Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif

Dongfang Wang

November 29, 2013

nature COMMUNICATIONS

ARTICLE

Received 9 Jul 2013 | Accepted 9 Oct 2013 | Published 13 Nov 2013

DOI: 10.1038/ncomms3730 OPEN

Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif

Mark P. Hamilton[1], Kimal Rajapakshe[1], Sean M. Hartig[1], Boris Reva[2], Michael D. McLellan[3], Cyriac Kandoth[3], Li Ding[3,4,5], Travis I. Zack[6], Preethi H. Gunaratne[7,8], David A. Wheeler[8], Cristian Coarfa[1] & Sean E. McGuire[1,9]

[1] Department of Molecular and Cellular Biology, Baylor College of Medicine, 1 Baylor Plaza Houston M822, Houston, Texas 77030, USA. [2] Computational Biology Center, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [3] The Genome Institute, Washington University, St Louis, Missouri 63108, USA. [4] Department of Genetics, Washington University, St Louis, Missouri 63110, USA. [5] Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. [6] The Eli and Edythe L Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. [7] Department of Biology and Biochemistry, University of Houston, 4800 Calhoun, Houston 77204, Texas, USA. [8] The Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [9] Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. Correspondence and requests for materials should be addressed to S.E.M. (email: sean.mcguire@bcm.edu).

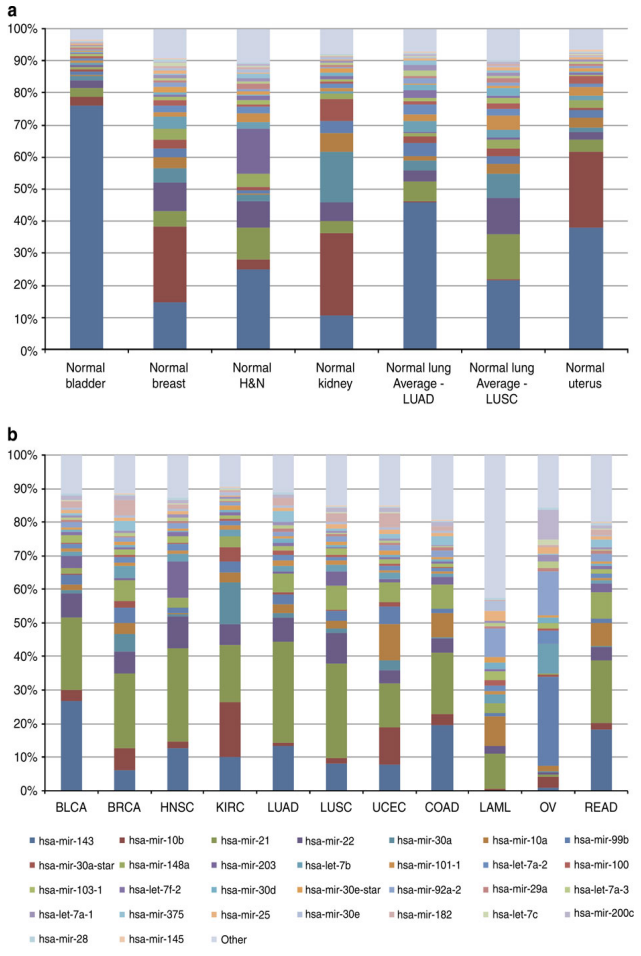
Introduction

1. **a core set of miRNAs:** exist in the overlapping oncogenic pathways of many tumor types.
2. TCGA pan-cancer data set and AGO-CLIP data
3. identify pan-cancer oncogenic cotargeting of miRNA "superfamily".

Outline

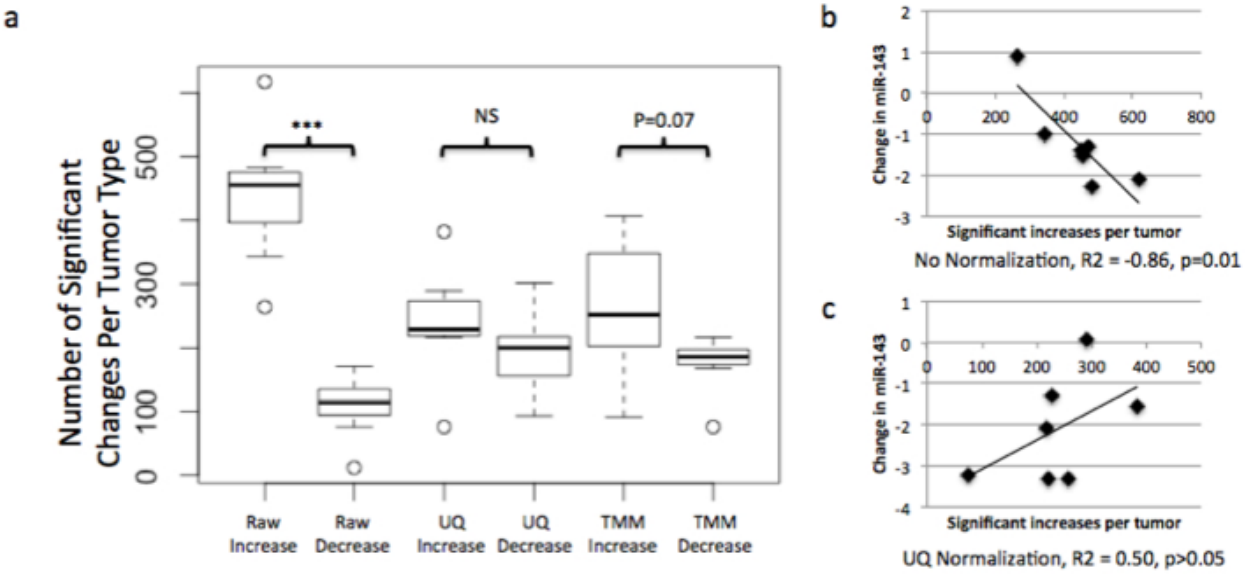
Outline

TCGA Global miRNA Expression Patterns



- top 30% constitute 90% expression across heterogeneous tumor types
- **miR-143, miR-21**

Analysis of TCGA Data



- more increased miRNAs than decreased miRNAs due to **loss of highly expressed miRNAs in tumors, especially miR-143**
- Normalize the data set : **Upper quantile and trimmed median of M-values**

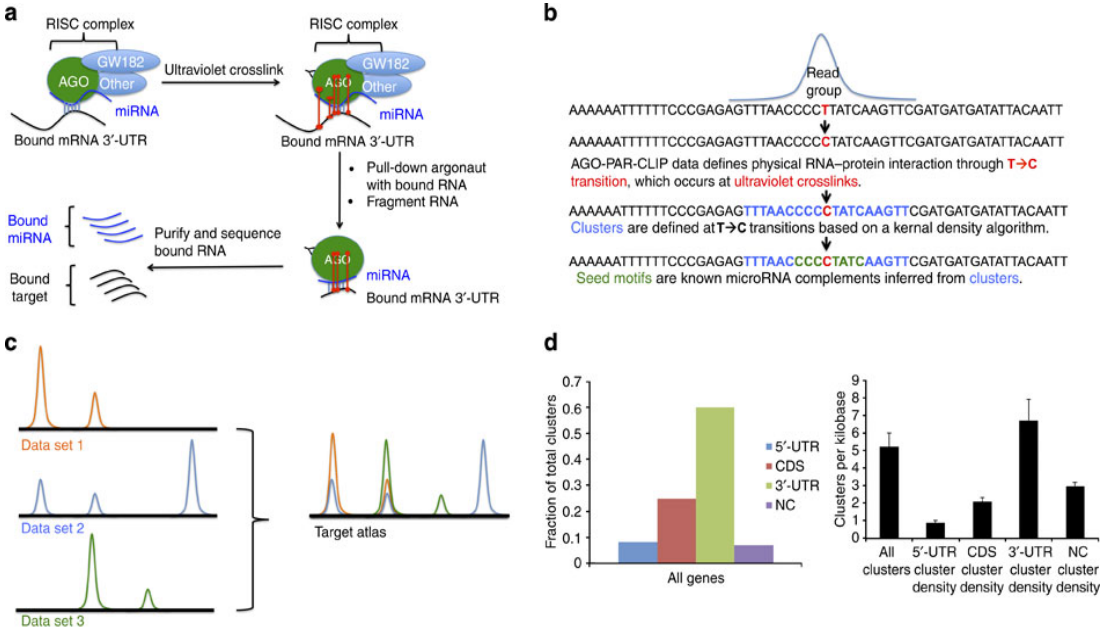
AGO-CLIP

1. Facilitated by Agronaute proteins, miRNAs bind target mRNAs in the **RNA-induced silencing complex**
2. **AGO-CLIP**: Argonaute Crosslinking Immunoprecipitation data sets experimentally identify miRNA-target interactions in a genome-wide manner through purification of Argonaute-protein-associated RNAs, which include bound miRNAs and their respective targets.

AGO-CLIP Atlas

AGO-CLIP technology

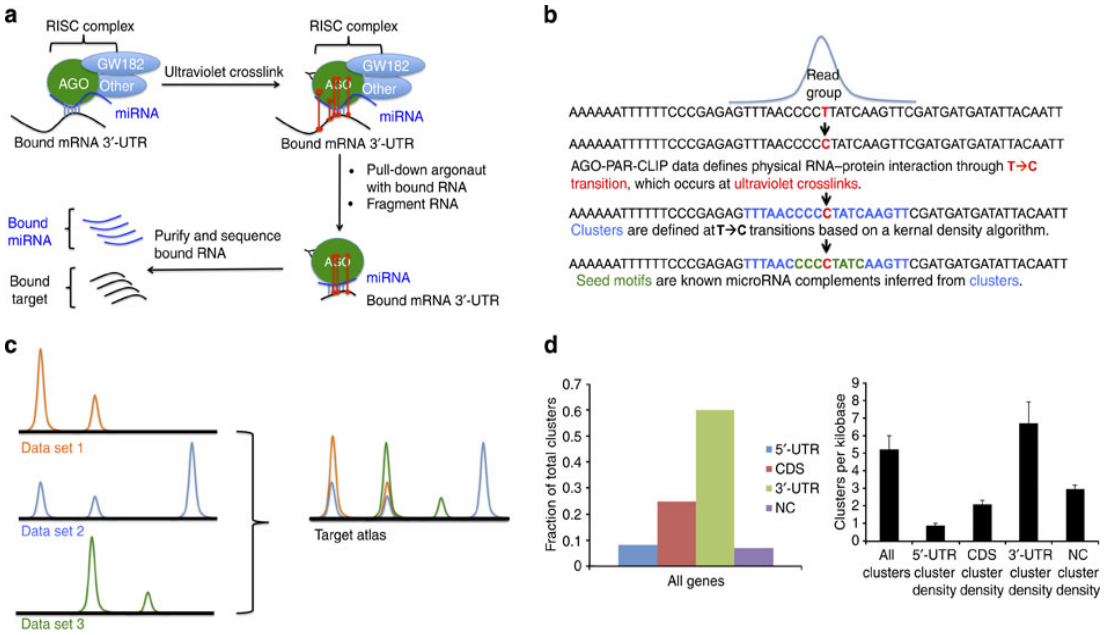
- ultraviolet crosslinking of RNA to protein
- immunoprecipitation to determine RNA species bound to the Argonaute protein
- CLIP is limited by low efficiency of the UV crosslinking step, since non-crosslinked RNA molecules are more readily reverse transcribed.



AGO-CLIP Atlas

AGO Photoactivatable-Ribonucleoside-Enhanced CLIP

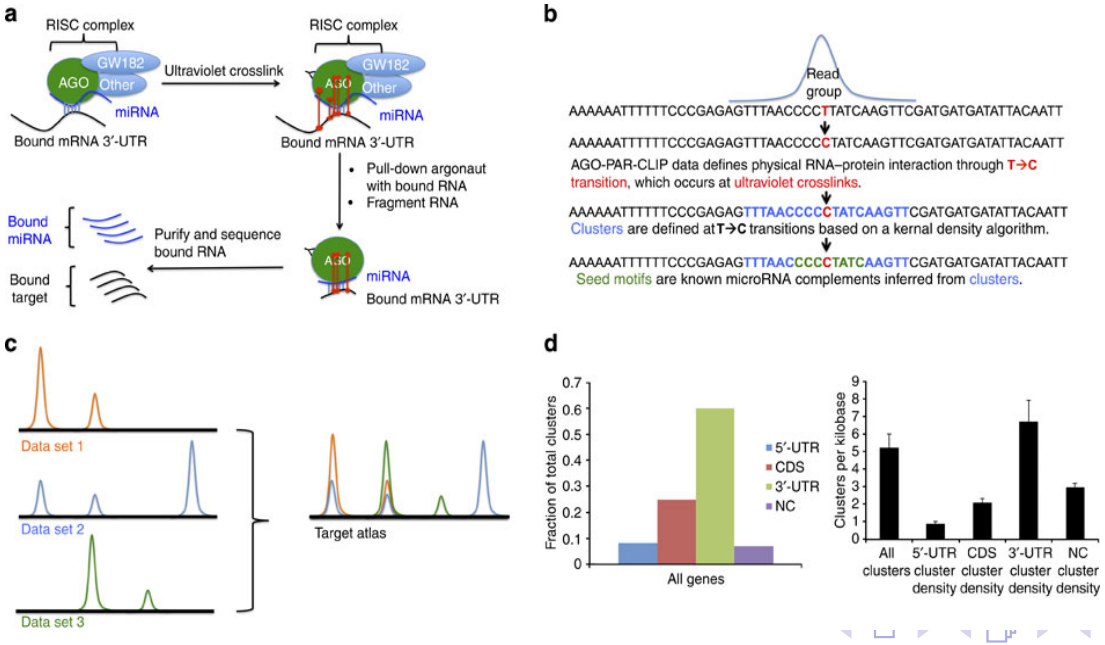
- nucleoside analogoues such as 4-thiouridine are induced before crosslinking
- **T-C transitions** during the reverse-transcription step of the AGO-CLIP experiment



AGO-CLIP Atlas

Seed recurrent across multiple data sets

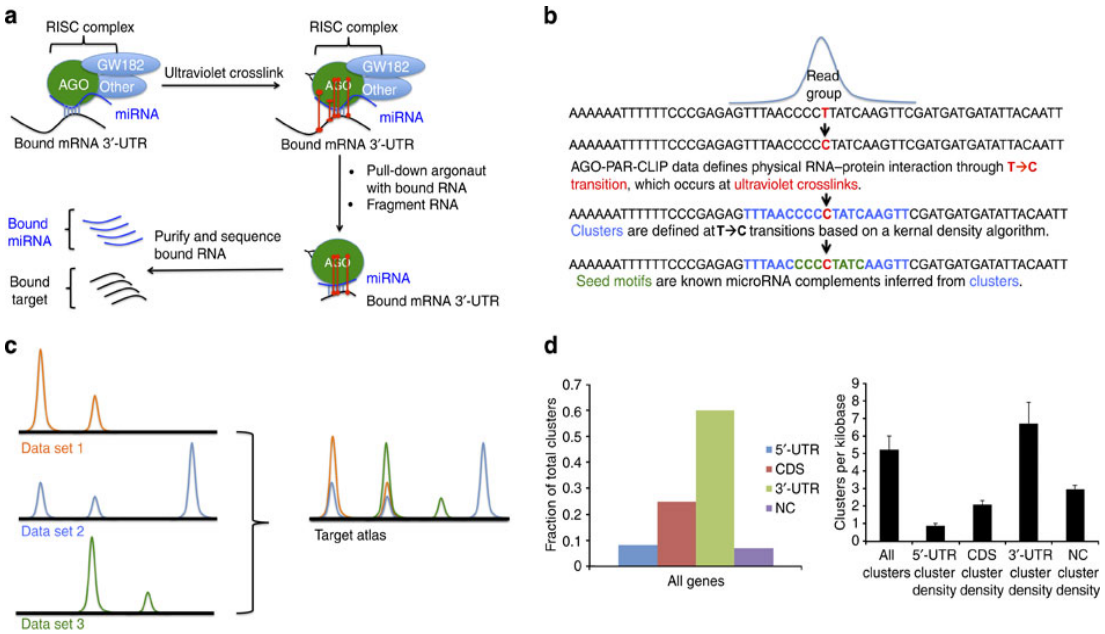
- 14 AGO-CLIP data sets; 124,000 target clusters; 300,000 putative seed motifs
- ≥ 3 occurrences of an AGO-CLIP peak on a given sites corresponding to a s significant event relative to a random distribution of clusters



AGO-CLIP Atlas

Cluster location by mRNA region

- 60% mappint to 3'UTR; 24.7% coding region; 8.2% 5'UTR; 7% ncRNAs
- unbiased platform; **MDM2,long ncRNA Xst**



TargetScan vs AGO-CLIP

Different spectrums of target prediction

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%

Outline

Define pan-cancer oncomiRs and miR suppressors

a

miRNA_ID	BLCA	BRCA	HNSC	KIRC	LUAD	LUSC	UCEC
hsa-mir-183	2.55	2.77	1.48		2.39	2.36	4.83
hsa-mir-182	1.88	2.03	1.02	0.47	2.03	1.61	4.03
hsa-mir-96	2.67	2.93	1.18	0.14	2.43	2.28	4.09
hsa-mir-210	3.90	3.27	2.36	3.41	3.51	4.17	3.09
hsa-mir-425	1.45	0.95	0.86	0.36	0.52	0.35	2.84
hsa-mir-130b	2.64	1.69	1.42	1.35	1.77	1.79	2.08
hsa-mir-18a	2.67	1.42	1.29	1.25		0.46	2.38
hsa-mir-93	1.34	0.90	1.17	1.31	0.10	0.60	1.77
hsa-mir-17	1.60	1.14	1.02	0.72	0.43	0.88	1.54
hsa-mir-106a	1.55	1.21	1.44	0.24	0.93	1.55	2.71
hsa-mir-135b	2.60	2.49	1.30	0.13	2.80	1.82	2.66
hsa-mir-301b	3.89	3.22	2.79	1.67	2.80	3.10	2.70
hsa-mir-192	2.08	1.50	0.53	0.24	2.63	0.64	1.95
hsa-mir-142	1.06	1.98	1.16	2.13	0.99	1.13	1.37
hsa-mir-301a	1.46	1.94	1.96	0.84	1.26	1.44	1.38
hsa-mir-19a	1.52	1.23	0.62	1.18	0.39	1.03	1.00
hsa-mir-33b	3.33	2.95	1.05	1.84	2.55	2.68	1.29
hsa-mir-590	1.44	1.27	0.47	1.32	0.86	1.20	0.71
hsa-mir-196a-1	2.23	3.17	4.32	0.66	5.87	5.26	2.41
hsa-mir-7-1	2.50	2.87	2.46	1.32	1.89	1.53	3.19
hsa-mir-21	1.20	1.94	1.18	2.35	1.01	0.38	0.69
hsa-mir-455	1.32	1.20	2.47	0.70	0.10	0.67	−0.65
hsa-mir-139	−3.06	−3.24	−1.97	−0.46	−3.18	−4.15	−3.20
hsa-mir-101-1	−1.11	−0.53	−1.86	−0.91	−1.75	−2.37	−2.67
hsa-mir-140	−1.39	−1.38	−1.20	−0.76	−1.86	−2.50	−2.22
hsa-mir-143	−3.21	−1.54	−1.30		−2.07	−3.31	−3.32
hsa-mir-145	−2.84	−2.36	−0.44	−0.18	−1.93	−2.21	−3.09
hsa-mir-27b	−1.00	−0.10	−0.91	−0.39	−0.41	−0.88	−1.97
hsa-mir-100	−2.75	−2.14	−1.63	0.36	−0.60	−1.86	−3.30
hsa-mir-99a	−2.77	−2.17	−1.77	−0.36	−1.53	−1.30	−2.89
hsa-mir-26a-2	−1.19	−0.59	−1.23	−0.19	−0.90	−2.02	−1.63
hsa-mir-1-2	−4.08	−5.50	−2.99	−1.34	−2.02	−3.28	−3.79
hsa-mir-133a-2	−3.93	−6.41	−2.63	−0.94	−2.72	−3.54	−3.71
hsa-let-7c	−2.68	−1.89	−1.71	−0.68	−1.86	−2.09	−2.20
hsa-mir-125b-1	−2.44	−1.94	−1.00	−0.83	−0.99	−1.42	−2.32
hsa-mir-29a	−1.32	−0.83	−1.33	−0.25	−0.40	−1.88	−1.89
hsa-mir-195	−2.09	−1.42	−1.72	0.67	−1.73	−2.53	−2.22
hsa-mir-10b	−1.54	−2.06	−0.94	−0.63	−0.36		−2.16
hsa-mir-101-2	−1.68	−1.40	−1.36	0.23	−2.00	−2.76	−2.61
hsa-mir-125a	−1.02	−0.61	−0.13	−0.70	−1.64	−2.35	−1.35
hsa-mir-204	−0.55	−2.72	−1.07	−1.35	−0.83	1.33	−2.82
hsa-let-7a-2	−0.66	−0.41	−0.44		−1.35	−2.50	−0.88
hsa-let-7a-1	−0.66	−0.41	−0.44		−1.35	−2.49	−0.88
hsa-let-7a-3	−0.66	−0.41	−0.44		−1.35	−2.49	−0.87
hsa-mir-26b	−0.56	−0.33	−0.52	−0.11	−0.17	−0.66	−0.66
hsa-let-7b	−0.67	−0.42	−0.30	1.15	−1.44	−1.97	−0.75
hsa-mir-451	−1.14	−3.82	−1.15	1.66	−3.27	−3.78	−0.69

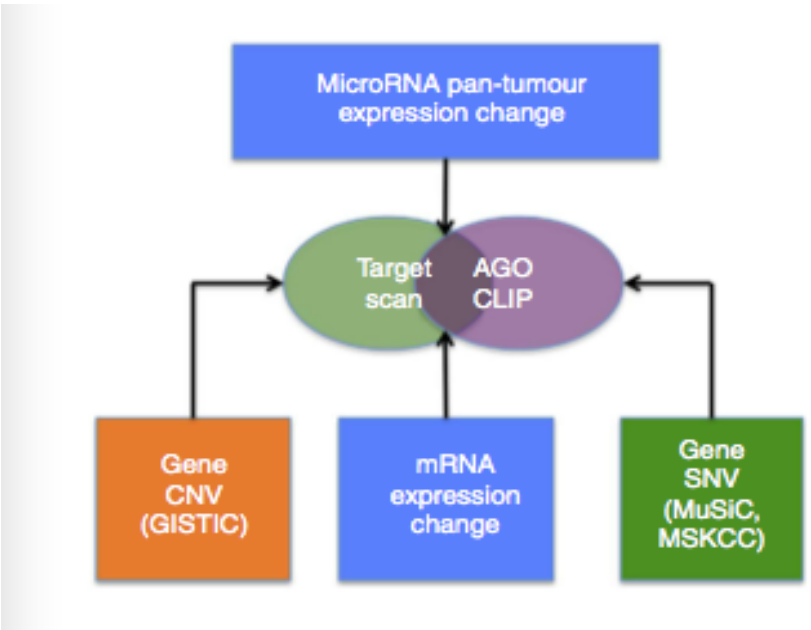
Log2 expression change per tumour type
−2 0 2

consistent expression changes across cancer types

- **Pan-cancer oncomiRs**: significant expression gain ($q < 0.05$, Fisher's exact test) in at least six out of seven pan-cancer tumor types
- **Pan-cancer miR suppressors**: significant expression loss in at least six out of seven tumor types
- focused on highly expressed miRNAs that had many conserverd target sites in the 3'UTR
  (1) 87 broadly conserverd miRNA families
  (2)Argonaute-bound groups in at least 3 out of 14 AGO-CLIP data sets

Nominate Cancer Genes

Using available pan-cancer data: SNV scores, CNV analysis, mRNA expression changes



TS and OC Definition

Utilize three external data sets generated for the purpose of pan-cancer analysis by the TCGA:

1. **MuSiC:**
   derive p-values to determine significantly mutated genes versus the background mutation rate
2. **MSKCC driver target analysis:**
   Create a binary definition of SNVs to stratify mutated genes as either OCs or TSs based on a functional impact score that weighs the probable impact of mutation at a specific amino acid residue
3. **GISTIC:**
   define CNV log ratios

TS and OC definition

Final Score

(mRNA increases)-(mRNA decrease)
+(0.5)(CNV amplification)-(0.5)CNV deletion
+(100)(mutation frequency across all tumors)
(± 1 MSKCC drivers)-(5)(truncation frequency)

1. **mRNA data:**
   - $+1$: given for a gene in each of seven tumors in which there was a significant increase(Student's t test, $q < 0.005$)
   - -1: significant decrease
2. **CNV data: GISTIC score**
   - set a threshold $0.3/-0.3$ for amplification/deletion
   - 0.5 : achieve amplification in 30% of samlpes for some tumor
3. **Mutation score:**
   - only considered based on MuSiC-determined Fisher's combine p-test FDR $q < 0.005$
   - MSKCC driver analysis: TS (-1) and OS (1)
   - truncating mutations (nonsense mutation)

Determine miRNA-target interactions

Four methods tested:

1. use all AGO-CLIP-defined sites without considering site conservation(TargetScan)
2. use only AGO-CLIP-defined sites with ≥ 3 occurences without considering TargetScan
3. use only TargetScan binding sites without considering AGO-CLIP data
4. combining:
   (1) AGO-CLIP-defined target sites with ≥ 3 occurences
   (2) AGO-CLIP-defined ≥ 1 occurences and a TargetScan call

MiRNA-target Enrichment Calculations

Intersect pan-cancer oncomiRs and miR suppressors with pan-cancer TSs versus pan-cancer OCs

Method

1. For pan-cancer oncomiRs and tumor suppressors respectively
2. For n number of OC or TS ranked in the top 3,000, $n = 1 \rightarrow 3,000$

x(TS targets per miRNA/total targets per miRNA)
versus (Students t-test)
x(OC targets per miRNA/total targets per miRNA)

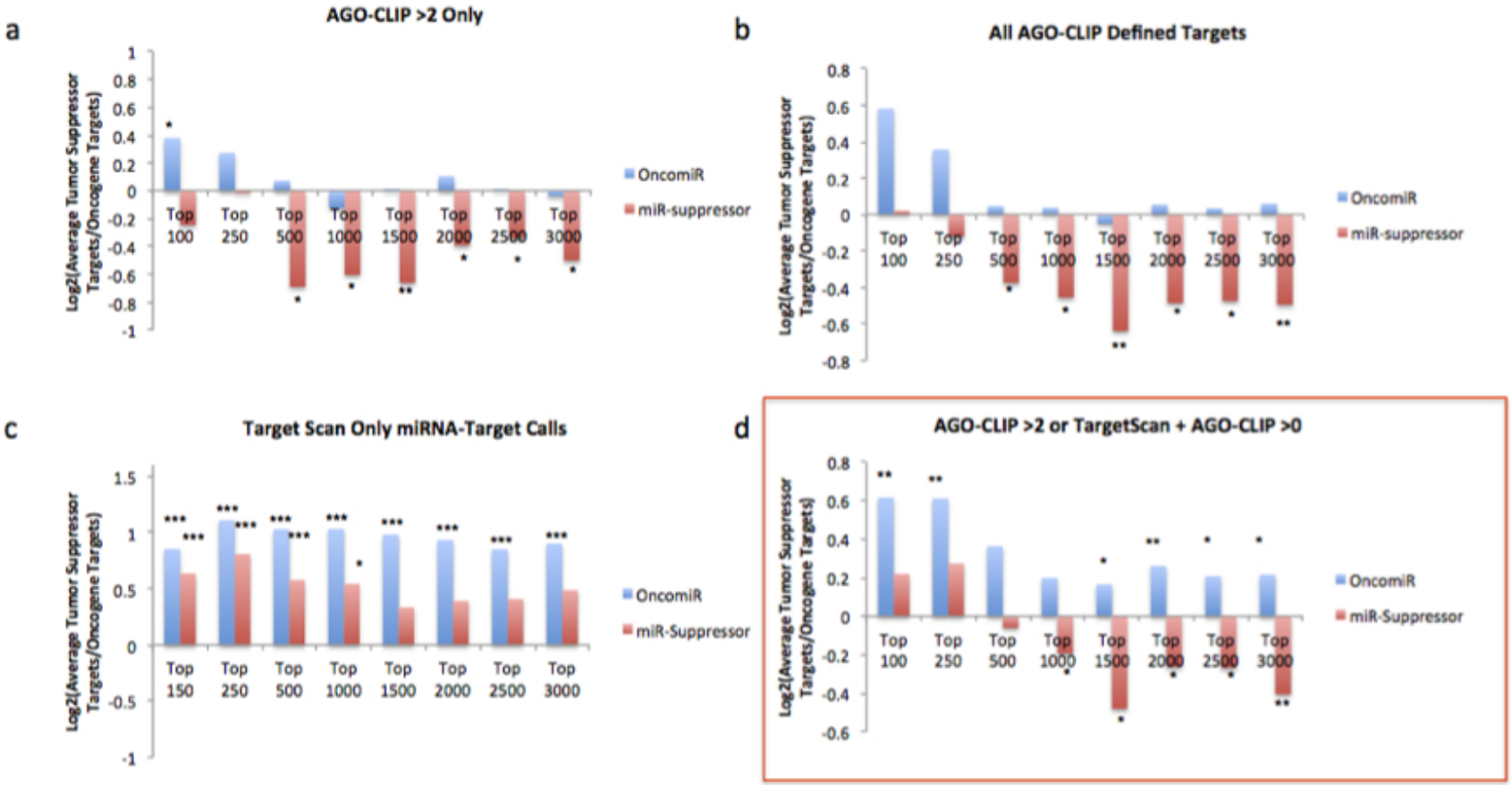


Figure : TS target versus oncomiR target enrichments for pan-cancer oncomirs and mir suppressors across the top 100-3000(10%) TS and OCs

- **AGO-CLIP alone: bias towards enrichment of OCs**

  top 3,000 OCs have 31% more AGO-CLIP clusters binding them than the top 3,000 TS.

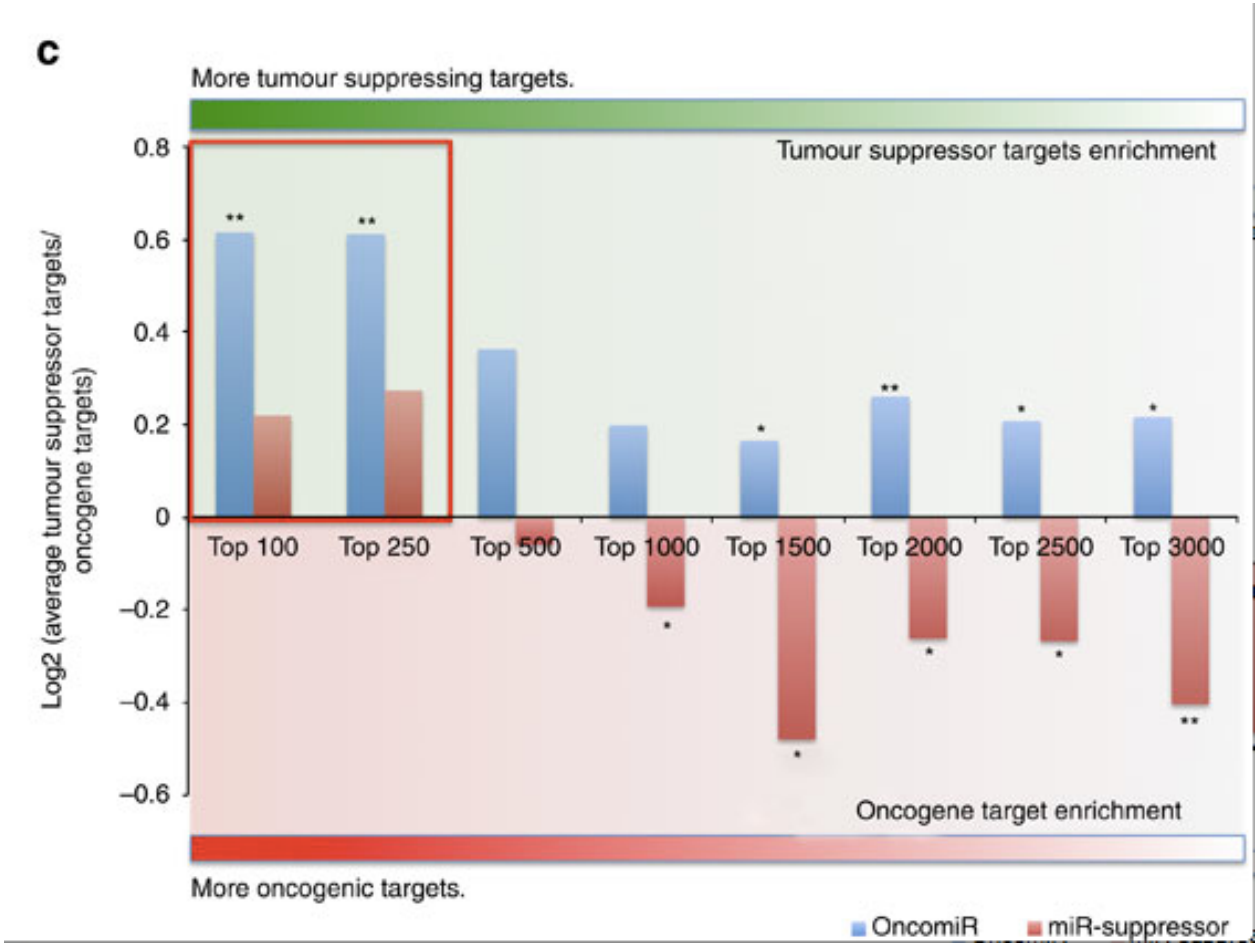
  overexpression of OCs in the cell lines used to perform the AGO-CLIP analyses,or greater miRNA binding of OCs in general

- **TargetScan alone: bias towards TS**

  top 3,000 TS have 40%larger 3'UTR lengths than the top 3,000 OCs

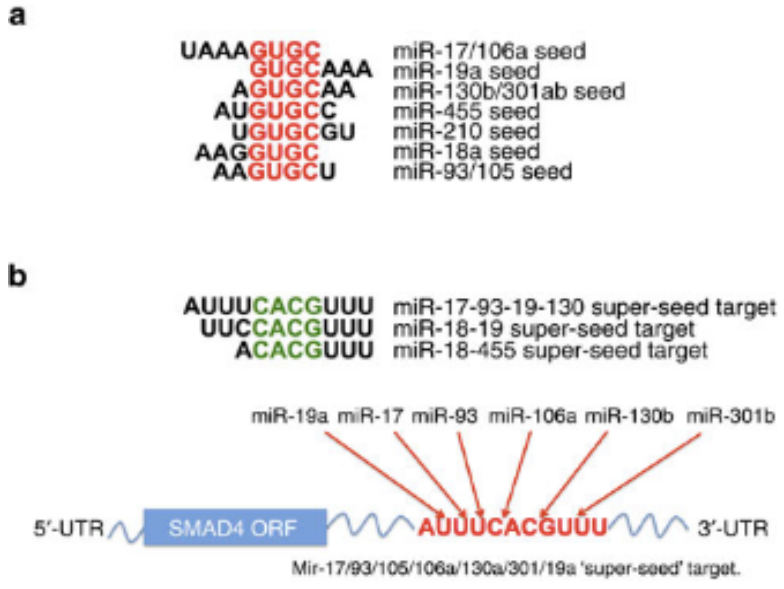
  relative size of 3'UTR directly determines the total number of predicted miRNA target sites associated with the 3'UTR

- **Ultimately, determining miRNA-target interaction using the final methods.**



Pan-caner oncomiRs are enriched for TS targets, and pan-cancer miR suppressors are enriched for oncogenic targets

Pan-cancer OncomiR Network



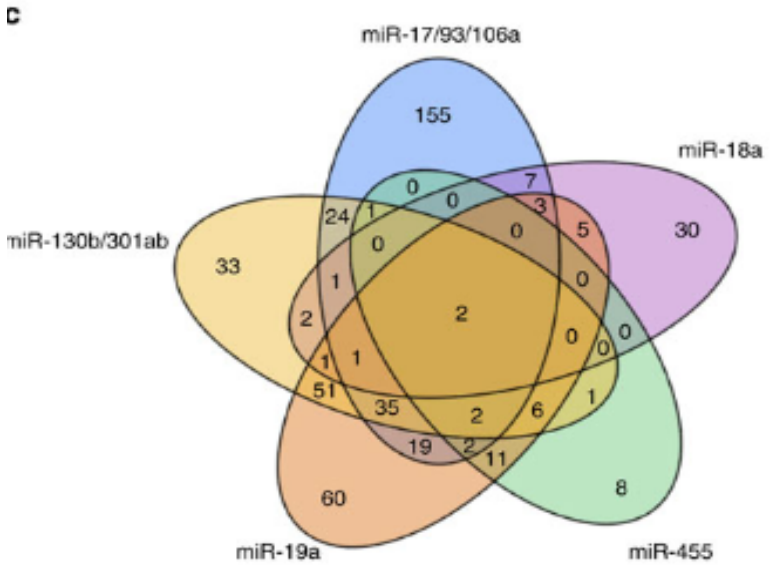
Seed similarity

- **GUGC motif 'superfamily'**: 10 of 22 pan-cancer oncomiRs 36 of 187 from broadly conserved seed families
- None of 25 pan-cancer miR suppressor contains a GUGC motif.

Hypothesis

pan-cancer oncomiRs may undergo coordinate regulation to mutually cotarget and suppress critical TS.

The majority of targets predicted in the top 3,000 TS are shared with at least one other superfamily member.



miR-17,-19,-130,-210,-18 superfamily

Coregulation of miR-17,19,130 families

Pan-cancer Correlations

1. strong positive miRNA-miRNA correlation(Pearson)
2. strong negative coorelation of the superfamily with many pan-cancer TS, $(PTEN, ZBTB4, TGFBR2)$

d

Tumour type	miRNA_ID	hsa-mir-106a	hsa-mir-17	hsa-mir-19a	hsa-mir-301a	hsa-mir-130b	hsa-mir-301b	hsa-mir-93
BLCA	hsa-mir-106a	1	0.42	0.476	0.183	0.314	0.076	0.629
	hsa-mir-17	0.42	1	0.792	0.496	0.352	0.332	0.623
	hsa-mir-19a	0.476	0.792	1	0.619	0.666	0.588	0.625
	hsa-mir-301a	0.183	0.496	0.619	1	0.458	0.744	0.561
	hsa-mir-130b	0.314	0.352	0.666	0.458	1	0.826	0.466
	hsa-mir-301b	0.076	0.332	0.588	0.744	0.826	1	0.446
	hsa-mir-93	0.629	0.623	0.625	0.561	0.466	0.446	1
BRCA	hsa-mir-106a	1	0.725	0.796	0.431	0.564	0.601	0.576
	hsa-mir-17	0.725	1	0.886	0.431	0.481	0.533	0.679
	hsa-mir-19a	0.796	0.886	1	0.419	0.549	0.581	0.594
	hsa-mir-301a	0.431	0.431	0.419	1	0.406	0.53	0.448
	hsa-mir-130b	0.564	0.481	0.549	0.406	1	0.755	0.451
	hsa-mir-301b	0.601	0.533	0.581	0.53	0.755	1	0.513
	hsa-mir-93	0.576	0.679	0.594	0.448	0.451	0.513	1
COAD	hsa-mir-106a	1	0.306	0.105	0.129	0.322	0.132	
	hsa-mir-17	0.306	1	0.408	0.333	0.419	0.343	0.527
	hsa-mir-19a	0.105	0.408	1	0.15	0.303	0.188	0.159
	hsa-mir-301a	0.129	0.333	0.15	1	0.489	0.532	0.319
	hsa-mir-130b	0.322	0.419	0.303	0.489	1	0.638	0.354
	hsa-mir-301b	0.132	0.343	0.188	0.532	0.638	1	0.376
	hsa-mir-93		0.527	0.159	0.319	0.354	0.376	1
	hsa-mir-106a	1	0.246	0.308	0.583	0.435	0.588	0.183
	hsa-mir-17	0.246	1	0.851	0.569	0.413	0.324	0.618

a

Gene_symbol	Tumor suppressor rank	Tumor type	hsa-mir-17	hsa-mir-106a	hsa-mir-19a	hsa-mir-130b	hsa-mir-301a	hsa-mir-301b	hsa-mir-93
PTEN	3	BLCA	−0.28	−0.156	−0.342	−0.311	−0.297	−0.324	−0.214
		BRCA	−0.25	−0.272	−0.234	−0.176	−0.191	−0.237	−0.286
		COAD	−0.033	0.137	0.074	0.109	0.077	0.045	−0.154
		HNSC	−0.211	−0.031	−0.134	−0.182	−0.07		−0.22
		KIRC	−0.055		−0.04	−0.132		−0.052	0.04
		LAML	−0.172	0.132	−0.178	−0.092		−0.041	0.322
		LUAD	−0.327	−0.252	−0.282	−0.228	−0.211	−0.242	−0.303
		LUSC	−0.196	−0.14	−0.095	−0.207	−0.097	−0.166	−0.211
		OV	−0.259	−0.107	−0.11	−0.157			−0.179
		READ							
		UCEC							

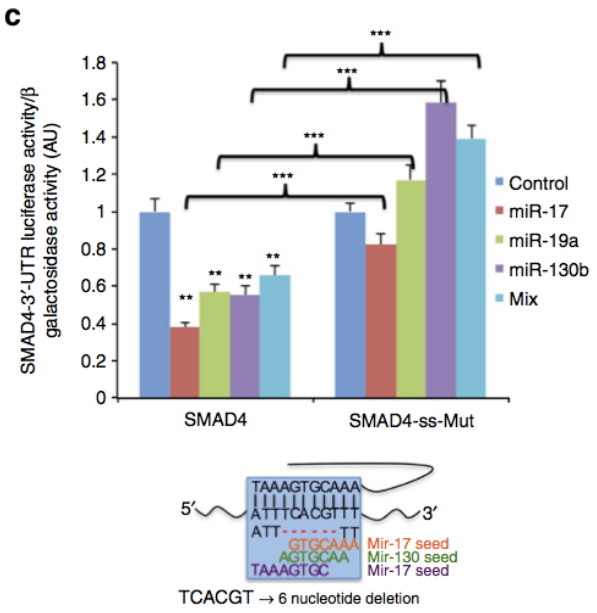
SMAD4 Gene

SMAD4	44	BLCA	0.444	0.337	0.343	0.114	0.371	0.229	−0.398
		BRCA	0.124	0.066	0.088	−0.031	−0.028	−0.009	0.028
		COAD	−0.263	−0.082	−0.034	−0.006	−0.026	−0.021	−0.236
		HNSC	−0.011	0.241	−0.027	0.189	−0.044	0.039	0.16
		KIRC	−0.09	−0.097	0.024	−0.312	−0.163	−0.14	−0.051
		LAML	−0.016	0.254	−0.072	−0.026	−0.024	−0.074	0.051
		LUAD	0.06	0.094	0.101		−0.011		
		LUSC	0.149	0.257	0.129	−0.124	0.094	0.062	0.173
		OV	0.093					0.051	0.098
		READ	−0.242		−0.29	−0.177	0.105	0.043	−0.067
		UCEC	0.149	−0.105	0.087	0.09	0.081	0.141	0.154

1. positively correlate with the superfamily in BLCA, but otherwise shows no significant correlation
2. potentially suggest a role for the miRNAs in translational repression

3'-UTR-luciferase fusions

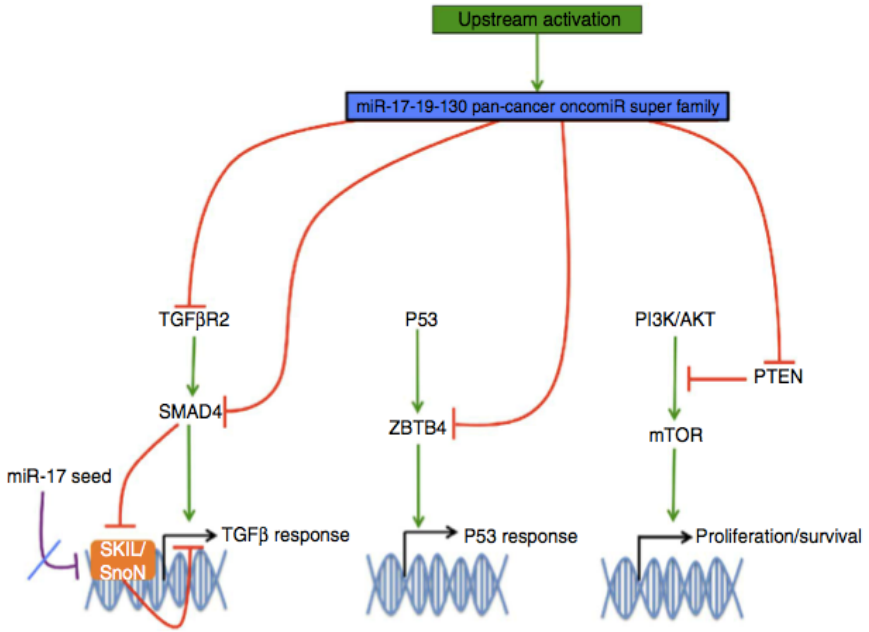
bind to the 3'UTR and significantly repress luciferase activity



- SMAD4: highly conserved and few potential compensatory sites
- delete the central six nucleosites of the SMAD4 super-seed site, strong ablation of each miRNA seed family's ablity to bind and regulate

Coordinately target multiple critical pathways

This study defines these pathway targets as significant across multiple tumor contexts



miR-17 seed family auto regulation through SKIL-repressor targeting.
SKIL 3′-UTR T deletion in miR-17 seed attenuates auto regulation through miR-17 seed family binding disruption.

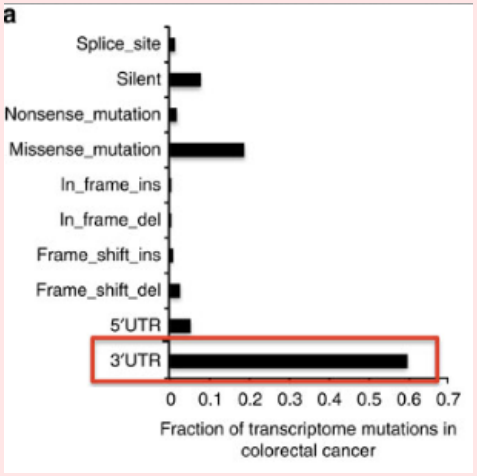
Outline

Aim

Integrate the AGO-CLIP data set with TCGA mutation data to identify somatic SNVs in miRNA target sites across tumors.

mutations outside the coding region

SNV analysis: relevant coding region ⇒ predicted functional impacts of amino acid changes



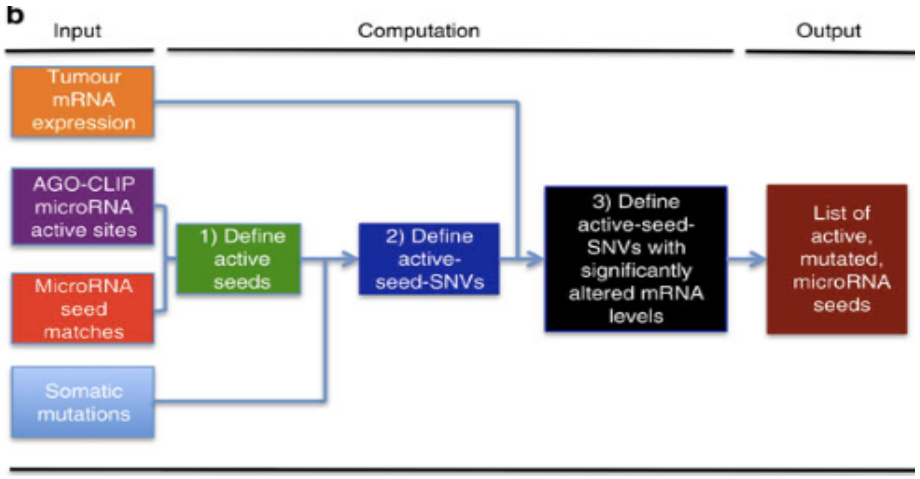
Possible interfere of miRNA binding

$$\begin{cases} \text{mutation in target site} \\ \text{target site complementary to the miRNA's seed} \end{cases}$$

$\Rightarrow$ such mutation attenuate miRNA's control

expand the search for relevant cancer mutations by imbuing silent mutations and 3'-UTR mutations with functional significance.

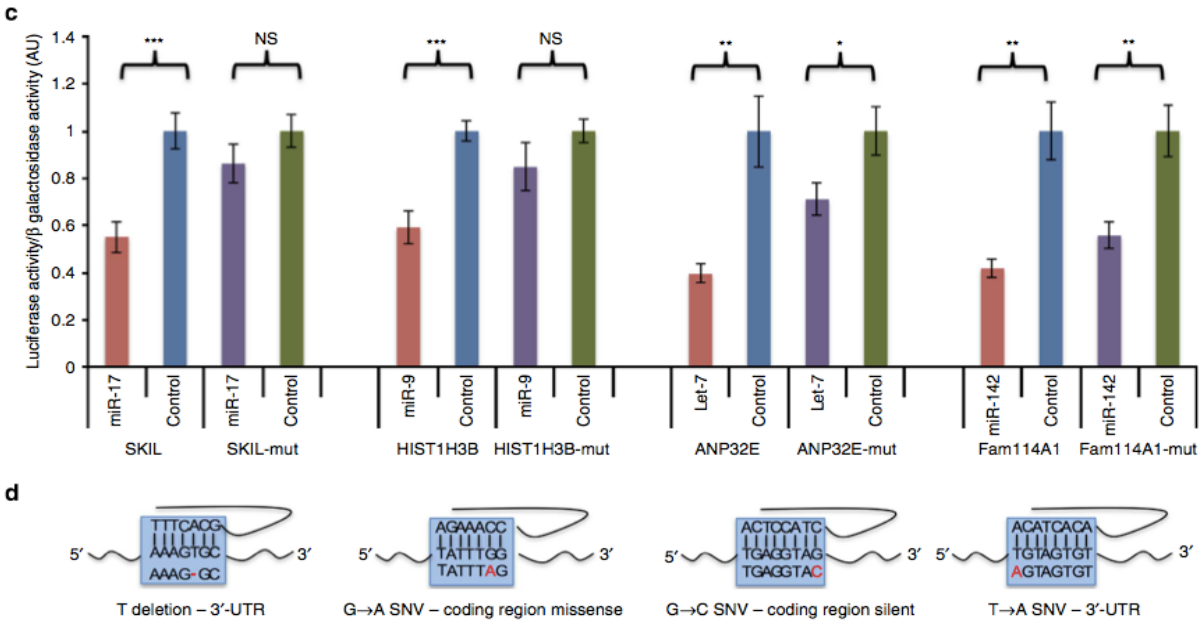
miSNP algorithm



two types of integration

1. mutation data and AGO-CLIP miRNA/mRNA binding sites
   reports at gene level both the miRNA binding sites and the mutations miRNAs targeting the genes;
2. RNA-seq gene expression data
   carry out a quantitative analysis of the effects of miRNAs and mutations on gene expression

report genes for which the expression associates with the mutation status in miRNA binding sites by comparing the gene expression with or without common sites.


c
Luciferase activity/β galactosidase activity (AU)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
***
NS
***
NS
**
*
**
**
miR-17
Control
miR-17
Control
miR-9
Control
miR-9
Control
Let-7
Control
Let-7
Control
miR-142
Control
miR-142
Control
SKIL
SKIL-mut
HIST1H3B
HIST1H3B-mut
ANP32E
ANP32E-mut
Fam114A1
Fam114A1-mut
d
5′
TTTCACG
AAAGTGC
AAAG-GC
3′
T deletion – 3′-UTR
5′
AGAAACC
TATTTGG
TATTTAG
3′
G→A SNV – coding region missense
5′
ACTCCATC
TGAGGTAG
TGAGGTAC
3′
G→C SNV – coding region silent
5′
ACATCACA
TGTAGTGT
AGTAGTGT
3′
T→A SNV – 3′-UTR

Discussion

1. **A Novel Resource:** integrate AGO-CLIP altas (experimentally defined miRNA binding sites) with TCGA tumor data
2. **A New Framework to Understand MiRNA Regulation of Cancer:**
   - Define a pan-cancer oncomiR superfamily
   - Determine accurate genome-wid miRNA-target interactions
   - Identify miRNA-binding sites mutations