PanomiR: A systems biology framework for analysis of multi-pathway targeting by miRNAs

Pourya Naderi Yeganeh[1,2,3], Yue Yang Teo[4,5], Dimitra Karagkouni[1,2,3,6], Yered Pita-Juárez[1,2,3,6], Sarah L. Morgan[7], Ioannis S Vlachos[1,2,3,6], and Winston Hide[1,2,3,*]

1 Harvard Medical School, Boston, MA, 02115
2 Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, 02215
3 Harvard Initiative for RNA Medicine, Boston, MA, 02115
4 National University of Singapore, Singapore
5 École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
6 Broad Institute of MIT and Harvard
7 Centre for Neuroscience, Surgery and Trauma, Blizard Institute, Queen Mary University of London, London, E1 2AT, UK

* To whom correspondence should be addressed. Email: whide@bidmc.harvard.edu

Abstract (200-word limit):

Charting miRNA regulation across multiple pathways is central to characterizing their role in disease. Yet, current methods expose only isolated miRNA-pathway interactions. We have developed a systems biology approach, Pathway networks of miRNA Regulation (PanomiR) that identifies miRNAs that coordinately target groups of pathways. PanomiR captures activity dynamics of pathways, detects dysregulated pathways, groups pathways by their co-expression, and detects miRNAs that specifically target these coordinated groups of dysregulated pathways. We have used PanomiR to explore disease-associated miRNAs and pathways in liver cancer. We show that it recapitulates known and novel liver cancer pathways, determines their up or down regulation, and sensitively captures biologically meaningful signals independent of detectable differential gene expression. PanomiR organizes liver cancer pathways into three broad groups that reflect pathogenic mechanisms of cancer: activation of transcription, cell cycle and proliferation, and dysregulated signalling. Using either experimentally-supported or predicted miRNA-mRNA interactions, PanomiR robustly detects and identifies miRNAs central to liver cancer, unbiased by the number of their gene targets-making it possible to place these cancer miRNAs using PanomiR into the context of co-activated pathways that they regulate. PanomiR is a sensitive, unbiased framework for detection of disease-specific multi-pathway programs under miRNA regulation. PanomiR is accessible as an open-source R/Bioconductor package: <https://bioconductor.org/packages/PanomiR>.

Keywords: microRNA, miRNA, Pathways, biological networks, systems biology, miRNA prioritization, pathway analysis.

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs acting as potent regulators of cellular functions and molecular pathways (1). They post-transcriptionally regulate gene expression and can orchestrate co-expressed genes to coordinate function across distinct pathways. miRNA dysregulation has been shown to be a central component of the pathogenesis of diverse diseases, including neoplastic conditions and Alzheimer's (2–19). Since a miRNA can target dozens of genes, characterizing their roles in health and disease requires the charting of this coordinated co-regulation across heterogeneous molecular cascades and pathways. To date, no framework exists to be able to analyze and prioritize multi-pathway dynamics of miRNA-orchestrated regulation.

Current best practice for the transcriptomic study of miRNA regulation relies on isolated miRNA-gene or one-to-one miRNA-pathway relationships. In the latter case (miRNA:pathway), techniques such as gene set enrichment and correlation are used to detect whether a pathway is potentially regulated by a miRNA. Enrichment analyses evaluate the presence (overlap) of targets of a single miRNA in a single pathway, aiming to identify pathways with higher number of targets than expected by chance (20–26). Correlation methods evaluate the association of the expression of a single miRNA with a gene or a proxy value representing the activity of a pathway (3, 27). However, large scale processes in health and disease cross pathways, and coordinate in multiple fashions, including gene-sharing, pathway co-activity, multi-pathway co-regulation, and cross-talk (28–33). Current approaches fail to account for these complex relationships and disease-specific expression dynamics– which in turn limits our ability to detect the potential of a miRNA to regulate highly specific or even extensive gene expression programs.

We have developed a framework to address these pressing needs. *Pathway networks of miRNA Regulation* (PanomiR) supports the discovery of central miRNA regulators based on their ability to regulate coordinated pathways, forming a transcriptional program. It determines if a miRNA concurrently regulates and targets a coordinate group of disease- or function-associated pathways, as opposed to investigating isolated miRNA-pathway events. In brief, PanomiR exploits predefined pathways, their co-activation, gene expression, and annotated miRNA-mRNA interactions to extract miRNA-pathway targeting events. It performs a series of analyses to derive these multi-pathway targeting events. (i) It captures the activity of pathways and identifies disease-specific dysregulated pathways using an activity-profiling technique that accounts for overall co-activity of genes and common biases (34–36); (ii) It constructs a co-expression network of dysregulated pathways (based on predetermined pathway networks) and deconvolves it into coordinate groups of pathways that act in concert using network clustering algorithms (34–36); and, (iii) It determines miRNAs that target coordinate pathway groups using a novel statistical test and pre-determined miRNA-mRNA interactions from either experimentally-supported or predicted databases (37, 38). Together, these steps produce broad-scale, multi-pathway, and disease-specific miRNA regulatory events. The whole process is depicted In Fig. 1.

In order to showcase PanomiR's ability to detect miRNAs regulating transcriptional programs in human disease, we applied it against the Hepatocellular Carcinoma dataset of The Cancer Genome Atlas (TCGA) (39). PanomiR unbiasedly detected multi-pathway miRNA regulation events in this well-documented disease, and enabled the generation of a comprehensive framework for the evaluation of miRNA-associated mechanisms (40). We detected dysregulated pathways and their regulating miRNAs in liver cancer, assessed

biological relevance of the readouts, and evaluated the statistical robustness of PanomiR using synthetic experimentation. PanomiR recapitulates known central miRNAs in Hepatocellular carcinoma with a biologically meaningful assignment of pathways under their regulation, unbiased by the number of genes targeted by miRNAs.

MATERIALS AND METHODS

overview and input datasets

The overall goal of PanomiR is to detect miRNAs that regulate multi-pathway condition-associated transcriptional programs (Figure 1). PanomiR ingests a user-provided gene expression (such as RNA-Seq) dataset to quantify pathway activity profiles using annotated pathway datasets from the Molecular Signatures Database (MSigDB) (Figure 1a-1b). Pathway activity profiles are compared between two conditions (e.g., cancer vs control, wild type vs KO, etc) to identify dysregulated pathways (Figure 1c). To find broad-scale condition-associated groups of functions, PanomiR constructs a co-activity network of dysregulated pathways and deconvolves the network into coherent functional groups using pathway co-expression networks (Figure 1d-1e). Subsequently, PanomiR uses annotated miRNA-mRNA interaction datasets (including predicted targets from TargetScan (37) or experimentally validated interactions from Tarbase (38)) and evaluates miRNA regulatory effects on the coordinated pathways (Figure 1f). The final output of PanomiR is an ordered list of central miRNA regulators along with statistical significance levels for each group of dysregulated pathways, providing an effective means for key miRNA prioritization, ranking, and target identification.

Capturing pathway activity dynamics

PanomiR ingests a user-provided gene expression dataset and summarizes it into pathway activity scores to capture its functional dynamics (Figure 1b). The scores are proxy values for the activity of genes in individual pathways, which in turn, represent biologically meaningful functional units of reduced complexity and increased tolerance to noise compared to single gene expression datasets (3, 35, 36, 41). Pathway activity scores leverage complex inter-relationships and co-activity of genes. They provide the means to examine biological functions in a continuum and are capable of detecting biological signals where standard differential gene expression analyses fail (3, 27, 34, 36, 41–43).

To capture pathway activity profiles, we have extended previously developed methods, as follows. In a two-step process: (a) we rank genes in each sample in descending order, according to their expression– the highest expressed gene gets the largest rank-score. (b) we calculate the average squared ranks of genes that belong to a pathway as the activity score. Form ally, for a pathway X with n genes, $Pathway_x = \{ g_1^x, \ldots, g_n^x \}$, the activity score, $Ac_{a,x}$, in sample a is:

$$Ac_{a,x} = \frac{1}{n}\sum_{i}^{n} rank_a(g_i^x)^2$$

where $rank_a(g_i^x)$ refers to the rank of gene g_i^x (descending order) in sample a based on expression values. We repeatedly generate activity profiles for each pathway of interest in each sample. Then, pathway profiles are normalized across the input samples (Supplementary Material). PanomiR uses the *canonical pathways* collection from *the Molecular Signatures database* (MSigDB) as its pathway reference. MSigDB is a carefully curated database that represents non-redundant pathways from established pathway repositories such as KEGG and Reactome (44).

Detection of dysregulated disease-associated pathways

PanomiR compares pathway activity profiles between case and control subjects to determine functional dynamics in disease. It uses linear models, implemented in the *Limma* package (45), to define dysregulated pathways by determining statistically significant differences in pathway activity profiles between cases and controls. (Figure 1c) (45). In contrast to enrichment analysis, the linear modeling framework determines the direction of dysregulation: whether a pathway is upregulated or downregulated in disease subjects, and accounts for confounding variables such as batch, center, or any other fixed effects and continuous covariates. PanomiR produces an ordered table of dysregulated pathways along with p-values of dysregulation, adjusted for multiple hypothesis testing using False Discovery Rate (FDR) (46).

Detection of groups of dysregulated pathways using their co-expression networks

Dysregulation of individual pathways is rarely an isolated event, since pathways share activity and are often co-regulated. PanomiR accounts for co-regulation to place dysregulated pathways into groups that represent coordinated high-level disease programs by exploiting the *Pathway co-expression network* (PCXN) (36), a tool that assesses the shared activity of pathways (Figure 1d). Panomir leverages PCxN co-activated pathways generated from a curated dataset of 3,207 expression profiles, resulting in an independent platform to query co-activity of all pathways in the MSigDB dataset (36, 47).

PanomiR masks PCxN to only include the subnetwork of dysregulated pathways identified from the disease gene expression data in the previous step. In the masked network, nodes represent dysregulated pathways and edges activity-correlation of pathways. PanomiR subsequently identifies densely interconnected pathway subnetworks using clustering algorithms (Figure 1e). The subnetworks denote clusters of highly correlated coordinate groups of dysregulated pathways driving disease or condition-specific functions.

miRNA prioritization within clusters of dysregulated pathways

We hypothesize that a subset of coordinated groups of disease-associated pathways have common miRNA regulators. PanomiR analyzes clusters of dysregulated pathways to define central miRNAs using annotated miRNA-mRNA interactions and an optimized test statistic (Figure 1f, Figure 2) capturing the extent to which the targets of a specific miRNA are present within a group of coordinated pathways.

PanomiR identifies miRNA regulatory events in three sequential steps (Figure 2). In brief,(i) by calculating individual miRNA-pathway overlap scores, (ii) by generalizing miRNA targeting scores to a group of pathways (i.e., a cluster of dysregulated pathways), and (iii) by estimating the statistical significance of miRNA targeting scores using an empirical approach, specific to the input dataset, for each miRNA and each cluster of dysregulated pathways.

In the first step, PanomiR derives the overlap scores for individual miRNA-pathway pairs using p-values of Fisher's Exact test, capturing overrepresentation of targets of a specific miRNA in each individual pathway. To make analysis disease or condition specific, PanomiR calculates overlap scores using only the genes expressed in the input experiment.

In the second step, an overall targeting score for a given cluster of pathways (Figure 2) is derived. The cluster of pathways are generated in the previous step using PCxN. Formally, for each cluster of dysregulated pathways C, the targeting score of a miRNA x is:

$$S_x^c = \frac{1}{|C|}\sum_{y \in C} \Phi^{-1}(1 - P_{xy})$$

where Φ^{-1}(.) denotes the inverse of the standard normal cumulative distribution function (CDF). P_{xy} denotes the Fisher's Exact test p-value overlap between targets of miRNA x and genes of pathway y. A targeting-score, S_x^c , is related to Stouffer's method for p-value aggregation. Using the inverse normal CDF avoids extreme cases in which a miRNA has many targets in one pathway and only a few targets in other pathways in a cluster.

In the third step, the statistical significance of the targeting score S_x^c is determined. This targeting-score does not constitute by itself an unbiased measure of miRNA-targeting as it might depend on the number of targets of a miRNA. To create an unbiased measure, PanomiR derives an empirical targeting p-value,$P(S_x^c)$, x with a score of S_x^c. This p-value denotes the probability of observing a larger targeting score from a random cluster of pathways ($|C|$ members). This empirical probability is derived using a bootstrap sampling approach by selecting randomized groups of pathways and re-calculating their cluster targeting score. The output p-values are then adjusted for multiple hypothesis comparison using the Benjamini-Hochberg False Discovery Rate (FDR).

Given the computational cost of bootstrap sampling, especially to calculate small p-values, PanomiR employs a Gaussian approximation approach to estimate $P(S_x^c)$. In clusters of large-enough size (>30 pathways), S_x^c values follow a normal distribution according to the Central Limit Theorem. PanomiR uses pre-calculated the Gaussian distribution estimates from 1,000 random S_x values to overcome the computational costs in these cases. In the last step, miRNAs are prioritized based on p-values for targeting each cluster. Details on assessments of the Gaussian estimation and robustness of $P(S_x^c)$-values using a jackknife estimation are provided in the Supplementary Material. PanomiR outputs cluster-specific lists of miRNAs ranked by targeting p-values.

RESULTS

We assessed PanomiR's utility to achieve a systematic, unbiased, and biologically-meaningful determination of regulatory miRNAs. We applied PanomiR to a liver cancer gene expression dataset from TCGA as a case study (48). Figures 3 and 4 portray PanomiR's recapitulation of liver cancer-associated pathways, their coordination, and the miRNAs that target them. We found three clusters of dysregulated pathways in liver cancer representing coherent function of high-level cancer mechanisms: transcription, cell replication, and signaling. PanomiR detected miRNAs that targeted each cluster using either experimentally-supported (TarBase v8.0) or predicted (TargetScan v7.2) miRNA-miRNA interactions (Tables 2 and 3). By comparing PanomiR's results with the relevant literature and with standard enrichment analysis, we show that PanomiR provides informative biological inference of multi-pathway targeting by miRNAs.

PanomiR detects multiple liver cancer-associated pathways

We generated and compared pathway activity profiles from normal tissues (TCGA Abbreviation: NT, n = 49) and primary solid tumors (TCGA Abbreviation: TP, n = 368) from liver cancer gene expression RNAseq data using the MSigDB pathway database. PanomiR detected 428 upregulated and 397 downregulated pathways in TP compared to NT (FDR < 0.01, Table 1 and Supplementary Table S1).

Prioritized pathways reflect well-established dysregulated functions in liver cancer (Table 1). For example, *NUCLEAR SIGNALING BY ERBB4* is downregulated in TP and activated in NT and has the highest statistical significance among all pathways (Figure 3a, Table 1). Downregulation of ERBB4 in tumors is in concordance with a well-established body of evidence on the roles of ERBB signaling as a tumor suppressor in liver cancer (49, 50). In addition, we found downregulation of *HDL MEDIATED LIPID TRANSPORT* in tumor tissues, corroborated by several reports on lipid disorders in liver cancer including decreased plasma levels of HDL (51, 52). These results suggest the utility of PanomiR in detecting dysregulated disease functions through pathway activity analysis.

We compared pathway readouts from PanomiR with pathway enrichment analysis of differentially expressed genes in the same liver cancer dataset. Enrichment analysis using Fisher's Exact Test identified 51 enriched pathways in differentially expressed (DE) genes (FDR <0.01, Supplementary Table S2), 50 of which were also determined as dysregulated by PanomiR. Significant overlap between the results suggests that PanomiR recapitulates the majority of enrichment analysis readouts (Fisher's Exact Test p-value = 3.5e-08). However, PanomiR also detected dysregulated liver cancer pathways that were missed by enrichment analysis. For example, the top liver cancer-associated pathway according to PanomiR, *NUCLEAR SIGNALING BY ERBB4,* was not detected by enrichment analysis (p-value = 1), since overrepresentation analysis prioritizes pathways with more DE genes than expected by chance, and missing pathways with significantly differential activity between pathways and controls but not increased DE gene proportions. Table 1 and Supplementary Table 1 show several other instances of pathways that were detected by PanomiR but were not identified in the standard enrichment analysis.

Synthetic data analysis shows PanomiR captures biologically-meaningful signals

To assess the recapitulation of biological signals by PanomiR, we employed two randomization tests (Supplementary Material). In the first, we asked to what extent PanomiR detected dysregulated pathways in a random assignment of samples to case and control groups in liver cancer (i.e., biologically meaningless classes). PanomiR found a very small number of dysregulated pathways (0.054 ± 1.2, Mean ± SD) via randomized case/control sample assignment (Supplementary Table S3). In the second test, we asked whether the use of biologically meaningful pathways (as annotated in the MSigDB) held advantages over using randomly assigned gene sets. We generated randomized pathways by permuting gene labels to conserve the overlap structure of the original MSigDB dataset. We found that annotated gene-sets generate a significantly larger number of dysregulated pathways (one-sided Z test p-value $< 3.34 * 10^{-5}$). We also compared the distribution of adjusted p-values from dysregulated pathways from MSigDB and randomized pathway collections, irrespective of FDR cut-offs. This experiment showed a significant difference between the two scenarios according to an one-sided Kolmogorov-Smirnov test (p-value < 2.86E-18 Supplementary Table S3, Supplementary Figure S1). Biologically meaningful gene sets were more likely to be dysregulated than randomized pathways and more likely to capture biological signals.

Identification of coordinate clusters of dysregulated pathways

Pathways coordinate and co-regulate through various mechanisms, including gene sharing. To detect coordinated groups of dysregulated pathways, we used the Pathway Co-expression Network (PCxN), where edges represent precalculated correlations between pathways based on independent gene expression data (36). We mapped the 200 most statistically significant dysregulated pathways onto the PCxN network and performed Louvain clustering to identify coordinate pathway groups. PanomiR identified 3 major clusters of dysregulated pathways with consistent functions (Figure 4).

The largest cluster of dysregulated pathways (Cluster A) contained pathways upregulated in TP such as *SPLICEOSOME, PROTEASOME, TRANSLATION, RNA POLL II TRANSCRIPTION, and SIGNALLING BY WNT* (Supplementary Table S4). Wnt signaling activation is a critical mechanism for transformation of precancerous lesions into liver cancer through proliferation (53). The second largest cluster (Cluster B) contained terms related to cell cycle and proliferation (Figure 4, Supplementary Table S4). The third cluster (Cluster C) contained liver cancer-associated signaling pathways that were either downregulated or upregulated in TP vs NT with terms related to ERBB signaling, IL signaling, and NOTCH signaling (Supplementary Table 4). Dysregulated pathways within clusters A and B showed a coherent direction of differential activation, suggesting coordinate multi-pathway dysregulation in driving high-order functions. We validated the robustness of pathway clustering using a variety of parameters and algorithms (Supplementary Figure S2). The results indicate that PanomiR successfully deconvolves distinct groups of dysregulated pathways that represent higher order functional programs of liver cancer.

Detection of regulatory miRNAs that target dysregulated pathway clusters

We hypothesized that coordinated clusters of dysregulated pathways have common miRNA regulators. In our case study, we examined separately experimentally-supported (TarBase v8.0) and predicted miRNA-mRNA interactions (Targetscan v7.2) to detect miRNAs that targeted each cluster of dysregulated pathways (Tables 2 and 3, Supplementary Tables S5 and S6). Our results show that PanomiR identified distinct miRNAs for each cluster of liver cancer-associated pathways.

When using experimentally-supported interactions, PanomiR detected 202, 104, and 1 miRNA regulators in clusters A, B, and C respectively (FDR < 10^{-5}, Table 2, Supplementary Table 5). These included known liver cancer-associated miRNAs with consistent modes of action with their targeted pathway clusters. Cluster A was targeted by miR-525-3p, miR-1307, miR-631, and miR-663a– these miRNAs have been previously shown for their role in tumor migration and invasion (54–57). Cluster B was targeted by miRNAs with established roles in regulating cell-cycle in liver cancer including miR-107, miR-124-3p, and miR-103a-3p. For example, miR-107 is a P53-associated regulator of cell cycle and proliferation, elevated in early-stage liver cancer (58–61); miR-124-3p is a tumor suppressor that regulates proliferation and invasion in liver cancer by inducing G1-phase cell-cycle arrest (62, 63); miR-103a-3p is a promoter of proliferation that is highly dysregulated in liver cancer (64). In cluster C, we found miR-410-3p as a central regulator of the relevant module. This miRNA has been shown to be a circulating biomarker of distant metastasis into the lung and the liver (65, 66), it also regulates adenomas via signaling pathways such as MAPK, PTEN/AKT, and STAT (67, 68). In cluster C, we also found a significant targeting by miR-552-3p, which has been associated with liver cancer and regulates various hallmarks of cancer (69). Supplementary material provides an examination of the relationship between PanomiR miRNAs with DE miRNAs in TP vs NT. While we did not find a significant association between prioritization by PanomiR and differential expression, PanomiR attributes distinct DE miRNAs to distinct groups of pathway targeting events– providing a knowledge-driven approach for functional characterization of data-driven disease miRNAs. Our results establish that PanomiR successfully detects key regulating liver-cancer miRNAs and their downstream dysregulated pathways.

We also assessed PanomiR using predicted miRNA-mRNA interactions (37). Although PanomiR detected multiple liver cancer-associated miRNAs from predicted interactions, the set of prioritized miRNAs were different than that of experimentally validated interactions (Table 3, Supplementary Table 6). For example, PanomiR prioritized miR-299-3p in cluster C, a regulator

of IL and STAT signaling pathways in liver cells, which have several associated annotated pathways in cluster C (70). Our results suggest that predicted and experimentally validated miRNA interactions databases produce complementary results and both should be considered as such for analyzing transcriptomic data.

We compared miRNA detection of PanomiR with miRNA-pathway enrichment analysis in our case study (Tables 2 and 3). For comparative purposes we employed three extensions of miRNA-enrichment analysis test to adapt to multi-pathway scenarios. (a) We extended enrichment analysis to a group (cluster) of pathways by asking how many pathways within a given pathway cluster were significantly enriched for targets of a miRNA. For example, if the targets of a miRNA x are significantly enriched in five pathways within a group of pathways, the miRNA gets a targeting score of 5. (b) We used Stouffer's method to obtain one single p-value that combines enrichment p-values of a miRNA within all pathways in a cluster. (c) We used Fisher's p-value aggregation method to combine all enrichment p-values of a miRNA as an alternative of Stouffer's method.

PanomiR successfully detected liver cancer-associated miRNAs that were not prioritized by extended enrichment tests. When using experimentally-supported miRNA-mRNA interactions, enrichment analysis of cluster A found miR-525-3p as enriched in only one and miR-1307-5p in none out of 65 pathways (Table 2). When using predicted miRNA-mRNA interactions, PanomiR detected several miRNAs that were not detected by the extended enrichment analysis (Table 3). It is of note that the enrichment tests (Tables 2 and 3) used a highly-relaxed FDR threshold (FDR <0.25) to enable a more sensitive detection. Using a conservative FDR cut-off (e.g., FDR < 0.05) would have resulted in an even lower detection rate of miRNAs. The results suggest that (a) PanomiR can detect liver cancer-associated miRNAs that are not detectable by simple enrichment tests, and (b) a subset of critical liver cancer miRNAs can be detected only by analyzing a group of pathways, and not by examining individual pathways.

Enrichment analyses have been shown to be biased towards detecting miRNAs with a larger number of targets (71). We examined the relationship of the number of targets of a miRNA with its prioritization ranking by PanomiR or enrichment analysis extensions (Table 3). The enrichment ranking of miRNAs significantly correlated with their number of gene targets whereas PanomiR was unbiased to the number of gene targets. Figure 5 displays the bias of enrichment analysis (including Stouffer's and Fisher's extensions) towards prioritizing miRNAs with more targets in cluster A, while miRNAs with a small number of targets did not rank highly. PanomiR did not show a correlation between ranking according to PanomiR and the number of targets (Figure 5), suggesting its ability to prioritize miRNAs irrespective of the number of their gene targets. Supplementary material provides additional evaluation of unbiased and robust miRNA prioritization by PanomiR. Using jackknifing and bootstrapping schemas, we show that PanomiR miRNA prioritization is rather based on collective targeting of all pathways and is not driven by individual pathways

DISCUSSION

We have built a framework to determine miRNA regulation of multiple coordinately regulated pathways. Most of the existing tools for miRNA-pathway analysis are focused on one-to-one miRNA-pathway relationships without the ability to infer relationships between miRNAs and groups of co-regulated pathways. Previous studies have proposed using p-value integration methods to address multi-pathway analysis, but none address pathway co-activity/coordination and account for expression dynamics (72). PanomiR addresses these challenges by deconvolving gene expression datasets into coordinate groups of pathways which exhibit condition-associated dynamics; measuring the extent to which miRNAs target these groups. In the case study of dysregulated pathway clusters in the liver cancer dataset, PanomiR captured

large-scale features of cancers such as dysregulated transcription, cellular replication, and signaling (Figure 3). These clusters represent coherent higher-order functional units that recapitulate specific, yet central, disease mechanisms.

The use of pathway activity profiles is a key component of PanomiR for granularity in defining coordinate functional groups by sensitively identifying dysregulated pathways (Figures 2 and 3, Table 1). In our case study, PanomiR detected critical known liver cancer pathways even with few associated DE genes (Table 1). Pathway activity profiling in PanomiR also facilitated the understanding of the directionality of pathway (de)activation in disease states. Methods that use pathway activity scores are often limited in generating explainable and biologically-meaningful pathway activity profiles. For example, nonlinear dimensionality reduction has been applied previously to miRNA-pathway relationships (3). Although the nonlinear approach detects relationships between miRNAs and pathways, it does not explain actual activation or deactivation of pathways. PanomiR's activity profiles are biologically informative as a higher (or lower) pathway activity indicates a higher (or lower) overall activity of associated genes. Pathway activity profiles in PanomiR are directly comparable and translatable across different datasets, which makes it possible to leverage co-expression of pathways to detect disease-specific functional dynamics and themes across datasets and platforms (34).

Deconvolution of coordinated pathway groups allowed PanomiR to robustly and unbiasedly detect miRNA regulatory events in liver cancer that are not detected by enrichment analysis (Figure 4, Tables 2 and 3). Comparison of our results with the liver cancer literature showed that PanomiR's prioritized miRNAs have distinct roles in the disease, concordant with the functional characteristic of the pathway cluster that they were discovered from. For example, miR-107 regulates cell cycle and proliferation and targets cluster B which includes cellular replication pathways (Figure 4, Table 2). These results highlight the ability to identify miRNAs that consistently target groups of pathways even with only a few targets from each pathway. Our results demonstrate that PanomiR can robustly detect miRNAs that regulate the broad, yet specific, transcriptional programs of liver cancer.

The complete landscape of miRNA binding to mRNAs is currently unknown. This gap contributes to the discrepancy in miRNA prioritization based on the background dataset of miRNA-mRNA interactions (Tables 2 and 3). But in both cases, using TargetScan and TarBase datasets, PanomiR discovered informative and complementary miRNA regulatory events (Tables 2 and 3). These results suggest that users could tailor the background datasets (predicted or experimentally-supported) to their study design and research questions. We have made PanomiR flexible to heterogeneous miRNA-mRNA interactions and gene-expression datasets. PanomiR can be expanded (in future development) to co-expression analysis of miRNAs and pathways, which has been proposed to provide informative pointers to biological programs of diseases (3).

In summary, PanomiR is a systems biology framework to study dysregulated pathways, their co-activity, and their regulating miRNAs. PanomiR accounts for co-expression of pathways and disease-specific expression dynamics to identify miRNA-regulatory events– which is an advancement to current practice of studying static, isolated miRNA-pathway interactions. We have made PanomiR available as an open-source R/Bioconductor package for the use of the community.

AVAILABILITY

PanomiR is available as a Bioconductor package. The development version of PanomiR can be accessed via <https://github.com/pouryany/PanomiR> and

<https://bioconductor.org/packages/PanomiR>. Additional scripts and analyses, specific to this manuscript are available via <https://github.com/pouryany/PanomiR_paper>. Predicted miRNA-mRNA interactions were all retrieved from <http://www.targetscan.org/vert_72/>.

FUNDING
This work was supported by the Harvard Medical School Aging Initiative Pilot Grant (WAH, IV, PNY). S.M. supported by Cure Alzheimer's Foundation.

ACKNOWLEDGEMENTS
We would like to thank Dr. Isabel Castanho for valuable feedback. We thank Cure Alzheimer's Foundation for their support.

CONFLICT OF INTEREST
Authors declare no conflict of interest.

REFERENCES

1. Friedman,R.C., Farh,K.K.-H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

2. Cai,Y., Yu,X., Hu,S. and Yu,J. (2009) A brief review on the mechanisms of miRNA regulation. *Genomics Proteomics Bioinformatics*, **7**, 147–154.

3. Wilk,G. and Braun,R. (2018) Integrative analysis reveals disrupted pathways regulated by microRNAs in cancer. *Nucleic Acids Res.*, **46**, 1089–1101.

4. Cui,Q., Yu,Z., Purisima,E.O. and Wang,E. (2006) Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.*, **2**, 46.

5. Artmann,S., Jung,K., Bleckmann,A. and Beissbarth,T. (2012) Detection of simultaneous group effects in microRNA expression and related target gene sets. *PLoS One*, **7**, e38365.

6. Werfel,S., Leierseder,S., Ruprecht,B., Kuster,B. and Engelhardt,S. (2017) Preferential microRNA targeting revealed by in vivo competitive binding and differential Argonaute immunoprecipitation. *Nucleic Acids Res.*, **45**, 10218–10228.

7. Peláez,N. and Carthew,R.W. (2012) Biological robustness and the role of microRNAs: a network perspective. *Curr. Top. Dev. Biol.*, **99**, 237–255.

8. Kehl,T., Backes,C., Kern,F., Fehlmann,T., Ludwig,N., Meese,E., Lenhof,H.-P. and Keller,A. (2017) About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget*, **8**, 107167–107175.

9. Pasquinelli,A.E. (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.*, **13**, 271–282.

10. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68-73.

11. Mishima,Y. (2012) Widespread roles of microRNAs during zebrafish development and beyond. *Dev. Growth Differ.*, **54**, 55–65.

12. Hashimoto,N. and Tanaka,T. (2017) Role of miRNAs in the pathogenesis and susceptibility of diabetes mellitus. *J. Hum. Genet.*, **62**, 141–150.

13. Deng,X., Su,Y., Wu,H., Wu,R., Zhang,P., Dai,Y., Chan,T.-M., Zhao,M. and Lu,Q. (2015) The role of microRNAs in autoimmune diseases with skin involvement. *Scand. J. Immunol.*, **81**, 153–165.

14. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–906.

15. Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.

16. Lima,R.T., Busacca,S., Almeida,G.M., Gaudino,G., Fennell,D.A. and Vasconcelos,M.H. (2011) MicroRNA regulation of core apoptosis pathways in cancer. *Eur. J. Cancer*, **47**, 163–174.

17. Sumazin,P., Yang,X., Chiu,H.-S., Chung,W.-J., Iyer,A., Llobet-Navas,D., Rajbhandari,P., Bansal,M., Guarnieri,P., Silva,J., *et al.* (2011) An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*, **147**, 370–381.

18. Gascon,E. and Gao,F.-B. (2012) Cause or effect: Misregulation of microRNA pathways in neurodegeneration. *Front. Neurosci.*, **6**, 48.

19. Rao,X., Di Leva,G., Li,M., Fang,F., Devlin,C., Hartman-Frey,C., Burow,M.E., Ivan,M., Croce,C.M. and Nephew,K.P. (2011) MicroRNA-221/222 confers breast cancer fulvestrant resistance by regulating multiple signaling pathways. *Oncogene*, **30**, 1082–1097.

20. Vlachos,I.S., Zagganas,K., Paraskevopoulou,M.D., Georgakilas,G., Karagkouni,D., Vergoulis,T., Dalamagas,T. and Hatzigeorgiou,A.G. (2015) DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.*, **43**, W460-6.

21. Chen,L., Heikkinen,L., Wang,C., Yang,Y., Knott,K.E. and Wong,G. (2018) miRToolsGallery: a tag-based and rankable microRNA bioinformatics resources database portal. *Database (Oxford)*, **2018**.

22. Chen,L., Heikkinen,L., Wang,C., Yang,Y., Sun,H. and Wong,G. (2019) Trends in the development of miRNA bioinformatics tools. *Brief. Bioinform.*, **20**, 1836–1852.

23. Backes,C., Khaleeq,Q.T., Meese,E. and Keller,A. (2016) miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res.*, **44**, W110-6.

24. Steinfeld,I., Navon,R., Ach,R. and Yakhini,Z. (2013) miRNA target enrichment analysis reveals directly active miRNAs in health and disease. *Nucleic Acids Res.*, **41**, e45.

25. Licursi,V., Conte,F., Fiscon,G. and Paci,P. (2019) MIENTURNET: an interactive web tool for microRNA-target enrichment and network-based analysis. *BMC Bioinformatics*, **20**, 545.

26. Li,J., Han,X., Wan,Y., Zhang,S., Zhao,Y., Fan,R., Cui,Q. and Zhou,Y. (2018) TAM 2.0: tool for MicroRNA set analysis. *Nucleic Acids Res.*, **46**, W180–W185.

27. Braun,R., Cope,L. and Parmigiani,G. (2008) Identifying differential correlation in gene/pathway combinations. *BMC Bioinformatics*, **9**, 488.

28. Khatri,P., Sirota,M. and Butte,A.J. (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.*, **8**, e1002375.

29. Rhodes,D.R., Kalyana-Sundaram,S., Tomlins,S.A., Mahavisno,V., Kasper,N., Varambally,R., Barrette,T.R., Ghosh,D., Varambally,S. and Chinnaiyan,A.M. (2007) Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia*, **9**, 443–454.

30. Merico,D., Isserlin,R., Stueker,O., Emili,A. and Bader,G.D. (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, **5**, e13984.

31. Huang,Y. and Li,S. (2010) Detection of characteristic sub pathway network for angiogenesis based on the comprehensive pathway network. *BMC Bioinformatics*, **11 Suppl 1**, S32.

32. Li,Y., Agarwal,P. and Rajagopalan,D. (2008) A global pathway crosstalk network. *Bioinformatics*, **24**, 1442–1447.

33. Karlebach,G. and Shamir,R. (2008) Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.*, **9**, 770–780.

34. Altschuler,G.M., Hofmann,O., Kalatskaya,I., Payne,R., Ho Sui,S.J., Saxena,U., Krivtsov,A.V., Armstrong,S.A., Cai,T., Stein,L., *et al.* (2013) Pathprinting: An integrative approach to understand the functional basis of disease. *Genome Med.*, **5**, 68.

35. Joachim,R.B., Altschuler,G.M., Hutchinson,J.N., Wong,H.R., Hide,W.A. and Kobzik,L. (2018) The relative resistance of children to sepsis mortality: from pathways to drug candidates. *Mol. Syst. Biol.*, **14**, e7998.

36. Pita-Juárez,Y., Altschuler,G., Kariotis,S., Wei,W., Koler,K., Green,C., Tanzi,R.E. and Hide,W. (2018) The Pathway Coexpression Network: Revealing pathway relationships. *PLoS Comput. Biol.*, **14**, e1006042.

37. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**.

38. Karagkouni,D., Paraskevopoulou,M.D., Chatzopoulos,S., Vlachos,I.S., Tastsoglou,S., Kanellos,I., Papadimitriou,D., Kavakiotis,I., Maniou,S., Skoufos,G., *et al.* (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.*, **46**, D239–D245.

39. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu and Cancer Genome Atlas Research Network (2017) Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*, **169**, 1327-1341.e23.

40. Szabo,G. and Bala,S. (2013) MicroRNAs in liver disease. *Nat. Rev. Gastroenterol. Hepatol.*, **10**, 542–552.

41. Fan,J., Salathia,N., Liu,R., Kaeser,G.E., Yung,Y.C., Herman,J.L., Kaper,F., Fan,J.-B., Zhang,K., Chun,J., *et al.* (2016) Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods*, **13**, 241–244.

42. Tomfohr,J., Lu,J. and Kepler,T.B. (2005) Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*, **6**, 225.

43. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S., *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15545–15550.

44. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdóttir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.

45. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

46. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.

47. McCall,M.N., Jaffee,H.A., Zelisko,S.J., Sinha,N., Hooiveld,G., Irizarry,R.A. and Zilliox,M.J. (2014) The Gene Expression Barcode 3.0: improved data processing and mining tools. *Nucleic Acids Res.*, **42**, D938-43.

48. Mounir,M., Lucchetta,M., Silva,T.C., Olsen,C., Bontempi,G., Chen,X., Noushmehr,H., Colaprico,A. and Papaleo,E. (2019) New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.*, **15**, e1006701.

49. Ito,Y., Takeda,T., Sakon,M., Tsujimoto,M., Higashiyama,S., Noda,K., Miyoshi,E., Monden,M. and Matsuura,N. (2001) Expression and clinical significance of erb-B receptor family in hepatocellular carcinoma. *Br. J. Cancer*, **84**, 1377–1383.

50. Liu,Y., Song,L., Ni,H., Sun,L., Jiao,W., Chen,L., Zhou,Q., Shen,T., Cui,H., Gao,T., *et al.* (2017) ERBB4 acts as a suppressor in the development of hepatocellular carcinoma. *Carcinogenesis*, **38**, 465–473.

51. Jiang,J., Nilsson-Ehle,P. and Xu,N. (2006) Influence of liver cancer on lipid and lipoprotein metabolism. *Lipids Health Dis.*, **5**, 4.

52. Jiang,J.-T., Xu,N. and Wu,C.-P. (2007) Metabolism of high density lipoproteins in liver cancer. *World J. Gastroenterol.*, **13**, 3159–3163.

53. Wang,W., Smits,R., Hao,H. and He,C. (2019) Wnt/β-catenin signaling in liver cancers. *Cancers (Basel)*, **11**, 926.

54. Pang,F., Zha,R., Zhao,Y., Wang,Q., Chen,D., Zhang,Z., Chen,T., Yao,M., Gu,J. and He,X. (2014) MiR-525-3p enhances the migration and invasion of liver cancer cells by downregulating ZNF395. *PLoS One*, **9**, e90867.

55. Eun,J.W., Seo,C.W., Baek,G.O., Yoon,M.G., Ahn,H.R., Son,J.A., Sung,S., Kim,D.W., Kim,S.S., Cho,H.J., *et al.* (2020) Circulating exosomal MicroRNA-1307-5p as a predictor for metastasis in patients with hepatocellular carcinoma. *Cancers (Basel)*, **12**, 3819.

56. Chen,B., Liao,Z., Qi,Y., Zhang,H., Su,C., Liang,H., Zhang,B. and Chen,X. (2020) MiR-631 inhibits intrahepatic metastasis of hepatocellular carcinoma by targeting PTPRE. *Front. Oncol.*, **10**, 565266.

57. Zhang,C., Chen,B., Jiao,A., Li,F., Sun,N., Zhang,G. and Zhang,J. (2018) miR-663a inhibits tumor growth and invasion by regulating TGF-β1 in hepatocellular carcinoma. *BMC Cancer*, **18**, 1179.

58. Böhlig,L., Friedrich,M. and Engeland,K. (2011) p53 activates the PANK1/miRNA-107 gene leading to downregulation of CDK6 and p130 cell cycle proteins. *Nucleic Acids Res.*, **39**, 440–453.

59. Zhang,J.-J., Wang,C.-Y., Hua,L., Yao,K.-H., Chen,J.-T. and Hu,J.-H. (2015) miR-107 promotes hepatocellular carcinoma cell proliferation by targeting Axin2. *Int. J. Clin. Exp. Pathol.*, **8**, 5168–5174.

60. Takahashi,Y., Forrest,A.R.R., Maeno,E., Hashimoto,T., Daub,C.O. and Yasuda,J. (2009) MiR-107 and MiR-185 can induce cell cycle arrest in human non small cell lung cancer cell lines. *PLoS One*, **4**, e6677.

61. Loosen,S.H., Castoldi,M., Jördens,M.S., Roy,S., Vucur,M., Kandler,J., Hammerich,L., Mohr,R., Tacke,F., Ulmer,T.F., *et al.* (2021) Serum levels of circulating microRNA-107 are elevated in patients with early-stage HCC. *PLoS One*, **16**, e0247917.

62. Lang,Q. and Ling,C. (2012) MiR-124 suppresses cell proliferation in hepatocellular carcinoma by targeting PIK3CA. *Biochem. Biophys. Res. Commun.*, **426**, 247–252.

63. Callegari,E., Gramantieri,L., Domenicali,M., D'Abundo,L., Sabbioni,S. and Negrini,M. (2015) MicroRNAs in liver cancer: a model for investigating pathogenesis and novel therapeutic approaches. *Cell Death Differ.*, **22**, 46–57.

64. Liu,Y., Zhang,Y., Xiao,B., Tang,N., Hu,J., Liang,S., Pang,Y., Xu,H., Ao,J., Yang,J., *et al.* (2021) MiR-103a promotes tumour growth and influences glucose metabolism in hepatocellular carcinoma. *Cell Death Dis.*, **12**, 618.

65. Wang,Y., Fu,J., Jiang,M., Zhang,X., Cheng,L., Xu,X., Fan,Z., Zhang,J., Ye,Q. and Song,H. (2014) MiR-410 is overexpressed in liver and colorectal tumors and enhances tumor cell growth by silencing FHL1 via a direct/indirect mechanism. *PLoS One*, **9**, e108708.

66. Liu,X. and Chu,K.-M. (2020) Exosomal miRNAs as circulating biomarkers for prediction of development of haematogenous metastasis after surgery for stage II/III gastric cancer. *J. Cell. Mol. Med.*, **24**, 6220–6232.

67. Wen,R., Umeano,A.C., Essegian,D.J., Sabitaliyevich,U.Y., Wang,K. and Farooqi,A.A. (2018) Role of microRNA-410 in molecular oncology: A double edged sword. *J. Cell. Biochem.*, **119**, 8737–8742.

68. Grzywa,T.M., Klicka,K., Rak,B., Mehlich,D., Garbicz,F., Zieliński,G., Maksymowicz,M., Sajjad,E. and Włodarski,P.K. (2019) Lineage-dependent role of miR-410-3p as oncomiR in gonadotroph and corticotroph pituitary adenomas or tumor suppressor miR in somatotroph adenomas via MAPK, PTEN/AKT, and STAT3 signaling pathways. *Endocrine*, **65**, 646–655.

69. Zou,Y., Zhao,X., Li,Y. and Duan,S. (2020) miR-552: an important post-transcriptional regulator that affects human cancer. *J. Cancer*, **11**, 6226–6233.

70. Servais,F.A., Kirchmeyer,M., Hamdorf,M., Minoungou,N.W.E., Rose-John,S., Kreis,S., Haan,C. and Behrmann,I. (2019) Modulation of the IL-6-signaling pathway in liver cells by miRNAs targeting gp130, JAK1, and/or STAT3. *Mol. Ther. Nucleic Acids*, **16**, 419–433.

71. Bleazard,T., Lamb,J.A. and Griffiths-Jones,S. (2015) Bias in microRNA functional enrichment analysis. *Bioinformatics*, **31**, 1592–1598.

72. Vlachos,I.S., Kostoulas,N., Vergoulis,T., Georgakilas,G., Reczko,M., Maragkakis,M., Paraskevopoulou,M.D., Prionidis,K., Dalamagas,T. and Hatzigeorgiou,A.G. (2012) DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res.*, **40**, W498-504.

73. Blondel,V.D., Guillaume,J.-L., Lambiotte,R. and Lefebvre,E. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.*, **2008**, P10008.

FIGURES

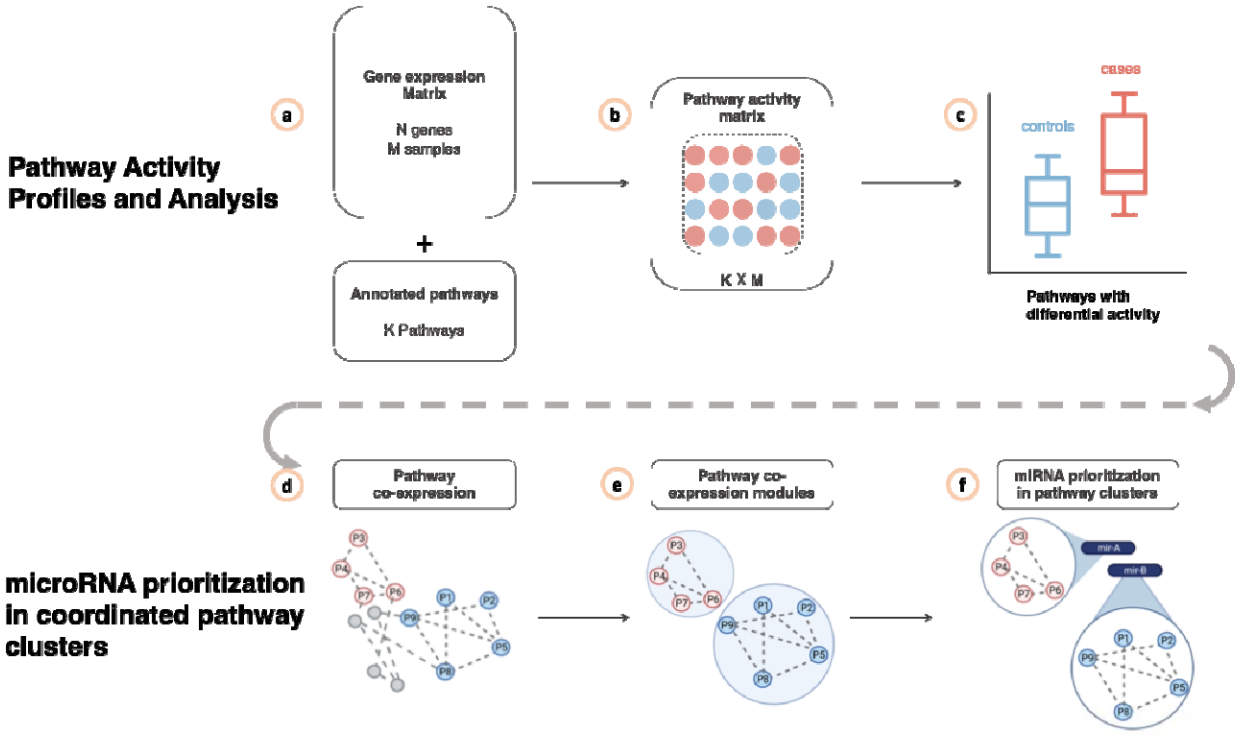
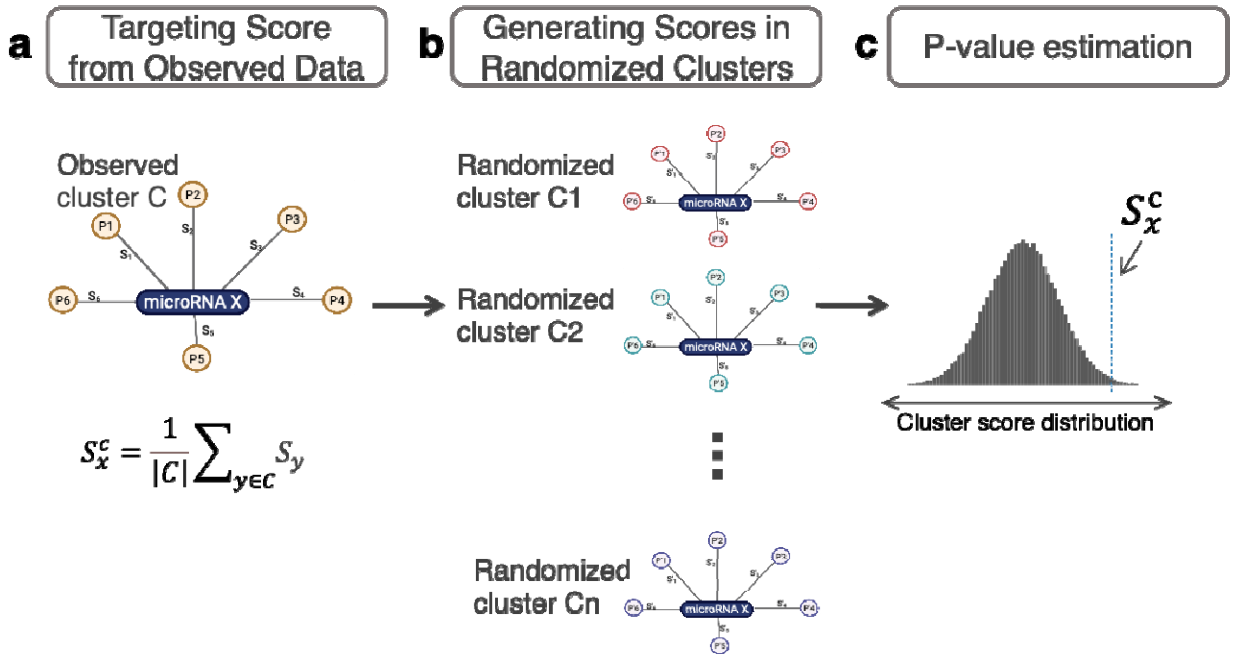


Figure 1. PanomiR workflow. PanomiR prioritizes miRNAs that target coordinated groups of pathways. **(a)** Input gene expression dataset and a set of annotated pathways **(b)** Gene expression data is summarized into pathway activity scores. **(c)** Pathway activity profiles are compared between disease and control subjects to discover dysregulated pathways. **(d)** Dysregulated pathways are mapped to the canonical pathway co-expression network (PCxN), where nodes denote pathways and the edges denote correlation of activity scores (36). **(e)** Within the network of dysregulated pathways, modules of coordinated pathways are identified using graph clustering algorithms **(f)** miRNAs are prioritized using annotated miRNA-mRNA interactions (known or predicted) for preferential targeting within each cluster of dysregulated pathways. The outputs of the pipeline are individuals lists of miRNAs with prioritization scores (targeting p-values) per each cluster of pathways.



$$S_x^c = \frac{1}{|C|}\sum_{y \in C} S_y$$

Figure 2. miRNA prioritization from pathway clusters. (a) PanomiR generates an observed targeting statistic ( ) for a miRNA X with respect to *C*, an observed cluster of pathways. The cluster-targeting statistic is an average individual overlap score for each miRNA-pathway pair. Individual overlap scores (e.g., S_1, S_2) are functions (inverse normal) of the overlap statistic (Fisher's exact test) between the miRNA target genes and the pathway member genes **(b)** PanomiR generates an empirical distribution of cluster-targeting scores for a miRNA X by randomly selecting a set of pathways and recalculating the cluster targeting score. **(c)** the prioritization p-value is calculated from comparing the observed to the null distribution of targeting score for the miRNA X. The p-value is used to rank the miRNAs.

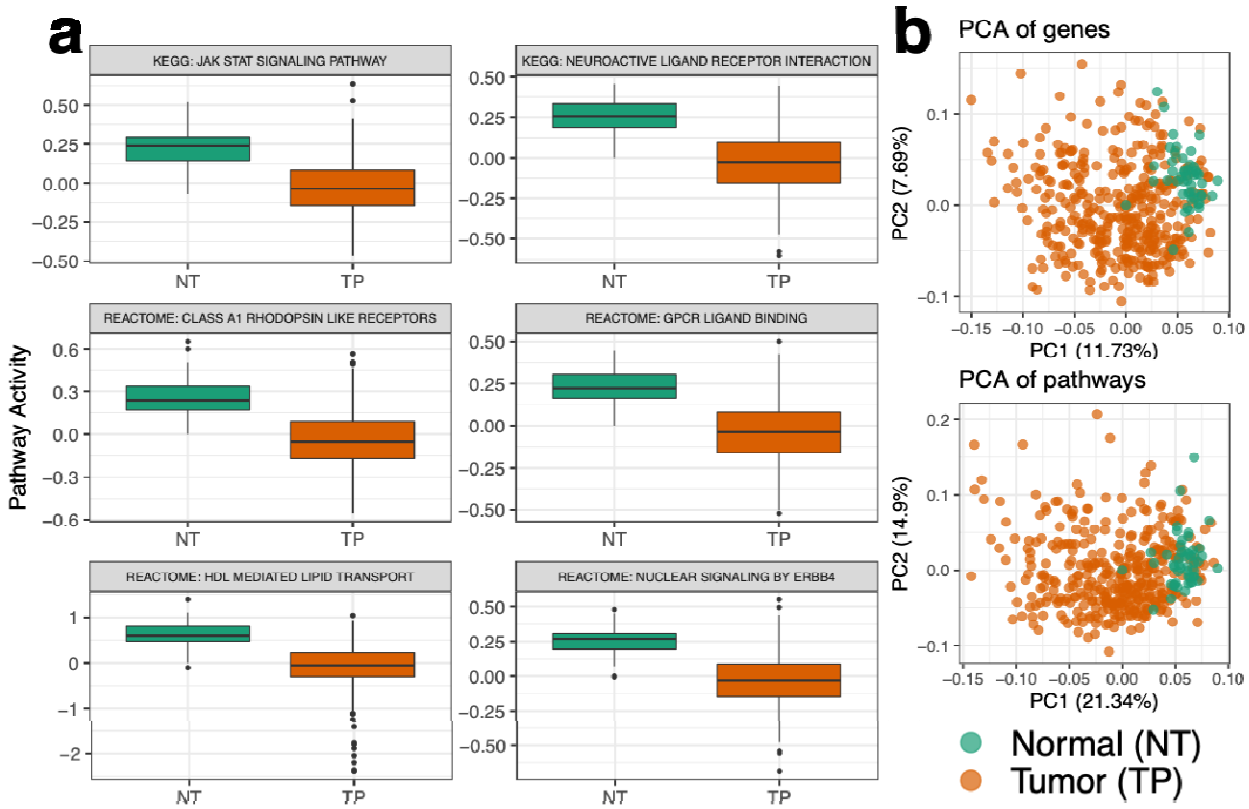


Figure 3. Pathway activity analysis of the Liver Hepatocellular Carcinoma dataset from the Cancer Genome Atlas. (a) Detection of dysregulated liver cancer pathways by comparison of pathway activity profiles between normal tissues (NT) and Tumor primary (TP) samples (39). Boxplots show the most significant dysregulated pathways selected based on p-values of difference between NT and TP (Table 1). (**b**) Principal component analysis (PCA) projection of the samples based on either genes or pathways. Pathway summarization in PanomiR allows to analyze the activity of pathways in a continuum. PCA of pathways conserves sample groups and captures a higher variation compared to the PCA of genes.

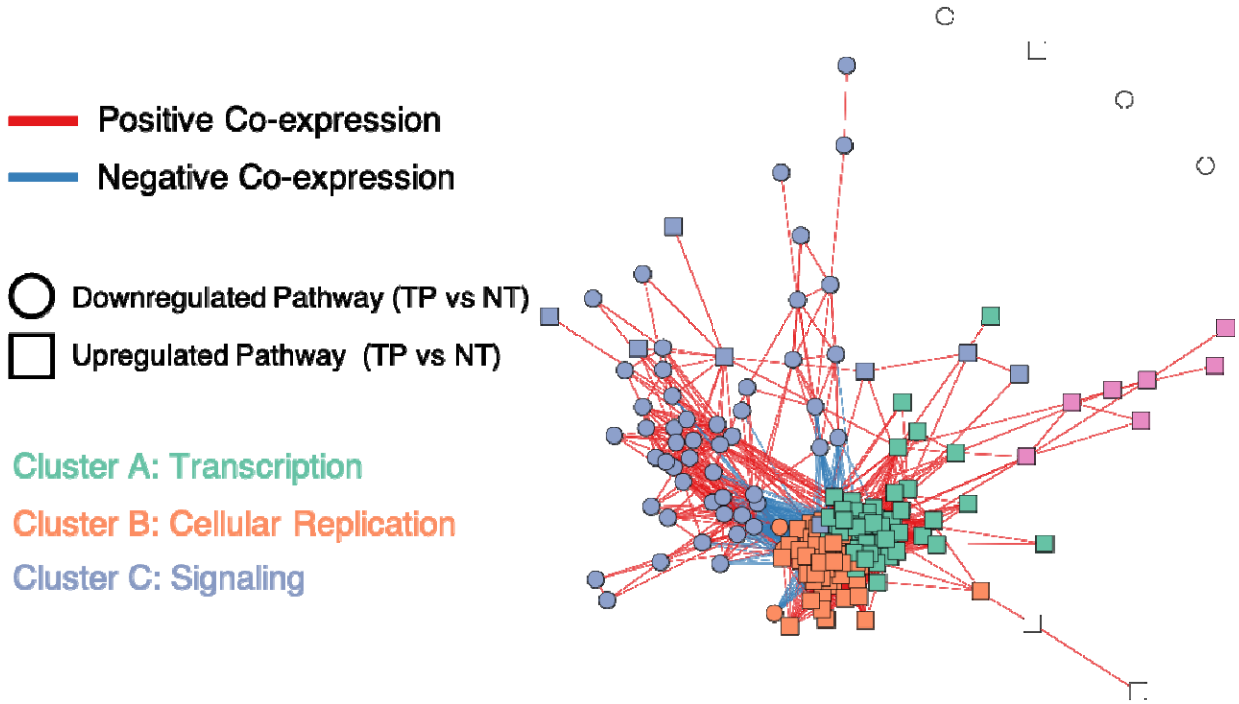


Figure 4. PanomiR deconvolutes coordinate clusters of dysregulated pathways in liver cancer. The network displays a pathway co-expression map of dysregulated liver cancer pathways. PanomiR detected three major groups of pathways, defined by direction of dysregulation and clusters of co-expression. The three classes are (i) activation of transcription in tumors (Cluster A) (ii) activation of cellular replication (Cluster B) (iii) deactivation of specific signaling pathways (Cluster C). Each node in the network represents a dysregulated pathway (Table 1). Edges represent canonical co-expression between two pathways, obtained from an independent compendium of gene expression data, as described in the PCxN method (36). Node colors represent unsupervised network clusters found by Louvain algorithm (73). Clusters were manually labeled according to the functional consensus of their pathways.

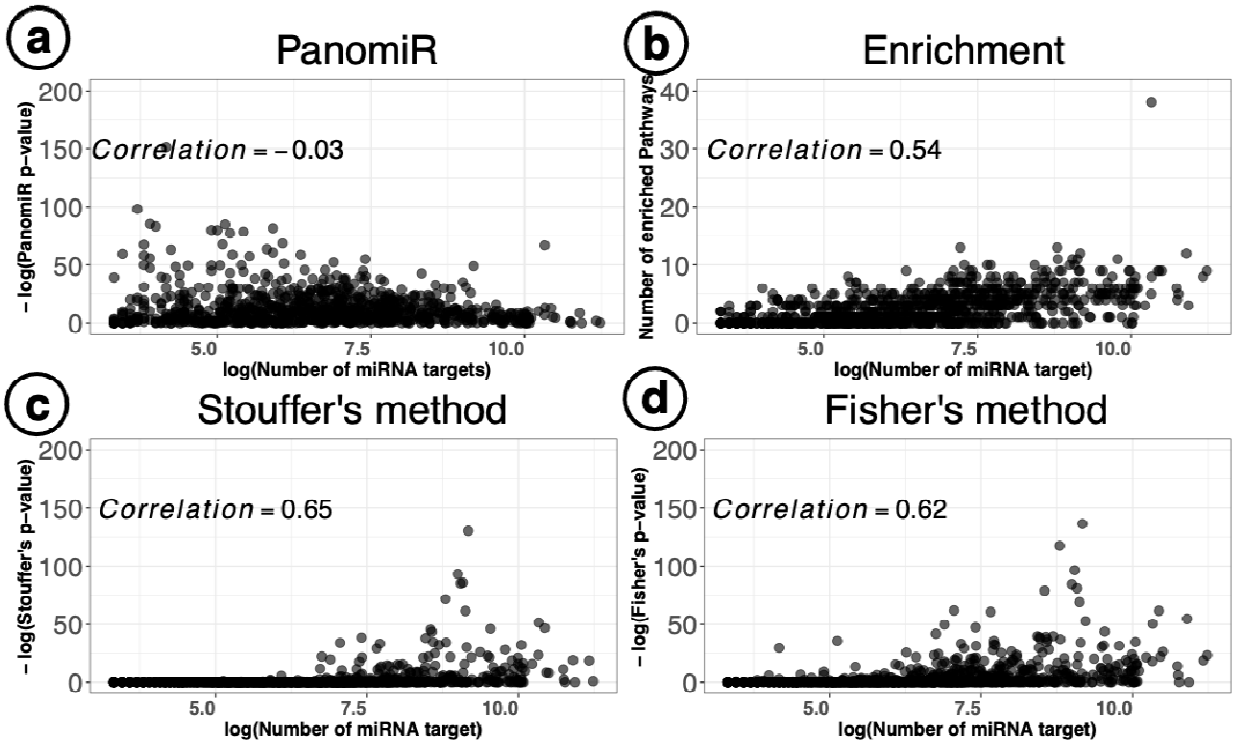


Figure 5. Unbiased prioritization of miRNAs by PanomiR. PanomiR prioritizes miRNAs with either a small or large number of annotated targets. In contrast, enrichment-based miRNA-prioritization methods are biased towards prioritization of miRNAs with larger numbers of gene targets. The figure displays correlation analysis of miRNA-prioritization rankings with the number of gene targets in Cluster A of the liver cancer dataset. Each point represents a miRNA annotated in the Tarbase dataset. (a) Spearman correlation analysis did not find a significant association between the number of targets and the prioritization ranking of miRNAs by PanomiR (correlation -0.03). (b) The number of enriched pathways for a miRNA significantly correlated with its number of gene targets. We also observed a significant correlation between the number of a miRNA's targets and its prioritization ranking based on (c) Stouffer's method and (d) Fisher's method for aggregation of enrichment p-value. X-axis denotes the log number of gene targets of miRNAs based on experimentally-validated miRNA-mRNA interactions from the TarBase database (38). Y-axes in panel b represents the number of significantly enriched pathways (Adjusted p-value < 0.25, Table 2).

TABLES

Table 1. Detection of dysregulated pathways in liver cancer. Most significant differentially activated pathways identified by PanomiR according to p-value of differential activity between tumor (TP) and normal tissues (NT). Dysregulation p-values were derived using linear models using the *limma* package by comparing pathway activity profiles of TP vs NT (45). Differential Activation adjusted P-values are for multiple hypothesis testing using FDR. Direction denotes upregulation or downregulation of pathway activity in TP vs NT. Enrichment adjusted p-values for pathways are provided for comparison. Enrichment p-values were derived from the differentially expressed genes (|FC| >1, FDR <0.05). The column "#DE genes" shows the number of differentially expressed genes (TP vs NT) that are incident in the pathway.

Dysregulated Pathways	Adjusted p-value	Direction TP vs NT	Enrichment adj. p-value	#DE genes
REACTOME: NUCLEAR SIGNALING BY ERBB4	1.13E-30	DOWN	1	5
KEGG: NEUROACTIVE LIGAND RECEPTOR INTERACTION	1.13E-30	DOWN	0.00109	27
KEGG: JAK STAT SIGNALING PATHWAY	3.19E-28	DOWN	1	12
REACTOME: CLASS A1 RHODOPSIN LIKE RECEPTORS	2.98E-27	DOWN	0.0114	28
REACTOME: GPCR LIGAND BINDING	4.2E-27	DOWN	0.0265	35
REACTOME: HDL MEDIATED LIPID TRANSPORT	7.57E-26	DOWN	0.53	4
PID: TCR CALCIUM PATHWAY	1.29E-25	DOWN	1	3
BIOCARTA: GATA3 PATHWAY	1.25E-24	DOWN	1	1
REACTOME: ASSEMBLY OF THE PRE REPLICATIVE COMPLEX	2.19E-24	UP	0.849	11
REACTOME: ORC1 REMOVAL FROM CHROMATIN	1.03E-23	UP	1	9
KEGG: TRYPTOPHAN METABOLISM	1.37E-23	DOWN	0.000547	15
BIOCARTA: ACTINY PATHWAY	4.77E-23	UP	1	0
REACTOME: PROTEIN FOLDING	5.59E-23	UP	1	0
KEGG: CYTOKINE-CYTOKINE RECEPTOR INTERACTION	1.71E-22	DOWN	0.0543	33
PID: ARF6 PATHWAY	2.55E-22	DOWN	1	5
BIOCARTA: IL1R PATHWAY	3.87E-22	DOWN	1	3
REACTOME: REGULATION OF INSULIN LIKE GROWTH FACTOR IGF ACTIVITY BY INSULIN LIKE GROWTH FACTOR BINDING PROTEINS IGFBPS	5.02E-22	DOWN	0.0474	6
REACTOME: SIGNALING BY GPCR	1.28E-21	DOWN	0.447	52
REACTOME: PREFOLDIN MEDIATED TRANSFER OF SUBSTRATE TO CCT TRIC	2.63E-21	UP	1	0
REACTOME: LIPOPROTEIN METABOLISM	2.64E-21	DOWN	0.0521	9
PID: IL1 PATHWAY	3.2E-21	DOWN	1	3
REACTOME: M G1 TRANSITION	3.92E-21	UP	0.447	15
BIOCARTA: TOLL PATHWAY	4.12E-21	DOWN	1	2
KEGG: UBIQUITIN MEDIATED PROTEOLYSIS	7.43E-21	UP	1	4
REACTOME: MRNA SPLICING MINOR PATHWAY	8.3E-21	UP	1	0

Table 2. PanomiR prioritizes regulatory miRNAs in liver cancer using experimentally-validated interactions. Prioritized miRNAs for each identified pathway cluster, ranked by PanomiR targeting p-value (Figure 2). miRNAs are prioritized based on experimentally validated miRNA-mRNA interaction from TarBase V8.0 (38). Enrichment analysis results are provided for comparison. The column "#Pathways enriched" denotes the number of pathways in the cluster with significant (FDR < 0.25) enrichment in the targets of each miRNA, derived using Fisher's Exact test.

Cluster A (n = 65)			Cluster B (n = 60)			Cluster C (n = 58)		
miRNA	**# Pathways Enriched**	**PanomiR Adjusted p-value**	**miRNA**	**# Pathways Enriched**	**PanomiR Adjusted p-value**	**miRNA**	**# Pathways Enriched**	**PanomiR Adjusted p-value**
hsa-miR-525-3p	1	2.7E-43	hsa-miR-107	40	1.92E-22	hsa-miR-410-3p	1	2.25E-07
hsa-miR-1307-5p	0	1.4E-27	hsa-miR-124-3p	38	2.04E-22	hsa-miR-552-3p	4	0.00057
hsa-miR-302c-5p	0	6.22E-24	hsa-miR-103a-3p	40	3.24E-21	hsa-miR-5187-5p	1	0.00057
hsa-miR-631	5	6.61E-24	hsa-miR-129-2-3p	37	2.31E-17	hsa-miR-612	2	0.00254
hsa-miR-663a	6	1.64E-23	hsa-miR-1-3p	36	2.43E-17	hsa-miR-198	7	0.0142
hsa-miR-595	4	4.03E-23	hsa-miR-23a-5p	3	8.12E-16	hsa-miR-621	2	0.0172
hsa-miR-933	1	1.04E-22	hsa-miR-663a	5	3.29E-15	hsa-miR-199b-5p	4	0.0209
hsa-miR-510-5p	1	1.04E-22	hsa-miR-449b-5p	33	2.26E-14	hsa-miR-204-3p	1	0.0249
hsa-miR-5009-5p	7	2.23E-22	hsa-miR-147a	41	2.26E-14	hsa-miR-4733-5p	1	0.0503
hsa-miR-2682-5p	0	4.65E-22	hsa-miR-193b-3p	35	3.26E-14	hsa-miR-506-3p	2	0.0503

Table 3. PanomiR prioritizes regulatory miRNAs in liver cancer using predicted interactions. Prioritized miRNAs for each identified pathway cluster, ranked by PanomiR targeting p-value (Figure 2). miRNAs are prioritized based on predicted miRNA:mRNA interaction from TargetScan V7.2 (37). The column "# Pathways enriched" denotes the number of pathways in the cluster with significant (FDR < 0.25) enrichment in the targets of each miRNA, derived using Fisher's Exact test.

Cluster A			Cluster B			Cluster C		
miRNA	**Pathways Enriched**	**PanomiR Adjusted p-value**	**miRNA**	**Pathways Enriched**	**PanomiR Adjusted p-value**	**miRNA**	**Pathways Enriched**	**PanomiR Adjusted p-value**
hsa-miR-371a-5p	0	1.06E-38	hsa-miR-191-5p	2	2.53E-19	hsa-miR-219a-2-3p	3	1.68E-09
hsa-miR-505-3p.2	0	8.96E-34	hsa-miR-892c-3p/hsa-miR-452-5p	1	8.81E-14	hsa-miR-376c-3p	4	9.51E-09
hsa-miR-1298-5p	0	1.88E-31	hsa-miR-4424	4	4.26E-12	hsa-miR-1249-3p	0	4E-08
hsa-miR-556-5p	0	3.4E-28	hsa-miR-339-5p	3	8.51E-11	hsa-miR-3605-3p	3	1.2E-07
hsa-miR-325-3p	3	5.95E-25	hsa-miR-944	0	8.51E-11	hsa-miR-143-3p	3	2.06E-07
hsa-miR-495-3p	0	1.05E-23	hsa-miR-345-5p	1	2.03E-10	hsa-miR-514b-5p/hsa-miR-513c-5p	1	6.57E-07
hsa-miR-1278	1	1.05E-23	hsa-miR-518c-3p	0	8.89E-10	hsa-miR-187-3p	3	1.39E-06
hsa-miR-651-5p	0	1.46E-22	hsa-miR-154-5p	2	3.17E-09	hsa-miR-625-3p	1	3.06E-06
hsa-miR-323b-3p	0	6.39E-18	hsa-miR-1251-5p	1	3.47E-09	hsa-miR-1306-5p	2	5.83E-06
hsa-miR-421	0	1.16E-17	hsa-miR-599	1	3.47E-09	hsa-miR-873-5p.1	3	7.82E-06