

Toward the blood-borne miRNome of human diseases

Andreas Keller^{1,2,21}, Petra Leidinger^{2,21},
 Andrea Bauer³, Abdou ElSharawy⁴, Jan Haas⁵,
 Christina Backes², Anke Wendschlag⁶, Nathalia Giese⁷,
 Christine Tjaden⁷, Katja Ott⁷, Jens Werner⁷,
 Thilo Hackert⁷, Klemens Ruprecht⁸, Hanno Huwer⁹,
 Junko Huebers¹⁰, Gunnar Jacobs⁴, Philip Rosenstiel⁴,
 Henrik Dommisch¹¹, Arne Schaefer⁴,
 Joachim Müller-Quernheim¹², Bernd Wullich¹³,
 Bastian Keck¹³, Norbert Graf¹⁴, Joerg Reichrath¹⁵,
 Britta Vogel⁵, Almut Nebel⁴, Sven U Jager¹⁶,
 Peer Staehler⁶, Ioannis Amarantos⁶, Valesca Boisguerin⁶,
 Cord Staehler⁶, Markus Beier⁶, Matthias Scheffler⁶,
 Markus W Buechler⁷, Joerg Wischhusen^{17,18},
 Sebastian F M Haeusler¹⁷, Johannes Dietl¹⁷,
 Sylvia Hofmann⁴, Hans-Peter Lenhof¹⁹,
 Stefan Schreiber^{4,20}, Hugo A Katus⁵, Wolfgang Rottbauer⁵,
 Benjamin Meder⁵, Joerg D Hoheisel³, Andre Franke^{4,21} &
 Eckart Meese^{2,21}

In a multicenter study, we determined the expression profiles of 863 microRNAs by array analysis of 454 blood samples from human individuals with different cancers or noncancer diseases, and validated this 'miRNome' by quantitative real-time PCR. We detected consistently deregulated profiles for all tested diseases; pathway analysis confirmed disease association of the respective microRNAs. We observed significant correlations ($P = 0.004$) between the genomic location of disease-associated genetic variants and deregulated microRNAs.

MicroRNAs (miRNAs) can regulate hundreds of genes post-transcriptionally and appear to regulate virtually all cellular processes. Owing to these properties, miRNAs have a critical role not

only in physiological but also in pathological processes¹. Although most reported miRNA expression profiles have been generated from solid tissues, there is growing evidence that miRNA profiles are readily accessible from body fluids, such as blood^{2,3}. The aim of our multicenter study was to elucidate and compare blood expression profiles of 863 miRNAs for different human diseases to test for disease-specific alterations. The generated blood-based 'miRNome' data have been deposited in the Gene Expression Omnibus and updated versions are available at <http://genetrail.bioinf.uni-sb.de/wholemirnomeproject/>. We applied identical standardized experimental and biostatistical procedures to the 454 analyzed blood samples from individuals 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 and from unaffected individuals (controls). All participating centers had to contribute samples to the control group (**Supplementary Table 1**). The different control cohorts had a high degree of reproducibility between the centers (**Supplementary Fig. 1**).

The platform we used is a highly specific primer extension-based microarray that shows a very small degree of cross-hybridization and can be used to distinguish between members of the *let-7* family⁴. To test for technical variance, we repeated the measurements on four samples (two blood samples and two tissue samples) and found a median correlation of 0.97. The correlation between different samples was significantly lower as shown by two-tailed unpaired Wilcoxon Mann-Whitney test ($P < 0.05$) (**Supplementary Fig. 2**). To estimate the biological variance, we analyzed blood samples taken from a healthy individual at three different time points during the day (9 a.m., 12 noon and 3 p.m.), with duplicate measurements at each time. Median correlation between the time points was 0.98 and between duplicates it was 0.99 (**Supplementary Fig. 3**).

On average, we found for each disease 103 deregulated miRNAs ($P < 0.05$; *t*-test after Benjamini-Hochberg adjustment). A total of 62 miRNAs (7.18% of all 863) were deregulated in at least six diseases in comparison to controls (**Supplementary Table 2**), and 24 miRNAs (2.78%) were deregulated in >50% of the 14 analyzed diseases. One miRNA (hsa-miR-320d) was deregulated in 11 diseases and three miRNAs (hsa-miR-423-5p, hsa-miR-146b-3p and hsa-miR-532-3p) were deregulated in nine of the tested

¹Biomarker Discovery Center, Heidelberg, Germany. ²Institute of Human Genetics, Saarland University, Medical Faculty, Homburg, Germany. ³German Cancer Research Center, Functional Genome Analysis, Heidelberg, Germany. ⁴Institute of Clinical Molecular Biology, Christian Albrechts University, Kiel, Germany. ⁵Department of Internal Medicine, University of Heidelberg, Heidelberg, Germany. ⁶febit group, Heidelberg, Germany. ⁷Department of General Surgery, University of Heidelberg, Heidelberg, Germany. ⁸Department of Neurology, Charité University Hospital, Berlin, Germany. ⁹Department of Cardiothoracic Surgery, Voelklingen Heart Center, Voelklingen, Germany. ¹⁰Department of Pneumology, Voelklingen Lung Center, Voelklingen, Germany. ¹¹Department of Periodontology, Operative and Preventive Dentistry, University Hospital Bonn, Bonn, Germany. ¹²Department of Pneumology, University Hospital Medical Center, Freiburg, Germany. ¹³Department of Urology, University Clinic, Friedrich-Alexander University, Erlangen-Nuremberg, Nuremberg, Germany. ¹⁴Department of Pediatric Hematology and Oncology, Saarland University, Medical Faculty, Homburg, Germany. ¹⁵Clinic for Dermatology, Venerology and Allergology, Saarland University, Medical Faculty, Homburg, Germany. ¹⁶Praxis für Dermatologie, Sulzbach, Germany. ¹⁷Department of Obstetrics and Gynecology, Medical School, University of Wuerzburg, Wuerzburg, Germany. ¹⁸Interdisciplinary Center for Clinical Research, Junior Research Group 'Tumor progression and immune escape', Medical School, University of Wuerzburg, Wuerzburg, Germany. ¹⁹Center For Bioinformatics, Saarland University, Saarbruecken, Germany. ²⁰First Medical Department, University Clinic Schleswig-Holstein, Kiel, Germany. ²¹These authors contributed equally to this work. Correspondence should be addressed to A.K. (ack@bioinf.uni-sb.de).

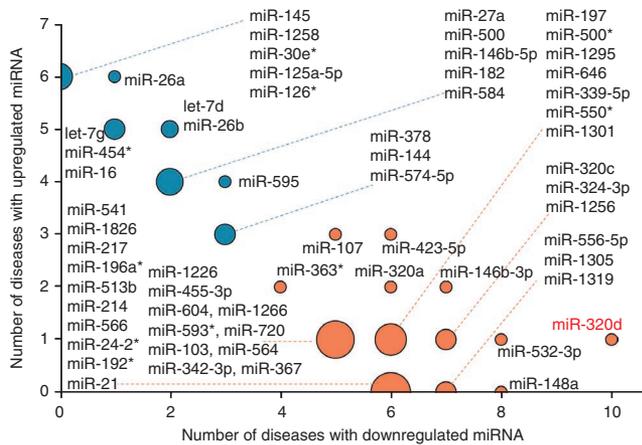


Figure 1 | Bubble plot of miRNAs that are up- or downregulated in several diseases. Bubble sizes correspond to the number of deregulated miRNAs. Orange bubbles denote miRNAs that are more often significantly down-regulated ($P < 0.05$) than upregulated. Blue bubbles denote miRNAs that are either more often upregulated or equally frequent up- and downregulated. *Homo sapiens* (hsa)-miR-320d was significantly deregulated ($P < 0.05$) in 11 diseases.

diseases. Known properties of these miRNAs are listed in **Supplementary Table 2**. Most miRNAs were consistently deregulated, that is, they were either up- or downregulated in the majority of diseases (**Fig. 1**). Analysis of the human microRNA disease database⁵ revealed that only a few of the miRNAs deregulated in blood were also previously reported as deregulated in solid tissues derived from individuals with the same diseases (**Supplementary Table 3**). A total of 121 miRNAs (14%) were not deregulated in any of the 14 analyzed diseases.

We carried out pathway analysis of putative target genes for miRNAs that were deregulated in at least six of 14 diseases ($n = 62$) and for miRNAs that were not deregulated in any disease ($n = 121$). We extracted the targets with $P < 0.001$ for both miRNA sets using GeneTrail^{6,7}. We found a total of 7,598 target genes for both miRNA sets. Of these genes, 27% were targets of miRNAs in both sets, 21% were targets of miRNAs that were frequently deregulated and 52% were targets of miRNAs that were not deregulated in our study. We applied an over-representation analysis relying on the hypergeometric distribution using GeneTrail to find significantly enriched ($P < 0.05$) biochemical pathways. For the set of frequently deregulated miRNAs, we found several disease-associated pathways (**Supplementary Table 4**) including 'pathways in cancer'. We did not detect any enriched pathway for the target genes of the 121 miRNAs that were not significantly deregulated in any disease. Pathways with significantly fewer ($P < 0.05$) targets than expected are indicated in **Supplementary Table 4**.

To explore whether the significantly deregulated miRNAs are in close genomic physical proximity to known susceptibility variants, we extracted 3,495 published single-nucleotide polymorphisms (SNPs) from the US National Institutes of Health genome-wide association study catalog (accessed 28 July 2010) and searched for the coding sequence of miRNAs in a genomic window of 250 kilobases (kb) around these SNPs. We detected 241 cases of physical proximity between SNPs and miRNAs. Of these, seven were related to diseases included in our study, representing interesting candidates for testing the hypothesis that miRNA deregulation depends on nearby genetic variants. Of the seven

SNPs, four are associated with heart diseases, including cardiac structure and function (rs7910620) and mean platelet volume (rs2393967, rs10914144 and rs10506328), two with multiple sclerosis (rs703842 and rs17445836) and one with melanoma. Notably, the relevant miRNA was significantly deregulated ($P < 0.05$) in the same disease, in six of the seven cases. To test whether these results could occur by chance, we carried out 10^6 non-parametric permutation tests. The proximity of genetic variants and deregulated miRNAs was significant ($P = 0.004$). All pairs of SNPs and adjacent miRNAs are summarized in **Supplementary Table 5** and one representative example is presented in **Figure 2**.

To distinguish individuals with disease from controls or from individuals with other diseases by miRNA profiling, we applied machine-learning techniques. Each of the 14 diseases was separated from controls with an average accuracy of 88.5%, ranging from at least 81.3% to up to 100% (**Supplementary Table 6**). By using only two miRNAs, we obtained an average accuracy of 72.5%, whereas the use of ten miRNAs resulted in an average accuracy of 80.6% ($P = 0.0002$, two-tailed unpaired Wilcoxon Mann-Whitney test) (**Supplementary Fig. 4**). Next, we performed pair-wise classification analyses between different diseases using samples collected at the same site to exclude between-institution bias. For the separation between pancreatic cancer and other pancreatic diseases, the accuracy was not significant ($P > 0.05$). However, this result does not necessarily imply a general similarity between miRNA profiles of malignant and nonmalignant diseases of the same organ. For example, we could distinguish lung cancer from COPD with an accuracy of 91.7%, corresponding to a highly significant classification ($P < 10^{-6}$). COPD is a common co-morbidity of lung cancer and also precedes tumors in 50–90% of cases⁸. Thus, a biomarker separating individuals with lung cancer from those with COPD but without cancer may prove useful.

We performed an independent validation of the miRNA profiles using different technologies and cohorts of individuals. In previous studies, we had confirmed 474 deregulated miRNAs in different diseases by performing quantitative real-time PCR (qRT-PCR) on samples from several individuals with lung cancer, melanoma, glioma and acute myocardial infarction^{8–11}. Here we additionally performed a large-scale validation for a larger dataset including data for 44 individuals with lung cancer and 41 with COPD. We selected 18 significantly deregulated ($P < 0.05$) miRNAs that separate both diseases in quadruplicate by qRT-PCR using the SmartChip

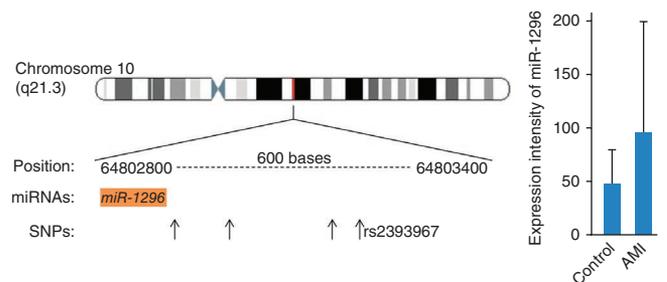


Figure 2 | Representative example for the physical proximity of a significantly deregulated miRNA and a known SNP. A schematic of the human chromosome 10q21 with *hsa-miR-1296* (magenta) and four SNPs (arrows) including SNP rs2393967 (SNP database (dbSNP) accession number) that is associated with heart diseases. The plot shows expression and s.d. of *hsa-miR-1296* in the blood of individuals with acute myocardial infarction (AMI, $n = 20$) compared to that in healthy controls ($n = 70$). $P = 0.006$.

Real-Time PCR System (WaferGen Biosystems). Of those 18 miRNAs, we validated 14, that is, these miRNAs were deregulated in a comparable manner in array and qRT-PCR experiments. The remaining four miRNAs were only rarely expressed as indicated by mean threshold cycle (Ct) values >28.5 . In **Supplementary Table 7** we list raw qRT-PCR data and the variance for the replicates. The overall correlation of the quantile normalized qRT-PCR and array results for the 45 analyzed miRNAs (27 miRNAs of previous studies and 18 miRNAs in the present study) was as high as 0.86 (**Supplementary Fig. 5**). We provide scatter plots and fold changes for all tested miRNAs (**Supplementary Table 8**).

We developed the concept of disease probability plots (DPPs) to determine the probability that a miRNA expression profile correctly indicates that an individual has one or several of the analyzed diseases. We computed the probabilities via a regression approach for each individual sample. Analyzing all DPPs, we predicted the correct disease in 67.45% of all individuals (exemplary DPPs are available in **Supplementary Fig. 6**). Assuming that all diseases are almost equally frequent in our dataset, this translates into an over eightfold increased accuracy of disease prediction by miRNA profiling as compared to random guessing.

Although our study supports the idea that blood cells have an miRNA pattern that varies between different diseases, there are several points to be considered when blood miRNA patterns are associated with diseases. Any association between a miRNA pattern and a disease can be confounded by co-morbidity for another disease. Furthermore, blood cells may not contribute equally to an miRNA pattern, with expression variation in a few cell types accounting for most of the pattern. Indeed, as recently shown for 27 different cell populations isolated from normal mouse hematopoietic tissues, different blood cell types have specific miRNA expression patterns¹². Distribution of the complete blood count (CBC) is known to vary in disease, for instance owing to cancers or diseases of the blood¹³ or bone marrow, cancers that spread to the bone marrow, autoimmune disease or side effects of medications. There are also variations in CBC in healthy individuals. It is possible that changes in miRNA profile in disease reflect shifts in the distribution of different blood-cell types. We tested this possibility using principal-component analysis; specifically, we carried out standard principal-component analysis on the expression matrix (<http://genetrail.bioinf.uni-sb.de/wholemironomeproject/>) and computed for each principal component the fraction of the overall data variance. Although it is likely that shifts in cell populations affect the overall miRNA profiles, we observed that even 27 different cell populations, represented by the first 27 principal components with highest variance, can account for only about 60% of the total variance in the miRNA profiles. Taken together, the ability to recognize systematic features in human blood cells and the relatively small normal CBC variation in healthy individuals

provides support for the feasibility of using miRNA expression patterns in peripheral blood as the basis for detection of disease¹³.

Identifying the complex relationships between disease and changes in miRNA expression patterns in blood cells could contribute not only to an understanding of the mechanism behind the pattern and of disease associations but provide insight into the pathological processes because miRNAs in turn influence the expression of thousands of genes.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturemethods/>.

Accession codes. Gene Expression Omnibus: GSE31568.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank F. Flachsbart, B. Noack and B. Loos for support. This work was financially supported by the German Ministry of Research Education (Bundesministerium für Bildung und Forschung 01EX0806), Hedwig Stalter Foundation, Homburger Forschungsförderungsprogramm and by the Deutsche Forschungsgemeinschaft (LE2783/1-1). Infrastructure support was received from the Deutsche Forschungsgemeinschaft cluster of excellence 'Inflammation at Interfaces'.

AUTHOR CONTRIBUTIONS

A.K. initiated the study; E.M., P.R., J.M.-Q., A.B., P.S., V.B., C.S., M.B., M.W.B., J.W., S.F.M.H., J.D., S.S., H.A.K., W.R., B.M., J.D.H. and A.F. designed the study; A.K., P.L., A.E., H.A.K., W.R., B.M., J.D.H., A.F., E.M., S.S. and B.V. wrote the manuscript; A.K., J.H., C.B., A.W., I.A., B.V. and H.-P.L. analyzed data; P.L., A.B., C.T., A.E., N.G., K.O., J.W., T.H., G.J., H.D., A.S., B.W., B.K., N.G., A.N., V.B., B.V., S.H. and B.M. performed experiments; C.T., K.O., T.H., K.R., H.H., J.H., G.J., H.D., A.S., B.W., B.K., J.R., S.U.J., N.G., M.S., M.W.B., J.W. and S.F.M.H. collected samples.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturemethods/>.

Published online at <http://www.nature.com/naturemethods/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Taft, R.J., Pang, K.C., Mercer, T.R., Dinger, M. & Mattick, J.S. *J. Pathol.* **220**, 126–139 (2010).
2. Mitchell, P.S. *et al. Proc. Natl. Acad. Sci. USA* **105**, 10513–10518 (2008).
3. Otaegui, D. *et al. PLoS ONE* **4**, e6309 (2009).
4. Vorwerk, S. *et al. New Biotechnol.* **25**, 142–149 (2008).
5. Lu, M. *et al. PLoS ONE* **3**, e3420 (2008).
6. Backes, C. *et al. Nucleic Acids Res.* **35**, W186–W192 (2007).
7. Backes, C., Meese, E., Lenhof, H.P. & Keller, A. *Nucleic Acids Res.* **38**, 4476–4486 (2010).
8. Keller, A. *et al. BMC Cancer* **9**, 353 (2009).
9. Leidinger, P. *et al. BMC Cancer* **10**, 262 (2010).
10. Meder, B. *et al. Basic Res. Cardiol.* **106**, 13–23 (2011).
11. Roth, P. *et al. J. Neurochem.* **118**, 449–457 (2011).
12. Petriv, O.I., Hansen, C.L., Humphries, R.K. & Kuchenbauer, F. *Cell Cycle* **10**, 2–3 (2011).
13. Garzon, R. *et al. Blood* **111**, 3183–3189 (2008).

ONLINE METHODS

Blood samples. The blood samples were collected and processed from five different institutions working closely together with the Heidelberg Biomarker Discovery Center (<http://www.bdc-heidelberg.com/biomarker-discovery/index.cfm>). The participating centers were the German Cancer Research Center (Deutsches Krebsforschungszentrum), Saarland University, Heidelberg University, Kiel University and Wuerzburg University. Groups at each of these centers provided samples from individuals with disease and from healthy individuals. Blood was extracted using PAXgene Blood RNA tubes (BD).

All blood donors participating in this study gave their informed consent. A complete list of screened samples is provided in **Supplementary Table 1**.

miRNA extraction and microarray screening. A total of 2.5 ml to 5 ml of blood were extracted in PAXgene Blood RNA tubes. The PAXgene Blood RNA tubes ensure stabilization of RNA and hence stabilization of the expression profiles. Blood cells were obtained by centrifugation at 5,000g for 10 min at room temperature (18–25 °C). The miRNeasy kit (Qiagen) was used to isolate total RNA including miRNA from the resuspended blood cell pellet according to the manufacturer's instructions. The eluted RNA was stored at –70 °C.

All samples were shipped overnight on dry ice and analyzed with the fully automated Geniom RT Analyzer (febit biomed) at febit's in-house genomic service department using the Geniom Biochip miRNA *Homo sapiens* version v12 to v14. Geniom biochips consist of a meandering microchannel that forms the so-called 'biochip'. Each biochip can be used to analyze eight different samples independently. The flexible oligomer synthesis is done *in situ* inside the microchannels using a light-directed process. The probes were designed as the reverse complements of the mature miRNA sequences as published in miRBase plus nucleotides at the 5'-end of the capture oligonucleotide as needed for the enzymatic extension (microfluidic primer extension assay; MPEA). For conventional miRNA hybridization assays the reverse complement of the miRNA sequences as published in the miRBase releases version 12.0 to 14.0 (ref. 14) (in total 863 mature miRNAs and miRNA star sequences) were synthesized with seven intraarray replicates⁴.

We mixed 250 ng of total RNA with 1 µl of 5 pM miRNA spike-in mix and dried it in a tabletop speedvac (Univapo 100H). Each RNA pellet was fully resuspended in 25 µl of hybridization buffer and denatured for 3 min at 95 °C. Until the hybridization, the denatured samples were kept on ice. Microarray hybridization was performed using the Geniom RT Analyzer and Geniom miRNA biochips *Homo sapiens*. The samples were loaded automatically and hybridization of unlabeled sample has been carried out for 16 h. On-chip sample labeling with biotin was carried out by MPEA⁴. Therefore, streptavidin R-phycoerythrin conjugate (SAPE) solution, antibody solution, equilibration buffer (1× NEB 2; New England Biolabs), stop buffer (6× SSPE; Applied Biosystems) and enzyme solution were placed into the RT Analyzer. The array equilibration was followed by incubation with enzyme solution. Enzyme incubation was stopped with stop buffer. SAPE staining, signal amplification and detection proceeded fully automated within the Geniom RT Analyzer. All steps from sample loading to miRNA detection were processed fully automatic inside the machine. As internal control standards five different probes labeled with Cy3 or biotin (bio)

were included: 5'-[Cy3]TCACTCATGGTTATGGCAGCACT GC-3' (80 nM), 5'-[bio]GTAGTTCGCCAGTTAATAGTTTTGCG-3' (12 nM), 5'-[bio]TCTTACCGCTGTTGAGATCCAGTTC-3' (4 nM), 5'-[bio]CCCACTCGTGCACCCAAGTATCTT-3' (0.4 nM) and 5'-[bio]CCATCCAGTCTATTAATTGTTGCCG-3' (0.04 nM).

The enzymatic MPEA together with the fully automated handling ensured a high degree of specificity as well as excellent reproducibility.

The detection pictures were evaluated using the Geniom Wizard Software. For each feature, the median signal intensity was calculated. Following a background correction step, the median of the seven replicates of each miRNA was computed. To normalize the data across different arrays, quantile normalization¹⁵ was applied and all subsequent analyses were carried out using the normalized and background subtracted intensity values. Since the miRBase has been upgraded twice in the past year from version 12.0 to version 14, we used for the final data analysis the 863 miRNAs that were consistently present in all three versions. The whole miRNome data are available for download from the project homepage (<http://genetrail.bioinf.uni-sb.de/wholemirnomeproject/>) and in the Gene Expression Omnibus¹⁶.

Statistical analysis. Single miRNA analyses were carried out using *t*-tests (unpaired, two-tailed) after verifying approximate normal distribution using Shapiro-Wilk test. The resulting *P* values were adjusted for multiple testing using Benjamini-Hochberg's adjustment¹⁷. In addition, the area under the receiver characteristic curve was computed.

Supervised classification of samples was carried out using support vector machines (SVM)¹⁸ as implemented in the R e1071 package¹⁹. As parameters of the SVM, we evaluated different kernel methods including linear, polynomial (degree 2 to 5), sigmoid and radial basis function kernels.

To detect miRNAs that contribute most diagnostic information and thus lead to accurate classifications, a subset selection technique has been applied. Specifically, an iterative filter approach based on the *t*-test was carried out. In each iteration, the *s* miRNAs with lowest *P* values were computed on the training set in each fold of a standard tenfold cross-validation, where *s* was sampled in regular intervals between 2 and 500 miRNAs. The respective subset was used to train the SVM and to carry out the prediction of the test samples in the cross validation. To compute probabilities for classes, a regression approach based on the output of the support vectors has been applied. To test for overtraining, nonparametric permutation tests were applied. All computations were carried out using the publicly available R statistical language¹⁹.

To evaluate the classification, we computed accuracy, specificity and sensitivity.

Pathway analysis. To detect biochemical networks that are putatively regulated by disease miRNAs, we carried out a so-called overrepresentation analysis. For a set of miRNAs, we extracted the targets using Genetrail (<http://genetrail.bioinf.uni-sb.de/>) via MicroCosm V5 (<http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/>) that uses the miRanda algorithm. To reduce the number of false positive miRNA targets, we applied a significance value threshold of 0.001 (ref. 6). The set of putative mRNA targets of disease relevant miRNAs was used as input for the web-based gene set analysis tool GeneTrail to find Kyoto Encyclopedia

of Genes and Genomes (KEGG) pathways that are significantly enriched with targets of disease relevant miRNAs²⁰. All significance values were corrected for multiple testing by Benjamini-Hochberg adjustment.

14. Griffiths-Jones, S. *Methods Mol. Biol.* **342**, 129–138 (2006).
15. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. *Bioinformatics* **19**, 185–193 (2003).
16. Edgar, R., Domrachev, M. & Lash, A.E. *Nucleic Acids Res.* **30**, 207–210 (2002).
17. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. *Behav. Brain Res.* **125**, 279–284 (2001).
18. Vapnik, V. *The Nature of Statistical Learning Theory*. 2nd edn. (Springer, New York, 2000).
19. Team, R. R. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2008).
20. Kanehisa, M. & Goto, S. *Nucleic Acids Res* **28**, 27–30 (2000).

