THE MECHANISM OF TRANSCRIPTION TERMINAITON BY THE PJW/PP1 COMPLEX IN TRYPANSOMATIDS

by

Yang Zhang

(Under the Direction of Robert Sabatini)

ABSTRACT

The Trypanosomatidae are early-diverged protozoan parasites, including human pathogens *Trypanosoma brucei* and *Leishmania major*. Within these parasites, hundreds of genes with unrelated functions are organized into long arrays throughout the genome, transcribed polycistronically by RNA polymerase II (RNAPII). Little is known about transcription termination in Trypanosomatids. Epigenetic marks, including base J, are enriched at RNAPII transcription termination regions in *T. brucei* and *Leishmania spp*., playing a critical role in transcription termination. The study of base J function led to the identification of the PJW/PP1 complex in *L. tarentolae*, composed of PP1, PNUTS, Wdr82, and the base J-binding protein JBP3. Ablation of these protein factors results in readthrough transcription at the end of gene arrays in *T. brucei* and *L. major*, indicating that the PJW/PP1 complex regulates transcription termination in Trypanosomatids. Unexpectedly, we discovered that in *T. brucei*, the PJW/PP1 complex is also essential for terminating antisense transcription from bidirectional promoters, ensuring transcriptional directionality, and for monoallelic expression of variant surface glycoprotein (VSG). Further exploration into the mechanism by which the complex regulates transcription termination revealed PP1-dependent phosphatase activity toward RNAPII C-

terminal domain (CTD), implicating CTD dephosphorylation at the end of gene arrays in regulating transcription termination. Additionally, our analysis of the intramolecular architecture of the PJW/PP1 complex focused on the role of PNUTS. PNUTS, as a scaffolding protein within the complex, mediates PP1 association with a central sequence containing conserved short linear motifs (SLiMs), specifically RVXF-ϕϕ-F motifs, and binds to Wdr82/JBP3 with its C-terminus. Despite the presence of eight PP1 homologs in *L. tarentolae*, PNUTS preferentially binds to PP1-8, discriminating between PP1 homologs based on their C-terminal tails and specific inserts within their catalytic core regions. Consistent with its role as a scaffolding protein, PNUTS protein levels must be finely tuned for complex integrity. These findings illuminate the process of transcription termination in trypanosomatids and establish a mechanistic link between base J and transcription termination.

INDEX WORDS: Trypanosomes, Leishmania, protein phosphatase 1, PJW/PP1 complex, base J, transcription termination, RNA polymerase II, polycistronic transcription

THE MECHANISM OF TRANSCRIPTION TERMINAITON BY THE PJW/PP1 COMPLEX IN TRYPANSOMATIDS

by

YANG ZHANG

BS, University of Wisconsin-Madison, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

THE MECHANISM OF TRANSCRIPTION TERMINAITON BY THE PJW/PP1 COMPLEX IN TRYPANSOMATIDS

by

YANG ZHANG

Major Professor: Robert Sabatini
Committee: David Garfinkel
Lance Wells
Zachary Lewis

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

DEDICATION

This dissertation is dedicated to Beina Yao, a soul singer, an ephemeral flower, and a forever gem.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep appreciation and gratitude to my advisor, Bob Sabatini, for his unwavering mentorship and support. He is always full of positive energy and his guidance has been invaluable in shaping my academic and research path. Next, I want to extend my heartfelt thanks to my labmate, Rudo Kieft, who has generously shared his reagents, experience, and invaluable advice with me. Without his support, this graduate school journey would have been much more challenging. I would also like to express my gratitude to my committee members for dedicating their time and providing valuable insights throughout my academic endeavors. Lastly, I want to thank my parents, family members, and my partner, L, for their unconditional love and care. Their support has been my rock, and it would have been impossible for me to reach this milestone without them by my side.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Transcription termination is important as it not only defines the gene borders, but also shapes the transcriptome by determining the cellular fate of the transcripts (1). Transcription termination defects in eukaryotes may have detrimental consequences for gene expression. For tandemly arranged protein-coding genes, extended readthrough transcripts will run into the promoter of the downstream gene, restricting its activity by transcriptional interference (2, 3). For genes arranged in convergent orientation on opposing DNA strands, defective transcription termination could lead to formation of double strand RNAs (dsRNAs), thus downregulating gene expression by RNA interference (RNAi) pathways (4). Moreover, not properly terminated transcription can result in a collision between elongating RNA polymerase II (RNAPII) on opposite DNA template strands (5, 6). It has been increasingly appreciated that RNAPII is engaged in pervasive transcription of the eukaryotic genome, generating pervasive transcripts (PTs) (7). Antisense transcripts transcribed from bidirectional promoters make up a significant fraction of PTs (8–11), and transcription termination is an important transcriptome surveillance mechanism to degrade them and ensure transcription directionality (12).

Although questions exist, one popular model for transcription termination is the torpedo model (13, 14), in which RNA cleavage by 3’ end processing factor after the poly(A) site provides an entry site for a 5’-3’ exoribonuclease that acts like a torpedo, co-transcriptionally degrading the nascent transcript until it catches up with RNAPII and displaces it from the DNA template. The mammalian PTW/PP1 complex, a protein-phosphatase 1-containing complex (15),

is essential in this process. It has been proposed that the PTW/PP1 complex dephosphorylates Spt5, an elongation factor, and RNAPII C-terminal domain (CTD), slowing down the elongating RNAPII after poly(A) sites (16). The altered kinetic makes it easier for the exoribonuclease to chase down the RNAPII, and facilitates transcription termination as a result.

Trypanosomatids are a group of parasitic protozoa that includes various species responsible for causing diseases in humans and animals, such as African sleeping sickness, Chagas disease, Leishmaniasis, etc. (17). They diverged from the main eukaryotic lineage around 200-500 million years ago (18), and carry many interesting genetic features, such as polycistronic transcription units, in which genes of unrelated functions are organized in tandem and transcribed altogether from a transcription start region constitutively by RNAPII (19). Individual mRNAs are trans-spliced from long polycistronic transcripts co-transcriptionally (20). Transcription termination, therefore, is decoupled from transcription termination in trypanosomatids to prevent premature transcription termination of the polycistronic transcription unit.

How transcription initiation and termination is regulated is little understood in trypanosomatids, but chromatin modifications have been regarded as important regulatory measures, including histone post-translational modifications (PTMs), histone variants, and base J (21). Base J is a hypermodified nucleobase, and is critical for regulating transcription termination. Its loss leads to transcription termination defects in multiple trypanosomatid species (22–24). How base J regulates transcription was not understood, but recent research has identified a base J-binding multiprotein complex, the PJW/PP1 complex that is involved in regulating transcription termination. Here , current understanding of transcription termination in trypanosomatids and recent findings about the PJW/PP1 complex will be discussed.

TRANSCRIPTION BY RNAPII

In eukaryotes, RNAPII transcribes both protein coding and non-protein coding genes (25). The transcription cycle can be divided roughly into 3 stages: initiation, elongation and termination (26). Regulatory checkpoints have been identified in each stage, playing a crucial role in controlling transcription and therefore gene expression (27–30). Initiation is a highly regulated process, where the RNAPII and ~80 different proteins have to be assembled on the promoter (31). Shortly after transcription initiation, ~30-60 nt downstream of the transcription start site comes another checkpoint, the promoter proximal pausing in mammalian cells (32). The last regulatory checkpoint is transcription termination, which is however, relatively less understood.

High-throughput DNA sequencing techniques have uncovered the pervasive transcription of genomes by RNAPII, which generates different types of non-coding RNAs (ncRNAs) (33). ncRNAs can be divided into short ncRNAs (<200 nt, including small nucleolar RNAs (snoRNAs), small interfering RNAs, microRNAs and PIWI-interacting RNAs) and long ncRNAs (lncRNAs, >200 nt that makes up the majority of ncRNAs) (34). Antisense transcripts are a class of long ncRNAs, and are transcribed from the strand opposite to that of the annotated sense transcript. Antisense transcripts are widespread in bacteria and eukaryotes (35–37). They are generally transcribed at a low level (38), retained in the nucleus (39), and do not code for proteins. More than 70% of the human and mouse genome generates antisense transcripts (40, 41). They are originated from independent promoters, bidirectional promoters (9–11, 42) or cryptic promoters (43–46), based on their position relative to sense transcripts. Bidirectional promoters are prevalent in species ranging from yeast (9, 11) to human cells (10). In yeast cells, pervasive transcription by RNAPII generates cryptic unstable transcripts (CUTs), and antisense

transcripts generated from bidirectional promoters make up a majority of CUTs of 200-600 nt long (9, 11, 47). In mammalian cells, antisense transcripts generated from bidirectional promoters are PROMPTs (promoter upstream transcripts). These transcripts are of low abundance, and usually rapidly degraded, leading to the speculation that they are just transcriptional noises; however, recent research has revealed their diverse involvements in cellular activities, including regulation of protein coding gene expression (48), and chromatin modifications (49). Regulation of these ncRNAs by transcription termination has been recognized as an important transcriptome surveillance mechanism.

CANONICAL TRANSCRIPTION TERMINATION MODELS

Transcription termination is critical to prevent readthrough transcription at 3' ends of genes, and to repress antisense transcription derived from bidirectional promoters at 5' ends, especially in the genome of organisms like yeast where neighboring genes are much closer than those in the human genome (50). Extended readthrough transcription into the neighboring genes could generate dsRNAs, triggering RNAi and gene repression, and causing detrimental effects. Transcription termination at 3' end of protein coding genes is relatively conserved in eukaryotes from yeast to mammalian cells, while termination of antisense transcripts is much less understood, and employs distinct mechanisms in yeast and mammalian cells.

Transcription termination at 3' end of protein coding genes

At the 3' end of protein coding genes, RNAPII encounters a polyadenylation signal (PAS), most commonly AAUAAA, and termination occurs (51). The transcription termination is coupled to 3' end processing of pre-mRNA involving pre-mRNA cleavage and polyadenylation to generate a mature transcript (Figure 1.1) (52, 53). In yeast, the 3' end processing machinery is

a multiprotein complex, including cleavage and polyadenylation factor (CPF) and cleavage factor complexes (CF, consisting of CFI and CFII subcomplexes). The large complex contains protein subunits that recognize the PAS in the nascent RNAs, CPF endonuclease Ysh1 cleaves the RNAs downstream of the PAS, and CPF-associated polyA polymerase (Pap1) adds a poly(A) tail onto the 3' end of the cleaved RNAs (54, 55). This process is also facilitated by subunits of the 3' end processing machinery that bind to RNAPII to bring the machinery close to the terminating complex. The mammalian counterpart of the yeast CPF-CF complex is slightly different, but most components are conserved (30, 56), including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), and cleavage factor (CF) complexes. Similar to yeast, the mammalian CPSF complex subunit CPSF73 cleaves the nascent RNA after the PAS is transcribed.

Two major transcription termination models have been proposed: the allosteric and the torpedo model (13, 14, 57, 58). In the allosteric model, transcription of PAS leads to a conformational change in RNAPII to slow it down and make it competent for termination. In the torpedo model (Figure 1.1), RNA cleavage by 3' end processing factor after PAS exposes 5' PO_4-end, providing an entry site for a 5'-3' exoribonuclease, XRN2/Rat1 (human/yeast) (59–62) that co-transcriptionally degrades the nascent transcript until it chases down RNAPII to dislodge it from the DNA template.

Antisense-transcription termination at the 5' end

Antisense transcription from bidirectional promoters at 5' end of genes are usually terminated by the Nrd1-Nab3-Sen1(NNS) pathway in yeast (63). The NNS complex is recruited early in transcription by Nrd1-RNAPII CTD interaction (64–67). In addition, Nrd1 and Nab3 are two RNA binding proteins that recognize specific motifs on nascent RNAs (68–70). When the

sequence motifs are transcribed, transcription termination is initiated in which Sen1, a superfamily I RNA and DNA helicase, possibly translocates to the nascent RNAs to catch up with the transcribing RNAPII by torpedo mechanism (1). In addition, Nrd1 recruits the TRAMP complex to polyadenylate, and exosome to rapidly degrade these antisense transcripts (71, 72).

Nrd1 and Nab3 homologs, however, are not identified in mammalian cells (73), suggesting mammalian antisense transcripts are processed by a different mechanism. Integrator is a metazoan-specific multi-subunit protein complex that carries out 3' end processing of non-coding RNAs, such as PROMPTs and snRNAs (74–76). PROMPTs are degraded rapidly by the integrator-associated nuclear exosome complex, after cleavage by the endonuclease component INTS11 (73, 77, 78). When the integrator is not recruited sufficiently, pervasive transcripts such as PROMPTs are processed by the cleavage and polyadenylation complex that could work redundantly with the integrator complex on certain types of transcripts (79). PROMOPTs 3' end are enriched in PASs that recruit all or part of the 3' processing machinery (80, 81). Furthermore, 3' end processing machinery possibly recruits NEXT complex (nuclear exosome targeting complex) (82), a functional counterpart of the yeast TRAMP complex, which recruits and activates the exosome for degrading PROMPTs (83). In contrast, the coding sequences of the sense transcription are relatively depleted of PASs, and the sense transcripts are bound by the U1 small nuclear ribonucleoprotein particles (snRNPs), which antagonize polyadenylation, thereby preventing premature transcription termination (81, 84).

Three factors that affect transcription termination

Three key factors that contribute to efficient transcription termination will be discussed here: RNAPII CTD PTMs, RNAPII kinetics, and epigenetic modifications. It is noteworthy that these factors are intricately interrelated as one factor could affect the other factors.

RNAPII CTD PTMs The smooth transition of the transcription cycle requires recruitment of certain protein factors in different stages. The largest subunit of RNAPII, Rpb1, carries a CTD tail that is critical in this process as it is a signaling and binding platform for proteins throughout the transcription cycle. The RNAPII CTD is a highly disordered polypeptide consisting of 26- or 52-heptad repeats (yeast/human), Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Figure 1.2) (85). Different PTMs have been found on CTD, such as phosphorylation, glycosylation, arginine methylation, ubiquitination, and proline isomerization (86). Despite its repetitive nature, the uniformity of the post-translational modification profile remains uncertain (85). CTD with different PTMs mediates recruitment of different protein factors (85, 87, 88). Phosphorylation is an important CTD PTM that has been intensively studied. Soon after initiation, RNAPII CTD is phosphorylated at Ser5 (Ser5P) and Ser7 (Ser7P) by a kinase of TFIIH, Kin28/CDK7 (yeast/human). CTD Ser5P recruits the RNA 5' capping machinery (89), and also recruits Nrd1 early in transcription for the NNS pathway to function in promoter-proximal regions (66, 67). CTD Ser5 is then dephosphorylated by a phosphatase, Ssu72 when RNAPII is released from promoter proximal pause and transitions into productive elongation (90, 91). CTD Ser2 phosphorylation (Ser2P) facilitates productive RNAPII elongation, and at the poly(A) site promotes efficient recruitment of transcription termination factors and RNA 3' processing factors (92). CTD Thr4P persists through the gene body, and its dephosphorylation is correlated with transcription termination at the poly(A) site (93). Ser7P is required for recruiting the integrator complex in metazoans (94, 95). Yeast and mammalian cells have similar RNAPII CTD phosphorylation patterns in transcription cycles except for Tyr1 (96). Tyr1P is enriched in the coding regions in yeast, which is thought to prevent premature transcription termination by antagonizing CTD Ser2P-mediated recruitment of transcription processing and termination

factors (97), and thus its dephosphorylation at the poly(A) site is required for recruitment of CPF component Pcf11 for transcription termination (30, 56, 98–100). In contrast, the mammalian CTD Tyr1P is enriched toward the 5’ end of genes, and to a lesser extent at 3’ ends. However, Tyr1 residue is critical for regulating 3’ end transcription termination and antisense transcription at 5’end in mammalian cells (101, 102). After transcription termination, RNAPII CTD has to be dephosphorylated before RNAPII can be recycled for another round of transcription (103). Therefore, RNAPII CTD phosphorylation and potentially other less studied PTMs need to be carefully regulated to orchestrate binding of protein factors to facilitate co-transcriptional events.

RNAPII kinetics There is a kinetic competition between RNAP II and the 5’ to 3’ exoribonuclease (XRN2 or Sen1) during transcription termination of different classes of RNAs, such as mRNA or CUTs (61). A slow-moving RNAPII, which could be readily caught up by 5’ to 3’ exoribonuclease, is prone to termination according to the torpedo model, while a fast-moving RNAPII delays transcription termination.

An important regulator of RNAPII kinetics is a highly conserved transcription elongation factor in all three kingdoms of life, 5,6-dichloro-1-b-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF). It consists of Spt4 and Spt5 subunits, and forms an intimate association with RNAPII during transcription (104–106). Spt5 has a C-terminal repeat (CTR) composed of 18 tandem repeats, TPAWNSGSK in yeast (Figure 1.3), which analogous to RNAPII CTD, acts as a scaffolding protein to recruit protein factors involved in transcription elongation (107). During transcription, phosphorylation of Spt5 CTR converts it from a negative transcription elongation factor to a positive one. At promoter proximal sites, unphosphorylated DSIF pauses RNAPII. Positive transcription-elongation factor b (P-TEFb) then phosphorylates Spt5 CTR among other proteins, releasing RNAPII and converting DSIF to a form that

stimulates productive RNAPII transcription elongation (32, 108). Spt5 phosphorylation persists until within 100 bp downstream of the poly(A) site where Spt5 phosphorylation decreases sharply and RNAPII decelerates (16), which is required for proper termination by torpedo mechanism in human cells. In *S. pombe*, Spt5 CTR dephosphorylation dissociates it from RNAPII and allows the recruitment of a transcription termination factor, Seb1 (109, 110).

Epigenetic modifications Epigenetic modifications include PTMs on histone tails, histone variants and DNA modifications (111). Epigenetic modifications impact genome accessibility through altering chromatin structure, or recruiting proteins with effector domains, thereby regulating transcription termination. The N-terminal tail of histone proteins are unstructured (112) and subject to a multitude of posttranslational modifications, forming histone marks (113). Distinct histone marks decorate the transcription unit (114, 115), and are crucial regulators of transcription by recruiting protein factors involved in transcription regulation. For example, H3K4me3 is enriched in the promoters of transcriptionally active genes (116), and recruits transcription factors with reader domains, such as TFIID to facilitate PIC assembly and transcription initiation (117, 118). Furthermore, H3K4Me3 is required for efficient termination by the NNS pathway (119). H3K36me3 is found associated with the gene bodies of actively transcribing genes, and is deposited co-transcriptionally by RNAPII-associated methyltransferase Set2/SETD2 (yeast/human) (120–122). In yeast cells, H3K36me3 recruits histone deacetylase (HDAC) complex Rpd3S to remove transcription-coupled histone acetylation, and thereby to prevent spurious transcription initiation in intragenic regions (44, 123). Histone marks also regulate transcription termination as represented by H3K9me2 that is enriched near the 3’ end of genes (124). R-loop formation near the transcription termination sites recruits histone methyltransferase G9a that deposits H3K9me2. H3K9me2-recruited heterochromatin 1ɣ (HP1ɣ)

promotes formation of heterochromatin and therefore reinforces Pol II pausing for efficient transcription termination by torpedo mechanism.

TRANSCRIPTION TERMINATION BY THE PTW/PP1 COMPLEX

PTMs on proteins regulate protein localization, stability, or protein-protein interactions. Phosphorylation is a major PTM, and plays a critical regulatory role for over 70% of eukaryotic cellular proteins (125). As described above, a tight regulation of phosphorylation on RNAPII CTD and Spt5 CTR is crucial for proper transcription termination, and their phosphorylation status is determined by the interplay between phosphatases and kinases. One phosphatase that has been receiving increasing attention that targets RNAPII CTD and Spt5 is protein phosphatase 1 (PP1). PP1 is found in association with PNUTS, TOX4 and Wdr82 in mammalian cells to form the PTW/PP1 complex. The PTW/PP1 complex interacts with RNAPII and components of 3' end processing machinery, suggesting its role in transcription (126, 127). Indeed, the PTW/PP1 complex regulates transcription termination at both 5' and 3' ends of genes in mammalian cells. Individual depletion of the PTW/PP1 complex components led to transcription termination defects, causing extended transcripts at 3' end and increased antisense transcription from bidirectional promoters (128, 129). At the 3' end, PP1-PNUTS-mediated dephosphorylation of Spt5 decelerates RNAPII, thereby facilitating efficient transcription termination by torpedo mechanism (Figure 1.4) (130). How the complex regulates antisense transcription is unclear, but the known target Spt5 has been implicated in repressing antisense transcription initiation in *S. pombe* (131, 132), hinting that a similar mechanism utilized at the 3' end could also be used at 5' end of genes.

The PNUTS-Wdr82-PP1 complex is also conserved in yeast cells. In *S. pombe*, there are two PP1 isoforms, Dis2 and Sds21. Dis2 associates with Ppn1 (PNUTS homolog) and Swd2.2 (Wdr82 homolog) within the CPF complex, and regulates transcription via dephosphorylation of RNAPII CTD Thr4P and Spt5 CTR (109, 110). Cdk9-mediated phosphorylation of Spt5 CTR accelerates RNAPII elongation, while Dis2-mediated dephosphorylation of Spt5 at the poly(A) site dissociates Spt5 and decelerates RNAPII, making it prone to transcription termination, similar to mammalian cells (16, 133, 134). The biological significance of Thr4P dephosphorylation by Dis2, however, is not clear.

The budding yeast genome encodes a single PP1 homolog, Glc7, which associates with Ref2 (PNUTS ortholog) (16, 135, 136), Swd2 (Wdr82 homolog), and other proteins, forming a subcomplex within the CPF, the APT complex (137). Glc7 regulates transcription termination of protein coding genes via dephosphorylation of RNAPII CTD Ty1P, thereby facilitating Ser2P-mediated recruitment of transcription termination factors, Rtt103 and Pcf11 (138). In addition, a recent research showed that the APT subcomplex promotes Pol II stalk-to-stalk homodimerization through CTD dephosphorylation (139), making Pol II unable to bind to transcription elongation factor and incompetent to transcription termination.

The following subsections will be dedicated to the PTW/PP1 complex components.

PP1

More than 420 serine/threonine kinases, which account for 98.2% of all phosphorylation events, are encoded by over 400 genes in the human genome; only fewer than 40 genes are responsible for encoding serine/threonine phosphatases (PSPs) (135, 140). The imbalance in the number of protein kinases versus phosphatases is evolutionarily conserved in yeast, plant and mammalian cells (141–143). PSPs are classified into various sub-families, including

phosphoprotein phosphatases (PPPs), metal-dependent protein phosphatases (PPMs) and aspartate-based phosphatases, with the largest one being the PPP family. The PPP family includes PP1, PP2A, PP2B, PP3, PP4, PP5, PP6, and PP7.

PP1 is a major serine/threonine phosphatase involved in a wide variety of cellular activities. It is estimated to catalyze one third of all dephosphorylation events in eukaryotic cells (144–146). Due to its heavy workload, PP1 is an abundant protein and ubiquitously expressed across tissues; however, PP1 is presumably always in association with other proteins, because it has little substrate specificity and has to be tightly regulated *in vivo* (147, 148). PP1 binds to over 200 confirmed PP1-interacting proteins (PIPs), forming highly specific holoenzymes in mammalian cells (135). PIPs regulate PP1 through multiple mechanisms, including being a targeting subunit, and substrate specifier or inhibitor (149). The PIP-PP1 holoenzymes are tailored to carry out specific functions in the cells, allowing PP1 to perform diverse regulatory functions.

PIPs have low overall sequence or structural conservation, and it is almost impossible to identify PP1 interacting proteins by sequence alignment alone. PIPs usually have extensive contact with PP1 using a combination of short linear motifs (SLiMs) of ~5-8 amino acids long, such as MyPhoNE, SILK, and ϕϕ (147). The SLiMs are short and degenerate sequences found in unstructured regions lacking bulky hydrophobic residues that could form a core. Instead, these regions are enriched in charged residues and short side chains (140, 150). The unstructured nature of PP1-interacting sequences allows PP1-interacting proteins to form novel and unexpected interactions with PP1, and the combined utilization of multiple SLiMs renders PIPs high affinity to PP1. The significance of each motif in PP1-PIP interaction, however, is interactor dependent (151), as mutation of one motif on a particular PP1 interactor could have a much more

severe effect on PP1 binding than on another PP1 interactor. Lastly, binding of PP1-interacting proteins to PP1 usually leads to only subtle changes in PP1 conformation, involving slight changes in loop conformation or orientation of a side chain, but affects PP1 substrate selectivity significantly.

The most well-known and widespread SLiM of PIPs is the RVxF motif found on 70% of PIPs (152, 153). This motif is typically defined by the consensus sequence [K/R][K/R][V/I][x][F/W], where x can be any residue except for Phe, Ile, Met, Tyr, Asp, or Pro (151, 152, 154). The RVxF motif-binding site is characterized by two deep hydrophobic pockets formed by residues on PP1. The hydrophobic pockets bury the “V” and “F” of the RVxF motif. The two hydrophobic residues at position 3 and 5, along with position 1, have the most conserved residues (151). The side chain of the first arginine is involved in hydrogen bonding with PP1 (135). Position 4 , while excluding F, I, M, Y, D and P, has a preference for R, K, S and T. Phosphorylation of the S or T at position 4 has been shown to play a regulatory role in PP1-PIP association (155). Moreover, N-terminal and C-terminal to the RVxF motif are enriched with basic and acidic residues, respectively, for interaction with electrostatic patches surrounding the RVxF-binding pocket. The RVxF motif acts as a primary anchor for PP1 binding, and doesn't affect PP1 overall structure or substrate specificity. It has been thought that the docking of the RVxF motif increases the local concentration of the PP1 interacting proteins and therefore promotes secondary interactions that dictate PP1 activity (reviewed in 157).

In mammalian genomes, three PP1 genes encode four PP1 isoforms, PP1α, PP1β, and PP1γ1 and PP1γ2, with the latter two being splice variants. Mammalian PP1 isoforms have a sequence identity ranging from 85% to 93%, and share a highly similar catalytic core structure, with only 5 polymorphic sites. Their N- and C-terminal extremities, however, are divergent in

sequence identity and are disordered in structure (140, 156). PNUTS has equal binding affinity to PP1 isoforms, and each one of them can be found associated in the PTW/PP1 complex (15). However, many PIPs have PP1 isoform selectivity, such as MYPT1 (140, 157), Spinophilin (158–161), Repoman and Ki-67 (162). So far, structural models have indicated two mechanisms that drive PP1 isoform selectivity: PP1 C-terminal tail and PP1**β**/ɣ specificity pocket. ASPP2 protein preferentially binds to PP1α by distinguishing the divergent PP1 C-terminal tails (163). The acidic N-Src loop of the SH3 domain in ASPP2 forms favorable electrostatic interactions with the basic residues following the PPII motif in PP1α C-terminal tail, driving isoform selectivity toward PP1α as a result. Ki-67, Repoman and RRP1B achieves PP1 isoform selectivity based on a single polymorphism in the catalytic core regions of PP1 isoforms: amino acid 20, which is a glutamine in PP1α, and an arginine in PP1**β**/ɣ (164). R20 in PP1β/γ forms a salt bridge with Q77 in PP1, exposing a hydrophobic pocket that accommodates the burial of Ser702 in RRP1B (164), or the KiR SLiM in RepoMan and Ki-67 (162). This is not possible with Q20 in PP1α. Therefore, the single amino acid difference within the catalytic core region drives isoform selectivity.

PNUTS

PP1-nuclear targeting subunit (PNUTS) was first discovered in a yeast two-hybrid screen with PP1α as a bait protein (165). As the name suggests, PNUTS targets PP1 to the nucleus (166). It is ubiquitously expressed across different tissues, with the highest expression in testis, brain and intestine (166). Human PNUTS (hPNUTS) is a multidomain protein, including an N-terminal transcription elongation factor S-II (TFS2N) domain implicated in regulating RNAPII transcription initiation and elongation (166), RGG motifs toward its C-terminus, and a putative zinc finger motif at its extreme C-terminus (165–167). PNUTS has been implicated in diverse

regulatory roles in regulating cell proliferation and apoptosis (168–171), in synaptic signal transduction (172), in telomere stability (173), DNA damage response (174), in cell cycle regulation (175) and transcription (128). PNUTS primarily facilitates these processes through PP1-dependent mechanisms, although it remains unclear whether PNUTS-PP1 functions within the PTW/PP1 complex to regulate these processes.

In vitro and *in vivo* studies have shown that PNUTS serves as a scaffolding protein that associates PP1, Wdr82 and Tox4 with distinct regions (15). Multiple SLiMs are combined within a central region (aa 396-424) of hPNUTS to bind to the top and side of PP1 remotely away from the PP1 catalytic center, forming an extensive interaction area with PP1 (136). Wdr82 binding regions on hPNUTS is mapped to aa 418-619, and Tox4 binds to the N-terminus. Consistent with its role as a scaffolding protein, depletion of PNUTS in HEK293 cell causes a complete loss of Tox4 and a significant reduction in Wdr82 protein level.

Wdr82

Wdr82 is a WD40 domain-containing protein. The WD40 repeat consists of about 40 amino acid residues, terminating with a tryptophan-aspartic acid dipeptide. The WD40 repeat forms a blade-like structure, and multiple together, usually seven, form a β-propeller shape domain with protein interacting surfaces on the top, bottom and sides (176). Therefore, a WD40 domain-containing protein usually mediates protein interactions within a large multiprotein complex (177). The WD40 domain can also recognize protein PTMs, such as histone H3K4 methylation or serine/threonine phosphorylation (178–181).

Wdr82 is a member of several multiprotein complexes, including the human Set1a and Set1b histone H3K4 methyltransferase complexes. It specifically recognizes RNAPII CTD Ser5P, and therefore tethers the histone methyltransferase complexes to the promoter regions of

actively transcribing genes where CTD Ser5P is enriched. The yeast Wdr82 homolog, swd2, is also part of the yeast H3K4 methyltransferase, Set1 complex, where Set1 is both the catalytic subunit and a scaffold protein in the complex, and cooperates with Swd2 for RNAPII CTD binding (182–184).

In addition to H3K4 methylation, Wdr82 regulates termination of sense transcription at 3' ends and antisense transcription at 5' end, as a component of the PTW/PP1 complex as described before. The role of Wdr82 in the PTW/PP1 complex has not been elucidated, but perhaps Wdr82 bridges the PTW/PP1 complex with RNAPII by interaction with CTD.

Recent studies have indicated that Wdr82 and ZC3H4, along with other associated proteins, form two distinct complexes involved in regulating transcription termination of non-coding genes such as PROMPTs (185, 186). Zinc finger CCCH-type containing 4 (ZC3H4) was initially identified as a novel interactor of Wdr82, and later shown to associate with RNAPII (15, 187). Little studies have been done on ZC3H4, but its depletion caused an accumulation of non-coding RNAs, especially PROMPTs (185–187). ZC3H4-Wdr82 are components of two different complexes, in both of which ZC3H4 is the scaffold protein. The first complex includes ZC3H4, Wdr82, ARS and nuclear exosome targeting complex (NEXT) (186). Turbo-ID identified PNUTS-PP1 in close proximity to the complex. It is not sure if they have any cross-talk in termination of PROMPTs. In the other complex, ZC3H4 associates with casein kinase 2 (CK2) and Wdr82 at N- and C-terminus, respectively to form the ZWC complex (186). CK2 consists of a catalytic CK2α and regulatory CK2**β** subunits. CK2 is highly conserved in eukaryotes and ubiquitously expressed (188, 189), and has kinase activity toward serine, threonine and tyrosine residues (190, 191). It has been implicated that the ZWC complex is recruited to the promoters

by Wdr82-RNAPII CTD S5P affinity where CK2 phosphorylates Spt5 at N-terminus, which suppresses divergent transcription (186).

Tox4

TOX HMG box family member 4 (TOX4) is a novel member of the HMG (high mobility group)-box protein family. HMG domain is involved in DNA binding (192). This is consistent with the report that TOX4 can recognize and bind to DNA adducts generated by platinum anticancer drugs (193). Its role in the PTW/PP1 complex is unclear, but due to its DNA binding ability, Tox4 has been speculated to target the PTW/PP1 complex to specific genomic loci. Recently, Tox4 has been shown to regulate T cell development in murine cells, and restricts promoter-proximal pausing and RNAPII recycling potentially by PP1-mediated dephosphorylation of RNAPII CTD Ser2P and Ser5P, and Spt5 (129, 194). It was assumed that the PTW/PP1 complex can recruit all isoforms of PP1 (15), and therefore the PTW/PP1 complex does not distinguish PP1 isoforms functionally. Interestingly, it was found that Tox4 preferentially binds to and activates PP1α over PP1**β**/ɣ *in vitro*, and its loss led to a greater impact on the chromatin occupancy of PP1α than PP1**β**/ɣ (129), suggesting Tox4 is involved in PP1 isoform selectivity.

TRYPANOSOMATID LIFE CYCLE

Trypanosomatids are early diverging eukaryotes with a single flagellum that include the genera Leishmania and Trypanosoma (195). They are extracellular or intracellular parasites of animals or plants (196). Parasites of this group that cause devastating diseases in vertebrates or lead to significant agricultural loss are usually transmitted by arthropods (197, 198), including Leishmania and Trypanosoma species, and plant parasites, *Phytomonas*. They have a complex

life cycle, alternating between different host species (199). *Leishmania* spp proliferates in the macrophage lysosomes of various vertebrates as amastigotes, and live in the midgut of sand flies as promastigotes (200). African trypanosomes. *Trypanosoma brucei*, proliferates in the midgut of tsetse flies as procyclic forms and express surface proteins, procyclins. Procyclic form *T. brucei* cells differentiate into metacyclic form when they migrate to the salivary glands of insects, and express different surface proteins, metacyclic variant surface glycoproteins (mVSGs). Through tsetse bites, mammalians can get the parasites that differentiate into the long slender and proliferative bloodstream forms in the mammalian bloodstream and tissue fluids. The bloodstream form *T. brucei* expresses a dense coat of VSG genes on its surface. When reaching high cell density, bloodstream form *T. brucei* then differentiates into non-dividing stumpy form with downregulated transcription and translation (201). The stumpy form *T. brucei* can be picked up by flies during a blood meal, and differentiates to the procyclic form, closing the cycle (202). To undergo differentiations and accommodate different environments, the parasites must undergo significant metabolic reprogramming, and express distinct surface proteins in different life stages. Therefore, a tight control over gene expression is critical.

TRYPANOSOMATID GENOMIC ORGANIZATION

In trypanosomatids, tens to hundreds of protein coding genes of unrelated functions are organized in tandem arrays, called the polycistronic transcription units (PTUs) that are transcribed constitutively by RNA Pol II (Figure 1.5) (203–206). The African trypanosome genome harbors around 190 PTUs that can be over 100 kb long (203, 207). Transcription initiation and termination occur only at a limited number of sites per chromosome, known as strand switch regions (SSRs). Transcription usually initiates within 5-10 kb regions at ~150 sites

(207, 208), known as transcription start regions (TSRs) or divergent strand switch regions (dSSRs) where the two neighboring PTUs diverge. Transcription usually terminates where two neighboring PTUs converge, known as transcription termination regions (TTRs) or convergent strand switch regions (cSSRs) (209). Transcription also initiates and terminates at head-tail (HT) regions where transcription of an upstream PTU terminates and initiation of a downstream PTU on the same strand occurs (210). Transcription of the PTUs generates long polycistronic transcripts. Individual mRNAs are generated by trans-splicing, which involves adding a capped spliced leader RNA (SLRNA) sequence at the 5' end, and coupled polyadenylation at the 3' end of upstream genes (210–214).

The surface proteins expressed in different life stages of *T. brucei* are transcribed by RNA polymerase I (RNAPI), which is another interesting feature for the trypanosomatid (215). Bloodstream form *T. brucei* cells develop a dense coat of VSG surface proteins. There are ~2500 VSG genes or gene fragments in the *T. brucei* genome, and most of them are found in long subtelomeric gene arrays (207). VSG genes can also be found in 15-20 telomeric bloodstream form expression sites (BESs) (Figure 1.5). A BES contains a promoter for RNAPI, several expression site-associated genes, 70 bp repeats, and a VSG gene immediately followed by the telomere repeat. At a given time, only one BES is allowed to be transcribed by RNAP I (215–217). The monoallelic expression of the VSG gene ensures that the cell surface is covered by a single type of VSG surface protein. The *T. brucei* cells can periodically switch the expression of VSG genes and express a different surface VSG protein, to which the host immune system has not tuned yet, by homologous recombination or activation of a silent BES, a phenomenon known as the antigenic variation (217). The monoallelic expression and antigenic variation allows the

parasites to evade the host immune response and achieve persistent infection, and therefore a tight regulation over gene expression is necessary.

TRANSCRIPTION IN TRYPANOSOMATIDS

Transcriptional regulation of gene expression is critical and widely used in eukaryotes, from yeast to mammalian cells, involving promoter recognition, and regulation over initiation and elongation (218). Due to the polycistronic nature of the gene arrays, however, transcriptional regulation of gene expression is challenging in trypanosomatids. Moreover, well-defined and canonical cis-acting elements for transcription initiation of PTUs have been lacking in trypanosomatids, although recent research suggests specific sequence drives transcription initiation in *T. brucei* (205, 219). Therefore, post-transcriptional regulation, instead of transcriptional regulation, has been considered to play a major role in expression of individual genes in trypanosomatids, including mRNA stability and translation efficiency (220).

Little is understood about transcription termination in trypanosomatids, and the polycistronic genome organization challenges the canonical model of transcription termination. Transcription termination of RNAPII is coupled to transcript cleavage and polyadenylation in opisthokonts. Cleavage of the nascent RNAs at the poly(A) site provides an entry site for a 5’ to 3’ exoribonuclease Rat1/Xrn2, facilitating transcription termination by torpedo model (221). Rat1/Xrn2 homolog also has a putative homolog in *T. brucei*, XRND (222). XRND has a conserved N-terminal exonuclease domain, and is predominantly localized in the nucleus. Although XRND is essential for cell viability, its detailed function has not been explored yet. However, this mode of transcription termination is obviously not compatible with polycistronic transcription, since if 3’ end formation is linked to transcription termination, this would result in

premature transcription termination within a PTU, leaving hundreds of downstream genes untranscribed. In trypanosomatids, the canonical polyadenylation signal, AAUAAA, and other cis-acting elements required for cleavage and polyadenylation are not conserved. It appears that the polyadenylation site of the upstream gene is determined by a polypyrimidine tract upstream a trans-splice site of the downstream gene (210, 211, 223). Tandem affinity purification of the largest subunit of CPSF complex has revealed that the majority of the components in the 3' end processing factors are conserved in *T. brucei* genome (224, 225). Depletion of the 3' end processing complex subunits led to reduced polyA tail length and splicing defects as revealed by accumulation of SLRNAs, indicating a general coupling of polyadenylation and trans-splicing, but uncoupling of 3' end formation and transcription termination.

What is more puzzling is the divergence in protein sequences of some key players in transcription. RNAPII CTD is a critical player in the RNAPII transcription cycle (85, 87, 226–228). All well-studied model organisms have a canonical CTD, which consists of repetitive heptapeptide sequences that are subject to PTMs, especially phosphorylation, thereby signaling and coordinating co-transcription activities (229). Pol II phosphorylation profiles are similar in fission yeast and mammalian cells (93, 109, 230, 231), indicating a high conservation of Pol CTD functionality. However, the trypanosomatids have a non-canonical CTD of about 300 amino acids long, which lacks any repetitive sequences and has little sequence identity to the CTDs of other organisms (232–235). The *T. brucei* RNAPII CTD is 284 amino acids long, and is essential for cell survival (236). The *T. brucei* RNAPII, like the canonical CTD, is rich in serine and phosphorylated (236, 237). It has been shown that the *T. brucei* noncanonical CTD is required for directing RNAPII to TSSs (238). Other than that, its biological function is largely unknown. Furthermore, the *T.brucei* Spt5 lacks the repetitive CTR that can be phosphorylated

and serves as a binding platform, although multiple phosphorylation sites are found on *T. brucei* Spt5 (Figure 1.3) (237).

In model organisms such as human cells and yeast cells, Antisense transcripts derived from antisense promoters constitutes a major fraction of non-coding RNAs generated by pervasive transcription of RNAPII (74, 239, 240). This is also observed in *T. brucei* genome where 99 of the 148 annotated *T. brucei* RNAPII promoters are bidirectional (241). How the antisense transcripts generated from bidirectional promoters are terminated to ensure sense transcripts predominate is not well-understood.

CHROMATIN MODIFICATIONS IN KINETOPLASTIDS

Although cis-regulatory elements are lacking, chromatin modifications have been recognized as important regulators of transcription in trypanosomatids. Through altering chromatin structure, chromatin modifications impact genome accessibility to limit transcription. Alternatively, chromatin modifications recruit proteins with specific effector domains to directly regulate transcription.

Chromatin modifications include histone PTMs, histone variants, and DNA modifications among others. A nucleosome is the fundamental unit of chromatin. The core of the nucleosome usually consists of histone H2A, H2B, H3 and H4, forming an octamer around which 147 bp of DNA is wrapped (112, 242). The primary sequence of histones is typically conserved across eukaryotes, but trypanosome histones are divergent in sequence from their counterparts in other organisms (21). In addition, the trypanosomatids have a variant for each core histone (H2A.Z, H2B.V, H3.V and H4.V). Many histone PTMs have been detected in *T. brucei* or *T. cruzi*.

However, the sequence divergence of histone proteins makes it difficult to compare histone PTMs between trypanosomatids and other eukaryotes.

Base J (β-D-Glucopyranosyloxymethyluracil) is a modified thymine residue found exclusively in kinetoplastids and some unicellular flagellates (243). Base J is synthesized in two steps: first, specific thymine residues are hydroxylated to produce hydroxymethyldeoxyuridine (HOMedU) by thymidine hydroxylases, either J-binding protein 1 or 2 (JBP1 or JBP2) (244); a glucose is then transferred to HOMedU by a glucosyltransferase (JGT) (245, 246).

TSRs and TTRs are enriched with distinct histone variants and epigenetic modifications (Figure 1.5) (204). They have been shown to play an important regulatory role in transcription. TSRs are characterized by the presence of multiple epigenetic marks, including H3K4me3, H4K10ac, H2AZ and H2BV. (205). H2AZ and H4 in the TSRs are acetylated by HAT1 and HAT2, respectively. Depletion of HAT2 reduces H2AZ deposition, and also shifts the sites of transcription initiation, while depletion of HAT1 reduces the total mRNA levels by 50% (247).

TTRs feature three epigenetic marks: base J, H3V and H4V, which are crucial regulators in transcription termination at TTRs in *T. brucei* and *Leishmania* spp (22–24, 248, 249). Base J is required for proper transcription termination in *Leishmania* (22, 248). In *T. brucei*, base J, H3V and H4V independently regulate transcription termination (22–24, 249). It has been found that base J and H3V are localized upstream of genes at the ends of some PTUs, thereby promoting premature transcription termination prior to the ends of these PTUs, and silencing the downstream genes as a result. Therefore, it has been proposed that transcription termination serves as a transcriptional regulation mechanism in trypanosomatids (22, 250). In *T. cruzi*, however, base J regulates transcription initiation (22) but not termination by RNAPII (251). Loss of base J in *T. cruzi* leads to increased RNAPII transcription and gene expression.

Furthermore, base J and H3V are enriched at subtelomeric and telomeric regions where BES and MES VSG genes are located (243). Depletion of base J, H3V, and H4V that is not found in subtelomeric or telomeric regions leads to loss of monoallelic expression and de-repression of BES and MES VSGs (216, 249). Hi-C and ChIP-seq have indicated that the silent and non-transcribed subtelomeric regions of the *T. brucei* genome are more compact than the transcribed core region, and silent bloodstream form VSG ESs are clustered together, indicating that a compact chromatin structure limits readthrough transcription into subtelomeric and telomeric regions in *T. brucei* (216). Epigenetic markers H3V and H4V are critical in regulating genome architecture, and their loss increases chromatin accessibility at both subtelomeric and telomeric regions, leading to VSG gene de-repression (217). Therefore, these epigenetic marks may define transcription termination sites by regulating chromatin accessibility.

THE PJW/PP1 COMPLEX

Besides histone marks, 15 proteins with chromatin regulatory domains are enriched in TSRs (241). Only a few of them have been studied, and whether/how they regulate transcription is unknown. J-binding protein 3 (JBP3) is a novel protein with base J-binding domain (252). It is associated with three other proteins, PNUTS, Wdr82 and PP1, forming the PJW/PP1 complex, where PP1 is key phosphatase subunit, PNUTS is a RVxF motif-containing protein, and Wdr82 is a WD40-repeat containing protein (252). The trypanosomatid PJW/PP1 complex is highly reminiscent of the human PTW/PP1 complex, and later characterization showed that it is involved in regulating transcription termination. Depletion of individual components in the PJW/PP1 complex from *T. brucei* causes accumulated antisense transcripts from bidirectional promoters, and readthrough transcription at TTRs, subtelomeric and telomeric regions, leading to

de-repression of genes that are located in these regions and are normally transcribed at a low level (252–254). Further characterization shows that the complex exhibits a PP1-mediated phosphatase activity toward Rpb1 CTD (254), suggesting a functional conservation between the PTW/PP1 and the PJW/PP1 complex. Therefore, it was proposed that the PJW/PP1 complex regulates transcription termination in trypanosomatids by PP1-mediated dephosphorylation of Rpb1 CTD and Spt5 at TTRs, slowing down the elongating RNAPII and facilitating transcription termination by torpedo mechanism (252).

The *Leishmania* genome has eight PP1 homologs that can be grouped into five clades, clade A to E, based on phylogenetic analysis (252). Unlike the PTW/PP1 complex where every PP1 isoforms can be associated into the complex (15), only PP1-8 of clade E is associated with the PJW/PP1 complex in *L. tarentolae* or *L. major*, indicating an isoform selectivity by PNUTS (252, 254). However, all eight PP1 homologs contain the RVxF binding pocket (255), so how PP1-8e confers isoform specificity to PNUTS was unknown in *Leishmania* spp. Interestingly, *T. brucei* genome also has eight PP1 homologs, but none of them belong to clades E, and LtPP1-8e lacks a *T. brucei* homolog (252). Consistent with the preference for PP1 of clade E, PNUTS purification from *T. brucei* cells recovered all the components of the complex except any PP1 homologs. The results suggest that the *T. brucei* PJW complex may form a weak or transient association with a particular PP1 homolog of another clade that cannot withstand the purification process. Alternatively, the *T. brucei* PJW complex may regulate transcription in a PP1-independent mechanism.

JBP3 is a member of two other complexes in addition to the PJW/PP1 complex. JBP3 immunoprecipitation revealed that JBP3 also interacts with SPARC complex implicated in chromatin remodeling or modification (discussed below), and with components of the

polymerase associated factor 1 complex (Paf1C) (253) that is involved extensively in regulating transcription in eukaryotes (256).

OTHER PROTEINS REGULATING TRANSCRIPTION TERMINATION

RBP33 is a nuclear localized protein essential for cell growth, and is a trypanosomatid-specific protein (257, 258). It is an RNA-binding protein. Its depletion leads to transcription termination defects at TTRs, and transcriptional readthrough into the neighboring PTUs, leading to dsRNA and thus RNAi silencing of protein coding genes at the end of PTUs (258). Furthermore, its depletion leads to de-repression of certain genes from the telomeric regions, such as MES VSGs. Therefore, RBP33 regulates transcription at similar genomic loci as the PJW/PP1 complex. It is not known if there is any crosstalk between them. However, RPB33 interacts with several splicing factors, and is therefore implicated in inhibiting trans-splicing of pervasive transcripts, which are then targeted for degradation by exosome.

SET27 promoter-associated regulatory complex (SPARC) consists of 6 core subunits, including SET27, CRD1, CSD1, PHD6, PBP1 and PWWP1 (259). SPARC complex is enriched in a narrow window within TSRs just upstream of transcription initiation and regulates transcription initiation (259). Multiple components (CRD1, PHD6 and PWWP1) show interaction with RNAPII subunits by immunoprecipitation, suggesting their roles in regulating transcription. Therefore, not surprisingly, deletion of SET27 leads to a shift of transcription initiation upstream and increases antisense transcription from bidirectional promoters, indicating that SPARC complex is required for accurate transcription initiation and transcription directionality. In addition, SET27 deletion leads to de-repression of subtelomeric genes, including VSG genes and expression site-associated genes (ESAGs). How the SPARC complex

regulates transcription remains to be investigated. However, multiple SPARC components contain histone writer and reader domains, such as SET methyltransferase domain, chromodomain and PHD finger histone methylation reader domain, and associate with histone or histone variants by immunoprecipitation, suggesting that the complex could be recruited to genomic loci by binding to specific histone modification, or play a role in regulating chromatin structure. Interestingly, all SPARC components except PHD6 associates with JBP3, suggesting there might be functional crosstalk between the PJW/PP1 complex and SPARC complex, although the detailed mechanism awaits further investigation. In addition, PHD6 and PWWP1 interact with components of the FACT complex components (POB3 and SPT16), which plays a role in VSG repression in *T. brucei* (260).

CONCLUSION AND DISSERTATION OVERVIEW

Transcriptional regulation in trypanosomatids remains poorly understood, and recent advancements in chromatin modifications have shed light on the role of base J in transcription termination and VSG monoallelic expression. Nonetheless, a fundamental question remained unanswered: how does base J exert its regulatory function on transcription termination? The characterization of a base J-containing protein complex, the PJW/PP1 complex, offers valuable insights into this question. We proposed that as a scaffolding protein, PNUTS allows independent association of PP1 and Wdr82/JBP3 to form the PJW/PP1 complex which is recruited to base J-enriched TSRs and TTRs and regulates transcription termination by dephosphorylation of RNAPII CTD and Spt5 (Figure 1.6).

Appendix A presents the initial identification and characterization of the trypanosomatid PJW/PP1 complex. To understand base J biogenesis, JGT-associated proteins were identified by

tandem affinity purification and mass spectrometry from *L. tarentolae*. This leads to the identification of the PJW/PP1 complex that contains PP1, an RVxF motif-containing protein (PNUTS), Wdr82 and a J-binding protein (JBP3). A similar complex, the PJW complex, is also found in *T. brucei* cells. Analysis of the PJW complex shows that it regulates transcription termination, and its loss leads to extended readthrough transcripts, accumulated antisense transcripts from bidirectional promoters, and de-repression of subtelomeric and telomeric genes.

Chapter 2 focuses on the role of PP1, and suggests RNAPII CTD as the substrates of the PJW/PP1 complex. Due to the lack of PP1 from the PJW complex in *T. brucei*, we were not able to address its role by RNAi. Therefore, we directly interrogated the role of PP1 by conditionally knocking it out in *L. major*. Its loss leads to readthrough transcription. Furthermore, we found that the PJW/PP1 complex shows PP1-dependent specific phosphatase activity toward RNAPII CTD *in vitro*, further confirming our proposal that PP1 is a vital component of the complex and RNAPII CTD dephosphorylation underlies the mechanism by which the PJW/PP1 complex regulates transcription termination.

Chapter 3 delves into the role of PNUTS. By structural modeling and biochemical assays, we show that *L. tarentolae* PNUTS binds to PP1-8 with a combination of SLIMs highly similar to human PNUTS, therefore confirming PNUTS as a PP1-interacting protein. Furthermore, we show that the N- and C-termini, and specific sequences within the PP1-8 catalytic core confer isoform specificity to PNUTS. Lastly, we explored the role of PNUTS as a scaffolding protein in the complex, and show that PNUTS regulate the complex stability.

REFERENCE

1. Porrua, O., and Libri, D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202

2. Greger, I. H., and Proudfoot, N. J. (1998) Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. *EMBO J.* **17**, 4771–4779

3. Shearwin, K. E., Callen, B. P., and Egan, J. B. (2005) Transcriptional interference--a crash course. *Trends Genet.* **21**, 339–345

4. Gullerova, M., and Proudfoot, N. J. (2008) Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell*. **132**, 983–995

5. Prescott, E. M., and Proudfoot, N. J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8796–8801

6. Hobson, D. J., Wei, W., Steinmetz, L. M., and Svejstrup, J. Q. (2012) RNA polymerase II collision interrupts convergent transcription. *Mol. Cell*. **48**, 365–374

7. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., and Steinmetz, L. M. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5320–5325

8. Core, L. J., Waterfall, J. J., and Lis, J. T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. **322**, 1845–1848

9. Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., and Jacquier, A. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*. **457**, 1038–1042

10. Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008) Divergent transcription from active promoters. *Science*. **322**, 1849–1851

11. Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L. M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*. **457**, 1033–1037

12. Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*. **155**, 1075–1087

13. Proudfoot, N. J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* **14**, 105–110

14. Connelly, S., and Manley, J. L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* **2**, 440–452

15. Lee, J.-H., You, J., Dobrota, E., and Skalnik, D. G. (2010) Identification and characterization of a novel human PP1 phosphatase complex. *J. Biol. Chem.* **285**, 24466–24476

16. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D. L. (2019) Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "sitting duck torpedo" mechanism. *Mol. Cell*. **76**, 896-908.e4

17. Horn, D. (2022) A profile of research on the parasitic trypanosomatids and the diseases they cause. *PLoS Negl. Trop. Dis.* **16**, e0010040

18. Simpson, A. G. B., Gill, E. E., Callahan, H. A., Litaker, R. W., and Roger, A. J. (2004) Early evolution within kinetoplastids (euglenozoa), and the late emergence of trypanosomatids. *Protist*. **155**, 407–422

19. Clayton, C. E. (2016) Gene expression in Kinetoplastids. *Curr. Opin. Microbiol.* **32**, 46–51

20. Günzl, A. (2010) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? *Eukaryot. Cell*. **9**, 1159–1170

21. Martínez-Calvillo, S., Romero-Meza, G., Vizuet-de-Rueda, J. C., Florencio-Martínez, L. E., Manning-Cela, R., and Nepomuceno-Mejía, T. (2018) Epigenetic Regulation of Transcription in Trypanosomatid Protozoa. *Curr. Genomics*. **19**, 140–149

22. Reynolds, D., Cliffe, L., Förstner, K. U., Hon, C.-C., Siegel, T. N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic Acids Res.* **42**, 9717–9729

23. Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H.-S. (2016) Base J and H3.V Regulate Transcriptional Termination in Trypanosoma brucei. *PLoS Genet.* **12**, e1005762

24. Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., and Sabatini, R. (2016) Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma brucei. *PLoS Genet.* **12**, e1005758

25. Hsin, J.-P., and Manley, J. L. (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137

26. Osman, S., and Cramer, P. (2020) Structural Biology of RNA Polymerase II Transcription: 20 Years On. *Annu. Rev. Cell Dev. Biol.* **36**, 1–34

27. Lee, T. I., and Young, R. A. (2000) Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137

28. Espinosa, J. M. (2010) The meaning of pausing. *Mol. Cell*. **40**, 507–508

29. Nechaev, S., and Adelman, K. (2011) Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim. Biophys. Acta*. **1809**, 34–45

30. Kuehner, J. N., Pearson, E. L., and Moore, C. (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* **12**, 283–294

31. Farnung, L., and Vos, S. M. (2022) Assembly of RNA polymerase II transcription initiation complexes. *Curr. Opin. Struct. Biol.* **73**, 102335

32. Adelman, K., and Lis, J. T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731

33. Pelechano, V., and Steinmetz, L. M. (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893

34. Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504

35. Wagner, E. G., and Simons, R. W. (1994) Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**, 713–742

36. Brown, T., Howe, F. S., Murray, S. C., Wouters, M., Lorenz, P., Seward, E., Rata, S., Angel, A., and Mellor, J. (2018) Antisense transcription-dependent chromatin signature modulates sense transcript dynamics. *Mol. Syst. Biol.* **14**, e8007

37. Vanhée-Brossollet, C., and Vaquero, C. (1998) Do natural antisense transcripts make sense in eukaryotes? *Gene*. **211**, 1–9

38. Ozsolak, F., Kapranov, P., Foissac, S., Kim, S. W., Fishilevich, E., Monaghan, A. P., John, B., and Milos, P. M. (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*. **143**, 1018–1029

39. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., Thomas, M., Davis, C. A., Shiekhattar, R., Gingeras, T. R., Hubbard, T. J., Notredame, C., Harrow, J., and Guigó, R. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789

40. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engström, P. G., Mizuno, Y., Faghihi, M. A., Sandelin, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Project Core Group), and FANTOM Consortium (2005) Antisense transcription in the mammalian transcriptome. *Science*. **309**, 1564–1566

41. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W. (2008) The antisense transcriptomes of human cells. *Science*. **322**, 1855–1857

42. Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., Horlings, H. M., Shah, N., Umbricht, C., Wang, P., Wang, Y., Kong, B., Langerød, A., Børresen-Dale, A.-L., Kim, S. K., van de Vijver, M., Sukumar, S., Whitfield, M. L., Kellis, M., Xiong, Y., Wong, D. J., and Chang, H. Y. (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629

43. Kim, T., Xu, Z., Clauder-Münster, S., Steinmetz, L. M., and Buratowski, S. (2012) Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell*. **150**, 1158–1169

44. Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*. **123**, 581–592

45. Kaplan, C. D., Laprade, L., and Winston, F. (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science*. **301**, 1096–1099

46. Whitehouse, I., Rando, O. J., Delrow, J., and Tsukiyama, T. (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature*. **450**, 1031–1035

47. Davis, C. A., and Ares, M., Jr (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3262–3267

48. Vance, K. W., and Ponting, C. P. (2014) Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.* **30**, 348–355

49. Angrand, P.-O., Vennin, C., Le Bourhis, X., and Adriaenssens, E. (2015) The role of long non-coding RNAs in genome formatting and expression. *Front. Genet.* **6**, 165

50. Wolfe, K. H. (2006) Comparative genomics and genome evolution in yeasts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 403–412

51. Eaton, J. D., and West, S. (2020) Termination of Transcription by RNA Polymerase II: BOOM! *Trends Genet.* **36**, 664–675

52. Kumar, A., Clerici, M., Muckenfuss, L. M., Passmore, L. A., and Jinek, M. (2019) Mechanistic insights into mRNA 3′-end processing. *Curr. Opin. Struct. Biol.* **59**, 143–150

53. Sun, Y., Hamilton, K., and Tong, L. (2020) Recent molecular insights into canonical pre-mRNA 3'-end processing. *Transcription*. **11**, 83–96

54. Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., and Tong, L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*. **444**, 953–956

55. Hill, C. H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M., Skehel, M., and Passmore, L. A. (2019) Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. *Mol. Cell*. **73**, 1217-1231.e11

56. Xiang, K., Tong, L., and Manley, J. L. (2014) Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Mol. Cell. Biol.* **34**, 1894–1910

57. Logan, J., Falck-Pedersen, E., Darnell, J. E., Jr, and Shenk, T. (1987) 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.* **84**, 8306–8310

58. Zhang, H., Rigo, F., and Martinson, H. G. (2015) Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol. Cell*. **59**, 437–448

59. Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*. **432**, 517–522

60. Eaton, J. D., Davidson, L., Bauer, D. L. V., Natsume, T., Kanemaki, M. T., and West, S. (2018) Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* **32**, 127–139

61. Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., Nguyen, T., Karp, S., and Bentley, D. L. (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell*. **60**, 256–267

62. West, S., Gromak, N., and Proudfoot, N. J. (2004) Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*. **432**, 522–525

63. Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*. **413**, 327–331

64. Heo, D., Yoo, I., Kong, J., Lidschreiber, M., and Mayer, A. (2013) The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing …. *Journal of Biological*

65. Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **26**, 1891–1896

66. Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., and Libri, D. (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell*. **55**, 467–481

67. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15**, 795–804

68. Creamer, T. J., Darby, M. M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S. J., and Corden, J. L. (2011) Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* **7**, e1002329

69. Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R. K., Stefl, R., and Libri, D. (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31**, 3935–3948

70. Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011) The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* **30**, 1790–1803

71. Arigo, J. T., Eyler, D. E., Carroll, K. L., and Corden, J. L. (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell*. **23**, 841–851

72. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., and Jacquier, A. (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*. **121**, 725–737

73. Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A., and Jensen, T. H. (2011) Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell*. **43**, 624–637

74. Lykke-Andersen, S., Žumer, K., Molska, E. Š., Rouvière, J. O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T. H. (2021) Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell*. **81**, 514-529.e6

75. Skaar, J. R., Ferris, A. L., Wu, X., Saraf, A., Khanna, K. K., Florens, L., Washburn, M. P., Hughes, S. H., and Pagano, M. (2015) The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* **25**, 288–305

76. Baillat, D., Hakimi, M.-A., Näär, A. M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*. **123**, 265–276

77. Andersen, P. R., Domanski, M., Kristiansen, M. S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., Sandberg, R., Hyman, A., LaCava, J., Rout, M. P., Andersen, J. S., Bertrand, E., and Jensen, T. H. (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* **20**, 1367–1376

78. Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J. S., Sandelin, A., and Jensen, T. H. (2016) Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell*. **64**, 520–533

79. Liu, X., Guo, Z., Han, J., Peng, B., Zhang, B., Li, H., Hu, X., David, C. J., and Chen, M. (2022) The PAF1 complex promotes 3’ processing of pervasive transcripts. *Cell Rep.* **38**, 110519

80. Ntini, E., Järvelin, A. I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P. R., Andersen, P. K., Preker, P., Valen, E., Zhao, X., Pelechano, V., Steinmetz, L. M., Sandelin, A., and Jensen, T. H. (2013) Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928

81. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B., and Sharp, P. A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*. **499**, 360–363

82. Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R., 3rd, Frank, J., and Manley, J. L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*. **33**, 365–376

83. Richard, P., and Manley, J. L. (2013) How bidirectional becomes unidirectional. *Nat. Struct. Mol. Biol.* **20**, 1022–1024

84. Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*. **468**, 664–668

85. Eick, D., and Geyer, M. (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490

86. Zaborowska, J., Egloff, S., and Murphy, S. (2016) The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777

87. Corden, J. L. (2013) RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem. Rev.* **113**, 8423–8455

88. Jeronimo, C., Bataille, A. R., and Robert, F. (2013) The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem. Rev.* **113**, 8491–8522

89. Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14**, 2435–2440

90. Bataille, A. R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S. D., and Robert, F. (2012) A universal RNA polymerase II CTD cycle is

orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol. Cell*. **45**, 158–170

91. Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M. M., Manley, J. L., and Tong, L. (2010) Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex. *Nature*. **467**, 729–733

92. Ahn, S. H., Kim, M., and Buratowski, S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell*. **13**, 67–76

93. Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N. J. (2017) Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell*. **65**, 25–38

94. Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*. **318**, 1777–1779

95. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., and Murphy, S. (2012) Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol. Cell*. **45**, 111–122

96. Rodríguez-Molina, J. B., West, S., and Passmore, L. A. (2023) Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol. Cell*. **83**, 404–415

97. Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*. **336**, 1723–1725

98. Richard, P., and Manley, J. L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.* **23**, 1247–1269

99. Mischo, H. E., and Proudfoot, N. J. (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta*. **1829**, 174–185

100. Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., and Cramer, P. (2014) Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol. Cell*. **55**, 745–757

101. Hsin, J.-P., Li, W., Hoque, M., Tian, B., and Manley, J. L. (2014) RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife*. **3**, e02112

102. Shah, N., Maqbool, M. A., Yahia, Y., El Aabidine, A. Z., Esnault, C., Forné, I., Decker, T.-M., Martin, D., Schüller, R., Krebs, S., Blum, H., Imhof, A., Eick, D., and Andrau, J.-C. (2018) Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol. Cell*. **69**, 48-61.e6

103. Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460

104. Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.-I. (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*. **357**, 921–924

105. Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018) Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*. **560**, 607–612

106. Vos, S. M., Farnung, L., Urlaub, H., and Cramer, P. (2018) Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*. **560**, 601–606

107. Hartzog, G. A., and Fu, J. (2013) The Spt4–Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. **1829**, 105–115

108. Decker, T.-M. (2021) Mechanisms of Transcription Elongation Factor DSIF (Spt4–Spt5). *J. Mol. Biol.* **433**, 166657

109. Kecman, T., Kuś, K., Heo, D.-H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C. V., and Vasiljeva, L. (2018) Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. *Cell Rep.* **25**, 259-269.e5

110. Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., and Fisher, R. P. (2018) A Cdk9–PP1 switch regulates the elongation–termination transition of RNA polymerase II. *Nature*. **558**, 460–464

111. Rothbart, S. B., and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta*. **1839**, 627–643

112. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. **389**, 251–260

113. Allis, C. D., and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500

114. Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K.,

and Young, R. A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*. **122**, 517–527

115. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318

116. ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C. J., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W.-K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss,

C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C.-L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaöz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., NISC Comparative Sequencing Program, Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, Children's Hospital Oakland Research Institute, Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W. H., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N. S., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., de

Bakker, P. I. W., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V. B., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., and de Jong, P. J. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. **447**, 799–816

117. Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., Hughes, S. H., and Roeder, R. G. (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell*. **152**, 1021–1036

118. Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W. M. P., van Schaik, F. M. A., Varier, R. A., Baltissen, M. P. A., Stunnenberg, H. G., Mann, M., and Timmers, H. T. M. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*. **131**, 58–69

119. Terzi, N., Churchman, L. S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011) H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol. Cell. Biol.* **31**, 3569–3583

120. Vakoc, C. R., Sachdeva, M. M., Wang, H., and Blobel, G. A. (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* **26**, 9185–9195

121. Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z.-W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2002) Set2 is a nucleosomal

histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell. Biol.* **22**, 1298–1306

122. Edmunds, J. W., Mahadevan, L. C., and Clayton, A. L. (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420

123. Venkatesh, S., Li, H., Gogol, M. M., and Workman, J. L. (2016) Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat. Commun.* **7**, 13610

124. Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N. J. (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*. **516**, 436–439

125. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3

126. Jerebtsova, M., Klotchenko, S. A., Artamonova, T. O., Ammosova, T., Washington, K., Egorov, V. V., Shaldzhyan, A. A., Sergeeva, M. V., Zatulovskiy, E. A., Temkina, O. A., Petukhov, M. G., Vasin, A. V., Khodorkovskii, M. A., Orlov, Y. N., and Nekhai, S. (2011) Mass spectrometry and biochemical analysis of RNA polymerase II: targeting by protein phosphatase-1. *Mol. Cell. Biochem.* **347**, 79–87

127. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. **149**, 1393–1406

128. Austenaa, L. M. I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B. A. M., de Pretis, S., Piccolo, V., Termanini, A., Prosperini, E., Pelizzola, M., de Laat, W., and Natoli, G. (2015) Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol. Cell*. **60**, 460–474

129. Liu, Z., Wu, A., Wu, Z., Wang, T., Pan, Y., Li, B., Zhang, X., and Yu, M. (2022) TOX4 facilitates promoter-proximal pausing and C-terminal domain dephosphorylation of RNA polymerase II in human cells. *Commun Biol*. **5**, 300

130. Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., Siepel, A., and Kraus, W. L. (2013) Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell*. **50**, 212–222

131. Shetty, A., Kallgren, S. P., Demel, C., Maier, K. C., Spatt, D., Alver, B. H., Cramer, P., Park, P. J., and Winston, F. (2017) Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol. Cell*. **66**, 77-88.e5

132. Sansó, M., Parua, P. K., Pinto, D., Svensson, J. P., Pagé, V., Bitton, D. A., MacKinnon, S., Garcia, P., Hidalgo, E., Bähler, J., Tanny, J. C., and Fisher, R. P. (2020) Cdk9 and H2Bub1 signal to Clr6-CII/Rpd3S to suppress aberrant antisense transcription. *Nucleic Acids Res.* **48**, 7154–7168

133. Eaton, J. D., Francis, L., Davidson, L., and West, S. (2020) A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145

134. Parua, P. K., and Fisher, R. P. (2020) Dissecting the Pol II transcription cycle and derailing cancer with CDK inhibitors. *Nat. Chem. Biol.* **16**, 716–724

135. Shi, Y. (2009) Serine/threonine phosphatases: mechanism through structure. *Cell*. **139**, 468–484

136. Choy, M. S., Hieke, M., Kumar, G. S., Lewis, G. R., Gonzalez-DeWhitt, K. R., Kessler, R. P., Stein, B. J., Hessenberger, M., Nairn, A. C., Peti, W., and Page, R. (2014) Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4097–4102

137. Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Hughes, T., Buratowski, S., Moore, C. L., and Greenblatt, J. (2003) Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3′-ends. *J. Biol. Chem.* **278**, 33000–33010

138. Nedea, E., Nalbant, D., Xia, D., Theoharis, N. T., Suter, B., Richardson, C. J., Tatchell, K., Kislinger, T., Greenblatt, J. F., and Nagy, P. L. (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol. Cell*. **29**, 577–587

139. Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. A. (2014) RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.* **21**, 175–179

140. Carminati, M., Cemre Manav, M., Bellini, D., and Passmore, L. A. (2022) A direct interaction between CPF and Pol II links RNA 3′-end processing to transcription. *bioRxiv*. 10.1101/2022.07.28.501803

141. Peti, W., Nairn, A. C., and Page, R. (2013) Structural basis for protein phosphatase 1 regulation and specificity. *FEBS J.* **280**, 596–611

142. Chen, M. J., Dixon, J. E., and Manning, G. (2017) Genomics and evolution of protein phosphatases. *Sci. Signal.* 10.1126/scisignal.aag1796

143. Schweighofer, A., and Meskiene, I. (2015) Phosphatases in plants. *Methods Mol. Biol.* **1306**, 25–46

144. Smoly, I., Shemesh, N., Ziv-Ukelson, M., Ben-Zvi, A., and Yeger-Lotem, E. (2017) An Asymmetrically Balanced Organization of Kinases versus Phosphatases across Eukaryotes Determines Their Distinct Impacts. *PLoS Comput. Biol.* **13**, e1005221

145. Cohen, P. T. W. (2002) Protein phosphatase 1--targeted in many directions. *J. Cell Sci.* **115**, 241–256

146. Ceulemans, H., and Bollen, M. (2006) A tighter RVxF motif makes a finer Sift. *Chem. Biol.* **13**, 6–8

147. Moorhead, G. B. G., Trinkle-Mulcahy, L., and Ulke-Lemée, A. (2007) Emerging roles of nuclear protein phosphatases. *Nat. Rev. Mol. Cell Biol.* **8**, 234–244

148. Verbinnen, I., Ferreira, M., and Bollen, M. (2017) Correction: Biogenesis and activity regulation of protein phosphatase 1. *Biochem. Soc. Trans.* **45**, 583–584

149. Virshup, D. M., and Shenolikar, S. (2009) From promiscuity to precision: protein phosphatases get a makeover. *Mol. Cell*. **33**, 537–545

150. Bollen, M., Peti, W., Ragusa, M. J., and Beullens, M. (2010) The extended PP1 toolkit: designed to create specificity. *Trends Biochem. Sci.* **35**, 450–458

151. Hurley, T. D., Yang, J., Zhang, L., Goodwin, K. D., Zou, Q., Cortese, M., Dunker, A. K., and DePaoli-Roach, A. A. (2007) Structural basis for regulation of protein phosphatase 1 by inhibitor-2. *J. Biol. Chem.* **282**, 28874–28883

152. Hendrickx, A., Beullens, M., Ceulemans, H., Den Abt, T., Van Eynde, A., Nicolaescu, E., Lesage, B., and Bollen, M. (2009) Docking motif-guided mapping of the interactome of protein phosphatase-1. *Chem. Biol.* **16**, 365–371

153. Wakula, P., Beullens, M., Ceulemans, H., Stalmans, W., and Bollen, M. (2003) Degeneracy and function of the ubiquitous RVXF motif that mediates binding to protein phosphatase-1. *J. Biol. Chem.* **278**, 18817–18823

154. Egloff, M. P., Cohen, P. T., Reinemer, P., and Barford, D. (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J. Mol. Biol.* **254**, 942–959

155. Meiselbach, H., Sticht, H., and Enz, R. (2006) Structural analysis of the protein phosphatase 1 docking motif: molecular description of binding specificities identifies interacting proteins. *Chem. Biol.* **13**, 49–59

156. Egloff, M. P., Johnson, D. F., Moorhead, G., Cohen, P. T., Cohen, P., and Barford, D. (1997) Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *EMBO J.* **16**, 1876–1887

157. Goldberg, J., Huang, H. B., Kwon, Y. G., Greengard, P., Nairn, A. C., and Kuriyan, J. (1995) Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature*. **376**, 745–753

158. Pinheiro, A. S., Marsh, J. A., Forman-Kay, J. D., and Peti, W. (2011) Structural signature of the MYPT1-PP1 interaction. *J. Am. Chem. Soc.* **133**, 73–80

159. MacMillan, L. B., Bass, M. A., Cheng, N., Howard, E. F., Tamura, M., Strack, S., Wadzinski, B. E., and Colbran, R. J. (1999) Brain actin-associated protein phosphatase 1

holoenzymes containing spinophilin, neurabin, and selected catalytic subunit isoforms. *J. Biol. Chem.* **274**, 35845–35854

160. Colbran, R. J., Bass, M. A., McNeill, R. B., Bollen, M., Zhao, S., Wadzinski, B. E., and Strack, S. (1997) Association of brain protein phosphatase 1 with cytoskeletal targeting/regulatory subunits. *J. Neurochem.* **69**, 920–929

161. Terry-Lorenzo, R. T., Carmody, L. C., Voltz, J. W., Connor, J. H., Li, S., Smith, F. D., Milgram, S. L., Colbran, R. J., and Shenolikar, S. (2002) The neuronal actin-binding proteins, neurabin I and neurabin II, recruit specific isoforms of protein phosphatase-1 catalytic subunits. *J. Biol. Chem.* **277**, 27716–27724

162. Carmody, L. C., Baucum, A. J., 2nd, Bass, M. A., and Colbran, R. J. (2008) Selective targeting of the gamma1 isoform of protein phosphatase 1 to F-actin in intact cells requires multiple domains in spinophilin and neurabin. *FASEB J.* **22**, 1660–1671

163. Kumar, G. S., Gokhan, E., De Munter, S., Bollen, M., Vagnarelli, P., Peti, W., and Page, R. (2016) The Ki-67 and RepoMan mitotic phosphatases assemble via an identical, yet novel mechanism. *Elife*. 10.7554/eLife.16539

164. Bertran, M. T., Mouilleron, S., Zhou, Y., Bajaj, R., Uliana, F., Kumar, G. S., van Drogen, A., Lee, R., Banerjee, J. J., Hauri, S., O'Reilly, N., Gstaiger, M., Page, R., Peti, W., and Tapon, N. (2019) ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. *Nat. Commun.* **10**, 771

165. Srivastava, G., Bajaj, R., Kumar, G. S., Gaudreau-Lapierre, A., Nicolas, H., Chamousset, D., Kreitler, D., Peti, W., Trinkle-Mulcahy, L., and Page, R. (2022) The ribosomal RNA processing 1B:protein phosphatase 1 holoenzyme reveals non-canonical PP1 interaction motifs. *Cell Rep.* **41**, 111726

166. Allen, P. B., Kwon, Y. G., Nairn, A. C., and Greengard, P. (1998) Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit. *J. Biol. Chem.* **273**, 4089–4095

167. Booth, V., Koth, C. M., Edwards, A. M., and Arrowsmith, C. H. (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J. Biol. Chem.* **275**, 31266–31268

168. Kim, Y.-M., Watanabe, T., Allen, P. B., Kim, Y.-M., Lee, S.-J., Greengard, P., Nairn, A. C., and Kwon, Y.-G. (2003) PNUTS, a protein phosphatase 1 (PP1) nuclear targeting subunit. Characterization of its PP1- and RNA-binding domains and regulation by phosphorylation. *J. Biol. Chem.* **278**, 13819–13828

169. Krucher, N. A., Rubin, E., Tedesco, V. C., Roberts, M. H., Sherry, T. C., and De Leon, G. (2006) Dephosphorylation of Rb (Thr-821) in response to cell stress. *Exp. Cell Res.* **312**, 2757–2763

170. De Leon, G., Sherry, T. C., and Krucher, N. A. (2008) Reduced expression of PNUTS leads to activation of Rb-phosphatase and caspase-mediated apoptosis. *Cancer Biol. Ther.* **7**, 833–841

171. Udho, E., Tedesco, V. C., Zygmunt, A., and Krucher, N. A. (2002) PNUTS (phosphatase nuclear targeting subunit) inhibits retinoblastoma-directed PP1 activity. *Biochem. Biophys. Res. Commun.* **297**, 463–467

172. Lee, S.-J., Lim, C.-J., Min, J.-K., Lee, J.-K., Kim, Y.-M., Lee, J.-Y., Won, M.-H., and Kwon, Y.-G. (2007) Protein phosphatase 1 nuclear targeting subunit is a hypoxia inducible gene: its role in post-translational modification of p53 and MDM2. *Cell Death Differ.* **14**, 1106–1116

173. Rose, M., Dütting, E., Schröder, N., Sticht, H., Brandstätter, J. H., and Enz, R. (2008) PNUTS forms a trimeric protein complex with GABAC receptors and protein phosphatase 1. *Mol. Cell. Neurosci.* **37**, 808–819

174. Kim, H., Lee, O.-H., Xin, H., Chen, L.-Y., Qin, J., Chae, H. K., Lin, S.-Y., Safari, A., Liu, D., and Songyang, Z. (2009) TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nat. Struct. Mol. Biol.* **16**, 372–379

175. Landsverk, H. B., Mora-Bermúdez, F., Landsverk, O. J. B., Hasvold, G., Naderi, S., Bakke, O., Ellenberg, J., Collas, P., Syljuåsen, R. G., and Küntziger, T. (2010) The protein phosphatase 1 regulator PNUTS is a new component of the DNA damage response. *EMBO Rep.* **11**, 868–875

176. Landsverk, H. B., Kirkhus, M., Bollen, M., Küntziger, T., and Collas, P. (2005) PNUTS enhances in vitro chromosome decondensation in a PP1-dependent manner. *Biochem. J.* **390**, 709–717

177. Schapira, M., Tyers, M., Torrent, M., and Arrowsmith, C. H. (2017) WD40 repeat domain proteins: a novel target class? *Nat. Rev. Drug Discov.* **16**, 773–786

178. Stirnimann, C. U., Petsalaki, E., Russell, R. B., and Müller, C. W. (2010) WD40 proteins propel cellular networks. *Trends Biochem. Sci.* **35**, 565–574

179. Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., 3rd, Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D., and Gamblin, S. J. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. **461**, 762–767

180. Migliori, V., Müller, J., Phalke, S., Low, D., Bezzi, M., Mok, W. C., Sahu, S. K., Gunaratne, J., Capasso, P., Bassi, C., Cecatiello, V., De Marco, A., Blackstock, W., Kuznetsov,

V., Amati, B., Mapelli, M., and Guccione, E. (2012) Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **19**, 136–144

181. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W., and Pavletich, N. P. (2007) Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell*. **26**, 131–143

182. Orlicky, S., Tang, X., Willems, A., Tyers, M., and Sicheri, F. (2003) Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell*. **112**, 243–256

183. Bae, H. J., Dubarry, M., Jeon, J., Soares, L. M., Dargemont, C., Kim, J., Geli, V., and Buratowski, S. (2020) The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. *Nat. Commun.* **11**, 2181

184. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev.* **15**, 3286–3295

185. Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12902–12907

186. Estell, C., Davidson, L., Eaton, J. D., Kimura, H., Gold, V. A. M., and West, S. (2023) A restrictor complex of ZC3H4, WDR82, and ARS2 integrates with PNUTS to control unproductive transcription. *Mol. Cell*. **83**, 2222-2239.e5

187. Park, K., Zhong, J., Jang, J. S., Kim, J., Kim, H.-J., Lee, J.-H., and Kim, J. (2022) ZWC complex-mediated SPT5 phosphorylation suppresses divergent antisense RNA transcription at active gene promoters. *Nucleic Acids Res.* **50**, 3835–3851

188. Estell, C., Davidson, L., Steketee, P. C., Monier, A., and West, S. (2021) ZC3H4 restricts non-coding transcription in human cells. *Elife*. 10.7554/eLife.67305

189. Litchfield, D. W. (2003) Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem. J.* **369**, 1–15

190. St-Denis, N. A., and Litchfield, D. W. (2009) Protein kinase CK2 in health and disease: From birth to death: the role of protein kinase CK2 in the regulation of cell proliferation and survival. *Cell. Mol. Life Sci.* **66**, 1817–1829

191. Vilk, G., Weber, J. E., Turowec, J. P., Duncan, J. S., Wu, C., Derksen, D. R., Zien, P., Sarno, S., Donella-Deana, A., Lajoie, G., Pinna, L. A., Li, S. S. C., and Litchfield, D. W. (2008) Protein kinase CK2 catalyzes tyrosine phosphorylation in mammalian cells. *Cell. Signal.* **20**, 1942–1951

192. Basnet, H., Su, X. B., Tan, Y., Meisenhelder, J., Merkurjev, D., Ohgi, K. A., Hunter, T., Pillus, L., and Rosenfeld, M. G. (2014) Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. *Nature*. **516**, 267–271

193. Stros, M., Launholt, D., and Grasser, K. D. (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci.* **64**, 2590–2606

194. Puch, C. B. M. du, Barbier, E., Kraut, A., Couté, Y., Fuchs, J., Buhot, A., Livache, T., Sève, M., Favier, A., Douki, T., Gasparutto, D., Sauvaigo, S., and Breton, J. (2011) TOX4 and its binding partners recognize DNA adducts generated by platinum anticancer drugs. *Arch. Biochem. Biophys.* **507**, 296–303

195. Wang, T., Zhao, R., Zhi, J., Liu, Z., Wu, A., Yang, Z., Wang, W., Ni, T., Jing, L., and Yu, M. (2023) Tox4 regulates transcriptional elongation and reinitiation during murine T cell development. *Commun Biol*. **6**, 613

196. Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., and Spiegel, F. W. (2012) The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–493

197. Lukeš, J., Skalický, T., Týč, J., Votýpka, J., and Yurchenko, V. (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* **195**, 115–122

198. Desjeux, P. (2004) Leishmaniasis: current situation and new perspectives. *Comp. Immunol. Microbiol. Infect. Dis.* **27**, 305–318

199. Kennedy, P. G. (2013) Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurol.* **12**, 186–194

200. Gull, K. (2001) The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. *Int. J. Parasitol.* **31**, 443–452

201. Bates, P. A. (2008) Leishmania sand fly interaction: progress and challenges. *Curr. Opin. Microbiol.* **11**, 340–344

202. Rico, E., Rojas, F., Mony, B. M., Szoor, B., Macgregor, P., and Matthews, K. R. (2013) Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei. *Front. Cell. Infect. Microbiol.* **3**, 78

203. Matthews, K. R. (2005) The developmental cell biology of Trypanosoma brucei. *J. Cell Sci.* **118**, 283–290

204. Kolev, N. G., Franklin, J. B., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010) The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS Pathog.* **6**, e1001090

205. Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X., Dewell, S., and Cross, G. A. M. (2009) Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev.* **23**, 1063–1076

206. Wedel, C., Förstner, K. U., Derr, R., and Siegel, T. N. (2017) GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J.* **36**, 2581–2594

207. Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P. J. (2003) Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell*. **11**, 1291–1299

208. Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T.-J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M. A., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M.-A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson,

A. J., Tallon, L., Turner, C. M. R., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E., and El-Sayed, N. M. (2005) The genome of the African trypanosome Trypanosoma brucei. *Science*. **309**, 416–422

209. Daniels, J.-P., Gull, K., and Wickstead, B. (2010) Cell biology of the trypanosome genome. *Microbiol. Mol. Biol. Rev.* **74**, 552–569

210. Clayton, C. (2019) Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.* **9**, 190072

211. Matthews, K. R., Tschudi, C., and Ullu, E. (1994) A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* **8**, 491–501

212. LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M. (1993) Coupling of poly(A) site selection and trans-splicing in Leishmania. *Genes Dev.* **7**, 996–1007

213. Kooter, J. M., De Lange, T., and Borst, P. (1984) Discontinuous synthesis of mRNA in trypanosomes. *EMBO J.* **3**, 2387–2392

214. Sutton, R. E., and Boothroyd, J. C. (1986) Evidence for trans splicing in trypanosomes. *Cell*. **47**, 527–535

215. Perry, K. L., Watkins, K. P., and Agabian, N. (1987) Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *Proceedings of the National Academy of Sciences*. **84**, 8190–8194

216. Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.-C., Chung, H.-M., Lee, P.-T., and Lee, M. G.-S. (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryot. Cell*. **2**, 542–551

217. Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A.-E., Sebra, R. P., and Siegel, T. N. (2018) Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*. **563**, 121–125

218. Horn, D. (2014) Antigenic variation in African trypanosomes. *Mol. Biochem. Parasitol.* **195**, 123–129

219. Cramer, P. (2019) Organization and regulation of gene transcription. *Nature*. **573**, 45–54

220. Cordon-Obras, C., Gomez-Liñan, C., Torres-Rusillo, S., Vidal-Cobo, I., Lopez-Farfan, D., Barroso-Del Jesus, A., Rojas-Barros, D., Carrington, M., and Navarro, M. (2022) Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Rep.* **38**, 110221

221. Haile, S., and Papadopoulou, B. (2007) Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Curr. Opin. Microbiol.* **10**, 569–577

222. Loya, T. J., and Reines, D. (2016) Recent advances in understanding transcription termination by RNA polymerase II. *F1000Res.* 10.12688/f1000research.8455.1

223. Li, C.-H., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., Estévez, A. M., and Clayton, C. (2006) Roles of a Trypanosoma brucei 5′→3′ exoribonuclease homolog in mRNA degradation. *RNA*. **12**, 2171–2186

224. Benz, C., Nilsson, D., Andersson, B., Clayton, C., and Guilbride, D. L. (2005) Messenger RNA processing sites in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **143**, 125–134

225. Koch, H., Raabe, M., Urlaub, H., Bindereif, A., and Preußer, C. (2016) The polyadenylation complex of Trypanosoma brucei: Characterization of the functional poly(A) polymerase. *RNA Biol.* **13**, 221–231

226. Swale, C., and Hakimi, M.-A. (2023) 3'-end mRNA processing within apicomplexan parasites, a patchwork of classic, and unexpected players. *Wiley Interdiscip. Rev. RNA.* **14**, e1783

227. Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. F., and Vannini, A. (2008) Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* **37**, 337–352

228. Darnell, J. E., Jr (2013) Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA.* **19**, 443–460

229. Egloff, S., Dienstbier, M., and Murphy, S. (2012) Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.* **28**, 333–341

230. Buratowski, S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell.* **36**, 541–546

231. Coudreuse, D., van Bakel, H., Dewez, M., Soutourina, J., Parnell, T., Vandenhaute, J., Cairns, B., Werner, M., and Hermand, D. (2010) A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Curr. Biol.* **20**, 1053–1064

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

233. Gilinger, G., and Bellofatto, V. (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.* **29**, 1556–1564

234. Smith, J. L., Levin, J. R., Ingles, C. J., and Agabian, N. (1989) In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. *Cell*. **56**, 815–827

235. Errata: Characterization of the RNA polymerases of Trypanosoma brucei: trypanosomal mRNAs are composed of transcripts derived from both RNA polymerase II and III (1989) *EMBO J.* **8**, 4359

236. Yurchenko, V. Y., Lukes, J., Jirku, M., Zeledón, R., and Maslov, D. A. (2006) Leptomonas costaricensis sp. n. (Kinetoplastea: Trypanosomatidae), a member of the novel phylogenetic group of insect trypanosomatids closely related to the genus Leishmania. *Parasitology*. **133**, 537–546

237. Das, A., and Bellofatto, V. (2009) The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in Trypanosoma brucei. *PLoS One*. **4**, e6959

238. Urbaniak, M. D., Martin, D. M. A., and Ferguson, M. A. J. (2013) Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. *J. Proteome Res.* **12**, 2233–2244

239. Das, A., Banday, M., Fisher, M. A., Chang, Y.-J., Rosenfeld, J., and Bellofatto, V. (2017) An essential domain of an early-diverged RNA polymerase II functions to accurately decode a primitive chromatin landscape. *Nucleic Acids Res.* **45**, 7886–7896

240. Jensen, T. H., Jacquier, A., and Libri, D. (2013) Dealing with pervasive transcription. *Mol. Cell*. **52**, 473–484

241. Villa, T., and Porrua, O. (2023) Pervasive transcription: a controlled risk. *FEBS J.* **290**, 3723–3736

242. Staneva, D. P., Carloni, R., Auchynnikava, T., Tong, P., Rappsilber, J., Jeyaprakash, A. A., Matthews, K. R., and Allshire, R. C. (2021) A systematic analysis of Trypanosoma brucei chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination. *Genome Res.* **31**, 2138–2154

243. Astolfi Filho, S., Martins de Sá, C., and Gander, E. S. (1980) On the chromatin structure of Trypanosoma cruzi. *Mol. Biochem. Parasitol.* **1**, 45–53

244. Borst, P., and Sabatini, R. (2008) Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.* **62**, 235–251

245. Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., and Sabatini, R. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.* **37**, 1452–1462

246. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014) Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J. Biol. Chem.* **289**, 20273–20282

247. Sekar, A., Merritt, C., Baugh, L., Stuart, K., and Myler, P. J. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **196**, 9–11

248. Kraus, A. J., Vanselow, J. T., Lamer, S., Brink, B. G., Schlosser, A., and Siegel, T. N. (2020) Distinct roles for H4 and H2A.Z acetylation in RNA transcription in African trypanosomes. *Nat. Commun.* **11**, 1498

249. van Luenen, H. G. A. M., Farris, C., Jan, S., Genest, P.-A., Tripathi, P., Velds, A., Kerkhoven, R. M., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., van Steensel, B., Myler, P. J., and Borst, P. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell*. **150**, 909–921

250. Kim, H.-S. (2021) Genetic Interaction Between Site-Specific Epigenetic Marks and Roles of H4v in Transcription Termination in Trypanosoma brucei. *Front Cell Dev Biol*. **9**, 744878

251. Ekanayake, D. K., Minning, T., Weatherly, B., Gunasekera, K., Nilsson, D., Tarleton, R., Ochsenreiter, T., and Sabatini, R. (2011) Epigenetic regulation of transcription and virulence in Trypanosoma cruzi by O-linked thymine glucosylation of DNA. *Mol. Cell. Biol.* **31**, 1690–1700

252. Kieft, R., Zhang, Y., Marand, A. P., Moran, J. D., Bridger, R., Wells, L., Schmitz, R. J., and Sabatini, R. (2020) Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. *PLoS Genet.* **16**, e1008390

253. Jensen, B. C., Phan, I. Q., McDonald, J. R., Sur, A., Gillespie, M. A., Ranish, J. A., Parsons, M., and Myler, P. J. (2021) Chromatin-Associated Protein Complexes Link DNA Base J and Transcription Termination in Leishmania. *mSphere*. 10.1128/mSphere.01204-20

254. Kieft, R., Zhang, Y., Yan, H., Schmitz, R. J., and Sabatini, R. (2023) Knockout of protein phosphatase 1 in Leishmania major reveals its role during RNA polymerase II transcription termination. *Nucleic Acids Res.* **51**, 6208–6226

255. Zhang, Y., and Sabatini, R. (2023) Leishmania PNUTS discriminates between PP1 catalytic subunits through a RVxF-ΦΦ-F motif and polymorphisms in the PP1 C-tail and catalytic domain. *bioRxiv*. 10.1101/2023.09.20.558696

256. Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., Jaehning, J. A., and Burton, Z. F. (1996) A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr. Purif.* **8**, 85–90

257. Fernández-Moya, S. M., Carrington, M., and Estévez, A. M. (2014) Depletion of the RNA-binding protein RBP33 results in increased expression of silenced RNA polymerase II transcripts in Trypanosoma brucei. *PLoS One*. **9**, e107608

258. Gómez-Liñán, C., Gómez-Díaz, E., Ceballos-Pérez, G., Fernández-Moya, S. M., and Estévez, A. M. (2022) The RNA-binding protein RBP33 dampens non-productive transcription in trypanosomes. *Nucleic Acids Res.* **50**, 12251–12265

259. Staneva, D. P., Bresson, S., Auchynnikava, T., Spanos, C., Rappsilber, J., Jeyaprakash, A. A., Tollervey, D., Matthews, K. R., and Allshire, R. C. (2022) The SPARC complex defines RNAPII promoters in Trypanosoma brucei. *Elife*. 10.7554/eLife.83135

260. Denninger, V., and Rudenko, G. (2014) FACT plays a major role in histone dynamics affecting VSG expression site control in Trypanosoma brucei. *Mol. Microbiol.* **94**, 945–962

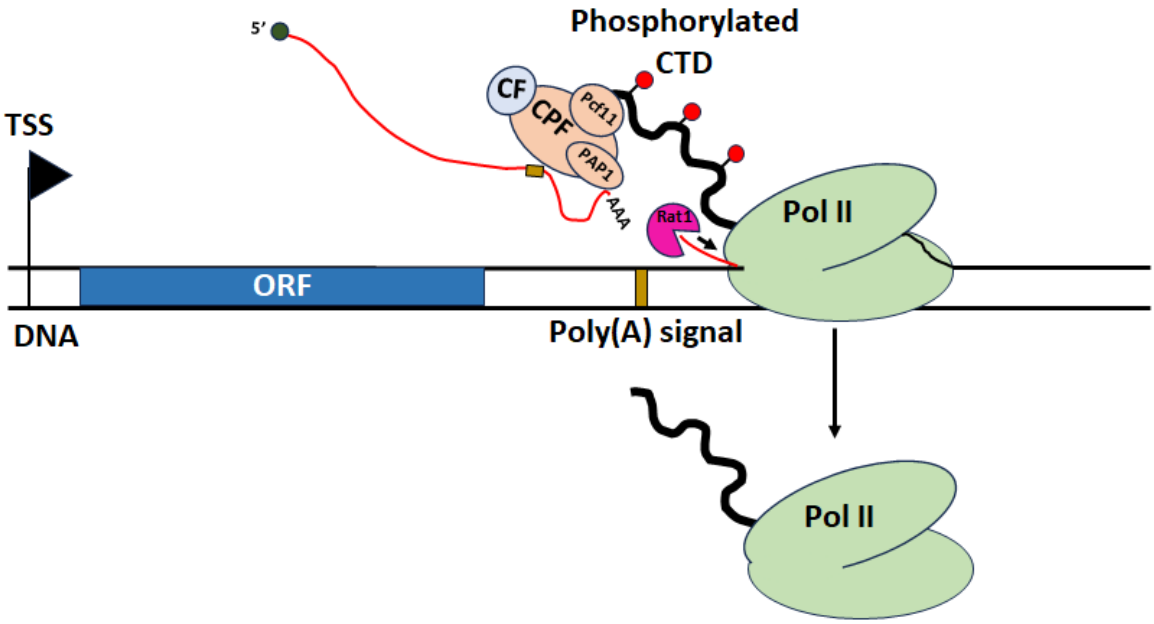


Figure 1.1 RNAPII transcription termination. In yeast cells, after the poly(A) signal is transcribed by RNAPII. 3' end processing machinery consisting of CPF and CF is brough to RNAPII by Pcf11 that recognize phosphorylated RNAPII CTD (indicated by red circle). 3' end processing machinery recognizes the PAS in nascent transcript, cleaves the nascent transcript, and polyadenylate the mRNA using associated PAP1. In the torpedo model, the 5' to 3' exoribonuclease, Rat1, degrades the nascent RNA from the exposed 5' PO_4, and eventually dislodges RNAPII from the DNA template.

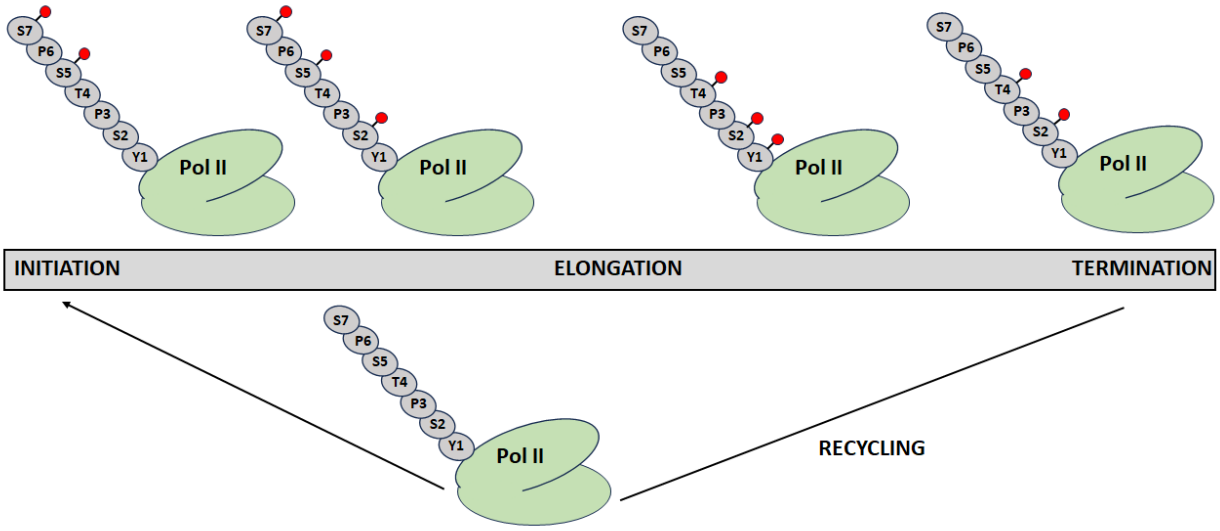


Figure 1.2 RNAPII CTD phosphorylation profile. The yeast RNAPII CTD consists of repetitive heptapeptide sequences. In different stages of the transcription cycle (Initiation, elongation and termination), particular residues on RNAPII CTD are phosphorylated (indicated by red circle) or dephosphorylated to generate RNAPII CTD phosphorylation profile and coordinate co-transcriptional activities. After transcription termination, RNAPII CTD is dephosphorylated to allow RNAPII preinitiation.

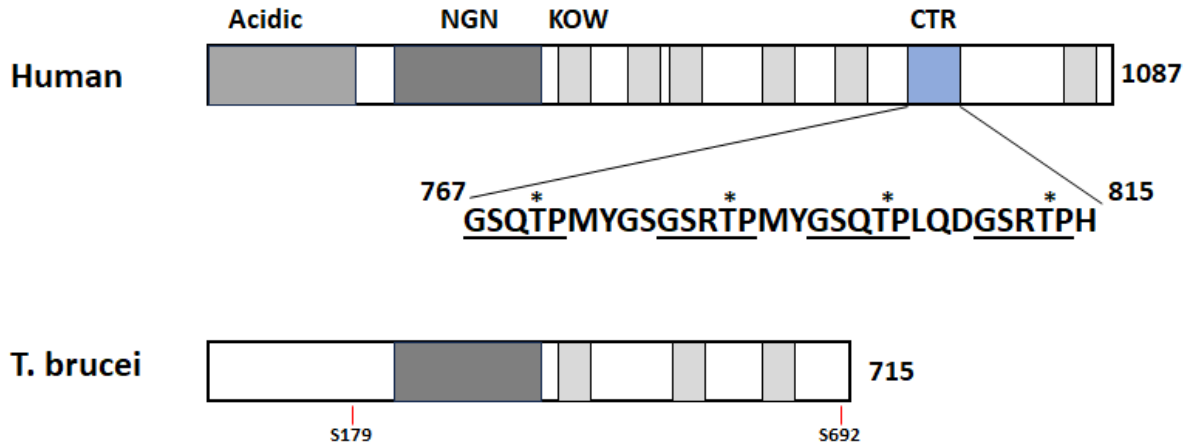


Figure 1.3 Human and *T. brucei* Spt5. The domain structures on human and *T. brucei* Spt5 are indicated. Human Spt5 has a CTR consisting of repetitive sequences. The CTR sequence is shown and underlined, and residues in the sequence that can be phosphorylated are indicated (*). Two phosphorylation sites were identified on *T. brucei* Spt5 and indicated below.

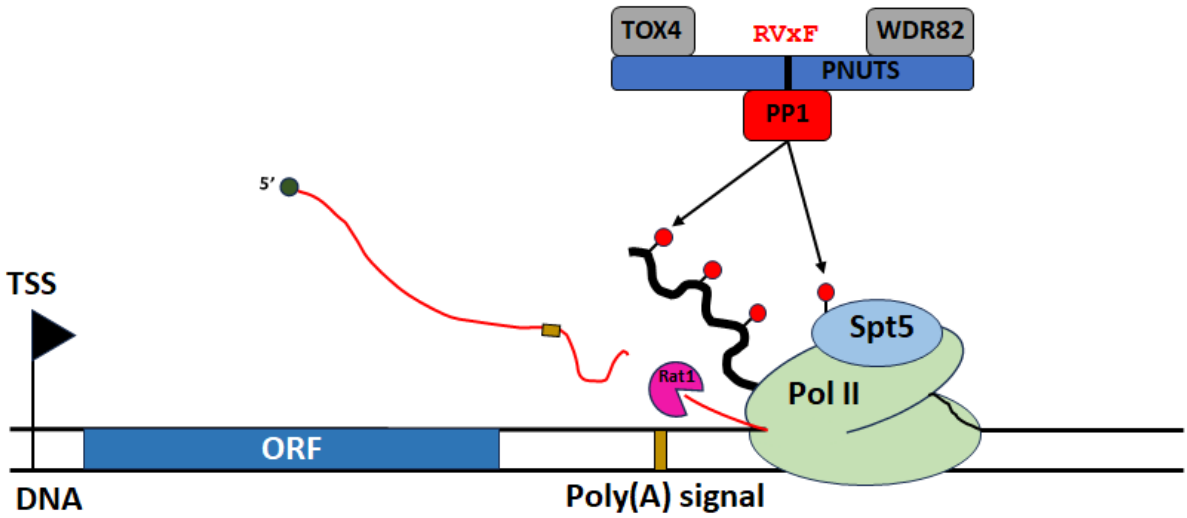


Figure 1.4 Transcription termination by the PTW/PP1 complex. PNUTS serves as a scaffolding protein, recruiting PP1 with a conserved RVxF motif, and binds to Tox4 and Wdr82 with N- and C-termini, respectively. The PTW/PP1 complex dephosphorylates Spt5 and RNAPII CTD at PAS, thereby pausing the elongating RNAPII and promoting transcription termination by torpedo mechanism.

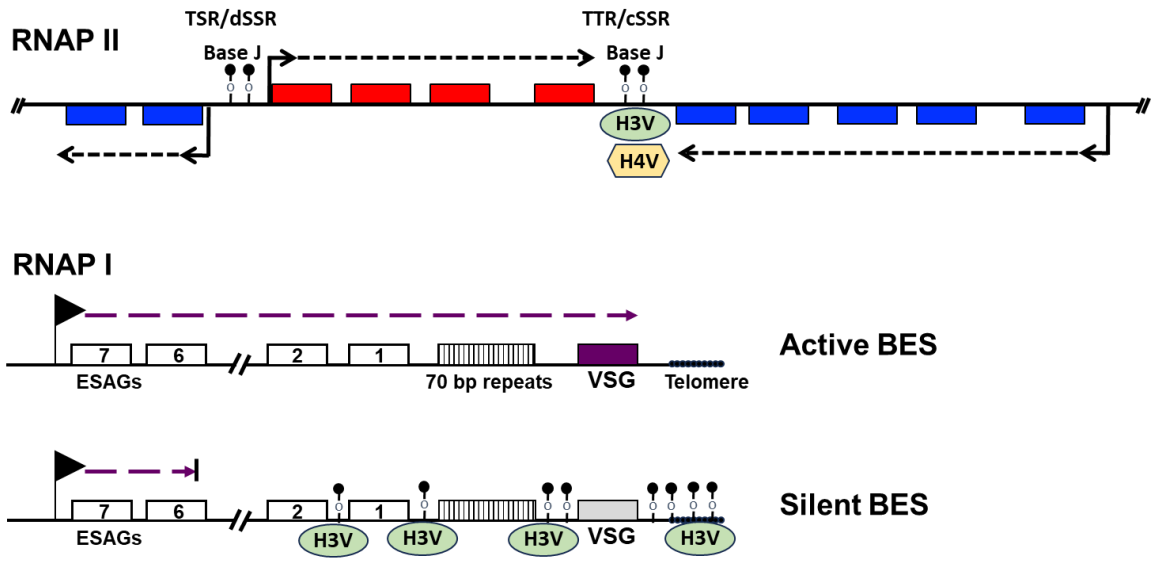


Figure 1.5 Trypanosomatid genomic organization. Genes are indicated by boxes. Genes encoded on the top strand and bottom strands are colored in red and blue, respectively. the direction of RNAPII or RNAPI transcription is indicated by dashed arrows. Enrichment of epigenetic modifications are indicated: base J in transcription start and termination regions and telomeric silent BESs; H3V in transcription termination region and silent BESs; H4V in transcription termination region.

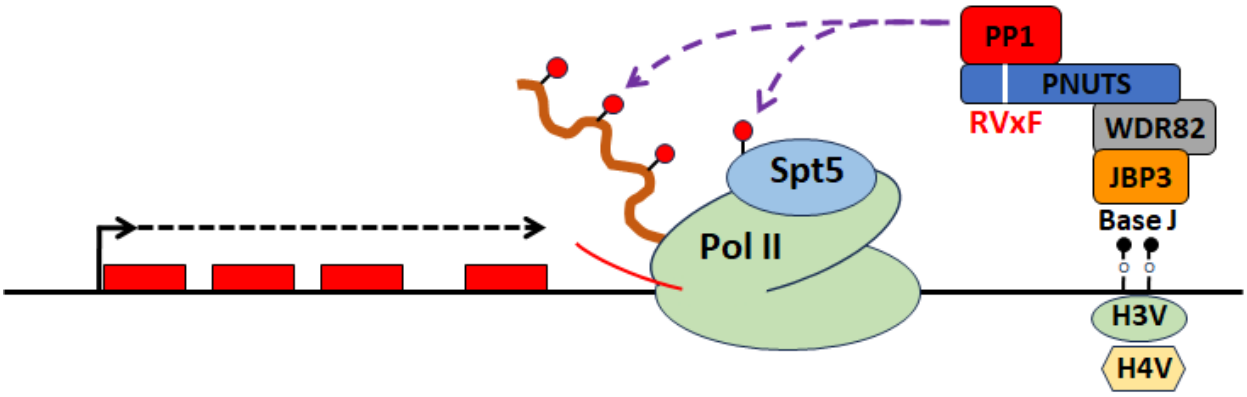


Figure 1.6 Transcription termination by the PJW/PP1 complex in trypanosomatids. In the PJW/PP1 complex, PNUTS is the scaffolding protein that associates with PP1 with the RVxF motif, and recruits Wdr82 and JBP3. The PJW/PP1 is recruited to transcription termination region by base J-JBP3 affinity, where PP1-mediated dephosphorylation of Spt5 and RNAPII CTD facilitates transcription termination.

CHAPTER 2

KNOCKOUT OF PROTEIN PHOSPHATASE 1 IN LEISHMANIA MAJOR REVEALS ITS ROLE DURING RNA POLYMERASE II TRANSCRIPTION TERMINATION[1]

[1]Kieft, R.*****, **Zhang, Y.***, Yan, H., Schmitz, R. J., & Sabatini, R. 2023. *Nucleic Acids Research*, gkad394. *co-first author

Reprinted here with permission of publisher

ABSTRACT

The genomes of kinetoplastids are organized into polycistronic transcription units that are flanked by a modified DNA base (base J, beta-D-glucosyl-hydroxymethyluracil). Previous work established a role of base J in promoting RNA polymerase II (Pol II) termination in *Leishmania major* and *Trypanosoma brucei*. We recently identified a PJW/PP1 complex in Leishmania containing a J-binding protein (JBP3), PP1 phosphatase 1, PP1 interactive-regulatory protein (PNUTS) and Wdr82. Analyses suggested the complex regulates transcription termination by recruitment to termination sites via JBP3-base J interactions and dephosphorylation of proteins, including Pol II, by PP1. However, we never addressed the role of PP1, the sole catalytic component, in Pol II transcription termination. We now demonstrate that deletion of the PP1 component of the PJW/PP1 complex in *L. major*, PP1-8e, leads to readthrough transcription at the 3'-end of polycistronic gene arrays. We show PP1-8e has *in vitro* phosphatase activity that is lost upon mutation of a key catalytic residue and associates with PNUTS via the conserved RVxF motif. Additionally, purified PJW complex with associated PP1-8e, but not complex lacking PP1-8e, led to dephosphorylation of Pol II, suggesting a direct role of PNUTS/PP1 holoenzymes in regulating transcription termination via dephosphorylating Pol II in the nucleus.

INTRODUCTION

Kinetoplastids are flagellated protozoans that are among the earliest diverging eukaryotes with a mitochondrion, whose members include pathogens responsible for multiple human diseases including human African trypanosomiasis (African sleeping sickness) and leishmaniasis. Kinetoplastid parasites, which include *Trypanosoma brucei* and *Leishmania major*, progress through life stages by cycling between an insect vector and a mammalian host.

Unlike most other eukaryotes, the entire genome of kinetoplastids is arranged into polycistronic transcription units (PTUs), consisting of tens to hundreds of genes co-transcribed from an initiation site at the 5'-end to the termination site at the 3'-end of the PTU (1,2). PTUs can be arranged in opposing directions in the genome where transcription terminates at convergent strand switch regions (cSSRs) (3). PTUs are also adjacently arranged on the same DNA strand in what is called a head-to-tail (H-T) or unidirectional arrangement, such that transcription of an upstream cluster terminates and initiation of a downstream gene cluster occurs (4–5). Pre-mRNAs are processed through trans-splicing with the addition of a 39 nucleotide spliced leader (SL) sequence to the 5' end of mRNAs (6–11) (reviewed in (12)), which is coupled to the 3' polyadenylation of the upstream transcript (13). This unique genomic organization has led to a model that gene expression is primarily controlled via post-transcriptional mechanisms such as mRNA processing, mRNA stability, and translation efficiency.

Regulation of RNA polymerase II (Pol II) transcription termination within PTUs provides a novel way to regulate protein coding gene expression in these parasites. Multiple chromatin modifications are enriched at sites of Pol II termination, including histone variants and the DNA modification base J that could be utilized by the parasites to regulate transcription (4,1,14,15). Base J consists of a glucosylated thymidine (16) and has been found only in the nuclear DNA of kinetoplastids, *Diplonema*, and *Euglena* (17,18). In kinetoplastids base J is found at almost all Pol II transcription termination sites (15) (19). J biosynthesis occurs through the hydroxylation of thymidine by either J-binding protein 1 (JBP1) or JBP2 (20), forming hydroxymethyluridine, followed by the transfer of a glucose by the glucosyltransferase enzyme JGT (21,22) (reviewed in (23,24)). Both JBPs have an N-terminal dioxygenase domain analogous to the TET proteins in mammals (25,26), and utilize 2-oxoglutarate, oxygen, and Fe^{2+} in the hydroxylation reaction

(20). JBP1 also contains a J-binding domain (27–29), while JBP2 contains a SWI/SNF2 helicase C-terminal domain presumably involved in chromatin binding/remodeling (30,31). Addition of the 2-oxoglutarate structural analog dimethyloxalylglycine (DMOG) to the growth medium inhibits hydroxylase activity of JBP1/2 and thus enables J reduction in cells without genetic modification (20). Reduction of J in Leishmania and *T. brucei* using genetic KO of JBPs or inhibition of hydroxylase activity via DMOG led to readthrough transcription at termination sites, suggesting a critical role for J in transcription termination (19,32–35). For several PTUs in *T. brucei* and *L. major*, we found J promoting Pol II termination prior to the end of the gene cluster, leading to silencing of the downstream genes (32,36). Loss of J from these premature termination sites results in readthrough transcription and derepression of the downstream genes. H3.V co-localizes with J at Pol II termination sites in *T. brucei* (15) and *L. major* (36) and has been shown to play a similar role in regulating transcription termination (33,35). J and H3.V independently function to promote termination, such that the combined loss of J and H3.V results in a synergistic increase in read through transcription, including at PTU internal termination sites (33,36). H4.V, which co-localizes with H3.V and J at termination sites in *T. brucei*, has recently been shown to play a similar synergistic role in transcription termination (34). Genes regulated by this 'premature termination' process include many that are developmentally regulated and, in the case for *T. brucei*, code for proteins involved in optimal growth and immune evasion during infection of the mammalian host (35,36).

During our studies of base J synthesis and function, we identified a PJW/PP1 complex in kinetoplastids composed of a J-binding protein (JBP3), protein phosphatase 1 (PP1), PP1 interactive-regulatory protein (PNUTS) and Wdr82 (37). Ablation of JBP3, Wdr82 or PNUTS in *T. brucei* causes read-through transcription at termination sites, indicating the role of the

complex in kinetoplastid transcription termination (37). JBP3 has also recently been shown to play a role in transcription termination in *L. tarentolae* (38). A similar PTW/PP1 complex (containing PP1, PNUTS, Wdr82 and the DNA binding protein Tox4) has been shown to play a role in regulating transcription in humans and yeast (39–46). Critical to this process is the regulation of protein phosphorylation by the major eukaryotic protein serine/threonine protein phosphatase, PP1. PP1-dependent dephosphorylation of Spt5 and Pol II leading to decreased Pol II elongation, enhances the capture by the torpedo exonuclease allowing Pol II dissociation and termination (39,41,46,47). PP1 function is modulated by its association with the PP1 regulatory factor PNUTS (PP1 nuclear targeting subunit) via the canonical PP1 RVxF interaction motif. Based on genetic and biochemical analysis of the structural components (PNUTS, JBP3 and Wdr82), we proposed the kinetoplastid PJW/PP1 complex regulates transcription termination by recruitment to termination sites via JBP3-base J interactions and dephosphorylation of specific proteins by PP1, similar to the control of termination in higher eukaryotes by PTW/PP1 (see Supplementary Figure 2.S1A). Central to this model is the sole catalytic subunit PP1. PP1, however, does not seem to be consistently included in the PJW complex across kinetoplastids (37). Purification of PNUTS from *T. brucei* extracts identified the associated proteins JBP3 and Wdr82, but not PP1. Furthermore, an obvious ortholog for the PP1 identified in the Leishmania PJW/PP1 complex is not present in the *T. brucei* genome. There are eight isoforms of PP1 in the Leishmania and Trypanosome genome distributed among five different clades (Supplementary Figure 2.S1B). The eight PP1 paralogs in *T. brucei* have been numbered 1–8 (48). In accordance with the numbering nomenclature introduced by Li and coworkers and labeling the clades A-E (38) in the phylogenetic analysis, we now refer to the PP1 component of the Leishmania PJW/PP1 complex as PP1-8e (Supplementary Figures 2.S1B and 2.S2). Interestingly, PP1-8e

belongs to a clade that lacks an ortholog in *T. brucei*, although present in other kinetoplastids including *T. cruzi* and more distantly related kinetoplastids such as *Bodo saltans* (37,38). TbPNUTS contains a 'RVxF' docking motif and presumably associates with at least one of the PP1 isotypes *in vivo* in a manner that is unstable during our purification methods and transcription termination proceeds by a similar mechanism among the kinetoplastids (37). However, due to the lack of an obvious PP1 component of the *T. brucei* complex, we sought to investigate whether the PJW/PP1 complex regulates transcription termination in a PP1-dependent manner by exploring the role of PP1-8e in Leishmania.

As stated above, PP1 mediated de-phosphorylation of both Pol II and Spt5 regulates the transition from transcription elongation to termination and Pol II release in humans and yeast (39,41,46,47). Interestingly, the only established direct substrate for PNUTS-PP1 is the C-terminal domain (CTD) of Pol II (39,49). During the transcription cycle, the largest subunit of Pol II, RPB1, becomes post-translationally modified in its CTD, which is an unstructured domain consisting of 52 repeats in humans (50). Coordinated reversible phosphorylation of the CTD regulates its association with factors involved in initiation, elongation and termination as well as co-transcriptional RNA processing (51–53). While the CTD of RPB1 in kinetoplastids is unique in that it does not contain the heptad or other repetitive motifs, 17 phospho-sites have been identified in this C-terminal serine-rich region in *T. brucei* (54,55). While this non-canonical CTD was shown to be essential for Pol II function (56,57), the functional significance of its phosphorylation remains unclear. Furthermore, a Pol II-CTD specific phosphatase has not been identified in kinetoplastids. While a CTD phosphatase RPAP2/Rtr1 homolog has been shown to associate with the *T. brucei* Pol II complex *in vivo* (58), phosphatase activity has not been studied.

Although PP1 phosphatases have been known to perform many essential functions in the life cycle of trypanosomatid parasites, including kinetoplastid segregation, cytoskeletal integrity, cytokinesis and nuclear positioning, little information is available regarding their role in transcription. No information is available regarding the function of the PP1-8e isotype in any kinetoplastid. In this work we sought to resolve the question regarding the essential role of PP1-8e in Leishmania and to test for its role in Pol II transcription termination by using a gene knockout (KO) strategy. Our data show that the PP1 isotype present in the PJW/PP1 complex has *in vitro* phosphatase activity and plays a central role in Pol II termination, where its catalytic activity is required for proper termination of Pol II transcription and repression of specific genes at the end of polycistronic units. Transcriptional and gene expression defects are similar to those seen upon the reduction in base J, directly linking DNA modification and PP1-8e protein phosphatase activity in the termination mechanism. Additionally, we show that Pol II is a direct substrate for PNUTS-PP1-8e *in vitro*. Together, these findings suggest a direct role of PNUTS/PP1 holoenzymes in regulating transcription termination via dephosphorylating Pol II in the Leishmania nucleus.

MATERIALS AND METHODS

Parasite culture

Promastigote *Leishmania major* strains were grown in M199 medium, supplemented with 10% FCS at 26°C. Transfections were performed with exponentially growing cells using the Amaxa electroporation system (Human T Cell Nucleofactor Kit, program U-033). After transfection, cells were split in two and allowed to recover for 24 hrs. before plating into 96 well plates to obtain clonal cell lines. Where appropriate, the following drug concentrations were

used: 15 μg/ml Blasticidin, 50 μg/ml Hygromycin, 10 μg/ml Puromycin and 50 μg/ml Nourseothricin. DMOG treatment of cells was performed by supplementing media with 2.5- or 5 mM DMOG for 5 days. Control cells were treated with an equal amount of DMSO. Promastigote form *L. tarentolae* were cultured in SDM79 medium and transfections performed as previously described (37). Where appropriate, the following drug concentrations were used: 50 μg/ml G418 and 10 μg/ml Puromycin.

DNA constructs and cell line generation

Endogenous HA-tagging in *L. tarentolae.* A background *L. tarentolae* cell line was established in which Cas9 and T7 polymerase are expressed from the tubulin array. WT cells were transfected with plasmid pTB007 (59), digested with PacI, to generate the Cas9/T7-expressing cell line. To tag the endogenous PP1-8e, RBP1 and PABP1 locus with 6xHA tag, this cell line was then used in a single round of transfection to generate the PP1-HA, RBP1-HA and PABP1-HA cell lines with gRNAs and donor fragments, as previously described (59). gRNAs were designed with LeishGEdit and generated *in vivo* upon transfection with the appropriate DNA fragment generated by PCR. The donor fragments were PCR-amplified from pGL2314 plasmid with 30 nt homology flanks specific to the target loci, as previously described (59).
For generation of C-terminal multi (Streptavidin Binding Protein, Protein A and FLAG) tagged constructs in *L. tarentolae*, the coding region of LtPNUTS without a stop codon was amplified and cloned into the BamH1 and XbaI sites of pSNSAP1. The resulting construct is referred to as PNUTS-Pd. The PNUTS-Pd plasmid was transfected into the PP1-8e-HA cell line and WT *L. tarentolae*. PP1-PNUTS fusions were generated, with a flexible linker between PP1 and PNUTS, as described previously for PP1-NIPP1 (60). The DNA was synthesized by Genewiz to include N-terminal Protein A, Streptavidin-Binding Protein (PS) tag. The synthesized fusion protein

DNA was sub-cloned into the pSNSAP1 plasmid by Gibson cloning to exclude the triple tag on the plasmid. The resulting construct is referred to as PP1-PNUTS fusion-PS. The sequences of all final constructs were confirmed by sequencing prior to transfection.

L. major PP1-8e KO. A background *L. major* cell line was established in which DiCre is expressed from the ribosomal locus and both Cas9 and T7 polymerase are expressed from the tubulin array. WT cells were transfected with plasmid pGL2399 (61), digested with PacI and PmeI, to generate Di-Cre-expressing cells. These cells were then transfected with plasmid pTB007 (59), digested with PacI, to generate the DiCre/Cas9/T7-expressing cell line. This cell line was then used to flank both copies of PP1-8e with LoxP sites, in a single round of transfection to generate the PP1-8e^{Flox} cell line. Donor fragment for Cas9-mediated replacement of endogenous PP1-8e was generated by PCR. WT PP1-8e was cloned into the NdeI and SpeI sites of the HA tagging/LoxP containing plasmid (pGL2341) (62). The resulting construct was used in a PCR reaction to generate the donor fragment flanked by 30 nucleotide sequence homology to the targeting integration sites, and transfected into the DiCre/Cas9/T7-expressing cell line. gRNAs were designed with LeishGEdit and generated *in vivo* upon transfection with the appropriate DNA fragment generated by PCR. PP1-8e KO cell lines were generated by adding 300 nM Rapamycin to the culture medium, reconstituting active Cre recombinase. After 14 days Rapamycin growth, clonal cell lines were obtained by limiting dilution in 96 well plates with the addition of 4.10^5 WT cells/ml. PP1-8e rescue constructs (FLAG-tagged and untagged) were generated by subcloning a *L. major* PP1 ORF-Sat^R fragment from the pXNG4 plasmid into the SpeI and BsiW I sites from the pTB007 plasmid, digested with PacI, resulting in constitutive expression from the tubulin locus. Untagged PP1-1 and PP1-7 over-expression constructs were

made the same way. All final constructs were sequenced prior to electroporation. Primers sequences used in the analysis are indicated in Supplementary Table S3.

Co-immunoprecipitation and peptide competition assay

5×10^8 of *L. tarentolae* cells were lysed and PNUTS-Pd was affinity purified using 50 µl IgG Sepharose beads as previously described (37). After incubation with cell extract, the beads were washed 3 times in 10 ml PA-150 buffer and boiled for 5 min in 1x SDS-PAGE sample buffer for western blot analysis with anti-protein A and anti-HA antibodies. For peptide competition assay, beads were resuspended in 0.3 ml PA-150 buffer and incubated with the indicated concentrations of WT (KPAEAPSRK**RVCW**ADEGHTDVSRGL) or mutant (KPAEAPSRK**RACA**ADEGHTDVSRGL) RVXF peptide while rotating at room temperature for 30 min. The IPs were then washed 3 times in 1 ml PA-150 buffer and proteins eluted by boiling for 5 min in 1× SDS-PAGE sample buffer and analyzed by western blot with anti-protein A and anti-HA antibodies.

pNPP phosphatase assay

Phosphatase activity was assayed by using the generic phosphatase substrate p-Nitrophenyl Phosphate (pNPP). PNUTS-Pd and PP1-PNUTS-PS fusion were affinity purified from 8×10^9 of *L. tarentolae* cells using the IgG Sepharose beads as above. The beads were washed 3 times in 10 ml PA-150 buffer. Following the final wash, the beads were resuspended in 10 ml PA-150 buffer. 0.1 ml bead slurry was taken, and beads collected by centrifugation and boiled in 1× sample buffer for western blot analysis with anti-protein A and anti-EF1α antibodies. The remaining 9.9 ml bead slurry was centrifuged to collect the beads, which were then resuspended in 0.2 ml assay buffer (50 mM Tris–HCl, pH 7.7, 1 mM DTT, 0.3 µM

$MnCl_2$). The phosphatase reactions were started by addition of 50 mM pNPP and were monitored by continuously following production of p-nitrophenol (absorbance at 405 nm).

Rpb1 phosphatase assay

Phosphorylation status of RBP1-HA was determined by western blot, following phosphatase treatment of cell lysates as previously described (63). Briefly, 4.5×10^8 cells were resuspended in 80 ul 0.1 M Tris–HCl, pH 6.8. Cells were lysed by addition of 20 ul 10% SDS and boiling for 5 min at 95°C. Cells lysates were then diluted 40-fold with dilution buffer (100 mM NaCl, 3 mM $MgCl_2$, 20 mM Tris–HCl, pH 7.5) to reduce the SDS concentration. Diluted cell lysates were then concentrated with Amicon spin column to 100 ul and treated with or without NEB Quick CIP and 1x PhosStop phosphatase inhibitor (Roche) for 30 min at 37°C. *In vitro* Rpb1 phosphatase assays were done by the addition of purified PNUTS-PP1 complex to purified Pol II complex from *L. tarentolae*. PNUTS was purified from *L. tarentolae* cells expressing PNUTS-Pd ($PNUTS^{WT}$ and $PNUTS^{RACA}$) using anti-FLAG magnetic beads (Invitrogen). To avoid interference of the cas9-Flag tagged protein in the PNUTS purification, we generated WT cells expressing PNUTS-Pd without HA-tagged PP1 through cas9 mediated editing. 8×10^9 cells were suspended in 20 ml lysis buffer (150 mM NaCl, 50 mM Tris–HCl pH 7.7, 1 mM EDTA, 0.5% NP-40, 10% glycerol) with protease inhibitors (8 μg/ml Aprotinin, 10 μg/ml Leupeptin, 1 μg/ml Pepstatin, 1 mM PMSF and 1x cOmplete Mini, EDTA free protease inhibitor cocktail; Roche). Cells were lysed by sonication for 5 times (15′ on/45′ off, 50% amplitude, large tip) on ice. The cell lysates were cleared by centrifugation at 21 000 × g at 4°C for 10 min and incubated with 200 μl anti-FLAG magnetic beads for 4 h at 4°C on rotor. Beads were then washed 3 times with 20 ml wash buffer (50 mM HEPES/KOH pH 7.5, 300 mM KCl, 2 mM $MgCl_2$, 0.5% NP-40) and PNUTS (and associated proteins) eluted off with 0.5 ml (400

ug/ml) 3× FLAG peptide twice at room temperature for 30 min while rotating. The two eluted fractions were pooled. The HA-tagged Rpb1 was purified from 3×10^9 *L. tarentolae* cells using 150 µl anti-HA magnetic beads as described for IgG Sepharose purification of PNUTS-Pd but with buffers supplemented with 1x PhosSTOP, 100 mM beta-glycerophosphate and 25 mM NaF to inhibit endogenous phosphatase activity. After incubation with cell extract, the beads were washed 3 times with PA-150 buffer. The bead-immobilized Rpb1 was incubated with 1 ml PNUTS eluents for indicated time periods at 30°C while rotating. After incubation, Rpb1 was sedimented and eluted by boiling for 5 min in 1× sample buffer for western blot analysis with anti-HA antibody.

Phosphorylation status of PABP1-HA was determined by western blot, following lysis of cells in 1.0% NP-40. 2.5×10^8 cells were resuspended in 0.2 ml lysis buffer (150 mM NaCl, 50 mM Tris–HCl pH 7.7, 1 mM EDTA, 1.0% NP-40, 10% glycerol) with protease inhibitors (8 µg/ml Aprotinin, 10 µg/ml Leupeptin, 1 µg/ml Pepstatin, 1 mM PMSF and 1× cOmplete Mini, EDTA free protease inhibitor cocktail; Roche) and incubated at 30°C with or without 1× PhosStop phosphatase inhibitor. Samples were taken at 0, 5, 15 and 30 min after incubation and reactions stopped by adding SDS-PAGE sample buffer and boiling for 5 min. *In vitro* PABP1 phosphatase assays were done by the addition of purified $PNUTS^{WT}$-PP1 complex to purified PABP1 from *L. tarentolae* as described above for the Rpb1 assay. Incubation of PABP1 in reaction buffer without addition of the PNUTS-PP1 complex was done as a control.

Western blotting

Proteins from 7.5×10^6 cell equivalents were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE 10% gel), transferred to nitrocellulose and probed with anti-HA antibodies (Sigma, 3F10, 1:3000) and anti-Elongation Factor 1A (Sigma,

05-235, 1:20 000) was used as a loading control. Bound antibodies were detected by an Alexa Fluor 800 labelled secondary goat anti-rat antibody and an Alexa Fluor 680 labelled secondary goat anti-mouse (LiCor) and analyzed with Image Studio software (LiCor)

RT-PCR analysis

Total RNA was isolated with Tripure Isolation Reagent (Roche). cDNA was generated from 0.5 to 1 μg Turbo™ DNase (ThermoFisher) treated total RNA with Superscript™ III (ThermoFisher) according to the manufacturer's instructions with either oligo dT primers or strand specific oligonucleotides. Strand specific RT reactions were performed with the strand specific oligonucleotide and an antisense blasticidin oligonucleotide. Equal amounts of cDNA were used in PCR reactions with Ready Go Taq Polymerase (Promega). A minus-RT control was used to ensure no contaminating genomic DNA was amplified.

Quantitative RT-PCR analysis

Total RNA was isolated and Turbo™ DNase treated as described above. Quantification of Superscript™ III generated cDNA was performed using an iCycler with an iQ5 multicolor real-time PCR detection system (Bio-Rad). Triplicate cDNA's were analyzed and normalized to enolase cDNA. qPCR oligonucleotide primers combos were designed using Integrated DNA Technologies software. cDNA reactions were diluted 20-fold and 5 μl was analyzed. A 15 μl reaction mixture contained 4.5 pmol sense and antisense primer, 7.5 μl 2× iQ SYBR green super mix (Bio-Rad Laboratories). Standard curves were prepared for each gene using 5-fold dilutions of a known quantity (100 ng/μl) of WT gDNA. The quantities were calculated using iQ5 optical detection system software. Primer sequences used in the analysis are indicated in Supplementary Table S3.

Phylogenetic analysis

Phylogenetic analysis was performed with Maximum Likelihood method and JTT matrix-based model, using MEGA11 software.

Determination of the genomic level of base J

To quantify the genomic J levels, DNA was isolated and utilized in the anti-J DNA immunoblot assay as described previously. Briefly, serially diluted genomic DNA was blotted onto a nitrocellulose membrane, followed by incubation with anti-J antisera. Bound antibodies were detected by an Alexa Fluor 800 labelled secondary goat anti-rabbit antibody. The membrane was stripped and incubated with Methylene Blue stain for DNA visualization.

Strand-specific RNA-seq library construction

For mRNA-seq, total RNA was isolated from wild-type and PP1-8e KO *L. major* cultures using TriPure. Six mRNA-seq libraries were constructed (triplicate samples for WT and PP1-8e KO) using Illumina TruSeq Stranded RNA LT Kit following the manufacturer's instructions with limited modifications. The starting quantity of total RNA was adjusted to 1.3 μg, and all volumes were reduced to a third of the described quantity. High throughput sequencing was performed on an Illumina Novaseq6000 instrument.

RNA-seq analysis

Raw reads from mRNA-seq were first trimmed using fastp with default settings (v0.23.2; (64)). Clean reads were locally aligned to the *L. major* Friedlin strain genome version 9.0 using bowtie2 tool (65) with ‘very high sensitivity’ parameter and further processed with samtools version 1.7 (66). To ensure proper read placement, alignments with multiple low-quality hits and mapping quality (MQ) scores less than 30 were removed. Strand-specific read coverage was calculated from BAM files obtained from Bowtie2 using customized pysam scripts

(https://github.com/pysam-developers/pysam). To compare read coverage for different sites the mean read coverage was calculated and normalized by total number of read number of sequencing library for each category. To compare read coverage of dSSRs, we analyzed the SSRs and the 5-kb flanking regions with DeepTools (3.5.0) using 100 bp bins flanking the SSRs and dividing each SSR into 50 equally sized bins (67). For each sample, FeatureCounts (v2.0.1) (68) was used to count reads for each reference transcript annotation, followed by normalization/variance stabilization using DESeq2 (v1.26.0) (69). Differentially expressed genes (DEGs) were identified using the DESeq2 by comparing WT and PP1 KO samples in triplicate by setting log2 fold change >1 and FDR-adjusted *P*-value <0.001 (Supplementary Table S1). The average mapping rate for mRNA-seq replicates of WT and PP1 KO was 0.92 (Supplementary Table S2). DMOG RNA-seq data shown here are from previously published work (36).

RESULTS

PP1-8e has *in vitro* phosphatase activity and associates with PNUTS via the RVxF motif

Leishmania PP1-8e contains all the invariant structural motifs (-GDXHG-, -GDXVXRG- and -GNH-) described in the members of the PPP family as well as all the kinetoplastid serine/threonine protein phosphatases characterized to date (70). As such, it possesses all six conserved residues in the catalytic site to coordinate two manganese ions (Supplementary Figure 2.S2). Since all attempts to purify soluble recombinant PP1-8e from *E. coli* failed, we utilized affinity purified enzyme expressed in *L. tarentolae* in order to examine *in vitro* protein phosphatase activity. Purification of PP1 from Leishmania extracts would also co-purify PNUTS as well as other PP1 regulatory proteins. To help ensure purified PP1-8e is associated

specifically with PNUTS, we expressed PS-tagged PP1-PNUTS fusions joined via a flexible linker (Figure 2.1A and B). In addition to the WT fusion (PP1-PNUTS), we used a fusion with an inactive catalytic moiety (PP1HK-PNUTS). PP1HK has a mutated metal-coordinating residue in the active site (H92K). Similar H66K mutation with the human PP1 catalytic site has been shown to eliminate its activity (71). Purified PP1-PNUTS fusion showed phosphatase activity with pNPP as a substrate (Figure 2.1C). As expected, the PP1HK-PNUTS fusion was largely inactive. The results indicate PP1-8e-PNUTS complex has PP1-dependent *in vitro* phosphatase activity.

To understand the function of the PP1-8e phosphatase in this complex, we initially sought to biochemically characterize the Leishmania PNUTS/PP1 complex. We verified the binary interaction between PNUTS and PP1-8e by co-immunoprecipitation experiments (Co-IP) *in vivo*. To do this we HA-tagged the endogenous locus of PP1-8e using Cas9 and expressed a Pd-tagged version of PNUTS from a plasmid in *L. tarentolae*. As expected, immunoprecipitation of PNUTS-Pd resulted in co-precipitation of PP1-8e-HA (Figure 2.2A). There is no precipitation of PP1-8e-HA in cells lacking PNUTS-Pd expression. To demonstrate the involvement of the canonical primary PP1-binding sequence, we created a cell line expressing PNUTS-Pd with V97A and W99A substitutions in the predicted RVXF motif (37). This $PNUTS^{RACA}$ mutant has minimal interaction with PP1 by Co-IP (Figure 2.2A). Furthermore, a short peptide from PNUTS that contains the RVXF motif is able to disrupt the PP1-PNUTS association, while the identical peptide with V97A and W99A substitutions is not (Figure 2.2B). We then purified the PNUTS complex from *L. tarentolae* cells expressing PNUTS-Pd (WT and RACA mutant) (Figure 2.1D) to ascertain the effect that loss of PP1-8e association had on *in vitro* phosphatase activity. The $PNUTS\text{-}Pd^{WT}$ immunoprecipitate has phosphatase activity, while the purified PNUTS-

Pd^{RACA} immunoprecipitate lacking PP1-8e does not (Figure 2.1D and E). Altogether, the data shows that Leishmania PP1-8e is a functional protein phosphatase and associates with the complex via the PNUTS RVXF PP1 binding motif.

PJW/PP1 dephosphorylates pol II *in vitro*

Dephosphorylation of the largest subunit of the Pol II complex, RPB1, has been shown to underlie the mechanism by which the PTW/PP1 complex regulates transcription termination in human cells. To investigate if the PJW/PP1 complex similarly regulates transcription termination by dephosphorylation of RPB1, we characterized the phosphatase activity of the PJW/PP1 complex from *L. tarentolea* expressing PNUTS-FLAG in an *in vitro* assay with Pol II that has been separately purified via HA-RBP1 expression. The phosphorylation level of Pol II RBP1 was first visualized by western blot analysis using anti-HA. In SDS-PAGE, trypanosome RBP1 migrates as a doublet where the upper band is phosphorylated RBP1 and the lower band is the unphosphorylated RBP1 (57,63,72). We found the Leishmania Pol II RBP1 appears as a doublet in SDS-PAGE gels, suggesting it is also phosphorylated (Figure 2.3A). To confirm the presence of these two forms was due to differences in the level of protein phosphorylation, extracts were treated with alkaline phosphatase (CIP). Phosphatase treatment led to the disappearance of the upper band and accumulation of the lower band (Figure 2.3A). This shift was completely blocked with phosphatase inhibitor. Therefore, we conclude that the upper and lower bands in our anti-HA western blots represented phosphorylated RBP1 (pRPB1) and dephosphorylated RBP1, respectively.

Purified HA-RBP1 was then tested as a substrate for purified PNUTS-PP1-8e complex. Analysis of the complex was necessary because isolated phosphatases generally have little specificity, and target-specific dephosphorylation often relies on cofactors/regulatory proteins.

As described above, PNUTS purification allows the isolation of the PJW/PP1 complex from *L. tarentolae* cells (37,38). Isolation of the PNUTS-PP1-8e component of the complex via PNUTS-Pd purification is confirmed in Figure 2.2A. We therefore utilized the PNUTS-Pd immunoprecipitate, via affinity purification of FLAG-PNUTS (Figure 2.3B), in an *in vitro* phosphatase assay with purified HA-RBP1. To test if PP1 is necessary for the phosphatase activity, PNUTS with a mutant RVxF motif (RACA) that lacks PP1-8e association (Figure (Figure 2.2A) was similarly purified (Figure 2.3B), and tested for Rpb1 *in vitro* phosphatase activity. Anti-HA immunoblotting clearly showed that the pRBP1 signal became diminished upon incubation with WT PNUTS-PP1-8e immunoprecipitate, whereas dephosphorylated RBP1 increased (Figure 2.3C). The dephosphorylation of Pol II by PNUTS-PP1-8e immunoprecipitate was completely blocked by the addition of phosphatase inhibitor and little to no dephosphorylation was seen with addition of PNUTS (RACA) mutant immunoprecipitate, lacking associated PP1-8e. These observations provide evidence that dephosphorylation of Pol II by PP1-8e was not mediated by a contaminating phosphatase and that the shift to the high mobility band was not an artifact resulting from proteolysis of Pol II.

To test if the *in vitro* phosphatase activity is specific, we characterized PP1-PNUTS activity on another phosphorylated protein, Poly-A binding protein 1 (LtPABP1). PABP1 orthologues from both *T. brucei* and Leishmania species have been shown to be targeted by serine/threonine phosphorylation events within its linker domain leading to multiple isoforms visible on the SDS-PAGE gel (73–76). LtPABP1 was represented by at least two distinctly migrating bands with apparent molecular masses of approximately 69 and 75 kDA, with the larger band representing the phosphorylated form of the protein, as previously described in Leishmania (73–75) (Figure 2.3D). To confirm the migration pattern of PABP1 on SDS gel was indeed due to

hyperphosphorylation, detergent lysis of parasites allows native phosphatases to convert the phosphorylated state to the dephosphorylated state, an activity that is inhibited by phosphatase inhibitors (Figure 2.3D). In contrast to the high dephosphorylation activity toward Pol II (Figure 2.3C), no dephosphorylation of purified HA-PABP1 was seen following incubation with WT PNUTS immunoprecipitate (Figure 2.3E). These results show the PP1-8e subunit of PJW/PP1 exhibits specific phosphatase activity toward Rpb1 *in vitro* and suggest that PP1-8e could be involved in the regulation of Pol II-CTD phosphorylation in the Leishmania nucleus.

DiCre approach for assessment of PP1 function in *L. major*

To investigate the phenotypes resulting from loss of PP1-8e in *L. major* promastigotes, an inducible knockout strain was generated using the DiCre system (Supplementary Figure 2.S3). An *L. major* strain expressing dimerizable split Cre recombinase was modified to carry 6xHA epitope-tagged alleles of PP1 flanked by LoxP sites, referred to as PP1::6xHA$^{-/\text{-flox}}$. Growth curves showed that the addition of loxP sites and the HA tag did not lead to any significant growth impairment (see below). KO induction of PP1-8e was attempted by rapamycin-mediated DiCre activation in logarithmically growing cultures (Figure 2.4). PCR analysis of these populations at 72 h after rapamycin addition revealed that the PP1::6xHA^{flox} alleles had been excised (Figure 2.4A and B). Controls without addition of rapamycin showed no gene excision (Figure 2.4B). The levels of PP1::6xHA protein were assessed by western blot, revealing a >90% reduction in the rapamycin-treated sample compared to the control sample (Figure 2.4C). After 15 days of growth in rapamycin a clone was obtained, referred to PP1 KO cB5. The PP1 KO cB5 grew normally as per the wild-type and parental DiCre strain (Figure 2.4D and data not shown). This phenotype was reproducible and observed in an independent clonal cell line (Supplementary Figure 2.S4A–C). The loss of PP1-8e gene products in both KO clones was confirmed by RT-

PCR (Supplementary Figure 2.S4D and E). We conclude that PP1-8e is not essential for Leishmania promastigote viability.

Loss of PP1-8e impairs pol II transcription in *L. major*

To assess the role of PP1-8e in transcription termination and mRNA gene expression, RNAs were isolated from the PP1 KO and WT cell lines and used to generate strand-specific RNA-seq libraries. For each condition (WT and PP1-8e KO) three independent mRNA-seq libraries were sequenced. Illumina sequencing reads were mapped to the *L. major* reference genome, and normalized read counts were calculated for every gene. First, we analyzed the read coverage for 5 kb on either side of all 146 transcription termination sites (TTSs) in the *L. major* genome. This includes convergent strand-switch regions (cSSRs) where the 3' termini of two PTUs converge and TTSs between head-to-tail (unidirectionally) oriented PTUs (Figure 2.5A–C). As expected, in WT cells the mean-normalized coverage on the top (coding) strand decreased sharply at the TTS of cSSRs (Figure 2.5A). However, when PP1-8e is deleted, the read coverage downstream of the TTS (top strand) was significantly higher (P value < 0.001), suggesting that loss of PP1-8e resulted in significant transcriptional readthrough. At these convergent sites, analysis of bottom (noncoding) strand also reveals significant differences between WT and the PP1-8e KO, where readthrough from the adjacent convergent PTU results in increased antisense transcripts past the TTS into the adjacent PTU (Supplementary Figure 2.S5A). For example, at a cSSR on chromosome 22 (cSSR 22.3) we see that the loss of PP1-8e leads to readthrough transcription on the top and bottom strands into the adjacent PTU (Figure 2.5B). RNA-seq read mapping data for all three replicates are shown in Supplementary Figure 2.S6. Quantitation of readthrough transcription at cSSRs genome-wide, by measuring the change in the ratio of antisense:sense reads in a 10kB window, indicates a 5.68-fold increase upon

deletion of PP1-8e (Supplementary Figure 2.S7). Readthrough transcription was also seen at TTS between head-to-tail (H-T) oriented PTUs (Figure 2.5C). While the gap between adjacent PTUs is small, termination defects at H-T sites are indicated by the increase in the top (coding) strand transcriptome reads from the 3' end of the last gene in the array into the downstream gene array. In many cases, the actual TTS is located prior to the final annotated gene in the PTU (32,36) (as discussed below). This would explain the increased strand coverage prior to the end of the upstream PTU at head-to-tail TTSs (Figure 2.5C). However, this increase in readthrough transcription did not occur to the same extent at different types of TTSs (Supplementary Figure 2.S5). One variable is the presence of noncoding RNA genes transcribed by Pol III (i.e. tRNA and 5S RNAs) (Supplementary Figure 2.S5B and C and Figure 2.5D). Readthrough was modest at cSSRs containing Pol III transcribed RNA genes compared with cSSRs that lack these RNA genes. Furthermore, there is little to no readthrough at centromeric cSSRs (Supplementary Figure 2.S5D) and telomeric localized TTSs (Supplementary Figure 2.S5I). Interestingly, while the presence of RNA gene at H-T sites had a similar (although smaller) effect on readthrough, centromeric localization had little to no effect (Supplementary Figure 2.S5F–H and Figure 2.5D). A similar analysis of transcript abundance surrounding transcription start sites (TSSs) revealed no significant changes in sense RNA downstream of TSSs at dSSRs due to the loss of PP1-8e (Supplementary Figure 2.S8), except for the small increase in top (coding) strand coverage at head-to-tail oriented PTUs that is presumably due to readthrough from the upstream PTU (Figure 2.5B). Furthermore, there was little to no increase in RNA upstream of TSSs at dSSRs that would correspond to bi-directional activity of Pol II promoters (Supplementary Figure 2.S8) (37).

To further characterize these transcription termination defects, normalized read counts mapped to the genome were calculated for every gene. Differential expression analysis (DESeq2 module) revealed that 34 genes had significantly higher mRNA abundances (>2-fold; $P < 0.001$) in the PP1-8e KO compared to the WT cell line (Supplementary Table S1). Interestingly, 24 of the 34 upregulated genes are located adjacent to transcription termination sites (TTSs). Therefore, the majority of the upregulated genes in the *L. major* PP1-8e KO are located at the end of a gene cluster immediately downstream or within a J peak, where J-mediated transcription termination attenuates transcription of downstream genes, similar to findings we previously described in *L. major* and *T. brucei* (32,36). This epigenetic regulated termination of Pol II transcription prior to the last ORF of the PTU we have referred to as premature termination. For example, on chromosome 9 only one gene is upregulated in the PP1 KO, and represents the last gene of a gene cluster at cSSR 9.1 (Figure 2.6A). Identical specific upregulation of this gene was seen following the loss of J at the TTS following DMOG treatment of WT cell (36). Interestingly, of the 21 genes that were upregulated upon the loss of base J (WT + DMOG) (36), 18 are also up in the PP1-8e KO (Supplementary Table S1). At the premature termination site at cSSR 9.1, loss of PP1-8e leads to readthrough transcription on the top strand that results in the complete transcription of the final annotated gene, which is processed to mature mRNA (Figure 2.6B). No expression change is detected for the genes immediately upstream or the final few genes of the adjacent convergent gene cluster in the PP1-8e KO. Gene expression changes were confirmed by RT-qPCR (Figure 2.6C and D). As previously demonstrated upon the loss of base J, readthrough transcription then extends down to a Pol III-transcribed gene in the cSSR (a 5S rRNA gene on the bottom strand and tRNA genes on the top strand), which are known to terminate Pol II transcription independent of J in *Leishmania* (19,32), as discussed above. Thus,

termination defects at this specific TTS in the PP1 KO are limited to changes in the top strand within the SSR and little to no change in antisense transcription into the adjacent PTU on either strand. Therefore, the presence of RNAP III-transcribed genes at this site seems to terminate transcription on both strands independent of PP1 (Figure 2.6B). This is in contrast to TTSs at cSSRs that lack Pol III genes (for example cSSR 22.3) where readthrough in the PP1 KO leads to antisense transcription into the adjacent PTU on both strands (Figure 2.5B). Additional example of PP1 regulation of gene expression via termination includes H-T region 26.5 (Figure 2.7) where J is found upstream of the last gene within the gene cluster. The expression of the downstream (and final) gene is increased upon the loss of J following DMOG treatment (36) or deletion of PP1-8e (Figure 2.7). Adjacent genes, upstream of the TTS or within the neighboring gene cluster, are not affected.

Additional evidence for readthrough transcription following PP1 loss is provided by strand-specific RT-PCR analysis (Figure 2.8A). Following the loss of PP1-8e we detect an increase in a nascent transcript that extends through the TTS, and further downstream, at a cSSR (Figure 2.8B). As we previously shown (32,36), the level of this RNA species increases along with decreasing levels of J in WT parasites treated with DMOG (Figure 2.8C and D). This increase in readthrough of nascent RNA leads to increased expression of downstream genes (Supplementary Figure 2.S9). Interestingly, reduction of base J levels in the PP1-8e KO upon treatment with DMOG results in further increase in readthrough transcription and appearance of a significant growth defect (Figure 2.8C–E and Supplementary Figure 2.S9).

To ensure the phenotype was specific to PP1 deletion and not due to off-target effects, an allele of PP1::Flag was reintroduced to the PP1 KO. Unfortunately, the use of Flag tag was not ideal with expression of Cas9-Flag in the same cell, and background not allowing detection of

PP1-Flag expression by western blot (data not shown). Thus, expression of PP1-Flag was confirmed at the RNA level (Supplementary Figure 2.S4D), with expression ~3-fold higher than WT endogenous levels. Termination of the PP1-Flag-complementation strain, measured by upregulation of the gene downstream of TTS at cSSR 9.1, was restored to levels seen in the PP1::6xHA (Figure 2.6D). While not 100% complementation (compared to WT), it potentially reflects the negative effect of C-terminal tag on PP1 function, demonstrating the requirement for PP1 for transcription termination. It appears that the replacement of WT PP1-8e locus with PP1-8e-6xHA tagged version already led to defects in termination at the level of nascent RNA and gene expression changes (Figures 2.6D and 2.8B). Rapamycin-mediated PP1-HA allele excision had additional effect on termination defects measured by gene expression and nascent RNA. Therefore, the presence of either C-terminal Flag or HA tag may significantly affect PP1-8e function. To explore the effect of C-terminal tags, we expressed an untagged version of PP1 in these cells. Expression of the untagged version of PP1-8e, with similar 3-fold higher levels of mRNA than WT (Supplementary Figure 2.S4D), resulted in an almost complete rescue in readthrough transcription measured by both gene expression and nascent RNA changes (Figures 2.6D and 2.8B). Taken together, these findings suggest PP1 phosphatase functions as part of the PP1/PJW complex in the promotion of Pol II termination in Leishmania.

To explore possible redundancy of PP1 isotype function, we over-expressed an untagged version of PP1-1a and PP1-7d isotypes in the PP1-8e KO. Among the 8 PP1 isotypes in *T. brucei*, PP1-1 and PP1-7 been shown to primarily localize within the nucleus (Tryptag.org; (77)). Interestingly, over-expression of LmPP1-1a or LmPP1-7d were able to partially rescue the PP1-8e KO (p-value of 0.35 and 0.035, respectively) (Supplementary Figure 2.S10). To determine whether PP1-8e catalytic activity is required for Pol II termination in *L. major*, we expressed an

untagged PP1-8e H92K mutant in the PP1 KO. H92K mutation of PP1-8e completely eliminated phosphatase activity toward pNPP *in vitro* (Figure 2.1C). Importantly, *L. major* cells that exclusively express PP1-8e H92K, at identical levels of mRNA as WT from the native locus, were unable to properly terminate Pol II transcription (Supplementary Figure 2.S10). Hence, termination of Pol II transcription requires both PP1-8e expression and activity.

DISCUSSION

These findings significantly expand our understanding of the mechanism of Pol II transcription termination in highly divergent organisms that utilize polycistronic transcription and therefore, need to decouple termination from 3'-end formation of individual genes. Pol II-mediated transcription termination of most protein-encoding genes in eukaryotes is directly linked to 3'-end formation where, according to the 'torpedo' model, cleavage of the nascent transcript at the poly(A) site provides access for the 5'-3' RNA exonuclease Xrn2/Rat1 (human/yeast) (46,78–81) that eventually leads to dissociation of the polymerase from the DNA template. Thus, termination is enhanced by mechanisms that decelerate Pol II and biases the kinetic competition between Pol II and the exonuclease torpedo chasing it down (78). Pol II speed is reduced by the PP1-mediated dephosphorylation (as a component of the mammalian PTW/PP1 complex) of the transcription elongation factor Spt5, resulting in deceleration of transcription downstream of poly(A) sites enhancing torpedo dislodgment of Pol II (46). PP1 mediated dephosphorylation of Pol II CTD coordinates the recruitment of factors involved in 3' end formation and termination (42). In yeast, PP1 dephosphorylation of both Pol II CTD and Spt5 is thought to orchestrate the recruitment of the termination factor Seb1 and the transition from elongation to termination and Pol II release (82). Identification of an analogous PJW/PP1

complex in Leishmania suggested a similar Pol II termination mechanism in kinetoplastids (37). Substitution of the mammalian Tox4 DNA binding protein with the base J binding protein (JBP3), presumably allows the complex to terminate Pol II transcription at specific sites at the end of PTUs marked by base J. According to this model, PP1-mediated Pol II deceleration would be stimulated upon reaching base J rather than poly(A) sites within the polycistronic gene array. The coupling of trans splicing and polyadenylation, preventing the generation of the 5' phosphate substrate for the 5'-exonuclease torpedo, may explain the ability to bypass the link between Pol II termination and 3' end formation within the gene array. Whether termination then proceeds via the 'torpedo' model upon recruitment of the PJW/PP1 complex remains to be tested. The analysis of the structural components of the PJW/PP1 complex (PNUTS, Wdr82 and JBP3) in *T. brucei* and JBP3 mutant in *L. tarentolae* indicated the critical role of the complex in transcription termination (37,38). While PP1 is the only catalytic component of the mammalian PTW/PP1 complex and dephosporylation by PP1-PNUTS is directly involved in regulating Pol II termination in human and yeast, analysis of PP1 function in kinetoplastid Pol II transcription has not been addressed until now (2,3,60,61). We now show that the PP1 component of the PJW/PP1 complex, PP1-8e, plays a key role in controlling the termination of Pol II transcription in kinetoplastids since the deletion of PP1-8e leads to defects in transcription termination at the 3' ends of PTUs in *L. major*, similar to phenotypes seen following the knockdown of PNUTS, JBP3 or Wdr82 in *T. brucei* (37) and JBP3 in *L. tarentolae* (38). We also show that PP1-8e is a protein phosphatase and is able to directly dephosphorylate Pol II *in vitro*. This is the first demonstration that the dephosphorylation of Pol II in kinetoplastids is mediated by PP1 (PJW/PP1 complex) *in vitro* and supports our overall model of PJW/PP1 complex function in Pol II termination (Supplementary Figure 2.S1).

Interestingly, termination defects are not seen to the same extent at all TTSs in the PP1-8e KO. The presence of Pol III transcribed RNA genes downstream of the TTS in cSSRs appears to effectively block Pol II readthrough transcription, as seen previously in Leishmania with reduced levels of base J (19,32), and there is essentially no readthrough at cSSRs at centromeric or telomeric locations. Similar differential defects in termination were recently described in the Leishmania JBP3 mutant (38). This suggests that other factors other than the PJW/PP1 complex and base J can play a role in reducing transcriptional readthrough at these loci. For example, apparent reduced read-through defects at H-T regions may reflect altered chromatin structures at termination sites adjacent to Pol II initiation complex assemblies. The impact of chromatin in Kinetoplastid transcription termination is illustrated by the role of H3V and H4V enriched at termination sites (32–34,36). Chromatin components associated with Pol II initiation may help prevent interference from upstream transcription elongation independent of PJW/PP1 complex function. The lack of defects at centromeric or telomeric locations may similarly reflect unique chromatin structures as well as compartmentalization within the nucleoplasm. Telomeres tend to be close to the nuclear periphery in trypanosomes. Compartmentalization of centromeric heterochromatin has been demonstrated in mammalian cells, often localized near nuclear lamina where chromatin is largely silenced. Furthermore, several histone markers have recently been characterized as uniquely associated with centromeres in Leishmania as well as being enriched for base J (83). Additional work is needed to identify other factors involved in termination. Transcriptional defects in the *T. brucei* PJW complex mutants included de-repression of genes located upstream of transcription start sites resulting from transcription between diverging PTUs (37). Apparently, similar to mammalian and yeast, Pol II transcription initiation sites are bi-directional in *T. brucei* giving rise to transcription in both sense and divergent antisense

directions where unidirectional transcription is ensured by early termination of antisense RNA by the PJW complex. In contrast to *T. brucei*, we find little evidence for antisense transcription between diverging PTUs in *L. major*, even after the loss of PJW/PP1 complex function in the PP1-8e KO. Similar lack of antisense transcription at the 5' ends of PTUs was also characterized in the LtJBP3 mutants (38). It is unclear why bi-directional activity and regulated antisense transcription in these regions would be restricted to Trypanosomes. However, in contrast to *T. brucei*, dSSRs of Leishmania lack the presence of transcriptionally regulated genes as well as being smaller in size and mostly lacking base J.

Importantly, similar to the transcription termination defects seen in base J and H3.V mutants in *T. brucei* and *L. major* (32,33,36), and PJW complex mutants in *T. brucei (37)*, the loss of PP1-8e results in the upregulation of mRNA levels for protein-coding genes downstream of base J and H3.V marked TTS at the 3'-end of PTUs. In fact, many of upregulated genes are shared between the base J mutant and the PP1 KO in *L. major*. Growth of WT *L. major* in DMOG resulted in 10-fold reduction of the modified DNA base and termination defects. Further reduction of base J levels at termination sites by DMOG treatment of the *H3.V* KO revealed greater termination defects, more significant gene expression changes, and greatly reduced cell growth (36). Interestingly, reducing J levels in the PP1-8e KO by DMOG treatment led to a similar additive increase in readthrough transcription and appearance of growth defects (Figure 2.8C-E and Supplementary Figure 2.S9). While readthrough transcription in *L. major* include the extension of Pol II onto the adjacent opposing gene cluster and dual strand transcription, we saw no evidence of transcription interference resulting in significant downregulation of mRNAs on the opposing gene cluster in cells with reduced base J (36). Similarly, we see no evidence of transcription interference in the PP1-8e KO here. Taken together these results indicate a

conserved role for J and the PJW/PP1 complex in regulating Pol II transcription termination and expression of genes within polycistronic gene clusters in kinetoplastids, and suggest that the essential nature of J and the PJW/PP1 complex (32,36,38,84) in Leishmania is related to their role in repressing specific genes at the end of gene clusters rather than the prevention of dual strand transcription. Overall, these data suggest that PP1-8e works with the PJW complex to terminate Pol II transcription at the end of PTUs, sometimes leading to premature termination thereby shaping the transcriptome.

However, while we now provide a mechanistic link between J modification of DNA at termination sites, Pol II dephosphorylation and transcription termination in kinetoplastids, how do we explain the lack of PP1 in the *T. brucei* PJW complex and the non-essential nature of LmPP1-8e? Especially when the other structural components (PNUTS, Wdr82 and JBP3) appear to be essential for termination as well as parasite growth (37,38). PP1-8e is the only PP1 isotype associated with the Leishmania PJW/PP1 complex *in vivo* (37,38). Sequence analysis suggests this is due to unique insertions within the PP1 catalytic subunit and C-terminal tail important for PNUTS association (Supplementary Figure 2.S2). While this hypothesis remains to be tested, this is reminiscent of the diversity of PP1 tails in Drosophila and mammals presumably involved in regulator protein interactions (85). In fact, the Drosophila PP1 regulatory protein ASPP (apoptosis-stimulating protein of p53) can discriminate between different PP1 isoforms based on the PP1 C-tail (85,86). The absence of these LmPP1-8e unique sequences in all 8 PP1 isoforms of *T. brucei* may explain why none is stably associated with the complex in *T. brucei* under our purification conditions. The finding that TbPNUTS contains the RVXF PP1-binding motif and is involved in termination, along with JBP3 and Wdr82, (37,38) combined with the data here showing LmPP1 function in termination, strongly suggests PP1 is involved in termination via the

PJW complex in *T. brucei*. TbPNUTS presumably associates with at least one of the PP1 isotypes *in vivo* in a manner that is unstable during our complex purification methods. Interestingly, PP1-1 is closely related to the PP1 orthologue (Glc7) involved in transcription termination in yeast (Supplementary Figure 2.S1) and has been shown to localize within the *T. brucei* nucleus (Tryptag.org; (77)) and associate with the PIP5Pase/RAP1 complex bound to telomeric VSG expression sites (87). The ability of LmPP1-1, or another isoform, to function in the PJW complex to a certain extent may explain why the LmPP1-8e KO is viable and the reduction of base J via DMOG results in additional defects in termination and defects in cell growth. The ability of LmPP1-1 and LmPP1-7 over-expression to partially rescue the PP1-8e KO, supports this idea. This redundancy is understandable since protein phosphatases use structurally related catalytic domains that are remarkably well conserved and shown to be relatively promiscuous *in vitro* (88). Furthermore, the ability of distinct phosphatases to functionally substitute has been demonstrated *in vivo*. For example, PP1 and PP2A-B56 phosphatases are recruited via their respective motif (RVxF and LxxIxE, respectively) containing regulatory proteins to allow control over different substrates and different mitotic processes (89–91). Removing either phosphatase produces markedly distinct phenotypic effects. However, modifying the PP1-binding motif of the regulatory protein to allow the alternative phosphatase be recruited in its place has shown they can functionally substitute for each other at kinetochores (92). Therefore, we believe some low level PNUTS complex association among the remaining PP1 isotypes allows sufficient termination control in the PP1-8e KO to remain viable.

Overall, this work provides new insight into the molecular mechanism utilized to control transcription termination at the end of PTUs in these divergent parasites. The data support conserved function of proteins involved in transcription termination among eukaryotes, despite

the need of kinetoplastids to bypass termination at the 3' end of every gene and terminate in base J specific manner within or at the end of polycistronic gene arrays. Further work is needed to understand details of this mechanism, including identifying other potential substrates of PP1-PNUTS.

DATA AVAILABILITY

All sequencing data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession numbers GSE200788.

ACKNOWLEDGEMENT

We are grateful for Erin Campbell for preparing the mRNA-seq libraries. We also thank Richard McCulloch, Jeziel Damasceno and Catarina De Almeida Marques for help setting up the Cas9 DiCre system.

FUNDING

National Institutes of Health [R01AI109108 to R.S.]; Office of the Vice President for Research (to R.J.S.). Funding for open access charge: Foundation for the National Institutes of Health [R01AI109108]

Conflict of interest statement. None declared.

REFERENCE

1. Thomas S., Green A., Sturm N.R., Campbell D.A., Myler P.J.. Histone acetylations mark origins of polycistronic transcription in Leishmania major. BMC Genomics [Electronic Resource]. 2009; **10**:152.

2. El-Sayed N.M., Myler P.J., Blandin G., Berriman M., Crabtree J., Aggarwal G., Caler E., Renauld H., Worthey E.A., Hertz-Fowler C.et al... Comparative genomics of trypanosomatid parasitic protozoa. Science. 2005; **309**:404–409.

3. Martinez-Calvillo S., Nguyen D., Stuart K., Myler P.J.. Transcription initiation and termination on Leishmania major chromosome 3. Eukaryot. Cell. 2004; **3**:506–517.

4. Siegel T.N., Hekstra D.R., Kemp L.E., Figueiredo L.M., Lowell J.E., Fenyo D., Wang X., Dewell S., Cross G.A.. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes Dev. 2009; **23**:1063–1076.

5. Kolev N.G., Franklin J.B., Carmi S., Shi H., Michaeli S., Tschudi C.. The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS Pathog. 2010; **6**:e1001090.

6. Boothroyd J.C., Cross G.A.. Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5' end. Gene. 1982; **20**:281–289.

7. Van der Ploeg L.H., Liu A.Y., Michels P.A., De Lange T.D., Borst P., Majumder H.K., Weber H., Veeneman G.H., Van Boom J.. RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucleic Acids Res. 1982; **10**:3591–3604.

8. De Lange T., Liu A.Y., Van der Ploeg L.H., Borst P., Tromp M.C., Van Boom J.H.. Tandem repetition of the 5' mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription?. Cell. 1983; **34**:891–900.

9. Nelson R.G., Parsons M., Barr P.J., Stuart K., Selkirk M., Agabian N.. Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in trypanosoma brucei. Cell. 1983; **34**:901–909.

10. Sutton R.E., Boothroyd J.C.. Evidence for trans splicing in trypanosomes. Cell. 1986; **47**:527–535.

11. Agabian N. Trans splicing of nuclear pre-mRNAs. Cell. 1990; **61**:1157–1160.

12. Borst P. Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. 1986; **55**:701–732.

13. LeBowitz J.H., Smith H.Q., Rusche L., Beverley S.M.. Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes Dev. 1993; **7**:996–1007.

14. Respuela P., Ferella M., Rada-Iglesias A., Aslund L.. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi. J. Biol. Chem. 2008; **283**:15884–15892.

15. Cliffe L.J., Siegel T.N., Marshall M., Cross G.A., Sabatini R.. Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. Nucleic Acids Res. 2010; **38**:3923–3935.

16. Gommers-Ampt J.H., Van Leeuwen F., de Beer A.L., Vliegenthart J.F., Dizdaroglu M., Kowalak J.A., Crain P.F., Borst P.. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell. 1993; **75**:1129–1136.

17. van Leeuwen F., Taylor M.C., Mondragon A., Moreau H., Gibson W., Kieft R., Borst P.. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid

protozoans and is abundant in their telomeres [see comments]. Proc. Nat. Acad. Sci. U.S.A. 1998; **95**:2366–2371.

18. Dooijes D., Chaves I., Kieft R., Dirks-Mulder A., Martin W., Borst P.. Base J originally found in kinetoplastid is also a minor constituent of nuclear DNA of Euglena gracilis. Nucleic Acids Res. 2000; **28**:3017–3021.

19. van Luenen H.G., Farris C., Jan S., Genest P.A., Tripathi P., Velds A., Kerkhoven R.M., Nieuwland M., Haydock A., Ramasamy G.et al... Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell. 2012; **150**:909–921.

20. Cliffe L.J., Hirsch G., Wang J., Ekanayake D., Bullard W., Hu M., Wang Y., Sabatini R.. JBP1 and JBP2 proteins are Fe2+/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. J. Biol. Chem. 2012; **287**:19886–19895.

21. Bullard W., da Rosa-Spiegler J.L., Liu S., Wang D., Sabatini R.. Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. J. Biol. Chem. 2014; **289**:20273–20282.

22. Sekar A., Merritt C., Baugh L., Stuart K., Myler P.J.. Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. Mol. Biochem. Parasitol. 2014; **196**:9–11.

23. Borst P., Sabatini R.. Base J: discovery, biosynthesis, and possible functions. Annu. Rev. Microbiol. 2008; **62**:235–251.

24. Reynolds D., Cliffe L., Sabatini R.. Schofield C.J., Hausinger R.P.. 2-Oxoglutarate-dependent hydroxylases Involved in DNA base J (β-D-glucopyranosyloxymethyluracil)

synthesis. 2-Oxoglutarate-Dependent Oxygenases. 2015; Cambridge: Royal Society of Chemistry; 309–323.

25. Iyer L.M., Tahiliani M., Rao A., Aravind L.. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle. 2009; **8**:1698–1710.

26. Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L.et al... Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; **324**:930–935.

27. Heidebrecht T., Christodoulou E., Chalmers M.J., Jan S., Ter Riet B., Grover R.K., Joosten R.P., Littler D., van Luenen H., Griffin P.R.et al... The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. Nucleic Acids Res. 2011; **39**:5715–5728.

28. Sabatini R., Meeuwenoord N., van Boom J.H., Borst P.. Recognition of base J in duplex DNA by J-binding protein. J. Biol. Chem. 2002; **277**:958–966.

29. Sabatini R., Meeuwenoord N., van Boom J.H., Borst P.. Site-specific interactions of JBP with base and sugar moieties in duplex J-DNA. J. Biol. Chem. 2002; **277**:28150–28156.

30. DiPaolo C., Kieft R., Cross M., Sabatini R.. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Mol. Cell. 2005; **17**:441–451.

31. Cliffe L.J., Kieft R., Southern T., Birkeland S.R., Marshall M., Sweeney K., Sabatini R.. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Res. 2009; **37**:1452–1462.

32. Reynolds D., Cliffe L., Forstner K.U., Hon C.C., Siegel T.N., Sabatini R.. Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. Nucleic Acids Res. 2014; **42**:9717–9729.

33. Reynolds D., Hofmeister B.T., Cliffe L., Alabady M., Siegel T.N., Schmitz R.J., Sabatini R.. Histone H3 variant regulates RNA polymerase II transcription termination and dual strand transcription of siRNA loci in Trypanosoma brucei. PLoS Genet. 2016; **12**:e1005758.

34. Kim H.-S. Genetic interaction between site-specific epigenetic marks and roles of H4v in transcription termination in Trypanosoma brucei. Front. Cell Dev. Biol. 2021; **9**:744878.

35. Schulz D., Zaringhalam M., Papavasiliou F.N., Kim H.S.. Base J and H3.V regulate transcriptional termination in Trypanosoma brucei. PLos Genet. 2016; **12**:e1005762.

36. Reynolds D.L., Hofmeister B.T., Cliffe L., Siegel T.N., Anderson B.A., Beverley S.M., Schmitz R.J., Sabatini R.. Base J represses genes at the end of polycistronic gene clusters in Leishmania major by promoting RNAP II termination. Mol. Microbiol. 2016; **101**:559–574.

37. Kieft R., Zhang Y., Marand A.P., Moran J.D., Bridger R., Wells L., Schmitz R.J., Sabatini R.. Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. PLoS Genet. 2020; **16**:e1008390.

38. Jensen B.C., Phan I.Q., McDonald J.R., Sur A., Gillespie M.A., Ranish J.A., Parsons M., Myler P.J.. Chromatin-associated protein complexes link DNA base J and transcription termination in Leishmania. Msphere. 2021; **6**:e01204-20.

39. Ciurciu A., Duncalf L., Jonchere V., Lansdale N., Vasieva O., Glenday P., Rudenko A., Vissi E., Cobbe N., Alphey L.et al... PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. PLos Genet. 2013; **9**:e1003885.

40. Jerebtsova M., Klotchenko S.A., Artamonova T.O., Ammosova T., Washington K., Egorov V.V., Shaldzhyan A.A., Sergeeva M.V., Zatulovskiy E.A., Temkina O.A.et al... Mass spectrometry and biochemical analysis of RNA polymerase II: targeting by protein phosphatase-1. Mol. Cell. Biochem. 2011; **347**:79–87.

41. Kecman T., Kuś K., Heo D.H., Duckett K., Birot A., Liberatori S., Mohammed S., Geis-Asteggiante L., Robinson C.V., Vasiljeva L.. Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. Cell Rep. 2018; **25**:259–269.

42. Schreieck A., Easter A.D., Etzold S., Wiederhold K., Lidschreiber M., Cramer P., Passmore L.A.. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nat. Struct. Mol. Biol. 2014; **21**:175–179.

43. Cheng H., He X., Moore C.. The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. Mol. Cell. Biol. 2004; **24**:2932–2943.

44. Nedea E., Nalbant D., Xia D., Theoharis N.T., Suter B., Richardson C.J., Tatchell K., Kislinger T., Greenblatt J.F., Nagy P.L.. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol. Cell. 2008; **29**:577–587.

45. Austenaa L.M., Barozzi I., Simonatto M., Masella S., Della Chiara G., Ghisletti S., Curina A., de Wit E., Bouwman B.A., de Pretis S.et al... Transcription of mammalian cis-regulatory elements is restrained by actively enforced early termination. Mol. Cell. 2015; **60**:460–474.

46. Cortazar M.A., Sheridan R.M., Erickson B., Fong N., Glover-Cutter K., Brannan K., Bentley D.L.. Control of RNA pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a “sitting duck torpedo” mechanism. Mol. Cell. 2019; **76**:896–908.

47. Parua P.K., Booth G.T., Sanso M., Benjamin B., Tanny J.C., Lis J.T., Fisher R.P.. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature. 2018; **558**:460–464.

48. Li Z., Tu X., Wang C.C.. Okadaic acid overcomes the blocked cell cycle caused by depleting Cdc2-related kinases in Trypanosoma brucei. Exp. Cell. Res. 2006; **312**:3504–3516.

49. Lee J.H., You J., Dobrota E., Skalnik D.G.. Identification and characterization of a novel human PP1 phosphatase complex. J. Biol. Chem. 2010; **285**:24466–24476.

50. Harlen K.M., Churchman L.S.. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. 2017; **18**:263–273.

51. Custodio N., Carmo-Fonseca M.. Co-transcriptional splicing and the CTD code. Crit. Rev. Biochem. Mol. Biol. 2016; **51**:395–411.

52. Buratowski S. Progression through the RNA polymerase II CTD cycle. Mol. Cell. 2009; **36**:541–546.

53. Hsin J.P., Manley J.L.. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 2012; **26**:2119–2137.

54. Nett I.R., Martin D.M., Miranda-Saavedra D., Lamont D., Barber J.D., Mehlert A., Ferguson M.A.. The phosphoproteome of bloodstream form trypanosoma brucei, causative agent of African sleeping sickness. Mol. Cell. Proteomics. 2009; **8**:1527–1538.

55. Urbaniak M.D., Martin D.M., Ferguson M.A.. Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. J. Proteome Res. 2013; **12**:2233–2244.

56. Das A., Banday M., Fisher M.A., Chang Y.J., Rosenfeld J., Bellofatto V.. An essential domain of an early-diverged RNA polymerase II functions to accurately decode a primitive chromatin landscape. Nucleic Acids Res. 2017; **45**:7886–7896.

57. Das A., Bellofatto V.. The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in Trypanosoma brucei. PLoS One. 2009; **4**:e6959.

58. Srivastava A., Badjatia N., Lee J.H., Hao B., Gunzl A.. An RNA polymerase II-associated TFIIF-like complex is indispensable for SL RNA gene transcription in Trypanosoma brucei. Nucleic Acids Res. 2017; **46**:1695–1709.

59. Beneke T., Madden R., Makin L., Valli J., Sunter J., Gluenz E.. A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. R. Soc. Open Sci. 2017; **4**:170095.

60. Winkler C., Rouget R., Wu D., Beullens M., Van Eynde A., Bollen M.. Overexpression of PP1-NIPP1 limits the capacity of cells to repair DNA double-strand breaks. J. Cell Sci. 2018; **131**:jcs214932.

61. Santos R., Silva G.L.A., Santos E.V., Duncan S.M., Mottram J.C., Damasceno J.D., Tosi L.R.O.. A DiCre recombinase-based system for inducible expression in Leishmania major. Mol. Biochem. Parasitol. 2017; **216**:45–48.

62. Duncan S.M., Myburgh E., Philipon C., Brown E., Meissner M., Brewer J., Mottram J.C.. Conditional gene deletion with DiCre demonstrates an essential role for CRK3 in Leishmania mexicana cell cycle regulation. Mol. Microbiol. 2016; **100**:931–944.

63. Badjatia N., Ambrosio D.L., Lee J.H., Gunzl A.. Trypanosome cdc2-related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of RNA polymerase II subunit RPB1. Mol. Cell. Biol. 2013; **33**:1965–1975.

64. Chen S., Zhou Y., Chen Y., Gu J.. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; **34**:i884–i890.

65. Langmead B., Salzberg S.L.. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 2012; **9**:357–359.

66. 1000 Genome Project Data Processing Subgroup Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.. The sequence alignment/map format and samtools. Bioinformatics. 2009; **25**:2078–2079.

67. Ramirez F., Dundar F., Diehl S., Gruning B.A., Manke T.. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; **42**:W187–W191.

68. Liao Y., Smyth G.K., Shi W.. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; **30**:923–930.

69. Love M.I., Huber W., Anders S.. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; **15**:550.

70. Brenchley R., Tariq H., McElhinney H., Szoor B., Huxley-Jones J., Stevens R., Matthews K., Tabernero L.. The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. BMC Genomics [Electronic Resource]. 2007; **8**:434.

71. Zhang J., Zhang Z., Brew K., Lee E.Y.. Mutational analysis of the catalytic subunit of muscle protein phosphatase-1. Biochemistry. 1996; **35**:6276–6282.

72. Ferreira L.R., Dossin Fde M., Ramos T.C., Freymuller E., Schenkman S.. Active transcription and ultrastructural changes during Trypanosoma cruzi metacyclogenesis. An. Acad. Bras. Cienc. 2008; **80**:157–166.

73. Bates E.J., Knuepfer E., Smith D.F.. Poly(A)-binding protein I of Leishmania: functional analysis and localisation in trypanosomatid parasites. Nucleic Acids Res. 2000; **28**:1211–1220.

74. da Costa Lima T.D., Moura D.M., Reis C.R., Vasconcelos J.R., Ellis L., Carrington M., Figueiredo R.C., de Melo Neto O.P.. Functional characterization of three leishmania poly(a) binding protein homologues with distinct binding properties to RNA and protein partners. Eukaryot. Cell. 2010; **9**:1484–1494.

75. de Melo Neto O.P., da Costa Lima T.D.C., Merlo K.C., Romao T.P., Rocha P.O., Assis L.A., Nascimento L.M., Xavier C.C., Rezende A.M., Reis C.R.S.et al... Phosphorylation and interactions associated with the control of the Leishmania Poly-A Binding Protein 1 (PABP1) function during translation initiation. RNA Biol. 2018; **15**:739–755.

76. Kramer S., Bannerman-Chukualim B., Ellis L., Boulden E.A., Kelly S., Field M.C., Carrington M.. Differential localization of the two T. brucei poly(A) binding proteins to the nucleus and RNP granules suggests binding to distinct mRNA pools. PLoS One. 2013; **8**:e54004.

77. Dean S., Sunter J.D., Wheeler R.J.. TrypTag.Org: a trypanosome genome-wide protein localisation resource. Trends Parasitol. 2017; **33**:80–82.

78. Fong N., Brannan K., Erickson B., Kim H., Cortazar M.A., Sheridan R.M., Nguyen T., Karp S., Bentley D.L.. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. Mol. Cell. 2015; **60**:256–267.

79. Connelly S., Manley J.L.. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 1988; **2**:440–452.

80. Kim M., Krogan N.J., Vasiljeva L., Rando O.J., Nedea E., Greenblatt J.F., Buratowski S.. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 2004; **432**:517–522.

81. West S., Gromak N., Proudfoot N.J.. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature. 2004; **432**:522–525.

82. Kecman T., Kus K., Heo D.H., Duckett K., Birot A., Liberatori S., Mohammed S., Geis-Asteggiante L., Robinson C.V., Vasiljeva L.. Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. Cell Rep. 2018; **25**:259–269.

83. McDonald J.R., Jensen B.C., Sur A., Wong I.L.K., Beverley S.M., Myler P.J.. Localization of epigenetic markers in Leishmania chromatin. Pathogens. 2022; **11**:930.

84. Genest P.A., ter Riet B., Dumas C., Papadopoulou B., van Luenen H.G., Borst P.. Formation of linear inverted repeat amplicons following targeting of an essential gene in Leishmania. Nucleic Acids Res. 2005; **33**:1699–1709.

85. Bertran M.T., Mouilleron S., Zhou Y., Bajaj R., Uliana F., Kumar G.S., van Drogen A., Lee R., Banerjee J.J., Hauri S.et al... ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. Nat. Commun. 2019; **10**:771.

86. Skene-Arnold T.D., Luu H.A., Uhrig R.G., De Wever V., Nimick M., Maynes J., Fong A., James M.N., Trinkle-Mulcahy L., Moorhead G.B.et al... Molecular mechanisms underlying the interaction of protein phosphatase-1c with ASPP proteins. Biochem. J. 2013; **449**:649–659.

87. Cestari I., McLeland-Wieser H., Stuart K.. Nuclear phosphatidylinositol 5-phosphatase is essential for allelic exclusion of variant surface glycoprotein genes in trypanosomes. Mol. Cell. Biol. 2019; **39**:e00395-18.

88. Ingebritsen T.S., Stewart A.A., Cohen P.. The protein phosphatases involved in cellular regulation. 6. Measurement of type-1 and type-2 protein phosphatases in extracts of mammalian tissues; an assessment of their physiological roles. Eur. J. Biochem. 1983; **132**:297–307.

89. Hertz E.P.T., Kruse T., Davey N.E., López-Méndez B., Sigurðsson J.O., Montoya G., Olsen J.V., Nilsson J.. A conserved motif provides binding specificity to the PP2A-B56 phosphatase. Mol. Cell. 2016; **63**:686–695.

90. Brautigan D.L., Shenolikar S.. Protein serine/threonine phosphatases: keys to unlocking regulators and substrates. Annu. Rev. Biochem. 2018; **87**:921–964.

91. Saurin A.T. Kinase and phosphatase cross-talk at the kinetochore. Front. Cell Dev. Biol. 2018; **6**:62.

92. Smith R.J., Cordeiro M.H., Davey N.E., Vallardi G., Ciliberto A., Gross F., Saurin A.T.. PP1 and PP2A use opposite phospho-dependencies to control distinct processes at the kinetochore. Cell Rep. 2019; **28**:2206–2219.

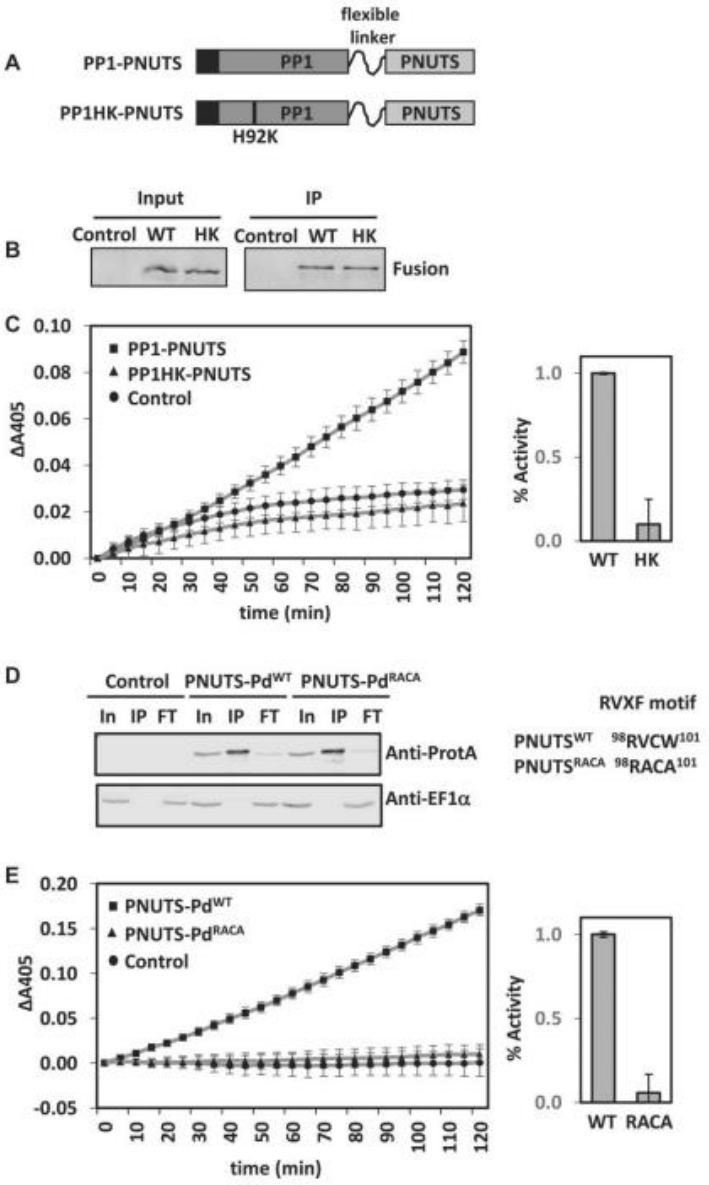


Figure 2.1 The PJW/PP1 complex exhibits phosphatase activity *in vitro*. (A–C) Phosphatase assay for PP1-PNUTS fusions. (**A**) Schematic of Ps-tagged PP1-PNUTS fusions. The black box indicates the N-terminal Protein A, Streptavidin Binding Protein tag. (**B**) PP1-PNUTS fusions were expressed in WT *L. tarentolae* and purified with anti-protein A resin. Control, WT cells; WT, cells expressing WT PP1-PNUTS fusion; HK, expressing H92K PP1 mutant fusion protein. (**C**) PP1-PNUTS fusion immunoprecipitates were assayed for phosphatase activity with pNPP as a substrate. Anti-protein A immunoprecipitates from WT cell extracts were included as a

negative control. Absorbances at 405 nm were measured. The changes in absorbance were plotted against time course. Equal level of input fusion protein was used in the phosphatase assay as shown by the western blot (B). The bar graph represents the relative changes in absorbance after 2-h incubation above the background control, with the value for WT fusion immunoprecipitates set as 1. (D, E) The PJW/PP1 complex exhibits *in vitro* phosphatase activity in a PP1-dependent manner. (**D**) Equal levels of pd-tagged WT PNUTS ($PNUTS^{WT}$) or PP1-8e binding deficient $PNUTS^{RACA}$ proteins were purified from cell extracts by anti-protein A affinity, as shown by the western blot with anti-protein A. Anti-EF1a serves as a loading control. Right, the conserved RVXF motif and mutations in $PNUTS^{RACA}$. (**E**) The immunoprecipitates were assayed for phosphatase activity against pNPP as described above. Error bars indicate standard error of the mean (SEM) from three experiments. The blots are representative of three independent experiments. In, Input; IP, bound fraction; FT, unbound fraction.

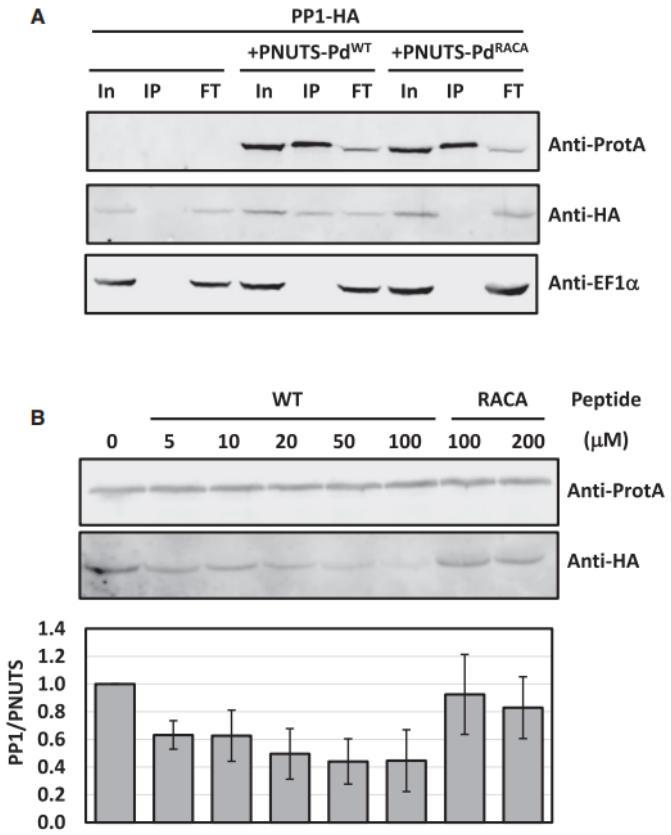


Figure 2.2 PP1-8e binds PNUTS via the RVXF motif. Analysis of the interactions between PNUTS and PP1-8e *in vivo*. (**A**) Co-immunoprecipitation analysis showing that V97 and W99 of the LtPNUTS 'RVXF' interaction motif is essential for interaction between PP1-8e and PNUTS. PP1-8e was endogenously tagged with HA tag, and Pd-tagged wild type or mutant PNUTS (with alanine substitutions for V and W in the RVXF motif) was over-expressed from the pSNSAP1 vector. Cell extracts from the indicated cell lines were purified by anti-protein A affinity resin and analyzed by western blot with anti-protein A and anti-HA. In; input (equivalent to the amount of protein added to the IP reaction mixture), IP; precipitated immunocomplexes, FT; flow through or non-bound supernatant. EF1α provides a loading control and negative control for the IP. (**B**) A peptide based on the PP1 interaction motif RVXF displaces PP1 from PNUTS-Pb bound affinity matrix. Pd-tagged PNUTS was purified from cell extract with anti-protein A

affinity resin, and IPs were incubated with indicated concentrations of synthetic 25-mer peptides with either WT (RVCW) or mutated (RACA) RVXF motif for 30 min. Dissociation of PP1 from the PNUTS IP was examined by western blotting of sedimented fractions. The blots shown are representative of three independent experiments. PP1 protein levels were quantified by densitometric analysis and normalized to PNUTS protein level in the IPs. The bar graph represents the mean ± SD for three independent experiments.

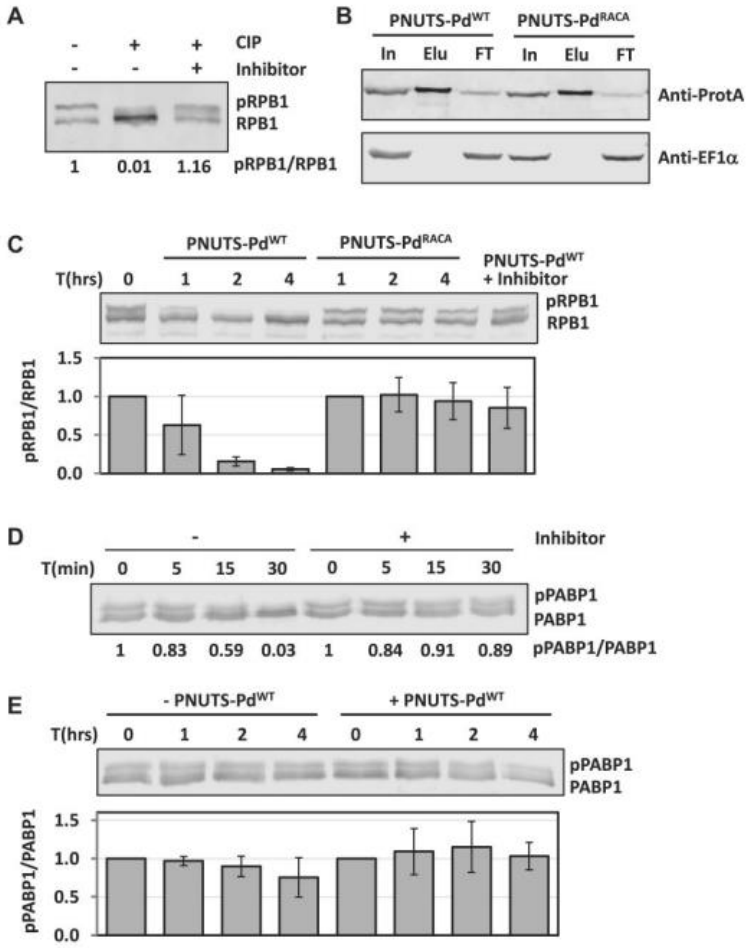


Figure 2.3 Pol II is a substrate for PP1-8e. PJW/PP1 subunit PP1-8e dephosphorylates Pol II *in vitro*. (**A**) LtRNA Pol II is phosphorylated. Western blot of parasite lysate expressing RPB1 endogenously tagged with HA-tag treated without (-) or with (+) calf intestinal phosphatase (CIP) in the absence or presence (+Inhibitor) of a phosphatase inhibitor cocktail (PhosSTOP). Blots were probed with anti-HA. Phosphorylated (pRPB1) and dephosphorylated (RPB1) forms of Rpb1 are indicated. Ratio of pRPB1/RPB1 from densitometric quantification of the blot is indicated below; with the ratio in initial extracts arbitrarily set to 1. (**B**) Purification of the PJW/PP1 complex. Pd-tagged WT PNUTS or RVXF mutant PNUTS (RACA) were purified with anti-FLAG affinity resin and eluted via 3× FLAG peptide and processed for Western blotting using anti-HA and anti-EF1α antibodies. In, Input; Elu, FLAG peptide elutant; FT, flow

through unbound fraction. (**C**) Pol II *in vitro* phosphatase assay. Equal inputs of immunoprecipitates from WT and mutant PNUTS, as shown in (B), were incubated with HA-immobilized Rpb1 for the indicated times. The last lane represents *in vitro* incubation of WT PNUTS-PP1-8e complex with Rpb1 for 4 hours with the addition of phosphatase inhibitor. HA-Immobilized Rpb1 proteins were processed for Western blotting using anti-HA. The blots shown are representative of at least three independent experiments. Densitometric quantification of pRPB1 and RPB1 forms from three independent *in vitro* phosphatase assays is shown below. The pRPB1/RPB1 ratio at T0 was arbitrarily set to 1. The bar graph represents the mean ± SD for three independent experiments. (**D**) LtPABP1 is phosphorylated. Western blot of parasite lysate expressing PABP1 endogenously tagged with HA-tag lysed in buffer II with 1% NP-40 and incubated at 30°C with or without phosphatase inhibitor for the indicated times. Ratio of pPABP1/PABP1 from densitometric quantification of the blot is indicated below; with the ratio in initial extracts (T0) arbitrarily set to 1. **(E)** PABP1 *in vitro* phosphatase assay. Experiments were performed and analyzed as in (C). HA-Immobilized PABP1 was incubated with and without immunoprecipitates from WT PNUTS for the indicated times.

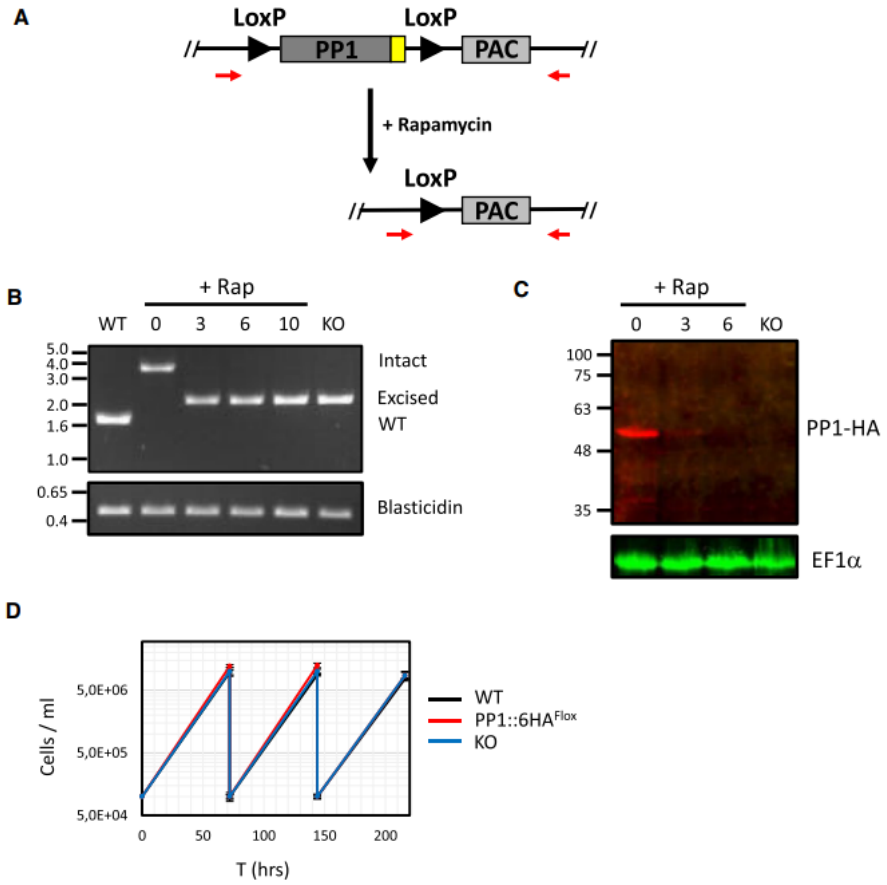


Figure 2.4 PP1-8e is non-essential in *L. major*. (**A**) Illustration of PP1-HA^{Flox} excision catalyzed by DiCre, as induced by rapamycin. Primers used in PCR analysis of genomic DNA, are shown in red. Please see material and methods (and Supplementary Figure 2.S3) for details on generation of the PP1-HA^{Flox} cell line. (**B**) PCR analysis of genomic DNA extracted from the PP1-HA^{Flox} without the addition (0) and 3–10 days after addition (+RAP) of rapamycin, leading to DiCre induction. Approximated annealing positions for the primers are shown in (A). DNA from wildtype cells (WT) is included as a control. KO clone obtained after 15 days growth in rapamycin. (Intact) and (Excised), PP1^{Flox} and PP1^{Flox} after excision, respectively. (**C**) Western blotting analysis of whole cell extracts of the PP1-HA^{Flox} without the addition (0) and 3–6 days after addition (+RAP) of rapamycin, leading to DiCre induction; extracts were probed with anti-HA antiserum and anti-EF1α was used as loading control (protein sizes are indicated, kDa). (**D**)

Growth curves of wild-type (WT, black), PP1-HA^{Flox} (red) and the PP1-8e KO clone 5 (blue); cells were seeded at ~10^5 cells/ml at day 0 and diluted back to that density every 3 days; cell density was assessed every 24 h, and error bars depict standard deviation from three replicate experiments.

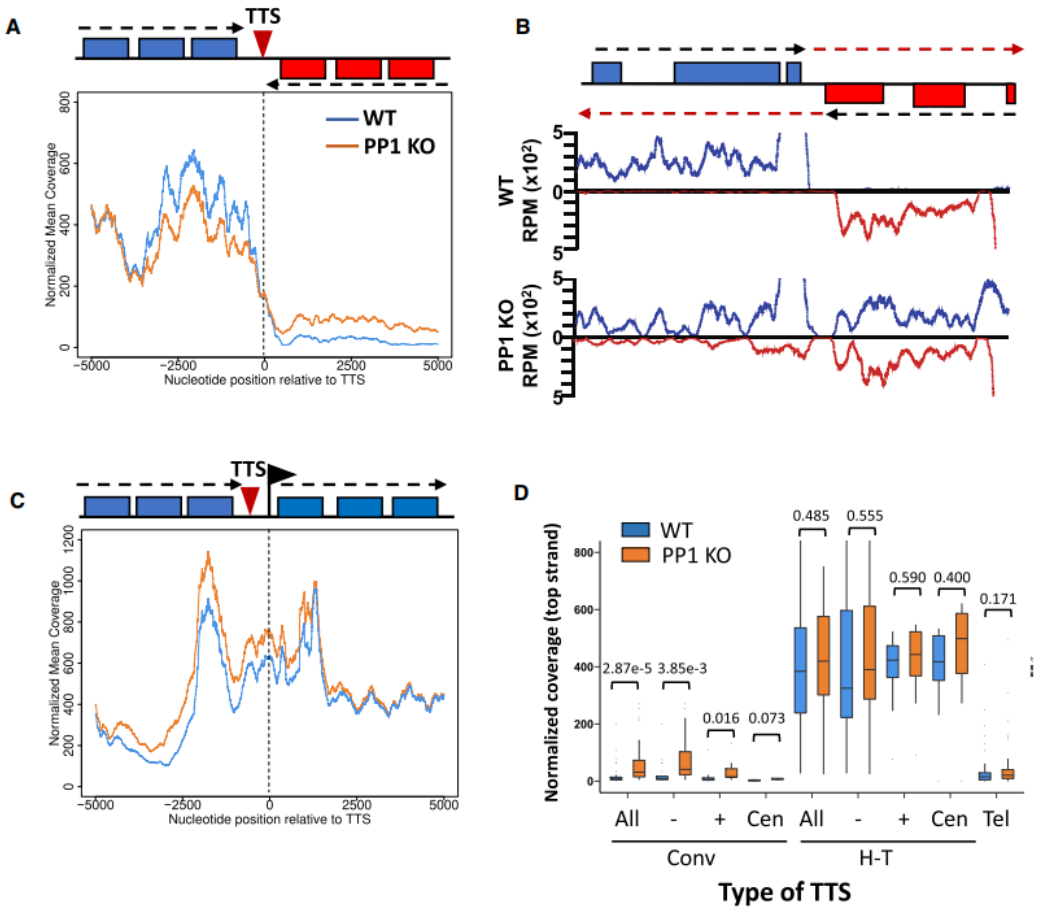


Figure 2.5 Depletion of PP1 results in readthrough at transcription termination sites. (**A**) Mean top strand coverage at each nucleotide position (bp) in the 10 kbp surrounding the transcription termination site (TTS) at all 39 cSSRs for WT cells (blue line) or PP1 KO (orange line). The schematic represents the protein-coding genes associated with each strand at an 'average' convergent TTS (cTTS). Plots are orientated that transcription proceeds from the left and terminates at '0', with the top strand being the coding strand on the left side of the TTS. Dashed arrows represent transcription direction. (**B**) A region on chromosome 22 from 503–513kb where J (36) and PP1 regulate transcription termination at a cSSR (cSSR 22.3) is shown. Top; map of the cSSR. ORFs are shown with the top strand in blue and the bottom strand in red. The red arrow indicates read through transcription following the loss of J (36) and loss of PP1.

Bottom; Strand-specific mRNA-seq reads from the indicated cell lines are mapped. Reads that mapped to the top strand are shown in blue and reads that mapped to the bottom strand in red. (**C**) Mean top strand coverage at each nucleotide position in the 10 kb surrounding the 52 TTSs between head-to-tail (unidirectional) PTUs. The schematic represents the protein-coding genes associated with each strand at an 'average' head-to-tail (HT) TTS. HT regions that are transcribed in the opposite direction of the diagram are reoriented so that the transcribed genes are represented on the top strand. The flag indicates the transcription start site for the downstream gene cluster as indicated by H3 acetylation localization (1). The region spans 5 kb flanking the transcription start site downstream of the TTS. (**D**) Box-and-whisker plots showing the median top strand coverage in the 5-kb region downstream of all 146 TTSs. Separate plots are shown for all the TTSs at cSSRs (Conv All), 24 TTSs at cSSRs that lack an RNA gene (Conv –), 8 TTSs at cSSRs that contain one or more RNA genes (Conv +), 7 TTSs at cSSRs adjacent to a centromere (Cent), TTSs between all head-to-tail PTUs (H-T All), 27 TTSs between head-to-tail PTUs that lack an RNA gene (H-T –),12 TTSs between head-to-tail PTUs that contain one or more RNA genes (H-T +), 13 TTSs between head-to-tail PTUs adjacent to a centromere (H-T Cen), and 55 TTSs at telomeres (Tel). Multiple comparisons were conducted by wilcoxon test. *P* values were presented on top of each compared group.

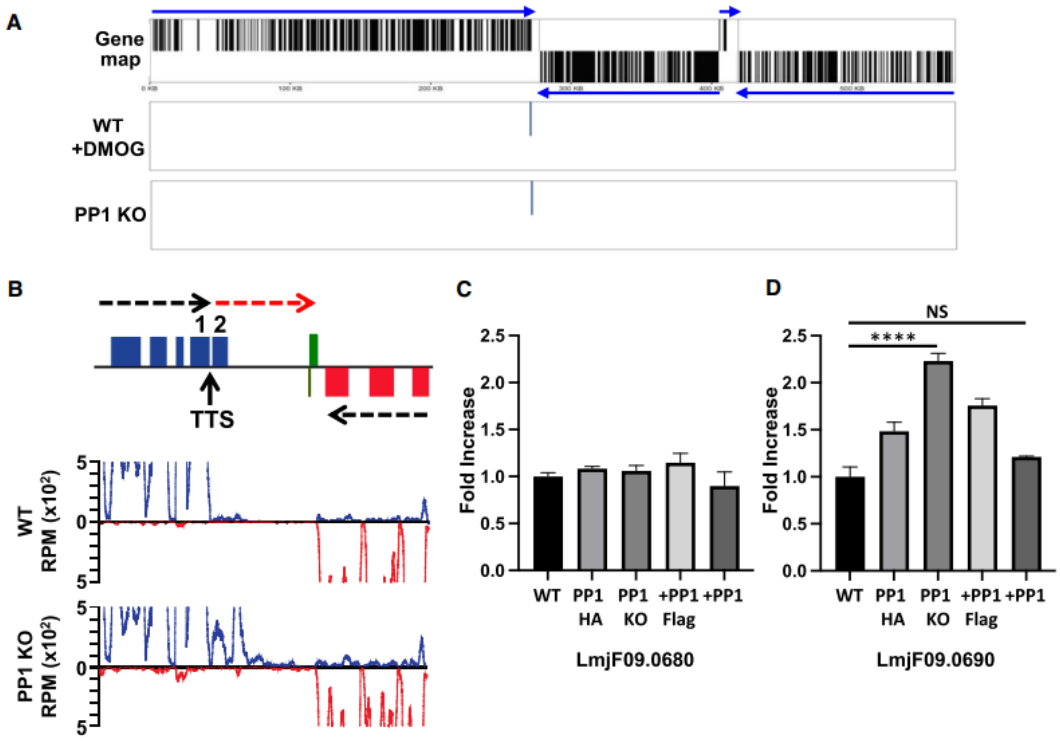


Figure 2.6 Decreased efficiency of RNAP II termination and increased gene expression following the loss of PP1. (**A**) Gene map of chromosome 9 is shown. mRNA coding genes on the top strand are indicated by black lines in the top half of the panel, bottom strand by a line in the bottom half. Genes on the top strand are transcribed from left to right and those on the bottom strand are transcribed from right to left, indicated by blue arrows. Panel below (WT + DMOG and PP1) indicates the location of the single mRNA (LmjF09.0690) found upregulated by at least two-fold in WT cells treated with DMOG relative to WT (36) and PP1 KO relative to WT. No other expression changes (up or downregulated) were detected. (**B**) A region on chromosome 9 from 263–285 kb where J (36) and PP1 regulates transcription termination and gene expression at a cSSR is shown (cSSR 9.1). Top; map of the cSSR. The vertical arrow indicates the proposed TTS. ORFs are shown with the top strand in blue and the bottom strand in red. The red arrow indicates read through transcription following the loss of J (36) and loss of PP1. Green boxes indicate RNAP III transcribed genes (tRNA and 5S rRNA).

The numbered genes (1) and (2) that flank the TTS refer to LmjF09.0680 and LmjF09.0690, respectively. Bottom; Strand-specific mRNA-seq reads from the indicated cell lines are mapped. Reads that mapped to the top strand are shown in blue and reads that mapped to the bottom strand in red. (C and D) Gene expression changes for the genes flanking the TTS and indicated (numbered) in the ORF map in B were confirmed by qRT-PCR in the indicated cell lines. mRNA levels in WT set to one. Error bars indicate the standard deviation between two biological replicates analyzed in triplicate. WT; wild-type, PP1 HA; PP1-HA^{Flox}, PP1 KO; PP1-HA^{Flox} excised, +PP1 Flag; PP1-KO + PP1-Flag tagged, +PP1,; PP1-KO + PP1 untagged. P values were calculated using Tukey's multiple comparisons test, ****P values < 0.0001; NS, not significant.

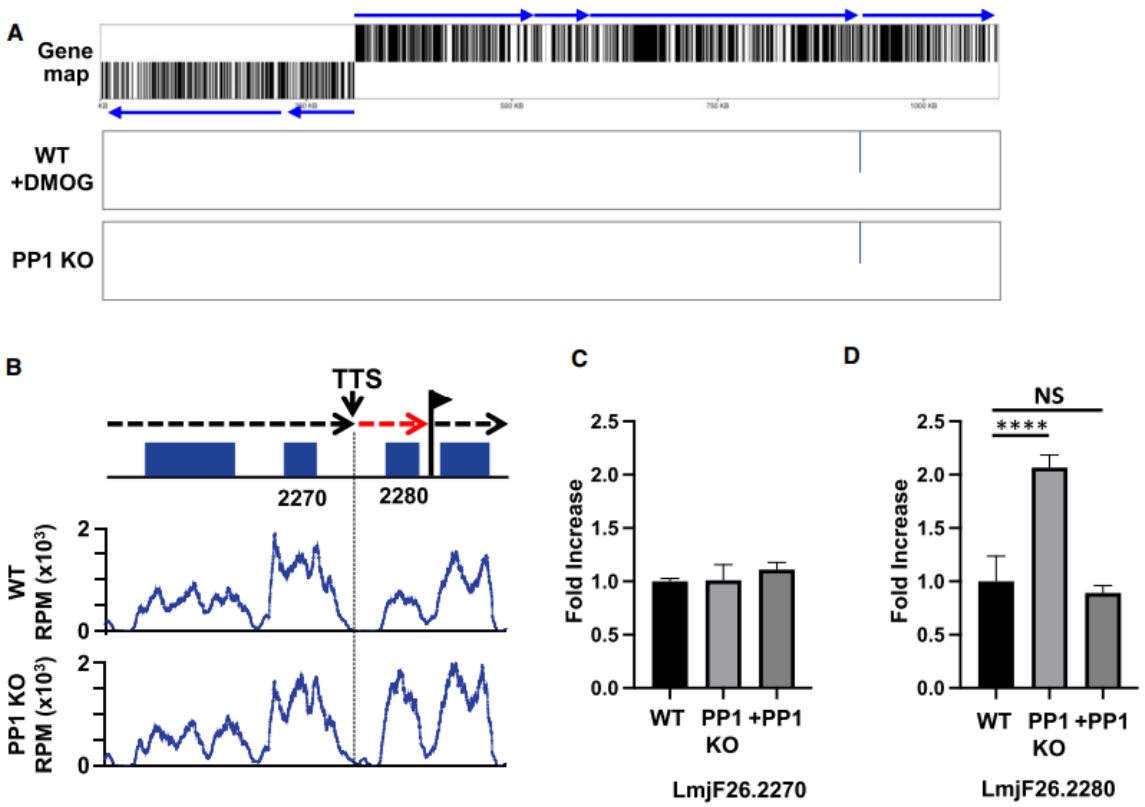


Figure 2.7 PP1 regulates RNAP II termination and gene expression at head-tail regions within gene clusters. (**A**) Gene map of chromosome 26 is shown where loss of J leads to upregulation of a single mRNA at the end of a gene cluster at a head-tail region. Labeling is as described in Figure 2.6. (**B**) Top; ORFs are plotted for the head-tail region on chromosome 26 from 912–922 kb, as described in Figure 2.3. Base J localizes at the transcription termination site (TTS). The vertical arrow indicates the proposed TTS (36). The black dashed arrow above the map indicates the direction of transcription and the dashed red arrow indicates read through transcription past the TTS. The upregulated gene, downstream of the TTS, is LmjF26.2280 (2280). The flag indicates the transcription start site for the downstream gene cluster as indicated by H3 acetylation localization (1). Bottom: Plot of the mRNA-seq data for the region above, as described in Figure 2.6. (**C** and **D**) Gene expression changes for the genes flanking the TTS and indicated in the ORF map in B were confirmed by RT-qPCR in the indicated cell lines, as

described in Figure 2.6 (C and D). Error bars indicate the standard deviation between two biological replicates analyzed in triplicate. P values were calculated using Tukey's multiple comparisons test, ****P values < 0.0001; NS, not significant.

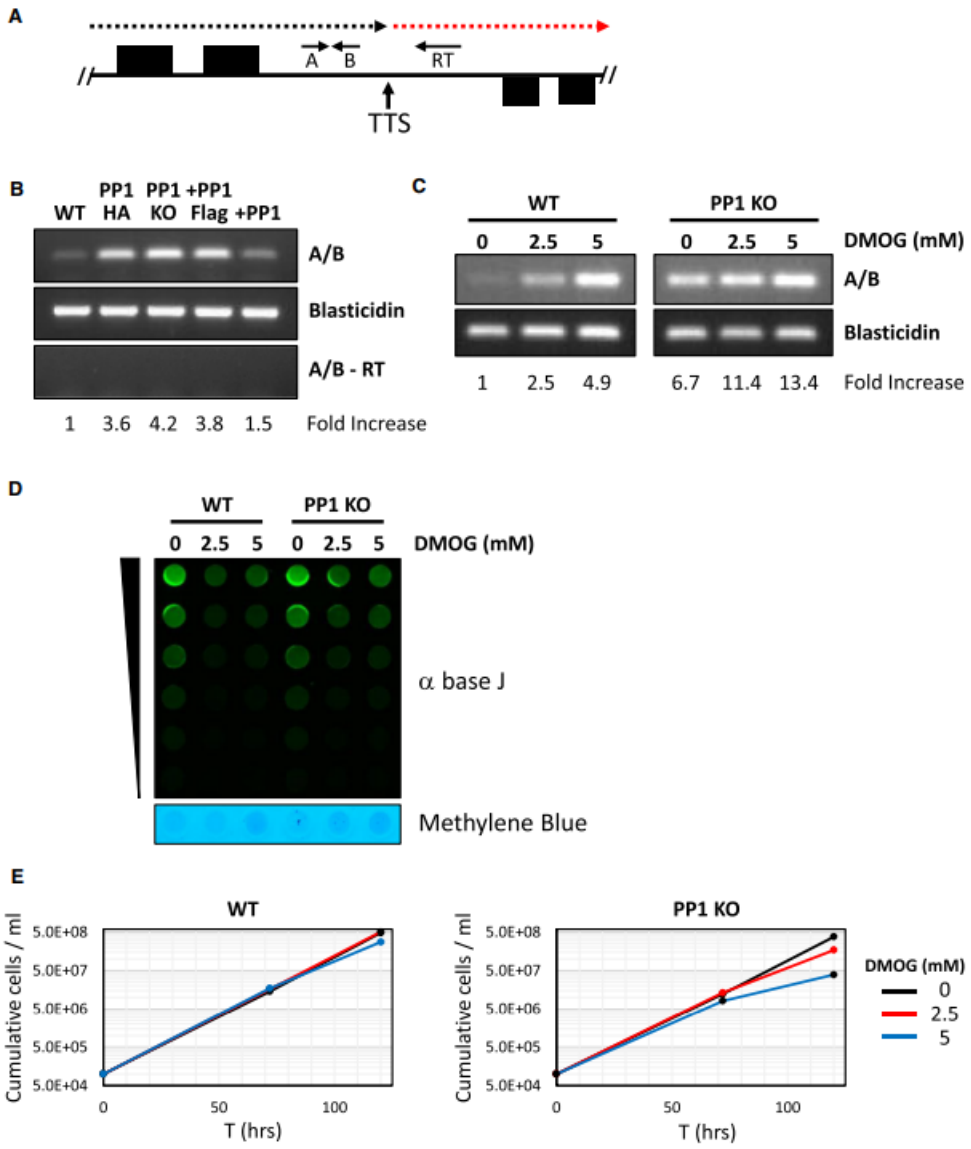


Figure 2.8 PP1 regulates nascent readthrough RNA synthesis. (**A**) Schematic representation (not to scale) of cSSR 22.3 illustrating the nascent RNA species assayed by RT-PCR. The dashed red arrow indicates read through transcription past the transcription termination site (TTS). Arrows indicate the location of primers utilized for strand-specific RT-PCR analysis. (**B**) Strand-specific RT-PCR analysis of readthrough transcription. Read through transcription on the top strand for the indicated cell lines was quantitated by performing site-specific cDNA synthesis using primer RT illustrated in the diagram above, followed by PCR using primers A and B.

Abundance was normalized using blasticidin marker (a gene specific primer against blasticidin was added to the same cDNA synthesis reaction with primer RT, followed by PCR using blasticidin primers). Fold increase in nascent readthrough RNA species relative to background levels in WT is based on A + B qPCR analysis, normalized to blasticidin qPCR. (**C**) Strand-specific RT-PCR analysis of readthrough transcription as in (B) for the indicated cell lines grown in the presence of DMOG. (**D**) Serially diluted genomic DNA from wild-type and PP1-8e KO cells, grown with 0, 2.5mM and 5mM DMOG for 5 days, were incubated with anti J antisera. DNA loading was verified by Methylene Blue staining. (**E**) Growth curves of wild-type and PP1-8e KO cells in the presence of 0, 2.5 mM and 5 mM DMOG. Cumulative cell numbers were calculated after passaging 10^5 cells/ml after 3 days of initial growth in medium with fresh DMOG.

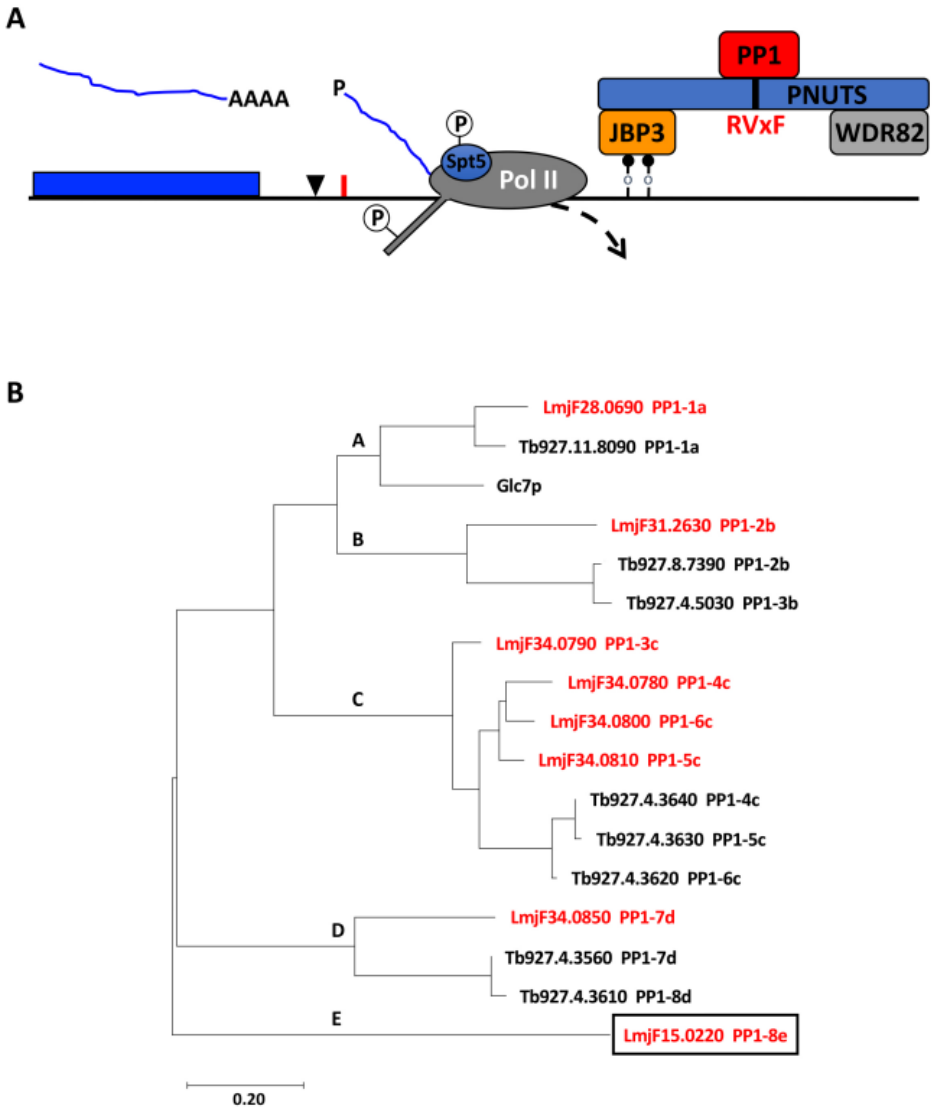


Figure 2.S1 Mechanism of transcription termination and phylogenic analysis of PP1 phosphatase in Leishmania and Trypanosomes. A. Model of the mechanism of termination. We have previously proposed a model where readthrough transcription at the end of polycistronic transcription units is controlled by the PJW/PP1 complex in kinetoplastids (38). According to this model, the complex is recruited to termination sites, at least partially, due to base J-JBP3 interactions. H3V and H4V localized at termination sites may play an additional role in complex localization via interactions with Wdr82 as well as by regulating base J synthesis. PNUTS is a scaffolding protein for the complex and regulates PP1 function via the

PP1 binding RVxF motif. PP1 dephosphorylation of Spt5 and RNA Pol II is critical for termination and dissociation of Pol II from the DNA template. The modified DNA base J are indicated by the lollipop diagram. The poly(A) site of the final gene prior to the termination site is indicated by the black arrow head, with the splice acceptor indicated by the red line. 3' end formation of mature RNAs is thought to occur via trans-splicing of the spliced leader sequence at the 5' end of the downstream gene followed by polyadenylation at the upstream gene.
This coupling, by preventing the generation of the 5' phosphate on the pol II engaged transcript, would prevent "torpedo" stimulated Pol II termination within the polycistronic gene array. We propose an additional function of the PJW/PP1 complex is to block trans splicing at termination sites allowing the link of 3' end formation to Pol II termination as described by the "torpedo model". B. Phylogenic tree for PP1 proteins from Leishmania major, Trypanosoma brucei and Saccharomyces cerevisiae. The tree was obtained as described in material and methods. The five clades of PP1 (A-E) are shown. Genes and species for proteins used in the alignment are *Leishmania major* (Gene IDs LmjF15.0220, LmjF28.0690, LmjF31.2630, LmjF34.0780, LmjF34.0790, LmjF34.0800, LmjF34.0810, and LmjF34.0850), *Trypanosoma brucei* (Gene IDs Tb927.4.3560, Tb927.4.3610, Tb927.4.3620, Tb927.4.3630, Tb927.4.3640, Tb927.4.5030, Tb927.8.7390, and Tb927.11.8090), and *Saccharomyces cerevisiae* (SGD:S000000935).

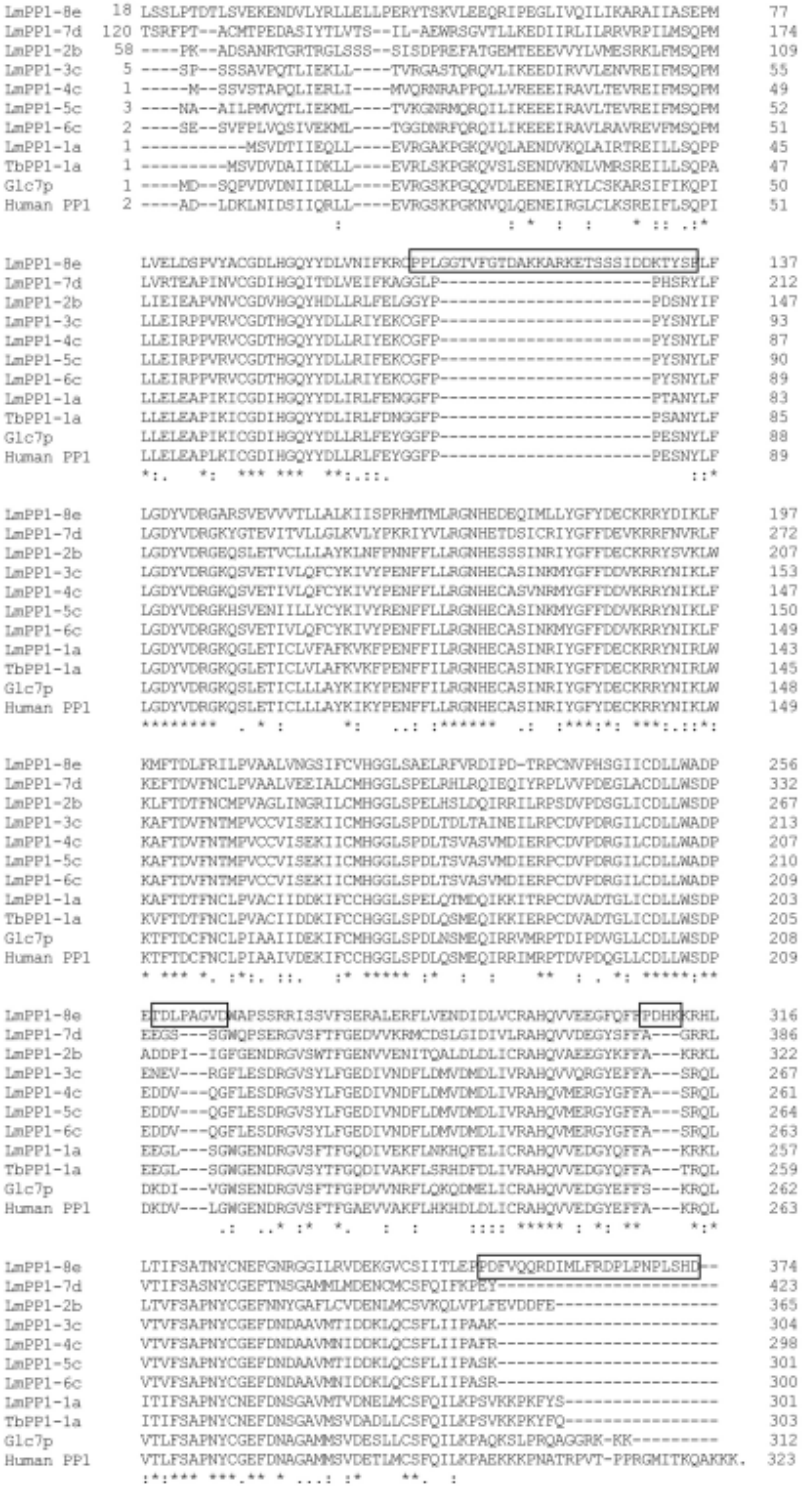


Figure 2.S2 PP1 sequence alignments from Leishmania major and yeast PP1 (Glc7p) and Human PP1 alpha (Human PP1). Sequences unique to LmPP1-8e are boxed. The six manganese-coordinating residues are underlined in PP1-8e.

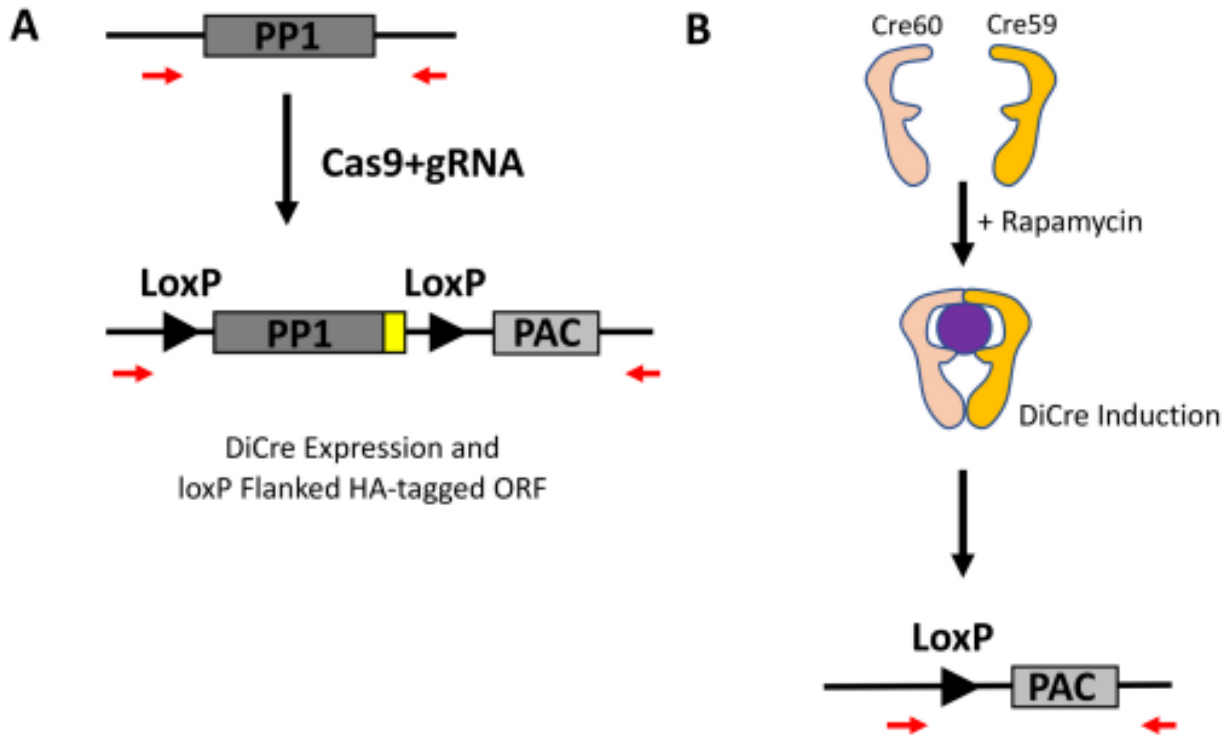


Figure 2.S3 CRISPR/Cas9 and DiCre assessment of PP1 function by conditional excision.

(A) A cell line was engineered to express both Cas9 and DiCre. Cas9 was used to rapidly replace all copies the PP1-8e gene by a version of the PP1 gene flanked by LoxP sites (PP1$^{`}$). (B) Illustration of PP1^{Flox} excision catalyzed by DiCre, as induced by rapamycin. Primers used in PCR analysis of genomic DNA, are shown in red.

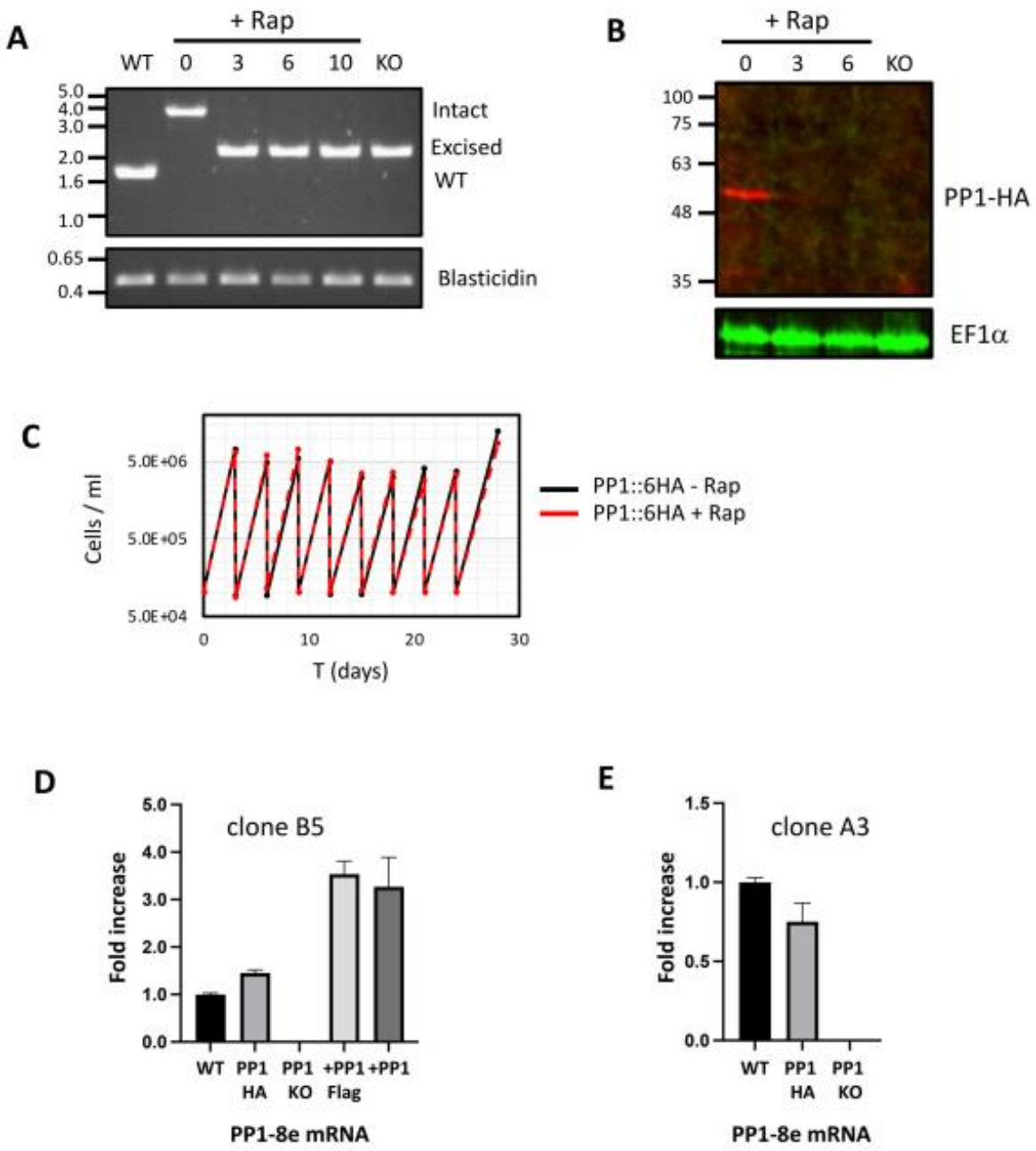


Figure 2.S4 PP1-8e KO clone A3 (A-C) Analysis of PP1-HA^{Flox} excision catalyzed by DiCre at the DNA and protein level and effects on cell growth as described in Figure 2.4. (D and E) Levels of PP1-8e mRNA measured by qRT-PCR in the indicated cell lines for clone B5 (D) and clone A3 (E). WT levels of PP1-8e is set to one. Error bars depict standard deviation from three replicate experiments.

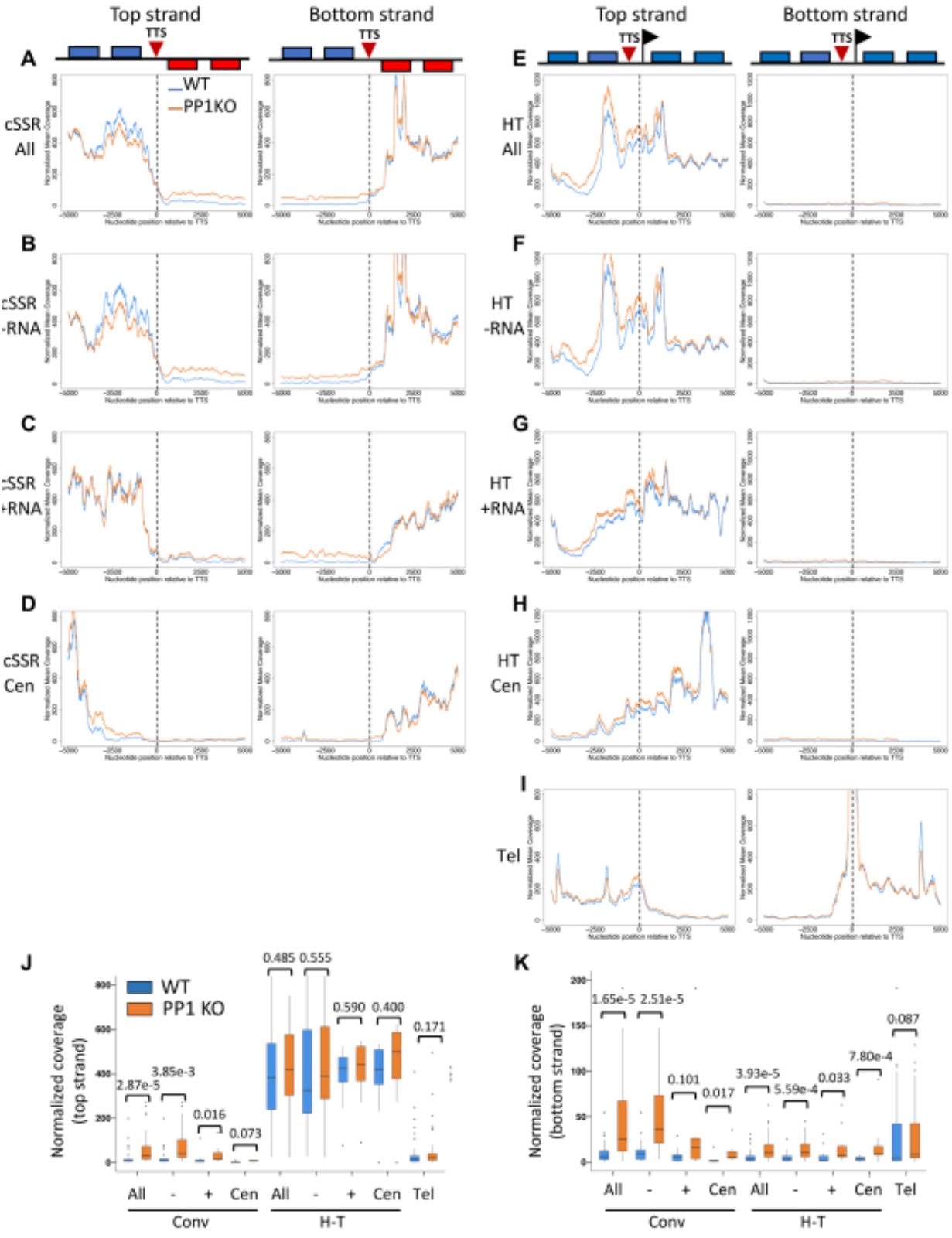

Top strand
Bottom strand
TTS
A
WT
PP1KO
cSSR
All
B
C
D
cSSR
Cen
E
HT
All
F
HT
-RNA
G
HT
+RNA
H
HT
Cen
I
Tel
J
WT
PP1 KO
Normalized coverage
(top strand)
2.87e-5
3.85e-3
0.016
0.073
0.485
0.555
0.590
0.400
0.171
All
-
+
Cen
Tel
Conv
H-T
K
Normalized coverage
(bottom strand)
1.65e-5
2.51e-5
0.101
0.017
3.93e-5
5.59e-4
0.033
7.80e-4
0.087

Figure 2.S5 Effect of PP1-8e KO on transcription termination.

The normalized read counts are shown for the 10 kb surrounding TTS for WT cells (blue line) or PP1 KO (orange line). Plots are orientated that transcription proceeds from the left and terminates at "0", with the top strand being the coding strand on the left side of the TTS. Panels on the left depict reads mapping to the top strand, and panels on the right depict reads mapping to the bottom strand. For head-to-tail TTSs, units that are transcribed in the opposite direction (right to left) are reoriented so that the transcribed genes are represented on the top strand. (A) All TTSs at cSSRs. (B) TTSs at cSSRs lacking either an RNA gene or a centromere. (C) TTSs at cSSRs with an adjacent Pol III RNA gene. (D) TTSs at cSSRs adjacent to a centromere. (E) TTS between all head-to-tail PTUs (HT). (F) TTSs at HT lacking either an RNA gene or a centromere. (G) TTSs at HT containing an RNA gene. (H) TTSs at HT adjacent to a centromere (Cen). (I) TTSs located at telomeres (Tel). (J-K) Box-and-whisker plots showing the median top strand (J) and bottom strand (K) coverage in the 5-kb region downstream of all 146 TTSs. Separate plots are shown for the different types of TTSs as described in Figure 2.5D. Multiple comparisons were conducted by wilcoxon test. *P* values were presented on top of each compared group.

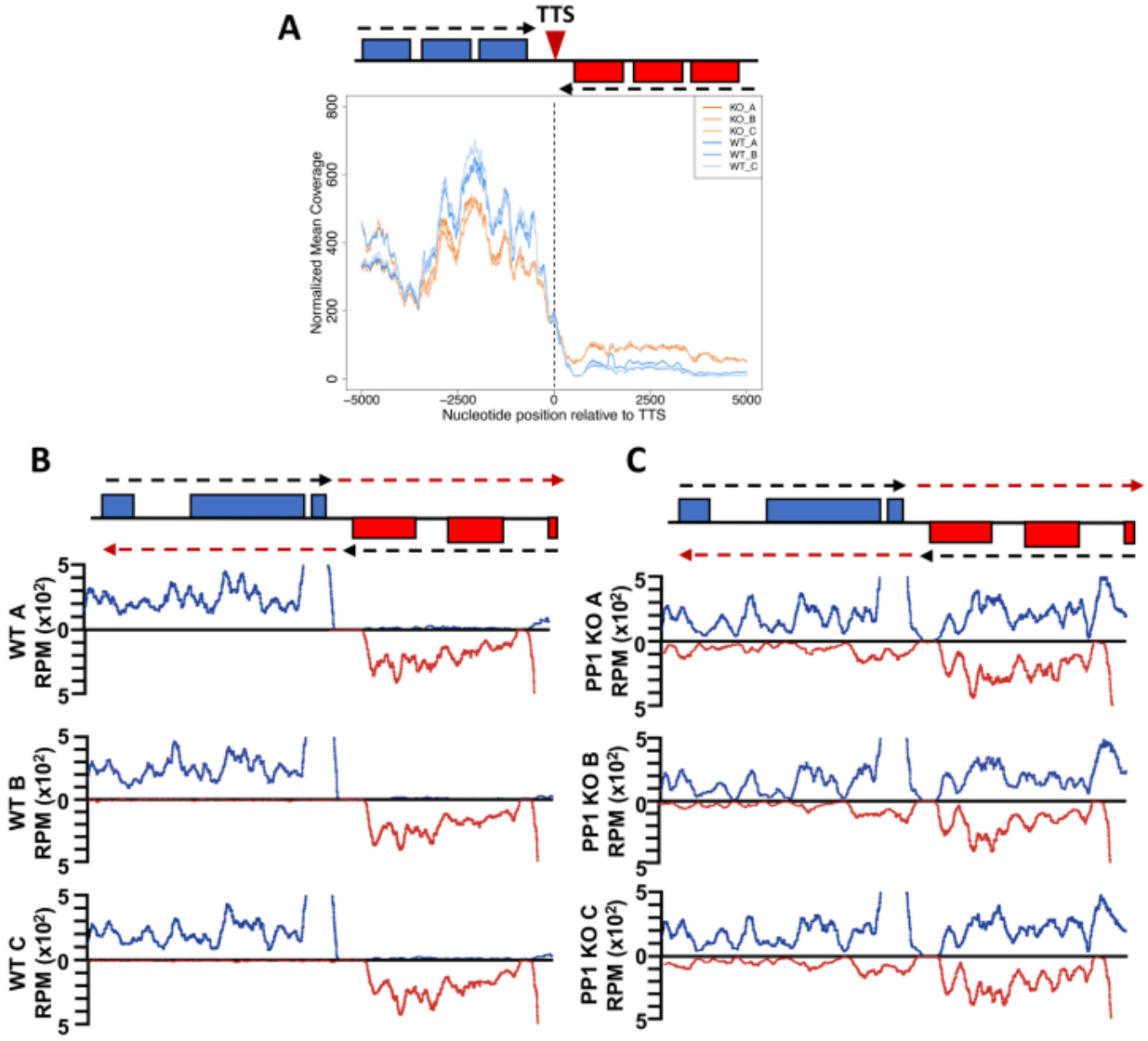


Figure 2.S6 Depletion of PP1 results in readthrough at cSSRs. (A) Mean top strand coverage at each nucleotide position in the 10 kb surrounding the transcription termination site (TTS) at all 39 cSSRs for WT cells (blue line) or PP1 KO (orange line), three biological replicates each, as described in Figure 2.5A. (B-C) Strand-specific mRNA-seq reads from the three replicates of WT (B) and PP1 KO (C) are mapped to cSSR 22.3 as described in Figure 2.5B.

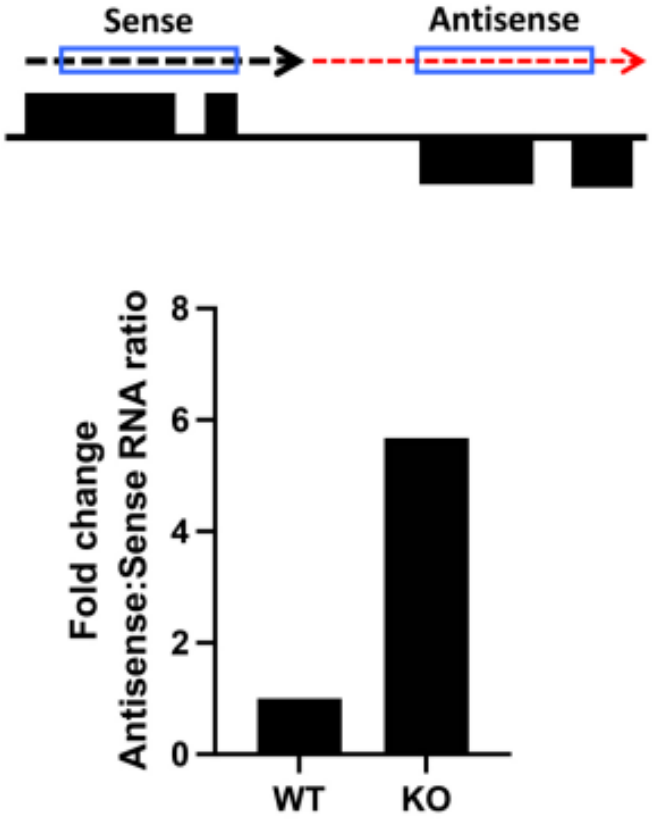


Figure 2.S7 Quantitation of read through at cSSRs by RNA-seq. Top: Diagram of a cSSR (not to scale), with genes indicated by black boxes. The black arrow indicates the direction of transcription on the top strand and the red arrow indicates read through transcription past the termination site. Blue boxes show the 10kb windows flanking the cSSR used to determine the antisense:sense RNA-seq RPM ratio. 10kb windows were set using the genomic positions for all cSSRs in the *L. major* genome that lack an RNA gene. Analysis of RNA-seq reads on the bottom strand was performed similarly. Bottom: The antisense:sense RNA-seq RPM ratio was determined at all cSSRs (excluding cSSRs 9.2 and 22.2, in which case the flanking gene cluster was less than 10kb, and 28.2, which is the only cSSR without J in WT *L. major* and thus read through occurs even in WT cells (33)). WT was set to one. Fold change relative to WT in the PP1 KO cells is plotted.

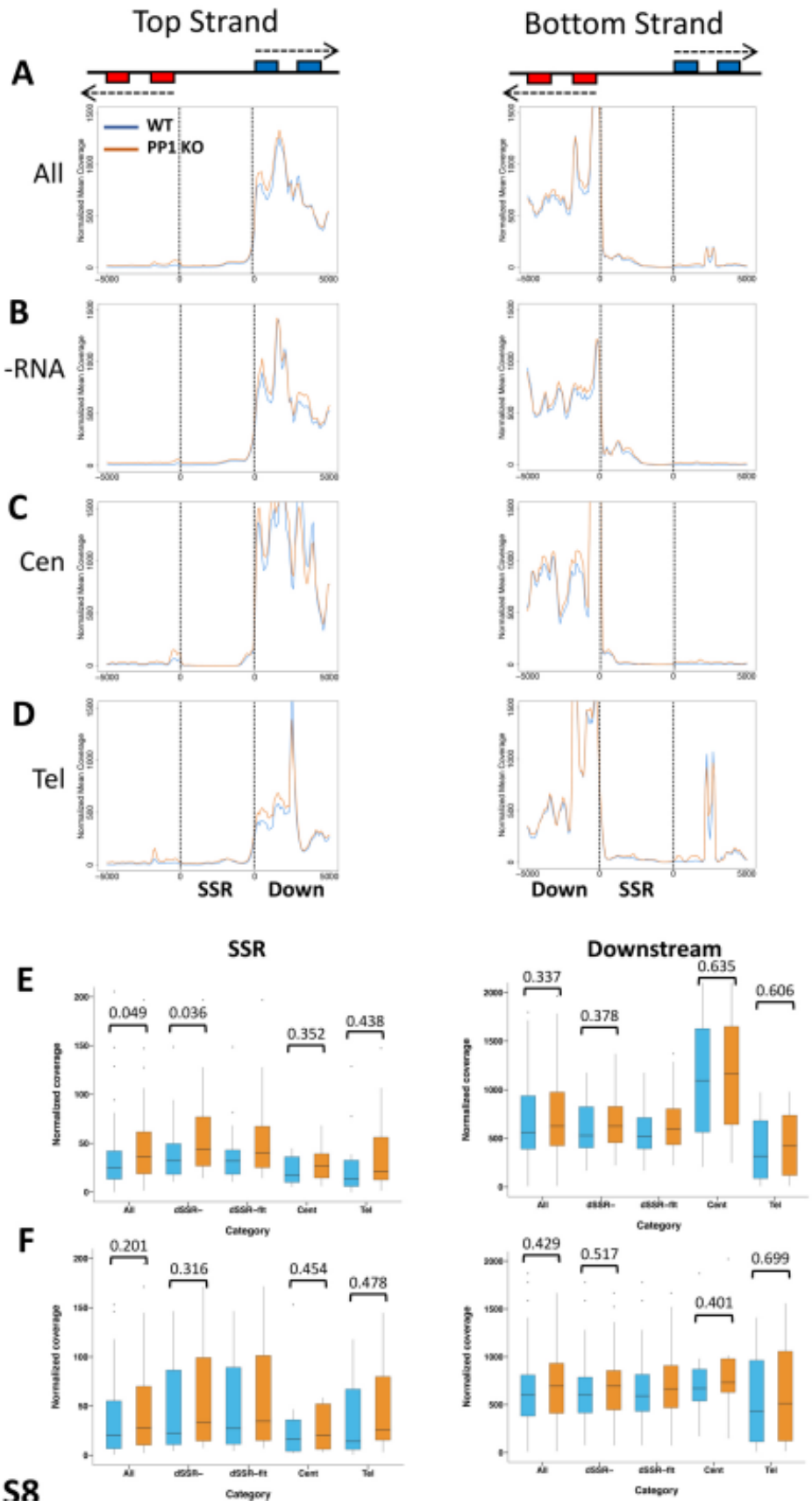

Top Strand
Bottom Strand
A
All
WT
PP1 KO
B
-RNA
C
Cen
D
Tel
SSR
Down
Down
SSR
E
SSR
Downstream
0.049
0.036
0.352
0.438
0.337
0.378
0.635
0.606
F
0.201
0.316
0.454
0.478
0.429
0.517
0.401
0.699
Normalized coverage
Category

Figure S8

Figure 2.S8 Effect of deletion of PP1-8e on transcription initiation at dSSRs. An analysis similar to the one for TTS in Figure 2.S5 is shown, displaying the normalized read coverage for the dSSR and 5kb region downstream of the transcription start sites (position "zero"). Arrows indicate the direction of transcription. Panels on the left depict reads mapping to the top strand, and panels on the right depict reads mapping to the bottom strand. (A) All TSSs. (B) TSSs in dSSRs lacking either an RNA gene or a centromere. (C)TSSs adjacent to a centromere. (D) TSSs adjacent to a telomere. Two dSSRs that contain rRNA gene arrays were not included for simplicity. (E and F) Box-and-whisker plots showing median coverage in the dSSR region and the 5-kB downstream of all TSSs (All), TSSs that lack an RNA gene in the SSR (-RNA), TSSs adjacent to a centromere (Cen) and TSSs at telomeres (Tel). (E) Top strand and (F) bottom strand.

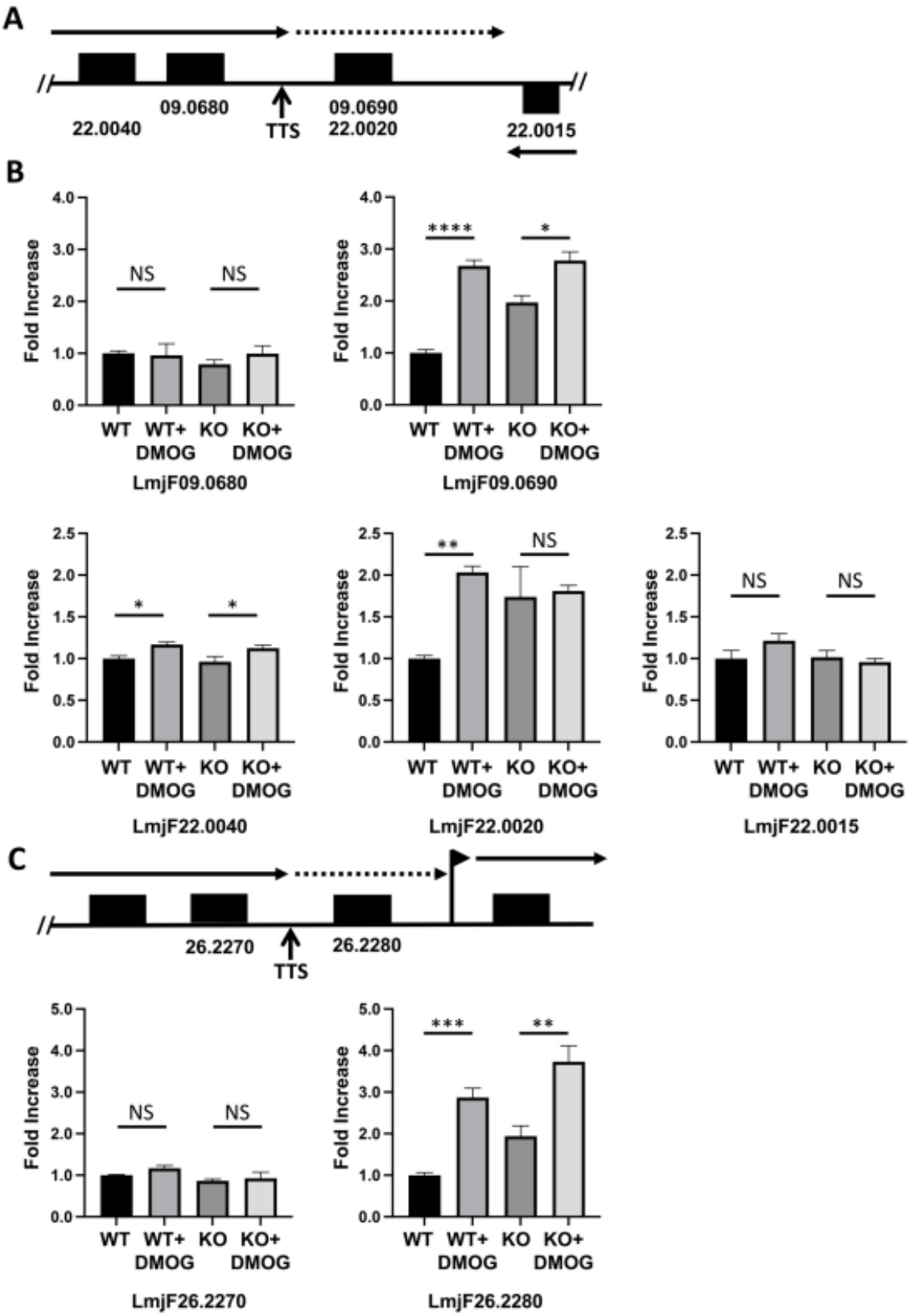

A
09.0680
09.0690
22.0020
22.0040
TTS
22.0015
B
Fold Increase
NS
NS
****
*
WT
WT+ DMOG
KO
KO+ DMOG
LmjF09.0680
LmjF09.0690
**
LmjF22.0040
LmjF22.0020
LmjF22.0015
C
26.2270
26.2280
TTS
***
LmjF26.2270
LmjF26.2280

Figure 2.S9 Loss of base J synthesis in the PP1-8e KO leads to further defects in transcription termination. (A) Schematic representation (not to scale) of two termination sites on chromosome 9 and chromosome 22. The dashed arrow indicates read through transcription past the transcription termination site (TTS). Genes that will be analyzed by RT-PCR are labelled. (B) Fold change in transcript abundance determined by RT-PCR of the indicated cell lines, with DMSO treated WT set to 1. WT cells grown in presence of 5mM DMOG, WT+DMOG; PP1-8e KO grown in presence of 5mM DMOG, KO+DMOG. Error bars indicate the standard deviation between two biological replicates analyzed in triplicate. P values were calculated using Tukey's multiple comparisons test. *, p values <0.05; **, p values <0.01; ***, p values <0.001; ****, p values <0.0001; NS, not significant.

(C) Gene expression changes measured by RT-PCR for the H-T site on chromosome 26 (see Figure 2.7) is analyzed as above.

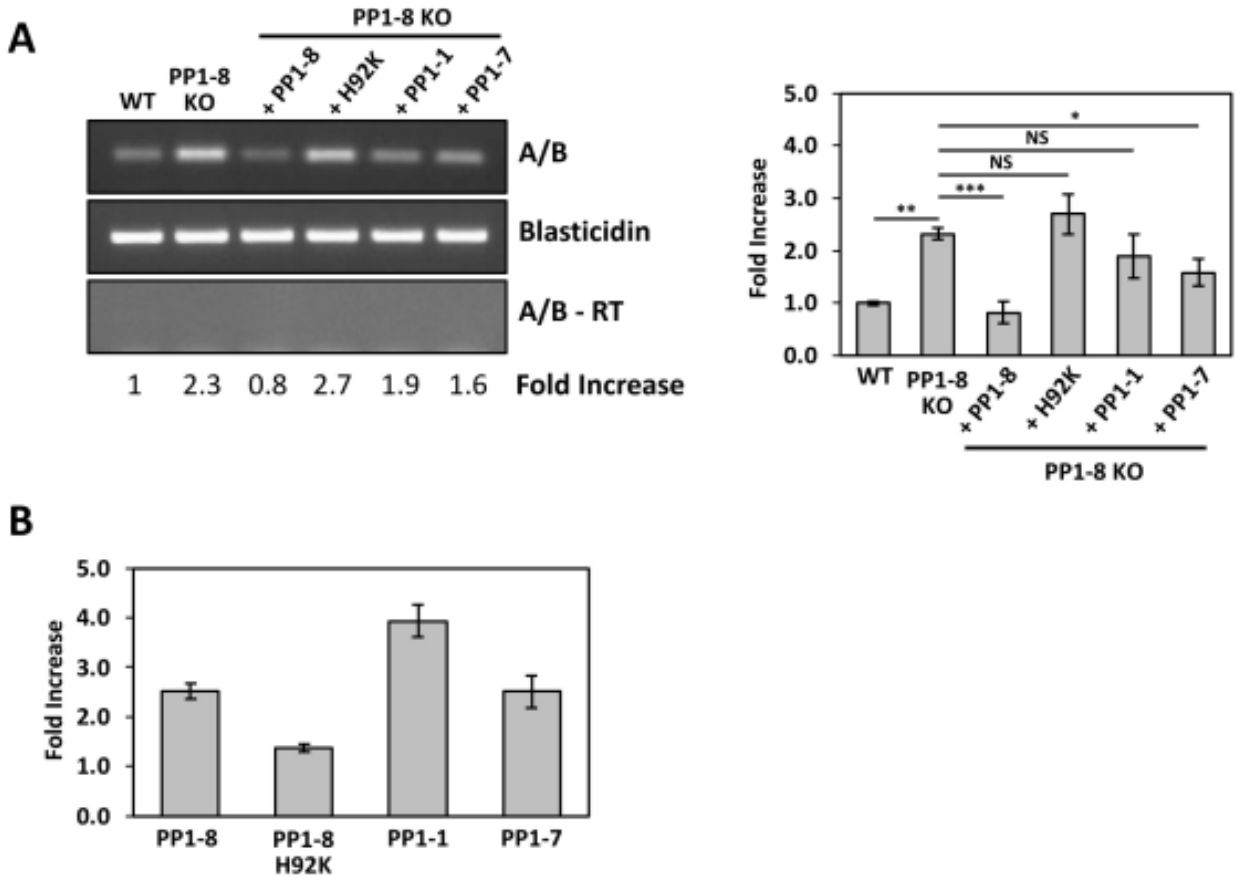


Figure 2.S10 PP1-8e catalytic activity is required for *L. major* termination. (A) Strand-specific RT-PCR analysis of readthrough transcription at cSSR 22.3 as described in Figure 2.8 A-B. Fold increase represents the generation of the nascent readthrough RNA species relative to background levels in WT determined by qRT-PCR (panel on the right). Error bars depict standard deviation from two replicate experiments analyzed in triplicate. P values were calculated using Tukey's multiple comparisons test. *, p values <0.05; **, p values <0.01; ***, p values <0.001; NS, not significant. PP1-8; LmPP1-8e, H92K; LmPP1-8e H92K mutant, PP1-1; LmPP1-1a, PP1-7; LmPP1-7d. (B) Levels of the indicated LmPP1 mRNA measured by RT-PCR in the PP1-8e KO expressing an untagged version of the protein from a plasmid. WT levels of the corresponding PP1 isotype is set to one. Error bars depict standard deviation from two replicate experiments analyzed in triplicate.

Supporting Information Table 2.S1. Genes with increased mRNA levels after depletion of PP1. The 34 genes showing the largest average (between the three replicates) significant (P_{adj}<0.001) increase in RNA abundance, as determined by RNA-seq analysis, are shown. Gene descriptions (annotation and accession number), fold upregulation and *p*-value adj are also listed. The "locus type" indicates whether the genes were immediately adjacent to convergent or head-to-tail transcription termination sites (cTTS and H-T, respectively) or located within a polycistronic transcription unit (PTU-internal) or adjacent to a divergent SSR (dSSR). "Up with DMOG" indicates if the gene is also upregulated >2-fold in WT cells treated with DMOG (37).

Accession	Annotation	Locus Type	Fold Upregulation	P_{adj}	Up with DMOG
LmjF09.0690	hypothetical protein, conserved	cSSR	11.5	<0.001	+
LmjF13.0700	Kinesin, putative	cSSR	2.6	<0.001	+
LmjF14.0470	UDP-glucoronosyl and UDP-glucosyl transferase, putative	cSSR	6.2	<0.001	+
LmjF16.1050	Protein of unknown function (DUF3184), putative	PTU internal	2.0	<0.001	
LmjF16.1280	hypothetical protein, conserved	PTU internal	2.0	<0.001	
LmjF17.0340.1	hypothetical protein, conserved	HT	2.8	<0.001	
LmjF19.0610	hypothetical protein, conserved	PTU internal	3.9	<0.001	
LmjF20.0820	hypothetical protein, conserved	HT	3.1	<0.001	+
LmjF20.0830	phosphopantetheinyl transferase-like protein	HT	4.6	<0.001	+
LmjF21.0015	histone H4	dSSR	3.5	<0.001	
LmjF22.0015	hypothetical protein (pseudogene)	cSSR	22.5	<0.001	
LmjF22.0020	hypothetical protein, unknown function	cSSR	6.6	<0.001	+
LmjF23.0510	hypothetical protein, conserved	HT	2.0	<0.001	
LmjF23.1590	oxidoreductase-like protein	cSSR	3.4	<0.001	
LmjF23.1600	oxidoreductase-like protein	cSSR	2.4	<0.001	
LmjF24.1770	inhibitor of cysteine peptidase	PTU internal	2.4	<0.001	
LmjF24.1800	hypothetical protein, conserved	PTU internal	2.1	<0.001	
LmjF24.1905	hypothetical protein, conserved	HT	4.3	<0.001	+
LmjF24.1910	chaperone protein DnaJ, putative	HT	5.7	<0.001	+
LmjF26.2280.1	nitrilase, putative	HT	22.3	<0.001	+
LmjF26.2550	hypothetical protein, conserved	PTU internal	3.1	<0.001	
LmjF27.1740	hypothetical protein, unknown function	cSSR	6.8	<0.001	+
LmjF30.2090	alcohol dehydrogenase, putative	cSSR	6.1	<0.001	
LmjF31.0350	amino acid transporter aATP11, putative	PTU internal	2.5	<0.001	
LmjF31.ncRNA3	unspecified product	HT	3.2	<0.001	+
LmjF33.1320	hypothetical protein, conserved	HT	2.5	<0.001	+
LmjF33.1760.1	OTT_1508-like deaminase, putative	cSSR	18.3	<0.001	+
LmjF33.1770	oxidoreductase-like protein (pseudogene)	cSSR	2.4	<0.001	
LmjF34.4230	hypothetical protein, conserved	PTU internal	2.1	<0.001	
LmjF35.2130	hypothetical protein, unknown function	HT	20.8	<0.001	+
LmjF35.2600	hypothetical protein, unknown function	cSSR	4.9	<0.001	+
LmjF35.2610	hypothetical protein, unknown function	cSSR	3.9	<0.001	+
LmjF36.2730	D-tyrosyl-tRNA deacylase, putative	HT	3.0	<0.001	+
LmjF36.3080	Octanoyltransferase, putative	HT	10.8	<0.001	+

Supporting Information Table 2. S2. High-throughput sequencing information. Information about all sequencing data generated in this study is listed.

Supporting Information Table S2. High-throughput Sequencing Information

Species	Genotype	Sample name	Type of RNA	Genome used for align	Raw read pair number	Mapped read pair number	Mapping rate	Average insert size	Average length of reads	Average mapping quality
L. major	WT	WT_A	polyA enriche	*L. major* Friedlin v9.0	89196563	78895412	0.884511794	234.2	146	35.2
L. major	WT	WT_B	polyA enriche	*L. major* Friedlin v9.0	94440997	87899211	0.930731502	228.7	146	35.3
L. major	WT	WT_C	polyA enriche	*L. major* Friedlin v9.0	131110228	122699202	0.935847675	249.3	148	35.2
L. major	PP1 KO	PP1_KO_A	polyA enriche	*L. major* Friedlin v9.0	60243756	56216398	0.933148956	246.2	147	35.2
L. major	PP1 KO	PP1_KO_B	polyA enriche	*L. major* Friedlin v9.0	85545285	78211979	0.914275743	271.9	147	35.2
L. major	PP1 KO	PP1_KO_C	polyA enriche	*L. major* Friedlin v9.0	83541119	76374190	0.914210761	256.9	147	35.2

Supporting Information Table 2.S3. Oligos used in these studies.

Name	Figure	Purpose	Orientation	Sequence (5' - 3'), restriction sites are underlined
LmjF-PP1_NdeI_pGL2314		Clone PP1 8e into pGL2314	sense	GATCGATCCATATGGCGCACACGAAGCGAGGGAG
LmjF-PP1_SpeI_pGL2314		Clone PP1 8e into pGL2314	antisense	GATCGATCACTAGTGTCGTGGCTCAGAGGGTTTG
LmjF-PP1_BclI_pXNG4		Clone PP1 8e into pXNG4	sense	GATCTGATCAATGGCGCACACGAAGCGAG
LmjF-PP1_BclI_pXNG4		Clone PP1 8e into pXNG4	antisense	GATCTGATCATCAGTCGTGGCTCAGAGGG
LmjF-PP1_Flag_pXNG4		Quick Change addition Flag tag onto PP1 8e in pXNG4	sense	GATCCGCTTCCAAACCCTCTGAGCCACGACACTAGTGACTACAAAGACGATGACGATAAGTAGAGATCTATGAGTCTTGTGATGTACTGGCTG
LmjF-PP1_Flag_pXNG4		Quick Change addition Flag tag onto PP1 8e in pXNG4	antisense	CAGCCAGTACATCACAAGACTCATAGATCTCTACTTATCGTCATCGTCTTTGTAGTCACTAGTGTCGTGGCTCAGAGGGTTTGGAAGCGGATC
LmjF-PP1_Xba_pTB007		Reclone PP1 8e into pTb007	sense	GATCGATCTCTAGAATGGCGCACACGAAGCG
LmjF-PP1_BsiWI_pTB007		Reclone PP1 8e into pTb007	antisense	GATCGATCCGTACGTCGACCACAAACAAAAATAC
LmjF-PP1_5'sgRNA		5' PP1 8e gRNA	forward	GAAATTAATACGACTCACTATAGGCTACGTTGCACACTCGAAAGGTTTTAGAGCTAGAAATAGC
LmjF-PP1_3'sgRNA		3' PP1 8e gRNA	forward	GAAATTAATACGACTCACTATAGGAAAATAGATGAGAGCGCACCGTTTTAGAGCTAGAAATAGC
sgRNA			reverse	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
LmjF-PP1_5'donor		5' PP1 8e donor fragment	sense	TCTTTTCTCTCCCTCACACGCATACTATTTCGTTGTAAAACGACGGCCAG
LmjF-PP1_3'donor		3' PP1 8e donor fragment	antisense	TTGTTTCTAAGGGAAGGGGGCGTGGCAACCCTCCATAGATCCATCCGCAC
LmjF-PP1_KO_verify	4, S4	PP1 8e Insertion and excision verify	sense	GGCGAGGAAGGACAGTAAA
LmjF-PP1_KO_verify	4, S4	PP1 8e Insertion and excision verify	antisense	CGTGGGCTTGTACTCGGT
cSSR22.3 Read Through RNA	8	RT primer cSSR 22.3	antisense	CCATGACTGTTGTCATCGCCATA
cSSR22.3 Up-TTS RNA (A)	8	RT PCR primer cSSR 22.3	sense	ACAGCACTACCGCAACGAAAG
cSSR22.3 Up-TTS RNA (B)	8	RT PCR primer cSSR 22.3	antisense	CACGGAAATGCAGGCATACAGAGGATCTTAACAGGG
blasticidin_HpaI	8	RT primer	antisense	GATCGATCGTTAACTTAGCCCTCCCACACATAACC
blasticidin_RTPCR	8	RT PCR primer	antisense	TTAGCCCTCCCACACATAAC
blasticidin_NdeI_RTPCR	8	RT PCR primer	sense	GATCGATCCATATGGCCAAGCCTTTGTCTCAAG
qPCR_LmjF33.0794 (B-tubulin)	6, 7, S9	qRT PCR primer	antisense	GAGACGAGGTCGTTCATGTT
qPCR_LmjF33.0794 (B-tubulin)	6, 7, S9	qRT PCR primer	sense	AGCAGTTCACGGGTATGTTC
qPCR_LmjF15.0220 (PP1)	S4	qRT PCR primer	sense	CTCGCTCCGTTGAGGTCGTTG
qPCR_LmjF15.0220 (PP1)	S4	qRT PCR primer	antisense	CCGCCGTGTACGCAGAAAATG
qPCR_LmjF09.0690	6, S9	qRT PCR primer	sense	CAAGCTCATCAATGCTGTCTG
qPCR_LmjF09.0690	6, S9	qRT PCR primer	antisense	GTAGGGTGGGATGAAGCAT
qPCR_LmjF09.0680	6, S9	qRT PCR primer	sense	GGGAGACTCAAAGAAAGACAGAG
qPCR_LmjF09.0680	6, S9	qRT PCR primer	antisense	CGCAGGGCAAACTCTACAT
qPCR_LmjF26.2270	7, S9	qRT PCR primer	sense	TCTACCGGCTTCAACATCTTC
qPCR_LmjF26.2270	7, S9	qRT PCR primer	antisense	TCGATGGCGGCTAAAGTAAC
qPCR_LmjF26.2280	7, S9	qRT PCR primer	sense	TCTGCATCAACACGGATACC
qPCR_LmjF26.2280	7, S9	qRT PCR primer	antisense	GCAAATAGCAACGCCAAACT
qPCR_LmjF22.0015	S9	qRT PCR primer	sense	AGGACTGCGACATCACAAAG
qPCR_LmjF22.0015	S9	qRT PCR primer	antisense	GACATGTAGGCTTCAGGGAAAT
qPCR_LmjF22.0020	S9	qRT PCR primer	sense	CCTCATGCAGCCCATCAA
qPCR_LmjF22.0020	S9	qRT PCR primer	antisense	GCGGACGTAGTGATCATAGAAC
qPCR_LmjF22.0040	S9	qRT PCR primer	sense	CGTGCATCTCACCTACCATC
qPCR_LmjF22.0040	S9	qRT PCR primer	antisense	CGACACCCTCACACCATTT

CHAPTER 3

LEISHMANIA PNUTS DISCRIMINATES BETWEEN PP1 CATALYTIC SUBUNITS THROUGH A RVXF-ΦΦ-F MOTIF AND POLYMORPHISMS IN THE PP1 C-TAIL AND CATALYTIC DOMAIN[1]

[1]**Zhang Y.** & Sabatini, R. Accepted by *Journal of Biological Chemistry*

ABSTRACT

PP1 phosphatases associate with specific regulatory subunits to achieve, among other functions, substrate selectivity. Among the eight PP1 isotypes in Leishmania, PP1-8e associates with the regulatory protein PNUTS along with the structural factors JBP3 and Wdr82 in the PJW/PP1 complex that modulates RNA polymerase II (Pol II) phosphorylation and transcription termination. Little is known regarding interactions involved in PJW/PP1 complex formation, including how PP1-8e is the selective isotype associated with PNUTS. Here, we show that PNUTS uses an established RVxF-ΦΦ-F motif to bind the PP1 catalytic domain with similar interfacial interactions as mammalian PP1-PNUTS and non-canonical motifs. These atypical interactions involve residues within the PP1-8e catalytic domain and N- and C-terminus for isoform specific regulator binding. This work advances our understanding of PP1 isoform selectivity and reveals key roles of PP1 residues in regulator binding. We also explore the role of PNUTS as a scaffold protein for the complex by identifying the C-terminal region involved in binding JBP3 and Wdr82, and impact of PNUTS on the stability of complex components and function in Pol II transcription *in vivo*. Taken together, these studies provide a potential mechanism where multiple motifs within PNUTS are used combinatorially to tune binding affinity to PP1, and the C-termini for independent binding of JBP3 and Wdr82, in the Leishmania PJW/PP1 complex. Overall, our data provide insights in the formation of the PJW/PP1 complex involved in regulating Pol II transcription in divergent protozoans where little is understood.

INTRODUCTION

Phosphorylation is a critical regulatory mechanism for over 70% of eukaryotic cellular proteins, and the majority of the phosphorylations occur on serine, threonine or tyrosine residues (125). More than 420 serine/threonine kinases target specific serine/threonine residues, which account for approximately 98% of all phosphorylation events. On the other hand, fewer than 40 serine/threonine phosphatases are involved in protein dephosphorylation (135, 140). Protein Phosphatase 1 (PP1) is a major serine/threonine phosphatase, estimated to catalyze one third of all dephosphorylation events in eukaryotic cells and involved in many essential cellular activities (including cardiac muscle contraction, glycogen metabolism, cell cycle transition, and transcription termination)(145, 146, 261). In contrast to protein serine/threonine kinases, although PP1 exhibits some intrinsic preference for pThr versus pSer and motifs surrounding the phosphorylation sites (262–264), the substrate specificity of PP1 is largely conferred by regulatory interactors of PP1 (RIPPOs) (previously referred to as PP1-interacting proteins or PIPs) (140, 265–268). Therefore, to carry out specific functions in a wide variety of cellular activities, PP1 binds over 200 confirmed RIPPOs, forming highly specific holoenzymes in mammalian cells (140, 149, 268, 269). These RIPPOs target PP1 to distinct cellular compartments and/or help direct its activity toward specific substrates (149, 270). RIPPOs usually associate with PP1 using a combination of short linear motifs (SLiMs). They bind in a largely extended manner at multiple sites across the top of PP1 (remote from the catalytic site), including the RVxF motif binding site and the ΦΦ motif binding site, both of which are used by a large number of RIPPOs. However, many studies have shown that RIPPO binding is usually more complex, with RIPPOs utilizing additional motifs beyond the RvXF and ΦΦ motif for PP1

holoenzyme formation (268). Characterizing these interactions is key to understanding how RIPPOs associate with PP1 and regulate specific biological processes such as transcription and gene expression.

One of the earliest characterized RIPPOs is PNUTS (PP1 nuclear targeting subunit), originally described as a nuclear regulator of PP1 that helps retain PP1 in the nucleus (165, 271, 272). PNUTS has been implicated in PP1-regulated processes including cell cycle regulation (175), RNA processing (165, 273), DNA repair (274), transcription (128) and telomere stability (173). Like most RIPPOs, PNUTS is a largely unstructured protein in the unbound state and included in a group of intrinsically disordered proteins (IDPs) (136, 149, 269, 275). This intrinsic flexibility is important for the formation of extensive interactions with PP1. PNUTS modulation of PP1 is mediated by a central region, employing RVxF-ΦΦ-Phe-Arg motifs (136). The most well-characterized motif is the RVxF motif ([K/R]-$X_{0–1}$-[V/I/L]-X-[F/W], where X can be any amino acid except proline) that is found in 90% of RIPPOs (151, 152, 154). ^{398}TVTW^{401} in human PNUTS (hPNUTS) constitutes the canonical RVxF PP1-binding motif, with the second and fourth residues burying deep in two hydrophobic pockets on the PP1 surface, providing an essential stabilizing force (136). As demonstrated for hPNUTS (15, 136), mutation of hydrophobic valine and phenylalanine/tryptophan positions in the RVXF-binding motif typically abolishes the ability of RIPPO to bind to PP1. Structure analyses of RIPPO:PP1 holoenzymes (including PNUTS) have identified several additional motifs that make contact with PP1(136). For example, the ΦΦ motif is a two-hydrophobic residue motif that is usually found 5-8 amino acids C-terminal to the RVXF motif on RIPPOs (136). hPNUTS-PP1 is found to be associated with two additional structural proteins, Wdr82 and the DNA binding protein Tox4, in a complex called PTW/PP1 (15). PNUTS is the scaffolding protein in the complex and mediates

independent associations of PP1, Wdr82 and Tox4. Tox4 interacts with an N-terminal TFIIS domain in hPNUTS, while Wdr82 binds to a C-terminal region in hPNUTS (aa 418-619). The PTW/PP1 complex is a negative regulator of RNA Pol II elongation rate and plays a key role in transcription termination. Depletion of individual components in human cells, or ortholog components in yeast, leads to RNA Pol II transcription termination defects (128, 129, 186, 276, 277). In the torpedo model of transcription termination, as Pol II reaches the poly(A) signal, pre-mRNA is cleaved, providing an entry site for the 5'-3' exoribonuclease Xrn2 to catch up with the Pol II and dislodge it from the DNA template, allowing for transcription termination (14, 59, 278). Dephosphorylation of the Pol II C-terminal domain (CTD) and Spt5, reducing the speed of the polymerase within the so-called termination zone, facilitates this process (109, 110, 279).

The Trypanosomatidae are early divergent protozoan parasites. Several members of the Trypanosomatidae including *Trypanosoma brucei* and *Leishmania major* are pathogenic to humans, causing Human African Trypanosomiasis (African sleeping sickness) and leishmaniasis. In these parasites, hundreds of genes of unrelated functions are arranged into polycistronic transcription units (PTUs) throughout the genome (280, 281). Genes in each PTU are co-transcribed from an initiation site at the 5' end to the termination site at the 3' end. Pre-mRNAs are processed through trans-splicing with the addition of a 39-nucleotide spliced leader sequence to the 5' end of mRNAs, which is coupled to the 3' polyadenylation of the upstream transcript (213, 282–288). Very little is understood regarding the RNA Pol II transcription cycle (initiation, elongation and termination) in these important eukaryotic pathogens. Epigenetic markers, such as histone variants (H3V and H4V) and the DNA modification base J, are enriched at Pol II transcription termination sites in Leishmania and *T. brucei* (204, 280, 289, 290). Base J is a glucosylated thymidine (291) and has only been identified in the nuclear DNA of kinetoplastids,

Diplonema, and *Euglena* (292, 293). The loss of base J (and H3V) in Leishmania and *T. brucei* led to readthrough transcription at termination sites, suggesting a critical role of base J in Pol II transcription termination (22, 24, 248, 249, 294). Exploring base J function further led to the identification of the PJW/PP1 complex in *Leishmania tarentolae* composed of PP1-PNUTS-Wdr82 and a base J-binding protein, JBP3 (252, 253). LtPNUTS is a predictively disordered 29 kDa protein with 23% sequence identity to hPNUTS, and contains a putative RVxF PP1 binding motif (^{97}RVCW^{99}) (252). Alanine substitution of the hydrophobic residues in the RVxF motif (^{97}RACA^{99}) has been shown to disrupt LtPNUTS-PP1 association (254). Additionally, short synthetic RVXF-containing peptides are sufficient to disrupt the LtPNUTS-PP1 association. Ablation of PNUTS, JBP3 and Wdr82 by RNAi in *T. brucei* (252), and deletion of PP1-8e and JBP3 in Leishmania (253, 254), has been shown to cause Pol II termination defects, similar to the defects following the loss of base J/H3V. These *in vivo* data, along with the recent demonstration that Pol II is a direct substrate for PP1-8e as a component of the Leishmania PJW/PP1 complex *in vitro* (254), supports a conserved PNUTS-PP1 regulatory mechanism from trypanosomatids to yeast and mammalian cells. We therefore proposed that similar to the PTW/PP1 complex, LtPNUTS is scaffolding protein that mediates independent binding of PP1, JBP3 and Wdr82, with JBP3 tethering the complex to the base J-enriched transcription termination sites for PP1-mediated dephosphorylation of Pol II.

Eight PP1 isoforms, grouped into five different clades (A-E), are identified in the Leishmania genome (Fig. 3.S1). Among these, only PP1-8e is found associated with the PJW/PP1 complex *in vivo* and shown to be involved in Pol II transcription termination (254). Although the *T. brucei* genome also harbors eight PP1 isoforms, no obvious PP1 isoform belongs to clade E as a homologue of LtPP1-8e (254). Furthermore, purification of TbPNUTS pulls down

JBP3 and Wdr82 but not PP1(252). Presumably, transient/weak association between a TbPP1 isotype and the PNUTS-Wdr-JBP3 complex via the conserved RVxF PP1-binding motif, allows a conserved transcription termination mechanism in *T. brucei* cells (252, 254). Unique sequences within PP1-8e may explain isotype selectivity of PNUTS binding in Leishmania (254). However, interactions involved in the selectivity of LtPNUTS for the PP1-8e isoform have not been explored. In fact, while PP1 isoform selectivity is thought to be an important feature of regulatory RIPPOs, limited mechanistic information exists on how this is achieved in any system. The mammalian PP1 isoforms (PP1α, PP1β, PP1γ) share a sequence identity ranging from 85% to 93%, and sequence variability mainly comes from the divergent N- and, most notably, C-termini, with only a few amino acid residues being different within the catalytic domains (140). Among the regulatory RIPPOs which display isoform preferences, such as MYPT1(295, 296), Spinophilin (297), RepoMan (162), Ki67(162), ASPP2 (163) and RRP1B (298), specificity is achieved via recognition of the PP1 C-terminus or a β/γ specificity pocket within the PP1 catalytic domain. The extreme C-terminus of PP1 ($PP1\alpha^{309\text{-}330}$) contains a SH3-binding motif (PPII – PxxPxR), that is conserved among all the mammalian PP1 isoforms, and a variable C-tail. The apoptosis stimulation proteins of p53 family (iASPP/ASPP1/ASPP2) utilize a SH3 domain to selectively bind the PP1 C-terminus via contacts in the PPII motif and residues in the variable C-tail region to achieve isoform selectivity (163, 299, 300). Ankyrin repeats of the myosin phosphatase targeting subunit MYPT1 associate with amino acids in the PP1 C-tail and drive selectivity towards PP1β (140). In the case of RRP1B, RepoMan and Ki-67, the SLiM (KiR or SLIV) immediately downstream of the RVxF motif determines the preference toward PP1γ through a single amino acid change in the catalytic domain of PP1(162, 298, 301).

Therefore, isoform specificity is mediated in these RIPPOs by a single amino acid difference in PP1 at position 20, which is an Arg residue in PP1β/γ and a Gln residue in PP1α.

In this study, we employ structural modeling and mutagenesis analysis to help define how LtPNUTS specifically recruits PP1-8e to the PJW/PP1 complex. First, we confirm that LtPNUTS demonstrates substrate specificity for PP1-8e among the identified LtPP1 isoforms *in vivo* by Co-IP analysis. We show that LtPNUTS binds to PP1 via a combination of well-characterized PP1-interacting motifs including the extended RVxF (RVXF-ϕ_R-$\phi\phi$) and Phe motif. We also identified unique termini and motifs within LtPP1-8e catalytic domain, including sites not previously shown to bind any PP1 regulator, that are important for PP1-PNUTS interaction. Lastly, we explore the scaffold function of PNUTS by mapping the Wdr82 and JBP3 binding domain at the C-terminus of PNUTS and demonstrate PNUTS protein level is critical for the integrity of the PJW/PP1 complex and function in Pol II termination. Together, these data support a model for extensive interactions between LtPNUTS and PP1-8e and provide key insights into the isoform selectivity of LtPNUTS and its scaffold function in overall stability of the PJW/PP1 complex.

RESULTS

LtPNUTS displays PP1 isoform selectivity

Our previous affinity purification-mass spectrometry data (252) indicated that PNUTS is part of a tightly interlinked protein network comprising the PP1 catalytic subunit PP1-8e, JBP3 and Wdr82 in *L. tarentolae* cells. While there are 8 PP1 isotypes in the Leishmania genome (Fig. 3.S1) (253, 254), only PP1-8e was associated with the Leishmania PJW/PP1 complex. In order to understand the specific association of PP1-8e in this complex, we first sought to verify the binary

interaction between PNUTS and the *L. tarentoae* PP1 catalytic subunits by co-immunoprecipitation (co-IP) *in vivo*. To do this, we HA-tagged the endogenous loci of LtPNUTS using cas9 and overexpressed various Pd-tagged LtPP1 isotypes from a plasmid (Fig. 3.1A). LtPP1-3 (LtaPh_3411201) is much smaller (167 aa) than other PP1 isotypes and is predicted to contain a partial catalytic core. We were unable to overexpress Pd-tagged LtPP1-3, suggesting that it could be a truncated pseudogene. Therefore, we analyzed 5 of the 7 complete PP1 isotypes in *L. tarentolae* (representing all 5 clades). Our results show that only PP1-8e can IP a significant fraction of PNUTS, while the other PP1 isotypes (PP1-1, PP1-2, PP1-4 and PP1-7) show no detectable interaction with PNUTS by Co-IP (Fig. 3.1B). These data confirm the MS analysis of the purified PNUTS-PP1 complex (252, 254) and directly demonstrate for the first time that PNUTS preferentially targets PP1-8e over other isoforms in intact leishmania cells.

Lt PNUTS associates with PP1-8e through an established RVxF-ϕ_R-ΦΦ -F motif

To determine the molecular basis of isoform specificity of PNUTS for PP1-8e in *L. tarentolae*, we used AlphaFold to help define the PNUTS-PP1 interaction interface. We first explored the predicted structure for the LtPP1 isotypes. The PP1 catalytic core is highly conserved across eukaryotes from human to yeast cells, consisting of 10 sets of α-helices (labelled A' to I) and 15 sets of β-sheets (numbered 1' to 14) (Fig. 3.S2) (153, 156). The catalytic core regions of the LtPP1 isotypes are predicted to be of high confidence by AlphaFold and their structural overlay to the determined human PP1 protein structure (PDB: 3E7a, Fig. 3.S3A) shows a high structural similarity. An example is the predicted LtPP1-1a structure (Fig. 3.S3B), which shows a high structural identity to hPP1 (Fig. 3.S3C) with a root mean square deviation (RMSD) of 0.580 Å. LtPP1-8e was also predicted with high confidence (Fig. 3.S3D) for the catalytic core region. The predicted LtPP1-8e structure aligns well to the hPP1 structure (Fig. 3.S3E) except

three regions within the catalytic core and the N- and C-terminus that we have identified as unique to PP1-8e (254) (Figs. 3.S2 and 3.S3E). Deletion of these unique regions in LtPP1-8e increases the structural similarity between LtPP1-8e and hPP1 (RMSD: 0.666 Å, Fig. 3.S3G). Thus, as previously predicted based on sequence conservation (254), the structural identity of PP1 catalytic subunits between mammals and Leishmania suggest strong functional conservation during evolution. However, unique sequences in PP1-8e may be important for PP1-8e specific functions in Leishmania.

We next submitted LtPNUTS and LtPP1-8e sequences together to AlphaFold2 to generate the predicted LtPP1:PNUTS structure (Fig. 3.S4). As expected, the majority of LtPNUTS is unstructured. While only a limited region of PNUTS is confidently predicted to become buried upon complex formation (Fig. 3.S4), this region binds in a largely extended manner at multiple sites across the top of PP1 in a way highly similar to several well-characterized PP1-interacting proteins (Fig. 3.2B and C), including human PNUTS (Fig. 3.S5) (136), spinophilin (302), and Gm (303). They share multiple well-charactered PP1-binding motifs, including the RVxF-ϕ_R-$\phi\phi$ binding motif (Fig. 3.2A and C). Furthermore, similar to hPNUTS, LtPNUTS is predicted to bind PP1 remotely away from the PP1 catalytic site, making it fully accessible to substrate. Consistently, we have recently demonstrated PP1 is catalytically active in the PNUTS:PP1-8e holoenzyme, capable of dephosphorylating model substrates, such as p-nitrophenyl phosphate, as well the LtPol II CTD (254). The first of the key interaction sites in the LtPNUTS:PP1-8e complex is bound by the RVxF-ϕ_R motifs (Fig. 3.2A-D). Nine residues of PNUTS (^{93}R to D^{101}) adopt an extended conformation and bind to a hydrophobic channel on the PP1 surface formed at the interface of the two β-sheets of the β-sandwich opposite to the catalytic site channel. PNUTS residues ^{96}RVCW^{99} form the RVxF motif, which binds the PP1

RVxF binding pocket, V and W are the anchoring hydrophobic residues that bind deeply in this pocket (Fig. 3.2D). The predicted LtPNUTS RVxF interaction is highly similar to those observed in other PP1 holoenzyme complexes, including mammalian PNUTS-PP1 (Fig. 3.S5B). Structural and functional studies of the mammalian PP1-PNUTS complex, and modeling of the LtPP1-PNUTS complex here suggest a dominant role for V97 and W99 in stabilizing the interaction between LtPNUTS and PP1-8e. We have recently shown the V97A-W99A double mutant is unable to bind PP1-8e (254). To test this hypothesis in more detail, we made single alanine mutations at each of these positions in the LtPNUTS expression plasmid (pSNSAP1) and tested the PNUTS mutants for interaction by Co-IP with endogenously HA-tagged PP1-8e. Alanine mutation of W99 completely abolished PP1-PNUTS association, and V97A decreased PP1-PNUTS association by 5-fold (Fig. 3.3A and 3B), indicating the importance of the hydrophobic association mediated by the RVxF motif. Inspection of the structure of hPNUTS in complex with PP1 highlighted interfacial PP1 amino acids I169, L243, F257, R261, V264, I266, M283, C291, F293 that are conserved in LtPP1-8e as I217, L293, F307, R314, L317, L336, V343, C344, and I346 that form the hydrophobic pocket and stabilize V97 and W99 in the PNUTS RVxF motif (Fig. 3.2D and Fig. 3.S5B). To test this, single alanine mutation of I217 was introduced into the LtPP1-8e expression construct and the PP1 mutants tested for interaction by Co-IP with endogenously HA-tagged PNUTS. Mutation of $I217_{PP1}$ to alanine significantly reduced (~50%) the PP1-PNUTS interaction (Fig. 3.3C and 3.S6B), supporting the importance of the hydrophobic interface with the conserved Val and Trp moieties of the LtPNUTS RVxF motif. We suggest that the VxW motif in LtPNUTS is the putative counterpart of a Vx(F/W) motif that comprises a key part of the PP1 phosphatase-binding site identified in several other PP1

regulatory subunits, including hPNUTS, where the VxW motif binds to a hydrophobic pocket of the phosphatase remote from the phosphatase active site (136) (Fig. 3.S5B).

A short 22-aa peptide from PNUTS that contains the RVxF motif is able to disrupt the PP1-PNUTS association, while the identical peptide with V97A and W99A substitutions is not (254), further confirming the importance of the RVxF motif in the LtPNUTS:PP1-8e complex. However, wild-type RVxF peptide did not elute all of the PP1 from PNUTS suggesting there may be additional interaction sites that stabilize the PNUTS–PP1 complex (254). PP1 phosphatase-regulatory proteins often have at least one, and often several, basic amino acids preceding the Vx(F/W) motif (155). It has been suggested that this basic region may interact with a negatively charged patch near the RVxF-binding pocket of PP1. In the case of hPNUTS, there is a run of five basic amino acids upstream of VxW (Fig. 3.2A). Two of which engage in salt bridges to acidic side chains of PP1 (Fig. 3.S5G). Similar interactions are predicted for LtPNUTS ^{94}RKR^{96}, which are predicted to have electrostatic interaction with PP1-8e residues D290, E340 and D292, respectively (Fig. 3.S5H). Alanine mutation of R94A$_{PNUTS}$, or its predicted interacting residue D290$_{PP1}$, leads to a ~50% reduction in PP1-PNUTS association (Figs. 3.3, B and C and 3.S6, A and B). Alanine mutation of K95$_{PNUTS}$ did not affect PP1-PNUTS interaction (Figs. 3.3B and 3.S6A). However, not all electrostatic interactions mediated by these basic residues contribute equally to the association, and sometimes, simultaneous alanine mutations of all the basic amino acid residues preceding the RVxF motif is required to affect PP1 binding, as observed for the fission yeast PNUTS (304). We were unable to generate the R96A$_{PNUTS}$ mutant, but alanine mutation of its interacting residue D292A$_{PP1}$ leads to 60% reduction in PP1-PNUTS interaction (Fig. 3.3C and 3.S6B), supporting the importance of R96$_{PNUTS}$. Acidic residues C-terminal to the RVxF motif are also present in other PP1 regulatory

subunits and important for binding PP1. The AlphaFold model shows D101 of LtPNUTS engaging in salt bridge interactions to $R314_{PP1}$ (Fig. 3.S5H). Similar interaction is observed on E405 of hPNUTS (Fig. 3.S5G). Consistent with the prediction, the $D101A_{PNUTS}$ mutant showed roughly 80% decreased interaction with PP1-8e (Fig. 3.3B and 3.S6A).

The LtPNUTS:PP1-8e model predicts $H114_{PNUTS}$ and $V115_{PNUTS}$ form the PNUTS ΦΦ motif, which binds the PP1 ΦΦ binding pocket (Fig. 3.2E). Like the RVxF interaction, the ΦΦ interaction is highly similar to those observed in other PP1 holoenzyme complexes (Fig. 3.2C). The ϕϕ motif usually consists of two hydrophobic residues of RIPPOs that are buried in a hydrophobic pocket on PP1, but can be degenerate, including sequences such as VS, VC, VK, IN, and HH (136). The ϕϕ motif of hPNUTS is represented by ^{410}YF^{411} located on a short β strand that hydrogen bonds with β strand β14 of PP1, extending one of its two central β sheets (Fig. 3.S5C). AlphaFold predicts a similar arrangement in the LtPNUTS-PP1 complex (Fig. 3.2E and 3.S5C). The predicted ϕϕ motif of LtPNUTS, ^{114}HV^{115}, is located on a short β strand formed by ^{112}VKHV^{115} that potentially H-bonds with PP1-8e's β strand 14 and the ϕϕ hydrophobic pocket on LtPP1-8e includes residues N100, R104, E350 and T348 (Fig. 3.2E). To test the significance of the ϕϕ motif, we mutated LtPNUTS ^{114}VH^{115} to alanine, and found that the mutation significantly weakens the PP1-PNUTS association (Fig. 3.3B and 3.S6A). While the structure of the hPNUTS:PP1 complex does not indicate any specific interactions between the ϕϕ motif of hPNUTS and the ϕϕ hydrophobic pocket of PP1, we noticed a potential salt bridge interaction between $H114_{PNUTS}$ and $T348_{PP1}$ in our model (Fig. 3.2E). Alanine mutation of $T348_{PP1}$, however, had minimal impact on PP1-PNUTS association (Fig. 3.3C and 3.S6B). Presumably, the interaction does not occur or the alanine mutation of T348 alone is not sufficient to disrupt the stabilizing β sheet interactions provided by the remaining residues in the pocket

(Fig. 3.2E). A distinctive feature of the predicted LtPNUTS-PP1-8e structure is the extended linker between the RVxF and ΦΦ motifs (Fig. 3.2A). In LtPNUTS these two motifs are separated by 14 residues, and would represent the longest insert observed thus far for any PP1 regulator. In G_M, these two motifs are separated by 10 residues (Fig. 3.2A). This "extended kink" is presumably stabilized by hydrophobic interactions made by the Φ_R motif, represented by $L111_{PNUTS}$, with the hydrophobic Φ_R pocket adjacent to the RVxF binding pocket in PP1-8e (Fig 3.2B and 3.2D). As such, $L111_{PNUTS}$ is stabilized by $F305_{PP1}$, $F307_{PP1}$, and $R314_{PP1}$ components of the Φ_R pocket in PP1-8e (Fig. 3.2D). The contact mediated by the Φ_R motif ($L407_{hPNUTS}$) is conserved in the hPP1:PNUTS structure (Fig. 3.S5B). Highlighting the importance of the ϕ_R motif in the LtPNUTS:PP1 structure, $L111A_{PNUTS}$ mutation significantly reduced (80%) the PNUTS-PP1 interactions (Figs. 3.3B and 3.S6A). Furthermore, alanine mutation of $F305_{PP1}$ decreased PP1-PNUTS association by 80%, and single alanine mutations of residues lining both the RVxF and Φ_R pockets of PP1-8e, F307, I346 and R314, completely disrupts the LtPP1-PNUTS association (Figs. 3.3C and 3.S6B).

In many cases, PP1 interactions can extend beyond the ΦΦ motif. For example, $F413_{hPNUTS}$ is the Phe motif that binds in a deep pocket immediately adjacent to $P298_{PP1}$ in the human complex (Fig. 3.S5D). This pocket is also frequently used by other regulators to bind PP1. For example, Gm ($F82_{GM}$) (303), spinophilin ($T461_{spino}$) (302), and RepoMan/Ki67 ($F404_{RM}$) (162) and RRP1B ($F696_{RRP1B}$) (298) bind this same pocket. A conserved F118 is present in LtPNUTS and predicted to bind a pocket adjacent to $P351_{PP1}$ in the LtPNUTS:PP1 model (Fig. 3.2F). $P351_{LtPP1}$ occupies a similar position as $P298_{hPP1}$ in the human PP1 pocket (Fig. 3.S5D). Additionally, the model predicts that Y96, N100, R104, L349, P352 and I360 form the Phe binding pocket in LtPP1-8e (Fig. 3.2F). While the $Y96A_{PP1}$ mutation had a minor effect,

alanine mutation of F118$_{PNUTS}$, or I360$_{PP1}$ significantly reduced LtPP1-PNUTS associations (Fig. 3.3B, 3C, S6A and S6B), supporting a similar involvement of the Phe motif in the LtPNUTS-PP1 complex.

An additional potential LtPP1-8e interaction beyond the ΦΦ motif is R125$_{PNUTS}$ (Fig. 3.2A). In hPNUTS R420 is involved in hydrophobic and electrostatic interactions with PP1, representing the so-called Arg motif (Fig. 3.S5E). R420$_{hPNUTS}$ is buried in a hydrophobic pocket formed by L296, P298, and P270 of hPP1. Additionally, E419$_{hPNUTS}$ and R420$_{hPNUTS}$ form bidentate salt bridges with R74$_{PP1}$ and D71$_{PP1}$, respectively (Fig. 3.S5E). However, this interaction is not predicted by AlphaFold in the LtPNUTS:PP1-8e complex. Rather, an alpha helix (residues 354-362) within the C-terminal tail of PP1-8e occupies the PP1 hydrophobic pocket involved in R420 hPNUTS binding (Fig. 3.S5F). Furthermore, while the Arg motif is presumably conserved on LtPNUTS as R125, the polar S124$_{PNUTS}$ replaces the negatively charged E419$_{PNUTS}$ in hPNUTS. Moreover, the interacting charged residue in hPP1, R74$_{PP1}$, is replaced by N100 in LtPP1-8e (Figure 3.S7). The replacement of charged residues with polar residues may prevent the formation of a bidentate salt bridge important for Arg motif binding. This concept along with the blocking of the Arg pocket by the C-terminal tail of LtPP1-8e could explain the divergence of the LtPNUTS-PP1 binding structure from hPNUTS at this region. R125 on LtPNUTS is therefore not predicted to bind to PP1-8e, and alanine mutation of R125 in LtPNUTS had no effect on PP1-PNUTS association (Figs. 3.3B and 3.S6A). Alanine mutation of R420$_{hPNUTS}$, however, does not affect PP1-PNUTS association in human cells (136), although the crystal structure indicates the importance of the Arg motif. Therefore, while the AlphaFold model clearly rules out the interaction, our Co-IP results do not completely exclude the possibility that R125$_{PNUTS}$ mediates interaction with PP1 in *L. tarentolae* cells.

Taken together, the predicted structure and mutagenesis analyses establishes that LtPNUTS, like a majority of PP1-specific regulators, binds LtPP1-8e, in part, using a general RVxF and ϕ_R-ΦΦ-F SLiMs. We noticed that a majority of the mutant PNUTS proteins tested here are over-expressed at a lower protein level than WT PNUTS protein (Fig. 3.S8A). This is not observed for PP1 mutants over-expressed from the same plasmid (Fig. 3.S8B), suggesting that PNUTS protein level is sensitive to mutations. However, reduction in the level of over-expressed PNUTS does not necessarily lead to reduction in PP1 binding in the co-IP. For example, $R125A_{PNUTS}$ is one of the lowest expressed PNUTS mutants (Fig. 3.S8A) but showed comparable PP1 association as WT PNUTS (Fig. 3.3B). Potentially, the reduced expression of the mutant PNUTS protein is sufficient for saturation binding of available PP1, allowing co-IP of PP1 to the same extent as WT PNUTS. Furthermore, confirmation of the PNUTS-PP1 interface based on mutation of PNUTS residues is supported by mutation analysis of the corresponding binding pocket on PP1 where expression levels are not affected by mutagenesis. Therefore, the reduced protein expression level of certain mutant PNUTS protein does not affect our overall conclusions regarding the RVxF-ϕ_R-ΦΦ-F motifs.

PP1-8e isoform specific residues are involved in PNUTS binding.

The mode of PP1-PNUTS interaction described above, via the established RVxF-ϕ_R-ΦΦ-F motif, is typical for a scaffolding function of regulatory proteins but likely does not affect selectivity toward PP1 isoforms. In fact, a majority of the PP1 residues characterize above as involved in PNUTS-PP1 binding are not restricted to the PP1-8e isotype (Fig. 3.S7), and thus, fail to explain the marked preference of PNUTS for PP1-8e. Therefore, we explored structural features present in LtPP1-8 that would confer specificity to LtPNUTS. As mentioned above, hPP1 isoforms share a high sequence identity with differences mainly limited to their extremities

and some RIPPOs take advantage of these differences to interact selectively with specific PP1 isoforms. As we recently noted (254), an interesting characteristic of the PP1 isotypes in Leishmania is the diversity of their N- and C-terminal tails and the insertion of several short sequence elements specifically within the catalytic subunit of PP1-8e (Fig. 3.S2). BLASTp search using the N-terminal (residues 1-32) and C-terminal (residues 241-264) sequences of *L. tarentolae* PP1-8 failed to identify any significant homology to any sequence in the NCBI database. To test the contribution of each of these unique PP1-8e characteristics to LtPP1-PNUTS association, we performed deletion and alanine mutagenesis (constructs used in this study are illustrated in Fig. 3.4A). As shown in Fig. 3.S8B, all PP1-8e mutants are over-expressed in leishmania cells to similar levels as WT PP1-8e. Upstream of the highly conserved catalytic domain, LtPP1-8e has a 32 amino N-terminal extension (Fig. 3.S2). Deleting the PP1-8e N-terminal region completely abolished PP1-PNUTS association by co-IP (Figs. 3.4B and 3.S9), suggesting that residues 1-32 of LtPP1-8e are essential for this interaction. Sequence differences between mammalian PP1-α and PP1-γ C-terminal ~25-amino-acid tails are implicated in isoform specific binding by ASPP2 (163) and MYPT1(295). Similarly, Lt PP1-8e has a unique extended C-tail of ~25 amino acids (Fig. 3.S2) that includes two residues (P352 and I360) that we have demonstrated above as important for PNUTS-PP1 binding, potentially via stabilization of the Phe motif (Fig. 3.2F). Deleting the PP1-8e C-terminal tail significantly impacted PNUTS binding *in vivo* (Fig. 3.4B). While deletion of the 23 amino acid C-terminal extension of PP1-8e (C□23) leads to ~80% loss in PNUTS binding, deletion of the final 11 amino acids (C□11) resulted in ~30% reduction in PNUTS binding (Figs. 3.4B and 3.S9). The 12 amino acid of the C-term region between these two deletions includes a predicted 9 amino acid alpha-helix (354-362) rich in charged or polar residues (Fig. 3.S2), potentially involved in

electrostatic interactions with PNUTS. To test this idea, we did alanine scanning mutagenesis of two regions within this C-terminal helical region. Alanine substitution of four residues within first half of this helix in PP1-8e (354-358A) resulted in 50% reduction in PNUTS binding and the 359-362A mutation of PP1-8e resulted in 80% reduction in PNUTS binding, similar to what we observed in the 23 amino acid deletion (CΔ23) (Fig. 3.4B). The $I360A_{PP1}$ mutant led to a similar 80% reduction in LtPNUTS-PP1-8e associations, indicating I360 is a key residue within this C-term 359-362 helical region. Thus, the unique C-terminal tail of PP1-8e, in particular residues 359-362, and the first 32 amino acids at the N-terminus are needed for PNUTS binding.

According to the LtPNUTS:PP1-8e model, while the N-terminus of PP1-8e is unstructured and thus, difficult to understand how it is involved in isoform selective binding to PNUTS, the C-terminus appears to provide additional stabilization to the Phe binding pocket. PP1-8e, and other LtPP1 isotypes, have an Phe binding pocket similar to the human PP1:PNUTS complex (Fig. 3.2F and Fig. 3.S5D and 3.S7). However, the unique C-terminus of PP1-8e provides additional residues (including P352 and I360) that may contribute to the Phe binding pocket. To examine this idea further we determined the AlphaFold model for the LtPNUTS:PP1-1a complex (Fig. 3.5). LtPP1-1a appears to have a majority of the conserved residues for the RVxF, ϕ_R, $\phi\phi$ and F motif binding pockets as the human PNUTS:PP1 complex and the predicted LtPNUTS:PP1-8e model (Figs. 3.5 and 3.S7). However, LtPP1-1a lacks the extended C-terminus present in LtPP1-8e (Figs. 3.5 and 3.S2) and, interestingly, is predicted to associate with LtPNUTS with the RVxF-ϕ_R-$\phi\phi$ motifs, but not the Phe motif (Fig. 3.5)

An additional characteristic of PP1-8e is the insertion of three unique sequence motifs within the catalytic domain; a 26 amino acid insertion (residues 109-134) near the N-terminus and two smaller (^{260}LPAGVD^{265} and ^{310}DHK^{312}) insertions near the C-terminus (Fig. 3.S2).

AlphaFold modeling shows the insertions are presented on the surface of PP1-8e at novel sites compared with the human PP1 structure and the LtPP1-1 isoform (Fig. 3.S3, E and F). To test the significance of these regions, we performed deletion and alanine mutagenesis. Deletion of the 26 amino acid insertion in PP1-8e (109-134Δ) results in severely reduced ability (80%) to associate with PNUTS (Fig. 3.4B). The 26 amino acid region is rich in charged and polar residues that are conserved among Leishmania PP1-8e homologs, potentially involved in electrostatic interactions with PNUTS. To test this idea, we did alanine mutagenesis in three regions of the 26 amino acid insertion: GGVFG (109-114A), DKKR (116-121A) and SDDYS (126-134A) (Fig. 3.4A). While the 116-121A and 126-134A mutations had little effect on PNUTS binding, mutation of five residues in 109-114A resulted in 80% reduction in PNUTS binding, similar to the effect of deleting the entire 26 amino insert (Fig. 3.4B). Similar alanine mutagenesis was performed for the two smaller PP1 insertions: ^{260}LPGVD^{265} and ^{310}DHK^{312} (Fig. 3.4B). The results show that while alanine mutagenesis of ^{310}DHK^{312} leads to a small decrease (~20%) in PP1-PNUTS association, alanine mutagenesis of ^{260}LP^{261} abolishes roughly 80% of PP1-PNUTS interaction (Figs. 3.4B and 3.S9). Alanine substitution of the remaining three residues of the ^{260}LPGVD^{265} (263-265A) led to approximately 90% reduction in PNUTS binding, and D265A$_{PP1}$ mutation only had a moderate effect on PP1-PNUTS interaction (Figs. 3.4B and 3.S9). Thus, unique sequences within the catalytic domain of PP1-8e, in particular residues ^{109}GGTVFG^{114} and ^{260}LPAGV^{264}, are needed for PNUTS interaction. Taken together these results suggest that LtPNUTS can discriminate between different PP1 isoforms based on the PP1 N- and C-terminus and unique sequence motifs within the catalytic domain. As such, these regions might underlie the mechanism by which LtPNUTS shows preferential binding to PP1-8e.

PNUTS as a scaffold for the PJW/PP1 complex

hPNUTS is a scaffolding protein in the human PTW/PP1 complex, binding Tox4 and Wdr82 with its N- and C-terminus regions, respectively, and PP1 via the centrally located RVxF motif (15). The hPNUTS is a 114 kDa protein (940 amino acids) with multiple identified protein domains (273). LtPNUTS, which lacks identifiable protein domains or motifs apart from the conserved PP1-interacting RVxF motif discussed above, is much smaller at 28.6 kDa, consisting of 264 amino acids. To test if LtPNUTS similarly serves as a scaffolding protein and binds to Wdr82 and JBP3 with distinct domains we over-expressed Pd-tagged PNUTS protein with various of N- and C-terminal truncations and studied the interaction between PNUTS truncations and endogenously HA-tagged JBP3/Wdr82 using Co-IP (Fig. 3.6A). We find that full length PNUTS allows significant co-IP of both JBP3 and Wdr82 (Figs. 3.6C and 3.S10), consistent with our previous studies of the PJW/PP1 complex in Leishmania and *T. brucei* (252). Confirming that the RVXF motif and PP1 binding are not required for JBP3 and Wdr82 association with PNUTS, mutation of the PP1 binding RVxF motif (RACA mutant) has little to no effect on Wdr82 or JBP3 binding to PNUTS (Figs. 3.6C and 3.S10). Interestingly, PNUTS proteins with three different N-terminal truncations (NΔ27, NΔ47, and NΔ75) are expressed at significantly lower levels than the full-length PNUTS control, and the major species run at lower molecular weights than expected on SDS-PAGE gel (Fig. 3.6B and 3.S8C). As an intrinsically disordered protein, hPNUTS is known to not run to the expected size on the SDS-PAGE gel (165) and we have characterized the altered mobility of TbPNUTS (252). Potentially, the deletion of an N-terminal sequence accentuates the disordered nature and altered mobility of the truncated LtPNUTS polypeptide. In this case, the major species represents the indicated truncated PNUTS protein. Alternatively, N-terminal deletions lead to PNUTS protein instability and further protein

cleavage. While it is difficult to obtain accurate measurement of binding with such low protein expression in the parasite, it seems that PNUTS with varying lengths of N-terminal truncations still immunoprecipitated a significant level of Wdr82 or JBP3 compared to the negative control, although not to the same extent as WT PNUTS (as shown in Fig. 3.S10).This is best represented by the NΔ75 PNUTS, with the highest level of expression among the N-terminally truncated PNUTS proteins (Fig. 3.6B). This would suggest the N-terminus of PNUTS is not essential for JBP3/Wdr82 binding. We noticed that similar N-terminal truncations of the PNUTS homolog in *T. brucei* does not result in decreased levels of expression (Fig. 3.6, D and E), allowing further studies of Wdr82/JBP3 association. To do this we tagged JBP3 and Wdr82 in *T. brucei* with HA and Myc tags respectively, and exogenously expressed protein A-tagged PNUTS via a Tet-inducible promoter. Supporting the LtPNUTS analysis, 72 aa deletion from the N-terminus (NΔ72) tested in TbPNUTS had little to no effect on Wdr82/JBP3 binding (Fig. 3.6F). In contrast, while all C-terminal truncations of LtPNUTS are expressed at levels similar to full-length in both HA tagged Wdr82 and JBP3 cell lines (Fig. 3.6B and 3.S8C), even the smallest 23 amino acid deletion (CΔ23) had a negative effect on both Wdr and JBP3 binding to LtPNUTS (Figs. 3.6C and 3.S10). While the 23 amino acid deletion led to complete loss of JBP3 binding, a small level of Wdr82 association remained that is subsequently lost upon further deletions of the C-term end (CΔ37 and CΔ66) (Figs. 3.6C and 3.S10). Similar to LtPNUTS, C-terminal deletion of TbPNUTS (CΔ82) results in complete loss of JBP3/Wdr82 association (Fig. 3.6F). CΔ23 and CΔ37 PNUTS had a minor but insignificant effect on LtPNUTS-PP1-8e association (Figs. 3.3B and 3.S6A), suggesting that PP1 binding into the complex is not dependent on Wdr82 or JBP3 interaction. Taken together the data suggests the C-terminus of LtPNUTS (and TbPNUTS) is

required for binding both Wdr82 and JBP3 and that binding is independent of PP1 binding at the central RVxF motif.

Thus far, we have been unable to produce soluble recombinant protein in *E. coli* to study the PJW/PP1 complex formation in vitro. Therefore, to further test the scaffold function of PNUTS in the complex and clarify its binding relationship with Wdr82 and JBP3, we utilized the RNAi system in *T. brucei*. This system would allow us to characterize, for example, the effect of PNUTS knock-down on the interaction between Wdr82 and JBP3 by Co-IP. Therefore, we tagged JBP3 and Wdr82 with HA and Myc, respectively, in the PNUTS RNAi cell line. We find that knock-down of TbPNUTS leads to decreased protein levels of both Wdr82 and JBP3 (Fig. 3.7A). JBP3 is particularly sensitive to PNUTS knockdown, with the majority (>90%) of JBP3 being lost within 24h of PNUTS RNAi induction. On the other hand, Wdr82 is less affected with 50% reduction in protein level within 24h, with levels decreasing to ~75% reduction upon 72 hr post induction. While this effect does prevent the analysis of JBP3/Wdr82 interactions by Co-IP, it is consistent with PNUTS knockdown in HEK293 cells which leads to loss of both Tox4 and Wdr82 (15), and further supports a scaffold function for LtPNUTS. Interestingly, knockdown of Wdr82 by RNAi similarly leads to a significant reduction in HA-tagged JBP3 protein level, but does not affect Myc-tagged PNUTS protein level (Fig. 3.7B). On the other hand, ablation of JBP3 by RNAi does not lead to any change in PTP-PNUTS or Myc-Wdr82 protein levels (Fig. 3.7C). PNUTS and Wdr82 association was analyzed by anti-protein A Co-IP with or without JBP3 RNAi induction. The result shows that JBP3 knockdown does not affect PNUTS-Wdr82 Co-IP (Fig. 3.7D), indicating that Wdr82 binds to PNUTS independently of JBP3. The results collectively further suggest that JBP3 associates into the complex via binding to Wdr82, and that complex integrity is essential to Wdr82 and JBP3 protein stability.

In setting up the PP1-PNUTS Co-IP analysis and over-expressing LtPNUTS from a plasmid in cells expressing a tagged version of PP1-8e from the endogenous locus, we noticed that transfection with the PNUTS expressing plasmid led to a ~50% decrease in PP1-8e abundance (Fig. 3.8, A and B). PNUTS over-expression has no effect on PP1-7 protein level, indicating an isotype specific effect. Interestingly, the effect of PNUTS over-expression on PP1-8e levels is not dependent on PP1 binding, since this occurs even upon over-expression of the PNUTS defective for PP1 association, such as RACA PNUTS (Fig. 3.8A), $L111A_{PNUTS}$, $F118A_{PNUTS}$, or $R125A_{PNUTS}$ (Fig. 3.S11A). However, over-expression of C-terminal truncated versions of PNUTS (CΔ23 or CΔ37) did not lead to reduced PP1 protein level to the same extent as other tested PNUTS mutants, indicating that the effect is dependent on the ability of PNUTS to bind Wdr82 and/or JBP3. Furthermore, while PNUTS over-expression had no effect on Wdr82 protein abundance, in a few clones it led to a shift in mobility of Wdr82 on the SDS-PAGE gel (Figs. 3.8C and 3.S11B). Endogenously HA tagged Wdr82 has a predicted molecular weight of 48 kDa, and a majority of the protein runs slightly above the 50 kDa protein ladder marker with a minor lower molecular weight species sometimes visible just below the marker (Fig. 3.8C). We observed that WT and RACA PNUTS over-expression caused the population of Wdr82 to shift to the lower molecular weight species (Fig. 3.8C) in 2 out of 9 and 6 clones analyzed, respectively (Fig. 3.S11B). The finding that expression of WT LtPNUTS and the RACA mutants had similar effects on the altered mobility of Wdr82 *in vivo*, indicates that the effect is independent of the ability of PNUTS to bind PP1-8e. Treatment of cell lysates with or without calf intestinal phosphatase and conditions we have demonstrated to dephosphorylate RNA Pol II (254), had no effect on Wdr82 gel mobility (data not shown), excluding the possibility that the observed shift in Wdr82 is due to changes in phosphorylation status. The

AlphaFold predicted Wdr82 structure indicates the N-terminus (1-27) of Wdr82 has low prediction confidence, followed by potentially solvent exposed ^{34}FYTGIN^{39} sequence susceptible for cleavage by chymotrypsin and thermolysin (Fig. 3.S12A), suggesting a disordered N-terminus region prone to proteolytic cleavage. Preliminary MS analyses to confirm the processing of the lower MW form of Wdr82 have been inconclusive. Regardless of the explanation for the different species of Wdr82 generated by LtPNUTS overexpression, both species bind LtPNUTS to similar degree. Co-IP studies show that the full-length and truncated Wdr82 species IP similar amounts of PNUTS (Fig. 3.8, C and D), indicating the potential cleavage of the Wdr82 N-terminus does not affect PNUTS association.

While it is unclear why LtPNUTS overexpression results in these effects on Wdr82 and PP1-8e, the data further support a scaffold function for PNUTS in the PJW/PP1 complex. Furthermore, as predicted based on previous studies of PJW/PP1 complex function *in vivo* in Leishmania and *T. brucei*, these defects correlate with defects in Pol II transcription termination (Fig. 3.8, E and F). Strand-specific RT-qPCR shows that compared to the parental cells (WT), cells that over-express WT or RACA PNUTS accumulated nascent transcripts downstream of the analyzed transcription termination site (Fig. 3.8, E and F). As a positive control, cells treated with DMOG, a drug that inhibits base J synthesis and induces transcription termination defects in Leishmania (22), also accumulated readthrough transcripts. The effect of PNUTS overexpression and corresponding decreased levels of PP1-8e on RNA Pol II termination seen here is consistent with the recently characterized role of PP1-8e in Pol II phosphorylation and transcription termination in *Leishmania* (254). To address the impact of Wdr82 cleavage that occurs following over-expression of PNUTS to the termination defects measured here, we repeated the analysis using HA-tagged Wdr82 cells and examined the degree of readthrough in cells with or without

Wdr82 cleavage (Fig.3.S11C). Compared to WT cells, C-terminal tagging of Wdr82 leads to increased readthrough transcripts (Fig. 3.S11D), possibly indicating an impaired function for Wdr82-HA, similar to what we observed for C-terminally tagged PP1-8e in *L. major* (254). However, we see no difference in the degree of readthrough transcription stimulated by PNUTS overexpression in cells that resulted in Wdr82 cleavage or not (Fig. 3.S11C). Therefore, altered processing of Wdr82 in the PNUTS expressing cell lines had no additional negative effect on Pol II transcription termination. Taken together, similar to the termination defects measured in the Leishmania PP1-8e KO (254), alterations in PJW/PP1 complex formation and levels of PP1-8e following PNUTS over-expression lead to defects in Pol II transcription termination.

DISCUSSION

RIPPOs are essential regulators of PP1 substrate specificity and cellular localization. RIPPOs share little sequence or overall structural identity but use short SLiMS (5-8 amino acids long) that are combined within an unstructured domain to render RIPPOs high affinity to PP1 (268, 270, 302). According to this PP1 binding code (268), the unique combination of PP1 binding motifs (SLiMS) allows RIPPOs to interact with PP1 in a highly specific manner. PNUTS-PP1 complex involved in regulating transcription termination is conserved from mammalian to yeast cells (15, 305, 306), and recent studies indicate that the binary interaction and function also exists in trypanosomatids (252, 253). Purification of the PNUTS complex from *L. tarentolae* identified a specific interaction with the PP1-8e isoform among the eight encoded in the Leishmania genome (252, 253) suggesting that PNUTS selectively targets PP1-8e to the complex. However, the isoform selectivity of PP1 targeting in intact parasites had not been established. Here we show that PNUTS selectively targets PP1-8e to the complex and targeting

requires both the non-isoform selective canonical PP1-binding motif and additional domains located throughout the PP1-8e sequence. Previous studies have shown that LtPNUTS is a highly disordered protein, and mutation of its putative RVxF motif disrupts its interaction with LtPP1-8e, indicating its importance in PP1-PNUTS interaction (252). In the current study, we used AlphaFold to predict the LtPNUTS:PP1-8e holoenzyme complex, and identified additional SLiMs beyond the canonical RVxF motif that are typically difficult to recognize based on sequence analysis alone, because they are short and highly degenerate. Our predicted LtPNUTS:PP1-8e holoenzyme complex and biochemical studies reveal that LtPNUTS binds PP1-8e using an extended RVxF-ϕ_R-$\phi\phi$-Phe motif used by several other RIPPOs including the human PNUTS:PP1 complex. Furthermore, our studies suggest additional interactions are involved that are atypical compared with any previously studied regulator. These include unique sequences at the ends and within the catalytic domain of PP1-8e that modulate isoform specific recruitment as well as increasing overall stability of the holoenzyme complex.

Compared to the other seven LtPP1 isoforms, LtPP1-8e has a long and unique C-tail with residues 354-362 predicted to form an α-helix secondary structure, and the remaining 12 residues (363-374) unstructured. The model indicates that two residues within the C-term α-helix (P352 and I360) accommodate the Phe motif in LtPNUTS ($F118_{PNUTS}$). Supporting this model, C-terminal deletion and alanine scanning mutagenesis of PP1-8e indicates the importance of residues 359-363 of this alpha-helical region in LtPNUTS-PP1-8e interactions. The strong negative effects of the $F118A_{PNUTS}$ and $I360A_{PP1\text{-}8e}$ mutants on complex formation further supports this idea. Therefore, although the residues that constitute the conserved RVxF, ϕ_R, $\phi\phi$ and F motif binding sites are present in all LtPP1 homologs, the PP1-8e C-tail may provide a stabilizing force to the PNUTS Phe SliM and represent a significant component of isoform

selectivity. While human PP1 isoforms have a short divergent N-terminus (~6 amino acids), a role of the N-terminus in RIPPO binding or isoform selection has not been described in other systems. Deletion of the LtPP1-8e N-terminus (1-32) leads to a dramatic decrease in LtPP1-PNUTS association, indicating its significance, but how it mediates association with LtPNUTS is unclear. The low confidence of the AlphaFold structure for this region makes it difficult to understand how the N-terminus is involved in isoform selectivity. We also now identify unique inserts within the PP1-8e catalytic region (^{260}LPAGVD^{265} and ^{310}DHK^{312} and the 26 amino acid 109-134 motif) where deletion or alanine mutagenesis completely abolishes or significantly decreases PP1-PNUTS association. Mutagenesis analysis has indicated residues ^{260}LPGV^{264} and ^{109}GGTVFG^{114} as key residues within these inserted motifs, essential for PNUTS-PP1-8e complex formation. Overall, the results suggest that polymorphisms within the PP1-8e catalytic domain and N- and C-terminus are essential for PNUTS binding. As such, these regions might underlie the mechanism by which LtPNUTS shows preferential binding to PP1-8e. However, the position/orientation of the LtPP1-8 polymorphisms were, in some cases, predicted with low confidence by AlphaFold. Therefore, how they contribute to PNUTS association cannot be easily inferred. Interestingly, the LtPNUTS:PP1-8e model predicts two of these unique regions of PP1-8e (C-term and the 26 amino acid internal motif) to be in close proximity to region 116-121 of PNUTS that includes the Phe SLiM (F118) (Fig. 3.5). The predicted role of the C-term forming an essential part of the Phe binding pocket is discussed above. Within the 26-amino acid insertion polymorphism in the PP1-8e catalytic domain, ^{113}FG^{114} is predicted to be in close proximity of $Y117_{PNUTS}$ (Fig. 3.5). The importance of this region is supported by our co-IP studies where alanine mutagenesis of residues 109-114 of PP1-8e, in contrast to mutagenesis of the remaining part of this 26-amino acid insert, significantly affected PP1-PNUTS association

(Fig. 3.4). The absence of both of these regions in LtPP1-1 could therefore explain the altered Phe motif interactions in the PNUTS:PP1-1 holoenzyme model (Fig. 3.5). Taken together, the data support two unique regions of the PP1-8e isotype making critical interactions with PNUTS Phe motif that may help explain the isotype specific stable association of the LtPNUTS:PP1-8e complex.

As mentioned above, our model predicts LtPNUTS binds PP1-8e via RVxF-ϕ_R-$\phi\phi$-Phe motifs, similar to the human PNUTS:PP1 complex. hPNUTS, like most RIPPOs is able to bind all PP1 isoforms. Presumably, the additional contacts with PP1-8e specific sequences we describe here allow isoform specific binding of LtPNUTS. However, the conservation of residues involved in interactions with the extended RVxF motif in all 8 LtPP1 isoforms (Fig. 3.S7) suggests, as described for mammalian isoform specific RIPPOs, some low level of PNUTS binding *in vivo* by the remaining PP1 isoforms. This characteristic would explain the ability of other PP1 isoforms to functionally compensate for the loss of PP1-8e in *Leishmania* (254). The PNUTS-PP1-8e complex has been shown to regulate transcription termination in *Leishmania* potentially through PP1-8e-mediated dephosphorylation of Pol II CTD (254). KO of PP1-8e in *L. major* causes transcription termination defects, which can be rescued, albeit to a limited degree, by over-expression of PP1-1 or PP1-7(254). Both proteins have conserved residues constituting the RVxF motif-binding pocket, and are predicted to interact with PNUTS through a majority of the extended RVxF motif. However, they do lack the PP1-8e unique motifs we demonstrate as critical for the PNUTS-PP1-8e Co-IP, including the C-tail and the ^{113}FG^{114} residues we predict essential for stable Phe SLiM binding and thus increase overall stability of the holoenzyme complex. Therefore, it is conceivable that while the enhanced affinity for PNUTS allows LtPP1-8e to outcompete other PP1 isotypes for PNUTS binding in the WT cell, in its absence the

remaining PP1 isotypes can form unstable or transient interaction with PNUTS to partially compensate for the loss of LtPP1-8e. Similarly, the lack of these LtPP1-8e specific polymorphisms essential for the PNUTS-PP1 co-IP in all eight TbPP1 isoforms may explain the failure to identify a PP1 isoform associated with PNUTS in *T. brucei*. While TbPNUTS has a conserved RVxF motif, purification of PNUTS from *T. brucei* cells identified the Wdr82 and JBP3 homologs but no catalytic PP1 component (252). Knockdown of TbPNUTS, TbJBP3 or TbWdr82 led to defects in Pol II transcription termination (252). Thus, we predict that a similar mechanism of Pol II termination involving PP1 mediated Pol II dephosphorylation via the PJW/PP1 complex exists in *T. brucei* as we characterized in *Leishmania*. The inability to demonstrate TbPNUTS-PP1 binding using co-IP suggests that the two proteins do not interact directly or interact in such a transient or weak manner that PP1 dissociates from the complex during affinity purification process. Although we cannot exclude the possibility that the *T. brucei* PNUTS complex functions without the association of the catalytic PP1 component, the presence of an RVXF motif in TbPNUTS and lack of the polymorphisms we demonstrate as critical for stable LtPNUTS-PP1-8e interactions in the Co-IP in all *T. brucei* PP1 isoforms support our model.

The PNUTS-PP1 complex in mammalian cells is found associated with structural factors Wdr82 and Tox4, forming the PTW/PP1 complex (15). hPNUTS is a large protein with multiple domains; including the RVxF motif (KSVTW) for PP1 binding and the N-terminal TFIIS-like domain required for Tox4 binding (273). hPNUTS serves as a scaffolding protein in the PTW/PP1 complex and its ablation in HEK293 cells causes a complete loss of Tox4 and a significant reduction in Wdr82 protein level (15). LtPNUTS, on the other hand, is much smaller with no recognizable domains other than the central RVxF motif and extended SLiMs identified

here involved in PP1 binding. For the first time, we now describe that PNUTS performs similar scaffolding function in the PJW/PP1 complex in kinetoplastids, representing a key regulator of complex formation/stability. We show that ablation of TbPNUTS leads to a complete loss of JBP3, the counterpart of Tox4, and a 50% reduction in Wdr82 protein level. Moreover, over-expression of LtPNUTS leads to reduction in PP1-8e levels and processing of Wdr82 (discussed below). We demonstrate that JBP3 and Wdr82 bind the C-terminus of LtPNUTS and TbPNUTS, independent of PP1 binding. The LtPNUTS defective for PP1 binding (RACA) has no detectable loss of binding to Wdr82 or JBP3, and C-terminal mutants, unable to bind Wdr82/JBP3, bind PP1 with WT level of efficiency. While there is no apparent interaction between Wdr82/JBP3 and PP1, C-terminal deletions had a significant negative effect on both Wdr82 and JBP3 association, suggesting interdependence of PNUTS binding by Wdr82 and JBP3. Alternatively, structural alteration in PNUTS caused by C-terminal deletion could explain the negative effects on the binding of both factors. While we are not able to rule this out, the effect would have to be localized to the C-terminus as the deletions have no measurable effect on PP1 binding. Furthermore, the use of RNAi in *T. brucei* supports the interdependence of PNUTS binding by Wdr82 and JBP3, where primary interactions between PNUTS and Wdr82 regulate JBP3 binding. While the ablation of JBP3 has no effect on Wdr82 levels or interactions with TbPNUTS, Wdr82 ablation leads to specific decrease in JBP3. Presumably, in *T. brucei*, the stability of JBP3 depends on interactions with Wdr82 (and PNUTS). Additional work is needed to fully elucidate specific interactions involved in PJW/PP1 complex formation. However, taken together, the results from the *in vivo* studies suggest that PNUTS is a scaffolding protein in the PJW/PP1 complex that mediates the independent binding of PP1 and Wdr82, and JBP3 association with the complex depends, at least partially, on interactions with Wdr82.

The effects of overexpression of LtPNUTS, and ablation of TbPNUTS, on the PJW/PP1 complex supports its key role as a scaffolding factor for the complex and indicates the concentration of PNUTS is finely tuned *in vivo* in kinetoplastids. Presumably, over-expression of PNUTS in Leishmania leads to stoichiometric imbalance that affects PJW/PP1 complex formation and stability of associated factors, including PP1-8e. LtPNUTS over-expression had no detectable effect on levels of PP1-7 isotype that is not associated with the PJW/PP1 complex. Interestingly, the specific decrease of LtPP1-8e protein level is not dependent on the ability of LtPNUTS to bind PP1, but on its ability to bind Wdr82/JBP3. Over-expression of C-terminally truncated LtPNUTS (CΔ23 and CΔ37) with significantly lower affinity to Wdr82/JBP3 did not lead to a loss of LtPP1-8 as seen following over-expression of WT PNUTS or PP1-8e binding mutants. These results suggest that the integrity of the PJW/PP1 complex is important for PP1-8e protein level. Excess LtPNUTS (regardless of its ability to bind PP1) could lead to decreased levels of Wdr82/JBP3 (or other unidentified co-factors) available to PNUTS-PP1 to form a stable functional complex. The shift in molecular weight of Wdr82 in a percentage of clones overexpressing PNUTS is currently unclear. We have addressed the possibility of a shift due to phosphorylation and proposed it represents proteolytic cleavage at the unstructured N-terminus. Further work is needed to understand the effect of LtPNUTS over-expression on Wdr82 processing. However, this effect is not linked to the ability of PNUTS to bind PP1. While it is unclear if this altered Wdr82 processing affects cellular function, it has no apparent consequence on the ability of Wdr82 to bind PNUTS.

Overall, the current study identified PP1-binding motifs on LtPNUTS and discovered novel sequences on PP1-8e that could confer isoform selectivity, thereby enhancing our understanding of the PP1 binding code modulating the interaction between PP1 and PP1-

interacting proteins. Moreover, our results indicate the conserved role of PNUTS as a scaffolding protein and that its protein level is critical for PJW/PP1 complex stability. The finding that PJW/PP1 complex defects associated with PNUTS over-expression led to readthrough transcription at Pol II termination sites provides additional support for the involvement of the complex in the mechanism of Pol II transcription termination in kinetoplastids. Additional studies regarding the PJW/PP1 complex formation will help dissect the novel RNA Pol II transcription cycle in these divergent eukaryotes.

EXPERIMENTAL PROCEDURES

Protein structure modeling with AlphaFold2

The predicted models were generated using the AlphaFold2 algorithm (307) via the ColabFold platform (308). In the open source Google CoLabFold platform, sequences were pasted in the query sequence box and the complex prediction was run with the default settings. The AlphaFold model was represented by five top-scored conformations along with estimates of prediction reliability (pLDDT), as described elsewhere (307). The protein models were analyzed and displayed with UCSF ChimeraX version: 1.5 (309)**.**

Parasite culture

Promastigote form *L. tarentolae* were grown in SDM79 medium at 27 ℃. Transfections were performed as previously described (252). Briefly, 1 x 10^8 cells were transfected using he BioRad GenePulser II (2 pulses at 1.4 kV/25 uF) in 0.4 cm cuvettes in 0.5 ml cytomix (2 mM EGTA, 120 mM KCl, 0.15 mM CaCl2, 10 mM KPi pH = 7.6, 25 mM HEPES pH = 7.6, 5 mM MgCl2, 0.5% glucose, 100 μg/ml BSA, 1 mM Hypoxanthine). After transfection, cells were plated into 96-well plates to obtain clonal cell lines by limiting dilution. Where appropriate, the

following drug concentrations were used: 50 g/ml G418 and 10 g/ml Puromycin. Bloodstream form *T. brucei* expressing T7 RNA polymerase and the Tet repressor ("single marker cells") (310) were cultured in HMI-9 medium at 37 °C. Transfections were performed using the Amaxa electroporation system (Human T Cell Nucleofactor Kit, program X-001) and clonal cell lines obtained as described (252). Where appropriate, the following drug concentrations were used: 2ug/ml G418, 2.5 ug/ml Hygromycin, 2.5 ug/ml Phleomycin, 5 ug/ml Blasticidin, 0.2 ug/ml Puromycin, and 2 ug/ml Tetracycline.

DNA constructs and cell line generation

Endogenous HA-tagging in *L. tarentolae.* A background *L. tarentolae* cell line was established to express Cas9 and T7 polymerase following transfection with PacI-digested pTB007 plasmid (59) as previously described (254). To tag the endogenous PP1-8e, PP1-7d, PNUTS, JBP3 or Wdr82 locus with 6xHA tag, the Cas9/T7-expressing cell line was transfected with gRNAs and donor fragments, as previously described (254). sgRNAs were designed with LeishGEdit. Appropriate DNA fragments were generated via PCR using sgRNA primers, and transfected to cells to generate gRNAs *in vivo*. The donor fragments were amplified via PCR from the pGL2314 plasmid with 30-nucleotide homology flanks specific to the target loci. For overexpressing C-terminal tagged proteins in *L. tarentolae*, the open reading frame of LtPNUTS or LtPP1 was PCR amplified without a stop codon and inserted into the pSNSAP1 vector at the BamH1 and Xba1 sites as previously described (254). The obtained constructs were referred to as PNUTS-Pd or PP1-Pd. The desired PP1 or PNUTS mutants were generated by oligonucleotide-mediated site-directed mutagenesis (QuikChange II XL Site-Directed Mutagenesis Kit, Agilent Technologies) following the manufacturer's instructions. All final constructs were sequenced prior to electroporation. PNUTS-Pd plasmid was transfected into the

PP1-8e-HA cell line and WT *L. tarentolae* and the PP1-Pd plasmid transfected into the PNUTS-HA and WT cell line.

Endogenous tagging in *T. brucei*. For tagging the 3' end of the TbPNUTS, Wdr82, and JBP3 with 3xHA or Myc tag in *T. brucei* cells, a PCR-based approach was used with the pMOTag4H or pMOTag3M vectors as described(252). For tagging PNUTS with the PTP tag in *T. brucei*, the 3' end of TbPNUTS was cloned in the ApaI and Not1 sites of the Pc-PTP-Neo vector (311) where the Neomycin resistance drug marker was replaced with a blasticidin resistance drug marker. The vector was then linearized by restriction enzyme digestion within the 3' end of the TbPNUTS gene, and used in transfection. For tetracycline regulated expression of PNUTS in *T. brucei*, the ORF with a C-terminal PTP tag was amplified by PCR and cloned into the HindIII and BamH1 sites of the pLew100V5 plasmid. The final construct (PNUTS-PTP-pLew100), was linearized with NotI prior to transfection. To induce PNUTS expression, tetracycline was added at 2 ug/ml. All final constructs were sequenced prior to electroporation. Primers sequences used are available upon request.

RNAi analysis

Conditional silencing of PNUTS, JBP3 and Wdr82 in *T. brucei* BF SMC was performed as previously described (252). Briefly, a fragment of the representative ORF was integrated into the BamHI site of the p2T7-177 vector. I-SceI linearized p2T7-177 constructs were transfected into BF SMC for targeted integration into the 177 bp repeat locus. All final constructs were sequenced prior to transfection. RNAi was induced with 2 μg/ml Tetracycline and growth was monitored daily in triplicate.

Co-immunoprecipitation

5×10^8 of *L. tarentolae* cells were lysed in lysis buffer and Pd-tagged protein was affinity purified using 50 ul IgG Sepharose beads as previously described (37). After incubation with cell extract for 4 hrs at 4 ℃, the IgG beads were washed 3 times in 10 ml PA-150 buffer. After the final wash, the beads were boiled for 5 min in 1x SDS-PAGE sample buffer. Samples were run on 10% sodium dodecyl sulfate polyacrylamide gel and transferred to nitrocellulose membrane for western blotting with anti-protein A and anti-HA antibodies. 1.2 x 10^8 of *T. brucei* cells expressing PTP-tagged protein was used for co-immunoprecipitation as described above. Samples were run on 10% sodium dodecyl sulfate polyacrylamide gel and transferred to nitrocellulose membrane for western blotting with anti-protein A, anti-HA and Anti-Myc antibodies.

Western blotting

Proteins from 1.4 x 10^7 cell equivalents of *L. tarentolae*, or 3 x 10^6 of *T. brucei* cells, were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS page 10% gel), transferred to nitrocellulose membrane (cytiva, 10600011). Membranes were blocked with Intercept Blocking Buffer (LiCor, 927-60001) for 1 hour at room temperature and then probed with anti-protein A (Sigma, P3775, 1:5000) for 1 hour at room temperature, anti-MYC (Santa Cruz, 9E10, 1:1000) overnight at 4 ℃, anti-HA antibodies (Sigma, 3F10, 1:3000) overnight at 4 ℃ or anti-Elongation Factor 1A (Sigma, 05-235, 1:20000) for 1 hour at room temperature. Blots were then washed in TBST (Tris-buffered saline + 0.05% Tween 20), and incubated with IRDye 800CW Goat anti-Rabbit (LiCor, 926-32211), IRDye 600RD Goat anti-Mouse (LicCor, 926-68070), or IRDye 600RD Goat anti-Rat (LicCor, 926-68076) for 1 hour at room temperature. The secondary antibodies were used at 1:10000 dilution in Intercept Blocking Buffer. Blots were

then washed in TBST, rinsed in TBS, and imaged with the Image Studio Lite imaging system and software (LiCor).

Strand-specific RT-qPCR

Total RNA was extracted using the Tripure Isolation Reagent (Roche). To synthesize cDNA, 1 ug of Turbo™ DNase-treated total RNA (ThermoFisher) was reverse-transcribed with strand-specific oligonucleotides using Superscript™ III kit (ThermoFisher), following the manufacturer's instructions. Quantification of the resulting cDNA was conducted using an iCycler with an iQ5 multicolor real-time PCR detection system (Bio-Rad). Triplicate cDNA samples were assessed and normalized against tubulin cDNA. For the qPCR reaction, a 15 ul mixture containing 5 ul of cDNA, 4.5 pmol each of sense and antisense primers, and 7.5 ul of 2× iQ SYBR green super mix (Bio-Rad Laboratories) was used. Standard curves were generated for each gene using 5-fold dilutions of a known quantity (100 ng/l) of WT gDNA. The quantities were determined using the iQ5 optical detection system software.

DATA AVAILABILITY

All data described in this study are presented in the article and accompanying supporting information.

FUNDING

This work was supported by the National Institutes of Health (grant number R01AI109108) to R.S. The contents is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

1. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., and Nigg, E. A. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signaling*. **3**, ra3–ra3
2. Peti, W., Nairn, A. C., and Page, R. (2013) Structural basis for protein phosphatase 1 regulation and specificity. *FEBS J.* **280**, 596–611
3. Shi, Y. (2009) Serine/threonine phosphatases: mechanism through structure. *Cell*. **139**, 468–484
4. Ceulemans, H., and Bollen, M. (2006) A tighter RVxF motif makes a finer Sift. *Chemistry & biology*. **13**, 6–8
5. Cohen, P. T. W. (2002) Protein phosphatase 1–targeted in many directions. *Journal of cell science*. **115**, 241–256
6. Moorhead, G. B. G., Trinkle-Mulcahy, L., and Ulke-Lemée, A. (2007) Emerging roles of nuclear protein phosphatases. *Nature reviews Molecular cell biology*. **8**, 234–244
7. Hoermann, B., Kokot, T., Helm, D., Heinzlmeir, S., Chojnacki, J. E., Schubert, T., Ludwig, C., Berteotti, A., Kurzawa, N., Kuster, B., Savitski, M. M., and Köhn, M. (2020) Dissecting the sequence determinants for dephosphorylation by the catalytic subunits of phosphatases PP1 and PP2A. *Nat. Commun.* **11**, 3583

8. Hoermann, B., and Köhn, M. (2021) Evolutionary crossroads of cell signaling: PP1 and PP2A substrate sites in intrinsically disordered regions. *Biochem. Soc. Trans.* **49**, 1065–1074

9. Rogers, S., McCloy, R., Watkins, D. N., and Burgess, A. (2016) Mechanisms regulating phosphatase specificity and the removal of individual phosphorylation sites during mitotic exit. *Bioessays*. **38 Suppl 1**, S24-32

10. Wu, D., De Wever, V., Derua, R., Winkler, C., Beullens, M., Van Eynde, A., and Bollen, M. (2018) A substrate-trapping strategy for protein phosphatase PP1 holoenzymes using hypoactive subunit fusions. *J. Biol. Chem.* **293**, 15152–15162

11. Köhn, M. (2020) Turn and face the strange: § a new view on phosphatases. *ACS central science*. **6**, 467–477

12. Felgueiras, J., Jeronimo, C., and Fardilha, M. (2020) Protein phosphatase 1 in tumorigenesis: is it worth a closer look? *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*. **1874**, 188433

13. Heroes, E., Lesage, B., Görnemann, J., Beullens, M., Van Meervelt, L., and Bollen, M. (2013) The PP1 binding code: a molecular-lego strategy that governs specificity. *The FEBS journal*. **280**, 584–595

14. Bollen, M., Peti, W., Ragusa, M. J., and Beullens, M. (2010) The extended PP1 toolkit: designed to create specificity. *Trends in biochemical sciences*. **35**, 450–458

15. Peti, W., Nairn, A. C., and Page, R. (2012) Folding of Intrinsically Disordered Protein Phosphatase 1 Regulatory Proteins. *Curr. Phys. Chem.* **2**, 107–114

16. Verbinnen, I., Ferreira, M., and Bollen, M. (2017) Biogenesis and activity regulation of protein phosphatase 1. *Biochemical Society Transactions*. **45**, 89–99

17. Allen, P. B., Kwon, Y.-G., Nairn, A. C., and Greengard, P. (1998) Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit. *Journal of Biological Chemistry*. **273**, 4089–4095
18. Kreivi, J.-P., Trinkle-Mulcahy, L., Lyon, C. E., Morrice, N. A., Cohen, P., and Lamond, A. I. (1997) Purification and characterisation of p99, a nuclear modulator of protein phosphatase 1 activity. *FEBS letters*. **420**, 57–62
19. Raha-Chowdhury, R., Andrews, S. R., and Gruen, J. R. (2005) CAT 53: a protein phosphatase 1 nuclear targeting subunit encoded in the MHC Class I region strongly expressed in regions of the brain involved in memory, learning, and Alzheimer's disease. *Molecular brain research*. **138**, 70–83
20. Landsverk, H. B., Kirkhus, M., Bollen, M., Küntziger, T., and Collas, P. (2005) PNUTS enhances in vitro chromosome decondensation in a PP1-dependent manner. *Biochemical Journal*. **390**, 709–717
21. Kim, Y.-M., Watanabe, T., Allen, P. B., Kim, Y.-M., Lee, S.-J., Greengard, P., Nairn, A. C., and Kwon, Y.-G. (2003) PNUTS, a protein phosphatase 1 (PP1) nuclear targeting subunit: characterization of its PP1-and RNA-binding domains and regulation by phosphorylation. *Journal of Biological Chemistry*. **278**, 13819–13828
22. Landsverk, H. B., Mora-Bermúdez, F., Landsverk, O. J. B., Hasvold, G., Naderi, S., Bakke, O., Ellenberg, J., Collas, P., Syljuåsen, R. G., and Küntziger, T. (2010) The protein phosphatase 1 regulator PNUTS is a new component of the DNA damage response. *EMBO reports*. **11**, 868–875
23. Austenaa, L. M. I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B. A. M., and de Pretis, S. (2015) Transcription of

mammalian cis-regulatory elements is restrained by actively enforced early termination. *Molecular cell*. **60**, 460–474

24. Kim, H., Lee, O.-H., Xin, H., Chen, L.-Y., Qin, J., Chae, H. K., Lin, S.-Y., Safari, A., Liu, D., and Songyang, Z. (2009) TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nature structural & molecular biology*. **16**, 372–379
25. Choy, M. S., Hieke, M., Kumar, G. S., Lewis, G. R., Gonzalez-DeWhitt, K. R., Kessler, R. P., Stein, B. J., Hessenberger, M., Nairn, A. C., and Peti, W. (2014) Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proceedings of the National Academy of Sciences*. **111**, 4097–4102
26. Choy, M. S., Page, R., and Peti, W. (2012) Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochemical Society Transactions*. **40**, 969–974
27. Hendrickx, A., Beullens, M., Ceulemans, H., Den Abt, T., Van Eynde, A., Nicolaescu, E., Lesage, B., and Bollen, M. (2009) Docking motif-guided mapping of the interactome of protein phosphatase-1. *Chemistry & biology*. **16**, 365–371
28. Meiselbach, H., Sticht, H., and Enz, R. (2006) Structural analysis of the protein phosphatase 1 docking motif: molecular description of binding specificities identifies interacting proteins. *Chemistry & biology*. **13**, 49–59
29. Wakula, P., Beullens, M., Ceulemans, H., Stalmans, W., and Bollen, M. (2003) Degeneracy and function of the ubiquitous RVXF motif that mediates binding to protein phosphatase-1. *Journal of Biological Chemistry*. **278**, 18817–18823

30. Lee, J.-H., You, J., Dobrota, E., and Skalnik, D. G. (2010) Identification and characterization of a novel human PP1 phosphatase complex. *Journal of Biological Chemistry*. **285**, 24466–24476

31. Nedea, E., Nalbant, D., Xia, D., Theoharis, N. T., Suter, B., Richardson, C. J., Tatchell, K., Kislinger, T., Greenblatt, J. F., and Nagy, P. L. (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular cell*. **29**, 577–587

32. Park, K., Zhong, J., Jang, J. S., Kim, J., Kim, H.-J., Lee, J.-H., and Kim, J. (2022) ZWC complex-mediated SPT5 phosphorylation suppresses divergent antisense RNA transcription at active gene promoters. *Nucleic acids research*. **50**, 3835–3851

33. Liu, Z., Wu, A., Wu, Z., Wang, T., Pan, Y., Li, B., Zhang, X., and Yu, M. (2022) TOX4 facilitates promoter-proximal pausing and C-terminal domain dephosphorylation of RNA polymerase II in human cells. *Communications Biology*. **5**, 300

34. He, X., and Moore, C. (2005) Regulation of yeast mRNA 3′ end processing by phosphorylation. *Molecular cell*. **19**, 619–629

35. Connelly, S., and Manley, J. L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development*. **2**, 440–452

36. Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*. **432**, 517–522

37. West, S., Gromak, N., and Proudfoot, N. J. (2004) Human 5′→ 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*. **432**, 522–525

38. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D. L. (2019) Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "sitting duck torpedo" mechanism. *Molecular cell*. **76**, 896-908. e4

39. Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., and Fisher, R. P. (2018) A Cdk9–PP1 switch regulates the elongation–termination transition of RNA polymerase II. *Nature*. **558**, 460–464

40. Kecman, T., Kuś, K., Heo, D.-H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C. V., and Vasiljeva, L. (2018) Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. *Cell reports*. **25**, 259-269. e5

41. Thomas, S., Green, A., Sturm, N. R., Campbell, D. A., and Myler, P. J. (2009) Histone acetylations mark origins of polycistronic transcription in Leishmania major. *BMC genomics*. **10**, 1–15

42. El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., and Hertz-Fowler, C. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science*. **309**, 404–409

43. Boothroyd, J. C., and Cross, G. A. M. (1982) Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5′ end. *Gene*. **20**, 281–289

44. Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P., Majumder, H. K., Weber, H., Veeneman, G. H., and Van Boom, J. (1982) RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. *Nucleic acids research*. **10**, 3591–3604
45. De Lange, T., Liu, A. Y. C., Borst, P., Tromp, M. C., Van Boom, J. H., and Van der Ploeg, L. H. T. (1983) Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? *Cell*. **34**, 891–900
46. Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983) Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. *Cell*. **34**, 901–909
47. Sutton, R. E., and Boothroyd, J. C. (1986) Evidence for trans splicing in trypanosomes. *Cell*. **47**, 527–535
48. Agabian, N. (1990) Trans splicing of nuclear pre-mRNAs. *Cell*. **61**, 1157–1160
49. Borst, P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. *Annual review of biochemistry*. **55**, 701–732
50. LeBowitz, J. H., Smith, H. Q., Rusche, L., and Beverley, S. M. (1993) Coupling of poly (A) site selection and trans-splicing in Leishmania. *Genes & development*. **7**, 996–1007
51. Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X., Dewell, S., and Cross, G. A. M. (2009) Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes & development*. **23**, 1063–1076

52. Respuela, P., Ferella, M., Rada-Iglesias, A., and Åslund, L. (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi. *Journal of Biological Chemistry*. **283**, 15884–15892

53. Cliffe, L. J., Siegel, T. N., Marshall, M., Cross, G. A. M., and Sabatini, R. (2010) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. *Nucleic acids research*. **38**, 3923–3935

54. Dharmalingam, K., and Goldberg, E. B. (1979) Beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. *Virology*. **96**, 393–403

55. Van Leeuwen, F., Taylor, M. C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., and Borst, P. (1998) β-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proceedings of the National Academy of Sciences*. **95**, 2366–2371

56. Dooijes, D., Chaves, I., Kieft, R., Dirks-Mulder, A., Martin, W., and Borst, P. (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. *Nucleic acids research*. **28**, 3017–3021

57. van Luenen, H. G., Farris, C., Jan, S., Genest, P.-A., Tripathi, P., Velds, A., Kerkhoven, R. M., Nieuwland, M., Haydock, A., and Ramasamy, G. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell*. **150**, 909–921

58. Reynolds, D., Cliffe, L., Förstner, K. U., Hon, C.-C., Siegel, T. N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic acids research*. **42**, 9717–9729

59. Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., and Sabatini, R. (2016) Histone H3 variant regulates RNA polymerase II transcription termination and dual strand transcription of siRNA loci in Trypanosoma brucei. *PLoS genetics*. **12**, e1005758

60. Kim, H.-S. (2021) Genetic interaction between site-specific epigenetic marks and roles of H4v in transcription termination in trypanosoma brucei. *Frontiers in Cell and Developmental Biology*. **9**, 744878

61. Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H.-S. (2016) Base J and H3. V regulate transcriptional termination in Trypanosoma brucei. *PLoS genetics*. **12**, e1005762

62. Kieft, R., Zhang, Y., Marand, A. P., Moran, J. D., Bridger, R., Wells, L., Schmitz, R. J., and Sabatini, R. (2020) Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. *PLoS Genetics*. **16**, e1008390

63. Jensen, B. C., Phan, I. Q., McDonald, J. R., Sur, A., Gillespie, M. A., Ranish, J. A., Parsons, M., and Myler, P. J. (2021) Chromatin-associated protein complexes link DNA base J and transcription termination in Leishmania. *Msphere*. **6**, 10.1128/msphere. 01204-20

64. Kieft, R., Zhang, Y., Yan, H., Schmitz, R. J., and Sabatini, R. (2023) Knockout of protein phosphatase 1 in Leishmania major reveals its role during RNA polymerase II transcription termination. *Nucleic Acids Research*

65. Pinheiro, A. S., Marsh, J. A., Forman-Kay, J. D., and Peti, W. (2011) Structural Signature of the MYPT1− PP1 Interaction. *Journal of the American Chemical Society*. **133**, 73–80

66. Terrak, M., Kerff, F., Langsetmo, K., Tao, T., and Dominguez, R. (2004) Structural basis of protein phosphatase 1 regulation. *Nature*. **429**, 780–784

67. Carmody, L. C., Baucum, A. J., Bass, M. A., and Colbran, R. J. (2008) Selective targeting of the γ1 isoform of protein phosphatase 1 to F-actin in intact cells requires multiple domains in spinophilin and neurabin. *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. **22**, 1660

68. Kumar, G. S., Gokhan, E., De Munter, S., Bollen, M., Vagnarelli, P., Peti, W., and Page, R. (2016) The Ki-67 and RepoMan mitotic phosphatases assemble via an identical, yet novel mechanism. *Elife*. **5**, e16539

69. Bertran, M. T., Mouilleron, S., Zhou, Y., Bajaj, R., Uliana, F., Kumar, G. S., van Drogen, A., Lee, R., Banerjee, J. J., and Hauri, S. (2019) ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. *Nature communications*. **10**, 771

70. Srivastava, G., Bajaj, R., Kumar, G. S., Gaudreau-Lapierre, A., Nicolas, H., Chamousset, D., Kreitler, D., Peti, W., Trinkle-Mulcahy, L., and Page, R. (2022) The ribosomal RNA processing 1B: protein phosphatase 1 holoenzyme reveals non-canonical PP1 interaction motifs. *Cell reports*

71. Llanos, S., Royer, C., Lu, M., Bergamaschi, D., Lee, W. H., and Lu, X. (2011) Inhibitory member of the apoptosis-stimulating proteins of the p53 family (iASPP) interacts with protein phosphatase 1 via a noncanonical binding motif. *Journal of Biological Chemistry*. **286**, 43039–43044

72. Zhou, Y., Millott, R., Kim, H. J., Peng, S., Edwards, R. A., Skene-Arnold, T., Hammel, M., Lees-Miller, S. P., Tainer, J. A., and Holmes, C. F. B. (2019) Flexible tethering of ASPP proteins facilitates PP-1c catalysis. *Structure*. **27**, 1485-1496. e4

73. Lesage, B., Beullens, M., Ceulemans, H., Himpens, B., and Bollen, M. (2005) Determinants of the nucleolar targeting of protein phosphatase-1. *FEBS letters*. **579**, 5626–5630

74. Egloff, M.-P., Cohen, P. T. W., Reinemer, P., and Barford, D. (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *Journal of molecular biology*. **254**, 942–959

75. Goldberg, J., Huang, H.-B., Kwon, Y.-G., Greengard, P., Nairn, A. C., and Kuriyan, J. (1995) Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature*. **376**, 745–753

76. Ragusa, M. J., Dancheck, B., Critton, D. A., Nairn, A. C., Page, R., and Peti, W. (2010) Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nature structural & molecular biology*. **17**, 459–464

77. Kumar, G. S., Choy, M. S., Koveal, D. M., Lorinsky, M. K., Lyons, S. P., Kettenbach, A. N., Page, R., and Peti, W. (2018) Identification of the substrate recruitment mechanism of the muscle glycogen protein phosphatase 1 holoenzyme. *Science Advances*. **4**, eaau6044

78. Egloff, M.-P., Johnson, D. F., Moorhead, G., Cohen, P. T. W., Cohen, P., and Barford, D. (1997) Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *The EMBO journal*. **16**, 1876–1887

79. Benjamin, B., Sanchez, A. M., Garg, A., Schwer, B., and Shuman, S. (2021) Structure-function analysis of fission yeast cleavage and polyadenylation factor (CPF) subunit Ppn1 and its interactions with Dis2 and Swd22. *PLoS Genetics*. **17**, e1009452

80. Carminati, M., Manav, M. C., Bellini, D., and Passmore, L. A. (2022) A direct interaction between CPF and Pol II links RNA 3′-end processing to transcription. *BioRxiv*

81. Vanoosthuyse, V., Legros, P., van Der Sar, S. J. A., Yvert, G., Toda, K., Le Bihan, T., Watanabe, Y., Hardwick, K., and Bernard, P. (2014) CPF-associated phosphatase activity opposes condensin-mediated chromosome condensation. *PLoS genetics*. **10**, e1004415

82. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., and Potapenko, A. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589

83. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nature methods*. **19**, 679–682

84. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*. **30**, 70–82

85. Wirtz, E., Leal, S., Ochatt, C., and Cross, G. M. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Molecular and biochemical parasitology*. **99**, 89–101

86. Schimanski, B., Nguyen, T. N., and Günzl, A. (2005) Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. *Eukaryotic cell*. **4**, 1942–1950

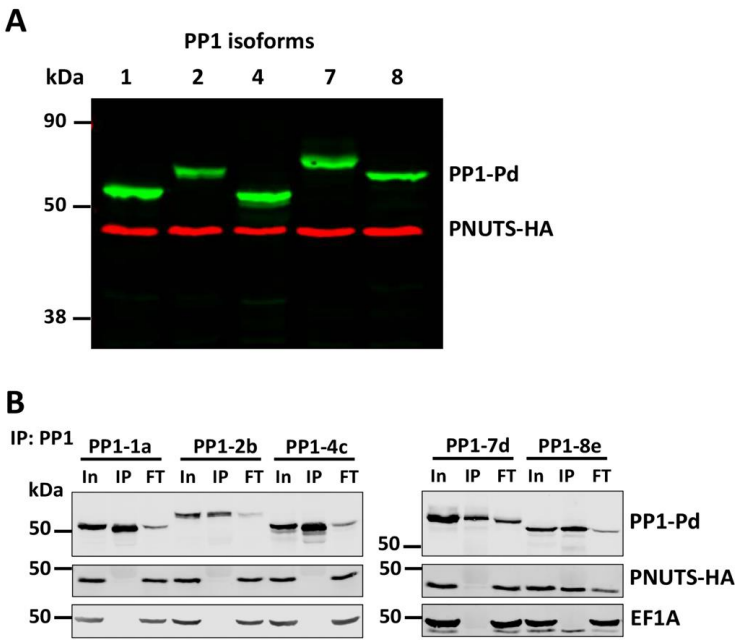


Figure 3.1. PNUTS binding is specific for the PP1-8e isoform. A, expression of PNUTS and PP1 isoforms in *L. tarentolae*. Cell extracts from *L. tarentolae* cells that endogenously express HA-tagged PNUTS, and exogenously express Pd-tagged PP1 isoforms from the pSNSAP1 vector were analyzed by western blotting with anti-HA and anti-protein A. B, PNUTS/PP1 Co-IP analysis. Lysates from the indicated cell lines were purified by anti-protein A affinity resin and analyzed by western blotting with anti-protein A and anti-HA. Equal cell equivalents of input (In), precipitated immunocomplexes (IP), and flow-through or unbound fraction (FT) were loaded on the gel. EF1A serves as a loading and negative control for the Co-IP.

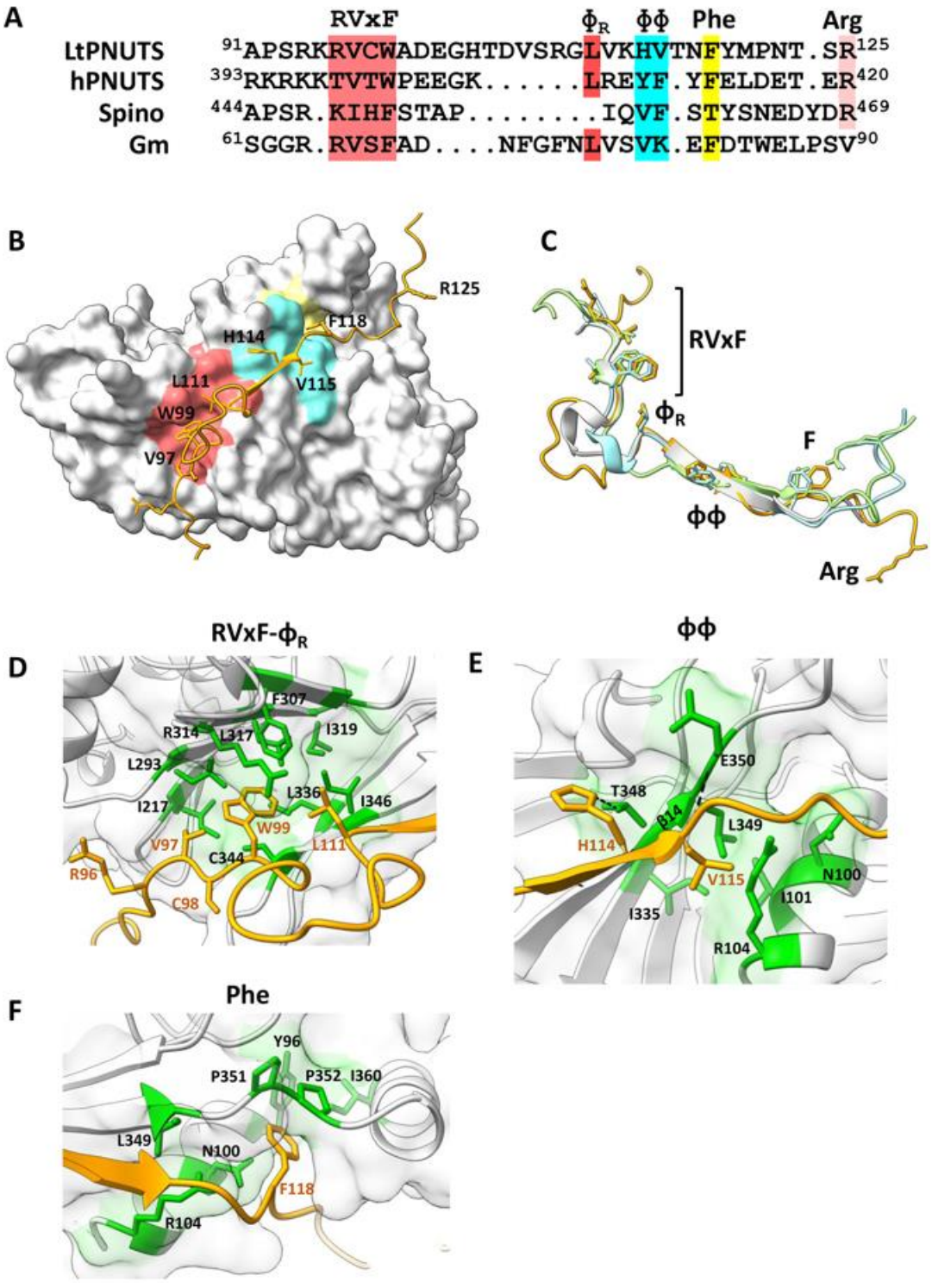


Figure 3.2. Predicted LtPNUTS-PP1-8e interacting motifs. A, structure-based sequence alignment of the PP1-interacting motifs of LtPNUTS, hPNUTS, spinophilin and Gm, with PP1 interacting residues indicated. Gm; muscle-specific glycogen-targeting subunit of PP1. B,

predicted structure of the LtPNUTS:PP1-8e complex. LtPNUTS is shown as orange ribbon with key interacting residues shown as sticks and LtPP1-8e is shown as a grey surface. LtPNUTS residues ^{96}RVCW^{99} and L111 are predicted to bind to the RVxF binding pocket (red), LtPNUTS residues ^{114}HV^{115} are predicted to bind to the PP1 ΦΦ binding pocket (cyan), and LtPNUTS residue F118 is predicted to binds to the Phe binding pocket (yellow). The colored regions of PP1-8e correspond to the zoomed in pockets shown in D, E and F. C, overlay of the RVxF and ΦΦ-F structures of four PP1 regulators, LtPNUTS (orange), hPNUTS (green, 4mp0), Spinophilin (blue, 3egg) and Gm (grey, 6dno), with residues binding RVxF-Φ_R, ΦΦ, and Phe pockets shown as sticks. The Arginine residue (R125) of LtPNUTS that deviates from the hPNUTS and Spinophilin structure is indicated. D-F, major binding interactions between LtPNUTS (orange sticks) and PP1 (surface). The well-established SLiM binding pockets (D, RVxF; E, ΦΦ; F, Phe) are shown. Key interacting residues in LtPP1-8 (black) and LtPNUTS (orange) are labelled. Predicted salt bridge interactions between PP1 residues E350 and T348 with the PNUTS ΦΦ motif indicated by dashed line in E.

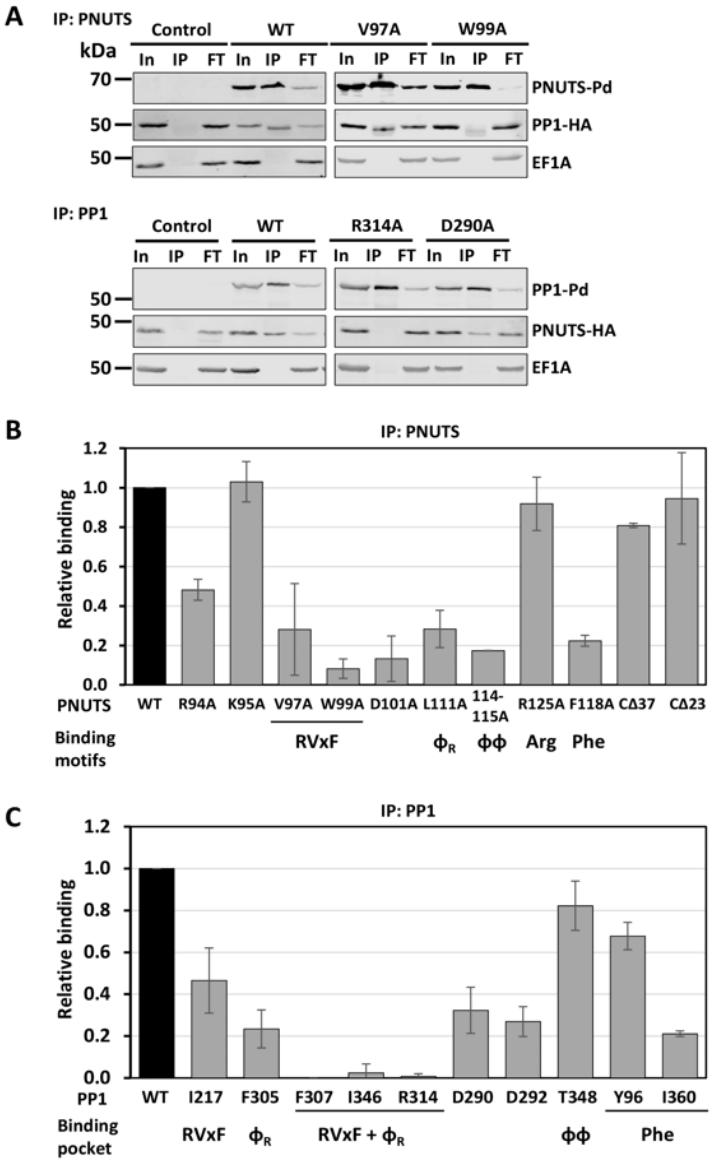


Figure 3.3. LtPNUTS binds LtPP1-8e using an extended RVxF- Φ_R-ϕϕ-Phe motif. A, co-immunoprecipitation assay of PP1-8e binding to PNUTS and their derivatives. PNUTS IP; PP1-8e was endogenously tagged with HA tag, and wild type or indicated PNUTS mutants with Pd tag were over-expressed from the pSNSAP1 vector. Cell extracts from the indicated cell lines were purified by anti-protein A affinity resin and analyzed by western blot with anti-protein A and anti-HA. In; input (equivalent to the amount of protein added to the IP reaction mixture), IP; 100% of the precipitated immunocomplexes, FT; flow through or non-bound supernatant. EF1α

provides a loading control and negative control for the IP. PP1 IP; PNUTS was endogenously tagged with HA tag, and wild type or indicated PP1-8e mutants with Pd tag were over-expressed from the pSNSAP1 vector. The levels of PNUTS pulled down in the PP1 IP were assessed by western blot as described above. Additional PNUTS and PP1-8e mutations analyzed by Co-IP are shown in Fig. 3.S6. B and C, the relative binding (%IP) between PNUTS and PP1-8e (WT and variants) determined by the ratio of the band intensity of IP to that of In. B, PNUTS-Pd IP. The bar graph represents the mean ± SD from three independent experiments, with the % IP of PP1 using WT PNUTS set to 1. The PNUTS binding motif that corresponds to the residue tested, according to the model in Fig. 3.2, is indicated at the bottom of the graph. CΔ23 and CΔ37 refer to C-terminal truncations of PNUTS described in Fig. 3.4. C, PP1-8e-Pd IP. The % IP of PNUTS from the PP1 pull-down (WT and mutants) was determined as in B. The PNUTS binding pocket represented by each residue of PP1 is indicated below. F307, I346 and R314 are predicted to be key residues of both the RVxF and Φ_R binding pocket of PP1-8e.

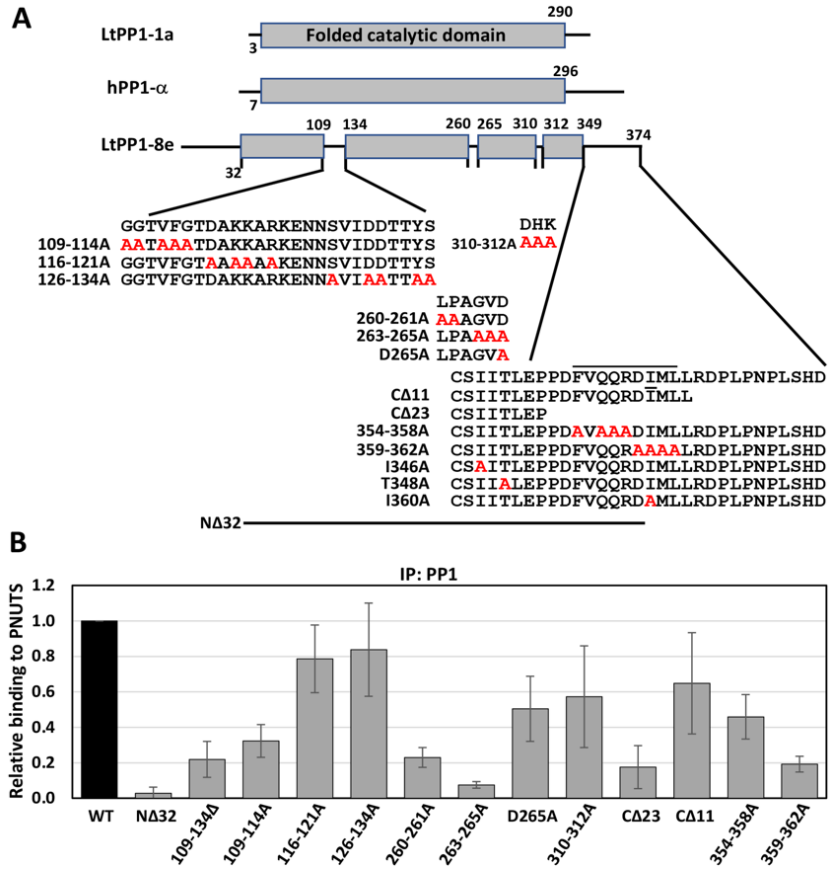


Figure 3.4. Non-canonical sites on PP1-8e are essential for PNUTS binding. A, PP1-8e constructs. The conserved PP1 catalytic domain is shown as grey boxes. Isoform differences between LtPP1-8e and LtPP1a and hPP1 are indicated by the lines within the catalytic domain and at the N- and C-terminus of LtPP1-8e. Sequence is provided for all these regions in LtPP1-8e, except the N-terminus, and residues subjected to alanine mutagenesis (red) or deletion are indicated. Residues in the predicted □-helix at the C-terminus are indicated by the line above the sequence. B, The % IP of PNUTS from the PP1 pull-down (WT and indicated variants) was determined as described in Fig. 3.3C. The bar graph represents the mean ± SD from three independent experiments, with the %IP from wild type PP1 set to 1. See Fig. 3.S9.

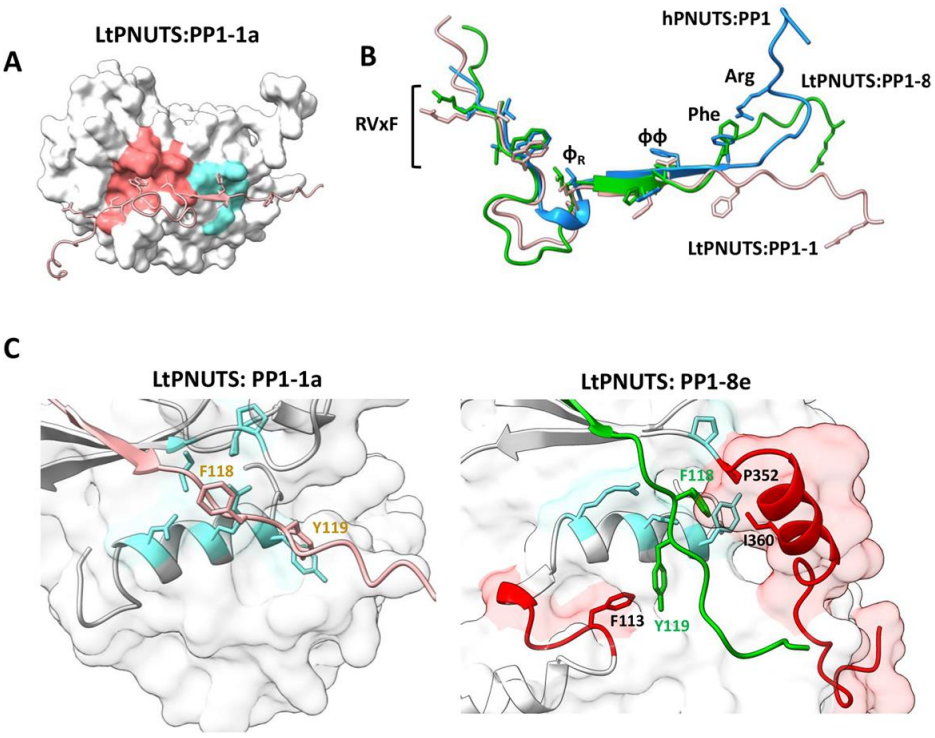


Figure 3.5. The predicted LtPNUTS-PP1-1a structure. A, The predicted holoenzyme structure of LtPNUTS (pink ribbon) and LtPP1-1a (white surface). The RVxF binding pocket (red) and ΦΦ binding pocket (cyan) are shaded on PP1 surface. B, Structural comparison of the PP1-binding domains of LtPNUTS in complex with PP1-8e (green) or PP1-1a (pink). Structure of hPNUTS (blue) bound to hPP1 is also shown. C, Close-up view of $F118_{PNUTS}$ (pink) in complex with LtPP1-1a (white surface, left) or (green) with LtPP1-8e (white surface, right). F118 and Y119 of PNUTS are shown as sticks and labelled. Conserved Phe-binding pocket residues (according the hPNUTS:PP1 structure) are shown in blue sticks. LtPNUTS:PP1-8e complex (right). The C-terminus of LtPP1-8e is shown in red carton, and key residues (P352 and I360) shown to be important for LtPNUTS binding are shown in sticks and labelled. Residues ^{109}GGTVFG^{114} within the PP1 catalytic motif and important for PNUTS binding are also shown in red, and residue ^{113}F shown as sticks and labelled.

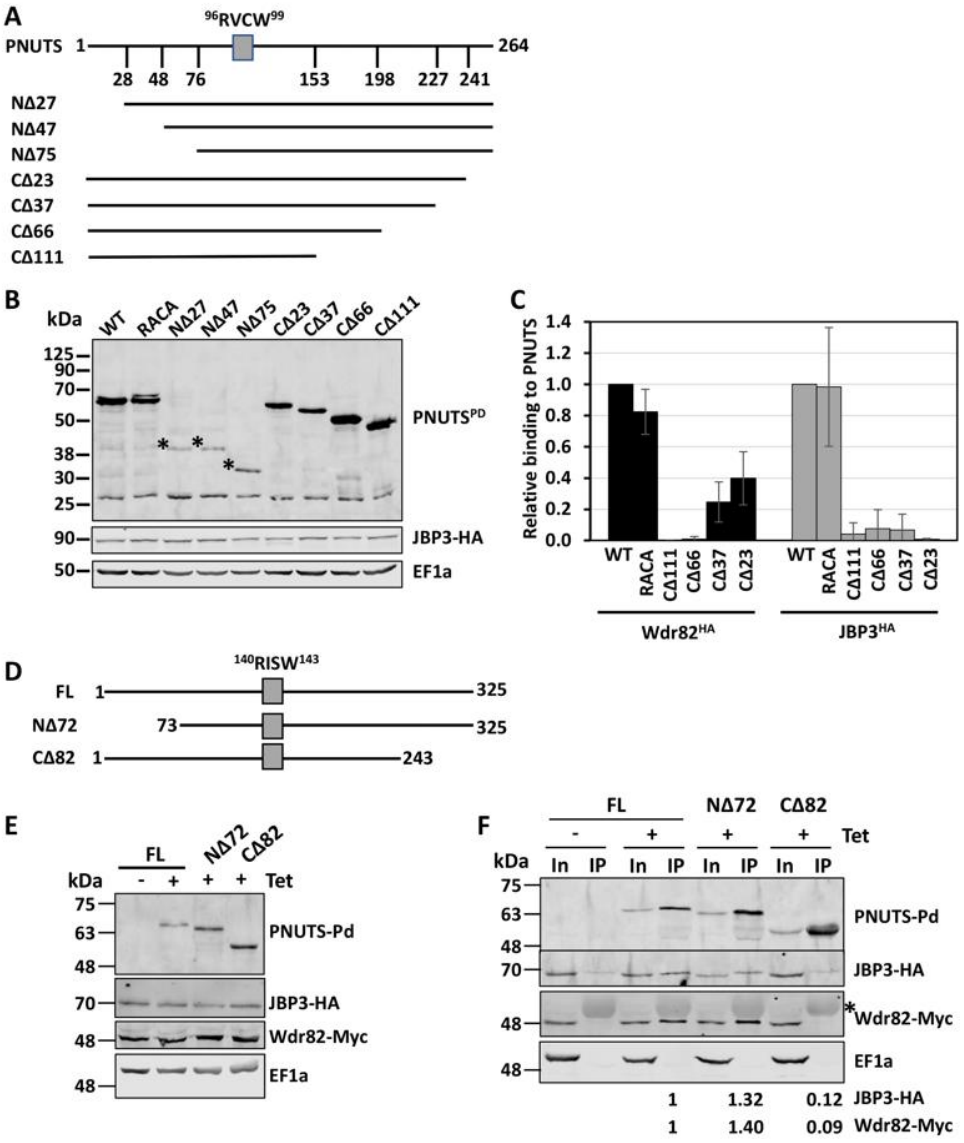


Figure 3.6. JBP3 and Wdr82 bind to the C-terminus of PNUTS. (A-C) Co-immunoprecipitation analysis of LtPNUTS and Wdr82 and JBP3. A, schematic diagram of PNUTS depicting the PP1-specific RVxF SLiM (RVCW). Constructs used in this study are illustrated. B, western blot showing the protein expression of Pd-tagged PNUTS (WT and truncation mutants) in JBP3-HA tagged *L. tarentolae* cells. Dots indicate the proposed products representing the indicated N-terminal truncations. Anti-EF1A western blot is shown as a loading control. C, analysis of JBP3/Wdr82 binding to PNUTS by Co-IP. PNUTS truncations (C) or

mutants (D) were tested for interaction with HA-tagged Wdr82 or JBP3 by Co-IP analysis. %IP of WT Wdr82 and JBP3 by PNUTS were set to one and relative %IP of the indicated mutants was determined as described in Fig. 3.3. The bar graph represents the mean ± SD from three independent experiments. See Fig. 3.S10. (D-F) Co-immunoprecipitation analysis of TbPNUTS and Wdr82 and JBP3. D, Schematic representation of the TbPNUTS truncations. The putative RVXF motif is indicated by a grey box. E, JBP3 and Wdr82 were endogenously tagged with HA and Myc tags, respectively. The protein expression of the indicated TbPNUTS was induced by addition of tetracycline (Tet) for 24 hrs and lysates analyzed by western blot with anti-protein A, anti-HA, anti-Myc, or anti-EF1a. EF1a serves as a loading control. F, Lysates of the indicated cell lines with or without tetracycline induction were purified by anti-protein A affinity resin and analyzed by western blot with anti-protein A, anti-HA, anti-Myc and anti-EF1a antibodies. FL, full-length (WT) TbPNUTS. Asterisk indicates the IgG cross-reactive signal in the IP fraction from anti-Myc. Protein A purification results in low background JBP3-HA signal in the absence of protein A-tagged PNUTS. %IP is quantified from two replicates and shown below for the corresponding cell lines.

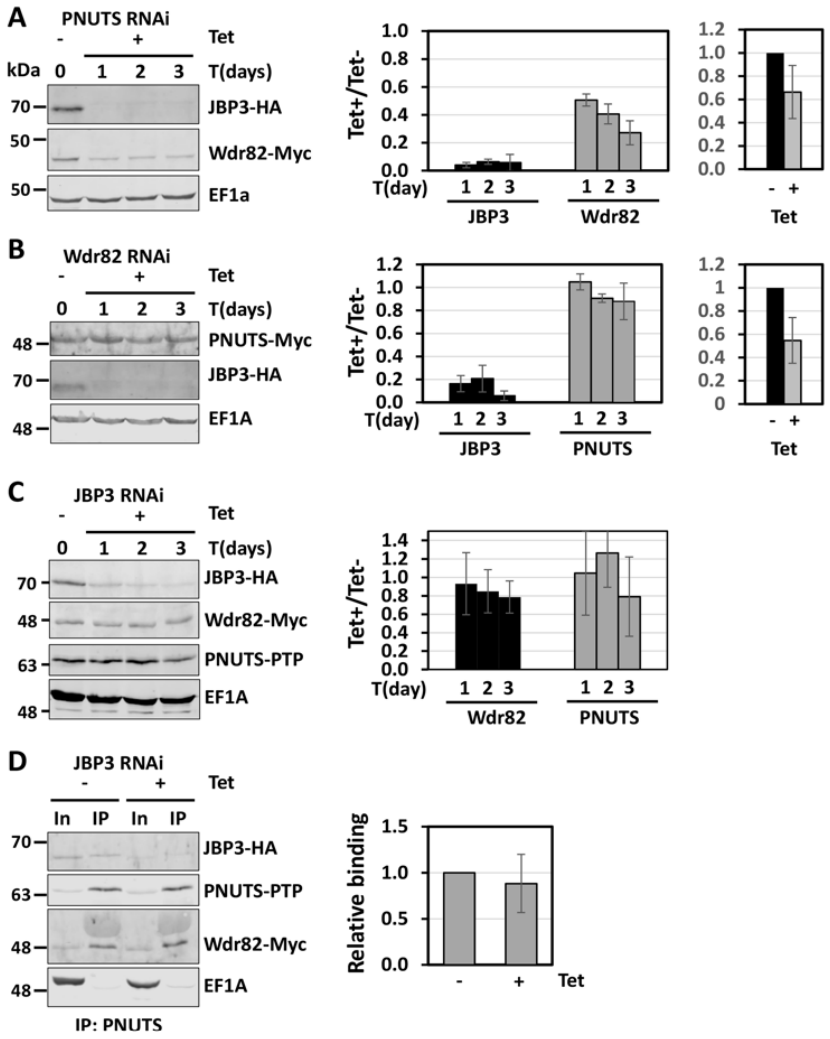


Figure 3.7. TbPNUTS functions as a scaffold factor. RNAi knockdown of the *T. brucei* PJW complex components. Endogenous loci of the indicated genes were tagged with HA, PTP, or Myc tags. Cells were then transfected with the indicated RNAi construct and knockdown of PNUTS (A), Wdr82 (B) or JBP3 (C) was induced by tetracycline (Tet) addition. Cell lysates were collected at the indicated time points and analyzed by western blot with anti-protein A, anti-HA or anti-Myc. Anti-EF1a serves as a loading control. Bands were quantified by densitometry. The middle bar graphs represent the mean ± SD from 3 independent experiments for the indicated protein level relative to protein level prior to the induction of RNAi. The bar graphs on the right show depletion of transcripts upon Tet induction of RNAi by qRT-PCR

analysis, and represent the mean ± SD from 3 independent experiments, with Tet- set to 1. D, effect of JBP3 knockdown on PNUTS-Wdr82 binding. JBP3 RNAi was induced for 48 hrs, and PNUTS-PTP was purified from cell extracts by anti-protein A affinity resin and analyzed by western blot. The %IP of Wdr82 by PNUTS IP with or without JBP3 RNAi induction was determined as described in Fig. 3.3. The bar graph on the right represents the mean ± SD from three independent experiments, with the %IP from the uninduced cells set to 1.

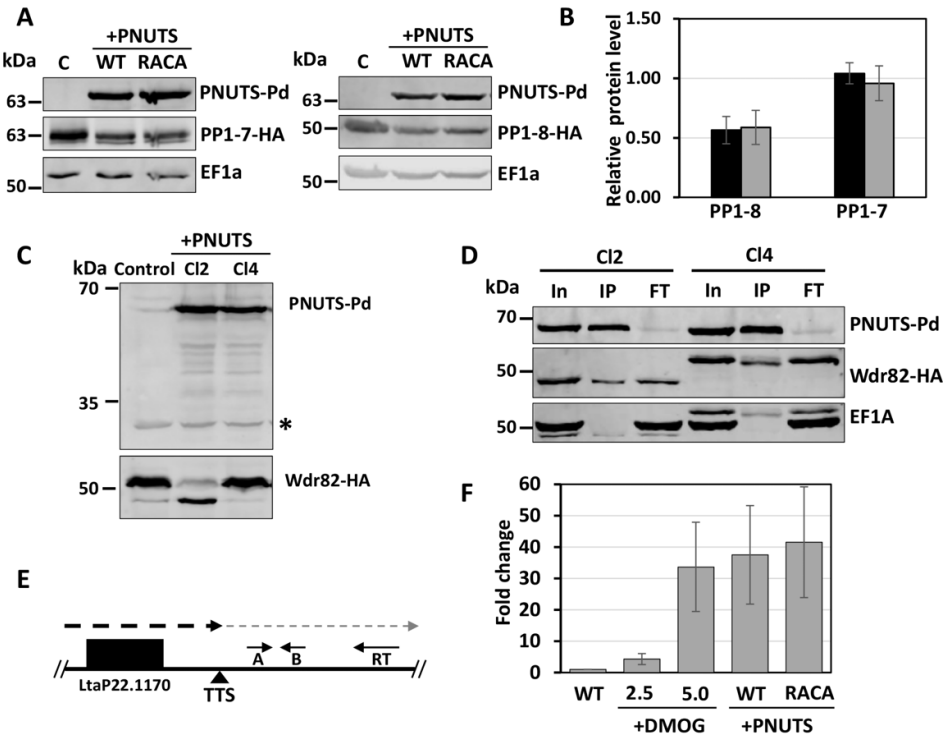


Figure 3.8. LtPNUTS overexpression alters PP1 and Wdr82 stability and transcription termination. A-D, effect of PNUTS overexpression on PP1 and Wdr82. A, PP1-1 or PP1-8 was tagged with HA-tag at its endogenous loci and either WT or RACA mutant PNUTS protein was exogenously overexpressed. Cell lysates were analyzed by western blot with anti-protein, anti-HA and anti-EF1a. Anti-EF1A serves as a loading control. PP1 tagged control cell lines not transfected with the PNUTS expression plasmid are indicated by the C for control. B, HA-tagged PP1-1 and PP1-8 band intensities were quantified by densitometry. The bar graph represents the mean ± SD of PP1-1 or PP1-8 protein level relative to control cells with no overexpression of PNUTS (WT, black bar; RACA mutant, grey bar). C, Wdr82 was tagged at its endogenous loci with HA tag with or without WT PNUTS overexpression and cell lines were analyzed by western blot with anti-protein A, and anti-HA. A non-specific product recognized by the anti-protein A antibody is indicated by an asterisk and serves as a loading control. Shown here are results from two clones (Cl2 and Cl4). See Fig. 3.S11B for results from multiple clones. D, cell

extracts from the cell lines in C were purified by anti-protein A affinity resin and analyzed by western blot with anti-protein A, anti-HA and anti-EF1a. E and F, effect of PNUTS overexpression on Pol II transcription termination. E, diagram of the termination site at the end of a polycistronic gene array on chromosome 22 illustrating the strand-specific RT-qPCR analysis of readthrough defects. The dashed line indicates the readthrough transcripts past the transcription termination site (TTS) that accumulate following a defect in Pol II termination. The location of primers for RT (RT) and qPCR (A and B) are indicated by the small arrows. F, RT-PCR analysis for the readthrough transcripts. Strand-specific cDNAs were synthesized from RNAs extracted from cells treated with the indicated concentrations of DMOG, or from cells with either WT or RACA mutant PNUTS overexpression using primer RT. Fold change of the readthrough transcripts relative to the WT ± SD is based on qPCR analysis with primer A and B, normalized to tubulin RNA.

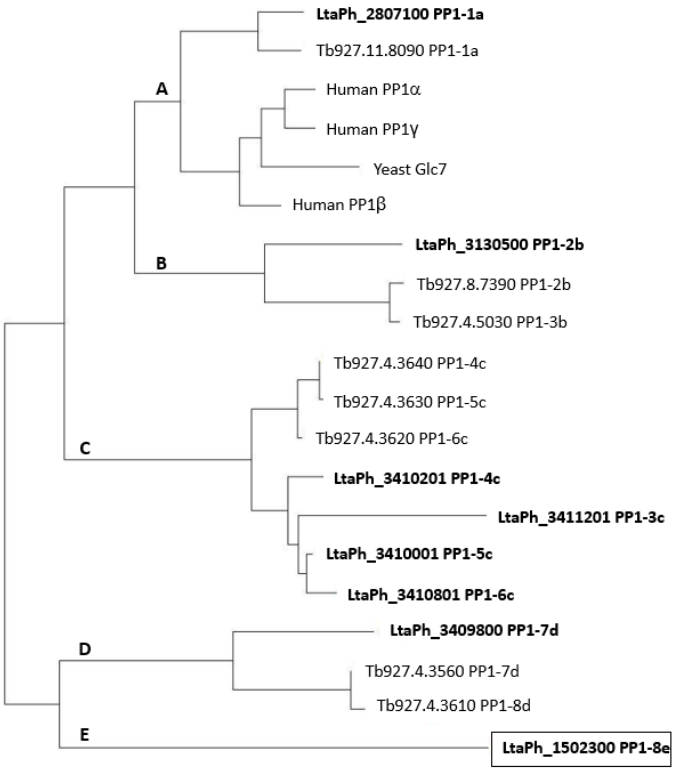
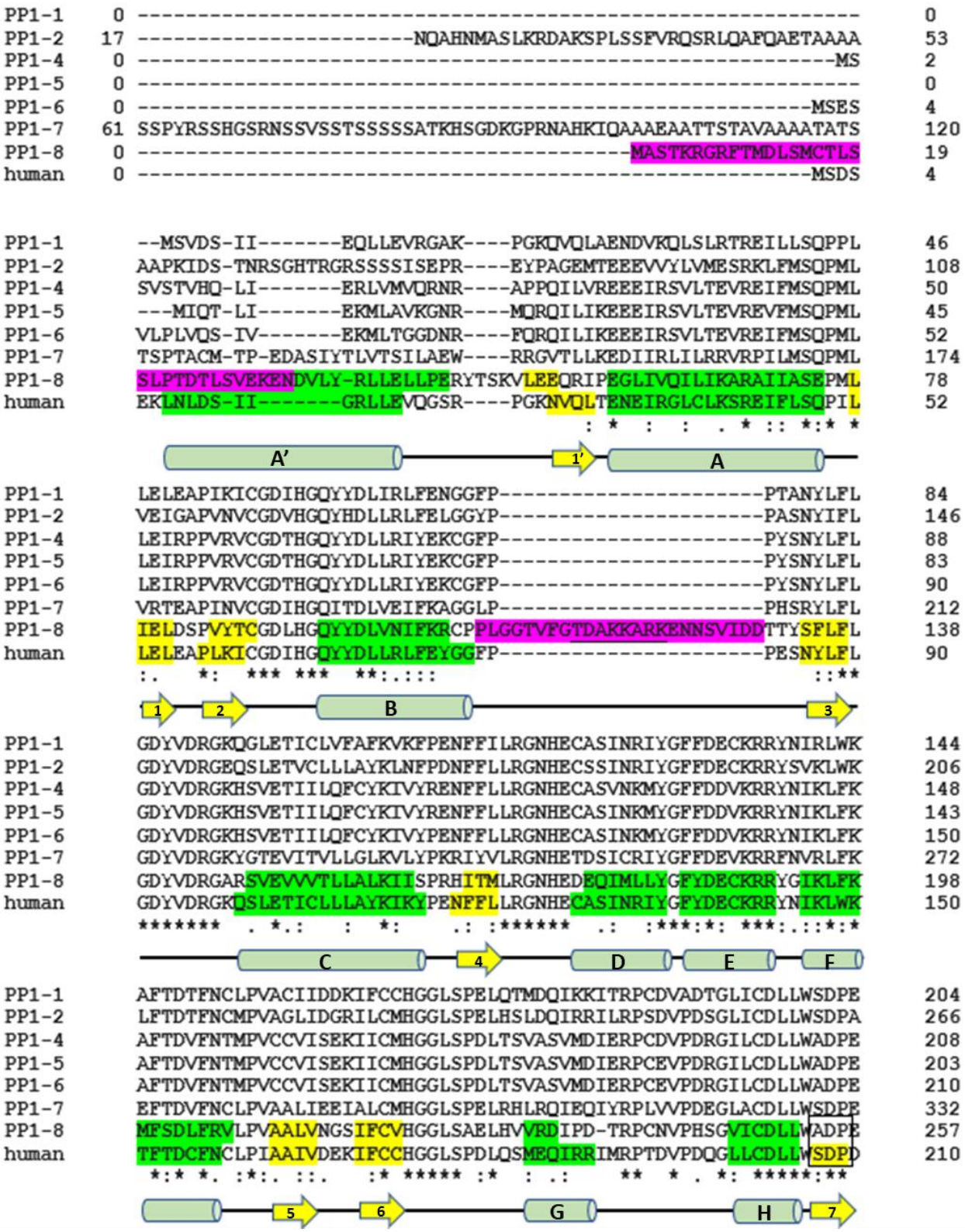


Figure 3.S1. Phylogenetic analysis of PP1 isotypes in *L. tarentolae* (bold), *T. brucei*, human (PP1α, PP1β and PP1γ) and yeast (Glc7). The tree was obtained with Maximum Likelihood method and JTT matrix-based model, using MEGA11 software. The five clades of PP1 (A-E) are shown. Genes for proteins used in the alignment for *L.* tarentolae and *T. brucei* are indicated. Glc7, *Saccharomyces cerevisiae* (SGD:S000000935).


PP1-1 0 0
PP1-2 17 NQAHNMASLKRDAKSPLSSFVRQSRLQAFQAETAAAA 53
PP1-4 0 MS 2
PP1-5 0 0
PP1-6 0 MSES 4
PP1-7 61 SSPYRSSHGSRNSSVSSTSSSSSATKHSGDKGPRNAHKIQAAAEAATTSTAVAAAATATS 120
PP1-8 0 MASTKRGRFTMDLSMCTLS 19
human 0 MSDS 4
PP1-1 --MSVDS-II-------EQLLEVRGAK----PGKQVQLAENDVKQLSLRTREILLSQPPL 46
PP1-2 AAPKIDS-TNRSGHTRGRSSSSISEPR----EYPAGEMTEEEVVYLVMESRKLFMSQPML 108
PP1-4 SVSTVHQ-LI-------ERLVMVQRNR----APPQILVREEEIRSVLTEVREIFMSQPML 50
PP1-5 ---MIQT-LI-------EKMLAVKGNR----MQRQILIKEEEIRSVLTEVREVFMSQPML 45
PP1-6 VLPLVQS-IV-------EKMLTGGDNR----FQRQILIKEEEIRSVLTEVREIFMSQPML 52
PP1-7 TSPTACM-TP-EDASIYTLVTSILAEW----RRGVTLLKEDIIRLILRRVRPILMSQPML 174
PP1-8 SLPTDTLSVEKENDVLY-RLLELLPERYTSKVLEEQRIPEGLIVQILIKARAIIASEPML 78
human EKLNLDS-II-------GRLLEVQGSR----PGKNVQLTENEIRGLCLKSREIFLSQPIL 52
A'
1'
A
PP1-1 LELEAPIKICGDIHGQYYDLIRLFENGGFP----------PTANYLFL 84
PP1-2 VEIGAPVNVCGDVHGQYHDLLRLFELGGYP----------PASNYIFL 146
PP1-4 LEIRPPVRVCGDTHGQYYDLLRIYEKCGFP----------PYSNYLFL 88
PP1-5 LEIRPPVRVCGDTHGQYYDLLRIYEKCGFP----------PYSNYLFL 83
PP1-6 LEIRPPVRVCGDTHGQYYDLLRIYEKCGFP----------PYSNYLFL 90
PP1-7 VRTEAPINVCGDIHGQITDLVEIFKAGGLP----------PHSRYLFL 212
PP1-8 IELDSPVYTCGDLHGQYYDLVNIFKRCPPLGGTVFGTDAKKARKENNSVIDDTTYSFLFL 138
human LELEAPLKICGDIHGQYYDLLRLFEYGGFP----------PESNYLFL 90
1
2
B
3
PP1-1 GDYVDRGKQGLETICLVFAFKVKFPENFFILRGNHECASINRIYGFFDECKRRYNIRLWK 144
PP1-2 GDYVDRGEQSLETVCLLLAYKLNFPDNFFLLRGNHECSSINRIYGFFDECKRRYSVKLWK 206
PP1-4 GDYVDRGKHSVETIILQFCYKIVYRENFFLLRGNHECASVNKMYGFFDDVKRRYNIKLFK 148
PP1-5 GDYVDRGKHSVETIILQFCYKIVYRENFFLLRGNHECASINKMYGFFDDVKRRYNIKLFK 143
PP1-6 GDYVDRGKHSVETIILQFCYKIVYPENFFLLRGNHECASINKMYGFFDDVKRRYNIKLFK 150
PP1-7 GDYVDRGKYGTEVITVLLGLKVLYPKRIYVLRGNHETDSICRIYGFFDEVKRRFNVRLFK 272
PP1-8 GDYVDRGARSVEVVVTLLALKIISPRHITMLRGNHEDEQIMLLYGFYDECKRRYGIKLFK 198
human GDYVDRGKQSLETICLLLAYKIKYPENFFLLRGNHECASINRIYGFYDECKRRYNIKLWK 150
C
4
D
E
F
PP1-1 AFTDTFNCLPVACIIDDKIFCCHGGLSPELQTMDQIKKITRPCDVADTGLICDLLWSDPE 204
PP1-2 LFTDTFNCMPVAGLIDGRILCMHGGLSPELHSLDQIRRILRPSDVPDSGLICDLLWSDPA 266
PP1-4 AFTDVFNTMPVCCVISEKIICMHGGLSPDLTSVASVMDIERPCDVPDRGILCDLLWADPE 208
PP1-5 AFTDVFNTMPVCCVISEKIICMHGGLSPDLTSVASVMDIERPCEVPDRGILCDLLWADPE 203
PP1-6 AFTDVFNTMPVCCVISEKIICMHGGLSPDLTSVASVMDIERPCEVPDRGILCDLLWADPE 210
PP1-7 EFTDVFNCLPVAALIEEIALCMHGGLSPELRHLRQIEQIYRPLVVPDEGLACDLLWSDPE 332
PP1-8 MFSDLFRVLPVAALVNGSIFCVHGGLSAELHVVRDIPD-TRPCNVPHSGVICDLLWADPE 257
human TFTDCFNCLPIAAIVDEKIFCCHGGLSPDLQSMEQIRRIMRPTDVPDQGLLCDLLWSDPD 210
5
6
G
H
7

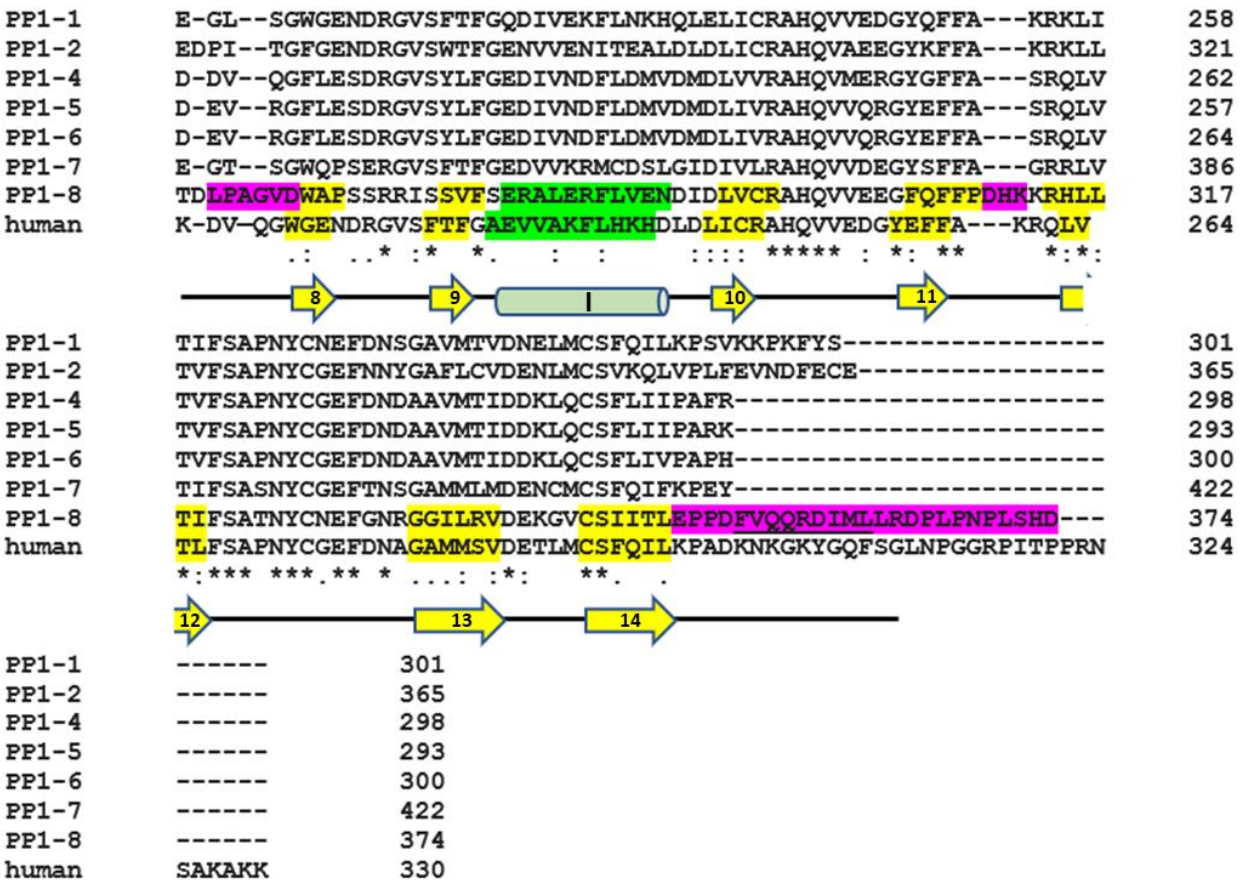


Figure 3.S2. Sequence alignment of *L. tarentolae* PP1 isoforms and human PP1a. Identical (*), conservative (:) and similar (.) residues between hPP1 and LtPP1-8 are indicated. Cylinders and arrows representing α-helices and β-strands of the hPP1a catalytic domain, respectively, are drawn under the sequence. The sequences corresponding to α-helices and β-strands are highlighted in the hPP1a sequence in green and yellow, respectively. The predicted secondary structure elements of Lt PP1-8e based on AlphaFold are also highlighted in green and yellow. The box at b-sheet 7 indicates that alanine substitution in the four amino acid element in LtPP1-8e leads to the inability of AlphaFold to predict a conserved b-sheet 7. But all the other a-helices and b-sheets within the catalytic domain are predicted. The distinct N- and C-termini and insertions within the catalytic domain of LtPP1-8e are highlighted in purple. Predicted α-helices within the 26-aa insert and C-terminus of LtPP1-8e are indicated by the underlined residues.

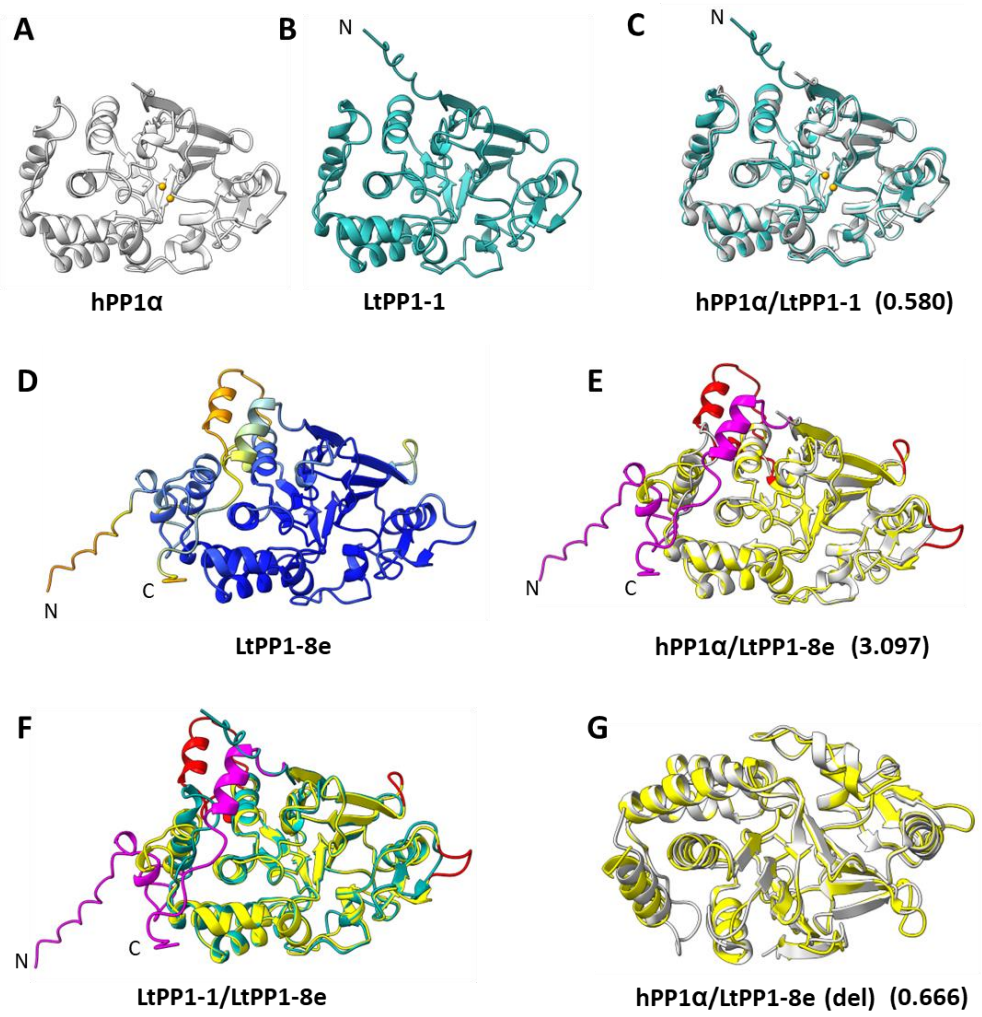


Figure 3.S3. Structural homology of *L. tarentolae* PP1-8e and Human PP1α catalytic domains. A, Human PP1α (7-300) structure (3E7A) shown as a carton in grey. Bound metal ions are shown as orange spheres. B, Predicted LtPP1-1a structure illustrated as a carton in blue. C, Structural overlay of hPP1 structure and predicted LtPP1-1a. The root-mean-square-deviation (RMSD), estimating the degree of structural similarity between the model and the crystal data, is indicated. D, The predicted Lt PP1-8e structure colored by pLDDT values, with blue representing high model confidence (pLDDT > 90) and orange representing low model confidence (70 > pLDDT > 50). E, Structural overlay of hPP1 (grey) and LtPP1-8e. LtPP1-8e is in yellow with unique sequences within the catalytic domain and the extremities are highlighted

in red and magenta, respectively. F, Structural overlay of LtPP1-1a and LtPP1-8e models with the PP1-8e unique sequences colored as in E. G, Structural overlay of hPP1 structure (grey) and LtPP1-8e model with the unique sequences within the catalytic motif and extremities deleted (del).

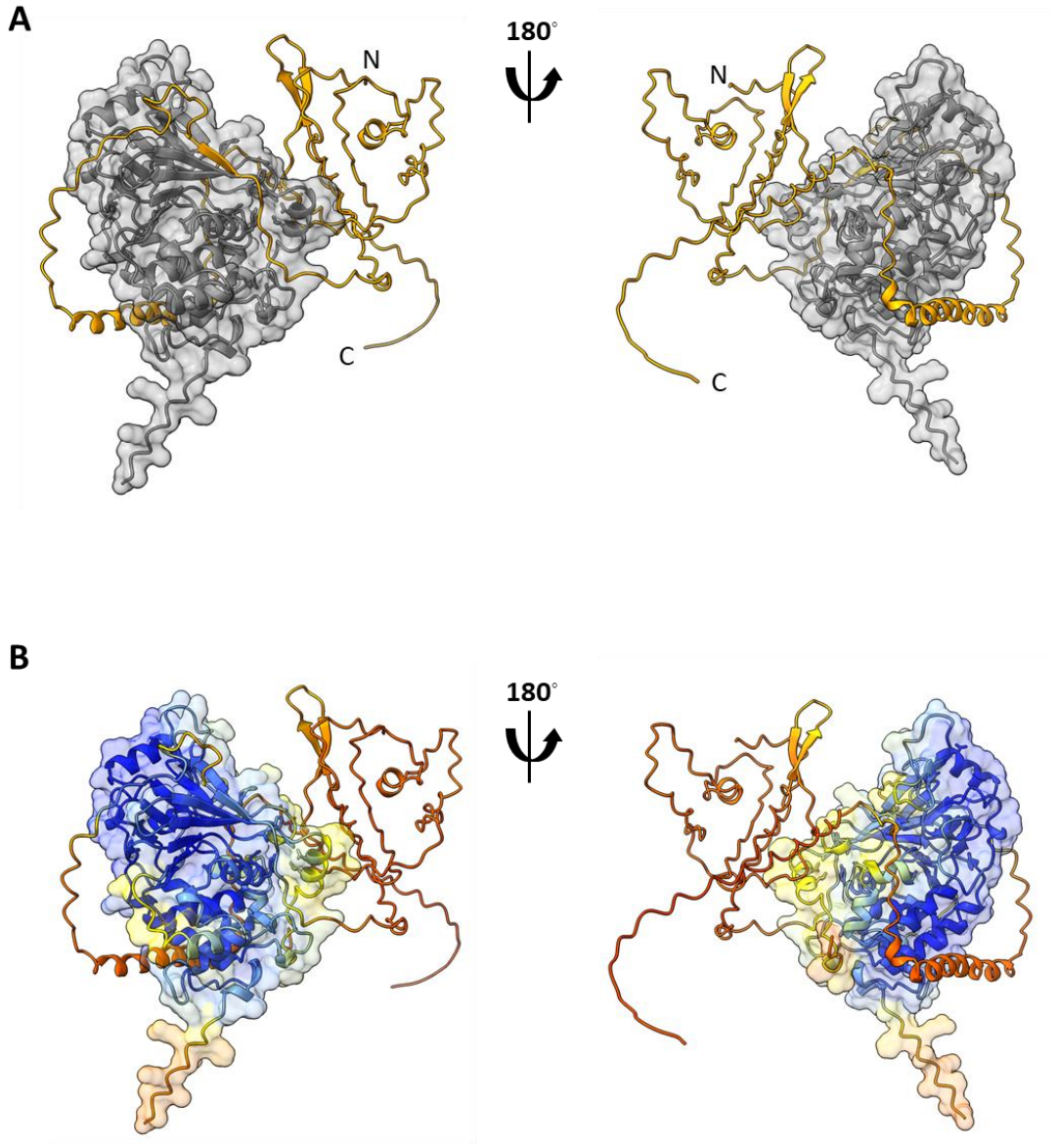
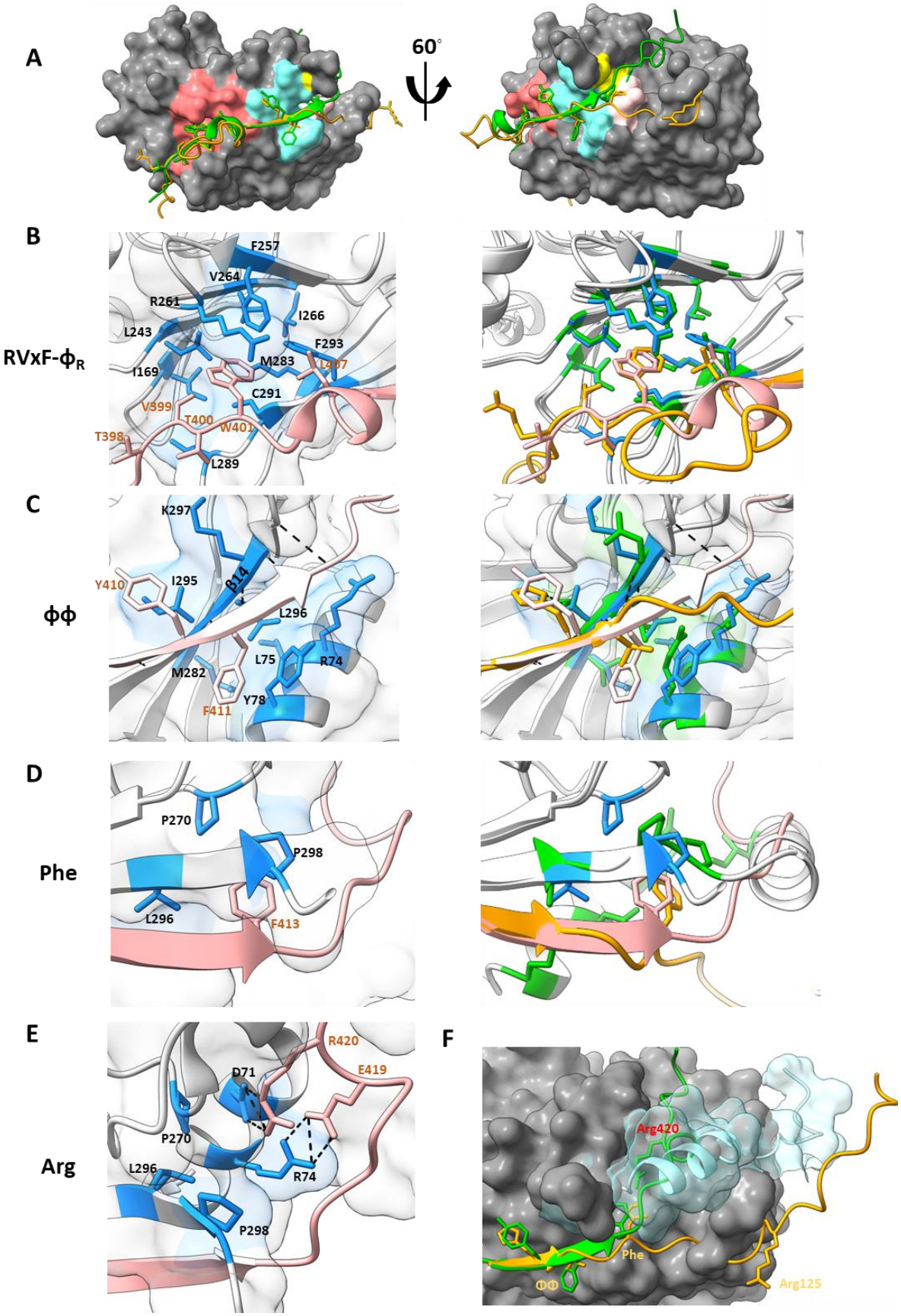


Figure 3.S4. AlphaFold model of the Lt PP1-8e:PNUTS complex. A, The predicted LtPNUTS (orange cartoon) and LtPP1-8e (gray surface) complex. B, The predicted LtPNUTS:LtPP1-8e complex colored by pLDDT values, with blue representing high model confidence (pLDDT > 90) and orange representing low model confidence (70 > pLDDT > 50).


A
60˚
B
RVxF-ϕR
F257
V264
R261
I266
L243
F293
M283
I169
L407
C291
V399
T400
W401
T398
L289
C
ϕϕ
K297
Y410
I295
β14
L296
L75
R74
M282
Y78
F411
D
Phe
P270
P298
L296
F413
E
Arg
R420
D71
E419
P270
L296
R74
P298
F
Arg420
Phe
ΦΦ
Arg125

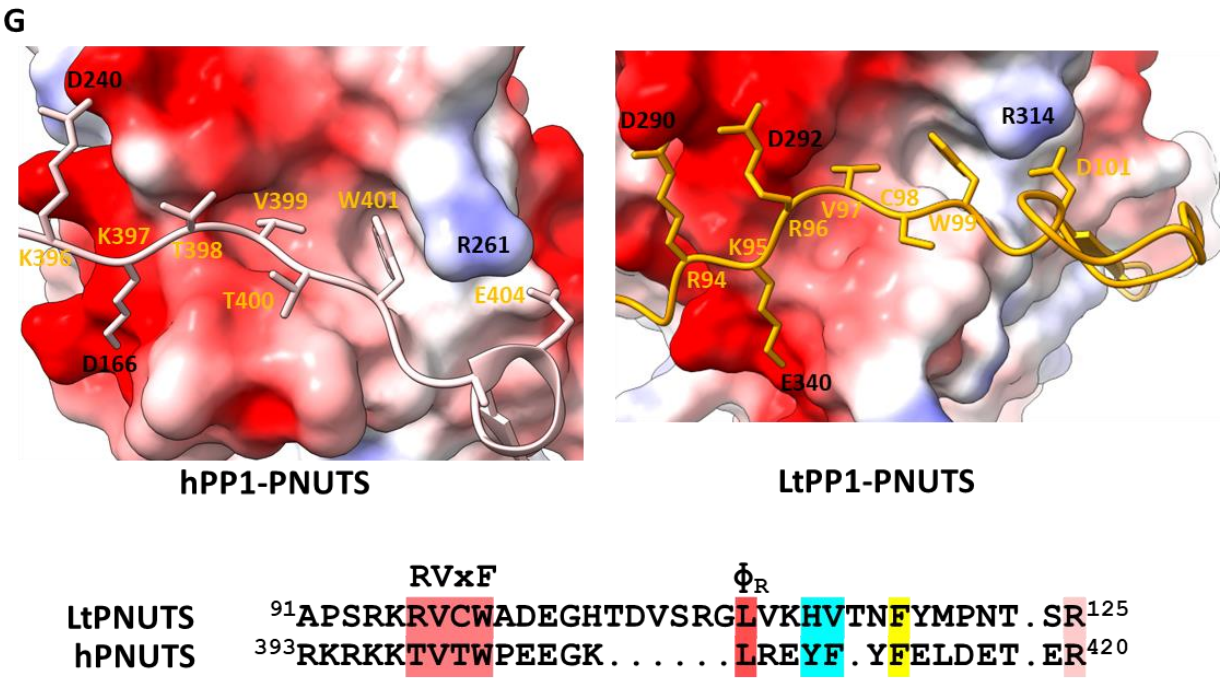


Figure 3.S5. LtPNUTS and hPNUTS share PP1-binding motifs. A, The predicted LtPNUTS:PP1-8e structure is superimposed to the human PP1:PNUTS structure (4mp0). Human PP1 is shown as grey surface. LtPNUTS (orange) and hPNUTS (green) are represented as ribbons with key interacting residues shown as sticks. The RVxF binding pocket (red), ΦΦ binding pocket (cyan), Phe binding pocket (yellow) and Arg binding pocket (pink) are shaded on the hPP1 surface and correspond to the zoomed-in pockets shown in B, C, D and E. Close-up of the RVxF binding pocket (B), ΦΦ binding pocket (C), and Phe (D) and Arg binding pockets (E) on the human PP1:PNUTS holoenzyme. B-D; Left, key interacting residues in hPP1 (blue) and hPNUTS (pink) are shown as sticks and labelled. Right, overlay of the Lt and human PP1-PNUTS structures, with key interactive residues between LtPP1-8e (green) and LtPNUTS (orange) shown as sticks. E, Arg binding pocket on the human PP1:PNUTS holoenzyme. Salt bridge interactions between PNUTS R420 and E419 with the corresponding residues of the Arg binding pocket in PP1 are indicated by the dashed line. F, structural overlay of the human PP1-

PNUTS structure and Lt PP1-PNUTS model. hPNUTS (green) is bound to hPP1 (grey surface) with Arg420 shown as stick. Arg125 in LtPNUTS is not predicted to bind into the Arg binding pocket. The Arg binding pocket in the human structure is occupied by the C-terminus of PP1-8e (shown as blue ribbon and surface). G and H, electrostatic surface potential representation (positive, blue; negative, red) of the RVxF-Or pocket in the hPNUTS:PP1 complex (G) and LtPNUTS:PP1-8e complex (H). The sequence alignments for these regions of PNUTS are provided below.

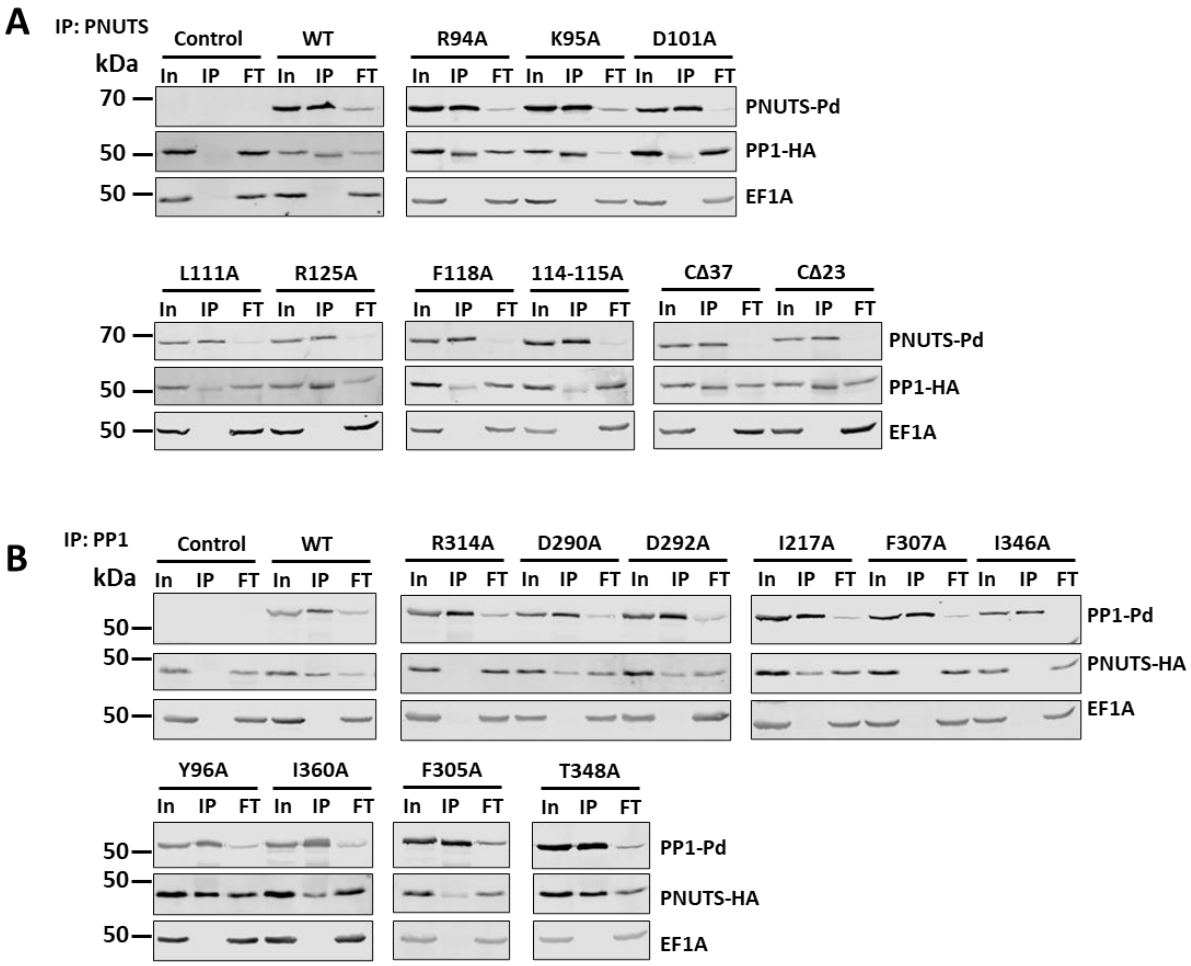
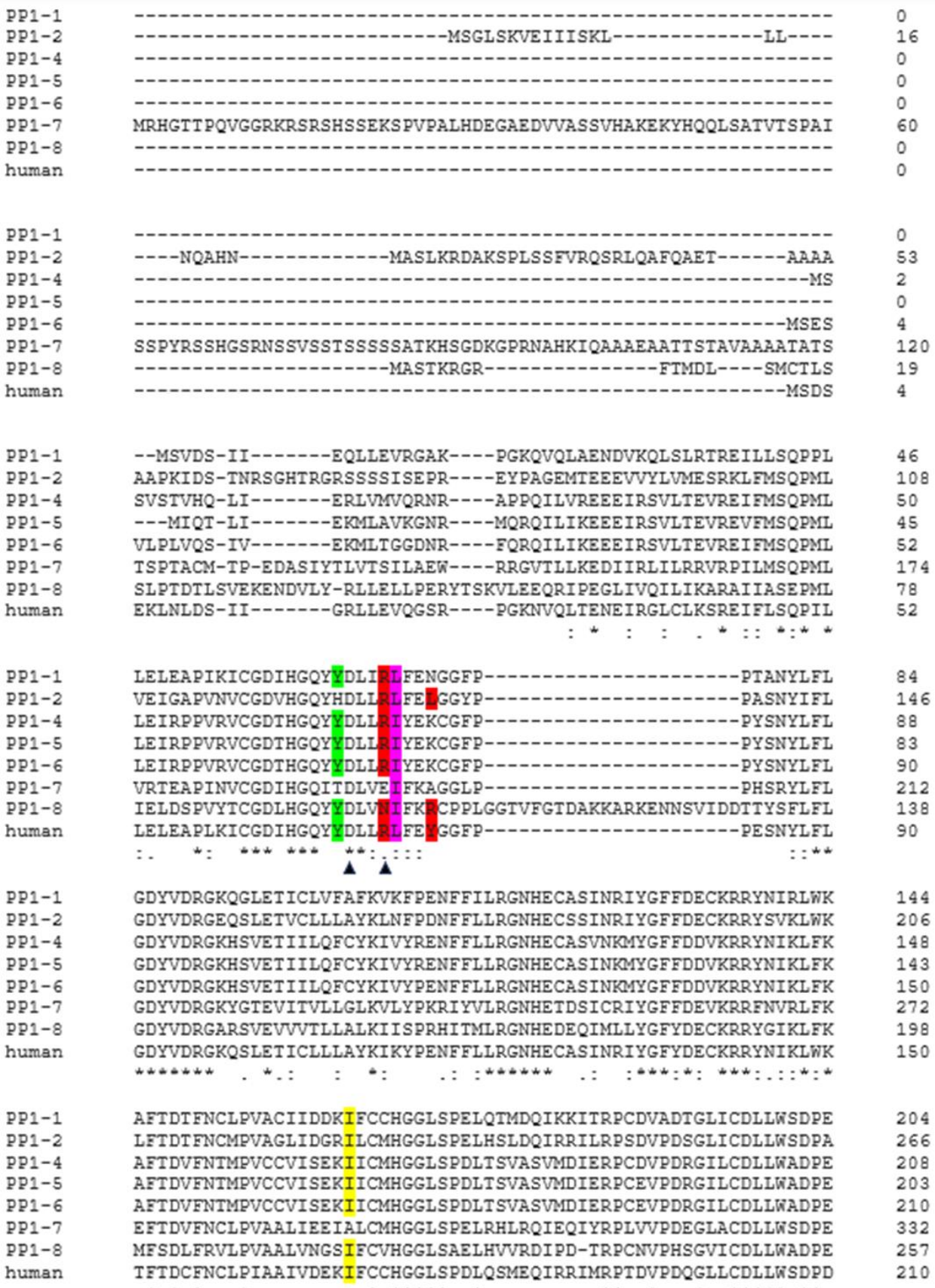


Figure 3.S6. Co-immunoprecipitation analysis of LtPP1-PNUTS. Western blots showing the Co-IP of Pd-tagged wild type or indicated PNUTS (A) or PP1 (B) mutants with HA-tagged PP1 or PNUTS, respectively, as described in Figure 3.3. A representative gel image from three independent experiments for each Co-IP is shown. Control and WT Co-IP gel blots are replicated here from Figure 3.3A.


PP1-1 0
PP1-2 16
PP1-4 0
PP1-5 0
PP1-6 0
PP1-7 60
PP1-8 0
human 0
PP1-1 0
PP1-2 53
PP1-4 2
PP1-5 0
PP1-6 4
PP1-7 120
PP1-8 19
human 4
PP1-1 46
PP1-2 108
PP1-4 50
PP1-5 45
PP1-6 52
PP1-7 174
PP1-8 78
human 52
PP1-1 84
PP1-2 146
PP1-4 88
PP1-5 83
PP1-6 90
PP1-7 212
PP1-8 138
human 90
PP1-1 144
PP1-2 206
PP1-4 148
PP1-5 143
PP1-6 150
PP1-7 272
PP1-8 198
human 150
PP1-1 204
PP1-2 266
PP1-4 208
PP1-5 203
PP1-6 210
PP1-7 332
PP1-8 257
human 210

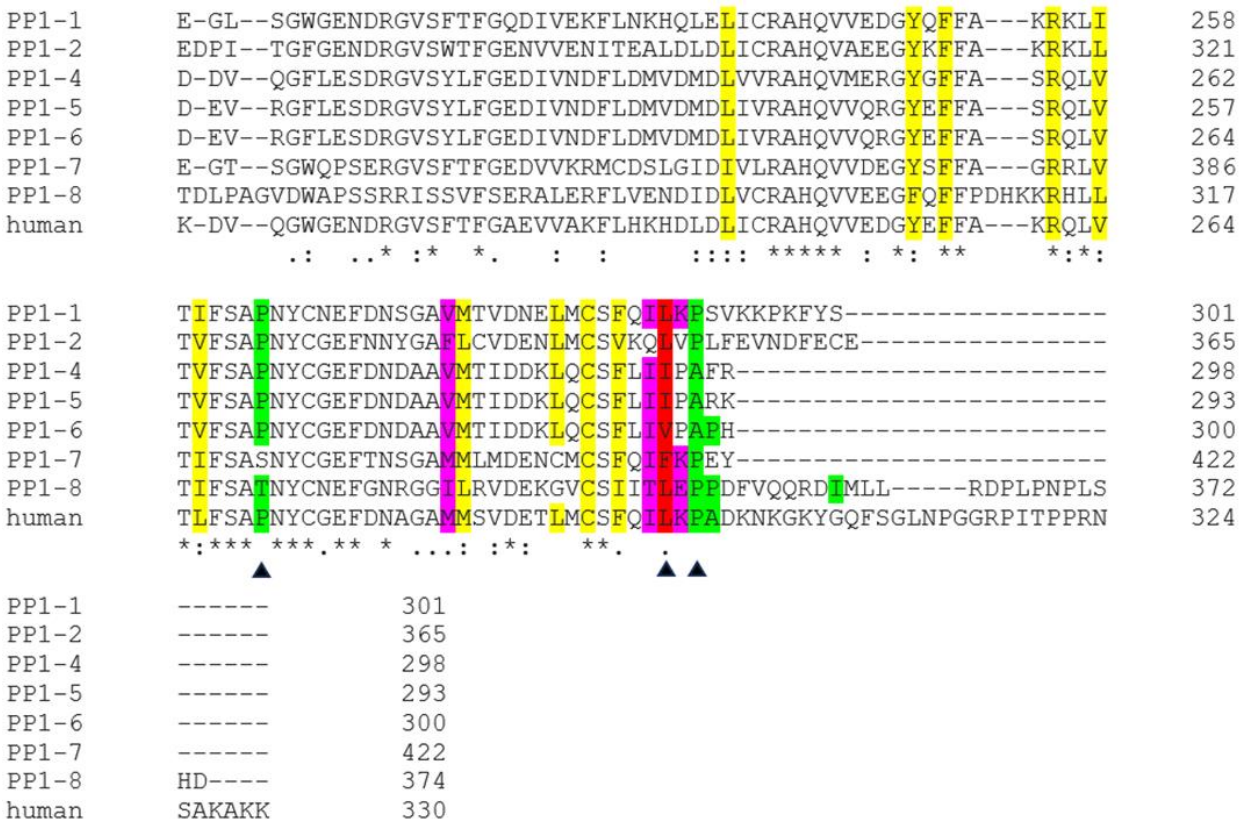


Figure 3.S7. Sequence alignment of *L. tarentolae* PP1 isoforms and human PP1a. Identical (*), conservative (:) and similar (.) residues are indicated. Residues that form the extended RVxF binding pocket (yellow), ϕϕ binding pocket (purple), and Phe binding pocket (green) in the hPNUTS:hPP1 structure and predicted LtPNUTS:PP1-8e structure are highlighted. Residue that is a component of both the ϕϕ binding pocket and Phe binding pocket is highlighted in red. Residues that form the Arg binding pocket in hPP1 are indicated by a triangle underneath. Conserved residues among the remaining isoforms are also highlighted. Alignment was generated with ClustalW.

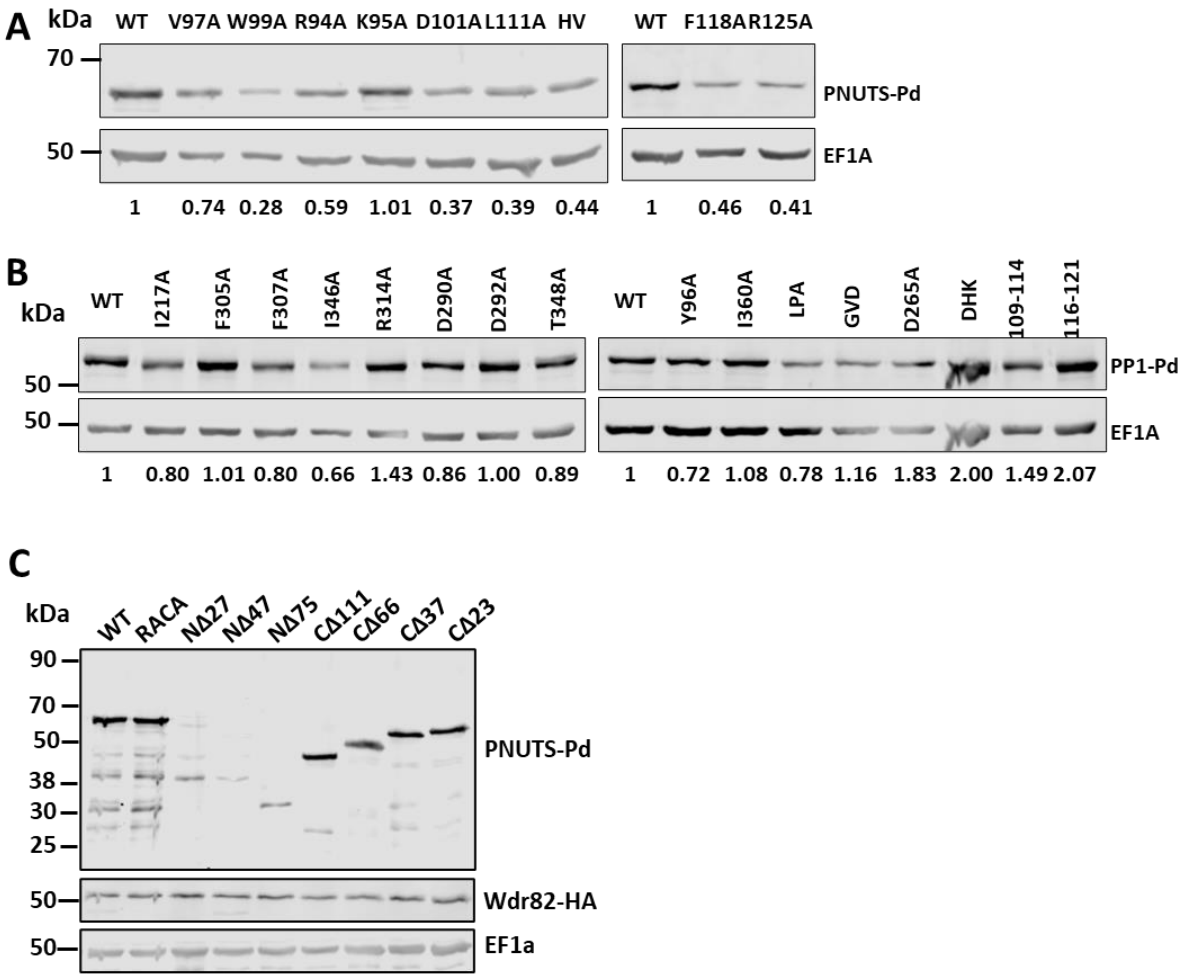


Figure 3.S8. Expression levels of PNUTS and PP1. Lysates from cells that over-express the indicated PNUTS (A and C) or PP1 (B) mutant were analyzed by western blot with anti-protein A and anti-EF1A. EF1A serves as a loading control. Bands were quantified by densitometry and protein expression for the indicated mutant were normalized to EF1A. The normalized values are shown below the gel with wild type set to 1.

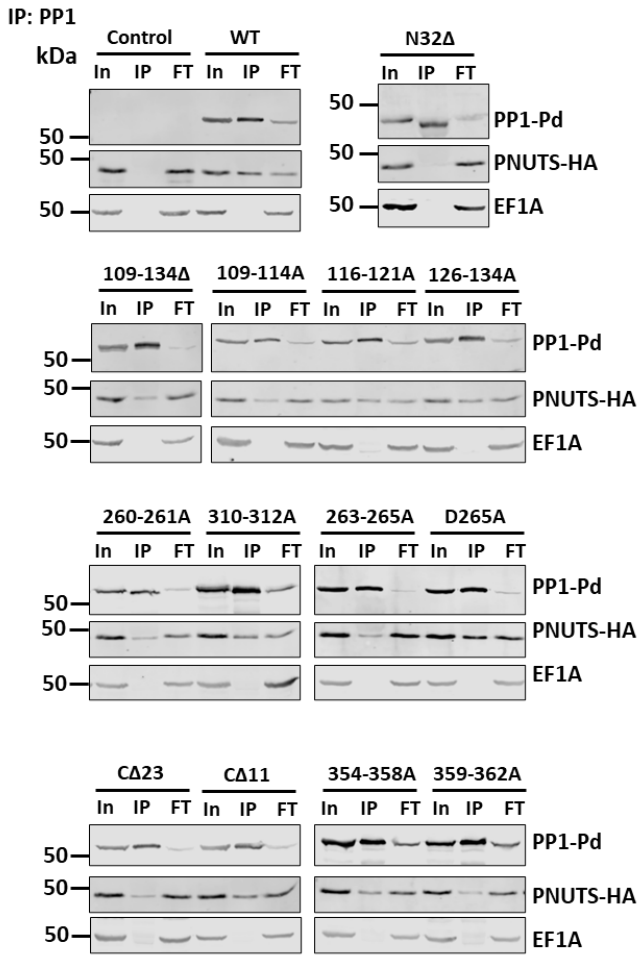


Figure 3.S9. Coimmunoprecipitation analysis of LtPP1-PNUTS. Western blots showing the Co-IP of Pd-tagged PP1-8e (WT and indicated variants) with PNUTS-HA as described in Figure 3.4. A representative gel image from three independent experiments for each Co-IP is shown.

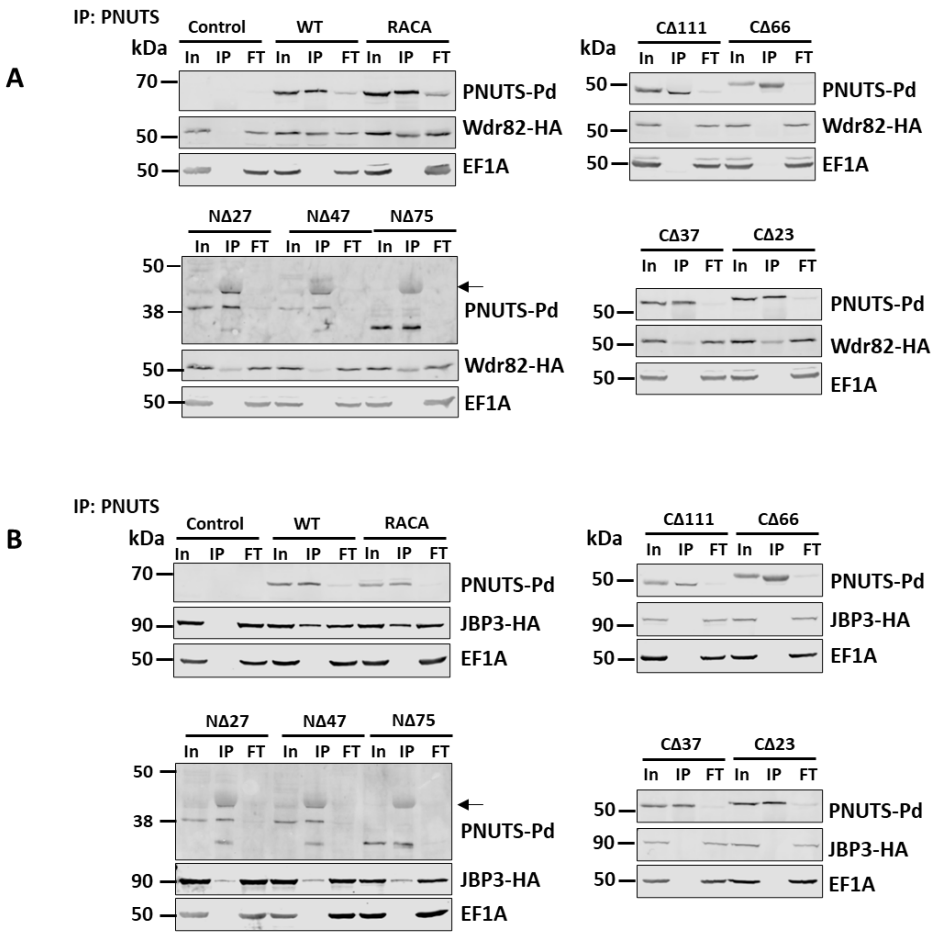


Figure 3.S10. Coimmunoprecipitation analysis of LPNUTS and Wdr82 and JBP3. Pd-tagged PNUTS (WT or truncations) were expressed in Lt cells and tested for interaction with HA-tagged Wdr82 (A) or JBP3 (B) by Co-IP analysis as described in Figure 3.5. A representative Western blot image from three independent experiments for each Co-IP is shown. Arrow indicates IgG contamination in the IP of N-terminal PNUTS mutants.

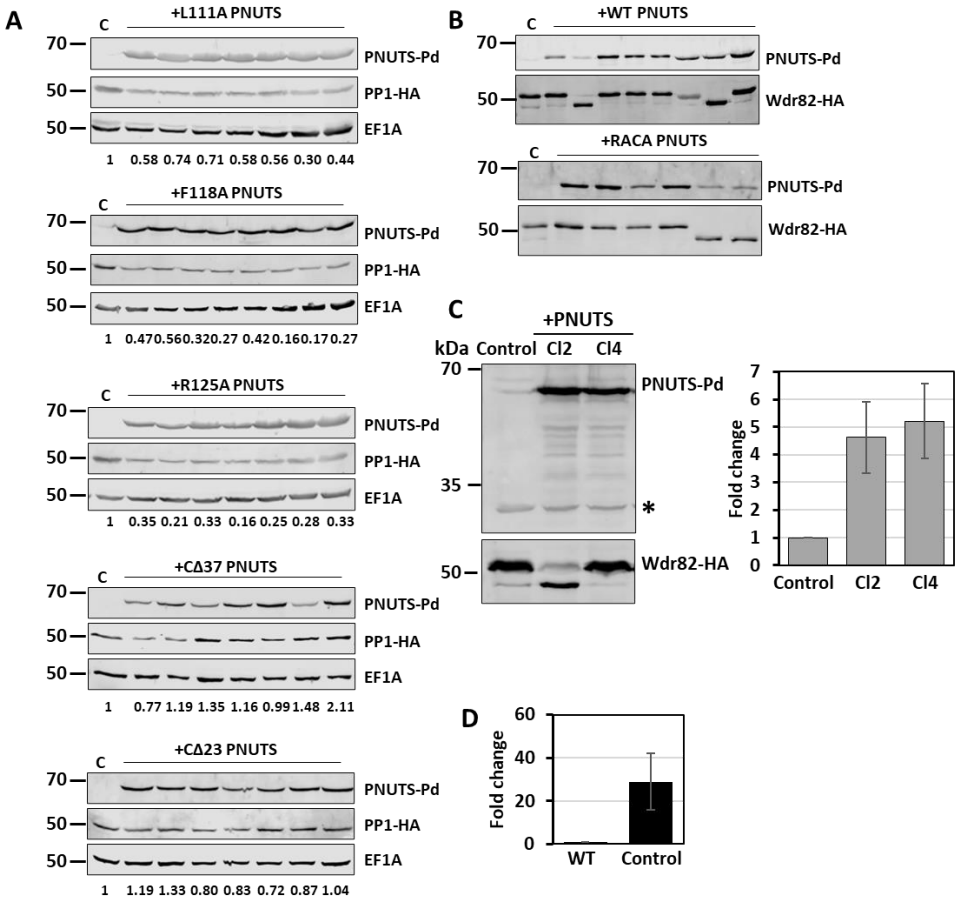


Figure 3.S11. Effects of PNUTS over-expression on PP1 and Wdr. A, Effect of PNUTS over-expression on PP1-8e levels. LtPP1-8e was endogenously tagged with the HA tag and lysates from clonal cell lines following transfection of the indicated PNUTS construct were analyzed by western blot with anti-protein A and anti-HA. Anti-EF1A serves as a loading control. Bands were quantified by densitometry and PP1 levels were normalized to EF1A. The normalized values are shown below the gel with the untransfected PNUTS control cell set to 1. Control cell line lacking the PNUTS expression plasmid is indicated by the C. B, effect of PNUTS over-expression on Wdr82. Wdr82 was endogenously tagged with HA and cells transfected with the indicated PNUTS expression plasmid. Lysates from clonal cell lines were analyzed by western blot with anti-protein A and anti-HA. C, transcription readthrough (Fold change) for the indicated Wdr82-HA tagged clonal cell lines was analyzed by RT-qPCR as in

Figure 3.8F. Gel blots are replicated here from Figure 3.8C. Control, Wdr82-HA tagged cell line untransfected with the PNUTS expression plasmid. D, transcription readthrough defects in WT versus Wdr82-HA tagged (Control) cell lines.

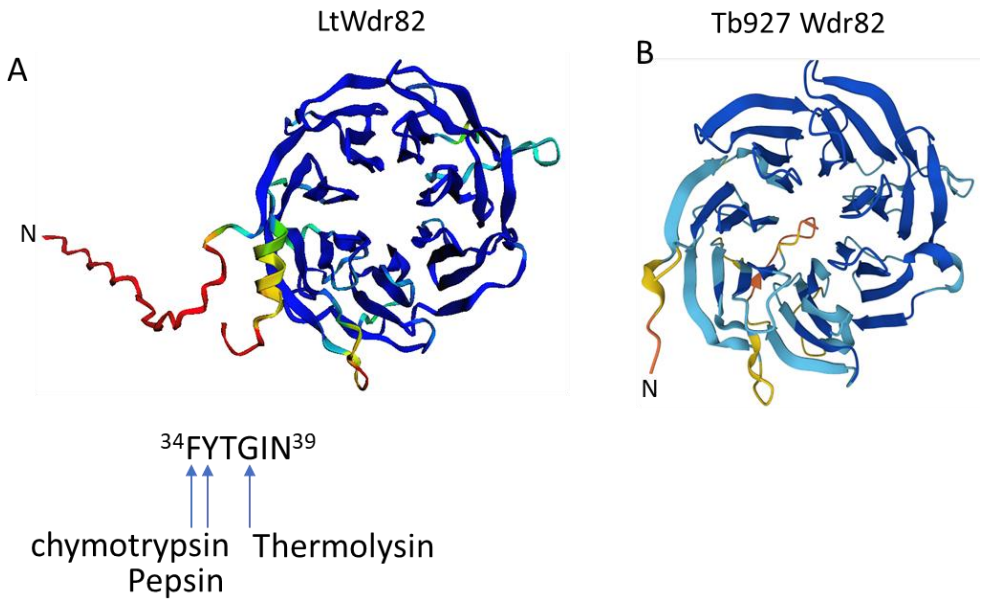


Figure 3.S12. AlphaFold model of LtWdr82 has an unstructured N-terminus. AlphaFold model of (A) LtWdr82, and (B) TbWdr82. The structures are colored by pLDDT values, with blue representing high model confidence (pLDDT > 90) and orange representing low model confidence (70 > pLDDT > 50). Potential protease cleavage sites in the N-terminus of LtWdr82 are indicated. N, N-terminus.

CHAPTER 4

CONCLUSION AND DISCUSSION

Because of the polycistronic genetic organization in trypanosomatids, post-transcriptional mechanisms have been regarded as the primary means of regulating gene expression (1, 2). Significant progress has been made in recent years to reveal the role of epigenetic modifications in regulating transcription, especially base J, H3V and H4V. These epigenetic marks independently regulate transcription termination in *T. brucei* or *Leishmania* spp, and promote transcription termination within a PTU prior to the end of some gene clusters (3–7). The precise mechanism by which these epigenetic marks regulate transcription has been elusive, but recent characterization of the trypanosomatid PJW/PP1 complex provides insight into the underlying mechanism. Individual depletion of PNUTS, Wdr82 or JBP3 by RNAi from *T. brucei* cells led to readthrough transcription at cSSRs and de-repression of the genes located at the end of some PTUs, accumulated antisense transcripts from bidirectional promoters, and upregulation of subtelomeric and telomeric VSG genes (Appendix A). Later, inducible PP1 KO in *L. major* also leads to extended readthrough transcription and de-repression of genes at the end of some PTUs (Chapter 2). These transcription termination defects induced by the loss of the PJW/PP1 complex are highly similar to what we observed after base J ablation. The PJW/PP1 complex includes a subunit that shows *in vitro* base J binding ability, JBP3. Therefore, it is proposed that base J regulates transcription termination by recruiting the PJW/PP1 complex via JBP3-tethering. The PJW/PP1 complex is reminiscent of the mammalian PTW/PP1 complex that regulates transcription termination by dephosphorylation of RNAPII CTD and Spt5, therefore allowing

recruitment of transcription termination factor and/or slowing down the RNAPII, facilitating transcription termination by torpedo model. Consistent with the speculation that PP1 is the key enzymatic subunit, the PJW/PP1 complex with the PP1 subunit, instead of the complex lacking PP1, can dephosphorylate RNAPII CTD *in vitro*, suggesting the trypanosomatid PJW/PP1 complex regulates transcription termination using a conserved mechanism as the mammalian PTW/PP1 complex (Chapter 2). Human PNUTS is a PP1-interacting protein with a central RVxF motif, and serves as a scaffolding protein for N- and C- terminal association of Tox4 and Wdr82 (8). *L. tarentolae* PNUTS (LtPNUTS) is also an RVxF motif-containing protein, with no other recognizable domains or motifs, and whether it similarly serves as a scaffolding protein in the PJW/PP1 complex was unknown. Structural modeling was employed to identify the other short and degenerate SLiMs in LtPNUTS that are difficult to identify by sequence alignment (9, 10). We show that LtPNUTS is specific for binding to LtPP1-8e using RVxF-ϕϕ-Phe motifs, similar to human PNUTS, and extremities and specific sequences within the catalytic core of PP1-8e confer isotype specificity to LtPNUTS (Chapter 3). In addition, consistent with its role as a scaffolding protein, LtPNUTS depletion reduces the protein levels of both JBP3 and Wdr82. However, truncation analysis of LtPNUST indicates that JBP3 and Wdr82 bind to similar regions in LtPNUST, suggesting that JBP3 and Wdr82 do not bind independently to PNUTS.

Based on these results, we propose that PNUTS is a scaffolding protein in the PJW/PP1 complex which is recruited to base J-enriched regions to regulate transcription termination by PP1-mediated RNAPII CTD dephosphorylation. However, several critical questions still remain and will be discussed below.

IS THE PJW/PP1 COMPLEX OPERATIONAL THROUGHOUT *T. BRUCEI* LIFE STAGES?

Base J synthesis is developmentally regulated during the *T. brucei* life cycle. Base J is absent in the procyclic life stage of *T. brucei* cells due to down-regulation of JBP1 and JBP2 in the procyclic stage (11, 12). Therefore, it raises the question of whether the PJW/PP1 complex is involved in regulating transcription termination in procyclic *T. brucei*, and if it does, how the complex is recruited to specific genomic loci in the absence of base J.

JBP3 shows an *in vitro* base J-binding ability by gel shift assay (13). However, it remains to be tested whether the J-binding domain is essential for JBP3 to regulate transcription termination, and whether the loss of JBP3 affects the genomic distribution of the PJW/PP1 complex *in vivo*. Furthermore, we need to figure out if the PJW/PP1 complex actually plays a role in regulating transcription termination in the procyclic stage of *T. brucei* cells. This can be tested by RNAi knockdown of the complex components to see if it leads to any transcription termination defects in procyclic *T. brucei* cells.

Base J may not be the only recruiter for the PJW/PP1 complex. In mammals and yeast, Wdr82 homologs are known to play a crucial role in tethering H3K4 methyltransferases to specific genomic loci by interacting with the Ser5P of RNAPII CTD through their WD40 domains. Interestingly, Wdr82 is a part of the PJW/PP1 complex and possesses WD40 repeat domains, which has the potential for recognizing specific protein or histone post-translational modifications (14–17). In *T. brucei*, strand switch regions exhibit an abundance of histone variants and post-translational modifications, including H3V and H4V, which are also essential for transcription termination. Therefore, it is conceivable that JBP3 and Wdr80 may function as dual tethering forces for the PJW/PP1 complex. Genomic loci with distinct chromatin

modification profiles might exhibit varying affinities for the PJW/PP1 complex, ultimately recruiting different levels of the complex.

IS PP1 AN INTEGRAL COMPONENT OF THE COMPLEX?

The roles of Wdr82, JBP3 and PNUTS in transcription termination were extensively tested in *T. brucei* by RNAi (Appendix A). Later, inducible knockout of *L. major* PP1-8 revealed that it also regulates transcription termination (Chapter 2). Moreover, the association of PP1 with the complex is responsible for the phosphatase activity of the complex toward RNAPII CTD (Chapter 3). These results suggest that the trypanosomatid PJW/PP1 complex regulates transcription termination through a conserved mechanism as in yeast and mammalian cells. However, we could not exclude the possibility that the *T. brucei* PJW complex regulates transcription termination using a PP1-independent mechanism, consistent with a lack of a clade E phosphatase in *T. brucei* genome (13).

Characterization of the binary interaction of leishmania PP1-PNUTS (Chapter 3) led us to identify multiple degenerate SLiMs present on PNUTS essential for the PP1-PNUTS interaction. We also found that the N- and C-termini of PP1-8e and specific sequences within PP1-8e are required for interaction with PNUTS. Interestingly, these unique sequences in LtPP1-8e are not found on any other Lt PP1 homologs or Tb PP1 homologs (18), explaining the specificity of PNUTS for LtPP1-8e and the lack of PP1 interaction with PNUTS in *T. brucei*. Interestingly, overexpression of the L. major PP1-1a or PP1-7d could partially rescue the transcription termination defects caused by PP1-8e KO. This led to the speculation that PP1-8e associates stably with the PJW complex due to the presence of binding pockets for PNUTS SLiMs, while in its absence, other PP1 homologs could form a transient or weak association with the PNUTS to

compensate for the loss of PP1-8e. Therefore, it is conceivable that a particular *T. brucei* PP1 homolog forms a weak or transient interaction with the PJW complex, regulating transcription termination by a PP1-dependent mechanism. Furthermore, the absence of a key enzymatic component from a multiprotein complex was also observed for *T. brucei* polyA polymerase, which was not purified with *T. brucei* cleavage and polyadenylation complex (19). In summary, we proposed that the PJW/PP1 complex regulates transcription termination in a PP1-dependent mechanism in both *T. brucei* and *Leishmania* spp.

HOW IS TRANSCRIPTION TERMINATION UNCOUPLED FROM 3' END FORMATION IN TRYPANOSOMATIDS?

According to the torpedo model, 3' end formation is coupled with transcription termination, and a lot of experimental evidence has supported this idea (20–23). However, the torpedo model poses significant challenges to polycistronic transcription in trypanosomatids, since a coupled 3' end formation and transcription termination would lead to premature transcription termination within a PTU, leaving hundreds of downstream genes un-transcribed. However, several interesting features in trypanosomatids potentially uncouple these two processes, and ensure transcription termination at specific genomic loci.

First, mRNAs are generated from the long polycistronic transcripts by trans-splicing, adding a hypermodified cap to the 5' end of the nascent RNAs still being synthesized by RNAPII (24). The 5' cap prevents the exposure of 5' PO_4-end to the putative exoribonuclease, XRND, and therefore interferes with transcription termination (25). Second, Tandem affinity purification of the largest subunit of CPSF complex and homolog search have revealed that the majority of the components in the 3' end processing machinery are conserved in *T. brucei* genome (19, 26),

including subunits from CPSF and CstF. However, certain subunits of the 3' end processing machinery are missing in the *T. brucei* genome, including Pcf11 and Rna14. In yeast, Pcf11, Rna14 and Rna15 recruit the Rat1 exoribonuclease for transcription termination by torpedo mechanism (21). The absence of two of the factors from the *T. brucei* 3' end processing complex suggests that the exoribonuclease cannot be stably associated with the complex within PTUs, and therefore uncouples 3' end formation with transcription termination. Lastly, the J-binding PJW/PP1 complex provides another possibility. The complex is potentially recruited to J-enriched transcription start and termination regions, which though remains to be tested, dephosphorylate substrates such as RNAPII CTD, halting elongating RNAPII and promote efficient transcription termination by torpedo model at TTRs.

WHAT'S THE ENDOGENOUS SUBSTRATES OF THE PJW/PP1 COMPLEX?

The PJW/PP1 complex displays PP1-dependent phosphatase activity toward RNAPII CTD *in vitro*, instead of another Leishmania phosphoprotein, indicating phosphatase specificity (Chapter 3). Various residues in RNAPII CTD have been implicated to be the target for PP1-PNUTS in other systems. PNUTS colocalizes with RNAPII on chromatin and its loss increases CTD Ser5P in *Drosophila* embryos (27). PNUTS-PP1 complex in human cells shows specific CTD pSer5 phosphatase activity both *in vivo* and *in vitro* (28, 29). Furthermore, Dis2-Ppn1 complex has been implicated to dephosphorylate CTD Thr4P in *S. pombe* (30), while PP1 depletion increased pTy1 in *S.cerevisiae* (31). The effect of the loss of PP1 or PNUTS on RNAPII CTD phosphorylation status, however, usually requires detection by antiserum against a specific phospho-residue and would not lead to an alteration to the ratio of total phosphorylated Rpb1 to unphosphorylated Rpb1 on SDS-PAGE gel. Therefore, it is surprising to us that almost

all the phosphorylated form of Rpb1 of lower SDS gel mobility was converted to the dephosphorylated form of higher SDS gel mobility in a time-dependent manner when incubated with the purified PJW/PP1 complex. In the future, it will be interesting to test if PNUTS or PP1 loss leads to a change in the phosphorylation status *in vivo* in trypanosomatids. Furthermore, Trypanosomatids RNAPII CTD has a non-canonical CTD as described in Chapter 1, and the biological significance of the PTMs on the non-canonical CTD has not been studied. Therefore, it will be interesting to identify the phospho-residues targeted by the PJW/PP1 complex. Lastly, it is also important to ask how RNAPII CTD dephosphorylation affects transcription termination. According to previous research, the underlying mechanism could be altered RNAPII kinetics, dissociation of elongation factors, or association of termination factors, which should be investigated in the future.

Another potential substrate is Spt5. Spt5 CTR is the target for PP1-PNUTS in human cells and fission yeast, and its dephosphorylation facilitates its release or RNAPII pausing at PAS (30, 32, 33). However, just like RNAPII CTD, the trypanosomatid Spt5 lacks a CTR consisting of repetitive sequences, although multiple phosphorylation sites have been mapped in *T. brucei* Spt5 (34). Spt5 does not run as a doublet on SDS-PAGE gel, preventing us from analyzing its phosphorylation states following *in vitro* incubation with the purified PJW/PP1 complex like RNAPII. Currently, the phos-tag gel technique is undertaken to solve this problem.

Identifying the endogenous substrates of the PJW/PP1 complex by a non-biased approach is critical for furthering our understanding of its biological function. A substrate-trapping strategy was used to identify the endogenous substrates of PP1-PIPs (28, 35). In addition, comparative phosphoproteomic analysis could be used to identify potential substrates (36).

HOW DOES THE PJW/PP1 COMPLEX REPRESS SUBTELOMERIC AND TELOMERIC VSG EXPRESSION?

A significant fraction of proteins that got de-repressed after *T. brucei* PNUTS depletion are the subtelomeric VSG genes, metacyclic VSG genes from MESs, and VSG genes from silent BESs. These genes are found in the subtelomeric and telomeric regions, and are transcribed by RNAP I. Transcription of silent expression sites is initiated at promoters, but non-productive transcription elongation or early termination prevents the transcription of the distal VSG genes (37, 38). Therefore, it has been proposed that the enrichment of base J at these regions may recruit the PJW/PP1 complex, which attenuates transcription elongation by RNAP I, although the mechanism remains to be investigated.

Multiple epigenetic marks are involved in silencing these genes, including H3V, H4V, base J and H3 (4, 6, 7, 39, 40). They potentially restrict chromatin accessibility at subtelomeric and telomeric regions, thereby repressing gene expression (40). Therefore, it remains a possibility that the PJW/PP1 complex may act as a chromatin remodeling complex, regulating chromatin structure. This possibility is further supported by the association of JBP3 with SPARC complex and FACT complex (41, 42).

Another possibility that is under investigation is that RNAPII transcription defects lead to transcription readthrough from the chromosomal core PTUs into subtelomeric and telomeric regions, thereby upregulating genes in these regions.

IS THERE ANY OTHER TRANSCRIPTION TERMINATION MECHANISM?

The polymerase associated factor 1 complex (Paf1C) is a highly conserved protein complex across eukaryotes, and was first identified in Pol II-associated proteins in *S. cerevisiae*

(43). The yeast Paf1C contains five member proteins: Paf1, Ctr9, Cdc73, Rtf1 and Leo1, and is enriched on gene bodies (44–46). The human Paf1C occupies the entire transcription unit, including promoters and regions downstream of PAS (47). Consistent with its widespread occupancy on transcription units, the Paf1C is involved in regulating different aspects of transcription. Paf1C associates with proteins known to regulate transcription elongation, such as FACT and Spt5 (48, 49), and stimulates transcription elongation *in vitro* and *in vivo* (50–54). Furthermore, Paf1C regulates transcription termination through its effects on CTD Ser2P and histone PTMs. Paf1C activates the CTD Ser2P kinase, Ctk1 to maintain Ser2P on the gene body (55). Ser2P is required for proper recruitment of 3' end processing factors (56–58). Ser2P CTD also recruits Set2, a histone methyltransferase that deposits H3K36me3 within ORFs to prevent cryptic transcription in yeast (59–63). Moreover, Paf1C recruits and stimulates Rad6-Bre1 complex to mono-ubiquitylate H2BK123/H2BK120 (yeast/human, H2Bub) (46, 64, 65). H2Bub is further required for di- and tri-methylation of H3K4 by the Set1 methyltransferase complex and K3K79 by the Dot1 methyltransferase complex (66). These histone markers are required for proper recruitment of 3' end processing and termination factors, including the NNS complex, and cleavage and polyadenylation complex (67–70). Therefore, loss of Paf1C has been observed to cause defective transcription termination for snoRNA genes and cryptic unstable transcripts (69, 71, 72). *L. tarentolae* JBP3 is associated with the Paf1C components (41), suggesting a potential of crosstalk between the PJW/PP1 complex and Paf1C.

WHAT IS THE INTRAMOLECULAR ARCHITECTURE OF THE PJW/PP1 COMPLEX?

Structural modeling and mutagenesis studies have identified and tested the motifs in LtPNUTS for binding LtPP1-8e, RVxF-ϕϕ-Phe; in addition, we also identified unique sequences in LtPP1-8e that confer isoform specificity to LtPNUTS (Chapter 3). However, how these unique sequences in LtPP1-8e, including N- and C-termini and inserts within the catalytic core structure, interact with LtPNUTS is not yet understood. In the future, determining the structure of the LtPP1-PNUTS complex will be essential to answer this question and expand the diversity of PP1-PIP interaction repertoire.

We found that deletion of the last 23 amino acids from PNUTS C-terminus almost disrupted the association with JBP3 and Wdr82, indicating they bind to C-terminus of PNUTS. Furthermore, the protein stability of both JBP3 and Wdr82 depends on the proper protein level of PNUTS, while depletion of either JBP3 or Wdr82 has no effect on PNUTS protein level in *T. brucei*, further confirming the scaffolding role of PNUTS in the complex. In addition, Wdr82 depletion leads to a reduction in JBP3 protein level, while JBP3 depletion does not affect Wdr82 protein level or Wdr82-PNUTS interaction in *T. brucei*, suggesting that JBP3 binds indirectly to PNUTS through Wdr82. However, this conclusion still needs confirmation through *in vitro* assembly assays using recombinant proteins. Finally, we showed that the cellular level of PNUTS has to be precisely regulated for the integrity of the PJW/PP1 complex. PNUTS overexpression leads to a reduction in PP1 protein level, and surprisingly we also observed that PNUTS overexpression leads to a change in Wdr82 mobility on SDS-PAGE gel. It is important to determine the identity of the lower molecular weight Wdr82 species and understand what caused the altered mobility of Wdr82.

REFERENCE

1. Clayton, C. E. (2002) Life without transcriptional control? From fly to man and back again. *EMBO J.* **21**, 1881–1888
2. Campbell, D. A., Thomas, S., and Sturm, N. R. (2003) Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect.* **5**, 1231–1240
3. Reynolds, D., Cliffe, L., Förstner, K. U., Hon, C.-C., Siegel, T. N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic Acids Res.* **42**, 9717–9729
4. Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., and Sabatini, R. (2016) Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma brucei. *PLoS Genet.* **12**, e1005758
5. Reynolds, D. L., Hofmeister, B. T., Cliffe, L., Siegel, T. N., Anderson, B. A., Beverley, S. M., Schmitz, R. J., and Sabatini, R. (2016) Base J represses genes at the end of polycistronic gene clusters in Leishmania major by promoting RNAP II termination. *Mol. Microbiol.* **101**, 559–574
6. Schulz, D., Zaringhalam, M., Papavasiliou, F. N., and Kim, H.-S. (2016) Base J and H3.V Regulate Transcriptional Termination in Trypanosoma brucei. *PLoS Genet.* **12**, e1005762
7. Kim, H.-S. (2021) Genetic Interaction Between Site-Specific Epigenetic Marks and Roles of H4v in Transcription Termination in Trypanosoma brucei. *Front Cell Dev Biol*. **9**, 744878
8. Lee, J.-H., You, J., Dobrota, E., and Skalnik, D. G. (2010) Identification and characterization of a novel human PP1 phosphatase complex. *J. Biol. Chem.* **285**, 24466–24476

9. Brauer, B. L., Moon, T. M., Sheftic, S. R., Nasa, I., Page, R., Peti, W., and Kettenbach, A. N. (2019) Leveraging New Definitions of the LxVP SLiM To Discover Novel Calcineurin Regulators and Substrates. *ACS Chem. Biol.* **14**, 2672–2682
10. Kliche, J., and Ivarsson, Y. (2022) Orchestrating serine/threonine phosphorylation and elucidating downstream effects by short linear motifs. *Biochem. J.* **479**, 1–22
11. DiPaolo, C., Kieft, R., Cross, M., and Sabatini, R. (2005) Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol. Cell*. **17**, 441–451
12. Genest, P.-A., Ter Riet, B., Cijsouw, T., van Luenen, H. G. A. M., and Borst, P. (2007) Telomeric localization of the modified DNA base J in the genome of the protozoan parasite Leishmania. *Nucleic Acids Res.* **35**, 2116–2124
13. Kieft, R., Zhang, Y., Marand, A. P., Moran, J. D., Bridger, R., Wells, L., Schmitz, R. J., and Sabatini, R. (2020) Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. *PLoS Genet.* **16**, e1008390
14. Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., 3rd, Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V., Pirrotta, V., Reinberg, D., and Gamblin, S. J. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*. **461**, 762–767
15. Migliori, V., Müller, J., Phalke, S., Low, D., Bezzi, M., Mok, W. C., Sahu, S. K., Gunaratne, J., Capasso, P., Bassi, C., Cecatiello, V., De Marco, A., Blackstock, W., Kuznetsov, V., Amati, B., Mapelli, M., and Guccione, E. (2012) Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **19**, 136–144

16. Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W., and Pavletich, N. P. (2007) Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell*. **26**, 131–143
17. Orlicky, S., Tang, X., Willems, A., Tyers, M., and Sicheri, F. (2003) Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell*. **112**, 243–256
18. Zhang, Y., and Sabatini, R. (2023) Leishmania PNUTS discriminates between PP1 catalytic subunits through a RVxF-ΦΦ-F motif and polymorphisms in the PP1 C-tail and catalytic domain. *bioRxiv*. 10.1101/2023.09.20.558696
19. Koch, H., Raabe, M., Urlaub, H., Bindereif, A., and Preußer, C. (2016) The polyadenylation complex of Trypanosoma brucei: Characterization of the functional poly(A) polymerase. *RNA Biol.* **13**, 221–231
20. Mariconti, L., Loll, B., Schlinkmann, K., Wengi, A., Meinhart, A., and Dichtl, B. (2010) Coupled RNA polymerase II transcription and 3' end formation with yeast whole-cell extracts. *RNA*. **16**, 2205–2217
21. Luo, W., Johnson, A. W., and Bentley, D. L. (2006) The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric–torpedo model. *Genes Dev.* **20**, 954–965
22. Richard, P., and Manley, J. L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.* **23**, 1247–1269
23. Grzechnik, P., and Kufel, J. (2008) Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell*. **32**, 247–258
24. Agabian, N. (1990) Trans splicing of nuclear pre-mRNAs. *Cell*. **61**, 1157–1160

25. Li, C.-H., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., Estévez, A. M., and Clayton, C. (2006) Roles of a Trypanosoma brucei 5′→3′ exoribonuclease homolog in mRNA degradation. *RNA.* **12**, 2171–2186
26. Swale, C., and Hakimi, M.-A. (2023) 3'-end mRNA processing within apicomplexan parasites, a patchwork of classic, and unexpected players. *Wiley Interdiscip. Rev. RNA.* **14**, e1783
27. Ciurciu, A., Duncalf, L., Jonchere, V., Lansdale, N., Vasieva, O., Glenday, P., Rudenko, A., Vissi, E., Cobbe, N., Alphey, L., and Bennett, D. (2013) PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. *PLoS Genet.* **9**, e1003885
28. Wu, D., De Wever, V., Derua, R., Winkler, C., Beullens, M., Van Eynde, A., and Bollen, M. (2018) A substrate-trapping strategy for protein phosphatase PP1 holoenzymes using hypoactive subunit fusions. *J. Biol. Chem.* **293**, 15152–15162
29. Landsverk, H. B., Sandquist, L. E., Bay, L. T. E., Steurer, B., Campsteijn, C., Landsverk, O. J. B., Marteijn, J. A., Petermann, E., Trinkle-Mulcahy, L., and Syljuåsen, R. G. (2020) WDR82/PNUTS-PP1 Prevents Transcription-Replication Conflicts by Promoting RNA Polymerase II Degradation on Chromatin. *Cell Rep.* **33**, 108469
30. Kecman, T., Kuś, K., Heo, D.-H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C. V., and Vasiljeva, L. (2018) Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. *Cell Rep.* **25**, 259-269.e5

31. Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. A. (2014) RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.* **21**, 175–179
32. Cortazar, M. A., Sheridan, R. M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D. L. (2019) Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "sitting duck torpedo" mechanism. *Mol. Cell*. **76**, 896-908.e4
33. Parua, P. K., Booth, G. T., Sansó, M., Benjamin, B., Tanny, J. C., Lis, J. T., and Fisher, R. P. (2018) A Cdk9–PP1 switch regulates the elongation–termination transition of RNA polymerase II. *Nature*. **558**, 460–464
34. Nett, I. R. E., Martin, D. M. A., Miranda-Saavedra, D., Lamont, D., Barber, J. D., Mehlert, A., and Ferguson, M. A. J. (2009) The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness. *Mol. Cell. Proteomics*. **8**, 1527–1538
35. Winkler, C., Rouget, R., Wu, D., Beullens, M., Van Eynde, A., and Bollen, M. (2018) Overexpression of PP1-NIPP1 limits the capacity of cells to repair DNA double-strand breaks. *J. Cell Sci.* 10.1242/jcs.214932
36. Song, T., Shen, Y., Jin, Q., Feng, W., Fan, L., and Cai, W. (2020) Comparative phosphoproteome analysis to identify candidate phosphoproteins involved in blue light-induced brown film formation in Lentinula edodes. *PeerJ*. **8**, e9859
37. Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Van Xong, H., and Pays, E. (2000) Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. *Mol. Microbiol.* **36**, 328–340

38. Kassem, A., Pays, E., and Vanhamme, L. (2014) Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in Trypanosoma brucei. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8943–8948

39. Stanne, T. M., and Rudenko, G. (2010) Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. *Eukaryot. Cell*. **9**, 136–147

40. Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A.-E., Sebra, R. P., and Siegel, T. N. (2018) Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*. **563**, 121–125

41. Jensen, B. C., Phan, I. Q., McDonald, J. R., Sur, A., Gillespie, M. A., Ranish, J. A., Parsons, M., and Myler, P. J. (2021) Chromatin-Associated Protein Complexes Link DNA Base J and Transcription Termination in Leishmania. *mSphere*. 10.1128/mSphere.01204-20

42. Staneva, D. P., Bresson, S., Auchynnikava, T., Spanos, C., Rappsilber, J., Jeyaprakash, A. A., Tollervey, D., Matthews, K. R., and Allshire, R. C. (2022) The SPARC complex defines RNAPII promoters in Trypanosoma brucei. *Elife*. 10.7554/eLife.83135

43. Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., Jaehning, J. A., and Burton, Z. F. (1996) A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr. Purif.* **8**, 85–90

44. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278

45. Fischl, H., Howe, F. S., Furger, A., and Mellor, J. (2017) Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. *Mol. Cell*. **65**, 685-698.e8

46. Van Oss, S. B., Shirra, M. K., Bataille, A. R., Wier, A. D., Yen, K., Vinayachandran, V., Byeon, I.-J. L., Cucinotta, C. E., Héroux, A., Jeon, J., Kim, J., VanDemark, A. P., Pugh, B. F., and Arndt, K. M. (2016) The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6. *Mol. Cell.* **64**, 815–825

47. Yang, Y., Li, W., Hoque, M., Hou, L., Shen, S., Tian, B., and Dynlacht, B. D. (2016) Correction: PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. *PLoS Genet.* **12**, e1005883

48. Squazzo, S. L., Costa, P. J., Lindstrom, D. L., Kumer, K. E., Simic, R., Jennings, J. L., Link, A. J., Arndt, K. M., and Hartzog, G. A. (2002) The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* **21**, 1764–1774

49. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992

50. Rondón, A. G., Gallardo, M., García-Rubio, M., and Aguilera, A. (2004) Molecular evidence indicating that the yeast PAF complex is required for transcription elongation. *EMBO Rep.* **5**, 47–53

51. Gaillard, H., Tous, C., Botet, J., González-Aguilera, C., Quintero, M. J., Viladevall, L., García-Rubio, M. L., Rodríguez-Gil, A., Marín, A., Ariño, J., Revuelta, J. L., Chávez, S., and Aguilera, A. (2009) Genome-wide analysis of factors affecting transcription elongation

and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. *PLoS Genet.* **5**, e1000364

52. Tous, C., Rondón, A. G., García-Rubio, M., González-Aguilera, C., Luna, R., and Aguilera, A. (2011) A novel assay identifies transcript elongation roles for the Nup84 complex and RNA processing factors. *EMBO J.* **30**, 1953–1964
53. Hou, L., Wang, Y., Liu, Y., Zhang, N., Shamovsky, I., Nudler, E., Tian, B., and Dynlacht, B. D. (2019) Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14583–14592
54. Kim, J., Guermah, M., and Roeder, R. G. (2010) The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*. **140**, 491–503
55. Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R. G. (2015) RNA polymerase II–associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science*. **350**, 1383–1386
56. Ahn, S. H., Kim, M., and Buratowski, S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell*. **13**, 67–76
57. Davidson, L., Muniz, L., and West, S. (2014) 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev.* **28**, 342–356
58. Eifler, T. T., Shao, W., Bartholomeeusen, K., Fujinaga, K., Jäger, S., Johnson, J. R., Luo, Z., Krogan, N. J., and Peterlin, B. M. (2015) Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts. *Mol. Cell. Biol.* **35**, 468–478

59. Fuchs, S. M., Kizer, K. O., Braberg, H., Krogan, N. J., and Strahl, B. D. (2012) RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the Set2 methyltransferase and histone H3 di-and trimethylation at lysine 36. *J. Biol. Chem.* **287**, 3249–3256

60. Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., and Strahl, B. D. (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* **25**, 3305–3316

61. Keogh, M.-C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., Chin, K., Punna, T., Thompson, N. J., Boone, C., Emili, A., Weissman, J. S., Hughes, T. R., Strahl, B. D., Grunstein, M., Greenblatt, J. F., Buratowski, S., and Krogan, N. J. (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*. **123**, 593–605

62. Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*. **123**, 581–592

63. Drouin, S., Laramée, L., Jacques, P.-É., Forest, A., Bergeron, M., and Robert, F. (2010) DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS Genet.* **6**, e1001173

64. Piro, A. S., Mayekar, M. K., Warner, M. H., Davis, C. P., and Arndt, K. M. (2012) Small region of Rtf1 protein can substitute for complete Paf1 complex in facilitating global histone H2B ubiquitylation in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10837–10842

65. Warner, M. H., Roinick, K. L., and Arndt, K. M. (2007) Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. *Mol. Cell. Biol.* **27**, 6103–6115

66. Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J. Biol. Chem.* **278**, 34739–34742

67. Kaczmarek Michaels, K., Mohd Mostafa, S., Ruiz Capella, J., and Moore, C. L. (2020) Regulation of alternative polyadenylation in the yeast Saccharomyces cerevisiae by histone H3K4 and H3K36 methyltransferases. *Nucleic Acids Res.* **48**, 5407–5425

68. Tomson, B. N., Davis, C. P., Warner, M. H., and Arndt, K. M. (2011) Identification of a role for histone H2B ubiquitylation in noncoding RNA 3'-end formation through mutational analysis of Rtf1 in Saccharomyces cerevisiae. *Genetics*. **188**, 273–289

69. Tomson, B. N., Crisucci, E. M., Heisler, L. E., Gebbia, M., Nislow, C., and Arndt, K. M. (2013) Effects of the Paf1 complex and histone modifications on snoRNA 3'-end formation reveal broad and locus-specific regulation. *Mol. Cell. Biol.* **33**, 170–182

70. Terzi, N., Churchman, L. S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011) H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol. Cell. Biol.* **31**, 3569–3583

71. Ellison, M. A., Lederer, A. R., Warner, M. H., Mavrich, T. N., Raupach, E. A., Heisler, L. E., Nislow, C., Lee, M. T., and Arndt, K. M. (2019) The Paf1 Complex Broadly Impacts the Transcriptome of Saccharomyces cerevisiae. *Genetics*. **212**, 711–728

72. Sheldon, K. E., Mauger, D. M., and Arndt, K. M. (2005) A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. *Mol. Cell*. **20**, 225–236

APPENDIX A

IDENTIFICATION OF A NOVEL BASE J BINDING PROTEIN COMPLEX INVOLVED IN RNA POLYMERASE II TRANSCRIPTION TERMINATION IN TRYPANOSOMES[1]

[1]Kieft, R., **Zhang, Y**., Marand, A.P., Moran, J.D., Bridger, R., Wells, L., Schmitz, R.J. and Sabatini, R. 2020. *PLoS Genetics*, 16(2), p.e1008390.

Reprinted here with permission of publisher

ABSTRACT

Base J, β-D-glucosyl-hydroxymethyluracil, is a modification of thymine DNA base involved in RNA Polymerase (Pol) II transcription termination in kinetoplastid protozoa. Little is understood regarding how specific thymine residues are targeted for J-modification or the mechanism of J regulated transcription termination. To identify proteins involved in J-synthesis, we expressed a tagged version of the J-glucosyltransferase (JGT) in *Leishmania tarentolae*, and identified four co-purified proteins by mass spectrometry: protein phosphatase (PP1), a homolog of Wdr82, a potential PP1 regulatory protein (PNUTS) and a protein containing a J-DNA binding domain (named JBP3). Gel shift studies indicate JBP3 is a J-DNA binding protein. Reciprocal tagging, co-IP and sucrose gradient analyses indicate PP1, JGT, JBP3, Wdr82 and PNUTS form a multimeric complex in kinetoplastids, similar to the mammalian PTW/PP1 complex involved in transcription termination via PP1 mediated dephosphorylation of Pol II. Using RNAi and analysis of Pol II termination by RNA-seq and RT-PCR, we demonstrate that ablation of PNUTS, JBP3 and Wdr82 lead to defects in Pol II termination at the 3'-end of polycistronic gene arrays in *Trypanosoma brucei*. Mutants also contain increased antisense RNA levels upstream of transcription start sites, suggesting an additional role of the complex in regulating termination of bi-directional transcription. In addition, PNUTS loss causes derepression of silent Variant Surface Glycoprotein genes involved in host immune evasion. Our results suggest a novel mechanistic link between base J and Pol II polycistronic transcription termination in kinetoplastids.

Author summary

Trypanosoma brucei is a parasitic protozoan that causes African sleeping sickness in humans. The genome of *T. brucei* is organized into polycistronic gene clusters that contain

multiple genes that are co-transcribed from a single promoter. We have recently described the presence of a modified DNA base J and variant of histone H3 (H3.V) at transcription termination sites within gene clusters where the loss of base J and H3.V leads to read-through transcription and the expression of downstream genes. We now identify a novel stable multimeric complex containing a J binding protein (JBP3), base J glucosyltransferase (JGT), PP1 phosphatase, PP1 interactive-regulatory protein (PNUTS) and Wdr82, which we refer to as PJW/PP1. A similar complex (PTW/PP1) has been shown to be involved in Pol II termination in humans and yeast. We demonstrate that PNUTS, JBP3 and Wdr82 mutants lead to read-through transcription in *T. brucei*. Our data suggest the PJW/PP1 complex regulates termination by recruitment to termination sites via JBP3-base J interactions and dephosphorylation of specific proteins (including Pol II and termination factors) by PP1. These findings significantly expand our understanding of mechanisms underlying transcription termination in eukaryotes, including organisms that utilize polycistronic transcription and novel epigenetic marks such as base J and H3.V. The studies also provide the first direct mechanistic link between J modification of DNA at termination sites and regulated Pol II termination and gene expression in kinetoplastids.

INTRODUCTION

Termination of RNA polymerase II (Pol II) transcription of mRNAs is a tightly regulated process where the polymerase stops RNA chain elongation and dissociates from the end of the gene or transcription unit. However, the underlying termination mechanism is not fully understood. Pol II termination of most protein-encoding genes in eukaryotes is tightly linked to the processing of the nascent transcript 3’ end (reviewed in [1]). This association ensures complete formation of stable polyadenylated mRNA products and prevents the elongating Pol II

complex from interfering with transcription of downstream genes. Transcription through the polyadenylation site results in an exchange of transcription factors, resulting in the regulation of the elongation-to-termination transition, in an ordered series of events: 1) dissociation of the elongation factor Spt5, 2) Pol II pausing, 3) changes in phosphorylation status of Pol II C-terminal domain (CTD), which promotes 4) recruitment of cleavage factors and termination factors, 5) transcript cleavage and 6) termination by the 5'-3' 'torpedo' exoribonuclease XRN2/Rat1.

Critical to this process is the regulation of protein phosphorylation by the major eukaryotic protein serine/threonine phosphatase, PP1. As recently demonstrated in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*), the PP1 phosphatase Dis2 regulates termination by de-phosphorylating both the Pol II-CTD as well as Spt5 [2]. This PP1-dependent dephosphorylation allows the efficient recruitment of the termination factor Seb1 as well as decreased Spt5 stimulation of Pol II elongation that enhances the ability of the torpedo exoribonuclease to catch up and destabilize the elongation complex. Transcription termination was also affected in the *S. cerevisiae* PP1 homologue Glc7 mutant [3]. PP1 action is modulated through the formation of heteromeric complexes with specific regulatory subunits [4]. These regulatory protein subunits regulate PP1 by targeting the protein to specific subcellular compartments, to particular substrates, or reduce its activity towards potential substrates. The vast majority of regulators bind PP1 via a primary PP1-binding motif, the RVxF motif [4–6]. One of the key PP1 regulatory proteins in the nucleus is the PP1 nuclear targeting subunit (PNUTS) [7, 8]. PNUTS is a multidomain protein that contains the canonical PP1 RVxF interaction motif, an N-terminal domain that interacts with the DNA binding protein Tox4 and a domain near the C-terminus that interacts with Wdr82 [9]. This stable multimeric complex in

humans is named the PTW/PP1 complex. Targeting of the complex to chromatin is presumably due in part through associations with Tox4. While the function of Wdr82 in this complex is not known, it may mediate interactions with Pol II by recognizing Ser5-phosphorylated CTD, as it does when it is associated with the Set1 complex [10]. In yeast, PP1 is associated with PNUTS and Wdr82 homologs in APT, a subcomplex of the cleavage and polyadenylation factor [11–14], and deletion of PNUTS and Wdr82 caused termination defects at Pol II-dependent genes [15, 16]. In mammals, PNUTS and Wdr82 mutant cells have defects in transcription termination at the 3' end of genes and 5' antisense transcription at bi-directional promoters [17].

Members of the Kinetoplastida order include the parasite *Trypanosoma brucei* that causes human and animal African trypanosomiasis. Kinetoplastids are protozoa with unique genome arrangements where genes are organized into polycistronic transcription units (PTU) that are transcribed by Pol II. Pre-messenger RNAs (mRNA) are processed to mature mRNA by coupled 5' RNA trans-splicing and 3' polyadenylation [18–21]. Given the close relationship between poly(A) processing and transcription termination of most Pol II transcribed protein-coding genes, it is not clear how multiple functional poly(A) sites within the trypanosome PTU can be transcribed without resulting in premature termination. While little is known regarding the mechanism of Pol II termination in kinetoplastids, two chromatin factors, base J and histone H3 variant (H3.V), have recently been shown to be involved. Base J is a modified DNA base found in kinetoplastids where a glucose moiety, linked via oxygen to the thymine base, resides in the major groove of DNA (reviewed in [22]). In *T. brucei* and *Leishmania major*, J and H3.V are enriched at sites involved in Pol II termination [23–26]. This includes sites within polycistronic gene clusters where (premature) termination silences downstream genes. Loss of J or H3.V leads to read-through transcription [24–29]. At PTU internal termination sites this leads to increased

expression of the downstream genes [24, 27, 28]. This epigenetic regulation of termination is thought to allow developmentally regulated expression of specific transcriptionally silent genes [24, 27, 28]. In *T. brucei*, this includes the expression of variant surface glycoprotein (VSG) genes involved in antigenic variation during bloodstream infections [24, 27].

It is currently unclear how base J and H3.V are involved in Pol II termination, since very little is understood regarding the mechanism of Pol II termination in kinetoplastids. Equally unclear is what regulates the specific localization of these epigenetic marks, including base J, in the genome. Base J is synthesized in a two-step pathway in which a thymidine hydroxylase (TH), JBP1 or JBP2, hydroxylates thymidine residues at specific positions in DNA to form hydroxymethyluracil (hmU) [30], followed by the transfer of glucose to hmU by the glucosyltransferase, JGT [31, 32] (reviewed in [22, 33]). The TH enzymes, JBP1 and JBP2, contain a TH domain at the N-terminus [30, 34–37]. JBP1 has a J-DNA binding domain in the C-terminal half of the protein that is able to bind synthetic J-DNA substrates *in vitro* and bind chromatin in a J-dependent manner *in vivo* [38–41]. The ability of JBP1 to bind J-DNA is thought to play a role in J propagation and maintenance. JBP2 does not bind the modified base directly, but is able to bind chromatin in a base J independent manner, presumably via the C-terminal SWI2/SNF2 domain [34, 38]. The JGT lacks DNA sequence specificity, and can convert hmU to base J *in vivo* regardless of where it is present [42–44]. This and other analyses of J synthesis indicate that JBP1 and JBP2 are the key regulatory enzymes of J synthesis.

In this study, we identify a new J-binding protein (called JBP3) in kinetoplastids, which is present in a complex containing PP1, Wdr82 and a putative orthologue of PNUTS. To characterize the role of this (PJW/PP1) complex in transcription termination, we investigated the consequence of mutants using RNAi in *T. brucei*. Ablation of JBP3, Wdr82 or PNUTS

in *T. brucei* causes read-through transcription at termination sites. As we previously demonstrated following the removal of base J and H3.V, these defects include transcription read-through at termination sites within Pol II transcribed gene arrays and the silent Pol I transcribed VSG expression sites, leading to de-repression of genes involved in parasite pathogenesis. Furthermore, ablation of JBP3, Wdr82 or PNUTS results in expression of genes upstream of Pol II transcription start sites. Presumably this represents a previously unappreciated role of termination of antisense transcription from gene promoters in trypanosomes. Overall these findings provide a first look at mechanisms involved in Pol II transcription termination in kinetoplastids and a direct link between base J and termination.

RESULTS

Affinity purification and identification of JGT-associated proteins in Leishmania

Base J is synthesized in a two-step pathway involving the hydroxylation of specific thymidine residues in the genome by a thymidine hydroxylase, JBP1 or JBP2, followed by the addition of the glucose moiety by the glucosyltransferase, JGT. In this study, we chose to identify the proteins that co-purify with JBP1, JBP2 and JGT, in order to understand the regulation of J-biosynthesis. *Leishmania tarentolae* was chosen as an experimental system, because it provides an easily grown source of high densities of parasites that synthesize base J. The proteins were cloned into the pSNSAP1 vector, carrying a C-terminal tag composed of a protein A domain, the TEV protease cleavage site, and the streptavidin binding peptide [45]. Separation of the final eluate from the JGT tagged cell line by SDS-PAGE and staining of the proteins by silver staining, revealed a dominant protein band of ~120 kDa, representing the JGT-strep fusion protein (JGT-S) and several co-purified protein bands that are not visible in the

control purification from untagged wildtype cells (Fig A.1A). In contrast, and unexpectedly, no co-purified proteins were identified from the JBP1-S or JBP2-S purification either by SDS-PAGE or mass spectrometry analysis. Therefore, we continued the analysis on the JGT purification. Gel based separations followed by in-gel digestion and liquid chromatography-tandem mass spectrometry (LC-MS/MS) of the JGT-S purification revealed a total of five proteins with at least 20 PSMs that were enriched at least 40 fold compared to the negative control purification (Table 1). A complete list of proteins identified by MS is shown in S1 Table. As expected, JGT-S was recovered in the eluate. The four additional potential JGT associated proteins were discovered with (hypothetical) molecular masses of 74, 42 and 29 kDa. One of these JGT-interacting proteins (42 kDa; LtaP15.0230) was protein phosphatase 1 (PP1), which contains a PP1 catalytic domain (Fig A.1B). The other 42-kDa protein (LtaP32.3990) contains three WD repeat domains and has been identified as a homologue of Wdr82/Swd2 (human/yeast). The remaining two JGT interacting proteins had not been previously characterized. We named the 74-kDa protein JBP3 (LtaP36.0380) because it has a domain with homology to the base J DNA binding domain of JBP1 (Fig A.1B and A.S1) and we demonstrate its ability to bind J-DNA (see below). The 29-kDa JGT interacting protein was named PNUTS (LtaP33.1440) since it contains a conserved RVxF PP1 interactive domain (Fig A.1C and A.S2) within an apparent intrinsically disordered region of the protein and is a part of a complex similar to the PTW/PP1 complex in humans (where JBP3 may represent a functional homolog of Tox4). PP1 interactive proteins such as PNUTS are highly disordered in their unbound state and fall in a group of intrinsically disordered proteins (IDPs) [4–6]. This intrinsic flexibility is important for the formation of extensive interactions with PP1 [6, 46, 47]. A bioinformatics analysis using the DISOPRED3 program [48, 49], which scores for the occurrence of disorder-inducing amino

acids, predicts a majority of TbPNUTS is disordered (Fig A.S2). Similarly, the Compositional profiler [50] shows PNUTS is enriched in major disorder-promoting residues and depleted in major order-promoting residues. This inherent disorder may explain why LtPNUTS and TbPNUTS migrate slower in SDS-PAGE than predicted (see Fig A.1A and A.S3A). To confirm the complex, we subsequently performed tandem affinity purifications with tagged JBP3 and Wdr82 followed by shotgun proteomics of the soluble fraction. Reciprocal purification of JBP3 and Wdr82 resulted in the identification of JGT, PNUTS, PP1, JBP3 and Wdr82 with at least 10-fold enrichment of each component compared to WT control purification (S2 Table). These subsequent purifications and shotgun MS analyses included a replicate of the tagged JGT pull-down. Interestingly, this JGT purification resulted in all components of the complex except PP1 (S2 Table), suggesting PP1 is the least stable component. These data indicate that in Leishmania JGT associates with a protein complex composed of PNUTS, PP1, JBP3 and Wdr82 similar to the PTW/PP1 complex in humans, shown to be involved in transcription termination (Fig A.1D). JBP3 may be a functional homologue of the human Tox4 DNA binding protein (see below). Based on this similarity, we now refer to this complex as PJW/PP1 in Leishmania.

JBP3 is a J-DNA binding protein

We noticed that a region of the LtaP36.0380 protein (amino acids 101–277) has sequence similarity with the J-binding domain of JBP1 (Fig A.S1A and A.S1B). A structural model of the Lt JBP1 JBD based on the X-ray crystallography-based structure (PDB ID 2xseA) is shown in Fig A.S1C. As expected from the sequence similarities between the kinetoplastid JBD for JBP1 and the putative JBP3, the 3D models generated by the comparative modeling program I-TASSER [51, 52] were of high quality, with TM-score value of 0.727. The sequence identity and similar domain composition to the JBD of JBP1 supported our contention that the LtaP36.0380

protein might be a J-binding protein, subsequently named JBP3. JBP3, similar to JBP1, appears to be conserved only in kinetoplastids (Fig A.S4).

We have previously developed a rapid isolation procedure for His-tagged recombinant JBP1 produced in *E. coli*, and gel shift assays to characterize J-DNA binding activities [40, 41]. We utilized a similar approach to investigate the specific interaction of JBP3 and J-modified DNA (J-DNA). To determine the ability of JBP3 to bind J-DNA, we used the gel shift assay to investigate the binding of recombinant JBP3 (Fig A.1E) to J-DNA duplex (VSG-1J) that has a single centrally located J modification compared with the same duplex without any base J (VSG-T). The J-DNA substrate was incubated with increasing amounts of JBP3 protein, and the complex was analyzed on native gels. The results of the gel shift assay in Fig A.1F, show that the amount of free J-DNA decreases with increasing concentrations of JBP3, concomitant with formation of the JBP3/J-DNA complex. In contrast, incubation of the unmodified DNA substrate with the highest concentration of JBP3 resulted in no visible complex. Therefore, JBP3 is a J-DNA binding protein.

Characterization of the putative phosphatase complex containing Wdr82, JBP3 and PNUTS in *T. brucei*

To study the function of the PJW/PP1 complex *in vivo*, we switched to *T. brucei* due to the benefits of forward and reverse genetics available in this system. However, while all other components are easily identified (see Table 1), no homologue of the *L. tarentolae* PP1 gene is present in the *T. brucei* genome [53, 54]. To characterize the complex in *T. brucei*, and identify the TbPP1 protein component, we used a TAP tagging approach with a PTP epitope tag [55] to identify TbPNUTS-interacting proteins. We generated a clonal procyclic form (PCF) *T. brucei* cell line expressing a C-terminal PTP-tagged TbPNUTS protein from its

endogenous locus. This cell line was used for the TAP procedure. Briefly, a crude protein extract was first purified by IgG affinity chromatography, and the TEV protease was used to cleave off the Protein A portion of the PTP tag. Subsequently, the TEV protease eluate underwent anti-Protein C affinity purification, and the final purified products were eluted with EGTA. The concentrated proteins were then trypsin-digested and analyzed by shotgun LC-MS/MS. As a control, we analyzed the eluate of a comparable purification of extract from wild-type *T. brucei*. If we subtract all proteins identified at 1% FDR in the WT control and only consider proteins with a score over 100, we identify seven proteins that include TbPNUTS, Wdr82 and JBP3 (S3 Table). The other four remaining proteins are commonly identified proteins in immunopurification / MS experiments, including tubulin, heat shock protein and elongation factor 1 [56]. However, neither JGT, nor a protein phosphatase, were identified as part of the TbPNUTS complex. In contrast to Leishmania, base J synthesis is developmentally regulated in *T. brucei*, where synthesis is lost upon differentiation to PCF due to the downregulation of JBP1 and JBP2 [22, 38, 44, 57]. While JGT has been shown to be part of the nuclear proteome in PCF *T. brucei* [58] and significant level of JGT activity is present in PCF, as shown by the presence of J in parasites grown in the presence of the J precursor hmU [38, 44], JGT mRNA is downregulated ~ 2-fold between BSF and PCF [59]. Therefore, to fully characterize the association of JGT with the TbPNUTS complex, we need to utilize BSF *T. brucei*.

Co-immunoprecipitation studies were performed to assess the PNUTS-containing complex in BSF *T. brucei*. Representative components of the putative PP1 complex were analyzed by western blot following immunoprecipitation to assess the authenticity of protein interactions identified by mass spectrometry. We generated BSF *T. brucei* cell lines expressing PTP-tagged versions of PNUTS, JBP3, Wdr82 and CPSF73 (negative control) (Fig A.S3A)

along with HA-tagged versions of JGT, Wdr82 and JBP3 (Fig A.2A and A.S3B). PTP-PNUTS immunoprecipitation recovers HA-Wdr82 and HA-JBP3, but not the La negative control (Fig A.2A). Similarly, PTP-JBP3 immunoprecipitation recovers HA-Wdr82 (Fig A. S3B). In contrast, no detectable HA-Wdr82 or HA-JGT is recovered in PTP-CPSF73 immunoprecipitates (Fig A.2A and A.S3B). Furthermore, JGT does not co-purify when PNUTS, JBP3 or Wdr82 is pulled down (Fig A.S3B), nor does Hsp70 co-purify when PNUTS or JBP3 is pulled down (Fig A.S3C).

To further characterize the PNUTS complex, extracts recovered from BSF *T. brucei* cells expressing epitope-tagged components were analyzed by western blot following sucrose fractionation. Analysis of *T. brucei* cells expressing PTP-PNUTS and HA-JBP3 or HA-Wdr82 indicate Wdr82, JBP3 and PNUTS co-migrate at <200kDa. These results indicate that JBP3, Wdr82 and PNUTS form a stable complex of <200kDa. Summation of the predicted size of the three complex components (~180 kDa) agrees with the observed mass of the PJW complex and suggests a 1:1 stoichiometry for the subunits. Taken together, data from co-immunoprecipitation, sucrose gradient analysis, and identification of PNUTS-, Wdr82-, and JBP3-associated proteins by mass spectrometry indicate that Wdr82, JBP3 and PNUTS comprise a stable PJW complex in kinetoplastids (Fig A.1D). In contrast to the complex in Leishmania, the stable purified *T. brucei* complex lacks both JGT and PP1.

LtPP1 is found to be stably associated in the PJW/PP1 complex, possibly through its putative PP1-interacting protein, PNUTS. PNUTS is present in both the *T. brucei* and Leishmania complexes and share a conserved RVxF motif (Fig A.1C). Moreover, a similar human PTW/PP1 complex suggests that PP1 confers the complex phosphatase activity critical for its regulation in Pol II termination [2, 3, 60, 61]. The apparent lack of PP1 in

the *T. brucei* complex is therefore surprising. To identify the TbPP1 component, we directly tested two of the eight PP1 genes in *T. brucei* for interactions with the complex by co-IP. PP1-1 has the highest sequence homology to the PP1 involved in termination in yeast and humans [54]. PP1-7 has been identified in the nucleus of *T. brucei* [58]. However, neither of these PP1 proteins associates with TbPNUTS in co-IP experiments (Fig A.S3C). Another possible explanation for the lack of PP1 in the PNUTS purification, or Co-IP, is that PP1 is not stably associated with the complex in *T. brucei*. This is consistent with one of four replicates of PJW/PP1 purification that resulted in all components of the complex except PP1 (Table 1), suggesting PP1 is the least stable component of the Leishmania complex.

Downregulation of PNUTS, Wdr82 and JBP3 causes defects in RNA Pol II transcription termination at the 3'end of PTUs

If PNUTS, Wdr82 and JBP3 interact with each other, one would predict that RNAi against the individual components would give the same phenotype. We therefore analyzed the role of the PJW/PP1 complex in transcription termination in BSF *T. brucei*. We have previously shown that base J and H3.V are present at termination sites within a PTU where loss of either epigenetic mark results in read-through transcription and increased expression of downstream genes [24, 27, 28]. Because Pol II elongation and gene expression is inhibited prior to the end of these PTUs, we refer to this as PTU internal termination. For example, base J and H3.V are involved in terminating transcription upstream of the last two genes (VSG; Tb927.5.3990 and Hypothetical protein; Tb927.5.4000) in a PTU on chromosome 5 (Fig A.3, top) where deletion of H3.V or inhibition of base J synthesis leads to read-through transcription [24, 27]. The resulting read-through RNAs become observable in a manner similar to those seen in other systems when termination mechanisms become incapacitated by experimental manipulation. The presence of

an open reading frame downstream of the termination site allows an additional measure of read-through where nascent RNA is processed to stable capped and polyadenylated mRNA species. As such, the loss of either epigenetic mark in *T. brucei* leads to generation of nascent RNA extending beyond the termination site and expression of the two downstream genes [24, 27]. To investigate the physiological function of the PJW complex, we analyzed inducible RNAi ablation of Wdr82, JBP3, and PNUTS in BSF *T. brucei*. As shown in Fig A.3, induction of RNAi against Wdr82, JBP3 and PNUTS and ablation of mRNA levels from 30–60% leads to reduced parasite growth, indicating that the proteins are important for normal cell proliferation in BSF *T. brucei*. In the case of PNUTS, ablation of mRNA levels to 50% leads to ~90% reduction at the protein level (Fig A.S5). We also detect evidence of read-through transcription at the representative PTU internal termination site on chromosome 5 upon ablation of the three factors. RT-PCR using oligos flanking the termination site (see diagram, Fig A.3A) detects increased RNA upon ablation of PNUTS, Wdr82 and JBP3 (Fig A.3C). As a control a separate RT-PCR utilized the same 5' primer and a 3' primer immediately upstream of the termination site. We have previously shown that an RNA species spanning the termination site is indicative of read-through transcription and is only detected following the loss of base J or H3.V, due to continued transcription elongation at termination sites [24, 27, 28]. Consistent with read-through transcription, both genes downstream of the termination site are significantly de-repressed upon the ablation of the three components of the PJW complex, in contrast to genes upstream (Fig A.3D). In contrast, no significant termination defects are detected upon ablation of a negative control, acidocalcisome VA a protein (Fig A.3C and A.3D). VA a has been shown to be an essential gene in *T. brucei* and ablation results in growth defects similar to those seen in PJW complex [62] (Fig A.3A). Therefore, the read-through defects measured in the PNUTS, Wdr82

and JBP3 mutants are presumably not the result of indirect effects of dying cells. These results suggest that PNUTS, Wdr82 and JBP3, possibly functioning in the PJW complex, are essential for Pol II termination.

To further explore the role of the PJW complex in the regulation of termination and whether the complex functions similarly across the genome, we performed stranded mRNA-seq to compare the expression profiles of PNUTS RNAi cells with and without tetracyclin induction. This led to the detection of 709 mRNAs that are increased at least 3-fold upon ablation of PNUTS ($P_{adj}<0.05$) (Fig A.S6A and A.S4 Table). In contrast, no mRNAs are downregulated. Of the 3-fold upregulated genes, a majority represent VSGs, ESAGs and Retrotransposon Hot Spot proteins (RHS) that are repressed in bloodstream *T. brucei* and are localized at the end of Pol II transcribed PTUs located within the chromosomes or in subtelomeres. The location of the genes with >3-fold upregulation was mapped (Fig A.4A and A.S7). Interestingly, these genes were closely located at regions flanking PTUs or subtelomeric regions, suggesting a role of PNUTS in genome-wide transcription, specifically at transcription initiation and termination sites, as well as subtelomeres. As evident in the genome-wide view (Fig A.S7), a significant fraction of the upregulated genes represent the silent subtelomeric VSG and RHS gene arrays while the majority of the regulated genes adjacent to PTU flanking regions in the chromosome core are annotated as hypothetical proteins of unknown function (Fig A.4A and A.S7 and A.S4 Table). The PTU flanking regions represent transcription termination sites (TTS) and transcription start sites (TSS) within the so-called convergent strand switch region (cSSR) and divergent strand switch region (dSSR), respectively. We have previously shown that base J is localized within the cSSRs and dSSRs flanking PTUs and involved in Pol II termination. Therefore, we wanted to first confirm the association with J at termination sites with changes in gene expression we

observed in the PNUTS mutant. If we analyze genes within 1kb of base J genome-wide, we see a correlation with upregulation of gene expression versus a similar analysis of the same number of randomly selected genes across the genome (Fig A.S6B). In contrast, the 4 genes upregulated in the VA a mutant lack any association with base J. The upregulated genes in the PNUTS mutant located at the 3' end of PTUs, including the VSG gene (Tb927.5.3990) analyzed in Fig A.3, map to regions downstream of base J and H3.V. The RNA-seq results confirm our initial RT-PCR analysis of nascent and steady-state RNA indicating a role of PNUTS, Wdr82 and JBP3 regulating termination and expression of downstream genes (Fig A.3). In these studies, we are measuring gene expression levels at day 2 of the RNAi to avoid potential secondary effects due to cell growth defects. While many of these changes reflect increases from silent or extremely low initial levels of RNA, and thus represent small total increases in mRNA levels, the changes are reproducible, significant and therefore, we believe to be biologically relevant. Consistent with our analysis of the J/H3.V mutants, we observe similar increases in the PNUTS mutant in the expression of genes downstream of PTU internal termination sites, which we previously demonstrated is caused by a defect in Pol II transcription termination resulting in readthrough transcription [24, 27, 28].

A remaining question was whether PNUTS also regulated termination at the 3'-end of gene clusters resulting in transcription into the SSR. Our previous RNA-seq analyses in *T. brucei* indicated that H3.V/J regulated termination at the 3'-end of PTUs, where mutants resulted in Pol II elongation into the dual strand transcription regions at cSSRs [24, 27]. To visualize the differences between WT and TbPNUTS mutant, fold changes between -Tet and +Tet induction of RNAi were plotted (Fig A.S8). While sense transcription remained largely unaffected throughout all 11 megabase chromosomes following the loss of PNUTS, significant

fold increases of antisense transcription were observed near transcription borders of PTUs, including downstream of normal transcription termination sites at the 3'-end. To take a closer look at the increased level of transcripts at these sites, and determine whether they are due to transcriptional readthrough, forward and reverse reads mapping to 5kb flanking and within cSSRs were counted and RPKM values were generated. As shown in Fig A.4B, cSSRs are computationally defined regions where coding strands switch based on the transcriptome. To see the difference between WT and PNUTS mutant, changes in transcript abundance upon PNUTS and VA a ablation were plotted. A metaplot summarizing the readthrough defect for all cSSRs is shown in Fig A.4B. We observed an increase in RNA extending into the cSSR in the PNUTS mutant, and no change in the VA a mutant control. Boxplots comparing the median RPKM values for SSRs also indicated that cSSRs were upregulated in the PNUTS mutant and the differences were statistically significant (Fig A.S9). All together these results indicate that the loss of PNUTS, JBP3, and Wdr82 result in defects in transcription termination at the 3'-end of PTUs as we described in the H3.V/J mutant.

PNUTS, Wdr82 and JBP3 are involved in Pol II transcription termination at 5' end of PTUs

Based on the initial RT-PCR analysis of the internal termination site in PNUTS, JBP3 and Wdr82 knockdowns (Fig A.3), we expected RNA-seq to illustrate defects in termination at 3'-end of PTUs genome-wide in the PNUTS mutant. However, we also detected accumulation of transcripts at dSSRs, illustrated by the genome-wide map of transcript levels (Fig A.S8) and the metaplot of dSSRs (Fig A.4C and A.S9). At these divergent TSSs, total RNA levels detected in the region upstream of the start site are increased in the PNUTS mutant when forward and reverse reads were analyzed separately, suggesting that transcription initiates upstream of its start

site upon the loss of PNUTS. In some cases, this leads to expression of genes present in the dSSR that are silent in WT cells. This explains the mapping of genes upregulated in the PNUTS mutant to 5'-end PTUs (Fig A.4A). A specific example, shown in Fig A.5A, includes a dSSR on chromosome 10 where a gene (Tb927.10.8340) located between the two divergent TSSs (mapped by tri-phosphate RNA sequencing of WT cells) is affected by the loss of PNUTS. Specific upregulation of this gene 4- to 13-fold following the loss of PNUTS, JBP3 and Wdr82 versus genes in the adjacent PTU is confirmed by RT-qPCR (Fig A.5B). Another example confirmed by RT-qPCR is shown in Fig A.S10 where the gene (Tb927.10.6430) located upstream of the TSS in WT cells is specifically upregulated 4- to 7.5-fold by the loss of PNUTS, JBP3 and Wdr82.

Many promoters for RNA Pol II are bidirectional in organisms from yeast to human [63–67]. Unidirectional transcription resulting in productive mRNAs is typically ensured since antisense transcription is susceptible to early termination linked to rapid degradation. Mapping TSSs using tri-phosphate RNA-seq has previously suggested bi-directional activity, with strong strand bias, of Pol II initiation sites in *T. brucei* [68, 69]. The RNA-seq data presented here is consistent with bi-directional activity of TSS, early termination and stimulation of divergent antisense transcription in the PNUTS knockdown leading to the activation of genes within the dSSR (Fig A.4C and A.5A and A.S10). As shown in Fig A.5A and A.5B, the silent 8340 gene within a dSSR on chromosome 10 is transcribed upon the loss of TbPNUTS. To see whether there is a correlation between the initiation of the antisense transcript with sense mRNA coding strand TSS, we analyzed the 5'-end of the nascent antisense transcript at this dSSR using RT-PCR (Fig A.5C). The significant drop of inability of a 5' primer to amplify the cDNA corresponding to the nascent 'antisense' transcript indicates the 5' end of the PNUTS regulated

transcript is adjacent to the TSS for the sense transcript on the opposing strand. Using various 3' primers for generating cDNA, antisense transcription is attenuated in uninduced cells as shown in the decline in cDNA with increasing length of the transcript (Fig A.5D). In the absence of PNUTS the antisense transcription fails to terminate, as seen in the maintained levels of cDNA for all primers tested in + Tet. The increasing effects of the loss of PNUTS on level of transcript with increasing length of the transcript supports the idea of PNUTS regulating early termination/elongation of the antisense transcript. Taken together the results suggest that the PJW complex (PNUTS, JBP3 and Wdr82) regulates premature termination of antisense transcription from bi-directional TSS and silencing of gene expression in *T. brucei*.

Another possibility is that increased transcription within dSSRs is due to transcription initiating upstream of the divergent TSSs in the absence of PNUTS. To address this possibility, we examined transcription at Head-to-Head (HT) boundaries of PTUs. HT sites are defined where transcription of one PTU terminates and transcription of another PTU on the same DNA strand initiates. In contrast to TSSs at dSSRs, HT sites contain a termination site for the upstream PTU, indicated by H3.V and H4.V, and a TSS marked by individual peaks of histone variants, such as H2A.Z, and histone modifications [25, 68, 69]. Similar to the metaplot analyses of dSSRs and cSSRs, we detected accumulation of transcripts at HT sites upon the ablation of PNUTS (Fig A.4D). At these non-divergent TSSs, the region upstream of the start site produced more transcripts in the PNUTS mutant indicating readthrough transcription, as expected, or possible initiating events further upstream. More interestingly, antisense transcripts are also increased at HT sites. The lack of a divergent TSS at these sites, indicated by the presence of single peak of H2A.Z, supports the conclusion that the antisense transcripts are the result of regulated bi-directional transcription activity.

Replication is not affected by pervasive transcription

Noncoding transcription, via defects in transcription termination, influences eukaryotic replication initiation. Transcription through origins located at 5'- and 3'-ends of Pol II transcription units leads to replication defects via dissociation of the prereplication complex (pre-ORCs) or sliding of MCM helicases [70–74]. Of the 40 early firing origins that have been mapped in *T. brucei*, 36 are upstream of TSS [75]. Analysis of *L. major* has indicated that replication initiation sites occur at the genomic locations where Pol II stalls or terminates, including sites precisely downstream of base J [76]. Therefore, increased transcription at PTU flanking regions in the absence of TbPNUTS may cause DNA replication defects. To see whether pervasive transcription has any effect on *T. brucei* replication, we analyzed whether TbPNUTS is required for proper cell-cycle progression. In *T. brucei*, DNA replication and segregation of kinetoplastid DNA (K) in the single mitochondrion precede those of nuclear DNA (N), so cells at different stages can be distinguished by their N and K configurations. 1N1K content indicates that cells are in G1, 1N2K indicates cells in S phase and 2N2K indicate post-mitotic cells. Representatives of cells with these DNA contents upon DAPI staining are shown in Fig A.S11A. We detect no change in cell populations following a 2 day induction of PNUTS RNAi ablation. To confirm the lack of cell cycle defects we monitored cell-cycle progression after RNAi ablation by flow cytometry, staining bulk DNA with propidium iodide. As shown in Fig A.S11B, uninduced cells show normal cell cycle profiles. Two days after RNAi ablation, there is no change in the cell-cycle profile or quantities of cells at each stage. The cell cycle profiles of the conditional RNAi ablation suggest that TbPNUTS is not required for proper cell-cycle progression and DNA replication. Furthermore, the increase in pervasive transcription in the PNUTS mutant has no measurable effect on DNA replication.

PNUTS regulates Pol I transcriptional repression of telomeric PTUs

In addition to Pol II termination sites distributed throughout the *T. brucei* genome, H3.V and base J localize within the ~14 Pol I transcribed polycistronic units located at the telomeric ends and involved in antigenic variation (so called bloodstream form VSG expression sites, BESs) (Fig A.6A)[22, 77–79]. Monoallelic expression of a VSG ES leads to the expression of a single VSG on the surface of the parasite, a key aspect of the strategy bloodstream form trypanosomes use to evade the host immune system. We have previously shown that the loss of H3.V and J leads to increased expression of VSG genes from silent telomeric BESs [24]. This effect is presumably due to the role of these epigenetic marks in attenuating transcription elongation Pol I within the silent VSG BESs, thereby preventing the transcription of silent VSGs. Differential gene expression analysis of RNA-seq reads mapped to the 427 genome indicates that the loss of PNUTS leads to increased VSG expression from silent BESs (S2 Table). In addition to the BES, depletion of TbPNUTS results in de-repression of VSGs from the silent telomeric metacyclic ES (MES), which are transcribed monocistronically by Pol I. A few of these VSG gene expression changes have been confirmed by RT-qPCR (Fig A.S12). To further explore the global function of TbPNUTS in VSG expression control, we mapped the RNA-seq reads to the VSGnome [80]. The VSGnome allows the analysis of VSG genes, such as those on minichromosomes, that were not included in the new *T. brucei* 427 genome assembly. As shown in Fig A.6, this analysis confirms the de-repression of Pol I transcribed VSG at BES and MES upon TbPNUTS depletion. On the other hand minichromosomal (MC) VSGs lacking a promoter [81] were not significantly affected; only 2 of 41 were upregulated ($P_{adj}<0.1$). These data indicate that transcription of (silent) telomeric VSGs in the PNUTS mutant is strongly dependent on the presence of a Pol I promoter. Interestingly, comparing the expression level of MES VSGs that

are adjacent to the promoter with BES VSGs over 10 kb downstream (Fig A.6A and Fig A.S12) suggested that the level of derepression is a function of distance from the Pol I promoter. As previously mentioned, the majority of VSG genes upregulated in the PNUTS mutant are chromosomal internal VSGs (S4 Table and A.S7 Fig). In the VSGnome, these (unknown) VSGs (Fig A.6B) were thought to be primarily located at subtelomeric arrays, but their exact positioning in the genome was not known [80]. These VSGs have now been mapped to the silent subtelomeric arrays assembled in the new 427 genome [79] and, as shown in A.S7 Fig and Fig A.6B, a significant fraction are upregulated upon the loss of PNUTS. Overall, these data indicate that TbPNUTS (PJW complex) regulates transcriptional silencing of telomeric Pol I transcribed telomeric VSG PTUs and Pol II transcription of VSGs within genome internal PTUs in *T. brucei*.

The derepression of ESAG 6 and 7 genes adjacent to the BES promoter along with the VSG ~40 kb downstream (S4 Table), suggests that PJW may function throughout the telomeric PTU. To examine this more closely, we analyzed the RNA-seq reads mapping to the 14 telomeric BES sequences [82]. RNA-seq reads mapping to BES were counted in 200bp windows with a 100bp steps. Read counts were converted into reads per million (RPM) and compared between +/- Tet to estimate log2 fold change and plotted in Fig A.6C and A.S13 Fig. The location of BES promoters is indicated by an arrow. The transcription of the active BES1 was not affected by the loss of TbPNUTS. However, when the remaining 13 silent BESs are analyzed we see derepression of ESAGs as well as the terminal VSG. In some cases, it seems that derepression extends 10–20 kb from the promoter to express ESAG 6 and 7, with no significant effect on the remaining genes (ESAGs) within the silent BES, and derepression of the VSG at the 3’ end. In other cases there is selective upregulation of the VSG gene, including the pseudo VSGs that are present upstream of some telomeric VSGs. The apparent selective VSG

upregulation may be due to the combined effect of the low level of transcription of the derepressed BES and enhanced stability of VSG mRNAs. For example, the increased VSG mRNA half-life (4.5 hrs) compared with ESAG 6 and 7 (1.8–2.8 hrs) [83, 84]. Transcripts of genes located approximately 5 kb upstream of the VSG have also been shown to be selectively rapidly degraded [85]. We also noticed that increased levels of RNA close to the promoter are significantly higher when there is an additional BES promoter upstream for ESAG10. In fact, the most significant gene expression changes are in BESs that have dual promoters (S4 Table). These data would suggest derepression of telomere-proximal VSG genes after PNUTS depletion is due to transcriptional activation of silent Pol I promoters. However, these results are also consistent with increased Pol I elongation along the BES. Repression of silent ESs is mediated at least in part by the inhibition of Pol I elongation within the BES preventing the production of VSG mRNA from the silent BESs [86–88]. Similar to its inhibition of Pol II transcription elongation at termination sites at the 5' and 3' end of PTUs genome-wide, PNUTS may also function at telomeric regions to attenuate Pol I transcription elongation within the silent ESs. The data here suggest PJW controls VSG silencing at BESs by regulating Pol I elongation (termination) and/or regulating access of the polymerase to silent promoter regions.

DISCUSSION

Current dogma in the field is that most, if not all, Pol II transcribed gene regulation is at the posttranscriptional level in kinetoplastids [89]. This is primarily based on the polycistronic arrangement of genes and identification of posttranscriptional regulons under the control of RNA binding protein regulated mRNA stability. Our previous studies on base J/H3.V function revealed regulated transcription termination at the 3'-end of PTUs as a novel mechanism of gene

silencing in kinetoplastids. The studies described here solidify this concept and open up another possible regulatory gene expression mechanism via early termination at the 5'-end of PTUs. These studies also provide the first direct mechanistic link between base J and transcription termination in kinetoplastids by the identification of a multi-subunit protein complex involved in termination that binds base J.

Here we identify a new base J binding protein, JBP3, and show that it is part of a PP1-PNUTS-Wdr82 containing protein module. We named this module the PJW/PP1 module based upon the mammalian PTW/PP1 module involved in transcription termination [9]. Mutation of its components, JBP3, Wdr82 and PNUTS, gives similar phenotypes in *T. brucei*, validating PJW is indeed a functional module. Our data strongly suggests that the kinetoplastid PJW/PP1 complex we identified here is reminiscent of human PTW/PP1 where PNUTS is the kinetoplastid functional homologue of human PNUTS and JBP3 the homologue of Tox4. However, while we demonstrate LtPP1 is a component of the Leishmania complex, the purified module in *T. brucei* lacks a PP1 homologue. PP1 is the only catalytic component of the human PTW/PP1 complex and dephosporylation by PP1 is directly involved in regulating Pol II termination [2, 3, 60, 61]. Therefore, the apparent lack of PP1 association with the PNUTS, Wdr82 and JBP3 complex involved in transcription termination in *T. brucei* is surprising. MS analysis of the TbPNUTS complex is the result of a single purification from 12 L of PC *T. brucei* cells. The finding that one of the four PJW/PP1 complex purifications in Leishmania also failed to pulldown PP1 suggests PP1 may be the least stable component of the complex in *T. brucei*. Once we increase the yield of TbPNUTS purification, and reduce the volume of cells needed, we can perform multiple pulldown/MS analyses to address this possibility. The specific loss of PP1 and JBP3 in one of the LtJGT pulldowns, may also reflect an effect of the tandem affinity tag on

LtJGT associations. Thus, future experiments should examine *T. brucei* complex purification using a different affinity tag or using TbJBP3 or TbWdr82 PTP-fusion proteins. We propose that PP1 is a functional component of the PJW/PP1 complex in both Leishmania and *T. brucei*, but association with TbPNUTS is less stable. However, further work needs to be done to determine whether PP1 is a component of the PTW complex and involved in termination in *T. brucei*. Therefore, we refer to the PJW complex when addressing functional analysis of the complex in *T. brucei* and, based on our studies in Leishmania, the PJW/PP1 complex when discussing the potential role of the complex in termination in kinetoplastids.

The association of JGT, the glucosyltransferase involved in base J synthesis, with the PJW/PP1 complex in Leishmania is also unexpected. Further work is needed to clarify the function of JGT in the Leishmania PJW/PP1 complex, for example, in regulating base J synthesis. Reciprocal tagging of JBP3 and Wdr82 pulls down the same PJW/PP1 components in Leishmania, including JGT. This helps confirm the nature of the PJW/PP1 complex and suggests that the tag itself does not significantly influence the protein interactions. As mentioned above, we have not directly addressed the possibility that the lack of PP1 in the TbPNUTS-PTP purification could be due to the PTP-tag. However, the lack of JGT in the *T. brucei* PJW complex is likely not due to the PTP-tag since co-IP analysis using tagged versions of PNUTS, JBP3 or Wdr82 fail to pull down JGT in BSF parasites. Furthermore, ablation of PNUTS, JBP3 and Wdr82 in BSF *T. brucei* has no significant effect on level of base J in the genome (A.S14 Fig). Not only does this support the lack of JGT of in the *T. brucei* PJW complex, but also addresses the possibility that termination defects we measure in the PJW mutants are due to the loss of base J.

We and others have previously shown that base J and H3.V co-localize at Pol II termination sites [23, 25] and are involved in transcription termination in *T. brucei* and *L. major*, where loss of base J and H3.V leads to readthrough transcription at the 3'-end of PTUs [24, 26–29]. For 'premature' termination sites within PTUs, read-through transcription resulting from the loss of base J and/or H3.V leads to transcription of silent genes downstream of the termination site [24, 27, 28]. We now show that depletion of components of the PJW complex in *T. brucei* leads to similar defects in Pol II termination at the 3'-end of PTUs, including the de-repression of downstream genes. We also uncover additional defects at the 5'-end of PTUs suggesting regulated early termination of antisense transcription from bi-directional transcription start sites. Genes in *T. brucei* have been shown to exhibit PTU positioning bias, where genes located near the 3'-end of a PTU exhibit relative increase in relative abundance of mRNAs upon heat shock [90]. To control for the stress of dying cells, we analyzed the ablation of the VA protein that results in a similar growth defect as the ablation of PJW complex components. In contrast to the 709 genes upregulated in the PNUTS RNAi, VA ablation led to the increased expression of only four genes and no significant changes in read-through at 3'-ends of PTUs (Fig A.4B). However, a small increase in RNA from dSSRs was evident in the VA RNAi (Fig A.4C). Interestingly, this significantly smaller effect corresponds with a slightly reduced growth defect in the VA RNAi compared to the PNUTS RNAi. We are unable to rule out the possibility that some of the increased RNAs we detect in these studies are in response to stress. For example, some treatments that inhibit *T. brucei* growth (such as cold shock and mild acid) have resulted in increased expression of procyclin [91, 92], a gene that is upregulated in both the PNUTS and VA mutant. However, the apparent increase in RNAs from dSSRs of the VA control RNAi does not lead to any gene expression changes, since none of the four upregulated genes localize to this

region of the genome. As we previously described following the loss of base J and H3.V in *T. brucei* and *Leishmania*, detection of increased nascent, unprocessed RNA by RT-PCR at the termination site supports the conclusion that the loss of PJW complex function leads to Pol II read-through transcription, rather than some Pol I or Pol II (re-) initiation event. The lack of identifiable Pol I promoter at these regions, and the low level of resulting transcripts where RNA Pol I transcription is known to be 5–10 times more active than transcription by RNA Pol II [93], supports this conclusion. Although we cannot exclude a potential role of the PJW complex in the regulation of RNA processing, the increase in both unprocessed (represented by nascent RNA spanning the termination site and RNA-seq reads corresponding to intergenic regions) and processed RNAs (mRNAs) strongly suggests the complex regulates RNA abundance at the level of transcription and that the defects we observe are not simply due to an alteration of RNA processing. We conclude that the gene expression changes we see in the PNUTS mutant reflect the role of the PJW complex in Pol II transcription termination in *T. brucei*.

Here, we propose a model where divergent transcription at the 5'-end and readthrough transcription at the 3'-end of gene arrays is affected by the PJW/PP1 complex in kinetoplastids (model in Fig A.7). According to this model, the PJW/PP1 complex is recruited to termination sites, at least partially, due to base J-JBP3 interactions. H3.V localized at 3'-end termination sites may play an additional role in complex localization since WDR5, a homolog of Wdr82, binds to the N-terminal tail of histone H3 [94]. Wdr82 is required for recruitment of the APT termination complex containing PNUTS-PP1 to snoRNA termination sites in yeast [95]. Wdr82 may also play a role in 5'-end termination site recognition since it binds to RNA Pol II CTD phosphorylated at Ser5 in yeast and humans [10, 96].

Similar to the mammalian complex, we propose PNUTS is a scaffolding protein for the entire PJW/PP1 complex and regulates PP1 function via the PP1 binding RVxF motif. Only three substrates have been identified for PNUTS/PP1: the Pol II elongation factor Spt5, the CTD of Pol II and MYC [2, 60, 97, 98]. MYC dephosphorylation by PP1 regulates chromatin binding and stability. PP1 dephosphorylation of Spt5 and Pol II has been directly implicated in regulating Pol II termination in other eukaryotes. Therefore, we propose that regulated phosphorylation of Spt5 and Pol II by the PJW/PP1 complex is critical for transcription termination in trypanosomatids. TbSpt5 has recently been shown to be associated with Pol II [99] and is phosphorylated at a single Ser residue [100]. The CTD of Pol II in trypanosomatids is unique in that it does not contain the heptad or other repetitive motifs that are conserved from yeast to humans [101]. However, the CTD was shown to be essential for Pol II transcription in *T. brucei* [102, 103] and 17 phosphorylated sites have been identified within the CTD [100]. Studies also suggest that CTD phosphorylation is required for Pol II association with trypanosome chromatin [104]. It is necessary to determine the role of the PJW/PP1 complex in the phosphorylation status of Spt5 and Pol II in trypanosomatids and whether phosphorylation of a non-canonical Pol II CTD is involved in transcriptional regulation in divergent eukaryotes thought to lack significant regulation.

An unexpected finding of this work was that, in the absence of PNUTS, Wdr82 or JBP3, genes located upstream of TSS are also now expressed, resulting from transcription between diverging PTUs. Mammalian and yeast promoters frequently give rise to transcription in both the sense and divergent antisense directions [63–67], giving rise to a (productive) sense transcript and a corresponding upstream noncoding RNA (ncRNA) [105–108]. Unidirectional transcription is typically ensured since the ncRNAs are susceptible to early termination linked to rapid

degradation [106, 109–111]. Early termination of divergent transcription at 5’ ends of mammalian genes occurs by similar mechanisms as termination at 3’ ends. In addition to regulating termination at the 3’ end of genes, Wdr82 and PNUTS have also been shown to be involved in enforcing early transcription termination at bi-directional promoters [17]. The role of PNUTS is thought to reflect the essential nature of PNUTS/PP1 since differential phosphorylation of the CTD of Pol II has been proposed to regulate the directionality of transcription at bi-directional promoters. Specifically, Tyr1 and Ser1 hyperphosphorylation of Pol II have been shown to be associated with antisense divergent transcription at mammalian promoters [112–114]. Spt5 regulation of Pol II elongation is involved in control of divergent antisense transcription as well as readthrough transcription at 3’ end of genes in yeast [2, 115], representing an additional target of PNUTS/PP1 regulation of transcription at both ends of genes. Similar to mammalian promoters where transcription is divergent and initiation is over a broad genomic region, previous studies have suggested that Pol II transcription initiation sites are intrinsically bi-directional in *T. brucei* [68, 69]. We found that loss of PNUTS, Wdr82 and JBP3 can have a major effect on divergent ‘antisense’ transcription, presumably reflecting the role of PJW in regulating termination of antisense transcription from bi-directional promoters in *T. brucei* (Fig A.7). Interestingly, decreased levels of base J in Leishmania led to detection of antisense RNAs corresponding to similar regions within divergent PTUs [26]. A possible explanation for antisense transcription at divergent TSSs is increased chromatin accessibility in the SSR, resulting in alternative TSS usage further upstream. However, the detection of significant antisense transcription at non-divergent TSSs (HT sites) upon the loss of PNUTS strongly supports the involvement of the PJW in regulating early termination of antisense transcription in *T. brucei*. Additional work is required to unambiguously confirm that the

increased nascent antisense RNA, and corresponding mRNAs, we detect from regions upstream of TSSs are in fact a result of bi-directional transcription activity.

Loss of J and H3.V in *T. brucei* resulted in similar readthrough defects at the 3'-end of PTUs (including telomeric BESs) and gene expression as seen here in PJW mutants, but without cell growth effects. We concluded that readthrough transcription at 3'-end and corresponding gene expression changes in the J/H3.V mutant are not lethal to the parasite. PJW mutants we analyze here lead to additional effects on transcription and gene de-repression at 5'-ends and decreased cell growth. The ability of the PJW complex to bind base J is consistent with its function at both ends of PTU's. However, it is unclear why the effects in the J mutant are limited to the 3'-end and whether specific function of the complex at the 5'-end can explain its essential nature. While J is present at both 5' and 3' PTU flanking regions and involved in transcription, it is apparently not the dominant mark since H3.V had more significant effects on 3'-end transcription and gene expression [24]. H3.V is limited to termination sites at the 3'ends of PTUs and not localized at TSSs in *T. brucei*. A significant role of H3.V in PJW complex function, as discussed above, would explain why H3.V and PJW mutants share termination defects at the 3'-end. Similarly, additional specific chromatin factors may be involved in recruitment of PJW complex to TSSs independent of base J, such as Pol II-Wdr82 interactions, as mentioned above, and modified and variant histones such as H2A.Z and H3K4Me3 (Fig A.7). These points may explain why PJW mutants lead to defects at both 5'- and 3'-ends and the J/H3.V mutants are limited to the 3'-end. We propose that the essential nature of the PJW complex is due to regulated expression of ncRNAs and/or mRNAs at the 5'-end of PTUs or additional unknown functions of JBP3, Wdr82 and PNUTS in *T. brucei*.

The possibility that the essential nature of the complex is due to regulation of transcription-replication conflicts at TSSs was directly addressed. We expected the pervasive transcription phenotype in the PJW/PP1 mutant to negatively impact DNA replication and explain the reduced cell growth. Replication origins tend to localize after TTS in yeast and upstream of promoters in humans, presumably to minimize transcription-replication conflicts [73, 116, 117]. The induction of transcription through origins, via defects in transcription termination at TTSs as well as at TSSs, leads to replication defects via dissociation or sliding of the pre-ORCs and MCMs [70–74] and changes in chromatin structure [118]. Furthermore, loss of origin function activates readthrough transcription in mammals and yeast. Clearly there is a relationship between transcription and DNA replication in eukaryotes. This relationship appears to exist in *T. brucei* as well, since origins flank the PTUs at TSS and TTS, and TbORC1 and TbMCM-BP deletions led to similar defects in transcription at PTU flanks as we illustrate here [75, 119]. In the recent analysis of the TbMCM-BP mutant [119], in addition to its role in termination at 3' ends, the antisense RNA at H-T sites suggested a role of TbMCM-BP in determining the direction of transcription, what we refer to as bi-directional activity, while at dSSRs they concluded it was due to alternative initiation events. Regardless of the mechanism involved, increased transcription upstream of TSSs in the PJW/PP1 mutants would presumably result in defects in replication. However, the TbPNUTS mutant does not indicate any alteration in cell cycle or DNA replication. Alternatively, the level of transcription induction at origins is too low at day two of the RNAi for any significant effects on replication. Further work is needed to explore the effects of increased transcription on ORC function in trypanosomes.

Functional interaction between replication and transcription machineries was further suggested by derepression of Pol I transcribed silent BESs and MESs in the TbORC1 and

TbMCM-BP mutants [75, 120]. How these operate mechanistically at ESs, regulating Pol I elongation via chromatin changes along silent BESs or BES promoter activity, remains unclear. In fact, whether VSG monoallelic expression control of BESs takes place at the initiation or elongation level is still debated. Low levels of transcripts from silent ESs upon the knockdown of factors such as ORC1, MCM-BP and PNUTS cannot resolve this issue, since the data are compatible with both models. While derepression of the BES in the PNUTS mutant would suggest the PJW complex has a direct effect on the activity of silent BES promoters, the proposed role of the complex in regulating Pol II elongation/termination at 5' and 3' end of PTUs genome-wide would suggest a similar role in regulating Pol I elongation. If so, presumably Pol I is regulated in a different manner, since there is no evidence for phosphorylation of Pol I. However, the Spt5 substrate for PNUTS-PP1 has been shown to bind and regulate Pol I transcription in mammalian cells [121, 122]. Further studies are necessary to understand how PJW complex regulation of Pol II transcription at 5' and 3' ends of PTUs genome-wide is related to silencing of the specialized telomeric Pol I PTUs. For example, derepression of BES promoters may be a functional consequence of non-coding RNA transcription generated by pervasive Pol II transcription in the PNUTS mutant. We have previously shown that readthrough transcription at 3' ends of PTUs lead to significant levels of siRNAs in *T. brucei* [24]. Regardless of the mechanism involved, the complex is required not only for repression of telomeric and subtelomeric VSGs but also VSGs scattered within the chromosome at 3' ends of PTUs. Depletion of PNUTS also increased the levels of procyclin and PAG RNAs, which are transcribed from a Pol I promoter that is repressed in BSF *T. brucei*. This transcription unit is located at the end of Pol II transcribed PTU and increased transcription in PNUTS mutant may be due to readthrough transcription as in the J/H3.V mutant [24]. Thus, the PJW complex is

required for repression of life-cycle specific genes transcribed by Pol I in the mammalian infectious form of *T. brucei*. We have therefore uncovered a possible functional link between transcriptional termination and Pol II- and Pol I-mediated gene silencing in *T. brucei*.

MATERIALS AND METHODS

Parasite cell culture

Bloodstream form *T. b. brucei* Lister 427 (MiTat 1.2) or "single marker cells (SMC)", expressing T7 RNA polymerase and the Tet repressor [123] were used in these studies and cultured in HMI-9 medium. Transfections were performed using the Amaxa electroporation system (Human T Cell Nucleofactor Kit, program X-001). Where appropriate, the following drug concentrations were used: 2 μg/ml G418, 5 μg/ml Hygromycin, 2.5 μg/ml Phleomycin, 2 μg/ml Tetracycline. Procyclic form *T. b. brucei* TREU667 and promastigote form *L. tarantolae* were cultured in SDM79 medium. Transfections were performed using the BioRad GenePulser II (2 pulses at 1.4 kV / 25 μF) in 0.4 cm cuvettes with 1 x 10^8 cells (*L. tarantolae*) or 3 x 10^7 cells (PCF *T. b. brucei*) in 0.5 ml cytomix (2 mM EGTA, 120 mM KCl, 0.15 mM $CaCl_2$, 10 mM KP_i pH = 7.6, 25 mM HEPES pH = 7.6, 5 mM $MgCl_2$, 0.5% glucose, 100 μg/ml BSA, 1 mM Hypoxanthine). Where appropriate, the following drug concentrations were used: 25 μg/ml G418 (PCF *T. b. brucei*) and 50 μg/ml G418, 10 μg/ml Puromycin (*L. tarantolae*).

RNAi analysis

For conditional PNUTS, JBP3 and Wdr82 silencing experiments in *T. brucei* a part of the ORF was integrated into the BamH I site of the p2T7-177 vector [124]. Sce-I linearized p2T7-177 constructs were transfected into BF SMC for targeted integration into the 177bp repeat

locus. RNAi was induced with 2 μg/ml Tetracycline and cell growth was monitored daily in triplicate. Primers sequences used are available upon request.

Epitope tagging of proteins

For generation of the dual (Protein A and Streptavidin Binding Protein) tagged constructs in *L. tarantolae*, the coding regions of LtJGT, LtJBP3, and LtWdr82 lacking stop codons were amplified and cloned into the BamHI and XbaI sites of pSNSAP1 [45]. The resulting constructs are referred to as JGT-SA, JBP3-SA and Wdr82-SA. In Lt. cells ectopically expressing the JGT-S, we performed an additional KO for a single JGT WT allele. For PTP tagging in *T. brucei*, the 3' end of *T. brucei* genes were cloned in the Apa I and Not I sites of the pC-PTP-Neo vector [55], resulting in dual (Protein A and Protein C) tagged proteins. Linearization of the constructs was performed using a unique restriction site within the 3' end of the cloned gene. All final constructs were sequenced prior to electroporation. Tagging the 3'-end of the TbPNUTS, TbJBP3 and TbWdr82 with 3x HA tag was performed using a PCR based approach with the pMOTag4H construct [125]. Primers sequences and construct information used are available upon request.

Determination of the genomic level of J

To quantify the genomic J levels, DNA was isolated and utilized in the anti-J DNA immunoblot assay as described previously [126]. Briefly, serially diluted genomic DNA was blotted to nitrocellulose, followed by incubation with anti-J antisera. Bound antibodies were detected by a secondary goat anti-rabbit antibody conjugated to HRP and visualized by ECL. The membrane was stripped and hybridized with a probe for the beta-tubulin gene to correct for DNA loading.

Tandem affinity purification (TAP) and co-immunoprecipitation

Tandem affinity purification was performed from whole cell extracts. 2 x 10^{11} cells (*L. tarantolae* and PCF *T. brucei*) were harvested, 1 time washed in 1 x PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 2 mM KH_2PO_4 pH = 7.4), 2 times washed in buffer I (20 mM Tris pH = 7.7, 100 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA) and 1 time washed in buffer II (150 mM Sucrose, 150 mM KCl, 3 mM $MgCl_2$, 20 mM K-glutamate, 20 mM HEPES pH = 7.7, 1 mM DTT, 0.2% Tween-20). Cell pellets were then adjusted to 20 ml (*L. tarentolae*) or 50 ml (PCF *T. brucei*) with buffer II with protease inhibitors (8 µg/ml Aprotinin, 10 µg/ml Leupeptin, 1 µg/ml Pepstatin, 1 mM PMSF and 2 tablets cOmplete Mini, EDTA free; Roche) and 2 times flash frozen in liquid Nitrogen. Lysates were then sonicated (Sonics, Vibra-Cell) for 5 times (15" on / 45" off, 50% amplitude, large tip) on ice. Extracts were cleared by centrifugation for 10 min at 21,000 x g at 4°C and incubated while rotating with 200 µl IgG Sepharose beads (GE Healthcare) for 4 hrs at 4°C. The beads were then washed with 35 ml PA-150 buffer (20 mM Tris pH = 7.7, 150 mM KCl, 3 mM $MgCl_2$, 0.5 mM DTT, 0.1% Tween-20) and 15 ml PA-150 buffer with 0.5 mM EDTA. The beads were then resuspended in 2 ml PA-150 buffer with 0.5 mM EDTA and 250 U TEV Protease (Invitrogen) while rotating for 16 hrs at 4°C. The supernatant was collected and for *L. tarantolae*, samples were incubated with 100 µl magnetic Streptavidin C1 Dynabeads (Invitrogen) for 4 hrs while rotating at 4°C. Beads were then washed with 50 ml PA-150 buffer, transferred into 1 ml elution buffer (100 mM Tris pH = 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM *d*-desthiobiotin), incubated for 2 hrs while rotating at room temperature. Eluted protein was then TCA precipitated and subjected to MS/MS analysis. To the *T. brucei* supernatant samples, $CaCl_2$ was added to a final concentration of 1.25 mM and incubated with 200 µl Anti-Protein C Affinity Matrix (Roche) for 4 hrs while rotating at 4°C.

Beads were then washed with 50 ml PA-150 buffer with 1.25 mM CaCl2. Bound protein was eluted in 5 ml elution buffer (5 mM Tris pH = 8.0, 5 mM EDTA, 10 mM EGTA), TCA precipitated and subjected to MS/MS analysis.

Pull-down experiments using PTP-tagged PNUTS, JBP3 and Wdr82 were performed using the first purification step of the TAP protocol on whole cell extracts from 2 x 10^8 cells in 200 µl buffer II. After the IgG Sepharose incubation, beads where washed in PA-150 buffer and boiled in Laemmli buffer. For Western analysis, 5 x 10^6 cell equivalents of each fraction (input, unbound and bound) were analyzed on 10% PAA / SDS gels and sequentially probed with anti-HA antibodies (Sigma, 3F10, 1:3000), anti-Protein A antibodies (Sigma, P3775, 1:5000) and anti-La antibodies (a gift from C. Tschudi, 1:500).

MS-MS analysis

For the initial JGT-S purification, purified proteins from JGT tagged cells and negative control cells were separated by SDS-PAGE gel and visualized by silver staining. Each lane of the silver-stained SDS-PAGE gel was divided into 5 large fractions before being cut down further into roughly 2mm x 2mm pieces and stored in separate tubes in preparation for digestion. Each fraction of gel was de-stained before undergoing denaturation in 10 mM dithiothreitol at 56°C for one hour. Denatured proteins were then alkylated by 55mM iodoacetamide for 45 minutes in the dark with intermittent vortexing. Sequencing grade trypsin (Promega, V5111) was then utilized to digest proteins overnight at 37°C with gentle vortexing. Peptides were extracted from each fraction separately by incubating gel pieces with a gradient of increasing concentrations of acetonitrile. All extracted peptides were concentrated, reconstituted in 0.1% formic acid in 5% acetonitrile, and passed through a 0.2um bio-inert membrane tube (PALL, ODM02C35) to remove any residual particulate. Each fraction was injected independently into an Orbitrap

Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) interfaced with an UltiMate 3000 RSLCnano HPLC system (Thermo Scientific). Peptides were resolved on an Acclai PepMap RSLC C18 analytical column (75um ID x 15cm; 2µm particle size) at a flow rate of 200nL/min using a gradient of increasing buffer B (80% acetonitrile in 0.1% formic acid) over 180 minutes. Data dependent acquisition was carried out using the Orbitrap mass analyzer collecting full scans every three seconds (300–2000 m/z range at 60,000 resolution). The most intense ions that met mono-isotopic precursor selection requirements were selected, isolated, and fragmented using 38% collision-induced dissociation (CID). Every precursor selected for ms/ms analysis was added to the dynamic exclusion list and precursors selected twice within 10 seconds were excluded for the following 20 seconds. Fragment ions were detected using the ion trap to increase the duty cycle and achieve more ms/ms scans per experiment. Raw data for each fraction was searched separately against the appropriate databases: GeneDB *Leishmania tarentolae* and T. *brucei* 927 database, using the Sequest search algorithm in Proteome Discoverer 2.2 (Thermo Fisher Scientific). Search parameters were set to allow for two missed tryptic cleavages, 20 ppm mass tolerance for precursor ions, and 0.3 Daltons mass tolerance for fragment ions. A fixed modification of carbamidomethylation on cysteine residues and a variable modification of oxidation on methionine residues were enabled to accurately match fragment ions. A fixed value peptide spectral match (PSM) validation node was used to validate each PSM at a maximum delta Cn of 0.05. True negative PSMs (decoy PSMs) were generated by searching the same raw data utilizing a decoy database, containing reversed protein sequences from the target protein database, and the exact same search parameters. All surviving PSMs were imported into ProteoIQ (v2.7, Premier Biosoft) for final validation. ProteoIQ used validated PSMs and decoy PSMs to generate protein identifications with a maximum false discovery rate

of 1%. Each protein present at 1% FDR was also required to have at least two unique peptides and 10 PSMs to be reported. For shotgun proteomics, eluted proteins were digested in-solution as previously described [127] and then analyzed by LC-MS/MS as described above.

3D structure prediction of JBP3

To analyze the putative J-DNA binding domain of JBP3, the representative region of LtJBP3 was threaded through the JBP1 JBD PDB structure to search for similar secondary structural folds using the comparative modeling program I-TASSER (iterative threading assembly refinement algorithm). (http://zhang.bioinformatics.ku.edu/I-TASSER) The initial 3D models generated by I-TASSER were of high quality, with a C-score of -0.76 and a TM-score of 0.62 (within the correct topology range of I-TASSER). TM-score is defined to assess the topological similarity of the two protein structures independent of size. A TM-score >0.5 corresponds approximately to two structures of the similar topology. The top predicted model was then aligned using TM-align from I-TASSER with the Lt JBD model 2xseA giving a TM-score of 0.727, RMSD of 1.46, and Cov score of 0.774.

Recombinant protein production

The *L. tarentolae* JBP3 gene was amplified from genomic DNA by PCR with primers containing 5' Nde I and 3' BamH I restriction sites. PCR fragments were digested with Nde I and BamH I and subcloned into the pet16b expression vector. Plasmids were transformed into Escherichia coli (BL21 DE3), and bacteria were cultured in defined autoinduction media to allow growing cultures to high densities and protein expression without induction, as previously described [128]. Briefly, LB media is supplemented with KH_2PO_4, Na_2HPO_4 and 0.05% Glucose, 0.2% Lactose and 0.6% Glycerol. Bacteria were cultured in this media in the presence of 100µg/ml ampicillin at 37°C for 2 h and then shifted to 18°C for 24 h. Cells were lysed and

JBP3 purified by metal affinity as previously described for JBP1 [40, 41]. The affinity-purified JBP3 was concentrated to 0.5 ml in a Centricon-100 apparatus and loaded onto a Sephadex S-200 (Amersham Biosciences 16/60) column equilibrated with buffer A (50 mM Hepes, pH 7.0, 500 mM NaCl, 1 mM DTT). The fraction containing JBP3 was concentrated to 200 ul by Centricon-100. Protein purity was analyzed by SDS-PAGE stained with Coomassie Brilliant Blue, and protein concentration was determined using BSA standards.

J-DNA binding

Electrophoretic Mobility Shift Assays were carried out essentially as described previously for JBP1 and VSG-1J DNA substrate [40, 41], with few changes. The binding reaction mixture (20 ul) contained 35 mM Hepes-NaOH, pH 7.9, 1mM EDTA, 1 mM DTT, 50 mM KCL, 5 mM MgCl, 10 ug BSA and 15 fmol radiolabeled DNA substrate containing a single modified J base (VSG-1J) or no modified base, and indicated JBP3 protein amounts. The reactions were incubated for 15 min at room temperature and analyzed on a 4.5% nondenaturing polyacrylamide gel at 4°C. After drying, the gels were exposed to film.

Sucrose sedimentation analysis

For the sedimentation analysis of the PJW/PP1 complex, extracts were made from the BF *T. brucei* cell lines expressing PNUTS-PTP and JBP3-HA or Wdr82-HA and loaded onto 10 ml, 5–50% sucrose gradients. Samples were centrifuged at 38000 rpm for 18 hours using a SW41 Ti rotor (Beckman). The gradient was fractionated from top to bottom in twenty aliquots of 500ul each. Proteins from each fraction were enriched by methanol: chloroform protein precipitation and resuspended in SDS loading buffer for electrophoresis.

RT-PCR analysis

Total RNA was isolated with Tripure Isolation Reagent (Roche). cDNA was generated from 0.5–2 μg Turbo DNase (ThermoFisher) treated total RNA with Superscript III (ThermoFisher) according to the manufacturer's instructions with either random hexamers, oligo dT primers or strand specific oligonucleotides. Strand specific RT reactions were performed with the strand specific oligonucleotide and an antisense Asf I oligonucleotide. Equal amounts of cDNA were used in PCR reactions with Ready Go Taq Polymerase (Promega). A minus-RT control was used to ensure no contaminating genomic DNA was amplified. Primer sequences used in the analysis are available upon request.

Quantitative RT PCR analysis

Total RNA was isolated and Turbo™ DNase treated as described above. Quantification of Superscript™ III generated cDNA was performed using an iCycler with an iQ5 multicolor real-time PCR detection system (Bio-Rad). Triplicate cDNA's were analyzed and normalized to Asf I cDNA. qPCR oligonucleotide primer combos were designed using Integrated DNA Technologies software. cDNA reactions were diluted 10 fold and 5 μl was analyzed. A 15 μl reaction mixture contained 4.5 pmol sense and antisense primer, 7.5 μl 2X iQ SYBR green super mix (Bio-Rad Laboratories). Standard curves were prepared for each gene using 5-fold dilutions of a known quantity (100 ng/μl) of WT gDNA. The quantities were calculated using iQ5 optical detection system software. Primers sequences used are available upon request.

Strand-specific RNA-seq library construction

For mRNA-seq, total RNA was isolated from *T. brucei* RNAi cultures grown in presence or absence of tetracycline for two days using TriPure. Six mRNA-seq libraries were constructed for PNUTS RNA (triplicate samples for plus and minus tetracyclin) and four libraries for VAa

RNA (duplicate samples) were constructed using Illumina TruSeq Stranded RNA LT Kit following the manufacturer's instructions with limited modifications. The starting quantity of total RNA was adjusted to 1.3 μg, and all volumes were reduced to a third of the described quantity. High throughput sequencing was performed at the Georgia Genomics and Bioinformatics Core (GGBC) on a NextSeq500 (Illumina).

RNA-seq analysis

Raw reads from mRNA-seq were first trimmed using fastp with default settings (v0.19.5; [129]). Remaining reads were locally aligned to the recently published long-read *T. brucei* Lister 427 2018 version 9.0 genome assembly (downloaded from [80]) and the Lister 427 BES sequences [82], using Bowtie2 version 2.3.4.1 [130]. With non-default settings (sensitive local) and further processed with Samtools version 1.6 [131]. To ensure proper read placement, alignments with multiple low-quality hits and mapping quality (MQ) scores less than 10 were removed. For each sample, HTSeq (v0.9.1) was used to count reads for each reference transcript annotation, followed by normalization/variance stabilization using DESeq2 (v1.18.1). Differential gene expression was conducted using DESeq2 by comparing TbPNUTs RNAi samples with and without tetracyclin in triplicate (log2 fold change and differential expression test statistics can be found in S3 Table). To express the transcripts levels for individual mRNA encoding genes as shown in S3 Table, we estimated transcript abundance as transcripts per million (TPM) by first normalizing the number of reads by kilobase of transcript, and then scaling each transcript per sample such that the sum of all transcript abundances within a sample was equal to 1 million. Due to incomplete gene annotation of the BES in the new *T. brucei* Lister 427 2018 genome assembly, gene expression changes for BES were determined by aligning reads to the Lister 427 BES sequences (S4 Table). To compare tetracyclin-treatment fold

changes for specific strands genome-wide, we counted reads from each strand in 200bp bins with a 100bp step. Mapping of differentially upregulated genes in a genome-wide context was determined by highlighting genes upregulated >3-fold in red for all 11 megabase chromosomes. Fold changes between Tet-untreated and Tet-treated PNUTS RNAi were plotted for all 11 megabase chromosomes as well. Tag counts in 200bp bins (100bp steps) were used to estimate correlations among samples (correlation coefficients among replicates were >0.99).

To analyze transcription defects at 3'- and 5'-end of PTUs, reads mapping to TTS (cSSRs) and TSS (dSSRs) were counted and reads per kilobase per million mapped reads (RPKM) values were generated. Similar to what we have previously done [24, 27, 28], lists of cSSRs and dSSRs were generated computationally as defined regions where coding strands switch based on the transcriptome as well as TSS mapped via triphosphate RNA sequencing. Head-to-Tail (HT) sites were defined were one PTU terminates (H3.V) and another PTU on the same strand initiates. The TSS at HT sites were further defined by a single peak of H2A.Z and the lack of an annotated gene on the antisense strand, distinguishing it from a TSS at a dSSR. Several SSRs located at subtelomeres were not included due to ambiguous nature of gene organization. SSRs and the 5-kb flanking regions were analyzed with DeepTools (v3.2.1) using 100bp bins flanking SSRs and dividing each SSR into 50 equally sized bins. Violin plots were generated using the R package vioplot, with the median and interquartile range illustrated by white circles and black boxes, respectively. Genes were considered adjacent to base J if the gene, according to the *T. brucei* Lister 427 annotation, was within 1-kb either upstream or downstream base J peaks. J IP-seq data shown here are from previously published work [23, 25].

To examine VSG expression, trimmed reads were bowtie-aligned to the VSGnome (retrieved from http://tryps.rockefeller.edu) [80] and differential expression was analyzed using

the DEseq2 identically to the analysis of the entire genome.

Cell cycle analysis

The cell cycle profile of the PNUTS RNAi cell line over time following induction was determined by staining cells with DAPI and cataloguing the nucleus/kinetoplastid (N/K) configurations of ~300 intact cells/time-point. Cells were harvested, 2 x washed in PBS, allowed to settle on glass slides and air dried for 10 min at RT. Cells were then fixed in -20°C methanol for 10 min and mounted with ProLong Gold antifade reagent with DAPI (Molecular Probes, P36935). Flow cytometry was carried out by washing the cells 2 times in PBS and resuspended at 1 x 10^7 / ml. Ice cold 100% ethanol was slowly added while shaking to a final concentration of 25% ethanol. Cells were then washed 2 times and resuspended in 0.5 ml PBS / 1% BSA / 1% Tween 20 with 5 μg / ml Propidium Iodide. Analysis was performed using a Cyan cytometer (DAKO).

Acknowledgments

We are grateful to Jessica Lopes da Rosa-Spiegler, David Reynolds, Nicolai Siegel, and Piet Borst for critical reading of the manuscript. We would also like to thank Shulamit Michaeli (Bar-IIan University) for the pSNSAP1 vector; Arthur Günzl (UConn Health) for the pC-PTP-NEO vector; Roberto DoCampo (University of Georgia) for VA a RNAi construct; Mike Cipriano (University of Georgia) for help with homology modeling; Christina L. Ethridge (University of Georgia) for generating the RNA-seq libraries.

Funding Statement

This work was supported by the National Institutes of Health [grant number

AI109108](to R.S). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

RNA-seq raw files and processed files have been deposited to the NCBI Gene Expression Omnibus (GEO) with accession number GSE135708.

REFERENCES

1. Shandilya J, Roberts SG. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta*. 2012;1819(5):391–400. 10.1016/j.bbagrm.2012.01.010

2. Kecman T, Kus K, Heo DH, Duckett K, Birot A, Liberatori S, et al. Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. *Cell Rep*. 2018;25(1):259–69 e5. 10.1016/j.celrep.2018.09.007

3. Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, et al. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol*. 2014;21(2):175–9. 10.1038/nsmb.2753

4. Bollen M, Peti W, Ragusa MJ, Beullens M. The extended PP1 toolkit: designed to create specificity. *Trends in biochemical sciences*. 2010;35(8):450–8. 10.1016/j.tibs.2010.03.002

5. Dancheck B, Nairn AC, Peti W. Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry*. 2008;47(47):12346–56. 10.1021/bi801308y

6. Ragusa MJ, Dancheck B, Critton DA, Nairn AC, Page R, Peti W. Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nat Struct Mol Biol*. 2010;17(4):459–64. 10.1038/nsmb.1786

7. Jagiello I, Beullens M, Stalmans W, Bollen M. Subunit structure and regulation of protein phosphatase-1 in rat liver nuclei. *J Biol Chem*. 1995;270(29):17257–63. 10.1074/jbc.270.29.17257

8. Kreivi JP, Trinkle-Mulcahy L, Lyon CE, Morrice NA, Cohen P, Lamond AI. Purification and characterisation of p99, a nuclear modulator of protein phosphatase 1 activity. *FEBS Lett*. 1997;420(1):57–62. 10.1016/s0014-5793(97)01485-3

9. Lee JH, You J, Dobrota E, Skalnik DG. Identification and characterization of a novel human PP1 phosphatase complex. *J Biol Chem*. 2010;285(32):24466–76. 10.1074/jbc.M110.109801

10. Lee JH, Skalnik DG. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol*. 2008;28(2):609–18. 10.1128/MCB.01356-07

11. Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, et al. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol Cell*. 2002;10(5):1139–50. 10.1016/s1097-2765(02)00707-4

12. He X, Khan AU, Cheng H, Pappas DL Jr., Hampsey M, Moore CL. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev*. 2003;17(8):1030–42. 10.1101/gad.1075203

13. Vanoosthuyse V, Legros P, van der Sar SJ, Yvert G, Toda K, Le Bihan T, et al. CPF-associated phosphatase activity opposes condensin-mediated chromosome condensation. *PLoS genetics*. 2014;10(6):e1004415 10.1371/journal.pgen.1004415

14. Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, et al. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J Biol Chem*. 2003;278(35):33000–10. 10.1074/jbc.M304454200

15. Cheng H, He X, Moore C. The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Mol Cell Biol*. 2004;24(7):2932–43. 10.1128/MCB.24.7.2932-2943.2004

16. Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, et al. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell*. 2008;29(5):577–87. 10.1016/j.molcel.2007.12.031

17. Austenaa LM, Barozzi I, Simonatto M, Masella S, Della Chiara G, Ghisletti S, et al. Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol Cell*. 2015;60(3):460–74. 10.1016/j.molcel.2015.09.018

18. Clayton C, Michaeli S. 3' processing in protists. *Wiley interdisciplinary reviews RNA*. 2011;2(2):247–55. 10.1002/wrna.49

19. Clayton CE. Gene expression in Kinetoplastids. *Curr Opin Microbiol*. 2016;32:46–51. 10.1016/j.mib.2016.04.018

20. Michaeli S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol*. 2011;6(4):459–74. 10.2217/fmb.11.20

21. Siegel TN, Gunasekera K, Cross GA, Ochsenreiter T. Gene expression in Trypanosoma brucei: lessons from high-throughput RNA sequencing. *Trends Parasitol*. 2011;27(10):434–41. 10.1016/j.pt.2011.05.006

22. Borst P, Sabatini R. Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol*. 2008;62:235–51. 10.1146/annurev.micro.62.081307.162750

23. Cliffe LJ, Siegel TN, Marshall M, Cross GA, Sabatini R. Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. *Nucleic Acids Res*. 2010;38(12):3923–35. 10.1093/nar/gkq146

24. Reynolds D, Hofmeister BT, Cliffe L, Alabady M, Siegel TN, Schmitz RJ, et al. Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription

of siRNA Loci in Trypanosoma brucei. *PLoS genetics*. 2016;12(1):e1005758 10.1371/journal.pgen.1005758

25. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, et al. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev*. 2009;23(9):1063–76. 10.1101/gad.1790409

26. van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, Velds A, et al. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell*. 2012;150(5):909–21. 10.1016/j.cell.2012.07.030

27. Reynolds D, Cliffe L, Forstner KU, Hon CC, Siegel TN, Sabatini R. Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic Acids Res*. 2014;42(15):9717–29. 10.1093/nar/gku714

28. Reynolds DL, Hofmeister BT, Cliffe L, Siegel TN, Anderson BA, Beverley SM, et al. Base J represses genes at the end of polycistronic gene clusters in Leishmania major by promoting RNAP II termination. *Mol Microbiol*. 2016;101(4):559–74. 10.1111/mmi.13408

29. Schulz D, Zaringhalam M, Papavasiliou FN, Kim HS. Base J and H3.V Regulate Transcriptional Termination in Trypanosoma brucei. *PLoS genetics*. 2016;12(1):e1005762 10.1371/journal.pgen.1005762

30. Cliffe LJ, Hirsch G, Wang J, Ekanayake D, Bullard W, Hu M, et al. JBP1 and JBP2 Proteins Are Fe2+/2-Oxoglutarate-dependent Dioxygenases Regulating Hydroxylation of Thymidine Residues in Trypanosome DNA. *J Biol Chem*. 2012;287(24):19886–95. 10.1074/jbc.M112.341974

31. Bullard W, da Rosa-Spiegler JL, Liu S, Wang D, Sabatini R. Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *JBC*. 2014.

32. Sekar A, Merritt C, Baugh L, Stuart K, Myler PJ. Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Mol Biochem Parasitol*. 2014;196(1):9–11. 10.1016/j.molbiopara.2014.07.005

33. Sabatini R, Cliffe L, Vainio L, Borst P. Enzymatic Formation of the Hypermodified DNA Base J In: Grosjean H, editor. *DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Function, Cellular Interactions and Evolution*. Texas: Landes Biosciences; 2009. p. 120–31.

34. Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, et al. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res*. 2009.

35. Kieft R, Brand V, Ekanayake DK, Sweeney K, DiPaolo C, Reznikoff WS, et al. JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form Trypanosoma brucei. *Mol Biochem Parasitol*. 2007;156(1):24–31. 10.1016/j.molbiopara.2007.06.010

36. Vainio S, Genest PA, ter Riet B, van Luenen H, Borst P. Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol Biochem Parasitol*. 2009;164(2):157–61. 10.1016/j.molbiopara.2008.12.001

37. Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, Kieft R, et al. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res*. 2007;35(7):2107–15. 10.1093/nar/gkm049

38. DiPaolo C, Kieft R, Cross M, Sabatini R. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol Cell*. 2005;17(3):441–51. 10.1016/j.molcel.2004.12.022

39. Heidebrecht T, Christodoulou E, Chalmers MJ, Jan S, Ter Riet B, Grover RK, et al. The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res*. 2011;39(13):5715–28. 10.1093/nar/gkr125

40. Sabatini R, Meeuwenoord N, van Boom JH, Borst P. Recognition of base J in duplex DNA by J-binding protein. *J Biol Chem*. 2002;277:958–66. 10.1074/jbc.M109000200

41. Sabatini R, Meeuwenoord N, van Boom JH, Borst P. Site-specific interactions of JBP with base and sugar moieties in duplex J-DNA. *Journal of Biological Chemistry*. 2002;277:28150–6. 10.1074/jbc.M201487200

42. Bullard W, Cliffe L, Wang P, Wang Y, Sabatini R. Base J glucosyltransferase does not regulate the sequence specificity of J synthesis in trypanosomatid telomeric DNA. *Mol Biochem Parasitol*. 2015;204(2):77–80. 10.1016/j.molbiopara.2016.01.005

43. Bullard W, Kieft R, Sabatini R. A method for the efficient and selective identification of 5-hydroxymethyluracil in genomic *DNA. Biol Methods Protoc*. 2017;2(1).

44. van Leeuwen F, Kieft R, Cross M, Borst P. Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in Trypanosoma brucei. *Molecular & Cellular Biology*. 1998;18(10):5643–51.

45. Aphasizhev R, Aphasizheva I, Nelson RE, Gao G, Simpson AM, Kang X, et al. Isolation of a U-insertion/deletion editing complex from Leishmania tarentolae mitochondria. *The EMBO Journal*. 2003;22(4):913–24. 10.1093/emboj/cdg083

46. Hurley TD, Yang J, Zhang L, Goodwin KD, Zou Q, Cortese M, et al. Structural basis for regulation of protein phosphatase 1 by inhibitor-2. *J Biol Chem*. 2007;282(39):28874–83. 10.1074/jbc.M703472200

47. Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R. Structural basis of protein phosphatase 1 regulation. *Nature*. 2004;429(6993):780–4. 10.1038/nature02582

48. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015;31(6):857–63. 10.1093/bioinformatics/btu744

49. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 2004;20(13):2138–9. 10.1093/bioinformatics/bth195

50. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*. 2007;8:211 10.1186/1471-2105-8-211

51. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5(4):725–38. 10.1038/nprot.2010.5

52. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015;12(1):7–8. 10.1038/nmeth.3213

53. Brenchley R, Tariq H, McElhinney H, Szoor B, Huxley-Jones J, Stevens R, et al. The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. *BMC Genomics*. 2007;8:434 10.1186/1471-2164-8-434

54. Li Z, Tu X, Wang CC. Okadaic acid overcomes the blocked cell cycle caused by depleting Cdc2-related kinases in Trypanosoma brucei. *Exp Cell Res*. 2006;312(18):3504–16. 10.1016/j.yexcr.2006.07.022

55. Schimanski B, Nguyen TN, Gunzl A. Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. *Eukaryot Cell*. 2005;4(11):1942–50. 10.1128/EC.4.11.1942-1950.2005

56. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*. 2013;10(8):730–6. 10.1038/nmeth.2557

57. Reynolds D, Cliffe L, Sabatini R. 2-Oxoglutarate-dependent hydroxylases involved In DNA base J (β-D-Glucopyranosyloxymethyluracil) synthesis In: Schofield CJ, Hausinger RP, editors. *2-Oxoglutarate-Dependent Oxygenases*. Cambridge: Royal Society of Chemistry; 2015.

58. Goos C, Dejung M, Janzen CJ, Butter F, Kramer S. The nuclear proteome of Trypanosoma brucei. *PLoS One*. 2017;12(7):e0181884 10.1371/journal.pone.0181884

59. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*. 2010;38(Database issue):D457–62. 10.1093/nar/gkp851

60. Ciurciu A, Duncalf L, Jonchere V, Lansdale N, Vasieva O, Glenday P, et al. PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. *PLoS genetics*. 2013;9(10):e1003885 10.1371/journal.pgen.1003885

61. Jerebtsova M, Klotchenko SA, Artamonova TO, Ammosova T, Washington K, Egorov VV, et al. Mass spectrometry and biochemical analysis of RNA polymerase II: targeting by protein phosphatase-1. *Mol Cell Biochem*. 2011;347(1–2):79–87. 10.1007/s11010-010-0614-3

62. Huang G, Ulrich PN, Storey M, Johnson D, Tischer J, Tovar JA, et al. Proteomic analysis of the acidocalcisome, an organelle conserved from bacteria to human cells. *PLoS Pathog*. 2014;10(12):e1004555 10.1371/journal.ppat.1004555

63. Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*. 2015;161(3):541–54. 10.1016/j.cell.2015.03.010

64. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*. 2009;457(7232):1038–42. 10.1038/nature07747

65. Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*. 2015;161(3):526–40. 10.1016/j.cell.2015.03.027

66. Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*. 2008;322(5909):1851–4. 10.1126/science.1164096

67. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature*. 2009;457(7232):1033–7. 10.1038/nature07728

68. Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS Pathog*. 2010;6(9).

69. Wedel C, Forstner KU, Derr R, Siegel TN. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J*. 2017;36(17):2581–94. 10.15252/embj.201695323

70. Gros J, Kumar C, Lynch G, Yadav T, Whitehouse I, Remus D. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Mol Cell*. 2015;60(5):797–807. 10.1016/j.molcel.2015.10.022

71. Mori S, Shirahige K. Perturbation of the activity of replication origin by meiosis-specific transcription. *J Biol Chem*. 2007;282(7):4447–52. 10.1074/jbc.M609671200

72. Nieduszynski CA, Blow JJ, Donaldson AD. The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Res*. 2005;33(8):2410–20. 10.1093/nar/gki539

73. Nieduszynski CA, Knox Y, Donaldson AD. Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev*. 2006;20(14):1874–9. 10.1101/gad.385306

74. Snyder M, Sapolsky RJ, Davis RW. Transcription interferes with elements important for chromosome maintenance in Saccharomyces cerevisiae. *Mol Cell Biol*. 1988;8(5):2184–94. 10.1128/mcb.8.5.2184

75. Tiengwe C, Marcello L, Farr H, Dickens N, Kelly S, Swiderski M, et al. Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei. *Cell Rep*. 2012;2(1):185–97. 10.1016/j.celrep.2012.06.007

76. Lombrana R, Alvarez A, Fernandez-Justel JM, Almeida R, Poza-Carrion C, Gomes F, et al. Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major. *Cell Rep*. 2016;16(6):1774–86. 10.1016/j.celrep.2016.07.007

77. Horn D. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol*. 2014;195(2):123–9. 10.1016/j.molbiopara.2014.05.001

78. Horn D, McCulloch R. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol*. 2010;13(6):700–5. 10.1016/j.mib.2010.08.009

79. Muller LSM, Cosentino RO, Forstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome

organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*. 2018;563(7729):121–5. 10.1038/s41586-018-0619-8

80. Cross GA, Kim HS, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Mol Biochem Parasitol*. 2014;195(1):59–73. 10.1016/j.molbiopara.2014.06.004

81. Wickstead B, Ersfeld K, Gull K. The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. *Genome Res*. 2004;14(6):1014–24. 10.1101/gr.2227704

82. Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, et al. Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One*. 2008;3(10):e3527 10.1371/journal.pone.0003527

83. Ehlers B, Czichos J, Overath P. RNA turnover in Trypanosoma brucei. *Mol Cell Biol*. 1987;7(3):1242–9. 10.1128/mcb.7.3.1242

84. Kabiri M, Steverding D. Studies on the recycling of the transferrin receptor in Trypanosoma brucei using an inducible gene expression system. *Eur J Biochem*. 2000;267(11):3309–14. 10.1046/j.1432-1327.2000.01361.x

85. Kerry LE, Pegg EE, Cameron DP, Budzak J, Poortinga G, Hannan KM, et al. Selective inhibition of RNA polymerase I transcription as a potential approach to treat African trypanosomiasis. *PLoS Negl Trop Dis*. 2017;11(3):e0005432 10.1371/journal.pntd.0005432

86. Vanhamme L, Berberof M, Le Ray D, Pays E. Stimuli of differentiation regulate RNA elongation in the transcription units for the major stage-specific antigens of Trypanosoma brucei. *Nucleic Acids Res*. 1995;23(11):1862–9. 10.1093/nar/23.11.1862

87. Vanhamme L, Pays E, McCulloch R, Barry JD. An update on antigenic variation in African

trypanosomes. *TRENDS in Parasitology*. 2001;17:338–43. 10.1016/s1471-4922(01)01922-5

88. Vanhamme L, Poelvoorde P, Pays A, Tebabi P, Van Xong H, Pays E. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. *Mol Microbiol*. 2000;36(2):328–40. 10.1046/j.1365-2958.2000.01844.x

89. Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol*. 2007;156(2):93–101. 10.1016/j.molbiopara.2007.07.007

90. Kelly S, Kramer S, Schwede A, Maini PK, Gull K, Carrington M. Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open Biol*. 2012;2(4):120033 10.1098/rsob.120033

91. Engstler M, Boshart M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. *Genes Dev*. 2004;18(22):2798–811. 10.1101/gad.323404

92. Rolin S, Hancocq-Quertier J, Paturiaux-Hanocq F, Nolan DP, Pays E. Mild acid stress as a differentiation trigger in Trypanosoma brucei. *Mol Biochem Parasitol*. 1998;93(2):251–62. 10.1016/s0166-6851(98)00046-2

93. Biebinger S, Rettenmaier S, Flaspohler J, Hartmann C, Pena-Diaz J, Wirtz LE, et al. The PARP promoter of Trypanosoma brucei is developmentally regulated in a chromosomal context. *Nucleic Acids Research*. 1996;24(7):1202–11. 10.1093/nar/24.7.1202

94. Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, et al. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol*. 2006;13(8):704–12. 10.1038/nsmb1119

95. Soares LM, Buratowski S. Yeast Swd2 is essential because of antagonism between Set1

histone methyltransferase complex and APT (associated with Pta1) termination factor. *J Biol Chem*. 2012;287(19):15219–31. 10.1074/jbc.M112.341412

96. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*. 2003;11(3):709–19. 10.1016/s1097-2765(03)00092-3

97. Dingar D, Tu WB, Resetca D, Lourenco C, Tamachi A, De Melo J, et al. MYC dephosphorylation by the PP1/PNUTS phosphatase complex regulates chromatin binding and protein stability. *Nature communications*. 2018;9(1):3502 10.1038/s41467-018-05660-0

98. Parua PK, Booth GT, Sanso M, Benjamin B, Tanny JC, Lis JT, et al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*. 2018;558(7710):460–4. 10.1038/s41586-018-0214-z

99. Srivastava A, Badjatia N, Lee JH, Hao B, Gunzl A. An RNA polymerase II-associated TFIIF-like complex is indispensable for SL RNA gene transcription in Trypanosoma brucei. *Nucleic Acids Res*. 2017.

100. Urbaniak MD, Martin DM, Ferguson MA. Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. *J Proteome Res*. 2013;12(5):2233–44. 10.1021/pr400086y

101. Guo Z, Stiller JW. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics*. 2004;5:69 10.1186/1471-2164-5-69

102. Das A, Banday M, Fisher MA, Chang YJ, Rosenfeld J, Bellofatto V. An essential domain of an early-diverged RNA polymerase II functions to accurately decode a primitive chromatin landscape. *Nucleic Acids Res*. 2017;45(13):7886–96. 10.1093/nar/gkx486

103. Das A, Bellofatto V. The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in Trypanosoma brucei. *PLoS One*. 2009;4(9):e6959 10.1371/journal.pone.0006959

104. Rocha AA, Moretti NS, Schenkman S. Stress induces changes in the phosphorylation of Trypanosoma cruzi RNA polymerase II, affecting its association with chromatin and RNA processing. *Eukaryot Cell*. 2014;13(7):855–65. 10.1128/EC.00066-14

105. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008;322(5909):1845–8. 10.1126/science.1162228

106. Preker P, Nielsen J, Schierup MH, Jensen TH. RNA polymerase plays both sides: vivid and bidirectional transcription around and upstream of active promoters. *Cell Cycle*. 2009;8(8):1106–7.

107. Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2013;110(8):2876–81. 10.1073/pnas.1221904110

108. Wei W, Pelechano V, Jarvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. *Trends Genet*. 2011;27(7):267–76. 10.1016/j.tig.2011.04.002

109. Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A*. 2011;108(26):10460–5. 10.1073/pnas.1106630108

110. Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, et al. Polyadenylation

site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol*. 2013;20(8):923–8. 10.1038/nsmb.2640

111. Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, et al. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*. 2013;155(5):1075–87. 10.1016/j.cell.2013.10.024

112. Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, et al. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife*. 2014;3:e02105 10.7554/eLife.02105

113. Fong N, Saldi T, Sheridan RM, Cortazar MA, Bentley DL. RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Mol Cell*. 2017;66(4):546–57 e3. 10.1016/j.molcel.2017.04.016

114. Shah N, Maqbool MA, Yahia Y, El Aabidine AZ, Esnault C, Forne I, et al. Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol Cell*. 2018;69(1):48–61 e6. 10.1016/j.molcel.2017.12.009

115. Shetty A, Kallgren SP, Demel C, Maier KC, Spatt D, Alver BH, et al. Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol Cell*. 2017;66(1):77–88 e5. 10.1016/j.molcel.2017.02.023

116. Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, et al. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res*. 2013;23(1):1–11. 10.1101/gr.142331.112

117. Miotto B, Ji Z, Struhl K. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci U S A*. 2016;113(33):E4810–9.

10.1073/pnas.1609060113

118. Soudet J, Gill JK, Stutz F. Noncoding transcription influences the replication initiation program through chromatin regulation. *Genome Res*. 2018;28(12):1882–93. 10.1101/gr.239582.118

119. Kim HS. Genome-wide function of MCM-BP in Trypanosoma brucei DNA replication and transcription. *Nucleic Acids Res*. 2019;47(2):634–47. 10.1093/nar/gky1088

120. Benmerzouga I, Concepcion-Acevedo J, Kim HS, Vandoros AV, Cross GA, Klingbeil MM, et al. Trypanosoma brucei Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. *Mol Microbiol*. 2013;87(1):196–210. 10.1111/mmi.12093

121. Anderson SJ, Sikes ML, Zhang Y, French SL, Salgia S, Beyer AL, et al. The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively. *J Biol Chem*. 2011;286(21):18816–24. 10.1074/jbc.M110.202101

122. Viktorovskaya OV, Appling FD, Schneider DA. Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *J Biol Chem*. 2011;286(21):18825–33. 10.1074/jbc.M110.202119

123. Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol Biochem Parasitol*. 1999;99(1):89–101. 10.1016/s0166-6851(99)00002-x

124. Wickstead B, Ersfeld K, Gull K. Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of Trypanosoma brucei. *Mol Biochem Parasitol*. 2002;125(1–2):211–6. 10.1016/s0166-6851(02)00238-4

125. Oberholzer M, Morand S, Kunz S, Seebeck T. A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. *Mol Biochem Parasitol*. 2006;145(1):117–

20. 10.1016/j.molbiopara.2005.09.002

126. van Leeuwen F, Wijsman ER, Kieft R, van der Marel GA, van Boom JH, Borst P. Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. *Genes & Development*. 1997;11(23):3232–41.

127. Lim JM, Sherling D, Teo CF, Hausman DB, Lin D, Wells L. Defining the regulated secreted proteome of rodent adipocytes upon the induction of insulin resistance. *J Proteome Res*. 2008;7(3):1251–63. 10.1021/pr7006945

128. Baez-Santos YM, Mielech AM, Deng X, Baker S, Mesecar AD. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J Virol*. 2014;88(21):12511–27. 10.1128/JVI.01294-14

129. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–i90. 10.1093/bioinformatics/bty560

130. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357–9. 10.1038/nmeth.1923

131. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. 10.1093/bioinformatics/btp352

132. UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506–D15. 10.1093/nar/gky1049

133. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W200–W4. 10.1093/nar/gky448

134. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.

10.1093/bioinformatics/btp348

135. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. 10.1093/bioinformatics/btu033

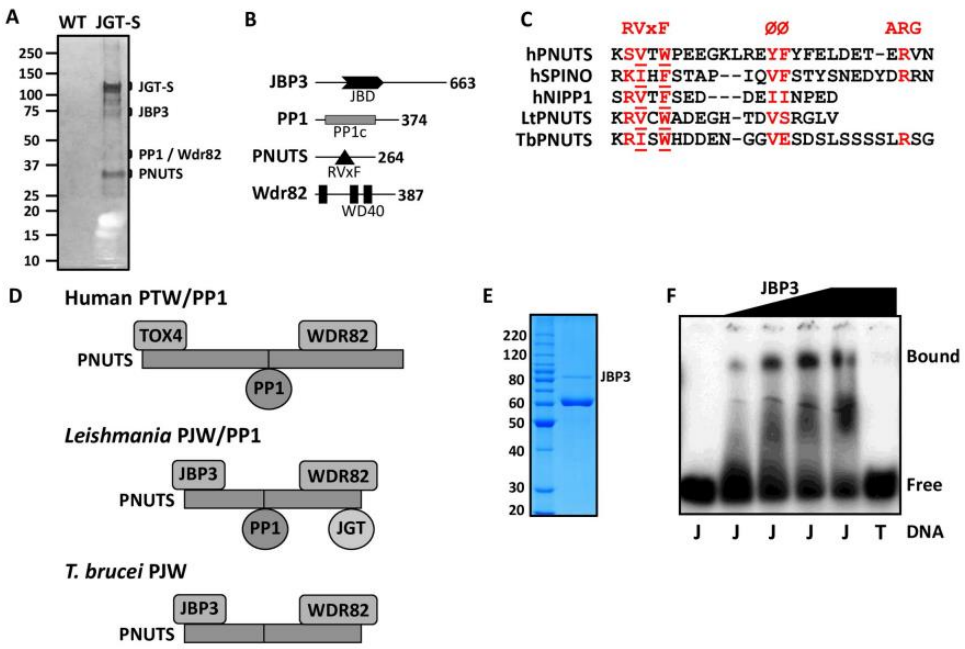


Fig. A.1 Identification of a novel phosphatase complex in *Leishmania tarentolae*. (A) Proteins recovered from tandem affinity purification from wild-type *L. tarentolae* extracts (WT) and from cells expressing the strep-tagged version of JGT (JGT-S) in a single JGT KO background, were analyzed on SDS-PAGE gel and silver stained. The location of the tagged JGT (confirmed by western blot) and associated proteins are indicated based upon MS analysis of the gel fragments and their predicted molecular weight. The inherent disorder in PNUTS may explain why it migrates slower in SDS-PAGE than predicted. The additional bands in the SDS-PAGE presumably represent degradation (primarily of JGT) as indicated by the MS analysis. (B) Summary of the PJW/PP1 complex. The domain structure of each component in the complex is schematically shown. JBD, J-DNA binding domain; RVxF motif, PP1 docking motif; PP1c, catalytic domain of protein phosphatase 1; WD40, WD40 repeat. The number of amino acids in each component is indicated. (C) Structure-based sequence alignment of the PNUTS, spinophilin PP1 and NIPP1 interactive domains in humans is compared with LtPNUTS (residues 95–112) and TbPNUTS (residues 139–164), where x is any amino acid and Ø represents a hydrophobic

amino acid. Critical residues in the RVxF motif are underlined. (D) Models for the PJW/PP1 complex in Leishmania and *T. brucei*. The models are based on the human PTW/PP1 complex where PNUTS acts as scaffold and the DNA binding protein and Wdr82 bind to the N- and C-terminus, respectively, and PP1 binds via the RVxF PP1 interaction motif (indicated by the line). PP1 is presumably not stably associated with the complex in *T. brucei* (see discussion) and is therefore labeled as the PJW complex. (E) Purification of recombinant LtJBP3. His-tagged rJBP3 expressed in *E. coli* was purified by metal affinity and size exclusion chromatography and analyzed by SDS-PAGE/Coomassie staining. The major copurified protein is the *E. coli* molecular chaperone GroEL. The migration of protein molecular mass standards (in kDa) is shown on the left. (F) Gel shift assays for modified and unmodified DNA substrates interacting with JBP3. 0.3 pmol radiolabeled J-DNA (J) was incubated with 0, 0.2, 0.38, 0.57 and 1 pmol of JBP3 and 0.3 pmol radiolabeled unmodified DNA (T) was incubated with 1 pmol of JBP3.

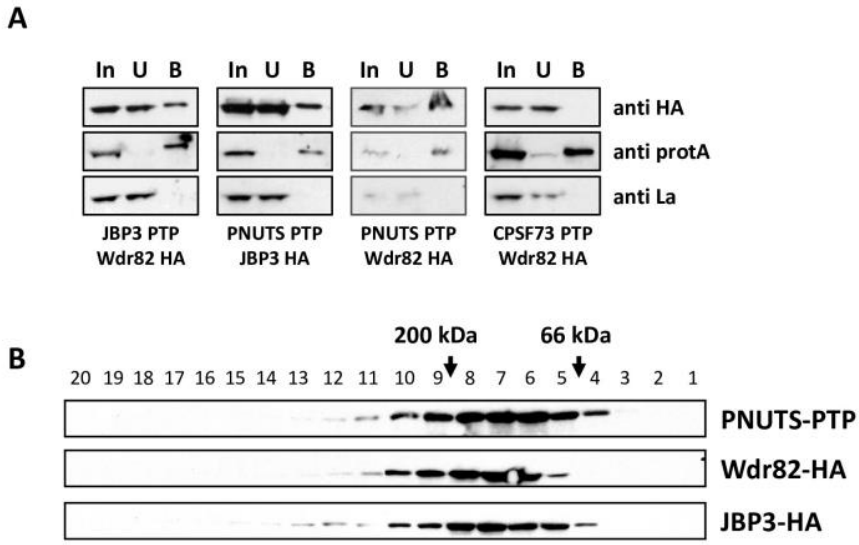


Fig. A.2 Characterization of the PJW/PP1 complex in *T. brucei*. (A) Co-immunoprecipitation of JBP3, Wdr82 and PNUTS. Cell extracts from bloodstream form *T. brucei* cells that endogenously express PTP- and HA-tagged versions of the indicated proteins were purified by protein A affinity and analyzed by western blotting with anti-HA, anti-protein A and anti-La. Equal cell equivalents of input (IP), unbound (U) and bound (B) fractions were loaded on the gel. (B) PNUTS, Wdr82, and JBP3 co-migrate following sucrose gradient fractionation. Cell extracts from BSF *T. brucei* cells that endogenously express PNUTS-PTP and JBP3-HA or express PNUTS-PTP and WDR82-HA were loaded onto 5–50% sucrose gradients and analyzed by density gradient centrifugation. Equal volumes of each fraction were analyzed by western blotting. Migration of PNUTS, JBP3 and WDR82 are shown in the gradient. The PNUTS and WDR82 signals were obtained from the same gradient and the JBP3 signal obtained from the other cell line applied to a parallel gradient. Molecular markers were applied to a parallel gradient.

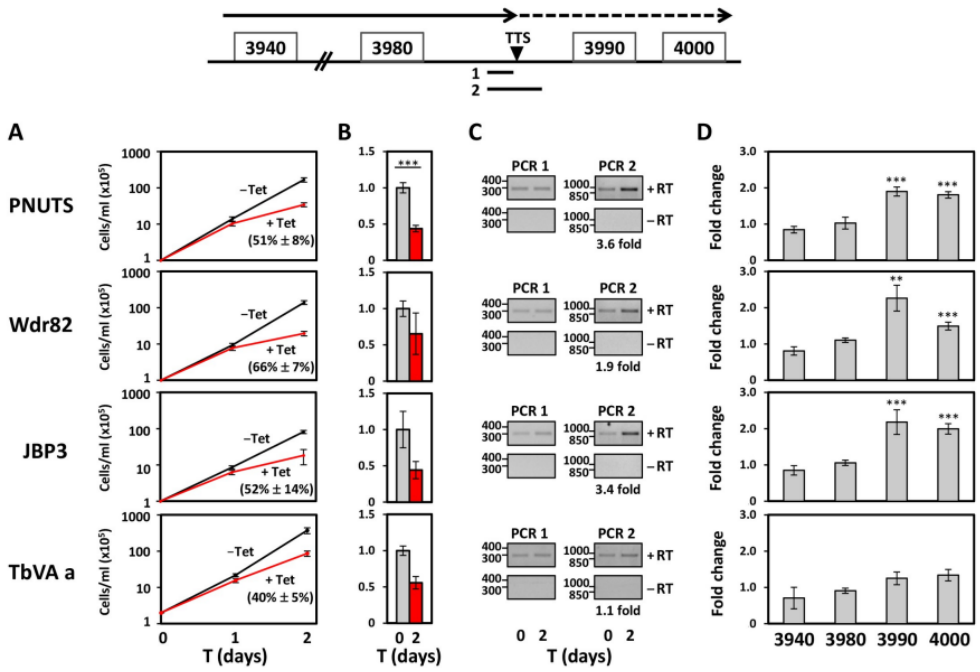


Fig. A.3 PNUTS, Wdr82 and JBP3 are involved in Pol II transcription termination in *T. brucei*. (A) PNUTS, JBP3 and Wdr82 are essential for cell viability of the infectious BSF *T. brucei*. Cell growth was arrested upon PNUTS, JBP3 and Wdr82 mRNA ablation by RNAi. VA a; acidocalcisome VA a protein. Reduction in cell growth rates upon mRNA ablation between day 1–2 are indicated. (B) Depletion of mRNA upon Tet induction of RNAi by qRT-PCR analysis. Results are arithmetic mean with error bars showing standard deviation from three biological replicates analyzed in triplicate. P values were calculated using Student's t test. ***, p value ≤ 0.001. (C and D) Analysis of Pol II termination. Above; Schematic representation of a termination site on chromosome 5 where Pol II has been shown to terminate prior to the last two genes (Tb927.5.3990 and Tb927.5.4000) in the PTU. The dashed arrow indicates readthrough transcription past the TTS that is regulated by base J and H3.V. Solid lines below indicate regions (1 and 2) analyzed by RT-PCR. (C) RT-PCR analysis of nascent RNA. cDNA was synthesized using random hexamers and PCR was performed using the appropriate reverse primer for each region (1 and 2) plus an identical forward primer. A minus-RT control (-RT) is

included. Quantitation (fold upregulation) of the PCR products for region 2 of the indicated gels using Gel Doc Quantity One software is provided below. (D) RT-qPCR analysis of genes numbered according to the ORF map above. Transcripts were normalized against Asf1 (which represents a PTU internal Pol II transcribed single copy gene as a negative control), and are plotted as the average and standard deviation of three biological replicates analyzed in triplicate. P values were calculated using Student's t test. **, p value ≤ 0.01; ***, p value ≤ 0.001. RT-PCR products for genes 3990 and 4000 were confirmed by cloning and sequencing of multiple clones.

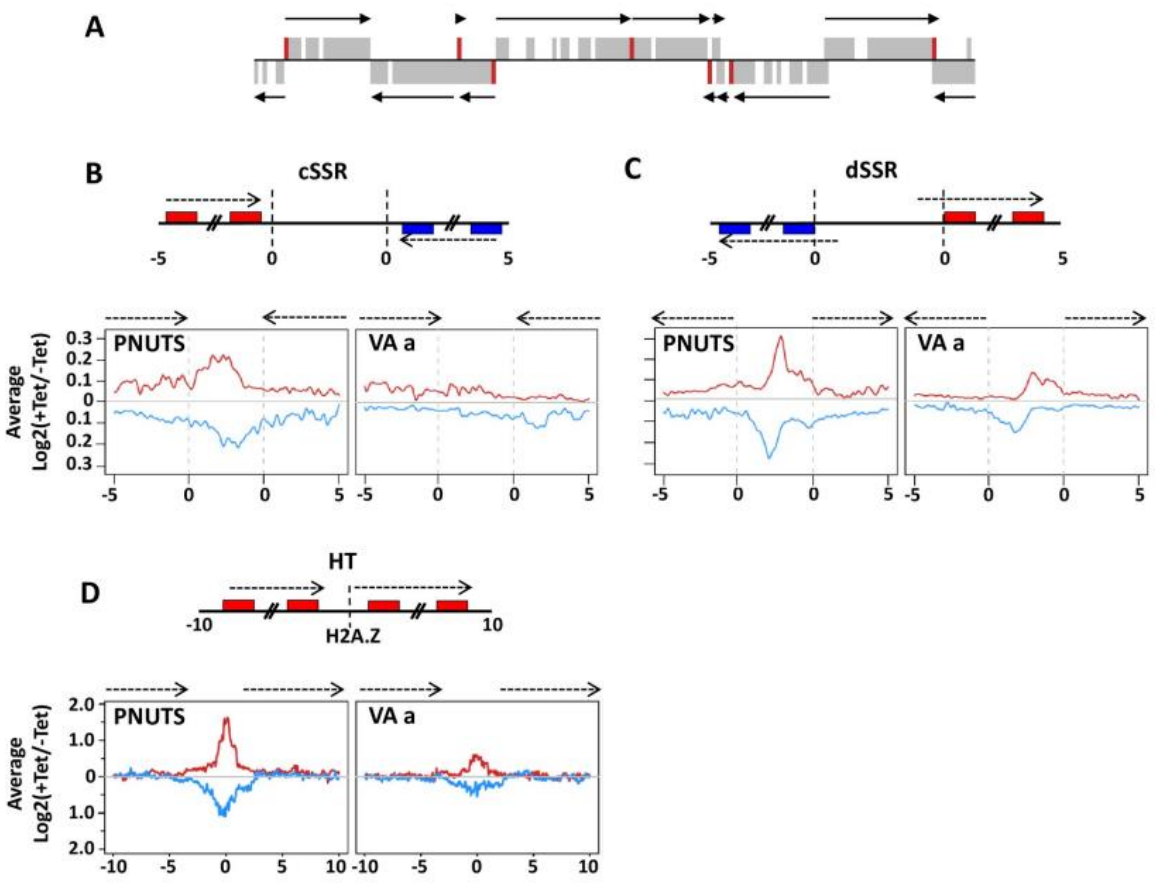


Fig. A.4 TbPNUTS affects transcription termination at 3' end and 5' end of PTUs.

Levels of transcripts downstream of TTSs and upstream of TSSs increased upon TbPNUTS ablation. (A) A core section of chromosome 10 is shown; genes that were upregulated >3-fold upon ablation of PNUTS are highlighted in red. Arrows indicate the direction of transcription of PTUs. (B and C) Diagram of a cSSR (B) and dSSR (C): boxes are genes in the PTUs and arrows indicate PTUs and direction of transcription. In cSSR, the poly A sites of the final gene in the PTU (indicated by the transcriptome) are marked by dotted line (0 kb). In dSSR, the dotted line (0 kb) indicates the 5' end of the first gene in the PTU (according to the transcriptome). Thus, the TSS is located further upstream, within the dSSR. Numbers refer to distance from SSR (kb). Below: Metagene profile of total sense and antisense RNA-seq signal over the SSRs and 5kb upstream and downstream regions into the PTUs. Fold changes comparing transcript levels

between day 0 and day 2 following induction of RNAi were calculated and plotted over the indicated regions genome-wide. Red and blue lines indicate RNAs from the top and bottom strand, respectively, as indicated on the diagram above. PNUTS, PNUTS RNAi; VA a, RNAi VA a. (D) Above: Diagram of a HT site. Boxes are the genes and arrows indicate direction of transcription as in C. The center of the H2A.Z peak is marked by a dotted line. Numbers refer to distance from center of H2A.Z peak (kb). Below: Metagene profile of the fold changes in RNA-seq reads over the HT sites and 10 kb upstream and downstream of the H2A.Z peak, as described in C.

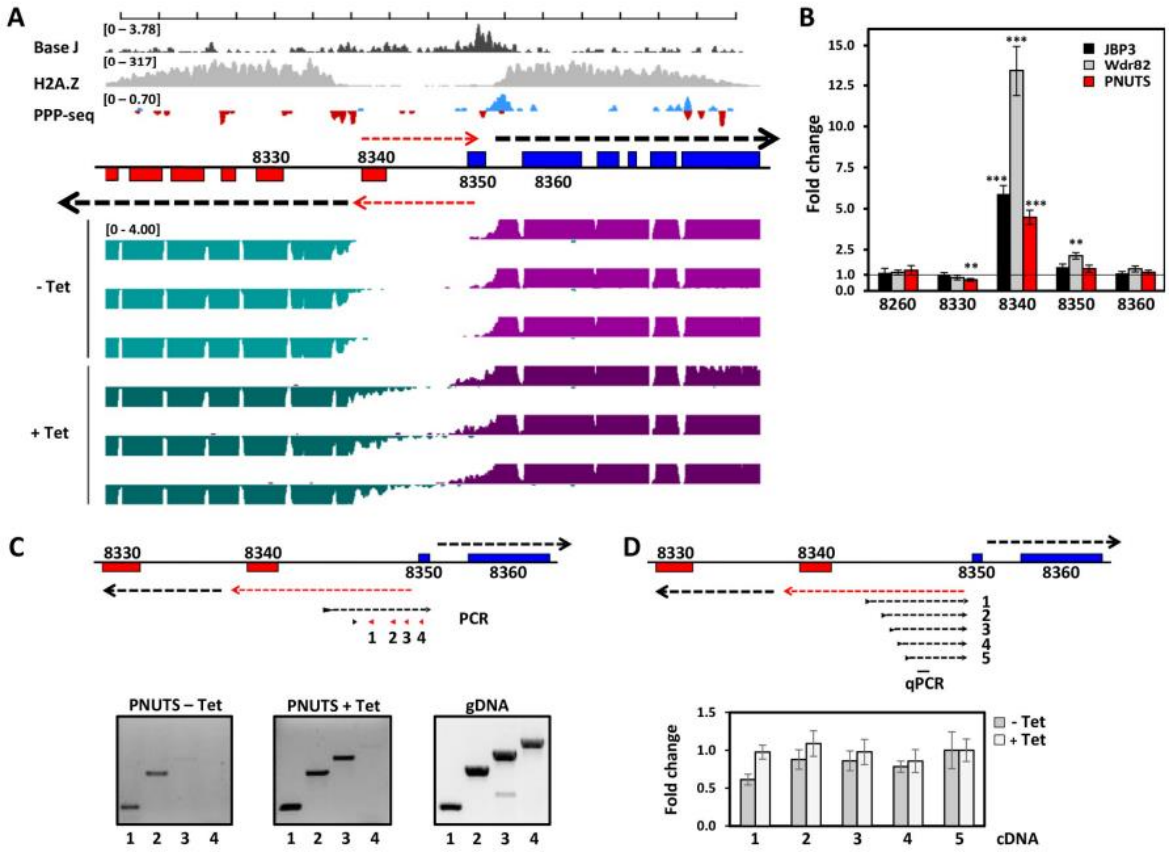


Fig. A.5 TbPNUTS affects early termination of antisense transcription at TSSs.

(A) Representative region of chromosome 10 illustrating bi-directional transcription at TSSs upon TbPNUTS ablation. TSSs are denoted by PPP-seq and H2A.Z ChIP-seq enrichment in wild-type *T. brucei*. PPP-seq track colors: Red, reverse strand coverage; blue, forward strand coverage. RNA-seq track colors: Green, reverse strand coverage; Purple, forward strand coverage. Black arrows indicate direction of sense transcription. Red arrows indicate stimulated antisense transcription in the PNUTS mutant that leads to de-repression of the annotated 8340 gene on the bottom strand and, to a lesser degree, the 8350 gene on the top strand. (B) Confirmation of mRNA-seq transcript changes in the PNUTS, Wdr82 and JBP3 RNAi by RT-qPCR. RT-qPCR analysis was performed for the indicated genes as described in Material and Methods. Transcripts were normalized against Asf1 mRNA, and are plotted as the average and

standard deviation of three replicates. P values were calculated using Student's t test. **, p value ≤ 0.01; ***, p value ≤ 0.001. RT-PCR products for gene 8340 were confirmed by cloning and sequencing of multiple clones. (C) Mapping of the 5' end of the antisense transcript by nested RT-PCR on strand-specific cDNA. PCR utilized constant 3' primer (indicated by black arrow-head) with varying 5' primers indicated in red (primers 1–4). cDNA levels utilized in the PCR reactions were normalized against strand-specific Asf1 mRNA. (D) Strand-specific RT-qPCR analysis of antisense nascent transcript. cDNA was generated using various strand specific 3' primers and RNA from–and + Tet. qPCR was then done using internal primers to amplify the region indicated by black bar. Transcript levels were normalized against strand-specific Asf1 mRNA. Error bars indicate standard deviation from at least three experiments.

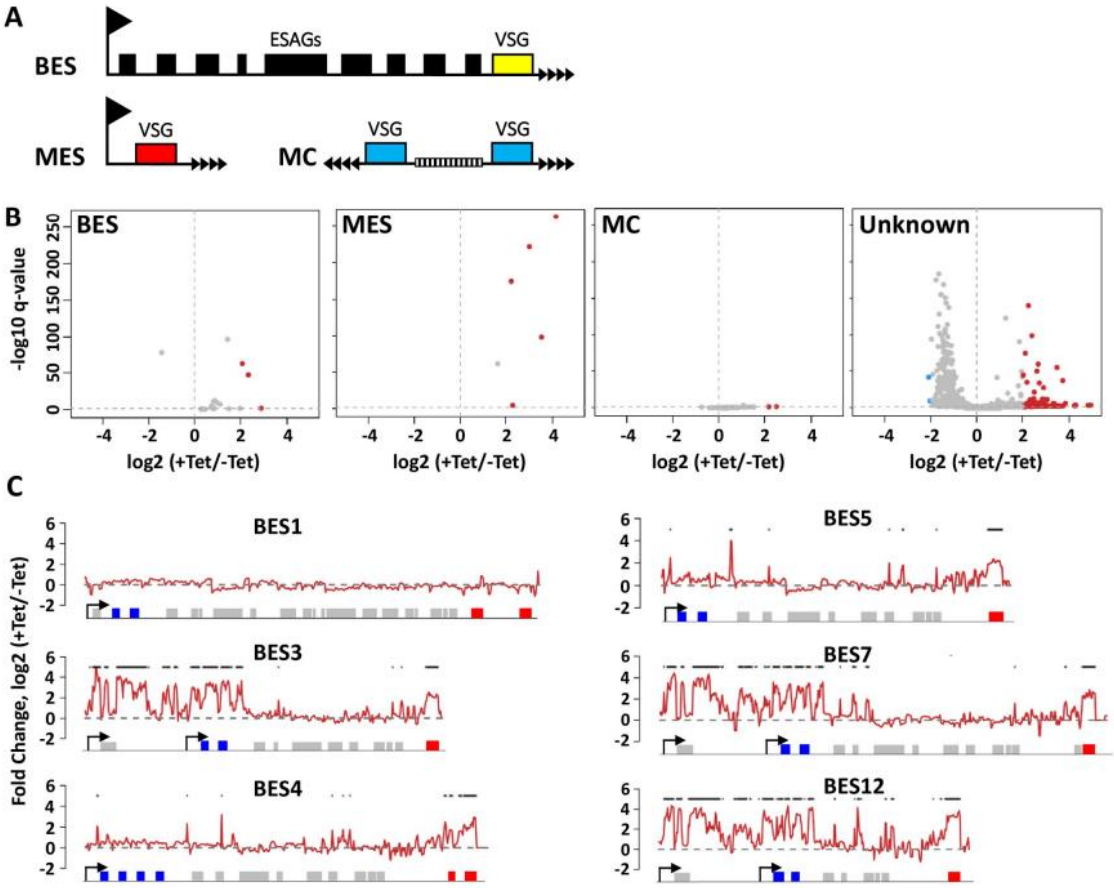


Fig. A.6 TbPNUTS is required for silencing of VSG genes. (A) Schematic diagrams of telomeric VSG genome locations; BES, Bloodstream-form Expression sites; MES, metacyclic expression sites; and MC, minichromosomal sites. ESAGs, expression site associated genes. (B) Reads from the RNA-seq experiment were aligned to the VSGnome database and raw reads mapping to each VSG was analyzed with DEseq as described in Materials and Methods. Fold changes comparing transcript levels between day 0 and day 2 following induction of PNUTS RNAi were calculated and plotted for BES, MES or MC VSGs. The rest of the VSGs ('Unknown') excluding BES, MC and MES were graphed separately. Red dots represent genes with greater than 3-fold change that are also significant with a Benjamini-Hochberg FDR test correction. (C) PNUTS regulates transcription of VSG BES. RNA-seq reads were aligned to the *T. brucei* 427 BES sequences (14 BESs). Fold changes comparing–and + Tet were plotted

over each BES. Six of the BES are shown here. See A.S11 Fig for all BES. Diagram indicates annotated genes, boxes, within the BES PTU. ESAG 6 and 7 are indicated in blue. The last gene on the right is the VSG gene (red). Some BESs have pseudo-VSG genes upstream. Promoters are indicated by arrows. Some BESs have two promoters. Asterix denote bins with greater than 3-fold change of expression following PNUTS ablation.

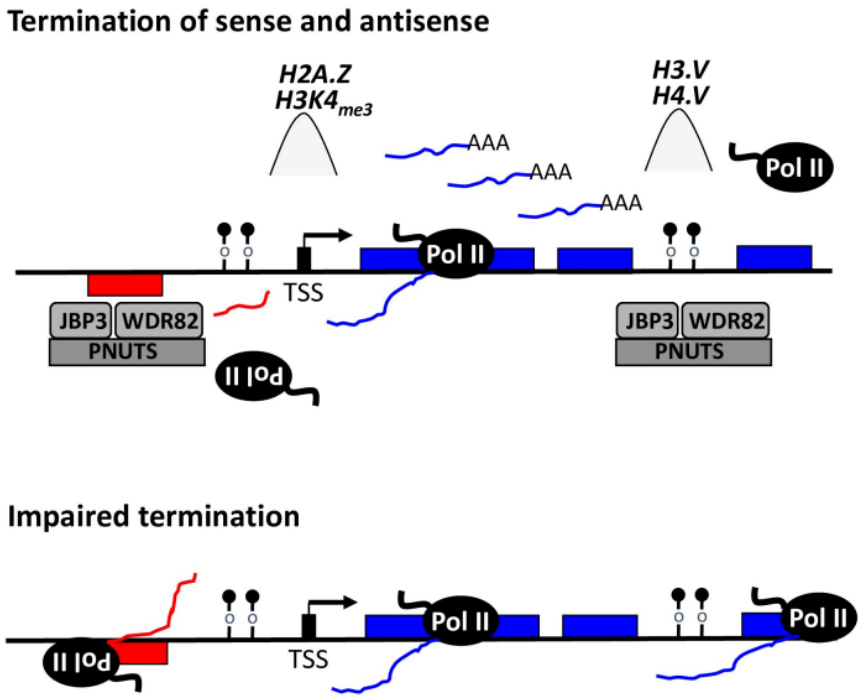


Fig. A.7 Regulation of termination by the PJW/PP1 complex in kinetoplastids. Schematic diagram of polycistronic RNA Pol II transcription. Transcription start sites (TSS), and direction of transcription, for the PTU on the top strand (blue genes) and bottom strand (red genes) are indicated by the black arrow. Peaks of histone variant (H2A.Z) and methylated histone H3 (H3K4me3) localized at TSS are indicated. According to the model, bi-directional transcription initiates at each TSS, but only the Pol II transcribing the 'sense strand' fully elongates and generates productive poly (A) mRNAs. Pol II terminates the 3'-end of the PTU marked with base J (ball and stick) and H3.V. In *T. brucei*, the PJW complex is recruited to the termination site via JBP3 recognition of J-DNA and somehow controls termination. In Leishmania, regulation of termination by the PJW/PP1 complex presumably involves dephosphorylation of the CTD of Pol II (as discussed in the text). While we propose a similar mechanism in *T. brucei*, these studies did not identify PP1 as a stable component of the PJW complex. The PJW complex is also recruited to the dSSR upstream of the TSS, leading to premature termination of antisense transcription. These short transcripts may be additionally targeted for degradation. Impaired

termination, following mutation of the PJW complex, leads to readthrough transcription at the 3'- and 5'-end of PTUs. Genes located downstream of these termination sites, at the 3'- and 5'-end (upstream of the TSS) of a PTU, can generate stabilized polyA mRNAs and be expressed.

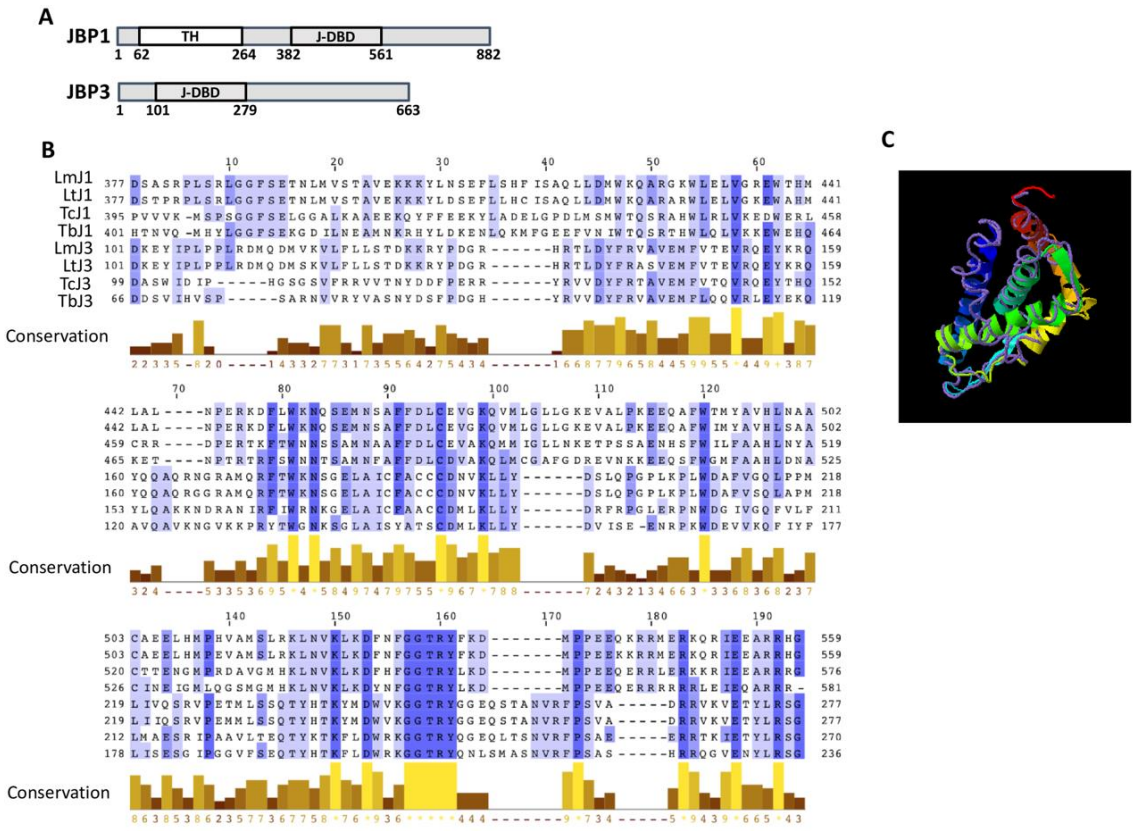


Fig. A.S1 JBP3 is a putative J-binding protein with a conserved JBD motif. (A) Schematic representation of the structure of JBP1 and JBP3 from *L. tarentolae* illustrating the presence of the conserved JBD and variable C-termini. (B) A multiple sequence alignment of the JBD of JBP1 homologues from *T. brucei* (Tb927.11.13640), *L. major* (LmjF.09.1480), *T. cruzi* (TcCLB.506753.120), and *L. tarentolae* (LtaP09.1510) and the conserved region of JBP3 is shown. The sequence alignment was generated using Maft and visualized with Jalview. Identical amino acids are indicated by highlighting; >80% agreement is highlighted in mid blue and >60% in light blue. Similar amino acids are indicated by hierarchical analysis of residue conservation shown below. (C) 3D structure prediction. Using the structure of the JBD of JBP1, JBP3 was run through I-TASSER and aligns with RMSD of 1.46, Cov score of 0.774, TM-score of 0.727. In the superposition, the thick backbones are the native JBP1 JBD structure and the thin backbone is the I-TASSER model of JBP3. Blue to red runs from N- to C-terminal.

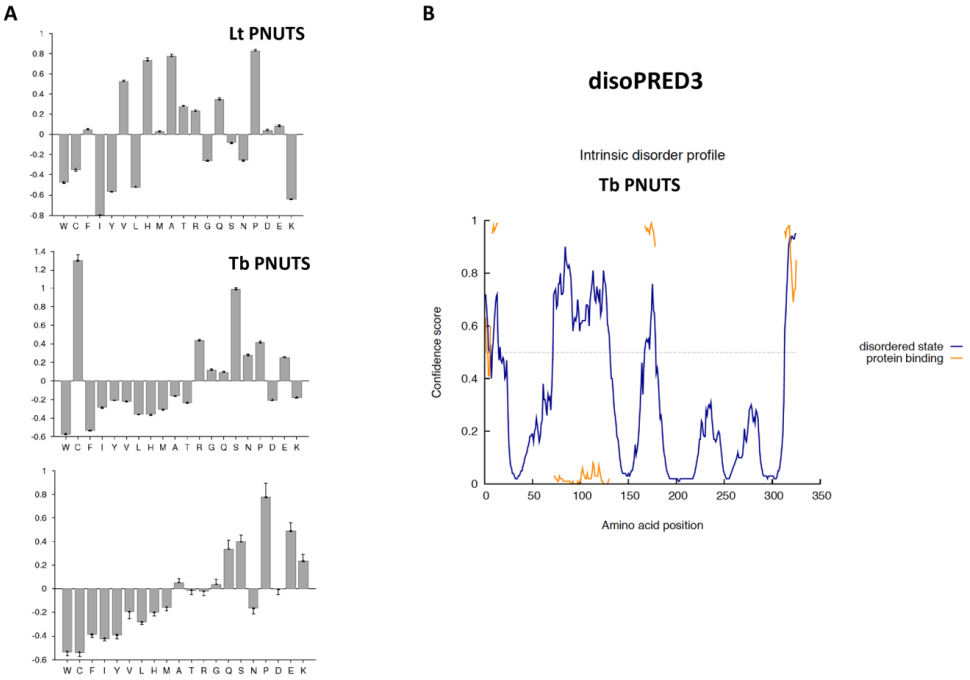


Fig. A.S2 PNUTS is a disordered protein. (A) Compositional profiling of Lt and Tb PNUTS showing the fractional amino acid composition in comparison with the compositional profile of typical ordered proteins. The compositional profile of typically disordered proteins from the DisProt database is shown for comparison below. (B) Analysis of TbPNUTS using the DISOPRED3 program for protein disorder prediction and for protein-binding site annotation within disordered regions.

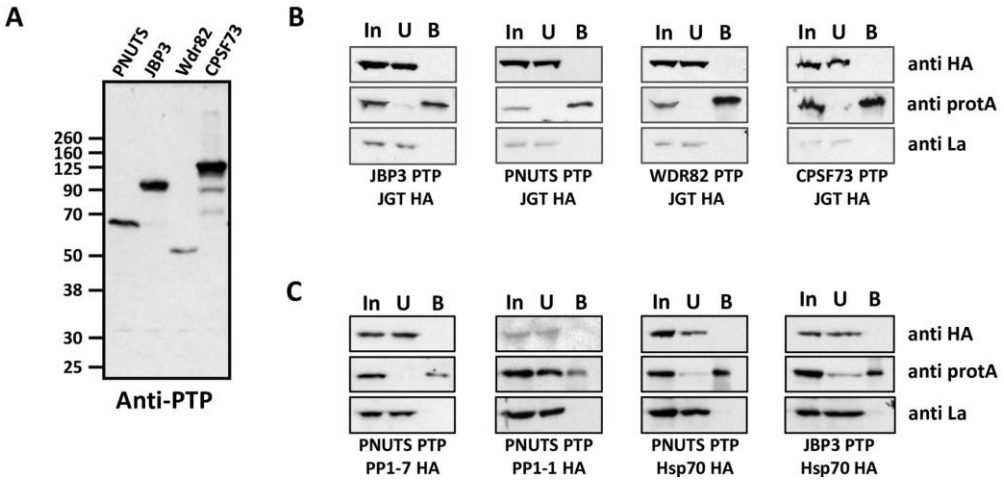


Fig. A.S3 Co-immunoprecipitation. (A) Endogenously PTP tagged PNUTS, JBP3, Wdr82 and CPSF73. (A and B) Co-IP experiments as described in Fig A.2A. (B) JGT is not associated with the *T. brucei* complex. (C) PP1-1, PP1-7 and HSP70 do not associate with PNUTS in *T. brucei*.

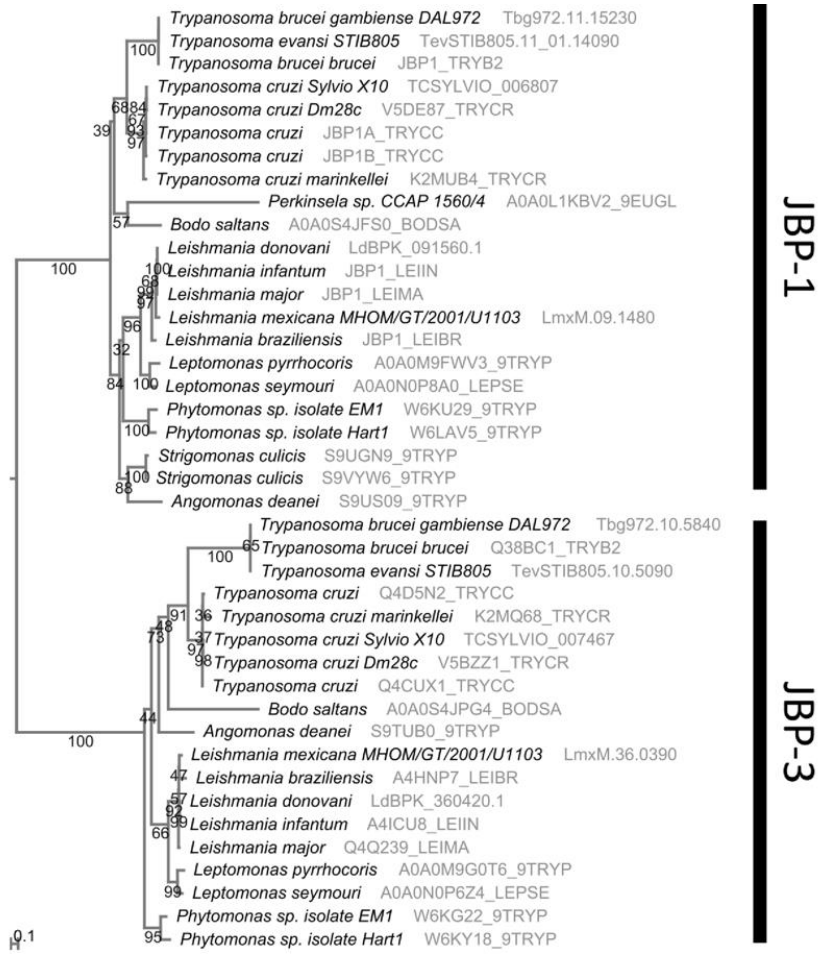


Fig. A.S4 JBP3 is restricted to kinetoplastids. A seed alignment of JBP-1 and JBP-3 was used to iteratively search UniProtKB [132] using jackhammer [133] until convergence with an e-value cut-off of 0.01 for sequence and 0.03 for hits. Full length protein sequences from representative species were chosen and aligned using hmmalign [133] and alignancer, trimmed using trimAL [134] with the automated1 flag, and a phylogenetic tree was made using raxml [135] (options -f a -x 12345 -p 12345 -N autoMRE -m PROTGAMMAJTTF). Matching sequences were found exclusively in species within the Kinetoplastida class. The sequences separated into two distinct groups with high bootstrap support. We were able to find JBP-3 family members in all kinetoplastid genomes where JBP-1 is found, apart from *Strigomonas culicis*, and *Perkinselia sp*. which only contain JBP-1 family members.

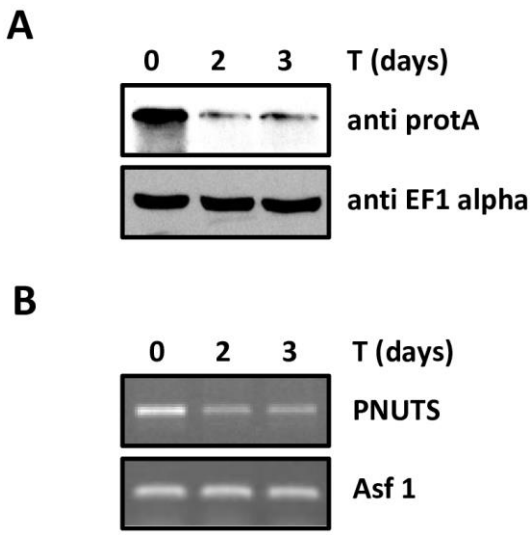


Fig. A.S5 PNUTS protein ablation upon RNAi induction. PNUTS RNAi cells with an endogenously WT PNUTS PTP tagged allele were used. Total protein and RNA was isolated after 2 and 3 days of RNAi induction and analyzed for the loss of PNUTS protein (protein A Western) and PNUTS mRNA (RT-PCR).

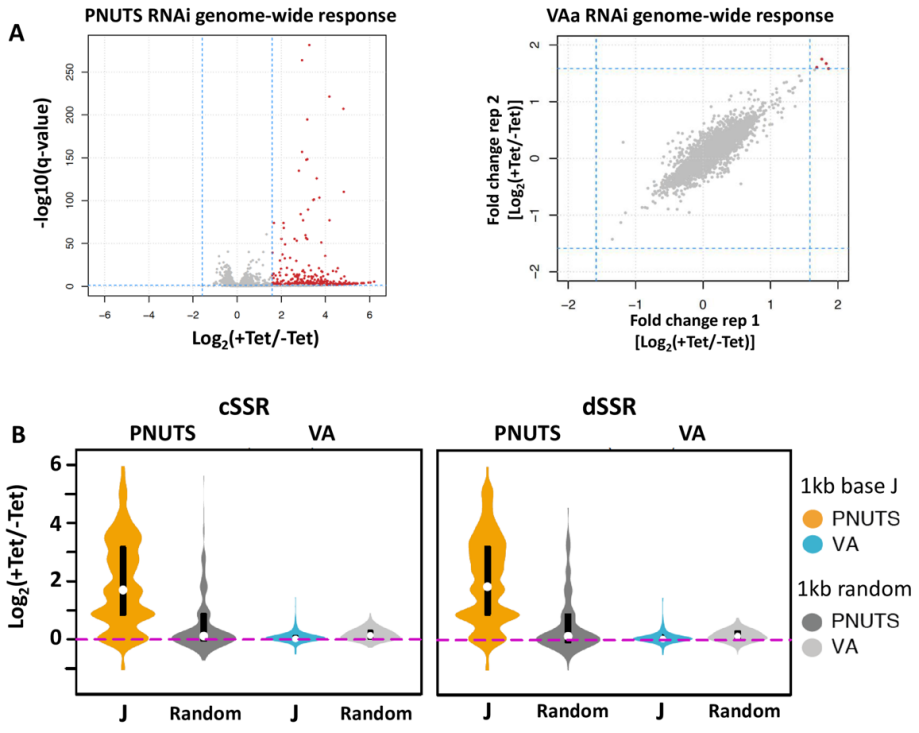


Fig. A.S6 Consequence of PNUTS depletion on the *T. brucei* transcriptome. (A) Gene expression changes upon RNAi ablation of PNUTS (left) and VA a (right) are plotted. Triplicate analysis of PNUTS and duplicate of VA a. On the left, red dots represent genes with greater than 3-fold change that are also significant with a Benjamini-Hochberg FDR test correction. On the right, red dots represent genes with greater than 2-fold change after VA a ablation in both replicates. (B) Gene expression changes at cSSRs (N = 193) and dSSRs (N = 197) that are within 1kb of base J matched with same number of random locations within the genome for ablation of PNUTS and VA a.

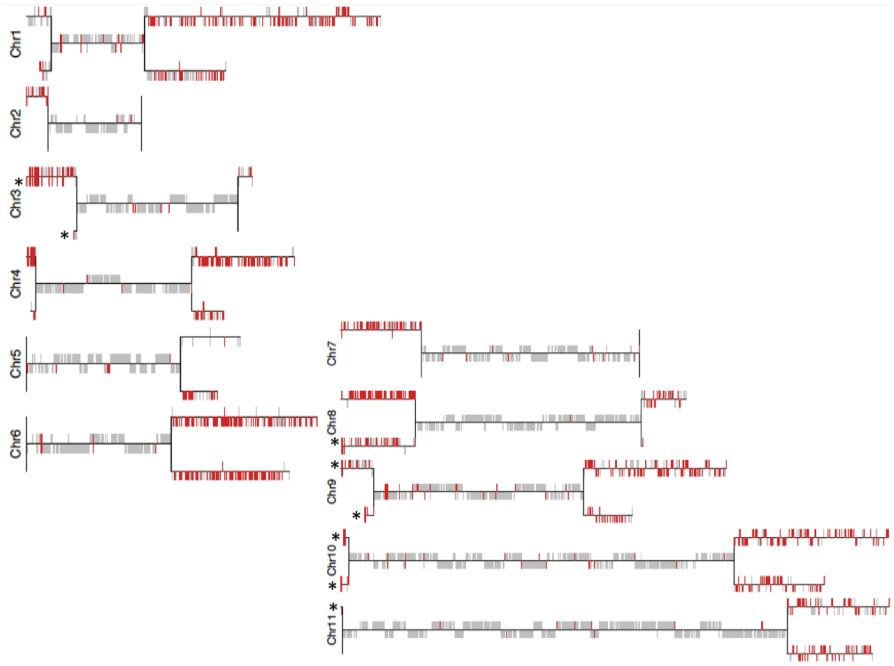


Fig. A.S7 Genes that are at least 3-fold upregulated upon the loss of PNUTS are located at the 3'- and 5'-end of PTUs and subtelomeric VSG clusters. Upregulated genes (>3-fold) upon PNUTS RNAi are indicated in red in the *T. brucei* 427 genome assembly. Only one of the two homologous chromosomes is depicted for the homologous core regions. Both chromosomes are shown for the heterozygous subtelomeric regions containing silent VSGs. The telomeric VSG expression sites are not included in this assembly. Metacyclic-form expression sites are marked with an asterisk.

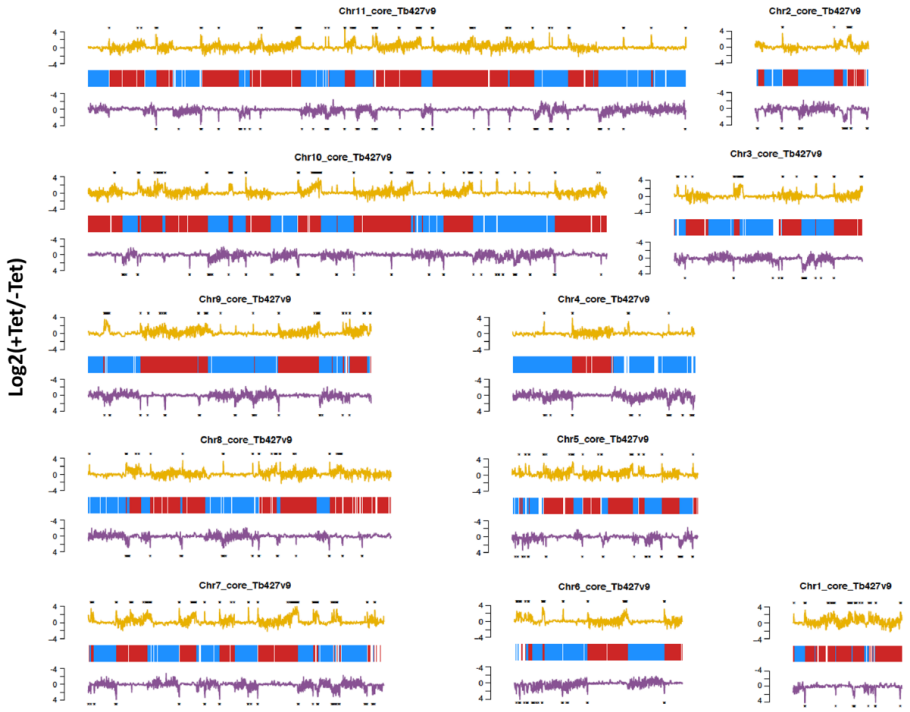


Fig. A.S8 Ablation of TbPNUTS accumulates antisense transcripts at PTU borders.

Transcription was measured by stranded RNA-seq. Fold changes comparing transcription levels between–and + tetracyclin induction of PNUTS RNAi were calculated in 200bp windows (100bp step) and plotted over the chromosome length. The core regions of the 11 chromosomes are shown. Forward (top strand) and reverse reads (bottom strand) were analyzed separately and plotted above and below the chromosome diagram, respectively. Blue genes are transcribed in reverse PTUs and red are forward PTUs. Asterix denote bins with greater than 2-fold change of expression following ablation.

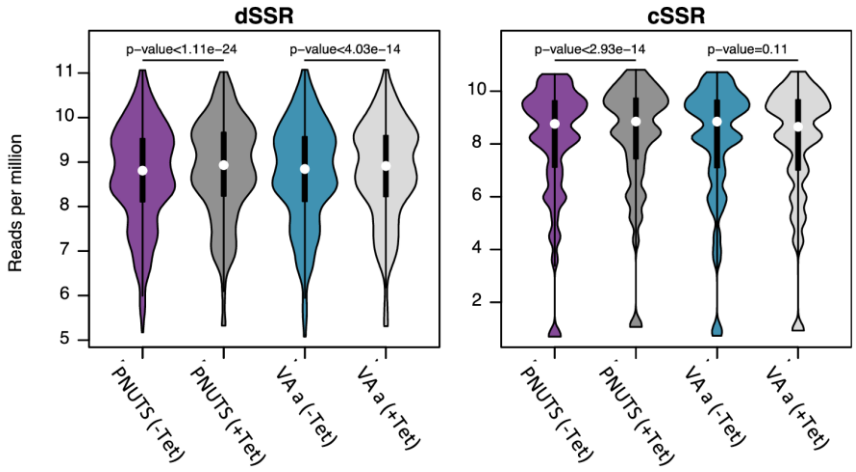


Fig. A.S9 Box plots comparing the levels of transcripts from dSSR and cSSR before and after TbPNUTS and VA a ablation. Normalized reads per million estimates were derived for dSSRs and cSSRs as the average across replicates per sample. Median values are indicated by white dots. Differences between + and–Tet were measured by a Mann-Whitney U statistical test.

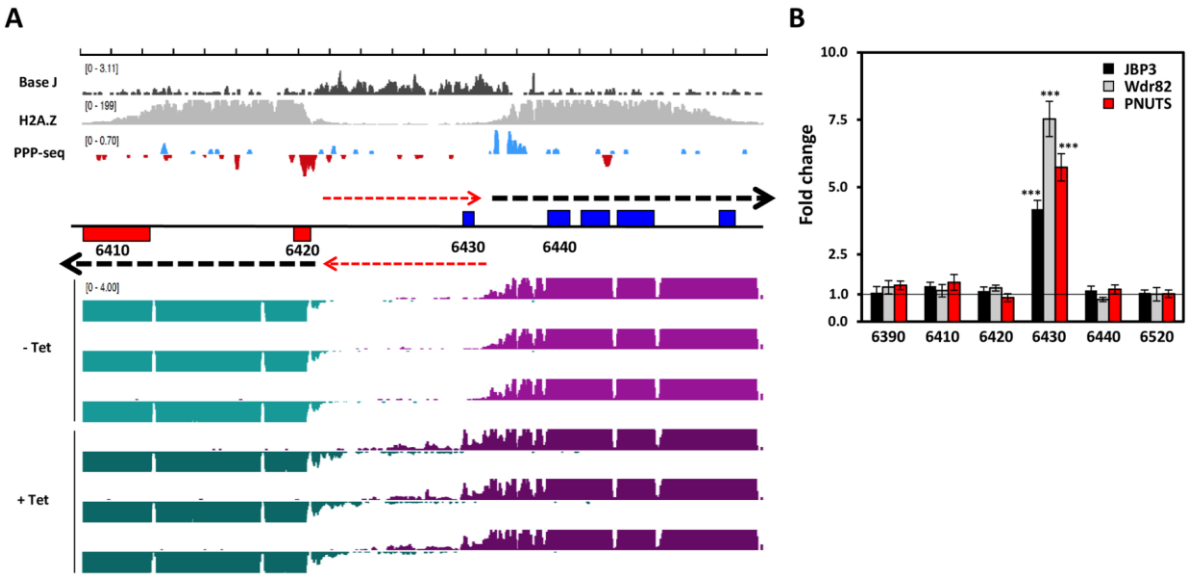


Fig. A.S10 TbPNUTS affects early termination of antisense transcription at TSSs. (A) Another divergent PTU region of chromosome 10 illustrating bi-directional transcription at TSSs upon TbPNUTS ablation. TSSs are denoted by PPP-seq and H2A.Z ChIP-seq enrichment in wild-type *T. brucei*. PPP-seq track colors: Red, reverse strand coverage; blue, forward strand coverage. RNA-seq track colors: Green, reverse strand coverage; Purple, forward strand coverage. Black arrows indicate direction of sense transcription. Red arrows indicate stimulated antisense transcription. In the case of the top strand, this 'antisense' transcription leads to derepression of the annotated 6430 gene (B) Confirmation of mRNA-seq transcript changes by RT-qPCR. RT-qPCR analysis was performed for the indicated genes as described in Fig A.5B. RT-PCR products for gene 6430 were confirmed by DNA sequencing.

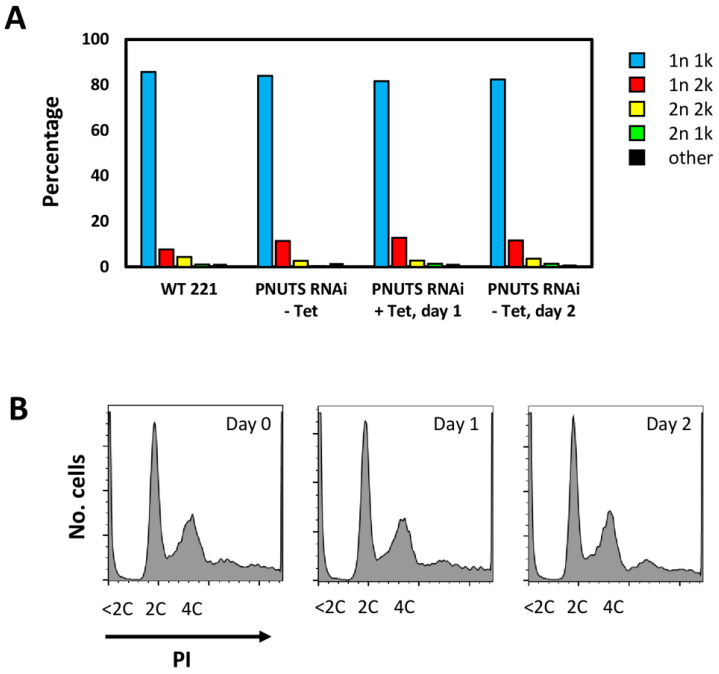


Fig. A.S11 Pervasive transcription does not affect DNA replication. (A) Cell cycle analysis of wild-type BSF *T. brucei* (WT 221) and PNUTS RNAi cells–and + Tet by DAPI. (B) Cell cycle analysis using flow cytometry. TbPNUTS RNAi cells were stained with Propidium Iodide (PI) and analyzed by flow cytometry.

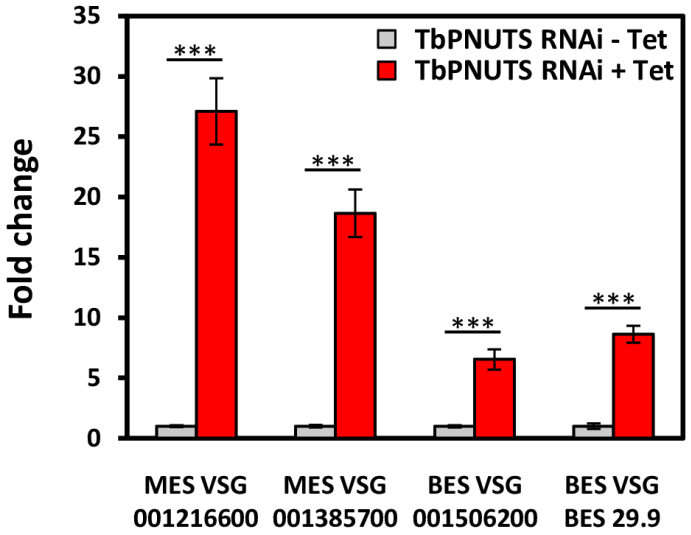


Fig. A.S12 Ablation of TbPNUTS results in de-repression of silent MES and BES VSGs.

qRT-PCR analysis of MES VSG 1954, MES VSG 559, BES VSG MITat 1.1 and BES VSG MiTat 1.8 expression upon TbPNUTS ablation. Error bars indicate standard deviation from at least three experiments. P values were calculated using Student's t test. ***, p value ≤ 0.001. All RT-PCR products were confirmed by DNA sequencing.

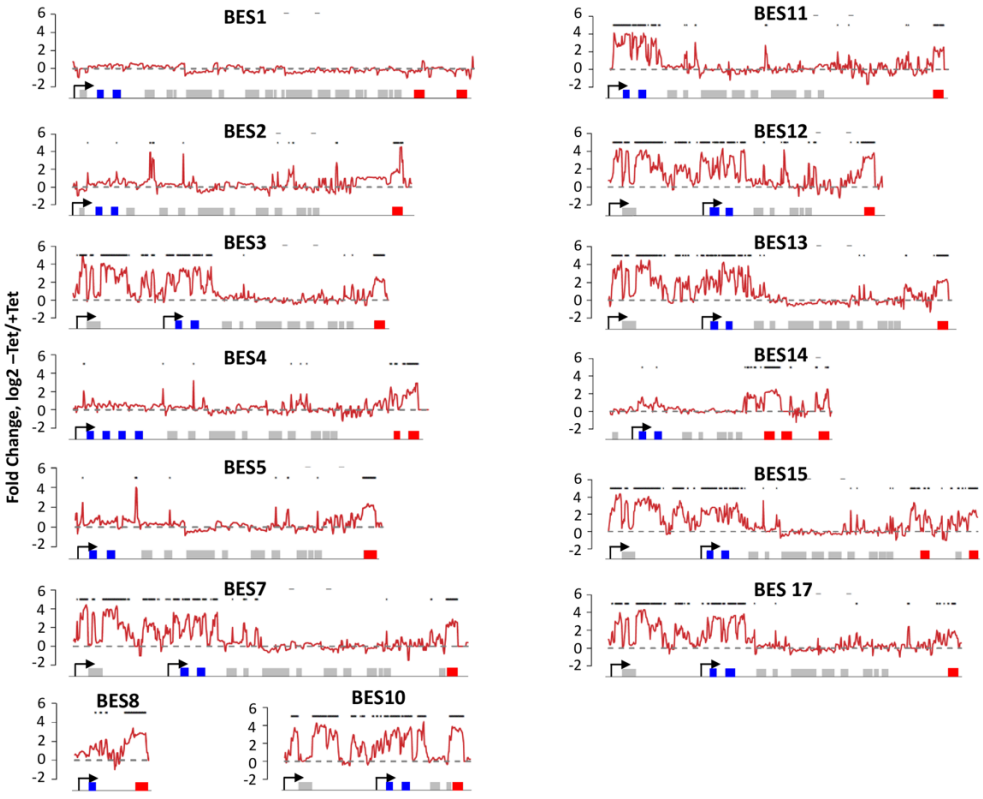


Fig. A.S13 TbPNUTS inhibits transcription of VSG BES. RNA-seq reads from the PNUTS RNAi were aligned to the T. brucei 427 BES sequences (14 BESs). Fold changes comparing plus and minus Tetracyclin were plotted over each BES as described in Fig A.6C.

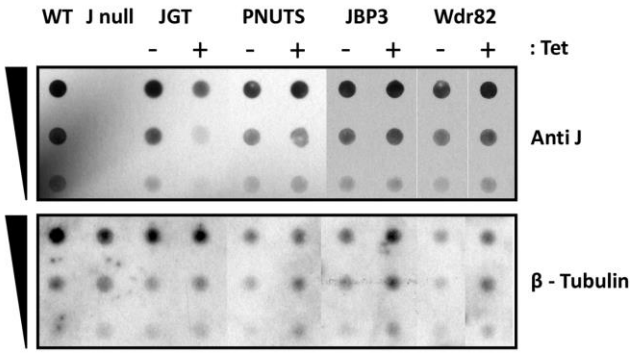


Fig. A.S14 PJW does not affect the synthesis of base J. DNA was isolated from the indicated T. brucei cell lines for anti-J dot blot analysis. +Tet indicates samples from day 2 RNAi induction. Samples were 2-fold serially diluted. The same blots were hybridized with a radioactive tubulin probe to control for DNA loading.

Table A1 Mass spectrometric identification of JGT purification products.

L. tarentolae proteins were identified by mass spectroscopy as described in Materials and Methods. Listed are proteins that were enriched at least 40-fold compared to the negative control purification of wildtype extract. The complete list of purified proteins is provided as S1 Table. Each protein is described by the systematic GeneDB name (http://www.genedb.org), annotation based on the genome database or homologies described in the text, molecular mass (kilodaltons), and *T. brucei* homologue.

Accession	Annotation	MW	*T. brucei* homologue
LtaP36.2450	JGT	101.3	Tb427.10.6900
LtaP33.1440	PNUTS	28.6	Tb427.10.11960
LtaP32.3990	Wdr82	41.5	Tb427tmp.01.8050
LtaP36.0380	JBP3	73.9	Tb427.10.4800
LtaP15.0230	PP1	42.3	Not identified

Table A.S1 JGT purification and gel based MS analysis in *L. tarentolae*. LtJGT was purified, fractionated on SDS-PAGE and proteins identified by mass spectrometry as in Table 1. Complete list of proteins identified in the JGT and the WT negative control purifications is shown. Proteins that were enriched at least 40-fold (based on PSMs) compared to the negative control purification are highlighted.

			JGT				WT			
Accession	Annotation	MW	Score	Peptides	Coverage	PSMs	Score	Peptides	Coverage	PSMs
LtaP36.2450	JGT	101.3	37058.3	60	63.9	3625	3957.3	10	11.6	75
LtaP33.1440	PNUTS	28.6	3698.4	8	42.8	61	0	0	0	0
LtaP32.3990	Wdr82	41.5	6721.1	14	36.4	185	215.8	1	2.6	1
LtaP36.0380	JBP3	73.9	5328.7	13	16.7	128	0	0	0	0
LtaP15.0230	PP1	42.3	4084.1	9	20.6	114	0	0	0	0
LtaP18.0230	RNA-binding protein, putative	43.5	1172.3	3	7.1	8	377.6	1	3.4	2
LtaP28.1290	Hypothetical protein, conserved	50.1	1066.7	3	7.9	12	0	0	0	0
LtaP28.1740	Sulfate transporter-like protein	217.8	470.3	2	0.9	2	575.4	2	0.9	14
LtaP27.1870	Hypothetical protein, unknown function (GO function: transcription initiation)	80.0	218.4	1	1.1	8	824.3	3	3.9	44
LtaP29.0900	High mobility group protein homolog tdp-1, putative	33.9	806.7	3	8.3	5	0	0	0	0
LtaP20.0260	Hypothetical protein, conserved	541.2	520.1	2	0.3	18	695.7	3	0.5	16

Table A.S2 Mass spectrometric identification of JGT, JBP3 and Wdr82 purification products. LtJGT, LtJBP3 and LtWdr82 proteins were purified and proteins in the soluble fraction identified by shotgun proteomics. Proteins that were enriched at least 10-fold (based on PSMs) compared to the negative control purification are highlighted. Pep: Peptides; Cov: Coverage.

			JGT				JBP3				Wdr82				WT			
Accession	Annotation	MW	Score	# Pep	% Cov	PSMs	Score	# Pep	% Cov	PSMs	Score	# Pep	% Cov	PSMs	Score	# Pep	% Cov	PSMs
LtaP36.2450	JGT	101.3	154.5	24	36.6	396	63.5	14	24.4	67	34.1	9	16.3	36	0	0	0	0
LtaP33.1440	PNUTS	28.6	24.3	6	35.2	35	40.5	9	39.8	127	53.1	10	46.6	310	0	0	0	0
LtaP32.3990	Wdr82	41.5	10.3	4	12.4	14	29.6	8	22.0	162	38.6	9	26.6	322	1.1	1	3.4	1
LtaP36.0380	JBP3	73.9	5.8	2	4.2	10	34.1	7	20.4	117	21.5	6	13.3	56	0	0	0	0
LtaP15.0230	PP1	42.3	0	0	0	0	12.3	4	8.6	18	28.0	6	16.3	56	0	0	0	0
LtaP33.0890	unspecified product	13.4	8.9	2	16.4	12	12.2	3	16.4	18	9.3	2	16.4	14	13.5	3	16.4	31
LtaP34.3340	Dna topoisomerase IB, large subunit	74.0	5.8	1	3.0	4	4.9	1	3.0	4	3.4	1	3.0	4	4	1	3.0	8

Table A.S3 PNUTS purification and MS analysis in *T. brucei*. TbPNUTS was purified and proteins in the soluble fraction identified by shotgun proteomics. List of PNUTS-PTP co-purified proteins, identified at 1% FDR by LC-MS/MS, with 10 or more PSMs are shown. Proteins that are enriched at least 40-fold compared to the negative control purification of wildtype extract and a score of at least 100 are highlighted. Included are proteins that are common contaminants of previous tandem affinity purifications (i.e. tubulin and elongation factor 1-alpha). Protein annotation and accession numbers are from the *T. brucei* 927 database at www.TriTrypDB.org.

			PNUTS				WT			
Accession	Annotation	MW	Score	Peptides	Coverage	PSMs	Score	Peptides	Coverage	PSMs
Tb927.10.11960	PNUTS	35.4	711.4	21	83	227	-	-	-	-
Tb927.1.2350	beta tubulin	49.7	596.3	22	60	185	-	-	-	-
Tb927.1.2360	alpha tubulin	49.8	461.1	24	61	132	-	-	-	-
Tb927.10.4800	JBP3	65.5	354.7	27	48	110	-	-	-	-
Tb927.11.16360	Wdr82	37.9	307.9	11	48	80	-	-	-	-
Tb11.v5.1035	heat shock protein 70 (hsp70), putative	75.2	156.5	21	40	44	-	-	-	-
Tb927.9.15400	ankyrin-repeat protein, putative	334.1	141.9	9	5	67	81.5	8	6	32
Tb927.10.2110	elongation factor 1-alpha	49.1	140.7	7	31	34	-	-	-	-
Tb927.6.4280	glyceraldehyde 3-phosphate dehydrogenase, glycosomal	39	94.8	6	24	35	484.3	3	12	204
Tb927.2.4950	hypothetical protein, conserved	155.8	59.6	5	7	25	73.7	5	7	32
Tb927.9.4350	hypothetical protein, conserved	94	73.4	2	5	24	40.0	4	11	13
Tb927.6.3750	heat shock 70 kDa protein, mitochondrial precursor, putative	71.4	84.7	11	23	23	-	-	-	-
Tb927.11.4500	hypothetical protein, conserved	101.9	52.5	5	10	20	94.6	5	9	35
Tb927.11.1130	calpain-like cysteine peptidase, putative	613.6	54.4	8	3	19	-	-	-	-
Tb927.10.15410	glycosomal malate dehydrogenase	33.7	50.9	3	16	18	59.6	4	19	20
Tb927.6.3220	hypothetical protein, conserved	115.6	47.4	1	2	18	-	-	-	-
Tb927.2.2650	hypothetical protein, conserved	367.5	54.6	5	3	17	60.4	5	3	17
Tb927.6.2580	hypothetical protein, conserved	77.9	40.1	3	6	17	-	-	-	-
Tb927.5.1450	receptor-type adenylate cyclase GRESAG 4, putative	135.7	50.7	8	13	16	-	-	-	-
Tb927.4.3530	hypothetical protein, conserved	84.3	38.8	4	10	16	-	-	-	-
Tb927.11.2990	KREPB4	46.4	33.1	1	2	16	-	-	-	-
Tb927.11.3250	dynein heavy chain, putative	530.8	42.3	12	5	15	36.5	9	4	12
Tb927.8.5580	N-terminal region of Chorein, a TM vesicle-mediated sorter/Integral peroxisomal membrane peroxin, putative	526.4	40.8	10	4	15	24.0	6	3	9
Tb927.10.4200	hypothetical protein, conserved	100	40.1	2	4	15	-	-	-	-
Tb927.11.1340	Flagella connector protein 1	127.8	39.1	5	10	14	-	-	-	-
Tb927.10.13350	hypothetical protein, conserved	121.3	37.4	2	4	14	-	-	-	-
Tb927.10.5350	dynein heavy chain, putative	474.4	40.2	7	4	13	-	-	-	-
Tb927.10.5850	N-terminal region of Chorein, a TM vesicle-mediated sorter, putative	519.4	38.6	10	4	13	-	-	-	-
Tb927.5.3230	hypothetical protein, conserved	240.8	37.7	5	4	13	-	-	-	-
Tb927.4.620	hypothetical protein, conserved	276.2	34.5	10	7	13	-	-	-	-