BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 1 2010, pages 91–97
doi:10.1093/bioinformatics/btp598

Systems biology

Identification of microRNA activity by Targets' Reverse EXpression

Stefano Volinia[1,2,3,*], Rosa Visone[3], Marco Galasso[1], Elda Rossi[4] and Carlo M. Croce[3]

[1]DAMA, Data Mining for Analysis of Microarrays, Department of Morphology and Embryology, University of Ferrara, Italy [2]Biomedical Informatics, [3]Comprehensive Cancer Center, Ohio State University, Columbus, OH, USA and [4]CINECA, Bologna, Italy

Received on June 25, 2009; revised on October 14, 2009; accepted on October 15, 2009

Advance Access publication November 6, 2009

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: Non-coding microRNAs (miRNAs) act as regulators of global protein output. While their major effect is on protein levels of target genes, it has been proven that they also specifically impact on the messenger RNA level of targets. Prominent interest in miRNAs strongly motivates the need for increasing the options available to detect their cellular activity.

Results: We used the effect of miRNAs over their targets for the detection of miRNA activity using mRNAs expression profiles. Here we describe the method, called T-REX (from Targets' Reverse EXpression), compare it to other similar applications, show its effectiveness and apply it to build activity maps. We used six different target predictions from each of four algorithms: TargetScan, PicTar, DIANA-microT and DIANA Union. First, we proved the sensitivity and specificity of our technique in miRNA over-expression and knock-out animal models. Then, we used whole transcriptome data from acute myeloid leukemia to show that we could identify critical miRNAs in a real life, complex, clinically relevant dataset. Finally, we studied 66 different cellular conditions to confirm and extend the current knowledge on the role of miRNAs in cellular physiology and in cancer.

Availability: Software is available at http://aqua.unife.it and is free for all users with no login requirement.

Contact: s.volinia@unife.it

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Characterization of genes that control the timing of larval development in *Caenorhabditis elegans* revealed two small regulatory RNAs, *lin-4* and *let-7* (Reinhart *et al*., 2000). Soon thereafter, *lin-4* and *let-7* were reported to represent a new class of small RNAs named microRNAs (miRNAs) (Lagos-Quintana *et al*., 2001; Lau *et al*., 2001; Lee and Ambros, 2001). miRNAs have since been found in plants, green algae, viruses and animals (Griffiths-Jones *et al*., 2008). The number of identified miRNA genes in human now surpasses 1000 (Landgraf *et al*., 2007; Ruby *et al*., 2006, 2007). miRNAs are involved in a variety of biological processes including cell cycle regulation, differentiation, development, metabolism, neuronal patterning and aging (Bartel, 2009). Alterations in miRNA expression are also involved in the initiation, progression and metastasis of human tumors and the consequences are just starting to be understood (Spizzo *et al*., 2009).

The number of putative miRNA targets is thought to be >60% of the total human genes, as novel algorithms allowed to increase by nearly 3-fold the number of conserved miRNA target sites (Friedman *et al*., 2009).

In principle from miRNA targets regulation it should be possible to infer miRNA expression. Farh and colleagues (Farh *et al*., 2005) firstly showed that tissue-specific miRNA activities could be predicted by analyzing mRNA expression profiles combined with miRNA seed analysis (a surrogate for target prediction). They showed that site depletion due to miRNA activity occurs specifically in tissue types expressing the corresponding miRNA. To explore the specificity of depletion, they used a modified Kolmogorov–Smirnov (KS) test to determine whether the set of genes with sites in either human or mouse orthologs were expressed at lower levels than controls. This finding was confirmed by Sood and coworkers (Sood *et al*., 2006). But not until recently the observation could be quantified at the whole proteome and transcriptome level (Baek *et al*., 2008; Selbach *et al*., 2008). Both groups used quantitative mass spectrometry to measure the proteome response as function of mRNA activity and showed that mRNA destabilization was a major component of miRNA activity. Furthermore, both groups massively tested different prediction algorithms for miRNA targets. According to Selbach's article DIANA-microT was found to be the most specific method (Maragkakis *et al*., 2009), with PicTar (Krek *et al*., 2005) and TargetScan (Lewis *et al*., 2005) close seconds.

We devised, validated and applied a technique for the generation of miRNA activity networks using messenger RNA profiles, therefore named Targets' Reverse Expression (T-REX). After Fahr (Farh *et al*., 2005), other groups have developed similar techniques (Arora and Simpson, 2008; Cheng and Li, 2008), considering mean absolute target gene expression, rank sum tests or 'ranked ratios'. Arora and Simpson confirmed that many miRNAs with reduced target gene expression corresponded to those known to be expressed in 8 cognate normal tissues. They validated their approach in two *in-vitro* experiment: inhibition of miR-122 by an antagomir (Krutzfeldt *et al*., 2005) and miR-124 transfection. Cheng and Li (2008) proposed the AC score, a generalization of the enrichment score in

*To whom correspondence should be addressed.

gene set enrichment analysis, or GSEA, (Subramanian *et al*., 2005). They again applied it to two *in-vitro* models: transfection of wild type and mutant miR-1 and miR-124. Huang *et al*. (2007) demonstrated that paired expression profiles of miRNAs and mRNAs can be used to identify functional miRNA-target relationships. They used a Bayesian data analysis algorithm, GenMiR++, to identify a network of 1597 high-confidence target predictions for 104 human miRNAs, which was supported by RNA expression data across 88 tissues and cell types. Compared to sequence-based predictions, GenMiR++ predictions were more accurate predictors for let-7b levels. Recently, a group used anti-correlation between expression of miRNA host genes and their putative targets (Gennarino *et al*., 2009). This last method can be used only for those miRNAs for which host genes are known. Finally, DIANA-mirExTra, available online (Maragkakis *et al*., 2009), considers a lists of regulated and one of background genes, but it has not been described further yet.

Using T-REX we have extensively screened a range of different physiological and pathological conditions. One of the main values of this study lies in the practical application of T-REX to clinical data, which led to some interesting novel observations. Among them, miRNA action in the control of lag log phase in mesenchymal stem cell growth, p53/TGF-beta pathways, chronic lymphocytic leukemia, acute myeloid leukemia and cancer metastasis.

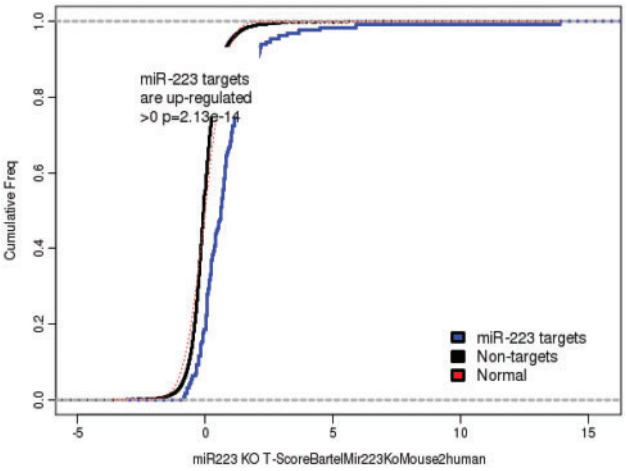

Fig. 1. Lack of miR-223 activity in miR-223 KO mice. The cumulative distribution function (ECDF) plot of the KS test correctly identifies target coding genes specifically controlled by miR-223. *t*-test values from miRNA KO cells versus control cells were analyzed. miR-223 was detected as the miRNA with the most up-regulated targets byT-REX. The blue curve (miR-223 target genes), on the right and below the black non-targets curve, indicates up-regulation of target messenger RNAs. The red dotted curve is that expected by random association and follows tightly that one of the non-target controls.

2 METHODS

2.1 Expression datasets

Expression profiles for messenger RNA were obtained by human or mouse genome wide Affymetrix assays. GDS datasets were directly imported from GEO (Barrett *et al*., 2007) by using BRB-ArrayTools (Simon *et al*., 2007), while for GSE records the CEL files were imported and RMA normalized. Genes whose expression differed by at least 1.5-fold from the median in at least 20% of the arrays were used. Values relative to each coding gene, for input to T-REX, were obtained by using the appropriate statistical test (for example, random-variance *t*-test for a two classes experiment or Spearman's correlation for quantitative traits). The random-variance *t*-test is an improvement over the standard separate *t*-test as it permits sharing information among genes about within-class variation without assuming that all genes have the same variance (Wright and Simon, 2003). Two classes profiles were subject to miRNA analysis only when the prediction rate of the profiles was >85% (the BRB Array Tools classifier was used). For quantitative traits (i.e. time, dosage, age), we tested the hypothesis that the Spearman's correlation between gene expression and the trait was zero. The tests *P*-values were used in a multivariate permutation test (Korn *et al*., 2007) in which the traits were randomly permuted among arrays. The multivariate permutation test was used to provide 90% confidence that the false discovery rate was <10%.

2.2 Targets' Reverse Expression (T-REX)

We used four algorithms and six target predictions: PicTar 5-ways (43 079 predictions for 130 miRNAs in 2513 target mRNAs) and PicTar 4-ways (75 968 predictions for 178 miRNAs in 9152 target mRNAs) were obtained from UCSC; TargetScan 5.1 [1 243 782 (55 028 conserved) predictions for 669 miRNAs (153 conserved miRNA families) in 17 689 (9448 conserved) target mRNAs]; DIANA-microT 3.0 (38 859 predictions for 165 miRNAs in 5100 target mRNAs) and DIANA-Union (472 428 predictions for 494 miRNAs in 18 362 target mRNAs). The T-REX results are reported for each one of the different predictions, thus the user can choose the preferred one(s). MicroT identifies miRNA targets by combining TargetScan and PicTar predictions (Maragkakis *et al*., 2009) and has been shown to be the most selective algorithm in a proteome scan (Selbach *et al*., 2008), thus we used it as the preferred one for the results reported in this paper. For each different miRNA the KS test was applied to the predicted target and non-target coding genes. The KS statistic is the maximum difference between the empirical cumulative distribution function (ECDF) of the two target and non-target distributions. The KS plots report the ECDF curves (Fig. 1, Supplementary Fig. 1A and C) and the box-plots (Supplementary Fig. 1B and D) for each significant miRNA. The Benjamini and Hochberg correction for multiple testing was applied to the miRNA *P*-values. Forty-six out of 66 experiments yielded at least one miRNA from T-REX. DAVID (EASE) was used for the functional annotation and statistical evaluation of GO terms associated to the regulated miRNA targets (Huang da *et al*., 2009). The T-REX results can be accessed at http://aqua.unife.it/T-REX.

2.3 Network clustering

Clustering algorithms are often used in biology in order to extract coherent groups of nodes from expression networks. Here we used a circular layout to portrait interconnected ring and star topologies. Circular layouts emphasize group and tree structures within a network. The graph-based clustering algorithms MCL (Enright *et al*., 2002) has been shown to enable good performances in extracting co-regulated genes from transcriptome networks (Brohee and van Helden, 2006). We applied it to detail the network of the most active miRNAs identified by T-REX in 66 different datasets.

3 RESULTS

3.1 Rationale and Implementation

Recent reports (Baek *et al*., 2008; Selbach *et al*., 2008) clearly indicated that coding genes target of miRNAs are specifically repressed both at protein and at messenger RNA level. Therefore, we developed an algorithm, named T-REX, to detect such targets' repression at mRNA level and to infer the specific controller

miRNA/s. The rationale was solely based on the fact that when a miRNA is active in the cell, its target mRNAs are repressed. Conversely, if the targets are over-expressed then the controlling miRNA is down-regulated.

First, a statistical value was obtained for the coding mRNAs in the dataset/experiment. For example, in a two-class experiment (i.e. treated versus non-treated cells, or disease versus control) we used the *t*-value for each measured mRNA (routinely between 5000 and 10 000). Alternatively, for a quantitative analysis (i.e. time course or dose-response) we used the Spearman's rank correlation coefficient. Then we applied the KS test to the respective values of target and non-target genes for each miRNA in each prediction algorithm.

3.2 Validation

We needed to prove that the T-REX algorithm could correctly detect the genome wide repression or de-repression of target messenger RNAs. Therefore we analyzed a number of expression profiles, where the perturbation was due to over-expression or absence of a specific miRNA. Firstly, we used two datasets from transfections of miR-124, in HepG2 and neuroblastoma (Supplementary Fig. 1A). In both experiments miR-124 was correctly detected as the miRNA with the most down-regulated targets. The blue curve (miR-124 target genes predicted by microT), on the left and above the black non-targets curve, indicated specific down-regulation of target messenger RNAs. Similarly, the box-plot (Supplementary Fig. 1B) showed an excess of repressed miR-124 targets, indicated by a *t*-value distribution in targets skewed towards negative values. The neuroblastoma data showed significant miR-124 activation with all the prediction algorithms (TargetScan, microT and Diana Union), but PicTar, for which no significant miRNA was identified (see the web site). Transfections were relatively ease for T-REX: for example, all the six predictors correctly identified miR-29 upon transfection of K562 cell lines (see online Results). Furthermore, no miRNAs were identified with any of the predictors in a negative control: Luciferase versus mock transfections (Supplementary Table 1). The transfections of a miRNA in cell cultures still represented controlled *in vitro* experiment, far from the complex physio-pathological *in vivo* conditions we were interested to unravel. Thus, we tested T-REX on a more relevant model, a miRNA knock-out (KO) mouse. Figure 1 describes the results for such a miR-223 KO mouse model (Baek *et al*., 2008).

Again our method proved to be both selective and specific: following the loss of miR-223 activity, its targets were the most up-regulated genes (Supplementary Table 2). The blue curve (miR-223 target genes), on the right and below the black non-targets curve, indicated up-regulation of target messenger RNAs. The relative box-plot (Supplementary Fig. 1D) also showed an excess of up-regulated miR-223 targets, with a *t*-value distribution skewed towards positive values.

As an additional element of validation, we tested liver cells from Dicer KO mice. In these cells the machinery for the production of mature miRNA is deficient and one would expect a general loss of miRNA activity. In complete agreement with this hypothesis, while the mRNA profile was balanced (comprising roughly an equal number of up- and down-regulated coding genes, ~3000, $P < 0.05$), we identified 15 losses and no gains of miRNA activity (Chi square, $P < 0.001$, Supplementary Table 3). Thus, all four tests on controlled experiments showed a robust performance of T-REX.

Since we were ultimately interested in deciphering the miRNome regulation in complex and clinically relevant samples, passing those controlled experiment was *conditio sine qua non* but not sufficient. Therefore we proceeded to validate T-REX by querying a 'real life' experiment, where conditions were not pre-determined: the overall survival in acute myeloid leukemia (AML). As a statistical value we used the $\log_2$ of hazard ratios derived from Cox regression. In Supplementary Table 4 we show the results of studying the miRNA activity associated to patients' overall survival in acute myeloid leukemia. We performed the KS test on the $\log_2$ of the hazard ratios derived from Cox regression. miR-181, miR-155 and miR-10 (Garzon *et al*., 2008; Marcucci *et al*., 2008; O'Connell *et al*., 2008) were correctly identified among the microRNAs with most significant association to patients' survival. In fact, miR-155 and miR-181 (Marcucci *et al*., 2008) and miR-10 (Garzon *et al*., 2008) were confirmed by miRNA chips and RT-PCR. It has to be noted that the queried AML dataset was unrelated to the confirmation AML datasets, underlining the sensitivity of T-REX.

3.3 The miRNA activity map

We applied T-REX to 66 different human or mouse studies to generate a map of miRNA activity. Forty-six out of 66 experiments (69.6%) yielded at least one significant miRNA (adjusted $P < 0.05$). The miRNAs and the associated cellular conditions are listed in Supplementary Table 5. The miRNA-cellular conditions networks for either activated or repressed miRNAs are shown in Figures 2

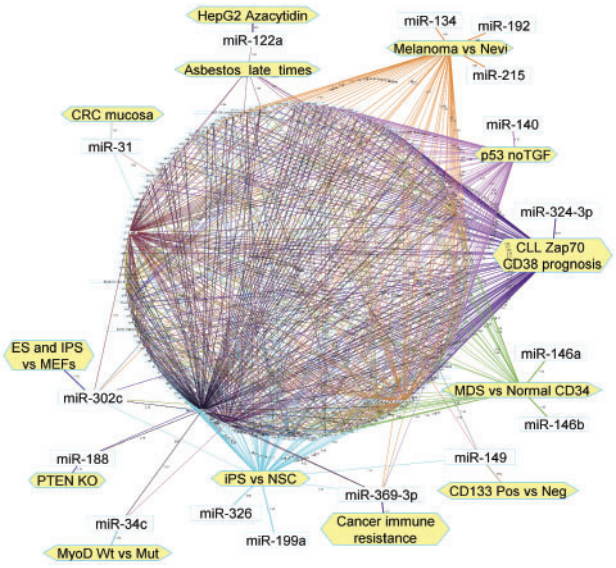

Fig. 2. The network of activated miRNAs in 35 different cellular conditions (868 edges, adjusted $P < 0.05$). Layout style is Circular (BCC isolated). Each edge color indicates a different cellular condition. External nodes were rearranged for clarity in the figure. Yellow hexagonal nodes represent the cellular conditions. This network representation emphasizes miRNAs which are associated to one cellular condition, or vice-versa. Abbreviations: CRC, colorectal carcinoma; ES, embryonic stem cells; IPS, induced pluripotent stem cells; MEFs, mouse embryonic fibroblasts; NSC, neural stem cells; TGF, tumor growth factor-beta; MDS, Myelodysplasia; CLL, chronic lymphocytic leukemia; Wt, wild-type; Mut, mutated.

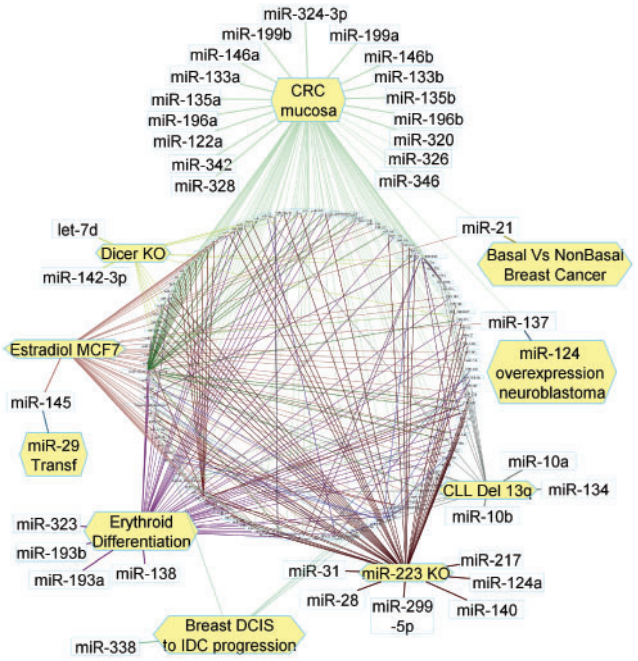

Fig. 3. The network of miRNAs with loss of activity in 24 different cellular conditions (418 edges, adjusted $P < 0.05$). Layout style is Circular (BCC isolated). Each edge color indicates a different cellular condition. Abbreviations: IDC, invasive/infiltrating ductal carcinoma; DCIS, ductal carcinoma *in situ*.

and 3, respectively. Figure 2 shows the network for miRNAs with gain of function in 35 different cellular conditions. Conversely, in Figure 3, is reported the network of miRNAs with loss of activity in 24 different cellular conditions.

To fine tune the selective threshold for activity, we investigated the frequency distribution of all significant miRNAs from T-REX (Supplementary Fig. 2). The resulting curve indicated a change in slope for $P < 0.0003$, due to an excess of highest scoring miRNAs. When using this P-value as a threshold, the controls still behaved as expected, i.e. miR-124 was identified in miR-124 transfections (in both the neuroblastoma and the HepG2 datasets). Thus we applied this very stringent threshold to select 176 *bona fide* modulated miRNAs across the 66 experiments.

Besides the results of the miR-124 and miR-223 controls, we detected an additional number of confirmatory signals in the T-REX miRNA map. For example, the members of the miR-302 cluster (a/b/c/d) were expressed in pluripotent hES (Landgraf *et al*., 2007) and iPS cells (Wilson *et al*., 2009) and repressed during erythroid differentiation. miR-34b was active in p53 experiment, in agreement to its known regulation (Chang *et al*., 2007; Corney *et al*., 2007; He *et al*., 2007). For a visual representation of the selected miRNAs, and their respective cellular conditions, we generated a miRNAs/conditions network (Fig. 4). Clusters of co-regulated miRNAs were detected by using MCL (Brohee *et al*., 2008) and were represented by edges of different colors. The functionally related experiments on pluripotent stem cells (iPS and hES) clustered together as expected.

miR-375 was active in lag–log phase of mesenchymal stem cells (MSCs), and miR-372 in induction of pluripotent iPS from neuronal stem cells (Voorhoeve *et al*., 2006). miR-96 was active in both iPS cells and lag–log phase of MSCs. miR-29 family members were induced in lag–log phase of MSCs and repressed in breast cancer progression and early colorectal cancer (CRC) mucosa (Fabbri *et al*., 2007; Mott *et al*., 2007). miR-27a/b (Mertens-Talcott *et al*., 2007) were activated in p53 experiment (irrespective of TGF-beta) and during cancer progression (melanoma versus benign nevi). miR-181a/b/c were repressed in early CRC mucosa and correlated to Zap 70 prognosis in CLL (Calin *et al*., 2005). miR-195 was activated by p53, positively correlated to Zap70 in CLL and repressed in early CRC mucosa (like miR-194). miR-17-5p was active in the p53 dataset (no TGF-beta) and repressed during erythroid differentiation. The miR-15/16 cluster was active in p53, irrespective of TGF-beta, correlated with CLL prognosis (Calin *et al*., 2008) and was repressed in early CRC mucosa (Bonci *et al*., 2008). miR-137 was active in the progression from benign nevi to melanoma (Bemis *et al*., 2008). When clusters of co-regulated miRNAs were detected by using MCL (Brohee *et al*., 2008), an interesting cluster contained miR-26a/b which were active in the basal subtype of breast cancer and repressed upon estradiol treatment of MCF7 breast cancer cell line.

Additional intriguing results could be extracted from the 1110 remaining significant miRNAs with scores <3.50 (Supplementary Table 5). For example, the fastest recovering patients from trauma had active miR-302 and miR-372 (pluripotent stem cells), active miR-205 [epithelial to mesenchymal transition (Gregory *et al*., 2008)] and miR-221/222 [positive regulators of cell cycle (le Sage *et al*., 2007)]. We then analyzed the transcriptome-wide expression profiles of 20 pulmonary metastases of clear cell renal cell (CCRCC) carcinoma in order to identify miRNAs associated with two important prognostic factors: the disease-free interval after nephrectomy (DFI) and the number of metastases per patient. In CCRC four miRNAs were negatively associated to disease free intervals, i.e. highly expressed in aggressive tumors with short DFI: miR-339, miR-221/222 and miR-188. In the same patients, miR-129, miR-29a/b/c, miR-30-3p and miR-299-3p correlated to the number of metastases present in the lungs. Myelodysplasia expressed higher levels of miR-181a/b/c when compared to CD34 positive cells. miR-34 was down-regulated upon p63 depletion by shRNA. The effect of cigarette smoke on the large airways included the down-regulation of miR-1 and miR-206 (Kim *et al*., 2006). The lack of FoxO in the LSK cell population enriched for hematopoietic stem cells (HSCs) (Tothova *et al*., 2007) led to the down-regulation miR-155 and miR-30 family. Zfx is a zinc finger protein of the Zfy family, whose members are highly conserved in vertebrates and is involved in embryonic and adult stem cells (SCs) self-renewal (Galan-Caridad *et al*., 2007). Analysis of embryonic and hematopoietic SCs lacking the transcription factor Zfx showed that miR-29s were down-regulated.

A side-product of our miRNA activity mining procedure was the identification of the regulated targets in the corresponding mRNA profile (Supplementary Material). The list of the regulated gene targets can be used to perform a Gene Ontology (GO) analysis and identify the affected cellular functions. For example, for the metastatic renal cancer (CCRCC) Supplementary Table 6 indicates the 41 genes which are targets of miR-221/222 and are regulated in the tumor profile. The GO analysis revealed that the miR-221/222 targets were significantly enriched for genes involved in the regulation of progression through cell cycle. For the same

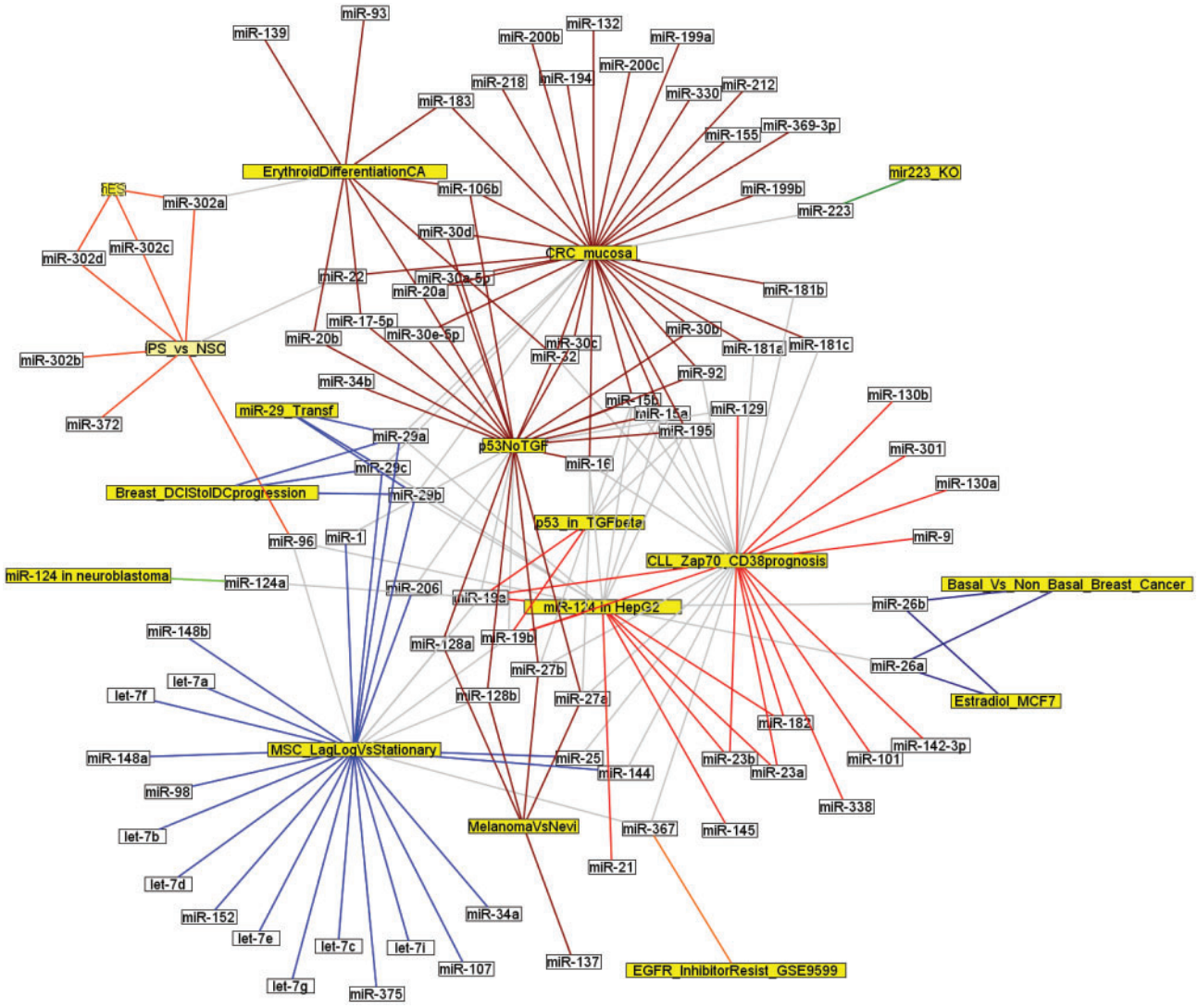

Fig. 4. The network of the 176 most significant miRNAs (adjusted $P < 0.0003$; score > 3.5) indentified from different cellular conditions. Out of 66 tested cellular conditions, 17 display highly significant miRNAs. The condition labels are colored in yellow. miRNA grouping was obtained by MCL clustering and is indicated by the colored edges. For example the orange connects correctly two different, but related, experiments: hES cells and iPS, both pluripotent human cell types, with activation of miR-302 cluster and miR-372. miR-26a/b are shared by two independent experiment on breast cancer: estradiol treatment of MCF7 cell line and difference between basal and non-basal subtypes. Abbreviations: MSC, mesenchymal stem cells. Additional details of the experiments are reported in the Supplementary Material.

disease and the miR-29 family members, the significant GO terms associated to the CCRCC profiles were extracellular matrix and focal adhesion (Supplementary Table 7).

4 DISCUSSION

We have devised, developed, implemented and validated, T-REX, a technique to infer the activity of miRNAs from expression profiles of messenger RNA. T-REX identifies regulated miRNAs via the modulation of their predicted targets. Predicted target lists might contain false positives and lack true positives. Nevertheless, we postulated that the large number of measured, although not validated, predicted targets for a miRNA should still result in efficient mining of its cellular activity. This hypothesis was confirmed by the positive performance on control experiments (two independent miR-124 over-expression experiments and the granulocytes from miR-223 KO mice). Additional confirmations of T-REX specificity were obtained from the Luciferase negative control (no regulated miRNAs, as expected) and from the conditional Dicer KO liver. While approximately the same number of up- and down-regulated coding mRNAs was present in the Dicer KO mice, only repressed, and no activated, miRNAs were detected. Such a highly unbalanced miRNA profile was indeed the expected one from Dicer KO cells, where the machinery for pre-miRNA processing was defective and thus no active/mature miRNAs should have been present.

Nevertheless passing controlled experiments was only a first validation step for T-REX. We further validated it on a real life experiment, investigating the miRNA activity associated to patients' survival in acute myeloid leukemia. The detection of miR-181 and miR-155 (Garzon *et al*., 2008; Marcucci *et al*., 2008; O'Connell *et al*., 2008) positively confirmed the detection power of T-REX in a complex, clinically relevant, dataset. It has to be noted here that the various prediction algorithms behave somehow differently. Only microT and TargetScan (ver 5.1, conserved sites) identified these two crucial miRNAs. Earlier TargetScan Ver 4.2 did not pick up miR-181, nor it did the TargetScan prediction based on the non-conserved sites. DIANA-Union did not detect miR-181, while PicTar 5ways did not detect miR-155. PicTar 4 ways did not detect any of the two miRNAs. The version of TargetScan with conserved sites contains around 50 000 predictions, while the one with the non conserved sites contains over 1 million. DIANA-Union is too composed of a very large number of predictions (472 428). Thus it appears that algorithms with a very large number of prediction sites are less reliable than those with fewer, possibly more reliable, sites. To conduct a systematic and rigorous evaluation of T-REX and other related applications large mRNA/miRNA datasets, run on homogeneous platforms, will need to be available.

Having proven the specificity and sensitivity of T-REX, we proceeded to reverse identify the miRNAs involved in more than 60 different physiological and pathological conditions, including cancers and leukemia (Supplementary Material). One hundred and seventy-six associations between miRNAs and cellular conditions were identified when using a very stringent threshold (Fig. 4). Some of our findings confirmed known activities, such as miR-302 in embryonic stem cells and miR-34 in the p53 pathway, while others revealed novel miRNAs/pathways associations, including control of lag-log phase in MSC growth, p53/TGF-beta pathways, CLL prognosis, cancer metastasis. In metastatic renal cancer (CCRCC) four miRNAs were highly expressed in aggressive tumors: miR-339, miR-221/222 and miR-188. Activation of miR-221/222 had been observed in other cancer types. In fact, miR-221/222 are cell cycle accelerators, since they work by controlling cell cycle inhibitors CDKN1C/p57 and CDKN1B/p27 (Fornari *et al*., 2008; Galardi *et al*., 2007; le Sage *et al*., 2007). No functions have yet been reported for the two other active miRNAs, miR-339 and miR-188. In the same renal cancer patients, miR-129, miR-29a/b/c, miR-30-3p and miR-299-3p were positively correlated to the number of lung metastases. Gebeshuber and colleagues (Gebeshuber *et al*., 2009) showed that over-expression of miR-29a led to epithelial to mesenchymal transition (EMT) and metastasis in cooperation with oncogenic Ras signaling. They also observed enhanced miR-29a in breast cancer patient samples. The roles of miR-129 and miR-299 in cancers have not been yet exposed. Myelodysplasia (MDS) expressed high levels of miR-181a/b/c when compared to CD34 positive cells, a novel finding as nothing is yet known about the role of miRNAs in MDS. miR-34s, which are downstream of p53 (He *et al*., 2007), appeared down-regulated upon p63 depletion by shRNA. The effect of cigarette smoke on the large airways included the down-regulation of miR-1 and miR-206, two genes induced during differentiation of C2C12 myoblasts *in vitro* (Kim *et al*., 2006). The lack of FoxO in the lineage-negative Sca-1+, c-Kit+ (LSK) cell population, enriched for hematopoietic stem cells (Tothova *et al*., 2007), led to the down-regulation of miR-155 and of the miR-30 family. Analysis of embryonic and HSCs lacking the transcription factor Zfx ,involved in embryonic and adult SCs self-renewal (Galan-Caridad *et al*., 2007), showed that the miR-29s were down-regulated when Zfx was absent.

In conclusion, the molecular dissection of miRNA activities, by using messenger RNA profiles and T-REX, yielded highly concise signatures from large mRNA experiments or patient cohorts, thus helping to unravel their functional comprehension and to escalate one level towards their molecular decoding. An exemplification of this feature was well represented by the trauma dataset. The samples from fastest recovering patients had active miR-302, miR-372 (both expressed in pluripotent stem cells) and miR-205 (epithelial to mesenchymal transition). The presence of these miRNAs strongly suggested a higher number of pluripotent or mesenchymal cells in the tissues which recovered faster from trauma.

ACKNOWLEDGEMENTS

Funding: AIRC (Italian Association for Cancer Research) and BioPharmaNet (Rete Alta Tecnologia dell' Emilia Romagna) grants (to S.V.).

Conflict of Interest: none declared.

REFERENCES

Arora,A. and Simpson,D.A. (2008) Individual mRNA expression profiles reveal the effects of specific microRNAs. *Genome Biol.*, **9**, R82.

Baek,D. *et al*. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.

Barrett,T. *et al*. (2007) NCBI GEO: mining tens of millions of expression profiles–database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.

Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

Bemis,L.T. *et al*. (2008) MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines. *Cancer Res.*, **68**, 1362–1368.

Bonci,D. *et al*. (2008) The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nature Med.*, **14**, 1271–1277.

Brohee,S. *et al*. (2008) Network Analysis Tools: from biological networks to clusters and pathways. *Nature Protocols*, **3**, 1616–1629.

Brohee,S. and van Helden,J. (2006) Evaluation of clustering algorithms for protein-protein interaction networks. *BMC bioinformatics*, **7**, 488.

Calin,G.A. *et al*. (2008) MiR-15a and miR-16-1 cluster functions in human leukemia. *Proc. Natl Acad. Sci. USA*, **105**, 5166–5171.

Calin,G.A. *et al*. (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *New Engl. J. Med.*, **353**, 1793–1801.

Chang,T.C. *et al*. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular Cell*, **26**, 745–752.

Cheng,C. and Li,L.M. (2008) Inferring microRNA activities by combining gene expression with microRNA target prediction. *PLoS ONE*, **3**, e1989.

Corney,D.C. *et al*. (2007) MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res.*, **67**, 8433–8438.

Enright,A.J. *et al*. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.

Fabbri,M. *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.

Farh,K.K. *et al*. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution, *Science*, **310**, 1817–1821.

Fornari,F. *et al*. (2008) MiR-221 controls CDKN1C/p57 and CDKN1B/p27 expression in human hepatocellular carcinoma, *Oncogene*, **27**, 5651–5661.

Friedman,R.C. *et al*. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Galan-Caridad,J.M. *et al*. (2007) Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell*, **129**, 345–357.

Galardi,S. *et al*. (2007) miR-221 and miR-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27Kip1. *J. Biol. Chem.*, **282**, 23716–23724.

Garzon,R. *et al*. (2008) Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proc. Natl Acad. Sci. USA*, **105**, 3945–3950.

Garzon,R. *et al*. (2008) MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*, **111**, 3183–3189.

Gebeshuber,C.A. *et al*. (2009) miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. *EMBO Rep.*, **10**, 400–405.

Gennarino,V.A. *et al*. (2009) MicroRNA target prediction by expression analysis of host genes. *Genome Res.*, **19**, 481–490.

Gregory,P.A. *et al*. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biol.*, **10**, 593–601.

Griffiths-Jones,S. *et al*. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.

He,L. *et al*. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130–1134.

Huang da,W. *et al*. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44–57.

Huang,J.C. *et al*. (2007) Using expression profiling data to identify human microRNA targets. *Nature Methods*, **4**, 1045–1049.

Kim,H.K. *et al*. (2006) Muscle-specific microRNA miR-206 promotes muscle differentiation. *J. Cell Biol.*, **174**, 677–687.

Korn,E.L. *et al*. (2007) An investigation of two multivariate permutation methods for controlling the false discovery proportion. *Stat. Med.*, **26**, 4428–4440.

Krek,A. *et al*. (2005) Combinatorial microRNA target predictions. *Nature Genetics*, **37**, 495–500.

Krutzfeldt,J. *et al*. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.

Lagos-Quintana,M. *et al*. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.

Landgraf,P. *et al*. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.

Lau,N.C. *et al*. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862.

le Sage,C. *et al*. (2007) Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation. *EMBO J.*, **26**, 3699–3708.

Lee,R.C. and Ambros,V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. *Science*, **294**, 862–864.

Lewis,B.P. *et al*. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Maragkakis,M. *et al*. (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.*, **37**, W273–W276.

Marcucci,G. *et al*. (2008) Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. *J. Clin. Oncol.*, **26**, 5078–5087.

Marcucci,G. *et al*. (2008) MicroRNA expression in cytogenetically normal acute myeloid leukemia. *New Engl. J. Med.*, **358**, 1919–1928.

Mertens-Talcott,S.U. *et al*. (2007) The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells. *Cancer Res.*, **67**, 11001–11011.

Mott,J.L. *et al*. (2007) mir-29 regulates Mcl-1 protein expression and apoptosis. *Oncogene*, **26**, 6133–6140.

O'Connell,R.M. *et al*. (2008) Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. *J. Exp. Med.*, **205**, 585–594.

Reinhart,B.J. *et al*. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, **403**, 901–906.

Ruby,J.G. *et al*. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C.elegans. *Cell*, **127**, 1193–1207.

Ruby,J.G. *et al*. (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.*, **17**, 1850–1864.

Selbach,M. *et al*. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.

Simon,R. *et al*. (2007) Analysis of gene expression data using BRB-array tools. *Cancer Inform.*, **3**, 11–17.

Sood,P. *et al*. (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA*, **103**, 2746–2751.

Spizzo,R. *et al*. (2009) SnapShot: MicroRNAs in cancer. *Cell*, **137**, 586–586.e1.

Subramanian,A. *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.

Tothova,Z. *et al*. (2007) FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell*, **128**, 325–339.

Voorhoeve,P.M. *et al*. (2006) A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*, **124**, 1169–1181.

Wilson,K.D. *et al*. (2009) MicroRNA profiling of human-induced pluripotent stem cells. *Stem Cells Dev.*, **18**, 749–758.

Wright,G.W. and Simon,R.M. (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, **19**, 2448–2455.