Phospho-site mutants of the RNA Polymerase II C-terminal domain alter subtelomeric gene expression and chromatin modification state in fission yeast

Maki Inada[1],[*], Robert J. Nichols[1], Jahan-Yar Parsa[2], Christina M. Homer[2], Ruby A. Benn[1], Reyal S. Hoxie[1], Hiten D. Madhani[2], Stewart Shuman[3], Beate Schwer[4] and Jeffrey A. Pleiss[5],[*]

[1]Biology Department, Ithaca College, Ithaca, NY 14850, USA, [2]Department of Biochemistry and Biophysics, UCSF, San Francisco, CA 94158, USA, [3]Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10065, USA, [4]Department of Microbiology, Weill Cornell Medical College, New York, NY 10065, USA and [5]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

ABSTRACT

Eukaryotic gene expression requires that RNA Polymerase II (RNAP II) gain access to DNA in the context of chromatin. The C-terminal domain (CTD) of RNAP II recruits chromatin modifying enzymes to promoters, allowing for transcription initiation or repression. Specific CTD phosphorylation marks facilitate recruitment of chromatin modifiers, transcriptional regulators, and RNA processing factors during the transcription cycle. However, the readable code for recruiting such factors is still not fully defined and how CTD modifications affect related families of genes or regional gene expression is not well understood. Here, we examine the effects of manipulating the $Y_1S_2P_3T_4S_5P_6S_7$ heptapeptide repeat of the CTD of RNAP II in *Schizosaccharomyces pombe* by substituting non-phosphorylatable alanines for Ser2 and/or Ser7 and the phosphomimetic glutamic acid for Ser7. Global gene expression analyses were conducted using splicing-sensitive microarrays and validated via RT-qPCR. The CTD mutations did not affect pre-mRNA splicing or snRNA levels. Rather, the data revealed upregulation of subtelomeric genes and alteration of the repressive histone H3 lysine 9 methylation (H3K9me) landscape. The data further indicate that H3K9me and expression status are not fully correlated, suggestive of CTD-dependent subtelomeric repression mechansims that act independently of H3K9me levels.

INTRODUCTION

The C-terminal domain (CTD) of the Rpb1 subunit of RNA Polymerase II (RNAP II) plays key roles in regulating gene expression (1). In humans and mice, the CTD consists of 52 heptapeptide repeats of the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$, or variants thereof. The CTDs of the fission yeast *Schizosaccharomyces pombe* and the budding yeast *Saccharomyces cerevisiae* comprise 29 and 26 heptad repeats, respectively, suggesting that the length of the CTD repeat array may be correlated with complexity of the organism (2). The CTD heptads are dynamically modified during the transcription cycle (3). The phosphorylation and dephosphorylation of serines 2, 5 and 7, threonine 4, and tyrosine 1, along with *cis-trans* isomerization of prolines 3 and 6, allow for a vast combinatorial complexity of CTD primary structures that inscribe a readable 'code' to orchestrate recruitment, activity, and egress of numerous cellular factors that govern gene expression (2).

Different phosphorylation or modification states of the CTD have been shown to play important roles in coordinating the coupling of RNA processing pathways. For example, RNAP II recruits to its CTD histone methyltransferases that generate the relaxed chromatin necessary for transcription initiation (2,3). Phosphorylation of Ser5 recruits enzymes that cap the 5′ end of nascent mRNAs. Indeed, in fission yeast, fusion of the capping enzymes RNA triphosphatase and RNA guanylyltransferase to RNAP II can rescue the lethality of mutants unable to be phosphorylated at Ser5, indicating that the chief essential function of Ser5 phosphorylation is the recruitment of the capping machinery (4). Ser2 and Ser5 phosphorylations are important in directing co-transcriptional methylation of lysines 4 and 36 of the histone H3 (5).

*To whom correspondence should be addressed. Tel: +1 607 254 8562; Fax: +1 607 254 8562; Email: jpleiss@cornell.edu
Correspondence may also be addressed to Maki Inada. Email: minada@ithaca.edu

During transcription elongation, Ser2 phosphorylation levels increase, and these have been shown to be important for recruitment of splicing factors such as Prp40, U2AF65 and U2 snRNA (6,7). Others have followed this finding by demonstrating splicing defects of specific reporter introns in human U2OS cells and chicken DT40 cells containing *S2A* mutations (8,9). Whereas global effects of the *S2A* mutation on gene expression have been examined previously (10–12), the methods used were not optimized to examine splicing, leaving unknown the full extent of the genome-wide impact of Ser2 phosphorylation on splicing. Here we used splicing-sensitive microarrays to examine pre-mRNA splicing in *S. pombe* cells mutated at CTD phosphorylation sites Ser2 and/or Ser7. Our results weigh against a role for CTD Ser2/7 phosphorylation in pre-mRNA splicing, but reveal an unexpected relationship with chromatin modification and gene expression in the subtelomeric regions of the genome.

MATERIALS AND METHODS

Cell growth

Strains were inoculated in 5 ml of YES medium and grown to saturation at $30°C$. Saturated cultures were backdiluted in 25 ml of YES medium and grown to mid-log phase (A_{600} between 0.4 and 0.7), flash frozen in liquid nitrogen then stored at $−80°C$.

Total RNA isolation, cDNA synthesis and microarray hybridization

Total RNA was isolated, converted to cDNA by reverse transcription, and labeled for microarray hybridization as described previously (13). Microarray probes were designed such that each protein coding gene was targeted by a single exonic probe (5098 in total). In addition, probes were designed targeting every annotated intron in the genome (4885 in total) as well as the junctions resulting from removal of those introns. Finally, a total of 126 probes were included that targeted a variety of non-coding RNAs, including snRNAs, snoRNAs and tRNAs. Additional details of the microarray platform are available through NCBI's GEO database using accession number GPL21065. Microarray experiments and data analyses were also conducted as described previously (13). Biological replicates were examined for each mutant, with a minimum of two technical replicates for each sample.

cDNA synthesis for quantitative PCR

Total RNA (4 μg) was treated with DNase (0.25 μl RQ1 DNase, Promega #M6101, in $1\times$ DNase buffer) at room temperature for 15 min to remove genomic DNA. RNA binding buffer (200 μl 2 M guanidinium, 75% isopropanol) was added. Samples were transferred to a DNA Clean and Concentrator 25 (Zymo #D4006) column and centrifuged at 14 000 rpm for 1 min. The flow-through was reloaded onto the column and centrifuged at 14 000 rpm for 1 min. Wash buffer (200 μl of 10 mM Tris–HCl, pH 8.0, 80% ethanol) was added to the column and centrifuged at 14 000 rpm for 1 min. Samples were washed a second time. The

column was transferred to a new microfuge tube, and the sample was eluted with 12 μl ddH$_2$O.

For cDNA synthesis, random primers (0.5 g/l dN$_9$) were annealed to purified, DNase-treated RNA (2 μg) in $1\times$ RT Buffer (50 mM Tris–HCl, pH 8.5, 75 mM KCl) at $60°C$ for 5 min and then cooled to room temperature for 5 min. RT mix was added ($1\times$ RT buffer, 3 mM MgCl$_2$, 10 mM DTT; 10 mM dNTPs, His6 MMLV RT (or commercial alternative)), samples were incubated for 2 h at $42°C$ and then stored at $−20°C$.

Primer design and optimization

All forward and reverse primers for quantitative PCR were designed using OligoWiz 2.0 (http://www.cbs.dtu.dk/services/OligoWiz/) (14). Representative subtelomeric genes were selected that met the following criteria: they were within 50 kb from the end of a chromosome; they had microarray intensities of greater than 1000; and primers with quality scores >0.9 could be generated by OligoWiz. Optimum annealing temperatures were determined via gradient qPCR experiments between 45 and $67°C$ using the gradient function of a BioRad MyCycler.

Quantitative PCR

Quantitative PCR reactions were performed in reactions containing 10 mM Tris–HCl, pH 8.5, 50 mM KCl, 1.5 mM MgCl$_2$, 0.2 mM dNTPs, 5% DMSO, $0.25\times$ SYBR Green, $1\times$ *Taq* DNA polymerase, and 250 nM forward and reverse primers. Data were collected using a Roche LightCycler 480. A 2-step protocol was used with a 3 min pre-incubation cycle at $95°C$, followed by 40 cycles of 10 s at $95°C$ and an annealing and extension period for 3 min at the primers' determined optimal annealing temperatures. Melting curves were generated to confirm the specificity of the amplicon.

Chromatin immunoprecipitation

Cultures (50 ml) of *S. pombe* were grown at $30°C$ to $A_{600} = 1.0$ and crosslinked by addition of formaldehyde (Sigma) to a final concentration of 1% with incubation at room temperature for 15 min. The crosslinking reaction was stopped with the addition of 125 mM glycine for 5 min. Cells were washed twice in $1\times$ TBS (20 mM Tris–HCl, pH 7.6, 0.15 M NaCl) and snap frozen. Samples were resuspended in 500 μl ChIP lysis buffer (50 mM HEPES–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and $1\times$ EDTA-free Complete protease inhibitor (Roche)) and were lysed in a Mini-Beadbeater (BioSpec Products) for 7 cycles of 1 min each with 2 min rests on ice. The cell lysate was transferred to a new tube and centrifuged for 10 min at 7000*g* at $4°C$. The resulting chromatin pellet was resuspended in 300 μl ChIP lysis buffer and sonicated for four 15 min cycles (High setting; 30 s on, 1 min off) in a Bioruptor waterbath sonicator (Diagenode). Cell debris was then removed by centrifugation at 21 000*g* for 10 min at $4°C$ and the supernatant was brought to 1.5 ml in ChIP lysis buffer. Immunoprecipitation was performed at $4°C$ overnight in 500 μl chromatin aliquots using 2 μg anti-H3K9me2 antibody (ab1220, Abcam). Pre-washed Protein

A Dynabeads (30 μl, Invitrogen) were used to bind the antibodies. Following a 2 h incubation at 4°C, the beads were washed twice with 1 ml ChIP lysis buffer, twice with 1 ml high salt ChIP lysis buffer (ChIP lysis buffer with 500 mM NaCl), twice with 1 ml wash buffer (10 mM Tris–HCl, pH 8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA), and once with 1 ml TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA). Each wash was performed for 5 min at room temperature with nutation. Beads were eluted with 100 μl elution buffer (50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 70°C for 20 min, followed by a second elution at room temperature for 5 min with 150 μl TE buffer with 0.67% SDS and 334 μg/ml proteinase K (Sigma). Input samples (50 μl) were combined with 200 μl TE buffer with 1% SDS and 250 μg/ml proteinase K. All samples were incubated overnight at 65°C, followed by addition of 1250 μl buffer NTB and DNA purification using a NucleoSpin Gel Clean-up column (Macherey-Nagel).

ChIP-seq library construction

For each genotype, library construction was performed for two biological replicates. Immunoprecipitation reactions or input samples were processed by treatment with the End-It DNA End-Repair Kit (Epicentre) and then A-tailed using Klenow fragment (New England Biolabs) according to manufacturers' instructions, with column purifications after each step using a NucleoSpin Gel Clean-up column (Marchery-Nagel) and MinElute Reaction Clean Up Kit (Qiagen) respectively. Next, barcoded Illumina adaptors (33 nM) were ligated to the DNA using 1200 U Rapid T4 DNA Ligase (Enzymatics) in a 30 μl reaction with incubation at room temperature for 30 min. The DNA was column purified using a NucleoSpin Gel Clean-up column (Marchery-Nagel) and PCR amplified using 1 U Phusion DNA polymerase (New England Biolabs), 2 M betaine, 75 μM dNTPs, and 0.4 μM each primer (TruSeq PCR enrichment primers Forward: 5'-AATGATACGGCGACCACCGAG-3'; Reverse: 5'-CAAGCAGAAGACGGCATACGAG-3') in a 50 μl reaction. Cycling conditions were: 98°C for 3 min, followed by 15 cycles of 98°C for 1 min 20 s, 65°C for 30 s, and 72°C for 30 s, with a final step of 72°C for 5 min. The reaction was size selected for products between 200 and 350 bp using SPRI beads (15). Library quality was determined by High Sensitivity DNA Bioanalyzer analysis (Agilent).

Sequencing data processing, read alignment and subtelomeric region analysis

ChIP-seq data were aligned using Bowtie1 (16). We allowed no mismatches within the seed sequence and, if a read could align equally well to multiple loci, a locus was chosen at random. Indexed, sorted bam files were created for each dataset using SAMtools (17) and bedgraph files were created using BEDtools (18). Using the bedgraph file, a sliding window algorithm was applied that averaged coverage over a 1000 bp window. This smoothed bedgraph was used for all downstream analysis.

The background signal level was defined as average coverage in all regions of the genome excluding the 80 kb adjacent to each chromosome end and the centromeric regions.

Telomeric and subtelomeric regions (80 kb adjacent to chromosome ends) were compared between datasets by normalizing each sample's coverage, on a per-base basis, to that of an input sample, consistent with field standards (19). To control for differences in read count, each sample was further normalized such that its background signal level (as defined above) would equal that of the input sample. This value corresponds to 1.0 in the plots.

Data availability

Microarray and ChIP-seq data are available through NCBI's GEO database (http://www.ncbi.nlm.nih.gov/geo/) using accession number GSE83426.

RESULTS

Global analyses using splicing-sensitive microarrays reveal changes in gene expression but not splicing in *rpb1-CTD-Ser2 and -Ser7* mutants

To examine the role of CTD serine phosphorylation marks on pre-mRNA splicing and global gene expression, we took advantage of *rpb1-CTD* mutant strains of *S. pombe* in which non-phosphorylatable alanines were substituted for every Ser2 (*S2A*), Ser7 (*S7A*) or both (*S2A/S7A*) in the context of a fully functional Rpb1 CTD array (comprising 14 consensus heptad repeats linked to the body of Rpb1 by a "rump" consisting of four degenerate heptads) (4). In addition, a fourth mutant in which each Ser7 was substituted by the phosphomimetic glutamic acid (*S7E*) was studied (4).

RNAseq has been employed previously to analyze changes in gene expression for a subset of these CTD mutants and no overt splicing defects were reported (12). However, while RNAseq is superior to microarrays for *de novo* discovery of introns, the low read depths typically obtained for either exon-exon or exon-intron boundaries can hinder meaningful quantitative analyses of genome-wide changes in splicing efficiency. Therefore, the genome-wide effects of these CTD mutations were examined here using custom-designed, splicing-sensitive microarrays that assay changes in gene expression for all annotated protein-coding genes in *S. pombe*, while simultaneously assaying splicing efficiency for every known intron in the genome (13). For each of the four CTD mutants, multiple replicate arrays were performed comparing relative RNA levels in the mutant and *WT* strains (see Materials and Methods). As a basis of comparison, we also examined the splicing defect associated with a mutation in the *prp2* gene, the *S. pombe* homolog of U2AF-65, which had been previously examined using this same microarray platform (20). As seen in Figure 1, none of the CTD mutants showed a global defect in splicing. The data, which are displayed in groups of three (total mRNA, pre-mRNA and mRNA) for each of the mutants, were arranged such that the ~4800 splicing events were ordered by the behavior of the pre-mRNA species, with those showing the largest increase appearing at the top and those with the largest decrease at the bottom. As expected, when examining the *prp2-1* data, pre-mRNA accumulation can be detected for the majority of introns in the genome. Accumulation of the pre-mRNA species in this mutant is accompanied, in most cases, by a decrease in the level of mature
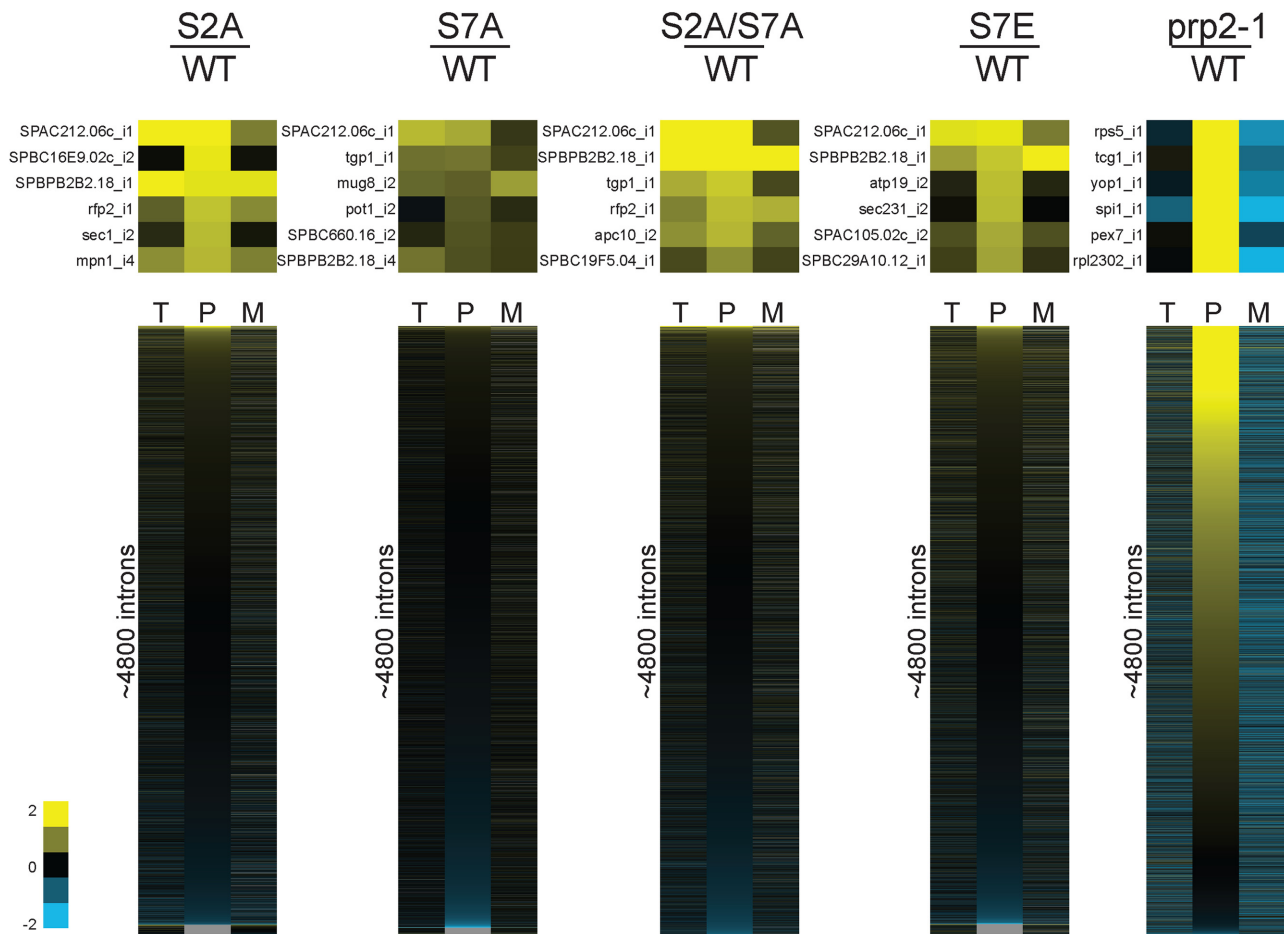
Figure 1. Effect of *rpb1*-CTD mutants on splicing. Genome-wide changes in pre-mRNA splicing for *rpb1*-CTD mutant strains *S2A*, *S7A*, *S2A/S7A*, and *S7E* compared to the isogenic *WT* strain were measured for ∼4800 introns via splicing-sensitive microarray analyses, and compared to the canonical splicing factor mutant strain *prp2-1*. Microarray data are displayed as Total mRNA (T), pre-mRNA (P), and mature mRNA (M) groupings for each strain. The data are ordered by largest to smallest accumulation of the pre-mRNA species. Insets show the top six pre-mRNA species for each strain.

mRNA, consistent with the *prp2-1* allele causing a *bona fide* splicing defect. By contrast, very few pre-mRNAs accumulated in the *rpb1-CTD* mutants. Moreover, as seen in the top insets of Figure 1, for the relatively small proportion of transcripts for which the pre-mRNA levels did increase in the *rpb1-CTD* mutants, the majority showed a simultaneous increase in the level of mature mRNA. These data suggest an overall increase in transcription rather than a change in splicing efficiency for these genes.

Spliceosomal U snRNA levels are unaffected by CTD Ser2 or Ser7 mutations

Prior studies had shown that mutations in the CTD, specifically substitutions at Ser7, affected the processing and steady-state levels of U2 snRNA in human cells (21). The microarray platform, which included probes targeting many non-coding RNAs, allowed us to monitor changes in the levels of U snRNAs in the CTD mutant strains. We found that there was little change in the levels of U1, U2, U4, U5 and U6 snRNAs in the *rpb1-CTD* mutant strains (Supplemental Figure S1).

Global analyses of gene expression in *rpb1-CTD-Ser2 and -Ser7* mutant strains

To characterize global changes in gene expression, we used the Significance Analysis of Microarrays (SAM) tool to identify transcripts with significantly different changes in gene expression in the replicate datasets of each of these mutants (22). Because SAM assesses the statistical significance of a change without necessarily considering the biological significance of those changes, we further restricted our analyses to those transcripts which both satisfied the SAM significance test and were at least two-fold changed in their overall gene expression. By these criteria, each of the mutants caused statistically significant changes in gene expression: a total of 192 genes were misregulated in the *S2A* mutant, 167 in the *S2A/S7A* mutant, 57 in the *S7A* mutant, and 137 in the *S7E* mutant (Supplemental Table S1). That more genes were misregulated in *S2A*, *S2A/S7A*, and *S7E* than in *S7A* correlates with the observations that *S7A* cells exhibited no discernible growth defect at temperatures between 18 and 37°C, whereas the other mutants did exhibit cold-sensitive or heat-sensitive growth defects (4). In general, increases in gene expression outnumbered decreases in
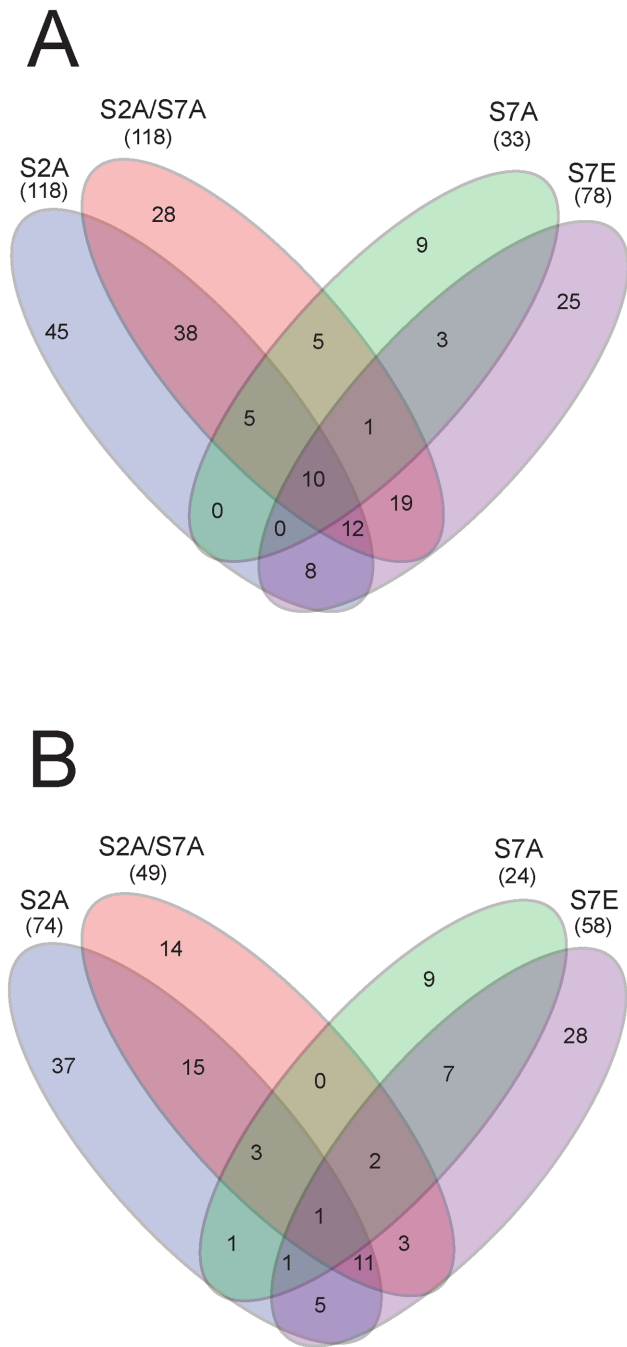
Figure 2. Effect of *rpb1*-CTD mutants on *S. pombe* transcriptomes. Protein-coding genes that are significantly upregulated (**A**) or downregulated (**B**) based on SAM analyses for the *rpb1-CTD* mutant strains relative to the *WT* strain and genes that overlap between mutants are shown by Venn diagram. Numbers in parentheses indicate total numbers of genes upregulated or downregulated for each *rpb1-CTD* strain.

gene expression. Moreover, significant overlap of gene expression targets was seen between the different mutants, in particular for *S2A* and *S2A/S7A*, which share more than half of their respective targets (Figure 2).

Comparison of microarray gene expression data with RNAseq data

Global gene expression can be monitored by microarray and RNAseq analyses and although the correlations seen between datasets generated by these two approaches might be limited based on many considerations (23,24), we sought to compare prior RNAseq data available for a subset of the CTD mutants (12) to our microarray data, which also include the *S7E* strain. In spite of technical differences in processing the samples (e.g. the use of total RNA versus poly(A)$^+$ RNA), reasonable correlations were seen between the RNAseq and microarray datasets comparing *WT* and *S2A* strains (Pearson correlation coefficient of 0.54), as well as the *WT* and *S2A/S7A* strains (Perason correlation coefficient of 0.59). Likely owing to the smaller number of genes showing misregulation in *S7A*, a lower correlation was seen for these datasets (Pearson correlation coefficient of 0.40). When only considering sets of genes that were identified as being statistically significantly misregulated by the two approaches, the correlation was high. For the *WT* versus *S2A* comparison, of the 155 genes that were identified in this work as significantly misregulated and for which RNAseq data were available, 61 genes (40%) were identified as significantly misregulated by RNAseq. For the *WT* versus *S2A/S7A* comparison, 55 out of the 114 genes (48%) were in common, and for the *WT* versus *S7A* comparison, 16 out of the 41 genes (39%) were found in common. Nearly half of the genes that were identified in our study as being significantly misregulated but which were not called significantly misregulated in the RNAseq dataset were characterized by low numbers of RNA-seq reads (Supplemental Table S2). Further characterizations of the similarities and differences seen between our data and the published RNAseq data are described below.

Global analyses of genes with altered expression in the *rpb1-CTD* mutants reveal limited enrichment by Gene Ontology

The protein-coding genes affected by the Rpb1-CTD mutations were examined to identify enrichment of functional gene categories as defined by Gene Ontology (GO) analysis. The sets of genes that were either significantly upregulated or downregulated by our SAM analyses for each mutant strain were compared to the background distribution. Whereas some functional categories were identified by this approach (see Supplemental Table S3), overall the analysis failed to identify large groups of functionally related transcripts. For those categories that showed statistically significant over-representation, only a small number of specific transcripts were identified, leaving the misregulated expression of most transcripts unexplained from a functional perspective. An earlier RNAseq-based analysis of some of these same mutants in *S. pombe* reported changes in functionally related clusters of affected genes including meiotic genes, iron response genes, and phosphate homeostasis genes (12). Although these multi-gene pathways did not exhibit overall enrichment within our dataset, an examination of the behavior of specific transcripts within these GO categories between the two datasets revealed broad overlap (Supplemental Tables S2 and S3), largely reflecting the overall similarities described above for these datasets. For exam-

ple, the meiotic genes *ste11* and *mei2* that were previously demonstrated by Northern analysis to be affected by serine phospho-site mutations (4), were also misregulated within our data, showing decreased expression in the *S2A* mutant, increased expression in the *S7A* and *S7E* mutants, but intermediate levels in the *S2A/S7A* double mutant (Supplemental Figure S2).

Gene expression in subtelomeric regions is specifically affected by RNAP II CTD mutants

While examining the functions of the genes for which expression was impacted by the CTD mutations, it became apparent that a number of galactose response genes were upregulated in the *S2A* strain, notwithstanding that the cells from which the RNA was isolated were grown in glucose-containing medium. Indeed, four out of six galactose response genes (*gal1, gal7, gal10* and *SPBPB2B2.11*) showed significant upregulation in at least one of the four CTD mutants. We noticed that these four upregulated genes were located next to one another, and quite close to the right end of chromosome 2, suggesting that their increased expression could be a consequence of their similar chromosomal locations rather than any functional relationship.

To examine whether chromosomal location was broadly correlated with changes in gene expression resulting from the CTD mutations, sliding windows of gene expression changes were determined by averaging the $\log_2$ ratio of mutant *vs WT* expression values for every coding transcript in 10 kbp windows along each chromosome. As seen in Figure 3A, patterns of upregulated gene expression were apparent near the ends of chromosomes 1 and 2 for all four CTD mutants: statistically significant upregulation is seen for the 50 kb from the ends except for that of the right arm of chromosome 1 for the *S7A* and *S2A/S7A* mutants. Whereas increased gene expression was not observed for either arm of chromosome 3, the ends of chromosome 3 in *S. pombe* are different from chromosomes 1 and 2 in that they comprise arrays of repeated rDNA, which are typically transcribed by RNA polymerase I. On chromosomes 1 and 2, the regions of upregulation were large, extending ∼50–100 kb from the telomeric ends of the chromosomes, and perhaps defining the subtelomeric regions of these chromosomes (Figure 3B). As shown in Figure 3B, large regions with downregulated gene expression were also observed on chromosomes 1 and 2 immediately adjacent to the regions of upregulation, ranging in size from ∼25 to 80 kb and present on all but the right arm of chromosome 2. In contrast to these subtelomeric regions, the gene expression values across the remainder of the chromosomes were generally unchanged from *WT*, with only few exceptions.

Previous studies of global gene expression changes of *rpb1-CTD* mutants in *S. pombe* via microarray analyses did not describe region-specific changes in gene expression (10,11). By independently examining the data from these studies, we detected similar increases in subtelomeric gene expression in the Saberianfar *et al.* data at the ends of chromosomes 1 and 2, but not chromosome 3 for the *S2A* mutant, the one allele common to all experiments (Supplemental Figure S3).
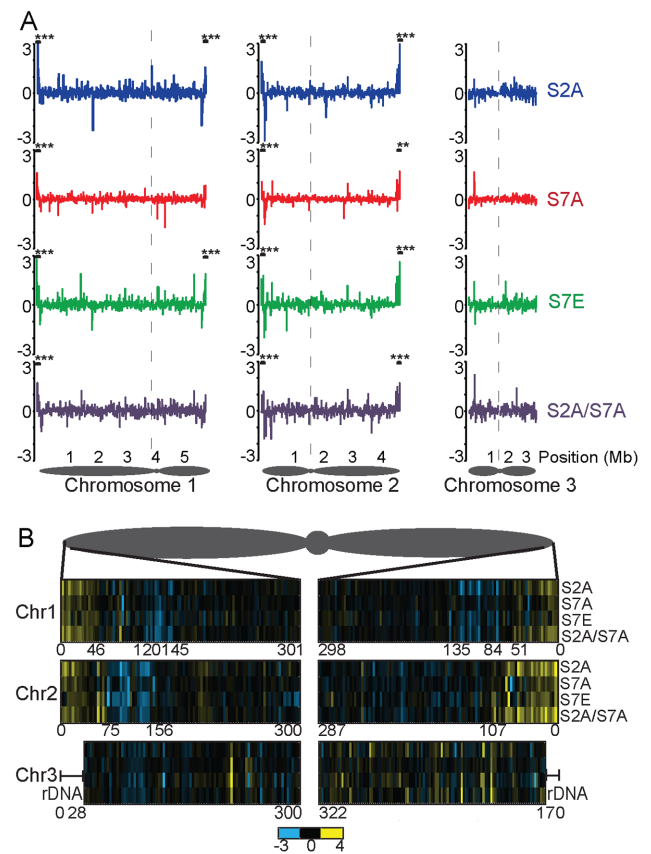

Figure 3. Chromosomal location of gene expression changes in *rpb1-CTD* strains. (**A**) Global changes displayed across each chromosome. Gene expression level changes were measured by microarray analyses, expressed as the $\log_2$ ratio of mutant *vs WT* for *S2A* (blue), *S7A* (red), *S7E*, (green) and *S2A/S7A* (purple) CTD mutants, averaged for every coding transcript within sliding 10 kb windows, and plotted *vs* chromosomal location. The dotted line indicates the location of the centromere. ***$P < 0.001$ and **$P < 0.01$ for the first 50 kb of the subtelomeric region. (**B**) The gene expression changes for every exon in the ∼300 kb region adjacent to the telomeric ends of chromosomes 1, 2 and 3 of *S. pombe*. Each bar represents an individual gene. Chromosomal distances are not to scale ($n = 3–5$).

Validation of gene expression changes by quantitative RT-PCR

Reverse transcription and quantitative PCR (RT-qPCR) analyses were performed to validate the changes in expression detected by microarray. Primers were designed targeting two or more genes within the subtelomeric regions of each arm for chromosomes 1, 2 and 3 that were within 50 kb from the end of the chromosome (Supplemental Table S4). In addition, genes outside of the subtelomeric regions for chromosomes 1 and 2 were examined. The RT-qPCR data confirmed significant upregulation of genes in the subtelomeric regions for each of the four CTD mutants, whereas expression of the non-subtelomeric genes were unchanged compared to *WT* (Figure 4 and Supplemental Figure S4). We further used RT-qPCR to examine changes in expression of the meiotically regulated genes *ste11* and *mei2*, each of which had been previously shown to misregulate in the context of CTD mutants (4). The RT-qPCR data corroborated both our microarray data and the previ-
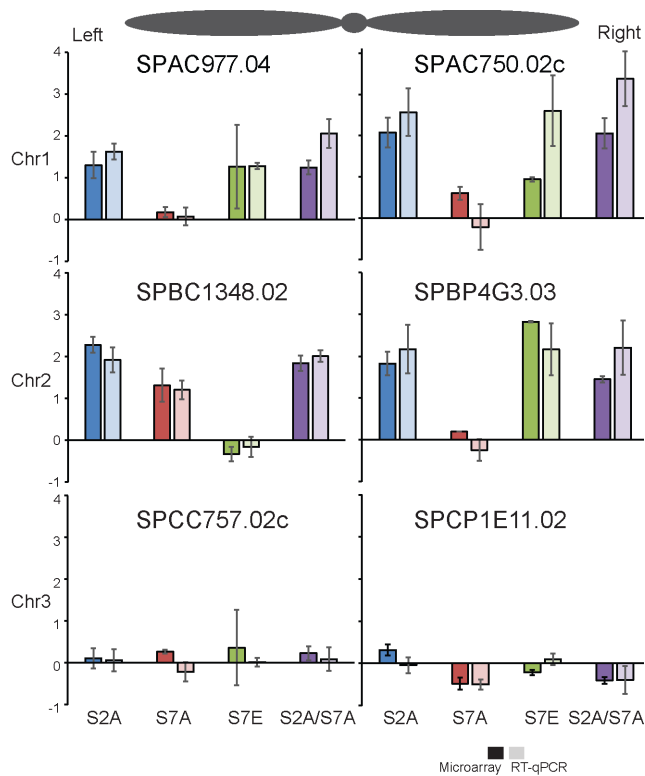
Figure 4. Effects of *rpb1-CTD* mutations on gene expression of individual genes by microarray and RT-qPCR analyses. Total mRNA levels for selected genes found to be upregulated within the 50 kb subtelomeric region by microarray analyses, as well as genes selected from the non-subtelomeric regions, were measured in the CTD mutants and compared to the *WT* strain by RT-qPCR analyses (see also Supplemental Figure S4). The $\log_2$ ratio of mutant *vs* WT were plotted for chromosomes 1, 2 and 3 for both microarray ($n = 3$–5) and RT-qPCR ($n = 2$) data. For the RT-qPCR experiments, gene expression levels were all normalized to *act1* mRNA levels.

ously published data, showing that *ste11* and *mei2* mRNA levels were decreased in the *S2A* strain, increased in the *S7A* strain, and near *WT* levels in the *S2A/S7A* strain. Importantly, these changes in gene expression match the known sterility phenotype of the *S2A* mutant that is reversed by the double mutant *S2A/S7A* (Supplemental Figure S2) (4).

Chromatin modifications are altered near the ends of the chromosomes

The telomeric and subtelomeric regions of chromosomes are typically heterochromatic, facilitating repression of gene expression in these regions. Given the increases in gene expression in these regions that accompanied the CTD mutations, we were interested in investigating whether changes in chromatin structure might be involved. ChIP-Seq was performed to query whether the repressive heterochromatin mark H3K9 methylation was altered in the *rpb1-CTD* mutant strains compared to *WT*. Consistent with our gene expression results, in the S2A mutant a general decrease in the H3K9 methylation levels was observed at the ends of chromosomes 1 and 2, but not chromosome 3 (Figure 5 and data not shown). The decrease in the H3K9 repressive marks appears within the first ~20 to ~45 kb from the chromoso-
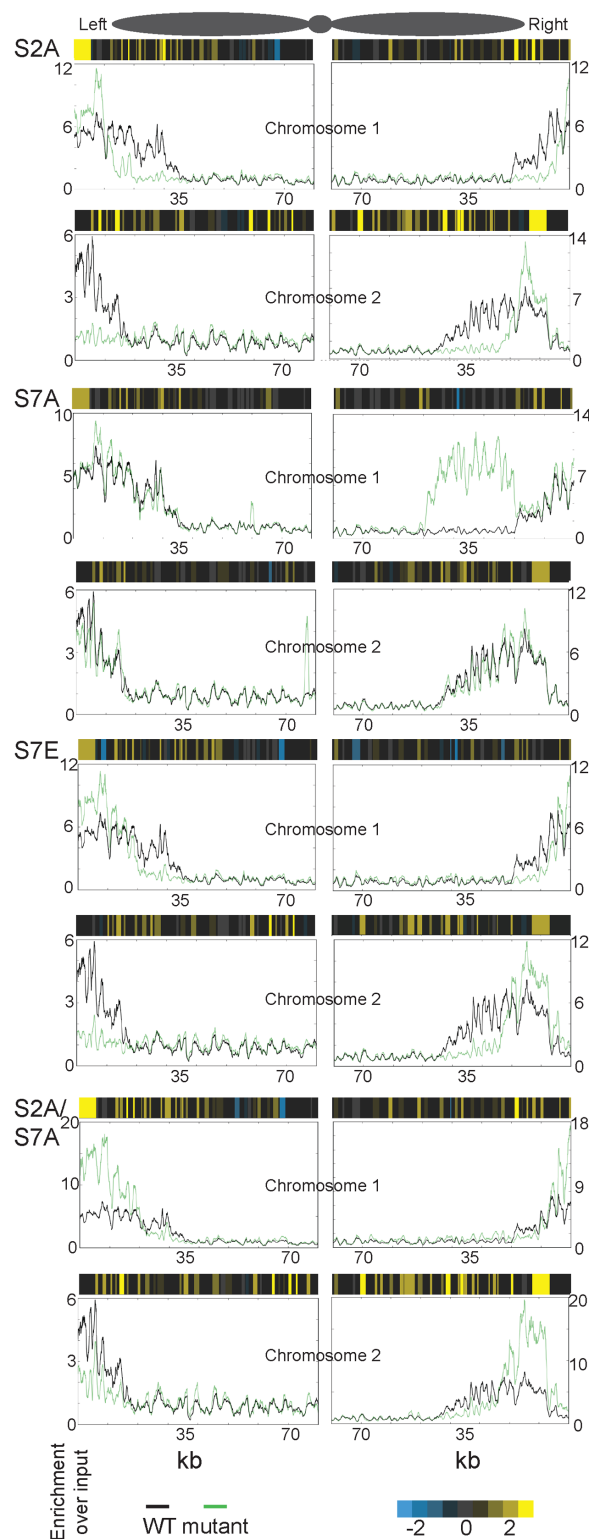
Figure 5. H3K9 methylated histone levels are altered in subtelomeric regions of *rpb1-CTD* mutants. H3K9-methylation modified histones were immunoprecipitated with associated chromatin. The DNA was isolated and sequenced by Illumina sequencing. Density of reads enriched over whole cell extract input (y-axis) was plotted for the first 80 kbp for Chromosome 1 and 2 for *S. pombe*. WT samples are in black, *rpb1-CTD* mutants are in green ($n = 2$). Heat maps of gene expression level changes compared to *WT* as measured by microarray are overlaid above the ChIP-Seq traces.

mal ends; by contrast, the regions of increased gene expression observed by microarray were larger, ranging from ~50 to ~100 kb (Figures 3B and 5). Nevertheless, genes with decreased H3K9 methylation overlap genes with increased gene expression in the subtelomeric regions in our mutant strains (*P*<0.001, Fishers Exact). Each of the other CTD mutations tested also altered the pattern of H3K9me, but in more complex ways. The *S7A* allele produced no clear decreases in H3K9me, but rather an increase in a subtelomeric region on right arm of chromosome 1. A similar phenotype has been reported in the *leo1Δ* mutant, which encodes a subunit of the PAF1 elongation complex (25). The *S7E* and *S2A/S7A* mutants produced mixed effects with increases at some sites and decreases at other (Figure 5). While the correlation between H3K9 methylation state and gene expression level were significant, there were also numerous regions of divergence. For example, for the regions directly adjacent to the telomere on the left arm of chromosome 1 and the right arm of chromosome 2, H3K9 methylation levels were higher than *WT* for the *S2A*, *S7E*, and *S2A/S7A* mutants. However, gene expression within these regions was derepressed in the mutant despite the increased levels of the repressive H3K9 mark (Figure 5). These data indicate that the CTD mutants impact subtelomeric repression mechanisms beyond those determined by H3K9me levels.

DISCUSSION

Our initial motivation in this work was to interrogate the relationship between CTD phosphorylation state and pre-mRNA splicing efficiency. Numerous lines of evidence over the past several years have demonstrated the co-transcriptional nature of pre-mRNA splicing (26). Whereas several reports have suggested that this may be facilitated by direct interactions between the CTD and specific splicing factors, others have cast doubt upon these direct connections (27,28). Using splicing-sensitive microarrays to assess the pre-mRNA splicing efficiency of nearly every annotated splicing event in the fission yeast genome, we detected almost no defect in pre-mRNA splicing in the background of any of the mutants examined (Figure 1). Although small numbers of splicing events can be identified that show weak apparent defects in splicing efficiency, from a global perspective these data strongly support the conclusion that phosphorylation of Ser2 or Ser7 in the *S. pombe* CTD is not required for efficient pre-mRNA splicing genome-wide.

Absent defects in pre-mRNA splicing, we queried changes in gene expression. Several of the mutants examined here have been studied by others using either microarrays or RNAseq (4,10–12). Overall, the global changes in gene expression detected here were reasonably well-correlated with those seen in previous studies, particularly given the differences in detection platform and sample processing (Pearson correlations between 0.4 and 0.59). As with previous studies, our data demonstrated mis-regulation of several genes involved in iron response, phosphate homeostasis, and meiosis. Nevertheless, a global analysis of the gene ontologies associated with regulated transcripts failed to provide a compelling functional relationship between mis-regulated transcripts. While the GO analysis did identify a handful of enriched functional categories, the total

number of transcripts contained within these categories represented only a fraction of the total transcripts misregulated by these mutants. Together, these data suggest that the changes in gene expression caused by these CTD mutations cannot be readily explained by alterations in functionally-related transcripts.

A common trend among many of the genes we found to be significantly upregulated in our genome-wide microarray experiment was their location within their chromosomes. Many of the genes we identified were clustered near the ends of chromosomes 1 and 2 (Figure 3). These areas may define a functionally distinct so-called subtelomeric region of their respective chromosomes. Chromosome 3 in *S. pombe* has a different structure than the other two chromosomes in that ribosomal RNA gene repeats exist near the ends of chromosome 3, perhaps explaining the distinct expression patterns seen in the CTD mutants for chromosomes 1 and 2, but not chromosome 3. Other groups have examined global changes in gene expression in the *S2A* mutant in *S. pombe* (10,11), albeit using different experimental platforms. While these studies identified and discussed expression changes in genes involved in cytokinesis and meiosis respectively, our analysis of the data generated by Saberianfar *et al.* show an up-regulation of genes in the subtelomeric regions of chromosomes 1 and 2, but not chromosome 3 (Supplemental Figure S3). Whereas a previous RNA-seq study with these CTD mutants (12) did not appear to uncover broad upregulation of gene expression within the subtelomeric regions, RNA-seq analyses typically remove from consideration sequences that align to repetitive regions such as the subtelomere. Indeed, many of the subtelomeric genes identified as being upregulated in our study have zero reads associated with them in the RNAseq dataset (Supplemental Table S2). Nevertheless, of the few subtelomeric genes that were preserved in that study, for example, the RecQ type DNA helicase *tlh2* was found to be similarly upregulated, matching our observation.

Given that the CTD mutations resulted in changes in gene expression within subtelomeric regions that are expected to exist in a heterochromatic state, we assayed the repressive heterochromatin mark H3K9 methylation using chromatin immunoprecipitation and subsequent sequencing (ChIP-seq). The H3K9 methylation levels were generally reduced in the first ~20–45 kb of chromosomes 1 and 2, but not chromosome 3, often overlapping the regions of increased gene expression in the CTD mutants. However, two regions were identified where the gene expression levels were high and yet the H3K9 methylation levels were also high. These regions, most prominently on the left arm of chromosome 1 and the right arm of chromosome 2, contain the *tlh1* and *tlh2* genes that encode RecQ type DNA helicases and are thought to nucleate heterochromatin assembly at telomeres (29,30). Given that expression of *tlh1* and *tlh2* is increased despite elevated H3K9 methylation, their expression may be necessary or perhaps even privileged for proper heterochromatin formation at these sites. Alternatively, there may exist subtelomeric repression mechanisms that are dependent on RNAP II CTD status, yet are independent of H3K9me. Supporting this view, the regions of upregulated gene expression identified here seem to extend ~50 kb further into the chromosome than the zones

of decreased H3K9 methylation, consistent with the idea that subtelomeric chromatin is distinct from euchromatin (31,32).

The precise mechanisms by which the RNAP II CTD mutations affect chromatin structure and gene expression remain to be elucidated. Several of the CTD repeat mutations examined here yielded similar yet distinct phenotypes, suggesting that the relevant CTD 'reader' protein(s) may recognize more than one CTD residue. While future work will be required to identify the relevant reader(s), it is is tempting to speculate that disruption of phosphorylation events on the CTD alters recruitment of chromatin modifying enzymes such as members of the Set family of histone methyltransferases. For example, Set2 associates with the phosphorylated form of the CTD (33), and Set2 protein levels have also been shown to be destabilized in CTD phosphomutants (34). Similarly, recent studies in *S. cerevisiae* have shown that Set1 and Set5 act together to modulate subtelomeric silencing (35). Whether these activities or components of the repressive machinery itself recognize the CTD in this current context will require a biochemical identification of the proximal recognition factor(s).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Hsin,J.-P. and Manley,J.L. (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.*, **26**, 2119–2137.
2. Eick,D. and Geyer,M. (2013) The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem. Rev.*, **113**, 8456–8490.
3. Buratowski,S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, **36**, 541–546.
4. Schwer,B. and Shuman,S. (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol. Cell*, **43**, 311–318.
5. Mbogning,J., Pagé,V., Burston,J., Schwenger,E., Fisher,R.P., Schwer,B., Shuman,S. and Tanny,J.C. (2015) Functional interaction of Rpb1 and Spt5 C-terminal domains in co-transcriptional histone modification. *Nucleic Acids Res.*, **43**, 9766–9775.
6. David,C.J., Boyne,A.R., Millhouse,S.R. and Manley,J.L. (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.*, **25**, 972–983.
7. Morris,D.P. and Greenleaf,A.L. (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.*, **275**, 39935–39943.
8. Gu,B., Eick,D. and Bensaude,O. (2013) CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.*, **41**, 1591–1603.
9. Hsin,J.-P., Xiang,K. and Manley,J.L. (2014) Function and control of RNA Polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. *Mol. Cell. Biol.*, **34**, 2488–2498.
10. Saberianfar,R., Cunningham-Dunlop,S. and Karagiannis,J. (2011) Global gene expression analysis of fission yeast mutants impaired in Ser-2 phosphorylation of the RNA pol II carboxy terminal domain. *PloS One*, **6**, e24694.
11. Coudreuse,D., van Bakel,H., Dewez,M., Soutourina,J., Parnell,T., Vandenhaute,J., Cairns,B., Werner,M. and Hermand,D. (2010) A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Curr. Biol.*, **20**, 1053–1064.
12. Schwer,B., Bitton,D.A., Sanchez,A.M., Bähler,J. and Shuman,S. (2014) Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 4185–4190.
13. Inada,M. and Pleiss,J.A. (2010) Genome-wide approaches to monitor pre-mRNA splicing. *Methods Enzymol.*, **470**, 51–75.
14. Wernersson,R., Juncker,A.S. and Nielsen,H.B. (2007) Probe selection for DNA microarrays using OligoWiz. *Nat. Protoc.*, **2**, 2677–2691.
15. Rohland,N. and Reich,D. (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.*, **22**, 939–946.
16. 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.
17. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.*, **25**, 2078–2079.
18. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.*, **26**, 841–842.
19. Flensburg,C., Kinkel,S.A., Keniry,A., Blewitt,M.E. and Oshlack,A. (2014) A comparison of control samples for ChIP-seq of histone modifications. *Front. Genet.*, **5**, 329.
20. Lipp,J.J., Marvin,M.C., Shokat,K.M. and Guthrie,C. (2015) SR protein kinases promote splicing of nonconsensus introns. *Nat. Struct. Mol. Biol.*, **22**, 611–617.
21. Egloff,S., O'Reilly,D., Chapman,R.D., Taylor,A., Tanzhaus,K., Pitts,L., Eick,D. and Murphy,S. (2007) Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. *Science*, **318**, 1777–1779.
22. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
23. Ho,J.W.K., Bishop,E., Karchenko,P.V., Nègre,N., White,K.P. and Park,P.J. (2011) ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC Genomics*, **12**, 134.
24. Nookaew,I., Papini,M., Pornputtapong,N., Scalcinati,G., Fagerberg,L., Uhlén,M. and Nielsen,J. (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **40**, 10084–10097.
25. Sadeghi,L., Prasad,P., Ekwall,K., Cohen,A. and Svensson,J.P. (2015) The Paf1 complex factors Leo1 and Paf1 promote local histone turnover to modulate chromatin states in fission yeast. *EMBO Rep.*, **16**, 1673–1687.
26. Bentley,D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–175.
27. Görnemann,J., Barrandon,C., Hujer,K., Rutz,B., Rigaut,G., Kotovic,K.M., Faux,C., Neugebauer,K.M. and Séraphin,B. (2011) Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. *RNA N. Y. N*, **17**, 2119–2129.
28. Tardiff,D.F., Lacadie,S.A. and Rosbash,M. (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol. Cell*, **24**, 917–929.

29. Kanoh,J., Sadaie,M., Urano,T. and Ishikawa,F. (2005) Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Curr. Biol.*, **15**, 1808–1819.

30. Hansen,K.R., Ibarra,P.T. and Thon,G. (2006) Evolutionary-conserved telomere-linked helicase genes of fission yeast are repressed by silencing factors, RNAi components and the telomere-binding protein Taz1. *Nucleic Acids Res.*, **34**, 78–88.

31. Matsuda,A., Chikashige,Y., Ding,D.-Q., Ohtsuki,C., Mori,C., Asakawa,H., Kimura,H., Haraguchi,T. and Hiraoka,Y. (2015) Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast. *Nat. Commun.*, **6**, 7753.

32. Tashiro,S., Handa,T., Matsuda,A., Ban,T., Takigawa,T., Miyasato,K., Ishii,K., Kugou,K., Ohta,K., Hiraoka,Y. *et al.* (2016) Shugoshin forms a specialized chromatin domain at subtelomeres that regulates transcription and replication timing. *Nat. Commun.*, **7**, 10393.

33. Krogan,N.J., Kim,M., Tong,A., Golshani,A., Cagney,G., Canadien,V., Richards,D.P., Beattie,B.K., Emili,A., Boone,C. *et al.* (2003) Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.*, **23**, 4207–4218.

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

35. Martín,G.M., King,D.A., Green,E.M., Garcia-Nieto,P.E., Alexander,R., Collins,S.R., Krogan,N.J., Gozani,O.P. and Morrison,A.J. (2014) Set5 and Set1 cooperate to repress gene expression at telomeres and retrotransposons. *Epigenetics Off. J. DNA Methylation Soc.*, **9**, 513–522.