Involvement of microRNA families in cancer

Stefan Wuchty[1],*, Dolores Arjona[2], Serdar Bozdag[3] and Peter O. Bauer[3]

[1]National Center of Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, [2]GeneDX Inc., Gaithersburg, MD 20877 and [3]Neuro-Oncology Branch, National Cancer Institute, National Institutes of Neurological Disorder and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

ABSTRACT

Collecting representative sets of cancer microRNAs (miRs) from the literature we show that their corresponding families are enriched in sets of highly interacting miR families. Targeting cancer genes on a statistically significant level, such cancer miR families strongly intervene with signaling pathways that harbor numerous cancer genes. Clustering miR family-specific profiles of pathway intervention, we found that different miR families share similar interaction patterns. Resembling corresponding patterns of cancer miRs families, such interaction patterns may indicate a miR family's potential role in cancer. As we find that the number of targeted cancer genes is a naïve proxy for a cancer miR family, we design a simple method to predict candidate miR families based on gene-specific interaction profiles. Assessing the impact of miR families to distinguish between (non-)cancer genes, we predict a set of 84 potential candidate families, including 75% of initially collected cancer miR families. Further confirming their relevance, predicted cancer miR families are significantly indicated in increasing, non-random numbers of tumor types.

INTRODUCTION

MicroRNAs (miRs) are small non-coding ribonucleic acids with mature transcripts of 18–25 nucleotides that interact with their gene target coding messenger RNA (mRNA). Such interactions cause mRNA degradation and putatively inhibit translation by direct and imperfect binding to the 3′- and 5′-untranslated regions (UTR). Although they are powerful tuners of mRNA translation themselves, miRs also exert control in combination with other regulatory elements such as transcription factors (1,2).

The elementary role of miRs in gene expression has been indicated in tissue- and organ-specific development (3) and the classification of tumors (4,5). Over-expressed miRs might diminish the expression levels of targeted tumor suppressor genes, reflecting the functionality of an oncomir. In turn, tumor suppressor miRs are known for targeting genes with oncogenic properties and for being either down regulated or deleted in tumor tissue, leading to a higher expression rate of targeted oncogenes (6). Recently, several miRs were identified as being involved in various steps of the metastatic process (7), a different group of cancer-related miRs that are known as metastamiRs.

From a genomic perspective, clusters of miRs are frequently located in common breakpoint regions and genomic areas of amplification and loss of heterozygosity (8). For example, patients suffering from B-cell chronic lymphocytic leukemia (CLL) often show down regulation of miR-15a and miR-16-1. In addition, these miRs are located at chromosome 13q14, a deleted genomic area in more than 65% of patients with CLL (8). In general, such genomic perturbations of miRs have been identified in the expression regulation of tumor-associated genes in many cancer types (6,9–11). Alternatively, genomically unperturbed miR clusters may be dysregulated by transcription factors that govern their expression. For example, transcription factor MYC shows disrupted expression regulation in cancers and transactivates the miR-17-92 cluster. Furthermore, dysregulation of MYC by chromosomal translocation and juxtaposition to an immunoglobulin enhancer causes Burkett's lymphoma (12,13).

Collecting cancer-related miRs from the literature, we used TargetScan (14) to group miRs into families according to their corresponding seed sequences. We showed that miR families containing cancer-related miRs generally tend to target many mRNAs. Although cancer miR families were further enriched among families that targeted many cancer genes, we statistically confirmed that cancer miR families indeed targeted cancer genes on a statistically significant level. Although cancer genes predominantly were highly connected in a web of protein

*To whom correspondence should be addressed. Tel: +1 301 402 9657; Fax: +1 301 480 2288; Email: wuchtys@ncbi.nlm.nih.gov

interactions, cancer miR families reinforced their enrichment among protein hubs, allowing broad access to signaling pathways. Analyzing their impact on signaling pathways, different miR families not only intervene in similar sets of signaling pathways through targeting cancer genes. We also observed that such interaction patterns resembled corresponding targeting patterns of other cancer miR families, potentially indicating a role in cancer. As we found that the sole number of targeted cancer genes alone reasonable classified a families involvement in cancer, we designed a method to predict candidate cancer miR families. Assuming that interaction patterns differed between (non-)cancer genes, we found a set of candidate cancer miR families that largely overlapped with initially collected cancer miR families.

MATERIALS AND METHODS

Cancer miRs and -genes

We collected 35 oncomiRs, 32 metastamiRs (7,15–17) and 42 tumor suppressor miRs (16–21) from the literature. Using family information from TargetScan (14), we assigned oncomiRs to 19 miR families with identical seed sequences, whereas we found 26 families containing 42 tumor suppressor miRs. Furthermore, 32 metastamiRs corresponded to 23 miR families. We used data from HMDD database, collecting experimental information about the involvement of single miRs in more than 90 cancer types from the scientific literature (22). As for cancer genes, we used 496 oncogenes and 876 tumor suppressor genes from the CancerGenes database (23).

miR–mRNA interactions

We assembled 72 770 predicted, conserved interactions between 153 miRNA families and 11 161 mRNAs, using human-specific data from TargetScan (14). Specifically, families were defined as groups of miRs that had the same seed sequence.

Human protein–protein interactions and signaling pathways

As a representative set of human protein–protein interactions, we pooled 73 869 interactions between 11 446 human proteins from Reactome (24), MINT (25) and HPRD (26). As a comprehensive collection of human canonical signaling pathways, we used information about 184 signaling pathways from the PID database (27).

Enrichment

We grouped N miR families according to their number of interactions where each group was represented by $N_{\geq k}$ families with a certain number of targeted mRNAs, $\overline{k}$. Out of a set of N_c families with cancer miRs we calculated the corresponding subset $N_{c,\geq k}$ in each group. As a control, we collected a set of non-cancer miR families, N_{nc}, that was larger than the set of families with cancer miRs, $|N_{nc}| > |N_c|$. Picking a random set of non-cancer control families of equal size N^r_{nc}, $|N^r_{nc}| = |N_c|$,

we defined $E_{c,\geq k} = \frac{|N_{c,\geq k}|}{|N^r_{nc,\geq k}|}$ as the enrichment of families with cancer miRs among families with k mRNA targets.

Using human protein interactions or signaling pathways, we grouped proteins according to their corresponding number k of interactions or involved pathways. Analogously, we determined the set of proteins with a feature i (e.g. being a cancer gene or targeted by a cancer miR family) in each group, $N_{i,\geq k}$. As a null model, we randomly sampled protein sets of equal size $|N_i| = |N^r_i|$ and defined the enrichment of proteins with feature i in a set of $N_{\geq k}$ proteins as $E_{i,\geq k} = \frac{|N_{i,\geq k}|}{|N^r_{i,\geq k}|}$.

Generally, we averaged E over 1000 randomizations. Note, that $E > 1$ points to an enrichment and vice versa (28), whereas the choice of a randomized control set of equal size allows us to obtain a normalized enrichment score, $E_{\{c,i\},\geq k} = 1$ when $k = 1$.

Random Forest algorithm

Random Forests is an ensemble learning method (29) where classification trees are constructed using N different bootstrap samples of the data ('bagging'). In addition, random forests change how trees are constructed by splitting each node, using the best among a subset of M randomly chosen predictors ('boosting'). New data are predicted by aggregating the predictions of N trees.

We represented all interactions between $m = 11\,161$ genes and $n = 153$ targeting miR families as a binary matrix and labeled each gene as (non-)cancer-related. Randomly picking $\sqrt{m}$ out of all m genes, we used $\sqrt{n}$ variables for the construction of each of 1000 trees. Specifically, we ensured that the number of randomly sampled (non-) cancer genes in each tree corresponded to the priors of (non-)cancer classes.

A miR family's i impact on the discrimination process was measured by their corresponding normalized importance defined as $\overline{I_i} = \frac{I_i}{\sigma_i}$, where I_i and σ_i are the importance and standard deviation of family i, respectively. Assessing their statistical significance, we permuted the targets of each miR family by keeping their initial number of targeted mRNAs constant. Subsequently, we calculated randomized, normalized importance values of each miR using the same parameters. Repeating the randomization process 100 times, we constructed null distributions of normalized importance scores for each miR. Fitting such distributions with a Z test, we calculated P values and considered all miR families with a FDR < 0.01 (30).

RESULTS

Characteristics of cancer miR families

We collected a total of 72 miRs from the literature, consisting of overlapping sets of 35 onco-, 42 tumorsuppressor- and 34 metastamiRs (16–21) (Supplementary Table S1). Using TargetScan (14), we grouped miRs in families that shared the same seed sequence. Assigning such cancer-related miRs to such groups, we obtained a total of 47 cancer miR families. On a more detailed level, we obtained 19 families with oncomiRs, whereas we found

26 families with tumor suppressor miRs and 23 miR families with metastamiRs (Supplementary Table S1). Notably, such families overlapped to a considerable degree (inset, Figure 1A).

To identify mRNA targets of miRs, we used computational predictions of interactions from TargetScan (14), assembling 72 770 interactions between 11 161 mRNAs and 153 miR families. Such a set of interactions included 40 of 47 cancer miR families that we collected from the literature. As cancer genes are particularly placed in groups of highly interacting proteins (31), we hypothesized that highly interacting miR families may increasingly harbor cancer miRs. In groups of miR families that interacted with at least k mRNA targets, we determined the corresponding number of families with cancer miRs. As for a non-cancer control set, we selected families that neither involved any literature-curated cancer miRs nor were otherwise reported in a cancer type. Using the HMDD database that collects literature about the involvement of single miRs in more than 90 tumor types (22), we carefully compiled a set of 51

non-cancer control families. Controlling for similar predicted miR target sites, we also demanded that such control miR families show similar seed match conservation levels of their corresponding UTRs compared with cancer miR families. As a conservation metric, we used the number of orthologous UTR sequences from different organisms that indicate conserved matches to the seed sequences of a given miR. Focusing on our sets of cancer and control miR families, we observed similar distributions in Supplementary Figure S1, indicating similar conservation levels. In Figure 1A, we calculated the enrichment of cancer miR families as a function of the number of interactions between miRs of a family and their corresponding mRNA targets. As our control set of non-tumor miR families was larger than the sets of cancer miR families, we selected random subsets of equal size out of 51 non-cancer families. Subsequently, we calculated enrichment as the ratio of the number of families with cancer miRs and non-cancer families in each group. Averaging over 1000 randomizations, we observed that families with onco-, tumor suppressor and
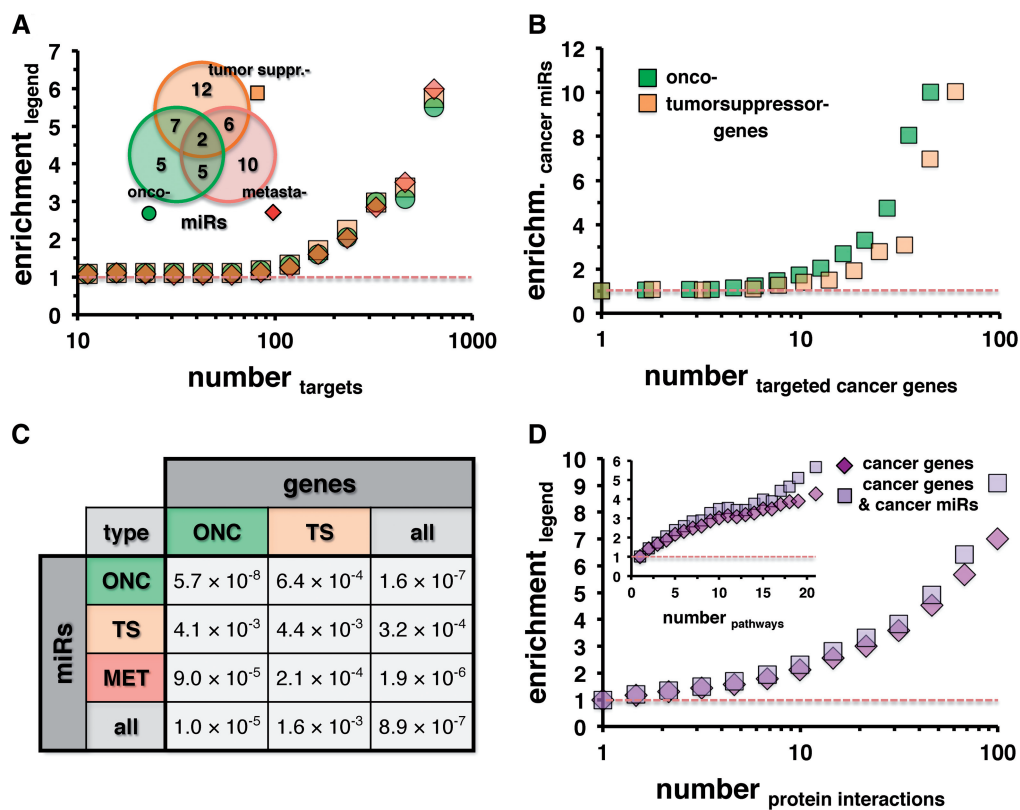

Figure 1. Characteristics of cancer miR families. (**A**) From the literature, we collected cancer miRs that were involved in 43 miR families. Considering onco-, tumor suppressor- and metastamiRs separately, we observed that their corresponding sets of miR families considerably overlapped (inset). In the main figure, we grouped miR families in bins where each family had at least k target genes. Compared with a non-cancer control set of families, we observed that families of onco-, tumor suppressor- and metastamiRs were enriched in groups of families that had an increasing number of targets. (**B**) Analogously, we grouped miR families according to their corresponding number of targeted onco- and tumor suppressor genes, respectively. Compared with a non-tumor control set of miR families, we observed that miR families involved in cancer tend to interact with an elevated number of cancer genes. (**C**) P values of Fisher's exact tests indicated that cancer miR families significantly interacted with cancer genes. (**D**) We determined the enrichment of cancer genes as a function of a gene's number of interaction partners in a human protein interaction network. Specifically, we grouped genes in bins where each gene was at least involved in k interactions. Although the enrichment of cancer genes increased in groups of highly interacting proteins, we observed an enforcement of the initial signal, considering cancer genes that were targeted by cancer miR families. In the inset, cancer-related genes were enriched in groups that appeared in an increasing number of pathways. Similarly, we observed a reinforcement of the initial trend considering cancer genes that interacted with cancer miR families.

metastamiRs were predominately enriched in groups of highly interacting miR families (Figure 1A).

Expecting that cancer miRs exert their influence on the expression of cancer genes through their interactions with corresponding mRNAs, we collected 496 onco- and 876 tumor suppressor genes from the CancerGenes database (23). Hypothesizing that cancer miR families may target many different cancer genes, we determined the enrichment of such families as a function of the number of targeted onco- and tumor suppressor genes. Using our cohort of non-cancer miR families as controls, we observed increasing trends in Figure 1B, suggesting that cancer miR families indeed tend to frequently target cancer genes. Testing the significance that cancer miR families indeed prefer to target cancer genes, we applied Fisher's exact test, allowing us to find a statistically significant tendency of families with onco-, tumor suppressor- and metastamiRs to predominately interact with onco- and tumor suppressor genes (Figure 1C).

As a corollary of these observations, we assumed that the enrichment of cancer genes among protein hubs in networks of interacting proteins may be reinforced by cancer miR families. Assembling a network of 73 869 protein–protein interactions between 11 446 human proteins (24–26), we grouped proteins in bins where each protein had at least k interactions. Pooling all 1259 onco- and tumor suppressor genes, we determined the number of such cancer genes in each group. As a baseline, we considered the enrichment of cancer genes as a function of the corresponding number of interaction partners by randomly picking cancer genes (Figure 1D). Focusing on all cancer genes that were targeted by cancer miR families, we observed an enhancement of the initial enrichment signals. As a consequence, we assumed that such a reinforcement signal may translate into an elevated involvement in pathways, as highly connected proteins tend to appear in an increasing number of pathways (32). Using 184 human signaling pathways from the PID database (27), we grouped proteins in bins where each protein participated in at least k pathways. Determining the number of cancer genes in each group, we randomly sampled sets of cancer genes, allowing us to find a strong enrichment signal in groups of genes that were involved in an increasing number of pathways (inset, Figure 1D). Focusing on cancer genes that were targeted by cancer miR families, we found a reinforcement of the initial trend.

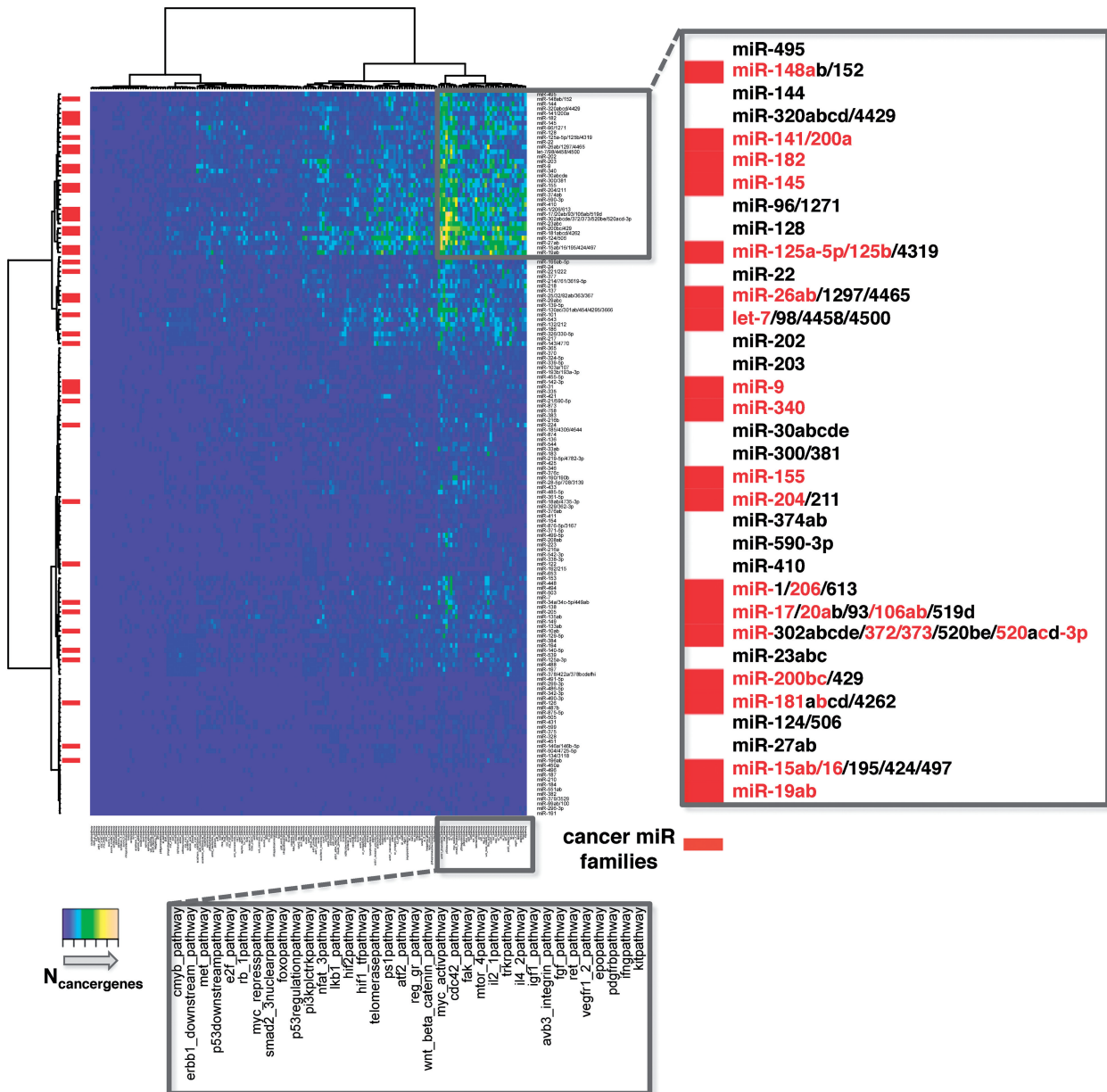### Prediction of families involved in cancer

Providing a bigger picture, we mapped the relationships between signaling pathways and miR families by counting the number of cancer genes in each pathway that were targeted by a given family. Applying ward clustering to such family-specific profiles of pathway intervention (Figure 2), we found a large cluster that significantly pooled a large fraction of cancer miR families and indicated miR families with a strong involvement in signaling pathways. For example, family miR-15ab/16/195/424/497 that involved cancer miRs-15ab/16 showed similar interaction patterns as co-clustered families

miR-27ab and miR-124/506. Such patterns may indicate that the latter families harbored miRs with a role in cancer as well. Assuming that similar interaction patterns of miR families may allow us to find families involved in cancer, we used the number of targeted cancer genes to predict cancer miR families. In a naïve approach, we predicted a miR family's involvement in cancer as a function of their number of targeted cancer genes. Using our set of literature-curated families with cancer miRs, the area under the receiver operating characteristic curve suggested that increased targeting of cancer genes is indeed a classifying criterion for the determination of cancer miR families (AUC = 0.74).

As a corollary, gene-specific profiles of targeting miR families may differ between (non-)cancer genes, prompting us to hypothesize that families which significantly contributed to the discrimination process may be potential families involved in cancer. Designing a simple heuristic, (i) we represented interactions between each of 11 161 genes and all 153 miR families as a binary interaction profile and labeled all cancer genes accordingly. (ii) Applying the random forest algorithm (29), we assessed each families' impact by determining its normalized importance, a measure that reflected the mean decrease in accuracy when the given miR family was unaccounted for in the discrimination process. (iii) To assess the statistical significance of normalized importance values, we permuted miR family profiles and calculated randomized, normalized importance scores. Determining a P value for each family with a Z test, we predicted 84 candidate cancer miR families with FDR < 0.01 (30), including 30 of 40 (75%) literature curated families with cancer miRs (Supplementary Table S2).

Obtaining experimental evidence to assess the biological relevance of candidate miR families, we mined data from the HMDD database (22). In Figure 3B, we constructed a bipartite matrix, indicating if at least one publication reported the involvement of a given miR of a certain family in a tumor type. Ward clustering such a matrix allowed us to find a cluster that accumulated the majority of predicted and curated cancer miR families. For example, we found mentioned families miR-27ab and miR124/506 in this cluster, families that both were predicted as involved in cancer. Nested in this cluster, we observed a small group of 23 miR families that appeared in most cancer types. Specifically, we observed that family miR-30abcdef/30abe-5p/384-5p was not only predicted to be involved in cancer but also appeared in 20 different cancer types. Furthermore, this family targeted numerous cancer genes in signaling pathways (Figure 2), indicating the family's potential to play a role in cancer.

Generally, we observed that both literature collected and predicted cancer miR families seemed to appear in a higher number of tumor types than remaining families in Figure 3B. We, therefore, hypothesized that the placement of miR families in given tumor types was a non-random process. Up to a certain year, we determined the average number of tumor types where literature collected and candidate cancer miR families were reported in. To assess the significance that a set of miR families appeared in an
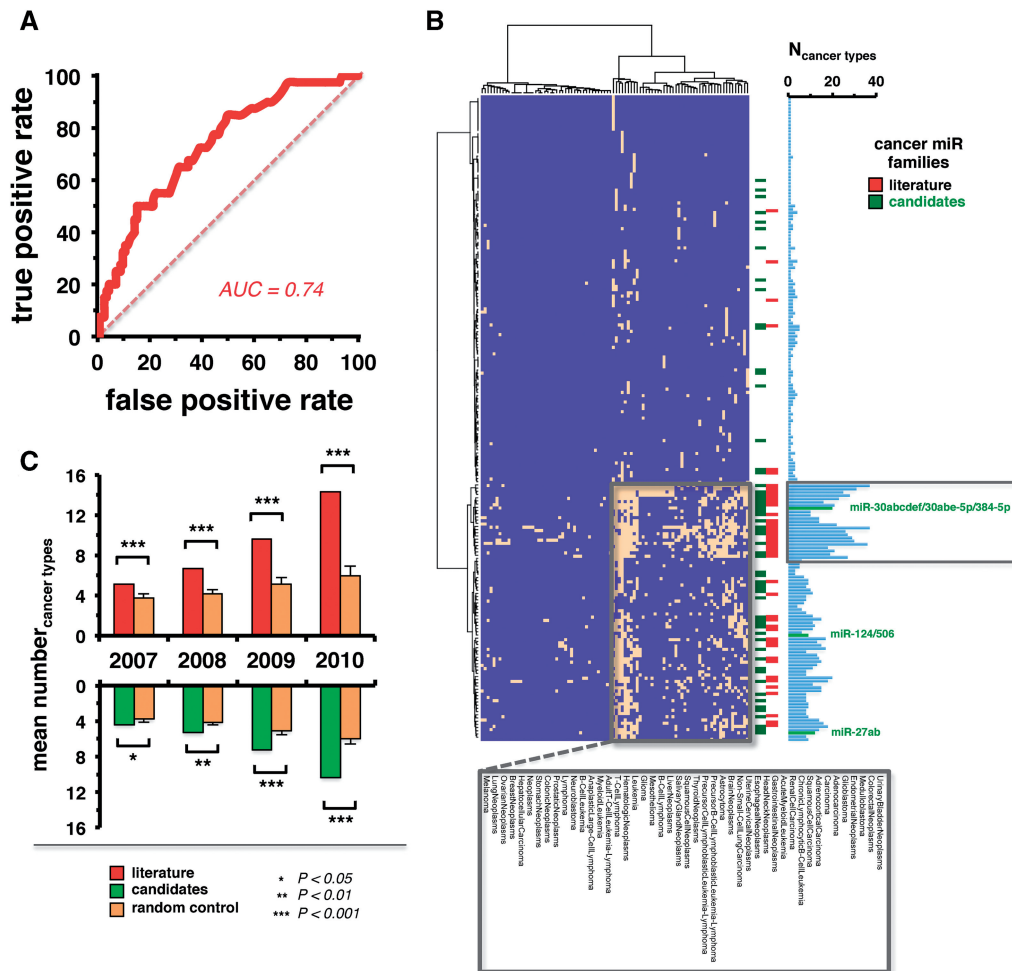
Figure 2. Involvement of miRs in signaling pathways. In the heatmap, we counted the number of cancer genes in a signaling pathway that were targeted by a given family of miRs. Ward clustering such a matrix, we found a large cluster of families that strongly interacted with cancer genes in signaling pathways and significantly pooled a large fraction of cancer miR families. Cancer miRs collected from the literature were indicated in red in their corresponding families.

observed average number of tumor types, we randomly sampled sets of families of equal size out of all 153 targeting families. Subsequently, we calculated corresponding average numbers of tumor types such families appeared in and repeated these steps 10^5 times. After obtaining empirical P values, we observed that literature-collected cancer miR families appeared in a significantly growing average number of tumor types (upper panel, Figure 3C). Analogously, we determined the significance of predicted families, suggesting a similarly increasing and significant trend of candidate families over the last couple of years (lower panel, Figure 3C).

DISCUSSION

Collecting onco-, tumor suppressor- and metastamiRs from the literature, we found that their corresponding miR families were significantly enriched in groups of families with an elevated number of mRNA targets. Such a result coincides with a well-known topological characteristic of cancer genes that are mostly found in groups of highly interacting human proteins. Focusing on onco- and tumor suppressor genes, we significantly observed that cancer miR families preferably targeted an increasing number of cancer genes, allowing us to

Figure 3. Prediction of candidate miR families involved in cancer. (**A**) In a naïve approach, we used the number of targeted cancer genes to detect miR families with cancer miRs. In the corresponding receiver operating characteristic curve, we observed a considerable AUC = 0.74, indicating the power of this simple classifier. (**B**) In a binary matrix, we indicated miR families that were reported in a certain tumor type. Applying ward clustering, we found a pronounced cluster that largely pooled candidate and curated cancer miR families. Furthermore, such families appeared in high numbers of different tumor types. (**C**) Up to a certain year, we determined the average number of tumor types where predicted and curated cancer miR families were reported in. The significance of observed trends over the last years was confirmed by randomly sampling sets of cancer miR families. Analogously, candidate cancer miR families gained considerable significance to appear in different tumor types over the last years, strongly indicating that candidates may indeed play a potential role in cancer.

conclude that such families exerted their biological role through perturbed fine regulation of the expression of cancer genes. As a consequence, enrichment of cancer genes targeted by cancer miR families in groups of highly interacting proteins was indeed enhanced. Such an observation suggests that the interactions between cancer miRs and genes add another layer to the discussion of a genes topological role in cancer. Although cancer genes alone predominately are hubs, their centrality in a cellular network is further emphasized by a small group of highly connected cancer miR families that modulate the expression of the underlying genes.

Notably, we determined enrichment compared with a non-tumor control set of miR families with no involvement in cancer. Furthermore, such a control set also has to address other biases. Specifically, conserved miRs are more widely studied and therefore may have a higher chance to be involved in cancers. Conversely, non-tumor

involved miRs may be non-conserved and may be predicted to interact with a low number of targets. To control for such a bias, we demanded that families in our non-tumor control set were involved in predicted conserved interactions between miRs and mRNAs.

miR families also secure a strong influence in signaling pathways through targeting cancer genes. Specifically, we observed that other miRs may resemble pathway interaction patterns of cancer miRs, suggesting that such patterns indicate novel cancer miR candidates. Resembling interaction patterns may arise from similar seed sequences of single miRs, leading to similar predicted mRNA targets. However, we grouped miRs in families of identical seed sequences, largely offsetting such effects. As a consequence, we observed that families miR-27abc, −124/506, −181abcd/4262 and −200bc/426 shared similar pathway interaction patterns in Figure 2, whereas their seed sequences differed strongly (miR-27abc: UCACAGU,

miR-124/506: AAGGCAC, miR-181abcd/4262: ACAUU CA and miR-200bc/429: AAUACUG).

On the simplest level, the number of targeted cancer genes indeed is a reasonable classifier to call a cancer miR family. However, the prediction of potential candidate cancer miRs would necessitate the determination of an optimized threshold in a supervised way by accounting for already known cancer miR families. Furthermore, such a naïve approach considers miR families as being independent from each other, therefore ignoring any composite effects between families. Considering whole gene-specific interaction profiles, we proposed a simple, unsupervised method, enabling us to predict potential cancer miR families. In particular, our previous observations prompted us to assume that miR interaction profiles of cancer and non-cancer genes considerably differ, allowing us to identify a subset of miR families that distinguishes between (non-)cancer genes. As a proof of concept, we notably found that the predicted and literature collected set largely overlapped, whereas candidate cancer miR families appeared in a non-random, growing number of different tumor types.

Although we do not account for any expression or genomic perturbation data, we solely introduced disease-specific information by choosing disease genes. Assuming that miR interaction profiles of (non-)disease genes may generally differ our method may be applicable to find families that play a role in other diseases as well.

Although miR families collected from the literature are on average indicated in more cancer types, such an observation is putatively caused by researchers bias, as well-studied miRs are usually considered first as experimental leads. However, the increasing number of tumor types candidate cancer miR families are indicated in suggest that predicted families already gained relevance as serious candidates.

As we grouped miRs into families according to their seed sequences, many predicted and literature-curated families can be quite large, suggesting that rather a subset of miRs contribute to cancer than a family of miRs as a whole. Therefore, miR families should be considered as groups that potentially harbor single miRs with a role in cancer. As such, predicted families may well serve as potential leads to find single cancer miRs and may significantly contribute to the determination of diagnostic miR signatures and therapeutic targets in different tumor types.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure 1 and Supplementary Tables 1 and 2.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Martinez,N.J. and Walhout,A.J. (2009) The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *Bioessays*, **31**, 435–445.
2. Shalgi,R., Lieber,D., Oren,M. and Pilpel,Y. (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput. Biol.*, **3**, e131.
3. Houbaviy,H.B., Murray,M.F. and Sharp,P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, **5**, 351–358.
4. Volinia,S., Calin,G.A., Liu,C.G., Ambs,S., Cimmino,A., Petrocca,F., Visone,R., Iorio,M., Roldo,C., Ferracin,M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA*, **103**, 2257–2261.
5. Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H., Ferrando,A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
6. Chen,C.Z. (2005) MicroRNAs as oncogenes and tumor suppressors. *New Engl. J. Med.*, **353**, 1768–1771.
7. Hurst,D.R., Edmonds,M.D. and Welch,D.R. (2009) Metastamir: the field of metastasis-regulatory microRNA is spreading. *Cancer Res.*, **69**, 7495–7498.
8. Calin,G.A., Sevignani,C., Dumitru,C.D., Hyslop,T., Noch,E., Yendamuri,S., Shimizu,M., Rattan,S., Bullrich,F., Negrini,M. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA*, **101**, 2999–3004.
9. Croce,C.M. and Calin,G.A. (2005) miRNAs, cancer, and stem cell division. *Cell*, **122**, 6–7.
10. Ciafre,S.A., Galardi,S., Mangiola,A., Ferracin,M., Liu,C.G., Sabatino,G., Negrini,M., Maira,G., Croce,C.M. and Farace,M.G. (2005) Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem. Biophys. Res. Commun.*, **334**, 1351–1358.
11. Silber,J., Lim,D.A., Petritsch,C., Persson,A.I., Maunakea,A.K., Yu,M., Vandenberg,S.R., Ginzinger,D.G., James,C.D., Costello,J.F. *et al.* (2008) miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med.*, **6**, 14.
12. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
13. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. and Bradley,A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.
14. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
15. Nicoloso,M.S., Spizzo,R., Shimizu,M., Rossi,S. and Calin,G.A. (2009) MicroRNAs—the micro steering wheel of tumour metastases. *Nat. Rev. Cancer*, **9**, 293–302.
16. Schickel,R., Boyerinas,B., Park,S.M. and Peter,M.E. (2008) MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death. *Oncogene*, **27**, 5959–5974.
17. Krutovskikh,V.A. and Herceg,Z. (2010) Oncogenic microRNAs (OncomiRs) as a new class of cancer biomarkers. *Bioessays*, **32**, 894–904.
18. Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
19. Budhu,A., Ji,J. and Wang,X.W. (2010) The clinical potential of microRNAs. *J. Hematol. Oncol.*, **3**, 37.
20. Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
21. Hammond,S.M. (2006) MicroRNAs as oncogenes. *Curr. Opin. Genet. Dev.*, **16**, 4–9.
22. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.

23. Higgins,M.E., Claremont,M., Major,J.E., Sander,C. and Lash,A.E. (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.*, **35**, D721–D726.

24. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.

25. Ceol,A., Chatr Aryamontri,A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.

26. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database— 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.

27. Schaefer,C.F., Anthony,K., Krupa,S., Buchoff,J., Day,M., Hannay,T. and Buetow,K.H. (2009) PID: the pathway interaction database. *Nucleic Acids Res.*, **37**, D674–D679.

28. Wuchty,S. (2004) Evolution and topology in the yeast protein interaction network. *Genome Res.*, **14**, 1310–1314.

29. Breiman,L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.

30. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B.*, **57**, 289–300.

31. Jonsson,P.F. and Bates,P.A. (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.

32. Wuchty,S. and Almaas,E. (2005) Peeling the yeast protein network. *Proteomics*, **5**, 444–449.