bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Crosstalk between RNA Pol II C-Terminal Domain Acetylation and Phosphorylation via RPRD Proteins

Ibraheem Ali[1,2], Diego Garrido Ruiz[3], Zuyao Ni[4], Jeffrey R. Johnson[1], Heng Zhang[6], Pao-Chen Li[1], Ryan J. Conrad[1,2], Xinghua Guo[4], Jinrong Min[6], Jack Greenblatt[4], Matthew Jacobson[3, 5], Nevan J. Krogan[1, 5], Melanie Ott[1,2]

Affiliations:

1 J. David Gladstone Institutes, San Francisco, California, 94158, USA

2 Department of Medicine, University of California San Francisco, San Francisco, California, 94143, USA

3 Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, 94143, USA

4 Donnelly Centre, University of Toronto, Toronto, Ontario, Canada

5 California Institute for Quantitative Biosciences (QBC) at UCSF, San Francisco, California, 94143, USA

6 Structural Genomics Consortium, University of Toronto, Ontario, Canada

Correspondence: melanie.ott@gladstone.ucsf.edu

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SUMMARY

Multiple posttranslational modifications of the RNA polymerase II C-terminal domain (CTD) coordinate passage through the transcription cycle. The crosstalk between different modifications is poorly understood. Here, we show how acetylation of lysine residues at position 7 of characteristic heptad repeats (K7ac), a modification only found in higher eukaryotes, regulates phosphorylation of serines at position 5 (S5p), a conserved mark of polymerases initiating transcription. Using mass spectrometry, we identified Regulator of Pre-mRNA Domain-containing (RPRD) proteins as reader proteins of K7ac. K7ac enhanced *in vitro* binding of CTD peptides to the CTD-interacting domain (CID) of RPRD1A and RPRD1B proteins and inversely regulated S5p levels genome-wide. Treatment with deacetylase inhibitors globally enhanced levels of K7ac- and decreased levels of S5-phosphorylated polymerases ≥500 base pairs downstream of transcription start sites of expressed genes, consistent with acetylation-dependent S5 dephosphorylation via a previously identified RPRD-associated S5 phosphatase. Consistent with this model, RPRD1B knockdown increased S5p, but also enhanced K7ac levels, indicating that RPRD proteins recruit a K7 deacetylase. Collectively, our data identify RPRD CIDs as K7ac reader domains and reveal auto-regulatory crosstalk between K7ac and S5p via RPRD proteins at the transition from transcription initiation to elongation in higher eukaryotes.

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

The RNA Polymerase II (Pol II) complex is highly conserved in all eukaryotic cells and responsible for the production of most gene expression products (Buratowski, 2003; Eick and Geyer, 2013). RPB1, the largest subunit of the complex, contains the catalytic core of the complex and a unique regulatory region called the C-terminal domain (CTD). In eukaryotes, the CTD is composed of twenty or more repeats with a heptad consensus sequence, $Y_1S_2P_3T_4S_5P_6S_7$, which is highly conserved from yeast to human. In multicellular eukaryotes, the CTD is expanded and contains a varying number of non-consensus repeats depending on the organism (Chapman et al., 2008). The 52 repeats of the mammalian CTD can be divided into 21 consensus repeats proximal to the enzymatic core, and 31 non-consensus repeats distal from the core with less fidelity to the consensus. Divergence from the consensus sequence most commonly occurs at position 7, which can be replaced with an asparagine (N), threonine (T), or a lysine (K) instead of the consensus serine (Eick and Geyer, 2013). The CTD is intrinsically disordered and functions as an interaction platform for accessory proteins required for transcription and transcription-associated RNA processing events (Buratowski, 2009; Jasnovidova and Stefl, 2013).

The heptad repeats within the CTD are extensively and dynamically post-translationally modified at different times during the transcription cycle. Of the seven consensus CTD residues, 5 can be phosphorylated (Y1, S2, T4, S5, and S7). The two remaining proline residues can undergo isomerization into *cis* or *trans* conformations (Heidemann et al., 2013). Serine-5 phosphorylation (S5p) and serine-2 phosphorylation (S2p) are the most thoroughly studied CTD modifications (Buratowski, 2009; Jasnovidova and Stefl, 2013). Serine-5 is phosphorylated by the cyclin-dependent kinase 7 (CDK7) subunit of general transcription factor TFIIH, is enriched at promoters, and decreases successively towards the 3' end of genes (Brookes et al., 2012; Ebmeier et al., 2017). The phosphorylated serine-2 mark, placed by several kinases (CDK9, CDK12, CDK13 and BRD4), starts to accumulate downstream of transcription start sites and steadily increases towards the 3' ends of genes, reflective of its critical role in productive polymerase elongation (Bartkowiak et al., 2010; Devaiah et al., 2012; Nechaev and Adelman, 2011). The distribution of consensus Pol II modifications is best studied in yeast, revealing a fixed transition point from S5p to S2p enriched, on average, 450 base pairs (bp) downstream of transcription start sites (TSS) (Bataille et al., 2012; Kim et al., 2010; Mayer et al., 2010; Tietjen et al., 2010). Similar to S5p, Serine-7 phosphorylation (S7p) is catalyzed by CDK7, is enriched near promoters and in gene bodies, and regulates the expression snRNA genes (Brookes et al.,

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2012; Egloff et al., 2012). Tyrosine-1 phosphorylation is enriched near promoters, and has been linked to enhancer and antisense transcription (Descostes et al., 2014). Threonine-4 phosphorylation is enriched in coding regions and is required for cell viability and transcription termination (Hintermair et al., 2012).

Posttranslational modifications (PTMs) specifically found in non-consensus repeats include asymmetric dimethylation of a single arginine (R1810me2), conserved among some metazoa, that regulates transcription of small nuclear and nucleolar RNAs (Sims et al., 2011). In addition, lysine residues at position 7 of eight heptad repeats are acetylated by the acetyltransferase p300/CBP (KAT3A/B) (K7ac), and were recently also found to be mono- and di-methylated by a yet unknown methyltransferase (Dias et al., 2015; Schroder et al., 2013; Voss et al., 2015; Weinert et al., 2018). These lysine residues evolved in higher eukaryotes in the common ancestor of the metazoan lineage, and are highly conserved among vertebrates (Simonti et al., 2015). While lysine-7 mono- and di-methlyation marks are found near promoters, K7ac is enriched in gene bodies (Dias et al., 2015). K7 residues are required for productive transcription elongation of immediate early genes in response to epidermal growth factor stimulation (Schroder et al., 2013). Importantly, K7ac marks are found at ~80% of actively transcribed genes, with a peak in signal +500 bp downstream of the TSS, indicating that the modification could more broadly regulate the transition from transcription initiation to productive elongation (Schroder et al., 2013). In a genetic model where all eight K7 residues were mutated to arginines (8KR), cells expressing 8KR RPB1 exhibited altered expression of genes relating to development, multicellularity and cell adhesion, underscoring a critical role of K7ac in the development of higher eukaryotes (Simonti et al., 2015).

Effector proteins interacting with differentially modified CTDs often contain a so-called CTD-interacting domain (CID), which is one of the best-studied CTD-binding modules and is conserved from yeast to humans (Ni et al., 2011). The mammalian Regulator of Pre-mRNA Domain-containing (RPRD) proteins 1A, 1B and RPRD2 proteins are homologues of the yeast transcription termination factor *Rtt103*, and each contains a CID (Ni et al., 2011). *Rtt103* and RPRD CIDs can bind CTD peptides carrying S2p, but not S5p; S7p and unmodified K7 residues reside at the edge of the CID binding cleft, and can be substituted without altering the binding affinity (Jasnovidova et al., 2017a; Meinhart and Cramer, 2004; Ni et al., 2014). RPRD1A and RPRD1B are found in macromolecular complexes that associate with Pol II and transcription regulatory factors, including the S5-phosphatase RPAP2 (Liu et al., 2015; Morales et al., 2014; Ni et al., 2011; Ni et al., 2014; Patidar et al., 2016). RPRD1A, also called P15RS, regulates

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

G1/S cell cycle progression and suppresses Wnt and β-catenin signaling via interactions with the class I lysine deacetylase HDAC2 and transcription factor 4 (TCF4) (Jin et al., 2018; Liu et al., 2015; Liu et al., 2002; Wu et al., 2010). RPRD1B, also called CREPT, was identified in a mass spectrometry-based screen for mammalian Pol II-interacting proteins; it is upregulated in various cancers, and regulates genome stability and transcription termination (Lu et al., 2012; Morales et al., 2014; Patidar et al., 2016; Zhang et al., 2018). Although the homology with *Rtt103* implies a conserved role in transcription termination and explains why the proteins are enriched in 3' ends of eukaryotic genes, an additional less well-defined role of RPRD proteins has emerged at 5' ends of genes in higher eukaryotes. This involves a mechanism to regulate genome stability through the resolution of R-Loops, which are DNA-RNA hybrids (Lu et al., 2012) as well as the recruitment of RPAP2 to initiating RNA Pol II (Ni et al., 2014).

In this study, we provide molecular insight into the role of RPRD proteins at the 5' ends of genes and newly connect RPRD proteins with K7ac. We find that RPRD proteins via their CIDs specifically interact with K7ac, and that this interaction promotes S5-dephosphorylation at and beyond +500 bp downstream of the TSS. This data supports a model where vertebrates evolved specific crosstalk between S5p and K7ac to ensure precise transcription initiation dynamics and a timely transition to a productive elongation phase at a defined distance to the TSS.

RESULTS

Preferential Binding of RPRD Proteins to Acetylated RPB1

To identify proteins that interact with Pol II K7ac, we performed stable isotope labeling with amino acids in cell culture (SILAC). We overexpressed HA-tagged RPB1 proteins, either wild type or 8KR mutant, in HEK293T cells. The proteins also contained a known α-amanitin resistance mutation enabling propagation of successfully transfected cells in the presence of α-amanitin, which induces the degradation of endogenous Pol II (Bartolomei and Corden, 1987). After culture of cells in differential metabolic labeling media, RPB1-containing complexes were purified via their HA tag and subjected to mass spectrometric analysis (**Fig. 1A**). We found all members of the RPRD family preferentially bound to wildtype RBP1, including RPRD1A, RPRD1B, RPRD2 along with several of their interacting partners such as RPAP2, RPAP3, MCM7 and RUVB1, which were previously identified by mass spectrometry (Ni et al., 2011; Patidar et al., 2016) (**Fig. 1B**).

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As RPRD1B is the most thoroughly studied of the RPRD family, we confirmed preferential binding of RPRD1B to wild type relative to 8KR mutated RPB1 in repeated co-immunoprecipitation experiments by performing either pull down of endogenous RPRD1B and blotting for HA-RPB1 protein or the reverse (**Fig. 1C**). The enrichment of endogenous RPRD1B proteins after wild type, and not mutant, HA-RPB1 immunoprecipitation was consistent among four independent experiments and statistically significant ($p = 0.0098$) (**Fig. 1D**). We also tested interaction between endogenous RPB1, RPRD1A and RPRD1B proteins in NIH3T3 cells treated with lysine deacetylase (KDAC) inhibitors. KDAC inhibitor treatment induced robust hyperacetylation of endogenous RPB1 in input material as tested with an antibody specific for K7ac (Schroder et al., 2013), but did not change total overall RPB1 protein levels (**Fig. 1E**). Following pulldown of endogenous Pol II, more RPRD1A and RPRD1B proteins were recovered when cells were treated with KDAC inhibitors as compared to vehicle-treated cells, confirming positive regulation of the RPB1:RPRD interaction by acetylation **(Fig. 1E and 1F)**.

Next, we tested *in vivo* recruitment of RPRD1B to a known target gene, *Leo1* (Ni et al., 2011). Using chromatin immunoprecipitation (ChIP) followed by quantitative PCR, we found RPRD1B recruitment to the *Leo1* promoter (+33 bp) consistently enhanced in NIH3T3 cells treated with KDAC inhibitors as compared to vehicle-treated cells (**Fig. 1G).** Similar to what we observed by western blotting, K7 residues were hyperacetylated at the Leo1 promoter in response to KDAC inhibition in ChIP analysis with the K7ac-specific antibody. Importantly, total Pol II occupancy did not increase under KDAC inhibition, confirming specific K7 hyperacetylation and enhanced RPRD1B recruitment in response to KDAC inhibition (**Fig. 1G**).

Direct Interaction of K7ac with RPRD CTD-Interacting Domains (CIDs)

To test whether K7ac modulates CTD binding to the RPRD CID domains, we performed isothermal titration calorimetry (ITC) to test the interaction between synthetic CTD peptides and purified CID domains from both RPRD1A and RPRD1B proteins. We generated CTD peptides spanning roughly 3 heptad repeats (20 amino acids) with repeat 39 at the center. This region was chosen as it is acetylated and phosphorylated *in vivo* (Voss et al., 2015; Weinert et al., 2018), and contains two consecutive K7 residues (**Fig. 2A**). Peptides were synthesized in an unmodified (UnM), acetylated (K7ac), or phosphorylated (S2p and S5p) state. S2p was included as a positive control as it enhances CTD:CID interactions, while S5p served as a negative control (Ni et al., 2014; Pineda et al., 2015). In addition, we combined S2p and S5p with K7ac to investigate potential combined effects. Compared to the unmodified CTD, binding of the RPRD CIDs to CTD peptides carrying K7ac had a significantly lower K_d (2.3-fold reduction for RPRD1A

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and 3.8-fold for RPRD1B), indicating enhanced binding. (**Fig. 2B–D**). S2p itself had a robust effect in enhancing CID binding, as previously observed, but combining K7ac with S2p further decreased the K_d by 2.8-fold and 4.2-fold, respectively. S5p-carrying peptides did not interact with CID proteins as expected (data not shown). Together these data indicate that K7ac enhances the interaction of RPRD proteins with the Pol II CTD with and without additional S2p marks.

To better understand the mechanism behind K7ac-dependent stabilization of binding between the CID and CTD peptides, we performed molecular modeling using previously published RPRD1B CID structures bound to CTD peptides [pdb: 4Q94 (dimer) and 4Q96 (tetramer)]. RPRD protein dimerization is believed to occur through coiled coil domain interactions that are not present in these structures (Mei et al., 2014; Ni et al., 2014), which nonetheless dimerize and tetramerize by domain swapping. We proceeded with *in silico* analyses using both structures and searched for consistencies between both. Because phosphorylation and acetylation change the net charge of the peptide fragment, we first calculated the electrostatic potential of the CID structure to investigate the charge distribution along the binding cleft (Dolinsky et al., 2004). We found that the recognition module within the CID, both in the dimeric and tetrameric structures, has a positively charged binding pocket (**Fig. 2E–H, Fig. S1A–D**), which will enhance the binding of peptides with S2p. Similarly, acetylation, and thus neutralization, of the positively charged lysine residues favored interaction with this binding pocket by enhancing electrostatic stability.

We also obtained 20 ns molecular dynamics trajectories to refine the models and identify residues within the CID that directly bind to S2p and K7ac, and thus contribute to the recognition of these post-translational modifications. Through these simulations, we reproduced the previously reported coordination between S2p and arginine-106 (R106), and observed that the two K7 residues in the acetylated state formed transient interactions with nearby CID residues (**Fig. 2I and 2J**). In particular, the first acetylated K7 residue in the CTD peptide formed transient hydrogen bonds with three CID residues (N18, N54 and Y61) in the same vicinity. The second acetylated K7 residue interacted with CID residues at the other end of the binding cleft (Q24, Q68, N69 and R72) (**Fig. 2I and 2J)**. Similar results were obtained in simulations of the tetrameric structure, which also showed coordination between S2p and R106 along with similar transient hydrogen-bonding between K7ac and several residues in the CID (N18, Q20, E92 and K96) (**Fig. S1E and S1F**). Together, the transient interactions coupled to the overall positive

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

electrostatic potential around the recognition module explain the stabilization of K7ac-modified CTD peptides over the non-acetylated or unmodified peptides.

Increased K7ac Correlates with Reduced S5p Downstream of Transcription Start Sites

RPDR1 proteins are known to interact with RPAP2, the mammalian homolog of yeast *Rtr1* and a known S5 phosphatase (Ni et al., 2011). We tested the influence of K7ac on S5p levels by performing ChIP-seq with antibodies specific for K7ac, S5p and total unmodified (8WG16) Pol II in chromatin isolated from NIH3T3 cells treated with KDAC inhibitors. Average TSS–anchored occupancy profiles were generated for all expressed genes (Ramirez et al., 2016) and normalized to genomic background signal (**Fig. S2A-C**). KDAC inhibition increased genome wide Pol II-K7ac occupancy as expected, but interestingly only from ≥500 bp downstream of the TSS onwards. It decreased K7ac levels proximal to this in an area, where total Pol II occupancy was strongly enhanced (**Fig. 3A and 3B**). In this region, S5p levels were also increased as compared to vehicle-treated cells, while beyond the ≥500 bp mark S5p levels decreased below the level of control cells, mirroring enhanced K7ac (**Fig. 3C**). While total Pol II occupancy increased immediately downstream of the TSS in response to KDAC inhibition, potentially explaining the TSS-proximal S5p enhancement, beyond +500–1000 bp, total Pol II occupancy on average remained unchanged, underscoring that the inverse relationship between K7ac and S5p at this point was not confounded by changes in total Pol II levels.

A focused analysis of known target genes of RPRD1B such as *Leo1* and *Cyclin D1* (Lu et al., 2012; Ni et al., 2011) showed corresponding profiles with lowered S5p and enhanced K7ac levels downstream of the TSS in response to KDAC inhibition (**Fig. 3D**). But ~10% of actively expressed genes, including *Tub1a1* and *Mdm2*, did not show a downregulation of S5p in response to KDAC inhibition despite strong upregulation of K7ac, indicating that these genes are not controlled by RPRD proteins but possibly alternative mechanisms (**Fig. 3E**). Unfortunately, we could not examine the occupancy of RPRD proteins genome-wide as the available antibodies did not show a sufficient signal-to-noise ratio in ChIP-seq experiments (data not shown).

RPRD1B Controls Genes Involved in Multicellularity, Development and Cell Adhesion

Overexpression of RPRD proteins has been shown to decrease S5p levels at the *Leo1* gene consistent with the model that RPRD proteins recruit the S5 phosphatase RPAP2 (Ni et al., 2011). We now performed the inverse experiment and knocked down RPRD1B in NIH3T3 cells using lentiviral shRNAs. A 50% knockdown efficiency was sufficient to induce global S5

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hyperphosphorylation as observed by western blotting, indicating a critical role of RPRD1B in overall S5 dephosphorylation. Surprisingly, we also observed a consistent upregulation of K7ac levels in RPRD1B knockdown cells, pointing to the recruitment of a K7 deacetylase by RPRD proteins in addition to the S5 phosphatase **(Fig. 4A)**. No change in total Pol II levels was observed.

RNA-seq on RPRD1B knockdown cells identified 271 differentially expressed genes as compared to control shRNA-treated cells (**Fig. 4B, Table S1**). RPRD1B was among the most significantly downregulated genes with an mRNA knockdown efficiency of 41% ($p = 0.0019$; **Fig. 4E**), similar to what was observed for protein expression. Gene Ontology analysis on dysregulated genes indicated that RPRD1B knockdown induced changes in genes related to developmental processes, multicellular organismal development and cell adhesion, consistent with previous findings that complete knockout of the factor causes embryonic lethality (Morales et al., 2014) **(Fig. 4C).** Furthermore, this is consistent with studies indicating K7ac specifically evolved in higher eukaryotes, and regulates developmental genes with significant enrichment for evolutionary origins in the early history of eukaryotes through early vertebrates (Schroder et al., 2013; Simonti et al., 2015).

The majority of dysregulated genes associated with multicellular organismal processes was upregulated in response to RPRD1B knockdown (69.7%). Examples include *Nov*, an immediate-early gene important for regulating proliferation and development, the Prolactin 2c2 gene *Prl2c2* involved in embryonal development, the fibroblast growth factor 10 gene *Fgf10* necessary for organogenesis, and the SLIT homolog 2 gene *Slit2* involved in neural ECM-mediated signaling (**Fig. 4D**). Examples of downregulated genes include the ECM-associated sulfatase 1 gene *Sulf1* and neural cell adhesion molecule *Ncam1* **(Fig. 4E)**. *Tnik,* an essential activator of the Wnt signaling pathway, was also down-regulated (**Fig. 4E**), but most other Wnt-regulated genes remained unchanged (**Table S1**). Together, these data identify RPRD1B as a regulator of genes involved in multicellular organismal development and further support the model that RPRD proteins are relevant reader proteins of the K7ac mark in higher eukaryotes.

RPRD1B Knockdown Perturbs both K7ac and S5p Marks Genome-Wide

Next, we performed ChIP-seq in RPRD1B knockdown cells using antibodies against K7ac, S5p and total Pol II. The most striking finding was the induction of a distinct TSS-proximal increase in K7 acetylation with minimal changes to total unmodified (8WG16) Pol II occupancy (**Fig. 5A and 5B, Fig. S3A and S3B**). This is consistent with the observation that K7ac was

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

induced upon RPRD1B knockdown in western blot experiments and underscores a model where RPRD1B recruits a K7 deacetylase that counterbalances K7 acetylation within the first 500–1000 bp of mRNA production (**Fig. 5A**). This peak in K7ac levels was mirrored by a TSS-proximal decrease in S5p, possibly through residual RPRD (and RPAP2) proteins remaining in the RPRD1B knockdown cells inducing S5 dephosphorylation in response to K7 hyperacetylation (**Fig. 5C, Fig. S3C).** Beyond this TSS-proximal region, S5p levels were elevated relative to cells treated with control shRNAs, consistent with the observation that reduced RPRD1B levels globally induce S5 hyperphosphorylation due to the lack of S5 phosphatase recruitment.

When RPRD1B knockdown cells were treated with KDAC inhibitors, the inverse relationship between K7ac and S5p observed after either KDAC inhibitor treatment or RPRD1B knockdown were no longer detected (**Fig. 5D-F).** This indicates that RPRD1B plays a critical role in mediating the changes in S5p levels observed after KDAC inhibition. It further supports the model that RPRD proteins recruit a KDAC in addition to a phosphatase, considering that knockdown of RPRD1B and KDAC inhibition each increased K7 acetylation to the same level, and no further increase was observed when both were applied together. Collectively, these results underscore the close link between acetylation of K7 residues and dephosphorylation of S5 via RPRD proteins at a unique distance from the TSS (+500-1000 bp) and point to the importance of an RPRD-associated K7 deacetylase.

DISCUSSION

In this study, we report a new molecular function for K7ac in terminating S5p in the transition from transcription initiation to elongation. We show that K7ac enhances the recruitment of RPRD proteins to the initiated Pol II complex in order to facilitate S5 dephosphorylation; this occurs presumably via RPAP2, their interacting S5 phosphatase **(Fig. 6A)**. Surprisingly, we found that lack of RPRD1B protein expression also increased K7ac levels, indicating that in addition to binding an S5 phosphatase, RPRD proteins may also recruit a K7 deacetylase. This provides a unique autoregulatory mechanism as binding to RPRD proteins to K7ac ultimately leads to the removal of the mark. As previous studies have highlighted the importance of S2p in enhancing the interaction between the CTD and RPRD CID domains (Ni et al., 2014; Pineda et al., 2015), the emergence of S2p downstream of K7ac may serve to maintain RPRD recruitment to complete S5 dephosphorylation during the early phase of

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transcription elongation (Ni et al., 2014). When levels of K7ac were perturbed –by KDAC inhibitor treatment (**Fig. 6B)** or RPRD1B knockdown (**Fig. 6C**)– recruitment of the S5 phosphatase was either enhanced, resulting in increased S5 dephosphorylation and lower S5p levels genome-wide, or recruitment was diminished, enhancing S5p levels, respectively. Therefore, these data support a model in which dynamics of K7 acetylation evolved to blunt the peak of S5 phosphorylation at a precise distance from the TSS in higher eukaryotes, likely facilitating the transition between transcription initiation and productive elongation.

We previously observed a transient enrichment in the occupancy of K7-acetylated Pol II located approximately +500 bp downstream of the TSS when we normalized K7ac peaks to total Pol II occupancy on expressed genes (Schroder et al., 2013). This fits well with the self-limiting nature of K7ac wherein the modification recruits its own terminating enzyme via the RPRD reader proteins. Here we observed characteristic changes in S5p levels at and beyond the +500 bp mark that support the model that K7ac and S5p are inversely correlated. The +500 bp mark further corresponds well with the proposed hand-off site between S5p and S2p previously determined in yeast, underscoring that higher eukaryotes may have evolved K7ac to maintain this transition at the same genomic position (Mayer et al., 2010). In our current study, peaks proximal to the +500 bp mark behaved on average differently than peaks at or beyond the mark. Possible explanations are: A) RPRD proteins are recruited specifically to the +500 bp location (at the peak of K7ac) and exert their effect on S5p in this region and beyond; B) the balance between S5 phosphorylation and K7 acetylation immediately downstream of the TSS is shifted towards S5p and efficient dephosphorylation can only occur after CDK7 levels are additionally lowered beyond the TSS (Ebmeier et al., 2017). C) It is the transition of Pol II into elongation and the occurrence of S2p at the transition point that allows for efficient RPRD association and efficient S5 dephosphorylation; D) the RPRD complex changes before and after the +500 bp mark with the K7 KDAC highly enriched closer to the TSS, and less enriched beyond +500 bp. We envision that multiple of the mechanisms might be in place to explain the observed changes. Interestingly, similar peak enrichments of CTD PTMs as we describe for the sense strand of transcription were observed along divergent transcription of the negative strand of DNA, the significance of which remains unknown.

An important finding of our study is that K7ac enhances binding of the RPRD CIDs to CTD peptides. We show that the affinity of the CID:K7ac interaction is ~90 μM, which lies in the range of acetyl-lysine interactions with bromodomains, the latter being the “classical” Kac recognition domain (Muller et al., 2011). Interestingly, the K7ac:CID interaction accommodates

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

additional phosphorylation marks such as S2p. This supports previous findings that the CID domain creates a positively charged "channel" in which the CTD peptide is dynamically situated depending on its PTM status (Jasnovidova et al., 2017a; Ni et al., 2014). K7ac recognition occurs by electrostatic and hydrogen bonding with various residues of the CID. Principles of specific recognition of phosphorylated amino acids have been well-studied and are consistent with the features we have highlighted for S2p. Lysine acetylation has received less study; our results suggest that side chains containing amide groups (N and Q) play an important role, forming transient hydrogen bonds with the amide group of acetylated K7 residues. It is especially interesting that N18, N69 and Q24, which form hydrogen bonds with the lysine amide in the dimer or tetramer models are conserved across RPRD proteins in mammals (Ni et al., 2014). This flexible mechanism of recognition could allow for the RPRD complex to be first recruited to the distal region of the CTD via K7ac alone. Furthermore, this could allow for the interaction between CID domains to accommodate for serine-to-asparagine substitutions at position 7, also in the distal region of the CTD. The complex can then migrate along the consensus repeats within the CTD using S2p and possibly also S7p to remove S5p along the full length of the vertebrate CTD (**Fig. 6A**) (Egloff et al., 2012; Schuller et al., 2016).

Serine-7 phosphorylation is the third-best studied PTM of the CTD. In mammals, S7p is enriched with S5p near promoters, but is uniquely stable in gene bodies (Descostes et al., 2014). Similar to K7ac, S7-phosphorylated residues are considered docking stations for RPAP2 regulating S5-dephosphorylation and expression of snRNA genes (Egloff et al., 2012). Interestingly, S7p also enhances CTD:CID interaction consistent with the enhancement of electrostatic stability we describe here for K7ac (Egloff et al., 2012; Ni et al., 2014). Recent studies have succeeded in examining individual repeat PTMs *in vivo* and showed that CTD heptads are generally phosphorylated at one position per repeat (Schuller et al., 2016; Suh et al., 2016). This supports a model of dynamic movement along the 52 mammalian CTD repeats proposed here where the RPRD complex may start at distal non-consensus regions and work its way up to more proximal consensus regions to reach all S5p marks and allow maximal placement of S2p. An interesting, yet unexplored connection may also exist between other minor phosphorylation sites (Y1 and T4) and K7 acetylation. Y1p in yeast is known to dissociate *Rtt103* during the initiation and elongation phases of transcription to prevent premature termination of transcription, and could interfere with K7ac recognition (Mayer et al., 2012). In contrast, T4p was shown to enhance the interaction with the *Rtt103* CID domain, and could be conserved in mammals (Jasnovidova et al., 2017b). As *Rtt103* and RPRD proteins are closely

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

related, Y1p and T4p may also regulate the recruitment of RPRD proteins in addition to K7ac, S2p and S7p.

While RPAP2 is a known S5 phosphatase, the nature of the RPRD-associated deacetylase remains unknown. Previous studies have identified HDAC2 as a RPRD1A-associated deacetylase, thus being a possible candidate for a K7 deacetylase (Liu et al., 2015). This is consistent with our previous observations that class I/II KDACs are involved in deacetylation of the hypophosphorylated form of Pol II during or after transcription initiation (Schroder et al., 2013). Notably, HDAC2, unlike HDAC1 and 3, is found at promoter-proximal regions in addition to gene bodies, underscoring its potential as the RPRD-associated K7ac deacetylase (Wang et al., 2009). Interestingly, RPRD1B has previously been shown to interact with p300/CBP to regulate gene expression in cancer (Zhang et al., 2018). This points to a fine-tuned balance between RPRD proteins in the recruitment of K7 acetyltransferases and deacetylases. Future studies will investigate whether differences exist between RPRD1A and RPRD1B proteins with respect to controlling K7 acetylation.

Gene expression changes as a consequence of RPRD1B knockdown were moderate, but the cellular pathways found to be altered in response to RPRD1B knockdown showed a relevant list of genes. These were mainly involved in development of multicellular organisms and were strikingly similar to differentially regulated pathways found in wildtype and 8KR- Pol II expressing cells (Simonti et al., 2015). We have previously shown that K7ac evolution in higher eukaryotes presented a unique mode by which transcription elongation is regulated in mammals. We propose that this regulation of K7ac is linked to the now reported recruitment of RPRD proteins and the corresponding S5 dephosphorylation, a step tightly controlled in its dynamics in yeast. The question of why the need arose to control S5p with K7ac in multicellular organisms at a defined distance from the TSS remains unanswered but will be further examined. At this point, our data underscore a key role of controlled CTD PTM regulation at the transition from initiation to elongation important for the expression of developmentally relevant genes; they further demonstrate that this control depends on precise interactions with the RPRD complex, which performs reader and effector functions at a well-defined time during the transcription cycle.

MATERIALS AND METHODS

Antibodies and Reagents

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

See **Table S2** (Antibodies and Usage); Dynabeads Protein G (ThermoFisher, 10003D), Dynabeads Protein A (ThermoFisher, 10001D), Bovine Calf Serum (Gemini, 100-506), 293T and NIH3T3 cells are from ATCC. Panobinostat (CAS 404950-80-7) and α-amanitin (CAS 23109-05-9) were purchased from Santa Cruz Biotechnology. All other chemicals and reagents were purchased from Sigma.

Cell Fractionation and Immunoprecipitation

Cell fractionation was performed using the Dignam & Roeder method with minor modifications. 293T or NIH3T3 cells were pelleted and washed in cold DPBS. Pellets were resuspended in 5 volumes of DR Buffer A (10mM HEPES-KOH pH 7.9, 10mM KCl, 1.5 mM $MgCl_2$, 0.5mM DTT, 1x HALT, 30nM Panobinostat and 5μM Nicotinamide). Cells were Dounce homogenized with 10 strokes using a tight pestle (Wheaton) and cytoplasmic lysates were set aside or decanted. Nuclear pellets were resuspended in DR Buffer C (20mM HEPES, 0.42M NaCl, 1.5mM MgCl2, 0.2mM EDTA, 25% glycerol, 0.5mM DTT, 1x HALT, 30nM Panobinostat and 5μM Nicotinamide), sonicated using the Sonic Dismembrator 500 (ThermoFisher Scientific). 500μg of nucleoplasm was precleared with normal IgG (Santa Cruz) conjugated to the appropriate beads and immunoprecipitation was performed using anti HA- agarose beads (Sigma, A2095) or antibodies bound to Dynabeads. Immunoprecipitates were eluted either by boiling in 2x Laemmli buffer (agarose) or incubating in Elution Buffer (50mM $NaHCO_3$, 1% SDS) and adding 2x Laemmli buffer (Dynabeads).

Lentiviral transduction of RPRD1B shRNAs

VSV-G pseudotyped lentiviruses were produced to contain a Puromycin resistance gene and a shRNA against RPRD1B (NM_027434.2-1003s21c1) or a scrambled control. Cells were transduced with 0.5mL unconcentrated virus and selected using 2μg/mL Puromycin for 1 week prior to experimentation.

Chromatin Immunoprecipitation in NIH3T3 cells

NIH3T3 cells were grown under normal conditions (10% BCS, 1x Penicillin and Streptomycin, 2mM L-Glutamine). We treated $6x10^7$ cells with a lysine deacetylase inhibitor cocktail (30nM panobinostat, 5uM Nicotinamide) or a vehicle control (DMSO, water) for 2h. Cells were fixed with 1% formaldehyde for 15 minutes, thoroughly washed with DPBS, and resuspended in ChIP lysis buffer #1 (10mM Tris pH 7.4, 10mM NaCl, 0.5% NP-40, 1x HALT, 30nM Panobinostat and 5μM Nicotinamide). After sitting on ice for 10 minutes, cells were briefly vortexed and nuclei

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were pelleted. Nuclei were treated with MNase (NEB, M0247S) for 25 minutes at RT, pelleted and resuspended on ice in ChIP lysis buffer #2 (50mM Tris HCl pH 8.0, 10mM EDTA, 0.5% SDS, 1x HALT, 30nM Panobinostat and 5μM Nicotinamide). Chromatin was further sheared by sonication using the Sonic Dismembrator 500 (ThermoFisher Scientific) and preserved at -80ºC until immunoprecipitation. 20-40ug chromatin was used for each IP with the antibody concentrations listed in **Table S2**. IPs were diluted into a final volume of 800uL with ChIP Dilution buffer (167mM NaCl, 16.7 mM Tris HCl pH 8.0, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS) and left at 4ºC overnight. IPs were washed then eluted in ChIP elution buffer (50mM $NaHCO_3$, 1% SDS) and decrosslinked at 65ºC for 16hrs. Samples were treated with RNAse A (Thermofisher, EN0531) for 20 minutes and DNA were purified using the QIAquick PCR purification kit (Qiagen, 28106). Primer sequences are available upon request. For samples that were deep-sequenced, 2ng of immunoprecipitated DNA from each reaction was used to create libraries using the Ovation Ultra-Low Library prep kit (Nugen, 0344-32) following manufacturer recommendations and libraries were deep sequenced on the HIseq 4000 or NextSeq 500 using single-end 50bp or single-end 75bp sequencing, respectively.

RNA sequencing

RNA was prepared from $1x10^6$ NIH3T3 cells using the QIAgen RNeasy Plus Kit. Libraries were prepared with the Ovation Ultralow System V2 kit pn: 7102-32 / 0344-32 and libraries were deep sequenced on NextSeq 500 using paired-end 75pb sequencing. RNA seq analysis was done using the Illumina RNAexpress application v 1.1.0.

ChIP-seq data analysis

Barcodes were removed and sequences were trimmed using Skewer (Jiang et al., 2014). For each ChIP 50-60 million reads were aligned to the *Mus musculus* mm10 genome assembly using Bowtie with the –a -l 55 -n 2 -m 1 parameter (Langmead et al., 2009). Peaks were called and sequence pileups normalized to reads per million using MACS2 -B -SPMR -g mm -no-model -slocal 1000 (Zhang et al., 2008). TSS profiling was done using -plotProfile on matrices generated with 10bp bins using the computeMatrix function found in the Deeptools 2.2.3 build (Ramirez et al., 2016). Normalized profiles are calculated as fold change in signal relative to the observed background signal which we define as the signal from the 10bp bin at -3000 relative to TSS. Reproducibility of data was assessed by principal component analysis (Ramirez et al., 2016).

Stable Isotope Labeling of Amino Acids in Culture (SILAC) of WT and 8KR Polymerases

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SILAC (Stable isotope-labeled amino acid) labeling was performed according to the manual of SILAC Protein Quantitation Kit (LysC) –DMEM (Thermo Scientific cat. no. A33969). In brief, 293T cells stably expressing Pol-II-WT-HA and Pol-II-8KR-HA were grown in the light medium (L-Lysine-2HCl) or heavy medium ($^{13}C_6$L-Lysine-2HCl), respectively. After growing 7 doubling time in the respective medium, incorporation efficiency of heavy L-lysine in 293T-Pol-II-8KR-HA cells was determined and the efficiency was more than 99%. To immunoprecipitate the HA proteins, 5 mg of total cell lysate from Pol-II-WT-HA and 5mg of total lysate from Pol-II-8KR-HA cells in p300 lysis buffer were mixed together (total 500 uL), and 100 uL of HA agarose (Roche) were added. After overnight immunoprecipitation at 4C, the HA-agarose was washed 4 times with 1 mL of cold p300 lysis buffer to remove non-specific binding proteins. The bound proteins were eluted twice by 100 ul of 0.1 M Glycine, pH 2.5 after 30 min incubation. Each elution was stored in separate tube. 10 uL of 1 M Tris-HCl pH 8.0 was added into each elution to neutralize the pH. The quality of the elution was monitored by Protein Silver Staining (Pierce). Two elusions were combined and 50 uL out of the 200 uL combined elusions were sent to Mass Spectrometry. Two independent biological repeats were performed.

Mass Spectrometry Analysis

Sample were analyzed on a Thermo Scientific LTQ Orbitrap Elite mass spectrometry system equipped with an Easy-nLC 1000 HPLC and autosampler. Samples were injected onto a pre-column (2cm x 100 um I.D. packed with 5 um C18 particles) in 100% buffer A (0.1% formic acid in water) and separated by a 120 minute reverse phase gradient from 5% to 30% buffer B (0.1% formic acid in 100% ACN) at a flow rate of 400 nl/min. The mass spectrometer continuously collected spectra in a data-dependent manner, acquiring a full scan in the Orbitrap (at 120,000 resolution with an automatic gain control target of 1,000,000 and a maximum injection time of 100 ms) followed by collision-induced dissociation spectra for the 20 most abundant ions in the ion trap (with an automatic gain control target of 10,000, a maximum injection time of 10 ms, a normalized collision energy of 35.0, activation Q of 0.250, isolation width of 2.0 m/z, and an activation time of 10.0). Singly and unassigned charge states were rejected for data-dependent selection. Dynamic exclusion was enabled to data-dependent selection of ions with a repeat count of 1, a repeat duration of 20.0 s, an exclusion duration of 20.0 s, an exclusion list size of 500, and exclusion mass width of + or - 10.00 ppm.

Raw mass spectrometry data were analyzed using the MaxQuant software package (version 1.2.5.8) (Cox and Mann, 2008). Data were matched to the SwissProt human proteins (downloaded from UniProt on 2/15/13, 20,259 protein sequence entries). MaxQuant was

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

configured to generate and search against a reverse sequence database for false discovery rate calculations. Variable modifications were allowed for methionine oxidation and protein N-terminus acetylation. A fixed modification was indicated for cysteine carbamidomethylation. Full trypsin specificity was required. The first search was performed with a mass accuracy of +/- 20 parts per million and the main search was performed with a mass accuracy of +/- 6 parts per million. A maximum of 5 modifications were allowed per peptide. A maximum of 2 missed cleavages were allowed. The maximum charge allowed was 7+. Individual peptide mass tolerances were allowed. For MS/MS matching, a mass tolerance of 0.5 Da was allowed and the top 6 peaks per 100 Da were analyzed. MS/MS matching was allowed for higher charge states, water and ammonia loss events. The data were filtered to obtain a peptide, protein, and site-level false discovery rate of 0.01. The minimum peptide length was 7 amino acids. Results were matched between runs with a time window of 2 minutes for technical duplicates.

Isothermal Titration Calorimetry

RNA Pol II CTD peptides were purchased from Peptide 2.0 (Chantilly, VA). ITC experiments were performed as previously described (Ni et al., 2014)

Molecular modeling

Crystallographic structures were used for the dimer (pdb:4Q94) and tetramer (pdb:4Q96) models. Electrostatic potential surfaces were calculated using an adaptive Poisson-Boltzmann solver (APBS) from the PDB2PQR server using the Amber force field and PROPKA to assign protonation states (Dolinsky et al., 2004). Amber's LEaP program was used with the Amber ff14SB force field and the following force field modifications: phosaa10 (phospates), ffptm (phosphorylated serines) and ALY.frcmod (acetyllysines). The TIP3P water model was used to solvate the system in a cubic periodic box, such that the closest distance between any atom in the system and the periodic boundary is 10 Å. Net positive charge in the box was neutralized by adding counterions (Cl-) until neutrality. Energy minimization was performed in two steps: using harmonic restraints on the protein (10.0 kcal mol^{-1} $Å^{-2}$) and an unrestrained minimization. For each minimization we ran 1000 steps of steepest descent and 1000 steps of conjugate gradient minimization at a constant volume with a non-bonded cutoff of 9 Å. The equilibration was done in three steps. First, the system was heated from 0 to 300 K with a restrained equilibration (10.0 kcal mol^{-1} $Å^{-2}$) for 20 ps at constant volume with a non-bonded cutoff of 9 Å, using the SHAKE algorithm to constrain bonds involving hydrogens, and the Andersen thermostat. The second round of equilibration was performed lowering the harmonic restraints (1.0 kcal mol^{-1} $Å^{-2}$) on the

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

system for 20 ps (other parameters identical). The third round was performed for 1 ns at constant pressure of 1.0 bar with non-bonded cutoff of 9 Å at 300 K with the Andersen thermostat. Simulations were performed without restraints using new velocities with random seeds at constant pressure of 1 bar with non-bonded cutoff distance of 9 Å. 20ns simulations were run with 2 fs timestep per construct. Coordinates and energy were saved every picosecond (500 steps) (Case et al., 2005). Molecular graphics and analyses were performed with the UCSF Chimera package (Pettersen et al., 2004).

ACKNOWLEDGEMENTS

We thank members of the Ott laboratory, J.J. Miranda, PhD and Bassem Al Sady, PhD, for helpful discussions, reagents and expertise. Natasha Carli, PhD and Jim McGuire from the Gladstone Genomics Core for library preparation and QCs for next-generation sequencing and for funding from the James B. Pendleton Charitable Trust. We are grateful for funding support from the NIH R01AI083139 (M.O.), P50-GM082250 (N.J.K.), and P01-CA177322 (J.R.J) UCSF Discovery Fellowship and American Society for Microbiology, Robert D. Watkins Graduate Research Fellowship (I.A.), and funding from the NSERC RGPIN-2016-06300 (J.M.). We thank John Carroll for graphics, Kathryn Claiborn for editing and Lauren Weiser for administrative support. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311)

AUTHOR CONTRIBUTIONS

I.A. designed the study, validated SILAC hits, conducted HDACi experiments, western blotting and quantification, performed ChIP-qPCR and ChIP-seq experiments and analyses, performed knockdown experiments, and statistical analyses. P.C.L. and J.J. performed SILAC experiments and mass-spectrometric data analysis. Z.N., H.Z., and J.M. performed ITC experiments. D.G.R. performed molecular modeling experiments. R.J.C. supported ChIP-seq analysis. X.G., J.G., M.J., and N.K. supervised experiments. M.O. supervised the study design and data collection. I.A. and M.O. wrote the manuscript.

REFERENCES

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev *24*, 2303-2316.
Bartolomei, M.S., and Corden, J.L. (1987). Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. Mol Cell Biol *7*, 586-594.
Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramee, L., Fortin, M.E., 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.
Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N.*, et al.* (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157-170.
Buratowski, S. (2003). The CTD code. Nat Struct Biol *10*, 679-680.
Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol Cell *36*, 541-546.
Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J Comput Chem *26*, 1668-1688.
Chapman, R.D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. Trends Genet *24*, 289-296.
Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.
Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I.*, et al.* (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife *3*, e02105.
Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J., 3rd, and Singer, D.S. (2012). BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. Proc Natl Acad Sci USA *109*, 6927-6932.
Dias, J.D., Rito, T., Torlai Triglia, E., Kukalev, A., Ferrai, C., Chotalia, M., Brookes, E., Kimura, H., and Pombo, A. (2015). Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. eLife *4*.
Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res *32*, W665-667.
Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D.*, et al.* (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep *20*, 1173-1186.
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.
Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.
Heidemann, M., Hintermair, C., Voss, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta *1829*, 55-62.
Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J *31*, 2784-2797.
Jasnovidova, O., Klumpler, T., Kubicek, K., Kalynych, S., Plevka, P., and Stefl, R. (2017a). Structure and dynamics of the RNAPII CTDsome with Rtt103. Proc Natl Acad Sci USA *114*, 11133-11138.
Jasnovidova, O., Krejcikova, M., Kubicek, K., and Stefl, R. (2017b). Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. EMBO Rep *18*, 906-913.
Jasnovidova, O., and Stefl, R. (2013). The CTD code of RNA polymerase II: a structural view. Wiley Interdiscip Rev RNA *4*, 1-16.
Jiang, H., Lei, R., Ding, S.W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics *15*, 182.
Jin, K., Chen, H., Zuo, Q., Huang, C., Zhao, R., Yu, X., Wang, Y., Zhang, Y., Chang, Z., and Li, B. (2018). CREPT and p15RS regulate cell proliferation and cycling in chicken DF-1 cells through the Wnt/beta-catenin pathway. J Cell Biochem *119*, 1083-1092.
Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol *17*, 1279-1286.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Liu, C., Zhang, Y., Li, J., Wang, Y., Ren, F., Zhou, Y., Wu, Y., Feng, Y., Zhou, Y., Su, F.*, et al.* (2015). p15RS/RPRD1A (p15INK4b-related sequence/regulation of nuclear pre-mRNA domain-containing protein 1A) interacts with HDAC2 in inhibition of the Wnt/beta-catenin signaling pathway. J Biol Chem *290*, 9701-9713.
Liu, J., Liu, H., Zhang, X., Gao, P., Wang, J., and Hu, Z. (2002). Identification and characterization of P15RS, a novel P15(INK4b) related gene on G1/S progression. Biochem Biophys Res Commun *299*, 880-885.
Lu, D., Wu, Y., Wang, Y., Ren, F., Wang, D., Su, F., Zhang, Y., Yang, X., Jin, G., Hao, X.*, et al.* (2012). CREPT accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. Cancer Cell *21*, 92-104.
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.
Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.
Mei, K., Jin, Z., Ren, F., Wang, Y., Chang, Z., and Wang, X. (2014). Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. Sci China Life Sci *57*, 97-106.
Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223-226.
Morales, J.C., Richard, P., Rommel, A., Fattah, F.J., Motea, E.A., Patidar, P.L., Xiao, L., Leskov, K., Wu, S.Y., Hittelman, W.N.*, et al.* (2014). Kub5-Hera, the human Rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. Nucleic Acids Res *42*, 4996-5006.
Muller, S., Filippakopoulos, P., and Knapp, S. (2011). Bromodomains as therapeutic targets. Expert Rev Mol Med *13*, e29.
Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimi Biophys Acta *1809*, 34-45.

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ni, Z., Olsen, J.B., Guo, X., Zhong, G., Ruan, E.D., Marcon, E., Young, P., Guo, H., Li, J., Moffat, J.*, et al.* (2011). Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. Transcription *2*, 237-242.
Ni, Z., Xu, C., Guo, X., Hunter, G.O., Kuznetsova, O.V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W.W.*, et al.* (2014). RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. Nat Struct Mol Biol *21*, 686-695.
Patidar, P.L., Motea, E.A., Fattah, F.J., Zhou, Y., Morales, J.C., Xie, Y., Garner, H.R., and Boothman, D.A. (2016). The Kub5-Hera/RPRD1B interactome: a novel role in preserving genetic stability by regulating DNA mismatch repair. Nucleic Acids Res *44*, 1718-1731.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.
Pineda, G., Shen, Z., de Albuquerque, C.P., Reynoso, E., Chen, J., Tu, C.C., Tang, W., Briggs, S., Zhou, H., and Wang, J.Y. (2015). Proteomics studies of the interactome of RNA polymerase II C-terminal repeated domain. BMC Res Notes *8*, 616.
Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
Schroder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D.A., Kaehlcke, K., Cho, S., Pollard, K.S., Capra, J.A.*, et al.* (2013). Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell *52*, 314-324.
Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell *61*, 305-314.
Simonti, C.N., Pollard, K.S., Schroder, S., He, D., Bruneau, B.G., Ott, M., and Capra, J.A. (2015). Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. BMC Evol Biol *15*, 35.
Sims, R.J., 3rd, Rojas, L.A., Beck, D.B., Bonasio, R., Schuller, R., Drury, W.J., 3rd, Eick, D., and Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science *332*, 99-103.
Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell *61*, 297-304.
Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidaemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S.*, et al.* (2010). Chemical-genomic dissection of the CTD code. Nat Struct Mol Biol *17*, 1154-1161.
Voss, K., Forne, I., Descostes, N., Hintermair, C., Schuller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M.*, et al.* (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription *6*, 91-101.
Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019-1031.
Weinert, B.T., Narita, T., Satpathy, S., Srinivasan, B., Hansen, B.K., Scholz, C., Hamilton, W.B., Zucconi, B.E., Wang, W.W., Liu, W.R.*, et al.* (2018). Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. Cell *174*, 231-244 e212.
Wu, Y., Zhang, Y., Zhang, H., Yang, X., Wang, Y., Ren, F., Liu, H., Zhai, Y., Jia, B., Yu, J.*, et al.* (2010). p15RS attenuates Wnt/{beta}-catenin signaling by disrupting {beta}-catenin.TCF4 Interaction. J Biol Chem *285*, 34621-34631.

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Y., Wang, S., Kang, W., Liu, C., Dong, Y., Ren, F., Wang, Y., Zhang, J., Wang, G., To, K.F.*, et al.* (2018). CREPT facilitates colorectal cancer growth through inducing Wnt/beta-catenin pathway by enhancing p300-mediated beta-catenin acetylation. Oncogene *37*, 3485-3500.

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FIGURE LEGENDS

Fig. 1: RPRD proteins interact with RPB1 in an acetylation-dependent manner. A. SILAC based mass spectrometry screen was used to identify factors that bind preferentially to WT Pol II compared to the 8KR mutant. **B**. Components of the RPRD1B/RPRD1A complex that were identified as WT interactors. **C**. Reciprocal IPs using antibodies against RPRD1B or hemagglutinin (HA) from WT and 8KR cells. **D**. Densitometry using ImageJ of n=4 independent RPRD1B western blots after HA- immunoprecipitation. **E**. NIH3T3 cells were treated with KDAC inhibitors (30nM Panobinostat and 5uM Nicotinamide) for 2 hours. IP from 500ug of nucleoplasm pre-cleared with IgG followed by western blotting with the indicated antibodies. Pol II (Total) blots were performed with an N-terminal H224 antibody **F**. Densitometry of RPRD1A and RPRD1B western blots produced from three independent Pol II (Total-H224) IP elutions. Error bars are SEM, * $p < 0.05$ for a one-tailed T test. **G.** ChIP-qPCR of *Leo1* at +33 nt downstream of TSS using the indicated antibodies from 4 independent chromatin preps, total Pol II ChIP experiments were done with the 8WG16 antibody. Values are represented as percent of input with IgG subtracted, error bars are SEM.

Fig. 2: RPRD CID domains recognize acetylated and phosphorylated CTD peptides. Isothermal titration calorimetry (ITC) experiments measuring in-vitro binding affinity between RPRD1A and RPRD1B CID domains and modified CTD peptides. **A.** Table of CTD-39 peptides with indicated post-translational modifications and predicted RPRD CID binding moieties. **B.** Representative ITC plots of CTD-CID interactions. **C** K_d values measured by ITC for RPRD1B CID and modified CTD-39 binding. **D**. K_d values measured by ITC for RPRD1A CID and modified CTD-39 binding. * $p < 0.05$, ** $p < 0.01$ *** $p < 0.005$ using a one tailed T test. **E.** Dimer model from RPRD1B crystal structure (pdb:4Q94) containing two recognition modules and two peptides fragments of the CTD. **F.** Electrostatic potential surface for RPRD1B CID. **G.** First CTD peptide fragment model from crystal structure superimposed with the electrostatic potential surface around the corresponding binding site. **H.** Second CTD peptide fragment model from crystal structure superimposed with the electrostatic potential surface around the corresponding binding site. **I.** Recognition elements around first K7ac and S2p in the CTD peptide fragment. **J.** Recognition elements around second K7ac in the CTD peptide fragment.

Fig. 3: An inverse relationship between K7ac and phosphorylation is induced upon KDAC inhibition. A-C. TSS profiles generated from ChIP-seq data of expressed genes in NIH3T3 cells. Normalized profiles are measured in reads per million and expressed as fold change relative to background signal. **D**. Occupancy profiles of RNA Pol II PTMs measured as reads

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

per million on selected RPRD1B occupied genes. S5p is measured using the RNA Pol II 4H8 antibody. **E.** Occupancy profiles of RNA Pol II PTMs on control genes. Green arrows indicate the direction of transcription relative to the TSS of the depicted gene. Black arrows indicate the site of affected PTMs in response to HDACi.

Fig. 4: RPRD1B knockdown dysregulates genes relating to development and multicellularity and disturbs Pol II PTM homeostasis. NIH3T3 cells were treated with shRNAs against RPRD1B (sh1B) or a scrambled sequence (shScr) and selected with puromycin for 1 week. **A.** Western blotting from whole cell lysates to monitor RPRD1B protein expression upon knockdown and the effect on Pol II K7ac and S5p. **B.** RNA-seq highlighting 271 significantly dysregulated genes in red. **C.** Gene ontology analysis of genes significantly dysregulated in response to RPRD1B knockdown. **D.** DESeq counts per million for selected upregulated genes associated with regulation of multicellular organismal process. **E** DESeq counts per million for selected downregulated genes in response to RPRD1B knockdown. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ using a one tailed T test.

Fig. 5: Increased S5p and K7ac levels in response to RPRD1B knockdown. TSS profiles of expressed genes in NIH3T3 cells. Normalized profiles are measured in reads per million normalized to background signal. **A-C** TSS profiles generated from ChIP-seq data from NIH3T3 cells expressing either scrambled or RPRD1B targeted shRNAs treated with Vehicle control. **D-F** TSS profiles generated from ChIP-seq data from NIH3T3 cells expressing either scrambled or RPRD1B targeted shRNAs treated with KDACi. S5p is measured using the RNA Pol II 4H8 antibody. Profiles are representative of two independent experiments.

Fig. 6: RPRD proteins are recruited to the RPB1 CTD via acetylation and phosphorylation to antagonize S5-phosphorylation. A. Model of RPRD complex reader and effector functions along modified residues within the CTD. 1B- RPRD1B 1A- RPRD1A, KDAC is a yet to be identified K7 deacetylase. B. KDAC inhibition induces K7 hyperacetylation and downregulation of S5-phosphorylation. C. Knockdown of RPRD1B perturbs the recruitment of the complex to result in both S5 hyperphosphorylation and K7 hyperacetylation. See text for details.

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Fig. S1: Dimer and tetramer models of RPRD1B:CTD complex share similar features. Comparison between dimeric and tetrameric models of RPRD1B:CTD complex. **A**. Dimer (left) and tetramer (right) models from crystal structures. **B**. Electrostatic potential surface for dimer (left) and tetramer (right) models. **C**. Dimer model CTD peptide fragment superimposed with the electrostatic potential surface around the corresponding binding site**. D.** Tetramer model CTD peptide fragment superimposed with the electrostatic potential surface around the corresponding binding site. **E**. Recognition elements around first K7ac and first S2p in the CTD peptide fragment in the tetramer model. **F**. Recognition elements around second K7ac and second S2p in the CTD peptide fragment in the tetramer model.

Supplemental Fig. S2: An inverse correlation between K7ac and phosphorylation on chromatin is induced upon KDAC inhibition. A-C. TSS profiles generated from ChIP-seq data of expressed genes in NIH3T3 cells measured in reads-per-million. S5p is measured using the RNA Pol II 4H8 antibody.

Supplemental Fig. S3: RPRD1B knockdown induces hyperacetylation and hyperphosphorylation on chromatin near TSSs. NIH3T3 cells were treated with shRNAs targeting a scrambled sequence (shSCR) or RPRD1B (shRPRD1B), then with KDACi or a Vehicle control for 2 hours followed by ChIP-seq. **A-C**. TSS profiles generated from ChIP-seq data shown as reads-per-million.

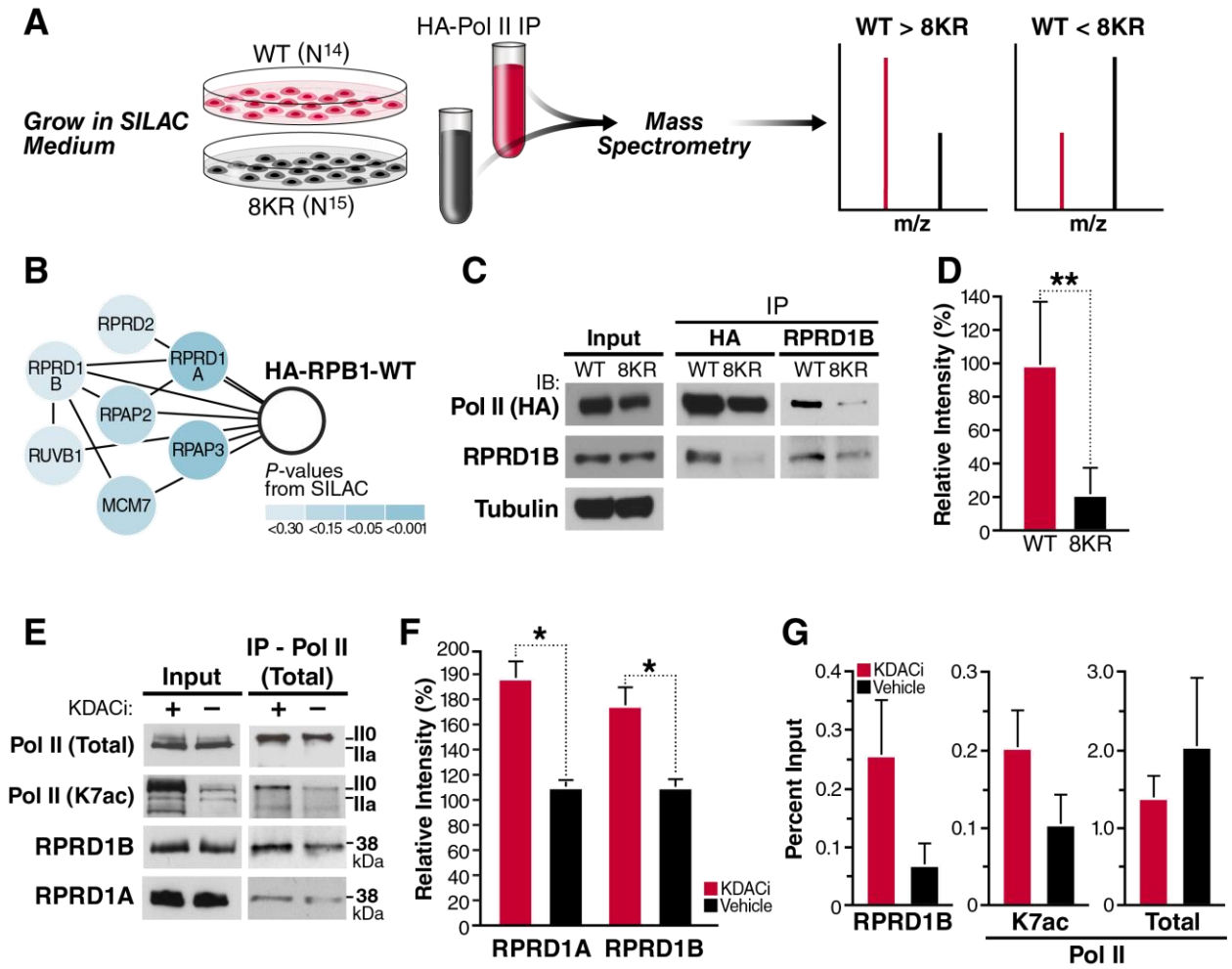
SUPPLEMENTAL TABLES

Table S1: Genes Dysregulated by RPRD1B Knockdown

Table S2: Antibodies and Usage

bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. 1



bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

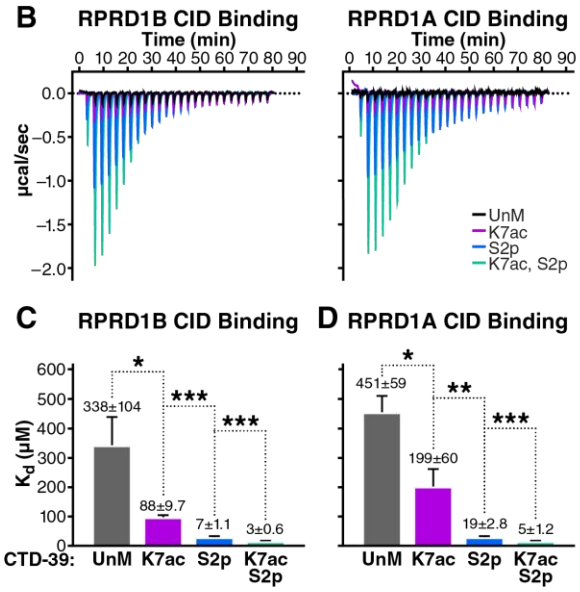
Fig. 2

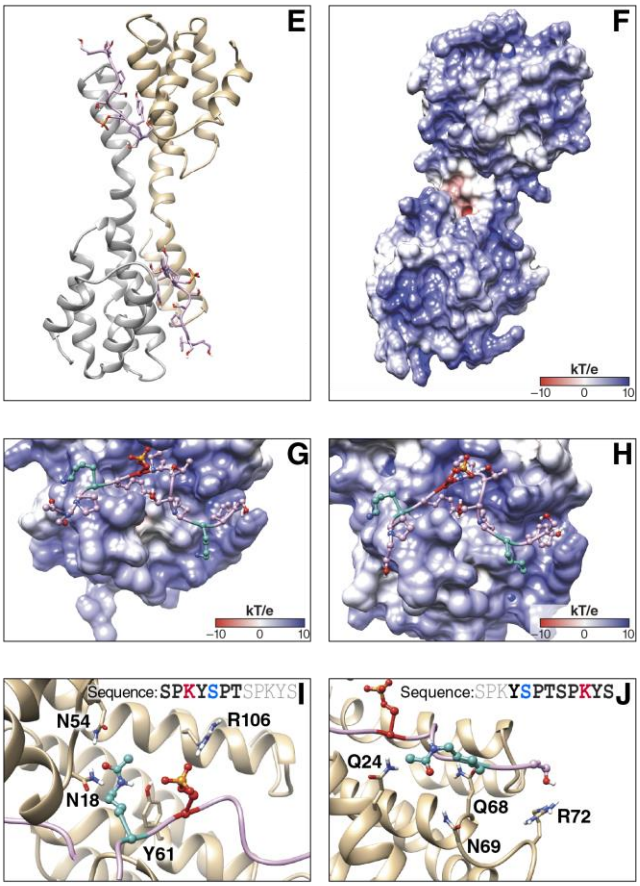
A

Heptad and Modifications	Sequence
CTD-39 UnM	E YTPTSPK YSPTSPK YSPTS
CTD-39 K7ac	E YTPTSPK YSPTSPK YSPTS
CTD-39 S2p	E YTPTSPK YSPTSPK YSPTS
CTD-39 K7ac-S2p	E YTPTSPK YSPTSPK YSPTS
CTD-39 S5p	E YTPTSPK YSPTSPK YSPTS
CTD-39 K7ac-S5p	E YTPTSPK YSPTSPK YSPTS

Amino-acids in red are acetylated; in blue are phosphorylated

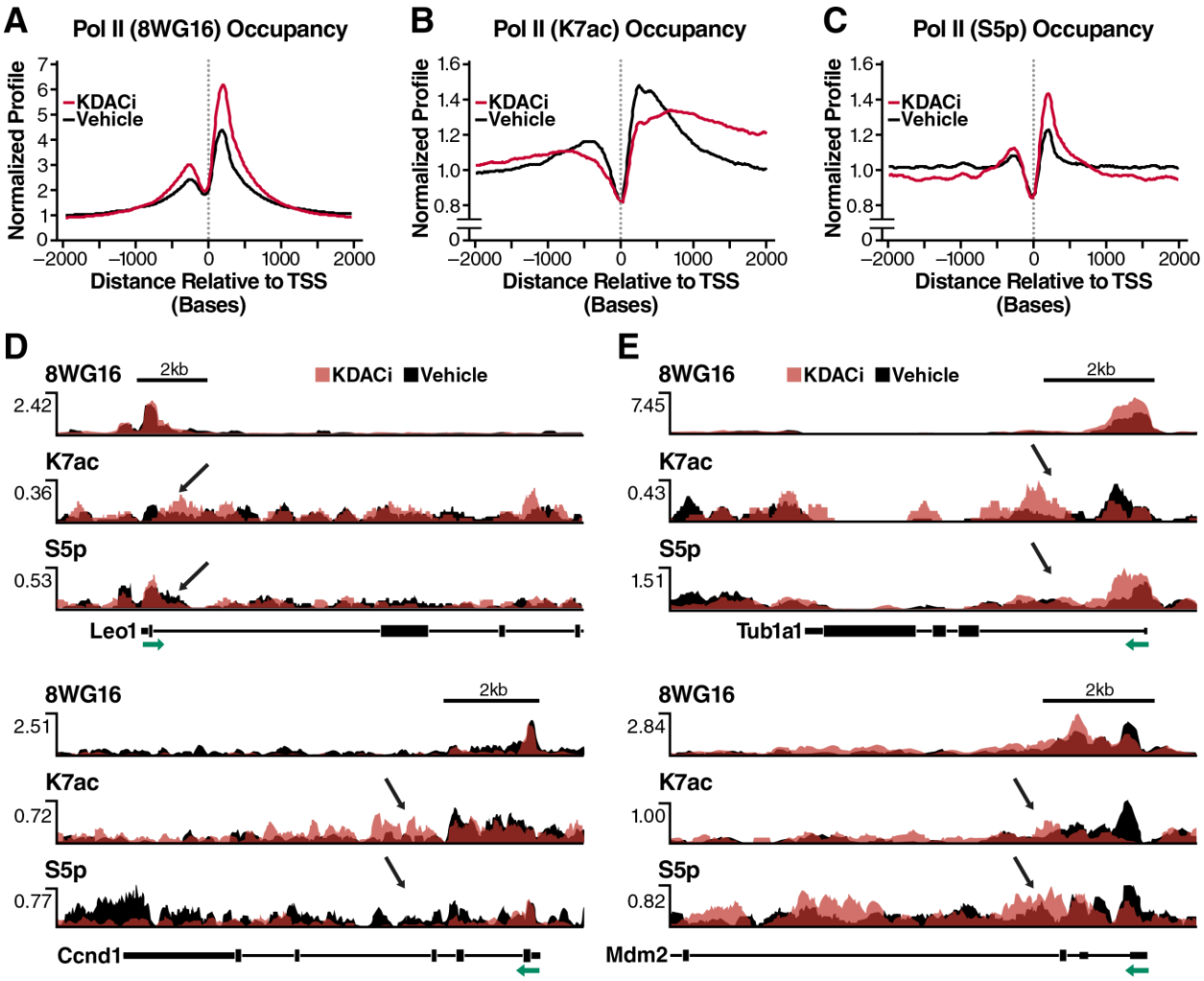
PREDICTED CID BINDING MOIETIES





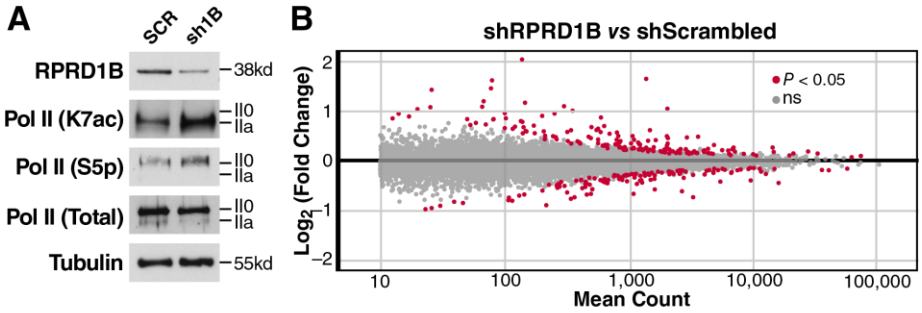
bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. 3



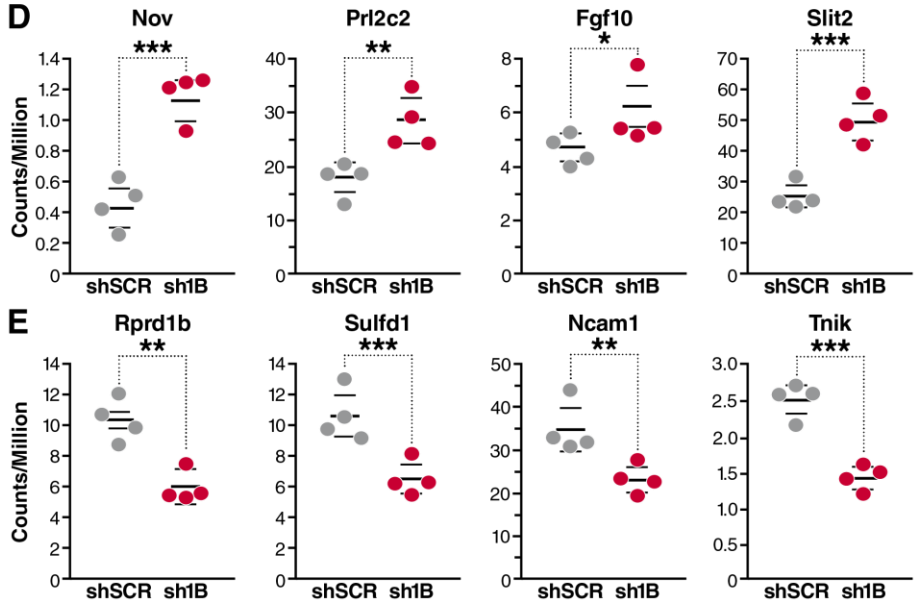
bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. 4



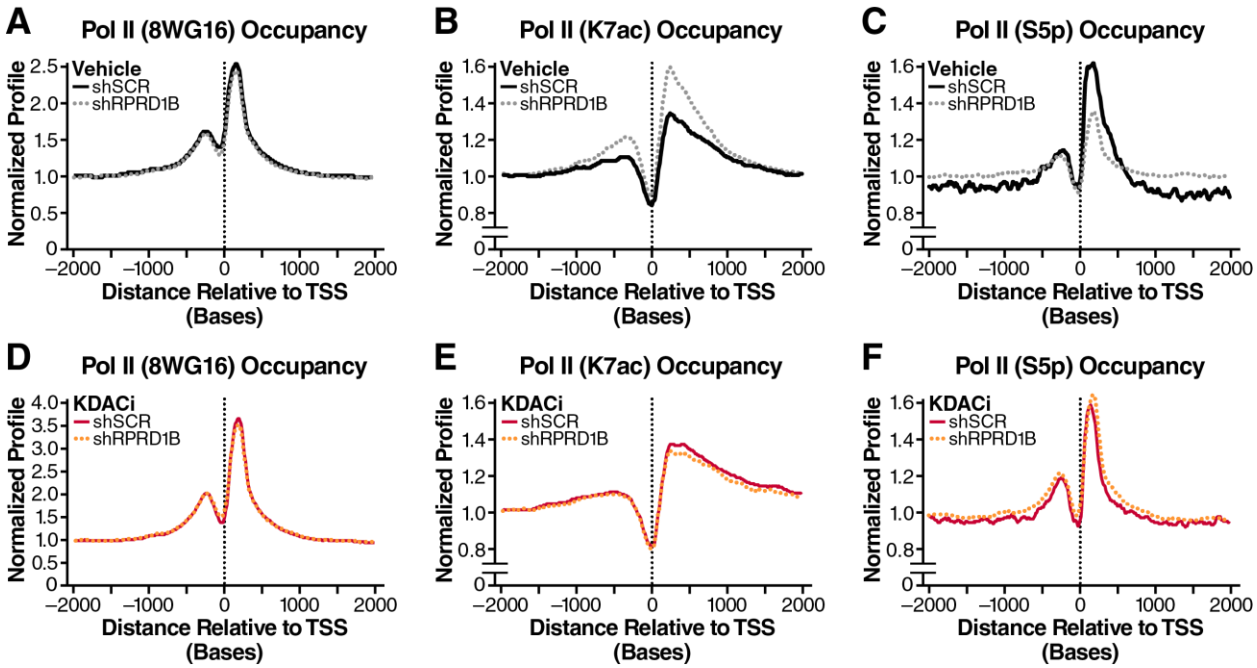
C

RPRD1B Dysregulated Genes (GO)	Number of Genes	Fold Enrichment	FDR q-value
Developmental Process	127	1.76	4.95E-9
Regulation of Multicellular Organismal Process	89	2.11	4.30E-9
Regulation of Developmental Process	81	2.07	1.46E-7
Regulation of Cellular Component Movement	42	2.87	1.56E-6
Regulation of Multicellular Organismal Development	64	2.13	4.40E-6
Cell Adhesion	34	3.07	7.59E-6
Anatomical Structure Development	87	1.79	8.59E-6



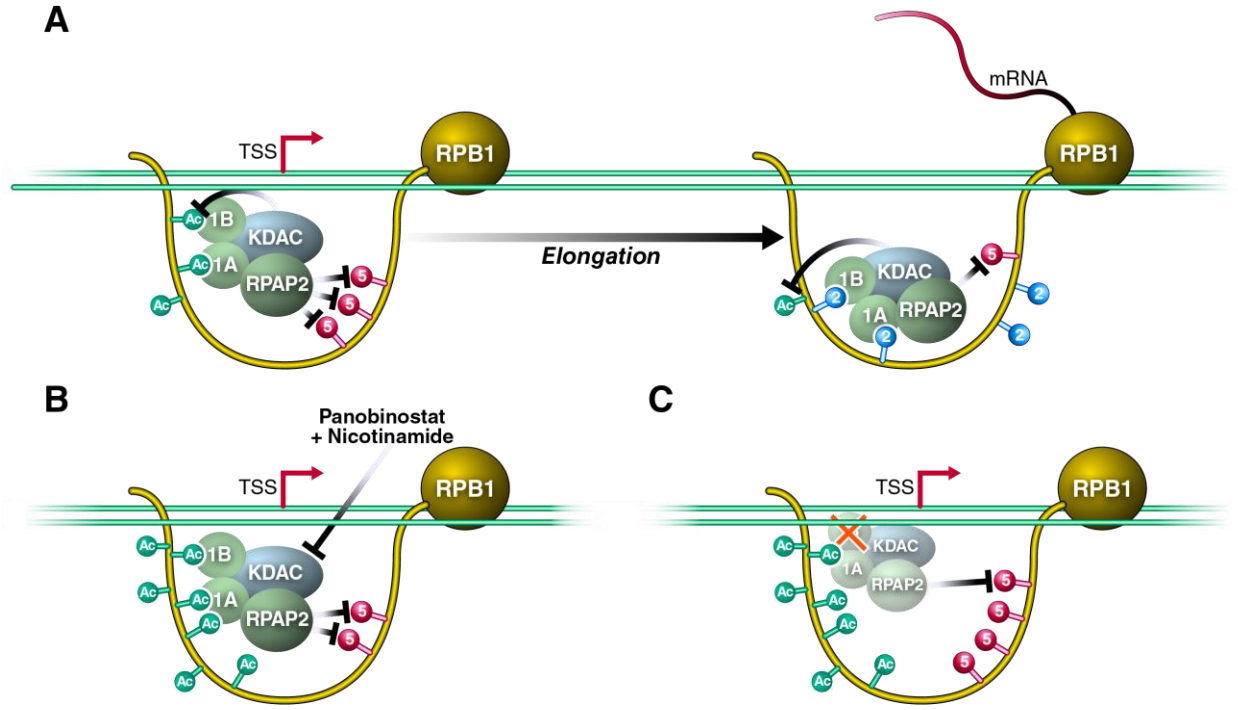
bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. 5



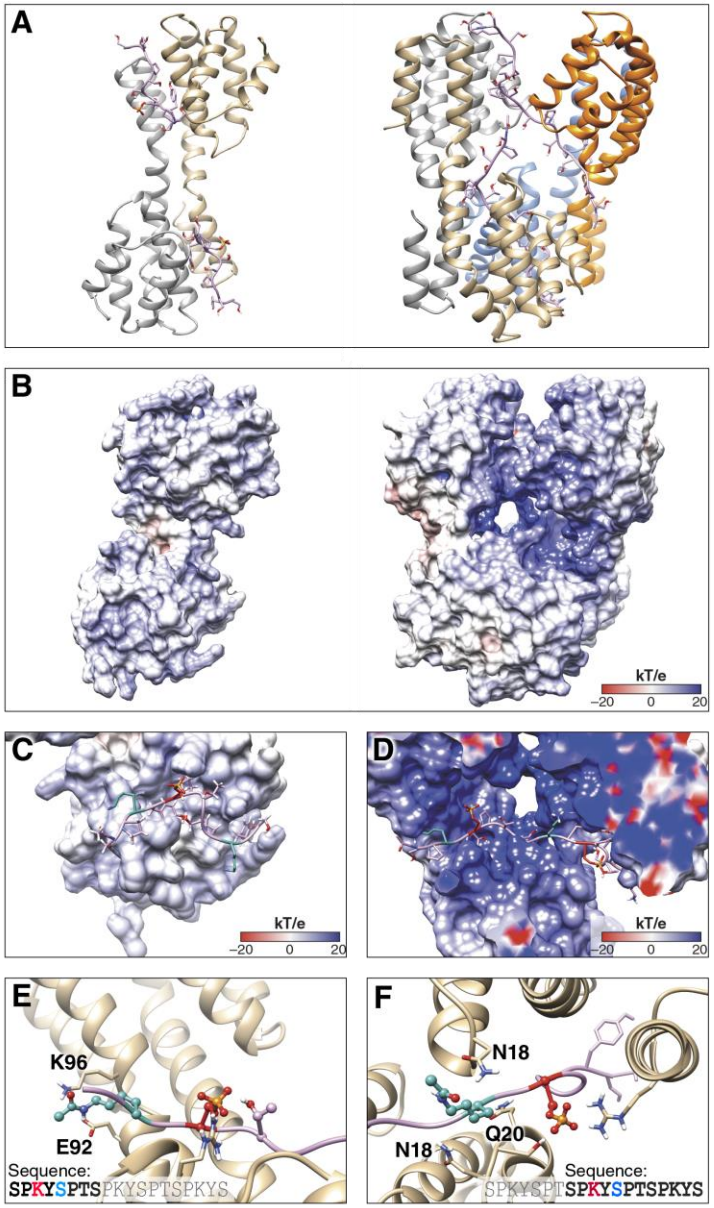
bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. 6



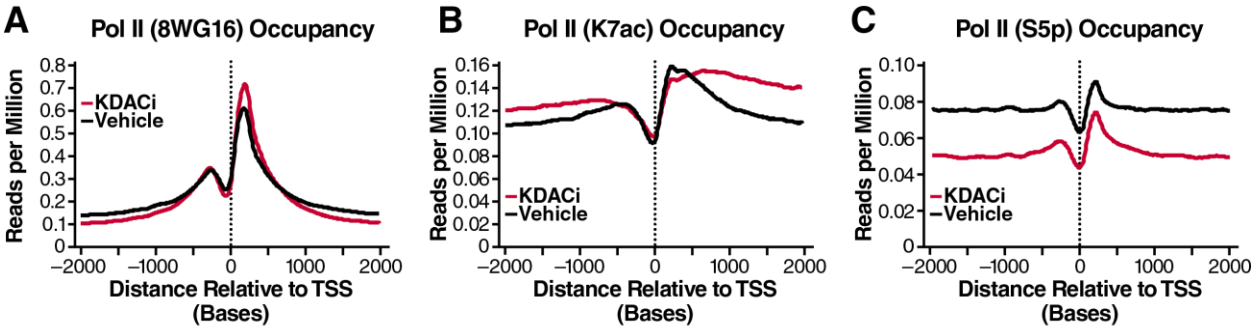
bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. S1



bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig. S2



bioRxiv preprint doi: https://doi.org/10.1101/442491; this version posted October 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

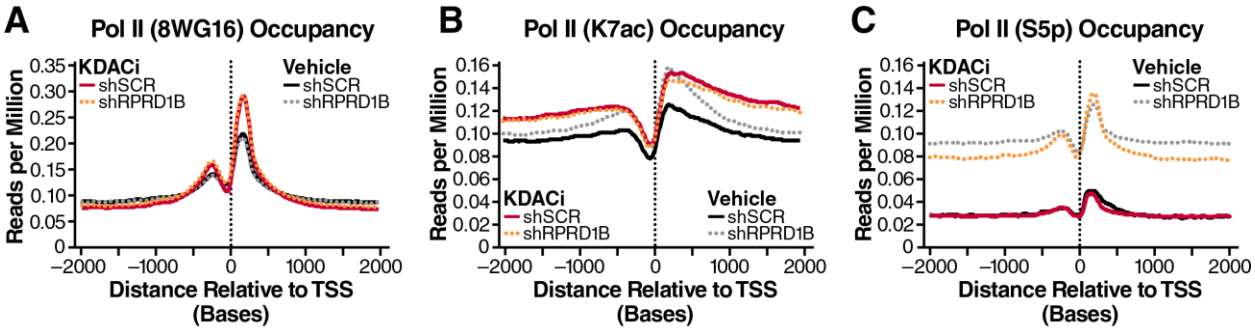



Fig. S3