Norwegian University of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Characterization of Chromatin-Bound Interactome of Cyclin-Dependent Kinase 12 (CDK12)

Sigrid Berg
Master of Science, Biotechnology

Characterization of Chromatin-Bound Interactome of Cyclin-Dependent Kinase 12 (CDK12)

Oslo University Hospital, Rikshospitalet, Department of Microbiology,

and

The Norwegian University of Life Sciences, Faculty of Chemistry, Biotechnology and Food Science

Acknowledgements

The work of this thesis was carried out at the Department of Microbiology at Oslo University Hospital (OUS), as part of a master's degree in Biotechnology at the Norwegian University of Life Sciences (NMBU).

Firstly, I would like to thank my supervisor Deo P. Pandey, PhD, for all supervision, support and enthusiasm throughout the year. Thank you for always being available for guidance in the laboratory and in the process of writing this thesis. I would also like to express my gratitude to my internal supervisor, Professor Dag Inge Våge, for his guidance and for devoting his time proofreading the thesis.

Thanks to Anna Lång, PhD, Emma Lång, PhD, and Stig Ove Bøe, PhD, for helping me out with immunofluorescent analyses and confocal microscopy. Moreover, thanks to Sachin Singh, PhD, and the Proteomic Core Facility at Oslo University Hospital, Rikshospitalet, for conducting liquid chromatography tandem mass spectrometry (LC-MS/MS) and assisting me in performing bioinformatic analyses.

My sincerest thanks go to my fellow students in the student's office. I have greatly appreciated the company, valuable discussions and friendship throughout the year. You have all brightened my day countless of times. Special thanks to Silje Lier and Kasturi Raorane for always being willing to devote their time for theoretical discussions.

I would like to express my gratitude to my closest ones. My sincerest thanks to my mum, dad and five siblings for the endless love and support, and for always believing in me. You are all truly inspiring people that I am incredibly grateful to have by my side in all aspects of life. Special thanks to my dad for introducing me to the world of biotechnology and for encouraging me to start this journey, and to my mum for being my greatest moral support. I would also like to thank my friends for the patience, support and joyful moments. Last, but absolutely not least – a sincere thanks to Sigve. Thank you for standing by my side every step of the way, and for always encouraging me to strive for greatness. You have been my greatest comfort and motivation.

Sigrid Berg
Ås, May 2020

Abstract

Proper gene expression relies on control by an abundant number of factors and mechanisms, and the orchestration of these fine-tuned programs is fundamental in cellular functionality. Transcription regulation by RNA polymerase II (RNA Pol II) carboxy-terminal domain (CTD) phosphorylation influence global transcription levels and manufacture the adjustment of final RNA products. Several kinases have been reported to be involved in regulation of these intracellular regulatory pathways by phosphorylation of Pol II CTD, and their functional disturbance is linked transcription mis-regulation and disease. Two of the kinases described are cyclin-dependent kinase 12 (CDK12) and CDK13. Research suggests for a role of CDK12 in phosphorylating CTD Serine 2 (Ser2) for productive elongation, while the evidence of CDK13 in its role as a CTD kinase is undefined. They both appear to possess roles in maintenance of genomic stability, but their exact roles are far from established.

We here aimed to elucidate the roles of chromatin-bound CDK12 and CDK13 expressed at endogenous levels. From technical causes, CDK13 was excluded from further analyses. This study consequently focused on CDK12 as a CTD Ser2 and Ser5 kinase. Moreover, we aimed to identify the interactome of chromatin-bound CDK12 to acquire knowledge on its role in transcription and RNA processing events. Heat shock treatment of cells was used as a tool to disturb RNA pol II transcriptional programs, and THZ531 was used to inhibit CDK12/13 catalytic activity.

Immunofluorescent analyses reveled changed distribution patterns of RNA Pol II and CDK12-GFP upon heat shock- and THZ531 treatment. CTD Ser2 phosphorylation levels decreased under both treatment methods. Obtained results of CTD Ser5 phosphorylation levels were diverging, hence undecidable. Moreover, mass spectrometry (MS) combined with western blot analyses revealed CDK12 to dissociate from chromatin after 30 minutes heat shock treatment. Bioinformatic analyses of the MS data further revealed mRNA processing to be among the most enriched functional groups of CDK12-GFP bound proteome under non-heat shock conditions. RNA splicing- and transcription regulatory factors, and proteins involved in the mRNA surveillance pathway were among the most enriched terms. Together, our data support previous research demonstrating a decrease in Ser2 phosphorylation state upon CDK12/13 inhibition, as well as CDK12 to be a factor involved in transcription and RNA processing events.

Sammendrag

Korrekt genekspresjon er avhengig av et stort antall faktorer, og en finjusert balanse av deres mekansimer er avgjørende for cellens funksjon. Transkripsjonsregulering ved RNA Pol II CTD fosforylering påvirker transkripsjon på et globalt nivå og justerer endelig nivå av cellens RNA produkt. Tidligere forskning hevder at flere proteiner, såkalte kinaser, er involvert i transkripsjonsregulering ved fosforylering av Pol II CTD, og svekket funksjonalitet av disse proteinene er knyttet til feilregulering av transkripsjon, og eventuell påfølgende sykdom. To av kinasene som er beskrevet er CDK12 og CDK13. Tidligere studier hevder at CDK12 fosforyliserer CTD Ser2 for å fremme produktiv elongering, mens CDK13 i sin rolle som CTD kinase fortsatt er udefinert. De ser begge ut til å være essensielle for å opprettholde genomisk stabilitet, men deres eksakte roller er til dags dato ikke etablert.

Målet med denne oppgaven var å belyse endogent uttrykt CDK12 og CDK13 i sine roller som kromatin-bundne CTD kinaser. CDK13 ble etter tekniske årsaker ekskludert fra videre analyser. Denne studien fokuserte som følger på CDK12 som CTD Ser2 og Ser5 kinase. Videre var målet å identifisere proteiner som naturlig interagerer med kromatin-bundet CDK12 for å oppnå kunnskap om dette proteinets rolle i transkripsjon og RNA-prosessering. Varmebehandling av celler ble brukt som et verktøy for å forstyrre RNA Pol II under transkripsjonen, og THZ531 ble brukt for å forstyrre CDK12/13 katalytisk aktivitet.

Immunofluorescerende analyser avslørte endrede fordelingsmønster av RNA Pol II og CDK12-GFP som følger av varme- og THZ531 behandling. Fluorescerende nivåer av CTD Ser2 sank under begge behandlingsmetodene, mens nivåene av CTD Ser5 var tvetydige. MS kombinert med western blot analyse avslørte at CDK12 frigjøres fra kromatin etter 30 minutter med varmebehandling. Bioinformatiske analyser demonstrerte videre at mRNA-prosesserende faktorer var blant de største funksjonelle gruppene som under normale forhold er bundet til CDK12-GPP. Faktorer involvert i RNA spleising- og transkripsjonsregulering, samt i kvalitetskontroll av mRNA, var blant de mest interessante funnene.

Våre data støtter tidligere forskning som viser en nedgang i CTD Ser2 fosfoyleringsnivå etter CDK12/13-inhibering, samt at CDK12 er involvert i transkripsjon og prosessering av RNA.

Abbreviations

3´UTR	3´- untranslated region
APA	alternative cleavage and polyadenylation
BAC	bacterial artificial chromosome
CAK	CDK-activity kinase
CDK	cyclin-dependent kinase
cDNA	complementary DNA
CF	cleavage factor
ChEP	chromatin enrichment for proteomics
ChIP	chromatin immunoprecipitation
CMTR1	cap methyltransferase 1
CPA	cleavage/polyadenylation apparatus
CPSF	cleavage and polyadenylation specificity factor
CS	cytoplasmic soluble
CstF	cleavage stimulation factor
CTD	carboxy-terminal domain
Ctrl	control
CycK	Cyclin K
DDR	DNA damage repair
DSIF	DRB sensitivity factor
EJC	exon junction complex
FACS	fluorescent activated cell sorting
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
HCC	hepatocellular carcinoma
HeLa	Henrietta Lacks
HGSOC	high grade serous ovarian carcinomas
HRP	horseradish peroxidase
HR repair	homologous repair
HS	heat shock
HSF	heat shock transcription factors
Hsp	heat shock protein
IP	immunoprecipitation
LC-MS/MS	liquid chromatography tandem mass spectrometry
mAb	monoclonal antibody
miRNA	micro RNA
mRNA	messenger RNA
MS	mass spectrometry
NEXT	nuclear exosome targeting

NHS	non-heat shock
NMBU	Norwegian University of Life Sciences
NP	nuclear pellet
NS	nuclear soluble
n.s	non-significant
OUS	Oslo University Hospital
pAb	polyclonal antibody
PAF	RNA polymerase-associated factor
PAS	polyadenylation signal
PAXT	poly(A)-tail exosome targeting
PCPA	premature cleavage and polyadenylation
PIC	pre-initiation complex
P-TEFb	positive transcription factor b
RNA	ribonucleic acid
RNAi	RNA interference
RNA Pol	RNA polymerase
RNGTT	RNA guanylyltransferase and 5′-phosphatase
RNMT	RNA guanine-7 methyltransferase
RPLP0	60S acidic ribosomal protein P0
rRNA	ribosomal RNA
RT	room temperature
RT-qPCR	quantitative reverse transcription polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser	serine
Ser2-p	phosphorylated serine 2
Ser5-p	phosphorylated serine 5
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particles
SRSF	Serine/arginine-rich protein-specific kinase
Suz12	polycomb protein SUZ12
TDP	tandem duplicator phenotype
tRNA	transfer RNA
Thr	threonine
Tyr	tyrosine
Xrn2	5′-3′exoribonuclease 2

Table of contents

1. Introduction

1.1 Gene regulation in eukaryotic cells

During the last few decades, it has become clear that gene regulation in higher eukaryotes is a complex and fine-tuned process comprising an abundant number of factors [1]. The orchestration of these processes is fundamental for survival and functionality of cell, and exploring the complexity of these mechanisms gives rise to the opportunity of preventing and treating many clinical pathologies. This regulation is controlled both at the level of transcription initiation and in subsequent steps [1].

Transcription is the first step in gene expression. The process is carried out by three unique nuclear polymerases designated RNA polymerase (RNA Pol) I, II and III [2]. These three enzymes traverse along gene bodies in terms of accurate synthesizing of complementary RNA transcripts with coding DNA strands as template [2].

1.2 Eukaryotic RNA polymerases

1.2.1 RNA polymerase I

RNA polymerase I (Pol I) is a 13-subunit enzyme that is solely devoted to the transcription of ribosomal RNA (rRNA) genes from the ribosomal DNA repeat loci [3]. These genes are first transcribed into 47S pre-rRNA, and subsequently processed into mature 28S, 18S and 5.8S rRNA which carry out fundamental structural and catalytic functions within the ribosomal core. Hence, Pol I is major director of rRNA gene expression, and key factor in regulation of ribosomal biogenesis [3].

1.2.2 RNA polymerase II

RNA polymerase II (Pol II) is a globular shaped, multiprotein complex consisting of 12 subunits, titled Rpb1 – Rpb12, in approximate order of their size [4]. X-ray crystallography has revealed structures of RNA Pol II in yeast, and recent studies have reported highly similar structures of mammalian RNA Pol II [4].

The largest subunit of Pol II, Rpb1, is in possession of a long, repetitive polypeptide extension termed carboxy-terminal domain (CTD) [5], which is a unique feature of Pol II and

distinguishes it from the other two polymerases. The CTD is responsible for recruitment of transcriptional factors, and is substantial in initiation of DNA transcription [5]. DNA is found to enter the cleft (Rpb1) and pass the upper and lower jaws (Rpb1, Rpb2, Rpb5, Rpb9) [6]. The clamp (Rpb1, Rpb2) works as the binding site of downstream DNA, and unwinding occurs here. The clamp and the jaw stabilize the downstream end of the DNA and allows for the cleft to open and close. A gap in the floor of the cleft works as an entrance for substrate nucleoside triphosphates in addition to an exit of nascent RNA [6].

Pol II transcribes messenger RNA (mRNA) from coding genes, as well as microRNA (miRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) from non-coding genes [7]. Post transcription, pre-mRNA is spliced into mature mRNA by removal of intronic sequences. Mature mRNAs are recruited to cytosolic ribosomes and subsequently translated into proteins that are involved in virtually every cell function. miRNAs can direct mRNA destabilization and translational repression [7], snRNAs are involved in pre-mRNA splicing and snoRNAs are involved in both translation and pre-mRNA splicing [8]. Hence, Pol II transcripts are prevalent in gene regulation and expression.

1.2.3 RNA polymerase III

RNA polymerase III (Pol III) is composed of 17 subunits and is the largest of the three eukaryotic RNA polymerases [9]. Like Pol I, Pol III synthetizes small series of highly expressed infrastructural RNAs. Its main task is to catalyze the synthesis of transfer RNAs (tRNAs), 5S rRNAs and U6 spliceosomal snRNAs. The 5S rRNAs are, like Pol I-transcribed rRNAs, components of the translational machinery, while tRNAs during translation serve as the adaptor required for proper translation of mRNA into proteins. U6 spliceosomal snRNAs are essential for pre-mRNA splicing [9, 10]. Hence, these transcripts are crucial for the protein synthesis.

1.2.4 RNA polymerase II and its transcriptome as main focus

Transcription by Pol II is one of the most profoundly regulated processes within the cell and is during the transcription cycle subject to multiple regulatory events [11]. Pol II CTD is dynamically modified during the transcription cycle and is in this manner essential in regulation of the multiple steps during transcription. These co-transcriptional actions influence global transcription levels and manufacture the adjustment of final RNA products

[5, 11]. Mis-regulation of these fine-tuned mechanisms can potentially cause deleterious alterations in RNA expression levels – the blueprint of life. Hence, I will further stay focused on the role of RNA Pol II CTD in regulation of the transcriptional cycle.

1.3 Regulation of transcription by RNA polymerase II CTD phosphorylation

1.3.1 RNA polymerase II CTD

RNA Pol II CTD comprise a tandem array of seven evolutionary preserved amino acid repeats with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (also termed YSPTSPS) [2, 11]. The number of heptad repeats differ between organisms and seems to correspond with genomic complexity. The mammalian RNA pol II and its corresponding CTD comprise 52 of these repeats, whereas the yeast *Saccharomyces cerevisiae* (*S. cerevisiae)* CTD comprise of 26 repeats [11, 12].

1.3.2 CTD modifications

RNA Pol II modifications are principal processes in transcription regulation and gene expression. Modifications of the CTD consensus heptapeptide regulate the RNA pol II machinery in the different phases of transcription and conjoin transcription with co-transcriptional processes as splicing [13], 3´end formation and processing [14, 15, 16].

The CTD is exposed to phosphorylation and glycosylation on Tyr1, Ser2, Thr4, Ser5 and Ser7 [17], and the two prolines are prone to be isomerized [18]. Modifications of the CTD throughout the transcription cycle is prerequisite in recruitment and coordination of transcription and RNA processing factors, as well as histone-modification enzymes [11]. I will here stay focused on phosphorylation of the CTD.

1.3.3 CTD phosphorylating events

Chromatin immunoprecipitation (ChIP)-assays are applied to investigate changes in Pol II localization and CTD phosphorylation state over gene bodies. In human cells, ChIP-data have revealed Tyr1, Ser5 and Ser7 to be phosphorylated at the 5´end of mammalian protein-coding genes, whereas Thr4 and Ser2 are phosphorylated over the gene body and near the 3´end (figure 1) [19]. The CTD phosphorylation state is quite distinct and definite to the different phases that takes place under transcription, and the correct movement of RNA pol II during

transcription relies on the phosphorylation state of the CTD [20, 21]. The phosphorylation patterns of Pol II CTD are similar in yeast and mammals. The eminent divergence in yeast is Tyr1-p, which is enriched over the gene bodies but decreases just before the 3´end of genes (figure 1) [22, 23].

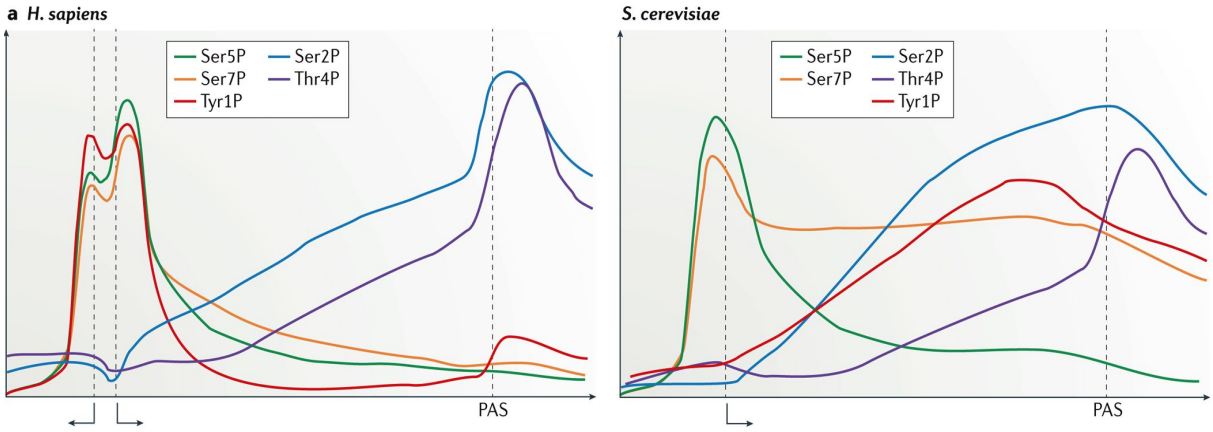


Figure 1. Average ChIP-profiles of phosphorylation state across gene-bodies in *H. sapiens* and *S. cerevisiae.*
In humans, Ser5, Ser7 and Tyr1 are phosphorylated near the transcription start site (TSS), whereas Ser2 and Thr4 are phosphorylated at the 3´end of gene bodies. The phosphorylation pattern in yeast is similar to in humans, except from Tyr1 which in yeast peaks just before the polyadenylation signal (PAS) [5].

1.4 Major steps of gene regulation and 3´end processing coupled to transcription

The CTD phosphorylation state regulates Pol II during the three major phases of transcription – initiation, elongation and termination.

1.4.1 Transcriptional initiation

During early transcription initiation, RNA Pol II with its unmodified CTD is recruited to promoters [24, 25], generating a pre-initiation complex (PIC) with the Mediator and the general tanscription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH [26]. Mediator is first mobilized and recruited to enhancer regions, of which Mediator has high affinity against unphosphorylated Pol II CTD [5, 27]. The Mediator-Pol II interaction must cease for Pol II to leave the promoter for productive elongation, and disruption of this contact is dependent on phosphorylation of Pol II CTD [28].

The mechanism behind Pol II release from Mediator still remains obscure in human cells. However, in yeast, evidence points towards that Pol II liberation requires the kinase subunit

of TFIIH to phosphorylate CTD Ser5 [29]. This event minimizes the affinity between Pol II CTD and Mediator. Pol II subsequently frees from PIC, allowing it to escape the promoter and step into the early elongation phase [28, 20].

1.4.2 Transcriptional elongation

Phosphorylation of the CTD is further involved in a network which is linked to regulatory events during elongation. A key regulatory event during the elongation-phase is promoter-proximal pausing, which takes place in the shift between early and productive elongation [30]. This mechanism is thought to be essential in facilitating rapid and synchronal transcriptional activity upon stimulation, as well as to serve as a checkpoint affecting downstream RNA processing factors required for splicing and capping [30].

In metazoan genes, promoter-proximal pause occurs after Pol II has synthetized around 20-100 nucleotides. It enters a paused state before it either terminates transcription or enters productive elongation, thus providing a window for capping to occur [1, 31]. Phosphorylation of Ser5 and Ser7 within Pol II CTD marks RNA Pol II initiation and is required for establishing pausing, as well as creating a binding scaffold for the 5´capping enzyme [32].

Two key regulators, the DRB sensitivity factor (DSIF) and the negative elongation factor (NELF), are required for this transiently halt of the elongation process [33, 34]. Elongation is proceeded by the means of positive transcription factor b (P-TEFb). P-TEFb phosphorylates Pol II CTD Ser2, in addition to the Spt5 subunit of DSIF and the E subunit of NELF, inducing the detachment of NELF which stimulates Pol II-pause release [30, 35]. During productive elongation, the phosphorylation state of Ser5 and Ser7 decrease, while phosphorylated Ser2 (Ser2-p) increase. P-TEFb dependent phosphorylation of CTD Ser2 stimulates recruitment of elongation and RNA-processing factors needed further downstream of the nascent transcript [5, 11].

To understand how pause Pol II is released into elongation, one can look into the Pol II-DSIF-PAF-SPT6 elongation complex. P-TEFb phosphorylation affects competition between NELF and elongation factor PAF1 complex (PAF), leading to discharge of NELF and binding of PAF [36]. This is in agreement with an earlier observation of PAF being required for pause release in cells [37].

1.4.3 The CTD and co-transcriptional mRNA processing

Before exported to nucleus for translation, most eukaryotic pre-mRNAs undergo 3′end processing and termination, where precursor mRNA (pre-mRNA) is developed into mature mRNA. Eukaryotic mRNA 3′end processing is series of co-transcriptional events, in which pre-mRNA is capped at the 5′end, spliced at intronic sites and polyadenylated at the 3′end [15]. The 3′end processing machinery also triggers transcription termination [38]. These enzymatic processes together promote mRNA stability, growth and viability [21], herein enhance translation of mRNA into protein.

Transcription and RNA processing are firmly interconnected processes as pre-mRNA 5′capping is affiliated to initiation, splicing to elongation, and 3′end processing to termination. Co-transcriptional phosphorylation events of the Pol II CTD allows for dynamical coupling of transcription with mRNA processing due to recruitment of various processing factors [39]. In this way, each Pol II phospho-isoform and its accompanying interactome takes place in Pol II journey across the gene body.

1.4.3.1 5′cap formation

During early elongation, rapid 5′end capping of nascent mRNA transcripts is required for subsequent mRNA processing events, including splicing, 3′end cleavage and polyadenylation, to occur [40]. This 5′cap further protects mRNA from degradation, recruit factors for export and translation, and marks mRNA to prevent recognition by the innate immune system [41]. Initiation of these RNA metabolic processes requires phosphorylation of the CTD.

The first nucleotide transcribed of nascent mRNA possesses a 5′triphosphate, in which 7-metylguanosione cap subsequently is co-transcriptionally added [42]. During the early phase of this process, the CTD is modified: phosphorylation of CTD Ser5 during transcription initiation mediates recruitment of RNA guanylyltransferase and 5′-phosphatase (RNGTT) to the nascent transcript [41]. RNGTT immediately removes the terminal phosphate group after the transcript emerges from Pol II, and subsequently adds an inverted guanosine cap [42].

This inverted guanosine cap acts as an enzymatic substrate for sets of methyltransferases. The 7-nitrogen of guanine is methylated by RNA guanine 7-methyltransferase (RNMT) [42].

RNMT is recruited to nascent RNA, probably in an indirectly RNA Pol II CTD-phospho dependent matter [42].

The initial transcribed nucleotides are subsequently methylated by Cap Methyltransferase 1 (CMTR1) [43]. CMTR1 interacts with Ser5-p CTD. This phospo-CTD mediated CMTR1-recruitment is thought to affect attachment of cap-binding protein complexes and promote ribosomal subunit binding [43].

Herein, Pol II CTD modifications unifies transcription and co-transcriptional RNA metabolism for 5´capping to occur. How transcription and RNA processing further occur in a dynamic manner, and at the same time, can be explained through the "RNA 3´end processing machinery" [38].

1.4.3.2 Recognition of the polyadenylation signal and endonucleolytic cleavage

Endonucleolytic cleavage of transcribed RNA requires co-transcriptional recognition of a polyadenylation signal (PAS) [15]. Mammalian pre-mRNA comprise three sequence elements which constitutes the polyadenylation signal, and four ancillary elements which regulates the 3´end processing. PAS consists of the polyadenylation signal sequence AAUAAA, and the four ancillary elements comprise of the cleavage site, a G/U-rich sequences downstream of the cleavage site, UGUA-containing sequences located upstream of the PAS, and G-rich sequences located downstream of the cleavage site [15].

The human mRNA 3´end processing machinery, also termed cleavage/polyadenylation apparatus (CPA), comprise the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors Im (CFIm) and IIm (CFIIm), and poly(A) polymerase [44]. These multi-subunit complexes are required for effective cleavage to occur [15, 44].

RNA Pol II transcription of the PAS promotes for recruitment of processing factors to nascent mRNA [45]. CPSF, CstF and CFIm respectively recognize the AAUAAA sequence, the U/GU-rich element and the UGUA sequence element of the nascent mRNA. Binding of these complexes recruits other factors, including CFIIm and poly(A) polymerase [46]. For these events to occur, co-transcriptional modification of the CTD is required. Pol II is paused approximately 1-5 kb downstream of the cleavage site, in which CTD Ser2 is

hyperphosphorylated (figure 1). Phospohorylation of CTD Ser2 on Pol II enhance binding of CPSF and CstF to the nascent RNA, in which CPSF subsequently catalyses an endonucleolytic cleavage [45, 46]. This cleavage reaction unveils a 3´OH terminus which is subsequently polyadenylated.

1.4.3.3 Polyadenylation

In eukaryotes, the poly(A)-tail of the protein-coding mRNA precursors is highly conserved [15]. This tail is crucial for mRNA stability and regulation of translation, and its length affect gene expression. Weakened polyadenylation can result in attenuated gene expression, and defective polyadenylation can lead to upregulated gene expression. These alterations can consequently lead to disease. For instance is an extended tail coupled to poor prognosis in certain types of cancer [47], demonstrating the importance of the precision in the poly(A)-dependent mechanism [15].

The polyadenylation process is initiated by the endonucleolytic cleavage of pre-mRNA, leading to a free hydroxyl group at the cleavage site and addition of adenosine monophosphate by poly(A) polymerase [48]. The polyadenylation mechanism, in light of the CTD, is not extensively studied in mammalian cells. However, in yeast, polyadenylation is less efficient upon deletion of the CTD [49]. Another coupling of CTD-dependent polyadenylation is found in Drosophila, where CTD Ser2 phosphorylation by P-TEFb/CDK9 was proposed to be prerequisite for polyadenylation reactions at *hsp70* genes to occur [48, 50, 51].

1.4.3.4 Alternative polyadenylation

The termination process at the 3´end of the gene is also subject to transcriptional regulation. The PAS defines the edge of the 3´UTR (3´-untranslated region), and selection of the cleavage site is hence important in gene regulatory events [15, 48]. Alternative cleavage and polyadenylation (APA) results in isoforms with different 3´UTRs and/or coding sequences [48]. mRNA 3´UTR length is directly linked to the cytoplasmic function of mRNA, including translatability, RNA stability and nuclear transportation [52]. Thus, PAS selection is a key feature in gene expression regulation.

This fashion of transcriptional regulation is intimately correlated with RNA Pol II CTD phosphorylation. Lemay et al. [51] described Seb1 in fission yeast *Schizosaccharomyces pombe*. Seb1 (homolog of human SCAF4/8) is found to be crucial for poly(A)-site selection. After connecting to the Pol II CTD, Seb1 in terms recognize specific sequence motifs downstream of the poly(A)-site on pre-mRNA, leading to alternative polyadenylation site selection [48, 51].

1.4.3.5 Splicing

Eukaryotic genes are comprised of segments of DNA, introns, located between two exons of the gene [10]. Introns are removed to generate mRNA with functional open reading frames in a process called pre-mRNA splicing. Splicing is predominantly a co-transcriptional process, which is catalyzed by a dynamic macromolecular ribonucleoprotein complex: the spliceosome [10, 53].

The spliceosome assembles on nascent RNA through sequential interference events of small nuclear ribonucleoprotein particles (snRNPs) which works in partnership with a number of non-snRNP auxiliary proteins [10]. As soon as introns are transcribed, U1 and U2 identifies 5´end splice sites at intron boundaries. As Pol II reaches the end of the intron, U4/U5 and U6 are recruited. Splicing involves two trans-esterification reactions leading to fusion of two exons and release of lariat-shaped intron [10]. Of relevance, Gu et al. [54] found that Pol II CTD mutated at Ser2-positions fails to recruit U2 snRNP to 5´splice site, further indicating an essential role of RNA Pol II CTD in splicing.

Co-transcriptional mRNA processing is in mammalian cells tightly coupled termination, which is discussed in the section below.

1.4.4 Transcriptional termination

For RNA Pol II to dismantle from chromatin and prepare for new transcription cycles, transcription termination must occur. As Pol II traverses the PAS of gene 3´end, it slows down and prepares for transcription termination [55]. Two CTD-dependent termination events are presently described for PAS recognition-triggered termination:

The allosteric model is based on Pol II recognition of the PAS, leading to a conformational change in Pol II active site, following Pol II release [55]. This recognition is likely based on CPA congregation on Pol II CTD. Conformingly, Zhang et al. [56] found CPA factor Pcf11 to cause dissociation of Pol II and the nascent transcript from chromatin through conformational changes of the CTD [55].

The torpedo model describes 5′-3′exoribonuclease activity of transcripts still being synthetized after PAS recognition and release of nascent RNA, in which 5′-3′exoribonucelase 2 (Xrn2) is recruited to PAS and degrades residual downstream RNA Pol II transcripts [55], [57]. Hence, Xrn2 degrades nascent RNA faster than Pol II synthetizes it. Xrn2, in addition to its yeast homologue Rat1, are both found to be recruited in a CTD-dependent manner. Rat1 is recruited to the CTD collectively with its associated protein Rtt103, in which Rtt103 connects to Ser2-p CTD [58]. In addition, increased levels of Thr4-p downstream of the PAS is in yeast coupled with Rtt103 binding to the Pol II [5, 50, 55]. Thus, these CTD phosphorylation events together collaborate in regulating transcription termination nascent mRNA.

The importance of these CTD-dependent regulatory events has stimulated exertion to identify the kinases behind. Research have revealed several kinases that are involved in these processes, and these findings are discussed in the following section.

1.5 Regulation of CTD phosphorylation by transcriptional cyclin-dependent kinases

Several protein kinases have been reported to be involved in these intracellular regulatory pathways. Cyclin-dependent kinases (CDKs) are one of the functional groups described [59]. The CDK family comprise 20 serine/threonine kinases which are divided into two subclasses: cell cycle regulating kinases (including CDK1, CDK2, CDK4 and CDK6), and transcription regulating kinases (CDK7, CDK8, CDK9, CDK12 and CDK13). They have in common their interaction with an activating cyclin-partner [59].

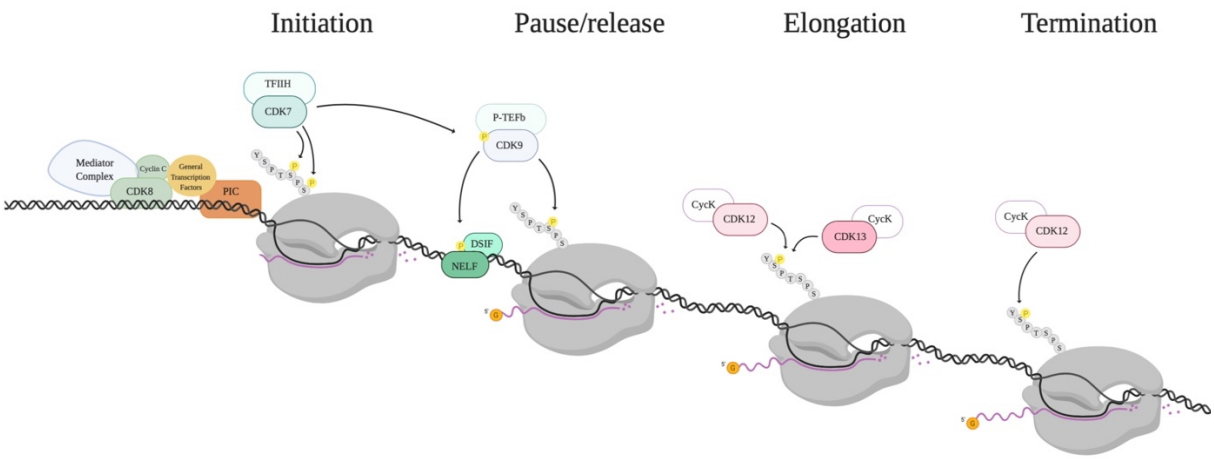


Figure 2. Schematic drawing of CDK-dependent phosphorylation of mammalian RNA Pol II CTD during initiation, elongation and termination. CDK8/Mediator together with other elements of the PIC recruit RNA Pol II to promoter regions. CDK7/TFIIH-phosphorylation of CTD Ser5 and Ser7 is required for promoter-proximal pause, whereas TFIIH also phosphorylates CDK9/P-TEFb in terms of activating the complex for further kinase activity. Phosphorylation of Ser5 also mediates 5´capping of nascent mRNA. CDK9/P-TEFb phosphorylates DSIF, NELF and CTD Ser2 for Pol II-pause release. CDK12 and CDK13 with their cofactor Cyclin K phosphorylates CTD Ser2 for productive elongation and, and CDK12/CycK acts on CTD Ser2 for recruitment of 3´end processing factors. Figure created with BioRender.com.

Members of the transcription-associated kinases phosphorylate the CTD of RNA Pol II in multiple steps during the transcription cycle, in terms of coordinating gene expression (figure 2) [59]. CDK8 is a subunit of the Mediator complex and help organize the PIC upon transcription initiation. On the other hand, CDK7 is a component of TFIIH, which is an essential in releasing Pol II from the PIC. Further transition into productive elongation, termination and 3´end processing demand the act of CDK9, CDK12 and CDK13 [59].

1.5.1 Cyclin-dependent kinase 8

CDK8 is a 464 amino acid protein that together with Cyclin C, MED12 and MED13 associate with the Mediator [28]. The Mediator complex is found to link the processes between RNA Pol II and transcription factors, promoters and enhancers [60]. It is prerequisite in most Pol II-generated transcripts in mammals as it binds to enhancers, recruiting Pol II to core promoters (figure 2). Supporting these aspects, the Mediator genome-wide co-localizes with CDK8, and CDK8 is found at enhancers as well [60].

As well as a role in positive transcription regulation, CDK8 is also involved in negative regulation. CDK8 is found to repress transcription initiation by sterically disrupt the interaction between Mediator and RNA Pol II [61].

1.5.2 Cyclin-dependent kinase 7

CDK7 is a 346 amino acid protein that together with Cyclin H and MAT1 form a CDK-activating kinase (CAK) [59]. During transcription, CDK7/CAK phosphorylates CTD Ser5 and Ser7 in terms of regulating transcription initiation and Pol II promoter escape (figure 2) [29, 32]. In addition, the CDK7/CAK complex is the catalytic subunit of transcription factor TFIIH. TFIIH phosphorylates CDK9/P-TEFb, indirectly activating promoter-proximal pause release of RNA Pol II [62]. CDK7/CAK-mediated CTD Ser5-p also aids the recruitment of capping enzymes to nascent mRNA [63]. This observation, in addition to CDK7 being involved in Pol II pausing, indicates a role of CDK7 in a checkpoint between Pol II pause and co-transcriptional mRNA capping [63].

1.5.3 Cyclin-dependent kinase 9

CDK9 is a 372 amino acid protein that binds to Cyclin T or Cyclin K (CycK) to gain its catalytic activity [64]. CDK9 and CycK together form the P-TEFb which is highly necessary in global regulation of gene transcription. After promoter-proximal pause, P-TEFb phosphorylates CTD Ser2, NELF and DSIF (figure 2), driving the RNA Pol II machinery into productive elongation [30]. CDK9 kinase activity is also found to be decisive in transcription termination, as it phosphorylates transcription termination factor Xrn2, enhancing the enzymatic activity of Xrn2 [65].

Apart from being required in Pol II pause release, CDK9 has during the last decade gained status as a multifunctional CTD-targeting kinase (figure 3). As mentioned, it is found to phosphorylate CTD Ser2 for Pol II release into productive elongation. Moreover, CDK9 is *in vivo* found to co-localize with Ser5 [66], and in chicken cells demonstrated to phosphorylate Thr4, which is required for histone mRNA 3´end processing [67].

1.5.4 Cyclin-dependent kinase 12

CDK12 is, compared to the other transcriptional CDKs, a large protein of 1,490 amino acids [68]. It connects with CycK to form an operative kinase complex. CDK12 comprise proline-rich motifs that are involved in protein-protein interactions, arginine/serine-rich motifs involved in pre-mRNA processing, a kinase domain and a carboxy-terminal domain that assists in its interaction with CycK [68, 69].

1.5.5 Cyclin-dependent kinase 13

CDK13 is the closest relative to CDK12 in the CDK family. It comprises 1,512 amino acids and share 43% sequence identity with CDK12 [68]. In contrast to its paralogue, it also contains a serine-rich domain and two alanine-rich domains, in which the functional relevance of these domains is to date not well understood [59]. Their kinase domains are 92% identical, and CDK13 also pairs up with CycK [59, 70]. CDK12 is the best studied of the two, while the function of CDK13 to date is more elusive.

Ever since the two paralogs CDK12 and CDK13 were found to be members of the CDK family, they have in several labs opened great research interest. They have during the past years emerged as kinases involved in CTD phosphorylation, as well as important characters during the transcription cycle. Apart from their structural similarities, research have revealed CDK12 and CDK13 to be functionally different.

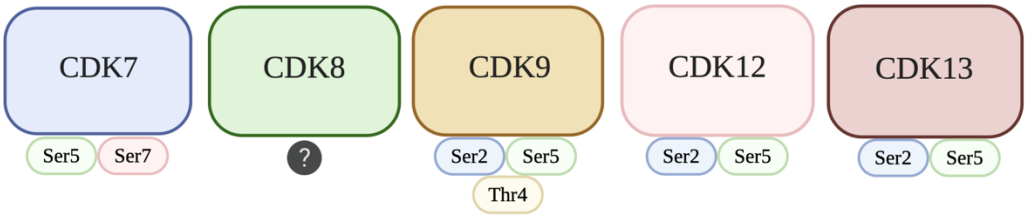


Figure 3. Summary of the transcriptional cyclin-dependent kinases and their target-residues of the mammalian CTD. CDK7 phosphorylates CTD Ser5 and Ser7 for Pol II to escape the promoter, as well as the Ser5 phosphorylation promotes from recruitment of capping enzymes to nascent mRNA. CDK9 phosphorylates CTD Ser2 for Pol II release into productive elongation, while its role in phosphorylation of Ser5 and Thr4 is poorly understood. CDK12 and CDK13 presumably phosphorylates Ser2, and possibly Ser5, although the purposes remain unclear. Figure created with BioRender.com.

1.6 Established data on CDK12 and CDK13

1.6.1 Early studies on CDK12 and CDK13 kinase activity

Early studies found CDK12 and CDK13 to phosphorylate Pol II CTD in Drosophila, yeast and insect cells. CDK12 was first found to possess phosphorylation activity against yeast CTD [71], before Bartkowiak et al. [72] found knockdown of CDK12 to reduce CTD Ser2 phosphorylation levels dramatically in Drosophila cells. Moreover, CDK12 was later found to phosphorylate CTD Ser2 and Ser5 at relatively similar levels in baculovirus-infected insect cells [68]. Baculovirus-infected insect cells was further used to study the kinase activity of

CDK13, which was found to phosphorylate CTD Ser2 and Ser5 [70]. What is known about their kinase activity in human cell lines will be discussed in the following subsections.

1.6.2 CDK12- and CDK13 kinase activity in human cells

Coinciding with their function, SR-4835-mediated dual inhibition of CDK12 and CDK13 lead to a slight decrease in bulk Ser2-p in triple-negative breast cancer cells [73]. THZ531-mediated CDK12/13 inhibition also exhibited a global decrease in Ser2-p in Neuroblastoma- and Jurkat cell lines [74]. On the other hand, CDK12 silencing in Jurkat cells did not show a global transcription defect, rather a gene-specific loss of Ser2-p [74, 75]. When it comes to the role of CDK12 and CDK13 in phosphorylating CTD Ser5, some studies have demonstrated both kinases to phosphorylate this residue of the CTD [70, 76], while others have not detected any correlation between the kinases and CTD Ser5 [74].

1.6.3 CDK12- and CDK13-mediated CTD phosphorylation during transcription

Phosphorylation of Ser2 is associated with productive transcriptional elongation and recruitment of 3´end processing factors [54]. Indeed, depletion of CDK12 and CDK13 have been shown to be involved in elongation and maturation of full-length mRNA. Zhang et al. [75] found THZ531-treatment to cause loss of elongation and gene expression in addition to impaired phosphorylation of Pol II CTD. Affirmatively, a recent study done on CDK12/13 depletion in Neuroblastoma cells lead to gene-length dependent elongation defects associated with early termination through premature cleavage and polyadenylation (PCPA) [74].

A role of CDK12 for productive elongation has therefore been suggested. However, the elongation defects appear to be gene specific. Blazek et al. [77] found depletion of CDK12/CycK to lead to decreased expression levels of long genes with a high number of exons, in which *BRCA1*, *ATR, FANCI* and *FANCD2* were among these genes – all known to take part in regulation of genomic stability. Eifler et al. [78] demonstrated that RNA interference (RNAi)-mediated depletion of CDK12 in HEK293 cells lead to a gene specific decrease in CTD Ser2-p in *c-MYC* and *c-FOS* genes. Moreover, Krajewska et al. [74] inhibited CDK12/13 with THZ531 and found induced PCPA of DDR genes. Thus, evidence suggests for a role of CDK12 in transcription elongation of a subset of protein-coding genes. However, other central features are assigned CDK12 and CDK13, and their roles in mRNA maturing events will be discussed below.

1.6.4 CDK12 and CDK13 in splicing

Beside their highly assumed roles as Pol II CTD kinases, CDK12 and CDK13 are found to be involved in splicing. Previous mass spectrometry (MS) studies presented data indicating the splicing machinery to be associated with CDK12 and CDK13 [59, 78]. Both kinases are found to interact with components of the spliceosome, specifically SRSFs (Serine/arginine-rich protein-specific kinases). In addition, they are both localized in nuclear speckles which is known to be sites for splicing factor storage [79, 80]. The RS domains in which both CDK12 and CDK13 are in possession of is a hallmark of splicing factors, and their additional common custody of CTD kinase domains is essential for a role in CTD-dependent co-transcriptional splicing [81, 82].

1.6.5 CDK12 in alternative polyadenylation

Moreover, CDK12 is assigned a role in regulation of APA. As previously described, APA gives rise to various transcript isoforms with differing 3´ends, and this fashion of gene expression has heavily impact on cell growth and development.

Dubbury et al. [83] showed that CDK12 suppresses intronic polyadenylation in DDR (DNA damage repair) gene regulation. DDR genes are found to harbor more intronic polyadenylation sites than other genes, in which these genes are more sensitive to loss-of-function CDK12 mutations. In agreement with this, CDK12 was in the same study found to globally suppress intronic polyadenylation in mouse embryonic stem cells [83], addressing a role of CDK12 in suppressing APA and consequently promoting full-length isoform expression, specifically in DDR genes [59].

1.6.6 CDK12 in 3´end formation

Moreover, CDK12 has been implied a role in 3´end processing. Eifler et al. [78] showed that including reduced levels of Ser2-p, CDK12 deficiency also lead to reduced levels of cleavage factor CstF64 on *c-FOS* genes, which in terms resulted in impaired 3´end processing. Moreover, MS data has demonstrated CDK12 to interact with 3´end processing factors as for instance CPSF1, Pcf11 and subunits of the RNA exosome complex [84].

Hence, CDK12 and CDK13 are both transcription-associated kinases evidently involved in multiple cellular processes. As transcription mis-regulation is contributor to a bunch of

cancers and other diseases [85, 86], CDK12 and CDK13 have sparked interest in research with respect to disease.

1.7 CDK12 and CDK13 in disease

1.7.1 CDK12 in disease

In consonance with the role of CDK12 in maintenance of genome stability, CDK12 loss-of-function is also key player in tumorigenesis. Joshi et al. [85] found that in ovarian cancer, four missense mutations in CDK12 is clustered in its protein kinase domain. They further found CDK12 depletion to abridge *BRCA1, FANCI* and *FANCD2* levels in ovarian cancer cells, and that CDK12 depletion lead to reduced formation of RAD51 foci formation and homologous repair (HR). Moreover, they proved CDK12 kinase domain mutations to abrogate its catalytic activity, in which these CDK12 mutants also disrupted HR activity [85].

CDK12 is also attributed a role in a genome-wide instability pattern found in cancer. Whole genome sequencing of human cancers has revealed a genome-wide configuration of repetitive DNA arrangements, in which the tandem duplicator phenotype (TDP) is one of the patterns described [59]. TDP is characterized by head-to-tail duplications of DNA segments, whereby specific types of tumors retain hundreds of these repeats [59].

Ng et al. [87] demonstrated that 12.8 % of high-grade serous ovarian carcinomas (HGSOC) harbored the TDP phenotype. CDK12 is known to be recurrently mutated in HGSOC and indeed, Popova et al. [88] found loss of CDK12 to be characterized by hundreds of tandem duplications in HGSOC and prostate cancers. In addition, Wu et al. [89] identified a novel type of metastatic castration-resistant prostate cancer, characterized by bi-allelic inactivation of CDK12 and consequently large number of tandem duplications. Presently, the mechanistic insight into how these CDK12-associated tandem duplications are generated remains elusive. These observations together lead to an assumption of CDK12 in role as tumor suppressor.

1.7.2 CDK13 in disease

CDK13 is likewise shown to be amplified in several human cancers, including hepatocellular carcinoma (HCC) and colorectal cancer. Kim et al. [86] screened a panel of 70 human cancer cell lines and found CDK13 to be amplified in 37.1% of them. In addition, they found loss of CDK13 to be significantly associated with early-onset of HCC, and amplification of CDK13

on the other side to be significantly associated with late-onset HCC. Herein, they labeled CDK13 as an oncogene with potent oncogenic activity [86].

Dong et al. [90] further analyzed a role of CDK13 in HCC. Two N-terminal residues of CDK13, Q103R and K96R, were mutated in respectively 60% and 65% in a total of 60 tumor samples. They found that these mutations were catalyzed by adenosine deaminases acting on RNA, in which adenosine (A) was converted to inosine (I), and inosine was subsequently recognized as a guanine (G) by the splicing- and translational machineries, leading to an amino acid substitution. The same A-to-I editing mechanism in CDK13 was also found in another study performed on A172 glioblastoma cells [91].

Recent evidence further points at CDK13 missense variants as the cause of congenital heart defects, facial dysmorphism and intellectual development disorder. This disorder, caused by missense mutations in the CDK13 kinase domain, is characterized by developmental delay, intellectual disability and a significant facial gestalt [92].

Taken together, proper expression of CDK12 and CDK13 appear to be essential to cell vitality. However, their exact roles are to date far from established. Research points towards a role of CDK12 in phosphorylating CTD Ser2, although evidence is inadequate. The role of CDK13 as a CTD kinase is even less established. Both kinases are nonetheless involved in maintenance of genomic stability and proper gene expression, and their genomic and proteomic disturbance is linked to oncogenesis and other diseases. Loss of CDK12 catalytic activity is directly linked to mis-procession of mRNA transcripts (mainly in DNA damage repair genes) due to a decreased recruitment of 3′end processing factors to nascent mRNA. It is however clear that to date, their full functions and activities remain elusive. Research remain to fulfill our understanding of their cellular functions.

2. Aim of study

In the present study, we wanted to elucidate the roles of chromatin-bound CDK12 and CDK13 expressed at endogenous levels. From technical causes, CDK13 was not taken into further analysis. We investigated the contribution of CDK12 in regulating aberrant Pol II, in terms of gaining a deeper understanding of CDK12 as a Pol II CTD kinase. This study focuses on CDK12 as CTD Ser2 and Ser5 kinase. We further wanted to identify the interactome of chromatin-bound CDK12 to acquire knowledge on its role in transcription and RNA processing events.

3. Materials and methods

3.1 Cell line and cultivation

3.1.1 Cell line

In this study, HeLa Kyoto (Riken, Japan) cells was used to investigate CDK12´s interaction partners and role in transcription. This cell line was the first human cell line, originally derived from Henrietta Lacks, a female patient who in 1951 died from an aggressive adenocarcinoma of the cervix [93].

3.1.2 Thawing

Cells were collected from a N_2- container and immediately thawed at 37°C for 5-10 minutes. The cell suspension was subsequently resuspended in 1mL Dulbecco´s Modified Eagle Medium (DMEM) - high glucose (Thermo Fisher Scientific), before transferred to a 15 mL tube, mixed in 7 mL DMEM and centrifuged in Megafuge 1.0 tabletop refrigerated centrifuge (Heraeus Instruments) at 1200*g* for 3 minutes. Supernatant was aspirated, before pellet was resuspended in 4mL DMEM and transferred to a 6-cm plate. Cells were incubated at 37°C in a humidified chamber containing 5% CO_2.

3.1.3 Cell cultivation

The HeLa cell line was cultured in Falcon® Cell Culture Dishes (Corning) of required size, and the cell line´s cultural needs was acquired in Dulbecco´s Modified Eagle Medium (DMEM) - high glucose, with a 10% concentration of fetal bovine serum ([FBS], Sigma-Aldrich®) and 1% Penicillin-Streptomycin (Thermo Fisher Scientific). Cells were cultivated at 37°C in a humidified chamber containing 5% CO_2. When reached a confluency of approximately 80%, cells were split at a ratio of 1/8.

3.1.4 Cell splitting

Adherent cell cultures have to be split in order to preserve cell viability, and to maintain exponential log-phase growth.

Cell culture medium (DMEM) was aspirated from the culture dish before cells were washed in 1x phosphate-buffered saline ([PBS], Oslo Universitetssykehus, avdeling for mikrobiologi). PBS was aspirated and substituted with Trypsin-EDTA (Thermo Fisher

Scientific) for complete detachment of cells. Cells incubated in Trypsin-EDTA for 15 minutes as room temperature (RT). When completely detached, cells were supplemented with DMEM to inhibit further enzymatic activity of Trypsin-EDTA. Cells were subsequently centrifuged, resuspended in culture medium and split at a ratio of 1/8.

3.2 Transfection and single-cell sorting

3.2.1 Lipid-mediated transfection of CDK9/12/13-GFP bacterial artificial chromosome constructs

HeLa cells were in this study transfected with recombinant bacterial artificial chromosome (BAC) CDK9-, CDK12- and CDK13-GFP (green fluorescent protein), in terms of investigating the functional properties of these epitope-tagged proteins. Transfection was carried out in a liposome-based approach, which allows for lipid/DNA complexes to fuse with cell membrane and subsequent delivery of genetic material to nucleus [94]. The theory behind will be described in detail in chapter 4, subsection 4.1.3. Two versions of the HeLa CDK12-GFP transfected cell line were generated: One expressing N-terminally tagged GFP, and the other expressing C-terminally tagged GFP. Both CDK9 and CDK13 were C-terminally tagged.

$8.0x10^5$ HeLa cells were counted with Trypan Blue Stain 0.4% (Thermo Fisher Scientific) in Countess™ Cell Counting Chamber Slides (Thermo Fisher Scientific) by the use of Countless™ II FL Automated Cell Counter (Thermo Fisher Scientific) and seeded in 6 cm-plates. The cells were incubated over night at 37°C in 5% CO_2. Lipofectamine mix was prepared in two separate Eppendorf tubes; One with 2 µg bacterial artificial chromosome (BAC) DNA (Emd Millipore) in 200 µL Opti-MEM (Thermo Fisher Scientific), and one with 6 µL of Lipofectamine 2000 (Thermo Fisher Scientific) in 200 µL of Opti-MEM. 200 µL from each tube was mixed into a final volume of 400 µL and incubated at room temperature for 25 minutes to allow for construction of Lipofectamine/BAC DNA complexes (figure 4). The HeLa cells in each 6-cm dish was subsequently transfected with 400 µL Lipofectamine mix. Cell culture medium was changed after 6-8 hours. The next day, cells were passed into a 15-cm dish containing 20 mL cell culture medium supplemented with 200 µg/ml Geneticin (Thermo Fisher Scientific), for selection of cells stably expressing the recombinant BAC-DNA.

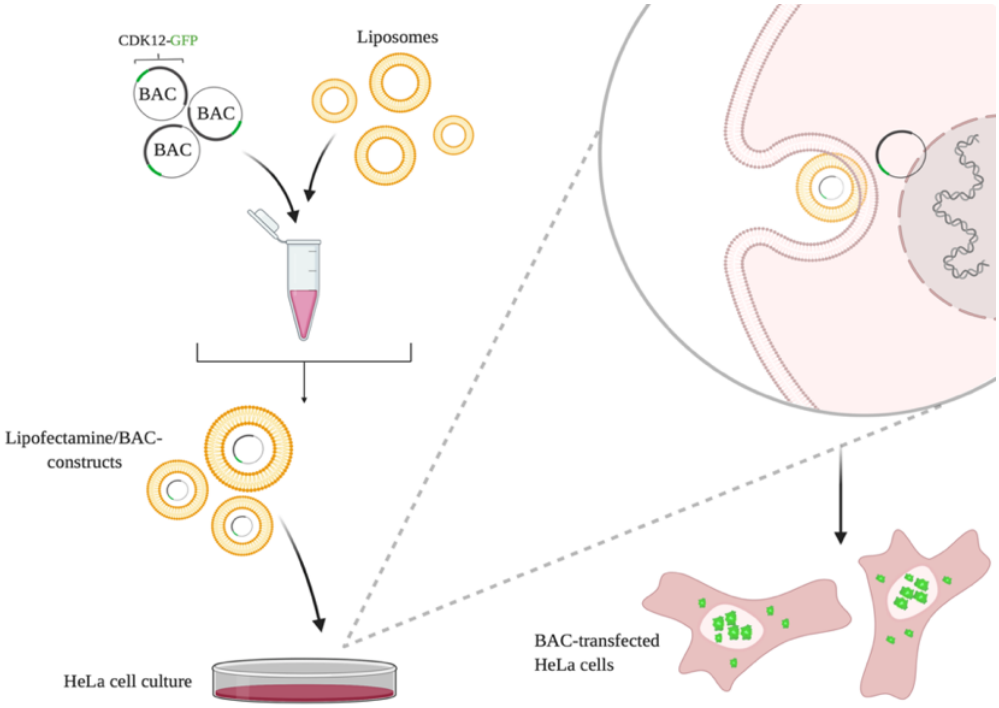


Figure 4. Simplified, schematic drawing of lipid-mediated BAC transfection. HeLa cells were incubated with Lipofectamine/BAC-DNA mix, which allowed for transfection of BAC-transgene complexes and subsequent expression of recombinant GFP-tagged protein in HeLa cells. Figure created with BioRender.com.

3.2.2 Fluorescence-activated cell sorting

Fluorescent-activated cell sorting (FACS) provides effective isolation of GFP-transfected cells from a heterogenous cell mixture [95]. The method utilizes the fluorescent characteristics of successfully GFP-transfected cells. It also provides single-cell sorting and subsequent growth of homogenous GFP-positive (GFP+) cell colonies in 96-well plates [95]. FACS was in this study utilized to separate GFP+ cells from untransfected HeLa cells, as well as to sort single-cell GFP+ HeLa cells in 96-well plates for GFP+ colonies to grow.

Adherent transfected and untransfected HeLa cells cultured in 15-cm dishes were washed in 1x PBS, trypsinized in 1.5 mL Trypsin-EDTA and centrifuged at 1200*g* for 5 minutes at RT. For control, untransfected HeLa cells were resuspended in 1 mL culture medium, in which 200 µL was transferred to Corning™ Falcon™ Test Tubes compatible for FACS. Transfected cells were resuspended in 500 µL culture medium and transferred to Corning™ Falcon™ Test Tubes. Tubes were handed to the Flow Cytometry Core Facility at Oslo University Hospital, Radiumhospitalet, for sorting.

3.3 Treatments

3.3.1 Heat shock treatment

Heat shock treatment changes global occupancy of RNA Pol II at mRNA genes [96]. Cells were treated with heat shock in order to investigate the contribution of CDK12 in regulating aberrant Pol II. The theory behind will be described in detail in chapter 4, subsection 4.1.1.

HeLa cell culture dishes were incubated at 42°C in humidified chamber containing 5% CO_2 for 10 or 30 minutes.

3.3.2 THZ531 treatment

THZ531 inhibits CDK12/13 catalytic activity [75]. The mechanism behind will be described in detail in chapter 4, subsection 4.1.2.

HeLa cells were treated with 3μM THZ531 (MedChemExpress) diluted in culture medium and incubated at 37ºC for 6 hours. 0.0003% dimethyl sulfoxide ([DMSO], Sigma-Aldrich®) in culture medium was used as control.

3.4 Cell lysis, protein isolation- and quantification

3.4.1 Cell lysis

Adherent HeLa cell cultures were washed in cold PBS, scraped off the plate and collected in cold 1.5 mL Eppendorf tubes. Cell suspension was resuspended before the tubes were centrifuged in Microcentrifuge 5424 R (Eppendorf®) at 1000*g* for 5 minutes at RT, and excessive PBS was decanted. Cells were lysed in High Salt (HS) buffer (Section B, Appendix) supplemented with protease inhibitors (1 μg/ml aprotinin ([AP] Sigma-Aldrich®) 1 μg/ml leupeptin ([LP], Sigma-Aldrich®), 0.1μM phenylmethanesulfonyl ([PMSF], Sigma-Aldrich®)) and incubated on ice for 5-10 minutes to allow for complete lysis of cells. Cell lysate was transferred to new cold Eppendorf tubes and sonicated in Bioruptor® Pico sonication device (Diagenode) 30s ON/OFF for 6 cycles at 4°C. Lysate was subsequently centrifuged at 20,000*g* for 20 minutes at 4°C in a microcentrifuge, and supernatant was transferred to a new cold 1.5 mL tube.

3.4.2 Cellular fractionation

Cellar fractionation is a technique used to separate cellular components to yield cytosolic, nuclear and chromatin bound proteins, in order to determine cellular localization of protein of interest [97]. Cellular fractionation was in this study performed to localize CDK12 in the cell.

3x15cm dishes of HeLa cells were washed in 1.5 mL cold PBS, scraped and collected in cold 15 mL tubes. The tubes were centrifuged at 400*g* for 5 minutes at 4°C. Cell volume was determined, resuspended gently in 5 volumes of cold hypotonic lysis buffer (Buffer A, Section B, Appendix) supplemented with protease inhibitors and allowed to swell on ice for 10 minutes. Cells were subsequently centrifuged at 400*g* for 5 minutes at 4°C and supernatant decanted. Due to osmotic swelling, cell volume was again determined, and cells were gently resuspended in 2 volumes of cold Buffer A supplemented with 0.5% IGEPAL CA-630 (Sigma-Aldrich®). Cells were transferred to a cold glass Kontes Dounce Homogenizer (Sigma-Aldrich®), and a pestle was carefully used in a total of ten up-and-down strokes. Lysate was subsequently transferred to cold Eppendorf tubes and centrifuged at 3300*g* for 15 minutes at 4°C. Supernatant, which made up the cytoplasmic fraction, was transferred to separate Eppendorf tubes, and the salt concentration was adjusted to 150 mM by addition of NaCl to the fractions. To remove residual amounts of cytoplasmic proteins, the residing pellet was washed in cold PBS and centrifuged for at 3300*g* for 5 minutes at 4°C.

Nuclear pellet was resuspended in 300 µL cold Buffer B supplemented with protease inhibitors (Section B, Appendix) and left on ice for 10 minutes. Samples were sonicated 20s on/off for 6 cycles. Next, 700 µL of cold Buffer B was added to sonicated sample to a total of 1 mL. Samples were tumbled for 45 minutes at 4°C, and subsequently centrifuged at 15,000*g* for 30 minutes at 4°C. Supernatant, which now made up the nuclear soluble (NS), was transferred to new, cold Eppendorf tubes.

1 mL of Buffer B supplemented with 12 µL Benzonase (Merck Millipore) was added to the remaining pellet, and tumbled over night at 4°C. This solution constituted of the nuclear pellet (NP). The protein yield of cytoplasmic fraction (CS), nuclear soluble (NS) and nuclear pellet (NP) was measured by Bradford Protein Assay (see subsection 3.4.4) before snap-frozen in N_2- container and stored at -80°C.

3.4.3 Co-immunoprecipitation

Co-immunoprecipitation is a biochemical method used to isolate proteins of interest from cell lysate, in interest of investigating the proteins binding the immunoprecipitated protein. The method takes advantage of antibody-recognition of target molecule [98]. GFP Trap Magnetic Agarose affinity beads (Chromotek) was in this study utilized to purify GFP-CDK12 from whole cell HeLa lysate.

Whole cell lysate was first thawed on ice. 60 µL beads GFP Trap Magnetic Agarose beads was transferred to a 1.5 mL Eppendorf tube and resuspended in 1 mL PBS, before centrifuged at 1200*g* for 1 minute. This washing step was repeated two times, before PBS was discarded. 1 mL HS buffer supplemented with protease inhibitors (Section B, Appendix) and 1 mg lysate (calculated by Bradford Protein Assay, subsection 3.4.4) was subsequently added to the beads, and tumbled at 4°C ON.

The day after, beads were washed in 1 mL ice cold HS buffer and centrifuged at 1200*g* for 1 minute at 4°C. HS buffer was decanted, and beads were subsequently washed x2 in 1 mL cold Washing buffer supplemented protease inhibitors (Section B, Appendix), before centrifuged at 1200*g* for 1 minute at 4°C. Washing buffer was discarded. Beads were washed x1 in cold PBS and again centrifuged at 1200*g* for 1 minute at 4°C. Bead-bound protein was eluted by incubation at 95°C, in 100 µL loading sample buffer ([LSB], (Section B, Appendix)) supplemented with 1 µL 2-Mercaptoetanol (Thermo Fisher Scientific) and 100 mM Dithiotreitol ([DTT], Saveen Werner AB), for 5 minutes. The Eppendorf tube was subsequently centrifuged at 1200*g* for 1 minute at RT, and eluate was transferred to a new Eppendorf tube.

3.4.4 Measurement of protein concentration

Protein concentration was measured with Bradford Protein Assay. 800 µL MQ-H_2O was in disposable cuvettes (Bio-Rad) supplemented with 200 µL Protein Assay Dye Reagent Concentrate (Bio-Rad) and 2 µL protein sample. As blank control, protein sample was replaced with buffer used in protein samples. Solutions were vortexed thoroughly. Absorbance was measured at 595 nm using a Smartspec Plus Spectrophotometer (Bio-Rad). Protein concentration was further calculated from absorbance values based on standard curve 6.2783*X+2.0083, as reference for calculation of unknown protein concentration.

3.5 Western blot

Western blotting is a technique in cell and molecular biology that is used to detect specific proteins in complex cell homogenates. Proteins are, based on molecular weight, separated by electrophoresis, transferred to a membrane and incubated with antibodies specific to the target protein [99].

Four rounds of blotting procedures were in this study executed. The first round of western blot experiments was performed to detect CDK9-, CDK12- and CDK13-GFP in HeLa cells. The second round was conducted to ascertain successful IP of CDK9-, CDK12- and CDK13-GFP. The third round was executed to localize endogenous CDK12 in the cells, while the last round was performed in order to validate results from MS analysis.

3.5.1 Sample preparation

HeLa cell lysates were thawed on ice, before diluted in MQ-H_2O and LSB to obtain equal protein concentration in all samples. Prepared samples contained 75% diluted protein and 25% LSB. 2 µL DTT was added to each sample to reduce disulfide bonds of proteins, and samples were heated at 5°C for 5 minutes for truly denaturation of protein tertiary structure, before loaded on the gel.

3.5.2 Protein separation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Protein extracts were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples were loaded on 6- or 8% NuPAGE® Novex Tris-Glycine gels (Thermo Fisher Scientific) or 10% Bolt™ Bis-Tris Plus gels (Thermo Fisher Scientific), respectively with 1X Tris-Glycine SDS Running Buffer (Thermo Fisher Scientific) or 1X NuPage® MOPS SDS Running Buffer (Thermo Fisher Scientific), both diluted in dH_2O. All gels were run in XCell SureLock Mini-Cell chamber (Thermo Fisher Scientific). Precision Plus Protein™ Dual Color Ladder (Bio-Rad) was added to the first well, and protein samples were loaded in the following wells. Proteins were separated at 200 Volt (V) for 30-45 minutes.

3.5.3 Blotting

Protein gel was subsequently transferred to Trans-Blot® Turbo™ Mini PVDF Membrane (Bio-Rad) and sandwiched in blot paper, before proteins were transmitted from gel to membrane in a Trans-Blot® Turbo™ Transfer System (Bio-Rad) at 25V for 10 minutes.

The protein-containing membrane was subsequently stained in Ponceau S solution (Sigma-Aldrich®) for 5 minutes at RT to assess for protein transformation, before briefly de-stained in TBS-T (Section B, Appendix) and incubated in blocking solution (5% skim milk powder (Sigma-Aldrich®) in TBS-T), for 1 hour at RT. Membrane was subsequently washed x1 in TBS-T containing 1% skim milk powder. Primary antibodies (table 1) were diluted in blocking solution, and membrane was subsequently incubated in primary antibody-solution on a shaking platform at 4°C ON. Membrane was relocated in TBS-T + 1% skim milk powder and washed 3x10 minutes. Secondary antibodies (table 2) were diluted in blocking solution. Membrane was incubated in secondary antibody solution on shaking platform for 45 minutes at RT, before again washed 3x10 minutes in TBS-T + 1% skim milk powder.

Table 1. Primary antibodies used in western blotting.

Antibody	Host species	Dilution	Producer	Catalog nr.
mAb CDK12	Rabbit	1:1000	Cell Signaling Technology, USA	11973S
mAb CDK9	Rabbit	1:1000	Cell Signaling Technology, USA	C127F
pAb Anti-GFP	Rabbit	1:50,000	Abcam, UK	ab6556
mAb Suz12	Rabbit	1:20,000	Cell Signaling Technology, USA	3737S
mAb Anti-Vinculin	Mouse	1:1000	Sigma-Aldrich®, Germany	V9131
pAb Anti-Histone H3	Rabbit	1:1000	Abcam, UK	ab9050
GAPDH	Rabbit	1:15,000	Santa Cruz Biotechnology, USA	sc-25778

Abbreviations: mAb = monoclonal antibody, pAb = polyclonal antibody, Suz12 = Polycomb repressive complex 2 subunit, GAPDH = glyceraldehyde 3-phosphate dehydrogenase.

Table 2. HRP (horseradish peroxidase)-conjugated secondary antibodies.

Target protein	Dilution	Producer	Catalog nr.
Anti-Rabbit IgG, HRP conjugate	1:5000	Sigma-Aldrich®, Darmstadt, Germany	12-348
Anti-Mouse IgG, HRP conjugate	1:5000	Sigma-Aldrich®, Darmstadt, Germany	A9917

Abbreviations: IgG = immunoglobulin G, HRP = horseradish peroxidase.

Detection of proteins on the membrane was conducted by SuperSignal™ West Pico PLUS Chemiluminiscent Substrate (Thermo Fisher Scientific). Membrane was incubated in these reagents before visualized in Gel Doc XR+ Documentation System (Bio-Rad).

3.6 Quantitative reverse transcription polymerase chain reaction analysis

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) is a method that can be utilized to measure levels of expressed genes at RNA levels [100]. The technique first transcribes RNA into complementary DNA (cDNA), before specific DNA targets are amplified in a PCR reaction. RT-qPCR takes advantage of fluorescent reporter molecules, in which the fluorescence is measured during each PCR cycle. The fluorescence intensity increases proportionally with amplified DNA. A fluorescent signal detected above the computer-chosen threshold is considered as valid. This data indicate information about starting levels of expressed RNA sequence of interest [100]. The technique was in this study assessed to determine levels of expressed recombinant GFP mRNA in BAC transfected HeLa cells.

3.6.1 RNA purification and cDNA construction

Total RNA was isolated by the use of RNeasy Plus Mini Kit (QIAGEN). RNA concentration was measured by NanoDrop™ Spectrophotometer (Thermo Fisher Scientific). 1 µg of purified RNA was diluted in a total of 10 µL RNAse Free H_2O (QIAGEN) and, in a total of 20 µL, mixed with reagents retrieved from High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) listed in table 4, before reverse transcribed to cDNA in 2720 Thermal Cycler (Thermo Fisher Scientific), with the following program:

Table 3. cDNA reverse transcription reaction conditions.

Temperature profile	Time profile
25 °C	10 minutes
37 °C	2 hours
85 °C	5 minutes
4 °C	Forever

Table 4. Reagents and volume per reaction in synthesis of cDNA.

Reagent	Volume/reaction
10X RT buffer	2 µL
25X dNTP mix	0.8 µL
10X Random Primers	2 µL
Reverse Transcriptase	1 µL
Nuclease free H_2O	4.2 µL
Total	10 µL

Abbreviations: RT buffer = Reverse Transcription buffer, dNTP = deoxynucleotide triphosphate.

3.6.2 RT-qPCR

To create the quantification of *GFP* RNA, a mixture of RT-qPCR reagents and cDNA (table 6) was pipetted in duplicate in a MicroAmp® Fast 96-well Reaction Plate (Thermo Fischer Scientific). The plate was sealed with MicroAmp™ Optical Adhesive Film (Thermo Fischer Scientific), centrifuged briefly in Microplate Centrifuge (VWR International) to eliminate excess air in sample, before the RT-qPCR reaction was carried out in a StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific) with reaction conditions listed in table 7.

Table 5. Primer sequences used for RT-qPCR.

Gene	Forward/Reverse	Sequence
RPLP0	Forward	TCCATTGTGGGAGAGCAGAC
	Reverse	CAGCAGTTTCTCCCAGACG
GFP	Forward	AGCAAAGACCCCAACGAGAA
	Reverse	TCGTCCATGCCGAGAGTGAT

Abbreviations: *RPLP0* = 60S acidic ribosomal protein P0, GFP = green fluorescent protein.

Table 6. RT-qPCR mixture reagents and volume per reaction.

PCR reaction reagent	Volume/reaction
PowerUp™ SYBR™ Green Master Mix (Thermo Fischer Scientific)	10 µL
RNAse Free H_2O (QIAGEN)	6 µL
Forward primer	2 µL
Reverse primer	2 µL
Sample cDNA	2 µL
Total	20 µL

Table 7. RT-qPCR reaction conditions

Temperature profile	Time profile
95 °C	10 min
95 °C	15 sec x 40
60 °C	25 sec
95 °C	15 sec
60 °C	1 min
95 °C	15 sec

3.6.3 Statistical analysis

Each sample was normalized to RPLP0 as housekeeping gene (see primer sequence in table 5). RT-qPCR data analysis was carried out in Microsoft Excel, and relative quantification method ($\Delta\Delta C_T$) was used to calculate the relative increase in gene expression of gene of interest.

3.7 Live cell imaging

In order to localize CDK12-GFP in the cells, as well as to study potential changes in protein dynamics upon THZ531- or heat shock treatment, we performed live imaging. Live cell imaging is a time-lapse based technique that is used to study the localization of proteins emitting fluorescence in cells [101].

10.000 HeLa cells (ctrl) and HeLa CDK12-GFP transfected cells were seeded in separate wells of a CytoSoft® Imaging 96-well Plate (Advanced BioMatrix). Cells were either treated in 0.0003% DMSO or 3µM THZ531 for 4 hours, or 10 minutes of heat shock.

Images were acquired on ImageXpress Micro Confocal (Molecular Devices**)** and processed by the help of the Fiji software [102].

3.8 Immunofluorescent staining

Immunofluorescent staining is a technique that is used to detect specific proteins within fixed cells or tissues by the use of fluorescently labeled antibodies [103]. The method involves image analysis, which allows for characterization of protein localization and distribution. The technique was in this study performed to gain insight into localization and phosphorylation state of RNA Pol II-phosphorylated CTD species.

3.8.1 Cell seeding and treatment

50.000 Hela cells were counted, and in 24-well plates seeded on 13mm cover slips (VWR International) in a total of 500 µL DMEM. 24 hours later, the cells were treated with 3µM THZ531 or 0.0003% DMSO for 6 hours, before given a heat shock at 42ºC for 10 minutes. 0.0003% DMSO solvent without THZ531 was used as control.

3.8.2 Fixation, blocking and primary antibody staining

Prior to antibody incubation, the cells were washed in cold 1x PBS and fixed in ice cold 3.7% formaldehyde (Section B, Appendix) for 10 minutes at RT. Cells were subsequently washed twice in cold PBS before permeabilized in PBS-T (0.2% Triton-X100 (Sigma-Aldrich®) in 1xPBS) for 15 minutes at RT and washed again. Cells were subsequently incubated in blocking solution, consisting of 0.5% bovine serum albumin ([BSA], Saveen & Werner) in PBS-T, for 1 hour at RT to block nonspecific binding. Primary antibody (listed in table 8) diluted in blocking solution was subsequently applied to cells and incubated for 1 hour at RT.

Table 8. Primary antibodies used in immunofluorescent staining

Primary antibody	Species	Dilution	Producer	Catalog nr.
Anti-RNA Polymerase II CTD	Mouse	1:1000	MBL Life science	MABI0601S
pAb Anti-GFP	Rabbit	1:5000	Abcam, UK	ab6556
Anti-Phospho RNA Polymerase II CTD (Ser2)	Mouse	1:1000	MBL Life science	MABI0602S
Anti-Phospho RNA Polymerase CTD (Ser5)	Mouse	1:1000	MBL Life science	MABI0603S

3.8.3 Secondary antibody staining and mounting

Cells were washed twice in cold PBS, before incubated in corresponding secondary antibody (listed in table 9) diluted in blocking solution for 45 minutes at RT. Cells were subsequently washed twice in PBS-T and once in PBS. As a reference for cell nuclei, the cover glasses were mounted in 12 µL Vectashield HardSet Antifade Mounting Medium with Dapi (Vector Laboratories) on Superfrost™ Microscope Slides (Thermo Fisher Scientific) and stored at 4°C in dark until further use.

Table 9. Fluorescent secondary antibodies used in immunofluorescent staining

Secondary antibody	Dilution	Producer	Catalog nr.
Goat Anti-Rabbit Alexa Fluor 555	1:1000	Thermo Fisher Scientific, USA	A27039
Goat Anti-Mouse Alexa Fluor 555	1:1000	Thermo Fisher Scientific, USA	A28180

3.8.4 Imaging and image-processing

Images were acquired on Leica TCS SP8 gSTED microscopy (Leica Microsystems) using a 40x oil immersion objective, equipped with Leica LASX software. All cells with the same antibody staining were imaged with the exact same acquisition parameters, and images were subsequently processed with equal brightness/contrast levels in the Fiji software.

3.9 Chromatin enrichment for proteomics

Chromatin enrichment for proteomics (ChEP) was performed to analyze chromatin on a proteomic scale. This protocol was developed by Kustatscher et al. [104]. The method enables isolation of transiently chromatin-bound factors and allows for subsequent protein analysis by MS. The protocol will be annotated as adjustments were made to the initial protocol.

3.9.1 Cell fixation and lysis

3.7% formaldehyde solution was added to 15 cm plates of HeLa cells cultures and incubated at 37ºC for 10 minutes. Cells were subsequently incubated in final concentration of 125mM glycine (Section B, Appendix) for 5 minutes at RT, before washed in PBS. 1mL cold PBS was added to each plate, and cells were detached by scraping. Cell suspension was resuspended and transferred to a 15 mL Falcon, before centrifuged at 1200*g* for 5 minutes at RT.

Cell pellet was subsequently mixed on ice in lysis buffer supplemented with protease inhibitors (Section B, Appendix), transferred to 2 mL Eppendorf tubes and carefully resuspended to homogenize suspension. Lysate was centrifuged at 2300*g* for 5 min at 4ºC. Supernatant (cytoplasmic fraction) was transferred to a new tube and stored at -80ºC until use.

3.9.2 Purification of cross-linked chromatin

For lysis of nuclei, pellet was carefully resuspended in 500µL SDS-buffer supplemented with protease inhibitors (Section B, Appendix) and incubated for 10 minutes at RT. 1.5 mL UREA-buffer complemented with protease inhibitors (Section B, Appendix) was added to suspension and tubes were inverted several times, before centrifuged at 16,100*g* for 30 minutes at RT. Supernatant was discarded. To wash out contaminants, pellet was washed in 500µL SDS-buffer. Next, 1.5 mL UREA-buffer was added, before centrifuged at 16,100*g* for 25 min at RT. Supernatant was discarded. To wash out the UREA-buffer, pellet was resuspended in 500µL SDS-buffer, before supplemented with another 1.5 mL SDS-buffer. Suspension was centrifuged at 16,100*g* for 25 minutes at RT.

Pellet was subsequently covered in 0.5 mL of storage buffer supplemented with protease inhibitors (Section B, Appendix), before tube was flicked until pellet dislodged from the bottom. Tube was thereafter kept on ice. Solution was subsequently sonicated in intervals of 30s on/off for 15 minutes at 4ºC. Supernatant, that should now supposedly contain sheared cross-linked chromatin, was transferred to a new tube. Bradford protein assay was performed to determine protein yield.

3.9.3 Reverse cross-linking, SDS-PAGE and Coomassie Blue staining

To verify the presence of chromatin-bound proteins in sample, SDS-PAGE followed by Coomassie Blue gel staining was conducted. For reverse formation of cross-links, 4x LSB was added to the chromatin-enriched sample to a final concentration of 1x, before mixture was incubated at 98ºC for 30 minutes. Tube was subsequently centrigued at 16,100*g* for 15s at RT. Further, an aliquot of the cytoplasmic fraction was likewise supplemented with 4x loading buffer and boiled at 98ºC, before both cytoplasmic fraction and chromatin-bound protein enriched samples were loaded to a 4-15% Novex™ Wedgewell™ Tris-Glycine mini protein gel (Thermo Fischer Scientific) with 1x Tris-Glycine SDS Running Buffer (Thermo

Fisher Scientific) in dH_2O. Electricity was sat to 120V, and bands were visualized by Coomassie Brilliant Blue (Sigma-Aldrich®).

3.10 Chromatin immunoprecipitation-mass spectrometry

Chromatin immunoprecipitation-mass spectrometry (ChIP-MS) is a technique used to identify chromatin-bound protein complexes associated with surrounding proteins [105]. Cells are cross-linked before chromatin and associated proteins are purified (ChIP), and proteins subsequently are identified by MS. The RIME protocol [105] was followed with certain adjustments. The method was in this study performed to identify CDK12-GFP-bound proteins, to gain insight into the role of CDK12 in transcription.

3.10.1 Cell fixation and lysis

HeLa cells were cultured in 245 x 245cm Square BioAssay Dishes (Thermo Fisher Scientific). Total amount of culture medium was measured, before a final concentration of 1% formaldehyde solution (Section B, Appendix) was added to the dish. Dishes were incubated at 37ºC for 5 minutes, then shaked for formaldehyde solution to circulate, and incubated at 37ºC for 5 more minutes. To quench the formaldehyde and terminate cross-linking reaction, a final concentration of 125mM glycine was added to the culture medium and incubated at 37ºC for 5 minutes. Cell culture medium was subsequently poured off, and cells were washed twice in 1x PBS. Plates were vertically placed on bench for excess PBS to accumulate, before PBS was aspirated. Cells were subsequently lysed in 7 mL ChIP-SDS lysis buffer supplemented with protease inhibitors (Section B, Appendix), collected with a cell scraper in 50 mL tubes and placed on ice. Tubes were stored at -80ºC until further use.

3.10.2 Sonication and chromatin enrichment

Tubes were thawed in water bath at 37ºC until solutions appeared to be transparent. Tubes were subsequently centrifuged for at 1200*g* for 10 minutes at RT, before put on ice, and supernatant discarded. The pellet was resuspended in 10 mL of Buffer 1 (Section B, Appendix) and rotated on HulaMixer™ Sample Mixer (Thermo Fisher Scientific) for 5 minutes at 4°C. Lysate was cleared by centrifugation at 2,000*g* for 5 minutes at 4°C, and supernatant decanted. Cell nuclei was suspended in 300 µL of Buffer 2 supplemented with protease inhibitors (Section B, Appendix) and transferred to 1.5 mL Bioruptor® microtubes.

Chromatin was then fragmented by sonication at 30s on/off for 10 minutes. 30 µL Triton-X100 was added to sonicated lysate. To separate non-fragmented material (high molecular weight) from fragmented chromatin, the tubes were centrifuged at 20,000*g* for 20 minutes at 4°C. Sonication efficiency was validated by agarose gel analysis of the DNA.

3.10.3 Reverse cross-linking, DNA purification and agarose gel electrophoresis

Before analysis of sheared chromatin, lysate was de-crosslinked. 30 µL lysate and 90 µL de-crosslinking buffer (Section B, Appendix) was added to an Eppendorf tube, and incubated ON at 65°C on a Grant Bio PCMT Thermoskaker (Keison Products).

The DNA in 100 µL of de-crosslinked sample was extracted by QIAquick PCR Purification Kit (QIAGEN). A 1.2% agarose gel was subsequently prepared by the addition of 0.365 g agarose powder (Thermo Fisher Scientific) to 30 mL 1x TAE electrophoresis buffer (Appendix). The solution was stirred and boiled in microwave until dissolved. 0.25 µL SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) was added to agarose solution for DNA to be visualized before agarose solution was solidified in suitable chamber (Thermo Fisher Scientific). 1x TAE buffer was subsequently added to chamber. 10 µL of sample was supplemented with 2 µL of 1x LSB, before 10 µL GeneRuler DNA Ladder Mix (Thermo Fisher Scientific) was loaded to the first well, and 15 µL of the protein samples were loaded in the following wells. The gel was run at 120V for 20-25 minutes.

3.10.4 Bead preparation and co-immunoprecipitation

The protein yield of samples from subsection 3.10.2 was measured by Bradford Protein Assay, and diluted by adding Buffer 2 to a final concentration of 2.5 mg/mL.

Before co-immunoprecipitation, a stock solution of 80 µL GFP-Trap Magnetic Agarose affinity beads was transferred to a 1.5 µL tube and put on pre-cooled magnetic rack. Beads were first resuspended x1 in blocking buffer (5 mg/mL BSA in PBS), before washed x3 in 1 mL of Buffer 2, and washed beads were subsequently incubated with 1.2 mL diluted lysate on HulaMixer™ Sample Mixer at 4°C over night. The next day, the beads were washed in a total of 10 times in RIPA buffer supplemented with protease inhibitors (Section B, Appendix) before handed to LC-MS/MS at the Proteomic Core Facility at Oslo University Hospital, Rikshospitalet, Institute of Immunology. The experiment was performed on an Easy nLC1000

nano-LC system connected to a quadrupole-Orbitrap (QExactive Plus) mass spectrometer (Thermo Fisher Scientific), with EasySpray column (C18, 2 µm beads, 100Å, 75 µm inner diameter) and solvent gradient as 2% to 30% in 60 minutes.

Analysis of MS data was performed by the Proteomic Core Facility in the MaxQuant software, version 1.6.1.2. Minimal unique peptides were set to 1, and false discovery rate allowed was 1% for protein and peptide identification.

3.11 Analysis of mass spectrometry data

3.11.1 Venn diagram analysis

Venn diagram was provided by the use of Venny 2.1.0 software [106].

3.11.2 Gene list annotation and analysis

The Metascape software [107] was used to interpret gene list provided from MS analysis. For functional annotation, enrichment analysis was performed. This is a tool comprising gene annotation portals to sort and visualize significantly enriched clusters. Terms with a p-value < 0.05, a minimal overlap of 3, and an enrichment factor > 1.5 were, based on their similarities, clustered.

3.11.3 Network analyses

Proteomic analysis was first conducted in Metascape software for sorting of proteins into functional groups, before visualized in Cytoscape software [108]. In order to gain a holistic understanding of the functional groups, Enrichr [109] was used. Moreover, the STRING protein network tool [110] was used to determine interactions between proteins.

3.12 Statistical analysis

All statistical analyses were performed in Microsoft® Excel software. Data was analyzed, and significant differences investigated using Student´s t-test, 2-tailed, equal variance,
* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

4. Results

4.1. Experimental setup

4.1.1 Heat shock treatment changes global occupancy of RNA Pol II at mRNA genes

The heat shock response (HSR) is composed of evolutionary conserved mechanisms found to be critical for cellular survival. During exposure to elevated temperatures, cells possess the ability to temporarily adapt to abnormal environments due to re-programming of the transcriptional and translational metabolic pathways, including alterations of cellular substructures [111].

An immediate response of heat shock is found to be at the transcriptional level. Recent studies revealed that upon heat shock, the expression levels of hundreds of genes are upregulated, as for instance heat shock proteins - while thousands of other genes are downregulated [112]. Thus, upon heat shock a rapid transcriptional reprogramming occur, causing an immediate transcriptional change.

The HSR is arranged by re-orchestration of gene expression and distal regulatory elements. In response to heat shock, the transcription levels of heat shock transcription factors (HSFs) and heat shock proteins (Hsps) are enhanced [111]. HSFs promote for stress-dependent transcription activation of Hsps. Under a heat shock response, one of four HSFs, HSF1, binds to heat shock regulatory elements (HSEs) located at hundreds of genome promoter regions. This binding promotes for recruitment of co-factors that enhance the expression of specific genes by increasing the release of paused RNA Pol II into productive elongation. However, HSF1 controls only a portion of heat shock induced genes, and the entirety remains to be fully understood [111].

Accompanying the events of transcription activation, thousands of other genes are repressed. A natural consideration for this event would be RNA Pol II, as it under normal conditions traverse over active gene bodies and is regulated in a CTD-dependent manner. Indeed, defective behavior of RNA Pol II upon heat shock has been reported:

Cardiello et al. [96] described global changes of RNA Pol II occupancy at mRNA genes as a consequence of heat shock treatment. They observed loss of Pol II accumulation at 3´end of coding genes, in conjunction with Pol II peaks that extended tens of kilobases downstream of

genes 3´ends. They further observed persuasive loss of termination which they found to be a result of impaired recruitment of the termination machinery. Put together, they concluded with a dual regulation of Pol II under heat shock conditions, in which they observed a decrease in levels of new Pol II occupancy, in as much as tolerance for polymerases in the act of transcription to cross gene body and traverse the normal point of termination [96].

Hence, these fine-tuned transcriptional programs can easily be disturbed in response to environmental changes, and herein be utilized as a method to investigate transcription on a mechanistic level. As previously described, Pol II CTD Ser2 phosphorylation is enriched downstream of the gene 3´end, and CDK12 is thought to phosphorylate CTD Ser2. By merging these clues, a possibility to investigate the contribution of CDK12 in regulating aberrant Pol II, arise.

In light of these theories, HeLa cells were in the present study given heat shock treatment in order to investigate the contribution of CDK12 in regulating aberrant RNA Pol II, as well as to observe the subsequent phosphorylation states of CTD Ser2 and Ser5.

4.1.2 THZ531 treatment induces covalent inhibition of CDK12/13

THZ531 is a specific inhibitor of CDK12/13 [75]. THZ531 is found to covalent inhibit both CDK12 and CDK13 by modification of the Cys-1039 residue, which is located nearby the active site, and by interacting with the ATP binding site to prevent ATP binding [75]. Together, these two features of THZ531 blocks the catalytic activity of CDK12/13. THZ531 is further found to downregulate expression of DDR genes in Jurkat cells, consistent with the role of CDK12 in DDR gene transcription, and to significant downregulate key Jurkat transcription factors [75]. Since then, THZ531 has been used regularly as selective inhibitor of CDK12/13 [69, 74].

In this study we wanted to investigate the global distribution of RNA Pol II, as well as the phosphorylation states of CTD Ser2 and Ser5 upon THZ531-induced CDK12/13 inhibition.

4.1.3 Lipid-mediated transfection of CDK9/12/13-GFP BAC constructs

In the present study, HeLa cells were transfected with bacterial artificial chromosome (BAC) transgenes, which is defined as vectors harboring entire loci with all regulatory elements

involved in expression regulation of the gene [113]. A limitation in copy numbers of integrated BAC transgene into the chromosome (1-3 copies) makes the BAC-based expression induce gene expression at near endogenous levels, in order to recapitulate *in vivo* conditions [113]. Under artificial overexpression of bait-proteins, non-interactors might be found to interact [114]. Thus, BAC-induced expression at near endogenous levels is considered an adequate approach in investigation of protein-protein interactions.

4.1.4 Quantification of transcript levels between BAC-transfected HeLa cell lines

For experiments to be conducted on the right basis, the expression of GFP-tagged recombinant genes was confirmed by RT-qPCR.

As previously described, CDK12 and CDK13 are both known to phosphorylate Pol II CTD during the transcription cycle, although their precise target residues are discussed. The roles of CDK9 in RNA CTD phosphorylation for cell cycle progression is more established, as research has revealed the ability of CDK9 to phosphorylate CTD Ser2 [19], and most likely Ser5 [66]. Thus, CDK9 was in this study used as a positive control.

RT-qPCR was assessed to quantify the transcript levels between the differentially transfected HeLa cell lines (figure 5). Moreover, as CDK12 was expressed with either N- or C-terminally tagged GFP, we wanted to investigate which version expressed the recombinant epitope tagged gene most efficiently. The analysis revealed different expression levels of BAC-transfected proteins. C-terminal tagged *CDK12 9128#4* demonstrated greater expression levels than the N-terminal tagged *CDK12 9129#5* version. For this reason, C-terminal tagged CDK12 was used in subsequent analyses.

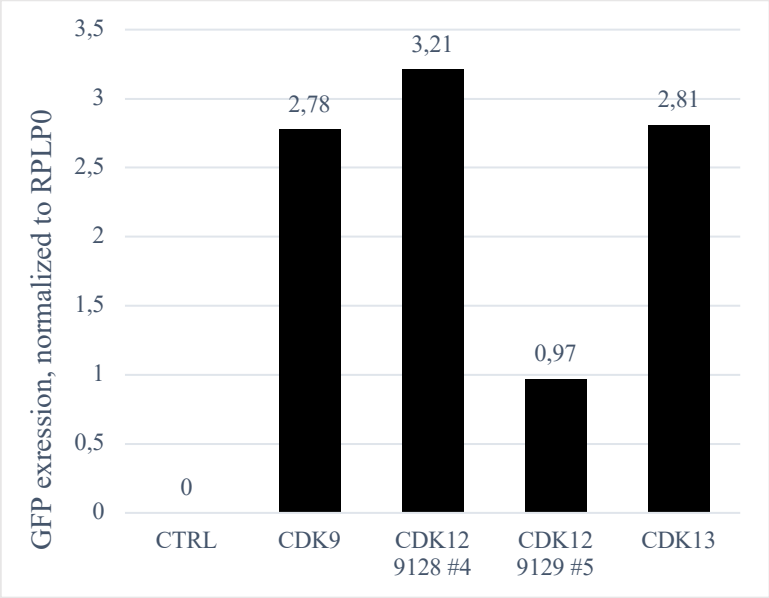


Figure 5. RT-qPCR analysis detecting mRNA levels of BAC-transfected GFP-tagged proteins. Detection of *CDK9-GFP*, *CDK12 9128#4-GFP*, *CDK12 9129#5-GFP* and *CDK13-GFP* transcript levels, normalized to internal *PRLP0* expression. Non-transfected cells was used as negative control (CTRL).

4.1.5 Validation of CDK9-GFP, CDK12-GFP and CDK13-GFP recombinant protein expression

Western blot analysis was performed in order to detect both endogenous- and GFP-epitope tagged variants of CDK9 (figure 6). Two bands were present at 50 kDa in CDK9-stained lysate. The band at 50 kDa should not be present in control lane (left panel, left lane) as the cells were untransfected, further indicating that the band in the CDK9-GFP transfected cells (left panel, right lane) could not be identified as CDK9-GFP. Moreover, no band for CDK9-GFP was to be detected in GFP-stained protein lysate (right panel, right lane). Hence, the transfection was not successful.

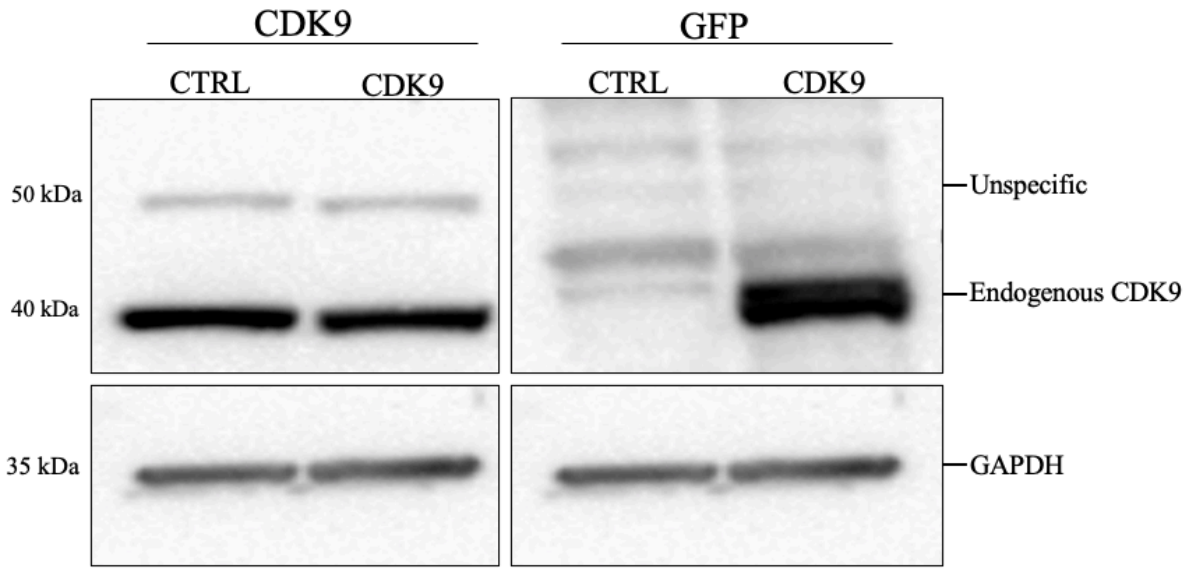


Figure 6. Western blot analysis detecting endogenous CDK9 and CDK9-GFP. Western blot analysis on untransfected (CTRL) and CDK9-GFP transfected (CDK9) HeLa cell lysates. Left panel represents anti-CDK9 stained lysate, and right panel represents anti-GFP stained lysate.

Western blot analysis was further assessed in order to detect endogenous CDK12, as well as GFP-tagged CDK12 and CDK13 recombinant proteins (figure 7). As demonstrated in figure 7, both endogenous CDK12 (left panel) and GFP-CDK12 (right panel) was detected in the protein analysis. On the contrary, CDK13-GFP was not detected.

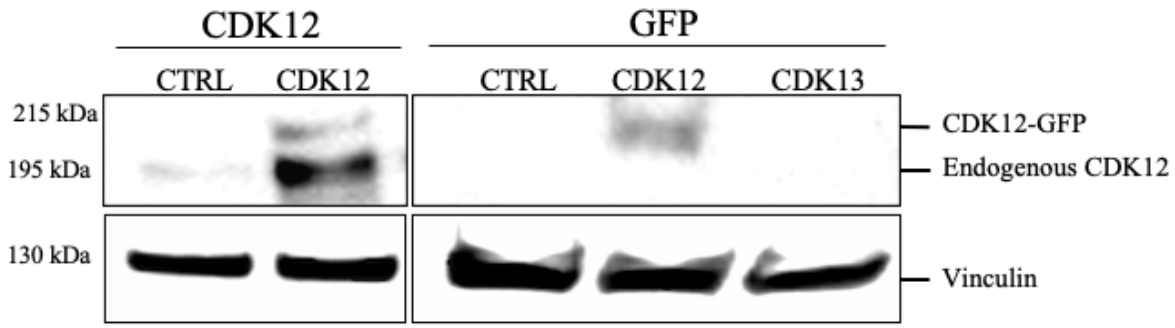


Figure 7. Western blot analysis detecting endogenous and epitope-tagged versions of CDK12 and CDK13. Left panel represents CDK12-GFP transfected HeLa cells (CDK12) compared to untransfected HeLa cells (CTRL) stained in anti-CDK12 primary antibody. Right panel represents protein analysis of untransfected cells (CTRL), CDK12-GFP transfected cells (CDK12) and CDK13-GFP transfected cells CDK13) stained in anti-GFP primary antibody.

4.1.6 Validation of CDK9-GFP, CDK12-GFP and CDK13-GFP immunoprecipitation

Immunoprecipitation was performed in another try of validating successful transfection of CDK9-GFP (figure 8). Two bands should have been present in CDK9-GFP transfected cell lysate in input-lane (left panel) indicating endogenous and GFP-tagged CDK9. Moreover, no bands were to be detected in immunoprecipitated CDK9-GFP transfected cell lysate (left panel, IP-marked lane of CDK9). CDK9-GFP was therefore excluded from further analysis.

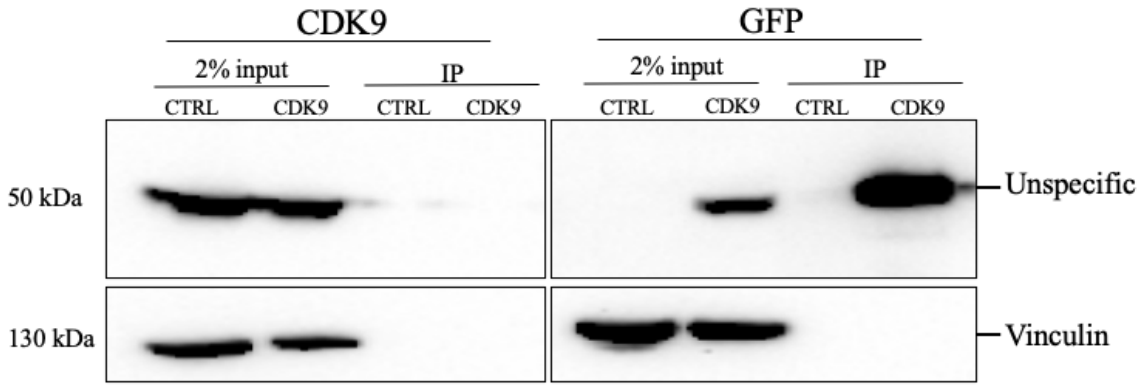


Figure 8. Western blot analysis detecting CDK9 and immunoprecipitated CDK9-GFP. Western blot analysis on untransfected (CTRL) and CDK9-GFP (CDK9) transfected HeLa cell lysates. Both panels represent protein analysis of 2% input lysate and immunoprecipitated (IP) proteins. Left panel represents anti-CDK9 stained lysate, while right panel represents anti-GFP stained lysate.

Immunoprecipitation was performed to further validate successful transfection of CDK12, in addition to a new attempt to detect CDK13-GFP in transfected HeLa cells (figure 9). Western blot analysis revealed successful immunoprecipitation of CDK12-GFP (right panel, IP-marked lane of CDK12), but no band was visible for CDK13-GFP (right panel, IP-marked lane of CDK13). Thus, the foundation for further detection of CDK13 was not present.

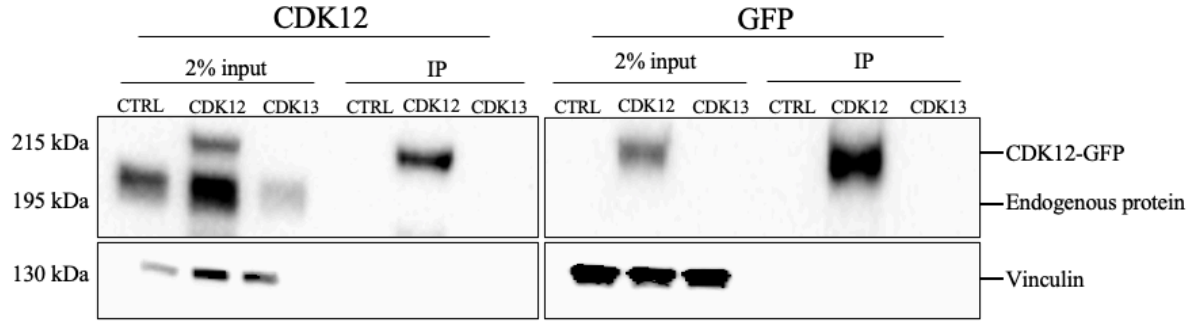


Figure 9. Western blot analysis detecting CDK12 and immunoprecipitated CDK12-GFP. Western blot analysis on untransfected (CTRL), CDK12-GFP transfected (CDK12) and CDK13-GFP transfected (CDK13) HeLa cell lysates. Both panels represent protein analysis of 2% input lysate and immunoprecipitated (IP) proteins. Left panel represents anti-CDK12 stained lysate, while right panel represents anti-GFP stained lysate.

4.1.7 Localization of endogenous CDK12 cells

To localize CDK12 in the cells, we performed time-lapse based live cell imaging. However, the experiment was unsuccessful due to constant background fluorescent signal.

To further localize CDK12 in the cells, we performed cellular fractionation. HeLa cells were fractionated and separated into cytosolic, nucleosolic and insoluble nuclear compartments. Fractions were analyzed by western blotting. As positive controls, specific antibodies with affinity towards the different cellular compartments were used. Histone H3 is chromatin-bound (NP), GAPDH is localized in cytoplasm (CS) and Suz12 is located in nuclear soluble (NS) and nuclear pellet (NP).

The upper lane in figure 10 demonstrates that CDK12 is mainly localized in cytoplasmic soluble (CS) and nuclear soluble (NS). However, a weak band was also observed on chromatin in nuclear pellet (NP). The finding of CDK12 present in the nucleus motivated our further work with CDK12 as a transcriptional CTD kinase.

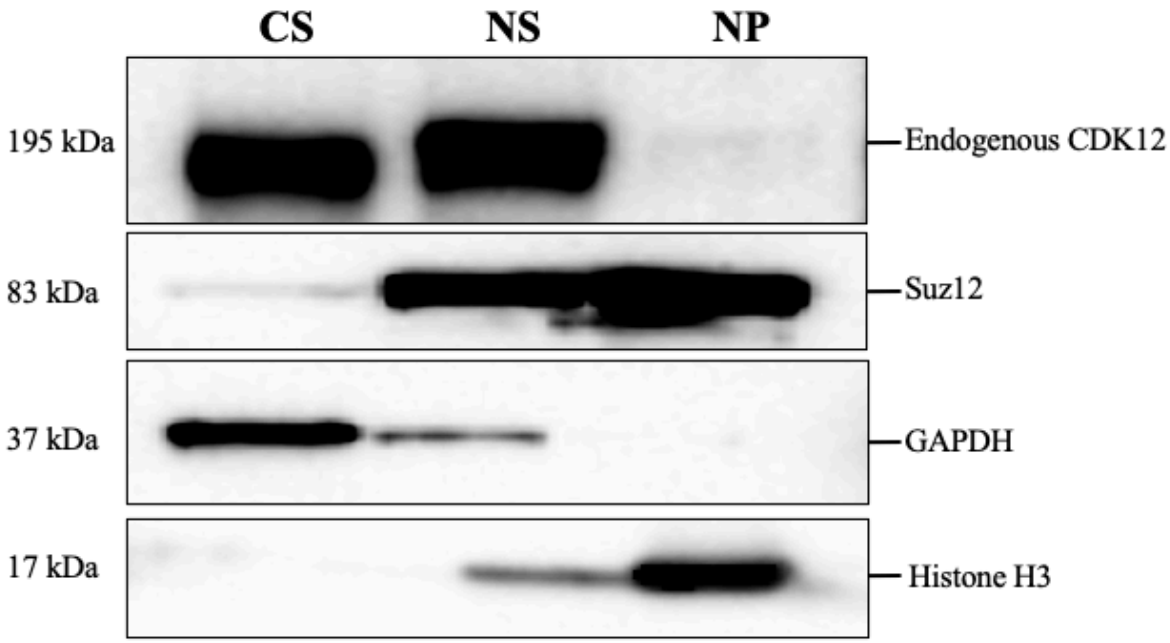


Figure 10. Western blot analysis of CDK12 in HeLa cellular compartments. Compartments from HeLa cellular fractionation was stained in anti-CDK12 primary antibody to localize CDK12 in the cell. Suz12, GAPDH and Histone H3 antibodes functioned as positive controls respectively for nuclear soluble (NS), cytoplasmic soluble (CS) and nuclear pellet (NP).

4.2 Identification of CDK12-bound chromatin interactome during normal and heat shock conditions

The association of CDK12 with other chromatin-bound proteins, under cross-linking conditions, has not been described before by MS analysis. We asked whether the interactome of chromatin-bound CDK12 could reveal new insights into the role of CDK12 in transcription and RNA processing. To assess this, we performed chromatin enrichment and co-immunoprecipitation followed by MS. The chromatin enrichment for proteomics (ChEP) protocol was first carried out, but we did not succeed to detect chromatin-associated proteins with SDS-PAGE gel electrophoresis followed by Coomassie Brilliant Blue protein staining. Thus, chromatin immunoprecipitation-mass spectrometry (ChIP-MS) was conducted.

ChIP-MS was in this study conducted three times. The first two runs were done as test runs, while the last run was performed in triplicate samples. For this reason, the first two runs included cells respectively treated with 0 and 10, and 0, 10 and 30 minutes of heat shock. The second run indicated that CDK12-GFP was lost from chromatin after heat shock treatment. As our initial intention was to study CDK12 on chromatin, the last run excluded cells treated with 30 minutes of heat shock.

During the last run of MS, which was carried out in triplicate samples, the mass spectrometer got clogged throughout the process, and samples were not examinable. Hence, the following results need to be seen in light of the fact that these results were acquired from two separate runs of MS.

4.2.1 Sonication efficiency on agarose gel electrophoresis

Sheared, de-crosslinked chromatin was analyzed on a 1.2% agarose gel. The size range ought to be at 200-600 bp. Figure 11a represents chromatin samples from the first run of ChIP-MS, where 200 kb bands were to be detected – however, predominantly in control samples. Figure 11b represents chromatin from the second run, where the agarose gel analysis was not successful. Figure 11c represents chromatin samples from the last run of MS, where samples were run in triplicates. The bands seem to have merged, but they were not evidently present at expected size range.

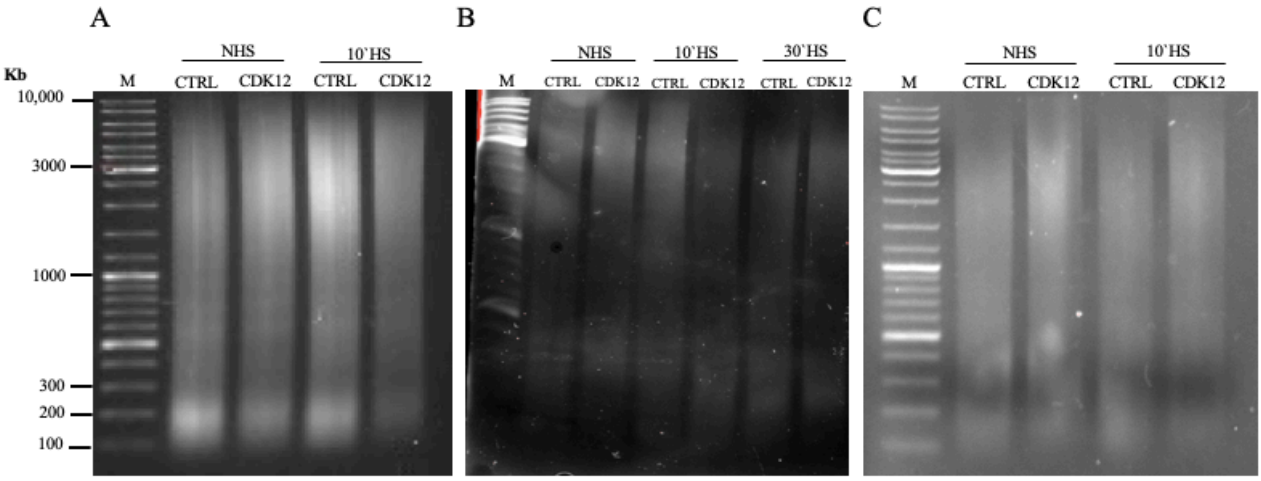


Figure 11. Size assessment of sheared chromatin on agarose gel. HeLa cells were fixed with formaldehyde and chromatin prepared according to the RIME protocol with certain adjustments (see section 3.10). Samples were sonicated 30s ON/OFF for 5 cycles, before analyzed on 1.2% agarose gel. Untransfected cells were used as control (CTRL) for CDK12-GFP transfected HeLa cells (CDK12). Treatments are indicated with non-heat shock (NHS), 10 minutes heat shock (10´HS) and 30 minutes heat shock (30´HS). Figure A represents samples from first run of ChIP-MS, figure B represents samples from the second run and figure C represents samples from the third run (triplicate samples). M = marker, GeneRuler Ladder Mix, 100 – 10,000 kb.

4.2.2 Alterations in CDK12-GFP bound chromatin interactome during normal and heat shock conditions

In order to organize and visualize CDK12-GFP bound interactome identified by MS, Venny 2.1.0 software was used. The Venn diagram in figure 12 represents proteins bound to CDK12-GFP under normal conditions, compared to 10- and 30 minutes of heat shock treatment from the second round of MS. CDK12-GFP bound proteome differ between the three conditions (NHS, 10´HS and 30´HS).

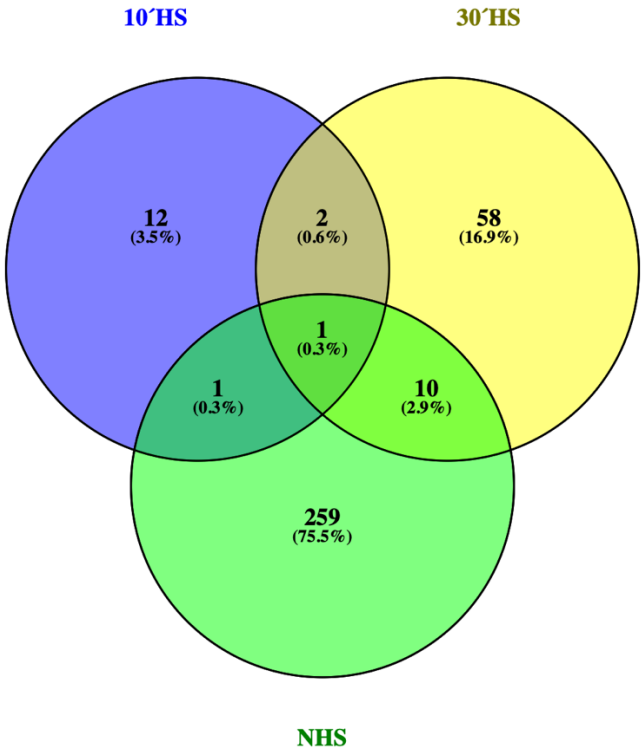


Figure 12. Venn diagram displaying shared and distinct CDK12-GFP bound proteins under normal and heat shock conditions. MS results from CDK12-GFP ChIP-MS, compiling CDK12-GFP-bound proteome under normal conditions (NHS), 10 minutes (10´HS) and 30 minutes (30`HS) of heat shock at 42 °C. The figure was obtained by the use of Venny 2.1.0 software.

4.2.3 CDK12 is lost from chromatin after 30 minutes of heat shock

CDK12-GFP MS data indicated that CDK12 is lost from chromatin upon heat shock treatment. To validate this, western blot analyses utilizing aliquots of reverse cross-linked, non-immunoprecipitated lysates from the second round of ChIP-MS, was conducted. Specific antibodies were used to detect the presence of endogenous CDK12 and CDK12-GFP on chromatin under normal conditions, and after 10- and 30 minutes of heat shock treatment. Non-transfected cells were used as control.

Figure 13a represents western blot analysis of lysates from non-transfected cells (CTRL) and CDK12-GFP transfected cells (CDK12) stained in GFP primary antibody. Bands are to be detected at expected size in non-heat shocked (NHS) cells and cells treated with 10 minutes of heat shock (10´HS), but no bands are visible in lysates from cells treated with 30 minutes of heat shock (30´HS).

A

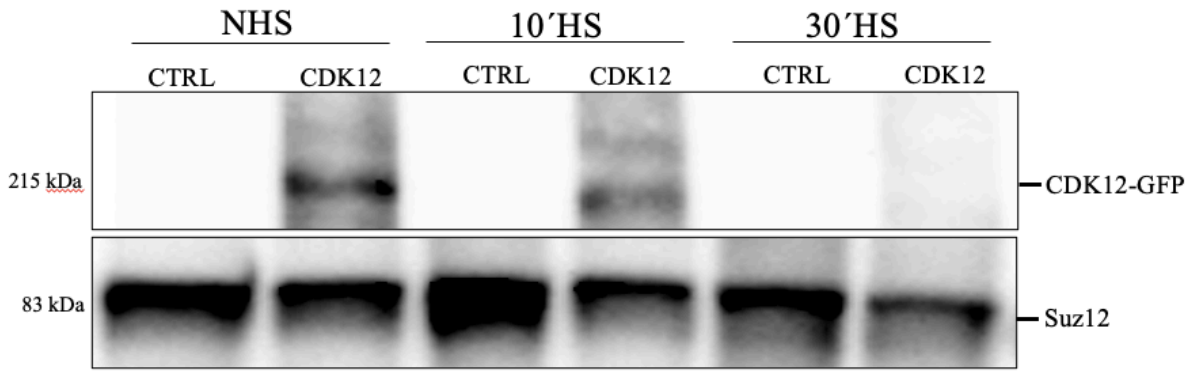


Figure 13A. Western blot analysis of chromatin-bound CDK12-GFP under normal and heat shock conditions. Protein analysis of reverse cross-linked ChIP-lysates from non-heat shocked versus heat shocked cells. Cell lysates from non-treated cells (NHS), compared to cells treated with 10 minutes (10´HS) and 30 minutes of heat shock (30´HS) at 42 °C were stained in anti-GFP primary antibody. Non-transfected cells were here used as negative control.

Figure 13b represents western blot analysis of lysates from non-transfected cells (CTRL) and CDK12-GFP transfected cells (CDK12) stained in CDK12 primary antibody. Faint bands are present in both non-heat shocked (NHS) and cells treated with 10 minutes of heat shock (10´HS). After 30 minutes of heat shock (30´HS), no bands are to be detected.

B

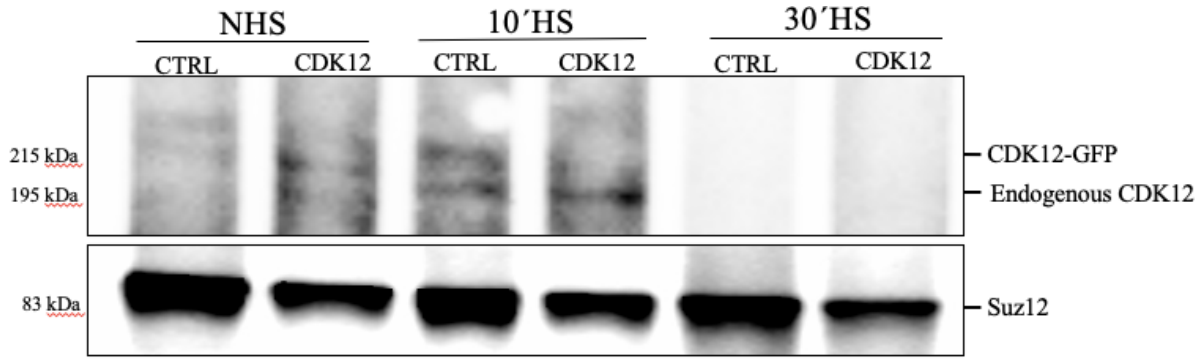


Figure 13B. Western blot analysis of chromatin-bound CDK12 under normal and heat shock conditions. Protein analysis of reverse cross-linked ChIP-lysates from non-heat shocked versus heat shocked cells. Cell lysates from non-treated cells (NHS), compared to cells treated with 10 minutes (10´HS) and 30 minutes of heat shock (30´HS) at 42 °C were stained in anti-CDK12 primary antibody. Non-transfected cells functioned as positive control.

4.2.4 Enrichment analysis of CDK12-GFP bound chromatin interactome during non-heat shock conditions

CDK12 was found to disassociate from chromatin after 30 minutes of heat shock, and as our initial intention in this study was to analyze the role of CDK12 on chromatin, further analysis was conducted on MS data from non-heat shocked cells. The first run identified a total of 1213 proteins, while the second run identified 1378 proteins. Proteins present in control-co-immunoprecipitations were filtered out when compared to CDK12-GFP samples. From the non-heat shocked samples, 37 proteins from the first run and 271 proteins from the second run remained for analysis after control-filtration. Moreover, identical proteins from the two runs were filtered out. Thus, the following analyses includes a total of 292 proteins (Table S6 and S8, Section C, Appendix).

Enrichment analysis was performed in the Metascape software in order to identify enriched functional groups of CDK12-GFP bound proteome (figure 14). Terms with a p-value < 0.05, a minimal overlap of 3, and an enrichment factor > 1.5 were, based on their similarities, clustered. Darker color and longer bars indicate a lower p-value, and a lower p-value indicate lower chance of that the observed enrichment is obtained by randomness. The input gene list was here compared to Gene Ontology (GO) Biological Processes and the TRRUST database (human and mouse transcriptional regulatory networks).

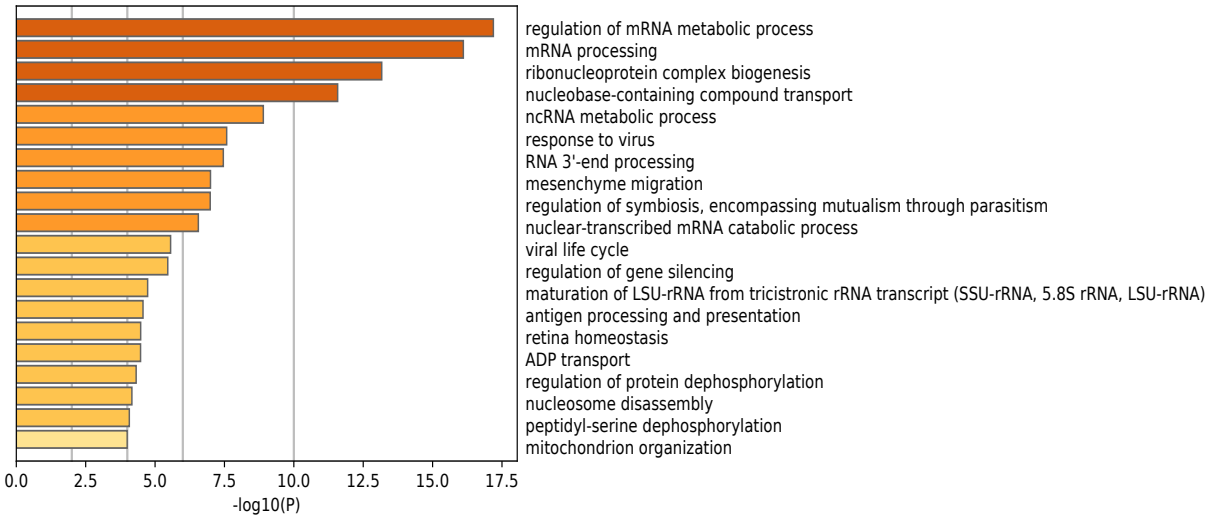


Figure 14. Enriched terms of CDK12-GFP-bound proteome identified by Metascape. Bar graph of Gene Ontology (GO) and TRRUST enriched terms. Statistical significance is indicated by p-value, in which darker colour and longer bars implies a lower p-value.

4.2.5 Visual representation of CDK12-GFP proteome

To gain insight into the role of CDK12 on chromatin, CDK12-GFP bound proteins identified by MS was visualized in an interaction network (figure 15). Enrichment analysis data from Metascape (figure 14) was plotted into the Cytoscape software. Edges connects nodes of functionally correlated proteins (based on the GO Biological Processes and TRRUST databases). Nodes of identical colors represent proteins involved in the same functional groups, and adjacent nodes of correlated functional profiles are indicated in similar shades of the same color. Nodes of less interest are downscaled in size and gray-colored. Seven proteins of interest found to harbor similar functional properties as the already existing nodes were added. These included SMARCE1, SMARCA4, SMARCC2, GATAD2, SAP18, PNN and RNF20. These ancillary interactions were determined by the use of STRING and UniProt databases.

Among interesting findings were RNA binding proteins, including components of the RNA surveillance pathway (EXOSC4, ECOSC5, EXOSC7 and SKIVL2L), proteins involved with the exon junction complex ([EJC], MAGOHB, RBM8A, SAP18 and PNN), splicing factors (SRSF4, SRSF5, BCAS2, DHX16, PRPF4, PQBP1, SF3A3, SNRPG, SNW1) and proteins involved in transcriptional regulation (CDC73, FIP1L1, NELFE, NELFCD and RNF20).

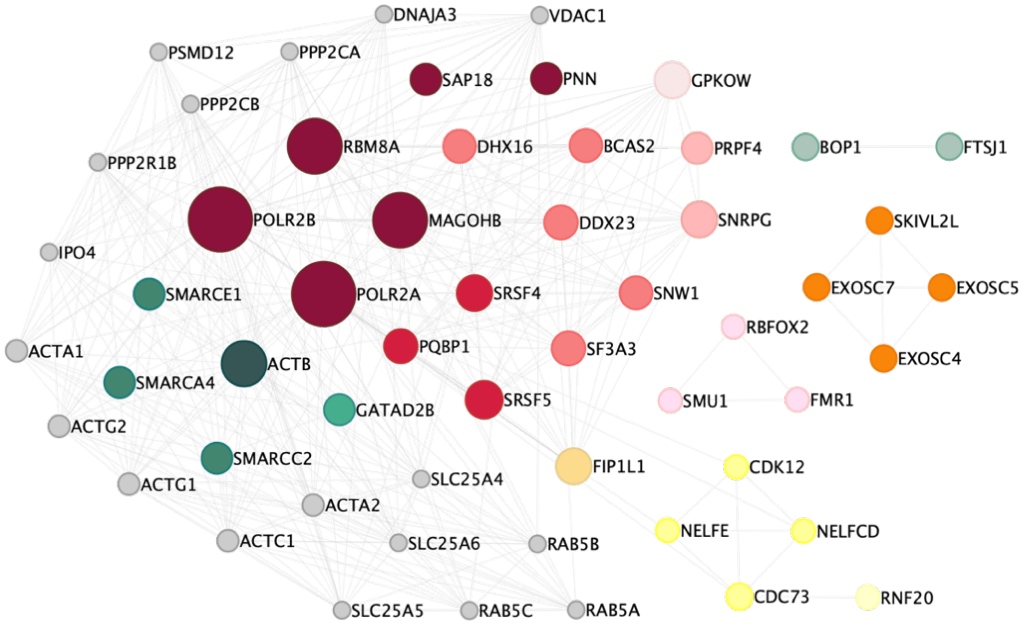


Figure 15. Protein interaction network representing chromatin-bound CDK12 interactome. Network representing proteins co-immunoprecipitated with chromatin-bound CDK12-GFP. The MS dataset was processed in Metascape before network was visualized in Cytoscape. Edges connects nodes of functionally correlated proteins (GO Biological Processes and TRRUST databases). Nodes of identical colors represent proteins involved in the same functional groups, and nodes of correlated functional profiles are indicated in similar shades of the same color. Gray nodes indicate proteins of less interest.

4.2.6 Visual representation of chosen protein groups of CDK12-GFP proteome

To gain a deeper understanding in the functional properties of CDK12 on chromatin, the proteome of CDK12-GFP involved in transcription and RNA processing was investigated in detail. The web-based gene enrichment analysis tool Enrichr was used to investigate the functional groups. Enriched terms of interest, based on GO Biological Processes, is visualized in figure 16. Proteins of interest can be generally divided into components involved in splicing (pink), proteins involved with the EJC (red), proteins involved in mRNA surveillance (green), regulatory transcription factors (yellow) and proteins involved in chromatin remodeling (turquoise).

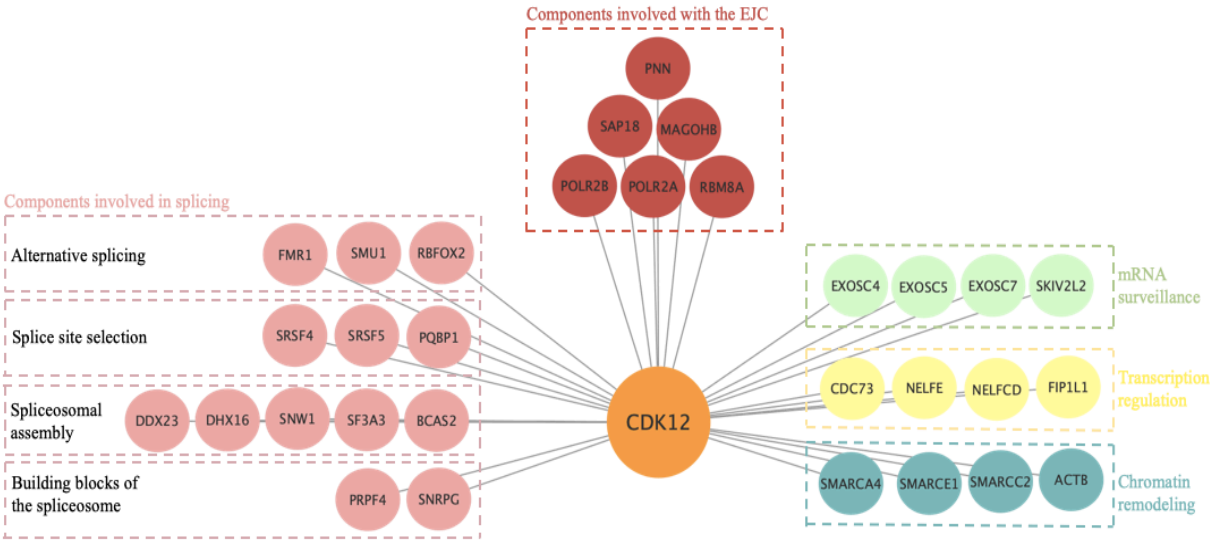
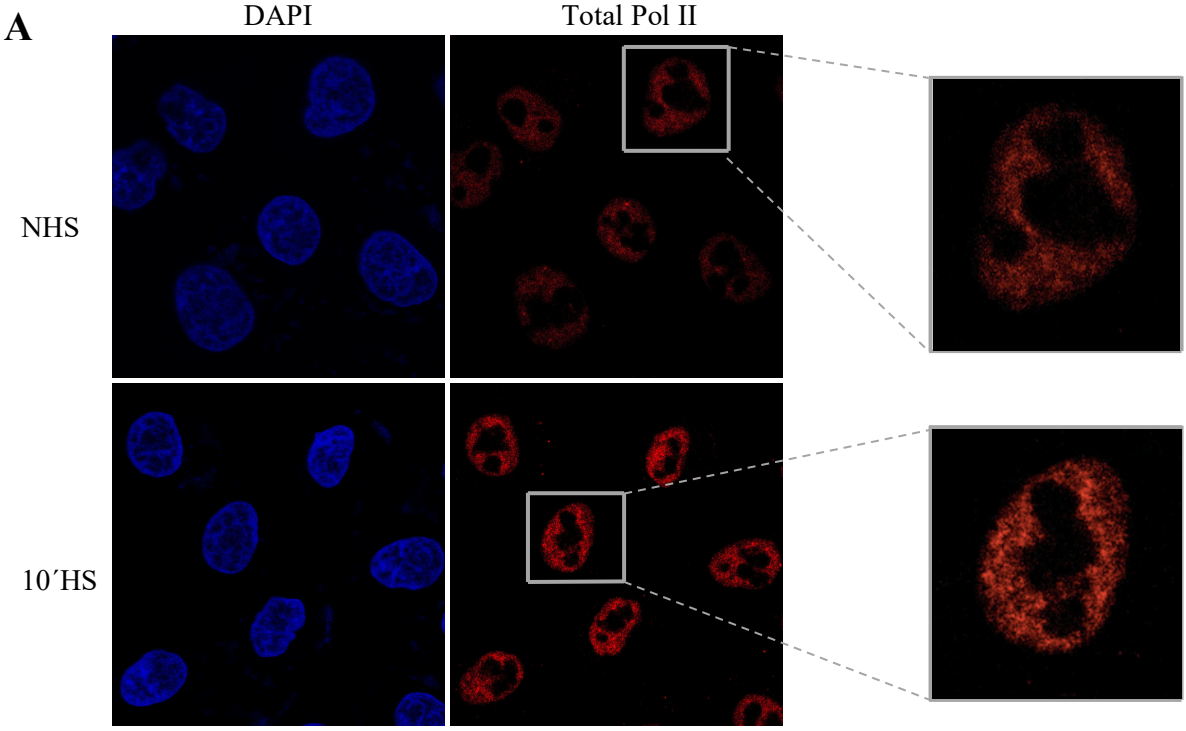


Figure 16. Visual representation of chosen proteins involved with chromatin-bound CDK12. Enriched terms of interest, based on GO Biological Processes, among proteins co-immunoprecipitated with chromatin-bound CDK12. Analysis was conducted by the web-based bioinformatics tool Enrichr. Functional groups of interest can be generally divided into five groups: splicing (pink), EJC (red), mRNA surveillance (green), transcription regulation (yellow) and chromatin remodeling (turquoise). Figure was created in Cytoscape.

4.3 The effect of heat shock- and THZ531 treatment on phosphorylation state of RNA Pol II-phosphorylated CTD species

4.3.1 Heat shock induces a global change in RNA Pol II distribution pattern

As previous data have demonstrated heat shock to induce transcriptional changes and stall RNA Pol II on the gene body, we wanted to investigate if Pol II exhibits a different distribution pattern after 10 minutes of heat shock compared to in non-heat shocked cells. Confocal imaging was used to address this. After 10 minutes of heat shock, a speckled pattern of RNA Pol II is observed (figure 17a). The same pattern is not to be observed in non-heat shocked cells.



Figur 17A. Effects on RNA Pol II distribution patterns after heat shock treatment. Immunofluorescence staining of anti-Pol II antibody recognizing the human RNA Polymerase II (right panel) in untransfected HeLa cells. Nuclei was counterstained with DAPI (left panel). Cells were either incubated under normal conditions at 36ºC (NHS) or 10 minutes at 42ºC (HS). Non-heat shocked cells were incubated in control media (DMEM + 0,0003% DMSO). The gray rectangles indicate cells shown in higher magnification. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.2 Heat shock induces global change in distribution pattern of CDK12-GFP

During the transcription cycle, CDK12 may phosphorylate RNA Pol II for productive elongation. This, including the heat shock-induced speckled pattern of Pol II (figure 17a), incentivized us to investigate the distribution pattern of CDK12-GFP upon heat shock treatment. Following 10 minutes of heat shock, a redistribution of CDK12-GFP was observed compared to in non-heat shocked cells (figure 17b).

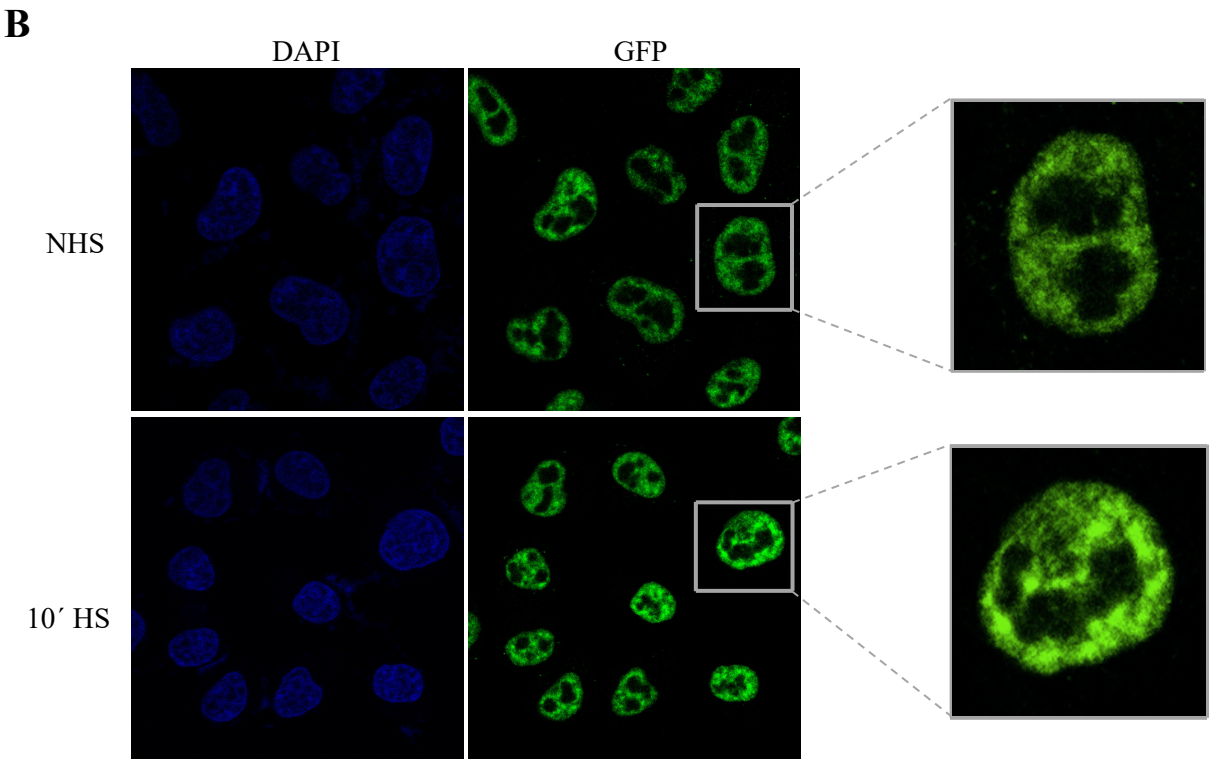


Figure 17B. Effects on GFP-CDK12 localization patterns upon heat shock treatment. Immunofluorescence detection of CDK12-GFP with specific antibody recognizing the GFP sequence of CDK12-GFP (right panel) in CDK12-GFP transfected HeLa cells. Nuclei was counterstained with DAPI (left panel). Cells were either incubated under normal conditions at 36ºC (NHS) or 10 minutes at 42ºC (HS). The gray rectangles indicate cells shown in higher magnification. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.3 Heat shock induces a decrease in CTD Ser2 phosphorylation

There are many reports suggesting that CDK12 phosphorylates Pol II CTD Ser2 for productive elongation, reviewed in [69]. The speckled pattern of Pol II (figure 17a) and the redistribution of CDK12-GFP upon heat shock (figure 17b) encouraged us to investigate the effect of heat shock on CTD Ser2 phosphorylation state. Upon heat shock, a decreased level of Ser2 phosphorylation is observed (figure 17c).

C

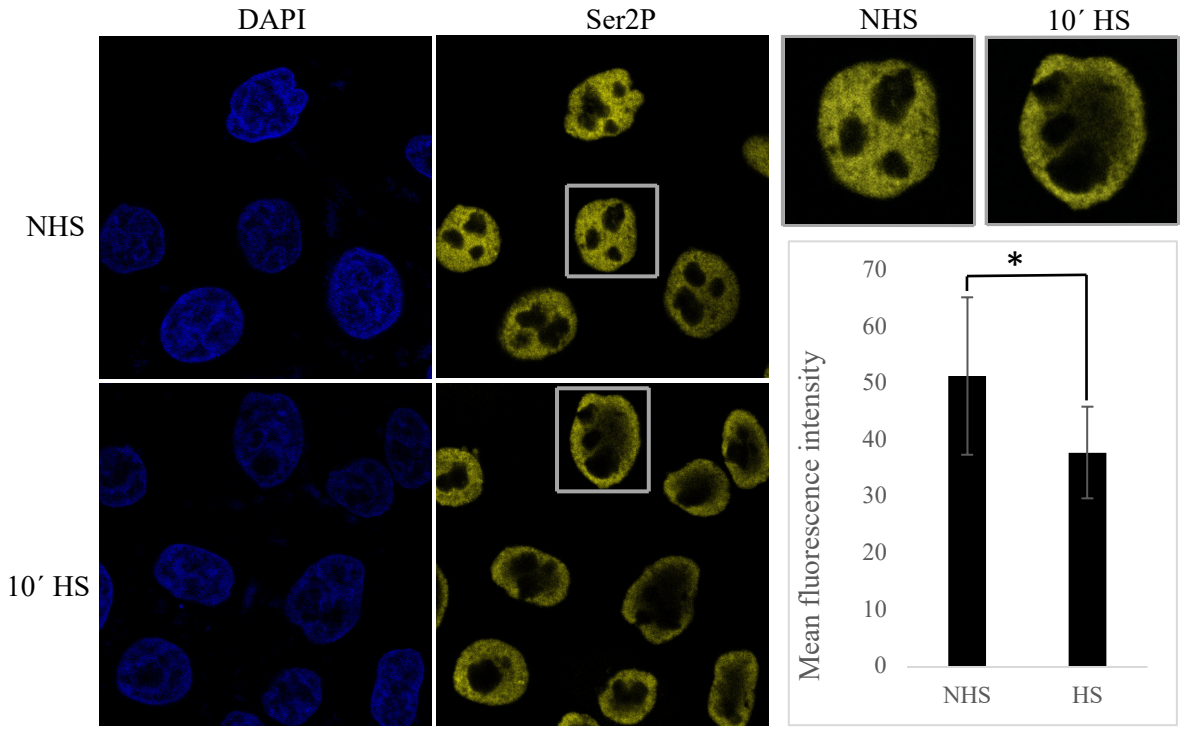


Figure 17C. Effects on phosphorylation state of Ser2 upon heat shock. Immunofluorescent staining of phospho anti-Ser2 specific antibody recognizing phosphorylated CTD Ser2 (right panel). Nuclei was counterstained with DAPI (left panel). Cells were either incubated under normal conditions at 36ºC (NHS) or 10 minutes at 42ºC (HS). The gray rectangles indicate cells shown in higher magnification. Statistical analysis by Student´s t-test, 2 tailed, equal variance, * = $p < 0.05$. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.4 Heat shock induces a decrease in CTD Ser5 phosphorylation

There are conflicting evidence over whether CDK12 phosphorylates CTD Ser5 [69]. The decrease in CTD Ser2 phosphorylation state upon heat shock (figure 17c) inspired us to investigate the phosphorylation levels of CTD Ser5 under the same conditions. After heat shock treatment, a decrease in CTD Ser5 phosphorylation is observed (figure 17d).

D

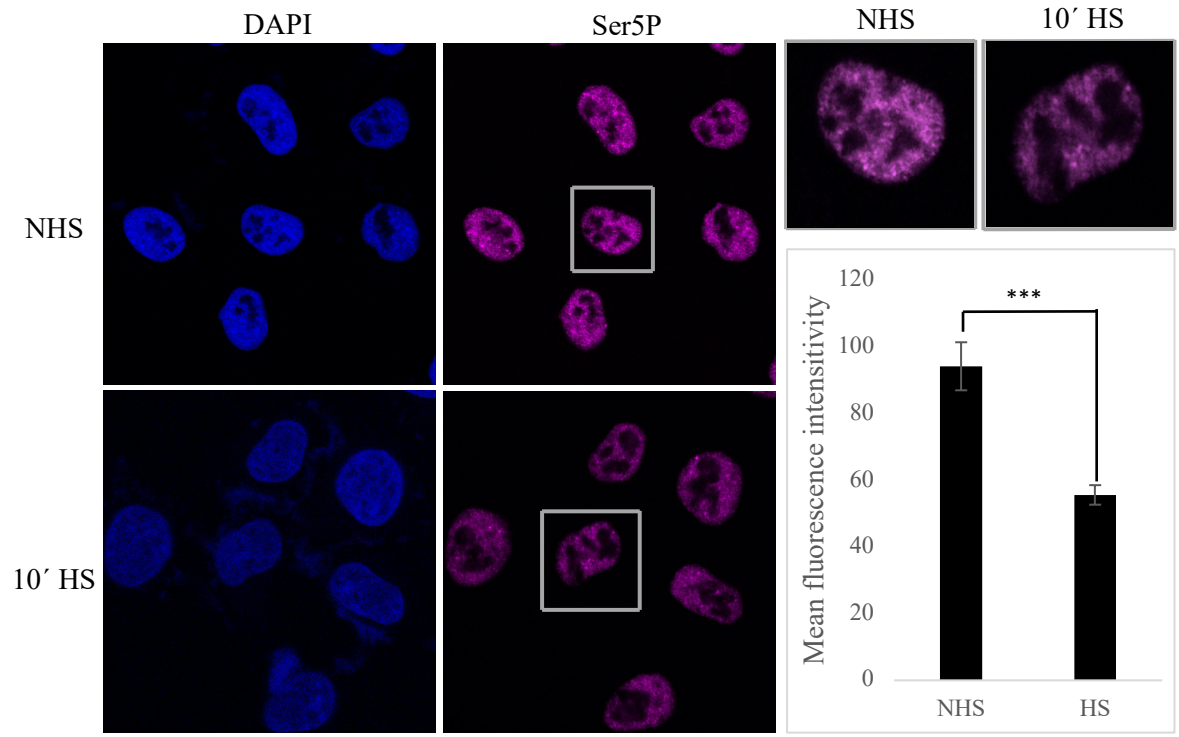


Figure 17D. Effects on global distribution patterns of phosphorylated Ser5 upon heat shock. Immunofluoresent staining of phospho anti-Ser5 specific antibody recognizing phosphorylated CTD Ser5 (right panel). Nuclei was counterstained with DAPI (left panel). Cells were either incubated under normal conditions at 36ºC (NHS) or 10 minutes at 42ºC (HS). The gray rectangles indicate cells shown in higher magnification. Statistical analysis by Student´s t-test, 2 tailed, equal variance, *** = $p < 0.001$. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.5 THZ531 treatment induces a global change in RNA Pol II distribution pattern

To investigate these heat shock-induced observations in view of CDK12/13, we conducted THZ531-induced inhibition of CDK12/13

CDK12 is associated with productive transcriptional elongation, which is linked to its ability to phosphorylate Pol II CTD. We wanted to investigate the global localization pattern of Pol II upon THZ531-induced inhibition of CDK12/13. After 6 hours of THZ531 treatment, a speckled pattern is to be observed compared to in non-heat shocked cells (figure 18a).

A

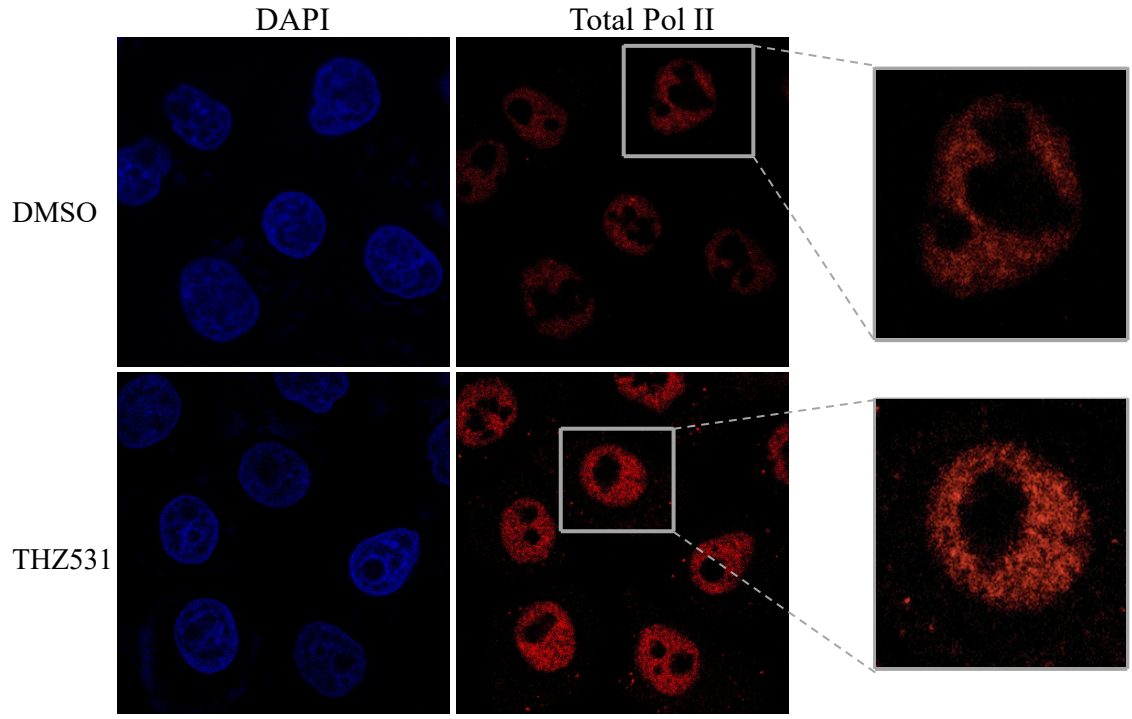


Figure 18A. Effect on global localization patterns of RNA Pol II upon THZ531-induced CDK12/13 inhibition. Immunofluorescenct staining of anti-Pol II recognizing RNA polymerase II (right panel). Nuclei was counterstained with DAPI (left panel). Cells were treated either with control media (0.0003% DMSO) or 3µM THZ531 for 6 hours. The gray rectangles indicate cells shown in higher magnification. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.6 THZ531 treatment induces a decrease in CTD Ser2 phosphorylation

Studies have indicated that CDK12 may have the ability to phosphorylate Pol II CTD for productive elongation, but its target residues are discussed. The fact that THZ531-treatment lead to a global change in RNA Pol II localization patterns (figure 18a), incentivized us to investigate the phosphorylation state of Ser2 upon CDK12/13 inhibition. After treatment, a global reduction of Ser2 phosphorylation levels is observed (figure 18b).

B

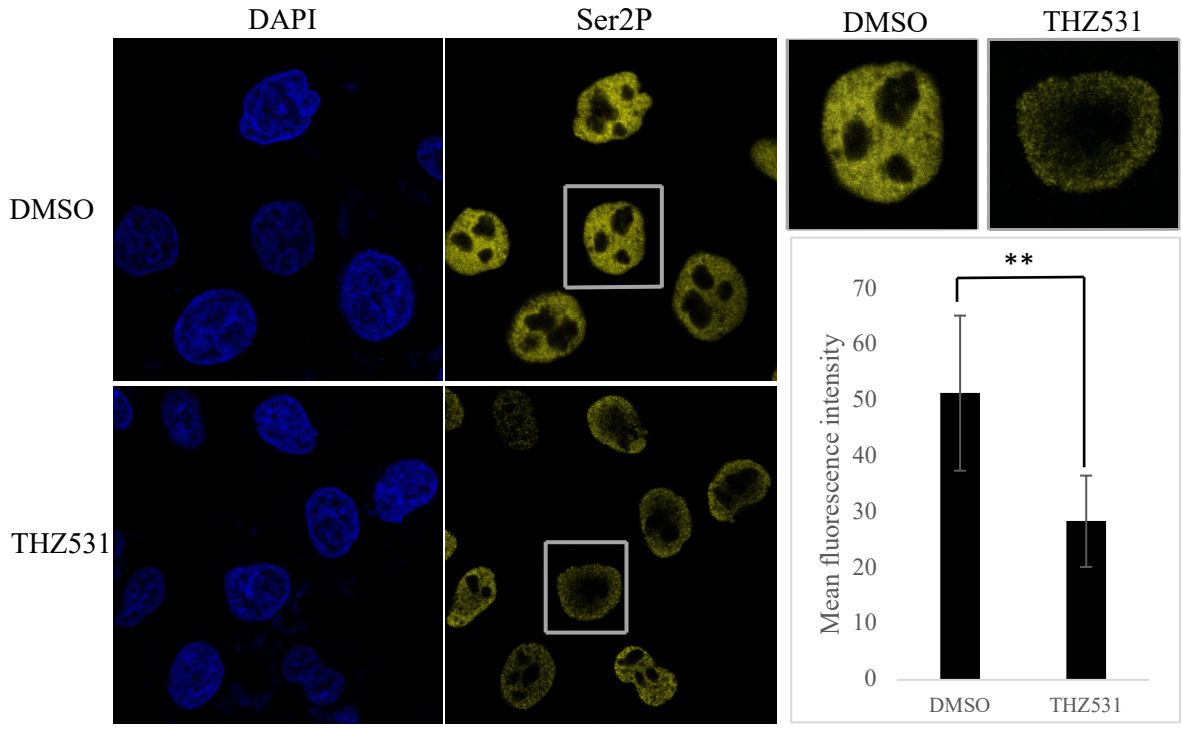


Figure 18B. Effects on CTD Ser2 phosphorylation state upon THZ531 induced CDK12/13 inhibition. Immunofluorescence staining of phospho anti-Ser2 antibody recognizing phosphorylated CTD Ser2 (right panel). Nuclei was counterstained with DAPI (left panel). Cells were treated either with control media (0.0003% DMSO) or 3µM THZ531 for 6 hours. The gray rectangles indicate cells shown in higher magnification. Statistical analysis by Student´s t-test, 2 tailed, equal variance, ** = $p < 0,01$. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

4.3.7 THZ531 treatment induce an increase in CTD Ser5 phoshorylation

To further investigate the role of CDK12 in regulating RNA Pol II by CTD phosphorylation, we investigated the phosphorylation state of CTD Ser5 upon THZ531-treatment. Interestingly, a slight increase in CTD Ser5 is to be observed after THZ531-induced CDK12/13 inhibition (figure 18c).

C

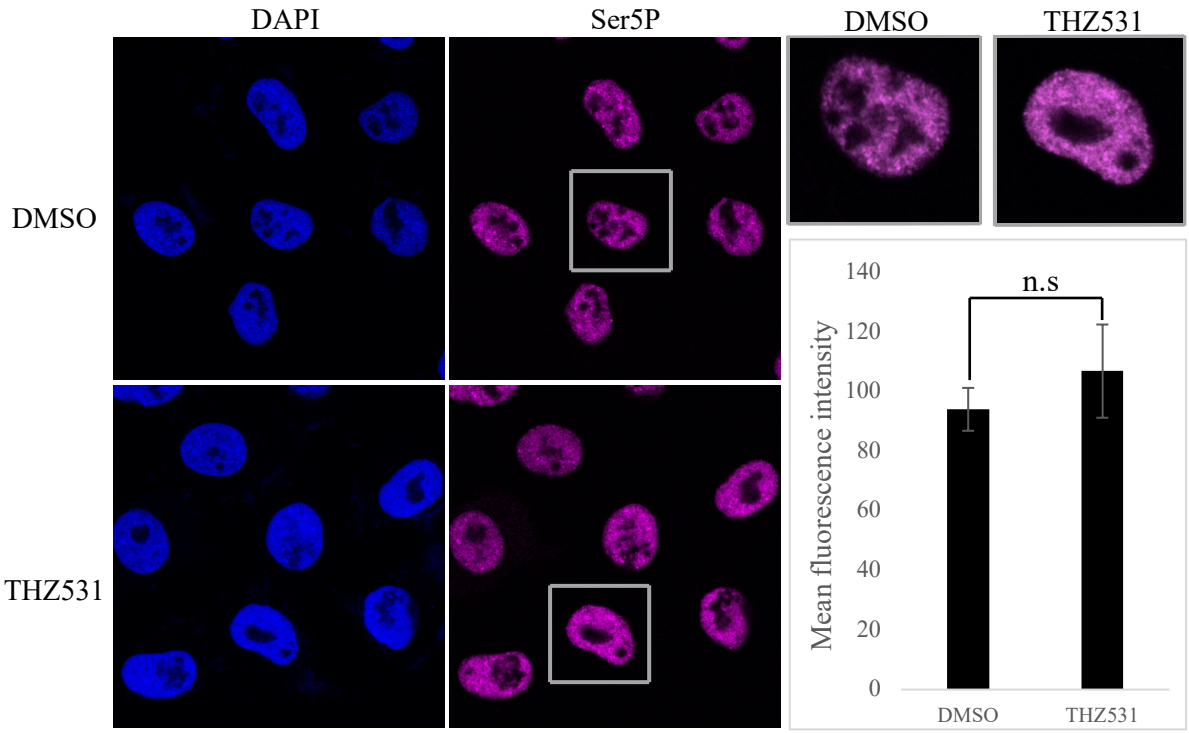


Figure 18C. Effects on CTD Ser2 phosphorylation state upon THZ531-induced CDK12/13 inhibition. Immunofluorescenct detection of phospho anti-Ser5 antibody recognizing phosphorylated CTD Ser5 (right panel). Nuclei was counterstained with DAPI (left panel). Cells were treated either with control media (0.0003% DMSO) or 3µM THZ531 for 6 hours. The gray rectangles indicate cells shown in higher magnification. Statistical analysis by Student´s t-test, 2 tailed, equal variance, n.s = p > 0,05. Images were captured under equal excitation conditions and presented at the same brightness/contrast levels in the ImageJ software.

5. Discussion

5.1 Unsuccessful attempt in detection of BAC-based expression of CDK9- and CDK13-GFP

RT-qPCR results (figure 5) validated successful transfection, as transcripts of CDK9-GFP, C- and N-terminally GFP-tagged CDK12 (9128#4 and 9129#5) and CDK13-GFP were present in the HeLa cell lines. Western blot analysis was further employed to confirm the presence of GFP-tagged proteins (figure 6-9). Clear bands were present in whole cell- and immunoprecipitated protein analysis of endogenous CDK12, as well as for CDK12-GFP. However, no bands were to be detected in CDK9- or CDK13 western blot analyses. As RT-qPCR indicates gene expression through mRNA copy numbers, this method does not imply the translation of mRNA into proteins. Hence, a possible scenario of the unsuccessful western blot analyses is improper translation of epitope-tagged mRNA. One way to assess if the problem is limited to the HeLa cell line would be to transfect a different cell line. Due to time limitations this troubleshooting event was not conducted, and solely CDK12-GFP was taken into further analysis.

5.2 Identification of CDK12-bound chromatin interactome by mass spectrometry analysis

Our MS data is based on two biological ChIP-MS replicates (samples from two different dishes of HeLa cells). The first and the second run gave output lists of respectively 37 and 271 proteins interacting with CDK12-GFP under non-heat shock conditions, demonstrating inter-sample variance and the importance of an experimental repeat.

The ChIP-technique is dependent on variables such as cross-linking conditions, antibody specificity and sonication efficiency [115]. These variables might be reason for the variance of CDK12-GFP binding proteins under non-heat shock conditions, between the two runs. However, as the agarose gel analysis from the second run of ChIP-MS (figure 11b) was not successful, the difference in efficiency of sonication cannot be specified. As we in the agarose gel analyses (figure 11) did not manage to detect bands at expected size range, our study could benefit from optimization of the cross-linking- and sonication conditions. As the bands seemed to merge at quite a high size range, an increase in sonication intervals could have been assessed for sufficient shearing of chromatin. Alternatively, the cross-linking time or formaldehyde concentration could be reduced. Excessive cross-linking reduces sonication

efficiency [116], and these two parameters require correlatively adjustment. To assign the issue, a sample with non-cross-linked cells could have been prepared and analyzed on agarose gel electrophoresis together with cross-linked samples to provide information of which step is in requirement of adjustment. Moreover, to further assure effective cross-linking and sonication conditions as reflected by protein yield, immunoprecipitated product could be analyzed by western blot analysis.

The last run of MS, which was performed in technical triplicates (three injections of the same sample), was lost in the mass spectrometer. Technical triplicates would provide information about intra-sample variations, and further improve the statistical significance of the experiment.

Hence, the ChIP-MS data should be seen in light of these sources of errors. However, our findings match output lists of previous MS studies done on CDK12 with regards to proteins involved in the same complexes and pathways. Our output data will be discussed in the following subsections.

5.2.1 Mass spectrometry data did not identify CDK12-GFP to interact with Cyclin K

Of notion, CycK was in the present study not found to co-immunoprecipitate with CDK12-GFP. CDK12 kinase activity is dependent on its interaction with CycK [77]. One possible reason could be that CDK12-GFP does not fold properly, which consequently does not allow for binding of CycK. One way to address this would be to perform cellular fractionation of CDK12-GFP to determine its cellular localization as compared to endogenous CDK12, and subsequently extract CDK12-GFP and probe for CycK in western blot analysis.

Another potential explanation is that we captured the role of CDK12 on chromatin under cross-linking conditions. It is possible that CDK12 operates in a cyclin-independent manner when bound to chromatin, interacting with associated RNA processing factors. Cyclin-independent function of a CDK in transcription is earlier seen in CDK6 [117].

5.2.2 Distinct divergence between CDK12-mass spectrometry studies

Four previous studies have to date performed co-immunoprecipitation on human CDK12 with subsequent MS analysis in terms of studying CDK12 interactome. Bartkowiak et al. [84] immunopurified CDK12 from HeLa cell nuclear extract and found an enrichment of mRNA splicing factors, mRNA-binding proteins and nuclear speckle components.
Eifler et al. [78] immunopurified Flag-tagged CDK12 from HEK293T whole-cell lysate and found CDK12 binding to components of the EJC, as well as to auxiliary splicing factors as SRSFs and the CDCL/PRPF19 complex. Tien et al. [118] used the SK-BR-3 breast cancer cell line to precipitate 3xFLAG-CDK12 from whole cell lysate, and detected RNA splicing factors. Finally, Liang et al. [119] purified Flag epitope-tagged CDK12 from HEK293T nuclear extract and found an enrichment of splicing factors, RNA processing factors, components of the spliceosome and nuclear speckle components. Accordingly, these studies coincide with our findings as regards to similar functional properties of the proteins, as well as engagement in the same complexes.

Venn diagrams were here created (figure 19) in interest of comparing MS-obtained datasets of CDK12-bound proteins from the above-mentioned studies, to the dataset obtained in the present study. Three of the four studies were supplemented with complete list of proteins, thus available for comparison. As indicated in figure 19a-c, few proteins are shared between our dataset and theirs.

One possible reason for the small overlap between the studies, despite the similarities between the enriched functional groups of the proteins, might be due to cell-line specific properties. Of relevance, CDK12 is amplified in the breast cancer SK-BR-3 cell line [118] which might affect its interaction partners in that particular study. Moreover, different experimental approaches might influence the outcome. An aspect of interest is that neither of these studies included cross-linking of cells before co-immunoprecipitation. Hence, one cannot claim their datasets to reflect the role of CDK12 on chromatin.

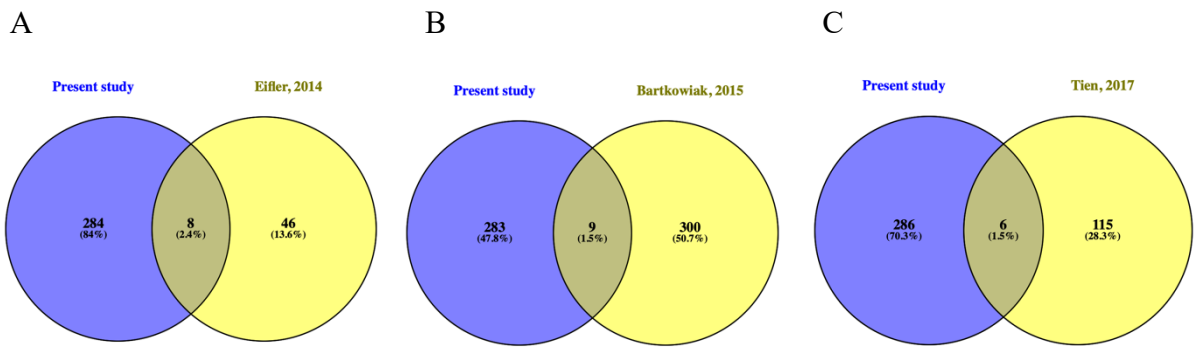


Figure 19. Venn diagram representing previous mass spectrometry analyses done on CDK12 compared to mass spectrometry data from the present study. Venn diagrams demonstrating shared and distinct proteins between MS data previously performed on CDK12, compared to MS data from the present study. Figure A represents data from Eifler et al. on CDK12 co-immunoprecipitation in whole cell lysate of HEK292 cells, figure B represents data from Bartkowiak et al. on CDK12 co-immunoprecipitation in HeLa nuclear extract, and figure C represents data from Tien et al. on CDK12 co-immunoprecipitation in whole-cell extract of SK-BR-3 cells. Venn diagrams were obtained by the use of Venny 2.1.0 software.

However, an in-depth analysis in the different MS datasets reveal similarities between them. Tien et al. [118] found CDK12 to co-immunoprecipitate with SNRPA and SNRPDG, while SNRPG was identified in the present study – all found to be building blocks of the spliceosome [120]. CDK12 was in the present study found to interact with SF3A3, while Eifler et al. [78] identified SF3B14 and Bartkowiak et al. [84] identified SF3B3 in interaction with CDK12. SF3A- and SF3B complexes are required for spliceosomal assembly on nascent RNA [121]. Moreover, splicing factors SRSF4 and 5 were identified in this study, whilst SRSF1-9 together were found to interact with CDK12 in the studies of Eifler, Bartkowiak and Tien. These observations reveal our MS data to complement previously discovered complexes with novel- and already identified subunits.

5.2.3 CDK12 interacts with mRNA processing factors

The enrichment analysis (figure 14) indicated mRNA processing factors to be among the most enriched functional groups of the CDK12-interacting proteins. The subsequent network analysis (figure 15) gave further insight into the mRNA processing factors involved with chromatin-bound CDK12. Moreover, the protein analysis performed in Enrichr (figure 16) gave a deeper understanding of the proteins involved in mRNA processing.

Proteins of interest from figure 16 that were found to interact in the same molecular pathways will, in light of CDK12, be further discussed. Protein interactions were here determined by the use of STRING and UniProt databases, as well as research papers (see in-text citations).

5.2.3.1 CDK12 interacts with transcription regulating factors

NELF-E and NELF-C/D were in this study found to interact with chromatin-bound CDK12 (figure 20). NELF-E and NELF-C/D are essential components of the NELF complex, which as initially described together with DSIF is known to negatively regulate transcription elongation by RNA Pol II. The P-TEFb/CDK9 complex is known to phosphorylate NELF-E for RNA Pol II release into productive elongation (figure 2), which further intrigues our observation of CDK12 in interaction with NELF-E, as CDK12 analogically to CDK9 is associated with transcription elongation and CTD Ser2 phosphorylation.

Bartkowiak et al. [122] recently found CDK12 to phosphorylate the NELF-E subunit in HeLa cells, but did not further elucidate their interaction. Moreover, the NELF complex was found to be lost from chromatin upon hydrogen peroxide (H_2O_2) treatment [123]. This observation is interesting in the view of our observation of CDK12 dissociating from chromatin upon heat shock treatment. The stress-induced coinciding behavior of these two transcription factors opens a door for further interest of research on the interaction between CDK12 and NELF.

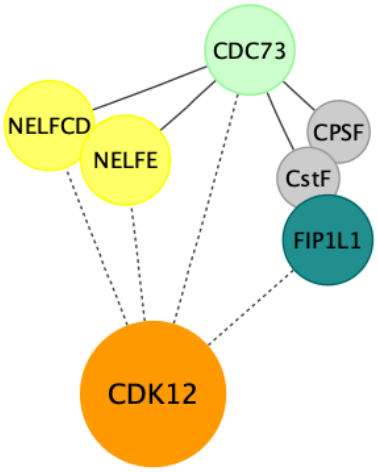


Figure 20. Interaction network representing interaction of CDK12 to transcription regulating factors. MS data revealed chromatin-bound CDK12 to interact with transcription regulation factors NELFE, NELFCD, CDC73 and FIP1L1. Dashed edges represent connections established from MS data in the present study and solid edges represent formerly determined connections. Grey nodes represent proteins of relevance which were not found to interact with CDK12-GFP in this study. The connections were determined by the use of STRING and UniProt protein databases, and already existing data from published articles (see in text citations). Figure was created in Cytoscape software.

CDK12 was in the present study further found to interact with CDC73 (figure 20). CDC73 is component of the PAF1 complex, which as previously described is required for Pol II pause release. Yu et al. [37] likewise found CDK12 to co-immunoprecipitate with CDC73, and further in their study suggested for a role of PAF1C (CDC73) in recruiting CDK12 to the transcription machinery for CTD Ser2 to be phosphorylated, for subsequent Pol II pause release to occur. NELF-C/D and NELF-E are also associated with CDC73, as both complexes are involved in the process of Pol II elongation [110].

Moreover, previous research has demonstrated knockdown of CDC73 to lead to significant shortening of transcripts through activation of intronic polyadenylation sites [124]. Hence, CDC73 is involved in regulation of alternative polyadenylation (APA). Interestingly, CDK12 is also shown to regulate APA, by globally suppressing intronic polyadenylation events, enabling the process of full-length transcripts [83]. Moreover, CDC73 is found to physically associate with cleavage and polyadenylation factor (CPSF) and cleavage stimulation factor (CstF), which are found to be global regulators of APA [125].

Interestingly, FIP1L1 was in this study co-immunoprecipitated with CDK12 (figure 20). FIP1L1 is subunit of the multiprotein complex CPSF [126]. It would be of major interest to further investigate the role of CDK12 in its interaction with CDC73 and FIP1L1 as potential interaction partners in regulation of APA.

Moreover, CPSF and CstF, as initially described, recognize and bind specific sequences of nascent mRNA, and binding of these complexes recruits other 3´end processing factors to nascent RNA. Rozenblatt-Rosen et al. [127] found the CDC73-CPSF-CstF complex to be necessary in 3´end processing. Indeed, Eifler et al. [78] concluded that CDK12, by phosphorylating CTD Ser2, possess a role in recruitment of CPSF73 and CstF64 3´end processing factors. An interesting point of view would be to investigate the role of interaction of CDK12 with CDC73 for 3´end processing to occur.

5.2.3.2 CDK12 interacts with subunits of the exon junction complex

During the last years, the exon junction complex (EJC) has emerged as a multiprotein complex of diverse functional assessments. The nuclear exon junction core complex, also termed the pre-EJC, consists of a stable tetrameric core, comprising the proteins eIF4A3,

MAGOHB, RBM8A and MLN51 [128], of which MAGOHB and RBM8A in the present study co-immunoprecipitated with CDK12 (figure 21). The EJC core is gravitated around the eIF4A3 protein, and an ATP-bound form of eIF4A3 closes around mRNA. Heterodimer MAGOHB-RBM8A connects with eIF4A3, preventing a conformational change and closes the complex on mRNA. MLN51 subsequently envelopes eIF4A3 and binds to MAGOHB, resultingly stabilizing the compounds [128].

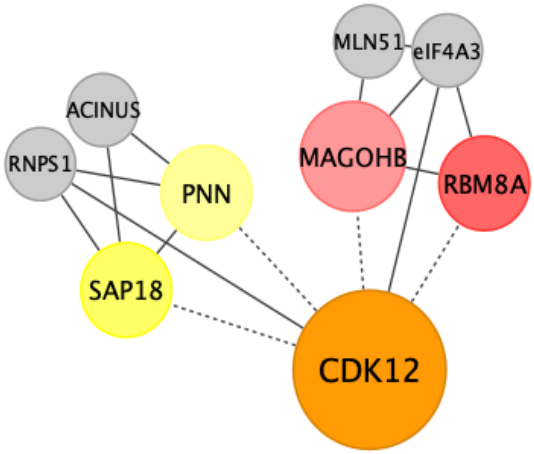


Figure 21. Interaction network representing interactions of CDK12 to components of the EJC and associated splicing factors. MS data revealed chromatin-bound CDK12 to interact with MAGOHB and RBM8A of the EJC, as well as SAP18 and PNN auxiliary splicing factors. Dashed edges represent connections established from MS data in the present study and solid edges represent formerly determined connections. Grey nodes represent proteins of relevance which were not found to interact with CDK12-GFP in this study. The connections were determined by the use of STRING and UniProt protein databases, and already existing data from published articles (see in text citations). Figure was created in Cytoscape software.

The EJC complex is interestingly found to be involved in splicing, transcription and 3´end formation. The role of the EJC in splicing was lately revealed. This complex was found to serve as a splicing enhancer to neighbor introns, thus participating to specific splicing choices around the genome [128]. During splicing, the EJC core is associated with peripheral splicing factors termed ACINUS, PNN, RNPS1 and SAP18, in which PNN and SAP18 in this study co-immunoprecipitated with CDK12. ACINUS and PNN work as scaffold proteins which binds to RNPS1, developing binding site for SAP18 for subsequent splicing to occur [128]. Depletion of components of the pre-EJC, RNPS1 or PNN all lead to exon skipping [129].

Noteworthy, both eIF4A3 and RNPS1 are found to be co-immunoprecipitated with CDK12 in the previous abovementioned MS studies performed [78, 84]. Thus, CDK12 is found to

interact with a variety of factors associated with the EJC, and our results further support the hypothesis that CDK12 is potentially involved in functional tasks of this complex.

Interestingly, Akhtar et al. [130] found that depletion of the pre-EJC components in Drosophila cells lead to genome-wide changes of the phosphorylation state of Pol II CTD, in addition to a global decrease in promoter proximal pausing. They further found the pre-EJC to associate with promoter-regions, suggesting a connection between the pre-EJC and the transcription machinery.

Moreover, they investigated the role of the pre-EJC complex in regulating Pol II during transcription, preventing premature Pol II release into elongation. In absence of the pre-EJC, the phosphorylation state of Ser2 CTD increased, and they further demonstrated that one of the roles of the EJC complex is to constrain the P-TEFb function by reducing its association with Pol II CTD. These observations are interesting in the light of CDK12, in its predicted role of phosphorylating Ser2 for productive elongation.

5.2.3.3 CDK12 interacts with subunits of the exosome complex

Our MS data further revealed interaction of chromatin-bound CDK12 with components of the exosome complex.

The human nuclear exosome core has a ring-like structure and serves as a scaffold in its association with catalytic subunits and accessory proteins [131]. The EXO9 is a six-subunit complex comprising the proteins EXOSC4, 5, 6, 7, 8 and 9. This complex is enveloped by a three-membered ring comprising the proteins EXOSC1, 2 and 3. This nine-subunit complex obtains its catalytic activity in contact with two enzymes performing enzymatic cleavage - DIS3 and EXOSC10 [131]. MS data from this study revealed EXOSC4, 5 and 7 to interact with chromatin-bound CDK12 (figure 22). Only Bartkowiak et al. [84] has previously found CDK12 to interact with components of the exosome complex.

The exosome complex is in the nucleus involved in 3´end processing of various types of RNA, in addition to being essential in RNA 3´-5´exonucleolytic activity. Foremost, the exosome complex is essential in mRNA decay: upon its depletion, aberrant pre-mRNAs accumulate in the cell [132]. More is still known in yeast than in human cells. Interestingly, nuclear RNA decay is in *S. cerevisiae* found to happen co-transcriptionally. Transcripts that

are failed to be terminated at the PAS are found to be co-transcriptionally terminated and handed to the exosome complex [133]. The underlying mechanisms of these processes in human cells have yet to emerge.

The human nuclear exosome is engaged with co-factors, which serve both as adaptors by recruitment of RNA substrates, and as activators by stimulation of RNA decay [134]. Most of these co-factors interplay with the essential RNA helicase SKIV2L2 – which in the present study was found to co-immunoprecipitate with CDK12.

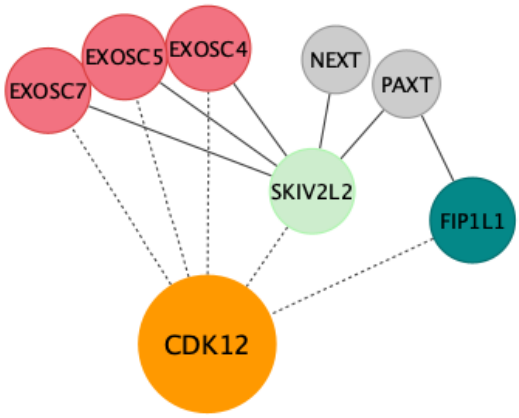


Figure 22. Interaction network representing interactions of CDK12 to components of the exosome complex. MS data revealed chromatin-bound CDK12 to interact with EXOSC4, 5 and 7, as well as SKIV2L2. Dashed edges represent connections established from MS data in the present study and solid edges represent formerly determined connections. Grey nodes represent proteins of relevance which were not found to interact with CDK12-GFP in this study. The connections were determined by the use of STRING and UniProt protein databases, and already existing data from published articles (see in text citations). Figure was created in Cytoscape software.

SKIV2L2 is in human cells involved with the exosome in nuclear processes (figure 22). It binds to the head of the exosome, bridging the transfer of target RNA strands to the exosome for subsequent RNA decay [134].

Moreover, SKIV2L2 is involved in three nuclear complexes harboring exosomal activity: nuclear exosome targeting (NEXT), human Trf4/Air2/Mtr4p polyadenylation, and poly(A)-tail exosome targeting (PAXT) [134]. Only NEXT is found to be directly linked to co-transcriptional decay and 3´end processing. The NEXT complex exhibits decay of promoter upstream transcripts and other non-coding RNAs. Furthermore, it is involved in

exosomal trimming of snoRNAs and snRNAs. Of interest, the human NEXT complex was found to interact with the cap-binding complex [135]. In *S. cerevisiae*, the NEXT complex is co-transcriptionally bridged to nascent transcripts by the help of the cap-binding complex for subsequent decay of transcripts, and recent research suggests for a similar mechanism in human cells [134].

Of notion, Wu et al. recently performed cross-linking immunoprecipitation and subsequent high-throughput sequencing (CLIP-seq) in HeLa cells, and found CPA components as the proteins most prone to bind to the PAXT complex [136]. Among these proteins, FIP1L1, which in this study was found to co-immunoprecipitate with CDK12, was found.

Hence, a role of CDK12 in co-transcriptional decay and 3´end processing through the exosome complex, in addition its interaction with the two complexes harboring exosomal activity, NEXT and PAXT, can be questioned upon.

5.3 Alterations in Pol II CTD Ser2 and Ser5 phosphorylation state after heat shock- and THZ531 treatment

The role of CDK12 in phosphorylating CTD Ser2 and Ser5 is to date discussed, as separate studies with different experimental approaches postulate various conclusions. A subject of discussion is whether CDK12 phosphorylates CTD Ser2 for productive elongation. To investigate the role of CDK12 in transcription, we performed immunofluorescence on cells treated with heat shock or THZ531.

5.3.1 Heat shock treatment decreases CTD Ser2 and Ser5 phosphorylation state

As heat shock treatment is demonstrated to disturb transcriptional programs, we utilized this method to investigate the role of CDK12 on a mechanistic level. Upon heat shock treatment, we observed a changed distribution pattern of RNA Pol II (figure 17a) and CDK12-GFP (figure 17b), as well as decreased levels of phosphorylated CTD Ser2 (figure 17c) and Ser5 (figure 17d). The decrease in Ser2-p and Ser5-p was by Student´s T-test considered significant (respectively $p < 0.05$ and $p < 0.001$). Different distribution pattern of RNA Pol II upon heat shock coincides with previous data demonstrating RNA Pol II to stall on the gene body upon elevated temperatures. The redistribution of CDK12-GFP corresponds with evidence of a known response of cells in exposure to elevated non-lethal temperatures, that is

protein aggregation (formation of stress granules) [137]. The following decrease in the phosphorylation state of CTD Ser2 and Ser5 might be an indication of CDK12 forming nuclear stress granules (figure 17b), inhibiting its catalytical activity, which in terms impairs its ability to phosphorylate Pol II CTD Ser5 and Ser2.

However, this analysis cannot exclude other factors to be involved in these alterations of CTD Ser2 and Ser5 phosphorylation states. Consequently, we further analyzed these observations in perspective of a decrease in CDK12/13 catalytic activity.

5.3.2 THZ531 treatment reduces the phosphorylation state of CTD Ser2, but not Ser5

To further investigate the heat shock-induced Ser2-p and Ser5-p decrease, we conducted THZ531-induced inhibition of CDK12/13. Previous studies have confirmed the efficiency of THZ531 in inhibiting CDK12 catalytical activity [74, 75]. This drug is also found to diminish levels of global Ser2 phosphorylation, as well as to affect productive RNA Pol II elongation [75]. As THZ531 inhibits the catalytic activity of both CDK12 and CDK13, a natural occasion would have been to likewise investigate CDK13-GFP. As we were not able to detect CDK13-GFP by western blot analysis, immunofluorescent staining of CDK13-GFP was not included in this study.

A speckled pattern of Pol II was observed upon THZ531-treatment (figure 18a), as well as a decrease in CTD Ser2 phosphorylation levels (figure 18b). On the contrary, an unexplainedly increase in Ser5 phosphorylation levels was observed (figure 18c). The decrease in Ser2-p was by Student´s T-test considered significant ($p < 0.01$), while the increase in Ser5-p was considered non-significant ($p > 0.05$). As CDK12 is suggested to phosphorylate Pol II CTD for productive elongation, the speckled pattern of RNA Pol II raised the question if RNA Pol II acts in an aberrant manner in absence of CDK12. However, as THZ531 inhibits both CDK12 and CDK13 catalytic activity, this analysis cannot exclude the phosphorylation activity of CDK13. The decreased levels of CTD Ser2 phosphorylation state coincides with the possibility of an aberrant behavior of RNA Pol II upon CDK12/13-inhibition, and moreover corresponds with previous data demonstrating a decrease in Ser2 phosphorylation state after SR-4835-mediated dual inhibition of CDK12/13 [73].

5.4 Future perspectives

Foremost, future studies would benefit from performing ChIP-MS in triplicate samples to gain insight into intra-sample variations and improve the statistical significance of the experiment. In order to further investigate the interactome of chromatin-bound CDK12-GFP, interaction partners of interest could henceforth be validated. Validation could be conducted by CDK12-GFP co-immunoprecipitation followed by western blot analyses with specific primary antibodies against proteins of interest. On basis of the proteins which were identified in interaction with CDK12-GFP in the present study, NELF-E and NELF-C/D, as well as CDC73 could have been interesting alternatives, as discussed above.

Regarding the phosphorylation state of RNA Pol II phosphorylated CTD species, additional validation is necessary. For further examination of Pol II occupancy on gene body upon heat shock- and THZ531 treatment, ChIP-seq analysis for RNA Pol II might be an informative approach. Moreover, validation of Ser2- and Ser5 phosphorylation states upon treatment could have been assessed in analyses such as western blotting or MS.

To further investigate catalytic activity of CDK12 or CDK13 as separate kinases, future studies could benefit from sequence-specific gene silencing. An RNA interference (RNAi) approach could have been assessed, with short hairpin RNA (shRNA) engineered to separately suppress expression of CDK12 and CDK13.

6. Concluding remarks

The aim of this study was to elucidate the roles of chromatin-bound CDK12 and CDK13 expressed at endogenous levels. As recombinant CDK13-GFP was not detectable in protein analyses, further experiments focused mainly on CDK12-GFP.

The MS data revealed novel and already identified interaction partners of chromatin-bound CDK12. CDK12 was found to disassociate from chromatin after 30 minutes of heat shock treatment. The interactome of chromatin-bound CDK12 under non-heat shock conditions was studied, and supported previous research pointing at CDK12 as a factor involved in transcription and RNA processing events. However, as these data were not retrieved from triplicate samples, firm conclusions cannot be drawn, and further validation is required.

To acquire knowledge on its role in transcription, the current theory of CDK12 as CTD Ser2 and Ser5 kinase was questioned. As (i) heat shock lead to redistribution of CDK12-GFP, (ii) nuclear speckles emerged both after heat shock- and THZ531 treatment in RNA Pol II stained cells, and (iii) both heat shock treatment and THZ531-induced inhibition of CDK12/13 resulted in a decrease in CTD Ser2 phosphorylation state, it is enticing to speculate that these observations together underpin the growing body of evidence pointing at CDK12 in the role as major CTD Ser2 kinase for productive elongation. However, these experimental approaches cannot exclude the catalytic activity of CDK13. Thus, our observations underpin previous research demonstrating a decrease in Ser2 phosphorylation state after inhibition of CDK12/13. Moreover, the roles of CDK12 and CDK13 in CTD Ser5 phosphorylation remains a matter of speculation.

In its entirety, these observations will together give direction for future research.

7. References

1. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nat Rev Genet, 2012. **13**(10): p. 720-31.
2. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.
3. Sharifi, S. and H. Bierhoff, *Regulation of RNA Polymerase I Transcription in Development, Disease, and Aging.* Annu Rev Biochem, 2018. **87**: p. 51-73.
4. Bernecky, C., et al., *Structure of transcribing mammalian RNA polymerase II.* Nature, 2016. **529**(7587): p. 551-4.
5. Harlen, K.M. and L.S. Churchman, *The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain.* Nat Rev Mol Cell Biol, 2017. **18**(4): p. 263-273.
6. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. **292**(5523): p. 1863-76.
7. Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* EMBO J, 2004. **23**(20): p. 4051-60.
8. Mourao, A., et al., *Structure and RNA recognition by the snRNA and snoRNA transport factor PHAX.* RNA, 2010. **16**(6): p. 1205-16.
9. Dieci, G., et al., *The expanding RNA polymerase III transcriptome.* Trends Genet, 2007. **23**(12): p. 614-22.
10. Wilkinson, M.E., C. Charenton, and K. Nagai, *RNA Splicing by the Spliceosome.* Annu Rev Biochem, 2019.
11. Zaborowska, J., S. Egloff, and S. Murphy, *The pol II CTD: new twists in the tail.* Nat Struct Mol Biol, 2016. **23**(9): p. 771-7.
12. Allison, L.A., et al., *The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function.* Mol Cell Biol, 1988. **8**(1): p. 321-9.
13. Hirose, Y., R. Tacke, and J.L. Manley, *Phosphorylated RNA polymerase II stimulates pre-mRNA splicing.* Genes Dev, 1999. **13**(10): p. 1234-9.
14. Hirose, Y. and J.L. Manley, *RNA polymerase II is an essential mRNA polyadenylation factor.* Nature, 1998. **395**(6697): p. 93-6.
15. Kumar, A., et al., *Mechanistic insights into mRNA 3'-end processing.* Curr Opin Struct Biol, 2019. **59**: p. 143-150.
16. Pillutla, R.C., et al., *Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes.* J Biol Chem, 1998. **273**(34): p. 21443-6.
17. Lu, L., et al., *Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II.* Biochemistry, 2016. **55**(7): p. 1149-58.
18. Kubicek, K., et al., *Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1.* Genes Dev, 2012. **26**(17): p. 1891-6.
19. Schuller, R., et al., *Heptad-Specific Phosphorylation of RNA Polymerase II CTD.* Mol Cell, 2016. **61**(2): p. 305-14.
20. Heidemann, M., et al., *Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription.* Biochim Biophys Acta, 2013. **1829**(1): p. 55-62.
21. Meininghaus, M., et al., *Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription.* J Biol Chem, 2000. **275**(32): p. 24375-82.
22. Nojima, T., et al., *Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing.* Cell, 2015. **161**(3): p. 526-540.

23. Buratowski, S., *Progression through the RNA polymerase II CTD cycle.* Mol Cell, 2009. **36**(4): p. 541-6.
24. Lu, H., et al., *The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10004-8.
25. Portz, B. and J. Shorter, *Switching Condensates: The CTD Code Goes Liquid.* Trends Biochem Sci, 2020. **45**(1): p. 1-3.
26. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
27. Flanagan, P.M., et al., *A mediator required for activation of RNA polymerase II transcription in vitro.* Nature, 1991. **350**(6317): p. 436-8.
28. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 155-66.
29. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.
30. Yamada, T., et al., *P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation.* Mol Cell, 2006. **21**(2): p. 227-37.
31. Ghosh, A., S. Shuman, and C.D. Lima, *Structural insights to how mammalian capping enzyme reads the CTD code.* Mol Cell, 2011. **43**(2): p. 299-310.
32. Glover-Cutter, K., et al., *TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II.* Mol Cell Biol, 2009. **29**(20): p. 5455-64.
33. Yamaguchi, Y., et al., *NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation.* Cell, 1999. **97**(1): p. 41-51.
34. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs.* Genes Dev, 1998. **12**(3): p. 343-56.
35. Zaborowska, J., N.F. Isa, and S. Murphy, *P-TEFb goes viral.* Bioessays, 2016. **38 Suppl 1**: p. S75-85.
36. Vos, S.M., et al., *Structure of activated transcription complex Pol II-DSIF-PAF-SPT6.* Nature, 2018. **560**(7720): p. 607-612.
37. Yu, M., et al., *RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II.* Science, 2015. **350**(6266): p. 1383-6.
38. Saldi, T., et al., *Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing.* J Mol Biol, 2016. **428**(12): p. 2623-2635.
39. Saldi, T., N. Fong, and D.L. Bentley, *Transcription elongation rate affects nascent histone pre-mRNA folding and 3′ end processing.* Genes Dev, 2018. **32**(3-4): p. 297-308.
40. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.* Nature, 1997. **385**(6614): p. 357-61.
41. Galloway, A. and V.H. Cowling, *mRNA cap regulation in mammalian cell function and fate.* Biochim Biophys Acta Gene Regul Mech, 2019. **1862**(3): p. 270-279.
42. Aregger, M. and V.H. Cowling, *Human cap methyltransferase (RNMT) N-terminal non-catalytic domain mediates recruitment to transcription initiation sites.* Biochem J, 2013. **455**(1): p. 67-73.
43. Inesta-Vaquera, F., et al., *DHX15 regulates CMTR1-dependent gene expression and cell proliferation.* Life Sci Alliance, 2018. **1**(3): p. e201800092.

44. Ruegsegger, U., K. Beyer, and W. Keller, *Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors.* J Biol Chem, 1996. **271**(11): p. 6107-13.
45. Davidson, L., L. Muniz, and S. West, *3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells.* Genes Dev, 2014. **28**(4): p. 342-56.
46. Fusby, B., et al., *Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation.* Mol Cell Biol, 2016. **36**(2): p. 295-303.
47. Scorilas, A., *Polyadenylate polymerase (PAP) and 3' end pre-mRNA processing: function, assays, and association with disease.* Crit Rev Clin Lab Sci, 2002. **39**(3): p. 193-224.
48. Larochelle, M., J. Hunyadkurti, and F. Bachand, *Polyadenylation site selection: linking transcription and RNA processing via a conserved carboxy-terminal domain (CTD)-interacting protein.* Curr Genet, 2017. **63**(2): p. 195-199.
49. Licatalosi, D.D., et al., *Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II.* Mol Cell, 2002. **9**(5): p. 1101-11.
50. Ni, Z., et al., *Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes.* Mol Cell, 2004. **13**(1): p. 55-65.
51. Lemay, J.F., et al., *The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection.* Genes Dev, 2016. **30**(13): p. 1558-72.
52. Tian, B., et al., *A large-scale analysis of mRNA polyadenylation of human and mouse genes.* Nucleic Acids Res, 2005. **33**(1): p. 201-12.
53. Misteli, T. and D.L. Spector, *RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo.* Mol Cell, 1999. **3**(6): p. 697-705.
54. Gu, B., D. Eick, and O. Bensaude, *CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo.* Nucleic Acids Res, 2013. **41**(3): p. 1591-603.
55. Proudfoot, N.J., *Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut.* Science, 2016. **352**(6291): p. aad9926.
56. Zhang, Z., J. Fu, and D.S. Gilmour, *CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11.* Genes Dev, 2005. **19**(13): p. 1572-80.
57. Kaneko, S., et al., *The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination.* Genes Dev, 2007. **21**(14): p. 1779-89.
58. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II.* Nature, 2004. **432**(7016): p. 517-22.
59. Chou, J., et al., *Transcription-Associated Cyclin-Dependent Kinases as Targets and Biomarkers for Cancer Therapy.* Cancer Discov, 2020. **10**(3): p. 351-370.
60. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.
61. Elmlund, H., et al., *The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15788-93.
62. Larochelle, S., et al., *Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II.* Nat Struct Mol Biol, 2012. **19**(11): p. 1108-15.
63. Ebmeier, C.C., et al., *Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications.* Cell Rep, 2017. **20**(5): p. 1173-1186.

64. Bacon, C.W. and I. D'Orso, *CDK9: a signaling hub for transcriptional control.* Transcription, 2019. **10**(2): p. 57-75.
65. Sanso, M., et al., *P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates.* Genes Dev, 2016. **30**(1): p. 117-31.
66. Ghamari, A., et al., *In vivo live imaging of RNA polymerase II transcription factories in primary cells.* Genes Dev, 2013. **27**(7): p. 767-77.
67. Hsin, J.P., A. Sheth, and J.L. Manley, *RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing.* Science, 2011. **334**(6056): p. 683-6.
68. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. **5**: p. 3505.
69. Kveta Pilarova, J.H., Dalibor Blazek, *CDK12: cellular functions and therapeutic potential of versatile player in cancer.* NAR Cancer, 2020. **Volume 2**(Issue 1).
70. Greifenberg, A.K., et al., *Structural and Functional Analysis of the Cdk13/Cyclin K Complex.* Cell Rep, 2016. **14**(2): p. 320-31.
71. Ko, T.K., E. Kelly, and J. Pines, *CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles.* J Cell Sci, 2001. **114**(Pt 14): p. 2591-603.
72. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1.* Genes Dev, 2010. **24**(20): p. 2303-16.
73. Quereda, V., et al., *Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer.* Cancer Cell, 2019. **36**(5): p. 545-558 e7.
74. Krajewska, M., et al., *CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation.* Nat Commun, 2019. **10**(1): p. 1757.
75. Zhang, T., et al., *Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors.* Nat Chem Biol, 2016. **12**(10): p. 876-84.
76. Cheng, S.W., et al., *Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II.* Mol Cell Biol, 2012. **32**(22): p. 4691-704.
77. Blazek, D., et al., *The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes.* Genes Dev, 2011. **25**(20): p. 2158-72.
78. Eifler, T.T., et al., *Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts.* Mol Cell Biol, 2015. **35**(2): p. 468-78.
79. Galganski, L., M.O. Urbanek, and W.J. Krzyzosiak, *Nuclear speckles: molecular organization, biological function and role in disease.* Nucleic Acids Res, 2017. **45**(18): p. 10350-10368.
80. Even, Y., et al., *CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo.* J Cell Biochem, 2006. **99**(3): p. 890-904.
81. Kohoutek, J. and D. Blazek, *Cyclin K goes with Cdk12 and Cdk13.* Cell Div, 2012. **7**: p. 12.
82. Mortillaro, M.J., et al., *A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix.* Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8253-7.
83. Dubbury, S.J., P.L. Boutz, and P.A. Sharp, *CDK12 regulates DNA repair genes by suppressing intronic polyadenylation.* Nature, 2018. **564**(7734): p. 141-145.
84. Bartkowiak, B. and A.L. Greenleaf, *Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex.* J Biol Chem, 2015. **290**(3): p. 1786-95.
85. Joshi, P.M., et al., *Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to*

cisplatin and poly(ADP-ribose) polymerase inhibitors. J Biol Chem, 2014. **289**(13): p. 9247-53.
86. Kim, H.E., et al., *Frequent amplification of CENPF, GMNN and CDK13 genes in hepatocellular carcinomas.* PLoS One, 2012. **7**(8): p. e43223.
87. Ng, C.K., et al., *The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer.* J Pathol, 2012. **226**(5): p. 703-12.
88. Popova, T., et al., *Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large Tandem Duplications.* Cancer Res, 2016. **76**(7): p. 1882-91.
89. Wu, Y.M., et al., *Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer.* Cell, 2018. **173**(7): p. 1770-1782 e14.
90. Dong, X., et al., *CDK13 RNA Over-Editing Mediated by ADAR1 Associates with Poor Prognosis of Hepatocellular Carcinoma Patients.* Cell Physiol Biochem, 2018. **47**(6): p. 2602-2612.
91. Sakurai, M., et al., *A biochemical landscape of A-to-I RNA editing in the human brain transcriptome.* Genome Res, 2014. **24**(3): p. 522-34.
92. Bostwick, B.L., et al., *Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders.* Genome Med, 2017. **9**(1): p. 73.
93. Lucey, B.P., W.A. Nelson-Rees, and G.M. Hutchins, *Henrietta Lacks, HeLa cells, and cell culture contamination.* Arch Pathol Lab Med, 2009. **133**(9): p. 1463-7.
94. Dalby, B., et al., *Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications.* Methods, 2004. **33**(2): p. 95-103.
95. Basu, S., et al., *Purification of specific cell population by fluorescence activated cell sorting (FACS).* J Vis Exp, 2010(41).
96. Cardiello, J.F., J.A. Goodrich, and J.F. Kugel, *Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination.* Mol Cell Biol, 2018. **38**(18).
97. Baghirova, S., et al., *Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells.* MethodsX, 2015. **2**: p. 440-5.
98. Nelson, J.D., O. Denisenko, and K. Bomsztyk, *Protocol for the fast chromatin immunoprecipitation (ChIP) method.* Nat Protoc, 2006. **1**(1): p. 179-85.
99. Taylor, S.C. and A. Posch, *The design of a quantitative western blot experiment.* Biomed Res Int, 2014. **2014**: p. 361590.
100. Arya, M., et al., *Basic principles of real-time quantitative PCR.* Expert Rev Mol Diagn, 2005. **5**(2): p. 209-19.
101. Baker, M., *Cellular imaging: Taking a long, hard look.* Nature, 2010. **466**(7310): p. 1137-40.
102. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
103. Maity, B., D. Sheff, and R.A. Fisher, *Immunostaining: detection of signaling protein location in tissues, cells and subcellular compartments.* Methods Cell Biol, 2013. **113**: p. 81-105.
104. Kustatscher, G., et al., *Chromatin enrichment for proteomics.* Nat Protoc, 2014. **9**(9): p. 2090-9.
105. Mohammed, H., et al., *Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes.* Nat Protoc, 2016. **11**(2): p. 316-26.
106. Oliveros, J.C. *VENNY. An interactive tool for comparing lists with Venn Diagrams*. 2007; Available from: http://bioinfogp.cnb.csic.es/tools/venny/.

107. Zhou, Y., et al., *Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.* Nat Commun, 2019. **10**(1): p. 1523.
108. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
109. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
110. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
111. Velichko, A.K., et al., *Mechanisms of heat shock response in mammals.* Cell Mol Life Sci, 2013. **70**(22): p. 4229-41.
112. Mahat, D.B., et al., *Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation.* Mol Cell, 2016. **62**(1): p. 63-78.
113. Hollenback, S.M., S. Lyman, and J. Cheng, *Recombineering-based procedure for creating BAC transgene constructs for animals and cell lines.* Curr Protoc Mol Biol, 2011. **Chapter 23**: p. Unit 23 14.
114. Xing, S., et al., *Techniques for the Analysis of Protein-Protein Interactions in Vivo.* Plant Physiol, 2016. **171**(2): p. 727-58.
115. Makoukji, J., *Chromatin Immunoprecipitation Assay for Analyzing Transcription Factor Activity at the Level of Peripheral Myelin Gene Promoters.* Methods Mol Biol, 2019. **2011**: p. 647-658.
116. Active Motif, E.E.R., *Complete Guide to Sonication of Chromatin for ChIP Assays.* 2020.
117. Fujimoto, T., et al., *Cdk6 blocks myeloid differentiation by interfering with Runx1 DNA binding and Runx1-C/EBPalpha interaction.* EMBO J, 2007. **26**(9): p. 2361-70.
118. Tien, J.F., et al., *CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion.* Nucleic Acids Res, 2017. **45**(11): p. 6698-6716.
119. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing.* Mol Cell Biol, 2015. **35**(6): p. 928-38.
120. *UniProtKB - P09012 (SNRPA_HUMAN)*. Available from: https://www.uniprot.org/uniprot/P09012.
121. Wang, K., et al., *A U2-snRNP-independent role of SF3b in promoting mRNA export.* Proc Natl Acad Sci U S A, 2019. **116**(16): p. 7837-7846.
122. Bartkowiak, B., et al., *CDK12 Activity-Dependent Phosphorylation Events in Human Cells.* Biomolecules, 2019. **9**(10).
123. Nilson, K.A., et al., *Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome.* Nucleic Acids Res, 2017. **45**(19): p. 11088-11105.
124. Yang, Y., et al., *PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation.* PLoS Genet, 2016. **12**(1): p. e1005794.
125. Yao, C., et al., *Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation.* Proc Natl Acad Sci U S A, 2012. **109**(46): p. 18773-8.
126. Kaufmann, I., et al., *Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase.* EMBO J, 2004. **23**(3): p. 616-26.
127. Rozenblatt-Rosen, O., et al., *The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 755-60.
128. Le Hir, H., J. Sauliere, and Z. Wang, *The exon junction complex as a node of post-transcriptional networks.* Nat Rev Mol Cell Biol, 2016. **17**(1): p. 41-54.

129. Leung, C.S. and T.L. Johnson, *The Exon Junction Complex: A Multitasking Guardian of the Transcriptome.* Mol Cell, 2018. **72**(5): p. 799-801.
130. Liu, M., et al., *The exon junction complex regulates the splicing of cell polarity gene dlg1 to control Wingless signaling in development.* Elife, 2016. **5**.
131. Januszyk, K., E.M. Weick, and C.D. Lima, *Reconstitution of the Human Nuclear RNA Exosome.* Methods Mol Biol, 2020. **2062**: p. 467-489.
132. Kilchert, C., S. Wittmann, and L. Vasiljeva, *The regulation and functions of the nuclear RNA exosome complex.* Nat Rev Mol Cell Biol, 2016. **17**(4): p. 227-39.
133. Colin, J., et al., *Roadblock termination by reb1p restricts cryptic and readthrough transcription.* Mol Cell, 2014. **56**(5): p. 667-80.
134. Schmid, M. and T.H. Jensen, *The Nuclear RNA Exosome and Its Cofactors.* Adv Exp Med Biol, 2019. **1203**: p. 113-132.
135. Andersen, P.R., et al., *The human cap-binding complex is functionally connected to the nuclear RNA exosome.* Nat Struct Mol Biol, 2013. **20**(12): p. 1367-76.
136. Wu, G., et al., *A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome.* Cell Rep, 2020. **30**(7): p. 2387-2401 e5.
137. Woolcock, K., *Reversible aggregation after heat shock.* Nat Struct Mol Biol, 2015. **22**(10): p. 758.

Appendix

Section A

Reagents, kits, consumables, instruments and softwares

Table S1. Reagents, producers and catalog numbers

Reagent	Producer	Catalog nr.
Dulbecco´s Modified Eagle´s Medium (DMEM) – high glucose	Sigma-Aldrich®, Darmstadt, Germany	D5671
Fetal bovine serum (FBS)	Sigma-Aldrich®, Darmstadt, Germany	F7524
100x Penicillin-Streptomycin	Thermo Fisher Scientific, Waltham, MA, USA	15140-22
0,05% Trypsin Ethylenediaminetetraacetic acid (EDTA)	Thermo Fisher Scientific, Waltham, MA, USA	25300-054
Phosphate-buffered saline (PBS)	Oslo Universitetssykehus, avdeling for mikrobiologi	-
10 mM THZ531	MedChemExpress	HY-103618
HeLa Kyoto	Earlier available in the group, Riken, Japan	-
CDK9/12/13-BAC DNA	Emd Millipore, MA, USA	-
Lipofectamine™ 2000 Transfection Reagent	Thermo Fisher Scientific, Waltham, MA, USA	11668027
Opti-Minimal Essential Medium™ (MEM)	Thermo Fisher Scientific, Waltham, MA, USA	31985062
0.4% Trypan Blue Stain	Thermo Fisher Scientific, Waltham, MA, USA	T10282
Ponceau S Solution	Sigma-Aldrich®, Darmstadt, Germany	P7170
>40% Tween® 20	Sigma-Aldrich®, Darmstadt, Germany	P2287
Coomassie Brilliant Blue G 250	Sigma-Aldrich®, Darmstadt, Germany	6104-58-1
Skim Milk Powder	Sigma-Aldrich®, Darmstadt, Germany	70166
1M Tris pH 7.5	Oslo Universitetssykehus, avdeling for mikrobiologi	-
5M NaCl	Oslo Universitetssykehus, avdeling for mikrobiologi	-
10x Tris-Glycine SDS Running Buffer	Thermo Fisher Scientific, Waltham, MA, USA	LC2675
20x NuPage® MOPS SDS Running Buffer	Thermo Fisher Scientific, Waltham, MA, USA	NP0001-02
IGEPAL® CA-630	Sigma-Aldrich®, Darmstadt, Germany	I3021
1 mM EDTA	Oslo Universitetssykehus, avdeling for mikrobiologi	-
5M MgCl2	Oslo Universitetssykehus, avdeling for mikrobiologi	-
60% Glycerol	Oslo Universitetssykehus, avdeling for mikrobiologi	-
50 mg/mL Geneticin	Thermo Fisher Scientific, Waltham, MA, USA	10131-027

Dimethyl sulfoxide (DMSO)	Sigma-Aldrich®, Darmstadt, Germany	D8418
95% Paraformaldehyde	Sigma-Aldrich®, Darmstadt, Germany	158127
Triton™ X-100	Sigma-Aldrich®, Darmstadt, Germany	T8787
Bovine serum albumin (BSA), Cohn fraction V, protease free	Saween & Werner, Limhamn, Sweden	B2000-1
Vectashield HardSet Antifade Mounting Medium with Dapi	Vector Laboratories, CA, USA	H-1200-NB
Ethylene glycol tetraacetic acid (EGTA)	Bioworld, OH, USA	50-255-956
>97% Sodium deoxycholate	Sigma-Aldrich®, Darmstadt, Germany	D6750
>98% N-lauroylsarcosine	Sigma-Aldrich®, Darmstadt, Germany	61739
>99.5% HEPES	Sigma-Aldrich®, Darmstadt, Germany	H3375
>99% LiCl	Sigma-Aldrich®, Darmstadt, Germany	L4408
1M KCl	Oslo Universitetssykehus, avdeling for mikrobiologi	-
20% SDS (sodium dodecyl sulphate)	Oslo Universitetssykehus, avdeling for mikrobiologi	-
98% UREA	Sigma-Aldrich®, Darmstadt, Germany	U-5378
100% UltraPure™ Agarose	Thermo Fisher Scientific, Waltham, MA, USA	16500-500
50x TAE	Oslo Universitetssykehus, avdeling for mikrobiologi	-
99% Glycine	Sigma-Aldrich®, Darmstadt, Germany	G8898
Protein Asssay Dye Reagent Concentrate	Bio-Rad Laboratories, Hercules, CA, USA	5000006
PowerUp™ SYBR™ Green Master Mix	Thermo Fisher Scientific, Waltham, MA, USA	A257776
RNAse Free H_2O	QIAGEN, Hilden, Germany	
50 mM 2-Mercaptoethanol	Thermo Fisher Scientific, Waltham, MA, USA	31350-010
SYBR™ Safe DNA Gel Stain	Thermo Fisher Scientific, Waltham, MA, USA	S33102
cOmplete, EDTA-free protease inhibitor cocktail tablets	Krackeler Scientific, Albany, NY, USA	11873580001
ROCHE cOmplete™ Protease Inhibitor Cocktail, Tablets provided in EASYpacks	Millipore Sigma, MA, USA	4693116001
Benzonase® Nuclease	Merch Millipore, MA, USA	70664-3
1,4-dithiothreitol (DTT)	Saween & Werner, Limhamn, Sweden	D1200-25
Leupeptin (LP)	Sigma-Aldrich®, Darmstadt, Germany	L0649
Aprotinin (AP)	Sigma-Aldrich®, Darmstadt, Germany	A4529
Phenylmethanesulfonyl (PMSF)	Sigma-Aldrich®, Darmstadt, Germany	P7626

Table S2. Kits, producers and catalog numbers

Kit	Producer	Catalog nr.
RNeasy Plus Mini Kit	QIAGEN, Hilden, Germany	72134
SuperSignal™ West Pico PLUS Chemiluminiscent Substrate	Thermo Fisher Scientific, Waltham, MA, USA	34580
High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific, Waltham, MA, USA	4368814
QIAquick PCR Purification Kit	QIAGEN, Hilden, Germany	28106

Table S3. Consumables, producers and catalog numbers.

Consumables	Producer	Catalog nr.
Novex™ Wedgewell™ 6% Tris-Glycine Gel	Thermo Fisher Scientific, Waltham, MA, USA	XP00062BOX
Novex™ Wedgewell™ 4-12%, Tris-Glycine Gel	Thermo Fisher Scientific, Waltham, MA, USA	XP04125BOX
Bolt™ Bis-Tris Plus, 8% Mini Protein Gel	Thermo Fisher Scientific, Waltham, MA, USA	NW00082BOX
Trans-Blot® Turbo™ Mini PVDF Membrane, 0.2 µm	Bio-Rad Laboratories, Hercules, CA, USA	1704156
Precision Plus Protein™ Dual Color Standards	Bio-Rad Laboratories, Hercules, CA, USA	1610374
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific, Waltham, MA, USA	SM0331
Protein Assay Dye Reagent Concentrate	Bio-Rad Laboratories, Hercules, CA, USA	5000006
GFP-Trap Magnetic Agarose affinity beads	Chromotek	gtma-10
Tube 50mL	LABORIMPEX, Nümbrecht, Germany	62.547.254
Tube 15mL	LABORIMPEX, Nümbrecht, Germany	62.553.042
Falcon® Tissue Culture Dish, 150 mm	Corning, NY, USA	353025
Falcon® Tissue Culture Dish, 100 mm	Corning, NY, USA	10212951
Falcon® Tissue Culture Dish, 60 mm	Corning, NY, USA	353004
Countess™ Cell Counting Chamber Slides	Thermo Fisher Scientific, Waltham, MA, USA	C10228
Cover slips	VWR International, Pennsylvania, USA	631-1578P
Superfrost™ Microscope Slides	Thermo Fisher Scientific, Waltham, MA, USA	12372098
Standard Disposable Polystyrene Cuvettes	Bio-Rad Laboratories, Hercules, CA, USA	2239950
1,5 mL Bioruptor Pico Microtubes	Diagenode, Liège, Belgium	C30010016
Corning™ Falcon™ Test Tubes with Cell Strainer Snap Cap	Thermo Fisher Scientific, Waltham, MA, USA	08-771-23
CytoSoft® Imaging 96-well Plate	Advanced BioMatrix, San Diego, CA, USA	5255
Nunclon™ Sphera™ 24-well Plate	Thermo Fisher Scientific, Waltham, MA, USA	174930
MicroAmp™ Optical Adhesive Film	Thermo Fisher Scientific, Waltham, MA, USA	10095714
MicroAmp® Fast 96-well Reaction Plate	Thermo Fisher Scientific, Waltham, MA, USA	4346907
245 x 245cm Square BioAssay Dishes	Thermo Fisher Scientific, Waltham, MA, USA	166508
HulaMixer™ Sample Mixer	Thermo Fisher Scientific, Waltham, MA, USA	15920D
Grant Bio PCMT Thermoskaker	Keison Produces	-

Table S4. Equipments and instruments

Equipment and instruments	Producer	Catalog nr.
Smart Spec Plus Spectrophotometer	Bio-Rad Laboratories, Hercules, CA, USA	170-2525EDU
StepOnePlus™ Real-Time PCR System.	Thermo Fisher Scientific, Waltham, MA, USA	4376600
Countless™ II FL Automated Cell Counter	Thermo Fisher Scientific, Waltham, MA, USA	AMQAF1000
Gel Doc XR+ Documentation System	Bio-Rad Laboratories, Hercules, CA, USA	1708195
Bioruptor® Pico sonication device	Diagenode, Liège, Belgium	B01060010
Kontes Dounce Homogenizer	Sigma-Aldrich®, Darmstadt, Germany	D8938
MegaFuge 1.0R refrigerated centrifuge	Heraeus Instruments	LV40799952
Microcentrifuge 5424 R	Eppendorf®, Hamburg, Germany	Z722960
XCell *SureLock* Mini-Cell™	Thermo Fisher Scientific, Waltham, MA, USA	EI0001
Trans-Blot® Turbo™ Transfer System	Bio-Rad Laboratories, Hercules, CA, USA	1704150EDU
NanoDrop™ Spectrophotometer	Thermo Fisher Scientific, Waltham, MA, USA	13-400-518
Applied Biosystems™ 2720 Thermal Cycler	Thermo Fisher Scientific, Waltham, MA, USA	12313653
ImageXpress Micro Confocal	Molecular Devices, San Jose, CA, USA	-
Leica TCS SP8 gSTED microscopy	Leica Microsystems, Santa Clara, CA, USA	-

Table S5. Softwares and developers

Sofware	Developer
Enrichr	Ma´ayan et al., Department of Pharmacology and Systems Therapeutics, New York, 10029, USA
Metascape	Chanda et al., Genomics Institute of the Novartis Research Foundation, San Diego, CA, 92121, USA
Venny 2.1.0	Oliveros et al., Centro Nacional de Biotecnologia, 28049 Madrid, Spania
Cytoscape	Ideker et al., Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA
Fiji/Image J	Rasband, National Institutes of Health (NIH), Maryland, USA
LASX software	Leica Microsystems, Santa Clara, CA, USA
Microsoft® Excel	Microsoft, Redmond, WA, USA

Section B

Buffers

Cell lysis and immunoprecipitation

High salt (HS) buffer

- 50 mM Tris-HCl
- 300 mM NaCl
- 0.5% IGEPAL CA-630
- 1 mM EDTA
- 1 mM DTT
- 0,1 µM phenylmethanesulfonyl (PMSF)
- 1 µg/ml aprotinin (AP)
- 1 µg/ml leupeptin (LP)

Washing buffer

- 50 mM Tris-HCl
- 150 mM NaCl
- 1 mM EDTA
- 0.5% IGEPAL CA-630

Cellular fractionation

Buffer A

- 10 mM Tris-Cl, pH 7.9
- 10 mM KCl
- 1.5 mM $MgCl_2$
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

Buffer B

- 20 mM Tris-Cl, pH 7.9
- 300 mM NaCl
- 2 mM $MgCl_2$
- 0.2 mM EDTA
- 0.5 mM DTT
- 20% glycerol
- 0.1% IGEPAL CA-630
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

Western blotting

20x TBS

- 200 mL 1M Tris pH 7.5
- 300 mL 5M NaCl

1x TBS-T

- 25 mL 20x TBS
- 475 mL dH_2O
- 500 µL Tween®20

1x Tris/Glycine/SDS buffer
- 100 mL 10 x Tris/Glycine/SDS buffer
- 900 mL MQ-H_2O

Cross-linking

3.7% formaldehyde solution
- 3.7g Paraformaldehyde powder
- Up to 100 mL 1x PBS

2M glycine
- 75g Glycine powder
- Up to 500 mL MQ-H_2O

Chromatin enrichment for proteomics (ChEP)

Lysis buffer
- 25 mM Tris, pH 7.4
- 0.1% Triton X-100
- 85 mM KCl
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

SDS-buffer
- 50 mM Tris pH 7.4
- 10 mM EDTA
- 4% SDS
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

UREA-buffer
- 10 mM Tris, pH 7.4
- 1 mM EDTA
- 8 M UREA powder

Storage-buffer
- 10 mM Tris, pH 7.4
- 1 mM EDTA
- 25 mM NaCl
- 10% glycerol
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

Chromatin immunoprecipitation (ChIP)

Buffer 1 (50 mL)
- 1 M Tris-HCl, pH 8
- 200 mM NaCl
- 1 mM EDTA
- 0.5 mM EGTA
- 0.1 µM PMSF
- 1 µg/ml AP

- 1 µg/ml LP
- Up to 50 mL MQ-H_2O
- 1x cOmplete, EDTA-free protease inhibitor cocktail tablet

Buffer 2 (20 mL)
- 10 mM Tris-HCl, pH 8
- 100 mM NaCl
- 1 mM EDTA
- 0.5 mM EGTA
- 0.1% sodium deoxycholat
- 0.5% N-lauroylsarcosine
- Up to 20 mL MQ-H_2O
- 1x cOmplete, EDTA-free protease inhibitor cocktail tablet
- 2x ROCHE cOmplete™ Protease Inhibitor Cocktail, Tabletsprovided in EASYpacks

RIPA buffer (20 mL)
- 50 mM HEPES
- 1 mM EDTA
- 0.7% sodium deoxycholat
- 1% IGEPAL CA-630
- 0.5M LiCl
- Up to 20 mL MQ-H_2O
- 1x cOmplete, EDTA-free protease inhibitor cocktail tablet

ChIP-SDS lysis buffer
- 100mM NaCl
- 50 mM Tris-HCl pH 8.1
- 5 mM EDTA
- 0.2% NaN_3
- 0.5% SDS
- 0.1 µM PMSF
- 1 µg/ml AP
- 1 µg/ml LP

De-crosslinking buffer
- 1% SDS
- 0.1 M $NaHCO_3$

1xTAE buffer
- 20 mL 50x TAE buffer
- 980 mL dH_2O

Multiple purpose solutions

Loading Sample Buffer (LSB)
- 50 mM Tris pH 6,8
- 2% SDS
- 0.1% Bromophenol Blue
- 10% Glycerol

Section C

Table S6. Mass spectrometry protein list of proteins interacting with CDK12-GFP under NHS conditions.
Proteins marked in yellow were included in MS functional analyses.

First round of ChIP-MS, non-heat shock conditions				
KRT6A	PRKCSH	TUBA4A	LANCL1	CPNE2
ALB	POR	FABP5	CS	CPNE9
SRGAP3	AZGP1	HN1L	BANF1	CPNE4
AP1G1	SET	SACS	LYZ	CPNE6
PDLIM5	ANP32A	CPNE3	CDK12	CPNE8
SIN3A	LARS	MAPK1IP1L	ANP32B	CPNE5
LACRT	EIF1	Sept 10	SETSIP	CPNE7
LTF	ATP5I	LYZ		

Table S7. Mass spectrometry protein list of proteins interacting with CDK12-GFP after 10 minutes of HS.

First round of ChIP-MS, 10 minutes of heat shock				
SNRPG	ANPEP	BCCIP	ETFA	DNM1
ALB	NEB	PEF1	HSP90AB2P	EIF5
KRT77	WARS	STOML2	SBSN	RPS15
IPO5	AZGP1	DIS3	CNTNAP3	SKP1
DNM1L	PSMA2	NOSIP	ANP32E	TUBA4A
CEP290	HMGB2	LUC7L2	ENY2	HBB
PSMD3	CANX	GOLT1B	EXOSC3	NAA25
DIAPH1	PPP2R1A	SRPRB	HYPK	SEC23A
BANF1	L1CAM	BZW2	UGGT1	SKIV2L
SSB	ATP6V1E1	SNRPGP15	CDK12	TOMM34
CTSD	EIF1	CNTNAP3B	Sept 10	DBN1
DNM3	GPD2	GPD2		
DNM2	CSNK1A1	CSNK1A1		

Table S8. Mass spectrometry protein list of proteins interacting with CDK12-GFP under NHS conditions.
Proteins marked in yellow were included in MS functional analyses.

Second round of ChIP-MS, non-heat shock conditions							
CNOT1	IGHA1	PPP2R1B	PPP1CA	PAFAH1B3	UBR2	NIFK	PACSIN3
SNRPG	C4BPA	ADSS	CNBP	PON2	NELFCD	RAB34	NOB1
TDRD15	RPN2	ADSL	PPP2CA	PPA1	RBM12B	TNKS1BP1	SSR3
TPM2	SLC25A5	CLIP1	RPS28	SAFB	FTSJ3	IMUP	TRIM33
KIF2A	ITGB1	POLR2B	RPL39P5	PPP1R7	SPG20	SLIRP	C14orf166
PDLIM1	ATP5B	CTNNA1	ACTG1	SKIV2L	LRRC47	FAM192A	HPSE
PSMD12	FH	STAT3	RAE1	OASL	MAPK1IP1L	EHD4	DRG1
SAP18	RHOC	USP8	HPCA	Sept 10	NUP43	DHX36	KDM2A
RTCA	SLC3A2	GARS	PURA	CCDC6	SMARCC2	PNN	CARHSP1
COPE	SRPR	SKIV2L2	SORD	DPYSL2	AGR3	WNK1	TBL2
PDCD5	ALDOC	ACAA2	RELA	KYNU	RAPGEF6	GOLPH3	PPME1
PPP1R12A	PRSS56	MSH2	GLO1	TXNRD1	IPO4	RPF2	UCHL5
XPO1	TROVE2	RECQL	PTK2	TSR1	SCFD1	LRRC40	GTF3C5
PRPF4	DLAT	QARS	FMR1	SMU1	PALLD	MOV10	RBM8A
RBFOX2	ACADM	ATP5O	HSPA14	DAK	STON2	EXOSC4	IKBKG
ASNA1	SLC25A6	CSNK1A1	ASPH	PYCRL	PSPC1	RIC8A	DDX19B
LANCL1	ACTN1	POLD2	SF3A3	SPECC1	GATAD2B	LUC7L	SRSF4
RRP9	TPT1	SRP9	SRSF5	DNTTIP2	STAM	NIT2	CDK3
RAD21	TRIM27	PRIM2	FKBP5	RNF20	GPKOW	EXOSC5	SLC25A4
DHX16	STMN1	HINT1	SQSTM1	ECM29	GLMN	NANS	RBFOX1
KDM1A	NELFE	MRE11A	SNW1	KIAA1429	SMARCE1	FAM207A	TPM1
USO1	XRCC1	FXR2	CUL1	DHX57	MAGOHB	RBM12	SNRPGP15
PQBP1	EIF2AK2	RAB5C	BOP1	CDC73	ZBTB9	ABCF3	CSNK1A1L
EDF1	EIF2S2	RAB7A	MLEC	EDC4	SEH1L	DDX19A	RAB5B
PPP1R11	CSRP1	SMARCA4	TTLL12	PDXDC1	DNAJA3	IMP3	RAB5A
ROCK2	VDAC1	SSR4	CTTN	TTC33	UHRF1	DDX18	ACTA1
CS	PCMT1	HCFC1	TRIM25	FIP1L1	DNAJC2	CDK12	ACTC1
BANF1	WARS	KIF11	PUM1	GIGYF2	DNAJC7	EIF3K	ACTG2
NPM3	ACP1	SMS	GANAB	LARP4	NCAPG	NXF1	ACTA2

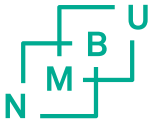
EIF3G	POLR2A	MRPL12	MVP	EIF3M	TUBA1C	PEF1	PPP2CB
EIF3J	CDK2	ARPC4	GOLGB1	BHLHA15	C7orf50	COPS7A	RPL39
BCAS2	PSMA4	ACTB	GAPVD1	NUP54	COPS4	ABCF2	HPCAL1
EML2	CANX	ACTR1A	NCAPH	PATL1	DDX23	SWAP70	EHD4
DHFR	CMPK1	LYZ	EXOSC7	SETD3	SPATA5L1	BAZ1B	

Table S9. Mass spectrometry protein list of proteins interacting with CDK12-GFP after 10 minutes of HS

Second round of ChIP-MS, 10 minutes of heat shock				
PRPF3	UQCRC2	NUP153	NUP153	NUP35
ASNA1	GARS	CHAF1B	CHAF1B	DDX54
COX5B	NUP153	WDR43	WDR43	SLK
PSMC3	PACSIN2	MTDH	MTDH	C8orf33

Table S10. Mass spectrometry protein list of proteins interacting with CDK12-GFP after 30 minutes of HS

Second round of ChIP-MS, 30 minutes of heat shock							
PFAS	APRT	MAPK1	**MAP1B**	RPS20	MLF2	ACAT2	SACS
ASNA1	SLC3A2	ADSL	ATP5O	SUMO1	ADRM1	POLDIP3	RBM27
NUDT21	SNRPA	CPS1	GCLM	DSC2	KYNU	XRN2	DACH1
DNAJA2	RAC1	SFN	HCFC1	LMNB2	INF2	OSBPL3	THRAP3
CBR3	GSPT1	PSMC2	ARHGDIA	PLP2	CEACAM16	MBNL1	PUS1
SNAPIN	GSTM3	RPS19	RAD23B	CUL1	NUP35	OLA1	PSAT1
HSPA4L	UQCRC2	SSR1	OXCT1	MLEC	STMN2	DDX19A	DDX19B
ATP1A1	CALR	MKI67	SEC61B	RRP1B	ZBTB9	FAM120A	MBNL2
GSTM5	RAC3	ATP1A3	GSTM2				



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway