

# Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques

Wan-Lin Su<sup>1,2</sup>, Robert R Kleinhanz<sup>2</sup> and Eric E Schadt<sup>3,\*</sup>

<sup>1</sup> Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA, <sup>2</sup> Department of Genetics, Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA, USA and <sup>3</sup> Department of Genomics and Multiscale Biology, Pacific Biosciences, Menlo Park, CA, USA

\* Corresponding author. Department of Genomics and Multiscale Biology, Pacific Biosciences, 1505 Adams Drive, Menlo Park, CA 94025, USA. Tel.: +1 650 521 8025; Fax: +1 650 847 1034; E-mail: eric.schadt@gmail.com

Received 19.7.10; accepted 8.4.11

**Integrative genomics and genetics approaches have proven to be a useful tool in elucidating the complex relationships often found in gene regulatory networks. More importantly, a number of studies have provided the necessary experimental evidence confirming the validity of the causal relationships inferred using such an approach. By integrating messenger RNA (mRNA) expression data with microRNA (miRNA) (i.e. small non-coding RNA with well-established regulatory roles in a myriad of biological processes) expression data, we show how integrative genomics approaches can be used to characterize the role played by approximately a third of registered mouse miRNAs within the context of a liver gene regulatory network. Our analysis reveals that the transcript abundances of miRNAs are subject to regulatory control by many more loci than previously observed for mRNA expression. Moreover, our results indicate that miRNAs exist as highly connected hub-nodes and function as key sensors within the transcriptional network. We also provide evidence supporting the hypothesis that miRNAs can act cooperatively or redundantly to regulate a given pathway and that miRNAs play a subtle role by dampening expression of their target gene through the use of feedback loops.**

*Molecular Systems Biology* 7: 490; published online 24 May 2011; doi:10.1038/msb.2011.23

**Subject Categories:** functional genomics

**Keywords:** causal associations; eQTL mapping; expression QTL; microRNA

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar license to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

## Introduction

The discovery of microRNAs (miRNAs), a class of endogenously produced small non-coding RNA molecules, revealed an additional mechanism by which genes are regulated and added yet another dimension of complexity to the regulation of biological systems (Lee *et al.*, 1993; Reinhart *et al.*, 2000). In just the past decade alone, miRNAs have been implicated in various biological processes including temporal development events, cell-cycle regulation, metabolism, immunity, and tumorigenesis (Esquela-Kerscher and Slack, 2006; Carleton *et al.*, 2007; Taganov *et al.*, 2007; Wilfred *et al.*, 2007; Zhao and Srivastava, 2007). These small single-stranded molecules of ~22 nucleotides have been found to regulate multiple pathways in several species, ranging from developmental timing and neuronal patterning in worms (Lee *et al.*, 1993; Reinhart *et al.*, 2000; Johnston and Hobert, 2003; Chang *et al.*, 2004); to apoptosis and fat metabolism in flies (Brennecke *et al.*, 2003; Xu *et al.*, 2003); to hematopoietic differentiation in

mice (Chen *et al.*, 2004); to leaf morphogenesis and adaptive stress response in plants (Palatnik *et al.*, 2003; Mallory and Vaucheret, 2006). In humans, over a third of the genome is predicted to be regulated by miRNAs (Lewis *et al.*, 2005) and computational methods estimate that the mammalian miRNA repertoire consists of close to 1000 miRNA genes—~3% of the human genome (Bentwich *et al.*, 2005; Berezikov *et al.*, 2005).

While many miRNAs have been discovered via a combination of experimental and computational approaches, the functional relevance of the majority of miRNAs remains unknown. Moreover, while miRNAs have been implicated in multiple disease areas, elucidating the mechanisms that detail the path from miRNA expression to phenotype remains a challenge. Because validation experiments are often time consuming and laborious, several groups have developed algorithms to predict the target RNA transcripts of miRNAs *in silico*. The majority of these programs utilize the high degree of complementarity between miRNA seed region and the 3' UTRs of their targets, predicted miRNA-messenger RNA

(mRNA) target thermodynamics, and the sequence conservation of the binding motif (reviewed by Sethupathy *et al*, 2006; Martin *et al*, 2007; Maziere and Enright, 2007). Moreover, other techniques ranging from brute-force motif-mining of enriched motifs within the 3' UTR of known genes, to sophisticated support vector machine learning techniques have also been applied to miRNA target prediction. Multiple groups have shown that the majority of these computational programs, when used in isolation, report high false-positive rates (~22–39%) and high false-negative rates (~35–52%) (Sethupathy *et al*, 2006).

Several groups have now reported successful use of integrative genomics and genetic approaches to investigate the genetic basis of intermediate molecular phenotypes such as gene expression and protein levels. Briefly, the approach combines classical genetics, where DNA sequence variation is examined and correlated with complex physical or behavioral traits, and gene expression profiling with the broad goal of elucidating the molecular underpinnings of complex traits. In addition, others have developed algorithms and procedures to identify statistically supported causal relationships between multiple traits of interests (Schadt *et al*, 2005). Recently, Yang *et al* (2009) validated eight of nine genes that were predicted to be causal for obesity-related traits. These results illustrate the tremendous success of using integrative genomics approaches to dissect the landscape of complex traits and predict with high accuracy the causal relationships between any given pair of traits.

We report here one of the first investigations into the genetic basis of miRNA gene expression and demonstrate how integrative genomics approaches can be used to explore the regulation of miRNAs and miRNA-mediated regulation. Using a highly sensitive and robust quantitative polymerase chain reaction (qPCR) assay, we examined expression levels of 187 (or ~30% registered miRNAs in miRBase (release 15) (Griffiths-Jones, 2004; Griffiths-Jones *et al*, 2006, 2008)) within the context of a mouse liver gene regulatory network. Our results suggest that the transcript abundances of these miRNA are under more complex regulation (i.e. more subtle regulation or are regulated by multiple loci) than those previously observed for mRNA abundances. Our data further suggest that many of the surveyed miRNAs are co-expressed with mRNAs involved in DNA replication and the cell cycle in liver tissue. Finally, we demonstrate that a number of miRNAs detected typically exist as highly connected hub-nodes within the liver transcriptional network that respond to perturbations in mRNA transcript levels and drive changes in mRNA expression profiles. Additionally, for the majority of surveyed miRNAs, the number of mRNA statistically supported as causal for regulating miRNA transcript exceeds the number of downstream mRNA transcripts predicted to respond to changes in miRNA levels, illustrating the complexity of miRNA-mediated regulation.

## Results

Integrating genotype, mRNA expression and phenotypic data from experimental mouse cross populations has proven to be a successful method for identifying genes supported as causal for complex traits like disease. We present here the results of a

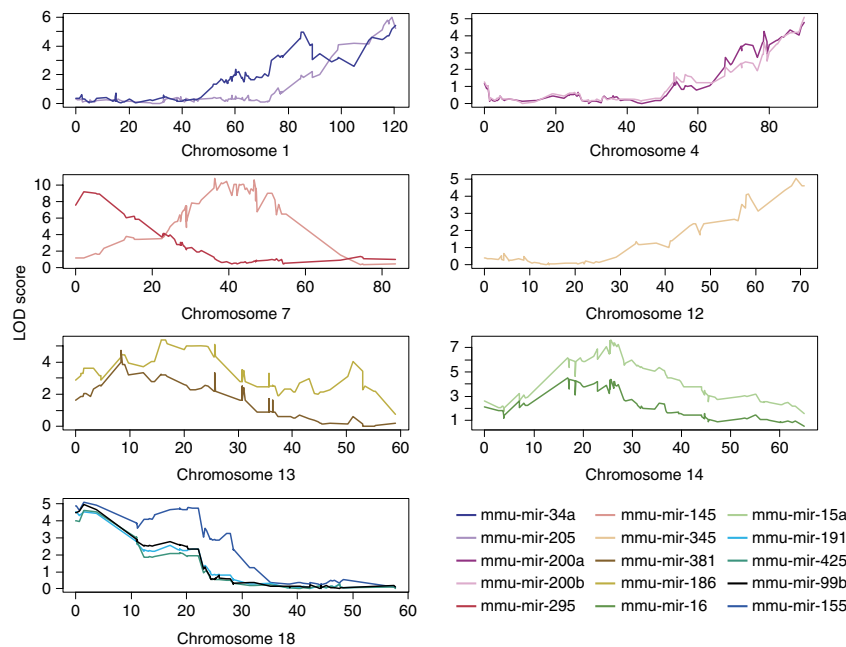
study designed to explore the association between miRNAs and their target mRNAs using an integrative genomics approach. In order to collect substantial tissue mass for all three profiling assays, we elected to perform this study in liver. Using an F<sub>2</sub> mouse cross, we collected both mRNA expression and genotype information from liver. Next, in order to focus on the additional layer of regulation provided by miRNAs, we profiled expression levels of 187 miRNAs. At the time the study was being designed, these miRNAs represented all known miRNA sequences. Taking into account the rapid pace of miRNAs discovered in the past decade, we note that the miRNAs surveyed currently represent a large subset of discovered miRNAs (~31% of registered miRNAs in miRBase database (release 15)).

Linkage analysis techniques were then applied to infer regulatory relationships between DNA loci and the two classes of expression traits, that is, mRNA and miRNAs. We next characterized the miRNA–mRNA relationships using a simple correlation analysis and applied a variation of a previously developed statistical inference technique to infer regulatory relationships between mRNA and miRNAs. The regulatory predictions from this statistical procedure were recently validated in a study where eight out of nine predictions were experimentally tested and successfully validated using either transgenic mice or knockouts mouse models of the predicted genes. Whereas a large number of studies have focused on the downstream targets of miRNAs, our results present a unique perspective on the regulatory roles of miRNAs by investigating predicted causal regulators of miRNAs in addition to the downstream mRNA targets of miRNAs.

## mRNA and miRNA eQTL mapping in the BXD mouse study

To study the genetic basis of variation in miRNA transcript abundance levels, we constructed an F<sub>2</sub> intercross population using two inbred mouse strains, C57BL/6J and DBA/2J, hereafter referred to as the BXD cross. To avoid sex-specific effects, we selected 127 male animals for miRNA and mRNA expression profiling. Each animal was genotyped using a dense panel of SNP markers (covering nearly 100% of the DNA variation segregating in the F<sub>2</sub> population) and RNA was isolated from liver samples from each F<sub>2</sub> animal and profiled on a whole-genome microarray consisting of 39 557 non-control probes. In addition, we profiled the intensity levels of 187 individual miRNAs using a highly sensitive and robust high-throughput primer extension (PE) qPCR assay (Raymond *et al*, 2005). Of the set of miRNAs investigated, transcripts with missing values in ≥60 animals and were presumed to be poorly detected and removed from the data set.

Using standard parametric linkage analysis techniques, we treated the expression levels of both mRNAs and miRNAs as quantitative traits to identify regulatory loci generally referred to as expression quantitative trait loci (eQTLs). In the case of miRNAs, at a LOD score significance threshold of >4.9 (corresponding to a false-discovery rate (FDR) threshold of 10%), we identified 10 eQTLs corresponding to 10 miRNA traits (~5% of the miRNA traits examined). At this threshold, we would at most expect a single false positive by random



**Figure 1** Top 15 microRNA expression quantitative trait loci (eQTL) plots. X-axis represents genomic coordinates in basepairs. Results shown are from the initial genome-wide scan. Of those shown, eQTLs on chromosome 18 for mmu-mir-99b, mmu-mir-191, and mmu-mir-425 were not replicated when prior knowledge of the eQTLs of the respective target mRNAs were incorporated into the linkage analysis.

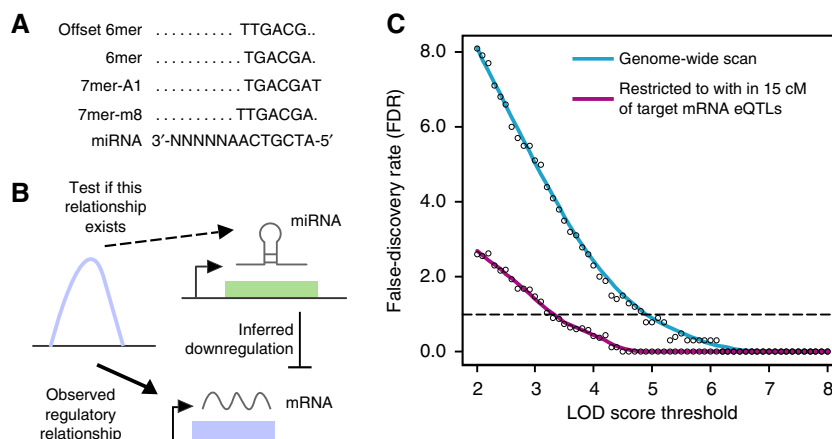
chance. Of the 10 eQTLs identified, 5 eQTLs map to the same location as their corresponding physical transcript (i.e. *cis* eQTLs (Doss *et al*, 2005)). Here, a *cis* eQTL is defined to contain the structural gene within the 15-cM region flanking the peak LOD score. A number of distinct miRNAs processed from the same polycistronic transcript (e.g. mmu-miR-15a/16, mmu-miR-200a/200b) also showed a high degree of correlation in expression levels and mapped to the same eQTL (Figure 1). The highest LOD score of 10.8 corresponded to mmu-miR-145 mapping to chromosome 7 and on average, each eQTL explained 21% of the observed variation.

In contrast, we identified 5293 eQTLs for 5107 of the 39 557 mRNA transcripts (~13%) profiled in the BXD cross at a LOD score threshold of >4.9 (corresponding to an FDR <5%), with a maximum LOD score of 85. Adjusting the LOD score threshold to achieve an FDR <10% (LOD score >4.4) resulted in a total of 7309 eQTLs for 6837 mRNA expression traits (~17%). Of these, 2712 (or 37%) were *cis* eQTLs. Thus by percentage, at the 10% FDR threshold, more than three times as many mRNA eQTL were detected when compared with the miRNA expression traits. On average, each eQTL explained 28% of the observed expression trait variation (the minimum percent variation explained was 15%).

There are a number of possible explanations for the reduced rate of eQTL detection observed for the miRNA. First, miRNA traits could potentially give rise to a more complex eQTL signature. Here, we refer to the complexity of the eQTL signature by the number of eQTLs within a given signature and the effect size accounted for by each locus. Hence, assuming that the effect produced by any single DNA variant is reduced in comparison to that of mRNA eQTLs, we might not be sufficiently powered to detect the bigger set of eQTL with smaller effect sizes. Alternatively, miRNAs may be more tissue

specific than mRNAs, where detecting the genetic variance component requires one to be in the right context. Finally, it is also possible that the genetic variance component for miRNAs is significantly less than the genetic variance component for mRNA traits, although our analyses herein suggest this is unlikely to be the case.

Because miRNAs are known to downregulate mRNAs, we developed a strategy to decrease the FDR of detecting miRNA eQTLs by restricting the mapping loci for each miRNA to regions of the genome that are *a priori* postulated to be weakly linked to miRNAs. In theory, the set of genes enriched for potential miRNA targets could share common QTL with the corresponding miRNAs (Figure 2A and B). Moreover, while miRNAs are known to bind imperfectly to the 3' UTR of target mRNAs, a perfect Watson-Crick pairing of ~6–8 nucleotides at the 5' region of the miRNA has been shown to be important in miRNA target recognition (Bartel, 2009). This region, known as the seed region, is typically conserved across species and spans nucleotides 1–8 of the mature miRNA transcript. Experimental perturbation studies using increased levels of miRNAs or knockdown of ectopic miRNA levels have resulted in decreased or increased levels of mRNAs that are enriched for the miRNA seed (Lim *et al*, 2005; Linsley *et al*, 2007). Hence, for each miRNA, we identified a set of mRNA expression traits that contained at least one hexamer region within the 3' UTR of the mRNA transcript that could potentially bind the seed of the given miRNA. These gene sets were then filtered to contain only genes that were significantly negatively correlated with the corresponding miRNA. Each set represented a set of mRNA expression traits that were highly enriched for potential direct target mRNAs of a given miRNA. We then searched for the corresponding miRNA eQTLs within regions that are significantly associated with one or more



**Figure 2** Detection thresholds of miRNA eQTLs. **(A)** Location of seed regions for miRNAs. **(B)** Illustration of procedure used to increase the statistical power of detecting miRNA eQTLs. Messenger RNA expression traits that were negatively correlated with a given miRNA and contained at least one corresponding hexamer seed region in the 3' UTR were inferred to be potentially downregulated by the miRNA. **(C)** False-discovery rates (FDRs) as a function of LOD score threshold for miRNA eQTLs.

mRNA transcript in the associated set of highly enriched potential direct targets (i.e. within 15 cM of the peak each potential target mRNA eQTL) (Figure 2).

To begin, we computed all pairwise correlation between all 39 557 mRNA transcripts and 183 miRNA transcripts. Next, we selected all pairs of significantly correlated miRNA and mRNA transcripts using a *P*-value threshold of  $<3.98 \times 10^{-4}$ , which corresponds to an FDR  $<1\%$ . At this threshold, 465 646 significant Pearson correlations or  $\sim 6\%$  of the correlation matrix were selected. For each of the 183 miRNAs, we selected all mRNA transcripts that were significantly negatively correlated with the given miRNA and contained at least one hexamer region matching the reverse complement of the miRNA seed in the 3' UTR of the mRNA transcript, resulting in 183 distinct sets of mRNA transcripts. On average, each set contained 261 mRNA transcripts. For each set, we then identified all eQTLs (LOD threshold  $>4.4$ , FDR  $<10\%$ ) associated with each mRNA transcript in the set. miRNA eQTL peaks that fell within 15 cM of the associated set of mRNA QTL peaks were considered to map to the same region as their potential direct targets.

By focusing only on the miRNAs eQTLs that overlapped with the eQTLs of their potential direct targets, we were able to decrease the FDR of detecting miRNA eQTLs (Figure 2C). Using this strategy, a LOD score threshold of  $>3.1$  corresponds to an FDR  $<10\%$ . At this threshold, we identified 72 miRNA eQTLs corresponding to 56 miRNAs and of these, only 11 (or 15%) were *cis* eQTLs with a large proportion of *cis* eQTLs ranking among the top 10 eQTLs with the strongest LOD scores (Supplementary Table 1A). Here, 6 out of the top 10 strongest eQTL signals were *cis* eQTLs. On average, each eQTL explained 13.5% of the variation observed in miRNA transcript abundance levels. We note that 8 out of the 10 miRNA eQTLs that were previously detected during the genome-wide search overlapped with the miRNA eQTLs that were identified in the more restricted analysis. Also, consistent with eQTLs detected for mRNA expression traits, the strongest LOD scores were realized at markers that were proximal to the physical location of the structural miRNA (i.e. were *cis* eQTL (Doss *et al*, 2005)).

The increased number of miRNA eQTL detected using the mRNA eQTL as a filter suggest that many miRNA may have a significant genetic variance component and that we simply were not well powered to detect them in this study. Mmu-miR-34a is particularly noteworthy given the size of its eQTL signature—that is, it maps to five eQTLs on different chromosomes demonstrating a strong genetic component driving variation of mmu-miR-34a expression levels in the BXD cross.

We next sought to determine if there were key loci involved in regulating many miRNAs. In analyzing the distribution of miRNA eQTLs across the genome (when restricted to those overlapping with target mRNA eQTLs), we identified a strong eQTL hotspot on chromosome 13 and a weaker hotspot on chromosome 17 (Supplementary Figure S1). Of the 72 eQTLs identified, 42% mapped to chromosome 13, suggesting the presence of a key regulator influencing the expression levels of many miRNAs. If we consider individual chromosomes as distinct bins, the *P*-value for the null hypothesis (assuming a Poisson distribution with 72 eQTLs spread across 20 bins) that this distribution occurred by chance would be  $<5.03 \times 10^{-18}$ . Similar analysis with mRNA eQTLs reveals the presence of many hotspots for the mRNA expression traits. Because many more mRNA expression traits were analyzed in this study, a bin size of 2 cM evenly spaced across the genome was used to detect mRNA eQTL hotspots. Overall, we detected seven mRNA eQTL hotspots where each hotspot is defined to comprised  $>1\%$  of the total number of eQTLs (computed using a Poisson distribution with mean=9.52). These hotspots localize to chromosomes 2, 4, 7, 9, 12, 13, and 17. The probability of detecting  $>1\%$  of eQTLs in one of the 768 bins is  $\sim 6.5 \times 10^{-40}$ .

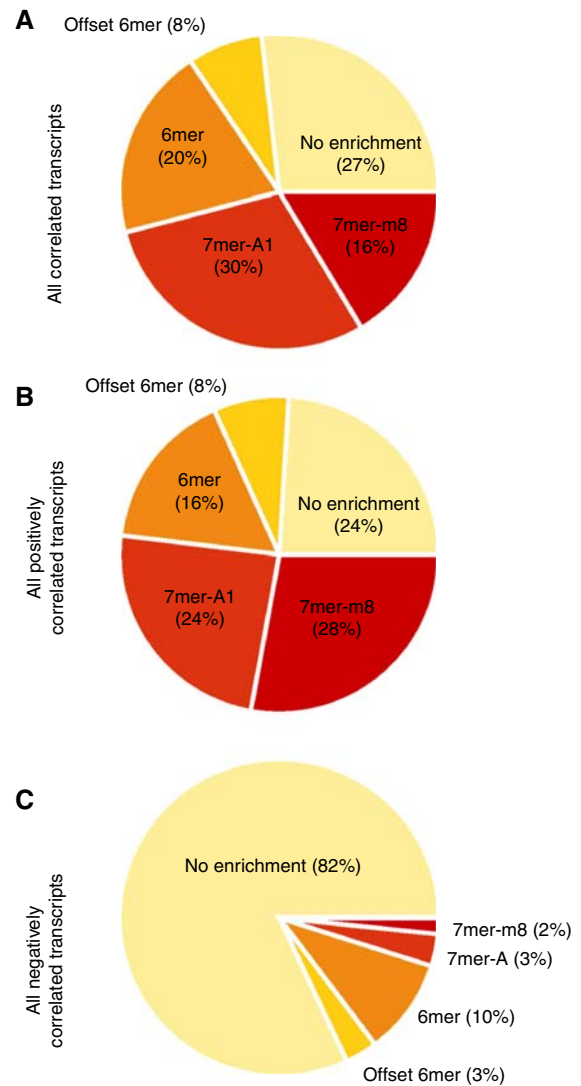
In order to better compare the location of miRNA eQTL hotspots to mRNA eQTL hotspots, we recomputed the probabilities of an miRNA eQTL hotspot using 2 cM bins. Our analysis indicates that the eQTL hotspots for miRNAs and mRNAs on chromosome 13 are  $<4$  cM apart. Interestingly, Dicer and Drosha, two key enzymes involved in miRNA biogenesis, are respectively located on chromosome 12 and 15, and hence do not overlap with the detected miRNA eQTL hotspots.



## Correlation analysis between miRNA and mRNA expression levels in mice

Using the same correlation matrix computed above (i.e. all pairwise correlation between the 39 557 mRNA and 183 miRNA transcripts), we identified 465 646 miRNA–mRNA trait pairs that were significantly correlated at an FDR < 1% ( $P$ -value <  $3.98 \times 10^{-4}$ ). By tabulating the number of significantly correlated mRNA transcripts per miRNAs, we found that a number of miRNAs were very broadly connected to tens of thousands of mRNAs, supporting the idea that some miRNAs behave as hub-nodes in the liver gene regulatory network. On average, each miRNA was significantly correlated with 2545 mRNA transcripts, with all 183 assayed miRNAs showing significant correlation to at least one mRNA transcript. In fact, eight miRNAs (mmu-miR-128b, mmu-miR-142-5p, mmu-miR-142-3p, mmu-miR-181a, mmu-miR-181c, mmu-miR-186, mmu-miR-212, and mmu-miR-342) were significantly correlated with > 10 000 mRNA transcripts, effectively tracking with the state of the liver gene regulatory network. Conversely, each mRNA transcript showed significant correlation to a mean of 19 miRNA transcripts. In fact, 24 869 out of 39 557 transcripts (63%) were significantly correlated with at least one miRNA. This large correlation signature between miRNAs and mRNAs suggests that the large majority of miRNAs affect the expression levels of many genes and that their own expression levels may themselves be subject to complex regulation. This explanation would be consistent with the weak genetic signature that we identified for the miRNAs. In this case, assuming that miRNAs are regulated by many factors, the expectation would be that many *trans*-acting eQTLs would influence the level of miRNA transcript abundances, each with a small effect, necessitating a larger sample sizes for miRNA eQTL detection.

As perfect reverse complementarity at the seed region has been shown to improve prediction of miRNA targets over background levels (Bartel, 2009), we tested each set of correlated mRNA transcripts for seed enrichment of their respective miRNA. Only seeds within the 3' UTR of the mRNA transcript were included in this analysis. Briefly, we started with the full set of correlated mRNA transcripts for each miRNA (referred to as the miRNA signature set) and computed the seed enrichment levels for each set using a hypergeometric distribution. Because there is a continuum in efficiency of various seed matches (e.g. longer seed matches are better predictors of miRNA targets), we calculated the seed enrichment of the various seeds (Figure 2) and then selected for the strongest seed region (assuming the following hierarchy in site efficacy: 7mer-m8 > 7mer-A1 > 6mer > offset 6mer) that yielded significant enrichment among the transcripts in the miRNA signature set at an initial Bonferroni-corrected  $P$ -value threshold of < 0.05. As expected, 73% of miRNA signature sets were enriched for at least one seed region (Figure 3A). Surprisingly, given that miRNAs have largely been shown to downregulate levels of their target transcripts (Lim *et al*, 2005; Linsley *et al*, 2007), our expectation was that the observed seed enrichment would be driven by mRNA transcripts that were negatively correlated with the given miRNA. However, when we partitioned the transcripts into those that were positively and negatively correlated with the miRNAs, mRNA transcripts



**Figure 3** Pie charts showing the distribution of seed enrichment for all 183 miRNA signature sets. For each miRNA signature set, the strongest seed match (assuming the following hierarchy in site efficacy: 7mer-m8 > 7mer-A1 > 6mer > offset 6mer) is chosen. **(A)** Distribution of seed enrichment using the full miRNA–mRNA correlation results. **(B)** Distribution of seed enrichment using only positive correlations between miRNA–mRNAs. **(C)** Distribution of seed enrichment using only negative correlations between miRNA–mRNAs.

that were positively correlated constituted the bulk of transcripts within the miRNA signature set that were driving the strong observed enrichment of miRNA seed regions (Figure 3B and C). In fact, only 18% (33/183) of negatively correlated miRNA signature sets were significantly enriched for the corresponding miRNA seed region compared with 73% (134/183) for positively correlated miRNA signature sets. This result suggests that a large proportion of miRNAs–mRNA interactions involve feedback loops as have been demonstrated in a number of studies (Martinez *et al*, 2008; Yu *et al*, 2008).

For each miRNA signature set, we performed an enrichment analysis using the Gene Ontology (GO) Biological Process categories and KEGG pathways. Here, we used the Fisher's exact test to compute the significance of the overlap between

**Table 1** MicroRNA signature sets with at least one corresponding hexamer per gene that are enriched for KEGG pathways cell-cycle genes

MicroRNA	Seed (6mer)	miRNA signature set size	Overlap (enrichment) <sup>a</sup>	Nominal P-value <sup>b</sup>	Corrected P-value <sup>b</sup>
mmu-mir-20a	CACTTT	377	33 (3.27)	2.50E-10	4.94E-08
mmu-mir-449	CACTGC	342	26 (2.84)	5.81E-07	1.15E-04
mmu-mir-19a	TTGCAC	306	21 (2.56)	4.27E-05	8.45E-03
mmu-mir-181b	GAATGT	622	38 (2.28)	2.29E-07	4.54E-05
mmu-mir-181c	GAATGT	693	40 (2.16)	4.51E-07	8.93E-05
mmu-mir-181a	GAATGT	698	40 (2.14)	5.49E-07	1.09E-04
mmu-mir-142-5p	CTTTAT	742	42 (2.11)	3.41E-07	6.75E-05
mmu-mir-18	CACCTT	495	28 (2.11)	6.60E-05	1.31E-02
mmu-mir-128b	CACTGT	811	43 (1.98)	1.49E-06	2.96E-04
mmu-mir-128a	CACTGT	782	41 (1.96)	4.14E-06	8.21E-04

<sup>a</sup>The number in parentheses is the fold-change enrichment. The fold-change enrichment is defined to be the observed overlap divided by the observed overlap.

<sup>b</sup>The nominal P-value represents the Fisher's exact test statistics under the null hypothesis that the number of overlapping genes between the two sets is the same as that from a background set of 4033 genes and the set of KEGG pathways cell-cycle genes consisting of 108 genes. The corrected P-value represents the Bonferroni-corrected P-value (nominal P-value multiplied by the number of sets searched).

each miRNA signature set and known sets of genes with functional and pathway information. A Bonferroni-corrected P-value of 0.05 (nominal P-value multiplied by the number of sets in each of the corresponding databases) was used as the threshold in determining significant overlap. In all cases, the total number of transcripts on the Agilent array was used as the background set. In addition, for GO Biological Process and KEGG Pathway enrichment analysis, sets with > 500 genes were not considered, given that sets of this size are too general to be of practical benefit. The top 10 results for each category ranked by fold-change enrichment are shown in Supplementary Table 2.

From the annotation results, we found that in liver tissue, a large fraction of the surveyed miRNAs (33 % of miRNAs with successful biological process annotations) are putatively involved in DNA replication and cell-cycle regulation (Supplementary Figure S3A). Of the sets tested, the top categories in terms of number of overlapping miRNA signature sets were DNA replication, cell division, and M phase of the mitotic cell cycle, mitosis, and mitotic cell cycle (Supplementary Figure S3B). These categories include miRNAs known to have a role in DNA replication and cell cycle, such as mmu-miR-34a and mmu-miR-15b, respectively.

When we analyzed the miRNA signature sets using KEGG pathways, we found similar results in the sense that the majority of miRNA signature sets are enriched for genes involved in cell cycle and DNA replication pathways (Supplementary Figure S3C). In addition, 18 miRNAs were linked to genes that have a role in the complement and coagulation cascade pathways. A further analysis of the miRNAs predicted to be involved in cell-cycle regulation revealed that 10 of the 19 miRNA signature sets were additionally enriched for the corresponding seed region of their respective miRNAs, providing putative regulatory links between the correlated set of mRNA transcripts and the corresponding miRNA transcript. These sets of significantly correlated mRNA transcripts that contain at least one 6mer sequence in the 3' UTR were also similarly enriched for genes involved in cell cycle (Table 1).

Using a Fisher's exact test, we computed the significance of association between all pairwise comparison of the 19 miRNA signature sets. Here, we found that despite being correlated to miRNAs with different seed regions, most of the miRNA signature sets that are predicted to be involved in cell-cycle

regulation were also significantly associated with one another. Hence, our results suggest that distinct miRNAs collectively function to modulate levels of the same set of mRNA transcripts that are all involved in the same cellular process. We further examined the other KEGG pathways that showed significant association with more than five miRNA signature sets and in each pathway, found that many, if not all of the miRNA signature sets that are associated with a given pathway to be highly significantly associated with one another. There were only a few examples where pairs of miRNA signature sets that were associated with the same KEGG pathway were not significantly associated with one another. For example, out of the 19 miRNA signature sets that show strong association with cell-cycle regulation, the miRNA signature sets that are significantly correlated with mmu-miR-25 and mmu-miR-296 share relatively few transcripts and no more than would be expected by chance alone (Fisher's exact test P-value ~0.99). Another example is the collection of transcripts within the miRNA signature set that is significantly correlated with mmu-miR-193, which differs from the collection of transcripts within the miRNA signature set mmu-miR-128 and mmu-miR-183 families despite all miRNA signature sets being enriched for the KEGG ribosome pathway.

Because a large majority of miRNA signature sets are enriched for genes containing the seed region of the given miRNA, we opted to annotate the sets of miRNA signature sets using only genes that contained at least one 6mer seed region in the 3' UTR region of the gene. That is, we wanted to identify those pairwise relationships that not only exhibited a significant correlation, but that also involved mRNA with a 3' UTR region that contained at least one 6mer seed region of the corresponding correlated miRNA. As expected, we were able to provide functional annotations for many of the miRNAs using this filter. The top 10 results for each database are shown in Supplemental Table 3. The most significant result was obtained for mmu-miR-20a using the KEGG pathway database. Previous studies have shown that this miRNA has a role in cell cycle; our analysis showed a three-fold enrichment in the number of overlapping genes involved in cell cycle, compared with what we would have expected by chance. Overall, using a simple correlation analysis, we identified functional categories for 73 different miRNAs and implicated 57 miRNAs in various pathways. Using both the correlation data and the presence of

a seed region, we have provided putative functions for 54 miRNAs using GO Biological Process categories, and 26 miRNAs using KEGG Pathways.

### Causal associations between miRNAs and mRNAs

Because miRNAs have been shown to have regulatory roles in many important biological pathways (Mallory and Vaucheret, 2006; Bartel, 2009; Kai and Pasquinelli, 2010), we were interested in distinguishing mRNA transcripts supported as inducing changes in miRNA trait expression from those serving as targets (both direct and indirect) of the miRNAs. Using integrative genomics approaches, previous studies have shown success in teasing apart the nature of the relationship between pairs of correlated quantitative traits such as mRNA and clinical phenotypes (Mehrabian *et al*, 2005; Schadt *et al*, 2005; Yang *et al*, 2009). Briefly, the approach leverages DNA sequence variation as a causal anchor to identify the best fitting model that describes the relationship between pairs of traits that are linked to the same genetic locus. More importantly, this approach has met with success at the experimental validation stage. For example, Yang *et al* (2009) recently validated eight out of nine predicted causal genes for obesity-related traits using transgenic and knockout mice.

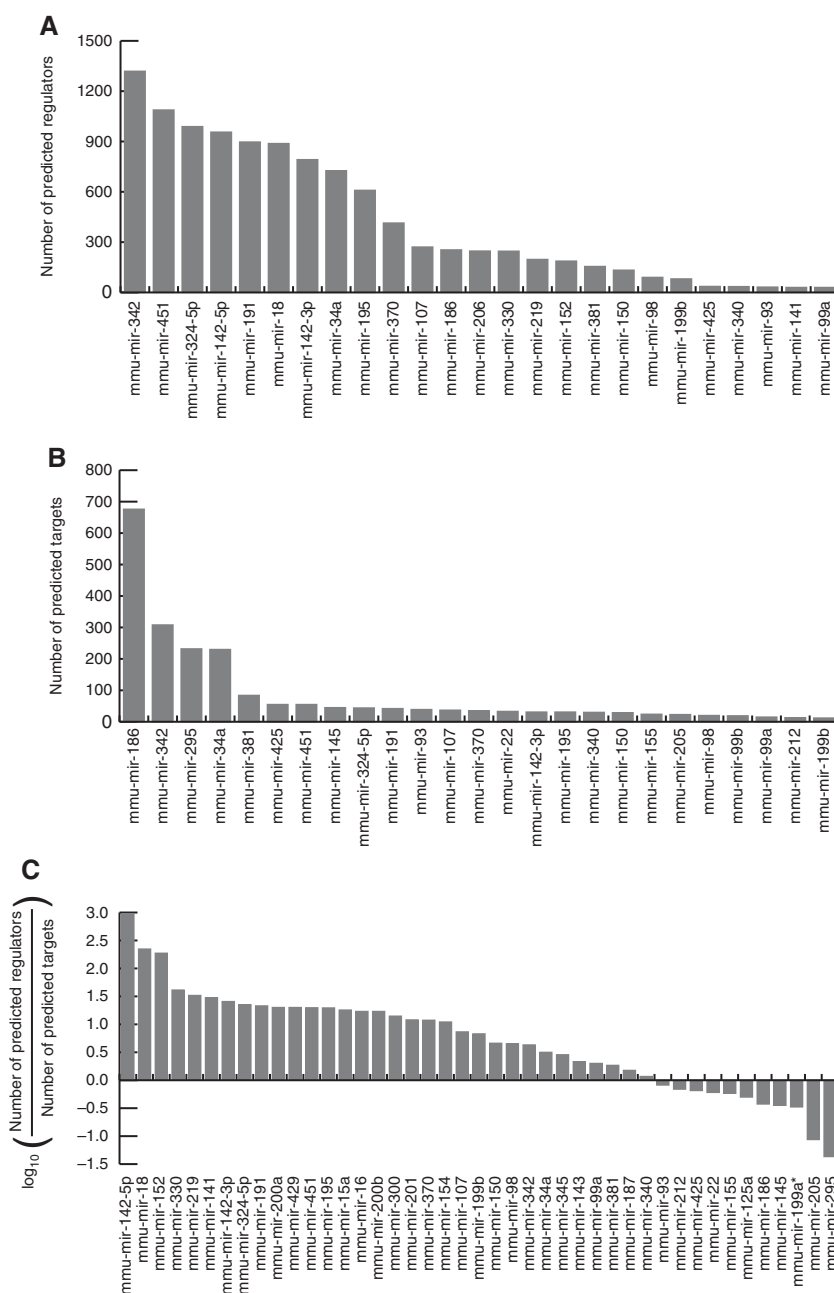
We applied a variation of a previously described statistical procedure (Schadt *et al*, 2005) to identify mRNAs that respond to changes in miRNA expression levels (miRNA targets), as well as mRNAs that perturb expression levels of miRNAs. First, we identified all miRNA and mRNA trait pairs linked to a common genomic region at an LOD score threshold of 3.4 (corresponding to a 10% FDR). Next, we identified 44 370 miRNA–mRNA trait pairs with closely linked eQTLs ( $<15$  cM). After filtering for significant correlations between the pairs of miRNA–mRNA trait pairs with closely linked eQTLs (at a  $P$ -value threshold of  $3.98 \times 10^{-4}$ , corresponding to a 1% FDR was used), we applied a variation of a previously described causal inference procedure (Schadt *et al*, 2005) to classify each miRNA–mRNA trait pair into one of three categories (with respect to the miRNA as the reference): (a) causal, where an eQTL for miRNA expression leads to changes in mRNA expression (miRNA targets); (b) reactive, where eQTL for mRNA levels leads to changes in miRNA expression (miRNA regulators); and (c) independent, eQTL independently drive miRNA and mRNA levels (independent). As we have shown previously, closely linked eQTLs associated with different traits, when treated as a single eQTL to assess the causal relationship between the corresponding traits, is inferred as an independent relationship with high power, thus greatly reducing the likelihood of wrongly inferring causal relationships in such cases (Schadt *et al*, 2005).

Surprisingly, instead of identifying large numbers of predicted causal miRNA targets, we identified many protein-coding genes supported as regulators of miRNA expression (Figure 4). A total of 2218 miRNA target relationships (miRNA  $\rightarrow$  mRNA) were identified while 11 114 ( $\sim 5$ -fold increase) were identified as causal for the inverse relationship (mRNA  $\rightarrow$  miRNA). For the set of miRNAs with predicted targets, a mean of 50 target mRNAs were identified. However, for the set of miRNAs with predicted regulators, we detected a mean of 198 protein-coding genes per miRNA. As shown in

Figure 4C, a substantial fraction of miRNAs are predicted to respond to many more mRNA targets as opposed to driving expression level changes in their downstream targets. Of the 43 miRNAs with both predicted regulators and targets, only 11 (26%) are predicted to induce downstream changes in more mRNA transcripts than to respond to changes in mRNA transcript abundances. On average, these 11 miRNAs regulate 2.2 mRNA transcripts per miRNA. However, for the remaining miRNAs where the mRNA regulators outnumber the target mRNAs, a mean of 55 regulating genes per miRNA was observed. Hence, within the mRNA–miRNA interaction network of the BXD liver tissue network, many miRNAs are predicted to be more strongly perturbed by changes in the mRNA expression network than to drive downstream changes in the network, suggesting that miRNAs serve as sensors of transcriptional network states; and then in response to network state changes, drive changes in more focused sets of transcripts underlying key biological processes.

As miRNAs have largely been demonstrated to down-regulate a large number of mRNA targets and there is comparatively less in the literature regarding protein-coding genes regulating miRNAs, we were motivated to investigate possible causes of bias in the methodology. Because differences in the error distribution of the two data sets (miRNA and mRNA data sets) are known to subtly affect the false positive and negative rates, we ran a simulation study to estimate the difference in measurement error between these two sets. We considered a range of measurement error differences, from no differences (0%) on up to a two-fold increase (100%). Of particular relevance to this study, differences in noise levels intrinsic to the type of assay being performed would result in a difference in statistical power to infer the correct relationship. For example, qPCR platforms are widely believed to be the gold standard for gene expression profiling providing greater sensitivity and a wider dynamic range. Using  $T_1$  and  $T_2$  as the quantitative traits of interest (for miRNA and mRNA, respectively), we performed 10 000 simulations for each of the three models: miRNA target (causal), miRNA regulator (reactive), and independent models. Here, we set  $\epsilon(T_1) < \epsilon(T_2)$  and plotted the results obtained. As seen in Figure 5, the power of the causality procedure that we applied to identify the causal model ( $L \rightarrow T_1 \rightarrow T_2$ ) increases slightly for small increases in error in  $T_2$ . Beyond that slight increase, the power to identify the causal model is independent of increases in the error in  $T_2$ . However, when we simulate the reactive model ( $L \rightarrow T_2 \rightarrow T_1$ ), an increase in measurement error in  $T_2$  leads to a dramatic decrease in power to detect reactive models. In fact, when the measurement error for  $T_2$  is twice that of  $T_1$ , the power of the procedure to detect the true simulated relationship drops by 20%. Hence, in relation to the miRNA and mRNA data sets presented here, assuming that microarray data carries a higher noise level than qPCR, the distribution of measurement error is likely larger in the mRNA data set than in the miRNA data set. Therefore, there should be little to no loss in power to detect the causal miRNA  $\rightarrow$  mRNA relationships (miRNA targets) while the number of predicted causal regulators of miRNA is likely to be an underestimation of the actual number.

Of greater concern, however, is the increase in the number of causal models falsely identified when  $\epsilon(T_1) < \epsilon(T_2)$ . As shown in Figure 5, in the case of the independent model, a two-fold



**Figure 4** Histograms of number of predicted targets and regulators. For the top two charts, only the top 25 miRNAs ranked in descending order are shown. (A) predicted regulators; (B) predicted targets; (C) log ratio of the number of predicted regulators over the number of predicted targets.

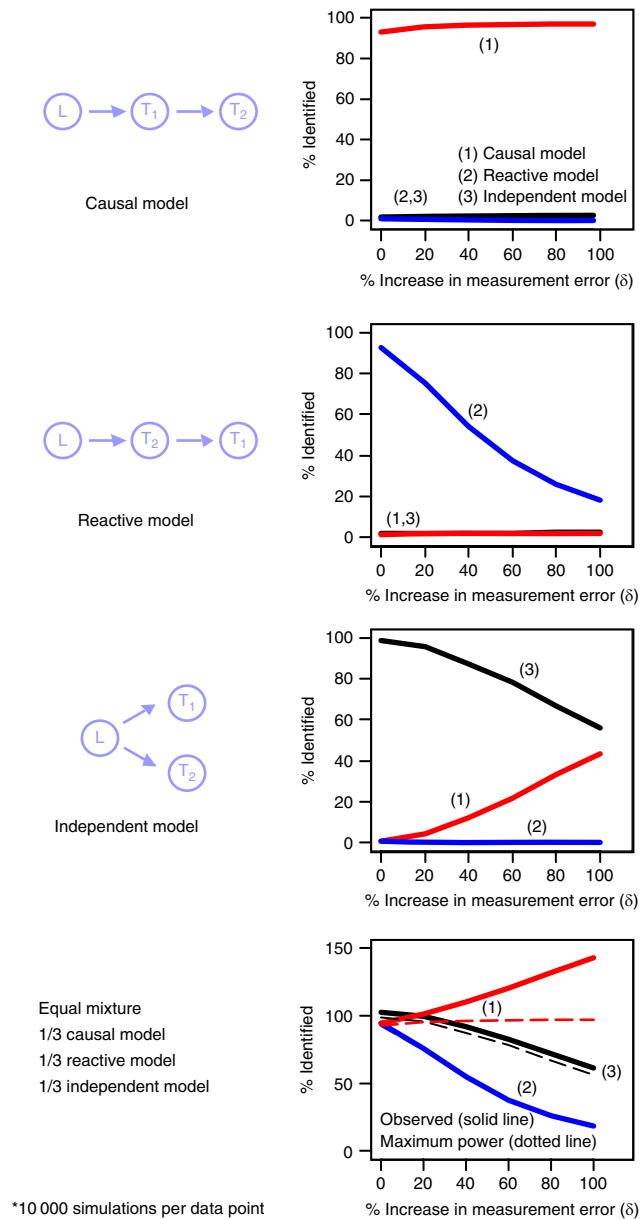
increase in measurement error would drive the procedure to falsely select the causal model as the best-fitting model. Hence, the false-positive rate of identifying the causal model increases with increasing difference in measurement error between the two data sets. Assuming an equal proportion of causal, reactive, and independent models, increasing error in  $T_2$  would lead to more causal calls and less reactive and independent calls. This increase in false causal calls and loss of power to detect true reactive and independent cases would lead to an artificial inflation of the number of causal cases identified relative to the number of reactive cases. However, we have instead observed in our data set the opposite trend;

that is, the number of reactive calls (mRNA  $\rightarrow$  miRNA) greatly out number the number of causal calls (miRNA  $\rightarrow$  mRNA). Hence, depending on the percent differences in measurement error between the two data sets, the earlier observation of a five-fold increase in the number of genes that lead to downstream changes in miRNA expression relative to the number of predicted miRNA targets is, at the worst, a conservative estimate.

## Discussion

The regulatory mechanisms of miRNA-mediated regulation present a complex picture. Initially discovered to inhibit





**Figure 5** Percent of true cases identified as a function of percent differences in measurement error between the two traits. In all cases shown, the error in  $T_2$  is larger than in  $T_1$ .

protein translation, miRNAs were later discovered to down-regulate mRNA transcript levels of their target mRNA. Further studies showed that in the right context, they could function to activate protein translation. Because miRNAs regulate their targets via an imperfect binding of the 22 nucleotide molecule to their target mRNAs, each miRNA has the potential to directly regulate hundreds of mRNA transcripts. Conversely, the length of most mRNA transcripts provide potentially many binding sites for the small non-coding regulatory molecule and miRNAs have in fact been shown to act cooperatively to regulate a given transcript (Saetrom *et al*, 2007; Hobert, 2008). miRNAs have also been theorized to act as buffers in regulating mRNA transcripts (Hornstein and Shomron, 2006; Herranz and Cohen, 2010).

In addition to the complexity of miRNA–mRNA relationships, there are numerous mechanisms regulating both overall content of miRNAs (e.g. factors involved in miRNA biogenesis) and specific miRNAs themselves. For example, uridylation of mature miR-26a has been shown to reduce miR-26a activity without significantly affecting its expression levels, while uridylation of the precursor form of let-7a targets the miRNA precursor for degradation (Kai and Pasquinelli, 2010). Another study demonstrated the buffering effect of mRNA from pseudogenes against miRNA activity (Poliseno *et al*, 2010). Here, the pseudogene PTENP1 was shown to act as a decoy for miRNAs targeting PTEN with decreasing levels of PTENP1 leading to increase activity of PTEN-targeting miRNAs on the PTEN transcript (Poliseno *et al*, 2010). These studies suggest not only complex regulation of biological pathways via miRNAs, but also intricate regulation of miRNAs transcript abundances themselves and similarly complex regulation of miRNA activity.

Our findings in this paper provide further support and characterization of the complexity of miRNA-mediated regulation in gene regulatory networks. Using an integrative genomics approach, we have explored the genetic basis of miRNA expression variation for a substantial fraction of known miRNAs (~30% of those registered in mirBase (version 15)). Despite surveying approximately a third of known mouse miRNAs today, many of the surveyed miRNAs were expressed in liver. We showed that variation in transcript abundances of many miRNA are linked to DNA sequence variation and that, in a number of ways, the genetics of miRNA gene expression is similar to mRNA expression traits. Most notably, the strongest LOD scores for miRNA expression traits are associated with *cis*-acting miRNA eQTLs, which is a frequent observation with mRNA expression traits. In addition, we were similarly able to detect the presence of miRNA eQTL hotspots, that is, regions that account for the trait variation of a disproportionately larger number of expression traits than would be expected by chance.

Because only a handful of miRNA eQTLs were detected using a genome-wide scan, we postulate that the effects of DNA variation on miRNAs are more subtle than the effects of sequence variation on mRNA transcripts. In this case, each miRNA may be regulated by multiple eQTLs, each with an effect size that is typically smaller than that detected for mRNA eQTLs. Alternatively, it is also possible that the most miRNAs are simply not affected by polymorphisms presence in the mouse genome or that the proportions observed in this study are simply a chance occurrence unintentionally resulting from surveying only a third of miRNAs that have been discovered. The former explanation would be concordant with miRNAs being regulated by an intricate network of multiple genes.

To address this ambiguity, we restricted the potential loci to those that only encompassed eQTLs for the set of potential target mRNAs of each miRNAs and noted an increase in statistical power to detect miRNA eQTLs. Overall, we were able to detect more than a seven-fold increase in miRNA eQTLs strongly suggesting that the variation in miRNA expression trait that is accounted for by DNA sequence variation is indeed smaller than that of mRNA expression trait. From the standpoint of gene regulatory networks, this observation

supports the notion of complex interplay between miRNAs and the genes located in those regions.

Of the set of miRNA eQTLs detected, 85% were *trans* eQTLs, suggesting that many miRNAs are subject to controlled extending beyond local sequence variants. In the simplest model, this could be viewed as a perturbation to an element or gene contained within the *trans* locus that indirectly then influences miRNA expression trait levels. Recent studies on *trans*-splicing events and the prevalence of chimeric RNA suggests alternative implications for *trans* eQTLs, where such events may directly lead to increase or decrease in miRNA transcript levels (Gingeras, 2009).

In examining the correlation structure between miRNA and mRNA, we noticed that counter to the widespread acceptance that miRNAs downregulate gene expression, a surprisingly large proportion of miRNAs signature sets (i.e. sets of mRNA transcripts that are significantly correlated with a given miRNA) that were negatively correlated with their respective miRNA showed poor enrichment for the corresponding miRNA seed in the 3' UTR of the mRNA transcripts within the set. Instead, many miRNAs signature sets that displayed a positive correlation with the individual miRNAs were observed to be enriched for the seed region of each respective miRNA in the 3' UTR of the mRNA. While unexpected, strong positive correlations between miRNAs and mRNAs have been previously reported in a number of studies (Liu *et al*, 2007; Tsang *et al*, 2007; Nunez-Iglesias *et al*, 2010). Additionally, using the expression of host genes with embedded miRNAs as a proxy for expression of the embedded miRNA, Tsang *et al* (2007) found significant positive and negative correlation between the embedded miRNAs and targets of the embedded miRNAs as predicted by TargetScanS algorithm, an miRNA prediction algorithm that utilizes the conserved Watson–Crick pairing (i.e. miRNA seed) between the miRNA and its target mRNA (Tsang *et al*, 2007).

One possible explanation for the occurrence of positive correlation between a given miRNA and their respective target that has been previously been postulated is the presence of feedback motifs. Feedback motifs are known to be common within gene regulatory networks and studies specific to miRNA–mRNA networks have shown an enrichment of feedback loops (Martinez *et al*, 2008). Others have elucidated and classified specific feedforward and feedback circuit motifs that could explain the both positive and negative miRNA target mRNA correlations (Tsang *et al*, 2007). While examples of miRNA activating transcription are not yet widespread, this represents another plausible explanation for our observation.

Further analysis of the causal relationship between miRNAs and mRNAs suggests that for many miRNAs, the set of protein-coding genes predicted to regulate levels of a given miRNA exceeds that of its downstream target mRNAs. Simulation studies that take into account the difference in noise levels between the two data sets indicate that the number of predicted miRNA targets is likely to be an overestimation while the number of predicted miRNA regulators is likely to be an underestimation. Overall, these results add to the increasing complexity of miRNA-mediated regulation.

Using GO Biological Process and KEGG pathways, we show that almost a quarter of the surveyed miRNAs are correlated

with mRNA transcripts involved in DNA replication and cell-cycle regulation within the context of a liver gene regulatory network. These miRNAs are often correlated with the sets of mRNA transcripts that are significantly associated with one another, suggesting that these miRNAs cooperatively or redundantly act to regulate a core set of mRNA transcript within a given pathway. In a few cases, two or more miRNAs were significantly correlated with two distinct sets of mRNA transcripts belonging to the same pathway.

To our knowledge, our results represent one of the first examinations of the genetic basis of variation in miRNA expression and exemplify how integrative genomics approaches may be used to elucidate multiple insights surrounding the regulatory circuitry of novel classes of RNA molecules. With advances in next-generation sequencing technologies, this approach can be extended to characterize the full set of transcriptional units (i.e. non-coding RNA, chimeric RNA, mRNA etc.) to decipher the myriad of molecular regulatory relationships that governs phenotypic behavior. From our study, given the abundance of miRNA–mRNA interaction, it seems that miRNAs may be sensing network states and responding to entire network changes in mRNA levels. We hypothesize that miRNAs then respond in a programmed manner to drive pathway changes via modulation of specific sets of mRNA.

## Materials and methods

### Animal and tissue collection

All procedures were performed with the approval of Merck & Co. (Whitehouse Station, NJ, USA) and the Institutional Animal Care and Use Committees at the Jackson Laboratories (JAX West, West Sacramento, CA, USA). A cohort of mice was derived from a standard F2 intercross constructed by breeding two standard inbred strains, C57BL/6J (B6) and DBA/2J (referred to as the BXD cross). All F2 breeding was performed at JAX West. Approximately 250 F2 mice were bred from F1 mice ( $N=12$  for each gender) constructed from breeding the B6 and DBA parental strains. Mice were weaned into cages of three same-sex pups per litter at 3 weeks of age. These three littermate mice remained together for the duration of the study. For our study, we selected 127 male animals that had been raised on a standard rodent chow diet (Purina Chow from Ralston-Purina Co., St Louis, MO) for 5 weeks and then switched to a high-fat diet for 12 additional weeks. For both diets, the mice were fed *ad libitum*. At 20 weeks of age, the mice were euthanized and liver tissues were collected and flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  before RNA isolation. Mice were fasted overnight before being euthanized.

### RNA sample preparation and microarray hybridization

RNA preparation and array hybridizations were performed at Rosetta Inpharmatics. A custom array consisting of 39 557 oligonucleotides was designed for this study using the mouse Unigene clusters and combined with RefSeq sequences and RIKEN full-length cDNA clones. The custom ink-jet microarrays were manufactured by Agilent Technologies (Palo Alto, CA). RNA sample preparation and microarray hybridization was performed as previously described (Chen *et al*, 2008). Briefly, 3  $\mu\text{g}$  of total RNA was reverse transcribed and labeled using Cy3 or Cy5 fluoro-chrome. Labeled complementary RNA from each F2 animal was hybridized against a control RNA pool. Here, the control RNA pool represents a cross-specific pool consisting of labeled cRNAs constructed from equal aliquots of RNA from all samples. Analysis of expression data was conducted as previously described (Chen *et al*, 2008). In summary, arrays were quantified on the basis of

spot intensity relative to background and adjusted for experimental variation between arrays using the average intensity over multiple channels and fitted to a previously described error model to determine significance (Chen *et al*, 2008). Gene expression measures are reported as the ratio of the mean  $\log_{10}$  intensity. All microarray data have been deposited in the NCBI GEO database under accession number GSE16886.

## miRNA profiling

We also assayed 187 miRNAs representing a wide dynamic range (<10 copies per cell to 300 000 copies per cell) using a previously published PE-qPCR technique (Raymond *et al*, 2005). These miRNAs represent the complete repository of miRNAs in miRBase (Griffiths-Jones, 2004; Griffiths-Jones *et al*, 2006, 2008) when the high-throughput miRNA profiling platform was designed. For each sample, we performed four technical replicates using an estimated total of 10  $\mu$ g of total RNA per sample. Each miRNA assay is then processed into a single reported value. miRNA assays with missing values in >67 individuals (i.e. present in  $\leq 60$  individuals) were discarded. Intensity values were normalized to reflect the number of miRNA copies per cell (assuming that the total mass of RNA per cell is 10 pg). Values were then transformed using  $\log_{10}$  transformation. The full miRNA expression data set can be found in Supplementary Table 4.

## Genotyping and statistical analysis

Genomic DNA was isolated from tail sections by Bioserve (Beltville, MD) using standard methods and genotyping was performed by Affymetrix (Santa Clara, CA) using the Affymetrix GeneChip Mouse Mapping 5K panel. From the panel of 5000 SNP markers, 2804 markers informative for the BXD cross and evenly spaced across all chromosomes, excluding the Y chromosome, were selected for use in all analyses. The complete data set can be found in Supplementary Table 5. For mRNA expression traits, we used the mean  $\log_{10}$  expression ratio; and for miRNA expression levels, we applied a  $\log_{10}$  transformation to the normalized intensity values. QTL analysis for both mRNA and miRNA expression traits were conducted as described previously (Schadt *et al*, 2003). Briefly, we generated a complete linkage map using MapMaker QTL (Lincoln *et al*, 1993). Using QTL cartographer (Basten *et al*, 1999), we then performed interval mapping to detect eQTLs for mRNA and miRNA expression traits. For the miRNA–mRNA correlation analysis, we computed all pairwise comparison the Pearson's correlation between the full set of 183 miRNA and 39 557 mRNA expression traits.

## Estimating FDRs

For both QTL analysis and correlation analysis, standard permutation testing was applied to determine significance thresholds. In the case of mRNA eQTL detection and miRNA–mRNA correlation analysis, five independent permutations were performed. Each permutation was performed in a way that preserves the gene correlation structure. For miRNA eQTL detection, we estimated the associated FDR values using 10 independent permutation runs.

## Causal associations

Causal associations were performed using the statistical procedure described previously (Schadt *et al*, 2005). Here, instead of using gene expression data and a phenotypic end point, we applied the described causal procedure to classify each miRNA–mRNA trait pair into one of three models (with respect to the miRNA as the reference): (a) eQTL for miRNA expression leads to changes in mRNA expression (miRNA targets); (b) eQTL for mRNA levels leads to changes in miRNA expression (miRNA regulators); and (c) eQTL independently drive miRNA and mRNA levels (independent). In addition, extensive simulations were performed in order to characterize the effect of performing the causal association procedure on data with varying precision and noise levels.

## Hexamer/heptamer enrichment analysis

Gene sets corresponding to the specific hexamer or heptamer seed regions were created. Each gene set represented a collection of transcripts containing at least hexamer (or heptamer) sequence in the 3' UTR of the mRNA transcript. Set enrichment was restricted to transcripts present on the microarray. A hypergeometric distribution was used to compute the significance of the enrichment.

## GO Biological Process and KEGG pathway annotations

We annotated each miRNA signature gene set identified using correlation or causal association methods using gene sets available in GO Biological Process and KEGG pathways. Here, we looked for significant overlap between our derived set and the annotated sets in GO Biological Process and KEGG pathways using a Fisher's exact test. A Bonferroni correction using the number of sets tested in each database was applied. After correcting for multiple hypotheses testing, we focused on sets with greater than 12 overlapping genes and expected counts of greater than 5.

## Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website ([www.nature.com/msb](http://www.nature.com/msb)).

## Acknowledgements

We thank The Jackson Laboratory, *In Vivo* Services, for their expert handling of the animals used in this study and Rosetta Gene Expression Laboratory for the execution of the sample preparation and hybridization experiments. We also thank Jason Egin for developing the computational tools that enabled this work; Xa Schildwachter for database assistance; and Michele Cleary and John Lamb for their insightful comments on earlier versions of this manuscript.

**Funding and author contributions:** WS was supported in part by a fellowship from the Merck Research Laboratories. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. WS and EES designed the statistical analysis, performed the experiments, analyzed the data, and wrote the paper. RK coordinated and managed the data collection of samples and microarray experiments. EES conceived the idea and supervised the project. All authors read and approved the final version of the paper.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233
- Basten CA, Weir BS, Zeng ZB (1999) *QTL Cartographer User's Manual*. Raleigh, NC: Department of Statistics, North Carolina State University
- Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**: 766–770
- Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**: 21–24
- Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**: 25–36



- Carleton M, Cleary MA, Linsley PS (2007) MicroRNAs and cell cycle regulation. *Cell Cycle* **6**: 2127–2132
- Chang S, Johnston Jr RJ, Frokjaer-Jensen C, Lockery S, Hobert O (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**: 785–789
- Chen CZ, Li L, Lodish HF, Bartel DP (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83–86
- Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA *et al* (2008) Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**: 429–435
- Doss S, Schadt EE, Drake TA, Lusis AJ (2005) Cis-acting expression quantitative trait loci in mice. *Genome Res* **15**: 681–691
- Esquela-Kerscher A, Slack FJ (2006) Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* **6**: 259–269
- Gingeras TR (2009) Implications of chimaeric non-co-linear transcripts. *Nature* **461**: 206–211
- Griffiths-Jones S (2004) The microRNA registry. *Nucleic Acids Res* **32**: D109–D111
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140–D144
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154–D158
- Herranz H, Cohen SM (2010) MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev* **24**: 1339–1344
- Hobert O (2008) Gene regulation by transcription factors and microRNAs. *Science* **319**: 1785–1786
- Hornstein E, Shomron N (2006) Canalization of development by microRNAs. *Nat Genet* **38**(Suppl): S20–S24
- Johnston RJ, Hobert O (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845–849
- Kai ZS, Pasquinelli AE (2010) MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nat Struct Mol Biol* **17**: 5–10
- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843–854
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20
- Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769–773
- Lincoln SE, Daly MJ, Lander ES (1993) *MAPMAKER/QTL User's Manual*. Cambridge, MA: Whitehead Institute of Biomedical Research
- Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, Johnson JM, Cummins JM, Raymond CK, Dai H, Chau N, Cleary M, Jackson AL, Carleton M, Lim L (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27**: 2240–2252
- Liu T, Papagiannakopoulos T, Puskas K, Qi S, Santiago F, Clay W, Lao K, Lee Y, Nelson SF, Kornblum HI, Doyle F, Petzold L, Shraiman B, Kosik KS (2007) Detection of a microRNA signal in an *in vivo* expression set of mRNAs. *PLoS One* **2**: e804
- Mallory AC, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants. *Nat Genet* **38**(Suppl): S31–S36
- Martin G, Schouest K, Kovvuru P, Spillane C (2007) Prediction and validation of microRNA targets in animal genomes. *J Biosci* **32**: 1049–1052
- Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJ (2008) A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**: 2535–2549
- Maziere P, Enright AJ (2007) Prediction of microRNA targets. *Drug Discov Today* **12**: 452–458
- Mehrabian M, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, Suh M, Armour C, Edwards S, Lamb J, Lusis AJ, Schadt EE (2005) Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat Genet* **37**: 1224–1233
- Nunez-Iglesias J, Liu CC, Morgan TE, Finch CE, Zhou XJ (2010) Joint genome-wide profiling of miRNA and mRNA expression in Alzheimer's disease cortex reveals altered miRNA regulation. *PLoS One* **5**: e8898
- Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. *Nature* **425**: 257–263
- Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**: 1033–1038
- Raymond CK, Roberts BS, Garrett-Engle P, Lim LP, Johnson JM (2005) Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA* **11**: 1737–1744
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–906
- Saetrom P, Heale BS, Srove Jr O, Aagaard L, Alluin J, Rossi JJ (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res* **35**: 2333–2342
- Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A *et al* (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* **37**: 710–717
- Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302
- Sethupathy P, Megraw M, Hatzigeorgiou AG (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat Methods* **3**: 881–886
- Taganov KD, Boldin MP, Baltimore D (2007) MicroRNAs and immunity: tiny players in a big field. *Immunity* **26**: 133–137
- Tsang J, Zhu J, van Oudenaarden A (2007) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell* **26**: 753–767
- Wilfred BR, Wang WX, Nelson PT (2007) Energizing miRNA research: a review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. *Mol Genet Metab* **91**: 209–217
- Xu P, Vernooij SY, Guo M, Hay BA (2003) The *Drosophila* microRNA *Mir-14* suppresses cell death and is required for normal fat metabolism. *Curr Biol* **13**: 790–795
- Yang X, Deignan JL, Qi H, Zhu J, Qian S, Torosyan G, Zhong J, Majid S, Falkard B, Kleinhans RR, Karlsson J, Castellani LW, Mumick S, Wang K, Xie T, Coon M, Zhang C, Estrada-Smith D, Farber CR, Wang SS *et al* (2009) Validation of candidate causal genes for abdominal obesity which affect shared metabolic pathways and networks. *Nat Genet* **41**: 415–423
- Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M, Wu K, Whittle J, Ju X, Hyslop T, McCue P, Pestell RG (2008) A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J Cell Biol* **182**: 509–517
- Zhao Y, Srivastava D (2007) A developmental view of microRNA function. *Trends Biochem Sci* **32**: 189–197



Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group. This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.