

New characteristics of miRNA and isomiR interactions with mRNA

Matthew Weston¹, Rony Chowdhury Ripan¹, Xiaoman Li^{2,*}, Haiyan Hu^{1,*}

¹Department of Computer Science, University of Central Florida, Orlando, Florida, United States of America,

²Burnett School of Biomedical Science, College of Medicine, University of Central Florida, Orlando, Florida, United States of America.

*Corresponding Authors:

Email: xiaoman@mail.ucf.edu, haihu@cs.ucf.edu

Table S1: Information on the 67 samples used in our study. Please refer to table_s1.tsv

sample	dataset	instrument	# bases	BioSample	GEO sample	cell type	donor
SRR95975 1	clash	Illumina Genome Analyzer Iix	2.9G	SAMN0233 9506	GSM1219 487	Flp-In T-REx 293	female fetus
SRR95975 2	clash	Illumina Genome Analyzer Iix	2.3G	SAMN0233 9502	GSM1219 488	Flp-In T-REx 293	female fetus
SRR95975 3	clash	Illumina Genome Analyzer Iix	2.3G	SAMN0233 9502	GSM1219 488	Flp-In T-REx 293	female fetus
SRR95975 4	clash	Illumina Genome Analyzer Iix	1.6G	SAMN0233 9509	GSM1219 489	Flp-In T-REx 293	female fetus
SRR95975 5	clash	Illumina Genome Analyzer Iix	1.6G	SAMN0233 9509	GSM1219 489	Flp-In T-REx 293	female fetus
SRR95975 6	clash	Illumina Genome Analyzer Iix	3.5G	SAMN0233 9504	GSM1219 490	Flp-In T-REx 293	female fetus
SRR95975 7	clash	Illumina Genome Analyzer Iix	3.5G	SAMN0233 9504	GSM1219 490	Flp-In T-REx 293	female fetus
SRR95975 8	clash	Illumina Genome Analyzer Iix	1.2G	SAMN0233 9510	GSM1219 491	Flp-In T-REx 293	female fetus
SRR95975 9	clash	Illumina Genome Analyzer Iix	981.2 M	SAMN0233 9508	GSM1219 492	Flp-In T-REx 293	female fetus
SRR24131 75	clear_c lip	Illumina HiSeq 2500	1.9G	SAMN0408 8102	GSM1881 504	Huh-7.5	Male 57y
SRR24131 76	clear_c lip	Illumina HiSeq 2500	645.5 M	SAMN0408 8103	GSM1881 505	Huh-7.5	Male 57y
SRR24131 77	clear_c lip	Illumina HiSeq 2500	1.1G	SAMN0408 8104	GSM1881 506	Huh-7.5	Male 57y
SRR24131 78	clear_c lip	Illumina HiSeq 2500	1.9G	SAMN0408 8105	GSM1881 507	Huh-7.5	Male 57y
SRR24131 79	clear_c lip	Illumina HiSeq 2500	589.6 M	SAMN0408 8106	GSM1881 508	Huh-7.5	Male 57y
SRR24131 80	clear_c lip	Illumina HiSeq 2500	333.4 M	SAMN0408 8107	GSM1881 509	Huh-7.5	Male 57y
SRR24131 81	clear_c lip	Illumina HiSeq 2500	651.4 M	SAMN0408 8077	GSM1881 510	Huh-7.5	Male 57y
SRR24131 82	clear_c lip	Illumina HiSeq 2500	439.1 M	SAMN0408 8078	GSM1881 511	Huh-7.5	Male 57y

SRR24131 83	clear_c lip	Illumina HiSeq 2500	1.1G	SAMN0408 8079	GSM1881 512	Huh-7.5	Male 57y
SRR24131 84	clear_c lip	Illumina HiSeq 2500	590.7 M	SAMN0408 8080	GSM1881 513	Huh-7.5	Male 57y
SRR24131 85	clear_c lip	Illumina HiSeq 2500	668.2 M	SAMN0408 8113	GSM1881 514	Huh-7.5	Male 57y
SRR24131 86	clear_c lip	Illumina HiSeq 2500	495.7 M	SAMN0408 8081	GSM1881 515	Huh-7.5	Male 57y
SRR18280 894		Illumina NovaSeq 6000		SAMN2654 2892	GSM5942 235	HEK293T	female fetus
SRR18280 895	eclip	Illumina NovaSeq 6000	10.7G	SAMN2654 2893	GSM5942 234	HEK293T	female fetus
SRR18280 896	eclip	Illumina NovaSeq 6000	8.5G	SAMN2654 2894	GSM5942 233	HEK293T	female fetus
SRR18280 897	eclip	Illumina NovaSeq 6000	8.7G	SAMN2654 2895	GSM5942 232	HEK293T	female fetus
SRR18281 007	eclip	Illumina NovaSeq 6000	2.4G	SAMN2654 2857	GSM5942 270	HEK293T	female fetus
SRR18281 008	eclip	Illumina NovaSeq 6000	2.3G	SAMN2654 2858	GSM5942 269	HEK293T	female fetus
SRR18281 011	eclip	Illumina NovaSeq 6000	2.6G	SAMN2654 2861	GSM5942 266	HEK293T	female fetus
SRR18281 013	eclip	Illumina NovaSeq 6000	3.1G	SAMN2654 2863	GSM5942 264	HEK293T	female fetus
SRR18281 057	eclip	Illumina NovaSeq 6000	4.4G	SAMN2654 3029	GSM5942 098	HEK293T	female fetus
SRR18281 059	eclip	Illumina NovaSeq 6000	2.3G	SAMN2654 3031	GSM5942 096	HEK293T	female fetus
SRR18281 060	eclip	Illumina NovaSeq 6000	1.2G	SAMN2654 3033	GSM5942 094	HEK293T	female fetus
SRR18281 061	eclip	Illumina NovaSeq 6000	2.4G	SAMN2654 3032	GSM5942 095	HEK293T	female fetus
SRR18281 062	eclip	Illumina NovaSeq 6000	1.1G	SAMN2654 3034	GSM5942 093	HEK293T	female fetus
SRR13415 087	fields	Illumina HiSeq 3000	20.7G	SAMN1729 3559	GSM5015 717	HCT116	Male 48y
SRR13415 088	fields	Illumina HiSeq 3000	6.9G	SAMN1729 3558	GSM5015 718	HCT116	Male 48y
SRR13415	fields	Illumina HiSeq 3000	12.9G	SAMN1729	GSM5015	HCT116	Male 48y

089					3557	719		
SRR13415					SAMN1729	GSM5015		
090	fields	Illumina HiSeq 3000	15G	3556	720	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
091	fields	Illumina HiSeq 3000	25.2G	3555	721	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
092	fields	Illumina HiSeq 3000	10.4G	3554	722	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
093	fields	Illumina HiSeq 3000	5.4G	3553	723	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
094	fields	Illumina HiSeq 3000	8G	3552	724	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
095	fields	Illumina HiSeq 3000	7.7G	3551	725	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
096	fields	Illumina HiSeq 3000	4.9G	3550	726	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
097	fields	Illumina HiSeq 3000	6.4G	3549	727	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
098	fields	Illumina HiSeq 3000	6.2G	3548	728	HCT116		Male 48y
SRR13415					SAMN1729	GSM5015		
099	fields	Illumina HiSeq 3000	2.5G	3547	729	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
301	fields	Illumina HiSeq 3000	17.3G	3004	904	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
302	fields	Illumina HiSeq 3000	5G	3003	905	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
303	fields	Illumina HiSeq 3000	18.6G	3002	906	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
304	fields	Illumina HiSeq 3000	4.4G	3013	907	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
305	fields	Illumina HiSeq 3000	11.6G	3011	908	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
306	fields	Illumina HiSeq 3000	4.6G	3010	909	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
307	fields	Illumina HiSeq 3000	2.9G	3009	910	HCT116		Male 48y
SRR14432					SAMN1902	GSM5282		
308	fields	Illumina HiSeq 3000	3.1G	2997	911	HCT116		Male 48y
ERR37611					SAMEA641			
53	kozer	Illumina HiSeq 3000	23.2G	3584		A375		Female 54y

ERR37611				SAMEA641				Female
54	kozer	Illumina HiSeq 3000	22.4G	3585		A375		54y
ERR37611				SAMEA641				Female
55	kozer	Illumina HiSeq 3000	22.4G	3586		A375		54y
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
47	qclash	Illumina HiSeq 2500	6.3G	9957	017	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
48	qclash	Illumina HiSeq 2500	12.5G	9956	018	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
49	qclash	Illumina HiSeq 2500	11G	9955	019	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
50	qclash	Illumina HiSeq 2500	2G	9954	020	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
51	qclash	Illumina HiSeq 2500	2.5G	9953	021	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
52	qclash	Illumina HiSeq 2500	7.1G	9951	022	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
53	qclash	Illumina HiSeq 2500	7.7G	9950	023	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
54	qclash	Illumina HiSeq 2500	9.7G	9952	024	endothelial cells	Unknown	
SRR58769				SAMN0741	GSM2720	human umbilical vein	Sex	
55	qclash	Illumina HiSeq 2500	4.6G	9949	025	endothelial cells	Unknown	

Table S2: Reads in our samples. Please refer to
https://figshare.com/articles/dataset/Table_S1/28958909?file=54310421

Table S3: Statistical summary and Mann-Whitney two-sided score of comparisons of nucleotide compositions of miRNAs based on various criteria

Nucleotide	Conditions	Min (%)	Max (%)	Median (%)	Average (%)	Std (%)	Mann-Whitney
A	<1000 targets	0	0.636	0.190	0.205	0.112	4.62E-5
	>=1000 targets	0	0.412	0.25	0.236	0.098	
	1 isomiR	0	0.545	0.195	0.201	0.116	0.04
	>=2 isomiRs	0	0.524	0.227	0.225	0.106	
	1 isomiR	0	0.545	0.195	0.201	0.116	0.013
	>=5 isomiRs	0	0.391	0.261	0.233	0.094	
	1 - 4 isomiRs	0	0.545	0.209	0.206	0.117	0.015
	>=5 isomiRs	0	0.391	0.261	0.233	0.094	
	0 isomiRs	0	0.636	0.190	0.208	0.110	0.006
	>= 5 isomiRs	0	0.391	0.261	0.233	0.094	
C	<1000 targets	0	0.65	0.227	0.253	0.142	0.0420
	>=1000 targets	0	0.545	0.217	0.223	0.116	
	1 isomiR	0	0.65	0.227	0.269	0.166	0.118
	>=2 isomiRs	0	0.571	0.217	0.227	0.127	
	1 isomiR	0	0.65	0.227	0.269	0.166	0.021
	>=5 isomiRs	0	0.529	0.19	0.203	0.106	
	1 - 4 isomiRs	0	0.65	0.227	0.264	0.159	0.01
	>=5 isomiRs	0	0.529	0.19	0.203	0.106	
	0 isomiRs	0	0.625	0.227	0.249	0.135	0.007
	>= 5 isomiRs	0	0.529	0.19	0.203	0.106	
G	<1000 targets	0	0.75	0.273	0.301	0.155	0.477
	>=1000 targets	0.045	0.7	0.273	0.282	0.115	
	1 isomiR	0.045	0.722	0.286	0.309	0.161	0.96
	>=2 isomiRs	0.045	0.75	0.286	0.31	0.151	
	1 isomiR	0.045	0.722	0.286	0.309	0.161	0.960
	>=5 isomiRs	0.087	0.75	0.292	0.31	0.133	
	1 - 4 isomiRs	0.045	0.722	0.286	0.31	0.164	0.805
	>=5 isomiRs	0.087	0.75	0.292	0.31	0.133	
	0 isomiRs	0	0.750	0.273	0.294	0.148	0.192

	>= 5 isomiRs	0.087	0.75	0.292	0.31	0.133	
T	<1000 targets	0	0.727	0.227	0.241	0.116	0.026
	>=1000 targets	0.048	0.478	0.261	0.259	0.097	
	1 isomiR	0	0.727	0.217	0.22	0.116	0.138
	>=2 isomiRs	0	0.478	0.227	0.237	0.106	
	1 isomiR	0	0.727	0.217	0.22	0.116	0.0178
	>=5 isomiRs	0	0.458	0.235	0.254	0.106	
	1 - 4 isomiRs	0	0.727	0.217	0.22	0.112	0.009
	>=5 isomiRs	0	0.458	0.235	0.254	0.106	
	0 isomiRs	0	0.591	0.238	0.249	0.114	0.604
	>= 5 isomiRs	0	0.458	0.235	0.254	0.106	

Table S4. Mutation rates across miRNA positions.

Position	num isomiRs w/mutation	percent isomiRs w/mutation	p-value
3	88	0.054	0.052
4	88	0.054	0.052
5	81	0.05	0.205
6	81	0.05	0.205
7	94	0.058	0.011
8	101	0.062	0.001
9	102	0.063	0.0008
10	124	0.077	2.39e-8
12	108	0.007	7.03e-5
13	93	0.057	0.014
14	106	0.065	0.0002
15	137	0.085	7.36e-12
16	175	0.108	0
17	136	0.084	1.43e-11
18	104	0.064	0.0004
19	128	0.079	2.30e-9
20	89	0.055	0.041

There were 1620 total isomiRs where the canonical miRNA sequence was available in mirBase and that appeared at least 10 times in all the datasets combined. Table 4 shows the number of those isomiRs in a given position i , what percent of isomiRs had a mutation in that position, and the p-value calculate from $1\text{-pbinom}(k-1, 1620, 1/22)$ where k is the number of isomiRs that have a mutation in position i . We skipped over any positions where fewer than 5% of the isomiRs have a mutation there. Based on these results, the seed region is less prone to mutations than the non-seed region. Upon reaching the non-seed region, mutation rates spike greatly with peaks at position 10 and 16.

Table S5: Information on miRNAs with 10000 mRNA targets.

miRNA	Homolog group	# homologs (miRBase)	# homologs (our study)	# mRNA targets	# isomiRs
hsa-let-7a-5p	hsa-let-7	18	17	11327	20
hsa-let-7f-5p	hsa-let-7	18	17	10485	16
hsa-miR-16-5p	hsa-miR-16	3	3	11176	38
hsa-miR-17-5p	hsa-miR-17	2	2	11036	43
hsa-miR-19b-3p	hsa-miR-19	5	5	10401	35
hsa-miR-20a-5p	hsa-miR-20	4	3	10984	44
hsa-miR-21-5p	hsa-miR-21	2	2	11129	29
hsa-miR-92a-3p	hsa-miR-92	5	4	12911	50
hsa-miR-3168	hsa-miR-3168	1	1	11266	6

The table includes each miRNA's homolog group, how many homologs were found by us and existed in miRBase, how many mRNA targets we identified, and how many isomiRs we identified.

Table S6: 127 isomiRs with canonical sequence unfound in our datasets. Refer to https://figshare.com/articles/dataset/Table_S6/29876195?file=57067805

Table S7: A summary of the # of chimeric reads supporting miRNA-mRNA and isomiR-mRNA interactions for each of the 307 miRNAs with isomiRs in our datasets. Refer to https://figshare.com/articles/dataset/Table_S7/29876192?file=57067802