Toward a combinatorial nature of microRNA regulation in human cells

Ohad Balaga[1], Yitzhak Friedman[2], and Michal Linial[2,*]

[1]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel

[2]Deptartment of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Israel

Supplementary Figures

Figure S1. Coverage of targeted genes and miRNAs by MDBs.

Table S2. Global performance of all MDBs and miRror2.0

Figure S3. CLIP data properties

Supplementary Tables

Table S1. Coverage of targeted genes and miRNAs by MDBs.

Table S2. List and data on miRNA overexpressed experiments (Agilent and Affymetrix platform)

Table S3. miRNA families.

Table S4. Down-regulated genes from has-miR-124 overexpression experiment.

Table S5. Prediction performance.

Table S6. Integrated prediction performance.

Table S7. Predictions for MDBs and miRror for CLIP data.

Table S8. Pairs of miRNAs that were identified by CLIP data.

Table S9. The list of 20 cell-lines transcriptomic experiments.

Table S10. Joint miRIS and JI for 200 miRNA-Duos (from 5 top predictions).
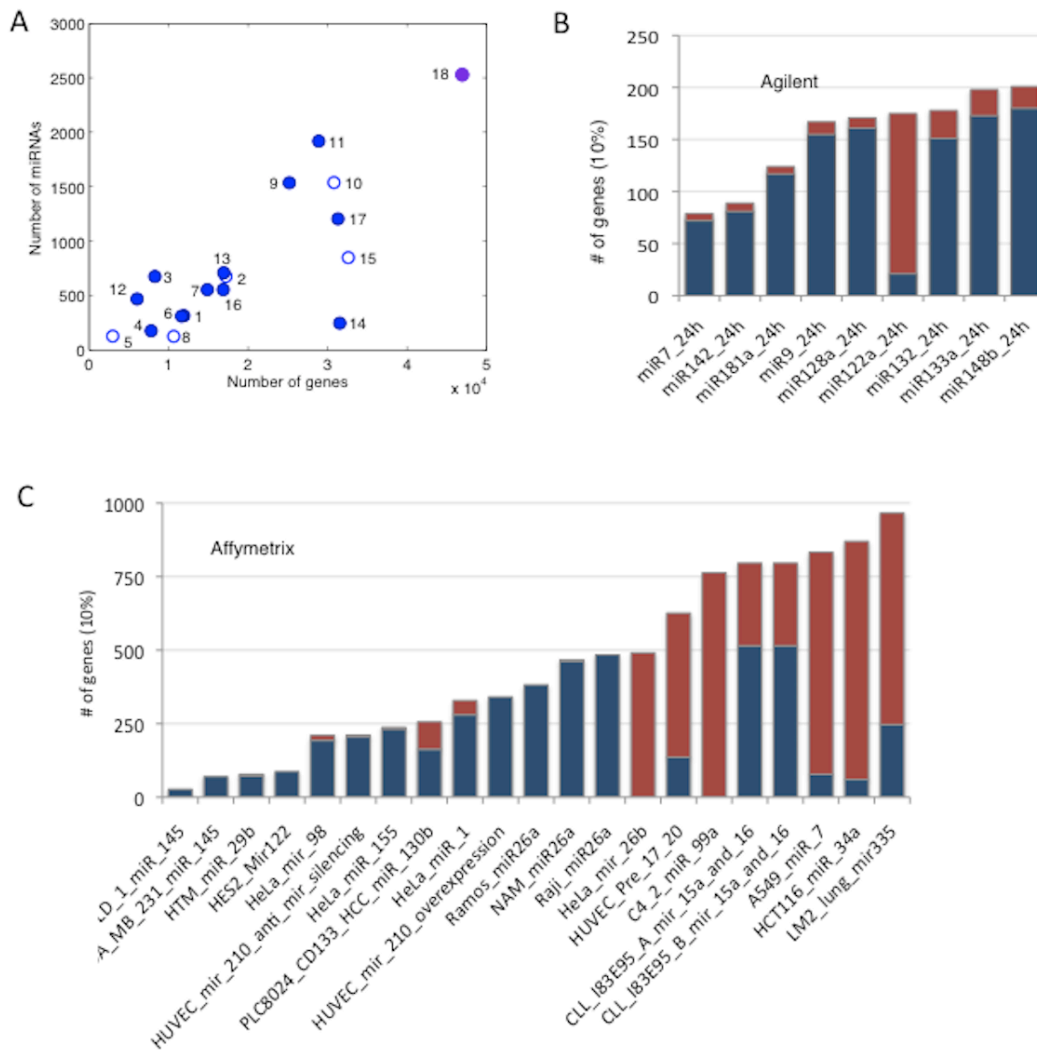
Supplementary Figures

Figure S1

Figure S1. miRNAs and target-genes and transcriptomic data. (A) A total of 17 pre-compiled sets are indicated. Number labels indicate each MDB: 1, MAMI; 2, PITA_all; 3, PITA_top; 4, PicTar 4way; 5, PicTar 5way; 6, RNA22; 7, TargetRank all; 8, TargetRank_conserved; 9, TargetScan_conserved; 10, TargetScan_nonconserved; 11, miRDB; 12, miRNAMap2; 13, microCosm; 4, microRNA.org_conserved; 15, microRNA.org_nonconserved; 16, microT; 17, MirZ; 18, A unified set that is supported by miRror (shown in purple). Filled circles are the MDBs that are further analyzed in this work. The coverage of miRNAs and gene-targets of each MDB differs significantly. For example, TargetScan covers ~1500 miRNA and ~25,000 genes, while miRNAMap2 covers 470 miRNA and ~6000 genes (*Supplementary Table S1)*. The down-regulated genes from miRNA overexpression experiments measured by Agilent (B) or Affymetrix (C) platforms (*Supplementary Table S2)*. The number of genes that were suppressed by a differential ratio >1.2 is shown. For example, for hsa-miR-145 (DLD-1 cells, GSE18625) only 270 genes meet the threshold while for hsa-miR-34a (HCT116 cells, GSE7754) this number reaches 8230 genes. The fraction of genes suppressed by a differential ≥3 is indicated in red. For most experiments, only few genes differentially expressed at this threshold (red bars). The amount of such genes is significantly lower in the data collected from Agilent platform (t-test, P value = 0.002).
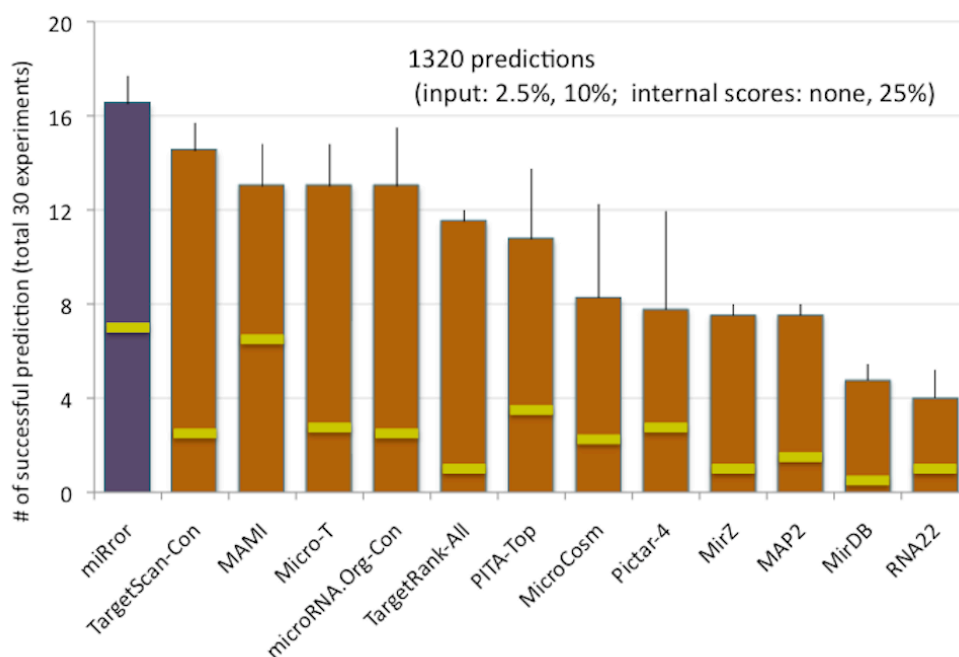
Figure 2

Figure S2. Integrated miRNA prediction success. The robustness of each MDB and miRror to parameters is tested. A total of 1440 prediction experiments are shown. Each bar is a summary of 120 predictions. Standard deviations were calculated from the average value of the prediction success for 4 sets (30 experiments, input of 2.5%, 10% with or without a filtration by internal score). The predictions are considered a success if at the top 9 predictions. Yellow line, the appropriate miRNA was the first in the ranked list. miRror predictions are colored purple. Source data for miRror performance in *Supplementary Table S6.*
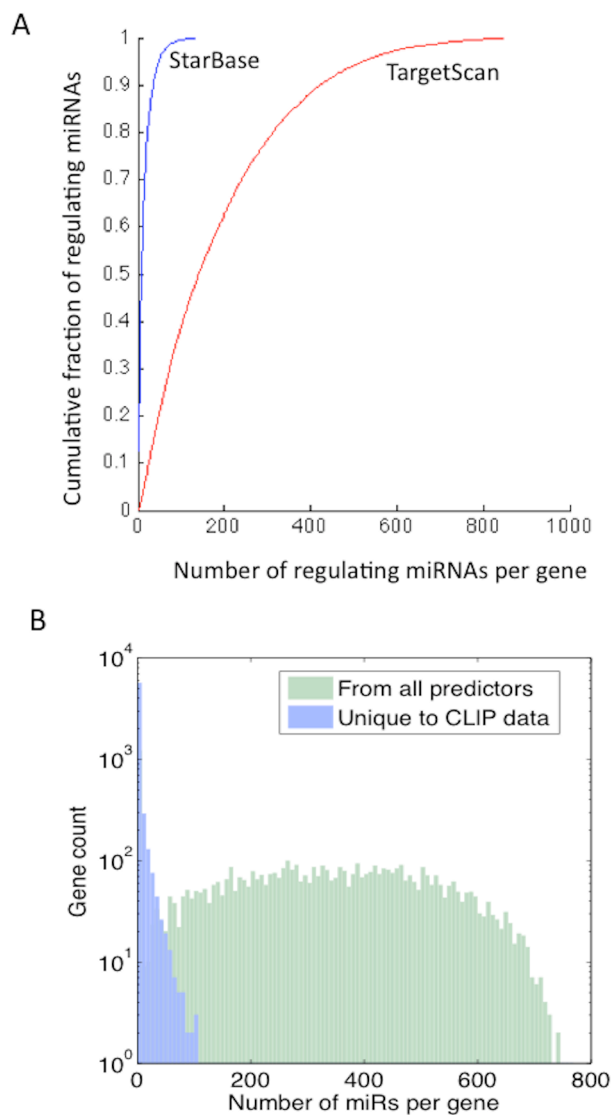
Figure S3

Figure S3. Statistics of the CLIP data. (A) The cumulative prediction of TargetScan in view of the actual numbers of miRNA that were collected from the unified human CLIP

data (from StarBase). The discrepancy between the real observation and the prediction is emphasized. (B) The number of genes according to the number of miRNAs per gene is plotted. Note that the distribution for the CLIP data differs from the union of the 12 MDBs. A predicted regulation of a gene by hundreds of miRNAs reflects the high number of false positives by the MDBs.

Supplementary Tables   *[Dataset] – Tables appear as unified xslx file*

Table S1. Coverage of targeted genes and miRNAs by all tested MDBs.

Table S2. [Dataset] Overexpressed miRNA experiments according to their platforms.

Table S3. miRNA families.

Table S4. [Dataset] Down-regulated genes from an overexpression experiment of hsa-miR-124.

Table S5. [Dataset] Prediction performance (top) and internal scoring (bottom)

Table S6. [Dataset] Integrated prediction performance (top) and internal scoring (bottom).

Table S7. [Dataset] Predictions for MDBs and miRror for CLIP data.

Table S8. Pairs of miRNAs that were identified by CLIP data.

Table S9. The list of the 20 transcriptome cell-lines transcriptomic experiments.

Table S10. [Dataset] Joint miRIS and JI for 200 miRNA-Duos (from 5 top predictions).

Table S1. Coverage of targeted genes and miRNAs by MDBs.

DB	Genes	miRNAs
MAMI	11914	317
PITA all	17216	677
PITA top	8292	677
PicTar 4way	7817	178
PicTar 5way	2972	129
RNA22	11641	313
TargetRank all	14874	555
TargetRank conserved	10652	127
TargetScan	16785	249
TargetScan conserved	25186	1536
TargetScan nonconserved	30802	1538
miRDB	28890	1918
miRNAMap2	6040	470
microCosm	16933	711
microRNA.org	15400	677
microRNA.org conserved	31526	248
microRNA.org nonconserved	32601	850
microT	16882	555
mirZ	31302	1205
Combined (miRror)	46921	2540

Table S3. Compression of miRNA by a family assignment for 12 MDBs.

Name	All miRNAs	miRNA families (Seed based)[a]	miRNA families (miRBase)	# of genes (for miRNAs)	# of genes (for families)
miRDB	1918	1238	303	28890	25337
TargetScan_Con	1536	1035	406	25186	22571
mirZ	1205	858	396	31302	31179
microCosm	711	505	261	16933	16626
PITA_TOP	677	481	348	8292	7922
microT	555	384	261	16882	16784
TargetRank_all	555	384	261	14874	14147
miRNAMap2	470	282	260	6040	6035
MAMI	317	162	150	11914	11578
RNA22	313	160	148	11641	11491
microRNA.org_Con	248	146	125	31526	31191
PicTar_4way	178	84	98	7817	7535

[a]Seed families are defined by a shard seed sequence (nucleotides 2-7).

Table S8. Pairs of miRNAs that were identified by CLIP data. The JI is shown.

DB	miRs	pairs	JI mean	JI variance
microT	555	153735	0.13428	0.00441
mirZ	1206	726615	0.12883	0.00454
miRNAMap2	470	110215	0.12765	0.00609
microRNA.org_Con	249	30876	0.09604	0.00572
TargetRank_all	555	153735	0.05547	0.00299
RNA22	313	48828	0.04680	0.00136
PicTar_4way	178	15753	0.04522	0.00627
PITA_TOP	677	228826	0.03211	0.00180
microCosm	711	252405	0.03097	0.00032
TargetScan_Con	1537	1180416	0.02765	0.00095
MAMI	317	50086	0.01946	0.00219
miRDB	1919	1840321	0.01713	0.00056
StarBase (CLIP)	653	212878	0.02001	0.00093

Table S9. The list of the 20 transcriptomics cell-culture transcriptomic experiments.

GEO acc.	Condition	Cell line	Genes	Best Pair miR A	Best Pair miR B	Joint-miRIS
GSE12792	Hypoxia	RPTEC	47	miR-624	miR-26	0.640957
GSE10043	Hyperthermia	U937	114	miR-586	miR-153	0.47807
GSE7355	C. albicans infection	HUVEC	123	miR-137	miR-571	0.395325
GSE10212	Ultrasound	U937	105	miR-19	miR-30	0.697619
GSE9714	Interferon	Nu61	111	miR-203	miR-144	0.310811
GSE9714	Interferon	SCC-61	124	miR-322	miR-193	0.251344
GSE6927	S. gordonii infection	HIGK	98	miR-132	miR-568	0.402211
GSE6927	F. nucleatum infection	HIGK	92	miR-203	miR-26	0.626812
GSE8562	XBP1	MCF7	37	miR-154	miR-8	0.547297
GSE6263	PTEN -/-	HCT116	81	miR-452	miR-578	0.308642
GSE2067	HCV Core	HepG2	33	miR-621	miR-128	0.465909
GSE3407	CSB mutant	fibroblasts	110	miR-25	miR-145	0.367424
GSE4817	S1P	glioblastoma	44	miR-150	miR-345	0.401515
GSE4425	Hypothermia	THP1	107	miR-8	miR-155	0.357866
GSE2450	atorvastatin + mevalonate	HUVEC	141	miR-340	miR-103	0.468972
GSE2450	mevalonate	HUVEC	138	miR-568	miR-10	0.23913
GSE2450	atorvastatin	HUVEC	134	miR-944	miR-367	0.462687
GSE2451	camptothecin	293T	121	miR-218	miR-374	0.639807
GSE2668	ATRA + antibiotics	OCI/AML-2	126	miR-19	miR-595	0.492063
GSE2668	VPA + antibiotics	OCI/AML-2	126	miR-630	miR-217	0.367063