ARTICLE

doi:10.1038/nature09652

Nascent transcript sequencing visualizes transcription at nucleotide resolution

L. Stirling Churchman[1] & Jonathan S. Weissman[1]

Recent studies of transcription have revealed a level of complexity not previously appreciated even a few years ago, both in the intricate use of post-initiation control and the mass production of rapidly degraded transcripts. Dissection of these pathways requires strategies for precisely following transcripts as they are being produced. Here we present an approach (native elongating transcript sequencing, NET-seq), based on deep sequencing of 3′ ends of nascent transcripts associated with RNA polymerase, to monitor transcription at nucleotide resolution. Application of NET-seq in *Saccharomyces cerevisiae* reveals that although promoters are generally capable of divergent transcription, the Rpd3S deacetylation complex enforces strong directionality to most promoters by suppressing antisense transcript initiation. Our studies also reveal pervasive polymerase pausing and backtracking throughout the body of transcripts. Average pause density shows prominent peaks at each of the first four nucleosomes, with the peak location occurring in good agreement with *in vitro* biophysical measurements. Thus, nucleosome-induced pausing represents a major barrier to transcriptional elongation *in vivo*.

Andres Veidenberg
Bioinfo JC, 7.02.2011

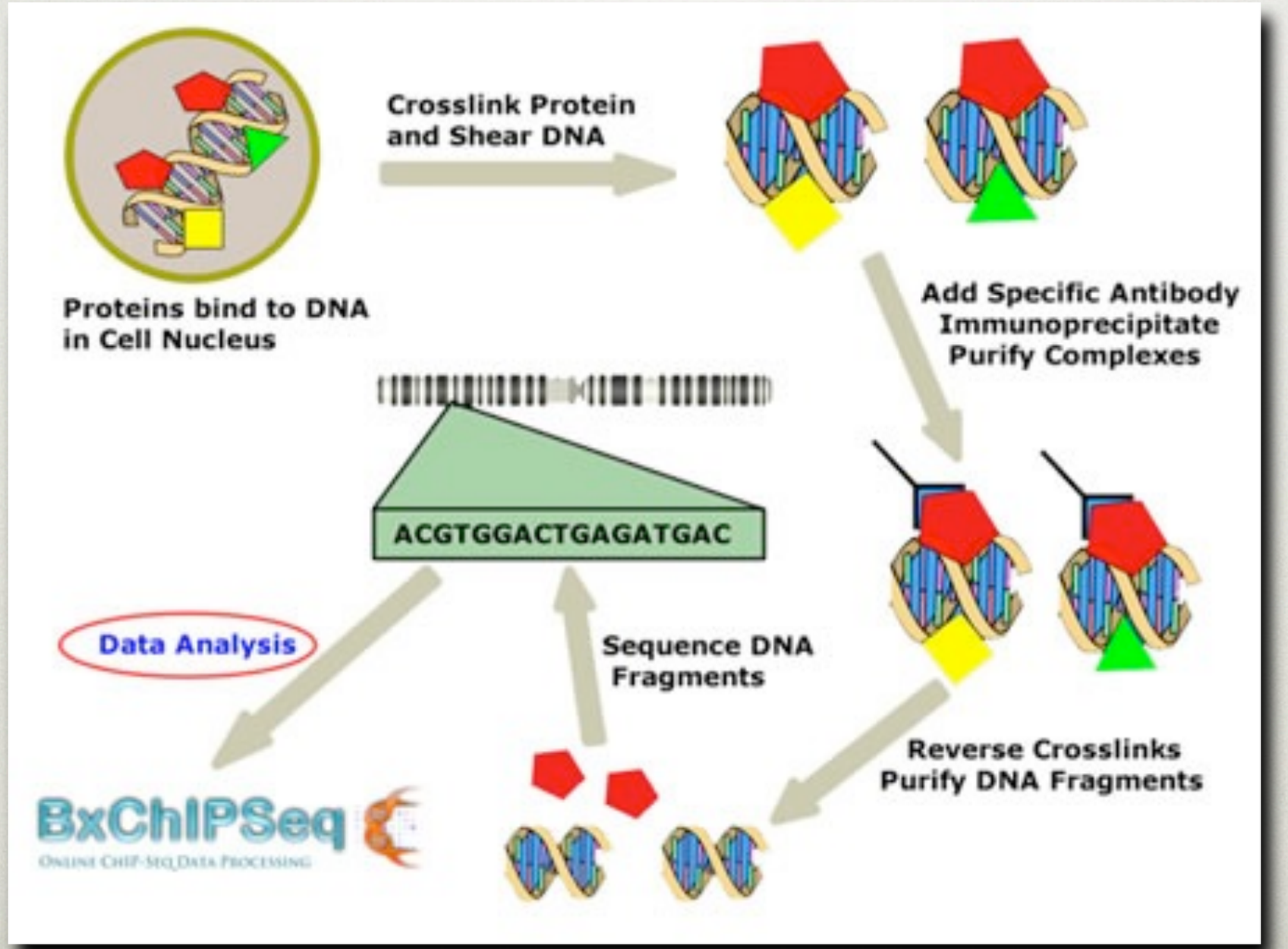
Problem

- Co-translational events
  - Rapid transcript degradation
- RNAP ubiquitous pausing
  - *in-vivo* mechanism largely unknown

NET-Seq

- Native elongating transcript sequencing
  - Monitors transcription at bp level
- Better at taking transcription snapshots
  - Doesn't use crosslinking (ChIP-Seq)
  - Doesn't manipulate physiological conditions

Chip-Seq


Crosslink Protein
and Shear DNA
Proteins bind to DNA
in Cell Nucleus
Add Specific Antibody
Immunoprecipitate
Purify Complexes
ACGTGGACTGAGATGAC
Data Analysis
Sequence DNA
Fragments
Reverse Crosslinks
Purify DNA Fragments
BxChIPSeq
Online ChIP-Seq Data Processing

Transcription quantification

- Flagged RNAPII purified via IP
- cDNA 3′ end deep-sequenced on Illumina GA2
- Alignment on S. Cerevisiae genome

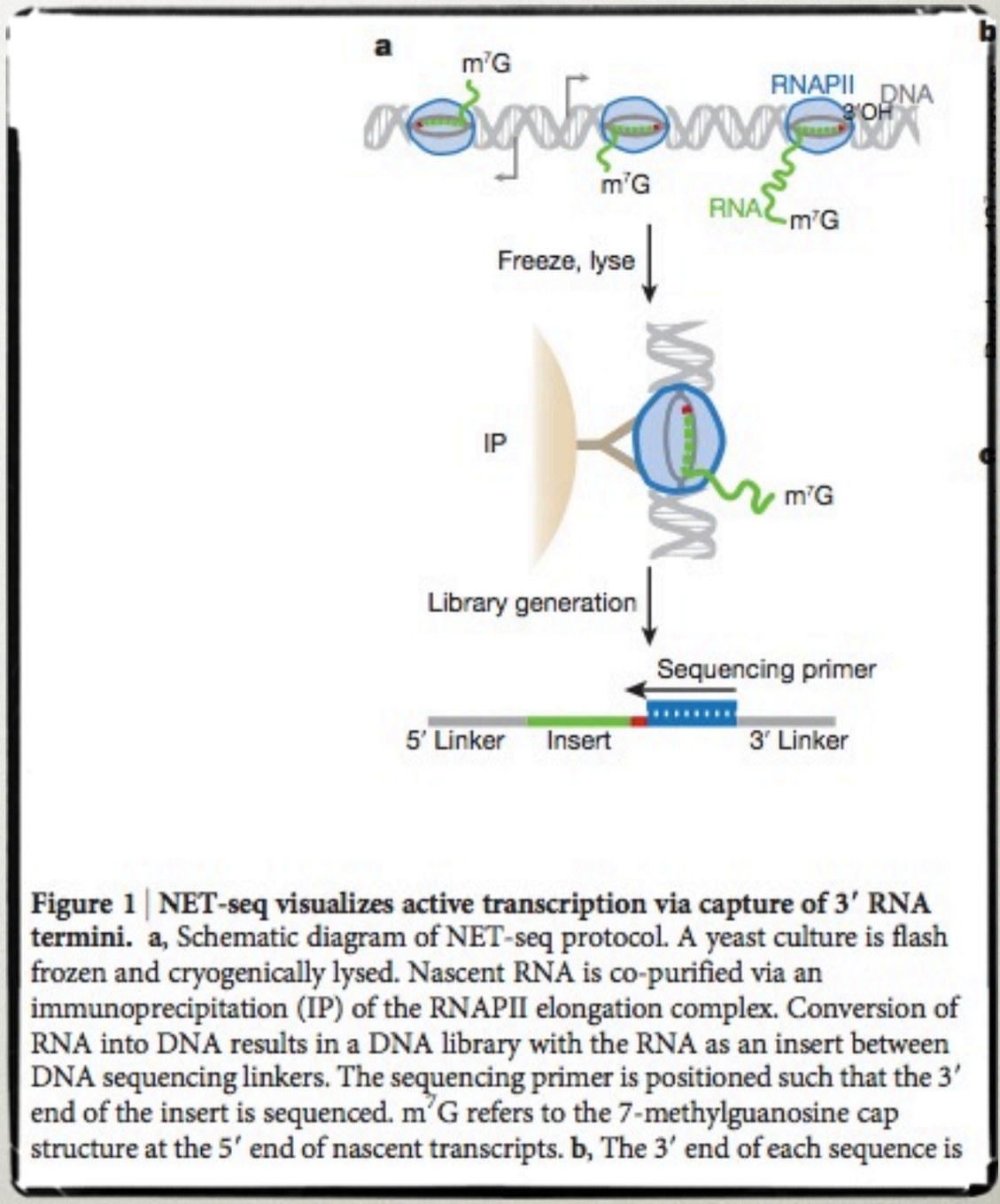


Figure 1 | NET-seq visualizes active transcription via capture of 3′ RNA termini. **a**, Schematic diagram of NET-seq protocol. A yeast culture is flash frozen and cryogenically lysed. Nascent RNA is co-purified via an immunoprecipitation (IP) of the RNAPII elongation complex. Conversion of RNA into DNA results in a DNA library with the RNA as an insert between DNA sequencing linkers. The sequencing primer is positioned such that the 3′ end of the insert is sequenced. m^7G refers to the 7-methylguanosine cap structure at the 5′ end of nascent transcripts. **b**, The 3′ end of each sequence is

RESULTS

- Higher RNAPII density for the first 700bp from 5’ end
- Captured nascent RNAs and splicing intermediates
  - RNA quantification
  - wide-spread existence of co-translational splicing

Antisense RNA

- Divergent transcription yield both mature sense RNA and antisense CUTs
- Sense transcription mostly over 8x higher than antisense
  - No correlation with nucleosome-free regions
  - Strong correlation with H4 acetylation levels

Antisense RNA

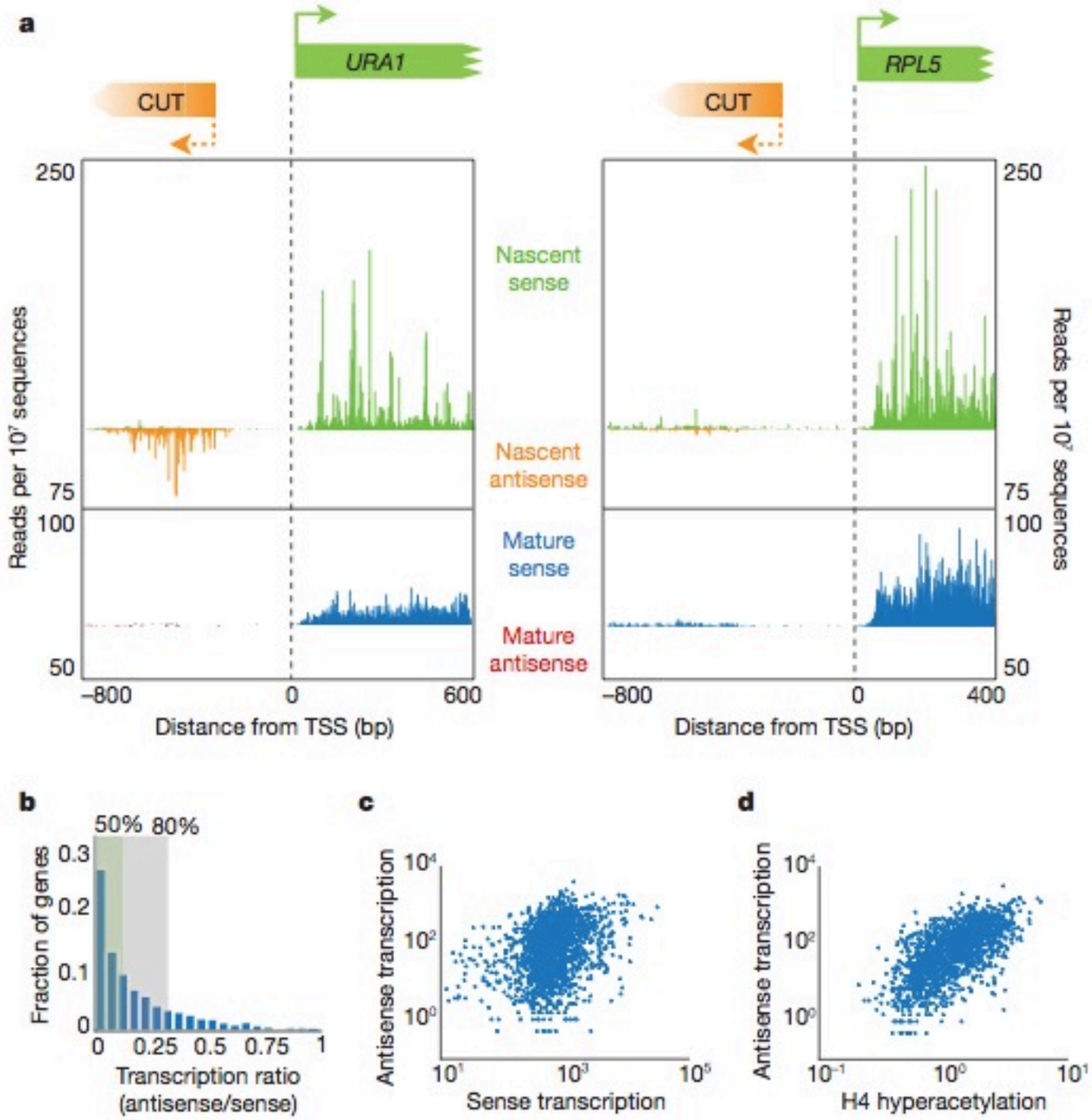


Figure 2 | Observation of divergent transcripts reveals strong directionality at most promoters. **a,** Nascent and mature transcripts initiating from *URA1* and *RPL5* promoters in the sense and antisense directions. Note that there are cryptic unstable transcripts (CUTs) in the antisense direction for *URA1* but not *RPL5*. **b,** A histogram of the transcription ratio (antisense/sense transcription levels) for 1,875 genes. The green and grey boxes indicate the subset of genes with a ratio of less than 1:8 and less than 1:3, respectively. **c,** Antisense transcription levels are plotted versus sense transcription for each tandem gene (Spearman correlation coefficient, $r_s = 0.34$). **d,** The level of antisense transcription for each promoter is plotted versus the local enrichment for H4 hyperacetylation using available data[24] ($r_s = 0.65$).

Transcription directionality

- Causative H4 acteylation
- Rpd3S deacetylase complex enforces promoter directionality
- Set2 recruitment to elongating RNAPII

Transcription pausing

- Developed an algorithm to find RNAPII oause positions
- Pauses occur frequenlty throughout the body of RNA messages
- Backtracking -> Dst1-stimulated cleavage

Backtracking

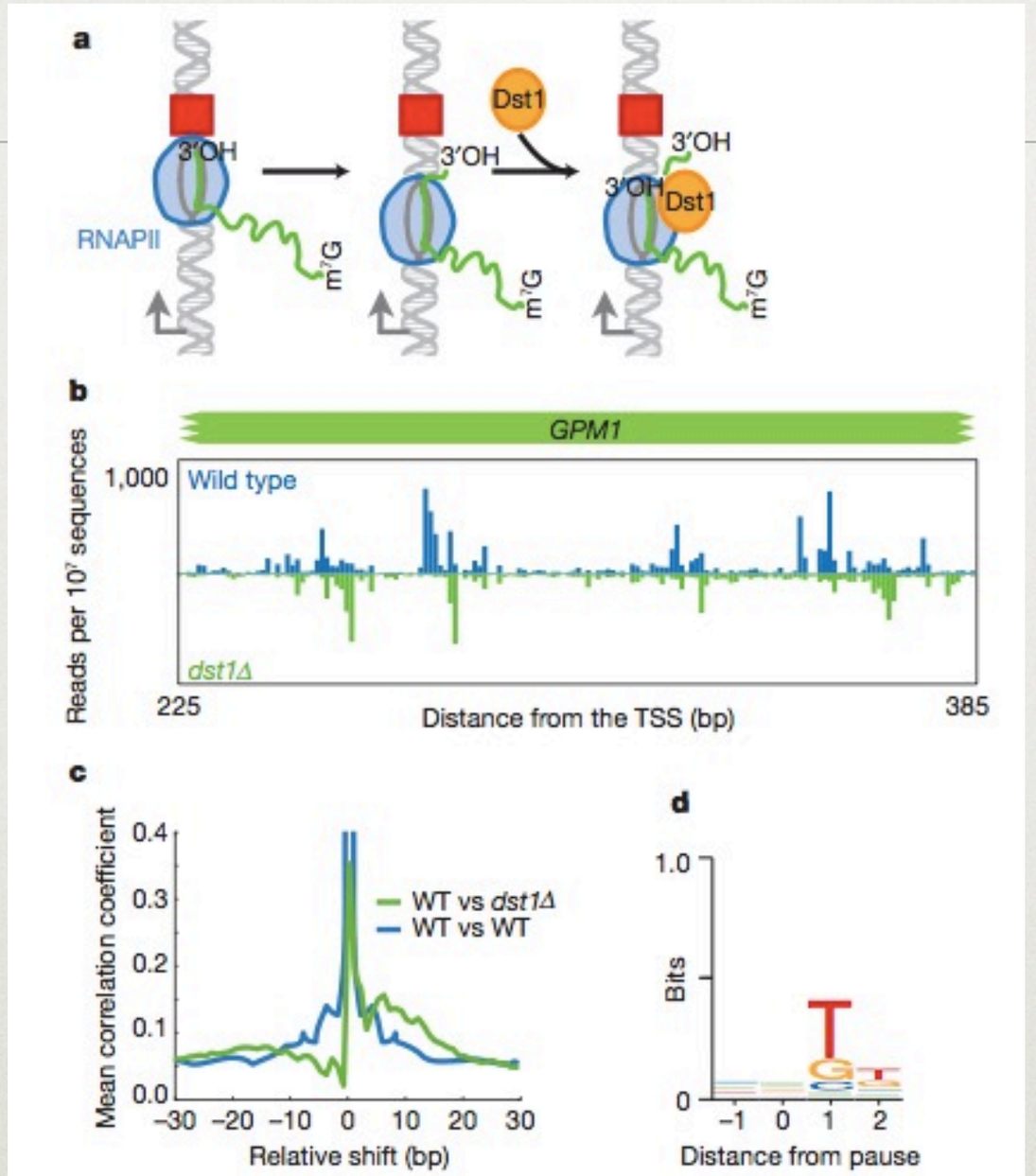


Figure 5 | Dst1 relieves RNAPII pausing after backtracking. **a**, A schematic describing an existing model for how RNAPII pauses at an obstacle (red square), backtracks and is induced to cleave its transcript through binding to Dst1 (refs 32, 33). **b**, A comparison of NET-seq data for wild-type and *dst1*Δ strains at the *GPM1* gene. **c**, Mean cross-correlation between the *dst1*Δ and wild-type data of well transcribed genes ($n = 770$, >2 reads per bp) (green line) was calculated by determining the Pearson's correlation coefficient at each gene between fixed *dst1*Δ data and shifted wild-type data followed by averaging over all genes. This analysis is compared to the mean autocorrelation of the wild-type data for well transcribed genes (blue line). **d**, The consensus sequence for all pauses observed in the *dst1*Δ strain.

Nucleosome barriers

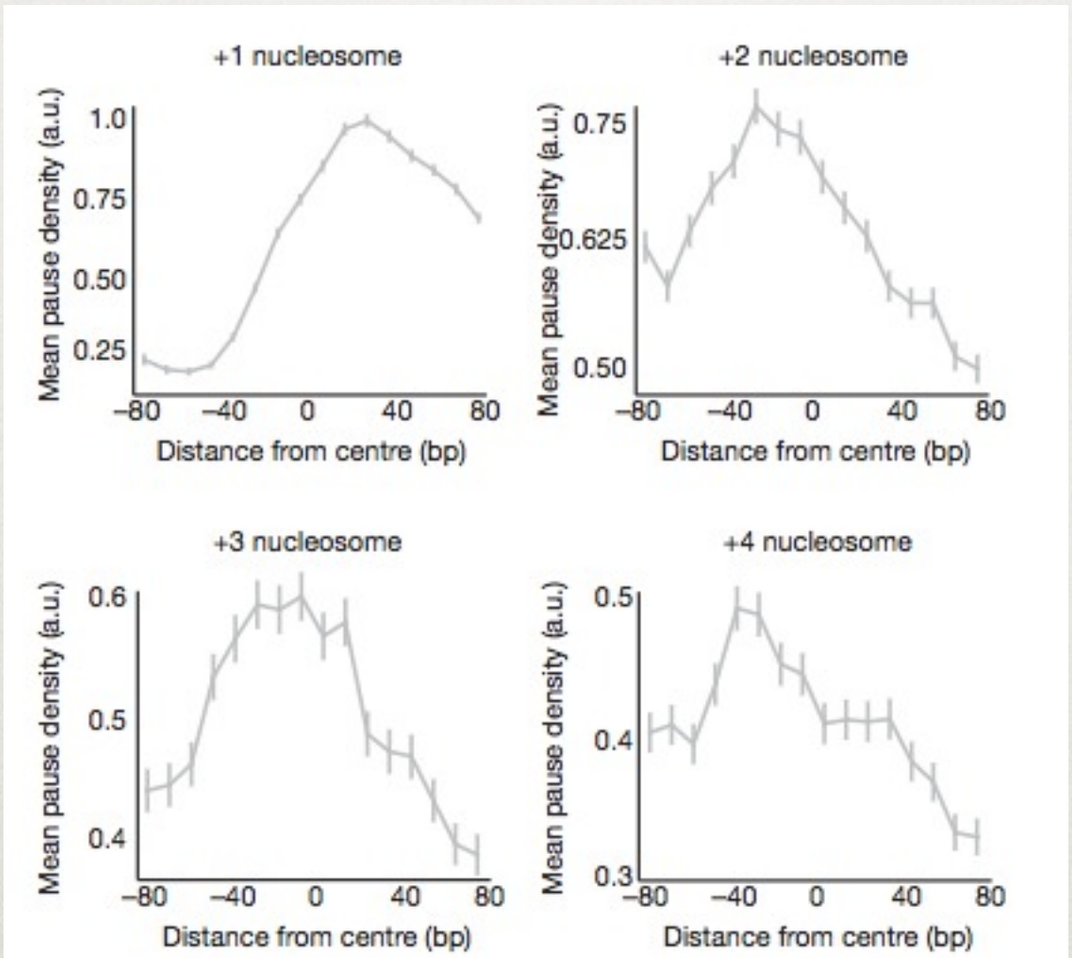
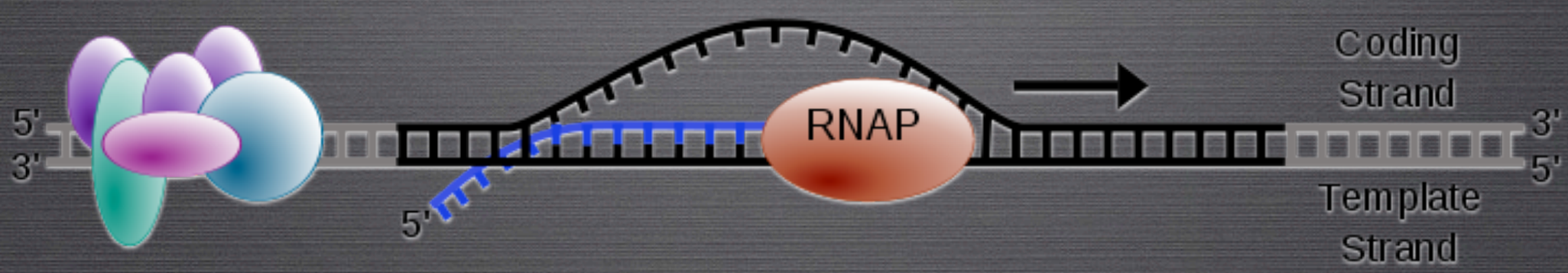


Figure 6 | Nucleosomes are a major barrier to transcription. Plot of mean pause densities in *dst1Δ* data relative to the first four nucleosomes after the transcription start site using available nucleosome positioning data[23]. Error bars represent one standard deviation.

Prespective

- Widespread obesrvation of divergent transcription
  - Most promoters show directionality
  - 2 mechanisms for antisense supression
    - Independence between sense and antisense transcripts
- First in-depth view of pausing in eucar. cells
  - Pauses shown throughout all RNA messages
  - Nuclesosomes induce pausing *in vivo*



AITÄH!