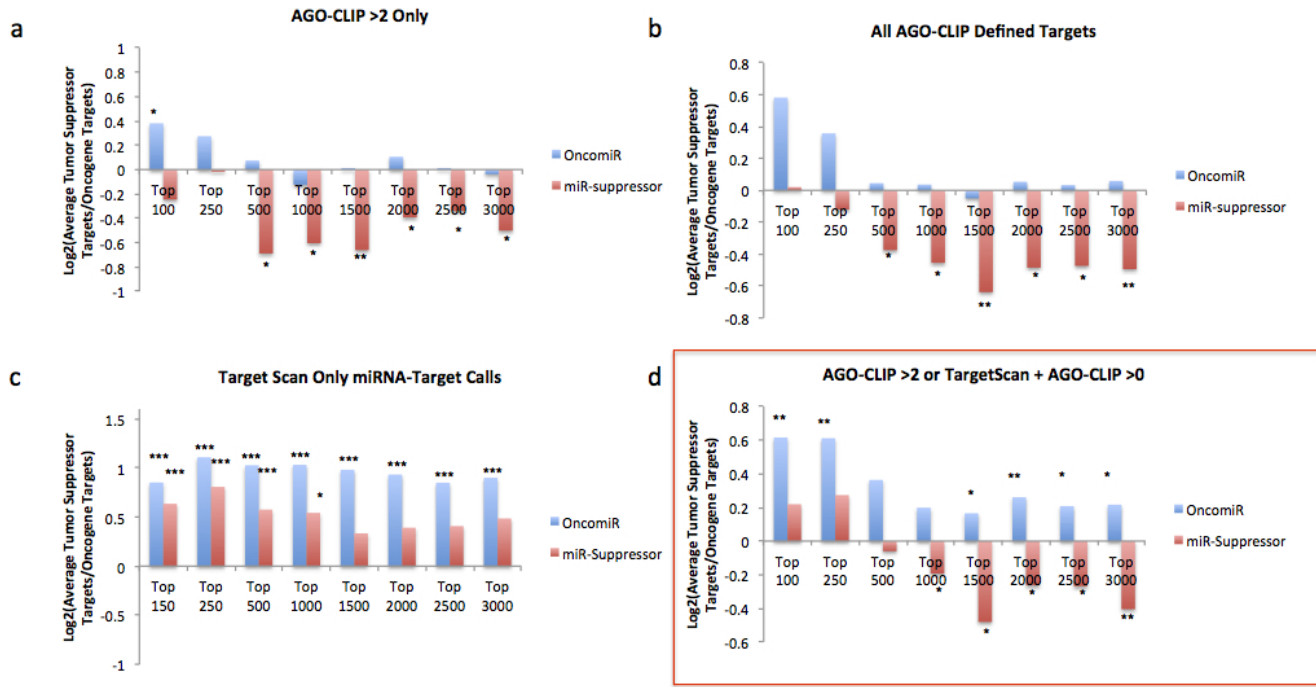
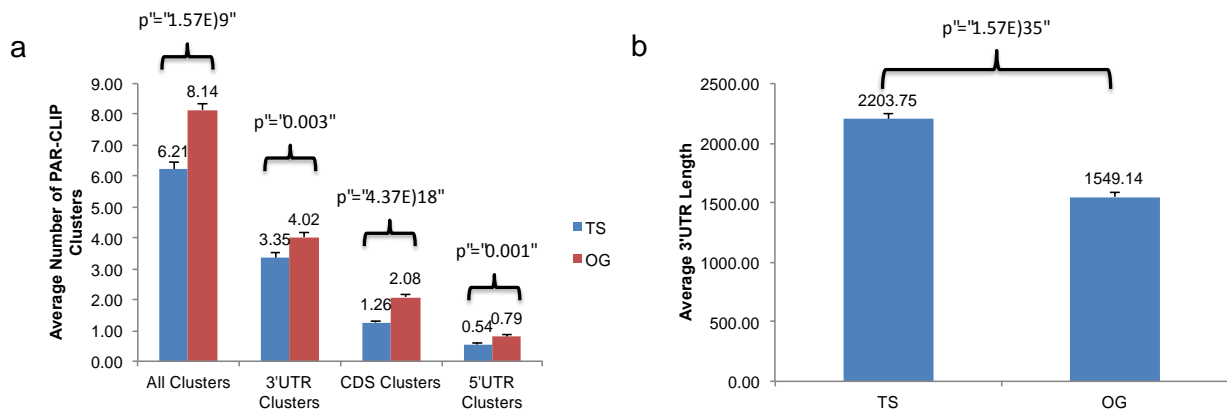


Supplementary Figure S1. Composition differences between tumor and normal due to large changes in highly expressed microRNAs are corrected through UQ or TMM normalization. A) There are significantly more significantly increased microRNAs than significantly decreased microRNAs in the TCGA miRNA data. Composition based normalization procedures correct for these deficits. Boxplot means are represented by horizontal line, boxes encompass the first and third quartile, whiskers represent 1.5x the inter-quartile range, outliers are represented as open circles outside the plot area. B) The total number of significant increases per tumor directly correlates with changes in *miR-143* levels (Pearson correlate, $n = 7$ tumor types, p -value calculated through T-Distribution). C) UQ normalization ablates the relationship between significant change and *miR-143* levels (Pearson correlate, $n = 7$ tumor types). D) Smear plots of representative UCEC data demonstrating successful normalization of miRNA sequencing data using normalization procedures that focus on correcting for composition differences. Red tags represent significant change ($p < 0.005$, Fisher's exact test). *** $p < 0.005$, UQ = upper-quartile normalization, TMM = Trimmed mean of M-values normalization, UCEC = Uterine corpus endometrioid carcinoma (representative tumor data). FC = Fold Change. CPM = Counts Per Million. Log values are log base two.



Supplementary Figure S2. MicroRNA-target enrichment analysis suggests a combination of AGO-CLIP and TargetScan data produces the best target nomination criteria as defined by significant enrichment of Tumor-Suppressor targets by oncomiRs and oncogene targets by miR-suppressors. Plots represent average percent tumor-suppressor versus percent oncomiR targets for Pan-Cancer oncomiRs and miR-suppressors. Graphs were generating by taking log2 ratio of tumor-suppressor over oncogene targets for each. A) AGO-CLIP defined targets with ≥ 3 occurrences enrich for oncogene targets of miR-suppressors only. B) All AGO-CLIP defined targets similarly enrich only for oncogene targets of miR-suppressors. C) TargetScan only target calls enrich for tumor suppressor targets. D) Combining AGO-CLIP data with TargetScan data (AGO-CLIP ≥ 3 or AGO-CLIP ≥ 1 + TargetScan) produces expected enrichments in tumor-suppressor targets for oncomiRs and oncogene targets for miR-suppressors. Student's T-Test, * $p < 0.05$, ** $p < 0.005$. N-values for enrichment reflect the total number of Pan-Cancer oncomiRs ($n=22$) and Pan-Cancer miR-suppressors ($n=25$). Further analysis was performed using method D (red box).



Supplementary Figure S3. Differential microRNA targeting and 3'UTR length of ranked tumor suppressors and oncogenes. A) Oncogenes have 31% more total clusters than tumor-suppressors, much of this differential targeting may derive from targeting of the coding region, which have on average 66% more total clusters than tumor suppressors. Data derives from the combination of 11 PAR-CLIP libraries used in this study, and reflects values for the top 3000 tumor suppressors and oncogenes. Error bars are s.e.m. . B) Tumor suppressors have, on average, 40% longer 3'UTRs than oncogenes. This may lead to increased identification of microRNA target sites on tumor suppressors by TargetScan. Data refers to the top 3000 tumor suppressors and oncogenes in our analysis (n=3000, p-values obtained with T-Test). Genomic coordinates for 3'UTR length derive from UCSC known genes. Error bars are s.e.m. TS = Tumor Suppressor, OG = Oncogene, 3'UTR = 3' untranslated region, CDS = coding sequence, 5'UTR = 5' untranslated region.

Supplementary Table S1. TCGA microRNA data inventory. All data was downloaded as isoform reads per million from the TCGA portal. List is fully up-to-date from the TCGA data portal as of 11-20-2012. Bolded tumors are included in Pan-Cancer Whitelist tumors included in this study.

Cancer Type	TCGA Acronym	Tumor	Matched Normal	Unmatched Normal
Acute Myeloid Leukemia	LAML	187	0	0
Bladder Urothelial	BLCA	124	16	0
Brain Low Grade Glioma	LGG	180	0	0
Brease Invasive Carcinoma	BRCA	886	98	2
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	CESC	122	2	0
Colon Adenocarcinoma	COAD	411	0	0
Head and Neck Squamous Cell Carcinoma	HNSC	309	38	0
Kidney Renal Clear Cell	KIRC	480	67	4
Kidney Renal Papillary Cell	KIRP	103	25	1
Liver Hepatocellular Carcinoma	LIHC	96	49	1
Lung Adenocarcinoma	LUAD	366	39	6
Lung Squamous Cell Carcinoma	LUSC	332	35	0
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	16	0	0
Ovarian Serous Cystadenocarcinoma	OV	462	0	0
Pancreatic Adenocarcinoma	PAAD	34	0	0
Prostate Adenocarcinoma	PRAD	174	50	0
Rectal Adenocarcinoma	READ	144	0	0
Sarcoma	SARC	30	0	0
Skin Cutaneous Melanoma	SKCM	241	0	0
Stomach Adenocarcinoma	STAD	151	24	0
Thyroid Carcinoma	THCA	352	43	1
Uterine Corpus Endometrioid Carcinoma	UCEC	485	17	12
Totals		5685	503	27

Supplementary Table S2. Percent expression of the top 30 TCGA defined microRNAs in AGO-CLIP datasets.

miRNA_ID	BC1	BC3	EF3D-AGO2	LCL-BAC	LCL-35	SDLCL	hESC	293T
hsa-mir-143	0.00000	0.00000	0.00012	0.00000	0.00000	0.00001	0.00030	0.00249
hsa-mir-10b	0.00000	#N/A	#N/A	#N/A	0.00001	0.00001	0.00000	0.00543
hsa-mir-21	0.07409	0.08265	0.12911	0.09119	0.25568	0.09334	0.09277	0.00831
hsa-mir-22	0.00002	0.00019	0.00209	0.00761	0.00095	0.00046	0.00285	0.00256
hsa-mir-30a	0.00003	0.00001	0.00004	0.00000	0.00022	0.00004	0.00156	0.00285
hsa-mir-10a	0.00004	0.00001	0.00015	0.00013	0.00091	0.00027	0.00000	0.00882
hsa-mir-99b	0.00037	#N/A	#N/A	#N/A	0.00000	0.00002	0.00040	0.00078
hsa-mir-30a*	0.00003	0.00001	0.00000	0.00000	0.00011	0.00001	0.00020	0.00285
hsa-mir-148a	0.00314	0.00096	0.01473	0.01241	0.01534	0.00520	0.00672	0.00468
hsa-mir-203	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	0.00005	0.00268
hsa-let-7b	#N/A	#N/A	0.00440	0.00221	0.00399	0.00056	0.00000	0.00895
hsa-mir-101-1	0.00243	0.00059	0.00169	0.00223	0.00427	0.00202	0.00758	0.00475
hsa-let-7a-2	#N/A	#N/A	0.05539	0.02518	0.02751	0.01960	0.00070	0.00603
hsa-mir-100	0.00000	#N/A	#N/A	#N/A	#N/A	#N/A	0.00010	0.00167
hsa-mir-103-1	0.01641	0.00144	0.03708	0.02917	0.02623	0.05206	0.01327	0.00498
hsa-let-7f-2	#N/A	#N/A	0.05320	0.02002	0.02590	0.01744	0.00030	0.00520
hsa-mir-30d	0.00737	0.00089	0.00270	0.00249	0.00395	0.00206	0.00135	0.00158
hsa-mir-30e*	0.00354	0.00058	0.00090	0.00111	0.00172	0.00145	0.00070	0.00148
hsa-mir-92a-2	0.02389	0.00937	0.00704	0.01380	0.01408	0.01247	0.00827	0.03653
hsa-mir-29a	0.00115	0.00303	0.01082	0.01848	0.01284	0.01058	0.00114	0.00212
hsa-let-7a-3	#N/A	#N/A	0.05539	0.02518	0.02751	0.01960	0.00070	#N/A
hsa-let-7a-1	#N/A	#N/A	0.05539	0.02518	0.02751	0.01960	0.00070	0.00603
hsa-mir-375	0.00000	#N/A	#N/A	#N/A	#N/A	#N/A	0.00000	0.00059
hsa-mir-25	0.00309	0.00171	0.00137	0.00266	0.00232	0.00219	0.00735	0.00778
hsa-mir-30e	0.00354	0.00058	0.00090	0.00111	0.00172	0.00145	0.00534	0.00148
hsa-mir-182	0.00265	0.00031	0.00004	0.00005	0.00005	0.00003	0.00083	0.00152
hsa-let-7c	#N/A	#N/A	0.00023	0.00014	0.00046	0.00018	0.00005	0.00178
hsa-mir-200c	0.00001	0.00002	0.00019	0.00023	0.00009	0.00012	0.00480	0.00077
hsa-mir-28	0.00000	0.00003	#N/A	#N/A	#N/A	#N/A	0.00050	0.00067
hsa-mir-145	#N/A	#N/A	0.00002	0.00000	0.00000	0.00000	0.00005	0.00025

Supplementary Table S3. AGO-CLIP and TargetScan miRNA-target prediction methods produce two different spectrums of target prediction for the Pan-Cancer microRNAs defined in this study.

Values for Pan-Cancer oncomiRs and miR-suppressors	Percentage
% AGO-CLIP Targets >2	31.65%
% TargetScan targets without AGO-CLIP peak	74.39%
% TargetScan targets with AGO-CLIP peak	25.61%
% AGO-CLIP targets Called by TargetScan	31.47%
% AGO-CLIP targets not called by TargetScan	68.53%
% AGO-CLIP targets outside 3'UTR (CDS, 5'UTR, ncRNA)	34.61%
% AGO-CLIP 3'UTRs called by Target Scan	48.27%
% of AGO-CLIP 3'UTR targets not called by TargetScan	51.73%

Supplementary Table S4. Synapse identification codes associated with this study.

File	Synapse ID
TCGA Pan-Cancer Project	syn300013
Complete AGO-CLIP Atlas Data (clusters and seeds)	syn1720733
miSNP Output (Individual tumors and union seed-snv calls)	syn1720734
GISTIC Scores	syn1703335
MuSiC Scores	syn1729383
MSKCC Driver Genes	syn1720070
Complete miRNA-seq Processed files	syn1856597