Isoform-specific translational control is evolutionarily conserved in primates

Jolene Draper[1]*, Julia Philipp[1]*, Zach Neeb[1], Richard Thomas[1], Solomon Katzman[2], Sofie Salama, David Haussler and Jeremy R. Sanford[1+]

[1] University of California Santa Cruz, Department of Molecular, Cellular and Developmental Biology, Department of Chemistry and Biochemistry CA, 95064, USA

[2] Center for Biomolecular Science and Engineering, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95060, USA

* these authors contributed equally

[+] *Corresponding Author*

Abstract (250 words max)

Alternative splicing (AS) alters messenger RNA (mRNA) coding capacity, localization, stability, and translation. Here we use comparative transcriptomics to identify *cis*-acting elements coupling AS to translational control (AS-TC). We sequenced total cytosolic and polyribosome-associated mRNA from human, chimpanzee, and orangutan induced pluripotent stem cells (iPSCs), revealing thousands of transcripts with splicing differences between subcellular fractions. We found both conserved and species-specific polyribosome association patterns for orthologous splicing events. Intriguingly, alternative exons with similar polyribosome profiles between species have stronger sequence conservation than exons with lineage-specific ribosome association. These data suggest that sequence variation underlies differences in the polyribosome association. Accordingly, single nucleotide substitutions in luciferase reporters designed to model exons with divergent polyribosome profiles are sufficient to regulate translational efficiency. We used position specific weight matrices to interpret exons with species-specific polyribosome association profiles, finding that polymorphic sites frequently alter recognition motifs for *trans*-acting RNA binding proteins. Together, our results show that AS can regulate translation by remodeling the *cis*-regulatory landscape of mRNA isoforms.

Introduction

The human genome encodes over 1,400 RNA binding proteins (RBPs), reflecting the emergence of elaborate post-transcriptional regulatory networks [1]. RBPs package transcripts into large complexes known as messenger ribonucleoprotein particles (mRNPs). As a nexus of post-transcriptional gene regulation, these heterogeneous complexes influence virtually every aspect of mRNA fate. Precursor RNA (pre-mRNA) splicing is central to this process. The splicing reaction, catalyzed by a macromolecular ribonucleonucleoprotein complex known as the spliceosome, results in elimination of intervening sequences (introns) and ligation of protein coding regions (exons). RBPs, RNA polymerase II and RNA structural elements contribute to a context-dependent regulatory system that enables the spliceosome to construct different messenger RNA sequences (mRNA isoforms) from a common precursor transcript [2, 3]. In addition to the obvious consequence of expanded protein coding capacity, the process of alternative splicing (AS) also finetunes subsequent steps of the gene expression pathway, including mRNA localization, translation and decay [4-11].

mRNAs contain a myriad of regulatory signals governing post-transcriptional gene regulation. These functional RNA elements can be remodeled within transcript isoforms by AS. For example, selection of exon sequences that contain premature termination codons (PTCs) or frame shifts can induce nonsense mediated mRNA decay (AS-NMD). Recently, polyribosome fractionation and high throughput mRNA sequencing (Frac-seq) revealed that mRNA isoforms can exhibit differential polyribosome association, suggesting that alternative splicing can be coupled to translational control (AS-TC) ((Sterne-Weiler et al. 2013; Floor and Doudna 2016; Blair et al. 2017; Reixachs-Solé et al. 2020; Wong et al. 2016). The inclusion or exclusion of microRNA targets sites, uORFs and protein binding sites through alternative promoter usage, AS, or alternative polyadenylation introduces the potential for isoform-specific translational

control (Sterne-Weiler et al. 2013; Floor and Doudna 2016; Sandberg et al. 2008; Spies et al. 2013; Taliaferro et al. 2016; Wang et al. 2012).

Comparative genomics and transcriptomics are powerful approaches for studying the evolution of gene regulatory mechanisms. For example, comparisons of tissue-specific gene expression and alternative splicing patterns across distantly related vertebrate species revealed that these modes of gene regulation evolve at different rates (Calarco et al. 2007; Mazin et al. 2018; Merkin et al. 2012; Barbosa-Morais et al. 2012). Comparative transcriptomic and proteomic analysis of primate cells also demonstrated that mRNA levels might evolve under less rigorous evolutionary pressure compared to protein expression levels (Khan et al. 2013). AS-NMD also appears to be strongly conserved through evolution. Hundreds of regions in the human genome are ultraconserved with the mouse and rat (Bejerano et al. 2004). Many of these sequences overlap alternative exon sequences associated with AS-NMD, suggesting that regulatory elements or the function of these mRNA isoforms are under strong purifying selection (Lewis et al. 2003; Pan et al. 2006; Lareau and Brenner 2015; Saltzman et al. 2008). Indeed, ablation or programmed mis-splicing of AS-NMD results in growth defects and loss of tumor suppression in cell culture and mouse embryo models (Thomas et al. 2020; Leclair et al. 2020). Taken together AS-NMD appears to be part of an evolutionarily conserved regulatory mechanism for fine tuning the expression of splicing factors.

By contrast to AS-NMD, the physiological and evolutionary significance of AS-TC is unknown. In this manuscript, we use comparative transcriptomics to investigate conservation of AS-TC and to identify exon sequences contributing to translational control. We generated Frac-seq data from human, chimpanzee, and orangutan induced pluripotent stem cells (iPSCs). Among thousands of mRNA isoforms with differential polyribosome association, we discovered orthologous alternative splicing events with evidence for either conserved or species-specific

translational control. We further showed that events with conserved translational control show higher sequence conservation specifically in cassette exons (and alternative first exons), indicating the functional relevance of these exons. We validated the ability of orthologous exons from isoforms with differential polyribosome association to affect mRNA translation in vivo using luciferase reporters. Taken together our data suggest that exonic sequences can regulate isoform-specific mRNA translation and that AS-TC is a conserved mechanism of gene regulation in primates.

Results

Frac-Seq analysis of primate iPSCs

We previously developed Frac-seq as a method to determine the association of alternative mRNA isoforms with the polyribosome (Sterne-Weiler 2013). To test the hypothesis that alternative splicing coupled to translational control is conserved in primates, we applied the Frac-seq methodology to human, chimpanzee, and orangutan iPSCs (Fig. 1A, Supplemental Table 1). We identified and quantified 6349, 5890 and 2417 alternative splicing events from all subcellular fractions isolated from human, chimpanzee and orangutan iPSCs, respectively (Fig 1B, Supplemental Table 2-4). Of these events we found 3579, 3445 and 1247 alternative splicing events with PSI values differing between fractions by more than 10%. We further separated these into events that generate non-productive NMD isoforms (AS-NMD; 792, 794 and 352 in human, chimpanzee, and orangutan iPSCs, respectively) and events maintaining the integrity of their open reading frame. The latter we consider alternative splicing events that are potentially implicated in translational control (AS-TC events; 2787, 2651 and 895, in human, chimpanzee and orangutan, respectively Fig 1C). Of these, 1061 splicing events exhibited differential sedimentation in both human and chimpanzee cells, whereas 264 events were classified as AS-TC in all three species (Fig 1D, Supplemental Table 5-7).

AS-TC is conserved across primate evolution

To identify patterns of polyribosome association in orthologous AS-TC events, we performed unsupervised hierarchical clustering on the Spearman correlation of the mean PSI for each species, across all fractions. As a control, we performed the same analysis using orthologous splicing events exhibiting uniform polyribosome association (ΔPSI <10%). These events cluster predominantly by species rather than by subcellular fraction (Fig 2A). By contrast, orthologous

AS-TC events show a stronger correlation for PSI between subcellular fractions rather than by species (Fig 2B). To identify orthologous isoforms with the most and least similar sedimentation profiles we calculated the difference in PSI between pairs of species and summed those differences into a single distance metric. The Manhattan distance for all orthologous AS-TC events results in a right-skewed distribution where the far left and right tails represent the AS-TC events with the most and least conserved sedimentation profile, respectively (Supp Fig 1A). We visualized these subsets of events in correlation heatmaps. The colors indicate the Spearman correlation of PSI values in events with low cumulative distance (similar sedimentation profiles) (Fig 2C) and in events with high cumulative distance (distinct sedimentation profiles) (Fig 2D). Interestingly, events with similar sedimentation profiles, cluster by fraction indicating the PSI values in these fractions are more similar to each other than to the other fractions within the same species. By contrast, in events with distinct sedimentation profiles, the different polyribosome fractions from each species cluster together, indicating more similarity within the species than the fractions. For example, C12orf29 exhibits a similar sedimentation pattern between species (Fig 2E, left panel), whereas CNN1 alternative splicing generates isoforms with species-specific sedimentation patterns (Fig 2d, right panel). In events with species-specific sedimentation patterns, the PSI values across the polysomal fractions differ between species, despite a well correlated nuclear output, represented by the PSI value in the cytoplasmic fraction (r= 0.90, 0.85, and 0.86 for human vs. chimp, human vs. orangutan and chimp vs orangutan, respectively. Supplemental fig 2).

Exon sequences associated with AS-TC are strongly conserved

To test the hypothesis that exonic sequence elements are determinants of AS-TC, we examined the sequence conservation of orthologous skipped exons (SE) mRNA isoforms with similar or species-specific polyribosome profiles. We also visualized the phastCons score for AS

events that were not associated with differential polyribosome association. As expected, skipped exons that are not associated with AS-TC are much less conserved compared to their flanking exons. By contrast, exons linked to AS-TC exhibit significantly higher phastCons scores and are more similar to their flanking exons. Further, AS-TC exons that have conserved sedimentation profiles between species have elevated phastCons scores relative to the other classes (Fig 3a). The high degree of sequence conservation within AS-TC exons suggests the presence of functional elements within these exons that influence polyribosome association. We next investigated the relationship between sequence variation and isoform-specific polyribosome sedimentation. Using pairwise alignments, we compared human and chimpanzee skipped (SE) AS-TC exons and calculated the SNP frequency within AS, conserved AS-TC and species-specific AS-TC events. After normalizing for sequence length, we observe the highest SNV frequency in AS-TC exons with species-specific polysome association patterns followed by canonical AS exons, followed by AS-TC exons with conserved sedimentation (Figure 3B). As expected, SNVs are enriched of variants around 50bp away from the splice sites in the exon sequence in SE events (Figure 3 C,D). This enrichment agrees with previously observed positional biases of SNPs around splice sites (Majewski and Ott, 2002). Interestingly, we observed a lower SNP density in AS-TC exons with similar sedimentation profiles between species. Taken together, these data suggest that exonic sequence variation can influence polyribosome association of mRNA isoforms.

Single nucleotide variants in orthologous AS-TC exons influence translation

To test the hypothesis that sequence differences associated with isoform-specific sedimentation patterns regulate mRNA translation, we created luciferase reporters from SE events from exhibiting AS-TC. Fig 4A shows the schematic for the skipped exon luciferase reporters, where

either the human (green) or chimpanzee (orange) skipped exon were inserted in-frame, upstream of the firefly luciferase. For each reporter, we measured luciferase activity and steady-state mRNA levels. Frac-seq analysis revealed differential sedimentation profiles for human and chimp isoforms from the GGCX, MELK and SUMF2 genes (Fig 4C-E left panel, respectively). The orthologous exon sequences for each of these events differ by only a single nucleotide (Supplemental Figure 3). Dual luciferase assays revealed that orthologous reporters for GGCX and MELK reporters exhibited significantly different activity (Figure 4C,D, middle panels). Frac-seq data indicates that human GGCX exon 2 inclusion is enriched in polyribosome fractions compared to chimp, where it is broadly distributed across the gradient. Consistent with this result, the human GGCX luciferase reporter is expressed more robustly than the chimp reporter. The opposite trend is observed for MELK exon 3, which is depleted from polyribosome fractions relative to the orthologous chimp isoform. In this case, the luciferase reporter containing the chimp sequence is more robustly expressed than the orthologous human reporter (Figure 4C-D, middle panels). This effect was likely due to translational control, as the steady-state mRNA levels are identical between the orthologous GGCX and MELK reporters (Figure 4 C-D, right panels). Additionally, the stability of luciferase enzyme levels following translational arrest with cycloheximide are not significantly different for the orthologous reporters, indicating that the single nucleotide substitution does not impact protein stability (supplemental figure 4). By contrast, orthologous reporters from SUMF2, exhibit nearly identical luciferase expression and reporter mRNA levels. These data are consistent with the Frac-seq profiles for human and chimp SUMF2 isoforms, which exhibit increased exon 2 inclusion in both species. Taken together, these data demonstrate that single nucleotide polymorphisms can influence translational control.

Alternative splicing alters the *cis*-regulatory landscape of mRNA isoforms

In order to test the effect of alternative splicing on the *cis*-regulatory landscape of mRNA isoforms, we performed pairwise alignments of alternatively spliced exons (e.g. skipped exons)

between two species at a time (human/chimpanzee and human/orangutan) and identified the sequence differences, predominantly single nucleotide variants (SNV), between the species. We then tested the effect of these sequence differences on predicted protein-RNA interactions sites. For that purpose, we scored the similarity of RNAcompete-derived positional weight matrices (PWMs) (Ray et al. 2017) against the sequences surrounding SNVs in a sliding window, collecting a predicted binding score for each tested RBP (Figure 5A). We performed this analysis on three groups of alternatively spliced events: events that are alternatively spliced but do not exhibit dynamic polysome association in one species but that are AS-TC events in the other species, events that were considered AS-TC in both species, and AS events without dynamic polysome association in both species. Next, we aimed to identify changes in the *cis*-regulatory landscape of mRNA isoforms that might be connected to dynamic, conserved polysome association or species-specific polysome association. For that, we compared the frequency of RBP binding sites that are changed in binding score due to the SNV in the different event groups (Figure 5B-C, Supplemental Tables 8-9). The AS events in both species serve as a control. Interestingly, both human/chimpanzee and human/orangutan sequence differences resulted in changes in putative RBP binding sites with frequencies significantly different to the AS/AS control group (Figure 5B,C). Out of the 10 PWMs that showed significant differences in frequency in the human/chimpanzee group and the 21 in the human/orangutan group, 2 were found in both analyses (Figure 5D). The alternatively spliced events analyzed for the human/chimpanzee and the human/orangutan comparison are predominantly distinct sets, with an overlap of 205 events (Figure 5E). Taken together these data suggest that SNPs associated with AS-TC exons with divergent Frac-seq profiles may alter the binding sites for RBPs associated with translational control.

Discussion

We applied a comparative transcriptomics approach to determine how genetic variation influences mRNA isoform polyribosome association in closely related primate stem cell models. We hypothesized that divergent sequences may change the cellular fate of orthologous mRNA isoforms and unveil functional elements important for post-transcriptional regulation of gene expression. Using Frac-Seq, we identified orthologous mRNA isoforms with similar or distinct distributions in sucrose gradients. At face value, these data demonstrated that polyribosome association can be regulated in an isoform-specific manner across primates. Using exon skipping events as a model, we found that orthologous isoforms with similar polyribosome sedimentation profiles exhibited strong sequence conservation across vertebrate phylogeny when compared to genes with species specific sedimentation profiles or non-AS-TC skipping events. We present evidence that single nucleotide polymorphisms within skipped exons are sufficient to alter reporter mRNA translation. These data suggest that functional elements within coding exons can regulate translation. We tested this hypothesis by leveraging publicly available RNAcompete data to investigate the impact of the SNVs on potential protein-RNA interactions (Ray et al. 2013). This analysis revealed several *trans*-acting RBPs that recognize sequences associated with AS-TC, suggesting that mRNP composition influences the cytoplasmic fate of mRNA isoforms.

Our data show that the coupling of alternative splicing to translational control is conserved across multiple primate cell lines, representing approximately 13 million years of primate evolution. We discovered 1000-3000 AS-TC events in human, chimp, and orangutan iPSC lines. Approximately 30-50% of AS-TC events were classified as AS-TC in another primate cell model and 362 alternative events exhibited AS-TC in all three species (Figure 1E). By comparing PSI between fractions and species, we found that isoform abundance was strongly correlated by

fraction, reflecting similar isoform-specific sedimentation patterns between species (Figure 2B). AS-TC events with shared sedimentation profiles also showed stronger alternative sequence conservation compared to species-specific AS-TC or AS events (Figure 3). The evolutionary conservation of these sequences indicates a potentially important biological function for this mode of gene regulation or the presence of potentially important *cis*-regulatory elements.

The mechanisms coupling alternative splicing to translational control are unclear. In the case of isoforms arising from alternative promoters distinct 5'UTR sequences can regulate isoform-specific translation (Niederer et al. 2022; Floor and Doudna 2016; Blair et al. 2017). It is less clear how alternative splicing of coding exons influences translational control. By analyzing orthologous skipped exons with similar or divergent sedimentation profiles in sucrose gradients we found significant differences in sequence conservation. Indeed orthologous alternative exons with similar sedimentation profiles between species exhibited a high degree of sequence conservation compared to exons with divergent sedimentation patterns (Fig. 3). Using a previously described luciferase reporter system we found that a single nucleotide substitution induced differences in reporter mRNA translation in HEK 293T cells (Fig. 4). For example, GGCX exon 2 and MELK exon 3 show different patterns of exon inclusion across sucrose gradients in human and chimp iPSCs. These differences in Frac-seq profiles are reflected in allele-specific expression of luciferase reporters in vivo. Together these data argue that *cis*-acting elements within the open reading frame provide another regulatory point for translational control.

One possibility is that the reach of exonic splicing enhancers extends to the cytosol . We previously demonstrated that the nucleo-cytoplasmic shuttling SR protein SRSF1 and a well characterized recognition motif from the *FN1* pre-mRNA can enhance translation both in vitro and in vivo (Sanford et al. 2004). SRSF1 promotes translation initiation through interactions with

mTORC1 components and inducing phosphorylation of the translational repressor 4E-BP (Blaustein et al. 2005; Michlewski et al. 2008); (Sanford 2004; Maslon et al. 2014)). Based on evidence presented here, we hypothesize that the sequences of AS-TC exons interact with *trans*-acting RNA binding proteins to drive isoform-specific translational control. We found that recognition motifs for several splicing factors such as SRSF1 and SRSF10 as well as translational control factors including LIN28A, FXR2, and DAZAP1 are frequently altered by SNPs within orthologous AS-TC exons with divergent sedimentation profiles (figure 5) . Additional work will be required to determine how exons such as MELK exon 3 or GGX exon 2 regulate translation.

Several mechanisms were not explored in our analysis and remain viable options for specific transcripts. Changes in mRNA sequence could contribute to isoform-specific differences in RNA secondary structure, which has previously been shown to regulate mRNA translation (Mao et al. 2014; Chemla et al. 2020; Mauger et al. 2019). For example, more stable secondary structures near the start codon attenuate translation initiation (Kudla et al. 2009). Likewise, codon optimality influences translation elongation and RNA stability in a variety of species (Gu et al. 2010; Tuller et al. 2010; Presnyak et al. 2015). In metazoans, codon optimality is important for enhancing the expression of intronless genes but could also influence codon usage within alternatively spliced isoforms, either through inducing a frameshift, modulating GC content, or by varied use of exons with suboptimal codons. Despite these open questions, the work presented here indicates that regulation of gene expression by AS-TC is evolutionarily conserved across the great apes and implicates RNA binding proteins as potential regulators of AS-TC in primates. Taken together, these findings suggest that alternative splicing influences mRNP composition by creating complexes with isoform-specific combinations of RNA binding proteins.

Materials & Methods

iPSC generation and culture

Human iPSC line C305 was derived from the GM12878 lymphoblastoid cell line using a proprietary episomal, plasmid-based reprogramming method by Cellular Dynamics. C305 cells were cultured on matrigel in mTeSR. Integration-free chimpanzee (Eip-8919-1A) and orangutan (Jos-C31) iPSCs were generated from primary fibroblasts as previously published by Field et al. (Field et al. 2017) and cultured on vitronectin in Essential 8 media.

Fractionation, polyribosome profiling, RNAseq

Frac-seq experiments were performed as previously published (Sterne-Weiler et al. 2013) using human, chimpanzee, and orangutan iPSC lines, with modification to the number of fractions collected. Cytosolic extracts from cell lines/tissues are fractionated by sucrose gradient centrifugation. We collected the total cytosolic lysate, the monoribosomal fraction (80s), as well as light (P2-4), medium (P5-8), and heavy (P9+) polyribosomal fractions. We then polyA+ selected RNA from each fraction and individually barcoded libraries using the Bioo NEXTflex Directional RNASeq kit. Multiplex-pooled libraries were sequenced on an Illumina HiSeq 4000 using paired-end 2x150bp sequencing, resulting in 75-150M reads per sample (Supplementary Figure 1B) with approximately 40-50% junction reads per sample.

Mapping of Illumina short read RNA sequencing

The reads were mapped to the human genome assembly hg38, the chimpanzee (Pan troglodytes) genome assembly panTro6, and the sumatran orangutan (Pongo abelii) genome assembly ponAbe3 using STAR v2.7 (Dobin et al. 2013). Repeat sequences were masked by mapping to repeatMasker sequences (Smit et al. RepeatMasker Open-4.0 at

http://repeatmasker.org) using Bowtie2 (Langmead et al. 2009). PCR duplicate removal was performed by collapsing fragments with common start and end positions and CIGAR strings using in house scripts. All data collection and parsing was done with bash and python2.7. Statistical analyses and data visualization were performed using R programming language version 3.5.1.

Identification and quantification of orthologous alternative splicing events

We used Stringtie (Pertea et al. 2015) to identify unannotated transcripts and used CAT transcriptomes (Fiddes et al. 2018) (for human (hg38), chimpanzee (panTro6), and orangutan (ponAbe3)), together with the stringtie merge command to generate the final transcriptomes for each species. Pairwise alternative splicing events were identified by pairwise comparison of all transcripts with at least one exon-intron junction in common. An alternative event was defined to be a set of exons unique to one transcript that are surrounded by two exon-intron junctions common to both transcripts or by one junction and the transcript terminus.

Alternative splicing events were quantified by counting the number of reads supporting the exon-exon junction and the number of reads supporting the exon-intron junction of each event. PSI (percent spliced in) values were calculated as the ratio of number of reads supporting the included isoform to the number of reads supporting both isoforms.

The event identification and quantification was implemented using in house python scripts. Orthologous alternative splicing events were identified by mapping the event sequences from one species to the genomic sequences of all others. The genes containing the orthologous event coordinates were determined using the CAT transcriptome annotations.

Cross-fraction and Cross-species comparison of conserved and species-specific orthologous events

All events identified were filtered to be supported by at least 15 junction reads (per comparison). Alternative splicing events undergoing translational control, or AS-TC events, were defined as events with a change in PSI value (delta PSI) between any two adjacent fractions of at least 0.1. Consequently, alternative splicing events not undergoing translational control were defined as events with a minimum PSI > 0 and delta PSI < 0.1. Alternative splicing events leading to nonsense-mediated decay (AS-NMD) were identified using *in silico* translation of raw transcripts and subsequent identification of premature termination codons (PTCs) (technically CDSinsertion).

For estimating the difference/conservation of the polysome association pattern we calculated the Manhattan distances for each event between each two[1] [2] species. The Manhattan Distance is the sum of differences in mean psi between two species across all fractions. Min/max normalization of the Manhattan distance allowed us to identify events with overall different sedimentation profiles as opposed to events with similar sedimentation profiles at a different y-axis intercept. We ranked all AS-TC and AS-NMD events based on their min/max normalized Manhattan distance and used the top and bottom 10% (= 350 events) for further analysis, considering them the least and most conserved set of events respectively.

Determination of Sequence conservation

To determine the sequence conservation of AS and AS-TC events with conserved or species-specific sedimentation profiles, phastCons (Siepel and Haussler; Siepel 2005) scores were obtained from the UCSC genome browser (Kent et al. 2002). For skipped exon events, phastCons scores were obtained for 100nt windows around both splice sites of the cassette exon as well as

around the upstream 5'ss and the downstream 3'ss. [3] [4] The scores were visualized with local nonlinear smoothing using a generalized additive model.

RNA purification and RT-qPCR

Total RNA was isolated using the Direct-zol RNA MiniPrep Kit (Zymo Research). 800ng of the RNA were treated with RQ DNase (per protocol). The DNase (Promega) treated RNA was reverse transcribed using the High-Capacity cDNA reverse transcriptase kit (Applied Biosystems). 1:200 dilutions of the cDNA were made. For the qPCR, we used Luna 2x SYBR premix (total volume 20μl per reaction), 0.25 nM primers and 5μl diluted cDNA. qPCR was performed on QuantStudio 6 Pro Real-Time PCR System (Applied Biosystems, Thermo Fisher) according to MIQE guidelines (Bustin et al. 2009).

Luciferase Reporters and reporter stability assays

HEK293T cells cultured in 6 well plates with DMEM supplemented with 10% FCS were cotransfected with either GGCX, MELK and SUMF2 reporters (500 ng) and a renilla luciferase plasmid (100 ng) as a transfection control. Luciferase activity was assayed 48 hours post transfection using Dual-Glo Luciferase Assay System (Promega). To measure reporter protein stability, HEK29T3 cells were cultured to 75% confluence in a 6-well plate were co-transfected with GGCX, MELK and SUMF2 Firefly luciferase reporter plasmid and Renilla luciferase plasmids as described above. 48 hours after transfection, cells were washed twice with 1x PBS and media containing 50 ug/mL cycloheximide (Santa Cruz Biotechnology) was then added to each well. Cells were harvested after incubating for 0-, 3-, and 6-hours with cycloheximide. Cells were washed once with 1x PBS and then lysed

in passive lysis buffer at room temperature for 15 minutes with vigorous shaking (Promega). 20 mL of each transfection cell lysate was added in triplicate to a 96-well plate and the Dual Luciferase Assay was performed according to the manufacturer's protocol (Promega). The levels of both Firefly luciferase and Renilla luciferase activity were then normalized to the 0 hour time point for all samples.

Alignment and identification of SNVs and indels

The sequences of alternative regions of AS-TC events from human and chimpanzee were globally (Needleman-Wunsch alignment) aligned using R Biostrings (Pagès et al. 2020) (function pairwiseAlignment) using default settings. Mismatches, insertions, and deletions were identified in the pairwise alignments using R Biostrings.

RBP binding prediction

RNAcompete (Ray et al. 2009, 2013) data sets available on ENCODE (ENCODE Project Consortium 2012) were used to predict RBP binding sites affected by single nucleotide sequence differences between human and chimpanzee AS-TC events. The matchPWM (Wasserman and Sandelin 2004) function from the R Biostrings package (Pagès et al. 2020) was used to score PWMs based on the RNAcompete data in a sliding window across the identified sequence differences. Matches achieving at least 80% of the maximum score were recorded for both human and chimpanzee datasets. The matches were compared between the species in form of a deltaPWMscore (e.g. PWMscore(human) - PWMscore(chimp)). [chisquare test for proportions]

Data Access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSExxx.

Figure Legends

Fig 1. Frac-seq reveals polyribosome associated mRNA isoforms. A) Alternative splicing can influence multiple post-transcriptional regulation pathways. B) Frac-seq (subcellular fractionation and subsequent sequencing of polyA+ selected RNA from fractions) was performed on human, chimpanzee, and orangutan iPSCs. RNA from the total cytosolic lysate, the mono-ribosome (80s), the light (P2-P4), medium (P5-P8), and heavy polyribosome (P9+) was sequenced. B) Identification and quantification of alternatively spliced events was performed using junctionCounts. This pipeline allowed the identification of 14 different event types. C) Alternative splicing events were further classified into AS, AS-TC, and AS-NMD events. The proportions of these three event groups are comparable between the three cell lines. D) Out of the events categorized as either AS, AS-TC, or AS-NMD, about 900 events were identified in all three cell lines. E) Out of the events categorized as AS-TC, over 300 events were identified in all three cell lines.

Fig 2. Orthologous AS-TC events exhibit either conserved or species-specific sedimentation profiles.

A) Heatmap of Spearman correlation of PSI values of all identified orthologous events. The columns and rows represent the total cytosolic lysate and the 4 subcellular fractions in each cell line. The colors represent the Spearman correlation of PSI values between pairs of fractions (red = high correlation, blue = low correlation). B) Heatmap of Spearman correlation of PSI values of all orthologous AS-TC events. C,D) Heatmaps of Spearman correlation of PSI values of orthologous AS-TC events. C) The events in this heatmap exhibit sedimentation profiles consistent across all three cell lines as shown in the example in panel E. D) The events in this heatmap exhibit species-specific sedimentation profiles as shown in the example in panel F. E) The skipped exon event within C12orf29 is an example of an alternative splicing event with

conserved sedimentation profiles across all three cell lines. F) The alternative first exon event of CNN1 is an example of an alternative splicing event with species-specific sedimentation profiles.

Fig 3. Alternative splicing events with sedimentation profiles consistent across species show higher sequence conservation. A) Sequence conservation of exon/intron boundaries of AS-TC and AS skipped exon events represented by phastCons scores. A lower score indicates less conservation. AS-TC events with conserved sedimentation profiles in blue, AS-TC events with species-specific sedimentation profiles in yellow, and simple AS events in black. B) Length normalized SNV density for skipped exon sequences associated with AS (black), conserved AS-TC profiles (blue) or species-specific AS-TC profiles (yellow). C,D) Positional distribution of SNP density in skipped exon sequences associated with AS (black), conserved AS-TC profiles (blue) or species-specific AS-TC profiles (yellow) relative to the 3' or 5' splice site, respectively.

Fig 4. AS-TC cassette exons drive isoform-specific expression. A) Schematic diagram of the pairs of luciferase reporter constructs containing either the human (green) or chimpanzee (orange) cassette exon from different genes exhibiting AS-TC. B) Schematic diagram of dual luciferase assay and RT-qPCR assay. C-E) Polyribosome sedimentation profiles from human and chimpanzee iPSCs (left panel), dual luciferase reporter assay data (middle panel) and RT-qPCR data (right panel) for isoforms from the GGCX, MELK and SUMF2 genes, respectively. Dual luciferase reporter assay and RT-qPCR analyses were performed on three independent transfections of orthologous reporter pairs. Statistical analysis by one-way ANOVA.

Fig 5. Alternative splicing alters the *cis*-regulatory landscape of mRNA isoforms. A) Schematic diagram of RBP binding analysis. B) Heatmap of log2 of relative frequencies of predicted RBP binding sites changed through single nucleotide variants between human and chimpanzee skipped exon sequences in AS/AS-TC events, AS-TC/AS-TC events and AS/AS

events compared to the AS/AS event control group. C) Heatmap of log2 of relative frequencies of predicted RBP binding sites changed through single nucleotide variants between human and orangutan skipped exon sequences in AS/AS-TC events, AS-TC/AS-TC events and AS/AS events compared to the AS/AS event control group. D) Euler diagram depicting the overlap of RBPs with significant differences in frequencies compared to the background between the human/chimpanzee and the human/orangutan analysis. E) Euler diagram depicting the overlap of analyzed events in the human/chimpanzee and the human/orangutan analysis.

Supplementary Fig 1. A) Distribution of cumulative Manhattan distance scores across AS-TC events. B) Number of uniquely mapped reads and junction reads mapped to each fraction collected from human, chimpanzee and orangutan iPSCs. C) Gene coverage in human, chimpanzee, and orangutan sequencing data. D) Pearson correlation of PSI values between replicates.

Supplementary Fig 2. A) Sashimi plot depicting junction reads mapped to a skipped exon event within C12orf29 (as shown in Figure 2E) across all fractions in human (green), chimpanzee (orange) and orangutan (purple). B) Sashimi plot depicting junction reads mapped to an alternative first exon event within CNN1 (as shown in Figure 2F) across all fractions in human (green) and chimpanzee (orange).

Supplementary Fig 3. Sequence alignments of orthologous skipped exons from human and chimpanzee frac seq data. (A) Alignment of GGCX exon 2. (B) Alignment of SUMF2 exon 2. (C) Alignment of MELK Exon 3. RBPs predicted to interact with either the human or chimpanzee allele are show in green or orange font, respectively.

Supplementary Fig 4. Luciferase enzyme activity stability measurements. HEK cells were transfected with human/chimp orthologous firefly reporters and a renilla luciferase reporter. 48 hours post transfection cells were treated with cycloheximide for 0, 3 or 6 hours. Luciferase levels were measured by dual luciferase assay and normalized to the 0h time point. Data were fit to a one phase decay nonlinear model and decay constants (K) for human and chimp ortholog pairs were compared using an Extra Sum of Squares F-test. The reporter pairs did not exhibit significant differences in decay constants.

Acknowledgements

We would like to thank our colleague Dr. Gina Mwala for critical feedback on the manuscript and Profs. Joshua Arribere, Angela Brooks, and Benedict Patten and Dr. Tim Sterne-Weiler for thoughtful discussion. We wish to thank Erin LaMontagne, and Lila Whitehead for their help culturing primate iPSCs. This research was made possible by a grant from the California Institute of Regenerative Medicine (GCIR-06673-A to JRS) and from the National Institutes of Health (R35GM130361 to JRS). The contents of this publication are solely the responsibility of the authors and not necessarily the official views of CIRM or any other agency of the State of California.

Author Contributions

Disclosure Declaration

No conflicts to disclose

References

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587–1593.

Bejerano G, Pheasant M, Makunin I, Stephen S, James Kent W, Mattick JS, Haussler D. 2004. Ultraconserved Elements in the Human Genome. *Science* **304**: 1321–1325. http://dx.doi.org/10.1126/science.1098119.

Blair JD, Hockemeyer D, Doudna JA, Bateup HS, Floor SN. 2017. Widespread Translational Remodeling during Human Neuronal Differentiation. *Cell Rep* **21**: 2005–2016.

Blaustein M, Pelisch F, Tanos T, Muñoz MJ, Wengier D, Quadrana L, Sanford JR, Muschietti JP, Kornblihtt AR, Cáceres JF, et al. 2005. Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* **12**: 1037–1044.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611–622.

Calarco JA, Xing Y, Cáceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM, Blencowe BJ. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev* **21**: 2963–2975.

Chemla Y, Peeri M, Heltberg ML, Eichler J, Jensen MH, Tuller T, Alfonta L. 2020. A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nat Commun* **11**: 4827.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, et al. 2018. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. *Genome Res* **28**: 1029–1038.

Field AR, Jacobs FMJ, Fiddes IT, Phillips APR, Reyes-Ortiz AM, LaMontagne E, Whitehead L, Meng V, Rosenkrantz JL, Haeussler M, et al. 2017. Structurally conserved primate lncRNAs are transiently expressed during human cortical differentiation and influence cell type specific genes. http://dx.doi.org/10.1101/232553.

Floor SN, Doudna JA. 2016a. Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**. http://dx.doi.org/10.7554/elife.10921.

Gu W, Zhou T, Wilke CO. 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol* **6**: e1000664.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler a. D. 2002. The Human Genome Browser at UCSC. *Genome Research* **12**: 996–1006. http://dx.doi.org/10.1101/gr.229102.

Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. 2013. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**: 1100–1104.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lareau LF, Brenner SE. 2015. Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* **32**: 1072–1079.

Leclair NK, Brugiolo M, Urbanski L, Lawson SC, Thakar K, Yurieva M, George J, Hinson JT, Cheng A, Graveley BR, et al. 2020. Poison Exon Splicing Regulates a Coordinated

Network of SR Protein Expression during Differentiation and Tumorigenesis. *Mol Cell* **80**: 648–665.e9.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* **100**: 189–192.

Mao Y, Liu H, Liu Y, Tao S. 2014. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic Acids Res* **42**: 4813–4822.

Maslon MM, Heras SR, Bellora N, Eyras E, Cáceres JF. 2014. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife* e02028.

Mauger DM, Cabral BJ, Presnyak V, Su SV, Reid DW, Goodman B, Link K, Khatwani N, Reynders J, Moore MJ, et al. 2019. mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci U S A* **116**: 24075–24083.

Mazin PV, Jiang X, Fu N, Han D, Guo M, Gelfand MS, Khaitovich P. 2018. Conservation, evolution, and regulation of splicing during prefrontal cortex development in humans, chimpanzees, and macaques. *RNA* **24**: 585–596.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**: 1593–1599.

Michlewski G, Sanford JR, Cáceres JF. 2008. The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell* **30**: 179–189.

Niederer RO, Rojas-Duran MF, Zinshteyn B, Gilbert WV. 2022. Direct analysis of ribosome targeting illuminates thousand-fold regulation of translation initiation. *Cell Syst* **13**: 256–264.e3.

Pagès H, Aboyoun P, Gentleman R, DebRoy S. 2020. Biostrings: Efficient manipulation of biological strings. R package version 2.48. 0.

Pan Q, Saltzman AL, Kim YK, Misquitta C, Shai O, Maquat LE, Frey BJ, Blencowe BJ. 2006. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev* **20**: 153–158.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290–295.

Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, et al. 2015. Codon optimality is a major determinant of mRNA stability. *Cell* **160**: 1111–1124.

Ray D, Ha KCH, Nie K, Zheng H, Hughes TR, Morris QD. 2017. RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods* **118-119**: 3–15.

Ray D, Kazan H, Chan ET, Peña Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q,

Hughes TR. 2009. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* **27**: 667–670.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172–177.

Reixachs-Solé M, Ruiz-Orera J, Albà MM, Eyras E. 2020. Ribosome profiling at isoform level reveals evolutionary conserved impacts of differential splicing on the proteome. *Nat Commun* **11**: 1768.

Saltzman AL, Kim YK, Pan Q, Fagnani MM, Maquat LE, Blencowe BJ. 2008. Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol Cell Biol* **28**: 4320–4330.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**: 1643–1647.

Sanford JR. 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes & Development* **18**: 755–768. http://dx.doi.org/10.1101/gad.286404.

Sanford JR, Gray NK, Beckmann K, Cáceres JF. 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* **18**: 755–768.

Siepel A. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research* **15**: 1034–1050. http://dx.doi.org/10.1101/gr.3715005.

Siepel A, Haussler D. Phylogenetic Hidden Markov Models. *Statistical Methods in Molecular Evolution* 325–351. http://dx.doi.org/10.1007/0-387-27733-1_12.

Spies N, Burge CB, Bartel DP. 2013. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res* **23**: 2078–2090.

Sterne-Weiler T, Martinez-Nunez RT, Howard JM, Cvitovik I, Katzman S, Tariq MA, Pourmand N, Sanford JR. 2013. Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res* **23**: 1615–1623.

Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS, et al. 2016. Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Mol Cell* **61**: 821–833.

Thomas JD, Polaski JT, Feng Q, De Neef EJ, Hoppe ER, McSharry MV, Pangallo J, Gabel AM, Belleville AE, Watson J, et al. 2020. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat Genet* **52**: 84–94.

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**: 344–354.

Wang ET, Cody NAL, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al. 2012. Transcriptome-wide regulation of pre-mRNA splicing and mRNA

localization by muscleblind proteins. *Cell* **150**: 710–724.

Wasserman WW, Sandelin A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics* **5**: 276–287. http://dx.doi.org/10.1038/nrg1315.

Wong QW-L, Vaz C, Lee QY, Zhao TY, Luo R, Archer SK, Preiss T, Tanavde V, Vardy LA. 2016. Embryonic Stem Cells Exhibit mRNA Isoform Specific Translational Regulation. *PLoS One* **11**: e0143235.

Rpackage citations (not found in paper pile)

1. Biostrings - Pagès H, Aboyoun P, Gentleman R, DebRoy S (2019). Biostrings: Efficient manipulation of biological strings. R package version 2.54.0
2. Xxx

Supp Data

---

larger data set is three species, most of the distance plots (literally every one except Corf) is two species. anyway, need to fix this sentence in one of those directions.
that brings me to my next comment..... do we discuss *why* most figures are two species? example: blame it on orangutan map/annotation confounded with low agreement of biological replicates. isn't this why people do triplicates.....
is this accurate..... @jsanfor2@ucsc.edu
Also, can we edit figure in ai to say AS-TC instead of ASTC for consistency? I still can't understand why the conservation of first two nucleotides into intron aren't near the highest phastcons score. anyway.
There isn't a data point for the human cytoplasm in the figure.

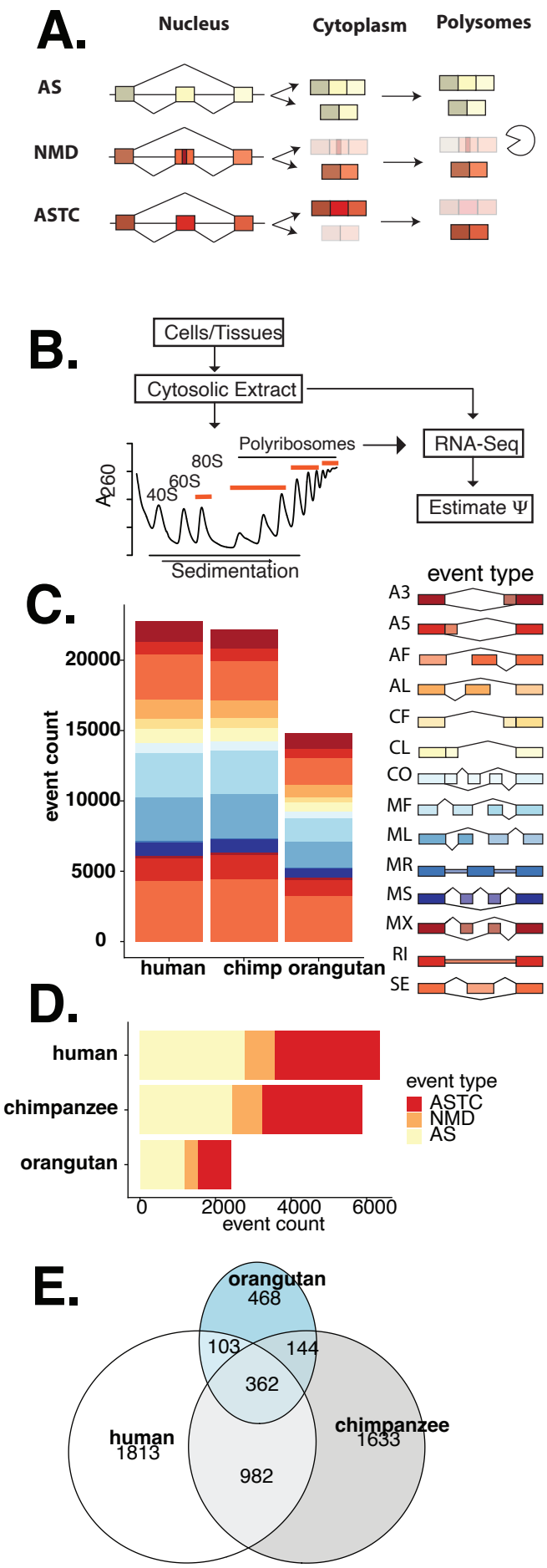

A.
Nucleus
Cytoplasm
Polysomes
AS
NMD
ASTC
B.
Cells/Tissues
Cytosolic Extract
Polyribosomes
RNA-Seq
Estimate Ψ
80S
60S
40S
A260
Sedimentation
C.
event type
A3
A5
AF
AL
CF
CL
CO
MF
ML
MR
MS
MX
RI
SE
event count
20000
15000
10000
5000
0
human
chimp
orangutan
D.
human
chimpanzee
orangutan
event type
ASTC
NMD
AS
0
2000
4000
6000
event count
E.
orangutan
468
103
144
362
human
1813
chimpanzee
1633
982

Figure 1

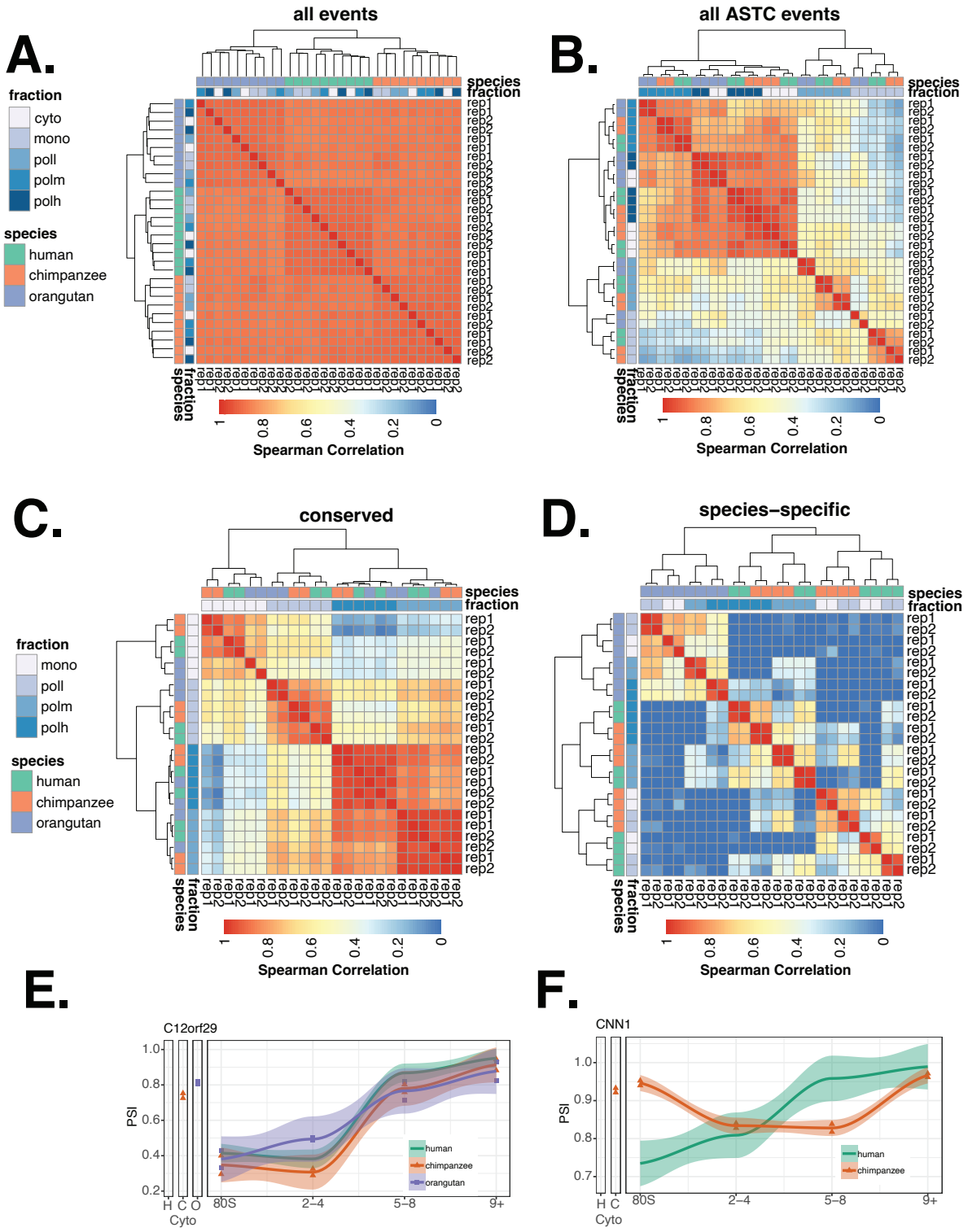

A.
all events
fraction
cyto
mono
poll
polm
polh
species
human
chimpanzee
orangutan
species
fraction
Spearman Correlation
B.
all ASTC events
C.
conserved
D.
species–specific
E.
C12orf29
PSI
H C O
Cyto
80S
2–4
5–8
9+
human
chimpanzee
orangutan
F.
CNN1
H C
Cyto

Figure 2

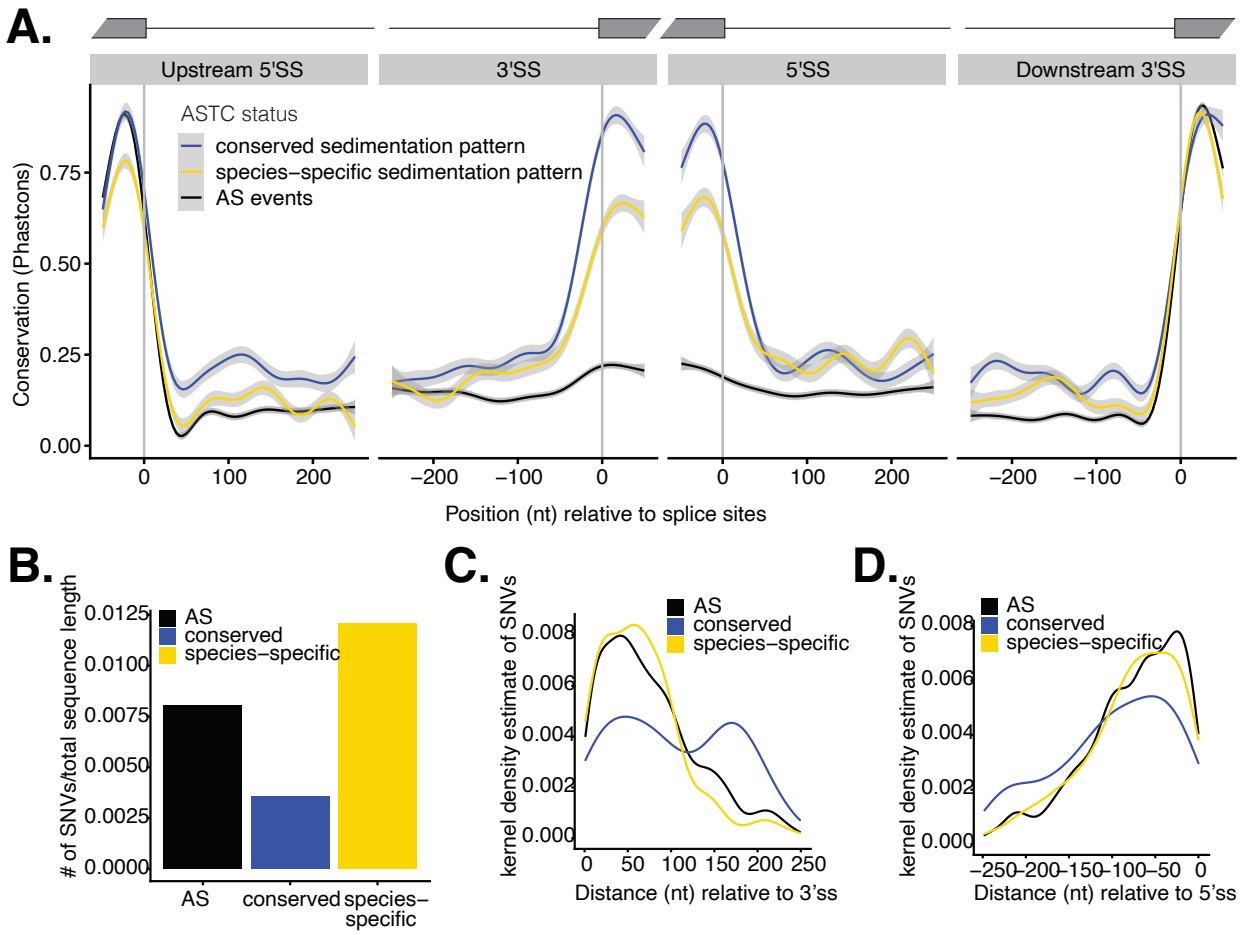

A.
Upstream 5'SS
3'SS
5'SS
Downstream 3'SS
ASTC status
conserved sedimentation pattern
species−specific sedimentation pattern
AS events
Conservation (Phastcons)
0.75
0.50
0.25
0.00
0
100
200
−200
−100
0
0
100
200
−200
−100
0
Position (nt) relative to splice sites
B.
# of SNVs/total sequence length
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
AS
conserved
species−specific
AS
conserved
species−specific
C.
kernel density estimate of SNVs
0.008
0.006
0.004
0.002
0.000
AS
conserved
species−specific
0
50
100
150
200
250
Distance (nt) relative to 3'ss
D.
kernel density estimate of SNVs
0.008
0.006
0.004
0.002
0.000
AS
conserved
species−specific
−250
−200
−150
−100
−50
0
Distance (nt) relative to 5'ss

Figure 3

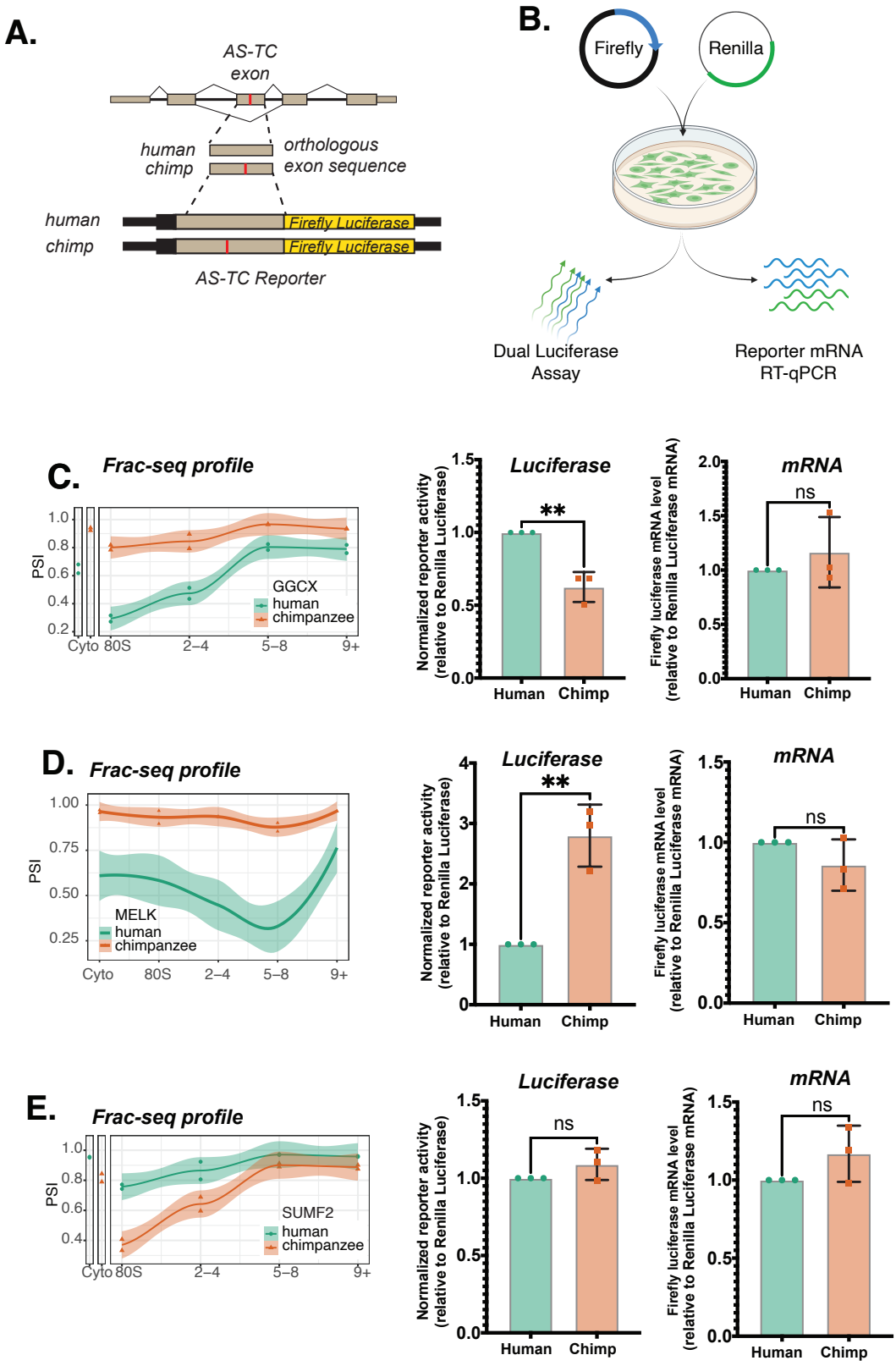

A.
AS-TC exon
human
chimp
orthologous exon sequence
human
chimp
Firefly Luciferase
Firefly Luciferase
AS-TC Reporter
B.
Firefly
Renilla
Dual Luciferase Assay
Reporter mRNA RT-qPCR
C.
Frac-seq profile
PSI
GGCX
human
chimpanzee
Cyto 80S 2–4 5–8 9+
Luciferase
Normalized reporter activity (relative to Renilla Luciferase)
**
Human Chimp
mRNA
Firefly luciferase mRNA level (relative to Renilla Luciferase mRNA)
ns
D.
Frac-seq profile
MELK
human
chimpanzee
Luciferase
**
mRNA
ns
E.
Frac-seq profile
SUMF2
human
chimpanzee
Luciferase
ns
mRNA
ns

Figure 4

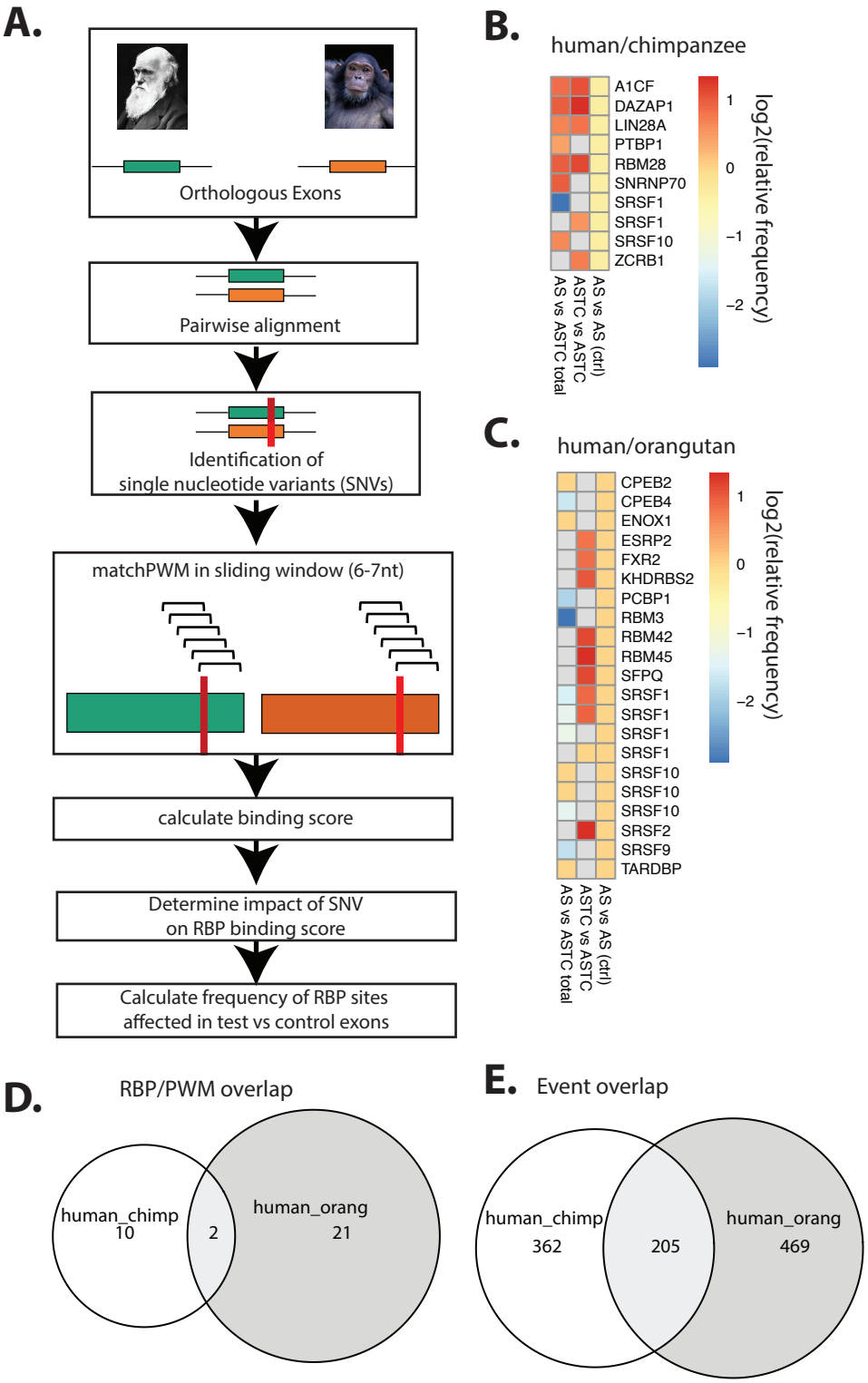
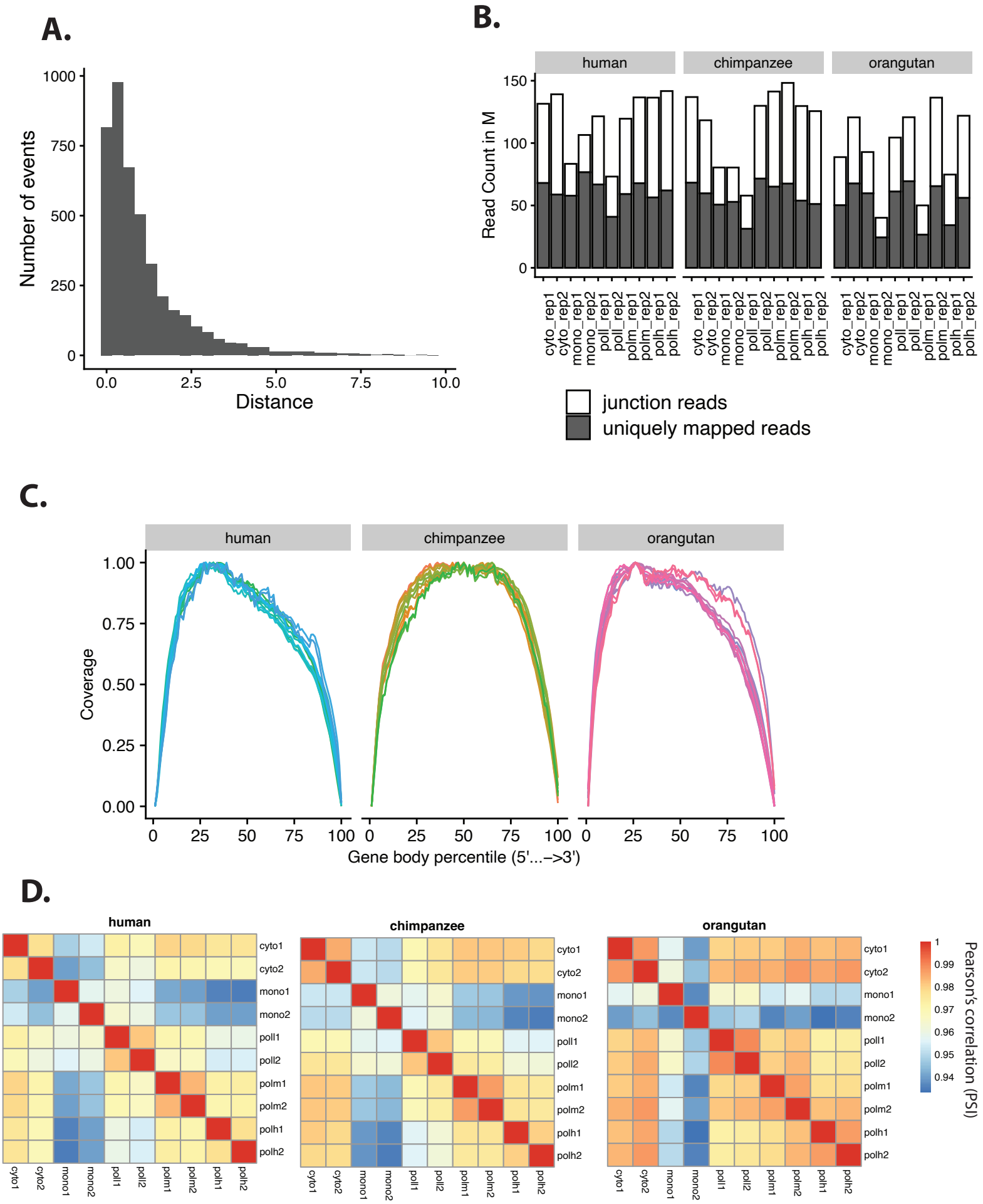
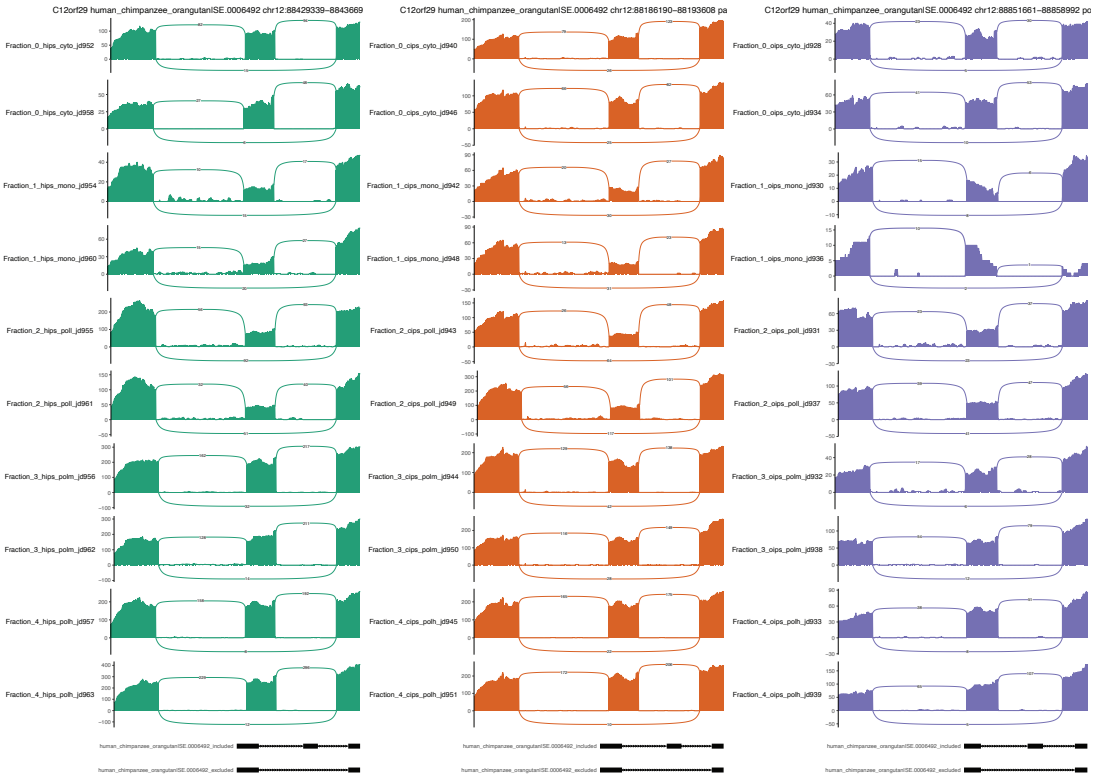

A.
Orthologous Exons
Pairwise alignment
Identification of
single nucleotide variants (SNVs)
matchPWM in sliding window (6-7nt)
calculate binding score
Determine impact of SNV
on RBP binding score
Calculate frequency of RBP sites
affected in test vs control exons
B.
human/chimpanzee
A1CF
DAZAP1
LIN28A
PTBP1
RBM28
SNRNP70
SRSF1
SRSF1
SRSF10
ZCRB1
AS vs AS (ctrl)
ASTC vs ASTC
AS vs ASTC total
log2(relative frequency)
1
0
−1
−2
C.
human/orangutan
CPEB2
CPEB4
ENOX1
ESRP2
FXR2
KHDRBS2
PCBP1
RBM3
RBM42
RBM45
SFPQ
SRSF1
SRSF1
SRSF1
SRSF1
SRSF10
SRSF10
SRSF10
SRSF2
SRSF9
TARDBP
AS vs AS (ctrl)
ASTC vs ASTC
AS vs ASTC total
log2(relative frequency)
1
0
−1
−2
D.
RBP/PWM overlap
human_chimp
10
2
human_orang
21
E.
Event overlap
human_chimp
362
205
human_orang
469

Figure 5

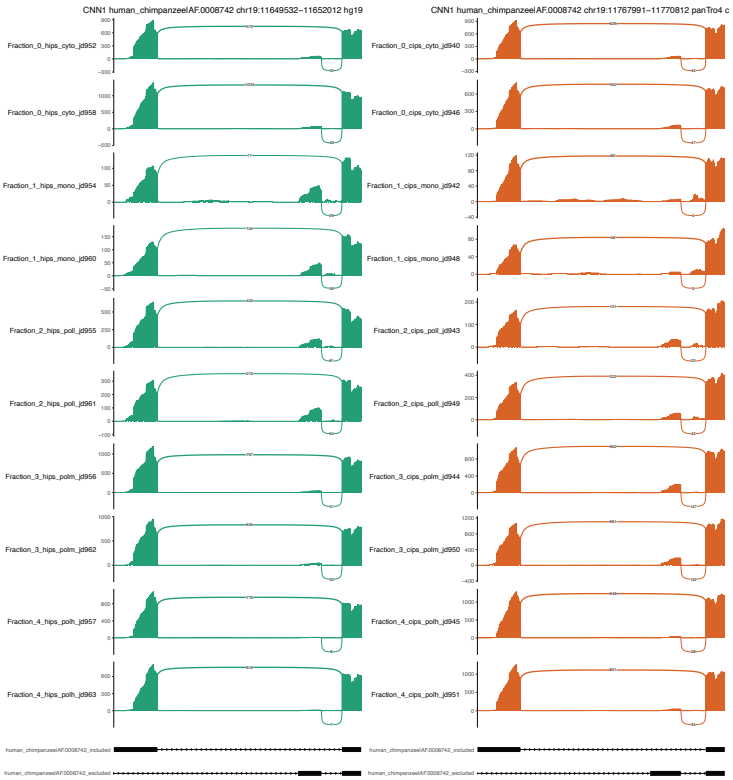


Supplementary Figure 1.

A.



B.



Supplementary Figure 2.

GGCX Exon 2

```
Human   ATAAAGTACAGAAAGACAAGGCTGAACTGATCTCAGGGCCCAGGCAGGACAGCCGAATAGGGAAACTCTTGGGTTTTGAG 80
Chimp   ATAAAGTACAGAAAGACAAGGCTGAACTGATCTCAGGGCCCAGGCAGGACAGCCGAATAGGGAAACTCTTGGGTTTTGAG 80

Human   TGGACAGATTTGTCCAGTTGGCGGAGGCTGGTGACCCTGCTGAATCGACCAACGGACCCTGCAAGCTTAGCTGTCTTTCG 160
Chimp   TGGACAGATTTGTCCAGCTGGCGGAGGCTGGTGACCCTGCTGAATCGACCAACGGACCCTGCAAGCTTAGCTGTCTTTCG 160

Human   TTTTCTTTTTG 171
Chimp   TTTTCTTTTTG 171
```

MSI1
MBLN1
SAMD4A

MELK Exon 3

```
Human   GTGGCTTTGCAAAGGTCAAACTTGCCTGCCATATCCTTACTGGAGAGATGCTAGCTATAAAAATCATGGATAAAAACACA 80
Chimp   GTGGCTTTGCAAAGGTCAAACTTGCCTGCCATATCCTTACTGGAGAGAcGCTAGCTATAAAAATCATGGATAAAAACACA 80

Human   CTAGGGGGATCCTTCAACTTC 101
Chimp   CTAGGGGGATCCTTCAACTTC 101
```

FXR2
RBM3
ANKHD1

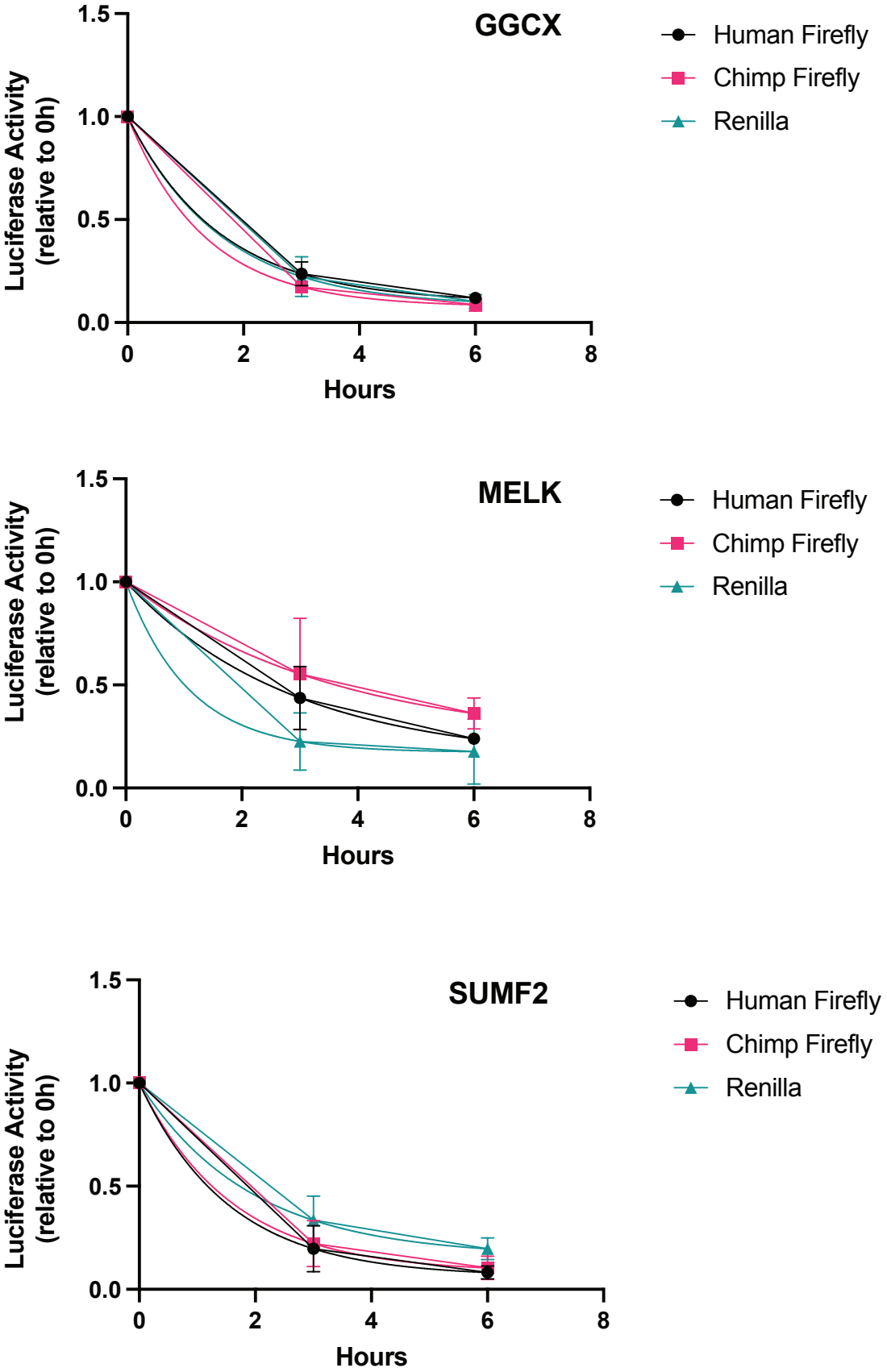
SUMF2 Exon 2

CNOT4
SRSF10
MSI1

```
Human   CCTGCAGGTCCTGGCTCTGGCATCCGAGAGAGACTGGAGCACCCAGTGTTACACGTGAGCTGGAATGACGCCCGTGCCTA 80
Chimp   CCTGCAGGTCCTGGCTCTGGCATCCTAGAGAGACTGGAGCACCCAGTGTTACACGTGAGCTGGAATGACGCCCGTGCCTA 80

Human   CTGTGCTTGGCGGGGAAAACGACTGCCCACGGAGGAAGAGTGGGAGTTTGCCGCCCGAGGGGGCTTGAAGG 151
Chimp   CTGTGCTTGGCGGGGAAAACGACTGCCCACGGAGGAAGAGTGGGAGTTTGCCGCCCGAGGGGGCTTGAAGG 151
```

Supplementary Figure 3



Supplemental Fig. 4