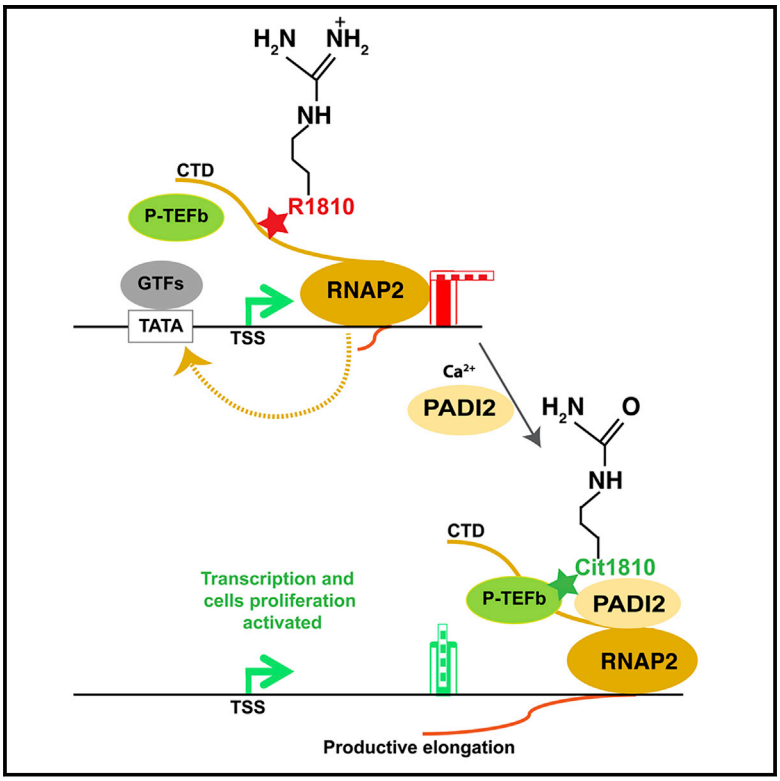
Article

Molecular Cell

Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription

Graphical Abstract



Authors

Priyanka Sharma, Antonios Lioutas, Narcis Fernandez-Fuentes, ..., Dirk Eick, Baldomero Oliva, Miguel Beato

Correspondence

miguel.beato@crg.eu

In Brief

Sharma et al. discovered that arginine1810 at the C-terminal domain of RNAP2 is deiminated to citrulline (cit1810) by the peptidyl arginine deiminase 2. The cit1810 is required to facilitate RNAP2 association with P-TEFb complex and consequently allows the efficient transcription of the highly transcribed genes relevant for cellular proliferation.

Highlights

- PADI2 citrullinates arginine1810 (cit1810) at CTD of RNA polymerase II (RNAP2)
- PADI2 and R1810 regulate transcription and proliferation of breast cancer cells
- Absence of cit1810 leads to RNAP2 accumulation at proximal promoter regions
- Cit1810 facilitates interaction of RNAP2 with P-TEFb complex

Sharma et al., 2019, Molecular Cell *73*, 1–13
January 3, 2019
https://doi.org/10.1016/j.molcel.2018.10.016

CellPress

Please cite this article in press as: Sharma et al., Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription, Molecular Cell (2018), https://doi.org/10.1016/j.molcel.2018.10.016

Molecular Cell
Article

Arginine Citrullination at the C-Terminal Domain Controls RNA Polymerase II Transcription

Priyanka Sharma,[1] Antonios Lioutas,[1] Narcis Fernandez-Fuentes,[2] Javier Quilez,[1] José Carbonell-Caballero,[1] Roni H.G. Wright,[1] Chiara Di Vona,[1] François Le Dily,[1] Roland Schüller,[3] Dirk Eick,[3] Baldomero Oliva,[4,5] and Miguel Beato[1,4,6,*]

[1]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona 08003, Spain
[2]IBERS, Institute of Biological, Environmental and Rural Science, Aberystwyth University, Aberystwyth SY23 3EB, UK
[3]Department of Molecular Epigenetics, Helmholtz Center Munich, Center of Integrated Protein Science, Munich, Germany
[4]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[5]Structural Bioinformatics Laboratory (GRIB-IMIM), Department of Experimental and Health Sciences, Barcelona 08003, Spain
[6]Lead Contact
*Correspondence: miguel.beato@crg.eu
https://doi.org/10.1016/j.molcel.2018.10.016

SUMMARY

The post-translational modification of key residues at the C-terminal domain of RNA polymerase II (RNAP2-CTD) coordinates transcription, splicing, and RNA processing by modulating its capacity to act as a landing platform for a variety of protein complexes. Here, we identify a new modification at the CTD, the deimination of arginine and its conversion to citrulline by peptidyl arginine deiminase 2 (PADI2), an enzyme that has been associated with several diseases, including cancer. We show that, among PADI family members, only PADI2 citrullinates R1810 (Cit1810) at repeat 31 of the CTD. Depletion of PADI2 or loss of R1810 results in accumulation of RNAP2 at transcription start sites, reduced gene expression, and inhibition of cell proliferation. Cit1810 is needed for interaction with the P-TEFb (positive transcription elongation factor b) kinase complex and for its recruitment to chromatin. In this way, CTD-Cit1810 favors RNAP2 pause release and efficient transcription in breast cancer cells.

INTRODUCTION

In mammals, the RNA polymerase II C-terminal domain (RNAP2-CTD) comprises 52 heptapeptide repeats, the first half of which (1–27) exhibit the consensus repeat sequence $Y_1S_2P_3T_4S_5P_6S_7$, whereas the second half (28–52) contains deviations from this consensus (Buratowski, 2009; Corden, 2013). Post-translational modification of the key residues at RNAP2-CTD dictates recruitment of protein complexes, which influence transcription elongation and the processing of the nascent transcripts (Jeronimo et al., 2016; Saldi et al., 2016; Harlen and Churchman, 2017). The CTD is evolutionary conserved, and dynamic phosphorylation of Y_1, S_2, T_4, S_5, and S_7 mediates selective recruitment of protein complexes that modulate various phases of transcription (Eick and Geyer, 2013; Zaborowska et al., 2016; Shah et al., 2018). Systematic studies using genetics and proteomics showed that phosphorylation of S_5 and S_2 is the most frequent modification and contributes to transcription efficiency (Schüller et al., 2016; Corden, 2016). However, modifications in non-consensus repeats have expanded the functional role of the CTD code (Voss et al., 2015; Dias et al., 2015), and recent work has focused on methylation of arginine 1810 (R1810) at repeat 31. Its asymmetrical dimethylation (me2a) by the methyltransferase CARM1 (or PRMT4) inhibits the expression of small nuclear RNAs (snRNAs) and nucleolar RNA (snoRNA) genes in human cells (Sims et al., 2011). This reaction is inhibited by phosphorylation of CTD serine residues, suggesting that it occurs before transcription initiation. In contrast, symmetric dimethylation (me2s) of R1810 by PRMT5 leads to recruitment of the survival of motor neuron protein (SMN) and to the interaction with senataxin, which enhances transcriptional termination (Zhao et al., 2016). The functional significance of dynamic post-translational deimination of arginine residues in pathophysiological conditions (Slade et al., 2014; Tanikawa et al., 2018) prompted us to investigate whether this modification occurs at R1810 of RNAP2-CTD and its possible implication in transcription regulation.

Citrullination is a deimination of protein-embedded arginine, which is converted to the non-coded amino acid citrulline (van Venrooij and Pruijn, 2000; Fuhrmann et al., 2015). Citrullination leads to a reduction in hydrogen-bond formation, affects histone-DNA interactions, and influences the chromatin organization. Citrullination also increases the hydrophobicity of proteins that affect the protein folding ability and therefore the functional activity of proteins (Vossenaar et al., 2003; Tanikawa et al., 2018). This reaction is catalyzed by enzymes called peptidyl arginine deiminases (PADIs), which have been associated with diverse disease conditions, such as thrombosis, prion disease, neurological disorders, autoimmune disease, and cancer (Witalison et al., 2015; György et al., 2006; Vossenaar et al., 2003; Baka et al., 2012). Among PADI family members, PADI2 is the most widely expressed isoform and is also overexpressed in breast cancer, where it regulates mammary carcinoma cell migration (Mohanan et al., 2012; Cherrington et al., 2012;

Horibata et al., 2017). Citrullination of core histones has been related to the gene expression, DNA damage responses, and pluripotency (Sharma et al., 2012; Tanikawa et al., 2012; Christophorou et al., 2014), although the underlying mechanisms are largely unknown.

RNAP2-mediated gene expression starts with binding to the gene promoters of basal transcription factors that recruit RNAP2 to form the transcription pre-initiation complex. Shortly after transcription initiation, RNAP2 pauses 30–50 bp downstream of the transcription start sites (TSSs) and requires the activation of P-TEFb (positive transcription elongation factor b) kinase complexes to continue with the productive elongation (Marshall and Price, 1995; Adelman and Lis, 2012; Jonkers and Lis, 2015). Promoter-proximal pausing affects the expression of many genes but is more prominent for highly expressed genes in responses to developmental and environmental stimuli (Zeitlinger et al., 2007; Core et al., 2008; Gilchrist et al., 2010; Day et al., 2016). Recently, RNAP2 pausing was found to inhibit transcriptional initiation, indicating that paused RNAP2 first needs to be released in order to allow a new cycle of transcription initiation (Shao and Zeitlinger, 2017; Gressel et al., 2017). However, despite the strong evidence of RNAP2 pausing, the nature of paused RNAP2 to allow efficient transcription still remains unclear.

Here, we report the discovery that PADI2 citrullinates the R1810 (cit1810) at RNAP2-CTD. The absence of PADI2-mediated cit1810 widely affects transcription and cell proliferation in breast cancer cells. PADI2 occupancy increases with the level of gene transcription. Further, we found that replacing wild-type RNAP2 with the R1810A mutant compromises transcription, reduces interaction with the P-TEFb complex, and leads to accumulation of RNAP2 on the proximal promoter of PADI2-dependent genes. Thus, citrullination of R1810 facilitates interaction with the P-TEFb complex, favoring RNAP2 pause release and promoting transcription of cell cycle genes and cell proliferation of breast cancer cells.

RESULTS

Citrullination of R1810 at RNAP2-CTD

Two arginine residues within non-consensus repeats in human RNAP2-CTD, R1603 and R1810, are conserved in vertebrates. Recently, R1810 within repeat 31 was found to be asymmetrically (Sims et al., 2011) or symmetrically (Zhao et al., 2016) dimethylated, leading to either reduced expression of snRNAs and snoRNA or to efficient transcription termination, respectively. To examine the possibility that R1810 at RNAP2-CTD could be citrullinated in cells, we immunoprecipitated nuclear extracts from the luminal breast cancer cell line T47D (Truss et al., 1995) with a citrulline antibody followed by western blot with an antibody to RNAP2. We found two specific bands migrating as the non-phosphorylated (IIA) and phosphorylated (IIO) forms of the large subunit POLR2A of RNAP2. The IIO band reacted preferentially with α-citrulline compared to the IIA band of RNAP2 (Figure S1A). We raised a polyclonal antibody against a 13-residue peptide centered on R1810, which was replaced by citrulline (Figure 1A, top). This antibody (α-Cit1810) was specific, as it reacted with the citrullinated peptide, but not with the wild-type, methylated (me2aR1810), or phosphorylated (S2 or S5) peptides (Figures S1B and S1C) and mainly recognized the phosphorylated form of RNAP2 in western blots of nuclear extracts from T47D cells (Figures 1A and S1D).

To validate that R1810 is citrullinated, we transiently transfected T47D cells with an α-amanitin-resistant hemagglutinin (HA)-tagged wild-type (WT^r) RNAP2 or with a $R1810A^r$ mutant of RNAP2, followed by α-amanitin treatment to deplete the endogenous RNAP2 (Figures S1E–S1G). Precipitation with anti-HA antibody followed by western blot showed that the WT^r RNAP2, but not the $R1810A^r$ mutant, reacts with α-Cit1810 (Figure 1B). In super-resolution immunofluorescence images of T47D cells, α-Cit1810 decorated bright clusters overlapping with RNAP2, preferentially in its S2 or S5 phosphorylated forms (Figures 1C and 1D). Thus, R1810 is citrullinated in cells prevalently on the phosphorylated actively transcribing form of RNAP2.

Citrullination of R1810 by PADI2

In a search for the responsible enzyme, we found that T47D cells express only *PADI2* and *PADI3* (Figure 2A) among family members. Depletion of PADI2, but not PADI3, reduces R1810 citrullination (Figures 2B, S2A, and S2B). In MCF7 breast cancer cells that express *PADI2* and *PADI4* (Cuthbert et al., 2004; Sharma et al., 2012), depletion of PADI2, but not PADI4, reduced R1810 citrullination (Figures S2C and S2D). To test whether PADI2 acts directly on the RNAP2-CTD, we incubated recombinant PADI2 with either a recombinant glutathione S-transferase (GST)-N-CTD (repeat 1–25.5, including R1603) or with GST-C-CTD (repeat 27–52, including R1810; Figure 2C, left panel). PADI2 citrullinates R1810 in the C-CTD much more efficiently than R1603 in N-CTD (Figure 2C, right panel). Thus, PADI2 is the enzyme responsible for citrullination of R1810 in breast cancer cells.

The affinity of PADI2 for the unmodified R1810 peptide measured by microscale thermophoresis (Jerabek-Willemsen et al., 2011) is $K_d = 220 \pm 54.5$ nM, whereas peptides phosphorylated at S2 or S5 were not bound (Figure 2D), suggesting that the observed S2/S5 phosphorylation in R1810 citrullinated CTD most probably occur outside of repeats 31 and 32. In co-immunoprecipitation experiments using T47D and MCF7 cell extracts, PADI2 but neither PADI3 nor PADI4 interacted with RNAP2 (Figures S2E and S2F). Similarly, an antibody against PADI2 precipitated Cit1810-RNAP2, along with S5P- and S2P-RNAP2 (Figures 2E and S2G). Monoclonal antibodies against RNAP2 phosphorylated at S2 and S5 (see STAR Methods) precipitated Cit1810 RNAP2 as well as PADI2, but not PADI3 (Figure 2F). In T47D nuclear extracts fractionated using size-exclusion chromatography, PADI2 eluted along with phosphorylated RNAP2 in the high-molecular-weight fractions, whereas PADI3 eluted in lower molecular weight range (Figure S2H). Finally, triple-labeling immunofluorescence microscopy showed that PADI2 co-localizes with Cit1810-RNAP2 and with S2P-RNAP2 (Figure S2I). These observations support the association of PADI2 with Cit1810-RNAP2 that is engaged in transcription.

Citrullination of R1810 Regulates Transcription and Cell Proliferation

We next performed mRNA sequencing in control and PADI2 depleted T47D cells (Figure S3A). Strikingly, in global differential

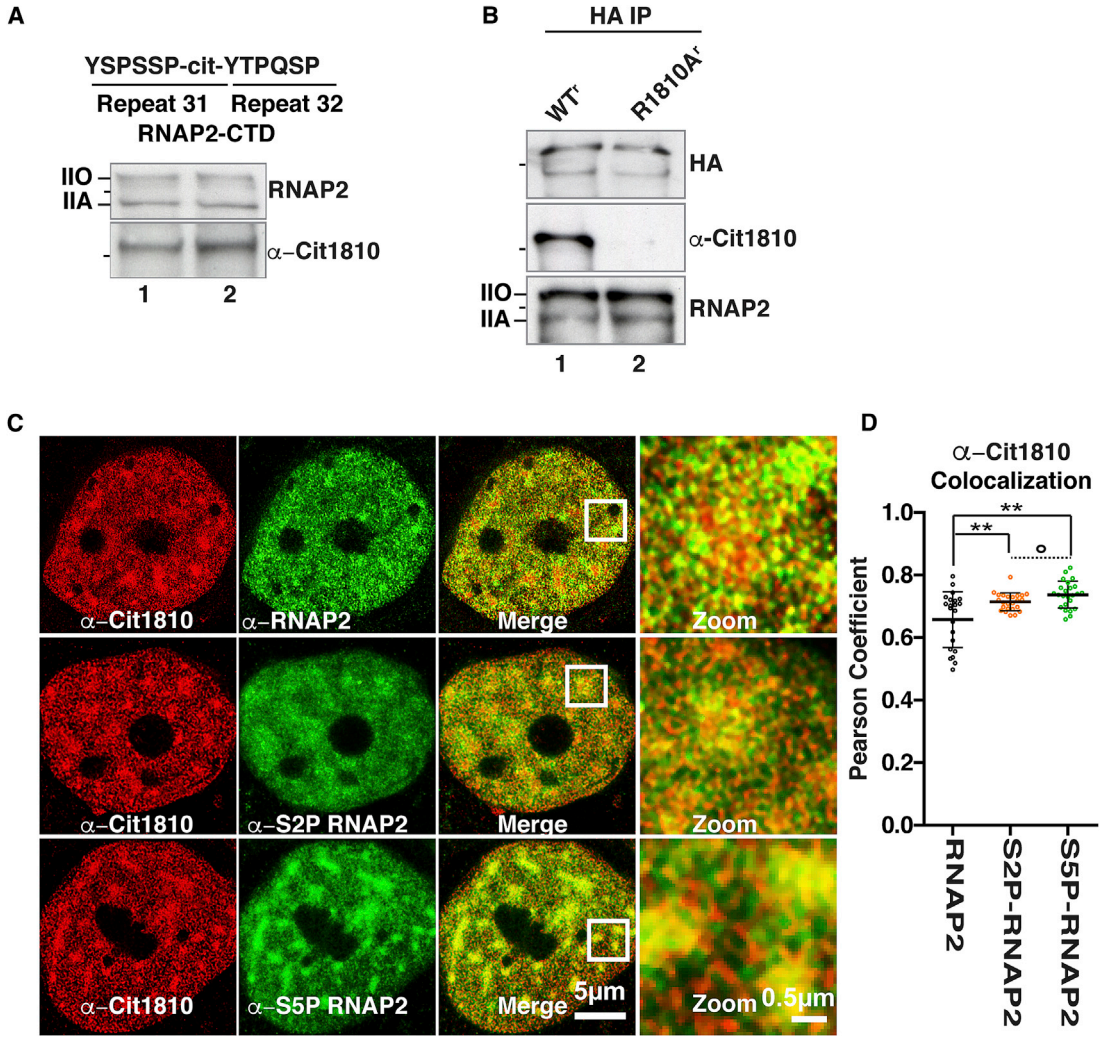


Figure 1. Citrullination of R1810 at RNAP2-CTD, Related to Figure S1

(A) Top: The epitope within repeats 31/32 of the CTD domain of RNAP2 used to generate α-Cit1810. Bottom: Duplicated western blot of T47D nuclear extract with α-Cit1810 and α-total-RNAP2 is shown. Line on the left mark is the migration of the 250-kDa size marker.

(B) Extracts from T47D cells expressing α-amanitin-resistant WT^r or R1810A^r mutant of RNAP2 were precipitated with α-HA antibody and probed with α-Cit1810 or α-RNAP2.

(C) Representative super-resolution images of T47D cells immunostained with α-Cit1810 (red) in combination with α-total-RNAP2 (green), α-S2P-RNAP2 (green), and α-S5P-RNAP2 (green).

(D) Plot representing the mean Pearson correlation coefficient of individual cells for α-Cit1810-RNAP2 with α-total-RNAP2 (n = 22), α-S2P-RNAP2 (n = 24), and α-S5P-RNAP2 (n = 24); values presented as the mean ± SEM. **p value < 0.005; °p value > 0.05.

expression (DEseq) analysis, PADI2 knockdown affected the expression of over 4,000 genes: downregulated (2,186) and upregulated (2,141; Figures 3A and S3B). Gene ontology analysis of the downregulated genes revealed enrichment in RNAP2-mediated transcription and cell proliferation (Figure S3C; Table S1). Reduced expression was validated by qRT-PCR for several genes, including *SERPINA6*, *c-MYC*, and *HMGN1* genes, and control genes *GSTT2* and *LRRC39* were not affected (Figures 3B and S3D). Depletion of CARM1 or PRMT5, which catalyze dimethylation of R1810 (Sims et al., 2011; Zhao et al., 2016), did not affect the expression of PADI2-dependent genes (Figures S3E and S3F).

To explore the direct effect of PADI2 on nascent transcription, we performed chromatin-associated RNA sequencing (ChrRNA-seq) (Nojima et al., 2016) in control and PADI2-depleted cells. We found that ~2,000 transcripts were significantly affected by the PADI2 knockdown, and the majority of them were downregulated (1,884; Figures 3C, 3D, and S3G). ChrRNA-seq changes were validated on PADI2-dependent genes by qRT-PCR (Figure 3E). We conclude that PADI2 is required for efficient transcription and that upregulation of mRNAs upon PADI2 depletion may be a consequence of downregulation of transcription-relevant genes, although we cannot exclude citrullination of other PADI2 substrates. To support this conclusion, we

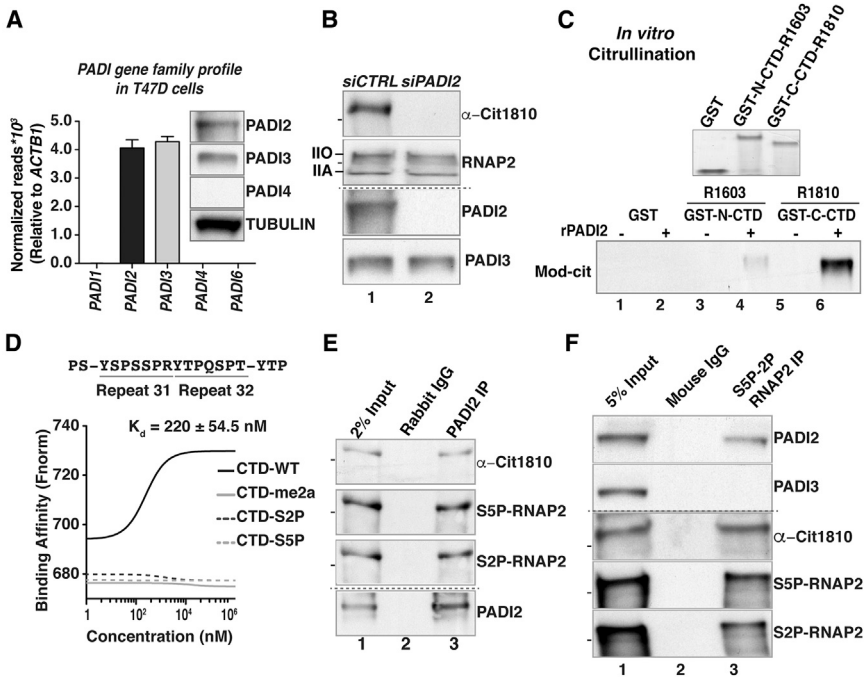


Figure 2. PADI2 Citrullinates R1810 at RNAP2-CTD, Related to Figure S2

(A) Bar plot showing the normalized reads of *PADI* gene family members from two RNA sequencing experiments performed in T47D cells. The normalized reads are represented relative to *ACTB1* gene. Values are means ± SEM. (Inset) Western blots performed on T47D nuclear extract with antibodies to PADI2, 3, and 4 and tubulin are shown.

(B) Nuclear extracts from T47D cells transfected with small interfering RNA (siRNA) control (si*CTRL*) or siRNA against PADI2 (si*PADI2*) are probed with α-Cit1810, α-total-RNAP2, α-PADI2, and α-PADI3.

(C) Top: Coomassie blue staining of SDS-PAGE with recombinant GST-tagged, GST-N-CTD, and GST-C-CTD proteins used for the citrullination assay. Bottom: *In vitro* citrullination immunoblot with or without recombinant PADI2 (rPADI2) using as substrate the N-terminal half of the CTD containing R1603 (lanes 3 and 4) or the C-terminal half containing R1810 (lanes 5 and 6), both linked to GST, is shown. As a control, GST was also tested (lanes 1 and 2).

(D) Microscale thermophoresis assay showing the affinity of recombinant PADI2 for the indicated wild-type and modified CTD-RNAP2 peptides encompassing the R1810. y axis represents the binding affinity as normalized fluorescence (Fnorm) (see STAR Methods).

(E and F) Immunoprecipitation with (E) α-PADI2, (F) α-S2P/S5P RNAP2, or non-immune mouse or rabbit immunoglobulin G (IgG) of T47D extracts followed by western blot with the indicated antibodies.

performed mRNA sequencing in T47D cells expressing only the α-amanitin-resistant HA-tagged WT^r or the R1810A^r mutant form of RNAP2. We found 1,392 downregulated genes in cells expressing R1810A^r mutant RNAP2, of which 939 (67.4%) were also dependent on PADI2. We confirmed this finding by qRT-PCR of a PADI2-dependent *SERPINA6*, *c-MYC*, and *HMGN1* and control genes *GSTT2* and *LRRC39* (Figure 3F). Thus, PADI2 and R1810 are required for efficient transcription.

Because many PADI2- and R1810-dependent genes are related to cell proliferation, we monitored T47D cell proliferation after PADI2 depletion (si*PADI2*), inhibition with Cl^- amidine, or in cells expressing only the R1810A^r mutant of RNAP2. In all cases, we found a significant reduction of cell proliferation (Figure 3G). PADI2-depleted cells were arrested at the G1 phase of the cell cycle (Figure S3H), as expected given the downregulation of genes critical for G1 phase progression, including *CCND1*, *PLK1* in presence of R1810A^r mutant as compared to WT^r RNAP2, or PADI2 depletion (Figures 3H and S3I–S3K; Table S2).

PADI2 Is Enriched on Active Genes

Chromatin immunoprecipitation sequencing (ChIP-seq) of PADI2 in T47D cells showed that 60% of chromatin-bound PADI2 was localized over protein-coding gene sequences, within 3 kb upstream of the TSS (transcription start site) and 3 kb downstream of the TTS (transcription termination site) (q value $\leq$ 0.005; Figure 4A, left panel). The highest enrichment (2.5-fold) was found in the coding exons, followed by the 3-kb region downstream of the TTS (1.6-fold; Figure 4A, right panel). Overall, PADI2 occupancy overlapped with RNAP2 binding measured by ChIP-seq (Figure 4B). To explore whether PADI2 binding is related to transcription, we separated genes in 4 quartiles according to their transcription level: high (100%–95%); medium (95%–50%); low (last 50%); and silent (non-significant expression; Baranello et al., 2016). We found that RNAP2 and PADI2 occupancy increased in parallel with the gene expression levels (Figures 4C and S4A), supporting a role of PADI2 in transcription.

To verify the specificity of the PADI2 occupancy, we performed PADI2 ChIP-qPCR in control (si*CTRL*) and PADI2 knockdown (si*PADI2*) over high (*SERPINA6* and c-*MYC*) and low expressed (*GSTT2*) genes and found that PADI2 depletion drastically decreased the levels (Figure S4B). PADI2 occupancy was significantly higher on genes downregulated by PADI2 depletion compared to those non-regulated (Figure S4C). Finally, RNAP2-ChIP followed by PADI2 re-ChIP revealed the association of RNAP2 and PADI2 at regulatory regions and gene bodies of the highly transcribed *SERPINA6* and c-*MYC* genes, but not in the low expressed *GSTT2* gene (Figure S4D). Thus, PADI2 seemed to be part of the transcription machinery in highly expressed genes.

Citrullination of R1810 Controls RNAP2 Pausing

Next, we analyzed RNAP2 occupancy by ChIP-qPCR in T47D cells prior and after PADI2 depletion. We observed accumulation of RNAP2 around the TSS of the highly expressed *SERPINA6*, *c-MYC*, and *HMGN1* genes (Figures S5A and S5B, right panel). This effect is also dependent on R1810, as it is observed in T47D cells expressing only the α-amanitin-resistant HA-tagged R1810A^r mutant form of RNAP2 in comparison with cells expressing only the WT^r RNAP2 (Figures 5A and S5B,

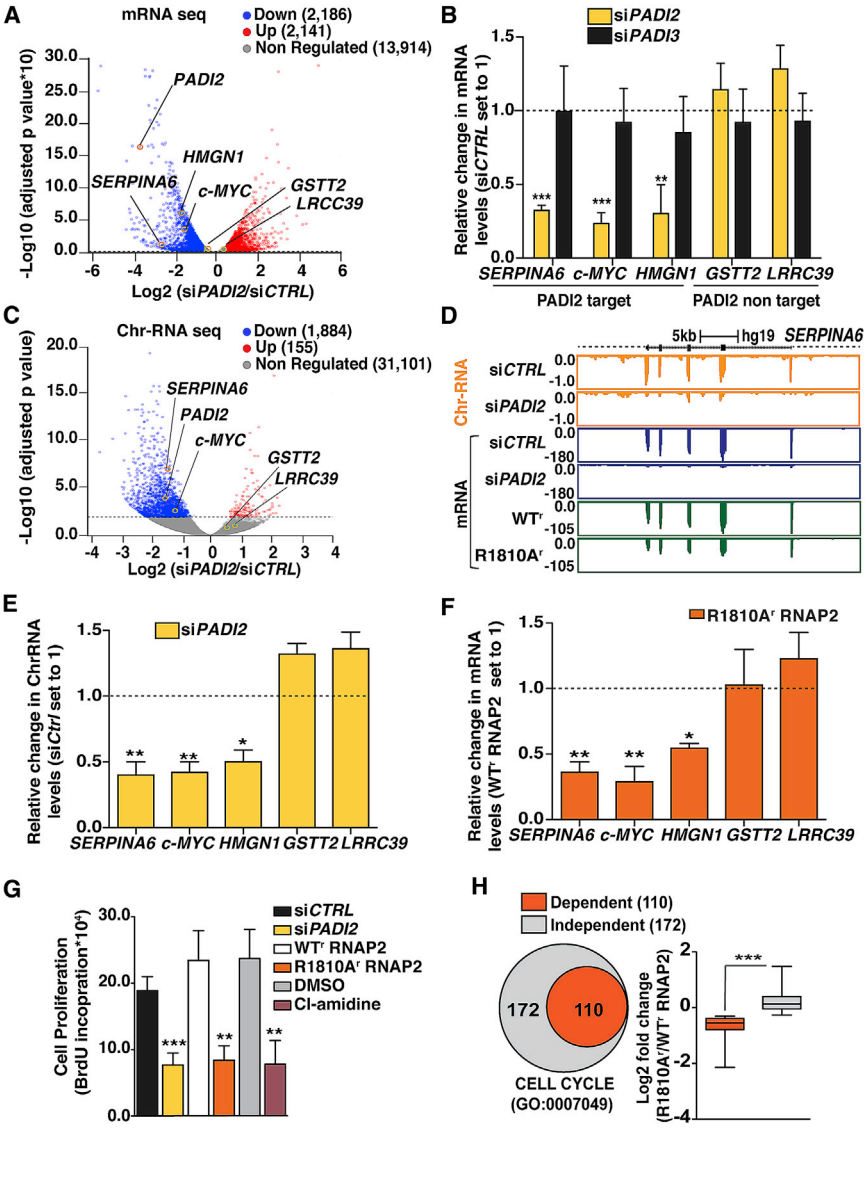


Figure 3. PADI2-Mediated cit1810 at RNAP2-CTD Regulates Transcription and Cell Proliferation in Breast Cancer Cells, Related to Figure S3

(A) Volcano plot showing genome-wide mRNA changes after PADI2 depletion from biological replicates. The x axis represents log2 expression fold changes (FCs), and the y axis represents the adjusted p values (as −log10). Differentially expressed genes (FC > 1.5 or <1/1.5 and p value < 0.01) are shown; the positions of *PADI2* and genes used for validation are also indicated.

(B) qRT-PCR validation in T47D cells transfected with si*CTRL*, si*PADI2*, and si*PADI3*. Changes in mRNA levels were normalized to *GAPDH* mRNA. Data represent mean ± SEM of at least three biological experiments as in other plots in the figure. **p value < 0.01; **p value < 0.001.

(C) Volcano plot showing genome-wide chromatin-associated RNAs changes before and after PADI2 depletion from independent replicates. The x axis represents the log2 FCs, and the y axis represents the adjusted p values (−log10). The dotted line indicates the cutoff p value < 0.01. Differential expressed genes (FC > 1.5 or 1/1.5 and the p value < 0.01) are shown.

(D) Browser snapshots of *SERPINA6* gene in T47D cells showing chromatin-associated RNAs sequencing profile (orange) and mRNA sequencing profiles after *PADI2* knockdown (blue) and expressing WT^r and R1810^r mutant form of RNAP2 (green). Scale is indicated on the top of the gene.

(E) qRT-PCR on chromatin-associated RNA (ChrRNA) in T47D cells transfected with si*CTRL* and si*PADI2* RNAs. Data are normalized to *GAPDH* ChrRNA expression level. Data represent mean ± SEM of at least three biological experiments. *p value ≤ 0.05; **p value ≤ 0.01.

(F) T47D cells expressing only α-amanitin-resistant HA-tagged WT^r or R1810A^r mutant form of RNAP2 were used for qRT-PCR of mRNA from PADI2-dependent genes (*SERPINA6*, *c-MYC*, and *HMGN1*) and for control genes (*GSTT2* and *LRRC39*). Data normalized to GAPDH mRNA expression level. Data represent mean ± SEM of at least three biological experiments. *p value ≤ 0.05.

(G) Cell proliferation of T47D cells in the absence (si*CTRL* or DMSO) or presence of PADI2 depletion (si*PADI2*), PADI2 inhibitor (Cl^- amidine in DMSO), and of cells expressing α-amanitin-resistant HA-tagged WT^r and R1810A^r mutant form of RNAP2. Data represent mean ± SEM of at least six biological experiments. **p value ≤ 0.01; ***p value ≤ 0.001.

(H) Left: Venn diagram showing the set of genes related to cell cycle (GO: 0007049; n = 315; out of them 282 genes expressed in T47D cells) that are downregulated in T47D cells expressing the R1810A^r mutant of RNAP2 in comparison to WT^r RNAP2 (R1810-dependent) versus R1810-independent gene (see Table S2). Right: Boxplot showing log2 fold change (R1810A^r/WT^r RNAP2) for R1810-dependent and independent cell cycle genes. Each box in the panel represents the interquartile range; whisker extends the box to the highest and lowest values, and horizontal lines indicate the median value. Dependent genes showed significant lower mRNA levels than independent genes (***p value < 0.0001; calculated by Wilcoxon-Mann-Whitney test).

left panel), suggesting absence of PADI2-mediated citrullination of R1810 leads to RNAP2 pausing. In T47D cells expressing only the R1810A mutant compared to cells expressing the WT^r RNAP2, ChIP-seq experiments showed remarkably high accumulation of RNAP2 at proximal promoters (Figure 5B) and a corresponding change in the pausing index that correlated with the gene expression levels (Figure 5C). Previously, Raji cells expressing R1810A also showed paused RNAP2 at proximal promoter for the 5% most highly expressed genes (Zhao et al., 2016).

Focusing on PADI2-dependent genes (n = 2,186), we found a pronounced accumulation of RNAP2 around the TSS with a significantly increased pausing index in presence of R1810A mutant as compared to wild-type form of RNAP2 (Figures 5D–5F). Similarly, genes downregulated in the presence of the R1810A^r RNAP2 mutant (n = 939) also showed significantly higher pausing index as compared to cells expressing WT^r RNAP2 (Figure S5D). Thus, PADI2 depletion or absence of R1810 leads to RNAP2 accumulation on the promoters of highly expressed genes and to reduced level of S2P and S5P forms of RNAP2 (Figure S5C).

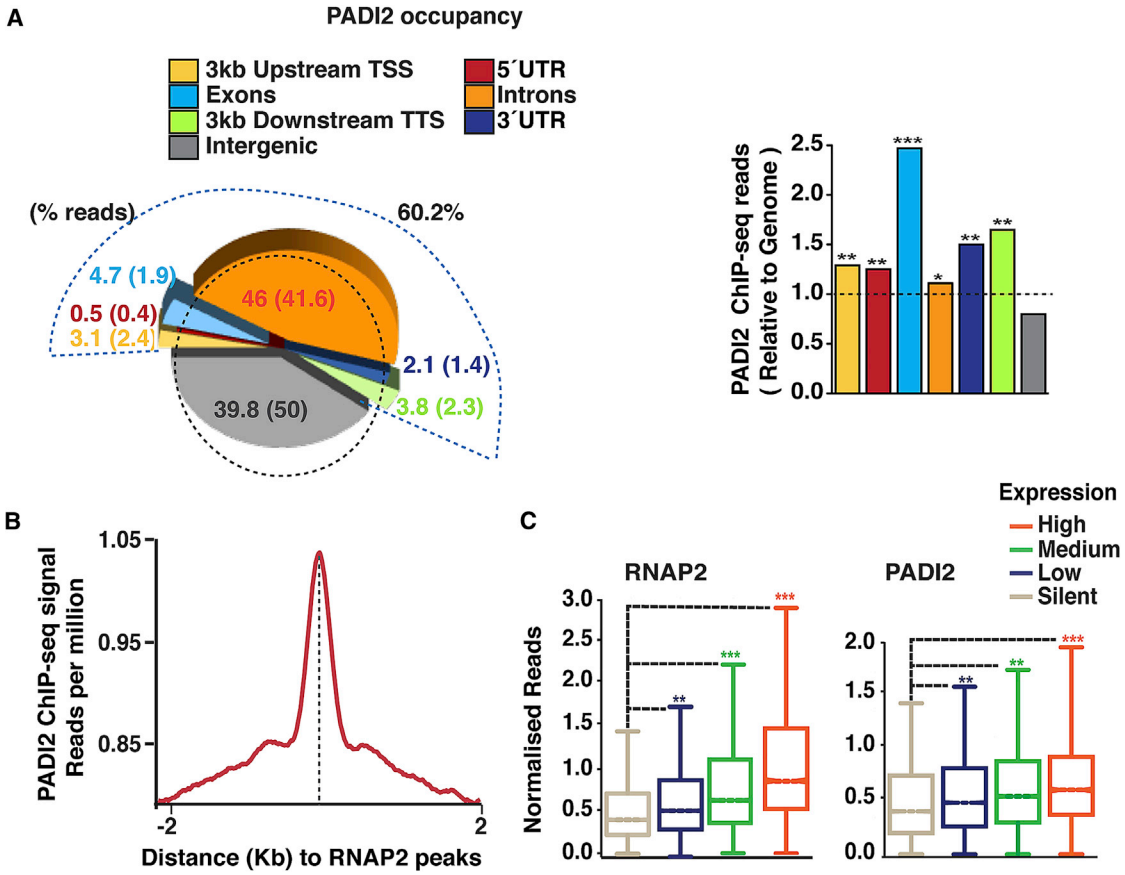


Figure 4. PADI2 Occupancy on Active Genes, Related to Figure S4

(A) Left: Spie chart showing the distribution of PADI2 ChIP-seq peaks over various genomic regions. A dashed curved line indicates the region from 3 kb upstream of TSS to 3 kb downstream of TTS (or polyadenylation site); the numbers in parentheses show the proportion occupied by each region in the genome. Right: Enrichment of normalized PADI2 reads in various genome regions relative to random distribution is shown (*p value < 10^{-2}; **p value < 10^{-3}; ***p value < 10^{-4}).

(B) Distribution of normalized PADI2 reads around the center of RNAP2 peaks in T47D cells.

(C) RNAP2 and PADI2 occupancy across genes classified with increasing levels of expression. p value was calculated by Wilcoxon-Mann-Whitney test in comparison to silent genes as indicated (**p value < 10^{-3}; ***p value < 10^{-5}).

We also found that genes upregulated upon depleting PADI2 or upon expressing the R1810A^{r} mutant RNAP2 exhibited lower pausing index (Figures S5E and S5F), suggesting that they do not need to overcome RNAP2 pausing to maintain their expression. In summary, we found that PADI2 citrullination of CTD R1810 is important for RNAP2 promoter pause release.

Cit1810 at RNAP2-CTD Recognized by P-TEFb

Citrullination is known to modulate functional protein-protein interactions (Tanikawa et al., 2012, 2018; Vossenaar et al., 2003). We wondered whether the Cit1810 influences the interaction of RNAP2 with the components of P-TEFb complex CDK9 and CCNTI (Cylin T1), which are required for RNAP2 pause release and productive elongation (Jonkers and Lis, 2015; Gressel et al., 2017). Immunoprecipitation of T47D cells extracts with a PADI2 antibody, precipitated CDK9 and CCNT1 (Figure 6A, left panel), and conversely a CDK9 antibody pulled down PADI2 (Figure 6A, right panel), indicating that PADI2 associates with the P-TEFb complex. Immunoprecipitation of extracts from PADI2-depleted cells (si*PADI2*) showed strong reduction in RNAP2 interaction with CDK9 and CCNT1 compared to control cells (si*CTRL*; Figure 6B). In T47D or Raji cells expressing only the α-amanitin-resistant HA-tagged WT^{r} RNAP2, CDK9 and CCNT1 were also immunoprecipitated with HA-tag antibody, and the interaction was significantly reduced in cells expressing the R1810A^{r} mutant RNAP2 (Figures 6C and 6D). Thus, PADI2 and R1810 are required for the association of RNAP2 with the CDK9-CCNT1 complex. Next, we performed peptide pull-down assays with T47D nuclear extracts using biotinylated wild-type (R1810) and cit1810 RNAP2-CTD peptides (Figure 6E, upper panel). The cit1810 peptide pulls down the CDK9-CCNT1 complex much more efficiently than the wild-type peptide (R1810; Figure 6E, lower panel), confirming that PADI2-mediated cit1810 is required to facilitate cooperation of RNAP2 with CDK9-CCNT1 complex.

To explore whether R1810 is important for CDK9 recruitment to the promoter region of genes, we performed ChIP-seq with CDK9 antibody in T47D cells expressing only the α-amanitin-resistant HA-tagged R1810A^{r} or the WT^{r} RNAP2. We found that CDK9

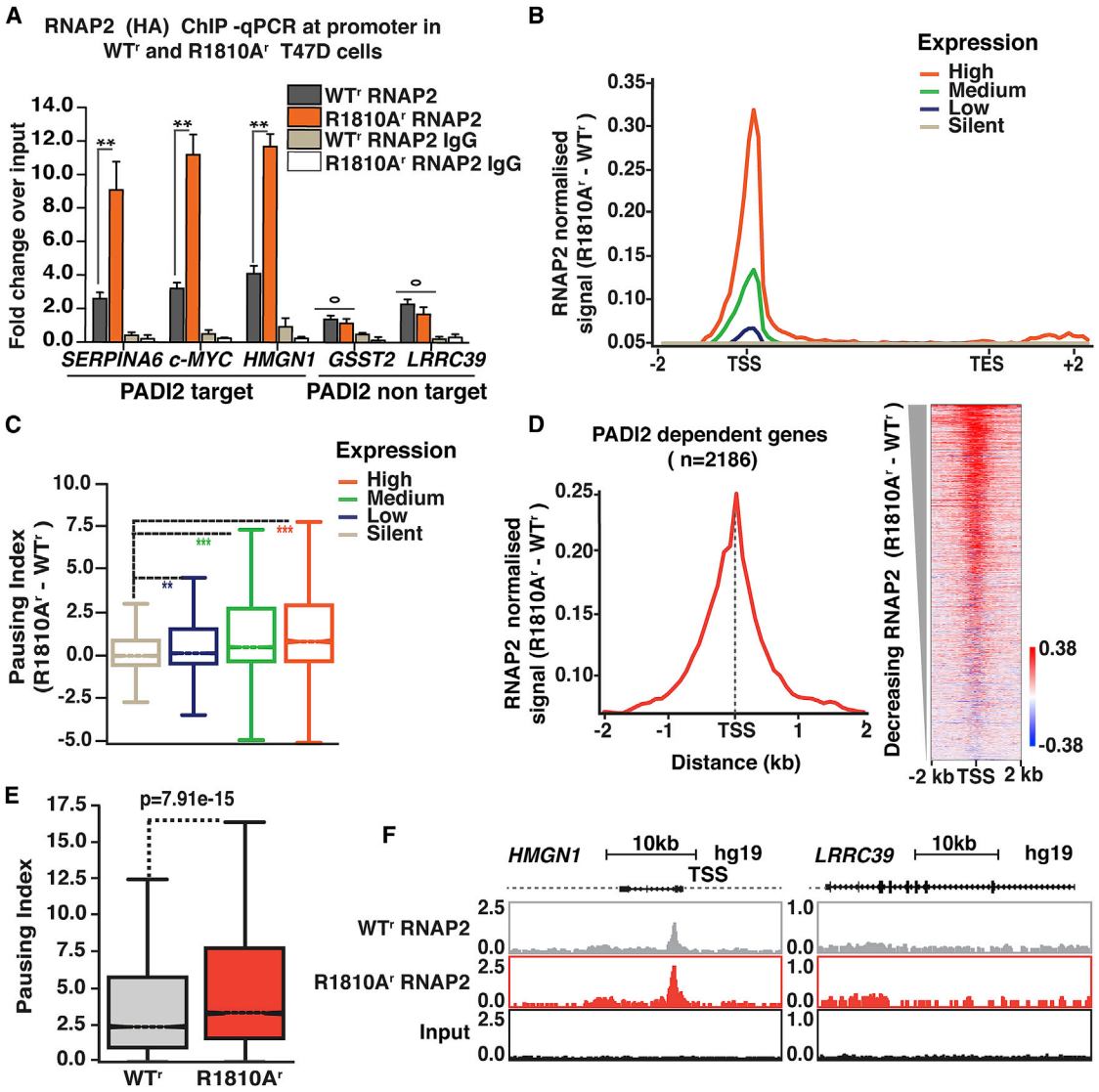


Figure 5. Cit1810 at CTD-RNAP2 Regulates Pausing in Breast Cancer Cells, Related to Figure S5

(A) RNAP2 ChIP qPCR assay performed in T47D cells expressing only the HA-tagged wild-type (WT^r) or R1810A^r mutant of RNAP2 with the HA antibody. Non-immune IgG was used as negative control. y axis: fold change over the input samples. Data represent mean ± SEM of at least three biological experiments, **p value $\leq$ 0.01; °p value > 0.05.

(B and C) Difference in RNAP2 density in T47D cells expressing HA-tagged R1810A^r mutant versus WT^r RNAP2 (B) across genes classified by expression.

(C) Pausing index of RNAP2 as indicated. p value was calculated by Wilcoxon-Mann-Whitney test in comparison to silent genes (**p value < 10^{-3}; ***p value < 10^{-5}).

(D and E) PADI2-dependent genes (n = 2,186) showing (D; left) average profile of difference in RNAP2 density (R1810A^r − WT^r) around TSS. (Right) Heatmap at TSS of genes ranked from highest to lowest RNAP2 density is shown (R1810A^r − WT^r).

(E) Higher pausing index in cell expressing R1810A^r mutant as compared to WT^r form of RNAP2.

(F) Browser snapshots showing RNAP2 occupancy for *HMGN1* and control gene *LRRC39* in cells expressing HA-tagged WT^r or R1810A^r form of RNAP2.

occupancy around the TSS decreased remarkably in the presence of R1810A mutant compared to wild-type form of RNAP2, particularly in highly expressed genes (Figure 6F), suggesting that the integrity of the R1810 is required for CDK9 recruitment. When we compared PADI2 dependent and non-dependent genes, we also found a significant decrease of CDK9 occupancy in the presence of R1810A mutant as compared to wild-type form of RNAP2 (Figure 6G). We validated the ChIP-seq results by ChIP-qPCR and confirmed that mutation of R1810 significantly decreases the recruitment of the CDK9 to the promoter of PADI2 target genes *SERPINA6*, *c-MYC*, and *HMGN1*, but not to the low expressed genes *GSST2* and *LRRC39* (Figure 6H). Altogether, these data support the idea that PADI2-mediated citrullination of R1810 at RNAP2-CTD facilitates the recruitment of P-TEFb complex that promotes RNAP2 pause release and transcription of actively expressed genes involved in cellular proliferation.

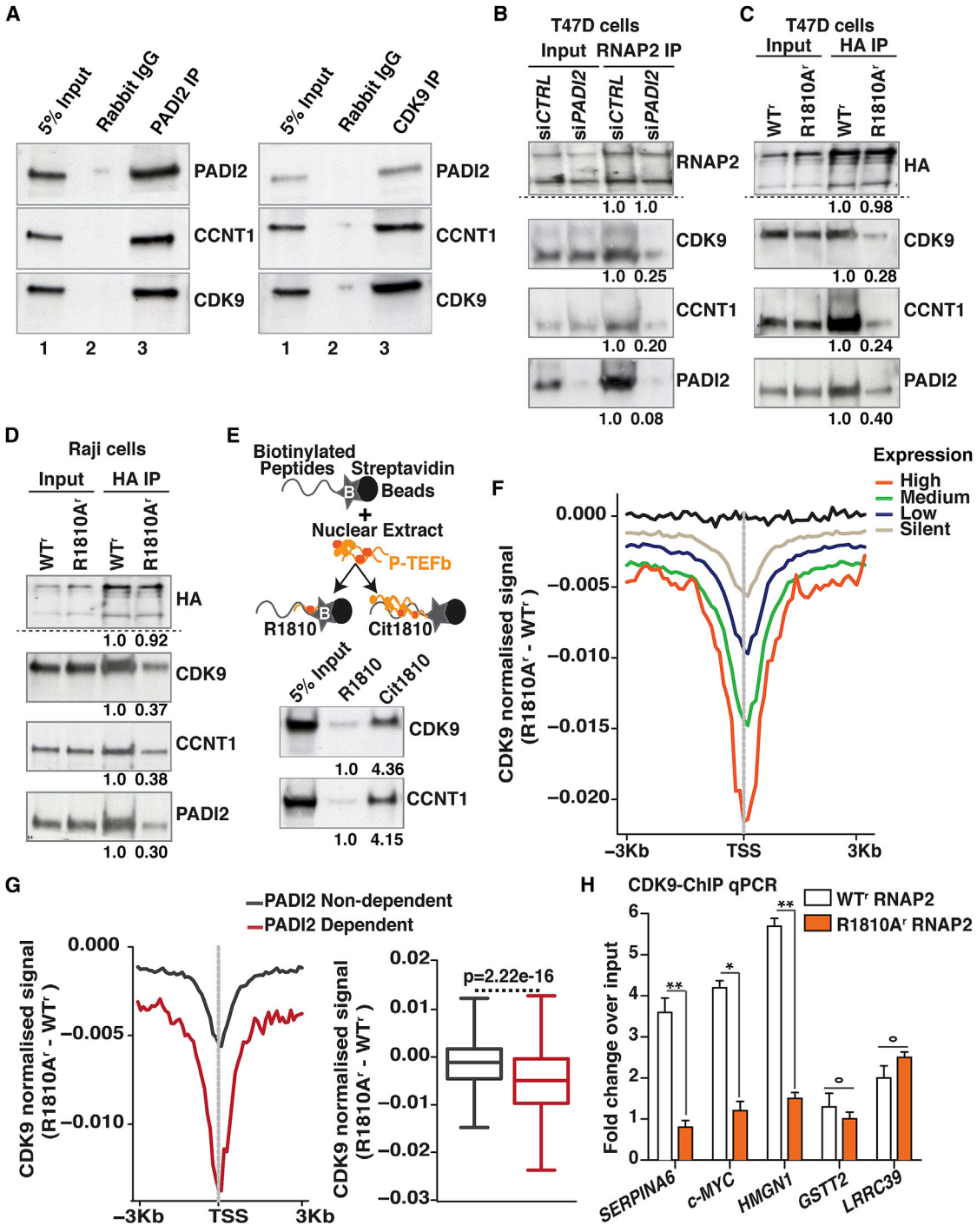


Figure 6. Cit1810 at RNAP2-CTD Is Recognized by P-TEFb

(A) Immunoprecipitation with α-CDK9 (left) and α-PADI2 (right) or non-immune rabbit IgG of T47D extracts followed by western blot with the indicated antibodies.

(B) Extracts from T47D cells in the presence (si*Ctrl*) or absence (si*PADI2*) of PADI2 were immunoprecipitated with α-total-RNAP2 followed by western blot for the indicated antibodies along with relative quantifications underneath.

(C and D) Immunoprecipitation of extracts from (C) T47D and (D) Raji cells expressing only the HA-tagged α-amanitin-resistant WT^r or R1810A^r mutant of RNAP2 were precipitated with α-HA antibody and probed with the indicated antibodies. The relative quantification is shown underneath each gel.

(E) (Top) Schematic representation of the pull-down assays with T47D nuclear extracts and RNAP2-CTD biotinylated peptides (wild-type R1810 or cit1810). (Bottom) Results of the pull-down experiments with wild-type (R1810) or cit1810 biotinylated RNAP2-CTD peptides shown as western blots probed with the indicated antibodies against CDK9 and CCNT1 are shown. Quantification of the increase with the Cit1810 relative to the wild-type R1810 is shown underneath.

(legend continued on next page)

PADI2 Unique Residues Contribute to Citrullinate R1810 at RNAP2-CTD

We wondered about the reason why only PADI2, but not PADI3 or PADI4, carries out citrullination of R1810 at RNAP2-CTD. To address this issue, we used the published structure of the PADI2 (Slade et al., 2015) to model the attachment of RNAP2-CTD peptide encompassing R1810 (Figure 7A) and performed a similar structural modeling with the amino acids of PADI3. Our analysis revealed that the predicted binding score energies calculated with Rosetta were significantly lower for PADI2 in comparison to PADI3 (Figure 7B; Videos S1 and S2; see STAR Methods), indicating a higher affinity of PADI2 for the R1810 peptide.

We next examined the PADI2 residues contributing most to the affinity for the R1810 peptide and analyzed their conservation in the PADI family. We found that non-conserved residues in PADI2 accumulate at the rim of the catalytic domain (Figure S6), most likely contributing to the R1810 peptide binding. We calculated Rosetta score for all PADI2 interface residues (including conserved and non-conserved) and ranked them according to conservation among PADI family members (Figure S7A; Table S3). Next, we choose PADI2-specific residues R580 and L642 and also two other PADIs conserved residues D374 and S401 that showed comparable low binding energy of PADI2 toward the R1810 peptide (Figures 7A and S7A) and introduced mutations that changed chemical properties of these four residues while maintaining the volume (Figure S7B). Using GST C-CTD that encompasses R1810 as a substrate for *in vitro* citrullination assays, we found that the R580E and L642T mutations drastically reduced citrullination activity, whereas the D374K and S401A have a non-significant effect (Figure 7C), confirming the specificity of PADI2 unique residues for citrullination of R1810.

DISCUSSION

In this study, we identify a novel post-translational modification of the RNAP2-CTD, namely the citrullination by PADI2 of R1810. This modification is coupled with the active form of RNAP2 and regulates the transcription of highly expressed genes involved in cell proliferation by mediating the interaction with P-TEFb complex and favoring RNAP2 pause release. We detected this modification initially in T47D breast cancer cells with a pan-citrulline antibody and confirmed it with Cit1810 antibody generated using a 13-mer CTD peptide centered on Cit1810. In breast cancer cells, PADI2, but not PADI3 or PADI4, specifically catalyze R1810 citrullination. Inhibition or depletion of PADI2 compromises transcription of thousands of highly expressed genes, as monitored by mRNA-seq analysis. Many of these PADI2-dependent genes are involved in key biological functions, including RNAP2 transcription and cell proliferation. In chromatin RNA-seq, we found that PADI2 is mainly involved in active transcription, leading us to conclude that PADI2 participates in facilitating active transcription. However, in mRNA-seq analysis, we also find genes upregulated upon PADI2 depletion. Their upregulation could be an indirect consequence of changes in the expression of PADI2-dependent genes or could be related to the need of R1810 dimethylation for proper transcription termination (Zhao et al., 2016). This remains to be directly demonstrated.

The previously reported asymmetrical (Sims et al., 2011) and symmetrical (Zhao et al., 2016) dimethylation of R1810 occurs mainly in hypo-phosphorylated RNAP2, as detected only after phosphatase treatment. In contrast, we find that cit1810 preferentially associates with the transcriptionally engaged phosphorylated RNAP2. Depletion of the methyltransferases responsible for R1810 dimethylation, CARM1 and PRMT5, did not affect the expression of PADI2-dependent genes, and depletion of PADI2 did not change the expression of a broad range of snRNAs, the targets of CARM1. As PADI2 cannot act on methylated R1810, and citrullination will preclude methylation by arginine methyltransferases, it seems that these are alternative types of modifications influencing different stages of transcription, as observed in other arginine residues that can undergo methylation and citrullination (Tanikawa et al., 2018; Cuthbert et al., 2004). This implies the dynamic nature of R1810 modifications, which change the docking surface for regulatory protein complexes to control various phases of transcription.

Our results of RNAP2 ChIP-seq show that PADI2-mediated cit1810 is required for RNAP2 pause release or high turnover (Krebs et al., 2017) at promoters of highly expressed genes that maintain cellular proliferation (Day et al., 2016; Gilchrist et al., 2010; Zeitlinger et al., 2007). In search for the molecular mechanism, we found that PADI2 interacts with the P-TEFb kinase complex, which is needed for RNAP2 pause release and productive transcript elongation. This interaction depends on R1810, supporting a function of cit1810 to facilitate transcription (Figure 7D; Video S3. Most likely, citrullination of R1810 at RNAP2-CTD is the key mechanism to maintain the transcription level of highly expressed genes involved in tumor progression and metastasis, essentially needed to overcome the RNAP2 pausing. This assumption is coherent with the recent findings that inhibition of PADI2 activity suppresses the mammary gland tumor invasion in mice (Horibata et al., 2017) and reduces the mammary cancer progression in dogs and cats (Ledet et al., 2018). Remarkably, PADI2-null mice are viable (van Beers et al., 2013), suggesting that PADI2-mediated cit1810 RNAP2-CTD is not needed during normal development or that this function is fulfilled by another member of the PADI family. This last

(F) Average profile of CDK9 density around TSS in ChIP-seq experiments using CDK9 antibody in T47D cells expressing the HA-tagged wild-type (WT^r) or $R1810A^r$ mutant of RNAP2 and difference ($R1810A^r - WT^r$) across genes classified by expression level. Black line representing signal difference from random regions is shown.

(G) Left: Similar as in (F) for PADI2-dependent genes versus non-dependent genes (siPADI2/si*CTRL*; FC < 1.5; FC > 1/1.5, p value > 0.05). Right: Boxplot shows average difference in CDK9 profiles in PADI2-dependent versus non-dependent genes. p value calculated by Wilcoxon-Mann-Whitney test.

(H) Fold change over input in CDK9-ChIP qPCR assay on the promoter region of three PADI2-dependent genes (*SERPINA6*, *c-MYC*, and *HMGN1*) and two non-dependent genes (*GSTT2* and *LRRC39*) in T47D cells expressing the HA-tagged wild-type (WT^r) or $R1810A^r$ mutant of RNAP2. Values are means ± SEM. *p value < 0.05; **p value < 0.01; °p value > 0.05.

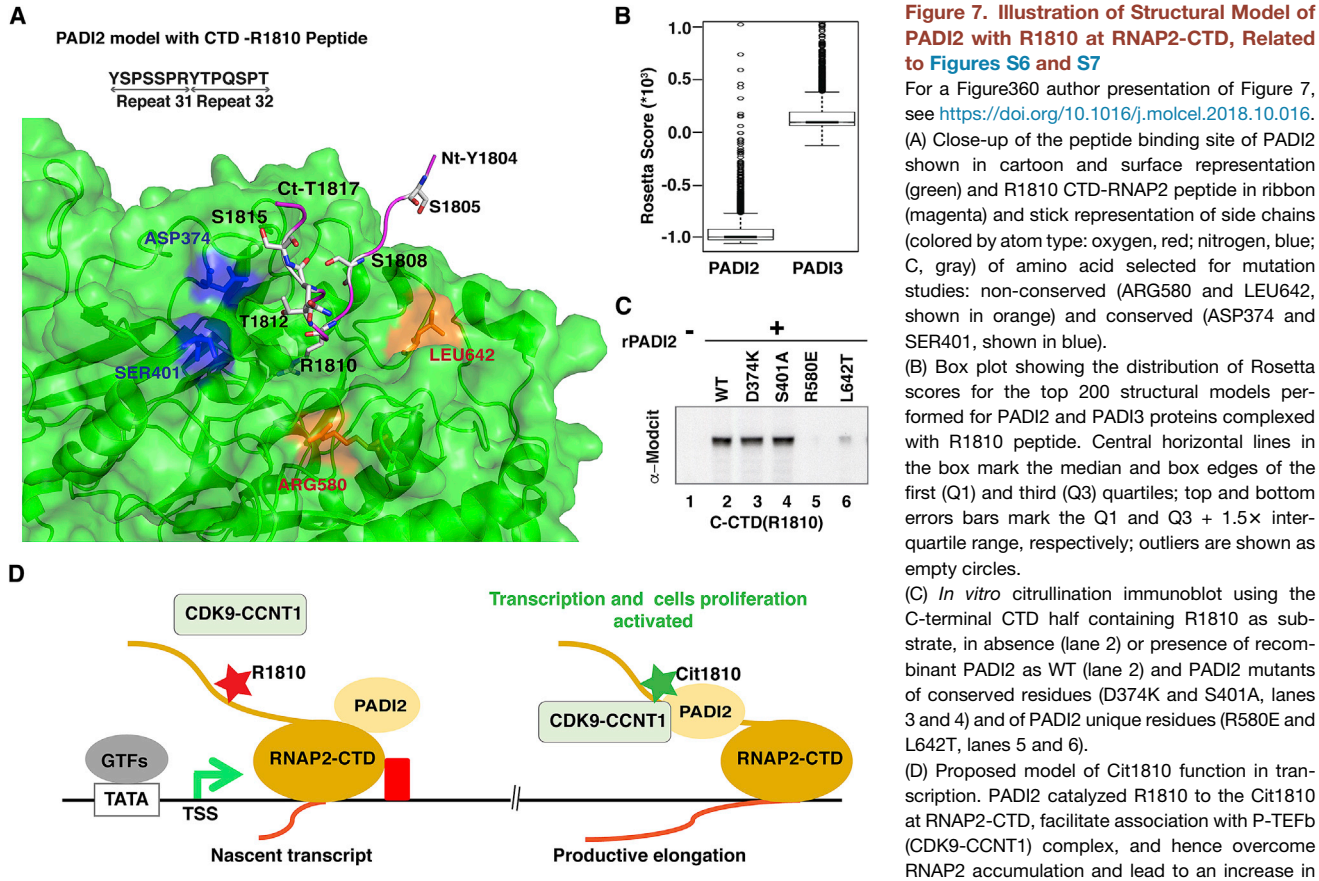


Figure 7. Illustration of Structural Model of PADI2 with R1810 at RNAP2-CTD, Related to Figures S6 and S7

For a Figure360 author presentation of Figure 7, see https://doi.org/10.1016/j.molcel.2018.10.016. Figure360▷

(A) Close-up of the peptide binding site of PADI2 shown in cartoon and surface representation (green) and R1810 CTD-RNAP2 peptide in ribbon (magenta) and stick representation of side chains (colored by atom type: oxygen, red; nitrogen, blue; C, gray) of amino acid selected for mutation studies: non-conserved (ARG580 and LEU642, shown in orange) and conserved (ASP374 and SER401, shown in blue).

(B) Box plot showing the distribution of Rosetta scores for the top 200 structural models performed for PADI2 and PADI3 proteins complexed with R1810 peptide. Central horizontal lines in the box mark the median and box edges of the first (Q1) and third (Q3) quartiles; top and bottom errors bars mark the Q1 and Q3 + 1.5× interquartile range, respectively; outliers are shown as empty circles.

(C) *In vitro* citrullination immunoblot using the C-terminal CTD half containing R1810 as substrate, in absence (lane 2) or presence of recombinant PADI2 as WT (lane 2) and PADI2 mutants of conserved residues (D374K and S401A, lanes 3 and 4) and of PADI2 unique residues (R580E and L642T, lanes 5 and 6).

(D) Proposed model of Cit1810 function in transcription. PADI2 catalyzed R1810 to the Cit1810 at RNAP2-CTD, facilitate association with P-TEFb (CDK9-CCNT1) complex, and hence overcome RNAP2 accumulation and lead to an increase in transcription and cell proliferation.

possibility seems unlikely, as by modeling the structure of PADI2 with the bound R1810 CTD peptide, we identified essential PADI2-specific amino acid residues that are not conserved in other PADI family members.

Although our results support a function of PADI2-mediated citrullination of CTD R1810 in transcription elongation, we cannot exclude an pleiotropic action of PADI2 on other substrates, including citrullination arginine 26 at histone H3 (H3R26), which could be involved in local chromatin decondensation and transcription activation (Zhang et al., 2012). Another intriguing open question concerns the implications of the proposed irreversibility of the citrullination reaction (Cuthbert et al., 2004; Wang et al., 2004). In the absence of enzymes that erase citrullination, alternative mechanisms may exit to replace the citrullinated RNAP2 before reaching transcription termination. Given that arginine-mediated interactions between intrinsically disordered protein domains, including the CTD of RNAP2, and RNAs or poly(ADP-ribose) are important in the formation of liquid droplets within the cell nucleus (Altmeyer et al., 2015; Hnisz et al., 2017; Harlen and Churchman, 2017), we cannot exclude that citrullination of R1810 could also participate in modulating these interactions and those influencing transcriptional output. Further work will be required to investigate these possibilities.

Many elongation factors and kinases are implicated in the control of RNAP2 transcription pause release, a mechanism that controls the expression of genes involved in cancer progression and metastasis, like *CDK9*, c-*MYC*, and *JMJD6* (Zhang et al., 2017; Bywater et al., 2013; Miller et al., 2017). *PADI2* is also found overexpressed in breast cancer (Cherrington et al., 2012) and other cancers (Guo et al., 2017). Indeed, we found that PADI2 depletion or mutation of R1810 reduced cell proliferation of breast cancer cells by modulating cell cycle progression. Also, among *PADI* family members, only *PADI2* is overexpressed in breast cancer and other cancers and its overexpression correlates with poor prognosis (Cherrington et al., 2012; Curtis et al., 2012; Richardson et al., 2006; Cancer Genome Atlas Network, 2012; Adib et al., 2004; Brune et al., 2008; Cho et al., 2011; Compagno et al., 2009; Hou et al., 2010; Murat et al., 2008; Győrffy et al., 2010, Győrffy et al., 2013; Szász et al., 2016; Vathipadiekal et al., 2015). Thus, our finding opens the possibility that specific inhibition of citrullination at R1810-RNAP2 may represent a suitable drug target.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, four tables, and three videos and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.10.016.

ACKNOWLEDGMENTS

We thank David Bentley for GST-CTD plasmids and Hiroshi Kimura for RNAP2 S2P/5P monoclonal antibodies. We thank CRG Genomics, Protein Technology, and the Advanced Light Microscopy facilities for all technical support. We are grateful to all the members of the chromatin and gene expression lab for useful suggestions. We acknowledge Juán Valcárcel, Guillermo P. Vicent, Enrique Vidal Ocabo, and Gwendal Dujardin from CRG for constructive criticism and advice during the course of this work. P.S. was supported by a Novartis fellowship and Beatriu de Pinós fellowship (co-founded by Marie Curie Action, 2013 BP_B 00061). Our work was supported by Spanish MEC (SAF2016-75006), the Catalan Government (2017-2019 SGR 747_MBeato), and the European Research Council Synergy Grant "4DGenome" (609989). We acknowledge the support of the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa," and the CERCA Programme/Generalitat de Catalunya.

AUTHOR CONTRIBUTIONS

Conceptualization, P.S. and M.B.; Methodology, P.S. and M.B.; Investigation, P.S., A.L., C.D.V., and F.L.D.; Formal Analysis, P.S., J.Q., J.C.-C., N.F.-F., and A.L.; Data Curation, P.S., J.Q., J.C.-C., N.F.-F., R.H.G.W., and B.O.; Visualization, P.S. and M.B.; Project Administration, P.S.; Writing – Original Draft, P.S. and M.B.; Writing – Review & Editing, P.S., M.B., D.E., A.L., N.F.-F., F.L.D., J.Q., and R.H.G.W.; Funding Acquisition, M.B. and P.S.; Resources, R.S. and D.E.; Supervision, P.S. and M.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 22, 2018
Revised: August 31, 2018
Accepted: October 9, 2018
Published: November 21, 2018

REFERENCES

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

Adib, T.R., Henderson, S., Perrett, C., Hewitt, D., Bourmpoulia, D., Ledermann, J., and Boshoff, C. (2004). Predicting biomarkers for ovarian cancer using gene-expression microarrays. Br. J. Cancer *90*, 686–692.

Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Rask, M.B., Streicher, W., Jungmichel, S., Nielsen, M.L., and Lukas, J. (2015). Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). Nat. Commun. *6*, 8088.

Arita, K., Shimizu, T., Hashimoto, H., Hidaka, Y., Yamada, M., and Sato, M. (2006). Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4. Proc. Natl. Acad. Sci. USA *103*, 5291–5296.

Baka, Z., György, B., Géher, P., Buzás, E.I., Falus, A., and Nagy, G. (2012). Citrullination under physiological and pathological conditions. Joint Bone Spine *79*, 431–436.

Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. Science *294*, 93–96.

Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B.N., Chung, H.J., Chan-Salis, K.Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., et al. (2016). RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. Cell *165*, 357–371.

Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. Curr. Opin. Cell Biol. *11*, 347–351.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brune, V., Tiacci, E., Pfeil, I., Döring, C., Eckerle, S., van Noesel, C.J., Klapper, W., Falini, B., von Heydebreck, A., Metzler, D., et al. (2008). Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. J. Exp. Med. *205*, 2251–2268.

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

Bywater, M.J., Pearson, R.B., McArthur, G.A., and Hannan, R.D. (2013). Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat. Rev. Cancer *13*, 299–314.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., Shiekhattar, R., and Shilatifard, A. (2015). PAFI, a molecular regulator of promoter-proximal pausing by RNA Polymerase II. Cell *162*, 1003–1015.

Cherrington, B.D., Zhang, X., McElwee, J.L., Morency, E., Anguish, L.J., and Coonrod, S.A. (2012). Potential role for PAD2 in gene regulation in breast cancer cells. PLoS ONE *7*, e41242.

Cho, J.Y., Lim, J.Y., Cheong, J.H., Park, Y.Y., Yoon, S.L., Kim, S.M., Kim, S.B., Kim, H., Hong, S.W., Park, Y.N., et al. (2011). Gene expression signature-based prognostic risk score in gastric cancer. Clin. Cancer Res. *17*, 1850–1857.

Christophorou, M.A., Castelo-Branco, G., Halley-Stott, R.P., Oliveira, C.S., Loos, R., Radzisheuskaya, A., Mowen, K.A., Bertone, P., Silva, J.C., Zernicka-Goetz, M., et al. (2014). Citrullination regulates pluripotency and histone H1 binding to chromatin. Nature *507*, 104–108.

Compagno, M., Lim, W.K., Grunn, A., Nandula, S.V., Brahmachary, M., Shen, Q., Bertoni, F., Ponzoni, M., Scandurra, M., Califano, A., et al. (2009). Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature *459*, 717–721.

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

Corden, J.L. (2016). Pol II CTD code light. Mol. Cell *61*, 183–184.

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.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J., and Kouzarides, T. (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545–553.

Day, D.S., Zhang, B., Stevens, S.M., Ferrari, F., Larschan, E.N., Park, P.J., and Pu, W.T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. Genome Biol. *17*, 120.

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, e11215.

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

Fernandez-Fuentes, N., Madrid-Aliste, C.J., Rai, B.K., Fajardo, J.E., and Fiser, A. (2007). M4T: a comparative protein structure modeling server. Nucleic Acids Res. *35*, W363–W368.

Fuhrmann, J., Clancy, K.W., and Thompson, P.R. (2015). Chemical biology of protein arginine modifications in epigenetic regulation. Chem. Rev. *115*, 5413–5461.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. eLife *6*, e29736.

Guo, W., Zheng, Y., Xu, B., Ma, F., Li, C., Zhang, X., Wang, Y., and Chang, X. (2017). Investigating the expression, effect and tumorigenic pathway of PADI2 in tumors. OncoTargets Ther. *10*, 1475–1485.

Győrffy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res. Treat. *123*, 725–731.

Gyorffy, B., Lánczky, A., and Szállási, Z. (2012). Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr. Relat. Cancer *19*, 197–208.

Győrffy, B., Surowiak, P., Budczies, J., and Lánczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS ONE *8*, e82241.

György, B., Tóth, E., Tarcsa, E., Falus, A., and Buzás, E.I. (2006). Citrullination: a posttranslational modification in health and disease. Int. J. Biochem. Cell Biol. *38*, 1662–1677.

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

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23.

Horibata, S., Rogers, K.E., Sadegh, D., Anguish, L.J., McElwee, J.L., Shah, P., Thompson, P.R., and Coonrod, S.A. (2017). Role of peptidylarginine deiminase 2 (PAD2) in mammary carcinoma cell migration. BMC Cancer *17*, 378.

Hou, J., Aerts, J., den Hamer, B., van Ijcken, W., den Bakker, M., Riegman, P., van der Leest, C., van der Spek, P., Foekens, J.A., Hoogsteden, H.C., et al. (2010). Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS ONE *5*, e10312.

Iannone, C., Pohl, A., Papasaikas, P., Soronellas, D., Vicent, G.P., Beato, M., and ValcárRcel, J. (2015). Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA *21*, 360–374.

Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P., and Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. Assay Drug Dev. Technol. *9*, 342–353.

Jeronimo, C., Collin, P., and Robert, F. (2016). The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. J. Mol. Biol. *428*, 2607–2622.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 167–177.

King, C.A., and Bradley, P. (2010). Structure-based prediction of protein-peptide specificity in Rosetta. Proteins *78*, 3437–3449.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. Mol. Cell *67*, 411–422.e4.

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.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283–291.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., et al. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. *487*, 545–574.

Ledet, M.M., Anderson, R., Harman, R., Muth, A., Thompson, P.R., Coonrod, S.A., and Van de Walle, G.R. (2018). BB-Cl-Amidine as a novel therapeutic for canine and feline mammary cancer via activation of the endoplasmic reticulum stress pathway. BMC Cancer *18*, 412.

Livingstone, C.D., and Barton, G.J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput. Appl. Biosci. *9*, 745–756.

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

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. *270*, 12335–12338.

Meininghaus, M., Chapman, R.D., Horndasch, M., and Eick, D. (2000). Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. J. Biol. Chem. *275*, 24375–24382.

Miller, T.E., Liau, B.B., Wallace, L.C., Morton, A.R., Xie, Q., Dixit, D., Factor, D.C., Kim, L.J.Y., Morrow, J.J., Wu, Q., et al. (2017). Transcription elongation factors represent *in vivo* cancer dependencies in glioblastoma. Nature *547*, 355–359.

Mohanan, S., Cherrington, B.D., Horibata, S., McElwee, J.L., Thompson, P.R., and Coonrod, S.A. (2012). Potential role of peptidylarginine deiminase enzymes and protein citrullination in cancer pathogenesis. Biochem. Res. Int. *2012*, 895343.

Murat, A., Migliavacca, E., Gorlia, T., Lambiv, W.L., Shay, T., Hamou, M.F., de Tribolet, N., Regli, L., Wick, W., Kouwenhoven, M.C., et al. (2008). Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. J. Clin. Oncol. *26*, 3015–3024.

Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nat. Protoc. *11*, 413–428.

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., and Huber, W. (2010). EBImage–an R package for image processing with applications to cellular phenotypes. Bioinformatics *26*, 979–981.

Pohl, A., and Beato, M. (2014). bwtool: a tool for bigWig files. Bioinformatics *30*, 1618–1619.

R Development Core Team (2017). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160–W165.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121–132.

Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. J. Mol. Biol. *428*, 2623–2635.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schüller, R., Forné, 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.

Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forné, I., Decker, T.M., Martin, D., Schüller, R., Krebs, S., et al. (2018). Tyrosine-1 of RNA polymerase II CTD controls global termination of gene trasncription in mammals. Mol. Cell *69*, 48–61.e6.

Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat. Genet. *49*, 1045–1051.

Sharma, P., Azebi, S., England, P., Christensen, T., Møller-Larsen, A., Petersen, T., Batsché, E., and Muchardt, C. (2012). Citrullination of histone H3 interferes with HP1-mediated transcriptional repression. PLoS Genet. *8*, e1002934.

Sims, R.J., 3rd, Rojas, L.A., Beck, D.B., Bonasio, R., Schüller, 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.

Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins *17*, 355–362.

Slade, D.J., Subramanian, V., Fuhrmann, J., and Thompson, P.R. (2014). Chemical and biological methods to detect post-translational modifications of arginine. Biopolymers *101*, 133–143.

Slade, D.J., Fang, P., Dreyton, C.J., Zhang, Y., Fuhrmann, J., Rempel, D., Bax, B.D., Coonrod, S.A., Lewis, H.D., Guo, M., et al. (2015). Protein arginine deiminase 2 binds calcium in an ordered fashion: implications for inhibitor design. ACS Chem. Biol. *10*, 1043–1053.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Szász, A.M., Lánczky, A., Nagy, Á., Förster, S., Hark, K., Green, J.E., Boussioutas, A., Busuttil, R., Szabó, A., and Győrffy, B. (2016). Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget *7*, 49322–49333.

Tanikawa, C., Espinosa, M., Suzuki, A., Masuda, K., Yamamoto, K., Tsuchiya, E., Ueda, K., Daigo, Y., Nakamura, Y., and Matsuda, K. (2012). Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. Nat. Commun. *3*, 676.

Tanikawa, C., Ueda, K., Suzuki, A., Iida, A., Nakamura, R., Atsuta, N., Tohnai, G., Sobue, G., Saichi, N., Momozawa, Y., et al. (2018). Citrullination of RGG motifs in FET proteins by PAD4 regulates protein aggregation and ALS susceptibility. Cell Rep. *22*, 1473–1483.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Truss, M., Bartsch, J., Schelbert, A., Haché, R.J., and Beato, M. (1995). Hormone induces binding of receptors and transcription factors to a rearranged nucleosome on the MMTV promoter in vivo. EMBO J. *14*, 1737–1751.

van Beers, J.J., Zendman, A.J., Raijmakers, R., Stammen-Vogelzangs, J., and Pruijn, G.J. (2013). Peptidylarginine deiminase expression and activity in PAD2 knock-out and PAD4-low mice. Biochimie *95*, 299–308.

van Venrooij, W.J., and Pruijn, G.J. (2000). Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. Arthritis Res. *2*, 249–251.

Vathipadiekal, V., Wang, V., Wei, W., Waldron, L., Drapkin, R., Gillette, M., Skates, S., and Birrer, M. (2015). Creation of a human secretome: a novel composite library of human secreted proteins: validation using ovarian cancer gene expression data and a virtual secretome array. Clin. Cancer Res. *21*, 4960–4969.

Vicent, G.P., Zaurin, R., Nacht, A.S., Li, A., Font-Mateu, J., Le Dily, F., Vermeulen, M., Mann, M., and Beato, M. (2009). Two chromatin remodeling activities cooperate during activation of hormone responsive promoters. PLoS Genet. *5*, e1000567.

Vicent, G.P., Nacht, A.S., Ballaré, C., Zaurin, R., Soronellas, D., and Beato, M. (2014). Progesterone receptor interaction with chromatin. Methods Mol. Biol. *1204*, 1–14.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, 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.

Vossenaar, E.R., Zendman, A.J., van Venrooij, W.J., and Pruijn, G.J. (2003). PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. BioEssays *25*, 1106–1118.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.H., Perlin, J.R., Leonelli, L., Sonbuchner, L.S., McDonald, C.H., Cook, R.G., Dou, Y., et al. (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. Science *306*, 279–283.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Witalison, E.E., Thompson, P.R., and Hofseth, L.J. (2015). Protein arginine deiminases and associated citrullination: physiological functions and diseases associated with dysregulation. Curr. Drug Targets *16*, 700–710.

Wright, R.H., Castellano, G., Bonet, J., Le Dily, F., Font-Mateu, J., Ballaré, C., Nacht, A.S., Soronellas, D., Oliva, B., and Beato, M. (2012). CDK2-dependent activation of PARP-1 is required for hormonal gene regulation in breast cancer cells. Genes Dev. *26*, 1972–1983.

Xu, S., Grullon, S., Ge, K., and Peng, W. (2014). Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. Methods Mol. Biol. *1150*, 97–111.

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

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. *39*, 1512–1516.

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

Zhang, X., Bolt, M., Guertin, M.J., Chen, W., Zhang, S., Cherrington, B.D., Slade, D.J., Dreyton, C.J., Subramanian, V., Bicker, K.L., et al. (2012). Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor α target gene activation. Proc. Natl. Acad. Sci. USA *109*, 13331–13336.

Zhang, Y., Zhou, L., Leng, Y., Dai, Y., Orlowski, R.Z., and Grant, S. (2017). Positive transcription elongation factor b (P-TEFb) is a therapeutic target in human multiple myeloma. Oncotarget *8*, 59476–59491.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J., et al. (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48–53.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-Cit1810 RNAP2-CTD	This study	N/A
Rabbit polyclonal anti-PADI2	Santa Cruz Biot.	sc-133877
Mouse monoclonal anti-PADI2	Sigma	WH0011240M1
Rabbit polyclonal anti-PADI2	Proteintech	12110-1-AP
Rabbit polyclonal anti-Citrulline	Millipore	AB5612
Rabbit polyclonal anti-CARM1	Millipore	09-818
Rabbit polyclonal anti-PRMT5	Millipore	07-405
Mouse monoclonal anti-TUBULIN	Sigma	T9026
Mouse monoclonal anti-GAPDH	Santa Cruz Biot.	sc-32233
Rabbit polyclonal anti-PADI4	Abcam	ab50247
Mouse monoclonal anti-PADI3	Santa Cruz Biot.	sc-393622
Rabbit polyclonal anti-HA	Abcam	ab9110
Rabbit polyclonal anti-RNAP2 (CTD4H8)	Millipore	05-623
Rabbit polyclonal anti-RNAP2 (N-20)	Santa Cruz	sc-899
Rabbit monoclonal anti-RNAP2 (D8L4Y)	Cell Signaling	14958
Rat monoclonal anti-Ser2P-RNAP2 (3E10)	Millipore	04-1571
Rabbit polyclonal anti-CDK9 (H-169)	Santa Cruz Biot.	SC-8338
Rabbit polyclonal anti-CCNT1	Bethyl Labs	A303-499A
IgG mouse	Millipore	12-371
IgG Rabbit	Cell Signaling	2729S
Mouse monoclonal anti-Ser2P RNAP2	Hiroshi Kimura, MBL Life science	CMA602
Mouse monoclonal anti-Ser5P RNAP2	Hiroshi Kimura, MBL Life science	CMA603
Chemicals, Peptides, and Recombinant Proteins		
Lipofectamine 3000	Invitrogen	L3000008
Cl¯amidine	Calbiochem	506282
α-amanitin	Sigma Aldrich	A2263
G418	Sigma Aldrich	000000004727878001
Doxycycline	Sigma Aldrich	D9891
SMARTpool On-target plus siRNAs for human PADI2	Dharmacon (Thermo Scientific)	M-019485-01
SMARTpool On-target plus siRNAs for human PADI3	Dharmacon (Thermo Scientific)	M-021051-01
CARM1 siRNA (h)	Santa Cruz Biot.	sc-44875
PRMT5 siRNA (h)	Santa Cruz Biot.	sc-41073
PADI4 siRNA (h)	Santa Cruz Biot.	sc-61283
Proteinase K	ThermoFisher Scientific	AM2546
TURBO DNase	ThermoFisher Scientific	AM2239
Propidium iodide	Molecular Probe	P-1304
DNase I (RNase-free)	ThermoFisher Scientific	AM2222
Cell proliferation ELISA BrdU Colorimetric assay	Roche	11647229001
Dynabeads MyOne Streptavidin T1	ThermoFisher	65601
Dynabeads M-280 Sheep Anti-Mouse IgG	ThermoFisher	11201D
Protein G Plus / Protein A Agarose	Millipore	IP05

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CTD-WT PSY**S$_2$**PS**S$_5$**P**R**YT$_2$PQ**S$_5$**PTYT$_2$**P	This study	N/A
CTD-Cit PSY**S$_2$**PS**S$_5$**P(**cit)**YT$_2$PQ**S$_5$**PTYT$_2$**P	This study	N/A
CTD-me2a PSY**S$_2$**PS**S$_5$**P**R(me2a)**YT$_2$PQ**S$_5$**PTYT$_2$**P	This study	N/A
CTD-S2P PSY**S$_2$(p)**PS**S$_5$**P**R**YT$_2$**(p)**PQ**S$_5$**PTYT$_2$**(p)**P	This study	N/A
CTD-S5P PSY**S$_2$**PS**S(p)$_5$**P**R**YT$_2$PQ**S$_5$(p)**PTYT$_2$**P	This study	N/A
CTD-WT (B) BIOTIN-PSY**S$_2$**PS**S$_5$**P**R**YT$_2$PQ**S$_5$**PTYT$_2$**P	This study	N/A
CTD-Cit ((B) BIOTIN-PSY**S$_2$**PS**S$_5$**P(**cit)**YT$_2$PQ**S$_5$**PTYT$_2$**P	This study	N/A
Critical Commercial Assays		
Anti-Citrullination detection kit	17-347	Millipore
Deposited Data		
ChIP-seq, RNA-seq, ChrRNA-seq data	This study	GEO: GSE105795
Additional data (this manuscript)	Mendeley Data	https://doi.org/10.17632/f3mdnsjbfg.3
Experimental Models: Cell Lines		
T47D cells	ATCC (Truss et al., 1995)	HTB-133
MCF7 cells	ATCC	HTB-22
Raji cells	ATCC	CCL-86
Oligonucleotides		
PADI2 D374K-GCTGGACTCTCCCCGAaAgGGAAA CCTAAAGGAC	This study	N/A
PADI2 S401A-GAGCCCCTCTTTGAGgCTGTCACCA GCCTTGAC	This study	N/A
PADI2 R580E-GATGGACGAGGACCACgaaGCCAGA GCCTTCTTCC	This study	N/A
PADI2 L642T-CTGCCTACCACAAATTTacGGGGGAA GTCCACTGTG	This study	N/A
Recombinant DNA		
pCoofy1	Addgene	43974
pET21aGST CTD (27-52)	Bentley, 1999	GST-C-CTD RNAP2
pET21aGST CTD (1-25.5)	Bentley, 1999	GST-N-CTD RNAP2
RBP1 gene cloned into LS* mock vector	Meininghaus et al., 2000	Wild-type (WT^r)
RBP1 gene (Arginine to alanine) cloned into LS* mock vector	Sims et al., 2011	R1810A^r Mutant
Human PADI2 wild type gene cloned in pCoofy1 vector	This study	WT-PADI2
D374K mutation in WT-PADI2	This study	D374K-PADI2
S401 mutation in WT-PADI2	This study	S401A-PADI2
L642T mutation in WT-PADI2	This study	L642T-PADI2
R580E mutation in WT-PADI2	This study	R580E-PADI2
Software and Algorithms		
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/
SICER	Xu et al., 2014	https://home.gwu.edu/~wpeng/Software.htm
deepTools2	Ramírez et al., 2016	https://github.com/deeptools/deepTools/wiki
Kallisto	Bray et al., 2016	https://pachterlab.github.io/kallisto/
Trimmomatic	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
ChIPpeakAnno	Zhu et al., 2010	http://www.bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html
Bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
FIJI	Schindelin et al., 2012	https://fiji.sc

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R	R Development Core Team (2017)	https://www.R-project.org/
GSEA	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
Kaplan Meier Survival	Győrffy et al., 2010; Gyorffy et al., 2012; Győrffy et al., 2013; Szász et al., 2016; Vathipadiekal et al., 2015	http://kmplot.com/analysis/index.php?p=service
JalView	Waterhouse et al., 2009; Livingstone and Barton, 1993	http://www.jalview.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further correspondence and requests for reagents should be directed to and fulfilled by the Lead Contact Miguel Beato (miguel.beato@crg.eu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

T47D, MCF, and Raji cells were grown routinely in their optimal medium according the ATCC recommendations. T47D cell line carrying a single copy of luciferase reporter gene driven by MMTV promoter (T47D-MTVL, Truss et al., 1995) were grown in RPMI-1640 medium without phenol red supplemented with 10% dextran coated charcoal treated FBS (DCC/FBS), 2 mM L-glutamine, 100 U/mL penicillin-streptomycin as reported previously (Vicent et al., 2009; Wright et al., 2012). MCF7 cells were grown in DMEM (without phenol red) with 10% DCC/FBS and 100 U/mL penicillin-streptomycin. All cell transfections were carried out using Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. Cells were treated with 200 μM citrulline inhibitor Cl^{-}amidine (506282, Calbiochem) or vehicle (DMSO) for 2 hr. T47D cells ($2.5x10^6$ cells per 8 μg plasmid) were transiently transfected with HA tagged α-amanitin resistant wild-type (WT^r) or R1810A^r RNAP2 plasmids for 24 hr followed by α-amanitin (A2263, Sigma Aldrich, 6 μg/mL) treatment for 12 hr to inhibit endogenous RNAP2.

Raji cells with stable expression of HA tagged WT^r or R1810A^r RNAP2 plasmids DNA were generated by electroporation (condition 250V and 950 μF; $2x10^7$ cells per 10 μg plasmid DNA) followed by selection with G418 (000000004727878001, Sigma Aldrich,1 mg/mL) and Doxycycline (D9891, Sigma Aldrich, 0.1 μg/mL). To induce the expression of recombinant RNAP2 cells were grown in tetracycline-free complete medium for 48 hr prior to doxycycline (D9891, Sigma Aldrich) addition at a concentration of 0.1 μg/mL. Endogenous RNAP2 was subsequently inhibited using α-amanitin (2 μg/mL) for 24 hr before downstream assays.

METHOD DETAILS

Experimental procedures

Generation of anti-Cit1810 RNAP2-CTD Antibody

The citrulline specific antibody was raised in rabbits by Eurogentec using a KLH coupled CTD peptide sequence (YSPSSP-cit-YTPQSP). Affinity purification was performed first on a column containing the citrullinated peptide, followed by removal of non-citrullinated specific antibodies on a column containing the non-citrullinated peptide.

Commercial antibodies used in this study were as follows: anti-PADI2 for ChIP and Immunoprecipitation assays: (sc-133877 lot no # E1214 and H0715, Santa Cruz Biot.); anti-PADI2 for western blots and Immunofluorescence (12110-1-AP from Proteintech and WH0011240M1 from Sigma); anti-citrulline (AB5612, Millipore), anti-citrulline detection kit (17-347, Millipore,), anti-CARM1 (09-818, Millipore), anti-PRMT5 (07-405, Millipore), anti-TUBULIN (T9026, Sigma), anti-GAPDH (sc-32233, Santa Cruz Biot.), anti-PADI4 (ab50247, Abcam), anti-PADI3 (sc-393622, Santa Cruz Biot), anti-HA (ab9110, Abcam), anti-RNAP2 for ChIP (CTD4H8,05-623, Millipore; Rpb1 NTD (D8L4Y), 14958, Cell Signaling), for western blot (N-20, sc-899, Santa Cruz Biot.), anti-phospho-S2 RNAP2 CTD (3E10, 04-1571 from Millipore), anti-CDK9 (H-169, SC-8338,lot no # C0415), anti-CCNT1(A303-499A,Bethyl Labs), IgG negative control for ChIP and immunoprecipitation assays (12-371, Millipore; 2729S Cell Signaling). Anti-S2P-RNAP2 (CMA602, MBL Life science) and anti-S5P-RNAP2 (CMA603, MBL Life science) were kindly provided by Hiroshi Kimura's laboratory.

Peptides

All peptides were synthesized and purified by Eurogentec. The CTD peptide PSY**S_2**PS**S_5**P**R**YT$_2$PQ**S_5**PTYT$_2$P was used for dot blots, and microscale thermophoresis (MST) experiments, either unmodified (CTD-WT) or with Cit1810, R1810me2a, S2-P (1805,1812,1819) or S5-P (1808, 1815) modifications. The N-terminal biotin labeled CTD peptides (CTD-WT with R1810) PSY**S_2**PS**S_5**P**R**YT$_2$PQ**S_5**PTYT$_2$P

and (CTD-cit1810) PSY**S**$_2$PS**S**$_5$P**cit**YT$_2$PQ**S**$_5$PTYT$_2$P were used for pull down assay. Peptides were quantified by amino acid analysis, and the presence of the modifications was confirmed by mass spectrometry.

Plasmids

The α-amanitin resistant HA-tagged wild-type (WT^r) or R1810A^r mutant RNAP2 plasmids were previously published (Meininghaus et al., 2000; Sims et al., 2011). The GST-N-CTD (repeats 1-25.5), the GST-C-CTD (repeats 27-52) of RNAP2 were kindly provided by David Bentley (Bentley, 1999).

RNA interference experiments

For siRNAs inhibition experiments T47D or MCF7 cells were transfected with 100 μM siRNA using Lipofectamine 3000 (Invitrogen) for 72 hr according to manufacturer's instructions. SMARTpool On-target plus siRNAs for PADI2 (M-019485-01) and PADI3 (M-021051-01) from Dharmacon (Thermo Scientific). siRNAs for CARM1 (sc-44875), PRMT5 (sc-41073) and PADI4 (sc-61283) from Santa Cruz Biot.

Chromatin RNA extraction

Chromatin RNA from T47D cells transfected with either siCTRL or si*PADI2* was prepared as described previously (Nojima et al., 2016). Briefly, T47D cells transfected with si*CTRL*or si*PADI2* were lysed for 5 min in 4 mL of ice-cold HLB+N [10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$ and 0.5% (vol/vol) NP-40], followed by addition of 1 mL of ice cold HB+NS buffer [10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% vol/vol, NP-40 and 10% wt/vol sucrose]. Cell nuclei were collected by centrifugation at 1,400 rpm for 5 min at 4°C, resuspended in 125 μL of NUN1 buffer [20 mM Tris-HCl pH 7.9, 75m M NaCl, 0.5 mM EDTA and 50% vol/vol glycerol]. After addition of 1.2 mL of ice cold NUN2 buffer [20 mM HEPES-KOH pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 7.5 mM $MgCl_2$, 1% vol/vol NP-40 and 1 M urea], samples were incubated on ice for 15 min, mixing by vortexing every 3 min. The chromatin pellet was resuspended in HSB buffer [10 mM Tris pH 7.5, 500 mM NaCl and 10 mM $MgCl_2$] and treated with TURBO DNase (AM2239, Thermo Scientific) at 37°C for 15 min, followed by Proteinase K (AM2546, Thermo Scientific) for 10 min at 37°C. Chromatin-RNA was purified by Trizol (Thermo Scientific), quantified with a Qubit 3.0 Fluorometer (Life Technologies).

RNA extraction and RT-qPCRs

RNA from T47D cells transfected with either siCTRL or si*PADI2* and expressing α-amanitin resistant HA-tagged wild-type (WT^r) or R1810A^r mutant RNAP2 was extracted using RNeasy (QIAGEN) according to manufacturer's instructions. 1 μg of purified RNA was used for DNase treatment (Thermo Scientific), quantified with a Qubit 3.0 Fluorometer (Life Technologies).

Reverse transcription was performed for chromatin RNA with random hexamers, for RNA with oligo (dT) using SuperScript III (Invitrogen) according to manufacturer's instructions. Complementary DNA was quantified by qPCR using Roche Lightcycler (Roche), as previously described (Vicent et al., 2009). For each gene product, relative RNA abundance was calculated using the standard curve method and expressed as relative RNA abundance after normalizing against the human *GAPDH* gene level. All the gene expression data generated by RT-qPCR represent the average and SEM of at least 3 biological replicates. Primers used for RT-qPCR are listed in Table S4.

PolyA RNA-seq

Purified RNA was analyzed on Bioanalyzer using an RNA Pico assay chip. PolyA plus RNA mRNA libraries were prepared using TruSeq Stranded RNA Library Prep Kit (Illumina) and sequenced using Illumina HiSeq 2500.

Chromatin RNA-seq

Chromatin associated RNA was prepared as mentioned above (Nojima et al., 2016). Before preparing chromatin RNA libraries, contaminant of rRNAs was depleted using Ribo-Zero rRNA removal kit. Libraries were prepared using TruSeq Stranded small RNA Library Prep Kit (Illumina) and sequenced using Illumina HiSeq 2500.

Protein extract preparation, Co-immunoprecipitation (IP), Peptides pull down, and western blots

Cells were prepared as described previously (Wright et al., 2012). Briefly, 5 × 10^6 to 10^7 cells were lysed on ice for 30 min in lysis buffer (1% Triton X-100 in 50 mM Tris pH 7.4-7.6, 130 mM NaCl) containing proteases inhibitors (11836170001, Roche) with rotation, followed by sonication for 7 min with every 30 s on and 30 s off. After centrifugation at 4°C and 13,000 rpm for 10 min, extracts were used for protein quantitation. For IP 2 mg of extract were incubated for 12 hr with protein G/A agarose beads (for rabbit antibodies, IP05, Millipore) or Dynabeads M-280 Sheep Anti-Mouse IgG (for mouse antibodies, 11201D, Thermo Scientific), previously coupled with 5-7 μg of the corresponding antibodies or a control IgGs. For RNAP2-S2P and -S5P 7 μg of each mouse monoclonal antibodies (CMA602 or CMA603, respectively) were coupled with Dynabeads, followed by 12 hr incubation with extract at 4°C. The samples were washed 6 times with lysis buffer and boiled for 5 min in SDS gel sample buffer. For detection of mentioned proteins of molecular weight (< 200 kDa) 4%–12% SDS-PAGE gels were used; while for the RNAP2 large subunits (> 200 kDa), we used 3%–8% SDS-PAGE.

For peptide pull down assay, 100 μg of each CTD-WT (with R1810) and CTD-cit1810 biotin labeled peptides were bound to 100 μL of the Dynabeads MyOne Streptavidin T1 (65601, Thermo Scientific) in 1 mL of binding buffer [150 mM NaCl, 50 mM Tris pH 8, 1% IGEPAL CA-630] by rotation at room temperature for 2 hr. The peptide bounds Dynabeads were incubated with 300 μg of T47D cells nuclear extract for 12 hr at 4°C along with rotation. Peptide bound protein complexes were then washed five times with wash buffer

[400 mM NaCl, 50 mM Tris pH 8, 1% IGEPAL CA-630] followed by two washes with binding buffer. All buffers were supplemented with freshly prepared protease inhibitors. Samples were eluted by incubation at 70°C for 5 min in SDS gel sample buffer, followed by protein detection of mentioned proteins by western blots.

For western blots primary antibodies were used at 1:250 to 1:1000 dilution (for α−cit1810; 1:50) and incubated overnight at 4°C, followed by 1 hr incubation with horseradish peroxidase conjugated anti-mouse (NA931V) or anti-rabbit (NA934V, Amersham) and blots were developed using ECL prime western blotting detection reagent (RPN2232, GE Healthcare) according to the manufacturer instructions.

Size exclusion chromatography

The size exclusion chromatography of T47D cell extracts were carried out using Superdex 200 10/300 mm columns (17517501, GE healthcare). As per manufacturer's instructions, for high molecular weight (Ferritin, 440 KDa) and for low molecular weight (Conalbumin,75 KDa and Carbonic anhydrase 29KDa) were run along with cell extracts. Samples were chosen according to chromatography profile and used for western blots.

BrdU (5′-bromo-2'-deoxyuridine) cell proliferation assay

T47D cells (1×10^4) were plated in a 96-well plate followed by transfection with control/PADI2 siRNAs or treated with the PADI inhibitor Cl^-amidine at concentration of 200 μM or DMSO or expressing α-amanitin resistant HA-tagged wild-type (WT^r) or $R1810A^r$ mutant form of RNAP2.The cell proliferation ELISA BrdU Colorimetric assay (Roche,11647229001) was performed as per manufacturer's instructions. The experiments were performed at least four biological replicates.

Fluorescence-activated cell sorting (FACS) experiments

FACS assay was performed in T47D cells transfected with control or PADI2 siRNAs from three biological replicates. Briefly, cells were trypsinized, washed three times with 1x PBS and fixed with cold absolute ethanol in suspension at 70% final concentration. Cells were stained with propidium iodide (P-1304, Molecular Probe) followed by DNase I (RNase-free) (AM2222, Thermo Scientific) treatment and stored for 24 hr at 4°C and DNA contents of cell cycle phases were analyzed using a BD LSR II flow cytometer.

In vitro citrullination assay with recombinant PADI2 wild-type and mutants (D374K, S401A, R580E, and L642T) and fragments of the RNAP2-CTD

The PADI2 open reading frame (ORF) was cloned into the HIS-tagged expression vector pCoofy1 (Addgene, 43974) and the wild-type (WT) plasmid sequence were verified. All four PADI2 mutants were generated by using WT PADI2 plasmid by performing site-directed mutagenesis as directed in Quick change mutagenesis kit and mutations were confirmed by sequencing. The mutagenic primers given in Key Resources Table.

Recombinant proteins were expressed in bacteria strain BL21 pRARE and purified following the standard method of histidine-tagged recombinant protein. Briefly, cells were lysed in Buffer A (50 mM Tris HCl pH 7.4, 500 mM NaCl, 10% glycerol, 2 mM DTT, 20 mM Imidazole, 1% and Triton X-100) complemented with proteinase inhibitors (11836170001, Roche). Purification was performed using the HiTrap TALON crude (28953766, GE Healthcare) according to manufacturer's instruction. Proteins eluted in buffer containing 50 mM Tris-HCl pH 7.4, 300 mM NaCl and 10% glycerol, were stored at −80°C until required. The GST-N-CTD (repeats 1-25.5), GST-C-CTD (repeats 27-52) of RNAP2 were expressed and purified following the standard glutathione bead purification (Bentley, 1999). *In vitro* citrullination was carried with recombinant His-PADI2 in deimination buffer (50 mM HEPES pH 7.5, 10 mM $CaCl_2$, 4 mM DTT) at 37°C for 1 hr. Samples were dissolved in sample Laemmli buffer for immunoblot analysis using an anti-citrulline antibody (Christophorou et al., 2014; Wang et al., 2004), Millipore,17-347).

Microscale Thermophoresis (MST) of recombinant PADI2 with RNAP2-CTD peptides

Wild-type peptide or peptides carrying modifications of R1810me2a CTD, S2P-CTD and S5P CTD (20 nM to 500 μM) were titrated against a fixed concentration of fluorescent recombinant His-PADI2 (50 nM). MST data were acquired at 20°C using the red LED at 20% and IR- Laser at 40% using a (Monolith NT.115, Nano Temper Technologies) according to manufacturer's instructions. The results are plotted as normalized fluorescence (Fnorm, representing binding affinity) against the concentration of the unlabeled ligand and fitted according to the law of mass action.

Immunofluorescence, image acquisition and analysis

T47D cells were grown on round 10 mm glass coverslips transfected with sicontrol or si*PADI2* prior to fixation with 4% paraformaldehyde in PBS for 5 min and permeabilized with PBS 0.1% Triton X-100 (PBST) at room temperature for 5 min. Coverslips were blocked with IF buffer (5% BSA, 0.1%Triton X-100 in PBS) for 20 min at room temperature and incubated overnight with primary antibodies diluted in IF buffer at 1:50 of α-Cit1810 (Rabbit), 1:50 of α-CTD4H8 (mouse, Santa Cruz Biot.) to detect the total RNAP2, 1:250 of α-S2P-RNAP2 (mouse, CMA602) and 1:250 of α-S5P-RNAP2 (mouse, CMA603). For triple staining 1:500 of α-PADI2 (Mouse, WH0011240M1, Sigma) was used with 1:50 of α-Cit1810 along with α-S2P-RNAP2 (3E10; Rat, Millipore; Figure S2I). For Figure S1G, α-HA (1:250, ab9910; Rabbit) and α-RNAP2 (1:50, CTD4H8; mouse, Santa Cruz Biot.). After 3x washes with PBST (1X PBS with Triton X-100 0.1%) samples were incubated with secondary antibodies at a dilution 1:500 (Alexa Fluor 594 anti-rabbit, Alexa Fluor 488 anti-mouse or Alexa Fluor 680 anti-mouse and Alexa Fluor 488 anti-rat, Invitrogen-Molecular Probes) for 1 hr at room temperature followed by three washes with PBST. Samples were mounted with Mowiol mounting medium. For quantification, DAPI (4 ', 6-Diamidino-2-Phenylindole, Dihydrochloride) or Hoechst fluorescent stains were used as reference. Confocal images of T47D cells (Figures S1G and S2B) were acquired with a Leica SP5 (DMI 6000) inverted microscope using an HCX PLAN APO λ blue 63x/1.4-0.6 Oil immersion lens, PMT detectors and diode and Argon lasers. Laser and spectral detection bands were chosen for the optimal imaging of Alexa 488, Alexa 594 and DAPI to obtain Z stacks of nuclear optical sections at a distance of 0.42 μm.

Super-resolution images were acquired with a Leica SP8 STED 3X microscope using a HC PL APO CS2 100x/1.4 Oil immersion lens, a pulsed supercontinuum light source (white light laser) and HyD detectors, using the Leica acquisition software. Laser and spectral detection bands were chosen for the optimal imaging of nuclear optical sections with a z-distance of 0.16 μm. Deconvolution was performed using the Huygens deconvolution software (Scientific Volume Imaging, SVI) for STED modes using shift correction to account for drift during stack acquisition. Adjustments of individual channels were applied to the whole image, pseudo-coloring, cropping, and different channel composite images were done with FIJI (https://fiji.sc, Schindelin et al., 2012). Raw cropped images were used to calculate the correlation of the fluorescent signal with R version 3.4.1 (https://www.R-project.org/; R Development Core Team, 2017). All images were imported to the R with the package EBImage (Pau et al., 2010).

ChIP-qPCRs

For ChIP assays (Vicent et al., 2014), 10 × 10^6 of T47D cells, transfected with either si*CTRL* or si*PADI2* and expressing α-amanitin resistant HA-tagged wild-type (WT^r) or $R1810A^r$ mutant RNAP2 were cross-linked for 10 min with 1% formaldehyde. The lysate was sonicated to a DNA fragment size range of 200-300 bp using a Biorupter sonicator (Diagenode). PADI2 was immunoprecipitated with 6 μg of PADI2 antibody (z-22):sc-133877 lot number E1214 and H0715 using 50 μg of chromatin and 42 μL of Protein A-Agarose Beads (Diagenode). For RNAP2, 150 μg chromatin was incubated with 20 μg of antibody (Rpb1 NTD (D8L4Y), 14958, Cell Signaling; CTD4H8, Millipore), 15 μg of anti-HA (ab9110) for HA-tagged wild-type (WT^r) or $R1810A^r$ mutant form of RNAP2, 15 μg of anti-CDK9 (H-169, SC-8338, lot no # C0415), 20 μg of S2P; CMA602 and S5P; CMA603) or control IgG in IP Buffer with 2X SDS buffer (100 mM NaCl, 50 mM TrisHCl, pH 8, 5 mM EDTA and 0.5% SDS) and 1X Triton buffer (100 mM Tris-HCl, pH 8.8, 100 mM NaCl, 5 mM EDTA and 5% Triton-X) with protease inhibitors (11836170001, Roche) for 16 hr at 4°C. Followed by incubating with Protein A Sepharose beads CL-4B (17-0780-01, GE Healthcare) or 50 μL of Dynabeads M-280 sheep anti-mouse IgG (11201D, Thermo Scientific) for 3 hr. Beads were washed 3 times with low salt buffer (140 mM NaCl, 50 mM HEPES, pH 7.4, 1% Triton X-100), 2 times with high salt buffer (500 mM NaCl, 50 mM HEPES, pH 7.4, 1% Triton X-100) followed by single wash of LiCl Buffer (10 mM Tris HCl pH 8.0, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholic acid and 1 mM EDTA) and 1X TE buffer in cold room. Subsequently, crosslinks were reversed at 65°C overnight, followed by RNAase treatment for 1.5 hr and bound DNA was purified by Phenol-Chloroform extraction. The resultant eluted DNA was quantified by Qubit 3.0 Fluorometer (Life technologies), and followed by real-time qPCR analysis and data represented as fold change over input fraction from at least 3 biological replicate experiments. Primers used for RT-qPCR are listed in Table S4.

ChIP-seq

PADI2 and RNAP2 ChIP-purified DNA were analyzed on Bioanalyzer using DNA1000 kit. At least 1 ng of purified DNA were used to prepare ChIP-seq libraries using Illumina ChIP Sample Library Preparation Kit. End repaired and adaptor ligated libraries samples were size selected using E-Gel Size Select 2% Agarose Gel (Thermo Scientific, USA) followed by 13 cycles of PCR amplification. Barcode libraries from several samples were pooled and sequenced using Illumina HiSeq 2000 in single end sequencing run to obtain ~80–100 million reads.

Bioinformatic Procedures

RNA-seq Data processing

Adaptor sequences were remove from raw paired-end reads PolyA plus mRNA and ChrRNA raw paired-end reads by using Trimmomatic (Bolger et al., 2014) with the parameter values single-end mode, seed mismatches = 2, palindrome clip = 30, simple clip threshold = 12, min adaptor length = 1, keep both reads, leading = 3, trailing = 3, target length = 0, strictness = 999 and min length = 36. Transcript-level quantification was performed by Kallisto (Bray et al., 2016) by using 100 bootstraps. Spearman pairwise correlation (R2) between the two biological replicates (E1 and E2) were calculated, for the mRNA-sequencing si*CTRL* (R^2 = 0.96), si*PADI2* (R^2 = 0.97), WT^r (R^2 = 0.98) and $R1810A^r$ (R^2 = 0.97) and for the ChrRNA-sequencing experiments si*CTRL* (R^2 = 0.95) and si*PADI2* (R^2 = 0.93).

Gene Expression Analysis

Differential expression analysis was performed using DESeq2 Bioconductor package (Love et al., 2014). The analysis was performed by using 196,520 number of annotated transcripts of hg19 (correspond to 57,280 number of genes). Out of this, we quantitated data for total 18,241 (mRNA seq) and 33,140 (ChrRNA seq) genes. Genes with FC < 1/1.50 and adjusted p value < 0.01 were considered as downregulated and genes with FC > 1.50 and adjusted p value < 0.01 as upregulated (Figures 3A and 3C).

ChIP-seq Data processing

For ChIP-seq datasets, first we performed the quality control analysis using fastQC tool with version GPLV3 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Single-ended sequences were trimmed to 50 bp and mapped to the human genome assembly hg19 using Bowtie (Langmead et al., 2009), keeping only tags that mapped uniquely and with no more than two mismatches. Fragment sizes were estimated using HOMER tools (Heinz et al., 2010). Normalized coverage profiles for regions of interest were obtained by using deepTools2 (Ramírez et al., 2016) with a 100bp bin size.

RNAP2 and CDK9 ChIP-sequences were analyzed as mentioned previously (Iannone et al., 2015) by using MACS2 (Zhang et al., 2008) with a FDR q-value < 0.05. We noted that it typically results in regions with a very modest enrichment. Therefore, by default, we applied a more stringent threshold of FDR q-value < 0.0001 and a 4-fold enrichment over the control sample to base the subsequent downstream analyses on a high-confidence set of peaks. Because of broad peak of PADI2, significant enrichment to background

compared to Input DNA were identified using SICER (Xu et al., 2014) with the following parameters: window size 200; fragment size 233 bp; gap size 600; and FDR 0.01. Genome-wide 10kb Spearman pairwise correlation of PADI2 ChIP-sequencing signal was $R^2 = 0.98$. We then applied ChIPpeakAnno R library (Zhu et al., 2010) to create a merged list of peaks present in both replicates. ChIP-seq RPM normalized profiles were used to generate average profiles over different genomic features using bwtool (Pohl and Beato, 2014) deepTools2 (Ramírez et al., 2016) with a 100 bp bin size. Similarly, ChIP-seq profiles from HA-tagged WT^r and $R1810A^r$ were obtained by deepTools2 with 100 bp bin size and their difference ($R1810A^r$-WT^r) was plotted by using in house custom R version 3.4.1 (R Development Core Team, 2017) scripts on the indicates set of transcripts.

Pausing Index (PI) of RNAP2 Analysis

RNAP2 pausing index was calculated as mentioned previously (Baranello et al., 2016; Zeitlinger et al., 2007; Chen et al., 2015). Briefly, RNAP2 pausing index represents the dynamics of RNAP2 assembly and promoter release and hence not only presence or absence of transcription. We calculated pausing index (PI) as the ratio of the RNAP2 read count 1kb (kilobase) flanking to the TSS divided by size and read count in the same gene body divided by the size of the gene body.

Gene Ontology (GO) analysis

Go Annotation was performed using the online tool GSEA (Subramanian et al., 2005) (http://software.broadinstitute.org/gsea/index.jsp Gene Set Enrichment Analysis,) collection database v5. The significant cut off p value and FDR q-value < 0.05. All 315 genes documented with the parent cell cycle GO:0007049 (http://software.broadinstitute.org/gsea/msigdb/cards/CELL_CYCLE_GO_0007049) were considered for analysis. Among them, 282 genes expressed in T47D cells; 101 genes classified as PADI2 dependent (si*CTRL*/si*PADI2* FC < 1/1.5, p value < 0.01) and rest 181 genes as PADI2 independent (Figure 3H and Table S1).

Structural modeling of PADI2 and PADI3 with R1810 RNAP2-CTD complexes and conservation analysis

In the case of PADI2 (UniProt ID: Q9Y2J8), there are several crystallographic structures available in apo form in the PDB databank (Berman et al., 2000) and the one with PDB: 4N2I (Slade et al., 2015) was used to model the PADI2-R1810 RNAP2-CTD complex. The structure of the apo-form of PADI3 (UniProt ID: Q9ULW8) was derived by homology modeling using M4T (Fernandez-Fuentes et al., 2007). PADI2 (PDB code: 4N2G (Slade et al., 2015) was used as template sharing over 51% sequence identity with (coverage above 95%), hence optimal for homology modeling (Baker and Sali, 2001). The quality of the PADI3 model was assessed using ProSa2 (Sippl, 1993) and PROCHECK (Laskowski et al., 1993).

The structural models of PADI2 and PADI3 bound to RNAP2 CTD peptide encompassing R1810 (sequence: YSPSSPRYTPQSPT) were derived using the pepspec application (King and Bradley, 2010) in the Rosetta Suite (Leaver-Fay et al., 2011) as follow. The structure of PADI4 in complex with histone H3 N-terminal tail (PDB code:2DEW, Arita et al., 2006) was used to position the central arginine of R1810 CTD RNAP2 peptides, i.e., anchor residue, both in PADI2 and PADI3 structures. 100 models were generated, selecting the one with the best score for the second part of the modeling process: peptide extension. The central anchor residue, i.e., Arginine, was extended both in Nt and Ct directions by adding one residue at a time to match the sequence of the RNAP2 CTD peptide. In total 5000 models were generated for both PADI2 and PADI3 bound to RNAP2 CTD peptide bearing R1810 residue and ranked according to Rosetta score (Leaver-Fay et al., 2011).

The analysis of sequence conservation was done as follow. The sequences of PADI1 (UniProt ID: Q9ULC6), PADI2 (UniProt ID: Q9Y2J8), PADI3 (UniProt ID: W9ULW8), PADI4 (UniProt ID: Q9UM07) and PADI6 (UniProt ID: Q6TGC4) were aligned using ClustalW (Thompson et al., 1994). From the multiple sequence alignment, a conservation score, ranging from 0 to 10, was computed for each residue using JalView (Waterhouse et al., 2009; Livingstone and Barton, 1993).

QUANTIFICATION AND STATISTICAL ANALYSIS

For super-resolution image analysis in Figures 1D and S2H, Pearson correlation was calculated for the fluorescent signal of the two channels of interest for each individual stack. Statistical analysis (two-tailed Student's t test) for the average z stack Pearson's correlation of each individual cell. For all experiments of RT-qPCR, ChIP-qPCR and cell proliferation, a Two-tailed unpaired Student's t test was used to determine statistical significance between the groups. Plots and indicated statistical analysis were done with the use of Prism (GraphPad Prism 6.0 for MacOS), unless otherwise stated. Correction between biological replicated of RNA and ChIP-sequencing was calculated by Spearman's correlation (R2). For all other experiments, significance between groups calculated by Wilcoxon-Mann-Whitney test. If exact p value are not shown or indicated in legend then p values are represented in all figures as follows: *, p value $\leq$ 0.05; **, p value $\leq$ 0.01; ***, p value $\leq$ 0.001; °, p value > 0.05.

DATA AND SOFTWARE AVAILABILITY

The accession number for the high throughput sequencing data reported in this paper is GEO: GSE105795. Data reported here was deposited here: https://doi.org/10.17632/f3mdnsjbfg.3.