The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain

Kevin M. Harlen and L. Stirling Churchman

Abstract | The carboxy-terminal domain (CTD) extends from the largest subunit of RNA polymerase II (Pol II) as a long, repetitive and largely unstructured polypeptide chain. Throughout the transcription process, the CTD is dynamically modified by post-translational modifications, many of which facilitate or hinder the recruitment of key regulatory factors of Pol II that collectively constitute the 'CTD code'. Recent studies have revealed how the physicochemical properties of the CTD promote phase separation in the presence of other low-complexity domains. Here, we discuss the intricacies of the CTD code and how the newly characterized physicochemical properties of the CTD expand the function of the CTD beyond the code.

Low complexity
Protein domains with low variability of amino acids.

Extending from the enzymatic core of RNA polymerase II (Pol II) is a long unstructured protein domain composed of tandem repeats of low complexity. This domain consists of heptapeptide repeats with the consensus sequence YSPTSPS and forms the carboxy-terminal domain (CTD) of the largest subunit of Pol II RNA polymerase B1 (RPB1). The number of repeats varies between species, ranging from as few as 5 in *Plasmodium yoelii* to 26 in budding yeast (*Saccharomyces cerevisiae*) and 52 in humans[1] (FIG. 1). The CTD is not required for the catalytic functions of Pol II, but it is crucial for the regulation of multiple steps in transcription, as well as for the coupling of transcription to a number of co-transcriptional processes[2–5]. The transcription process in eukaryotes consists of three major stages: initiation, elongation and termination. Both the CTD in its entirety and the individual residues within the heptapeptide have crucial roles in transcription regulation throughout the transcription process.

The sequence of the CTD consists of residues that can be subjected to multiple post-translational modifications (PTMs), including phosphorylation, proline isomerization and *O*-GlcNAcylation[2,5–8]. In addition, non-consensus Lys and Arg residues within the CTDs of some organisms can be methylated, ubiquitylated or acetylated[9–14] (TABLE 1). These PTMs constitute the 'CTD code' and are key to regulating processes such as pre-mRNA capping, splicing and the PTM of histones[2,4,6].

By far the most studied and best-characterized CTD modification is phosphorylation. In general, CTD phosphorylation promotes or inhibits the binding of proteins to the CTD, thereby enabling the specific recruitment of appropriate proteins to transcribing Pol II throughout the transcription process[15–17]. The CTD is frequently referred to as a landing pad at which myriad transcription factors, chromatin modifiers and RNA processing enzymes are recruited to the actively transcribing polymerase through specific interactions with modified CTD residues[16,17]. Although the repetitive nature of the CTD increases the potential complexity of the code, recent work has demonstrated that many repeats within a given CTD are not phosphorylated[18,19] (BOX 1). Thus, combinatorial calculations imply that the number of actual CTD states is experimentally manageable: in the hundreds[18,19] rather than astronomical.

The size of the CTD varies dramatically between species[1,20]. At the low end, the *Trypanosoma brucei*[21] Pol II lacks a CTD entirely, whereas the *P. yoelii* CTD has only five heptapeptide repeats, which exhibit minimal conservation. *S. cerevisiae* has 26 repeats, most of which are consensus[1,20]. Of the 52 repeats that constitute the human CTD, the first 26 are highly conserved whereas the remainder are less so[1] (FIG. 1). The entire CTD code is not essential for the function of Pol II, as a full-length CTD is not required for viability or even for sustaining wild-type growth rates in budding yeast. Genetic studies have revealed that the minimal length required for viability is eight repeats and that only ten repeats are needed for growing wild-type yeast on rich media[22]. To overcome temperature sensitivity, cold sensitivity and auxotrophic phenotypes, 13 repeats are required[23]. Thus, to maintain viability, only half of the yeast CTD is required. However, recent work has demonstrated that a reduced repeat number is unstable

Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

Correspondence to L.S.C.
churchman@genetics.med.harvard.edu

doi:10.1038/nrm.2017.10
Published online 1 Mar 2017

Multivalent interactions
Weak non-covalent interactions between biological molecules.

Liquid–liquid phase separation
The process of forming fluid, liquid-like droplets or compartments.

in yeast, and that DNA recombination and resolution of G-quadruplex–like DNA structures that are present in the CTD maintain the wild-type number[24]. Thus, although a handful of CTD repeats are sufficient for viability, a full-length CTD is preferred and actively maintained by the cell. Furthermore, mutations of some CTD residues in yeast are tolerated or lead to mild phenotypes in other cases[22,25–30]. This finding demonstrates that the CTD code works in concert with many other processes, such as the formation of protein–RNA complexes, to control the production and processing of RNA. Thus, the CTD code is not a stand-alone code like the genetic code, but is instead a component of a larger biological mechanism that regulates transcription and co-transcriptional processes.

The repetitive nature and perhaps the length of the CTD also contribute to its physical properties, which may be important for regulating transcription. The amino acid repeats within the CTD confer low-complexity protein domain properties to this region of RPB1. Low-complexity domains are often a hallmark of proteins found in membraneless cellular compartments, which display liquid-like properties owing to numerous weak, non-covalent interactions between their various components[31,32]. The presence of such multivalent interactions can facilitate liquid–liquid phase separation, thereby contributing to the formation of the membraneless compartments. The CTD has recently been shown to form multivalent interactions that can manifest as liquid–liquid phase separation[33,34], suggesting another mechanism by which the CTD may regulate transcription.

The CTD has been reviewed extensively[1–4,6,35–38]. Therefore, in this Review, we only briefly summarize the established roles of the CTD in transcription. We then focus on recent work demonstrating the dynamic functional states of the CTD and the relationships between these dynamic states and transcription. In addition, we discuss recent work suggesting that the CTD serves to regulate transcription through its intrinsic ability to form phase-separated structures through multivalent interactions with low-complexity domains of other regulatory factors[33,34], implying that the role of the CTD extends beyond the CTD code.

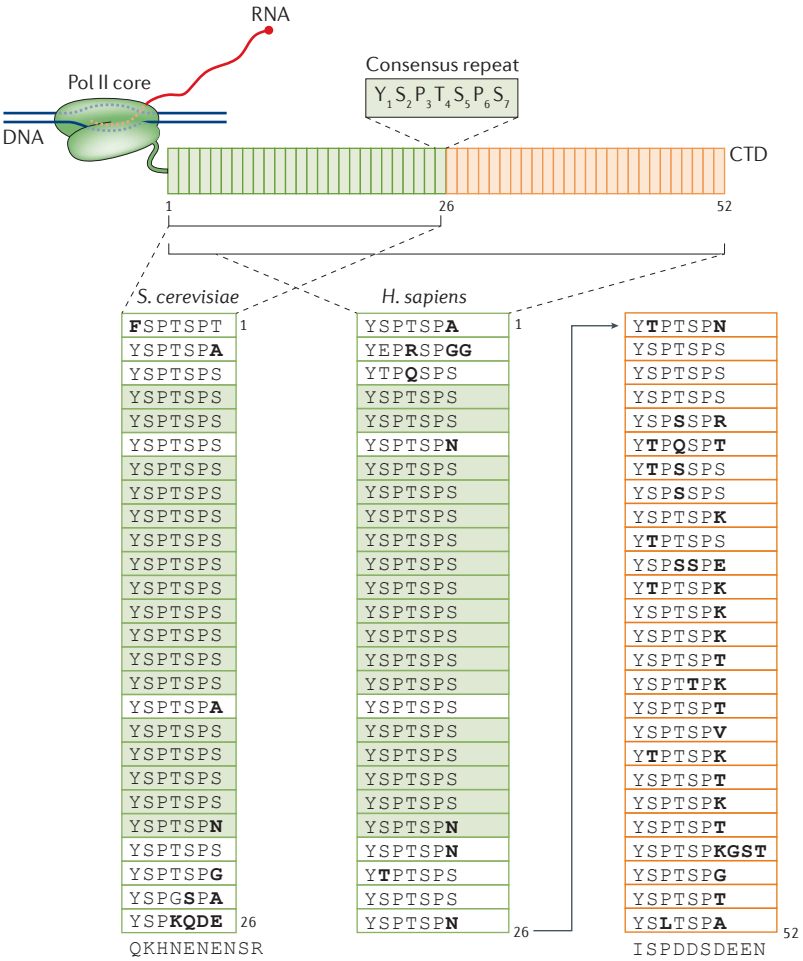

Figure 1 | **The composition and conservation of the CTD.** The core and the carboxy-terminal domain (CTD) of *Saccharomyces cerevisiae* and human RNA polymerase II (Pol II) are depicted. Each rectangle represents one CTD repeat and the repeat-sequence consensus is shown above. The yeast CTD is composed of 26 mostly consensus repeats, whereas the human CTD is composed of 52 repeats, the first 26 being highly similar to the yeast CTD whereas the remaining 26 are more divergent from the consensus sequence. Identical repeat sequences between yeast and human CTDs are highlighted in green; note the high degree of conservation between *S. cerevisiae* and the first half of the human CTD.

Transcription and related processes

The CTD regulates each step in the transcription process, from initiation to termination. In addition, the CTD is important for regulating several co-transcriptional processes, such as splicing and chromatin modification. In this section we summarize the roles of the CTD during transcription.

Transcription initiation. Pol II is recruited to promoters with an unmodified CTD[39,40] and forms the pre-initiation complex through interactions with general transcription factors. The CTD is important for pre-initiation complex formation because the unmodified CTD has a high affinity for the Mediator complex[39,41,42], a transcription co-activator that is important for the formation of bridging interactions between transcription activators at enhancers and the general transcription machinery at promoters[43]. The high affinity of the unmodified CTD for the Mediator complex is due to the formation of multiple hydrogen bonds and hydrophobic interactions with subunits of the Mediator complex[39,41,42]. Promoter escape by Pol II and the transition from transcription initiation to elongation is partly regulated by phosphorylation of the CTD[44]. Residues within the CTD are referred to by their order in the heptapeptide: Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Cyclin-dependent kinase 7 (CDK7; known as Kin28 in *S. cerevisiae*), which is a component of the general transcription factor TFIIH, phosphorylates Ser5 and Ser7 and aids in promoter escape by Pol II. The phosphorylation of Ser5 lowers the affinity of Pol II for the Mediator complex, probably by disrupting the hydrogen bonding between them[41,45–47]. After phosphorylation of Ser5 and Ser7, other CTD residues are dynamically phosphorylated and dephosphorylated by several CTD kinases

Table 1 | **Post-translational modifications of the CTD and their associated processes**

Post-translational modification	Position in the CTD	Organisms	Associated process or processes	Refs
Ser5 phosphorylation	Multiple repeats	• *Saccharomyces cerevisiae* • *Schizosaccharomyces pombe* • *Homo sapiens*	Transcription initiation, mRNA capping and splicing, non-coding RNA transcription termination and chromatin modification	15,25,28,52, 103,105,108, 110,133
Ser2 phosphorylation	Multiple repeats	• *S. cerevisiae* • *S. pombe* • *H. sapiens*	Transcription elongation, promoter-proximal pause and release, splicing, transcription termination and DNA topology	25,53,64,72, 76,88,102, 114,129,131
Ser7 phosphorylation	Multiple repeats	• *S. cerevisiae* • *S. pombe* • *H. sapiens*	snRNA expression, interaction with the Integrator complex and P-TEFb recognition	58,156,157
Thr4 phosphorylation	Multiple repeats	• *S. cerevisiae* • *S. pombe* • *Gallus gallus* • *H. sapiens*	Transcription elongation and termination, post-transcriptional splicing, processing of histone mRNA and chromatin remodelling	27,28,54,55
Tyr1 phosphorylation	Multiple repeats	• *S. cerevisiae* • *S. pombe* • *G. gallus* • *H. sapiens*	Inhibition of recruitment of transcription termination factors, CTD stability, antisense and enhancer transcription	49–51
Arg methylation	Arg1,810 of human RPB1	• *Mus musculus* • *H. sapiens*	snRNA and snoRNA regulation, R-loop resolution and transcription termination	9,14
Lys methylation	Lys7 in the non-consensus region of human CTD	• *H. sapiens* • *M. musculus* • *Drosophila melanogaster* • *Caenorhabditis elegans*	Supports nucleosome occupancy at promoters; negatively regulates gene expression	11,13
Lys acetylation	Lys7 in the non-consensus region of murine CTD; Lys7 in repeats 39, 42, 47 and 49 of human CTD	• *H. sapiens* • *M. musculus*	Induction of growth-factor response genes, transcription elongation; maintains balance between Lys methylation and acetylation and affects mRNA expression levels	10,11,13
O-GlcNAcylation	Ser5 and/or Ser7 in multiple repeats	*H. sapiens*	Pre-initiation complex assembly	7,8
Ubiquitylation	RPB1 Lys residues 859, 1866, 1873, 1887, 1908, 1922	*M. musculus*	RPB1 degradation	12

CTD, carboxy-terminal domain of RNA polymerase II; P-TEFb, positive transcription elongation factor B; RPB1, RNA polymerase B1; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA.

and phosphatases[2] throughout the transcription process. Thus, the unmodified CTD has been predominantly implicated in transcription initiation, whereas post-translationally modified CTD exists predominantly during transcription elongation[44,48].

CTD phosphorylation dynamics during transcription. In both budding yeast and humans, each of the heptapeptide residues Tyr1, Ser2, Thr4, Ser5 and Ser7 is phosphorylated at some point during transcription, and the relative abundance of these modifications changes along the gene[15,49–58] (FIG. 2a). The first modifications that were studied in detail were the phosphorylation of Ser5 and Ser2. These studies revealed that the levels of phosphorylated Ser5 (Ser5P) peak near the promoter and transcription start site (TSS) of genes, whereas Ser2P levels increase across gene bodies before decreasing sharply after the polyadenylation site (PAS)[15,53,59]. These patterns appear throughout the genome and are conserved across multiple species, including humans[2,6,55,60–62]. The phosphorylation patterns of each residue along protein-coding genes have been extensively characterized in yeast[28,51,60–63] and in mammalian cells[6,49,50,55] and are mostly similar across species (FIG. 2a). One exception is Tyr1P in budding yeast, which has a similar phosphorylation profile to that of Ser2P, although its levels decrease before the PAS. By contrast, Tyr1P levels in metazoans are highest near the promoter, and Tyr1P peaks are associated with paused Pol II in both the sense and antisense directions[49,50] (FIG. 2a). Broadly speaking, changes in the CTD phosphorylation state are coupled with transcription stage (FIG. 2b), with major changes in CTD phosphorylation occurring at transitions between transcription initiation, elongation and termination[6]. These phosphorylation patterns are not coincidental; they are crucial for the recruitment of key regulatory factors that control these transitions as well as the modification of chromatin and processing of the nascent RNA, as we discuss next.

Transcription elongation. The regulation of transcription elongation is strongly linked to the CTD phosphorylation state. In metazoans, a key regulatory event during elongation is promoter-proximal pausing, which occurs 20–100 bp downstream of the TSS, at the transition between early and productive elongation[64]. Pol II pausing is induced when Pol II is bound by negative elongation factor (NELF) and DRB sensitivity-inducing

Promoter-proximal pausing
In metazoans, highly regulated pausing of RNA polymerase II downstream of promoters, which controls the transition into productive transcription elongation.

Box 1 | **Mapping and quantifying CTD phosphorylation**

Much of the available data regarding the carboxy-terminal domain (CTD) code was obtained using antibodies that specifically recognize different phosphorylated residues of the CTD. Many monoclonal antibodies recognizing single, double or unphosphorylated CTD have been developed[6,105,154]. The best antibodies target a specific phosphorylated residue (for example, phosphorylated Ser5 (Ser5P)) within a diheptad, which is the minimal functional unit of the CTD[155]. Epitope masking, cross-reactivity and interference by other modifications can complicate the use of phospho-CTD-specific antibodies. Many CTD-specific antibodies have been characterized *in vitro* in an effort to understand their limitations[6,105], although the degree to which these limitations hold true *in vivo* is unclear. Nonetheless, these phospho-CTD-specific antibodies have been used with great success to examine CTD function. Chromatin immunoprecipitation (ChIP) with phospho-CTD-specific antibodies has been the method of choice for mapping levels of phospho-CTD modifications across individual genes and entire genomes[20]. Phospho-CTD-specific antibodies have also been used to monitor changes in CTD phosphorylation during transcription activation in live-cell imaging[154] and to purify distinct transcription complexes associated with each phospho-CTD modification[28].

The affinity and avidity differences between phospho-CTD-specific antibodies prevent comparison of phosphorylation levels between different CTD residues. Furthermore, it is not yet possible to use these antibodies to determine which heptapeptides are phosphorylated. To address these limitations by mass spectrometry, it is necessary to create genetically modified CTD constructs that enable trypsin digestion of the CTD, resulting in unique peptides whose location within the CTD molecule can be deduced from their mass[18,19]. In two studies, mass spectrometry analysis of modified CTDs revealed the relative abundance of each modification across the CTD, as well as similar trends in CTD phosphorylation between yeast and humans. Ser5P and Ser2P were the most abundant modifications in both yeast and humans, followed by Thr4P and Tyr1P. Ser7P exhibited the lowest level of phosphorylation, although Ser7P levels may have been artificially reduced by the genetic manipulation of the CTD required for mass spectrometry analysis. Interestingly, the CTD is phosphorylated in a relatively uniform manner with individual phospho-modifications being evenly distributed across the heptapeptides. Most phosphorylated repeats contain a single phospho-modification, and many repeats are entirely unphosphorylated[18,19]. The possibility that multiple phospho-modifications in close proximity could clash sterically or electrostatically might explain the observed dearth of phosphorylation. Another possibility is that unphosphorylated CTD repeats are important for CTD function throughout the transcription process.

factor (DSIF), which is a heterodimer of SPT4 and SPT5 (REFS 64–66). Positive transcription elongation factor B (P-TEFb)[64,66,67], which is composed of CDK9 and cyclin T, phosphorylates NELF, causing its dissociation, and phosphorylates DSIF, converting it into a positive transcription elongation factor, thereby triggering Pol II-pause release[68] (FIG. 2b). P-TEFb also phosphorylates Ser2, which is coincidental with promoter-proximal pause release[66,69], and also directly recruits several elongation and chromatin-modifying factors[70–73]. Recent work has expanded our knowledge of the regulation of this transition by identifying other key factors, such as bromodomain-containing protein 4 (BRD4), which is an atypical CTD kinase[74], and the transcription elongation complex RNA polymerase II-associated factor (PAF1)[75,76]. Much of this regulation is orchestrated by CTD kinases[66,74,77], emphasizing the importance of CTD phosphorylation in this transition. In addition to its role in promoter-proximal pausing, the phosphorylated CTD promotes the recruitment of the PAF1 complex and other transcription elongation factors, such as SPT6 and the human transcription elongation regulator 1 like (TCERG1)[28,72,78–81].

Although the roles of some CTD kinases are clear, the roles of other CTD kinases remain controversial. For example, *Drosophila melanogaster* P-TEFb is the major Ser2 kinase at the 5′ region of genes, whereas CDK12 is the major Ser2 kinase during elongation[82]. In human cells, two CDKs can phosphorylate the CTD during elongation: CDK12 and its homologue CDK13. CDK12 appears to be promiscuous and is able to phosphorylate Ser5, Ser7 and Ser2 *in vitro*[83], whereas CDK13 is able to phosphorylate Ser5 and Ser2 *in vitro*[84]. Adding to the complexity, knockdown or chemical inhibition of CDK12 or CDK13 has very little effect on global CTD phosphorylation *in vivo* and modest effects on gene expression[82,84–86]. Thus, the exact roles and interplay between P-TEFb, CDK12 and CDK13 remain elusive. Finally, although it is clear that CTD modifications affect the regulation of transcription elongation, it remains unknown whether the reverse is true. That is, whether transcription elongation also affects CTD modifications and why transitions in CTD state occur at specific stages of gene transcription.

Directing co-transcriptional processes. The phosphorylation state of the CTD couples Pol II to a suite of factors that regulate transcription elongation, mRNA processing, chromatin modification and even DNA topology[2,71,87,88] (TABLE 1). Consequently, the CTD contributes to co-transcriptional processes by granting regulatory factors intimate access to Pol II, the nascent RNA and chromatin. The influence of the CTD on RNA processing is facilitated by the positioning of the CTD proximal to the RNA exit channel[89,90]. To illustrate the importance of this proximity, the CTD can function when it is attached to Pol II subunits other than RPB1, but only in configurations that place the CTD near its original position[89]. Factors that are involved in numerous complexes and processes (reviewed in REFS 2,68) interact with the CTD, and new factors are continually being added to this list. For some co-transcriptional processes, the role of the CTD is well established, whereas for others the role of the CTD is emerging.

The CTD is important for the regulation of mRNA capping, which is the first step of mRNA processing. Ser5P levels peak early in transcription, near the TSS (FIG. 2a), and Ser5P contributes to the recruitment of the mRNA capping complex[15,52], thereby coupling RNA capping to early transcription elongation. Once recruited, the interaction of Ser5P with the capping complex allosterically aids in its activation[91]. Rapid capping of the transcript is crucial because downstream mRNA processing, including splicing and 3′ end cleavage and polyadenylation, require a properly capped transcript[92–95]. The crucial role of Ser5 in mRNA capping is highlighted by the lethal phenotype caused by Ser5-to-Ala substitutions in fission yeast (*Schizosaccharomyces pombe*), which can be overcome by tethering the capping complex to the CTD[25]. Thus, the essential role of Ser5 in fission yeast is the recruitment of the capping complex to Pol II.

Bromodomain-containing protein 4
(BRD4). A transcription regulation factor that binds to acetylated chromatin and has atypical carboxy-terminal domain kinase activity.

RNA polymerase II-associated factor 1
(PAF1). A transcription elongation complex that regulates both transcription and chromatin modification.

Another process that is regulated by the CTD is co-transcriptional splicing, which is tightly coupled to transcription elongation[87]. In budding yeast, splicing machineries are co-transcriptionally recruited: a majority of splicing occurs co-transcriptionally and splicing occurs almost immediately after the 3′ splice site is transcribed[28,96–106]. Taken together, these observations suggest that the recruitment of the splicing machinery to Pol II is highly regulated. In mammals, the phosphorylated CTD promotes splicing[101], and mutating Ser2 decreases spliceosome recruitment[102]. In yeast, the spliceosome interacts with Ser5P[28,103]. The recent advent of higher resolution chromatin immunoprecipitation (ChIP) techniques[28] and strand-specific high-resolution mapping of Pol II revealed dynamic remodelling of CTD phosphorylation at specific genomic loci[104,105]. In particular, there is strong modification of the CTD within a narrow region around splice sites[28,104,105]. In budding yeast, the 3′ splice site is associated with low levels of Thr4P and elevated levels of Ser5P and Ser7P[28,104]. In mammalian cells, Pol II containing high levels of Ser5P co-purified with splicing intermediates, suggesting that the spliceosome is associated with Ser5P during transcription[105]. Moreover, the ChIP profiles of Ser5P and U1 small nuclear ribonucleoprotein A, which is a component of the U1 spliceosomal subcomplex, are nearly identical at 3′ splice sites in budding yeast[28]. Given that Ser5P interacts with the spliceosome, these data suggest that elevated Ser5P levels contribute to the recruitment of the spliceosome near 3′ splice sites[28,103,105] and assist in the regulation of co-transcriptional splicing. Furthermore, Pol II pauses directly at splice sites in both yeast and humans[28,106], suggesting that Pol II elongation dynamics, CTD phosphorylation and co-transcriptional splicing are coupled.

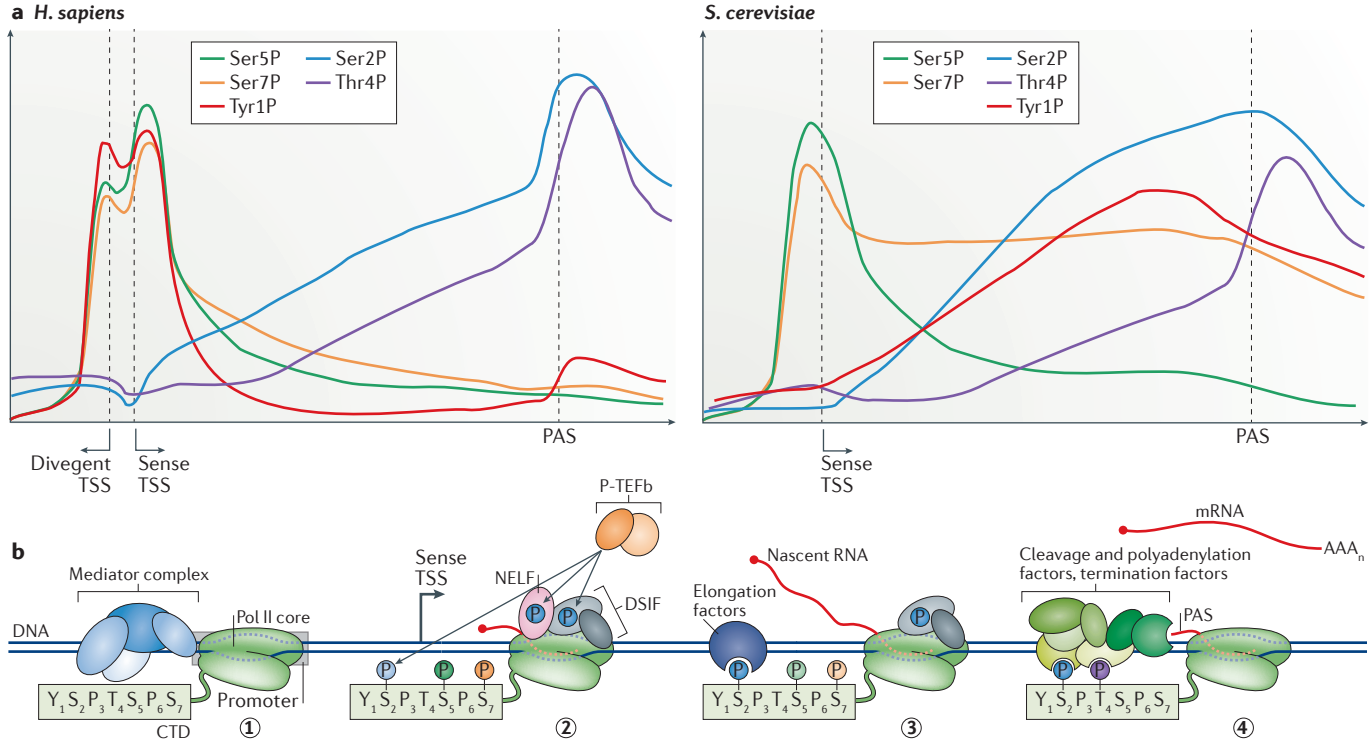

Figure 2 | **Transcription regulation by the CTD code. a** | Average chromatin immunoprecipitation (ChIP) profiles of phosphorylated residues of the carboxy-terminal domain (CTD) of RNA polymerase II (Pol II) across protein-coding genes in humans (*Homo sapiens*) and in budding yeast (*Saccharomyces cerevisiae*). The profiles of budding yeast and humans are highly similar, except for phosphorylated Tyr1 (Tyr1P). See the main text for details about the functions of each modification. **b** | Transcription in eukaryotes consists of several stages, including initiation, elongation and termination, which coincide with changes in the phosphorylation state of the CTD. Four steps of the transcription cycle are depicted with the corresponding CTD phosphorylation state and associated factors. Step 1 depicts the recruitment of the core Pol II enzyme to the promoter with an unphosphorylated CTD that interacts with the Mediator complex. The Mediator complex has a high affinity for unphosphorylated CTD and upon phosphorylation of Ser5 of the CTD by cyclin-dependent kinase 7 (CDK7; not shown), this affinity is lost and Pol II escapes from the promoter. Step 2 depicts Pol II during promoter-proximal pausing, which regulates the transition into productive transcription elongation in metazoans. At this step, Pol II is highly phosphorylated at Ser5 and Ser7, is paused downstream of the transcription start site (TSS) and is bound by negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). The arrival of positive elongation factor B (P-TEFb) leads to the phosphorylation of NELF, DSIF and Ser2. These phosphorylation events are followed by the release of NELF and the transition into productive transcription elongation. During productive elongation (step 3), the CTD contains lower levels of Ser5P and Ser7P and higher levels of Ser2P, which promotes the recruitment of many transcription elongation, chromatin-modifying and RNA-processing factors that regulate co-transcriptional processes. Finally, at step 4, Pol II transitions from transcription elongation to termination. Ser2P and Thr4P levels peak, thereby promoting the recruitment of cleavage and polyadenylation factors, as well as termination factors that release Pol II from the DNA. PAS, polyadenylation site. Part **a** is modified with permission from REF. 3. Part **b** is modified with permission from REF. 20 (http://pubs.acs.org/doi/full/10.1021/cr400071f).

Cryptic unstable transcripts (CUTs). Non-coding RNAs that are rapidly degraded by the nuclear exosome.

Chromatin modification and DNA topology. The CTD also contributes to the recruitment and regulation of histone modifiers. For example, Ser5P recruits complex proteins associated with Set1 (COMPASS) to the 5′ region of genes, thereby inducing methylation of histone 3 at Lys4 (H3K4me), which in turn recruits transcription elongation factors[107–110]. Ser5 and Ser2 phosphorylation is important for recruiting the H3K36 methyltransferase SET domain-containing protein 2 (SET2)[111–113]. H3K36 methylation is crucial for regulating transcription elongation and chromatin stability and for preventing intragenic transcription initiation, partly by recruiting and activating the histone deacetylase complex (HDAC) RPD3S[111,112,114–117]. The SET3 complex and HDACs such as RPD3S can also be directly recruited to Pol II by the CTD[118,119], as can the histone acetyltransferase (HAT) complexes NUA4 and SAGA[120,121]. These findings highlight the importance of transcription and CTD phosphorylation in the regulation of histone acetylation. However, as both HDACs and HATs can be recruited to Pol II, the exact mechanisms by which the CTD, and more generally transcription, regulate histone acetylation levels across gene bodies require further exploration. Recent work has shown that the CTD can also control DNA topology during transcription elongation by recruiting and stimulating topoisomerase I through Ser2 phosphorylation[88]. This work expands the known roles of the CTD, strengthening the link between transcription and DNA structure and topology[122–128].

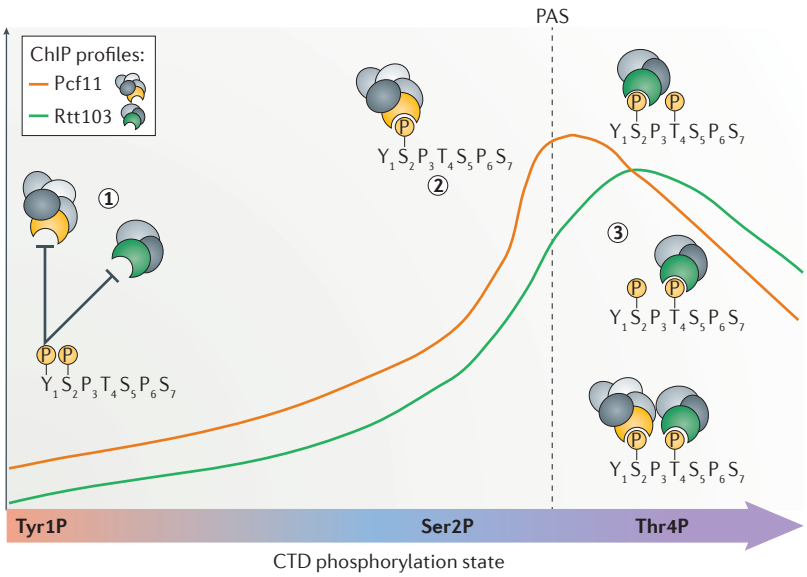

Figure 3 | **Regulation of transcription termination in *Saccharomyces cerevisiae* by the CTD.** Shades along the *x* axis represent the phosphorylation state of the carboxy-terminal domain (CTD) of RNA polymerase II (Pol II) around the polyadenylation site (PAS) of protein-coding genes. Shown are the chromatin immunoprecipitation (ChIP) profiles of the cleavage and polyadenylation complex component protein 1 of cleavage and polyadenylation factor I (Pcf11) and the transcription termination factor regulator of Ty transposition protein 103 (Rtt103). Three positions along the gene body (steps 1–3) depict the phosphorylation state of the CTD and its interactions with Pcf11 and Rtt103, together suggesting a model for the precise recruitment of Pcf11 and Rtt103 during transcription. During transcription elongation (step 1), phosphorylated Tyr1 (Tyr1P) prevents the recruitment of Pcf11 and Rtt103 to Pol II. Loss of Tyr1P and elevated Ser2P levels (step 2) recruit Pcf11 to Pol II for cleavage and polyadenylation of the nascent transcript. Ser2P and Thr4P recruit Rtt103 (step 3) and transition Pol II into the termination stage. Because Rtt103 can bind to both Ser2P and Thr4P, three possible states of CTD-bound Rtt103 are represented downstream of the PAS. In one state, Rtt103 binds to Ser2P, and in another, Rtt103 binds to Thr4P. In the third state, both Pcf11 and Rtt103 may be co-recruited by binding to different CTD modifications to coordinate cleavage and polyadenylation with transcription termination.

Transcription termination. The CTD is associated with transcription termination and 3′ end processing of coding genes in both yeast and humans. In yeast, Ser2P binds directly to members of the cleavage and polyadenylation and termination machinery, such as protein 1 of cleavage and polyadenylation factor I (Pcf11) and regulator of Ty1 transposition protein 103 (Rtt103)[129–132]. However, Ser2P is not solely responsible for the recruitment of these factors, because although Ser2P levels are high across gene bodies (FIG. 2), Pcf11 and Rtt103 levels are highly concentrated at the 3′ ends of genes[51]. Tyr1P is a likely candidate for some of the observed specificity. *In vitro*, Tyr1P prevents the binding of termination factors to the CTD[51], and Tyr1P levels decrease before PAS in yeast cells. However, overall Tyr1P is detected at much lower levels than Ser2P[18,19], suggesting that the tight recruitment window of termination factors to regions near the PAS probably requires additional CTD modifications. Indeed, Thr4P levels were recently demonstrated to peak downstream of the PAS (FIG. 2), and Thr4P interacts with Rtt103 (REF. 28). Thus, three CTD modifications are implicated in the recruitment of transcription termination factors to Pol II: first, Ser2P binds to and presumably helps to recruit Rtt103 and Pcf11; second, Thr4P binds to Rtt103; and third, Tyr1P prevents the binding of Rtt103 and Pcf11 *in vitro*[28,51] (FIG. 3). These CTD modifications, along with sequences in the RNA, are likely to collaborate to regulate transcription termination and cleavage and polyadenylation in yeast, illustrating the complexity of the CTD code.

In budding yeast, Ser5 phosphorylation is associated with transcription termination of non-coding RNAs through the NNS complex, which is responsible for terminating the transcription of small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and cryptic unstable transcripts (CUTs)[133,134]. The NNS complex consists of the RNA-binding proteins nuclear pre-mRNA downregulation protein 1 (Nrd1) and nuclear polyadenylated RNA-binding protein 3 (Nab3) and the helicase splicing endonuclease 1 (Sen1)[135]. The NNS complex is partly recruited to Pol II through binding of Nrd1 to Ser5P[136]. Transcription termination by the NNS complex is tightly coupled to RNA processing because the NNS complex interacts with the nuclear exosome to process snRNAs and snoRNAs and rapidly degrades CUTs[137].

Other CTD residues are associated with the termination of multiple classes of transcript. For example, Ser7P was implicated in regulating 3′ processing at snRNA genes in mammals[58], although the exact mechanism is unclear, and Thr4P was implicated in 3′ processing of histone genes in vertebrates[54]. Thus, the CTD is linked to termination and 3′ end processing of both coding and non-coding transcripts from yeast to mammals.

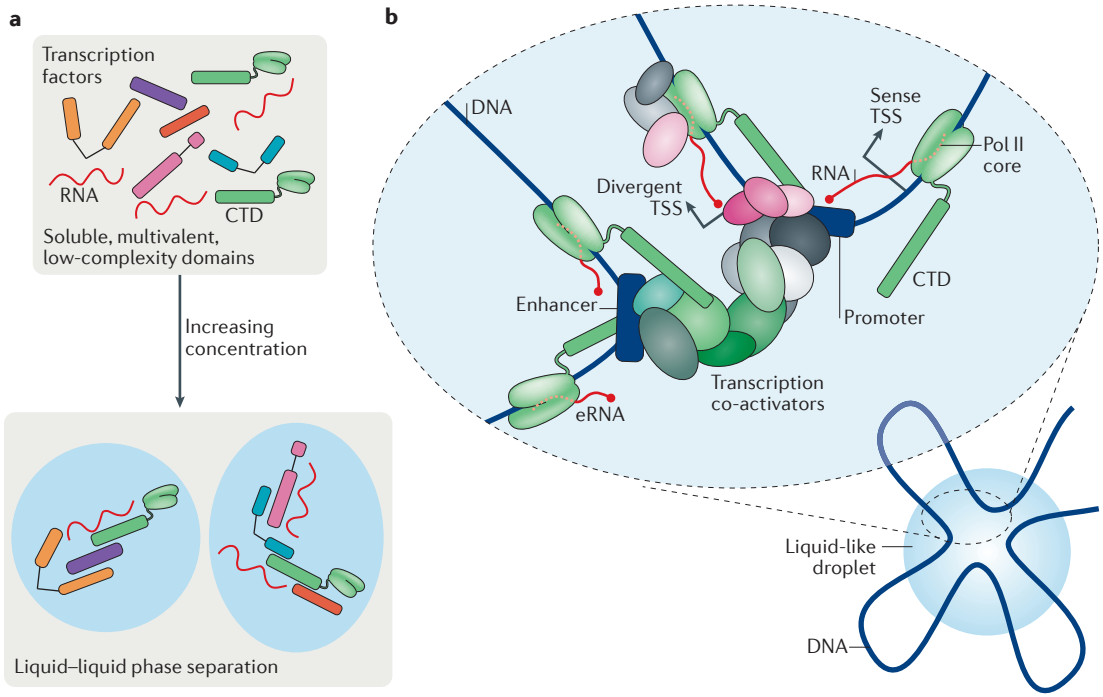

Figure 4 | **Formation of compartments by liquid–liquid phase-separation. a** | Many biological molecules such as proteins and RNA contain unstructured, disordered regions or regions of low complexity. These regions are capable of forming multivalent interactions with one another and, upon increases in their local concentration, can form liquid–liquid phase-separated compartments that behave like liquid-like droplets to regulate biological processes. The carboxy-terminal domain (CTD) of RNA polymerase II (Pol II), transcription factors and RNA all possess low-complexity domains or regions, suggesting that these molecules may be able to form liquid-like droplets. **b** | Do liquid-like droplets form at promoters? The formation of liquid–liquid phase-separated droplets can be nucleated by high local concentrations of intrinsically disordered, low-complexity biological molecules such as RNA and low-complexity protein domains. Multivalent interactions between these molecules form liquid-like droplets, which can act as cellular compartments. Near promoters, the high levels of Pol II complexes and thus of their low-complexity CTDs, which support sense, divergent antisense and enhancer transcription, along with the nascent transcripts and transcription factors replete with low-complexity domains, may nucleate the formation of liquid-like droplets. These droplets could regulate transcription by further concentrating Pol II and transcription factors near promoters. eRNA, enhancer RNA; TSS, transcription start site.

The CTD in liquid–liquid phase separation

Many biological processes are partly regulated by liquid–liquid phase separations, which exploit multivalent interactions between low-complexity protein domains and RNA[31,138–142] (FIG. 4a). Low-complexity domains and intrinsically disordered domains and proteins form liquid-like droplets throughout the cell[141], and these droplets act as compartments that trap and localize proteins[140,141]. Examples of such droplets include Cajal bodies, nucleoli and P bodies[140,141]. Recent *in vitro* studies have established that the Pol II CTD can be incorporated into and form liquid phase-separated compartments and hydrogels, revealing potential roles for the CTD beyond factor recruitment[33,34]. Two transcription-related proteins that contain low-complexity domains, the fused in sarcoma (FUS) protein and the general transcription factor TATA-binding protein-associated factor 2N (TAF2N; also known as TAF15), enabled the phase separations observed in these studies. This result suggests another mechanism for gene activation, whereby Pol II is recruited to promoters through the incorporation of the CTD into liquid phase-separated compartments that are formed by low-complexity domains of transcription factors[33,34].

TAF15 hydrogels that were formed *in vitro* could interact with the CTD in a manner that depended on its length and phosphorylation state: a CTD with only five heptapeptide repeats was not incorporated into the hydrogels, whereas CTDs with ten or more repeats were efficiently incorporated[33]. Notably, the minimal CTD length for viability in yeast is eight to ten repeats, the size required for efficient hydrogel incorporation. Phosphorylation also modulated CTD incorporation into the TAF15 hydrogels, whereby highly phosphorylated CTD was not retained. These results suggest a potential Pol II recruitment and transcription initiation mechanism, whereby Pol II with an unphosphorylated CTD is partly recruited to promoters via its incorporation into a liquid–liquid phase-separated compartment, and is released from this compartment into transcription elongation following phosphorylation of the CTD. Moreover, the degenerate carboxy-terminal region of the CTD (found in mammals) was more efficiently retained by

Cajal bodies
Membrane-less nuclear compartments comprising mainly RNA and protein and where RNA processing and maturation occurs.

P bodies
Membrane-less cytoplasmic compartments where several RNA processing events occur.

Hydrogels
Macromolecular polymers composed of a network of multivalent interactions.

TAF15 hydrogels than the conserved amino-terminal region where the Mediator complex binds[41,42]. Thus, the recruitment of Pol II to promoters through CTD hydrogel formation would not interfere with Mediator binding; hence, both liquid–liquid phase separation and the Mediator complex could contribute to Pol II recruitment. The CTD is also incorporated into FUS hydrogels *in vitro*[33,34], and mutations in *FUS* that alter the capacity of the protein to form hydrogels also hamper its capacity to activate gene expression[33]. The CTD can nucleate liquid-like phase separation of the FUS low-complexity domain *in vitro*[34]; FUS also associates with the CTD *in vivo*, and FUS depletion prevents inappropriate hyperphosphorylation of Ser2 (REF. 143). However, there is currently no evidence to indicate that FUS and the CTD engage in phase separation *in vivo*. Moreover, as the CTD-induced TAF15 and FUS phase-separation experiments were performed *in vitro*, it remains unclear whether and to what extent they occur *in vivo*. Nevertheless, it is possible that at least under some conditions, cells can take advantage of the unique physicochemical properties of the Pol II CTD to regulate transcription and co-transcriptional processes.

Given that the CTD and transcription-related factors can be incorporated into liquid-like droplets, it is intriguing to speculate that the CTD is important for liquid-like phase separation during transcription initiation and other stages of transcription *in vivo*. A recent study demonstrated the essential function of Thr4 during mitosis, possibly by interacting with other low-complexity domain proteins and forming liquid-like droplets[144]. Clusters of Pol II, whereby numerous Pol II molecules are localized to a specific nuclear region, have been observed during transcription[145–147]. Interestingly, the existence of these clusters is linked to transcriptional output as cluster lifetimes correlate with nascent RNA production[146]. Moreover, activation of transcription at the heat shock loci in *D. melanogaster* results in the formation of a transcription 'compartment', where Pol II molecules are recycled during the transcription process[147,148]. Intriguingly, these compartments require active poly(ADP-ribose) polymerase (PARP). The polymer product of PARP is structurally similar to RNA and can nucleate liquid–liquid phase separation in the presence of intrinsically unstructured domains or repeated domains[149,150]. Thus, an interesting possibility is that these compartments are the result of *in vivo* phase separation at the heat shock loci. Although the formation and composition of these transient transcription clusters remain largely elusive, the abundance of low-complexity domains among transcription factors[138,139], and the presence of the CTD, the nascent RNA and probably other polymers in the clusters, indicate that these compartments could be liquid-like droplets formed by a phase separation. A compelling possibility is that some of these clusters are nucleated by the colocalization of multiple enhancers with their target promoter. Enhancers and promoters are bidirectionally transcribed, which could lead to high local concentration of Pol II, transcription factors and RNA, thereby contributing to the formation of liquid-like droplets[64,146,151–153] (FIG. 4b). Thus, an interesting hypothesis is that enhancers and antisense transcription may partially regulate transcription through nucleating droplet formation.

It remains unexplored whether the CTD can be involved in the formation of liquid-like droplets after transcription initiation. Many liquid-like droplets contain RNA, and nascent RNA might encourage the formation of such droplets with the CTD or other transcription and RNA-processing factors. *In vitro,* the CTD must be unphosphorylated to efficiently associate with TAF15 hydrogels[33,143]. However, given that FUS has been shown to regulate CTD phosphorylation *in vivo* it is possible that the CTD and other regulatory factors may phase separate in a way that tolerates or even requires CTD phosphorylation[33]. This could be induced perhaps by recruiting factors that regulate liquid–liquid phase separation or by increasing the hydrogen-bonding potential of the CTD. Furthermore, even during transcription elongation, regions of individual CTDs will be unphosphorylated or minimally phosphorylated (BOX 1), potentially allowing multiple modes of interaction with liquid-like droplets[18,19]. Clearly, more studies are required to investigate these hypotheses; nevertheless, these recent observations certainly imply that the CTD has roles beyond that of a simple landing pad for transcription-associated factors.

Concluding remarks

The presence of multiple heptapeptide repeats and numerous potential PTMs in the CTD implied the existence of a vast number of possible CTD functional states. However, the combinatorial complexity of the CTD code is limited by the fact that on average, each repeat contains less than one phosphorylated residue[18,19] (BOX 1). Despite this considerable reduction in complexity, the number of possible states is still large; consequently, our understanding of the code remains incomplete. Many questions remain regarding how the CTD is spatiotemporally modified, as well as how various modifications work together to promote or block association of regulatory factors. Furthermore, the distribution of PTMs across the many heptapeptide repeats of individual CTD molecules and the number of factors that occupy an individual CTD remain unknown. To address these questions, it will be necessary to analyse single CTD molecules.

Another challenge is to reconcile how the CTD serves as a protein landing pad, with its propensity to form separated liquid droplets and/or hydrogels. Answering these questions will require *in vivo* assays capable of monitoring liquid–liquid phase separation at small scales. In addition, *in vitro* studies must be performed to determine how the physicochemical properties of phosphorylated and non-phosphorylated CTDs relate to their interactions with regulatory molecules. These studies will shed new light on the roles of the CTD, improve our understanding of its function beyond the CTD code and enable a deeper appreciation of transcriptional regulation.

1. Chapman, R. D., Heidemann, M., Hintermair, C. & Eick, D. Molecular evolution of the RNA polymerase II CTD. *Trends Genet.* **24**, 289–296 (2008).
2. Jeronimo, C., Bataille, A. R. & Robert, F. The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem. Rev.* **113**, 8491–8522 (2013).
This review lists and categorizes all known CTD-interacting proteins.
3. Jeronimo, C., Collin, P. & Robert, F. The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *J. Mol. Biol.* **428**, 2607–2622 (2016).
4. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
5. Corden, J. L. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chem. Rev.* **113**, 8423–8455 (2013).
6. Heidemann, M., Hintermair, C., Voß, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta* **1829**, 55–62 (2013).
7. Lu, L. *et al.* Distributive O-GlcNAcylation on the highly repetitive C-terminal domain of RNA polymerase II. *Biochemistry* **55**, 1149–1158 (2016).
8. Ranuncolo, S. M., Ghosh, S., Hanover, J. A., Hart, G. W. & Lewis, B. A. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription *in vitro* and *in vivo*. *J. Biol. Chem.* **287**, 23549–23561 (2012).
9. Sims, R. J. *et al.* The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**, 99–103 (2011).
10. Schröder, S. *et al.* Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Mol. Cell* **52**, 314–324 (2013).
11. Voss, K. *et al.* Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. *Transcription* **6**, 91–101 (2015).
12. Li, H. *et al.* Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol. Cell. Biol.* **27**, 5296–5305 (2007).
13. Dias, J. D. *et al.* Methylation of RNA polymerase II non-consensus lysine residues marks early transcription in mammalian cells. *eLife* **4**, e11215 (2015).
14. Zhao, D. Y. *et al.* SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**, 48–53 (2016).
This study defines a direct role for Arg methylation at the CTD in transcription termination.
15. Komarnitsky, P., Cho, E. J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460 (2000).
This work demonstrates different CTD phosphorylation states during transcription and introduces the CTD code.
16. Egloff, S. & Murphy, S. Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**, 280–288 (2008).
17. Buratowski, S. The CTD code. *Nat. Struct. Biol.* **10**, 679–680 (2003).
18. Suh, H. *et al.* Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. *Mol. Cell* **61**, 297–304 (2016).
This work determines the abundance and composition of CTD phosphorylation states *in vivo* in *S. cerevisiae*.
19. Schüller, R. *et al.* Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol. Cell* **61**, 305–314 (2016).
This work determines the abundance and composition of CTD phosphorylation *in vivo* in human cells.
20. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).
21. Chapman, A. B. & Agabian, N. *Trypanosoma brucei* RNA polymerase II is phosphorylated in the absence of carboxyl-terminal domain heptapeptide repeats. *J. Biol. Chem.* **269**, 4754–4760 (1994).
22. West, M. L. & Corden, J. L. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223–1233 (1995).
This publication describes the number of CTD repeats required for yeast viability and wild-type growth.
23. Nonet, M. L. & Young, R. A. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* **123**, 715–724 (1989).
24. Morrill, S. A., Exner, A. E., Babokhov, M., Reinfeld, B. I. & Fuchs, S. M. DNA instability maintains the repeat length of the yeast RNA polymerase II C-terminal domain. *J. Biol. Chem.* **291**, 11540–11550 (2016).
25. Schwer, B. & Shuman, S. Deciphering the RNA polymerase II CTD code in fission yeast. *Mol. Cell* **43**, 311–318 (2011).
26. Schwer, B., Bitton, D. A., Sanchez, A. M., Bähler, J. & Shuman, S. Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast. *Proc. Natl Acad. Sci. USA* **111**, 4185–4190 (2014).
27. Rosonina, E. *et al.* Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc. Natl Acad. Sci. USA* **111**, 11924–11931 (2014).
28. Harlen, K. M. *et al.* Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep.* **15**, 2147–2158 (2016).
This work determines that Pol II with different CTD phosphorylation states interacts with unique sets of proteins.
29. Cassart, C., Drogat, J., Migeot, V. & Hermand, D. Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast. *Transcription* **3**, 231–234 (2012).
30. Schwer, B., Sanchez, A. M. & Shuman, S. Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proc. Natl Acad. Sci. USA* **109**, 18024–18029 (2012).
31. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid–liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
32. Mitrea, D. M. & Kriwacki, R. W. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal.* **14**, 1 (2016).
33. Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
This study demonstrates that the CTD can form hydrogels and liquid-like droplets *in vitro*.
34. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-residue view of *in vitro* FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell* **60**, 231–241 (2015).
35. Schüller, R. & Eick, D. Getting access to low-complexity domain modifications. *Trends Biochem. Sci.* **41**, 894–897 (2016).
36. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
37. Jasnovidova, O. & Stefl, R. The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip. Rev. RNA* **4**, 1–16 (2013).
38. Zaborowska, J., Egloff, S. & Murphy, S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23**, 771–777 (2016).
39. Myers, L. C. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45–54 (1998).
40. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl Acad. Sci. USA* **88**, 10004–10008 (1991).
This publication describes how unphosphorylated CTD is associated with the transcription pre-initiation complex.
41. Robinson, P. J. *et al.* Structure of a complete mediator-RNA polymerase II pre-initiation complex. *Cell* **166**, 1411–1422.e16 (2016).
42. Robinson, P. J. J., Bushnell, D. A., Trnka, M. J., Burlingame, A. L. & Kornberg, R. D. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc. Natl Acad. Sci. USA* **109**, 17931–17935 (2012).
43. Hahn, S. & Young, E. T. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705–736 (2011).
44. Payne, J. M., Laybourn, P. J. & Dahmus, M. E. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J. Biol. Chem.* **264**, 19621–19629 (1989).
45. Wong, K. H., Jin, Y. & Struhl, K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Mol. Cell* **54**, 601–612 (2014).
46. Jeronimo, C. & Robert, F. Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.* **21**, 449–455 (2014).
References 45 and 46 reveal that phosphorylation of the CTD by TFIIH dissociates the Mediator complex and induces promoter escape by Pol II.
47. Søgaard, T. M. M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282**, 14113–14120 (2007).
48. Cadena, D. L. & Dahmus, M. E. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**, 12468–12474 (1987).
49. Hsin, J.-P., Li, W., Hoque, M., Tian, B. & Manley, J. L. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife* **3**, e02112 (2014).
50. Descostes, N. *et al.* Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* **3**, e02105 (2014).
51. Mayer, A. *et al.* CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**, 1723–1725 (2012).
52. Schroeder, S. C., Schwer, B., Shuman, S. & Bentley, D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14**, 2435–2440 (2000).
53. Licatalosi, D. D. *et al.* Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. *Mol. Cell* **9**, 1101–1111 (2002).
54. Hsin, J.-P., Sheth, A. & Manley, J. L. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. *Science* **334**, 683–686 (2011).
55. Hintermair, C. *et al.* Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J.* **31**, 2784–2797 (2012).
56. McCracken, S. *et al.* 5′-capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306–3318 (1997).
This publication describes the binding of an RNA capping protein to a specific phosphorylation state of the CTD.
57. Kim, M., Suh, H., Cho, E.-J. & Buratowski, S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J. Biol. Chem.* **284**, 26421–26426 (2009).
58. Egloff, S. *et al.* Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777–1779 (2007).
59. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
This work demonstrates that the CTD is important for coupling transcription to co-transcriptional RNA processing.
60. Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278 (2010).
This study identifies how general transitions in the transcription process are coupled to changes in CTD phosphorylation.
61. Kim, H. *et al.* Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1279–1286 (2010).
62. Tietjen, J. R. *et al.* Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1154–1161 (2010).
63. Bataille, A. R. *et al.* A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol. Cell* **45**, 158–170 (2012).
64. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
65. Wu, C.-H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414 (2003).
66. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).

67. Ni, Z. *et al.* P-TEFb is critical for the maturation of RNA polymerase II into productive elongation *in vivo*. *Mol. Cell. Biol.* **28**, 1161–1170 (2008).
68. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
69. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3**, e02407 (2014).
70. Li, J., Moazed, D. & Gygi, S. P. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.* **277**, 49383–49388 (2002).
71. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
72. Sun, M., Larivière, L., Dengl, S., Mayer, A. & Cramer, P. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J. Biol. Chem.* **285**, 41597–41603 (2010).
73. Fuchs, S. M., Laribee, R. N. & Strahl, B. D. Protein modifications in transcription elongation. *Biochim. Biophys. Acta* **1789**, 26–36 (2009).
74. Devaiah, B. N. *et al.* BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc. Natl Acad. Sci. USA* **109**, 6927–6932 (2012).
75. Chen, F. X. *et al.* PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015 (2015).
76. Yu, M. *et al.* RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).
77. Rodrĩguez-Molina, J. B., Tseng, S. C., Simonett, S. P., Taunton, J. & Ansari, A. Z. Engineered covalent inactivation of TFIIH-kinase reveals an elongation checkpoint and results in widespread mRNA stabilization. *Mol. Cell* **63**, 433–444 (2016).
78. Liu, J., Fan, S., Lee, C.-J., Greenleaf, A. L. & Zhou, P. Specific interaction of the transcription elongation regulator TCERG1 with RNA polymerase II requires simultaneous phosphorylation at Ser2, Ser5, and Ser7 within the carboxyl-terminal domain repeat. *J. Biol. Chem.* **288**, 10890–10901 (2013).
79. Qiu, H., Hu, C., Gaur, N. A. & Hinnebusch, A. G. Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J.* **31**, 3494–3505 (2012).
80. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* **21**, 160–174 (2007).
81. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434 (2008).
82. Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316 (2010).
83. Bösken, C. A. *et al.* The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* **5**, 3505 (2014).
84. Liang, K. *et al.* Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol. Cell. Biol.* **35**, 928–938 (2015).
85. Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **25**, 2158–2172 (2011).
86. Bartkowiak, B., Yan, C. & Greenleaf, A. L. Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. *Biochim. Biophys. Acta* **1849**, 1179–1187 (2015).
87. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
88. Baranello, L. *et al.* RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371 (2016).
89. Suh, H., Hazelbaker, D. Z., Soares, L. M. & Buratowski, S. The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. *Mol. Cell* **51**, 850–858 (2013).
90. Chen, C.-Y., Chang, C.-C., Yen, C.-F., Chiu, M. T.-K. & Chang, W.-H. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl Acad. Sci. USA* **106**, 127–132 (2009).
91. Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* **3**, 405–411 (1999).
92. Cooke, C. & Alwine, J. C. The cap and the 3′splice site similarly affect polyadenylation efficiency. *Mol. Cell. Biol.* **16**, 2579–2584 (1996).
93. Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W. & Gilmartin, G. M. Participation of the nuclear cap binding complex in pre-mRNA 3′ processing. *Proc. Natl Acad. Sci. USA* **94**, 11893–11898 (1997).
94. Schwer, B. & Shuman, S. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing *in vivo*. *RNA* **2**, 574–583 (1996).
95. Topisirovic, I., Svitkin, Y. V. & Sonenberg, N. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* **2**, 277–298 (2011).
96. Carrillo Oesterreich, F. *et al.* Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**, 372–381 (2016).
97. Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
98. Tardiff, D. F., Lacadie, S. A. & Rosbash, M. A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol. Cell* **24**, 917–929 (2006).
99. Moore, M. J., Schwartzfarb, E. M., Silver, P. A. & Yu, M. C. Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. *Mol. Cell* **24**, 903–915 (2006).
100. Carrillo Oesterreich, F., Preibisch, S. & Neugebauer, K. M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* **40**, 571–581 (2010).
101. David, C. J., Boyne, A. R., Millhouse, S. R. & Manley, J. L. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* **25**, 972–983 (2011).
102. Gu, B., Eick, D. & Bensaude, O. 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 (2013).
103. Morris, D. P. & Greenleaf, A. L. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275**, 39935–39943 (2000).
104. Milligan, L. *et al.* Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol. Syst. Biol.* **12**, 874 (2016).
105. Nojima, T. *et al.* Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
106. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
107. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11**, 709–719 (2003).
108. Krogan, N. J. *et al.* The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* **11**, 721–729 (2003).
109. Lee, J.-H. & Skalnik, D. G. Wdr82 Is a C-terminal domain-binding protein that recruits the Setd1A histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.* **28**, 609–618 (2008).
110. Terzi, N., Churchman, L. S., Vasiljeva, L., Weissman, J. & Buratowski, S. H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1–Nab3–Sen1 pathway. *Mol. Cell. Biol.* **31**, 3569–3583 (2011).
111. Li, B. *et al.* The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **278**, 8897–8903 (2003).
This work reveals that co-transcriptional histone methylation is coupled to transcription by the CTD.
112. Xiao, T. *et al.* Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* **17**, 654–663 (2003).
113. Schaft, D. *et al.* The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. *Nucleic Acids Res.* **31**, 2475–2482 (2003).
114. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **23**, 4207–4218 (2003).
115. Kizer, K. O. *et al.* A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* **25**, 3305–3316 (2005).
116. Vojnic, E., Simon, B., Strahl, B. D., Sattler, M. & Cramer, P. Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *J. Biol. Chem.* **281**, 13–15 (2006).
117. Smolle, M. & Workman, J. L. Transcription-associated histone modifications and cryptic transcription. *Biochim. Biophys. Acta* **1829**, 84–97 (2013).
118. Govind, C. K. *et al.* Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol. Cell* **39**, 234–246 (2010).
119. Drouin, S. *et al.* DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS Genet.* **6**, e1001173 (2010).
120. Ginsburg, D. S., Govind, C. K. & Hinnebusch, A. G. NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. *Mol. Cell. Biol.* **29**, 6473–6487 (2009).
121. Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K. & Hinnebusch, A. G. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol. Cell* **25**, 31–42 (2007).
122. Jeppsson, K., Kanno, T., Shirahige, K. & Sjögren, C. The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nat. Rev. Mol. Cell Biol.* **15**, 601–614 (2014).
123. Kouzine, F., Levens, D. & Baranello, L. DNA topology and transcription. *Nucleus* **5**, 195–202 (2014).
124. Durand-Dubief, M., Svensson, J. P., Persson, J. & Ekwall, K. Topoisomerases, chromatin and transcription termination. *Transcription* **2**, 66–70 (2011).
125. King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58–62 (2013).
126. Kanagaraj, R. *et al.* RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. *Nucleic Acids Res.* **38**, 8131–8140 (2010).
127. Islam, M. N., Fox, D. III, Guo, R., Enomoto, T. & Wang, W. RecQL5 promotes genome stabilization through two parallel mechanisms — interacting with RNA polymerase II and acting as a helicase. *Mol. Cell. Biol.* **30**, 2460–2472 (2010).
128. Li, M., Xu, X. & Liu, Y. The SET2–RPB1 interaction domain of human RECQ5 is important for transcription-associated genome stability. *Mol. Cell. Biol.* **31**, 2090–2099 (2011).
129. Lunde, B. M. *et al.* Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **17**, 1195–1201 (2010).
130. Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517–522 (2004).
131. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* **430**, 223–226 (2004).
132. Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* **13**, 67–76 (2004).
This publication describes how the specific phosphorylation of CTD Ser2 couples transcription to mRNA cleavage and polyadenylation.
133. Arigo, J. T., Eyler, D. E., Carroll, K. L. & Corden, J. L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23**, 841–851 (2006).
134. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* **23**, 853–864 (2006).
135. Carroll, K. L., Ghirlando, R., Ames, J. M. & Corden, J. L. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* **13**, 361–373 (2007).

136. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15**, 795–804 (2008).
137. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202 (2015).
138. Radó-Trilla, N. & Albà, M. Dissecting the role of low-complexity regions in the evolution of vertebrate proteins. *BMC Evol. Biol.* **12**, 155 (2012).
139. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, 43 (2010).
140. Berry, J., Weber, S. C., Vaidya, N., Haataja, M. & Brangwynne, C. P. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl Acad. Sci. USA* **112**, E5237–E5245 (2015).
141. Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *EMBO J.* **35**, 1603–1612 (2016).
142. Xiang, S. *et al.* The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**, 829–839 (2015).
143. Schwartz, J. C. *et al.* FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* **26**, 2690–2695 (2012).
144. Hintermair, C. *et al.* Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Sci. Rep.* **6**, 27401 (2016).
145. Cisse, I. I. *et al.* Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).
146. Cho, W.-K. *et al.* RNA polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, e13617 (2016).
147. Yao, J., Ardehali, M. B., Fecko, C. J., Webb, W. W. & Lis, J. T. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol. Cell* **28**, 978–990 (2007).
148. Zobeck, K. L., Buckley, M. S., Zipfel, W. R. & Lis, J. T. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* **40**, 965–975 (2010).
149. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015).
150. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
151. Liu, Z. *et al.* 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *eLife* **3**, e04236 (2014).
152. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
153. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
154. Stasevich, T. J. *et al.* Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* **516**, 272–275 (2014).
155. Stiller, J. W. & Cook, M. S. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot. Cell* **3**, 735–740 (2004).
This work describes how two heptapeptide repeats constitute the functional unit of the CTD.
156. Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780–1782 (2007).
157. Czudnochowski, N., Bösken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, 842 (2012).

Acknowledgements
The authors thank S. Buratowski, A. Mayer, C. Patil, B. Portz and F. Winston for their critical reading of the manuscript. This work was supported by US National Institutes of Health National Human Genome Research Institute (NHGRI) grants R01HG007173 (to L.S.C.) and a Burroughs Wellcome Fund Career Award at the Scientific Interface (to L.S.C.). K.M.H. was supported by a US National Science Foundation Graduate Research Fellowship (DGE1144152).

Competing interests statement
The authors declare no competing interests.