

Integrated Gene Set Analysis for microRNA Studies

Estimation of the type I error rate

2015-10-22

Methods

The estimation of the type I error rate have been computed from a permutation strategy of the data.

We have created a table of all miRNA - gene target pairs, and then randomly permuted the gene column. This would have the effect of preserving the number of genes each miRNA targets, and the genes and the number of miRNAs each gene is associated with, but would remove all biological association within and between miRNAs.

All scripts are available in <https://github.com/dmontaner-papers/gsa4mirna> (folder: scripts_permutation).

Results

For each cancer type, this table shows the percentage of significant results.

The amount of False Positives (FP) found in the permuted analysis remains under the expected thresholds according to the p-value cutoff selected.

Cancer	% FP in paired studies	% FP in unpaired studies
BLCA	0	0
BRCA	0	0
CESC	0.13542271	0.01934610
COAD		4.99129425
ESCA	0.03869220	0.17411492
HNSC	0.13542271	0.27084542
KICH	0.17411492	0.01934610
KIRC	0.19346102	2.01199458
KIRP	0.40626814	0.09673051
LIHC	0.01934610	4.39156510
LUAD	0.61907526	0
LUSC	1.02534339	0.90926678
PAAD	0	0
PCPG	0	0.07738441
PRAD	0.01934610	0
READ		0.85122848
SKCM		1.29618882
STAD	0.19346102	0.01934610
THCA	0	0
UCEC	0.09673051	1.45095763