PROGTar: A database of Prognostically Inversely Correlated miRNAs and Genes (PICs) in multiple cancers

Chirayu Pankaj Goswami

Molecular and Genomic Pathology, Thomas Jefferson University Hospital, Philadelphia, USA

ABSTRACT

PROGTar is a database of Prognostically Inversely Correlated miRNA-mRNA pairs (PIC's) in 23 cancer types. Partner miRNA and mRNA in a PIC show inverse correlation of expression and opposite hazards. We analyzed miRNA and mRNA expression data downloaded from The Cancer Genome Atlas (TCGA) in a 3 step approach to identify PICs in different cancer types. In first step we performed correlation analysis between miRNAs and mRNAs for each cancer type. This was followed by performing hazard analysis separately for miRNAs and mRNAs using expression data and survival related clinical variables. In the third step we merged the correlation and hazard result sets. Resultant miRNA and mRNA pairs were filtered to retain only pairs that had negative correlation between miRNA and mRNA expression and opposite hazards for miRNA and mRNA, at a statistically significant level (p <= 0.05).

Results from our pan cancer analysis are available on the web based application PROGTar. Users can search for miRNA/mRNA of interest on the database to find inversely correlated partners. Users can also create prognostic plots for the PICs of interest. Prognostic plots created with PROGTar show arms for high and low expression of target molecule and its corresponding partner in the PIC, bifurcated at median of expression. The plots also show arms for a combined prognostic signature calculated using expression levels of both partners in the PIC. The application is available freely for non-commercial use at www.xvm145.jefferson.edu/progtar

METHODOLOGY

DATA

We Downloaded miRNA and mRNA sequencing data for 23 cancer types from The Caner Genome Atlas, along with clinical data. For Glioblastoma, array based expression data were downloaded.

Clinical and Expression data were preprocessed separately.

PREPROCESSING OF CLINICAL DATA

For Clinical data, latest follow up information was used. Survival related variables (Time to death, Death event) were retained and other variables were discarded.

Samples with <10 days followup were removed from the list

Duplicate instances of samples were removed by retaining the one with highest precedence according to TCGA guidelines.

PREPROCESSING OF EXPRESSION DATA

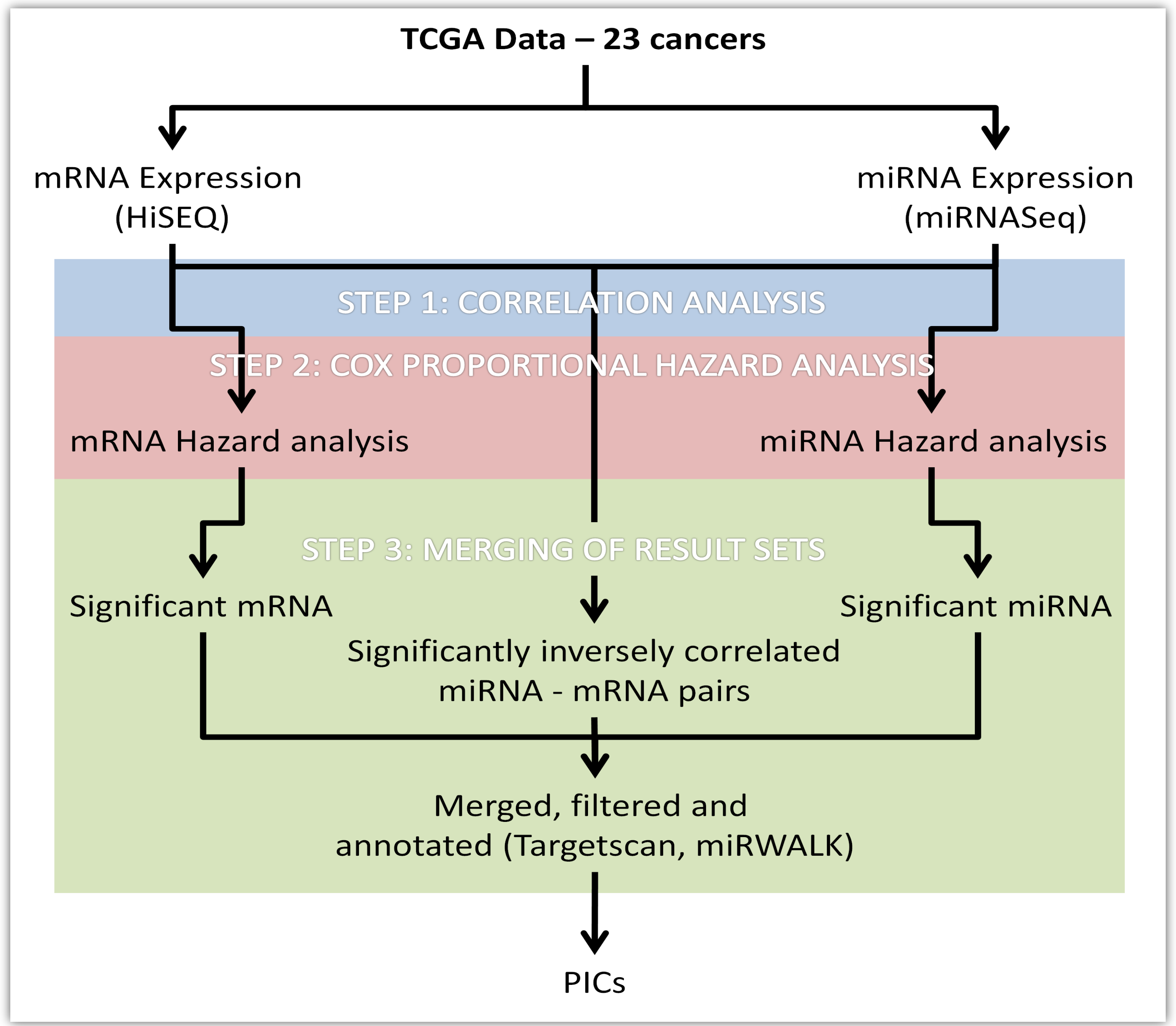
miRNA/mRNA expressed in <20% of samples for each cancer type were removed.

Any duplicate miRNA/mRNA were removed by retaining the instance with highest coefficient of variation. Final data were Log2 transformed.

MERGING OF CLINICAL AND EXPRESSION DATA

Clinical and Expression datasets were then merged to obtain final datasets. Any samples for which either clinical, miRNA expression of mRNA expression data were missing, were omitted from the final analysis.

PICs were identified from Clinical and Expression datasets using a three step approach (Figure 1).



TCGA Data – 23 cancers

mRNA Expression (HiSEQ)

miRNA Expression (miRNASeq)

STEP 1: CORRELATION ANALYSIS

STEP 2: COX PROPORTIONAL HAZARD ANALYSIS

mRNA Hazard analysis

miRNA Hazard analysis

STEP 3: MERGING OF RESULT SETS

Significant mRNA

Significant miRNA

Significantly inversely correlated miRNA - mRNA pairs

Merged, filtered and annotated (Targetscan, miRWALK)

PICs

Step1: CORRELATION ANALYSIS

Pair wise correlation analysis was performed for all possible combinations of miRNA and mRNA for each cancer type. Correlation coefficients and p values were calculated.

Step 2: HAZARD ANALYSIS

Cox proportional hazards analyses by using clinical and expression data were performed for all miRNAs and mRNAs separately for each cancer type using the 'coxph' function of R library 'survival'. Hazard ratio and log likelihood p values were calculated for each miRNA and mRNA.

Step 3: MERGING OF CORRELATION AND HAZARD DATA

Results generated from correlation analysis were merged with hazard analysis results for miRNA and mRNA. At this step, we retained only the pairs which showed

a) Significant negative correlation between partners in the pair (p<0.05)

b) Significant hazard ratio for miRNA and mRNA (p<0.05)

c) Opposite direction of hazards for the partners (<1 or >1).

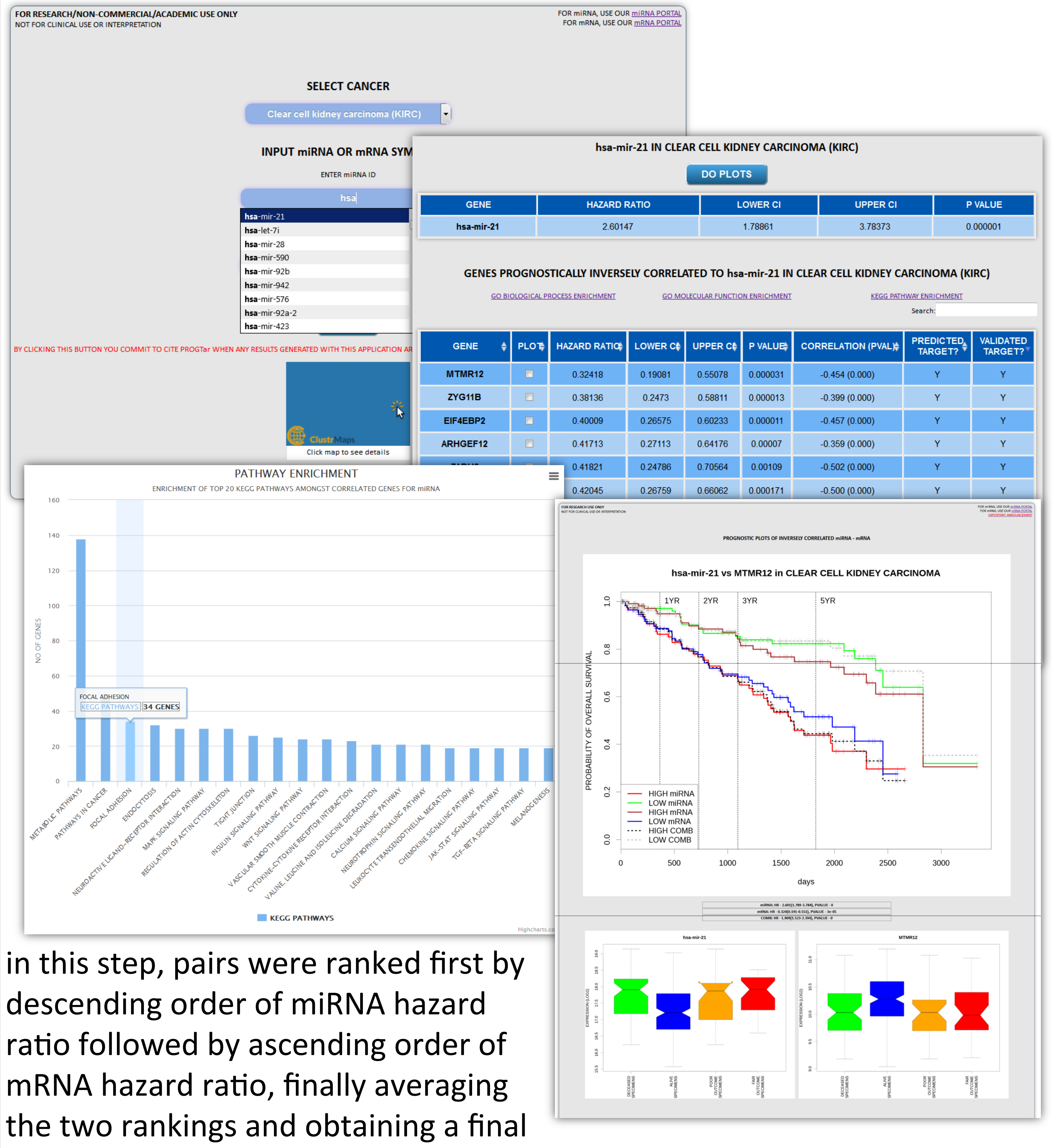
Any PICs which did not satisfy the above mentioned criteria were discarded

ANNOTATION AND RANKING OF PICs

We used miRWALK and TargetScan databases for annotating the partners in our results for experimentally validated or predicted target relationship respectively. Final ranking of PICs was done using a 2 step ranking process.

In first step (mRNA based ranking) we ranked each pair by descending order of mRNA hazard ratio (data not shown). This was followed by ranking pairs by ascending miRNA hazard ratio. The two rankings so obtained were averaged and a final mRNA based ranking was obtained by ascending order of this average.

In second step (miRNA based ranking) we repeated the process except that in this step, pairs were ranked first by descending order of miRNA hazard ratio followed by ascending order of mRNA hazard ratio, finally averaging the two rankings and obtaining a final rank by ascendingly ordering the average rank. miRNA and genes are displayed by ascending ranks on the web application.



FOR RESEARCH/NON-COMMERCIAL/ACADEMIC USE ONLY
NOT FOR CLINICAL USE OR INTERPRETATION

FOR miRNA, USE OUR miRNA PORTAL
FOR mRNA, USE OUR mRNA PORTAL

SELECT CANCER

Clear cell kidney carcinoma (KIRC)

INPUT miRNA OR mRNA SYM

ENTER miRNA ID

hsa

hsa-mir-21
hsa-let-7i
hsa-mir-28
hsa-mir-590
hsa-mir-92b
hsa-mir-942
hsa-mir-576
hsa-mir-92a-2
hsa-mir-423

BY CLICKING THIS BUTTON YOU COMMIT TO CITE PROGTar WHEN ANY RESULTS GENERATED WITH THIS APPLICATION AR

Click map to see details

hsa-mir-21 IN CLEAR CELL KIDNEY CARCINOMA (KIRC)

DO PLOTS

GENE	HAZARD RATIO	LOWER CI	UPPER CI	P VALUE
hsa-mir-21	2.60147	1.78861	3.78373	0.000001

GENES PROGNOSTICALLY INVERSELY CORRELATED TO hsa-mir-21 IN CLEAR CELL KIDNEY CARCINOMA (KIRC)

GO BIOLOGICAL PROCESS ENRICHMENT   GO MOLECULAR FUNCTION ENRICHMENT   KEGG PATHWAY ENRICHMENT

Search:

GENE	PLOT	HAZARD RATIO	LOWER CI	UPPER CI	P VALUE	CORRELATION (PVAL)	PREDICTED TARGET?	VALIDATED TARGET?
MTMR12	□	0.32418	0.19081	0.55078	0.000031	-0.454 (0.000)	Y	Y
ZYG11B	□	0.38136	0.2473	0.58811	0.000013	-0.399 (0.000)	Y	Y
EIF4EBP2	□	0.40009	0.26575	0.60233	0.000011	-0.457 (0.000)	Y	Y
ARHGEF12	□	0.41713	0.27113	0.64176	0.00007	-0.359 (0.000)	Y	Y
[illegible]	□	0.41821	0.24786	0.70564	0.00109	-0.502 (0.000)	Y	Y
[illegible]	[illegible]	0.42045	0.26759	0.66062	0.000171	-0.500 (0.000)	Y	Y

PATHWAY ENRICHMENT

ENRICHMENT OF TOP 20 KEGG PATHWAYS AMONGST CORRELATED GENES FOR miRNA

FOR RESEARCH USE ONLY
NOT FOR CLINICAL USE OR INTERPRETATION

PROGNOSTIC PLOTS OF INVERSELY CORRELATED miRNA - mRNA

hsa-mir-21 vs MTMR12 in CLEAR CELL KIDNEY CARCINOMA

WEB APPLICATION

The application is available at URL mentioned in the abstract. The **home page** lists the **available cancer types** and input boxes for miRNA and gene symbols. The input boxes allow searching of potential symbol matches by typing a few characters and the results displayed are in descending order of ranking for PICs in which molecule partners appear. Users can input either a miRNA or Gene symbol in this page to view hazard related statistics and a list of inversely correlated partners of the molecule on the **results page**. The results page also allows users, when searching for mRNA partners for a miRNA of interest to **map** the miRNAs into **GO Biological Processes, GO Molecular functions, or to KEGG pathways.**

Available for each inversely correlated molecule on the results page is correlation coefficient with p value, hazard ratio with p value and confidence intervals and predicted and/or validated target information.

On the results page, users can select a few partners of interest and create prognostic plots for the pair on the plots page. The plots page displays **Kaplan Meier plots** for high and low expression arms of both partners .

Also displayed on the plots page are **expression plots** for both partners.