Article

Cell

The Ground State and Evolution of Promoter Region Directionality

Yi Jin,[1,3] Umut Eser,[2,3] Kevin Struhl,[1,*] and L. Stirling Churchman[2,4,*]
[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: kevin@hms.harvard.edu (K.S.), churchman@genetics.med.harvard.edu (L.S.C.)
http://dx.doi.org/10.1016/j.cell.2017.07.006

SUMMARY

Eukaryotic promoter regions are frequently divergently transcribed in vivo, but it is unknown whether the resultant antisense RNAs are a mechanistic by-product of RNA polymerase II (Pol II) transcription or biologically meaningful. Here, we use a functional evolutionary approach that involves nascent transcript mapping in *S. cerevisiae* strains containing foreign yeast DNA. Promoter regions in foreign environments lose the directionality they have in their native species. Strikingly, fortuitous promoter regions arising in foreign DNA produce equal transcription in both directions, indicating that divergent transcription is a mechanistic feature that does not imply a function for these transcripts. Fortuitous promoter regions arising during evolution promote bidirectional transcription and over time are purged through mutation or retained to enable new functionality. Similarly, human transcription is more bidirectional at newly evolved enhancers and promoter regions. Thus, promoter regions are intrinsically bidirectional and are shaped by evolution to bias transcription toward coding versus non-coding RNAs.

INTRODUCTION

Eukaryotic promoter regions are nucleosome-depleted regions that contain binding sites for transcriptional activator proteins and core promoters bound by general transcription factors and RNA polymerase II (Pol II) assembled into a preinitiation complex (Burke et al., 1998; Smale, 1997, 2001; Struhl, 1987; Weis and Reinberg, 1992). Activator-binding sites in promoters and enhancers can function bidirectionally, but the preinitiation complex is intrinsically asymmetric and hence promotes transcription in only one direction. Nevertheless, most eukaryotic promoter regions generate divergent transcripts, many of which are antisense non-coding RNAs that are rapidly degraded by the nuclear exosome (Almada et al., 2013; Core and Lis, 2008; Flynn et al., 2011; Kilchert et al., 2016; Neil et al., 2009; Ntini et al., 2013; Rege et al., 2015; Seila et al., 2008; Vera and Dowell, 2016). Although a given coding transcript and the divergent upstream non-coding transcript share the same promoter region, each transcript originates from a different preinitiation complex (Rhee and Pugh, 2012) and thus are initiated by different core promoters. In addition, divergent non-coding transcripts are often observed in enhancers that can be located far upstream or downstream of the promoter region.

Several possible functions, such as maintenance of nucleosome-depleted regions (NDRs) and de novo gene formation (Scruggs et al., 2015; Wu and Sharp, 2013), have been proposed for divergent transcription but none has been demonstrated experimentally. On the other hand, divergent transcription may be the by-product of an open chromatin region and thus represent transcriptional noise (de Boer et al., 2014; Seila et al., 2009; Struhl, 2007). At the heart of this debate lies the question of intrinsic directionality. Are promoter regions intrinsically unidirectional and then shaped by evolution to support divergent transcription, or are they intrinsically bidirectional? Consistent with the unidirectional model, divergent transcription is not observed equally across eukaryotes (Core et al., 2012) and sense and antisense divergent transcription rates do not correlate (Churchman and Weissman, 2011). Further, directionality is controlled by a number of regulators (Churchman and Weissman, 2011; Marquardt et al., 2014; Tan-Wong et al., 2012; Whitehouse et al., 2007). On the other hand, pervasive divergent transcription across fungal and mammalian genomes (>80% of promoter regions) supports the idea that promoter regions are intrinsically bidirectional.

Functional analyses of native, and hence highly evolved, organisms in vivo cannot distinguish whether bidirectional promoter regions and non-coding transcripts are a mechanistic consequence of transcription or an evolved biological function. In principle, this distinction can be addressed by analyzing transcription of evolutionarily irrelevant DNA.

In this study, we use a functional evolutionary approach to investigate the intrinsic directionality of yeast promoter regions and how promoter region directionality evolves (Hughes et al., 2012). Specifically, we compare Pol II occupancy across native *S. cerevisiae*, *K. lactis*, and *D. hansenii* genomes with *S. cerevisiae* strains containing large regions of these foreign yeast species genome. Of particular note, we previously described nucleosome-depleted regions that fortuitously occur in *D. hansenii* coding regions when they are present in *S. cerevisiae* (Hughes et al., 2012). These regions, which presumably arise by fortuitous binding of *S. cerevisiae* activators that recruit nucleosome remodelers, often function as promoters,

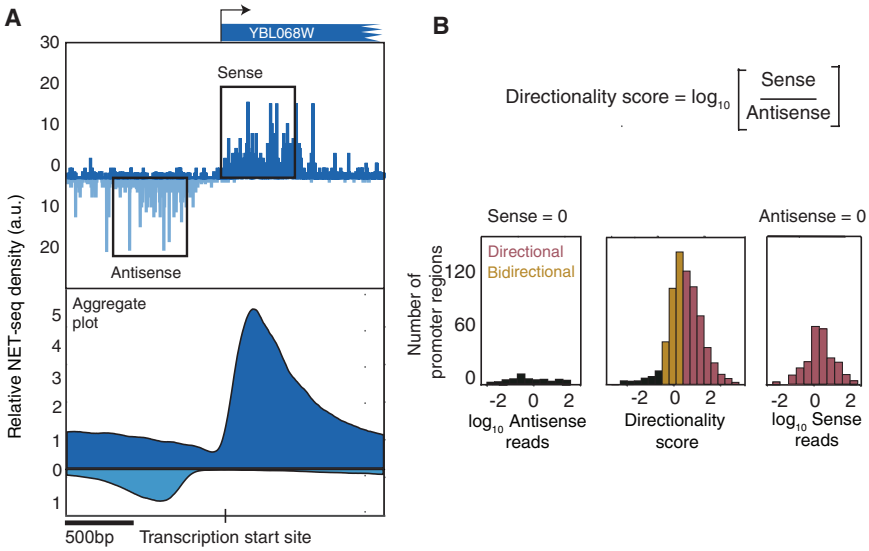

Figure 1. Promoter Region Directionality across the *Saccharomyces cerevisiae* Genome

(A) NET-seq analysis of the *Saccharomyces cerevisiae* genome. The upper panel shows the distribution of active Pol II for the promoter region of the YBL068W gene. Transcription in the sense and antisense directions are plotted above and below the horizontal axis, respectively. The bottom panel shows the aggregate plot of NET-seq reads averaged over all gene promoter regions by aligning to their TSS. Only promoter regions between tandemly oriented genes were included to ensure that the antisense transcript is non-coding. (B) Directionality score is defined as the log10 ratio of sense and antisense reads measured within 500-bp windows situated upstream and downstream of TSS (shown as boxes in Figure 1A). Genome-wide distribution of the directionality score is displayed in the middle panel. For promoter regions lacking sense (left panel) or antisense reads (right panel), distributions of the antisense or sense reads are displayed instead. Promoter regions were categorized as directional (yellow) if the sense to antisense ratio was ≥ 3 and bidirectional (pink) if the ratio was <3.
See also Figure S1.

although transcription was minimally characterized. As *D. hansenii* coding regions are evolutionarily irrelevant for transcriptional initiation in *S. cerevisiae*, they represent an ideal case in which to mechanistically examine the issue of bidirectional transcription in the absence of evolutionary constraints. We then combine this information with evolutionary analysis of yeast species to address how promoter directionality has evolved.

RESULTS

Transcription Is Biased toward the Coding Direction

While coding mRNAs are relatively long-lived, the corresponding antisense transcripts are rapidly degraded non-coding RNAs. Consequently, to quantitate promoter directionality, nascent transcription in each direction must be monitored. Using native elongating transcript sequencing (NET-seq) to precisely and quantitatively map engaged Pol II complexes across the *S. cerevisiae* genome (Churchman and Weissman, 2011), we define a directionality score based on the ratio of sense and antisense reads (Figures 1A and 1B). Because of the compact yeast genome, we focused our analysis on promoter regions between tandemly oriented genes in order to avoid analyzing coding transcription in the antisense direction. Consistent with previous reports (Churchman and Weissman, 2011), *S. cerevisiae* promoter regions exhibit higher sense transcription than divergent antisense transcription on average (Figures 1B, S1A, and S1B) The majority (>70%) of promoter regions are "directional," defined by at least three times more sense transcription than antisense transcription, including highly directional cases in which no antisense is detected. On the other hand, approximately a quarter of promoter regions are "bidirectional" that we define as sense: antisense ratios between 1/3 and 3 (Figure 1B). Importantly, NET-seq measures of directionality positively correlate with directionality measures using TFIIB ChIP-exo data (Rhee and Pugh, 2012) that correspond to initiation rates and 4tU-seq (Schulz et al., 2013) data that correspond to synthesis rates, indicating that bias in directionality largely arises from biases in initiation, not in elongation (Figure S1C). Thus yeast promoter regions are largely directional but exhibit substantial variability.

Directionality Loss in Foreign Environment

To address whether promoter regions are intrinsically unidirectional or bidirectional, we analyzed promoter directionality in *K. lactis*, in *D. hansenii*, and in five *S. cerevisiae* strains each containing a yeast artificial chromosome (YAC) harboring a ~150-kb piece of *K. lactis* or *D. hansenii* DNA (Figures 2A and S1D; see STAR Methods) (Hughes et al., 2012). Because of the lack of transcription start site annotation for these yeast species, we developed FIDDLE (flexible integration of data with deep learning), an integrative deep learning tool that leverages multiple types of available genomics data to predict genome-wide transcription start sites (Eser and Churchman, 2016). FIDDLE is capable of predicting transcription start site (TSS) with nearly the same accuracy as 5′ end mapping techniques, such as TSS-seq (Malabat et al., 2015) (Figures S2A and S2B).

Like in *S. cerevisiae*, promoter regions in both *K. lactis* and *D. hansenii* are predominantly directional (Figures 2B and 2C). Interestingly, a global-scale comparison of native *K. lactis* and *D. hansenii* with the *S. cerevisiae* strains containing the corresponding foreign yeast DNA reveals a reproducible loss in promoter directionality when DNA is in a foreign environment (p values $< 10^{-4}$ and 2.3×10^{-3} for *K. lactis* and *D. hansenii*, respectively, using a Kolmogorov-Smirnov [KS] test) (Figures 2B, 2C, and S2C–S2E). We do not observe an enrichment of any *S. cerevisiae* transcription factor motifs at promoter regions that change in directionality (Figure S2F). Thus, the decrease in

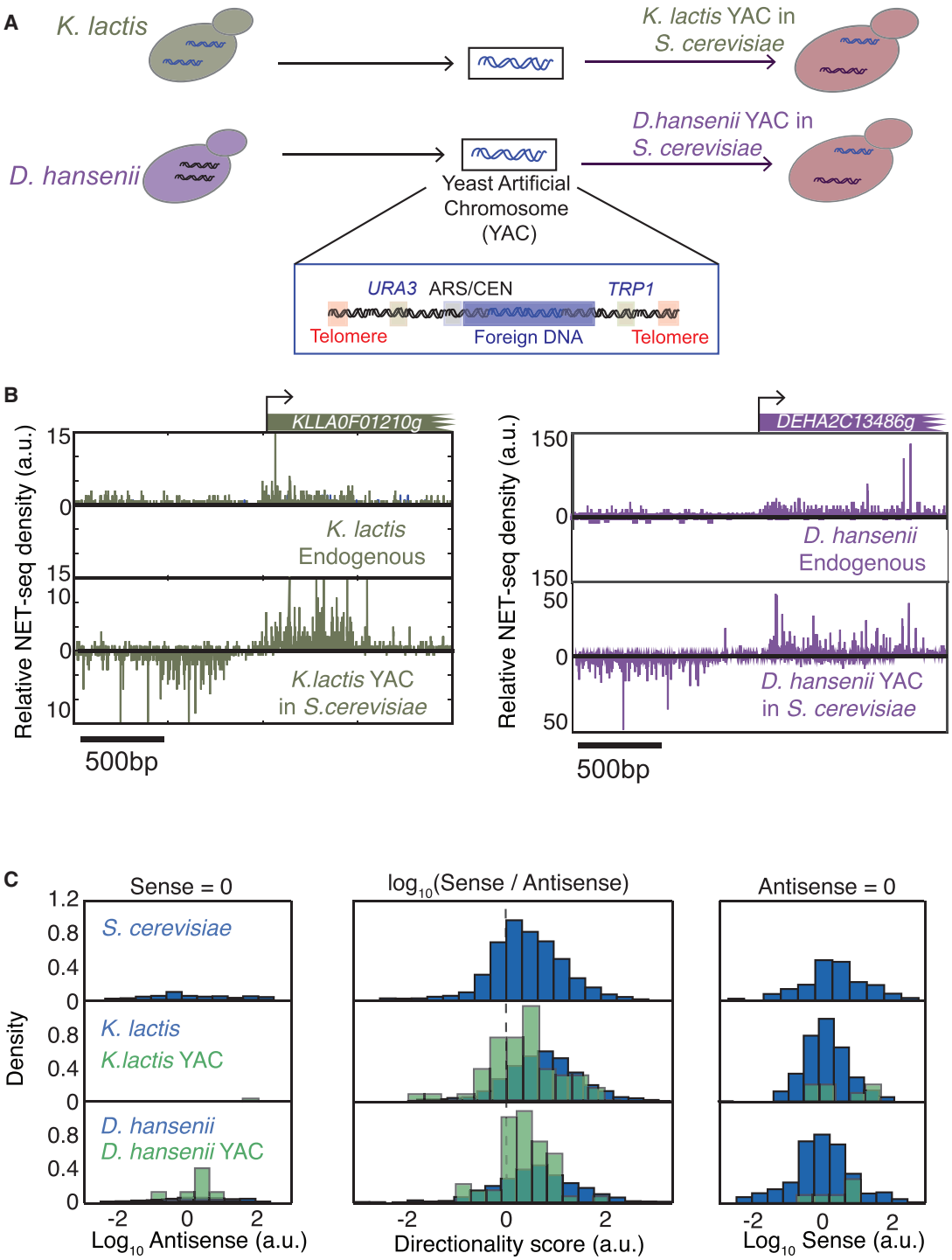

Figure 2. Promoter Region Directionality Is Lost in a Foreign Environment

(A) Chromosome pieces extracted from *K. lactis* and *D. hansenii* were inserted into yeast artificial chromosomes (YACs) containing centromere and telomere sequences and selective markers on both arms.

(B) NET-seq reads for two promoter regions from *K. lactis* (left) and *D. hansenii* (right) are shown in their native environments and in the foreign (*S. cerevisiae*) environment.

(C) Genome-wide distributions of the directionality score native species and YAC *S. cerevisiae* strains are displayed in the middle panel. For promoter regions lacking sense (left panel) or antisense reads (right panel), distributions of the antisense or sense reads are displayed instead. The p values of two sample Kolmogorov-Smirnov (KS) test for YAC and native distributions are 3.8×10^{-7} (*K. lactis*) and 2.0×10^{-9} (*D. hansenii*).

See also Figures S1 and S2.

directional bias suggests that some DNA sequences on the heterologous promoter regions are no longer recognized by *S. cerevisiae* proteins, and consequently are unable to promote directionality in the foreign environment. Therefore, the overall decrease in directionality and the absence of highly directional transcription from the heterologous promoter regions suggests that the ground-state of transcription is bidirectional.

Fortuitous Promoter Regions Generate Equal Bidirectional Transcription

In the experiments described above, promoter directionality is not entirely lost in a foreign environment, due either to residual detection of foreign DNA sequences by *S. cerevisiae* proteins or an intrinsically directional property of promoter regions. To discriminate between these possibilities, we examined de novo promoter regions that do not exist in the native organism and hence are evolutionarily irrelevant in *S. cerevisiae*. Using FIDDLE, we identified in an unbiased manner 43 *D. hansenii* coding sequences that act as fortuitous promoter regions when placed in the foreign environment of the *S. cerevisiae* nucleus (Figures 3A, 3B, and S3A). Importantly, we observed that fortuitous promoter regions are predominantly bidirectional, certainly more bidirectional than directionality in native *D. hansenii* (p value < 10^{-12} using a KS test) or *D. hansenii* YACs (p value < 10^{-5} using a KS test) (Figures 3C and S3B). These fortuitous promoter regions are depleted for nucleosomes with phased nucleosomes on both sides and are enriched for the general transcription factor, TFIIB (Figure 3B) (Hughes et al., 2012). The majority of de novo nucleosome-depleted regions (71%) co-occur with a fortuitous promoter region. Core promoter elements, TATA-like motifs, are not enriched in fortuitous promoter regions (Table S1). While certain core promoter elements produce strong levels of transcription, a wide variety of DNA sequences are capable of inducing initiation (Lubliner et al., 2013, 2015; Smale and Kadonaga, 2003). Thus, we propose that the de novo removal of nucleosomes by activators would expose DNA to the transcription machinery, resulting in modest levels of transcription initiation on both strands. In sum, these observations demonstrate that promoter regions are intrinsically bidirectional and hence that directional promoters are molded by evolution.

Newly Evolved Promoter Regions Are Less Directional Than Are More Evolved Promoter Regions

The idea that evolutionary pressure drives promoter regions away from their intrinsic bidirectionality predicts that directionality should increase in accord with evolutionary time. Using sequence alignment from seven *Saccharomyces* species (Siepel et al., 2005), we calculated genomic evolutionary rate profiling (GERP) scores that reflect the deficit in nucleotide substitution arising from selective pressure on a particular element that constrains the DNA sequence (Cooper et al., 2005a). This analysis reveals that in *S. cerevisiae* directional promoter regions contain more constrained elements than do bidirectional promoter regions, indicating that they have experienced higher levels of evolutionary selection (p value of 0.02 using a KS test) (Figure 4A). In addition, we coarsely categorized genes by evolutionary time by comparing *S. cerevisiae* genes whose orthologs are found only in the *Saccharomyces sensu stricto* genus to all other genes (Carvunis et al., 2012). Promoter regions of *sensu stricto* only genes are less directional than are those genes also found in other yeast species (p value < 10^{-12} using a KS test) (Figures 4B and 4C). Together, these analyses suggest that the promoter regions evolve to support directional transcription.

To investigate which sequences may have been selected for and may confer directionality, we searched for elements that show differential enrichment between directional and bidirectional promoter regions. Elements that make strong core promoters, such as the TATA box, showed similar enrichments across both classes of promoter regions with little to no effect (Figure S4). Instead, we postulated that the selection for transcription factor binding motifs might be responsible for altering promoter region directionality. We determined the preferential enrichment of annotated yeast TF motifs across directional and bidirectional promoter regions. We found that 12 transcription factor motifs were statistically more enriched at directional promoter regions and zero motifs were more enriched at bidirectional promoter regions (Table S2). These analyses suggest that directionality is orchestrated by a set of asymmetrical activators (or repressors) that increase sense transcription and/or repress antisense transcription.

Evolutionary Resolution of Fortuitous Promoter Regions

Although the fortuitous promoter regions described here arise in the context of an artificial experiment, they are analogous to new promoter regions that inevitably arise during evolution via fortuitous changes in DNA sequences and/or transcription factors. After the generation of such a novel promoter region, the organism can take either of two possible evolutionary paths (Figure 5A): purge or retain the novel promoter region. To look for such events, we identified *S. cerevisiae* transcription factor binding motifs that are more frequently observed at fortuitous versus native *D. hansenii* promoter regions (Figures 3D and S5; Table S3). At the top of the list are Reb1 and Abf1, constitutive regulatory factors that bind to many genes in *S. cerevisiae* and function through the recruitment of chromatin remodelers that displace nucleosomes (Ganapathi et al., 2011; Hartley and Madhani, 2009; Raisner et al., 2005); these proteins are present in *K. lactis*, but in *D. hansenii* are not present (Abf1) or serve a different function (Reb1) (Tsankov et al., 2010; Wapinski et al., 2007). Consistently, fortuitous promoter regions are present in *D. hansenii* YACs, but not present in the *K. lactis* YACs, suggesting that fortuitous promoter regions arise when DNA is placed into a foreign environment in which the set of transcription factors differ from the endogenous environment. In addition, the frequencies of Reb1 and Abf1 binding sites across coding sequences in 23 yeast species vary as a function of phylogenetic branching point relative to *S. cerevisiae*. Binding site frequencies are systematically lower in coding regions after the whole-genome duplication (WGD) event, when these binding sites became functional (Figure 5B), suggesting that these sites were purged from coding regions to discourage the formation of fortuitous promoter regions.

Conversely, a fortuitous promoter region and the new transcripts arising within a coding region might be utilized and evolutionarily selected, particularly after the WGD when the other copy of the original gene would remain. In this scenario, transcription factor binding sites found at fortuitous promoter regions should

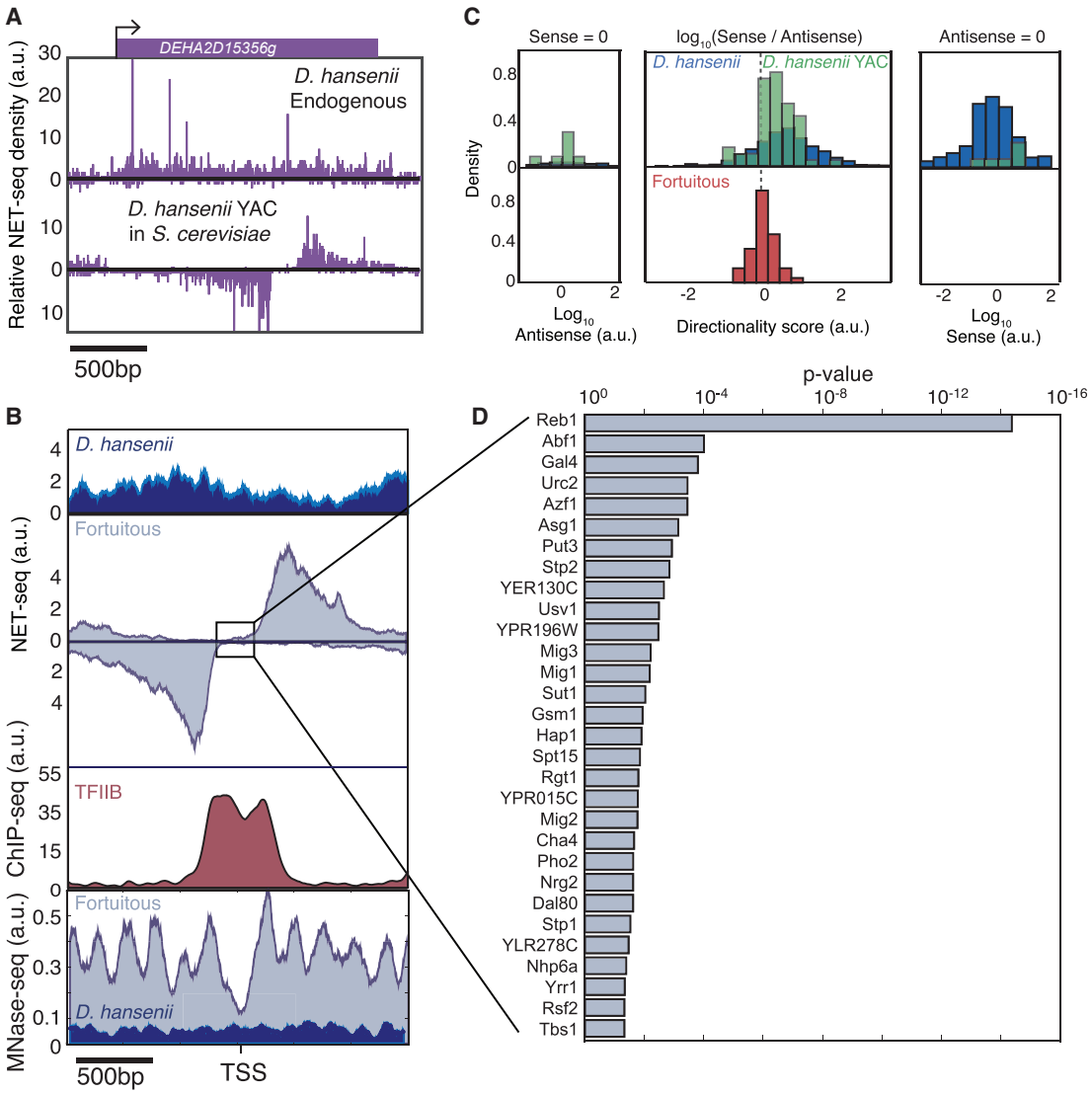

Figure 3. Fortuitous Promoter Regions Arise in Foreign Environments and Produce Bidirectional Transcription

(A) Example of a fortuitous promoter region emerging within the coding sequence of the *D. hansenii* gene, DEHA2D15356g when in a foreign (*S. cerevisiae*) environment.

(B) Aggregate plot of the average NET-seq reads over the fortuitous promoter regions in native and YAC strains (upper two panels). Transcription in the sense and antisense directions are plotted above and below the horizontal axis, respectively. Aggregate plots are shown for TFIIB chromatin immunoprecipitation sequencing (ChIP-seq) in YAC strains (dark red) and MNase-seq in YAC (blue gray) and native (dark blue) strains over the fortuitous promoters are shown (bottom two panels).

(C) Histogram of directionality scores for native *D. hansenii* (blue, upper), corresponding YACs (green, upper), and the fortuitous promoter regions (bottom). Genome-wide distributions of the directionality score are displayed in the middle panels. For promoter regions lacking sense (left panels) or antisense reads (right panels), distributions of the antisense or sense reads are displayed instead. Two sample KS test p values are 7.1×10^{-12} and 2.1×10^{-4} for when comparing fortuitous distribution to *D. hansenii* native and YAC distributions, respectively.

(D) Transcription factors whose binding sites are significantly enriched at fortuitous promoter regions. p values are determined through comparison of binding site density at fortuitous promoter regions compared to *D. hansenii* native promoter regions. Table S3 summarizes the data.

See also Figures S3 and S5 and Tables S1 and S3.

also be found at the promoters of newly evolved genes. We identified a set of transcription factor binding sites that are specifically enriched at *S. cerevisiae* promoter regions as compared to coding regions and found that only a subset of these are located at promoters of *sensu stricto* only genes. In addition, there is high overlap between the transcription factor binding sites enriched in the promoters of newly evolved (*sensu stricto* only) genes and fortuitous promoter regions that are highly distinct from those enriched at older genes (Figure 5C), suggesting that some newly evolved genes may have started as a

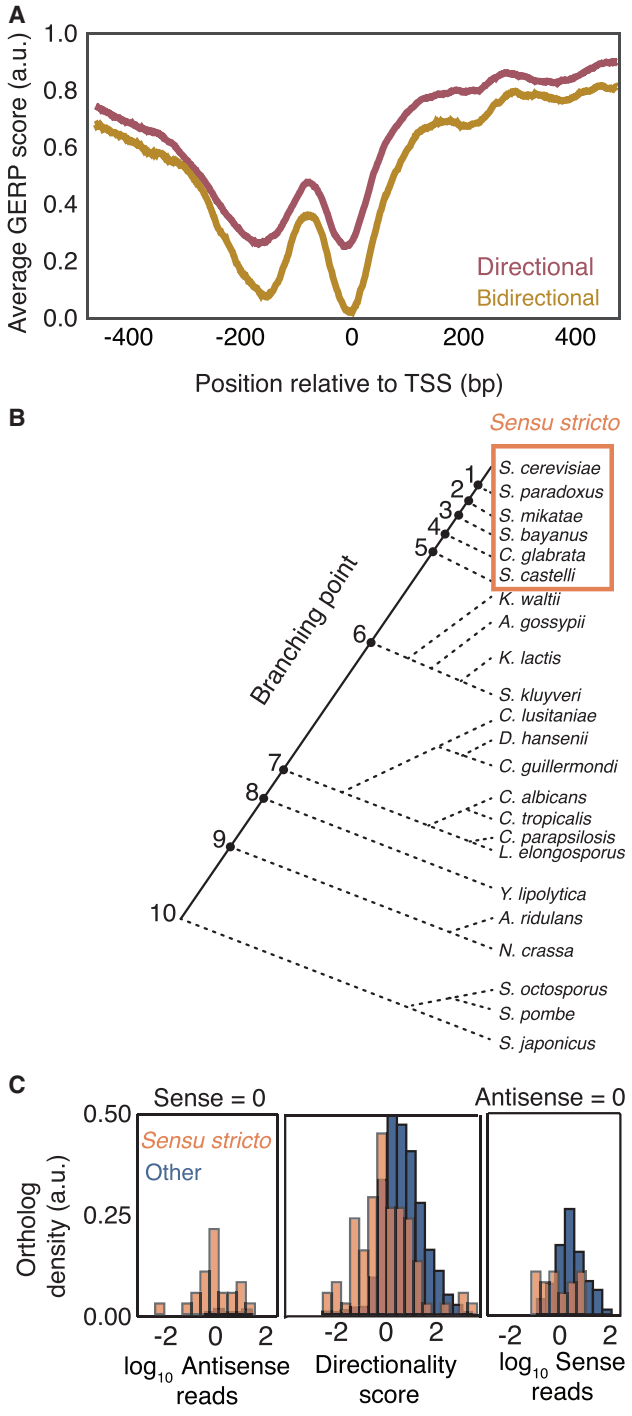

Figure 4. Evolutionary Analysis of Promoter Region Directionality

(A) Aggregate plots of genomic evolutionary rate profiling (GERP) scores, determined by multiple alignments of seven *Saccharomyces* genomes, for directional and bidirectional promoter regions as defined in Figure 1B. The p value is 0.02 calculated by Kolmogorov-Smirnov test for the distributions of average GERP scores over directional and bidirectional promoters, i.e., 500 bp upstream of TSS.

(B) Evolutionary tree displaying the relationship between 23 yeast species. *Saccharomyces sensu stricto* species are boxed.

(C) Directionality distributions for the promoter regions of the genes whose orthologs are present only in *Saccharomyces sensu stricto* (orange) and those whose orthologs are also present in other species (blue). The distributions are significantly different according to the KS test (p value < 3.5×10^{-12}).

See also Figure S4 and Table S2.

fortuitous promoter region. Lastly, when fortuitous promoter regions that arise in coding sequences are retained through evolution, they might split the coding region in half to generate two separate genes. We found 148 possible "gene splitting" events across the *S. cerevisiae* genome by asking whether two tandem genes show strong homology to a single gene in one of the other 22 sequenced yeast genomes (Table S4). Four tandem *S. cerevisiae* gene pairs show high homology to a single *D. hansenii* coding region, and all of these are separated by either a Reb1 or Abf1 binding site, which is not expected by chance (p value = 0.047). This suggests that these gene splitting events may have been formed by the arrival of new transcription factors. Thus, the bidirectional fortuitous promoter regions emerging in YAC coding regions represent a naive state that likely reflects how promoters arise during natural evolution.

Human Transcription Is Bidirectional at Newly Evolved Regulatory Regions

As widespread divergent transcription also occurs in human cells (Core et al., 2008; Preker et al., 2008; Seila et al., 2008), we asked whether the promoter region ground state is conserved to humans. Analysis of NET-seq data from HeLa S3 cells reveals that transcription from human promoter regions is also strongly biased toward the coding direction (Figures 6A and 6B) (Mayer et al., 2015). A comparative epigenomic data of the livers of 20 mammalian species identified a small set of newly evolved liver promoter regions as regions that are functionally active (defined by histone modifications) in human liver and none of the other 19 mammalian livers (Villar et al., 2015). Consistent with our analysis in yeast, we find that newly evolved liver human promoter regions are more bidirectional than highly conserved promoter regions (Figure 6C). The modest effect size is due in part to the tissue-specific nature of the Villar et al. (2015) classification, because some of the promoter regions classified as newly evolved might actually be more highly conserved promoter regions, and simply active in other tissues.

In contrast to human promoter regions, most human enhancers are newly evolved, arising from the exaptation of ancestral DNA (Villar et al., 2015). Thus enhancers could arise in a similar manner as the fortuitous promoter regions in yeast, because in large genomes with low gene density, fortuitous changes in sequences will occur more frequently within intergenic regions, and some of these changes will create new regulatory regions. We postulated that the transcription produced by enhancer regions (eRNAs) would be bidirectional, similar to that of fortuitous promoter regions in yeast. Consistent with what has been described (Andersson et al., 2014), we find that transcription directionality of human enhancers is bidirectional and indistinguishable to the transcription directionality of fortuitous yeast promoter regions (p value of 0.59 using a KS test) (Figure 6B). Thus across yeast and human genomes, newly evolved nucleosome depleted regions produce bidirectional transcription,

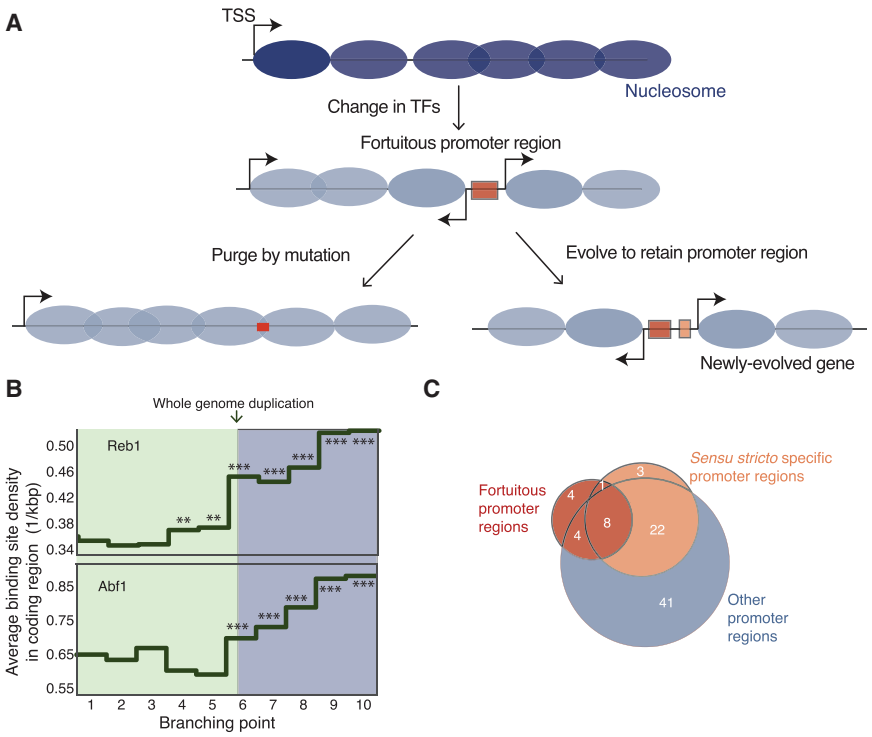

Figure 5. Evolutionary Trajectories of De Novo Promoter Regions

(A) Schematic showing possible paths that a cell can take after an emergence of a fortuitous promoter region. The first option is purging the transcription factor binding site found in the coding sequence by mutation (left). Alternatively, the new transcripts produced by the fortuitous promoter regions could be retained (right).

(B) Coding sequence binding site densities are calculated for Reb1 and Abf1 for the genomes of 23 yeast species, averaged across each clade and plotted against the branching point in the evolutionary tree relative to *S. cerevisiae* (see Figure 4B). Difference between the binding site densities for the genomes of the species at each branch point and densities for the *S. cerevisiae* genome was determined by a two-sample Poisson intensity test (Gu et al., 2008). **p value $< 10^{-4}$, ***p value $\ll 10^{-10}$.

(C) Venn diagram shows the overlap between the transcription factors whose binding sites are enriched at fortuitous promoter regions and endogenous promoter regions of *sensu stricto* specific and other genes in *S. cerevisiae*.

See also Table S4.

indicating that the promoter region ground state is a conserved feature of transcription mechanics.

DISCUSSION

Analysis of transcriptional events that occur in evolutionarily irrelevant DNA make it possible to determine the ground state that reflects basic mechanistic properties of Pol II transcription in vivo. The observation that fortuitous promoter regions (*D. hansenii* coding sequences in *S. cerevisiae* cells) give rise to equal transcription in both directions indicates that bidirectionality is the transcriptional ground state. The ground state is mediated primarily by activator proteins, which generate nucleosome-depleted regions via recruitment of nucleosome remodeling complexes and stimulate transcription in both directions. As a consequence, functional core promoters that support transcription to similar extents invariably occur on both sides of the activator-binding sites within the nucleosome-depleted region. This latter conclusion, though perhaps unexpected, is consistent with the observation that the sequence of the core promoter has little effect when transcriptional activation occurs at low to moderate levels (Iyer and Struhl, 1995). Our results also suggest that, in native organisms, many and perhaps nearly all of the non-coding antisense transcripts from bidirectional promoter regions arise as a mechanistic consequence of Pol II transcription and are evolutionarily irrelevant. Some individual antisense transcripts may have a biological function, but the mere existence of such transcripts is expected and does not imply any functional role. Similarly, the bidirectionality of eRNAs is likely a consequence of the transcriptional ground-state, which may question the direct biological significance of enhancer RNAs in mammalian cells.

It has been suggested that divergent transcription promotes new gene formation, and there are promoter regions that seem to have been derived from enhancers (Engreitz et al., 2016; Wu and Sharp, 2013). Thus, rather than serving immediate functional roles, a subset of eRNAs and antisense RNAs could be acting as an RNA reservoir that can be shaped by evolutionary pressures to serve physiological functions in descendants (Churchman, 2017; Wu and Sharp, 2013).

The fact that promoter regions are intrinsically bidirectional means that directional transcription in native organisms is an evolved trait. As transcriptional activator proteins are generally bidirectional and nucleosome-depleted regions are non-directional, this evolutionary process could occur via DNA sequences (and interacting proteins) on one side of the activator binding sites that increase and/or decrease transcriptional activity in one direction. For example, two core promoter regions that mediate divergent transcription might evolve to differentially respond to the activator protein(s), and such a mechanism occurs in *S. cerevisiae*, because strong activator proteins often require a canonical TATA element for high levels of transcription (Iyer and Struhl, 1995; Struhl, 1986). Alternatively, the binding of asymmetric activators (e.g., monomers) or repressors within a promoter region could also confer higher transcription directionality. Our analysis suggests that this is the dominant mechanism by which directionality is conferred in yeast as we identified a subset of DNA-binding protein motifs that are preferentially enriched at directional promoter regions (Table S2). Other mechanisms of directional transcription could involve binding sites for repressors that block the connection between the activator and the basic Pol II machinery (Brent and Ptashne, 1984), sequences that affect nucleosome stability, or sequences that affect transcriptional

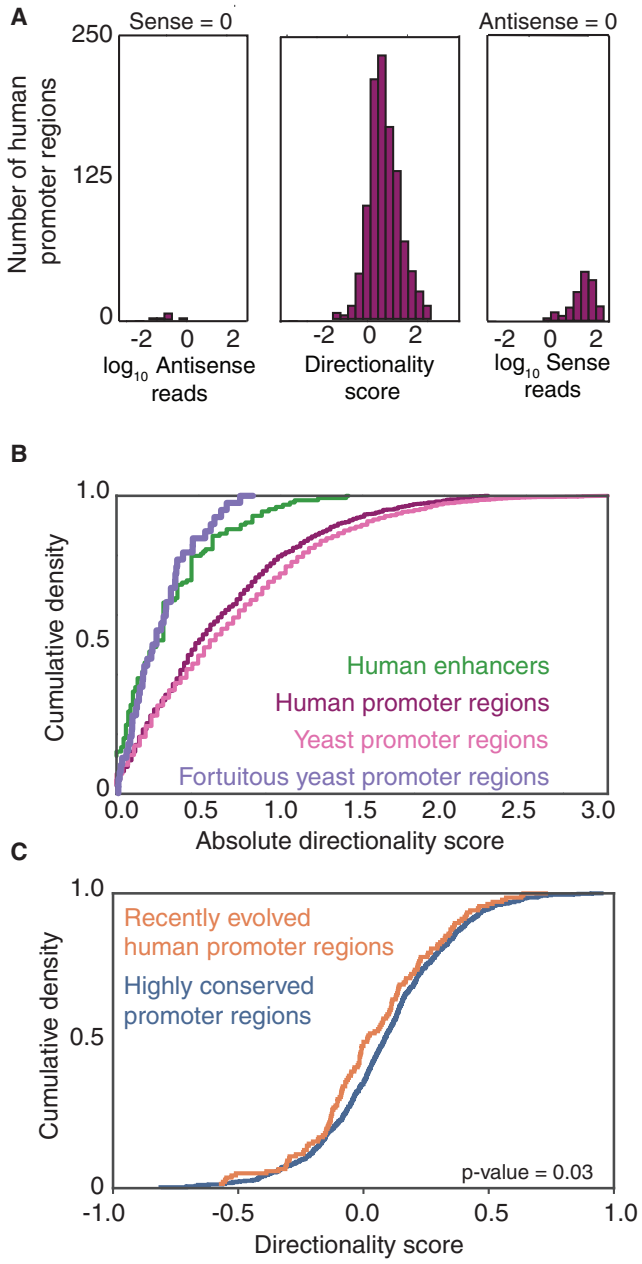

Figure 6. Human Transcription Is More Bidirectional at Newly Evolved Regulatory Regions

(A) Directionality score histogram of human coding sequence promoter regions is shown. The directionality scores of the human promoter regions are calculated the same way as for yeast promoter regions, using NET-seq data from HeLa S3 cells (Mayer et al., 2015), with one alteration. The length of the upstream and downstream windows around the TSS is 1 kb instead of 500 bp, due to the ambiguity of human TSS annotation. Non-overlapping human CDS were curated as described in Mayer et al. (2015).

(B) Absolute values of directionality scores for enhancers (HeLa S3), human coding promoter regions (HeLa S3), and yeast coding promoter regions are plotted as cumulative distribution. Enhancer regions were identified as described in Mayer et al. (2015). Fortuitous promoter regions and enhancers are not statistically significantly different (p value = 0.59 using a KS test).

(C) Directionality score distribution of recently evolved and older promoter regions are shown (p = 0.03 using a KS test). The list of recently evolved and older promoter regions is obtained from Villar et al. (2015).

elongation, reinitiation, or termination. Whatever mechanisms are involved, our results strongly suggest that directional transcription in native organisms involves co-evolution and selection of DNA sequences and transcription factors for some biological function(s).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Strains
  - Growth Conditions
- METHOD DETAILS
  - NET-Seq
  - RNA-Seq
  - ChIP-Seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Processing and Alignment of Sequencing Reads
  - TSS detection for native and YAC species
  - Determining nucleosome depleted regions within the coding sequence of D. hansenii YACs
  - Directionality score calculation
  - Evolutionary rate profiling
  - Transcription factor enrichment
  - Evolutionary retention of fortuitous promoter regions
  - Evolutionary purging of fortuitous promoter regions
  - Directionality change boxplots for individual transcription factors
  - Motif match score
  - Discriminative motif match analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.07.006.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.J., U.E., K.S., and L.S.C.; Methodology, Y.J., U.E., K.S., and L.S.C.; Investigation, Y.J. and U.E.; Software, U.E.; Formal Analysis, Y.J., U.E., K.S., and L.S.C.; Writing – Original Draft, U.E. and L.S.C.; Writing – Review & Editing, Y.J., U.E., K.S., and L.S.C.; Funding Acquisition, K.S. and L.S.C.

ACKNOWLEDGMENTS

We thank Y. Niu, Z. Moqtaderi, A. Mayer, and K. Harlen for technical advice; J. Geisberg for sharing TFIIB antisera; C. Patil for advice and editing; and S. Eddy, C. Weisman, and members of the Churchman and Struhl labs for valuable discussions. This work was supported by grants from the NIH (R01-HG007173 and R01-GM117333 to L.S.C.; R01-GM30186 to K.S.) and a Burroughs Wellcome Fund Career Award at the Scientific Interface (to L.S.C.).

Received: December 3, 2016
Revised: April 6, 2017
Accepted: July 7, 2017
Published: August 10, 2017

REFERENCES

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360–363.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699–709.

Brent, R., and Ptashne, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. Nature *312*, 612–615.

Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E., and Kadonaga, J.T. (1998). The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb. Symp. Quant. Biol. *63*, 75–82.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. Nature *487*, 370–374.

Churchman, L.S. (2017). Not just noise: genomics and genetics bring long noncoding RNAs into focus. Mol. Cell *65*, 1–2.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr. Protoc. Mol. Biol. *Chapter 4*, Unit 4.14.1-17.

Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A.; NISC Comparative Sequencing Program (2005a). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep. *2*, 1025–1035.

de Boer, C.G., van Bakel, H., Tsui, K., Li, J., Morris, Q.D., Nislow, C., Greenblatt, J.F., and Hughes, T.R. (2014). A unified model for yeast transcript definition. Genome Res. *24*, 154–166.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

Eser, U., and Churchman, L.S. (2016). FIDDLE: An integrative deep learning framework for functional genomic data inference. bioRxiv. http://dx.doi.org/10.1101/081380.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3′-end formation. Proc. Natl. Acad. Sci. USA *107*, 17945–17950.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc. Natl. Acad. Sci. USA *108*, 10460–10465.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2011). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res. *39*, 2032–2044.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gu, K., Ng, H.K.T., Tang, M.L., and Schucany, W.R. (2008). Testing the ratio of two poisson rates. Biom. J. *50*, 283–298.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each Phospho-CTD residue. Cell Rep. *15*, 2147–2158.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445–458.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol. Cell *48*, 5–15.

Iyer, V., and Struhl, K. (1995). Mechanism of differential utilization of the his3 TR and TC TATA elements. Mol. Cell. Biol. *15*, 7059–7066.

Kilchert, C., Wittmann, S., and Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. Nat. Rev. Mol. Cell Biol. *17*, 227–239.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lubliner, S., Keren, L., and Segal, E. (2013). Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res. *41*, 5569–5581.

Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., and Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. Genome Res. *25*, 1008–1017.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. eLife *4*, e06722.

Marquardt, S., Escalante-Chong, R., Pho, N., Wang, J., Churchman, L.S., Springer, M., and Buratowski, S. (2014). A chromatin-based mechanism for limiting divergent noncoding transcription. Cell *157*, 1712–1723.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Mayer, A., and Churchman, L.S. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. Nat. Protoc. *11*, 813–833.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

Moura, G.R., Lousado, J.P., Pinheiro, M., Carreto, L., Silva, R.M., Oliveira, J.L., and Santos, M.A.S. (2007). Codon-triplet context unveils unique features of the Candida albicans protein coding genome. BMC Genomics *8*, 444.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol. *20*, 923–928.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851–1854.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. Cell *123*, 233–248.

Rege, M., Subramanian, V., Zhu, C., Hsieh, T.-H.S., Weiner, A., Friedman, N., Clauder-Münster, S., Steinmetz, L.M., Rando, O.J., Boyer, L.A., and Peterson, C.L. (2015). Chromatin dynamics and the RNA exosome function in concert to regulate transcriptional homeostasis. Cell Rep. *13*, 1610–1622.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295–301.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

Scruggs, B.S., Gilchrist, D.A., Nechaev, S., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2015). Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. Mol. Cell *58*, 1101–1112.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849–1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle *8*, 2557–2564.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. Biophys. Acta *1351*, 73–88.

Smale, S.T. (2001). Core promoters: active contributors to combinatorial gene regulation. Genes Dev. *15*, 2503–2508.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu. Rev. Biochem. *72*, 449–479.

Struhl, K. (1986). Constitutive and inducible Saccharomyces cerevisiae promoters: evidence for two distinct molecular mechanisms. Mol. Cell. Biol. *6*, 3847–3853.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell *49*, 295–297.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103–105.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671–675.

Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Gonçalves, J.P., Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P., et al. (2014). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. Nucleic Acids Res. *42*, D161–D166.

Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A., and Rando, O.J. (2010). The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. *8*, e1000414.

Vera, J.M., and Dowell, R.D. (2016). Survey of cryptic unstable transcripts in yeast. BMC Genomics *17*, 305.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54–61.

Weis, L., and Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J. *6*, 3300–3309.

Wheeler, T.J., and Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics *29*, 2487–2489.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031–1035.

Wong, K.H., Jin, Y., and Moqtaderi, Z. (2001). Multiplex illumina sequencing using DNA barcoding. Curr. Protoc. Mol. Biol. *Chapter 7*, Unit 7.11.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990–996.