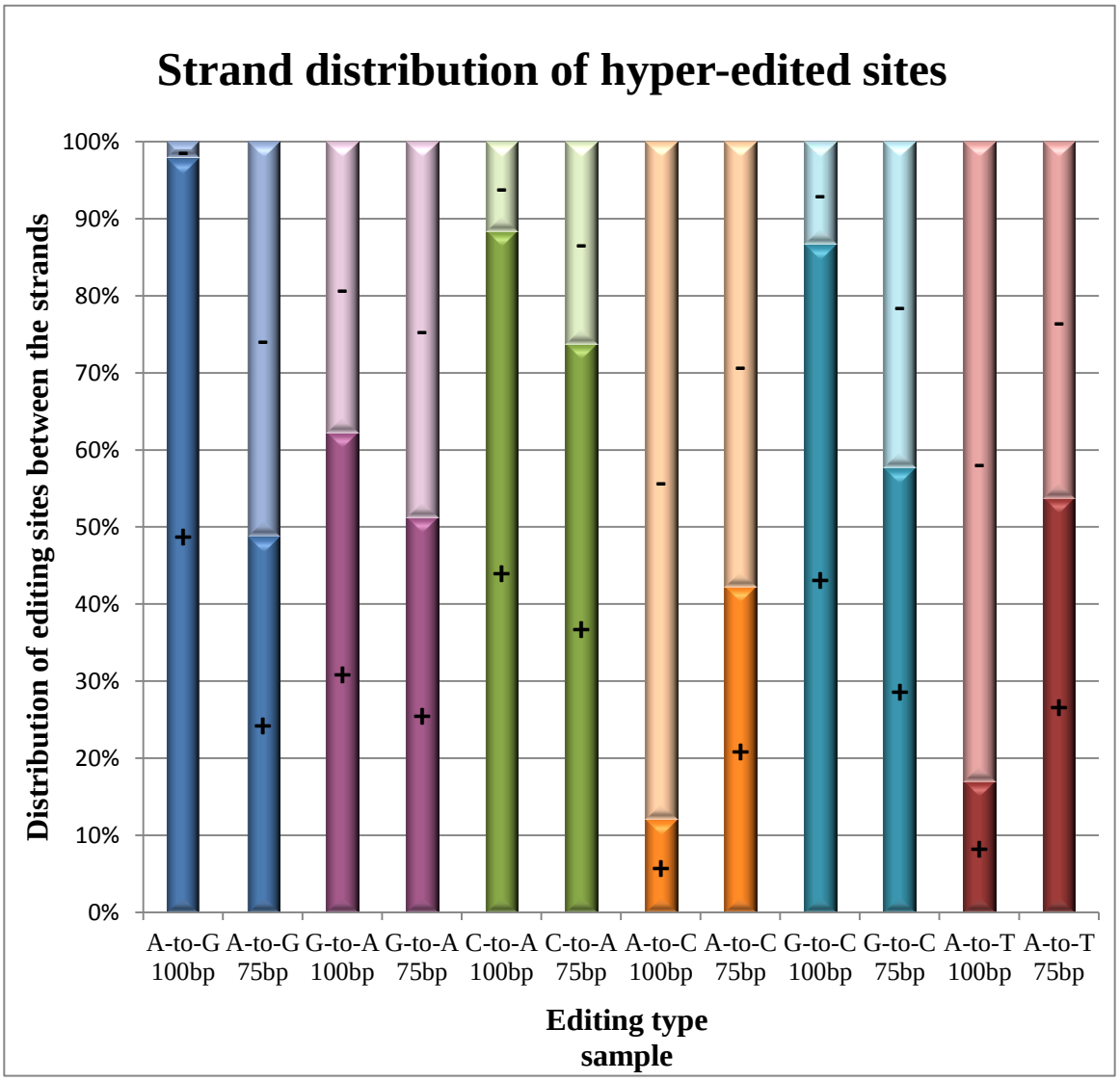
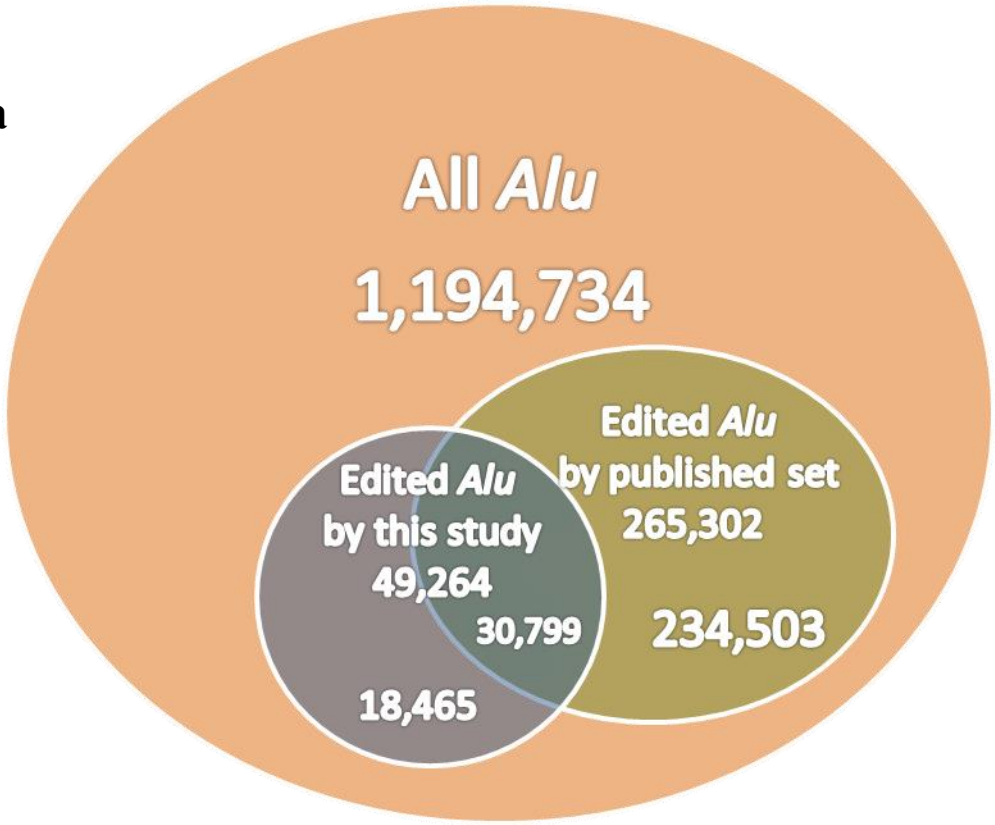


Supplementary Figure 1. **Strand distribution of the detected editing sites.** Each pair of bars of the same color refers to a specific type of mismatch and to either the 100bp or the 75bp dataset. Each bar shows the fraction of editing sites that were either (+), the specified mismatch (e.g., A-to-G), or (-), the complementary mismatch (e.g., T-to-C). The 100bp dataset was stranded (the sequenced strand was the expressed strand), whereas the 75bp dataset was not (the sequenced strand was arbitrary). Only the A-to-G sites show the expected behavior from true editing sites, namely, near 100% A-to-G preference at the stranded sample (100bp) and random distribution (≈50/50%) at the non-stranded sample.

a



b

	All *Alu* elements in the genome[3]	Moderately edited *Alu* elements[3]	Hyper-edited *Alu* elements[3]	P-Value (Wilcoxon rank sum test)[4]
No. of *Alu* elements	1,194,734	234,503	49,264	
Distance between the edited *Alu* and the nearest inverted *Alu* (bp)[1]	1845 ± 2099	1130±1449	814±1103	$<1\times10^{-300}$
Minimum of (no. *Alu*+, no. *Alu*−) in the region[2]	6.03 ± 4.81	7.28±4.54	8.13±4.60	$<1\times10^{-300}$

Supplementary Figure 2. The hyper-edited *Alu*s are part of particularly long dsRNA structure. (a) A Venn diagram showing the overlap between the 49,264 hyper-edited *Alu*s and the 265,302 edited *Alu*s that were identified by Ramaswami et al. 2013, as well as the total number of *Alu* elements in the human genome (RepeatMasker). (b) Comparison of the properties of hyper-edited and moderately edited *Alu*s.

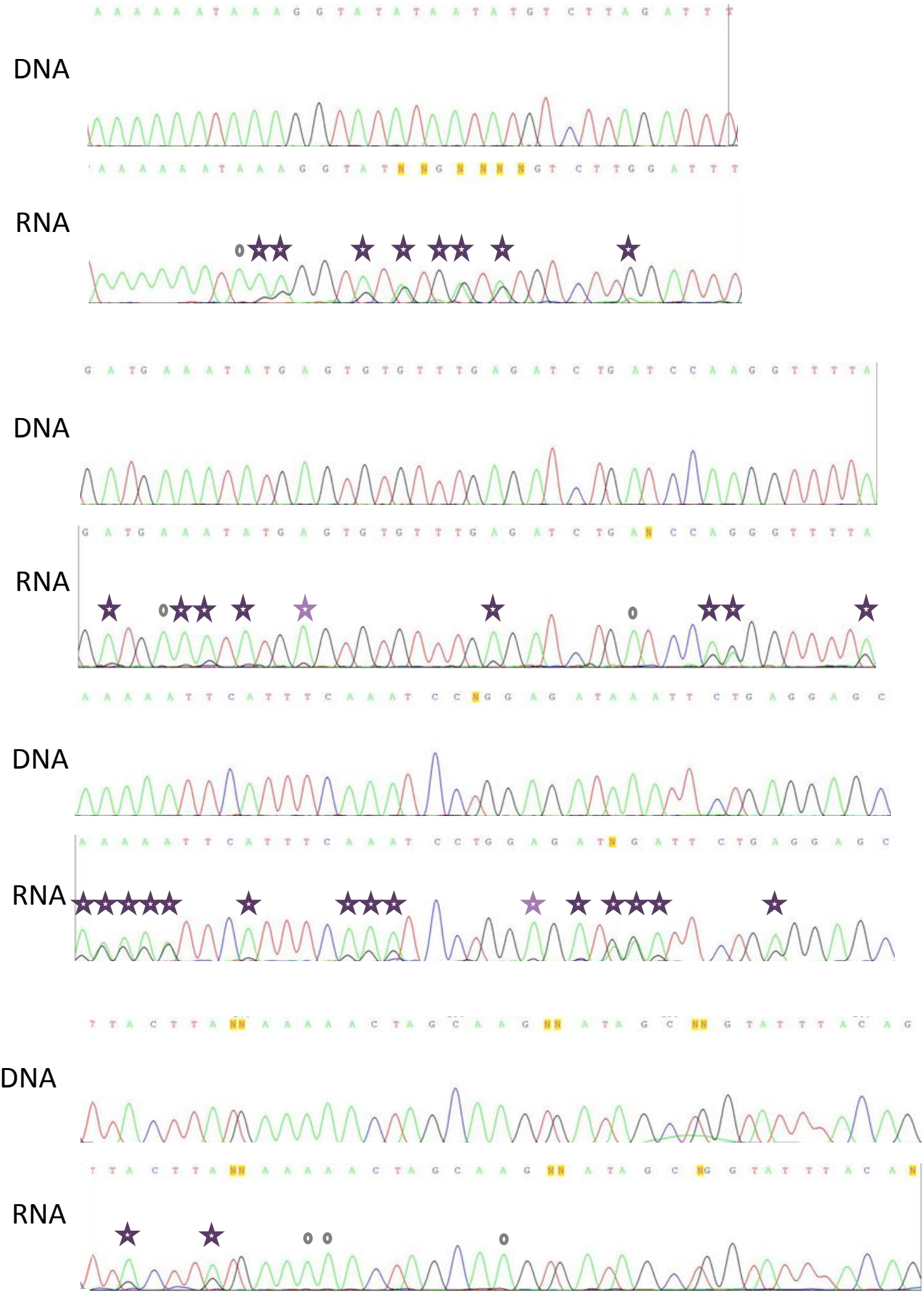
[1]Nearest inverted *Alu* at distance <10kbp upstream or downstream.

[2]Regions used: ±10kb of the edited *Alu* element.

[3]Means are reported along with standard deviations.

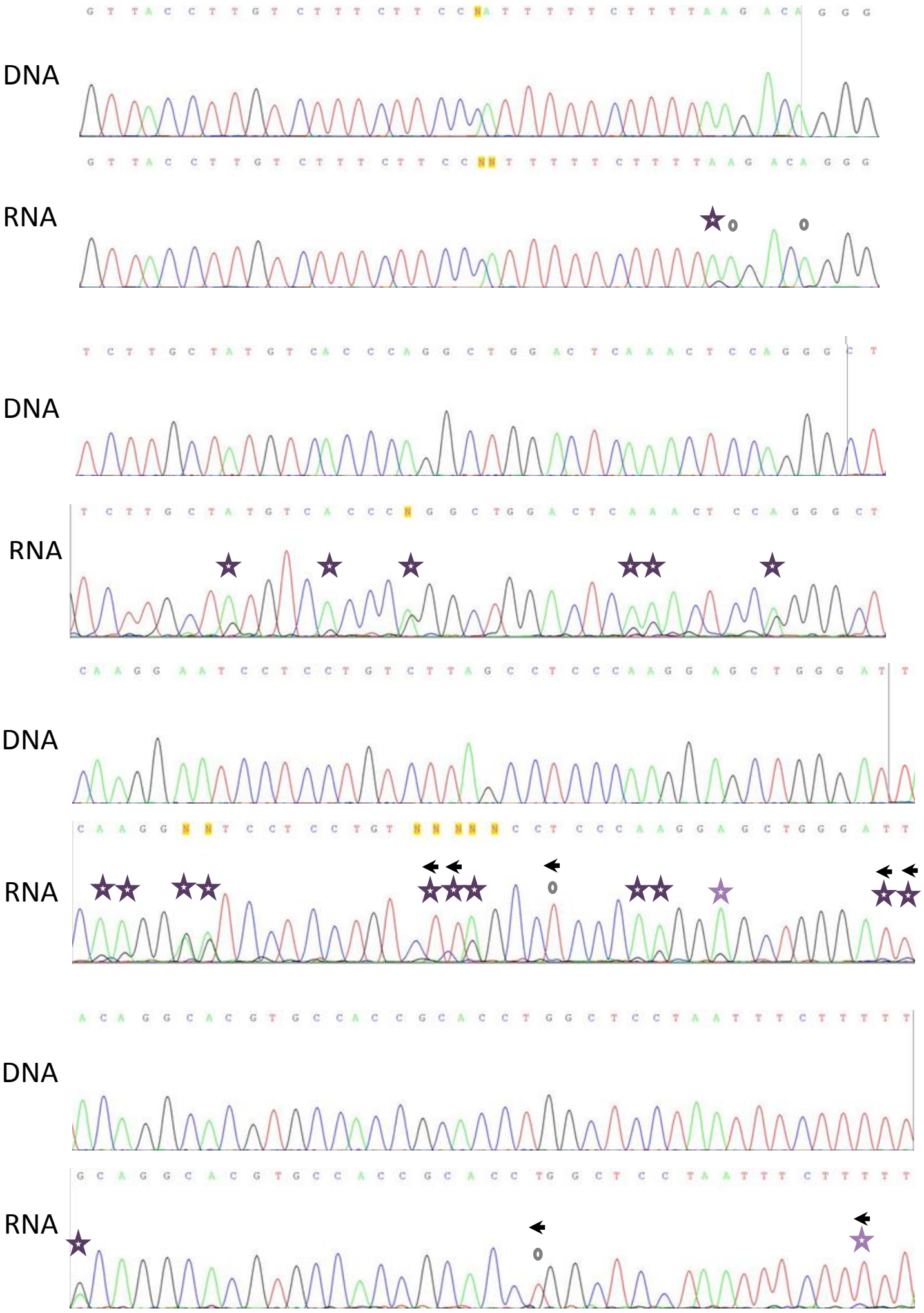
[4]Compring the moderately edited *Alu* to the hyper-edited *Alu*.

a **> hg19 chrX:84346270-84346424 (+) ; APOOL 3' UTR**



b

>hg19 chr12:99128730-99128890(-) ; ANKS1B 3' UTR (sense) ; APAF1 3' UTR (antisense)



Supplementary Figure 3. Sanger sequencing chromatograms of hyper-edited regions. Chromatograms of genomic DNA are presented along with those of the corresponding cDNA (RNA). Bases are colored as: A- green, C- blue, G- black, T- red. Editing sites are evident as nucleotides having an A in the DNA and a G (or a mixed A/G signal) in the RNA. We annotated the observed editing sites that were computationally predicted with dark purple stars, and new observed editing sites with light purple stars. Grey ellipses indicate predicted sites that were not validated. **(a)** A chromatogram of chrX:84346270-84346424 (+), in the 3' UTR of the APOOL gene, which is the non-repetitive hyper-edited region presented in Figure 4 of the main text. We identified 34 editing sites (of which 32 were predicted); six sites were predicted but not validated. At the edited cluster (between the first and the last edited A), 34/39 (87.2%) of the adenosines were edited. **(b)** A chromatogram of chr12:99128730-99128890 (−), in the 3' UTR of both the ANKS1B (sense) and the APAF1 (antisense) genes. At ANKS1B (sense), where editing appears as an A in the DNA and a G (or A/G) in the RNA, we identified 16 edited sites (of which 15 were predicted); two sites were predicted but not validated. At APAF1 (antisense), where editing appears as a T in the DNA and a C (or T/C) in the RNA (indicated by backward pointing arrows), we identified five editing sites (of which four were predicted); two sites were predicted but not validated.

Supplementary Table 1. Hyper-editing detected in the Illumina BodyMap 75bp dataset.

Editing type	Hyper-edited reads (n)	Merged regions (n)	Editing sites (n)	Unique editing sites (n)	Unique editing sites (% of all types)	Sites overlapping with dbSNP (n)	Sites overlapping with dbSNP (% of all sites, for each type)
A-to-G	390,881	62,860	2,080,519	455,014	97.25%	7,448	1.64%
G-to-A	19,163	1,169	88,275	7,664	1.64%	831	10.84%
A-to-C	2,421	272	10,880	1,580	0.34%	100	6.33%
A-to-T	491	282	2,580	1,528	0.33%	68	4.45%
C-to-A	452	231	2,014	1,092	0.23%	63	5.77%
G-to-C	1,490	197	6,196	1,016	0.22%	115	11.32%

Supplementary Table 2. Differential hyper-editing among tissues.

Tissue	Hyper-edited reads	Hyper-edited regions	Enrichmnet[1]	Unique editing sites
Brain	22,117	12,650	2.37	75,984
Adipose	32,074	10,302	1.64	68,677
Adrenal	28,005	10,163	1.61	60,789
Prostate	48,479	10,734	1.56	71,255
Testes	28,166	9,918	1.47	63,565
Lung	46,136	8,796	1.33	57,611
Colon	24,801	7,217	1.08	45,174
Lymph	36,081	6,551	1.00	38,702
Ovary	19,529	6,205	0.92	38,876
Breast	17,232	5,420	0.85	34,492
Thyroid	20,274	5,384	0.82	34,171
Kidney	18,580	3,783	0.58	23,068
Liver	7,369	2,158	0.34	13,723
Heart	7,542	1,928	0.30	10,636
White blood cells	19,142	1,714	0.25	10,347
Skeletal muscle	15,354	916	0.14	5,230
Total	**390,881**	**103,839**		**652,300**
Unique		**62,860**		**455,014**

[1]The enrichment is equal to the number of hyper-edited regions in each tissue divided by the expected number, which was computed by multiplying the total number of hyper-edited regions (over all tissues) by the ratio of the number of mapped reads in the tissue to the number of mapped reads in all tissues [2].

[2] Carmi, S., Borukhov, I. & Levanon, E. Y. Identification of widespread ultra-edited human RNAs. PLoS Genet. 7, e1002317 (2011).

Supplementary Table 3. Published methods discard the vast majority of editing sites

Article	Description of dataset	Accession numbers	Source reads[1]	Editing sites, as reported by authors	Hyper-editing sites we detected (%A-to-G of total) [novel]	Ratio of number of sites: Our screen /published article
HUMAN						
Identifying RNA editing sites using RNA sequencing data alone (Ramaswami et al. 2013)	Illumina Human BodyMap 2.0 75bp SE and 50bp PE	GSE30611: ERR030888-903, ERR030872-887	3,821,002,610	370,623	546,732 (95.71%) [478,739]	1.48
RNA editing in the human ENCODE RNA-seq data (Park et al. 2012)	ENCODE Project cell line GM12878 75bp PE	GSM958728: SRR521447-456	424,451,564	1,716	157,077 (96.04%) [156,104]	91.54
RNA editing in the human ENCODE RNA-seq data (Park et al. 2012)	ENCODE Project cell line GM12891 75bp PE	GSM958747: SRR521531-33	199,366,592	1,885	97,681 (98.28%) [96,712]	51.82
RNA editing in the human ENCODE RNA-seq data (Park et al. 2012)	ENCODE Project cell line GM12892 75bp PE	GSM958748: SRR521534-37	270,307,390	843	125,146 (98.35%) [124,626]	148.45
Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome (Peng et al. 2012)	Lymphoblastoid cell line (LCL) of an anonymous male Han Chinese individual (YH) PolyA+ 75 bp PE and 100 bp PE	SRS254654: SRR324678-685	319,075,474	10,343	73,463 (98.73%) [73,462]	7.10
Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome (Peng et al. 2012)	Lymphoblastoid cell line (LCL) of an anonymous male Han Chinese individual (YH) PolyA- (ss) 90 bp PE	SRS254691: SRR324687, SRR325616 and SRR324688	843,673,098	10,770	270,853 (98.81%) [270,850]	25.15
Accurate identification of A-to-I RNA editing in human by transcriptome sequencing (Bahn et al., 2012)	U87MG cell line derived from a human grade IV glioma ADAR+ (ctrl) 60bp PE	GSE28040 (GSM693746) SRR388226-7	111,160,762	4,141	27,124 (94.64%) [25,945]	6.55
MOUSE						
High levels of RNA-editing site conservation amongst 15 laboratory mouse strains (Danecek et al. 2012**)**	Mouse brains C57BL/6NJ strain 76bp PE	ERP000614: ERR033015-16	114,374,684	4,869	11,849 (96.38%) [11,811]	2.43
FLY						
Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila (Rodriguez et al. 2012)	Drosophila melanogaster head nascent RNA of ADAR+ strains: yw and FM7 males 101bp SE	GSE37232: GSM914114-7	257,255,489	1,350	39,472 (99.87%) [39,334]	29.24

[1] Paired-end reads were counted as two single-end reads.

Supplementary Table 4. Identification of hyper-edited RNAs in different species

Species	Reference	Description	Accession numbers	Source reads[1] [Mapped reads]	Editing sites (% A-to-G of total)	Editing sites per mapped read
Human	Illumina Human BodyMap project 2.0	Homo sapiens brain[2] 75bp SE	ERR030890	64,313,204 [59,130,196]	75,984 (98.93%)	$1.3 \cdot 10^{-3}$
Mouse	High levels of RNA-editing site conservation amongst 15 laboratory mouse strains (Danecek et al. 2012)	Mus musculus brain 76bp PE	ERP000614: ERR033015-6	114,374,684 [103,507,244]	11,849 (96.38%)	$1.1 \cdot 10^{-4}$
Rat	Evolutionary dynamics of gene and isoform regulation in Mammalian tissues (Merkin et al. 2012)	Rattus norvegicus brain 80bp PE	GSM1020666	238,077,800 [199,514,577]	21,761 (93.61%)	$1.1 \cdot 10^{-4}$
Opossum	The evolution of gene expression levels in mammalian organs (Brawand, D. et al. 2011)	Monodelphis domestica brain 76bp SE	GSM752590	22,273,667 [14,411,371]	1,563 (89.88%)	$1.1 \cdot 10^{-4}$
Drosophila	Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila (Rodriguez et al. 2012)	Drosophila melanogaster head 101bp SE	GSE37232: GSM914114-7	257,255,489 [248,984,465]	39,472 (99.87%)	$1.6 \cdot 10^{-4}$

[1] Paired-end reads were counted as two single-end reads.

[2] Values are reported only for the brain-derived subset of reads, to facilitate comparison with the other species.

Supplementary Table 5. Targets and primers for direct sequencing validation.

Amplified region [size]	Primer sequences (5' to 3') (top row: left, bottom row: right)	Functional region [Repeat element]	Annealing Tm	Gene name	Number of observed editing sites [cluster size[1]]
chr4: 89181173-89181565 [393bp]	GCATTCAAAACAAAAATGAAACCTGC	3'UTR [Alu]	66°c	PPM1K	23 [252bp]
	CCAAACCAACAGGCAAAGGA				
chr1: 145511836-145512201 [366bp]	TGTGAGGAGTATAAAAATAGTCAAAGC	3'UTR/ intron [Alu]	62°c	NBPF10/ RBM8A	21 [252bp]
	TGGACCAAACTATAAATCCAAGAGC				
chr14: 31563509-31563791 [283bp]	AAGCTTTGTAATGTGGGGCT	3'UTR [Alu]	62°c	AP4S1	11 [113bp]
	TTTCACTCTGTCACCCAGGC				
chr12: 99128610-99128962 [353bp]	GGGGCTCATCTCATGTAGGC	3'UTR [Alu]	66°c	ANKS1B (−)	16 (-)[2] [92bp]
	CCTCTTGACCGACTGCATGA			APAF1 (+)	5 (+)[2] [63bp]
chrX: 84346226-84346481 [256bp]	AGCCTGGGCACAAGAACAAA	3'UTR [NonRep]	62°c	APOOL	34[2] [161bp]
	CAAGGAACCTTGTCTGTGAGA				
chr10: 15118871-15119501 [631bp]	AGCCAAAAGAACAAACAGCTGT	3'UTR/ intron [RepNonAlu]	64°c	ACBD7	10 [179bp]
	AATTGCTCATCCCTCACCCC				
chr1: 149913606-149913817 [212bp]	TCTCAGTCCTGTCAGATCCTGT	3'UTR [RepNonAlu]	62°c	OTUD7B	11 [43bp]
	GCAGTGAGATTCCAAAGGAAGG				

[1] The number of bases between the first and the last edited A.
[2] Chromatograms of the Sanger sequencing are presented in Supplementary Figure 3.