HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2016 August 15.

Published in final edited form as:

Nat Struct Mol Biol. 2016 March ; 23(3): 231–238. doi:10.1038/nsmb.3176.

7SK-BAF axis controls pervasive transcription at enhancers

Ryan A. Flynn[1,2], **Brian T. Do**[1,2], **Adam J. Rubin**[2], **Eliezer Calo**[3], **Byron Lee**[1,2], **Hannes Kuchelmeister**[4], **Michael Rale**[5], **Ci Chu**[1,2], **Eric T. Kool**[4], **Joanna Wysocka**[3], **Paul A. Khavari**[2], and **Howard Y. Chang**[1,2]

Howard Y. Chang: howchang@stanford.edu

[1]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA

[2]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA, USA

[3]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

[4]Department of Chemistry, Stanford University, Stanford, CA, USA

[5]The Genome Institute, Washington University in St. Louis, Saint Louis, MO, USA

Abstract

RNA functions at enhancers remain mysterious. Here we show that the 7SK small nuclear RNA (snRNA) inhibits enhancer transcription by modulating nucleosome position. 7SK occupies enhancers and super enhancers genome-wide in mouse and human cells, and 7SK is required to limit eRNA initiation and synthesis in a manner distinct from promoter pausing. Clustered elements at super enhancers uniquely require 7SK to prevent convergent transcription and DNA damage signaling. 7SK physically interacts with the BAF chromatin remodeling complex, recruit BAF to enhancers, and inhibits enhancer transcription by modulating chromatin structure. In turn, 7SK occupancy at enhancers coincides with Brd4 and is exquisitely sensitive to the bromodomain inhibitor JQ1. Thus, 7SK employs distinct mechanisms to counteract diverse consequences of pervasive transcription that distinguish super enhancers, enhancers, and promoters.

Introduction

Eukaryotic genomes are extensively transcribed[1,2], but unfettered transcription alters gene expression and leads to genome damage by several means[3]. RNA polymerase II (Pol II)

Correspondence to: Howard Y. Chang, howchang@stanford.edu.

Accession Codes: All sequencing data is deposited in the Gene Expression Omnibus SuperSeries GSE69143 and a summary of all sequencing experiments is provided in (Supplementary Table 5).

Author Contributions: R.A.F. and H.Y.C. conceived and designed the study. R.A.F carried out the majority of the experiments and analysis. R.A.F. and M.R. performed ChIRP-seq and R.A.F. and B.T.D. analyzed ChIRP-seq data. R.A.F and C.C. analyzed ChIRP-MS data. R.A.F., E.C., and B.L. performed ChIP-seq and analyzed the data. R.A.F., H.K., and E.T.K. synthesized icSHAPE reagents and performed icSHAPE experiments. R.A.F. performed GRO-seq and R.A.F., B.T.D., E.C., and J.W. analyzed GRO-seq data. R.A.F. performed ATAC-seq and R.A.F., B.T.D., B.L., P.A.K., and A.J.R. analyzed the data. R.A.F. and H.Y.C. wrote the manuscript with input from all co-authors.

transcribes functional regulatory elements, such as enhancers and super enhancers. Super enhancers (SE, also known as stretch enhancers) are distinguished from enhancers by the disproportionate concentration of transcriptional and chromatin modification machinery bound and appear to be particularly important for regulating gene expression networks and controlling cell state[4-6]. Nearly all active enhancers generate bidirectional transcripts termed eRNAs[2], and they are also marked by flanking, phased nucleosomes with specific histone modifications[7-9]. Genome-wide characterization of promoter and enhancer elements revealed commonalities between promoters and enhancers, such as divergent transcription start site (TSS) pairs, well phased (+) 1 and (-) 1 nucleosomes flanking the TSSs, and a central transcription factor (TF) binding site[10]. Notably, while enhancer transcription is one of the earliest steps of gene activation[11] and some enhancer RNAs (eRNAs) participate in gene regulation[12,13], far less is known about the control of eRNA transcription.

Most of our understanding of how polymerases are regulated *in vivo* has focused on protein factors (reviewed by[14,15]). However, several long noncoding RNAs (lncRNAs) have enhancer-like activities that can recruit transcription activating complexes to positively regulate genes *in cis* or *in trans*[16]. Short Alu containing RNAs can directly bind the preinitiation complex (PIC) *in trans* and repress Pol II transcription [17]. Another, *trans* acting RNA, the 7SK snRNA, has been shown to uniquely control Pol II pausing at promoters, but its role genome-wide has yet to be fully elucidated.

7SK is a highly abundant snRNA (~200,000 copies per cell[18]) that serves as a scaffold for Larp7, Hexim1, Mepce, and Cdk9-CyclinT1 (P-TEFb), forming the canonical 7SK snRNP. A key step in RNA Pol II transcription is promoter-proximal pausing, which occurs bidirectionally ~25-60 nucleotides downstream of TSSs. Promoter-proximal pause release is gated by the positive transcription elongation factor b (P-TEFb)-7SK snRNA pathway; release from 7SK allows for P-TEFb phosphorylation of Pol II (as well as other kinases[19,20]) and subsequent elongation[21]. The 7SK small nuclear ribonucleoprotein (snRNP) is thought to reside in the nucleoplasm, but it has been suggested that 7SK could operate physically on chromatin[22-26]. While these studies have focused on the specific functions of the protein-components of the 7SK snRNP, a comprehensive analysis of the RNA component of the 7SK snRNP at chromatin has not been determined.

RNA is ideally suited to control the flow of information in the nucleus for several reasons. RNA transcription or occupancy can mark unique allele or spatial positions in the nucleus[27]. Many lncRNAs can act as RNA scaffolds that bring together multiple distinct protein complexes into physical proximity to regulate gene expression or other functions[28]. Further, RNA can base pair with itself to form complex secondary and tertiary structures, and one RNA sequence can adopt multiple physical conformations with similar folding energies[29]. Here we show that 7SK is a multifaceted RNA scaffold for co-transcriptional control. By combining recently described and novel RNA-centric technologies, we characterize 7SK's *in vivo* binding partners, chromatin occupancy sites, and RNP-specific conformations, and reveal a novel role for 7SK as an important regulator of ATP-dependent chromatin remodeling at enhancers.

Results

Conserved 7SK chromatin occupancy at super enhancers

We mapped the genomic occupancy of 7SK using chromatin isolation by RNA purification followed by deep sequencing (ChIRP-seq) in mouse embryonic stem (ES) cells. Using two orthogonal probe sets (“Even” and “Odd”) targeting 7SK, we recovered most of the cellular 7SK (Supplementary Fig. 1a and 1b). Genomic DNA recovery was specific over other abundant nuclear and cytoplasmic RNAs, and was sensitive to RNaseA treatment (Supplementary Fig. 1b). Surprisingly, we observed extensive 7SK occupancy across the entire transcribed loci of mRNA genes, which mirrors the binding profile of Pol II (Fig. 1a, Supplementary Fig. 1c). Outside genic regions, 7SK also associated with super enhancer (SE) and typical enhancer (TE) elements (Fig. 1a-c). At these elements, 7SK overlaps regions bound by Pol II pausing factors (Nelf-a, Supt5h), a reader of histone acetylation (Brd4), a Pol II initiation mark (TBP) and open chromatin sites (as measured by the assay of transposase-accessible chromatin using sequencing, ATAC-seq) (Fig. 1a). The typical pattern of occupancy is a focal peak of open chromatin bound by Brd4 and other transcription factors, with 7SK extensively occupying the transcribed regions. “Super enhancers” are recently reported designations for clustered enhancer elements[4]; whether SE possess distinct biochemical or regulatory properties has been debated[30]. 7SK ChIRP-seq signal is strongly enriched at SE, similar to findings for other core transcriptional machinery (Fig. 1b and 1c). These data identify 7SK as a chromatin bound ncRNA, occupying regulatory regions actively transcribed by Pol II.

ChIRP followed by quantitative PCR (ChIRP-qPCR) and other controls confirmed the specificity of these findings. We selectively recovered 7SK-bound regions in a RNA-dependent manner (Supplementary Fig. 1d). 7SK ChIRP-seq signal is not simply a result of highly active transcription because we observed no enriched 7SK signal at the ribosomal DNA locus (transcribed by Pol I and III but not Pol II), consistent with 7SK being excluded from the nucleolus (Supplementary Fig. 1e). 7SK is highly conserved at the primary sequence level between mouse and human. Consistently, 7SK ChIRP-seq in human ES (H1) and HeLa cells each identified over 5000 sites, with similar distributions of peaks over genic and enhancer regions (Supplementary Fig. 1f-i). These results suggest 7SK's role on chromatin is conserved across mammalian species and cell-types.

We developed “factor ratio analysis” to compare co-occupancy of 7SK with transcriptional and chromatin regulators at promoters, SE, and TE (Supplementary Fig. 2a). Because many factors involved in transcriptional initiation occupy both promoters and enhancers[10], regulatory features that may distinguish these classes of elements are actively sought. Globally, 7SK associates with actively transcribed chromatin and components of the Pol II transcriptional apparatus (Supplementary Fig. 2b). In the factor ratio analysis, the chromatin occupancy signal of each factor is normalized to TBP, thus the proportion of different factors relative to Pol II initiation per base can be documented (**Methods**). We used Start-Seq[31] data to precisely align experimentally measured transcription start sites (TSS) at promoters, SE, and TE for all analyses (Supplementary Table 1, **Methods**) Further, in order to directly compare SE and TE, we used ATAC-seq to identify individual peaks of open chromatin in

SE domains (Supplementary Table 1), which can facilitate the analysis of functional regions within SE[32].

Factor ratio analysis revealed that promoters, SE, and TE each have distinct and characteristic ratios of factor occupancy, and components of the 7SK snRNP are key distinguishing features (Fig. 1d and Supplementary Fig. 2). Promoters are enriched for the pausing factor Nelf-a and Pol II Ser5p (initiated but paused Pol II). Additionally, canonical 7SK snRNP proteins like Hexim1 and Ddx21[33] are also biased for promoters. In contrast, SE and TE both have higher levels of mediator Med1 and histone acetylation reader Brd4, with SE greater than TE. Finally, SE have disproportionately more 7SK than promoters or TE, and the 7SK at SE is not associated with proportional levels of pausing (Nelf-a or Supt5h) nor canonical 7SK snRNP (Hexim1 or Ddx21) factors as promoters (Fig. 1d). Factor ratio analysis using two independent TBP data sets or the basal transcription factor TAF1 as the denominator yielded the same findings, confirming the robustness of this approach (Supplementary Fig. 2c). The different ratios between 7SK and these chromatin factors at promoters and enhancers suggest that 7SK may mediate additional novel functions at enhancer elements.

An orthogonal analysis interrogating binding patterns revealed that 7SK and Nelf-a associate with divergently transcribed polymerase complexes at promoters, tracking Pol II and TBP[10] (Supplementary Fig. 2d). A similar trend is seen at enhancer regions. However, at SE the 7SK binding is broad, more similar to that of Pol II Ser5p (Supplementary Fig. 2e and Fig 1c). To better understand how the colocalization of 7SK at promoters and enhancers relates to function, we extended the factor ratio analysis to 41 genome-wide datasets and performed principal component analysis (PCA). We found that the first two principal components of the factor ratio analysis can differentiate promoters, SE, and TE (Supplementary Fig. 2f). Interestingly, we find that SE exhibit an intermediate pattern between promoters and TE, but that TE and SE in general cluster together (Supplementary Fig. 2f). These data reveal individual SE peaks as a special class of elements, which have features of both promoters and enhancers, and that SE attract the most 7SK per Pol II initiation of all three elements. Additionally, while 7SK marks all three regions, pausing factor ratios suggest that Pol II pausing operates differently at enhancer elements.

7SK, ConvT, and DNA-damage signaling at super enhancers

7SK association with different factors at specific elements genome-wide raises the possibility that 7SK has multiple functions. We therefore directly assayed the effects of 7SK on Pol II transcription by applying Global Run On sequencing (GRO-seq) in mouse ES cells. We utilized antisense oligonucleotides (ASOs)[34] that robustly depleted 7SK (Supplementary Fig. 3a). Loss of 7SK globally altered transcription but with several distinct patterns. At promoters, loss of 7SK resulted in a reduction of promoter-proximal and a gain of gene body transcription, consistent with Pol II pause release (Fig. 2a, 2b, and Supplementary Fig. 3b). In contrast, SE and TE experienced consistent increases in both TSS-proximal and distal eRNA transcription (Fig. 2a and 2b). Quantification of individual elements showed over 1600 enhancers with significantly increased transcription; 400 enhancers increased eRNA synthesis by 200% or more (131 enhancers at ≥ 300% and 44

enhancers at ≥ 400% elevated, respectively). In contrast, fewer enhancers (243) show significantly decreased eRNA production (Supplementary Fig. 3c). 7SK depletion also caused increased histone H3 lysine 36 trimethylation (H3K36me3) at enhancers, a mark associated with active Pol II transcription (Supplementary Fig. 3d). Importantly, enhancers that gained GRO-seq and H3K36me3 signal after 7SK knockdown are enriched for similar ES specific GO terms above the background of previously defined ES enhancer elements (Supplementary Fig. 3e and 3f, **Methods**). Taken together, these data suggest that 7SK acts to repress transcription at enhancers and promoters with distinct mechanisms.

One important consequence of unfettered transcription is convergent transcription: when one region of the genome is transcribed on both the Watson and Crick strands (Supplementary Fig. 4a). Convergent transcription (ConvT) can lead to Pol II collisions and activation-induced cytidine deamination (AID)-mediated DNA damage[3]; thus surveillance mechanisms to reduce this phenomenon are important for genome stability. Using an established metric of ConvT based on GRO-seq data (**Methods**), we find transcribed regions experience more ConvT as compared to insulator (Ctcf) or random regions, and SE in particular are the most enriched (Supplementary Fig. 4b) mirroring recent reports in B-cell lymphoma[3]. Loss of 7SK causes significant increases in ConvT at promoters, SE, and TE (Fig. 2c and Supplementary Fig. 4c). Despite P-TEFb's previously reported activity at both downstream and upstream Pol II complexes[35], we observed a gain in ConvT signal in the downstream (mRNA) direction and loss in the upstream direction (Fig. 2c), These results strengthen the notion that promoters experience specific directional regulation, and suggest that promoter pause-release results in a more permissive environment for ConvT, possibly allowing downstream convergent Pol II complexes[36] to initiate and transcribe more readily.

At enhancers, elevated ConvT was seen emanating in both directions. Notably, in control cells there was ~8 fold more ConvT at SE than at TE, and 7SK depletion resulted in broad increases in ConvT (Fig. 2c). While SE have more transcription than TE, they are particular enriched for 7SK-dependent ConvT changes even when compared to promoters which have higher levels of transcription as judged by GRO-seq (Supplementary Fig. 4d). This effect is likely a direct consequence of the clustered arrangement of open chromatin sites (each a TSS) within SE; the peak-to-peak distance from one open chromatin site to the next is at least 10-fold shorter in SE than TE or promoters (Supplementary Fig. 4e). Finally, intragenic enhancers, which are located within the mRNA transcriptional unit, also experience more ConvT than intergenic TE (median ConvT values of 0.067 and 0.025, respectively, Supplementary Fig. 4f). Therefore, 7SK plays an active role in controlling the balance of transcription genome-wide.

The observation that 7SK modulates ConvT, especially at SE, raised the possibility that downstream signaling events, such as DNA damage signaling might occur after 7SK loss. We monitored levels of S139-phosphorylated gamma-Histone 2A.X ("γ-H2AX" hereafter) by immunofluorescence and found elevated levels of γ-H2AX at the single cell level after 7SK depletion (Fig. 3a). To gain a higher resolution picture, we performed ChIP-seq of γ-H2AX in mouse ES cells with and without 7SK depletion. 7SK depletion caused significant increases of γ-H2AX at promoters, SE, and TE (Supplementary Fig. 4g). Notably, there is a positive correlation between changes in ConvT and γ-H2AX changes at all three elements

(Fig. 3b), and this correlation is strongest for SE (Spearman correlation, ρ = 0.677). Together, these data suggest that 7SK is important for controlling the directional bias of transcription to prevent DNA damage at clustered SE elements.

7SK forms a novel snRNP with the BAF complex

To discover how 7SK might functionally discriminate promoters from enhancers, we employed ChIRP-mass spectrometry (ChIRP-MS) to identify novel *in vivo* protein partners of the 7SK snRNA[37]. Canonical 7SK snRNP factors were highly enriched, including Larp7, Mepce, Ccnt1, Cdk9, Hexim1, Ddx21, and Brd4 (Fig. 4a). Notably, 7SK retrieved five BAF complex subunits, including the central ATPase subunit, Brg1, as well as Baf155, Baf60a, Baf53a, and Baf47 (Fig. 4a). The number peptides of BAF subunits were comparable to that of a known 7SK snRNP partner (Brd4), and 4 of 5 BAF subunits were only retrieved by 7SK by not by three other nuclear ncRNAs U1, U2, or Xist (Fig. 4a). BAF is a mammalian ATP-dependent nucleosome-remodeling complex, and has been linked to major human diseases, including in cancer and autism[38,39]. In mouse ES cells, BAF has recently been shown to regulate nucleosome occupancy and inhibit eRNA synthesis at enhancers[40], making BAF an attractive partner for 7SK action. Further, recent reports suggest that specific noncoding RNAs can impart locus-specific control of the BAF complex[41,42].

ChIP-seq analysis of Baf155 and other 7SK partners revealed that 7SK resides in two distinct complexes. Enhancer-bound 7SK is co-associated with Baf155 and Brd4, while promoter-bound 7SK is associated with Hexim1 and Ddx21, as exemplified by the *Klf4* locus (Fig. 4b). Global analysis of 7SK ChIRP-seq peaks revealed that ChIP-seq of Baf155 and Brd4 clustered together and away from Hexim1 ChIP-seq (Fig. 4c). To further characterize these factors at the peak-level, we analyzed the intersection of 7SK, Baf155, and Hexim1 peaks across the genome (Supplementary Table 2). We find that most enhancers have BAF occupancy (84%) but no HEXIM. Conversely, most promoters (73%) have HEXIM occupancy but no BAF. We next characterized these complexes biochemically: we separately immunoprecipitated BAF and Hexim1. Analyzing the BAF-enriched proteome by MS, we robustly recovered subunits of the BAF complex (Fig. 4d, range 14-253 peptides per protein, Supplementary Table 3). However little Hexim1 or P-TEFb was observed (Fig. 4d). Western blot analysis of Hexim1-enriched proteins confirmed the lack of association between BAF and Hexim1 (Supplementary Fig. 5a). Despite the poor co-recovery between BAF and HEXIM, we found that both complexes could separately retrieve 7SK. Native RNA immunoprecipitation (RIP) of Hexim1 or BAF subunit Arid1a recovered ~65% and ~6% of the cellular 7SK, respectively (Supplementary Fig. 5b). Given 7SK's high abundance, estimated at ~200,000 copies per cell[18], the 7SK-BAF complex could exist at more than 10,000 copies per cell and thus potentially regulate thousands of loci across the genome.

We discovered that 7SK exists in two mutually exclusive RNA structures in association with BAF vs. Hexim1. We measured the secondary structure of endogenous 7SK in association with either Hexim1 or BAF RNP complexes using icSHAPE[43]. Immunoprecipitated Hexim1 or BAF complexes were treated with the icSHAPE probe to obtain a snapshot of 7SK structure in each RNP context. We find many single nucleotide reactivity differences in Stem loops 1 and 2, as well as large contiguous differences between BAF- and Hexim1-associated

structures in Stem loop 3 (Fig. 5a and Supplementary Table 4). By subtracting the icSHAPE values in BAF from Hexim1, we observe nucleotides 219-250 of 7SK are substantially more reactive (single-stranded) in association with BAF, while nucleotides 284-300 are more reactive in association with Hexim1 (Fig. 5b). Mapping these reactivities onto the predicted secondary structure of 7SK's Stem loop 3 and 4 (Fig. 5c) further confirms the distinct regional differences in BAF- and Hexim1-associated 7SK structural profiles. Altogether, ChIP-seq, biochemical purification, and RNA structural data suggest that 7SK exists in two separable complexes—a 7SK-Hexim1 complex at promoters and a 7SK-BAF complex at enhancers.

7SK facilitates BAF action at enhancers

We next characterized the step-wise interplay of the 7SK-BAF complex and enhancer chromatin. As transcription occurs on a nucleosomal template and 7SK associates with the BAF complex at enhancers (Fig. 6), we hypothesized that 7SK may change chromatin organization. Using ATAC-seq we inferred nucleosome positions in mouse ES cells (**Methods**) and obtained high quality nucleosome positioning at promoter and enhancer elements (Supplementary Fig. 2c and 2d). The organization and density of nucleosomes was nearly unchanged at promoters after the loss of 7SK (Fig. 6a). In contrast, at TE and SE elements 7SK depletion causes stronger central nucleosome positioning (Fig. 6b). Histone H3 ChIP-seq revealed little change in histone occupancy at enhancers and promoters after the loss of 7SK (Supplementary Fig. 5c and 5d). Thus, 7SK affects relative nucleosome positioning but not occupancy.

As an RNA component of BAF, we hypothesized that 7SK could bridge interactions important for the BAF complex's activity. Glycerol gradient analysis showed that 7SK depletion led to decomposition the canonical 7SK snRNP (Cdk9, Ccnt1, and Hexim1) to lower molecular weight (MW) fractions (Supplementary Fig. 6a), serving as positive control[44]. Importantly, 7SK depletion also shifted both Pol II and BAF complex subunits (Brg1, Arid1a, and Baf155) towards lower MW fractions (Supplementary Fig. 6b). These data suggest that 7SK could function to connect BAF to the transcriptional apparatus. In IP experiments with BAF subunit Arid1a, 7SK depletion reproducibly decreased Pol II co-IP with the BAF complex, but did not affect the mutual co-IPs of BAF subunits (e.g. Baf47, Fig. 6c). These data support 7SK's role in scaffolding BAF and Pol II, thereby facilitating a chromatin state compatible with transcribed loci.

Our biochemical data raise the possibility that 7SK might facilitate the *in vivo* recruitment to, or stabilization of BAF on chromatin. To determine *in vivo* roles of 7SK and BAF, we depleted cells of 7SK and performed Baf155 ChIP-seq. Loss of 7SK resulted in global reduction of Baf155 occupancy at enhancers (Fig. 6d, Supplementary Fig. 6c and 6d). ChIP-qPCR validated Baf155 occupancy at multiple mouse ES super enhancers in a 7SK-dependent manner (Fig. 6e). Enhancers with 7SK-dependent BAF occupancy are enriched for selective activity during early embryogenesis, indicating relevance to ES cell biology (Supplementary Fig. 6e). The identification of BAF, a known player in genome stability[45], as a 7SK-enhancer partner also adds to the mechanism of 7SK protection from DNA damage signaling. Multivariate regression analysis showed that 7SK-dependent control of enhancer

transcription, convergent transcription, and BAF recruitment are each independent and significant contributors to diminish DNA damage (Supplementary Fig. 6f). The ConvT signal provided additional predictive power on top of the GRO-seq, despite being derived from the GRO-seq data (Supplementary Fig. 6f), suggesting that overlapping read density is indeed important for γ-H2AX deposition.

As nucleosome positioning can modulate transcription factor binding to the genome[46], ChIP-seq analysis of Oct4 (also known as Pou5f1) in mouse ES cells revealed that 7SK depletion also led to loss of Oct4 occupancy (Supplementary Fig. 7a-e). Analysis of transcription factor footprints using ATAC-seq and the bioinformatics tool PIQ[47] (**Methods**), also showed 7SK dependence for Oct4 and Sox2 binding, but not for Ctcf (Supplementary Fig. 7f-h). To directly compare 7SK loss to BAF loss, we analyzed RNA-seq data collected after Brg1 depletion[40]. Loss of Brg1 or 7SK resulted in strongly concordant changes in eRNA synthesis at the same SE elements (281 of 349 SE, $p = 1.85\times10^{-17}$, Fisher's exact test), and TE elements (2210 of 3739 TE, $p = 2.75\times10^{-15}$, Fisher's exact test). Thus, 7SK depletion phenocopies BAF loss. These results suggest that 7SK is important to recruit BAF, a known inhibitor of eRNA transcription[40], to enhancers across the genome.

In turn, 7SK-BAF is targeted to enhancers via Brd4 bromodomain interaction. JQ1 is a small molecule bromodomain inhibitor and promising anti-cancer drug candidate[48]. We found that brief treatment (1 hour) with JQ1 dissociated 7SK but not Pol II from enhancer chromatin genome-wide, as measured by ChIRP-seq and ChIP-seq (Supplementary Fig. 8). No compensatory expression of 7SK snRNP factors was observed in these JQ1 treatment conditions (Supplementary Fig. 8b). At the *Nanog* promoter and associated SE, Pol II's chromatin occupancy was largely stable, except for a loss in the 3′-end of the *Nanog* locus, suggesting that JQ1 was indeed leading to promoter-proximal pausing of Pol II (Fig. 7a). Increased Pol II pausing at genes was seen globally (Supplementary Fig. 8c and 8d). Quantitative analysis of 7SK occupancy, normalized for changes of Pol II association post JQ1 treatment, revealed that 7SK occupancy was most sensitive to JQ1 at SE over TE and promoters (Fig. 7b, Supplementary Fig. 8c-f). This is consistent with Brd4's relative enrichment at SE and TE over promoters (Fig. 1d).

Discussion

Our results reveal the distinct control mechanisms and consequences of pervasive transcription at enhancers vs. promoters, notably distinguishing super enhancers. Enhancers are the earliest transcriptional responders to cellular differentiation and stress, and therefore are critical for orchestrated gene expression[11]. As a chromatin bound noncoding RNA, 7SK represses transcription at enhancers via BAF chromatin remodeling but at promoters via Pol II pause release (Fig. 7c). Enhancer-specific linkage of 7SK to BAF allows cells to focally inhibit eRNA production, and resolve undesired consequences of enhancer transcription, such as convergent transcription at super enhancers. Thus, manipulation of 7SK may impact diseases including cancer and autism where BAF and enhancer dysfunction play important roles. Enhancers and promoters are often in spatial proximity during gene expression, yet distinct mechanisms and small molecule sensitivities (e.g. JQ1) control each element's transcription and other post-transcriptional features[49]. 7SK potentially serves as a single

integrator to globally regulate transcription. Future work will need to decipher how the cell discriminates these elements and what role 7SK or other noncoding RNAs play in this choice.

Online Methods

Cell culture and Antisense Oligonucleotide Knockdown of 7SK

V6.5 mouse ES cells (Kind gift, P.A. Sharp) were cultured on gelatinized plates with serum and LIF as previously described[35]. H1 human ES cells (kind gift V. Sebastiano) were cultured in mTeSR1 on matrigel coated dishes. All cells were negative for mycoplasma contamination. Cells were passaged as single cells using Accutase. HeLa cells were cultured in DMEM, 10% fetal bovine serum, and 1% Penicillin-Streptomycin. To deplete mouse ES cells of 7SK, chimeric DNA-2′O-methylated RNA oligonucleotides were used as previously described[34] (Supplementary Table 6). Briefly, cells were trypsinzed and nucleofected using the Amaxa mES cell kit (Lonza) with 1nmole of ASO per 2 million mouse ES cells. After nucleofection cells were grown for 2-24hrs and used for downstream analysis.

Western blotting and Antibodies

Cells were lysed in lysis buffer (50mM HEPES, 200mM NaCl, 1mM EDTA, 10% Glycerol, 0.1% NP-40, 0.2% TritonX-100, 0.5% N-lauroylsarcosine; supplemented with Protease Inhibitor Cocktail (Roche)), briefly sonicated to solubilize chromatin, and spun for 10 minutes at 4°C at 13,000 rpm. Clarified protein lysates were quantified with the BCA Protein Assay Reagent Kit (Pierce). Antibodies used: Pol II N-term (Santa Cruz, sc-899), Ccnt1 (Santa Cruz, sc-10750), Cdk9 (Santa Cruz, sc-484), Brg1 (Santa Cruz, sc-17796), Hexim1 (Abcam, ab25388), Baf47 (Santa Cruz, sc-166165), Arid1a (Santa Cruz, sc-32761), Baf155 (Gift from G. Crabtree).

Glycerol Gradient Analysis

Mouse ES cells were depleted of 7SK as described above and biological duplicates of control and 7SK ASO were collected. After 12 hours of knockdown cells were lysed in Glycerol Gradient Lysis Buffer (10mM HEPES, 2mM $MgCl_2$, 10mM KCl, 0.5% NP-40, 0.5mM EDTA, 150mM NaCl; supplemented with Protease Inhibitor Cocktail (Roche) and SUPERaseIn (Life Technologies)) on ice for 15 minutes. Insoluble material was pelleted by centrifugation for 10 minutes at 4°C at 13,200 rpm. Whole cell protein lysates were quantified as above and 400μg of lysate was used for each individual gradient. For each sample, 10mL of a 10-30% glycerol gradient solution was mixed and 400μg of lysate (in 300μL) was loaded and spun for 16 hours at 4°C at 40,000 rpm in a Beckman SW41Ti rotor. After centrifugation and fractionation (300μL per fraction, 30 fractions, collected from the top) samples were analyzed by western blotting.

Start-seq centering of promoters and enhancers

Promoters—24,064 mm9 genes were obtained from RefSeq and all microRNA and snoRNA genes were removed. The most upstream TSS was obtained and a +/- 1kb window was calculated around each. For each gene, the 2kb window was first scanned along the sense strand to find the highest Start-seq[31] peak. The gene was discarded if it had no

position with at least 5 reads on the sense strand. Next, the 2kb window was scanned on the antisense strand to find the highest antisense peak upstream of the sense peak. A minimum threshold was not set for finding an antisense peak. If no antisense peak was found, the gene was labeled unidirectional. If this antisense peak was the sense peak for another gene both genes were labeled bidirectional. Otherwise, the gene was labeled divergent. Only genes with unique sense peaks were retained, resulting in a final list of 14,234 genes with Start-seq signal. Of these 14,234 genes, 2060 were bidirectional, 1443 were unidirectional, and 10731 were divergent. We then calculated the distance (TSS-pair width) between sense and antisense peaks for all non-unidirectional genes.

Traditional enhancers—8,563 mm9 enhancers were obtained from White et al.[4] enhancer calls and +/- 1kb windows were obtained around their centers. For each enhancer, the sense and antisense strands were simultaneously scanned to find the pair of sense and antisense nucleotides that maximized the sum of Start-seq signal on both strands and where the sense peak was downstream of the antisense peak (consistent with divergent transcription). Enhancers were discarded if Start-seq signal was not found on both strands, resulting in a final list of 5,356 enhancers. As with promoters, we then calculated the distance between sense and antisense peaks for all enhancers.

Super enhancers—231 mm9 super enhancer regions were obtained from White et al.[4] SE calls and intersected with peak calls from ATAC-seq chromatin occupancy data to obtain 415 high confidence ATAC-seq peaks within SE regions. All peaks (known as SE peaks) within 1kb of a RefSeq TSS were removed, resulting in a list of 361 SE peaks. The Start-seq centering algorithm we ran with regular enhancers did not result in better centered peaks, so we used these 361 SE peaks in our final list.

ChIRP-seq Assay

The ChIRP-seq assay was performed as described previously[50]. Mouse ES cells were cultured as above and treated with Dimethyl sulfoxide (DMSO) or 500nM JQ1[48] (Gift from J. Bradner) for 1 hour at 37°C. Isolated RNA was used in qRT-PCR analysis (Stratagene) to quantify enrichment of 7SK and depletion of other cellular RNAs. Isolated DNA was used for qPCR analysis or to make deep sequencing libraries with the NEBNext DNA Library Prep Master Mix Set for Illumina (NEB). Library DNA sequenced from a single end for 75 cycles on an Illumina HiSeq 2500 (Summary of all sequencing experiments are listed in Supplementary Table 5).

Sequencing reads were first trimmed of adaptors (FASTX Toolkit) and mapped using Bowtie to a custom bowtie index containing repetitive RNA elements (rRNA, snRNAs, and y-RNAs[51]). Subsequent reads were then mapped to mm9. Mapped reads were separately shifted towards the 3′ end using MACS and normalized to a total of 10 million reads. Even and Odd replicates were merged as described previously[50] by taking the lower of the two read density values at each nucleotide across the entire genome. The full pipeline is available at https://github.com/bdo311/chirpseq-analysis.

ChIRP-seq Data Analysis

Transcription start site (TSS) regions were obtained from the RefSeq annotations of genes in the hg19 genome for H1/HeLa ChIRP-seq data. For each type of genomic feature (promoters, SE, and TE), metagene plots were made by plotting the column means of an n×p matrix where the n^{th} row represents a single example of the genomic feature and the p^{th} column represents the read density at position p of the genomic feature. The full pipeline for producing metagene plots is available at github.com/bdo311/metagene-maker. 7SK ChIRP-seq peaks were called with MACS2 and defined as genic if the center of the peak was within 1kb of an annotated gene (intergenic otherwise).

Global Run On sequencing and Data Processing

Mouse ES cells were depleted of 7SK as described above and biological duplicates of control and 7SK ASO treated cells were collected. GRO-seq was performed as previously described[52] with the following modifications. After 5-bromouridine labeled RNA was isolated, deep sequencing libraries were constructed as described for FAST-iCLIP[51]. Libraries sequenced from a single end for 75 cycles on an Illumina HiSeq 2500 or NextSeq 500 machine.

For each sample, 3′ adaptor trimming, quality filtering, PCR duplicate collapsing, and 5′ barcode trimming was performed as previously described[51]. Processed reads were then mapped using Bowtie to a custom bowtie index containing single-copy loci of repetitive RNA elements (rRNA, snRNAs, and y-RNAs[51]). Reads that did not map to the custom index were then mapped to mm9. These steps were similar to the ChIRP-seq data processing above. Metagenes were made with metagene-maker and with a custom Perl script after separating sense from antisense reads. Statistics were calculated in a +500bp window surrounding the center of each element.

To calculate convergent transcription (ConvT) in Control and 7SK ASO-treated cells, we took the minimum of the sense signal and the antisense normalized GRO-seq read density at each nucleotide, such that if a nucleotide had GRO-seq signal only on one strand ConvT = 0. To calculate the proportion of ConvT blocks before and after 7SK knockdown for different regulatory elements, we averaged the WT and ASO ConvT tracks at 100-bp intervals. For each treatment, we marked each interval as a ConvT interval if its mean ConvT signal exceeded the genome-wide mean. Observed versus expected fold change was calculated by dividing the fraction of ConvT intervals mapping to each regulatory element type by the genomic fraction of that element type. Finally, metagene-maker was used to quantify average ConvT profiles in Control and ASO for each regulatory element. Statistics were calculated in a +/- 1000bp window surrounding the center of each element. A summary of all changes observed at promoter and enhancers can be found in Supplementary Table 7.

To calculate enhancer peak-to-peak distances, we took the distance from an enhancer to its closest TE or individual SE peak. Peaks were centered using Start-seq signal (TE) or ATAC-seq signal (SE) as described previously.

ATAC Sequencing and Data Processing

ATAC-seq was performed essentially as previously described[53]. Briefly, mouse ES cells were depleted of 7SK as above. Transposition was performed for 30 minutes at 37°C after which cells were lysed and DNA isolated Library DNA was sequenced from a single end for 75 cycles on an Illumina HiSeq 2500.

ATAC-seq reads were processed as previously described[53]. Once mapped reads were obtained, paired end ATAC-seq fragments from two length ranges: fragments shorter than 100bp, and fragment lengths between 147-180bp were isolated for separate analysis. These ranges were used by Buenrostro et al.[53] to distinguish transposase insertion into nucleosome free DNA and insertions flanking nucleosomes, respectively. At promoters, TE, and SE the number of fragment centers (single base pair) overlapping each bin was counted for each of the two fragment size ranges. We plotted the average fragment count in each bin normalized to the average fragment count in the first five bins in order to set the background signal to one. Statistics were calculated in a +/- 250bp window surrounding the center of each element

Transcription factor footprint detection with PIQ

We used PIQ[47] to quantify DNA binding of TFs based on ATAC-seq data. In order to maximize sensitivity we merged ATAC-seq data from replicates. To eliminate detection biases based on sequencing depth, we sampled from the higher depth merged library in order to achieve equal depth from control and 7SK-depleted libraries (resulting in ~89 million single end reads in each condition). Purity thresholds for footprint calls in control samples were adjusted to identify ten thousand binding events for each TF.

Immunofluorescence (IF)

IF was performed as described previously[33]. Briefly, mouse ES cells were treated with Control of 7SK ASO and seeded into 12-well plates containing 18-mm glass cover slips and cultured for 12 hours. After which cells were fixed in 4% PFA for 10min at room temperature, washed 3 × 5min with PBS. Cells were permeabilized in PBS containing 0.3% (v/v) Triton X-100 for 5min, and blocked overnight at 4°C in PBT buffer (PBS with 1%BSA, 0.1% Triton X-100 (v/v), 0.05% sodium azide (w/v)). After blocking, cover slips were incubated in PBT with 1:200 dilution of the S139-phosphorylated gamma-Histone 2A.X (γ-H2AX) antibody (Abcam, ab11174) at room temperature for 2 hours. Cover slips were wash 3× 5min with PBT and incubated with the Alexa-Fluor 568 secondary antibody (1:1000; Life Technologies) for 1 hour. Cells were washed 3× 5min with PBT, 2× 5min with PBS, rinsed briefly with water and mounted onto glass slides using VECTASHIELD (Vectorlabs, H-1200) mounting medium with DAPI. All images were taken and processed using a Zeiss LSM700 confocal microscope.

In vivo CLICK SHAPE (icSHAPE)

icSHAPE was performed largely as described previously[43] with the following modifications. Mouse ES cells were grown as above, cell lysates were made as for Glycerol Gradient analysis, quantified, and 400μg of each sample was used for further processing. For Hexim1 enrichment, 30μL of Protein A Dynabeads (Life Technologies) were prepared overnight at

4°C with 8μg of anti-Hexim1 antibody (Abcam, ab25388). For BAF (Arid1a) enrichment, 30μL of Protein A Dynabeads (Life Technologies) were prepared overnight at 4°C with 5μg of Rabbit anti-Mouse secondary antibody and 8μg of anti-Arid1a antibody (Santa Cruz, sc-32761). After overnight incubation, beads were washed twice in 1mL of NT2 buffer (50mM Tris pH 7.5, 150mM NaCl, 1mM EDTA, 0.05% NP-40) and added to freshly prepared lysate and immunoprecipitated overnight at 4°C. Samples were then washed three times in Glycerol Gradient Lysis buffer each for 2 minutes on ice. After washing, beads were resuspended in 30μL of icSHAPE reaction buffer (100mM HEPES, 6mM $MgCl_2$, 100mM NaCl, 1U SUPERaseIn (Life Technologies), 50mM NAI-N_3 [or DMSO as mock]) and incubated at 37°C for 12 minutes. After modification RNA was immediately isolated with TRIzol (Life Technologies) followed by RNA cleanup with RNeasy Mini Columns (Qiagen). Subsequently, RNA was treated as standard DMSO or NAI-N_3 samples for standard icSHAPE library preparation[43]. DNA library samples were submitted for deep sequenced on Illumina NextSeq 500 machines for single-end 75bp cycle runs.

Data processing steps were performed as described previously[43] with the following modifications. We generated biological replicates (BAF) and biological (Hexim1) for DMSO and NAI-N_3 samples. Reads were collapsed to remove PCR duplicates, barcodes removed, and mapped specifically to the full-length murine 7SK snRNA sequence (GeneID: 19817). In NAI-N_3 libraries we define the -1 position of the 5′ the mapped reads as RT (Reverse Transcription) stops, corresponding to structure modified positions. We defined RT stop coverage as the number of times a base is mapped as a RT stop. We constructed the background base density profile for 7SK as the sequencing depth of each base in the DMSO libraries. We defined reactivity score as the subtraction of background RT stops (DMSO libraries) from RT stops of the modified NAI-N_3 libraries, and then adjusted by the background base density. The score is then scaled into the range of [0, 1], after removing the outliers by 90% Winsorization (the top 5th percentile is set to 1 and the bottom 5th percentile is set to 0.

Native Immunoprecipitation (IP) and RNA-IP

Mouse ES cell extract was prepared for analysis of native protein-protein and RNA-protein interaction analysis and described previously[33]. Hexim1 (Abcam, ab25388) or BAF (Arid1a, Santa Cruz, sc-32761) were enriched with bead-coupled antibodies for four hours at 4°C. Enriched samples were washed three times on ice in Glycerol Gradient Lysis Buffer. Proteins or RNA pulled down with each IP were analyzed by western blot, mass spectrometry, or qRT-PCR.

ChIP-seq

Sample and Library preparation—ChIP-seq was performed as described previously[33]. DNA was purified for subsequent qPCR analysis or deep sequencing library construction (as above for ChIRP-seq). The following antibodies were used for ChIP studies: Pou5f1 (Santa Cruz, sc-8629), Hexim1 (Abcam, ab25388), Ddx21 (Novus, NB100-1781), and Baf155 (Gift from G. Crabtree), γ-H2AX (Abcam, ab11174), and Brd4 (Bethyl, A301-985A).

Factor Ratio Analysis—For all TE, SE, and promoter regions (-1kb to +1kb around their centers), we calculated a mean read density for all ChIP-seq, ATAC-seq, and 7SK ChIRP-seq datasets by dividing total coverage by the number of nucleotides (2000) and obtained enrichment values by dividing these values by the values for their corresponding input samples. Enrichment values for each factor at each region were all normalized by the enrichment of TBP[54], TBP[55], or TAF1[55] at that region. To make heatmaps, for all factors we averaged these TBP- or TAF1-normalized enrichment values across promoters, TE, and SE and displayed all factors or a subset of factors using R's pheatmap library. To show that promoters, TE, and SE have different factor ratios as a class, we ran principal component analysis on the TBP[54]-normalized enrichment values for all regions with the prcomp function in R and plotted the first two principal components using R's ggplot2 library.

Data Analysis—Raw ChIP-seq data were obtained from the Gene Expression Omnibus (Supplementary Table 8). All ChIP-seq data was mapped to the genome and normalized to 10 million reads to facilitate comparison with ChIRP-seq data. The format of ChIP-seq data is identical to the format of ChIRP-seq data: for each factor of interest, each position in the genome has a value that represents its binding level relative to other positions. Metagenes for ChIP-seq data were produced as described above and statistics were calculated in a +/-250bp window surrounding the center of each element.

In order to assess co-binding of Hexim1, Baf155, and 7SK we called peaks from 2 Hexim1 and 2 Baf155 ChIP-seq replicates and Even and Odd replicates of 7SK ChIRP-seq using MACS2 under the follow parameters: 7SK parameters = --broad --broad-cutoff 0.3 -p 1e-3. We obtained 323,145 overlapping peaks, of which 85,230 had signal values within 50% of each other in both datasets. We took the top 50,000. Baf155 parameters = --broad --broad-cutoff 0.3 -p 1e-3. We obtained 1,006,367 overlapping peaks, of which 959,238 had signal values within 50% of each other in both datasets. We took the top 10,000. Hexim1 parameters = -p 1e-3. We obtained 346,242 overlapping narrow peaks, of which 19,634 had IDR < 0.01. We took the top 10,000. These peaks were subsequently analyzed in two was. First each peak was assigned to either a promoter or enhancer element.

To derive a linear model to explain γ-H2AX fold change after 7SK knockdown, we used $\log_2$ ASO/Scr fold change values at TE and SE for GRO-seq, Baf155 ChIP-seq, γ-H2AX ChIP-seq, and ConvT. We used the R library edgeR to process the GRO-seq and ChIP-seq data and in-house Python scripts to process ConvT data. GRO-seq, Baf155, and ConvT values were then regressed onto γ-H2AX fold change one factor at a time using R's lm function, and the adjusted R-squared was calculated for each regression.

We used edgeR and displayed the data as volcano plots to quantitatively analyze the changes in ChIP-seq or GRO-seq signal. Specifically, we counted ChIP-seq or GRO-seq reads falling in specified windows around promoters and enhancer TSSs and used edgeR with two biological replicates for each condition (scramble or 7SK-targeting ASO) to identify the subset of regions exhibiting significantly altered signal in 7SK depleted cells. Read counts were normalized by library size for each sample. We used a threshold of false discovery rate < 0.1 to call significantly altered regions. Volcano plots represent the $\log_2$ fold change for the particular experiment (GRO-seq or ChIP-seq) plotted against $-\log_{10}$(FDR).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of Chang labs for discussion, and the following individual for reagents and advice. BAF: D. Hargreaves and G. Crabtree (Stanford University), GRO-seq: I. Jonkers and J. Lis (Cornell University), JQ1: J. Bradner (Dana Farber Cancer Institute), human ES cell culture: V. Sebastiano (Stanford University), traveling ratio: P. Rahl (Syros Pharmaceutical) and C. Lin (Baylor College of Medicine), and P. Batista and R.C. Spitale for critical reading of the manuscript. Supported by Stanford Medical Scientist Program and US National Institutes of Health (NIH) 1F30CA189514-01 (R.A.F.); NIH GM068122 and GM110050 (E.T.K.); Helen Hay Whitney Foundation (E.C.); NIH (P50-HG007735, R01-HG004361), California Institutes for Regenerative Medicine, and Howard Hughes Medical Institute (H.Y.C.).

References

1. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–461. [PubMed: 24670763]
2. Djebali S, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–108. [PubMed: 22955620]
3. Meng FL, et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell. 2014; 159:1538–1548. [PubMed: 25483776]
4. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
5. Hnisz D, et al. Super-Enhancers in the Control of Cell Identity and Disease. Cell. 2013; 155:934–947. [PubMed: 24119843]
6. Parker SCJ, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proceedings of the National Academy of Sciences. 2013; 110:17921–17926.
7. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]
8. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS. 2010; 107:21931–21936. [PubMed: 21106759]
9. He HH, et al. Nucleosome dynamics define transcriptional enhancers. Nature Genetics. 2010; 42:343–347. [PubMed: 20208536]
10. Core LJ, et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nature Genetics. 2014
11. Arner E, Daub CO, Vitting-Seerup K, Andersson R. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. 2015
12. Li W, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013; 498:516–520. [PubMed: 23728302]
13. Lam MTY, et al. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature. 2013; 498:511–515. [PubMed: 23728303]
14. Kwak H, Lis JT. Control of transcriptional elongation. Annu Rev Genet. 2013; 47:483–508. [PubMed: 24050178]
15. Campos EI, Reinberg D. Histones: annotating chromatin. Annu Rev Genet. 2009; 43:559–599. [PubMed: 19886812]
16. Ørom UA, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. Cell. 2013; 154:1190–1193. [PubMed: 24034243]
17. Kassube SA, et al. Structural Insights into Transcriptional Repression by Noncoding RNAs That Bind to Human Pol II. J Mol Biol. 2013; 425:3639–3648. [PubMed: 22954660]
18. Gurney T, Eliceiri GL. Intracellular distribution of low molecular weight RNA species in HeLa cells. J Cell Biol. 1980; 87:398–403. [PubMed: 6159360]

19. Bartkowiak B, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes & Development. 2010; 24:2303–2316. [PubMed: 20952539]
20. Yu M, et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science. 2015; 350:1383–1386. [PubMed: 26659056]
21. Zhou Q, Li T, Price DH. RNA polymerase II elongation control. Annu Rev Biochem. 2012; 81:119–143. [PubMed: 22404626]
22. D'Orso I, Frankel AD. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat Struct Mol Biol. 2010; 17:815–821. [PubMed: 20562857]
23. Ji X, et al. SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. Cell. 2013
24. Liu W, et al. Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. Cell. 2013; 155:1581–1595. [PubMed: 24360279]
25. McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. CellReports. 2013; 5:1256–1268.
26. McNamara RP, Reeder JE, McMillan EA, Bacon CW. KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Molecular Cell. 2015
27. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell. 2013; 152:1298–1307. [PubMed: 23498938]
28. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–693. [PubMed: 20616235]
29. McGinnis JL, Dunkle JA, Cate JHD, Weeks KM. The mechanisms of RNA SHAPE chemistry. J Am Chem Soc. 2012; 134:6617–6624. [PubMed: 22475022]
30. Pott S, Lieb JD. What are super-enhancers? Nature Genetics. 2015; 47:8–12. [PubMed: 25547603]
31. Williams LH, et al. Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. Molecular Cell. 2015
32. Hnisz D, et al. Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. Molecular Cell. 2015
33. Calo E, et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature. 2014
34. Castelo-Branco G, et al. The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. Genome Biol. 2013; 14:R98. [PubMed: 24044525]
35. Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proceedings of the National Academy of Sciences. 2011; 108:10460–10465.
36. Mayer A, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell. 2015; 161:541–554. [PubMed: 25910208]
37. Chu C, et al. Systematic Discovery of Xist RNA Binding Proteins. Cell. 2015
38. Kadoch C, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nature Genetics. 2013; 45:592–601. [PubMed: 23644491]
39. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]
40. Hainer SJ, et al. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. Genes & Development. 2015; 29:362–378. [PubMed: 25691467]
41. Prensner JR, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nature Genetics. 2013; 45:1392–1398. [PubMed: 24076601]
42. Han P, et al. A long noncoding RNA protects the heart from pathological hypertrophy. Nature. 2014:1–19.
43. Spitale RC, et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015; 519:486–490. [PubMed: 25799993]

44. Krueger BJ, et al. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res. 2008; 36:2219–2229. [PubMed: 18281698]
45. Dykhuizen EC, et al. BAF complexes facilitate decatenation of DNA by topoisomerase IIα. Nature. 2013; 497:624–627. [PubMed: 23698369]
46. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012; 13:613–626. [PubMed: 22868264]
47. Sherwood RI, et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol. 2014; 32:171–178. [PubMed: 24441470]
48. Filippakopoulos P, et al. Selective inhibition of BET bromodomains. Nature. 2010; 468:1067–1073. [PubMed: 20871596]
49. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nature Publishing Group. 2015; 17:47–62.
50. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Molecular Cell. 2011; 44:667–678. [PubMed: 21963238]
51. Flynn RA, et al. Dissecting noncoding and pathogen RNA-protein interactomes. RNA. 2014
52. Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife. 2014; 3:e02407–e02407. [PubMed: 24843027]
53. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods. 2013; 10:1213–1218. [PubMed: 24097267]
54. Rahl PB, et al. c-Myc regulates transcriptional pause release. Cell. 2010; 141:432–445. [PubMed: 20434984]
55. Liu Z, Scannell DR, Eisen MB, Tjian R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell. 2011; 146:720–731. [PubMed: 21884934]
56. Marz M, et al. Evolution of 7SK RNA and its protein partners in metazoa. Mol Biol Evol. 2009; 26:2821–2830. [PubMed: 19734296]
57. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]

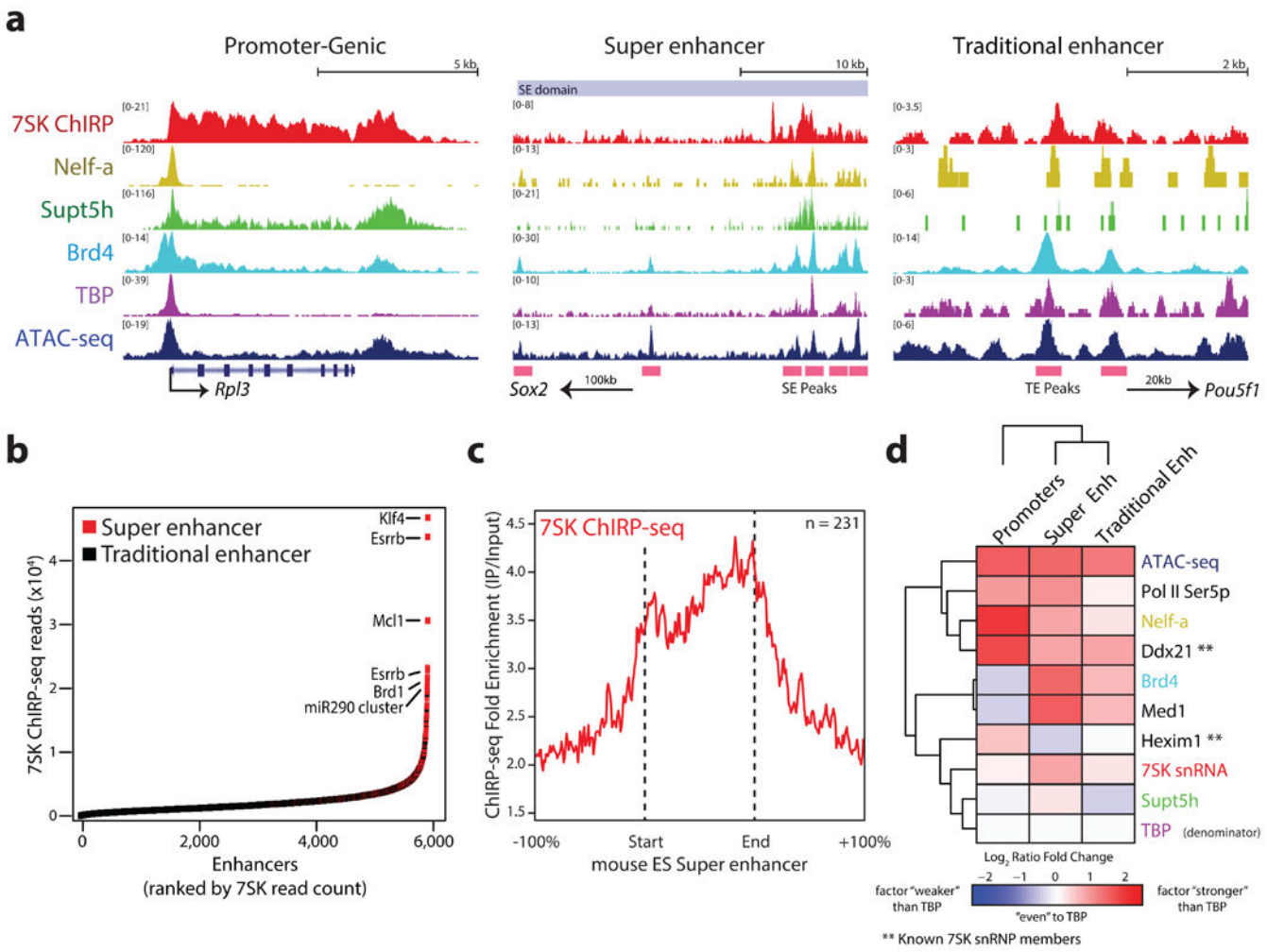

Figure 1. 7SK binds transcribed regions and prefers super enhancers

a, UCSC genome browser views of the *Rpl3* locus (left), *Sox2*-proximal super enhancer (SE, middle), and *Pou5f1*-distal typical enhancer (TE, right). 7SK snRNA (red), Nelf-a (yellow), Sup5th (green), Brd4 (turquoise), TBP (purple), and ATAC-seq open chromatin (blue) and shown for each locus. Enhancer peaks are marked with pink bars. **b**, Hockey-plot of 7SK ChIRP-seq across active mouse ES cell enhancers. Regions defined as SE[4] are marked in red. **c**, Metagene analysis of 7SK ChIRP-seq enrichment at 231 (number of elements, "n") previously reported SE elements[4]. The up and downstream regions of each SE element are scaled to display the same width of each element. **d**, Heat map of Factor Ratio Analysis at promoter, SE, and TE elements. Colors indicate $\log_2$ ratio fold change relative to TBP.

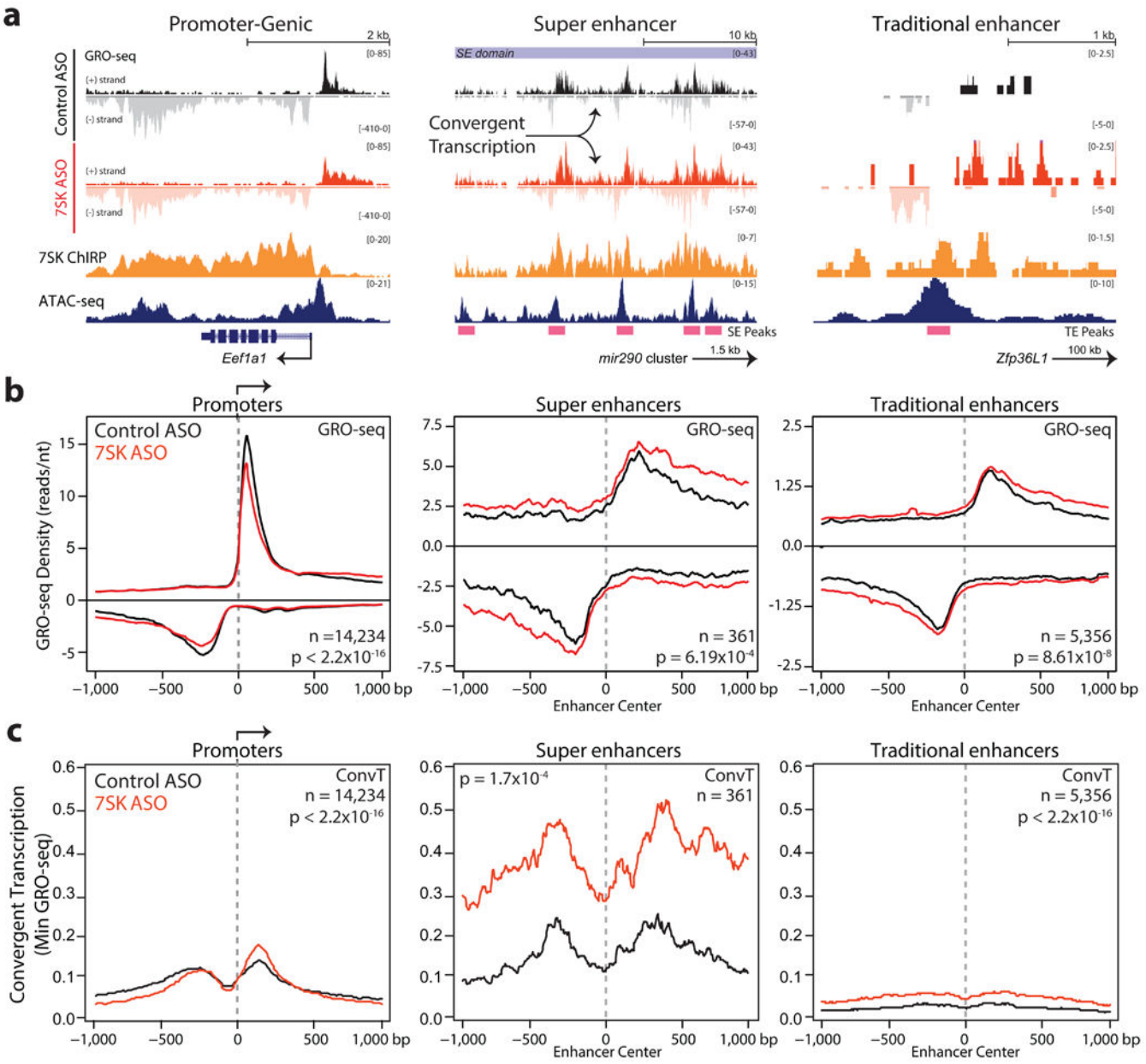

Figure 2. 7SK differentially regulates transcription at enhancers and promoters, controlling convergent transcription and DNA damage signaling

a, UCSC genome browser views of the *Eef1a1* locus (left), *microRNA-290* family SE (middle) and the *Zfp36L1* distal enhancer (right). Sense (black = control ASO [antisense oligonucleotide] and red = 7SK ASO) and antisense (grey = control ASO and light red = 7SK ASO grey) GRO-seq reads, 7SK ChIRP-seq (orange), and ATAC-seq open chromatin (blue) are shown with enhancer peaks marked as in Fig. 1. **b**, Metagene analysis of GRO-seq signal at promoters (left), SE (middle), and TE (right) centered at each element +/- 1kb. Control and 7SK ASO conditions are shown in black and red, respectively. P-values are calculated with the Hotelling's t-test. **c**, Metagene analysis of convergent transcription

(ConvT) as measured by GRO-seq, plotted as in (**b**), and scaled to the same ConvT value (y-axis).

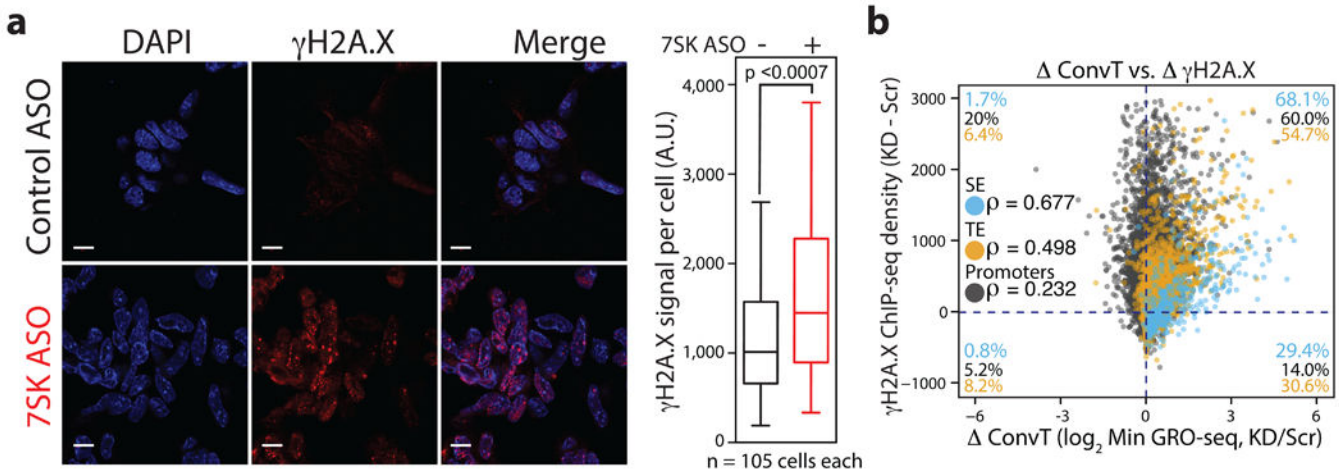

Figure 3. 7SK prevents DNA damage at enhancers and super enhancers

a, Representative immunofluorescence images (left) of paraformaldehyde-fixed mouse ES cells. Scale bars, 10 μm. DAPI, 49,6-diamidino- 2-phenylindole. Quantification (right) of 105 individual cells from Control ASO or 7SK ASO treated mouse ES cells. Boxes represent 25th and 75th percentile and median values are plotted. P-values calculated with the K-S test. **b**, Scatter plot analysis of the 7SK ASO/Control ASO change in γ-H2AX ChIP-seq vs. $\log_2$ change in ConvT at promoters, SE, and TE. Spearman correlation values are noted for each element class.

a

ChIRP-MS peptide count

Protein	*7SK*	*U1*	*U2*	*Xist*	*Protein Complex*
Larp7	1,027	0	0	0	7SK snRNA
Mepce	657	0	0	0	7SK snRNA
Ccnt1	655	0	0	0	7SK snRNA
Cdk9	451	0	0	0	7SK snRNA
Hexim1	342	0	0	0	7SK snRNA
Ddx21	116	77	29	9	7SK snRNA/snoRNP
Ctr9	69	24	29	0	Pol II Elongation
Smarcc1/Baf155	69	0	0	0	BAF
Smarca4/Brg1	47	35	35	0	BAF
Brd4	42	0	0	0	7SK snRNA
Smarcd1/Baf60a	27	0	0	0	BAF
Actl6a/Baf53a	15	0	0	0	BAF
Smarcb1/Baf47	8	0	0	0	BAF

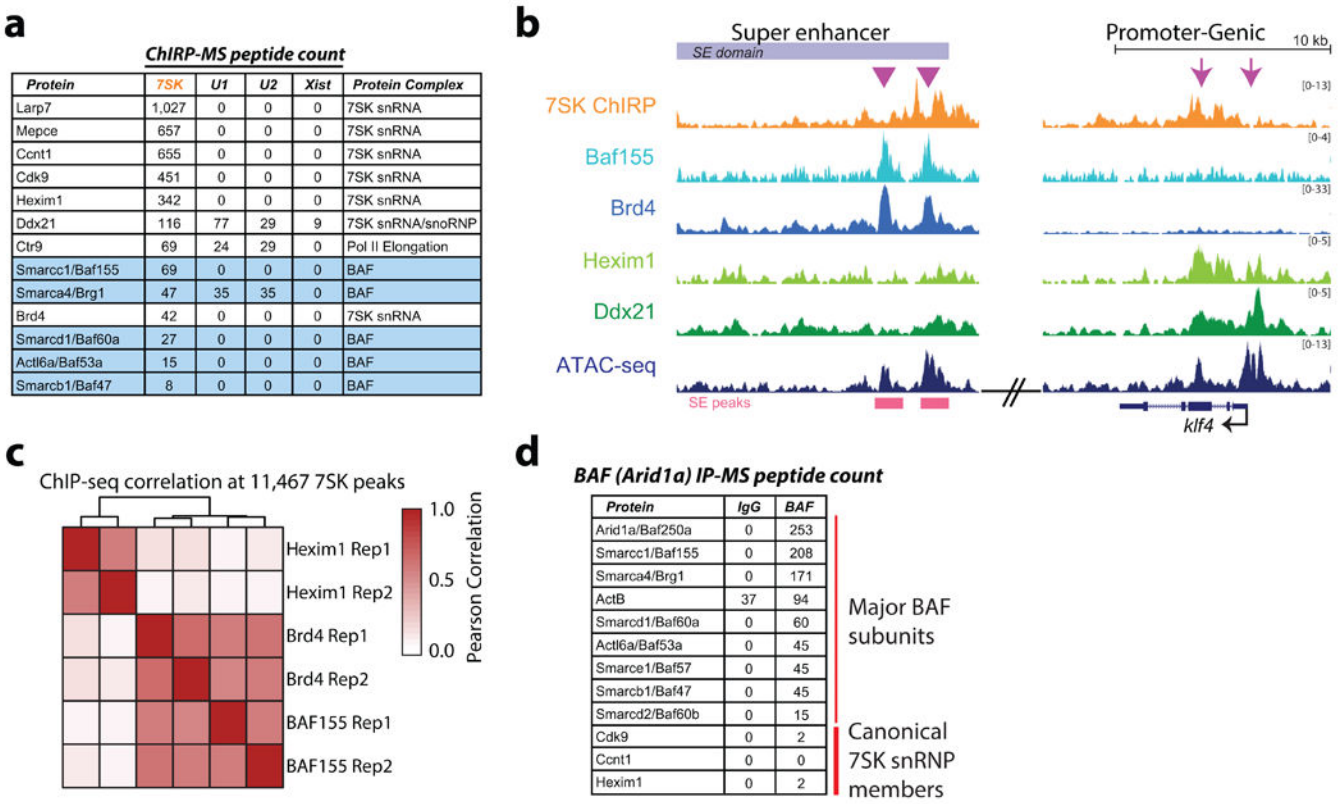

Figure 4. 7SK interacts with BAF in a distinct snRNP

a, Proteins identified in ChIRP-MS of the 7SK snRNA, U1 snRNA, U2 snRNA, and Xist RNA[37]. For each associated protein the ranked value of its enrichment from a ChIRP-MS is shown. **b**, UCSC genome browser view of the *Klf4* SE (left) and genomic locus (right). Triangles (left) and arrows (right) denote regions of common binding for 7SK-SE and 7SK-genic associated complexes. **c**, Pearson correlation analysis of Hexim1, Baf155, and Brd4 ChIP-seq density at 11,467 7SK ChIRP-seq peaks. **d**, Proteins identified in IP-MS of the BAF (Arid1a) complex. Values of BAF or 7SK snRNP subunits are shown (full list in Supplementary Table 3).

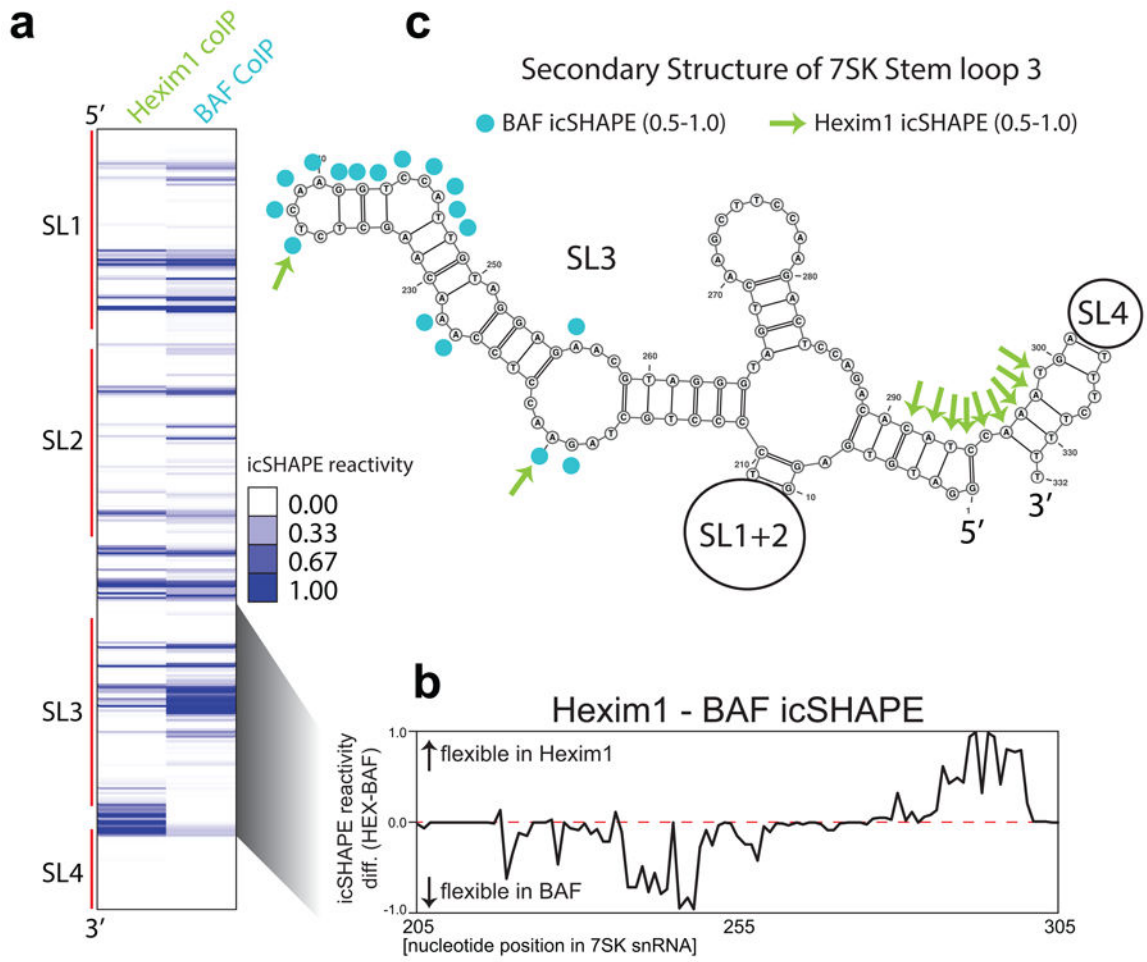

Figure 5. Two distinct 7SK RNA conformations in 7SK-BAF vs 7SK-Hexim complexes

a, icSHAPE of Hexim1- or BAF-associated 7SK snRNA. The first 5 and last 15 nucleotides are not analyzed for structural reactivity and highly reactive bases are shown as blue. **b**, icSHAPE difference analysis between Hexim1- or BAF-associated 7SK from nucleotides 205 to 305. Positive values indicate more single stranded in Hexim1 and negative values are bases more reactive in BAF. **c**, Secondary structure prediction of the 7SK snRNA with nucleotides differentially reactive in Hexim1- or BAF-associated structures (above icSHAPE values of 0.5). Stem loops (SL) 1, 2, and 4 are cartooned as loops.

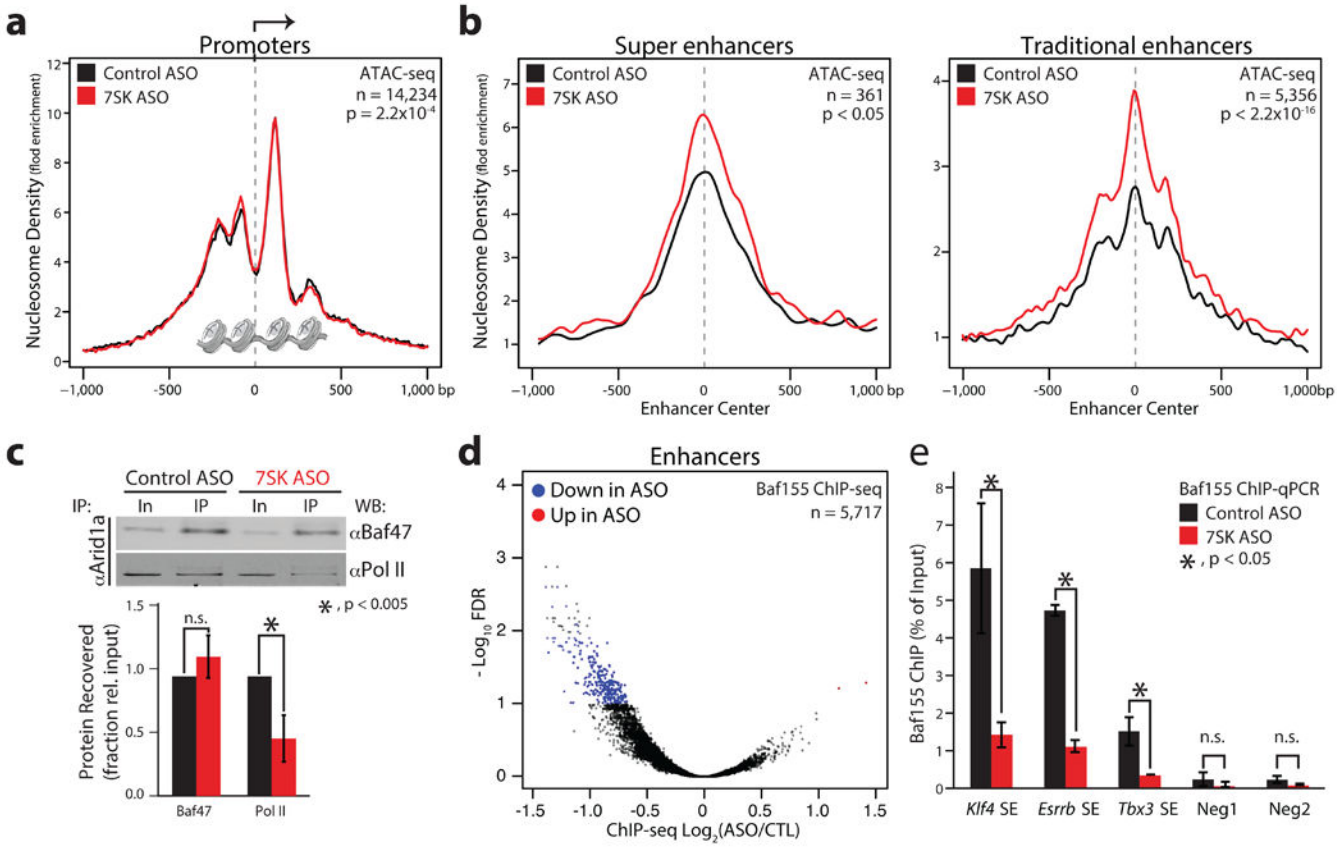

Figure 6. 7SK recruits BAF and modulates nucleosome phasing at enhancers

a,**b**, Metagene analysis of mononucleosome ATAC-seq signal at promoters (**a**), SE (**b**, left) and TE (**b**, right) +/- 1kb in mouse ES cells treated with control (black) or 7SK (red) ASOs. Nucleosomes are cartooned in grey. P-values calculated with the Hotelling's T-test. **c**, Co-immunopreciptiation of Arid1a followed by western blot analysis (top) and quantification (bottom). Quantification was performed from three biological replicates mean values +/- S.E.M are shown. * $p < 0.005$, two sided t-test. **d**, Volcano plot analysis of all enhancers with significant (FDR < 0.1) increased (red) or decreased (blue) Baf155 ChIP-seq signal 7SK depletion. **e**, Baf155 ChIP-qPCR analysis of mouse ES cells. Data are mean and s.d. of two biological samples collected from separate cultures on different days. * $p < 0.05$, two sided t-test.

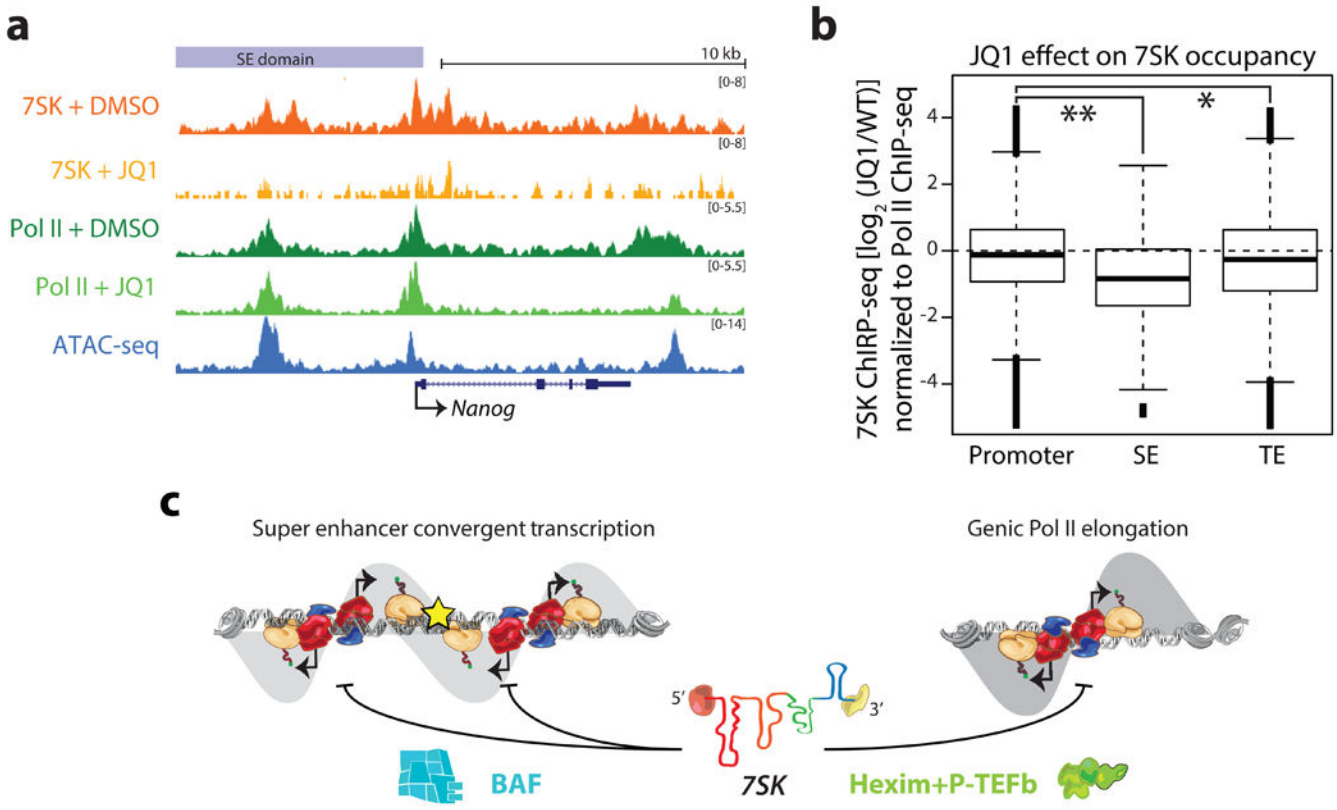

Figure 7. 7SK is recruited to enhancers via bromodomain interaction

a, UCSC genome browser view of the *Nanog* SE and genomic locus. 7SK ChIRP and Pol II ChIP-seq in mouse ES cells treated for 1 hour with DMSO or JQ1 are shown. **b**, Quantitative analysis of 7SK sensitivity to JQ1 treatment at promoters, SE, and TE, normalized to the levels of Pol II ChIP-seq. * $p < 6.5\times10^{-14}$, ** $p < 2.2\times10^{-16}$, K-S test. **c**, Model of 7SK's concerted action to repress transcription at enhancers with the BAF complex and at promoters via the Hexim1-P-TEFb snRNP.