

Supplementary material for the article :

“RNA editing level in the mouse is determined by the genomic repeat repertoire“

Yossef Neeman^{1,2}, Erez Y. Levanon^{1,3}, Michael F. Jantsch⁴ and Eli Eisenberg⁵

¹*Compugen Ltd., 72 Pinchas Rosen St., Tel-Aviv 69512, Israel*

²*Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel.*

³*Department of Pediatric Hemato-Oncology, Safra Children's Hospital, Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.*

⁴*Max F. Perutz Laboratories, Dept. of Chromosome Biology, University of Vienna, Dr. Bohrgasse 1, A-1030 Vienna, Austria*

⁵*School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

Supplementary Table 1	2
Supplementary Table 2	3

Supplementary Table 1. Mismatches distribution within different human repeat families

		Alu	L1	MIR	L2	MaLR	MER1_type
AG	Within repeat	28540	842	153	95	313	82
	Paired repeats	25122	585	94	27	160	25
GA	Within repeat	123	35	20	21	16	16
	Paired repeats	43	12	9	8	0	4
CT	Within repeat	100	16	15	10	5	3
	Paired repeats	42	6	8	3	0	0
TC	Within repeat	390	97	30	28	29	17
	Paired repeats	282	28	17	6	6	3

Same as table 2 for human. The number of mismatches within each repeat-family is presented. Only mismatches that are part of clusters (at least three consecutive identical mismatches) are included. Paired repeats are repeats within exons for which the closest inverted repeat of the same family resides within 2000nt.

Supplementary Table 2: Characteristics of the six main repeat families in human.

Family		All repeats			All repeats within RNAs			All repeats within exons			Paired repeats within exons			Paired repeat with a significant BLAST hit		
		Mean	STD	Median	Mean	STD	Median	Mean	STD	Median	Mean	STD	Median	Mean	STD	Median
	Number	1169709			278512			14248			8147			6247		
Alu	Length	261	72	295	259	75	294	263	70	296	261	71	295	274	62	298
	Divergence	13	4	12	13	4	12	12	4	12	12	4	12	12	4	11
	Number	910257			132632			5807			1855			51		
L1	Length	547	832	287	472	729	258	605	980	293	626	945	319	1908	2075	1072
	Divergence	21	8	22	21	8	22	22	8	23	21	8	22	17	11	16
	Number	576712			148219			7170			3199			7		
MIR	Length	142	58	140	142	59	141	139	58	138	139	57	138	141	58	184
	Divergence	29	5	29	28	5	29	28	5	28	28	5	28	28	5	32
	Number	418535			91425			4112			1152			2		
L2	Length	222	223	138	215	217	133	213	217	131	227	232	142	1107	756	932
	Divergence	30	5	31	30	5	30	29	5	30	29	5	30	29	2	15
	Number	329494			36499			2798			576			23		
MaLR	Length	325	248	343	330	246	349	350	247	356	335	243	352	408	374	374
	Divergence	21	6	21	20	7	20	20	6	21	21	6	21	17	6	17

Same as table 3 for human. The data in this table was extracted from RepeatMasker, (<http://www.repeatmasker.org/>). Number is the copy-number of repeats of the given family. The length is in nucleotides, and the divergence values are given in percents.