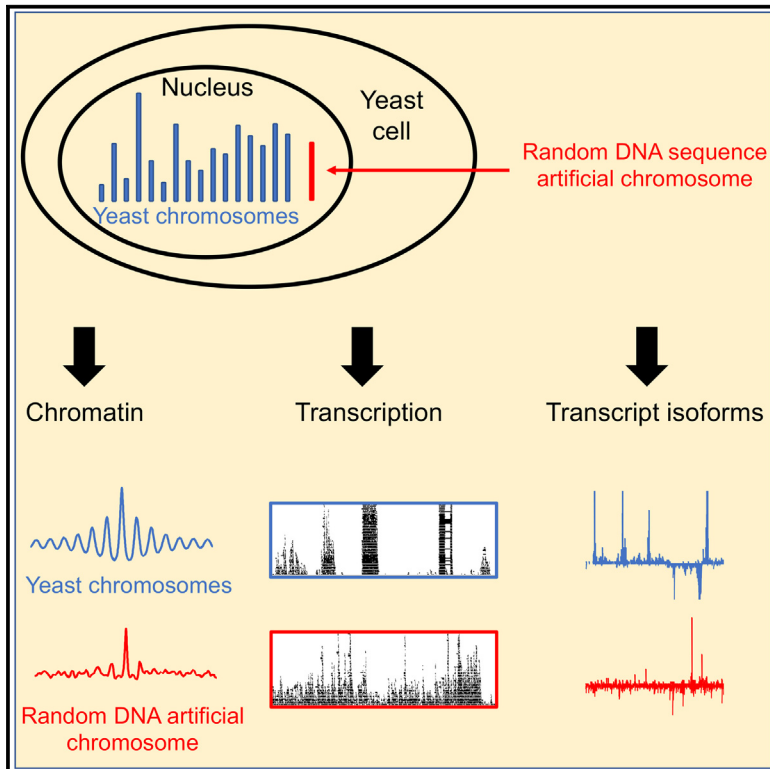


Functional analysis of a random-sequence chromosome reveals a high level and the molecular nature of transcriptional noise in yeast cells

Graphical abstract



Authors

Zlata Gvozdenov, Zeno Barcutean,
Kevin Struhl

Correspondence

kevin@hms.harvard.edu

In brief

“Biological noise” is defined as functionally insignificant events that occur in living cells due to imperfect fidelity of biological processes. By analyzing chromatin structure and transcription of a randomly generated sequence in yeast cells, Gvozdenov et al. show that transcriptional noise occurs at levels comparable to typical genes.

Highlights

- Analyzing chromatin and transcriptional noise using a random-sequence chromosome
- Transcriptional noise in yeast occurs at a level comparable to a typical gene
- Numerous 5' ends indicate a very low intrinsic specificity of transcriptional initiation
- Poly(A) profiles suggest limited evolutionary constraints on 3' end formation



Article

Functional analysis of a random-sequence chromosome reveals a high level and the molecular nature of transcriptional noise in yeast cells

Zlata Gvozdenov,¹ Zeno Barcutean,¹ and Kevin Struhl^{1,2,*}¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA²Lead contact*Correspondence: kevin@hms.harvard.edu<https://doi.org/10.1016/j.molcel.2023.04.010>

SUMMARY

We measure transcriptional noise in yeast by analyzing chromatin structure and transcription of an 18-kb region of DNA whose sequence was randomly generated. Nucleosomes fully occupy random-sequence DNA, but nucleosome-depleted regions (NDRs) are much less frequent, and there are fewer well-positioned nucleosomes and shorter nucleosome arrays. Steady-state levels of random-sequence RNAs are comparable to yeast mRNAs, although transcription and decay rates are higher. Transcriptional initiation from random-sequence DNA occurs at numerous sites, indicating very low intrinsic specificity of the RNA Pol II machinery. In contrast, poly(A) profiles of random-sequence RNAs are roughly comparable to those of yeast mRNAs, suggesting limited evolutionary restraints on poly(A) site choice. Random-sequence RNAs show higher cell-to-cell variability than yeast mRNAs, suggesting that functional elements limit variability. These observations indicate that transcriptional noise occurs at high levels in yeast, and they provide insight into how chromatin and transcription patterns arise from the evolved yeast genome.

INTRODUCTION

Biological function, while often obvious, is an elusive concept. Are all events that occur in living cells biologically meaningful? The flip side of biological function is a concept termed “biological noise,” which is defined by reproducible events (i.e., not experimental errors) that occur in living cells due to imperfection (i.e., lack of fidelity) of a biological process.¹ Biological noise is distinct from stochastic variations that occur in individual cells.

The classic view of transcription by RNA polymerase II (RNA Pol II) is that initiation from the promoter and cleavage/polyadenylation in the 3′ untranslated region (3′ UTR) leads to discrete and functional RNA species, most of which are mRNAs. There is considerable heterogeneity at the 5′ and 3′ ends of these mRNAs, leading to hundreds of different isoforms.² However, there are numerous RNA Pol II-generated transcripts that are not initiated from canonical promoters, do not terminate at the termination regions, and can be transcribed in the opposite direction of classic mRNAs.^{3–5} Many “cryptic” RNAs are unstable in wild-type cells, and they are only observed in strains defective for genes involved in RNA degradation^{6–8} or in maintaining repressive chromatin structure.^{9–13} Are these nonclassical RNAs functional or do they represent biological noise?

Based on the genome-wide distribution of RNA Pol II and quantitative information in yeast cells, it has been estimated that only ~10% of the ~12,000 elongating molecules at any given time are engaged in the generation of conventional

mRNAs.¹ Calculations of RNA Pol II specificity suggest a 10⁴-fold difference between an optimal promoter vs. a random genomic site. Interestingly, this level of specificity is comparable with that of a typical sequence-specific DNA-binding protein and with the frequency at which incorrect nucleotides or amino acids are incorporated into RNA and protein polymers. From these observations, it was proposed that ~90% of elongating RNA Pol II molecules represent transcriptional noise, producing transcription with no biological role.¹ However, transcriptional noise has never been experimentally defined or measured.

RNA Pol II transcription is inextricably intertwined with chromatin structure. As RNA Pol II traverses the gene, there is a rapid equilibrium between histone eviction to facilitate RNA Pol II elongation and histone deposition to restore normal chromatin structure.^{14–16} This dynamic process is mediated by histone chaperones and chromatin-modifying proteins such as FACT, Spt6, Asf1, the Rpd3-S histone deacetylase complex, Set2 histone methylase, and the Swi/Snf nucleosome remodeling complex.^{9–13,17,18} Many of these proteins travel with RNA Pol II, either via direct or indirect interactions or by recognizing the altered structure generated by RNA Pol II transcription. Mutations that inhibit histone deposition result in aberrant open chromatin regions that generate cryptic transcripts that are observed at much lower levels in wild-type strains. Genes that are very highly transcribed have reduced levels of nucleosome occupancy in the coding region.^{14–16,19,20}

Eukaryotic promoters that drive transcription of conventional genes are typically depleted of nucleosomes, thereby permitting increased access to transcription factors.^{19–23} These nucleosome-depleted regions (NDRs) are flanked by highly positioned +1 (downstream) and –1 (upstream) nucleosomes. Nucleosomes in the mRNA coding region are statistically positioned from the +1 nucleosome, with the degree of positioning decreasing progressively at more downstream locations. The position of the +1 nucleosome and the statistically positioned nucleosomes downstream are linked to transcriptional initiation and elongation, respectively.²²

The characteristic chromatin pattern is also observed when NDRs are fortuitously generated in coding regions of foreign yeast DNA sequences present in *Saccharomyces cerevisiae*.²⁴ Moreover, these fortuitous NDRs act as bidirectional promoters that give rise to equal levels of transcription in both directions.²⁵ Similarly, enhancers located far from mRNA coding regions are also associated with NDRs and equal levels of bidirectional transcription.^{25–27} Taken together, these considerations indicate that chromatin structure will influence both the amount and location of transcriptional noise.

The best experimental definition of biological noise involves molecular events that occur on random-sequence DNA. Here, we address transcriptional noise by analyzing an 18-kb segment of DNA whose sequence was randomly generated. We analyze chromatin structure (nucleosome positioning and occupancy by MNase-seq), transcription (by RNA-seq), 3' end formation by cleavage/polyadenylation (3' READS), and 5' end formation (5'-seq). Our results indicate that (1) chromatin forms efficiently on random-sequence DNA but with fewer NDRs and with a lower degree of nucleosome positioning, (2) transcriptional noise occurs at levels comparable with that of typical mRNAs, (3) 5' ends occur with very little specificity, and (4) poly(A) sites of random-sequence RNAs have features in common with natural mRNAs, although they are somewhat more heterogeneous.

RESULTS

Generating a yeast strain with a new chromosome containing random-sequence DNA

Random-sequence DNA is the optimal form of biologically irrelevant DNA, both for measuring transcriptional noise and for identifying sequences important for a biological function in an unbiased manner. We computationally generated an 18-kb stretch of random-sequence DNA. A yeast strain containing this random-sequence DNA was generated by homologous recombination *in vivo* of introduced DNAs encoding the computational sequence.^{28,29} Specifically, overlapping oligonucleotides were co-transformed with a transformation-associated recombination (TAR) vector to assemble a ~27-kb circular chromosome (Data S1), dubbed “ChrXVII,” capable of propagating in yeast cells (Figure S1). For unknown reasons, the resulting strain grows slightly more slowly than the parental strain (155 vs. 140 min doubling time in synthetic medium).

Nucleosomes form efficiently on random-sequence DNA

As chromatin is the physiological template for transcription, the assessment of biological noise requires understanding the chro-

matin structure of the random-sequence chromosome. Poor chromatin assembly on random-sequence DNA would likely lead to artificially high levels of transcription and hence an exaggerated level of transcriptional noise. We investigated the nucleosomal pattern by performing MNase-seq (paired end reads) on isolated mononucleosomes. As a control, we treated naked DNA from the same cells with MNase to generate DNA fragments of a comparable size. For each biological replicate, we obtained the nucleosome:naked DNA ratio of sequence reads for both ChrXVII and yeast genomic DNA. Importantly, this ratio is independent of the copy number of ChrXVII and it is not affected by the sequence-specificity of MNase cleavage. In addition, nucleosome patterns on random-sequence and genomic DNA are analyzed in the same samples.

Biological replicates are highly reproducible, and nucleosome positions in yeast genomic DNA are indistinguishable from those reported in multiple studies (Figure S2).^{15,30} We combined the replicates to generate a total of ~90 million single nucleosomes and ~40 million fragments from the naked DNA controls. Of these, ~300,000 nucleosomes (16.6 alignments and 2,350 hits per bp) and 80,000 control DNA fragments (4.5 alignments and 640 hits per bp) belong to ChrXVII, which is sufficient coverage to study chromatin structure on random-sequence DNA.

As determined by the length of the paired end sequencing reads, nucleosomes on random and genomic DNAs are structurally similar (Figure S3A). In both cases, the major class of nucleosomes average 147 bp (range 143–180), i.e., the expected size (Figures S3A and S3B). There is also a minor class averaging 129 bp (range 120–142) that presumably arises from MNase cleavage roughly two helical turns within the nucleosome itself. These two size classes of MNase-generate fragments, and an even more minor and smaller one (range 80–119 bp), behave similarly with respect to nucleosome positioning (Figures S3C and S3D). With respect to nucleosome occupancy, ChrXVII produces a different MNase-seq profile than its purified DNA control (Figures 1A and S2C). In contrast, MNase patterns of *in vivo* and purified mitochondrial DNA are indistinguishable. This indicates that nucleosome locations on ChrXVII are not random.

Based on the combined replicates, there are 2.24-fold more nucleosomes and 1.57-fold more naked DNA fragments on random-sequence DNA vs. genomic DNA. This 2.24:1.57 ratio suggests that nucleosome occupancy on ChrXVII is 1.42-fold higher than on yeast genomic DNA; values of independent replicates range between 1.33 and 1.51. At first glance, this seems physically impossible because most yeast genomic DNA is associated with nucleosomes. However, in accord with many studies, ~15% of the yeast genome (promoter and terminator regions) is largely devoid of nucleosomes (fewer than 15 midpoints; see below), whereas ChrXVII has no such regions. In addition, ~15% of genomic DNA has low nucleosome occupancy (defined by 15–25 midpoints), whereas only 6% of ChrXVII has such regions. From these observations, we estimate that yeast genomic DNA has a ~20% deficit in nucleosomes, meaning that the observed number of nucleosomes on genomic DNA needs to be multiplied by 1.25 to obtain full genome coverage. When this deficit is accounted for, we estimate that nucleosome occupancy on ChrXVII to be 1.13-fold higher than on yeast genomic DNA. Thus, nucleosomes form on random-sequence

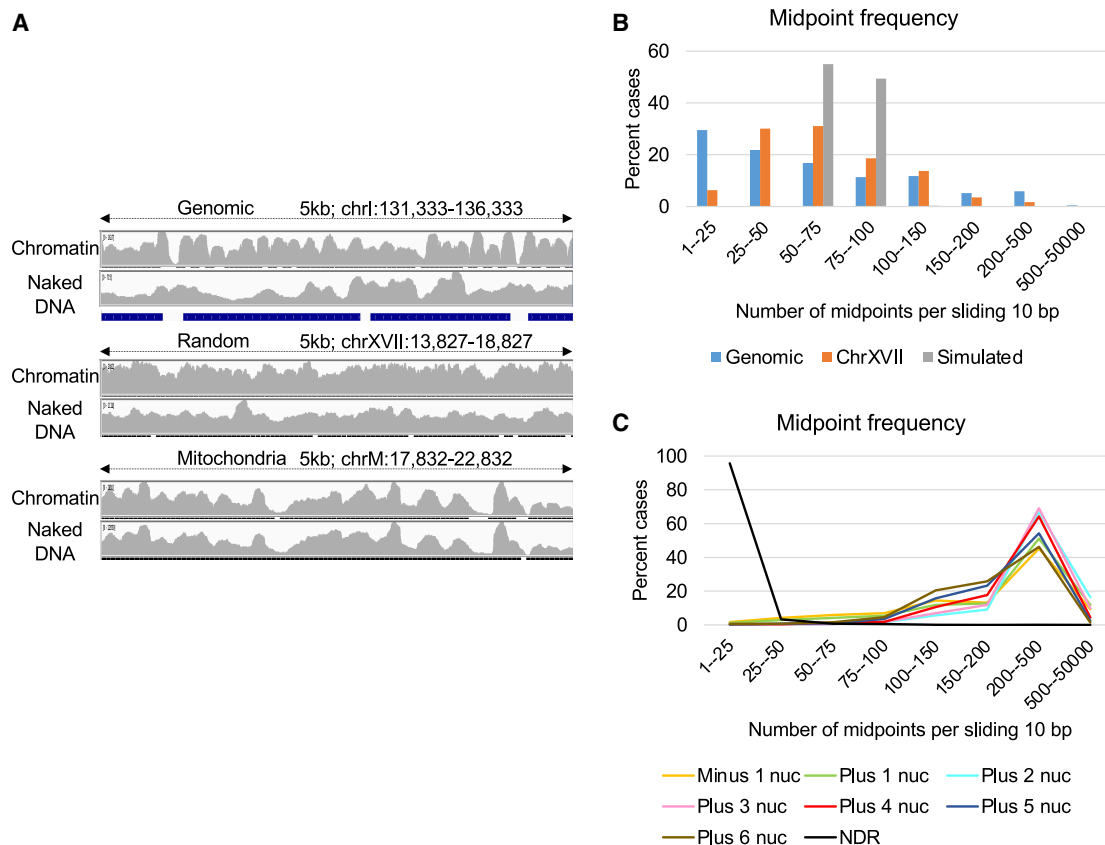


Figure 1. Chromatin structure on random-sequence DNA is similar, but not identical, to yeast genomic chromatin

(A) Browser snapshot representing 5-kb windows of aligned MNase-seq fragments from genomic, random, and mitochondrial DNA, both for chromatin and naked DNA.

(B) Midpoint frequencies per sliding 10-bp window classified into distinct 8 categories for genomic (blue), random (ChrXVII; orange) and randomly simulated data (gray) displayed as percent of the total.

(C) Midpoint frequency representing percent of the midpoints found within 10 bp of nucleosome-depleted regions (NDRs) and the -1 to $+6$ nucleosome centers.

DNA with an efficiency comparable with, and perhaps slightly greater than, that on genomic DNA.

Random-sequence DNA has fewer NDRs and a lower degree of nucleosome positioning and phasing

To investigate nucleosome positioning, we analyzed the frequencies of nucleosome midpoints that are defined by the central position of an MNase-generated sequence. Such nucleosome midpoints depend on the precise location of MNase cleavage within flanking linker sequences, so a nucleosome located at a precise position will inevitably be observed as multiple adjacent midpoints. We therefore consolidated midpoints within a moving average of 10-bp windows. With 90 million fragments per 12 million bp yeast genome and 10-bp sliding windows, 75 nucleosome midpoints are expected to occur by chance, which is seen from the peak of the computationally simulated nucleosomes (Figure 1B).

The occurrence of midpoints per nucleotide position for genomic and random-sequence chromatin was compared with these computationally randomized midpoint locations (Figure 1B). Virtually all computationally simulated nucleosomes

have 50–100 midpoints, which corresponds to random positioning. As expected from the clear pattern of nucleosome positioning in virtually all eukaryotic species, only 25% of nucleosomes in yeast genomic chromatin have 50–100 midpoints. In random-sequence chromatin, about half of the nucleosomes are poorly positioned—50% have 50–100 midpoints—and the other half shows varying levels of positioning.

Yeast chromatin is characterized by nucleosome-depleted promoter and terminator regions as well as by phased and positioned nucleosomes in coding regions. As expected, 30% of nucleotide positions in the yeast genome have fewer than 25 midpoints and hence are nucleosome-depleted to various extents, whereas 6.4% have >200 midpoints indicative of well-positioned nucleosomes (Figure 1B). The highest number of midpoints per position is associated with the $+1$ to $+3$ nucleosomes, and this value decreases gradually for nucleosomes at more downstream positions (Figure 1C).

In contrast, within random-sequence DNA, far fewer (6%) positions have <25 midpoints, indicating a paucity of NDRs (Figure 1B). Furthermore, there are considerably fewer well-positioned nucleosomes—only 1.7% have >200 midpoints—on the

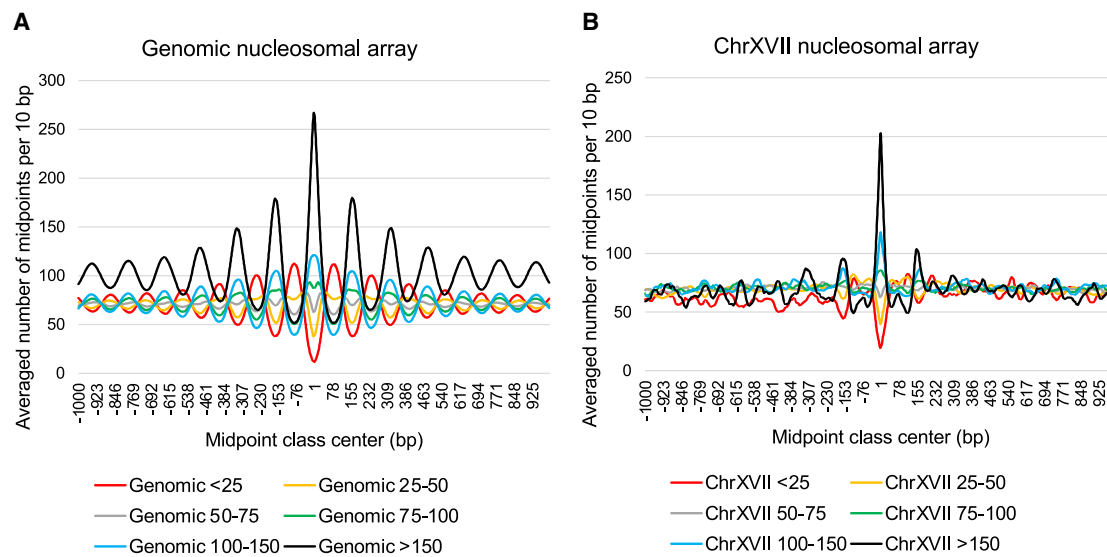


Figure 2. Phased nucleosome arrays form poorly on random-sequence chromatin

Meta-nucleosome position profiles of (A) genomic and (B) random-sequence DNA (ChrXVII). The average number of midpoints per 10 bp that are located 1 kb upstream or downstream of aligned nucleosomes (midpoints defined at coordinate 0) with the indicated number of midpoints (various colors).

random-sequence DNA than on genomic DNA. Thus, the striking chromatin pattern in yeast genomic DNA is considerably muted in random-sequence DNA. However, nucleosome positioning on the random-sequence DNA differs from that of computationally randomized positions (Figure 1B), presumably because of intrinsic DNA sequence preferences and/or spacing constraints imposed by nucleosome remodeling complexes.

To address nucleosome phasing on genomic and random-sequence DNA, we aligned the most highly positioned nucleosomes by their midpoints and analyzed midpoint frequencies of adjacent nucleosomes. As expected, we observe the classic nucleosome-phasing pattern on genomic DNA: i.e., multiple positioned nucleosomes with decreasing levels of positioning at increasing distance from the most positioned nucleosome (Figure 2A). In contrast, on random-sequence DNA, adjacent nucleosomes are very poorly positioned, and indeed midpoint frequencies are essentially flat after the adjacent nucleosomes (Figure 2B). In a similar vein, when regions of low nucleosome occupancy (<25 midpoints) are aligned, regions corresponding to adjacent nucleosomes have very limited positioning. Thus, when compared with yeast genomic chromatin, ChrXVII chromatin is characterized by fewer NDRs, fewer highly positioned nucleosomes, and little nucleosome phasing, although nucleosome positioning is nonrandom.

Levels of RNAs expressed from random-sequence DNA are comparable with those of typical yeast RNAs

We performed RNA-seq to measure steady-state levels of poly(A)-containing RNA in the strain harboring ChrXVII. Transcription occurs readily in both directions throughout the random-sequence DNA with few, if any, regions where RNAs are absent (Figure 3A). In contrast, and as expected, mRNAs transcribed from genomic DNA are organized into distinct regions with directionality (sense gene alignments are blue and

antisense are red). In addition, very little stable RNA is observed between mRNA coding regions.

We quantified RNA levels expressed from random-sequence and genomic DNA in 1-kb windows and plotted the results as \log_2 (nucleotide coverage) in ascending order (Figure 3B). As expected, most RNAs from most genomic DNA are expressed at comparable levels, although 3.6% of the genome is expressed at considerably lower levels (presumably in intergenic or silenced regions), and 7.6% of the genome is expressed at considerably higher levels (primarily in ribosomal and glycolytic genes). Transcription from all regions of random-sequence DNA leads to similar RNA levels that are roughly comparable to that of the majority of mRNA expressed from yeast genomic DNA (Figure 3B). When compared with various functional genomic regions, the amount of transcription of random DNA is slightly lower than that of mRNA coding regions (genes), but it is much higher than that of promoters, terminators, antisense within genes, NDRs, and intergenic regions (Figure 3C; Table S1). Thus, as assayed by steady-state RNA levels, transcriptional noise from random-sequence DNA occurs at roughly comparable levels to those of typical mRNAs, but it is suppressed at other many non-coding regions of the yeast genome.

Transcription from random-sequence DNA is more variable than from genomic DNA

We analyzed biological replicates to estimate the extent of transcriptional variability within a population of cells for genomic and random DNA. Interestingly, the correlation coefficient for genomic-derived RNAs (0.99) is higher than for random-sequence-derived RNAs (0.94). As these correlations involve the identical samples, and RNAs are expressed at roughly comparable levels, this observation suggests that transcription from random-sequence DNA is more variable. To expand upon this observation, we scanned the two replicates for intervals with at

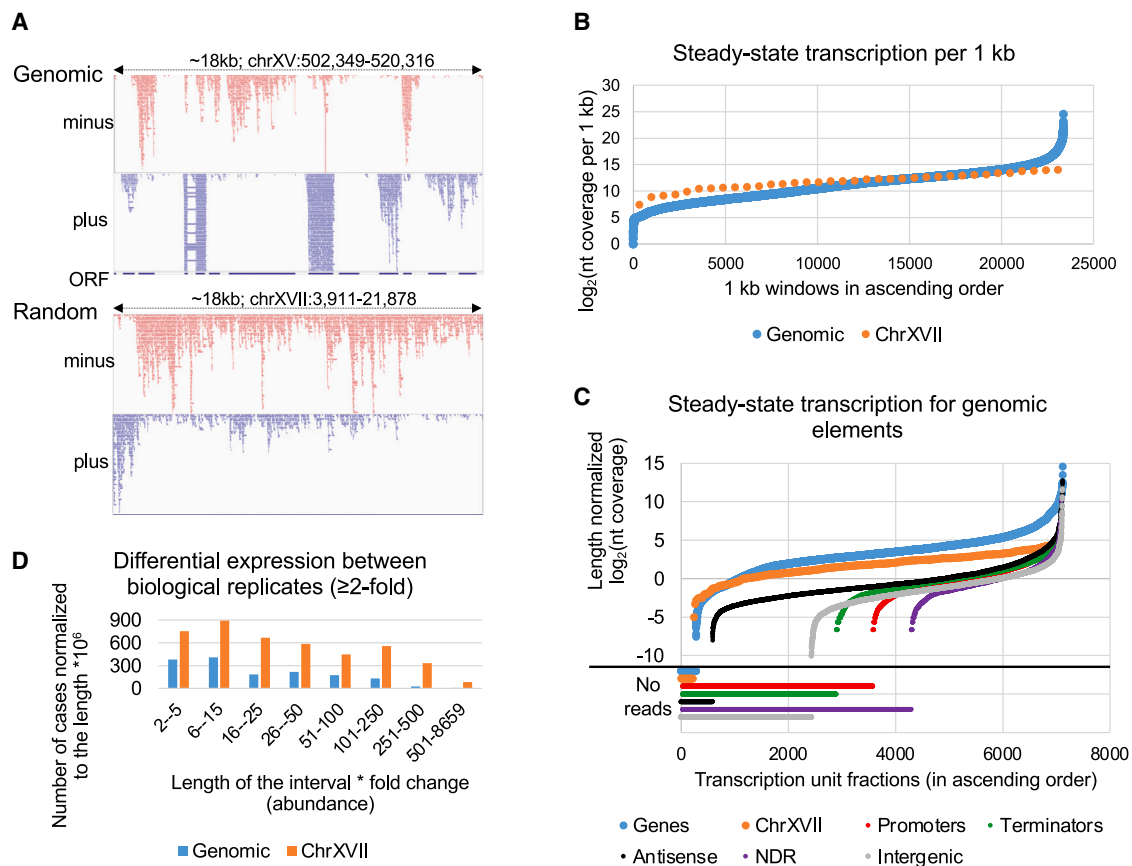


Figure 3. Steady-state levels of random-sequence RNAs are comparable to most yeast mRNAs but show more cell-to-cell variability

(A) Snapshot from the IGV browser showing aligned RNA-seq fragments from genomic and random-sequence DNA, with sense (blue) and antisense (red) transcripts within an 18-kb window.

(B) Steady-state levels of RNA expressed from genomic (blue) and random-sequence (orange) DNA as measured by nt coverage per 1-kb windows sorted in ascending order.

(C) Steady-state RNA levels for ChrXVII (calculated for 100-nt windows) and for indicated (various colors) genomic elements. Nucleotide coverages were normalized to the lengths of all windows before deriving \log_2 values which are plotted in ascending order.

(D) Number of cases with differential (>2 -fold) nt coverage between replicates in the indicated 8 categories defined by the length of the interval multiplied by the fold increase and normalized to the length of genomic and random-sequence DNA.

least a 2-fold difference in nucleotide coverage. The length of the interval with differential coverage for the replicates depends on the continuity of ≥ 2 -fold difference between adjacent nucleotides. The variation between the replicates is expressed as the fold-change multiplied by the length of the interval (Figure 3D). From this analysis, the sum of the differential intervals is 2.8 times higher for ChrXVII-encoded RNAs as opposed to genomic-encoded RNAs. Thus, transcription from random-sequence DNA is more variable than from genomic DNA.

Transcription levels are higher and RNAs are less stable when expressed from random-sequence DNA

Measurements of steady-state RNA levels reflect a balance between synthesis and degradation. We measured transcription rates using 4-thiouracil (4tU)-seq, an approach involving a 5-min pulse of 4tU, resulting in metabolic incorporation into newly synthesized RNA,^{31–33} which is subsequently enriched via activated disulfides conjugated to biotin.³⁴ Levels of newly

synthesized RNA and hence rates of transcription are roughly comparable throughout the random-sequence DNA, but they are 3-fold higher than that of typical yeast mRNAs (Figures 4A and S4). Interestingly, active transcription from random-sequence DNA is much higher than from antisense, NDRs, promoters, terminators, and intergenic regions (Figure 4B; Table S1), suggesting that transcriptional noise from these regions is suppressed.

As observed with steady-state RNA levels, newly synthesized RNA and hence transcription rates within cell populations are more variable on random-sequence DNA than on genomic DNA. On newly synthesized RNA, the nucleotide-level correlation between biological replicates is 0.97 on yeast genomic DNA and only 0.86 on random-sequence DNA. In addition, the sum of differential intervals on random-sequence DNA is 3.6-fold higher than on genomic DNA (Figure 4C). A comparison of populational differences between random-sequence and genomic transcription reveals that actively transcribed RNA

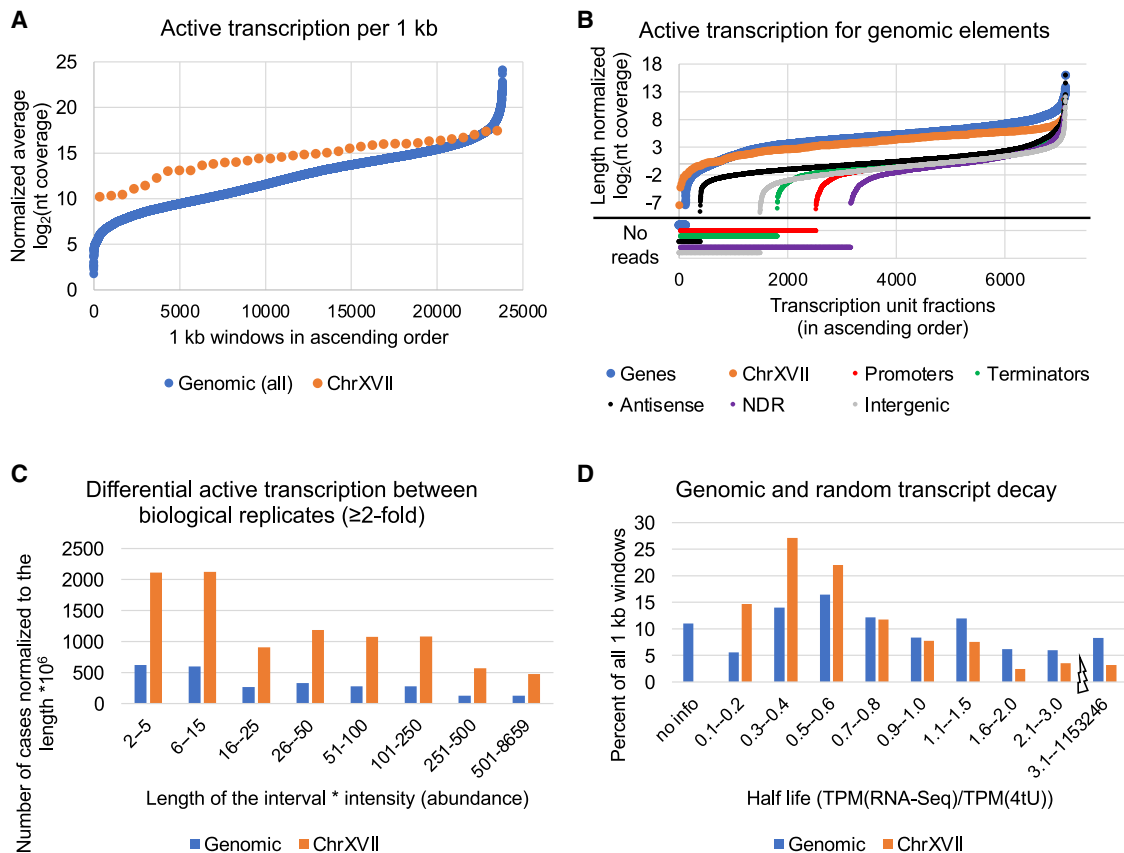


Figure 4. Active transcription and RNA degradation rates are higher for RNAs expressed from random-sequence vs. genomic DNA

(A) The amount of active transcription from genomic (blue) and random-sequence DNA (orange) as measured by background-subtracted 4tU-seq nt coverage per 1-kb windows that are plotted in ascending order.

(B) Levels of active transcription for the indicated genomic elements (normalized to their lengths with ChrXVII split into 100-nt windows).

(C) Numbers of cases with differential (>2 -fold) nt coverage between replicates in the indicated 8 categories defined by the length of the interval multiplied by the fold increase and normalized to the length of genomic and random-sequence DNA.

(D) RNA half-lives of genomic and random-sequence RNAs were calculated as $\text{TPM(RNA-seq)}/\text{TPM(4tU)}$. The numbers of cases of the indicated 9 categories are shown as the percentage of the total. 1 kb windows for which either RNA-seq or 4tU-seq = 0 were excluded (no information).

levels are more variable than steady-state RNA levels (Figures 3D and 4C).

Steady-state RNA levels (M) are determined by the rate of transcription (β) and the rate of RNA decay (α) according to the formula: $M = \beta/\alpha$.^{33,35,36} As the steady-state levels of RNAs generated from random-sequence DNA are roughly comparable with those of the majority of genomic-encoded RNAs (Figure 3B), this means that decay rates of RNAs expressed from random-sequence DNA are ~ 3 -fold faster (Figure 4D). Thus, the level of transcriptional noise is somewhat higher than the level of transcription of typical yeast genes.

Similar, but not identical, polyadenylation patterns of RNAs expressed from random-sequence and genomic DNA

We used 3' READS³⁷ to compare the polyadenylation patterns of RNAs expressed from random-sequence and yeast genomic DNA (Figure 5A). As expected, polyadenylation of yeast mRNAs is almost completely restricted to 3' UTRs, with each gene hav-

ing an average of ~ 50 3' isoforms.³⁸ In contrast, RNAs expressed from ChrXVII are polyadenylated throughout the entire region of random-sequence DNA (Figures 5A, S5A, and S5B; Table S1).

To compare 3' isoform features between RNAs expressed from random-sequence and genomic DNA, we examined the number of reads per isoform (relative isoform levels) and adjacent isoform distances (isoform clusters as compared with random simulation of isoform locations). Both the range of 3' isoform levels (Figure 5B) and cluster formation (Figure 5C) are similar for 3' UTRs of yeast mRNAs and random-sequence RNAs. However, the maximally expressed isoform and maximum isoform intensities within 400-nt windows tend to be slightly lower for RNAs expressed from random-sequence DNA than from genomic DNA (Figures S5C and S5D). Consistent with these observations, the numbers of 3' isoforms per 400-nt window are higher for ChrXVII RNAs than for 3' UTRs (Figure S5E). The distance between maximum isoforms in adjacent, nonoverlapping 200-nt windows is larger in RNAs from ChrXVII

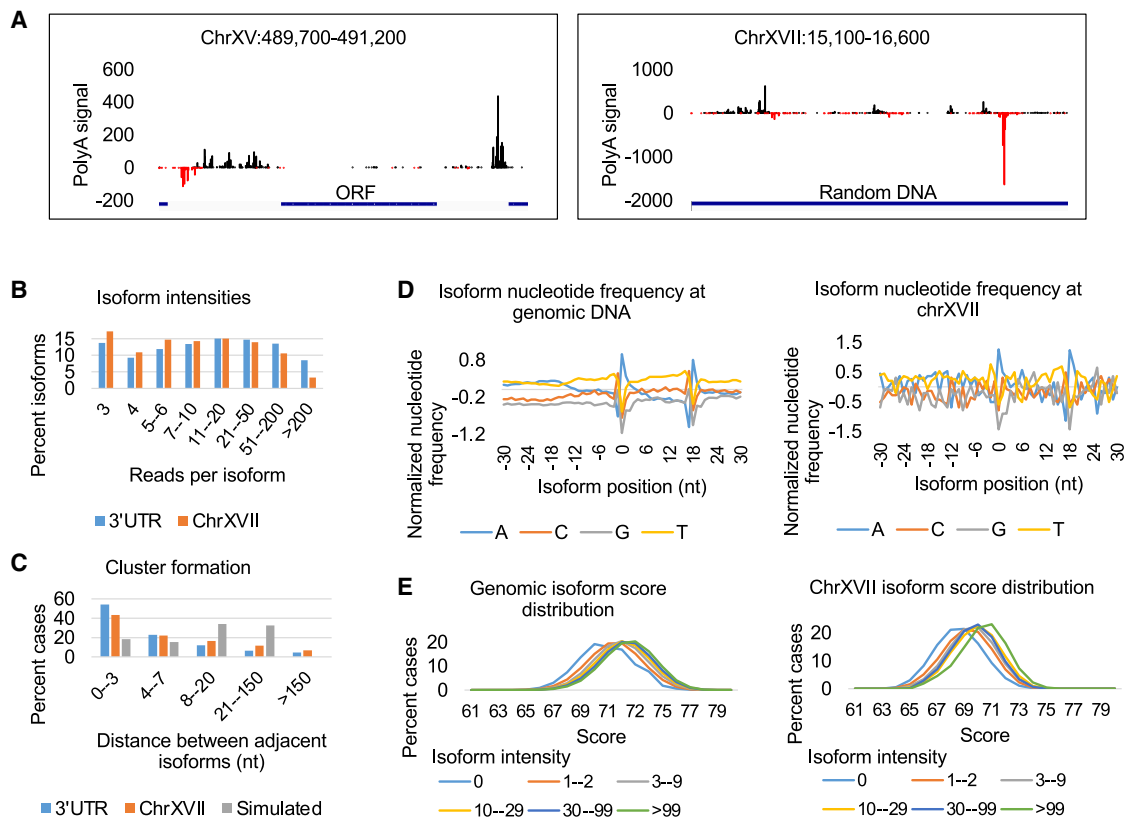


Figure 5. Poly(A) site profiles of random-sequence and yeast mRNAs are roughly similar

(A) Poly(A) reads aligned to representative 1.5 regions of genomic and random-sequence DNA.

(B) Isoform levels (reads per isoform) for poly(A) sites in genomic 3' UTRs (blue) and random-sequence RNAs (orange).

(C) Frequency of distances between adjacent 3' isoforms for 3' UTRs (blue), random-sequence RNAs (orange), and computationally simulated data that randomizes positions (gray).

(D) Nucleotide frequencies at positions -30 to $+30$ for the most highly expressed (top 10%) 3' isoforms from genomic (normalized to nucleotide frequency in the yeast genome) and ChrXVII RNAs.

(E) Isoform score distributions³⁸ of genomic and ChrXVII RNAs in 6 bins (different colors) based on isoform read intensity.

than from genomic DNA (Figure S5F). The distance between the first and last isoforms within a 400-nt sliding window have a propensity to be below 100 nt for 3' UTRs and above 200 nt for random-sequence DNA (Figure S5G). Taken together, ChrXVII and yeast mRNAs have roughly similar polyadenylation patterns, although ChrXVII 3' isoforms tend to be more spread out, slightly more frequent, and relatively less abundant than 3' isoforms in yeast mRNAs.

The nucleotide composition around poly(A) sites of random-sequence RNAs mimic those within the 3' UTRs of yeast mRNAs (Figure 5D): there is a spike of A $+1$ to $+3$ nt with respect to the isoform position (63% nucleotide frequency for genomic and 80% for random); $+18$ to $+20$ nt downstream A peak for genomic (61%) and for random (42%); both $+8$ to $+16$ and -4 to -2 regions have a tendency to be enriched in A residues (38%–34% for genomic and 44%–33% for random); comparable GC nt “rise and fall” profiles for the first -2 to $+1$ nt peak and $+16$ to $+18$ downstream (Figure 5D). Lastly, when compared with the nucleotide sequence matrix of yeast 3' isoforms,³⁸ the most highly expressed ChrXVII isoforms have scores that are

nearly as high as those of yeast 3' isoforms, and the isoform score distributions drop similarly in both random-sequence and genomic RNAs in accord with their level of expression (Figure 5E). Thus, the choice of poly(A) sites by the cleavage/polyadenylation machinery is roughly similar for RNAs expressed from random-sequence and genomic DNAs.

5' isoform profiles from random-sequence DNA show very low specificity, comparable to non-5' UTR regions of the yeast genome

We mapped 5' isoforms with 5'-seq—a procedure we developed by modifying a previously described dephosphorylation-decapping method^{39,40} and obtained the expected metagene profile for endogenous yeast genes (Figure S6A). We note that 5' isoforms are generated during the process of transcriptional initiation, but levels of 5' isoforms are also influenced by stability of the transcripts; 5' isoforms emanating from ChrXVII are scattered at low levels throughout the entire stretch of 18-kb of random-sequence DNA with few, if any, well expressed isoforms (Figure 6A). This pattern is very different from 5' isoforms from

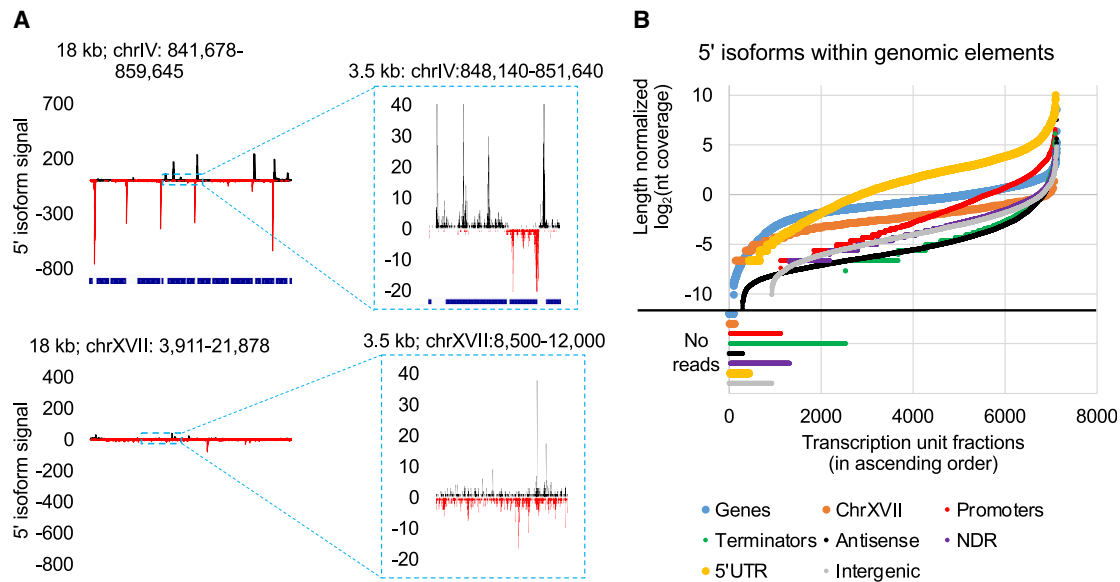


Figure 6. Numerous 5' isoforms expressed from random-sequence DNA and from yeast genomic regions outside 5' UTRs

(A) Aligned 5' isoform reads for genomic and random-sequence RNA within 18-kb (left) and 3.5-kb (right, enlarged by a factor of 20) windows. (B) Number of 5' isoforms for random-sequence RNAs (calculated for 100-nt windows) and for various functional genomic regions. Nucleotide coverages were normalized to the lengths of all windows before deriving \log_2 values and plotting the values in ascending order.

genomic DNA, which are heavily biased to 5' UTRs (Figures 6A, left and 6B). However, increased magnification reveals that random-sequence DNA and genomic regions outside 5' UTRs express many 5' isoforms at low levels (Figure 6A, right). We then measured the numbers of 5' isoforms per various functional genomic elements, including the whole genome split into 200-nt windows. The 5' isoform pattern from random-sequence DNA resembles that observed from yeast genomic regions not corresponding to 5' UTRs, and it is slightly lower than within open reading frames (ORFs) (Figures 6A, 6B, and S6B; Table S1). This suggests that transcriptional noise is initiated with very low specificity and that 5' isoforms initiated from non-5' UTRs within genomic DNA are likely to be transcriptional noise. TSS nucleotide frequencies for ChrXVII 5' isoforms display some similarities to 5' isoforms of yeast mRNAs (Figure S6B). Unlike 3' isoforms, which are typically clustered, the distances between adjacent 5' isoforms for ChrXVII RNAs, are largely expected by chance, indicating weak clustering (Figure S6D).

DISCUSSION

Studying biological noise using random-sequence DNA

Previously, we addressed the issue of biological function by assaying nucleosome positioning,²⁴ transcription,²⁵ and polyadenylation³⁸ in the yeast *S. cerevisiae* on evolutionarily irrelevant DNA from the distant yeast species *Debaryomyces hansenii*. However, while *D. hansenii* DNA is evolutionarily irrelevant in *S. cerevisiae* cells, it is not evolutionarily neutral and hence not suitable for measuring biological noise. Here, we measure biological noise in yeast cells by analyzing chromatin structure and various aspects of transcription of an 18-kb region of DNA whose sequence was randomly generated. Molecular events

on random-sequence DNA are not only functionally irrelevant per se, but we assume that they mimic biological noise that occurs on yeast genomic DNA. Conversely, we presume that events that occur on genomic, but not random-sequence, DNA are biologically meaningful. In addition, this approach complements, but is distinct from, experiments that define sequences required for a specific molecular function via genetic or biochemical selections from libraries of molecules with short stretches of random DNA sequences.

Chromatin structure on random-sequence DNA

Essentially all of ChrXVII is nucleosomal, but chromatin on random-sequence DNA does differ from chromatin of yeast genomic DNA. First, genomic DNA has many more regions of low nucleosome occupancy than random-sequence DNA, providing direct evidence that such regions are subject to evolutionary selection and hence functional. Indeed, these NDRs are almost always in promoters and terminators. Second, chromatin on random-sequence DNA shows a lower degree of nucleosome positioning, but there are some highly positioned nucleosomes, presumably due to DNA sequence specificity for nucleosome formation.⁴¹ Third, nucleosomes on random-sequence DNA are poorly phased, even when flanking a highly positioned nucleosome, arguing against a simple statistical positioning model.⁴²

NDRs of the yeast genome arise from (1) poly(dA:dT) sequences,⁴³ which have intrinsic nucleosome destabilizing properties^{24,44,45} and are preferred substrates of the RSC nucleosome remodeling complex^{46,47}; and (2) binding sites for activator proteins that recruit nucleosome modelers and histone acetylase complexes. NDRs support localized transcriptional initiation that leads to a highly positioned +1 nucleosome followed by phased but progressively less positioned nucleosomes

further downstream.²² This nucleosome-positioning pattern and bidirectional transcription occurs even when NDRs are generated fortuitously on evolutionarily irrelevant DNA.^{24,25} Importantly, the specific locations of functional sequences within the genome result in a nucleosome positioning pattern that is similar (though not identical) among cells in a population.

In contrast, random-sequence DNA lacks functional elements at specific locations, and most nucleosomes are weakly positioned. As such, nucleosome positions on ChrXVII are very likely to differ considerably from cell to cell, except for the small minority of well-positioned nucleosomes. Furthermore, and as discussed below, transcription is initiated at numerous sites throughout ChrXVII, yielding RNAs that differ considerably among cells in the population. Thus, while the classic nucleosome-positioning pattern should occur upon transcription from a given site in an individual cell, the overall cell population will have +1 and subsequent nucleosomes at numerous positions. In addition, highly positioned nucleosomes on ChrXVII reflect intrinsic sequence preferences of histone octamers, which is insufficient for nucleosome phasing. Formation of phased nucleosome arrays requires transcriptional initiation from a defined location²² and transcription-dependent spacing restraints of the Chd1 and Isw1 nucleosome remodelers.⁴⁸

Transcriptional noise in yeast occurs at high levels but is suppressed at some genomic locations

The nucleosomal nature of chromosome XVII is important for interpreting the transcriptional properties of random-sequence DNA and hence measurements of transcriptional noise. As expected for random-sequence DNA, steady-state RNA levels are roughly similar across the 18-kb of ChrXVII. Interestingly, the steady-state levels of RNAs transcribed from random-sequence DNA are comparable with those of most yeast mRNAs. Moreover, levels of newly synthesized random-sequence RNAs (from 4tU-seq) are ~4-fold higher than those of typical mRNAs, indicating a high level of transcriptional noise. The difference between steady-state and newly synthesized RNA levels indicates that RNAs generated from random-sequence DNA are considerably less stable than conventional mRNAs.

The relative instability of random-sequence RNAs is likely due to at least three reasons. First, 3 out of 64 codons are nonsense codons, so translation of random-sequence RNAs will generally yield very short proteins, hence subjecting these RNAs to nonsense-mediated decay.^{49,50} Second, due to the first ATG rule for translational initiation and the fact that only 1 of 64 codons are bound by the methionine initiator tRNA, many 5' UTRs are likely to be long and possibly subject to increased degradation.⁵¹ Third, the Nrd1-Nab3-Sen1 transcription termination pathway degrades short noncoding transcripts.^{52,53} Although we did not measure the lengths of random-sequence RNAs, they are very likely to be short due to their translated products (see above) and very heterogeneous due to the extreme diversity of 5' ends.

The level of newly synthesized RNA from random-sequence DNA is in excellent accord with calculated estimates that only 10%–20% of elongating RNA Pol II molecules generate mRNAs, such the remaining 80%–90% represent transcriptional noise.¹

However, unlike random-sequence RNAs that are synthesized across ChrXVII at roughly comparable levels with those of yeast mRNAs, the synthesis of antisense RNAs and transcripts from NDRs, promoters, terminators, and intergenic regions occurs at much lower levels. In addition, many individual regions of the yeast genome show little if any transcription. These observations strongly suggest that yeast cells have mechanisms to suppress biological noise at many genomic locations, including those (e.g., promoters and terminators) that have biological function in the native organism. One such suppression mechanism might involve preferential degradation of antisense transcripts due to increased numbers of sequence recognition sites for the Nrd1-Nab3-Sen1 termination pathway.⁵⁴

We are unable to distinguish between synthesis of true mRNAs and biological noise on the sense strand within mRNA coding regions. It is likely that some of the measured levels of synthesis within mRNA coding regions are due to internal initiation. This suggestion is consistent with our observation of many 5' ends within coding regions and the existence of many internal transcripts that are only observed in strains defective in mRNA degradation.^{6,7,55} In this regard, we speculate that chromatin structural changes associated with transcription of classically defined mRNAs permits increased levels of internal transcription.

Inherent specificity of transcriptional initiation and polyadenylation

In yeast, 5' and 3' ends are generated by independent molecular processes that are not mechanistically connected.² Profiles of 5' and 3' ends of random-sequence RNAs, respectively, provide information on the inherent specificity of transcriptional initiation and polyadenylation that is unbiased by evolutionary selection. Although 5' ends of RNAs are not randomly distributed across ChrXVII, there are numerous 5' ends and no strongly favored initiation sites. These results indicate that the inherent specificity of transcriptional initiation by the basic RNA Pol II machinery is low. For polyadenylation, only a small subset of potential poly(A) sites is used, and there is a considerable range in the levels of 3' isoforms. This indicates that the inherent specificity of polyadenylation is higher than that of transcriptional initiation, although the large number of distinct poly(A) sites suggests a modest level of specificity.

Evolutionary and mechanistic implications for transcription in yeast cells

Unlike the case for random-sequence DNA, transcriptional initiation and poly(A) sites are, respectively, strongly biased to promoters and 3' UTRs. For transcriptional initiation, this bias reflects poly(dA:dT) sequences and activator-binding sites that lead to NDRs that create windows for the RNA Pol II machinery.^{24,25} Within these open windows, DNA sequences make only a minor contribution to where transcription is initiating. This limited specificity is consistent with previous observations in yeast that (1) TATA element quality is relatively unimportant for low and moderate levels of constitutive transcription typical of yeast genes^{56,57}; and (2) the initiator element makes only a modest contribution to the level of transcription, although it is important for 5' site choice.^{58–60} An important exception to this limited specificity is the importance of a consensus TATA

element for high levels of transcription mediated by activator proteins under appropriate environmental conditions.⁵⁶

For polyadenylation, the mechanism of strong 3' bias for polyadenylation is poorly understood, but it is highly likely to reflect evolutionarily selected sequences in the 3' UTR. In yeast, the most likely candidate is the long AT-rich stretch that immediately follows the ORF.⁶¹ In many other organisms, an AAUAAA motif plays a key role in poly(A) site selection,^{62–64} but such motifs occur throughout the genome, and hence can't be the sole basis of 3' UTR bias and poly(A) site selection. The overall similarity in polyadenylation profiles of random-sequence RNAs and yeast mRNAs suggests that the choice of poly(A) sites primarily reflects intrinsic sequence preferences of the cleavage/polyadenylation machinery, not evolutionarily selected sequences. However, there is some evolutionary selection for poly(A) sites because poly(A) sites in random-sequence RNAs tend to be more spread out, slightly more frequent, and relatively less abundant than 3' isoforms in yeast mRNAs.

Cleavage/polyadenylation of yeast mRNAs does not occur within long protein-coding regions, and it is associated with a large structural motif that includes a long AT-rich stretch.³⁸ As such, the frequency of poly(A) sites in RNAs transcribed from random-sequence DNA is surprising. One possibility is that phased nucleosome arrays on transcribed regions inhibit cleavage/polyadenylation. As random-sequence chromatin lacks such phased arrays, perhaps cleavage/polyadenylation occurs on chromatin with longer linker regions between nucleosomes. Alternatively, but not mutually exclusive, mRNA translation might indirectly affect the poly(A) profile by selective degradation, and in this regard random-sequence RNAs have longer 5' UTRs (due to the low frequency of ATG codons), shorter translation products (due to high frequency of stop codons), and hence longer UTRs. As poly(A) profiles are based on steady-state RNA levels, perhaps the observed 3' isoforms of random-sequence RNAs represent a subset with short 5' and 3' UTRs.

Functional elements constrain cell-to-cell variability

For all RNA analyses described here, correlation coefficients of biological replicates are clearly higher for yeast genomic DNA than for random-sequence DNA. This dichotomy cannot be due to a technical issue because the correlation coefficients are derived from the identical samples. Instead, this observation suggests that RNA profiles of individual cells within a population are more variable on random-sequence DNA than on yeast genomic DNA. We suggest that this reflects the fact that functional elements in promoters and terminator regions create windows for transcriptional initiation and polyadenylation that are the same in every cell. These windows represent functional constraints that limit the variability of RNA profiles among individual cells within a population.

Limitations of the study

Although the 18-kb randomly generated sequence is sufficiently long for the analyses here, it is not long enough to identify motifs for DNA binding and other proteins that occur relatively infrequently. The randomly generated sequence had a GC content of 50%, which is slightly higher than the 40% GC content of yeast genomic DNA. All transcription experiments involved poly(A)-

containing RNA and hence we could not investigate RNA species having short or no poly(A) tails.

STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
 - Assembly of random sequence chromosome
 - Nucleosome mapping by MNase-seq
 - Measuring steady-state RNA levels by RNA-seq
 - Measuring RNA synthesis by 4tU-seq and half-life calculations
 - Mapping polyadenylation sites by 3'READS
 - Mapping 5' ends
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Nucleosome analysis
 - Analysis of steady-state RNAs
 - Measuring RNA synthesis and half-life calculations
 - Polyadenylation site analysis
 - Analysis of 5' ends

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2023.04.010>.

ACKNOWLEDGMENTS

We thank Dan Gibson for kindly providing pCC1BAC-LC yeast vector and Joseph Geisberg and Zarnik Moqtaderi for helpful discussion. Work performed by Z.B., an employee of Google LLC, was done during his personal time. This work was supported by a Postdoctoral fellowship to Z.G. (1F32GM140555) and research grants to K.S. (GM30186 and GM131801) from the National Institutes of Health.

AUTHOR CONTRIBUTIONS

K.S. proposed and supervised the project. Z.G. designed and conducted experimental and theoretical work and analyzed data. Z.G. and Z.B. wrote the python scripts. Z.G. and K.S. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 18, 2022

Revised: February 18, 2023

Accepted: April 10, 2023

Published: May 2, 2023

REFERENCES

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

2. Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127–131.
3. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* 103, 5320–5325.
4. 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.
5. Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037.
6. Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* 23, 841–851.
7. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* 23, 853–864.
8. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737.
9. Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096–1099.
10. Mason, P.B., and Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation *in vivo*. *Mol. Cell Biol.* 23, 8323–8333.
11. Carrozza, M.J., Li, B., Florens, L., Sugauma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., and Workman, J.L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581–592.
12. Keogh, M.C., Kurdistan, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., et al. (2005). Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive rpd3 complex. *Cell* 123, 593–605.
13. Joshi, A.A., and Struhl, K. (2005). Interaction of the Eaf3 chromodomain with methylated histone H3-K36 mediates preferential histone deacetylation at mRNA coding regions. *Mol. Cell* 20, 971–978.
14. Kristjuhan, A., and Svejstrup, J.Q. (2004). Evidence for distinct mechanisms facilitating transcript elongation through chromatin *in vivo*. *EMBO J.* 23, 4243–4252.
15. Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* 39, 1235–1244.
16. Schwabish, M.A., and Struhl, K. (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.* 24, 10111–10117.
17. Schwabish, M.A., and Struhl, K. (2006). Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell* 22, 415–422.
18. Schwabish, M.A., and Struhl, K. (2007). The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation *in vivo*. *Mol. Cell Biol.* 27, 6987–6995.
19. Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic structure. *EMBO J.* 14, 2570–2579.
20. Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* 18, 735–748.
21. Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosome throughout the yeast genome. *Genome Res.* 18, 1073–1083.
22. Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273.
23. Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630.
24. 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.
25. Jin, Y., Eser, U., Struhl, K., and Churchman, L.S. (2017). The ground state and evolution of promoter regions directionality. *Cell* 170, 889–898.e10.
26. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.
27. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* 46, 1311–1320.
28. Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and Hutchison, C.A., 3rd (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. USA* 105, 20404–20409.
29. Kouprina, N., and Larionov, V. (2008). Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* 3, 371–377.
30. Brogaard, K.R., Xi, L., Wang, J.P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. *Nature* 486, 496–501.
31. Barrass, J.D., Reid, J.E., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biol.* 16, 282.
32. Miller, C., Schwab, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay *in yeast*. *Mol. Syst. Biol.* 7, 458.
33. Sun, M., Schwab, B., Pirk, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals *xrn1*-dependent buffering of transcript levels. *Mol. Cell* 52, 52–62.
34. Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol. Cell* 59, 858–866.
35. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
36. Blumberg, A., Zhao, Y., Huang, Y.F., Dukler, N., Rice, E.J., Chivu, A.G., Krumholz, K., Danko, C.G., and Siepel, A. (2021). Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. *BMC Biol.* 19, 30.
37. Jin, Y., Geisberg, J.V., Moqtaderi, Z., Ji, Z., Hoque, M., Tian, B., and Struhl, K. (2015). Mapping 3' mRNA isoforms on a genomic scale. *Curr. Protoc. Mol. Biol.* 110. 4.23.21–4.23.17.
38. Moqtaderi, Z., Geisberg, J.V., Jin, Y., Fan, X., and Struhl, K. (2013). Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. *Proc. Natl. Acad. Sci. USA* 110, 11073–11078.
39. Machida, R.J., and Lin, Y.Y. (2014). Four methods of preparing mRNA 5' end libraries using the Illumina sequencing platform. *PLoS One* 9, e101812.

40. Pelechano, V., Wei, W., and Steinmetz, L.M. (2016). Genome-wide quantification of 5'-phosphorylated mRNA degradation intermediates for analysis of ribosome dynamics. *Nat. Protoc.* *11*, 359–376.
41. Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* *186*, 773–790.
42. Fedor, M.J., Lue, N.F., and Kornberg, R.D. (1988). Statistical positioning of nucleosomes by specific protein binding to an upstream activating sequence in yeast. *J. Mol. Biol.* *204*, 109–127.
43. Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. USA* *82*, 8419–8423.
44. Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.* *16*, 847–852.
45. Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* *19*, 65–71.
46. Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2014). Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. *Genes Dev.* *28*, 2492–2497.
47. Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic nucleosome organization reconstituted with pure proteins. *Cell* *167*, 709–721.e12.
48. Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* *333*, 1758–1760.
49. He, F., and Jacobson, A. (2015). Nonsense-mediated mRNA decay: degradation of defective transcripts Is Only Part of the Story. *Annu. Rev. Genet.* *49*, 339–366.
50. Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* *20*, 406–420.
51. Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* *7*, e32536.
52. Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* *16*, 190–202.
53. Arndt, K.M., and Reines, D. (2015). Termination of transcription of short noncoding RNAs by RNA polymerase II. *Annu. Rev. Biochem.* *84*, 381–404.
54. 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.
55. Rougemaille, M., and Libri, D. (2010). Control of cryptic transcription in eukaryotes. *Adv. Exp. Med. Biol.* *702*, 122–131.
56. Iyer, V., and Struhl, K. (1995). Mechanism of differential utilization of the *his3* T_R and T_C TATA elements. *Mol. Cell. Biol.* *15*, 7059–7066.
57. Struhl, K. (1986). Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* *6*, 3847–3853.
58. Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. *EMBO J.* *4*, 3273–3280.
59. Hahn, S., Hoar, E.T., and Guarente, L. (1985). Each of three "TATA elements" specifies a subset of transcription initiation sites at the *CYC1* promoter of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* *82*, 8562–8566.
60. Nagawa, F., and Fink, G.R. (1985). The relationship between the "TATA" sequence and transcription initiation sites at the *HIS4* gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* *82*, 8557–8561.
61. Lui, K.H., Geisberg, J.V., Moqtaderi, Z., and Struhl, K. (2022). 3' untranslated regions are modular entities that determine polyadenylation profiles. *Mol. Cell. Biol.* *42*, e0024422.
62. Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* *263*, 211–214.
63. Wickens, M., and Stephenson, P. (1984). Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* *226*, 1045–1051.
64. Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* *25*, 1770–1782.
65. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
66. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.
67. Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* *352*, 1225–1228.
68. Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* *11*, 53–55.
69. Kourennaia, O.V., Tsujikawa, L., and deHaseth, P.L. (2005). Mutational analysis of *Escherichia coli* heat shock transcription factor Sigma 32 reveals similarities with Sigma 70 in recognition of the -35 promoter element and differences in promoter DNA melting and -10 recognition. *J. Bacteriol.* *187*, 6762–6769.
70. Wong, K.-H., Jin, Y., and Moqtaderi, Z. (2013). Multiplex Illumina sequencing using DNA barcoding. *Curr. Protoc. Mol. Biol.* *Chapter 7*, unit 7.11.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
DH10B (MAX Efficiency™ DH10B Competent Cells)	Thermo Fisher Scientific	Cat# 18297010
Chemicals, peptides, and recombinant proteins		
Micrococcal nuclease	Worthington Biochemicals	Cat# LS004797
Oligo d(T)25 magnetic beads	NEB	Cat# S1419S
NEBNext® Magnesium RNA Fragmentation Module	NEB	Cat# E6150S
T4 RNA ligase 1	NEB	Cat# M0437M
T4 RNA ligase 2 truncated	NEB	Cat# M0242L
MyOne Streptavidin C1 Dynabeads	Thermo Fisher Scientific	Cat# 65001
ShortCut RNase III	NEB	Cat# M0245S
4-Thiouracil	Sigma Aldrich	SKU: 440736-1G CAS: 591-28-6
Biotium Biotin-XX MTSEA	Biotium	Cat# 90066
MDE (mRNA decapping enzyme)	NEB	Cat# M0608S
Quick CIP	NEB	Cat# M0525S
Critical commercial assays		
Agilent Bioanalyzer High Sensitivity DNA Kit	Agilent	Part# 5067-4626
Deposited data		
Raw and analyzed data	This paper	GEO: GSE216450
Experimental models: Organisms/strains		
<i>S. cerevisiae</i> : Strain background: BY4741	ATCC	ATCC 4040002
<i>S. cerevisiae</i> : Strain background: BY4741 pCC1BAC-LCyeast-20-kb-random-DNA	This paper	N/A
Oligonucleotides		
1000 bp random DNA fragments (GeneArt™ Custom DNA Fragments 751-1000 bp)	Thermo Fisher Scientific	Cat# 815030DE
Recombinant DNA		
pCC1BAC-LCyeast	Dan Gibson (Gibson et al. ²⁸)	N/A
pCC1BAC-LCyeast-20-kb-random-DNA	This paper	N/A
Software and algorithms		
Bowtie2	Langmead and Salzberg ⁶⁵	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
bedtools version 2.24.0	Quinlan and Hall ⁶⁶	https://sourceforge.net/projects/bedtools/
Python version 3.8	Python Software Foundation	https://www.python.org
Python scripts	This paper	https://github.com/zlata-gvozdennov/biological-noise https://doi.org/10.5281/zenodo.7804010
IGV version 2.10	Broad Institute ⁶⁷	https://software.broadinstitute.org/software/igv/
Other		
2100 Bioanalyzer system	Agilent	Model G2939BA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kevin Struhl (kevin@hms.harvard.edu).

Materials availability

Reagents generated in this study are available upon request.

Data and code availability

- All raw and processed sequencing data generated in this paper have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) database under accession number GSE216450.
- Python scripts can be found at <https://github.com/zlata-gvozdencov/biological-noise> (<https://doi.org/10.5281/zenodo.7804010>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments were performed in yeast strain BY4741 containing an artificial circular chromosome with 18 kb of a randomly generated sequence. See [method details](#) for generation and growth of this strain.

METHOD DETAILS

Assembly of random sequence chromosome

Random DNA sequence was generated computationally and split in 1 kb fragments, with 100 bp overlap from each end, and these sequence fragments were synthesized by Thermo Fisher Scientific. The constructs were assembled *in vivo* using transformation associated recombination (TAR) cloning.^{28,29} Briefly, overlapping DNA fragments and cut vector were co-transformed in competent yeast spheroplasts from strain BY4741⁶⁸ to generate a circular chromosome *via* homologous recombination.⁶⁹ Competent yeast spheroplasts were prepared from 400 ml logarithmic phase ($OD_{595} \sim 0.6$) cells.²⁹ The cells were washed with 20 ml of cold 1 M sorbitol and treated with 30 μ g/ml zymolyase 20T (US Biological) and 40 μ l β ME (Sigma) in 20 ml SPE solution (1 M sorbitol, 0.01 M sodium phosphate, 0.01 M Na_2EDTA pH 7.5) by gently nutating for 1 h at 30 °C. Spheroplasts were washed with 50 ml of cold 1 M sorbitol (twice) by gently rocking to resuspend, centrifuged for 10 min at 300 g, and resuspended in 2 ml STC solution (1 M sorbitol, 0.01 M Tris·Cl pH 7.5, 0.01 $CaCl_2$). Spheroplasts were mixed with overlapping DNA fragments (40 ng per 1 kb fragment) and 20 ng *Eco*RI-cut TAR vector (pCC1BAC-LCyeast) and incubated for 10 min at room temperature. 800 μ l PEG 8000 was added to the spheroplasts and the sample was incubated for another 10 min at room temperature. The spheroplasts were centrifuged for 5 min at 300 g, resuspended with 800 μ l SOS solution (1 M sorbitol, 6.5 mM $CaCl_2$, 0.25% yeast extract, 0.5% peptone) and recovered for 40 min at 30 °C. Spheroplasts were mixed with 7 ml melted SORB-TOP-His (1 M sorbitol, 3% bacto agar, complete synthetic medium without histidine) and poured on SORB-His plates (1 M sorbitol, 2% bacto agar, complete synthetic medium without histidine). Colonies were ready for selection after 5 days. The assembled chromosome was verified with pairs of diagnostic primers that detect successful homologous recombination at the junctions ([Figure S1](#)). The final 30 kb circular chromosome was isolated from yeast and electroporated into *E. coli* (strain DH10B, Invitrogen), and re-transformed into BY4741 using the standard lithium acetate procedure.

Nucleosome mapping by MNase-seq

Half of an inoculation loop of BY4741 strain with random sequence (artificial chromosome) colonies were inoculated directly into 500 ml liquid medium (complete synthetic medium without histidine) and grown overnight to logarithmic phase ($OD_{595} \sim 0.55$). The cells were crosslinked with 27 ml 37% formaldehyde by nutating at 100 rpm for 30 min at 30C. The crosslinked samples were neutralized with 30 ml 2.5 M glycine by nutating at 100 rpm for 10 min at 30C, spun down at 4000 rpm for 10 min at 4C, and washed with 50 ml cold water. The crosslinked cells were spheroplasted in 20 ml Tris/sorbitol buffer (50 mM Tris·Cl pH 7.5, 1 M sorbitol) containing 25 mg zymolyase 20T (US Biological) and 14 μ l β ME (Sigma) at 60 rpm for 45 min at 30C. The spheroplasts were spun down at 570 g for 5 min at 4C, supernatant was carefully removed, and the spheroplasts were resuspended with ~ 3 ml cold NP-S buffer (1 M sorbitol, 50 mM NaCl, 10 mM Tris·Cl pH 7.4, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.0075% NP40/IGEPAL, 1 mM β ME, 500 μ M spermidine). 600 μ l (100 ml original culture) was used per one MNase titration (MNase digestion time and MNase concentration). MNase (Worthington Biochemical Corporation) was dissolved in NP buffer (NP-S without sorbitol with 30% glycerol) at 40 U/ μ l. Crosslinked spheroplasts from 100 ml cultures were treated with 1600 U MNase for 1-1.5 h at 37C. Reaction was stopped with 100 μ l stop buffer (10% SDS, 0.5 M Na_2EDTA pH 8), reverse crosslinked with 30 μ l Proteinase K (20 mg/ml stock) at 65C overnight. DNA was purified with phenol-chloroform extraction and ethanol precipitation and mononucleosomal-sized DNA was prepared following gel electrophoresis. As a control, genomic DNA from 100 ml culture was treated with ~ 20 -200 nU MNase for 20 min at 37C and mono-nucleosome-sized fragments were isolated. Library preparation was performed with minor modifications of a procedure described previously.⁷⁰ Starting with 1 μ g DNA, fragment ends were repaired and 5' phosphorylated in 50 μ l reaction volume in 1x T4 DNA Ligase buffer (NEB) with 2 μ l 10 mM dNTP mix, 1 μ l T4 DNA Polymerase (3 U/ μ l, NEB), 1 μ l T4 PNK (10 U/ μ l, NEB), and 0.2 μ l DNA pol I, large Klenow (5 U/ μ l, NEB) incubating for 45-50 min at room temperature. DNA was phenol-chloroform extracted and ethanol precipitated. 3' overhangs were added in 30 μ l reaction volume in 1x NEB2 with 2 μ l 10 mM dATP and 1 μ l Klenow

exo- (5 U/ μ l, NEB) for 2 h at 37°C. DNA was phenol-chloroform extracted and ethanol precipitated. Illumina combinatorial dual indexing strategy (with T overhang at 3' of the 5' adapters) was used to ligate annealed oligos to DNA fragments in 20 μ l reaction volume with 2 μ l of 5 mM annealed index combination and 2.5 μ l T4 DNA Ligase (2,000 U/ μ l, NEB). Library was purified with SizeSelector-I beads (Aline Biosciences) as described.³⁷ Amplified libraries (12 PCR cycles using 5% of the ligated material as a PCR template) were purified with SizeSelector-I beads again. Purified barcoded libraries were then examined and quantified via Bioanalyzer (Agilent Technologies), pooled, and further gel purified by 8% PAGE (fragment sizes with ~225-335 bp were excised). Samples were sequenced with Illumina NextSeq High for 75 cycles (paired end reads).

Measuring steady-state RNA levels by RNA-seq

The yeast strain containing ChrXVII was grown as described under MNase-Seq. Total RNA was obtained by standard hot-phenol extraction and DNase treated according to the manufacturer's instructions (Turbo DNase by Thermo Fisher Scientific). Total RNAs from two biological replicates were subjected to poly(A) selection. 25 μ g total, denatured RNA (5 min at 65°C and briefly cooled on ice) were incubated with oligo d(T)25 magnetic beads (NEB) in 400 μ l annealing buffer (10 mM Tris·Cl pH 7.5, 60 mM NaCl, 1 mM Na₂EDTA pH 8) at 500-550 rpm for 30 min at room temperature. Messenger RNA annealed to the magnetic beads was washed with the annealing buffer twice, and then eluted from the beads with 100 μ l TE buffer by incubation for 3 min at 55°C. Denaturation of eluted RNA, incubation with oligo d(T)25 magnetic beads, and two annealing buffer washes were repeated. Poly(A) RNA was finally eluted with 18 μ l water for 3 min at 55°C. 2 μ l of fragmentation buffer from NEBNext® Magnesium RNA Fragmentation Module (NEB) were added to the purified mRNA, which was fragmented for 3 min at 94 °C. The fragmentation was stopped with 2 μ l fragmentation stop buffer (NEB) and mRNA fragments were ethanol precipitated. To remove 3' phosphoryl groups and to phosphorylate 5' ends, mRNA fragments were treated with T4 PNK (NEB). Libraries were prepared as described previously.³⁷ 1 μ l of 1 μ M pre-adenylated 3' adapter with 5' chain terminator was ligated to mRNA fragments with 1 μ l T4 RNA Ligase 2 truncated (200 U/ μ l, NEB) in 15 μ l 1x T4 RNA Ligase buffer (NEB) containing RNase inhibitor and 15% PEG8000 for 2 h at 25°C. The reaction was heat inactivated (2 min at 70°C) and cooled on ice. 1 μ l of denatured 3' adapter³⁷ was added to the reaction together with 1.4 μ l 10 mM ATP and 1 μ l T4 RNA Ligase 1 (10 U/ μ l, NEB), and the reaction was incubated for 2 h at 25°C. The heat-inactivated reaction was used for reverse transcription and the library was amplified for 17 cycles. SizeSelector-I beads purified, barcoded libraries were examined and quantified via Bioanalyzer (Agilent Technologies), pooled, and further gel purified by 8% PAGE (fragment sizes with ~175-500 bp were excised). Libraries were subjected to the paired-end, strand specific high-throughput sequencing with Illumina NextSeq Mid for 150 cycles (75 nt read length) equally apportioned between R1 and R2.

Measuring RNA synthesis by 4tU-seq and half-life calculations

The yeast strain containing ChrXVII was grown as described under MNase-Seq. At the log phase, the cells were treated with either 4-thiouracil (4tU, Sigma Aldrich; 5 mM final concentration) or DMSO as a control for 5 min. Spike-in control cells, untreated *S. pombe* cells and *K. lactis* cells treated with 4tU for 5 min, were added to the sample before RNA extraction. Poly(A) RNA enrichment, fragmentation, and RNA fragment ends repair (T4 PNK treatment) were performed as described under RNA-Seq. RNA was then biotinylated with activated disulfide methane thiosulfonate (MTS) biotin (Biotium Biotin-XX MTSEA) as previously described.³⁴ Briefly, 38.9 μ l fragmented poly(A) RNA (derived from 25-100 μ g total RNA), 1 μ l 1 M HEPES pH 7.4, and 0.1 μ l 0.5 M Na₂EDTA pH 8 were incubated with 10 μ l 82 μ M freshly made (dissolved in dimethylformamide) MTSEA-biotin-XX (16.4 μ M final) for 30 min at room temperature with rotation (in the dark). Unreacted MTSEA-biotin-XX was removed from the RNA sample with acid phenol-chloroform extraction and RNA was recovered with ethanol precipitation. Glycogen-blocked MyOne Streptavidin C1 Dynabeads (Invitrogen) were incubated with biotinylated RNA by rotating the sample in 150 μ l binding buffer (10 mM Tris·Cl pH 7.4, 1 mM Na₂EDTA pH 8, 100 mM NaCl, 0.005% Tween-20) for 20 min at room temperature (light protected) and washed 4 times with high salt wash buffer (100 mM Tris·Cl pH 7.4, 10 mM Na₂EDTA pH 8, 1 M NaCl, 0.05% Tween-20). 4tU-containing RNA was eluted from the beads with 25 μ l elution buffer (100 mM DTT, 20 mM HEPES pH 7.4, 1 mM Na₂EDTA, 100 mM NaCl, 0.05% Tween-20) by rotating the sample for 15 min at room temperature twice, and ethanol precipitated. Adapter ligations to 4tU-enriched RNA, reverse transcription, library amplification, library purification, sequencing, and data analysis were performed as described under RNA-Seq.

Mapping polyadenylation sites by 3' READS

The yeast strain containing ChrXVII was grown as described under MNase-Seq. 3'READS was performed as described previously.³⁷ Prior to adapter ligation, which was performed as under RNA-Seq, mRNA was subjected to the series of poly(A) enrichments and fragmentations. 25 μ g denatured, total RNA (DNase treated) were incubated with oligo d(T)25 magnetic beads once, as described under RNA-Seq. Beads with annealed poly(A) RNA were resuspended in 50 μ l RNase III digestion buffer (10 mM Tris·Cl pH 8.3, 60 mM NaCl, 10 mM MgCl₂, 1 mM DTT) and digested with 1.25 μ l RNase III (2 U/ μ l, NEB) for 15 min at 37°C to leave short 3' fragments upstream of the poly(A) tails. Poly(A) fragments were eluted with 200 μ l TE for 3 min at 55°C, extracted with acid phenol-chloroform, and ethanol precipitated. Heat-denatured poly(A) fragments were enriched with Biotin-T45U5-coupled MyOne Streptavidin C1 Dynabeads (Invitrogen) in annealing-hybridization buffer (20 mM Tris·Cl pH 7.5, 300 mM NaCl, 2 mM Na₂EDTA pH 8, 0.05% Tween-20) at 450-500 rpm for 30 min at 25°C and washed three times with stringent wash buffer (10 mM Tris·Cl pH 7.5, 1 mM NaCl, 1 mM Na₂EDTA pH 8, 0.05% Tween-20). Fragments were digested on the beads with 1 μ l RNase H (5 U/ μ l, NEB) in RNase H reaction buffer (50 mM Tris·Cl pH 8.3, 1 mM NaCl, 3 mM MgCl₂, 10 mM DTT) for 30 min at 37°C, eluted with 200 μ l elution buffer (1 mM NaCl, 1 mM

Na₂EDTA pH 8, 0.05% Tween-20), and acid phenol-chloroform extracted. Samples were paired-end sequenced using Illumina NextSeq Mid for 150 cycles (75 read length).

Mapping 5' ends

The yeast strain containing ChrXVII was grown as described under MNase-Seq. 50–100 μg poly(A)-containing RNA (DNase I treated) was fragmented with NEBNext® Magnesium RNA Fragmentation Module, as described under RNA-Seq, but scaled appropriately (80 μl fragmentation reaction per 100 μg starting total RNA for 4–5 min). Ethanol precipitated and denatured RNA fragments were dephosphorylated with 2 μl Quick CIP (5 U/μl, NEB) in 15.5 μl reaction volume containing 1.5 μl T4 RNA Ligase buffer (NEB) and 2 μl RNase inhibitor (20 U/μl) for 30 min at 37°C. To the heat-inactivated reaction, 2 μl 5 μM of denatured, pre-adenylated 3' adapter,³⁷ 1.5 μl T4 RNA Ligase buffer (NEB), 9 μl 50% PEG8000, and 2 μl truncated T4 RNA Ligase 2 (200 U/μl, NEB) were added and the sample was incubated for 2 h at 25°C. Heat-inactivated, resulting products were treated or not treated with mRNA de-capping enzyme, MDE (100 U/μl, NEB), by splitting the reaction into two tubes, adding 2.5 μl MDE or water to the tubes, and incubating for 1 h at 37°C. 1 μl of 25 μM denatured 5' adapter,³⁷ 0.5 μl 100 mM ATP, and 1 μl T4 RNA Ligase 1 (30 U/μl, NEB) were added to the previously heat-inactivated samples and incubated for 2 h at 25°C. The samples were reverse transcribed, amplified, and purified as described under RNA-Seq. Both MDE treated and control sample were paired-end sequenced with NextSeq Mid and NextSeq High.

QUANTIFICATION AND STATISTICAL ANALYSIS

Nucleosome analysis

We combined seven different samples representing four biological replicates that had one or two MNase digestion time points (with a ~30 min difference). The reads were aligned to the reference genome, sacCer3 (UCSC, 2011) containing chrXVII (pCC1BAC-LCyeast-20-kb-random-DNA) using bowtie2⁶⁵ with the options: `-no-mixed -no-discordant -no-dovetail -no-contain -no-overlap`, and bam files were converted to bed using bedtools bamtobed (v.2.24.0).⁶⁶ These bed files were used to compute fragment locations for correctly aligned paired-end reads (fragment_location.py), fragment size distribution (fragment_distribution.py) and size separation (fragment_location_size_separation.py), fragment midpoints locations (num_of_mid_point.py), midpoint frequency distributions (num_of_mid_point.py), midpoint background subtractions (middle_point_bg_subtraction.py), moving averages (num_of_mid_point_mov_ave.py), MNase cutting frequencies (cutting.py), deriving +2 to +6 nucleosomes (max_midpoint_count.py, find_which_gene.py), meta-analysis of phased nucleosomal arrays (averaging_mp_aligned_fragments.py), and nucleosomal/NDR nucleotide scores (matrices_generator_mono.py). MNase cutting frequency was calculated for genomic and random-sequence chromatin and naked DNA by counting per nt MNase-Seq fragment ends and normalizing to the total genomic or ChrXVII values. Per nucleotide coverages from the fragment locations were derived with fragments_to_single_pos.py, and counts per defined intervals with count_per_defined_interval.py.

Analysis of steady-state RNAs

RNA-seq reads (10.4 and 8.1 million reads for the two replicates) were mapped to the *S. cerevisiae* genome and the introduced random sequence using bowtie2 `-5 25 -3 25 -no-mixed -no-discordant -dovetail`. Bam files were converted to bed using bedtools bamtobed (v.2.24.0).⁶⁶ Fragment locations were derived with fragment_location_directionality.py, per nucleotide coverages from the fragment locations with fragments_to_single_pos.py, counts per defined intervals with count_per_defined_interval.py (minus reads align to plus strand and vice versa, plus reads align to minus strand), and differentially expressed intervals with interval_identifier.py.

Measuring RNA synthesis and half-life calculations

From five biological replicates we obtained ~20 million paired-end sequencing reads (excluding spike-ins). Duplicate reads were removed with remove_duplicate_reads_bothfastq.py. Spike-in normalized nt coverages were used to subtract background from the 4tU sample to obtain active transcription measurements (4tu_dmso_normalization.py). Half-life calculations were performed from the combined RNA-Seq and 4tU-Seq values (Schwalb et al., 2016).³⁶ Briefly, unitless constant of RNA half-life was estimated from the number of transcripts (RNA-Seq) divided by the production of new mRNA (4tU-Seq), with RNA-Seq and 4tU-Seq expressed as transcripts per million (TPMs) (tpm.py):

$$\text{TPM} = \frac{\text{nt coverage}}{\text{TU length}} \cdot \frac{10^6}{\sum \frac{\text{nt coverage}}{\text{TU length}}}$$

Polyadenylation site analysis

Sequence reads from the 3'READS data were aligned with bowtie2 `-x sacCer3random_spmbe -5 25 -3 25 -no-mixed -no-discordant -dovetail -X 1000`. Duplicates based on both read pairs were removed with remove_duplicate_reads_bothfastq.py. This procedure generates new fastq files, with filtered, non-duplicate reads from which Ts (3' As) were shortened. These reads were aligned with bowtie `-p24 -n0 -l17 -m1`. Only reads with polyTs, i.e., reads that come from the truthful polyadenylation were considered. The reads ending with T, such that it is not clear whether the T comes from the genomes or is a part of the polyadenylation process, were

excluded. 3' isoforms were further derived with `3reads_locations.py` and `num_of_mid_point.py` (for each strand separately). We obtained ~18 million paired-end sequence reads from 6 biological replicates. The number of isoform reads for each isoform was normalized to the transcript coverage per 50 nt window (± 25 nt from the isoform location). Using sliding windows, maximum isoforms were identified and following associated features were quantified: percent signal at the most used (maximum) coordinate; maximum isoform intensities; distances between maximum isoforms; distance between the first and the last isoform; and isoform frequencies (number of distinct coordinates) (`3prime_isof_sliding_window_stat.py`, `major_coord_distances.py`). Alone standing isoforms were excluded from the sliding windows. *De novo* isoform-specific nucleotide frequencies were calculated (`matrices_generator_mono.py`) and isoform scores were based on isoform frequency matrices³⁸ (`predicting3prime.py`).

Analysis of 5' ends

Sequence reads from the 5' end analysis were aligned with following command: `bowtie2 -x sacCer3random_spombe -5 4 -3 16 -no-mixed -no-discordant -dovetail -X 1000 (-5 4 -3 16 was used for 37 nt read length, -5 4 -3 55 was used for 76 nt long reads)`. From 5 replicates, we obtained 32.7 million sequence reads for the de-capped sample and 25.9 million reads for the control sample (excluding duplicate reads, which were removed with `remove_duplicate_reads_bothfastq.py`). Bowtie-generated bam files were converted to bed using `bedtools bamtobed (v.2.24.0)`.⁶⁶ The 5' isoforms were determined by the first position past the 5' adapter after which we computed the 5' reads per genomic coordinate (`fragment_location_directionality.py`, `select_plus_5isoform.py`, `select_minus_5isoform.py`, `num_of_mid_point.py`). Negative control values per genomic coordinate were normalized using the internal *S. pombe* control and subtracted from the values of the de-capped sample (`5isoforms_MDE_minus_normalization.py`).

Molecular Cell, Volume 83

Supplemental information

**Functional analysis of a random-sequence
chromosome reveals a high level and the molecular
nature of transcriptional noise in yeast cells**

Zlata Gvozdenov, Zeno Barcutean, and Kevin Struhl

Figure S1

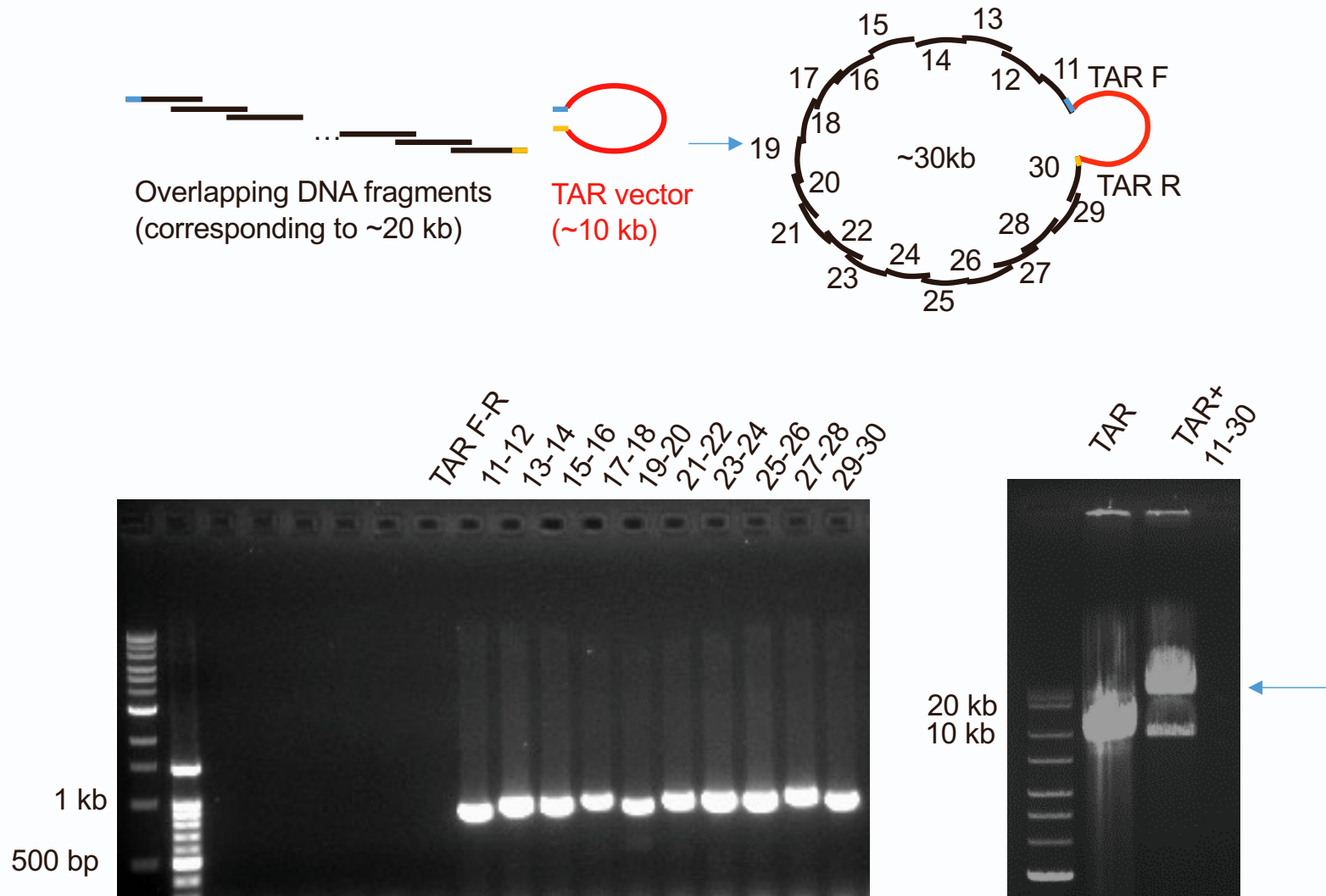
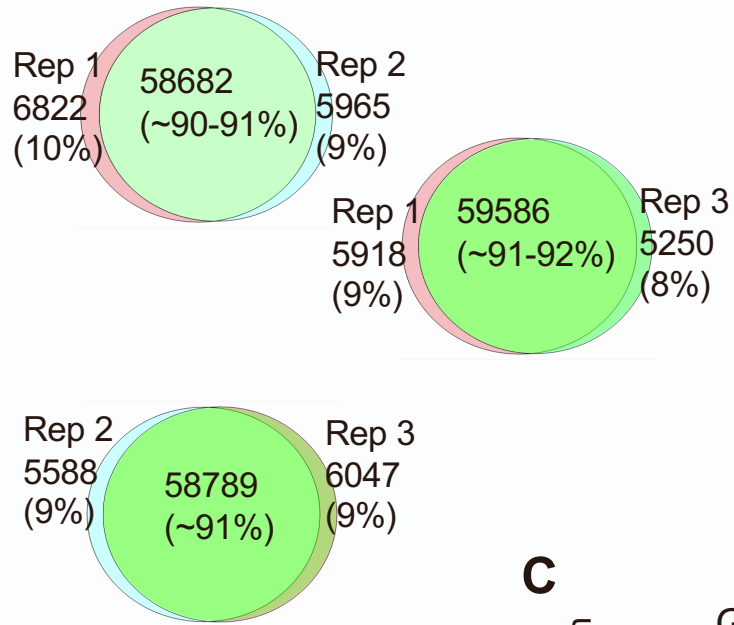


Figure S1. Construction of the yeast strain harboring a random-sequence chromosome.

Related to Figure 1. Overlapping DNA fragments, each containing a predefined, computationally randomized sequence, and the TAR vector were introduced into yeast cells with chromosome assembly occurring *in vivo* via homologous recombination (top panel). The assembled chromosome was verified with PCR amplification of the indicated fragment junctions (bottom left panel). For overall size verification, the final construct was excised from the TAR vector with the restriction sites at random-sequence ends (bottom right panel).

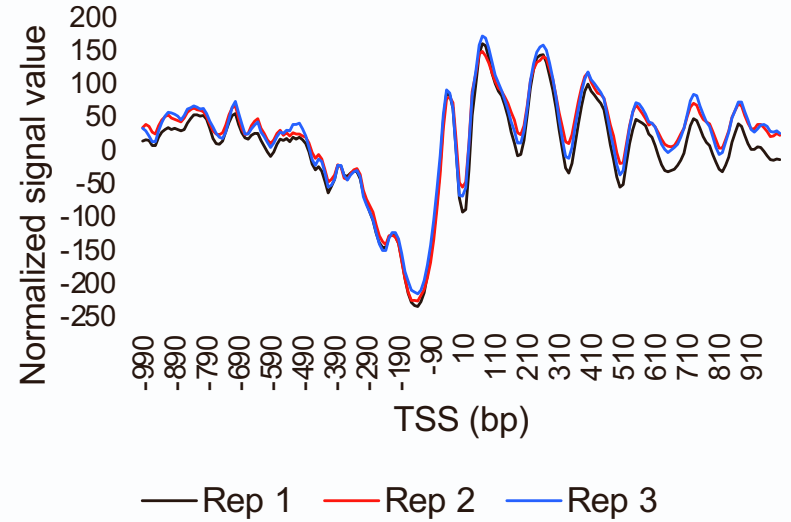
A

Nucleosomes that overlap by at least 80% (116 bp)



B

Metagene nucleosome profile



C

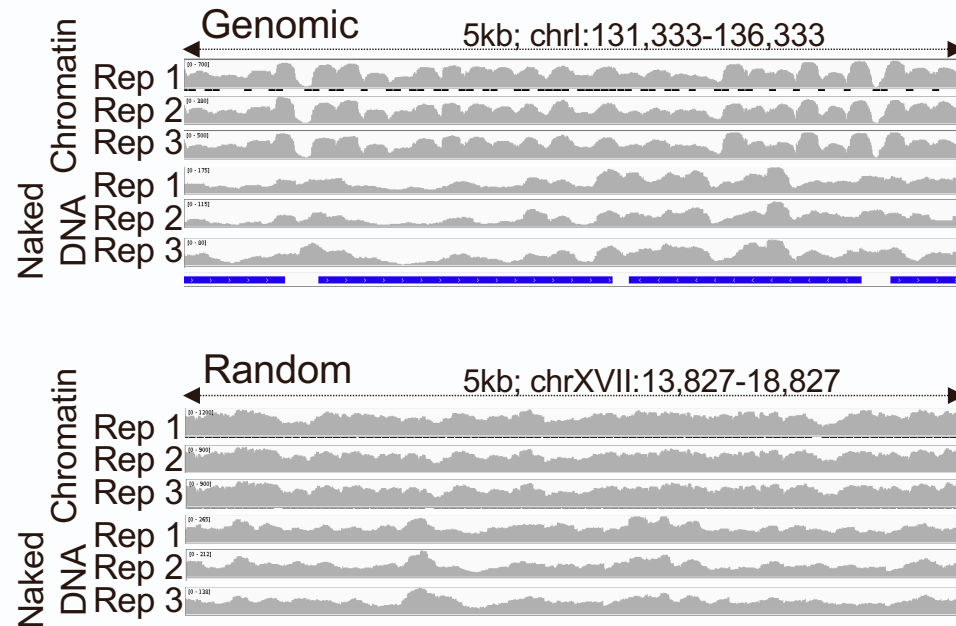
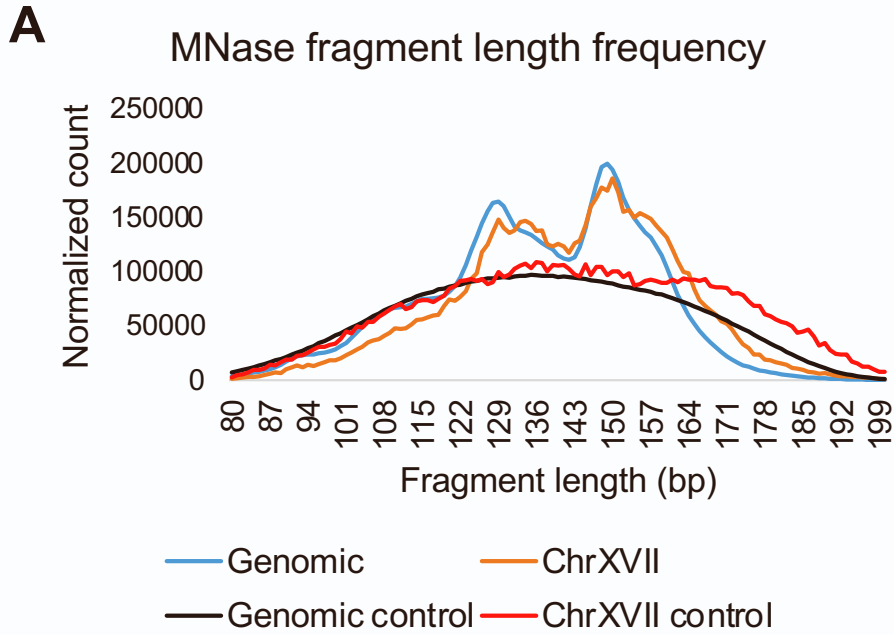


Figure S2. Evaluation of MNase-Seq replicates. Related to Figure 1. (A) Overlapping nucleosome positions in pairs of biological replicates. Nucleosomes with intersecting coordinates > 116 bp (80% in coverage; 90% of all nucleosomes) between pairs of biological replicates were considered to be the same. (B) Nucleosome metagene profile for three biological replicates generated by aligning normalized MNase-Seq reads with respect to TSS are similar to each other and to previously reported profiles (Lee et al., 2007, Brogaard et al., 2012). (C) Snapshot of 5-kb windows of genomic, random-sequence, and mitochondrial chromatin and naked DNAs of 3 biological replicates.



B Abundance of each fragment classes

Fragments	Genomic	Control
All sizes (total)	90969720	40191940
80-119 bp	19.3% (17537027)	22.3% (8969531)
120-142 bp	37% (33638247)	27.3% (10978508)
143-180 bp	42.3% (38446782)	33.8% (13571491)

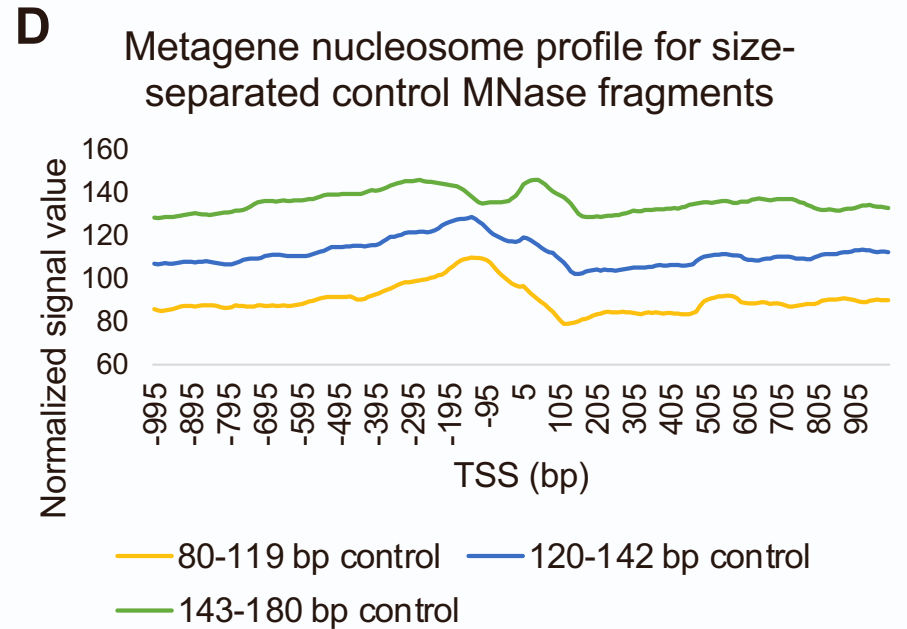
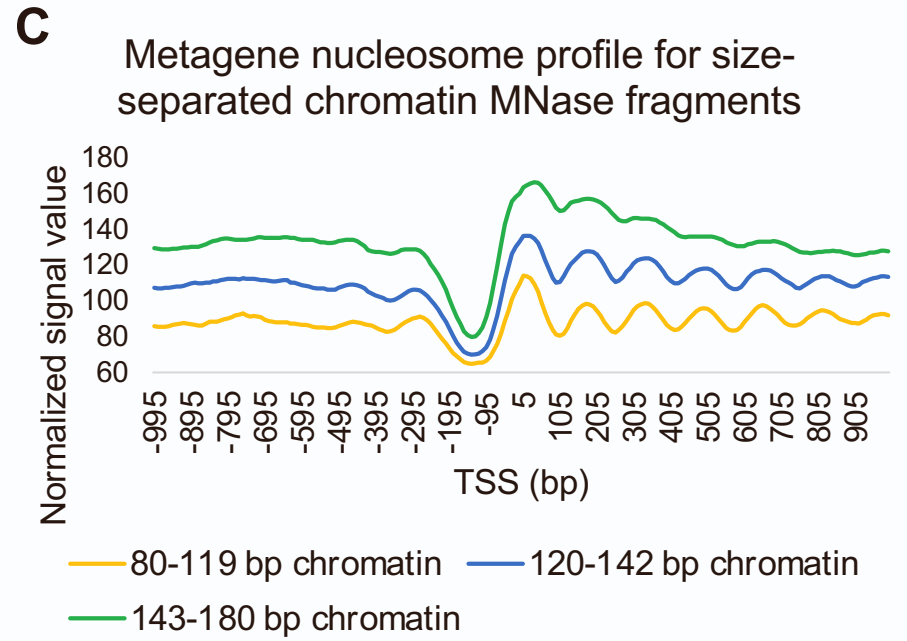


Figure S3. Assessment of MNase-Seq fragment classes. Related to Figure 1. (A) Distribution of MNase-Seq fragment lengths for genomic and random-sequence (ChrXVII) chromatin and naked DNA. (B) Number of MNase-Seq reads in the indicated size classes of chromatin and naked DNA control. (C) Metagene nucleosome profiles of the three different MNase-seq size classes. (D) Metagene profiles of the indicated size classes of MNase-Seq fragments of the naked DNA controls.

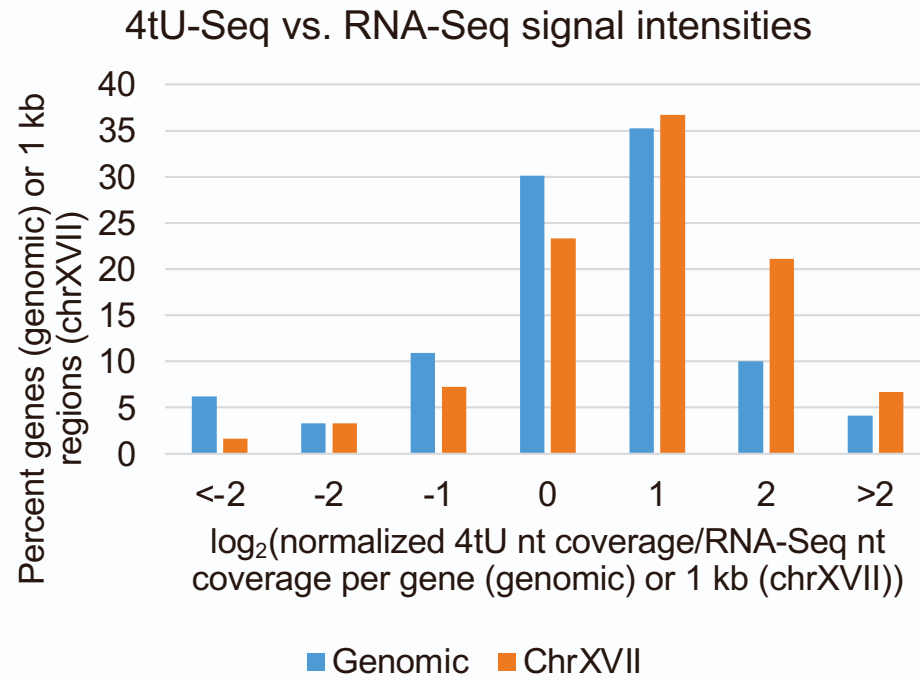
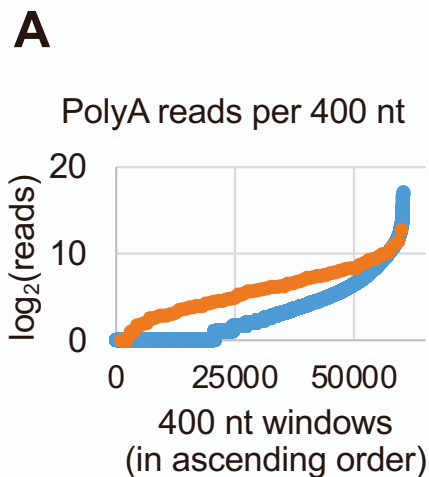
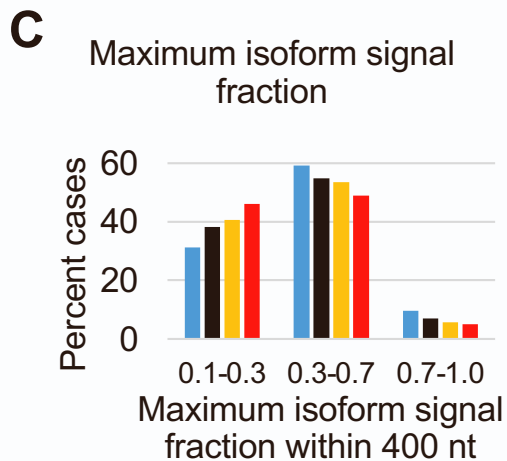


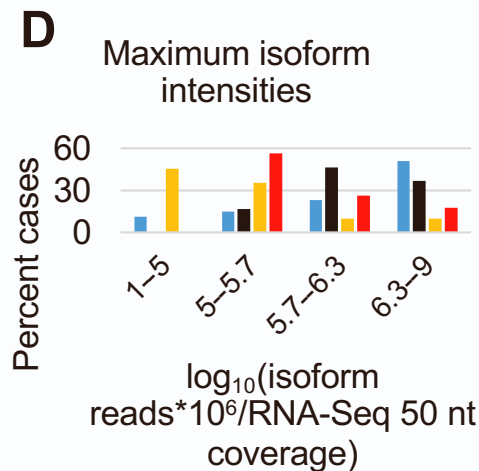
Figure S4. Active transcription from random-sequence DNA is higher than from genomic DNA. Related to Figure 4. Percent of the genes (for genomic) or 1 kb regions (for ChrXVII) with differential active to steady-state transcript levels expressed as $\log_2(4\text{tU-Seq/RNA-Seq})$.



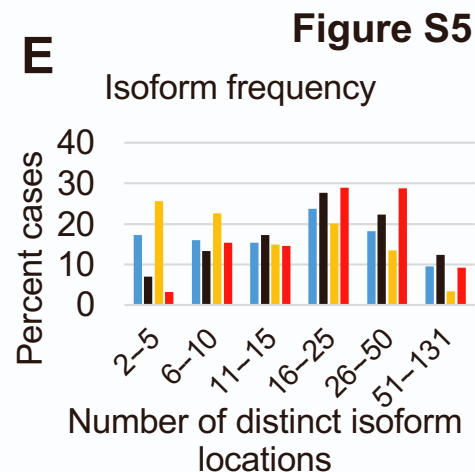
• Genomic • ChrXVII



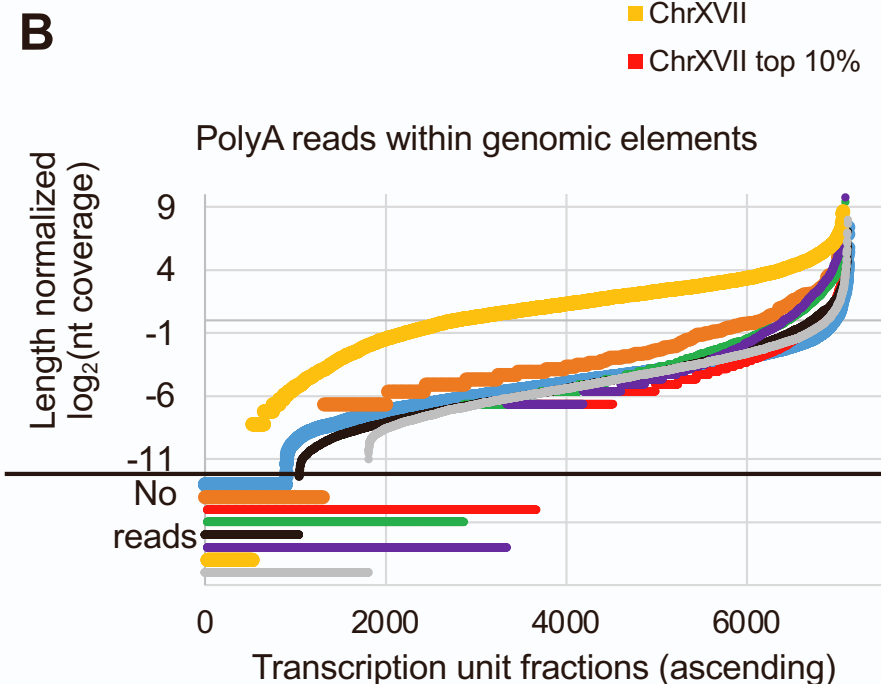
■ 3'UTR
■ 3'UTR top 10%
■ ChrXVII
■ ChrXVII top 10%



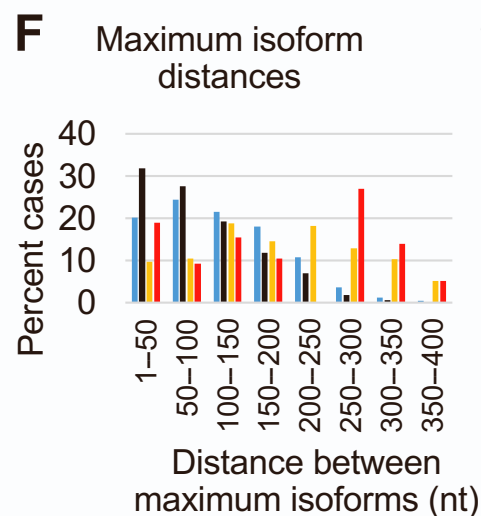
■ 3'UTR
■ 3'UTR top 10%
■ ChrXVII
■ ChrXVII top 10%



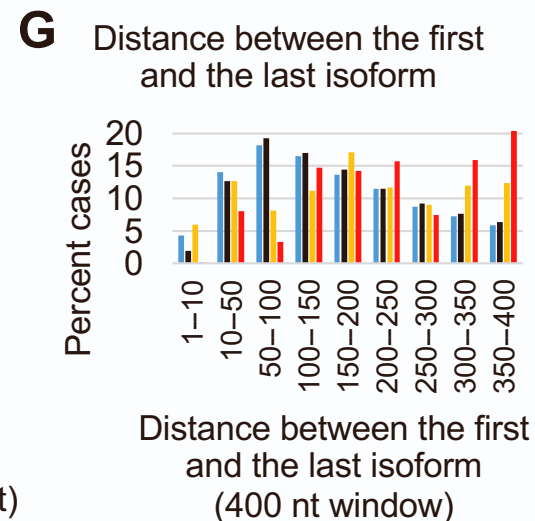
■ 3'UTR
■ 3'UTR top 10%
■ ChrXVII
■ ChrXVII top 10%



• Genes • ChrXVII • Promoters
• Terminators • Antisense • NDR
• 3'UTR • Intergenic



■ 3'UTR
■ 3'UTR top 10%
■ ChrXVII
■ ChrXVII top 10%



■ 3'UTR
■ 3'UTR top 10%
■ ChrXVII
■ ChrXVII top 10%

Figure S5

Figure S5. Properties of genomic and random-sequence 3' mRNA isoforms. Related to Figure 5. (A) Number of poly(A) reads (\log_2) per 400 nt windows in ascending order for genomic (blue) and random-sequence RNAs (normalized for the chromosome copy number; orange). (B) Number of poly(A) reads per 100 nt windows for the indicated functional genomic elements. (C) Percent of cases having the indicated ranges of maximum isoform signals (in percent of total signals) within 400 nt sliding windows in 3'UTRs and random-sequence RNAs (top 10% refers to the most highly expressed isoforms). Only isoforms with > 2 reads and isoform values after normalization to transcription were included in the analysis. (D) Percent of cases having the indicated ranges of maximum isoform intensities within 400 nt sliding windows in 3'UTRs and random sequence RNAs (total and top 10%). (E) The number of 3' isoforms within the windows. (F) Distances between maximum isoforms within 3'UTR and ChrXVII within non-overlapping adjacent 200 nt windows for every nt as a starting window. (G) Distance between first and last isoform within 400 nt sliding windows.

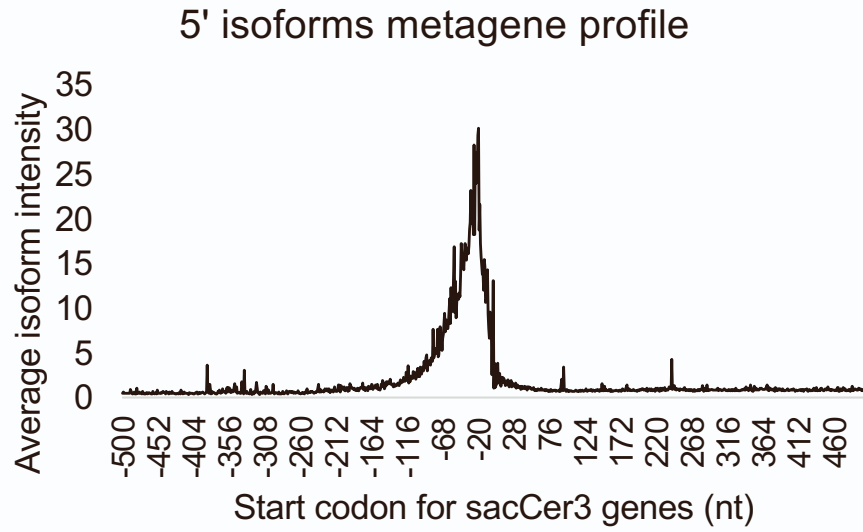
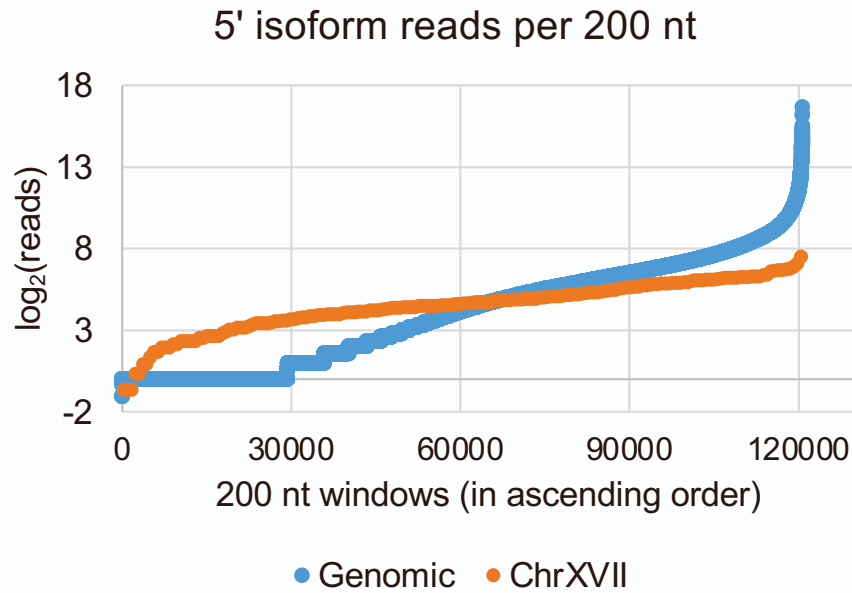
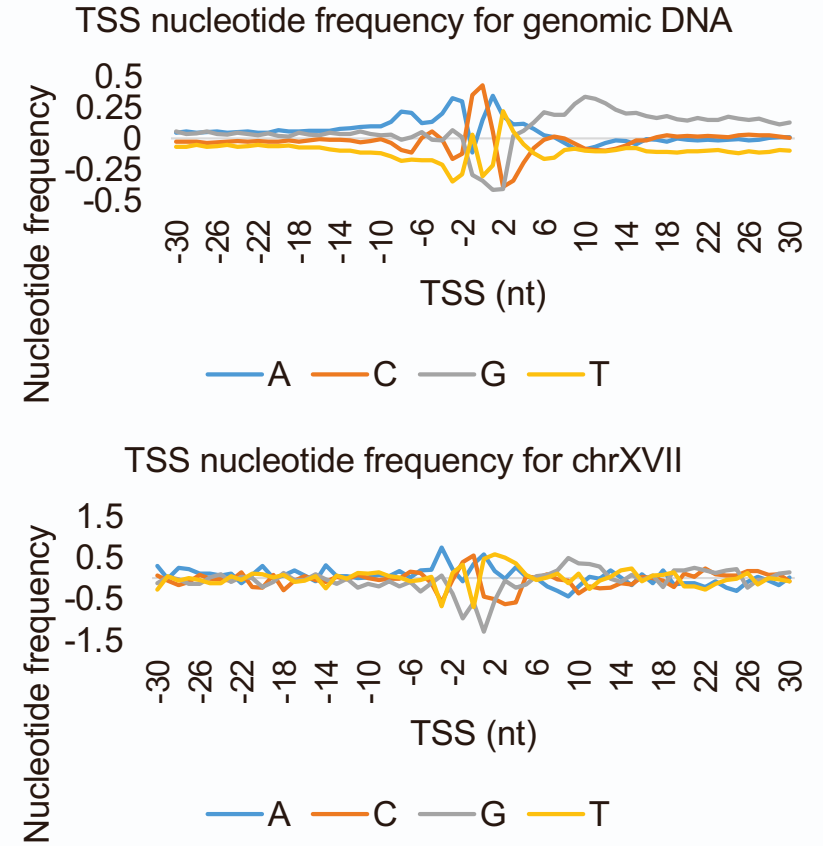
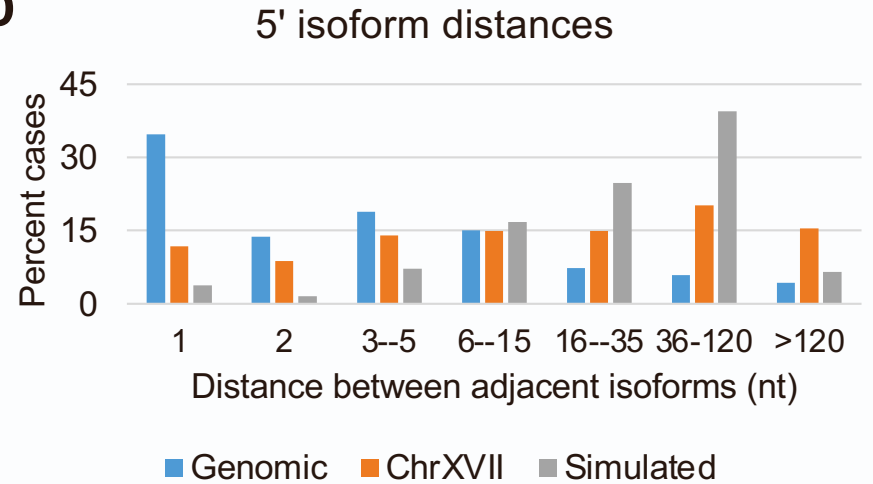
Figure S6**A****B****C****D**

Figure S6. Properties of 5' isoforms for random-sequence and genomic RNAs. Related to Figure 6. (A) Metagene 5' isoform signal with respect to the start codon for all yeast genes. (B) 5' isoform reads per 200 nt windows for genomic (blue) and random-sequence RNAs (normalized for the chromosome copy number; orange). (C) Nucleotide frequencies (normalized to each nucleotide in the genome) of the region ± 30 nt adjacent to the transcription start site (TSS) for 5' isoforms expressed from genomic and random-sequence DNA with > 3 reads per isoform after background subtraction. (D) The frequency of distances between adjacent 5' isoforms for genomic (blue), random-sequence (orange) and computationally randomized RNAs (gray).

Data S1. Random DNA chromosome (ChrXVII) sequence. Related to Figure 1. Random DNA sequence starts at 3911 and ends at 21878.

>chrXVII

GCTAGTGATAATAAGTGACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAA
CAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAA
AAAACGCAAAGCAATGATTAAGGATGTTTCAGAATGAACTCATGGAAACACTTAACCAAGTGCATAA
ACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTAATGATGACAGCCCCGGAAGCGAG
GAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTTCTTCTCAGGCTATCA
GAGATGCCGAGAAAAGCAGGGCGACTACCCCGATATGGAAATTCGAGGACGGGTTGAGCAACG
TGTTGTTTATACAATTGAACAAATTAATCATATCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGA
AGACGTATTTCCACCGGTGATCGGGGTTGCTGCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGT
TCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGG
GAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTT
TCTATCTTGGGGAAAAGGACGATGTCACCTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATT
ATTCTTCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTG
CCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTT
ATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTGCTGATGTGCTGATTGTT
CCCACGCCTGCTGAGTTGTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCA
AGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGCAAT
GGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGT
TGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAAGTGTGTTTGAACAGGCCATTG
ATCAACGCTCTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTT
TCGATCGTCTGATTAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTGTTATTCCAAAACATA
CGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTTCGTTAA
TTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAG
TTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACG
TGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACA
GACACCGGCGTTCGGTCAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAG
CTGCTGCACTTACCGAAAGTGATTATCGTGTCTGGTTGGCGAGCTGGATGATGAGCAGATGGCTGCA
TTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCG
ATTGCAGAATGAATTTGCTGGAATATTTCTGCGCTGGCTGATGCGGAAAATATTTACGTAAGATTAT
TACCCGCTGATCAACACCGCAAATTCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCTGGTGAAC
ATCTCCCCGTGATGAGTGCACCTCAAAAAGCCTTTACAGATAAAGAGGAATTAAGTAAAGCAGCAGG
CATCTAACCTTTCATGAGCAGAAAAAAGCTGGGGTGATTTTGAAGCTGAAGAAGTTATCACTCTTTTA
ACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAAGCTCACGACATCAGTTTGCTCCTGGA
GCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACTGGACAGGTCTCGTGTCCAACCTGAGTG
TATAGAGAAAATTGAGGCCATTCTTAAAGGAACCTGAAAAGCCAGCACCTGATGCGACCACGTTTTAG
TCTACGTTTATCTGTCTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCT
CTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCCTCGTAT
CGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCCTCGTATCGTCGGTCTGATTATTAGTCTGGGAC
CACGGTCCCCTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCCTCGTATCGTCGGTCT
GATTATTAGTCTGGGACCATGGTCCCCTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCC
ACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCCTCGTATCGTCGGTCTGATTATTAG
TCTGGGACCACGGTCCCCTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCCTCGTGT
GTCGGTCTGATTATCGGTCTGGGACCACGGTCCCCTTGTATTGTCGATCAGACTATCAGCGTGAGACT
ACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCTGTAGAACGGAGTAACC
TCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCACG
GTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGATGCAA
GTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTACAGTGTCGCTGATTGTT
ATTGCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGAT
TATTTGACGTGGTTTGATGGCTCCACGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGG
GTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTT
CCGGTTTAAAGCGTTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGA
AAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCTTCTGTTTTTGTCCGTG

GAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATAACATTATACGAAGTTATAT
TCGATGCGGCCGCAAGGGGTTTCGCGTCAGCGGGTGTGGCGGGTGTGGGGGCTGGCTTAACTATGCGG
CATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCACACAGATGCGTAAGGAG
AAAATACCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGC
CTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAG
GGTTTCCCAGTCACGACGTTGTA AAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAA
TTCTACTAAGGGGCAGGAAGACAGACTGGACGCGATGCGACAGACAAAACGGTGTGATCCTACTGCCT
ACGGGCAGAGTTAAGGCCTCAACGCAGGATTTCAAGGAGGAAGCACGCGCACGAAAACAGCGGCACG
CGAGCCGAAATCTCGTAGCGGCGCACGGTTACGTCATACTGGTATAAGGACGATTAGGAATCAGAAG
GATTCGCGGCCATCGTGCATTTGTGCCCGTGAGCATCTGGGGAGCGCACCCCTGCCAGACTTTCTGTGC
GATGTCGGGCATATTACTCTCGGGTAAGTGAACACCTAGTAGACTTGCTCAACAGACTCGTATGAAT
GGATATAGAAGTCCAAGGTTACCCCGGAGAAAATAGATCGAGTCATGCTTCATGGAAATCTTATGAT
ATCCTGGTCTGTGGTTGTAAGTCTGATTAGGTACCAATCTCCTCCATGTTATGTCCGAGCGAATCAATC
ACCGAGCGTTGAGCTGTCACGATAGCAGGCTCGGTTATGAGCGGCTGTGCTGGCTCGGGATTTGCTTG
GCGCATTGCGGAAGTGCAGCAATTAACGCTATCCGCGTCCAGTCATATCCCTGAATTGGTAAGCTATTA
GTAGTGGCGCTGGGAAACAGCGAGCTATGAAGAAAACCGTGGGAACAAGAGAGCTCTAAGGCTCGC
AATGGCTAACGGCGATACTCAAAGGCACTACTCTTGGGGTGGCTTCCCGCCGAGAGCACAACCTGATCT
AGGCCGTGATAGTGCAGGGCGTTTACCGTTGGGACAAAATCTACTAAACCACCGCCGAACACCCCAT
CTGCAGTGCTCAACGGAGGTTTAGGTA AAAATCATTTCGGTGGTGTATCTATACCGGAAGTGCCCAAGT
GGGGGGTGCACAGGAGCACATGACACTCTGTATGATGGAACCGAATGTTACGGAGGGGGGTTTTTCG
CGAATCGGAGTTAGATTAACGGTCCCCAGCGTTATACTGAAGTAGCGGTTTTGGTAATTCGATGCGGT
GGGACCGCGAGAATATTTCTGAATCGTTGTA ACTACTCCCATCTCAATGAATCCAGGTATTTAATCCAG
TATTAAGATTGAGGTGATAAGATATTGTTCTTAAGCGTGTTC CCAACTAATAACGGAACGGCTAAAAA
CCCAACGTCTTGCCGCCCCCTCTGTGGCAATTGCTGTCTACGCTATCTAGTTGGAAGCTAAGTGCTCT
GCTTGATGACAGAAAATAAGTTAATTAGTGTACATAAGAGCGAATCGTCATCCTCAGTCGATGCT
TCTATTATATTAGGCTGGTCCAAGTACCAGGTTCCGGAATGTAGGCTTTGGCGTTAACAAAATAGCTCG
TAGGTGCATACATTCTAATCGTATCGCGTTTCATTAAGGCTACGCGCGACCTGTTACGCGGGATAGAG
AACC GGCTCAGATGGTTGTGCAATGTAGAGAGGTCAAACCAGTAGTATACCTCTTGCCACCATTTAA
GGCATAAGAAATTATACCCCGCAAACGTGGACAGGAGACGAACCCGTCCTTACGTTGTCACACAC
CGCGGAAAGTGATAGGGTGGGCTCGCCA ACTATAATTGGTTCAGCTAAAATCTATACTAGAAATATGC
CAGATACCGTTACGCATCCACTGTCCAAATTCCGGGAGGCTAAATTTTAAAGTGACGGGTAAATTAAT
TTCGTTTTGATTTTACCAGGACCATCCTGATCGGCCGGAATCGGTGGTTAATCGGTTTTATAGGCTCCC
TGCTAGTAAGGTAAGTGAAGCAGCCCCATTTGAGATGGAAAGGGGATATGCTCCGAGGGTACGTTGC
TTGGCGCTAAATAGGTGGTGTACGACGTTGATCGGCCGCGCCCCTAGGACGAAGGACGGCCCCGGCG
ATGGTAGGATAACCTCCAAGATTGCGTTAGGGGCGAGGGCTAAGTACCAGGTTACAGGGTGGGGTCT
TGAGCTGTGGACATAAGGTTCCAAGCACACGGCTCACTAGTCTTCGGCGCGGGAGCTACGTGATCAGC
GATTCATTGTGCCACTTACATGCGGTAGCCAAGTTATGCGCGATTTTCGGTCTTTAGAAGGAGACCCGA
ATACCGATCTGGTAGGTTTAGTATGACGTCGAAATCAAATCGCGCATAGGCGGGACATTCTTGCAAT
TTATTGAATGCGCAAATATGCTCCAGCCTCCTGAAGAATAAACGCAATGATTTATCCGGACCCCTCCT
CCTAATGTCCCTAGTGCTTCGCCCGGAGAAGGGGTTACTCCTTAAGATTATTGTTCTGGATGCCTACA
TAGTGTGAAGGTGCGCTGAGTTATCGTTTGCTTCAATTCCAGTCATTCTGCGTGTGTAACACGCACC
GCGCCCTCTGCCAGATCAACA ACTCCTAGGCTGATACGGGTGACTGTCTTACATGAATAGCAGAACAT
ATTATTAGATTATTGGACAATCAGGATTGCATCGACATTTTATGCCTGCAATTGAATCAACTGACGGCC
GAGTTCGCGCCATGGA AATTTCGGTCTAGCTTATCTGTGCAAGCCAGTGGGGTACTACGAATTGACTT
CCAAAGCTAGTTTCTTAACGTGATCTGAGCGGCAGGAACTTTTCTGCTACCGCTTGACGGACCTATGTA
TGCTGAGCGCGGATTGGCAGTGGCCACGCCGTACCATGACAGCTATATCACGTACCCAGCCTTAGCTA
TTTCTTGAGGTGCAATTA CTGTGGCTGTACGATAGGATCAAGTAGCGTTTCTCCACGTTTCTCATAATC
CATCGGTGCGCGGAACTGACGCTTTTGTAGAGGCTGGTACGTATTACAGTCGATCGGATACTGTTA
CTCTGTATAGTAGCGCAAATGATGGGTTGACACTAACTTTTGAAGCACGACATAGTCGTTTAAACGAA
CAATCGCTGGACGCGCCCGTTAATCCTGGAGACCTGCAGCGCGGTGAGCCCTGTGCGCCGTTGCGCG
CGTGAGTGGTCATAGCACATTGACATGCGAAACACTTTAGCTAGTAAAAGGTACGTGCATTTTGTCCA
GGCATAACAAATCATCTCTTCCAATATTGCCACGTTGCACAAACTCGGGCACAGAATAATACCACAAC
GCCGTTCTCGGTA ACTTGGTCTGGCTCGCTCGGATTATGCGCGTTGTTGGGTAGCGAGCTCCCCAGTA
ACAAAATACTAATCTGCCC GCGCCTTCATCGGT CATCTGCGCAGCCATCACGTCTAGGAATTGGTCC
CGCATACCAATTTCGTA ACTTGATCCCTATGGTACTCGGCGACTCAGCACCATTGTCTATACTGAAAGG
TCCATACAAAGCCTCTAAGATTGCGATGCATTTTCGATGGCATGGATTCTTTGGTCGTTTTGCTTGCC

TATAATCATTATGCTGAAACGACAATCGCTAGCCCGCTCTTTCTACCTGAGTACGCTAATAACAGCAA
GCGGGCGCAGCCTCCCAACTCCCCATGCGGAAGGCATGAAGATTACGGTGTAAATAGTAATCGTGAAG
GAGCTAGGATGATGGGAACCGATTTTCAAGCCAACCTACGGGTCTCGAGTCGTCGCTATCTAGTTGTC
GCGGATAATGGGAGAGCTTGGTGTACCCTACAACATGACCGCGGCATTTAGGCTGAGATGCAATC
ACGAGCCCTTATAGCAGCGACTAAAATTGCCAACTGATTAAGGGTTTCTAGCTGGTTCGTCTATTTAA
TGCAATCAGGTACCCTGCTTACATACAGTGTCCGAGTATCTTTATTTACTGGGACGCCTGCGGAAAGC
AGGGGGAGCACCGAGCAATTTCTAACTGTGAGTACTTGTATGCACGTATTGCTCCGCTCTAGCGTAAAC
TTCGCGCGGCTCAAACCTGGGTCCAAGCCTGCTCTGGTCGTCAGCTCAGCCAAGATGTTTCAGAGAGCCT
GCTTGGTCATCGACCACACTTGTAGCGAGCCCAAGGCGAGGGCAACGCCAGCCCCTTTGATGTTGCT
TCGTAGAGGCTGTGAAAATACTCTCTTCCCTATCAACTCTACACCGCTTCTCTATATCGGCCGAACGAA
CTCGCGAGCGGGACGAACAGGCTAATGCATGGCACATTATTCTCAGTGTGGATTGTGGTTTTCAAGTC
AAATGTAGAATCATTTAATAATCCTGTCTCGACGTGCCTGTGCAATTTTCGCGCTAAAATGATGCCGAG
CTGGTAACCCATCCGCAGATTATTAAGGCCTTGGCAATCTACTTCCGATTACGCGCCATAGGTTTTGTA
TAATACTTCTTGATTTTCGTTTGCATTCTGCACATAGAAGTACGAGCTAAGATGGGCGCCAGCTGAGAGAC
TAAGCCTTGCAGACAACAAAGAAGGGTACGTCCTCACCTCAGCAACATGACCTTGTGATATCCCACGG
GTTAAGGTGATAAGAATTATAGCGAACTCAACTCCACATTGCCGTAGCGAGCCAATCCCCTCCAAC
AGCGAGTAACGAAAATGGTAATCCAGCTCCAGGGTTATCGTCTTTAGTACCTGCACACATTCAGGGTC
CCCCGCACGCCGGTAGGCCAGGGTACGATCTTTTGTGCATAACTGTAGGGGAGTCCTCGAAACAGGG
CTCGCTGAGACCACCCATTGAACCTTGGGGAGGCCGTTAACATTCGAACAGGATCTGCAGCCGGTG
GGGTGACCGCAGGGATATGGTGTGGTGTAGTAGATGCCGCATTTTTGGCGCATCCTGGAACCTCGTTAAAC
CACAAGAAGGACACTCGATTGTTCAAGTGTATGAGCGAGTGAAGCTTTTAAACAATCAGAACTATGTT
GTGATAAATCCCCCAATTTATAACCTATGAGAGGGACGTTGGTGTGATGCAGTTTTGCCGTGTGCATA
TCCTAAGTACTGCTTATTTGCAATCTCATGTACAATCGTAGCGACAATCACTATCGCTCACCTCACC
AACACACCGCGGGGTGATGCATTACCGTGCACCATCATGCACGGCCTTGCCTCGGGTCCAGGAGATT
TACAGAGCTACCCTGTGTTAGACCCTTCTGGCTCTCGCATCGCTTTAGTCTACTGATCCCTCTAACATA
ATTTCCCTCAGATAGACCCCTTGGTCTGCTGCTCATCTTCTGCAGAGCTAGGGTCTAAGGAGTAGCG
GATGCCACATCGATTTTGGGCGTAACAGCGCCTAGGCCCTACTCATGCCCTCCTGGCAATGAATACA
CCGATTATCGTTTTCGCCAGATACGAAACGAAACGGCGTACATCGTATATCAACCACCAGCGCCGCGG
TGAGTGGCTGAGCGGAGATAGTCTCATAAGTTGTCAAACTTTGGACTTTAGGGTACGTTCTCGA
AGGTACATCCCTCACGGGGTCCCTACACCTCACAGGAGAAAATTTAGCGCAATCGTATGGAAGAGCC
GCTGTTTTTATTGGGCGCAAAACGCCCGTACGTGTGTGCATAGACCACGCTTACCATGCAATCTAGG
ATAATCGAGGGGTCTAATGGAACCCCAATCTGTAAGACCGTATTTGAGGTGACTTGAGCTCCCCGCCA
TCGGGAGTGCTGAGAGCGATACGCATAAAATCAACTACACAACCTAACATCAGTATCCTCCGGGTTGACG
CATGAGGTGAAGTCCGTGCTGTGGCAAGTAACCACGCGTTCGTATACTAACCTTGATCCACATTTAG
AGTTGCCCGATGCCGTGCGTCCGAGGCGGGCACGTGGTTCACTGACCTTACCTGTGCTTTATTGTATG
TCGGCAAAAGCGGCTGAGCGGGTCCACATTTTACACGGGCTCTGTGGGATTCTCCGCGACCCATCG
GCTACTGTTTATGTTTCCACTCTTCATATTTTCAAGGGTATCTCAATTTTCATCGGTAAGCAGAATTGTC
CTATGGCGCGTATGTGGATTGGGTCAATTTATTGAATGAAGTATCGGTACTAGCTCTGTTTGCCTGTGA
CTCTAGACGTGGGCGAAGTTTCTTTGAAAACACTGCTTTCTTGTGTCAACAGCCCCGGTTAAGATATCTA
TAGAGATCAAATCCGACCATCTTCTTCCCAACATTTGAGTGTCTTCTCGGGCACTTCTCCCCCTCACC
GGCATGACTCGCACGCTCCTTCGCTACCTTGGAGCGACACTGTTGCGTGAGACTTGTACCCTAGTGGAT
AGCGTAAGAAAAAATTTTCGTGACGAACGACGACCGTCATTGCATGGCCGATTGCTATCTAATA
GTATTTCTTAGTTAGGATATAATACCACTGGGAAATCATCTTAGCCGAGTCATGGAACCTAACCGGTTT
TACAGTCCGGTTGGGTACGTGTCCGCTCGACCAACCGTTCTGATGACGATTAGCCTCTAGTGTGTATC
GATTGTAATCAAAACAGACTGTAAATCCATTTAGGGTTCTTGGCTACACGCTTTGAGCTACAATCTATA
AAGTTGGCCTAACTGAGTCCGACCCAAGCGGCTGCTGTGTCCGATCTGCCCGGCGTTTGAAGCCGGT
GGTGTCCACACGAACCTTAAACGTTTATGTTGGGCCCGTGTCTTTGCTCATTCTTCAAGCGTAGGCTTGC
CACTGATATGCCACTAGTGAACGCCGGGCAGTTCCTCTGCTCGCGTTATTGAATGGCCAACCTCTCCCT
GGCCCCGGTATCCCATGATATTTGGTGCAGAGAAAGGTGGCCTTGCAGCTCAACGCTAGCATTAT
TGAAGGCCATACTCGAATTTCAACGCCGTAATGTGACGTTTGTTCATAACTTCATATGAAGTTTCTCA
TCCAGCTGTGTGCAACCGCCTTACGTGCTTAAATAAGCCGCGGATTACTGTACTAAGGTCGTTGTGCG
CGCCAATGGAGAATCGAAAAAAGTAGTTGATCTGGTACCTAGAAAGCGATCTAACATTCAACTAG
CAATTCAGACACCGATGATTTACAAAATACTTGCCCAACCCGATAATCTTGATTTTCCGGTCTTTTC
CGACCACAAATCTGGGTGAATAGCACTTTAAAATACAGGGTTGAACTCGAAAAGGAGGCTGCACTGA
TACACCGGGCGTGCCTTGTACTGTCTTCGGTAAATACGCGCGCGGACTTTAGGTAGACATCACGGAA
AAGTAAGAGGAATAAATGGGTAGCCGGTCCGGATGAAGGGCACTCGTTGTTTAAAGAGTGCCTCCTTGG

GCTATCCAGAGTACCCAGTAGATACTCGACGAGGAGCATTCTGAGTCTCCGGATTCCGGTTCCGTTGAA
ATTTACTACTCGGTAATGGGCCGGACTATCATATCGAGTAGCGCCGAGTCTCATGCTCGATAGTTCCAAT
GGAAGAGTTGCAGCAAGCTAGGAACCCCGAAGCGCGATTCTATTTCCATAACGTTGTAGCGGTACTG
CAATGCTCAGGCCTGCTCGTACTAACACGACGCGGGCATTACAGTTATTACTCGCGATTAGCGGGATAG
TAAGTCGGCCGTTTCGCGAAACTCGGTTGGTGCCTGGCAAACCCGGCCAGCACTACGGGATCAGGTATAC
TATCTCTCCCGCAATTTGGATCGGGAAGAAGCGTTGTTGCGCCCGACATGGGGTCAGTCGAGTCTATC
AACGGGTTTTACTTTGAATTGGGTGTAACCCAATTTATGACTACTGAAGTCATGGCACGGACTGCCTTA
AGTATCTGAGGACTCAGCAATTCGTCGAATGGCATGTAATAATTCCTGATAAGAGACGCCCGTGTCTA
GAAACTGGTTTTGACTTTTACATATTACGGACTAGATTGCCCATCCAGGGTGCCTGCAGTACTGCTTGT
CTTCATAGATTACAACCTTCCCTAACACCAATCTAGTCTCCCACTTTATGCTTCTGGGCAGCCACATGG
ATAACGCCCTCTAAAGCATCATTACAGCTGCTGCGTTTGATAGTAGACTGTAATGGCAGAGCGGGTAG
TGGTTTTCCGCTGAGTCGCCGTGACCCACAATCTTCCAGGCGATCAGACTTAATCGGAGGAACCCGGT
AGTCGTCGAGGCAAGTAAGGCCACACAGATAAAGTCCTTGGCCGGTCTTTGTTCAATCGGAATGTTCT
AGAACGGCGTAGGAGATCGATTTAGGGTGGCGCCGAATAGGACACTACAAGCAGGTGTGCTAGGACA
AGCCAGCGCCTTGTGCTTTTCTTTGTGATCAAGTGTGGGCACACATCCTTTTCTTCTAAACTAGTCTG
GGGTTGACGACGTAGCGGTAACCTAGTGCCTATACAGGCTGAGTCATCCGGATAGTGTCTGGCTTA
GAATCCACTAACTAGGGTATTCTGGTAACCACCATTACAGCGCAGTGCCTAGGAATGCTGAAGGCTACC
CCTGACGAGACAACCTTACCCTCCATCGCCGCTTTCGCGTACAACCTACTCTACTCTGCATCGACAAGCT
TATGCGGGTTTTGATCCAGATACAGCGACCCTCACAGAACGGCAATCTCAGACGGTAGGATCGGACATA
AGTAATTCCTTGGTTACCAATGGGCGTATTAGGGCTAAGGACCTAGTCCACAGCAACCCCGCAGCTCTT
CCGTGCCAGCACGACTGTGGCCTCAGACAAAGGATTAGAGCGTGGTCCGTATGAGCCACCCATTGAA
CACACCGTGATGGATAGAAGATAAGTAGCCACATTCGTCTTTATAGAGTGACAGTAGTCGTTAAGT
CGACAATGTTCTGACGATAGGGTAACCGCAGGGTACGAAGCAAGAATAGAGCCCTAACGCTGGTAT
TTGGCAATAATAAAAGGGTCGAACGGCTCTGGTAACTCGGCTGAGGGAAGCCGAAGGCGCCGCTGAG
TTGAAGTCTCGTACTGTGACGCCAGCTCTAACCCGCGAACAAACAACACCCGAACCAAGCCACAGATG
CGAAGCCTTTTCCACGTAAGCGGAGGCGCGTGGATAAATTTATACTTAACGTAGTTAATCGGTCTGC
AGCCTTGCATTCCCACGCGAAGCTACGTCACCTTCTCTATCAATATGACCCGACATGCCCTTGAAGAT
GAGATGAATGATTAGAGAGGGCGCATACTAACAGATTACAATCGAGCGCAACATAGGCTATTTCTGTA
AGTTAGATTAATGGTAAATAGCTGGGAGTACTGCGAATCCACCGGTGCTATGCGACCACGACAGGAA
GCATCGAAATAGTGACGGTTTTTAATAGTTCGAGCTTAAAGCTAGTGTCTCGAAATGACACGCTTCGTTG
TTCGGCTTGGTCCAAGCGACTACGACAGCGACTCTCAAACTCTCTGCTGAAGAAGGCGGAGAGGTTA
CAGTCTCTAGCGTTGAGACGTTCCGCTAACGCCTTCATCTAACCATATTGACTAAGGGCGTCTCGTCCG
TTAAACCTCGCCCTGCCTAGCCGCATCCACTCACGCATATGGGCATCTAAGAAATTAGATAAACGATA
GATACCTCGGATTCTCGGATGTACATTCGATGTATCGAGTAGTCGCTGATTCCAACGGTAGTCCCA
GCCATCGGCAAGGTCTGACTATACTCGAGATTGAAGCTCTGCCAACGGTGGGGGCAATACATTGCTG
GCGATTTGGACTTCATATGGCAATGTATAACTCTTTTAGCGAAGGCTGTTAGTATGGCGGACTTTAAAA
GTCGGGTTTTCGCTGTGTATAATCGTGTGAGTGAAGGTGCGTATTGCGACACTGTTAAACATCTGGGGC
CCGGGACCCCAACAGTATACTTTAAGGATACTAAAACGCGTATCGGGGGTCTTCCCCCTTACCCT
AACGTATCTGGCTAAGAATCTAGTTTTGATGGCTAGTATTGTCAGTATTACTATATGCACTGTATGGC
ACGTCCTTACATGCTAATGCCCATAGTACGATGAATCTACAATGCACGGCTACGGTCCGGCGGGCGG
GGAACACTTACGTAGAGCGTCCAATTTTATTGGCAGGTTCCAACCTCTATCGGCGGGCTTTACGGTAAA
GCGCCTTTGCCGTAGCCGACTAGTGGGCTTAAACAACGGCGTAGGTTACAAGTTGTCTTGGGTTTAGTGT
CCAGCTTCGACTTTGGTCCCTAGTTCGATTATGAACACCCTCCCTCAGGGACCTTTTAAATACCCACGCG
TGATCTTTGTCTGCCCCGGTACAATGTATTTCTGTGACCCTAAGAACGTAATTTGATCCCACAAGATG
CGAGCACTGAACCGCGCGGTTTCATTTAGGATTGGGTCGAGCTTTTCCCTAGTGACACGGGTCAAACA
TTAATCATACAACTAGGGTAGTAGCAAGTTGGCACCTGGTTAATATGCGGTATCGTTCCACCCGGA
CACTAGACCCGAGACCTTCTTAGTCAGTCAGAAGGAGAGCGGAGACCTTAGTTCAGGTCTCGGTCCTGA
TGAGTCAGTACCAAAGGCTGATTTACGCCGGAGGACACGTGGATCGGGGGTACTACGGTCGCTACAGT
GAATAACGCCGCACCTTTATATGGCAACTAAGCGGCCTTAACCGGCTCGTATCTGGCGCCGGACCTCG
AACTTGCATTATGTGCGCCGTGAGGAGAACTCTCCTCCGGTCACGGTCGGTGATTTCTTACAACCGG
GCGGCCGGTACTCTGGATAGCGTGCCCTCTGGACTATTTCCATACCCTGAATATAAACCATAAGCG
CCGCAAGCGGGTATCATAGGGAGGATCAAAGAACCGCACCCGGCCTTATCATTAGTCGTTTCAGA
TAGGGAATTGTAAACCCCTTCAAGGTTTCGCGGTACCCTGAAGCCGATAACAATATCAATGCTCGAACT
ACTTACGTCAATCGTAATCTCCATTTGATGTCCTTCGTCAAGTATACGTCGGATATAGATAGGTCACTG
CGGCCAGACCGCAACTGGGAGTTTGAAGGTACAACCTGTTGTGTGCAAGAAAGCTGAAGTCGCGCGCG
CTAACCAACATGCTCTTACCTTGGGCGGCCCTCTCGCGGATTGACCTGACATACTGAATGCTAACAG

CATGCCACGTTTCGTAATTGAACTGCTACCGATACCTTCGTAACGCTAGATACGGTATTCTCTATTCTAC
TCACCAAAAGTATGTATCAGTGC GCGATAGGCATCGGCCCTCGATTGGCAAACGTTCCCTAGCTGATTTT
CGTATGTTTCATCGGGTACTAAAAGCCTCTGCGTGGAGGCGACATGACGCAACATACGGTATTTTGGCT
CTAACTGGTGGCAGGGTGTGCCACTTTTATATCATTGTCTATTTAAAGCCAAAGGTGAACGCCCTAGA
AAGAGAACGGTATGGCAGTCTTTGGGGTGAACGAGGCCAGCGTAGGCACAAGCAAATCAGTCGTC
CAATTAATATACTCGACGAAATATTGAGGGAAGCGCCAGCGCGTACCCATAAGCTATATGTCACGCCA
TGCACCATCGACATTAACACTCTCAACGTAATAGGTATATACAATTATCGACATTACGACATATCGGT
AAGATATCCGAGGTCTCAGTAAGGGTGGCAGGCCCTGGAGGTGAGCGGATTATTCTCCGGGAAAA
TGACGACGGATATGAGAGAACACGATCGGATTGGGTTGCCTACCTGTCGTCTTTGAGCCGCATTCCCG
AGAATGCTCAAAGACCCTCTGTGCGGGCCTTGCCGGCCAGAACAGTCCCCTGGTTAGAGAACATTTG
GTCGCGAAGGTGCAGGTTTGACAGTATCTCCATGTTAGGGCCAAGGACCTTCGTATGTATCCAGACTTT
CGCAGTCGCATTTCTCCGATAGCAAAGCTCGTAAGAGGAATTATCTTGTGCGACCTCCGGCTGTAGA
GGAGCGTTCATCATGCCAAGAAACAAGGCGAATGTGTATGATACGTCGCGGGAGGAGAGGTGCCCAT
GTCACGTTTGTACGAAGTTGTACACGGTAATTATACTTGTATCCGAAACTAAGTTCGAAACCGGGG
GTAGCACACCCTCCCATCGACTCAGAGGCAATTACAGACACACGTCGCGGAAGTAGATTTGGTCTCTC
TAGTCATACCGGAGGTGAAAGCCTGCTTACGCGCTGGCGTCTACCTCGTCAATAACAACCTATTG
GCCAGAACCGTATTGCTTGGTCAATCCGGCATCTGCGAGCAGCATTGCGCGACGCTATAGGAAGCG
GTTTCGAATGAGCGATGCCATTAAGTTACTTGTAGACCAGTACTACGTGTGGGGTATGCGGTAATAC
GAATAGTCTAACCGGTATGGGGACCATTACGTGTGCGGAGCATCCGAAACAAGAGATATCGAATAA
TCGCAGAATCGTAGGGAAGTTCGTGTAAAAGACATGGGCTGCTGTATACTTGTGTAGAACTCCACTTA
TATCTCCCAAGCCAACCGACATTGAACGCTAGACGGACGCACAGACCGGCCATGTTTCGCGCTCGATT
TGAACGAGCATCATTCTGTCTCGCTGCCAAGGACGCACCAGTGGCGACGAACTTCAGCCGAAAGTAA
ACAGGTTACCAGCGGGCAGCCTTGATACTATGCCTGTACGTAGAACTTCCAGGCTAGCAGACAAGTTA
GACAAGTACAAGACCAGCTATGACTAGGCTTACACTCGGTTCTAATCCATGGCAGGAATTTGCTTATG
CGTAGAACTGAGAACTGTCGCCCGCAGAGATTACTACTATTCAGCCCCTACTAATACGCTACCCCG
TTAGAAATCCCAATTTTACTACAAAATCGATACCCTGGTTTCTGACTCCAACATCTCTTCGTCTACTGA
TTTCTGAAGCCCTTGCAAAGGGCTAGTATGGCATCTTTATATCAGTGGCGCAAGGCTTACCACTTATCG
CACTGCGACGATCGCGGGTACGCGCACCTTATGCTGACCGTTAAGTTAGATTAACACAGTTTCGGGA
CCTTGAATGACGTTTCGAGACCAACTTTAGCGAGTGTCTTATAGGGTGGAGCTGAAACTTCTTTTAGCA
CTAGGTGGGTACGTGCGCCAAATCAGCAATCATCTATGTAACAATACGGACATGAATCGTTTTGGGA
ACCCGTCTAAATTCAAGAGCAGGATGTGCTTTCGACTCAAATCTCGGCACATTATGATTGGGTGCCA
CGCGTGCCTTCTGTGTTTCAAGACTTTAGCTCTAGCTTAGCACAAAACAAGGAGCTGGCCCCCTAGGT
CTACTACCAATGGAATATACTGACGAGGCTGATGCGTATCCTTAACAATTGACAGGCGCGGGCGAGGA
ATAACGGGATTTAGGATAGGGGATTTCTCTCTGTCATCGAGTCCGCGCCAGTCAGCAAGACGAGCG
CTTAGAGAGAACTATTGCTGTCAACCCGATGCTTAAGTTAGCCGAGTGTAGCAGATACAGCCACCAAA
TAACACCGACGTCGCGAGTGCCTCAATCTTTAATGATCATCAAGCACACAGAGAACTACTTCGATC
TCCTCGTACTGGAGGAGGGAGCTCGCGAGGACGACTAGTCATACGTCTGAATGTGCAGTGTAACTAC
GTAAAGACAGTGGGAGTTGAGTGGCCCTACCGTACTCTGCCGCATCGTAAAGCCGGGTTAGATAGACT
CCGCGATTGCCACCAATGACCCATTTACCACGCCGGTTCGACTTGGTGCCTTATGCATACGGATT
GCCATGGTGCATTATATTGGAACCAATTGATATCGTCTGCGAAAGCAGTAAGGGTTTGAACGCTCC
GATGCGACAGATCCCGCATAGTTGATTGTTGGGTCTCCACGATACGTTTCTCAACCAGGCACCTCCAA
AAAGAGTAAAAGAGCAATGTCCTTACGTGTGTAGTTAACAACCTCGTATGTATCCGGGGTTCGGCAA
CCTGCATTTGAATATCGCAATGCGCGTGTCAATTTCTGCATCGTAGCACCCGCTTCTCGCTTTTTACGTA
TTGGCGTTGCCTCAAGACCAGTCCGTCAATTTAAGGCGTGTGGTTCTACCAGTTGCCTTTTTGAACTG
TCGAGTGATAATAATGTACGAAAGTCGGGTCAGGGTTTTCCGCTAGCGCGATTAGGGTACCTAGGGTG
GTAACCTGCTCAGGTGCCGTATCTGCAACGCCGTCGTGCGCGTCAAGGTTATGTCCGCTCATACTCT
AACTGAGCCCTGGTGTGTTGTACATTACTCGGCATCTCACGAGAACGCCAACGAAAGGATTTTATA
ACGGTGTCTAGATGTCGGGGTTCGCGGCACCTTAGGTGAATCAACTTTTCGGAGTACTACCATTGAGTT
TCGATAGTTGGGAGAGGCGGGTCTGCGTCAATCGCCGAAGAACATGTCTCTATGTGCCTAAGTTGGGT
CTTCATCAGCGTCGATTAGGTGAAGAGAGGAGCTGGGTGCAAAGTCCGTTACGGACAGCCAAAAGAT
ACGTGCTACGCGCTGAGTAAATGGATAGAGCCGGGTTATGATAACCCCCAAAATTGGTGGATCTAAAT
TAGACTTATGTTCCACAGCTAGGCTTTCCACCCCTGGCACAAAGTTGTATATACAAGATATTGGCCT
GGTGTAGCTTTTGTCTCCCGGTAAGAAAAGCCGCATATCTCACTTCTAGCGGCTGTTAATGTAAAGGT
GCTGCATGGAACAGAAGTTCGACCTTGATACCGGCATTGATTCGATATGAATTGTAGGCAGCCTTTA
AAGCTTTCCAGATATAGCATGTAGCAGTTGGACCCTGTGACGACTCTAAGCCACCGACTCAGCATCCA
GTAAAAATGCGCCTAGCCACCAGTCCGTGTGTACCCGCGCTGTGCTACAAATTTTGTCCGTACCGCT

ACCTCCGTCCATCTGCATAGTTGGAGGTCTCTGAGGAATTAATAAGCTTTACTAGCGCTTCTTTATGAC
CCTACGTGACACAGGCCGACGCGCCAGGGATCTATACTCTCAGCTTTTCGAGTTGGCGCCCCTGCTGA
GACCATCGTCATCAGGCAAATCGGATCGCACGTCAGCCTTTTTCCAAGAATCGCTCGGCTAAATGTAC
ATTGCAAAAGAGGGCGAAAGGGCTCTAAGCCATATAGTCCTGGCAAAGGGAGGCAAGGACGGAGGG
CGGCTGATAATCGTAATACCAAGCACAATGCGTTTCGGAGCATGTTGGATCACGAATTGCCACTTCGTT
GACTCGGATGACTAATCCTTCACACGCAACCTCCATATGCGACTGAGGGCAATCAGTGCTGGCAATTA
TATTTCCCCTTCCGCACAGAATATCAACGATAGTACGTCATTGCGGCCGGTACAAGCTGTGTCAAACA
GCACGCTGCGAGGACGCGTTAACAGATGTGGTCTACGGACGGATTGAATTTGTTCTAACTTCGGTAA
TGGAGTTTATTAACAATTCCCTTACGGAATATGCGCAAGAGTCCGTATCACTTGATTTAAAACCTGAACA
CAGGTCGGCGCTTCTATGCGCCATAGGGGTTGTGCCTCCGAACTAAGAGCTGAGCCGTATTTGCGGAC
CTACCTGTGTAGGTGTGGGGCGGTCATCGTCTACGCTGACGCACTCATGATGTGGCCCGCACAAG
ACGCATGCTCTATGTGGCCGATATCCGGAAGGTTCCACGATAATCAATCTGTCGAGCCCTCCTTCTAG
AGCGTCCAAGGGCCCCATTGGCCCCATTGTAGAATCCAGCTATCCATGTCAGGATATTGATAAGACTA
TTGTCATGCCGAACATAACCGAAGGTCGCTCGAACAGCCAGGGGGCAATCAGGAGGCAGGGGGCCCAA
ATGGCCACTAAGAGAGTTAACGGGAGAGACTGAGAACGGCTAGGTGATGGTACCGACCGTCTGTGGC
TCGCCCTCTAGGATGCTTAGATTCTCCCGAGTGCCTTTGGGGCCGTAGCGTTGTGATTTCCGCCCA
GCGCCGGTCAAATTTGTACAAATCGCAATGAGGGGTTCCAGATGTTCAAAGCTGTTCTTAGACCCTC
TGCTCGGTCTCTAATGCAGAGTTGAGAATTGCGTGACATTTCTGTGCACAGCCAGCAGACCCGTACG
GGCCCCAGTTGCTTACGGGACGACCTTACCTCCATCCGATCAGGTCCGAGCCAGGTAACCTTGATCCAC
ACACCCACGTGAGCGTGGGGAACCTAAGATCATCCATAAAATAACAGGTCTATACTTTTTCTCCTTCG
TGATCGCAGACTGTCTGAGGTATCATGCCGTCCAATACAACGATGACCATGGCAGTCGCTCCGACAGG
GTAGGCTACTATTGGTAATCCTGACTACGAGCTTCGTCAGGGTGTTCAGTCTTAAAGTGGGCTGTCTT
GGCACTTACGATGTCCATTGATCCGGTGGCGATTGGGTTGTACTGCGATGAAACCGTTCTTAAACGTCGT
TCCGTTAATTCAGATCGTAGATGGGCGATTTGCCACAAGTCCGAGGGGTCCCGACTAAGAGAGTGGGG
GAGGAAGTCCCGTAGACGCGGTCTGATAAGAAGCGCGTAATATGTTGATAGCGATTTAAAACCCCTAA
GCTCATCGCCACGCGCACCAAGCCTGTTTTAGGACCGTTTTAAATCTGCATGACTCGAAGGAGGTATTG
TGCTAACAGAGAAGTTTAGCGTCCGCTCTGTCTCCAGTCTTCCGTCCAGTGGGGGAAAATTCCCCCTC
AGGAGTTAAATGTGCCGAAGCGTCTTATCTCGTTACGCGTCGGCGAACATCTCGAATCTAGTGGAT
TCATCATAAGATGTTTGACAATGCGTGTGAGGATGGCATTTCGACCATGCACGTGGGAAGGGGCGGCC
GTTCAAACCTATTATGGGCACGCAGTTCAGAGGCGAAACCAGGGCGAGCCTATAGATGTCCTTGGCAT
TGAAGGCCGAGTCTATCTCCTCTACATAAATCACGTATAGAGAAGAGAGAGCAATTCCTTTCACAAGC
ATATCAGTACGCCTCAATCTCCATGTGATTCACTTTTGTGAGCGTATACTCAGTTCTTACCGGCCACA
GAGGACTTGAACGCTCTTCATGGTGAGCTAGTCTTTTATAATATGATGTTTCTTACTCTAAGGCGCTCC
ATTATACGGTTCCAATACTAACTTCGATCTCGGACGGCCTATTTGTGAGTGTGCAGAGGCAGTAGTGT
AAACCCAAGAAGTGAGCCCCGCGTTGCAATCTATCGAGATCTGCGCCTTTAGACGGACGACCGACCTG
GACGAGAACATTTAAAGCGCGATGGATCATCGTTCCGGACCAGTGGACTGCCAGGTTGATGCCTACCA
AAACCTTTGGAGGTTACATCGATGAGCTGAATCGGTAATAGATACTTAGTATCAAGCGTTCGGTCCAT
GGGACTCTAGTCTGTATATCCGGCGCAGTAACTACGCGAACGCAACTTACCCTCTGCAGATGTCT
GTTTCCCTTAGACACGAACGTTACACCCCGATCGATATTTTCGCATAACGTAATGAATATAGTGTAGTAA
ACGACGCTACTTTATGTGAGCACGTATAACAGCTCGTGAACAAAAAGCAGCTAGGAGAAGCGTGC
GTGGAGGGATGTCTAAAACCCGGGGTCGCCCGCAGCGCCAGTAGAATACGCAGATGGTTCGCCCGAA
CGTCGTGGGGGATACACTGTATATAACGCGTACTATCAATGAACGCGACAATAATTGGATGCGTGTA
TCGAGAGCGCAGTTTGTTAATTCCTCAGGCGGCCACGAGAGGAACTCCGCTAAAGACTTGTACCGAGC
CGGGACTACTATTGGTAACGAGTCTGGAGAGTGACTGGTAGCACGGGGTAATAGCAGTTGCTTGTCTG
GCCCTTGATCCTGTCAACCCACAATCCGCTTCCTTGAAACCGATAGCGGTTTTAACTAGGACTATCAAA
CAAATCGTCTCTATGGCGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATG
CAAGCTTGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGT
GAAATTGTTATCCGCTACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGT
GCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCCGCCGGGCGTGCACCAATT
CTCATGTTTGACAGCTTATCATCGAATTTCTGCCATTTCATCCGCTTATTATCACTTATTACGGCGTAGCA
ACCAGGCGTTAAGGGCACCAATAACTGCCTTAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGT
ACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAACGGCATGATGAACCTGAATC
GCCAGCGGCATCAGCACCTTGTGCCTTGCCTATAATATTTGCCCATGGTGAACCGGGGGCGAAGAA
GTTGTCCATATTGGCCACGTTTAAATCAAACCTGGTGAACCTCACCCAGGGATTGGCTGAGACGAAAA
ACATATTCTCAATAAACCCCTTAGGGAAATAGGCCAGGTTTTACCGTAACACGCCACATCTTGC

TATATGTGTAGAACTGCCGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTG
CTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTTCATTGCCA
TACGAAATTCGGATGAGCATTATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTG
CTTATTTTTCTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTG
AGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATC
CAGTGATTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAATACGCCCG
GTAGTGATCTTATTTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCC
AAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCT
TCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGGCTAACCGTCGCACAGGAAGGACAGAGA
AAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCG
TGCTGCTGACGGTGTGACGTTCTCTGTTCCGGTACACCACATACGTTCCGCCATTCTATGCGATGCA
CATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAA
CGAAGTCTACACGAAGTTTTTGCCTGGATGTGGCTGCCCGCACCCGGTGCAGTTTGCATGCCG
GAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGAAA
TGTGGAATGAGTGGATATGCTGTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAG
CGAACGAAACAGTCGGGAAAATCTCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTG
TAGCGTTTATAGGAAGTAGTGTCTGTCTATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGC
CTTCAGGAACAATAGAAATCTTCGTGCGGTGTACGTTGAAGTGGAGCGGATTATGTCAGCAATGGAC
AGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCTTTTTACAGCCAGTAGTGCTCGCCGAG
TCGAGCGACAGGGCGAAGCCCTCGAGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGCTATGGCCCT
GCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCAGCGGCCGCGCCGCGGCTTGTGGA
TACCTCGCGAAAACCTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTC
ACCCGGCGCGGCGTTGACAGATGAGGGGCAGGCTCGATTTCCGCCGGCGACGTGGAGCTGGCCAGCC
TCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCACAGATGATGTGGACAAGCCTGGGGAT
AAGTGCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACAC
TTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAG
AAAATCCAGCATTGCAAGGGTTTCCGCCGTTTTTCGGCCACCCTAACCTGTCTTTTAACTGCTTTT
AAACCAATATTTATAAACCTGTTTTTAAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAA
GGGGGGTGCSCCCTTCTCGAACCTCCCGGTGAGTGAGCGAGGAAGCACCAGGGAACAGCACTT
ATATATTCTGCTTACACACGATGCCTGAAAAAATCTCCCTTGGGGTTATCCACTTATCCACGGGGATAT
TTTTATAATTATTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCT
GGTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCTCAAATGACAGTCTT
GTCTGTGACAAATTGCCCTTAAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTA
TCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAAATCAAAAACCTTGTACACTTACATGGATCTGT
CATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAAAAATAGCCCGCAATCGTCCAGTCAAAC
GACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGTATCTGTTTCGTTGACCA
GATCAGAAAATCTGATGGCACCCACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATG
CTGAAATATTCGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTTGCGG
GGGAAGGAAGTGGTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAGGCTATGAATCTTTTCC
TTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCAT
TCCCTTCTTATCGGGTTACAGAACCAGGTTTACGCAAGTTTCGGCTTAGTGAAACAAAAGAAATCACCA
ATCCGTATGCCATGCGTTTATACGAATCCCTGTGTGATGATCGTAAGCCGGATGGCTCAGGCATCGTCT
CTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTC
CGCCGCCGCTTCTGACAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATAACAT
GAGAAAAAGAAAGGCCCGCCAGACGACTCATATCGTATTTTCTTCCGCGATATCACTTCCATGACGAC
AGGATAGTCTGAGGGTTATCTGTACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGG
GTAATTTGTACAGTTTTGCTGTTTCTTCCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTT
TAAGGAAGCCAAATTTGAGGGCAGTTTGTACAGTTGATTTCTTCTTCTTCCCTTCGTCATGTGACCT
GATATCGGGGTTAGTTCGTCATCATTGATGAGGGTGGATTATCACAGTTTATTACTCTGAATTGGCTA
TCCGCGTGTGTACCTCTACCTGGAGTTTTTCCACGGTGGATATTTCTTCTTGCCTGAGCGTAAGAGC
TATCTGACAGAACAGTTCTTCTTTGCTTCCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGG
CGAGCATCACGTGCTATAAAAATAATTATAATTTAAATTTTTTAAATATAAATATATAAATAAAAATAG
AAAGTAAAAAAGAAATTAAGAAAAAATAGTTTTGTTTTCCGAAGATGAAAAGACTCTAGGGGG
ATCGCCAACAAATACTACCTTTTATCTTGTCTTCCCTGCTCTCAGGTATTAATGCCGAATTGTTTCATCT
TGTCTGTGTAGAAGACCACACGAAAATCCTGTGATTTTACATTTTACTTATCGTTAATCGAATGTAT
ATCTATTTAATCTGCTTTTCTTGTCTAATAAATATATATGTAAAGTACGCTTTTTGTTGAAAATTTTTAA

ACCTTTGTTTATTTTTTTTTTCTTCATTCCGTAACCTCTTCTACCTTCTTTATTTACTTTCTAAAATCCAAAT
ACAAAACATAAAAATAAATAAACACAGAGTAAATTCCCAAATTATTCCATCATTAAAAGATACGAGG
CGCGTGTAAGTTACAGGCAAGCGATCCTAGTACACTCTATATTTTTTTATGCCTCGGTAATGATTTTCA
TTTTTTTTTTTTCCACCTAGCGGATGACTCTTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTAT
ACATTATATAAAGTAATGTGATTTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAA
GGCAAAGATGACAGAGCAGAAAGCCCTAGTAAAGCGTATTACAAATGAAACCAAGATTCAGATTGCG
ATCTCTTTAAAGGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCCAGAAAAAGAGGCAGAAGCAGT
AGCAGAACAGGCCACACAATCGCAAGTGATTAACGTCCACACAGGTATAGGGTTTCTGGACCATATGA
TACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGTGACTTACACATAGACG
ACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTTTAAAGAGGCCCTACTGGCGCGT
GGAGTAAAAAGGTTTGGATCAGGATTTGCGCCTTTGGATGAGGCACTTCCAGAGCGGTGGTAGATCT
TTCGAACAGGCCGTACGCAGTTGTCGAACTTGGTTTGCAAAGGGAGAAAGTAGGAGATCTCTTTGCG
AGATGATCCCGCATTTTCTTGAAAGCTTTCAGAGGCTAGCAGAATTACCCTCCACGTTGATTGTCTGC
GAGGCAAGAATGATCATCACCGTAGTGAGAGTGCGTTCAAGGCTCTTGCGGTTGCCATAAGAGAAGCC
ACCTCGCCCAATGGTACCAACGATGTTCCCTCCACCAAAGGTGTTCTTATGTAGTTTTACACAGGAGTC
TGGACTTGAC