U.PORTO

DOUTORAMENTO
BIOLOGIA MOLECULAR E CELULAR

Regulation of Alternative Polyadenylation in *Drosophila melanogaster*

Marta Sofia dos Santos Oliveira

D 2019

U.PORTO

Marta Sofia dos Santos Oliveira. Regulation of Alternative Polyadenylation in *Drosophila melanogaster*

D.ICBAS **2019**

Regulation of Alternative Polyadenylation in *Drosophila melanogaster*

Marta Sofia dos Santos Oliveira

SEDE ADMINISTRATIVA INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
FACULDADE DE CIÊNCIAS

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Regulation of alternative polyadenylation in *Drosophila melanogaster*

Marta Sofia dos Santos Oliveira

Tese do Programa Doutoral em Biologia Molecular e Celular

2019

MARTA SOFIA DOS SANTOS OLIVEIRA

Regulation of alternative polyadenylation in Drosophila melanogaster

Tese de Candidatura ao grau de Doutor em Biologia Molecular e Celular;

Programa Doutoral da Universidade do Porto (Instituto de Ciências Biomédicas de Abel Salazar e Faculdade de Ciências)

Orientadora: Doutora Maria Alexandra Marques Moreira Mourão do Carmo

Categoria: Investigadora Principal

Afiliação: Instituto de Ciências Biomédicas Abel Salazar e Instituto de Biologia Molecular e Celular e Universidade do Porto, Porto, Portugal

Co-orientadora: Prof. Dr. Ana Maria Pires Pombo

Categoria: Full Professor

Afiliação: Berlin Institute for Molecular Systems Biology, Max-Delbrück Center for Molecular Medicine e Humboldt University of Berlin, Berlim, Alemanha

Co-orientador: Doutor Jaime Pais de Freitas

Categoria: Pós-doutorado

Afiliação: Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

Este trabalho iniciou-se em Setembro de 2013, foi co-financiado pela FEDER (FCOMP-01-0124-FEDER-022718-PEst-C/SAU/LA0002/2013) durante o período de Abril a Agosto de 2014 e financiado pela Fundação para a Ciência e Tecnologia (SFRH/BD/102002/2014) com início a Janeiro de 2015 e fim em Dezembro de 2018.

De acordo com o disposto no nº 2, alínea a, do artigo 31º do Decreto-Lei nº 230/2009, este trabalho apresenta resultados descritos e aceites para publicação com as seguintes referências:

MARTA S. OLIVEIRA, Jaime Freitas, Pedro AB Pinto, Ana Jesus, Joana Tavares, Mafalda Pinho, Carla Lopes, Carlos Conde, Claudio E. Sunkel and Alexandra Moreira (2019) The cell cycle kinase Polo is controlled by a conserved 3′UTR regulatory sequence in *Drosophila melanogaster. Molecular and Cellular Biology*. doi:10.1128/MCB.00581-18.

Liu, X., Freitas, J., Zheng, D., Hoque, M., **OLIVEIRA, M. S.**, Martins, T., Henriques, T., Tian, B., Moreira, A. (2017) Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *RNA, 23*(12), 1807-1816. doi:10.1261/rna.062661.117

Eu, a autora Oliveira, M. S. e autora desta tese de doutoramento, declaro que planeei e executei a maioria do trabalho experimental, interpretação e respectiva discussão dos resultados descritos ao longo desta dissertação e nos manuscritos acima citados com o apoio e ajuda dos restantes autores.

As contribuições a nível experimental e de análise não realizadas por mim referidas neste tese são as seguintes: os autores Liu, X., Freitas, J., Zheng, D. e Hoque, M. realizaram o protocolo de 3’ Region Extraction And Deep Sequencing e identificação e análise de locais de poliadenilação; o autor Pinto, PAB analisou a conservação do sinal pA1 do *polo* e respectivo alinhamento e criou a estirpe de mosca com o génotipo w^{1118};*gfp-polo*Δ*USE*;$polo^{9}$/*TM6B*; a autora Tavares, J criou e executou o scrip no software R; o autor Henriques, T analisou e compilou os resultados de ChIP-seq; a autora Lopes C balanceou a estirpe de mosca $heph^{2}$/*TM3* com o balanceador *TM6B*; e a aquisição de imagens de imunofluorescência e aneuploidia foi realizada pelo autor Conde C.

TABLE OF CONTENTS

ABSTRACT

Alternative polyadenylation (APA) is a mechanism that contributes to the complexity of the transcriptome by generating mRNA isoforms that differ in their coding sequence and/or in the 3' untranslated region (UTR) by the selection of different polyadenylation signals. APA occurs co-transcriptionally, therefore, we sought to understand the molecular mechanisms behind the influence of RNA polymerase II (RNAPII) transcription elongation rate on polyadenylation site selection at a genome-wide scale in *Drosophila melanogaster*. Using 3'READS, we analysed APA mRNA isoforms in fly bodies and heads. We showed that APA differences observed between bodies and heads may be due to the distance between proximal and distal polyadenylation sites and also due to differential gene expression of important elongation and mRNA 3'end formation factors. Several genes coding for proteins with functions in cleavage, polyadenylation, elongation and termination, including *Ssu72*, *Cdk9*, *CsfF64* and *Pcf1*, are upregulated in *Drosophila melanogaster* heads, which indicates that there is an environment-specific regulation of this class of proteins. Using the *Drosophila melanogaster* mutant strain *RpII215*^{C4}, which possesses a point mutation in RNAPII largest subunit causing a 50% slower transcription elongation rate, we showed that APA is affected by a slower RNAPII elongation rate in a context-dependent manner. For a significant number of genes in the fly mutant body, there is an increase in the expression of mRNAs containing 3'UTRs produced by proximal and generally weaker polyadenylation sites. We also showed that *Ssu72*, which is a phosphatase that acts on the carboxy-terminal domain (CTD) of RNAPII, is upregulated in the fly mutant bodies. Taken together, these results highlight the genome-wide and tissue-specific impact of RNAPII elongation rate in *Drosophila melanogaster* APA. To dissect the molecular mechanisms underlying these observations, we focused on the *polo* gene, the *Polo Like Kinase-1* orthologue. *polo* encodes the conserved cell cycle regulator Polo and has two polyadenylation signals in the 3'UTR, producing two distinct mRNAs. Previous work from our group showed that transgenic flies without the distal polyadenylation signal of *polo* die during metamorphosis. This occurs because the longer *polo* mRNA is the main responsible for Polo protein production necessary for the correct proliferation of abdominal histoblasts during metamorphosis and viability of the fly. Using bioinformatic tools, we searched for potential auxiliary *cis* elements in *polo* 3'UTR and we found a conserved pyrimidine-rich sequence upstream of the proximal polyadenylation signal of *polo* that we named Upstream Sequence Element (USE) as it is similar to the previously described

human and viral USEs. We observed that the USE is present in approximately 5% of the 3'UTRs of all *Drosophila melanogaster* genes and that it is more prevalent when located upstream of weak polyadenylation signals. We also showed that the USE has an *in vivo* effect on *polo* expression. Transgenic flies without this element in *polo* 3'UTR present an abdominal phenotype, show impaired *polo* APA, and low Polo protein levels and activity at the kinetochores. This results in mitotic aberrations and aneuploidies due to the lower activity of the mitotic kinases Aurora B and Mps1, two known Polo targets. These findings highlight an *in vivo* role for this conserved USE in the regulation of *polo* mRNA 3'end processing and expression. We identified that Hephaestus (Heph) RNA binding protein binds to *polo* USE and showed that Heph depletion causes a decrease in the levels of the longest *polo* mRNA and Polo protein. Accordingly, fly mutants with low Heph levels ($heph^2$ hypomorph) show altered *polo* mRNA 3'end formation and significantly reduced Polo, Aurora B and Mps1 levels at the kinetochores of proliferating cells, which also present a higher tendency to become aneuploid. This is the first finding of a *cis* regulatory sequence in *Drosophila melanogaster* capable of regulating *polo* APA by binding of Heph, thus affecting Polo activity and mitosis. It had been previously shown by our group that Polo protein levels control *polo* APA in an auto-regulatory feedback loop mechanism, suggesting a yet unknown function for Polo. To investigate the hypothesis of Polo being involved in polyadenylation site selection, we used two *polo* mutant fly strains: $polo^1$, a kinase dead mutant and $polo^9$, a null hypomorph. We unveiled that impaired Polo activity lowers the expression levels of *CG6024*, *abd-b*, *rp49* and *U6*, either directly or indirectly, and that low Polo levels lead to a predominantly hypophosphorylated RNAPII CTD profile in comparison to wild type, indicating that this cell cycle kinase might have a novel role in RNAPII-mediated transcription or co-transcriptional mechanisms in *Drosophila melanogaster*. In summary, we showed that RNAPII transcription elongation rate modulates APA at the genome-wide level in *Drosophila melanogaster* in a context-dependent mode. We also dissected the molecular mechanisms involved in *polo* APA, characterizing a novel USE that binds Heph and controls Polo activity at the kinetochores and cell cycle progression. Additionally, we described a possible new function for Polo in RNAPII CTD phosphorylation. The work described in this thesis thus provides new insight into the basic molecular mechanisms that occur during transcription and APA *in vivo* and how they act synergistically with RNAPII to efficiently coordinate mRNA 3'end processing, gene expression as well as cell cycle progression in *Drosophila melanogaster*.

SUMÁRIO

A poliadenilação alternativa (APA) é um mecanismo que contribui para a complexidade do transcriptoma ao gerar isoformas de mRNA que diferem na sua sequência codificante e/ou região 3' não traduzida (UTR) via a selecção de differentes sinais de poliadenilação. A APA ocorre co-transcricionalmente, logo, tentamos compreender os mecanismos moleculares resultantes da influência da velocidade de elongação da transcrição da RNA polimerase II (RNAPII) na selecção do local de poliadenilação *genome-wide* em *Drosophila melanogaster*. Usando a técnica 3'READS, analisamos isoformas de mRNA produzidas por APA em corpos e cabeças de moscas. Nós mostramos que as diferenças de eventos de APA observadas entre os corpos e as cabeças podem dever-se à distância entre os locais de poliadenilação proximal e distal e também à expressão diferencial de importantes factores de elongação e proteínas envolvidas na formação da extremidade 3' do mRNA. Vários genes que codificam para proteínas com funções na clivagem, poliadenilação, elongação e terminação, incluindo *Ssu72*, *Cdk9*, *CsfF64* e *Pcf1*, estão *upregulated* nas cabeças de *Drosophila melanogaster*, o que indica uma regulação desta classe de proteínas dependente do contexto. Utilizando a estirpe mutante de *Drosophila melanogaster RpII215*^{C4}, que tem uma mutação pontual na subunidade maior da RNAPII que causa uma velocidade de elongação da transcrição 50% mais lenta, nós mostramos que a APA é afectada por uma velocidade de elongação da RNAPII mais lenta dependendo do contexto. Para um número significativo de genes no corpo da mosca mutante, há um aumento na expressão de mRNAs com 3'UTRs produzidos por locais de poliadenilação proximais que são geralmente mais fracos. Também mostramos que *Ssu72*, uma fosfatase que actua no domínio carboxi-terminal (CTD) da RNAPII, está *upregulated* nos corpos da mosca mutante. Estes resultados enfatizam o impacto *genome-wide* e dependente do contexto celular da velocidade de elongação da RNAPII na APA de *Drosophila melanogaster*. De forma a dissecar os mecanismos moleculares subjacentes, nós concentramo-nos no gene *polo*, o ortólogo da *Polo Like Kinase-1*. O gene *polo* codifica a cinase Polo, um regulador do ciclo celular altamente conservado, e tem dois sinais de poliadenilação no 3'UTR, produzindo assim dois mRNAs distintos. Trabalhos anteriores do nosso grupo mostraram que moscas transgénicas sem o sinal de poliadenilação distal de *polo* morrem durante a metamorfose. Isto deve-se ao facto do mRNA mais longo de *polo* ser o principal responsável pela produção da proteína Polo necessária para a correcta proliferação dos histoblastos abdominais durante a metamorfose e para a viabilidade da mosca. Usando metodologias bioinformáticas, procuramos por potenciais elementos *cis*

auxiliares no 3'UTR do *polo* e encontramos uma sequência conservada rica em pirimidinas a montante do sinal de poliadenilação proximal do *polo* que denominamos de Upstream Sequence Element (USE) devido às semelhanças com USEs descritos anteriormente em humanos e vírus. Nós observamos que o USE existe em aproximadamente 5% dos 3'UTRs de todos os genes de *Drosophila melanogaster* e que é mais prevalente a montante de sinais de poliadenilação fracos. Também mostramos que o USE tem um efeito *in vivo* na expressão do *polo*. Moscas transgénicas sem este elemento no 3'UTR do *polo* demonstram um fenótipo abdominal, apresentam APA alterada do *polo*, e níveis baixos de proteína Polo e sua respectiva actividade nos cinetocoros. Por consequência, isto causa aberrações mitóticas e aneuploidias devido à baixa atividade das cinases mitóticas Aurora B e Mps1, dois dos alvos de Polo. Estes resultados enfatizam uma função *in vivo* na regulação da expressão e processamento das extremidades 3' dos mRNAs do *polo* para este USE conservado. Demonstramos que a proteína Hephaestus (Heph) é capaz de se ligar ao USE do *polo* e mostramos que a depleção de Heph causa uma diminuição nos níveis da isoforma mais longa do *polo* e da proteína Polo. Mutantes de mosca com baixos níveis de Heph (hipomorfo $heph^{2}$) demonstram alterações na formação das extremidades 3' dos mRNAs de *polo* e uma redução significativa dos níveis de Polo, Aurora B e Mps1 nos cinetocoros de células em divisão, que também apresentam uma maior tendência para se tornarem aneuploides. Esta é a primeira descoberta de uma sequência reguladora *cis* em *Drosophila melanogaster* capaz de regular a APA do *polo* via a ligação de Heph, afectando assim a actividade do Polo e a mitose. Anteriormente, o nosso grupo demonstrou que os níveis de proteína Polo controlam a APA do *polo* via um mecanismo de auto-regulação, o que sugeria uma função ainda desconhecida para Polo. De forma a investigar a hipótese de Polo estar envolvido na selecção de locais de poliadenilação, foram utilizadas duas estirpes de mosca mutantes para *polo*: $polo^{1}$, um mutante *kinase dead* e $polo^{9}$, um hipomorfo nulo. Nós revelamos que Polo com uma baixa actividade reduz os níveis de expressão de *CG6024*, *abd-b*, *rp49* e *U6*, directa ou indirectamente, e que baixos níveis de Polo causam um perfil predominantemente hipofosforilado da CTD da RNAPII em comparação com o controlo, o que indica que esta cinase do ciclo celular pode ter uma nova função na transcrição mediada pela RNAPII ou nos mecanismos co-transcripcionais em *Drosophila melanogaster*. Em suma, nós mostramos que a velocidade de elongação da transcrição da RNAPII modula a APA ao nível *genome-wide* em *Drosophila melanogaster* dependendo do contexto. Nós também dissecamos os mecanismos moleculares envolvidos na APA do *polo* ao caracterizar um USE ao qual se liga Heph e que assim controla a actividade de Polo nos cinetocoros e a progressão do ciclo

celular. Finalmente, descrevemos uma potencial nova função para o Polo na fosfosrilação da CTD da RNAPII. O trabalho descrito nesta tese fornece assim uma nova percepção sobre os mecanismos moleculares básicos que ocorrem durante a transcrição e APA *in vivo* e como estes actuam sinergisticamente com a RNAPII de forma a coordenar eficientemente o processamento das extremidades 3' dos mRNAs, a expressão genética e a progressão do ciclo celular em *Drosophila melanogaster*.

List of Abbreviations

Δ: genetic deletion

3'RACE: rapid amplification of cDNA ends

3'READS: 3' region extraction and deep sequencing

Abd-b: Abdominal B

APA: alternative polyadenylation

APC/C: anaphase-promoting complex/cyclosome

Ars2: arsenic resistance protein 2

bp: base pair(s)

BubR1: Bub1-related kinase

Cdk: cyclin-dependent kinase

cDNA: complementary DNA

CFIm: cleavage factor I

CFIIm: cleavage factor II

ChIP: chromatin immunoprecipitation

COX-2: cyclooxygenase-2

CS: cleavage site

CSTF: cleavage stimulation factor

CPSF: cleavage and polyadenylation specificity factor

CTD: Carboxy Terminal Domain

CyO: *curly of oster*

DGET: *Drosophila* gene expression tool

DNA: deoxyribonucleic acid

DNase: deoxyribonuclease

dNTPs: deoxynucleoside-triphosphate

dPAS: distal polyadenylation site

DSE: downstream sequence element

DSIF: DRB sensitivity inducing factor

DTT: dithiothreitol

ECL: enhanced chemiluminescence

EDTA: ethylenediaminetetraacetic acid

Elav: embryonic lethal abnormal vision

Emb: Embargoed

FBS: fetal bovine serum

FDR: false discovery rate

FoxM1: Forkhead Box M1

FOXO1: Forkhead Box O1

GFP: green fluorescent protein

Heph: Hephaestus

hnRNPK: heterogeneous nuclear ribonucleoprotein K

IgG-HRP: immunoglobulin G-horseradish peroxidise

IgM: immunoglobulin M

kb: kilobase(s)

INCENP: Inner centromere protein

LC: liquid chromatografy

MOPS: 3-(N-morpholino)propanesulfonic acid

Mps1: Monopolar spindle 1

Mps1^{T490Ph}: Mps1 phosphorylated at T490

mRNA: messenger RNA

MS: mass spectrometry

MT: mutant

NELF: negative elongation factor

nt: nucleotide(s)

pA: polyadenylation

PAGE: polyacrylamide gel electrophoresis

PAP: poly(A) polymerase

PBD: Polo-box domain

PBS: phosphate buffered saline

PCR: polymerase chain reaction

PLK: Polo-like kinase

PMSF: phenylmethylsulfonyl fluoride

pPAS: proximal polyadenylation site

pre-mRNA: precursor messenger RNA

PTBP1: polypyrimidine tract binding protein

P-TEFβ: positive transcription elongation factor b

RBP: RNA binding protein

RE: relative expression

RED: relative expression difference

RNA: ribonucleic acid

RNAPI: RNA Polymerase I

RNAPII: RNA Polymerase II

RNAPIII: RNA Polymerase III

RNAPs: RNA polymerases

RNase: ribonuclease

ROI: region of interest

rp49: Ribosomal protein 49

RPM: reads per million

rpm: rotations per minute

rRNA: ribosomal RNA

RT-qPCR: real time-quantitative polymerase chain reaction

SAC: spindle assembly checkpoint

SDE: standard deviation error

SDS: sodium dodecyl sulfate

Ser-2-P: phosphorylated RNAPII CTD Serine 2

Ser-5-P: phosphorylated RNAPII CTD Serine 5

Ser-7-P: phosphorylated RNAPII CTD Serine 7

Tyr-1-P: phosphorylated RNAPII CTD Tyrosine 1

Thr-4-P: phosphorylated RNAPII CTD Threonine 4

snRNP: small nuclear ribonucleoprotein

SSC: saline sodium citrate

Ssu72: suppressor of *sua7* gene 2

TBS: Tris buffered saline

TFIIIB: transcription factor IIIB

TM6B: Third multiple 6B

Tris: trishydroxymethylaminomethane

tRNA: transfer RNA

USE: upstream sequence element

USEmt: mutated upstream sequence element

USE-N: upstream sequence element with a G>N substitution

UTR: unstranslated region

WT: wild type

INTRODUCTION

Three functionally and structurally related [1] DNA-dependent RNA polymerases (RNAPs) generate the entire transcriptome of all eukaryotes. Ribosomal RNA (rRNA) genes are by far the most transcribed genes, a function performed by RNA polymerase I (RNAPI). RNA polymerase II (RNAPII) is responsible for the transcription of protein coding genes and some non-coding RNAs, such as long non-coding RNAs and micro-RNA precursors. 5S rRNA, transfer RNAs (tRNAs) and some non-coding RNAs (like 7SK RNA and U6 snRNA) are transcribed by RNA polymerase III (RNAPIII). 90% of all transcription in a cell is performed by RNAPI and RNAPIII while RNAPII is responsible for the remaining transcriptional output [2].

1. RNA Polymerase II

RNAPII is the smallest of eukaryotic RNAPs and comprises 12 subunits (RPB1-12) [3]. Ten of them form a highly conserved structure around the catalytic site [4], which is surrounded on opposite sides by subunits RPB1 and RPB2. These two, together with RPB3 and RPB11 are similar in sequence and structure to other RNAPs [5, 6] and RPB5, RPB6, RPB8, RPB10 and RPB12 subunits are shared by all RNAPs [1]. A peripheral heterodimer around RNAPII made by RPB4 and RPB7 forms a stalk structure during the transcription cycle that binds to RPB1, RPB2 and RPB6 [7] and has regulatory potential [8].

1.1. The RNA Polymerase II Carboxy Terminal Domain code

Different from any other RNAP [9], the largest subunit of RNAPII, RPB1, has a long, disordered and mobile Carboxy Terminal domain (CTD) consisting of a simple heptapeptide repeat or heptarepeat with the following amino acid consensus sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser [10]. This CTD is unique to higher eukaryotes [11, 12], evolutionarily conserved [13] and many deletions and/or mutations to this domain are incompatible with life [14-16].

Mammals have a highly conserved RNAPII CTD consisting of 52 tandem heptarepeats while *Drosophila melanogaster* has 45 repeats [10], the latter of which are significantly divergent in sequence from its mammalian counterpart ([13, 17] and see **FIGURE 1**). RNAPII requires a CTD with a minimal number of repeats for normal function, viability and development in both mammalian cells and *Drosophila melanogaster* [13, 14, 16, 18, 19]. Expansion of CTD repeats is associated with an increase in organism complexity [12, 20] and evolutionary optimization for functional efficiency [10, 18].

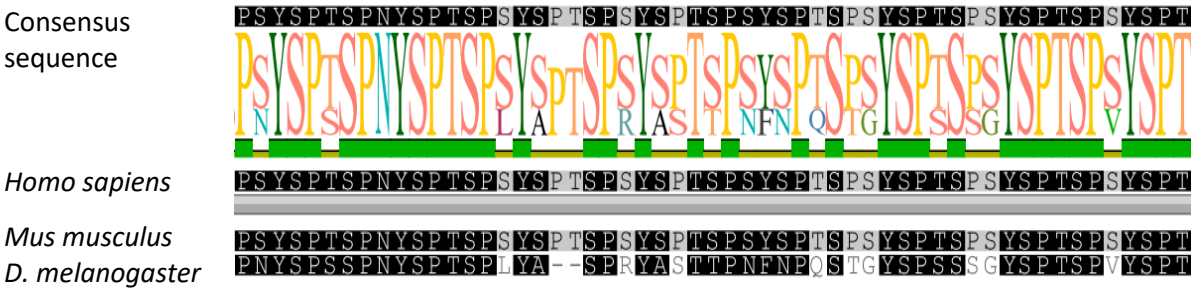


FIGURE 1 | Partial sequence (from the 40th to the 94th amino acid) of the CTD tail of the RPB1 subunit in *Homo sapiens*, *Mus musculus* and *Drosophila melanogaster*.

The key function of the RNAPII CTD tail is to serve as a signalling platform where several proteins with a myriad of roles in chromatin remodelling, mRNA processing, transcription initiation, elongation and termination dynamically and co-transcriptionally assemble by recognizing one or more specific RNAPII CTD modifications and/or conformations in a timely and sequential manner [21-23]. Also, the pattern of RNAPII CTD modifications created by various enzymes (mostly kinases and phosphatases, which will be explored in the next section) in an orderly fashion is associated to specific steps of the transcriptional cycle, thus creating the CTD code [24, 25].

Every residue of the heptarepeat may undergo reversible post-translational modifications during the transcription cycle. These modifications include phosphorylation [26-31], glycosylation, isomerisation [32], *O*-GlcNAcylation [33], methylation [34, 35], ubiquitination [36] and acetylation [35, 37]. An unmodified RNAPII CTD is flexible, unstructured and condensed, allowing it to evade certain modifications that would recruit specific proteins in an untimely fashion [38]. A modified CTD is longer and more rigid in structure [26, 39] and by avoiding certain conformations incompatible with the binding of certain proteins [40], this form of CTD may be a more competent signalling platform than the unmodified version of CTD.

Interestingly, it has been reported that on average, RNAPII CTD heptarepeats usually tend to be single phosphorylated instead of undergoing many potential post transcriptional modifications, indicating that the CTD does not tend to be heavily phosphorylated/modified at any potential site [41]. This also suggests that the potential combinations of RNAPII CTD modification patterns, i.e., the complexity of the CTD code, are not as intricate as first expected. This is in line with the fact that it would be challenging for a kinase to add a second phosphate to an already negatively charged heptarepeat and as such, it is more likely that two modifications reside in two neighbouring repeats [29]. Nevertheless, considering the wide combination of potential modifications, the number of heptarepeats, their occasional deviation from the canonical

sequence and the dynamic signalling, activity and interaction of the many CTD-binding proteins, the RNAPII CTD is a hotspot for gene expression control and coordination of the transcriptional machinery throughout the entire transcription cycle.

1.2. Transcription cycle

To express or not to express a protein-coding gene relies on several variables: the corresponding DNA sequence must be accessible, RNAPII-mediated transcription together with excision of introns and exon ligation by splicing must proceed without mishaps and the resulting precursor mRNA (pre-mRNA) must be properly processed, cleaved and polyadenylated in order to become stable and mature before being transported to the cytoplasm. The careful regulation, control checkpoints and co-transcriptional coupling of all these steps governs cellular identity by accurately and efficiently determining which, when, where and how much each gene is expressed. In the next few sections, transcription initiation, elongation and termination will be summarily described.

1.2.1. Initiation

By itself, RNAPII cannot initiate transcription. Instead, RNAPII with an unmodified CTD assembles upon promoters (see **FIGURE 2**, panel 1) in response to activation signals and basal transcription factors (reviewed in [42]), which can in turn recruit chromatin remodelers to facilitate promoter access [43] and recognize specific promoter regions that aid its identification and the assembly of a stable pre-initiation complex [22, 44]. Recruited by the Mediator complex [45], the helicase TFIIH then separates the template DNA strands and RNA synthesis of the first 20 nucleotides (nt) begins. As soon as this nascent pre-mRNA transcript emerges from RNAPII, it is immediately decorated with different proteins [46].

TFIIH includes CDK7, which phosphorylates RNAPII CTD Ser-5 and then Ser-7 close to the transcription start site [45, 47-52]. In turn, this releases the stabilizing Mediator complex [53] and contributes to the co-transcriptional recruitment and activation of the capping machinery [52, 54, 55] to immediately modify the nascent 5'end of the pre-mRNA and protect it from degradation [56, 57]. Phosphorylated by the ABL1/2 kinase [58], RNAPII CTD Tyr-1-P also interacts with the capping enzyme [59]. This stage is known as promoter clearance (see **FIGURE 2**, panel 2) and many interactions between RNAPII and basal transcription factors are lost [60]. Once one RNAPII escapes the promoter and frees it, another pre-initiation complex may now form.

However, after promoter clearance, most Ser-5-P RNAPIIs generate short transcripts, pause and terminate [61]. In fact, a significant percentage of promoter proximal regions of several mammalian and *Drosophila melanogaster* genes are occupied by these paused RNAPII [62-65] after transcribing approximately 20-60 nt past the transcription start site whether the gene is being actively transcribed or not [57, 62, 64, 66-68]. At this stage, there is a tightly controlled [69] checkpoint between initiation and productive elongation: promoter proximal pausing [66].

There are two main factors responsible for the promoter proximal pausing. DRB sensitivity inducing factor (DSIF), which includes the SPT4 and SPT5 subunits [70], binds to the RNAPII in regions previously covered by basal transcription factors [71]. The negative elongation factor complex (NELF, comprised of NELF-A, B, D and E) sequentially binds to DSIF [72-75]. Both complexes stabilize the paused RNAPII in an inactive state [76], cause its enrichment near promoters [62, 74, 77] and inhibit transcription elongation [70, 73, 78, 79] by distorting the catalytic site and hindering polymerization of the nascent strand [76].

Several attempts to transcribe-pause-and-backtrack are made by RNAPII before this rate-limiting step is successful, implying that this transition is a central step of gene expression control. Firstly, Positive-Transcription Elongation Factor β (P-TEFβ) [61, 80-82], which comprises cyclin dependent kinase (CDK) 9 and cyclin T, is recruited to promoter-paused RNAPIIs via several transcription factors [26, 83]. P-TEFβ phosphorylates RNAPII CTD Ser-2, DSIF and NELF [26, 28, 84], which promotes the release of RNAPII from its paused status [26] and transition into productive elongation. RNAPII CTD Ser-7-P also promotes efficient RNAPII CTD Ser-2-P by P-TEFβ [29]. A phosphorylated NELF no longer associates with DSIF while phosphorylated SPT5 becomes a positive regulator of elongation throughout the remainder of the transcription cycle [72, 77, 85]. Like RNAPII CTD Ser-5-P, phosphorylated SPT5 co-transcriptionally recruits and activates the capping machinery as well [86].

Secondly, TFIIS also efficiently promotes the escape and elongation of promoter-paused RNAPII caused by DSIF/NELF inhibition [57, 81, 87] as there is evidence that both this transcription factor and NELF use mutually exclusive RNAPII sites [76]. If NELF is bound to RNAPII, TFIIS is unable to realign the RNAPII active site to promote elongation [57, 81, 87]. Once NELF is phosphorylated by CDK9, it is released from RNAPII and TFIIS is now able to bind.

In turn, RNAPII CTD Ser-2-P recruits other proteins with various functions in chromatin remodelling and mRNA processing [22, 88-93] and will subsequently cause the release of most of the proteins that comprised the pre-initiation complex to give rise to new binding

sites for different proteins [26]. RNAPII CTD Ser-2-P is prevalent throughout gene bodies and their 3'ends [49].

Promoter proximal pausing may thus serve to ensure 5'end capping is complete and RNAPII is properly modified and decorated with the proper mRNA processing proteins before proceeding into productive elongation ([72, 94-96] and see **FIGURE 2**, panel 3) while concomitantly bypassing the rate-limiting RNAPII recruitment to the pre-initiation complex and facilitating a quick transcription. This way, RNAPII CTD Ser-5 and Ser-7 are modified after successful transcription initiation and only then can RNAPII CTD Ser-2 be phosphorylated at the correct stage [22]. Additionally, promoter proximal RNAPII pausing and the proteins present in the pre-initiation complex that remain after RNAPII promoter clearance aid in maintaining the promoter accessible [43, 97] and may ensure a quicker formation of new pre-initiation complexes [95]. When the promoter proximal pausing control checkpoint is surpassed, RNAPII is now primed for productive elongation [89, 98].

1.2.2. Elongation

Throughout the transcriptional cycle, some phosphorylation marks are dynamically modulated and removed. During the transition to productive elongation, RNAPII CTD Ser-5-P and Ser-7-P are gradually dephosphorylated by the suppressor of *sua7* gene 2 (Ssu72) phosphatase [23, 99-101] while RNAPII CTD Ser-2-P is increasingly phosphorylated by CDK9 (see **FIGURE 2**, panel 4). Small CTD phosphatase 1 and RNAPII-associated protein 2 can also dephosphorylate RNAPII CTD Ser-5-P [102, 103].

During elongation, CDK12 and CDK13 also phosphorylate RNAPII CTD Ser-2-P [69, 84, 104-106]. In *Drosophila melanogaster*, Cdk9 is found enriched at the 5'ends of genes whereas Cdk12 can be detected along the gene bodies and at their 3'ends, suggesting that they preferentially phosphorylate RNAPII CTD Ser-2 at different stages [105, 106]. In humans, this distinction is not as clear as both CDK12 and CDK13 are orthologs of the fruit fly Cdk12 [106, 107] while BRD4 is also a RNAPII CTD Ser-2 kinase able to stimulate CDK9 activity [108-110]. CDK12 activity is also spurred by the presence of RNAPII CTD Ser-7-P [111].

RNAPII CTD phosphorylation does not overall interfere with the transcription elongation rate [22]. However, the rate of transcription elongation affects gene expression. The *Drosophila melanogaster* mutant *RpII215*^{C4} has a 50% slower transcription elongation rate in comparison to wild type due to a point mutation in the Rpb1 subunit of RNAPII where an arginine was replaced by a histidine (R741H) [112-114]. It has been shown that a slower transcription elongation rate deregulates mRNA 3'end processing of *polo*, *abd-b,*

lace, stlk, CG6024, *cyclin D* and histones [115, 116] and can cause premature transcription termination [117]. Splicing in numerous genes is also altered by a slower RNAPII [118, 119], including splicing of the *ultrabithorax Hox* gene in the fly [120].

RNAPII usually moves forward as it polymerizes RNA, but it can also move backwards at certain DNA sequences: this is called backtracking. A higher frequency of backtracking has been attributed to the slower fly mutant RNAPII [121]. Certain pause sites along the DNA template exist where RNAPII backtracks, which may result in transcription arrest [122-125]. As mentioned, TFIIS can reactivate RNAPII elongation by realigning the RNAPII active site and promote elongation [57, 87]. RNAPII tends to backtrack often [126], which highlights the relevance of the role of TFIIS in enhancing transcriptional elongation and maintaining paused RNAPIIs in a ready-to-elongate state [76, 126].

Concomitant with the observation that RNAPII CTD Thr-4-P is associated with the hyperphosphorylated isoform of RNAPII, this mark appears in gene bodies (see **FIGURE 2**, panel 4) and is enriched at the 3'ends of genes [30]. Indeed, all genes being transcribed contain this CTD modification [127]. Curiously, in mitotic cells, a hyperphosphorylated RNAPII only phosphorylated upon CTD Thr-4 has been found tethered to centrosomes and midbodies with the responsible kinases being PLK-1 [128], PLK-3 and CDK9 [30]. While the function of this modification in this context is still unclear, RNAPII CTD Thr-4 mutants show mitotic abnormalities [128], elongation anomalies or even lethality, with RNAPII enrichment past the transcription start site and overall low levels across gene bodies and their respective 3'ends [30]. The TFIIF-associated phosphatase F-Cell Production 1 can dephosphorylate RNAPII CTD Ser-2-P, Ser-5-P and Thr-4-P [129, 130].

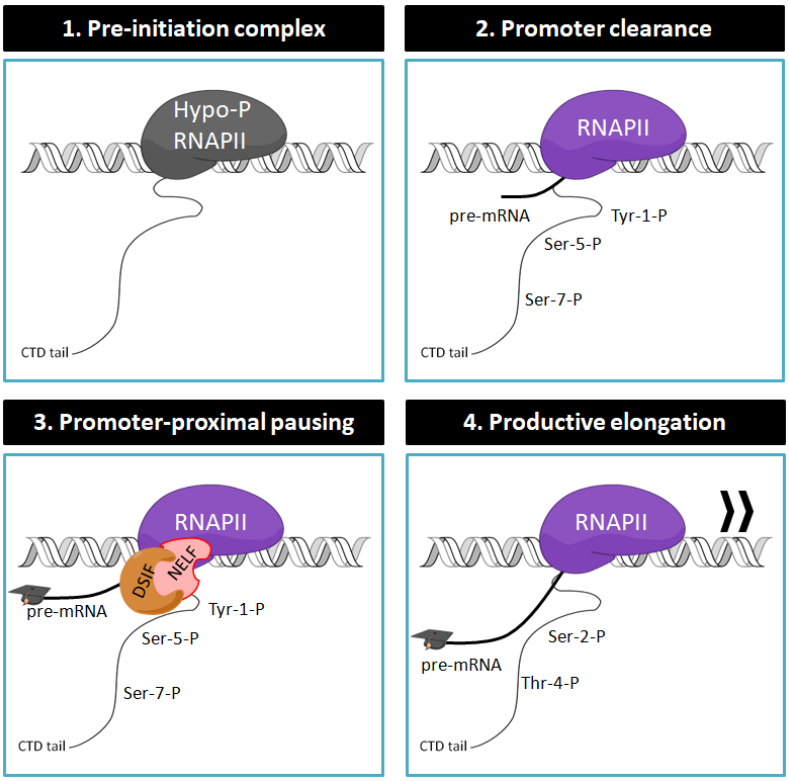


FIGURE 2 | The hyper- and hypophosphorylated RNAPII CTD tail along the transcription cycle of a protein-coding gene. A hypophosphorylated RNAPII CTD is recruited to the promoter of the gene to be transcribed (panel 1). Once stable as part of a pre-initiation complex, RNAPII CTD Tyr-1, Ser-5 and Ser-7 are phosphorylated and RNA polymerization begins (panel 2), but RNAPII is quickly paused while still close to the promoter by the DSIF/NELF complex to allow extra CTD modifications and the 5'end capping of the nascent pre-mRNA (panel 3). As RNAPII starts to elongate productively, CTD Ser-2 and Thr-4 are also phosphorylated (panel 4). RNAPII CTD is then targeted by phosphatases, which returns it to its original hypophosphorylated state so it can start a new transcription cycle (panel 1). Adapted from [131].

Additionally, it has been reported that across the entire CTD, the Ser-2-P and Ser-5-P marks are present in identical amounts whereas Tyr-1-P and Thr-4-P are more seldom to find [41].

1.2.3. Termination

The release of the newly transcribed transcript and disengaging of RNAPII from the DNA – transcription termination – is a phenomenon intertwined with mRNA 3'end formation. There are currently three models that attempt to explain transcription termination: the allosteric model [132, 133], the torpedo model [134-136] or a combination of the two [137-142].

The allosteric/anti-terminator model postulates that after RNAPII transcribes past the cleavage and polyadenylation (pA) site, the lack of interaction of several elongation and/or termination factors (such as Pcf11 [139, 143]) or the pA site itself may be destabilizing and lead to conformational changes that cause RNAPII to disengage [132, 144], consequently leading to transcription termination.

The torpedo model [145] suggests that once the pre-mRNA is cleaved, the elongating RNAPII continues to transcribe past the pA signal in the presence of CDK9 and phosphorylated SPT5 [146], now connected to a new transcript whose uncapped 5'end matches the 3'end of the previously cleaved pre-mRNA [147]. Such an uncapped transcript can be targeted by Xrn2 [137], an exonuclease recruited close to the transcription start site [148] which rapidly degrades RNA [135, 144] and competes with the elongating RNAPII. Hypothetically, if Xrn2 catches up with RNAPII, the removal of the nascent RNA from the catalytic site of RNAPII will cause a conformational change that dissociates it from the DNA and induces transcription termination [134, 147]. A dominant-negative mutation of Xrn2 delays transcription termination genome-wide [117]. Curiously, a slower RNAPII produces shorter transcripts while a fast RNAPII originated longer transcripts [117], which supports the premises of this model: Xrn2 can more easily catch up with a slower RNAPII and induce termination than with a faster RNAPII.

Several reports show that RNAPII occupancy is not only high at promoter regions, but also downstream of functional pA signals [77, 149-153], which correlates with higher levels of gene expression [154, 155] and depends on the interaction of termination factors (such as CSTF77 [156]) and capping enzymes [151]. As mentioned, transcription termination involves mRNA 3'end processing factors [101, 157-160] whose occupancy also peaks at these sites (such as PCF11, CSTF and the CPSF complexes [151, 153, 160]) and they also contribute to the accumulation of RNAPII downstream of the pA signal [161]. Regions downstream of functional pA signals are considered to be RNAPII pause sites and enhance transcription termination [162-164] mediated by Xrn2 degradation [137, 149]. Stimulated by this RNAPII pausing past the pA signal, phosphorylation of RNAPII CTD Ser-2 by CDK12 also peaks at the 3'ends of genes. RNAPII pausing by the pA signal and increased levels of RNAPII CTD Ser-2-P may coordinate to ensure proper recruitment of the mRNA 3'end processing machinery [156] and/or termination [151, 156].

Downstream of the pA signals, RNAPII also backtracks, which is thought to promote premature termination via Xrn2 degradation. As expected, TFIIS is present in these regions [126]. Interestingly, inhibiting TFIIS caused a 50% decrease in elongation rate [126] and the slower RNAPII caused by the *RpII215*^{C4} mutation, which also possesses a

50% decrease in transcriptional rate [112-114], is thought to backtrack more slowly than its wild type counterpart [121].

At the 3'end of genes, as each transcription round finishes, the hyperphosphorylated CTD of RNAPII must be dephosphorylated to return it to an initiation-competent hypophosphorylated state (see **FIGURE 2**, panel 1). This allows RNAPII recycling [96] so that a new transcription cycle may begin. As such, it is of the utmost importance that the phosphorylation/dephosphorylation events are sequential and abide by this specific order for proper gene expression control.

1.3. Transcription and pre-mRNA 3'end processing - co-transcriptional coupling

Transcription can, in part, be described as a streamlined "RNA factory" in which every co-transcriptional step is elaborately interdependent and coupled to each other seamlessly [165]. Initiation is coupled to 5'end capping, elongation is intertwined with splicing and termination is tethered to mRNA 3'end formation [166], but virtually every step of mRNA processing is interwoven, thus bestowing both feed forward and feedback coordination and including numerous checkpoints in between to ultimately maximize RNA production efficiency [167].

The dynamic and transient interactions of CTD-binding proteins also clearly emphasizes the co-transcriptional coupling of the different stages of the transcription cycle [22]. As the enzymes responsible for the numerous steps in mRNA processing are recruited to RNAPII CTD together with their substrates and co-localize, this not only increases the local concentration of these factors nearby the nascent RNA, but also enhances the efficiency and yield of their respective enzymatic reactions [167]. One of the most important functions of RNAPII CTD tail is precisely this co-localization [168], which is thought to ensure that RNAPII is competent to engage in the various mRNA processes as soon as they occur.

Transcription initiation and termination are linked through phosphorylation of TFIIB, which takes place after phosphorylation of RNAPII CTD Ser-5 [169]. TFIIB binds to mRNA 3'end processing proteins and the 3'ends of genes [89, 166, 168, 170]. Its mutation abrogates the assembly of a productive transcription complex and proper recruitment of mRNA 3'end processing proteins [169]. TFIID can also recruit mRNA 3'end processing proteins to the CTD tail [151, 171]. Additionally, transcription initiation is hindered through mutation of a pA signal [172].

mRNA 3'end processing proteins have been identified in promoters and the transcription initiation complex [151, 170, 171, 173, 174] as well as the RNAPII CTD tail [89, 151, 171, 175, 176]. mRNA 3'end formation is stimulated by RNAPII [91], affected by transcription initiation and elongation rates [120, 177, 178] and regulated by a plethora of proteins: the basal pre-mRNA 3'end machinery, but also by transcription and splicing factors [179, 180]. Phosphorylation of RNAPII CTD Ser-2 also promotes mRNA 3'end formation by early recruitment of termination factors [89, 93, 160] and mRNA 3'end processing proteins affect this mark at the transcription start site [161], hence linking mRNA 3'end formation to other transcription steps [181].

As previously mentioned, transcription termination promotes efficient mRNA 3'end processing [182], both coupled by the cleavage and polyadenylation events [183]. Cleavage also requires the RNAPII CTD [88]. Consequently, RNAPII disengagement from DNA together with its recycling, and that of other proteins (such as CDK9) from the 3'end regions and return to the promoter, is also dependent on proper mRNA 3'end formation [158], which consequently promotes the initiation of new rounds of transcription and enhances optimal gene expression [172].

2. Pre-mRNA 3'end processing

pre-mRNA 3'end processing and formation is viewed as a highly regulated step in transcription involved in gene expression and maintenance of important biological pathways [180, 184]. This process may also have a deleterious impact on human diseases if misregulated, such as cancer, thrombophilia, thalassemias, occulopharyngeal muscle dystrophy and neurodegenerative diseases [180, 185].

The core and auxiliary mRNA 3'end processing machinery is quite numerous and includes proteins of the transcription, splicing and translation machinery [181, 184, 186-189]. It is also well conserved between mammals and *Drosophila melanogaster* [190-192], with several paralogues of cleavage and pA factors found and characterized [193-199]. The molecular mechanisms behind the cleavage and pA events are reviewed in detail in [184, 187, 188, 200, 201]. Briefly, these two co-transcriptional processes begin with transcription termination by RNAPII and the binding of the Cleavage and Polyadenylation Specificity Factor (CPSF), the Cleavage Stimulation Factor (CSTF) and Cleavage Factor I (CFIm) to the newly transcribed pre-mRNA, which will sequentially recruit Cleavage Factor II (CFIIm), Symplekin and pA polymerase (PAPα). This basal multi-protein complex will then enable efficient cleavage and pA of the newly transcribed pre-mRNA [144, 173, 184, 186, 202-207].

Once properly capped at the 5'end and spliced out of introns [158], two co-transcriptional processes, the now stable and mature mRNA is exported to the cytoplasm and finally translated into a functional protein, which can undergo post-translational processing (reviewed in [208]).

2.1. Auxiliary regulatory sequences

A combination of both *trans*-acting proteins (such as CSTF, CFIm, CFIIm, Symplekin and PAPα for example) and *cis* auxiliary elements present in the pre-mRNA 3' untranslated region (UTR) operate synergistically to generate a ready-to-export mature mRNA [188, 209].

2.1.1. *cis*-acting elements

The hexameric sequence that comprises the pA signal is necessary for cleavage and pA of the pre-mRNA [210-212] that occurs 10-30 nt downstream at the cleavage site (CS, **FIGURE 3**), making it the most relevant *cis*-acting element in pre-mRNA 3'end processing. The most efficient pA [213] consists of the highly conserved [200] sequence AAUAAA for 48% of *Drosophila melanogaster* genes [214] and 53% of human genes [215] (**FIGURE 3**).

Computational analyses have shown that the most frequent variation of this canonical signal is the AUUAAA sequence, the second most efficient pA signal [213] present in 10% of fly genes [214] and 17% of human genes [215] (**FIGURE 3**). A particularly interesting pA signal is AAUAUA, which occurs in only 2% of mammalian genes [215], but it is as common as AUUAAA in *Drosophila melanogaster* [214, 216] (**FIGURE 3**), which indicates that it must be biologically relevant for this species. As different pA signal sequences have different efficiencies [213], this suggests that each may have different impacts on mRNA 3'end formation, mRNA and protein levels of the respective transcribed gene. Additionally, genes may or may not possess more than one functional pA signal, potentially generating several transcripts according to the usage of each pA signal [217], a mechanism denominated alternative polyadenylation (APA, further detailed in section 2.2).

Other *cis* regulatory sequences are found both up- and downstream of each pA signal (**FIGURE 3**). They can also affect the pA signal efficiency and overall mRNA 3'end processing [184, 188, 218-222]. These sequences have been consistently predicted by bioinformatic studies [223-226]; of note, *cis* elements are mostly absent near silent pA signals, indicating that they may assist in defining and regulating nearby pA signals [227]. These *cis* regulatory sequences often contain accessible binding sites for mRNA 3'end processing factors, RNA Binding Proteins (RBPs) and micro-RNAs [188, 228-231] (**FIGURE 4**). These sequences, together with the mRNA 3'end processing machinery and regulatory proteins, cooperate synergistically to define the efficiency of a particular pA signal [188, 232, 233], its usage [234] and enhance its recognition [173, 235].

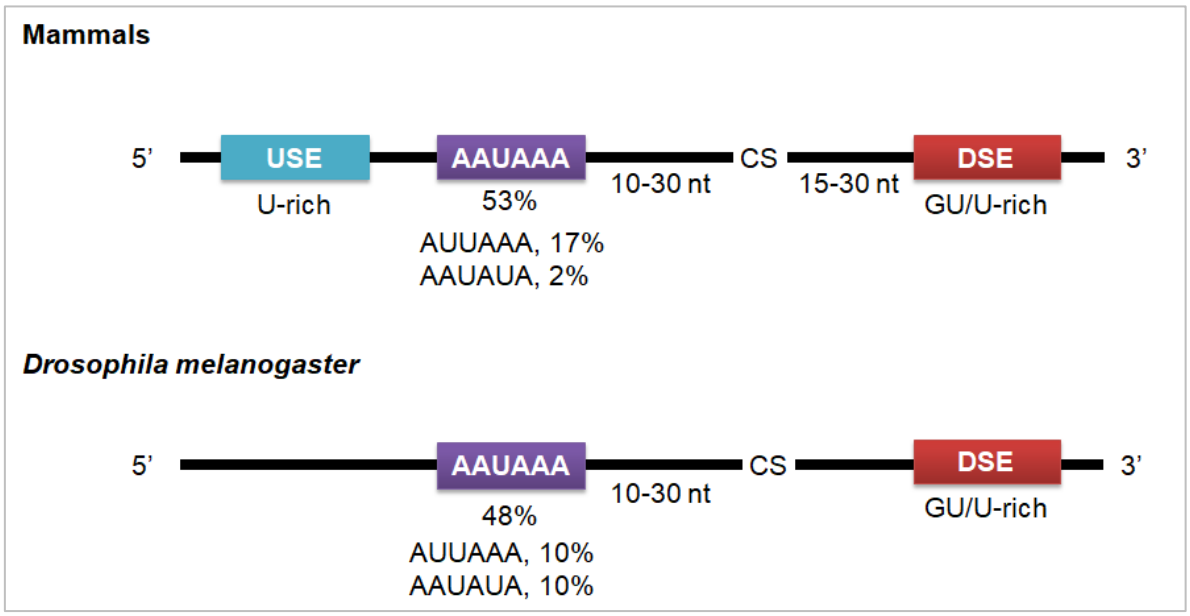


FIGURE 3 | *cis* acting elements that have key roles in pre-mRNA 3'end processing and formation. There are clear similarities in the machinery between mammals and

Drosophila melanogaster, such as the same canonical pA signal (AAUAAA) and common variations (AUUAAA and AAUAUA) and GU/U-rich downstream sequence elements (DSEs) 10-30 nt downstream of the cleavage site (CS) [214-216]. Upstream Sequence Elements (USEs) have often been reported in viral and human genes, but not in the fruit fly.

USEs are predicted to be widespread [223, 227, 236] and have been found in several viral and human genes: late pA signal of simian virus 40 [237, 238], adenovirus major late transcription units L1, L3 and L4 pA signals [239-242], ground squirrel hepatitis virus [243, 244], Human Immunodeficiency Virus Type 1 Long Terminal Repeat [245-248], *C2 complement* [249, 250], *lamin B2* [251], *collagen type I Alpha 1* [252], *collagen type I Alpha 2* [252, 253], *collagen type II Alpha 1* [252], *cyclo-oxygenase-2* (*COX-2*) [254-256], *prothrombin F2* [236, 257], *MECP2* [258], *ADD1* [259], immunoglobulin M (IgM) [260], *OAS1* [261], β-globin [262] and *JunB Proto-Oncogene* [263]. USEs are known to be generally U-rich [184], but without a consensus motif [219], poorly conserved and capable of regulating mRNA 3'end efficiency in a structure-, orientation- and position-dependent manner [184, 219, 264, 265] by serving as an extra platform for core mRNA 3'end processing factors [236, 256] or by recruiting more *trans*-acting elements [249, 250]. For example, the expression of the Ser protease thrombin encoded by the *prothrombin F2* gene is stimulated by a protein complex comprised of U2AF35, U2AF65 and PTBP1 bound to its USE in response to inflammation and stress [236] while its mRNA levels remain unchanged [266]. Moreover, the U2AF35 and U2AF65 proteins are involved in splicing, a process often linked with mRNA 3'end formation [267-270]. As such, the USE-dependent assembly of this class of proteins also suggests that these *cis* elements promote crosstalk between different mRNA processing steps [158].

Downstream sequence elements (DSE) also tend to be poorly conserved; nonetheless, they are present in approximately 80% of genes containing the AAUAAA or AUUAAA pA signals [271] and are normally U- or GU-rich [184, 205, 207, 227]. Additionally, they help define the cleavage and pA site [205, 263, 272] and influence pA signal efficiency [200, 263, 273] in a position- and distance-dependent manner [272, 274, 275]. DSE functionality also seems to depend on the nt composition and not its particular sequence [207]. In contrast to USEs, DSEs appear to associate more frequently with more efficient pA signals [227] and less efficient DSEs seem to be associated with proximal and suboptimal pA signals [216, 227]. Some examples include the adenovirus L3 and adeno-associated virus [273], rabbit angiotension converting enzyme [276], simian virus 40 late pA signal [275, 276], *MC1R* [277], IgM [260], *MC4R* and *JunB Proto-Oncogene* [263].

To our knowledge, there were no studies reporting the presence of USEs in *Drosophila melanogaster* until the work described in this thesis [278], but poorly conserved GU-rich DSEs exist [214, 216, 232, 279, 280]. The predicted distal DSE in the *CG11699* gene significantly promotes the usage of the canonical distal pA signal and controls *CG11699* expression [281]. Disruption of this DSE by a transposable element ablates the production of the *CG11699* long mRNA isoform, which in turn increases the expression of this gene by promoting the usage of the inefficient proximal pA signal.

Curiously, a recent study has revealed that only a small percentage of mRNA *cis* elements are occupied by RBPs *in vivo* [282], suggesting that there is a higher regulatory potential inherent to a large amount of transcripts that may be further explored by *trans*-acting proteins.

2.1.2. *trans*-acting elements

Correct mRNA 3'end processing requires the well-orchestrated participation of *trans*-acting multi-protein complexes [181, 184, 186-189] (many of them RBPs) whose components act cooperatively/competitively ([231, 283] and reviewed in [144, 221, 284-286]).

The core cleavage and pA factors are prime examples of *trans*-acting elements. As soon as RNAPII transcribes the pA signal, CFIm binds to UGUA sequences upstream of the cleavage site, promoting PAPα and CPSF binding [173, 206, 287]. CFIIm, which also contains PCF11, mediates the interaction between CFIm and CPSF [288]. The CPSF complex is comprised by six subunits of 30, 73, 100 and 160 kDa (CPSF30, CPSF73, CPSF100 and CPSF160, respectively) [203, 289], FIP1L1 [264] and WDR33 [181] and binds to the pA signal via CPSF30 and WDR33 [290, 291] and to PAPα through CPSF160 [292]. PAPα adenylation activity [289, 293-295], which adds approximately 200-300 adenines to the pre-mRNA 3'end, is stimulated by FIP1L1 [264]. CSPF73, CFIm and CFIIm are responsible for cleaving the pre-mRNA preferentially 3' of a CA dinucleotide [202, 288, 296-298]. The CSTF complex contains the 50, 64 and 77 kDa subunits (CSTF50, CSTF64 and CSTF77, respectively). CSTF64 binds to DSEs [205] and the CSTF77 subunit cooperatively interacts with CPSF160 to promote cleavage between the pA signal and this *cis* element [220, 292]. CSTF50 and CSTF77 interact directly with RNAPII CTD [176], which together with Symplekin, are thought to be scaffold proteins. Symplekin is also able to interact with CSTF64 [299]. Interestingly, the individual binding of these proteins and their interactions are generally stabilized and complemented by their combined assembly [273, 275], thus highlighting the plasticity and coupling of each step of

mRNA 3'end formation. The detailed molecular mechanisms of the process are reviewed in [184, 187, 188, 200, 201].

Other *trans*-acting proteins have regulatory roles over the mRNA 3'end processing mechanism while also participating in other cellular pathways, which may or may not be transcription-related. Embryonic Lethal Abnormal Vision (ELAVL1/Elav) is a well-studied RBP with functions in APA, mRNA stability and location, translation and splicing [300-304]. In *Drosophila melanogaster*, Elav modulates *hox* APA and splicing [305] and promotes 3'UTR extension [305-308]. Elav is also involved in the alternative splicing of *erect wing*, *armadillo* [309] and *neuroglian* [310] and mRNA 3'end formation of *rbp9* and *fne* [306]. The human polypyrimidine tract binding protein (PTBP1) is known to be involved in mRNA 3'end processing, alternative splicing, APA and translation of several genes [250, 256, 262, 266, 311-318]. Interestingly, PTBP1 competes with CSTF64 in the *C2 complement* DSE binding site to promote preferential choice of the distal pA signal [313], but PTBP1 can also recruit the heterogeneous nuclear ribonucleoprotein H [319] to the *C2 complement* DSE, which will in turn promote CSTF64 recruitment [273, 275, 320]. This indicates that PTBP1 is capable of both enhancing and inhibiting APA by interfering with the assembly of the mRNA 3'end machinery [250, 313]. Hephaestus (Heph, the *Drosophila melanogaster* PTBP1 paralogue) regulates embryo dorso-ventral patterning and germline-soma signalling [321-324], spermatid individualization [325, 326], the splicing and expression of transcripts encoding for proteins involved in spermatogenesis, such as Mlc1 [326], *oskar* alternative splicing, mRNA levels and translation [327] and Gurken protein location [322], but there is no known function in APA for this species.

The consensus binding motifs of RBPs are small and have a notoriously low complexity and limited diversity [328], i.e., they are generally GU- or U-rich, but with no defined sequence with the exception of the pA signal hexamer already mentioned. This suggests that the same or overlapping *cis* sequences are often targeted by different competing RBPs [226] that can interact with other *trans*-acting proteins either positively or negatively, leading to a dynamic and complex post-transcriptional regulation ([313, 329, 330] and reviewed in [331]). This indicates that each network of *trans* elements that transiently assemble and re-assemble on the *cis*-acting elements of each transcript is probably unique and highly plastic, depending on different stimuli or cellular context [329, 332-334], 3'UTR secondary/tertiary structure [282, 335-337] and expression [287, 338, 339] and/or post-translational modifications of RBPs and/or basal mRNA 3'end processing proteins [340, 341].

Together with their complementing *trans*-acting elements, the joint and highly ordered regulation of all *cis* sequences (efficient, inefficient or a mixture of both [232]) of a particular 3'UTR will consequently confer different properties and functions to the mRNA molecule [342], thus modulating its expression [335], decay [343, 344], stability [345-348], translation [330, 349-351], protein-protein-mRNA interactions [304, 352, 353] and cellular localization [304, 354-360] and will ultimately determine the fate of any given transcript (**FIGURE 4**).

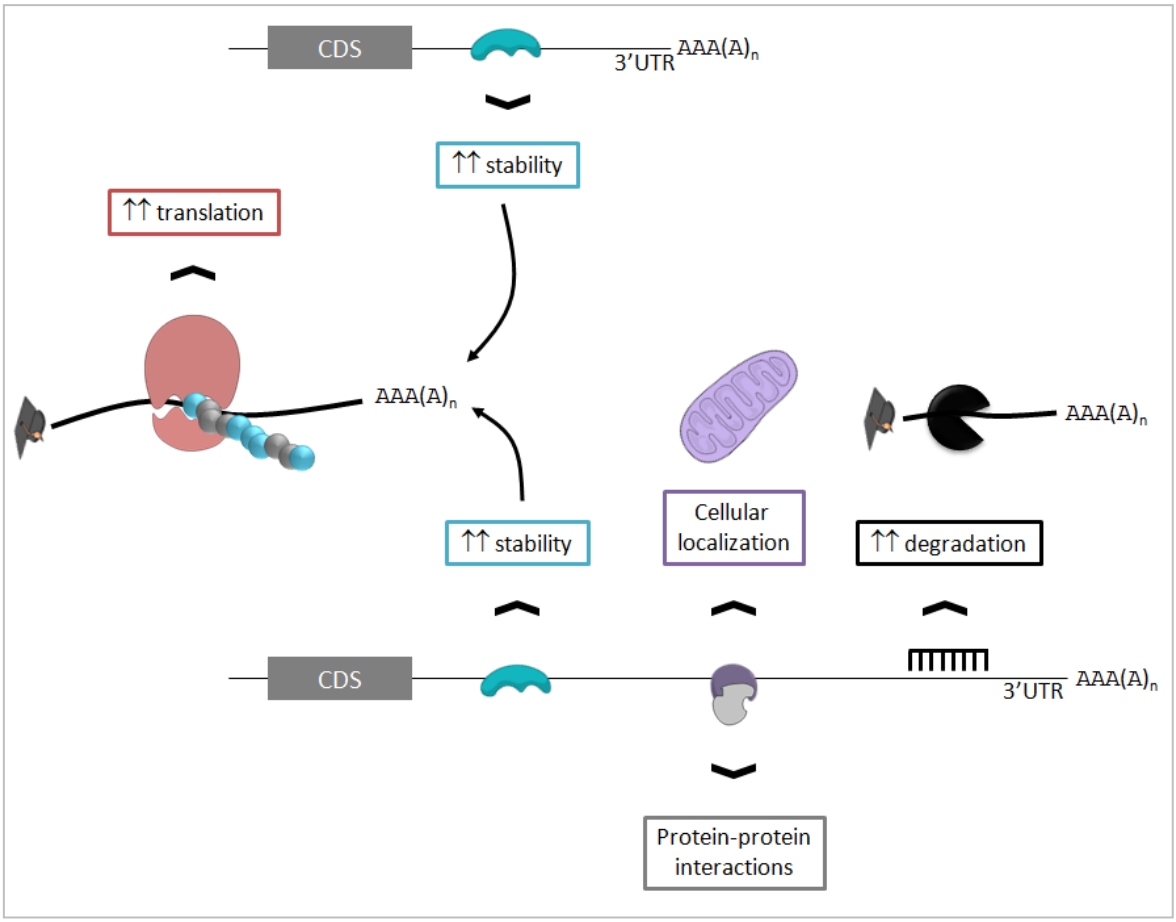


FIGURE 4 | Regulation of mRNA metabolism via the various *cis*- and *trans*-acting elements present in two 3'UTRs. Core and auxiliary proteins act cooperatively or competitively for *cis* element binding to aid in the efficient cleavage, pA and processing of the new pre-mRNA. An RBP that enhances transcript stability (in blue) increases its half-life, which in turn promotes its translation into protein. Other RBPs may lead the mRNA to specific subcellular locations (in purple) via protein-protein interactions (in grey). microRNAs (represented in black) may induce mRNA degradation via exonucleases. Adapted from [361].

For instance, during embryonic development, 71% of *Drosophila melanogaster* transcripts present spatially distinct patterns which correlate with the subcellular localization and

function of their respective proteins [362], indicating that mRNA-dependent localization is crucial from early developmental stages. *oskar* mRNA local translation at the posterior pole in *Drosophila melanogaster* oocytes promotes the posterior patterning of the embryo [363-365] and involves two independent *cis*-acting elements with different functions and different *trans*-acting proteins. This transcript is first transported from nurse cells to the oocyte by the interaction between the microtubule motor protein Dynein and a stem loop located in the *oskar* 3'UTR [366]. Once located in the oocyte, another interaction between a localization element in *oskar* 3'UTR allows its binding to another microtubule motor protein, Kinesin-1, which then transports *oskar* to the posterior pole where it will be locally translated [367, 368].

2.2. Alternative Polyadenylation

As previously mentioned, APA occurs when there is more than one functional pA signal along the gene, thus potentially producing several mRNAs according to the usage of each pA signal and contributing to the fine-tuning of gene expression control [217]. Contrary to alternative splicing, which is ubiquitous in humans, but atypical in *Saccharomyces cerevisiae* [369], APA is an evolutionarily conserved phenomenon across all eukaryotes [370], occurring in over 70% of human genes [215, 371, 372] and 78% of *Drosophila melanogaster* genes [280].

APA may take place in the coding region of a gene if pA signals are located along this sequence (coding region-APA). In coding region-APA, both the coding and non-coding sequences of the transcripts are different and they will likely synthesize different proteins that contribute to proteome diversity from a single gene (**FIGURE 5**, top panel) as already reported for alternative splicing [373, 374].

APA can also occur in the 3'UTR if there are pA signals located at this region (3'UTR-APA), which leads to the generation of mRNA isoforms with different 3'UTR lengths. In 3'UTR-APA, the various transcripts will produce the same protein as their coding capacity is unaffected (**FIGURE 5**, bottom panel), but they may have different mRNA metabolisms and regulate gene expression through the presence and/or absence of different *cis* regulatory elements (and respective *trans*-acting elements) in their different 3'UTRs [342], which may influence protein levels. Thus, transcripts with longer 3'UTRs have more *cis* regulatory sequences and are more likely to be post-transcriptionally regulated. The generation of diverse mRNAs with the same coding region but different 3'UTR lengths through APA could therefore result in extensive post-transcriptional mRNA 3'end regulation [201, 234, 285, 342, 361, 370, 375].

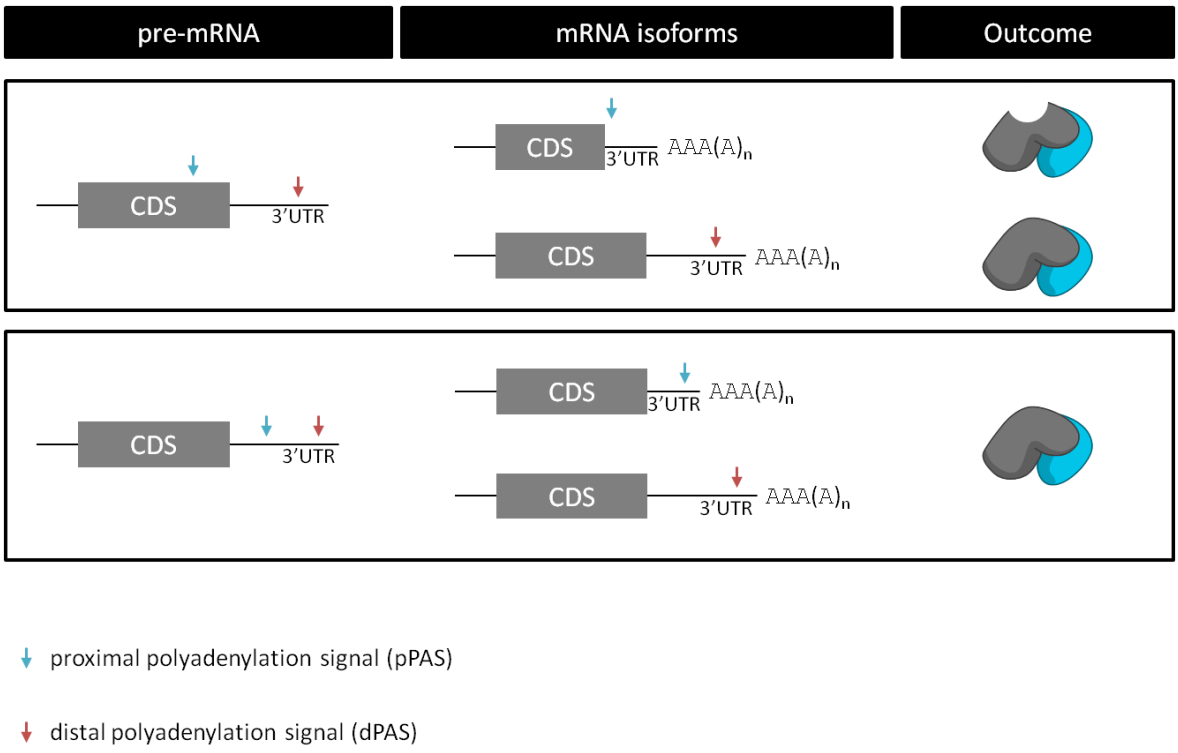


FIGURE 5 | Coding region-APA and 3'UTR-APA. Different proteins are produced when APA affects the coding region of the gene (coding region-APA, top panel). If the pA signals are in the 3'UTR, APA generates various mRNA isoforms that differ in the length of their 3'UTR, but the same protein is produced (3'UTR-APA, bottom panel). Adapted from [234].

Although the mechanism behind the usage of one pA signal over another is still unclear, the 'first come, first served' model is generally applied: proximal pA signals are chosen in detriment of distal ones because RNAPII and the mRNA 3'end processing machinery recognize them first and will assemble and cleave the transcript near the proximal pA signal instead of at distal pA signals [165, 220]. Curiously, distal pA signals are more prone to have canonical and therefore more efficient sequences than proximal pA signals, which tend to have variants of the AAUAAA hexamer [215, 263, 376] and a higher frequency of nearby DSEs [263]. Hypothetically, the less efficient proximal pA signals allow more flexibility for regulation and the influence of more *cis*- and *trans*-acting elements [207, 232] whilst distal pA signals ensure correct termination of the transcript [285]. Indeed, the lack of efficiency of a pA signal can be surpassed by the presence of *cis* elements [234, 263]. Additionally, this level of regulatory fine-tuning not only confers different means to control gene expression, but it also minimizes the chances of mRNA 3'end processing defects and/or overcomes the potentially fatal flaw of having one unique sequence for mRNA 3'end processing, pA and cleavage governed by a single enzyme as both could be more or less easily eliminated by mutation overtime [232].

As such, not only is APA deterministic, but it is linked to many important physiological events as reported by several genome-wide studies. Preferential choice of proximal pA signals is commonly associated with a proliferative cellular state [377, 378], including cancer [348]. In contrast, differentiated cells have the opposite trend, preferentially using more distal pA signals and producing transcripts with longer 3'UTRs [369, 372, 379-382]. APA is also known to be development- [371, 383, 384] and cell-type specific [261, 385, 386] and, similar to alternative splicing [387], it is also tissue-specific [255, 280, 324, 329, 369, 382, 388], a characteristic shown to be conserved within the same tissues from different species [279, 372].

While APA is widespread across eukaryotes, its genome-wide impact is still ambiguous, but it is a clearly important mechanism of gene regulation for many genes.

An example of coding region-APA in cellular differentiation is the immunoglobulin heavy chain class shift [389-391]. Plasma cells secrete IgM by selecting a proximal pA signal and immature B cells express choose a more distal pA signal that encodes the IgM form with a transmembranar domain [332]. This pA switch is also splicing dependent [392] and modulated by CSTF64 levels. CSTF64 levels are low in immature B cells, allowing the recognition of the more efficient distal pA. In contrast, plasma cells express high levels of CSTF64, promoting the usage of the proximal, but less efficient pA signal via the influence of the IgM DSE [338, 393].

COX-2 regulation by 3'UTR-APA depends of a tripartite USE upstream of its proximal pA signal and is tissue specific [255]. PTBP1, p54(nrb), PSF and U1A are *trans*-acting proteins that recognize the *COX-2* USE and modulate the mRNA 3'end formation and pA signal usage of this gene by promoting the production of the shorter *COX-2* transcript [256] despite *COX-2* pA1 (AUUAAA) presenting lower pA efficiency [213] than the canonical *COX-2* pA2.

In *Drosophila melanogaster*, the *trans*-acting RBPs CSTF64 and Sex-lethal share similar consensus binding sequences [207, 394] and compete to modulate *enhancer of rudimentary* 3'UTR-APA [395]. In somatic cells, CSTF64 promotes preferential choice of the inefficient proximal pA signal and protein production by binding to three GU-rich DSEs [329]. In the female germline, the female-specific Sex-lethal competes with CSTF64 for the three DSEs downstream of the proximal pA signal, promoting distal AAUAAA [395] signal usage and resulting in repression of protein translation of the transcript [329].

Another example of 3'UTR-APA and its biological consequences in *Drosophila melanogaster* is the *polo* gene, which encodes a key mitotic kinase [396, 397]. This gene

produces two mRNA isoforms through a tightly regulated differential usage of two pA signals [397]. Although the two mRNAs differ in their 3'UTR lengths despite producing the same protein, the longest *polo* isoform is more physiologically relevant as it is both necessary and sufficient for fly viability and known to produce the majority of Polo protein in the cell ([115] and **FIGURE 6**). Flies without this isoform die at the pupae stage due to a lethal failure in abdominal morphogenesis and flies with a mutated *polo* pA1, hence sustaining deficient *polo* APA, present an abdominal phenotype consistent with low levels of Polo protein.

polo APA regulators are still mostly unknown, but RNAPII elongation rate is one of them: there is a preferential shift to the proximal pA signal [115] in the slower RNAPII mutant fly, *RpII215*^{C4} [113, 398]. Polo protein levels alone also regulate *polo* APA. Overexpression of Polo in third instar larvae leads to a preferential usage of *polo* pA1 signal, producing the short *polo* mRNA isoform that produces low amounts of Polo ([115] and **FIGURE 6**). These data indicate that Polo protein levels regulate *polo* expression via an autogenous feedback loop by promoting *polo* pA1 choice when too much Polo protein is present in the cell. This type of auto-regulation is also a characteristic reported for several mammalian and fruit fly RBPs and their respective post-transcriptional mRNA regulation [399, 400].

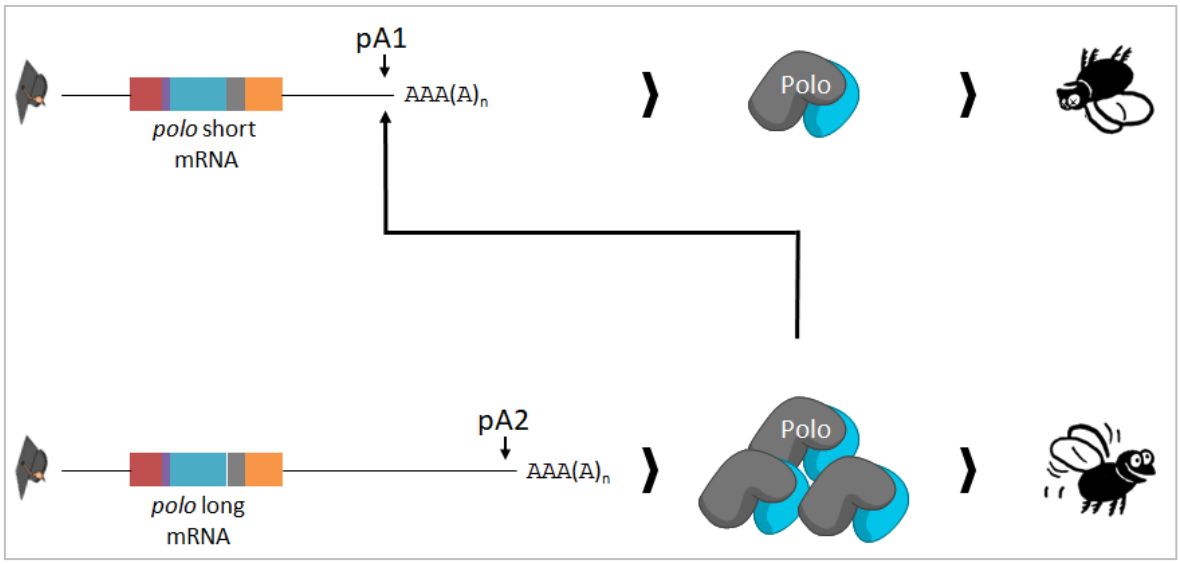


FIGURE 6 | *Drosophila melanogaster polo* APA regulation. The *short* polo isoform does not generate enough Polo protein to be compatible with life. Instead, most Polo protein is produced from the translation of the longest *polo* mRNA, but when Polo is overexpressed, the cell preferentially generates the shorter *polo* mRNA to compensate.

3. A Story about Polo

Polo is a cell cycle protein kinase involved in the regulation of cell division. Polo is conserved in all eukaryotes and its amino-terminal domain is most closely related to the catalytic domain of a Ser/Thr kinase [397, 401], but it also contains two Polo-Boxes at the carboxy terminal half which determine subcellular localization and regulate the catalytic kinase domain via auto-inhibition [402-406]. In addition, their unique sequence motif is responsible for recognizing phosphopeptides [407-409].

Polo was initially identified by screening of female sterile mutations along the *polo* locus that showed aberrant mitoses during embryonic [396] and larvae development [397]. Homozygous *polo*^{1} females, a kinase dead mutant originated by a single point mutation [396], are capable of laying eggs when crossed with homozygous males, however the embryos never cellularize. They are thought to be blocked during the syncytial mitotic cycles, which ultimately leads to their death [396]. It was therefore postulated that *polo*^{1} homozygous larvae reach adulthood because their heterozygous mother provided enough Polo protein for its development that even while it is gradually replaced by the mutant protein, it is sufficient for the organism to survive until later stages [396]. This indicates that this protein kinase is crucial at all stages of development in *Drosophila melanogaster* [397].

The null mutants *polo*^{9} and *polo*^{10} were characterized as the two strongest hypomorphic alleles for Polo [410]. Polo protein is barely detectable in either mutants' neuroblasts while the *polo*^{1} variant Polo levels are similar to those of wild type [411], and this depletion culminates in lethality at the third instar larval stage.

3.1. Phosphorylating: What, When and Where?

Expression of *polo* is abundant in tissues and developmental stages in which there is extensive proliferation [115, 412] and is especially high in syncytial embryos and adult females (as seen by the orange bars in **FIGURE 7**). This is a pattern that has been previously observed for other cell cycle-specific proteins [413-420]. However, during mitosis in the syncytial *Drosophila melanogaster* embryos, a cyclical activity increment far surpasses the amount of protein that can be accounted by *de novo* synthesis, implying that post-translational modifications most likely have an important role in regulating Polo activity [411, 421]. Accordingly, Polo protein does not appear to significantly decrease between cell cycles [422].

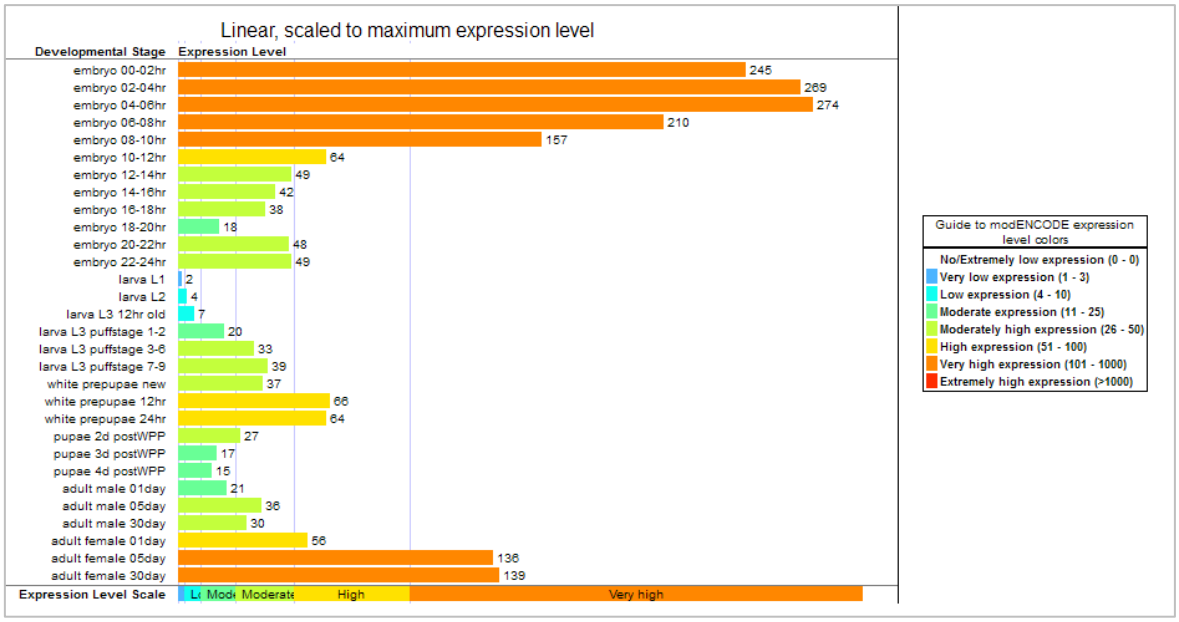


FIGURE 7 | modENCODE developmental expression data for *polo* by RNA-Seq [412]. Polo kinase is most expressed in early embryonic stages and adult females (orange coloured bars depict very high expression levels).

Polo has multiple roles during mitosis [423] and displays a highly dynamic localization pattern regulated in a cell cycle-dependent manner. Polo remains mostly in the cytoplasm during interphase and associates with condensing chromosomes and centrosomes at the onset of prophase, which is suggestive of a role in both the formation and maintenance of the bipolar spindle [397, 424]. As prometaphase begins, Polo plays an important role by accumulating at the kinetochores and assisting in kinetochore-microtubule attachments [425, 426]. During anaphase, it localizes to the spindle midzone [424, 427-429] in agreement with a specific association between Polo and microtubules [411]. Finally, during telophase, Polo remains concentrated along the post-mitotic bridge and midbody [427, 428].

Lack of Polo leads to a spindle assembly checkpoint (SAC)-independent mitotic arrest in the *Drosophila* S2 cell line [430] as co-depletion of Bub1-related kinase (BubR1) and Mad2 (both of which are part of the SAC complex [431-435]) did not release cells from their arrested state. In fact, inhibition of Polo led to a failure in Mad2 recruitment and assembly of the mitotic checkpoint complex, which are known to be Monopolar spindle 1 (Mps1) dependent [430, 436, 437]. This suggests that Polo is necessary for the SAC signalling pathway, but not for its activation. To further test this hypothesis, the authors observed a failure in Mps1 recruitment at unattached kinetochores after depleting Polo. The Mps1 kinase accumulates at kinetochores to reach its full activation state via *trans*-phosphorylation [438, 439], which in turn culminates in an efficient onset of SAC [432-

434]. In contrast, Polo localization and activation were unaffected after depletion of Mps1 or other components of the mitotic checkpoint complex [430].

A similar effect was shown upon Aurora B inhibition. Aurora B is necessary for effective activation of Polo by phosphorylation on its activation T-loop (Thr182) at the onset of mitosis at centromeres [440] and by promoting its dissociation from its inhibitor, Map205 [422, 441], among other roles [442-444]. In the absence of Aurora B, a constitutively active Polo mutant alone was able to correctly recruit key components of the mitotic checkpoint complex, such as Mps1, to kinetochores. This suggests that active Polo is responsible for this role in SAC signalling and that it does not need further action from Aurora B to perform its downstream functions. It is plausible that Polo and Aurora B are part of a positive feedback loop so as to ensure a correct mitotic checkpoint complex recruitment, which then leads to a full SAC activation due to the presence of Mps1 at unattached kinetochores [430].

Polo also co-localizes and co-immunoprecipitates with Hsp90, which increases protein stability and interactions by modulating its physiological conformation [445, 446]. Furthermore, Polo is destabilized by loss of Hsp90, decreasing by 90% in 48h [447]. The Hsp90 co-chaperone, Sgt1, also seems to be essential to this stabilization as lack of this protein derived the same effects as the loss of Hsp90 function, which suggests that Hsp90 probably requires Sgt1 to properly stabilize Polo [448].

3.2. Polo *versus* Polo-like kinases

Polo-like kinases (PLKs) are part of a highly conserved family of enzymes with a large variety of roles during the entire the cell cycle, most particularly during mitosis [397, 449-453]. Many paralogues in several eukaryotes have been described, with varying degrees of similarities to *polo* (**FIGURE 8**): the Cdc5 protein in *Saccharomyces cerevisiae* [454], Plo-1 in *Schizosaccharomyces pombe* [455], Plx-1 in *Xenopus* [456] and several mouse and human PLKs [449, 457-460].

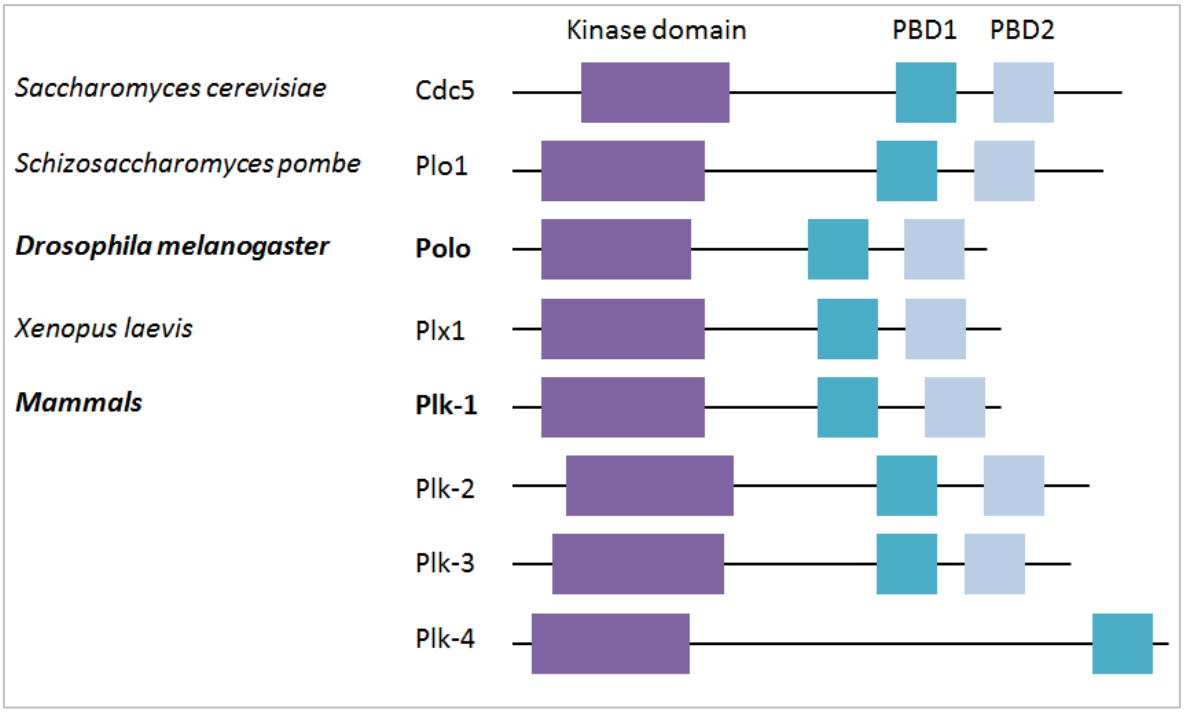


FIGURE 8 | Schematic representation of Polo-like kinases from several different eukaryotic species: *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Xenopus laevis* and mammals. PBD1 (in turquoise) and PBD2 (in light blue) refer to Polo-box domain 1 and Polo-box domain 2, respectively. PLK-5 is not included as it lacks a kinase domain [461]. PLK-1 (in bold) is the most similar paralogue to the original Polo kinase (also in bold). Adapted from [462].

PLKs undergo several dynamic changes regarding their abundance, activity and localization during cell cycle progression, most likely reflecting their different roles at different timings (**FIGURE 9** and [450, 463]). PLK activity increases during G2 phase and peaks at the onset of mitosis [427, 428, 459, 464] where they act upon centrosomes, spindle formation and contribute to both the activation and inactivation of the cyclin-dependent kinase 1 (CDK1)-cyclin B complex [421, 452, 465-469].

During interphase and early prophase, Polo-like kinase 1 (PLK-1) is involved in centrosome maturation and during prophase and metaphase, it also promotes microtubule nucleation at the spindle poles (**FIGURE 9** and [427, 470-472]). At this stage, the kinase co-localizes at kinetochores and appears to be a regulator of SAC components and the metaphase-anaphase transition (**FIGURE 9** and [473, 474]). At the onset of anaphase, PLK-1 is found within post-mitotic bridges (**FIGURE 9**), suggesting a role in cytokinesis [427, 428, 449, 475-480] and a substantial fraction of PLK-1 remains in the new G1 phase [481]. *PLK-1* expression correlates with its activity levels and is equally cell cycle-dependent: cells in interphase show basal expression which then starts to increase at the

end of the S phase and throughout the G2 phase, with a peak in mitosis [428, 458, 482, 483], suggesting that one of the key aspects of PLK-1 regulation is at the transcriptional level.

Recently, it has been shown that *PLK-1* expression is upregulated by heterogeneous nuclear ribonucleoprotein K (hnRNPK) and downregulated by two microRNAs (miR-149-3p and miR-193b-5p) via competitive binding to the same C-rich sequence in the *PLK-1* mRNA 3'UTR [484]. This dynamic post-transcriptional regulatory mechanism is also dependent on the levels of hnRNPK and each of the two microRNAs.

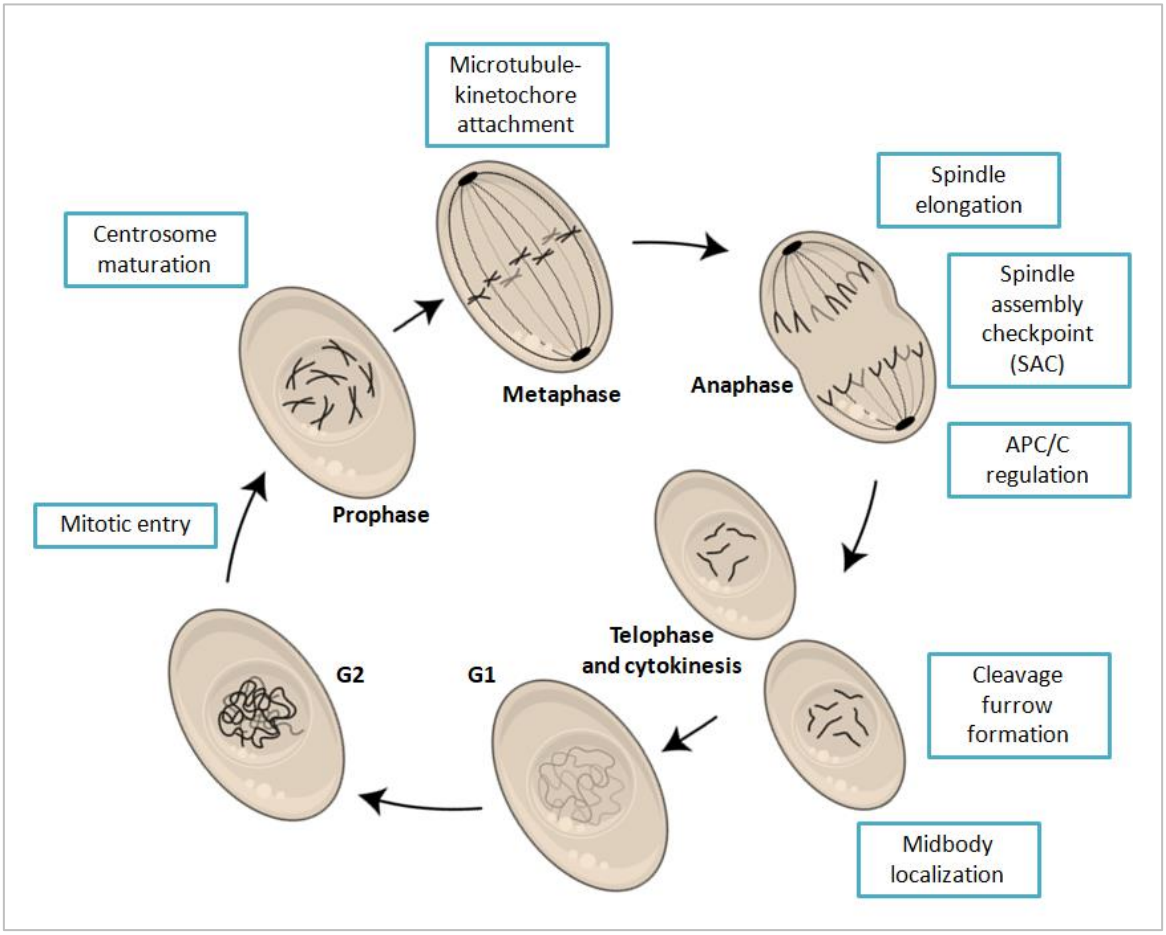


FIGURE 9 | The various functions of PLK-1 during mitosis. Adapted from [463].

Interfering with these kinases using antibodies or Polo paralogue mutations led to abnormal spindles in *Xenopus* [485], *Schizosaccharomyces pombe* [455] and *Saccharomyces cerevisiae* [454, 486], as well as with HeLa (immortalized) and Hs68 (non-immortalized) human cells [470], highly similar to the effect of the *polo*^{1} mutation in *Drosophila melanogaster* [396]. In the absence of PLKs, mitotic cyclins fail to be destroyed, indicating that they are important regulators of the anaphase-promoting complex/cyclosome (APC/C) [451, 465-468, 487, 488]. Nevertheless, it appears that even

in the absence of PLK-1, APC/C is still activated [426, 489]. As such, a backup mechanism must be present to compensate for the lack of phosphorylation by this kinase.

Interestingly, Polo kinase has a peak in enzymatic activity only during late mitosis [421], an observation that questions Polo requirement in the maintenance and/or formation of a functional bipolar spindle [397, 424]. However, this difference may not be incongruous: in *Drosophila melanogaster*, centrosome separation begins during telophase of the previous cell cycle, corresponding to the peak of Polo activity. This process occurs during prophase in mammalian cells, which also coincides with the peak of the PLK-1 enzymatic activity, the most similar PLK of the family in relation to Polo as shown in **FIGURE 8** [475].

In contrast with Polo, the role of PLK-1 in SAC signalling is not yet fully understood. Some authors [473, 490] state that PLK-1 is a SAC regulator as its downregulation significantly diminished the levels of kinetochore-associated proteins (such as Mad2), while others showed that inhibition or depletion of PLK-1 led to Mad2 and BubR1 accumulation as well as a SAC-dependent prometaphase arrest. This suggests that PLK-1 might be required at the onset of anaphase, but it does not seem to be essential for SAC [425, 426, 480, 491-495]. With the checkpoint silenced, PLK-1-depleted cells do exit mitosis, bypassing anaphase and cytokinesis [426, 480, 491].

Human PLK-1 is ubiquitinated in late mitosis [481, 496, 497]. In *Drosophila melanogaster* however, Polo kinase levels do not fluctuate significantly during cell division in the early embryo, although this difference might be due to the absence of the G1 phase in syncytial mitotic cycles [498]. Similarly to *Drosophila melanogaster* embryos, cyclical increase in PLK-1 protein activity during mitosis in cultured human cells also exceeds the amount of protein able to be produced *de novo*, implying that post-translational modifications have a focal role in regulating its activity [411, 421, 428]. Indeed, while PLK-1 protein levels increased four-fold after a G1/S phase arrest, its enzymatic activity increased about 26-fold [428].

The enzymatic activity of PLKs can be regulated by auto-inhibition via their Polo-box and kinase domains [408, 476, 499], SUMOylation [500], phosphorylation, either through other protein kinases [409, 428, 476, 501-508] or even auto-phosphorylation [427] or an antagonistic phosphatase [501, 506].

For optimal activation, PLK-1 requires the highly conserved Thr210 residue [498] to be phosphorylated [409, 509-512] which will induce a conformational change [502]: hindering this process results in the abrogation of PLK-1 activity [503]. Lee and collaborators [513] showed that by mimicking a constitutive phosphorylation on Thr210 in *S. cerevisiae*

increased the activity of the Cdc5 (the paralogue of PLK-1) protein by four-fold. Moreover, co-depletion of a phosphatase adaptor in PLK-1-depleted cells minimized the mitotic aberrations and resulted in the increase of Thr210 phosphorylation [506].

Two separate studies [503, 505] in mammalian cells unveiled that Aurora A, necessary for the G2/M transition [514], as well as an auxiliary protein, Bora, are both responsible for Thr210 phosphorylation. Mitotic entry was delayed and Thr210 phosphorylation was suppressed [503, 505] by depleting or inhibiting Aurora A alone. Bora knockdown delayed mitotic entry as well as Thr210 phosphorylation, but contrary to Aurora A, it failed to activate PLK-1 [505]. This implies that Bora is only responsible for promoting phosphorylation at this site, which then leads to the activation of the enzyme. Nevertheless, combining both Bora and Aurora A increased PLK-1 activity synergistically by seven-to-nine-fold, suggesting that binding of Bora controls the accessibility to PLK-1 Thr210 by Aurora A [503, 505] via a change in the conformation of PLK-1 where the Polo-box domain is inhibiting the catalytic domain [499, 513]. Although *Drosophila melanogaster* Bora has been described as required for proper Aurora A activation and performance [515], PLK-1 appears to be the single target of Bora in mammalian cells [503, 505]. Further analysis showed that both the kinase and Polo-box domains of Polo can interact directly with Bora. Interestingly, PLK-1 is also capable of regulating Aurora A by promoting Bora degradation [516, 517]. In the absence of Bora or Aurora A, PLK-1 is still activated in late mitosis [503]. As PLK-1 is necessary for cytokinesis following Bora degradation, it is plausible that Thr210 should be targeted once again by another kinase [498].

In *Drosophila melanogaster*, Polo is phosphorylated in Thr182 at its activation loop during early mitosis, which allows optimal activation and normal Polo functions at kinetochores. This action is catalysed by both Aurora B and Inner centromere protein (INCENP) at the centromere [440], where both of these proteins accumulate [518, 519] and control crucial cell cycle processes [520-524]. After Polo is recruited to the centromeres of chromosomes whose kinetochores are not under tension, it binds to INCENP and promotes its phosphorylation by Aurora B. Once fully activated, Polo locates at the outer kinetochores in prophase and prometaphase. Thus, INCENP may serve as a platform to link the roles of both Aurora B and Polo at the kinetochores [440]. Interestingly, the levels of Thr182-phosphorylated Polo are not affected at the centrosomes [440], suggesting that another kinase may be responsible for the activation of Polo at this location similar to what has been proposed for the Aurora A-independent phosphorylation PLK-1 requires during cytokinesis [498].

3.3. Polo-like kinases: more than just cell cycle regulators

Fairley and colleagues [525] demonstrated that overexpressing PLK-1 led to an increase of RNAPIII-transcribed tRNA and 5S rRNA levels. Considering the difference in the overall enzymatic activity during cell cycle progression, PLK-1 was added to extracts of cells arrested either in the S phase or mitosis. tRNA gene levels increased in the S-arrested cells, but not in mitotic cellular extracts. These results suggest that PLK-1 is involved in the transcription of these genes, hence RNAPIII regulation, in a cell cycle-dependent manner. Overexpression of a subunit of the Transcription Factor for RNAPIII B (TFIIIB), Brf1, also resulted in an increase in tRNA and 5S rRNA levels in interphase cells caused by PLK-1-dependent phosphorylation, which enhances Brf1 transcriptional activity. In mitotic cells, PLK-1 phosphorylates another site on Brf1, repressing this effect. These results suggest that, according to the level of PLK-1 activation, this kinase has different effects upon RNAPIII regulation.

It has also been demonstrated that RNAPII CTD Thr-4 is phosphorylated by PLK-1 specifically during mitosis and that this hyperphosphorylated exclusively binds to centrosomes and the midbody [128], known locations of PLK-1 in metaphase and anaphase [427, 428, 470-472, 475, 480]. Cell cycle progression in RNAPII CTD Thr-4 mutants is compromised, suggesting that the PLK-1-dependent phosphorylation of RNAPII CTD Thr-4 and its tethering to centrosomes are a requirement for normal mitosis [128]. However, even though Thr-4-P is a modification mark associated with an elongation-committed RNAPII common at the 3'end of genes and proven to be essential for cell survival [30] and mitotic progression [128], the mechanism behind this recruitment to the centrosome and its physiological role is still unclear.

More functions for PLK-1 outside of mitosis have emerged. Plo-1, the Polo paralogue of fission, yeast can control its own expression, along with other cell cycle genes [526]. PLK-1 directly activates the transcription factor Forkhead Box M1 (FoxM1) needed for the correct expression of several cell cycle regulators such as PLK-1 itself [527-529]. It also inhibits the anti-proliferative transcription factor Forkhead Box O1 (FOXO1) and the respective transcription of its pro-apoptotic target genes [530]. Inhibiting PLK-1 enhances non-viral transgene expression in cancer cell lines [531]. PLK-1 phosphorylates the CCCTC-Binding Factor, consequently affecting the expression of many genes [532]. Interestingly, it has also been reported that PLK-1 binds to the same binding region in RNAPII CTD tail as cyclin T [533], a subunit of the P-TEFβ, thus possibly affecting RNAPII transcription elongation. CDK9 can also be separated from its cyclin by PLK-1-mediated

phosphorylation of the latter [533], suggesting a link between cell cycle-dependent transcriptional silencing control and promoter proximal pausing.

With growing evidence that PLK-1 directly phosphorylates the RNAPII CTD tail [128], regulates RNAPIII activity [525] and modulates the activity of some genes [526-532], it is plausible that *Drosophila melanogaster* Polo may also phosphorylate RNAPII, thus affecting its activity, APA and/or transcription in general (**FIGURE 10**), especially considering that Polo levels alone regulate *polo* APA [115], a co-transcriptional event in which RNAPII plays a part [115, 153, 280, 534, 535].

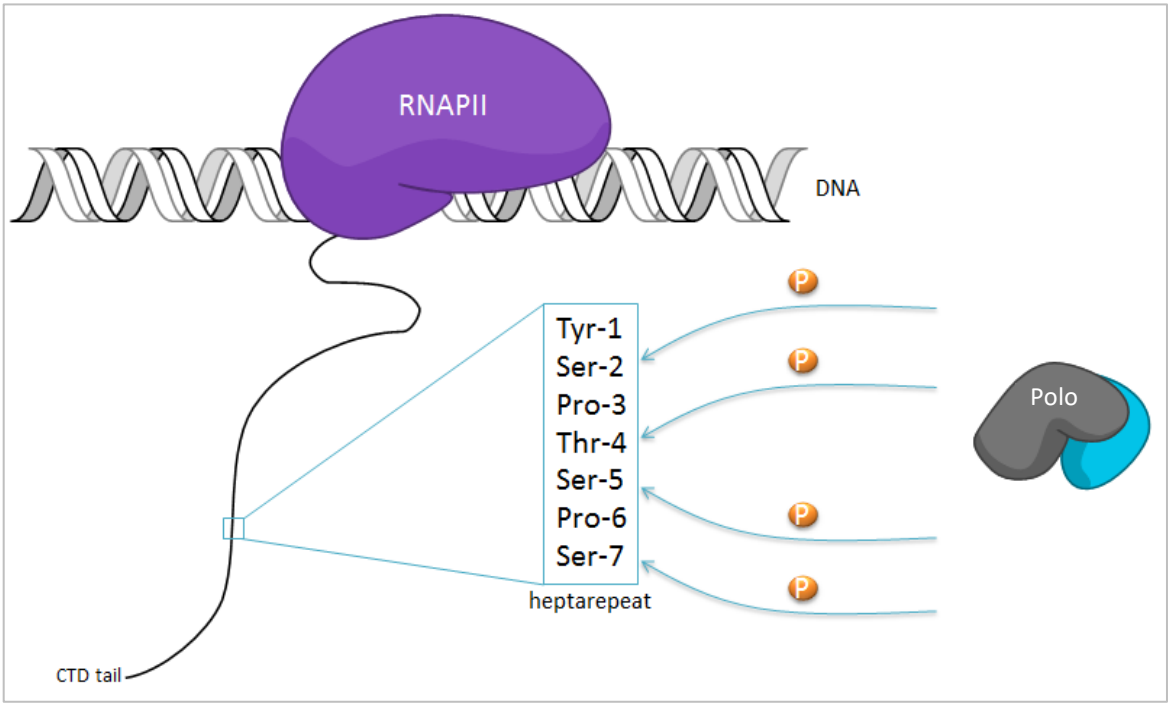


FIGURE 10 | A possible role of Polo in post-translational modifications on the heptapeptide repeat of RNAPII CTD tail. We hypothesise that like PLK-1, Polo may phosphorylate one or more amino acids of RNAPII CTD tail, in particular Ser and/or Thr.

4. Aims

To our knowledge, the genome-wide effect of a slower elongation transcription rate in *Drosophila melanogaster* mRNA 3'end processing has not been yet investigated. In addition, mRNA 3'end processing of the fundamental Polo kinase has been poorly described and its role upon transcription and beyond mitosis has rarely been the focus of the latest research. This project aims to study APA in *Drosophila melanogaster* at the genome-wide level, to unveil the molecular mechanisms behind the regulation of *polo* APA and gene expression, and to investigate a possible new function for Polo. Specifically:

1. To describe the genome-wide impact of a slower RNAPII transcription elongation rate on APA;
2. To disclose putative *cis* regulatory elements and *trans* acting factors involved in polo APA;
3. To investigate if Polo has a function in RNAPII CTD phosphorylation, transcription and APA.

MATERIALS AND METHODS

1. *Drosophila melanogaster* animal model

1.1. Stocks and maintenance

The *RpII215*^{C4}, carrying the RNAPII *C4* allele [114], and the *heph*^{2} hypomorph mutant flies were obtained from the Bloomington *Drosophila* Stock Center (Indiana University). The *heph*^{2} mutant was balanced with TM6B. *polo*^{9}*/TM6C* flies were kindly provided by David Glover (Department of Genetics, University of Cambridge). Wild type *w*^{1118}, and *polo* kinase dead mutant *w*^{1118}*;If/CyO*; *polo*^{1} *eB/TM6B* flies were kindly provided by Claudio Sunkel. The transgenic *Mz1061-Gal4;gfp-polo;polo*^{9}*/TSTL* (also referred to as *gfp-polo;polo*^{9}/TSTL) as well as *w*^{1118};*gfp-polo*Δ*pA1*;*polo*^{9}/TM6B (referred to as *gfp-polo*Δ*pA1*;*polo*^{9}/TM6B) flies were previously described [115, 448]. All stocks were grown at either 18 or 25°C using standard culture conditions and media, with or without antibiotics.

1.2. Generation of *gfp-poloΔUSE;polo*$^{9-/-}$ flies

The transgenic strain *w*^{1118};*gfp-polo*Δ*USE1*;*polo*^{9}*/TM6B* was created by Pedro Pinto and obtained by injecting *w*^{1118} embryos with the respective transgene according to [536] and selected by mating the transgenic flies with a strain carrying a balancer chromosome and dominant markers, *w*^{1118};*Sco/SM6*. These were then mated with *w*^{1118};*If/CyO*;*MKRS/TM6B* and *w*^{1118};*If/CyO*;*polo*^{9}*/TM6C* to generate the *w*^{1118};*gfp-polo*Δ*USE*;*polo*^{9}*/TM6B* (also referred to as *gfp-polo*Δ*USE*;*polo*^{9}/TM6B) strain. Two viable and fertile homozygous lines were obtained with the transgene inserted on the second chromosome.

1.2.1. *gfp-polo*Δ*USE*;*polo*$^{9-/-}$ abdomen phenotype analysis

gfp-poloΔUSE;polo$^{9-/-}$ individuals were identified by the Hu marker and every abdominal defect in comparison to heterozygous individuals was considered as an abnormality. These defects included fewer or missing bristles and missing, malformed or nicked tergites in both males and females.

1.2.2. *gfp-polo*Δ*USE*;*polo*$^{9-/-}$ abdomen preparation

Flies were dissected between the thorax and the abdomen, carefully removing all appendages. Abdomens were then incubated in a lactic acid:ddH_2O (3:1) solution and incubated overnight at 60ºC, mounted in fresh lactic acid:ddH_2O (3:1) solution and incubated overnight in a 60ºC oven.

1.3. *Drosophila melanogaster* embryo collection

Adults were placed in a collection cage with apple juice-agar plates supplemented with yeast. They were incubated for 24 h at 25°C, then the 0-24 h embryos were washed with deionized water and exposed to 50% bleach with rotation for no more than 5 minutes to remove the chorion. Dechorionized embryos were then washed with Phosphate Buffer Saline (PBS) and transferred to an eppendorf to be further processed or stored at -20°C.

1.4. *Drosophila melanogaster* third instar larvae brain collection

Wandering third instar transgenic or mutant larvae were collected and compared to w^{1118} larvae to ensure homozygous individuals were obtained. After 4-60 individuals were selected from each strain, brains were dissected using tweezers and transferred to an eppendorf with either PBS or TRIzol (Ambion) to be processed immediately after or stored at -20°C.

2. Protein Analyses

2.1. Western blotting

Drosophila melanogaster embryos or 20 third instar larvae brains were homogenized in 50 mM trishydroxymethylaminomethane (Tris)-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 10 glycerol, 50 mM NaF, 5 mM sodium pyrophosphate, 1 Triton-X 100, 1 mM 1,4-dithiothreitol (DTT, Thermo Fisher Scientific), 0.1 mM phenylmethylsulfonyl fluoride (PMSF, Sigma-Aldrich), 1:100 Na_3VO_4 (Merck) and protease inhibitor cocktail (Roche or Sigma-Aldrich). The extracts were incubated for 20-30 minutes at 4˚C with rotation and centrifuged at 8000 rotations per minute (rpm) for 5 minutes to remove debris. Total protein concentration was assessed by the Bradford protein assay (Bio-Rad Protein Assay Dye Reagent Concentrate). All gels were loaded with 30-50 µg of protein extract per well.

Western blots for Polo detection were separated in 7.5% (resolving) and 5% (stacking) Tris-glycine sodium dodecyl sulfate (SDS)-polyacrylamide (Bio-Rad) gels at 120 V for 1h30 and transferred to a nitrocellulose membrane (Novex, ThermoFisher Scientific) with the iBlot Gel Transfer Device (Invitrogen) using the P3 parameters for 7 minutes. After 1h blocking in 5% non-fat dried milk in Tris-Buffered Saline (TBS, Merck) and 0.01% Tween20 (VWR), membranes were incubated overnight at 4ºC with mouse anti-Polo antibody (MA294, 1:40, kind gift from Claudio Sunkel) and mouse anti-α-tubulin DM1A (1:20000, Sigma-Aldrich) and then incubated with goat anti-mouse immunoglobulin G-horseradish peroxidase (IgG-HRP) (1:20000, Santa Cruz Biotechnology) in 5% non-fat

dried milk in TBS and 0.01% Tween20 for 1h. Signal was detected with the Amersham enhanced chemiluminescence (ECL) Prime Western Blotting Detection Reagent (GE Healthcare), the Amersham Hyperfilm ECL (GE Healthcare) and the Fuji Medical Film Processor FPM-100A together with the Anatomix developer and X-Fix fixer (Fujifilm Europe GmbH). Protein semi-quantification was performed using the Molecular Imager GS-800 Calibrated Densitometer (Bio-Rad) and analysed in the Quantity One software (Bio-Rad).

Western blots for detecting RNAPII modifications were performed pre-cast using NuPAG 3-8% Tris-Acetate SDS-polyacrylamide gel electrophoresis (PAGE) Gels (Life Technologies) at 150 V for 1h30 and transferred to a nitrocellulose membrane (Bio-Rad) overnight at 4ºC at 250 V and 350 mA. The membrane was then blocked for 1 h in 5% non-fat dried milk in blocking buffer (10 mM Tris HCl pH 8, 150 mM NaCl and 0.1% Tween20) and incubated with mouse RNAPII CTD Ser-5-P (4H8 clone, 1:200000, BioLegend) and rat RNAPII CTD Ser-7-P (1:10, kind gift from Dirk Eick, German Research Center for Environmental Health, Munich) and mouse anti-α-tubulin DM1A (1:20000) for at least 2h. After 1 h washing in blocking buffer, membranes were incubated with goat anti-mouse or anti-rat IgG-HRP (1:2500 and 1:10000, respectively, Jackson ImmunoResearch) for 1 h. Signal was detected with Amersham ECL Western Blotting Detection Reagent (GE Healthcare).

2.2. Immunofluorescence assays in third Instar Larvae Brain Squashes

Immunofluorescence assays were prepared with the help of João Barbosa (Cell Division & Genomic Stability, i3S, University of Porto). Briefly, dissected larvae brains from each strain (see section 1.4) were placed in fresh 50 µM colchicine in PBS for 1h30 at 25ºC and then fixated for 5 minutes with 1.8% formaldehyde and 45% acetic acid. Each fixated brain was squashed three times onto a slide, then frozen in liquid nitrogen. Brain squashes were incubated with absolute ethanol for 10 min, then 10 min with PBS 0.1% Triton X-100 and a 10 minute wash with PBS. Fixated brains were blocked with 10% fetal bovine serum (FBS) in PBS-T 0.05% for 1h at room temperature and then incubated overnight at 4ºC in the same solution with the following antibodies: rat anti-Spc105 (1:150), mouse anti-Polo (1:10), rabbit anti-P-Aurora (1:500, Rockland, Limerik, PA), rabbit anti-Thr676-P-Mps1 (1:10000) and ginea pig anti-total Mps1 (1:250, Gp15 RRID:AB_2567774). This was followed by three 5 minute washes in PBS 0.05% Tween20 with gentle agitation. Slides were then incubated with the following secondary antibodies (Thermo Fischer) diluted 1:250 in PBS 0.05% Tween20 for 2h at room temperature: goat anti-rat AlexaFluor© 647, goat anti-mouse AlexaFluor© 488, goat anti-rabbit AlexaFluor©

568 and goat anti-ginea pig AlexaFluor© 568 and washed three times in PBS 0.05% Tween20 with gentle agitation for 5 minutes. Vectashield with DAPI was then added.

Image acquisition was performed using a Laser Scanning confocal microscope Leica TCS SP5 II and the LAS 2.6 software (Leica Microsystems, Germany). Image analyses were performed using Fiji [537]. Neuroblasts were identified as the larger cells based on the background signal from the different antibodies used that made the cell shape visible. For immunofluorescence quantification, each individual kinetochore was detected by their Spc105 mean fluorescence intensity and a region of interest (ROI) that befitted every kinetochore in each cell was defined using the maximum projection fluorescence intensity images. The same ROIs per cell were then used to detect the mean fluorescence intensity for Polo, P-Aurora and P-Mps1 and these values were then normalized to Spc105. Background was defined by the average of the mean fluorescence intensity of several other regions inside the cell free of kinetochores and subtracted to each mean fluorescence intensity. Control values were averaged and used for normalization of values determined in the different biological conditions tested. Quantifications were done using GraphPad Prism 7.0c (La Jolla California USA, www.graphpad.com).

2.3. Aneuploidy immunofluorescence in third Instar Larvae Brain Squashes

Dissected third instar larvae brains with the *gfp-polo;polo*$^{9-/-}$, *gfp-poloΔUSE;polo*$^{9-/-}$, *w*^{1118} and *heph*^{2}*/TM6B* genotypes were squashed and incubated with colchicine as previously described to obtain chromosome spreads that facilitate the visualization of chromosome content. Spc105 was used as a kinetochore reference.

2.4. Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were performed as previously described [538], with a few adaptations. Homogenized and dechorionated 0-24h *Drosophila melanogaster* embryos were cross-linked for 30 minutes with rotation using 1% formaldehyde (Panreac or Sigma-Aldrich). The reaction was quenched using 1 M glycine (Merck or Fisher BioRegeants) for 5 minutes with rotation and the cells were washed twice using cold PBS to remove all traces of formaldehyde, glycine and media.

After the last wash, cells were resuspended in lysis buffer (1% SDS, 50 mM Tris-HCl pH 8.1, 10 mM EDTA pH 8) with EDTA-free protease inhibitor (Roche) and phosphatase inhibitor (Sigma-Aldrich) cocktails and incubated for 20 minutes at 4°C with rotation. Cell lysates were then sonicated 20 times at 4°C (Bioruptor, Diagenode, 30/30 cycles) and

centrifuged at 13200 rpm for 1 minute at 4°C to remove cellular debris. An input aliquot was taken at this stage.

Input samples were diluted with 337 mM NaCl, 14.5 mM Tris pH 8.1, 0.9% Triton X-100, 1.0 mM EDTA pH 8 and 0.01% SDS and heated overnight at 65°C to revert the cross-linking, then incubated for 1 h with 9 mM EDTA pH 8, 3.6 mM Tris-HCl pH 6.8 and 36 µg/mL proteinase K (NZYTech) at 45°C. Input DNA purification was performed using phenol-chloroform (Sigma-Aldrich) at 4°C, then two 5 minutes centrifugations at 16100 x *g* and the chromatin was precipitated for 4 h with absolute ethanol, 150 mM $NaCH_3CO_2$ pH 5.2 (Merck) as well as 1 µL glycogen (Ambion) at -80°C. Samples were further centrifuged at 4°C for 20 minutes at 16100 x *g*, washed with 70% ethanol and resuspended in DNase/RNase-free HyClone water (GE Healthcare). To test for chromatin fragmentation and quality, inputs from each sample were separated in 1.5% *SeaKem* LE agarose (Lonza) gel electrophoresis and the ones with a chromatin smear between 200 and 500 base pairs (bp) were considered.

Samples were then diluted two fold with ChIP dilution buffer (167 mM NaCl, 16.7 mM Tris pH 8.1, 1.1% Triton X-100, 1.2 mM EDTA pH 8 and 0.01% SDS) and pre-cleared for 1 h at 4°C with rotation using previously washed protein A sepharose beads 4B Fast Flow from *S. aureus* (P9424, Sigma-Aldrich). These beads were discarded after a 1 minute long centrifugation at 1200 x *g* and the RNAPII IP (4 µL of Rpb3, a kind gift from John Lis, Cornell University, US) was performed overnight.

Washed protein A sepharose beads were then added to each IP for at least 4 h at 4°C with rotation before being washed with low salt buffer (150 mM NaCl, 2 mM Tris pH 8.1, 0.1% SDS, 1% Triton X-100 and 2 mM EDTA pH 8), high salt buffer (500 mM NaCl, 20 mM Tris pH 8.1, 0.1% SDS, 1% Triton X-100 and 2 mM EDTA pH 8), LiCl salt buffer (0.25 mM LiCl (Sigma-Aldrich), 10 mM Tris pH 8.1, 1% NP-40, 1% sodium deoxycholate and 1 mM EDTA pH 8) and finally Tris-EDTA buffer (10 mM Tris pH 8.1 and 1 mM EDTA pH 8). Chelex-100 (Sigma-Aldrich) was added to the immune complex, which was then boiled and incubated for 45 minutes at 55°C with 36 µg/mL proteinase K and boiled once again. Samples were centrifuged at 13400 x *g* for 1 minute and the beads washed one last time with DNase/RNase-free water (HyClone) to collect any remaining chromatin.

2.5. Chromatin Immunoprecipitation-sequencing analyses on Upstream Sequence Element-N-containing genes

RNAPII ChIP-seq was previously published in [539] and accessible through GEO (GSE20000). Datasets were mapped as described in the original work against the dm6

version of the *Drosophila melanogaster* genome. The genomic location of mapped reads was compiled using custom scripts and visually examined using the UCSC Genome Browser in bedGraph format. ChIP-seq hit locations were filtered based on fragment length. The ChIP-seq datasets were binned in 25 bp windows for visualization in bedGraph files.

2.6. RNA-protein pull-down assay

1 µg of oligonucleotides containing a T7 anchor (GTAATACGACTCACTATAGGG) followed by the wild type or mutated USE (USE RNA or USEmt RNA, NZyTech) were used as DNA templates for *in vitro* transcription using the MEGAScript T7 transcription kit (Thermo Fisher Scientific) and the Magnetic RNA-Protein Pull-Down Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Briefly, 50 pmol of the *in vitro* transcribed RNAs were biotinylated and bound to streptavidin magnetic beads for 30 minutes with agitation. The complex was incubated with 200 µg of *Drosophila melanogaster* 0-24h embryo protein extract for 45 minutes at 4˚C with agitation (RNA-protein binding buffer included 20 mM Tris pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$ and 0.1% Tween20), then washed before the bound proteins were eluted from the RNAs with agitation. Half of the eluate was separated in 12% (resolving) and 5% (stacking) Tris-glycine SDS-polyacrylamide gel at 120 V for 1h30 and stained with BlueSafe (NZyTech) for 1h to determine protein recovery. The other half was processed for identification by mass spectrometry.

2.7. Mass Spectrometry and Gene Ontology analyses of Upstream Sequence Element RNA Binding Proteins

Protein eluates were first reduced, alkylated and digested with trypsin and then separated by a 15 cm liquid chromatography (LC) C_{18} column with a 2h run and eluted into electrospray ionization high-resolution accurate-mass Orbitrap mass spectrometer (MS, Q Exactive, Thermo Fisher Scientific). Results were acquired in data-dependent positive acquisition mode alternating between a full scan and subsequent HCD MS/MS of the 10 most intense peaks from full scan. Protein identification was performed based on the 8634 *Drosophila melanogaster* proteins classified as mRNA binding in UniProt database [540] using the Proteome Discoverer 2.2 software (Thermo Fisher Scientific). The Precursor Ions Quantifier node of Proteome Discoverer software calculated the protein abundances of the samples, normalized the peptide groups and scaled them. Normalization was based on total peptide amount by summing the peptide group abundances for each sample and determining the maximum sum within a sample. The sample with the highest abundance

is taken as a reference and the abundance values of the other samples are corrected by a constant factor per sample so that at the end the total abundance is the same for all samples. Proteins identified by five or more unique peptides according to Proteome Discoverer were considered reliable hits and their accession codes were uploaded onto the PANTHER platform [541, 542] (Version 13.1, released on February 3rd 2018) to analyse their GO molecular functions.

3. mRNA Analyses

3.1. RNA extraction

0-24h embryos, third instar larvae brains or whole adult fly RNA extractions were performed using the same protocol. For each strain, total RNA of 20 third instar larvae brains was extracted with Trizol (Thermo Fisher Scientific) according to the manufacturer's protocol. RNA integrity was analyzed either using an Experion RNA StdSens Analysis Kit (Bio-Rad) with an Experion Automated Electrophoresis System or in a 1-1.5% agarose gel electrophoresis.

3.2. cDNA synthesis

Complementary DNA (cDNA) was synthesised according to the manufacturer's instructions, using 100 U of SuperScript III or SuperScript VI reverse transcriptase (Thermo Fisher Scientific), 0.5 mM denucleotide mix (dNTPs, Thermo Fisher Scientific), 2.5 µM random hexamers (5'-OH dN_6, Sigma-Aldrich), First-Strand Buffer/SSIV Buffer (Thermo Fisher Scientific), 5 mM DTT and 20 U of RiboLock. 0.5-1 µg of total RNA was used and each reaction was complemented with a negative control, RT^-, to which no reverse transcriptase was added.

The SuperScript III synthesis program using a Biometra thermocycler went as follows: 65°C for 5 minutes, 4°C for 5 minutes, 25°C for 5 minutes, 50°C for 1 h and 70°C for 15 minutes. The SuperScript IV synthesis program went as follows: 65°C for 5 minutes, 4°C for 5 minutes, 23°C for 10 minutes, 50°C for 10 minutes and 80°C for 10 minutes. Samples were then stored at -20°C.

When deemed necessary, DNAse treatment was also performed. 0.5-1 µg of total RNA was treated with 2.4 U of DNase I recombinant and DNase I Incubation Buffer (Roche) for 30 minutes at 37°C. Enzyme inactivation was achieved by adding EDTA pH 8 up to 8 mM and heating samples up to 75°C for 10 minutes.

3.3. 3'Rapid Amplification of cDNA Ends

cDNAs for 3'Rapid Amplification of cDNA Ends (3'RACE) were synthesized using SuperScript IV Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions, with 2.5 µM 3'RACE adapter primer (GCGAGCACAGAATTAATACGACTCACTATAGGT15VN) and 0.5-1 µg of total RNA from w^{1118} and *heph*^{2}*/TM6B* mutant embryos and *gfp-polo;polo*^{9}*/TSTL* and w^{1118};*gfp-polo*Δ*USE*;*polo*$^{9-/-}$ female adult abdomens. The 3'RACE polymerase chain reaction (PCR) was performed using GoTaq DNA Polymerase (Promega) according to the manufacturer's instructions, an anchor primer (which annealed to the 3'RACE adapter primer, GCGAGCACAGAATTAATACGACT) and a specific primer on the coding sequence of *polo* (CCGTACAACATGTGCCGTAG). Reactions were performed in a Biometra 48-well Personal Thermocycler with the following program: 5 minutes at 95ºC, 35 cycles of 30 seconds at 95ºC, 90 seconds at 52°C, 1 minute at 72ºC and a final elongation step for 10 minutes at 72ºC. The PCR products were then separated on a 0.8-2% agarose gel, the bands cut and kept at -80ºC overnight. The bands were then incubated at 42ºC for 4 minutes and the products were purified using sephadex columns.

3.4. Northern blotting

The DNA probes were generated from w^{1118} or w^{1118};*gfp-polo*Δ*USE*;*polo*^{9}*/TM6B* larvae brain cDNA, producing two fragments with 543 and 176 bp for *polo* and another one with 550 bp for *gfp* (the oligonucleotides used are depicted below). The 176 bp and 550 probes were labelled with α-^{32}P dCTP (10 mCi/mL, PELSBLU013H250UC, PerkinElmer) in a RT-PCR performed using GoTaq DNA Polymerase according to the manufacturer's instructions in a Biometra 48-well Personal Thermocycler. The RT-PCR program for the 176 bp probe was: 5 minutes at 95ºC, 35 cycles of 30 seconds at 95ºC, 30 seconds at 58°C, 30 seconds at 72ºC and a final elongation step at 72ºC for 10 minutes. The RT-PCR program for the 550 bp probe was: 5 minutes at 95ºC, 35 cycles of 30 seconds at 95ºC, 30 seconds at 60°C, 45 seconds at 72ºC and a final elongation step at 72ºC for 7 minutes. The probe corresponding to the 543 fragment was double digested with *Sal*I and *Bgl*II to create 5' overhangs at both ends, which were then separated on a 0.8-2% agarose gel, the bands cut and kept at -80ºC overnight. The bands were then incubated at 42ºC for 4 minutes and the products were purified using sephadex columns. 3'end labelling with α-^{32}P dCTP (10 mCi/mL) of these fragments was achieved using the Klenow fragment (Thermo Fisher) according to the manufacturer's instructions. 20 µg of total RNA from w^{1118}, *heph*^{2}*/TM6B*, *gfp-polo*Δ*USE*;*polo*^{9}*/TM6B* and *gfp-polo*Δ*pA1*;*polo*^{9}*/TM6B* larvae brains were loaded onto a 1% agarose gel with 3-(N-morpholino)propanesulfonic acid

(MOPS) and separated by electrophoresis for 6-7h at 75 V. The gel was washed for 10 minutes in 10x standard saline citrate (SSC) and the mRNA was transferred to an Hybond-N^+ nylon membrane (Amersham Biosciences) overnight, which was then cross-linked in a Hoefer UVC 500 ultra-violet crosslinker (700 J/cm^2), washed and pre-hybridized for 2h. After denaturation, each labelled probe was then added and incubated overnight to allow hybridization (the hybridization temperature for the *polo* probe was 46.6ºC and 51.5ºC for the *gfp* probe). The membrane was washed with 2x SSC and 0.1% SDS at room temperature, twice with 1x SSC and 0.1% SDS at the respective hybridization temperatures and several times for 15 minutes with 0.5x SSC and 0.1% SDS at 65ºC. The northern blot was developed using a Fuji Medical X-Ray film Super RX-N (Fujifilm) after exposure to the membrane for 4h at -80ºC.

3.5. Real Time-quantitative Polymerase Chain Reaction

For Real Time-quantitative Polymerase Chain Reaction (RT-qPCR) experiments using cDNA, 1 µL of cDNA was used in a final volume of 10 µL per well, as well as 5 µL SYBR Select Master Mix (Applied Biosystems) and variable concentrations of each forward and reverse primer pair according to their respective efficiencies (see section 3.7). For experiments using sheared chromatin from ChIP, 4 µL of sample were used per well.

Each experiment condition was made at least in duplicate to ensure reproducibility. The annealing temperature corresponded to the one chosen during each primer pair optimization and efficiency testing (section 3.7).

Reactions were run using the StepOne Real-time PCR System or the 7500 Fast Real-Time PCR System (Applied Biosystems) and collected data were analysed using the StepOne or 7500 v2.0.6 software (Applied Biosystems). C_t values between replicates were considered if variation was less than 0.5 and the threshold was manually adjusted when necessary.

The program used for the experiments in the StepOne Real-time PCR System was the following: 50°C for 2 minutes, 95°C for 2 minutes, 40 cycles of 95°C for 15 seconds, optimal annealing temperature for 15 seconds and 72°C for 1 minute, 95°C for 15 seconds and 55.5°C for 10 seconds. The program used in the 7500 Fast Real-Time PCR System was the following: 50°C for 2 minutes, 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds, optimal annealing temperature for 1 minute and 72°C for 1 minute.

3.6. Primer optimization and efficiency

Several annealing temperatures for each primer pair were tested using a gradient thermal cycler (Veriti 96 Well Thermal Cycler, Applied Biosystems) using the program described in section 3.4. The annealing temperature that was used for the PCR product showing the most intense band after 1.5% agarose gel electrophoresis separation was chosen as the optimal annealing temperature under these conditions.

After this optimization, primer efficiency for RT-qPCR was performed using serial dilutions (1, 10, 100, 1000 and 10000) of the cDNA or chromatin samples. These dilutions were plotted against the obtained C_t values to create a linear trend line. If the slope of this trend line was approximately -3.33, the primer pair was considered efficient. Some exceptions for a few primer pairs previously used in the laboratory were, however, used even if they showed lower efficiencies. The RT-qPCR program used to test primer efficiency was the same as depicted in section 3.5. All primers used in this work showed no significant dimerization issues.

For ChIP analysis, the optimization trend lines for each sample and for each primer pair were used to determine an approximate chromatin occupancy percentage for each IP using linear regression, with the input chromatin occupancy set to 100% in each condition.

For the remaining experiments using cDNA, a relative gene expression data analysis was performed by applying the $\Delta\Delta C_t$ method [543, 544]. *rp49* or *7SL* were used as housekeeping genes.

3.7. Oligonucleotide sequences

3.7.1. Chromatin Immunoprecipitation

TABLE 1 | Primer pair sequences used in the ChIP experiments. Primer pairs 1 through 5 were designed to anneal along each locus as depicted by **FIGURE 7**.

Primer name	Primer sequence
polo 1F	GCTTTGTGCTTGGTTTTCGT
polo 1R	TTTACTACGGACTGCCCCTTT
polo 2F	GTTCTTGCCCAGCTCTTGTC
polo 2R	AGATTGGCCTTGAGGAAGGT
polo 3F	CCGTACAACATGTGCCGTAG
polo 3R	CCAGATGTACATGATGCCGA

polo 4F	AAGGCCGAATGTTAGTTTAACG
polo 4R	TTTCGATATGAAGGGGAAGG
polo 5F	ACGTGTTTCGAAATGCCTAT
polo 5R	ACACTTAAACACTTTGCAGCAG

3.7.2. mRNA expression levels

TABLE 2 | Primer pair sequences used for the mRNA level measurements. Total primer pairs were designed to anneal to and amplify the coding sequence of the respective gene, therefore both isoforms when applicable, and the pA2 primer pairs were designed to anneal at the 3'UTR of the longest isoform of the respective genes.

Primer name	Primer sequence
polo total F	CCGTACAACATGTGCCGTAG
polo total R	CTTTAGACACGCCGTTCTCC
polo pA2 F	TACTGCTGCAAAGTGTTTAAGTG
polo pA2 R	CGCTTTTAGTCAAAAGCATTTAC
lace total F	GCTGGTGTACCTGTTTTCCAA
lace total R	TCCTTCCATGATGGGTGTG
lace pA2 F	GCGGGCAGATGAATCTTAAA
lace pA2 R	GCGTCGAGTATCCAACTTCTG
abd-b total F	GCTAGTCCAGCGATTGGAAG
abd-b total R	GTCGGTTGGTCACACATCAG
abd-b pA2 F	TCCGTACAACACCATTTTCG
abd-b pA2 R	AGTGGCGATTACGAGCTGAT
CG6034 total F	CACCGCACTCCACACAACTA
CG6034 total R	ATTGGGATGTCCGGTTCC
CG6034 pA2 F	CAGTAACGGAAGACCCGAAA
CG6034 pA2 R	GGTCCAAAGGAGGGTGAAAT
rp49 F	ATCGGTTACGGATCGAACAA
rp49 R	GACAATCTCCTTGCGCTTCT
18s F	TGGTCTTGTACCGACGACAG
18s R	GCTGCCTTCCTTAGATGTGG
U6 F	TTGGAACGATACAGAGAAGATTAGC
U6 R	TCACGATTTTGCGTGTCATC
Ssu72 F	GGCACCAAATACGAGGACA
Ssu72 R	CAGGCCGTTCTGTGTGTAGA

elav F	GGCTTTGTTGGTCTTGAAGC
elav R	AGGATCCCACAACGAATCAG
heph F	ATCACACGTATCGGCTTTCC
heph R	CACAGCCATGTCTCACTT
Pcf11 F	TGCCATGGACACACTAATCAA
Pcf11 R	CGTCATCGTCGTCTTCAAAA
fne F	CGTGACCATGACCAACTACG
fne R	GCCCAGGGTATAACCATTCA
Nelf-E F	CATTTTCCCAACATGGTTTACA
Nelf-E R	GCTTGCAGTGCCTTTTTCTT
Rrp6 F	AAGAGGAATCGGCCCAAG
Rrp6 R	ATTGCATTCTTGAACCCTTTG
Spt6 F	GGCCGTCTCCGATAGTAGC
Spt6 R	TCGATCAGATCTTTGAGCTCTTC
7SL F	TTGGCTAAGGAGGGATGAAC
7SL R	CTACTGCCTACCACGGGAAC

3.7.3. Northern blot DNA probes

TABLE 3 | Primer pair sequences used for the generation of the northern probes.

Primer name	Primer sequence
polo northern probe small F	CCGTACAACATGTGCCGTAG
polo northern probe small R	CTTTAGACACGCCGTTCTCC
polo northern probe large F	CGGGTTTGCAAAATGTTACG
polo northern probe large R	AGATTGGCCTTGAGGAAGGT
gfp northern probe F	GGAGAGGGTGAAGGTGATGC
gfp northern probe F	TCGAAAGGGCAGATTGTGTG

4. *In silico* Studies

4.1. Assessment of the *RpII215*^{C4} R741H mutation

After downloading the *Saccharomyces cerevisiae* RNAPII structure from the Protein Data Bank (PDB ID 4A3F [87]), Arg726 (which corresponds to Arg741 in *Drosophila melanogaster*) was manually mutated to His726 (R726H) using the Visual Molecular Dynamics (http://www.ks.uiuc.edu/Research/vmd/) software [545], thus mimicking the

RNAPII point mutation of the C4 fly mutants [112, 113, 398]. Distance between the mutated residue and the nearest nucleotide was also measured using VMD.

4.2. 3' Region Extraction And Deep Sequencing and polyadenylation site identification

Heads and bodies were dissected from w^{1118} and $RpII215^{C4}$ 1-5 days old adult male flies using standard procedures. Total RNA was extracted and purified as described in section 3.1. The 3' region extraction and deep sequencing (3'READS) and 3'READS+ methods were previously described in [546] and in [547], respectively. Data processing was based on the methods previously described [547]. Briefly, reads from 3'READS were mapped to the fly genome (BDGP5, Ensembl v70) using Bowtie2 [548]. Uniquely mapped reads (with MAPQ score > 10) that had at least two non-genomic Ts at the 5' end were considered as pA site containing reads. pA sites located within 24 nt from each other were clustered together as previously described. pA sites mapped to the genome were further assigned to genes using gene models defined by the Ensembl database. The 3′ ends of the gene models were extended by 2 kilobases (kb) to include downstream pA sites, but the extension did not go beyond the transcription start site of the downstream gene. To eliminate spurious pA sites, we further required that the number of pA sites reads for a pA site was ≥ 5% of all pA site reads for the gene in at least two samples.

4.3. Alternative polyadenylation analysis

APA analysis was carried out following largely the methods previously described [371, 549, 550]. Briefly, relative expression (RE) of two pA site isoforms in the 3'-most exon, e.g., proximal pA site (pPAS) and distal pA site (dPAS), was calculated by log2(RPM) of dPAS vs. pPAS, where RPM was reads per million pA site-containing reads. Relative expression difference (RED) of two isoforms in two comparing samples was based on difference in RE of the two isoforms between two samples. DEXSeq was used to derive statistically significant APA events (false discovery rate (FDR) < 0.05) [551] .

4.4. Compilation of RNA binding proteins, cleavage/polyadenylation, elongation and termination factors

For all known cleavage/polyadenylation, elongation and termination factors across several species [180, 181, 188, 220, 234, 285, 375, 552, 553], *Drosophila melanogaster* orthologs were found using FlyBase QuickSearch tool [554] when possible. The expression of the genes that encoded these proteins was then assessed based on the 3'READS data in wild type heads and bodies [280].

4.5. USE conservation analysis

The *polo* 3'UTR sequences of 18 *Drosophila* species were obtained from FlyBase [554] and aligned using the SnapGene software (from GSL Biotech, Chicago, Illinois, USA; available at snapgene.com). The coloured sequences above the alignments represent the consensus sequence.

4.6. Sequence motif search and Gene Ontology enrichment of Upstream Sequence Element-containing genes

NCBI RefSeq transcripts of *Drosophila melanogaster* (assembly dm6) and *Homo sapiens* (assembly hg38) were retrieved from UCSC Table Browser as *.bed* files. Using a R-based script [555], we queried the 3'UTR of the *Drosophila melanogaster* and human genomes for the presence of the conserved TTGTTTTT sequence. The search was first restricted to 450 nt upstream of three different pA signals: AATAAA, ATTAAA or AATATA (Supplementary Tables 5 and 6, [278]). The query was also performed with a single nt substitution (G>N), TTNTTTTT (USE-N) for *Drosophila melanogaster*, restricted to 90 nt upstream of the three same pA signals. Gene ontology enrichment analysis of the USE-containing genes was performed using the PANTHER platform [541, 542] (Version 13.0, released on November 12th 2017).

4.7. Frequency, distance to polyadenylation signals and expression analyses of Upstream Sequence Element-containing genes

The 3'UTRs of USE-containing genes were evenly divided in 90 nt slots up to 450 nt and the genes were sorted according to distance between the USE and each pA signal. The percentage of the USE and USE-N-containing genes was calculated using the frequency of each pA (AATAAA, ATTAAA or AATATA) in the *Drosophila melanogaster* [214] and human [215] genomes. We used the *Drosophila* Gene Expression Tool (DGET [556, 557], https://www.flyrnai.org/tools/dget/web/) to determine the expression levels of USE-N-containing genes in different tissues and developmental stages. As control, we used genes with AATAAA, ATTAAA or AATATA that did not have USE-N in their 3'UTR. In the control subset of genes with ATTAAA, genes with the canonical AATAAA signal were excluded. In the control subset of genes with AATATA, genes with a higher pA efficiency (AATAAA, ATTAAA, AGTAAA, CATAAA, TATAAA, GATAAA, ACTAAA and AATACA [213]) were also excluded.

5. Statistical Analysis

For each assay, at least three independent experiments were performed. Results were expressed as mean values ± standard deviation error (SDE) determined by GraphPad Prism 7.0c (La Jolla California USA, www.graphpad.com). Differences in *p* values below 0.05 calculated via a two-tailed paired Student's *t*-test were considered statistically significant. Statistical significance in the immunofluorescence assays was assessed by the Mann-Whitney test in GraphPad Prism 7.0c.

RESULTS

1. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*

Some of the results presented in this chapter were published in [280].

RNAPII elongation rate is known to have an impact on 3'UTR-APA and pA signal selection in six genes in *Drosophila melanogaster* [115]. To investigate this effect on a genome-wide scale, we mapped all mRNA 3'ends using the 3' Region Extraction and Deep Sequencing (3'READS) methodology [546, 547] in the head and body of *Drosophila melanogaster* wild type and the mutant *RpII215*^{C4} strain, which possesses a slower RNAPII elongation rate [113].

1.1. 3'UTR-Alternative polyadenylation pattern varies between the head and body in *Drosophila melanogaster*

It has been reported in several organisms that terminally differentiated cells tend to produce mRNA isoforms with longer 3'UTRs than proliferative cells in several organisms [369, 372, 379, 380, 382], including *Drosophila melanogaster* [279]. Using 3'READS and wild type flies, we found approximately 17000 novel pA sites in the downstream region of annotated 3'ends [280], which suggests poor annotation of mRNA 3'ends using less adequate sequencing methods [279]. We therefore re-annotated these pA sites and were able to extend the 3'ends of 10112 genes using our data [280]. We have also found that there is a significant abundance of distal pA sites (dPASs) that are selected in wild type fly heads (**FIGURE 11**, red dots) whereas proximal pA sites (pPASs) are preferentially chosen in fly bodies (**FIGURE 11**, blue dots).

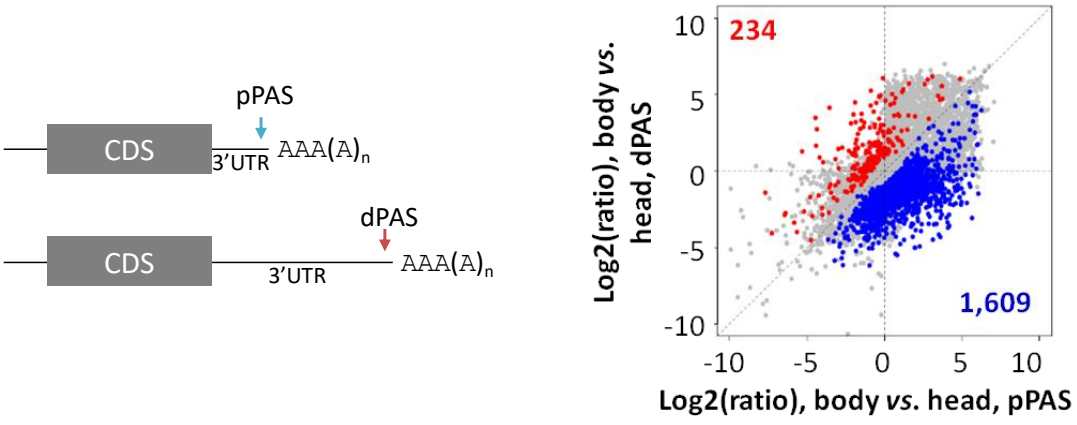


FIGURE 11 | 3'UTR-APA differences between *Drosophila melanogaster* body and head in the 3'UTR. On the left, schematic representation of two mRNA isoforms

produced by selecting a pPAS or a dPAS: pPAS choice leads to a short 3'UTR and dPAS choice leads to a long 3'UTR. On the right, scatter plot showing pPAS and dPAS isoform abundance differences between fly head and body (two biological replicates used). Genes with significantly (FDR < 0.05, DEXseq analysis) higher abundance of pPAS isoforms in the body *vs*. head are shown in blue (1609 genes), and those with higher abundance of dPAS isoforms are in red (234 genes).

It remains unclear why body tissues tend to select pPASs and head tissues predominantly choose dPASs. One of the potential factors we investigated was the distance between the pPAS and dPAS and how it relates to 3'UTR shortening *versus* lengthening in each compartment.

First, we sorted expressed genes that undergo APA into five bins according to the distance between their pPAS and dPAS (**FIGURE 12**, table). Then, we plotted the five different categories against the relative expression difference (RED) found by 3'READS between the head and the body, i.e., the difference in pPAS *vs*. dPAS usage between both compartments. As it can be observed in the graph of **FIGURE 12**, there is a clear correlation between the pPAS-dPAS distance and the RED values between the head and body: the longer the distance between pPAS and dPAS, the higher the difference between pPAS *vs*. dPAS usage in each compartment. As APA occurs co-transcriptionally, the distance between pPAS and dPAS is a direct reflection of the time period between the selection and processing of the pPAS or the dPAS. Thus, these results suggest that differences in the kinetic competition between pPAS *vs*. dPAS selection may ultimately lead to APA event changes observed between the fly head and body.

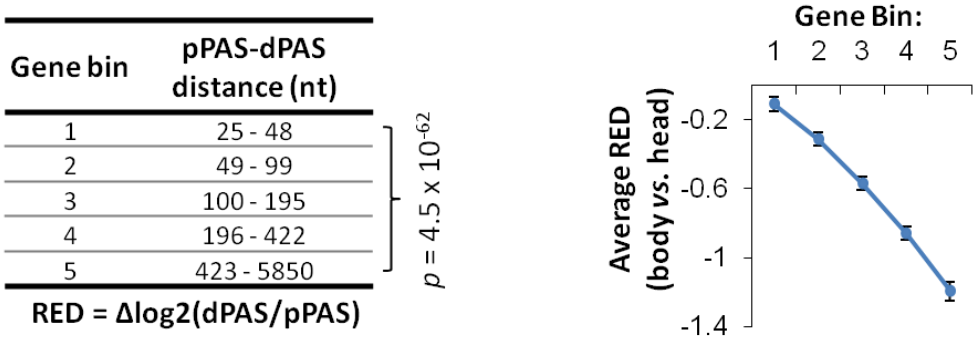


FIGURE 12 | Correlation between pPAS *vs*. dPAS usage in body and head and distance between pA sites. Expressed genes that undergo APA were evenly divided into five bins according to their distance between pPAS and dPAS, resulting in ~1300 genes in each bin. pPAS-dPAS distance is displayed in the table. pPAS *vs*. dPAS usage between body and head is represented by relative expression difference (RED) values, which corresponds to the difference in log2(ratio) of dPAS isoform abundance to pPAS isoform

abundance between body and head. Error bars are standard error of mean. Values for genes in bin #1 were compared with those in bin #5 by the Wilcoxon rank sum test and the *p*-value is shown. Only pA sites with ≥ 5 reads were used for analysis.

1.2. RNA processing genes are upregulated in *Drosophila melanogaster* heads

To understand the mechanistic differences involved in the APA changes found between wild type fly head and body, we compiled a list of all known mRNA 3'end processing proteins, elongation and termination factors [180, 181, 188, 220, 234, 285, 375, 552, 553] and asked how they were expressed using the 3'READS data obtained from wild type heads and bodies.

Most of these genes are intriguingly upregulated in wild type head in comparison to the body as shown in **TABLE 4**, namely genes encoding for proteins that are directly involved in transcription and mRNA 3'end processing as well as elongation and termination (highlighted in bold). As with the previously reported prevalence of longer mRNA isoforms in the brain [379, 380, 382], these results may imply that a distinct level of regulation of these factors is necessary in such a highly differentiated tissue that may ultimately result in the production of mRNA isoforms with long 3'UTRs.

TABLE 4 | Differential expression of genes coding for cleavage and pA, elongation and termination factors between *w*^{1118} head and body by 3'READS analysis. Gene expression in the head was used as reference, meaning that genes upregulated in the head are termed 'Upregulated' and are downregulated in the body and genes downregulated in the head are termed 'Downregulated' and are upregulated in the body.

mRNA 3'end processing proteins	*hfp* *nonA* *PNUTS* *mle* *CG10077* *eIF4A* *CG3689* ***Pcf11*** *CG11454* *snRNP-U1-70K* *Pp1-87B* *ku80* ***CstF64*** *sf3a1* *fip1* *mtr4* *eIF3a* *hel25e* *eIF3b* *eIF4G1* ***elav***	**UPREGULATED**
	gem3 *bel* *pabp2* *rhau* *rin* *rbfox1*	**DOWNREGULATED**

Elongation factors	***NELF-B*** *SC35* ***NELF-A*** *fs(1)h* ***Cdk9*** *TH1* *ctr9* *hay* ***Ssu72*** *Chd1* *CycK* ***MED26*** *snf* *hpr1* *eaf* *TfIIFalpha* ***EloB*** ***EloC*** *SF1* *kto*	**UPREGULATED**
	LARP7 ***EloA*** *Spn42Dd* *HEXIM* *rtf1*	**DOWNREGULATED**
Termination factors	*pasha* *Trf4-1* *csul* ***RpII140*** ***MED16*** *CG34183*	**UPREGULATED**
	MED20	**DOWNREGULATED**

1.3. *In silico* analysis of the *RpII215^{C4}* R741H mutation suggests stereochemical hindrance between RNA Polymerase II and the DNA

Within the scope of further understanding the impact of RNAPII elongation rate in APA and pA site usage [115], we were interested in studying why the RNAPII of the *RpII215^{C4}* fly mutant is 50% slower than their wild type counterpart [113, 398]. Early studies of the original *RpII215^{C4}* RNAPII suggested an alteration in structure due to the arginine-to-histidine (R741H) point mutation [112], but the cause for the slower transcription elongation rate remained elusive. We first sought to answer this question with an *in silico* study using the crystallized structure of RNAPII from *Saccharomyces cerevisiae* [558] (the structure for the *Drosophila melanogaster* RNAPII is not yet available).

Using the VMD software [545], we mimicked the R741H point mutation found in the Rpb1 subunit of RNAPII of the C4 flies (see section 4.1 of Materials and Methods), which

corresponds to R726H in the *Saccharomyces cerevisiae* RNAPII, and then measured the distance between both amino acid residues and the DNA strand (**FIGURE 13**).

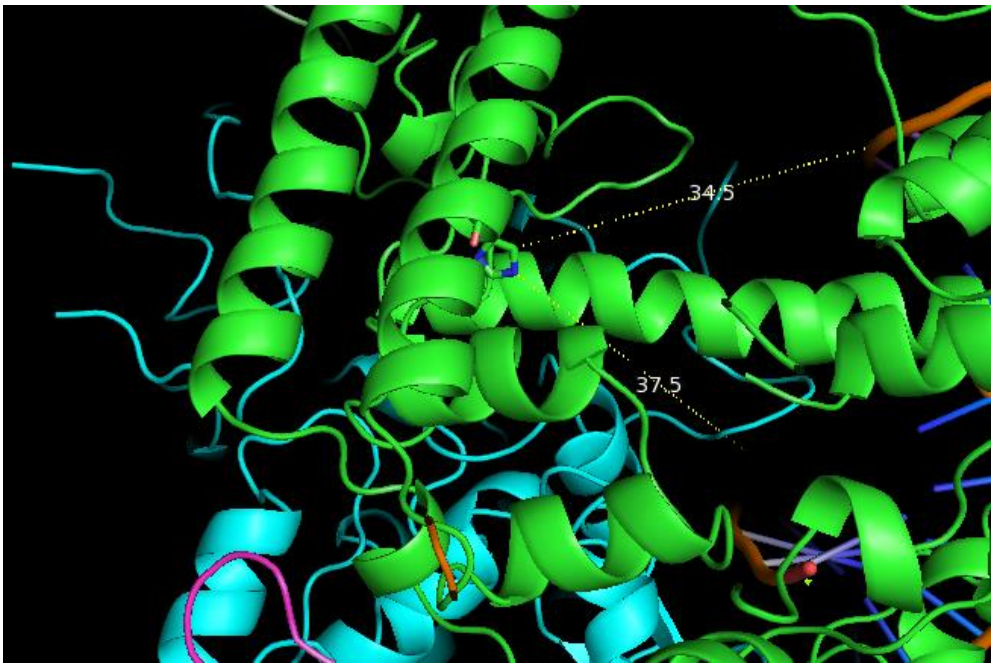


FIGURE 13 | Visual representation of the wild type RNAPII and a variant with a R726H point mutation from *Saccharomyces cerevisiae* [87]. This image was made with the VMD software [545] support (http://www.ks.uiuc.edu/Research/vmd/). VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

In the presence of a histidine instead of the arginine, the distance between the catalytic centre of the protein complex and the DNA strand to be transcribed increased 2-3 Å (from 34.5 Å to 37.5 Å, see **FIGURE 13**).

We thus hypothesize that the positively charged arginine (centre, side-chain coloured in green and orange, **FIGURE 13**) is more prone to attract the negatively charged DNA and RNA strands and allow them an easier and faster passage through the RNAPII catalytic centre when compared to the neutral histidine (centre, side-chain coloured in green and blue, **FIGURE 13**) due to possible stereochemical hindrance. As the efficiency of the enzymatic reaction is dependent on the optimal distance between the catalytic site and the substrate, the increase in the distance introduced by the histidine may cause the slower transcription elongation rate described for this mutant strain [113].

Productively elongating RNAPIIs must be phosphorylated at the CTD Ser-2 residue [89, 98]. Although the general structure of the transcriptionally active RNAPII CTD is probably

unaffected by the Arg-to-His replacement, CTD Ser-2 sites may not be available for phosphorylation and the RNAPII activity may be compromised in the *RpII215*^{C4} strain.

1.4. A slower transcription elongation rate affects alternative polyadenylation in a tissue-specific manner

The *Drosophila melanogaster* mutant strain *RpII215*^{C4} had shown a preferential usage of proximal pA signals in six genes [115]. We now examined genome-wide pA site usage and APA profiles by 3'READS in *RpII215*^{C4} flies and compared them to our observations for the wild type strain.

The slower *RpII215*^{C4} mutant strain shows a mild preference for expressing mRNAs with shorter 3'UTRs in the adult male body (360 genes with upregulated pPASs against 301 genes with upregulated dPASs, as seen in **FIGURE 14**), but not in the head (**FIGURE 14**, 567 genes with upregulated pPASs against 541 genes with upregulated dPASs). This indicates that the effect on APA caused by a slower transcription elongation rate is possibly context- or tissue-specific.

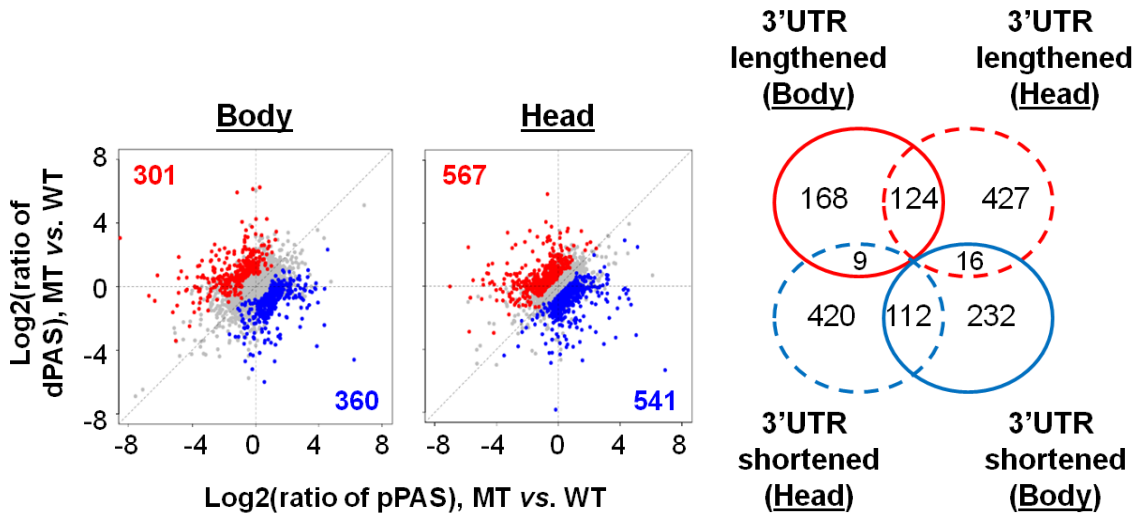


FIGURE 14 | APA regulation in the 3'UTR of the slower *RpII215*^{C4} fly mutant. (A) Scatter plots comparing APA isoform abundance between wild type *w*^{1118} (WT) and the slower *RpII215*^{C4} mutant (MT) in the body (left) *versus* the head (right). **(B)** Venn diagram comparing APA event changes between WT and MT in the body and head compartments.

The fact that we do not observe differences in APA events in the head in the presence of a slower transcription elongation rate in comparison to wild type heads also suggests that APA regulation in the more differentiated fly head tissues, namely the preferential dPAS selection, is robust and more impervious than tissues in the fly bodies when dealing with a slower transcription elongation rate.

1.5. Gene expression of RNA Binding Proteins, elongation and termination factors is altered by a slower transcription elongation rate

One potential explanation for the context-specific effect of a slower transcriptional elongation rate on APA events may be the differential expression of genes coding for proteins with functions in pre-mRNA cleavage and pA, such as the core *trans*-acting factors, elongation and termination factors and/or RBPs between the wild type and the slower fly strains.

Using the list of genes encoding for known mRNA 3'end processing proteins, elongation and termination factors [180, 181, 188, 220, 234, 285, 375, 552, 553] (see section 1.2) and our 3'READS data, we asked how these genes were expressed in wild type heads and bodies in comparison the slower *RpII215*^{C4} mutant heads and bodies.

As seen in **TABLE 5**, we did not observe a clear difference in the expression of these genes between wild type and mutant bodies (left-side table) or wild type and mutant heads (right-side table): there are approximately as many genes up- or downregulated in the mutant body/head in comparison to the wild type counterpart in contrast to the general upregulation of these proteins seen in wild type heads *versus* bodies (**TABLE 4**). There is a higher number of genes encoding for mRNA 3'end processing proteins (approximately 77%) upregulated in the slower mutant heads when compared to wild type heads, but we did not observe differences in APA events between wild type and mutant heads by 3'READS (**FIGURE 14**). This suggests that the upregulation of these genes does not have an effect on APA in head tissues with a slower transcription elongation rate.

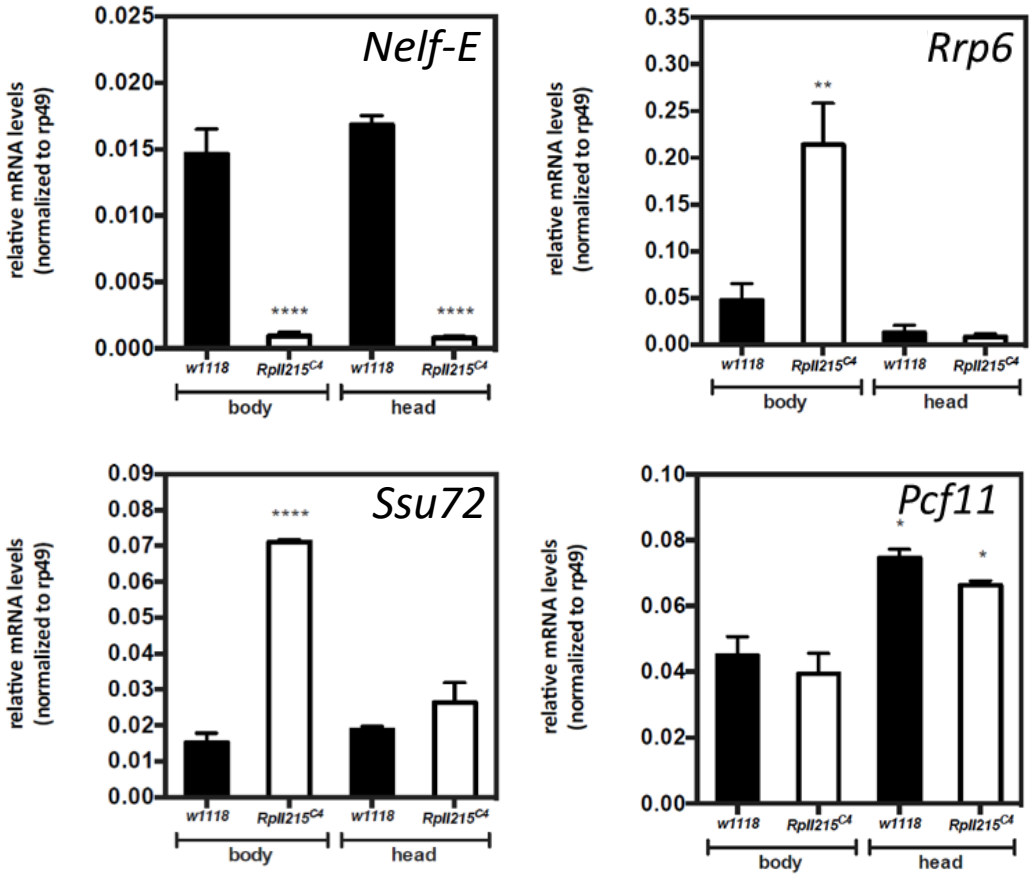
TABLE 5 | Differential expression of genes coding for cleavage and pA, elongation and termination factors between wild type (WT) head and body *versus* *RpII215*^{C4} (MUT) head and body by 3'READS analysis. Gene expression in the mutant was used as reference, meaning that in both the body (left-side table) and head (right-side table) compartments, genes upregulated in the mutant strain are termed 'Upregulated' and are downregulated in wild type and genes downregulated in the mutant are termed 'Downregulated' and are upregulated in wild type.

	MUT body *versus* WT body		MUT head *versus* WT head	
3'end processing proteins	bel CG11454 CG3689 Cstf64 Fip1 Gem3	**UPREGULATED**	bel CG10077 CG11454 CG3689 Cstf64 eIF4A	**UPREGULATED**

	hfp nonA Pabp2 Pcf11 PNUTS Pp1-87B rin Sf3a1 snRNP-U1-70K		eIF4G1 Fip1 Gem3 Hel25E hfp Ku80 mle nonA Pabp2 Pp1-87B Rbfox1 rin Sf3a1 snRNP-U1-70K	
	CG10077 eIF3a eIF3b eIF4A eIF4G1 Hel25E Ku80 mle Mtr4 Rbfox1 Rhau	**DOWNREGULATED**	eIF3a eIF3b Mtr4 Pcf11 PNUTS Rhau	**DOWNREGULATED**
Elongation factors	Cdk9 Chd1 EloC hay HEXIM Hpr1 Kto LARP7 Ssu72 TfIIFalpha TH1	**UPREGULATED**	Chd1 eaf EloA EloC hay HEXIM Hpr1 LARP7 MED26 NELF-A Rtf1 SC35 snf Spn42Dd Ssu72	**UPREGULATED**
	Ctr9 CycK eaf EloA EloB fs(1)h MED26 NELF-A NELF-B Rtf1 SC35 SF1 snf Spn42Dd	**DOWNREGULATED**	Cdk9 Ctr9 CycK EloB fs(1)h kto NELF-B SF1 TfIIFalpha TH1	**DOWNREGULATED**

Termination factors	CG34183 csul pasha RpII140 snRNP-U1-70K	**UPREGULATED**	csul pasha RpII140 snRNP-U1-70K	**UPREGULATED**
	Med16 Med20 Trf4-1	**DOWNREGULATED**	CG34183 Med16 Med20 Trf4-1	**DOWNREGULATED**

To complement this analysis, we quantified by RT-qPCR the expression levels of some genes encoding proteins with a well-established role in mRNA processing in the heads and bodies of adult *RpII215*^{C4} male flies in comparison to wild type *w*^{1118}. We focused on the 3'-5' exonuclease *Rrp6* [559], the RNAPII CTD Ser-5-P and Ser-7-P phosphatase *Ssu72* [560], the termination factor *Pcf11* [143], the RBPs *fne* and *elav* [305, 307, 561, 562], the elongation factor *Spt6* [72, 563] and the negative elongation factor *Nelf-E* [72-75].



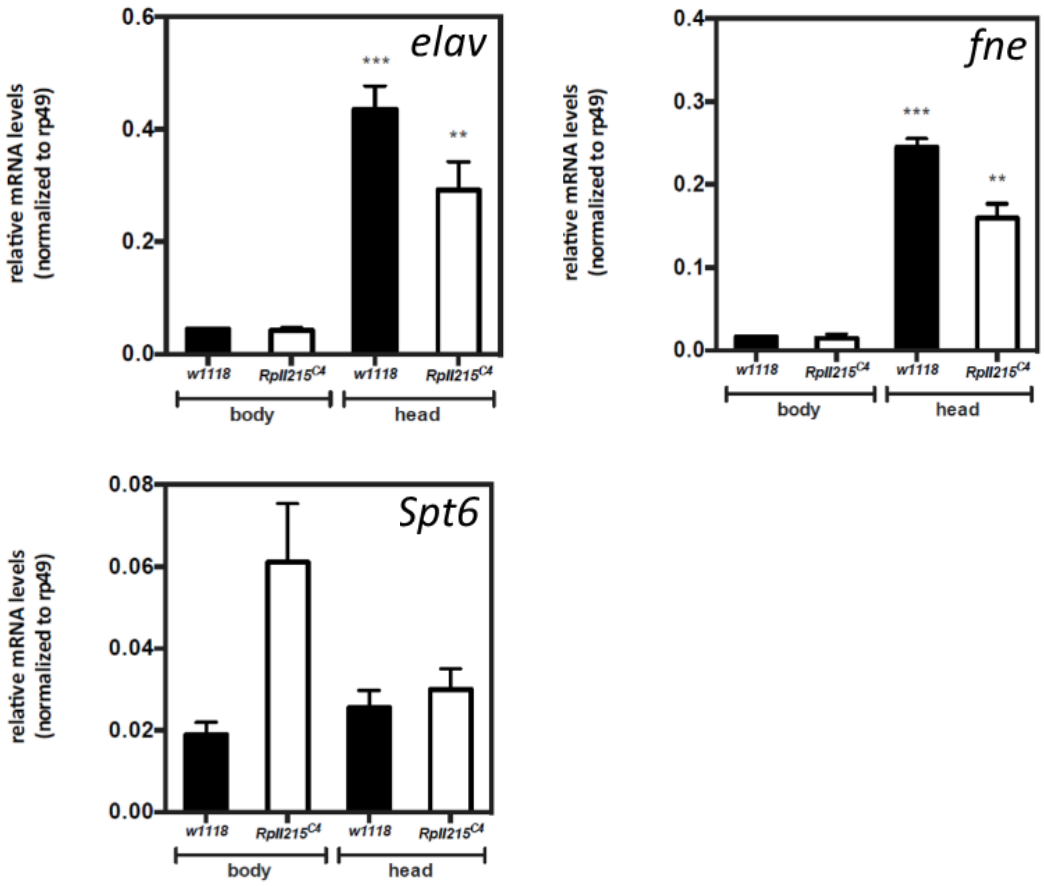


FIGURE 15 | Relative mRNA expression levels of mRNA 3'end processing factors in wild type (*w*^{1118}) and *RpII215*^{C4} male bodies and heads. Expression was quantified by RT-qPCR. *rp49* was used as a housekeeping gene. Error bars show SEM from three independent experiments.

Interestingly, *Nelf-E* expression is drastically reduced in *RpII215*^{C4} (**FIGURE 15**). Mutant flies also show a marked increase in the expression of *Rrp6* and *Ssu72* in the body, but not in the head, in comparison to wild type, while *Pcf11* shows no statistical difference between both strains. *elav* and *fne*, which have been shown to be highly expressed in the head [305, 307, 561, 562], present a small decrease in their expression in the head of mutant flies. The expression levels of *Spt6* are not statistically different between both strains, but nonetheless, its expression tends to be higher in the mutant bodies.

The differential expression of *Nelf-E*, *Rrp6* and *Ssu72* in *RpII215*^{C4} flies may explain the changes in APA observed between wild type and mutant bodies (**FIGURE 14**). Together with the data showing that there is a differential regulation of mRNA 3'end processing factors in the wild type *Drosophila melanogaster* head *versus* body (**TABLE 4**), these results further suggest that these regulatory genes are either upregulated or downregulated according to cellular context (head or body compartments) and/or an altered transcriptional elongation rate (**FIGURE 15**).

2. The cell cycle kinase Polo is controlled by a conserved 3' untranslated region regulatory sequence in *Drosophila melanogaster*

Most of the results presented in this chapter were published in [278].

The *polo* gene produces two mRNAs depending on the selection of two pA signals in *polo* 3'UTR ([397] and **FIGURE 16**). The distal pA signal produces the longest *polo* isoform, which is both necessary and sufficient to ensure correct development and viability in flies as it is the main responsible for Polo protein production. The proximal *polo* pA signal generates the shortest *polo* mRNA and modulates Polo protein levels by an autoregulatory feedback mechanism [115].

2.1. *Drosophila melanogaster polo* 3' untranslated region has a conserved U-rich regulatory Upstream Sequence Element

Conserved non-coding regions between different species may predict putative regulatory elements [236, 564, 565] while functionally irrelevant sequences are less conserved. To identify potential *cis* regulatory elements in *polo* APA, we aligned the *polo* 3'UTR of 18 different *Drosophila* species. We found a 28-nt-long U-rich sequence (over 53% of U content) 127 nt upstream of *polo* pA1 (**FIGURE 16** and **FIGURE 17**) similar to USEs, a class of *cis* elements that regulate mRNA 3'end formation that have been previously described in human [236, 249-257, 263] and viral [237-245, 247, 248] genomes and thus, we named this sequence USE. We found that the TTGTTTTT in the USE is a consistently conserved region between *Drosophila melanogaster* and the 50 million year apart *Drosophila albomicans* (**FIGURE 17)**, which hints at a potentially regulatory role for this small sequence. Supporting this hypothesis, *Drosophila simulans* and *Drosophila sechellia* are the closest related species to *Drosophila melanogaster* and their 28-nt-long USE only differs in one single nt from the latter. This conservation is closely followed by the 11 million years-distant *Drosophila yakuba*, whose USE differs from the one in *Drosophila melanogaster* by two nt.

FIGURE 16 | *polo* 3'UTR mRNA sequence. The USE is highlighted within a black box and depicted in blue. The two mapped pA signals are highlighted by black boxes and depicted in red. The pre-mRNA cleavage sites (CS) are demarked by the grey arrows.

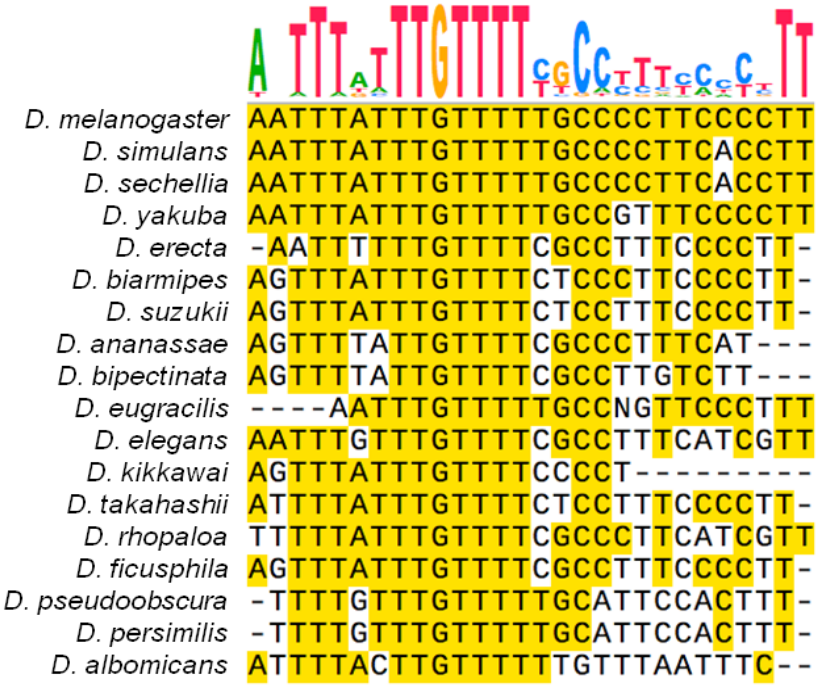


FIGURE 17 | *polo* USE conservation and alignment in 18 *Drosophila* species. Nucleotides are coloured in yellow if they match the corresponding *Drosophila melanogaster polo* USE nucleotides. The coloured sequence above the alignment represents the consensus sequence, with the conservation level defined by letter size.

Interestingly, seven *Drosophila* species also show conservation of *polo* pA1 and the region directly upstream (**FIGURE 18**). In *Drosophila pseudoobscura* and *Drosophila persimilis*, both 37 million years apart from *Drosophila melanogaster*, we also observed that the *polo* pA1 sequence corresponds to the canonical pA signal, AATAAA (**FIGURE 18**) instead of the less efficient ATTAAA variation [213].

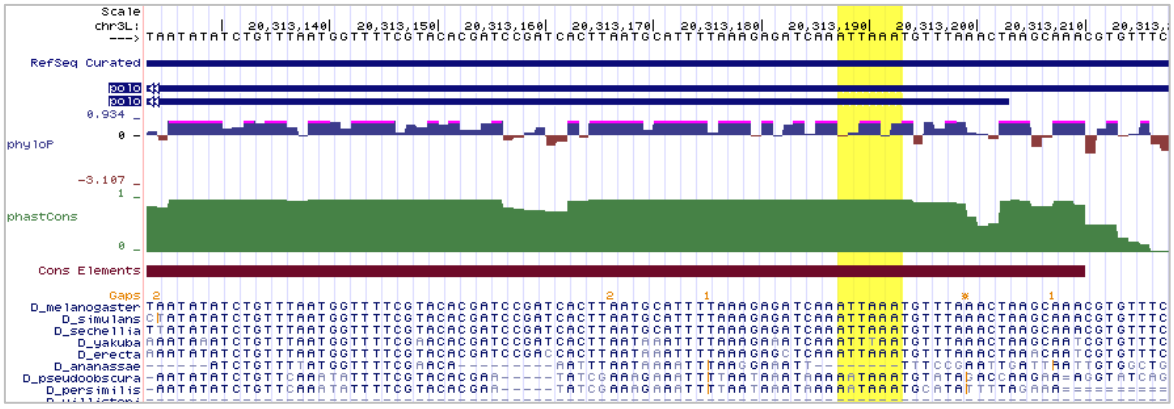


FIGURE 18 | UCSC browser snapshot of *polo* pA1 conservation and alignment in seven *Drosophila* species. *polo* USE is highlighted in yellow across the species in display. The sequence above the alignment represents the consensus *Drosophila melanogaster polo* sequence.

Apart from the USE, we were unable to find an overall conservation of the entirety of *polo* 3'UTR across these 18 fly species and curiously, no conservation of *polo* pA2 or the surrounding region despite the fact that this pA signal is the main responsible for Polo protein production and vital for *Drosophila melanogaster* development [115].

2.2. The Upstream Sequence Element is more prevalent near polyadenylation signals and upstream of non-canonical signals

To study the genome-wide prevalence of the USE across the *Drosophila melanogaster* as well as the human genomes, we developed a bioinformatic script that allowed us to question whether the eight most conserved nt of the USE, TTGTTTTT, are associated with pA signals in 3'UTRs. Considering that *polo* pA2 is found 380 nt downstream of the USE, we chose a range of 0-450 nt between the TTGTTTTT and the pA signal.

We have selected three different pA signals with different pA signal efficiencies for this study: AATAAA, ATTAAA or AATATA. AATAAA is the canonical pA signal in both humans [215] and flies [214] while the ATTAAA signal is the second most efficient signal [213] in both species and also corresponds to *polo* pA1. AATATA, *polo* pA2, shows poor pA signal efficiency in human cells [213], but is known to be physiologically relevant in insects [214,

216]. The *polo* mRNA produced by *polo* pA2 signal is also the longest isoform which is responsible for producing most of Polo protein and is vital for fly viability [115].

After excluding repeated hits corresponding to USE-containing genes with more than one pA signal (AATAAA and ATTAAA or AATAAA and AATATA or ATTAAA and AATATA), we found that the USE is present in the 3'UTRs of 889 and 1140 genes in *Drosophila melanogaster* and humans, respectively. This corresponds to 5.2% of *Drosophila melanogaster* genes and 2.7% of human genes (**TABLE 6**).

TABLE 6 | Number of *Drosophila melanogaster* and human genes containing the conserved USE upstream of AATAAA, ATTAAA or AATATA, which corresponds to approximately 5.2% and 2.7% of the fly and human genomes, respectively.

	Drosophila melanogaster	**Human**
AATAAA	625	663
ATTAAA	397	446
AATATA	396	322
Total #	889	1140
Total%	5.2%	2.7%

Gene ontology analyses show that many of these USE-containing genes in *Drosophila melanogaster* are involved in fundamental physiological processes, such as nervous system development, morphogenesis, tissue development and, interestingly, mRNA 3'UTR binding and RNA metabolic processes (**TABLE 7**). Curiously, nine out of 10 terms are related to nucleic acid metabolism in genes containing both the conserved USE and the AATATA pA signal (**TABLE 7**).

TABLE 7 | Top 10 molecular function GO terms enriched in the *Drosophila melanogaster* USE-containing genes upstream of AATAAA, ATTAAA or AATATA pA signals.

GO term	**USE-450 nt-AATAAA (fold Enrichment)**
Peripheral nervous system development (GO:0007422)	3.69
Motor neuron axon guidance (GO:0008045)	3.65
Neuronal system (R-DME-112316)	3.45
Dendrite development (GO:0016358)	3.15

Axonogenesis (GO:0007409)	3.14
Neuron recognition (GO:0008038)	3.13
Dendrite morphogenesis (GO:0048813)	3.11
Regulation of embryonic development (GO:0045995)	3.03
Regulation of protein kinase activity (GO:0045859)	3.03
Regulation of cell development (GO:0060284)	2.62

GO term	**USE-450 nt-ATTAAA (fold Enrichment)**
Regulation of axon extension involved in axon guidance (GO:0048841)	16.2
Regulation of neuron maturation (GO:0014041)	13.74
Regulation of neuron remodeling (GO:1904799)	13.74
Presynaptic active zone (GO:0048786)	8.59
Regulation of extent of cell growth (GO:0061387)	8.59
Striated muscle tissue development (GO:0014706)	8.59
Developmental growth involved in morphogenesis (GO:0060560)	6.61
Alzheimer disease-amyloid secretase pathway (P00003)	6.52
Apical cortex (GO:0045179)	6.45
Developmental cell growth (GO:0048588)	5.86

GO term	**USE-450 nt-AATATA (fold Enrichment)**
Messenger ribonucleoprotein complex (GO:1990124)	37.79
mRNA 3'UTR binding (GO:0003730)	7.56
Transcription regulatory region sequence-specific DNA binding (GO:0000976)	3.17
mRNA binding (GO:0003729)	3.03
Sequence-specific double-stranded DNA binding (GO:1990837)	2.98
Regulatory region nucleic acid binding (GO:0001067)	2.94
Double-stranded DNA binding (GO:0003690)	2.81
Positive regulation of RNA metabolic process (GO:0051254)	2.78
Positive regulation of transcription by RNAPII (GO:0045944)	2.75
Instar larval or pupal development (GO:0002165)	2.57

However, it is important to note that the genome-wide prevalence of each of the pA signals analyzed is rather different. In humans and *Drosophila melanogaster*, about half of the genes contain the AATAAA canonical signal (48% in flies and 53% in humans, see **FIGURES 19** and **20** [214, 215]). Predictably, we found a higher number of genes

containing the canonical pA signal associated with the USE across the human and fly genomes (625 genes in *Drosophila melanogaster* and 663 in *Homo sapiens*, see **TABLE 6**).

In contrast, both ATTAAA and AATATA are found in a more modest percentage of genes (10% for both ATTAAA and AATATA in flies [214, 216] *versus* 17% for ATTAAA and 2% for AATATA in humans [215], see **FIGURES 19** and **20**).

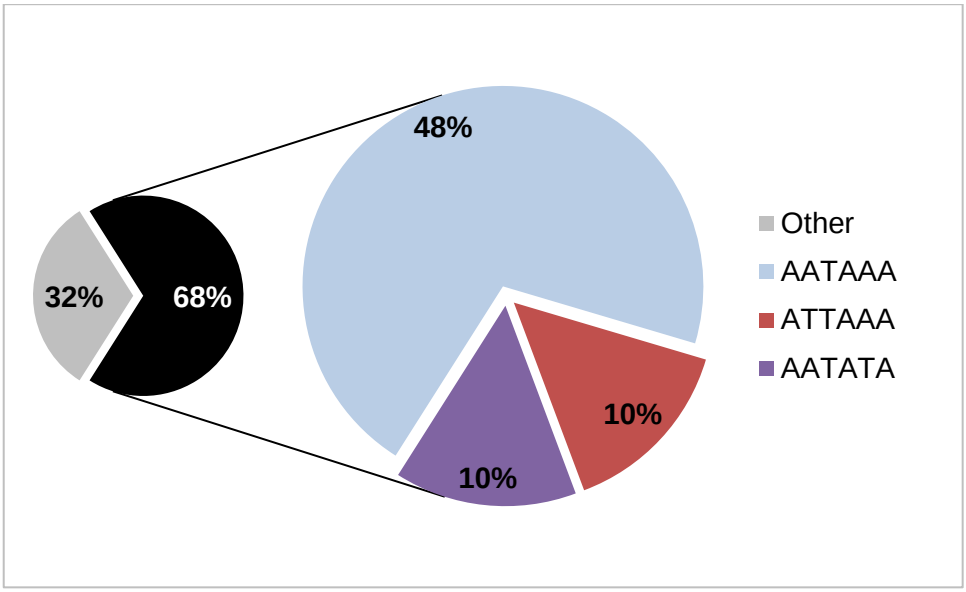


FIGURE 19 | Percentage of AATAAA, ATTAAA or AATATA-containing genes depicted for *Drosophila melanogaster* [214, 216], which approximately corresponds to 68% of the fly genome.

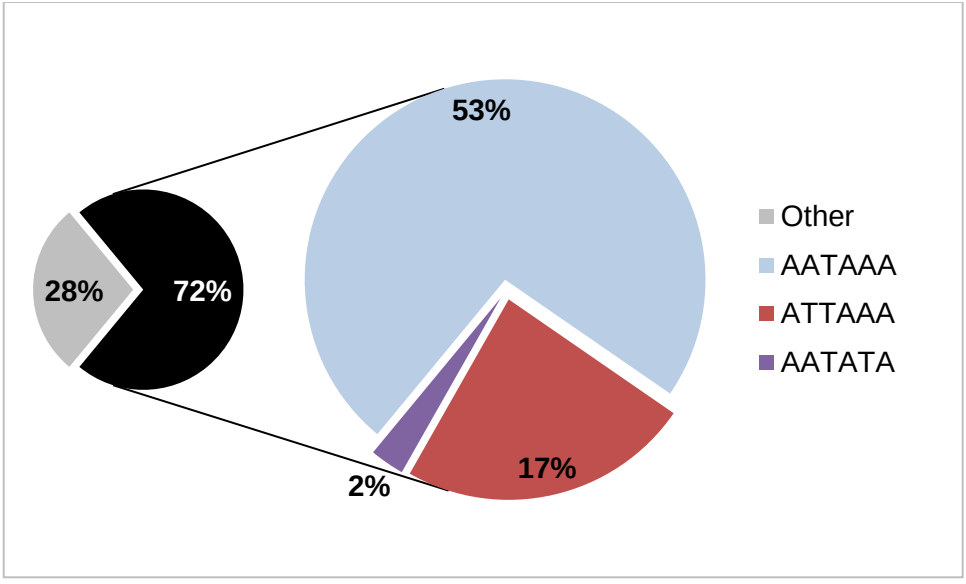


FIGURE 20 | Percentage of AATAAA, ATTAAA or AATATA-containing genes depicted for humans [215], which approximately covers 72% of our genome.

When we used the different genome-wide frequencies for each pA signal (**FIGURES 19** and **20**) and the absolute numbers of genes found with the script (**TABLE 6**) to normalize the different frequencies of USE-containing genes accordingly, we observe that it is more likely to find an USE associated with a non-canonical pA signal (ATTAAA or AATATA) than with the canonical signal in humans and flies (**FIGURE 21**). As less efficient pA signals are more prone to be regulated by *cis* regulatory elements [285], our findings that USEs are more prevalent upstream of these signals argues that they may act as regulatory *cis* elements of weak pA signals.

pA signal	% *Drosophila melanogaster* USE-containing genes	% human USE-containing genes
AATAAA	7.5	2.9
ATTAAA	22.8	6.1
AATATA	22.7	37.7

FIGURE 21 | Genome-wide percentage of *Drosophila* and human USE-containing genes normalized to each pA signal frequency. The TTGTTTTT sequence is more common upstream of weaker pA signals in the 3'UTR (ATTAAA and AATATA in comparison to AATAAA) both in *Drosophila melanogaster* and human genomes.

Auxiliary elements with a function in APA tend to be near the pA signal they regulate [227, 274, 275] and USE function upon pA signals is known to be distance-dependent [236, 255]. To investigate if the position of the TTGTTTTT sequence is dependent on the distance to the pA signal in flies and humans, we next studied the distribution of the USE in relation to the pA signal across the 450 nt we delimited in the script. We observed that regardless of the pA signal efficiency, the USE is indeed more often found in the vicinity of a pA signal both in flies and humans, preferably within 90 nt (**FIGURE 22**). This is indicative that the USE might have a function in pA signal recognition in both species.

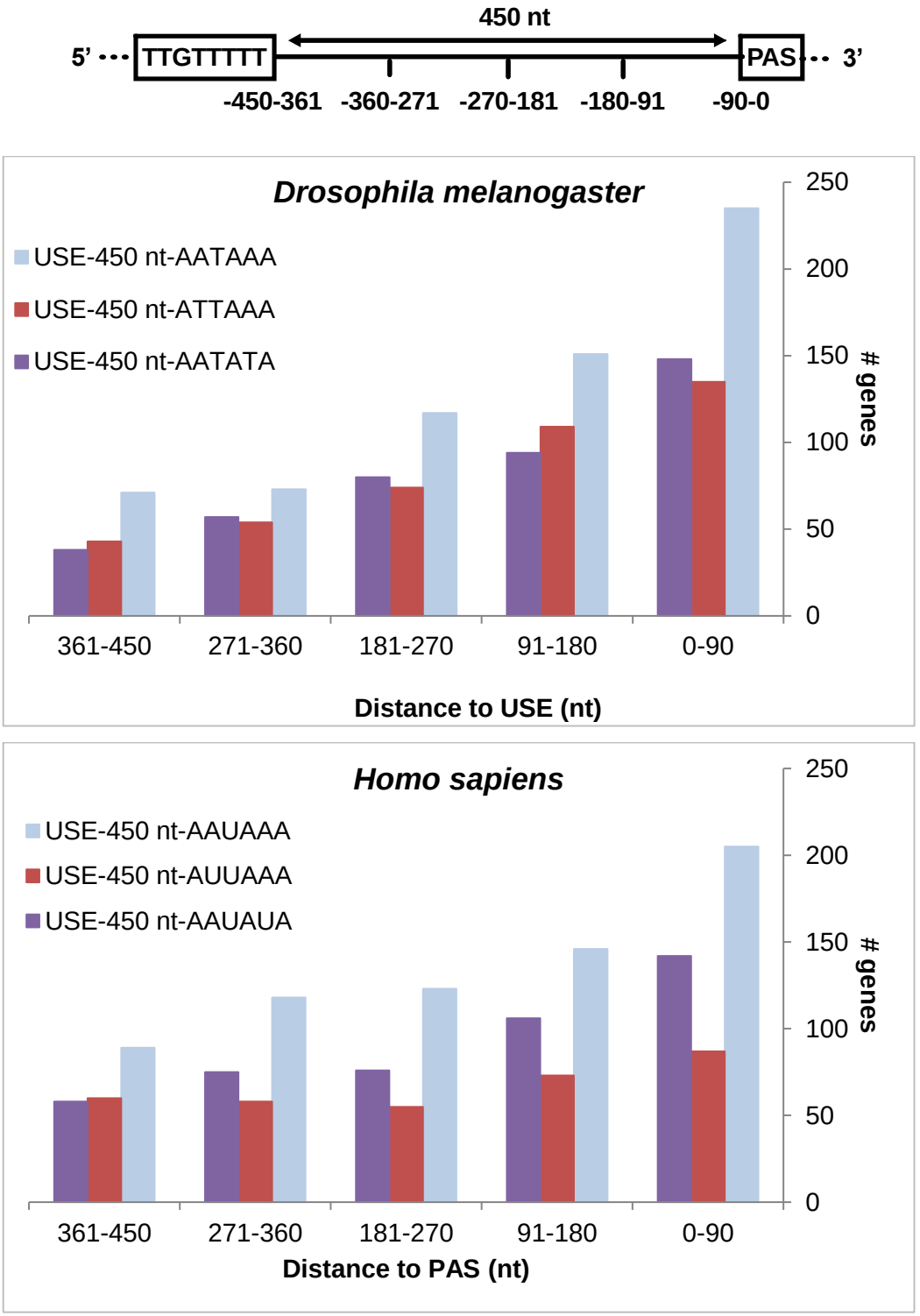


FIGURE 22 | The USE is located near pA signals both in human and fly genes. Genes were sorted according to distance (in nt) between the USE and each pA signals (AATAAA, ATTAAA or AATATA, distance divided in 90 nt slots: 0-90, 91-180, 181-270, 271-360 and 361-450, depicted right to left).

As the TTGTTTTT sequence is conserved across the *Drosophila* genus (**FIGURE 17**) and 3'UTRs tend to be AT-rich, we investigated the relevance of the G nt by performing a new

bioinformatic query to search for the TTNTTTTT sequence (USE-N) in the 3'UTRs of *Drosophila melanogaster* upstream of AATAAA, ATTAAA or AATATA. Within a distance of 90 nt, which corresponds to the distance range in which the USE is more often observed (**FIGURE 22**), we found that the USE-N is present in 7.1% of all the 3'UTRs across the fly genome (**TABLE 8**). Performing a similar analysis as in **FIGURE 21**, we normalized the USE-N-containing gene frequencies according to the respective genome-wide frequencies for each pA signal (**FIGURE 19**) and the absolute numbers of genes found with the new script (**TABLE 8**). We noted that the USE-N is equally more prevalent upstream of non-canonical pA signals (ATTAAA or AATATA) in *Drosophila melanogaster* (**TABLE 8**).

TABLE 8 | Number of *Drosophila melanogaster* USE-N-containing genes upstream of AATAAA, ATTAAA or AATATA and respective percentage normalized to each pA signal frequency. The TTNTTTTT sequence is more common upstream of weaker pA signals in the 3'UTR (ATTAAA and AATATA in comparison to AATAAA).

	# USE-N-containing genes	% USE-N-containing genes
AATAAA	738	9.0
ATTAAA	422	24.8
AATATA	454	26.7
Total	1616	7.1

To explore whether the presence of the USE influences gene expression, we used the *Drosophila* Gene Expression Tool (DGET, [557]). We analysed the expression of USE-N-containing genes in different developmental stages of the fruit fly (embryos, third instar larvae central nervous system and females, see **FIGURE 23**) and found that regardless of pA signal efficiency, there is an increase in the number of actively expressed genes with the USE-N than without the USE-N (compare the bars with and without the USE-N for each pA signal, **FIGURE 23**) in all developmental stages. Interestingly, we always observe a more obvious trend in females, followed by third instar larvae central nervous system and then in 06-08h embryos (**FIGURE 23**). These results suggest that the USE may enhance gene expression in a sex- and tissue-dependent manner in *Drosophila melanogaster*.

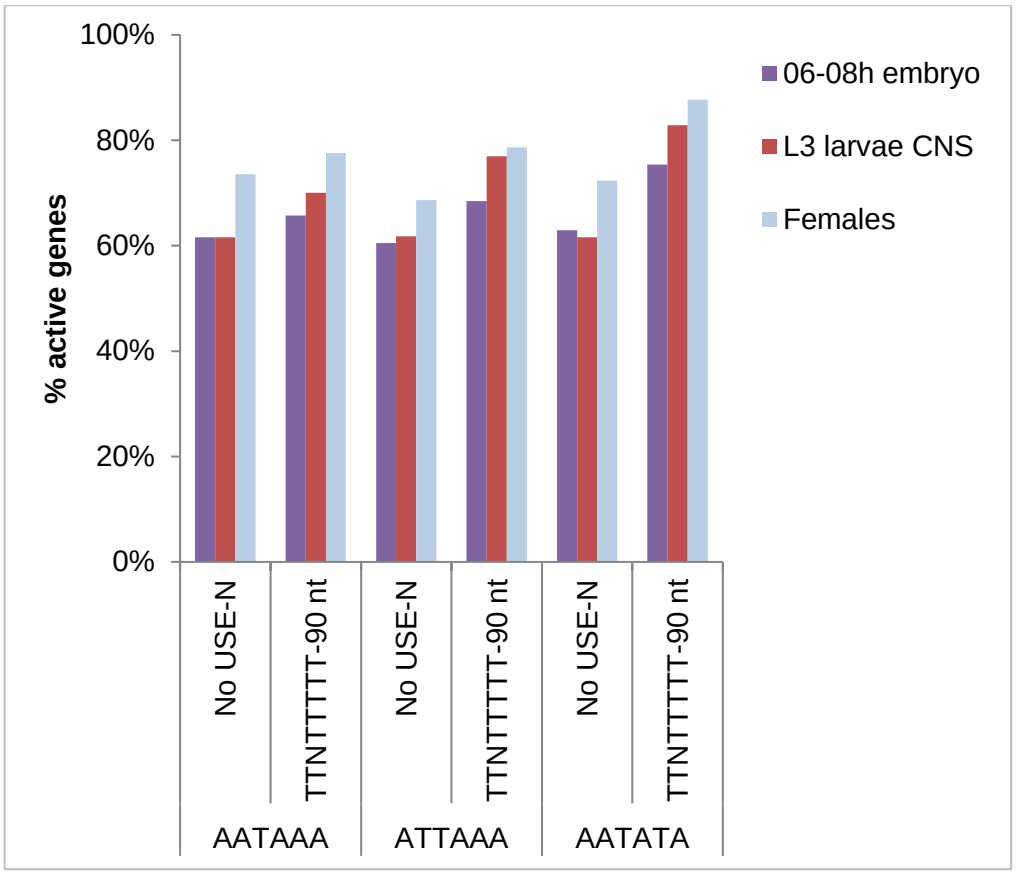


FIGURE 23 | Percentage of active *Drosophila melanogaster* genes with and without the USE-N up to 90 nt upstream of each of the three pA signals in the 3'UTR (AATAAA, ATTAAA or AATATA). Histograms show sorted data from DGET [556, 557] for 06-08 h embryos, third instar larvae CNS and one day old adult females.

2.3. The *polo* Upstream Sequence Element affects RNA Polymerase II occupancy along the *polo* gene

mRNA 3'end formation occurs co-transcriptionally [151, 170, 179, 566] and RNAPII transcriptional state has also been shown to have a role in APA [115, 153, 280, 534, 535]. We thus investigated the chromatin occupancy profiles of RNAPII along the *polo* locus by Chromatin Immunoprecipitation (ChIP) [538] in *Drosophila melanogaster* wild type embryos using a specific antibody against the RNAPII Rpb3 subunit (Rpb3 antibody).

We observe that RNAPII occupancy is highest in the region of primer pair 1. This suggests an accumulation of RNAPII which is characteristic of promoter proximal RNAPII pausing [57, 62, 64, 66-68]. Indeed, publically available ChIP-seq data [539] from 20-24h embryos using a specific antibody against the unmodified isoform of RNAPII (8WG16 antibody) confirms that *polo* contains promoter proximal paused RNAPII. Additionally, we noted that RNAPII occupancy is lowest along the 3'UTR of *polo* (**FIGURE 24**, primer pairs 4-5). This result is expected considering that *polo* transcription is terminating in this region

and we also corroborated this abrupt drop of the RNAPII signal in the *polo* 3'UTR by comparing the same RNAPII ChIP-seq data [539].

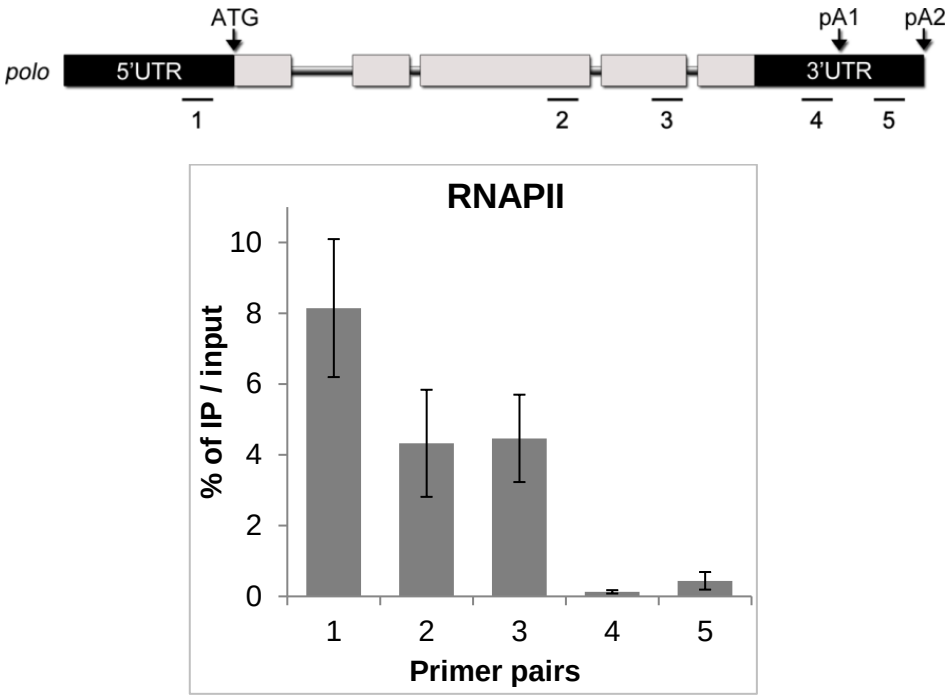


FIGURE 24 | RNAPII occupancy along the *polo* gene. ChIP analysis along the *polo* locus in *Drosophila melanogaster* 0-24h embryos using the Rpb3 antibody. Numbers below each bar set represent the respective primer pair as depicted in the schematic representation of the *polo* locus above (adapted from [567]). Results are the means ± SDE of three independent experiments.

Interestingly, *polo* USE is located within the region where the loss of RNAPII signal is more evident, defined by primer pair 4 (**FIGURE 24**). We hypothesized whether the reduction in RNAPII occupancy in this region was associated to the presence of *polo* USE. To study this at a genome-wide scale, we compared RNAPII chromatin occupancy profile in genes with and without the USE-N in the previously analysed dataset [539] (**FIGURE 25**). Remarkably, we found that the same abrupt decrease in the RNAPII signal in the 3'UTR is characteristic of USE-N-containing genes in comparison to genes without this sequence, regardless of pA signal strength (**FIGURE 25**, compare the RNAPII signal between the solid lines with the dashed lines of each colour). These genes contain the USE-N up to a distance of 90 nt upstream of the AATAAA, ATTAAA or AATATA pA signals, which approximately corresponds to the region where the loss of RNAPII signal is detected (**FIGURE 25**). These results suggest that the USE-N may not only be an APA regulatory sequence, but it may also aid in the disengaging of RNAPII from the chromatin, thus revealing a potential novel role for USEs in promoting RNAPII transcription termination in *Drosophila melanogaster*.

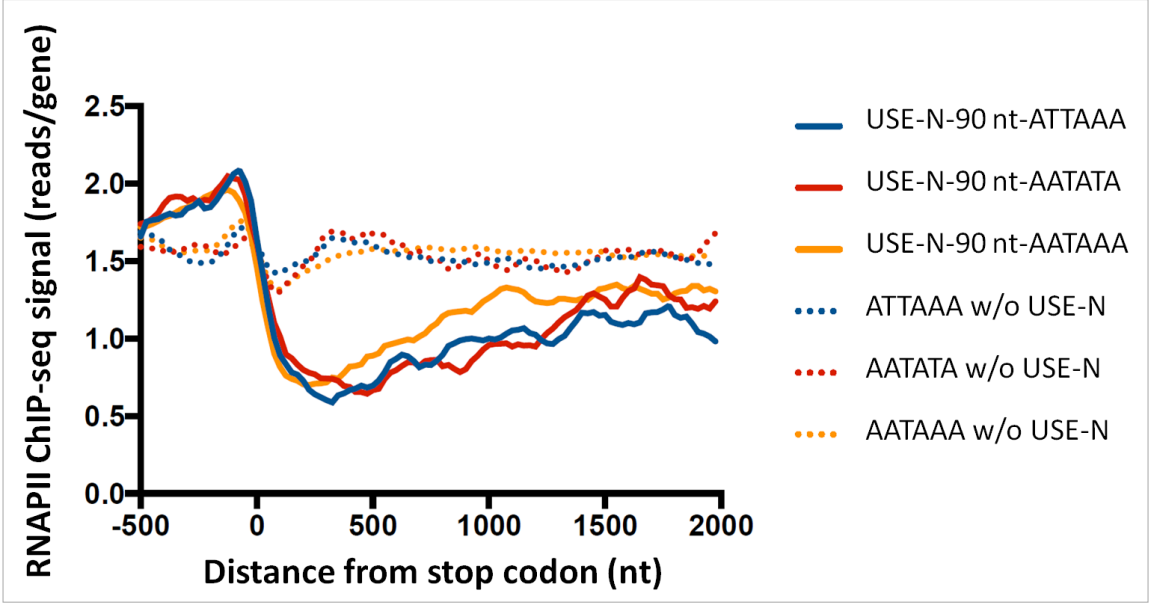


FIGURE 25 | Meta gene analysis of RNAPII ChIP-seq signal [539] surrounding the stop codon in USE-N-containing genes drops dramatically in comparison to genes without this element with AATAAA (orange lines), ATTAAA (blue lines) or AATATA (red lines) pA signals. 588, 297 and 319 USE-N-containing genes upstream of AATAAA, ATTAAA or AATATA (**TABLE 8**) and 4114, 1206 and 410 genes without this element upstream of AATAAA, ATTAAA or AATATA were used, respectively.

2.4. Deletion of *polo* Upstream Sequence Element causes a prevalent abdominal phenotype and impairs Polo activity and *polo* pA signal selection *in vivo*

Genome-wide analyses suggest that the USE may be a regulator of gene expression, APA and transcription termination in *Drosophila melanogaster* (see sections 2.2 and 2.3). To understand the role of the USE in *polo* gene regulation and mRNA 3'end formation, we created a transgenic fly strain in which the USE is deleted in a *polo* null background (*polo*^{9}, [410]). These transgenic flies contain the *polo* gene tagged with the *gfp* reporter (see **FIGURE 26**), they are viable and the *gfp-poloΔUSE* transgene can rescue the lethality induced by the *polo*^{9} background.

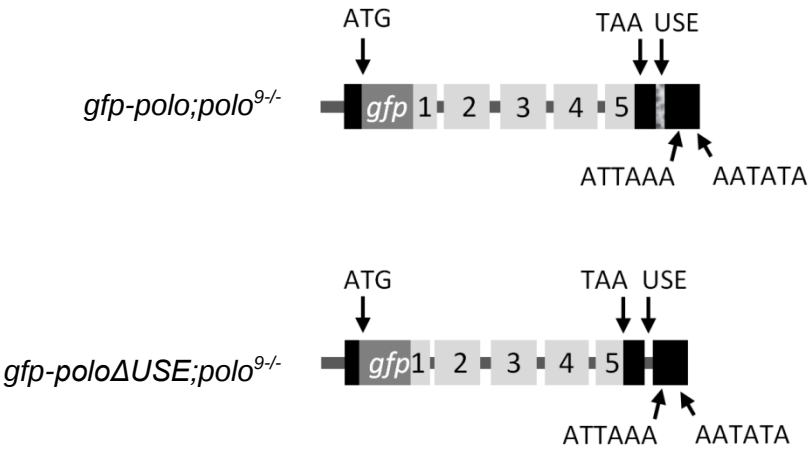


FIGURE 26 | Schematic representation of the transgene constructed to generate *gfp-poloΔUSE* transgenic flies in a *polo*^{9} background *(gfp-poloΔUSE;polo*$^{9-/-}$*). The 1-5 grey boxes represent the five *polo* exons with the black boxes representing the 5' and 3'UTRs. ATG refers to the translation start site and TAA to the stop codon. The *gfp* box represents the *gfp* coding sequence inserted in frame with the *polo* initiation codon. The textured box in the 3'UTR represents the USE, which was deleted in *gfp-poloΔUSE;polo*$^{9-/-}$ flies. The ATTAAA and AATATA depict the proximal (pA1) and distal (pA2) pA signals, respectively.

As previously shown, adult homozygous transgenic flies with a mutated and non-functional *polo* pA1 (ATTAAA > GTTAAC, *gfp-poloΔpA1;polo*$^{9-/-}$ flies) present abdominal defects [115]. Interestingly, *gfp-poloΔUSE;polo*$^{9-/-}$ homozygous transgenic flies show a remarkably identical phenotype to that of *gfp-poloΔpA1;polo*$^{9-/-}$ flies (**FIGURES 27** and **28**, as denoted by the black arrowheads). As seen in **FIGURES 27** and **28**, this phenotype is restricted to the abdomen and includes abnormalities in the bristles and tergites in both males and females. In contrast, the heterozygous individuals from both strains (*gfp-poloΔUSE;polo*^{9}*/TM6B* and *gfp-poloΔpA1;polo*^{9}*/TM6B*) display no phenotypic differences from the *gfp-polo;polo*$^{9-/-}$ control.

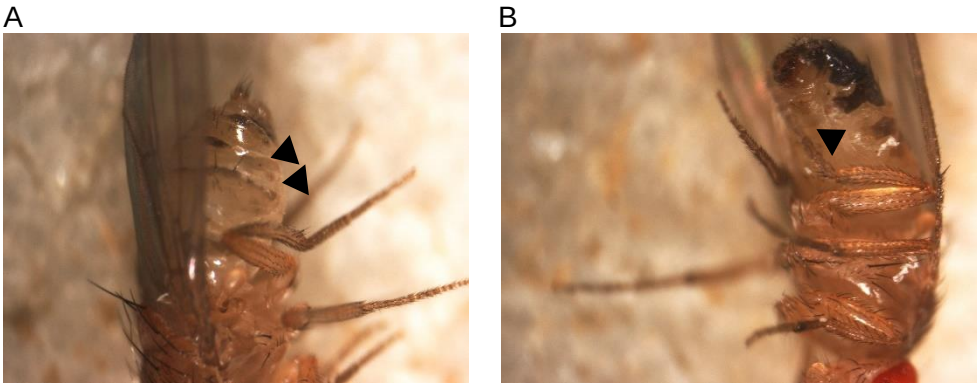


FIGURE 27 | Representative images of female (A and C) and male (B and D) *gfp-poloΔUSE;polo*$^{9-/-}$ adult abdomens. *gfp-poloΔUSE;polo*$^{9-/-}$ flies display an abdominal phenotype characterized by tergite malformation with pigmentation nicks and loss of bristles (indicated by the black arrow heads). Females (A, C) display a stronger phenotype than males (B, D).

FIGURE 28 | Dorsal view of abdomen preparations from *gfp-polo;polo*$^{9-/-}$ (panel 1), *gfp-poloΔUSE;polo*^{9}*/TM6B* (panel 2) and *gfp-poloΔUSE;polo*$^{9-/-}$ (panel 3) adult females. Anterior side is up.

We quantified these defects in the *gfp-poloΔpA1;polo*$^{9-/-}$ and *gfp-poloΔUSE;polo*$^{9-/-}$ strains in both males and females. We found that 86% of *gfp-poloΔUSE;polo*$^{9-/-}$ adults present an abdominal phenotype (**TABLE 9**). Interestingly, this is a sex-specific effect as 100% of *gfp-poloΔUSE;polo*$^{9-/-}$ females display phenotypic defects, in particular heavily malformed tergites (see **FIGURES 27** and **28**) against only 82% of *gfp-poloΔUSE;polo*$^{9-/-}$ males, which tend to present more moderate defects (**TABLE 9** and **FIGURE 28**). The phenotype in the *gfp-poloΔpA1;polo*$^{9-/-}$ flies is milder, with only 47% of adults displaying abdominal malformations (**TABLE 9**) and suggesting that the USE is more relevant for the phenotype presented than pA1; 66% of females display phenotypic abnormalities *versus* only 30% of males, with both sexes being fertile. In contrast to the males and the *gfp-poloΔpA1;polo*$^{9-/-}$

strain, *gfp-poloΔUSE;polo*$^{9-/-}$ females represent only 20% of the homozygous population and are sterile, which is in agreement with a *polo* loss of function such as the *polo*^{1} kinase dead condition [396] and *polo*^{3}, *polo*^{4}, *polo*^{5}, *polo*^{7} and *polo*^{8} mutants. Of note, *polo*^{7} and *polo*^{8} strains were also reported to present abdominal defects [568]. These results suggest that the USE has a function in *polo* gene expression and that deletion of *polo* USE *in vivo* (*gfp-poloΔUSE;polo*$^{9-/-}$ strain) has similar physiological repercussions as incorrect *polo* pA1 signal selection (*gfp-poloΔpA1;polo*$^{9-/-}$ strain).

TABLE 9 | Prevalence of the abdominal phenotype found in *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ adult flies compared to heterozygous individuals.

	gfp-poloΔUSE;polo$^{9-/-}$	*gfp-poloΔpA1;polo*$^{9-/-}$
Adult females with phenotype	25	59
Adult males with phenotype	84	29
Flies with phenotype	109	88
Total	127	186
% Adult females with phenotype	100%	65.6%
% Adult males with phenotype	82.4%	30.2%
% phenotype	85.8%	47.3%

We then asked if the absence of *polo* USE affects the biological pathways controlled by Polo. We focused on two well-characterized mitotic kinases, Aurora B and Mps1, which are Polo targets at the kinetochores of dividing neuroblasts that allow for correct cell cycle progression [430, 569-572]. In *gfp-polo*Δ*USE*;*polo*$^{9-/-}$ kinetochores, we observed that GFP-Polo levels decreased two-fold in comparison to the *gfp-polo;polo*$^{9-/-}$ control (**FIGURE 29**, left graph). Consequently, T-loop phosphorylation of Aurora B and Mps1 shows a similar reduction (**FIGURE 29**, middle and right graphs), indicating that kinetochore activation of these kinases is compromised upon deletion of *polo* USE.

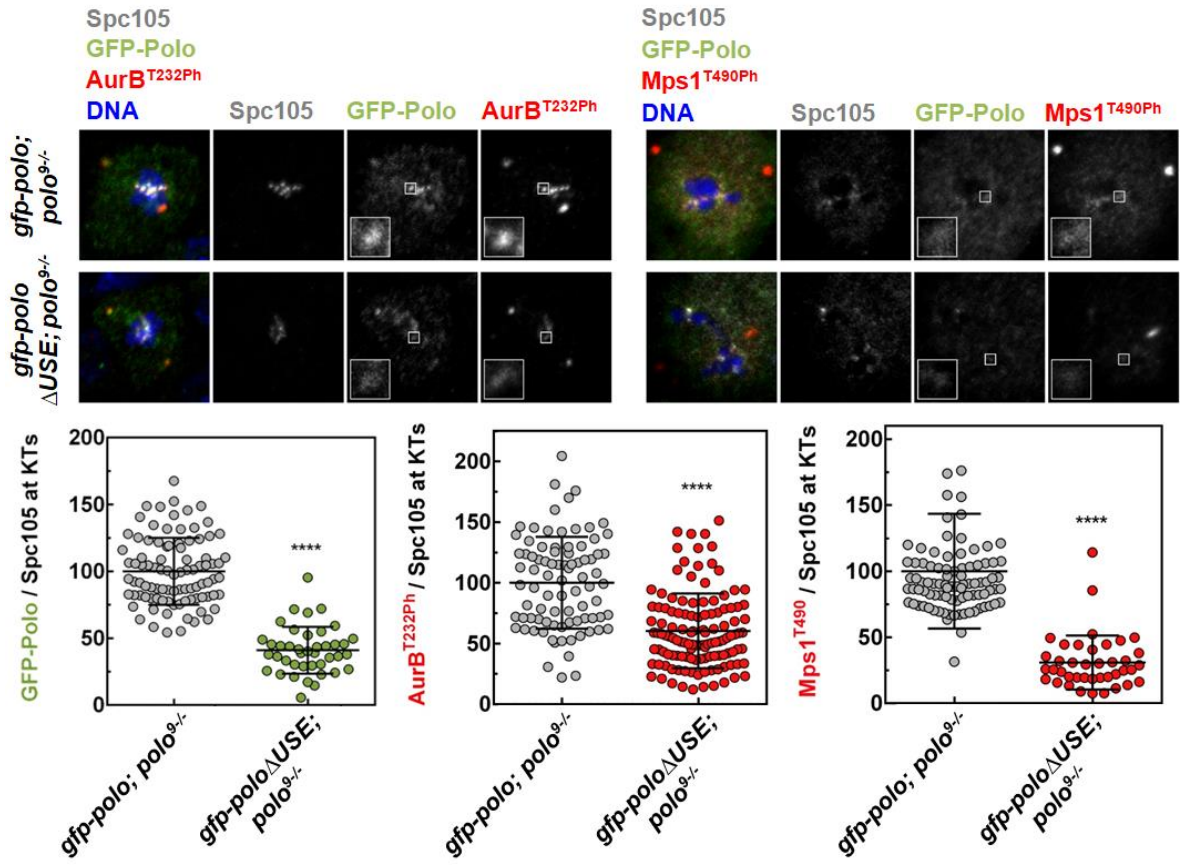


FIGURE 29 | Representative immunofluorescence images and respective quantifications of GFP-Polo, Aurora B^{T232Ph} and Mps1^{T490Ph} levels at kinetochores of dividing third instar larvae neuroblasts expressing *gfp-polo;polo$^{9-/-}$* and *gfp-poloΔUSE;polo$^{9-/-}$*. GFP-Polo, Aurora B^{T232Ph} and Mps1^{T490Ph} fluorescence intensities were determined relative to Spc105 signal, which was used as a kinetochore marker. All values were normalized to the control (*gfp-polo;polo$^{9-/-}$* neuroblasts) mean fluorescence intensity, which was set to 100% (n≥7 kinetochores from at least 10 neuroblasts for each condition). Results are expressed as mean values ± SDE and statistical significance was assessed by the Mann-Whitney test.

These results clearly show that *in vivo* deletion of *polo* USE has an impact in the expression and activity of the cell cycle kinases Polo, Aurora B and Mps1 at the kinetochores.

Although *gfp-poloΔUSE;polo$^{9-/-}$* individuals are viable, it is interesting to note that they present a stronger mitotic phenotype (more cells in mitotic arrest and fewer cells) than *gfp-poloΔpA2;polo$^{9-/-}$* or *polo$^{9-/-}$*, which are lethal at the pupae stage [115] and late third instar stage [410], respectively. These observations further reveal that *polo* USE has a relevant *in vivo* function in the normal development of the larvae brain by controlling *polo* expression.

Reduced activity of Mps1 and Aurora B in mitosis often leads to chromosome segregation errors. Accordingly, we detected *gfp-poloΔUSE*;*polo*$^{9-/-}$ (**FIGURE 30**) neuroblasts containing an abnormal number of chromosomes as opposed to the regular karyotype observed in every control neuroblast analysed (8 chromosomes, *gfp-polo;polo*$^{9-/-}$, **FIGURE 30**), indicating that USE-less neuroblasts are more prone to mitotic errors than control cells.

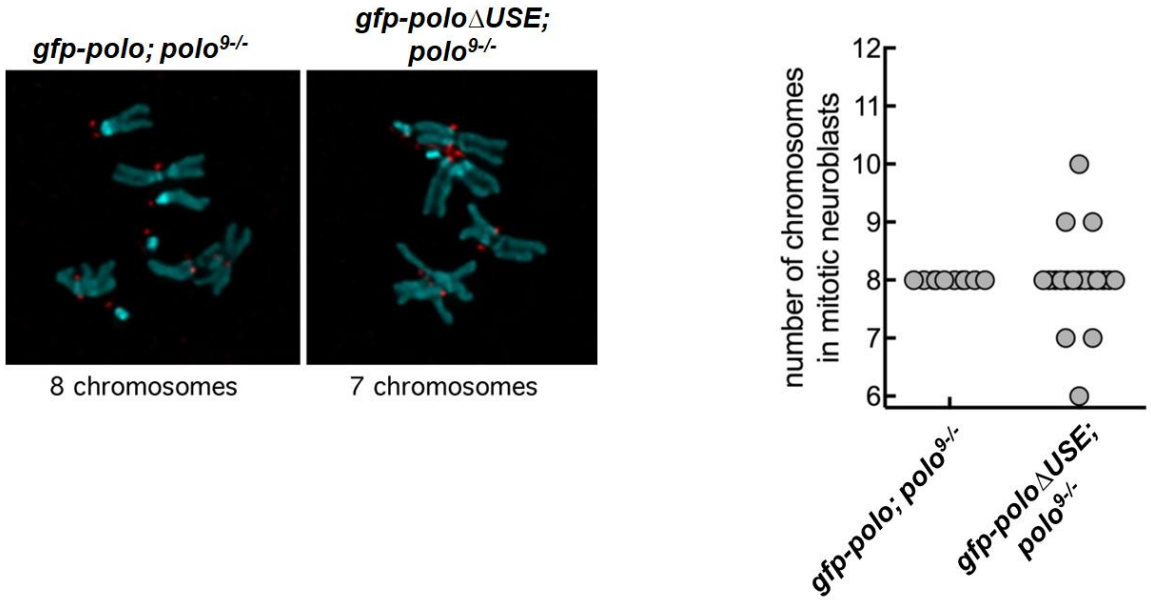


FIGURE 30 | Representative immunofluorescence images of mitotic neuroblasts from squashed *gfp-polo;polo*$^{9-/-}$ and *gfp-poloΔUSE;polo*$^{9-/-}$ larvae brains. Brains were incubated with colchicine to obtain chromosome spreads that facilitate the visualization of chromosome content. Spc105 was used as a kinetochore reference. Chromosome content is shown for each representative neuroblast. The graph represents the quantification of chromosome numbers in *gfp-polo;polo*$^{9-/-}$ and *gfp-poloΔUSE;polo*$^{9-/-}$ mitotic neuroblasts.

Taken together, our results indicate that the decreased activity of Polo caused by the *in vivo* deletion of the USE (**FIGURE 29**) leads to abnormal cell cycles and chromosome instability, which ultimately culminates in mitotic arrests, errors and aneuploidies (**FIGURE 30**). While these abnormalities are not enough to be lethal, they clearly have an effect in dividing cells (neuroblasts), and are likely the cause of the abdominal phenotype seen in *gfp-poloΔUSE;polo*$^{9-/-}$ adults. This is perhaps a milder failure of incorrect proliferation of the abdominal histoblasts which was observed in *gfp-poloΔpA2;polo*$^{9-/-}$ individuals [115].

To investigate the function of the USE in *polo* mRNAs expression, we quantified the levels of both *polo* mRNA isoforms in *gfp-poloΔUSE;polo*$^{9-/-}$ by RT-qPCR using two different primer pairs as depicted by the schematics in **FIGURE 31**: one that amplifies both mRNAs and another one able to specifically amplify only the longest *polo* isoform. **FIGURE 31**

shows that there is a reduction of both *polo* mRNA isoforms in *gfp-poloΔUSE;polo*$^{9-/-}$ neuroblasts when compared to the *gfp-polo;polo*$^{9-/-}$ control.

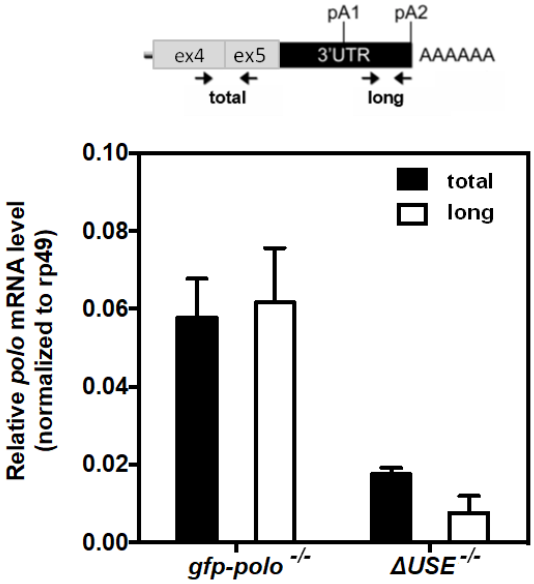


FIGURE 31 | Expression levels of *polo* mRNA isoforms of *gfp-poloΔUSE;polo*$^{9-/-}$ (*ΔUSE*$^{-/-}$) third instar larvae brains compared to the *gfp-polo;polo*$^{9-/-}$ control (*gfp-polo*$^{-/-}$). Expression was quantified by RT-qPCR using two distinct primer pairs: the total primer pair amplifies both mRNA isoforms and the long primer pair amplifies only the longest isoform. *rp49* was used as a housekeeping gene. Error bars show SEM from three independent experiments.

Additionally, the expression of the longest *polo* mRNA, which corresponds to the essential isoform responsible for producing most of Polo protein [115], is more compromised. This further indicates that deleting a potentially regulatory element from the 3'UTR of *polo* deregulates its mRNAs expression, hinders Polo kinase activity and consequently leads to mitotic delay, aneuploidies and abnormal defects.

Knowing that *polo* expression is altered in the *gfp-poloΔUSE;polo*$^{9-/-}$ strain, we then investigated whether the *in vivo* deletion of the USE influenced *polo* mRNA 3'end formation. For this purpose, we mapped *polo* mRNA 3'ends by 3'RACE in *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ females, the developmental stage in which we observed the predominant abdominal phenotype (**FIGURES 27** and **28** and **TABLE 9**). *gfp-polo;polo*^{9}*/TSTL* females were used as a control and produce the two *polo* mRNA isoforms that have been previously described [397]. In contrast, flies without the USE (*gfp-poloΔUSE;polo*$^{9-/-}$) and with a mutated *polo* pA1 (*gfp-poloΔpA1;polo*$^{9-/-}$) both show impaired pA signal selection, using two cryptic pA signals in addition to *polo* pA1 and pA2 to produce two new *polo* mRNA isoforms (# and * bands in **FIGURE 32**). This indicates that

the USE has a function in *polo* pA signal selection, contributing to proper *polo* mRNA 3'end formation and the recognition of *bona fide polo* pA signals *in vivo*.

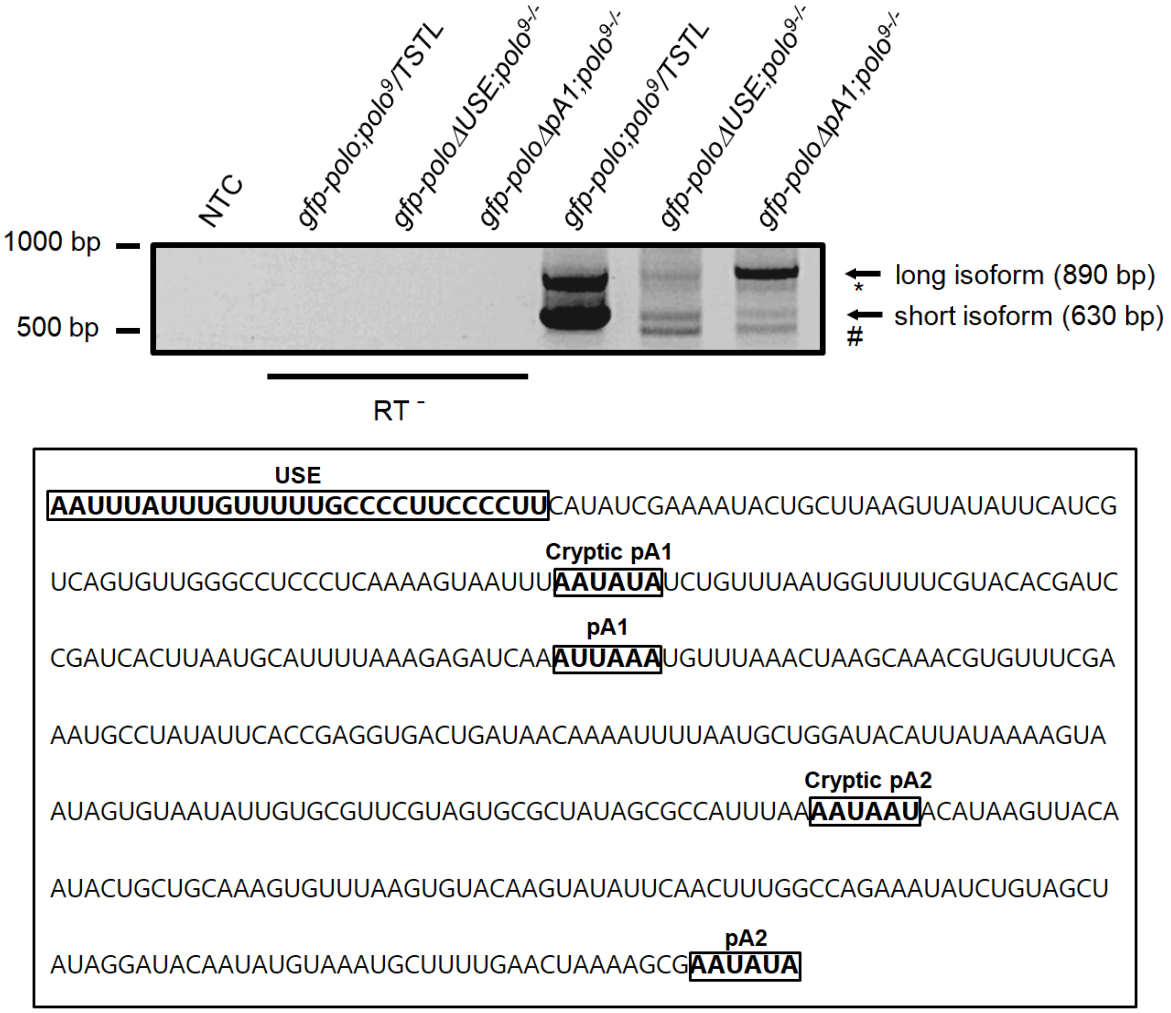


FIGURE 32 | mRNA 3'end mapping of *polo* in the abdomen of *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ female adults by 3'RACE. *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ flies produce two additional *polo* mRNA (depicted by * and #) isoforms other than the expected short (630 bp) and long (890 bp) mRNAs seen in the control (*gfp-polo;polo*^{9}*/TM6B*). NTC is a no template control and RT^{-} are the negative controls. In the 3'UTR sequence of *polo*, we highlighted the position of the USE, both pA signals (pA1 and pA2), the cryptic pA1 (AAUAUA) 57 bp upstream of pA1 and the cryptic pA2 (AAUAAU) 113 bp upstream of pA2.

2.5. Hephaestus is an RNA Binding Protein that binds to *polo* Upstream Sequence Element

Our previous results clearly show that the USE has a physiological role in correct *polo* expression. We next focused on identifying the RBPs that specifically bind to the USE RNA as they are the true effectors of USE-dependent *polo* regulation. For this purpose,

we performed an RNA-protein pull-down assay associated to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis (**FIGURE 33**) using as templates an *in vitro* transcribed 28-nt long USE RNA and a mutated USE (USEmt) in which 17 nt were specifically mutated to disrupt the most conserved nucleotides of the sequence, including TTGTTTTT (**FIGURE 34**).

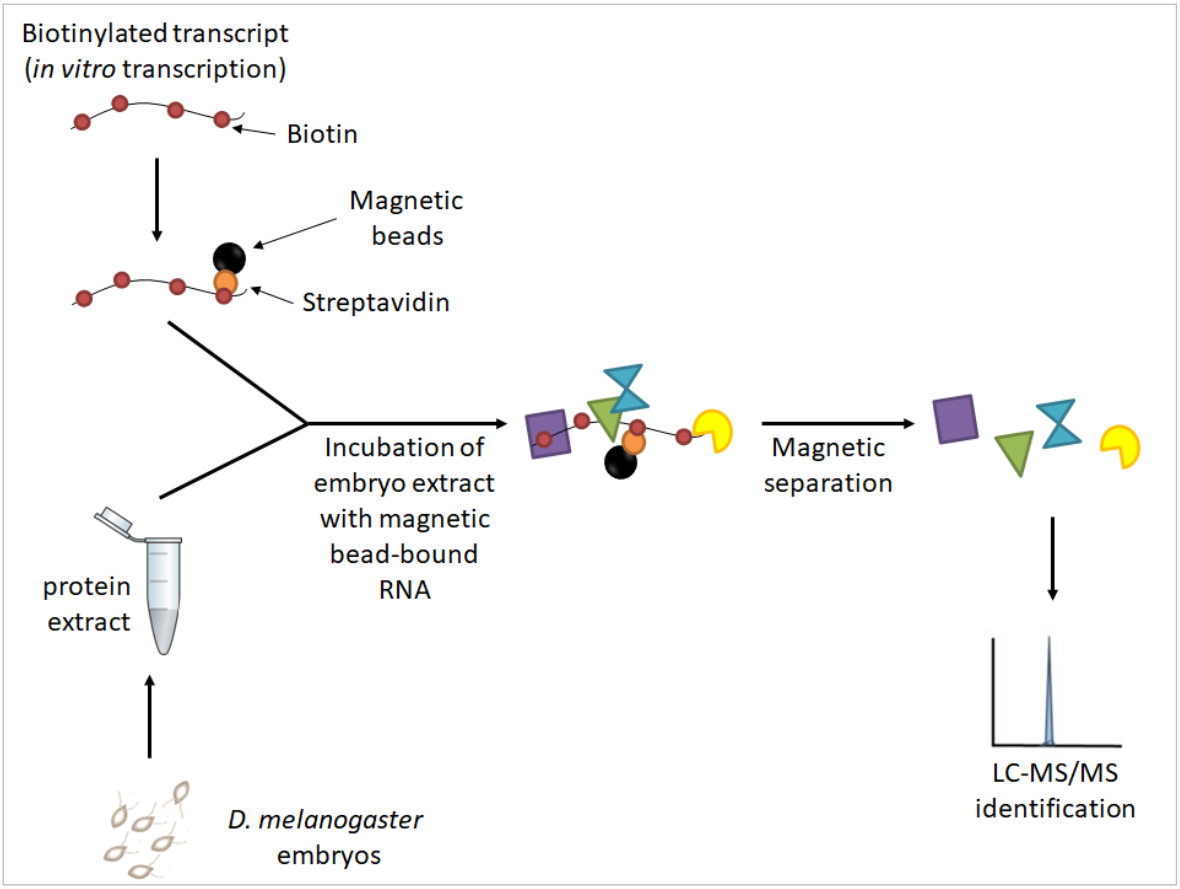


FIGURE 33 | Schematic representation of the RNA-protein pull-down assay protocol. *Drosophila melanogaster* embryo proteins were exposed to biotin-labelled USE or USEmt RNAs, purified and later identified by MS analysis. Further detail can be found in sections 2.6 and 2.7 of Materials and Methods.

USE: AAUUUAUUUGUUUUUGCCCCUUCCCCUU

USEmt: AAUUCAUAGACGACAAACGCAAGAUCUU

FIGURE 34 | Sequences of the USE and USEmt RNAs used in the RNA-protein pull-down assay (5'-3'). * above the USEmt sequence indicate the 17 nt single point mutations in relation to the USE RNA.

We normalized both the USE and USEmt eluates to the same total amount of protein before loading the LC-MS/MS (see section 2.7 of Materials and Methods for further detail) and protein profiles from the eluates of USE and USEmt RNAs before LC-MS/MS separation show clear differences in the Coomassie-stained gel of **FIGURE 35**.

The proteins able to bind to the USE but not to USEmt were considered positive hits if at least five unique peptides for the USE RNA (#unique peptides$_{USE}$ ≥ 5) and less than five unique peptides for the USEmt RNA (#unique peptides$_{USEmt}$ < 5) had been identified by LC-MS/MS.

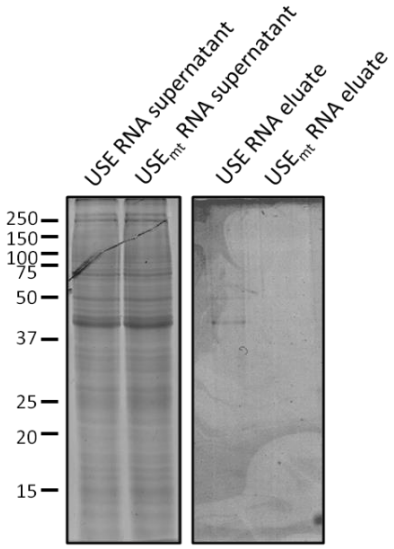


FIGURE 35 | Representative Coomassie-stained gel showing the RBPs from *Drosophila melanogaster* embryo extracts pulled down with the USE and USEmt RNAs using the Pierce RNA-Protein pull-down assay kit. Samples from both the supernatants and eluates and all washing steps (not shown) were run to assess protein recovery in each condition.

GO analyses show an enrichment of functions related to mRNA 3'end processing for many of the USE-binding proteins (**TABLE 10**) such as poly(A)$^+$ mRNA export from nucleus, regulation of mRNA processing and mRNA polyadenylation, which is in agreement with the expected roles of RBPs.

TABLE 10 | Top 10 RNA-related GO terms enriched in the proteins that bind to the USE RNA detected by LC-MS/MS.

Top 10 RNA-related GO terms	USE-binding proteins (fold enrichment)
RNA cap binding complex	6.60
poly(A)$^+$ mRNA export from nucleus	5.03
mRNA binding	3.78
regulation of mRNA processing	3.77
mRNA polyadenylation	3.52
regulation of mRNA metabolic process	3.46
mRNA export from nucleus	3.42
mRNA-containing ribonucleoprotein complex export from nucleus	3.42
mRNA processing	3.38
RNA export from nucleus	3.32

TABLE 11 displays the top 10 USE-binding proteins involved in RNA metabolism sorted by the number of unique peptides identified per protein and the complete list of USE-binding proteins can be found in **TABLE 12** (see Appendix). Heph is involved in Gurken protein location [322] and *oskar* expression regulation [327]. Its human ortholog, PTBP1, is an RBP involved in mRNA 3'end processing, APA, alternative splicing and translation [250, 256, 262, 266, 311-318], although no role in APA has been described for Heph. Arsenic resistance protein 2 (Ars2) is involved in the biogenesis of small RNAs [573] and Tho2 has a function in exporting mRNAs from the nucleus [574]. Splicing factor 3b subunit 3 (Sf3b3) is part of the U2 small nuclear ribonucleoprotein particle and also part of a transcriptional co-activator complex [575]. There is insufficient information regarding the CG7728, Suppressor of white-apricot [Su(w^a)] and CG5728 proteins to understand if they may have a function related to mRNA metabolism while Embargoed (Emb), Dynactin 1 p150 subunit (Dctn1-p150) and Spindle E (SpnE) have no reported RNA-related function.

Heph is the top positive hit, specifically binding to the USE RNA and not to the USEmt (#unique peptides$_{\text{USE RNA}}$ = 15 and #unique peptides$_{\text{USEmt RNA}}$ = 4), therefore we decided to further investigate the *in vivo* function of Heph in regulation of *polo* APA and/or expression.

TABLE 11 | List of the top 10 RNA-related proteins obtained in LC-MS/MS that bind to the USE (#unique peptides ≥ 5) and do not bind to USEmt (#unique peptides < 5) sorted from largest to smallest #unique peptides identified/protein. The total number of positive hits comprises 282 proteins and can be found in **TABLE 12** (see Appendix).

Protein	#Unique peptides pulled down with USE RNA	#Unique peptides pulled down with USE_{mt} RNA
Heph	15	4
Emb	15	4
CG7728	14	0
Ars2	14	3
Dctn1-p150	13	4
Tho2	13	4
Su(w^a)	12	3
CG5728	12	4
Sf3b3	12	4
SpnE	11	3

2.6. Hephaestus modulates efficient *polo* mRNA 3'end formation and Polo protein production

To study the *in vivo* function of Heph on *polo* expression and APA, we used a hypomorph fly strain for the respective gene, the *heph*^{2}*/TM6B* mutant. These flies do not have abdominal abnormalities as the *gfp-poloΔUSE;polo*$^{9-/-}$ strain, but they have a phenotype in testis [325] and display defects in spermatogenesis [324] that lead to male sterility [388].

heph^{2} is mostly lethal for most of the homozygous population, originating few escapers. Consequently, we chose to work with heterozygous individuals as this allele had previously been reported to produce low levels of *heph* mRNA and protein [324, 388]. To validate the use of this heterozygous model, we measured *heph* mRNA levels in *heph*^{2}*/TM6B* larvae brains by RT-qPCR and as seen in **FIGURE 36**, *heph* expression is reduced by approximately half the amount found in *w*^{1118} control individuals.

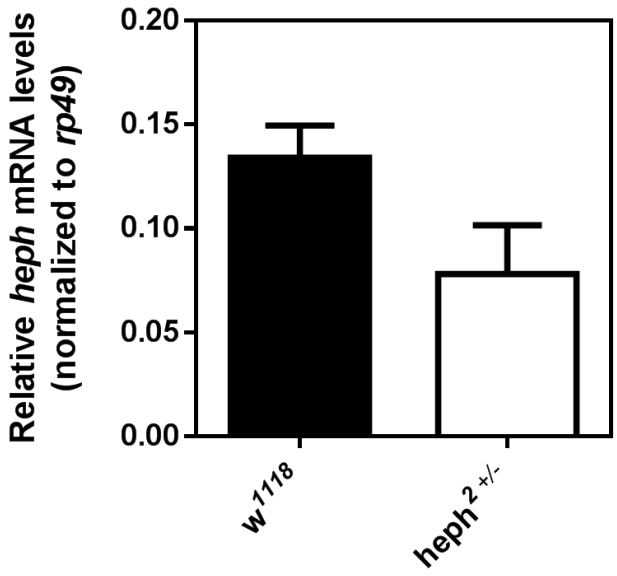


FIGURE 36 | Expression levels of *heph* mRNA in *heph²/TM6B* third instar larvae brains quantified by RT-qPCR. *rp49* was used as a housekeeping gene. Error bars show SD from three independent experiments.

By RT-qPCR and adopting a similar methodology as that described in section 2.4, we investigated the levels of *polo* mRNA isoforms in *heph²/TM6B* mutant larvae brains and observed that the longest *polo* isoform is significantly reduced in comparison to the wild type control (**FIGURE 37**), similarly to what we had previously observed with the *gfp-poloΔUSE;polo$^{9-/-}$* strain. Using *Drosophila* S2 cells treated with *heph* dsRNA, we also quantified *polo* mRNA levels and showed a decrease of the longest *polo* isoform levels when *heph* is knocked down [278], corroborating our *in vivo* results. These results show that low levels of Heph lead to a decrease in the expression of the longest *polo* mRNA isoform in a similar manner as deleting the USE *in vivo*, an effect that is possibly due to the lack of binding of Heph to *polo* USE.

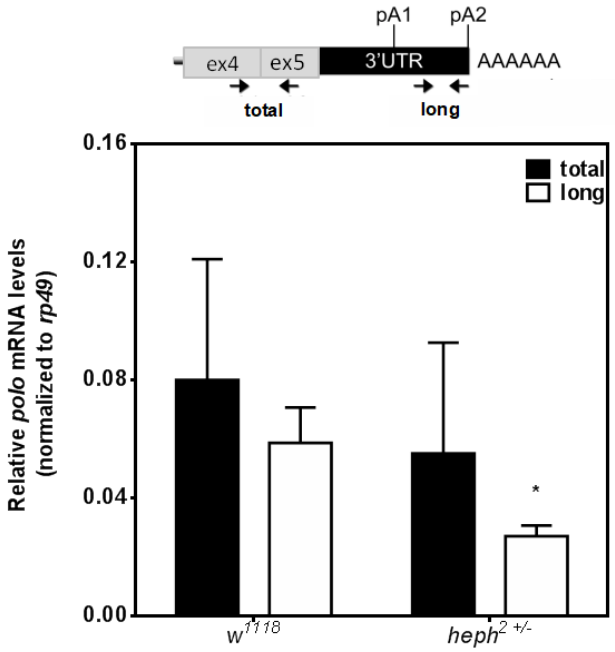


FIGURE 37 | Expression levels of total *polo* mRNA in *heph²/TM6B* third instar larvae brains quantified by RT-qPCR. *rp49* was used as a housekeeping gene. Error bars show SD from three independent experiments. * indicates *p* value < 0.05.

To further understand if the absence of Heph has an impact on *polo* mRNA 3'end formation analogous to *gfp-poloΔUSE;polo*$^{9-/-}$, we proceeded to map the 3'ends of *polo* mRNAs by 3'RACE in *heph*^{2}*/TM6B* flies. While the *heph*^{2}*/TM6B* mutant displays a different 3'RACE pattern (three bands, **FIGURE 38**) from *gfp-poloΔUSE;polo*$^{9-/-}$ transgenic flies (four bands, **FIGURE 32**), this mutant clearly presents abnormal *polo* mRNA 3'end formation as well: *heph*^{2}*/TM6B* mutants do not utilize the cryptic pA signal upstream of *polo* pA1 that both *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ do, but they use the cryptic *polo* pA signal located upstream of *polo* pA2 (*, **FIGURE 38**) also used by both the *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ strains (compare the *heph*^{2}*/TM6B* lane in **FIGURE 38** with the *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ lanes in **FIGURE 32**). This result is in agreement with the reduced levels of the distal *polo* mRNA isoform in *heph*^{2}*/TM6B* larvae brains, both *in vivo* and *in vitro* (**FIGURE 37** and [278]) and corroborates our hypothesis that the lack of Heph and *polo* USE deletion both have a deleterious effect on normal *polo* mRNA 3'end formation.

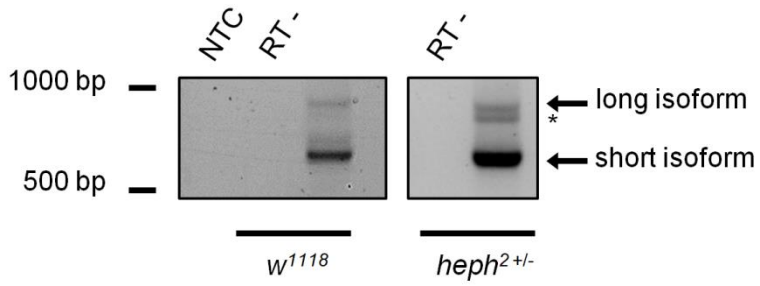


FIGURE 38 | mRNA 3'end mapping of *polo* in *heph^2/TM6B* 0-24h embryos by 3'RACE. *polo* pA signal usage in the *heph^2/TM6B* mutant shows an additional mRNA isoform (*) compared to the *w^{1118}* control. NTC is the no template control and RT$^-$ are the negative controls.

To complement our analysis of *polo* mRNA 3'end formation and mRNA production regulated by the USE and Heph, we have also performed northern blot in *gfp-poloΔUSE;polo^9/TM6B*, *heph^2/TM6B* and *gfp-poloΔpA1;polo^9/TM6B* third instar larvae brains (**FIGURE 39**). *heph^2/TM6B* mutants show a decrease of both *polo* mRNAs in comparison to the *w^{1118}* control (**FIGURE 31**, compare lanes 1 and 3), which is in agreement with our previous observations.

gfp-poloΔpA1;polo^9/TM6B larvae brains can only generate the *gfp-polo* pA2 isoform due to *polo* pA1 mutation [115] (**FIGURE 39**, compare lanes 2 and 4), and *gfp-poloΔUSE;polo^9/TM6B* individuals present a single band at the same molecular size, indicating that *in vivo* deletion of the USE hinders *polo* pA1 selection and leads to the production of only the *gfp-polo* pA2 isoform. While the endogenous *polo* mRNAs are visible in all strains, we were not able to detect the bands of weaker intensity detected by 3'RACE produced by usage of cryptic pA signals by northern blot.

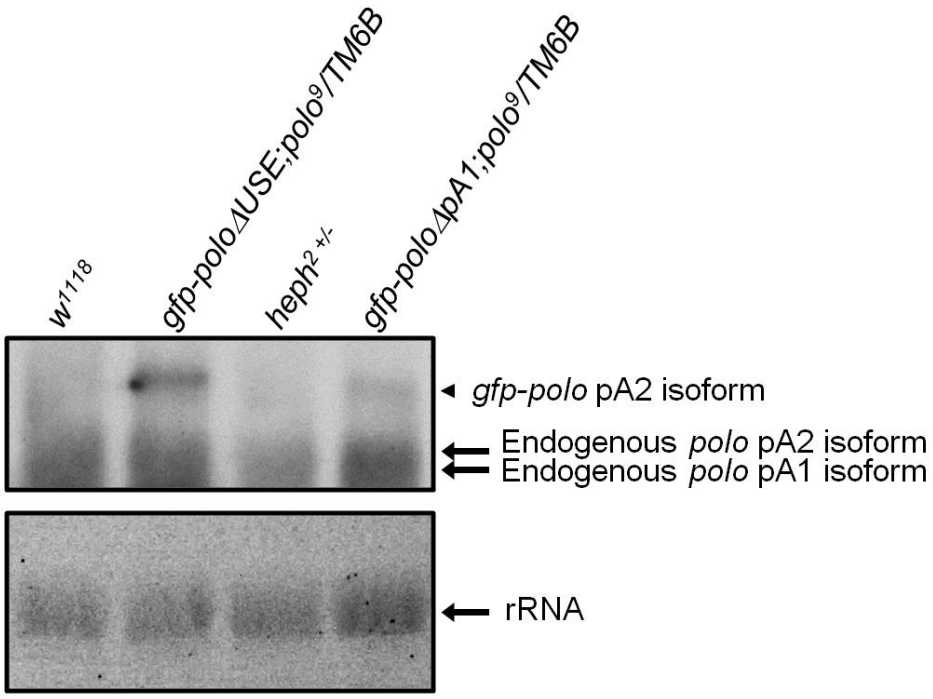


FIGURE 39 | Northern blot using total RNA from w^{1118} (lane 1), *gfp-poloΔUSE;polo^{9}/TM6B* (lane 2), *heph^{2}/TM6B* (lane 3) and *gfp-poloΔpA1;polo^{9}/TM6B* (lane 4) third instar larvae brains. The two endogenous *polo* mRNAs present in every lane (endogenous *polo* pA1 and pA2 isoforms) are indicated by black arrows. Only the *gfp-polo* pA2 isoform is expressed by the *gfp-poloΔUSE;polo^{9}*/TM6B and *gfp-poloΔpA1;polo^{9}*/TM6B strains as indicated by the black arrowhead. rRNA served as loading control.

We then asked how was GFP-Polo and Polo protein production in the wild type *w^{1118}*, *heph^{2}/TM6B*, *gfp-polo;polo$^{9-/-}$*, *gfp-poloΔUSE;polo$^{9-/-}$*, *gfp-poloΔpA1;polo$^{9-/-}$* and *gfp-poloΔpA2;polo$^{9-/-}$* strains. A remarkable reduction in total Polo protein levels in the *heph^{2}/TM6B* mutant in comparison to wild type is clearly seen (**FIGURE 40**), confirming that Heph is necessary for Polo protein expression. Curiously, we do not observe a significant decrease in total GFP-Polo levels in *gfp-poloΔUSE;polo$^{9-/-}$* in comparison to the *gfp-polo;polo$^{9-/-}$* control. This is an unexpected result as we found low levels of GFP-Polo in the kinetochores of *gfp-poloΔUSE;polo$^{9-/-}$* neuroblasts (**FIGURE 29**). As *gfp-poloΔUSE;polo$^{9-/-}$* neuroblasts still generate the longer *gfp-polo* isoform (**FIGURE 31** and **39**), which is main responsible for GFP-Polo protein production [115], total GFP-Polo levels remain mostly unaltered (**FIGURE 40**), but GFP-Polo activity and recruitment to the kinetochores is affected in *gfp-poloΔUSE;polo$^{9-/-}$* by a still unknown mechanism (**FIGURE 29**). Without the distal *gfp-polo* isoform, total GFP-Polo levels are noticeably diminished as seen in *gfp-poloΔpA2;polo$^{9-/-}$* neuroblasts when compared to *gfp-polo;polo$^{9-/-}$* (**FIGURE 40**), corroborating our previous results: the shorter *gfp-polo* mRNA cannot compensate for

the loss of the longer isoform regarding Polo protein production during metamorphosis, which eventually culminates in the death of *gfp-poloΔpA2;polo*$^{9-/-}$ individuals [115]. In contrast, loss of Heph compromises *polo* distal pA signal selection (as seen in **FIGURES 37** and **38**), subsequently inducing a more detrimental outcome on Polo protein production (**FIGURE 40**). Taken together, our results reveal a novel role for Heph in *polo* distal pA signal selection and Polo protein expression.

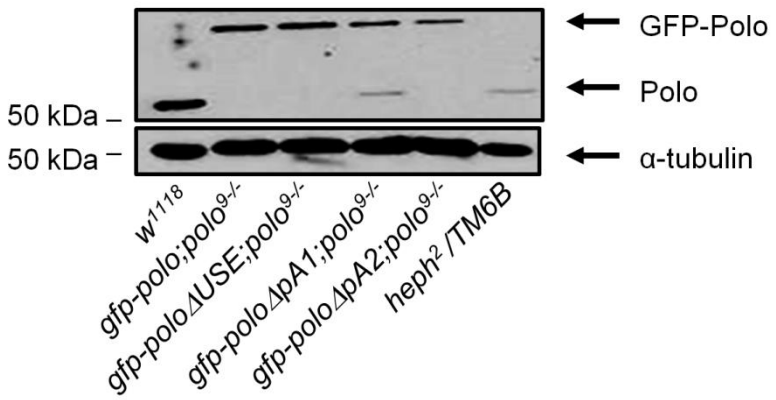


FIGURE 40 | Representative western blot showing the total GFP-Polo and Polo protein levels (anti-Polo MA294, 1:40) in *w*^{1118}, *gfp-polo;polo*$^{9-/-}$, *gfp-poloΔUSE;polo*$^{9-/-}$, *gfp-poloΔpA1;polo*$^{9-/-}$, *gfp-poloΔpA2;polo*$^{9-/-}$ and *heph*^{2}*/TM6B* mutants. α-tubulin (anti-α-tubulin DM1A, 1:20000) was used as loading control.

To further understand the physiological function of Heph in Polo-dependent pathways, we investigated the levels of Polo and two Polo downstream targets, Aurora B and Mps1, in dividing *heph*^{2}*/TM6B* neuroblasts. Similarly to *gfp-poloΔUSE;polo*$^{9-/-}$ (**FIGURE 29**), *heph*^{2}*/TM6B* cells also display a suboptimal accumulation of Polo at the kinetochores (Polo panel, left graph, **FIGURE 41**) and its kinase activity is also hindered since the phosphorylation levels of both Aurora B and Mps1 are equally reduced (AurB^{T232Ph} and Mps1^{T490Ph} panels, middle and right graphs, **FIGURE 41**). These results show that the absence of Heph is similar to the *in vivo* deletion of *polo* USE and illustrate a new *in vivo* effect of Heph on Polo and downstream Polo targets with a potentially adverse outcome in mitotic progression.

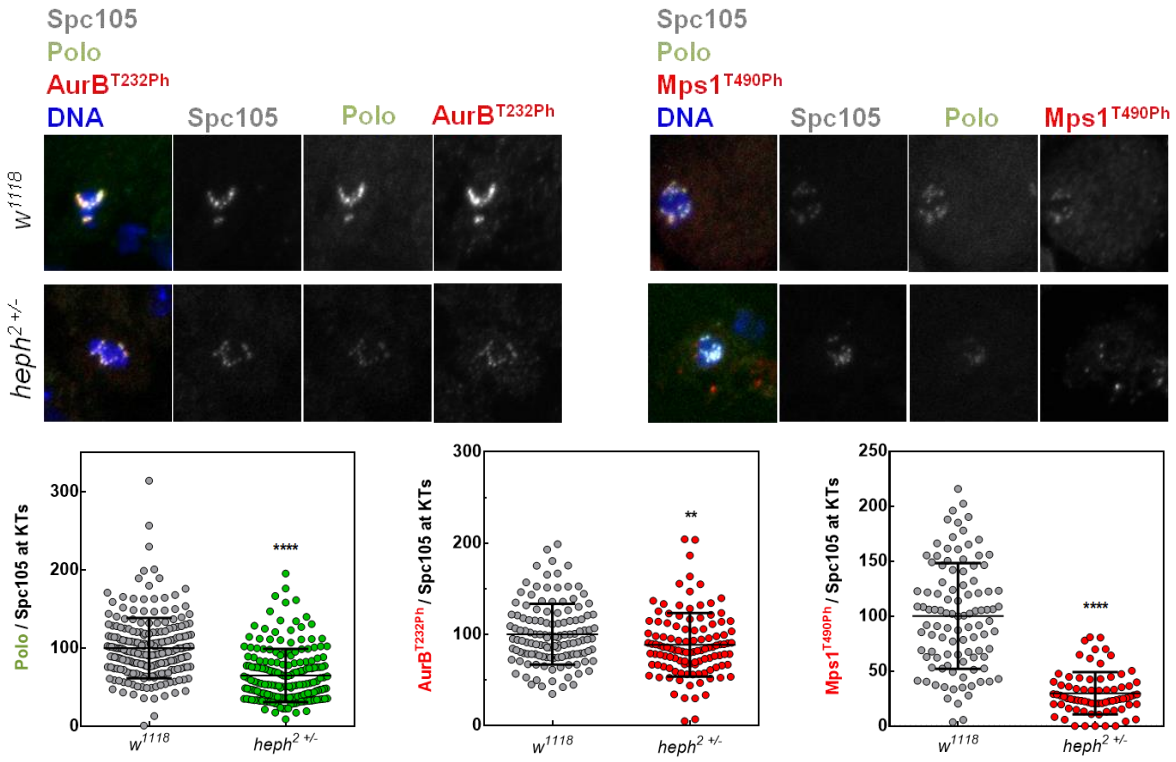


FIGURE 41 | Representative immunofluorescence images and respective quantifications of Polo, Aurora B^{T232Ph} and Mps1^{T490Ph} levels at kinetochores of dividing third instar larvae neuroblasts from *w^{1118}* (control) and *heph^{2}/TM6B* mutants. Polo, Aurora B^{T232Ph} and Mps1^{T490Ph} fluorescence intensities were determined relative to Spc105 signal, which was used as a kinetochore reference. All values were normalized to the control (*w^{1118}*) mean fluorescence intensity, which was set to 100% (n ≥ 7 kinetochores from at least 10 neuroblasts for each condition). Results are expressed as mean values ± SDE and statistical significance was assessed by the Mann-Whitney test.

Taking this novel cell cycle role of Heph into consideration, we asked if mitotic progression proceeded correctly in *heph^{2}/TM6B* neuroblasts. Most neuroblasts in the *heph^{2}/TM6B* mutant contain a normal chromosome content (8) as in the wild type (*w^{1118}*), but they do present more aneuploidy events (**FIGURE 42**) as we had previously observed for *gfp-poloΔUSE;polo$^{9-/-}$* neuroblasts. This indicates that the absence of Heph has a physiological impact on correct cell cycle progression in a similar manner as the absence of *polo* USE.

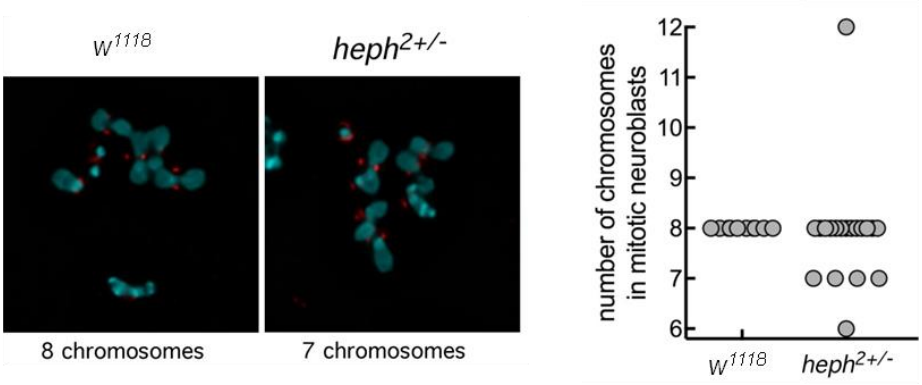


FIGURE 42 | Representative immunofluorescence images of mitotic neuroblasts from squashed *Drosophila melanogaster* larval brains with the *w*^{1118} and *heph*^{2}*/TM6B* genotypes. Spc105 was used as a kinetochore reference. Chromosome content is shown for each representative neuroblast. The graph represents the quantification of chromosome numbers in *w*^{1118} and *heph*^{2}*/TM6B* mitotic neuroblasts.

3. The effect of Polo activity on alternative polyadenylation and genome-wide transcription

Polo kinase levels can determine *polo* APA [115]. There is preferential usage of *polo* pA1 signal when Polo is overexpressed, which implies that high levels of Polo modulate *polo* expression through an auto-regulatory feedback loop that promotes *polo* pA1 choice and the production of the short *polo* isoform. We thus asked if Polo could have a new function in APA and/or transcription of other genes *in vivo.* To study this, we used two *polo* mutant fly strains, the kinase dead *polo*^{1} mutant that has a single point mutation and normal Polo protein levels [396, 411] and the null *polo*^{9} [410] mutant, and analysed APA and the expression levels of several genes.

3.1. Gene expression is altered in *polo* mutants

Initially, we analysed APA profiles and mRNA expression levels of genes with a similar genomic structure of *polo*: *polo*, *CG6024*, *lace* and *abdominal-b* (*abd-b*) all have five exons and two pA signals.

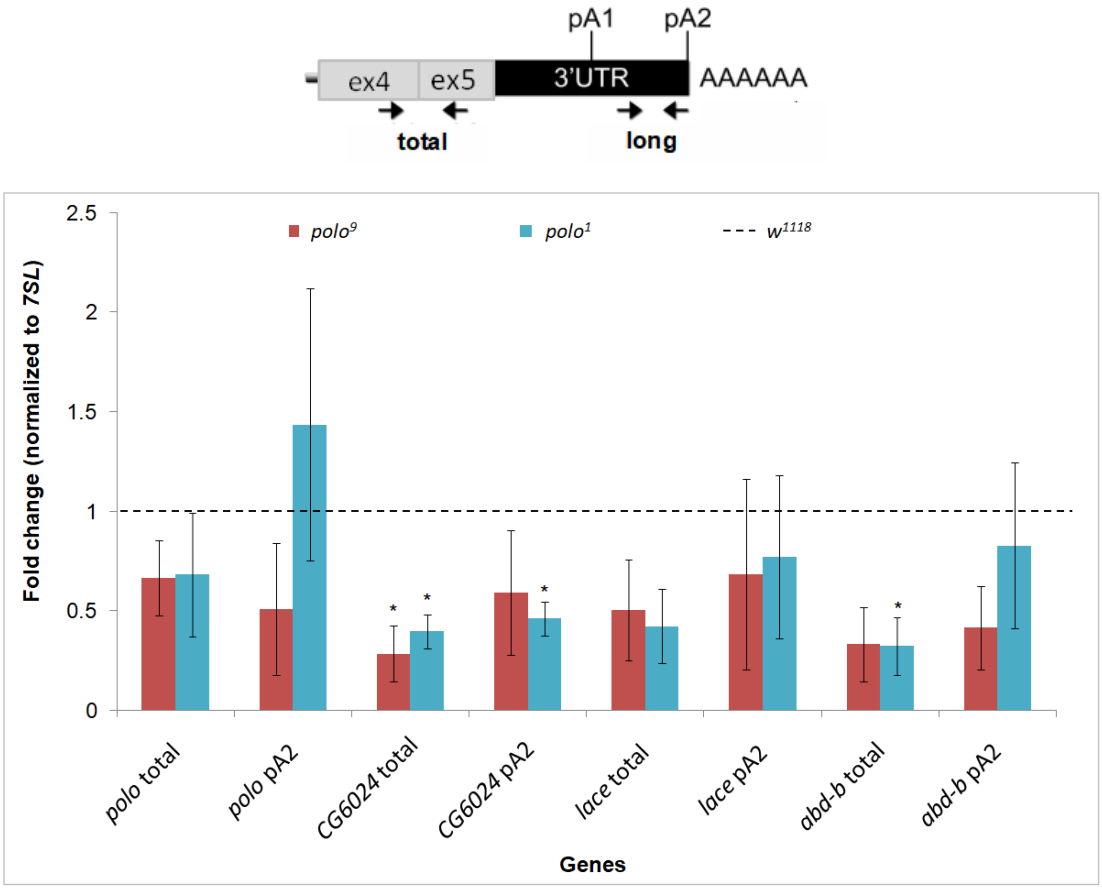


FIGURE 43 | Expression levels of the mRNA isoforms of *polo*, *CG6024*, *lace* and *abd-b* genes in wild type, heterozygous *polo*^{9} (null mutant, red bars) and heterozygous *polo*^{1} (kinase dead mutant, blue bars) male adults measured by RT-qPCR. Results

are the means ± SDE of three independent experiments and statistical significance was tested with a two-tailed paired Student's *t*-test. * indicates $p \leq 0.05$. *7SL* was used as a housekeeping gene.

Our results suggest that Polo levels and/or activity do not affect APA in most of the genes, but have a clear effect in *CG6024* and *abd-b* expression (**FIGURE 43**). The *CG6024* gene function is still unknown and its expression is moderate [412, 554, 576]. Curiously, the total levels of *CG6024* mRNAs are reduced by over 70% in the *polo* null mutant and over 60% in the *polo* kinase dead mutant. However, levels of the *CG6024* longest isoform are only statistically diminished to approximately half in the kinase dead mutant (**FIGURE 43**). This suggests that the expression of the shorter *CG6024* isoform is reduced in the absence of Polo while the long isoform levels are reduced only when Polo activity is faulty.

We observe no differences in the expression between wild type and either *polo* mutant regarding the long isoform of *abd-b*, but total expression levels are reduced by nearly 70% in the *polo* kinase dead mutant (**FIGURE 43**), indicating that the production of the shorter isoform of this gene is particularly diminished in *polo* mutants. As one of the three *hox* genes of the bithorax complex, *abd-b* has functions in adult development, including the posterior abdomen and genitalia [577-582].

Although the P-element insertion located in the *polo* promoter in the null mutant ($polo^{9}$) should ablate most of *polo* mRNA expression [410] in this mutant, we found that *polo* mRNA levels are not statistically different from the wild type control (**FIGURE 43**). This may be explained by the nature of the samples (heterozygous adult males), in which the wild type allele may compensate the loss of function of the mutated one.

Taken together, these results suggest that Polo levels and/or activity do not affect APA in general, although a more extensive study involving a larger number of genes should be performed.

All genes analysed so far are transcribed by RNAPII. To further explore a potential regulatory role of Polo kinase in transcription, we were also interested in studying the expression levels of genes transcribed by other RNAPs and whether they are altered in *polo* mutants. We selected three highly expressed genes to analyse: RNAPI-transcribed *18s*, RNAPII-transcribed *rp49* and RNAPIII-transcribed *U6* (**FIGURE 44**).

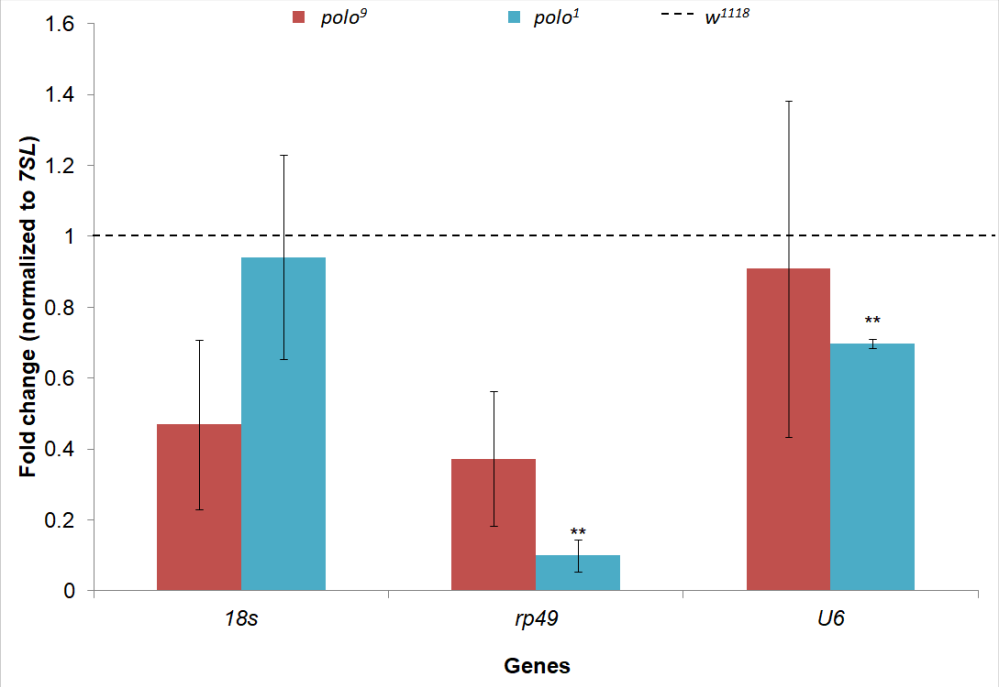


FIGURE 44 | mRNA expression levels of the *18s*, *rp49* and *U6* genes in wild type, *polo*^{9} (null, red bars) and *polo*^{1} (kinase dead, blue bars) mutants measured by RT-qPCR. Results are the means ± SDE of three independent experiments and statistical significance was tested with a two-tailed paired Student's *t*-test. ** indicates $p \leq 0.01$. *7SL* was used as a housekeeping gene.

We observed that *rp49* expression is remarkably downregulated by 90% while *U6* expression is diminished by 30% in the kinase dead mutant when compared to the wild type control (**FIGURE 44**).

Taken together, these results suggest that Polo is not a general regulator of APA or gene expression in *Drosophila melanogaster*, but its activity does affect the expression of several RNAPII-transcribed genes.

3.2. *polo* null mutant affects RNA Polymerase II Carboxy Terminal Domain phosphorylation pattern

A slower RNAPII transcription elongation rate alters *polo* APA, promoting the choice of *polo* pA1. Additionally, RNAPII chromatin occupancy profile is altered along the *polo* gene in *RpII215*^{C4} mutants [115]. Having shown that *polo* mutants affect the expression of *CG6024*, *abd-b* and *rp49*, we then asked if Polo activity affects that of RNAPII. For that, we analysed the RNAPII CTD phosphorylation patterns in *polo* kinase dead and null mutant embryos by western blot using antibodies that specifically detect two RNAPII CTD

modifications: Ser-5-P and Ser-7-P. These two CTD modifications are present during transcription initiation and the transition to promoter proximal pausing [45, 47-52].

As it can be seen, the RNAPII CTD phosphorylation pattern for the heterozygous *polo* null mutant (*polo*$^{9+/-}$) is clearly different from the control with the RNAPII hypophosphorylated (hypo-P RNAPII, **FIGURE 45**) isoform bereft of Ser-5-P and Ser-7-P being predominant and not primed for elongation. This suggests that transcription initiation/productive elongation in the null mutant may be compromised together with the developmental abnormalities related to a deficient cell cycle progression inherent to this null mutant [410]. These results suggest that the absence of Polo causes a suboptimal phosphorylation of the RNAPII CTD tail. Considering that Polo is a Ser/Thr kinase, it is possible that Polo is involved in RNAPII CTD phosphorylation.

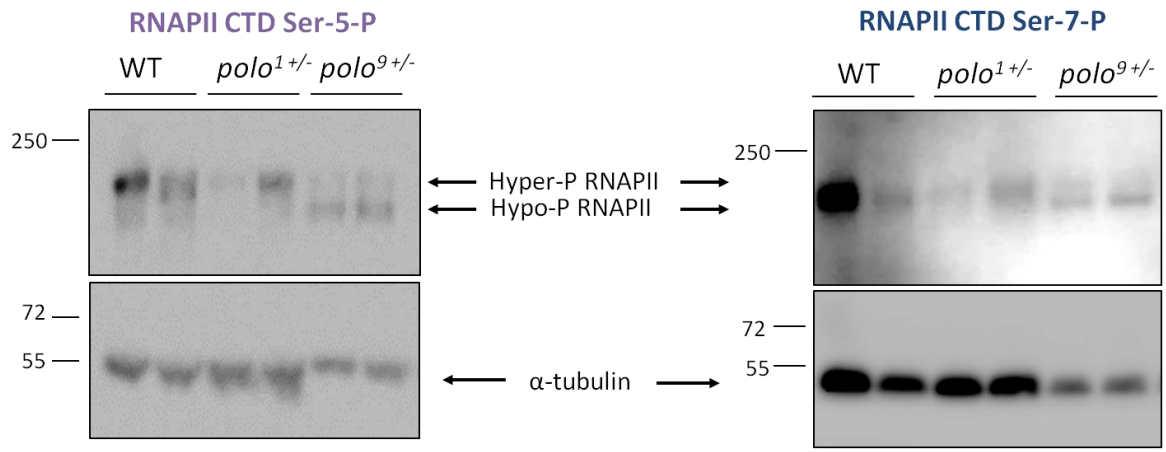


FIGURE 45 | The null mutant shows more hypophosphorylated RNAPII in comparison to wild type. Western blot using 0-24 h *Drosophila melanogaster* embryo protein extracts from *w*^{1118}, *polo*^{1} (kinase dead) and *polo*^{9} (null) strains against RNAPII CTD Ser-5-P (4H8, left panel) and RNAPII CTD Ser-7-P (right panel). α-tubulin served as loading control.

Taken together, these data show preliminary evidence that *Drosophila melanogaster* Polo kinase may affect RNAPII activity, which may have physiological repercussions at several levels during transcription and consequently, in gene expression control.

GENERAL DISCUSSION

In humans, 70% of all genes undergo APA [215, 371, 372]. Through this evolutionarily conserved [370] mechanism, the cell can produce mRNA isoforms that have different coding sequences and/or 3'UTRs, which not only contributes to transcriptome diversity, but also adds another level of gene expression control.

In the fly genome, we have determined that APA occurs in approximately 78% of genes using 3'READS [280]. This observation had been previously noted in fly, but in a less broad APA study using polo(A)$^+$ RNA-seq [279] whereas 3'READS is capable of mapping the characteristically A-rich pA sites [546, 547] that commonly go undetected using other, less specific sequencing methods. Indeed, we were able to properly re-annotate the 3'ends of over 10000 genes using our new data in comparison to the previous study. We have also found that there is a clear preference for *Drosophila melanogaster* wild type heads to choose dPASs in detriment of more pPASs. This finding is in agreement with genome-wide studies in other organisms, which have shown that there is a general prevalence of terminally differentiated cells to produce mRNA transcripts with long 3'UTRs, thus indicating that dPASs are preferentially selected in neural tissues [369, 372, 379, 380, 382]. We have observed these head-body differences in males. It would be equally interesting to perform this study in females or use different developmental stages since there are known RBPs that have germline-dependent roles such as Sex-lethal [329] and Heph [324, 388] that could therefore have a role in APA in highly specialized tissues such as ovaries and development-dependent 3'UTRs have been characterized [216].

The molecular mechanisms of how or why the cell chooses one pA signal over another are still unclear. To investigate the mechanisms behind APA in the fly head and body, we analysed the expression levels of genes encoding proteins involved in mRNA processing as modulating the levels of these proteins strongly correlates with pA site selection [173, 287, 313, 329, 338, 339, 393]. We found that many of these genes, such as *Pcf11* (termination factor that is part of the CFIIm complex), *CstF64* (part of the CSTF complex and a basal mRNA 3'end processing protein), *Cdk9* (responsible for priming RNAPII for productive elongation [26, 28, 84]) and *Ssu72* (RNAPII CTD Ser-5 and Ser-7-phosphatase [23, 99-101]), are upregulated in the wild type fly head while a select few (the transcription elongation factor elongin A – *EloA* [583] - and *MED20*, a subunit of the Mediator complex, for example) are downregulated in comparison to the body. Taken together, these results suggest that there is a different type of APA regulation in highly differentiated cells in the fly head in comparison to more proliferative cells in the fly body. Although further experiments must be performed to confirm this, we can hypothesize that

a more rigorous transcriptional control via upregulation of certain mRNA processing proteins in *Drosophila melanogaster* head tissues is needed whereas body tissues do not appear to present such strict control. This hypothesis is corroborated by several studies that show that the expression deregulation of genes encoding mRNA processing proteins may lead to further transcriptional deregulation of other genes as well as developmental abnormalities. Loss of Heph affects actin [321], Gurken [322] and Oskar [327] regulation, the splicing events of several genes [323, 326] and spermatogenesis [324, 388]. Deregulation of *su(f)*, which is the fly ortholog of CSTF77 [193], impairs neural differentiation and mitosis [196]. In zebrafish, PCF11 attenuates its own expression and that of other genes encoding transcription regulators [584].

polo APA is deregulated in the slower *RpII215*^{C4} mutants, with adult male flies preferably using the proximal *polo* pA1 3.5 times more efficiently than the wild type control concomitant with differences in RNAPII chromatin occupancy along *polo* [115]. APA of a small number of genes similarly structured to *polo* is equally defective in the slower *RpII215*^{C4} mutant fly strain [115]: *abd-b, lace, stlk, CG6024* and *cyclin D* all have five exons and two functional pA signals in their 3'UTRs as does *polo* and there is a preferential choice of their proximal pA signal in the presence of the slower RNAPII of the *RpII215*^{C4} mutant fly. Beyond APA, it has also been shown that a slower elongation rate causes premature transcription termination [117], deregulates mRNA 3'end formation of histones [116] and alters numerous splicing events [118, 119], including the resplicing of the *ultrabithorax Hox* gene in *RpII215*^{C4} flies [120]. To further understand if a slower transcription elongation rate affected APA at a genome-wide scale, we used 3'READS and found that while *RpII215*^{C4} flies present a higher preference for pPAS selection in the body, this trend is not observed in the head. This indicates that the elongation rate-dependent regulation is markedly distinct and impacts APA differently in heads and bodies. Our results indicate that neuronal cells are unresponsive to a 50% slower transcription elongation rate and that body tissues are more reactive to a slower transcription elongation rate, tending to adhere by the 'first come, first served' model of APA regulation in *Drosophila melanogaster* [115]. A recent study has mimicked the *RpII215*^{C4} R741H mutation using mouse embryonic stem cells [585] and found that a slower transcriptional elongation rate hinders neural differentiation via compromised gene expression and alternative splicing in long genes involved in synapse signalling and causes embryonic lethality in mouse. The authors also observed that undifferentiated embryonic cells were more impervious to the change in transcription elongation rate [585]. Together with our results, this new data indicates that the impact of a slower elongation rate varies according to the developmental stage and species analysed. Thus, this should

be considered when analysing different genome-wide datasets obtained from different tissues, cells and/or species.

The distinct head/body response to the slower RNAPII elongation rate may be due to the differential expression of genes encoding for RBPs, proteins involved in mRNA 3'end formation, and transcription elongation and termination factors in the mutant fly strain. We found that there is an increase of over four-fold in the expression of *Rrp6* and *Ssu72* in the *RpII215*^{C4} fly bodies and also a remarkable downregulation of over 15-fold of *Nelf-E* expression in the mutant fly. The downregulation of the negative elongation factor *Nelf-E* may be a compensatory mechanism to the inherent slower transcription rate in the mutant fly. In the mutant bodies, the marked upregulation of the RNAPII CTD Ser-5 and Ser-7 phosphatase *Ssu72*, which is also a component of the yeast mRNA 3'end machinery [100], and downregulation of the negative elongation factor *Nelf-E*, together with the mild upregulation of the *Spt6* elongation factor, may explain the preferential usage of pPAS observed in *RpII215*^{C4} bodies. Accordingly, it has been shown that genome-wide 3'UTR length inversely correlates with upregulation of mRNA 3'end processing factors in mammalian cells [586] and *Drosophila melanogaster* [192], namely core cleavage and pA factors such as CFIm. Our results thus suggest that *Ssu72 and Spt6* upregulation and downregulation of *Nelf-E* in *RpII215*^{C4} bodies lead to the usage of less efficient pPAS in *Drosophila melanogaster*.

It has been reported that the slower *RpII215*^{C4} mutant flies show altered RNAPII chromatin occupancy profile for the *polo* gene [115]. Although the structural differences in the slower RNAPII may explain this observation, it is also possible that an incorrect recruitment or function of transcription elongation and/or termination factors such as the NELF complex [74, 587] or Ssu72 [23, 588], whose expression is respectively down- and upregulated in *RpII215*^{C4} mutant bodies, are responsible for the defective RNAPII occupancy observed.

Polo kinase is crucial for correct cell cycle progression [397] and *polo* expression is known to be tightly regulated by APA and essential for correct fly development [115]. Our next goal was to identify the regulators of *polo* APA. Genome-wide studies predict that a significant percentage of human pA signals are flanked by U-rich sequence elements both upstream (USE) and downstream (DSE) [227]. USEs are known to be U-rich *cis* auxiliary elements localized upstream of pA signals [227] that help to modulate the mRNA 3'end formation efficiency of the genes in which they are found [184, 249, 256, 260, 267]. We have identified a particularly well-conserved sequence in *polo* 3'UTR with over 53% of U content which we named *polo* USE due to the similarities of this element to previously

described USEs in human [236, 249-257, 263] and viral [237-245, 247, 248] genes. This is the first description and characterization of a USE in *Drosophila melanogaster*, which indicates that this class of *cis* auxiliary sequences is widespread among different species and hints at a possible functionality.

We have also found that approximately 5% of all fly 3'UTRs contain the eight most conserved nt of *polo* USE, TTGTTTTT, upstream of the AATAAA, ATTAAA or AATATA pA signals. For comparison, the USE in the human *prothrombin F2* gene, TATTTTTGT, is present in the 3'UTR of 4% of human transcripts and in close proximity to AATAAA or ATTAAA [236], which highlights the relevance of the USE in the *Drosophila melanogaster* genome.

polo, *mirror*, *still-life* and *sex-lethal* are four examples of USE-containing genes that clearly show that this *cis* regulatory sequence can be found in the 3'UTR of vital genes with relevant roles in fly and neural development, gametogenesis and mRNA 3'end processing, in agreement with the functions of human USE-containing genes [236, 249, 255]. *polo* is an essential gene and contains the USE upstream of an ATTAAA. *mirror* USE is upstream of the AATAAA, and has important functions in eye formation [589], oogenesis [590] and peripheral nervous system development [591]. *still-life* encodes a G nt exchange factor for Rho-GTPases that specifically localizes to presynaptic terminals [592] and contains the USE upstream of the ATTAAA. *sex-lethal*, which encodes an RBP specific to the female germline involved in *enhancer of rudimentary* APA [329], contains the USE upstream of an AATATA.

We found that it is three times more likely to find the USE associated to non-canonical pA signals (ATTAAA or AATATA) than to the AATAAA pA signal. These signals tend to be less efficiently processed than the canonical pA signal [213], thus making them likely to be regulated [252, 260, 376] by the USE. Moreover, we revealed that USEs tend to be close to pA signals regardless of their efficiency, which is also in agreement with the distance-dependent effect already reported for USEs [236, 255]. Finally, we also observed that genes with the USE-N (TTNTTTTT) in their 3'UTRs tend to be more transcriptionally active than genes without this element in a developmental stage-dependent manner. Taken together, our results support the hypothesis that the USE is an auxiliary sequence of the pA signal, influencing cleavage and pA efficiency and acting as a widespread regulator of gene expression in *Drosophila melanogaster*.

To study the function of *polo* USE, we generated transgenic flies without *polo* USE. *gfp-poloΔUSE;polo*$^{9-/-}$ adult flies show a predominant phenotype characterized by the malformation of the abdominal epidermis identical to the one found in flies with a mutated

and non-functional *polo* pA1, the *gfp-poloΔpA1;polo*$^{9-/-}$ strain (ATTAAA > GTTAAC) [115]. Interestingly and in contrast to *gfp-poloΔpA1;polo*$^{9-/-}$ flies, *gfp-poloΔUSE;polo*$^{9-/-}$ females are sterile and all of them display the phenotype. These characteristics are reminiscent of the female sterility found in several *polo* mutant strains [396, 568], and the abdominal defects found in *polo*^{1}/*polo*^{2}, *polo*^{7} and *polo*^{8}.

The identical phenotype between *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ adults suggests that flies without *polo* USE also have incorrect *polo* APA, which we confirmed by mapping *polo* mRNA 3'ends in the two strains. Both *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ flies show erroneous *polo* pA signal choice, opting to select two cryptic and particularly inefficient pA signals together with *polo* pA1 and pA2 and revealing *polo* APA deregulation. This is in agreement with the literature: if a pA signal is mutated or deleted, there is an activation of cryptic pA signals located in the vicinity [318]. It has also been described that relevant *cis* regulatory sequences (such as *polo* USE) retain pA activity even in the absence of a functional pA signal [260], which is probably why there is still detectable *gfp-polo* pA1 mRNA in *gfp-poloΔpA1;polo*$^{9-/-}$ adult flies by 3'RACE. Additionally, USEs are known to stabilize the pA machinery on inefficient and non-canonical pA signals (both *polo* pA1 and pA2 are non-canonical), which has an impact on pA and mRNA 3'end formation efficiency [184, 249, 256, 260, 267]. Interestingly, USE mutations result in a significant reduction in cleavage and pA of the *COX-2* AUUAAA signal [255], which corresponds to the *polo* pA1 sequence. This may explain the activation of cryptic pA signals and low levels of *polo* after *in vivo* deletion of *polo* USE. The low *polo* mRNA levels and abnormal polo mRNA 3'end formation in *gfp-poloΔUSE;polo*$^{9-/-}$ flies clearly show that the USE has a physiological function in *polo* APA related to fly viability and normal development and highlight the importance of selecting the two *polo* pA signals – AUUAAA and AAUAUA – to accurately regulate *polo* expression.

Both *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ flies present an abdominal phenotype characterized by tergite malformation and misregulated *polo* APA while the few *gfp-poloΔpA2;polo*$^{9-/-}$ escapers show an aggravated abdominal phenotype. This phenotype is due to a deficit in Polo protein, required for the rapid proliferation of the abdominal histoblasts that occurs when the organism enters metamorphosis [593-597]. Interestingly, the slower *RpII215*^{C4} mutant fly strain does not have an abdominal phenotype even though the slower RNAPII affects *polo* APA; instead, they present a much more complex phenotype related to the mutation of the *ultrabithorax Hox* gene [598] in which the capitellum of the haltere of heterozygous individuals partially transforms into the wing blade [114, 599]. A possible explanation is the wider number of genes affected by a slower RNAPII.

Polo kinase has several key functions during the different phases of the cell cycle [423]. It is needed at various locations throughout the cell as mitosis progresses [397, 411, 424, 427, 428], particularly at the kinetochores during metaphase [425, 426] where it phosphorylates numerous targets that contribute to proper mitotic progress such as Aurora B and Mps1 kinases [430, 569, 571, 572]. We discovered that loss of the USE not only has a direct impact on *polo* at the transcriptional level, but this consequently compromises Polo levels and kinase activity at the kinetochores during metaphase, as well as the activation of the Aurora B and Mps1 kinases. Although *gfp-poloΔUSE;polo*$^{9-/-}$ individuals are capable of producing the longest *gfp-polo* mRNA, which is the main isoform responsible for effective GFP-Polo protein production that allows organism viability [115], the suboptimal accumulation of key kinases (GFP-Polo, Aurora B and Mps1) at the kinetochores indicates that *gfp-poloΔUSE;polo*$^{9-/-}$ neuroblasts present a deficit in the normal function of GFP-Polo protein that eventually hinders mitotic fidelity and leads to aneuploidies, which we also observed in the neuroblasts of these flies. Our results are in agreement with the abnormalities in mitosis and meiosis described for several *polo* mutants: *polo*^{1}/*polo*^{2}, *polo*^{7}, *polo*^{8} and *polo*^{9} [396, 568].

It is possible that *polo* USE is particularly needed to increase Polo kinase levels at the kinetochores of dividing cells, namely in highly proliferative developmental stages such as the embryo, third instar larvae brains and metamorphosis. The peak of *polo* mRNA detection is during the 0-4h embryonic stage [397], which is in agreement with the need for Polo protein during the first rapid cell divisions. At this stage, most of *polo* mRNAs are inherited from ovaries, which are enriched in the maternal longest *polo* isoform [412], before zygotically-dependent transcription begins [600]. In the early pupae stage, histoblasts become highly proliferative [593-597] and are particularly sensitive to Polo levels, requiring the longest mRNA isoform of *polo* to form the adult epidermis [115]. If *gfp-polo* pA2 signal is used, but its selection is faulty (as seen for both the *gfp-poloΔUSE;polo*$^{9-/-}$ and *gfp-poloΔpA1;polo*$^{9-/-}$ strains), individuals show abdominal defects.

It is relevant to highlight the fact that the *gfp-poloΔUSE;polo*$^{9-/-}$ transgenic strain contains a deleted and not mutated USE, two very different modifications to *polo* 3'UTR that may or may not have similar repercussions. This deletion may disrupt possible secondary structures present in both *polo* isoforms that may lead to different mRNA stabilities, functions and obstruct potential RBP interactions. This deletion may also hypothetically place other putative *cis* regulatory elements in a different position and/or conformation, thus altering their potential influence on *polo* pA signal selection. Regardless, the USE is only comprised by 28 nt, a rather small region considering the 2.3 and 2.5 kbp sizes of the shorter and longer *polo* mRNA isoforms, respectively. As such, it is unlikely that its

deletion will have a relevant impact on the overall structure of the two mRNAs, but it could be interesting to generate a new *Drosophila melanogaster* strain with a mutated USE and validate our results with the *gfp-poloΔUSE;polo*$^{9-/-}$ transgenic strain.

USEs impact pA signal choice and pA efficiency by interacting with and recruiting numerous RBPs and core mRNA 3'end processing factors, either directly or indirectly [236, 249, 250, 256]. These *trans*-acting proteins are the true effectors of USE-dependent gene regulation. We identified Heph as the top RBP that specifically binds to the USE RNA. Heph is involved in *oskar* expression [327], Gurken protein location [322] and the splicing of Mlc1 and several other genes involved in spermatogenesis [326]. While no known role for Heph has been described regarding APA in *Drosophila melanogaster*, its human ortholog, PTBP1, affects pA signal efficiency and mRNA 3'end formation by recognizing the USEs of *COX-2, prothrombin F2* and *C2 complement* pre-mRNAs [236, 250, 256] and consequently, we next explored the role of Heph on *polo* gene expression control.

We showed that hypomorphic *heph*^{2}*/TM6B* mutants present abnormal *polo* mRNA 3'end formation and decreased production of the longest *polo* mRNA isoform. Concomitantly, we verified that Polo protein production was remarkably reduced in this mutant. It is possible that binding of Heph to the USE blocks the binding of cleavage and pA factors to *polo* pA1 in a similar manner as ELAVL1/Elav [307, 601], thus activating *polo* pA2 usage. Taken together, these results strongly suggest a new *in vivo* function for Heph in *Drosophila melanogaster* APA by promoting usage of *polo* distal pA signal, therefore enhancing Polo protein production.

Our data indicate that both the USE and Heph are necessary for correct *polo* mRNA 3'end formation and mRNA levels. However, while we did observe incorrect *polo* mRNA 3'end formation in both the *gfp-poloΔUSE;polo*$^{9-/-}$ and *heph*^{2}*/TM6B* strains, the mRNA species produced by these strains are different. The lack of *polo* USE and *polo* pA1 mutation lead to the production of two new *polo* mRNAs and the lack of Heph only generates one. This suggests that absence of the USE affects correct *polo* pA1 and pA2 usage *in vivo* while the absence of Heph mostly hinders *polo* distal pA signal selection. As suggested by our list of proteins that specifically bind to the USE RNA and not to the USEmt RNA, it is plausible that the USE has more interacting partners than just Heph. These proteins may probably contribute to the remaining USE regulatory functions linked to *polo* expression and may compensate for the lack of Heph and its respective roles.

If there are more RBPs acting upon the USE beyond Heph, and considering that the USE is comprised of 28 nt and each RBP can, on average, bind to approximately six nt at a

time, there is the possibility of a protein complex assembling on this *cis* element and modulating its function, hence *polo* expression. This is the case of the *COX-2* [256] and *prothrombin F2* genes [236], for which at least four (*COX-2*) or three (*prothrombin F2*) different RBPs have been identified to interact with their respective USEs, with PTBP1 common to both. Taking into consideration the similarities of these two USEs (*COX-2* USE is AUUUCUUA and *prothrombin F2* USE is UAUUUUUGU) to *polo* USE (UUGUUUUU), it is plausible that similar RBPs will recognize these similar sequences (like Heph/PTBP1, with CUCUCU and UCUU as their consensus binding sequences [602]). In the future, it would be interesting to test if the other proteins that specifically bind to the USE RNA influence *polo* expression and characterize this hypothetical protein network in search of rate-limiting or synergistically-acting elements as well as stoichiometry. For instance, the Tho2 protein (#unique peptides$_{\text{USE RNA}}$ = 13 and #unique peptides$_{\text{USEmt RNA}}$ = 4) has a crucial function in the nuclear export of some mRNAs, such as those encoding for heatshock proteins [574] and it is possible that it has also a role in *polo* mRNA export. The splicing factor Sf3b3 (#unique peptides$_{\text{USE RNA}}$ = 12 and #unique peptides$_{\text{USEmt RNA}}$ = 4) is part of a transcriptional co-activator complex [575] and is recruited to the prespliceosome [603], which may suggest further functions in other mRNA processing steps. Rbp9 (#unique peptides$_{\text{USE RNA}}$ = 11 and #unique peptides$_{\text{USEmt RNA}}$ = 2) is an RBP related to Elav [604] with a function in *Drosophila melanogaster* oogenesis via binding to the 3'UTR mRNA of *bag-of-marbles* and regulating its translation [604] and downregulation of *extramacrochaetae* mRNA [605] and would also be an interesting USE-binding protein to study.

Although we detected a total of 282 RBPs that specifically bind to the USE RNA and do not bind to the USEmt RNA, a potential technical drawback to the methodology employed is whether the USE needs a specific conformation to allow the binding of its RBPs *in vivo*. This could be circumvented by using the full length *polo* 3'UTR to allow possible secondary structures to form. We must also consider intrinsic *versus* biological specificities for these new potential USE RBPs: a promising positive hit *in vitro* may not specifically bind to the USE *in vivo* and *in vivo* USE-binding proteins may act less specifically *in vitro* [606]. Additionally, RBPs have variable expression levels [287, 338, 339] that may depend on different stimuli or cellular context [329, 332-334], which can consequently condition the binding ability of the RBP to the USE. Such is the case of CSTF64 levels, which modulate IgM APA differently in immature B cells (low CSTF64 concentration) and plasma cells (high CSTF64 concentration) [338, 393] and Sex-lethal binding to *enhancer of rudimentary* DSE that specifically occurs in the female germline [329]. As such, we might have only identified a fraction of putative *polo* USE interactors.

Some Polo-dependent biological pathways are similarly altered in *gfp-poloΔUSE;polo*$^{9-/-}$ and *heph*^{2}*/TM6B* individuals. Similarly to *gfp-poloΔUSE;polo*$^{9-/-}$, the *heph*^{2}*/TM6B* mutant shows reduced accumulation of Polo at the kinetochores of dividing cells. Additionally, the low phosphorylation levels of Aurora B and Mps1 kinases corroborate that the activity levels of these three key cell cycle kinases are equally hindered at the risk of compromising normal mitotic progression. Accordingly, both the *gfp-poloΔUSE;polo*$^{9-/-}$ and *heph*^{2}*/TM6B* strains show higher abundance of aneuploidies, with their chromosomes more prone to missegregate than expected. These results reveal a novel role for *polo* USE and Heph in the mitotic pathways controlled by Polo and cell cycle regulation at the kinetochores while also denoting that the absence of either the USE or Heph has a similar effect in these events.

Surprisingly, we found a reduction in RNAPII chromatin occupancy in the 3'UTR of USE-containing genes, including *polo*, that is not nearly as evident in genes without this element. It is tempting to speculate that the USE favours displacement of RNAPII from the chromatin of actively transcribed genes. We propose that during *polo* transcription, Heph may assemble with RNAPII along the *polo* gene. When RNAPII reaches the 3'end of the gene and disengages from the chromatin, Heph is recruited to the USE on the nascent pre-mRNA, modulating *polo* pA signal usage and promoting distal pA signal selection when more Polo protein is required for the cell (**FIGURE 46**).

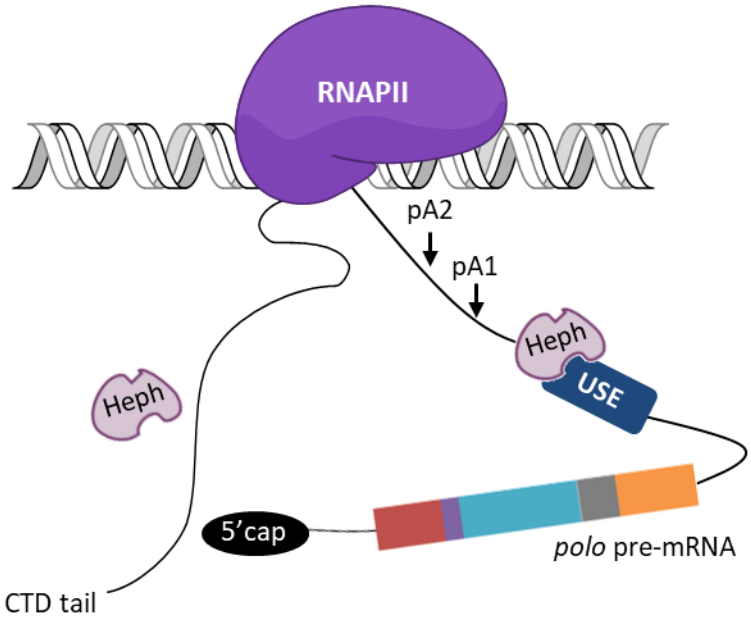


FIGURE 46 | Working model for the molecular mechanism of Heph-dependent mRNA 3'end formation of *polo* mediated by *polo* USE. At the 3'end of the *polo* gene, Heph disengages from RNAPII and moves to the *polo* pre-mRNA, binding to the USE as the

pre-mRNA is transcribed. Here, together with the basal cleavage and pA machinery, Heph is necessary to correctly select the pA signal.

Polo protein levels regulate *polo* APA, which is in turn also regulated by the USE [278] and the RNAPII elongation rate [115]. We therefore investigated the genome-wide effect of Polo activity on APA and transcription.

Neither the lack of Polo nor its activity was found to be a general regulator of APA in *Drosophila melanogaster*, but *abd-b* total expression levels are significantly reduced in the *polo* kinase dead mutant. *abd-b* is one of the three *hox* genes of the bithorax complex with a vital function in adult development, including the posterior abdomen, genitalia and gonads [577-582] despite always having low expression levels [412]. While there is no described physical interaction between *abd-b* and *polo*, we have shown that abdominal defects are present in flies that lack the USE of *polo* and also that females are sterile [278]. Additionally, homozygous *polo* kinase dead mutant females are also notoriously sterile [396], with Polo also being involved in spermatogenesis [607], meiosis and mitosis [396, 570, 608, 609]. Considering these similarities, it is tempting to speculate a possible interaction between *polo* and *abd-b*.

When the expression of genes transcribed by different RNAPs was analysed, the levels of *rp49* and *U6* were significantly decreased in the *polo* kinase dead mutant while the expression of *18s* remained unaffected by the absence of Polo or its activity. Interestingly, *rp49* is transcribed by RNAPII and *U6* by RNAPIII, two RNAPs previously reported to be regulated by PLK-1 [30, 525]. These new results seem to suggest that both RNAPII and RNAPIII activities may be also regulated by Polo in *Drosophila melanogaster*. On the other hand, RNAPI is not known to be regulated by PLK-1, which may further explain why *18s* expression is not affected by the lack of Polo or its activity. Together with the gene expression data obtained with *CG6024* and *abd-b*, this suggests that the expression of several RNAPII-transcribed genes is prone to be modulated by Polo activity.

RNAPII is the only RNAP with a CTD tail that affects its transcriptional activity [21-25]. To further explore a potential regulatory function of Polo on RNAPII activity, we analysed global RNAPII CTD phosphorylation patterns in the *polo* null and kinase dead mutant embryos. Polo-deficient individuals present an altered RNAPII CTD phosphorylation pattern in the form of a largely hypophosphorylated RNAPII bereft of Ser-5-P and Ser-7-P in comparison to the more hyperphosphorylated isoform of RNAPII found in the wild type strain.

There is a technical limitation associated with the signal strength of the RNAPII western blots: it is highly dependent on the number of accessible RNAPII CTD marks. This also means that the absence of an RNAPII CTD mark or isoform can indicate either its physical absence or its masking by other modifications [30]. This may hold true for the mostly undetected hyperphosphorylated RNAPII in the *polo* null extracts, but it is unlikely considering that we were able to successfully detect a hyperphosphorylated RNAPII in the wild type strain.

Hypophosphorylated RNAPII is commonly found at the pre-initiation complex during transcription initiation [131]. At the promoters, the RNAPII CTD Ser-5 residue is the first to be phosphorylated, followed by Ser-7 [45, 47-52]. The lack of these two modifications in the $polo^{9+/-}$ mutant suggests that transcription initiation and consequent co-transcriptional recruitment of the capping machinery may be hindered [52, 54, 55], which may have severe consequences in the transcription cycle of many genes. The developmental stage chosen for this experiment is also relevant. During embryonic development, there is a great need for a burst of transcription and for correct, reliable cell division as the organism grows. If Polo protein is deficient and consequently hinders proper phosphorylation of RNAPII, both mitosis and correct transcription are compromised in the *polo* null mutant.

Several modifications including Ser-2-P [49, 89, 98] and Thr-4-P [30, 127] occur as RNAPII becomes productively elongating, thus generating a hyperphosphorylated isoform. RNAPII CTD Thr-4-P is PLK-1-dependent and a hyperphosphorylated RNAPII with this modification is found tethered to locations where PLK-1 normally is in mitotic cells, such as centrosomes and the midbody [128]. If this is also true in *Drosophila melanogaster*, the hypophosphorylated RNAPII found in Polo-deficient individuals may also be due to low levels of Polo-dependent CTD Thr-4-P. The antibodies used in the western blots specifically detect RNAPII CTDs with low (hypophosphorylated) or high (hyperphosphorylated) levels Ser-5-P and Ser-7-P, but RNAPII CTD may also be phosphorylated/unphosphorylated in other residues, such as Thr-4. Taken together, our results suggest that *polo* null mutants present 1) atypical transcription initiation and capping due to low levels of RNAPII CTD Ser-5-P and Ser-7-P and 2) anomalous productive elongation, also because of low RNAPII CTD Ser-5-P and Ser-7-P and possibly from deficient RNAPII CTD Thr-4-P, which in turn also hinders mitosis [128]. Although we did not observe a significant reduction of mRNA levels in six out of seven genes in the $polo^{9+/-}$ mutant strain, our results suggest a new function for Polo in transcription and maintenance of proper RNAPII CTD phosphorylation patterns. It would be interesting to investigate how Polo affects gene expression at the genome-wide level.

Ssu72 phosphatase activity is downregulated by four-fold in the presence of an RNAPII CTD phosphorylated on both Ser-5 and Thr-4 [610]. It is tempting to speculate that, as the CTD Ser-5 and/or Thr-4 residues are not properly phosphorylated in the *polo* mutant, Ssu72 activity should be abnormally upregulated. In turn, Ssu72 could promote further dephosphorylation of RNAPII, which could ultimately leave RNAPII in a non-productive stage in the *polo* null mutant.

Throughout this thesis, I have shown that APA in *Drosophila melanogaster* relies on the concerted action of RNAPII kinetics in a tissue-dependent manner [280]. In the particular case of *polo*, it also relies on Heph binding to the USE, which controls Polo function at the kinetochores and proper cell cycle progression [278]. I have also shown that Polo alters the phosphorylation pattern of RNAPII CTD, suggesting a yet unknown regulatory role for Polo in transcription. In short, this work sheds new insight into the molecular mechanisms that occur during transcription and APA *in vivo*, such as the contribution of RNAPII kinetics, the first report of a USE in *Drosophila melanogaster* and a potential novel role for an old cell cycle kinase, Polo.

APPENDIX

TABLE 12 | List of the 282 RNA-related proteins obtained in LC-MS/MS that bind to the USE (#unique peptides ≥ 5) and do not bind to USEmt (#unique peptides < 5) sorted from largest to smallest #unique peptides identified/protein. The top 10 proteins identified are highlighted and included in **TABLE 11**.

Protein	#Unique peptides pulled down with USE RNA	#Unique peptides pulled down with USE_{mt} RNA
Heph	15	4
Emb	15	4
CG7728	14	0
Ars2	14	3
DCTN1-P150	13	4
Tho2	13	4
Su(w^a)	12	3
CG5728	12	4
Sf3b3	12	4
SpnE	11	3
Nonc	11	0
Piwi	11	1
Rangap	11	1
CG16916	11	2
Rbp9	11	2
CG11123	11	4
Mif2	11	4
Mle	11	4
Mrna-Cap	11	4
Sm	11	4
Pp2c	10	3
Asnrs	10	0
Caper	10	0
Cpsf160	10	0
Pea	10	0
Beta-Phers	10	2
Upf2	10	2
Armi	10	3
CG14230	10	3
CG9107	10	4
Mrpl13	10	4
Nup93-1	10	4
CG7483	9	0
Eif2beta	9	0
Nup98-96	9	0
Rpt4R	9	1
Spt5	9	1

Asprs-M	9	2
CG32533	9	2
CG8915	9	2
Nup153	9	2
Psc	9	2
Sym	9	2
CG11414	9	3
CG4901	9	3
CG6745	9	3
Nelf-A	9	3
Eif4g1	9	4
Eif4h2	9	4
Nito	9	4
Ythdc1	9	4
Dcr-1	8	0
Nxf2	8	0
Rae1	8	0
CG8833	8	1
Clbn	8	1
Erf3	8	1
Hfp	8	1
CG1582	8	2
Eip93F	8	2
Nop5	8	2
Rnp4F	8	2
Suv3	8	2
Waw	8	2
Apt	8	3
Asprs	8	3
CG2875	8	3
Lwr	8	3
Mbo	8	3
Rpi1	8	3
Scaf6	8	3
Serrs	8	3
Smb	8	3
Snama	8	3
Tdrd3	8	3
CG4038	8	4
CG4896	8	4
CG7246	8	4
Dcr-2	8	4
Eif3e	8	4
Fib-RA	8	4
Kra	8	4
L(2)K09022	8	4

Peng	8	4
Rin	8	4
Rpi135	8	4
Spn	8	4
CG6379	7	0
Ire1	7	0
Scm	7	0
Vav	7	0
CG1316	7	1
CG18596	7	1
Cysrs-M	7	1
Gtp-Bp	7	1
Atx2	7	2
CG10418	7	2
CG12259	7	2
CG17540	7	2
CG31441	7	2
CG5382	7	2
CG6712	7	2
Mrpl1	7	2
Mura	7	2
Nlg4	7	2
Phax	7	2
Rabex-5	7	2
Tra2	7	2
Xpac-RA	7	2
CG10214	7	3
CG12942	7	3
CG14207	7	3
CG3198	7	3
Eif5	7	3
L(1)G0007	7	3
Mxt	7	3
Nst	7	3
Rlua-2	7	3
Rpl34b	7	3
Rpt4	7	3
Bgcn	7	4
Cap	7	4
CG31156	7	4
Cn-IIIB	7	4
Hisrs	7	4
Noi	7	4
Pum	7	4
Rhogap18b	7	4
Srl	7	4

Tfap-2	7	4
CG1218	6	0
CG6951	6	0
Dhx15	6	0
Eif3d2	6	0
Fkbp13	6	0
Fus-RE	6	0
Her	6	0
Nelf-E	6	0
Nep3	6	0
Nona-L	6	0
Or85e	6	0
Poldip2	6	0
Sf3a1	6	0
U4-U6-60k	6	0
Ublcp1	6	0
V	6	0
CG10887	6	1
CG10909	6	1
CG2091	6	1
CG6163-Ra	6	1
CG7564	6	1
CG8349	6	1
Eif6	6	1
Hsc70-1	6	1
Ada2a	6	2
Arfgap1	6	2
CG1571	6	2
CG4787	6	2
CG5116	6	2
Dscam3	6	2
Hex-C	6	2
L(1)G0004	6	2
Or71a-RB	6	2
Ran-Like	6	2
Rnps1	6	2
Sec6	6	2
Sodh-1	6	2
Brv2	6	3
CG10384	6	3
CG11147	6	3
CG30403	6	3
CG5715	6	3
CG9518	6	3
Ct32316	6	3
Dnd	6	3

Gcn2	6	3
Glob3	6	3
Hp1b	6	3
Nop56	6	3
Rfc38	6	3
Trprs-M	6	3
Tsh	6	3
CG33099	6	4
Ct13580	6	4
Gcs2alpha	6	4
Grsm	6	4
Pgant35A-RA	6	4
Wrnexo	6	4
Yrt	6	4
Pka-Like	5	3
Anp32a	5	4
Ance-3	5	0
Apc	5	0
Bap60	5	0
CG13643	5	0
CG17493	5	0
CG18262	5	0
CG5756	5	0
Cn	5	0
Cyp303a1	5	0
Cyp4d14-RA	5	0
Cyp4d2	5	0
Drak	5	0
Eag	5	0
Eip78C	5	0
Gce	5	0
Gefmeso	5	0
Gwl	5	0
Hip1	5	0
Lar	5	0
Mrp	5	0
Mrps9	5	0
Nude	5	0
Or49b	5	0
P	5	0
Patj	5	0
Patr-1	5	0
Phkgamma	5	0
Rpn13R	5	0
Sak	5	0
Sdt	5	0

Snrnp-U1-70K	5	0
Spn88Eb	5	0
Swim	5	0
Ttll12	5	0
Wek	5	0
Bub3	5	1
CG10862	5	1
CG12129	5	1
CG2921	5	1
Daw	5	1
Gstd10	5	1
Ica69	5	1
Jhdm2	5	1
Orc2	5	1
Pcs	5	1
Rbp1-Like	5	1
Rhau	5	1
Serrs-M	5	1
Cdk4	5	2
CG10445	5	2
CG17309	5	2
CG44249	5	2
CG6994	5	2
CG9727	5	2
Desat2	5	2
Fj	5	2
Gs1l	5	2
Idgf6	5	2
Idh	5	2
Mei-P26	5	2
Pex12	5	2
Pnut	5	2
Rhobtb	5	2
Sba	5	2
Sec13	5	2
Srpk79D	5	2
Atbp	5	3
CG10077	5	3
CG11533	5	3
CG1814-Ra	5	3
CG4500	5	3
Ct31087	5	3
Eif3d1	5	3
Gro	5	3
Mav	5	3
Nurf-38	5	3

Or22b	5	3
Sk	5	3
Traf4	5	3
CG14322	5	4
Cip4	5	4
Ets21C	5	4
Hsc20	5	4
Kat-60L1	5	4
Lgs	5	4
Mttfb1	5	4
Nd-39	5	4
Rpd3	5	4
Spn28F	5	4
Tio	5	4
Top1	5	4

REFERENCES

1. Cramer, P., et al., *Structure of eukaryotic RNA polymerases.* Annu Rev Biophys, 2008. **37**: p. 337-52.
2. Rudra, D. and J.R. Warner, *What better measure than ribosome synthesis?* Genes Dev, 2004. **18**(20): p. 2431-6.
3. Werner, F. and D. Grohmann, *Evolution of multisubunit RNA polymerases in the three domains of life.* Nat Rev Microbiol, 2011. **9**(2): p. 85-98.
4. Edwards, A.M., et al., *Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro.* J Biol Chem, 1991. **266**(1): p. 71-5.
5. Zhang, G., et al., *Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution.* Cell, 1999. **98**(6): p. 811-24.
6. Vassylyev, D.G., et al., *Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution.* Nature, 2002. **417**(6890): p. 712-9.
7. Bushnell, D.A. and R.D. Kornberg, *Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6969-73.
8. Runner, V.M., V. Podolny, and S. Buratowski, *The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors.* Mol Cell Biol, 2008. **28**(6): p. 1883-91.
9. Ovchinnikov Yu, A., et al., *The primary structure of E. coli RNA polymerase, Nucleotide sequence of the rpoC gene and amino acid sequence of the beta'-subunit.* Nucleic Acids Res, 1982. **10**(13): p. 4035-44.
10. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD.* Trends Genet, 2008. **24**(6): p. 289-96.
11. Stump, A.D. and K. Ostrozhynska, *Selective constraint and the evolution of the RNA polymerase II C-Terminal Domain.* Transcription, 2013. **4**(2): p. 77-86.
12. Yang, C. and J.W. Stiller, *Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. 5920-5.
13. 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.
14. West, M.L. and J.L. Corden, *Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations.* Genetics, 1995. **140**(4): p. 1223-33.

15. Brickey, W.J. and A.L. Greenleaf, *Functional studies of the carboxy-terminal repeat domain of Drosophila RNA polymerase II in vivo.* Genetics, 1995. **140**(2): p. 599-613.
16. Bartolomei, M.S., et al., *Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II.* Mol Cell Biol, 1988. **8**(1): p. 330-9.
17. Zehring, W.A., et al., *The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro.* Proc Natl Acad Sci U S A, 1988. **85**(11): p. 3698-702.
18. Nonet, M., D. Sweetser, and R.A. Young, *Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II.* Cell, 1987. **50**(6): p. 909-15.
19. Lu, F., B. Portz, and D.S. Gilmour, *The C-Terminal Domain of RNA Polymerase II Is a Multivalent Targeting Sequence that Supports Drosophila Development with Only Consensus Heptads.* Molecular Cell, 2019.
20. Corden, J.L., *RNA polymerase II C-terminal domain: Tethering transcription to transcript and template.* Chem Rev, 2013. **113**(11): p. 8423-55.
21. Egloff, S. and S. Murphy, *Cracking the RNA polymerase II CTD code.* Trends Genet, 2008. **24**(6): p. 280-8.
22. Buratowski, S., *Progression through the RNA polymerase II CTD cycle.* Mol Cell, 2009. **36**(4): p. 541-6.
23. Eick, D. and M. Geyer, *The RNA polymerase II carboxy-terminal domain (CTD) code.* Chem Rev, 2013. **113**(11): p. 8456-90.
24. Buratowski, S., *The CTD code.* Nat Struct Biol, 2003. **10**(9): p. 679-80.
25. Corden, J.L., *Transcription. Seven ups the code.* Science, 2007. **318**(5857): p. 1735-6.
26. Peterlin, B.M. and D.H. Price, *Controlling the elongation phase of transcription with P-TEFβ.* Mol Cell, 2006. **23**(3): p. 297-305.
27. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7.* Science, 2007. **318**(5857): p. 1780-2.
28. Bres, V., S.M. Yoh, and K.A. Jones, *The multi-tasking P-TEFβ complex.* Curr Opin Cell Biol, 2008. **20**(3): p. 334-40.
29. Czudnochowski, N., C.A. Bosken, and M. Geyer, *Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFβ recognition.* Nat Commun, 2012. **3**: p. 842.

30. Hintermair, C., et al., *Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation.* EMBO J, 2012. **31**(12): p. 2784-97.
31. Hsin, J.P., et al., *RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells.* Elife, 2014. **3**: p. e02112.
32. Heidemann, M. and D. Eick, *Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code.* RNA Biol, 2012. **9**(9): p. 1144-6.
33. Ranuncolo, S.M., et al., *Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo.* J Biol Chem, 2012. **287**(28): p. 23549-61.
34. Sims, R.J., 3rd, et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation.* Science, 2011. **332**(6025): p. 99-103.
35. Dias, J.D., et al., *Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells.* Elife, 2015. **4**.
36. Li, H., et al., *Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells.* Mol Cell Biol, 2007. **27**(15): p. 5296-305.
37. Schroder, S., et al., *Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells.* Mol Cell, 2013. **52**(3): p. 314-24.
38. Portz, B., et al., *Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain.* Nat Commun, 2017. **8**: p. 15231.
39. Yonezawa, Y., *Molecular dynamics study of the phosphorylation effect on the conformational states of the C-terminal domain of RNA polymerase II.* J Phys Chem B, 2014. **118**(17): p. 4471-8.
40. Lenz, P. and P.S. Swain, *An entropic mechanism to generate highly cooperative and specific binding from protein phosphorylations.* Curr Biol, 2006. **16**(21): p. 2150-5.
41. Schuller, R., et al., *Heptad-Specific Phosphorylation of RNA Polymerase II CTD.* Mol Cell, 2016. **61**(2): p. 305-14.
42. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
43. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.
44. Grunberg, S. and S. Hahn, *Structural insights into transcription initiation by RNA polymerase II.* Trends Biochem Sci, 2013. **38**(12): p. 603-11.

45. Boeing, S., et al., *RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion.* J Biol Chem, 2010. **285**(1): p. 188-96.
46. Singh, G., et al., *The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion.* Annu Rev Biochem, 2015. **84**: p. 325-54.
47. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II.* Nature, 1992. **358**(6388): p. 641-5.
48. Akoulitchev, S., et al., *Requirement for TFIIH kinase activity in transcription by RNA polymerase II.* Nature, 1995. **377**(6549): p. 557-60.
49. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes Dev, 2000. **14**(19): p. 2452-60.
50. 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.
51. Imasaki, T., et al., *Architecture of the Mediator head module.* Nature, 2011. **475**(7355): p. 240-3.
52. Ebmeier, C.C., et al., *Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications.* Cell Rep, 2017. **20**(5): p. 1173-1186.
53. Max, T., M. Søgaard, and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator.* Journal of Biological Chemistry, 2007. **282**(19): p. 14113-14120.
54. Ho, C.K. and S. Shuman, *Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme.* Mol Cell, 1999. **3**(3): p. 405-11.
55. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II.* Genes Dev, 1997. **11**(24): p. 3306-18.
56. Rasmussen, E.B. and J.T. Lis, *In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7923-7.
57. Nechaev, S., et al., *Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila.* Science, 2010. **327**(5963): p. 335-8.
58. Baskaran, R., M.E. Dahmus, and J.Y. Wang, *Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain.* Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11167-71.

59. 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.
60. Roeder, R.G., *Transcriptional regulation and the role of diverse coactivators in animal cells.* FEBS Lett, 2005. **579**(4): p. 909-15.
61. Marshall, N.F. and D.H. Price, *Control of formation of two distinct classes of RNA polymerase II elongation complexes.* Mol Cell Biol, 1992. **12**(5): p. 2078-90.
62. Muse, G.W., et al., *RNA polymerase is poised for activation across the genome.* Nat Genet, 2007. **39**(12): p. 1507-11.
63. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.
64. Zeitlinger, J., et al., *RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo.* Nat Genet, 2007. **39**(12): p. 1512-6.
65. Gilmour, D.S., *Promoter proximal pausing on genes in metazoans.* Chromosoma, 2009. **118**(1): p. 1-10.
66. Gilmour, D.S. and J.T. Lis, *RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells.* Mol Cell Biol, 1986. **6**(11): p. 3984-9.
67. Rougvie, A.E. and J.T. Lis, *The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged.* Cell, 1988. **54**(6): p. 795-804.
68. Krumm, A., et al., *The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region.* Genes Dev, 1992. **6**(11): p. 2201-13.
69. Guo, J. and D.H. Price, *RNA polymerase II transcription elongation control.* Chem Rev, 2013. **113**(11): p. 8583-603.
70. 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.
71. 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.
72. Andrulis, E.D., et al., *High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation.* Genes Dev, 2000. **14**(20): p. 2635-49.
73. Yamaguchi, Y., et al., *Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA.* Mol Cell Biol, 2002. **22**(9): p. 2918-27.

74. Wu, C.H., et al., *NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila.* Genes Dev, 2003. **17**(11): p. 1402-14.
75. Missra, A. and D.S. Gilmour, *Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex.* Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11301-6.
76. Vos, S.M., et al., *Structure of paused transcription complex Pol II-DSIF-NELF.* Nature, 2018. **560**(7720): p. 601-606.
77. Rahl, P.B., et al., *c-Myc regulates transcriptional pause release.* Cell, 2010. **141**(3): p. 432-45.
78. 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.
79. Yamaguchi, Y., et al., *Structure and function of the human transcription elongation factor DSIF.* J Biol Chem, 1999. **274**(12): p. 8085-92.
80. Renner, D.B., et al., *A highly purified RNA polymerase II elongation control system.* J Biol Chem, 2001. **276**(45): p. 42601-9.
81. Palangat, M., et al., *A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS.* Proc Natl Acad Sci U S A, 2005. **102**(42): p. 15036-41.
82. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.
83. Galbraith, M.D., A.J. Donner, and J.M. Espinosa, *CDK8: a positive regulator of transcription.* Transcription, 2010. **1**(1): p. 4-12.
84. Marshall, N.F., et al., *Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase.* J Biol Chem, 1996. **271**(43): p. 27176-83.
85. Yamada, T., et al., *P-TEFβ-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation.* Mol Cell, 2006. **21**(2): p. 227-37.
86. Mandal, S.S., et al., *Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II.* Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7572-7.
87. Cheung, A.C., S. Sainsbury, and P. Cramer, *Structural basis of initial RNA polymerase II transcription.* EMBO J, 2011. **30**(23): p. 4755-63.
88. Hirose, Y. and J.L. Manley, *RNA polymerase II is an essential mRNA polyadenylation factor.* Nature, 1998. **395**(6697): p. 93-6.
89. 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.

90. Ahn, S.H., M. Kim, and S. Buratowski, *Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing.* Mol Cell, 2004. **13**(1): p. 67-76.
91. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 272-8.
92. Bentley, D.L., *Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors.* Curr Opin Cell Biol, 2005. **17**(3): p. 251-6.
93. 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.
94. Kwak, H. and J.T. Lis, *Control of transcriptional elongation.* Annu Rev Genet, 2013. **47**: p. 483-508.
95. 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.
96. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it.* Genes Dev, 2004. **18**(20): p. 2437-68.
97. Gilchrist, D.A., et al., *Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation.* Cell, 2010. **143**(4): p. 540-51.
98. Meinhart, A. and P. Cramer, *Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors.* Nature, 2004. **430**(6996): p. 223-226.
99. Krishnamurthy, S., et al., *Ssu72 Is an RNA polymerase II CTD phosphatase.* Mol Cell, 2004. **14**(3): p. 387-94.
100. Rosado-Lugo, J.D. and M. Hampsey, *The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition.* J Biol Chem, 2014. **289**(49): p. 33916-26.
101. Zhang, D.W., et al., *Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination.* J Biol Chem, 2012. **287**(11): p. 8541-51.
102. Yeo, M., et al., *A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5.* J Biol Chem, 2003. **278**(28): p. 26078-85.
103. Egloff, S., et al., *Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes.* Mol Cell, 2012. **45**(1): p. 111-22.
104. Peng, J., et al., *Identification of multiple cyclin subunits of human P-TEFβ.* Genes Dev, 1998. **12**(5): p. 755-62.
105. Bartkowiak, B. and A.L. Greenleaf, *Phosphorylation of RNAPII: To P-TEFβ or not to P-TEFβ?* Transcription, 2011. **2**(3): p. 115-119.

106. 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.
107. 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.
108. Devaiah, B.N., et al., *BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain.* Proc Natl Acad Sci U S A, 2012. **109**(18): p. 6927-32.
109. Itzen, F., et al., *Brd4 activates P-TEFβ for RNA polymerase II CTD phosphorylation.* Nucleic Acids Res, 2014. **42**(12): p. 7577-90.
110. Lu, H., et al., *Compensatory induction of MYC expression by sustained CDK9 inhibition via a BRD4-dependent mechanism.* Elife, 2015. **4**: p. e06535.
111. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. **5**: p. 3505.
112. Coulter, D.E. and A.L. Greenleaf, *Properties of mutationally altered RNA polymerases II of Drosophila.* J Biol Chem, 1982. **257**(4): p. 1945-52.
113. Coulter, D.E. and A.L. Greenleaf, *A mutation in the largest subunit of RNA polymerase II alters RNA chain elongation in vitro.* J Biol Chem, 1985. **260**(24): p. 13190-8.
114. Chen, Y., et al., *Mapping mutations in genes encoding the two large subunits of Drosophila RNA polymerase II defines domains essential for basic transcription functions and for proper expression of developmental genes.* Mol Cell Biol, 1993. **13**(7): p. 4214-22.
115. Pinto, P.A., et al., *RNA polymerase II kinetics in polo polyadenylation signal selection.* EMBO J, 2011. **30**(12): p. 2431-44.
116. 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.
117. Fong, N., et al., *Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition.* Mol Cell, 2015. **60**(2): p. 256-67.
118. Fong, N., et al., *Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate.* Genes Dev, 2014. **28**(23): p. 2663-76.
119. Honkela, A., et al., *Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays.* Proc Natl Acad Sci U S A, 2015. **112**(42): p. 13115-20.

120. de la Mata, M., et al., *A slow RNA polymerase II affects alternative splicing in vivo.* Mol Cell, 2003. **12**(2): p. 525-32.
121. Chen, Y., et al., *Drosophila RNA polymerase II mutants that affect transcription elongation.* J Biol Chem, 1996. **271**(11): p. 5993-9.
122. Izban, M.G. and D.S. Luse, *Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing.* Genes Dev, 1991. **5**(4): p. 683-96.
123. Izban, M.G. and D.S. Luse, *Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates.* J Biol Chem, 1992. **267**(19): p. 13647-55.
124. Kireeva, M.L., et al., *Nature of the nucleosomal barrier to RNA polymerase II.* Mol Cell, 2005. **18**(1): p. 97-108.
125. Guermah, M., et al., *Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates.* Cell, 2006. **125**(2): p. 275-86.
126. Sheridan, R.M., et al., *Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate.* Mol Cell, 2019. **73**(1): p. 107-118 e4.
127. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II.* Science, 2012. **336**(6089): p. 1723-5.
128. Hintermair, C., et al., *Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression.* Sci Rep, 2016. **6**: p. 27401.
129. Lin, P.S., M.F. Dubois, and M.E. Dahmus, *TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II.* J Biol Chem, 2002. **277**(48): p. 45949-56.
130. Hsin, J.P., K. Xiang, and J.L. Manley, *Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing.* Mol Cell Biol, 2014. **34**(13): p. 2488-98.
131. Brookes, E. and A. Pombo, *Modifications of RNA polymerase II are pivotal in regulating gene expression states.* EMBO Rep, 2009. **10**(11): p. 1213-9.
132. Logan, J., et al., *A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8306-10.
133. Osheim, Y.N., N.J. Proudfoot, and A.L. Beyer, *EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage.* Mol Cell, 1999. **3**(3): p. 379-87.

134. Connelly, S. and J.L. Manley, *A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II.* Genes Dev, 1988. **2**(4): p. 440-52.
135. West, S., N. Gromak, and N.J. Proudfoot, *Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites.* Nature, 2004. **432**(7016): p. 522-5.
136. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II.* Nature, 2004. **432**(7016): p. 517-22.
137. 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.
138. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model.* Genes Dev, 2006. **20**(8): p. 954-65.
139. West, S. and N.J. Proudfoot, *Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination.* Nucleic Acids Res, 2008. **36**(3): p. 905-14.
140. Rosonina, E., S. Kaneko, and J.L. Manley, *Terminating the transcript: breaking up is hard to do.* Genes Dev, 2006. **20**(9): p. 1050-6.
141. Richard, P. and J.L. Manley, *Transcription termination by nuclear RNA polymerases.* Genes Dev, 2009. **23**(11): p. 1247-69.
142. Loya, T.J. and D. Reines, *Recent advances in understanding transcription termination by RNA polymerase II.* F1000Res, 2016. **5**.
143. Zhang, Z. and D.S. Gilmour, *Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript.* Mol Cell, 2006. **21**(1): p. 65-74.
144. Proudfoot, N.J., *Ending the message: poly(A) signals then and now.* Genes Dev, 2011. **25**(17): p. 1770-82.
145. Kuehner, J.N., E.L. Pearson, and C. Moore, *Unravelling the means to an end: RNA polymerase II transcription termination.* Nat Rev Mol Cell Biol, 2011. **12**(5): p. 283-94.
146. Laitem, C., et al., *CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes.* Nat Struct Mol Biol, 2015. **22**(5): p. 396-403.
147. Proudfoot, N.J., *How RNA polymerase II terminates transcription in higher eukaryotes.* Trends Biochem Sci, 1989. **14**(3): p. 105-10.

148. Brannan, K., et al., *mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription.* Mol Cell, 2012. **46**(3): p. 311-24.
149. Gromak, N., S. West, and N.J. Proudfoot, *Pause sites promote transcriptional termination of mammalian RNA polymerase II.* Mol Cell Biol, 2006. **26**(10): p. 3986-96.
150. Core, L.J. and J.T. Lis, *Transcription regulation through promoter-proximal pausing of RNA polymerase II.* Science, 2008. **319**(5871): p. 1791-2.
151. Glover-Cutter, K., et al., *RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes.* Nat Struct Mol Biol, 2008. **15**(1): p. 71-8.
152. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
153. 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.
154. Grosso, A.R., et al., *Dynamic transitions in RNA polymerase II density profiles during transcription termination.* Genome Res, 2012. **22**(8): p. 1447-56.
155. Ji, Z., et al., *Transcriptional activity regulates alternative cleavage and polyadenylation.* Mol Syst Biol, 2011. **7**: p. 534.
156. 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.
157. Birse, C.E., et al., *Coupling termination of transcription to messenger RNA maturation in yeast.* Science, 1998. **280**(5361): p. 298-301.
158. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.
159. Buratowski, S., *Connections between mRNA 3' end processing and transcription termination.* Curr Opin Cell Biol, 2005. **17**(3): p. 257-61.
160. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code.* Nat Struct Mol Biol, 2010. **17**(10): p. 1279-86.
161. Nojima, T., et al., *Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing.* Cell, 2015. **161**(3): p. 526-540.
162. Enriquez-Harris, P., et al., *A pause site for RNA polymerase II is associated with termination of transcription.* EMBO J, 1991. **10**(7): p. 1833-42.

163. Plant, K.E., et al., *Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene.* Mol Cell Biol, 2005. **25**(8): p. 3276-85.
164. Zhang, H., F. Rigo, and H.G. Martinson, *Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site.* Mol Cell, 2015. **59**(3): p. 437-48.
165. Barabino, S.M. and W. Keller, *Last but not least: regulated poly(A) tail formation.* Cell, 1999. **99**(1): p. 9-11.
166. Moore, M.J. and N.J. Proudfoot, *Pre-mRNA processing reaches back to transcription and ahead to translation.* Cell, 2009. **136**(4): p. 688-700.
167. Maniatis, T. and R. Reed, *An extensive network of coupling among gene expression machines.* Nature, 2002. **416**(6880): p. 499-506.
168. Perales, R. and D. Bentley, *"Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions.* Mol Cell, 2009. **36**(2): p. 178-91.
169. Wang, Y., J.A. Fairley, and S.G. Roberts, *Phosphorylation of TFIIB links transcription initiation and termination.* Curr Biol, 2010. **20**(6): p. 548-53.
170. Calvo, O. and J.L. Manley, *Strange bedfellows: polyadenylation factors at the promoter.* Genes Dev, 2003. **17**(11): p. 1321-7.
171. Dantonel, J.C., et al., *Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA.* Nature, 1997. **389**(6649): p. 399-402.
172. Mapendano, C.K., et al., *Crosstalk between mRNA 3' end processing and transcription initiation.* Mol Cell, 2010. **40**(3): p. 410-22.
173. Venkataraman, K., K.M. Brown, and G.M. Gilmartin, *Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition.* Genes Dev, 2005. **19**(11): p. 1315-27.
174. 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.
175. Fong, N. and D.L. Bentley, *Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD.* Genes Dev, 2001. **15**(14): p. 1783-95.
176. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.* Nature, 1997. **385**(6614): p. 357-61.
177. Cramer, P., et al., *Functional association between promoter structure and transcript alternative splicing.* Proc Natl Acad Sci U S A, 1997. **94**(21): p. 11456-60.

178. Ardehali, M.B., et al., *Spt6 enhances the elongation rate of RNA polymerase II in vivo.* EMBO J, 2009. **28**(8): p. 1067-77.
179. Nagaike, T., et al., *Transcriptional activators enhance polyadenylation of mRNA precursors.* Mol Cell, 2011. **41**(4): p. 409-18.
180. Hollerer, I., et al., *mRNA 3'end processing: A tale of the tail reaches the clinic.* EMBO Mol Med, 2014. **6**(1): p. 16-26.
181. Shi, Y., et al., *Molecular architecture of the human pre-mRNA 3' processing complex.* Mol Cell, 2009. **33**(3): p. 365-76.
182. West, S. and N.J. Proudfoot, *Transcriptional termination enhances protein expression in human cells.* Mol Cell, 2009. **33**(3): p. 354-64.
183. Proudfoot, N.J., *Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut.* Science, 2016. **352**(6291): p. aad9926.
184. Zhao, J., L. Hyman, and C. Moore, *Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 405-45.
185. Danckwardt, S., M.W. Hentze, and A.E. Kulozik, *3' end mRNA processing: molecular mechanisms and implications for health and disease.* EMBO J, 2008. **27**(3): p. 482-98.
186. Edmonds, M., *A history of poly A sequences: from formation to factors to function.* Prog Nucleic Acid Res Mol Biol, 2002. **71**: p. 285-389.
187. Mandel, C.R., Y. Bai, and L. Tong, *Protein factors in pre-mRNA 3'-end processing.* Cell Mol Life Sci, 2008. **65**(7-8): p. 1099-122.
188. Millevoi, S. and S. Vagner, *Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation.* Nucleic Acids Res, 2010. **38**(9): p. 2757-74.
189. Di Giammartino, D.C., et al., *RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs.* Genes Dev, 2014. **28**(20): p. 2248-60.
190. Mount, S.M. and H.K. Salz, *Pre-messenger RNA processing factors in the Drosophila genome.* J Cell Biol, 2000. **150**(2): p. F37-44.
191. Darmon, S.K. and C.S. Lutz, *mRNA 3' end processing factors: a phylogenetic comparison.* Comp Funct Genomics, 2012. **2012**: p. 876893.
192. Vallejos Baier, R., J. Picao-Osorio, and C.R. Alonso, *Molecular Regulation of Alternative Polyadenylation (APA) within the Drosophila Nervous System.* J Mol Biol, 2017. **429**(21): p. 3290-3300.
193. Takagaki, Y. and J.L. Manley, *A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein.* Nature, 1994. **372**(6505): p. 471-4.

194. Bai, C. and P.P. Tolias, *Drosophila clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters.* Nucleic Acids Research, 1998. **26**(7): p. 1597-1604.
195. Salinas, C.A., et al., *Characterization of a Drosophila homologue of the 160-kDa subunit of the cleavage and polyadenylation specificity factor CPSF.* Mol Gen Genet, 1998. **257**(6): p. 672-80.
196. Audibert, A. and M. Simonelig, *The suppressor of forked gene of Drosophila, which encodes a homologue of human CstF-77K involved in mRNA 3'-end processing, is required for progression through mitosis.* Mech Dev, 1999. **82**(1-2): p. 41-50.
197. Benoit, B., et al., *The Drosophila poly(A)-binding protein II is ubiquitous throughout Drosophila development and has the same function in mRNA polyadenylation as its bovine homolog in vitro.* Nucleic Acids Res, 1999. **27**(19): p. 3771-8.
198. Murata, T., et al., *The hiiragi gene encodes a poly(A) polymerase, which controls the formation of the wing margin in Drosophila melanogaster.* Dev Biol, 2001. **233**(1): p. 137-47.
199. Benoit, B., et al., *Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10593-8.
200. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes & Development, 1997. **11**(21): p. 2755-2766.
201. Tian, B. and J.L. Manley, *Alternative polyadenylation of mRNA precursors.* Nat Rev Mol Cell Biol, 2017. **18**(1): p. 18-30.
202. Takagaki, Y., L.C. Ryner, and J.L. Manley, *Four factors are required for 3'-end cleavage of pre-mRNAs.* Genes Dev, 1989. **3**(11): p. 1711-24.
203. Bienroth, S., et al., *Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors.* J Biol Chem, 1991. **266**(29): p. 19768-76.
204. Keller, W., et al., *Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA.* EMBO J, 1991. **10**(13): p. 4241-9.
205. MacDonald, C.C., J. Wilusz, and T. Shenk, *The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location.* Mol Cell Biol, 1994. **14**(10): p. 6647-54.
206. 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.

207. Takagaki, Y. and J.L. Manley, *RNA recognition by the human polyadenylation factor CstF.* Mol Cell Biol, 1997. **17**(7): p. 3907-14.
208. Alberts, B., *Molecular biology of the cell*. 4th ed. 2002, New York: Garland Science. xxxiv, 1548 p.
209. Matoulkova, E., et al., *The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells.* RNA Biol, 2012. **9**(5): p. 563-76.
210. Wickens, M. and P. Stephenson, *Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation.* Science, 1984. **226**(4678): p. 1045-51.
211. Wilusz, J., S.M. Pettine, and T. Shenk, *Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal.* Nucleic Acids Res, 1989. **17**(10): p. 3899-908.
212. Sheets, M.D., S.C. Ogg, and M.P. Wickens, *Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro.* Nucleic Acids Res, 1990. **18**(19): p. 5799-805.
213. Wickens, M., *How the messenger got its tail: addition of poly(A) in the nucleus.* Trends Biochem Sci, 1990. **15**(7): p. 277-81.
214. Retelska, D., et al., *Similarities and differences of polyadenylation signals in human and fly.* BMC Genomics, 2006. **7**: p. 176.
215. 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.
216. Sanfilippo, P., J. Wen, and E.C. Lai, *Landscape and evolution of tissue-specific alternative polyadenylation across Drosophila species.* Genome Biol, 2017. **18**(1): p. 229.
217. Edwalds-Gilbert, G., K.L. Veraldi, and C. Milcarek, *Alternative poly(A) site selection in complex transcription units: means to an end?* Nucleic acids research, 1997. **25**(13): p. 2547-2561.
218. Gilmartin, G.M., et al., *CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition.* Genes Dev, 1995. **9**(1): p. 72-83.
219. Graveley, B.R., E.S. Fleming, and G.M. Gilmartin, *RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor.* Mol Cell Biol, 1996. **16**(9): p. 4942-51.
220. Proudfoot , N.J., *Ending the message: poly(A) signals then and now.* Genes Dev, 2011. **25**(17): p. 1770-82.

221. Tian, B. and J.H. Graber, *Signals for pre-mRNA cleavage and polyadenylation.* Wiley Interdiscip Rev RNA, 2012. **3**(3): p. 385-96.
222. Shi, Y. and J.L. Manley, *The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site.* Genes Dev, 2015. **29**(9): p. 889-97.
223. Hu, J., et al., *Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation.* RNA, 2005. **11**(10): p. 1485-93.
224. Vlasova, I.A., et al., *Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1.* Mol Cell, 2008. **29**(2): p. 263-70.
225. Gruber, A.J., et al., *A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation.* Genome Res, 2016. **26**(8): p. 1145-59.
226. Plass, M., S.H. Rasmussen, and A. Krogh, *Highly accessible AU-rich regions in 3' untranslated regions are hotspots for binding of regulatory factors.* PLoS Comput Biol, 2017. **13**(4): p. e1005460.
227. Legendre, M. and D. Gautheret, *Sequence determinants in human polyadenylation site selection.* BMC Genomics, 2003. **4**(1): p. 7.
228. Xie, X., et al., *Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals.* Nature, 2005. **434**(7031): p. 338-45.
229. Ghosh, T., et al., *MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic {beta}-actin gene.* Nucleic Acids Res, 2008. **36**(19): p. 6318-32.
230. Hogan, D.J., et al., *Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system.* PLoS Biol, 2008. **6**(10): p. e255.
231. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.* Cell, 2012. **149**(6): p. 1393-406.
232. Graber, J.H., et al., *In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 14055-60.
233. Neve, J., et al., *Cleavage and polyadenylation: Ending the message expands gene regulation.* RNA Biol, 2017. **14**(7): p. 865-890.
234. Di Giammartino, D.C., K. Nishida, and J.L. Manley, *Mechanisms and consequences of alternative polyadenylation.* Mol Cell, 2011. **43**(6): p. 853-66.
235. Yang, Q. and S. Doublie, *Structural biology of poly(A) site definition.* Wiley Interdiscip Rev RNA, 2011. **2**(5): p. 732-47.

236. Danckwardt, S., et al., *Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals.* EMBO J, 2007. **26**(11): p. 2658-69.
237. Carswell, S. and J.C. Alwine, *Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences.* Mol Cell Biol, 1989. **9**(10): p. 4248-58.
238. Schek, N., C. Cooke, and J.C. Alwine, *Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses.* Molecular and Cellular Biology, 1992. **12**(12): p. 5386-5393.
239. DeZazzo, J.D. and M.J. Imperiale, *Sequences upstream of the AAUAAA influence poly(A) site selection in a complex transcription unit.* Mol Cell Biol, 1989. **9**: p. 4951-4961.
240. Prescott, J.C. and E. Falck-Pedersen, *Varied poly(A) site efficiency in the adenovirus major late transcription unit.* J Biol Chem, 1992. **267**(12): p. 8175-81.
241. Prescott, J. and E. Falck-Pedersen, *Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites.* Mol Cell Biol, 1994. **14**(7): p. 4682-93.
242. Sittler, A., H. Gallinaro, and M. Jacob, *Upstream and downstream cis-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2.* Nucleic Acids Res, 1994. **22**(2): p. 222-31.
243. Russnak, R. and D. Ganem, *Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses.* Genes & Development, 1990. **4**(5): p. 764-776.
244. Russnak, R.H., *Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive.* Nucleic Acids Research, 1991. **19**(23): p. 6449-6456.
245. Brown, P.H., L.S. Tiley, and B.R. Cullen, *Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences.* Journal of Virology, 1991. **65**(6): p. 3340-3343.
246. DeZazzo, J.D., J.E. Kilpatrick, and M.J. Imperiale, *Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation.* Mol Cell Biol, 1991. **11**(3): p. 1624-30.
247. Cherrington, J. and D. Ganem, *Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences.* EMBO J, 1992. **11**(4): p. 1513-24.
248. Valsamakis, A., N. Schek, and J.C. Alwine, *Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for*

efficient polyadenylation in vitro. Molecular and Cellular Biology, 1992. **12**(9): p. 3699-3705.

249. Moreira, A., et al., *Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved.* EMBO J, 1995. **14**(15): p. 3809-19.
250. Moreira, A., et al., *The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms.* Genes Dev, 1998. **12**(16): p. 2522-34.
251. Brackenridge, S. and N.J. Proudfoot, *Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal.* Mol Cell Biol, 2000. **20**(8): p. 2660-9.
252. Natalizio, B.J., et al., *Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals.* J Biol Chem, 2002. **277**(45): p. 42733-40.
253. Wang, L., S. Tanaka, and F. Ramirez, *GATA-4 binds to an upstream element of the human alpha2(I) collagen gene (COL1A2) and inhibits transcription in fibroblasts.* Matrix Biol, 2005. **24**(5): p. 333-40.
254. Sully, G., et al., *Structural and functional dissection of a conserved destabilizing element of cyclo-oxygenase-2 mRNA: evidence against the involvement of AUF-1 [AU-rich element/poly(U)-binding/degradation factor-1], AUF-2, tristetraprolin, HuR (Hu antigen R) or FBP1 (far-upstream-sequence-element-binding protein 1).* Biochem J, 2004. **377**(Pt 3): p. 629-39.
255. Hall-Pogar, T., et al., *Alternative polyadenylation of cyclooxygenase-2.* Nucleic Acids Res, 2005. **33**(8): p. 2565-79.
256. Hall-Pogar, T., et al., *Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR.* RNA, 2007. **13**(7): p. 1103-15.
257. Sachchithananthan, M., et al., *The relationship between the prothrombin upstream sequence element and the G20210A polymorphism: the influence of a competitive environment for mRNA 3'-end formation.* Nucleic Acids Res, 2005. **33**(3): p. 1010-20.
258. Newnham, C.M., et al., *Alternative polyadenylation of MeCP2: Influence of cis-acting elements and trans-acting factors.* RNA Biol, 2010. **7**(3): p. 361-72.
259. Nedeljkovic, M., et al., *Long-distance regulation of Add2 pre-mRNA3'end processing.* RNA Biol, 2013. **10**(4): p. 516-27.
260. Phillips, C. and A. Virtanen, *The murine IgM secretory poly(A) site contains dual upstream and downstream elements which affect polyadenylation.* Nucleic Acids Res, 1997. **25**(12): p. 2344-51.

261. Aissouni, Y., et al., *The cleavage/polyadenylation activity triggered by a U-rich motif sequence is differently required depending on the poly(A) site location at either the first or last 3'-terminal exon of the 2'-5' oligo(A) synthetase gene.* J Biol Chem, 2002. **277**(39): p. 35808-14.
262. Millevoi, S., et al., *A physical and functional link between splicing factors promotes pre-mRNA 3' end processing.* Nucleic Acids Res, 2009. **37**(14): p. 4672-83.
263. Nunes, N.M., et al., *A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence.* EMBO J, 2010. **29**(9): p. 1523-36.
264. 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.
265. Darmon, S.K. and C.S. Lutz, *Novel upstream and downstream sequence elements contribute to polyadenylation efficiency.* RNA Biol, 2012. **9**(10): p. 1255-65.
266. Danckwardt, S., et al., *p38 MAPK controls prothrombin expression by regulated RNA 3' end processing.* Mol Cell, 2011. **41**(3): p. 298-310.
267. Lutz, C.S., et al., *Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro.* Genes Dev, 1996. **10**(3): p. 325-37.
268. Millevoi, S., et al., *An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries.* EMBO J, 2006. **25**(20): p. 4854-64.
269. Martinson, H.G., *An active role for splicing in 3'-end formation.* Wiley Interdiscip Rev RNA, 2011. **2**(4): p. 459-70.
270. Misra, A. and M.R. Green, *From polyadenylation to splicing: Dual role for mRNA 3' end formation factors.* RNA Biol, 2016. **13**(3): p. 259-64.
271. Zarudnaya, M.I., et al., *Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures.* Nucleic Acids Res, 2003. **31**(5): p. 1375-86.
272. Chen, F., C.C. MacDonald, and J. Wilusz, *Cleavage site determinants in the mammalian polyadenylation signal.* Nucleic Acids Res, 1995. **23**(14): p. 2614-20.
273. Chen, F. and J. Wilusz, *Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs.* Nucleic Acids Res, 1998. **26**(12): p. 2891-8.
274. Gil, A. and N.J. Proudfoot, *Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation.* Cell, 1987. **49**(3): p. 399-406.
275. Bagga, P.S., et al., *The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA*

processing through a trans-acting factor. Nucleic acids research, 1995. **23**(9): p. 1625-1631.

276. Arhin, G.K., et al., *Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals.* Nucleic acids research, 2002. **30**(8): p. 1842-1850.
277. Dalziel, M., N.M. Nunes, and A. Furger, *Two G-Rich Regulatory Elements Located Adjacent to and 440 Nucleotides Downstream of the Core Poly(A) Site of the Intronless Melanocortin Receptor 1 Gene Are Critical for Efficient 3' End Processing.* Molecular and Cellular Biology, 2007. **27**(5): p. 1568-1580.
278. Oliveira, M., et al., *The cell cycle kinase Polo is controlled by a conserved 3'UTR regulatory sequence in Drosophila melanogaster*. 2019.
279. Smibert, P., et al., *Global patterns of tissue-specific alternative polyadenylation in Drosophila.* Cell Rep, 2012. **1**(3): p. 277-89.
280. Liu, X., et al., *Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster.* RNA, 2017. **23**(12): p. 1807-1816.
281. Mateo, L., A. Ullastres, and J. Gonzalez, *A transposable element insertion confers xenobiotic resistance in Drosophila.* PLoS Genet, 2014. **10**(8): p. e1004560.
282. Taliaferro, J.M., et al., *RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation.* Mol Cell, 2016. **64**(2): p. 294-306.
283. Baltz, A.G., et al., *The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.* Mol Cell, 2012. **46**(5): p. 674-90.
284. Wahle, E. and U. Ruegsegger, *3'-End processing of pre-mRNA in eukaryotes.* FEMS Microbiol Rev, 1999. **23**(3): p. 277-95.
285. Tian, B. and J.L. Manley, *Alternative cleavage and polyadenylation: the long and short of it.* Trends Biochem Sci, 2013. **38**(6): p. 312-20.
286. Xiang, K., L. Tong, and J.L. Manley, *Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery.* Mol Cell Biol, 2014. **34**(11): p. 1894-910.
287. Brown, K.M. and G.M. Gilmartin, *A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im.* Mol Cell, 2003. **12**(6): p. 1467-76.
288. de Vries, H., et al., *Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors.* EMBO J, 2000. **19**(21): p. 5895-904.
289. Murthy, K.G. and J.L. Manley, *Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus.* J Biol Chem, 1992. **267**(21): p. 14804-11.

290. Chan, S.L., et al., *CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing.* Genes Dev, 2014. **28**(21): p. 2370-80.
291. Schonemann, L., et al., *Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33.* Genes Dev, 2014. **28**(21): p. 2381-93.
292. Murthy, K.G. and J.L. Manley, *The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation.* Genes Dev, 1995. **9**(21): p. 2672-83.
293. Wahle, E., *Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors.* J Biol Chem, 1991. **266**(5): p. 3131-9.
294. Bienroth, S., W. Keller, and E. Wahle, *Assembly of a processive messenger RNA polyadenylation complex.* EMBO J, 1993. **12**(2): p. 585-94.
295. Takagaki, Y., L.C. Ryner, and J.L. Manley, *Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation.* Cell, 1988. **52**(5): p. 731-42.
296. Ruegsegger, U., D. Blank, and W. Keller, *Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits.* Mol Cell, 1998. **1**(2): p. 243-53.
297. Ryan, K., O. Calvo, and J.L. Manley, *Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease.* RNA, 2004. **10**(4): p. 565-73.
298. Mandel, C.R., et al., *Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.* Nature, 2006. **444**(7121): p. 953-6.
299. Takagaki, Y. and J.L. Manley, *Complex protein interactions within the human polyadenylation machinery identify a novel component.* Mol Cell Biol, 2000. **20**(5): p. 1515-25.
300. Brennan, C.M. and J.A. Steitz, *HuR and mRNA stability.* Cell Mol Life Sci, 2001. **58**(2): p. 266-77.
301. Hinman, M.N. and H. Lou, *Diverse molecular functions of Hu proteins.* Cell Mol Life Sci, 2008. **65**(20): p. 3168-81.
302. Dai, W., G. Zhang, and E.V. Makeyev, *RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage.* Nucleic Acids Res, 2012. **40**(2): p. 787-800.
303. Diaz-Munoz, M.D., et al., *The RNA-binding protein HuR is essential for the B cell antibody response.* Nat Immunol, 2015. **16**(4): p. 415-25.
304. Berkovits, B.D. and C. Mayr, *Alternative 3' UTRs act as scaffolds to regulate membrane protein localization.* Nature, 2015. **522**(7556): p. 363-7.

305. Rogulja-Ortmann, A., et al., *The RNA-binding protein ELAV regulates Hox RNA processing, expression and function within the Drosophila nervous system.* Development, 2014. **141**(10): p. 2046-56.
306. Soller, M. and K. White, *ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA.* Genes Dev, 2003. **17**(20): p. 2526-38.
307. Hilgers, V., S.B. Lemke, and M. Levine, *ELAV mediates 3' UTR extension in the Drosophila nervous system.* Genes Dev, 2012. **26**(20): p. 2259-64.
308. Oktaba, K., et al., *ELAV Links Paused Pol II to Alternative Polyadenylation in the Drosophila Nervous System.* Mol Cell, 2014.
309. Koushika, S.P., M. Soller, and K. White, *The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein.* Mol Cell Biol, 2000. **20**(5): p. 1836-45.
310. Koushika, S.P., M.J. Lisbin, and K. White, *ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform.* Curr Biol, 1996. **6**(12): p. 1634-41.
311. Valcarcel, J. and F. Gebauer, *Post-transcriptional regulation: the dawn of PTB.* Curr Biol, 1997. **7**(11): p. R705-8.
312. Wagner, E.J. and M.A. Garcia-Blanco, *Polypyrimidine tract binding protein antagonizes exon definition.* Mol Cell Biol, 2001. **21**(10): p. 3281-8.
313. Castelo-Branco, P., et al., *Polypyrimidine tract binding protein modulates efficiency of polyadenylation.* Mol Cell Biol, 2004. **24**(10): p. 4174-83.
314. Blechingberg, J., et al., *Regulatory mechanisms for 3'-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript.* Nucleic Acids Res, 2007. **35**(22): p. 7636-50.
315. Coutinho-Mansfield, G.C., et al., *PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation.* Genes Dev, 2007. **21**(13): p. 1573-7.
316. Sawicka, K., et al., *Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein.* Biochem Soc Trans, 2008. **36**(Pt 4): p. 641-7.
317. Kafasla, P., et al., *Defining the roles and interactions of PTB.* Biochem Soc Trans, 2012. **40**(4): p. 815-20.
318. Costessi, L., et al., *Characterization of the distal polyadenylation site of the ss-adducin (Add2) pre-mRNA.* PLoS One, 2013. **8**(3): p. e58879.
319. Mikula, M., et al., *Heterogeneous nuclear ribonucleoprotein (HnRNP) K genome-wide binding survey reveals its role in regulating 3'-end RNA processing and transcription termination at the early growth response 1 (EGR1) gene through XRN2 exonuclease.* J Biol Chem, 2013. **288**(34): p. 24788-98.

320. Veraldi, K.L., et al., *hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells.* Mol Cell Biol, 2001. **21**(4): p. 1228-38.
321. Wesley, C.S., et al., *Loss of PTB or negative regulation of Notch mRNA reveals distinct zones of Notch and actin protein accumulation in Drosophila embryo.* PLoS One, 2011. **6**(7): p. e21876.
322. McDermott, S.M. and I. Davis, *Drosophila Hephaestus/polypyrimidine tract binding protein is required for dorso-ventral patterning and regulation of signalling between the germline and soma.* PLoS One, 2013. **8**(7): p. e69978.
323. Heimiller, J., et al., *Drosophila polypyrimidine tract-binding protein (DmPTB) regulates dorso-ventral patterning genes in embryos.* PLoS One, 2014. **9**(7): p. e98585.
324. Robida, M.D. and R. Singh, *Drosophila polypyrimidine-tract binding protein (PTB) functions specifically in the male germline.* EMBO J, 2003. **22**(12): p. 2924-33.
325. Castrillon, D.H., et al., *Toward a molecular genetic analysis of spermatogenesis in Drosophila melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis.* Genetics, 1993. **135**(2): p. 489-505.
326. Sridharan, V., et al., *High Throughput Sequencing Identifies Misregulated Genes in the Drosophila Polypyrimidine Tract-Binding Protein (hephaestus) Mutant Defective in Spermatogenesis.* PLoS One, 2016. **11**(3): p. e0150768.
327. Besse, F., et al., *Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation.* Genes Dev, 2009. **23**(2): p. 195-207.
328. Dominguez, D., et al., *Sequence, Structure, and Context Preferences of Human RNA Binding Proteins.* Mol Cell, 2018. **70**(5): p. 854-867 e9.
329. Gawande, B., et al., *Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline.* EMBO J, 2006. **25**(6): p. 1263-72.
330. Tiedje, C., et al., *The p38/MK2-driven exchange between tristetraprolin and HuR regulates AU-rich element-dependent translation.* PLoS Genet, 2012. **8**(9): p. e1002977.
331. Dassi, E., *Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins.* Front Mol Biosci, 2017. **4**(67): p. 67.
332. Peterson, M.L., *Immunoglobulin heavy chain gene regulation through polyadenylation and splicing competition.* Wiley Interdiscip Rev RNA, 2011. **2**(1): p. 92-105.
333. Hollerer, I., et al., *The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates*

mammalian mRNA expression in a quantitative and qualitative fashion. RNA, 2016. **22**(9): p. 1441-53.

334. Boukhatmi, H. and S. Bray, *A population of adult satellite-like cells in Drosophila is maintained through a switch in RNA-isoforms.* Elife, 2018. **7**: p. e35954.
335. Wissink, E.M., E.A. Fogarty, and A. Grimson, *High-throughput discovery of post-transcriptional cis-regulatory elements.* BMC Genomics, 2016. **17**: p. 177.
336. Wu, X. and D.P. Bartel, *Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability.* Cell, 2017. **169**(5): p. 905-917 e11.
337. Moqtaderi, Z., J.V. Geisberg, and K. Struhl, *Extensive Structural Differences of Closely Related 3' mRNA Isoforms: Links to Pab1 Binding and mRNA Stability.* Mol Cell, 2018. **72**(5): p. 849-861 e6.
338. Takagaki, Y., et al., *The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation.* Cell, 1996. **87**(5): p. 941-52.
339. Kubo, T., et al., *Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs.* Nucleic Acids Res, 2006. **34**(21): p. 6264-71.
340. Ryan, K. and D.L. Bauer, *Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation.* Int J Biochem Cell Biol, 2008. **40**(11): p. 2384-96.
341. Blee, Tajekesa K.P., Nicola K. Gray, and M. Brook, *Modulation of the cytoplasmic functions of mammalian post-transcriptional regulatory proteins by methylation and acetylation: a key layer of regulation waiting to be uncovered?* Biochemical Society Transactions, 2015. **43**(6): p. 1285-1295.
342. Mayr, C., *Regulation by 3'-Untranslated Regions.* Annu Rev Genet, 2017. **51**: p. 171-194.
343. Chen, C.Y. and A.B. Shyu, *AU-rich elements: characterization and importance in mRNA degradation.* Trends Biochem Sci, 1995. **20**(11): p. 465-70.
344. Barreau, C., L. Paillard, and H.B. Osborne, *AU-rich elements and associated factors: are there unifying principles?* Nucleic Acids Res, 2005. **33**(22): p. 7138-50.
345. Brewer, G., *An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro.* Mol Cell Biol, 1991. **11**(5): p. 2460-6.
346. Fan, X.C. and J.A. Steitz, *Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs.* EMBO J, 1998. **17**(12): p. 3448-60.
347. Stoecklin, G., et al., *A novel mechanism of tumor suppression by destabilizing AU-rich growth factor mRNA.* Oncogene, 2003. **22**(23): p. 3554-61.

348. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673-84.
349. Mazan-Mamczarz, K., et al., *RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8354-9.
350. Lau, A.G., et al., *Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF).* Proc Natl Acad Sci U S A, 2010. **107**(36): p. 15945-50.
351. Jia, J., et al., *Regulation and dysregulation of 3'UTR-mediated translational control.* Curr Opin Genet Dev, 2013. **23**(1): p. 29-34.
352. Ma, W. and C. Mayr, *A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions.* Cell, 2018. **175**(6): p. 1492-1506 e19.
353. Lee, S.-H. and C. Mayr, *Gain of Additional BIRC3 Protein Functions through 3ʹ-UTR-Mediated Protein Complex Formation.* Molecular Cell, 2019. **74**(4): p. 701-712.e9.
354. Ephrussi, A., L.K. Dickinson, and R. Lehmann, *Oskar organizes the germ plasm and directs localization of the posterior determinant nanos.* Cell, 1991. **66**(1): p. 37-50.
355. Huang, Y. and G.G. Carmichael, *Role of polyadenylation in nucleocytoplasmic transport of mRNA.* Mol Cell Biol, 1996. **16**(4): p. 1534-42.
356. An, J.J., et al., *Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons.* Cell, 2008. **134**(1): p. 175-87.
357. Andreassi, C. and A. Riccio, *To localize or not to localize: mRNA fate is in 3'UTR ends.* Trends Cell Biol, 2009. **19**(9): p. 465-74.
358. Jambor, H., et al., *Systematic imaging reveals features and changing localization of mRNAs in Drosophila development.* Elife, 2015. **4**.
359. Braz, S.O., et al., *Expression of Rac1 alternative 3' UTRs is a cell specific mechanism with a function in dendrite outgrowth in cortical neurons.* Biochim Biophys Acta Gene Regul Mech, 2017. **1860**(6): p. 685-694.
360. Martin, K.C. and A. Ephrussi, *mRNA localization: gene expression in the spatial dimension.* Cell, 2009. **136**(4): p. 719-30.
361. Lutz, C.S. and A. Moreira, *Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression.* Wiley Interdiscip Rev RNA, 2011. **2**(1): p. 22-31.

362. Lecuyer, E., et al., *Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.* Cell, 2007. **131**(1): p. 174-87.
363. Lehmann, R. and C. Nusslein-Volhard, *Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila.* Cell, 1986. **47**(1): p. 141-52.
364. Snee, M.J., et al., *A late phase of Oskar accumulation is crucial for posterior patterning of the Drosophila embryo, and is blocked by ectopic expression of Bruno.* Differentiation, 2007. **75**(3): p. 246-55.
365. Sinsimer, K.S., et al., *A late phase of germ plasm accumulation during Drosophila oogenesis requires lost and rumpelstiltskin.* Development, 2011. **138**(16): p. 3431-40.
366. Jambor, H., et al., *A stem-loop structure directs oskar mRNA to microtubule minus ends.* RNA, 2014. **20**(4): p. 429-39.
367. Brendza, R.P., et al., *A function for kinesin I in the posterior transport of oskar mRNA and Staufen protein.* Science, 2000. **289**(5487): p. 2120-2.
368. Ghosh, S., et al., *Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA.* Nat Struct Mol Biol, 2012. **19**(4): p. 441-9.
369. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.
370. Shi, Y., *Alternative polyadenylation: new insights from global analyses.* RNA, 2012. **18**(12): p. 2105-17.
371. Hoque, M., et al., *Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing.* Nat Methods, 2013. **10**(2): p. 133-9.
372. Derti, A., et al., *A quantitative atlas of polyadenylation in five mammals.* Genome Res, 2012. **22**(6): p. 1173-83.
373. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.
374. Ben-Dov, C., et al., *Genome-wide analysis of alternative pre-mRNA splicing.* J Biol Chem, 2008. **283**(3): p. 1229-33.
375. Elkon, R., A.P. Ugalde, and R. Agami, *Alternative cleavage and polyadenylation: extent, regulation and function.* Nat Rev Genet, 2013. **14**(7): p. 496-506.
376. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome Res, 2000. **10**(7): p. 1001-10.
377. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites.* Science, 2008. **320**(5883): p. 1643-7.

378. Elkon, R., et al., *E2F mediates enhanced alternative polyadenylation in proliferation.* Genome Biol, 2012. **13**(7): p. R59.
379. Zhang, H., J.Y. Lee, and B. Tian, *Biased alternative polyadenylation in human tissues.* Genome Biol, 2005. **6**(12): p. R100.
380. Hilgers, V., et al., *Neural-specific elongation of 3' UTRs during Drosophila development.* Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15864-9.
381. Shepard, P.J., et al., *Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq.* RNA, 2011. **17**(4): p. 761-72.
382. Miura, P., et al., *Widespread and extensive lengthening of 3' UTRs in the mammalian brain.* Genome Res, 2013. **23**(5): p. 812-25.
383. Ji, Z., et al., *Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7028-33.
384. West, S.M., et al., *Developmental dynamics of gene expression and alternative polyadenylation in the Caenorhabditis elegans germline.* Genome Biol, 2018. **19**(1): p. 8.
385. Velten, L., et al., *Single-cell polyadenylation site mapping reveals 3' isoform choice variability.* Mol Syst Biol, 2015. **11**(6): p. 812.
386. Hwang, H.W., et al., *cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation.* Neuron, 2017. **95**(6): p. 1334-1349 e5.
387. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.* Nat Rev Mol Cell Biol, 2009. **10**(11): p. 741-54.
388. Robida, M., et al., *Drosophila polypyrimidine tract-binding protein is necessary for spermatid individualization.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(28): p. 12570-12575.
389. Early, P., et al., *Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways.* Cell, 1980. **20**(2): p. 313-9.
390. Rogers, J., et al., *Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain.* Cell, 1980. **20**(2): p. 303-12.
391. Alt, F.W., et al., *Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends.* Cell, 1980. **20**(2): p. 293-301.
392. Peterson, M.L., *Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development.* Immunol Res, 2007. **37**(1): p. 33-46.

393. Takagaki, Y. and J.L. Manley, *Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation.* Mol Cell, 1998. **2**(6): p. 761-71.
394. Singh, R., J. Valcarcel, and M.R. Green, *Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins.* Science, 1995. **268**(5214): p. 1173-6.
395. Wojcik, E., et al., *Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene.* Genetics, 1994. **138**(4): p. 1163-70.
396. Sunkel, C.E. and D.M. Glover, *polo, a mitotic mutant of Drosophila displaying abnormal spindle poles.* J Cell Sci, 1988. **89 (Pt 1)**: p. 25-38.
397. Llamazares, S., et al., *polo encodes a protein kinase homolog required for mitosis in Drosophila.* Genes Dev, 1991. **5**(12A): p. 2153-65.
398. Greenleaf, A.L., et al., *Alpha-amanitin-resistant D. melanogaster with an altered RNA polymerase II.* Cell, 1979. **18**(3): p. 613-22.
399. Matia-Gonzalez, A.M., E.E. Laing, and A.P. Gerber, *Conserved mRNA-binding proteomes in eukaryotic organisms.* Nat Struct Mol Biol, 2015. **22**(12): p. 1027-33.
400. Stoiber, M.H., et al., *Extensive cross-regulation of post-transcriptional regulatory networks in Drosophila.* Genome Res, 2015. **25**(11): p. 1692-702.
401. Hanks, S.K., A.M. Quinn, and T. Hunter, *The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.* Science, 1988. **241**(4861): p. 42-52.
402. Lee, K.S., et al., *Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase Plk.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9301-6.
403. Cheng, K.Y., et al., *The crystal structure of the human polo-like kinase-1 polo box domain and its phospho-peptide complex.* EMBO J, 2003. **22**(21): p. 5757-68.
404. Reynolds, N. and H. Ohkura, *Polo boxes form a single functional domain that mediates interactions with multiple proteins in fission yeast polo kinase.* J Cell Sci, 2003. **116**(Pt 7): p. 1377-87.
405. García-Alvarez, B., et al., *Molecular and structural basis of polo-like kinase 1 substrate recognition: Implications in centrosomal localization.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3107-12.
406. Park, J.E., et al., *Polo-box domain: a versatile mediator of polo-like kinase function.* Cell Mol Life Sci, 2010. **67**(12): p. 1957-70.
407. Elia, A.E., L.C. Cantley, and M.B. Yaffe, *Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates.* Science, 2003. **299**(5610): p. 1228-31.

408. Elia, A.E., et al., *The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain.* Cell, 2003. **115**(1): p. 83-95.
409. Barr, F.A., H.H. Sillje, and E.A. Nigg, *Polo-like kinases and the orchestration of cell division.* Nat Rev Mol Cell Biol, 2004. **5**(6): p. 429-40.
410. Donaldson, M.M., et al., *Metaphase arrest with centromere separation in polo mutants of Drosophila.* J Cell Biol, 2001. **153**(4): p. 663-76.
411. Tavares, A.A., D.M. Glover, and C.E. Sunkel, *The conserved mitotic kinase polo is regulated by phosphorylation and has preferred microtubule-associated substrates in Drosophila embryo extracts.* EMBO J, 1996. **15**(18): p. 4873-83.
412. Graveley, B.R., et al., *The D. melanogaster transcriptome: modENCODE RNA-Seq data*. 2010.
413. Edgar, B.A. and P.H. O'Farrell, *Genetic control of cell division patterns in the Drosophila embryo.* Cell, 1989. **57**(1): p. 177-87.
414. Lehner, C.F. and P.H. O'Farrell, *Expression and function of Drosophila cyclin A during embryonic cell cycle progression.* Cell, 1989. **56**(6): p. 957-68.
415. Whitfield, W.G., et al., *Transcripts of one of two Drosophila cyclin genes become localized in pole cells during embryogenesis.* Nature, 1989. **338**(6213): p. 337-40.
416. Lehner, C.F. and P.H. O'Farrell, *The roles of Drosophila cyclins A and B in mitotic control.* Cell, 1990. **61**(3): p. 535-47.
417. Whitfield, W.G., et al., *The A- and B-type cyclins of Drosophila are accumulated and destroyed in temporally distinct events that define separable phases of the G2-M transition.* EMBO J, 1990. **9**(8): p. 2563-72.
418. Jimenez, J., et al., *Complementation of fission yeast cdc2ts and cdc25ts mutants identifies two cell cycle genes from Drosophila: a cdc2 homologue and string.* EMBO J, 1990. **9**(11): p. 3565-71.
419. Lehner, C.F. and P.H. O'Farrell, *Drosophila cdc2 homologs: a functional homolog is coexpressed with a cognate variant.* EMBO J, 1990. **9**(11): p. 3573-81.
420. Edgar, B.A. and P.H. O'Farrell, *The three postblastoderm cell cycles of Drosophila embryogenesis are regulated in G2 by string.* Cell, 1990. **62**(3): p. 469-80.
421. Fenton, B. and D.M. Glover, *A conserved mitotic kinase active at late anaphase-telophase in syncytial Drosophila embryos.* Nature, 1993. **363**(6430): p. 637-40.
422. Archambault, V., et al., *Sequestration of Polo kinase to microtubules by phosphopriming-independent binding to Map205 is relieved by phosphorylation at a CDK site in mitosis.* Genes Dev, 2008. **22**(19): p. 2707-20.
423. Bruinsma, W., J.A. Raaijmakers, and R.H. Medema, *Switching Polo-like kinase-1 on and off in time and space.* Trends in Biochemical Sciences, 2012. **37**(12): p. 534-542.

424. Logarinho, E. and C.E. Sunkel, *The Drosophila POLO kinase localises to multiple compartments of the mitotic apparatus and is required for the phosphorylation of MPM2 reactive epitopes.* J Cell Sci, 1998. **111 (Pt 19)**: p. 2897-909.
425. Sumara, I., et al., *Roles of polo-like kinase 1 in the assembly of functional mitotic spindles.* Curr Biol, 2004. **14**(19): p. 1712-22.
426. Lenart, P., et al., *The small-molecule inhibitor BI 2536 reveals novel insights into mitotic roles of polo-like kinase 1.* Curr Biol, 2007. **17**(4): p. 304-15.
427. Golsteyn, R.M., et al., *Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function.* J Cell Biol, 1995. **129**(6): p. 1617-28.
428. Lee, K.S., et al., *Plk is an M-phase-specific protein kinase and interacts with a kinesin-like protein, CHO1/MKLP-1.* Mol Cell Biol, 1995. **15**(12): p. 7143-51.
429. Moutinho-Santos, T., *In vivo localisation of the mitotic POLO kinase shows a highly dynamic association with the mitotic apparatus during early embryogenesis in Drosophila.* Biology of the Cell, 1999. **91**(8): p. 585-596.
430. Conde, C., et al., *Drosophila Polo regulates the spindle assembly checkpoint through Mps1-dependent BubR1 phosphorylation.* EMBO J, 2013. **32**(12): p. 1761-77.
431. Kulukian, A., J.S. Han, and D.W. Cleveland, *Unattached kinetochores catalyze production of an anaphase inhibitor that requires a Mad2 template to prime Cdc20 for BubR1 binding.* Dev Cell, 2009. **16**(1): p. 105-17.
432. Santaguida, S., et al., *Evidence that Aurora B is implicated in spindle checkpoint signalling independently of error correction.* EMBO J, 2011. **30**(8): p. 1508-19.
433. Saurin, A.T., et al., *Aurora B potentiates Mps1 activation to ensure rapid checkpoint establishment at the onset of mitosis.* Nat Commun, 2011. **2**: p. 316.
434. Heinrich, S., et al., *Mph1 kinetochore localization is crucial and upstream in the hierarchy of spindle assembly checkpoint protein recruitment to kinetochores.* J Cell Sci, 2012. **125**(Pt 20): p. 4720-7.
435. Luo, X. and H. Yu, *Mitosis: short-circuiting spindle checkpoint signaling.* Curr Biol, 2012. **22**(4): p. R128-30.
436. Maciejowski, J., et al., *Mps1 directs the assembly of Cdc20 inhibitory complexes during interphase and mitosis to control M phase timing and spindle checkpoint signaling.* J Cell Biol, 2010. **190**(1): p. 89-100.
437. Sliedrecht, T., et al., *Chemical genetic inhibition of Mps1 in stable human cell lines reveals novel aspects of Mps1 function in mitosis.* PLoS One, 2010. **5**(4): p. e10251.

438. Kang, J., et al., *Autophosphorylation-dependent activation of human Mps1 is required for the spindle checkpoint.* Proc Natl Acad Sci U S A, 2007. **104**(51): p. 20232-7.
439. Jelluma, N., et al., *Chromosomal instability by inefficient Mps1 auto-activation due to a weakened mitotic checkpoint and lagging chromosomes.* PLoS One, 2008. **3**(6): p. e2415.
440. Carmena, M., et al., *The chromosomal passenger complex activates Polo kinase at centromeres.* PLoS Biol, 2012. **10**(1): p. e1001250.
441. Kachaner, D., et al., *Interdomain allosteric regulation of Polo kinase by Aurora B and Map205 is required for cytokinesis.* J Cell Biol, 2014. **207**(2): p. 201-11.
442. Hauf, S., et al., *The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint.* J Cell Biol, 2003. **161**(2): p. 281-94.
443. Lens, S.M., et al., *Survivin is required for a sustained spindle checkpoint arrest in response to lack of tension.* EMBO J, 2003. **22**(12): p. 2934-47.
444. Emanuele, M.J., et al., *Aurora B kinase and protein phosphatase 1 have opposing roles in modulating kinetochore assembly.* J Cell Biol, 2008. **181**(2): p. 241-54.
445. Nathan, D.F. and S. Lindquist, *Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase.* Mol Cell Biol, 1995. **15**(7): p. 3917-25.
446. Buchner, J., *Hsp90 & Co. - a holding for folding.* Trends Biochem Sci, 1999. **24**(4): p. 136-41.
447. de Carcer, G., et al., *Requirement of Hsp90 for centrosomal function reflects its regulation of Polo kinase stability.* EMBO J, 2001. **20**(11): p. 2878-84.
448. Martins, T., et al., *Sgt1, a co-chaperone of Hsp90 stabilizes Polo and is required for centrosome organization.* EMBO J, 2009. **28**(3): p. 234-47.
449. Golsteyn, R.M., et al., *Cell cycle analysis and chromosomal localization of human Plk1, a putative homologue of the mitotic kinases Drosophila polo and Saccharomyces cerevisiae Cdc5.* J Cell Sci, 1994. **107 (Pt 6)**: p. 1509-17.
450. Glover, D.M., I.M. Hagan, and A.A.M. Tavares, *Polo-like kinases: a team that plays throughout mitosis.* Genes & Development, 1998. **12**(24): p. 3777-3787.
451. Nigg, E.A., *Polo-like kinases: positive regulators of cell division from start to finish.* Curr Opin Cell Biol, 1998. **10**(6): p. 776-83.
452. Dai, W., X. Huang, and Q. Ruan, *Polo-like kinases in cell cycle checkpoint control.* Front Biosci, 2003. **8**: p. d1128-33.
453. Archambault, V., G. Lepine, and D. Kachaner, *Understanding the Polo Kinase machine.* Oncogene, 2015. **34**(37): p. 4799-807.

454. Kitada, K., et al., *A multicopy suppressor gene of the Saccharomyces cerevisiae G1 cell cycle mutant gene dbf4 encodes a protein kinase and is identified as CDC5.* Mol Cell Biol, 1993. **13**(7): p. 4445-57.
455. Ohkura, H., I.M. Hagan, and D.M. Glover, *The conserved Schizosaccharomyces pombe kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells.* Genes Dev, 1995. **9**(9): p. 1059-73.
456. Kumagai, A. and W.G. Dunphy, *Purification and molecular cloning of Plx1, a Cdc25-regulatory kinase from Xenopus egg extracts.* Science, 1996. **273**(5280): p. 1377-80.
457. Clay, F.J., et al., *Identification and cloning of a protein kinase-encoding mouse gene, Plk, related to the polo gene of Drosophila.* Proc Natl Acad Sci U S A, 1993. **90**(11): p. 4882-6.
458. Lake, R.J. and W.R. Jelinek, *Cell cycle- and terminal differentiation-associated regulation of the mouse mRNA encoding a conserved mitotic protein kinase.* Mol Cell Biol, 1993. **13**(12): p. 7793-801.
459. Hamanaka, R., et al., *Cloning and characterization of human and murine homologues of the Drosophila polo serine-threonine kinase.* Cell Growth Differ, 1994. **5**(3): p. 249-57.
460. Holtrich, U., et al., *Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1736-40.
461. Andrysik, Z., et al., *The novel mouse Polo-like kinase 5 responds to DNA damage and localizes in the nucleolus.* Nucleic Acids Res, 2010. **38**(9): p. 2931-43.
462. Dai, W., *Polo-like kinases, an introduction.* Oncogene, 2005. **24**(2): p. 214-6.
463. Zitouni, S., et al., *Polo-like kinases: structural variations lead to multiple functions.* Nat Rev Mol Cell Biol, 2014. **15**(7): p. 433-52.
464. Tsvetkov, L. and D.F. Stern, *Phosphorylation of Plk1 at S137 and T210 is inhibited in response to DNA damage.* Cell Cycle, 2005. **4**(1): p. 166-71.
465. Charles, J.F., et al., *The Polo-related kinase Cdc5 activates and is destroyed by the mitotic cyclin destruction machinery in S. cerevisiae.* Curr Biol, 1998. **8**(9): p. 497-507.
466. Descombes, P. and E.A. Nigg, *The polo-like kinase Plx1 is required for M phase exit and destruction of mitotic regulators in Xenopus egg extracts.* EMBO J, 1998. **17**(5): p. 1328-35.

467. Kotani, S., et al., *PKA and MPF-activated polo-like kinase regulate anaphase-promoting complex activity and mitosis progression.* Mol Cell, 1998. **1**(3): p. 371-80.
468. Shirayama, M., et al., *The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in Saccharomyces cerevisiae.* EMBO J, 1998. **17**(5): p. 1336-49.
469. Gheghiani, L., et al., *PLK1 Activation in Late G2 Sets Up Commitment to Mitosis.* Cell Rep, 2017. **19**(10): p. 2060-2073.
470. Lane, H.A. and E.A. Nigg, *Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes.* J Cell Biol, 1996. **135**(6 Pt 2): p. 1701-13.
471. Rapley, J., et al., *Coordinate regulation of the mother centriole component nlp by nek2 and plk1 protein kinases.* Mol Cell Biol, 2005. **25**(4): p. 1309-24.
472. De Luca, M., P. Lavia, and G. Guarguaglini, *A functional interplay between Aurora-A, Plk1 and TPX2 at spindle poles: Plk1 controls centrosomal localization of Aurora-A and TPX2 spindle association.* Cell Cycle, 2006. **5**(3): p. 296-303.
473. Ahonen, L.J., et al., *Polo-like kinase 1 creates the tension-sensing 3F3/2 phosphoepitope and modulates the association of spindle-checkpoint proteins at kinetochores.* Curr Biol, 2005. **15**(12): p. 1078-89.
474. Goto, H., et al., *Complex formation of Plk1 and INCENP required for metaphase-anaphase transition.* Nat Cell Biol, 2006. **8**(2): p. 180-7.
475. Glover, D.M., H. Ohkura, and A. Tavares, *Polo kinase: the choreographer of the mitotic stage?* J Cell Biol, 1996. **135**(6 Pt 2): p. 1681-4.
476. Mundt, K.E., et al., *On the regulation and function of human polo-like kinase 1 (PLK1): effects of overexpression on cell cycle progression.* Biochem Biophys Res Commun, 1997. **239**(2): p. 377-85.
477. Brennan, I.M., et al., *Polo-like kinase controls vertebrate spindle elongation and cytokinesis.* PLoS One, 2007. **2**(5): p. e409.
478. Burkard, M.E., et al., *Chemical genetics reveals the requirement for Polo-like kinase 1 activity in positioning RhoA and triggering cytokinesis in human cells.* Proc Natl Acad Sci U S A, 2007. **104**(11): p. 4383-8.
479. Petronczki, M., et al., *Polo-like kinase 1 triggers the initiation of cytokinesis in human cells by promoting recruitment of the RhoGEF Ect2 to the central spindle.* Dev Cell, 2007. **12**(5): p. 713-25.
480. Santamaria, A., et al., *Use of the novel Plk1 inhibitor ZK-thiazolidinone to elucidate functions of Plk1 in early and late stages of mitosis.* Mol Biol Cell, 2007. **18**(10): p. 4024-36.

481. Lindon, C. and J. Pines, *Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells.* J Cell Biol, 2004. **164**(2): p. 233-41.
482. Uchiumi, T., D.L. Longo, and D.K. Ferris, *Cell cycle regulation of the human polo-like kinase (PLK) promoter.* J Biol Chem, 1997. **272**(14): p. 9166-74.
483. Martin, B.T. and K. Strebhardt, *Polo-like kinase 1: target and regulator of transcriptional control.* Cell Cycle, 2006. **5**(24): p. 2881-5.
484. Shin, C.H., et al., *Regulation of PLK1 through competition between hnRNPK, miR-149-3p and miR-193b-5p.* Cell Death Differ, 2017. **24**(11): p. 1861-1871.
485. Qian, Y.W., et al., *Activated polo-like kinase Plx1 is required at multiple points during mitosis in Xenopus laevis.* Mol Cell Biol, 1998. **18**(7): p. 4262-71.
486. Schild, D. and B. Byers, *Diploid spore formation and other meiotic effects of two cell-division-cycle mutations of Saccharomyces cerevisiae.* Genetics, 1980. **96**(4): p. 859-76.
487. Golan, A., Y. Yudkovsky, and A. Hershko, *The cyclin-ubiquitin ligase activity of cyclosome/APC is jointly activated by protein kinases Cdk1-cyclin B and Plk.* J Biol Chem, 2002. **277**(18): p. 15552-7.
488. May, K.M., et al., *Polo boxes and Cut23 (Apc8) mediate an interaction between polo kinase and the anaphase-promoting complex for fission yeast mitosis.* J Cell Biol, 2002. **156**(1): p. 23-8.
489. Kraft, C., et al., *Mitotic regulation of the human anaphase-promoting complex by phosphorylation.* EMBO J, 2003. **22**(24): p. 6598-609.
490. Chi, Y.H., et al., *Requirements for protein phosphorylation and the kinase activity of polo-like kinase 1 (Plk1) for the kinetochore function of mitotic arrest deficiency protein 1 (Mad1).* J Biol Chem, 2008. **283**(51): p. 35834-44.
491. van Vugt, M.A., et al., *Polo-like kinase-1 is required for bipolar spindle formation but is dispensable for anaphase promoting complex/Cdc20 activation and initiation of cytokinesis.* J Biol Chem, 2004. **279**(35): p. 36841-54.
492. McInnes, C., et al., *Inhibitors of Polo-like kinase reveal roles in spindle-pole maintenance.* Nat Chem Biol, 2006. **2**(11): p. 608-17.
493. Peters, U., et al., *Probing cell-division phenotype space and Polo-like kinase function using small molecules.* Nat Chem Biol, 2006. **2**(11): p. 618-26.
494. Steegmaier, M., et al., *BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth in vivo.* Curr Biol, 2007. **17**(4): p. 316-22.
495. Petronczki, M., P. Lenart, and J.M. Peters, *Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1.* Dev Cell, 2008. **14**(5): p. 646-59.

496. Ferris, D.K., S.C. Maloid, and C.C. Li, *Ubiquitination and proteasome mediated degradation of polo-like kinase.* Biochem Biophys Res Commun, 1998. **252**(2): p. 340-4.
497. Bassermann, F., et al., *The Cdc14B-Cdh1-Plk1 axis controls the G2 DNA-damage-response checkpoint.* Cell, 2008. **134**(2): p. 256-67.
498. Archambault, V. and D.M. Glover, *Polo-like kinases: conservation and divergence in their functions and regulation.* Nat Rev Mol Cell Biol, 2009. **10**(4): p. 265-275.
499. Jang, Y.J., et al., *Functional studies on the role of the C-terminal domain of mammalian polo-like kinase.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 1984-9.
500. Wen, D., et al., *SUMOylation Promotes Nuclear Import and Stabilization of Polo-like Kinase 1 to Support Its Mitotic Function.* Cell Rep, 2017. **21**(8): p. 2147-2159.
501. Hamanaka, R., et al., *Polo-like kinase is a cell cycle-regulated kinase activated during mitosis.* J Biol Chem, 1995. **270**(36): p. 21086-91.
502. Kothe, M., et al., *Structure of the catalytic domain of human polo-like kinase 1.* Biochemistry, 2007. **46**(20): p. 5960-71.
503. Macurek, L., et al., *Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery.* Nature, 2008. **455**(7209): p. 119-23.
504. Maroto, B., et al., *P21-activated kinase is required for mitotic progression and regulates Plk1.* Oncogene, 2008. **27**(36): p. 4900-8.
505. Seki, A., et al., *Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry.* Science, 2008. **320**(5883): p. 1655-8.
506. Yamashiro, S., et al., *Myosin phosphatase-targeting subunit 1 regulates mitosis by antagonizing polo-like kinase 1.* Dev Cell, 2008. **14**(5): p. 787-97.
507. Archambault, V. and M. Carmena, *Polo-like kinase-activating kinases: Aurora A, Aurora B and what else?* Cell Cycle, 2012. **11**(8): p. 1490-5.
508. Caron, D., et al., *Mitotic phosphotyrosine network analysis reveals that tyrosine phosphorylation regulates Polo-like kinase 1 (PLK1).* Sci Signal, 2016. **9**(458): p. rs14.
509. Qian, Y.W., E. Erikson, and J.L. Maller, *Mitotic effects of a constitutively active mutant of the Xenopus polo-like kinase Plx1.* Mol Cell Biol, 1999. **19**(12): p. 8625-32.
510. Jang, Y.J., et al., *Phosphorylation of threonine 210 and the role of serine 137 in the regulation of mammalian polo-like kinase.* J Biol Chem, 2002. **277**(46): p. 44115-20.
511. Kelm, O., et al., *Cell cycle-regulated phosphorylation of the Xenopus polo-like kinase Plx1.* J Biol Chem, 2002. **277**(28): p. 25247-56.

512. van de Weerdt, B.C., et al., *Uncoupling anaphase-promoting complex/cyclosome activity from spindle assembly checkpoint control by deregulating polo-like kinase 1.* Mol Cell Biol, 2005. **25**(5): p. 2031-44.
513. Lee, K.S. and R.L. Erikson, *Plk is a functional homolog of Saccharomyces cerevisiae Cdc5, and elevated Plk activity induces multiple septation structures.* Molecular and Cellular Biology, 1997. **17**(6): p. 3408-3417.
514. Hirota, T., et al., *Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells.* Cell, 2003. **114**(5): p. 585-98.
515. Hutterer, A., et al., *Mitotic activation of the kinase Aurora-A requires its binding partner Bora.* Dev Cell, 2006. **11**(2): p. 147-57.
516. Chan, E.H., et al., *Plk1 regulates mitotic Aurora A function through betaTrCP-dependent degradation of hBora.* Chromosoma, 2008. **117**(5): p. 457-69.
517. Seki, A., et al., *Plk1- and beta-TrCP-dependent degradation of Bora controls mitotic progression.* J Cell Biol, 2008. **181**(1): p. 65-78.
518. Cooke, C.A., M.M. Heck, and W.C. Earnshaw, *The inner centromere protein (INCENP) antigens: movement from inner centromere to midbody during mitosis.* J Cell Biol, 1987. **105**(5): p. 2053-67.
519. Earnshaw, W.C. and R.L. Bernat, *Chromosomal passengers: toward an integrated view of mitosis.* Chromosoma, 1991. **100**(3): p. 139-46.
520. Adams, R.R., et al., *Essential roles of Drosophila inner centromere protein (INCENP) and Aurora B in Histone H3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation.* Journal of Cell Biology, 2001. **153**(4): p. 865-880.
521. Chang, C.J., et al., *Drosophila Incenp is required for cytokinesis and asymmetric cell division during development of the nervous system.* Journal of Cell Science, 2006. **119**(6): p. 1144-1153.
522. Cheeseman, I.M., et al., *The conserved KMN network constitutes the core microtubule-binding site of the kinetochore.* Cell, 2006. **127**(5): p. 983-97.
523. DeLuca, J.G., et al., *Kinetochore microtubule dynamics and attachment stability are regulated by Hec1.* Cell, 2006. **127**(5): p. 969-82.
524. Welburn, J.P., et al., *Aurora B phosphorylates spatially distinct targets to differentially regulate the kinetochore-microtubule interface.* Mol Cell, 2010. **38**(3): p. 383-92.
525. Fairley, J.A., et al., *Direct regulation of tRNA and 5S rRNA gene transcription by Polo-like kinase 1.* Mol Cell, 2012. **45**(4): p. 541-52.

526. Papadopoulou, K., et al., *Regulation of gene expression during M-G1-phase in fission yeast through Plo1p and forkhead transcription factors.* J Cell Sci, 2008. **121**(Pt 1): p. 38-47.
527. Laoukili, J., et al., *FoxM1 is required for execution of the mitotic programme and chromosome stability.* Nat Cell Biol, 2005. **7**(2): p. 126-36.
528. Fu, Z., et al., *Plk1-dependent phosphorylation of FoxM1 regulates a transcriptional programme required for mitotic progression.* Nat Cell Biol, 2008. **10**(9): p. 1076-82.
529. Zhang, J., et al., *Polo-like kinase 1-mediated phosphorylation of Forkhead box protein M1b antagonizes its SUMOylation and facilitates its mitotic function.* J Biol Chem, 2015. **290**(6): p. 3708-19.
530. Yuan, C., et al., *The function of FOXO1 in the late phases of the cell cycle is suppressed by PLK1-mediated phosphorylation.* Cell Cycle, 2014. **13**(5): p. 807-19.
531. Christensen, M.D., et al., *Kinome-level screening identifies inhibition of polo-like kinase-1 (PLK1) as a target for enhancing non-viral transgene expression.* J Control Release, 2015. **204**: p. 20-9.
532. Del Rosario, B.C., et al., *Exploration of CTCF post-translation modifications uncovers Serine-224 phosphorylation by PLK1 at pericentric regions during the G2/M transition.* Elife, 2019. **8**: p. e42341.
533. Jiang, L., et al., *Polo-like kinase 1 inhibits the activity of positive transcription elongation factor of RNA Pol II β (P-TEFβ).* PLoS One, 2013. **8**(8): p. e72289.
534. Calvo, O. and J.L. Manley, *The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription.* EMBO J, 2005. **24**(5): p. 1009-20.
535. Martincic, K., et al., *Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing.* Nat Immunol, 2009. **10**(10): p. 1102-9.
536. Roberts, D.B., *Drosophila: a Practical Approach*. IRL Press at Oxford University Press. 1986.
537. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
538. Nelson, J.D., O. Denisenko, and K. Bomsztyk, *Protocol for the fast chromatin immunoprecipitation (ChIP) method.* Nat Protoc, 2006. **1**(1): p. 179-85.
539. Negre, N., et al., *A cis-regulatory map of the Drosophila genome.* Nature, 2011. **471**(7339): p. 527-31.
540. The UniProt, C., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(D1): p. D158-D169.

541. Mi, H., et al., *Large-scale gene function analysis with the PANTHER classification system.* Nat Protoc, 2013. **8**(8): p. 1551-66.
542. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Res, 2017. **45**(D1): p. D183-D189.
543. Winer, J., et al., *Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro.* Anal Biochem, 1999. **270**(1): p. 41-9.
544. Schmittgen, T.D., et al., *Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods.* Anal Biochem, 2000. **285**(2): p. 194-204.
545. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
546. Hoque, M., W. Li, and B. Tian, *Accurate mapping of cleavage and polyadenylation sites by 3' region extraction and deep sequencing.* Methods Mol Biol, 2014. **1125**: p. 119-29.
547. Zheng, D., X. Liu, and B. Tian, *3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA.* RNA, 2016. **22**(10): p. 1631-9.
548. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
549. Li, W., et al., *Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3? End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation.* PLoS Genet, 2015. **11**(4): p. e1005166.
550. Li, W., et al., *Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control.* BMC Biology, 2016. **14**(1): p. 1-17.
551. Anders, S., A. Reyes, and W. Huber, *Detecting differential usage of exons from RNA-seq data.* Genome Res, 2012. **22**(10): p. 2008-17.
552. Meyer, V., B. Oliver, and D. Pauli, *Multiple Developmental Requirements of Noisette, the Drosophila Homolog of the U2 snRNP-Associated Polypeptide SF3a60.* Molecular and Cellular Biology, 1998. **18**(4): p. 1835-1843.
553. Kirchner, J., et al., *Essential, Overlapping and Redundant Roles of the Drosophila Protein Phosphatase 1α and 1β Genes.* Genetics, 2007. **176**(1): p. 273-281.
554. Gramates, L.S., et al., *FlyBase at 25: looking to the future.* Nucleic Acids Res, 2017. **45**(D1): p. D663-D671.
555. RDevelopment CORE TEAM, R., *R: A language and environment for statistical computing*. 2008, R foundation for statistical computing Vienna, Austria.

556. Graveley, B.R., et al., *The developmental transcriptome of Drosophila melanogaster.* Nature, 2011. **471**(7339): p. 473-9.
557. Hu, Y., et al., *The Drosophila Gene Expression Tool (DGET) for expression analyses.* BMC Bioinformatics, 2017. **18**(1): p. 98.
558. Cheung, A.C. and P. Cramer, *Structural basis of RNA polymerase II backtracking, arrest and reactivation.* Nature, 2011. **471**(7337): p. 249-53.
559. FlyBase, C., M. Swiss-Prot Project, and M. InterPro Project, *Gene Ontology annotation in FlyBase through association of InterPro records with GO terms*. 2004.
560. Werner-Allen, J.W., et al., *cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase.* Journal of Biological Chemistry, 2011. **286**(7): p. 5717-5726.
561. Oktaba, K., et al., *ELAV links paused Pol II to alternative polyadenylation in the Drosophila nervous system.* Mol Cell, 2015. **57**(2): p. 341-8.
562. Zaharieva, E., et al., *Concentration and Localization of Coexpressed ELAV/Hu Proteins Control Specificity of mRNA Processing.* Mol Cell Biol, 2015. **35**(18): p. 3104-15.
563. Kaplan, C.D., et al., *Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster.* Genes & Development, 2000. **14**(20): p. 2623-2634.
564. Audibert, A. and M. Simonelig, *Autoregulation at the level of mRNA 3′ end formation of the suppressor of forked gene of Drosophila melanogaster is conserved in Drosophila virilis.* Proceedings of the National Academy of Sciences, 1998. **95**(24): p. 14302-14307.
565. Pan, Z., et al., *An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism.* Gene, 2006. **366**(2): p. 325-34.
566. Hsin, J.P. and J.L. Manley, *The RNA polymerase II CTD coordinates transcription and RNA processing.* Genes Dev, 2012. **26**(19): p. 2119-37.
567. Henriques, T., et al., *Transcription termination between polo and snap, two closely spaced tandem genes of D. melanogaster.* Transcription, 2012. **3**(4): p. 198-212.
568. White-Cooper, H., et al., *Mutations in new cell cycle genes that fail to complement a multiply mutant third chromosome of Drosophila.* Genetics, 1996. **144**(3): p. 1097-1111.
569. Jelluma, N., et al., *Release of Mps1 from kinetochores is crucial for timely anaphase onset.* J Cell Biol, 2010. **191**(2): p. 281-90.

570. Carmena, M., et al., *Polo kinase regulates the localization and activity of the chromosomal passenger complex in meiosis and mitosis in Drosophila melanogaster.* Open Biol, 2014. **4**(11): p. 140162.
571. von Schubert, C., et al., *Plk1 and Mps1 Cooperatively Regulate the Spindle Assembly Checkpoint in Human Cells.* Cell Rep, 2015. **12**(1): p. 66-78.
572. Moura, M., et al., *Protein Phosphatase 1 inactivates Mps1 to ensure efficient Spindle Assembly Checkpoint silencing.* Elife, 2017. **6**.
573. Sabin, L.R., et al., *Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila.* Cell, 2009. **138**(2): p. 340-351.
574. Rehwinkel, J., et al., *Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster.* Nat Struct Mol Biol, 2004. **11**(6): p. 558-66.
575. Stegeman, R., et al., *The Spliceosomal Protein SF3B5 is a Novel Component of Drosophila SAGA that Functions in Gene Expression Independent of Splicing.* Journal of Molecular Biology, 2016. **428**(18): p. 3632-3649.
576. Gelbart, W.M. and D.B. Emmert, *FlyBase High Throughput Expression Pattern Data*. 2013.
577. Freeland, D.E. and D.T. Kuhn, *Expression patterns of developmental genes reveal segment and parasegment organization of D. melanogaster genital discs.* Mech Dev, 1996. **56**(1-2): p. 61-72.
578. Moore, L.A., et al., *Identification of genes controlling germ cell migration and embryonic gonad formation in Drosophila.* Development, 1998. **125**(4): p. 667-78.
579. Lo, P.C., et al., *Homeotic genes autonomously specify the anteroposterior subdivision of the Drosophila dorsal vessel into aorta and heart.* Developmental Biology, 2002. **251**(2): p. 307-319.
580. DeFalco, T., S. Le Bras, and M. Van Doren, *Abdominal-B is essential for proper sexually dimorphic development of the Drosophila gonad.* Mechanisms of Development, 2004. **121**(11): p. 1323-1333.
581. Foronda, D., et al., *Requirement of Abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule.* Development, 2006. **133**(1): p. 117-27.
582. Mihaly, J., et al., *Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex.* Development, 2006. **133**(15): p. 2983-93.
583. Gerber, M., et al., *In vivo requirement of the RNA polymerase II elongation factor elongin A for proper gene expression and development.* Molecular and Cellular Biology, 2004. **24**(22): p. 9911-9919.
584. Kamieniarz-Gdula, K., et al., *Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination.* Molecular Cell, 2019.

585. Maslon, M.M., et al., *A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes.* The EMBO Journal, 2019: p. e101244.
586. Ji, Z. and B. Tian, *Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types.* PLoS One, 2009. **4**(12): p. e8419.
587. Natarajan, M., et al., *Negative elongation factor (NELF) coordinates RNA polymerase II pausing, premature termination, and chromatin remodeling to regulate HIV transcription.* J Biol Chem, 2013. **288**(36): p. 25995-6003.
588. Gibbs, E.B., et al., *Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain.* Nat Commun, 2017. **8**: p. 15233.
589. McNeill, H., et al., *mirror encodes a novel PBX-class homeoprotein the functions in the definition of the dorsal-ventral border in the Drosophila eye.* Genes & Development, 1997. **11**(8): p. 1073-1082.
590. Atkey, M.R., et al., *Capicua regulates follicle cell fate in the Drosophila ovary through repression of mirror.* Development, 2006. **133**(11): p. 2115-2123.
591. Karim, M.R. and A.W. Moore, *Convergent local identity and topographic projection of sensory neurons.* J Neurosci, 2011. **31**(47): p. 17017-27.
592. Sone, M., et al., *Still life, a protein in synaptic terminals of Drosophila homologous to GDP-GTP exchangers.* Science, 1997. **275**(5299): p. 543-547.
593. Mandaravally Madhavan, M. and H.A. Schneiderman, *Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development ofDrosophila melanogaster.* Wilehm Roux Arch Dev Biol, 1977. **183**(4): p. 269-305.
594. Roseland, C.R. and H.A. Schneiderman, *Regulation and metamorphosis of the abdominal histoblasts ofDrosophila melanogaster.* Wilehm Roux Arch Dev Biol, 1979. **186**(3): p. 235-265.
595. Madhavan, M.M. and K. Madhavan, *Morphogenesis of the epidermis of adult abdomen of Drosophila.* J Embryol Exp Morphol, 1980. **60**: p. 1-31.
596. Ninov, N. and E. Martin-Blanco, *Live imaging of epidermal morphogenesis during the development of the adult abdominal epidermis of Drosophila.* Nat Protoc, 2007. **2**(12): p. 3074-80.
597. Ninov, N., C. Manjon, and E. Martin-Blanco, *Dynamic control of cell cycle and growth coupling by ecdysone, EGFR, and PI3K signaling in Drosophila histoblasts.* PLoS Biol, 2009. **7**(4): p. e1000079.

598. Mortin, M.A., W.J. Kim, and J. Huang, *Antagonistic interactions between alleles of the RpII215 locus in Drosophila melanogaster.* Genetics, 1988. **119**(4): p. 863-73.
599. Greenleaf, A.L., et al., *Genetic and biochemical characterization of mutants at an RNA polymerase II locus in D. melanogaster.* Cell, 1980. **21**(3): p. 785-92.
600. Thomsen, S., et al., *Genome-wide analysis of mRNA decay patterns during early Drosophiladevelopment.* Genome Biology, 2010. **11**(9): p. R93.
601. Zhu, H., et al., *Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences.* J Biol Chem, 2007. **282**(4): p. 2203-10.
602. Perez, I., et al., *Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo.* RNA, 1997. **3**(7): p. 764-78.
603. Martinez, E., et al., *Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo.* Mol Cell Biol, 2001. **21**(20): p. 6782-95.
604. Kim-Ha, J., J. Kim, and Y.J. Kim, *Requirement of RBP9, a Drosophila Hu homolog, for regulation of cystocyte differentiation and oocyte determination during oogenesis.* Mol Cell Biol, 1999. **19**(4): p. 2505-14.
605. Park, S.J., et al., *Down regulation of extramacrochaetae mRNA by a Drosophila neural RNA binding protein Rbp9 which is homologous to human Hu proteins.* Nucleic Acids Res, 1998. **26**(12): p. 2989-94.
606. Jankowsky, E. and M.E. Harris, *Specificity and nonspecificity in RNA-protein interactions.* Nat Rev Mol Cell Biol, 2015. **16**(9): p. 533-44.
607. Herrmann, S., I. Amorim, and C.E. Sunkel, *The POLO kinase is required at multiple stages during spermatogenesis in Drosophila melanogaster.* Chromosoma, 1998. **107**(6-7): p. 440-51.
608. Mirouse, V., E. Formstecher, and J.L. Couderc, *Interaction between Polo and BicD proteins links oocyte determination and meiosis control in Drosophila.* Development, 2006. **133**(20): p. 4005-13.
609. Riparbelli, M.G., et al., *Inhibition of Polo kinase by BI2536 affects centriole separation during Drosophila male meiosis.* Cell Cycle, 2014. **13**(13): p. 2064-72.
610. Luo, Y., et al., *novel modifications on C-terminal domain of RNA polymerase II can fine-tune the phosphatase activity of Ssu72.* ACS Chem Biol, 2013. **8**(9): p. 2042-52.

ACKNOWLEDGEMENTS

I'd like to start this (longer than I expected) section by thanking my two main supervisors, Alexandra Moreira and Ana Pombo. This is because they had equal parts in the making of this long thesis, of sharing both my frustration and excitement for an experiment, my reluctance and then acceptance of having to re-write scientific reports, abstracts, posters and articles, for spurring my hesitance in thinking up new experiments. Thank you for all the moments we've shared and experienced, and especially, for the opportunity to learn and to grow both as a Scientist and as a person. I cannot say it was always an easy process or that I enjoyed every moment for the past five years for it would be a lie, but I am still enormously grateful for the chance you've given me, for allowing me to immerse myself into the world of Gene Regulation, transcription and signalling pathways, in my beautiful and amazing hometown of Porto and in great, awesome and fun Berlin.

Jaime Freitas, you've been my co-supervisor as well as my mentor and friend and probably the one that helped me out the most at the bench. We didn't always agree on every topic, but we always got along and without you, I wouldn't have gotten as far as I did (trust me, I know, and perhaps you do too). Thank you for the long philosophical chats we've had late at the IBMC and then i3S, thank you for your saint-like patience (just ask Mafalda) and thank you as well for your keenly-aware perception of asking the exact questions I don't know how to answer (but learned along the way).

I also need to thank Mafalda Pinho, who started working on her Master thesis when I began my PhD. Mafy, you're an amazing person, an extremely happy and effusive person, so very different from how and who I am, but perhaps that's why I've enjoyed having you nearby so much those first two years. Not only for your help, but as a friend, for all our long chats complaining about everything and nothing and for just being there whenever I needed you.

Joana Wilton, I probably met you at the lowest of my lows during my PhD, but I remember you telling me that I still smiled and laughed, so I seemed okay. Thank you for your keen eye, thank you for all the talks, all the lunches, thank you for understanding and just for being you. I needed that, I really did. You're an awesome woman and an even greater Scientist, so if anyone ever tells you otherwise, send them my way and I'll set them straight.

Thank you, Isabel Pereira-Castro, both as a Scientist and as a good friend. You've helped me to grow in many ways, to be smarter, to look beyond, to save some (lunch) time and weekends for me, to be super practical and that was and still is priceless to me. I

cherished our discussions, your tips, advice and suggestions to cut this and that and then this part too. You are and always will be GR's Mommy and you've made me feel right at home from the start.

Thank you so very much to all the members of both the Gene Regulation and the AG Pombo labs for each and every contribution you've done to the making (and writing) of this manuscript, even if it only consisted of a happy good morning: Joana Maia, Sandra Braz, Ana Curinha, Ana Keating, Andrea Cruz, Ana Jesus, Beatriz Garcia, Andreia Pinho, Olena Kutsenko (I miss you, you awkward person!), Sasha, João Dias, Izabella Harabula, Carmelo Ferrai (I miss your good humour so very much!!!), Rieke Kempfer, Gesa Loof, Robert Beagrie, Marta Slimak-Mastrobuoni, Ana Miguel, Nadina Skourti-Stathaki, Markus Schüler, Dorothee Krämer, Tiago Rito, Giulia Caglio, Mariano Barbieri and Elena Torlai.

I will also take the time to give my profuse thanks to the members of the Cell Division & Genomic Stability group: Augusta Monteiro (thank you for all the heads-up for when fly food was going to be made), Carlos Conde (thank you for the hours at the confocal microscope and advice, you were a precious help), Margarida Moura (I annoyed you so much!), João Barbosa (thanks for the brain squashes, I'd – very much literally – never do them without you), Mariana Osswald, Sofia Moreira, Pedro Resende and Eurico Sá. And last but not least, I am incredibly thankful to Professor Claudio Sunkel, the father of Polo, for inspiring me during my First and Master Degrees and for the good advice you've given me to make this thesis what it is today.

To my tight-knitted Biochemistry group that consists of Sara Moreira, Emiliana Pereira (I. Simply. Adore you.), Mariana Pereira and Pedro Melo (Yeti das Neves, you made it, SHOCK-U!). For years, we have all taken a long, hard journey together and then took our separate, but necessary turns. It is however wonderful and refreshing to know that our friendship stays as it is, just where we left it, exactly the same, years later once more. Thank you for the long talks, the long nights, the dinners, the lunches, the Killer Note, the themed picnics, Los Dorks, those scary Pharmacology flashcards, everybody else's bashing, and everything in between. I would not be where or who I am without you guys.

I also dedicate a full paragraph of this section to my cats, because they deserve it. I don't write properly when I'm not around them, I don't work properly without them nearby, it doesn't feel like coming home in Berlin when I'm not greeted at the door by something furry and purring. You don't know it and you never will, but you were a great help in grounding me, in relaxing me and overall just helping me enjoy the simple pleasures in life such as eating and sleeping, which are always great and lacking in a PhD student's life. Nininha (my forever 19 year old dear sister), Gil and Mel, you were always there for me

and I love you for it even if one of you opted to leave me right in the midst of this hard path.

And though more of a distraction from my work than a helpful hand, I am also thankful to my boyfriend. João Sousa, you've been a pillar of strength and calm and relaxation and mental replenishment to me from day 1 (although I really wanted to hit you when you went to London without me and increased the distance between us from 351 to 2026km. Yeah, I checked that). Still, I love you and you love me and that's all that really matters.

While they haven't had an active role in helping me complete another chapter of my career, I also have to thank my family, mostly for putting up with my crazy hours. To my parents, thank you for coming along with me at the beginning of my first three months of living completely on my own in Berlin, for helping out in the apartment and for giving me a head's up on how to cook the basic side dishes. Thank you.

Finally, I'd like to take a short moment to dedicate this thesis to myself. I now look back and truly value my own determination, sheer stubbornness and hardcore resilience, because without those, I certainly wouldn't have finished this work. It involved a lot of patience during the down (deep down, rock bottom, really) moments and those were definitely the hardest to push through. Not that hard qPCR that would just not work, not that ChIP result, weird because of a crappy antibody. No, all of that is what being a Scientist is all about. The hardest part is definitely the waiting between experiments and results and when you do have results, the hard part becomes trying to make sense of them and building a story with half-hazy pieces of a crooked puzzle for which you lost the final image of. *That's* the hard part.

Just smile and wave, guys. Smile and wave.