

Using expression profiling data to identify human microRNA targets

Jim C Huang^{1,7}, Tomas Babak^{2,7}, Timothy W Corson^{2,3,6}, Gordon Chua⁴, Sofia Khan³, Brenda L Gallie^{2,3}, Timothy R Hughes^{2,4}, Benjamin J Blencowe^{2,4}, Brendan J Frey^{1,4,5} & Quaid D Morris^{2,4,5}

We demonstrate that paired expression profiles of microRNAs (miRNAs) and mRNAs can be used to identify functional miRNA-target relationships with high precision. We used a Bayesian data analysis algorithm, GenMiR++, to identify a network of 1,597 high-confidence target predictions for 104 human miRNAs, which was supported by RNA expression data across 88 tissues and cell types, sequence complementarity and comparative genomics data. We experimentally verified our predictions by investigating the result of let-7b downregulation in retinoblastoma using quantitative reverse transcriptase (RT)-PCR and microarray profiling: some of our verified let-7b targets include *CDC25A* and *BCL7A*. Compared to sequence-based predictions, our high-scoring GenMiR++ predictions had much more consistent Gene Ontology annotations and were more accurate predictors of which mRNA levels respond to changes in let-7b levels.

miRNAs are short endogenous noncoding transcripts with a widespread role in post-transcriptional regulation in higher eukaryotes. miRNAs downregulate the expression of their target mRNAs by complementary sequence binding that either causes the inhibition of translational initiation or leads to mRNA degradation, presumably after translocation to P-bodies¹. It is thought that miRNAs make an important contribution to the regulation of gene expression and that their misregulation is implicated in cancer onset and progression. Of the hundreds of known human miRNAs, however, only a handful have been experimentally linked to specific functions. Accurately determining the full repertoire of miRNA targets is now a major bottleneck in miRNA functional characterization. Complete specification of the distinguishing targeting features remains elusive, although measures of hybridization affinity, phylogenetic conservation of target sites and target-site accessibility have proved to be informative^{2–6}. One of the most accurate predictors of targeting is the presence of a conserved 6–8-bp ‘seed region’ of exact Watson-Crick complementary between the 3′ untranslated region (UTR) of the mRNA and the 5′ end of the miRNA, typically between nucleotides 2–7 (ref. 4). However, as

many as 40% of seed region matches that are conserved between human and chicken are false positives⁴. Target predictions based on less conservation have higher sensitivity but even higher false positive rates, necessitating laboratory verification to warrant functional claims about individual miRNA-mRNA interactions.

Precise identification of miRNA targets is critical to advancing our understanding of human diseases arising from miRNA misregulation, and accurate identification of physiologically active miRNA targets is now a considerable impediment to the functional characterization of individual miRNAs. The regulatory role of miRNAs makes them strong candidate oncogenes and tumor suppressors: many examples have been found of miRNA misregulation in tumors⁷, and miRNA expression levels are often more predictive of cancer subtypes than corresponding mRNA levels⁸. Example of both oncogenic miRNAs⁹ and those that suppress tumors^{10,11} have been reported.

Simultaneous profiling of miRNA and mRNA expression may be a timely strategy to achieve the required precision in the identification of functional miRNA targets. Many miRNAs cause degradation of their targets¹² and a large number of mRNAs are regulated in this way^{13,14}. Expression profiling of miRNAs and the mRNAs that they target for degradation should reveal an inverse relationship between the expression profile of the miRNA and that of its targets. This inverse expression pattern has been observed in distributions of expression profiles of predicted targets of tissue-specific miRNAs and more widely expressed miRNAs^{13–16}.

To our knowledge, there has been no practical demonstration that paired miRNA and mRNA expression profiles can be used to support miRNA target prediction. This demonstration is needed because inverse expression signals previously detected when examining the average behavior of all of the predicted targets of a miRNA may be masked by individual differences in transcriptional and post-transcriptional regulation.

We applied a highly sensitive data analysis procedure that accounts for individual differences in regulation by scoring candidate miRNA targets using paired miRNA-mRNA expression profiling

¹Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada. ²Department of Molecular and Medical Genetics, University of Toronto, 1 King's College Rd., Toronto, Ontario M5S 1A8, Canada. ³Division of Applied Molecular Oncology, Ontario Cancer Institute/Princess Margaret Hospital, University Health Network, Toronto, Ontario M5G 2M9, Canada. ⁴Banting and Best Department of Medical Research, University of Toronto, 160 College Street, Toronto, Ontario M5G 1L6, Canada. ⁵Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada. ⁶Present address: Department of Molecular, Cellular and Developmental Biology, Yale University, P.O. Box 208103, New Haven, Connecticut 06520, USA.

⁷These authors contributed equally to this work. Correspondence should be addressed to Q.D.M. (quaid.morris@utoronto.ca) or B.J.F. (frey@psi.toronto.eu).

data. Our procedure, called GenMiR++ (Generative model for miRNA regulation), achieves its high sensitivity by considering all other predicted miRNA regulators of an mRNA and balancing different sources of uncertainty when calculating its score¹⁵ (Supplementary Figs. 1–2 online). Through the use of our procedure, we demonstrate that the miRNA degradation signal can be used to vet individual miRNA sequence-based target predictions; by doing so, we established the GenMiR++ network of 1,597 high-confidence targets for 104 human miRNAs using a Bayesian inference algorithm. We demonstrate the biological validity of this network by focusing on let-7b, which we identified as a potential tumor-suppressor in retinoblastoma.

RESULTS

Scoring degradation signatures using GenMiR++

We compiled miRNA and mRNA expression data (Supplementary Methods online) for 151 human miRNAs and 16,063 mRNAs across a mixture of 88 normal and cancerous tissue samples common to both datasets^{8,17}. We obtained target predictions for 114 of the profiled miRNAs, covering a total of 890 unique profiled mRNAs, from the TargetScanS track of the University of California Santa Cruz (UCSC) genome browser. In total, we associated both miRNA and mRNA expression profiles with 6,387 of the predicted target pairs. We scored each miRNA-mRNA pair for the presence of an inverse expression pattern using the GenMiR++ procedure, which evaluates the degree to which the expression of the miRNA explains the pattern of downregulation of the mRNA expression levels under the approximation that the mRNA transcription rate in a given sample is constant across all mRNAs. The scores that GenMiR++ assigns to each pair depend not only on the expression profiles of the predicted miRNA regulator and the mRNA target but also on the profiles of all other predicted miRNAs for the same target (Supplementary Methods). GenMiR++ also balances different sources of uncertainty before scoring potential relationships. In general, miRNA-mRNA target pairs are penalized if both the miRNA and mRNA were highly expressed in the same tissue and are rewarded if the miRNA was highly expressed in tissues that the mRNA had low expression in, especially if few of its other predicted miRNA regulators were highly expressed therein.

Gene Ontology enrichment of GenMiR++-predicted target sets

We reasoned that if GenMiR++ had successfully identified the functional miRNA targets among the sequence-based predictions, then the set of high-scoring targeted mRNAs for each miRNA should have more consistent Gene Ontology annotations than random subsets of the sequence-based predictions. To test this, we downloaded Gene Ontology Biological Process (GO-BP) annotations from the Gene Ontology Annotation Database¹⁸. For each miRNA, we compare the Gene Ontology categories enrichments of two subsets of its TargetScanS-predicted targets: those with GenMiR++ scores in the top 50th percentile and a random subset of equal size. We scored the Gene Ontology enrichment within the target sets using Fisher's exact test (Supplementary Methods) for all GO-BP categories associated with at least five Ensembl proteins.

As a control to ensure that the observed enrichment was not due to coexpression of the GenMiR++ targets or to indirect regulation of the candidate targets by the miRNA, we also computed enrichment scores from sets of mRNAs which were coexpressed with the GenMiR++ targets but were not targeted by any miRNAs. We

found significantly greater consistency in Gene Ontology annotation among the set of GenMiR++ targets than in the random subsets (Supplementary Fig. 3a online; $P < 10^{-100}$, Wilcoxon-Mann-Whitney test) as well as significantly less consistency in the Gene Ontology annotation among the coexpressed sets (Supplementary Fig. 3b; $P < 10^{-100}$, Wilcoxon-Mann-Whitney test). A list of significant miRNA-Gene Ontology associations is available in Supplementary Fig. 3c and Supplementary Table 1 online.

The high-confidence GenMiR++ miRNA target network

Having established that GenMiR++ scoring enriches for more consistent GO-BP annotation, we thresholded the GenMiR++ score to select a high-confidence set of predicted functional targets from the sequence-based predictions. We set our threshold so that our high-confidence set covered 25% (1,597 of 6,387) of the sequence-based predictions, and we estimate based on random relabelings of the mRNAs that our set has a false positive rate of 3.5% (Supplementary Methods and Supplementary Table 2 online). The miRNA-mRNA network whose edges are predicted interactions from our high-confidence set covers 104 miRNAs and 316 unique mRNAs (Supplementary Fig. 4 online).

Validating GenMiR++-predicted let-7b targets

To experimentally validate the predictive accuracy of our method, we used our high-confidence GenMiR++-predicted targets for let-7b (Fig. 1) to predict the outcome of *let-7b* misregulation in retinoblastoma. Notably, no neural tissue was represented in the expression data used to build our network, so this test evaluates how well our predictions generalize to other unrepresented tissues and thus reveals the value of GenMiR++ and our resource for broader scientific exploration. We used microarrays to profile three retinoblastoma samples and one retina sample from a healthy individual (healthy retina). Of 245 interrogated miRNAs, 25 were

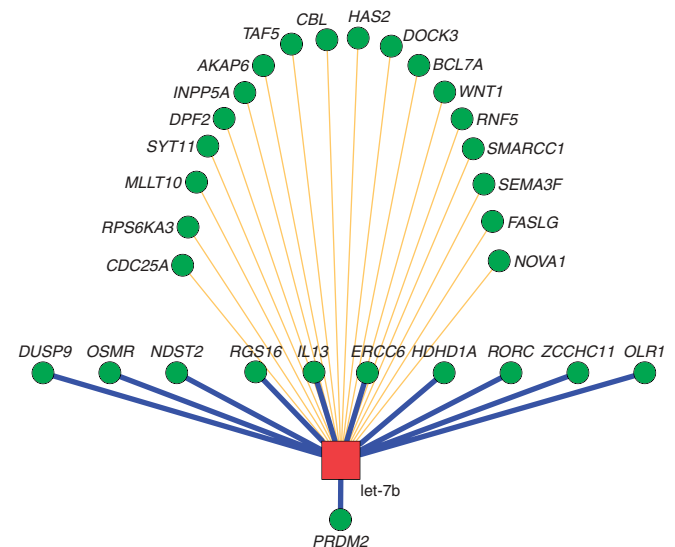


Figure 1 | The GenMiR++ network of let-7b targets. Nodes representing mRNAs are green circles. Messenger RNAs connected to let-7b by blue edges are predicted by GenMiR++ to be regulated solely by miRNAs from the let-7 miRNA family, of which *PRDM2* is predicted to be regulated solely by let-7b. Yellow lines connect let-7b to all other predicted mRNA targets, thus indicating that the mRNA target is predicted to also be regulated by other miRNAs from one or more additional families.

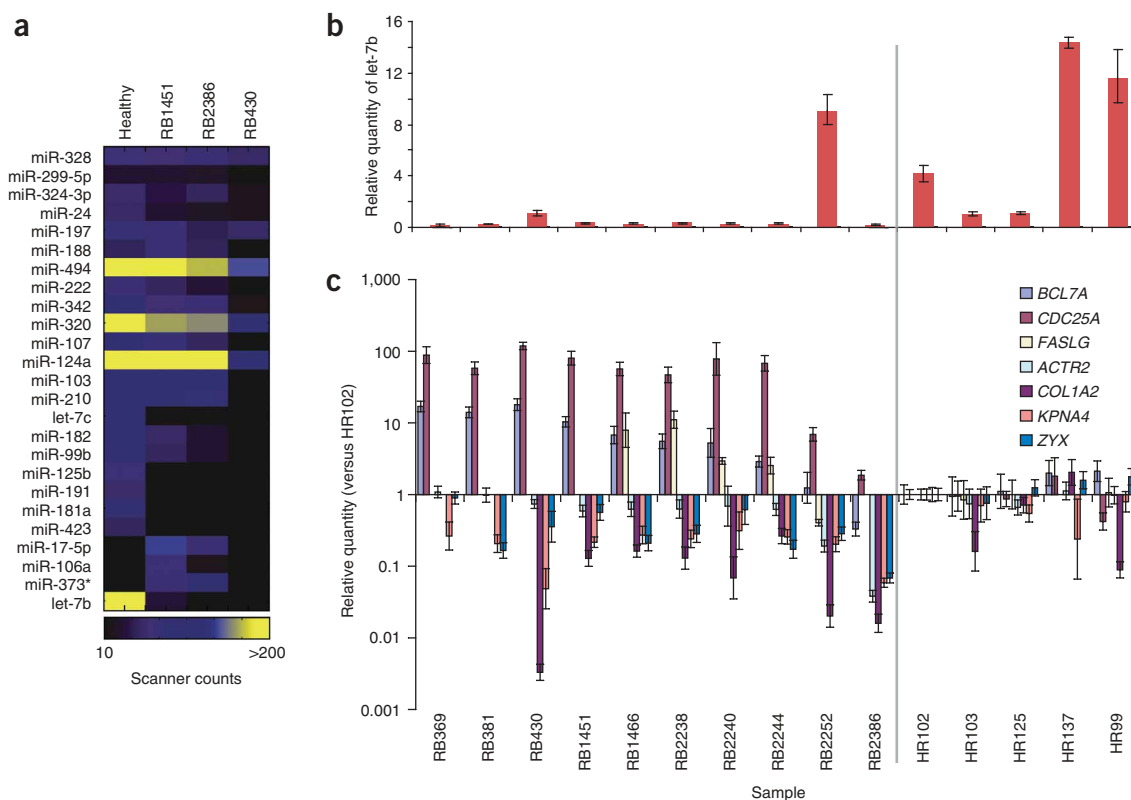


Figure 2 | Reduced levels of let-7b parallel increased levels of GenMiR++ targets in retinoblastoma. **(a)** Expression profiles of 25 (of 245) miRNAs in healthy retina and retinoblastoma samples. Let-7b is downregulated ~50-fold. **(b)** Quantitative real-time reverse transcriptase PCR assays reveal general downregulation of let-7b across retinoblastoma (RB) samples compared to healthy retina (HR). Relative quantities are calibrated to HR102. Error bars represent s.d. of the relative fold expression level from triplicate samples. **(c)** Quantitative reverse transcriptase PCR validation with three GenMiR++-predicted let-7b targets (*CDC25A*, *BCL7A*, *FASLG*) and three TargetScanS-predicted let-7b targets not confirmed by GenMiR++ (*ACTR2*, *COL1A2*, *KPNA4*). The negative control (*ZYX*) is not predicted to be a let-7b target by either method. Shown are the fold-enrichment of expression of the seven transcripts in 15 RNA samples, 10 from retinoblastoma (left; labeled 369–2386) and five retina samples from healthy individuals (right; labeled HR102–HR99). For 6 of the transcripts, expression in HR102 was used as a baseline to calculate fold enrichment. For *FASLG*, expression in HR125 was used as its baseline. Error bars represent the s.d. of the expression level over 4 technical replicates. Absence of error bars is an indication that the transcript did not amplify in the sample.

detected above background (that is, above 99% of negative control probes consisting of 1,000 random probe sequences per array). Of these, the most apparent differential expression was evident for let-7b, which was on average ~50-fold lower in abundance in retinoblastoma versus healthy retina (**Fig. 2a**). This downregulation is widespread across retinoblastoma; 9/10 of these samples from different individuals had lower let-7b levels than the average let-7b expression in five retina samples from healthy individuals (**Fig. 2b**).

We then used quantitative real-time PCR to measure the abundance of five predicted let-7b targets (**Fig. 2c**) in our high-confidence GenMiR++ network. Two of these predicted targets, *RORC* and *IL13*, did not amplify in more than 2 out of the 15 samples, and we excluded them from subsequent analysis. As negative controls, we also assayed three TargetScanS-predicted let-7b targets with GenMiR++ scores in the bottom 50th percentile. We found that two of the GenMiR++-predicted targets (*BCL7A* and *CDC25A*) are significantly upregulated ($P < 0.01$) in at least eight of the ten retinoblastoma samples compared to the healthy retina (sample 102; HR102) baseline. Another GenMiR++-predicted target, *FASLG*, was significantly upregulated ($P < 0.01$) in four of the five retinoblastoma samples in which it amplified. Thus, when a positive target amplified in retinoblastoma, its expression was significantly higher than the retinal baseline at least 80% of the

time, whereas only once were any of the three negative controls upregulated in any of the retinoblastoma samples.

let-7b downregulates GenMiR++-predicted targets

To ensure that the observed upregulation of the GenMiR++-predicted targets of let-7b was not caused by other changes in retinoblastoma, we used microarrays to profile the transcriptomic response to exogenous let-7b. We transfected a synthetic RNA duplex of the mature let-7b sequence into WERI-Rb1 cells and compared their expression profile to cells transfected in parallel with a scrambled let-7b duplex. In total, 9,659 genes were expressed above the median expression intensity in the control retinoblastoma cells, and of these, we focused on the most downregulated 5% (483 targets). To ensure that this screen was enriched for targets of let-7b, we confirmed that the reverse complement of the let-7b seed region was overrepresented in the 3' UTRs of the downregulated genes.

We found sequence-based predictions supported by a high GenMiR++ score had much higher precision than those that were not, with only a moderate loss of sensitivity. Twelve of the 34 WERI-Rb1 cell-expressed TargetScanS-predicted let-7b targets are included in the high-confidence GenMiR++ network; of these, 5 (42%) were scored as downregulated in the assay (*BCL7A*, *HDHD1A*, *RNF5*, *SEMA3F*, *SMARCC1*) whereas only 2 of the 22

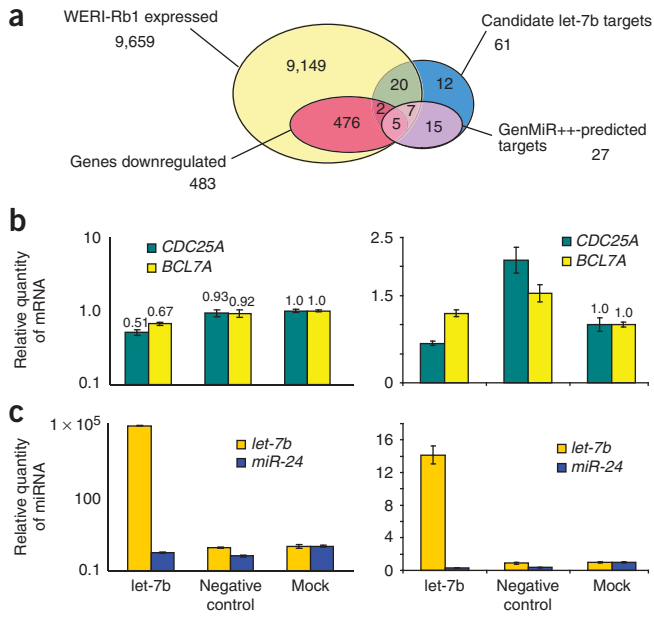


Figure 3 | Increasing levels of let-7b reveals enrichment of GenMiR++ targets among downregulated genes. **(a)** Venn diagram showing overlap of WERI-Rb1 expressed genes, downregulated genes, GenMiR++ target predictions and sequence-based predictions. **(b)** TaqMan real-time RT-PCR results of *CDC25A* and *BCL7A* in WERI-Rb1 (left) and HeLa cells (right), normalized against 5S expression and calibrated to endogenous levels. Both genes were significantly downregulated in both WERI-Rb1 and HeLa cells compared to transfection of the Ambion negative control ($P < 0.01$). **(c)** Relative let-7b levels (Taqman real-time RT-PCR) after let-7b transfection in WERI-Rb1 cells (left) and HeLa cells (right), normalized against 5S expression and calibrated to endogenous (mock) levels. miR-24 levels were similarly analyzed, as a negative control. WERI-Rb1 let-7b and miR-24 levels are ~900- and 16-fold lower, respectively, compared to healthy fetal retina. Error bars, s.d. of the expression level over four technical replicates.

(9%) other TargetScanS predictions (*CDC34*, *EIF4G2*) scored as such. This represents a significant enrichment for confirmed targets among the GenMiR++-predicted targets (Fig. 3a; $P = 0.038$, Fisher's exact test) with only a 28.5% loss in sensitivity (2/7) compared to all of the TargetScanS-predicted targets.

The precision of TargetScanS is lower in our assay, 21% (7/34), than in previous reports, suggesting that we may have missed some let-7b targets. Indeed, using quantitative real-time PCR, we identified *CDC25A* downregulation and reconfirmed *BCL7A* downregulation in the let-7b transfected WERI-Rb1 cells (Fig. 3b,c). However, based on our results, we would expect 71% (5/7) of the false negatives in our microarray assay to also be among the GenMiR++-predicted set. So, for example, if there were seven false negatives in our assay, bringing the actual precision of TargetScanS closer to previous reports at 41% (14/34), then the expected precision of the GenMiR++-predicted set would double to 83% (10/12).

DISCUSSION

We demonstrated that miRNA-mRNAs paired expression profiles can be used on a large-scale by GenMiR++ to improve the accuracy of sequence-based miRNA-target predictions. Our target predictions have more consistent Gene Ontology annotations than random subsets of target predictions based on sequence alone and contain targeting relationships that expand our understanding of the roles of miRNAs in disease. Furthermore, by providing a surrogate for direct experimental validation, GenMiR++ can be used to help identify 3' UTR sequence features that are responsible for target degradation because *cis* elements predictive of a high GenMiR++ score should also be predictive of target degradation.

Our demonstration that let-7b targets *CDC25A*, and that *CDC25A* is overexpressed in let-7b-depleted retinoblastoma, suggests another mechanism for let-7b-mediated tumor suppression in humans. In breast cancer, *CDC25A* overexpression^{19,20} is sufficient to overcome the G1-S checkpoint resulting from DNA damage²¹, possibly through triggering the activation of the cyclin-dependent kinase Cdk2 required for progression from G1 to S phase. Here we report that *CDC25A* is overexpressed in retinoblastoma, suggesting

a similar mechanism may be at work therein. Notably, *CDC25A* is sufficient for oncogenic transformation of mouse fibroblasts lacking a functional *Rb1* gene or in cooperation with active Ras.

Although it is impossible for us to rule out, it is extremely unlikely that the inverse expression patterns we observed between miRNAs and their GenMiR++-predicted mRNA targets are due to indirect regulation. The Gene Ontology enrichments for GenMiR++-predicted target sets are considerably larger than those of transcripts with similar expression patterns; if the miRNA regulation were indirect, these should be approximately equal. We do not, however, rule out the possibility that there are additional redundant indirect pathways also contributing to their degradation.

At present our GenMiR++ scores only predict which miRNA targets are regulated by transcript degradation; we have not attempted to predict targets regulated by translational repression. In an effort to avoid ruling out the latter type of interaction, we only considered candidate targets with exact seed region matches that generally lead to transcriptional degradation^{13,14} (Supplementary Fig. 5 online). However, extending our approach to address translation repression would be straightforward if protein expression data were available. As with transcriptional degradation, if miRNA regulates target expression by translational repression, then the miRNA expression profiles should explain any observed downregulation of protein abundances compared to what would be expected based on the mRNA expression profiles.

The current version of GenMiR++ assumes a constant transcription input for each mRNA; although this assumption is an oversimplification, it is surprisingly effective, suggesting that modeling transcriptional rate may not be necessary for miRNA target prediction. Nonetheless, we could extend the basic GenMiR++ model to incorporate available data or computational predictions on the transcriptional rate for specific mRNAs. Extensions to the GenMiR++ model that incorporate protein abundance and transcription rate data are described in **Supplementary Methods**.

Methods such as GenMiR++ that account for additive regulation by miRNAs should consider all other predicted miRNAs when scoring miRNA-mRNA pairs. However, all such methods are sensitive to the number of false positives in the initial set of sequence-based predictions; higher numbers of false positives provide more opportunities for one of the false positives to show a spurious inverse expression relationship. Indeed, GenMiR++ was not as successful when applied to the union of five datasets of sequence-based targets that contained a larger proportion of false positives per mRNA (data not shown). It may be possible to make GenMiR++ more robust to very noisy sets of candidate targets by using sequence features like

accessibility^{5,6} to assign a prior probability to each of the candidate targets and incorporating that prior into the GenMiR++ model, so that, for example, a less accessible target site would require more support in the expression data to achieve a high score.

Based on our functional analyses and biological testing, we believe that the GenMiR++ network is the most accurate miRNA target interaction network currently available, and the methods underlying its production will become widely implicated and further developed. Our results highlight the potential for GenMiR++ as a principled algorithmic framework in which functional regulatory networks of miRNAs and functional target mRNAs can be discovered by combining miRNA and mRNA sequence and expression data; and through these discoveries the pivotal role of miRNAs in cancer onset and progression can be better understood.

METHODS

Collection and mapping of the sequence-based target predictions. We downloaded TargetScanS⁴ predictions from build 17 of the human genome on UCSC genome browser²² for the 114 miRNAs represented in our expression compendium. We then mapped the European Molecular Biology labels of the profiled mRNA transcripts in our compendium to the Human Gene Nomenclature Committee labels of the predicted targets. Through this mapping, we unambiguously associated 890 mRNAs with an expression profile and a set of predicted miRNA regulators. Though GenMiR++ is designed to work with any sequence-based predictions, we chose the TargetScanS over a number of similar methods for two reasons: a comparison of protein, miRNA and mRNA expression data in mouse showed that the interactions predicted by TargetScanS are more likely to lead to mRNA degradation rather than translational repression (**Supplementary Fig. 5**). Also, the TargetScanS target sets showed the greatest Gene Ontology enrichment compared with four other sequence-based methods (**Supplementary Fig. 6** online).

The GenMiR++ model. GenMiR++¹⁵ uses miRNA and mRNA expression profiles from the same sets of tissues and cell types to identify candidate miRNA-mRNA target pairs that are best supported by the expression data. For each mRNA, its candidate miRNA regulators were scored according to how much the miRNA expression profile contributed to explaining downregulation of the mRNA expression, given all other miRNA candidate regulators. GenMiR++ calculates the scores by attempting to reproduce an mRNA's profile by a weighted combination of the genome-wide average normalized expression profile and the negatively weighted profiles of a subset of the miRNA regulators. miRNAs that often appear in subsets whose reproductions are good fits to the mRNA profile are assigned the highest scores.

Additional methods. Descriptions of the collection and renormalization of the preexisting microarray data and new microarray data from WERI and HeLa cells; the tests for Gene Ontology enrichment; the design of the miRNA expression profiling array; the extraction and preparation of RNA for quantification; the let-7b transfection; the quantitative real-time RT-PCR; the calculation of the false detection rate of GenMiR++ scores; the source and handling of the human tissues; the GenMiR++ model and its learning algorithm; and extensions to the GenMiR++ model to incorporate protein abundance and transcriptional rate data are

available in **Supplementary Methods**. Studies with human tissues were approved by the Research Ethics Boards of the University Health Network (Toronto) and the Hospital for Sick Children (Toronto). The complete GenMiR++ miRNA-mRNA interaction network is available at <http://www.psi.toronto.edu/genmir/>.

Accession codes. Gene Expression Omnibus (GEO): GSE7072 (miRNA data) and GSE7185 (mRNA data).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

J.C.H. and T.B. were supported by Natural Science and Engineering Research Council postgraduate scholarships. T.W.C. was supported by a Canada Graduate Scholarship from the Canadian Institutes for Health Research (CIHR). This study was supported by a Natural Sciences and Engineering Research Council operating grant and Canadian Foundation for Innovation and Ontario Research Fund infrastructure grants to Q.D.M.; a CIHR grant to B.J.F. and T.R.H.; an Ontario Genomics Institute and Genome Canada grant to B.J.F. and B.J.B.; a CIHR and National Cancer Institute of Canada grant to B.J.B.; and a US National Institutes of Health grant to B.L.G. B.J.F. is a Fellow of the Canadian Institute for Advanced Research.

Published online at <http://www.nature.com/naturemethods/>
Reprints and permissions information is available online at
<http://ngp.nature.com/reprintsandpermissions>

- Engels, B.M. & Hutvagner, G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **25**, 6163–6169 (2006).
- Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
- Enright, A.J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
- Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Long, D. *et al.* Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.* **14**, 287–294 (2007).
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
- Wu, W., Sun, M., Zou, G.M. & Chen, J. MicroRNA and cancer: Current status and prospective. *Int. J. Cancer* **120**, 953–960 (2006).
- Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
- He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).
- Johnson, S.M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
- Hammond, S.M. MicroRNAs as oncogenes. *Curr. Opin. Genet. Dev.* **16**, 4–9 (2006).
- Bagga, S. *et al.* Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**, 553–563 (2005).
- Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
- Farh, K.K. *et al.* The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
- Huang, J.C., Morris, Q.D. & Frey, B.J. Bayesian inference of MicroRNA targets from sequence and expression data. *J. Comput. Biol.* **14**, 550–563 (2007).
- Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3' UTR evolution. *Cell* **123**, 1133–1146 (2005).
- Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA* **98**, 15149–15154 (2001).
- Camon, E. *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* **32**, D262–D266 (2004).
- Cangi, M.G. *et al.* Role of the Cdc25A phosphatase in human breast cancer. *J. Clin. Invest.* **106**, 753–761 (2000).
- Wu, W., Fan, Y.H., Kemp, B.L., Walsh, G. & Mao, L. Overexpression of cdc25A and cdc25B is frequent in primary non-small cell lung cancer but is not associated with overexpression of c-myc. *Cancer Res.* **58**, 4082–4085 (1998).
- Mailand, N. *et al.* Rapid destruction of human Cdc25A in response to DNA damage. *Science* **288**, 1425–1429 (2000).
- Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).