Wiley and Sons Inc. https://doi.org/10.1002/bies.201670912

Zhang, David., Mosley, A. L., Ramisetty, S. R., Rodríguez-Molina, J. B., Washburn, M. P., & Ansari, A. Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M111.335687

Zhang, G., & Neubert, T. A. (2009). Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 527, p. 79). NIH Public Access. https://doi.org/10.1007/978-1-60327-834-8_7

Zhang, Hanghang, Pandey, S., Travers, M., Sun, H., Morton, G., Madzo, J., Chung, W., Khowsathit, J., Perez-Leal, O., Barrero, C. A., Merali, C., Okamoto, Y., Sato, T., Pan, J., Garriga, J., Bhanu, N. v., Simithy, J., Patel, B., Huang, J., … Issa, J. P. J. (2018). Targeting CDK9 Reactivates Epigenetically Silenced Genes in Cancer. *Cell*, *175*(5), 1244-1258.e26. https://doi.org/10.1016/j.cell.2018.09.051

Zhang, Huimin, Rigo, F., & Martinson, H. G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Molecular Cell*, *59*(3), 437–

448. https://doi.org/10.1016/j.molcel.2015.06.008

Zhao, D., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F. W., Zhong, G., Liu, K., Li, W., Moffat, J., Vedadi, M., Min, J., Pawson, T. J., Blencowe, B. J., & Greenblatt, J. F. (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*. https://doi.org/10.1038/nature16469

Zhou, Q., Li, T., & Price, D. H. (2012). RNA polymerase II elongation control. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-052610-095910

VI Appendix

Acknowledgments

Throughout the five-year lifetime of my Ph.D., I have received a great deal of support and assistance.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Dirk Eick, for saving my Ph.D. life by accepting me as a student and for his unwavering support of my Ph.D. study and related research, as well as his patience, motivation, scientific advice and knowledge, and many insightful discussions and suggestions. He's my go-to person for science problems, and he was instrumental in helping me finish this thesis. His patient support was invaluable during the research and writing of this thesis. I could not have asked for a greater supervisor for my Ph.D.

Aside from my supervisor, I'd like to express my gratitude to the other members of my thesis committee: Prof. Dr. Axel Imhof and Prof. Dr. Heinrich Leonhardt, not only for their generous collaborations and insightful comments but also for the challenging question that prompted me to broaden my research to include multiple perspectives.

I also thank my project collaborators, For the proteomics study, I appreciate Prof. Dr. Axel Imhof, Dr. Ignasi Forne, and Dr. Tobias Straub, who were all part of my project. For the RNA-seq study, I thank Dr. Stefan Krebs, Prof. Dr. Caroline Friedel, and Dr. Michael Kluge. For the TT-seq experiment, I appreciate Prof. Dr. Patrick Cramer, Dr. Björn Schwalb, Dr. Anna Sawicka, and Le Xiong. The interferon experiment was helped by PD. Dr. Josef Mautner and Dr. Dinesh Adhikary.

My heartfelt gratitude also goes out to the Eick lab's present and previous members: Dr. Michaela Rohrmoser, Dr. Nilay Shah, Dr. Yousra Yahia, and Hesham Elsaman. I especially thank Dr. Tim-Micael Decker, the project's initiator, for introducing me to the research and allowing me to participate in experiments. Even if you are out of the lab and in a new position, your grateful remote support for this project continues. This research would not be feasible without his invaluable assistance. I'd also want to thank Anita Gruber-Eber, who has spent the most time in the lab with me over the previous

two years. Anita is a hilarious and generous friend who has helped me immensely in both the lab and life. I like her cheerful attitude and her capacity to grin in the face of difficulties.

I'd like to thank Lara Ceglarek and Cara Sickinger, my intern students with whom I collaborated. They're all talented, motivated, and joyful young scientists that remind me of myself when I first started a research in the lab.

I gratefully acknowledge the funding received towards my PhD from the China Scholarship Council (CSC), Helmholtz Zentrum München (HMGU) and SFB1064 collaborative research center (CRC).

In addition, I could not have finished this dissertation without the help of my friends, who offered delightful distractions from my study by enjoying wonderful experiences in the Alps and the city.

Last but not least, I'd like to express my gratitude for my parents' unwavering love and support with numerous video calls. You're always there with me.