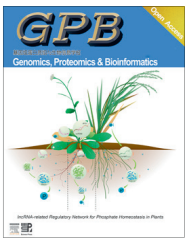
Genomics Proteomics Bioinformatics 13 (2015) 159–168

HOSTED BY

ELSEVIER

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com



ORIGINAL RESEARCH

Regulatory MicroRNA Networks: Complex Patterns of Target Pathways for Disease-related and Housekeeping MicroRNAs

Sachli Zafari [1,a], Christina Backes [1,b], Petra Leidinger [2,c], Eckart Meese [2,d], Andreas Keller [1,*,e]

[1] *Chair for Clinical Bioinformatics, Saarland University, University Hospital, Saarbrücken 66123, Germany*
[2] *Department of Human Genetics, Saarland University, University Hospital, Homburg 66421, Germany*

Received 28 January 2015; revised 27 February 2015; accepted 27 February 2015
Available online 11 July 2015

Handled by William C.S. Cho

KEYWORDS

miRNA;
miRNA targeting;
Systems biology;
Regulatory networks;
miRNA housekeepers

Abstract Blood-based **microRNA** (miRNA) signatures as biomarkers have been reported for various pathologies, including cancer, neurological disorders, cardiovascular diseases, and also infections. The regulatory mechanism behind respective **miRNA** patterns is only partially understood. Moreover, "preserved" **miRNAs**, *i.e.*, **miRNAs** that are not dysregulated in any disease, and their biological impact have been explored to a very limited extent. We set out to systematically determine their role in **regulatory networks** by defining groups of highly-dysregulated **miRNAs** that contribute to a disease signature as opposed to preserved housekeeping **miRNAs**. We further determined preferential targets and pathways of both dysregulated and preserved **miRNAs** by computing multi-layer networks, which were compared between housekeeping and dysregulated **miRNAs**. Of 848 **miRNAs** examined across 1049 blood samples, 8 potential housekeepers showed very limited expression variations, while 20 **miRNAs** showed highly-dysregulated expression throughout the investigated blood samples. Our approach provides important insights into **miRNAs** and their role in **regulatory networks**. The methodology can be applied to systematically investigate the differences in target genes and pathways of arbitrary **miRNA** sets.

Introduction

MicroRNAs (miRNAs) are a class of small non-coding RNAs consisting of around 22 nucleotides. They are known to regulate translations post-transcriptionally [1–3]. There is ample evidence that miRNAs play crucial roles not only in physiological but also in pathological processes [2,4,5]. Depending on the genes that are targeted by selected miRNAs, they can

* Corresponding author.
E-mail: andreas.keller@ccb.uni-saarland.de (Keller A).
[a] ORCID: 0000-0003-3054-1935.
[b] ORCID: 0000-0001-9330-9290.
[c] ORCID: 0000-0002-7778-1437.
[d] ORCID: 0000-0001-7569-819X.
[e] ORCID: 0000-0002-5361-0895.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

http://dx.doi.org/10.1016/j.gpb.2015.02.004

either act as tumor suppressors or as oncogenes. Using high-throughput profiling, dysregulation of miRNAs has been widely observed in cancers at different stages, not only in the tissue but also in body fluids such as whole blood or serum [6–8]. Furthermore, miRNA expression can contribute to the maintenance of cancer stem cells, as proposed by Lu and co-workers in 2005 [9]. Furthermore, miRNAs were more accurate compared to mRNAs in classifying poorly-differentiated tumors [9]. Over the last years, miRNAs have been increasingly proposed as specific and sensitive non-invasive biomarker candidates [10]. A key advantage of miRNA signatures as compared to other RNAs is their stability in various body fluids such as peripheral blood, urine, and saliva [11,12]. More importantly, bioinformatics data suggest that each single miRNA potentially regulates very large sets of genes (even hundreds of putative gene targets are known) [13]. This highlights the potential influence of miRNAs on almost every metabolic and regulatory pathway. miRNAs potentially regulate more than 60% of mammalian mRNAs [10,14]. Several thousand interactions between miRNAs and mRNAs have already been validated by reporter assays. These findings underline the importance of an integrative network analysis of miRNAs and mRNAs [13,15]. The general concept of our previous studies was the interpretation of dysregulated miRNAs in small case–control studies. First, we extended this concept to general disease markers. A meta-analysis revealed miR-144* as miRNA that seems to be generally altered in diseases in the expression abundance [16]. In this study, we extended the concept in two directions. Besides focusing just on dysregulated miRNAs, we were also interested in the miRNAs whose expression is not altered in patients with different diseases. Secondly, we extended the approach from a pure miRNA perspective toward a systemic view including target genes and pathways. Respective networks of the unaltered miRNAs may, for instance, be of special interest for maintaining physiological processes of living cells.

We implemented a computational pipeline integrating analysis and visualization of miRNA, mRNA and pathway data. To find miRNA–mRNA networks, we evaluated expression profiles composed of 848 human miRNAs, which were previously determined by microarrays of 1049 blood samples [8]. The data set contains profiles from patients with 19 different diseases including 10 types of cancer and healthy controls. All signatures were obtained from whole blood collected in PAXgene blood tubes (BD, USA). We set out to define groups of highly-dysregulated miRNAs that show altered expression among different diseases in comparison with healthy controls, contributing most to a disease signature, and groups of preserved miRNAs that show limited expression variations. We next tried to build networks specific for the dysregulated miRNAs and networks specific for the miRNAs with very low expression variation [8]. To this end, we also determined preferential targets and pathways of both dysregulated and preserved miRNAs using validated interactions between miRNAs and mRNAs. Beyond the generation of general networks for preserved and dysregulated miRNAs, we further tested the applicability of our approach to diseases affecting specific organs. Our largest data collection for this task has been done for diseases affecting the lung. We thus analyzed and compared miRNA data from patients with lung cancer or chronic obstructive pulmonary disease (COPD) using our novel approach.

Results and discussion

Determination of preserved and dysregulated miRNAs

The first goal of this study was to identify miRNAs that show a significantly-altered expression in different diseases and healthy controls, and miRNAs that by contrast show very low expression variability in diseases and in controls. The two groups of miRNAs are referred to as "dysregulated" or "preserved" miRNAs from now onward. miRNA profiles have been obtained from 1049 blood samples in total [8], all of which were collected in PAXgene blood tubes using the same protocol. In each case, the profiles were generated from the expression data of 848 miRNAs as previously determined by microarray analysis [7,8]. For discovering dysregulated miRNAs, multiple biostatistical or information theory based approaches are used. Moreover, different cut-off values for the respective methods are applied (*e.g.*, different alpha levels for hypothesis tests) to define dysregulated miRNAs. The different approaches combined with different cut-offs potentially have a substantial influence on the set of dysregulated miRNAs. Since we do not want to perform all downstream calculations using just on one criterion, we evaluated different measures (such as adjusted P value and area under the receiver operator) with different threshold values and performed pathway analyses for the different sets. Specifically, we used t-test and area under the curve (AUC) values to identify dysregulated or preserved miRNA sets. By applying 20 different thresholds for P values or AUC values (summarized in Table S1), we obtained 20 different sets of dysregulated and preserved miRNAs. To minimize the influence of false positive miRNAs that result from using a specific criterion or threshold value, both the preserved and the dysregulated miRNAs were analyzed for their involvement in functional categories. To this end, the dysregulated and preserved miRNA criteria were each analyzed by Tool for Annotations of miRNAs (TAM) that statistically evaluates whether a set of miRNAs shows a significant enrichment in a functional category. For each of the analyzed TAM categories (Table S2), we counted the number of significant associations with each of the two miRNA groups ($P < 0.05$). The functional categories that were significant for at least two of the 20 criteria are summarized in **Table 1** [17,18], including the number of significant associations in the 20 criterion categories. If functional categories were found for at least two preserved and two dysregulated miRNA criteria in a similar frequency, they were considered also for both miRNA groups (one example is the category angiogenesis, which is found with three preserved and three dysregulated miRNA sets). By this consensus approach we defined the two final sets of dysregulated and preserved miRNA categories. All further analyses were done on the consensus pathways that are presented in Table 1.

To further check the validity of the different sets, we carried out non-parametric permutation tests, *i.e.*, we performed the same calculations with randomly-assigned distributions in cases and controls. Moreover, we also performed a stability analysis. As an example for permutation tests, we here describe the results obtained for randomly-permuted data that have been evaluated with the t-test. In more detail, the original P values of the t-test for each miRNA were compared to 1000 P values obtained for 1000 permutations of class labels. For

Table 1 Significant miRNA functional categories for dysregulated and preserved miRNAs

miRNA functional categories	No. of significant associations with dysregulated miRNAs	No. of significant associations with preserved miRNAs
Akt Pathway	4	
Angiogenesis	**3**	**3**
Apoptosis	2	–
Bone regeneration	3	–
HIV latency	4	–
Hormones regulation	6	–
Human embryonic stem cell regulation	6	–
Cell proliferation [17]	3	–
Muscle development	2	–
Immune system [18]	6	–
Immune response	**2**	**6**
mRNA tumor suppressors	2	–
Onco-miRNAs	**5**	**3**
Cell cycle related	–	6

Note: Different sets of dysregulated and preserved miRNAs were obtained by applying 20 different thresholds for *t*-test, *P* values or AUC values (summarized in Table S1). miRNA groups were analyzed by TAM to evaluate whether a set of miRNAs shows a significant enrichment in a functional category and number of significant associations with each of the two miRNA groups was counted. The functional categories that were significant for at least two of the 20 criteria are summarized here. Categories that were found with at least two sets significantly enriched for similar frequency are highlighted in bold.

the original data, 9.4% of miRNAs remained significant after adjustment for multiple testing, whereas 0.08% miRNAs were discovered for the permutation tests. To test the stability, we left out single diseases and repeated all calculations. Here, the median correlation with the original data was 0.99 with the standard deviation of 0.03, indicating a high stability when leaving out single diseases from the analysis.

According to our consensus approach, dysregulated miRNAs were associated with 13 significant categories and preserved miRNAs were associated with four categories only. Furthermore, categories including hormone regulation, human embryonic stem cell regulation, the cell cycle, and the immune system were significantly associated with six different miRNA criteria each. Associations with both dysregulated and preserved miRNAs were also revealed for some functional categories, including the immune response, onco-miRNAs, and angiogenesis ($P < 0.05$).

To determine dependencies between the identified miRNAs, we calculated a network with the miRNAs as the first layer and the functional categories as the second layer of the network. For the dysregulated miRNAs, the network consists of 89 nodes, including 77 miRNAs and 12 functions, and 206 edges (*i.e.*, connections between functions and miRNAs). For the preserved miRNAs, the network consists of 56 nodes, including 53 miRNAs and three functions, and 80 edges between them. To discover potential key players, we calculated the networks with the highest average degree (highest number of edges relative to the number of nodes) by removing miRNAs with a low degree in both networks. This analysis of the dysregulated network resulted in identification of 20 miRNAs including 11 downregulated and 9 upregulated miRNAs (**Table 2**), and 8 miRNAs were found to be preserved by analyzing the preserved network (**Table 3**). The sub-networks for the dysregulated and the preserved miRNAs are shown in Figure S1. The highest number of associations between miRNAs and functional categories was found for the dysregulated miRNAs hsa-miR17 and hsa-miR20a, which were connected to 10 and to 9 different categories, respectively.

Determination of validated target genes of preserved and dysregulated miRNAs

As described in the Materials and methods section, the third layer of the network consists of validated target genes of miRNAs. miRecords [19] has been used as a predicted and experimentally-validated target resource. Since we searched for specific interactions, we used validated gene targets and left out all predicted ones. In total, we identified 58 target genes for 16 of the 20 dysregulated miRNAs and 93 target genes for 4 of the 8 preserved miRNAs (**Table 4**). On average, the 20 dysregulated miRNAs had 2.9 validated targets, while the 8 preserved miRNAs had 11.6 validated targets. The target genes for the miRNAs were, however, not uniformly distributed. For example, miR-16 has 77 targets (Table 4). Correcting for this factor, we recalculated the target distribution by excluding miR-16 and found that the preserved miRNAs showed on average 2.3 miRNAs targets, which is below the average number of the dysregulated miRNAs. A total of 9 target genes were identified both for preserved and for dysregulated miRNAs. These include *JAK1*, *VEGFA*, and *BCL2*, which are known to be regulated by hsa-miR17 and hsa-miR20a. The sub-networks with the preserved and the dysregulated miRNAs along with their target genes are shown in Figure S2.

Again, we carried out permutation tests and randomly selected target gene sets of the same size, counting for each gene how often it was randomly selected in 1000 repetitions. While most genes were not discovered in any or just very few random target sets, *ESR1*, *CCND1*, *BCL2*, and *VEGFA* were discovered in more than 50% of all permutation test. This in fact means that there is a high chance to pick one of these

Table 2 Significantly-dysregulated miRNAs

miRNAs	Node degree	Raw *P* value	Adjusted *P* value	Change of expression
hsa-miR-17	10	1.55E−09	5.04E−08	↓
hsa-miR-20a	9	2.81E−07	3.84E−06	↓
hsa-miR-18a	8	9.84E−05	6.13E−04	↓
hsa-miR-222	8	7.27E−08	1.26E−06	↓
hsa-miR-92a-1*	8	8.56E−05	5.51E−04	↑
hsa-miR-19b	7	2.37E−05	1.78E−04	↑
hsa-miR-106b	5	5.06E−06	5.10E−05	↓
hsa-miR-126	5	1.88E−07	2.75E−06	↓
hsa-miR-24-2*	5	1.05E−04	6.47E−04	↓
hsa-miR-93	5	1.46E−06	1.70E−05	↓
hsa-miR-223	5	2.73E−09	8.26E−08	↑
hsa-miR-93*	5	9.42E−06	8.37E−05	↑
hsa-miR-106a	4	5.43E−08	1.00E−06	↓
hsa-miR-20b	4	1.84E−11	1.11E−09	↓
hsa-miR-26a	4	4.51E−04	2.20E−03	↓
hsa-miR-126*	4	2.24E−15	4.74E−13	↑
hsa-miR-25*	4	1.89E−06	2.08E−05	↑
hsa-miR-26a-1*	4	9.54E−05	5.99E−04	↑
hsa-miR-27a*	4	2.60E−06	2.76E−05	↑
hsa-miR-34a	4	5.35E−08	1.00E−06	↑

Note: Node degree refers to the number of edges connected to the node. Dysregulated miRNAs with downregulated and upregulated expression are indicated with ↓ and ↑, respectively. Asterisk indicates a miRNA with lower expression than its counterpart, when two mature miRNAs originate from the opposite arms of the same pre-miRNA. miRNAs are ranked based on their degrees in this table.

Table 3 Significantly-preserved miRNAs

miRNA	Node degree	Raw *P* value	Adjusted *P* value
hsa-miR-150*	3	0.87668	0.90883
hsa-miR-16	3	0.78348	0.84635
hsa-miR-19a*	3	0.65080	0.75345
hsa-miR-92a	3	0.65099	0.75345
hsa-miR-15a*	3	0.62203	0.72918
hsa-miR-27b	3	0.54836	0.67198
hsa-miR-21*	3	0.47290	0.60988
hsa-miR-19a	3	0.41498	0.55204

Note: Node degree refers to the number of edges connected to the node. miRNAs are ranked based on their adjusted *P* values. Asterisk indicates a miRNA with lower expression than its counterpart, when two mature miRNAs originate from the opposite arms of the same pre-miRNA.

genes randomly by chance. One reason for this may be that the respective genes are targeted by many different miRNAs. For example, miRecords lists 23 entries for *VEGFA*.

Calculation of target gene pathways

The fourth layer of the network consists of the pathways of validated target genes. To find pathways significantly enriched for the target genes, we applied our gene set analysis toolkit GeneTrail [20]. All layers that we calculated are summarized in **Figure 1**. With respect to the fourth layer we found significant enrichment for 20 and 12 KEGG pathways for the dysregulated and preserved miRNAs, respectively (0.3 pathways per target for dysregulated miRNAs and 0.13 pathways per target on average for preserved miRNAs). Of these, 11 were significant for both dysregulated and preserved miRNAs, leaving 9 pathways specific for targets of dysregulated miRNAs and a single pathway, namely NOD-like receptor signaling, specific for the targets of preserved miRNAs (Figure 1). Interestingly, a previous work [8] suggested that miRNAs that are not dysregulated between cases and controls do not regulate biochemical pathways. Our data however suggest that there might be significant enrichment for target genes of preserved miRNAs on specific pathways. This might be however due to a more stringent definition of preserved miRNAs in this work (*t*-test >0.05 and AUC ranging 0.35–0.65, just 12 as compared to 121 in [8]) and the limit to experimentally-validated targets in this study.

Since preserved miRNAs could be considered as a kind of control for dysregulated miRNAs, the 11 pathways shared by preserved and dysregulated miRNAs may potentially represent just a background. As for the dysregulated miRNAs, we carried out permutation tests. Our results showed no significant enrichment for the respective 11 pathways. Nevertheless, given the positive target genes in the permutation test in the

Table 4 Validated gene targets of the significantly-dysregulated and preserved miRNAs

miRNA group	miRNA name	No. of target genes	Name of target genes
Dysregulated	hsa-miR-17	15	*BCL2*, *NCOA3*, *CCND1*, *VEGFA*, *CDKN1A*, *BMPR2*, *RUNX1*, *IL-8*, *BIM*, *p21*, *EDG1*, *VIM*, *MEF2D*, *MAP3K12*, *JAK1*
	hsa-miR-20a	11	*BCL2*, *E2F1*, *CCND1*, *VEGFA*, *RUNX1*, *IL-8*, *BIM*, *BMPRII*, *MEF2D*, *MAP3K12*, *JAK1*
	hsa-miR-222	13	*KIT*, *TIMP3*, *CDKN1B*, *CDKN1C*, *ESR1*, *PPP2R2A*, *PTEN*, *BIM*, *SOD2*, *MMP1*, *FOS*, *BBC3*, *STAT5A*, *p27*
	hsa-miR-126	7	*VCAM1*, *IRS1*, *PIK3R2*, *CRKL*, *VEGFA*, *TOM1*, *PIK3R2*
	hsa-miR-106a	6	*RB1*, *VEGFA*, *RUNX1*, *APP*, *CDKN1A*, *RB1*
	hsa-miR-106b	4	*E2F1*, *VEGFA*, *CDKN1A*, *ITCH*
	hsa-miR-18a	4	*THBS1*, *CTGF*, *ESR1*, *BIM*
	hsa-miR-126*	1	*SLC45A3*
	hsa-miR-223	8	*NFIA*, *IRS1*, *gag-pol*, *LMO2*, *STMN1*, *RHOB*, *FBXW7*, *EPB41L3*
	hsa-miR-27a*	3	*SP1*, *SP3*, *SP4*
	hsa-miR-93	3	*E2F1*, *VEGFA*, *CDKN1A*
	hsa-miR-19b	3	*CTGF*, *ESR1*, *FMR1*
	hsa-miR-93*	2	*CDKN1A*, *VEGFA*
	hsa-miR-20b	2	*VEGFA*, *ESR1*
	hsa-miR-26a-1*	1	*SERBP1*
	hsa-miR-26a	5	*SMAD1*, *PLAG1*, *TGFBR2*, *SERBP1*, *EZH2*
Preserved	hsa-miR-27b	6	*CYP1B1*, *NOTCH1*, *ADORA2B*, *PPARG*, *MMP13*, *ST14*
	hsa-miR-19a	10	*PTEN*, *CTGF*, *THBS1*, *BIM*, *ESR1*, *NR4A2*, *ERBB4*, *CCND1*, *BMPRII*, *TNF*
	hsa-miR-16	77	*TPPP3*, *BCL2*, *VEGFA*, *CCND1*, *PDCD4*, *RAB21*, *CADM1*, *SKAP2*, *WT1*, *RAB9B*, *ACTR1A*, *TPI1*, *CFL2*, *H3F3B*, *MCL1*, *ASXL2*, *C10orf104*, *C14orf109*, *CARD8*, *CDC14B*, *CENPJ*, *CEP63*, *CREBL2*, *ECHDC1*, *CCDC76*, *NPAL2*, *C4orf27*, *C2orf43*, *CCDC111*, *GOLGA5*, *GOLPH3L*, *GTF2H1*, *HACE1*, *HDHD2*, *HERC6*, *C17orf80*, *HRSP12*, *HSDL2*, *HSPA1A*, *JUN*, *PWWP2A*, *FAM122C*, *LOC339804*, *FAM69A*, *MSH2*, *NT5DC1*, *OMA1*, *OSGEPL1*, *PDCD6IP*, *PHKB*, *PMS1*, *PNN*, *PRIM1*, *RAD51C*, *RHOT1*, *RNASEL*, *SLC35A1*, *SLC35B3*, *TIA1*, *HSP90B1*, *UGDH*, *UGP2*, *VPS45*, *WIPF1*, *ZNF559*, *CRHBP*, *CSHL1*, *CDK6*, *CCNE1*, *CCND3*, *HMGA1*, *CAPRIN1*, *RECK*, *BMI1*, *VEGFR2*, *FGFR1*, *VEGF*
	hsa-miR-92a	2	*ITGA5*, *BMPRII*

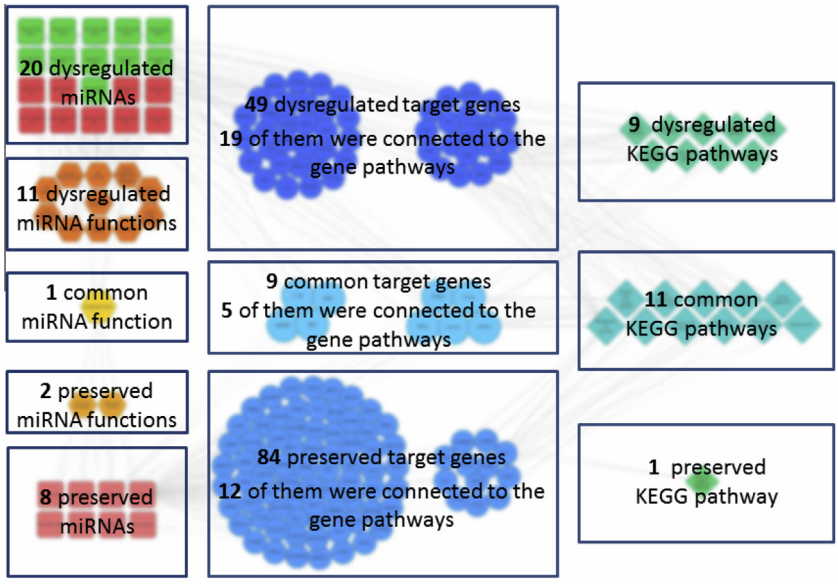


Figure 1 Overview of the full network with both preserved and dysregulated miRNAs
The upper part represents dysregulated miRNAs and the associated functions, target genes, and pathways, while same content is shown for the preserved miRNAs in the lower part. The functions, targets and pathways shared by these two groups of miRNAs are shown in the middle.

preceding subsection, certain likelihood remains that these networks are generated due to noise or are influenced at least partially by non-specific targets. Thus, we compared the significance values instead of differentiating between enriched and not-enriched networks. For the 11 significant pathways that were shared by targets of dysregulated and preserved miRNAs, 9 were more significant for the network of dysregulated miRNAs as compared to the network of preserved miRNAs. An example is "cancer pathways" with P values of 10^{-6} for the targets of preserved miRNAs and of 10^{-12} for targets of dysregulated miRNAs. As summarized in **Figure 2**, our data indicate that the dysregulated miRNAs have a stronger influence on most of the 9 pathways than the preserved miRNAs. Two pathways namely the "p53 signaling cascade" and "focal adhesion" appear to be influenced to a comparable extent by targets both of dysregulated and of preserved miRNAs. Pathways that are downstream of both preserved and dysregulated miRNAs may have a higher inherent robustness and may be less suited to be targeted by exogenous interventions aimed to change disease-related miRNA pathways.

The sub-networks containing target genes and significant pathways are shown in Figure S3. This network shows that only a fraction of all target genes was of relevance for the enrichment analysis. Of the 58 target genes of the dysregulated miRNAs, just 25 targets participated in at least one significant KEGG pathway (43%). This was even more obvious for the 93 target genes of preserved miRNAs with only 17 targets participating in at least one significant KEGG pathway (18%), which suggest that KEGG pathways are more significant in dysregulated miRNAs than preserved miRNAs.

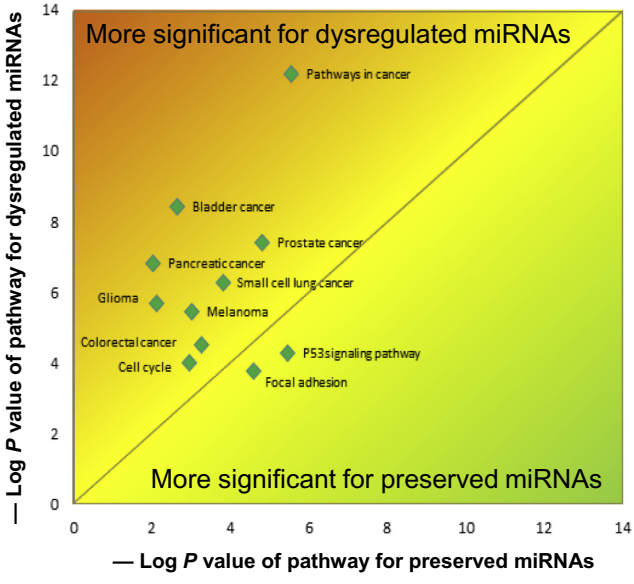


Figure 2 P values of the 11 overlapping KEGG pathways shared by preserved and dysregulated miRNAs

P values of KEGG pathways shared by targets of dysregulated and preserved miRNAs were log-transformed and plotted with the P values for dysregulated miRNAs on the Y axis and preserved miRNAs on the X-axis. The pathways in the upper left corner are more significant in dysregulated miRNAs while the ones in the lower right part are more significant for the preserved miRNAs. The higher fraction of miRNAs in the upper left part indicates that dysregulated miRNAs have a stronger regulatory influence than the preserved miRNAs.

The overall network of preserved and dysregulated miRNAs

We merged the networks for dysregulated and preserved miRNAs to generate a comprehensive visualization including all four layers of miRNAs, miRNA functional categories, validated target genes, and the target gene KEGG pathways. A schematic version of the network is presented in Figure 1, a detailed version of the complete network in Figure S4. In both visualizations, the upper part represents the dysregulated portion of the network, the lower part represents the preserved portion and the middle part represents the overlap of both sub-networks. As aforementioned, we discovered 20 dysregulated miRNAs and 8 preserved miRNAs, which are enriched in 12 and 3 functional categories, respectively, sharing one of them, namely angiogenesis.

The dysregulated miRNAs had 58 target genes, 49 of which were not found for preserved miRNAs. On the other hand, the preserved miRNAs had 93 targets genes, 84 of which are not targets of dysregulated miRNAs. Dysregulated and preserved miRNAs share 9 targets including *CTGF*, *ESR1*, *BMPRII*, *BIM*, *THBS1*, *PTEN*, *BCL2*, *CCND1*, and *VEGFA*. Out of 49 target genes specific to dysregulated RNAs, 19 targets were significantly enriched in KEGG pathways, whereas only 12 out of the 84 target genes specific to preserved miRNAs were enriched in KEGG pathways. Among them, 11 pathways are shared by the preserved and dysregulated miRNAs, which include pathways in cancer, small cell lung cancer, p53 signaling pathway, focal adhesion, cell cycle, bladder cancer, glioma, colorectal cancer, prostate cancer, pancreatic cancer, and melanoma. Except for the pathways for p53 signaling and focal adhesion, the remaining 9 pathways showed a higher significance for the targets of the dysregulated miRNAs as compared to the preserved miRNAs (Figure 2). Especially the pathways for bladder cancer and for cancer in general were substantially more enriched for the targets of dysregulated miRNAs.

Our results indicated that dysregulated miRNAs have a higher regulatory influence as compared to preserved miRNAs. Interestingly, we also found a certain degree of regulation for miRNAs that are stably expressed throughout the blood samples. These miRNAs appear less influenced by diseases and may be of increased relevance for pathways that are central to maintaining basic cell biology functions. Pathways that are downstream of both preserved and dysregulated miRNAs may have a higher inherent robustness and may be less suited to be targeted by exogenous interventions aimed to change disease-related miRNA pathways. The network analysis also indicated that some specific miRNAs are involved in many different functions. For example, miR-17 and miR-20a are relevant in 10 and 9 functional categories with 15 and 11 target genes, respectively, which are relevant for disease pathways such as cancer, focal adhesion, or the p53 signaling cascade.

Network for specific diseases

To demonstrate that methodology can be applied to any comparison in case–control studies, we investigated the networks

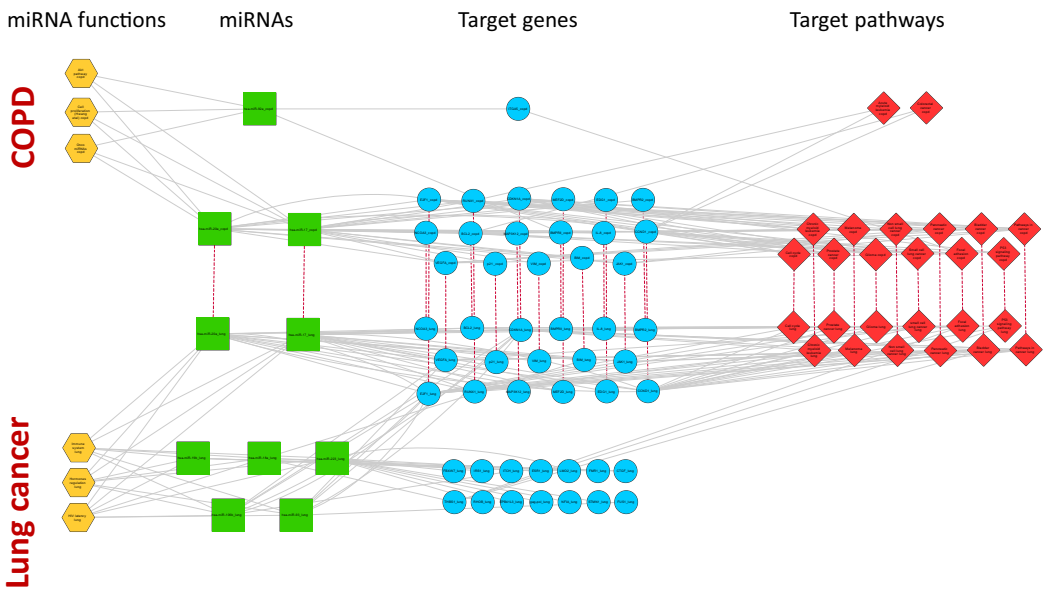


Figure 3 Network alignment between COPD and lung cancer diseases
The upper part represents miRNAs and the associated functions, target genes, and pathways for COPD, while the same content is shown for the lung cancer in the lower part. The functions, targets and pathways shared by these two groups of miRNAs are shown in the middle. Nodes shared by both networks are connected by red dotted lines. COPD, chronic obstructive pulmonary disease.

for lung cancer and for COPD patients. After network was constructed analogously to the methodology described above, we performed a network alignment to find differences in pathways between both diseases (Figure 3). As a result, we identified 5 dysregulated miRNAs, miR-19b, miR-18a, miR-223, miR-106b, and miR-93, which were expressed for lung cancer but not COPD. And conversely, we found that miR-92a was detected for COPD but not lung cancer. We also identified 2 dysregulated miRNAs, miR-17 and miR-20a, both for lung cancer and COPD. Analyses of the target genes specific for either COPD or lung cancer indicated that *ITGA5* was as the only specific target for the COPD, whereas 14 specific target genes were found for lung cancer with 16 target genes shared by both diseases. A literature research in PubMed revealed that the majority of lung cancer-specific target genes, including *IRS1*, *ITCH*, *ESR1*, *CTGF*, *THBS1*, *RHOB*, *STMN1*, and *FUS1*, have previously been related to lung cancer. Notably, for target pathway, we did not find any pathway specific for lung cancer and only two pathways specific for COPD were revealed. The majority of the identified pathways was shared by both COPD and lung cancer. These pathways were mostly identified by the general approach as it was described above for lung cancer ad COPD. Our analysis indicates that the target genes of miRNAs regulate similar pathways in lung cancer and COPD. The main differences between both are found at the target gene level.

Conclusion

Biomarkers for the early and specific detection of various diseases could satisfy an urgent clinical need. To fully explore the value of a biomarker, it is essential to understand its biological role. miRNAs are increasingly recognized as valuable biomarkers for various diseases. Here, we systematically analyzed potential regulatory patterns of miRNAs. We build networks that contained significantly-enriched miRNA functions, validated target genes and target pathways to allow identification of miRNAs with disease-related expression variability, of miRNAs with nearly-unaltered expression in different diseases, and of downstream target genes and targeted pathways of both dysregulated and preserved miRNAs. To demonstrate that our methodology can be applied for specific diseases, we compared blood-borne miRNA profiles of COPD patients to those of lung cancer. The respective analysis results indicate that we can derive specific target genes for the different diseases that may be of relevance for molecular pathological mechanisms. In the case of lung cancer and COPD, target pathways were however similar to each other. The comparative analysis can thus be used to investigate miRNA signatures from different diseases across multiple scales, from single miRNA and target gene level up to enriched pathways.

Materials and methods

Core concept of our approach is to define two sets of miRNAs, dysregulated and preserved miRNAs. Dysregulated miRNAs are those, which are significantly altered in the expression abundance in diseases in comparison to controls. In contrast, preserved miRNAs are those whose expression show a limited variability (see the criteria that we used to define "preserved") between diseases and controls and thus are mostly preserved expressed. In this study, dysregulated and preserved miRNA

data were calculated in comparing profiles of patients to controls using different parameters such as *t*-tests, adjusted *P* value, raw *P* value, and AUC. Based on these data, multi-layer networks have been constructed, analyzed and visualized, partially relying on ideas from previous publications [7,8]. As mentioned, the foundation of this study are 848 miRNAs profiled in 1049 samples [8] from both patients with cancer and non-cancer diseases, and healthy controls, which were obtained from whole blood collected in PAXgene blood tubes (BD, USA). In addition to unaffected individuals (controls), the subjects also included patients with lung cancer, prostate cancer, pancreatic ductal adenocarcinoma, melanoma, ovarian cancer, gastric tumors, Wilms tumor, pancreatic tumors, multiple sclerosis, chronic obstructive pulmonary disease (COPD), sarcoidosis, periodontitis, pancreatitis, or acute myocardial infarction. Detailed information on the data set was presented previously [16].

Microarray measurement and miRNA extraction have been carried out as described previously [7,8]. In brief, we collected blood in PAXgene Blood RNA tubes containing chemistry for stabilizing RNAs. Total RNA was obtained using the miRNeasy kit (Qiagen) and the RNA was stored at −70 °C. The RNA was screened for miRNAs using the Geniom Biochip miRNA *Homo sapiens* (versions v12 and higher). Of each available 3′ and 5′ mature form, seven identical replicates were synthesized on the biochips, of which the median expression was calculated. The unlabeled samples were hybridized for 16 h before on-chip labeling with Biotin was performed. Besides miRNA probes, five controls labeled with Cy3 or biotin were included in the analysis for quality control.

In the following sections we focus on the bioinformatics analyses, including the basic statistical analysis, the step-wise generation of the multi-layer networks and the visualization and comparison of results.

Statistical analysis

The basic statistical analyses, *e.g.*, detection of differentially-regulated miRNAs have been carried out using the publicly-available statistical language R (http://www.r-project.org/). First, standard quantil normalization has been carried out to account for intra-array variations. Following tests for the approximate normal distribution of the 848 miRNAs, two tailed unpaired *t*-tests have been calculated resulting in raw *P* values. These have been adjusted for multiple testing using Benjamini–Hochberg procedure. Beyond the *t*-test, the area under the receiver operator characteristics curve (AUC) with the respective 95% confidence intervals for all patients (cancer and non-cancer diseases) versus all healthy controls was calculated in order to discover potentially-dysregulated miRNAs. To define miRNAs as dysregulated or preserved, different *P* value thresholds and AUC values have been systematically explored. The dysregulated and preserved miRNAs are the nodes of the first layer of the network.

miRNA annotation

In order to find functional categories of dysregulated and preserved miRNAs, TAM [21], a web-based tool for annotating human miRNAs has been applied. In this tool, miRNAs are summarized in various biological categories and TAM statistically evaluates whether a set of miRNAs shows a significant enrichment in the considered categories. Altogether, TAM offers 257 different categories, including many disease associations, functions, families and clusters of miRNAs. Since a substantial part of this functionality is not of relevance for our study, we specifically analyzed 24 miRNA functional categories available in TAM, which are listed in Table S2. The respective categories represent the second layer of the network, edges between a miRNA ***i*** and a category ***y*** were added if the category ***y*** was significantly enriched and ***i*** belongs to that category.

miRNA–target interactions

In order to define interactions between miRNAs and target genes, we evaluated different prediction approaches. Generally, they predicted a weak concordance between each other and importantly also to validated target sets. For these analyses, we parsed all available information from miRecords [19]. In order to achieve the highest certainty in miRNA – mRNA target interactions, we used validated miRNA – mRNA interactions, although validated targets potentially show a bias toward more frequently-studied miRNAs. This bias is at least partially addressed in this study since we did comparative analysis on preserved and dysregulated miRNAs. mRNAs represent the third layer of the network. Edges are added for the respective validated interactions between miRNAs and mRNAs.

Target gene set analysis

After specifying the validated miRNA target genes, the next step was to identify target gene pathways. To do this, we employed GeneTrail [20], a web-based application that allows specifying enriched functional categories and pathways in gene sets. GeneTrail relies on the same basic principle as TAM. Over ten thousand gene categories can be analyzed in an integrative manner. Since we were most interested in biochemical pathways, we focused our analysis on annotated pathways from the KEGG database. The respective pathways represent the fourth layer of the network. Edges between a gene *i* and a pathway *y* were added if the pathway *y* was significantly enriched with target genes and the gene *i* participates in the pathway.

Network analysis, visualization, and alignment

Finally, in order to build the four-layer miRNA and gene interaction networks and to compare the outputs between dysregulated and preserved networks or between lung cancer and COPD, we utilized the open source software Cytoscape [22]. This tool is tailored for visualization of molecular interaction networks and biological pathways and for integration of the networks. Importantly, Cytoscape allows for any type of attribute data to highlight relevant parts of the network. Additionally, Cytoscape also offers a Java-based plugin system that enables easy analysis of networks, starting from very basic algorithms up to complex network alignment tools. The multi-layer networks are schematically presented in **Figure 4**.

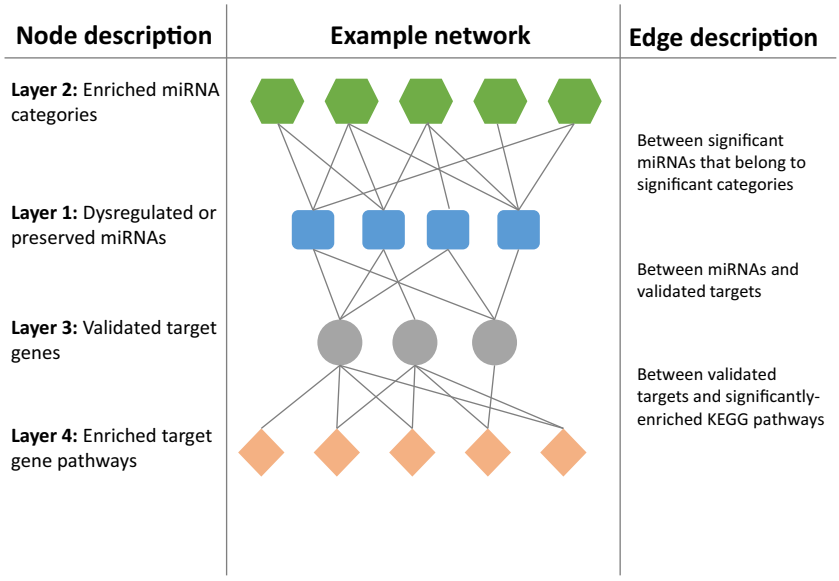


Figure 4 Schema of the multi-layer regulatory network involving miRNAs
A multi-layer network composed of miRNAs (Layer 1) and the associated functional categories (Layer 2), validated target genes (Layer 3), and enriched pathways (Layer 4) is shown as an example network. The four layers are described on the left and connecting edge types in between the corresponding layers are shown on the right.

Authors' contributions

SZ and CB implemented the software package. PL provided data and supported data analysis and interpretation. EM and AK contributed in study set-up, evaluated and interpreted the data, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was funded by the European Union (FP7 BestAgeing, 6031), Saarland University, Medical Faculty. Authors acknowledge the contribution of Comprehensive Biomarker Center (CBC), Heidelberg, in funding the study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2015.02.004.

References

[1] Pasquinelli AE, Hunter S, Bracht J. MicroRNAs: a developing story. Curr Opin Genet Dev 2005;15:200–5.
[2] Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 2004;5:396–400.
[3] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281–97.
[4] Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 2005;433:769–73.
[5] Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006;6:857–66.
[6] Keller A, Leidinger P, Gislefoss R, Haugen A, Langseth H, Staehler P, et al. Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. RNA Biol 2011;8:506–16.
[7] Leidinger P, Backes C, Deutscher S, Schmitt K, Mueller SC, Frese K, et al. A blood based 12-miRNA signature of Alzheimer disease patients. Genome Biol 2013;14:R78.
[8] Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, et al. Toward the blood-borne miRNome of human diseases. Nat Methods 2011;8:841–3.
[9] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834–8.
[10] Etheridge A, Lee I, Hood L, Galas D, Wang K. Extracellular microRNA: a new source of biomarkers. Mutat Res 2011;717:85–90.
[11] Hanson EK, Lubenow H, Ballantyne J. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. Anal Biochem 2009;387:303–14.
[12] Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev 2012;32:326–48.
[13] Backes C, Meese E, Lenhof HP, Keller A. A dictionary on microRNAs and their putative target pathways. Nucleic Acids Res 2010;38:4476–86.
[14] Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009;19:92–105.
[15] Laczny C, Leidinger P, Haas J, Ludwig N, Backes C, Gerasch A, et al. MiRTrail–a comprehensive webserver for analyzing gene

and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases. BMC Bioinformatics 2012;13:36.

[16] Keller A, Leidinger P, Vogel B, Backes C, ElSharawy A, Galata V, et al. MiRNAs can be generally associated with human pathologies as exemplified for miR-144*. BMC Med 2014;12:224.

[17] Hwang HW, Mendell JT. MicroRNAs in cell proliferation, cell death, and tumorigenesis. Br J Cancer 2006;94:776–80.

[18] Xiao C, Rajewsky K. MicroRNA control in the immune system: basic principles. Cell 2009;136:26–36.

[19] Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. MiRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 2009;37:D105–10.

[20] Keller A, Backes C, Al-Awadhi M, Gerasch A, Kuntzer J, Kohlbacher O, et al. GeneTrailExpress: a web-based pipeline for the statistical evaluation of microarray experiments. BMC Bioinformatics 2008;9:552.

[21] Lu M, Shi B, Wang J, Cao Q, Cui Q. TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. BMC Bioinformatics 2010;11:419.

[22] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.