Published online 2 December 2012

Nucleic Acids Research, 2013, Vol. 41, No. 3 ***e45***
doi:10.1093/nar/gks1142

miRNA target enrichment analysis reveals directly active miRNAs in health and disease

Israel Steinfeld[1,*], Roy Navon[2], Robert Ach[3] and Zohar Yakhini[1,2]

[1]Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, [2]Agilent Laboratories, Agilent Technologies, Tel Aviv 49527, Israel and [3]Agilent Laboratories, Agilent Technologies, Santa Clara, CA 95051, USA

Received April 18, 2012; Revised October 23, 2012; Accepted October 24, 2012

ABSTRACT

microRNAs (miRNAs) are short non-coding regulatory RNA molecules. The activity of a miRNA in a biological process can often be reflected in the expression program that characterizes the outcome of the activity. We introduce a computational approach that infers such activity from high-throughput data using a novel statistical methodology, called minimum-mHG (mmHG), that examines mutual enrichment in two ranked lists. Based on this methodology, we provide a user-friendly web application that supports the statistical assessment of miRNA target enrichment analysis (miTEA) in the top of a ranked list of genes or proteins. Using miTEA, we analyze several target prediction tools by examining performance on public miRNA constitutive expression data. We also apply miTEA to analyze several integrative biology data sets, including a novel matched miRNA/mRNA data set covering nine human tissue types. Our novel findings include proposed direct activity of miR-519 in placenta, a direct activity of the oncogenic miR-15 in different healthy tissue types and a direct activity of the poorly characterized miR-768 in both healthy tissue types and cancer cell lines. The miTEA web application is available at http://cbl-gorilla.cs.technion.ac.il/miTEA/.

INTRODUCTION

microRNAs (miRNAs) are short (usually ~22 nt) non-coding regulatory RNA molecules. Hundreds of miRNAs have been discovered in recent years and several have been functionally characterized (1). In mammals, miRNAs are well known to take part in regulating tissue differentiation (2) and for several miRNAs a well-defined tissue specific signature is known (3). As such, miRNAs are known to regulate major biological processes such as development, cancer (4,5) and heart function (6).

Many studies attempted to elucidate the mechanism by which miRNAs act to regulate target genes. With recent experimental studies, many of the major factors that partake in the recognition mechanism of miRNA targeting have been revealed (7,8). Better understanding of the miRNA mechanism of regulation led to the development of a large variety of computational tools designed to predict which genes are targeted by any miRNA of interest (9). We herein refer to these tools as miRNA target prediction algorithms (miTPAs). The current study uses the publicly available predictions of several such miTPAs.

The refined characterization of miRNA targets enables better understanding of the role of miRNAs in various biological processes by combining measurement in relevant samples with analysis that takes information about targets into account. Several groups have developed computational tools to infer miRNA activity by analyzing their targets in mRNA transcription profiles (10–15). Sood *et al*. (10) used the Pictar miTPA (16) and defined cell-type specific signatures of miRNAs by searching for enriched miRNA targets in expression profiles using the Wilcoxon rank sum test. Cheng and Li (11) used the miRanda miTPA (17) to identify miRNA activity in miRNA transfected HeLa cells. The statistical approach of Cheng and Li (11) employs a generalization of the enrichment score used by GSEA (18) which requires a permutation step to infer the significance level. Liang *et al*. (14) have used Targetscan (19) to report miRNA activity in a breast cancer data set and in miRNA transfected HeLa cells. Their method assesses miRNA activity in every sample in the cohort. These per-sample activity scores, based on *t*-test variants, are then compared for different classes of samples in the cohort, using variants of the Kruskal–Wallis test. In general, current approaches, including the methods mentioned earlier, work with a fixed set of genes as the target set of a specific miRNA, either by using miTPA predictions or simply by considering the miRNA

*To whom correspondence should be addressed. Tel: +972 3 928 8545; Fax: +972 3 928 8501; Email: israels@cs.technion.ac.il

seed sequence (20,21). Subsequently, the expression distribution of this set is tested for divergence compared to the overall transcription profile. When significant divergence in observed the miRNA itself can be deduced to be active. A few groups have also provided online services to infer miRNA activity (14,15). While miRNA targets behave in a coherent manner in synthetic and controlled environments or experimental systems, where the miRNA is artificially induced or repressed, their coordinated activity is obscured in most biological conditions. Therefore, there is a need for more sensitive analysis approaches that can capture more subtle trends in the data.

In accordance with the notion that transcription networks are not discrete and should be modelled as quantitative relations (22) and to enable the performance of more accurate and sensitive analysis of miRNA target enrichment, we avoid the use of a predefined set of mRNAs to describe the targets of a specific miRNA. We leverage the additional information provided by the miTPA, in the form of a prediction score or significance level. We thereby produce better predictions of miRNA activity, including subtle trends as earlier. In this article, we address mutual enrichment in two ranked lists of elements and develop a statistical framework and software for this purpose, including miRNA target enrichment analysis (miTEA)—a publically available web-based application. We apply our methods to synthetic expression profiles to show the strength of our approach as well as to assess the robustness level of different miTPAs, in the context of these specific experiments. We further demonstrate an analysis of an integrative biology high-throughput data set of mRNA and miRNA expression profiling and shed light on several aspects of miRNA mechanisms of regulation in healthy human tissue data sets as well as in cancer data sets.

MATERIALS AND METHODS

Expression profiling

Protein and mRNA expression profiles for miRNA transfected HeLa cells were taken from (23,24). Matched mRNA and miRNA expression profiles for cancer cell lines and primary tumours were taken from (25,26), respectively. miRNA profiles from nine human tissue types were described in Ach *et al*. (27). For mRNA profiles from these tissue types, the same total RNA preps used for miRNA profiling in Ach *et al*. (27) were labelled in duplicates and profiled on human whole genome gene expression DNA microarrays from Agilent Technologies, according to the manufacturer's protocols (www.agilent.com). Data were deposited in GEO with the accession number GSE31904.

Repeat measurements of the same tissue type were averaged resulting in one profile for each tissue type. To obtain a tissue type-specific ranked list of genes, the expression of each gene was standardized (mean = 0, SD = 1) across samples and for each tissue type genes were ranked according to their standardized expression signal.

Minimum-mHG

The mmHG statistics is a generalization of the mHG statistics (28–32). While the mHG statistics quantifies the enrichment level of a set of elements in the top of a ranked list of elements, the mmHG statistics quantifies the mutual enrichment level for two ranked list of elements. As such, the mmHG statistics is applicable to any two ranked lists of common elements. While any parametric or non-parametric correlation statistics (e.g. Spearman's correlation coefficient), that takes the same input, calculates the overall agreement between the two ranked lists, the mmHG statistic focuses only on agreement at the top of the two ranked lists. The mmHG calculates how many elements are common in the top of both lists, without predefining what the top is. Its output is simply the chance for getting the obtained size of intersection at the top of the two ranked lists of elements at random (the enrichment *P*-value). In this article, we assess the mutual enrichment in two ranked lists of genes—one ranked according to an expression-based measurement and one according to a miTPA score. Another example of an application is to assess mutual enrichment when genes are ranked according to differential expression in two types of disease or other biological condition.

As our main application will be for genes we will, from here on, focus on genes as the ranked elements. A formal definition of the mmHG statistics follows. Given a single permutation $\pi \in S_N$ and for every $i = 1 \ldots N$ we define a binary vector λ_i in which exactly i entries are 1 and $N - i$ are 0, as follows:

$$\lambda_i(j) = 1 \; iff \; \pi(j) \leq i \tag{1}$$

We define the mmHG score of a permutation π as:

$$\mathrm{mmHG}(\pi) = \min_i P\text{−value}(\mathrm{mHG}(\lambda_i)) \tag{2}$$

We have previously introduced the mHG statistics to evaluate the enrichment of a fixed gene set within a ranked list of genes (28,29). For completeness, we define the mHG score of a ranked binary vector λ as:

$$mHG(\lambda) = \min_{1 \leq n \leq N} HGT(N, B, n, b_n(\lambda)) \tag{3}$$

Where,

$$N = |\lambda|, b_n(\lambda) = \sum_{i=1}^{n} \lambda_i, B = b_N(\lambda) \tag{4}$$

and

$$HGT(N, B, n, b) = \Pr(X \geq b) = \sum_{i=b}^{\min(n, B)} \frac{\binom{n}{i}\binom{N-n}{B-i}}{\binom{N}{B}} \tag{5}$$

is the tail probability for a random variable X, having an appropriate hypergeometric distribution. mHG *P*-values used earlier, denoted *P*-value(mHG(λ)), assume a uniform null distribution of binary vectors with weight B. These *P*-values are exact and do not require correction

for the multiple thresholds tested. See details of the underlying methodology in Eden *et al.* (28).

For two permutations, $\pi_1 = (\pi_1(1), \ldots, \pi_1(N))$ and $\pi_2 = (\pi_2(1), \ldots, \pi_2(N))$, over a universe of N genes, the relative permutation π, of π_2 w.r.t. π_1, is defined by

$\pi(\pi_1(j)) = \pi_2(j)$, for every $j = 1 \ldots N$ or simply, using operations in the permutation group S_N:

$$\pi = \pi_2 \cdot \pi_1^{-1} \quad (6)$$

We are now ready to define $\mathrm{mmHG}(\pi_1, \pi_2)$ representing the mutual enrichment of two ranked list of genes as:

$$\mathrm{mmHG}(\pi_1, \pi_2) = \mathrm{mmHG}(\pi) \quad (7)$$

where π is the relative permutation of π_2 w.r.t. π_1 as defined earlier.

Although the mmHG score is a minimum taken over a space of mHG *P*-values, it cannot be considered as a significance measure, due to the additional multiple testing. To assess the statistical significance of an mmHG result obtained in actual analysis, we work over a null model that consists of a uniform measure on the group S_N. This is equivalent to a uniform measure over $S_N \times S_N$ by the conversion of any π_1, π_2 to their relative permutation $\pi = \pi_2 \bullet \pi_1^{-1}$ as earlier.

For mHG, there is a dynamic programming process that provides a full characterization of the distribution of the statistics under the null model. For mmHG, we do not currently have an efficient process that allows for calculating exact *P*-values. We can however use a union bound approach as follows:

$$P\text{-value}(\mathrm{mmHG}(\pi)) < \mathrm{mmHG}(\pi) \times N \quad (8)$$

mmHG for miRNA target enrichment in this study

In this article, we use gene scores given by any miTPA to define gene ranking π_1. For a specific miRNA, genes are therefore ranked according to how likely they are to be targeted by that miRNA. The following ranking schemes are used in the analyses performed in this study to define the gene ranking π_2:

(A) Protein expression values and related changes, derived from SILAC/pSILAC experiments.
(B) mRNA expression values derived from high-throughput transcription profiling .and
(C) In integrated miRNA–mRNA data—for a specific miRNA, called the pivot miRNA, mRNAs were ranked according to the anti-correlation level (Pearson's *r*) of their expression patterns, across the entire cohort of samples, to the pivot miRNA expression pattern. See Enerly *et al.* (26) for an example of a similar analysis.

When using any specific miTPA we reduce the gene universe to the genes for which the miTPA reports results (genes targeted by any miRNA). In addition, for each expression experiment the gene universe was further reduced to the genes that were actually measured in the experiment and reported to have produced an interpretable signal. The same is true for the protein expression experiments.

Practical implementation

Some modifications to the abstract definition of mmHG were used in this study, to accommodate practical considerations. These are described in this section. First, instead of computing the *P*-value of an mHG result as required in Equation (2), we used the following bound which follows from Eden *et al.* (28):

$$P\text{-value}(\mathrm{mHG}(\lambda_i)) \leq \mathrm{mHG}(\lambda_i) \cdot i \quad (9)$$

Thus, instead of computing the mmHG score we define the mmHG^* score to be:

$$\mathrm{mmHG}^*(\pi) = \min_{1 \leq i \leq N} \mathrm{mHG}(\lambda_i) \cdot i \quad (10)$$

Notice that from Equation (9) it follows that:

$$\mathrm{mmHG}(\pi) \leq \mathrm{mmHG}^*(\pi) \quad (11)$$

and thus from Equation (8) we get:

$$P\text{-value}(\mathrm{mmHG}(\pi)) \leq \mathrm{mmHG}(\pi) \cdot N \leq \mathrm{mmHG}^*(\pi) \cdot N \quad (12)$$

As a second practical modification, the optimization processes described in Equations (2) and (3) are not carried out in their full exhaustive scope. We only run the outer loop up to a reasonable threshold—*n_max*. The altered equations are therefore:

$$\mathrm{mmHG}_{\mathrm{n\_max}}(\pi) = \min_{1 \leq i \leq n\_max} P\text{-value}(\mathrm{mHG}_{n\_max}(\lambda_i)) \quad (13)$$

$$\mathrm{mHG}_{\mathrm{n\_max}}(\lambda) = \min_{1 \leq n \leq n\_max} \mathrm{HGT}(N, B, n, b_n(\lambda)) \quad (14)$$

$$P\text{-value}(\mathrm{mmHG}_{n\_max}(\pi)) \leq \mathrm{mmHG}_{n\_max}(\pi) \cdot n\_max. \quad (15)$$

Defining the bound-based version:

$$\mathrm{mmHG}^*_{n\_max}(\pi) = \min_{1 \leq i \leq n\_max} \mathrm{mHG}_{n\_max}(\lambda_i) \cdot i \quad (16)$$

we now get:

$$P\text{-value}(\mathrm{mmHG}_{n\_max}(\pi)) \leq \mathrm{mmHG}^*_{n\_max}(\pi) \cdot n\_max \quad (17)$$

For all parts of this study, we used the bound presented in the right hand side of Equation (17) to represent the significance level of an mmHG result. The bound used in the analysis was $n\_max = 2000$ excluding the integrative analysis part which used $n\_max = 3000$.

Since the complexity of computing the mHG *P*-value is $O(N^2)$, we get that the complexity of computing $\mathrm{mmHG}(\pi)$ is $O(N^3)$. Using the bounds presented earlier, we get a reduced complexity of $O(N^2)$ for computing $\mathrm{mmHG}^*(\pi)$ or $O(n\_max^2)$ if we consider Equation (17).

miRNA mutual exclusion

A scored list of miRNAs were mutually excluded according to miRNA family or similar 6-bp seed, as obtained from miRBase (1), giving precedence to high-scoring miRNAs.

Tool description (website)

A web-based application miTEA is publically available, free for non-commercial use, at http://cbl-gorilla.cs.technion.ac.il/miTEA/

miTEA is designed to statistically deduce miRNA activity from high-throughput measurement results. The input for miTEA is a ranked list of genes for which the mmHG statistics is used. The miTPA ranking is embedded into miTEA and is updated periodically. The web interface for miTEA is shown in Supplementary Figure S1. miTEA supports the following formats of gene nomenclature: gene symbol, protein/gene RefSeq, Uniprot, Unigene or Ensembl. miTEA automatically removes gene redundancies keeping the highest ranking occurrence (in each of the two lists, when applicable). This includes dealing with duplicates that hide behind different nomenclatures. miTEA currently supports the following organisms: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster* and *Danio rerio*, each with miTPAs that support it. The output consists of a colour-coded miRNA network, where the nodes are the enriched miRNAs and an edge between two nodes is present if the corresponding miRNAs have substantial overlap in their target genes. The network's nodes are colour-coded according to the detected enrichment levels. The output also includes a table consisting of the enriched miRNAs (with web links to additional information), statistical characterization of the results and the set of targeted genes found to be enriched in the top. Results can be exported to Excel. A running example can be found in miTEA's website. A single run of miTEA, spanning all human miRNAs, on the web server, using default parameters, takes ~13 s.

Simulated data

Each simulated data set, E, was composed of 100 samples, where in each sample we generated a simulated expression profile spanning all 17 297 genes that are available in Targetscan (V5.2). The expression level of each gene in each sample was independently randomly drawn from a standard normal distribution, $E(g,s) \sim N(0,1)$ for gene g and sample s. In each simulated data set 10 random miRNAs were simulated to be active at different levels of influence: the first repressing its top 100 targets, the second repressing its top 200 targets, ..., the last repressing its top 1000 targets. In addition, for each simulated data set, the level of activity, α, of the miRNAs is changed. The miRNA repression was simulated for only 50 of the 100 samples. The expression levels of the miRNA targets in the affected samples were reduced by $(\alpha + \varepsilon)$, where ε is drawn from a standard normal distribution. That is, $E(g,s) \sim (N(0,1) - (\alpha + N(0,1)))$ for a target g of an active miRNA, in one of the affected samples s. For each level of α, three simulated data sets were generated.

While mirAct uses the entire data set as input, DIANA-mirExTra takes only a target set and background set of genes as input and miTEA uses a ranked list of genes as input. Therefore, when testing the simulated data set using miTEA and DIANA-mirExTra genes were ranked according to their t-test P-value down-regulation in the effected samples. For DIANA-mirExTra, genes with $P < 0.05$ were assigned to the target set and the rest to the background set. The use of miTEA, mirAct and DIANA-mirExTra was carried out with default parameters as provided in the online versions.

RESULTS

Web-based miTEA tool

In this study, we introduce miTEA, an online web application designed to detect and measure miRNA regulatory activity in high-throughput measurement results. miTEA takes as input a ranked list of genes that represents the results of a high-throughput measurement experiment. It employs the output of any selected miTPA to define, in a robust manner, a ranked list of genes representing the targets of each miRNA in the relevant organism. Using mmHG, a novel statistical enrichment method (see 'Materials and Methods' section), miTEA finds mutual enrichment in the two ranked lists of genes and thus statistically infers miRNA activity (see Figure 1 and 'Materials and Methods' section for more details). The main advantages of miTEA, compared to other similar tools (10–15) are:

- The use of all prediction scores given by the miTPA rather than setting a threshold and predefining the set of genes targeted by a specific miRNA.
- A novel statistical approach that provides an assessment of the statistical significance of the obtained results without using simulations.
- Its availability as an efficient web tool easily accessible to the community.

Evaluation through a comparison of miTPAs

As a benchmark for miRNA activity, we first tested miTEA on two seminal studies that measured global mRNA as well as protein expression profiles in response to a perturbation of a single miRNA, in a controlled environment (24,23). The data span 8 protein expression profiling and 13 mRNA expression profiling results following constitutive miRNA over-expression experiments. To better assess the sensitivity of our approach, we applied miTEA using eight different miTPAs: Targetscan (v5), microCosm (v5), PITA, PicTar, MicroT (v3), RNA22, targetRank and EIMMo (v4) (1,16,19,33–37). In addition, we also applied miTEA using a Targetscan version of only conserved sites (Figure 2). In each of the experiments, all genes were ranked according to their down-regulation and these ranked lists of genes were used as input to miTEA (Figure 1). In the majority of the experiments and for many of the miTPAs, miTEA detects a significant enrichment of the targets of the perturbed miRNA in the down-regulated genes—as expected (Figure 2). Furthermore, miTEA finds the perturbed miRNA targets to be most enriched compared to targets of all other non-perturbed

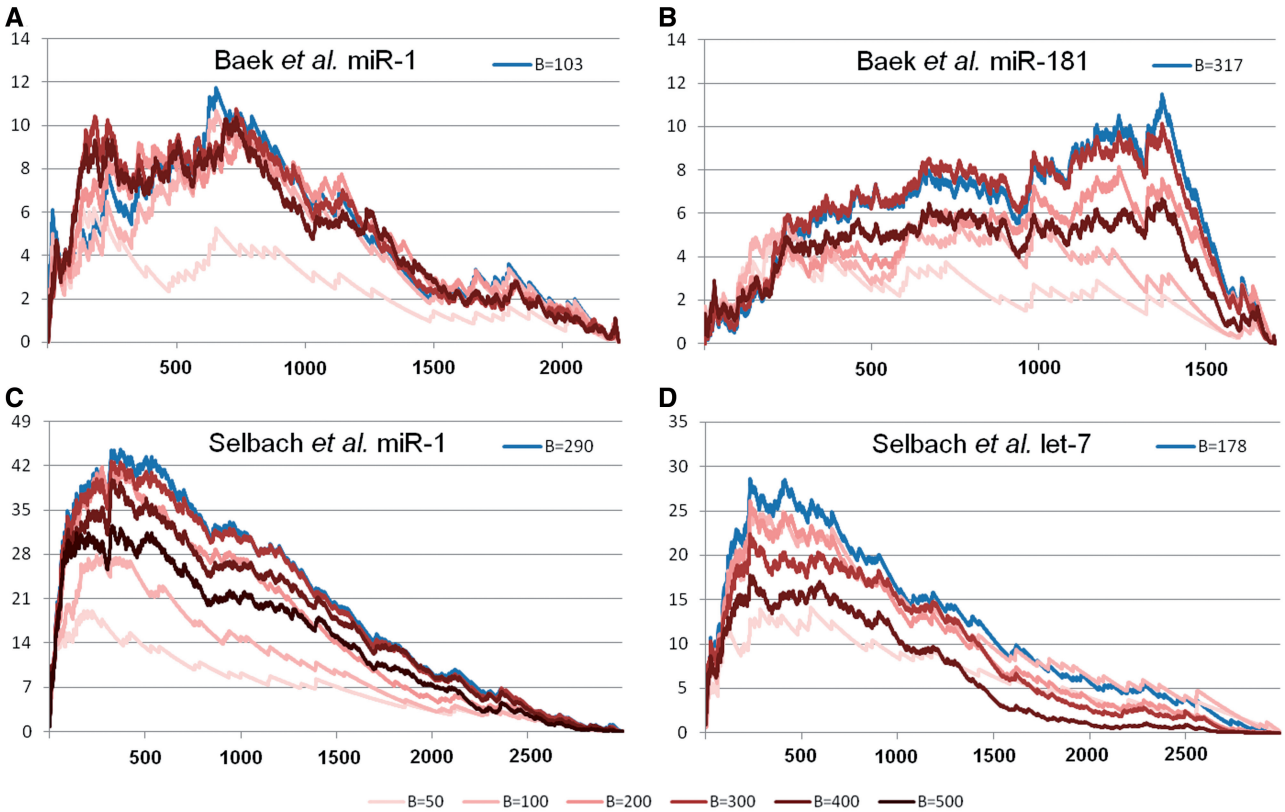

Figure 1. mmHG applied to lists of genes ranked according to their down-regulation following a specific miRNA transfection. mmHG circumvents the need to define the set of affected genes in the experiment and the set of genes targeted by a specific miRNA by exhaustingly testing all options and selecting the optimal thresholds. Hypergeometric *P*-values (*y*-axis) are plotted in −log10 scale as a function of the threshold used to define the list of down-regulated genes (*x*-axis). Lines with different shades of red represent different selection of top targeted genes by the relevant miRNA, as defined by Targetscan (V5.2). For example, the lightest shade of red represents the selection of the top 50 targeted genes as the set of genes targeted by the miRNA. The blue line represents the optimal threshold selection for top targeted genes. Different panels represent different protein expression profiling experiments: (**A**) miR-1 transfection (24), (**B**) miR-181 transfection (24), (**C**) miR-1 transfection (23) and (**D**) let-7 transfection (23). As can be observed, the selection of the set of genes targeted by the miRNA or the set of genes down-regulated in the experiment can vastly influence the enrichment results, thus requiring a more flexible approach.

miRNAs in most of the experiments (Figure 2). In particular, when the miTPA used was Targetscan miTEA detected the correct perturbed miRNA in 20 out of the 21 experiments tested.

The high range of enrichment scores obtained for different miTPAs could be a result of the different level of accuracy in the target predictions provided by the miTPAs. Similarly, different level of noise and robustness can explain the range of enrichment scores obtained for the different experiments. In this respect, two additional interesting insights arise. First, Targetscan shows the best specificity and sensitivity for miRNA detection by yielding the highest enrichment scores in most experiments. The proteomics work was done using pSILAC (23) and SILAC (24) technologies. A second insight is that the enrichment scores in the pSILAC data are more significant than the enrichment scores in the SILAC data. In particular, the two studies measure response to miR-1 over-expression. The targets (as predicted by Targetscan) of miR-1 are enriched in the pSILAC data with mmHG $P < 10^{-71}$ while in the SILAC experiment they are enriched with mmHG $P < 10^{-9}$.

Comparing miTEA to existing online tools

To evaluate miTEA performance in comparison to other existing online tools we compared miTEA to mirAct (14) and mirExTra (15) using a simulated data set as well as using the miRNA over-expression experiments used earlier. The simulated data sets were designed to model the activity of a small set of miRNAs in a subset of the samples in different levels of activity (see 'Materials and Methods' section for complete description of data simulation process). Our results show that miTEA is able to detect the active miRNAs with higher level of sensitivity and specificity even in cases of low activity (Figure 3). For the over-expression experiments miTEA is able to find the activity of the perturbed miRNA with higher level of significance but more importantly with higher specificity—finding the perturbed miRNA to be most active (Supplementary Table S1). For example, for the mRNA profiling data measured 32 h after let-7 transfection (23) mirExTra finds miR-608 to yield better results ($P < 3.1 \times 10^{-9}$, data not shown) than those of let-7. Similarly, in the same experiment, mirAct finds miR-491 to be more active (data not shown). Both are inconsistent

Selbach protein	TS	TScon	PITA	Miranda	PicTar	EIMMo	MicroT	RNA22	TRank
let-7	26.72	23.53	14.48	4.11	13.86	25.00	21.25	3.13	7.69
miR-1	43.25	36.00	18.03	8.54	23.05	43.61	41.90	3.09	25.37
miR-155	72.89	17.38	50.37	6.86	9.12	68.71	63.71	4.59	35.30
miR-16	34.09	27.81	16.91	4.15	17.45	35.25	41.15	4.12	24.36
miR-30	23.49	17.92	11.48	4.33	12.39	21.37	26.37	1.65	20.47
Selbach mRNA									
let-7b_8h	1.49	0.00	0.00	0.34	0.00	0.00	0.00	1.53	0.00
let-7b_32h	12.05	8.01	3.41	0.87	6.14	10.88	9.79	0.00	3.16
miR-155_8h	10.11	2.74	4.61	0.44	0.00	7.41	5.59	0.00	5.92
miR-155_32h	63.33	12.44	34.71	7.70	5.51	64.00	55.02	1.13	40.61
miR-16_8h	10.79	6.82	3.87	1.03	2.66	8.63	5.55	0.00	3.60
miR-16_32h	46.48	30.21	20.15	4.30	17.75	48.87	39.04	0.78	28.31
miR-1_8h	39.99	31.04	15.55	7.53	23.91	34.35	28.35	1.88	28.06
miR-1_32h	75.76	60.88	26.86	8.78	36.05	73.54	66.35	8.49	61.10
miR-30a_8h	2.94	2.65	2.02	1.41	4.52	0.00	0.44	0.00	0.00
miR-30a_32h	47.41	42.13	11.69	7.26	21.86	39.59	40.16	1.64	39.52
Baek protein									
miR-1	9.28	8.07	5.60	2.36	4.04	10.17	11.10	1.58	4.29
miR-124	13.25	11.66	7.93	3.47	3.88	12.69	10.17	0.71	5.99
miR-181	8.81	4.79	4.11	2.63	1.11	2.91	7.03	2.22	6.02
Baek mRNA									
miR-1	37.18	31.55	13.23	4.38	18.86	35.35	26.10	2.97	27.77
miR-124	143.73	107.01	44.38	40.45	15.47	110.24	114.25	4.34	70.82
miR-181	50.10	27.30	21.94	6.16	13.83	24.80	28.62	3.50	39.15

Figure 2. Results of applying miTEA using a data set of miRNA transfection experiments. The table provides the enrichment scores [given in $-\log_{10}$(mmHG *P*-value)] obtained for two high-throughput data sets of miRNA transfection experiments (23,24). Each row represents a different experiment, for either transcriptomics or proteomics and each column represents an application of miTEA with a different miTPA. Each row is coloured in gradient to red according to the enrichment score compared to other enrichment scores obtained for the relevant transfected miRNA experiment. Enrichment scores are marked in blue when the targets of the transfected miRNA, in that specific experiment, were not found to be most enriched compared to targets of other miRNAs. TS_con stands for Targetscan representation of only conserved sites. Higher scores could be due to better target predictions or due to less noisy and more robust profiling experiments. As can be observed, using Targetscan (V5.2), miTEA obtains the highest enrichment scores in 15 out of the 21 experiments and also finds the correct perturbed miRNA in 20 out of 21 experiments. Thus, Targetscan exhibits the highest sensitivity and specificity in detecting miRNA activity compared to the other miTPAs.

with the experimental set-up. miTEA clearly identifies let-7 as the most active miRNA ($P < 9 \times 10^{-13}$).

Tissue-specific miRNA activity

The strength of miTEA is very well established in the highly controlled miRNA over-expression experiments described earlier. In the more practical contexts of less controlled data sets, such as clinical ones, we should consider a milder effect of miRNAs on their targets. Therefore, to better capture miTEA sensitivity in detecting miRNA activity, we applied it to mRNA profiling data set of human tissue types measured for the purpose of this study. In this analysis, we applied miTEA in combination with Targetscan (V5.2), since among the different miTPAs the latter demonstrates the highest specificity and sensitivity in detecting miRNA activity in the controlled experiments (Figure 2) and since it was comparable to other version of Targetscan (Supplementary Table S1).

The data set we used consist of 20 human tissue samples spanning nine different healthy tissue types of origin (see 'Materials and Methods' section). To test miRNA activity in the different tissue types, we ranked all the genes, for each tissue type separately, according to the level of their down-regulation in that specific tissue. Applying miTEA on each of the resulting tissue-specific ranked list of genes, we composed a miRNA activity map for the different tissue types (see Figure 4 and Supplementary Table S2). The results confirmed the specific miRNA activity for miR-124 in brain (38) (mmHG $P < 4 \times 10^{-11}$), miR-122 in liver (39) (mmHG $P < 2 \times 10^{-6}$) and miR-1/miR-206 in skeletal muscle (40,41) (mmHG $P < 6 \times 10^{-10}$).

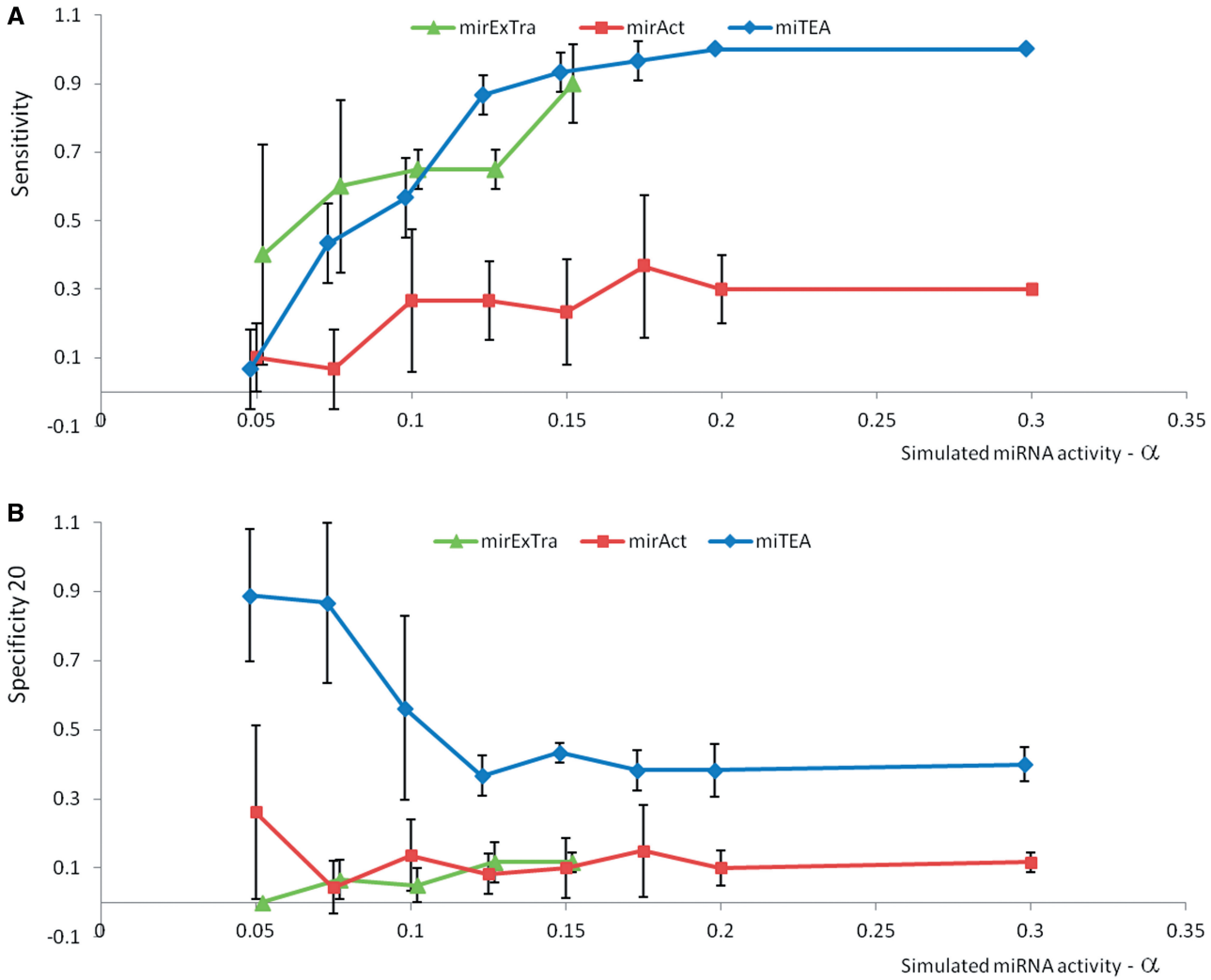

Figure 3. Comparing miTEA to alternative online tools. We compared miTEA to mirAct (14) and mirExTra (15) using simulated data (see 'Materials and Methods' section for full description of the simulation process). The data simulated the activity of 10 miRNAs in different levels of activity—α. The active miRNA affected their targets in a different manner—some regulated only their top 100 targets while others regulated their top 1000 targets. For each level of α three simulated data sets were generated and the different tools were applied on each of the simulated data set. Panel **A** and **B** depict the observed sensitivity and specificity levels, respectively, as a function of the simulated activity—α. The measure of specificity that was used in this analysis is the percent of truly active miRNAs in the top 20 significantly detected miRNAs. For $\alpha > 1.5$, the targets set when using mirExTra was too big and did not give any results. As can be noted miTEA achieves higher sensitivity and specificity levels and these are significant even in low simulated activity levels.

Our analysis also suggests an activity of miR-519a in placenta (mmHG $P < 3 \times 10^{-5}$, Supplementary Figure S2a). miR-519 was observed to be specifically expressed in placenta (42,43). A direct effect on its targets, however, has not been reported so far, to our knowledge.

DA-miRs in human tissues types

miRNA regulatory contributions to differentiation are not expected to be manifested only in single tissue specificity, but rather in differences that pertain to the entire differentiation programs that lead to distinct tissue types. Using miRNA expression profiling data from the same cohort of healthy tissue samples, we next took a more exploratory approach to detect complex miRNA activity, as appose to tissue specific miRNA activity. We selected each miRNA as a pivot and ranked all the genes according to the level of anti-correlation between each gene's mRNA expression profile and the pivot miRNA expression profile across the cohort of samples. We then applied miTEA, using Targetscan (V5.2), to the above-ranked list to test the enrichment of pivot targets in its anti-correlated genes. We call miRNAs for which such enrichment is observed directly active miRNAs (*DA-miRs*). As expected, our results indicate that the list of *DA-miRs* forms a superset of the tissue-specific active miRNA described earlier. Specifically, we see that of the 470 profiled miRNAs that are also covered by Targetscan database, 71 are found to be *DA-miRs* ($P < 0.05$ after Bonferroni multiple testing correction, see Supplementary Table S3 and 'Materials

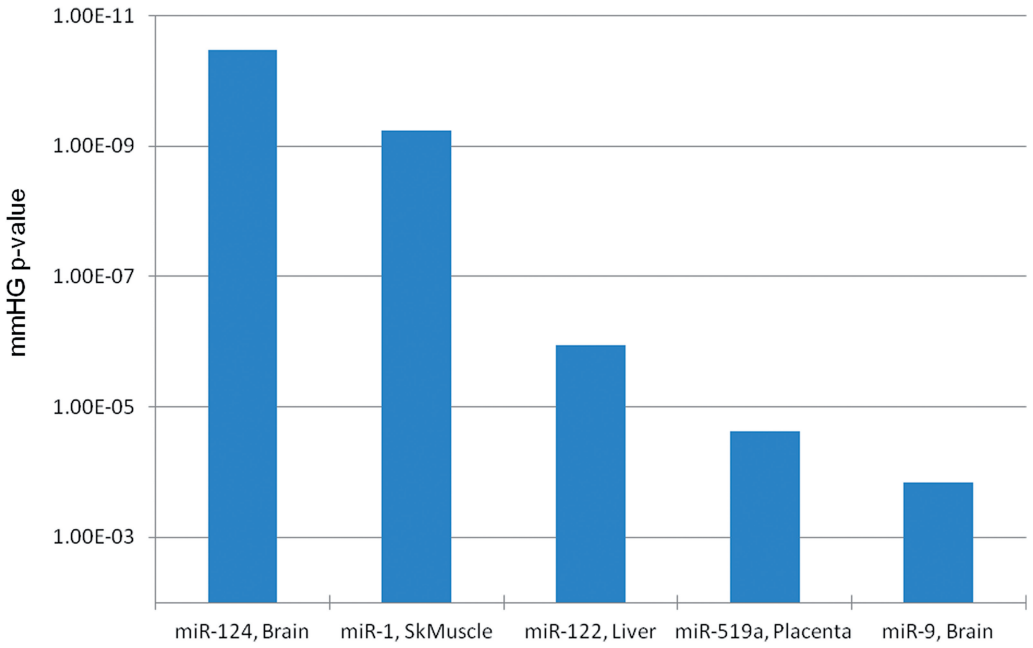

Figure 4. Tissue-specific miRNA activity. mRNA genes were ranked according to their tissue-specific down-regulation (see 'Materials and Methods' section). The tissue-specific ranked list of genes was given as input to miTEA. The figure presents results for top pairs of miRNAs and tissue types. Most of the observed miRNA activity is consistent with known miRNA tissue-specific activities (e.g. miR-124 and brain as well as 122 and liver).

and Methods' section). After mutually excluding miRNAs from the same family we end up with 30 *DA-miRs* found in this data set of human tissue-type samples (Supplementary Table S3). Further examination of these *DA-miRs* showed that, although found in healthy human tissues, many were either shown to have elevated expression levels in cancer [e.g. miR-135b and miR-182 (44)] or directly implicated in cancer development and metastasis [e.g. miR-15b (45), miR-200 (46) and the oncogenic miR-17/92 cluster (47)]. However, to our knowledge, not all of the *DA-miRs* have been implicated in association with cancer (e.g. miR-512-3p, miR-768). Their direct activity in different human tissues might imply a role in differentiation, a role that when impaired or disrupted might lead to cancer development.

Study of miRNA regulation in health and disease

The link between the *DA-miRs* in healthy human tissues and oncogenicity has led us to investigate the miRNA activity map in cancer samples. The NCI-60 panel is composed of 60 cell lines representing nine different cancer types (48). We applied our miTEA-based analysis pipeline presented in the previous section to the NCI-60 matched miRNA–mRNA expression data set (49) to obtain the NCI-60 miRNA activity map. Out of 338 profiled miRNAs, 47 were found to be *DA-miRs* ($P < 0.05$ after Bonferroni multiple testing correction, see Supplementary Table S4). After mutually excluding miRNAs from the same family, we end up with 27 *DA-miRs* in the NCI-60 data set (see Supplementary Table S4). In line with our observation regarding *DA-miRs* in the healthy tissue samples, many of the NCI-60 *DA-miRs* are also known to be associated with cancer development (e.g. miR-19, miR-429 and miR-20). We also note that some of the *DA-miRs* observed in this cancer model data set were not previously shown to be associated with cancer (e.g. miR-194 and miR-595).

There are only three tissue types overlapping between the healthy and NCI-60 data sets (breast, ovary and brain), yet we further explored the similarity between the two miRNA activity maps. Out of 258 miRNAs that can be commonly mapped in both data sets, 42 and 44 are *DA-miRs* in the healthy and in the NCI-60 data sets, respectively, with a significant overlap of 14 *DA-miRs* (hypergeometric $P < 0.004$). It is possibly more interesting to note those miRNAs that loose/gain their direct activity when comparing normal to highly proliferative cellular conditions (see Figure 5 and Supplementary Table S5). We find 30 miRNAs that are specifically active in cancer (e.g. miR-142 and miR-29b) and 28 miRNAs that are specifically active in healthy tissue types (e.g. miR-15b and miR-148a). Moreover, deviating from the widely recognized role of miRNAs as repressors, we see a small set of miRNAs (19 in the tissue type data set and 5 in the NCI-60 data set) that are significantly correlated with their targets rather than anti-correlated. For example, miR-377 yields enrichment levels of mmHG $P < 6 \times 10^{-11}$ and mmHG $P < 9 \times 10^{-6}$ in the tissue types and NCI-60 data sets, respectively, when considering positive correlatees.

DA-miRs in breast cancer

Thus far, we describe miTEA results for heterogeneous data sets composed of several different tissue types. To further develop our understanding of miRNA direct activity and test miTEA in a more homogenous context,

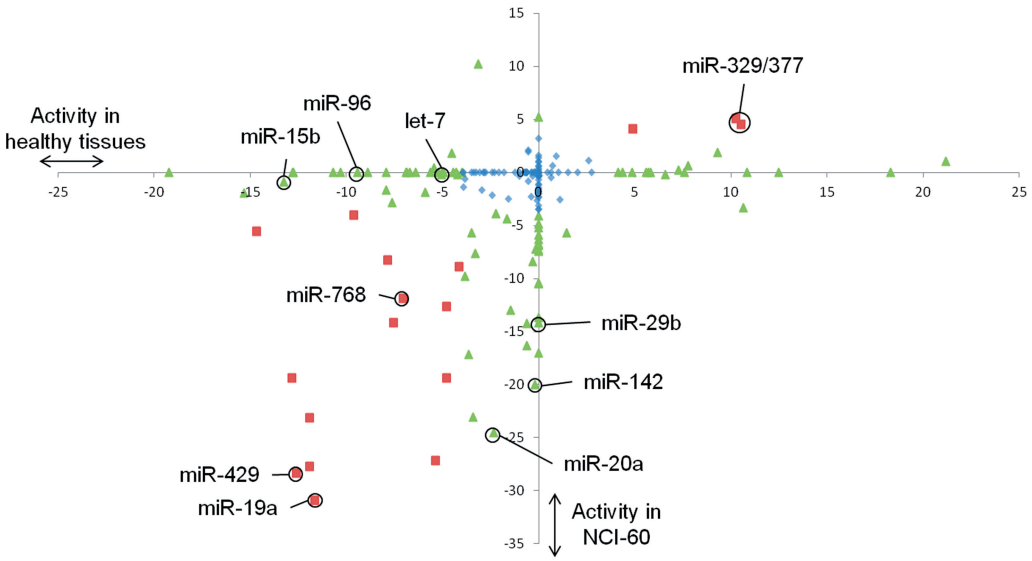

Figure 5. miRNA activity scores in healthy and cancer data sets. miRNA activity scores [signed log(mmHG *P*-value)]) are plotted for each miRNA. *X*-axis represents the activity scores of the miRNAs in the healthy tissue type data set where the *Y*-axis represents the activity scores in the NCI-60 data set. Marked in green are miRNAs that were found to be directly active in only one of the data sets. Marked in red are the 14 miRNAs that were found to be directly active in both data sets (hypergeometric $P < 0.004$, for the intersection) and 3 miRNAs that were found to be positively correlated with their targets in both data sets (hypergeometric $P < 0.004$, for the intersection).

we ran the above workflow on a matched integrative mRNA–miRNA data set of 100 primary breast cancer samples (26). Generating the miRNA activity map for this cohort, we found eight miRNAs from five miRNA families to be *DA-miRs*, expanding the findings reported in Enerly *et al.* (26) (Supplementary Table S6). We compared these results to miRNA high-throughput transfection assays that measured cell proliferation in MCF7 breast cancer cell lines (26). We found that all transfected *DA-miRs* play a direct role in cell proliferation in the cell lines (Supplementary Figure S3).

DISCUSSION

In this study, we introduce miTEA—a framework for miRNA target enrichment analysis. miTEA takes as input a ranked list of genes and then finds miRNAs of which targets are enriched in the top of the input list. miTEA uses a novel statistical analysis that takes into account the rich information available from high-throughput experiments as well as from the different miTPAs. We show how miTEA can be applied to detect miRNA activity in different experiments and shed light into miRNA activity map in healthy and cancer samples. miTEA is a web-based tool allowing the community an easy and efficient free access.

Most results obtained from recent high-throughput measurement technologies are naturally given as ranked lists of genes rather than as fixed sets of genes. The question of mutual enrichment in two ranked lists of genes is thus highly relevant to the analysis of such data sets. The mmHG approach focuses on commonalities in the top ends of the two analyzed lists. Statistical properties of such commonalities are not adequately addressed by other models.

To assess the accuracy and robustness of the miTEA approach, we first applied it to study well-designed and controlled experiments where mRNA and protein expression profiling were performed following over-expression of specific miRNAs. Using Targetscan as the underlying miTPA, miTEA specifically detected the activity of the over-expressed miRNA in 20 out of the 21 tested cases with high significance levels. Among the various miTPAs used in this comparison, Targetscan yielded the most accurate and robust result, even when compared to the version of Targetscan with only conserved sites and thus was used for further analysis in this study. We note that this comparison is limited to the particular data sets analyzed in our investigation. The strong target enrichment results obtained in the mRNA expression experiments shed light on the mode of mRNA regulation by miRNAs. It points to mRNA degradation rather than translation inhibition, which is consistent with several recent studies (24,50), but also justifies the use of miTEA to detect miRNA activity using mRNA expression profiles, as we describe in this article.

We next utilized miTEA to study a data set of healthy human tissue samples and confirmed a direct activity for known tissue-specific miRNAs (e.g. miR-122 in liver). We note that not all miRNAs with tissue specific expression are found to be active (e.g. miR-215 and liver, Supplementary Figure S2b), supporting the need for a robust enrichment analysis. We also observed a direct activity for miR-519a in placenta (mmHG $P < 3 \times 10^{-5}$, Supplementary Figure S2a). miR-519a was previously shown to be specifically expressed in placenta alongside with numerous other miRNAs (42,43). Our finding is, to

our knowledge, the first to show a direct regulatory involvement, by means of effect on targets, of this miRNA in placenta. This direct activity observed for miR-519a was not shared with the other reported miRNAs that are specifically expressed in placenta, further emphasizing the need for a joint analysis approach rather than only miRNA expression analysis. miR-519a is known to target the developmental related gene HuR (51) and further research is needed to better understand the relation of miR-519a to the intricate placental transcriptional programs.

We introduce a new approach to analyzing integrated miRNA/mRNA data, based on miTEA. Applying this approach to a matched data set of nine healthy human tissue types, we found 71 significant *DA-miRs* originating from 30 unique miRNA families. This result provides the first comprehensive miRNA activity map for these human tissue types. Moreover, the high prevalence of observed directly active miRNAs corroborates the involvement of miRNAs in differentiation and development. In addition, our results show that a high overall expression level of a miRNA does not necessarily indicate a significant regulatory effect on its targets (e.g. miR-125b or let-7a, Supplementary Figure S4). This is also the case for miRNAs that have high-variable expression levels (e.g. miR-205, Supplementary Figure S4). Although lack of miRNA activity could be a consequence of missing target information provided by the miTPA, this result indicate that high activity level of a miRNA is not a direct corollary of it being highly or variably expressed in the samples.

Interestingly, we observed that most of the *DA-miRs* found in our data set have been previously implicated in cancer. This includes well-established oncomirs such as miR-15 (45) (Supplementary Figure S2c) and the miR-17/92 cluster (47). We would expect that disregulation of active miRNAs may lead to a compromise in the control of transcriptional programs driving differentiation, which might eventually lead to cancer pathogenesis. We note that not all *DA-miRs* are known cancer-related miRNAs and further examination is therefore necessary to see whether those *DA-miRs*, such as miR-512-3p and miR-768 (Supplementary Figure S2d), contribute to the development of cancer when disrupted.

We further applied our workflow to a matched miRNA/mRNA data set publicly available for the NCI-60 cell lines data set. The resulting miRNA activity map consists of 47 cancer *DA-miRs* originating from 27 unique miRNA families. The cancer miRNA activity map is of similar size to the healthy tissue type miRNA activity map and, as expected, the two maps share a significant number of *DA-miRs*. The shared *DA-miRs* include known oncomirs such as miR-19a (47) (Supplementary Figure S5) and miR-429 (46) as well as less characterized miRNAs such as miR-768. In addition, we find 30 miRNAs that are *DA-miRs* in the NCI-60 data set only and 28 that are *DA-miRs* in the healthy data set only. This discrepancy could arise from the fact that while both data sets span nine different tissue types, only three tissue types are common. Another explanation may relate to specific functions of the miRNAs that are not related to cancer or to tissue differentiation. Tumour suppression is one such attribute that can explain the cancer-specific direct activity of miRNAs. Indeed many of the miRNAs which we found to be specifically directly active in the NCI-60 data set were shown to be tumour suppressors in various cancer types [e.g. miR-29b/c (52,53), miR-142 (54,55) and miR-101 (56,57)]. It would be interesting to further examine other such miRNAs for their potential role as tumour suppressors (e.g. miR-30e, miR-203). Specific direct activity of miRNAs in the healthy data set can be attributed to specific tissue type regulation. Our results also show that there is a subset of miRNAs that have a significant positive correlation with their targets (19 and 5 in the healthy and cancer cohorts, respectively). It was previously shown that miRNAs can up-regulate translation of their targets in quiescent cells (58) which is in line with our observation in the healthy cohort as compared to the cancer cohort. Positive correlation between miRNAs and their targets and even more so, miRNA roles as activators, has weak support in existing literature. It is also possible that these results represent false positive results in these cohorts. Therefore, more work is required to put our positive correlation results in broader perspective.

miR-768 is located within the sequence of a known snoRNA—HBII-239. Since the snoRNA has a better evolutionary support, miR-768 annotation was discarded from miRBase (1). It is thus not clear whether miR-768 functions as a miRNA. Recent studies have shown that snoRNA-like miRNAs should be examined with care (59). Specifically, an indication for miR-768 expression was observed in HeLa cells (60). In our study, we used miR-768 predicted targets, taken from Targetscan (V5.2) and compared their expression profiles to the measured expression profile of miR-768. Our results clearly show its predicted targets to be significantly anti-correlated to its measured expression profile across the cohort of samples in both data sets of healthy human tissues and cancer cell lines. Thus, we provide evidence not only for its existence but also for its regulatory activity in the two independent sample cohorts.

We also analyzed a homogenous integrated mRNA/miRNA data sets of 100 breast cancer samples. Interestingly, we detected eight *DA-miRs* in spite of the homogeneity of the data set. To validate direct activity, we compared these results to miRNA transfection assays that measured cell proliferation in MCF7 breast cancer cell lines. The cell lines showed elevated or reduced levels of proliferation when transfected with these breast cancer *DA-miRs*, in agreement with their observed direct activity in the breast cancer cohort (26). This homogenous data context demonstrates miTEA's high sensitivity in detecting miRNA activity.

miTEA's novel statistical approach, called mmHG, enables the detection of miRNA activity without predefining the set of genes targeted by the miRNA nor the set of genes that are over/under expressed in the expression experiment. In this respect, miTEA expands and extends the mHG approach (26,28,30) and uses two ranked lists of genes. The statistical approach used by miTEA finds the two thresholds in the top of the two

ranked lists of genes that maximize the mutual enrichment and thereby allows the data to define the set of affected genes in the experiment and the set of relevant miRNA targets. This approach is bound to be more sensitive than previous approaches that fix the threshold on the prediction score provided by the miTPA or define the miRNA targets according to a 6/7/8-seed match (20,21), as these approaches use a fixed set of genes to characterize the miRNA targets. Similarly, using regression or a simple correlation approach to compare two ranked lists of genes may result in such mutual enrichment being obscured by the majority of the genes further down the lists (see Supplementary Table S1). This notion is exemplified in Figure 1, where fixing a threshold to either define the down-regulated genes in the experiment or to define the genes targeted by the perturbed miRNA in the experiment might result in loss of observed significance. Moreover, a pair of thresholds that optimizes enrichment in one configuration will not be optimal in a different miRNA and expression configuration. Thus, the different thresholds selected in each of the experiments tested much improve the enrichment analysis results, supporting the need for a statistical tool that enables flexible threshold selection. Indeed, when compared to alternative online tools miTEA was able to obtain results with higher specificity and sensitivity in both simulated data sets as well in a miRNA perturbation data set (Figure 3).

The threshold optimization process obviously introduces statistical multiple testing. We address multiple testing correction through a combination of an exact approach to mHG (28) and use of a union bound. We provide a rigorous bound on the *P*-value of the resulting enrichment. Further characterization of the distribution of the mmHG statistics under the null model (a uniform measure over S_N—the group of all permutations) will significantly contribute to improved analysis of mutual enrichment in two ranked lists.

To conclude, the findings presented in this study represent the tip of the iceberg, in terms of detecting miRNA direct activity in health and disease. Making miTEA accessible to the community, as a user-friendly and efficient web application, we anticipate it will be used more broadly to analyze high-throughput molecular data. This will hopefully expand the scope of experimental results and finings, as related to the role of miRNA in health and disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–6 and Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

We thank Limor Leibovich (Technion) for useful discussion of the mmHG methodology. We thank the anonymous reviewers for their constructive comments.

FUNDING

EU FP7 project GlycoHIT (to I.S., in part). Funding for open access charge: Agilent Technologies.

Conflict of interest statement. None declared.

REFERENCES

1. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2007) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
2. Goymer,P. (2008) Stem cells: MicroRNAs promote differentiation. *Nat. Rev. Cancer*, **8**, 245.
3. Babak,T., Zhang,W., Morris,Q., Blencowe,B.J. and Hughes,T.R. (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. *RNA*, **10**, 1813–1819.
4. Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H., Ferrando,A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
5. Volinia,S., Calin,G.A., Liu,C.-G., Ambs,S., Cimmino,A., Petrocca,F., Visone,R., Iorio,M., Roldo,C., Ferracin,M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA*, **103**, 2257–2261.
6. van Rooij,E., Sutherland,L.B., Qi,X., Richardson,J.A., Hill,J. and Olson,E.N. (2007) Control of stress-dependent cardiac growth and gene expression by a MicroRNA. *Science*, **316**, 575–579.
7. Brodersen,P. and Voinnet,O. (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nat. Rev. Mol. Cell. Biol.*, **10**, 141–148.
8. Leibovich,L., Mandel-Gutfreund,Y. and Yakhini,Z. (2010) A structural-based statistical approach suggests a cooperative activity of PUM1 and miR-410 in human 3′-untranslated regions. *Silence*, **1**, 17.
9. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
10. Sood,P., Krek,A., Zavolan,M., Macino,G. and Rajewsky,N. (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA*, **103**, 2746–2751.
11. Cheng,C. and Li,L.M. Inferring MicroRNA activities by combining gene expression with MicroRNA target prediction. *PLoS One*, **3**, e1989.
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. Madden,S.F., Carpenter,S.B., Jeffery,I.B., Björkbacka,H., Fitzgerald,K.A., O'Neill,L.A. and Higgins,D.G. (2010) Detecting microRNA activity from gene expression data. *BMC Bioinformatics*, **11**, 257.
14. Liang,Z., Zhou,H., He,Z., Zheng,H. and Wu,J. (2011) mirAct: a web tool for evaluating microRNA activity based on gene expression data. *Nucleic Acids Res.*, **39**, W139–W144.
15. Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Simmosis,V.A., Zhang,L. and Hatzigeorgiou,A.G. (2010) The DIANA-mirExTra web server: from gene expression data to microRNA function. *PLoS One*, **5**, e9171.
16. Krek,A., Grun,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C., Stoffel,M. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
17. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
18. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting

genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
19. 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.
20. Farh,K.K.-H., 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.
21. 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.
22. Biggin,M.D. (2011) Animal transcription networks as highly connected, quantitative continua. *Dev. Cell*, **21**, 611–626.
23. 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.
24. Baek,D., Villén,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.
25. Shankavaram,U.T., Varma,S., Kane,D., Sunshine,M., Chary,K.K., Reinhold,W.C., Pommier,Y. and Weinstein,J.N. CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics*, **10**, 277–277.
26. Enerly,E., Steinfeld,I., Kleivi,K., Leivonen,S.-K., Aure,M.R., Russnes,H.G., Rønneberg,J.A., Johnsen,H., Navon,R., Rødland,E. *et al.* (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One*, **6**, e16915.
27. Ach,R., Wang,H. and Curry,B. (2008) Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. *BMC Biotechnol.*, **8**, 69.
28. Eden,E., Lipson,D., Yogev,S. and Yakhini,Z. (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.*, **3**, e39.
29. Steinfeld,I., Navon,R., Ardigò,D., Zavaroni,I. and Yakhini,Z. (2008) Clinically driven semi-supervised class discovery in gene expression data. *Bioinformatics*, **24**, i90–i97.
30. Eden,E., Navon,R., Steinfeld,I., Lipson,D. and Yakhini,Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48.
31. Straussman,R., Nejman,D., Roberts,D., Steinfeld,I., Blum,B., Benvenisty,N., Simon,I., Yakhini,Z. and Cedar,H. (2009) Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.*, **16**, 564–571.
32. Avraham,R., Sas-Chen,A., Manor,O., Steinfeld,I., Shalgi,R., Tarcic,G., Bossel,N., Zeisel,A., Amit,I., Zwang,Y. *et al.* (2010) EGF decreases the abundance of microRNAs that restrain oncogenic transcription factors. *Sci. Signal.*, **3**, ra43.
33. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
34. Maragkakis,M., Alexiou,P., Papadopoulos,G.L., Reczko,M., Dalamagas,T., Giannopoulos,G., Goumas,G., Koukis,E., Kourtis,K., Simossis,V.A. *et al.* (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
35. 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.
36. 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.
37. Miranda,K.C., Huynh,T., Tay,Y., Ang,Y.-S., Tam,W.-L., Thomson,A.M., Lim,B. and Rigoutsos,I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
38. Makeyev,E.V., Zhang,J., Carrasco,M.A. and Maniatis,T. (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell*, **27**, 435–448.
39. Chang,J., Nicolas,E., Marks,D., Sander,C., Lerro,A., Buendia,M.A., Xu,C., Mason,W.S., Moloshok,T., Bort,R. *et al.* (2004) miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol.*, **1**, 106–113.
40. Chen,J.-F., Mandel,E.M., Thomson,J.M., Wu,Q., Callis,T.E., Hammond,S.M., Conlon,F.L. and Wang,D.-Z. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.*, **38**, 228–233.
41. Kim,H.K., Lee,Y.S., Sivaprasad,U., Malhotra,A. and Dutta,A. (2006) Muscle-specific microRNA miR-206 promotes muscle differentiation. *J. Cell Biol.*, **174**, 677–687.
42. Liang,Y., Ridzon,D., Wong,L. and Chen,C. (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, 166.
43. Kotlabova,K., Doucha,J. and Hromadnikova,I. (2011) Placental-specific microRNA in maternal circulation—identification of appropriate pregnancy-associated microRNAs with diagnostic potential. *J. Reprod. Immunol.*, **89**, 185–191.
44. Navon,R., Wang,H., Steinfeld,I., Tsalenko,A., Ben-Dor,A. and Yakhini,Z. (2009) Novel rank-based statistical methods reveal MicroRNAs with differential expression in multiple cancer types. *PLoS One*, **4**, e8003.
45. Aqeilan,R.I., Calin,G.A. and Croce,C.M. (2010) miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ.*, **17**, 215–220.
46. 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.
47. Mendell,J.T. (2008) miRiad roles for the miR-17-92 cluster in development and disease. *Cell*, **133**, 217–222.
48. Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, **6**, 813–823.
49. Liu,H., D'Andrade,P., Fulmer-Smentek,S., Lorenzi,P., Kohn,K.W., Weinstein,J.N., Pommier,Y. and Reinhold,W.C. (2010) mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. *Mol. Cancer Ther.*, **9**, 1080–1091.
50. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
51. Abdelmohsen,K., Kim,M.M., Srikantan,S., Mercken,E.M., Brennan,S.E., Wilson,G.M., de Cabo,R. and Gorospe,M. (2010) miR-519 suppresses tumor growth by reducing HuR levels. *Cell Cycle*, **9**, 1354–1359.
52. Garzon,R., Heaphy,C.E.A., Havelange,V., Fabbri,M., Volinia,S., Tsao,T., Zanesi,N., Kornblau,S.M., Marcucci,G., Calin,G.A. *et al.* (2009) MicroRNA 29b functions in acute myeloid leukemia. *Blood*, **114**, 5331–5341.
53. Pass,H.I., Goparaju,C., Ivanov,S., Donington,J., Carbone,M., Hoshen,M., Cohen,D., Chajut,A., Rosenwald,S., Dan,H. *et al.* (2010) hsa-miR-29c* is linked to the prognosis of malignant pleural mesothelioma. *Cancer Res.*, **70**, 1916–1924.
54. Sun,W., Shen,W., Yang,S., Hu,F., Li,H. and Zhu,T.-H. (2010) miR-223 and miR-142 attenuate hematopoietic cell proliferation, and miR-223 positively regulates miR-142 through LMO2 isoforms and CEBP-[beta]. *Cell Res.*, **20**, 1158–1169.
55. Liu,X., Sempere,L.F., Galimberti,F., Freemantle,S.J., Black,C., Dragnev,K.H., Ma,Y., Fiering,S., Memoli,V., Li,H. *et al.* (2009) Uncovering growth-suppressive MicroRNAs in lung cancer. *Clin. Cancer Res.*, **15**, 1177–1183.
56. Hao,Y., Gu,X., Zhao,Y., Greene,S., Sha,W., Smoot,D.T., Califano,J., Wu,T.-C. and Pang,X. (2011) Enforced expression of miR-101 inhibits prostate cancer cell growth by modulating the COX-2 pathway in vivo. *Cancer Prev. Res.*, **4**, 1073–1083.
57. Friedman,J.M., Liang,G., Liu,C.-C., Wolff,E.M., Tsai,Y.C., Ye,W., Zhou,X. and Jones,P.A. (2009) The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res.*, **69**, 2623–2629.

58. Vasudevan,S., Tong,Y. and Steitz,J.A. (2007) Switching from repression to activation: MicroRNAs can up-regulate translation. *Science*, **318**, 1931–1934.
59. Ender,C., Krek,A., Friedländer,M.R., Beitzinger,M., Weinmann,L., Chen,W., Pfeffer,S., Rajewsky,N. and Meister,G. (2008) A human snoRNA with microRNA-like functions. *Mol. Cell*, **32**, 519–528.
60. Ono,M., Scott,M.S., Yamada,K., Avolio,F., Barton,G.J. and Lamond,A.I. (2011) Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Res.*, **39**, 3879–3891.