JBC Papers in Press. Published on March 14, 2011 as Manuscript M110.196451
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M110.196451

A Two-Tiered Approach Identifies a Network of Cancer and Liver Diseases Related Genes Regulated by miR-122

Daniel R. Boutz[1#]*, Patrick Collins[2#], Uthra Suresh[3#], Mingzhu Lu[4], Cristina M. Ramírez[5], Carlos Fernández-Hernando[5], Yufei Huang[3,4], Raquel de Sousa Abreu[3], Shu-Yun Le[6], Bruce A. Shapiro[6], Angela M. Liu[7,8], John M. Luk[7,8,9,10], Shelley Force Aldred[2], Nathan Trinklein[2], Edward M. Marcotte[1,11] and Luiz O. F. Penalva[3,6,12*]

Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712,[1] Switchgear Genomics, Menlo Park, California 94025,[2] Children's Cancer Research Institute, University of Texas Health Science Center, San Antonio, Texas 78229,[3] Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, Texas 78249,[4] Department of Medicine, Leon H. Charney Division of Cardiology, New York University School of Medicine, New York, New York 10016,[5] Center for Cancer Research Nanobiology Program, National Cancer Institute-Frederick, Frederick, Maryland 21702,[6] Department of Surgery, The University of Hong Kong, Queen Mary Hospital, Hong Kong,[7] Department of Pharmacology, National University of Singapore, Singapore,[8] Department of Surgery, National University of Singapore, Singapore,[9] Cancer Science Institute, National University of Singapore, Singapore,[10] Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712,[11] Department of Cellular and Structural Biology, University of Texas Health Science Center, San Antonio, Texas 78229[12]

# These authors contributed equally to this work

* Corresponding authors

Correspondence to: Daniel R. Boutz[1#]* Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, 2500 Speedway, MBB 3.210, Austin, TX 78712, USA. Tel.: +1-512-232-3919; Fax: +1 512 471 2149; E-mail: dboutz@mail.utexas.edu

Correspondence to: Luiz O. F. Penalva[3,6,8]* Children's Cancer Research Institute, University of Texas Health Center, San Antonio, TX 78229, USA. Tel.: +1-210-562-9049; Fax: +1-210-562-9014; E-mail: penalva@uthscsa.edu

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. Amongst 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.

Copyright 2011 by The American Society for Biochemistry and Molecular Biology, Inc.

Downloaded from www.jbc.org by guest, on March 21, 2011

MicroRNAs (miRNAs) are small (20-24 nt) endogenously expressed non-coding 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 (RISC) to mRNAs by Watson-Crick base pairing between the miRNA and target mRNA (reviewed in (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 (4), and many questions remain, however it is clear that sequence complementarity lies at the heart of miRNA function. In animals, sequence complementarity between a 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). While 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 case for miR-122, which has been implicated in liver related diseases and hepatocellular carcinoma (HCC). miR-122 is a 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 (10)). The use of anti-sense 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 to 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 this miRNA's function. 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 (more details at http://switchgeargenomics.com/products/utr-reporter-collection/). 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).

Downloaded from www.jbc.org by guest, on March 21, 2011

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, non-genic, non-conserved 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 6,000 HT-1080 cells (ATCC) 18-24 hours before transfection to achieve 80% confluence at the time of transfection. Each transfection included 0.15µl of DharmaFECT DUO transfection reagent (Dharmacon), 100ng of 3'UTR reporter and sufficient mimic or non-targeting control miRNA (Dharmacon) to yield a final concentration of 20nM in a total volume of 100µl/well. Each construct was transfected in triplicate separately with either the miR-122 mimic or the non-targeting control. Plates were incubated at 37°C for 24 hours post-transfection before being removed. 100µl of Steady Glo luciferase assay reagent (Promega) were added to each well, plates were incubated at room temperature for 30 minutes 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 non-targeting 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 non-targeting control for each wild type and mutant construct pair was also calculated.

Proteomics sample preparation- HT1080 cells were transfected with 20nM miR-122 mimic or mock transfection in T-25 flasks using the transfection reagent Dharmafect 4 (Dharmacon). Cells were harvested 24 hours post-transfection and lysed by dounce homogenization in low-salt buffer (10mM Tris-HCl, pH8.0/10mM KCl/1.5mM MgCl2) with protease inhibitor cocktail (Calbiochem) and centrifuged at 1000xg 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 (TFE) was added to 50% (vol/vol). Samples were reduced with 15mM DTT at 55˚C for 45 minutes, then alkylated with 55mM iodoacetamide at room temperature for 30 minutes. Following alkylation, samples were diluted in digestion buffer (50mM Tris-HCl, pH8.0/2mM $CaCl_2$) to a final TFE concentration of 5% (vol/vol). Proteomics grade trypsin (Sigma) was added to a 1:50 (enzyme:protein) concentration and samples were digested at 37˚C for 5 hours. The digestion was quenched with 1% formic acid and sample volume was reduced to 20µl 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 10kDa 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 online 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 CID fragmentation spectrum acquisition (MS2) in the LTQ, with a maximum of 12 MS2 scans per MS1. Dynamic exclusion was activated, with a 45 seconds exclusion time for ions selected more than twice in a 30-second window.

Proteomics data analysis- Spectra were searched against the Ensembl release 54 (21) human protein-coding sequence database using Sequest

Downloaded from www.jbc.org by guest, on March 21, 2011

(Bioworks ver. 3.3.1, Thermo Scientific). A 1% FDR was determined against a reversed-concatenated decoy database, with specific filters (i.e. deltaCN, XCorr) selected to maximize the number of protein IDs in the forward database while maintaining the percentage of reversed IDs at 1% of the total. The protein list was curated by collapsing into groups proteins for which there was identical evidence of observation, and removing 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- HepG2, Huh7 cells were obtained from American Type Tissue Collection and maintained in Dulbecco's Modified Eagle 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 non-targeting control mimic sequence (CM) to control for non-sequence-specific effects in miRNA experiments.

Cell were lysed in ice-cold buffer containing 50 mM Tris-HCl, pH 7.5, 125 mM NaCl, 1% NP-40, 5.3 mM NaF, 1.5 mM NaP and 1mM orthovanadate, 175 mg/ml octylglucopyranoside and 1 mg/ml of protease inhibitor cocktail (Roche) and 0.25 mg/ml AEBSF (Roche). Cell lysates were rotated at 4ºC for 1 h before the insoluble material was removed by centrifugation at 12000 x 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 transfered 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 NIH (http://rsbweb.nih.gov/ij/).

Antibodies- Rabbit polyclonal anti-Vimentin (1:1000) and rabbit polyclonal antibody against G6PC3 (1:1000) were obtained from Cell Signaling and Sigma-Aldrich 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 hepatocellular carcinoma (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 7mer 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 bps 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). Since 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. A 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 paring 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

Downloaded from www.jbc.org by guest, on March 21, 2011

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 $D_{KL}(P \| Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$, 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 a 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 = $(E - E_w)/SDT_w$, 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 and standard deviation 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 one 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. $E_h = E_1 - E_{cost}$, 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 +/- 40nts 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 was 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 (UFR) in the target site flanking regions were selected. Finally, the corresponding RNA secondary structures of the UFRs 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 the Fisher's exact test. A similar analysis was performed on targets from the proteomics dataset containing a 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 has "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 has 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.

Downloaded from www.jbc.org by guest, on March 21, 2011

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: 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 while 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: PicTar (18) and MicroCosm (2,19). Individual luciferase reporter constructs were employed in co-transfection experiments with 20nM miR-122 mimic or a non-targeting 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) - **Figure 1**. 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 **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 2,** 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 while 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. and Baek et al. examined miRNA-induced changes in protein levels by MS, using the isotopic labeling technique SILAC to quantify protein abundance (29,33). 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 20nM miR-122 mimic or mock transfection for 24 hours were lysed and split into cytosolic and nuclear fractions, with each analyzed independently across three replicate samples. In total, 2,422 proteins were observed in at least two replicates, with 1,704 and 1,903 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. – **Table S2**.

As a first step towards confirming the putative targets, we examined the significantly down-

Downloaded from www.jbc.org by guest, on March 21, 2011

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 7 sequential matches complementary to positions 1-7 or 2-8 of the miRNA. 75 of the 226 identified targets contained a 7mer 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 8mer seed match, consistent with previous studies showing a greater repressive effect for 8mers over other seed lengths (34-36) (**Figure 2a**). 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 (**Figure 2b**). Including non-conserved TargetScan predictions showed little improvement over the 7-8mer 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 ALDOA (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). Though not previously identified as a direct target, the vimentin 3'UTR does contain a 7mer-m8 seed site, suggesting it may indeed be a direct miR-122 target.

Two additional proteins identified in our down-regulated set, citrate synthase (CS) 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 ((41), reviewed in (42)). These examples demonstrate the effectiveness of our experimental approach towards 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, although the change in expression did not surpass the 1.5-fold repression we established as the cut-off 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 1**). 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, while the proteomics revealed 65 additional direct targets (containing a 7mer or greater miR-122 seed site) and 151 indirect targets (lacking a 7mer miR-122 seed site) in the down-regulated set (**Figure 3**).

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 (**Figure 4**).

Two recent studies identified changes in

Downloaded from www.jbc.org by guest, on March 21, 2011

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 **(Table S3**). Indeed, of the 41 targets empirically derived from the luciferase based strategy which 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 while 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 miR-122 mimic - **Figure 5.** To better understand this regulation, we determined the secondary structure of target mRNAs (wild type and mutant) in the presence of miR-122 – **Figure 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 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. This data indicates 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 sequence (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 (**Figure 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 7mer 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 (**Figure 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

Downloaded from www.jbc.org by guest, on March 21, 2011

sequence as expected, but lacks a consistent binding motif to complement the miRNA 3'-end due to frequent mismatches, indicating a lack of consensus with regards to bulge length. Thus, while 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 **Figures S1d-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 to 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 – (**Figure 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 – **Figure 8 and Table S7.** A larger number of connections were obtained when indirect targets (lacking a 7 mer seed site) were also included – **Figure S2 and Table S9.** A more comprehensive list of genes related to diabetes and cancer can be found in **Tables 3 and S5**.

While 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 which becomes dysregulated upon down-regulation of miR-122 in HCC tumors, leading to loss of mitochondrial metabolic function (23). Interestingly, the authors found that miR-122 responsive genes were strongly enriched for cell cycle associated processes, a finding which 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.

In order 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 (**Figure 9 and Table S8**) containing down-regulated genes with either a predicted miR-122 target site or a direct connection to predicted target. The genes with highest connectivity are JUN, a well establish oncogene with defined roles in transcription regulation (reviewed in (47)) and RAC1 (Ras-related C3 Botulinum Toxin Substrate 1), a member of the RAS superfamily of small GTP-binding proteins which are implicated in the control of cell growth, cytoskeletal reorganization, and the activation of protein kinases (reviewed in (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 protein, 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 tied to oncogenesis (reviewed in (49). CREB1 does contain a predicted miR-122 target site,

Downloaded from www.jbc.org by guest, on March 21, 2011

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 (**Figure 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 (50)); MAPK1 and MAPK3, members of the MAP kinase family which are also known as extracellular signal-regulated kinases (ERKs) (reviewed in (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 (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 (53)); and VEGFA, a member of the platelet-derived growth factor (PDGF)/vascular endothelial growth factor (VEGF) family that acts on endothelial cells increasing vascular permeability and growth, and is also implicated in angiogenesis, cell migration and apoptosis (reviewed in (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 establishes 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 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 (10)). Our data suggests 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 non-receptor type 1 (PTPN1) is an enzyme that is the founding member of the protein tyrosine phosphatase (PTP) family. PTPs regulate numerous cellular processes including cell growth, differentiation, cell cycle, and oncogenic transformation (reviewed in (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 (67)). In cancer, vimentin over-expression 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).

Downloaded from www.jbc.org by guest, on March 21, 2011

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 (69)). MMP7 expression was shown to be anti-correlated with miR-122 in HCC patients, and levels of MMP expression and the stage of tumor progression are frequently correlated. Moreover, it has been suggested that MMPs 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 (69)).

Paxillin (PXN) has also been implicated in metastasis. Paxillin is a multi-domain protein that is primarily present in sites of cell adhesion to the extracellular matrix (ECM) 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, 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 the present 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 (ALDOA), a direct target of miR-122, cleaves fructose-1,6-bisphosphate to generate dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP). DHAP can be used to make glycerol-3-phosphate, which can then be converted to triacylglycerol. (79-82), GAP 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, while 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 towards 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 (CS), a direct target of miR-122, catalyzes the rate-limiting first step in the citrate cycle, 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 (FASN) (79). Ultimately, fatty acids and glycerol are combined to form triacylglycerols which 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. While miRNA regulation is often thought of as a method for shutting down protein expression, these examples demonstrate the more nuanced role

Downloaded from www.jbc.org by guest, on March 21, 2011

of miRNAs as chemostats, allowing for modulated control over expression 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. While 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 ASO-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.

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**, D154-158
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

Downloaded from www.jbc.org by guest, on March 21, 2011

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., Grun, 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**, D140-144
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., Graf, S., Haider, S., Hammond, M., Howe, K., Jenkinson, A., Johnson, N., Kahari, 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., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Smith, J., and Searle, S. M. (2010) *Nucleic Acids Res* **38**, D557-562
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) *Proc Natl Acad Sci U S A*, Proceedings of the 6th Conversations in Biomolecular Stereodynamics, Structures and Methods
28. Le, S. Y., Liu, W. M., and Maizel, J. V. (2002) *Knowledge Based System* **15**, 243-250

Downloaded from www.jbc.org by guest, on March 21, 2011

29. Baek, D., Villen, 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., Schwanhausser, 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., and et al. (1992) *J Natl Cancer Inst* **84**, 165-174
40. Krutzfeldt, 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., Ruskeepaa, 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

Downloaded from www.jbc.org by guest, on March 21, 2011

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) *ScientificWorldJournal* **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., Bieche, I., Dargere, 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**, E231-236
75. Turner, C. E. (2000) *J Cell Sci* **113 Pt 23**, 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. Elmen, J., Lindow, M., Schutz, S., Lawrence, M., Petri, A., Obad, S., Lindholm, M., Hedtjarn, 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. Cahova, M., Vavrinkova, H., and Kazdova, 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

Downloaded from www.jbc.org by guest, on March 21, 2011

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 Suppl**, S20-24

FOOTNOTES

*The authors would like to thank Dr. Christine Vogel for critical comments on the text. Research in LOFP lab was supported by the Children's Cancer Research Institute –UTHSCSA. DRB acknowledges funding from the Welch (F-1515) and Packard Foundations and the NIH (GM076536, GM067779, and GM088624) to EMM. Research in BAS lab was supported by the Intramural Research Program of the NIH, Nation Cancer Institute, Center for Cancer Research. The authors NDT, SFA, PJC 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. SwitchGear Genomics contributed funding to this research and the preparation of the manuscript.

FIGURE LEGENDS

Fig 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-value) < -1.3] (miR-122 treated vs. control) were considered significantly down-regulated (bottom-left quadrant).

Fig 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: seed site complementarity of 7mer-A1 and 7mer-m8 motifs, 8mer, and at either 7 or 8mer (7-8mer) seeds, as well as the TargetScan 5.1 with strict conservation of target sites (TS-Con). Results were plotted as (a) the occurrence of a predicted target site versus the ranking of the proteins based on increasing down-regulation, and (b) the level of enrichment as the ratio of the frequency of predicted targets within each subset versus the total experimental dataset.

Fig 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 7mer or greater miR-122 seed complementary site, "Indirect" refers to significantly down-regulated genes lacking such a site.

Downloaded from www.jbc.org by guest, on March 21, 2011

Fig 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.

Fig 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.

Fig 6. Consensus binding motifs calculated for targets identified by luciferase and proteomics analyses. The probability a nucleotide's involvement 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 trait (not shown).

Fig 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 (blue bars). Total human transcriptome was used as background (grey bars). (*) designates a P-value < 0.05, (**) designates a P-value < 0.005.

Fig 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. In green, genes identified in the proteomics analysis containing at least one miR-122 seed sequence; in pink, genes identified by the luciferase analysis; in blue, genes identified by both methods. Liver function and disease associations are displayed in purple.

Fig 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 while indirect targets lacking such sites are brown; in pink are targets identified by the luciferase analysis; in blue are high confidence targets identified by both methods.

Downloaded from www.jbc.org by guest, on March 21, 2011

Table 1. Summary of ten high-confidence validated miR-122 direct targets.

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

Downloaded from www.jbc.org by guest, on March 21, 2011

Table 2. Comparison of computational target predictions against luciferase dataset.

	TargetScan Conserved	TargetScan No Conservation	PicTar	MicroCosm	Number of Predictions by Different Algorithms		
					≥ 1	≥ 2	3
Total Predictions	27	85	54	100	139	84	16
Downregulated (% of Total)	14 (52%)	35 (41%)	21 (39%)	25 (25%)	37 (37%)	35 (42%)	9 (56%)

Downloaded from www.jbc.org by guest, on March 21, 2011

Table 3. Distribution of identified miR-122 targets according to biological function and diseases.

(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**

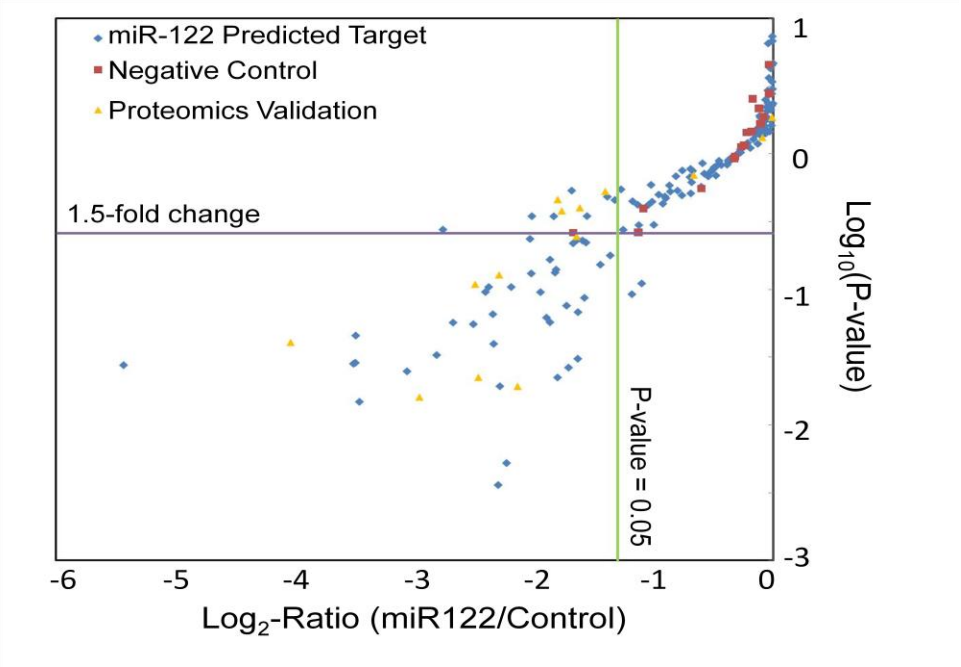
(C) Genes from Luciferase and Proteomics (with seed sequences) enriched for Diabetes mellitus:

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

Green - direct targets identified with proteomics analysis.
Pink - direct targets identified by luciferase assay.
Blue - high confidence targets identified by both methods.

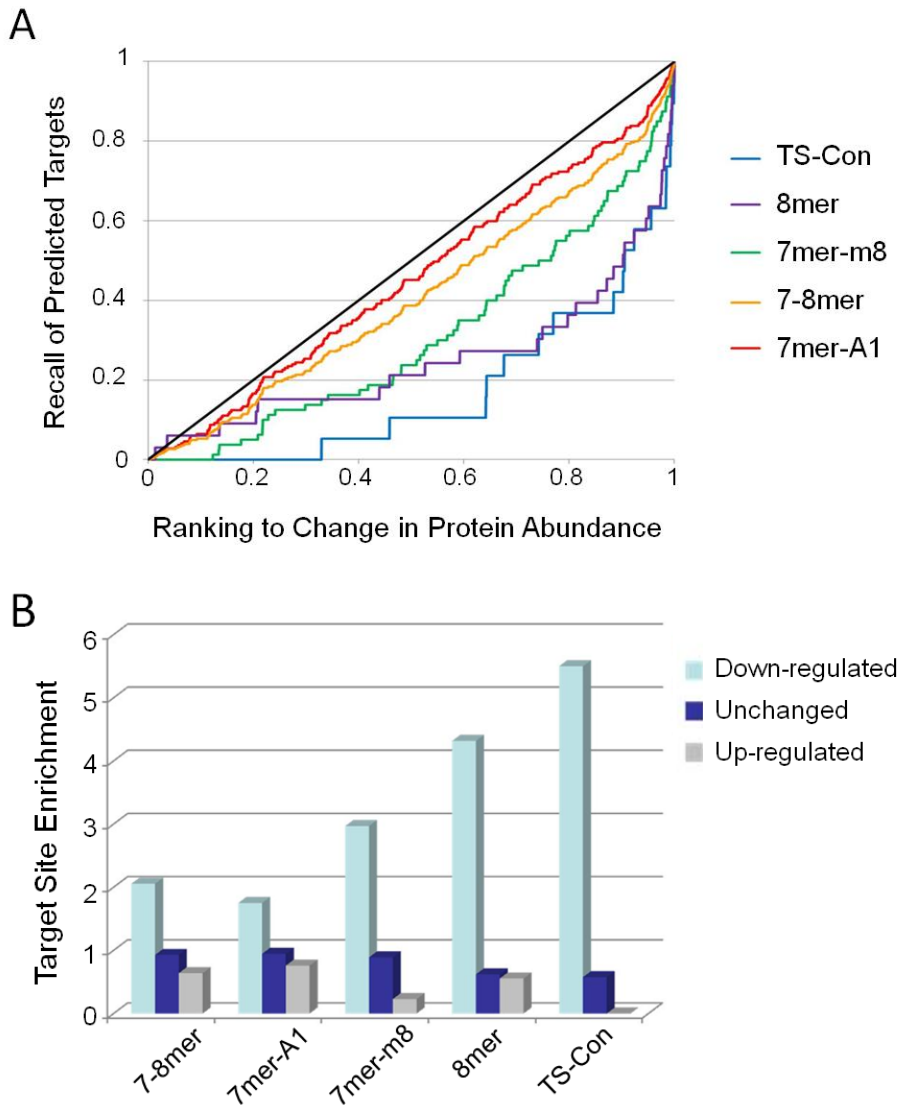
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 1



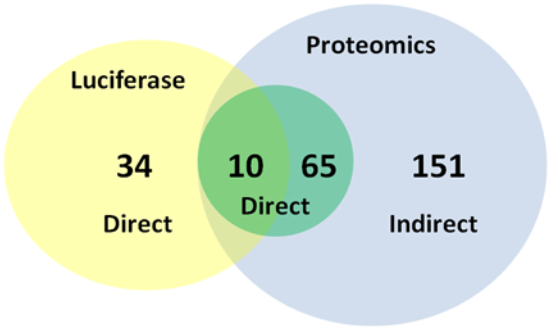
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 2



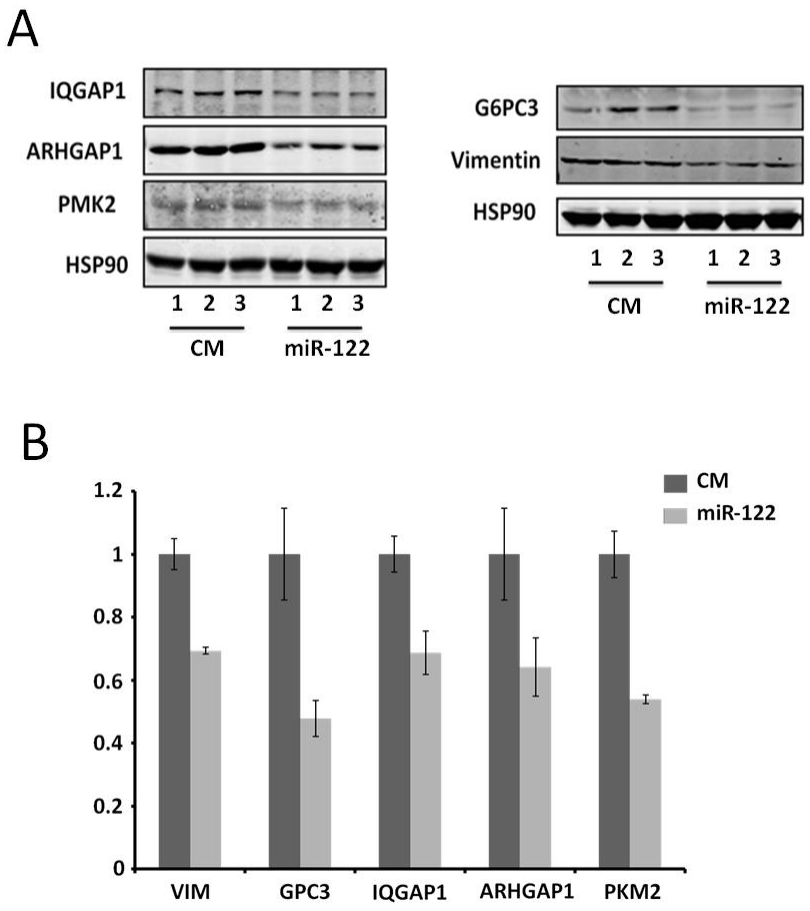
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 3



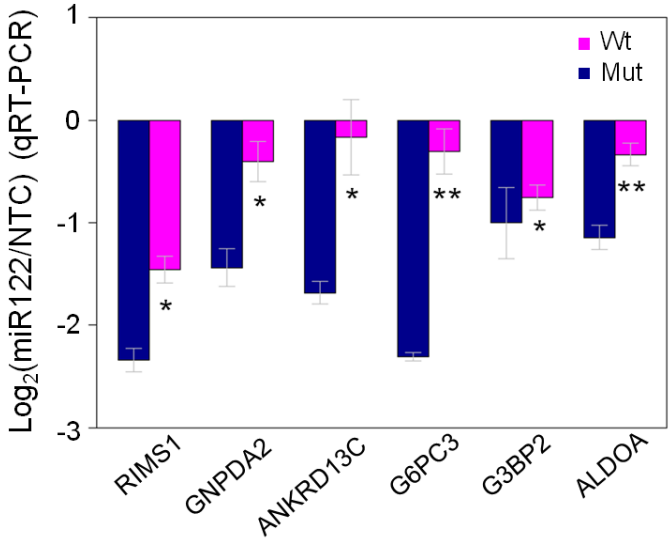
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 4



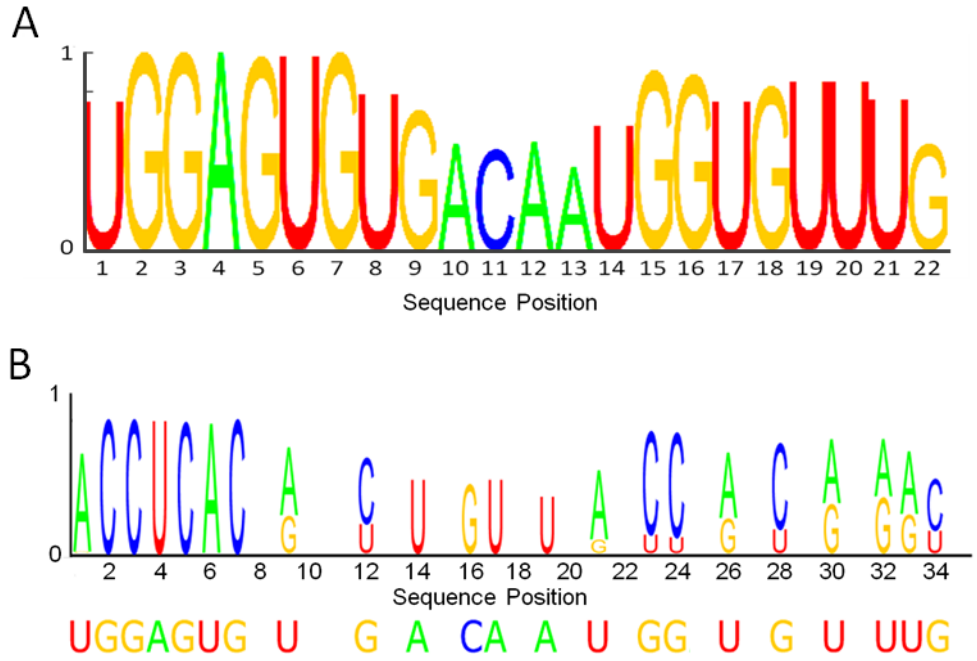
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 5



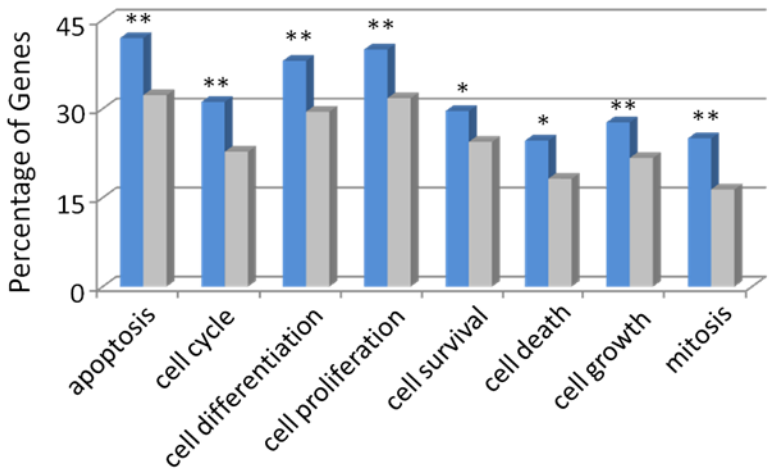
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 6



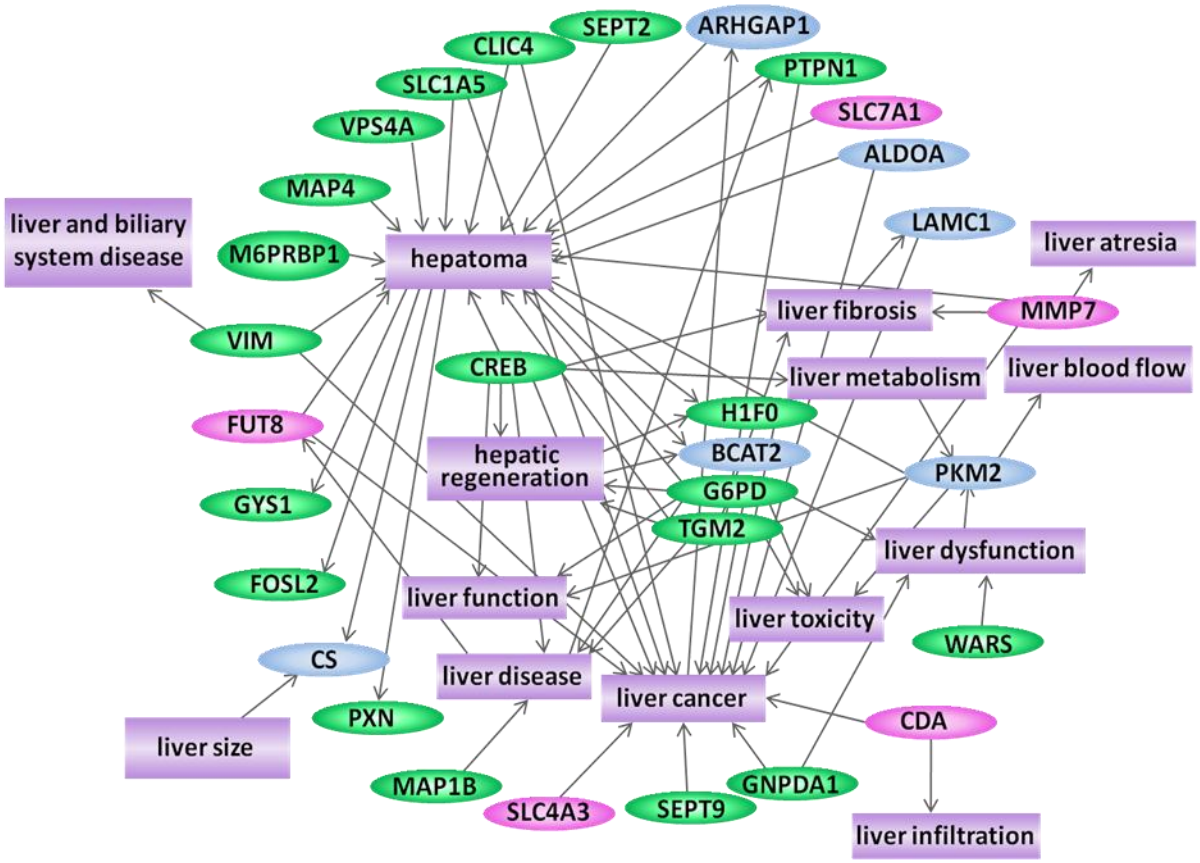
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 7



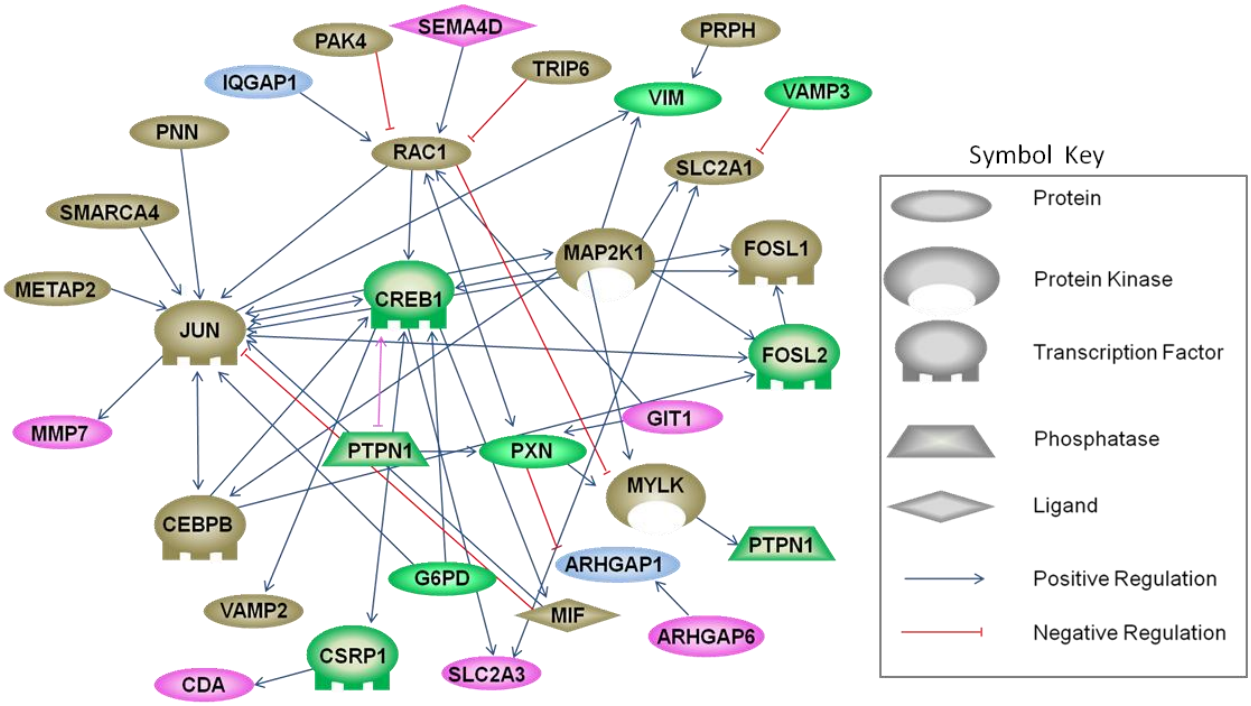
Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 8



Downloaded from www.jbc.org by guest, on March 21, 2011

Figure 9



Downloaded from www.jbc.org by guest, on March 21, 2011