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[1,7], Frank J. Slack[1,2,3], Ioannis S. Vlachos[1,2,3,6], and Winston A. Hide[1,2,3*]

1 Harvard Medical School, Boston, MA, USA
2 Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA
3 Harvard Medical School Initiative for RNA Medicine, Boston, MA, UA
4 National University of Singapore, Singapore
5 École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
6 Broad Institute of MIT and Harvard, Cambridge, MA, USA
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:

Charting microRNA (miRNA) regulation across pathways is central to characterizing their role in disease. Yet, current methods reveal only individual miRNA-pathway interactions. We have developed a systems biology approach, *Pathway networks of miRNA Regulation* (PanomiR), that overcomes these limitations to identify miRNA targeting of groups of interacting pathways using gene expression. The approach does not depend on statistically significant enrichment of miRNA target genes in individual pathways or significant differentially expressed genes. Rather, it directly captures differential activity of pathways between states, determining their up-or-down regulation while sensitively detecting biologically-meaningful signals. PanomiR analyzes the co-activity of differentially regulated pathways to determine coordinate functional groups and uses these co-activated grouped pathways to uncover miRNAs that target them. Incorporating both experimentally-supported or predicted miRNA-mRNA interactions, PanomiR robustly identifies miRNAs central to the regulation of disease functions. We applied PanomiR to a liver cancer dataset and showed that it can organize liver cancer pathways and their regulating miRNAs into coordinated transcriptional programs, reflecting the pathogenic mechanisms of hepatocellular carcinoma. PanomiR recapitulated known central miRNAs in liver cancer with a biologically meaningful assignment of pathways under their regulation, unbiased by the number of genes targeted by each miRNA. PanomiR is a granular framework for detecting broad-scale multi-pathway programs under miRNA regulation. It 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 that act as potent regulators of cellular functions and molecular pathways (1). They post-transcriptionally regulate gene expression and can coordinate gene 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 Disease (2–20). Because miRNAs can target dozens of genes, the characterization of their roles in health and disease requires charting of coordinate co-regulation across heterogeneous molecular cascades and pathways. Despite extensive progress in the field to map the effects of miRNAs on one or more pathway activities (4, 21–27) or the effect of pathway activity on a miRNA, no framework exists for characterization and prioritization of the multi-pathway dynamics of miRNA-orchestrated regulation that form driving transcriptional programs in both healthy and diseased states.

Current best practice for the transcriptomic study of miRNA regulation relies on miRNA-gene or one-to-one miRNA-pathway relationships. miRNA-pathway analysis techniques such as gene set enrichment and correlation are used to detect whether a pathway is potentially regulated by a miRNA (4, 21). Enrichment analyses evaluate the presence (overlap) of targets of a single miRNA in a single pathway, aiming to identify pathways with a higher number of targets than expected by chance (21, 22, 28–32). Alternatively, correlation methods evaluate the association of the expression of a single miRNA with a gene representing the activity of a pathway (4, 33). Table 1 describes some of the most widely used methods for miRNA pathway analyses, their scope and approach. Large-scale functional processes in health and disease coordinate across pathways in multiple ways, including gene-sharing, pathway co-activity, multi-pathway co-regulation, and cross-talk (34–39). 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 broadly-acting gene expression programs.

To uncover how multiple pathways are coordinated by miRNAs to form gene expression programs, we have developed a framework to address existing limitations from a systems perspective. *Pathway networks of miRNA Regulation* (PanomiR) enables discovering central miRNA regulators based on their ability to control coordinate pathways forming a transcriptional program. PanomiR 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. PanomiR derives these multi-pathway targeting events using predefined pathways, their co-activation, gene expression, and annotated miRNA-mRNA interactions. Its framework (i) captures the activity of pathways and identifies disease-specific differentially regulated pathways using pathway activity profiling, a technique that accounts for overall co-activity of genes and commonly observed biases (40–42); (ii) constructs a co-expression network of differentially regulated pathways (based on a reference of pathway co-expression networks) and deconvolves it into coordinate groups of pathways that act in concert using network clustering algorithms (40–42); (iii) determines miRNAs targeting these coordinate pathway groups using a novel statistical test and pre-determined miRNA-mRNA interactions from experimentally-supported or prediction databases (43, 44). Taken together, these steps produce broad-scale, multi-pathway, and disease-specific miRNA regulatory events (Figure 1).

In order to highlight PanomiR's ability to detect miRNAs regulating gene expression programs in human disease, we applied it to the hepatocellular carcinoma dataset of The Cancer Genome Atlas (TCGA) comprising 368 primary tumor samples and 49 controls(45).

PanomiR unbiasedly detected multi-pathway miRNA regulation events in this well-documented disease and generated a comprehensive framework for evaluating miRNA-associated mechanisms (46). We identified differentially regulated pathways and uncovered their regulating miRNAs in liver cancer, assessed biological relevance of the readouts, and evaluated the statistical robustness of PanomiR using *in silico* experimentation. PanomiR recapitulated 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. PanomiR is available as an easy-to-use Bioconductor R package, enabling its application in research projects, inclusion into *in silico* tools, and augmentation of analysis pipelines.

MATERIALS AND METHODS

Overview and input datasets

The overarching goal of PanomiR is to detect miRNAs regulating multi-pathway condition-associated gene expression programs (Figure 1). PanomiR uses as input a user-provided gene expression dataset (e.g., RNAseq) to quantify pathway activity profiles by utilizing annotated pathway datasets from the Molecular Signatures Database (MSigDB) (Figure 1A-1B)(47). Pathway activity profiles are then compared between two conditions (e.g., cancer *vs* control, wild type *vs* knockout) to identify and prioritize differentially regulated pathways (Figure 1C). To determine broad-scale condition-associated groups of functions, PanomiR constructs a co-activity network of differentially regulated (or disease dysregulated) pathways and deconvolves the network into coherent functional groups using reference pathway co-expression networks; using our previously-described pathway activity methods (Figure 1D-1E) (40–42). Subsequently, PanomiR integrates miRNA-mRNA interactions provided by the user (such as predicted targets from TargetScan (43) or experimentally validated interactions from TarBase (44)) to evaluate miRNA regulatory effects on coordinate pathway groups (Figure 1F). The final output of PanomiR is a ranked list of central miRNAs, together with statistical significance levels for each group of differentially regulated pathways, providing an effective means for identification of pathway groups, and for key miRNA prioritization, ranking, and target detection. PanomiR identifies differentially coordinated transcriptional programs between two conditions to provide a direct prioritization of the miRNAs responsible for their coordination.

Capturing pathway activity dynamics

Extending the approach developed in our previous methodology, Pathprint (40–42), PanomiR ingests a user-provided gene expression dataset and calculates pathway activity scores to capture pathway 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. By capturing gene expression levels as pathway activity scores, inherent complexity is reduced while tolerance to noise is increased when compared to gene-centric analyses (4, 41, 42, 48). Pathway activity scores leverage the complex inter-relationships and co-activity of genes. They provide the means to examine biological functions in a continuum and detect biological signals where standard differential gene expression analyses fail (4, 33, 40, 42, 48–50).

To capture pathway activity profiles, in a two-step process: (a) we rank genes in each sample in descending order, according to their expression, i.e., 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. Formally, 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 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 MSigDB as its pathway database reference (47). MSigDB is a carefully curated database that represents non-redundant pathways from established pathway repositories such as KEGG and Reactome (47, 51, 52).

Detection of differentially regulated disease-associated pathways

PanomiR compares pathway activity profiles between case and control subjects to determine functional dynamics in disease. PanomiR defines differentially regulated pathways by determining statistically significant differences in pathway activity profiles between cases and controls using linear models, implemented in the *Limma* package (Figure 1C) (53). In contrast to enrichment analysis, the linear modeling framework of PanomiR determines the directionality of differential regulation: It defines whether a pathway is upregulated or downregulated in disease subjects (or experimental conditions), and accounts for confounding variables such as batch, sequencing center, or any other fixed effects and continuous covariates. PanomiR outputs an ordered table of differentially regulated pathways along with p-values of differential regulation, adjusted for multiple hypothesis testing using False Discovery Rate (FDR) (54).

Detection of groups of differentially regulated pathways via their co-expression networks

Dysregulation of an individual pathway is rarely an isolated event since pathways share activity and are often co-regulated. PanomiR accounts for co-regulation to place differentially regulated pathways into groups that represent high-level disease programs by exploiting the *Pathway co-expression network* (PCXN) (42): a reference tool that organizes and assesses the shared activity of pathways (Figure 1D). PanomiR leverages PCxN's network, generated from a curated dataset of 3,207 expression profiles, providing an independent platform, to query co-activity of all pathways in the MSigDB dataset (42, 55).

PanomiR masks PCxN to contain only the subnetwork of differentially regulated pathways that were identified from the two-group data analysis in the previous step. In the masked network, nodes represent differentially regulated pathways and edges activity-correlation of pathways. PanomiR subsequently identifies densely interconnected differentially regulated pathway subnetworks using graph clustering algorithms (Figure 1E). The default clustering algorithm of PanomiR is Louvain, but PanomiR can use other clustering methods that are available in the igraph R-package (56). The subnetworks denote clusters of highly correlated coordinate groups of differentially regulated pathways driving disease or condition-specific functions.

miRNA prioritization within clusters of differentially regulated pathways

PanomiR exploits the concept that a coordinate group of disease-associated pathways has common miRNA regulators. Using annotated miRNA-mRNA interactions and an empirical statistical test (Figure 2), it analyzes clusters of differentially regulated pathways, to define central miRNAs, and captures the extent to which the targets of a specific miRNA are present within a group of coordinate pathways. miRNA regulatory events are then identified in three sequential steps (Figure 2): (i) by calculating individual miRNA-pathway overlap scores, (ii) by generalizing miRNA targeting scores to a group of pathways (i.e., a cluster of differentially regulated pathways), and (iii) by estimating the statistical significance of miRNA targeting scores

using an empirical approach. The empirical statistical tests are specific to the input dataset, for each miRNA and each cluster of differentially regulated 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-, condition-, tissue-, or cell-type-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 clusters of pathways are generated in the previous step using PCxN. Formally, for each cluster of differentially regulated 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 $\Phi^{-1}(.)$ denotes the inverse of the standard normal cumulative distribution function (CDF) and P_{xy} denotes the Fisher's Exact test p-value of overlap between targets of miRNA x and genes of pathway y. The targeting-score, S_x^c , is related to Stouffer's method (with equal weights) for p-value aggregation. 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 in order to produce cluster-specific lists of miRNAs ranked by targeting p-values. The 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 also derives an empirical targeting p-value, $P(S_x^c)$, for a score of S_x^c. This p-value denotes the probability of observing a larger targeting score from a random cluster of pathways (with $|C|$ members) than the one observed. This empirical probability is derived using a bootstrap sampling approach by selecting randomized groups of pathways and re-calculating their cluster targeting score. This approach directly tackles known or unknown biases in gene annotations for miRNA targets, as have been discussed by our group (21) and others (57, 58). The output p-values are then adjusted for multiple hypothesis comparison using the Benjamini-Hochberg False Discovery Rate (FDR) (54).

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 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. We provide detailed assessments of the Gaussian estimation and robustness of $P(S_x^c)$-values using a jackknife estimation as Supplementary Materials (Supplementary Figures S3 and S4).

RESULTS

PanomiR's utility is presented in a case-study showing its ability to provide a systematic, unbiased, and biologically meaningful determination of regulatory miRNAs. We applied PanomiR to a liver cancer gene expression dataset from TCGA (59). Figures 3 and 4 portray PanomiR's recapitulation of liver cancer-associated pathways (Table 2), their coordination, and the miRNAs that target them. We found three clusters of differentially regulated 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 3 and 4). By comparing PanomiR's results with the relevant literature and with enrichment analysis, we show that PanomiR provides informative and novel 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 and using the MSigDB pathway database. PanomiR detected 428 upregulated and 397 downregulated pathways in TP compared to NT (FDR < 0.01, Total pathways 1220; Table 2 and Supplementary Table S1). The large-scale differences in pathway activity profiles closely mirror the differential expression results at the gene level: more than 50% of the genes were differentially expressed based on a similar statistical design (FDR <0.01; n = 7801; total genes= 14212).

Differentially regulated pathways reflected well-established dysregulated functions in liver cancer (Table 2). For example, *NUCLEAR SIGNALING BY ERBB4* was downregulated in TP and activated in NT and has the highest statistical significance among all pathways (Figure 3a, Table 2). 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 (60, 61). 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 (62, 63). These results suggest the utility of PanomiR in detecting differentially regulated disease functions through pathway activity analysis.

We compared pathway readouts from PanomiR with pathway enrichment analysis of differentially expressed genes from the same liver cancer dataset. Enrichment analysis using Fisher's Exact Test and comparable cut-offs identified 51 enriched pathways (FDR <0.01, Supplementary Table S2) from differentially expressed genes (differential gene expression: FDR < 0.05; |LogFC| > 1; supplementary material). Of these enriched pathways, 50 were also determined as differentially regulated 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). PanomiR also detected 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 misses pathways with significantly differential activity between pathways and controls but not increased DE gene proportions. Table 2 and Supplementary Table S1 show several other instances of pathways that were detected by PanomiR but were not identified in the standard enrichment analysis. These results highlight the ability of PanomiR to detect significant functional dysregulation in disease even in absence of significant differential gene expression (Table 2).

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 differentially regulated 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 differentially regulated pathways (0.054 ± 1.2, Mean ± SD) via randomized case/control sample assignment (Supplementary Table S3). In the second test, we interrogated

whether the use of biologically meaningful pathways (as annotated in the MSigDB) held advantages versus 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 differentially regulated pathways (one-sided Z test p-value < $3.34 * 10^{-5}$; mean = 693.785 pathways at an FDR < 0.01; sd = 39.1). We also compared the distribution of adjusted p-values from differentially regulated pathways from MSigDB and with that of randomized pathway collections, irrespective of FDR cut-offs. This experiment showed a significant difference between the two scenarios according to a 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 differentially regulated than randomized pathways and were more likely to capture biological signals.

Identification of coordinate clusters of differentially regulated pathways

Pathways coordinate and co-regulate through various mechanisms, including gene sharing. To detect coordinate groups of differentially regulated pathways, we used the Pathway Co-expression Network (PCxN), where edges represent precalculated correlations between pathways based on independent gene expression data (42). We mapped the 200 most statistically significant differentially regulated pathways onto the PCxN network and performed Louvain clustering to identify coordinate pathway groups. PanomiR identified 3 major clusters of differentially regulated pathways with consistent functions (Figure 4).

The largest cluster of differentially regulated 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 (64). 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 down or upregulated in TP vs NT with terms related to ERBB signaling, IL signaling, and NOTCH signaling (Supplementary Table 4). Differentially regulated pathways within clusters A and B showed a coherent direction of differential regulation in cancer vs normal tissues, suggesting a 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 differentially regulated pathways that represent higher-order functional programs of liver cancer.

Detection of regulatory miRNAs that target clusters of differentially regulated pathways

We evaluated whether the coordinate clusters of differentially regulated pathways have common miRNA regulators. In our case study, we examined separately experimentally supported (TarBase v8.0; >500K interactions) and predicted miRNA-mRNA interactions (Targetscan v7.2; >113K interactions) to detect miRNAs that target each cluster of differentially regulated pathways (Tables 3 and 4, Supplementary Tables S5 and S6). Our results showed that PanomiR identified distinct miRNAs for each cluster of liver cancer-associated pathways.

With the use of experimentally supported interactions, PanomiR detected 202, 104, and 1 miRNA regulators in clusters A, B, and C respectively (FDR < 10^{-5}, Table 3, Supplementary Table S5). 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 to have a role in tumor migration and invasion (65–68). 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 (69–72); miR-124-3p is a tumor suppressor that regulates proliferation and invasion in liver cancer by inducing G1-phase cell-cycle arrest (73, 74); miR-103a-3p is a promoter of proliferation that is highly dysregulated in liver cancer (75). 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 (76, 77), it also regulates adenomas via signaling pathways such as MAPK, PTEN/AKT, and STAT (78, 79). In cluster C, we also found a significant targeting role for miR-552-3p, which has been associated with liver cancer and regulates various hallmarks of cancer (80). 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 differentially regulated pathways.

PanomiR was also assessed using predicted miRNA-mRNA interactions (43). Although PanomiR detected multiple liver cancer-associated miRNAs from predicted interactions, the set of prioritized miRNAs were different than that of experimentally supported interactions (Table 4, Supplementary Table S6). 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 (81). Supplementary tables and results provide information on processing predicted interactions and additional evaluations of PanomiR using varying parameters for selection of predicted interactions. Our results suggest that predicted and experimentally validated miRNA interactions databases produce complementary results, and both should be considered for the downstream analysis of transcriptomic data.

We compared PanomiR's results with a standard miRNA-pathway enrichment analysis in our case study (Tables 3 and 4). For comparative purposes, we employed three extensions of miRNA-enrichment analysis tests to adapt to multi-pathway scenarios. (a) We initially extended enrichment analysis to a group (cluster) of pathways by interrogating the number of pathways within a given cluster that 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 revealed miR-525-3p as enriched in only 1 and miR-1307-5p in none out of 65 pathways (Table 3). When using predicted miRNA-mRNA interactions, PanomiR detected several miRNAs that were not detected by the extended enrichment analysis (Table 4). It is of note that the enrichment tests (Tables 3 and 4) 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 retained 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 are biased towards detecting miRNAs with a larger number of targets (80). We examined the relationship of the number of targets of a miRNA with its prioritization ranking by PanomiR or enrichment analysis extensions (Table 4). The enrichment ranking of miRNAs significantly correlated with their number of gene targets while PanomiR was unbiased to this number. 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. Additional evaluation of unbiased and robust miRNA prioritization by PanomiR is provided in the Supplementary material. Using jackknifing and bootstrapping, we showed that PanomiR miRNA prioritization is rather based on collective targeting of all pathways and is not driven by individual pathways

DISCUSSION

We have built PanomiR, a framework able 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 use of p-value integration methods to address multi-pathway analysis, but none of them determine pathway co-activity/coordination and account for expression dynamics (82). PanomiR addresses these challenges by deconvolving gene expression datasets into coordinate groups of pathways with condition-associated dynamics and by measuring the extent to which miRNAs target these groups. In the case study of 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; It sensitively detects differentially regulated pathways and provides granularity in definition of coordinate functional groups (Figures 2 and 3, Table 2). In a case study, PanomiR detected critical known liver cancer pathways even with few associated differentially expressed genes (Table 2). 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 as they may use nonlinear dimensionality reduction approaches (3). PanomiR fills this gap by providing biologically meaningful measurements of changes in pathway activity profiles where 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, platforms, and species (40).

Deconvolution of coordinate pathway groups allowed PanomiR to detect miRNA regulatory events in liver cancer robustly and unbiasedly, many of which were not detectable by conventional analyses (Figure 4, Tables 3 and 4). PanomiR's prioritized miRNAs have distinct roles in liver cancer, concordant with the functional characteristic of the pathway clusters 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 3). 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, gene expression programs of liver cancer.

The complete landscape of miRNA-mRNA binding events is currently unknown. This gap contributes to the discrepancy in miRNA prioritization based on the background dataset of miRNA-mRNA interactions (Tables 3 and 4). By using 113K high-confidence predicted miRNA interactions (TargetScan) and more than 500K experimentally supported (TarBase) miRNA targets, PanomiR discovered informative and complementary miRNA regulatory events (Tables 3 and 4). Users have the ability to tailor the background datasets (miRNA-mRNA integration or pathway gene sets) 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 differentially regulated pathways, their co-activity, and their regulating miRNAs. It accounts for co-expression of pathways and disease-specific expression dynamics to identify miRNA-regulatory events, providing an advance over the current practice of studying static and isolated miRNA-pathway interactions. PanomiR is 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>.

FUNDING

This work was supported by the Harvard Medical School Aging Initiative Pilot Grant (WAH, ISV, FJS) and R01CA258776 (ISV). FJS was supported by grants from the US National Institute of Health (R35 CA232105; R01 AG058816-01). This study was partially supported by the Cure Alzheimer's Foundation (SM and WAH)

ACKNOWLEDGEMENTS

We would like to thank Dr. Isabel Castanho and Ms. Saloni Shah for valuable feedback. We thank the Cure Alzheimer's Foundation and Harvard Medical School for their sponsorship.

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. Slack,F.J. and Chinnaiyan,A.M. (2019) The role of non-coding RNAs in oncology. *Cell*, **179**, 1033–1055.

3. 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.

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

5. 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.

6. 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.

7. 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.

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

9. 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.

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

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

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

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

14. 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.

15. 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.

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

17. 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.

18. 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.

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

20. 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.

21. 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.

22. 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.

23. Kehl,T., Kern,F., Backes,C., Fehlmann,T., Stöckel,D., Meese,E., Lenhof,H.-P. and Keller,A. (2020) miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database. *Nucleic Acids Res.*, **48**, D142–D147.

24. Alaimo,S., Giugno,R., Acunzo,M., Veneziano,D., Ferro,A. and Pulvirenti,A. (2016) Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. *Oncotarget*, **7**, 54572–54582.

25. Hsu,J.B.-K., Chiu,C.-M., Hsu,S.-D., Huang,W.-Y., Chien,C.-H., Lee,T.-Y. and Huang,H.-D. (2011) miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics*, **12**, 300.

26. Zagganas,K., Vergoulis,T., Paraskevopoulou,M.D., Vlachos,I.S., Skiadopoulos,S. and Dalamagas,T. (2017) BUFET: boosting the unbiased miRNA functional enrichment analysis using bitsets. *BMC Bioinformatics*, **18**, 399.

27. Preusse,M., Theis,F.J. and Mueller,N.S. (2016) MiTALOS v2: Analyzing tissue specific microRNA function. *PLoS One*, **11**, e0151771.

28. 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**.

29. 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.

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

31. 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.

32. 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.

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

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

35. 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.

36. 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.

37. 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.

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

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

40. 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.

41. 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.

42. 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.

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

44. 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.

45. 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.

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

47. 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.

48. 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.

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

50. 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.

51. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.

52. Jassal,B., Matthews,L., Viteri,G., Gong,C., Lorente,P., Fabregat,A., Sidiropoulos,K., Cook,J., Gillespie,M., Haw,R., *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.

53. 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.

54. 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.

55. 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.

56. Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *InterJournal, complex systems*, **1695**, 1–9.

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

58. Godard,P. and van Eyll,J. (2015) Pathway analysis from lists of microRNAs: common pitfalls and alternative strategy. *Nucleic Acids Res.*, **43**, 3490–3497.

59. 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.

60. 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.

61. 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.

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

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

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

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

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

74. 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.

75. 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.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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

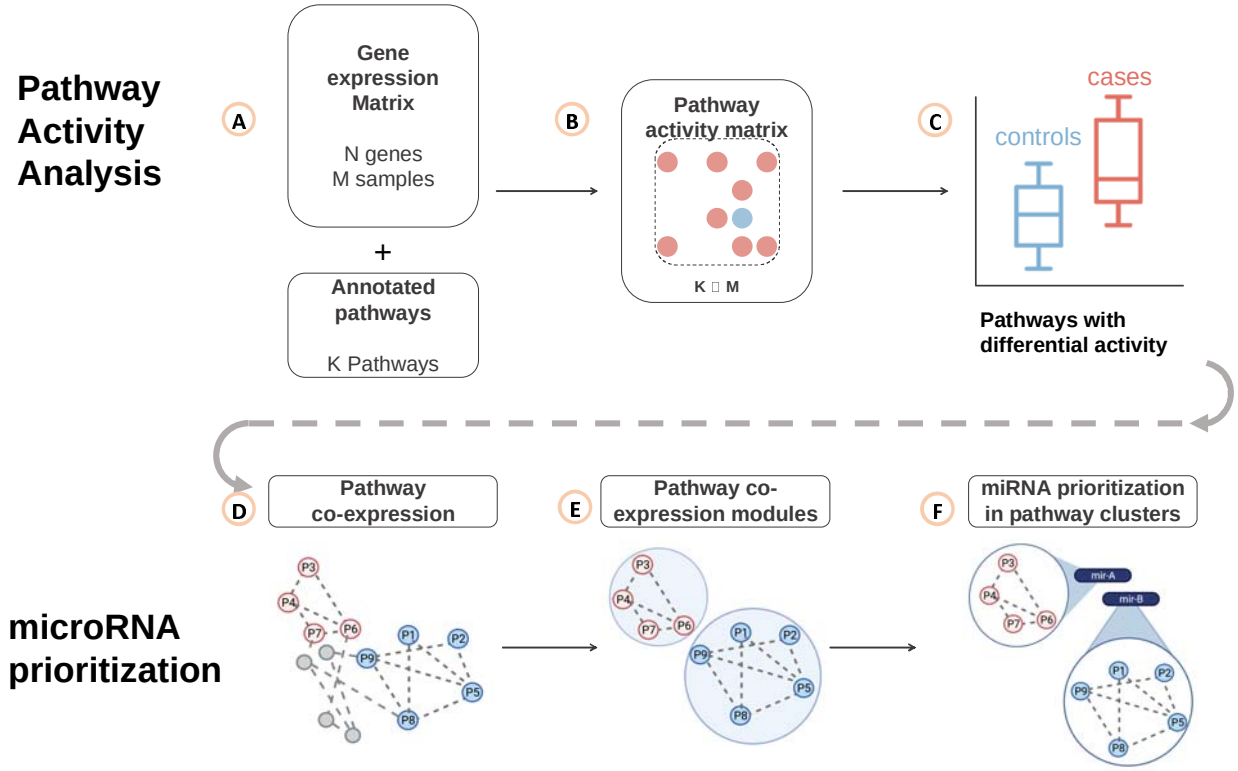

Figure 1. PanomiR workflow. PanomiR prioritizes miRNAs that target coordinate groups of pathways. **(A)** Input gene expression dataset and a set of annotated pathways **(B)** Gene expression data is summarized into pathway activity scores. **(C)** Pathway activity profiles are compared between disease and control subjects to discover differentially regulated pathways. **(D)** Differentially regulated pathways are mapped to the canonical pathway co-expression network (PCxN), where nodes denote pathways and the edges denote correlation of activity scores. **(E)** Within the network of differentially regulated pathways, modules of coordinate 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 differentially regulated pathways. The outputs of the pipeline are individuals lists of miRNAs with prioritization scores (targeting p-values) per each cluster of pathways.

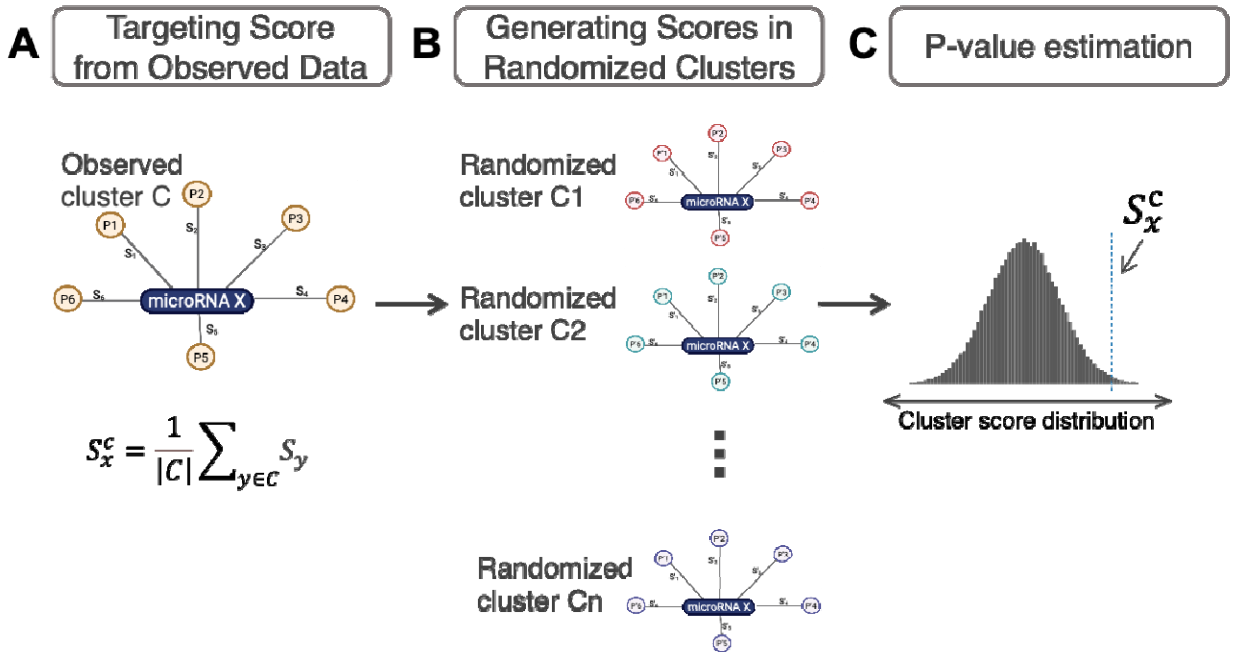

Figure 2. miRNA prioritization from pathway clusters. (A) PanomiR generates an observed targeting statistic, S_X^C , for a miRNA X with respect to *C*, an observed cluster of pathways. The cluster-targeting statistic is an average individual overlap score for each miRNA-pathway pair. Individual overlap 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 targeting statistic, S_X^C, to the null distribution of targeting score for the miRNA X. The p-value is used to rank the miRNAs.

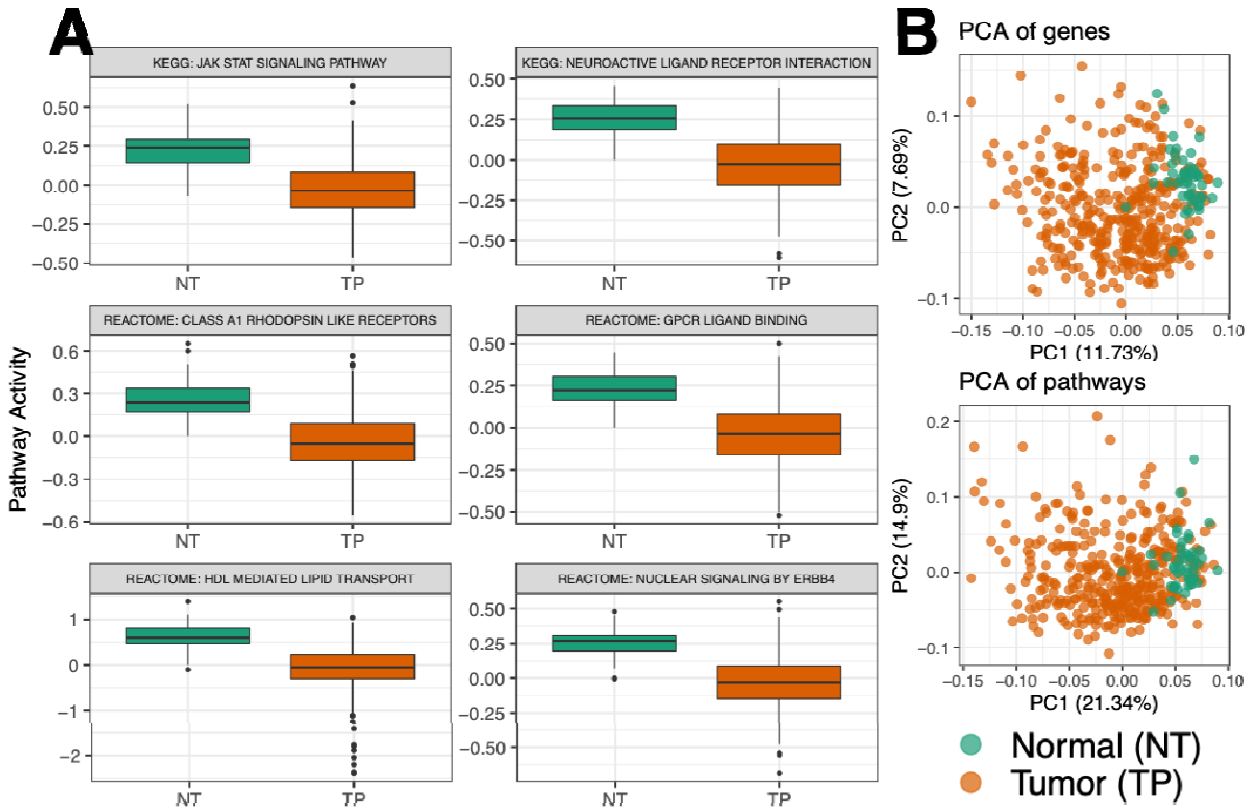

Figure 3. Pathway activity analysis of the Liver Hepatocellular Carcinoma dataset from the Cancer Genome Atlas. (A) Detection of differentially regulated liver cancer pathways by comparison of pathway activity profiles between normal tissues (NT) and Tumor primary (TP) samples (45). Boxplots show the most significant differentially regulated pathways selected based on p-values of difference between NT and TP (Table 2). **(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.

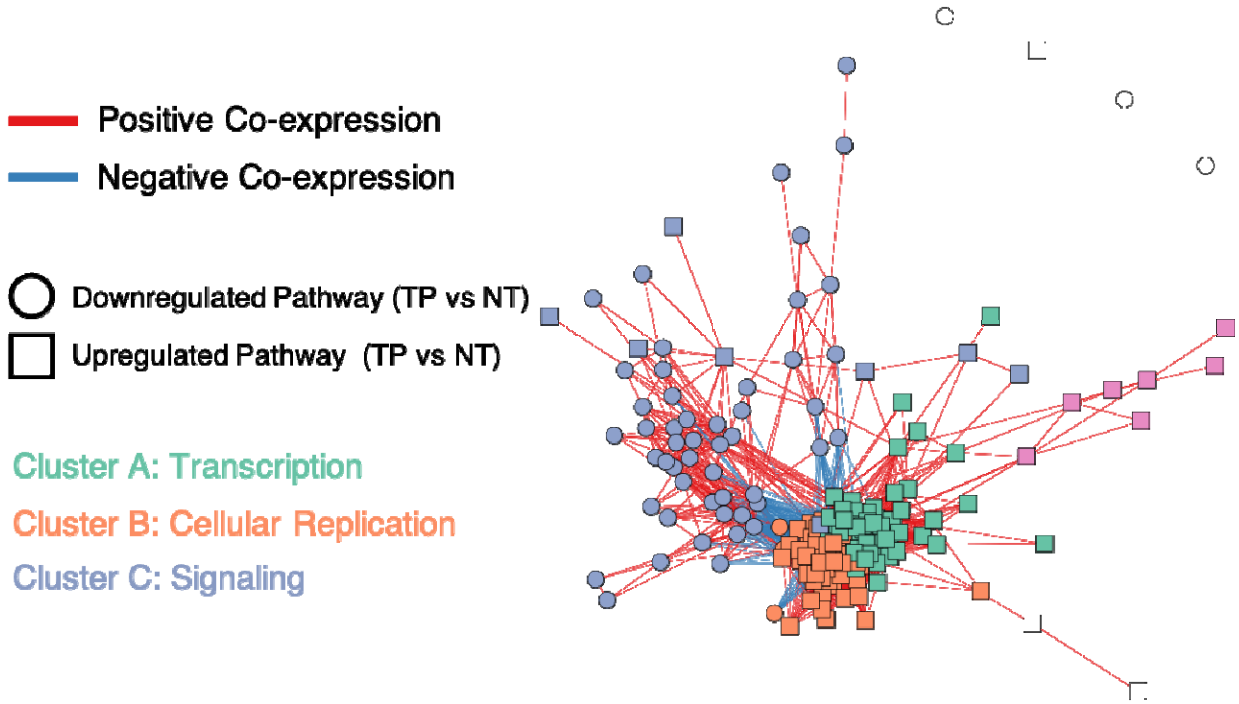

Figure 4. PanomiR deconvolutes coordinate clusters of differentially regulated pathways in liver cancer. The network displays a pathway co-expression map of liver cancer pathways. PanomiR detected three major groups of pathways, defined by direction of differential regulation and clusters of co-expression. 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 differentially regulated pathway (Table 2). Edges represent canonical co-expression between two pathways, obtained from an independent compendium of gene expression data, as described in the PCxN method (42). Node colors represent unsupervised network clusters found by Louvain algorithm (83). Clusters were manually labeled according to the functional consensus of their pathways.

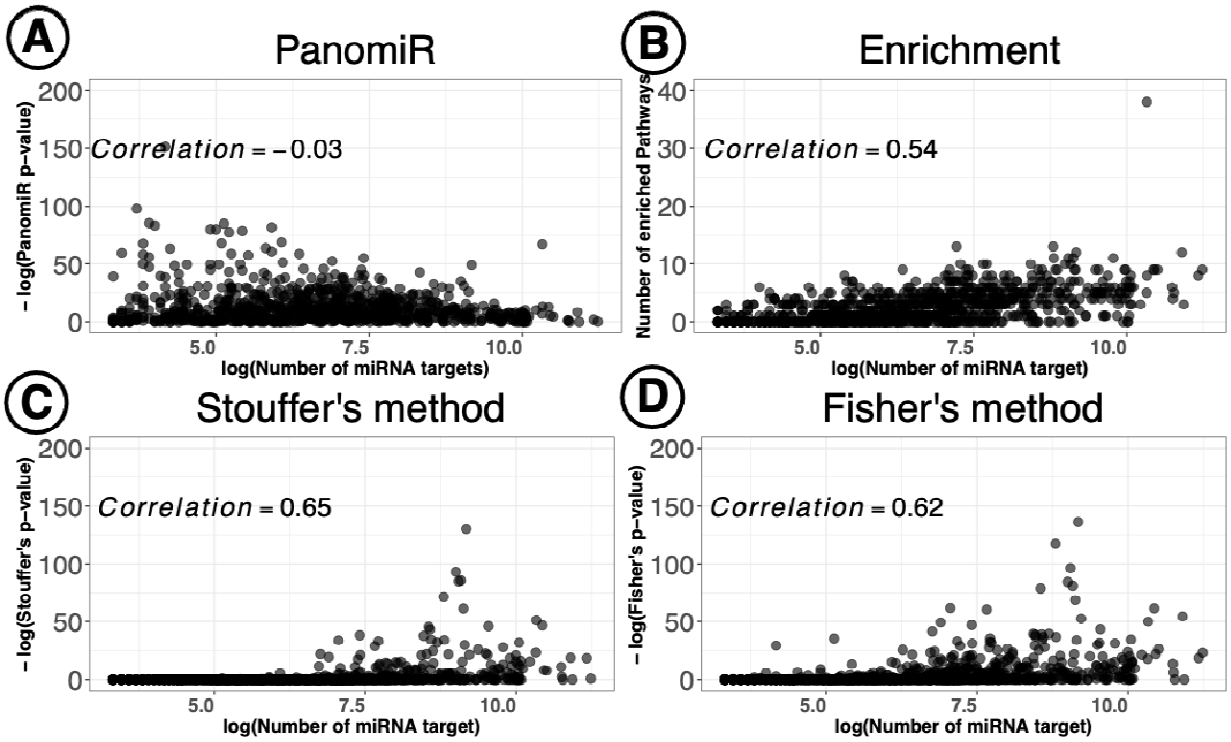

Figure 5. Unbiased prioritization of miRNAs by PanomiR. PanomiR prioritizes miRNAs with either a small or large number of annotated targets. In contrast, enrichment-based miRNA-prioritization methods are biased towards prioritization of miRNAs with larger numbers of gene targets. The figure displays 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 (44). Y-axes in panel b represents the number of significantly enriched pathways (Adjusted p-value < 0.25, Table 3).

TABLES

Table 1. Overview of standard miRNA-Pathway analysis methods and PanomiR.

Method/ Reference	Multi-pathway targeting	Pathway activity dynamics	Pathway coordination/ interaction	Open-source software	Tissue-specific customization
PanomiR This work	X	X	X	X	X
miRPath v3 [PMID 25977294]	X Meta-analysis				
Wilk and Braun [PMID 29294105]		X		X	X
miRPathDB 2 [PMID 31691816]				X precalculated	X precalculated
MITHrIL [PMID 27275538]		X (via DEG)			X
miRTar [PMID 21791068]				X Web portal non-functional	
BUFET [PMID 28874117]				X	X
miTalos [PMID 26998997]					X precalculated

Table 2. Detection of differentially regulated pathways in liver cancer. Most significant differentially regulated pathways identified by PanomiR according to p-value of differential activity between tumor (TP) and normal tissues (NT). Differential regulation p-values were derived using linear models using the *limma* package by comparing pathway activity profiles of TP vs NT (53). 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 present in the pathway.

Differentially Regulated 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 3. 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 (44). 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 4. 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 (43). 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