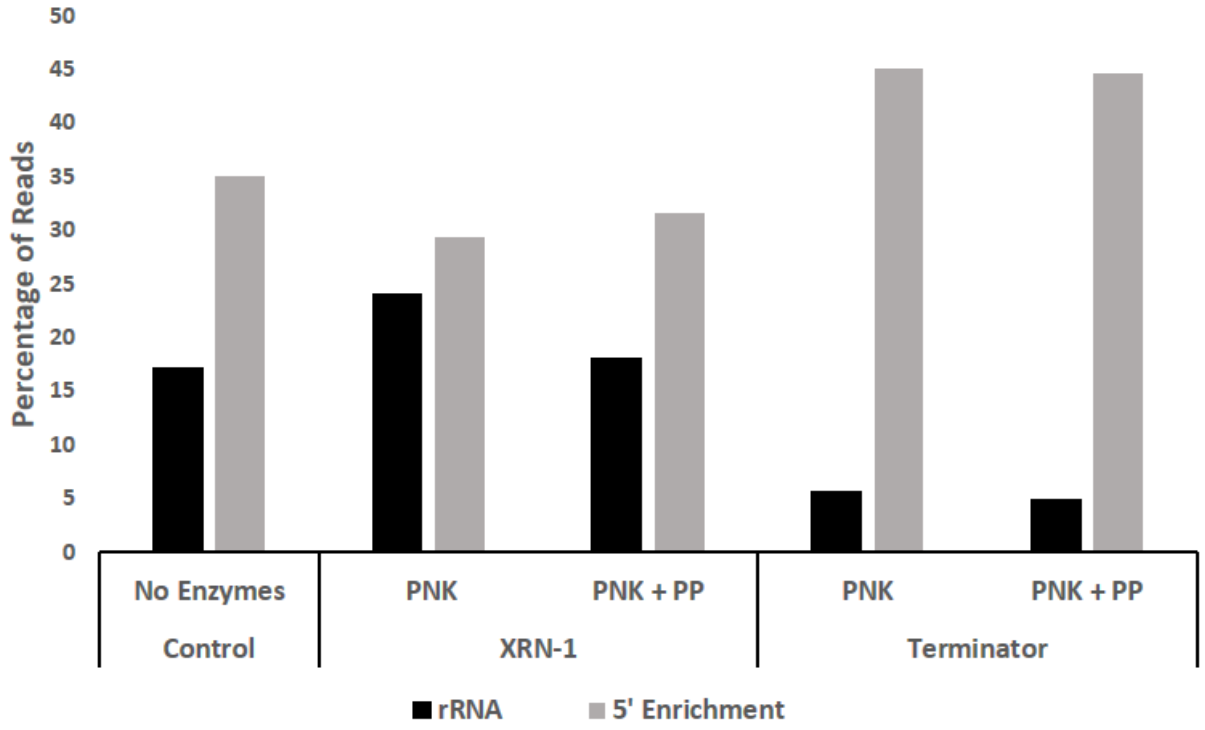
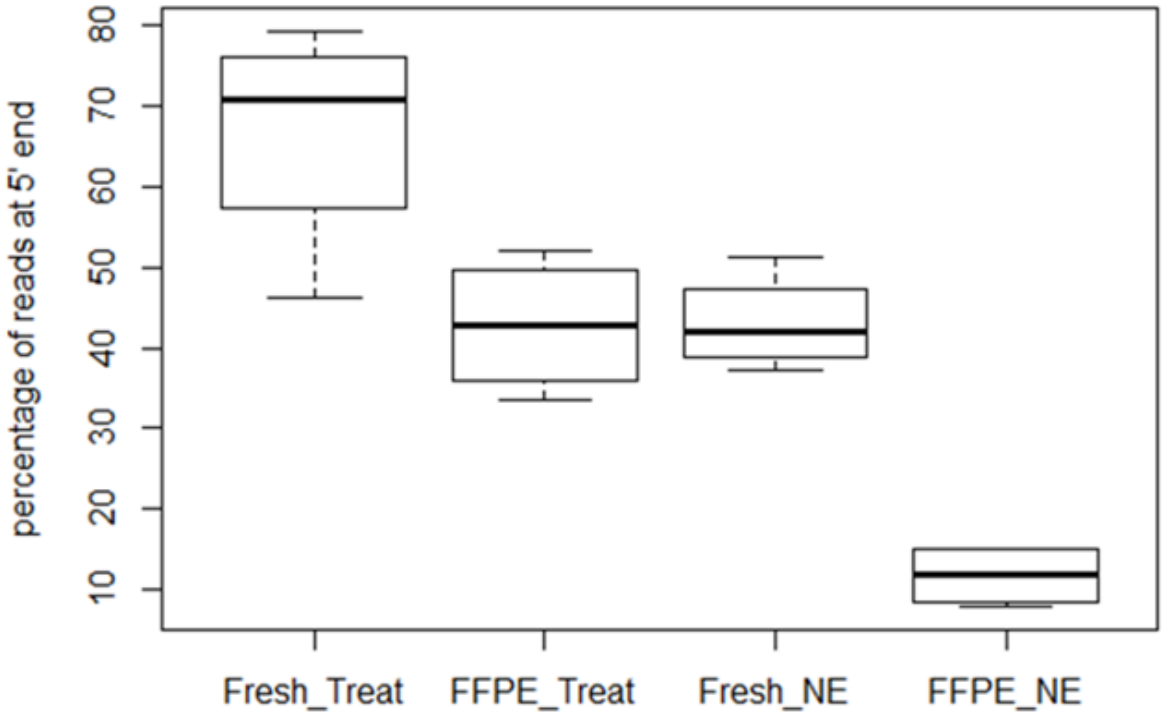
Supplemental materials for **FFPEcap-seq: a method for sequencing capped RNAs in formalin-fixed paraffin-embedded samples**

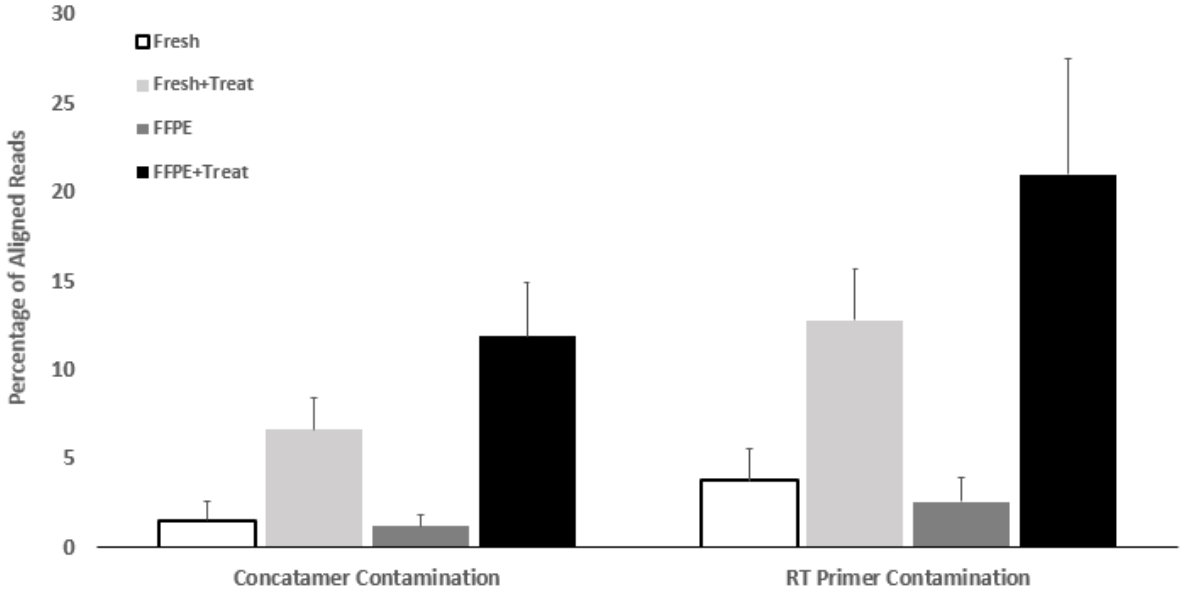
Jeffery M. Vahrenkamp, Kathryn Szczotka, Mark K. Dodson, Elke A. Jarboe, Andrew P. Soisson, and Jason Gertz



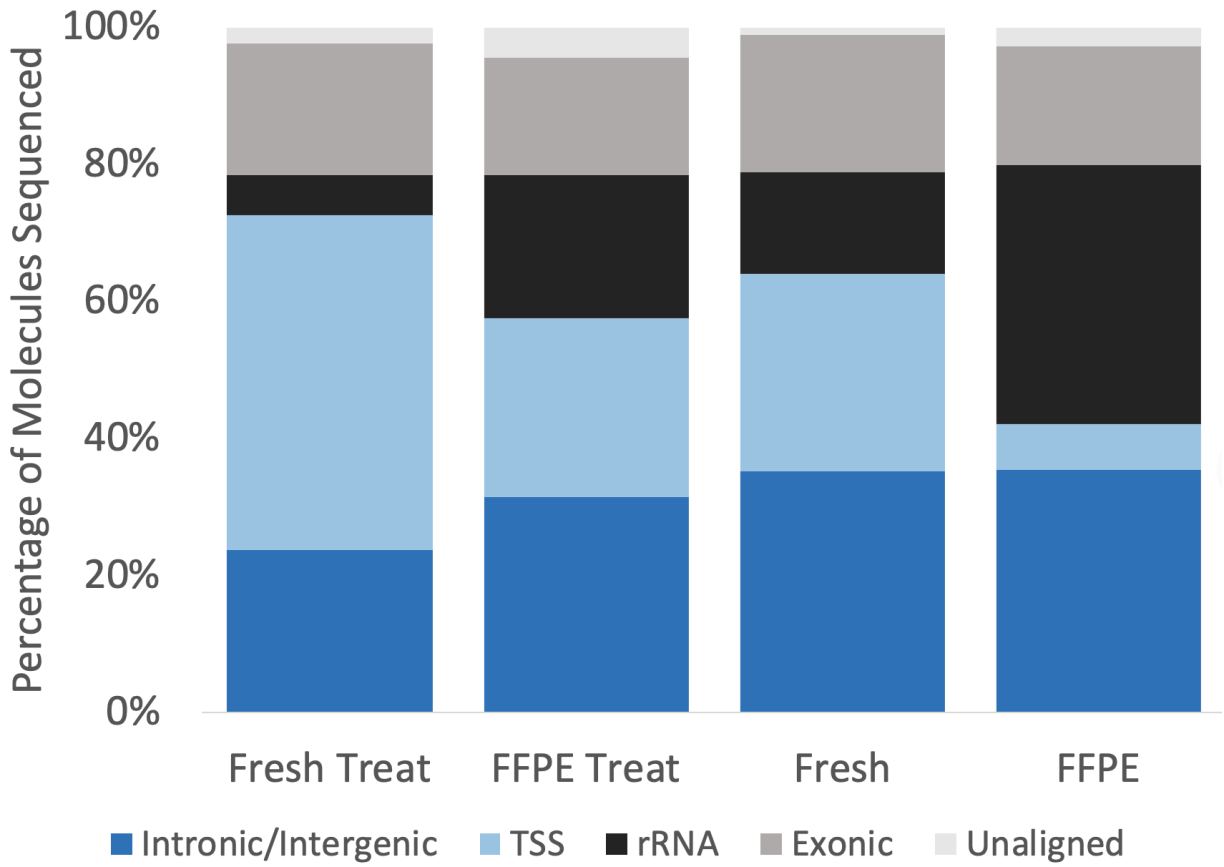
Supplemental Figure S1. Comparison of Enzymatic Treatments. Mock treated samples had rRNA contamination of 17% and 5' enrichment of 35%. Samples treated with XRN-1 and Polynucleotide kinase (PNK) or XRN-1, PNK and polyphosphotase (PP) did not show any decrease in rRNA or increase in 5' enrichment. Samples treated with Terminator and PNK or Terminator, PNK, and polyphosphotase had a decrease in rRNA and an increase in 5' enrichment. Polyphosphotase did not have a significant effect on Terminator treated samples.



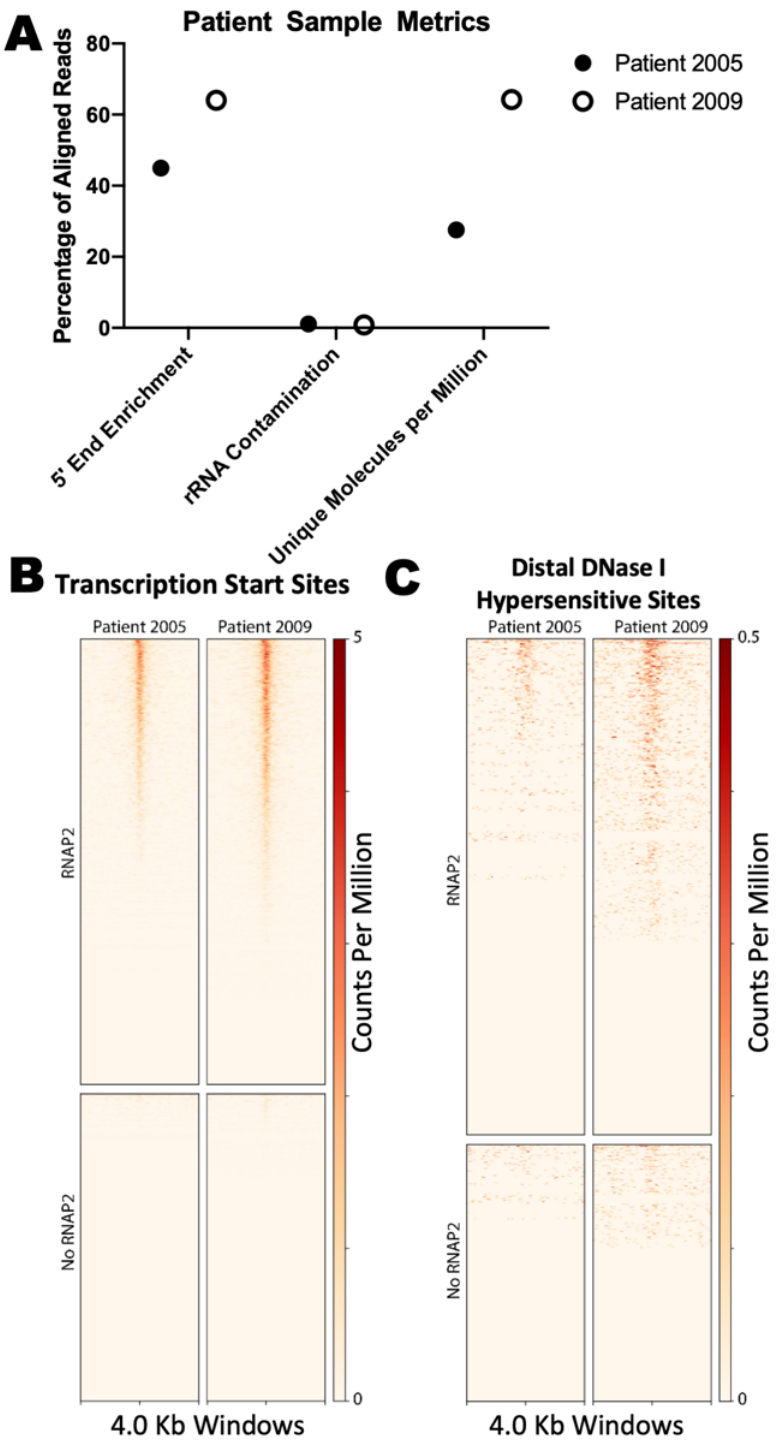
Supplemental Figure S2. Distribution of 5' Enrichment by Sample. Box plots show the distribution of 5' enrichment, as defined as the percent of reads within the 5' most 10% of mRNAs. Dark line represents median, boxes represent 25th and 75th percentiles and whiskers represent 5th and 95th percentiles.



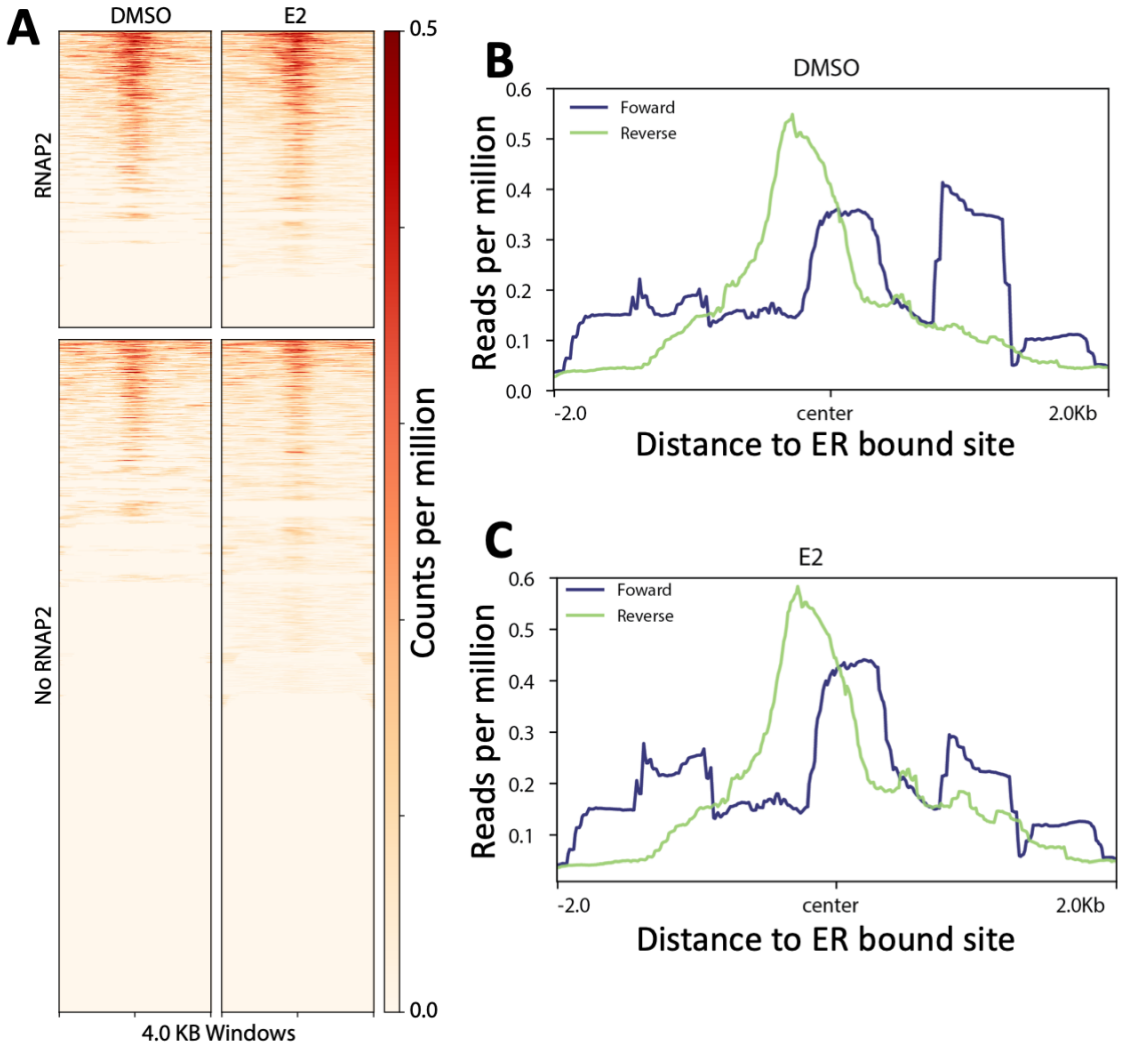
Supplemental Figure S3. Effects of Treatment on Primer Contamination. Concatemer contamination is from template switching oligos being added consecutively during reverse transcription. RT Primer contamination is due to the Revere Transcription primer hybridizing to the template switching primer without any RNA captured between them. Both contaminants increase as the number of target RNAs decrease due treatment or low RNA input.



Supplemental Figure S4. The distribution of molecules across genomic regions changes with the source of RNA and enzymatic pre-treatment. Bar graphs show the breakdown of read mapping for freshly derived and FFPE derived RNA with and without enzymatic pre-treatment. TSS = transcription start sites.



Supplemental Figure S5. Quality metrics and evidence of eRNAs in patient samples. A) Quality metrics are shown for FFPEcap-seq libraries made with RNA derived from two patient samples: a 14 year old sample (2005) and a 10 year old sample (2009). Heatmaps show FFPEcap-seq signal from patient-derived RNA at transcription start sites (B) and promoter-distal DNaseI hypersensitive sites (C).



Supplemental Figure S6. Detection of eRNAs after a 17β-estradiol (E2) induction. A) Heatmap shows the FFPEcap-seq signal at ESR1 bound sites after a 1 hour DMSO or E2 induction. ESR1 binding sites are separated into RNAP2 bound (top panels) and not RNAP2 bound (bottom panel). FFPEcap-seq signal is shown for positive and negative strand reads, at ESR1 bound sites, in cells treated with DMSO (B) or E2 (C). The data exhibits a strand shift consistent with eRNA production. Off center peaks are due to a few high signal loci. Data in figure panels A-C is a composite of two replicate FFPEcap-seq libraries for each condition.

Supplemental Table S1. Sequences of oligonucleotides used in FFPE Cap-seq

Oligonucleotide name	Sequence
Template Switching Oligo	A/iSp18/TAGTCGAACTGAAGGTCTCCAGCNNNNNNNNNArGrGrG
Iso Template Switching Oligo	/5isodG//iisodC/ATAGTCGAACTGAAGGTCTCCAGCNNNNNNNNNArGrGrG
Nanocage_RT_primer	TAGTCGAACTGAAGGTCTCCGAACCGCTCTTCCGATCTNNNNNN
Nanocage_2ndStrand_F	TAGTCGAACTGAAGGTCTCCAGC
Nanocage_2ndStrand_R	TGACGTCGTCTAGTCGAACTGAAGGTCTCCGAACC
Final Library primer F	AATGATACGGCGACCACCGAGATCTACACTAGTCGAACTGAAGGTCTCCAGC
Final Library Primer R	CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAACCGCTCTTCCGATCT
Custom Sequencing Primer	TAGTCGAACTGAAGGTCTCCAGC

/5isodG/ designates a iso 5' G base
/iisodG/ designates an internal iso C base
/iSp18/ designates an internal 18 carbon spacer
N designates an equal mixing of all bases.
r designates a ribonucleotide base instead of a dioxyribonucleotide base