ARTICLE

doi:10.1038/nature10799

Genome-wide structure and organization of eukaryotic pre-initiation complexes

Ho Sung Rhee[1] & B. Franklin Pugh[1]

Transcription and regulation of genes originate from transcription pre-initiation complexes (PICs). Their structural and positional organization across eukaryotic genomes is unknown. Here we applied lambda exonuclease to chromatin immunoprecipitates (termed ChIP-exo) to examine the precise location of 6,045 PICs in *Saccharomyces*. PICs, including RNA polymerase II and protein complexes TFIIA, TFIIB, TFIID (or TBP), TFIIE, TFIIF, TFIIH and TFIIK were positioned within promoters and excluded from coding regions. Exonuclease patterns were in agreement with crystallographic models of the PIC, and were sufficiently precise to identify TATA-like elements at so-called TATA-less promoters. These PICs and their transcription start sites were positionally constrained at TFIID-engaged downstream +1 nucleosomes. At TATA-box-containing promoters, which are depleted of TFIID, a +1 nucleosome was positioned to be in competition with the PIC, which may allow greater latitude in start-site selection. Our genomic localization of messenger RNA and non-coding RNA PICs reveals that two PICs, in inverted orientation, may occupy the flanking borders of nucleosome-free regions. Their unambiguous detection may help distinguish bona fide genes from transcriptional noise.

Assembly of the PIC and its post-assembly control are critical early steps in the transcription of eukaryotic genes. TBP (TATA-binding protein) arrives at most promoters as part of the multi-subunit TFIID complex that includes TBP-associated factors (TAFs)[1]. Together these proteins help recruit RNA polymerase (Pol) II and its entourage of general transcription factors (GTFs) to the transcription start sites (TSSs) of genes[2–4]. These PICs assemble in nucleosome-free promoter regions (NFRs) that are flanked by an upstream −1 nucleosome and a downstream +1 nucleosome[5]. PICs have largely been defined biochemically using purified GTFs at a few model genes[2–4], but little is known about their assembly and organization *in vivo*, particularly at near-base-pair resolution on a genome-wide scale.

An oddity of TBP is that when it is part of the TFIID complex, it tends to bind promoters that lack the TATA box consensus TATAWAWR (W indicates A/T; R indicates A/G)[6]. Approximately 80–90% of all *Saccharomyces* genes are thus designated as 'TATA-less', and have a predominant PIC assembly mechanism and chromatin architecture that differs substantially from those in the 'TATA box' class of genes[6–9]. So far, no TBP-binding motif has been identified at TATA-less promoters, and so the origins of TFIID-promoter specificity have been rather enigmatic[10]. When TBP is not part of the TFIID complex, the SAGA complex directs TBP to TATA-box-containing Pol II promoters[11–13].

TFIIA and TFIIB clamp TBP to DNA, and make DNA contacts immediately upstream and downstream of the TATA box. TFIIB is a linchpin between TBP and Pol II[14,15]. Its intimate contact with Pol II directs how far downstream Pol II productively initiates transcription[16,17]. TFIIF enhances the interaction of Pol II with TFIIB, assists in recruiting TFIIE, and promotes downstream elongation events[3,18]. TFIIE then stimulates DNA strand separation by Pol II at the TSS, and enhances the activity of TFIIH. TFIIH holoenzyme is multi-functional, having ATP-dependent helicase (Ssl2 and Rad3) and kinase (Kin28) activities that reside on biochemically separable sub-complexes (TFIIH and TFIIK, respectively), both of which are key to efficient open complex formation and transcription initiation[19–21].

We examined the structural organization of PICs and their specificity on a genomic scale using ChIP-exo[22]. This novel strategy substantially improved mapping resolution and eliminated many false positives. The exonuclease processively degrades a DNA strand in the 5′–3′ direction until a crosslinking point is encountered (Supplementary Fig. 1a). The crosslinking inefficiency inherent to ChIP allows multiple crosslinking points to be detected in a population by deep sequencing. When applied to the GTFs on a genomic scale, we obtained detailed and comprehensive information on PIC structure and genomic organization.

Genome-wide PIC structure

We applied ChIP-exo genome-wide to Pol II and each GTF (Fig. 1a), and verified binding for TFIIB by locus-specific PCR using a series of tiled primers (Supplementary Fig. 1b). When exonuclease stop sites were mapped over all annotated mRNA promoters that contained a TATA box consensus, a distinctive pattern was observed (Fig. 1b). Importantly, each GTF displayed a strand-specific composite pattern of exonuclease stop sites that occurred only when TATA boxes, but not TSSs (Supplementary Fig. 2a and data not shown), were aligned, indicating that PICs are positioned with respect to the TATA box.

For TFIIB, we detected four DNA crosslinking points (pairs of exonuclease stops), designated B1–B4 (Fig. 1c). The TATA box was precisely centred between B1 and B2, which were separated by 20 ± 3 base pairs (bp). Crosslinking point B3 and the diffuse B4 region indicate that TFIIB further crosslinked over a broad region downstream of the TATA box towards the TSS. We compared the four TFIIB crosslinking points to crystallographic-based models of open and closed TBP–TFIIB–Pol-II–promoter complexes (Fig. 1c)[14,15]. An important caveat of the crystallographic models is they were built from multiple independent structures of truncated TFIIB–TBP–TATA, TFIIB–Pol II and Pol II elongation complexes. Thus, the combined structures represent a hypothetical organization.

Within the modelled closed and open structures, crosslinking site B1 precisely (± 3 bp) mapped to where the TFIIB carboxy-terminal core straddles the upstream DNA-binding stirrup of TBP (11 bp

[1]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

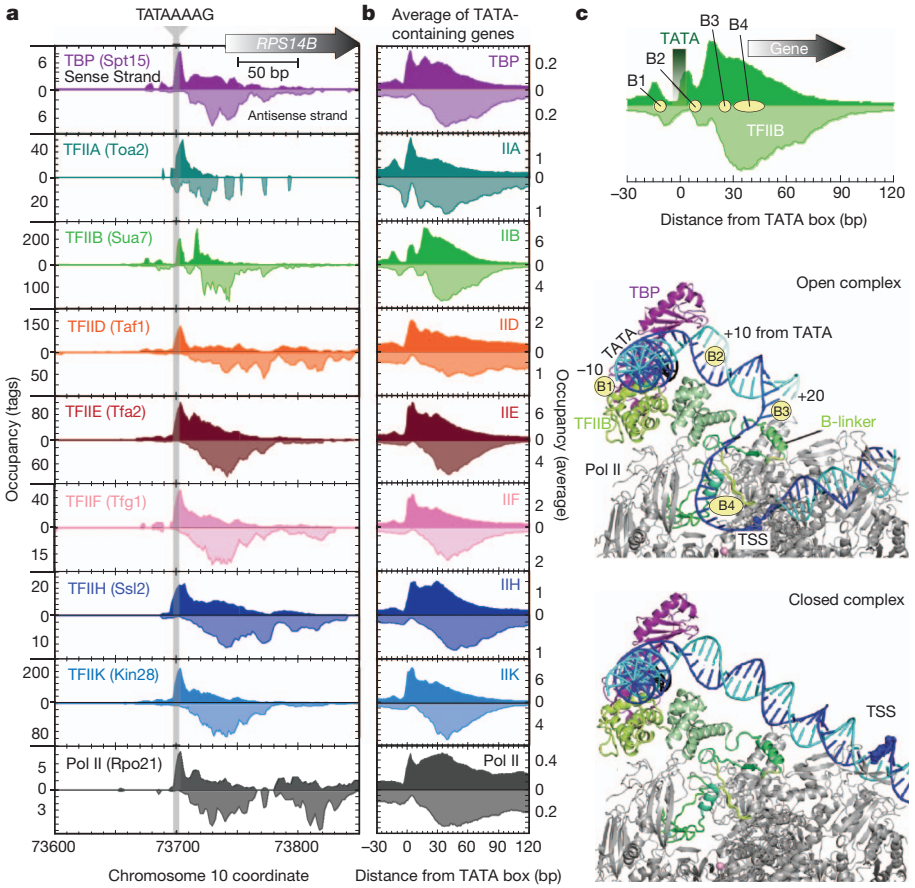

Figure 1 | Genome-wide structural organization of PICs. **a**, Raw ChIP-exo tag distribution for GTFs and Pol II around the *RPS14B* gene. Filled plots represent unfiltered 5′ ends of sequencing tags on the sense (darker fill) and antisense strand (lighter fill). **b**, Average GTF and Pol II occupancy around the TATA box of 676 annotated mRNA genes. Plots are as in panel **a**. **c**, Relationship of four TFIIB crosslinking points to crystallographic-based models of the PIC[14]. The top panel is expanded from panel **b** for TFIIB. The middle and bottom panels show modelled open and closed TBP–TFIIB–Pol-II–promoter DNA complexes, respectively.

upstream of the TATA box midpoint). B2 mapped precisely (± 3 bp) to where the TFIIB core amino-terminal cyclin fold encounters DNA just downstream of TBP's other stirrup (9 bp downstream of the TATA box midpoint). B3 mapped to where the TFIIB linker region is closest to the DNA, which was 19 bp downstream of the TATA box midpoint. B4 corresponded to a broad region defined by close proximity of the TFIIB reader (or finger) domain to single-stranded DNA within the modelled open complex, but was not evident within the closed complex. Similar broad regions of crosslinking were observed with the other GTFs (Fig. 1b), and may reflect indirect crosslinking. Support that these PICs represent open complexes is provided by permanganate reactivity studies of the *GAL1*, *GAL10* and *HSP82* loci[23,24]. Taken together, the entirety of the genomic crosslinking sites observed with the GTFs and Pol II fits remarkably well with the crystallographic models of the PIC open complex[14,15], and with many aspects of *in vitro* chemical crosslinking of these proteins[19,25–27].

TATA-like elements at TATA-less genes

An apparent paradox of so-called TATA-less promoters is their utilization of TFIID, which has long been described as the TATA-box-binding complex[4]. However, it is unclear whether the TBP subunit of TFIID recognizes specific DNA sequences at TATA-less promoters. Inasmuch as TBP is expected to be found at all TATA-less promoters, and motif searching algorithms failed to identify candidate TBP-binding sites, we instead opted to search for sequence elements with up to two mismatches to the TATAWAWR consensus. We also limited our search to measured PIC locations. Remarkably, 99% of the PICs at TATA-less promoters contained a sequence having two or less mismatches to the TATA box consensus (Fig. 2). We refer to these mismatched elements as

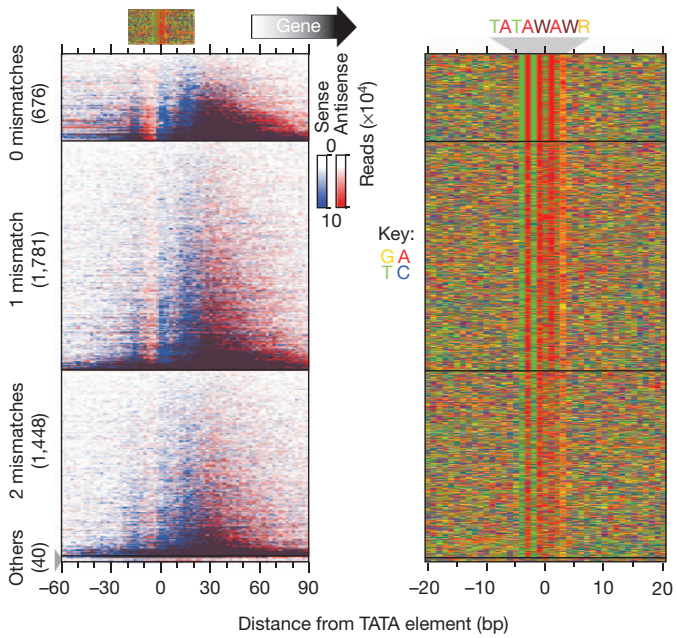

Figure 2 | Identification of TATA-like elements at TATA-less genes. Left, TFIIB occupancy around individual TFIIB-enriched TATA elements of mRNA genes (rows, $n = 3{,}945$), sorted by occupancy level. Occupancy on the sense (blue) and antisense (red) strands is shown with respect to TSS orientation. Right, a colour chart representation of the DNA sequence located ± 20 bp from the TATA element midpoint and ordered as shown in the left panel.

'TATA-like', as they did not form a consensus, whereas those conforming to the consensus retain the 'TATA box' designation. We refer to the two elements together as 'TATA elements'.

To assess whether TFIIB was positioned around these TATA-like elements in a canonical manner as seen at bona fide TATA boxes, strand-specific ChIP-exo tags were plotted around each element, separated into panels by 0, 1 or 2 mismatches to the TATA box consensus (Fig. 2). Notably, regardless of its occupancy level, the distribution of TFIIB crosslinking and thus its positioning relative to these TATA-like elements was quite similar to the positioning observed at bona-fide TATA boxes. When the other GTFs were examined, their patterns relative to TATA-like elements were also similar to those found at TATA boxes (Supplementary Fig. 2b), although some downstream differences were observed (addressed later). Thus, as previously seen at the three yeast TATA-box-containing genes *GAL1*, *GAL10* and *HSP82* (refs 23, 24), and in mammalian *in vitro* systems[28], at least the upstream portion of most PICs are positioned with respect to resident TATA elements nearly identically, regardless of their Pol II promoter classification as TATA-box containing or TATA-less. Although the 'TATA-less' designation may be a misnomer, this class of genes is not simply a slight variation of the TATA class, but instead has predominant regulation by TFIID versus SAGA, positive versus negative regulation by chromatin, and lower plasticity of expression[6–9].

TATA–less TSS positioning by nucleosomes

Permanganate reactivity experiments detect open complex formation upstream of the TSS at the *GAL1*, *GAL10* and *HSP82* TATA-box-containing promoters[23,24]. These and other studies[17,29] have led to the notion that Pol II scans downstream from the open complex to find the TSS. In agreement with this being a general mechanism, we find that PICs of TATA-box-containing genes generally reside upstream of the TSS (Supplementary Fig. 2a).

TAFs are largely depleted at TATA-box-containing promoters, although they are not entirely absent (Fig. 3a). The low level of Taf1 that was present tended to be positioned similarly to TBP and other GTFs (Fig. 1b and Supplementary Fig. 2a, b). In contrast, Taf1-enriched/TATA-less promoters (which are related, as shown in Supplementary Fig. 3) showed additional interactions downstream of the TSS that exactly coincided with the size and location of the +1 nucleosome (Fig. 3a). Indeed, Taf1 displayed a more uniform positioning pattern in relation to the TSS and +1 nucleosome than to TATA elements, which suggests that at least part of the TFIID TAF complex engages and is positioned by the +1 nucleosome at TATA-less promoters. Consistent with this, Bdf1, which is considered to be a missing piece of Taf1 (ref. 30), binds to the +1 nucleosome[31]. Furthermore, Bdf1 showed a nearly identical ChIP-exo pattern to that of Taf1 (Supplementary Fig. 4). TFIID–nucleosomal interactions have also been reported in mammalian systems, although the details may differ[32].

If TFIID binds simultaneously to both the +1 nucleosome and a TATA element, then the intervening Pol II would seem to be fenced in by TFIID, thereby limiting its ability to scan DNA. This model predicts that the TSS would reside closer to the TATA element and be positionally restricted relative to the +1 nucleosome, compared to Taf1-depleted/TATA-box-containing promoters. Indeed, the TSS at TATA-less promoters resided ~10–20 bp closer to the TATA element than at TATA-box-containing promoters (Supplementary Fig. 2a).

We also compared the position of TATA elements and TSSs in relation to the +1 nucleosome. We separately examined individual Taf1-depleted/TATA-box-containing and Taf1-enriched/TATA-less promoters (Fig. 3b). Strikingly, at the Taf1-enriched promoters, the TSS was tightly positioned at the border between the 5′ NFR and the +1 nucleosome in comparison to Taf1-depleted/TATA-box-containing promoters. The latter had TSSs distributed across the adjacent nucleosome position, and these nucleosomes were relatively depleted compared to the Taf1-enriched class (Fig. 3a).

Taken together, we interpret these observations as reflecting distinct functions of the +1 nucleosome at the two classes of genes. Nucleosomes and PICs of the TFIID-enriched/TATA-less class might cooperatively assembly, in which case the nucleosome may be instructive for TSS selection by impeding Pol II scanning (Fig. 3c). In contrast, nucleosomes and PICs of the Taf1-depleted/TATA-box-containing class may competitively assemble. This may allow for the greater stochasticity or plasticity of expression that is characteristic of this class[9], in which nucleosome loss would prime the gene for a high level of transcription. A nucleosome competition mechanism removes an impediment to Pol II scanning. Pol II could scan further, thereby allowing productive initiation at specific DNA elements[33]. The transition into a scanning state may be rate limiting in the scanning cycle, as the PIC is detected upstream of the TSS.

GTF depletion in genes and their termini

Although it is clear that GTFs assemble in promoter regions of mRNA genes and disengage Pol II within ~150 bp of the TSS[34–37], it is less

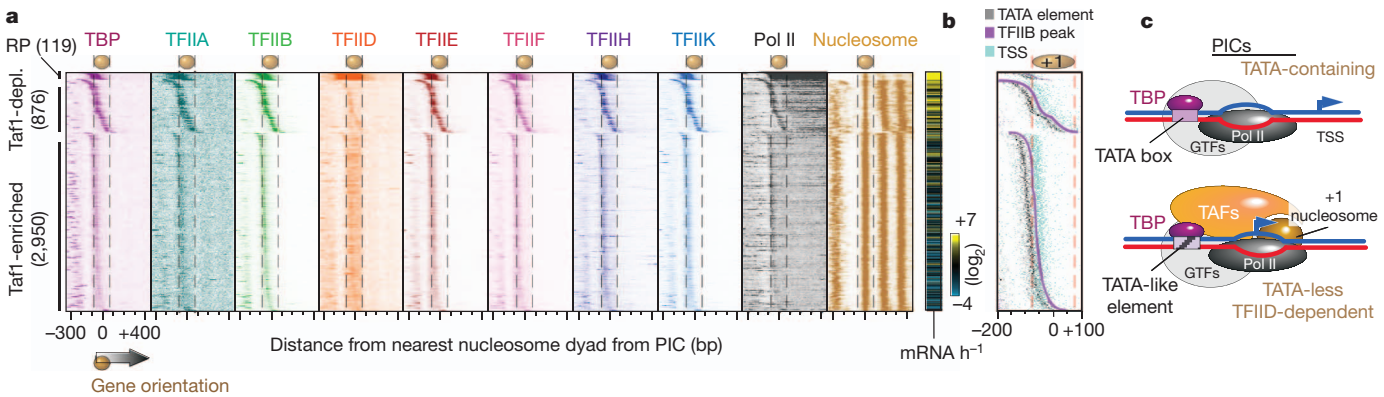

Figure 3 | PIC organization in relation to TFIID and the +1 nucleosome. **a**, GTF occupancy around the nearest nucleosome position (essentially +1) to an mRNA PIC, which were sorted by the distance between the two. Unfiltered tags on each strand were shifted in the 3′ direction by a fixed distance (~8 bp depending on each GTF, 73 bp for nucleosomes), so as to reflect better the points of crosslinking. Taf1-depleted (Taf1-depl.) and Taf1-enriched genes were determined as being distinct clusters when GTF occupancies of all genes were clustered by *k*-means (see Fig. 5a). For all graphs of this type, image resolution is less than the number of rows, resulting in some averaging and thus the appearance of less variance across adjacent rows. See Supplementary Data 2 for underlying values, which can be visualized in Treeview. RP indicates ribosomal protein genes. The right panel shows transcription frequency[48]. The nucleosome borders are denoted by vertical dashed black lines. **b**, Same as panel **a**, but showing an overlay of TATA elements, TFIIB and TSS. **c**, Model of PIC organization at TATA-box-containing and TATA-less/TFIID-dependent genes.

clear to what extent they assemble across the body of genes, or at genes that are either transcriptionally silent or classified as 'dubious'. A whole genome view of GTFs and Pol II is presented in Fig. 4a. Remarkably, PICs were almost entirely excluded from coding regions, regardless of gene activity, whereas Pol II was enriched across gene bodies, as expected. Approximately 90% of dubious open reading frames (ORFs) lacked a canonical PIC organization or contained PICs within the ORF, and thus are unlikely to be coding. Thus, coding region PIC-driven initiation, whether in the sense or antisense direction, is infrequent on the scale of what is seen at mRNA promoters. Moreover, the observed GTF pattern makes clear that Pol II disengages all GTFs at the promoter.

Much less is known of the fate of Pol II at the ends of genes, as it undergoes termination. To examine the 3′ ends of genes, without the complications associated with nearby mRNA promoters, we separated 3′ ends into those having nearby 3′ or 5′ ends of an adjacent gene (Fig. 4b). Within terminal intergenic regions, GTFs were highly depleted, indicating that PICs rarely exist at the 3′ ends of genes at levels seen for mRNA genes (although lower levels do exist).

Remarkably, we find a highly correlated enrichment of Pol II and TFIIH (Ssl2) but not TFIIK (Kin28), at the end of genes within 3′ NFRs (Fig. 4b). Such a physical separation of the TFIIH/Ssl2 and TFIIK/Kin28 submodules of holo-TFIIH in genome-wide binding experiments has not previously been reported, but may be in accord with their biochemistry[19–21]. However, Ssl2 is biochemically separable from the TFIIH core, which therefore prompted us to examine additional core TFIIH subunits, Ssl1 and Tfb1. Surprisingly, both were absent from the ends of genes (Supplementary Fig. 6), although they were present within PICs at promoters. These results suggest that Ssl2, a 3′–5′ helicase, operates independently of holo-TFIIH at the ends of genes. Consistent with a possible role of Ssl2 in transcription termination, Ssl2 has functional interactions with the Hsp90 protein chaperone[38], which has been implicated in the disassembly of the transcription machinery[39].

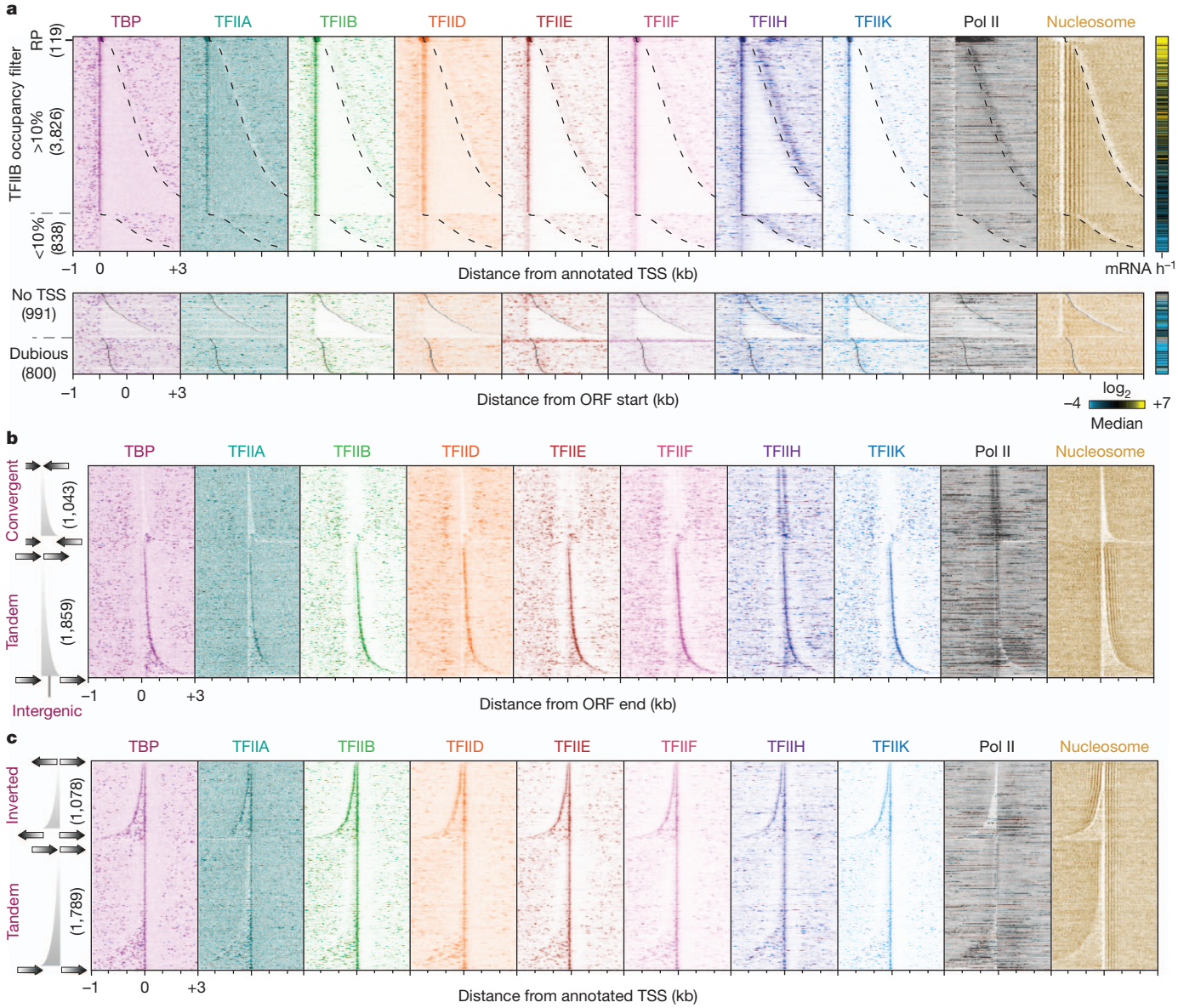

Figure 4 | Genomic view of PICs in relation to genes. **a**, GTF occupancy around transcript and ORF start sites[47], sorted by gene length. See Supplementary Data 3–5 for underlying values. Transcript or ORF ends are indicated by black dashed and solid lines, respectively. The right panel shows transcription frequency[48]. **b**, GTF distribution around the 3′ ends of mRNA genes, sorted by intergenic length, and sectioned by convergent versus tandem gene-pairs. Occupancy at eight reported looped genes[49,50] is shown in Supplementary Fig. 5. **c**, GTF distribution around the TSS of mRNA genes, sorted by intergenic length, and sectioned by inverted versus tandem gene-pairs.

Divergent transcription from distinct PICs

In contrast to coding regions, PICs were abundant in intergenic regions, far beyond what could be accounted for at mRNA promoters (Fig. 4c). Divergent transcription, in which mRNA and non-coding (nc)RNA initiation occurs within the same region but elongates in opposite directions, is well established in eukaryotes[40,41]. However, it has been unclear whether divergent transcription originates from the same PIC site. Conceivably, the entire PIC or a portion thereof might assemble in either direction. As shown in Fig. 4c, even the shortest (~120 bp) 5′ intergenic regions of mRNA genes with inverted orientation were associated with two PICs, one for each mRNA direction. Thus, divergent mRNA transcription originates from two distinct PICs, even when arising from the same NFR.

We next examined the composition and location of PICs associated with mRNA and ncRNA (variously classified as cryptic unstable transcripts (CUTs), stable uncharacterized transcripts (SUTs), Xrn1-sensitive unstable transcripts (XUTs) and antisense (AS))[42–44]. We also examined orphan PICs, which we defined as being >160 bp from any annotated TSS or ORF start site. Nearly all had the same relative composition of GTFs, including Taf1 depletion or enrichment (Fig. 5a), although mRNA PICs generally had higher occupancy levels. GTFs had highly correlated occupancies at all PICs (see Supplementary Fig. 7 for mRNA PICs). ncRNA PICs were generally organized around an adjacent nucleosome, as seen for mRNA PICs (Supplementary Fig. 8). Thus, all mRNA, ncRNA and orphan PICs are compositionally homogeneous with regards to the GTFs (excluding TAFs).

To visualize better the context of the low-occupancy ncRNA PICs with mRNA genes, we marked ncRNA PIC locations by their TATA element, and plotted their directionality with respect to nearby mRNA (Fig. 5b). We observed a general trend where ncRNA and mRNA PICs were positioned in opposite directions 150–200 bp apart. This places both PICs within the same NFR, and thus within the same canonical nucleosome architecture, as seen for two divergent mRNA PICs that share the same NFR (Supplementary Fig. 9).

Low-occupancy ncRNA PICs were also found towards the 3′ ends of mRNA genes, of which the majority were antisense to the mRNA (Fig. 5b), and associated with low expression of the sense mRNA

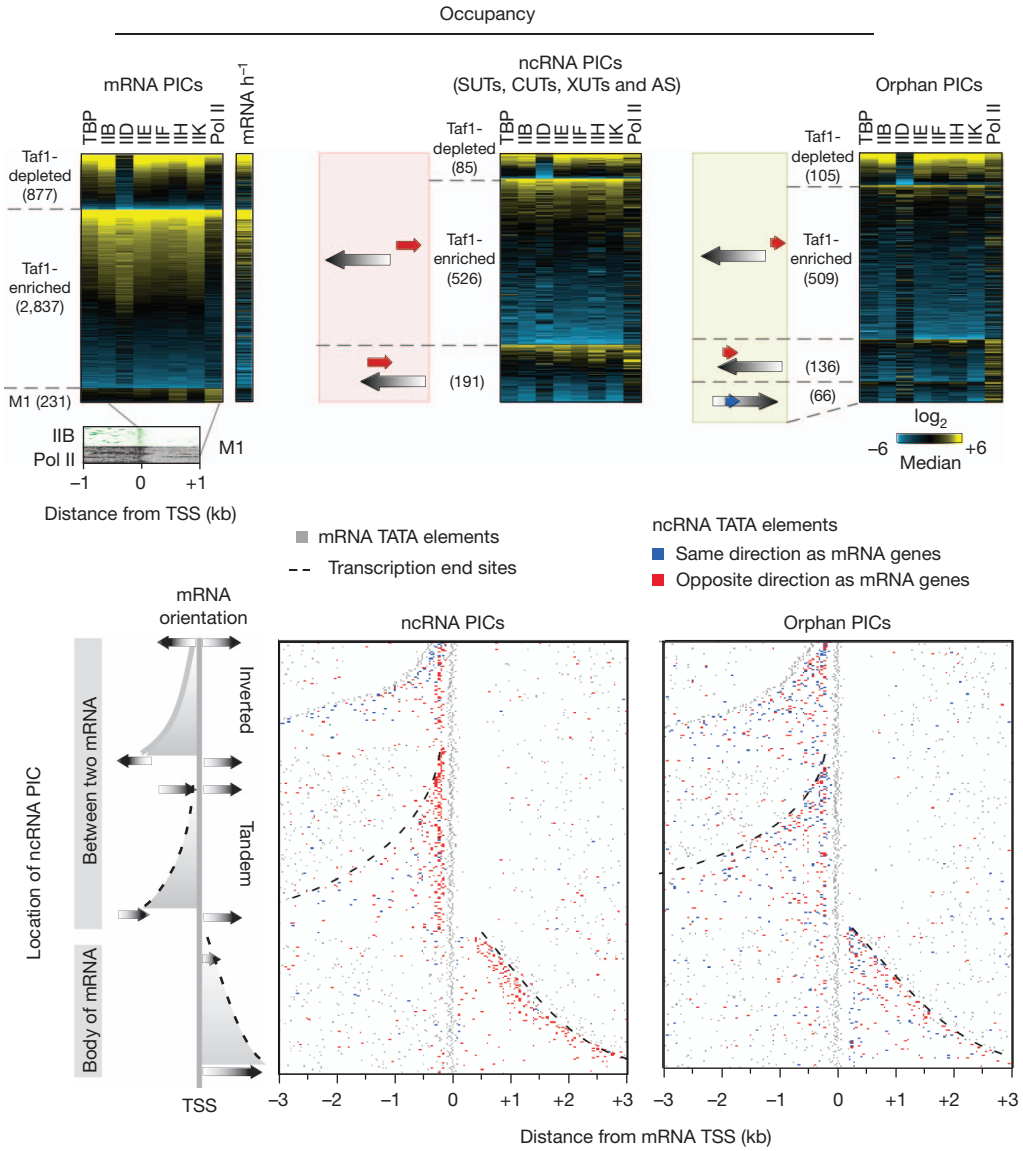

Figure 5 | Distribution of ncRNA PICs. **a**, GTF occupancy levels at PICs for mRNA, ncRNA and orphans, sectioned by *k*-means clustering. Occupancy levels of each GTF were median normalized, $\log_2$ transformed, then sorted by row median. The small M1 group represents a terminating polymerase originating from upstream. **b**, Distribution of ncRNA PICs relative to mRNA genes. mRNA genes were filtered to retain only those having a nearby ncRNA-associated PIC (defined as having >10% of the genome-wide TFIIB average). Plotted are the locations of TATA elements associated with the mRNA (grey), sense-directed ncRNA (blue), and antisense-directed ncRNA (red). Additional plot details are as described in Fig. 4.

(Supplementary Fig. 10a). Thus, ncRNAs, which tend to be antisense[40], are generally associated with repression when residing in gene bodies.

In total, we identified ~6,000 PICs in rapidly growing yeast cells, in which the PICs had an occupancy level of >10% of the genome average. More than 98% of these PICs had a TATA element precisely where TBP bound. Approximately 70% of the identified PICs were associated with mRNA genes (Supplementary Fig. 10b). The remaining ~30% were divided evenly between ncRNA and orphans. At lower detection thresholds, many more low-occupancy PICs could be identified. We do not believe that they represent technical noise, as they are highly enriched in NFRs in which mRNA and ncRNA PICs are found. They might produce low levels of promoter-specific basal transcription.

Unifying principles of PICs

Our data suggest that with the exception of TAFs, PICs are compositionally homogeneous in regards to GTFs at coding and non-coding Pol II transcription units in the yeast genome. PICs differ markedly in occupancy levels, which is in accord with their transcription frequency. PICs tend to form at NFR/nucleosome borders at the 5′ end (and to some extent at the 3′ end) of genes, where they direct either mRNA or ncRNA transcription away from the NFR. As such, an NFR may normally accommodate two divergently oriented PICs at markedly different occupancy and transcription levels. These occupancy levels do not strictly correlate, which suggests largely independent control of the two PICs.

PIC assembly, orientation and positioning may be contributed to in part by the resident TATA element, as well as through sequence-specific factors and co-factors. TFIID-regulated promoters may rely less on TATA element strength, and more on an NFR-adjacent nucleosome for PIC assembly, orientation and positioning. The adjacent nucleosome might also serve to impede a scanning Pol II so that it can productively select a TSS at a focused position just inside the nucleosome. At SAGA-regulated promoters, which tend to contain a consensus TATA box, nucleosome occupancy may be more competitive with PIC assembly, wherein the strength of TBP/TATA interactions would be more important for PIC assembly, orientation and positioning. As such, there would be no nearby nucleosome to impede polymerase scanning, which would allow TSS selection to be controlled by other factors, including DNA sequence.

The emergent concept of ncRNA and the difficulty of distinguishing random transcriptional noise from specific initiation raise the question as to what constitutes a gene[45]. The unambiguous and precise mapping of PIC locations across a genome, as described here, might help define the start of individual genes.

METHODS SUMMARY

Saccharomyces strains (BY4741) bearing TAP-tagged GTF or Pol II subunits (or untagged TBP) were grown to exponential phase in yeast extract peptone dextrose (YPD) media (30 °C to $OD_{600\,nm} = 0.8$), then subjected to 1% formaldehyde crosslinking for 15 min at 25 °C. Cells were harvested and washed. Sonicated chromatin was prepared by standard methods. Standard ChIP methods were used, followed by lambda exonuclease treatment and library construction, as described elsewhere[22]. Libraries were sequenced by an ABI SOLiD sequencer. Figures displaying strand-specific sequencing tags represent the raw data without normalization to input. TFIIB peak calls were made with GeneTrack software[46]. PICs ($n = 6{,}045$) were identified as having a TFIIB peak-pair in at least two out of four biological replicates and having ≥33 sequence tags (>10% of average TFIIB occupancy)[22]. PICs were assigned to the nearest TSS within ± 200 bp, with mRNA[47] having precedence over ncRNA. For this purpose, ORFs lacking a TSS (from the *Saccharomyces* Genome Database) were assigned a hypothetical TSS based on the genome-wide consensus. PICs of ncRNA were assigned to the nearest TSS within ± 200 bp of SUTs, CUTs, ASs and XUTs[42–44], with SUTs/CUTs having precedence over AS/XUTs. To assign directionality to orphan PICs, we compared nucleosome occupancy on the lower versus higher coordinate side of TFIIB locations. If the higher coordinate had higher nucleosome occupancy it was classified as 'sense', otherwise it was 'antisense'. We validated this method by applying it to mRNA PICs, and found >91% of the assignments to be correct. We searched for TATA elements between −80 to +20 bp of the midpoint of 6,045 TFIIB-bound locations on the sense strand, first by searching for the consensus TATAWAWR, then for one and then two mismatches to the consensus. The TATA element closest to −28 bp of a TFIIB peak had precedence if multiple elements were found.

Received 8 October; accepted 20 December 2011.
Published online 18 January 2012.

1. Green, M. R. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem. Sci.* **25,** 59–63 (2000).
2. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56,** 549–561 (1989).
3. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10,** 2657–2683 (1996).
4. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21,** 327–335 (1996).
5. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature Rev. Genet.* **10,** 161–172 (2009).
6. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116,** 699–709 (2004).
7. Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13,** 573–585 (2004).
8. Lee, T. I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405,** 701–704 (2000).
9. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18,** 1084–1091 (2008).
10. Sugihara, F., Kasahara, K. & Kokubo, T. Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **39,** 59–75 (2011).
11. Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22,** 2994–3006 (2008).
12. Dudley, A. M., Rougeulle, C. & Winston, F. The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step *in vivo*. *Genes Dev.* **13,** 2940–2945 (1999).
13. Bhaumik, S. R. & Green, M. R. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters *in vivo*. *Mol. Cell. Biol.* **22,** 7365–7371 (2002).
14. Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462,** 323–330 (2009).
15. Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science* **327,** 206–209 (2010).
16. Li, Y., Flanagan, P. M., Tschochner, H. & Kornberg, R. D. RNA polymerase II initiation factor interactions and transcription start site selection. *Science* **263,** 805–807 (1994).
17. Pardee, T. S., Bangur, C. S. & Ponticelli, A. S. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J. Biol. Chem.* **273,** 17859–17864 (1998).
18. Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274,** 35668–35675 (1999).
19. Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288,** 1418–1421 (2000).
20. Keogh, M. C., Cho, E. J., Podolny, V. & Buratowski, S. Kin28 is found within TFIIH and a Kin28–Ccl1–Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. *Mol. Cell. Biol.* **22,** 1288–1297 (2002).
21. Svejstrup, J. Q., Feaver, W. J. & Kornberg, R. D. Subunits of yeast RNA polymerase II transcription factor TFIIH encoded by the *CCL1* gene. *J. Biol. Chem.* **271,** 643–645 (1996).
22. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein–DNA interactions detected at single nucleotide resolution. *Cell* **147,** 1408–1419 (2011).
23. Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261,** 759–762 (1993).
24. Giardina, C. & Lis, J. T. Dynamic protein–DNA architecture of a yeast heat shock promoter. *Mol. Cell. Biol.* **15,** 2737–2744 (1995).
25. Forget, D., Langelier, M. F., Therien, C., Trinh, V. & Coulombe, B. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell. Biol.* **24,** 1122–1131 (2004).
26. Lagrange, T. *et al.* High-resolution mapping of nucleoprotein complexes by site-specific protein–DNA photocrosslinking: organization of the human TBP–TFIIA–TFIIB–DNA quaternary complex. *Proc. Natl Acad. Sci. USA* **93,** 10620–10625 (1996).
27. Chen, H. T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119,** 169–180 (2004).
28. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19,** 101–110 (2005).

29. 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).
30. Matangkasombut, O., Buratowski, R. M., Swilling, N. W. & Buratowski, S. Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev.* **14,** 951–962 (2000).
31. Koerber, R. T., Rhee, H. S., Jiang, C. & Pugh, B. F. Interaction of transcriptional regulators with specific nucleosomes across the *Saccharomyces* genome. *Mol. Cell* **35,** 889–902 (2009).
32. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131,** 58–69 (2007).
33. Faitar, S. L., Brodie, S. A. & Ponticelli, A. S. Promoter-specific shifts in transcription initiation conferred by yeast TFIIB mutations are determined by the sequence in the immediate vicinity of the start sites. *Mol. Cell. Biol.* **21,** 4427–4440 (2001).
34. Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nature Struct. Mol. Biol.* **17,** 1272–1278 (2010).
35. 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).
36. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Struct. Mol. Biol.* **11,** 394–403 (2004).
37. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408,** 225–229 (2000).
38. Flom, G., Weekes, J. & Johnson, J. L. Novel interaction of the Hsp90 chaperone machine with Ssl2, an essential DNA helicase in *Saccharomyces cerevisiae. Curr. Genet.* **47,** 368–380 (2005).
39. Freeman, B. C. & Yamamoto, K. R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296,** 2232–2235 (2002).
40. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10,** 833–844 (2009).
41. Wei, W., Pelechano, V., Jarvelin, A. I. & Steinmetz, L. M. Functional consequences of bidirectional promoters. *Trends Genet.* **27,** 267–276 (2011).
42. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–1037 (2009).
43. Granovskaia, M. V. *et al.* High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol.* **11,** R24 (2010).
44. van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475,** 114–117 (2011).
45. Gerstein, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17,** 669–681 (2007).
46. Albert, I., Wachi, S., Jiang, C. & Pugh, B. F. GeneTrack—a genomic data processing and visualization framework. *Bioinformatics* **24,** 1305–1306 (2008).
47. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103,** 5320–5325 (2006).
48. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95,** 717–728 (1998).
49. Ansari, A. & Hampsey, M. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19,** 2969–2978 (2005).
50. Singh, B. N. & Hampsey, M. A transcription-independent role for TFIIB in gene looping. *Mol. Cell* **27,** 806–816 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Albert and Y. Li for bioinformatic support, and members of the Pugh laboratory and the Penn State Center for Eukaryotic Gene Regulation for valuable discussions. Sequencing was performed at the Penn State Genomics Core Facility. This work was supported by National Institutes of Health grant GM059055.

Author Contributions H.S.R. performed the experiments and conducted data analyses. H.S.R. and B.F.P. conceived the experiments, analyses, and co-wrote the manuscript.

Author Information Sequencing data have been deposited at the NCBI Sequence Read Archive under accession number SRA046523. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.F.P. (bfp2@psu.edu).