THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 20, pp. 18066–18078, May 20, 2011
Printed in the U.S.A.

Two-tiered Approach Identifies a Network of Cancer and Liver Disease-related Genes Regulated by miR-122*[S]

Received for publication, October 21, 2010, and in revised form, March 11, 2011 Published, JBC Papers in Press, March 14, 2011, DOI 10.1074/jbc.M110.196451

Daniel R. Boutz,[a1,2] Patrick J. Collins,[b1] Uthra Suresh,[c1] Mingzhu Lu,[d] Cristina M. Ramírez,[e] Carlos Fernández-Hernando,[e] Yufei Huang,[c,d] Raquel de Sousa Abreu,[c] Shu-Yun Le,[f] Bruce A. Shapiro,[f] Angela M. Liu,[g,h] John M. Luk,[g,h,i,j] Shelley Force Aldred,[b] Nathan D. Trinklein,[b] Edward M. Marcotte,[a,k] and Luiz O. F. Penalva[c,f,l3]

From the [a]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, [b]SwitchGear Genomics, Menlo Park, California 94025, the [c]Children's Cancer Research Institute, University of Texas Health Science Center, San Antonio, Texas 78229, the [d]Department of Electrical and Computer Engineering, University of Texas, San Antonio, Texas 78249, the [e]Department of Medicine, Leon H. Charney Division of Cardiology, New York University School of Medicine, New York, New York 10016, the [f]Center for Cancer Research Nanobiology Program, NCI-Frederick, National Institutes of Health, Frederick, Maryland 21702, the [g]Department of Surgery, University of Hong Kong, Queen Mary Hospital, Hong Kong, China, the [h]Department of Pharmacology, National University of Singapore, Singapore, the [i]Department of Surgery, National University of Singapore, Singapore, the [j]Cancer Science Institute, National University of Singapore, Singapore 117597, Singapore, the [k]Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712, and the [l]Department of Cellular and Structural Biology, University of Texas Health Science Center, San Antonio, Texas 78229

MicroRNAs function as important regulators of gene expression and are commonly linked to development, differentiation, and diseases such as cancer. To better understand their roles in various biological processes, identification of genes targeted by microRNAs is necessary. Although prediction tools have significantly helped with this task, experimental approaches are ultimately required for extensive target search and validation. We employed two independent yet complementary high throughput approaches to map a large set of mRNAs regulated by miR-122, a liver-specific microRNA implicated in regulation of fatty acid and cholesterol metabolism, hepatitis C infection, and hepatocellular carcinoma. The combination of luciferase reporter-based screening and shotgun proteomics resulted in the identification of 260 proteins significantly down-regulated in response to miR-122 in at least one method, 113 of which contain predicted miR-122 target sites. These proteins are enriched for functions associated with the cell cycle, differentiation, proliferation, and apoptosis. Among these miR-122-sensitive proteins, we identified a large group with strong connections to liver metabolism, diseases, and hepatocellular carcinoma. Additional analyses, including examination of consensus binding motifs for both miR-122 and target sequences, provide further insight into miR-122 function.

MicroRNAs (miRNAs)[4] are small (20–24 nt) endogenously expressed noncoding RNAs that regulate the translational efficiency and/or degradation of specific mRNAs. First discovered in *Caenorhabditis elegans* (1), miRNAs have since been identified in a diverse set of eukaryotic organisms as well as viruses, with over 700 miRNAs currently identified in humans (2). miRNAs function via the RNAi pathway, guiding the RNA-induced silencing complex to mRNAs by Watson-Crick base pairing between the miRNA and target mRNA (reviewed in Ref. 3). The resulting interaction leads to translational repression of the mRNA, although how this repression is achieved remains unclear. Several mechanisms have been proposed (reviewed in Ref. 4), and many questions remain; however, it is clear that sequence complementarity lies at the heart of miRNA function. In animals, sequence complementarity between an miRNA and its target mRNA is rarely perfect. The vast majority of binding sites contain mismatches between strands, and these mismatches have been shown to play an important functional role in target repression (5, 6). Imperfect complementarity allows for a greater variability of target sequences, thus increasing the number of potential binding sites for a given miRNA, with estimates of 300–400 targets on average per miRNA (7). Although many potential miRNA targets can be identified through predictions, no computational approach is all-inclusive, and even highly conservative predictions must be validated by experimental means to verify functional relevance.

Certain miRNAs have caught the attention of scientists due to their strong connections to diseases and cancer. Such is the

* This work was supported, in whole or in part, by National Institutes of Health Grants GM076536, GM067779, and GM088624, and Grant F-1515 from the Welch Foundation and by the Packard Foundation. (to E. M. M.) and the NCI Center for Cancer Research Intramural Research Program (to the B. A. S. laboratory). This work was also supported by the Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio (to the L. O. F. P. laboratory), and by SwitchGear Genomics. N. D. T., S. F. A., and P. J. C. are employees of and hold stock in SwitchGear Genomics. SwitchGear Genomics sells 3′UTR reporter vectors commercially. SwitchGear Genomics is the owner by assignment of patents or patent applications related to the 3′UTR reporter platform.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S10 and Figs. S1–S3.

[1] These authors contributed equally to this work.
[2] To whom correspondence may be addressed: Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, 2500 Speedway, MBB 3.210, Austin, TX 78712. Tel.: 512-232-3919; Fax: 512-471-2149; E-mail: dboutz@mail.utexas.edu.
[3] To whom correspondence may be addressed. Tel.: 210-562-9049; Fax: 210-562-9014; E-mail: penalva@uthscsa.edu.

[4] The abbreviations used are: miRNA, microRNA; nt, nucleotide; HCC, hepatocellular carcinoma.

case for miR-122, which has been implicated in liver-related diseases and hepatocellular carcinoma (HCC). miR-122 is highly abundant in the liver, accounting for 70% of total liver miRNA expression (8), and liver specificity seems to be conserved, at least from mouse to human (8, 9). miR-122 has been associated with the regulation of liver metabolism as well as hepatitis C infection and is often down-regulated in HCCs (reviewed in Ref. 10). The use of antisense miR-122 down-regulated several genes implicated in liver metabolism and produced an increase in expression of hundreds of genes that are normally repressed in hepatocytes, suggesting that miR-122 functions as an important player in maintaining the liver phenotype (11, 12). Moreover, silencing of miR-122 in high fat fed mice produced a reduction of hepatic steatosis, which can be linked to a reduction in cholesterol synthesis and stimulation of fatty acid oxidation (11).

miR-122 dysregulation has also shown a strong association with tumorigenesis. A reduction in miR-122 expression has been observed both in a rat model for HCC and in human HCC samples compared with pair-matched control tissues (13). Restoration of miR-122 expression in HCC cell lines impaired *in vitro* migration, anchorage-independent growth, invasion, angiogenesis, and intrahepatic metastasis (14). Similar findings have been obtained by other research groups (15, 16). In addition, miR-122 has been shown to influence apoptosis; transfections of the hepatoma cell line Huh-7 with a miR-122 mimic produced an up-regulation in apoptosis levels as indicated by both flow cytometry and TUNEL assay (17).

Given the importance of miR-122 to proper liver function in health and disease, a more extensive knowledge of miR-122 targets would greatly improve our understanding of the function of this miRNA. To this end, we have developed a combined high throughput screen for miRNA-targeted genes that uses a luciferase-based assay and label-free quantitative proteomic mass spectrometry. Our combined approach revealed that miR-122 controls a network of genes with strong connections to liver metabolism, diseases, and cancer-related processes.

EXPERIMENTAL PROCEDURES

Human 3′UTR Luciferase Fusion Reporters—Human 3′-untranslated regions (3′UTRs) were systematically identified and cloned into an optimized luciferase reporter vector system (SwitchGear Genomics). The reporter protein contains a PEST protein degradation sequence that enables a more sensitive measure of repression. A total of 139 3′UTR luciferase fusion reporters were selected from this genome-wide collection based on the presence of one or more predicted miR-122 target sites as determined by the computational prediction sets identified by PicTar (18) and microCosm (formerly miRBase Targets) (2, 19). An additional 16 reporters were used as negative controls, determined by a lack of predicted miR-122 target sites. This negative set included the empty vector, 11 constructs from the genome-wide collection, and 4 controls containing a random, nongenic, and nonconserved sequence in place of the 3′UTR. An additional 14 reporters were selected from the genome-wide collection for validation of putative miR-122 targets identified in the proteomics experiment.

Cell Growth, Transfection, and Luciferase Assays—96-Well plates were seeded with 6000 HT-1080 cells (ATCC) 18–24 h before transfection to achieve 80% confluence at the time of transfection. Each transfection included 0.15 μl of DharmaFECT DUO transfection reagent (Dharmacon), 100 ng of 3′UTR reporter, and sufficient mimic or nontargeting control miRNA (Dharmacon) to yield a final concentration of 20 nM in a total volume of 100 μl/well. Each construct was transfected in triplicate separately with either the miR-122 mimic or the nontargeting control. Plates were incubated at 37 °C for 24 h post-transfection before being removed. 100 μl of SteadyGlo luciferase assay reagent (Promega) was added to each well, and plates were incubated at room temperature for 30 min and finally read on a LmaxII-384 luminometer (Molecular Devices).

To identify which genes were significantly repressed, we calculated a p value (one-tail t test) and $\log_2$ ratio for each reporter from the average luminescence values of the miR-122 mimic and nontargeting control transfections. We then established threshold values for significance based on the normal distribution of the negative control set, a conservative p value < 0.05 along with a minimum of a 1.5-fold repression.

Mutagenesis Studies—Mutations of seed sequences were generated using a modification of the QuikChange (Stratagene) protocol (20). miR-122 target sites were predicted in the cloned sequences of six miR-122-responsive 3′UTRs identified by experimental screens. Two or three nucleotides were mutated within a single target site in each 3′UTR reporter construct. After sequence confirmation, mutant reporters were tested along with wild-type controls in quadruplicate as described above. Luminescence values of wild-type and mutant constructs were compared using the one-tailed t test. The $\log_2$ ratio of miR-122 mimic and nontargeting control for each wild-type and mutant construct pair was also calculated.

Proteomics Sample Preparation—HT1080 cells were transfected with 20 nM miR-122 mimic or mock transfection in T-25 flasks using the transfection reagent Dharmafect 4 (Dharmacon). Cells were harvested 24 h post-transfection and lysed by Dounce homogenization in low salt buffer (10 mM Tris-HCl, pH 8.0, 10 mM KCl, 1.5 mM $MgCl_2$) with protease inhibitor mixture (Calbiochem) and centrifuged at 1000 $\times$ g to separate into crude soluble (cytosolic) and insoluble (nuclear) fractions. Nuclear fractions were washed once and resuspended in low salt buffer, at which point both fractions were treated the same. For each sample, 2,2,2-trifluoroethanol was added to 50% (v/v). Samples were reduced with 15 mM DTT at 55 °C for 45 min and then alkylated with 55 mM iodoacetamide at room temperature for 30 min. Following alkylation, samples were diluted in digestion buffer (50 mM Tris-HCl, pH 8.0, 2 mM $CaCl_2$) to a final 2,2,2-trifluoroethanol concentration of 5% (v/v). Proteomics grade trypsin (Sigma) was added to a 1:50 (enzyme/protein) concentration, and samples were digested at 37 °C for 5 h. The digestion was quenched with 1% formic acid, and sample volume was reduced to 20 μl by SpeedVac centrifugation. Digested peptides were bound and washed on HyperSep C-18 SpinTips (Thermo), resuspended in peptide buffer (95% H_2O, 5% acetonitrile, 0.1% formic acid), and filtered through Microcon 10-kDa centrifugal filter (Millipore), with the digested peptides collected as flow-through.

Peptides were separated on a Zorbax reverse-phase C-18 column (Agilent) using a 5–38% acetonitrile gradient over 230 min and analyzed on line by nanoelectrospray-ionization tandem mass spectrometry on an LTQ-Orbitrap (Thermo Scientific). Data-dependent ion selection was activated, with parent ion (MS1) scans collected at high resolution (100,000). Ions with charge $>+1$ were selected for collision-induced dissociation fragmentation spectrum acquisition (MS2) in the LTQ, with a maximum of 12 MS2 scans per MS1. Dynamic exclusion was activated, with a 45-s exclusion time for ions selected more than twice in a 30-s window.

Proteomics Data Analysis—Spectra were searched against the Ensembl release 54 (21) human protein-coding sequence database using Sequest (Bioworks version 3.3.1, Thermo Scientific). A 1% false discovery rate was determined against a reversed concatenated decoy database, with specific filters (*i.e.* Δ*CN*, XCorr) selected to maximize the number of protein identifications in the forward database while maintaining the percentage of reversed identifications at 1% of the total. The protein list was curated by collapsing into groups the proteins for which there was identical evidence of observation and by removing the proteins for which observed peptides could be accounted for by other proteins with additional unique observations. Preference was given to sequences with predicted miR-122 seed sites in the 3′UTR of the associated mRNA transcript. Differential expression of proteins across miR-122-treated and control samples was calculated from the spectral count based on the APEX method of quantitation (22). Proteins with a Z-score >1.65 (one-tailed) were considered significantly down-regulated in miR-122-treated samples.

Cell Culture, Transfection, and Western Blot Analysis in Liver Cell Lines—Huh7 cells were obtained from American Type Tissue Collection and maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS and 2% penicillin/streptomycin.

Huh7 cells were transfected with 40 nM mi*RIDIAN* miRNA mimics (miR-122) (Dharmacon) utilizing Oligofectamine (Invitrogen). All control samples were treated with an equal concentration of nontargeting control mimic sequence to control for nonsequence-specific effects in miRNA experiments.

Cell were lysed in ice-cold buffer containing 50 mM Tris-HCl, pH 7.5, 125 mM NaCl, 1% Nonidet P-40, 5.3 mM NaF, 1.5 mM NaP, and 1 mM orthovanadate, 175 mg/ml octyl glucopyranoside, and 1 mg/ml protease inhibitor mixture (Roche Applied Science) and 0.25 mg/ml 4-(2-aminoethyl)benzenesulfonyl fluoride (Roche Applied Science). Cell lysates were rotated at 4 °C for 1 h before the insoluble material was removed by centrifugation at 12,000 × *g* for 10 min. After normalizing for equal protein concentration, cell lysates were resuspended in SDS sample buffer before separation by SDS-PAGE. Proteins were transferred onto nitrocellulose membranes and probed with the indicated antibodies. Protein bands were visualized using the Odyssey Infrared Imaging System (LI-COR Biotechnology). Densitometry analysis of the gels was carried out using ImageJ software from the National Institutes of Health (rsbweb.nih.gov).

Antibodies—Rabbit polyclonal anti-vimentin (1:1000) and rabbit polyclonal antibody against G6PC3 (1:1000) were obtained from Cell Signaling and Sigma, respectively. Rabbit polyclonal antibody against PMK2 (1:1000), goat polyclonal antibody against IQGAP1 (1:500), and mouse polyclonal anti-ARHGAP1 (1:500) were obtained from Abnova. Hsp90 mouse monoclonal antibody from Cell Signaling was used as an internal loading control in each experiment. Secondary fluorescently labeled antibodies were from Molecular Probes (Invitrogen).

Comparison of Experimental Results with HCC Microarray Data—Correlations between expression of miR-122 and predicted targets were analyzed using miRNA and mRNA profiling data from a cohort of 94 HCC patients as described (23, 24). Correlation of each of the predicted targets was evaluated using Pearson correlation analysis.

Identification of miR-122-binding Motif—The miR-122-binding motif was investigated in luciferase and proteomics data separately. Genes identified as direct targets were selected for binding site analysis. Putative target sites were identified by the presence of a 7-mer or greater seed complementary site. Ensembl (release 58) and RefSeq (GRCh37/hg19 assembly) 3′UTR sequences were both analyzed, with the sequence containing the greatest number of predicted target sites selected for the analysis. For each binding site identified by a 7-mer seed, a stretch of 40 bp from the 5′ of the site was extracted. The miR-122-binding secondary structure with the 40-bp site was then generated by RNAduplex (25). Because miRNA binding is imperfect, the motif was predicted in terms of probability as regards the likelihood that each nucleotide in the miRNA sequence binds to the target site. An miR-122-binding matrix was constructed with a value of 1 for the *ij*-th element if the *j*th nucleotide of miR-122 paired to the *i*-th site, and a value of 0 otherwise. Based on the binding matrix, the empirical probability of binding was obtained for each miR-122 nucleotide. For each miR-122 nucleotide, the pairing nucleotides in each binding site were tallied, and the empirical frequency of corresponding perfectly paired nucleotides or a bulge or mismatch was then calculated. Next, the average frequency of a bulge or mismatch between every adjacent perfectly paired site was determined. To compare the similarity of motifs derived independently from luciferase and proteomics datasets, the KL distance between the two motif-binding probabilities was calculated as shown in Equation 1,

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \qquad \text{(Eq. 1)}$$

where $P(i)$ is the binding probability of the i-th nucleotide of miRNA-122 obtained from the luciferase dataset, and $Q(i)$ is the corresponding probability obtained from the proteomic dataset.

Target Site Accessibility Prediction for miRNA Targets—The computational prediction of miRNA targets relied upon a set of computer programs, including Target, SigStb, SegFold, and Scanfd (26–28). We initially used Target to search for putative target regions containing complementary sequences with an miR-122 seed sequence (P2–P8) in which only one wobble base pair G:U or U:G was allowed in P3–P8. A target site accessibility computation was then performed to eliminate unfavorable

sites, based on the hypothesis that a favorable target site should have an unstable folding region, have high target accessibility of the seed sequence, and have a distinct RNA secondary structure in the region immediately downstream of the complementary seed sequence.

To calculate the thermodynamic stability of the miRNA target site, SigStb was employed to compute a smoothly moving average stability score (Stbscr) for the region ±50 nts from first position of the predicted seed sequence (P1), using a 50-nt sliding window. Stbscr is defined as a standard Z-score, Stbscr = $(E - E_w)/\mathrm{SDT}_w$, where E is the lowest free energy computed from a local segment of 50-nt, and E_w and STD_w are the sample mean ± S.D. of the lowest free energy computed by folding all segments of the same size that are generated by taking successive overlapping segments of 50 nt stepped 1 nt at a time from the start to end positions of the sequences.

The target accessibility of a seed sequence was measured by the hybridization energy E_h of base pairing between the miRNA seed and the complementary seed sequence of the targeted mRNA. $E_h = E_1 - E_{\mathrm{cost}}$ was computed by Scanfd, where E_1 is the energy contribution from the entire base-paired region between the miRNA seed and the complementary seed sequence; E_{cost} is the energy cost of opening the complementary seed sequence into a single-stranded state in the local folding region ±40 nts around P1.

The distinct RNA secondary structures found in the flanking regions of the computed target sequence were characterized by SegFold and Scanfd. First, a significance score and stability score for each overlapping segment were computed by sliding a fixed length window 1 nt at a time along the complete 3′UTR sequence from the 5′ to 3′ end using the program SegFold. Using a Monte Carlo simulation, the window size was systematically increased from 40 to 100 nt by a 2-nt step. 100 randomly shuffled sequences were generated for each overlapped wild-type sequence, and the lowest free energies of each overlapped segment were calculated for both wild-type and random sequences. For each sequence, the most significant unusual folding regions in the target site flanking regions were selected. Finally, the corresponding RNA secondary structures of the unusual folding regions were computed by Scanfd.

Gene Ontology and Network Analyses—An in-depth literature search was performed in Pathway Studio 6 by the "add neighbors" algorithm to identify cell processes enriched among significantly repressed genes in the luciferase dataset. Significance was determined by Fisher's exact test. A similar analysis was performed on targets from the proteomics dataset containing an miR-122 seed complementary sequence in the 3′UTR. The add neighbors algorithm in Pathway Studio 6 was again used to identify targets that are associated with liver-related diseases. This analysis was performed for the combined luciferase and proteomics dataset with targets containing miR-122 seed sequence and also for the combined luciferase and the entire proteomics dataset.

The "add direct interactions" algorithm in Pathway Studio 6 was used to create a network of miR-122 predicted targets that have "expression" and/or "regulation" relations among other targets in the combined luciferase and proteomics set. Another kind of network was created by the "add common targets" algorithm that identifies nodes that have high connectivity to the genes in the dataset. The top 15 nodes showing high connectivity to the miR-122 predicted targets in the combined dataset were selected.

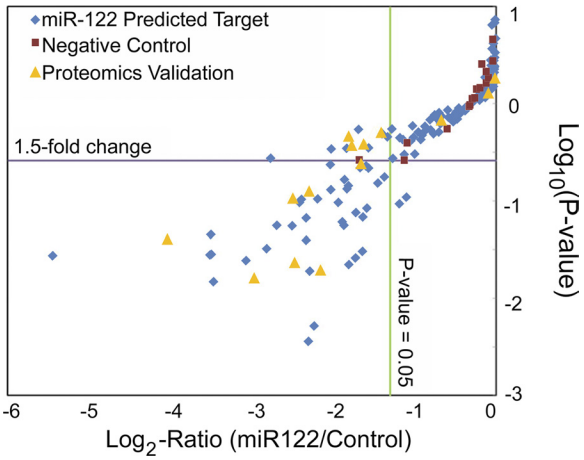


FIGURE 1. **Luciferase-based screening to identify miR-122 targets.** SwitchGear luciferase reporter constructs containing 3′UTRs of 139 genes predicted to be miR-122 target sites (*blue* ◆) were assayed for repression in response to miR-122 treatment. Negative controls (*red* ■) consisted of 11 3′UTRs from human genes lacking predicted miR-122 target sites and five scrambled sequence controls. An additional 14 genes (*yellow* ▲) identified as putative targets in the proteomic analysis were also assayed. Constructs exhibiting greater than 1.5-fold repression ($\log_2$ − ratio < −0.58) and a p value <0.05 ($\log_{10}(p\ \mathrm{value}) < -1.3$) (miR-122-treated *versus* control) were considered significantly down-regulated (*bottom-left quadrant*).

RESULTS

Identification and validation of miRNA targets can be a complicated task. There are several alternatives available starting with numerous target prediction methods and continuing with different biological procedures that encompass reporter-based screenings, shotgun proteomics, and Ago2-based immunoprecipitation methods (29–33). We have conducted two independent high throughput approaches as follows: a luciferase-reporter based screen and a quantitative shotgun proteomics analysis to identify a large set of genes under the influence of miR-122. These two approaches are very distinct in nature; the luciferase screen focuses on a list of genes derived from computational predictions, whereas the proteomics approach is open-ended. Datasets from both analyses were subsequently compared and analyzed with bioinformatics tools to determine particular connections between identified mRNA targets and specific biological processes.

Luciferase-based Strategy to Map miR-122 Targets—A genome-wide clone collection of luciferase reporters containing human 3′UTRs was used to conduct a screen of 139 computationally predicted miR-122 target genes to identify experimentally responsive targets. Target predictions were obtained from two commonly used sources, the PicTar (18) and MicroCosm (2, 19). Individual luciferase reporter constructs were employed in co-transfection experiments with 20 nM miR-122 mimic or a nontargeting mimic control in the HT1080 fibrosarcoma cell line. Thresholds of $p < 0.05$ (one-tailed t test) and down-regulation greater than 1.5-fold were determined from a set of negative controls to define the statistically significant subset of miR-122-responsive targets (>95% confidence) (Fig. 1).

TABLE 1
Comparison of computational target predictions against luciferase dataset

					No. of predictions by different algorithms		
	Target scan conserved	TargetScan no conservation	PicTar	MicroCosm	≥1	≥2	3
Total predictions	27	85	54	100	139	84	16
Down-regulated (% total)	14 (52%)	35 (41%)	21 (39%)	25 (25%)	37 (37%)	35 (42%)	9 (56%)

This repressed subset contains 37 of the 139 predicted targets screened (27%), with 24 of 37 (65%) repressed >2-fold. The complete list of results can be found in supplemental Table S1.

The use of multiple computational predictions for the selection of our test set allowed us to examine how each algorithm performed in the luciferase assay. In addition to PicTar and MicroCosm predictions, we also considered predictions from TargetScan 5.1 (7). Although TargetScan was not used in the initial selection of predicted targets, several genes in our test set overlap with TargetScan predictions. To best evaluate prediction methods, we considered each method independently, as well as in combination for genes predicted by more than one method. Of the 139 genes in the predicted target set, 85 were identified by TargetScan (27 if only conserved sites are considered), 54 by PicTar, and 101 by MicroCosm. The results, summarized in Table 1, indicate that TargetScan predictions (with and without conservation included) were the most likely to be validated as targets of miR-122 in the luciferase assay, and predictions by multiple algorithms increased the likelihood of repression, and genes predicted by MicroCosm alone performed quite poorly in our assay.

Mapping of miR-122 Targets by Shotgun Proteomics—For our second independent approach to identify miR-122 targets, we turned to mass spectrometry-based quantitative proteomics. Recent studies by Selbach *et al.* (33) and Baek *et al.* (29) examined miRNA-induced changes in protein levels by MS, using the isotopic labeling technique SILAC to quantify protein abundance. These papers report measurable repression in both mRNA and protein levels; however, the effect was consistently greater at the protein level, emphasizing the importance of measuring changes in cellular protein abundance. For this study, we adapted the APEX method of label-free protein quantitation by mass spectrometry (22).

For proteomics experiments, HT1080 cells were also used to maintain consistency with the luciferase assay. Cells treated with 20 nM miR-122 mimic or mock transfection for 24 h were lysed and split into cytosolic and nuclear fractions, with each analyzed independently across three replicate samples. In total, 2422 proteins were observed in at least two replicates, with 1704 and 1903 proteins observed in the cytosolic and nuclear fraction, respectively. 271 proteins were significantly repressed in at least one fraction of the miR-122-treated samples (Z-score >1.65, fold-change >1.3). After discarding proteins for which apparent miR-122-induced down-regulation in one fraction was contradicted by an increase in the other, we arrived at a final list of putative miR-122 targets containing 226 proteins identified as significantly down-regulated across the total cellular pool, *i.e.* in both fractions (supplemental Table S2).

As a first step toward confirming the putative targets, we examined the significantly down-regulated set for features of miR-122 targeting. The simplest feature of most miRNA-binding sites is the "seed" site, encoding a sequence in the mRNA 3′UTR that is perfectly complementary to nucleotides 2–7 at the 5′-end of the mature miRNA guide strand. To decrease false-positive predictions, we used a comparatively strict definition of the seed site, requiring at least seven sequential matches complementary to positions 1–7 or 2–8 of the miRNA. 75 of the 226 identified targets contained a 7-mer or greater seed complementary sequence in the 3′UTR. Down-regulated proteins containing at least one seed site showed a 2-fold enrichment over the total distribution of seeds. This enrichment increased to 3.8-fold for genes containing the even stricter 8-mer seed match, consistent with previous studies showing a greater repressive effect for 8-mers over other seed lengths (Fig. 2*A*) (34–36). To further support the claim that our down-regulated set is enriched for direct targets, we mapped conserved miR-122 target predictions from TargetScan 5.1 onto our dataset. Of the 124 conserved targets in the TargetScan database, 19 were mapped to the 2422 proteins, with 10 of the 19 (53%) found in the down-regulated set, a 5.5-fold enrichment (Fig. 2*B*). Including nonconserved TargetScan, predictions showed little improvement over the 7–8-mer seed prediction. Together, these results indicate that the down-regulated set is significantly enriched for real miR-122 targets.

Next, we looked to the literature for miR-122 targets that have been experimentally validated at the mRNA and/or protein level, providing a set of positive controls. Indeed, we observed several of these validated targets in the down-regulated set of proteins, including GYS1 and aldolase A (11). We also observed down-regulation of vimentin, which is commonly up-regulated in cancers, including HCC, and was shown to decrease significantly in HCC cells upon miR-122 expression (14, 16). Vimentin is a marker for mesenchymal cells and is positively associated with invasiveness and metastatic potential (37–39). Although not previously identified as a direct target, the vimentin 3′UTR does contain a 7-mer-m8 seed site, suggesting it may indeed be a direct miR-122 target.

Two additional proteins identified in our down-regulated set, citrate synthase and IQ-motif-containing GTPase-activating protein 1 (IQGAP1), have been previously implicated as miR-122 targets (7, 18, 40). IQGAP1 is particularly intriguing, as it was recently identified along with vimentin as a factor in hepatotumorigenesis (Ref. 41 and reviewed in Ref. 42). These examples demonstrate the effectiveness of our experimental approach toward identifying a subset of proteins enriched in miR-122 targets.

Comparison of Luciferase and Proteomic Data Sets—Following the initial screen, we selected an additional 14 miR-122 target genes identified through the proteomic analysis and evaluated their 3′UTR response to miR-122 by the luciferase

reporter assay. The results showed a very good correlation between the two methods employed; of the 14 tested, 7 genes exhibited significant down-regulation, and an additional 3 were significantly repressed with p values < 0.05, but the change in expression did not surpass the 1.5-fold repression we established as the cutoff for significance. Combining the 7 new targets with 3 genes identified in the initial screening, a total of 10 proteins were validated as high confidence direct targets with significant down-regulation in both analyses (Table 2). All 10 genes contain predicted target sites in their 3′UTRs; however, only *ALDOA*, *CS*, and *IQGAP1* have been previously validated as direct miR-122 targets. Furthermore, all previous validations of these three targets have been carried out in mice (11, 40, 43), making this the first validation of these three targets in a human cell line. An additional 34 targets were significantly down-regulated in the luciferase analysis, although the proteomics revealed 65 additional direct targets (containing a 7-mer or greater miR-122 seed site) and 151 indirect targets (lacking a 7-mer miR-122 seed site) in the down-regulated set (Fig. 3).

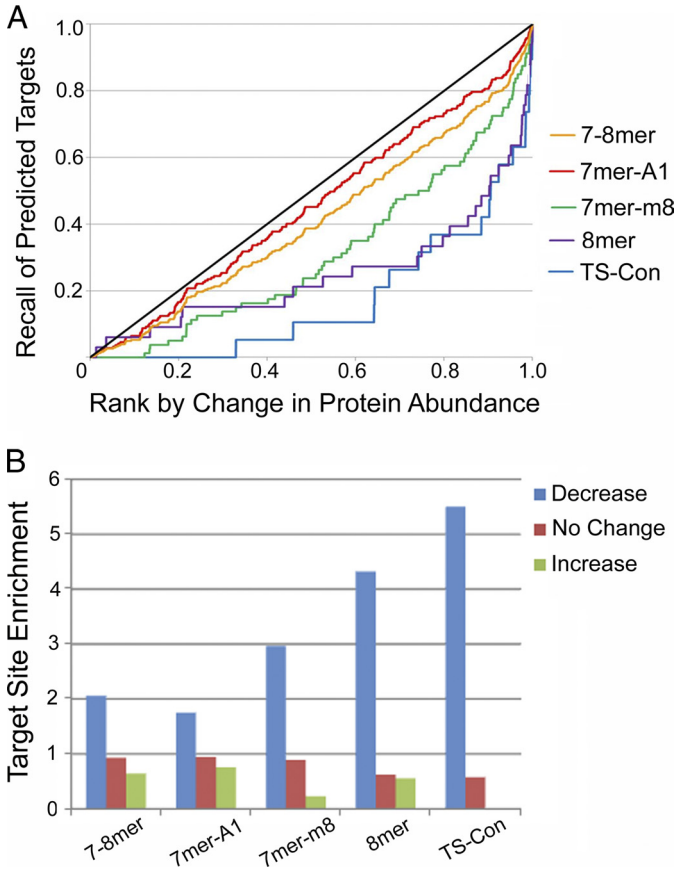


FIGURE 2. **Proteomic analysis identifies down-regulated proteins with strong enrichment for miR-122 target site prediction.** Predicted target sites were identified within 3′UTRs of proteins down-regulated in our proteomic analysis using various target prediction algorithms as follows: seed site complementarity of 7-mer-A1 and 7-mer-m8 motifs, 8-mer, and at either 7- or 8-mer (7–8-mer) seeds, as well as the TargetScan 5.1 with strict conservation of target sites (*TS-Con*). Results were plotted as the occurrence of a predicted target site *versus* the ranking of the proteins based on increasing down-regulation (*A*), and level of enrichment as the ratio of the frequency of predicted targets within each subset *versus* the total experimental dataset (*B*).

Confirmation of Target Down-regulation in Liver Cells—Our dual approach to target identification revealed many proteins responsive to miR-122. To determine the importance of context and confirm that targets identified in HT1080 cells showed similar response in a liver cell line, we first analyzed changes in abundance for five proteins in the Huh-7 hepatocellular carcinoma cell line. Western blot analysis revealed all five identified targets to be significantly down-regulated in response to miR-122 transfection (Fig. 4).

Two recent studies identified changes in expression patterns for a subset of genes that were anti-correlated with miR-122 expression as revealed by microarray analysis of HCC patients (23, 24). To further validate the biological relevance of targets identified in this study, and particularly in the context of HCC, we looked at whether our identified targets exhibited a strong anti-correlation with miR-122 as described in these recent studies (supplemental Table S3). Indeed, of the 41 targets empirically derived from the luciferase-based strategy that we mapped to the microarray dataset, 18 were negatively correlated with miR-122 expression in HCC tissues (Pearson correlation < -0.4, p value < 0.0001). For targets identified by the proteomic approach, 27 of 68 mapped targets were negatively correlated with expression of miR-122 (Pearson correlation < -0.4, p value < 0.0001), including 7 of the 10 targets cross-validated by luciferase. In total, 38.4% of mapped targets showed strong negative correlation with miR-122, although only 2 of 99 showed significant positive correlation, suggesting that the observed changes in protein levels as determined by independent assays in HT1080 cells are indicative of functional miR-122 targets under biologically relevant conditions.

miR-122-binding Site Analysis—To further validate our findings, we selected six miR-122 direct targets identified from our study for mutagenesis analysis. 2–3 bases were mutated in the seed recognition sequence of each 3′UTR reporter. In 5 of 6 cases, mutating the miR-122 seed recognition sequence resulted in significantly decreased repression by miR-122 (p value < 0.05), measured by luminescence in the presence of the

TABLE 2
Summary of 10 high confidence validated miR-122 direct targets

Target gene	Luciferase		Proteomics		No. of target site predictions	Target site type
	p value	Fold-change	Z-score	Fold-change		
ALDOA	2.32E-02	−2.25	7.32	−1.38	1	8-mer
ARHGAP1	6.81E-03	−3.13	5.42	−17.91	1	7-mer-m8
BCAT2	3.45E-03	−2.11	2.73	−2.52	1	7-mer-m8
CS	1.02E-02	−1.85	3.85	−1.36	1	7-mer-m8
GNPDA2	5.86E-03	−4.86	1.91	−4.84	1	7-mer-m8
IQGAP1	2.19E-03	−3.46	9.34	−2.12	2	7-mer-m8, 7-mer-A1
LAMC1	4.51E-02	−1.52	5.41	−4.02	2	7-mer-m8, 7-mer-A1
LMNB2	1.45E-02	−3.28	5.52	−5.73	2	7-mer-m8, 8-mer
MTHFD2	6.42E-03	−1.94	6.26	−15	1	7-mer-A1
PKM2	1.83E-04	−2.62	13.71	−1.59	1	8-mer

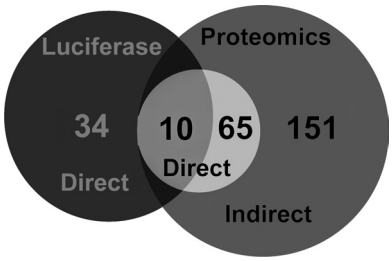


FIGURE 3. **Summary of the combined down-regulated proteomics and luciferase datasets.** A combined 260 genes were identified as significantly down-regulated in at least one of two experimental approaches. Significance was defined as p value < 0.05 and fold-change >1.5 for luciferase experiments and Z-score >1.65 and fold-change >1.3 for proteomics experiments. *Direct* refers to significantly down-regulated genes containing a 7-mer or greater miR-122 seed complementary site; *Indirect* refers to significantly down-regulated genes lacking such a site.

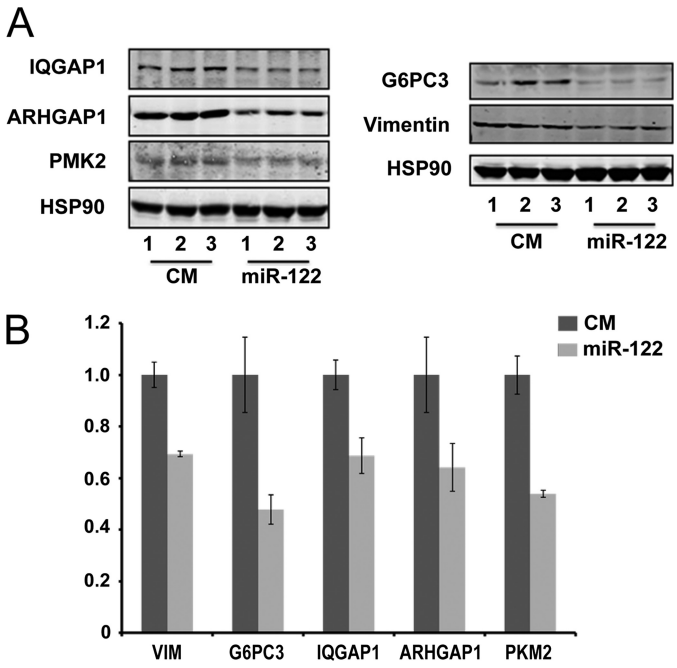


FIGURE 4. **Identified targets show miR-122-induced repression of protein levels in Huh-7 cells.** *A,* Western blots of five proteins identified as targets in our screen, from control mock-transfected (*CM*) and miR-122-transfected Huh-7 cells. *B,* relative protein abundance levels plotted from Western data show significant down-regulation of all five target proteins tested. *VIM*, vimentin.

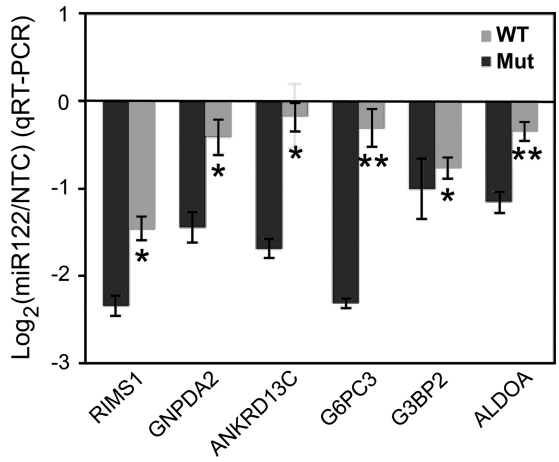


FIGURE 5. **Mutagenesis of miRNA seed recognition sequence disrupts miR-122 repression.** Six different genes identified as targets of miR-122 were selected for a mutagenesis study to partially validate our analysis and determine the importance of the seed sequence in miR-122 function. Two to three nucleotides were altered in the 3′UTR region of each gene matching the seed sequence and analyzed in luciferase assays. * designates a p value $<$ 0.005; ** designates a p value $<$0.005. *Mut*, mutant; *ALDOA*, aldolase A.

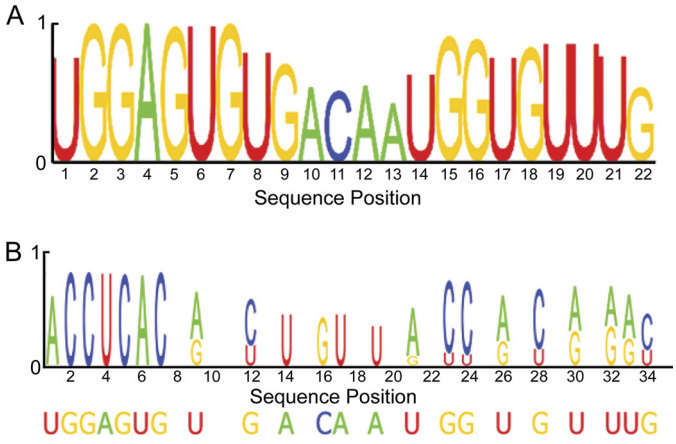


FIGURE 6. **Consensus binding motifs calculated for targets identified by luciferase and proteomics analyses.** The probability of involvement of a nucleotide in base pairing is displayed as the *height* of the corresponding letter for the miR-122 sequence (*A*) and predicted mRNA target site (*B*). Binding motifs were determined for the proteomics datasets. Binding motifs for the luciferase dataset show a very similar trend (not shown).

miR-122 mimic (Fig. 5). To better understand this regulation, we determined the secondary structure of target mRNAs (wild type and mutant) in the presence of miR-122 (supplemental Fig. S1 and Table S4). The miRNA accessibility in the seed complementary region of mRNAs appears to be very high for the six wild-type mRNAs, but it is decreased at least 40% for all mutated seed complementary sequences, as determined by the contribution of hybridization energy between the miRNA seed and mRNA seed complementary sequence. The data indicate that the thermodynamic stability of the local mRNA fold around the seed complementary sequence is below average (the stability score is greater than zero) and is increased for all mutated seed complementary sequences (stability score is decreased). These results also indicate that the seed complementary sequence is not involved in any local distinct RNA structure (within 50 nt), although neighboring regions do appear to have significant structures (supplemental Fig. S1). It remains unclear whether these adjacent secondary structures play any role in miRNA-mediated regulation.

The two independently derived sets of experimentally determined targets allowed us to examine whether miR-122 target sites share any common features by identifying consensus binding motifs for the miRNA and mRNA alike. Given that the selection of luciferase targets involved computational predictions of target sites, which could bias the motif search, the initial motif analysis focused on the proteomics dataset alone. Motifs were calculated from target sites containing a 7-mer or greater miR-122 seed complementary sequence in the significantly down-regulated dataset. Each sequence position was scored probabilistically as the likelihood of involvement in target site binding (Fig. 6). The miR-122 5′-end corresponding to the seed sequence shows high probability as expected based on selection criteria. However, motif characteristics of the middle and

3′-end were not affected by selection bias and thus reflect the general interaction features of the miRNA and target site. Interestingly, despite possible selection bias, the luciferase-binding motif was highly similar to the proteomics derived motif, with a calculated KL distance of 0.0021; the small KL distance quantifies the close similarity of the two binding motifs (data not shown).

The miR-122-binding motif contains two regions with high binding probability separated by a "bulge" region of four bases corresponding to nts 10–13 exhibiting substantially decreased binding probability. This bulge region has been identified in other miRNAs and may contribute to proper miRNA-target interaction (5, 44). The 3′-end of the miRNA shows a high frequency of binding, especially in nts 15–21. Previous studies have shown binding in this region to be common but not essential for other miRNAs (34, 35), and our results confirm the significance of this region to provide stability to the overall interaction. Analysis of the mRNA-binding motif shows strong binding complementary to the miRNA seed sequence as expected, but it lacks a consistent binding motif to complement the miRNA 3′-end due to frequent mismatches, indicating a lack of consensus with regard to bulge length. Thus, although the miRNA-binding motif implicates extensive involvement of the 3′-binding region, the exact placement of where this binding occurs in the target strand varies greatly between target sites. Examples of this can be found in supplemental Fig. S1, *d–f*, where the target mRNA secondary structures contain stem loops of varying sizes within the bulge region, thereby greatly affecting the distance between the seed complementary sites and 3′ complementary sites within the target strand.

Gene Ontology and Biological Association Analysis of Datasets—We expect miRNA-mediated regulation to display functional network characteristics similar to those observed for RNA-binding proteins (45, 46). In the case of miR-122, we anticipated the identification of genes associated with liver metabolism, liver diseases, and HCC. To define the biological nature of genes obtained in our study, we performed gene ontology and biological association analyses. First, we looked for enrichment of specific biological processes in both the proteomics and the luciferase datasets. Pathway Studio 6 (Ariadne Genomics) was used to identify biological processes that were enriched for the down-regulated set obtained from the combined luciferase and proteomics approaches. The add neighbors algorithm was used to obtain cell processes downstream of the targets. The highly connected cell processes were compared with the background targets that were not affected by miR-122. We used the Fisher's exact test with p value ≤ 0.05 to select significantly enriched cell processes. Apoptosis, cell cycle, cell death, cell differentiation, cell growth, cell proliferation, and mitosis were the top processes determined to be significantly enriched in the down-regulated set in respect to background (Fig. 7). Further analyses identified 30 genes from the "direct target set" (luciferase and proteomics combined), which have multiple associations with liver diseases, including HCC, metabolism, and function (Fig. 8 and supplemental Table S7). A larger number of connections were obtained when indirect targets (lacking a 7-mer seed site) were also included (supplemental Fig. S2 and Table S9). A more comprehensive list of genes

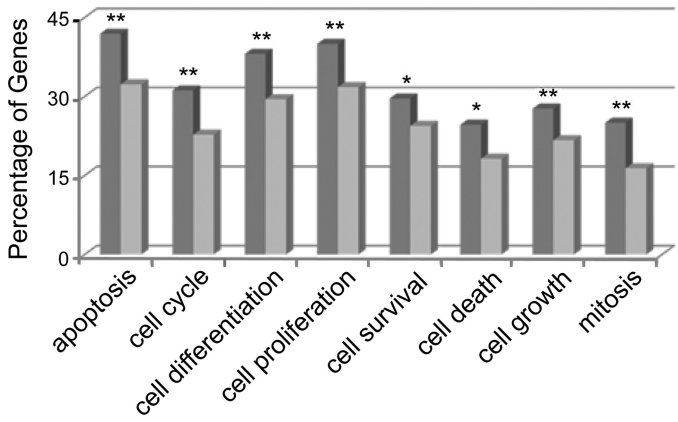


FIGURE 7. **Gene ontology analysis of miR-122 identified targets.** Pathway Studio 6 (Ariadne Genomics) was used to determine enrichment of cellular processes in the combined luciferase/proteomics down-regulated direct target set (*dark gray bars*). Total human transcriptome was used as background (*light gray bars*). * designates a p value <0.05; ** designates a p value <0.005.

related to diabetes and cancer can be found in Table 3 and supplemental Table S5.

Although these associations are not conclusive, they are certainly consistent with a role for miR-122 regulation in HCC development. In further support of this role, a recent study comparing mRNA and miRNA profiles across tumor and non-tumor tissue samples of HCC identified a network of mitochondrial genes responsive to miR-122 expression that becomes dysregulated upon down-regulation of miR-122 in HCC tumors, leading to loss of mitochondrial metabolic function (23). Interestingly, the authors (23) found that miR-122-responsive genes were strongly enriched for cell cycle-associated processes, a finding that is consistent with our results despite a limited overlap in genes identified between the two studies. These associations suggest a role for miR-122 in regulating not just liver metabolism but general liver function and tissue identity.

To identify gene functional networks that might be modulated by miR-122, we searched for biological connections between genes within the entire down-regulated dataset. Connections were established based on regulation of expression, function, or protein association described in the literature and retrieved by Pathway Studio 6. A network of 33 genes was identified (Fig. 9 and supplemental Table S8) containing down-regulated genes with either a predicted miR-122 target site or a direct connection to the predicted target. The genes with highest connectivity are *JUN*, a well establish oncogene with defined roles in transcription regulation (reviewed in Ref. 47) and RAC1 (Ras-related C3 botulinum toxin substrate 1), a member of the RAS superfamily of small GTP-binding proteins that are implicated in the control of cell growth, cytoskeletal reorganization, and the activation of protein kinases (reviewed in Ref. 48). Neither *JUN* nor *RAC1* contain predicted miR-122 target sites; however, both engage in functional interactions with several miR-122 targets identified in our screen. Another interesting gene, cAMP response-element binding protein (*CREB1*), was also shown to be strongly connected in our network, including connections to both *RAC1* and *JUN*. *CREB1* is a transcription factor involved in the regulation of a wide variety of cellular processes and is tied to oncogenesis (reviewed in Ref. 49).

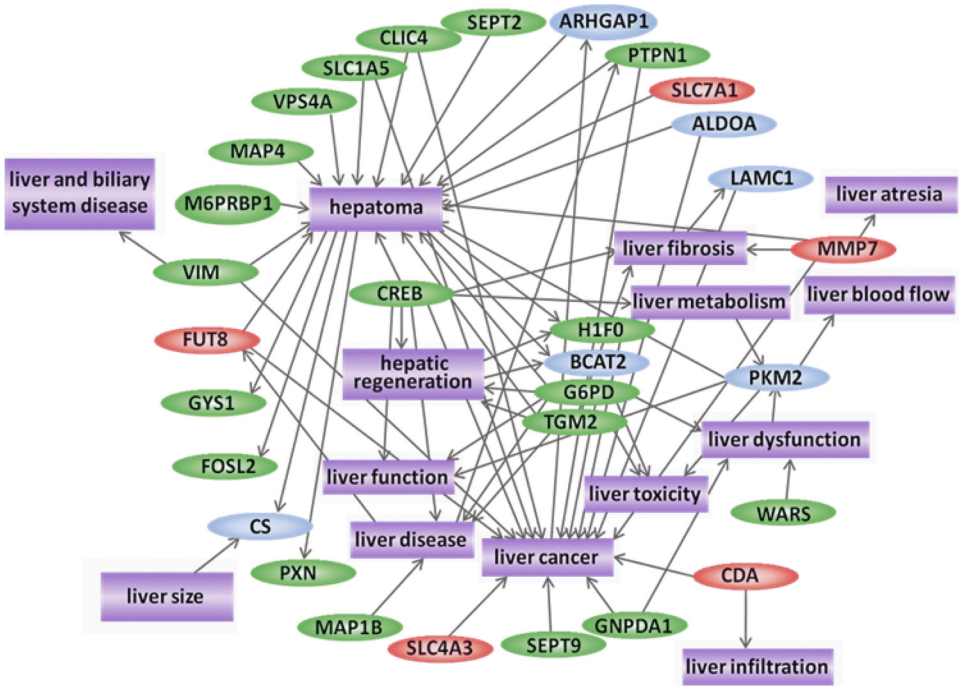


FIGURE 8. **miR-122 and liver-related functions and diseases.** The association between identified miR-122 targets and liver-related processes and diseases was investigated with Pathway Studio 6. *Green* indicates the genes identified in the proteomics analysis containing at least one miR-122 seed sequence; *red* indicates the genes identified by the luciferase analysis; *blue* indicates the genes identified by both methods. Liver function and disease associations are displayed in *purple*. *ALDOA*, aldolase A; *Vim*, vimentin; *CS*, citrate synthase; *PXN*, paxillin.

TABLE 3

Distribution of identified miR-122 targets according to biological function and diseases

The colors used are as follows: green, direct targets identified with proteomics analysis; pink, direct targets identified by luciferase assay; blue, high confidence targets identified by both methods.

(A) Genes from Luciferase and Proteomics (with seed sequences) enriched for cancer-related cell processes.

Cell Process	Genes
Apoptosis	ARHGAP1, BAI2, CACYBP, CDA, CDC42BPB, CLIC4, CSRP1, FBLIM1, FKBP1A, FOSL2, GIT1, H1F0, KHDRBS1, KIF5B, LAMC1, LETM1, LMNB2, M6PRBP1, MAP1B, MAP4, MMP7, NPEPPS, NUMBL, NUTF2, PKM2, PTPN1, PXN, RAD21, RBM3, SEMA4D, SEPT2, SEPT9, SERPINB6, SFRS2IP, SH3BGRL3, SLC1A5, SLC2A3, SLC7A1, TPD52L2, VAMP3, VIM, WARS, YARS
Cell Cycle	ARHGAP1, BCAT2, CLIC4, CSRP1, FOSL2, GNPNAT1, H1F0, KHDRBS1, KIF5B, LMNB2, MAP4, MAPRE1, MMP7, NPEPPS, NUDC, NUTF2, PKM2, PPP2CA, PTPN1, PXN, RAD21, RBM3, RCC2, SEPT2, SEPT9, SLC7A1, VIM, VPS4B
Cell Death	ALDOA, CLIC4, FKBP1A, FOSL2, H1F0, KIF5B, LAMC1, LMNB2, MAP1B, MAP4, MMP7, NUMBL, PKM2, PTPN1, PXN, RAD21, RBM3, RDH8, SEPT9, SFRS2IP, SH3BGRL3, TPD52L2, VIM, VPS4B
Cell Differentiation	ALDOA, ARHGAP1, BICD1, CACYBP, CDA, CLIC4, CSRP1, FBLIM1, FKBP1A, FOSL2, FUT8, GIT1, GNPNAT1, GTF3C2, H1F0, IQGAP1, KHDRBS1, KIF5B, LAMC1, M6PRBP1, MAP1B, MMP7, NUDC, PKM2, PPP2CA, PTPN1, PXN, RBM3, SEMA4D, SEPT2, SEPT9, SF3B4, SH3GL1, SLC1A5, SLC2A3, SLC4A3, SLC7A1, TNPO2, VAMP3, VIM, VPS4B, WARS
Cell Proliferation	ALDOA, ARHGAP1, BCAT2, CACYBP, CDA, CLIC4, CSRP1, DR1, FBLIM1, FKBP1A, FOSL2, GNPNAT1, GTF3C2, H1F0, IQGAP1, KHDRBS1, KIF5B, LAMC1, LMNB2, M6PRBP1, MAP1B, MAP4, MMP7, NUDC, NUMBL, PKM2, PPP2CA, PTPN1, PTPN14, PXN, RAD21, RBM3, RIMS1, SEMA4D, SEPT2, SEPT9, SH3GL1, SLC1A5, SLC4A3, SLC7A1, TPD52L2, VIM, WARS, YARS
Cell Survival	ARHGAP1, CDA, CSRP1, FBLIM1, FKBP1A, FOSL2, GYS1, H1F0, KHDRBS1, KLK5, LAMC1, MAP1B, MMP7, NUMBL, PKM2, PTPN1, PXN, RBM3, RIMS1, SEMA4D, SERPINB6, SLC1A5, SLC2A3, SLC4A3, SLC7A1, TPD52L2, VIM, WARS
Mitosis	ACTR1A, ALDOA, ARHGAP1, CLIC4, CSRP1, FKBP1A, GIT1, H1F0, IQGAP1, KHDRBS1, KIF5B, LMNB2, MAP1B, MAP4, MAPRE1, NUDC, NUTF2, PKM2, PPP2CA, PTPN1, PXN, RAD21, RBM3, RCC2, SLC7A1, VIM

(B) Genes from Luciferase and Proteomics (with seed sequences) enriched for lipid/cholesterol /glucose metabolism:

Lipid/ Cholesterol/ Glucose Metabolism	ALDOA, CDA, CREB1, CS, G6PD, GYS1, H1F0, MMP7, M6PRBP1, PKM2, PTPN1, PXN, VIM, VPS4A, WARS

(C) Genes from Luciferase and Proteomics (with seed sequences) enriched for Diabetes mellitus:

Diabetes mellitus	ALDOA, ARHGAP1, CDA, CREB1, CS, G6PD, GNPDA1, GYS1, KHDRBS1, LAMC1, MMP7, PKM2, PTPN1, PXN, SH3BGRL3, SLC2A3, TGM2, VAMP3, VIM, WARS

CREB1 does contain a predicted miR-122 target site, implicating it as a hub for the indirect regulation of numerous genes by miR-122.

Expanding upon our functional network, we identified several genes outside of our dataset with strong connectivity to our down-regulated gene set, including several that have been strongly implicated in apoptosis, cell growth, and proliferation and are key regulators of tumorigenesis (supplemental Fig. S3 and Tables S6 and S10). Most notably, *p53*, the extensively characterized tumor suppressor gene whose functions include cell cycle regulation and DNA repair (reviewed in Ref. 50); *MAPK1* and *MAPK3*, members of the MAPK family that are also known as ERKs (reviewed in Ref. 51); *MYC*, an important proto-oncogene that functions either as a positive or negative regulator of transcription and modulates cell cycle progression, apoptosis, and cellular transformation (reviewed in Ref. 52); *EGFR*, a transmembrane glycoprotein that functions as a receptor for members of the epidermal growth factor family and is involved cell proliferation (reviewed in Ref. 53); and VEGFA, a member of the platelet-derived growth factor (PDGF)/vascular endothelial growth factor (VEGF) family that acts on endothelial cells increase vascular permeability and growth and are also implicated in angiogenesis, cell migration, and apoptosis (reviewed in Ref. 54).

DISCUSSION

Our combined high throughput approach identified 260 genes repressed in response to miR-122, several of them key players in liver metabolism, disease, and cancer. Our data establish miR-122 as a regulatory node in a functional network of genes involved in liver metabolism and disease. Our results indicate that the number of proteins affected by miR-122

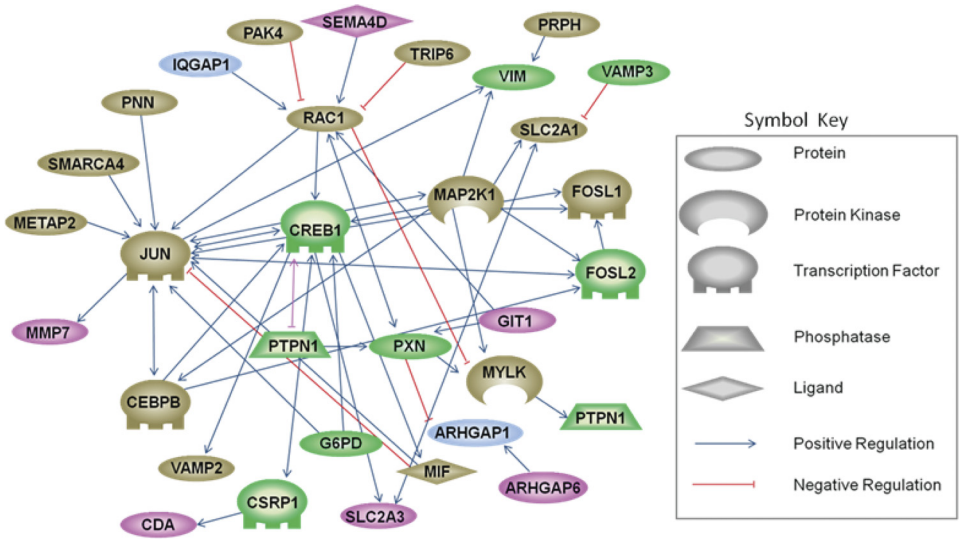


FIGURE 9. **Associations among miR-122-identified targets.** Pathway Studio 6 was used to establish connections among genes that responded to miR-122. Genes identified in the proteomics analysis as direct miR-122 targets containing predicted target sites are represented in *green,* and indirect targets lacking such sites are *brown*; *pink* indicates targets identified by the luciferase analysis; and *blue* indicates high confidence targets identified by both methods. *PXN*, paxillin.

extends far beyond direct targets to include indirect but functionally related targets. To highlight the extensive information contained within these functional networks, we will discuss a few highly relevant genes identified in this study in the context of liver processes and disease.

miR-122 has an established role in hepatocarcinoma/hepatoma (13, 15, 55, 56), functioning as a tumor suppressor gene, and is frequently down-regulated in tumor samples and HCC cell lines (reviewed in Ref. 10). Our data suggest that miR-122 controls a complex network of genes involved in cell cycle, proliferation, apoptosis, survival, and mutagenesis; therefore, miR-122 down-regulation could promote tumor formation in multiple ways. We will summarize the role of several direct targets we have identified in the context of tumorigenesis.

Tyrosine-protein phosphatase nonreceptor type 1 (PTPN1) is an enzyme that is the founding member of the protein-tyrosine phosphatase family. Protein-tyrosine phosphatases regulate numerous cellular processes, including cell growth, differentiation, cell cycle, and oncogenic transformation (reviewed in Ref. 57). PTPN1 (also known as PTP1B) has been explored as a potential target to control type 2 diabetes and obesity (58, 59) and has been shown to regulate glucose homeostasis, body weight, and energy expenditure thanks to its function as a negative regulator of insulin- and leptin receptor-mediated signaling pathways (60, 61). Additionally, PTPN1 has been suggested to function as an oncogene in the context of breast cancer. PTPN1 is up-regulated in HER2/Neu-transformed cells, and 90% of all breast tumor samples tested overexpress both PTPN1 and HER2/Neu (62–64).

A few miR-122 targets are associated with microtubules, including SEPT2 and SEPT9, members of the septin family. Septins were initially determined to be involved in cytokinesis and cell cycle control but more recently have been shown to have a role in microtubule-dependent processes, such as karyokinesis, exocytosis, and maintenance of cell shape (65). Both SEPT2 and SEPT9 were observed to be up-regulated in a variety of tumor types, including HCC (66). Another microtubule-associated target is vimentin (VIM), a member of the intermediate filament family. Intermediate filaments, along with microtubules and actin microfilaments, make up the cytoskeleton. Vimentin is implicated in the maintenance of cell shape, via stabilization of cytoskeletal interactions, and plays a role as an organizer of proteins implicated in attachment, migration, and cell signaling (reviewed in Ref. 67). In cancer, vimentin overexpression has been associated with HCC metastasis (68). Moreover, proteomic analysis indicated that circulating vimentin is higher in patients with small HCC than normal non-neoplastic controls, suggesting its use as a potential surrogate marker (37).

Another direct target identified in our dataset is the matrix metalloproteinase MMP7, which plays a role in the breakdown of extracellular matrix in normal physiological processes as well as in metastasis (reviewed in Ref. 69). MMP7 expression was shown to be anti-correlated with miR-122 in HCC patients, and levels of matrix metalloproteinase expression and the stage of tumor progression are frequently correlated. Moreover, it has been suggested that matrix metalloproteinases are also necessary for the creation and maintenance of a microenvironment that helps tumor growth and angiogenesis (70). MMP7 was determined to be up-regulated in cirrhotic nodules (potential precursors of HCC), compared with normal liver, as well as in HCC (71). Furthermore, elevated MMP7 expression is correlated with decreased survival and increased recurrence and liver metastasis of colon cancer (72) and pancreatic carcinoma (73). Most importantly, MMP7 was demonstrated to promote *in vitro* invasiveness of cancer cells of the stomach, colon, and pancreas (reviewed in Ref. 69).

Paxillin has also been implicated in metastasis. Paxillin is a multidomain protein that is primarily present in sites of cell adhesion to the extracellular matrix known as focal adhesions. Paxillin interacts with many proteins involved in the organization of the actin cytoskeleton, which are required for cell motility and implicated in a variety of biological processes, including

tumor metastasis (74, 75). Expression of paxillin protein in HCC may affect the invasive and metastatic ability of the tumor. In this regard, paxillin up-regulation was found to correlate with the presence of extrahepatic metastasis in hepatocellular carcinoma (76) and lymph node metastasis in breast tumors (77). Positive paxillin protein expression was associated with low differentiation, with the presence of portal vein thrombosis, and with extra-hepatic metastasis (76).

In addition to liver cancer, miR-122 has a direct role in the regulation of cholesterol and lipid metabolism (11, 78). The expression of miR-122 is highly restricted to the liver, where it is believed to maintain the differentiated state (8, 9). Silencing of miR-122 down-regulates genes implicated in cholesterol biosynthesis and triglyceride metabolism leading to reduction of total cholesterol levels (11, 78), making miR-122 a viable candidate for therapeutic inhibitors to lower cholesterol in humans.

In this study, we have identified several miR-122-responsive genes involved in glucose homeostasis and the citrate cycle, the regulation of which can ultimately alter lipid metabolism. The liver synthesizes triacylglycerols from fatty acids when glucose levels are high and acetyl-CoA production exceeds the energy requirements of the cell (79–82). Glucose provides the necessary substrates for triacylglycerol synthesis (acetyl-CoA for fatty acid synthesis and glycerol) using reactions in the glycolytic pathway and the citrate cycle (79–82). Aldolase A, a direct target of miR-122, cleaves fructose 1,6-bisphosphate to generate dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. Dihydroxyacetone phosphate can be used to make glycerol 3-phosphate, which can then be converted to triacylglycerol (79–82). Glyceraldehyde 3-phosphate is processed into pyruvate through the glycolysis pathway, ending with the conversion of phosphoenolpyruvate to pyruvate with the help of pyruvate kinase. We identified the muscle isoform of pyruvate kinase (PKM2) as a direct miR-122 target, whereas the liver form, PKLR, was not repressed in our luciferase assay. Unlike PKM2, PKLR responds to regulation by epinephrine and glucagon, allowing the liver to shift toward gluconeogenesis in response to stimuli such as low blood glucose levels (83, 84).

In order for pyruvate to enter the citrate cycle, it is first converted through oxidative decarboxylation to form acetyl-CoA. Citrate synthase, a direct target of miR-122, catalyzes the rate-limiting first step in the citrate cycle and the condensation of oxaloacetate and acetyl-CoA to form citrate (79–82). Citrate can also be cleaved to re-generate acetyl-CoA and oxaloacetate by citrate lyase (ACLY), an indirect target of miR-122. Acetyl-CoA can also be converted to malonyl-CoA, the building block for fatty acid synthesis by fatty-acid synthase (79). Ultimately, fatty acids and glycerol are combined to form triacylglycerols that are packaged into VLDL particles in the liver and transported to the adipose tissue where they are stored in lipid droplets. It is worth noting that many of these genes are in pathways that are still functional, although highly regulated, in liver. Although miRNA regulation is often thought of as a method for shutting down protein expression, these examples demonstrate the more nuanced role of miRNAs as chemostats, allowing for modulated control overexpression and dampening of stochastic noise to stabilize protein levels (85–87).

We also identified the muscle isoform of glycogen synthase (GYS1) as a direct target of miR-122. Glycogen synthase catalyzes the rate-limiting step in glycogenesis. Although the liver-specific isoform GYS2 lacks miR-122 target sites, the GYS1 3′UTR contains three sites, suggesting miR-122 might play an important role in maintaining the tissue-specific expression of glycogen synthase isoforms by suppressing the non-liver form. This is the likely scenario for pyruvate kinase isoforms PKM2 and PKLR as well.

The aforementioned evidence shows the vital role of miR-122 in lipid metabolism. In its regulation of multiple genes, including *PKM2*, *ALDOA*, *CS*, and *GYS1*, miR-122 is shown to regulate glucose homeostasis and ultimately lipid metabolism. As a result, the previous effects observed in miR-122 antisense oligonucleotide-treated mice on lipoprotein metabolism may be due at least in part to an alteration on glucose homeostasis caused by miR-122 depletion.

As additional miRNAs are identified and investigated, the importance of their function in the regulation of expression grows more evident, especially as pertains to control of cell growth and maintenance of the differentiated state. The extent to which miRNAs have been implicated in tumorigenesis and disease progression across virtually all cancer types is indicative of this importance, yet our understanding of how miRNAs are involved remains quite limited. The development of high throughput screens such as ours will lead to a more comprehensive identification of the large numbers of biologically relevant miRNA targets. These identifications will in turn allow us to map out regulatory networks and reveal the molecular mechanisms through which miRNAs function, or in the case of disease, dysfunction. Our approach has generated the largest dataset of experimentally tested miR-122 targets currently available. In addition, our network analysis has led to the identification of many interactions through which miR-122 could affect the development and progression of hepatocellular carcinoma. By using the data produced through approaches such as ours, we expect to see the development of improved target prediction algorithms, the validation of a larger number of targets, and through validated targets, a greater understanding of miRNA function.

Acknowledgment—We thank Dr. Christine Vogel for critical comments on the text.

REFERENCES

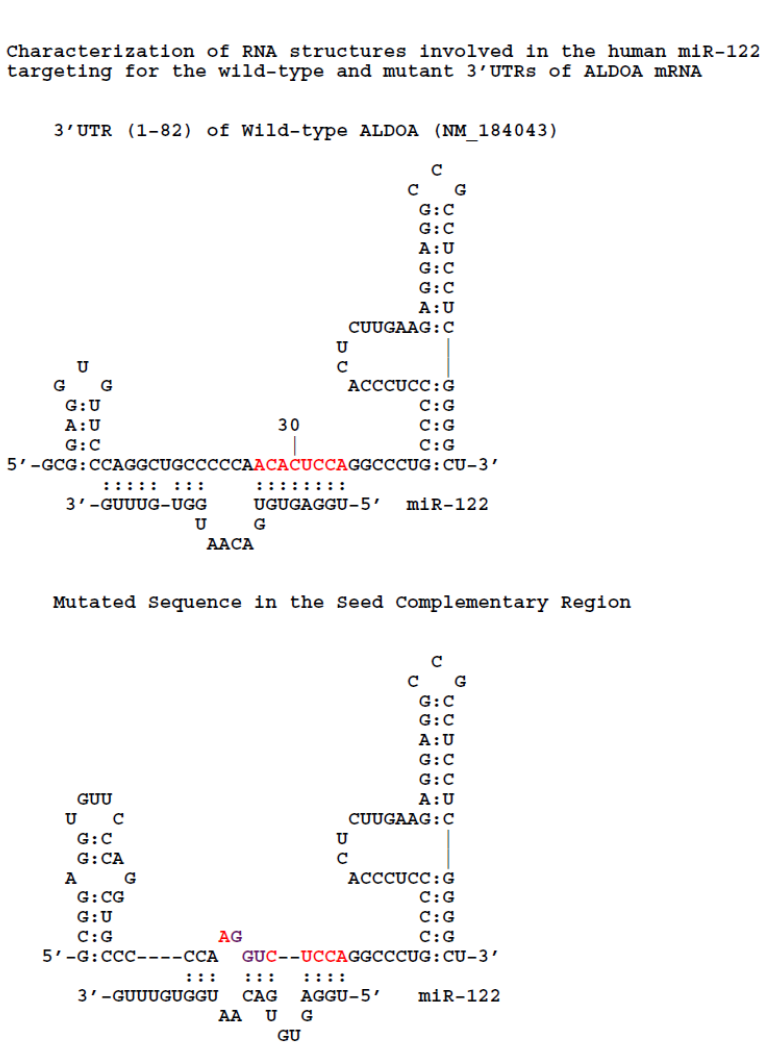
1. Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993) *Cell* **75,** 843–854
2. Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008) *Nucleic Acids Res.* **36,** D154–D158
3. Bartel, D. P. (2004) *Cell* **116,** 281–297
4. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008) *Nat. Rev. Genet.* **9,** 102–114
5. Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004) *Genes Dev.* **18,** 1165–1178
6. Ye, W., Lv, Q., Wong, C. K., Hu, S., Fu, C., Hua, Z., Cai, G., Li, G., Yang, B. B., and Zhang, Y. (2008) *PLoS One* **3,** e1719
7. Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009) *Genome Res.* **19,** 92–105
8. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002) *Curr. Biol.* **12,** 735–739

9. Chang, J., Nicolas, E., Marks, D., Sander, C., Lerro, A., Buendia, M. A., Xu, C., Mason, W. S., Moloshok, T., Bort, R., Zaret, K. S., and Taylor, J. M. (2004) *RNA Biol.* **1,** 106–113
10. Girard, M., Jacquemin, E., Munnich, A., Lyonnet, S., and Henrion-Caude, A. (2008) *J. Hepatol.* **48,** 648–656
11. Esau, C., Davis, S., Murray, S. F., Yu, X. X., Pandey, S. K., Pear, M., Watts, L., Booten, S. L., Graham, M., McKay, R., Subramaniam, A., Propp, S., Lollo, B. A., Freier, S., Bennett, C. F., Bhanot, S., and Monia, B. P. (2006) *Cell Metab.* **3,** 87–98
12. Czech, M. P. (2006) *N. Engl. J. Med.* **354,** 1194–1195
13. Kutay, H., Bai, S., Datta, J., Motiwala, T., Pogribny, I., Frankel, W., Jacob, S. T., and Ghoshal, K. (2006) *J. Cell. Biochem.* **99,** 671–678
14. Tsai, W. C., Hsu, P. W., Lai, T. C., Chau, G. Y., Lin, C. W., Chen, C. M., Lin, C. D., Liao, Y. L., Wang, J. L., Chau, Y. P., Hsu, M. T., Hsiao, M., Huang, H. D., and Tsou, A. P. (2009) *Hepatology* **49,** 1571–1582
15. Coulouarn, C., Factor, V. M., Andersen, J. B., Durkin, M. E., and Thorgeirsson, S. S. (2009) *Oncogene* **28,** 3526–3536
16. Bai, S., Nasser, M. W., Wang, B., Hsu, S. H., Datta, J., Kutay, H., Yadav, A., Nuovo, G., Kumar, P., and Ghoshal, K. (2009) *J. Biol. Chem.* **284,** 32015–32027
17. Wu, X., Wu, S., Tong, L., Luan, T., Lin, L., Lu, S., Zhao, W., Ma, Q., Liu, H., and Zhong, Z. (2009) *Scand. J. Gastroenterol.* **44,** 1332–1339
18. Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y. L., Dewey, C. N., Sood, P., Colombo, T., Bray, N., Macmenamin, P., Kao, H. L., Gunsalus, K. C., Pachter, L., Piano, F., and Rajewsky, N. (2006) *Curr. Biol.* **16,** 460–471
19. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006) *Nucleic Acids Res.* **34,** D140–D144
20. Zheng, L., Baumann, U., and Reymond, J. L. (2004) *Nucleic Acids Res.* **32,** e115
21. Flicek, P., Aken, B. L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Gräf, S., Haider, S., Hammond, M., Howe, K., Jenkinson, A., Johnson, N., Kähäri, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Koscielny, G., Kulesha, E., Lawson, D., Longden, I., Massingham, T., McLaren, W., Megy, K., Overduin, B., Pritchard, B., Rios, D., Ruffier, M., Schuster, M., Slater, G., Smedley, D., Spudich, G., Tang, Y. A., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S. P., Zadissa, A., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernández-Suarez, X. M., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Smith, J., and Searle, S. M. (2010) *Nucleic Acids Res.* **38,** D557–D562
22. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) *Nat. Biotechnol.* **25,** 117–124
23. Burchard, J., Zhang, C., Liu, A. M., Poon, R. T., Lee, N. P., Wong, K. F., Sham, P. C., Lam, B. Y., Ferguson, M. D., Tokiwa, G., Smith, R., Leeson, B., Beard, R., Lamb, J. R., Lim, L., Mao, M., Dai, H., and Luk, J. M. (2010) *Mol. Syst. Biol.* **6,** 402
24. Hao, K., Luk, J. M., Lee, N. P., Mao, M., Zhang, C., Ferguson, M. D., Lamb, J., Dai, H., Ng, I. O., Sham, P. C., and Poon, R. T. (2009) *BMC Cancer* **9,** 389
25. Schuster, P., Fontana, W., Stadler, P. F., and Hofacker, I. L. (1994) *Proc. Biol. Sci.* **255,** 279–284
26. Le, S. Y., and Maizel, J. V., Jr. (1989) *J. Theor. Biol.* **138,** 495–510
27. Le, S. Y., Chen, J. H., and Maizel, J. V. (1990) *Proceedings of the 6th Conversations in Biomolecular Stereodynamics, Structures and Methods*, State University of New York, Albany, NY, June 6–10, 1989 (Editors, R. H. Sarma and M. H. Sarma), Adenine Press, Inc., New York
28. Le, S. Y., Liu, W. M., and Maizel, J. V. (2002) *Knowledge Based System* **15,** 243–250
29. Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008) *Nature* **455,** 64–71
30. Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009) *Nature* **460,** 479–486
31. Hendrickson, D. G., Hogan, D. J., Herschlag, D., Ferrell, J. E., and Brown, P. O. (2008) *PLoS One* **3,** e2126
32. Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005) *Nature* **433,** 769–773
33. Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008) *Nature* **455,** 58–63
34. Brennecke, J., Stark, A., Russell, R. B., and Cohen, S. M. (2005) *PLoS Biol.* **3,** e85
35. Doench, J. G., and Sharp, P. A. (2004) *Genes Dev.* **18,** 504–511
36. Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007) *Mol. Cell* **27,** 91–105
37. Sun, S., Poon, R. T., Lee, N. P., Yeung, C., Chan, K. L., Ng, I. O., Day, P. J., and Luk, J. M. (2010) *J. Proteome Res.* **9,** 1923–1930
38. Kokkinos, M. I., Wafai, R., Wong, M. K., Newgreen, D. F., Thompson, E. W., and Waltham, M. (2007) *Cells Tissues Organs* **185,** 191–203
39. Hendrix, M. J., Seftor, E. A., Chu, Y. W., Seftor, R. E., Nagle, R. B., McDaniel, K. M., Leong, S. P., Yohem, K. H., Leibovitz, A. M., Meyskens, F. L., Jr., *et al.* (1992) *J. Natl. Cancer Inst.* **84,** 165–174
40. Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005) *Nature* **438,** 685–689
41. Tsubota, A., Matsumoto, K., Mogushi, K., Nariai, K., Namiki, Y., Hoshina, S., Hano, H., Tanaka, H., Saito, H., and Tada, N. (2010) *Carcinogenesis* **31,** 504–511
42. White, C. D., Brown, M. D., and Sacks, D. B. (2009) *FEBS Lett.* **583,** 1817–1824
43. Gatfield, D., Le Martelot, G., Vejnar, C. E., Gerlach, D., Schaad, O., Fleury-Olela, F., Ruskeepää, A. L., Oresic, M., Esau, C. C., Zdobnov, E. M., and Schibler, U. (2009) *Genes Dev.* **23,** 1313–1326
44. Vella, M. C., Reinert, K., and Slack, F. J. (2004) *Chem. Biol.* **11,** 1619–1623
45. Keene, J. D. (2007) *Nat. Rev. Genet.* **8,** 533–543
46. Sanchez-Diaz, P., and Penalva, L. O. (2006) *RNA Biol.* **3,** 101–109
47. Shaulian, E. (2010) *Cell. Signal.* **22,** 894–899
48. Bosco, E. E., Mulloy, J. C., and Zheng, Y. (2009) *Cell. Mol. Life Sci.* **66,** 370–374
49. Siu, Y. T., and Jin, D. Y. (2007) *FEBS J.* **274,** 3224–3232
50. Farnebo, M., Bykov, V. J., and Wiman, K. G. (2010) *Biochem. Biophys. Res. Commun.* **396,** 85–89
51. Boutros, T., Chevet, E., and Metrakos, P. (2008) *Pharmacol. Rev.* **60,** 261–310
52. Soucek, L., and Evan, G. I. (2010) *Curr. Opin. Genet. Dev.* **20,** 91–95
53. Burgess, A. W. (2008) *Growth Factors* **26,** 263–274
54. Hirashima, M. (2009) *Anat. Sci. Int.* **84,** 95–101
55. Meng, F., Henson, R., Wehbe-Janek, H., Ghoshal, K., Jacob, S. T., and Patel, T. (2007) *Gastroenterology* **133,** 647–658
56. Gramantieri, L., Ferracin, M., Fornari, F., Veronese, A., Sabbioni, S., Liu, C. G., Calin, G. A., Giovannini, C., Ferrazzi, E., Grazi, G. L., Croce, C. M., Bolondi, L., and Negrini, M. (2007) *Cancer Res.* **67,** 6092–6099
57. Tonks, N. K. (2006) *Nat. Rev. Mol. Cell Biol.* **7,** 833–846
58. Koren, S., and Fantus, I. G. (2007) *Best Pract. Res. Clin. Endocrinol. Metab.* **21,** 621–640
59. Liu, S., Zhou, B., Yang, H., He, Y., Jiang, Z. X., Kumar, S., Wu, L., and Zhang, Z. Y. (2008) *J. Am. Chem. Soc.* **130,** 8251–8260
60. Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A. L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C. C., Ramachandran, C., Gresser, M. J., Tremblay, M. L., and Kennedy, B. P. (1999) *Science* **283,** 1544–1548
61. Klaman, L. D., Boss, O., Peroni, O. D., Kim, J. K., Martino, J. L., Zabolotny, J. M., Moghal, N., Lubkin, M., Kim, Y. B., Sharpe, A. H., Stricker-Krongrad, A., Shulman, G. I., Neel, B. G., and Kahn, B. B. (2000) *Mol. Cell. Biol.* **20,** 5479–5489
62. Bentires-Alj, M., and Neel, B. G. (2007) *Cancer Res.* **67,** 2420–2424
63. Zhai, Y. F., Beittenmiller, H., Wang, B., Gould, M. N., Oakley, C., Esselman, W. J., and Welsch, C. W. (1993) *Cancer Res.* **53,** 2272–2278
64. Zhai, Y. F., Wirth, J. J., Welsch, C. W., and Esselman, W. J. (1996) *Cancer Treat. Res.* **83,** 107–125
65. Silverman-Gavrila, R. V., and Silverman-Gavrila, L. B. (2008) *Scientific World Journal* **8,** 611–620
66. Liu, M., Shen, S., Chen, F., Yu, W., and Yu, L. (2010) *Mol. Biol. Rep.* **37,** 3601–3608
67. Ivaska, J., Pallari, H. M., Nevo, J., and Eriksson, J. E. (2007) *Exp. Cell Res.* **313,** 2050–2062
68. Hu, L., Lau, S. H., Tzang, C. H., Wen, J. M., Wang, W., Xie, D., Huang, M., Wang, Y., Wu, M. C., Huang, J. F., Zeng, W. F., Sham, J. S., Yang, M., and Guan, X. Y. (2004) *Oncogene* **23,** 298–302

69. Ou, L., Ma, J., Zheng, X., Chen, X., Li, G., and Wu, H. (2006) *Protein Expr. Purif.* **47,** 367–373
70. Nelson, A. R., Fingleton, B., Rothenberg, M. L., and Matrisian, L. M. (2000) *J. Clin. Oncol.* **18,** 1135–1149
71. Colombat, M., Paradis, V., Bièche, I., Dargère, D., Laurendeau, I., Belghiti, J., Vidaud, M., Degott, C., and Bedossa, P. (2003) *J. Pathol.* **201,** 260–267
72. Fang, Y. J., Lu, Z. H., Wang, G. Q., Pan, Z. Z., Zhou, Z. W., Yun, J. P., Zhang, M. F., and Wan, D. S. (2009) *Int. J. Colorectal Dis.* **24,** 875–884
73. Shi, W. D., Meng, Z. Q., Chen, Z., Lin, J. H., Zhou, Z. H., and Liu, L. M. (2009) *Cancer Lett.* **283,** 84–91
74. Turner, C. E. (2000) *Nat. Cell Biol.* **2,** E231–E236
75. Turner, C. E. (2000) *J. Cell Sci.* **113,** 4139–4140
76. Li, H. G., Xie, D. R., Shen, X. M., Li, H. H., Zeng, H., and Zeng, Y. J. (2005) *World J. Gastroenterol.* **11,** 1445–1451
77. Vadlamudi, R., Adam, L., Tseng, B., Costa, L., and Kumar, R. (1999) *Cancer Res.* **59,** 2843–2846
78. Elmén, J., Lindow, M., Schütz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjärn, M., Hansen, H. F., Berger, U., Gullans, S., Kearney, P., Sarnow, P., Straarup, E. M., and Kauppinen, S. (2008) *Nature* **452,** 896–899
79. Saggerson, D. (2008) *Annu. Rev. Nutr.* **28,** 253–272
80. Cahová, M., Vavrínková, H., and Kazdová, L. (2007) *Physiol. Res.* **56,** 1–15
81. Frayn, K. N. (2003) *Biochem. Soc. Trans.* **31,** 1115–1119
82. Langhans, W. (2003) *Curr. Opin. Clin. Nutr. Metab. Care* **6,** 449–455
83. Ishibashi, H., and Cottam, G. L. (1978) *J. Biol. Chem.* **253,** 8767–8771
84. Nagano, M., Ishibashi, H., McCully, V., and Cottam, G. L. (1980) *Arch. Biochem. Biophys.* **203,** 271–281
85. Bartel, D. P., and Chen, C. Z. (2004) *Nat. Rev. Genet.* **5,** 396–400
86. Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S., and Carthew, R. W. (2009) *Cell* **137,** 273–282
87. Hornstein, E., and Shomron, N. (2006) *Nat. Genet.* **38,** S20–S24

Supplementary Figure S1

A



B

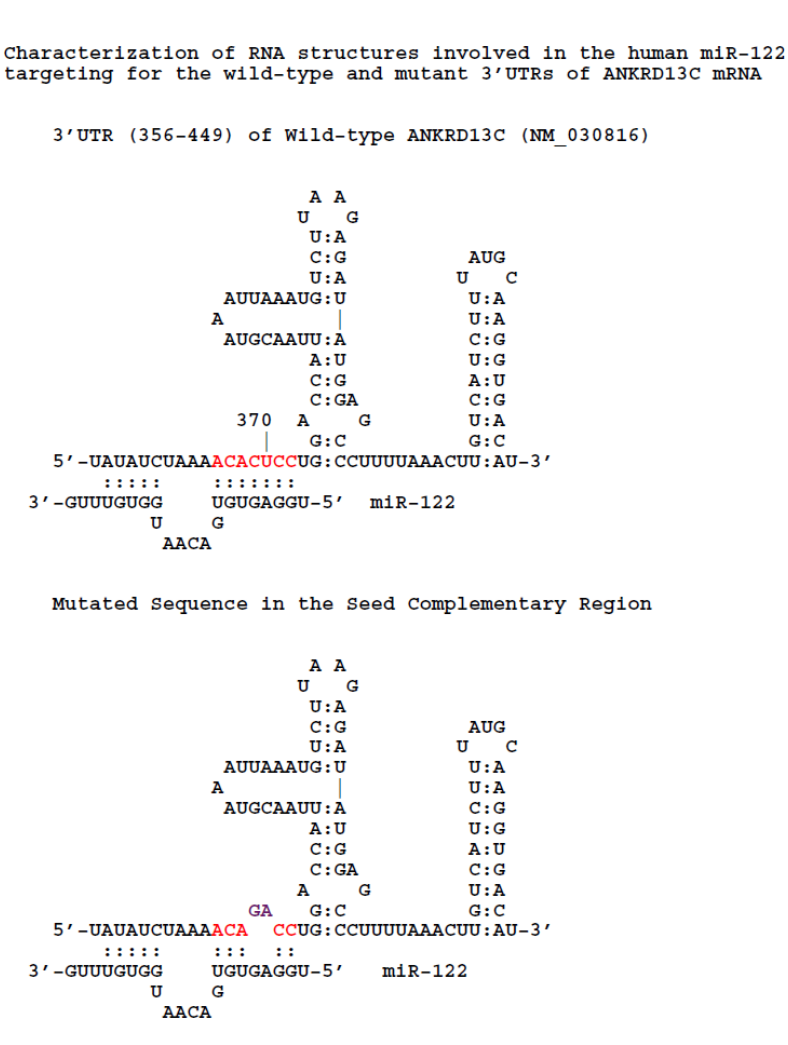


Figure S1. Secondary structure prediction of miR-122:target mRNA binding. The six target sites analyzed by mutagenesis (Figure 4) were selected for computational prediction of RNA secondary structure using a set of computer programs including Target, SigStb, SegFold, and Scanfd (55-57). (A) ALDOA; (B) ANKRD13C; (C) G6PC3; (D) G3BP2; (E) GNPDA2; (F) RIMS1.

Supplementary Figure S1

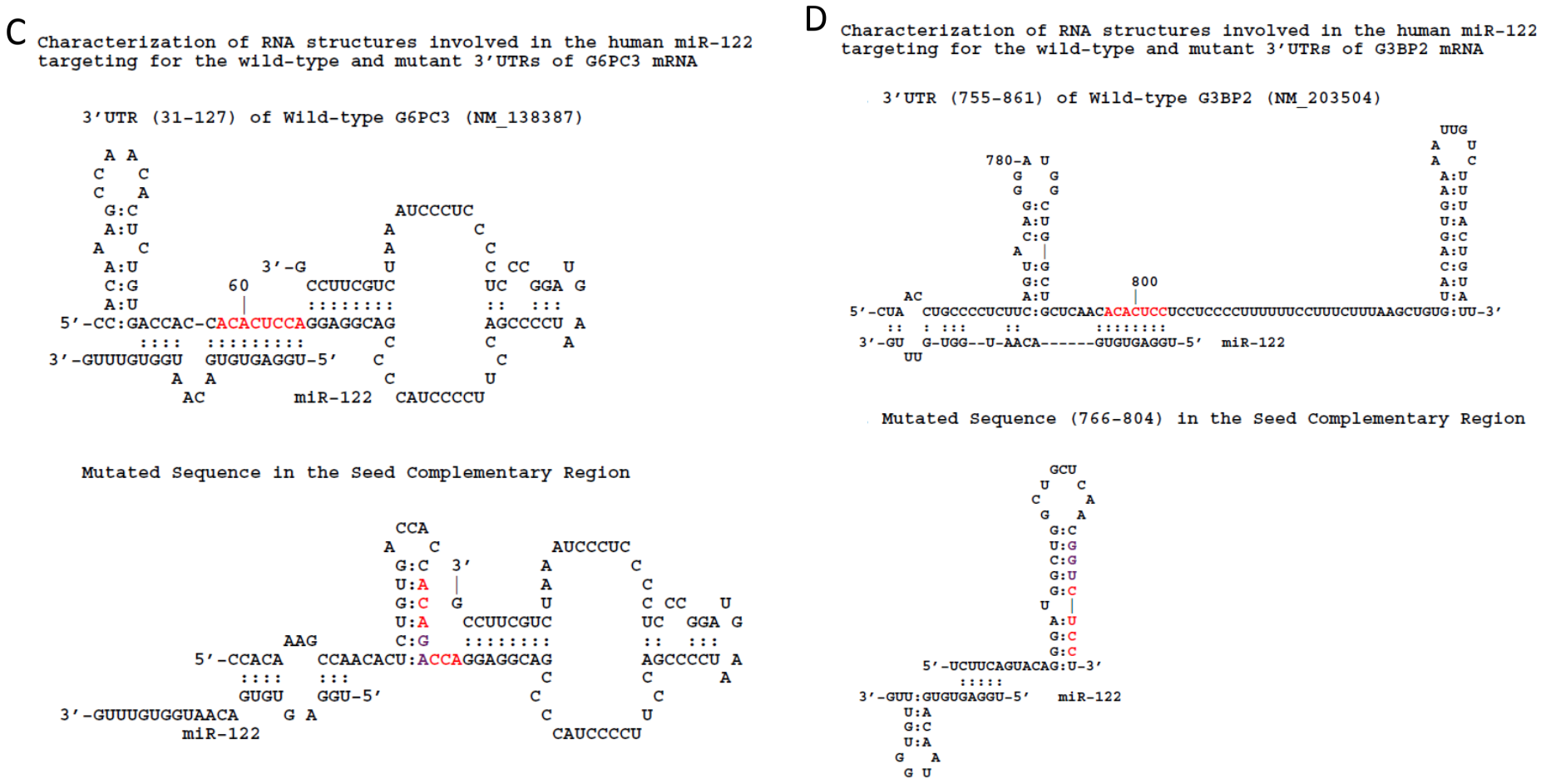


Figure S1. Secondary structure prediction of miR-122:target mRNA binding. The six target sites analyzed by mutagenesis (Figure 4) were selected for computational prediction of RNA secondary structure using a set of computer programs including Target, SigStb, SegFold, and Scanfd (55-57). (A) ALDOA; (B) ANKRD13C; (C) G6PC3; (D) G3BP2; (E) GNPDA2; (F) RIMS1.

Supplementary Figure S1

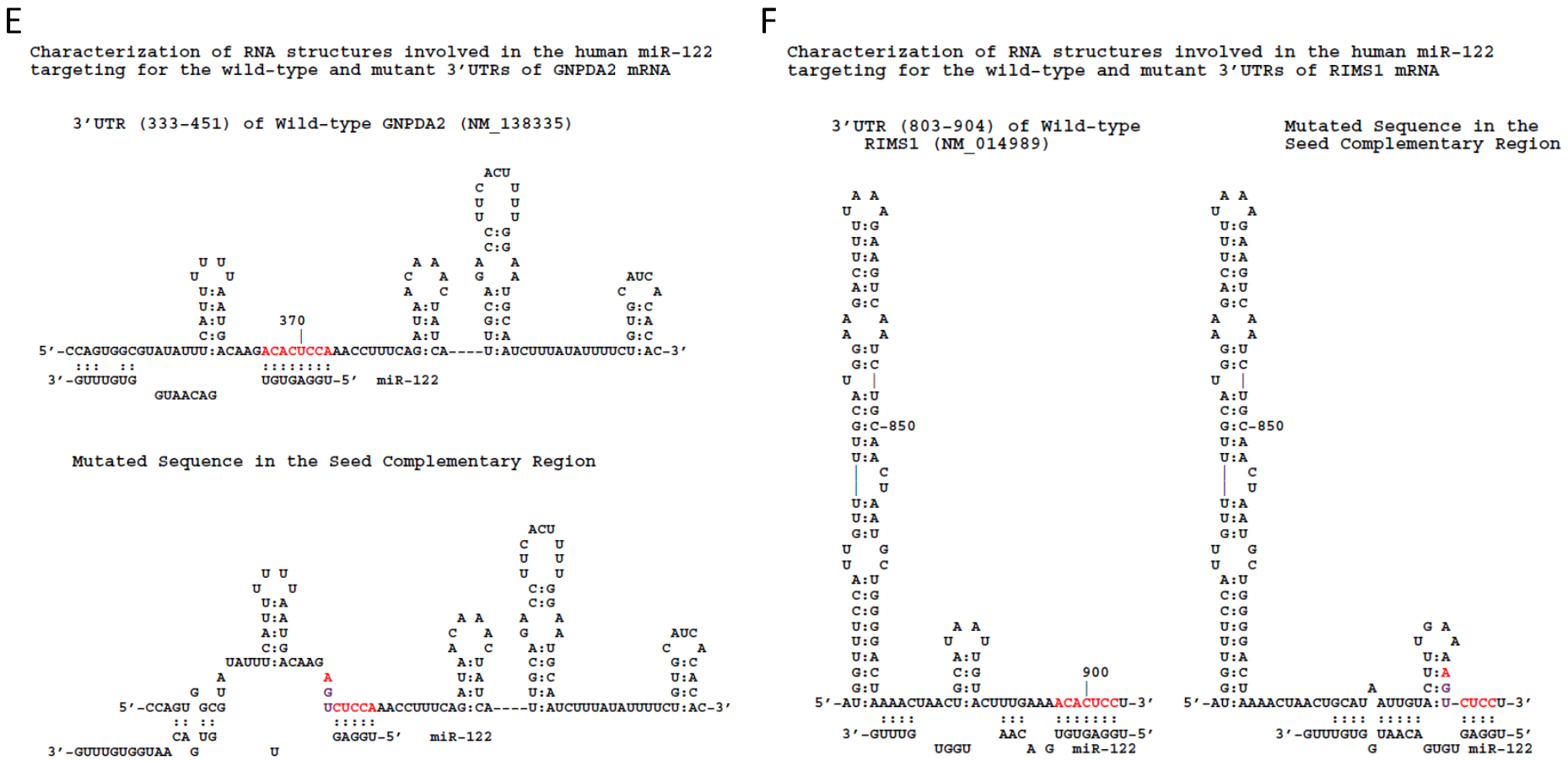


Figure S1. Secondary structure prediction of miR-122:target mRNA binding. The six target sites analyzed by mutagenesis (Figure 4) were selected for computational prediction of RNA secondary structure using a set of computer programs including Target, SigStb, SegFold, and Scanfd (55-57). (A) ALDOA; (B) ANKRD13C; (C) G6PC3; (D) G3BP2; (E) GNPDA2; (F) RIMS1.

Supplementary Figure S2

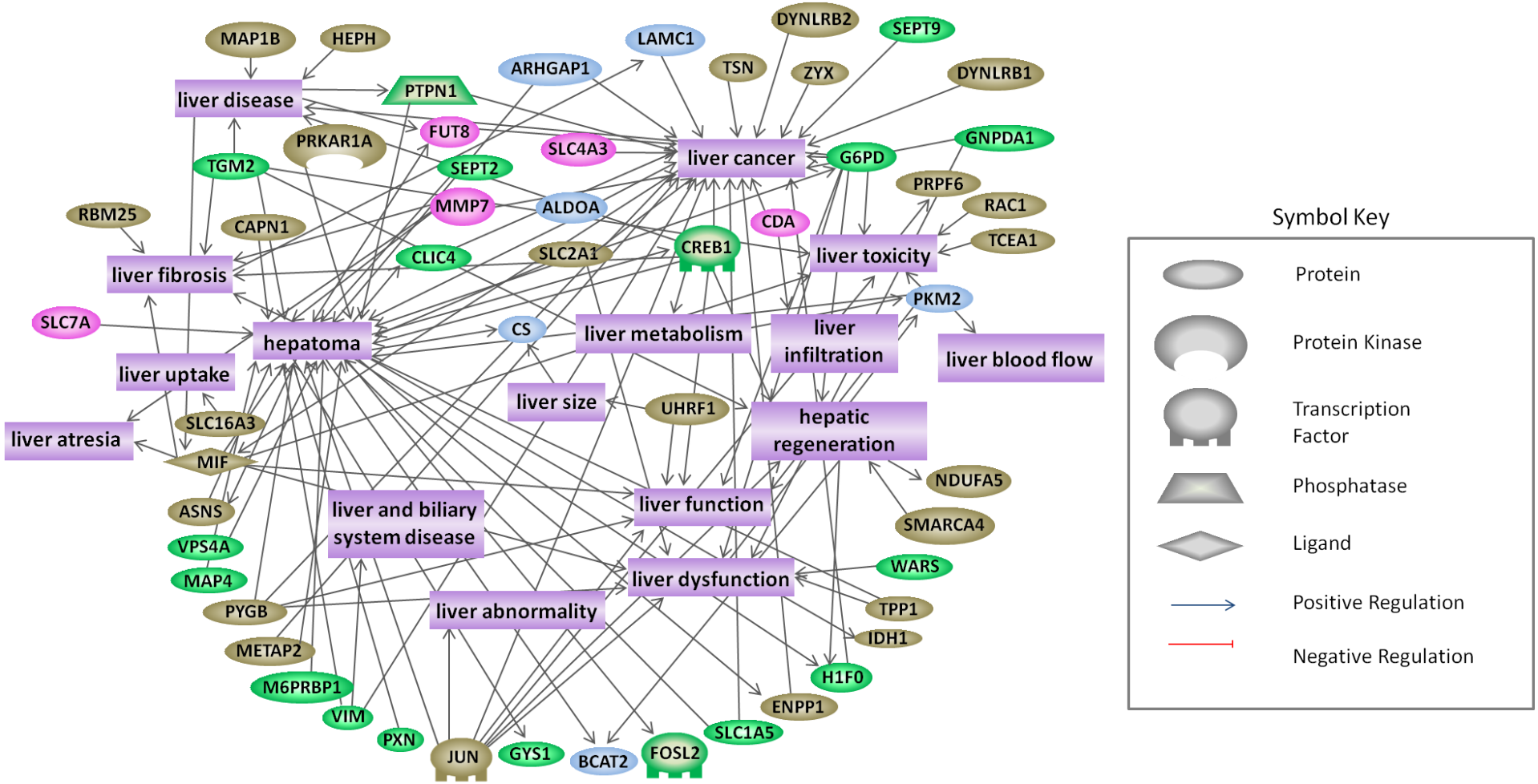


Figure S2. miR-122 and liver related functions and diseases. The association between identified miR-122 targets and liver related processes and diseases was investigated with Pathway Studio 6. In green, genes identified in the proteomics analysis containing at least one miR-122 seed sequence; in brown, genes identified in the proteomics analysis without miR-122 seed sequence; in pink, genes identified by the luciferase analysis; in blue, genes identified by both methods.

Supplementary Figure S3

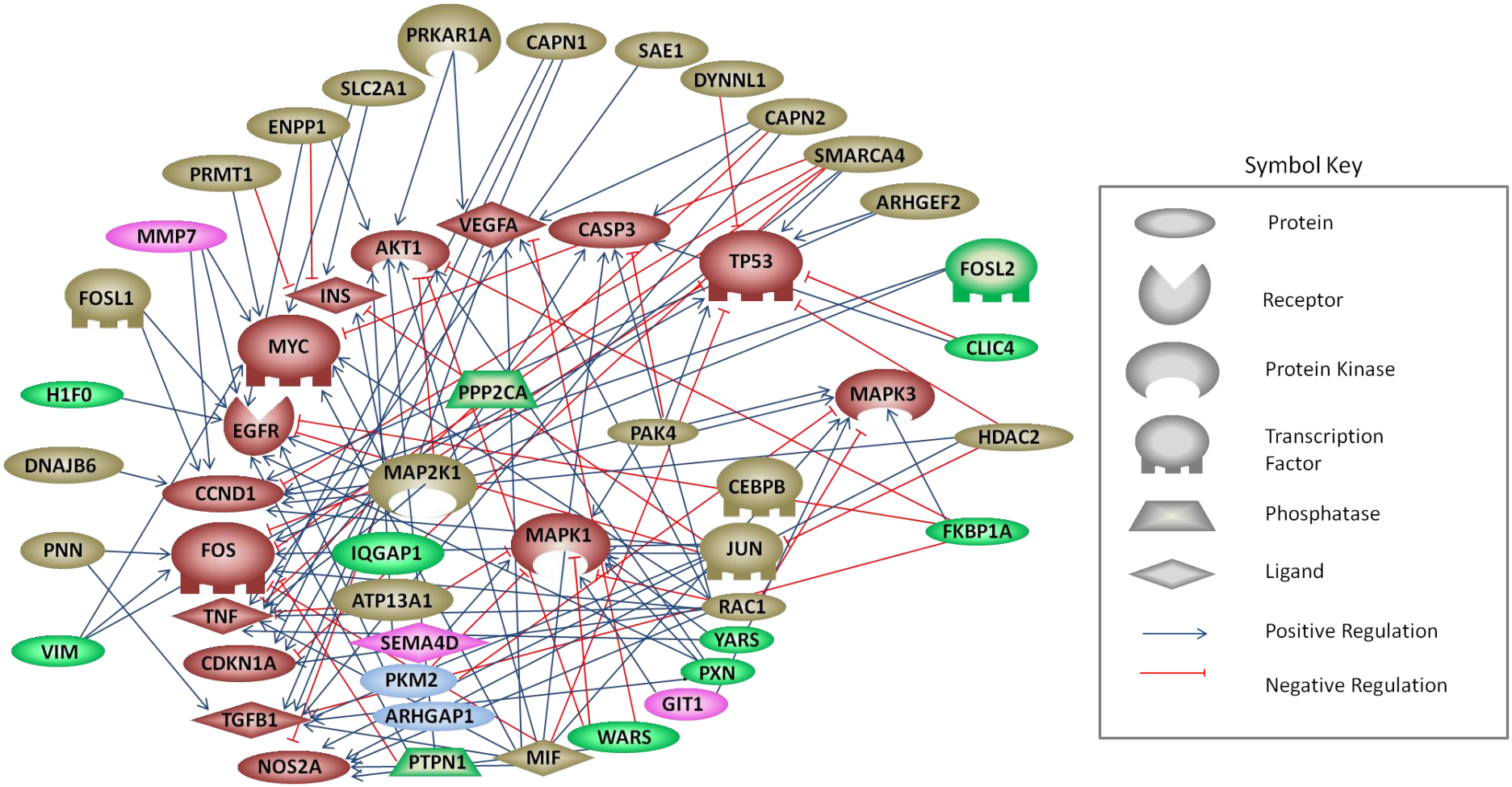


Figure S3. Expanded functional network analysis of miR-122 associated genes. Pathway Studio 6 was used to establish connections among genes that responded to miR-122 and main regulators. In green, genes identified in the proteomics analysis containing at least one miR-122 seed sequence; in brown, genes identified in the proteomics analysis without miR-122 seed sequence; in pink, genes identified by the luciferase analysis; in blue, genes identified by both methods; in red non target genes determined to have a high number of connections to genes that responded to miR-122.

Table S4. Summary of RNA accessibility and local RNA structural features around predicted human miR-122 target sites of wild-type and mutant mRNAs.

mRNAs	Potential Target sites	Hybridization Energy in Seed (kcal/mol)	Moving StbScr	Average of WFS Z_score	Significant RNA Folding
ALDOA	34	-16.9	-0.04	0.24	40-85 (-2.94)
Mutant		-7.9	-0.12	-0.41	
ANKRD13C	373	-17	1.06	-0.56	397-454 (-3.48)
Mutant		-6.8	0.9	-0.48	
G3BP2	804	-17.9	0.68	-0.34	830-869 (-4.74)
Mutant		-3.8	-0.01	-0.18	
G6PC3	65	-16.7	0.26	-0.16	76-135 (-1.98)
Mutant		-2.7	-0.26	0.65	
GNPDA2	373	-16.5	0.8	0.31	395-453 (-2.06)
Mutant		-8.4	0.47	0.03	
RIMS1	904	-15.8	1.2	-0.63	804-868 (-3.12)
Mutant		-9.5	0.86	-0.15	

Table S5. Distribution of identified miR-122 targets according to biological function and diseases.

(A) Genes from Proteomics (down-regulated) enriched for cancer-related cell processes.

Cell Process	Genes
Apoptosis	ACTL6A, ARHGAP1, ASNS, ATXN10, ATXN2L, C14orf153, CACYBP, CAPN1, CAPN2, CAPNS1, CLIC4, COX7C, CSRP1, DCTN1, DPYSL2, DYNLL1, ENPP1, FBLIM1, FKBP1A, FOSL1, FOSL2, FUBP3, H1F0, HDAC2, HEPH, HMOX2, IDH1, IRF2BP2, JUN, KHDRBS1, KIF5B, LAMB1, LAMC1, LETM1, LMNB2, M6PRBP1, MAP1B, MAP2K1, MAP4, METAP2, MIF, MTX1, NDUFA5, NEDD4L, NPEPPS, NUTF2, PAK4, PFAS, PKM2, PLRG1, PNN, PNPLA8, POLR2A, PPP2R4, PRKAR1A, PRMT1, PRPF6, PRPH, PSMD14, PSMD9, PTPN1, PXN, RAC1, RAD21, RBM25, RBM3, RFC4, RPL10, SAE1, SEPT2, SEPT7, SEPT9, SERPINB6, SFRS2, SFRS2IP, SH3BGRL3, SLC16A3, SLC1A5, SLC2A1, SLC2A3, SLC7A1, SMARCA4, SPTBN4, SRP72, TAF10, TMPO, TP53RK, TPD52L2, TPP1, TSN, TUBG1, UBE2M, UBQLN1, UHRF1, UHRF2, UTP14A, VAMP3, VIM, WARS, YARS, ZNRD1, ZYX
Cell Cycle	ARHGAP1, ARHGEF2, ASNS, BAZ1B, BCAT2, CAPN2, CD3EAP, CEP170, CLIC4, COPS8, CSRP1, DDX20, DNAJB6, DYNLL1, EEF1B2, ENPP1, ERH, FKBP1A, FOSL1, FOSL2, GNPNAT1, H1F0, HDAC2, JUN, KHDRBS1, KIF5B, LMNB2, MAP2K1, MAP4, MAPRE1, METAP2, MIF, NEDD4L, NPEPPS, NUDC, NUTF2, PAK4, PKM2, PNN, POLR2A, PPP2CA, PRKAR1A, PRMT1, PSMD14, PSMD9, PTPN1, PXN, RAC1, RAD21, RAE1, RBM3, RCC2, RFC4, SAE1, SAP18, SEPT2, SEPT7, SEPT9, SF3B4, SFRS2, SLC2A1, SMARCA4, SRP72, SUGT1, TAF10, TMPO, TP53RK, TPP1, TSN, TUBG1, UBE2S, UBQLN1, UBQLN2, UHRF1, UHRF2, VIM, VPS4B, ZNRD1, ZYX
Cell Death	ACTL6A, ALDOA, ATXN2L, C14orf153, CAPN1, CAPN2, CAPNS1, CLIC4, DCTN1, DPYSL2, DYNLL1, ENPP1, FIGNL1, FKBP1A, FOSL1, FOSL2, H1F0, HDAC2, IDH1, JUN, KIF5B, LAMC1, LMNB2, MAP1B, MAP2K1, MAP4, MIF, MTX1, NEDD4L, PAK4, PKM2, PNN, PPP2R4, PRPF6, PRPH, PTPN1, PXN, PYGB, RAC1, RAD21, RAE1, RBM3, RDH8, SCAMP3, SEPT9, SFRS2, SFRS2IP, SH3BGRL3, SLC2A1, SMARCA4, SPTBN4, SUGT1, TAX1BP3, TMPO, TPD52L2, TPP1, TSN, UBQLN1, UBQLN2, VIM, VPS4B, ZYX
Cell Differentiation	ALDOA, ARHGAP1, ATXN10, ATXN2L, CACYBP, CAPN1, CAPN2, CAPNS1, CD3EAP, CLIC4, COPS8, CSRP1, DCTN1, DDX20, DPYSL2, ENPP1, FBLIM1, FIGNL1, FKBP1A, FOSL1, FOSL2, FUBP1, FUBP3, GNPNAT1, H1F0, HDAC2, HMOX2, IDH1, IQGAP1, JUN, KHDRBS1, KIF5B, LAMB1, LAMC1, M6PRBP1, MAP1B, MAP2K1, MAP4, METAP2, MIF, MTX1, NDUFA5, NUDC, NUP50, PAK4, PKM2, PNN, PNPLA8, POLR2A, PPP2CA, PRKAR1A, PRMT1, PRPH, PSMD14, PSMD9, PTPN1, PXN, PYGB, RAC1, RAE1, RBM3, RPL10, SAE1, SEPT2, SEPT9, SF3B4, SFRS2, SH3GL1, SLC1A5, SLC2A1, SLC4A3, SMARCA4, SPTBN4, TAF10, TCEA1, TMPO, TNPO2, TOMM20, TOMM7, TPBG, TRIP6, TUBG1, UHRF1, UTP14A, VAMP2, VAMP3, VIM, VPS4B, WARS, ZYX
Cell Proliferation	ALDOA, ARHGAP1, ARHGEF2, ASNS, ATXN2L, BCAT2, CACYBP, CAPN2, CAPNS1, CD3EAP, CLIC4, COL6A3, COPS8, CSRP1, DCTN1, DPYSL2, DR1, DYNLL1,

	DYNLRB1, ENPP1, FBLIM1, FIGNL1, FKBP1A, FOSL1, FOSL2, FUBP1, GNPNAT1, H1F0, HDAC2, HEPH, HMOX2, IQGAP1, JUN, KHDRBS1, KIF5B, LAMB1, LAMC1, LMNB2, M6PRBP1, MAP1B, MAP2K1, MAP4, METAP2, MIF, NDUFA5, NUDC, PAK4, PHPT1, PKM2, PNN, PNPLA8, POLR2A, PPP2CA, PPP2R4, PRKAR1A, PRMT1, PRPF6, PSMD14, PTPN1, PTPN14, PXN, RAC1, RAD21, RAE1, RBM3, RFC4, RPL10, RPL35A, SCAMP3, SEMA4D, SEPT2, SEPT7, SEPT9, SFRS2, SH3GL1, SLC1A5, SLC2A1, SMARCA4, SPTBN4, TAF10, TAX1BP3, TCEA1, TP53RK, TPD52L2, TRIP6, TSN, TUBG1, UBQLN1, UHRF1, UHRF2, UTP14A, VAMP2, VIM, WARS, YARS, ZNRD1, ZYX
Cell Survival	ACTL6A, ARHGAP1, ASNS, ATXN10, BAZ1B, CAPN1, CAPN2, CSRP1, DCTN1, DDX20, DPYSL2, DYNLL1, ENPP1, FBLIM1, FKBP1A, FOSL1, FOSL2, GYS1, H1F0, HDAC2, HMOX2, IDH1, JUN, KHDRBS1, LAMC1, MAP1B, MAP2K1, METAP2, MIF, MTX1, NDUFA5, NEDD4L, PAK4, PKM2, PNN, PNPLA8, POLR2A, PPP2R4, PRKAR1A, PRMT1, PRPF6, PRPH, PSMD14, PSMD9, PTPN1, PXN, RAC1, RAE1, RBM3, RPL35A, RRS1, SAE1, SERPINB6, SLC16A3, SLC1A5, SLC2A1, SMARCA4, SPTBN4, TAF10, TPBG, TPD52L2, TPP1, TUBG1, UBQLN2, UHRF1, VIM, WARS, ZYX
Mitosis	ACTL6A, ACTR1A, ALDOA, ARHGAP1, ARHGEF2, CAPN2, CD3EAP, CEP170, CLIC4, COPS8, CSRP1, DCTN1, DCTN2, DPYSL2, ENPP1, FKBP1A, FOSL1, H1F0, HDAC2, IQGAP1, JUN, KHDRBS1, KIF5B, LMNB2, MAP1B, MAP2K1, MAP4, MAPRE1, MIF, NUDC, NUP35, NUTF2, PKM2, POLR2A, PPP2CA, PRKAR1A, PRMT1, PTPN1, PXN, RAC1, RAD21, RAE1, RBM3, RCC2, RFC4, SAE1, SEPT7, SFRS2, SLC2A1, SMARCA4, SUGT1, TAF10, TMPO, TPP1, TSN, TUBG1, UBE2S, UBQLN1, UBQLN2, UHRF1, UTP14A, VIM, ZYX

(B) Genes from Proteomics (down-regulated) enriched for lipid/cholesterol /glucose metabolism:

Lipid/ Cholesterol/ Glucose Metabolism	ALDOA, CEBPB, CREB1, CS, ENPP1, G6PD, GRSF1, GYS1, H1F0, IDH1, JUN, M6PRBP1, MAP2K1, MIF, MYLK, NCOR2, PKM2, PNPLA8, PTPN1, PXN, RAC1, RPL10, SLC2A1, TOMM20, VAMP2, VIM, VPS4A, WARS

(C) Genes from Proteomics (down-regulated) enriched for Diabetes mellitus:

Diabetes mellitus	ALDOA, ARHGAP1, CEBPB, CREB1, CS, ENPP1, G6PD, GNPDA1, GYS1, KHDRBS1, LAMC1, MIF, NCOR2, NUTF2, PKM2, PRPH, PSMD9, PTPN1, PXN, RAC1, SH3BGRL3, SLC2A1, TGM2, VAMP3, VIM, WARS

Table S6. Summary of associations described in Figure S3.

Targets	Genes
AKT1	ENPP1, FKBP1A, JUN, MAP2K1, MIF, PPP2CA, PRKAR1A, PTPN1, PXN, RAC1, SEMA4D, SLC2A1, WARS
CASP3	CAPN1, CAPN2, CLIC4, JUN, MAP2K1, MIF, PAK4, RAC1
CCND1	ARHGEF2, DNAJB6, FKBP1A, FOSL1, FOSL2, HDAC2, JUN, MAP2K1, MIF, MMP7, RAC1, SMARCA4
CDKN1A	CAPN1, H1F0, HDAC2, JUN, MAP2K1, RAC1, SMARCA4
EGFR	ENPP1, FKBP1A, FOSL1, H1F0, JUN, MAP2K1, MIF, MMP7, PKM2, PTPN1, PXN
FOS	FOSL1, FOSL2, JUN, MAP2K1, MIF, PNN, PTPN1, RAC1, SMARCA4, VIM
INS	ENPP1, JUN, MAP2K1, MIF, PRMT1, PTPN1, RAC1, SLC2A1
MAPK1	ARHGAP1, ATP13A1, GIT1, JUN, MAP2K1, MIF, PAK4, PKM2, PPP2CA, PTPN1, PXN, RAC1, SEMA4D, WARS, YARS
MAPK3	ARHGAP1, FKBP1A, GIT1, MAP2K1, MIF, PAK4, PKM2, PPP2CA, RAC1
MYC	MAP2K1, MIF, MMP7, PRMT1, RAC1, SLC2A1, SMARCA4, VIM
NOS2A	CAPN1, DYNLL1, HDAC2, JUN, MAP2K1, MIF, RAC1, WARS
TGFB1	CAPN1, CAPN2, ENPP1, FKBP1A, JUN, MAP2K1, MIF, PNN, PXN
TNF	ATP13A1, CAPN1, CAPN2, DNAJB6, MAP2K1, MIF, MMP7, PTPN1, RAC1, SAE1, SMARCA4, YARS
TP53	ARHGEF2, CAPN2, CLIC4, DYNLL1, FOSL2, HDAC2, JUN, MAP2K1, MIF, PAK4, RAC1, SAE1, SMARCA4, VIM
VEGFA	CAPN2, JUN, MAP2K1, MIF, PPP2CA, PRKAR1A, RAC1, WARS

Green - genes identified in the proteomics analysis containing at least one miR-122 seed sequence.
Brown - genes identified in the proteomics analysis without miR-122 seed sequence.
Pink - genes identified by the luciferase analysis.
Blue - genes identified by both methods.
Red – additional genes determined to have a high number of connections to genes that responded to miR-122.

Table S7. Pathway Studio 6 reference files referent to Figure 8.

Relation	# of References	MedLine Reference
ALDOA ---> hepatoma	1	6322658:4
ALDOA ---> liver cancer	1	6183164:6
ARHGAP1 ---> hepatoma	2	18996642:3, 16217026:10233
CDA ---> liver cancer	1	9797364:0
CDA ---> liver infiltration	1	17565640:8
CLIC4 ---> hepatoma	1	6310497:0
CLIC4 ---> liver cancer	1	12810631:10040
CREB1 ---> hepatic regeneration	8	9199295:9, 12167624:10021, 15084473:10447, 16415092:10248, 9077547:10856, 9727068:11162, 11447268:10007, 15169930:10342
CREB1 ---> hepatoma	7	14973073:4, 17277233:10045, 131103658:10216, 8886030:3, 1824683:0, 11553511:10201, 131103658:11216
CREB1 ---> liver cancer	1	14973073:10007
CREB1 ---> liver diseases	1	9077542:10017
CREB1 ---> liver fibrosis	1	16293380:1
CREB1 ---> liver function	1	9614149:10034
CREB1 ---> liver metabolism	2	16415092:10247, 16497994:10198
FUT8 ---> hepatoma	5	10232614:10236, 9455807:8, 12499776:4, 10580126:6, 14559815:10032
FUT8 ---> liver cancer	2	16236725:10020, 10232614:10119
G6PD ---> hepatic regeneration	1	8645338:3
G6PD ---> hepatoma	4	6085329:0, 7378065:4, 10567366:10164, 10728670:10077
G6PD ---> liver cancer	3	11856576:10, 3230041:0, 15563548:10210
G6PD ---> liver diseases	1	11903742:3
G6PD ---> liver dysfunction	2	1265775:6, 1955295:6
G6PD ---> liver function	3	3884006:0, 9228459:3, 11110846:10149
G6PD ---> liver toxicity	2	10918522:7, 61145:5
GNPDA1 ---> liver cancer	1	18549886:4
GNPDA1 ---> liver dysfunction	1	17785839:3
H1F0 ---> liver cancer	1	1747124:8
hepatic regeneration ---> BCAT2	1	16194549:3
hepatic regeneration ---> H1F0	1	1510879:3
hepatoma ---> BCAT2	1	16194549:5
hepatoma ---> CS	1	12198131:10153
hepatoma ---> FOSL2	1	9670973:10293
hepatoma ---> GYS1	1	11882651:0
hepatoma ---> H1F0	1	6279172:0
hepatoma ---> PXN	1	12941817:10188
LAMC1 ---> liver cancer	1	9403717:9
liver cancer ---> ARHGAP1	5	19440389:0, 16322093:10241, 15976441:10419, 16024604:10409, 15297371:10219
liver diseases ---> FUT8	2	10052590:7, 14559815:10226
liver diseases ---> PTPN1	1	17030176:8
liver dysfunction ---> PKM2	1	18501810:1
liver fibrosis ---> LAMC1	1	8611150:0
liver metabolism ---> PKM2	1	6780202:11
liver size ---> CS	1	18840537:5
M6PRBP1 ---> hepatoma	1	16489205:10122
MAP1B ---> liver diseases	1	9659171:1
MAP4 ---> hepatoma	1	9234808:7
MMP7 ---> hepatoma	1	16474169:10152
MMP7 ---> liver atresia	1	15696117:10
MMP7 ---> liver cancer	12	11408348:10086, 14647460:1, 11394006:3, 10089944:1, 9626346:9, 10357559:6, 19421758:6, 19421758:5, 16267009:10248, 14744783:10507 <more data available...>
MMP7 ---> liver fibrosis	2	15696117:8, 16226986:9

PKM2 ---> hepatoma	1	9224771:10156
PKM2 ---> liver blood flow	1	929878:0
PKM2 ---> liver function	1	7338204:3
PKM2 ---> liver toxicity	2	2506674:6, 10092053:5
PTPN1 ---> hepatoma	5	14976221:5, 7544790:1, 15561934:10014, 16269466:10200, 11297536:10020
PTPN1 ---> liver cancer	1	12554649:10109
SEPT2 ---> hepatoma	1	19165576:4
SEPT9 ---> liver cancer	3	19024104:9, 19024104:14, 19024104:2
SLC1A5 ---> hepatoma	3	14563674:10036, 12381519:10016, 11533296:10036
SLC1A5 ---> liver cancer	2	12381519:10308, 14563674:10020
SLC4A3 ---> liver cancer	1	17018624:10015
SLC7A1 ---> hepatoma	1	11406552:10150
TGM2 ---> hepatic regeneration	3	18008394:1, 16707846:1, 10801782:10020
TGM2 ---> hepatoma	2	10801782:10209, 16357165:10327
TGM2 ---> Liver Diseases	3	16279902:1, 15638421:3, 154101372:11360
TGM2 ---> liver fibrosis	3	17708605:0, 12651621:5, 14752834:0
TGM2 ---> liver toxicity	4	12651621:7, 16763214:10392, 154101372:10351, 9864372:10190
VIM ---> hepatoma	1	2474862:1
VIM ---> liver and biliary system diseases	1	1285354:7
VIM ---> liver cancer	3	8853042:0, 15205331:10182, 15604094:10199
VPS4A ---> hepatoma	1	16920176:2
WARS ---> liver dysfunction	1	19837917:0

Table S8. Pathway Studio 6 reference files referent to Figure 9

Relation	Type	# of References	MedLine Reference
CEBPB --+> FOSL2	Expression	1	12896981:10142
CEBPB --+> JUN	Expression	2	12446786:10038, 12896981:10142
CREB1 --+> CSRP1	Expression	1	15917302:10108
CREB1 --+> JUN	Expression	9	17482134:2, 14741741:7, 1331648:10, 9658406:10230, 9448004:10310, 11832338:10291, 11832337:10231, 12702768:10175, 164100930:11026
CREB1 --+> MIF	Expression	3	9605932:9, 17142669:10207, 17182862:10162
CREB1 --+> SLC2A3	Expression	1	15054091:5
CSRP1 --+> CDA	Expression	1	1327747:1
CSRP1 --+> CREB1	Expression	1	12704194:1
FOSL1 --+> JUN	Expression	3	11019821:6, 14625389:10073, 16199889:10267
FOSL2 --+> FOSL1	Expression	2	13679379:9, 160101907:11045
FOSL2 --+> JUN	Expression	1	10359014:10141
G6PD --+> JUN	Expression	1	12446786:10146
GIT1 --+> RAC1	Expression	1	162106263:11085
IQGAP1 --+> RAC1	Expression	3	14699063:10227, 12900413:10169, 12070151:10162
JUN --+> CEBPB	Expression	6	12215258:7, 12446786:10038, 15299028:10028, 12215258:5, 12896981:10284, 14684744:10276
JUN --+> CREB1	Expression	1	11784328:10188
JUN --+> FOSL1	Expression	8	13679379:9, 15806162:10, 10784353:7, 11280724:10146, 7721415:10046, 16199889:10218, 15528491:10245, 160101907:11045
JUN --+> FOSL2	Expression	4	7862446:7, 15618352:10037, 160101907:11045, 160102519:11091
JUN --+> MMP7	Expression	1	11158322:10203
JUN --+> VIM	Expression	8	16831196:5, 14654785:8, 17215518:10591, 16214953:10430, 18206971:10268, 160104971:10177, 160105031:10239, 10978519:5
MAP2K1 --+> FOSL1	Expression	6	9858557:6, 11756554:10254, 9858556:10167, 15615716:10366, 160105102:10111, 15615716:10245
MAP2K1 --+> JUN	Expression	10	8570188:5, 9858557:6, 11522656:10132, 12414794:10049, 16446383:10806, 12359764:10393, 16410316:10254, 160105102:10111, 9687508:10194, 12060661:10148
MAP2K1 --+> SLC2A1	Expression	5	10085148:5, 11108276:10125, 12554784:10105, 11279172:10108, 10085148:10120
MAP2K1 --+> VIM	Expression	3	10359013:10038, 11029295:10035, 12183438:10250
MIF --+> JUN	Expression	1	12881477:10203
MIF ---| JUN	Expression	1	19478200:6
PNN --+> JUN	Expression	1	10542237:10048
PRPH --+> VIM	Expression	1	131101960:11066
PTPN1 --+> PXN	Expression	1	10625692:10025
PXN --+> GIT1	Expression	1	10938112:10262
PXN --+> MYLK	Expression	1	14576084:10208
PXN --+> RAC1	Expression	1	110003439:11058
PXN ---| ARHGAP1	Expression	1	110002351:11114
PXN ---| JUN	Expression	1	16740738:10145
RAC1 --+> JUN	Expression	12	10856308:6, 9488437:10243, 12954641:10280, 11283263:10024, 10523665:10221, 9520479:10032, 163100271:10135, 160104327:10072, 9448004:10164, 12810717:10193 <more data available...>
RAC1 --+> PXN	Expression	2	15308668:4, 110003439:11029
SLC2A1 --+> SLC2A3	Expression	3	11739520:10206, 11080093:10171, 131101295:11125
SLC2A3 --+> SLC2A1	Expression	1	7474941:11
SMARCA4 --+> JUN	Expression	1	14673169:10135
ARHGAP6 --+> ARHGAP1	Regulation	1	15492870:0
CEBPB --+> CREB1	Regulation	1	12468538:10292

CREB1 --+> CSRP1	Regulation	1	15917302:10198
CREB1 ---| PTPN1	Regulation	1	12649327:10215
GIT1 --+> RAC1	Regulation	2	10938112:10246, 17122362:10030
JUN --+> MAP2K1	Regulation	1	9687508:10194
MAP2K1 --+> CEBPB	Regulation	1	163101697:11224
MAP2K1 --+> CREB1	Regulation	4	11123304:10176, 11532940:10162, 131103539:11132, 131104294:11109
MAP2K1 --+> FOSL2	Regulation	1	9858556:10228
MAP2K1 --+> MYLK	Regulation	1	15371522:10167
MAP2K1 --+> SLC2A1	Regulation	1	11279172:10108
METAP2 --+> JUN	Regulation	3	16984890:10069, 165100836:10042, 165100836:11042
MIF ---| JUN	Regulation	1	12626594:10017
PAK4 ---| RAC1	Regulation	1	15827085:10266
PTPN1 --+> CREB1	Regulation	1	12649327:10215
PXN --+> RAC1	Regulation	2	18392556:9, 18392556:7
RAC1 --+> CREB1	Regulation	1	11290522:8
RAC1 --+> PXN	Regulation	2	18392556:7, 18813807:6
RAC1 ---| MYLK	Regulation	1	15123662:10052
SEMA4D --+> RAC1	Regulation	1	164102010:11054
TRIP6 ---| RAC1	Regulation	1	18348201:10
VAMP3 ---| SLC2A1	Regulation	1	10564264:10221

Table S9. Pathway Studio 6 reference files referent to Figure S2.

Relation	# of References	MedLine Reference
ALDOA ---> hepatoma	1	6322658:4
ALDOA ---> liver cancer	1	6183164:6
ARHGAP1 ---> hepatoma	2	18996642:3, 16217026:10233
ARHGAP1 ---> liver cancer	8	19826001:4, 10649492:2, 17190795:10446, 15976441:10872, 16951145:10262, 16204057:10299, 16024604:10328, 15126338:10198
ASNS ---> hepatoma	1	12368390:10037
CAPN1 ---> hepatoma	1	7539218:9
CDA ---> liver cancer	1	9797364:0
CDA ---> liver infiltration	1	17565640:8
CLIC4 ---> hepatoma	1	6310497:0
CLIC4 ---> liver cancer	1	12810631:10040
CREB1 ---> hepatic regeneration	8	9199295:9, 12167624:10021, 15084473:10447, 16415092:10248, 9077547:10856, 9727068:11162, 11447268:10007, 15169930:10342
CREB1 ---> hepatoma	7	14973073:4, 17277233:10045, 131103658:10216, 8886030:3, 1824683:0, 11553511:10201, 131103658:11216
CREB1 ---> liver cancer	1	14973073:10007
CREB1 ---> liver diseases	1	9077542:10017
CREB1 ---> liver fibrosis	1	16293380:1
CREB1 ---> liver function	1	9614149:10034
CREB1 ---> liver metabolism	2	16415092:10247, 16497994:10198
DYNLRB1 ---> liver cancer	1	11750132:4
DYNLRB2 ---> liver cancer	1	11750132:4
ENPP1 ---> liver cancer	1	12441313:10176
FUT8 ---> hepatoma	5	10232614:10236, 9455807:8, 12499776:4, 10580126:6, 14559815:10032
FUT8 ---> liver cancer	2	16236725:10020, 10232614:10119
G6PD ---> hepatic regeneration	1	8645338:3
G6PD ---> hepatoma	4	6085329:0, 7378065:4, 10567366:10164, 10728670:10077
G6PD ---> liver cancer	3	11856576:10, 3230041:0, 15563548:10210
G6PD ---> liver diseases	1	11903742:3
G6PD ---> liver dysfunction	2	1265775:6, 1955295:6
G6PD ---> liver function	3	3884006:0, 9228459:3, 11110846:10149
G6PD ---> liver toxicity	2	10918522:7, 61145:5
GNPDA1 ---> liver cancer	1	18549886:4
GNPDA1 ---> liver dysfunction	1	17785839:3
H1F0 ---> liver cancer	1	1747124:8
hepatic regeneration ---> BCAT2	1	16194549:3
hepatic regeneration ---> H1F0	1	1510879:3
hepatic regeneration ---> NDUFA5	1	131101087:11169
hepatoma ---> BCAT2	1	16194549:5
hepatoma ---> CREB1	1	10629058:10334
hepatoma ---> CS	1	12198131:10153
hepatoma ---> ENPP1	2	6102383:0, 201068:2
hepatoma ---> FOSL2	1	9670973:10293
hepatoma ---> FUT8	3	9455807:8, 10580126:6, 10232614:10013
hepatoma ---> G6PD	3	179698:4, 2849577:0, 2846237:0
hepatoma ---> GYS1	1	11882651:0
hepatoma ---> H1F0	1	6279172:0
hepatoma ---> IDH1	1	16636309:10245
HEPH ---> liver diseases	1	11897618:10123
JUN ---> hepatic regeneration	22	9727068:10804, 14761977:10355, 15870258:10580, 15231690:10031, 11927562:10040, 15242989:10180, 15523045:10021, 16912279:10006, 16858288:12, 10395942:7 <more data available...>

		17573710:3, 12135113:5, 15075337:10269, 11120824:10231, 11739718:10137, 7728992:10229, 15059897:10196,
JUN ---> hepatoma	8	131104515:11363
JUN ---> liver abnormality	2	12771026:10030, 160101903:11048
		15927205:0, 15492232:10024, 16061660:10038, 16618719:10201, 14555611:10038, 10406935:10129, 16912279:10145, 12168099:1,
JUN ---> liver cancer	21	10960769:2, 7510474:7 <more data available...>
JUN ---> liver dysfunction	1	12853483:10228
JUN ---> liver function	1	11927562:10034
JUN ---> liver toxicity	6	8827825:5, 18182393:5, 7943321:2, 12842828:10237, 11927562:10041, 11470752:10015
LAMC1 ---> liver cancer	1	9403717:9
liver cancer ---> MIF	7	15770393:9, 19009023:2, 19254456:0, 16707461:10279, 15319268:10409, 10878343:10155, 15466195:10250
liver diseases ---> FUT8	2	10052590:7, 14559815:10226
liver diseases ---> MIF	1	11391079:30
liver diseases ---> PTPN1	1	17030176:8
liver dysfunction ---> PKM2	1	18501810:1
liver fibrosis ---> LAMC1	1	8611150:0
liver metabolism ---> PKM2	1	6780202:11
liver size ---> CS	1	18840537:5
liver toxicity ---> PRPF6	1	12960737:14
M6PRBP1 ---> hepatoma	1	16489205:10122
MAP1B ---> liver diseases	1	9659171:1
MAP4 ---> hepatoma	1	9234808:7
METAP2 ---> hepatoma	1	14633714:10027
METAP2 ---> liver cancer	1	17957724:6
MIF ---> hepatoma	1	6752005:1
MIF ---> liver atresia	1	16385258:2
MIF ---> liver dysfunction	2	362901:1, 12695559:6
MIF ---> liver fibrosis	1	14706927:1
MIF ---> liver function	1	11086030:10247
MIF ---> liver toxicity	8	12051698:7, 16584398:1, 3334006:0, 16522786:1, 15879138:10662, 17909632:10356, 133100951:11042, 133100787:11042
MMP7 ---> hepatoma	1	16474169:10152
MMP7 ---> liliary atresia	1	15696117:10
MMP7 ---> liver cancer	12	11408348:10086, 14647460:1, 11394006:3, 10089944:1, 9626346:9, 10357559:6, 19421758:6, 19421758:5, 16267009:10248, 14744783:10507 <more data available...>
MMP7 ---> liver fibrosis	2	15696117:8, 16226986:9
PKM2 ---> hepatoma	1	9224771:10156
PKM2 ---> liver blood flow	1	929878:0
PKM2 ---> liver function	1	7338204:3
PKM2 ---> liver toxicity	2	2506674:6, 10092053:5
PRKAR1A ---> hepatoma	3	2342460:0, 1889088:0, 15331577:10069
PTPN1 ---> hepatoma	5	14976221:5, 7544790:1, 15561934:10014, 16269466:10200, 11297536:10020
PTPN1 ---> liver cancer	1	12554649:10109
PXN ---> hepatoma	3	18997105:7, 18695939:3, 15191880:10347
PYGB ---> hepatoma	3	222482:0, 2337960:4, 2158961:0
PYGB ---> liver cancer	2	16525674:8, 7927248:10
PYGB ---> liver dysfunction	1	2783160:5
PYGB ---> liver function	1	2157500:0
RAC1 ---> liver toxicity	1	12488232:10037
RBM25 ---> liver fibrosis	3	19476212:3, 19476212:1, 19476212:2
SEPT2 ---> hepatoma	1	19165576:4
SEPT9 ---> liver cancer	3	19024104:9, 19024104:14, 19024104:2

SLC16A3 ---> liver uptake	1	9458768:8
SLC1A5 ---> hepatoma	3	14563674:10036, 12381519:10016, 11533296:10036
SLC1A5 ---> liver cancer	2	12381519:10308, 14563674:10020
SLC2A1 ---> hepatoma	1	10400647:10179
SLC2A1 ---> liver cancer	1	18250992:0
SLC2A1 ---> liver dysfunction	1	16374780:5
SLC4A3 ---> liver cancer	1	17018624:10015
SLC7A1 ---> hepatoma	1	11406552:10150
SMARCA4 ---> hepatic regeneration	1	11870262:10163
TCEA1 ---> liver toxicity	1	7151718:0
TGM2 ---> hepatic regeneration	3	18008394:1, 16707846:1, 10801782:10020
TGM2 ---> hepatoma	2	10801782:10209, 16357165:10327
TGM2 ---> liver diseases	3	16279902:1, 15638421:3, 154101372:11360
TGM2 ---> liver fibrosis	3	17708605:0, 12651621:5, 14752834:0
TGM2 ---> liver toxicity	4	12651621:7, 16763214:10392, 154101372:10351, 9864372:10190
TPP1 ---> liver dysfunction	1	19760908:3
TPP1 ---> liver fibrosis	1	8986240:6
TSN ---> liver cancer	1	12532453:0
UHRF1 ---> hepatic regeneration	2	17242348:10146, 17242348:10122
UHRF1 ---> liver size	1	17242348:10005
VIM ---> hepatoma	1	2474862:1
VIM ---> liver and biliary system diseases	1	1285354:7
VIM ---> liver cancer	3	8853042:0, 15205331:10182, 15604094:10199
VPS4A ---> hepatoma	1	16920176:2
WARS ---> liver dysfunction	1	19837917:0
ZYX ---> liver cancer	1	16680155:9

Table S10. Pathway Studio 6 reference files referent to Figure S3.

Relation	Type	# of References	MedLine Reference
ARHGEF2 --+> CCND1	Expression	2	19730435:7, 19730435:6
ARHGEF2 --+> TP53	Expression	1	18394899:2
ATP13A1 ---| TNF	Expression	1	15472516:8
CAPN1 --+> CDKN1A	Expression	1	11807952:5
CAPN1 --+> TNF	Expression	1	19318376:8
CAPN2 ---| TNF	Expression	1	15501405:9
DYNLL1 ---| NOS2A	Expression	1	9368577:10183
DYNLL1 ---| TP53	Expression	1	15611139:8
ENPP1 --+> EGFR	Expression	1	7909745:0
FOSL1 --+> CCND1	Expression	12	15143165:1, 16061661:10452, 19020303:10196, 17145963:10341, 16966382:10288, 16266986:10306, 16151051:10225, 15374964:10231, 17114644:10279, 160101467:10268 <more data available...>
FOSL2 --+> CCND1	Expression	1	12052899:10242
FOSL2 --+> FOS	Expression	1	8815874:10200
GIT1 --+> MAPK1	Expression	1	14701758:10188
GIT1 --+> MAPK3	Expression	1	14701758:10188
HDAC2 --+> CCND1	Expression	1	16985053:6
HDAC2 --+> NOS2A	Expression	1	15356168:10235
JUN --+> CCND1	Expression	62	15016640:10192, 16618766:10304, 12097301:10446, 10340949:10218, 15574764:10368, 16051644:10340, 12702582:10250, 15172983:10262, 16778192:10235, 12969971:10206 <more data available...>
JUN --+> FOS	Expression	44	12392213:10, 14561487:3, 10858536:3, 8934539:9, 15668325:10231, 11927562:10133, 10598580:10037, 10359014:10026, 10464319:10025, 10026157:10083 <more data available...>
JUN --+> MAPK1	Expression	1	12738796:10067
JUN --+> NOS2A	Expression	3	9776360:8, 11087760:10129, 10215633:10203
JUN --+> TGFB1	Expression	1	19151758:5
JUN ---| INS	Expression	15	8264634:2, 7862133:9, 9328343:10023, 16549443:10067, 14578290:10307, 11799123:10251, 14557546:10112, 12011047:10231, 14578290:10, 1310538:5 <more data available...>
MAP2K1 --+> CCND1	Expression	44	12085346:8, 10504490:8, 10419534:10214, 9618377:10138, 11751416:10245, 17409427:10035, 9891068:10218, 10066798:10283, 10910097:10214, 11029295:10044 <more data available...>
MAP2K1 --+> CDKN1A	Expression	7	17321721:5, 12867429:4, 10748202:10229, 10570262:10036, 10913180:10186, 15284233:10243, 160100315:11072
MAP2K1 --+> EGFR	Expression	1	12556561:10305
MAP2K1 --+> MYC	Expression	3	9891045:2, 10037749:10032, 12644583:10325
MAP2K1 --+> TNF	Expression	1	18187448:0
MAP2K1 --+> VEGFA	Expression	11	16327989:5, 12440226:4, 12482909:10318, 15705901:10334, 15899904:10245, 16849522:10364, 10667605:10221, 16103054:10256, 16327989:4, 15665520:7 <more data available...>
MAP2K1 ---| FOS	Expression	1	18473919:7
MAP2K1 ---| NOS2A	Expression	4	11131279:6, 11171562:10227, 14581482:10099, 9794431:10349
MAP2K1 ---| TP53	Expression	1	14729959:10104
MIF --+> CCND1	Expression	11	12297513:3, 15840582:1, 17142775:10717, 16982923:10310, 17210698:10281, 17045821:10387, 16835407:10501, 19938192:10, 12297513:10171, 16982923:10024 <more data available...>
MIF --+> INS	Expression	2	16873699:10118, 11086030:10182
MIF --+> MAPK1	Expression	1	18355352:11

MIF --+> MYC	Expression	1	2407738:3
MIF --+> NOS2A	Expression	5	16697942:12, 18064633:5, 15790730:10173, 9126926:10203, 10956264:10015
MIF --+> TNF	Expression	16	18064633:5, 15276025:9, 18081874:0, 9892616:10099, 9725234:10107, 14633847:10182, 12480958:10054, 11994438:10014, 10727256:10029, 12946935:10245 <more data available...>
MIF --+> VEGFA	Expression	13	12925952:10, 17854909:5, 18022162:7, 19009023:6, 18097062:10170, 16709839:10406, 16982923:10323, 17114481:10759, 17620429:10333, 16618748:10279 <more data available...>
PAK4 --+> TP53	Expression	1	16227603:10197
PNN --+> FOS	Expression	1	10542237:10048
PNN --+> TGFB1	Expression	1	160102075:11118
PRKAR1A --+> VEGFA	Expression	1	17442730:3
PRMT1 --+> MYC	Expression	1	12704081:10121
PTPN1 --+> AKT1	Expression	1	11280739:10216
PTPN1 --+> EGFR	Expression	1	16391241:10174
PTPN1 ---| FOS	Expression	1	10652360:10
PXN --+> EGFR	Expression	1	18172305:2
PXN --+> MAPK1	Expression	2	16014377:10126, 17118962:10221
RAC1 --+> CDKN1A	Expression	1	12052868:10227
RAC1 --+> FOS	Expression	9	15976459:13, 9584171:10309, 10849427:10213, 15078869:10200, 160104327:10072, 11150863:3, 15976459:8, 9755162:10037, 9488437:10176
RAC1 --+> MYC	Expression	1	16943436:10253
RAC1 --+> NOS2A	Expression	1	160102349:11074
RAC1 --+> VEGFA	Expression	3	16440308:5, 12068011:10032, 16230378:10135
RAC1 ---| MAPK1	Expression	1	16798728:8
RAC1 ---| MAPK3	Expression	1	16798728:8
SAE1 --+> TNF	Expression	1	11035048:10186
SLC2A1 --+> INS	Expression	3	1468312:8, 11562503:10035, 10753892:10038
SMARCA4 --+> TNF	Expression	1	19556365:8
SMARCA4 --+> TP53	Expression	3	11950834:10104, 14673169:10167, 11108719:10178
SMARCA4 ---| CCND1	Expression	1	18239461:3
SMARCA4 ---| CDKN1A	Expression	2	15286705:6, 11950834:10177
SMARCA4 ---| FOS	Expression	22	12372840:4, 19081374:2, 10082538:3, 16717115:10535, 17666433:10427, 16923960:10681, 14657023:10250, 11238380:10038, 12058073:10309, 11734557:10027 <more data available...>
SMARCA4 ---| MYC	Expression	1	14729964:10281
WARS --+> NOS2A	Expression	2	15526382:9, 15526382:8
ATP13A1 ---| MAPK1	Regulation	1	15472516:4
CAPN1 --+> TGFB1	Regulation	1	9070300:0
CAPN2 --+> CASP3	Regulation	8	11124942:4, 16405498:10313, 9576764:11094, 16412092:10203, 17166174:10195, 17711427:10021, 131102803:10231, 11124942:10010
CAPN2 --+> TGFB1	Regulation	1	9070300:0
CAPN2 --+> VEGFA	Regulation	1	16484616:10198
CLIC4 --+> CASP3	Regulation	1	14610078:10210
CLIC4 ---| TP53	Regulation	1	11997498:4
DNAJB6 --+> CCND1	Regulation	1	18373498:7
ENPP1 --+> AKT1	Regulation	1	16452497:10176
ENPP1 ---| INS	Regulation	13	16865358:0, 18948963:0, 19046915:0, 11027689:10020, 18071025:10025, 18184924:10152, 18426862:10189, 15793263:10124, 16607460:0, 15209435:3 <more data available...>
FKBP1A --+> MAPK3	Regulation	1	11784851:10257
FKBP1A ---| AKT1	Regulation	2	10455142:10029, 12084817:10110

FKBP1A ---| EGFR	Regulation	2	9545268:3, 9545268:2
FKBP1A ---| TGFB1	Regulation	3	16720724:0, 10566134:14, 15047702:10393
FOSL1 --+> EGFR	Regulation	2	18288638:8, 18288638:7
H1F0 --+> EGFR	Regulation	1	7961591:5
HDAC2 ---| TP53	Regulation	1	10777477:2
JUN ---| EGFR	Regulation	1	18302141:4
MAP2K1 --+> CASP3	Regulation	4	17531086:11, 17108320:10139, 11101507:10124, 14688022:10024
MAP2K1 --+> INS	Regulation	3	12417588:10072, 12364332:10207, 9645686:10225
MAP2K1 --+> MAPK3	Regulation	2	18533112:3, 19671683:7
MAP2K1 --+> TGFB1	Regulation	4	10862759:5, 11401854:10148, 12457461:5, 16547158:7
MAP2K1 ---| AKT1	Regulation	1	18059341:4
MIF --+> AKT1	Regulation	6	17310986:7, 17911626:10158, 12881477:10157, 16317091:10030, 18355352:11, 19737008:8
MIF --+> CASP3	Regulation	1	18439909:4
MIF --+> EGFR	Regulation	1	18474653:7
MIF --+> MAPK3	Regulation	9	17142775:1, 17045821:3, 19136583:6, 17316137:2, 11978785:10009, 11751895:10308, 12782713:10030, 16186482:10015, 11773615:10080
MIF --+> TGFB1	Regulation	1	15187193:14
MIF ---| FOS	Regulation	1	15371880:8
MIF ---| TP53	Regulation	36	12239594:2, 18493321:8, 18502749:9, 16709839:6, 12538581:0, 11756671:4, 15968736:0, 16125329:1, 10562311:10066, 16982923:10421 <more data available...>
MMP7 --+> CCND1	Regulation	1	10825188:10234
MMP7 --+> EGFR	Regulation	1	15623600:2
MMP7 --+> MYC	Regulation	1	10825188:10234
PAK4 --+> MAPK1	Regulation	4	19435802:4, 16227603:10265, 16227603:10273, 11313478:10134
PAK4 --+> MAPK3	Regulation	1	19435802:4
PAK4 ---| CASP3	Regulation	2	14560027:10123, 11278822:10224
PKM2 --+> EGFR	Regulation	1	3131034:5
PKM2 ---| MAPK1	Regulation	3	12446705:10239, 10856308:10123, 10915788:10270
PKM2 ---| MAPK3	Regulation	1	12446705:10239
PRKAR1A --+> AKT1	Regulation	1	17442730:4
PRMT1 ---| INS	Regulation	1	19467247:7
PXN --+> AKT1	Regulation	1	110003451:11051
PXN --+> TGFB1	Regulation	1	12595498:10079
RAC1 --+> CASP3	Regulation	4	16092944:5, 15919807:10015, 11102528:10268, 131104039:11003
RAC1 --+> TNF	Regulation	3	18218673:3, 18218673:6, 18218673:7
SEMA4D --+> AKT1	Regulation	5	16799460:2, 16055703:10097, 17875711:10269, 16055703:10192, 16702230:10329
SEMA4D --+> MAPK1	Regulation	3	17688882:8, 110003426:10104, 110003426:11104
SLC2A1 --+> MYC	Regulation	1	10823814:10082
VIM --+> FOS	Regulation	1	3801408:3
VIM --+> MYC	Regulation	1	3801408:3
VIM --+> TP53	Regulation	1	3801408:3
WARS ---| AKT1	Regulation	3	15579907:10121, 14630953:10092, 14630953:10093
WARS ---| MAPK1	Regulation	2	15579907:6, 14630953:10071
WARS ---| VEGFA	Regulation	3	11773626:10102, 14630953:10021, 11956181:10106
YARS --+> MAPK1	Regulation	1	10395653:10101
YARS --+> TNF	Regulation	2	12391224:10217, 10102815:10028