

Posttranscriptional Regulatory Networks: From Expression Profiling to Integrative Analysis of mRNA and MicroRNA Data

Swanhild U. Meyer, Katharina Stoecker, Steffen Sass,
Fabian J. Theis, and Michael W. Pfaffl

Abstract

Protein coding RNAs are posttranscriptionally regulated by microRNAs, a class of small noncoding RNAs. Insights in messenger RNA (mRNA) and microRNA (miRNA) regulatory interactions facilitate the understanding of fine-tuning of gene expression and might allow better estimation of protein synthesis. However, in silico predictions of mRNA–microRNA interactions do not take into account the specific transcriptomic status of the biological system and are biased by false positives. One possible solution to predict rather reliable mRNA–miRNA relations in the specific biological context is to integrate real mRNA and miRNA transcriptomic data as well as in silico target predictions. This chapter addresses the workflow and methods one can apply for expression profiling and the integrative analysis of mRNA and miRNA data, as well as how to analyze and interpret results, and how to build up models of posttranscriptional regulatory networks.

Key words mRNA, miRNA, Microarray, Multiple linear-regression, TaLasso, Pathway analysis, Quantitative real-time PCR, Gene ontology, R language, Genomatix pathway system

1 Introduction

Gene expression is regulated at the posttranscriptional level by small noncoding RNA species. One prominent class of small noncoding RNAs are microRNAs (miRNAs), which are 19–24 nt in length [1, 2]. miRNAs are transcribed by RNA polymerase II from independent genes or represent introns of messenger RNA (mRNA) transcripts [3, 4]. In the canonical miRNA biogenesis, processing of primary miRNAs to precursor miRNAs (~70 nt) is catalyzed by DROSHA in complex with dsRNA-binding proteins [3, 5]. Alternatively, precursor miRNAs can be generated by splicing and debranching of introns (mirtrons) [4, 6]. Precursor miRNAs are exported to the cytoplasm and are further processed by DICER-TRBP complex to form the miRNA duplex (~20 base pairs).

The miRNA strand, which is loaded into the miRNA-induced silencing complex leads to translational repression, destabilization, and degradation of target mRNAs [3, 7]. Target mRNAs are recognized by partial base pairing in the 3'-untranslated region [8, 9], within the protein coding sequence, or 5'-untranslated region [10, 11]. A predominant function of miRNAs is to negatively regulate gene expression by decreasing mRNA levels [12]. There is evidence of relatively few miRNAs having a positive effect on target gene expression in certain cellular conditions [13, 14].

Insights in mRNA and miRNA regulatory interactions facilitate the understanding of fine-tuning of gene expression and might allow better estimation of protein synthesis. As miRNA deregulation is a hallmark of several diseases [1] the understanding of miRNA–mRNA relations is of high interest from a scientific as well as medical [15] therapeutic point of view. Currently, computational prediction of miRNA targets is biased by a high false positive rate due to the short target site sequence. Moreover, certain mRNAs are not expressed or targeted in specific biological conditions. However, holistic experimental analyses of miRNA–mRNA relations are time consuming, as for example argonaute cross-linking immunoprecipitation [16]. Using mRNA as well as miRNA transcriptomics data together with results from target prediction algorithms can be a suitable first step in revealing mRNA–miRNA relationships based on real data. Several approaches have been suggested for the joint analysis of miRNA and mRNA data [17]. Integrating expression data increases the chance of identifying functionally relevant RNA–interactions. Thus, this chapter focuses on how to perform expression profiling and how to integrate mRNA and miRNA data together with in silico target predictions as well as how to interpret results of joint expression data analysis (Fig. 1).

2 Materials

2.1 Sample Preparation

1. *Cell culture.* Always use the appropriate base medium, e.g., Dulbecco's Modified Eagle's Medium, and the corresponding supplements, e.g., amino acids and growth factors for culturing of cells. Growth conditions, such as 37 °C and 5 % CO₂ are suitable for most mammalian cells. A sterile working atmosphere and sterile working materials such as flasks, pipettes, filter tips etc. are mandatory.
2. *RNA extraction.* For total RNA purification you can utilize ready to use kits from Qiagen, Life Technologies, Promega, or PEQlab for example. We recommend the use of miRNeasy Mini Kit (Qiagen) as described by the supplier. Utilize RNase-free water for RNA elutions.

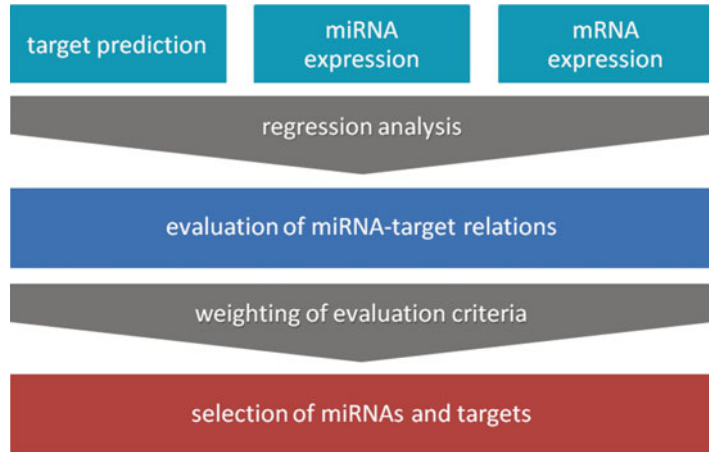


Fig. 1 Workflow of integrated mRNA and miRNA analysis. Integrated analysis starts with the input of target predictions and miRNA as well as mRNA data. Data from multiple regression analysis is further analyzed to make a final selection of miRNA–target interactions of high interest

3. *RNA quantity and quality.* RNA concentration and purity is determined by utilizing the Spectrophotometer NanoDrop1000 (Thermo Scientific) and RNA quality is analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies).
4. *cDNA synthesis and RT-qPCR.* Both, conversion of RNA to cDNA (miScript II RT Kit) and the quantification of microRNAs (miScript SYBR Green PCR Kit and miScript Primer Assay) can be performed by utilizing the miScript System of Qiagen.

2.2 mRNA Profiling by Hybridization Arrays

For mRNA profiling an oligonucleotide hybridization-based platform such as the Gene 1.0 ST Array System from Affymetrix can be used. The Affymetrix GeneChip Mouse Gene 1.0 ST Array detects 28 853 well annotated genes and consists of a square glass substrate enclosed in a plastic cartridge implying 770 317 distinct 25-mer oligonucleotide probes, based on the February 2006 mouse genome sequence (UCSC mm8, NCBI build 36) with an extensive acquisition of RefSeq, putative complete CDS GenBank transcripts, all Ensembl transcript classes and RefSeq NMs from human and rat [18].

Use Affymetrix GeneChip® Whole Transcript (WT) Sense Target Labeling Assay and the corresponding Kits: GeneChip® Eukaryotic Poly-A RNA Control Kit, GeneChip® WT cDNA Synthesis and Amplification Kit, GeneChip® Sample Cleanup Module, GeneChip® WT Terminal Labeling Kit, GeneChip Hybridization, Wash and Stain Kit, and GeneChip® IVT cRNA Cleanup Kit. With this Kit system, samples are labeled and hybridized by generating amplified and biotinylated sense-strand DNA targets

from the whole expressed genome of interest. Utilize the Affymetrix GeneChip® Fluidics Station 450 for the 169 array format applying the FS450_0007 fluidics protocol of Affymetrix. Use Affymetrix GeneChip® Scanner 3000 7G or a higher version for scanning and generation of optical images of the hybridized chips which are called DAT files.

After image exposure proceed the generated CEL files by using the Affymetrix GeneChip® Operating System (GCOS), Affymetrix GeneChip® Command Console (AGCC). Affymetrix Expression Console™ Software is part of the Affymetrix Power Tool (APT) and can be used for normalization. Perform statistical analysis and pathway analysis with GeneChip®-compatible™ Software, NetAffx Analysis Center, Integrated Genome Browser (IGB) and further third-party platforms such as Multi Experiment Viewer (MeV) [19, 20] and R programming language.

2.3 *microRNA* Profiling by Hybridization Microarrays

miRNA profiling can be performed by using an oligonucleotide hybridization-based platform such as the Mouse miRNA Microarray Release 15.0, 8 × 15 K from Agilent Technologies. The array consists of one glass slide formatted with eight high-definition 15 K arrays containing probes for 696 distinct miRNAs based on Sanger miRBase (release 15.0) (Agilent Technologies product information). Samples are labeled and hybridized by using the miRNA Complete Labeling and Hybridization Kit (Agilent Technologies). Signal intensities are acquired by using the Agilent Microarray Scanner G2505C and further processed by applying the Feature Extraction Software 10.7.3.1 (Agilent Technologies). Agilent's Feature Extraction software automatically reads and processes raw microarray image files. The software finds and places microarray grids, rejects outlier pixels, accurately determines feature intensities and ratios, flags outlier pixels, and calculates statistical confidences (Agilent technologies product information). R programming language can be utilized for statistical analysis and visualization of the data.

2.4 *microRNA* Profiling by qPCR Low-Density Arrays

qPCR profiling is facilitated by the TaqMan Rodent MicroRNA Array cards A and B 144 (Applied Biosystems by Life Technologies) which comprehensively cover Sanger miRBase v10. Together both cards (A and B) contain a total of 585 miRNA assays. Moreover, each array contains six control assays (Applied Biosystems product information). TaqMan MicroRNA Arrays are used in conjunction with Megaplex™ RT Primers that are predefined pools of up to 381 RT primers. Megaplex™ PreAmp Primers are used for pre-amplification step. Low-density arrays (385-well format) are run on the 7900 HT Fast Real-Time PCR System (Applied Biosystems by Life Technologies). Quality control and derivation of C_q-values can be done using RQ Manager 1.2 (Applied Biosystems by Life Technologies). The data normalization, comprising analysis and

visualization, can be performed by using RealTime StatMiner (Integromics) software as well as R programming.

2.5 Bioinformatics and Databases

For in silico target prediction use the latest version of TargetScan (currently TargetScan Release 6.2, June 2012) (<http://www.targetscan.org/>) [21, 22]. Moreover, use miRanda (currently the latest release of microrna.org is August 2010; <http://www.microrna.org/>) [23, 24] for target prediction.

For analyzing the inverse relation of expressed miRNA and mRNAs in conjunction with target predictions we recommend using a Lasso regression model [25]. miRNA–mRNA relations derived from the regression analysis can be further processed by testing for enrichment in gene ontology (GO) terms [26], or KEGG pathways (<http://www.genome.jp/kegg/pathway.html>) amongst others. Applying Genomatix Pathway System (GePS) (Genomatix) facilitates the comprehensive analysis and visualization of enriched canonical pathways, GO terms, disease terms, and transcription factors based on information extracted from public and proprietary databases [27] and co-citation in the literature. GePS also facilitates the creation and extension of networks based on literature data.

3 Methods

Sophisticated analysis of transcriptomics data together with bioinformatic data mining does not only reveal novel players in biological processes, but in addition generates a comprehensive view on posttranscriptional networks. Prior to generating transcriptomic regulatory networks including protein coding transcripts as well as small regulatory RNAs, the expression of mRNAs and microRNAs should be monitored. For expression studies it is important to use standardized conditions including reasonable numbers of technical and biological replicates and standard preparation protocols for sample processing as well as tangible aims of the study.

3.1 Sample Preparation

1. *Cell Culture.* Cell cultures are utilized in medical and molecular laboratories for diagnostics as well as research. In most cases, cells are cultivated for days or weeks to receive sufficient amounts of cells for analysis. For RNA extraction from cells the minimum amount of starting material is usually 100 cells [28]. The maximum amount of cells depends on the RNA content of the cell type. Parameters influencing the reproducibility of results are as follows:
 - Cell culture reagents and working materials.
 - Experience of the operator himself or herself.

For culturing of cells one should always use the same batch and number of cells and if possible even the same passage. It is also recommended to use cells not longer than 20 passages [29]. Some cells lose their characteristics rather rapidly when taken into culture. In these cases, only a few passages are advisable. In cell culture experiments technical replicates ($n > 2$) and biological replicates ($n > 3$) are usually utilized for the whole experiment. The vitality of cells should be monitored during the experiment. This can be visually verified by using microscopes and additionally with cell viability tests (e.g., Promega) or electric cell-substrate impedance sensing [30]. The reproducibility of results strongly depends on highly standardized workflows for each independent experiment. For RNA isolation you should also use equal volumes of cell lysis buffer. Lysed cells should always be kept on dry ice and stored at $-80\text{ }^{\circ}\text{C}$.

2. *RNA Extraction.* Since both mRNA and miRNA should be included in expression profiling experiments, total RNA purification kits need to be used for isolating RNA from cultured cells or various tissues. It is very important that the kit recovers RNA molecules smaller than 200 nucleotides. Homogenized cell lysates should be thawed at $37\text{ }^{\circ}\text{C}$ and subsequently incubated at room temperature (RT) for 5 min. The miRNeasy Mini kit (Qiagen) combines phenol/guanidine-based sample lysis with chloroform-based separation of nucleic acids from proteins and other cell constituents. For RNA purification silica membrane columns are used. Due to the toxicity of Phenol and Chloroform it is very important to work carefully under a fume. After sample homogenization with the Qiazol lysis reagent and the addition of chloroform the aqueous phase is separated from the organic phase by centrifugation. RNA is included in the upper aqueous phase while DNA is located in the interphase and proteins in the lower organic phase or interphase. Only the upper aqueous phase is separated and contamination with the interphase and lower phase should be strongly avoided. The RNA should be isolated as described by the supplier [28]. DNA digest on the spin column is also not recommended, because of material loss and less inefficiency compared to DNase treatment protocols of solved RNA. Total RNA is then eluted in $40\text{ }\mu\text{l}$ RNase-free water and stored at $-80\text{ }^{\circ}\text{C}$. To obtain a higher total RNA concentration, it is recommended to repeat the elution step by using the same RNeasy spin column and the first eluate [28].
3. *RNA quantity and quality.* After RNA extraction, RNA concentration and purity are determined. Analyze $1.5\text{ }\mu\text{l}$ RNA solution by utilizing the Spectrophotometer NanoDrop1000. All RNA samples should occupy a 260/280 ratio within a range of 2.8–2.1. The RNA quality can be further confirmed

by using the Agilent 2100 Bioanalyzer. The Agilent 2100 Bioanalyzer is a chip-based platform that uses microcapillary electrophoresis to analyze proteins, nucleic acids or cells [31].

4. *cDNA synthesis and RT-qPCR*. Microarray technology (see Subheadings 3.2 and 3.3) is one of the most powerful tools for assumption based large-scale expression profiling. However, microarray profiling results have to be validated by using quantitative real-time PCR (qPCR). qPCR utilizes an RNA-dependent DNA polymerase to synthesize complementary DNA. For the reverse transcription of miRNA and mRNA we used a polyadenylation based approach (miScript II RT Kit by Qiagen). For the master mix add 4 μ l 5 \times miScript HiFlex buffer, 2 μ l 10 \times miScript Nucleic Mix, 2 μ l miScript Reverse Transcriptase Mix, and RNA. The amount of RNA is in the range of 10 pg to 1 μ g and depends on the experimental design and number of target reactions. Fill the reaction up with water to a total volume of 20 μ l. The RT-reactions are incubated for 1 h at 37 $^{\circ}$ C and the cDNA reaction is stopped by incubation at 95 $^{\circ}$ C for 5 min. Samples should be stored at -20 $^{\circ}$ C [32].

The microRNA and gene expression can be analyzed with the CFX384 Touch 244 Real-Time Detection Cycler (Bio-Rad Laboratories). For specific quantification of miRNA and mRNA expression the miScript Primer Assay and the miScript SYBR Green-based RT-PCR Kit can be used. The reaction mix should be prepared as described by the supplier. The following cycling conditions should be used: after initial activation for 15 min at 95 $^{\circ}$, the cycle steps of denaturation for 15 s at 94 $^{\circ}$ C, annealing for 30 s at 55 $^{\circ}$ C, and an extension phase for 30 s at 70 $^{\circ}$ C should be repeated for 39 cycles [32] (see **Notes 1** and **2**).

3.2 mRNA Profiling and Validation

1. *mRNA hybridization microarrays*. Microarray technology uses the natural attraction between nucleotides, consisting of probe-target hybridization, detection and quantification of labeled targets to determine the relative amount of nucleic acid sequences and enables high-throughput analysis of transcripts. Therefore, the major applications of microarray technology are gene expression profiling and genetic variation analysis [33]. Using in situ synthesized oligonucleotides as probes and in silico designed microarrays, Affymetrix pioneered the understanding of total transcript activity. This chapter focuses on gene expression profiling with Affymetrix Gene Chip Gene 1.0 ST Array System that offers an option for whole-transcript coverage. When performing mRNA expression profiling by using Gene Chip Mouse Gene 1.0 ST Array (Affymetrix) follow the manufacturer's instructions [34] (see **Notes 3–5**). The manufacturer's instructions are summarized below.

2. *Isolation of total mature RNA from cells (see Subheadings 2.1 and 3.1.)*
3. *Preparation of total RNA with T7-(N)6 Primers and Poly-A RNA Controls.* For this step the GeneChip® Poly-A RNA Control Kit is required. In the first step the dilutions of Poly-A RNA Controls is prepared. Add 2 µl of 269 Poly-A RNA Control stock to 38 µl of Poly-A control Dil Buffer (first dilution 1:20). Mix gently and spin down the tube. Transfer 2 µl of the first dilution to 98 µl of Poly-A Control Dil Buffer (second dilution: 1:50). Transfer 2 µl of the second dilution to 98 µl of Poly-A Control Dil Buffer (third dilution 1:50). Next, prepare the solution of T7-(N)6 Primers/Poly-A RNA Controls. Add 2 µl of the T7-(N)6 Primers (stock conc. 2.5 µg/µl), 2 µl diluted Poly-A RNA Control (3rd dilution, 1:50) to 16 µl RNase-free water in a non-stick RNase-free tube. Mix the solution, spin down and place it on ice. Performing GenChip Gene 1.0 ST Arrays use 100 ng total sample RNA, add 2 µl of the prepared T7-(N)6 Primers/Poly-A RNA Controls solution and fill up to a total volume of 5 µl with RNase-free water. The mix is flicked, followed by centrifugation. The reaction mix is then incubated at 70 °C for 5 min and at 4 °C for 2 min. The mix is placed on ice.
4. *First-Cycle, First-Strand and second strand cDNA Synthesis.* The GeneChip® WT cDNA Synthesis Kit is required for this preparation step. For the first-cycle of first strand cDNA synthesis mix 2 µl of 5× 1st Strand Buffer, 1 µl DTT (0.1 M), 0.5 µl dNTP Mix (10 mM), 0.5 µl RNase Inhibitor, 1 µl SuperScript II enzyme in one tube. Add 5 µl of the master mix to the T7-(N)6 Primers/Poly-A RNA Controls solution, mix, and centrifuge the tube. The total volume for first strand cDNA synthesis is 10 µl. The reaction is incubated at 25 °C for 10 min, at 42 °C for 60 min, at 70 °C for 10 min, and finally at 4 °C for 2 min. Continue with the second strand cDNA synthesis and prepare the second strand master mix: To do so mix 4.8 µl RNase-free water with 4 µl MgCl₂ (17.5 mM), 0.4 µl dNTP Mix (10 mM), 0.6 µl DNA Polymerase I, and 0.2 µl RNase H, and 10 µl of the first-cycle second-strand master mix was added to the reaction tube of the first-strand cDNA synthesis reaction. Gently vortex the tube and centrifuge. The reaction is incubated at 16 °C for 120 min, at 75 °C for 10 min, and at 4 °C for further 2 min.
5. *First-cycle, cRNA synthesis and cleanup.* This procedure requires the GeneChip® WT cDNA Amplification Kit and the GeneChip® Sample Cleanup Module. Prepare the IVT master mix, including 5 µl of 10× IVT, 20 µl IVT NTP Mix, and 5 µl IVT enzyme mix. Transfer this 30 µl of the IVT master mix to

the first-cycle cDNA synthesis reaction. Mix the sample again and centrifuge. The reaction is then incubated for 16 h at 37 °C. Proceed to the cleanup procedure for cRNA using the GeneChip Sample Cleanup Module. Add 50 µl of RNase-free water to each IVT reaction, followed by adding 350 µl of cRNA Binding Buffer to each reaction. Vortex the samples for 3 s. Add 250 µl of 100 % EtOH to each reaction and mix. Transfer the sample to the IVT cRNA Cleanup Spin Column and centrifuge for 15 s at $\geq 8,000 \times g$. Discard the flow-through. The spin column is transferred to a new 2 ml collection tube and the column is washed by adding 500 µl of cRNA wash Buffer. The column is centrifuged again for 15 s at $\geq 8,000 \times g$. Discard the flow-through. The column is washed again with 500 µl of 80 % EtOH. Leave the cap of the tube open and centrifuge the column for 5 min at $\leq 25,000 \times g$. The spin column is transferred to a new 1.5 ml collection tube and 15 µl of RNase-free water are added to the membrane of the column directly. Incubate the reaction for 5 min at room temperature. Subsequently, centrifuge for 1 min at $\leq 25,000 \times g$. The concentration can be determined by measuring the absorbance in a spectrophotometer (e.g., NanoDrop).

6. *Second-cycle, first-strand cDNA synthesis.* For this step the GeneChip® WT cDNA Synthesis Kit is used. In a strip tube the cRNA sample (max. 10 µg) is mixed with 1.5 µl Random Primers (3 µg/µl) and the reaction is filled up with RNase-free water to a total volume of 8 µl. Mix and spin down the tube. The cRNA/Random Primer mix is incubated at 70 °C for 5 min, followed by incubation at 25 °C for 5 min and at 4 °C for 2 min. In a second tube the second cycle, reverse transcription master mix is prepared. Combine 4 µl of 5× 1st Strand Buffer with 2.0 µl DTT (0.1 M), 1.25 µl dNTP+dUTP (10 mM), and 4.75 µl SuperScript II. This 12 µl of second-cycle, first-strand cDNA synthesis master mix is transferred to the second-cycle, cRNA/Random Primers Mix. After flicking, the tube is centrifuged briefly. The reaction is incubated at 25 °C for 10 min, at 42 °C for 90 min, at 70 °C for 10 min, and at 4 °C for 2 min.
7. *Hydrolysis of cRNA and cleanup of single-stranded DNA.* For this preparation step the GeneChip® WT cDNA Synthesis Kit and the GeneChip® Sample Cleanup Module are required again. 1 µl of RNase H is added to each sample and the reaction is incubated at 37 °C for 45 min, at 95 °C for 5 min, and at 4 °C for 2 min. Add 80 µl of RNase-free water to each sample, followed by adding 370 µl cDNA Binding Buffer. Vortex the sample for 3 s. Convey the total volume of 471 µl to a cDNA Spin Column from the GeneChip Sample Cleanup Module. Spin for 1 min at $\geq 8,000 \times g$. Discard the flow-through.

The spin column is transferred to a new 2 ml collection tube and 750 μl of cDNA Wash Buffer are added. Centrifuge the tube at $\geq 8,000 \times g$ for 1 min and discard flow-through. Open the cap of the cDNA Cleanup Spin Column and spin at $\leq 25,000 \times g$ for 5 min, discard the flow-through, and place the column in a new 1.5 ml collection tube. Add 15 μl of the cDNA Elution Buffer directly to the spin column membrane and centrifuge at $\leq 25,000 \times g$ for 1 min. Add another 15 μl of cDNA Elution Buffer to the column membrane and incubate at room temperature for 1 min, followed by spinning at $\leq 25,000 \times g$ for 1 min. Determine the yield by spectrophotometric UV measurement (e.g., NanoDrop).

8. *Fragmentation of single-stranded DNA.* This step requires the GeneChip[®] WT Terminal Labeling Kit. Use 0.2 ml strip tubes for the fragmentation. First, add 5.5 μg single-stranded DNA to a tube and fill up with RNase-free water to a total volume of 31.2 μl . Second, prepare the fragmentation master mix in a fresh tube, by adding 10 μl RNase-free water, 4.8 μl 10 \times cDNA Fragmentation Buffer, 1 μl UDG (10 U/ μl), and 1 μl APE 1 (1,000 U/ μl) per reaction. This 16.8 μl fragmentation master mix is added to the prior prepared single-stranded DNA, vortex gently and centrifuge the sample. The reaction is incubated at 37 °C for 60 min, at 93 °C for 2 min and in the end at 4 °C for 2 min. After incubation transfer 45 μl of the reaction mix to a new 0.2 ml strip tube. Examine the RNA quality using 2 μl of the residual RNA with the Bioanalyzer (as described in further detail in Chapter 5). The fragmented single stranded DNA should be approximately 40–70 nt.
9. *Labeling of fragmented single-stranded DNA.* The GeneChip[®] WT Terminal Labeling Kit will be used. First, prepare a labeling master mix by mixing 12 μl of the 5 \times TdT buffer, 2 μl TdT, and 1 μl DNA Labeling Reagent in a 0.2 ml tube and aliquot 15 μl of the prepared master mix (total volume 60 μl). After adding 15 μl of the master mix to the 45 μl fragmented single stranded DNA, the tubes are mixed and centrifuged. Incubate the reactions at 37 °C for 60 min, at 70 °C for 10 min, and finally at 4 °C for 2 min.
10. *Hybridization.* For hybridization the GeneChip Hybridization, Wash and Stain Kit is required. For the format array a total volume of 100 μl Hybridization Cocktail is needed. Add 27 μl fragmented and labeled DNA target, 1.7 μl (3 nM) Control Oligonucleotide B2, 5 μl of 20 \times Eukaryotic Hybridization Controls, 50 μl of 2 \times Hybridization Mix, 7 μl of DMSO, and 9.3 μl nuclease-free water to a 1.5 ml RNase-free tube. Gently vortex the tube and heat the hybridization cocktail at 99 °C for 5 min, subsequently cool the reaction for 5 min to 45 °C and centrifuge at full speed for 1 min. Equilibrate the GeneChip

ST Array to room temperature and inject 80 μ l of the Hybridization Cocktail, including the fragmented and labeled DNA, into the array. The array is then placed into a rocker hybridization oven (45 °C, 600 rpm) for 17 h.

11. *Wash, Stain and Scan.* For washing, staining and scanning of the arrays, the Gene Chip Hybridization, Wash and Stain Kit is required. After hybridization vent the array and extract the Hybridization Cocktail. Refill the array with 100 μ l wash buffer A. Gently tap the bottles of staining reagents and aliquot 600 μ l of staining cocktail 1 into a 1.5 ml amber tube, 600 μ l of staining cocktail 2 into a further (clear) microcentrifuge tube and 800 μ l of Array Holding Buffer into a new 1.5 ml microcentrifuge tube. All tubes are centrifuged to remove any air bubbles. For washing and staining the probe array, the Fluidics Station 450/250 is used. Please, follow the detailed manufacturer's instructions [35].
12. *Scanning.* If there are no air bubbles inside the array, it is ready to be scanned on the GeneChip Scanner 3000. The scanner is also controlled by Affymetrix® GeneChip Command Console (AGCC). Optical images of the hybridized chip, called DAT files, are generated.
13. *Data file generation.* Proceed CEL files from the prior generated DAT files which contain intensity information by using the Affymetrix GeneChip® Operating System (GCOS) and Affymetrix GeneChip® Command Console (AGCC). The Affymetrix Expression Console™ Software can be used for background adjustment, normalization, data quality control and gene-level signal detection, supporting also probe set summarization and CHP file generation for 3' WT expression arrays.
14. *Statistical analysis and pathway analysis.* For the statistical as well as for the pathway analysis you can use 3rd party platforms as well as R programming language. The correlation coefficient can be calculated. High correlation coefficients >0.90, indicate a high reproducibility of the data. The analysis of variance (ANOVA) is extremely important during microarray processing. One-way analysis of variance tests significance of one condition, e.g., time, whereas two-way analysis of variance measures the significance of two factors simultaneously, e.g., time and treatment. For statistical analysis you can use different statistical test, e.g., the Mann–Whitney Test. It is also recommended to filter for low intensity values, because counts <30 are close to the background and therefore should be removed from the subsequent analysis. For visual illustration of the data you can use heat maps and scatter plots. It is also recommended doing cluster analysis such as principle component analysis. Gene expression profiling for specific marker-genes should be

performed as well. For validation of microarray results it is also recommended to use quantitative real-time PCR (*see* **Notes 1, 2, 6–8**).

3.3 microRNA Profiling by Hybridization Microarrays

When performing miRNA expression profiling by using Mouse miRNA Microarrays (Agilent Technologies) (*see* **Note 9**) follow the manufacturer's instructions specified for the miRNA Complete Labeling and Hybridization Kit:

1. *Labeling*. Dephosphorylate and spike 100 ng total RNA per sample by mixing 4 μ l of RNA with 0.7 μ l 10 \times Calf Intestinal Phosphatase Buffer, 1.1 μ l Labeling Spike-in, 0.5 nuclease-free water, and 0.7 μ l Calf Intestinal Phosphatase. Incubate the reaction mixture at 37 °C for 30 min. Denature RNA by adding 5 μ l DMSO at 100 °C for 8 min and cool at 0 °C for 5 min. For fluorophor ligation add 2 μ l 10 \times T4 RNA Ligase Buffer, 2 μ l nuclease-free water, 3 μ l pCp-Cys, and 1 μ l T4 RNA Ligase and incubate at 16 °C for 2 h. Purify samples by using Bio-Spin 6 Chromatography Columns (Bio-Rad) and dry in a SpeedVac Concentrator (Thermo Scientific) at 50 °C.
2. *Hybridization and scanning*. For hybridization of labeled miRNAs resuspend the dried samples in 17 μ l nuclease-free water and mix with 1 μ l Hybridization Spike-in, 4.5 μ l 10 \times GE Blocking Agent, and 22.5 μ l 2 \times Hi-RPM Hybridization Buffer and incubate at 100 °C for 5 min with subsequent cooling at 0 °C for 5 min and centrifugation. Load the samples into the array and hybridize at 55 °C for 20 h. Wash miRNA microarrays and scan in a single pass mode with a scan resolution of 3 μ m, 20 bit mode. Extract signal intensities and background and log₂ transform the data by using Feature Extraction Software 10.7.3.1 (Agilent Technologies).
3. *Evaluation and analysis of data*. Retain miRNAs that show a signal greater than zero in at least two of the replicates within one group. For miRNAs that meet these detection criteria define a value equal to zero as outlier and replace it by the mean value of the other replicates. Normalize the data by applying loess M normalization [36, 37]. Microarray data should be registered into ArrayExpress database [38], or Gene Expression Omnibus [39] which are publicly available repositories consistent with the MIAME guidelines [40]. For further processing of the data use GeneSpring GX Software (Agilent Technologies) or R programming (*see* **Note 10**). To detect the significant differential expression use significance analysis of microarray (SAM) [41]. SAM is an assumption free approach adapted to microarray analysis that identifies differentially expressed miRNAs by permutation and performing false discovery rate (FDR) correction of *p*-values.

3.4 *microRNA* Profiling by qPCR Low-Density Arrays

Validation of miRNA profiling results derived from hybridization-based arrays can be performed by qPCR based analysis (*see Note 11*). We recommend setting up three reverse transcription reactions per sample as the reverse transcription itself introduces bias. When using the TaqMan Rodent MicroRNA Array system follow the manufacturer's instructions.

1. *Reverse transcription.* Prepare the reverse transcription reaction mix by adding the following amounts per sample: 0.8 μ l Megaplex RT Primers (10 \times), 0.2 μ l dNTPs with dTTP (100 mM), 1.5 μ l MultiScribe Reverse Transcriptase (50 U/ μ l), 0.8 μ l RT Buffer (10 \times), 0.9 μ l magnesium chloride (25 mM), 0.1 μ l RNase Inhibitor (20 U/ μ l), 0.2 μ l nuclease-free water. Add 1–350 ng total RNA or water for the no template control reaction, respectively. Incubate the reaction mixture on ice for 5 min. Using the 9700HT Systems (Applied Biosystems by Life Technology) set up the following run method: 40 cycles of 16 $^{\circ}$ C for 2 min, 42 $^{\circ}$ C for 1 min, and 50 $^{\circ}$ C for 1 s. Then set up a final step of 85 $^{\circ}$ C for 5 min followed by cooling to 4 $^{\circ}$ C.
2. *Pre-amplification.* Pre-amplify the reverse transcription product by mixing 12.5 μ l TaqMan PreAmp Master Mix (2 \times), 2.5 μ l Megaplex PreAmp Primers (10 \times), and 7.5 μ l nuclease-free water and mix by inverting the tube. Subsequently add 2.5 μ l reverse transcription product and 22.5 μ l of the pre-amplification mixture and incubate on ice for 5 min. Set up the following running conditions on a 9700HT platform: 95 $^{\circ}$ C for 10 min, 55 $^{\circ}$ C for 2 min, 72 $^{\circ}$ C for 2 min followed by 12 cycles consisting of 95 $^{\circ}$ C for 15 s and 60 $^{\circ}$ C for 4 min. To facilitate enzyme inactivation heat it to 99.9 $^{\circ}$ C for 10 min and cool down to 4 $^{\circ}$ C. To dilute the pre-amplification product add 75 μ l of 0.1 \times TE Buffer (8.0 pH) to each reaction. When having three reverse transcription reactions per sample we recommend to aequimolarly pool the reaction products by unifying 4 μ l each (*see Note 12*).
3. *Real-time quantitative PCR.* Run the real-time PCR reaction by mixing the following volumes per array: 450 μ l TaqMan Universal PCR Master Mix (no AmpErase UNG, 2 \times), 9 μ l of the diluted and pooled pre-amplification product, 441 μ l nuclease-free water. Load the array by dispensing 100 μ l of the PCR reaction mix into each port of the TaqMan MicroRNA Array, centrifuge and seal the array. Run the array on the 7900HT System by choosing "Relative Quantification" and the 384-well TaqMan Low Density Array default thermal-cycling conditions of the SDS software.
4. *Evaluation and analysis of data.* To review the results, transfer the SDS files into a Comparative CT (RQ) study using the RQ Manager software (Applied Biosystems) (*see Note 13*). Applied Biosystems recommends analyzing the study with "Automatic

Baseline” and “Manual CT” set to 0.2. View each amplification plot manually and adjust the threshold settings for individual assays if necessary. It is important to use the same threshold settings across all samples or arrays within a study for a given assay. For detailed downstream analysis use software such as Real-Time StatMiner Software (Integromics) which amongst many other analysis options allows to for example omit miRNAs which do not show cycle of quantification (Cq) values smaller 32 in at least two of the corresponding replicates of a group. For normalization of the data use modified loess method [36] (*see Note 14*). Calculate fold changes of relative expressions (*see Note 8*) and identify significant differential expression by applying SAM and FDR correction.

3.5 Integrative Analysis of miRNA and mRNA Expression Data

The main assumption in the joint analysis of miRNA and mRNA expression data is that regulation of gene expression by miRNAs primarily takes place in the form of mRNA degradation. It was shown that this actually holds for mammalian cells [12]. Therefore, we can assume that changes in gene expression on mRNA level can be explained by expression changes of targeting miRNAs.

A straightforward approach for detecting these regulatory relations would be to calculate the Pearson correlation coefficient (r) between the mRNA expression and the expression of miRNAs, which are predicted to target the respective gene, over certain conditions. Afterwards one is able to determine the statistical significance of the correlation coefficient that can be utilized to select significantly anticorrelated miRNA–mRNA relations [42]. However, there is one major issue in the correlation analysis of mRNA and miRNA expression data, since a gene can also be targeted by more than one miRNA. A linear relationship between the expression profiles of several miRNAs and the expression of a common predicted target gene cannot be detected in a one-to-one fashion, because each of these miRNAs may have individual influences on the gene expression (Fig. 2). A more suitable method for quantifying the down-regulation by predicted miRNAs is therefore to use a multiple linear regression approach. Here, the goal is to simultaneously incorporate the multiple expression profiles of the predicted miRNAs in order to assess their individual contribution in explaining the mRNA expression. Since target predictions usually consist of many false positives, it is desirable to select for each gene only these miRNAs out of the set of predicted ones that have an actual influence on the mRNA expression. In order to achieve this a Lasso regression model can be used that performs a feature selection on the explanatory variables. The LASSO regression model is therefore suitable to identify potential regulatory relations between miRNAs and genes [25]. Since we are mainly interested in the down-regulation of genes, a Lasso regression model with non-positive constraints appears to be the most appropriate approach. This is implemented for example in TaLasso [17].

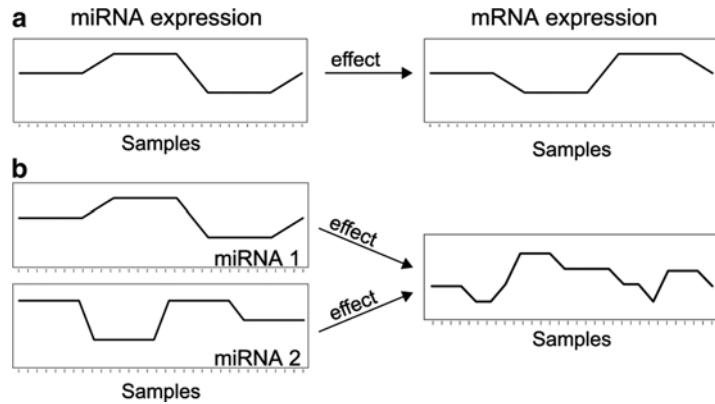


Fig. 2 MiRNA–target relations depending on target prediction algorithms. **(a)** The mRNA expression profile of a gene is expected to be anti-correlated to a single targeting miRNA over different conditions if the respective miRNA has a regulatory effect on the gene. **(b)** If more than one miRNA has a regulatory effect on the gene, the anticorrelation becomes less clear, since all targeting miRNAs can have individual influences on the mRNA expression

An overview of different approaches for the determination of miRNA–mRNA relations based on target predictions and combined expression data is given in the review of Muniategui et al. [17] (*see* **Notes 15** and **16**). The integrative analysis of mRNA and miRNA expression can be visualized by the open-source software Cytoscape 2.8.3 (<http://www.cytoscape.org/>) (Fig. 3).

TaLasso is available as an easy-to-use Web interface that can deal with miRNA and mRNA expression matrices. Two tab-separated text files must be provided as input. These files have to contain the respective expression values as well as the miRNA and mRNA identifier as row names. The columns of the expression matrices correspond to the samples and must be matched. Examples for such kind of input files are given on the Web site (<http://talasso.cnb.csic.es/>) (Fig. 4). Several target predictions resources can be chosen as underlying network. As mentioned above, we propose TargetScan as target prediction tool (*see* **Note 17**). Furthermore, a set of experimentally validated target interactions can be selected in order to assess the amount of validated miRNA–mRNA relationships in the result. Standard settings can be used for all other parameters. After submitting the job, the result page will show up (Fig. 5). It contains all predicted target relationships that were predicted by the algorithm. The score and the p -value indicate the significance of the interaction. The result can be downloaded by right-clicking on the download links above the result list and selecting “Save as...”. All three result files should be downloaded. The network of miRNA–mRNA interactions is stored as a comma-separated text file containing the assignment matrix. In order to prepare this matrix for importing into graph visualization tools, it has to be preprocessed.

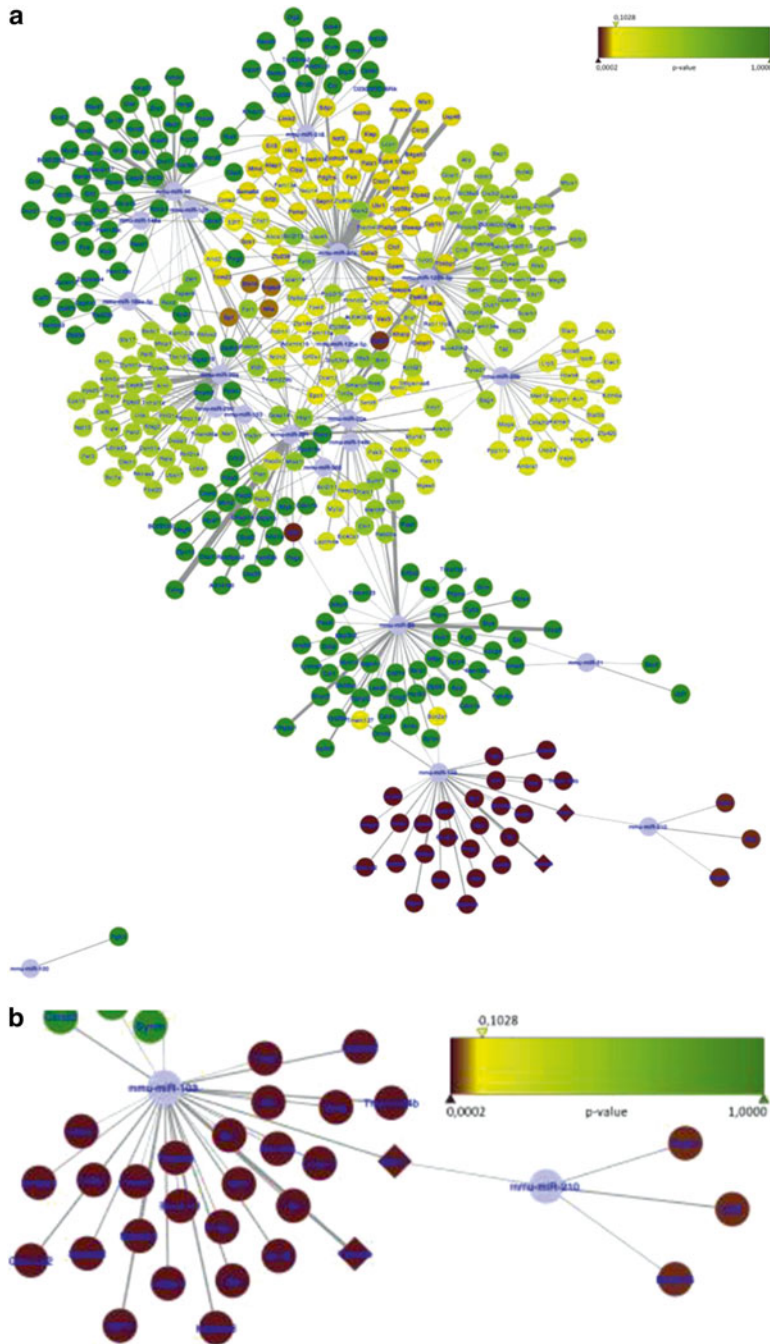


Fig. 3 Pathway dependent visualization of miRNA–mRNA relations with Cytoscape 2.8.3. The enrichments include TargetScan-based miRNA–target predictions as well as mRNA and microRNA expression data. The blue nodes represents microRNAs, whereas the *green*, *yellow* and *red* nodes implying genes. Color intensity (*red*= significant *green*= nonsignificant) indicates the significance of a network area to be overrepresented in a certain pathway or biological context. **(a)** The results for the Neurotrophin signaling pathway are exemplarily depicted. **(b)** Image section of **(a)** The microRNAs 103 and 210 are highly involved in the regulation of the Neurotrophin signaling pathway, corresponding to the overrepresented gene environment



TALASSO

Main Help

1. **Gene Expression matrix** geneExpression.txt Gene example file

2. **MiRNAs Expression matrix** mimaExpression.txt Mirna example file

3. **Data Type**

4. **Gene - MiRNAs putative targets**

microma mirbase mirecords 2007

mirecords 2010 (mirgen) DIANAmicroT (mirgen) I mirandaXL pictar4way targetscans

(mirgen) I pictar4way targetscans (mirgen) Union (mirgen) microma

(mirgen) mirbase (mirgen) pictar4way (mirgen) pictar5way

(mirgen) targetscans mirwalk tarbase

5. **Gene - MiRNAs putative targets for validation**

mirecords 2010

mirwalk

tarbase

6. **Algorithm**

7. **Tuning Factor (only for TaLasso algorithm)**

8. **Name your job (optional)**

If you find TaLasso useful, please include the following publication in your references:

Muniategui A, Nogales-Cadenas R, Vázquez M, L. Aranguren X, Agirre X, et al. (2012) Quantification of miRNA-mRNA Interactions. *PLoS ONE* 7(2): e30766. doi:10.1371/journal.pone.0030766

[Medline] [Online version]

Fig. 4 Talasso Web interface. Screenshot of the Talasso Web interface, which implements a sequential workflow and different analysis parameters

Therefore, the R environment for statistical computing (<http://www.R-project.org/>) can be used. This software can be downloaded for free.

Within the R environment, the downloaded assignment matrix (“geneExpression-X_targets.txt”) can be imported via `m=as.matrix(read.csv(“c:\pathtofile\geneExpression-X_targets.`

Results for geneExpression-186

Download results files: [Targets] [Genes] [MiRNAs]

Table of results

First [1] >> Last

Gene	miRNA	Score	pValue	Mirecords	Mirwalk	Tarbase
<input type="checkbox"/> HECW1 [ENSG00000002746]	<input type="checkbox"/> hsa-miR-92	0.10203	0.003072			
<input type="checkbox"/> SLC7A2 [ENSG00000003989]	<input type="checkbox"/> hsa-miR-15b	0.095237	0.015322			
<input type="checkbox"/> EXTL3 [ENSG00000012232]	<input type="checkbox"/> hsa-miR-124a	0.065003	4.4876e-05			
<input type="checkbox"/> NADK [ENSG00000008130]	<input type="checkbox"/> hsa-miR-1	0.037987	0.00103			
<input type="checkbox"/> BTBD7 [ENSG00000011114]	<input type="checkbox"/> hsa-miR-130a	0.03297	0.099781			
<input type="checkbox"/> BTBD7 [ENSG00000011114]	<input type="checkbox"/> hsa-miR-30b	0.01116	0.19897			
<input type="checkbox"/> CALCR [ENSG00000004948]	<input type="checkbox"/> hsa-miR-30b	0.0044773	0.11517			
<input type="checkbox"/> PAF1 [ENSG00000006712]	<input type="checkbox"/> hsa-miR-17-5p	0.0035542	0.15875			
<input type="checkbox"/> LASP1 [ENSG00000002834]	<input type="checkbox"/> hsa-miR-1	0.0018174	0.33767	✓	✓	✓
<input type="checkbox"/> MAP4K5 [ENSG00000012983]	<input type="checkbox"/> hsa-miR-92	0.0016009	0.022455			

Fig. 5 Talasso output. The Talasso result page contains all miRNA–mRNA relationships, which were predicted by the algorithm as well as the score and the *p*-value indicating the significance of the interaction

txt”, *header = F*)). The miRNA identifiers can be imported by the command `colnames = scan("c:\path\to\file\geneExpression-X_mirna.txt")` and the mRNA identifiers by `rownames = scan("c:\path\to\file\geneExpression-X_gene.txt")`. Afterwards, the row and column names of the assignment matrix *m* can be set by `dimnames(m) = list(rownames, colnames)`. If you need further help for these commands, just type `? command`, e.g. `? read.csv`. In order to select only those miRNAs and genes in the network that have actual interactions partners, type `m = m[apply(m, 1, function(x) sum(x != 0) > 0), apply(m, 2, function(x) sum(x != 0) > 0)]`. The interactions are indicated by the respective *p*-value in the result file which should be logarithmized first. Therefore, type `m[m != 0] = -log10(m[m != 0])`. Afterwards, the processed assignment matrix can be exported as *xlsx* file by `write.xlsx(m, "c:\path\to\file\assignment.xlsx")`. This function requires the package “*xlsx*” that can be installed by `install.package("xlsx")`. You are now able to import this assignment matrix into graph visualization tools like yEd (<http://www.yworks.com>). Therefore, open the file by File > Open and choose “assignment matrix”. We refer to the yEd user manual for further information on importing and visualizing the network (<http://yed.yworks.com/support/manual/index.html>).

3.6 Interpretation of Integrated Expression Data

An important step in the processing of results from integrated miRNA–mRNA analysis is the interpretation of miRNA–target relations. To properly select targets of high significance in the specific biological context and the hypothesis tested for one should

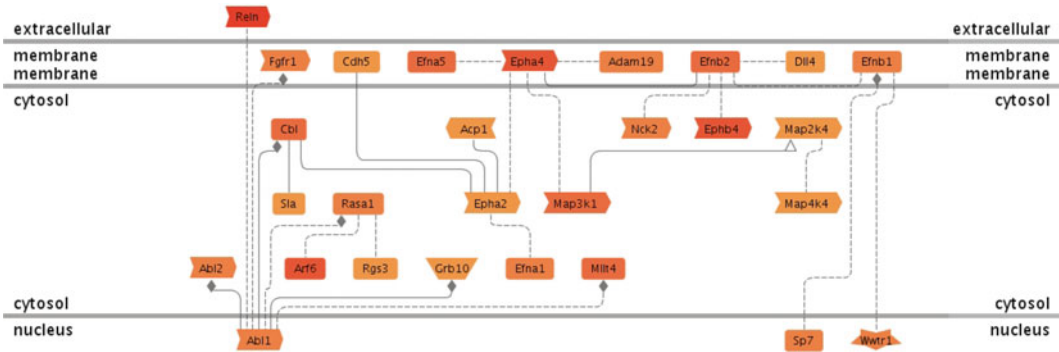


Fig. 6 Enrichment of miRNA targets in signaling pathways based on co-citation. Predicted miRNA targets derived from integrated analysis were analyzed by Genomatix Pathway System algorithm. The results for ephrin are exemplarily depicted. The *red shading* intensity increases with the number of distinct miRNAs, which are inversely correlated to as well as predicted to target the respective gene

acquire quantitative and qualitative parameters on the results from integrated miRNA–mRNA analysis. Parameters include:

1. Number of total targets per miRNA.
2. Number of targeting miRNAs per target.
3. Number of targeted transcription factors per miRNA.
4. Enrichment of targets in canonical pathways, KEGG pathways or pathways based on co-citation (Fig. 6).
5. Enrichment of targets in GO terms, or disease terms etc.
6. Strength of miRNA or mRNA regulation.

Prioritize and weight parameter based on the biological question you look at and use these information for selection of miRNA–target relations you might want to experimentally validate, e.g., transcription factors which are targeted by even several inversely correlated miRNAs or targets of a specific pathway (*see Note 18*).

4 Notes

1. The miScript Reverse Transcription Kit contains 2 buffers. One buffer is the 5× miScript HiSpec Buffer for cDNA preparation. This buffer is used when one intends to do mature miRNA profiling. The other buffer is the 5× miScript HiFlex Buffer for cDNA preparation. This buffer is used when one aims at quantifying mature miRNAs in parallel with precursor miRNAs or mRNAs.
2. A major advantage of the miScript PCR System is the generation of cDNA for both, microRNA and gene expression profiling within the same reaction. On the other hand the miScript PCR

system is very cost intensive. Alternatively, you can quantify the expression of miRNAs and mRNA independently. Latter implies the use of the miScript PCR System for miRNA quantification as described above and the use of gene expression profiling RT kits and SYBR Green kits suited for reverse transcription of mRNAs only.

3. The equipment such as the Affymetrix GeneChip® Fluidics Station 450 and Affymetrix GeneChip® Scanner 3000 7G are cost intensive. At the same time the microarray performance should be standardized according to the MIAME guidelines [40]. Therefore, we recommend to use providers which proceed the arrays (e.g., core facility of the EMBL, Heidelberg).
4. Probes of the GeneChip Gene 1.0 ST Array cap the whole length of the detected genes. This provides a more comprehensive and proper image of gene expression compared to classical 3' based microarrays. Most 3'-based expression arrays are addicted to transcript's poly-A tails and probes are located to the 3'end of the detected genes. Under certain conditions some genes may not be properly represented on a classical 3'-based expression array such as partially degraded RNA samples, truncated transcripts, or alternative splicing at the 3'end of the gene or polyadenylation sites (www.affymetrix.com).
5. The Gene ST Arrays are the latest of the Affymetrix expression arrays. With the next generation of GeneChip ST Arrays, the GeneChip 2.0 ST Arrays, it is even possible to detect and measure long intergenic RNAs (lincRNA) (www.affymetrix.com).
6. For gene expression analysis profiling, cDNA can be also synthesized by using 1 ng–1 µg. For gene expression analysis of total RNA and 5× First Strand Buffer, 10 mM of each dNTP, 50 µM Hexamere Primer, and 100 U of Moloney Murine Leukemia Virus Reverse Transcriptase; M-MLV RT [H-] (Promega) per reaction. The reaction mix is incubated for 20 min at 21 °C. Then the cDNA synthesis step is performed at 120 min at 48 °C and the reaction is stopped by 2 min incubation at 90 °C. To quantify the gene expression you can use SSo Fast Eva Green Supermix (Bio-Rad Laboratories). The final reaction volume is 10 µl. Use 5 µl SsoFast EvaGreen Supermix, 400 nM forward and 400 nM reverse primers. A total cDNA amount of 50 ng to 50 fg is recommended. The following cycling conditions are verified: After initial activation for 30 s at 98 °C the cycle steps of denaturation for 5 s at 95 °C and annealing/elongation for 20 s at 60 °C are repeated for 39 cycles. Afterwards, a melting curve analysis is performed for each run. The melting curve is generated from 65 to 95 °C with an increment of 5 °C, 5 min.

7. According to the MIQE guidelines [43] normalization of qPCR results is crucial for reducing technical variance. Genes or miRNAs considered as reference should always be evaluated in the specific experimental context prior to applying as normalizers. For evaluation of possible reference genes or microRNAs you can use geNorm [44] or NormFinder [45], both part of the gene expression analysis software suite GenEx (by MultiD Analyses AB).
8. Relative expression changes should be represented as fold-changes (FC) according to the following formula: $FC = 2^{-\Delta\Delta Ct}$. Including the following calculation steps:

$$\Delta Ct = Ct(\text{target gene}) - Ct(\text{reference gene})$$

$$\Delta\Delta Ct = \Delta Ct(\text{treatment}) - \Delta Ct(\text{control})$$

[46].

9. Hybridization based miRNA microarrays may detect not only mature miRNAs but as well precursor miRNAs. Agilent microRNA microarrays favor the detection of mature miRNAs over precursor forms because of a hairpin-like structure of the probe that sterically reduces the probability of longer oligonucleotide sequences to bind.
10. Agilent's GeneSpring GX software [47] provides statistical tools for visualization and analysis of microarray data such as clustering and principal component analysis, and pathway analysis. For users who are familiar with R language and programming the respective R scripts for statistical analysis and visualization such software is recommended. The R environment is freely available and its use is more flexible in terms of applying different methodologies compared to a given software interface.
11. qPCR assays provide significant advantages over microarrays with respect to sensitivity, dynamic range, and specificity. Probe based qPCR systems such as TaqMan MicroRNA Assays are highly specific and should be favored over DNA intercalating fluorescent dyes. However, probe design is time and cost intensive compared to using universal intercalating dyes.
12. When assay sensitivity is of the utmost importance, or when sample is limiting, a preamplification step using Megaplex™ PreAmp Primers can be added. The preamplification significantly enhances the ability to detect low expressed miRNAs, enabling the generation of a comprehensive expression profile using as little as 1 ng of input total RNA [48]. However, one should keep in mind that preamplification is an additional step which introduces bias.

13. RQ Manager is limited to the simultaneous analysis of 10 cards. This is a disadvantage for larger studies because the analysis has to be partitioned into packages containing no more than 10 cards. Thus, one cannot use for example the automatic threshold function. When having more than 10 cards the automatic adjustment cannot take into account the results of all cards and would calculate different values according to the subset of cards analyzed in parallel. Use fixed threshold option, e.g., threshold at 0.2 instead.
14. Evaluate different normalization methods for your data as inappropriate normalization can strongly affect data quality and the detection of differential expression [43, 44, 49].
15. Consider that in order to calculate reliable relations between miRNA and mRNA expression, an appropriate number of samples in your dataset is necessary. As a rule of thumb, the data should be derived from at least three different conditions or time points with at least three replicates.
16. Application of different methods for the joint analysis of miRNA and mRNA data and comparison of results allow for estimation of the stability of anticorrelated miRNA and mRNA relations over different approaches.
17. Target prediction algorithms do not necessarily contain the latest miRNA annotations according to the latest miRBase version. One needs to assure that miRNAs of interest are in the prediction dataset. Alternatively, one can choose less conservative prediction datasets and use the intersection of results from different prediction algorithms. Results of regression analysis strongly depend on the choice of prediction algorithm and whether one uses the set unity of predictions by different algorithms or the intersection. In general, rather conservative prediction algorithms should be used.
18. Depending on the biological system and hypothesis one wants to test one can implement additional parameters. The priority and weight addressed to each parameter is rather subjective and should be documented carefully to make the individual selection strategy comprehensible.

Acknowledgement

The authors gratefully acknowledge the contribution of Helmut Blum, Stefan Bauersachs, and Stefan Krebs to the conduction of expression profiling. The work was supported by a grant from the German Federal Ministry of Education and Research and the Bavarian State Ministry of Sciences, Research and the Arts.

References

1. Esteller M (2011) Non-coding RNAs in human disease. *Nat Rev Genet* 12:861–874
2. Fabian MR, Sonenberg N, Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 79:351–379
3. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11:597–610
4. Westholm JO, Lai EC (2011) Mirtrons: microRNA biogenesis via splicing. *Biochimie* 93:1897–1904
5. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10:126–139
6. Curtis HJ, Sibley CR, Wood MJA (2012) Mirtrons, an emerging class of atypical miRNA. *Wiley Interdiscip Rev RNA* 3:617–632
7. Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336:237–240
8. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233
9. Chi SW, Hannon GJ, Darnell RB (2012) An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* 19:321–327
10. Kumar A, Wong AK, Tizard ML et al (2012) miRNA_Targets: a database for miRNA target predictions in coding and non-coding regions of mRNAs. *Genomics* 100:352–356
11. Forman JJ, Legesse-Miller A, Collier HA (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* 105:14879–14884
12. Guo H, Ingolia NT, Weissman JS et al (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466:835–840
13. Lee S, Vasudevan S (2013) Post-transcriptional stimulation of gene expression by microRNAs. *Adv Exp Med Biol* 768:97–126
14. Vasudevan S, Tong Y, Steitz JA (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318:1931–1934
15. Enerly E, Steinfeld I, Kleivi K et al (2011) miRNA–mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* 6:e16915
16. Chi SW, Zang JB, Mele A et al (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* 460:479–486
17. Muniategui A, Pey J, Planes FJ et al (2013) Joint analysis of miRNA and mRNA expression data. *Brief Bioinform* 14:263–278
18. Affymetrix I GeneChip® Gene 1.0 ST Array System for Human, Mouse and Rat. A simple and affordable solution for advanced gene-level expression profiling. Data sheet. http://affy.arabidopsis.info/documents/gene_1_0_st_datasheet.pdf
19. Saced AI, Sharov V, White J et al (2003) TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* 34:374–378
20. Howe E, Holton K, Nair S et al (2010) MeV: MultiExperiment viewer. In: Ochs MF, Casagrande JT, Davuluri RV (eds) *Biomedical informatics for cancer research*. Springer, New York, USA, pp 267–277
21. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15–20
22. Friedman RC, Farh KK, Burge CB et al (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19:92–105
23. Betel D, Wilson M, Gabow A et al (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36(Database issue):D149–D153
24. Enright AJ, John B, Gaul U et al (2003) MicroRNA targets in *Drosophila*. *Genome Biol* 5:R1
25. Lu Y, Zhou Y, Qu W et al (2011) A Lasso regression model for the construction of microRNA–target regulatory networks. *Bioinformatics* 27:2406–2413
26. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29
27. Schaefer CF, Anthony K, Krupa S et al (2009) PID: the pathway interaction database. *Nucleic Acids Res* 37(Database issue):D674–D679
28. Qiagen (2012) miRNeasy Mini Handbook – (EN). For purification of total RNA, including miRNA, from animal and human cells and tissues. <http://www.qiagen.com/Knowledge-and-Support/Resource-Center/Resource-Search/>
29. Hartung T, Balls M, Bardouille C et al (2002) Good cell culture practice. ECVAM good cell culture practice task force report 1. *Altern Lab Anim* 30:407–414
30. Wegener J, Keese CR, Giaever I (2000) Electric cell-substrate impedance sensing (ECIS) as a

- noninvasive means to monitor the kinetics of cell spreading to artificial surfaces. *Exp Cell Res* 259:158–166
31. Agilent Technologies DNA, RNA, protein and cell analysis. Agilent 2100 bioanalyzer. Lab-on-a-chip technology. http://www.icmb.utexas.edu/core/DNA/Information_Sheets/Bioanalyzer/bioanalyzer.pdf
 32. Qiagen (2011) miScript PCR System Handbook. For real-time PCR analysis of microRNA using SYBR Green detection. <http://www.qiagen.com/Knowledge-and-Support/Resource-Center/Resource-Search/?q=miScript+PCR+System+Handbook%3b&cl=en%3b>
 33. Dufva M (2009) Introduction to microarray technology. *Methods Mol Biol* 529:1–22
 34. Affymetrix I GeneChip® Whole Transcript (WT) Sense Target Labeling Assay. User Manual. http://www.affymetrix.com/estore/browse/products.jsp?categoryIdClicked=&productId=131467#1_3
 35. Affymetrix I (2008) GeneChip® Expression Wash, Stain and Scan User Manual. For Cartridge Arrays. http://www.affymetrix.com/estore/browse/products.jsp?categoryIdClicked=&productId=131467#1_3
 36. Risso D, Massa MS, Chiogna M et al (2009) A modified LOESS normalization applied to microRNA arrays: a comparative evaluation. *Bioinformatics* 25:2685–2691
 37. Meyer SU, Kaiser S, Wagner C et al (2012) Profound effect of profiling platform and normalization strategy on detection of differentially expressed microRNAs—a comparative study. *PLoS One* 7:e38946
 38. Brazma A, Parkinson H, Sarkans U et al (2003) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31:68–71
 39. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210
 40. Brazma A, Hingamp P, Quackenbush J et al (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29:365–371
 41. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98:5116–5121
 42. van der Auwera I, Limame R, van Dam P et al (2010) Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. *Br J Cancer* 103:532–541
 43. Bustin SA, Benes V, Garson JA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55:611–622
 44. Vandesompele J, Preter K, de Pattyn F et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034
 45. Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 64:5245–5250
 46. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25:402–408
 47. Agilent Technologies GeneSpring GX 9.0 QuickStartGuide. http://www.chem.agilent.com/library/usermanuals/Public/GeneSpringGX9_QuickStartGuide.pdf
 48. Applied Biosystems Megaplex™ Pools For microRNA Expression Analysis. Protocol. http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/general-documents/cms_053965.pdf
 49. Meyer SU, Pfaffl MW, Ulbrich SE (2010) Normalization strategies for microRNA profiling experiments: a “normal” way to a hidden layer of complexity? *Biotechnol Lett* 32:1777–1788