Published online 24 June 2010

Nucleic Acids Research, 2010, Vol. 38, No. 15 ***e160***
doi:10.1093/nar/gkq570

Towards computational prediction of microRNA function and activity

Igor Ulitsky[1,*], Louise C. Laurent[2,3] and Ron Shamir[1]

[1]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel, [2]Center for Regenerative Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037 and [3]Department of Reproductive Medicine, University of California, San Diego, 200 West Arbor Drive, San Diego, California 92103, USA

Received January 25, 2010; Revised June 3, 2010; Accepted June 7, 2010

ABSTRACT

While it has been established that microRNAs (miRNAs) play key roles throughout development and are dysregulated in many human pathologies, the specific processes and pathways regulated by individual miRNAs are mostly unknown. Here, we use computational target predictions in order to automatically infer the processes affected by human miRNAs. Our approach improves upon standard statistical tools by addressing specific characteristics of miRNA regulation. Our analysis is based on a novel compendium of experimentally verified miRNA-pathway and miRNA-process associations that we constructed, which can be a useful resource by itself. Our method also predicts novel miRNA-regulated pathways, refines the annotation of miRNAs for which only crude functions are known, and assigns differential functions to miRNAs with closely related sequences. Applying our approach to groups of co-expressed genes allows us to identify miRNAs and genomic miRNA clusters with functional importance in specific stages of early human development. A full list of the predicted mRNA functions is available at http://acgt.cs.tau.ac.il/fame/.

INTRODUCTION

MicroRNAs (miRNAs) are small (19–25 nt), non-coding RNAs that can reduce the abundance and translational efficiency of mRNAs, and play a major role in regulatory networks, influencing diverse biological phenomena (1). In metazoans, this repression is generally conferred by the binding of miRNAs to the 3′ UTRs of their targets. Some miRNAs were shown to repress translation of anywhere from tens to hundreds of mRNAs (2,3). Several miRNAs have been shown to affect multiple members of the same pathway (4–7). Determining the role of individual miRNAs in cellular regulatory processes poses a major challenge. The function of the vast majority of miRNAs is currently unknown, and even for relatively well studied miRNAs, only a handful of targets have been rigorously characterized. Knock-out studies in model organisms have had only limited success in delineating miRNA function, possibly because redundant miRNAs exist at saturating levels in wild-type cells, or because of compensatory effects in downstream signaling pathways (8).

Analyzing properties of miRNA targets is a promising approach to predicting miRNA function. A large number of algorithms for sequence-based prediction of miRNA targets have been described in the literature [reviewed in (9)]. As the number of validated targets is currently limited, methods for target-based inference of miRNA function must rely on these predictions. If the targets of a specific miRNA are enriched with genes annotated with some biological process or pathway, it is reasonable to infer that the miRNA is involved in the same process. This suggests the following simple algorithm for genome-wide inference of miRNA function: Predict that a miRNA is involved in every process/pathway for which the number of miRNA targets taking part in the process is statistically significant. Several studies used this approach. Gaidatzis *et al.* (10) applied a log-likelihood test to look for enrichment or depletion of targets of specific miRNAs in KEGG pathways. Similar algorithms using Gene ontology (GO), KEGG and BioCarta pathways were implemented in miRgator (11) and SigTerms (12), both of which evaluate statistical significance using a hypergeometric (HG) test.

*To whom correspondence should be addressed. Tel: +1 617 258 8346; Fax: +1 617 258 6768; Email: ulitsky@wi.mit.edu
Correspondence may also be addressed to Ron Shamir. Tel: +972 3 640 5383; Fax: +972 3 640 5384; Email: rshamir@tau.ac.il
Present address:
Igor Ulitsky, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA, USA 02142.

The previously described methods for target-based miRNA function prediction have three limitations. First, none of them was systematically tested for its ability to recover known miRNA functions. Second, they treat equally all predicted miRNA targets. Several recent studies have shown that, in fact, numerous factors, such as the local context within the 3′ UTR and the relative distance from the stop codon, influence the efficacy of individual miRNA target sites (13,14). These studies also offered 'context scores' for ranking the predicted targets of each miRNA. Third, existing methods do not take into account the very uneven distribution of 3′ UTR lengths; for example, the 3′ UTRs of genes expressed in brain and neural systems are almost twice as long as those of other genes (Figure 1B), and proliferating cells express genes with relatively short 3′ UTRs (15). Accordingly, genes highly expressed in the neural lineage harbor more predicted miRNA target sites (Figure 1B). It is not surprising, therefore, that the pathway most commonly predicted by Gaidatzis *et al.* (10) was 'axon guidance'; and that many seemingly unrelated miRNA families, such as mir-17, known to be primarily involved in cell-cycle regulation (6), and myeloid cell differentiation (16), were predicted by them to be related to neuronal pathways.

miRNAs frequently appear as co-localized clusters that are also co-expressed, and even transcribed as a single polycistron. Therefore, it is likely that co-localized miRNAs share similar functions, which may be revealed by a joint analysis of their targets. Xu and Wong (17) used the HG test followed by random resampling to look for over-representation of miRNA cluster targets in BioCarta pathways. Their analysis identified the mouse miR-183–96–182 cluster as a regulator of the insulin-signaling pathway. However, their method did not take into account how many of the miRNAs in the cluster regulate each pathway member.

Here we introduce FAME (functional assignment of miRNAs via enrichment), a new permutation-based statistical method that tests for over- or under-representation of miRNA targets in a designated set of target genes (Figure 1A). Unlike previous studies, FAME utilizes weights (confidence values) for miRNA-target pairs, accounts for the number of miRNAs regulating each target, and can be used for analysis of any group of miRNAs. Here, we focus on three main applications of FAME: direct inference of miRNA function using sets of genes sharing a common annotation, indirect inference of miRNA function using matched miRNA/mRNA expression data, and prediction of function for genomic clusters of miRNAs.

In order to compare our method to other methods and to test its ability to recover known miRNA functions, we assembled a compendium of 83 experimentally validated miRNA-function associations. We show that our method is superior to the currently used HG test. We describe the novel functions suggested by FAME for several miRNA

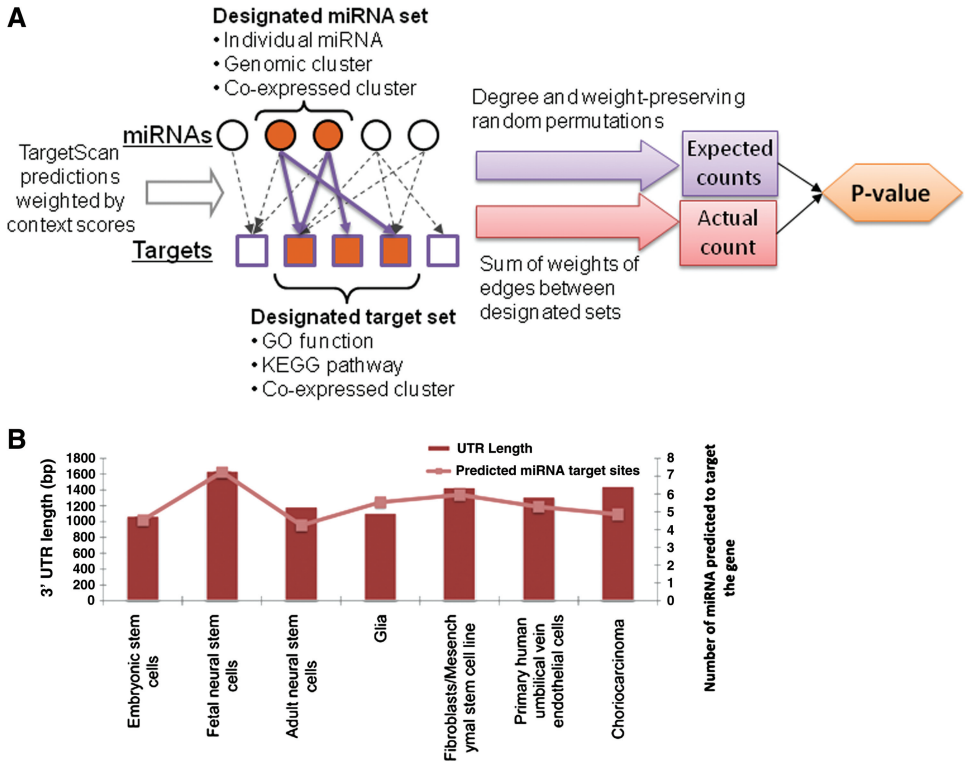

Figure 1. (**A**) FAME outline. A bipartite graph is constructed with edges corresponding to miRNA–target predicted pairs and edge weights determined by TargetScan context scores. Degree-preserving graph randomization is used to evaluate the significance of the total weight of the edges connecting a designated set of miRNAs to a designated set of targets, by computing an empirical *P*-value. (**B**) UTR lengths and the average number of miRNA regulators for genes highly expressed in various stem cell-related cell lines. See methods for the description of the mRNA dataset. The numbers are averages over the 200 genes with the highest average expression levels in each group of cell lines.

families, and show how it can be used to refine miRNA function in cases where only a relatively general miRNA function is known. We focus in particular on two families with similar seed sequences, mir-17 and mir-106/302, and demonstrate that they are likely to have both shared and unique functions.

Analysis of enrichment or depletion of miRNA targets in a set of co-expressed genes is an indirect yet potent way of providing clues to miRNA function. Previous studies used it to identify significant impact of miRNAs on tissue-specific gene expression patterns (18). Motif finding in 3′ UTR sequences has also been used for this task (19). We use our method to identify 68 miRNA families and 27 genomic clusters regulating 21 gene co-expression clusters in diverse human stem cell lines. Clusters enriched with the targets of a specific miRNA tend to be anti-correlated with the miRNA expression, whereas clusters depleted of miRNA targets are co-expressed with it. Finally, we use FAME's results to predict novel miRNA functions related to stem cell biology. We hypothesize that two miRNAs of unknown function, mir-499 and mir-544, play a pivotal role in early development.

An implementation of FAME with a graphical user interface is available as part of the Expander 5.0 microarray data analysis suite (60) (http://acgt.cs.tau.ac.il/expander). This implementation supports analysis of over- and under-representation of miRNA targets in gene sets from human, mouse, fly and worm. In addition, a full list of the GO 'biological process' terms and KEGG pathways predicted to be targeted by each human miRNA appear in http://acgt.cs.tau.ac.il/fame/. This website also allows the display of the targeted genes as part of KEGG pathway maps.

MATERIALS AND METHODS

miRNA target predictions and 3′ UTR sequences

Human miRNA target predictions for conserved miRNA families were taken from the TargetScan 5.0 database (21). Following the suggestion in (20), conserved target sites were used when testing for enrichment of miRNA targets, and both conserved and non-conserved predicted target sites were used when testing for depletion. Lengths of 3′ UTR sequences were taken from the TargetScan website. Spurious enrichments were avoided by filtering out targets with similar 3′ UTRs. If two 3′ UTRs shared an identical subsequence of at least 75 nt, and >80% of the miRNAs predicted to regulate them, only the gene with the longer 3′ UTR was retained in the target set. A total of 101 genes were filtered out.

Construction and randomization of bipartite graphs

The bipartite graph $G = (M,T,E,W)$ was constructed as follows. The context scores for all miRNA-target sites in the database were ranked and normalized to the range (0,1). For each miRNA m and target t, for each target site reported in TargetScan, an edge (m,t) was added to E with a weight of $1+(\lfloor k \cdot a \rfloor \cdot b)$, where k is the relative rank of the highest ranking conserved target site of m in t, and a and b are two parameters. Using this weighting scheme, the edge weights attain a discrete values $w_1,\ldots,w_a$, and the parameter b controls the relative contribution of the context scores to the enrichment/depletion significance. We used $a = 5$ and $b = 3$ throughout this study, but got similar results when these parameters were altered (Supplementary Figure S3). In particular these tests showed that using context-score weights ($b > 1$) improved the performance of FAME (Supplementary Figure S3).

The random graphs are generated by performing, for each possible edge weight w_i, a long sequence of independent edge shuffle operations (22), which preserves the number of edges with weight w_i incident on each node. Throughout the study, the total number of random edge shuffles for each random graph was set to $5 \cdot |E|$.

GO annotations

Human GO annotations were taken from the Entrez Gene database. GO annotations from the 'biological process' namespace with between 10 and 2000 predicted miRNA targets were used. In order to remove redundancy, we filtered out terms that differed by less than four genes, retaining in the dataset only the GO set that was assigned to more genes. Applying these filters resulted in 1499 GO sets.

miRNA and mRNA expression data

The miRNA expression data are described in (23). A total of 705 miRNAs were profiled using the Illumina human miRNAs version 1 microarray. A total of 21 060 mRNAs were profiled using Illumina human WG-6 version 1 microarrays. Briefly, the data compare gene expression in 26 cell lines representing 16 cell types, including five embryonic stem cell (ESC) lines, five fetal neural stem cell (fNSC) lines, four adult surgery derived neural stem (aNSC) lines, one extraembryonic endoderm-like (XE) cell line differentiated from the WA09 hESC line, two glial cell lines, three fibroblast cell lines, two mesenchymal stem cell (MSC) lines, two umbilical cord vein (UCV) endothelial cell lines and two choriocarcinoma cell lines. For each cell line, the same RNA preparations were used to generate the mRNA and miRNA expression data. miRNAs were profiled in quadruplicate, and mRNAs were profiled in duplicate. All replicates were averaged prior to subsequent analysis.

Identifying co-expression clusters

We extracted the 4000 genes with the highest variance among the genes that had at least two samples with at least 3-fold difference over the minimal level across the profiles. These genes were clustered using CLICK (24), which resulted in 21 clusters. Assignment of genes to clusters is presented in Supplementary File S4.

Identifying genomic clusters of miRNAs

Following (23,25), we defined a genomic cluster of miRNAs as a maximal segment such that every two consecutive miRNAs were separated by <50 000 bp. Genomic

positions of miRNAs were taken from MiRBase (26). We considered only clusters that contained representatives from at least two different TargetScan families. Finally, we united any pair of clusters that contained exactly the same set of TargetScan families. This resulted in 27 clusters containing from 2 to 27 distinct TargetScan families, with an average of 3.3 families per cluster (Supplementary Table S1).

RESULTS

A novel framework for detection of enrichment or depletion of miRNA targets

Our goal was to compute the significance of the overlap between a given set of predicted miRNA targets and a designated set of genes. Ideally, the computation should account for the strength of each predicted miRNA-target pair (or for our confidence in its biological relevance), and for the number of miRNAs regulating each gene (Figure 1A). FAME constructs a weighted directed bipartite graph $G = (M,T,E,W)$ in which miRNAs (M) are connected with their predicted targets (T). An edge (m,t) appears in E for every target site for m that appears in the 3′ UTR of t (hence, parallel edges between the same pair are possible). We used TargetScan 5.0 for prediction of miRNA targets, as it was recently shown to be superior to other target predictors (2). miRNAs that belong to the same TargetScan family (a set of miRNAs sharing the same seed sequence) are grouped together into a single node in M. T contains a node for each gene (represented by an Entrez Gene entry) that is predicted to be targeted by at least one miRNA. Edges in E are assigned discrete edge weights based on the TargetScan context scores ('Materials and Methods' section). Spurious enrichments were avoided by excluding from T genes with similar 3′ UTR sequences ('Materials and Methods' section). As proposed in (18), we used only evolutionarily conserved miRNA target sites when testing for over-representation, and both conserved and non-conserved sites when testing for under-representation.

Following the construction of G, we used degree-preserving permutations to generate N random graphs $G_1,\ldots,G_N$, in which, for each possible edge weight w, the number of outgoing edges with weight w from each miRNA and the number of incoming edges with weight w for each target were the same as in G ('Materials and Methods' section). We used $N = 10\,000$ throughout this study. These graphs were used for evaluating the significance of the overlap between the targets of a set M' of miRNAs (in this study, a TargetScan family or a set of families represented in a genomic miRNA cluster) and a set of targets T' (e.g.: a set of genes sharing a GO annotation). Let $W_G(M',T')$ be the total weight of the edges connecting M' and T' in G. We compared $W_G(M',T')$ with all $W_{Gi}(M',T')$, and computed an empirical P-value and a z-score for (M',T'). All the (M',T') pairs were then ranked by their P-values, and FDR was assessed by the Benjamini–Hochberg procedure (27).

A compendium of validated miRNA targets

Rigorous evaluation of any prediction algorithm requires a 'gold standard': in our case a set of miRNAs with known functions. As we know of no available resource describing validated miRNA functions, we carried out an extensive literature survey and constructed a compendium of miRNAs with experimentally established functions in mammals. We included in the compendium only cases in which at least one target relevant to the pathway or function was experimentally validated (i.e. functions suggested based solely on phenotypes resulting from the perturbation of the miRNA were not included). In each case, we manually assigned the KEGG pathway and GO annotation that was closest to the reported function. The compendium, with references to the original publications, appears in Supplementary File S1. It contains a total of 31 miRNA–KEGG pathway associations and 52 miRNA–GO set associations.

Direct prediction of miRNA functions

Enrichment or depletion of miRNA targets in a set of genes involved in a specific process or pathway is the most direct clue to miRNA function prediction. We used two data sets to test this approach, one based on pathways taken from KEGG, and one based on GO. Each TargetScan miRNA family m was tested for over-representation of its targets in each KEGG pathway or GO annotation set that contained at least three targets of m.

We first describe the results on KEGG pathways. Using the compendium, we compared FAME with the HG test, and with the log-likelihood ratio (LLR) scores used by Gaidatzis *et al.* (10). For each miRNA m associated with a KEGG pathway P in the compendium, we ranked all 140 tested KEGG pathways according to the significance of their enrichment with the targets of m (Figure 2A). The success of each method in predicting a specific function was measured by the rank of P in this list. Eighteen compendium miRNA–pathway pairs met the criterion of at least three genes in P being predicted targets of m, and they were ranked by each of the three methods. In six cases the known pathway corresponded to the top FAME prediction, compared to just four cases when the HG test was applied, and three cases when the LLR test was used (Figure 2A). The average position of the known function across all the 18 pairs was higher for FAME than for the HG and LLR tests (Figure 2B), although the difference was not statistically significant, perhaps due to the small size of the compendium. Performance of the HG test was similar when only the top 25, 50 or 75% of the miRNA–target pairs (as determined by the context score) were used (Supplementary Figure S1A), and it never placed more than four correct pathways as top predictions (results not shown).

The top KEGG pathway predictions for each miRNA family are shown in Table 1. This analysis allowed us to predict novel functions for a number of miRNAs:

- mir-122 is a conserved liver-specific miRNA (28) important for normal metabolic function of the liver.

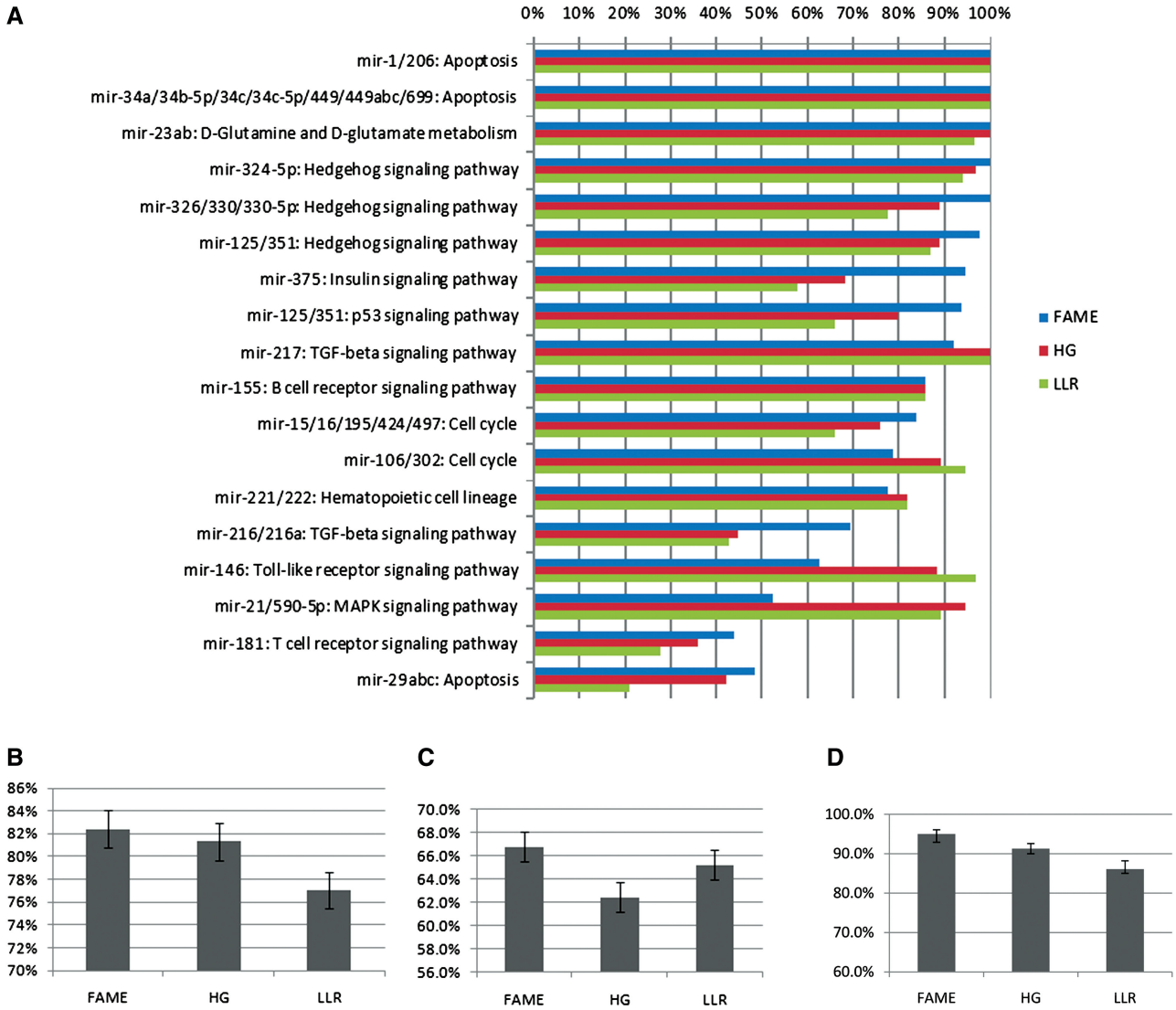

Figure 2. Comparison of methods for detection of enrichment of miRNA targets. (**A**) For each miRNA family, all the KEGG pathways were tested for enrichment of miRNA targets and ranked in increasing order of *P*-value. In case of ties, annotations were ranked in decreasing order of z-score. The chart shows the relative position of the compendium function in each list. (**B**) Average location of the known KEGG pathway in the ranked lists obtained by using FAME and the HG and LLR tests. Error bars represent one standard error. (**C**) Average location of the known GO 'biological process' annotation in the ranked lists of the three methods. (**D**) Same as C, but taking into account only annotations that were placed in the top 10% by at least one of the methods.

mir-122 inhibition in mice led to reduced cholesterol biosynthesis and stimulation of hepatic fatty-acid oxidation (29); and mir-122 ablation led to decreased plasma cholesterol levels in mice, suggesting that mir-122 could be an effective therapeutic target (30). However, no related targets or pathways affected by mir-122 have been characterized. FAME predicted that the 'glycolysis/gluconeogenesis' pathway is regulated by mir-122 ($P = 5.5 \times 10^{-4}$, FDR < 0.05). This suggests that the mir-122 regulation of cholesterol biosynthesis is mediated by direct regulation of the carbohydrate metabolism enzymes PKM2, G6PC and ALDOA, which are predicted targets of mir-122 involved in glycolysis and gluconeogenesis.

- The well-studied mir-21 family is known to regulate MAPK signaling through SPRY1 (31). FAME predicted that mir-21 regulates three pathways: 'cytokine-cytokine receptor interaction', 'Jak-STAT signaling' and 'MAPK signaling' (FDR < 0.05, listed in decreasing order of statistical significance). Our top prediction thus implicates mir-21 in cytokine signaling. Consistent with this hypothesis, the expression levels of mir-21 are upregulated following treatment with LPS, which induces inflammation (32).

Using GO to predict miRNA-regulated processes

While the results with KEGG were promising, the specificity of KEGG pathways is rather limited, and some biological processes, such as development, are poorly represented. A much more comprehensive repository of gene sets in human is GO (http://www.geneontology

Table 1. KEGG pathways predicted by FAME to be regulated by miRNAs

miRNA	KEGG pathway	Number of targets	*P*-value	Weight enrichment factor
let-7/98	Aminoacyl-tRNA biosynthesis	3	1.3×10^{-3}	9.73
mir-1/206	SNARE interactions in vesicular transport	6	6.5×10^{-4}	3.64
mir-103/107	Hedgehog-signaling pathway	8	2.0×10^{-4}	3.81
mir-122	Glycolysis / gluconeogenesis	3	5.0×10^{-4}	17.45
mir-124/506	Metabolic pathways	70	1.0×10^{-4}	1.54
mir-125/351	Tyrosine metabolism	3	8.0×10^{-4}	8.13
mir-125a-3p	Cytokine–cytokine receptor interaction	4	6.3×10^{-3}	3.44
mir-129/129-5p	Cardiac muscle contraction	6	1.0×10^{-4}	8.89
mir-132/212	TGF-beta-signaling pathway	10	5.5×10^{-4}	3.10
mir-138	Axon guidance	13	8.5×10^{-4}	2.34
mir-139-5p	Purine metabolism	4	3.9×10^{-3}	3.60
mir-140/140-5p/876-3p	Notch-signaling pathway	4	4.3×10^{-3}	4.49
mir-142-3p	Regulation of actin cytoskeleton	12	2.5×10^{-4}	2.61
mir-143	Natural killer cell mediated cytotoxicity	4	4.1×10^{-3}	3.38
mir-145	Axon guidance	17	1.0×10^{-4}	2.55
mir-146	Toll-like receptor-signaling pathway	3	1.0×10^{-4}	10.47
mir-148/152	Basal transcription factors	4	4.5×10^{-3}	4.41
mir-15/16/195/424/497	Cell cycle	17	1.0×10^{-4}	2.19
mir-150	Wnt-signaling pathway	7	6.9×10^{-3}	2.80
mir-155	T cell receptor-signaling pathway	10	1.0×10^{-3}	3.17
mir-185/882	GnRH-signaling pathway	4	8.6×10^{-3}	3.28
mir-190	Cell adhesion molecules (CAMs)	5	3.1×10^{-3}	5.03
mir-194	TGF-beta-signaling pathway	9	4.3×10^{-3}	2.67
mir-202/202-3p	ECM-receptor interaction	11	1.3×10^{-3}	2.69
mir-203	Insulin-signaling pathway	14	3.0×10^{-3}	1.98
mir-205	PPAR-signaling pathway	4	3.3×10^{-3}	4.84
mir-208/208ab	Wnt-signaling pathway	6	7.3×10^{-3}	3.02
mir-21/590-5p	Cytokine-cytokine receptor interaction	11	1.0×10^{-4}	4.33
mir-217	Gap junction	4	8.6×10^{-3}	2.98
mir-218	Heparan sulfate biosynthesis	6	1.0×10^{-4}	4.82
mir-219/219-5p	Ether lipid metabolism	3	4.3×10^{-3}	7.21
mir-23ab	Glycosphingolipid biosynthesis—lacto and neolacto series	6	2.0×10^{-4}	4.48
mir-24	Alanine and aspartate metabolism	3	1.4×10^{-3}	8.17
mir-27ab	Neuroactive ligand-receptor interaction	16	1.2×10^{-3}	2.09
mir-28/28-5p/708	Jak-STAT-signaling pathway	4	2.4×10^{-3}	3.25
mir-299/299-3p	Focal adhesion	3	5.9×10^{-2}	2.66
mir-29abc	ECM-receptor interaction	21	1.0×10^{-4}	6.13
mir-324-5p	TGF-beta-signaling pathway	4	2.5×10^{-2}	3.18
mir-326/330/330-5p	Arachidonic acid metabolism	3	2.5×10^{-4}	17.91
mir-33/33ab	Antigen processing and presentation	3	7.7×10^{-3}	5.98
mir-339-5p	ErbB-signaling pathway	3	3.2×10^{-2}	3.03
mir-346	Wnt-signaling pathway	5	1.0×10^{-2}	3.32
mir-34a/34b-5p/34c/34c-5p/449/449abc/699	*N*-Glycan biosynthesis	4	1.3×10^{-3}	4.62
mir-361/361-5p	Nucleotide excision repair	3	1.0×10^{-4}	23.84
mir-365	Apoptosis	4	2.2×10^{-3}	5.04
mir-374/374ab	Retinol metabolism	4	1.0×10^{-4}	10.75
mir-375	Purine metabolism	4	9.3×10^{-3}	4.10
mir-376/376ab/376b-3p	Neuroactive ligand-receptor interaction	4	5.8×10^{-3}	3.66
mir-377	Ubiquitin mediated proteolysis	9	3.1×10^{-3}	2.44
mir-378/422a	Hedgehog-signaling pathway	4	2.0×10^{-3}	7.22
mir-379	Adherens junction	3	2.9×10^{-2}	3.92
mir-384/384-3p	Lysine degradation	4	5.5×10^{-4}	8.32
mir-410	Heparan sulfate biosynthesis	4	1.6×10^{-3}	4.29
mir-411	Ubiquitin mediated proteolysis	3	3.6×10^{-2}	3.43
mir-431	Adherens junction	3	8.5×10^{-2}	2.37
mir-433	Cell cycle	7	6.0×10^{-4}	3.93
mir-485/485-5p	Metabolic pathways	17	3.5×10^{-4}	2.49
mir-486/486-5p	Focal adhesion	6	1.8×10^{-3}	3.28
mir-490/490-3p	Adipocytokine-signaling pathway	3	1.9×10^{-2}	4.33
mir-496	CAMs	3	7.3×10^{-3}	5.05
mir-503	p53-signaling pathway	6	1.0×10^{-4}	5.85
mir-543	Circadian rhythm—mammal	4	8.0×10^{-4}	5.33
mir-592/599	mTOR-signaling pathway	3	7.5×10^{-3}	4.02
mir-7/7ab	Purine metabolism	5	1.8×10^{-3}	3.55
mir-758	Toll-like receptor-signaling pathway	3	5.7×10^{-3}	5.31
mir-874	Calcium-signaling pathway	5	1.0×10^{-2}	2.99

(continued)

Table 1. Continued

miRNA	KEGG pathway	Number of targets	*P*-value	Weight enrichment factor
mir-875-5p	Cytokine–cytokine receptor interaction	3	3.7×10^{-2}	3.32
mir-96/1271	Glycosphingolipid biosynthesis—ganglio series	4	5.0×10^{-4}	5.86
mir-99ab/100	Melanogenesis	3	1.1×10^{-3}	8.65

Only the top prediction for each miRNA family and with FDR < 0.1 are shown. 'Weight enrichment factor' is the ratio between the total weight of the edges between the miRNA and the pathway genes in the bipartite graph G, and the average weight of such edges in 10 000 random graphs.

.org). Since sets of genes sharing a GO annotation (henceforth referred to as *GO sets*) frequently overlap, and can be very general or very specific, we focused on 1499 non-redundant GO sets, containing between 10 and 2000 genes (Supplementary File S2, see 'Supplementary Methods' for details). The predictions of GO annotations for miRNA families appear in Supplementary File S3. For 36 compendium miRNA–GO set pairs, the GO set contained at least three predicted miRNA targets, and these pairs where used further (Supplementary File S1). The average ranking of the known miRNA–GO set pairs was higher when using FAME than when using the HG or the LLR tests (Figure 2C, Supplementary Figure S2). When we considered only pairs for which the known function was ranked in the top 10%: FAME significantly outperformed the HG test and the LLR test ($P = 0.015$ and 0.002 respectively, Figure 2D). Once again, performance of the HG test was not altered by using only the top 25, 50 or 75% of the predictions (Supplementary Figure S1B–C).

Since our compendium consisted of relatively broad and non-specific functions, representing the current limited knowledge of miRNA functions, its precision for the evaluation of the performance of FAME is limited. It is possible that related, but more specific, functional terms that correspond to the real function of the miRNA were ranked higher than the compendium functions. Indeed, manual inspection of the results suggested several such cases (Table 2). For example, mir-146 was shown to be involved in the innate immune response (33,34), and therefore labeled 'immune response' in our compendium. However, only two genes annotated in GO with 'immune response' are predicted by TargetScan to be regulated by mir-146. One of the top FAME predictions for mir-146 is 'I-κB kinase/NF-κB cascade', a pathway that contains three mir-146 targets (CARD10,IRAK1 and TRAF6). Indeed, mir-146 was shown to affect the activity of the NF-κB pathway (35). Interestingly, the expression of mir-146 was also shown to be regulated by NF-κB (34), suggesting that mir-146's function in immune response is regulated by a feedback loop. In another example, mir-205 was shown to regulate epithelial-to-mesenchymal transition (EMT) by targeting the transcription factors ZEB1 and ZEB2 (36). FAME predicts that mir-205 regulates 'establishment or maintenance of cell polarity' (ranked 12th). Loss of apical-basal polarity is one of the key steps in EMT (37). Notably, ZEB1 and ZEB2 are not annotated with this term in GO, despite the fact that several polarity-related genes, such as CRB3, PATJ and LGL2, are known ZEB1 targets (38). FAME prediction thus suggests that mir-205 regulates EMT mainly through regulation of apical-basal polarity genes, both by direct repression and via ZEB1 and ZEB2.

FAME analysis highlights the differences between miRNAs with similar seed sequences

miRNA genes tend to appear in multiple copies in the genome, and can be grouped into families sharing similar mature sequences. According to the currently accepted model, the 'seed sequence', nucleotides 2–8 of the mature miRNA sequence, is the main determinant of miRNA targeting specificity (2,21,39). Several miRNA families share similar, but not identical, seed sequences. We previously observed that at least 18 different miRNAs that have the AAGUGC hexamer in their seed sequence are highly expressed in ESCs, (23). These miRNAs belong to two TargetScan 5.0 families: mir-17-5p/20/93.mr/106/519.d (henceforth referred to as mir-17, seed sequence A<u>AAGUGC</u>), and mir-106/302 (seed sequence <u>AAGUGC</u>U). Since TargetScan predictions are based on the conservation of seed matches, and these seeds overlap, TargetScan predicted numerous common targets for both families. However, the question of whether these two groups have entirely identical functions has not been resolved. Several members of both families have been studied and some evidence exists on their function. Members of both the mir-17 and mir-302 families were found to regulate the G1/S cell-cycle checkpoint (6,40–45) and the TGFβ-signaling pathway (46). Deletion of mir-17 in mice led to inhibited B cell development (16), and mir-17 was shown to control monocytopoiesis by targeting RUNX1 (47). Recently, the mir-302 family was shown to regulate the mesendodermal cellular fate specification, and repression of this family in human ESCs inhibited the formation of neuroectoderm during embryogenesis (48).

FAME predictions for the two families appear in Supplementary File S3. FAME predicted that both families regulate cell-cycle progression: 'regulation of cell cycle' was enriched in both target sets ($P = 0.5.0 \times 10^{-4}$ for mir-17 and $P = 0.00465$ for mir-106/302). mir-17 targets were more significantly enriched for 'negative regulation of progression through cell cycle' ($P = 4.5 \times 10^{-4}$), concordant with the role mir-17 plays in accelerating progression through cell cycle (6). However, additional FAME predictions point to differences in the developmental functions of the two families. The targets of mir-17 but not mir-106/302 were enriched for 'regulation of myeloid

Table 2. Refinement of known miRNA functions

miRNA family	Known function (rank)	Proposed refined function (rank)
mir-146	Immune response (–)	I-κB kinase/NF-κB cascade (2)
miR-21/590-5p	Protein kinase cascade (178)	Negative regulation of MAP kinase activity (9)
mir-192/215	Regulation of cell cycle (15)	Regulation of progression through cell cycle (13)
mir-17-5p/20/93.mr/106/519.d	Regulation of cell cycle (2)	Negative regulation of progression through cell cycle (1)
mir-205	Epithelial cell differentiation (–)	Establishment or maintenance of cell polarity (12)
mir-141/200a	Epithelial cell differentiation (62)	Morphogenesis of embryonic epithelium (4)
mir-1/206	Glucose metabolism (40)	Glucose catabolic process (8)
mir-1/206	Regulation of apoptosis (52)	Anti-apoptosis (7)
mir-9	Neuron development (26)	Peripheral nervous system development (12)
mir-130/301	Angiogenesis (26)	Blood vessel morphogenesis (6)
mir-29abc	Regulation of apoptosis (377)	Apoptotic program (23)

Relative ranks of known non-specific miRNA functions and proposed specific functions. (–) in the 'Known function' column indicates that the GO set corresponding to that function contained less than three miRNA targets, and thus this function was not ranked.

leukocyte differentiation' ($P = 0.0799$ versus $P = 0.503$), with five mir-17 targets annotated with this GO term. In contrast, only the targets of mir-106/302 were enriched with 'central nervous system neuron differentiation' ($P = 0.011$ versus $P = 0.25$ for mir-17). These findings suggest that in addition to their common role in cell-cycle regulation, the two families have distinct roles in cell differentiation: mir-17 regulates development of leukocytes while mir-106/302 regulates development of the nervous system.

Using matched miRNA and mRNA expression for detection of miRNA regulation

Identifying over- or under-representation of miRNA targets in a set of co-expressed genes is an indirect but powerful method for studying miRNA function (18). Here, we utilized this approach to study a collection of ~130 simultaneous miRNA and mRNA expression profiles from cell lines designed to identify regulatory pathways critical for self-renewal, pluripotency and differentiation of human stem cells ('Materials and Methods' section) (49). We refer to this collection as the stem cell data set (SCD). SCD contained expression profiles of ESCs, fNSCs, aNSCs, glia cells, fibroblasts, MSCs, umbilical vein endothelial cells, and two choriocarcinoma cell lines. We clustered the mRNA expression patterns in SCD using CLICK (24) and obtained 21 clusters of co-expressed genes ('Materials and Methods' section, Supplementary File S4). We tested each cluster and each miRNA for enrichment and depletion of miRNA targets using FAME and the HG test.

According to the currently accepted model, miRNAs function as repressors of gene expression, and thus their expression patterns are expected to be anti-correlated with those of their targets (50,51) and correlated with their anti-targets [genes depleted of miRNA target sites (18)]. However, it is often difficult to identify such effects in matched miRNA/mRNA expression datasets. For example, in the SCD, the average Pearson correlation between expression of miRNAs and their TargetScan targets is 0.009. The uneven 3′ UTR lengths of genes highly expressed in different stem cell types (Figure 1B) could be one of the reasons for this observation. In order to test this, we analyzed the results of FAME and the HG test using the miRNA expression data in SCD. When we detected over-representation of miRNA targets in a cluster, we tested whether the miRNA expression pattern and the average expression pattern of the mRNA cluster were significantly anti-correlated (Figure 2C). Similarly, we tested cases of miRNA target depletion for a significant positive correlation. In cases where the miRNA family contained more than one miRNA with expression data in SCD, we chose as a representative the miRNA that had the highest absolute value of expression correlation with the cluster. We found that the most significant enrichments identified by FAME were consistently better supported by the miRNA expression data (Figure 3): FAME yielded evidence of a significant positive correlation of miRNAs and sets of genes depleted of their targets in 23% of the cases, and evidence of a negative correlation of miRNAs and their targets in 18% of the cases. These results suggest that miRNA target depletion is more effective than enrichment in identifying functionally relevant miRNAs using co-expression data. Indeed, as described below, for several miRNAs with a known function in specific differentiation-related processes, we found evidence of depletion of target sites in genes expressed during the same developmental stage, but no evidence of enrichment of target sites in genes expressed at other stages.

In addition, we evaluated the correlation between the enrichment of miRNA targets in a cluster and the similarity of the expression patterns of the miRNA and the cluster. To this end, we used the *P*-values computed using FAME and the HG test to assign every miRNA–cluster pair with a relative rank of their enrichment *P*-value (highest ranks were assigned to pairs in which the targets of the miRNA were most significantly enriched in the cluster). Using FAME, we found a strong negative correlation between the enrichment rank and the similarity of the gene expression patterns (evaluated using Pearson correlation, $r = -0.27$). The same correlation was significantly weaker when using the HG-test ($r = -0.065$, $P = 0.026$ for the difference between the two correlation coefficients). Similarly, the correlation between the significance of the depletion of miRNA targets and the similarity of gene expression profiles was significantly higher when using FAME compared to the

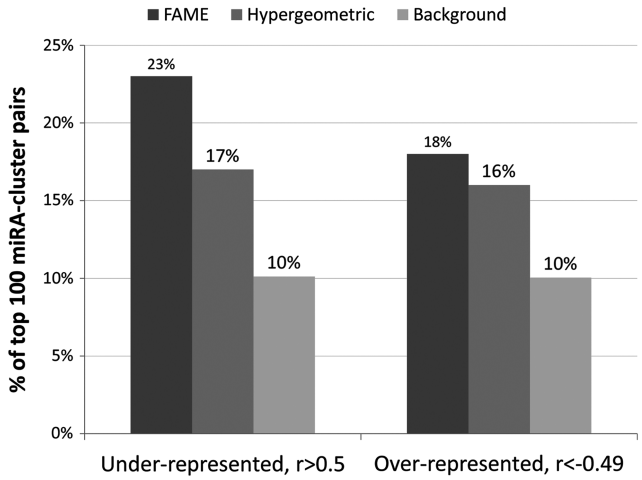

Figure 3. Performance of methods for enrichment detection on co-expression clusters. Out of the 1323 possible miRNA–cluster pairs, those with a correlation of $r > 0.5$ or $r < -0.5$ between the miRNA and the average mRNA expression were marked as 'high' (~10% for each direction). The plots show the fraction of the 100 most significant miRNA–cluster pairs found by FAME and the HG test that fell into the 'high' category.

HG test ($r = 0.037$ versus $r = -0.019$, $P = 0.0377$ for the difference between the two correlation coefficients).

Individual miRNAs and genomic clusters involved in stem cell biology

By combining weak signals coming from individual miRNAs in a genomic cluster, one can uncover the function of the whole cluster (17). To test this concept we identified genomic clusters of miRNAs in the human genome (Supplementary Table S1; 'Materials and Methods' section) and repeated the analysis of the 21 SCD co-expression clusters using the genomic miRNA clusters.

Overall, at FDR < 0.1, we identified enrichment or depletion of targets of 68 miRNA families and 27 genomic clusters in 21 co-expression clusters (Figure 4A). Of the 68 miRNA families, 16 are known to be related to stem cell biology [out of 25 stem cell-related families taken from (52), $P = 0.027$], indicating that our cluster-based analysis is capable of revealing functionally relevant miRNAs. In comparison at FDR < 0.1, the HG test reported significant enrichment or depletion for 77 miRNA families, but the overlap with the stem cell-related families was not significant ($P = 0.344$).

Analysis with FAME revealed several known miRNA regulations that are supported by miRNA expression data: miR-9 and miR-124 targets are enriched in cluster 2, which is downregulated in fNSCs (and also in ESCs), and miR-9 targets are depleted in cluster 3 (Figure 4B). Targets of miR-17 family were enriched in clusters 7 and 8, which show low expression in ESCs, and depleted in clusters 1 and 13, which are upregulated in ESCs (Figure 4C). Targets of mir-106/302, which share the AAGUGC hexamer with mir-17 (see below), were also enriched in clusters 7 and 8, albeit less significantly ($P = 0.0603$ and $P = 0.0056$, respectively, FDR < 0.1). Accordingly, miR-9 and miR-124 miRNAs were upregulated in fNSCs, and miR-17 and related miRNAs were strongly upregulated in ES cells [Figure 4B and C, (23)]. Interestingly, we identified more significant enrichment (compared to using individual miRNA families) with the four genomic clusters that contain members of the miR-17 and miR-106/320 families: gc:17-92a, gc:371-527, gc:106b-93 and gc:302a-367 (Figure 4C). These findings underscore the power of analyzing genomic clusters of miRNAs in addition to analyzing individual miRNAs.

miR-145 was recently shown to be an important regulator of differentiation by repressing the key ESC transcription factors OCT4, SOX2 and KLF4 (53). Increased miR-145 expression inhibited hESC self-renewal and induced lineage-restricted differentiation (53). In our data, the expression of miR-145 was strongly induced in differentiated cells, but also in two of the ESC lines (H1 and HSF6, Figure 4D), which may reflect their heterogeneity. FAME identified a significant depletion of miR-145 targets in cluster 8, which is upregulated in ESCs and in fNSCs ($P = 0.0091$). In addition, the targets of the gc:143-145 cluster, which contains miR-145 along with miR-143, were even more significantly depleted in cluster 8, which contains genes downregulated in ESCs (and therefore upregulated following ESC differentiation, $P = 3.0 \times 10^{-4}$). This suggests that mir-143 is also likely to be related to ESC differentiation. We found similar depletions for mir-499 and mir-544 families ($P = 0.0023$ and $P < 1.0 \times 10^{-4}$, respectively), both of which are also downregulated in ESCs (Figure 4D), suggesting that these families play an important role during ESC differentiation and early human development.

FAME identified significant depletion of the targets of let-7, mir-125 and genomic clusters containing these miRNAs, in cluster 3, which contains genes upregulated in fNSCs. Depletion of let-7 targets in brain-specific genes was reported earlier (18). Interestingly, members of the let-7 family were expressed at similar levels in the various non-ESC lines in our data set, but significant depletion of their targets was observed only in the cluster upregulated in fNSCs. We also identified significant enrichment of let-7 targets (but not of mir-125 targets) in cluster 5, which is upregulated in one of the choriocarcinoma lines (BEWO). Consistently, all members of the let-7 family were strongly downregulated in this cell line (Figure 4E). let-7 miRNAs are known tumor suppressors downregulated in various cancers (54). As members of the let-7 family appear in multiple genomic locations, our results suggest that their repression, either through coordinated events of transcriptional regulation or post-transcriptionally, leads to upregulation of their targets, which may contribute to malignancy in choriocarcinoma.

gc:134-758 is a large miRNA cluster of unknown function (55) located on chromosome 14, and is significantly downregulated in ESCs (23) (Figure 4F). We identified significant depletion of the targets of gc:134-758 in cluster 2, upregulated in differentiated cells, and in particular in aNCSs. The targets of two

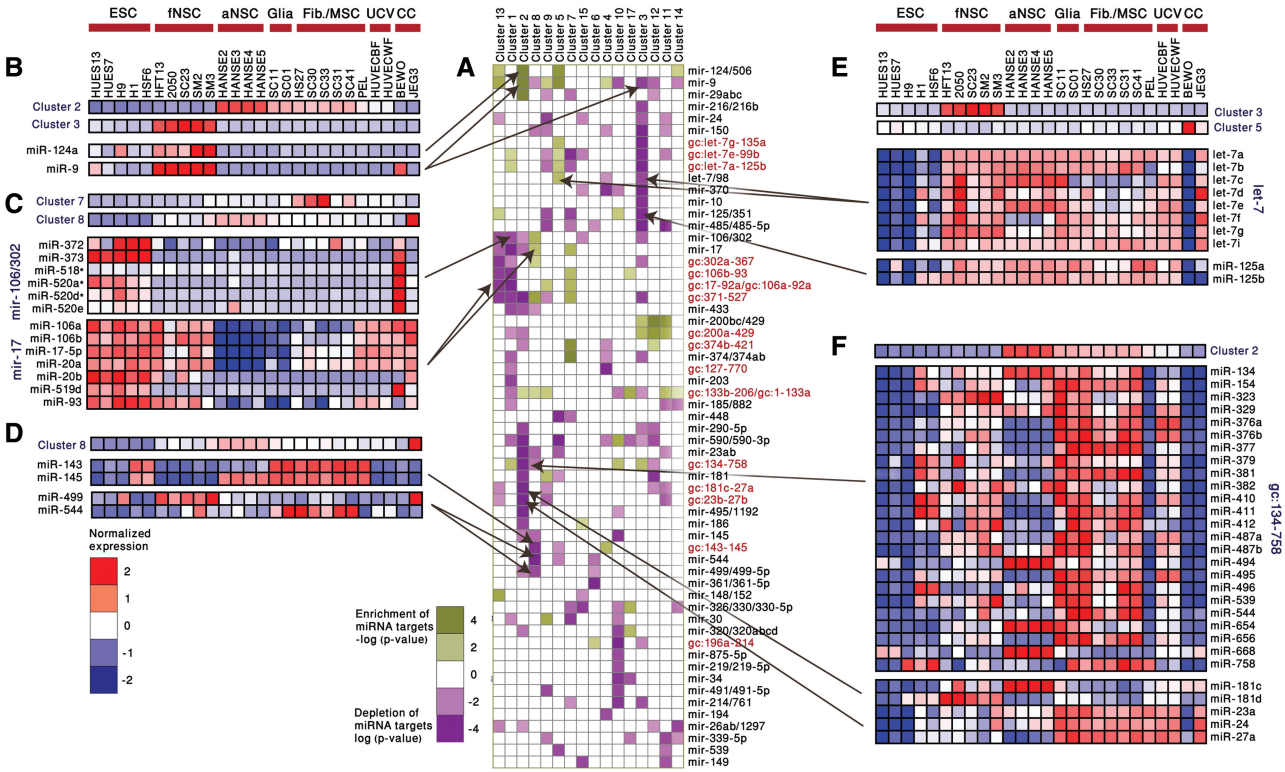

Figure 4. mRNA co-expression clusters and miRNA regulation in stem cell lines. (**A**) Enrichment and depletion of miRNA targets in co-expression clusters. Purple (green) squares indicate over- (under-) representation of miRNA targets in a cluster. Names of genomic clusters of miRNAs (Supplementary Table S1) are written in red. Only clusters with at least 30 genes that were enriched with targets of at least one miRNA with $P < 3 \times 10^{-3}$ and FDR < 0.1 are shown. Cluster-miRNA pairs with $P > 0.05$ are not shown (white squares). (**B–F**) Average expression levels of the mRNAs in co-expression clusters and of miRNAs in different families. The top rows in each subfigure show average mRNA expression of the co-expression clusters and the matrices below them show the expression of the miRNA families under the same conditions. The expression pattern of each miRNA and each mRNA were normalized to mean 0 and SD of 1. Fib., fibroblasts; CC, choriocarcinoma (placental cancer).

additional miRNA clusters, gc:181c-27a and gc:23b-27b, which were also downregulated in ESCs (Figure 4F), were also significantly depleted in this cluster. These results suggest that the main function of all three clusters takes place in differentiated cells and in aNSCs (rather than in fNSCs).

FAME reduces the 3′ UTR length bias

One of our goals in designing FAME was to overcome the bias introduced by the variability in 3′ UTR lengths. The average 3′ UTR lengths in the SCD varied greatly: the average 3′ UTR length of genes in cluster 2 (genes up regulated in fNSCs) was 2342, compared to just 1160 in cluster 2. In order to test whether FAME was able to alleviate this bias, we divided the 21 clusters into four bins based on average 3′ UTR length and compared the total number of significant enrichments (FDR < 0.1) in each bin (Figure 5). The number of significant enrichments found with the HG test was correlated with UTR length ($r = 0.59$); this correlation was significantly reduced with FAME ($r = 0.18$).

DISCUSSION

We have presented FAME, a novel method for detecting enrichment or depletion of miRNA targets in sets of genes. This method has two main applications at the present time: direct inference of miRNA functions using sets of co-annotated genes, and prediction of miRNA-based regulation using mRNA expression data. To allow rigorous evaluation of FAME in the first task, we assembled a compendium of 83 miRNA–pathway and miRNA–process pairs. To the best of our knowledge, this is the first time the performance of such a method has been rigorously validated against experimentally tested functions. While it is still quite modest, this compendium can also be useful for evaluating future approaches to miRNA function prediction, and will improve as experimental evidence on miRNA function accumulates. Our compendium complements an existing database that lists the involvement of miRNAs in human disease (56).

The use of functional annotations of predicted targets for inference of miRNA function is a promising concept, but algorithms for this problem must cope with numerous obstacles, including the limited accuracy of miRNA target prediction methods, biases in 3′UTR length and composition, limitations in the existing systems for functional annotations, and the fact that most miRNAs have only a limited effect on the expression levels of their targets (2,3). A wealth of methods have been developed for *cis*-regulatory motif finding, typically applied to

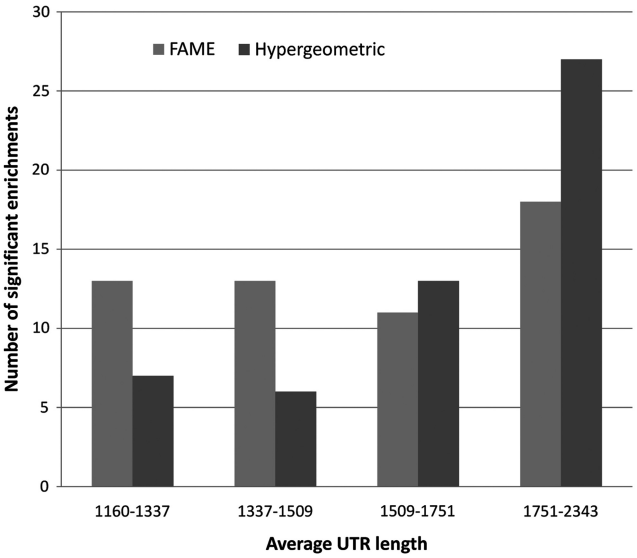

Figure 5. The 3′ UTR length bias. The 21 clusters in the SCD were divided into four bins, with five to six clusters in each bin, based on the average UTR length in each cluster. The total number of significant enrichments (FDR < 0.1) is shown for each bin.

promoter sequences of co-expressed genes (19,57,58), but more recently also to 3′ UTRs (19,59). Some of these methods address key issues affecting transcription factor and miRNA binding, such as GC-content and distance from the transcription start site. However, they do not address key in miRNA target analysis, such as the number of miRNAs targeting each 3′ UTR and the influence of the 3′ UTR context around the miRNA target site. The results described here suggest that our statistical analysis is capable of overcoming some of these difficulties as it correctly infers the miRNA function in many cases. Such analysis can be improved further in the future by directly addressing differences in the composition of the 3′ UTRs.

Prediction of relatively low-resolution functions (e.g. on the level of KEGG pathways) seems to be easier than prediction of the precise biological process affected by each miRNA (represented by a GO term). However, assessment of the success of FAME on the latter task is limited by the currently crude knowledge of miRNA functions. Our analysis is also limited by the quality of the target prediction algorithms. The most successful predictors use information about target site conservation, but they miss targets (and therefore functions) that are less well conserved. In addition, it is currently difficult to efficiently predict functional miRNA target sites in coding sequence, even though a considerable fraction of them is estimated to regulate gene expression (2,3).

Data describing miRNA and mRNA expression in the same samples are now collected in multiple systems. We suggest the following three-step strategy for analysis of such data: (i) clustering of mRNA expression, (ii) detection of over- and under-representation of miRNA targets in clusters using FAME and (iii) analysis of the correlation between the expression patterns of the miRNAs and of the mRNA clusters in which they are implicated. As we show, this analysis recovers a significant number of miRNAs with key roles in the studied system. If the expression data describe a comparison between two biological conditions, differential expression analysis (e.g. using *t*-test) can substitute for the clustering of the mRNA patterns. Both types of analysis are made possible by our implementation of FAME as part of the Expander 5.0 microarray data analysis suite (http://acgt.cs.tau.ac.il/expander/). Using Expander, which contains pre-compiled TargetScan 5.0 target predictions, it is possible to load expression data, identify co-expressed or differentially expressed genes, and use FAME to detect over- or under-representation of miRNA targets [see the guidelines for using FAME in Expander in (60)]. Our implementation is quite efficient and analysis of 21 co-expression clusters with 1000 random iterations (including over 150 million network rewiring operations) takes 20 min on a standard laptop PC with 2.6 GHz processor and 2 GB of RAM. We believe that this type of analysis will be of immense value for future joint mRNA/miRNA expression studies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Chaim Linhart for helpful discussions and comments on an early version of this manuscript.

FUNDING

Edmond J. Safra Bioinformatics program at Tel Aviv University, Legacy Heritage Fund and EMBO long-term fellowship (to I.U.); Israel Science Foundation (grant no. 802/08); European Community's Seventh Framework Programme (TRIREME project, grant HEALTH-F4-2009-223575); Wolfson Family Charitable Trust. Funding for open access charge: Israel Science Foundation (grant no. 802/08); European Community's Seventh Framework Programme (TRIREME project, grant HEALTH-F4-2009-223575); Wolfson Family Charitable Trust; NIH 5K12HD001259-10 Women's Reproductive Health Research Career Development Program (to L.C.L.).

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
2. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
3. Selbach,M., Schwanhausser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
4. Fabbri,M., Garzon,R., Cimmino,A., Liu,Z., Zanesi,N., Callegari,E., Liu,S., Alder,H., Costinean,S., Fernandez-

Cymering,C. *et al.* (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl Acad. Sci. USA*, **104**, 15805–15810.
5. Korpal,M., Lee,E.S., Hu,G. and Kang,Y. (2008) The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.*, **283**, 14910–14914.
6. Cloonan,N., Brown,M.K., Steptoe,A.L., Wani,S., Chan,W.L., Forrest,A.R., Kolle,G., Gabrielli,B. and Grimmond,S.M. (2008) The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. *Genome Biol.*, **9**, R127.
7. Valastyan,S., Reinhardt,F., Benaich,N., Calogrias,D., Szasz,A.M., Wang,Z.C., Brock,J.E., Richardson,A.L. and Weinberg,R.A. (2009) A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell*, **137**, 1032–1046.
8. Miska,E.A., Alvarez-Saavedra,E., Abbott,A.L., Lau,N.C., Hellman,A.B., McGonagle,S.M., Bartel,D.P., Ambros,V.R. and Horvitz,H.R. (2007) Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genet.*, **3**, e215.
9. Rajewsky,N. (2006) microRNA target predictions in animals. *Nat. Genet.*, **38**, S8–S13.
10. Gaidatzis,D., van Nimwegen,E., Hausser,J. and Zavolan,M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
11. Nam,S., Kim,B., Shin,S. and Lee,S. (2008) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.
12. Creighton,C.J., Nagaraja,A.K., Hanash,S.M., Matzuk,M.M. and Gunaratne,P.H. (2008) A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA*, **14**, 2290–2296.
13. Nielsen,C.B., Shomron,N., Sandberg,R., Hornstein,E., Kitzman,J. and Burge,C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
14. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
15. Sandberg,R., Neilson,J.R., Sarma,A., Sharp,P.A. and Burge,C.B. (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643–1647.
16. Ventura,A., Young,A.G., Winslow,M.M., Lintault,L., Meissner,A., Erkeland,S.J., Newman,J., Bronson,R.T., Crowley,D., Stone,J.R. *et al.* (2008) Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell*, **132**, 875–886.
17. Xu,J. and Wong,C. (2008) A computational screen for mouse signaling pathways targeted by microRNA clusters. *RNA*, **14**, 1276–1283.
18. Farh,K.K., Grimson,A., Jan,C., Lewis,B.P., Johnston,W.K., Lim,L.P., Burge,C.B. and Bartel,D.P. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.
19. Halperin,Y., Linhart,C., Ulitsky,I. and Shamir,R. (2009) Allegro: analyzing expression and sequence in concert to discover regulatory programs. *Nucleic Acids Res.*, **37**, 1566–1567.
20. Shamir,R., Maron-Katz,A., Tanay,A., Linhart,C., Steinfeld,I., Sharan,R., Shiloh,Y. and Elkon,R. (2005) EXPANDER–an integrative program suite for microarray data analysis. *BMC Bioinformatics*, **6**, 232.
21. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
22. Shen-Orr,S.S., Milo,R., Mangan,S. and Alon,U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, **31**, 64–68.
23. Sharan,R. and Shamir,R. (2000) CLICK: a clustering algorithm with applications to gene expression analysis. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **8**, 307–316.
24. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
25. Laurent,L.C., Chen,J., Ulitsky,I., Mueller,F.J., Lu,C., Shamir,R., Fan,J.B. and Loring,J.F. (2008) Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells*, **26**, 1506–1516.
26. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
27. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B*, **57**, 289–300.
28. Girard,M., Jacquemin,E., Munnich,A., Lyonnet,S. and Henrion-Caude,A. (2008) miR-122, a paradigm for the role of microRNAs in the liver. *J. Hepatol.*, **48**, 648–656.
29. Esau,C., Davis,S., Murray,S.F., Yu,X.X., Pandey,S.K., Pear,M., Watts,L., Booten,S.L., Graham,M., McKay,R. *et al.* (2006) miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.*, **3**, 87–98.
30. Czech,M.P. (2006) MicroRNAs as therapeutic targets. *N. Engl. J. Med.*, **354**, 1194–1195.
31. Thum,T., Gross,C., Fiedler,J., Fischer,T., Kissler,S., Bussen,M., Galuppo,P., Just,S., Rottbauer,W., Frantz,S. *et al.* (2008) MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature*, **456**, 980–984.
32. Moschos,S.A., Williams,A.E., Perry,M.M., Birrell,M.A., Belvisi,M.G. and Lindsay,M.A. (2007) Expression profiling in vivo demonstrates rapid changes in lung microRNA levels following lipopolysaccharide-induced inflammation but not in the anti-inflammatory action of glucocorticoids. *BMC Genomics*, **8**, 240.
33. Dai,R., Phillips,R.A., Zhang,Y., Khan,D., Crasta,O. and Ahmed,S.A. (2008) Suppression of LPS-induced Interferon-gamma and nitric oxide in splenic lymphocytes by select estrogen-regulated microRNAs: a novel mechanism of immune modulation. *Blood*, **112**, 4591–4597.
34. Taganov,K.D., Boldin,M.P., Chang,K.J. and Baltimore,D. (2006) NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc. Natl Acad. Sci. USA*, **103**, 12481–12486.
35. Bhaumik,D., Scott,G.K., Schokrpur,S., Patil,C.K., Campisi,J. and Benz,C.C. (2008) Expression of microRNA-146 suppresses NF-kappaB activity with reduction of metastatic potential in breast cancer cells. *Oncogene*, **27**, 5643–5647.
36. Gregory,P.A., Bert,A.G., Paterson,E.L., Barry,S.C., Tsykin,A., Farshid,G., Vadas,M.A., Khew-Goodall,Y. and Goodall,G.J. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.*, **10**, 593–601.
37. Thiery,J.P. (2003) Epithelial-mesenchymal transitions in development and pathologies. *Curr. Opin. Cell Biol.*, **15**, 740–746.
38. Aigner,K., Dampier,B., Descovich,L., Mikula,M., Sultan,A., Schreiber,M., Mikulits,W., Brabletz,T., Strand,D., Obrist,P. *et al.* (2007) The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene*, **26**, 6979–6988.
39. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
40. Pickering,M.T., Stadler,B.M. and Kowalik,T.F. (2009) miR-17 and miR-20a temper an E2F1-induced G1 checkpoint to regulate cell cycle progression. *Oncogene*, **28**, 140–145.
41. Petrocca,F., Visone,R., Onelli,M.R., Shah,M.H., Nicoloso,M.S., de Martino,I., Iliopoulos,D., Pilozzi,E., Liu,C.G., Negrini,M. *et al.* (2008) E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell*, **13**, 272–286.
42. Sylvestre,Y., De Guire,V., Querido,E., Mukhopadhyay,U.K., Bourdeau,V., Major,F., Ferbeyre,G. and Chartrand,P. (2007) An E2F/miR-20a autoregulatory feedback loop. *J. Biol. Chem.*, **282**, 2135–2143.
43. Yu,Z., Wang,C., Wang,M., Li,Z., Casimiro,M.C., Liu,M., Wu,K., Whittle,J., Ju,X., Hyslop,T. *et al.* (2008) A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J. Cell. Biol.*, **182**, 509–517.

44. Card,D.A., Hebbar,P.B., Li,L., Trotter,K.W., Komatsu,Y., Mishina,Y. and Archer,T.K. (2008) Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol. Cell. Biol.*, **28**, 6426–6438.
45. Wang,Y., Baskerville,S., Shenoy,A., Babiarz,J.E., Baehner,L. and Blelloch,R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.*, **40**, 1478–1483.
46. Petrocca,F., Vecchione,A. and Croce,C.M. (2008) Emerging role of miR-106b-25/miR-17-92 clusters in the control of transforming growth factor beta signaling. *Cancer Res.*, **68**, 8191–8194.
47. Fontana,L., Pelosi,E., Greco,P., Racanicchi,S., Testa,U., Liuzzi,F., Croce,C.M., Brunetti,E., Grignani,F. and Peschle,C. (2007) MicroRNAs 17-5p-20a-106a control monocytopoiesis through AML1 targeting and M-CSF receptor upregulation. *Nat Cell Biol*, **9**, 775–787.
48. Rosa,A., Spagnoli,F.M. and Brivanlou,A.H. (2009) The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Dev. Cell*, **16**, 517–527.
49. Müller,F.J., Laurent,L.C., Kostka,D., Ulitsky,I., Williams,R., Lu,C., Park,I.H., Rao,M.S., Shamir,R., Schwartz,P.H. *et al.* (2008) Regulatory networks define phenotypic classes of human stem cell lines. *Nature*, **455**, 401–405.
50. Shkumatava,A., Stark,A., Sive,H. and Bartel,D.P. (2009) Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev.*, **23**, 466–481.
51. Stark,A., Brennecke,J., Bushati,N., Russell,R.B. and Cohen,S.M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell*, **123**, 1133–1146.
52. Gangaraju,V.K. and Lin,H. (2009) MicroRNAs: key regulators of stem cells. *Nat. Rev. Mol. Cell. Biol.*, **10**, 116–125.
53. Xu,N., Papagiannakopoulos,T., Pan,G., Thomson,J.A. and Kosik,K.S. (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*.
54. Slack,F. (2009) let-7 microRNA reduces tumor growth. *Cell Cycle*, **8**, 1823.
55. Glazov,E.A., McWilliam,S., Barris,W.C. and Dalrymple,B.P. (2008) Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.*, **25**, 939–948.
56. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
57. Das,M.K. and Dai,H.K. (2007) A survey of DNA motif finding algorithms. *BMC Bioinformatics*, **8 (Suppl. 7)**, S21.
58. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al.* (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
59. van Dongen,S., Abreu-Goodger,C. and Enright,A.J. (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat. Methods*, **5**, 1023–1025.
60. Ulitsky,I., Maron-Katz,A., Shavit,S., Sagir,D., Linhart,C., Elkon,R., Tanay,A., Sharan,R., Shiloh,Y. and Shamir,R. (2010) Expander: from expression microarrays to networks and functions. *Nat. Protoc.*, **5**, 303–322.