He et al. Supplementary Online Material

Contents

Supplementary Materials and Methods

Samples and RNA preparation

DNA-free total RNA from Jurkat T cell-leukemia line and MRC5 diploid lung cell line were purchased from Ambion (Austin, Tx). The colorectal cancer cell line HCT116 and pancreatic cancer cell line MiaPaCa2 were purchased from ATCC (Manassas, VA) and were grown in McCoy's 5A medium with 10% fetal calf serum. Normal human peripheral blood mononuclear cells (PBMC) from a healthy volunteer were isolated from fresh peripheral blood with Histopaque-1077 (Sigma, St. Louis, MO). Total RNA from PBMC, HCT116 and MiaPaCa2 was isolated with the RNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The DNA was removed from the total RNA preparation with the *DNA-free* kit (Ambion).

Total RNA was enriched for the non-ribosomal fraction by treating 20-24 ug of DNA-free total RNA with the RiboMinus transcriptome isolation Kit (Invitrogen, Carlsbad, CA). To remove as much ribosomal RNA (rRNA) as possible, two rounds of rRNA reduction were performed. Subsequently, the RNA was ethanol-precipitated, washed in 70% ethanol, re-suspended in RNase-free water, and finally eluted in RNase-free water from a RNeasy column. To confirm that each RNA sample was DNA-free and to evaluate the effectiveness of rRNA removal, RT-PCR was performed with reagents from SuperScript Vilo cDNA synthesis Kit (Invitrogen) according to the manufacturer's recommendations. The efficiency of RNA removal was evaluated by qPCR using primers that resulted in the amplification of either 18S ribosomal RNA or specific mRNA species. We found that ~90% of 18S and 28S rRNA was successfully removed by the procedure described above.

Bisulfite conversion of RNA

RNA enriched for the rRNA-depleted fraction was treated with bisulfite according to a modified protocol based on the EpiTect Bisulfite Kit (Qiagen). Twenty ul of rRNA-depleted RNA (corresponding to 12-15 ug total RNA) was mixed with 85 ul of Bisulfite Mix and 35 ul of Protect Buffer. The conversion of C to U was performed by heating the solution under the following conditions: 99^{o}C, 5min; 60^{o}C, 25min; 99^{o}C, 5min; 60^{o}C, 85min; 99^{o}C, 5min; 60^{o}C, 155 min. Following completion of the reaction, the converted RNA was desalted by following steps: (i) the mix was diluted with RNase-free water to a total volume of 15 ml and concentrated to 0.67 ml using Centriprep Ultracel YM-3 15ml filter (Millipore, Bedford, MA) by three sequential centrifugations at 3000 g, room temperature for 95 min, 30 min and 10 min; (ii) the RNA was ethanol-precipitated, washed in 70% ethanol and re-suspended in 15 ul RNase-free water; (iii) the recovered RNA was mixed with 300 ul of 0.5 M Tris-HCl (pH 9.0) and incubated at 37^{o}C for 60 min for de-sulfonation (SOM ref *1*); and finally, (iv) the treated RNA was ethanol-precipitated, washed in 70% ethanol, and resuspended in 30 ul RNase-free water.

Library preparation for sequencing by synthesis

Double stranded cDNA was synthesized from the converted RNA using regular random hexamer (for RNA-seq library construction) or random octamers comprised of only three bases (A, C, T) (for ASSAGE library construction) and Supersctipt III reverse transcriptase (Invitrogen), following the manufacturer's recommendations for preparation of double-stranded cDNA. Specifically, for ASSAGE library construction: (i) we used a random octamer mix of $(H)_7A$: $(H)_7C$: $(H)_7T$, (where H = A,C, or T) at 2:1:1 molar ratio as primer for first-strand cDNA synthesis: 30 ul of converted RNA was mixed with 4.5ul of water and 1.5

ul containing 400 ng of random octamer mixture; the mixture was incubated at 70^{o} C for 10 min and cooled in ice for 3 min; (ii) the denatured RNA/oligo was mixed with 12 ul of 5 x first strand buffer, 6 ul of 0.1M DTT, 3 ul of 10 mM dNTP; 0.3 ul of 100mM dATP and 1.5 ul of RNaseOut; (iii) the mixture was incubated at 45^{o}C for 2 min before 3 ul of Supersctipt III retro-transcriptase was added; (iv) the mixture was incubated at 45^{o}C for 1 hour before being cooled at ice for 3 min; (v) the first-stranded cDNA was mixed with 61.5 ul of water, 15 ul of GEX 2nd strand buffer (Illumina), 4.5 ul of 10 mM dNTP, 0.45 ul of 100mM dTTP, 6 ul of DNA polymerase I (Illumina), 1.5 ul of RNaseH, and 1.5 ul of E. coli DNA ligase (Invitrogen); (vi) total of 150 ul mixture was incubated at 16^{o}C for 2 hours; (vii) 3 ul of T4 DNA polymerase was added to the mix and the reaction was incubated at 16^{o}C for 5 min to complete the double-stranded cDNA synthesis. Double-stranded cDNA was cleaned up by QIAquick PCR purification column (Qiagen) and eluted in 70 ul of elution buffer (EB).

The double-stranded cDNA was used to construct libraries for sequencing following Illumina's standard genomic DNA sample preparation instructions. Briefly, this procedure consists of six steps: (i) double-stranded cDNA in 70 ul of EB was mixed with 700 ul of nebulization buffer and fragmented by nebulization; (ii) the ends of the fragmented double-stranded cDNA were blunt-ended with T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase; (iii) a dA was added to the 3' and of each strand by Klenow (exo -) polymerase; (iv) adapters designed for library construction from genomic DNA available from Illumina were ligated to the cDNA fragments; (v) ligation product was gel-purified to select for ~120-200bp fragments; and (vi) PCR amplification to enrich ligated fragments. For sequencing, the library was denatured with 0.1 M NaOH to generate single-stranded DNA molecules, captured on Illumina flow cells, amplified in situ and subsequently the ends of the fragments were sequenced for 36 cycles on an Illumina Genome Analyzer. A similar library (RNA-seq library) made from the rRNA-depleted RNA of PBMC in the absence of bisulfite treatment was used as control to determine bisulfite-conversion efficiency and the potential effect of bisulfite on the sense transcriptome (see main text).

For paired-end (PE) sequencing, PE-ASSAGE libraries were constructed using Illumina PE-genomic DNA library kit, following the protocol for ASSAGE library described above with modification. Specifically, PE adaptors were used at the ligation step; and the pre-amplified library after the ligation was resolved by agarose gel to select fragments of ~300 bp in length (including adapters), which was then gel-purified for final library amplification. Sequencing of PE-ASSAGE library was performed following Illumina's Solexa protocol.

Efficiency of Bisulfite Conversion of C to U

First, the A, C, G, and T content of the transcriptome was determined through sequencing 3.27 million quality-controlled tags (118 million bases) generated from PBMC RNA in the absence of bisulfite treatment. These tags had high chastity scores according to Illumina criteria and matched the human genome. The C/G and A/T contents of these tags were 48% and 52%, respectively. Next, random subsets of 50,000 of these 36-base sequences were used to determine the fraction of 36-base tags that should contain Gs but no Cs after various conversion efficiencies using Monte Carlo simulations. The fraction of tags that contain Gs but do not contain Cs provides a measure of conversion efficiency. From these simulations, a 95% conversion rate (± 1%) was determined to be most consistent with the number of observed for 36-base tags that contained Gs but not Cs in the actual experimental libraries prepared from bisulfite-treated RNA.

Data Analysis

The sequences generated were aligned independently to the human genome (release hg18) and transcript databases (RefSeq release 35) using the Illumina alignment software Eland. The reference sequence for the alignment was modified in silico to simulate a genome in which all Cs had been converted to Ts by bisulfite treatment. Two versions of the human genome were generated: one in which all of the Cs were changed to Ts (representing the converted Plus-strand of DNA of each chromosome according to hg18) and the other in which all of the Gs were changed to As (representing the complement of the converted Minus-strand of each chromosome). We similarly created two converted versions of the RefSeq sequences. All experimentally-identified sequences were matched to both versions of the modified genome, both versions of the RefSeq sequences, and in some cases to the normal (unconverted) genome and RefSeq sequences. The alignment was performed with the Eland software using the eland_extended module which matches 32 bases and then explores the alignment of the remaining 4 bases of the 36 base tags. Although the matching is based on 32 bases, the sequences with 36 perfect matches receive better scores than those that match 32 to 35 bases. The following criteria were used to filter the tags for further analysis: Each tag was required (i) to pass the Illumina chastity filter; (ii) not to match to rRNAs or t-RNAs; and (iii) to match uniquely to one of the two converted versions of the genome with no more than two mismatches (thus defining a "unique" tag).

Analysis of tags and transcript mapping

Each tag was assigned to a gene from the Ensembl ensGene database (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/ensGene.txt.gz) by virtue of its unique genomic position. Promoter and terminator regions of the genes were defined as those sequences which mapped one kb upstream or one kb downstream from transcription 'start' and 'end' sites, respectively. The tags were classified as sense or antisense (Table S1). Sense tags were defined as those that matched genomic positions between the 'start' and 'end' sites of the same strand of a known transcript annotated in the Ensembl database or in it promotor or terminator as defined above. Antisense tags were defined as those that matched positions on the opposite strand of a known gene. Distinct tags that matched to regions of the genome that were known to produce transcripts from both strands were considered ambiguous and were excluded from further analysis (Table S1). Tags that could not be classified in any of the above categories were labeled as non-annotated in Table S1. Tag densities in exons, introns, and other genomic regions were compared using t-tests.

To estimate the theoretical percentage of converted tags that could be uniquely matched to the reference genome, two sets of virtual 36-base tags were generated in silico. The first set included 1.6 million tags covering the sequences of each strand of the transcripts that were used as examples in Figures 2, S3 and S5 as well as arbitrarily selected transcripts from elsewhere in the genome. The second set included 4.8 million tags representing the two strands from random areas of each chromosome (100,000 sequences from each strand of every chromosome). In each of these sets, all C's were changed to Ts to mimic the effects of bisulfite-conversion. We performed a similar in silico experiment to determine the fraction of tags that should map to the genome in the absence of bisulfite conversion. For this determination we used virtual tags derived from the set of 4.8 million tags described above but did not change the Cs to Ts. We found that 52% of the tags expected to be derived from ASSAGE could be assigned to a unique genomic

position. 71% of virtual tags generated in the absence of conversion could be uniquely assigned, with the remainder assigned to more than one genomic position. The proportion of converted tags that could not be uniquely matched was higher because of the three-base rather than four-base code, allowing a portion of closely-related gene sequences to produce identical virtual tags. Nevertheless, the process of converting a four-base to a three-base code only moderately reduced the tag information content and still allowed assignment of tags to unique genomic positions.

The procedure for matching PE-ASSAGE tags was performed similarly to that described above in the Data Analysis section. To assess potential transcript splicing, we required that the two tags (tag1 and tag2) representing each end of the cDNA fragment were unique and matched to the same gene and the same orientation (sense or antisense). The average distances separating tag1 and tag2 in sense transcripts was 169 and 182 bp in the Jurkat and MRC5 libraries, respectively, as determined by mapping to the human transcript (not genomic) database. Analysis of the distribution of PE tags with respect to the genomic distances separating tag1 and tag2 from individual cDNA fragments provided an estimate of the fraction of sense or antisense tags that resulted from splicing.

RNA microarray expression analysis

Fractions of the same RNA preparations used to construct ASSAGE libraries were used for the generation of cRNA for microarray analysis on Agilent chips. Briefly, total RNA was reverse transcribed by MMLV-RT using an oligo-dT primer that incorporated a T7 promoter sequence. The cDNA was then used as a template for in vitro transcription using T7 RNA polymerase and Cy-3-labeled CTP. Labeled cRNA samples were used for hybridization to Agilent 4x 44K microarrays and scanned using an Agilent Scanner. Microarray expression levels were compared to sense transcript tag levels for corresponding genes among all five samples analyzed. All genes that were expressed at a minimum level of at least five tags in any ASSAGE library were assessed. The average correlation coefficient between transcript profiles assessed by microarray and ASSAGE was 0.67. None of 100,000 simulations between microarray expression levels and ASSAGE expression levels of the same genes, randomly shuffled between the five analyzed samples, resulted in average correlation coefficients > 0.67 (the maximal average value observed in any of the 100,000 simulations was 0.28).

RT-PCR detection of sense and antisense transcripts

To validate sense and antisense transcripts identified by ASSAGE, total RNA (1-2 ug) conversion by bisulfite treatment and RNA cleanup was performed as described above. Converted RNA (750 to 1500 ng) was resuspended in ~15 ul of water. A standard first-strand cDNA synthesis was performed by using a random octamer $(H)_7$A/C/T mix and Supersctipt III reverse transcriptase. Primers used for RT-PCR were: for gene ENSG00000206028 antisense transcript: forward primer - GGTTTAAGGTAGGGGATGGTTT (matched to converted sequence of GGTCTAAGGCAGGGGATGGCCC at chr22: 25398333-25398354) and reverse primer – TCCACACTCACATCCCAAAA (matched to converted sequence of CTCTGGGACGTGAGTGTGGA at chr22: 25398479-25398498); for gene ENSG00000159496 antisense transcript: forward primer – TTTGGAAAGATGATTGTTGTGG (matched to converted sequence of TCCGGAAAGATGACCGTTGCGG at chr22: 22364583-22364604) and reverse primer – CAAAACCACACAAAAATACCCTAA (matched to converted sequence of

CCAGGGCACCCCTGCGTGGTTCTG at chr22: 22364443-22364466); for gene ENSG00000162576 sense transcript: pair#1 forward primer - TGGATGTGGGGTTGTATATTTG (matched to converted sequence of CGGACGCGGGGCTGTACACCTG at chr1: 1280555-1280576) and reverse primer – CACCACCAACACCTCCTTCT (matched to converted sequence of AGAAGGAGGTGCTGGCGGTG at chr1: 1280346-1280365); pair#2 forward primer: TTGGTTGTTTGTTTGGAGGTTA (matched to converted sequence of CTGGCCGTCCGCCTGGAGGTCA at chr1: 1280496-1280517) and reverse primer – CCAATCCCAATACACCACCT (matched to converted sequence of AGGTGGTGCACTGGGACCGG at chr1: 1280247-1280266); antisense transcript: pair#1 forward primer –TTGTGGGTGGTTAGGAGGAT (matched to converted sequence of CTGCGGGCGGCCAGGAGGAC at chr1: 1279608-1279627) and reverse primer – CCCACAAACCCCACACTAAC (matched to converted sequence of GCCAGTGTGGGGTCTGCGGG at chr1: 1279740-1279759); pair#2 forward primer: GGAGGAAGTTGTGTGTGAGTTTT (matched to converted sequence of GGAGGAAGCCGCGCGTGAGCCTT at chr1: 1280600-1280622) and reverse primer – CCTCAACCTTCAACAACAACAA (matched to converted sequence of TTGCCGTCGTCGAAGGCCGAGG at chr1: 1280713-1280734). Forward or reverse oligo in each pair of primer was attached with M13-forward sequencing primer – GTAAAACGACGGCCAGT to facilitate the sequencing of each RT-PCR product.

All primers were synthesized by Invitrogen. RT-PCR for converted RNA was performed in 20 ul reactions containing 1xPCR buffer (67mM Tris-HCT, pH 8.8, 6.7mM $MgCl_2$, 16.6 mM NH_4SO_4, 10mM 2-mercaptoethanol), 0.5 mM dNTPs, 0.5uM forward and 0.5uM reverse primers, 5% DMSO and 1 u Plantium Taq (Invitrogen), and cDNA generated from 0.4-4 ng of total RNA. PCR reactions were carried out using a touchdown PCR protocol (1 cycle of 94^{o}C for 2 min; 3 cycles of 94^{o}C for 10 sec, 64^{o}C for 15 sec, 70^{o}C for 15 sec; 3 cycles of 94^{o}C for 10 sec, 62^{o}C for 15 sec, 70^{o}C for 15 sec; 3 cycles of 94^{o}C for 10 sec, 60^{o}C for 15 sec, 70^{o}C for 15 sec; 41 (for gene ENSG00000206028 and ENSG00000159496 in PBMC) or 33 cycles (for gene ENSG00000162576 in MiaPaCa2 and MRC5) of 94^{o}C for 10 sec, 58^{o}C for 15 sec, 70^{o}C for 15 sec). RT-PCR products were gel-purified by QIAquick gel purification kit (Qiagen) and sequenced using conventional Sanger dideoxy terminators.

Analysis of antisense transcript splicing

To investigate if splicing occurred in the antisense transcripts, total RNA without conversion were first used for RT-PCR (selecting genes that had many antisense tags and no sense tags in the ASSAGE libraries). Briefly, reverse transcription was performed using SuperScript VILO cDNA synthesis kit (Invitrogen) and first strand cDNA was directly used for PCR. To investigate the splicing of antisense transcripts for gene ENSG00000162576 (in MiaPaCa2 cell line), the RT-PCR primers used were: ENSG00000162576 forward primer - GTGCAGGCCACAGTAATGGT (chr1: 1280028-1280047) and reverse primer - CTTCGACGACGGCAACTT (chr1: 1280708-1280727); RT-PCR products were sequenced to confirm the expected splice event. Second, bisulfate-treated RNA-derived RT-PCR was performed with two primer pairs that matched to converted antisense transcripts and each PCR product was sequenced to confirm that the splicing event was in the expected antisense strand. The primers used were: pair#1 forward primer – TGTGTGGTTGTGTGGGATTT (matched to converted sequence of CGCGCGGTCGTGCGGGACCC at chr1: 1280217-1280236) and reverse primer-

CCACCTCTACAAAAACCTAACCAT (matched to converted sequence of ACGGCCAGGCTCTCGTAGAGGTGG of at chr1: 1280510-1280533); pair#2 forward primer – TTTGGGTGGTTGTTGGTTTT (matched to converted sequence of CCCGGGCGGCTGCCGGTCCC at chr1: 1280235-1280254) and reverse primer- AAACTCACACACAACTTCCTCCT (matched to converted sequence of AGGAGGAAGCCGCGCGTGAGCCT of at chr1: 1280599-1280621).

Another approach to identify spliced antisense transcript was to examine the distances between tag1 and tag2 in the PE-ASSAGE antisense tags from Jurkat and MRC5 cells. Using a minimal distance of 600 bp cutoff, we identified 79 and 86 genes as potentially spliced antisense transcripts from Jurkat and MRC5 cells, respectively. We selected six genes (three from each cell line) for RT-PCR screening and four of them generated products; each of the products was ~200 bp in size. These four genes were ENSG00000157483, ENSG00000198624, ENSG00000105679 from Jurkat cells; and ENSG00000121454 from MRC5 cells. Primers used for these four genes' RT-PCR were: for gene ENSG00000157483's antisense transcript: forward primer – AGACCACAAGGAGGAGAAGC (chr15: 57238352-57238371) and reverse primer – GCTTTCTTCAGAATGGAACATTT (chr15: 57240413-57240435); for gene ENSG00000198624's antisense transcript: primer pair#1 forward primer- TTCCTGAGTCAACGGAAACTT (chr5: 150571883-150571903) and reverse primer – CTGTAGATGACACGCCAGCA (chr5: 150572535-150572554); primer pair#2 forward primer – TGCTGGCGTGTCATCTACA (chr5: 150572535-150572554) and reverse primer – CGGGGGTTAAAGGCTGATA (chr5: 150583287-150583306); for gene ENSG00000105679's antisense transcript: forward primer- GTTGCTGAAACAGCCAAGGT (chr19: 40725671-40725690) and reverse primer – CACAGTCACAGAGTCCACGTC (chr19: 40724701-40724721); for gene ENSG00000121454's antisense transcript: forward primer – CAGGCCAAGGAGAAAAACAA (chr1: 178509749-178509768) and reverse primer: CTAGAGGGCAGCCTCCTCTG (chr1: 178509024-178509043).

RT-PCR for gene ENSG00000162576 was performed in 30 ul reactions containing the same components described above. PCR reactions were carried out using a standard PCR protocol (1 cycle of 94^{o}C for 2 min; 40 cycles of 94^{o}C for 15 sec, 58^{o}C for 15 sec, 70^{o}C for 1 min). RT-PCR for gene ENSG00000157483, ENSG00000198624, ENSG00000105679 and ENSG00000121454 contained all components as described above plus 2 mM ATP. PCR reactions were carried out using a touchdown PCR protocol (1 cycle of 94^{o}C for 2 min; 3 cycles of 94^{o}C for 10 sec, 64^{o}C for 15 sec, 70^{o}C for 15 sec; 3 cycles of 94^{o}C for 10 sec, 62^{o}C for 15 sec, 70^{o}C for 15 sec; 3 cycles of 94^{o}C for 10 sec, 60^{o}C for 15 sec, 70^{o}C for 15 sec; 36 cycles of 94^{o}C for 10 sec, 58^{o}C for 15 sec, 70^{o}C for 15 sec). All RT-PCR products were sequenced to confirm the occurrence of splicing; comparison with genomic databases showed consensus splice sites at the expected positions of each spliced antisense cDNA fragment.

1. W. Gu, R. L. Hurto, A. K. Hopper, E. J. Grayhack, E. M. Phizicky, *Mol Cell Biol* **25**, 8191 (2005).

Figure S1 - Principles of ASSAGE

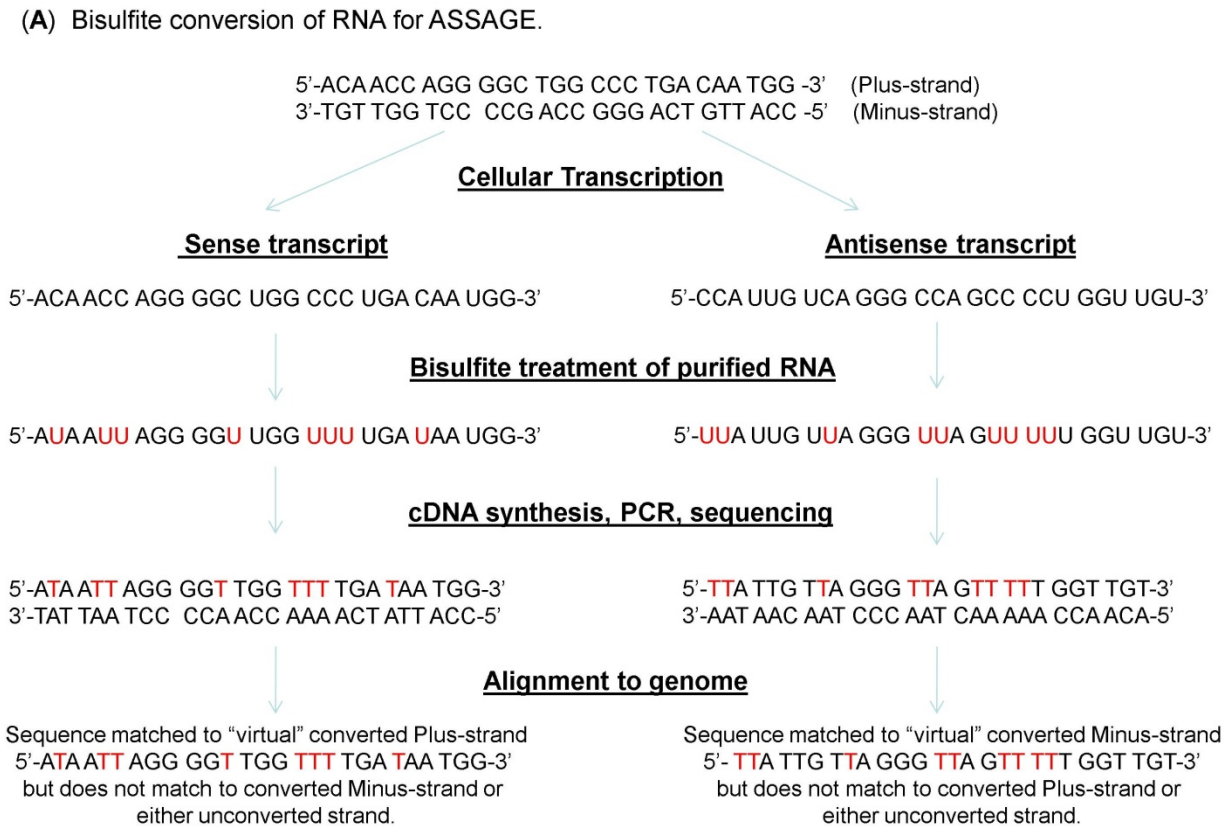


Figure S1 - continued

(B) ASSAGE to analyze sense and antisense transcripts across the genome.

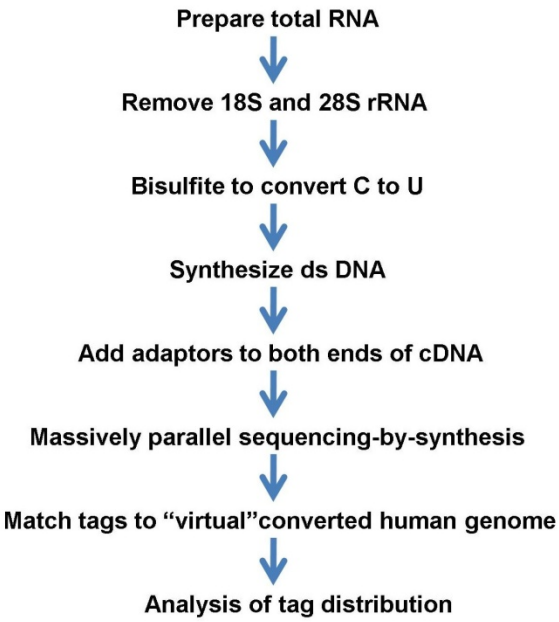


Figure S2 - Distribution of tags located in promoter and terminator regions of genes in the indicated samples.

(**A**) PBMC

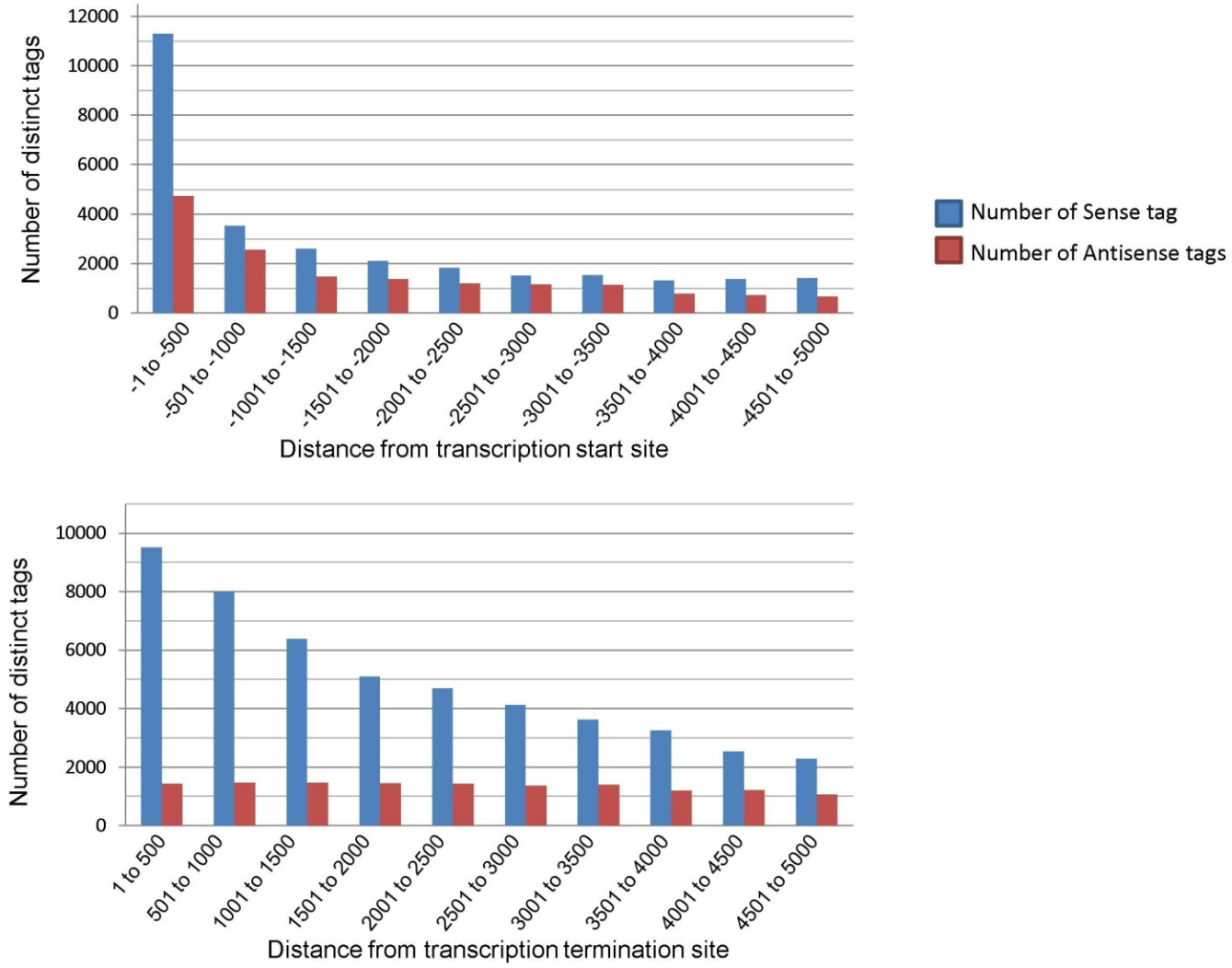
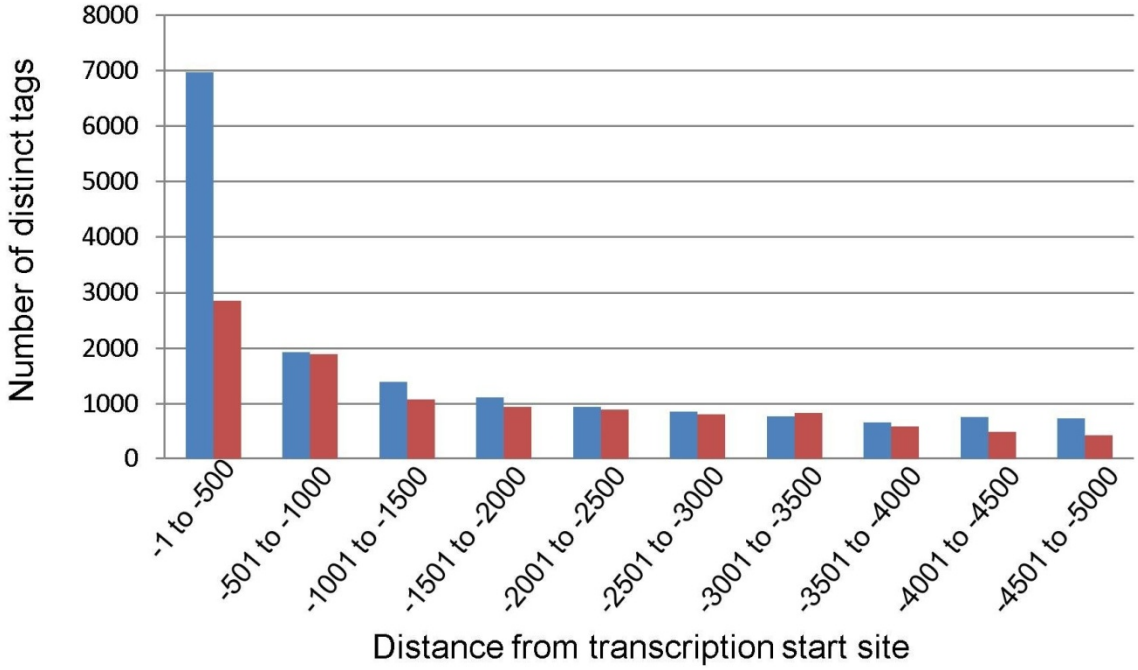


Figure S2 - continued

(**B**) Jurkat



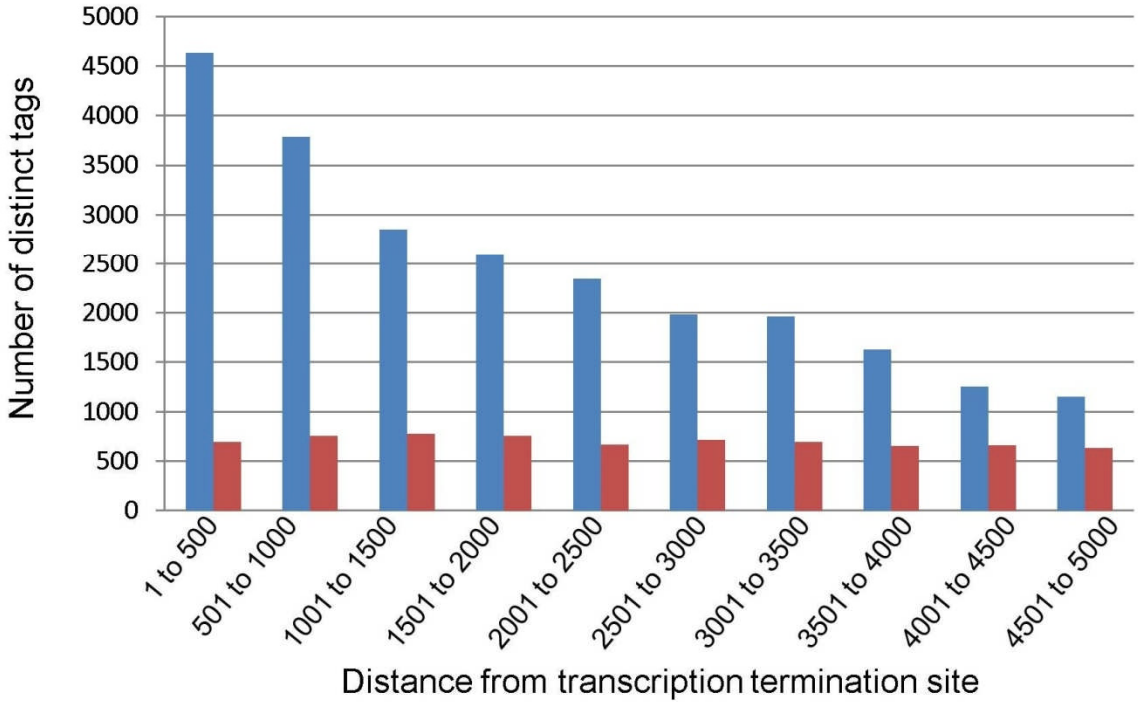


Figure S2 - continued

(**C**) HCT116

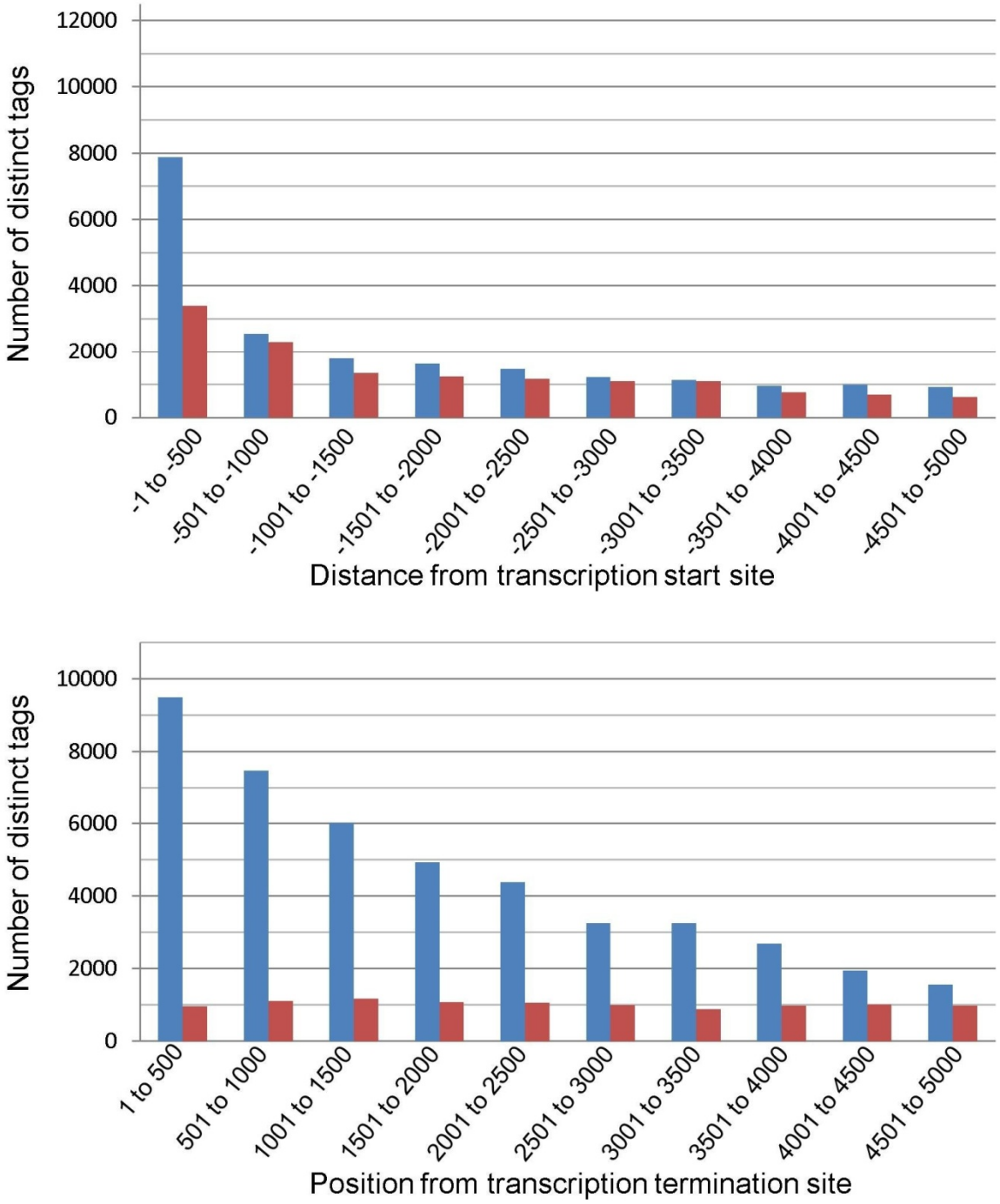
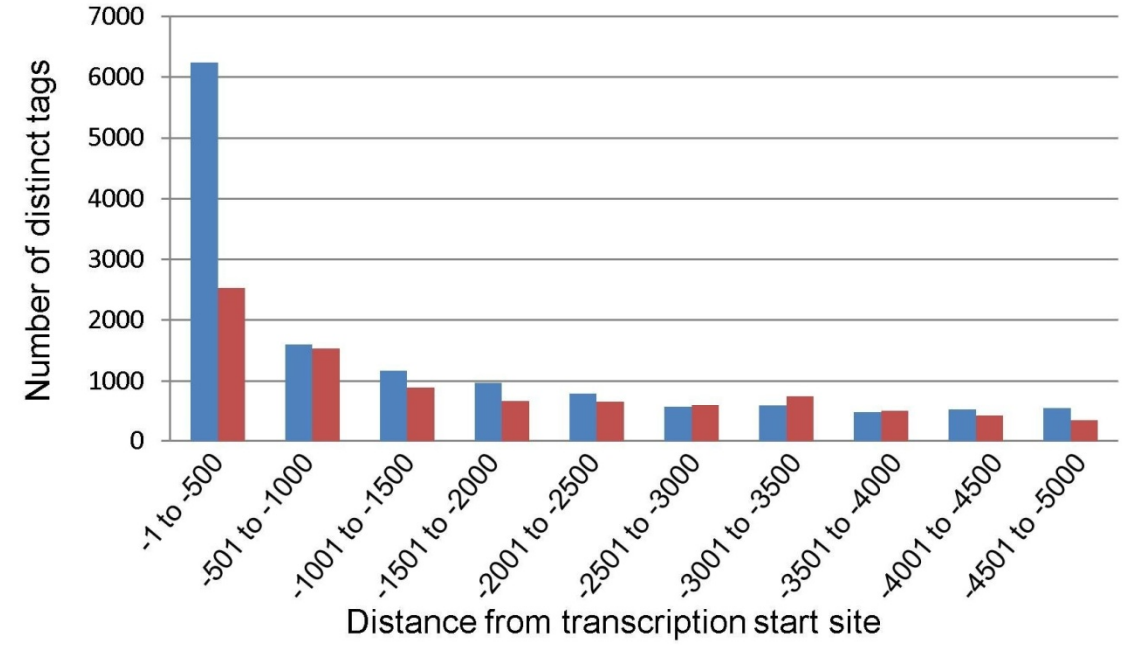


Figure S2 - continued

(**D**) MiaPaCa2



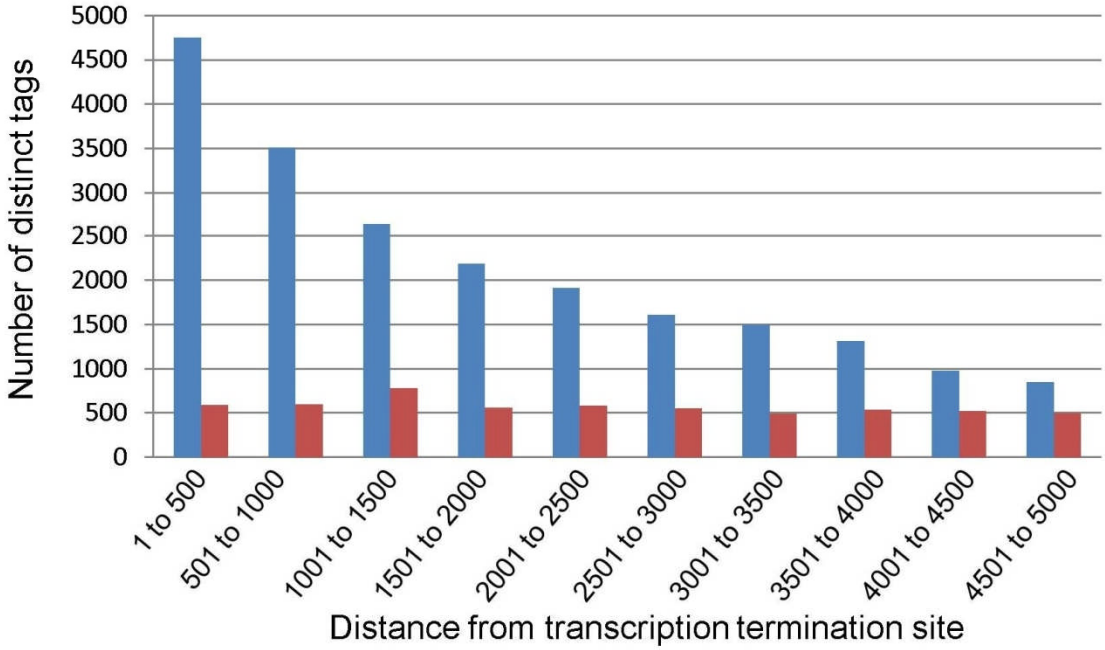
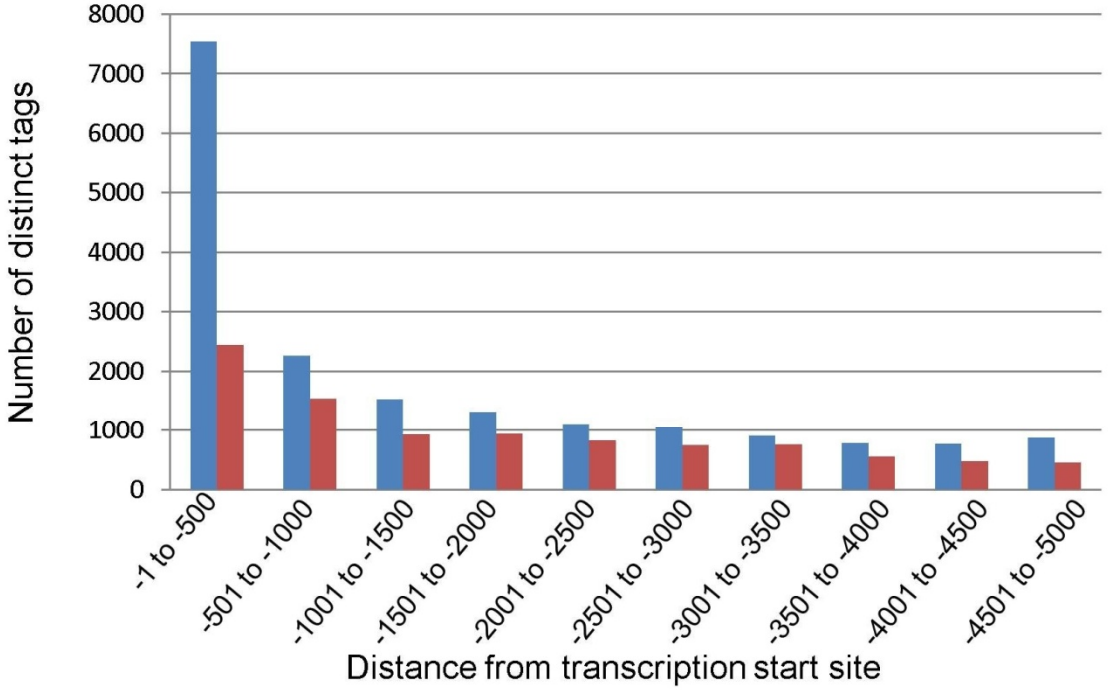


Figure S2 - continued

(E) MRC5



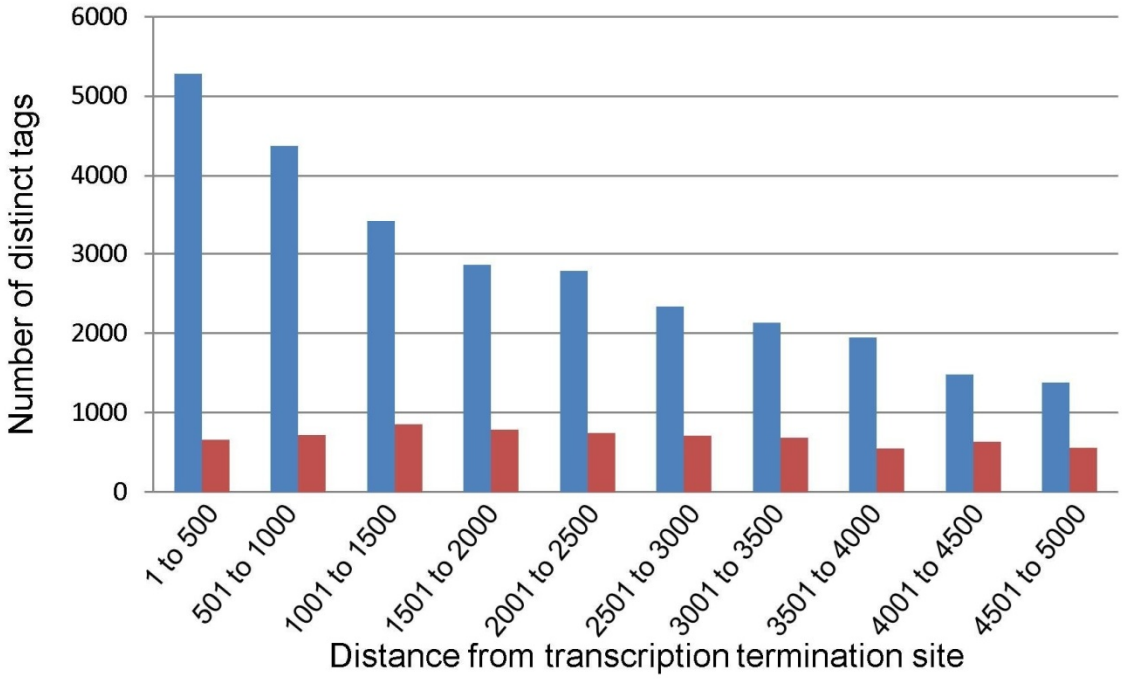


Figure S3 - Examples of tag distribution in a S gene (A), an AS gene (B) and a S gene with promotor AS tags (C) genes from PBMC.

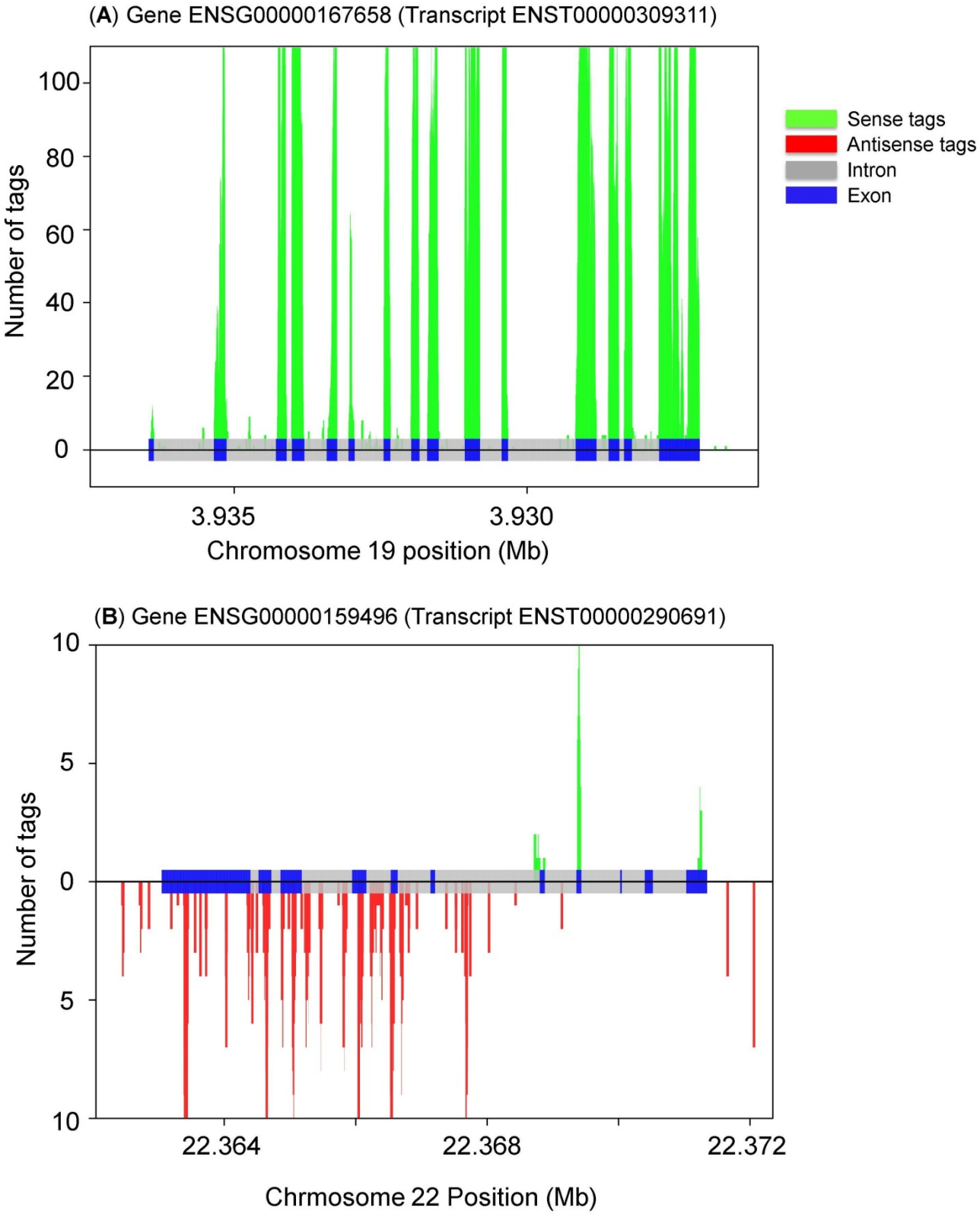


Figure S3 - continued

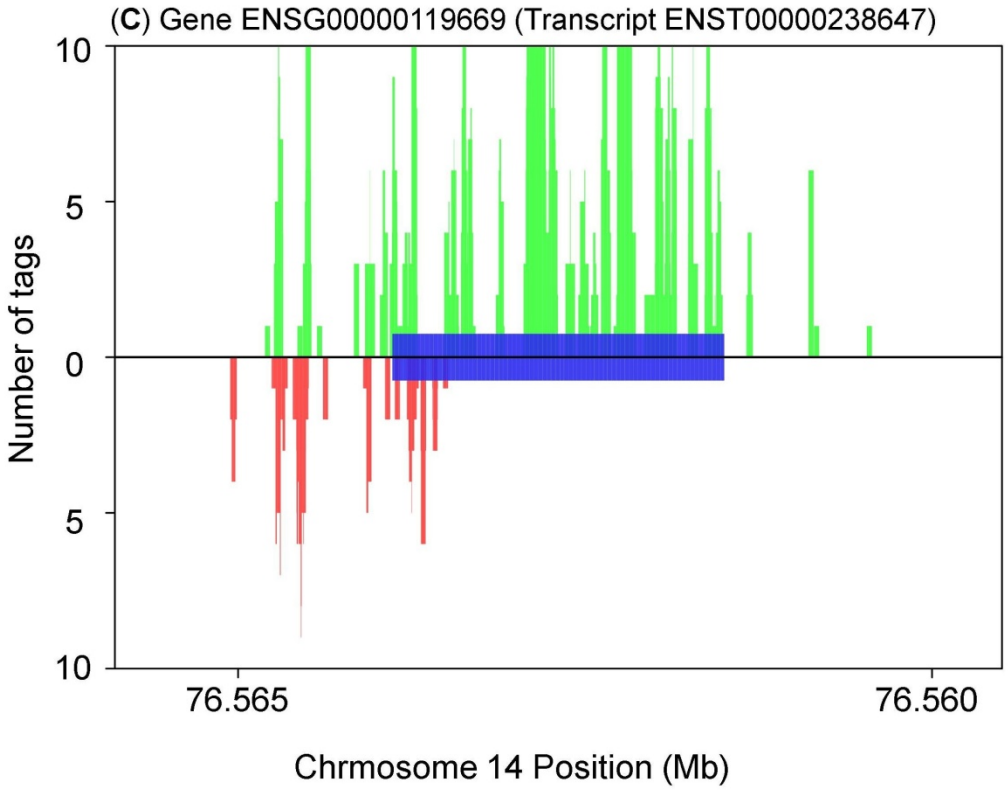


Figure S4 - RT-PCR confirmation of antisense transcripts in PBMC RNA. Bisulfite treated RNA was used as a template for RT-PCR with AS transcript specific primers as described in the Supplementary Materials and Methods. The expected specific products were observed in the reverse transcriptase treated RNA (+ lanes) and not in the minus reverse transcriptase control lanes (- lanes) confirming that the products were derived from RNA.

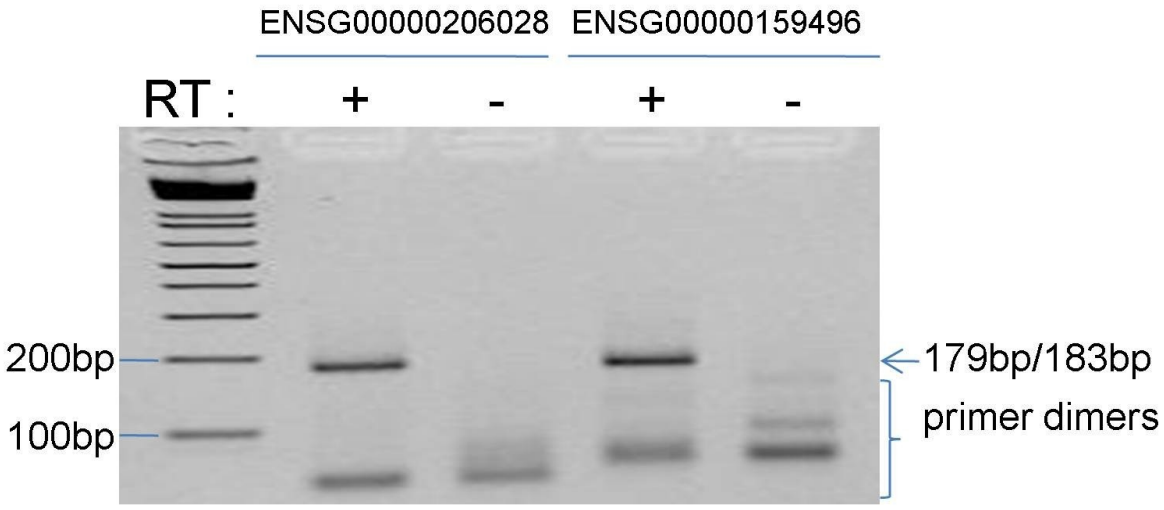
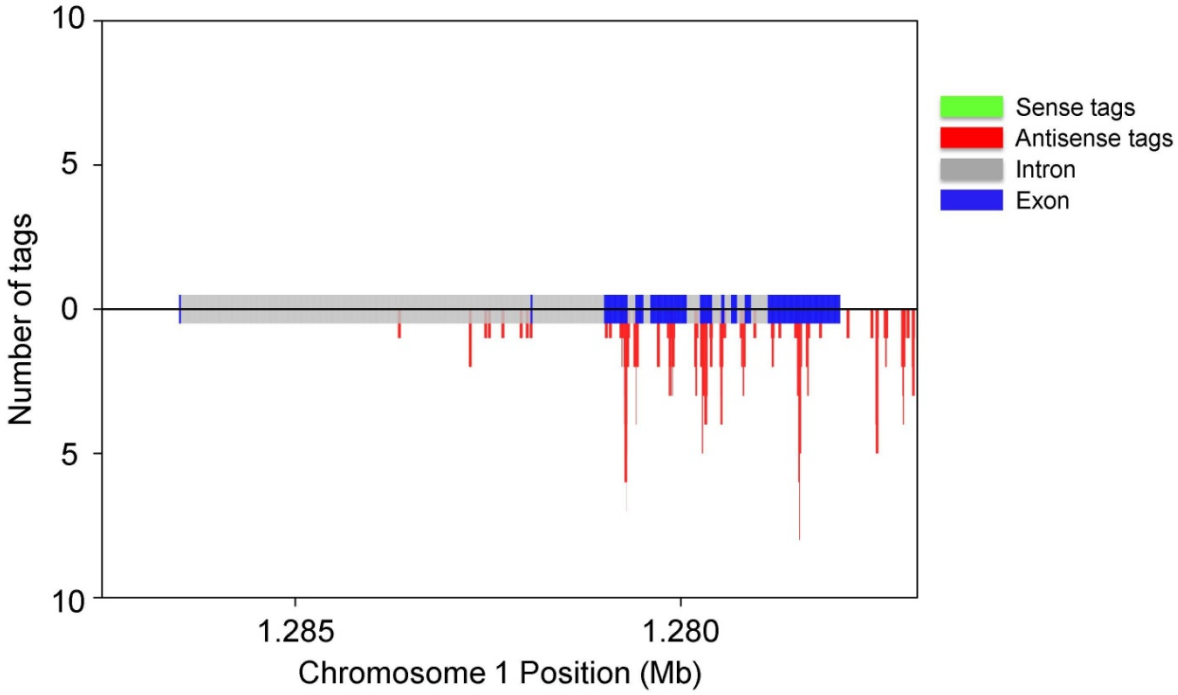


Figure S5 - Differential regulation of sense and antisense transcripts in different cell types. Gene ENSG00000162576 (transcript ENST00000378864) is an example of a gene that is AS in one cell line (A) and S in another cell line (B).

(**A**) Gene ENSG00000162576 (Transcript ENST00000378864) in MiaPaCa2 cells



(**B**) Gene ENSG00000162576 (Transcript ENST00000378864) in MRC5 cells

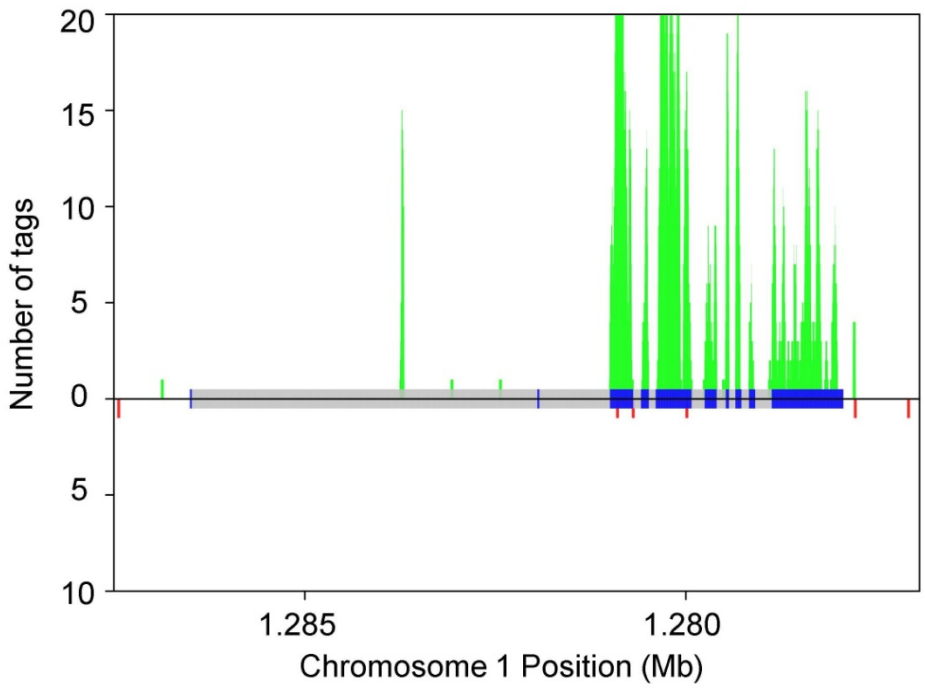


Figure S6 - Gene expression profiles from two independent ASSAGE experiments. For Jurkat cells, 11076 genes that had a combination of at least 5 unique sense and antisense tags in the first experiment were analyzed (A, B); 11379 genes that met the same selection criteria from MRC5 cells were analyzed (C, D). Each tag count was normalized to one million total tags for plotting.

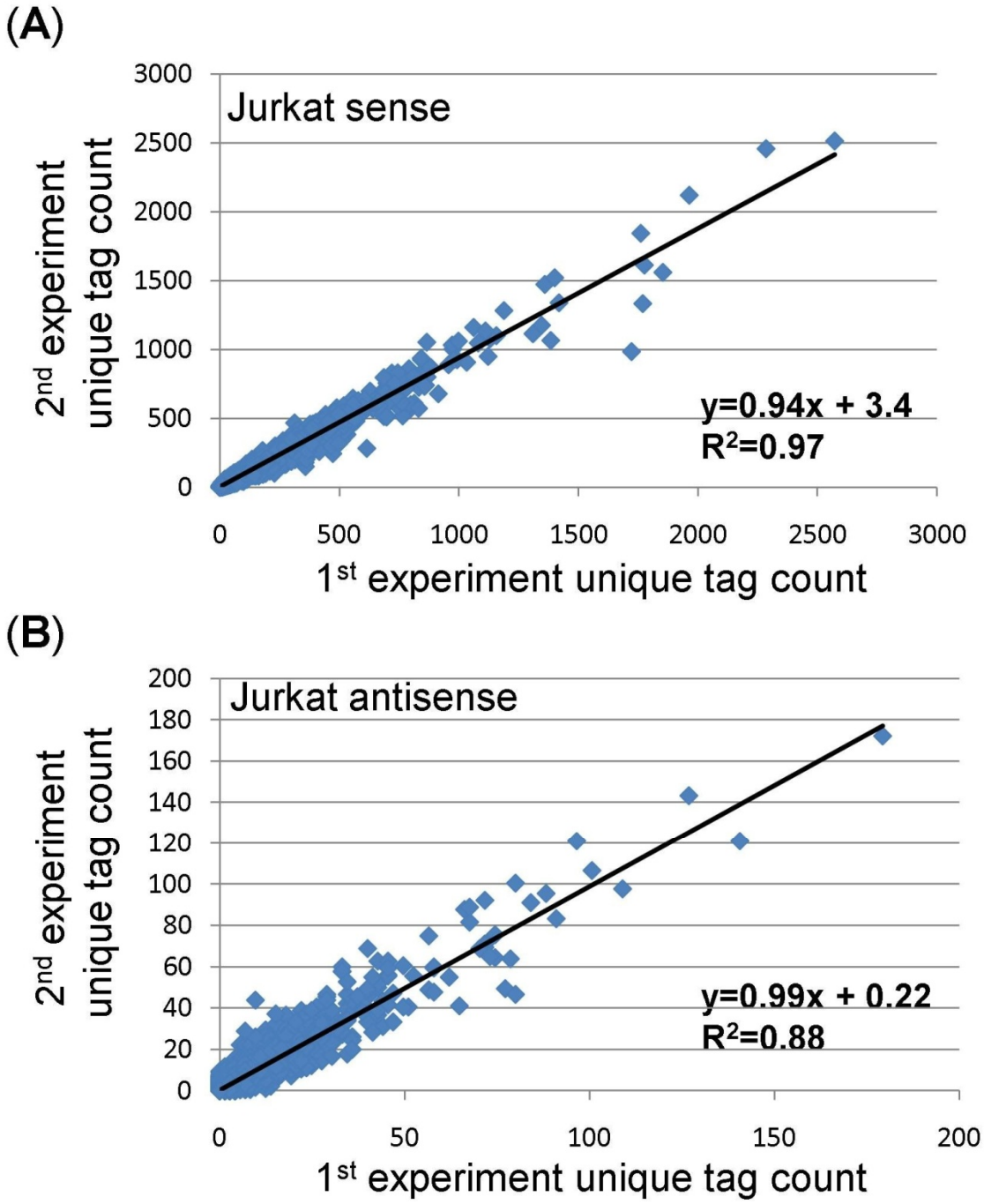


Figure 6 - continued

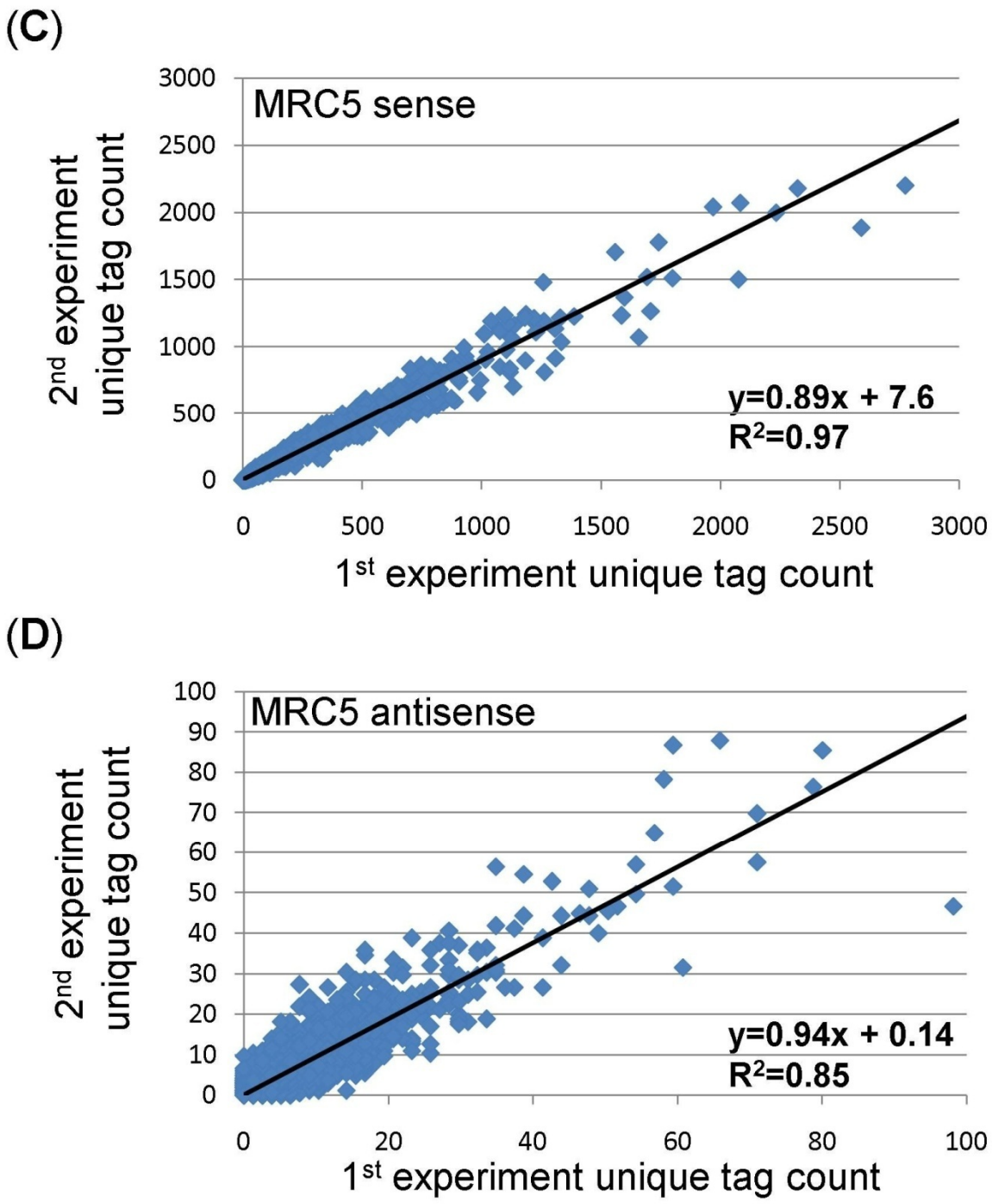


Figure S7 - Confirmation of differential regulation of sense and antisense transcripts for gene ENSG 00000162576 in different cell types. Bisulfite treated RNA from MiaPaCa2 or MRC5 cells was used as a template for RT-PCR with sense (A) and antisense (B) transcript specific primers as described in the Supplementary Materials and Methods. The expected specific products were observed in the reverse transcriptase treated RNA (+ lanes) and not in the minus reverse transcriptase control lanes (- lanes) confirming that the products were derived from RNA. As predicted by ASSAGE, predominantly sense transcripts were expressed in MRC5 cells and antisense transcripts in MiaCaPa2 cells.

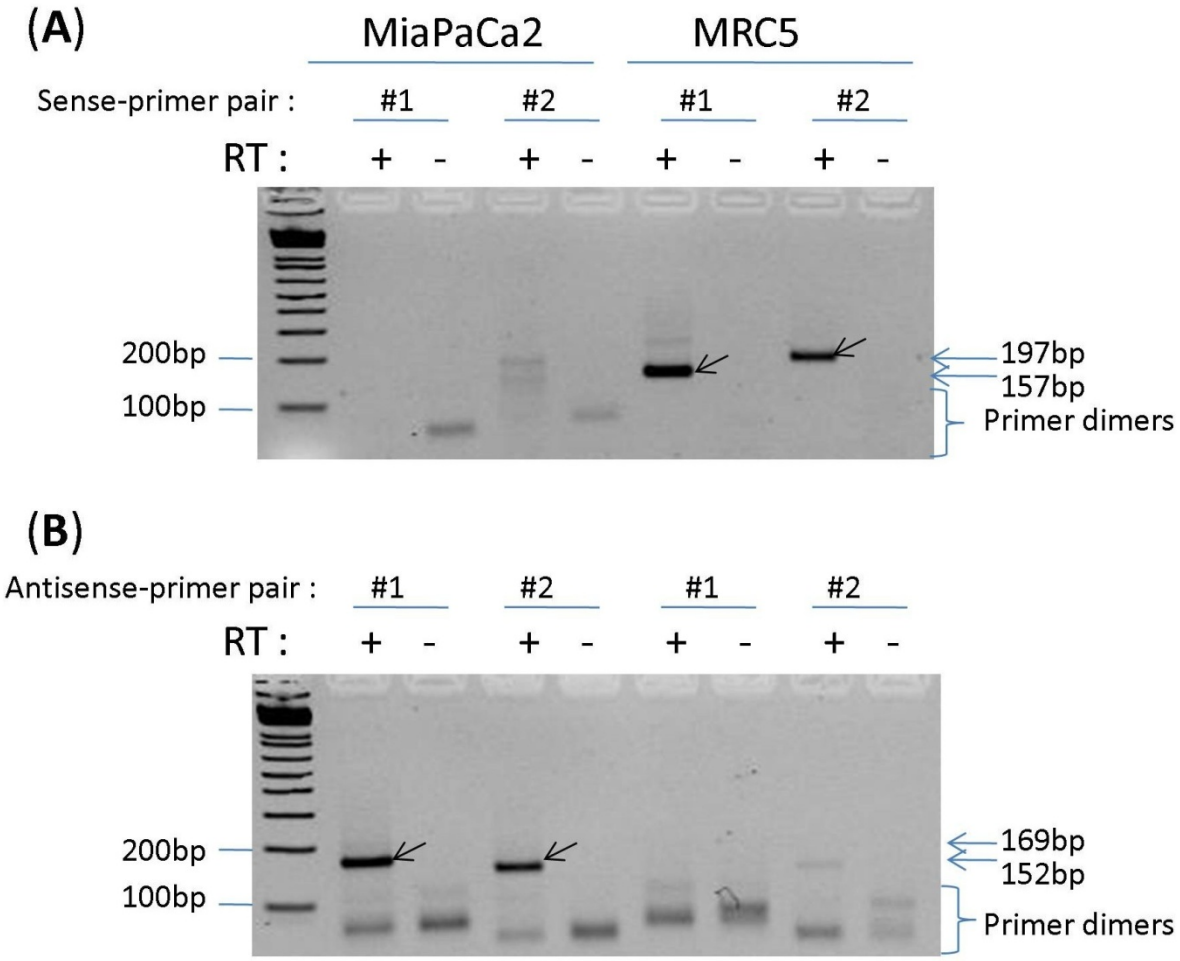
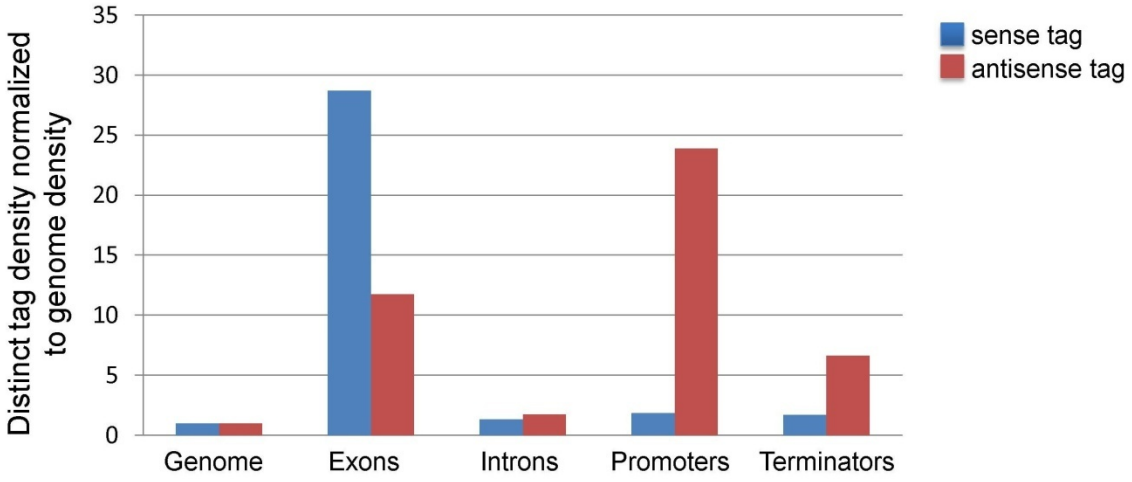


Figure S8 - ASSAGE tag densities in indicated cell lines. The density of distinct sense and antisense tags in the indicated regions were normalized to the overall genome tag density. The promoter and terminator regions were defined as the 1 kb of sequence that were upstream or downstream, respectively, of the transcript start and end sites.

(**A**) Jurkat



(**B**) HCT116

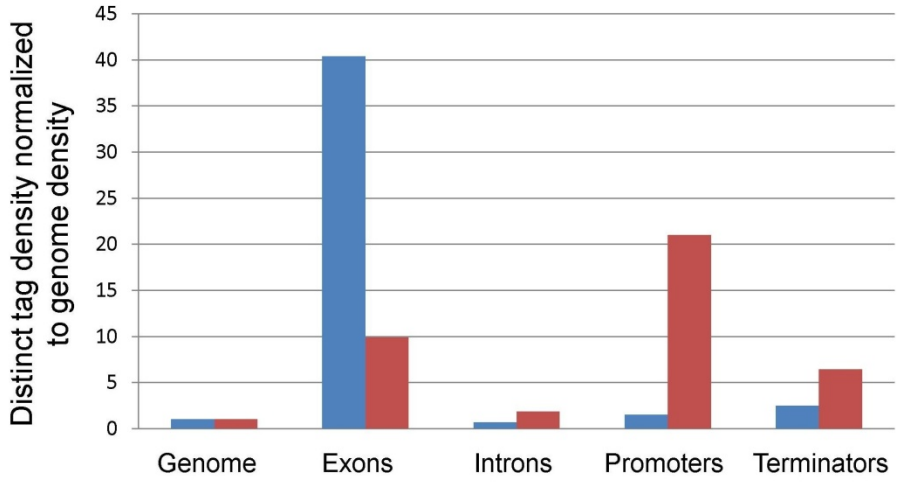
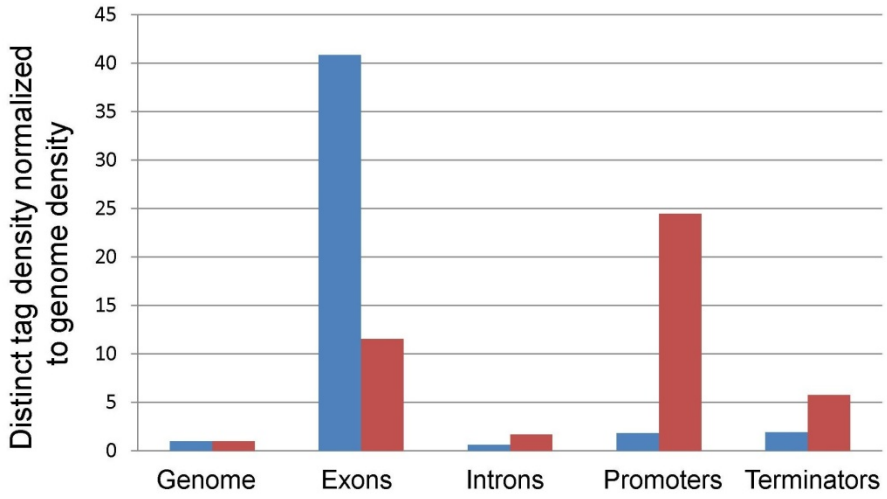


Figure S8 - continued

(**C**) MiaPaCa2



(**D**) MRC5

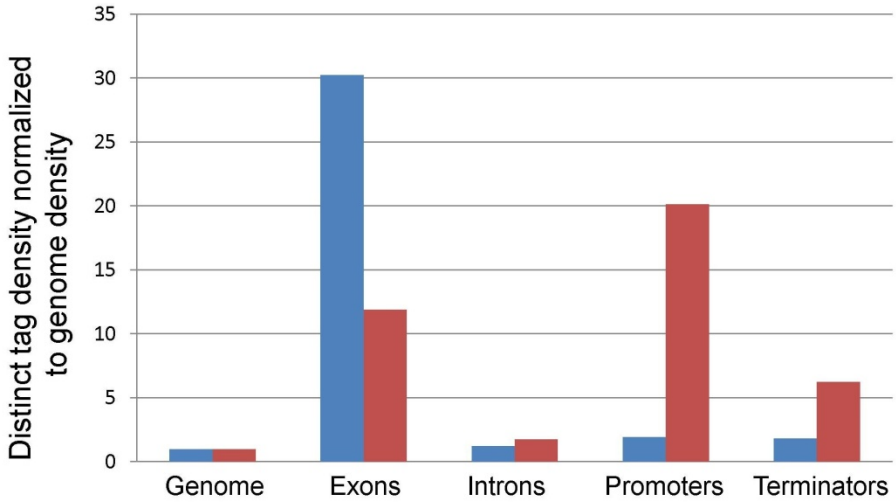


Figure S9 - Gene expression profiles from paired-end (PE) and regular (non-PE) ASSAGE experiments for Jurkat cell line (A-D) and MRC5 cell line (E-H). Non-PE ASSAGE experiment data was the same dataset as 2nd-experiment data in fig.S6. Only genes that had at least one unique tag from PE or non-PE experiment were included in each analysis. Each tag count was normalized to one million total tags for plotting.

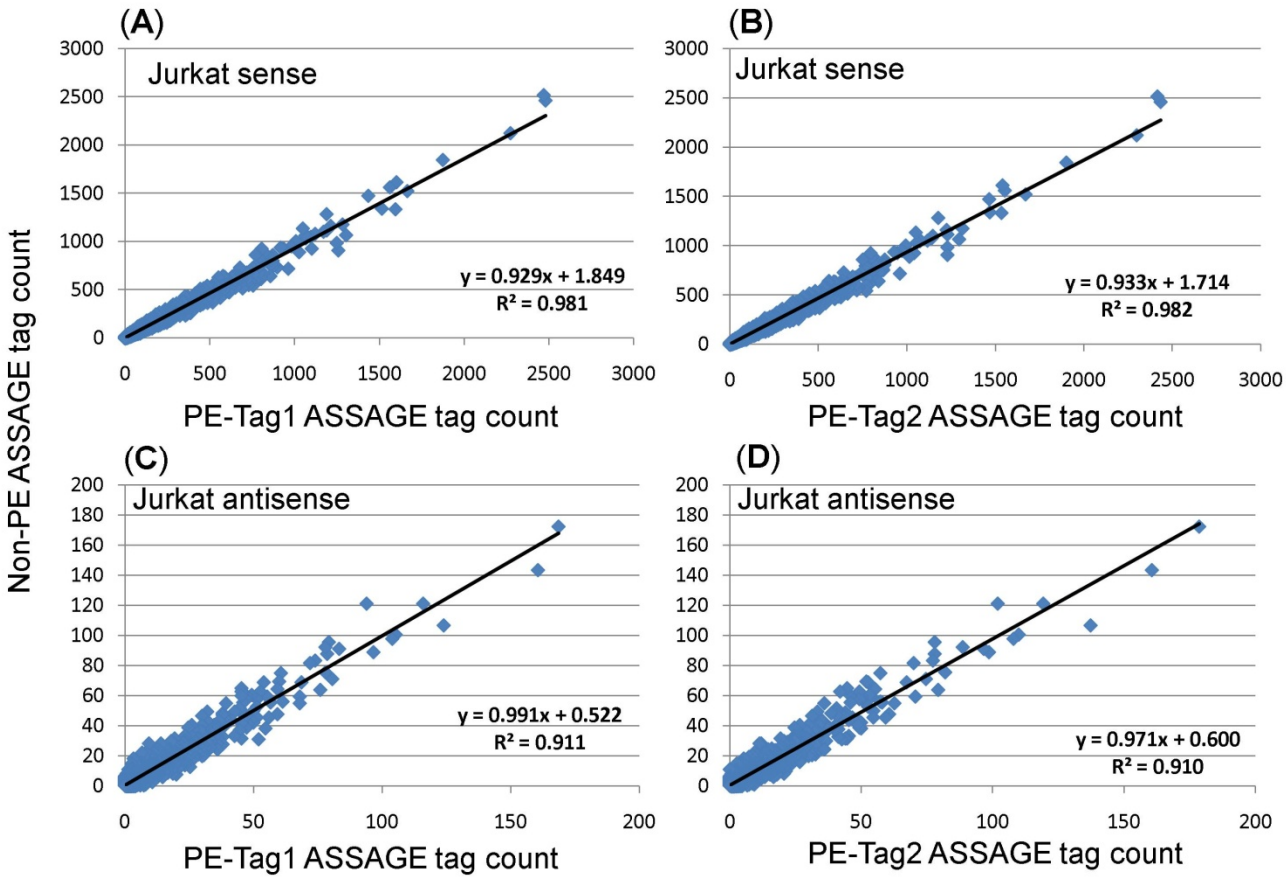


Figure S9 - continued

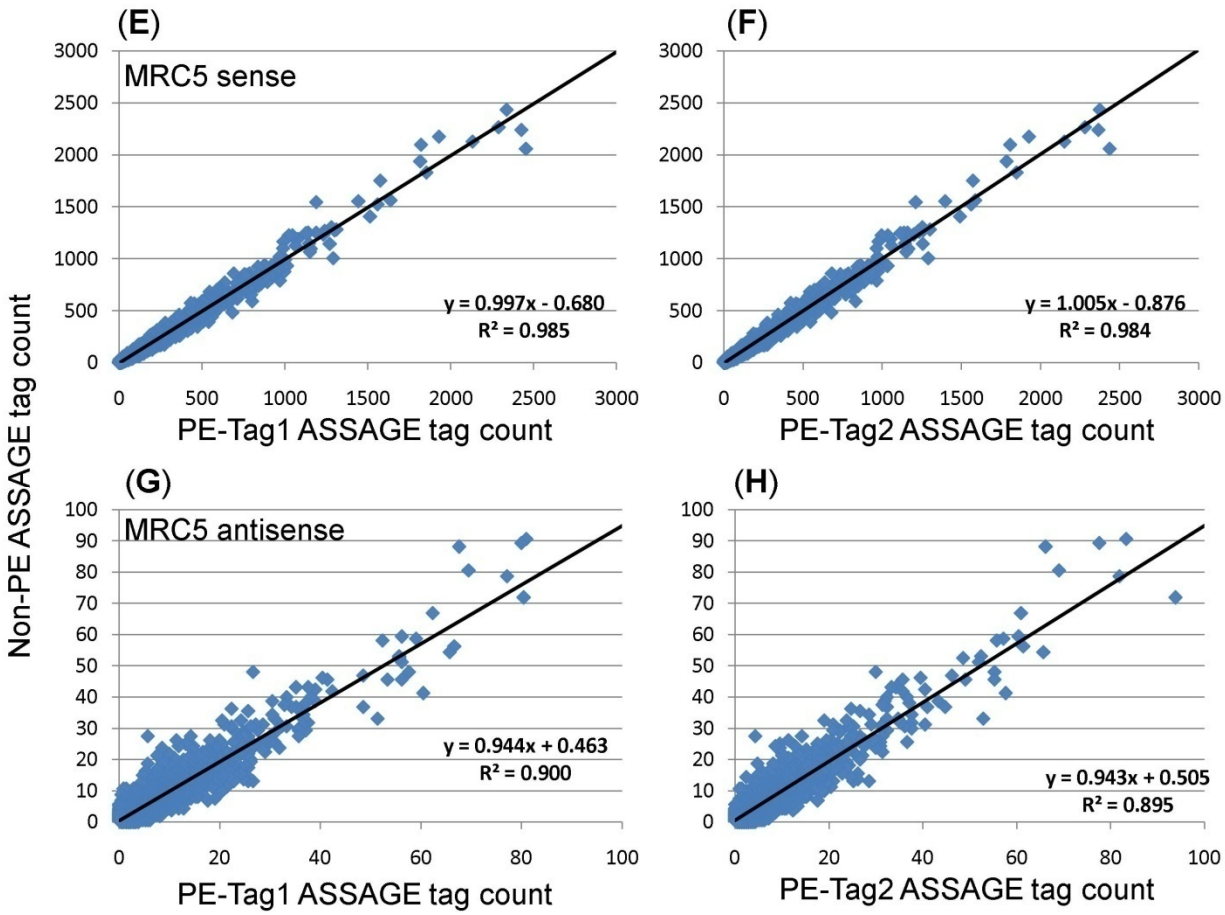
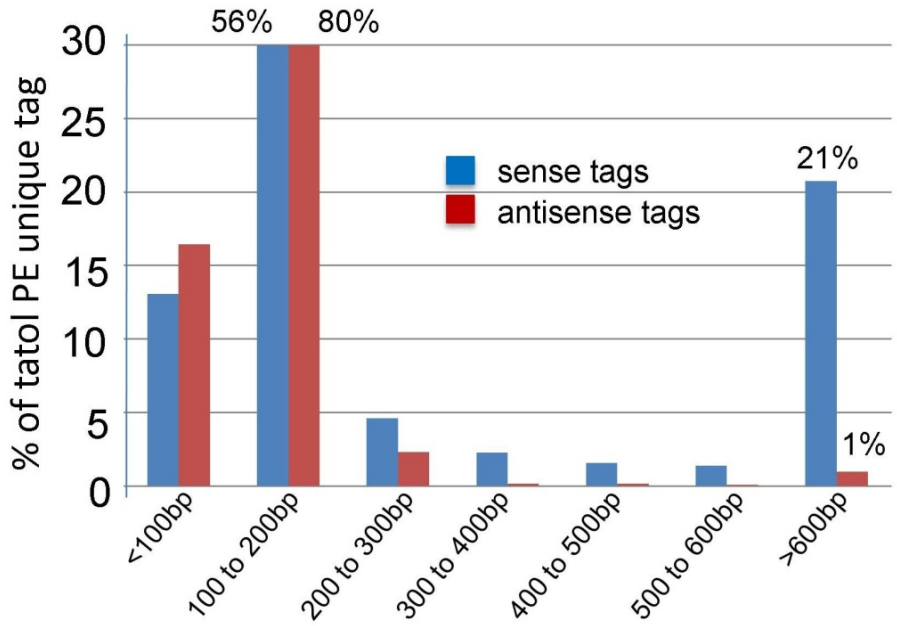


Figure S10 - Distribution of PE-ASSAGE tags from Jurkat (A) and MRC5 (B) classified by distances between genomic positions of paired tag1 and tag2. Note the percentages of PE tag with tag1to tag2 distances of 100-200bp or >600bp are indicated on the top of each bar.

(A) Jurkat



(B) MRC5

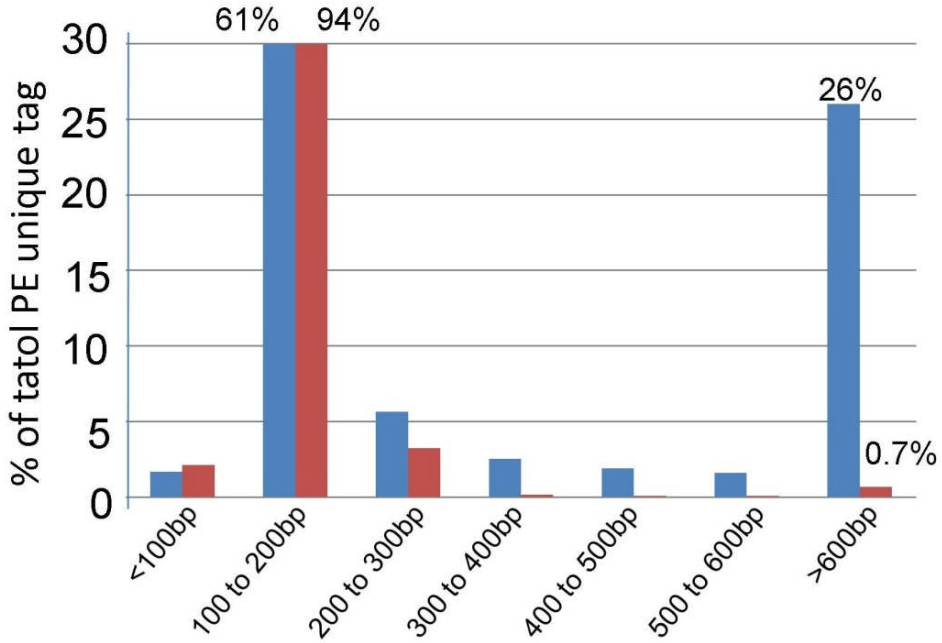


Figure S11 - An example of splicing of antisense transcript of gene ENSG00000162576 in MiaPaCa2 cells. RT-PCR identified a portion of the transcript that was spliced (A). Sequencing of the isolated PCR product and comparison with genomic sequence revealed that the splice junction was flanked by canonical splice sites on the DNA strand encoding the antisense transcript (B). The splice event was distinct from the known splicing of sense transcript (C). Similar splicing events were confirmed by PCR and sequencing for AS transcripts to genes ENSG00000157483, ENSG00000198624 and ENSG00000105679 from Jurkat cells and to gene ENSG00000121454 from MRC5 cells.

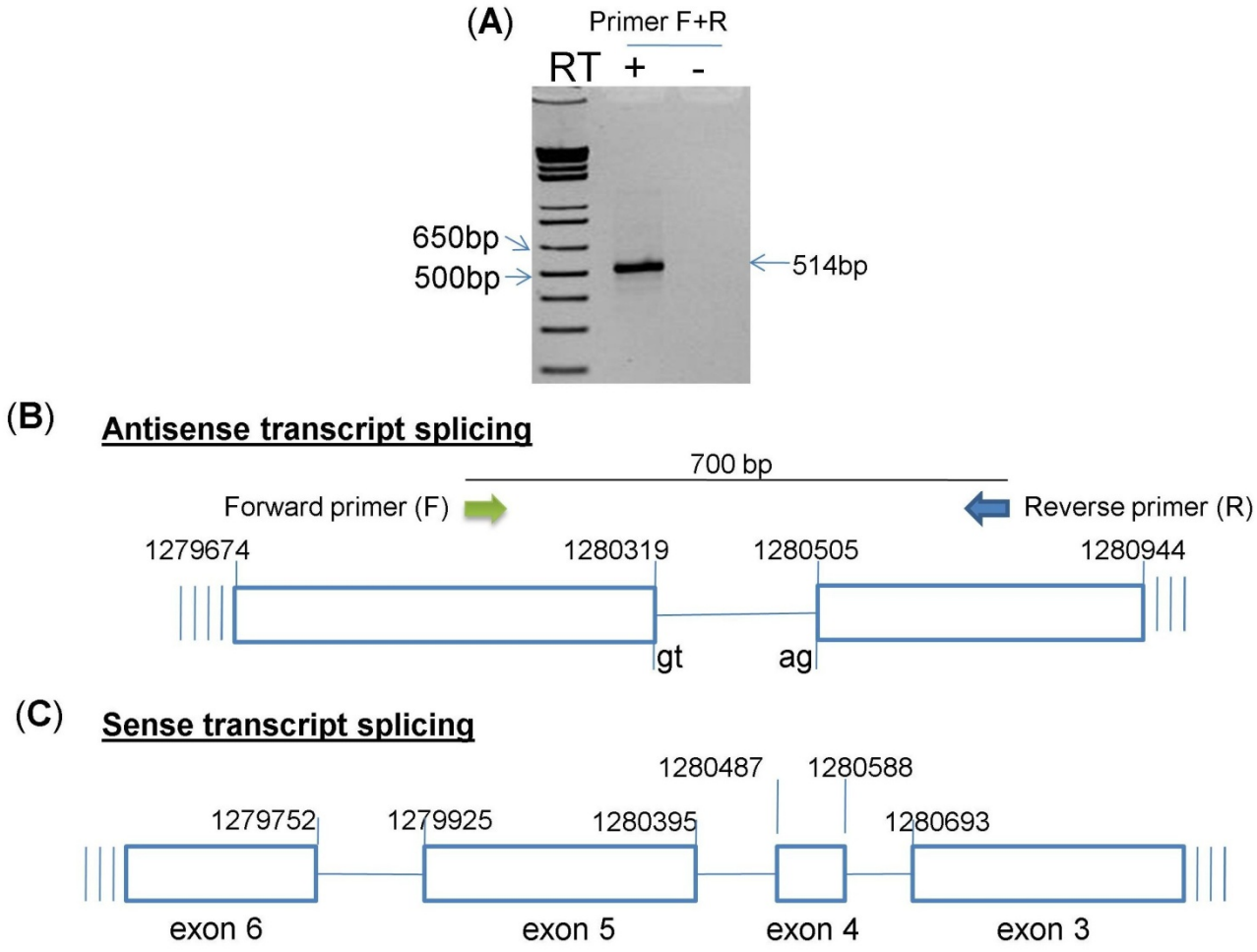
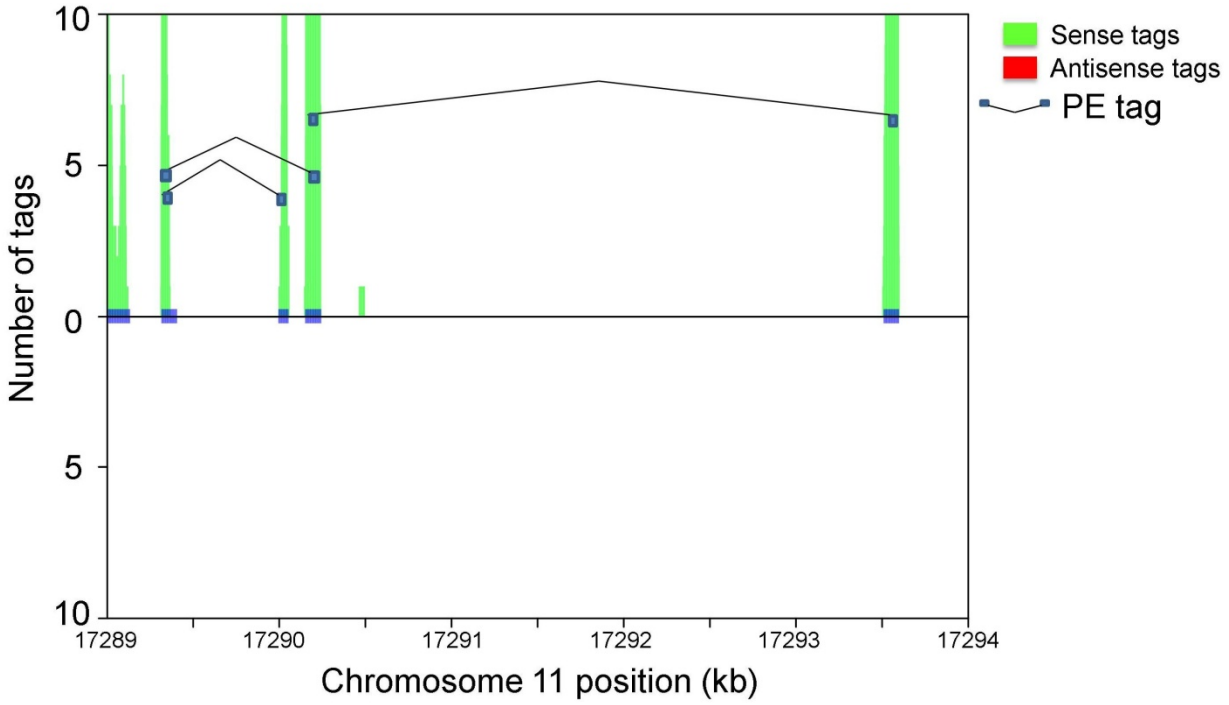


Figure S12 - Transcript splicing revealed by PE-ASSAGE tag analysis. One example of sense (**A**) and two of antisense (**B, C**) transcripts splicing events identified in Jurkat cells are shown. ASSAGE tag data for mapping was from an independent regular ASSAGE experiment (experiment #2). Each pear-end clone illustrated in the map represents one or multiple tags in which distances between tag1 and tag2 were greater than 600bp.

(**A**) Part of gene gene ENSG00000070081 (ENST00000323688)



(**B**) Part of gene ENSG00000198624 (ENST00000355417) intron 1-2

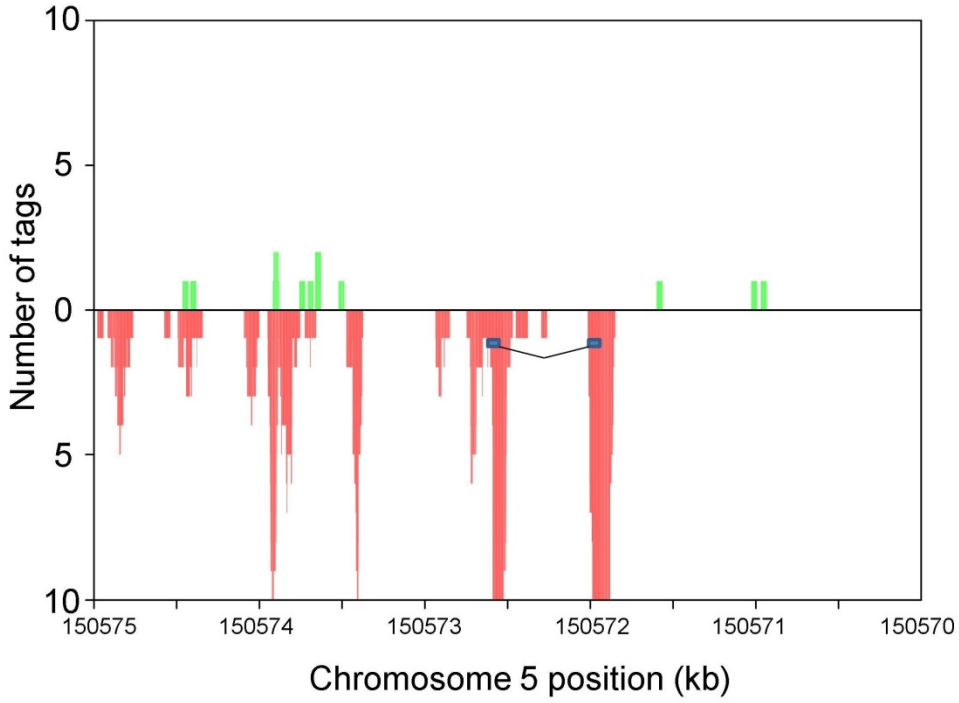


Figure S12 - continued

(**C**) Part of gene ENSG00000105679 (ENST00000222286)

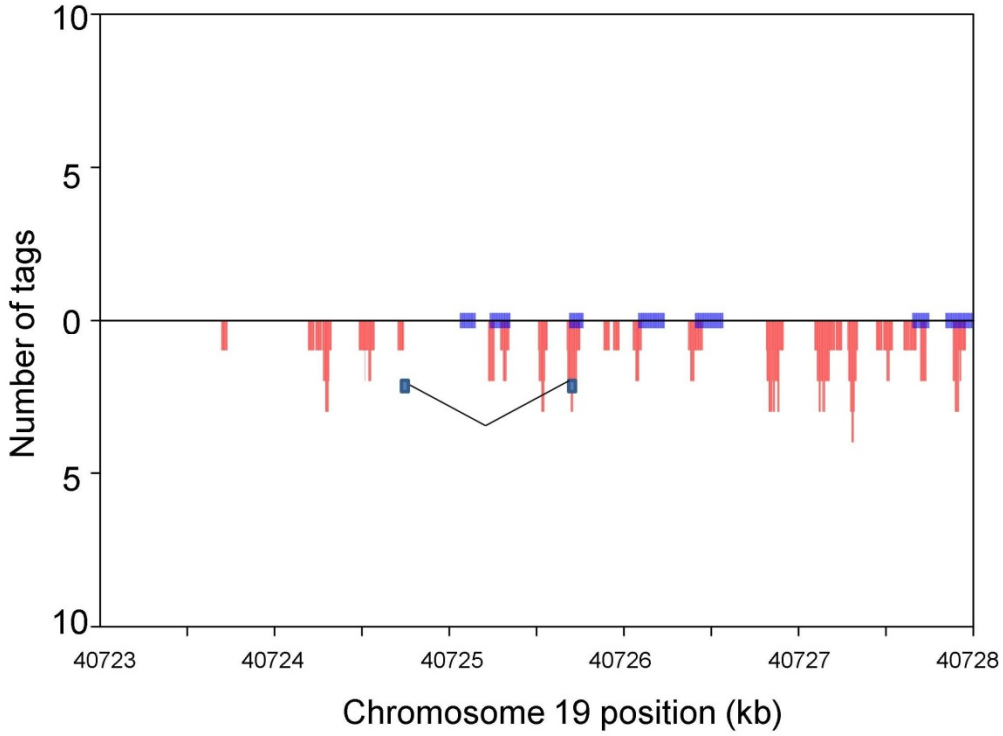


Figure S13 - Lack of correlation between density of distinct antisense tags of gene #1 and sense tags from gene #2 in the indicated cell lines. Gene #2 is defined as the closest neighboring gene arranged tail-to-tail with gene #1. A total of 28,269 genes were evaluated in each cell sample.