Supplemental Material

Simultaneous sequencing of coding and non-coding RNA reveals a human transcriptome dominated by a small number of highly expressed non-coding genes

Boivin and Deschamps-Francoeur et al., 2018

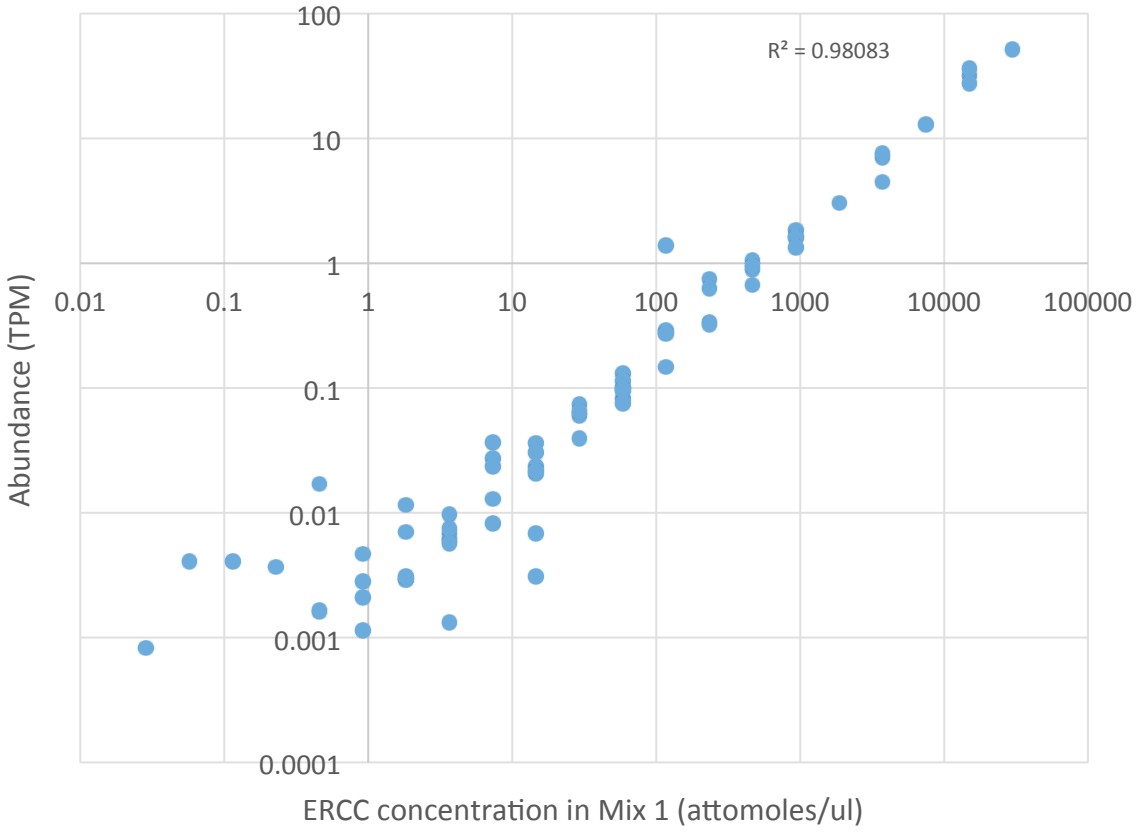


Figure S1. Comparison between the concentration of ERCC (External RNA Controls Consortium; from Thermo Fisher Scientific) spike-in RNAs added to the library and the estimated abundance of these molecules as detected by FRT-seq. The Pearson correlation is 0.99.

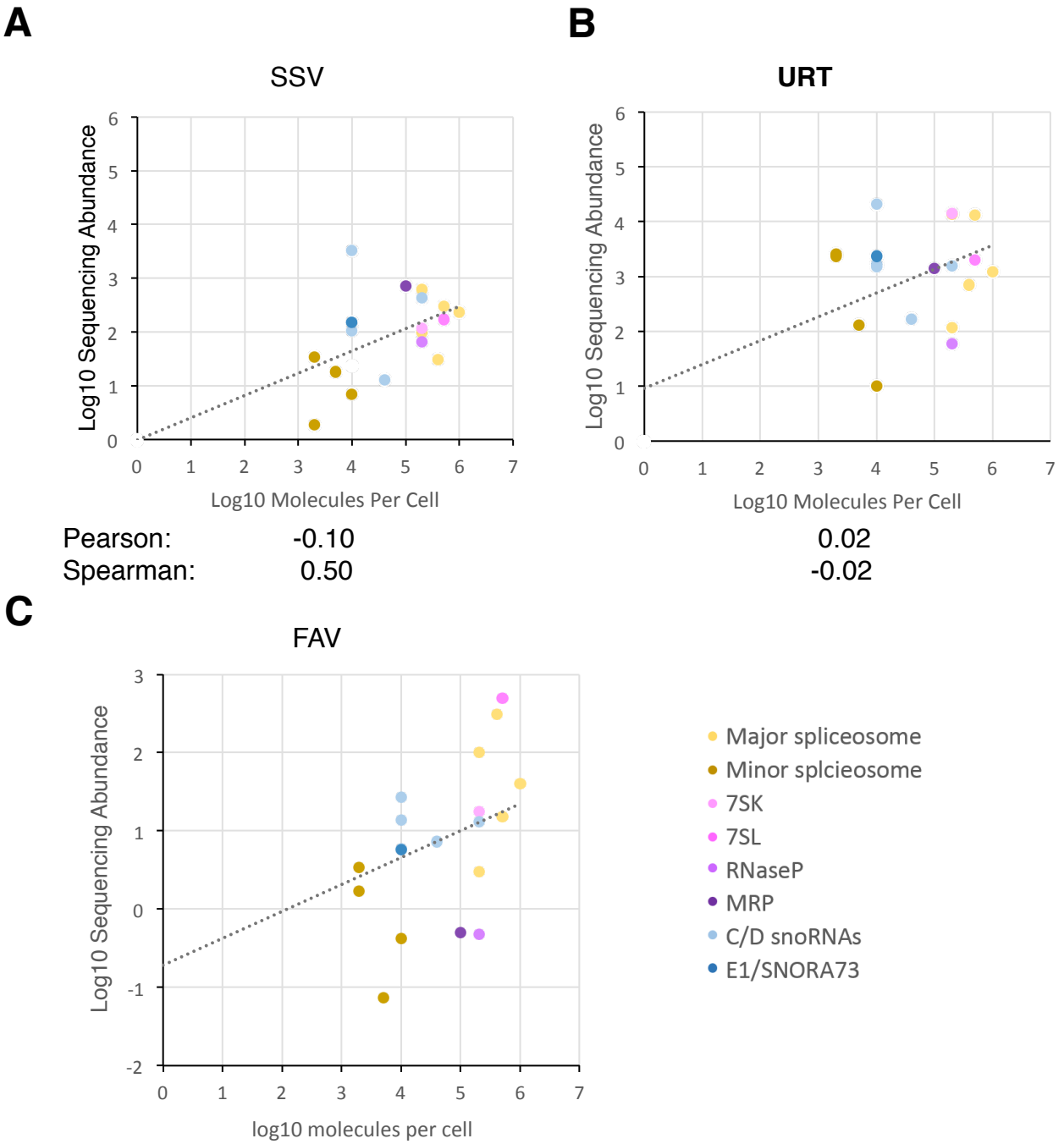


Figure S2 (related to Figure 1)**.** Comparison between the capacity of different sequencing methods to detect the abundance of the RNA component of major ribonucleoprotein particles. The abundance (TPM) of eight different classes of non-coding RNA was determined using the different sequencing methods described in Figure 1B and compared to the established number of molecules per cell previously estimated using immunoprecipitation of labeled RNA (Tyc et al (2006) The Ever-Growing World of Small Nuclear Ribonucleoproteins. In *The RNA World (Third Edition)*). The Pearson and Spearman correlation coefficients are indicated at bottom.

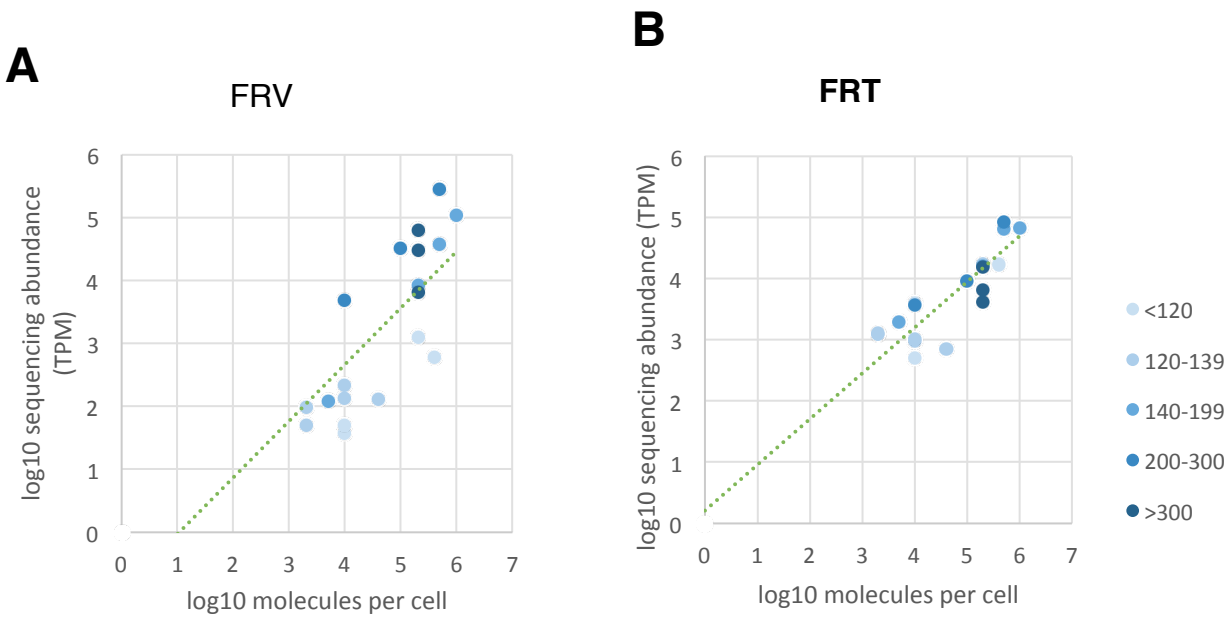


Figure S3 (related to Figure 1C). Evaluation of the size bias of (A) viral (FRV) and (B) TGIRT (FRT) reverse transcriptase based sequencing methods using a standard set of non-coding RNA. The non-coding RNA abundance obtained by the viral RT or TGIRT based sequencing methods FRV or FRT was plotted against established estimates of the number of molecules per cell as in Figure 1C. The size of the different RNA molecules used is indicated on the right (in nucleotides). The dotted line represents the linear trendline with best fit for the data points.

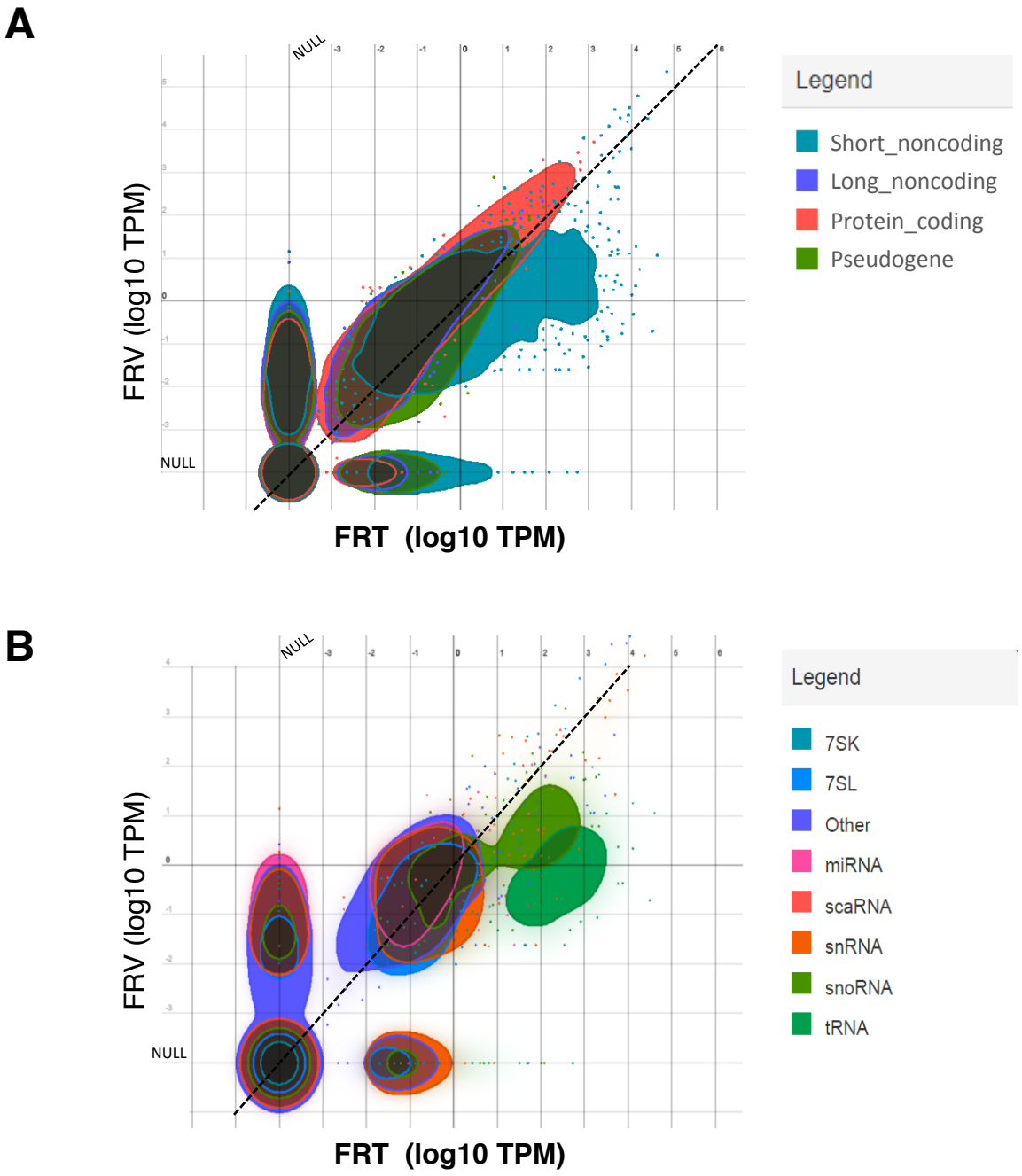


Figure S4 (related to Figures 1 and 2)**.** Comparison between the RNA abundance values generated by viral (FRV) and TGIRT (FRT) reverse transcriptase based sequencing methods. The values obtained by FRV and FRT for either all transcripts (A) or only short non-coding RNA (B) were compared using splatterplots (see https://www.ensembl.org/Help/Faq?id=468 for extensive description of the different RNA types groups in (A)). The different types of RNAs compared are indicated on the right. The diagonal line shows a perfect positive correlation regression as a visual help. The splatterplots were produced using the Splatterplots JS browser-based tool (available online at http://graphics.cs.wisc.edu/Projects/SplatterJs).

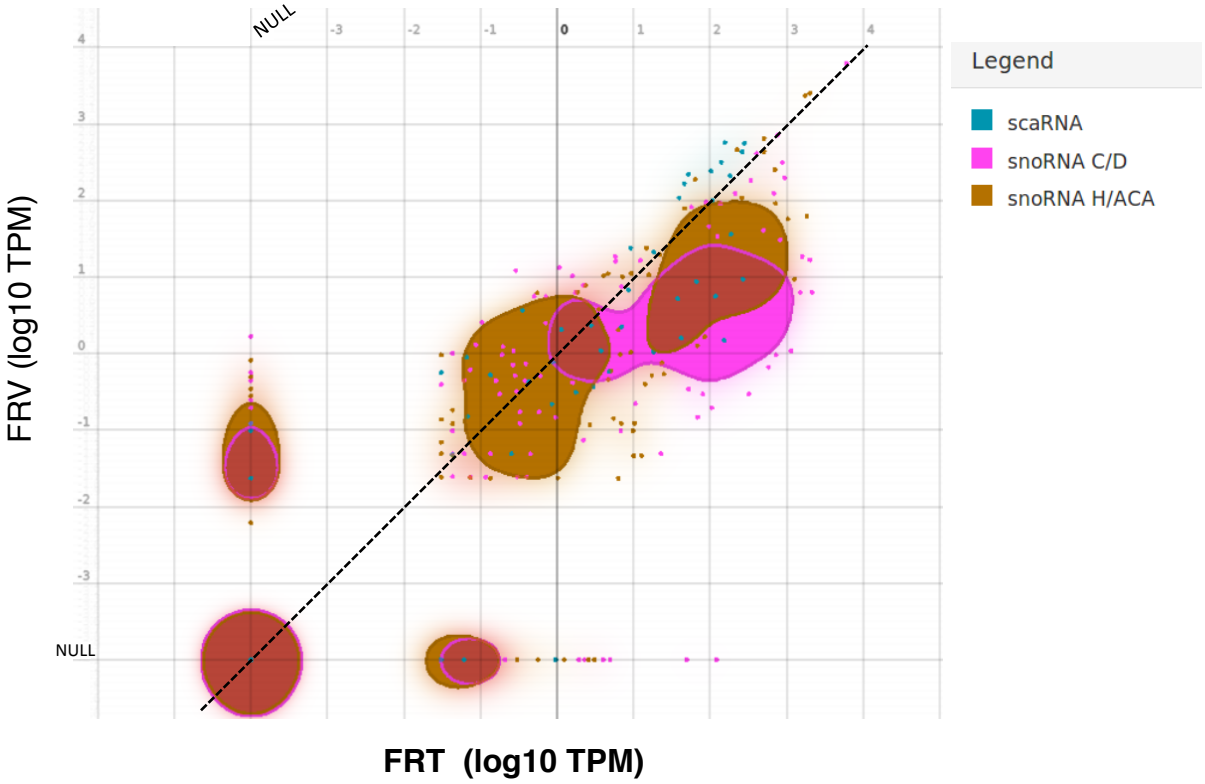


Figure S5 (related to Figures 1 and 4)**.** Comparison between the snoRNA abundance values generated by viral (FRV) and TGIRT (FRT) reverse transcriptase based sequence methods. The abundance of C/D, H/ACA and scaRNA obtained using FRV and FRT were compared and presented in the form of a splatterplot. The different types of RNAs compared are indicated on the right. The diagonal line shows a perfect positive correlation regression as a visual help. The splatterplot was produced using the Splatterplots JS browser-based tool (available online at http://graphics.cs.wisc.edu/Projects/SplatterJs).

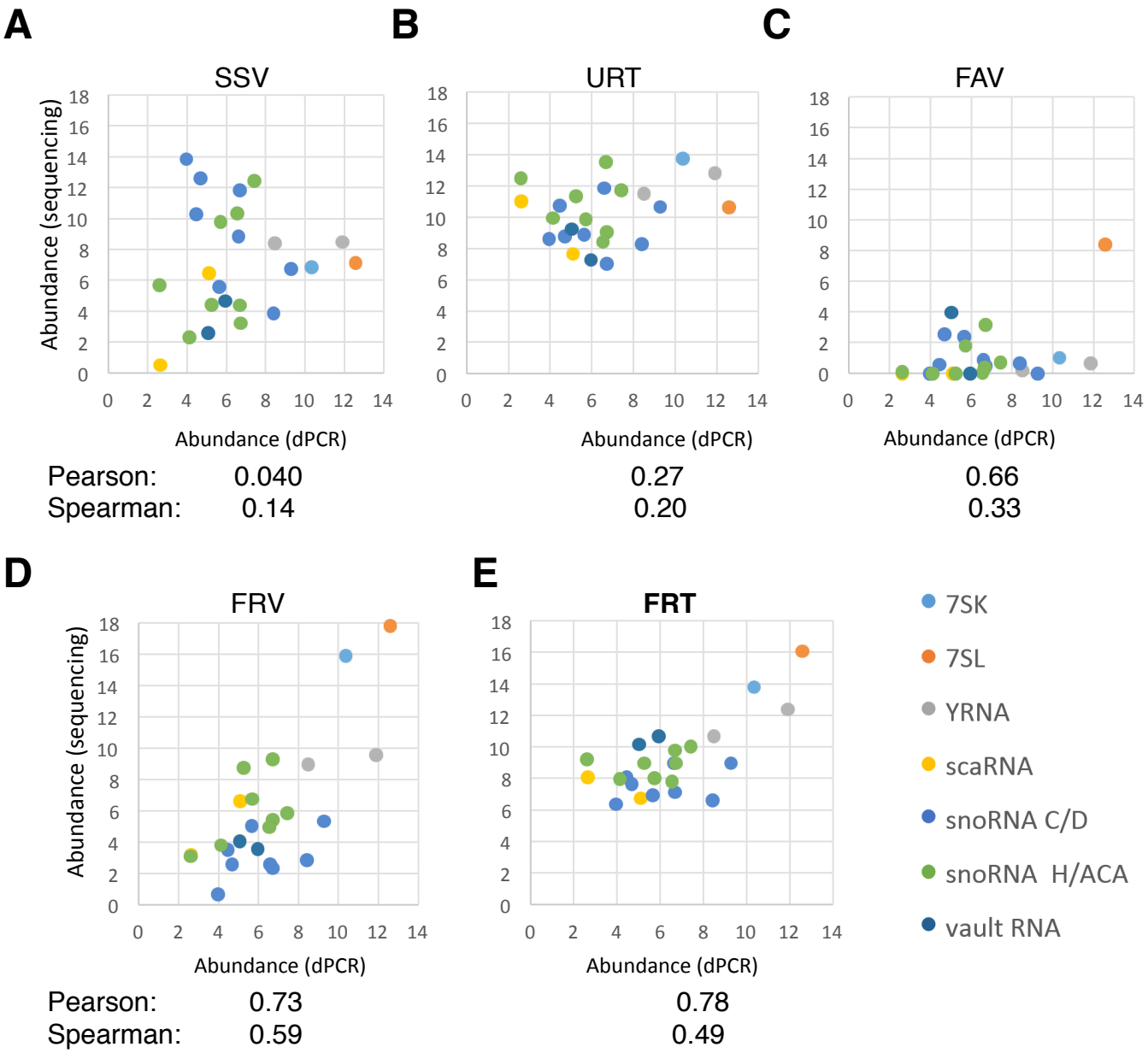


Figure S6 (related to Figure 1)**.** Comparison between the RNA abundance values detected by digital PCR (dPCR) and those detected using different sequencing methods. The RNA abundance was detected with different sequencing protocols using different RNA selection methods and reverse transcriptases including: Size selected viral reverse transcriptase sequencing (SSV), Ribodepleted TGIRT sequencing (URT), Fragmented poly A selected viral reverse transcriptase sequencing (FAV), Fragmented ribodepleted viral reverse transcriptase sequencing (FRV), Fragmented ribodepleted TGIRT sequencing (FRT). The resulting digital PCR values (expressed as log2 of the number of copies/ul detected) were compared to the abundance of the same molecules by the different sequencing approaches (shown as log2 of TPM). The values used for the sequencing are the average of two biological replicates. The Pearson and Spearman coefficients for each pair of comparisons are indicated at the bottom of each scatter plot. The names of the different RNA considered are shown on the bottom right corner. The primers used for the dPCR are listed in Table S10.

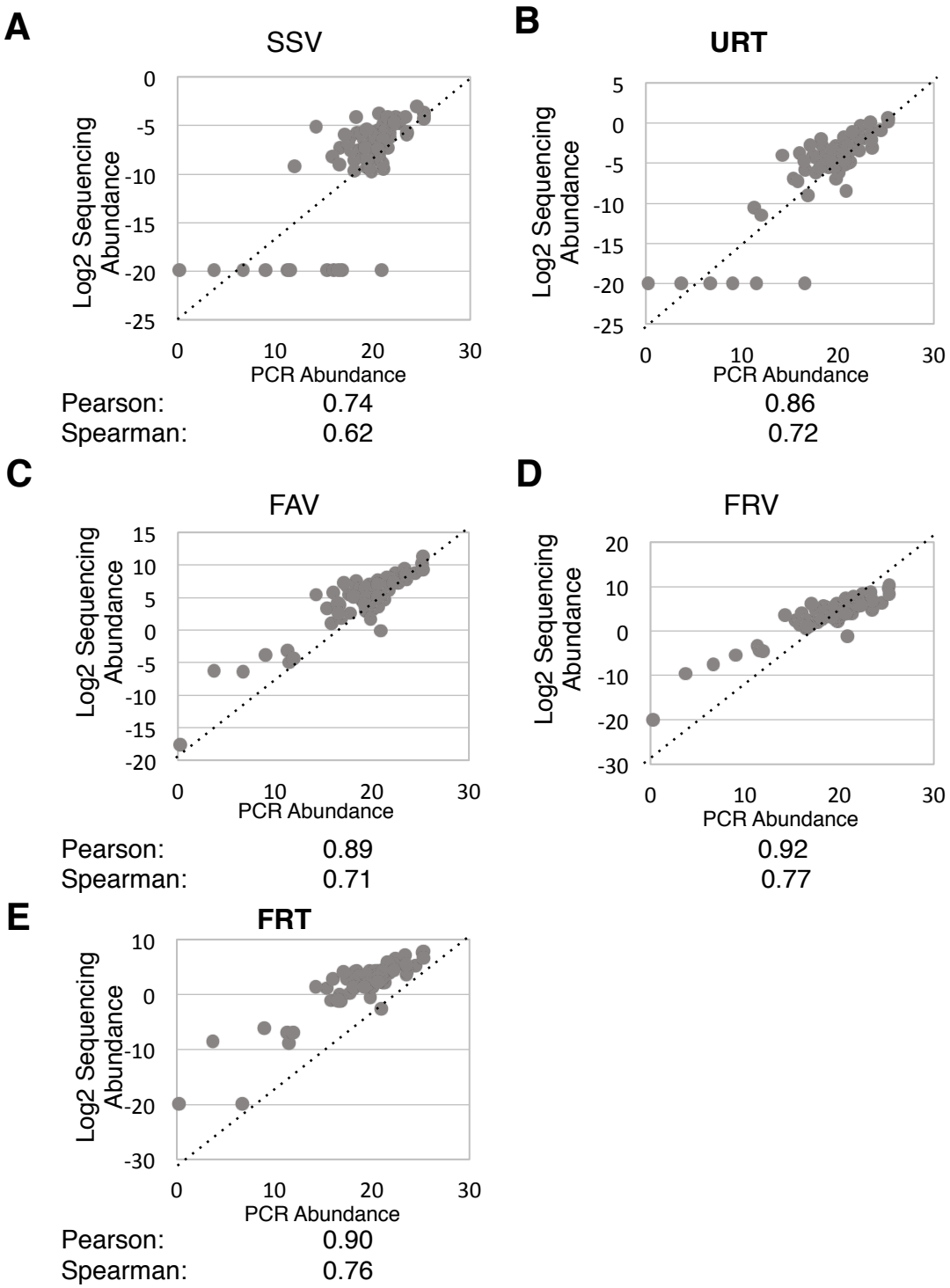


Figure S7 (related to Figures 1, 2 and 3)**.** Comparison between the capacity of the different sequencing methods and quantitative RT-PCR to predict mRNA abundance. The protein-coding mRNA abundance was detected with different sequencing protocols using different RNA selection methods and reverse transcribed as described in Figure 1 and the resulting values compared to those detected by quantitative RT-PCR using viral reverse transcriptases in the same RNA sample. The values used are the average of two biological replicates. The Pearson and Spearman coefficients for each pair of comparisons are indicated at the bottom of each scatter plot.

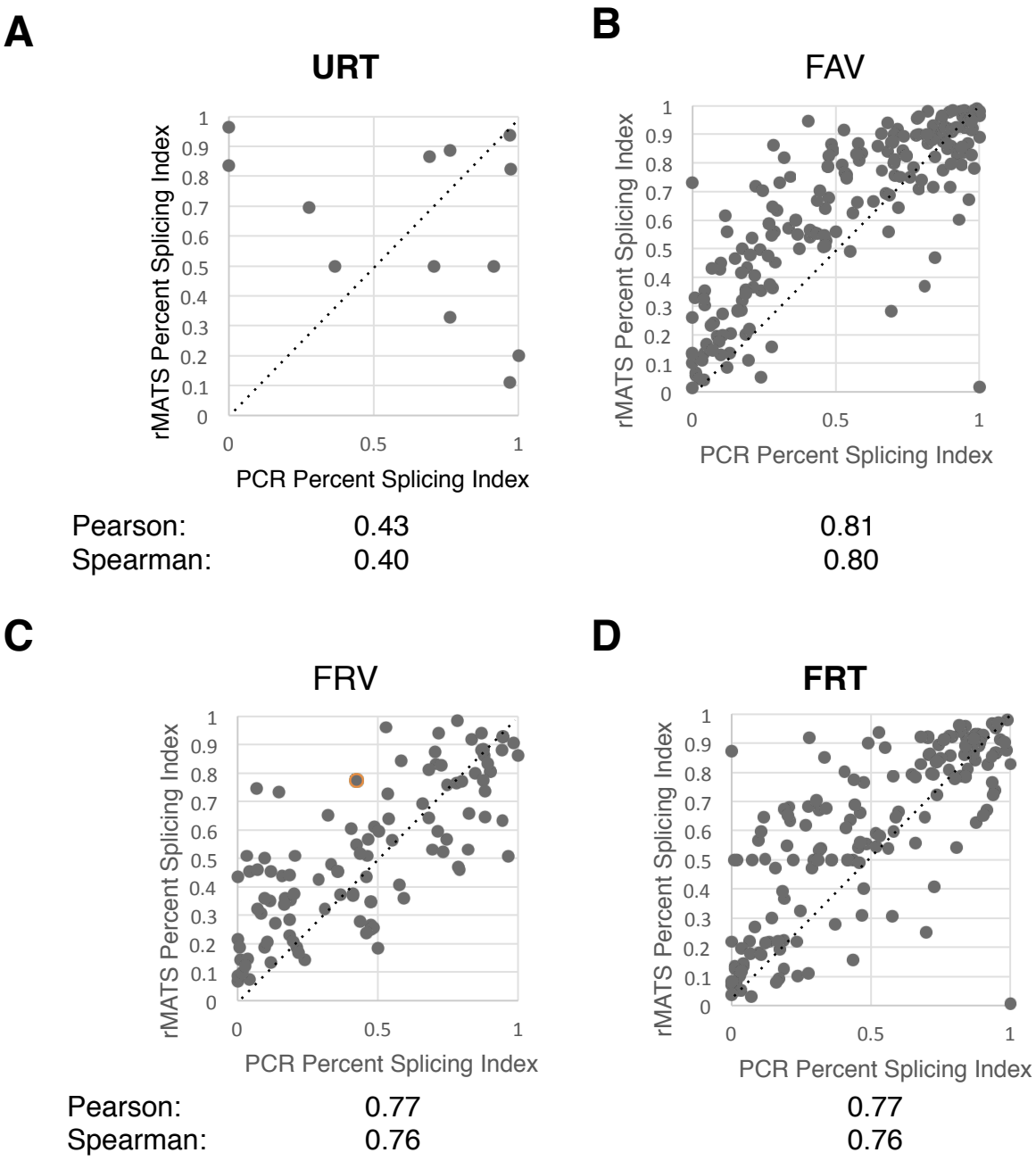


Figure S8 (related to Figure 1)**.** Comparison between the capacity of the different sequencing methods and quantitative RT-PCR to predict the ratio of splicing isoforms. The splicing index predicted by rMATS using data generated from different sequencing protocols described in Figure 1 was compared with that obtained by ASPCR (alternative splicing analysis by end-point RT-PCR) using viral reverse transcriptases in the same RNA sample. The values used are the average of two biological replicates. The Pearson and Spearman coefficients for each pair of comparisons are indicated at the bottom of each scatter plot.

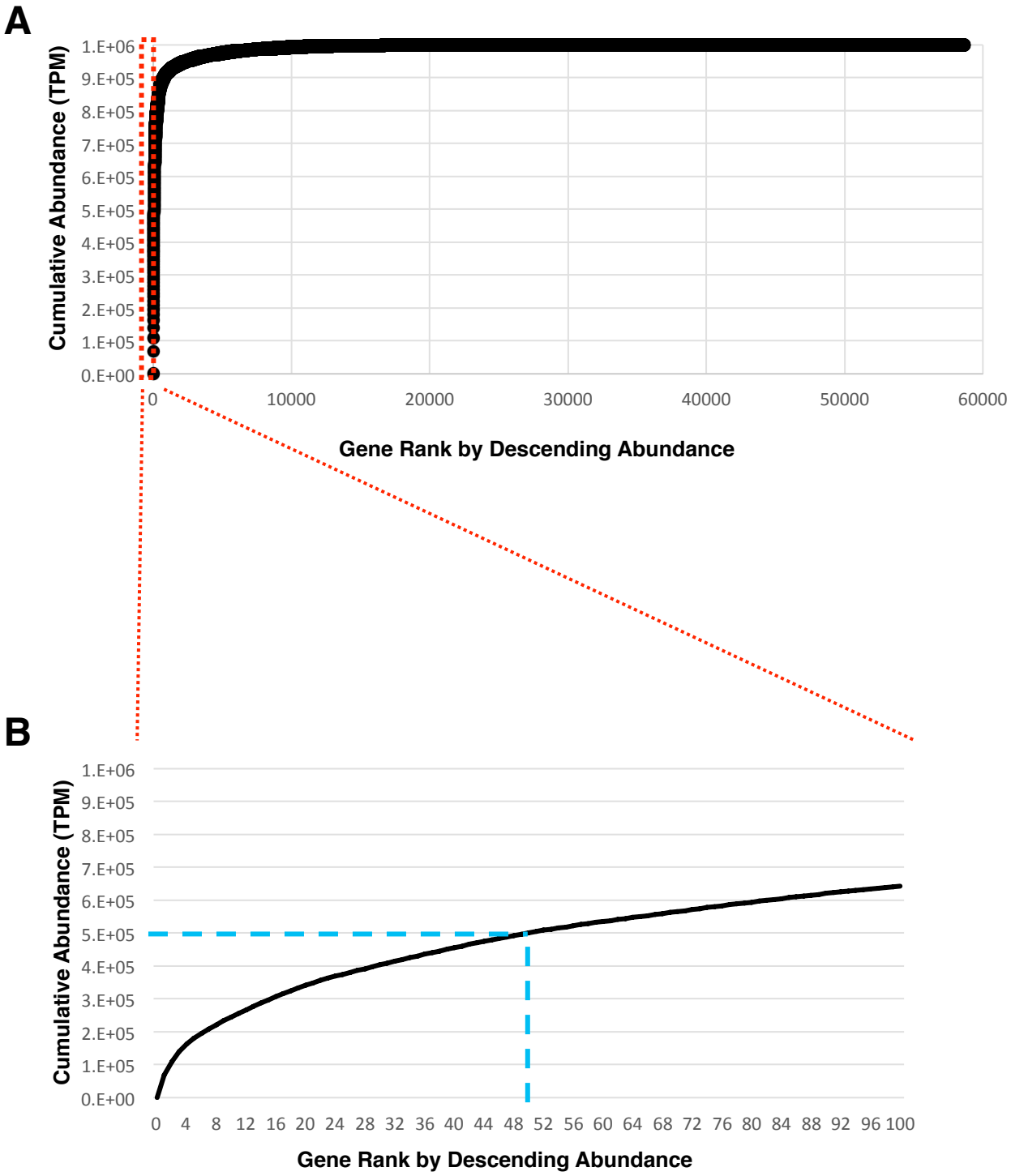


Figure S9 (related to Figures 2-4). Fifty genes generate 50% percent of the human transcriptome in terms of number of transcripts. (**A**) Genes were ranked in descending order based on RNA abundance and plotted against the cumulative number of transcripts per million. (**B**) The region of the curve depicting the most highly expressed genes is enlarged for clarity. Note that no coding gene is present in the top 100 most expressed genes (the coding gene with the highest transcript abundance is ranked 136 over all genes).

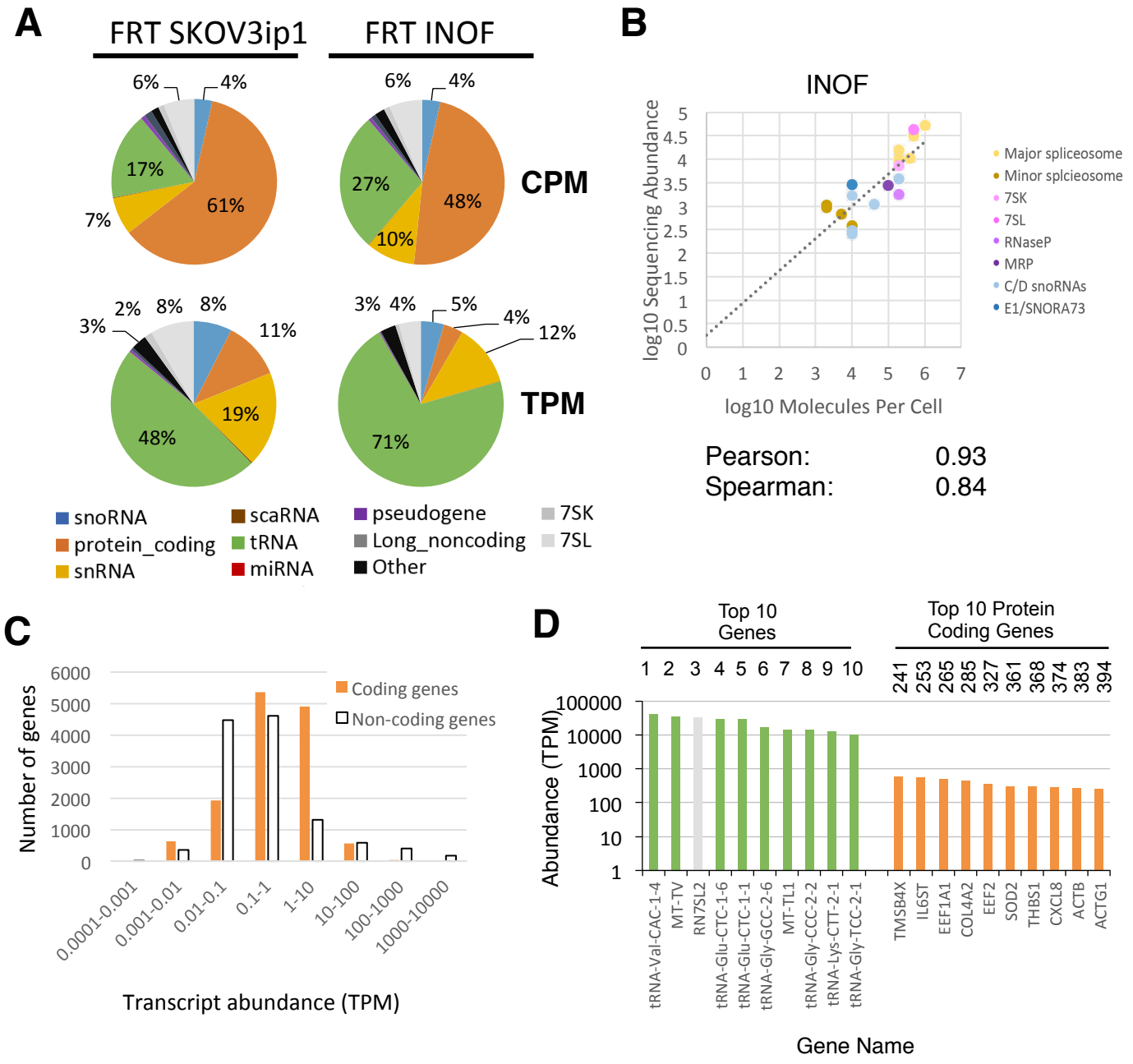


Figure S10 (related to Figures 1 and 2). Comparison between results obtained from FRT RNA-Seq in SKOV3ip1 and INOF cell lines. (A) Distribution of RNA in the human transcriptome as detected by the FRT sequencing method in SKOV3ip1 and INOF. The percentage of the main classes (≥ 2%) is indicated. The color legend for the different RNA classes is shown at bottom. (B) Comparison between FRT sequencing of INOF and the established number of molecules per cell previously estimated using immunoprecipitation of labeled RNA (as in Figure 1C right panel). The Pearson and Spearman correlation coefficients are indicated at bottom. (C) The abundance of both coding and non-coding genes in FRT INOF, separated in bins based on transcript abundance and the number of genes per bin illustrated in the form of a bar graph. (D) The genes producing the top ten overall most abundant transcripts and the top ten most abundant protein coding RNA in FRT INOF are indicated in the form of a bar graph. The rank of each transcript based on abundance in transcript per million (TPM) is indicated on top.

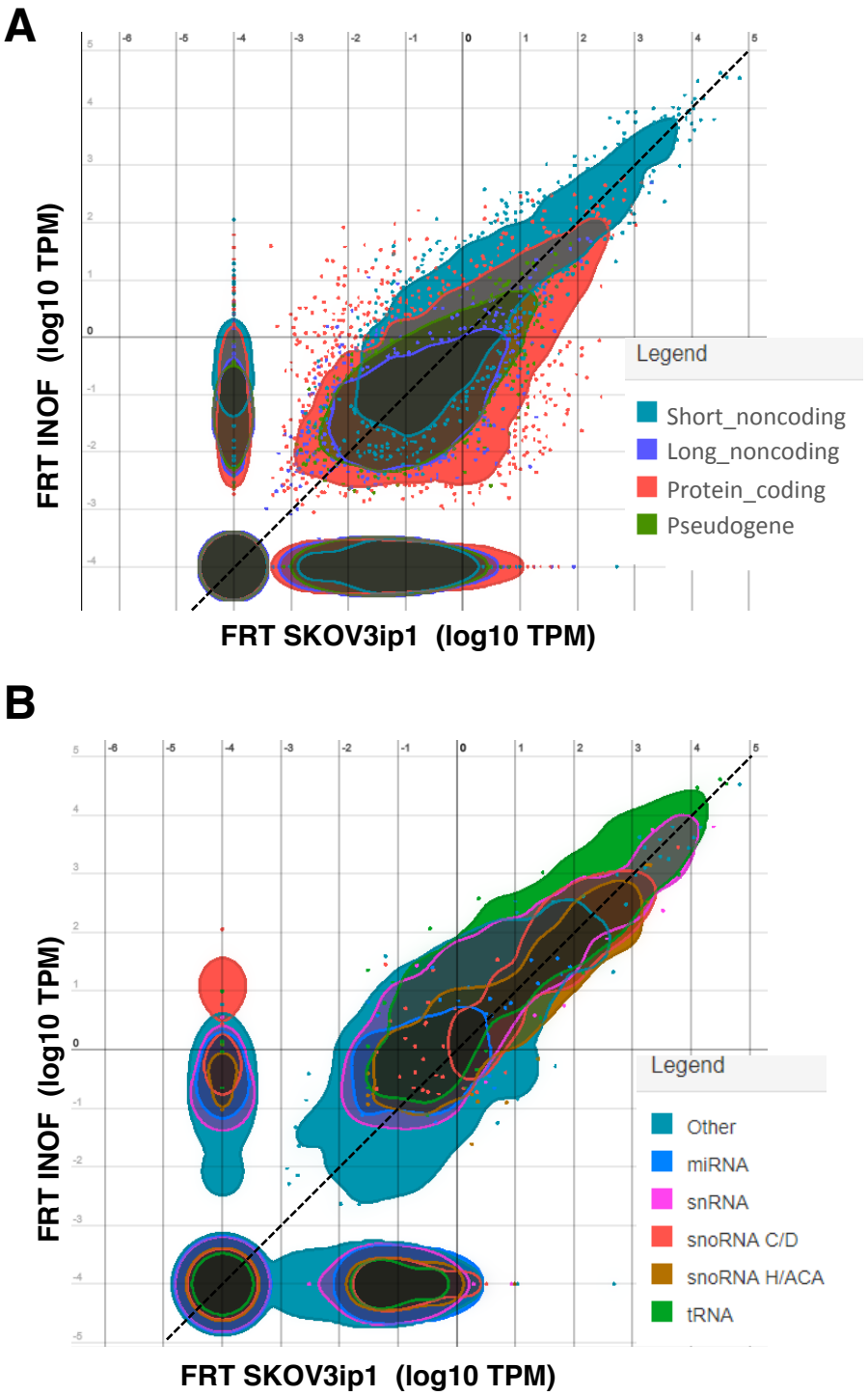


Figure S11 (related to figures S10, S4). Comparison between the RNA abundance values generated by FRT sequencing method for SKOV3ip1 and INOF cell lines. The values obtained for SKOV3ip1 and INOF for either all transcripts (A) or only short non-coding RNA (B) were compared using splatterplots (see https://www.ensembl.org/Help/Faq?id=468 for extensive description of the different RNA types groups in (A)). The different types of RNAs compared are indicated on the right. The diagonal line shows a perfect positive correlation regression as a visual help. The splatterplots were produced using the Splatterplots JS browser-based tool (available online at http://graphics.cs.wisc.edu/Projects/SplatterJs).

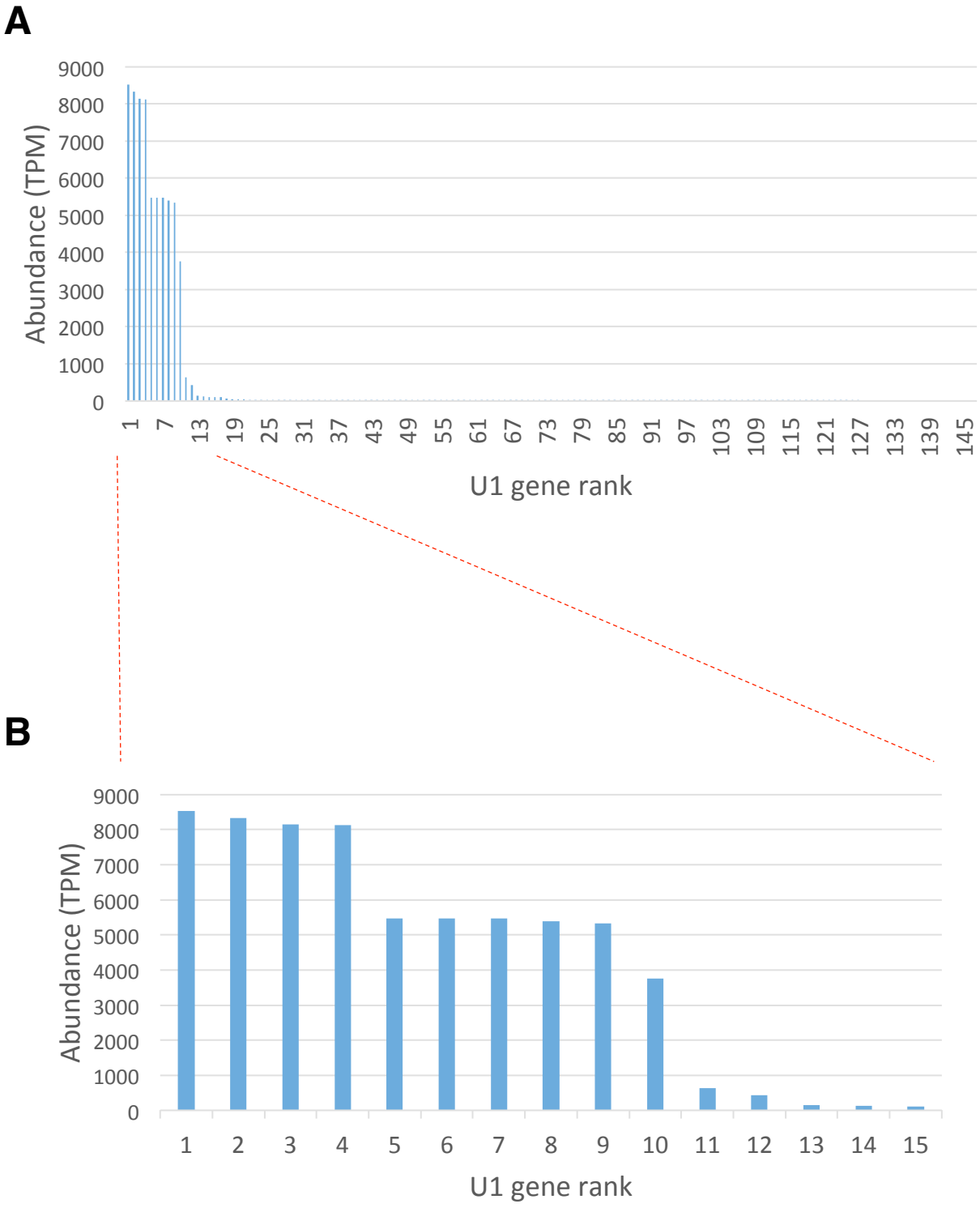


Figure S12 (related to Figure 2). U1 snRNA is a multi-copy gene whose ten most highly expressed copies produce >96% of U1 transcripts. (**A**) 147 copies of the U1 snRNA are annotated in the human genome, most of which are not detected or are detected at very low levels. (**B**) Close up of the ten most abundant U1 transcripts are detected at levels between 3000 and 9000 TPM.

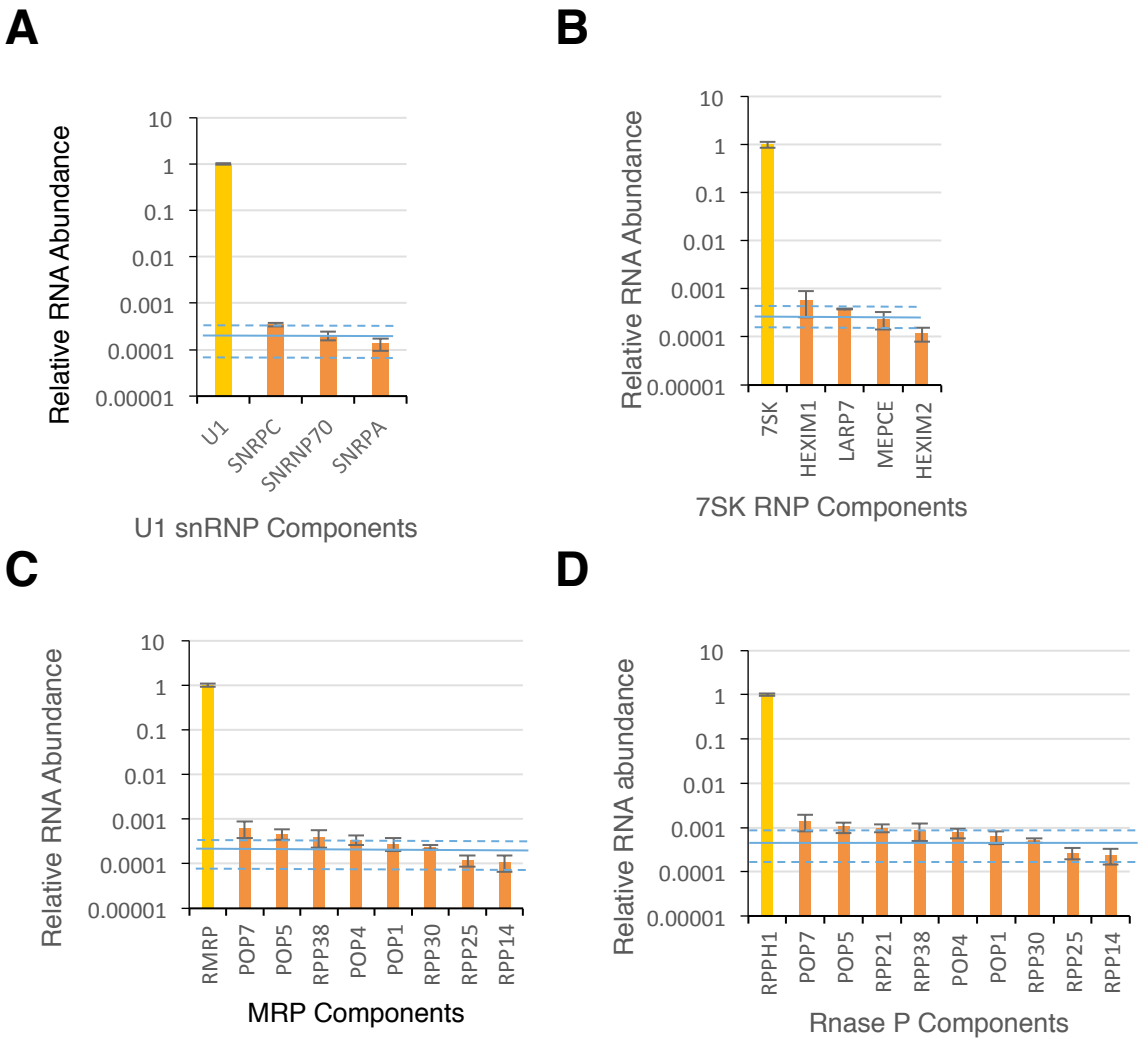
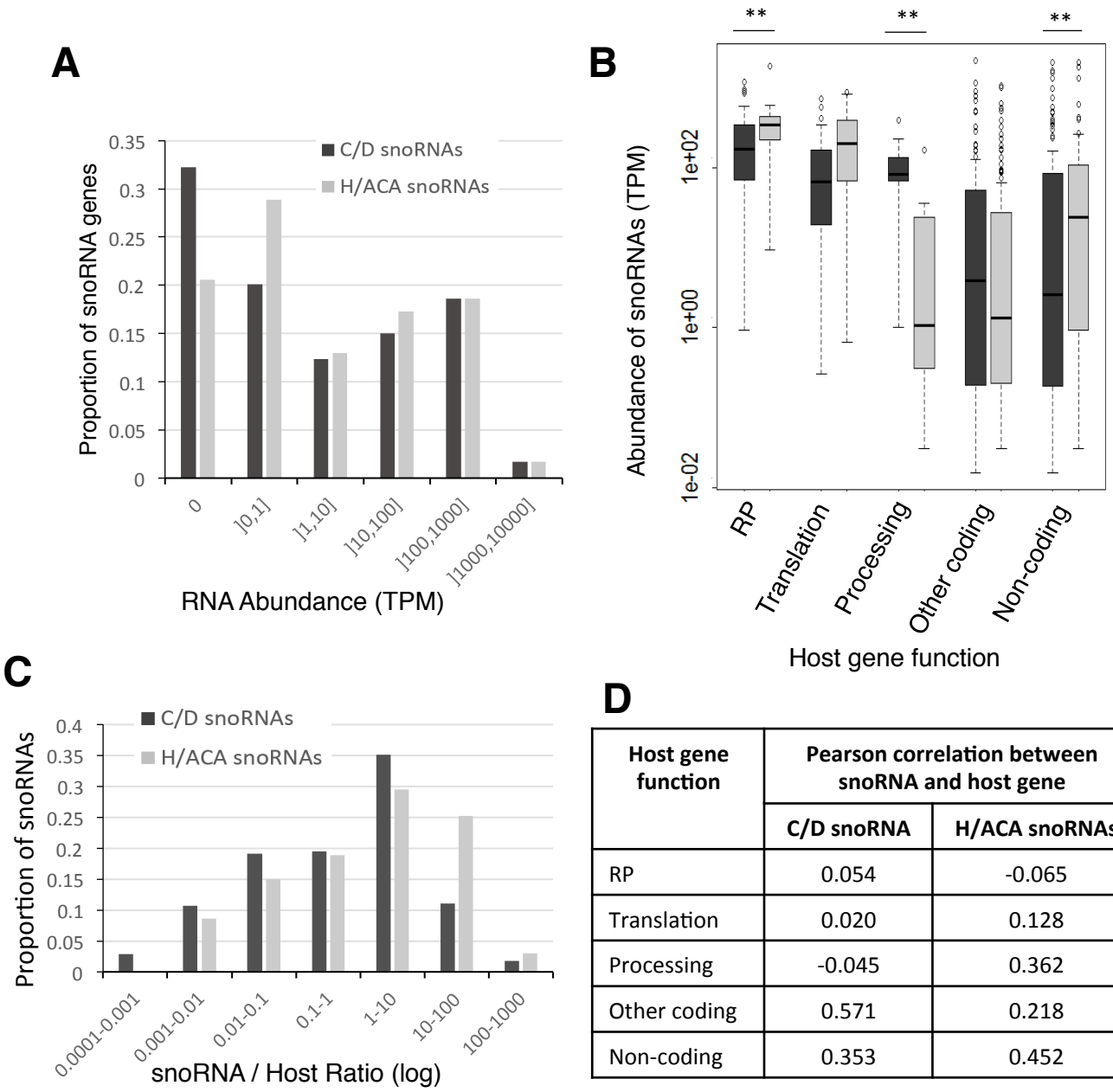


Figure S13 (related to figure 3)**.** Relative abundance of the coding and non-coding RNA components of ribonucleoprotein particles as detected by FRT**.** The abundance of mRNA coding for the protein components of U1 snRNP (A), 7SK RNP (B), Ribonuclease MRP (C) and the ribonuclease RNase P (D) were detected and plotted relative to their respective non-coding RNA. The solid line indicates the mRNA average abundance level of the complex and the dashed lines indicate the 5% and 95% confidence interval. The standard deviation of two biological replicates is indicated in the form error bars.



D

Host gene function	Pearson correlation between snoRNA and host gene	
	C/D snoRNA	H/ACA snoRNAs
RP	0.054	-0.065
Translation	0.020	0.128
Processing	-0.045	0.362
Other coding	0.571	0.218
Non-coding	0.353	0.452

Figure S14 (related to figure 4). (A) The abundance of both C/D and H/ACA snoRNA was determined using FRT, binned and represented in the form of a bar graph. (B) Box plot illustrating the distribution of C/D and H/ACA snoRNA abundance separated by the function of the host gene. RP indicates host genes coding for ribosomal proteins, translation indicates host genes implicated in translation and ribosome biogenesis, processing indicates host genes involved in splicing and RNA maturation, other coding includes all protein-coding host genes not present in other categories and non-coding indicates host genes with no coding sequence, as defined by Ensembl annotations. The black boxes indicate C/D snoRNA, while grey boxes indicate H/ACA snoRNA. The solid line indicates the median value while ** indicates a p-value <0.01 as obtained using a Kolmogorov-Smirnov goodness-of-fit test (C) Bar graph representing the proportion of C/D and H/ACA snoRNA exhibiting different snoRNA / host RNA ratios. (D) Table indicating the Pearson correlations between the abundance of snoRNAs and of their host genes, grouped according to host gene functions and abbreviated as in (B).

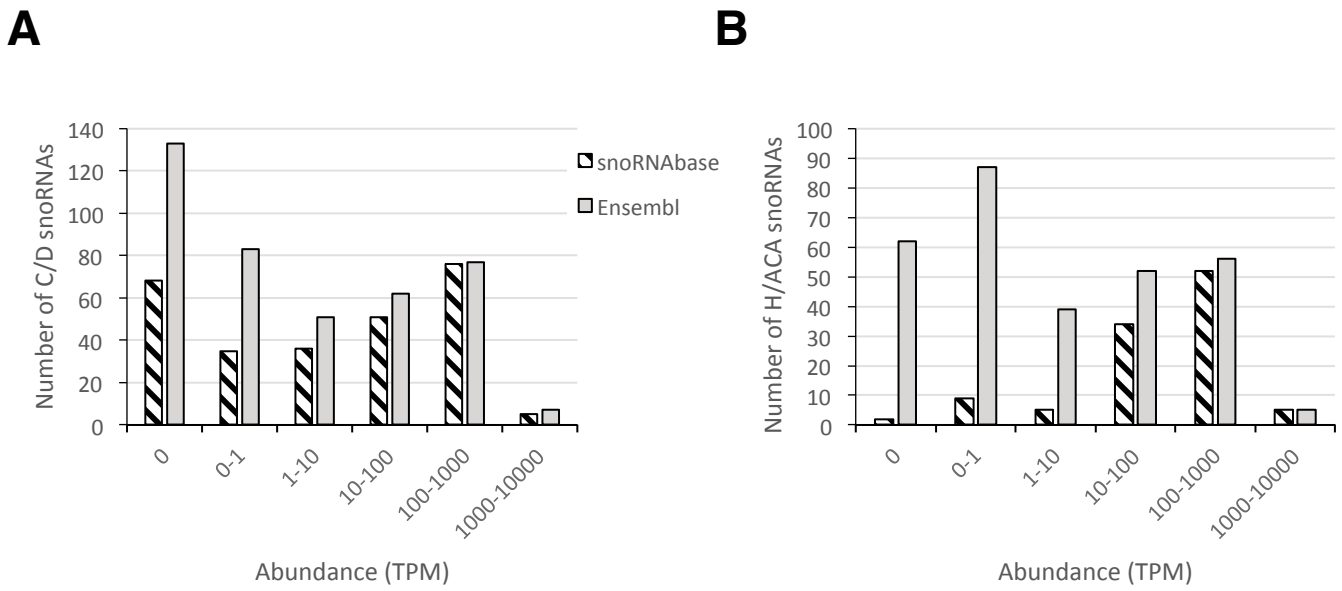


Figure S15 (related to figure 4)**.** Number of snoRNAs annotated in the snoRNAbase and Ensembl databases detected by FRT. The abundance of both C/D (A) and H/ACA (B) snoRNAs was determined using FRT and the number of genes annotated in the snoRNAbase (striped columns) and Ensembl databases (gray columns) corresponding to the different levels of gene expression is illustrated in the form of bar graphs.

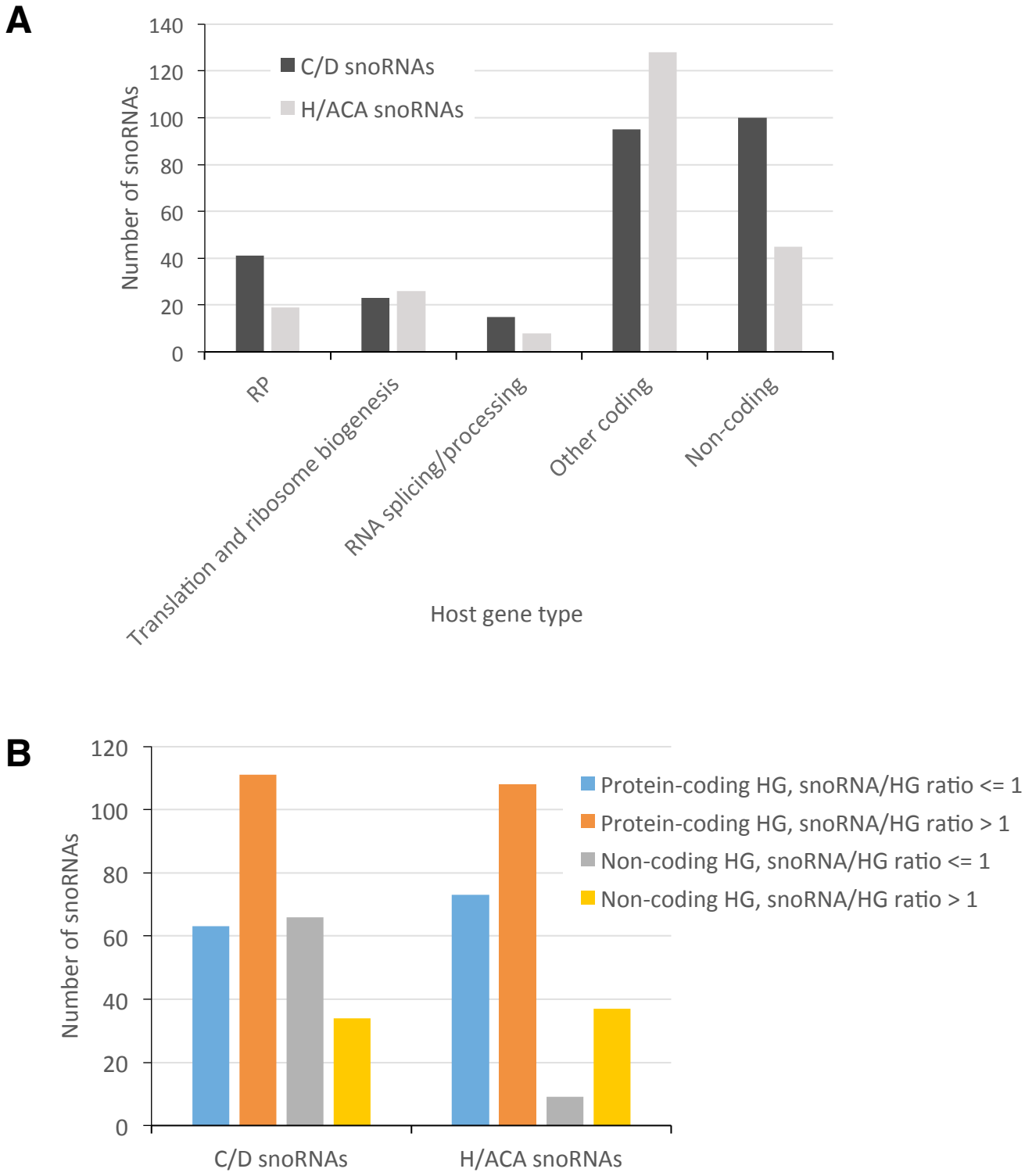


Figure S16 (related to figure 4)**.** (A) The number of snoRNAs encoded in host genes of different types is shown for both C/D snoRNAs (black columns) and H/ACA snoRNAs (pale grey columns). (B) The number of snoRNAs in coding or non-coding host genes (HG) and with abundance lower (snoRNA/HG ratio <= 1) or higher (snoRNA/HG ratio >1) than their host gene (HG) is shown for both C/D and H/ACA snoRNAs.

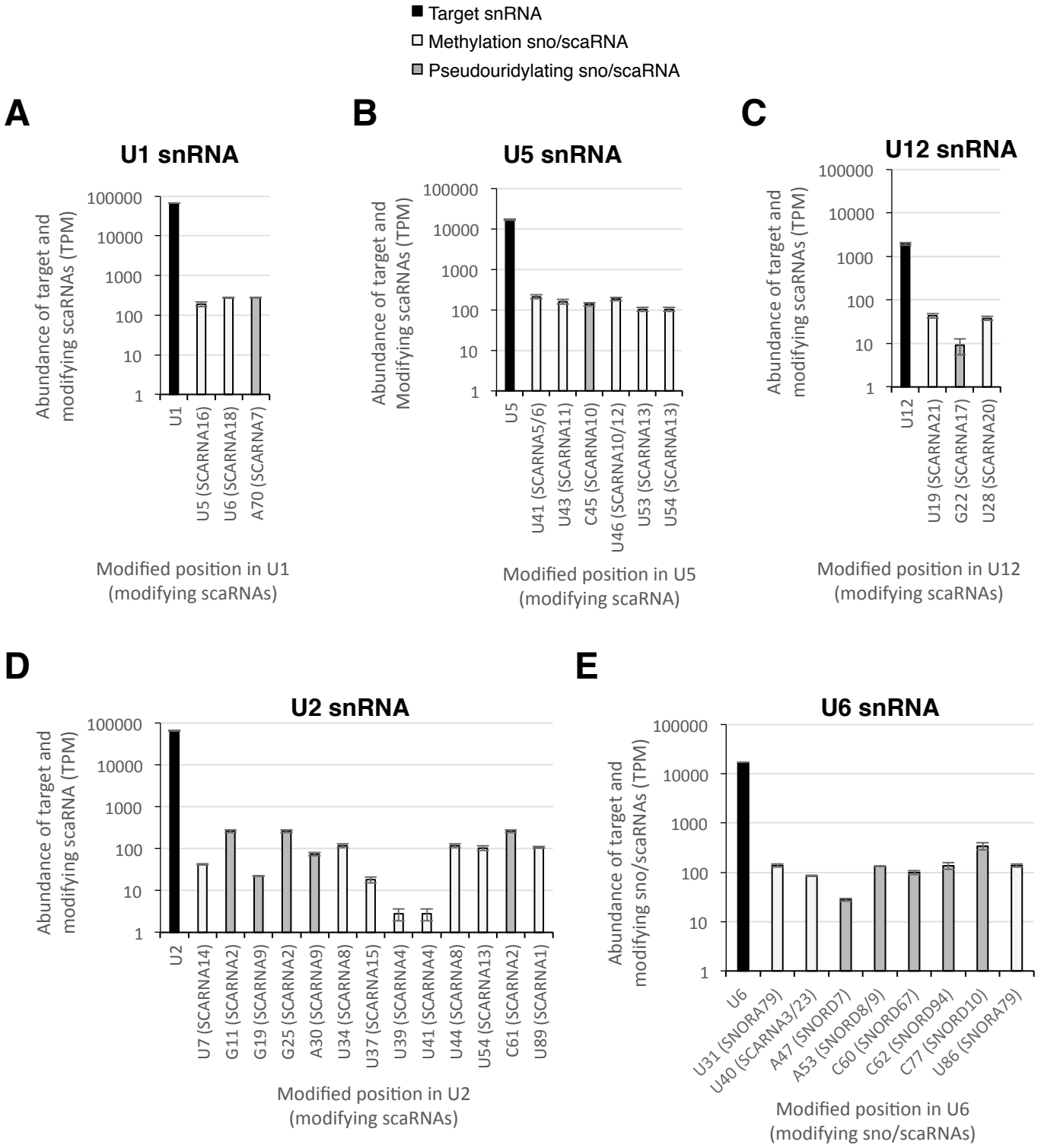


Figure S17 (related to figure 5). Relative abundance of the scaRNA/snoRNA and their snRNA modification targets. (A-E) Relative abundance of scaRNAs/snoRNAs that modify U1 (A), U5 (B), U12 (C), U2 (D) and U6 (E) snRNAs. The abundance of the snRNA (black column) and that of scaRNAs/snoRNAs corresponding to different methylation (white columns) or pseudouridylation sites (grey columns) within the snRNA are indicated in the form of bar graphs.

Dataset	Number of raw sequencing reads	Alignment rate
SSV_1	17116221	98.7%
SSV_2	15351481	98.6%
URT_1	49471618	99.2%
URT_2	37421215	99.3%
FAV_1	173632990	96.1%
FAV_2	200274722	97.6%
FRV_1	265305890	99.3%
FRV_2	292651704	99.3%
FRT_1	60971932	99.3%
FRT_2	41464929	98.8%

Table S1 (related to Figure 1). Statistics summarizing the datasets considered in this study. The alignment rate represents the proportion of raw reads aligned to the genome (see Methods).

	URT	FRT	SSV	FRV	FAV
Pearson	0.995	0.994	0.998	0.996	0.920
Spearman	0.791	0.905	0.696	0.933	0.919

Table S2 Pearson and Spearman correlations between replicates for every sequencing method investigated.

Gene name	RNA type	Average total RNA abundance (TPM)
U4	snRNA (tri-snRNP) major spliceosome	15736.99
U5	snRNA (tri-snRNP) major spliceosome	16777.28
U6	snRNA (tri-snRNP) major spliceosome	17080.25
U4atac	Minor spliceosome	1260.513
U6atac	Minor spliceosome	1233.739

Table S3 (related to Figure 1). Average total abundance (TPM) of the major and minor spliceosome tri-snRNP RNAs, according to FRT.

Table S5 (related to Figure 2 and Table S4). List of all protein-coding genes with abundance greater than 100 TPM in SKOV3ip1.

Gene id	Gene name	Gene length	SKOV3ip1 FRT average abundance (TPM)	SKOV3ip1 protein coding rank	INOF FRT average abundance (TPM)	INOF protein coding rank
ENSG00000156508	EEF1A1	8047	1339.887	1	502,8535	3
ENSG00000197956	S100A6	1645	1052.295	2	90,9755	40
ENSG00000124942	AHNAK	122693	904.1208	3	196,0134	14
ENSG00000167996	FTH1	7942	895.2919	4	145,6419	22
ENSG00000167658	EEF2	9413	697.0203	5	365,7956	5
ENSG00000034510	TMSB10	1050	647.222	6	249,5061	11
ENSG00000205542	TMSB4X	2119	642.129	7	605,5317	1
ENSG00000134352	IL6ST	59898	573.8624	8	556,8915	2
ENSG00000074800	ENO1	18247	442.8957	9	58,3131	83
ENSG00000136527	TRA2B	22230	437.183	10	61,6651	77
ENSG00000184009	ACTG1	13876	432.2191	11	263,0431	10
ENSG00000124486	USP9X	150944	404.7235	12	113,1335	29
ENSG00000108298	RPL19	4444	366.8748	13	63,3646	71
ENSG00000089157	RPLP0	4549	363.3004	14	104,8120	32
ENSG00000229117	RPL41	1357	353.9832	15	101,3086	35
ENSG00000067225	PKM	32793	328.9003	16	83,5513	45
ENSG00000134871	COL4A2	207215	305.281	17	448,2319	4
ENSG00000231500	RPS18	4500	299.9977	18	114,3638	28
ENSG00000112306	RPS12	3123	295.687	19	88,7502	41
ENSG00000092841	MYL6	5335	278.0255	20	193,9721	15
ENSG00000196154	S100A4	6523	269.9995	21	0,0910	10953
ENSG00000152409	JMY	91026	268.4495	22	7,8367	797
ENSG00000140988	RPS2	2808	266.6582	23	61,2506	79
ENSG00000182718	ANXA2	55749	264.5948	24	62,5462	74
ENSG00000145425	RPS3A	5079	263.4244	25	87,7159	43
ENSG00000187837	HIST1H1C	641	259.7633	26	37,9529	134
ENSG00000008988	RPS20	7286	242.7897	27	75,9861	52
ENSG00000141736	ERBB2	42512	242.4466	28	0,8758	5955
ENSG00000092148	HECTD1	107692	240.1952	29	55,2704	92
ENSG00000134333	LDHA	14037	237.9606	30	34,6958	150
ENSG00000138326	RPS24	23052	235.3905	31	81,4192	48
ENSG00000105974	CAV1	36399	234.7079	32	30,1625	175
ENSG00000026025	VIM	9334	229.3214	33	142,4967	23
ENSG00000136942	RPL35	4101	227.7768	34	54,3355	95
ENSG00000111640	GAPDH	4447	226.3353	35	132,7998	24
ENSG00000273542	HIST1H4K	311	224.319	36	31,8492	164
ENSG00000087086	FTL	1577	221.6377	37	196,0283	13
ENSG00000197061	HIST1H4C	434	220.2348	38	20,1610	265
ENSG00000172757	CFL1	39004	219.8038	39	91,8462	38
ENSG00000196262	PPIA	27884	219.6427	40	74,2481	53
ENSG00000147403	RPL10	19213	217.8391	41	131,2604	25
ENSG00000206503	HLA-A	4624	215.6137	42	76,3862	51
ENSG00000123416	TUBA1B	3615	214.2238	43	48,1335	112
ENSG00000133112	TPT1	7899	214.0906	44	104,1897	34
ENSG00000274267	HIST1H3B	410	213.8559	45	26,3025	195
ENSG00000163682	RPL9	4824	209.5798	46	66,7768	64
ENSG00000147604	RPL7	5522	201.5211	47	57,3043	85
ENSG00000142541	RPL13A	4754	201.2846	48	87,8047	42
ENSG00000198804	MT-CO1	1541	199.4792	49	91,5667	39
ENSG00000065978	YBX1	19922	196.4635	50	22,4775	231
ENSG00000181163	NPM1	24021	196.055	51	34,6802	151
ENSG00000177954	RPS27	1391	195.9607	52	72,4902	56
ENSG00000136813	KIAA0368	124053	195.1797	53	39,0286	130
ENSG00000117450	PRDX1	12011	193.7347	54	28,6494	181
ENSG00000198938	MT-CO3	783	192.1585	55	67,8074	61
ENSG00000089009	RPL6	13648	190.5738	56	49,3331	110
ENSG00000131469	RPL27	4686	189.6372	57	49,6294	109
ENSG00000162434	JAK1	133275	188.1401	58	52,9018	99
ENSG00000166710	B2M	7400	187.1137	59	154,4836	21

Table S5 (related to Figure 2 and Table S4). List of all protein-coding genes with abundance greater than 100 TPM in SKOV3ip1 (continued).

Gene id	Gene name	Gene length	SKOV3ip1 FRT average abundance (TPM)	SKOV3ip1 protein coding rank	INOF FRT average abundance (TPM)	INOF protein coding rank
ENSG00000142937	RPS8	3528	185,2385944	60	83,8907501	44
ENSG00000112715	VEGFA	16303	184,2909183	61	54,0231466	96
ENSG00000149273	RPS3	22794	180,6468841	62	66,4748811	65
ENSG00000164587	RPS14	6566	180,3798057	63	62,2891248	75
ENSG00000187514	PTMA	6646	179,747785	64	27,9701715	186
ENSG00000150991	UBC	5764	177,5098571	65	50,0029955	106
ENSG00000156482	RPL30	21618	177,3616229	66	56,5569255	86
ENSG00000142676	RPL11	4646	174,1782902	67	58,2572684	84
ENSG00000171863	RPS7	5714	174,0813192	68	65,5897689	68
ENSG00000091527	CDV3	16531	173,4036012	69	30,2522297	174
ENSG00000177600	RPLP2	3233	173,1158956	70	32,6971761	159
ENSG00000152818	UTRN	567333	172,4630631	71	51,4099762	102
ENSG00000075624	ACTB	36633	171,7495056	72	274,734432	9
ENSG00000197238	HIST1H4J	372	167,0785797	73	20,9283363	251
ENSG00000182774	RPS17	3811	166,4324179	74	55,481299	88
ENSG00000130255	RPL36	16929	166,2935791	75	25,9065965	197
ENSG00000142534	RPS11	3324	165,8401624	76	60,5340947	80
ENSG00000114391	RPL24	5691	162,4745028	77	43,0789838	121
ENSG00000213741	RPS29	28180	159,6994516	78	72,9076766	54
ENSG00000171858	RPS21	1404	157,6144421	79	43,3450998	120
ENSG00000168298	HIST1H1E	753	157,2683117	80	16,8638879	312
ENSG00000276368	HIST1H2AJ	386	156,5591782	81	9,66324206	621
ENSG00000197958	RPL12	3731	155,3231812	82	49,6685886	108
ENSG00000148303	RPL7A	3212	154,946237	83	66,2836093	66
ENSG00000105193	RPS16	2741	153,791609	84	47,0460768	113
ENSG00000135486	HNRNPA1	6895	148,4653045	85	21,1746443	248
ENSG00000197746	PSAP	35077	148,1646493	86	33,0682135	158
ENSG00000265972	TXNIP	4165	146,2784482	87	0,19873687	9890
ENSG00000161016	RPL8	2822	145,5460882	88	36,5322115	143
ENSG00000150093	ITGB1	105473	145,2016204	89	154,960716	20
ENSG00000198918	RPL39	5139	145,2014496	90	68,4338141	60
ENSG00000135046	ANXA1	18636	143,1383588	91	21,445788	242
ENSG00000143384	MCL1	5187	142,7862826	92	45,5809586	116
ENSG00000197903	HIST1H2BK	380	142,0714021	93	13,8727438	404
ENSG00000198755	RPL10A	2377	141,9584609	94	55,1202809	93
ENSG00000188643	S100A16	6259	141,7992867	95	16,2330151	333
ENSG00000204628	RACK1	11187	141,2429964	96	55,3500504	90
ENSG00000198034	RPS4X	21621	140,5275915	97	50,0375071	105
ENSG00000054654	SYNE2	373450	140,2683116	98	6,6372242	946
ENSG00000105640	RPL18A	4277	139,8018831	99	48,3306504	111
ENSG00000149925	ALDOA	17367	138,7425634	100	25,8808696	198
ENSG00000104904	OAZ1	3981	138,3287992	101	66,0837119	67
ENSG00000006327	TNFRSF12A	3938	137,2926352	102	24,1175077	216
ENSG00000092199	HNRNPC	60358	136,8907026	103	26,1170762	196
ENSG00000185567	AHNAK2	41113	133,0718495	104	6,23526232	1002
ENSG00000143947	RPS27A	3950	132,0883115	105	36,3898426	144
ENSG00000100316	RPL3	7507	128,3814989	106	36,9111206	139
ENSG00000183255	PTTG1IP	24318	125,8833367	107	12,9315733	437
ENSG00000109971	HSPA8	5741	124,1112858	108	69,9055331	58
ENSG00000100644	HIF1A	52746	123,7407058	109	15,6979608	350
ENSG00000137154	RPS6	4539	120,7080629	110	66,7939485	63
ENSG00000110955	ATP5B	7893	120,4608544	111	20,3948204	257
ENSG00000174718	KIAA1551	33737	120,2723751	112	0,64673628	7000
ENSG00000113013	HSPA9	20562	120,107456	113	9,08617756	674
ENSG00000110700	RPS13	3398	120,0083213	114	37,7496648	136
ENSG00000119335	SET	12976	119,438906	115	11,6716049	492
ENSG00000125968	ID1	1232	118,7424858	116	0,4814278	7878
ENSG00000272398	CD24	5796	118,6112141	117	0,02349353	12204
ENSG00000180596	HIST1H2BC	9053	118,3785749	118	8,44549448	730

Table S5 (related to Figure 2 and Table S4). List of all protein-coding genes with abundance greater than 100 TPM in SKOV3ip1 (continued).

Gene id	Gene name	Gene length	SKOV3ip1 FRT average abundance (TPM)	SKOV3ip1 protein coding rank	INOF FRT average abundance (TPM)	INOF protein coding rank
ENSG00000162402	USP24	148754	116,6951044	119	19,1912693	279
ENSG00000198830	HMGN2	3522	116,2212637	120	7,5537243	823
ENSG00000105185	PDCD5	6384	115,0676603	121	36,7063287	141
ENSG00000092820	EZR	53671	114,6510492	122	5,87333322	1066
ENSG00000127603	MACF1	405861	113,5020901	123	53,9057797	98
ENSG00000168003	SLC3A2	32773	113,2986823	124	6,84660323	923
ENSG00000168028	RPSA	5853	113,1458517	125	56,2421888	87
ENSG00000180530	NRIP1	104701	113,1254473	126	34,1441842	152
ENSG00000143933	CALM2	16519	110,804847	127	94,0137843	37
ENSG00000080824	HSP90AA1	58961	110,7757948	128	16,6242727	320
ENSG00000135404	CD63	4384	109,8503961	129	55,4410028	89
ENSG00000277157	HIST1H4D	311	109,3168177	130	18,0852479	298
ENSG00000106682	EIF4H	22856	108,250627	131	13,1474828	429
ENSG00000096384	HSP90AB1	7722	106,951343	132	20,3043054	260
ENSG00000083857	FAT1	138939	106,3755969	133	33,9005472	155
ENSG00000124635	HIST1H2BJ	6853	104,4675178	134	11,3798253	506
ENSG00000196924	FLNA	26114	102,2070771	135	62,1351628	76
ENSG00000137331	IER3	1355	101,8533498	136	20,5307306	255
ENSG00000178913	TAF7	60596	101,8260774	137	15,7953277	347
ENSG00000174444	RPL4	26517	101,3494604	138	45,8068504	114
ENSG00000132475	H3F3B	9459	101,3069759	139	14,6550447	381
ENSG00000136810	TXN	12829	100,5645483	140	21,2680071	244

GO term	GO_ID	# genes from background with this annotation (total background genes considered)	# genes with >100 TPM and with this annotation (total genes with >100 TPM considered)	Benjamini corrected p-value
translational elongation	GO:0006414	101 (13528)	47 (130)	1.6 E-65
translation	GO:0006412	331 (13528)	48 (130)	3.0 E-40
ribosomal small subunit biogenesis	GO:0042274	11 (13528)	7 (130)	1.3 E-07
nucleosome assembly	GO:0006334	84 (13528)	11 (130)	2.2 E-06
chromatin assembly	GO:0031497	87 (13528)	11 (130)	2.5 E-06
protein-DNA complex assembly	GO:0065004	91 (13528)	11 (130)	3.3 E-06
nucleosome organization	GO:0034728	93 (13528)	11 (130)	3.5 E-06
ribosome biogenesis	GO:0042254	122 (13528)	12 (130)	3.7 E-06
cellular macromolecular complex assembly	GO:0034622	318 (13528)	17 (130)	8.6 E-06
DNA packaging	GO:0006323	117 (13528)	11 (130)	2.2 E-05
cellular macromolecular complex subunit organization	GO:0034621	357 (13528)	17 (130)	3.4 E-05
chromatin assembly or disassembly	GO:0006333	127 (13528)	11 (130)	4.0 E-05
ribonucleoprotein complex biogenesis	GO:0022613	180 (13528)	12 (130)	1.2 E-04
cell motion	GO:0006928	475 (13528)	17 (130)	1.1 E-03
rRNA processing	GO:0006364	92 (13528)	8 (130)	2.4 E-03
rRNA metabolic process	GO:0016072	96 (13528)	8 (130)	3.0 E-03
ribosomal large subunit biogenesis	GO:0042273	10 (13528)	4 (130)	7.6 E-03
anti-apoptosis	GO:0006916	206 (13528)	10 (130)	1.2 E-02
response to calcium ion	GO:0051592	55 (13528)	6 (130)	1.2 E-02
macromolecular complex assembly	GO:0065003	665 (13528)	18 (130)	1.3 E-02
regulation of apoptosis	GO:0042981	804 (13528)	20 (130)	1.4 E-02
regulation of programmed cell death	GO:0043067	812 (13528)	20 (130)	1.5 E-02
regulation of cell death	GO:0010941	815 (13528)	20 (130)	1.5 E-02
homeostatic process	GO:0042592	751 (13528)	19 (130)	1.5 E-02
maintenance of location	GO:0051235	64 (13528)	6 (130)	1.9 E-02

Table S6 (related to Figure 2C). Top Gene Ontology (GO) enriched biological process terms for the group of protein-coding genes with highest abundance (>100 TPM).

Forward Name	Reverse Name	Forward Sequence	Reverse Sequence
RN7SK.q.F1	RN7SK.q.R1	CGGTCTTCGGTCAAGGGTATACG	AGCGCCTCATTTGGATGTGTCT
RN7SL2.q.F1	RN7SL2.q.R1	AGGTCGGAAACGGAGCAGGTC	CGGGGTCTCGCTATGTTGCT
RNY1.q.F1	RNY1.q.R1	TTGATTGTTCACAGTCAGTTACAGATCG	AGTCAAGTGCAGTAGTGAGAAGGG
RNY3.q.F1	RNY3.q.R1	TGGTGTTTACAACTAATTGATCACAACCAG	AGCAGTGGGAGTGGAGAAGGAACAAAG
SCARNA18.q.F1	SCARNA18.q.R1	TGCTTCTCATTCCTTGGGAGCA	TGTTGGAGAAATATACTACCACTCAACCT
SCARNA1.q.F1	SCARNA1.q.R1	ACCGAGCTGTCTATATCCTAGCCT	ACTGGGCTTAAAAGACTCATGGCT
SCARNA22.q.F1	SCARNA22.q.R1	GTCCTGACCTGTCCTCTGTGAGC	TGTACTGAAAACCGTGGTGTCCT
SNORA28.q.F1	SNORA28.q.R1	GGCAGATGATCAAAACTGTCTGACAC	AGTCTATATAACGGCTTGTCTCATGGG
SNORA34.q.F1	SNORA34.q.R1	AGACCAGCAGTTGTACTGTGGC	GCCATTCCCTACTGAGGTCCCA
SNORA44.q.F1	SNORA44.q.R1	GGGCTGTGGCTGGTCATAGC	AAAGCTGAGTGGCAGCTTGCAG
SNORA46.q.F1	SNORA46.q.R1	TCCCCATTCTTGGTTACGCTGT	TGTTCCTTAACTCTATACAGCAACAGCA
SNORA63.q.F1	SNORA63.q.R1	TAAGTGCTGTGTTGTCGTTCCCC	TATGAGACCAAGCGTCCCTGGC
SNORA64.q.F1	SNORA64.q.R1	AGTTGCACTTGGCTTCACCCG	GCACCCCTCAAGGAAAGAGAGG
SNORA68.q.F1	SNORA68.q.R1	GAATCACTGTTTCTTATAGCGGTGGTT	AAATTCACTTTGAGGGGCACGG
SNORD16.q.F1	SNORD16.q.R1	AATTTGCGTCTTACTCTGTTCTCAGC	TCAGTAAGAATTTTCGTCAACCTTCTGTAC
SNORD32A.q.F1	SNORD32A.q.R1	AACATTCACCATCTTTCGTTTGAGTCTCAC	GTCTCAGAGCGGTGCATGGG
SNORD46.q.F1	SNORD46.q.R1	AAAAGAATCCTTAGGCGTGGTTGTG	CAGTCAGTGTAACTATGACAAGTCCTTG
SNORD67.q.F1	SNORD67.q.R1	GTTGCACACTGGTGGAGCCATG	GAGTCAGATGGCCCCTGTGC
SNORD83A.q.F1	SNORD83A.q.R1	AGGCTCAGAGTGAGCGCTGG	GTTCTCAGAAGGAAGGCAGTAGAGAA
SNORD89.q.F1	SNORD89.q.R1	ACAAGAAAAGGCCGAATTGCAGT	GAGGTCAGACTAGTGGTTCGCTT
VTRNA1-1.q.F1	VTRNA1-1.q.R1	TCAGCGGTTACTTCGACAGTTCT	AGGACTGGAGAGCGCCCG
VTRNA1-2.q.F1	VTRNA1-2.q.R1	TCAGCGGTTACTTCGAGTACATTGT	AGAGCTGGAAAGCACCCGCG
SNORA23.q.F1	SNORA23.q.R1	GATCTTGCTATCCACACAAACATCATGC	TCCAGAGACAACACTAGACCAGTACT
SNORD124.q.F1	SNORD124.q.R1	GGATGATGTTCCAGTTGAGACTCAAGAA	GGTCAGGGACCAAGTGGCTCC
SNORD13.q.F1	SNORD13.q.R1	AGCGTGATGATTGGGTGTTCATACG	CAGACGGGTAATGTGCCCACG
SNORD88C.q.F1	SNORD88C.q.R1	AGCACTGGGCTCTGATCACCC	CCTCAGACCCCCAGGTGTCAA

Table S10 (related to Figure S6). List of primers used for digital PCR analysis of sncRNA