(SUPPLEMENTAL DATA)

CeFra-seq reveals broad asymmetric mRNA and non-coding RNA distribution profiles in *Drosophila* and human cells.

Louis Philip Benoit Bouvrette[1,2], Neal A.L. Cody[1], Julie Bergalet[1], Fabio Alexis Lefebvre[1,2], Cédric Diot[1,2], Xiaofeng Wang[1], Mathieu Blanchette[4], Eric Lécuyer[1,2,3,5]

1- Institut de Recherches Clinique de Montréal (IRCM), Montréal, Canada.
2- Département de Biochimie, Université de Montréal, Montréal, Canada.
3- Division of Experimental Medicine, McGill University, Montréal, Canada.
4- McGill School of Computer Science, McGill University, Montréal, Canada.

5-Corresponding author:

Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada, H2W 1R7
Email: eric.lecuyer@ircm.qc.ca

Table S1: RNA-seq read statistics for Human HepG2 and *Drosophila* D17 subcellular fractions.

Library	Species	Fraction	Replicate	Number of reads	Duplicate (%)	Number of Aligned reads	Alignment (%)	Multiple alignment	Average FPKM
rRNA-depletion	*Drosophila*	Cytosolic	1	20,698,062	68.32	18,876,632	91.2	53.4	$1.80x10^6$
rRNA-depletion	*Drosophila*	Cytosolic	2	24,285,303	71.73	22,488,190	92.6	55.2	
rRNA-depletion	*Drosophila*	Membrane	1	21,843,697	43.58	19,069,547	87.3	17.1	$0.61x10^6$
rRNA-depletion	*Drosophila*	Membrane	2	21,134,180	44.12	18,513,541	87.6	18.4	
rRNA-depletion	*Drosophila*	Insoluble	1	22,607,476	30.85	19,781,541	87.5	23.3	$0.34x10^6$
rRNA-depletion	*Drosophila*	Insoluble	2	24,059,152	32.19	20,979,580	87.2	24.4	
rRNA-depletion	*Drosophila*	Nuclear	1	22,996,271	47.03	18,902,934	82.2	56.8	$0.44x10^6$
rRNA-depletion	*Drosophila*	Nuclear	2	24,902,681	43.5	20,694,127	83.1	52.8	
rRNA-depletion	Human	Cytosolic	1	32,552,634	64.76	27,344,212	84	24.6	$3.20x10^6$
rRNA-depletion	Human	Cytosolic	2	30,564,649	67.55	25,246,400	82.6	24.3	
rRNA-depletion	Human	Membrane	1	35,909,087	40.04	30,774,087	85.7	17.5	$0.62x10^6$
rRNA-depletion	Human	Membrane	2	35,061,741	39.84	30,819,270	87.9	16.5	
rRNA-depletion	Human	Insoluble	1	34,097,076	20.53	31,539,795	92.5	8.1	$0.26x10^6$
rRNA-depletion	Human	Insoluble	2	35,451,428	22.57	32,757,119	92.4	7.6	
rRNA-depletion	Human	Nuclear	1	35,747,174	10.16	32,065,215	89.7	7	$0.24x10^6$
rRNA-depletion	Human	Nuclear	2	37,390,540	11.56	33,913,219	90.7	6.8	
PolyA+	*Drosophila*	Cytosolic	1	22,668,517	53.28	20,628,350	91	7.4	$2.10x10^6$
PolyA+	*Drosophila*	Cytosolic	2	19,499,648	53.13	17,569,182	90.1	8	
PolyA+	*Drosophila*	Membrane	1	25,331,304	55.33	22,139,559	87.4	7.8	$1.30x10^6$
PolyA+	*Drosophila*	Membrane	2	21,260,027	51.94	18,474,963	86.9	9.8	
PolyA+	*Drosophila*	Insoluble	1	20,515,878	42.19	17,664,170	86.1	34.7	$0.42x10^6$
PolyA+	*Drosophila*	Insoluble	2	20,678,208	41.61	17,783,258	86	32.9	
PolyA+	*Drosophila*	Nuclear	1	23,076,204	48.43	18,460,963	80	55.7	$0.25x10^6$
PolyA+	*Drosophila*	Nuclear	2	21,195,707	47.52	16,998,957	80.2	54.5	
PolyA+	Human	Cytosolic	1	31,719,232	53.46	30,101,551	94.9	6.8	$0.99x10^6$
PolyA+	Human	Cytosolic	2	24,755,754	49.65	22,948,583	92.7	6.4	
PolyA+	Human	Membrane	1	24,235,894	37.53	21,885,012	90.3	13.5	$0.46x10^6$
PolyA+	Human	Membrane	2	20,236,753	35.82	18,840,417	93.1	6.4	
PolyA+	Human	Insoluble	1	25,382,072	17.75	22,996,157	90.6	12.6	$0.21x10^6$
PolyA+	Human	Insoluble	2	23,337,888	17.33	21,354,167	91.5	12.2	
PolyA+	Human	Nuclear	1	21,059,227	11.22	19,795,673	94	6.5	$0.19x10^6$
PolyA+	Human	Nuclear	2	19,884,506	12.69	18,731,204	94.2	6.6	

Table S2. Cell component gene ontology (GO) enrichments of HepG2 cell fraction-specific proteins.

Cytosolic		Membrane	
Gene Category	**p-Value**	**Gene Category**	**p-Value**
cytosol	3.37E-81	endomembrane system	3.34E-47
cytoplasm	6.01E-57	bounding membrane of organelle	3.43E-39
extracellular vesicle	2.48E-50	endoplasmic reticulum	4.40E-37
membrane-bounded vesicle	3.94E-36	vesicle	1.70E-23
cytoplasmic part	1.05E-34	extracellular exosome	2.18E-22
vesicle	8.38E-34	extracellular membrane-bounded organelle	2.24E-22
extracellular region	3.46E-22	extracellular vesicle	3.06E-22
nucleus	1.64E-11	extracellular organelle	3.14E-22
nucleoplasm	1.65E-05	cytoplasm	5.34E-18
nuclear lumen	6.17E-05	membrane-bounded organelle	2.34E-17
nuclear part	7.35E-05	extracellular region part	2.77E-16
intracellular membrane-bounded organelle	4.80E-04	Golgi apparatus	3.99E-14
mitochondrion	6.90E-04	integral component of membrane	6.22E-14
intracellular organelle	1.64E-03	intrinsic component of membrane	2.55E-13
intracellular organelle lumen	2.52E-03	endosome	2.02E-08
		cytoplasmic vesicle	5.05E-08
		vacuole	6.52E-07
		vacuolar part	1.14E-05
		lysosome	4.41E-05
		lytic vacuole	4.71E-05
		whole membrane	5.83E-05

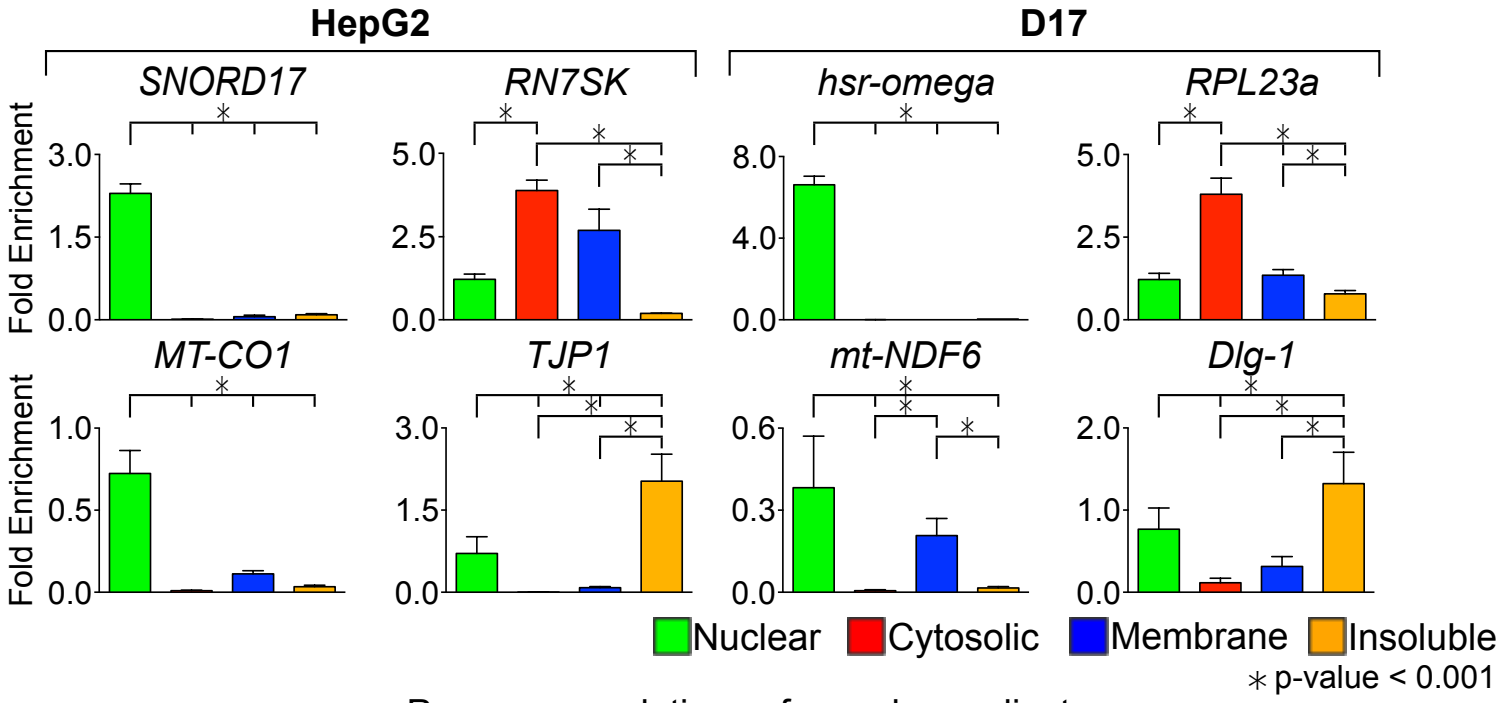
Insoluble		Nucleus	
Gene Category	**p-Value**	**Gene Category**	**p-Value**
macromolecular complex	1.89E-08	membrane-enclosed lumen	3.48E-97
cytoplasmic part	4.84E-08	nuclear part	3.74E-92
ribonucleoprotein complex	4.86E-08	nucleus	8.07E-66
intracellular ribonucleoprotein complex	4.86E-08	membrane-bounded organelle	9.55E-64
intracellular organelle part	6.79E-06	nucleoplasm	5.96E-44
organelle part	2.38E-05	nucleolus	1.34E-42
cytoplasm	2.59E-05	chromosome	4.33E-36
intracellular non-membrane-bounded organelle	1.76E-04	organelle envelope	4.86E-34
cytosol	1.87E-04	envelope	1.51E-33
non-membrane-bounded organelle	1.92E-04	chromosomal part	9.82E-32
intracellular part	1.95E-04	protein complex	1.11E-27
organelle	2.77E-04	mitochondrial part	4.09E-26
membrane-bounded organelle	4.48E-04	preribosome	1.86E-25
RNAi effector complex	4.94E-04	chromatin	1.36E-20
cytoplasmic ribonucleoprotein granule	5.25E-04	chromosomal region	1.39E-19
intracellular membrane-bounded organelle	6.33E-04	nuclear chromosome	2.14E-19
ribonucleoprotein granule	7.52E-04	ribonucleoprotein complex	2.88E-15
polysome	1.98E-03	intracellular ribonucleoprotein complex	2.88E-15
protein complex	2.71E-03	chromosome, centromeric region	5.05E-15
		nuclear pore	3.38E-14
		SWI/SNF superfamily-type complex	4.31E-14
		condensed chromosome	2.23E-08
		kinetochore	2.64E-08
		nuclear envelope	2.71E-07
		histone methyltransferase complex	3.18E-07
		nuclear pore outer ring	5.18E-07
		chromosome, telomeric region	2.11E-06
		PcG protein complex	4.45E-06
		transcriptional repressor complex	5.11E-06
		methyltransferase complex	5.11E-06
		transferase complex	7.19E-06

Table S3. Cell component gene ontology (GO) enrichments of HepG2 cell fraction-specific mRNAs.

Cytosolic		Membrane	
Gene Category	**p-Value**	**Gene Category**	**p-Value**
ribosome	3.50E-50	cytosolic ribosome	3.90E-31
cytosolic part	3.20E-46	ribosomal subunit	3.00E-22
ribonucleoprotein complex	5.80E-35	cytosolic part	7.20E-22
small ribosomal subunit	2.10E-28	ribosome	1.40E-18
large ribosomal subunit	5.70E-27	cytosol	3.50E-13
cytosol	6.10E-19	large ribosomal subunit	5.00E-13
mitochondrion	5.30E-15	small ribosomal subunit	2.30E-09
mitochondrial part	3.90E-11	ribonucleoprotein complex	5.50E-09
mitochondrial membrane	4.90E-10	endoplasmic reticulum	3.40E-05
mitochondrial envelope	1.10E-09	endosome	8.10E-05
mitochondrial ribosome	3.80E-05	transcription factor complex	4.30E-04
non-membrane-bounded organelle	1.00E-04	nucleoplasm part	8.70E-04
respiratory chain	4.10E-04	extrinsic to membrane	1.10E-03
		Golgi apparatus part	7.10E-03
		organelle subcompartment	7.10E-03

Insoluble		Nuclear	
Gene Category	**p-Value**	**Gene Category**	**p-Value**
nucleoplasm	1.20E-43	extracellular matrix part	5.40E-06
nucleolus	1.00E-31	nuclear lumen	6.60E-06
chromosome	3.00E-24	membrane-enclosed lumen	1.40E-05
microtubule cytoskeleton	2.50E-22	basement membrane	2.20E-05
nucleoplasm part	2.10E-20	nuclear speck	2.70E-05
nuclear body	3.00E-20	nucleoplasm	6.50E-05
spliceosome	1.10E-19	endomembrane system	8.40E-05
spindle	3.70E-19	endoplasmic reticulum part	9.30E-05
chromosomal part	6.50E-18	collagen	1.00E-04
condensed chromosome	1.10E-17	nucleoplasm part	2.90E-04
ribonucleoprotein complex	9.10E-16	anchoring collagen	2.30E-03
cytoskeleton	1.20E-15	cell-cell adherens junction	2.40E-03
kinetochore	7.10E-14	nuclear body	3.40E-03
nuclear speck	1.70E-12	anchoring junction	4.50E-03
nuclear chromosome	2.60E-12	microtubule cytoskeleton	4.60E-03
nuclear pore	3.00E-10	nuclear envelope	4.80E-03
microtubule organizing center	6.20E-10	cell-cell junction	7.30E-03
centrosome	9.30E-10		
nuclear envelope	1.90E-08		
proteasome complex	5.50E-08		

A



B

Pearson correlations of samples replicate

	HepG2		D17	
	PA	RD	PA	RD
Cytosolic	0.999	0.999	0.994	0.997
Membrane	0.998	0.999	0.995	0.996
Insoluble	0.998	0.999	0.995	0.985
Nuclear	0.996	0.994	0.997	0.996

Figure S1: **Fractionation validation and inter-replicate correlations.**
(A) RT-qPCR of RNAs sample controls show fractionation efficiency.
The accumulation of the indicated RNA fraction-specific markers was assessed in HepG2 and D17 cells.
(B) Summary table of inter-replicates Pearson correlations of transcript per million (TPM) values within each fractions for HepG2 and D17 cell RNA-seq samples. RD= rRNA-depletion, PA= poly-A+.

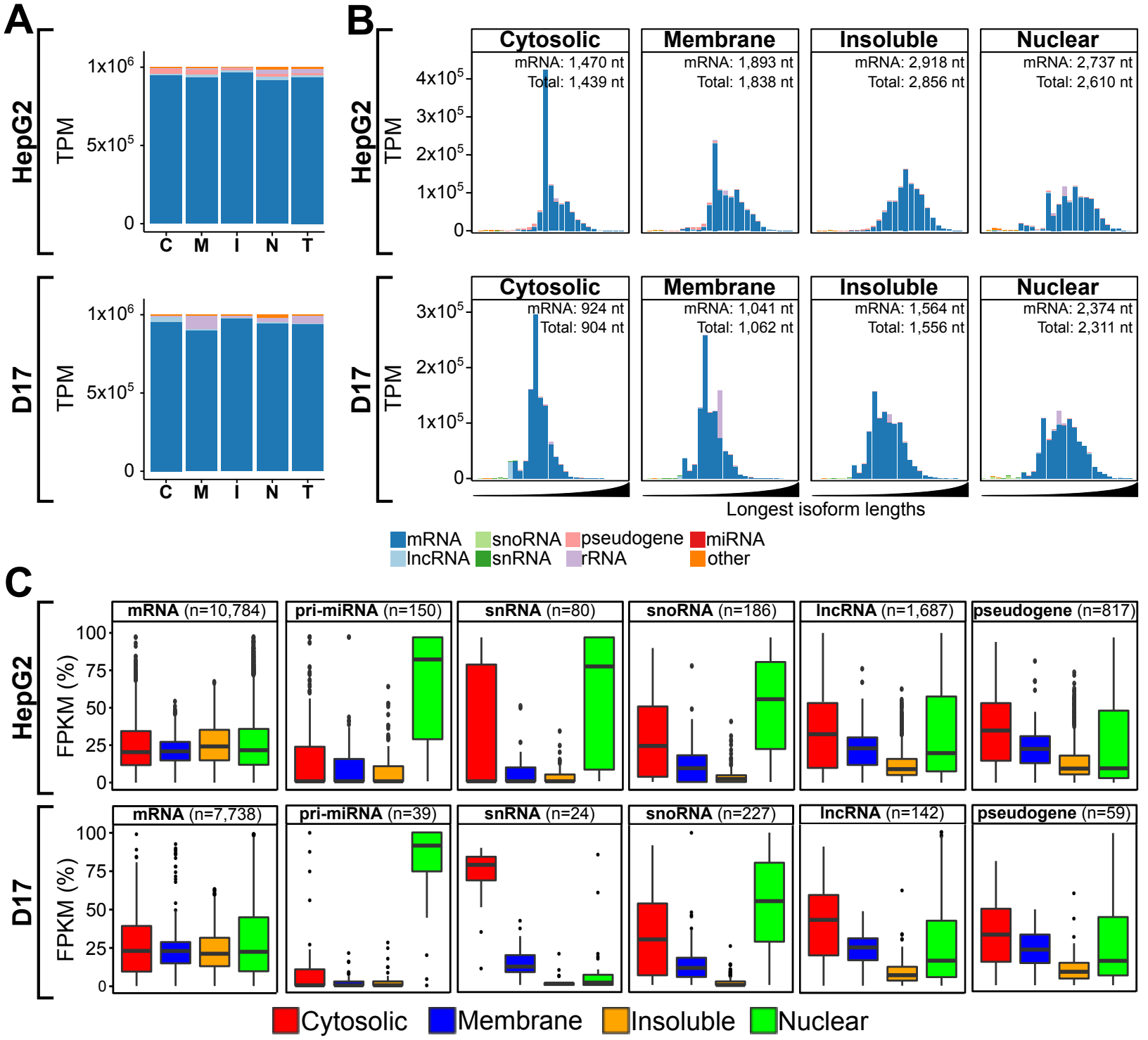


Figure S2: **Distinctive transcript composition of subcellular fractions in poly-A+ dataset.** (A) Histograms depicting the RNA biotype content, in TPM, detected via PA sequencing of cytosolic (C), membrane (M), insoluble (I), and Nuclear (N) fractions or whole-cell RNA (T=total) from HepG2 (upper panel) and D17 cells (lower panel). (B) Histograms of the RNA biotype content of HepG2 and D17 cell fractions, binned according to the length of the longest annotated isoform of detected RNA species, following a log10 scale from 1.5-5 (i.e. ranging from 31-100,000 nt). The expected lengths for mRNA and total RNA populations are indicated for each fraction. For A and B, biotypes accounting for less than 1% of the overall TPMs were grouped as “other”. (C) Boxplots showing the fraction distribution profiles of different RNA biotypes in percent FPKM (pFPKM) for HepG2 (upper plots) and D17 (lower plots) cells. The number (n) of transcripts analyzed for each biotype is indicated. TPM: Transcripts Per Millions; FPKM: Fragment per kilobase per million; lncRNA: long noncoding RNA; mRNA: messenger RNA; miRNA: microRNA; miscRNA: miscellaneous other non-coding RNA; snoRNA: small nucleolar RNA; snRNA: small nuclear RNA; rRNA: ribosomal RNA.

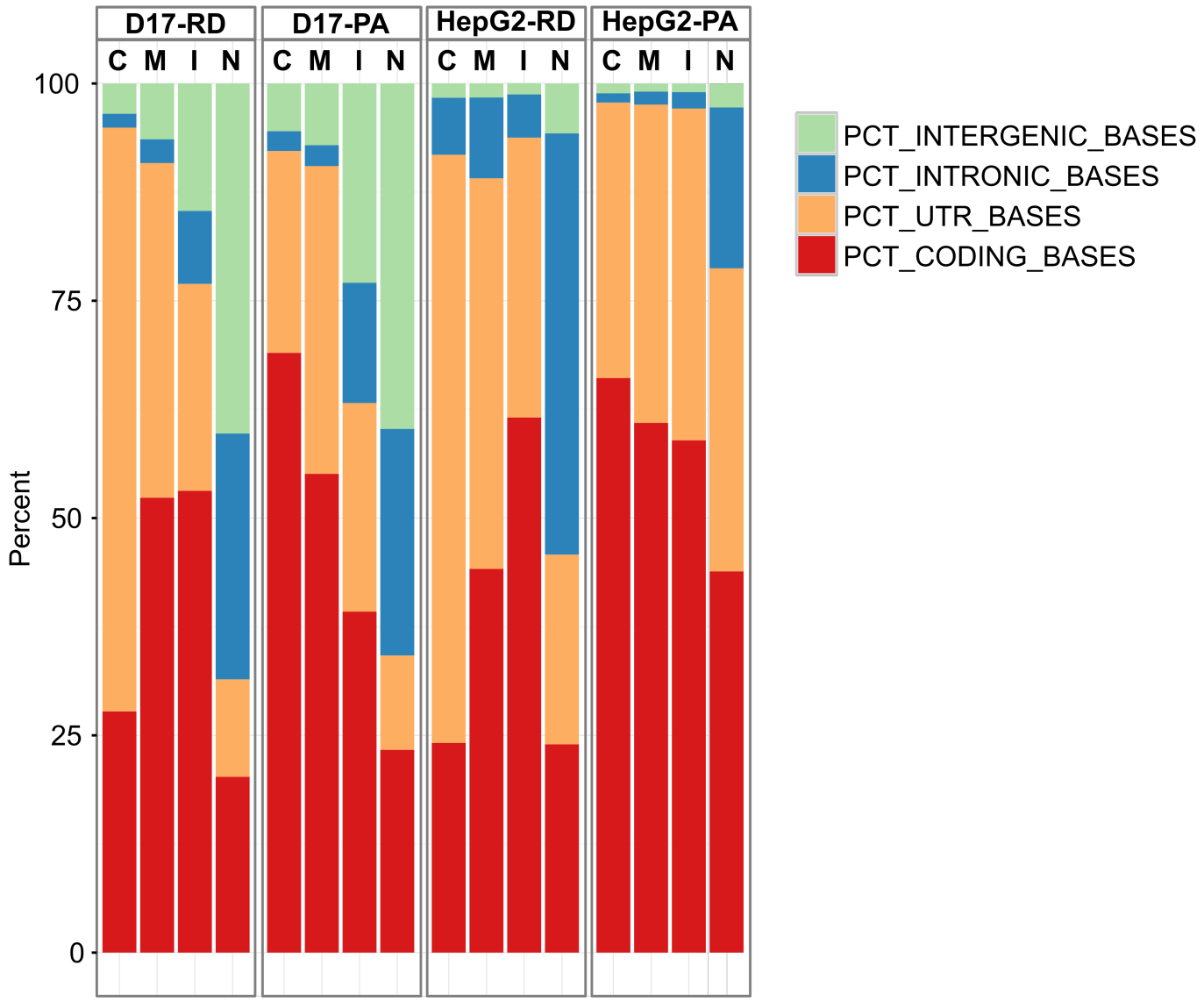


Figure S3: **RNA-seq read distribution profiles varies between fractions.**
Fraction of the total number of aligned bases that mapped to eiter the protein coding regions of genes, untranslated regions (UTR) of genes, gene introns, or intergenic regions of genomic DNA in HepG2 and D17 cells either assessed from polyA-enriched (PA) or rRNA-depleted (RD) sequencing datasets. C= Cytosolic, M= Membrane, I= Insoluble, N= Nuclear.

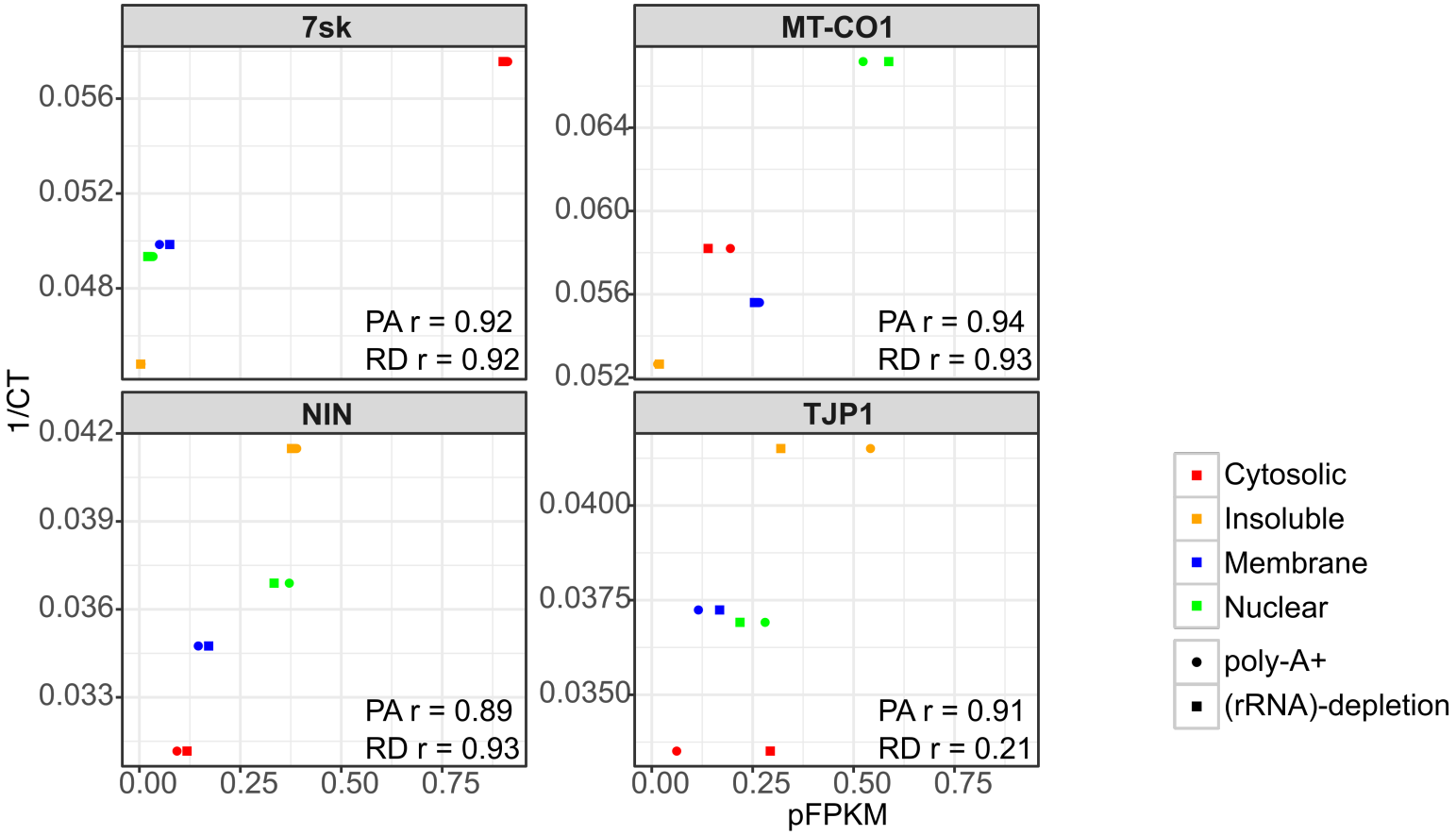


Figure S4: **pFPKM values are strongly correlated with the results of RT-qPCR.**
Scatter plot of RT-qPCR cycle threshold (CT) values and pFPKM for fraction-specific RNA markers demonstrate that CT values are strongly correlated with pFPKM values. Similar results were obtained with rRNA-depletion and poly-A+ RNA sequencing datasets.
Pearson correlations are indicated on each graph.

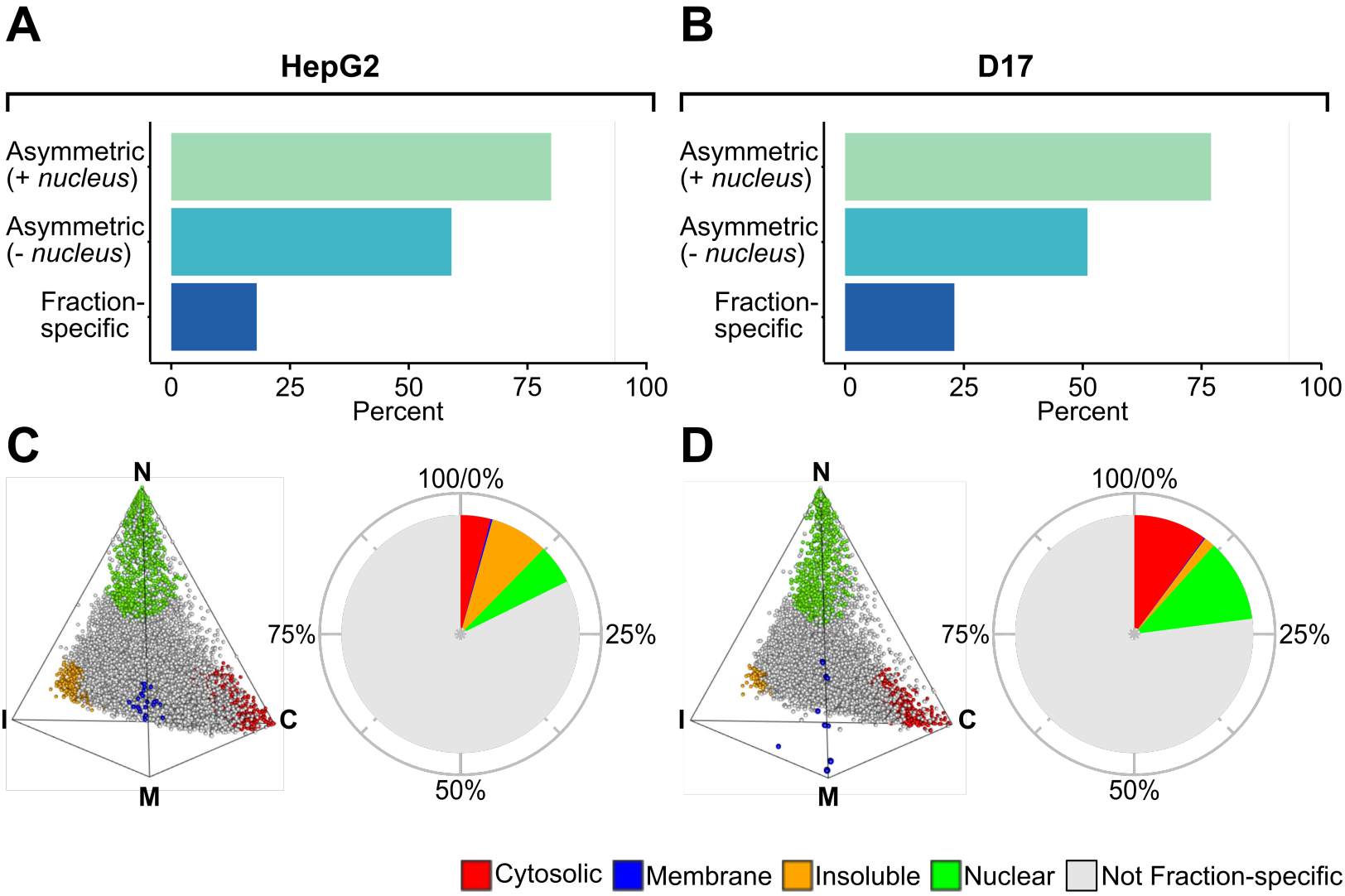


Figure S5: **Distribution profile of mRNAs in rRNA-depleted sequencing datasets.**
(A-B) Histograms showing the percent of asymmetrically distributed and fraction-specific mRNAs in HepG2 (A) and D17 (B) cells.
(C-D) Simplex graphs (left panels) and pie charts (right panels) depicting the relative distribution and proportion of fraction-specific mRNAs, coloured according to the fraction they are enriched in, relative to the total mRNA population in HepG2 (C) or D17 (D) cells.
C= Cytosolic, M= Membrane, I= Insoluble, N= Nuclear.

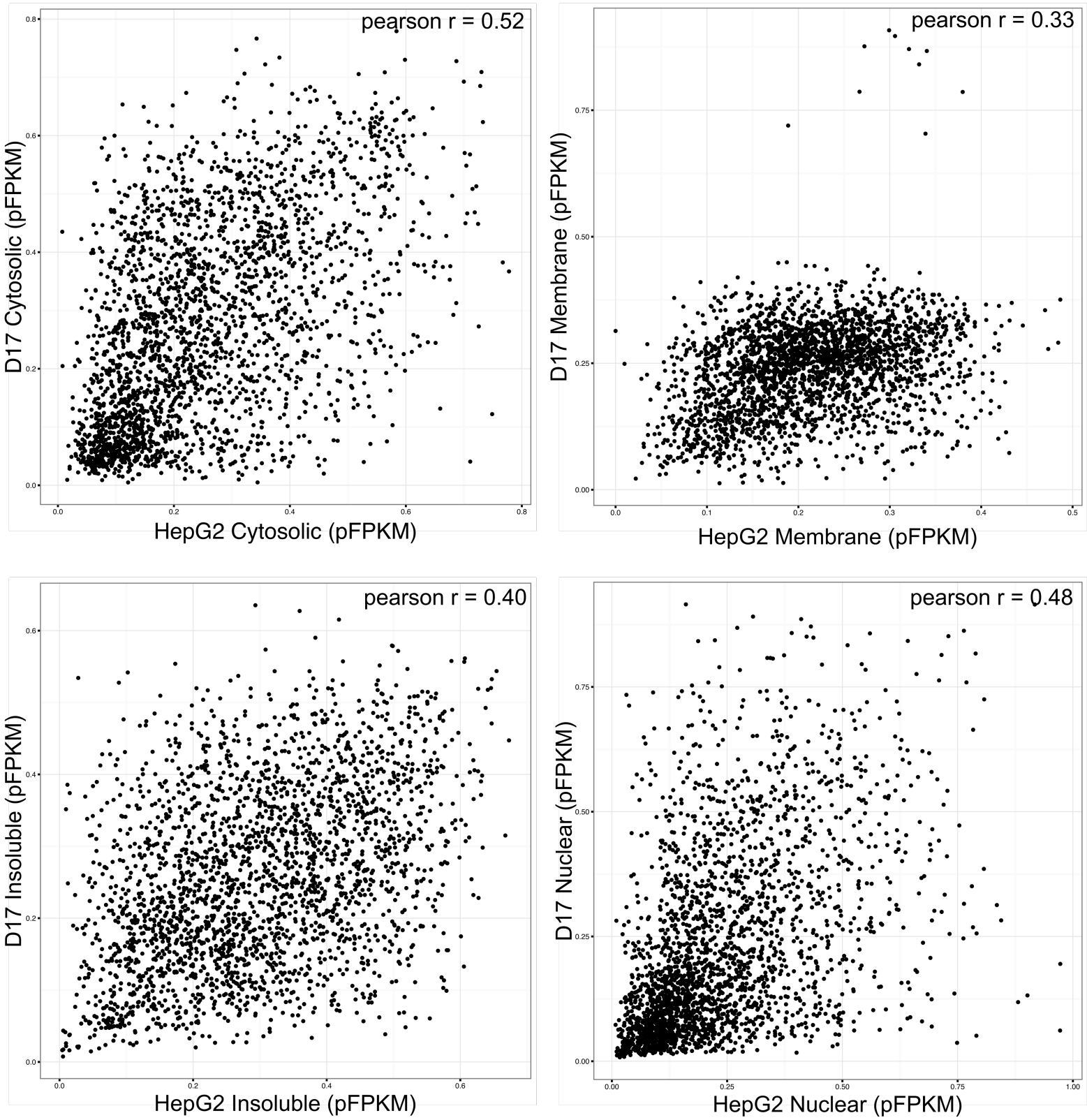


Figure S6: **Human and fly cell mRNA localization is strongly correlated between the same fraction.** Scatter plot of mRNA relative localization in pFPKM between the same fraction in HepG2 and D17 cells. Pearson correlations are indicated on each graph. All p-values $<2.2 \times 10^{-16}$.

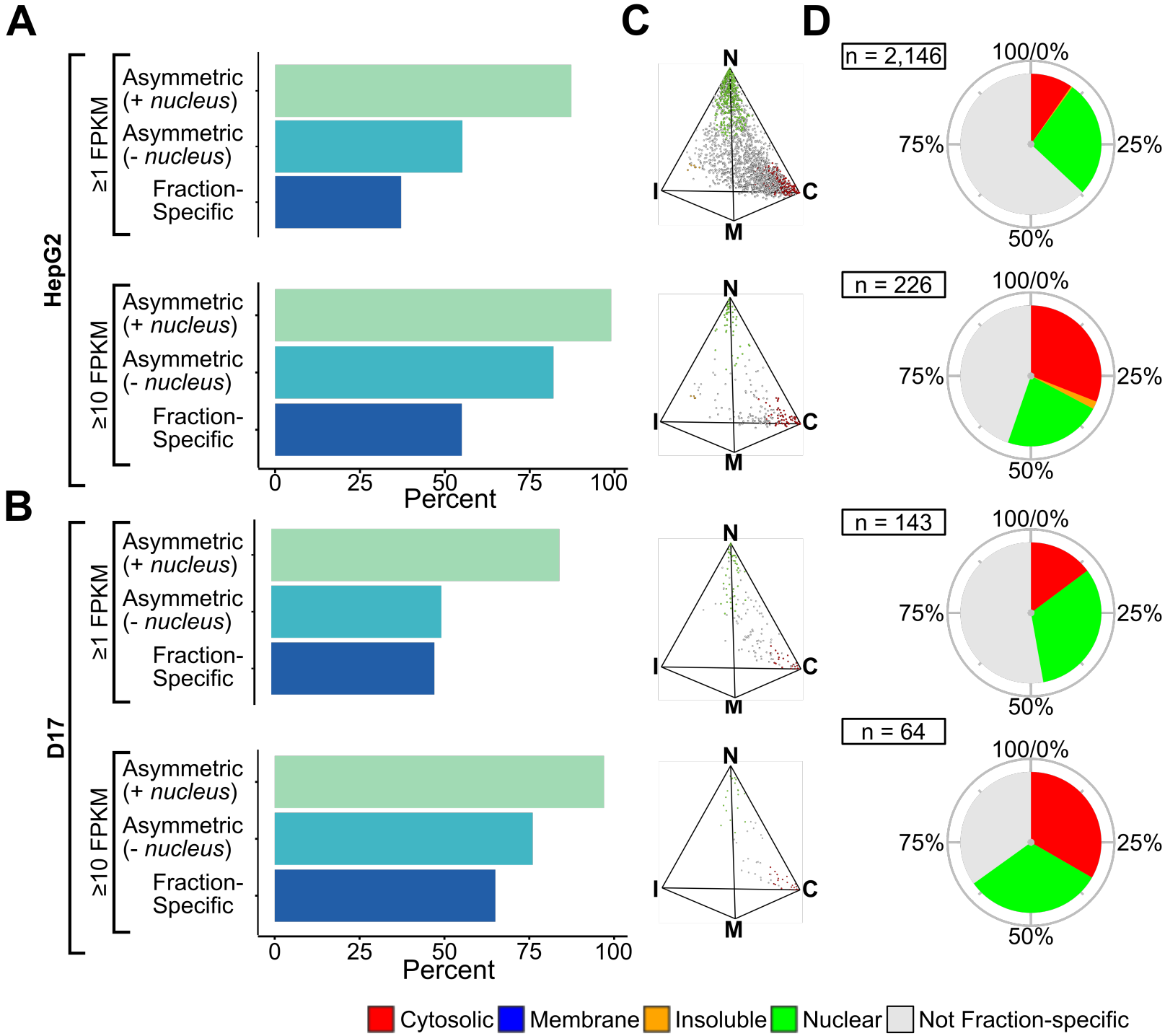


Figure S7: LncRNAs from rRNA-depleted sequencing datasets are asymmetrically distributed and exhibit preferential polarization towards the nucleus and cytosol.
(A-B) Histogram showing the percent of asymmetrically distributed and fraction-specific lncRNAs, obtained from RD-sequencing, expressed in HepG2 (A) and D17 (B) cells, either using a standard expression threshold (≥1FPKM) or focusing on highly expressed transcripts (≥10 FPKM).
(C-D) Simplex graphs (C) and pie charts (D) depicting the relative distribution and proportion of fraction-specific lncRNAs, coloured according to the fraction they are enriched in, relative to the total lncRNA populations detected in HepG2 or D17 cells at the expression thresholds indicated in (A-B).

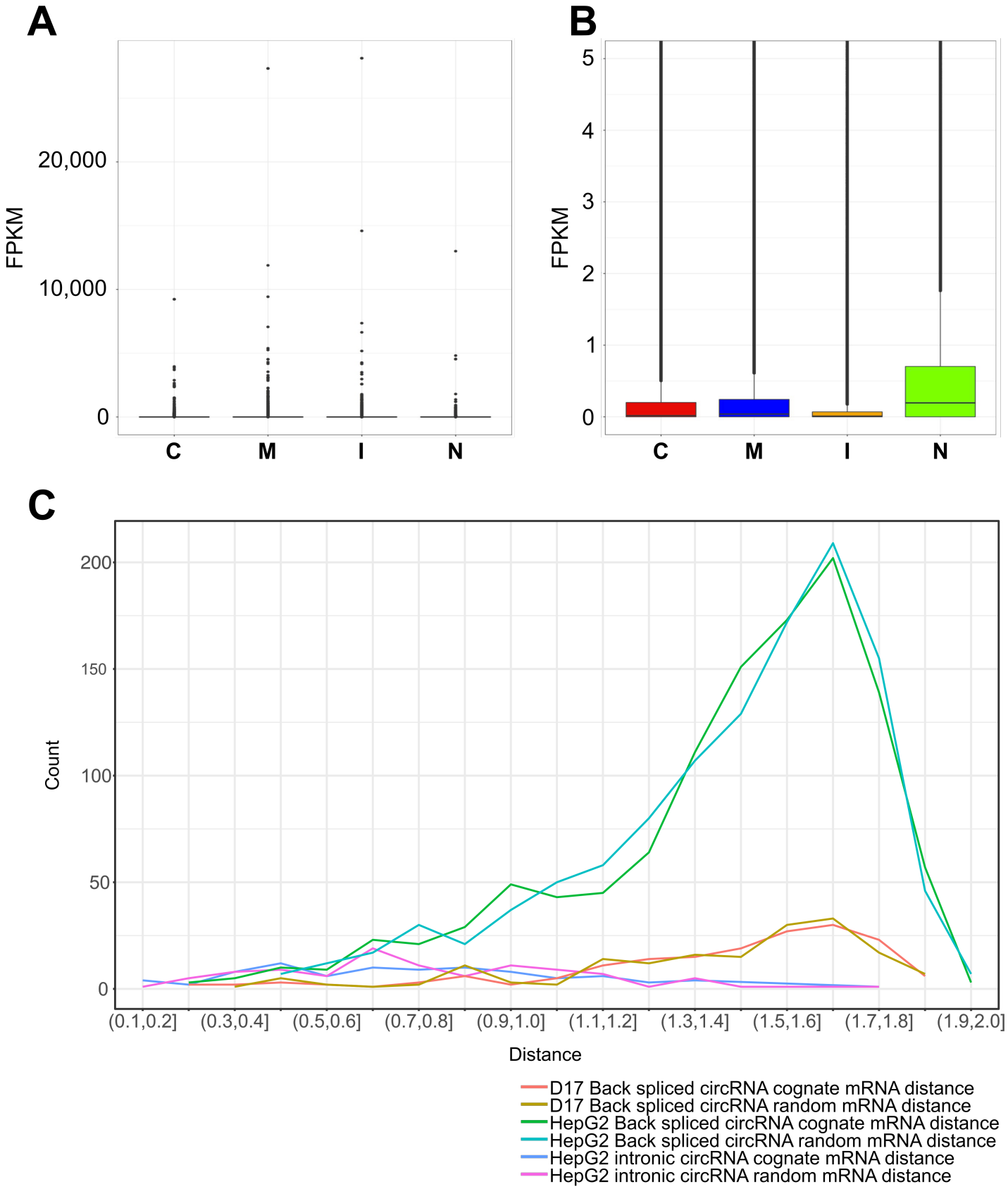


Figure S8: CircRNA exhibit distinct distribution compared to other intronic regions or their host mRNAs.

(A) Boxplot of the expression, in FPKM, of 1,040,283 introns as annotated by Ensembl in HepG2 cells.

(B) Zoomed view of the boxplot described in (A).

(C) The relative localization distance between a circRNA and its cognate mRNA is similar to that of a random pairing of an equal number of cirRNA and a mRNA encompassing a putative circRNA for both back-spliced and intronic circRNA in both human and fly cells.

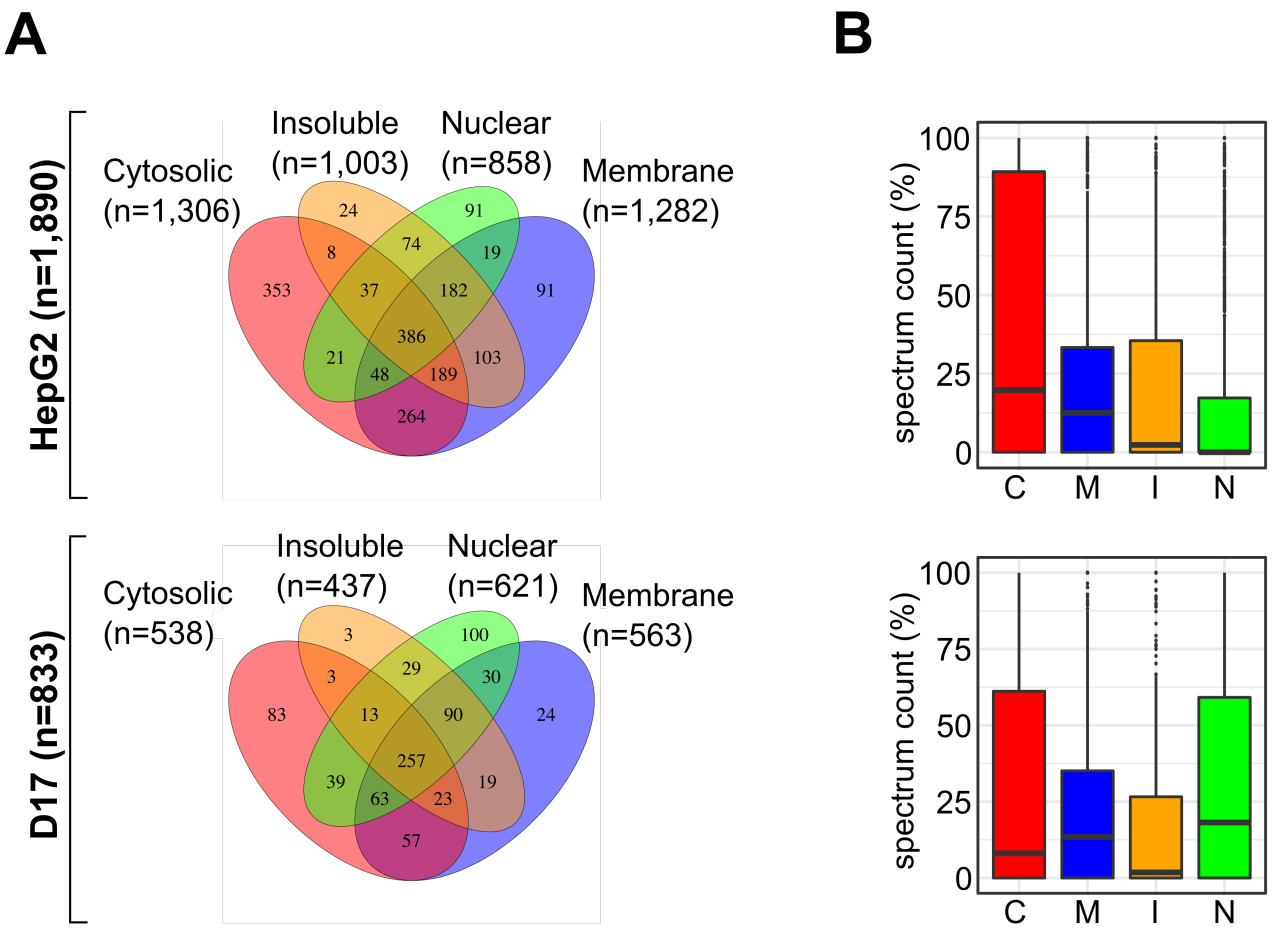


Figure S9: **Proteome distribution of each fraction.**
(A) Venn diagram depicting the relative distribution of proteins in HepG2 (upper panel) and D17 (lower panel) cell fractions following tandem mass spectrometry (LC-MS/MS), assessed by measuring their total spectrum count.
(B) Boxplot showing the relative distribution, in percent spectrum count, of all proteins identified in our LC-MS/MS datasets for HepG2 (upper panel) and D17 (lower panel) cell fractions.
C= Cytosolic, M= Membrane, I= Insoluble, N= Nuclear.

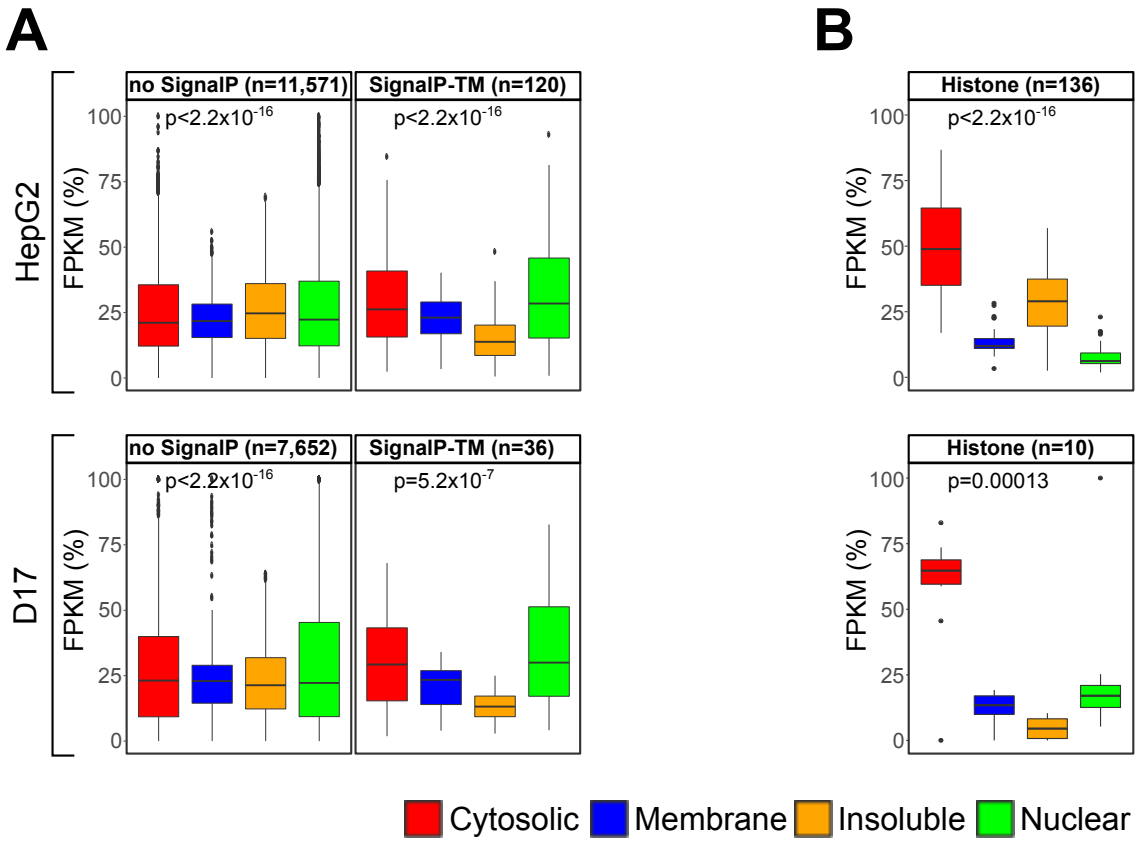


Figure S10: mRNA bearing various motifs exhibit asymmetric disributions.
(A) Boxplots showing the fraction distribution profiles of mRNAs bearing a signal peptide cleavage sites, in pFPKM, for HepG2 (upper plots) and D17 (lower plots) cells.
(B) Boxplots showing the fraction distribution profiles of mRNAs canonical histones, in pFPKM, for HepG2 (upper plots) and D17 (lower plots) cells.
The number (n) of transcripts analyzed and Kruskal-Wallis p-value is indicated for each motifs.