doi:10.1038/nature25757

**Supplementary Note 1:** The observed effects between isoforms are likely smaller than the expected ~2-fold effect of a single SIRLOIN element due to several factors: (a) errors in isoform annotation; (b) difficulties in quantifying isoform-specific expression using short-read RNA-seq; and (c) isoforms do not typically differ just by the SIRLOIN element, but also by additional sequences that likely affect/inhibit nuclear export.

**Supplementary Note 2:** Since the HNRNPK consensus binding motif and the region we found to be essential for SIRLOIN activity in both JPX and PVT1 were C-rich motifs, we tested whether nuclear enrichment can be predicted from sequence alone. We counted the occurrences of all possible hexamers and correlated them with nuclear/cytoplasmic ratios in ENCODE cell lines. C-rich hexamers were among the best correlated with nuclear enrichment in human ENCODE cell lines (**Extended Data Figure 7a-b**) and with reduction in nuclear enrichment following HNRNPK knockdown (**Extended Data Figure 6g**), and correlation between C-rich hexamer content and nuclear enrichment was observed also in mouse liver, mouse embryonic stem cells and MIN6 cells (using data from[3, 1]), with the correlation typically stronger when considering occurrences only in internal exons (**Extended Data Figure 7b**). The nuclear enrichment predicted from sequence alone was quite substantial – 311 genes expressed in HepG2 cells with ≥50 occurrences of C-rich hexamers (>4 Cs) in their internal exons, and at least twofold more C-rich than G-rich motifs, were on average 3.4-fold more nuclear than genes with <50 occurrences (e.g, there are 84 C-rich hexamers in *MLXIPL* and only 10 of G-rich ones). These 311 genes with numerous C-rich hexamers were enriched for transcriptional regulators (GO categories "chromatin organization" and "positive regulation of transcription", q-values $2.7 \cdot 10^{-4}$ and $4.8 \cdot 10^{-3}$, respectively, GOrilla analysis[2]), suggesting that such protein-coding genes, which are known to be generally less stable on both mRNA and protein levels[3], are also more likely to be regulated by the HNRNPK-mediated nuclear enrichment pathway. Notably, a single consensus AluSx element in an antisense orientation has 44 hexamers with >3 Cs, and 5 with >4 Cs (compared to 14 with >3 Gs, and none with >4 Gs), and so integration of Alu elements in an antisense orientation substantially increases the C-rich hexamer content of the host transcript.

**Supplementary Note 3:** We tested the effect of HNRNPK knockdown on localization of transcripts from the NucLibB library. Consistently with the effects on endogenous targets, AcGFP with SIRLOIN-containing tiles became less enriched in the nucleus following HNRNPK depletion, predominantly through reduction in nuclear expression levels (22% on average), slighter increase in cytoplasmic levels (12% on average), and with no significant changes in overall gene expression (1% increase on average, **Figure 4h**). These results suggest that HNRNPK binding drives nuclear enrichment driven by SIRLOIN elements, but other factors likely contribute to decrease in overall expression levels associated with SIRLOIN integration.

**Supplementary Note 4:** The nuclear/cytoplasmic ratios were strongly correlated between mouse liver and human HepG2 hepatocellular carcinoma cells (**Extended Data Figure 9a**, Spearman R=0.51, $P<10^{-50}$) and there was a significant correlation between number of HNRNPK eCLIP clusters in HepG2 and nuclear enrichment in the mouse liver (Spearman R=0.05, $P=3.3 \cdot 10^{-7}$). Also, there is evidence that some of the divergence between nuclear enrichments in the two species can be attributed to SIRLOIN elements. Alu elements are primate-specific, but the mouse B1 repeat family is also derived from 7SL RNA and contains a sequence similar to the Alu SIRLOIN, and which we predict to also be regulated by HNRNPK. Indeed, Mlxipl#71 and Mlxipl#72 in NucLibA, two tiles from mouse Mlxipl 3' UTR that overlaps a B1 element in a negative-strand orientation, were associated with nuclear localization and reduced expression levels (**Extended Data Figure 9b**). These tiles had the SIRLOIN-like sequence

(TGGCCTCCAACTCAGAGATCCACCCACCCCTGCCTCTGG) closer to the 3' end compared to other Mlxipl tiles, and only affected localization and expression when cloned into the 3' UTR, suggesting that the efficacy of this SIRLOIN-like B1-derived element, at least in human cells, is more restricted than that of human SIRLOIN. Both elements appear to have contributed to divergence in expression levels and localization between human and mouse – when homologous genes are compared (see Methods), transcripts that adopted SIRLOIN elements from Alus are more nuclear

and lowly expressed in human hepatic cells, while those that adopted SIRLOIN-like B1 elements in mouse are more nuclear and lowly expressed in mouse (**Extended Data Figure 9c**).

**Supplementary Note 5:** Since HNRNPK is interacting with splicing factors and was shown to affect splicing of specific genes[4], and intron retention is associated with nuclear retention[5], the HNRNPK-induced nuclear enrichment could be mediated by effects on RNA splicing. This is unlikely to be the case, since: (i) The AcGFP mRNA we use as backbone for the libraries is not spliced and yet its localization is affected by HNRNPK knockdown; (ii) Transcripts enriched in the nucleus that are HNRNPK targets or contain SIRLOIN elements were generally well spliced (90% of the 423 genes with >2-fold nuclear enrichment and either a SIRLOIN element or >2 HNRNPK eCLIP clusters had splicing efficiency of >80% in the WCE sample); (iii) HNRNPK knockdown which affected localization of hundreds of genes resulted in very few differential splicing events (1.5 significant events with $P<0.05$ on average per experiment, MISO analysis, see Methods); (iv) HNRNPK knockdown resulted in very mild increase in splicing efficiencies of HNRNPK targets in the nuclear fraction (splicing efficiency increased by 1% on average, $P=0.001$, no significant changes in the WCE and Cytoplasmic fractions), and those genes that changed their localization following the knockdown (at least twofold decrease in nuclear enrichment) were affected similarly to those that did not ($P=0.85$, Wilcoxon rank-sum test).

**Supplementary Note 6:** RNA editing was suggested to play a causal role in nuclear retention of IRAlu-containing RNAs[7]. However, the RNAs we studied are not predicted to form long intramolecular double stranded RNA (dsRNA) regions that would be effective substrates for A-to-I editing (for example, see **Extended Data Figure 5f** for the predicted secondary structures of AcGFP mRNAs with JPX#9 and PVT1#22 in their 3' UTRs). In order to evaluate the extent of RNA editing in our library inserts, we analyzed the signatures of A-to-I RNA editing in the sequencing data of NucLibB. A→G mutations indicative of A-to-I editing were more prevalent than other mutations in cDNA reads when normalized to the input plasmids (**Extended Data Figure 10a**). The overall editing levels of SIRLOIN-containing sequences were typically low (<3%), but higher than the levels in other NucLibB sequences (**Extended Data Figure 10b-d**), and fragments with higher editing levels were enriched in the nucleus and expressed at significantly lower levels (**Extended Data Figure 10e**, 14% average increase, $P=1.8\cdot10^{-4}$; and 43% average decrease, $P=3.5\cdot10^{-9}$ for localization and expression, respectively). However, it is unlikely that A-to-I editing plays a causal role in nuclear enrichment of SIRLOIN-containing sequences, as the per-position editing levels were very low, and >97% of reads sequenced from the nuclear fractions and mapping to SIRLOIN-containing tiles did not contain any edited bases. It is possible that the plausibly longer duration that transcripts enriched in the nucleus spend in the nucleus (compared to those that are efficiently exported) leads to a higher probability of RNA editing, which in the case of SIRLOIN-containing tiles may act on inter-molecular dsRNA formed with RNAs encompassing 'sense strand' Alu elements; alternatively, some other structure or element associated with SIRLOIN elements affects their susceptibility to undergo editing.

## Supplementary References


1 Tan, J. Y. *et al.* Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. *Genome research* **25**, 655-666, doi:10.1101/gr.181974.114 (2015).

2 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics* **10**, 48, doi:10.1186/1471-2105-10-48 (2009).

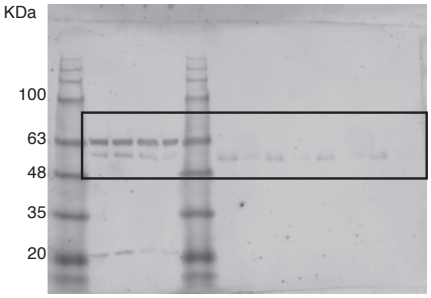
3 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:nature10098 [pii]
10.1038/nature10098 (2011).

4 Bomsztyk, K., Denisenko, O. & Ostrowski, J. hnRNP K: one protein multiple processes. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**, 629-638, doi:10.1002/bies.20048 (2004).

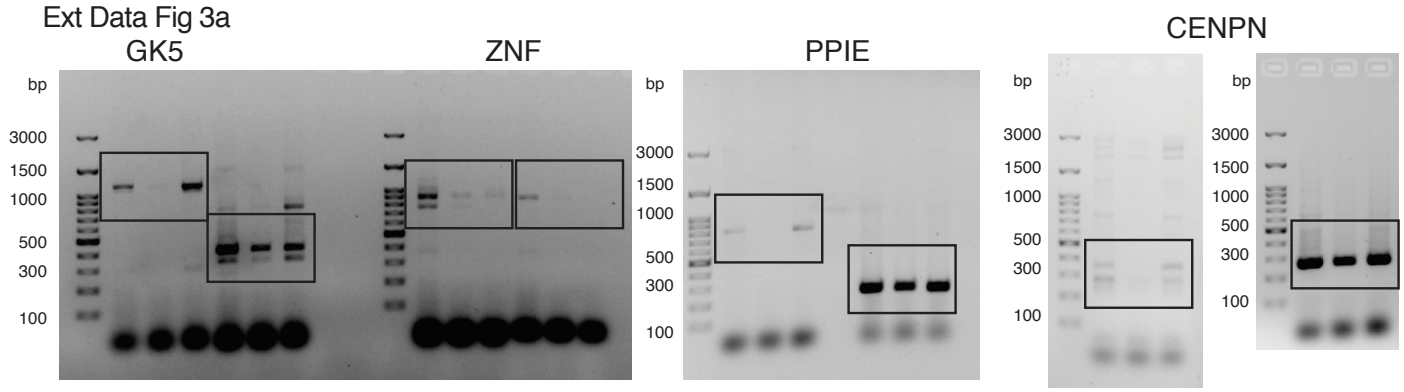
5 Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome research* **24**, 1774-1786, doi:10.1101/gr.177790.114 (2014).

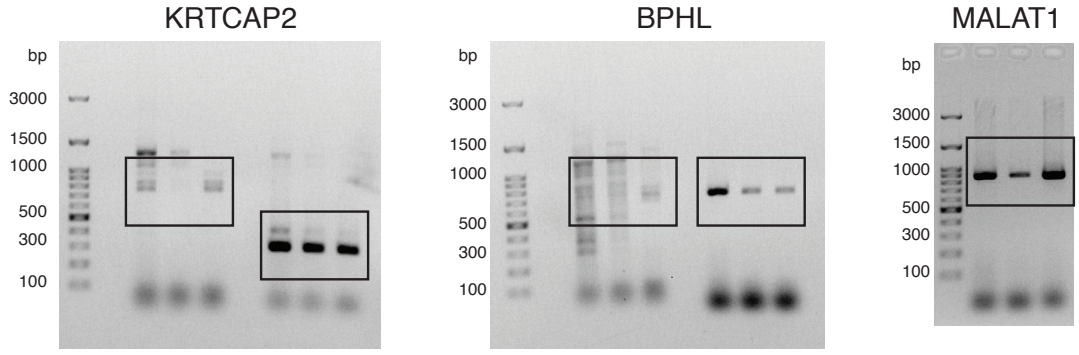
Supplementary Figure - Uncropped scans with size marker indications

Fig 4c

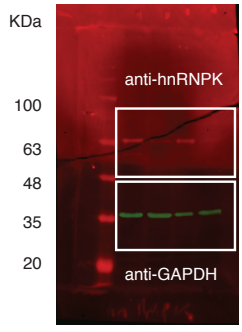



Ext Data Fig 3a




Ext Data Fig 7b




Ext Data Fig 9a