Uniform transitions of the general RNA polymerase II transcription complex

Andreas Mayer[1,2], Michael Lidschreiber[1,2], Matthias Siebert[1,2], Kristin Leike[1], Johannes Söding[1] & Patrick Cramer[1]

We present genome-wide occupancy profiles for RNA polymerase (Pol) II, its phosphorylated forms and transcription factors in proliferating yeast. Pol II exchanges initiation factors for elongation factors during a 5′ transition that is completed 150 nucleotides downstream of the transcription start site (TSS). The resulting elongation complex is composed of all the elongation factors and shows high levels of Ser7 and Ser5 phosphorylation on the C-terminal repeat domain (CTD) of Pol II. Ser2 phosphorylation levels increase until 600–1,000 nucleotides downstream of the TSS and do not correlate with recruitment of Spt6 and Pcf11, which bind the Ser2-phosphorylated CTD *in vitro*. This indicates CTD-independent recruitment mechanisms and CTD masking *in vivo*. Elongation complexes are productive and disassemble in a two-step 3′ transition. Paf1, Spt16 (part of the FACT complex), and the CTD kinases Bur1 and Ctk1 exit upstream of the polyadenylation site, whereas Spt4, Spt5, Spt6, Spn1 (also called Iws1) and Elf1 exit downstream. Transitions are uniform and independent of gene length, type and expression.

Gene transcription begins with the assembly of Pol II and its initiation factors on promoter DNA. Pol II then starts mRNA synthesis and exchanges initiation factors for elongation factors, which are required for chromatin passage and RNA processing[1–3]. Whereas Pol II is unphosphorylated during initiation, it is phosphorylated at its CTD during elongation. The CTD is phosphorylated at Ser5 residues in the 5′ region of a gene and at Ser2 residues in the 3′ region[4,5]. The phosphorylated CTD recruits elongation factors to ensure cotranscriptional RNA processing and chromatin modification[2,3,6–9]. A genome-wide study has shown that initiation factors are present at all active Pol II gene promoters[10], but it is unknown whether all elongation factors are recruited to all active genes, and no genome-wide studies have examined whether factor recruitment correlates with specific Pol II phosphorylations.

To address these questions, we used high-resolution genome-wide occupancy profiling by chromatin immunoprecipitation (ChIP) of Pol II, its phosphorylated forms, its elongation factors and components of the Pol II initiation and termination machinery in proliferating yeast cells. Statistical analysis provides strong evidence for a general elongation complex—that is, one composed of all elongation factors—that mediates chromatin transcription and mRNA processing at all Pol II genes. The general elongation complex is apparently established during a 5′ transition within a narrow window just downstream of the TSS, and it is disassembled in two major steps during a 3′ transition around the polyadenylation (pA) site. The results also show that CTD phosphorylation patterns previously observed at individual genes occur globally and that levels of CTD phosphorylation do not

correlate with the *in vivo* occupancy of two factors that bind the phosphorylated CTD *in vitro*. General elongation complexes are active, as their gene occupancy predicts mRNA expression levels.

RESULTS

Genome-wide profiling reveals Pol II on a majority of genes

We determined genome-wide occupancy profiles by ChIP in exponentially growing *Saccharomyces cerevisiae* strains expressing tandem affinity purification (TAP)–tagged proteins (Online Methods). Chromatin immunoprecipitation was performed as described[11], with modifications (Online Methods and **Supplementary Methods**). Enriched DNA fragments of an average size of 250 nucleotides (nt; **Supplementary Fig. 1**) were analyzed with tiling microarrays that cover the yeast genome at 4-nt resolution[12]. For data normalization, we developed a procedure that corrects for nonspecific antibody binding by using input measurements as well as mock immunoprecipitations (**Supplementary Methods**). Data from two or three highly reproducible replicates were averaged (**Supplementary Table 1**). The profile for the Pol II subunit Rpb3 (**Fig. 1**) matched previous profiles[13] obtained with different strains, experimental protocols and array platforms, but the new profile showed more details (**Supplementary Fig. 2**).

Pol II was observed at genes encoding proteins, small nuclear RNA and small nucleolar RNA, and at regions producing cryptic unstable and unannotated transcripts[14], but was lacking at genes transcribed by Pol I and Pol III (**Fig. 1a** and **Supplementary Fig. 3**). Of 4,366 yeast genes with annotated TSS and pA sites[15], 2,465 (56%) showed Pol II peak occupancies above 20%, consistent with transcription of most of the

[1]Gene Center Munich and Department of Biochemistry, Center for Integrated Protein Science, Ludwig-Maximilians-Universität München, Munich, Germany. [2]These authors contributed equally to this work. Correspondence should be addressed to J.S. (soeding@LMB.uni-muenchen.de) or P.C. (cramer@LMB.uni-muenchen.de).
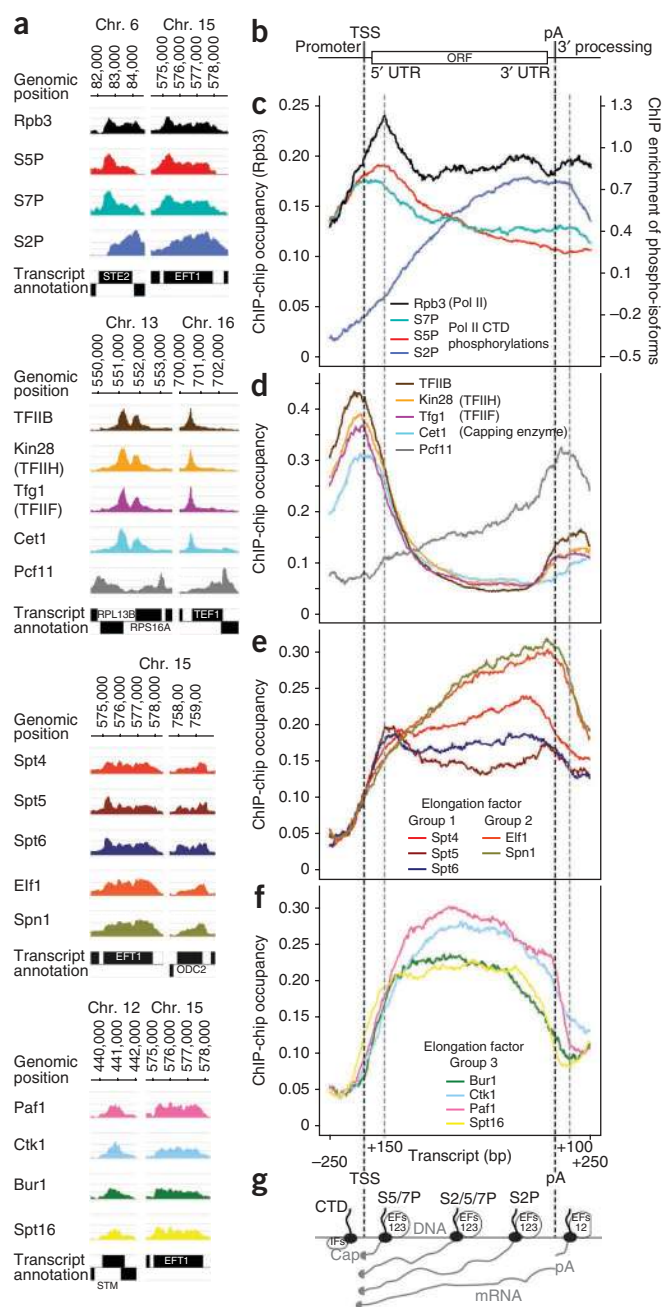
Figure 1 Genome-wide occupancy profiling of the Pol II machinery. (**a**) Factor occupancy on selected genes. Colored profiles represent normalized factor occupancies smoothed by a 150-nt window running median. The color code is used throughout figures. Black boxes indicate transcripts[12] on the Watson (top) and Crick strands (bottom). (**b**) DNA frame with promoter, 5′ untranslated region (UTR), open reading frame (ORF) and 3′ UTR. Dashed black lines indicate the TSS and pA site. Dashed gray lines mark the positions 150 nt downstream of the TSS and 100 nt downstream of the pA site. (**c**) Gene-averaged profiles for the median gene length class (1,238 ± 300 nt, 339 genes) of Pol II and its phosphorylated forms. Profiles of other length classes are generally similar (**Supplementary Fig. 9**). Occupancies and signal intensities are given for Rpb3 and phosphorylated Pol II on the left and right y axes, respectively. For details, see Online Methods. (**d**) Gene-averaged profiles as in **c** for initiation (TFIIB, TFIIF, TFIIH), 5′ capping (Cet1) and termination (Pcf11) factors. (**e**,**f**) Gene-averaged profiles as in **c** for elongation factors of groups 1 (Spt4, Spt5, Spt6), 2 (Elf1, Spn1) and 3 (Spt16, Paf1, Ctk1, Bur1). (**g**) Cartoon representation of Pol II (black dots) and its CTD (black lines) transcribing along DNA (horizontal gray line) from left to right, to produce mRNA (gray lines). IFs, initiation factors; EFs 123, elongation factors of groups 1, 2 and 3; S2/5/7P, phosphorylation of serines 2, 5 and 7.

genome[12]. To average Pol II profiles over genes, we examined the 50% most highly expressed genes[16] that were at least 200 nt away from neighboring genes These were sorted into four main length classes, scaled to adjust for length differences and aligned by their TSS and pA sites (Online Methods). The pA site marks the point of RNA 3′ cleavage and polyadenylation, but transcription continues beyond this site until termination. Consistent with this, the gene-averaged Rpb3 profile revealed Pol II occupancy through the transcribed region and into the region flanking the pA site on the 3′ side (**Fig. 1b,c** and **Supplementary Fig. 4**).

Initiation and termination factors flank the transcribed region

Gene-averaged profiles for the initiation factors TFIIB, TFIIF and TFIIH showed a single strong peak 50–30 nt upstream of the TSS (**Fig. 1d**, **Supplementary Fig. 4** and **Supplementary Table 2**). This indicates the presence of initiation complexes at promoters and is

consistent with a scanning mechanism for TSS location in yeast[17,18]. TFIIF was found only at promoters and not within transcribed regions, indicating that its reported elongation-stimulatory activity *in vitro*[19] is restricted *in vivo* to early RNA elongation and to downstream sites of transient association. The weaker peaks for initiation factors observed downstream of the pA site are mostly due to residual spillover effects from closely spaced genes on the same strand. When we averaged only over convergently transcribed genes, the peaks were reduced two- to three-fold (**Supplementary Fig. 5**). The remaining peaks may indicate gene looping at selected genes. Occupancy of the capping enzyme subunit Cet1 peaked just downstream of the TSS, consistent with capping when the nascent RNA appears on the Pol II surface. The symmetric peaks of averaged occupancy of initiation factor and capping enzyme indicate that these factors are restricted to defined locations just upstream and downstream, respectively, of the TSS. Occupancy of the 3′ processing and termination factor Pcf11 peaked downstream of the pA site, consistent with transcription and completion of mRNA 3′-end formation downstream of the pA site. Thus, representative initiation and termination factors show peak occupancies outside the transcribed region and are apparently not present during mRNA chain elongation.

Elongation factors enter during a single 5′ transition

Elongation factor profiles did not correlate with profiles of initiation or termination factors (**Fig. 2**). Elongation factors were absent at the promoter, but their occupancies sharply increased downstream of the TSS within a narrow window of ~50 nt, indicating coordinated elongation complex assembly during a single 5′ transition (**Fig. 1e,f** and **Supplementary Table 2**). Spt16 was an exception, entering ~30 nt further upstream. Elongation factors showed characteristic distributions over the transcribed region. We observed three distinct profile shapes and used them to group the factors (**Fig. 2b**). Group 1 includes Spt4, Spt5 and Spt6, group 2 includes Spn1 and Elf1, and group 3 includes Spt16, Paf1 and the CTD kinases Bur1 and Ctk1.

Spn1 and Elf1 interact within a Pol II complex

The similar gene-averaged profiles of the poorly characterized factors Spn1 and Elf1 suggested that these factors interact. To test this, we purified Spn1 from yeast using a TAP tag. Spn1 copurified with Elf1 and Pol II (**Supplementary Fig. 6**), consistent with an interaction between Spn1 and Elf1. To probe for a direct Spn1-Elf1 interaction, we
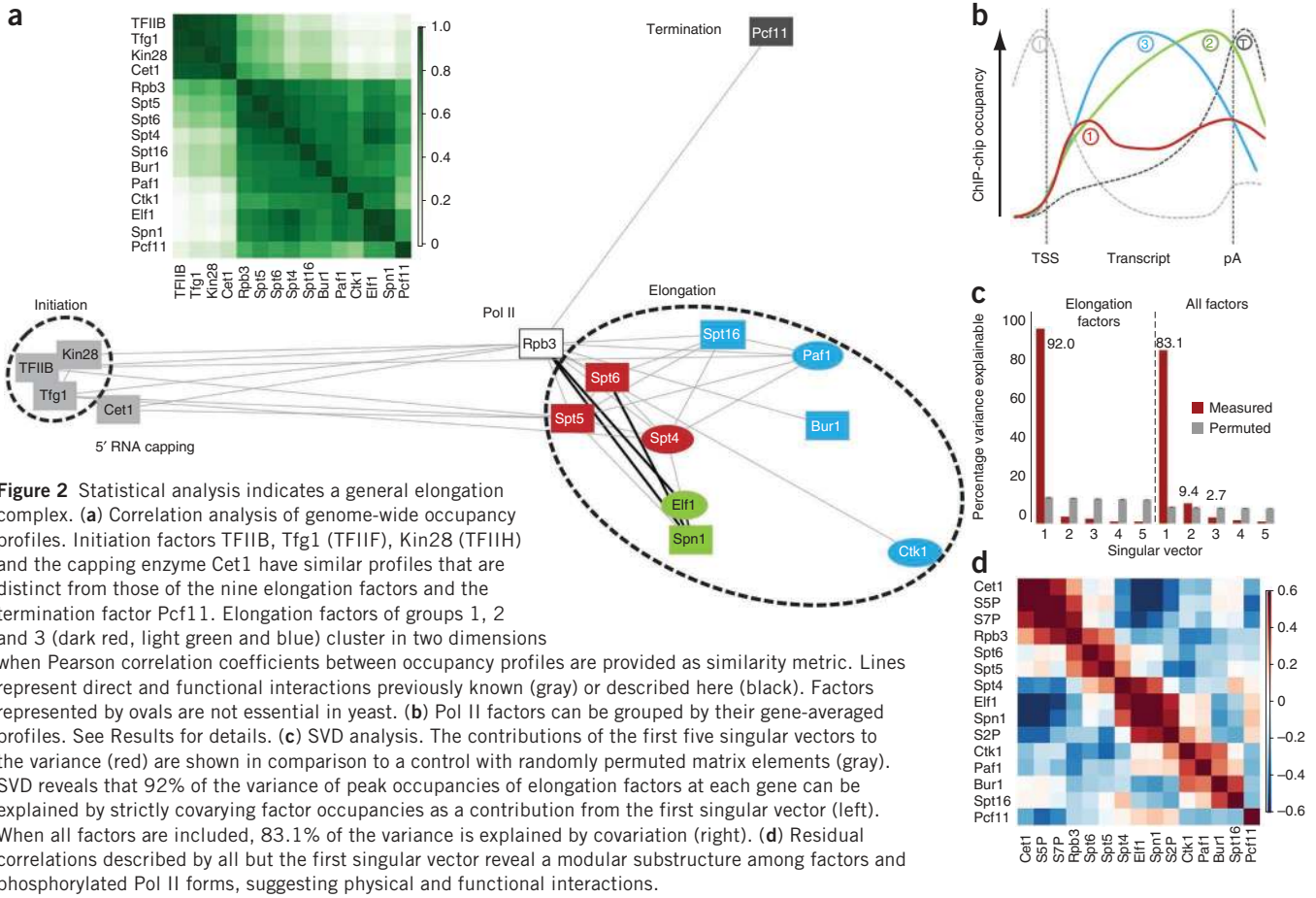
Figure 2 Statistical analysis indicates a general elongation complex. (**a**) Correlation analysis of genome-wide occupancy profiles. Initiation factors TFIIB, Tfg1 (TFIIF), Kin28 (TFIIH) and the capping enzyme Cet1 have similar profiles that are distinct from those of the nine elongation factors and the termination factor Pcf11. Elongation factors of groups 1, 2 and 3 (dark red, light green and blue) cluster in two dimensions when Pearson correlation coefficients between occupancy profiles are provided as similarity metric. Lines represent direct and functional interactions previously known (gray) or described here (black). Factors represented by ovals are not essential in yeast. (**b**) Pol II factors can be grouped by their gene-averaged profiles. See Results for details. (**c**) SVD analysis. The contributions of the first five singular vectors to the variance (red) are shown in comparison to a control with randomly permuted matrix elements (gray). SVD reveals that 92% of the variance of peak occupancies of elongation factors at each gene can be explained by strictly covarying factor occupancies as a contribution from the first singular vector (left). When all factors are included, 83.1% of the variance is explained by covariation (right). (**d**) Residual correlations described by all but the first singular vector reveal a modular substructure among factors and phosphorylated Pol II forms, suggesting physical and functional interactions.

coexpressed the two factors in bacteria. Spn1 and Elf1 did not copurify after coexpression (data not shown). These results suggest that Spn1 and Elf1 interact indirectly within a Pol II complex, and the profiling data suggest that their recruitment and functions during the transcription cycle are distinct from those of other elongation factors.

Elongation factors exit during a two-step 3′ transition
Around the pA site, two steps of a 3′ transition could be distinguished. The two-step transition was most easily seen at genes with high factor occupancies, such as ribosomal protein genes (**Fig. 3** and
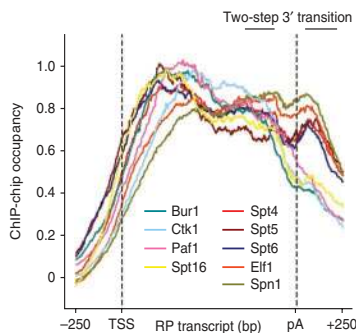
Figure 3 Two-step 3′ transition observed at ribosomal protein (RP) genes. Shown are averaged elongation factor profiles on selected ribosomal protein genes. Dashed lines indicate the TSS and pA sites. The regions of the two-step 3′ transition are indicated. In contrast to the averaged profiles in **Figure 1**, the very high occupancies did not allow us to align profiles with their promoter minima along the *y* axis.

Supplementary Table 2). Whereas group 3 factor occupancies sharply decreased upstream of the pA site, factors from groups 1 and 2 apparently exited further downstream, suggesting they are present during RNA 3′-end formation and possibly during transcription termination. Spn1 and Elf1 peaked just upstream of the pA site, and 100 nt downstream of this site they were still present at about 80% of their peak occupancies (**Fig. 1e** and **Supplementary Table 2**).

A general elongation complex for chromatin transcription
High correlations between elongation factor profiles (**Fig. 2a** and **Supplementary Fig. 7**) suggested that all elongation factors co-occupy active genes. To investigate this, we measured covariation in the data sets by singular value decomposition (SVD). We calculated peak occupancies for nine elongation factors within 4,366 genes (Online Methods). After subtracting the row mean of the $9 \times 4,366$ matrix from each element, we subjected the resulting matrix to SVD. The first singular value, which describes strict covariation, explained 92% of the total variance (**Fig. 2c**, left, and **Supplementary Fig. 8a**). Thus, elongation factor occupancies covaried over all genes, consistent with a general composition of the elongation complex. The apparent elongation complex composition and coordinated assembly during a 5′ transition were independent of gene length, expression, function, transcript type, size of the nucleosome-depleted promoter region and the presence of introns (**Fig. 4** and **Supplementary Fig. 9**). Although differences in the composition of elongation complexes in individual cells cannot be ruled out, these results indicate a general initiation-elongation transition and a general elongation complex composition on Pol II genes.
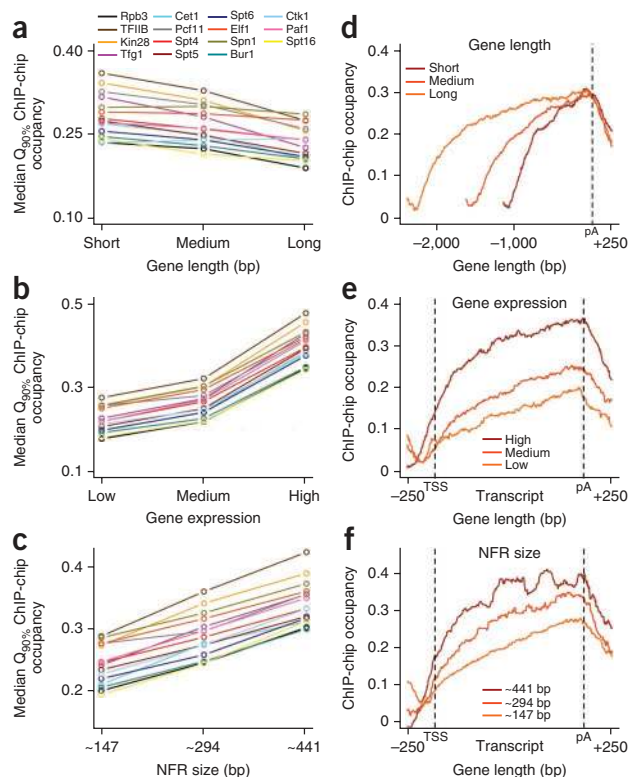
Figure 4 Transcription complex composition and transitions are independent of gene length, expression and NFR size. (**a–c**) Medians of the peak factor occupancies covary between different length classes (**a**), expression level classes (**b**) and nucleosome-free promoter region (NFR) size classes (**c**)[53,54]. $Q_{90\%}$, 90% quantiles of gene occupancies, used as a proxy for peak occupancies. (**d–f**) Gene-averaged profiles of the representative elongation factor Elf1 have shapes and transition points that are independent of gene length (**d**), expression level (**e**) and NFR size (**f**). The same holds for all other profiled factors and also for genes grouped by transcript type and functional class (**Supplementary Fig. 9**).

indicate that CTD phosphorylation patterns previously observed on individual genes[4,5] occur globally and depend on TSS location.

### Recruitment of CTD kinases explains CTD phosphorylations

The Ser7 and Ser5 peaks just downstream of the TSS are consistent with the presence of the Ser7 and Ser5 kinase Kin28 (refs. 21,22) just upstream of the TSS (**Fig. 5a**). Furthermore, the early peak of Ser7 phosphorylation is consistent with dependence of Ser7 phosphorylation on the co-regulatory Mediator complex[23], which binds at promoters. The subsequent increase in Ser2 phosphorylation is consistent with the Ser2 kinases Bur1 and Ctk1 (refs. 24,25) entering during the 5′ transition and remaining present in transcribed regions (**Fig. 5a**). Thus, specific CTD phosphorylations can be explained by the recruitment of corresponding specific CTD kinases at defined distances from the TSS. Unfortunately, we were not able to obtain satisfactory profiles of CTD phosphatases to compare their recruitment with the observed decreases in CTD phosphorylation.

CTD phosphorylation and factor recruitment

To clarify the relationships between CTD phosphorylations and factor occupancies, we subjected all profiles to SVD and correlated residual profiles lacking the contributions of the first SVD term; the eliminated first term described 85.6% of the covariation of factor occupancies (**Supplementary Fig. 8b**). Ser7 and Ser5 phosphorylation correlated with the occupancy of the capping enzyme Cet1 (**Fig. 2d**), as expected from binding of capping enzyme to Ser5-phosphorylated CTD *in vitro*[26]. As Ser5 phosphorylation levels peaked more than 100 nt downstream from the Cet1 peak, the capping enzyme may already be bound when the first Ser5 residues are phosphorylated. Cet1 occupancy dropped very sharply further downstream, whereas Ser5 phosphorylation levels remained high, suggesting an active mechanism to release the capping enzyme from the CTD. Ser2 phosphorylation correlated strongly with

CTD phosphorylation profiles depend on TSS location

To investigate how the observed profiles and transitions correlate with CTD phosphorylation, we determined ChIP-chip profiles for Pol II phosphorylated at CTD residues Ser7, Ser5 and Ser2 using site-specific antibodies[20] (Online Methods). The averaged profiles revealed broad peaks of Ser7 and Ser5 phosphorylation at around 20 and 120 nt downstream of the TSS (**Fig. 1c**, **Supplementary Fig. 4** and **Supplementary Table 2**). Ser7 and Ser5 phosphorylation decreased over the transcribed region, whereas Ser2 phosphorylation increased, saturated at 600–1,000 nt downstream of the TSS and sharply decreased 100–200 nt downstream of the pA site (**Fig. 5**, **Supplementary Figs. 4** and **9** and **Supplementary Table 2**). The point where full Ser2 phosphorylation was reached did not depend on pA site location, but rather on TSS location (**Fig. 5b**). Although we cannot rule out changes in ChIP efficiency due to the accessibility of the phosphorylated CTD to antibodies, our results overall
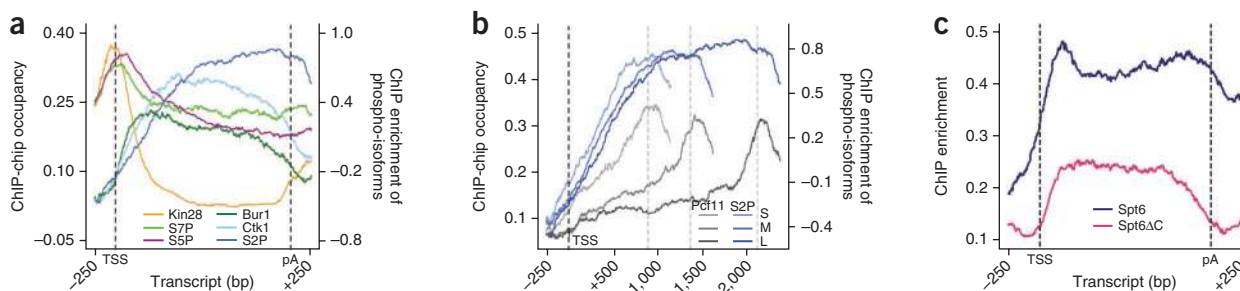
Figure 5 Pol II phosphorylation and factor occupancy. (**a**) Gene-averaged profiles (for length class only) of CTD phosphorylations, CTD Ser5 kinase Kin28 and Ser2 kinases Bur1 and Ctk1. (**b**) Profiles for Ser2-phosphorylated Pol II and Pcf11 aligned at the TSS and pA site, respectively, and averaged for length classes short (S), medium (M) and long (L). Occupancy and signal intensity for Pcf11 and S2P are plotted on the left and right *y* axes, respectively. (**c**) Gene-averaged profiles of Spt6 and the C-terminal deletion variant Spt6ΔC (lacking the 202 C-terminal residues)[16]. As 100% occupancy levels are not expected for Spt6ΔC, the *y* axis shows ChIP enrichments obtained by normalization with input measurements as well as mock IPs (**Supplementary Methods**) without scaling to 100% occupancy.
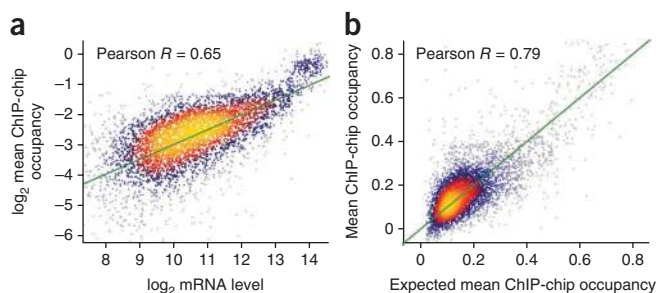
Figure 6 Elongation complex occupancy predicts mRNA expression. (**a**) The logarithm of the average elongation factor and Rpb3 transcript occupancy is highly correlated with the logarithm of mRNA levels. For constant mRNA synthesis rates (elongation complex speeds) v and decay rates r, we expect a linear relationship between elongation factor occupancy o and mRNA levels c, because at equilibrium the rates of mRNA synthesis and decay are equal, and thus $ov = cr$. This would result in a linear dependence between log o and log c with slope 1: log o = log c + log (r/v). The actual slope of 0.49 (green) implies that the ratio of Pol II speed to decay rate increases slightly with increasing mRNA level ($v/r \propto c^{0.51}$). For ribosomal protein genes, v/r is about three-fold higher than average. (**b**) The averaged transcript occupancy of Rpb3 and the nine elongation factors is highly correlated with the occupancy expected on the basis of the relationship between occupancies and observed mRNA levels in **a**.

Spn1 and Elf1 occupancy (**Fig. 2d** and **Supplementary Fig. 8b**), suggesting these factors are stabilized within the elongation complex by Ser2 phosphorylation.

Possible CTD masking and CTD-independent recruitment

CTD Ser2 phosphorylation did not correlate with occupancy of Pcf11 and Spt6, although these factors bind to the Ser2-phosphorylated CTD *in vitro*[27,28]. Pcf11 was recruited mainly at the pA site (**Fig. 5b** and **Supplementary Fig. 9**), consistent with the known role of Pcf11 in RNA 3′ processing. This may be explained if the Ser2-phosphorylated CTD becomes accessible to Pcf11 only after pA site passage and 3′ RNA cleavage. Alternatively, Pcf11 cross-linking may be increased by cooperative interactions of factors and RNA around the pA site or by conformational changes in the elongation complex. Spt6 entered early, during the 5′ transition, suggesting a recruitment mechanism independent of CTD Ser2 phosphorylation. To investigate this, we determined the ChIP-chip profile of a variant of Spt6 lacking its CTD-binding C-terminal domain (Spt6ΔC), using a yeast strain that expresses only a truncated Spt6 lacking the last 202 residues[16]. Deletion of the Spt6 CTD-binding domain led to much less recruitment of Spt6 but did not abolish its entry during the 5′ transition (**Fig. 5c** and **Supplementary Fig. 10**). Thus, Spt6 is apparently recruited in a CTD-independent manner during the 5′ transition, but full recruitment requires the CTD-binding domain. The CTD-binding domain was required for retaining Spt6 until the pA site was reached (**Fig. 5c**), consistent with Spt6's preference for binding the Ser2-phosphorylated CTD. These results indicate that binding of a factor to the phosphorylated CTD *in vitro* cannot predict factor recruitment *in vivo*. This suggests that the CTD may be transiently masked and its accessibility regulated, and that factor recruitment includes CTD-independent and CTD phosphorylation type–independent recruitment.

No evidence for promoter-proximally stalled Pol II

In higher eukaryotes, Pol II is often stalled early during elongation near the promoter[29–31] and can be released by activators[32]. Our data do not provide evidence for the presence of such promoter-proximally stalled Pol II in growing yeast. Genes with stalled Pol II would show Ser5 but not Ser2 phosphorylation, or at least more Ser5 than Ser2 phosphorylation. However, we did not find genes with a peak for Ser5 phosphorylation and no peak for Ser2 phosphorylation (data not shown). SVD analysis of initiation and elongation factor profiles revealed a high covariance of 83.1% (**Fig. 2c**, right, and **Supplementary Fig. 8c**), suggesting that initiation complexes are generally efficiently converted to elongation complexes. Although Rpb3 occupancy peaks around 150 nt downstream of the TSS (**Fig. 1c**), this does not indicate polymerase stalling, as stalling generally occurs closer to the TSS. Instead, this Pol II peak may be explained by the 5′ transition being slow because of capping, phosphorylation and assembly events, leading to an apparent accumulation of Pol II. Alternatively or additionally, the peak may reflect transient pausing of Pol II between the +1 and +2 nucleosomes, which are positioned around 40 and 210 nt downstream of the TSS, respectively[33]. Our data from exponentially growing yeast also do not show evidence for polymerase peaks upstream of the TSS, as observed in yeast during stationary growth[34].

General elongation complexes are productive

To investigate whether the general elongation complexes are active on most genes, we correlated averaged Rpb3 and elongation factor occupancies with mRNA levels[16]. The mRNA level should be proportional to the mRNA synthesis rate of a single elongation complex times its occupancy, divided by the mRNA decay rate (see legend of **Fig. 6**). For constant mRNA synthesis and decay rates, we would therefore expect a linear dependence of occupancy on mRNA level, corresponding to a slope of 1 in a log-log plot. We found a robust correlation of 0.65 between the log occupancy and the log mRNA levels (**Fig. 6a**); however, the slope was only ~0.5. A correlation of 0.71 was obtained when we used the distance-filtered gene set. This shows that increased mRNA levels are due not only to greater elongation complex occupancy, but also to a higher ratio of mRNA synthesis rates over decay rates. The same dependence leads to a high correlation of 0.79 between the observed average occupancy and the expected occupancy calculated from the mRNA level (**Fig. 6b**). These correlations indicate that most general elongation complexes are active in producing mRNA and that occupancy of genes by the general elongation complex is a good predictor for gene expression level.

DISCUSSION

Here we have established an improved protocol for obtaining high-resolution genome-wide occupancy profiles for components of the Pol II transcription machinery, to investigate the chromatin transcription cycle *in vivo* (**Fig. 1g**). We demonstrate that Pol II elongation factors, which are required for chromatin passage and RNA processing on individual genes, associate with all transcribed Pol II genes in proliferating yeast cells. Elongation factors show three distinct, nonrandom patterns of distribution over genes, and these distribution patterns are independent of gene length, type, function or nucleosome structure. The underlying general elongation complex is established and disassembled during uniform transitions at the beginnings and the ends of genes. Elongation factors enter during a sharp 5′ transition just downstream of the TSS and exit in a two-step 3′ transition around the pA site. Our genome-wide Pol II phosphorylation profiles match patterns observed at individual genes and are explained by the recruitment of specific CTD kinases at defined distances from the TSS. Analysis of CTD phosphorylation profiles does not support the existence of promoter-proximally stalled Pol II in growing yeast. CTD phosphorylation is not predictive of the recruitment of factors that

bind the phosphorylated CTD *in vitro*. Instead, we obtained evidence that CTD accessibility is regulated by transient CTD masking and that recruitment mechanisms are CTD independent. Finally, occupancy of genes by the general elongation complex predicts the resulting mRNA levels, suggesting that most or all elongation complexes are active.

Published biochemical and genetic data suggest that the 5′ transition corresponds to a coordinated conversion of a general initiation complex to a general elongation complex. The conversion includes initiation factor dissociation, which liberates the Pol II clamp domain[35] for binding Spt5 (ref. 36). Spt5 could coordinate entry of group 1 factors because it binds Spt4 *in vitro*[37] and associates with Spt6 *in vivo*[38]. Group 1 factors could recruit group 2 factors, as Spt6 binds Spn1 (refs. 38,39). Consistent with this, group 1 factors interact genetically with Elf1 (ref. 40) and, as we show here, Spn1 and Elf1 interact within a Pol II complex. Recruitment of group 3 factors may commence with CTD Ser5 phosphorylation; this recruits Bur1 (ref. 41), which in turn phosphorylates Spt5, thereby recruiting Paf1 (refs. 25,42,43). Spt16 enters around 30 nt upstream from other elongation factors (**Fig. 1f** and **Supplementary Table 2**), perhaps because it binds to the +1 nucleosome[44]; this is consistent with its role as a histone chaperone[45]. Initiation factors are not present when the 5′ transition is completed around 150 nt downstream of the TSS, consistent with Ctk1 promoting dissociation of initiation factors[46].

The general two-step 3′ transition we observed is consistent with ChIP data obtained at individual genes. The early exit of group 3 factors Paf1, Ctk1 and Bur1, and the continued presence of Spt4, Spt5 and Spt6, have previously been observed[47,48]. Our results, however, show that the reported Spt16 occupancy downstream of the pA site[47] does not occur globally. Also, our observation of peak levels of the bona fide 3′-processing factor Pcf11 downstream of the pA site challenges the idea of an early loading of 3′-processing factors in the 5′ region of genes[49]. We observed continued presence of group 1 and group 2 factors downstream of the pA site, consistent with their reported roles in mRNA 3′ processing[50], mRNA export[27] and re-establishment of chromatin structure after Pol II passage[51].

Our results also provide insights into the role of CTD phosphorylation during transcription complex transitions and in the coordination of transcription-coupled events. First, peak levels of Ser7 and Ser5 phosphorylation were generally observed in the 5′ regions of genes, and peak Ser2 phosphorylation in the 3′ regions of genes. Second, the 5′ transition occurs before any substantial Ser2 phosphorylation, suggesting that the assembly of the general elongation complex is independent of Ser2 phosphorylation, consistent with the observation that the Ser2 kinase Ctk1 is not required for association of elongation factors with transcribing Pol II[52]. Third, peak levels of Ser2 phosphorylation are always reached 600–1,000 nt downstream of the TSS, regardless of the position of the pA site. This argues against a role of Ser2 phosphorylation in triggering the 3′ transition, although Ser2 phosphorylation is required for cotranscriptional 3′ RNA processing[52]. Fourth, the recruitment of Pcf11 and Spt6, which both bind the Ser2-phosphorylated CTD *in vitro*, cannot be explained solely by factor binding to the Ser2-phosphorylated CTD *in vivo*. Instead, late Pcf11 entry suggests CTD masking within the transcribed region and an increase in CTD accessibility upon RNA cleavage at the pA site, allowing for Pcf11 binding. Furthermore, Spt6 enters during the 5′ transition even when it lacks its CTD-binding domain, indicating that a CTD-independent recruitment mechanism exists. The CTD-binding domain seems to be more important for retaining Spt6 until the pA site is reached than for recruiting it during the 5′ transition.

In conclusion, our data support the following view of a productive chromatin transcription cycle. The initiation complex forms ~30–50 nt upstream of the TSS and contains unphosphorylated Pol II and initiation factors. The complex then scans for the TSS downstream, begins RNA synthesis and triggers RNA 5′ capping. Next, the complex is converted into a general elongation complex during a sharp and efficient but presumably slow 5′ transition that is completed around 150 nt downstream of the TSS, where Ser5 phosphorylation levels peak. During subsequent elongation, Ser2 phosphorylation increases until it reaches peak levels 600–1,000 nt downstream of the TSS. During a two-step 3′ transition, a group of elongation factors exits upstream of the pA site, whereas another group persists downstream, where it is joined by additional factors such as Pcf11, resulting in mRNA 3′ processing and transcription termination.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

1. Pokholok, D.K., Hannett, N.M. & Young, R.A. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol. Cell* **9**, 799–809 (2002).
2. Orphanides, G. & Reinberg, D. RNA polymerase II elongation through chromatin. *Nature* **407**, 471–475 (2000).
3. Orphanides, G. & Reinberg, D. A unified theory of gene expression. *Cell* **108**, 439–451 (2002).
4. 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).
5. 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).
6. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
7. Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**, 178–191 (2009).
8. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19**, 1401–1415 (2005).
9. Hirose, Y. & Manley, J.L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429 (2000).
10. Venters, B.J. & Pugh, B.F. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res.* **19**, 360–371 (2009).
11. Aparicio, O. *et al.* Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. in *Current Protocols in Molecular Biology* (eds. Ausubel, F.A. *et al.*) Ch. 21, Unit 21.3 (Wiley, 2005).
12. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* **103**, 5320–5325 (2006).
13. Jasiak, A.J. *et al.* Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J. Biol. Chem.* **283**, 26423–26427 (2008).

14. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
15. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
16. Dengl, S., Mayer, A., Sun, M. & Cramer, P. Structure and in vivo requirement of the yeast Spt6 SH2 domain. *J. Mol. Biol.* **389**, 211–225 (2009).
17. Kuehner, J.N. & Brow, D.A. Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* **281**, 14119–14128 (2006).
18. Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
19. Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D.H. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276**, 42601–42609 (2001).
20. Chapman, R.D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780–1782 (2007).
21. 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).
22. Akhtar, M.S. *et al.* TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393 (2009).
23. Boeing, S., Rigault, C., Heidemann, M., Eick, D. & Meisterernst, M. RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a Mediator-dependent fashion. *J. Biol. Chem.* **285**, 188–196 (2010).
24. Murray, S., Udupa, R., Yao, S., Hartzog, G. & Prelich, G. Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase. *Mol. Cell. Biol.* **21**, 4089–4096 (2001).
25. Liu, Y. *et al.* Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol. Cell. Biol.* **29**, 4852–4863 (2009).
26. Rodriguez, C.R. *et al.* Kin28, the TFIIH-Associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.* **20**, 104–112 (2000).
27. 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).
28. Barillà, D., Lee, B.A. & Proudfoot, N.J. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **98**, 445–450 (2001).
29. Nechaev, S. *et al.* Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
30. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
31. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
32. Rahl, P.B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
33. Jiang, C. & Pugh, B.F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
34. Radonjic, M. *et al.* Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol. Cell* **18**, 171–183 (2005).
35. Chen, H.-T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
36. Hirtreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
37. Guo, M. *et al.* Core structure of the yeast Spt4-Spt5 complex: a conserved module for regulation of transcription elongation. *Structure* **16**, 1649–1658 (2008).
38. Lindstrom, D.L. *et al.* Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell. Biol.* **23**, 1368–1378 (2003).
39. Krogan, N.J. *et al.* RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992 (2002).
40. Prather, D., Krogan, N.J., Emili, A., Greenblatt, J.F. & Winston, F. Identification and characterization of Elf1, a conserved transcription elongation factor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 10122–10135 (2005).
41. Qiu, H., Hu, C. & Hinnebusch, A.G. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol. Cell* **33**, 752–762 (2009).
42. Zhou, K., Kuo, W.H.W., Fillingham, J. & Greenblatt, J.F. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci. USA* **106**, 6956–6961 (2009).
43. Laribee, R.N. *et al.* BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Curr. Biol.* **15**, 1487–1493 (2005).
44. Stuwe, T. *et al.* The FACT Spt16 "peptidase" domain is a histone H3–H4 binding module. *Proc. Natl. Acad. Sci. USA* **105**, 8884–8889 (2008).
45. Belotserkovskaya, R. *et al.* FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090–1093 (2003).
46. Ahn, S.H., Keogh, M.-C. & Buratowski, S. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. *EMBO J.* **28**, 205–212 (2009).
47. Kim, M., Ahn, S.-H., Krogan, N.J., Greenblatt, J.F. & Buratowski, S. Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J.* **23**, 354–364 (2004).
48. Keogh, M.-C., Podolny, V. & Buratowski, S. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol. Cell. Biol.* **23**, 7005–7018 (2003).
49. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D.L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–78 (2008).
50. Kaplan, C.D., Holland, M.J. & Winston, F. Interaction between transcription elongation factors and mRNA 3′-end formation at the *Saccharomyces cerevisiae* GAL10–GAL7 locus. *J. Biol. Chem.* **280**, 913–922 (2005).
51. Kaplan, C.D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**, 1096–1099 (2003).
52. 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).
53. Hesselberth, J.R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
54. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).

ONLINE METHODS

Yeast strains. *Saccharomyces cerevisiae* BY4741 strains with C-terminally TAP-tagged genes (Open Biosystems) were tested for expression of the correct tagged protein. The BY4741 wild-type strain was used for profiling with phospho-CTD–specific antibodies 3E8, 3E10 and 4E12 (ref. 20) in the presence of phosphatase inhibitors.

Chromatin immunoprecipitation. Chromatin immunoprecipitation was performed as described[11,55], with modifications (**Supplementary Methods**). Strains were grown to mid-log phase and treated with 1% (v/v) formaldehyde for 20 min at 20 °C. Cells were disrupted by vortexing and chromatin was sheared with a Bioruptor UCD-200 device. Cross-links were reversed by heating with proteinase K, DNA was purified and RNA was digested. DNA was amplified with GenomePlex Whole Genome Amplification 2 Kit. DNA quantity and quality were controlled with a NanoDrop Photometer. Over 1 μg of DNA was typically obtained from 600 ml culture. At least two biological replicates were analyzed. DNA labeling and microarray analysis were performed as recommended by Affymetrix. For data normalization, we used a new procedure that included both mock immunoprecipitation and input measurements (**Supplementary Methods**). Resulting factor occupancies were validated by quantitative PCR at selected loci (**Supplementary Fig. 1**).

Gene-averaged profiles. Of the 5,769 nondubious protein-coding nuclear genes in the *Saccharomyces* Genome Database, we analyzed 4,366 genes (76%) with assigned TSS and pA site[15]. For the calculation of gene-averaged profiles, genes were excluded when their TSS was downstream or over 200 nt upstream of the start codon, when their pA site was upstream or over 200 nt downstream of the stop codon, or when neighboring genes were less than 200 nt away (**Supplementary Methods**). Of the resulting 1,786 genes, 1,140 showing expression levels higher than the median expression level of all 5,769 yeast genes[16] were divided into four classes by open reading frame length (**Supplementary Fig. 9a**). There were 266, 339 and 299 genes in the three major length classes, short, medium and long, respectively. Profiles within each class were scaled to median gene length and median-averaged over genes. For SVD we used 90% quantiles of gene occupancies as a robust proxy for peak occupancies.

55. Fan, X., Lamarre-Vincent, N., Wang, Q. & Struhl, K. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res.* **36**, e125 (2008).