nature microbiology

ARTICLES

PUBLISHED: 1 MARCH 2017 | VOLUME: 2 | ARTICLE NUMBER: 17021

A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase

Katherine Smollett[1], Fabian Blombach[1], Robert Reichelt[2], Michael Thomm[2] and Finn Werner[1]*

The archaeal transcription apparatus is closely related to the eukaryotic RNA polymerase (RNAP) II system, while archaeal genomes are more similar to bacteria with densely packed genes organized in operons. This makes understanding transcription in archaea vital, both in terms of molecular mechanisms and evolution. Very little is known about how archaeal cells orchestrate transcription on a systems level. We have characterized the genome-wide occupancy of the *Methanocaldococcus jannaschii* transcription machinery and its transcriptome. Our data reveal how the TATA and BRE promoter elements facilitate recruitment of the essential initiation factors TATA-binding protein and transcription factor B, respectively, which in turn are responsible for the loading of RNAP into the transcription units. The occupancies of RNAP and Spt4/5 strongly correlate with each other and with RNA levels. Our results show that Spt4/5 is a general elongation factor in archaea as its presence on all genes matches RNAP. Spt4/5 is recruited proximal to the transcription start site on the majority of transcription units, while on a subset of genes, including rRNA and CRISPR loci, Spt4/5 is recruited to the transcription elongation complex during early elongation within 500 base pairs of the transcription start site and akin to its bacterial homologue NusG.

Transcription is a fundamental process in biology, and RNA polymerases (RNAPs) are closely related in all domains of life[1]. The archaeal and eukaryotic systems are near-identical in terms of RNAP subunit composition and architecture, regarding transcription initiation, elongation factors and the molecular mechanisms that govern their activity[2]. The universally conserved core of RNAP resembles a crab-claw-like structure made of the large catalytic subunits Rpo1/2 and the assembly platform including Rpo3/11. The archaeal RNAP shares five to six additional subunits with eukaryotic RNAPII that are absent in bacterial RNAP (ref. 3). This includes the Rpo4/7 stalk module that protrudes from the core enzyme, binds to the nascent RNA and modulates transcription processivity and termination[4]. Archaeal transcription has been studied extensively *in vitro*, but relatively little is known about the genome-wide distribution of RNAP and basal transcription factors and how this correlates with promoter elements and transcription output. A limited number of archaeal promoters have been functionally characterized and seem to rely TATA boxes, B-recognition (BRE) and Initiator elements (Inr)[5,6]. The former two are binding sites for the two basal transcription factors TATA-binding protein (TBP) and transcription factor B (TFB), respectively[2]. Both are strictly required for promoter-directed transcription *in vitro*[7] and homologous to eukaryotic TBP and TFIIB with identical functions but faster dynamics in terms of promoter binding[8]. The third basal transcription factor, transcription factor E (TFE), is homologous to TFIIE, it enhances the stability of the transcription pre-initiation complex (PIC) by catalysing the isomerization of the closed to open complex, during which the DNA strands are separated and the template strand is loaded into the active site of RNAP (refs 9 and 10). The elongation factor Spt4/5, NusG in bacteria, is the only RNAP-associated factor that is conserved throughout the three domains of life. Spt4/5 enhances transcription processivity and possibly functions during promoter escape[11]. Interestingly, *in vitro* experiments revealed that Spt4/5 and NusG are denied access to the PIC by TFE and σ^{70}, respectively[10,12]. Chromatin immunoprecipitation (ChIP) experiments show that yeast Spt4/5 is recruited to RNAP proximal to the promoter, suggesting a role in transition from initiation to elongation[13], whereas *E. coli* NusG is recruited to RNAP during elongation in a stochastic fashion[14].

We applied ChIP followed by high-throughput sequencing (ChIP-seq) to characterize the whole-genome distribution of *Methanocaldococcus jannaschii* (Mja) RNAP and initiation factors TBP and TFB, and to examine the recruitment patterns of Spt4/5 in archaea. To orientate the transcription machinery within the genome, we mapped and analysed global transcription start sites (TSSs) and steady-state RNA levels. We identified positive correlations between BRE/TATA motif strength; binding of TBP and TFB to the promoter; occupancy of RNAP and Spt4/5 within the gene and RNA levels. The elongation factor Spt4/5 showed two different modes of recruitment: early, promoter-proximal recruitment to RNAP (similar to yeast Spt4/5), and a later recruitment during early elongation on ribosomal RNA (rRNA) and CRISPR loci more akin to bacterial NusG.

Results

Organization of the Mja transcriptome. The workflow of the RNA-seq analysis is illustrated in Supplementary Fig. 1a. To characterize the Mja transcriptome we first mapped the genome-wide TSSs using a terminator exonuclease (TEX) RNA-seq approach. We mapped 1,508 TSSs (Supplementary Sections 'Transcription start site mapping of *Methanocaldococcus jannashii*' and 'Organisation of the Mja transcription units') and used our TSS map to annotate 976 transcription units (TUs) that we defined as the sequence spanning from the primary TSS to the stop codon (on mRNA genes) or the annotated 3′ end (on noncoding RNA genes) of the last cistron. A further 138 TUs

[1]University College London, Institute for Structural and Molecular Biology, Gower Street, London WC1E 6BT, UK. [2]Institute of Microbiology and Archaea Center, Universität Regensburg, 93053 Regensburg, Germany. *e-mail: f.werner@ucl.ac.uk

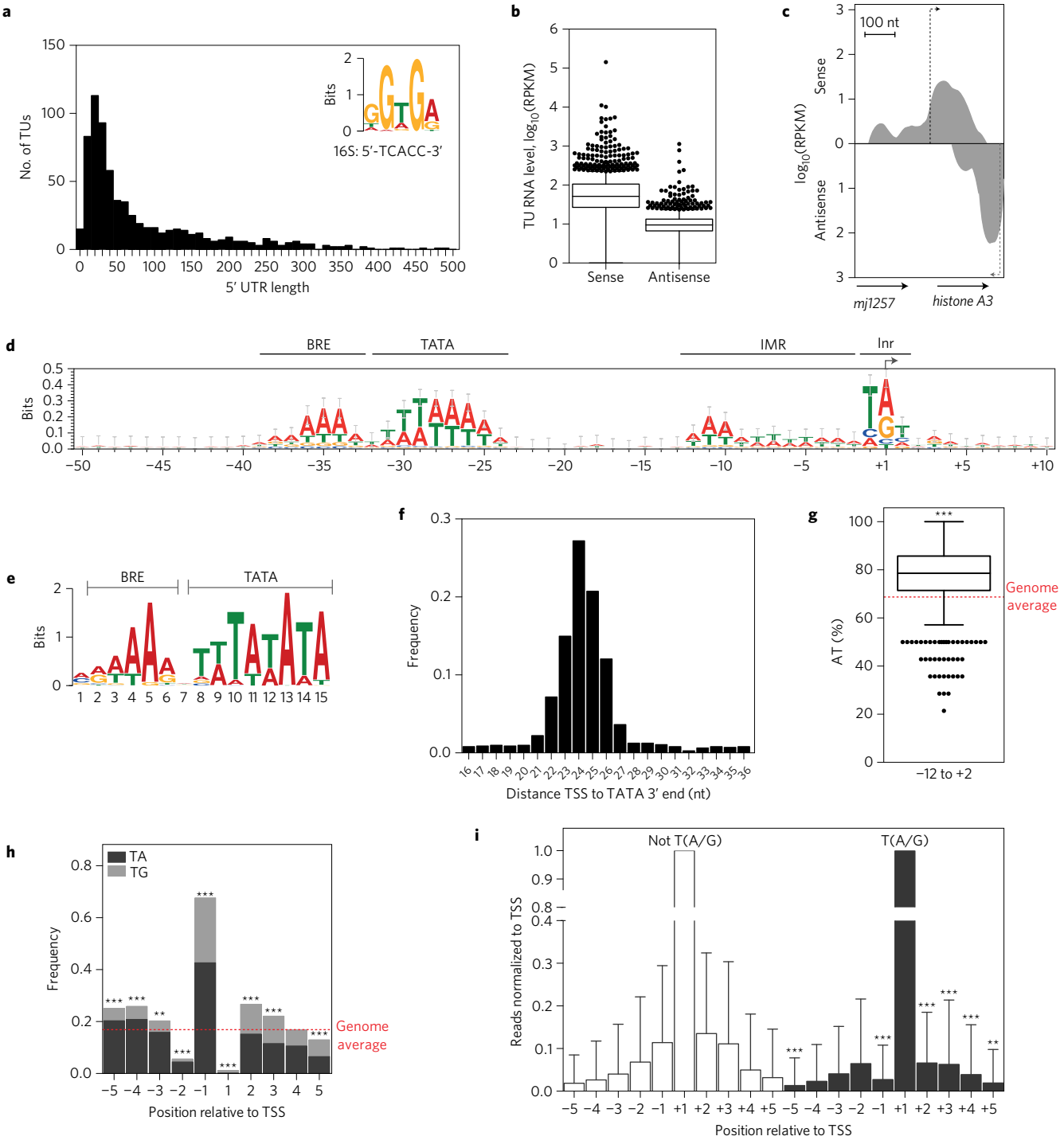

Figure 1 | TSS map and promoter motif analysis. a, The 5′ UTR distance distribution from the primary TSS to the start codon of Mja mRNAs ($n = 689$). Inset: ribosome binding site (RBS) sequence motif identified by the MEME algorithm. For comparison, the complementary sequence of the Mja 16S RNA is shown. **b**, Comparison of sense and antisense RNA levels at all TUs ($n = 1{,}138$). Horizontal lines represent the median and whiskers indicate 1.5× the interquartile range, and individual RPKM values represent the average of two biological replicates. **c**, Strand-specific RNA profiles reveal sense and antisense transcripts on the histone A3 locus. Dotted lines indicate TSS, the average of two biological replicates. **d**, Promoter DNA sequence alignments centred on the TSS ($n = 1{,}508$) reveal regions with a sequence bias corresponding to the BRE/TATA elements, the initially melted region (IMR) and the initiator (Inr) of the promoter. **e**, The BRE/TATA consensus motif identified by MEME-ChIP. **f**, The distance between the 3′ end of the TATA motif and the TSS is centred on 24 nt (TATA at $P < 1 \times 10^{-3}$, $n = 1{,}129$). **g**, AT content distribution of the IMR that exceeds the genome average of 68.7% (red dotted line); the horizontal line within the box represents the median and whiskers indicate 1.5× the interquartile range. Significance according to Wilcoxon signed-rank test ($P < 1 \times 10^{-10}$, $n = 1{,}508$). **h**, Dinucleotide frequency of TA and TG motifs surrounding the TSS. The red dotted line indicates the genome-wide frequency of 0.15, and the significance was assessed by Fisher's exact test ($n = 1{,}507$). **i**, The T(A/G) motif increases the precision of TSS selection. The read counts of all 5′ ends from TEX-treated RNA surrounding assigned TSSs were identified (averaged across two biological replicates) and the reads normalized to the TSS at each position. Data are presented as mean + s.d., $n = 447$ not T(A/G) or 762 T(A/G). Initiation immediately upstream and downstream is fourfold and twofold lower, respectively, for TSS with T(A/G) compared to those without by Wilcoxon rank-sum test. In **g–i**, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

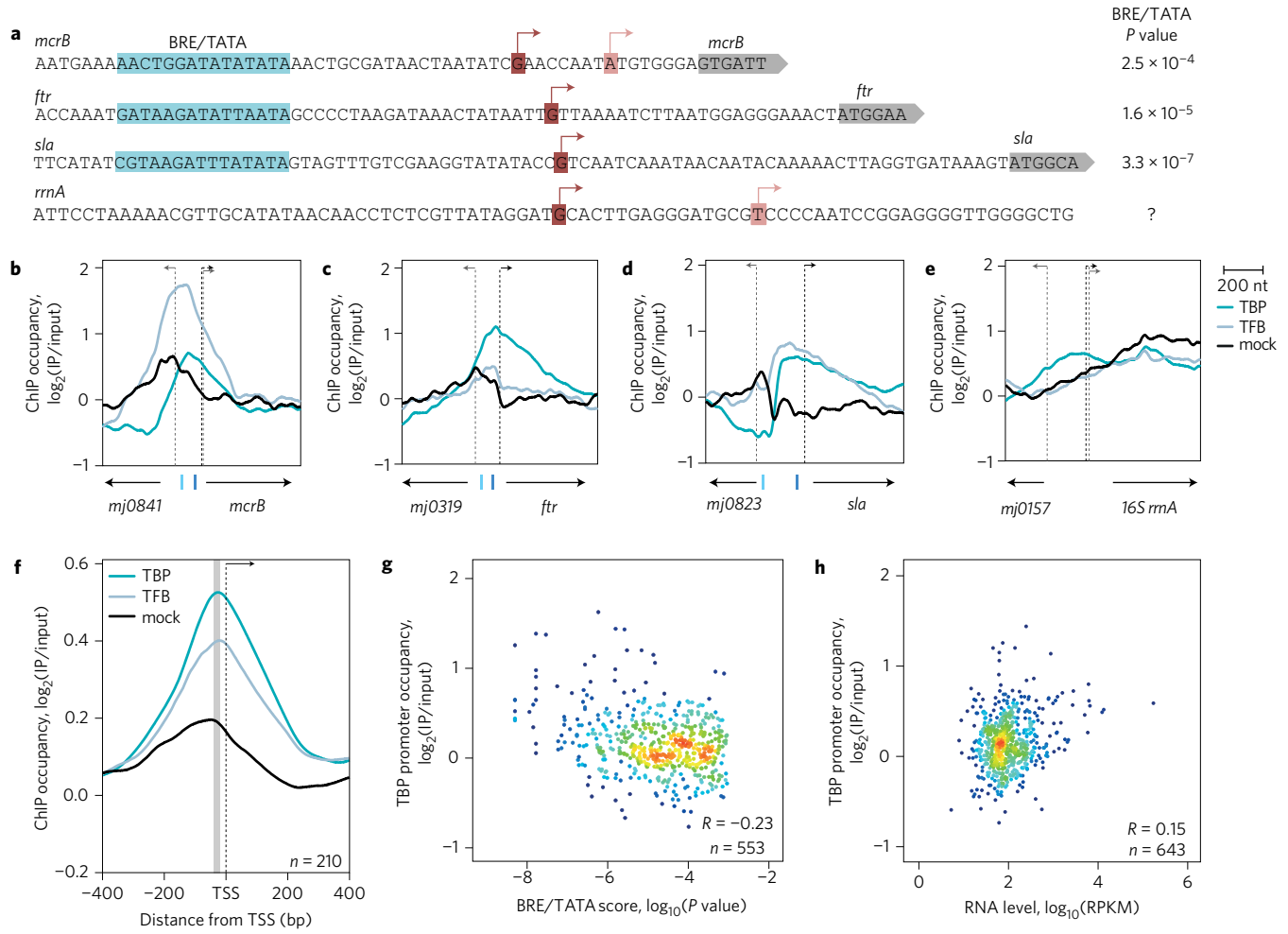

Figure 2 | Correlation between TBP/TFB binding to the promoter and RNA levels. a, The BRE/TATA motifs (highlighted in blue), primary and secondary TSS (red and pink, respectively) and the coding region (grey) of three selected mRNA (*mcrB*, *ftr* and *sla*) and the rRNA promoter. The confidence score (*P* value) for the BRE/TATA motif is indicated to the right of the sequence. **b**–**e**, TBP, TFB and mock control occupancy profiles at the *mcrB* (**b**), *ftr* (**c**), *sla* (**d**) and *rrnA* (**e**) promoter. TSSs are indicated as arrows, with the primary TSS in black. **f**, A metadata analysis shows that the averaged occupancy profiles of TBP and TFB of the top 25% of mRNA TUs (by sense RPKM, $n = 210$) co-locate with the predicted BRE/TATA motif (grey). **g**, Correlation between BRE/TATA score (*P* value) and TBP occupancy. Spearman correlations are indicated on TBP $R = -0.23$, $P = 6 \times 10^{-8}$, $n = 553$. Points are coloured using a density gradient (ranging from blue (low) to red (high)). **h**, Correlation between the TBP occupancy and RNA levels (sense RPKM for all TUs with detectable transcript; average of two biological replicates). Spearman correlations indicated on TBP $R = 0.15$, $P = 1 \times 10^{-4}$, $n = 643$. Occupancy data in **a**–**h** represent the average of four (TBP) or two (TFB and mock) technical replicates.

were predicted based on gene orientation but were not associated with a TSS. We identified several novel genes encoding open reading frames (ORFs) and ncRNAs (listed in Supplementary Tables 3 and 4). Mja TUs are organized into a combination of single- and multicistronic operons (Supplementary Fig. 2e). The majority of protein-encoding genes encode long untranslated leader regions (5′ UTR) with only 16 mRNAs (1.9%) being defined as leaderless (<5 nucleotides (nt), Fig. 1a). Within the 5′ UTRs we identified ribosome binding sites (RBSs) in 54% of mRNA genes (Fig. 1a). To determine the global steady-state RNA levels, we next calculated the reads per kilobase of transcript per million mapped reads (RPKM) values for each TU. Using a cutoff value of RPKM > 1, we defined 63% of the TUs as transcriptionally active (adjusted $P < 0.05$; Supplementary Section 'The Mja transcriptome' and Supplementary Table 3). The two rRNA operons had the highest RPKM values and account for 80% of all mapped reads. Several small ncRNA genes including tRNAs were detected at low levels, but may be misrepresented due to loss during size selection of library preparation. We could detect antisense transcription in Mja (Fig. 1b), but the majority of antisense transcripts were not associated with a TSS, possibly due to their rapid degradation. We identified 12 antisense TUs with assigned TSS, including the Mja histone A3 gene (Fig. 1c and Supplementary Table 4). Both sense and antisense A3 transcripts were highly abundant, hinting at a possible regulation of A3 expression by antisense transcription. Northern blotting confirmed the presence of both sense and antisense A3 transcripts covering the histone A3 ORF (Supplementary Fig. 2f).

Promoter sequence elements and start site selection. Alignment of DNA sequences surrounding the TSSs identified two regions with a sequence bias, corresponding to the BRE/TATA elements and the initially melted region (IMR) that includes the initiator (Inr) surrounding the TSS (Fig. 1d). Sequence motif analysis of these DNA sequences revealed a global BRE/TATA consensus (Fig. 1e). These elements could be identified upstream of 76% of TSSs using a stringent motif confidence score (motif $P < 1 \times 10^{-3}$, Supplementary Fig. 3a), including all primary TSSs of TUs defined as transcriptionally active. BRE/TATA motifs are centred on register +24 relative to the TSS; this distance is conserved from

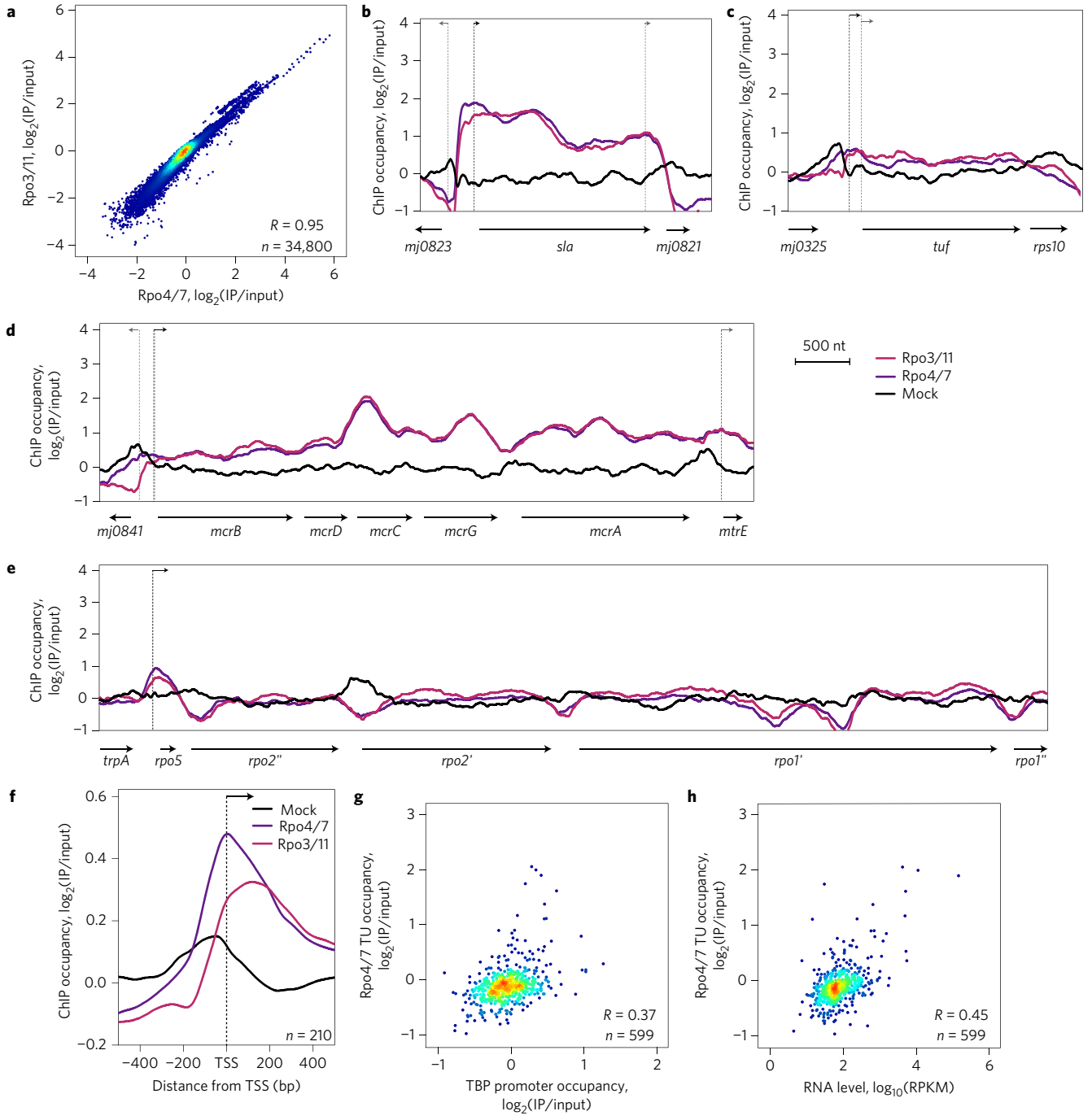

Figure 3 | The Rpo4/7 stalk and RNAP core remain associated through the transcription cycle. a, The correlation between the occupancy of RNAP subunit complexes Rpo4/7 and Rpo3/11 is very strong across the genome, as indicated by Spearman correlations ($P < 1 \times 10^{-10}$, $n = 34{,}800$). **b**–**e**, RNAP occupancy profiles on representative TUs: the *sla* (**b**), *tuf* (**c**), *mcr* (**d**) and RNAP subunit operon (**e**). Arrows indicate TSS (primary in black). **f**, Averaged occupancy profiles of Rpo4/7, Rpo3/11 and mock control at the top 25% of mRNA TU (by sense RPKM, $n = 210$). **g**, Correlation between the TBP promoter occupancy (TSS ± 250 bp) and RNAP TU occupancy (TSS + 250 to TU end) for all TUs (RPKM > 1). Spearman correlation $R = 0.37$, $P < 1 \times 10^{-10}$, $n = 599$. **h**, Correlation between steady-state RNA levels (sense RPKM for all TU RPKM > 1, average of two biological replicates) and RNAP (Rpo4/7) occupancy within the body of each TU, Spearman correlation $R = 0.45$, $P < 1 \times 10^{-10}$, $n = 599$. Occupancy data in **a**–**h** represent the average of four (Rpo4/7), three (Rpo3/11) or two (mock) technical replicates.

archaea to metazoans[15] (Fig. 1f). During open complex formation the two DNA strands of the IMR of the promoter from −12 to +2 are separated[9,16–18]. Alignments show that this region is enriched in A and T residues (80 ± 12% AT, genome average 69% AT, Fig. 1g). The AT content of the IMR does not correlate with RNA levels (Supplementary Fig. 3b). The Inr element formed by the bases surrounding the TSS showed a strong bias for the sequence T(A/G) at position −1/+1 (Fig. 1d), but, similar to the IMR, did not correlate with RNA levels (Supplementary Fig. 3c). Examining the dinucleotide frequency within this region revealed that TA and TG are not only highly enriched at position −1/+1 (combined >60%, compared to the genome average of 15%), but also strongly disfavoured at the neighbouring positions (−2/−1 and +1/+2, Fig. 1h). Conservation of the T(A/G) motif is independent of the

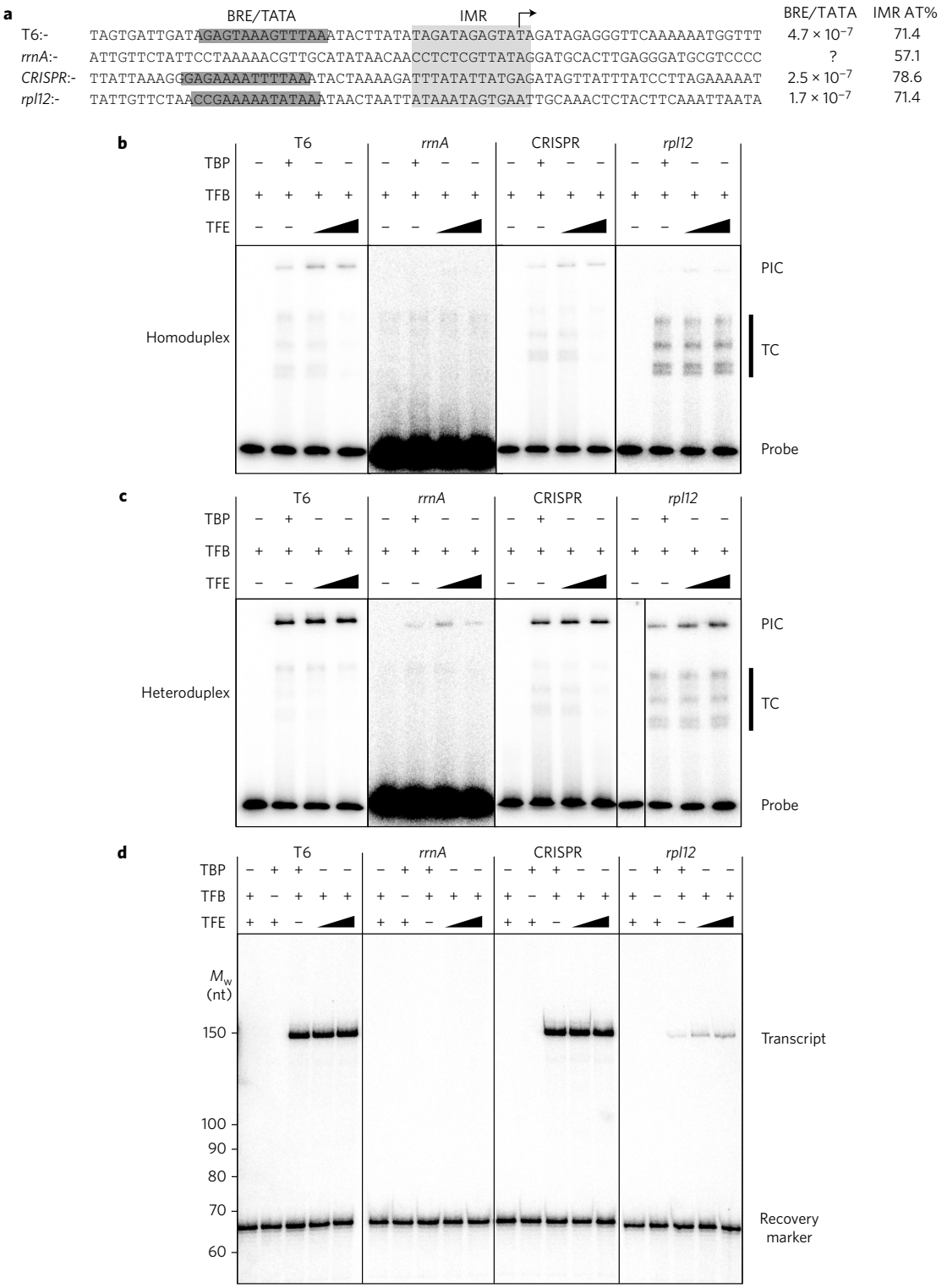

Figure 4 | PIC formation and promoter strength *in vitro*. **a**, Alignment of SSV T6 model promoter and representative Mja promoters including ribosomal RNA (*rrnA*), CRISPR and mRNA (ribosomal protein *rpl12*) promoters. BRE/TATA motifs are shown in dark grey with *P* values indicated. IMR is highlighted in light grey with AT% indicated. **b**, EMSA showing PIC formation on promoter templates in **a**. **c**, EMSAs using heteroduplex promoter variants. PIC indicates the transcription PIC, and TC the ternary DNA–TBP–TFB complexes. Exposure is adjusted to account for diverse signal intensities. **d**, Promoter-directed *in vitro* transcription assays. Promoter templates shown in **a** were fused to C-less cassette, resulting in transcripts with lengths of 150 nt (T6), 157 nt (*rrnA*) and 152 nt (CRISPR and *rpl12*). Representative examples of two technical replicates are shown.

distance between the TATA box and the TSS (Supplementary Fig. 3d). Because these results suggest that the Inr dictates TSS selection, we analysed the TSS specificity on promoters with and without the Inr motif. Promoters with an Inr sequence T(A/G) showed up to fourfold lower levels of transcription initiation at neighbouring positions compared to promoters without the T(A/G) motif (Fig. 1i). In summary, while the BRE/TATA motifs facilitate transcription PIC assembly, the Inr fine tunes TSS selection. A

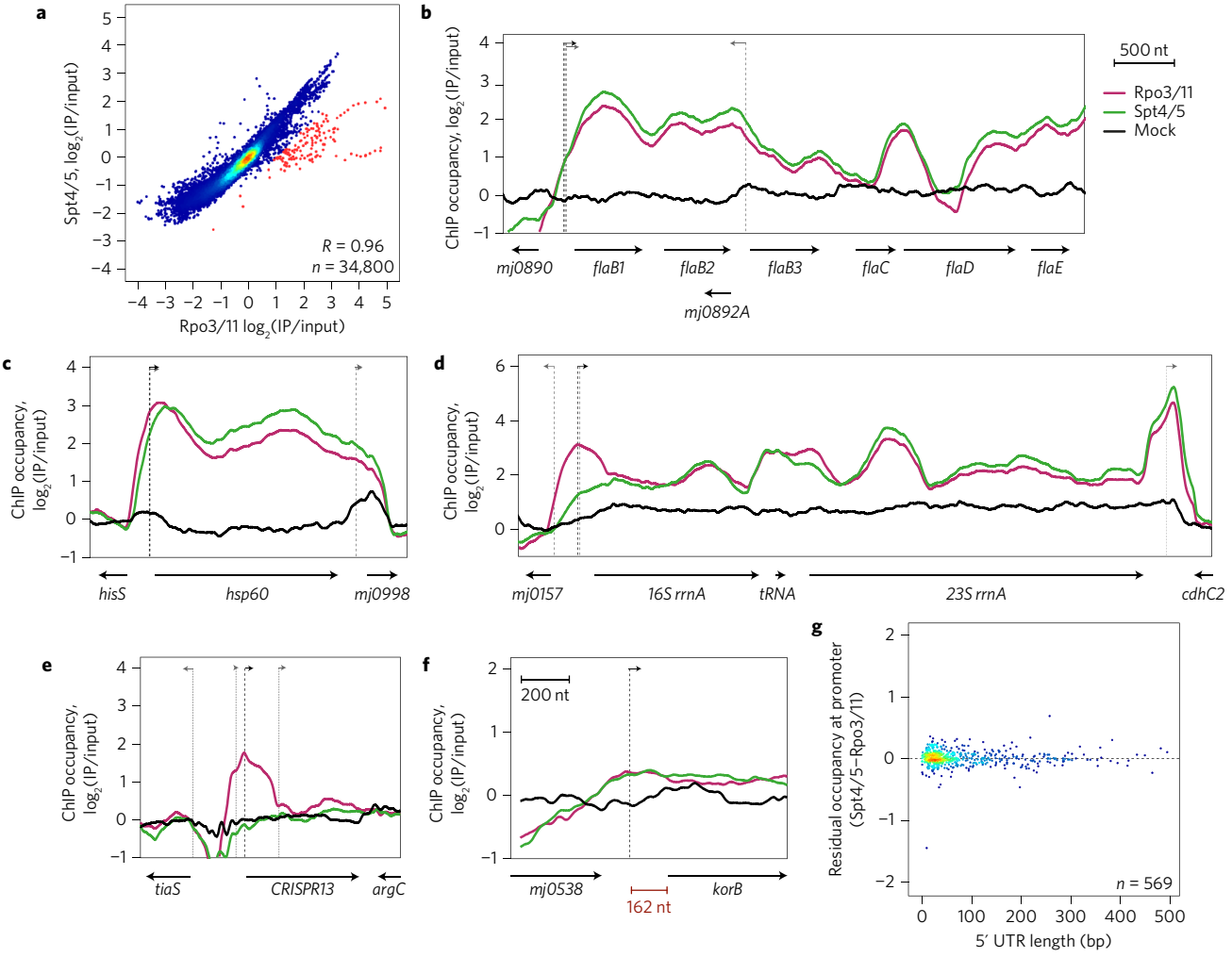

Figure 5 | Archaeal Spt4/5 is a general elongation factor that is recruited to RNAP via two distinct modes. a, Spt4/5 and RNAP occupancy correlates very strongly across the whole genome. Data points of sub-stoichiometric Spt4/5:RNAP occupancy, with Spt4/5 occupancy more than 1 $\log_2$(IP/input) lower than RNAP occupancy, are indicated in red, Spearman correlations $R = 0.96$, $P < 1 \times 10^{-10}$, $n = 34{,}800$. **b**–**f**, Spt4/5 occupancy profiles reflect two recruitment modes of Spt4/5 exemplified by the archaealleum (**b**) and rRNA operons (**d**). Representative RNAP and Spt4/5 occupancy profiles on the *fla* (**b**), *hsp60* (**c**), *rrnA* (**d**) and *CRISPR13* operon (**e**) and larger-scale plot of the long 5′ UTR gene *korB* (**f**). Arrows indicate TSS. **g**, 5′ UTR length does not affect the difference between Spt4/5 and Rpo3/11 occupancy proximal to the promoter of TUs (RPKM > 1), $n = 569$. Occupancy data in **a**–**g** represent the average of three (Rpo3/11 and Spt4/5) or two (mock) technical replicates.

comparison with other archaeal promoters[19–24] (Supplementary Fig. 4) reveals that the TATA consensus is largely conserved across the archaea, but the significance of IMR and Inr are subject to variation[25].

TBP and TFB binding to the Mja BRE/TATA motifs. We determined the global occupancy of the essential initiation factors TBP and TFB by ChIP using polyclonal antibodies raised against recombinant proteins followed by high-throughput sequencing (ChIP-seq). The workflow and detailed methods are described in Supplementary Section 'Occupancy profiling of the Mja general transcription machinery using ChIP-seq' and Supplementary Fig. 1b. Figure 2a–e shows the ChIP-seq profiles of four representative promoters, ranging from promoters that show a distinct and defined increased TBP and TFB occupancy centred on the BRE/TATA motifs (*mcrB* and *ftr*, Fig. 2b,c), those that display broader profiles, but are distinct from the mock control (*sla*, Fig. 2d), to promoters that do not show any increased occupancy at all (*rrnA*, Fig. 2e). Averaging the TBP/TFB occupancy profiles centred on the TSS of the top 25% expressed mRNA TUs (by RPKM) shows distinct TBP and TFB peaks (Fig. 2f). The apex of both peaks concurs with the location of the BRE/TATA motifs, which confirms the validity of our TBP/TFB profiling analysis (Fig. 1f). The profile of the mock immunoprecipitation (IP) control demonstrates that, although the mock shows a slight increase in signal, both TBP and TFB signals are above the background (Fig. 2f). To validate our results we compared the data to a subset of experimentally characterized promoters. A total of 19 transfer RNA (tRNA) and 12 mRNA Mja promoters have been analysed quantitatively *in vitro* with respect to the formation of DNA–TBP–TFB complexes using electrophoretic mobility shift assays (EMSAs)[6]. There is a strong correlation between the published *in vitro* binding data and the *in vivo* occupancy across their promoter regions (TSS ± 250 base pairs (bp); TBP $R = 0.7$, $P = 1.1 \times 10^{-5}$; TFB $R = 0.61$, $P = 2.6 \times 10^{-4}$; Supplementary Fig. 5c), which also implies that *in vitro* EMSAs are a good indicator for the binding of TBP and TFB to promoters *in vivo*. To relate the strength of the TBP/TFB binding to the sequence of the BRE/TATA motifs, we compared the confidence score (*P* value) of the BRE/TATA motif of each promoter to the TBP and TFB ChIP signal (Fig. 2g). The BRE/TATA score showed a weak but significant correlation with

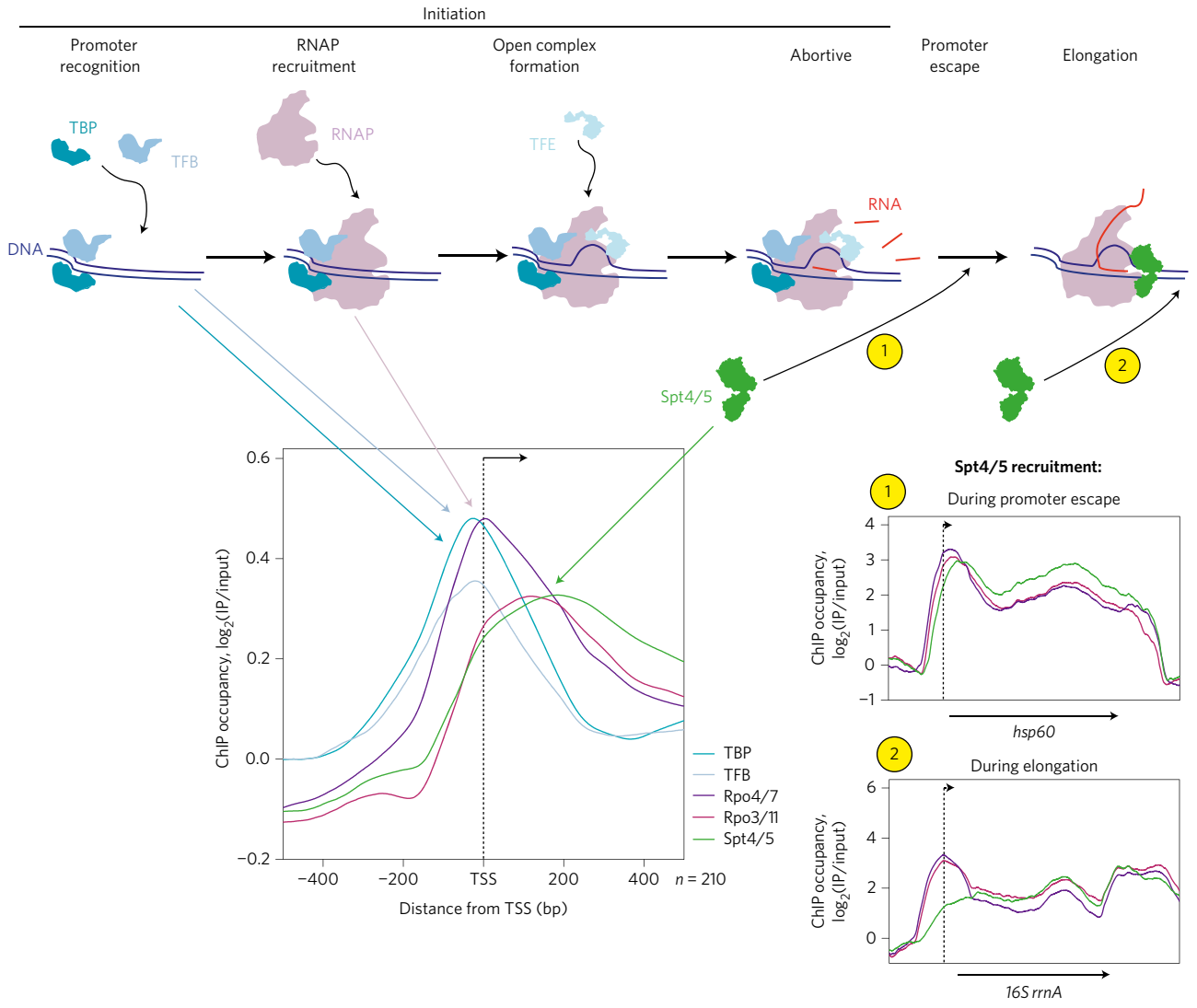

Figure 6 | Initial stages of the transcription cycle in archaea. The average occupancy profiles of TBP, TFB, RNAP and Spt4/5 on the top 25% of mRNA TUs reflect the initial stages of the transcription cycle. TBP and TFB are bound to the TATA and BRE promoter elements 24 nt upstream of the TSS, which in turn recruit RNAP to form the PIC. Subsequently, two modes of Spt4/5 recruitment could be distinguished on different genes: (1) on the majority of genes Spt4/5 is recruited 'early', probably during promoter escape; (2) on the rRNA operons and CRISPR Spt4/5 is recruited 'later' offset from TSS in the downstream direction, probably occurring during transcription elongation.

the TBP/TFB occupancy (TBP $R = -0.23$, $P = 6 \times 10^{-8}$; TFB $R = -0.30$, $P < 1 \times 10^{-10}$; mock $R = -0.08$, $P = 0.03$), but only a very weak correlation with TU RNA levels (Fig. 2h, TBP $R = 0.15$, $P = 1 \times 10^{-4}$; TFB $R = 0.15$, $P = 8.1 \times 10^{-5}$; no correlation with mock, $P > 0.05$).

RNAP occupancy correlates with RNA levels. We characterized the global occupancy of RNAP with two polyclonal antibodies directed against two distinct RNAP subcomplexes. The pairwise genome-wide correlation between occupancy of Rpo4/7 stalk and Rpo3/11 assembly platform subunits was calculated using 250 base pairs (bp) windows with a 50 bp overlap. The Rpo4/7 and Rpo3/11 signals correlate very strongly with each other ($R = 0.95$, $P < 1 \times 10^{-10}$, Fig. 3a). To visualize the RNAP occupancy within TUs we plotted the ChIP-seq profile as occupancy per nucleotide across the genome. The RNAP ChIP-seq profiles of individual loci emphasize very diverse profiles on different genes (Figs 3b–e and 5); for example, occupancy is high on the *sla* and *mcr* TUs (Fig. 3b,d), but is low on the *tuf* and *rpo* operons (Fig. 3c,e). A metadata analysis averaging the RNAP occupancy centred on the TSS reveals that the Rpo4/7 signal appears ~100 bp upstream of the Rpo3/11 signal (Fig. 3f). Promoter-bound TBP and TFB are strictly required for the recruitment and subsequent loading of RNAP into the TU *in vitro*. In good agreement, the occupancy of TBP and TFB at the promoter correlated with RNAP occupancy within the TU (Fig. 3g; Rpo4/7 compared to TBP, $R = 0.37$, $P < 1 \times 10^{-10}$; Rpo4/7 to TFB $R = 0.3$, $P < 1 \times 10^{-10}$; mock $R = 0.1$, $P = 0.02$). Finally, the RNAP occupancy within TUs correlated moderately well with RNA levels (Fig. 3h; Rpo4/7, $R = 0.45$, $P < 1 \times 10^{-10}$; Rpo3/11 $R = 0.48$, $P < 1 \times 10^{-10}$; mock $R = -0.15$, $P = 3.4 \times 10^{-4}$).

***In vitro* PIC assembly.** Surprisingly, the two Mja rRNA promoters (*rrnA* and *rrnB*) have no identifiable BRE/TATA motifs and do not show a strong TBP/TFB ChIP signal (Fig. 2a,e). This suggests that they are weak promoters, which is in stark contrast to the high RNAP occupancy and RNA levels. To probe the strength of Mja *rrn* promoters *in vitro*, we monitored PIC formation on the *rrnA* promoter using EMSA and promoter activity using transcription assays. For comparison, we included a representative Mja mRNA

promoter (*rpl12*), which is associated with high RNAP occupancy and RNA level, an Mja CRISPR promoter, which has high RNAP occupancy and the well-characterized viral SSV (Sulfolobus spindle-shaped virus) T6 promoter (Fig. 4a)[7,16,26,27].

The SSV T6 and CRISPR promoters recruit RNAP in a TBP/TFB-dependent fashion, and the addition of TFE stimulated the PIC in EMSA experiments (Fig. 4b). The *rpl12* promoter, which has a similar BRE/TATA consensus but lower IMR AT% than the CRISPR promoter, formed a weak PIC in the absence of TFE. In contrast, the *rrnA* promoter was not able to form a stable PIC. Heteroduplex promoter variants include a 4 bp non-complementary region (−3 to +1), mimic the open complex, and enhance PIC stability[16,26]. These variants enabled PIC formation at all four promoters, including *rrnA* (Fig. 4c). Introducing mutations into the TATA sequence abolished or dramatically reduced PIC formation on all promoters (Supplementary Fig. 6a,b). We used promoter-directed *in vitro* transcription experiments to complement the promoter-binding experiments. The results from both assays mirrored each other; although the SSV T6, *rpl12* and CRISPR promoters resulted in large amounts of transcripts with the correct size, the *rrnA* promoter was inactive (Fig. 3d). In conclusion, in contrast to the *in vivo* analysis, the *in vitro* transcription experiments show a direct link between promoter motifs, the recruitment of stable PIC, and promoter strength.

Spt4/5 is a general elongation factor with two distinct recruitment modes. We carried out a ChIP-seq analysis to characterize the global occupancy of the transcription elongation factor Spt4/5. The pairwise correlation between genome-wide occupancies of Spt4/5 and RNAP is very strong (Fig. 5a; Rpo3/11 $R = 0.96$, $P < 1 \times 10^{-10}$; Rpo4/7 $R = 0.95$, $P < 1 \times 10^{-10}$; mock $R = 0.035$, $P < 1 \times 10^{-10}$). Furthermore, a comparison of RNAP and Spt4/5 ChIP-seq profiles on individual TUs (by plotting their per nucleotide occupancy) demonstrates that Spt4/5 closely mirrors the undulating pattern of RNAP occupancy, which probably reflects pausing and varying transcription processivity (Fig. 5b,c). This behaviour suggests that Spt4/5 stably associates with the transcription elongation complex (TEC) *in vivo*. To detect any potential heterogeneity in the genome occupancy of RNAP and Spt4/5, we identified genome locations characterized by a lower Spt4/5:RNAP occupancy ratio (red dots in Fig. 5a). The overlapping 250 bp windows were merged to identify 23 separate genome regions with significantly lower Spt4/5 than RNAP occupancy (adjusted $P < 0.05$, Supplementary Table 5). These regions included 18 of the 20 CRISPR loci, both rRNA operons (*rrnA* and *rrnB*), two annotated small non-coding RNA genes, and *mj0496* (uncharacterized ORF). Closer scrutiny of these regions revealed that the lower Spt4/5:RNAP occupancy ratio is restricted to the promoter-proximal region of the gene, with the Spt4/5 profile matching that of RNAP from ~500 bp downstream of the promoter onwards (Fig. 5d,e and Supplementary Table 5).

The bacterial Spt5 homologue NusG aids the coupling of transcription and translation by interacting with the RNAP and the ribosome[28,29]. Similarly, transcription and translation are coupled in archaea[30]. We tested whether the recruitment of Spt4/5 to TECs on protein-encoding genes was influenced by recruitment of the ribosome to the RBS by analysing Spt4/5 occupancy on mRNA genes with long 5′ UTRs. The 5′ UTR of the *korB* gene is 162 bp long, but Spt4/5 is recruited symmetrically with RNAP close to the TSS and not further downstream at the RBS (Fig. 5f). To explore this globally we subtracted the RNAP- from the Spt4/5 occupancy at each mRNA promoter and plotted the value against the length of the 5′ UTR. If Spt4/5 recruitment was aided by the ribosome we would expect the difference in occupancy to increase with 5′ UTR length; however, no difference was observed (Fig. 5g). In conclusion, Spt4/5 follows two modes of recruitment (Fig. 6), in proximity to the promoter on the majority of TUs and several hundred base pairs downstream of the TSS on a subset of genes.

Discussion

We present the first comprehensive genome-wide analysis of transcription in archaea by characterizing (1) the occupancy of RNAP and basal transcription factors, (2) the transcriptome, including a TSS map, and (3) a promoter motif analysis, all in the same organism.

We identified 1,508 TSSs in *Mja* and could account for 88% of TSSs of the 1,114 predicted TUs. TSS analysis reveals that *Mja* mRNAs have long 5′ UTRs indicative of extensive riboregulation by sRNA and riboswitches. This pattern is similar to other methanogens, including *Methanosarcina mazei*, *Methanolobus psychrophilus*, *Thermococcus kodakarensis* and *Pyrococcus furiosus* and different from Sulfolobales and halophilic archaea that are characterized by leaderless mRNAs[19–23,31–34]. The assembly of the PIC *in vitro* is strictly dependent on the binding of TBP and TFB to TATA and BRE motifs of archaeal promoters, respectively. Our *in vivo* analysis reveals the prevalence of BRE/TATA motifs, suggesting that they are the dominant promoter elements in archaea. This is in contrast to eukaryotes, where conventional TATA motifs are absent at the majority of promoters[35]. We also reveal the importance of downstream sequences including the IMR and the 3 bp Inr element that increases the accuracy of TSS selection, while not correlating with RNA levels. Thus far, the role of the archaeal Inr has only been studied *in vitro*, mainly with mutated variants of the viral SSV1 T6 model promoter[36,37]. Our systems data reveal that the Mja Inr has a bias for T(A/G) at registers −1/+1. This preference for pyrimidine and purine nucleotides is a universally conserved promoter feature, which reflects the high degree of conservation between the RNAP active site architectures in the three domains of life[15,38,39]. The elevated AT content of the IMR favours local DNA melting, and experimental evidence shows that the IMR sequence affects promoter strength at individual promoters *in vitro*[9,25]. However, on a global level, the AT content of the Mja promoter IMR does not correlate with RNA levels, and it is thus unlikely that the IMR's AT content alone limits promoter strength *in vivo*.

Having explored the sequence characteristics of archaeal promoters, we characterized the association of RNAP, TBP, TFB and the elongation factor Spt4/5 with the genome. The averaged occupancy profiles of highly expressed genes illustrate the early stages of the archaeal transcription cycle with the stepwise assembly of the PIC, RNAP and Spt4/5 recruitment and promoter escape (Fig. 6). The individual RNAP profiles in different TUs are very diverse, including regions of high and low occupancy proximal to the promoter motifs and within TUs, which probably reflects variations in promoter recruitment, efficiency of escape, processivity and pausing[40]. It has been proposed that the yeast RNAPII RPB4/7 stalk reversibly associates with the RNAP core. Our ChIP-seq results demonstrate that both Rpo4/7 and Rpo3/11 are co-localized across the genome, suggesting that the stalk remains associated with the RNAP core as it progresses through the transcription cycle. The fact that Rpo4/7 is slightly offset upstream from Rpo3/11 signals at TSSs is probably due to epitope occlusion of the latter in the PIC (refs 11, 16). The molecular mechanisms of archaeal Spt4/5 have been characterized in some detail *in vitro*[10,17,41]. Our ChIP-seq results demonstrate that Spt4/5 associates with elongating RNAPs throughout the genome, behaving like an 'honorary' RNAP subunit on all genes, protein-encoding as well as non-coding RNA genes, meaning that Spt4/5 fulfils the criteria of a general elongation factor. By comparing the ChIP-seq profiles of RNAP and Spt4/5, two distinct modes of Spt4/5 recruitment become apparent, either (1) proximal to promoter and just offset from the TSS, or (2) further downstream within the first 500 bp of the TU (Fig. 6). All multi-subunit RNAP face a similar mechanical engineering challenge: a network of interactions

between promoter-bound initiation factors (TBP/TFB/TFE) and RNAP is crucial to enable efficient recruitment of RNAP during early initiation. However, these interactions need to be disrupted to allow RNAP to escape from the promoter[11]. As Spt4/5 and the initiation factor TFE bind to the RNAP clamp in a mutually exclusive manner *in vitro*[10,11], Spt4/5 recruitment proximal to the TSS could assist promoter escape of RNAP by displacing TFE. Our attempts to ChIP TFE were unsuccessful, despite the use of several independent antibody preparations, so we could not directly characterize the swapping of Spt4/5 and TFE *in vivo*. However, Spt4/5 mode (1) does support recruitment during promoter escape, and not during elongation. ChIP analyses from eukaryotic systems are in agreement with promoter-proximal recruitment of Spt4/5 (ref. 13) and the swapping with TFIIE proximal to the promoter[42,43]. Our results show notable exceptions to mode (1). In mode (2) the Spt4/5 occupancy does not match RNAP occupancy until several hundred base pairs downstream of the TSS; these include the two ribosomal RNA operons that account for 80% of the total RNA in the cell and the abundant CRISPR loci. In contrast to Mja Spt4/5, *E. coli* NusG is recruited during elongation at most TUs, but proximal to rRNA promoters due to the assembly of antitermination complexes including NusA, B and E, other ribosomal proteins, some of which are conserved in archaea[14,44]. rRNA operons and CRISPR regions differ from coding genes as templates for transcription in several regards, such as absence of coupled translation, strong secondary-structure content, co-transcriptional processing, and ribosome biogenesis. Unidentified rRNA and CRISPR promoter-specific transcription activators could enhance RNAP recruitment, stabilize the PIC, or interact with the RNAP clamp and possibly enhance promoter escape. This notion is supported by our finding that Mja rRNA promoters have surprisingly poor BRE/TATA motifs and have very low activity *in vitro*, in apparent conflict with the high steady-state levels of rRNA and RNAP occupancy on rRNA operons *in vivo*. The *Sulfolobus solfataricus* and *P. furiosus* rRNA promoters have defined BRE/TATA motifs and are very strong *in vitro*[9,27,45], whereas bacterial rRNA promoters tend to form unstable PICs, making them more amenable to regulation[46].

A quantitative analysis of the transcriptome reveals that 700 of the 1,114 TUs (63%) contain detectable transcripts, under the optimal growth conditions used. We found only a weak correlation between BRE/TATA motif scores or TBP/TFB occupancy, and no correlation with RNA levels. Steady-state RNA levels do not take into account factors such as RNA stability, but, as a good correlation was found between RNAP occupancy and RNA levels, it seems a reasonable proxy for transcription output for most Mja genes. The lack of a strong correlation between promoter motifs and RNA levels illustrates the importance of additional factors such as the chromatin context as well as gene-specific regulators[47]. For example, TBP recruitment to the Mja *rb2* promoter TATA element is enhanced by the adjacent binding of the Ptr2 activator *in vitro*[48]. Based on the BRE/TATA score of the *rb2* promoter, the relative TBP promoter occupancy can be predicted by linear regression as 0.14 $\log_2$(IP/input), while the observed value is much higher at 1.01, in line with a Ptr2-enhancement of TBP binding *in vivo*. A nascent elongating transcript (NET)-seq (refs 49,50) approach would allow a direct determination of transcription output *in vivo*, and could provide insights into the manifold factors that regulate transcription within archaea in the future.

Methods

Culture conditions. Mja strain DSM 2661 (ref. 51) was grown in large-scale 100 l fermenters in minimal medium containing 0.3 mM K_2HPO_4, 0.4 mM KH_2PO_4, 3.6 mM KCl, 0.4 M NaCl, 10 mM $NaHCO_3$, 2.5 mM $CaCl_2$, 38 mM $MgCl_2$, 22 mM NH_4Cl, 31 μM $Fe(NH_4)_2(SO_4)_2$, 1 mM $C_6H_9NO_6$, 1.2 μM $MgSO_4$, 0.4 mM $CuSO_4$, 0.3 μM $MnSO_4$, 36 nM $FeSO_4$, 36 nM $CoSO_4$, 3.5 nM $ZnSO_4$, 4 nM $KAl(SO_4)_2$, 16 nM H_3BO_3, 42 μM Na_2SeO_4, 0.3 nM Na_2WO_4, 11 μM $NaMoO_4$, 44 μM $(NH_4)_2Ni(SO_4)_2$ and 2 mM Na_2S. Fermenters were mixed at 250 r.p.m. and with H_2:CO_2 gas in a 4:1 ratio at 85 °C.

RNA preparation. RNA for sequencing was prepared from Mja cell pellets by Vertis Biotechnologies using the mirVana RNA isolation kit (Ambion). For TSS mapping, total RNA was treated with TEX (Epicentre) to remove 5′ mono-phosphate RNA. RNA for northern blot analysis was prepared from Mja cell pellets using peqGOLD TriFast reagent (PeQlab) according to the manufacturer's instructions.

ChIP. All antibodies used in ChIP experiments were rabbit antisera produced by Davids Biotechnologie using recombinant proteins prepared as in ref. 52. The specificity of antibodies was determined by western blot. Mock control IPs used pre-immune sera. ChIP was performed on cultures of Mja that were grown to late log phase as measured by a cell count of ~ 1×10^8 cells ml^{-1} and crosslinked by the addition of 0.1% formaldehyde for 1 min before quenching with 12.5 mM glycine. Similar crosslinking conditions have been used successfully for the thermophile *Pyrococcus*[53,54]. Fixed cell pellets were washed three times in PBS and then resuspended in lysis buffer (0.1% sodium deoxycholate, 1 mM EDTA, 50 mM HEPES pH 7.5, 140 mM NaCl, 1% Triton-X-100) plus 10% glycerol and protease inhibitor (cOmplete mini, EDTA-free protease inhibitor cocktail, Roche). DNA was sheared by sonication to ~300 bp fragments using a cup horn sonicator (Qsonica Q700) before mixing overnight at 4 °C with the appropriate antibody prebound to Dynabeads M-280 sheep anti-rabbit IgG (Life Technologies). Beads were washed twice with lysis buffer, once with lysis buffer 500 (0.1% sodium deoxycholate, 1 mM EDTA, 50 mM HEPES pH 7.5, 500 mM NaCl, 1% Triton-X-100), once with LiCl buffer (0.5% sodium deoxycholate, 1 mM EDTA, 250 mM LiCl, 0.5% nonidet P-40, 10 mM Tris pH 8) and a final wash with TE buffer (10 mM Tris pH 7, 0.1 mM EDTA). DNA–protein complexes were eluted with ChIP elution buffer (10 mM EGTA, 1% SDS, 50 mM Tris pH 8) at 65 °C for 10 min and the remaining complexes were eluted in TE (10 mM Tris pH 7, 0.1 mM EGTA) containing 0.67% SDS. Input samples were prepared by mixing sheared DNA–protein mix with TE (10 mM Tris pH 7, 0.1 mM EGTA) containing 1% SDS. Crosslinks were reversed and protein removed by treatment of samples with 0.05 mg ml^{-1} RNase A and 0.5 mg ml^{-1} proteinase K at 37 °C for 2–4 h followed by overnight incubation at 65 °C. DNA fragments were purified using MinElute columns (Qiagen) and quantified using the Qubit dsDNA HS kit (Life Technologies).

Illumina sequencing. A summary of steps is provided in Supplementary Fig. 1. Library preparation and Illumina sequencing of total- and TEX-treated RNA was performed by Vertis Biotechnologies. For the TEX-treated samples, RNA adapters were ligated to the 5′ ends, and 3′ ends were poly(A) tailed before first-strand cDNA synthesis and PCR amplification. The resulting cDNA was fractionated by ultrasound and 5′ ends were selected and further amplified after ligation of TruSeq 3′ end adapter primer (Illumina). For RNA-seq of total RNA, samples were fragmented with ultrasound and first-strand cDNA synthesis was performed using randomized N6 primer before ligation of strand-specific TruSeq adapters (Illumina) to the 5′ and 3′ end of the cDNA and PCR amplification. cDNA samples were pooled, subjected to size selection of 150–500 bp using Agencourt AMPure XP beads (Beckman Coulter), and sequenced on an Illumina HiSeq 2000 with single-end 50 bp read length followed by adapter trimming and filtering by quality score. ChIP-seq library preparation was performed using a NEBNext ChIP-seq library preparation set for Illumina and NEBNext multiplex adaptor oligos (New England Biolabs) including size selection to ~250 bp using Agencourt AMPure kit and sequenced on an Illumina HiSeq (library 1) or MiSeq (libraries 2 and 3) with single-end 50 nt read length followed by adapter trimming and quality filter. The quality of the sequences was further assessed by FastQC (ref. 55).

TSS mapping. For TSS analysis TEX-treated RNA sequences were aligned to the Mja genome using Bowtie[56] allowing for no mismatches in the first 28 nt of the read and filtering out any read that aligned to more than one location (for mapping statistics see Supplementary Table 1). BEDTools[57] was used to create strand-specific nucleotide resolution histograms of the 5′ nucleotide of each read across the entire genome for each replicate. The R statistical program[58] with findPeaks function from the package quantmod was used to determine the genome positions containing TSSs as peaks, that is, the highest position in any continuous sequence of counts. These TSSs were further filtered as detailed in Supplementary Section 'Transcription start site mapping of *Methanocaldococcus jannaschii*', and identified TSSs are listed in Supplementary Table 2 along with the read count for each replicate at the TSS coordinate.

TU mapping. The TSS list and list of annotated and novel genes (Supplementary Tables 2–4) was used to determine the TUs for single gene cistrons, multi gene operons and non-coding RNA genes. TU coordinates were defined as the TSS to the stop codon of the last cistron for coding TU, or the annotated end for non-coding RNA. Where multiple TSSs occur for a single TU, the primary TSS, the TSS with the highest read count, was used (for details see Supplementary Section 'Organisation of the Mja transcription units').

Fidelity of TSS selection. To assay their fidelity, the TSSs were first filtered so that where multiple assigned TSSs occurred within 5 nt, the one with the highest read

count was retained. The number of reads from the TEX-treated samples where the 5′ end mapped to each position −5 to +5 relative to the assigned TSS was then determined and averaged over the two replicates. For each individual region the read count was normalized to the read count at the +1 position of the assigned TSS. Significance between the same relative positions for assigned TSSs with an Inr of T(A/G) compared to those without was determined by Wilcoxon rank-sum test.

Transcriptome analysis. For transcriptome analysis, random primed RNA sequences were aligned to the Mja genome using Bowtie[56], allowing for no mismatches in the first 28 nt of the read. Reads that align to more than one location were found to only affect 1.8% of the genome, so these were included and each mapped to one location so that regions containing repeats (such as the rRNA operons) were not misrepresented in the data set. Mapping statistics are provided in Supplementary Table 1. For expression analysis the number of strand-specific reads across the length of each TU was determined using BEDTools[57] and used to calculate the strand-specific RPKM. The RPKM values were averaged over the two replicates (Supplementary Table 3). To assess whether a TU contains detectable transcript sense, RPKM values for each replicate were first log-transformed to approximate a normal distribution, then a one-sample *t*-test was applied for $\log_{10}$(RPKM) greater than 0 (that is, RPKM greater than 1) followed by Benjamini–Hochberg false discovery rate adjustment. An adjusted *P* value of <0.05 was used to define a detectable transcript.

ChIP occupancy analysis. An outline of the sequencing analysis is shown in Supplementary Fig. 1b. ChIP sequenced reads were aligned to the genome using Bowtie[56], allowing for no mismatches within the first 28 nt. BAM files were read into the R statistical program[58] with packages ShortRead and GenomicRanges. The package chipseq was used to extend the 50 bp reads in the sense orientation to reflect the average fragment size of 250 nt. Mapping statistics are shown in Supplementary Table 1 (for additional details see Supplementary Section 'Occupancy profiling of the Mja general transcription machinery using ChIP-seq' and Supplementary Fig. 1b.).

Genome-wide occupancy, overlapping windows across entire genome. For pairwise genome-wide comparison of occupancies, the genome was split into overlapping windows of 250 bp to reflect the average DNA fragment length of the ChIP fragments. The reads per window for each IP and input sample was determined using BEDTools[57] and normalized to individual read depth by dividing by total mapped reads per sample and multiplying by 1,000,000. Each IP sample was divided by the input, resulting in the normalized (IP/input) read count. The normalized read count was averaged across replicates and log-transformed to provide $\log_2$(IP/input) for each region.

Genome-wide occupancy, TU occupancy. To determine TU occupancy, each TU with detectable transcript levels (sense RPKM > 1 with adjusted $P < 0.05$) was first separated into a promoter region corresponding to TSS ± 250 nt (average fragment length) and an intra-TU region starting at the TSS + 250 nt and to the end of the TU, excluding those TUs smaller than 250 nt. The reads per segment for each IP and input sample were determined using BEDTools[57] and normalized to an individual read depth by dividing by total mapped reads per sample and multiplying by 1,000,000. Each IP sample was divided by the input, resulting in the normalized (IP/input) read count. The normalized read count was averaged across replicates and log-transformed to provide the $\log_2$(IP/input) for each region.

Occupancy at specific loci. For comparison of specific genomic intervals, BEDTools[57] was used to create per nucleotide read counts for the extended reads of IP and input samples across the entire genome. The reads were normalized to the individual read depth at each position by dividing by total mapped reads per sample and multiplying by 1,000,000. Each IP sample was divided by the input, resulting in a normalized (IP/input) read count. The normalized read count was averaged across replicates and log-transformed to provide the $\log_2$(IP/input) for each position. For individual genomic intervals the histograms at specific genome coordinates were extracted, replicates were averaged, and plots were smoothed using sliding 40 bp windows.

Metadata analysis plots. To prepare average occupancy profiles, the read counts surrounding the regions of interest (for example, TSSs for the top 25% of mRNA genes by RPKM) were extracted from the per nucleotide occupancy histograms normalized to read depth and input. The occupancy at each position relative to the site of interest was averaged across each TU. Replicates were averaged and plots smoothed by averaging over sliding 60 bp windows.

Occupancy RNAP versus Spt4/5. To detect variations in Spt4/5 recruitment pattern on different TUs, we calculated the difference between Spt4/5 and RNAP occupancy for each 250 bp window across the genome, as described above. We extracted the coordinates for windows with a difference <−1, that is, where Spt4/5 $\log_2$(IP/input) occupancy was at least 1 lower than RNAP occupancy. Overlapping windows were merged to determine the coordinates of these regions of difference, and the read counts for each complete region of difference were calculated and normalized to read depth and input, as described above. The significance between RNAP and Spt4/5 occupancies at these regions was determined by applying Welch's *t*-test followed by Benjamini–Hochberg false discovery rate adjustment. To determine whether differences between RNAP and Spt4/5 related to 5′ UTR length of coding TU genome-wide, the difference between Spt4/5 and RNAP occupancy was calculated for each mRNA TU promoter region (see Section 'Genome-wide occupancy, TU occupancy' for the calculation of promoter occupancy) and correlated to the length of the 5′ UTR.

Sequence motif analysis. To identify promoter elements, DNA sequences ranging from −50 to +10 nt relative to the identified TSSs were extracted using BEDTools[57] and direct alignments were visualized using WebLogo 3 (ref. 59). Putative promoter motifs were determined using MEME-ChIP (Motif Analysis of Large Nucleotide Datasets)[60], restricting the search to motifs 6–15 nt wide on the sense strand. The position weight matrix of the resulting 15 nt BRE/TATA motif was used with FIMO (Find Individual Motif Occurrences)[60] to identify matches in the sequences upstream of the TSSs and provide confidence scores as *P* values. Due to high AT content of the Mja genome, FIMO was also used to identify matches to the BRE/TATA motif in a control set of seven randomly generated sets of 1,508 sequences of the same length from the Mja genome using BEDTools[57] (Supplementary Fig. 3a). For identification of the Mja RBS motif, the DNA sequences corresponding to −20 to +20 around the start codons were analysed using MEME-ChIP and restricting the search to motifs of 4–5 nt on the sense strand. For analysis of the dinucleotide frequencies, the proportion of TA or TG at each position relative to the TSS was calculated. This was compared to the genome average occurrence of TA/TG dinucleotides using Fisher's exact test of significance. For analysis of the IMR, the percentage of AT at positions −12 to +2 relative to the TSS was calculated using BEDTools[57], and significance calculated by Wilcoxon signed rank test.

EMSA and *in vitro* transcription assays. Recombinant mjRNAP was prepared as in ref. 52 and EMSA assays were performed as in ref. 61. Oligonucleotides are listed in Supplementary Table 6. *In vitro* transcription reactions with plasmids bearing Mja promoters fused to C-less cassettes were carried out analogous to ref. 9, with the promoter region including 15 bp upstream of the identified BRE/TATA motifs and 8–13 bp downstream of the TSS. For construction of the C-less fusions, the following oligos (Supplementary Table 6) were used: *rrnA* fw, CRISPR TSS1 fw, CRISPR TSS2 fw and *rpl12* fw, all with the C-less rev. Buffer conditions and Mja transcription factor concentrations for Mja *in vitro* transcription assays were as described in ref. 61 with 300 ng of SacI-linearized plasmid, heparin concentration reduced to 5 μg ml^{-1} and a single incubation step at 65 °C for 15 min. A recovery marker was included to monitor possible losses during the nucleic acid purification before gel loading.

Northern blotting. Northern blotting was carried out as in ref. 62 using a low-range RiboRuler RNA ladder (Fermentas) and probes constructed from oligonucleotide templates A3 sense and A3 antisense (Supplementary Table 6).

Statistical analysis. All graphs were produced using GraphPad Prism version 5 and the R Statistical program[58] and package ggplot2 (ref. 63). Correlations and statistical tests were performed using R base install; specific tests are detailed as appropriate throughout the manuscript.

Data availability. The sequencing data sets generated during this study have been deposited in the NCBI Sequence Read Archive (SRA) under accession codes SRP089683 (ChIP) and SRP089689 (RNA). The Supplementary Information includes TSSs and promoter mapping data (Supplementary Table 2) and Mja operon organization, gene expression and occupancy data (Supplementary Table 3) in Excel spreadsheet format. The data that support the findings of this study are available from Finn Werner upon request.

Received 8 September 2016; accepted 24 January 2017; published 1 March 2017

References

1. Werner, F. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol.* **16,** 247–250 (2008).
2. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).
3. Korkhin, Y. *et al.* Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* **7,** e1000102 (2009).
4. Hirtreiter, A., Grohmann, D. & Werner, F. Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity *in vitro*. *Nucleic Acids Res.* **38,** 585–596 (2010).
5. Li, E., Reich, C. I. & Olsen, G. J. A whole-genome approach to identifying protein binding sites: promoters in *Methanocaldococcus* (*Methanococcus*) *jannaschii*. *Nucleic Acids Res.* **36,** 6948–6958 (2008).
6. Zhang, J., Li, E. & Olsen, G. J. Protein-coding gene promoters in *Methanocaldococcus* (*Methanococcus*) *jannaschii*. *Nucleic Acids Res.* **37,** 3588–3601 (2009).
7. Werner, F. & Weinzierl, R. O. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell* **10,** 635–646 (2002).
8. Gietl, A. *et al.* Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res.* **42,** 6219–6231 (2014).

9. Blombach, F. *et al.* Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife* **4,** e08378 (2015).
10. Grohmann, D. *et al.* The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43,** 263–274 (2011).
11. Werner, F. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.* **417,** 13–27 (2012).
12. Sevostyanova, A., Svetlov, V., Vassylyev, D. G. & Artsimovitch, I. The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *Proc. Natl Acad. Sci. USA* **105,** 865–870 (2008).
13. Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17,** 1272–1278 (2010).
14. Mooney, R. A. *et al.* Regulator trafficking on bacterial transcription units *in vivo*. *Mol. Cell* **33,** 97–108 (2009).
15. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 40–51 (2012).
16. Nagy, J. *et al.* Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Commun.* **6,** 6161 (2015).
17. Schulz, S. *et al.* TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl Acad. Sci. USA* **113,** E1816–E1825 (2016).
18. Bell, S. D., Jaxel, C., Nadal, M., Kosa, P. F. & Jackson, S. P. Temperature, template topology, and factor requirements of archaeal transcription. *Proc. Natl Acad. Sci. USA* **95,** 15218–15222 (1998).
19. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15,** 684 (2014).
20. Jäger, D. *et al.* Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc. Natl Acad. Sci. USA* **106,** 21878–21882 (2009).
21. Li, J. *et al.* Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon *Methanolobus psychrophilus*. *Sci. Rep.* **5,** 9209 (2015).
22. Wurtzel, O. *et al.* A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20,** 133–141 (2010).
23. Babski, J. *et al.* Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* **17,** 629 (2016).
24. Seitzer, P., Wilbanks, E. G., Larsen, D. J. & Facciotti, M. T. A Monte Carlo-based framework enhances the discovery and interpretation of regulatory sequence motifs. *BMC Bioinformatics* **13,** 317 (2012).
25. Blombach, F., Smollett, K. L., Grohmann, D. & Werner, F. Molecular mechanisms of transcription initiation—structure, function, and evolution of TFE/TFIIE-like factors and open complex formation. *J. Mol. Biol.* **428,** 2592–2606 (2016).
26. Werner, F. & Weinzierl, R. O. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell. Biol.* **25,** 8344–8355 (2005).
27. Qureshi, S. A., Bell, S. D. & Jackson, S. P. Factor requirements for transcription in the archaeon *Sulfolobus shibatae*. *EMBO J.* **16,** 2927–2936 (1997).
28. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science* **328,** 501–504 (2010).
29. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328,** 504–508 (2010).
30. French, S. L., Santangelo, T. J., Beyer, A. L. & Reeve, J. N. Transcription and translation are coupled in Archaea. *Mol. Biol. Evol.* **24,** 893–895 (2007).
31. Brenneis, M., Hering, O., Lange, C. & Soppa, J. Experimental characterization of Cis-acting elements important for translation and transcription in halophilic Archaea. *PLoS Genet.* **3,** e229 (2007).
32. Torarinsson, E., Klenk, H. P. & Garrett, R. A. Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* **7,** 47–54 (2005).
33. Koide, T. *et al.* Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol. Syst. Biol.* **5,** 285 (2009).
34. Toffano-Nioche, C. *et al.* RNA at 92 °C: the non-coding transcriptome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *RNA Biol.* **10,** 1211–1220 (2013).
35. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389,** 52–65 (2007).
36. Qureshi, S. A. Role of the *Sulfolobus shibatae* viral T6 initiator in conferring promoter strength and in influencing transcription start site selection. *Can. J. Microbiol.* **52,** 1136–1140 (2006).
37. Bell, S. D. & Jackson, S. P. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **275,** 12934–12940 (2000).
38. Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res.* **35,** 771–788 (2007).
39. Basu, R. S. *et al.* Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289,** 24549–24559 (2014).
40. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154,** 713–715 (2013).
41. Hirtreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38,** 4040–4051 (2010).
42. Diamant, G., Bahat, A. & Dikstein, R. The elongation factor Spt5 facilitates transcription initiation for rapid induction of inflammatory-response genes. *Nat. Commun.* **7,** 11547 (2016).
43. Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat. Struct. Mol. Biol.* **19,** 1108–1115 (2012).
44. Arnvig, K. B. *et al.* Evolutionary comparison of ribosomal operon antitermination function. *J. Bacteriol.* **190,** 7251–7257 (2008).
45. Micorescu, M. *et al.* archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol.* **190,** 157–167 (2008).
46. Jensen, K. F. & Pedersen, S. Metabolic growth rate control in *E*scherichia coli may be a consequence of subsaturation of the macromolecular biosynthetic apparatus with substrates and catalytic components. *Microbiol. Rev.* **54,** 89–100 (1990).
47. Peeters, E., Driessen, R. P., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* **13,** 333–341 (2015).
48. Ouhammouch, M., Dewhurst, R. E., Hausner, W., Thomm, M. & Geiduschek, E. P. Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc. Natl Acad. Sci. USA* **100,** 5097–5102 (2003).
49. Churchman, L. S. & Weissman, J. S. Native elongating transcript sequencing (NET-seq). *Curr. Protoc. Mol. Biol.* **4,** 1–17 (2012).
50. Nojima, T. *et al.* Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161,** 526–540 (2015).
51. Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136,** 254–261 (1983).
52. Smollett, K., Blombach, F. & Werner, F. Transcription in Archaea: preparation of *Methanocaldococcus jannaschii* transcription machinery. *Methods Mol. Biol.* **1276,** 291–303 (2015).
53. Reichelt, R., Gindner, A., Thomm, M. & Hausner, W. Genome-wide binding analysis of the transcriptional regulator TrmBL1 in *Pyrococcus furiosus*. *BMC Genomics* **17,** 40 (2016).
54. Liu, W., Vierke, G., Wenke, A. K., Thomm, M. & Ladenstein, R. Crystal structure of the archaeal heat shock regulator from *Pyrococcus furiosus*: a molecular chimera representing eukaryal and bacterial features. *J. Mol. Biol.* **369,** 474–488 (2007).
55. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data (2010); http://www.bioinformatics.babraham.ac.uk/projects/fastqc
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
57. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
58. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014); http://www.R-project.org
59. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. Weblogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
60. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).
61. Smollett, K., Blombach, F. & Werner, F. Transcription in Archaea: *in vitro* transcription assays for mjRNAP. *Methods Mol. Biol.* **1276,** 305–314 (2015).
62. Arnvig, K. B. & Young, D. B. Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **73,** 397–408 (2009).
63. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).

Acknowledgements

The authors thank J. Bähler and D. Bitton for advice throughout this project. The authors also thank T. Arnvig, D. Grohman and other members of the RNAP laboratory for encouragement and critical reading of the manuscript. Research in the RNAP laboratory at University College London is funded by Wellcome Trust Investigator Award WT096553MA (to F.W.).

Author contributions

K.S. designed and performed experiments, analysed data and wrote the manuscript. F.B. performed experiments and wrote the manuscript. R.R. and M.T. helped with fermenter growth and crosslinking, and provided biomass. F.W. conceptualized the study, designed experiments and wrote the manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.W.

How to cite this article: Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* **2,** 17021 (2017).

Competing interests

The authors declare no competing financial interests.