

SCIENTIFIC REPORTS



OPEN

Novel miRNA-mRNA interactions conserved in essential cancer pathways

Eduardo Andrés-León^{1,†}, Ildefonso Cases², Sergio Alonso³ & Ana M. Rojas¹

Received: 04 November 2016

Accepted: 08 March 2017

Published: 07 April 2017

Cancer is a complex disease in which unrestrained cell proliferation results in tumour development. Extensive research into the molecular mechanisms underlying tumorigenesis has led to the characterization of oncogenes and tumour suppressors that are key elements in cancer growth and progression, as well as that of other important elements like microRNAs. These genes and miRNAs appear to be constitutively deregulated in cancer. To identify signatures of miRNA-mRNA interactions potentially conserved in essential cancer pathways, we have conducted an integrative analysis of transcriptomic data, also taking into account methylation and copy number alterations. We analysed 18,605 raw transcriptome samples from The Cancer Genome Atlas covering 15 of the most common types of human tumours. From this global transcriptome study, we recovered known cancer-associated miRNA-targets and importantly, we identified new potential targets from miRNA families, also analysing the phenotypic outcomes of these genes/mRNAs in terms of survival. Further analyses could lead to novel approaches in cancer therapy.

Unrestrained cell proliferation is the principal hallmark of cancer, provoked by DNA insults and other events. These insults produce mutations in genes that might alter their expression and/or function, deregulating many physiological pathways and provoke chromosomal damage, events that drive oncogenic transformation and tumour progression¹. A physiological response is triggered by cells to restrain genomic instability by removing or repairing DNA lesions, widely known as the DNA damage response pathway (DDR)². Different levels of damage produce different responses, such that mild DNA damage induces a reparative response, while severe damage activates cell death in a regulated manner. Distinct proteins are involved in these processes. For instance, p53 contributes to the differential expression of pro-survival and pro-apoptosis genes, while ATM/ATR³ promotes cell death when there are many double strand breaks (DSB), activating CASP2, E2F1, P73 or CHEK1. In addition, several pathways are also relevant to restrain oncogenesis, by activating or repressing genes that control cell proliferation and programmed cell death. In particular, the pathways controlling apoptosis, necrosis and senescence are of special interest, as are those involved in the DDR, the cell cycle, DNA replication and telomere elongation, all of which constitute an intricate network in which there is extensive cross-talk that influences events like cancer initiation and progression⁴. Indeed, the intense research carried out into these pathways has helped characterize a large number of mechanisms directly implicated in tumour development⁵.

To further our understanding of cancer and to develop new treatments, several international initiatives have set out to generate complete libraries of tumour types. For example, The Cancer Genome Atlas (TCGA) consortium provides information extracted from a wide variety of cancers using different techniques, housing datasets containing DNA methylation samples, whole genome and whole exome sequences, transcriptome data (RNASeq and small RNASeq), etc. Previous studies identified genes related to cancer malignancy by comparing tumour tissue with normal tissue samples⁶. Such studies yielded relevant genes, of which oncogenes and tumour suppressor genes received most attention as they have the potential to convert a normal cell into a cancer cell. Oncogenes are often mutated or expressed at high levels in tumour cells, while tumour suppressor genes may also be mutated or expressed more weakly than normal. *TP53*⁷ or *FOS*⁸ are examples of tumour suppressor genes where mutations,

¹Computational Biology and Bioinformatics Group, Institute of Biomedicine of Seville, HUVR/US/CSIC, 41013 Sevilla, Spain. ²REDgene Bioinformatics, Sevilla, Spain. ³Program of Predictive and Personalized Medicine of Cancer (PMPPC), Germans Trias I Pujol Research Institute (IGTP), Can Ruti Campus, Badalona 08916, Barcelona, Spain. [†]Present address: Bioinformatics Unit, Instituto de Parasitología y Biomedicina "López Neyra", Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), Granada, Spain. Correspondence and requests for materials should be addressed to S.A. (email: salonsou@igtp.cat) or A.M.R. (email: arojas-ibis@us.es)

deletions and/or repression are implicated in processes related to cancer development. Similarly, the oncogenes *E2F3*³ and *AURKA*⁹ induce cell growth and proliferation when mutated or over-expressed, driving cells through the cell cycle checkpoints.

In the past decade, the role of microRNAs (miRNAs) in cancer and in cell proliferation has gained significance given their critical role in regulating target genes. These are single stranded non-coding RNAs of 19–22 nucleotides commonly involved in mRNA destabilization and degradation^{10,11}. miRNAs from the same family share a high degree of sequence homology and they are commonly found in clusters, the components of which are expressed simultaneously, in turn resulting in a tendency to regulate genes with similar functions¹². Thus, the inappropriate expression of miRNAs that regulate key genes like oncogenes or tumour repressor genes can lead to tumour development, and families such as miR-15, miR-1 and let-7 have frequently been related with cancer growth and metastasis.

The specific miRNAs that are capable of transforming normal cells into tumour cells are known as oncomiRs. For instance, the miR-17/92 cluster, a polycistron RNA that encodes for 6 different miRNAs, was the first oncomiR described. Its over-expression is related to *E2F1*¹³ and *AIB1*¹⁴ down regulation, which in turn produces tumours by inducing cell growth and proliferation. Conversely, there is another subset of miRNAs, the tumour suppressor miRNAs or anti-oncomiRs, among which we can distinguish the miR-143/145 cluster that is transcribed as a bicistronic unit. Relative to normal cells, these miRNAs are expressed more weakly in colon cancer cells and in breast, lung, head and neck, and bladder cancers¹⁵. The miR-143/145 cluster targets a large number of genes and their deregulation is thought to be one of the earliest events in cancer development¹⁶.

The relationships between miRNAs and their mRNA targets are difficult to study experimentally. Current methods are complex, expensive and must overcome technical challenges like tissue specificity, weak expression, 3' UTR selection and miRNA stabilization¹⁷. Some techniques measure mRNA expression upon the modification of miRNA expression using luciferase assays or other proteomic assays like pulsed stable isotope labelling with amino acids in cell culture (pSILAC). These approaches serve to quantify the protein derived from the mRNA regulated by different levels of miRNA¹⁸. Given the known level of complexity, computational predictions of putative miRNA-mRNA targets emerge as a complementary approach to facilitate the experimental characterization of relevant miRNAs. The most frequent procedures are based on seed sequence complementarity, evolutionary conservation and thermodynamic stability¹⁹. Moreover, additional steps can be implemented to reduce the unavoidable number of false positives obtained from software prediction. In this regard, miRNA target prediction can be improved by using inverse correlated profiles from over-expressed miRNAs and down-regulated genes from the same tumour samples, and vice versa²⁰.

One relatively unexplored issue is the potential identification of specific miRNA-mRNA interactions within well-established cancer pathways that are amenable for therapeutic targeting. To investigate whether such interactions are associated with cancer, an integrated analysis of several different tumour types within a comprehensive statistical framework could aid the discovery of specific cancer related miRNA-mRNA signatures. In addition, the analyses of these signatures in different tumours, accounting for additional factors like gene methylation (MET) and copy number alterations (CNAs), could provide relevant information regarding the signature's stability. In the context of survival, such information is likely to be useful to define novel therapeutic strategies²¹.

Results

Genome-wide analysis of individual tumours. Thousands of samples were examined for each of the 15 tumour types analysed here (see Supplementary Table 1 for details), performing differential expression analyses to identify deregulated genes and miRNAs. Only those exhibiting reliable FDR values (FDR < 0.05) and significantly altered expression (absolute log₂FC > 1) were selected for further analysis. Details regarding the analysis of RNASeq and small RNASeq samples from each individual tumour type are described in the Supplementary Text.

Integrative data analysis for the set of 15 tumours. We performed a comprehensive statistical analysis on a set of 15 tumours to identify novel miRNA-mRNA associations that might commonly be deregulated. To this end, we extracted data on differentially expressed genes and miRNAs, comparing samples from healthy subjects and tumour patients from each tumour (see Fig. 1 for the number of significantly deregulated genes and miRNAs).

Pathway enrichment. In order to estimate the deregulation of a pathway as a whole, we adopted the net effect of all the genes in a pathway rather than the individual contribution of only a few genes. To this end, deregulated genes from each tumour type were extracted and pathway enrichment was studied. Consequently, we inferred that pathways differed mildly across the distinct tumour types (Fig. 2). However, when the pathways were compared in all the cancer types, we observed that significantly altered pathways known to be related to cancer inhibition (apoptosis, necrosis and senescence, or DDR) frequently appeared to be depleted in differentially expressed genes. Adopting this approach means that genes may exhibit the opposite tendency in terms of expression to that of the pathway as a whole. For instance, although *TP53* is up-regulated in CHOL, de-regulation of the remaining genes in the pathway produce a net silencing of the pathway in this particular cancer. On the other hand, *TP53* is down-regulated in several tumour types (BRCA, ESCA, KIRP, LUAD, LUSC, and PRAD) in agreement with our observations. Indeed, while certain genes may frequently be mutated in cancer, in some cases these mutated forms are not functional. One such example is the aforementioned *TP53*, which is often over-expressed in cancer when it carries mutations, although many of these overexpressed mutated variants have lost their original function²². As such, non-functional overexpressed forms of *TP53* are not capable of triggering apoptosis. We observed a similar trend for other genes and while we found *MDM2* to be up-regulated in most tumours (Supplementary Table 6), the overall senescence pathway was dampened (Fig. 2). The same applies to *RBI*, which is up-regulated in 2 tumour types (CHOL and HNSC) but it is down-regulated in another 4 tumours

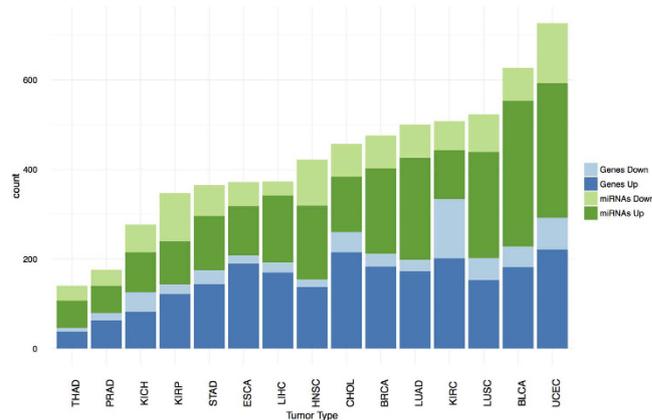


Figure 1. Number of differentially expressed genes and miRNAs (over-expressed and repressed) in each tumour type.

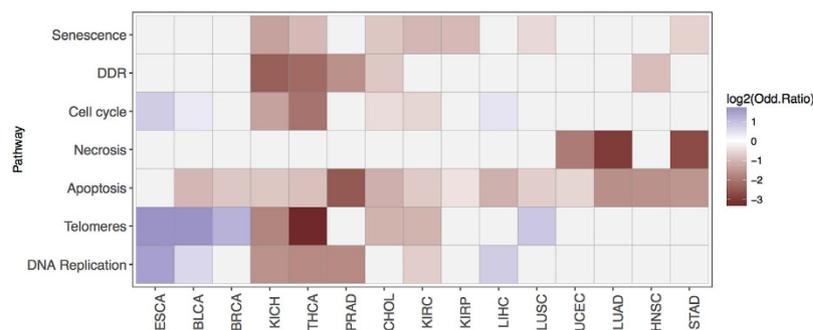


Figure 2. Pathway enrichment analysis of the differentially expressed genes in each tumour type. Statistical significance threshold was set at $P < 0.05$. Enrichment is indicated by \log_2 of the odd ratio. In blue, positively enriched pathways; in red, negatively enriched pathways (more altered than expected and less altered, respectively).

(BLCA, LIHC, LUAD, LUSC), consistent with the effect on the whole pathway. Other pathways exhibited distinct trends depending exclusively on the tumour type, such as DNA replication, telomere elongation and cell cycle (more details are available in Supplementary Table 2).

Genes differentially expressed across all tumour types. Significantly deregulated genes were retrieved from all types of tumours (for details see Supplementary Text), yet in keeping with other procedures^{23,24}, only those genes differentially expressed in at least 8 different tumour types were considered for further analysis. This cut-off was established as it reflected the best combination of tumour types, and the best enrichment of cancer-related genes and miRNA-related genes. To this end we compiled known cancer genes from COSMIC²⁵ and the network of cancer genes (NCG 5.0²⁶), which offers a comprehensive list of known cancer genes (more than 600). Using different thresholds for the number of tumour types (from 5 to 10), the figure of 8 tumours appeared to be the cut-off with the highest odds ratio (enrichment in known cancer genes) with a significant right-tailed Fisher false discovery rate (FDR) of $1.54E-02$ for COSMIC and $1.87E-03$ for NCG (for more details, see Supplementary Tables 3 and 4).

From the total of 147 significantly deregulated genes (see Supplementary Table 6), we identified *SCP24* as the only gene over-expressed in all tumour types. We also found *SYNE1* (relevant to the cell cycle) to be the most strongly downregulated gene in the 15 tumours studied (see Fig. 3 for the distribution of both *SCP24* and *SYNE1*). From the remaining differentially expressed genes, 11 belonged to the apoptosis pathway, in which *E2F1* and *CLSPN* were deregulated in 14 and 12 tumour types, respectively. In other pathways, 126 genes appeared to be misregulated in cell cycle pathways and 21 were altered in the DDR, where the oncogenes *CDT1*, *PLK1* and *DTL* were among the most prominent deregulated genes. In the DNA replication pathway, 26 genes were de-regulated including the previously identified *E2F1*, *CDT1* and the oncogene *GINS2* as the most significantly affected genes. Of the 29 genes for senescence that were deregulated, *Ubc10/UBE2C* was the most strongly deregulated. Finally, the polymerases *DNA2* and *POLE2* appeared to be deregulated in the telomere pathway in a large number of tumours.

miRNAs differentially expressed in tumours. We identified 95 miRNAs that were significantly deregulated in at least 8 of the 15 different tumour types studied (see Supplementary Table 7). As indicated above, this cut-off was selected by performing enrichment at different thresholds using data from the cancer related miRNAs stored in

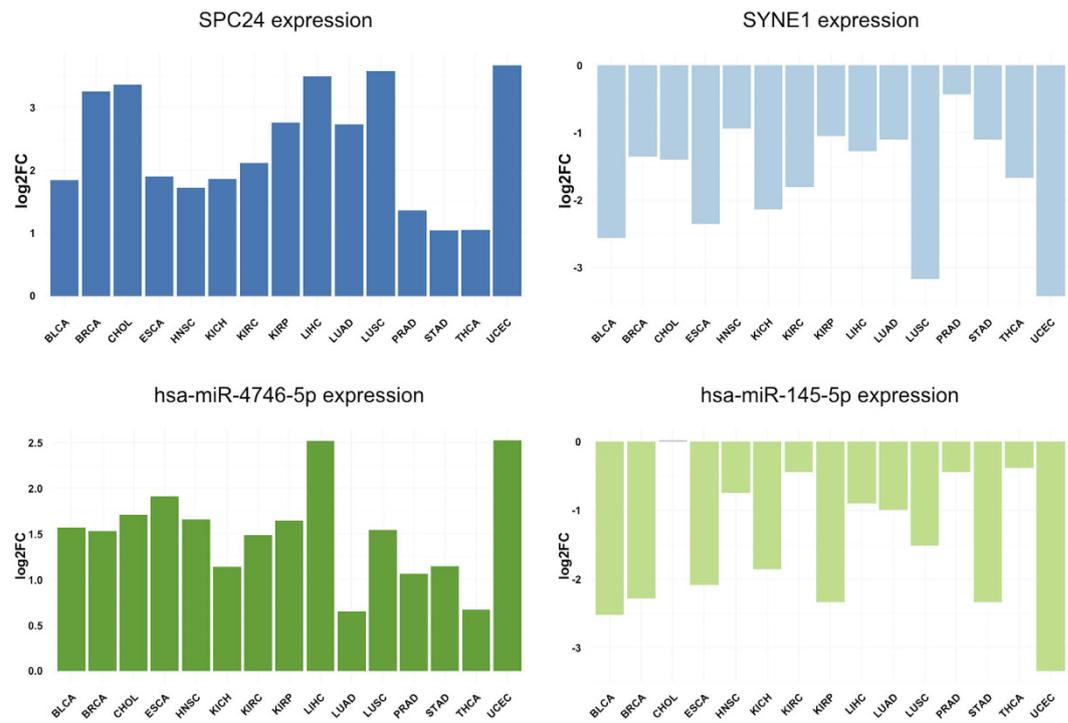


Figure 3. Genes and miRNAs deregulated almost all of the studied tumour types: *SPC24* is differentially overexpressed in all tumour types; *SYNE1* is repressed in all tumour types; miR-4746-5p up-regulated in all the tumour types; and miR-145-5p, down-regulated in all tumour types except Cholangiocarcinoma (CHOL).

OncomirDB²⁷, a database that is a high-confidence reference resource for studying miRNAs related to cancer. Among the up-regulated miRNAs that appear in the vast majority of tumours analysed, we found an oncogenic cluster that was over-expressed in 14 tumours, formed by miR-182, miR-96 and miR183, and miR-4746 that was over-expressed in all the tumours analysed (Fig. 3). Conversely, the anti-oncomiRs miR-145, miR-139 and miR-195 were the most strongly down-regulated miRNAs in our dataset. Interestingly, miR145 and miR143, which were also strongly downregulated (see Fig. 3), belong to the tumour-suppressor miRNA cluster miR143/145 implicated in the initial steps of tumorigenesis¹⁵. Moreover, miR-195 and miR-139 are also well-known tumour suppressor miRNAs²⁸.

Conserved deregulated miRNA-mRNA target interactions. The 147 genes and 95 miRNAs differentially expressed in the majority of tumour types were used to study potential miRNA-mRNA interactions. As miRNAs commonly repress gene expression through the RNA interference pathway, significant negative correlations in expression were expected to reflect the reliability of possible interactions. Accordingly, only inverse differential expression of miRNA-mRNA pairs predicted from miRGate²⁹ were evaluated with a genomic agreement ≥ 2 (for details see methods).

We analysed the correlation between gene expression and miRNA expression of the predicted pairs using linear regression models that included the log-transformed gene expression values as the response, and the copy number alterations (CNAs), gene methylation (MET) and log-transformed miRNA expression values as explanatory variables (Supplementary Table 11). For each of the 41 significant gene-miRNA pairs, individual models were constructed for every cancer type, and a joint analysis was also performed that included all the samples from all the different types of cancer. In the latter, “cancer type” was also included as an explanatory variable (see methods). Thirty-six out of the 41 pairs (Fig. 4) exhibited a statistically significant negative correlation in the joint analysis (Supplementary Figure 1).

From these 36 pairs, we found 25 novel interactions not described in other specialized databases like OncomirDB²⁷, miRCancer³⁰ and miRTarBase³¹. Most of these interactions involved cell cycle related genes, like *SPC24* and *CDC20* (regulated by the anti-oncomiR miR-139), and *PKMYT1* (regulated by the tumour suppressor miR-195). The senescence pathway was also represented and while most of the associations identified were known, novel ones were found like *HMG2A2*/miR-139 and *EZH2*/miR-195. In the DNA replication pathway, the interaction of the *MCM2* gene with the anti-oncomiR miR-139 was notable in the majority of tumours. Finally, the *PRKAR2B*/miR-135b pair was identified that includes the apoptosis related *PRKAR2B* gene.

To infer the pairs in which genes are targeted by few miRNAs (henceforth “high specificity” pairs), we identified 17 pairs with least secondary interactions that were positively enriched in the maximum possible number of tumours (see Methods). Interestingly, from these 17 significant pairs (adjusted p-value ≤ 0.05) several down-regulated tumour-suppressor miRNAs (e.g., miR-let-7c, miR-139, miR-145 or miR-195) appeared to regulate

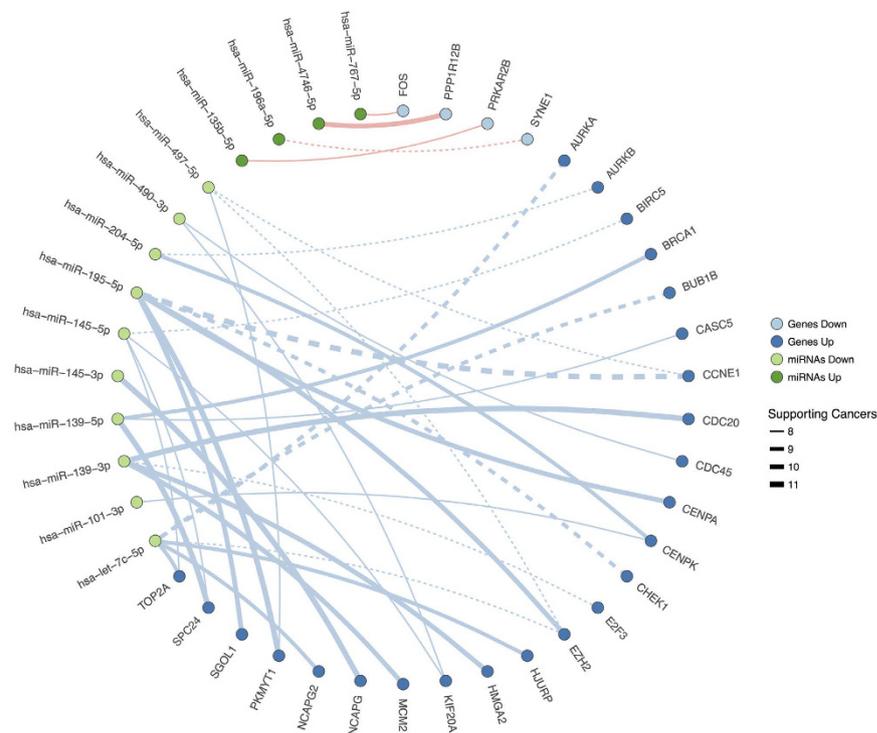


Figure 4. Relevant miRNA-mRNA targets found in the majority of the tumours studied. Lines represent interactions among miRNAs (green) and genes (blue). Continuous line indicates novel interactions, while dashed lines indicate validated interactions.

oncogenic genes related to the cell cycle and DDR pathways (*BUB1B*, *AURKA*, *BIRC5*, *CENPK*, *BRCA1* and *CHEK1*). Eleven of these 17 resulting pairs had not been published previously (Fig. 5 and Supplementary 9). Indeed, these pairs are particularly interesting since the low promiscuity of the miRNA's interactions favours their potential application as anti-miRNA oligonucleotides (AMOs) targets.

We also observed mRNA-miRNA pairs that only appeared in the two different lung tumour types (LUAD and LUSC). Linear regression models were established for these pairs to account for the CNAs and the MET status of the genes, as indicated above (Supplementary Figure 2, and Supplementary Table 12). We found 40 miRNA-mRNA pairs that involved 35 differentially expressed genes (4 being repressed, see Supplementary Figure 3) and 13 statistically deregulated miRNAs. The genes differentially expressed in these “exclusive lung” pairs were also differentially expressed in other tumours. Moreover, in terms of the 13 miRNAs involved in these lung exclusive interactions, 16 of the 40 associations involve miR-miR-1976, which regulates 16 genes involved in cell cycle, apoptosis and the DDR. The second most strongly represented miRNA in these exclusive lung pairs was miR-let-7b-5p, acting as a regulator in 10 interactions where 8 of the genes implicated also belong to cell cycle pathways, and another 2 to senescence pathways (Supplementary Figure 3). Notably, these two down-regulated miRNAs, which are in the most frequently obtained pairs, are apparently exclusive to these two tumour types as they don't appear to be significantly repressed in any other tumour.

Survival Analysis. The 36 pairs that exhibited a statistically significant negative correlation in the multivariate linear regression joint analysis (Supplementary Figure 1) were selected for a subsequent survival analyses (Fig. 6). The effect of gene expression on patient survival was explored separately for each cancer type using Cox proportional hazard models that included tumour stage. In these models, both tumour stage (encoded as one of two categories, i.e. Stage I-II vs Stage III-IV) and the log-transformed gene expression values were used as explanatory variables. Information on tumour stage was not available for prostate cancer (PRAD), so in this case tumour grade was employed as the explanatory variable (Gleason score 6–7 vs 8–10). Pairs that exhibited both a significant correlation between miRNA and gene expression, and a significant tumour stage-independent association between gene expression and patient survival, support the hypothesis that miRNA expression regulates gene expression and that the latter affects patient survival. We found miRNA-Gene Survival associations in 11 of the 36 selected pairs.

In 4 pairs, we found significant associations for at least two distinct tumour types. Then, the *HJURP*-miR-let.7c.5p in KIRP and LUAD, the *CDC20*/miR.139.3p in KICH and LUAD, the *KIF20A*/miR145.5.p pair in HNSC and LUAD, and the *CASC5*/miR.139.5p in BRCA and LUAD. The other 7 pairs were found in only one type of cancer each (Fig. 6 and Supplementary Table 13).

We repeated a similar analysis on the lung-exclusive predicted pairs identified in lung adenocarcinomas (LUAD) and squamous cell lung tumours (LUSC). Of the 74 gene-miRNA pairs originally predicted, 40 exhibited a significant negative correlation in the multivariate linear regression joint analysis that included

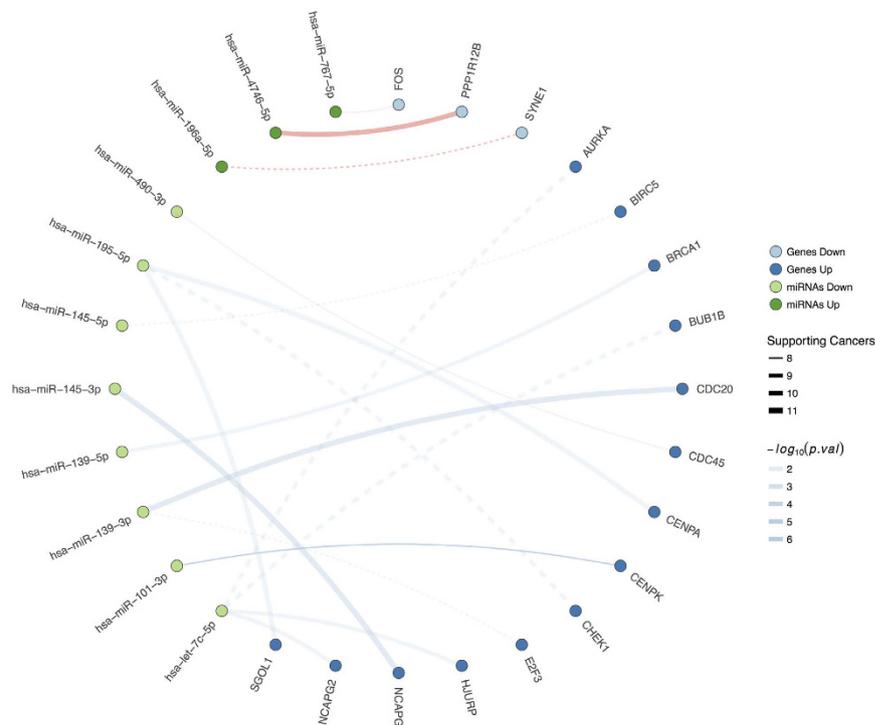


Figure 5. “Highly specific” pairs of interactions. Lines represent interactions among miRNAs (green) and genes (blue). Continuous line indicates novel interactions, while dashed lines indicate validated interactions.

tumour type, CNAs, MET and log-transformed miRNA expression as explanatory variables of gene expression (Supplementary Table 12 and Supplementary Figure 3). Of the 40 lung-exclusive pairs that passed the correlation analysis, we found 14 in which gene expression was also associated with patient survival after accounting for tumour stage (Supplementary Figure 4 and Supplementary Table 14). Notably, 13 were found in LUAD and only one in LUSC.

Discussion

The identification of miRNA-mRNA interaction pairs is gaining significant attention given their potential implications in devising novel therapies³². As such studies pose important experimental challenges, careful computational studies addressing this issue may help to identify novel therapeutic targets, as successfully described elsewhere for individual cancers³³. A widely used method to evaluate the recurrence of miRNA-target associations is a rank-based method (REC, recurrent score³⁴). Like PanMira³⁵, this approach uses processed RNAseq and/or microarray data from tumour samples, excluding samples from healthy subjects. By contrast, alternative analyses using TCGA^{36,37} also include healthy samples as this provides a more accurate framework to understand target/miRNA relationships in cancer. As such, we have included samples from healthy subjects as controls to identify miRNA/mRNA associations from differentially expressed genes and miRNAs, avoiding other spurious interactions in principle not related to cancer.

Although much work has been done in individual tumours, a systematic analysis to infer conserved miRNA-mRNA interactions in different cancer types based on differential expression has yet to be performed on a large number of raw clinical samples. Thus, to determine if different tumours exhibit distinct miRNA-mRNA interactions, we have studied 15 of the most common tumour types in-depth. We analysed 18,605 RNAseq and miRNAseq raw data samples, following recommended procedures³⁸ and applying restrictive filters to reduce false positives.

Our rationale was that miRNAs can act as onco-modulators of relevant target genes influencing their expression in tumours. Thus, by restricting the analyses to relevant cancer pathways, we could identify more suitable candidates. In the pathway analyses we found a general trend towards deregulation, although some genes within the pathways occasionally exhibited the opposite trend, even though the net effect of the remaining genes was to drive an overall “dampening” phenotype.

When considering individual tumour types, our results are consistent with previous findings where individual genes behaved as expected. *SYNE1* is the most significantly repressed gene and its dysregulation has been related to glioblastoma³⁹ and ovarian cancer⁴⁰, among others. From the miRNA perspective, we identified important species previously related to tumour progression. For instance, we found miR-4746 to be over-expressed in all the tumours analysed, in accordance with previous studies⁴¹. Conversely, tumour suppressor miRNAs (miR-145, miR-139 and miR-195) are the most strongly down-regulated miRNAs in our dataset.

To identify suitable relations, as a first filter we only considered de-regulated genes and de-regulated miRNAs that exhibited suitable inverted expression profiles. We next optimized the identification of miRNA binding to precise 3'

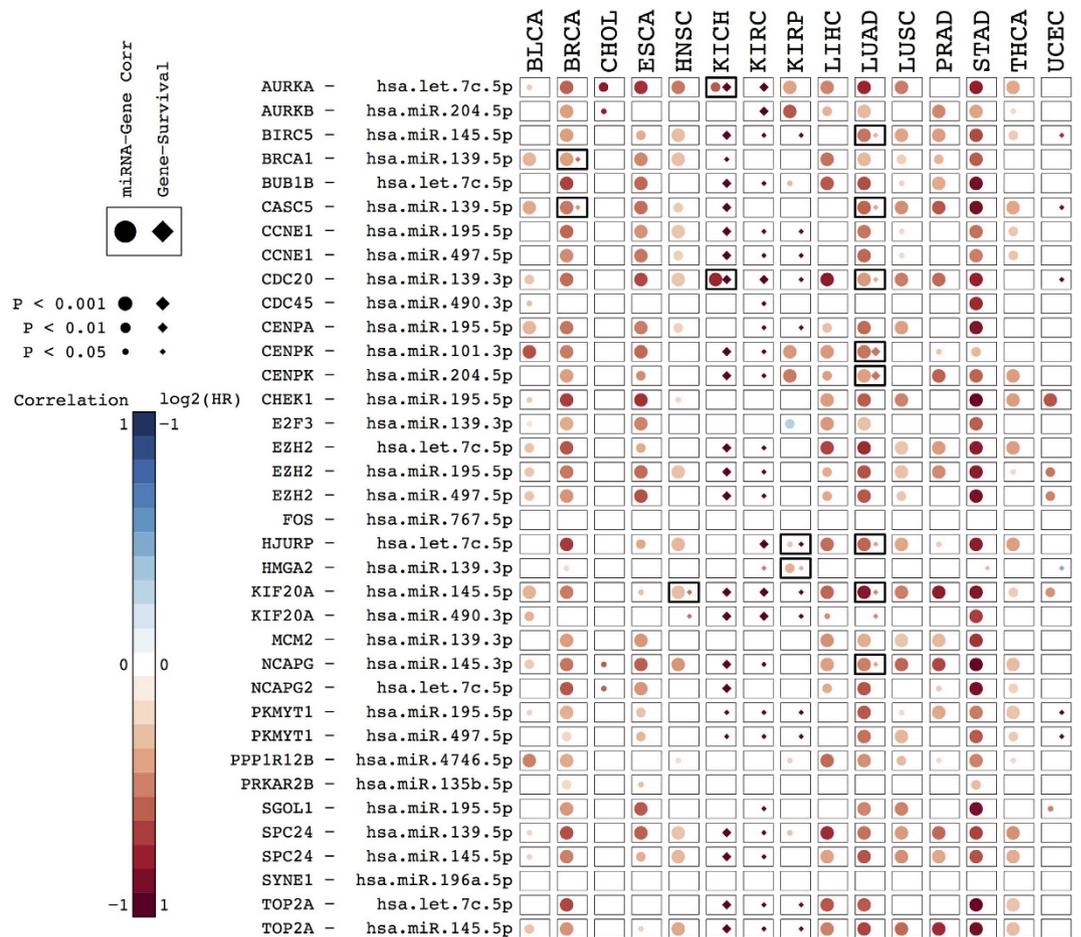


Figure 6. Survival with multifactorial Cox models in 15 tumours. Combined correlation and survival analysis for the selected 36 general miRNA-gene pairs. Every box summarizes the results of the correlation and survival analysis of a miRNA-gene pair (rows) in a particular cancer type (columns). Circles represent miRNA-gene expression correlations, after accounting for CNAs and gene methylation. Negative correlations in red, and positive correlations in blue. Diamonds represent the \log_2 of the hazard ratio (HR) of Cox proportional hazards models that included tumor stage (or tumor grade, in the case of PRAD) and the log-transformed gene expression values as explanatory variables. Positive $\log_2(\text{HR})$, i.e. shorter survival, in shades of red. Negative $\log_2(\text{HR})$, i.e. longer survival, in shades of blue. All P-values were corrected for multi-hypothesis testing by the FDR method. Pairs reaching statistical significance (Adj. $P < 0.05$) in both miRNA-gen expression correlation and gene expression-survival association are indicated with thicker boxes.

UTR sequences that should in principle trigger gene silencing. This effect is well accepted, given that most of the interactions involving miRNAs included the seed regions at 3' UTRs⁴². To support this observation, additional studies in mammals revealed that there is a tendency to diminish target mRNA levels¹¹. However, it is important to note that alternative mechanisms may be at play, including activation due to binding at the promoter or other regions of the gene⁴³. Although non-canonical binding sites^{42,44} have been described for miRNAs, these are limited and they are not included in most of the available prediction software used here to identify target genes. Gene expression is also influenced by other important “pleiotropic” factors. Moreover, indirect interactions involving complex transcriptional networks of other genes and RNA species may also influence the expression of a given gene. While these influences deserve deeper exploration, their complexity and the lack of reliable data to model these elements in cancer makes this an issue beyond the scope of this work. Thus, for the purposes of simplicity we focus on the most parsimonious explanation, which is gene silencing by binding to 3' UTR regions.

It is well known that methylation and copy number alterations affect gene expression, especially in cancer. From the original 41 pairs exhibiting significant inverted expression profiles, we filtered out some pairs by accounting for these factors with 36 consistent interactions remaining as both partners appeared to be deregulated together. Of these 36 pairs, 25 had not previously been inferred, although individually most of the genes and miRNAs involved are known to be related to cancer development. An illustrative example is the *SCP24*/miR-139 couplet, where involvement of the individual components is well-established (*SCP24* and miR-139-5p^{45,46}, while no experimental evidence is available for the pair. A subset of these 36 pairs (17 pairs), appear to be “highly specific” as both partners were not found with many other alternative interactors in the distinct tumour types. These specific associations are noteworthy, as they may be useful for the development of therapies due to the limited

off-target effects. From these pairs, 11 interactions had not previously been described, while among associations that had already been experimentally confirmed were those of *AURKA*⁴⁷ and *BUB1B*⁴⁸ with miR-let-7c.

When comparing miRNA-mRNA interactions among tumour types, we observed that tumours of the same origin exhibited interactions not identified in the remaining tumours (Supplementary Figure 3), in particular lung cancers (LUAD and LUSC), and that these were mostly mediated by two types of miRNAs, miR-1976 and let-7b-5p. These miRNAs act as master regulators of genes belonging to the cell cycle and apoptosis pathways. Moreover, they are involved in the majority of these lung exclusive interactions, suggesting a potential value of these miRNAs as biomarkers for lung cancer. In fact miR-1976 is a potential tumour suppressor in lung cancer, as described in relation with the oncogene *PLCE1*⁴⁹.

There are many factors known to affect cancer prognosis. Cancer survival is ultimately the result of a very complex network of mechanisms that include tumour aggressiveness, operability, routine primary treatment, available secondary therapies, possible drug-resistance, patient age, their physical status, comorbidity, etc. All these factors differ radically among cancer types and affect their prognosis profoundly. For many years, clinicians have employed a number of clinical-pathological characteristics that have proven to be very useful to estimate cancer patient survival. Most notably, tumour stage and degree of differentiation have been used as prognostic factors in the diagnosis and management of most cancer types. For instance, patients with Stage III or IV cancers, where the malignant cells have already metastasized to lymph nodes or distant organs, exhibit much worse prognosis than patients with Stage I-II tumours. Similarly, patients with poorly differentiated tumours exhibit worse prognosis than patients with moderately or well-differentiated tumours. Patient age and gender are also typically employed as predictive markers of survival. In recent years, molecular markers of patient survival or of response to treatment have been steadily incorporated into clinical practice (ASCO, <https://www.asco.org/practice-guidelines> and ESMO, <http://www.esmo.org/Guidelines/Guidelines-News>). Therefore, a single factor (gene, miRNA) should not be expected to produce a large effect on survival by itself, independently of the many confounding factors that potentially affect patient survival, especially across many different tumour types. As such, caution should be exercised regarding the associations between gene regulation by miRNAs and patient prognosis. Nonetheless, we have recovered known associations in our survival analyses, some of which have been studied extensively in many tumours, like the roles on prognosis of genes *AURKA*^{50–53} or *BRCA1*^{54,55}. We also found novel stage-independent prognostic genes not described in the literature like *CASC5*. This gene was previously described in lung cancer⁵⁶, but not in breast cancer, which is the association found in this study. It is worthy to mention that its regulating miRNA (miR-139-5p) has been itself associated to prognosis in lung cancer⁵⁷ but again not in breast. Therefore, our study revealed a novel miRNA-gene-survival association in breast cancer, that is supported by previous findings in lung cancer. Similar findings are the oncogene *BIRC5*, found relevant in LUAD here but only previously described in breast⁵⁸, bladder⁵⁹, chondrosarcoma⁶⁰ and sarcoma⁶¹. However, its regulating miRNA (miR-145-5p) has been associated to prognosis in gastric⁶² and kidney⁶³ tumours. Moreover *CDC20*, described as prognostic in colon⁶⁴ and breast⁶⁵, has been identified as relevant in KICH and LUAD in our study. Finally *NCAPG* and *KIF20A* genes of prognostic value in gliomas^{66,67} were found relevant here in LUAD. We also identified a few novel associations in lung tumours (LUAD), for instance *CENPK* has not previously been associated with this cancer type, even though the gene had already been associated to ovarian cancer⁶⁸. When addressing the “exclusive lung interactions” the genes *CDT1*, *CENPN*, *DAPK2*, *HJURP*, *KIF20A*, *NACPD2*, *NCAPG*, *NCAPH*, *SMC1B*, and *SPC24* have so far not been associated to survival in lung cancer, although most of its associated miRNAs have been. For instance miR-1976 and miR-let-7b-5p miRNAs had already been associated to prognosis in lung cancer^{49,69}, and in other tumours like breast⁷⁰, prostate⁷¹, and gastric⁷². In conclusion, our analysis has revealed novel miRNA-gene correlations that associate with patient prognosis, supported by previous findings on different tumour types.

To summarize, in addition to producing comparable results at the individual level, our integrative analysis provided a confident statistical approach to identify novel miRNA-target relationships in different tumour types within cancer-relevant pathways. The identification of lung exclusive mRNA/miRNA interactions that are significantly associated with survival opens potentially avenues for clinical research into cancers. This exploratory analysis, followed by future experimental validation, could be significant to further decipher the role of miRNAs in tumorigenesis, confirming the validity of our approach to design of miRNA-targeted cancer drugs focusing on the gene/pathway. By confidently identifying specifically altered miRNA-mRNA pairs and interactions, off-target effects could be minimized, reducing side-effects and enhancing already existing treatments.

Methods

TCGA Transcriptome samples. A considerable number of tumour types, and the maximum number of control and tumour samples, are necessary to identify miRNA-mRNA associations related to common processes of tumorigenesis. As such, raw data sample files (fastq or bam files) from the TCGA database were used. These restricted samples were obtained by applying to the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) from the National Institutes of Health (USA) for dbGAP-authorized access. Currently, TCGA includes data on 33 different tumour types obtained from different platforms. Since we are interested in the differential expression of genes and miRNAs implicated in tumour phenotypes, and their interactions, we selected those tumours for which both kinds of transcriptome samples existed. After applying this initial filter, 19 tumour types with over 22,000 paired-end RNASeq and single-end miRNASeq samples were selected. Subsequently, to carry out a comprehensive statistical study by comparing controls with cancer samples, tumours with less than 10 healthy samples and those whose total number of healthy samples did not reach 5% of the total, were discarded. After applying these filters, 15 tumours were selected: Kidney Chromophobe (KICH), Head and Neck squamous cell carcinoma (HNSC), Oesophageal carcinoma (ESCA), Kidney renal papillary cell carcinoma (KIRP), Liver hepatocellular carcinoma (LIHC), Kidney renal clear cell carcinoma (KIRC), Lung Adenocarcinoma (LUAD), Thyroid carcinoma (THAD), Prostate adenocarcinoma (PRAD), Bladder Urothelial

Carcinoma (BLCA), Breast invasive carcinoma (BRCA), Lung squamous cell carcinoma (LUSC), Stomach adenocarcinoma (STAD), Cholangiocarcinoma (CHOL) and Uterine Corpus Endometrial carcinoma (UCEC). A total of 18,605 (6,867 RNASeq and 11,738 small-RNASeq) samples were recorded from these tumours and prepared for downloading from December 21st 2015 at the Cancer Genomics Hub (<https://cghub.ucsc.edu/>). Detailed information about the number of samples for each tumour is shown in Supplementary Table 1.

Selected pathways. In this study, we selected seven pathways that are highlighted as hallmarks of cancer⁴: the DDR and cell cycle pathways were chosen as characteristic of cancer checkpoints and development; telomere elongation and DNA replication are characteristic of tumour progression; and finally, apoptosis, necrosis and senescence were chosen given their role in inhibiting cancer. A total of 1,264 genes (861 unique) belonging to these pathways were identified, mainly from the Reactome⁷³ and KEGG databases⁷⁴. In the case of the DDR pathway, the genes included were obtained from other published studies⁷⁵ and retrieved from the DDRpro database⁷⁶. All gene names and pathways, in which they participate, are shown in Supplementary Table 5.

Hyper-geometric pathway enrichment. In this work, seven cancer-related pathways were defined and the genes implicated were chosen from three different data sources. Since the classification of our pathways differs from those offered by enrichment analysis methods designed for NGS transcriptome sample analysis, we developed a statistical procedure based on an established approach for pathway analysis of cancer genomes⁷⁷, enabling custom pathways to be integrated. Given our hypergeometric distribution, we used a Fisher's exact test in a 2×2 contingency table to generate an odds ratio (OD) that quantifies the strength of enrichment, as well as a p-value corrected for multiple testing that estimates the proportion of enriched gene sets that would occur by chance given the number of gene sets tested based on the Benjamini-Hochberg method. While an OD > 1 for a given pathway means that the number of deregulated genes (either up-regulated or down-regulated) are positively enriched relative to all the genes expressed in the experiment (deregulated genes appear more than expected), an OD < 1 means that the given pathway is depleted in altered genes and consequently, it behaves like a healthy sample. As a control, differentially expressed genes equal in size to the number of genes within our described pathway were randomly selected to compare the Ors and adjusted p-values as shown in Supplementary Table 2. Ods for each of the tumour types are shown for any individual pathway in Fig. 2.

Transcriptome analysis of miRNAs and mRNAs. A total of 11,738 miRNA and 6,787 mRNA experiments were downloaded from the controlled-access Cancer Genomics Hub. These samples were analysed following the protocol proposed previously³⁸ and using a pipeline for NGS studies called miARma-Seq⁷⁸. Briefly, this tool brings together well-established software in a single bundle, allowing a complete analysis of the raw data. Quality analysis, adapter removal, read alignment, read count summarizing, differential expression analysis and gene pathway enrichment were performed, among others. For miRNA analysis, we employed miARma-Seq from the miRBase version 20 annotation database⁷⁹, obtaining aligned reads for each mature miRNA. Similarly, the latest GRCh37 EnsEMBL version⁸⁰ was used to acquire the total number of paired aligned reads for each human gene.

Statistical analysis of the transcriptome samples. To measure the miRNAs or genes differentially expressed in an independent cancer type, we compared the number of aligned reads associated with each miRNA/gene in the control and tumour samples for each individual tumour. To achieve this, edgeR⁸¹ was used from the differential expression module included in miARma-Seq⁷⁸. As recommended³⁸, weakly expressed elements (miRNAs or genes) were filtered out if the CPM (counts per million reads) was < 1. Reads were normalized using the TMM method (trimmed mean of M-values) and the exact tests to calculate differences between two groups of negative-binomial counts was performed to gather the log₂ fold change, and the Benjamini and Hochberg false discovery rate (FDR) for each miRNA/gene. The miRNAs or genes with a FDR < 0.05 were considered significantly and reliably differentially expressed (DE). From these, those with an absolute log₂ fold change (log₂FC) value > 1 were selected for subsequent analyses (Supplementary Tables 6 and 7).

The miRNA-mRNA target prediction. Differentially expressed genes belonging to our seven defined pathways and miRNAs (FDR < 0.05) for each tumour analysed with an absolute log₂FC ≥ 1 were chosen for target prediction. As miRNAs disrupt and reduce the expression of their target mRNAs, we expect miRNAs and mRNAs to exhibit an inverse expression profile. Thus, down-regulated mRNAs for a given tumour were used in the target prediction step for up-regulated miRNAs from the same tumour samples, and vice-versa. For miRNA-mRNA target predictions, we used miRGate²⁹, a database of novel predicted miRNA-mRNA pairs computed from a common source of sequences. This repository includes a feature called "genomic agreement" that represents the number of unique predictions obtained from five different methods that occur in the same genomic position. Therefore, they are predicted from different prediction algorithms. To obtain reliable miRNA-mRNA pairs, only targets with an agreement value ≥ 2 were chosen.

Integrative analysis of all tumour types. Once all samples for each of the 15 tumour types were analysed, these results were studied in a global framework. As most of the oncogenes and tumour suppressor genes, and their oncomiR and anti-oncomiR regulators, appear to be misexpressed in tumours⁸², statistically deregulated genes and miRNAs (FDR ≤ 0.05 and log₂FC ≥ 1) were extracted for each tumour type. Moreover, to decrease the number of false positives obtained from prediction algorithms, the miRNA target calculation step was improved by using inverse correlated profiles as recommended elsewhere²⁰. Furthermore, genes and miRNAs were selected when both appeared to be differentially expressed in at least eight common tumour types, and only targets with a genomic agreement ≥ 2 were considered (Supplementary Table 8).

In addition, in order to obtain highly specific miRNA-target associations conserved in the maximum number of tumour types as a representation of a shared deregulated mechanisms in cancer, we applied the following statistical approach: for each partner in a miRNA-mRNA interaction, we computed all the possible deregulated genes and miRNAs, respectively, to obtain the pair that is statistically enriched in the maximum number of tumours with fewer co-interactors. So for each miRNA, we obtained the gene(s) that could be potentially regulated in the maximum number of tumours and conversely, for each gene we obtained regulator miRNA(s) in the maximum number of tumours. This is implemented by computing a 2×2 contingency right-tailed Fisher exact test, measuring all existing miRNA-target associations obtained through the differentially expressed genes and miRNAs in the majority of tumours included with a genomic agreement ≥ 2 . P-values from fisher exact test were corrected from multiple testing using the Benjamini-Hochberg method⁸³. For details of these pairs see Supplementary Table 9.

Identification of pairs belonging to the same tumour type. Differentially expressed genes and miRNAs for each individual tumour type were collected (Fig. 1), and like all the aforementioned analysis, we only considered those genes and miRNAs having a $FDR \leq 0.05$ and a significant variation in expression (absolute $\log_2FC \geq 1$). Once we had the set of statistically deregulated genes and miRNAs in each tumour type, we computed all the possible miRNA-mRNA interactions using miRGate²⁹ and selecting only those associations predicted by more than one algorithm in the same genomic position in the 3'UTR (genomic agreement ≥ 2). Finally, once we have inferred this set of confident interactions, we first compare them among the 15 tumour types, to obtain a list of exclusive miRNA-mRNA interactions in each tumour type.

Analyses of methylation and copy number alterations in the pairs identified. The association between gene and miRNA expression was analysed by multifactorial linear regression for every selected gene-miRNA pair, with the gene log-transformed RPKM values as the response and the miRNA log-transformed RPKM as the predictor (Supplementary Tables 11 and 12). To correct for putative confounding factors affecting gene expression, copy number alterations (CNAs) and whole gene methylation (MET) were included in the models. For CNAs, Gistic2-preanalyzed data⁸⁴ was downloaded directly from cBioportal^{85,86} and the information was coded as a categorical ordered factor with five different levels: homozygous deletion (-2), hemizygous deletion (-1), no change (0), gain (1) and high level amplification (2). DNA methylation data was obtained from *Wanderer*, a web-based tool that offers convenient and fast access to gene-centred TCGA methylation data⁸⁷. Methylation of cholangiomas (CHOL) was not available from *Wanderer* at the time. Thus, CHOL methylation data was directly downloaded from the cBioportal. For simplicity, all DNA methylation probes mapping within a particular gene locus were included in its regression models as independent factors. Methylation data was transformed from β -values to M-values to reduce heterocedasticity before being included in the analyses⁸⁸. All methylation probes used in the analysis are shown in Supplementary Table 10. For each mRNA-miRNA pair, we performed individual regression analyses for every individual cancer type in the study, plus an additional joint analysis that included all cancer types. To account for differences in gene expression and miRNA expression among cancer types, cancer type, and the statistical interaction between cancer type and miRNA expression were included as factors in the joint analyses. In all analyses, the gene-miRNA correlation was calculated as the squared root of the fraction of variance in gene expression explained by miRNA expression obtained by ANOVA with Type-III sum of squares, after accounting for the other factors, i.e.: cancer type and its interaction with miRNA expression (where applicable), CNAs and MET. The direction of the correlation, either positive or negative, as well as the t -statistic and the resulting P-value were calculated from the coefficients of the linear models. P-values were adjusted by the FDR method, taking into consideration that every mRNA-miRNA association was analysed in 16 different datasets⁸³. Regression could not be calculated in some cases where either the gene or the miRNA were not detected in the RNASeq data of a particular tumour type. In those cases, the correlation between gene and miRNA in that tumour type was considered to be non-existent. Details are available in Supplementary Tables 8 and 1, Supplementary Figures 1 and 2. Only pairs exhibiting negative correlation between gene and miRNA expression, and with a P value < 0.05 after multi-hypothesis correction in the joint analysis, were selected for subsequent analyses (Supplementary Tables 11 and 12).

Survival analysis. For each gene-miRNA pair, the effect of gene and miRNA expression on patient survival was analysed by Cox's proportional hazards modelling separately in every cancer type. All the models included tumour stage as a co-factor, coded into two categories: good prognosis (stages I or II) and bad prognosis (stages III, IV and X). In prostate cancer, the classification was performed based on tumour grade instead of tumour stage, where good prognosis involved tumours with Gleason score of 6 or 7, and bad prognosis tumours with Gleason score ≥ 8 . In these analyses, gene or miRNA expression values were introduced as quantitative log-transformed variables. Multivariate Cox proportional hazards analysis provides an estimate of the effect of the gene/miRNA expression on survival, independently of the effect of tumour stage (or tumour grade, in the case of prostate cancer). P-values were adjusted by the FDR method⁸³ and survival analyses are shown in Fig. 6 and Supplementary Figure 4. For details see Supplementary Tables 13 and 14.

The code to analyse the data can be found at <https://github.com/amrojasmendoza/mRNA-miRNAs-cancer-pathways>.

References

1. Miller, E. C. & Miller, J. A. Mechanisms of chemical carcinogenesis. *Cancer* **47**, 1055–1064 (1981).
2. Roos, W. P. & Kaina, B. DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Lett* **332**, 237–248, doi: 10.1016/j.canlet.2012.01.007 (2013).
3. Xu, Y. & Baltimore, D. Dual roles of ATM in the cellular response to radiation and in cell growth control. *Genes Dev* **10**, 2401–2410 (1996).

4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674, doi: 10.1016/j.cell.2011.02.013 (2011).
5. Arora, H., Qureshi, R., Rizvi, M. A., Shrivastava, S. & Parihar, M. S. Study of apoptosis-related interactions in colorectal cancer. *Tumour Biol*, doi: 10.1007/s13277-016-5363-9 (2016).
6. Wen, M., Kwon, Y., Wang, Y., Mao, J. H. & Wei, G. Elevated expression of UBE2T exhibits oncogenic properties in human prostate cancer. *Oncotarget* **6**, 25226–25239, doi: 10.18632/oncotarget.4712 (2015).
7. Levine, A. J. The p53 tumor suppressor gene and gene product. *Princess Takamatsu Symp* **20**, 221–230 (1989).
8. Preston, G. A. *et al.* Induction of apoptosis by c-Fos protein. *Mol Cell Biol* **16**, 211–218 (1996).
9. Bischoff, J. R. *et al.* A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* **17**, 3052–3065, doi: 10.1093/emboj/17.11.3052 (1998).
10. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
11. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840, doi: 10.1038/nature09267 (2010).
12. Wang, Y., Luo, J., Zhang, H. & Lu, J. microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. *Mol Biol Evol*, doi: 10.1093/molbev/msw089 (2016).
13. Hammond, S. M. MicroRNAs as oncogenes. *Curr Opin Genet Dev* **16**, 4–9, doi: 10.1016/j.gde.2005.12.005 (2006).
14. Hossain, A., Kuo, M. T. & Saunders, G. F. Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. *Mol Cell Biol* **26**, 8191–8201, doi: 10.1128/MCB.00242-06 (2006).
15. Akao, Y., Nakagawa, Y. & Naoe, T. MicroRNAs 143 and 145 are possible common onco-microRNAs in human cancers. *Oncol Rep* **16**, 845–850 (2006).
16. Cordes, K. R. *et al.* miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* **460**, 705–710, doi: 10.1038/nature08195 (2009).
17. Huttenhofer, A. & Vogel, J. Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res* **34**, 635–646, doi: 10.1093/nar/gkj469 (2006).
18. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63, doi: 10.1038/nature07228 (2008).
19. Min, H. & Yoon, S. Got target? Computational methods for microRNA target prediction and their extension. *Exp Mol Med* **42**, 233–244, doi: 10.3858/emmm.2010.42.4.032 (2010).
20. Vishnubalaji, R. *et al.* Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. *Cell Death Dis* **6**, e1614, doi: 10.1038/cddis.2014.556 (2015).
21. Leiserson, M. D. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet* **47**, 106–114, doi: 10.1038/ng.3168 (2015).
22. Hinds, P. W. *et al.* Mutant p53 DNA clones from human colon carcinomas cooperate with ras in transforming primary rat cells: a comparison of the “hot spot” mutant phenotypes. *Cell Growth Differ* **1**, 571–580 (1990).
23. Yang, Z., Jones, A., Widschwendter, M. & Teschendorff, A. E. An integrative pan-cancer-wide analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer. *Genome Biol* **16**, 140, doi: 10.1186/s13059-015-0699-9 (2015).
24. Sanchez-Vega, F., Gotea, V., Margolin, G. & Elnitski, L. Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. *Epigenetics Chromatin* **8**, 14, doi: 10.1186/s13072-015-0007-7 (2015).
25. Forbes, S. A. *et al.* COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805–811, doi: 10.1093/nar/gku1075 (2015).
26. An, O., Dall’Olio, G. M., Mourikis, T. P. & Ciccarelli, F. D. NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic Acids Res* **44**, D992–999, doi: 10.1093/nar/gkv1123 (2016).
27. Wang, D., Gu, J., Wang, T. & Ding, Z. OncoMiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics* **30**, 2237–2238, doi: 10.1093/bioinformatics/btu155 (2014).
28. Yonemori, M. *et al.* Dual tumor-suppressors miR-139-5p and miR-139-3p targeting matrix metalloproteinase 11 (MMP11) in bladder cancer. *Cancer Sci*, doi: 10.1111/cas.13002 (2016).
29. Andres-Leon, E., Gonzalez Pena, D., Gomez-Lopez, G. & Pisano, D. G. miRGate: a curated database of human, mouse and rat miRNA-mRNA targets. *Database (Oxford)* **2015**, bav035, doi: 10.1093/database/bav035 (2015).
30. Xie, B., Ding, Q., Han, H. & Wu, D. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics* **29**, 638–644, doi: 10.1093/bioinformatics/btt014 (2013).
31. Chou, C. H. *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res* **44**, D239–247, doi: 10.1093/nar/gkv1258 (2016).
32. Kim, B. G. *et al.* Transcriptome-wide analysis of compression-induced microRNA expression alteration in breast cancer for mining therapeutic targets. *Oncotarget* **7**, 27468–27478, doi: 10.18632/oncotarget.8322 (2016).
33. Matamala, N. *et al.* MicroRNA deregulation in triple negative breast cancer reveals a role of miR-498 in regulating BRCA1 expression. *Oncotarget* **7**, 20068–20079, doi: 10.18632/oncotarget.7705 (2016).
34. Jacobsen, A. *et al.* Analysis of microRNA-target interactions across diverse cancer types. *Nat Struct Mol Biol* **20**, 1325–1332, doi: 10.1038/nsmb.2678 (2013).
35. Li, Y. & Zhang, Z. Potential microRNA-mediated oncogenic intercellular communication revealed by pan-cancer analysis. *Sci Rep* **4**, 7097, doi: 10.1038/srep07097 (2014).
36. Cancer Genome Atlas Research, N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49, doi: 10.1038/nature12222 (2013).
37. Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582, doi: 10.1038/nature14129 (2015).
38. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols* **8**, 1765–1786, doi: 10.1038/nprot.2013.099 (2013).
39. Serao, N. V., Delfino, K. R., Southey, B. R., Beever, J. E. & Rodriguez-Zas, S. L. Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. *BMC Med Genomics* **4**, 49, doi: 10.1186/1755-8794-4-49 (2011).
40. Doherty, J. A. *et al.* ESR1/SYNE1 polymorphism and invasive epithelial ovarian cancer risk: an Ovarian Cancer Association Consortium study. *Cancer Epidemiol Biomarkers Prev* **19**, 245–250, doi: 10.1158/1055-9965.EPI-09-0729 (2010).
41. Zhou, R. *et al.* Tumor invasion and metastasis regulated by microRNA-184 and microRNA-574-5p in small-cell lung cancer. *Oncotarget* **6**, 44609–44622, doi: 10.18632/oncotarget.6338 (2015).
42. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665, doi: 10.1016/j.cell.2013.03.043 (2013).
43. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci USA* **105**, 1608–1613, doi: 10.1073/pnas.0707594105 (2008).
44. Lytle, J. R., Yario, T. A. & Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc Natl Acad Sci USA* **104**, 9667–9672, doi: 10.1073/pnas.0703820104 (2007).
45. Thiru, P. *et al.* Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. *Mol Biol Cell* **25**, 1983–1994, doi: 10.1091/mbc.E14-03-0837 (2014).

46. Guo, J. *et al.* Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. *J Gastroenterol Hepatol* **24**, 652–657, doi: 10.1111/j.1440-1746.2008.05666.x (2009).
47. Weier, H. U. & Mao, J. H. Meta-analysis of Aurora Kinase A (AURKA) Expression Data Reveals a Significant Correlation Between Increased AURKA Expression and Distant Metastases in Human ER-positive Breast Cancers. *J Data Mining Genomics Proteomics* **4**, 127, doi: 10.4172/2153-0602.1000127 (2013).
48. Xiong, H. *et al.* Integrated microRNA and mRNA transcriptome sequencing reveals the potential roles of miRNAs in stage I endometrioid endometrial carcinoma. *PLoS One* **9**, e110163, doi: 10.1371/journal.pone.0110163 (2014).
49. Chen, G., Hu, J., Huang, Z., Yang, L. & Chen, M. MicroRNA-1976 functions as a tumor suppressor and serves as a prognostic indicator in non-small cell lung cancer by directly targeting PLCE1. *Biochem Biophys Res Commun* **473**, 1144–1151, doi: 10.1016/j.bbrc.2016.04.030 (2016).
50. Zhang, J., Li, B., Yang, Q., Zhang, P. & Wang, H. Prognostic value of Aurora kinase A (AURKA) expression among solid tumor patients: a systematic review and meta-analysis. *Jpn J Clin Oncol* **45**, 629–636, doi: 10.1093/jjco/hyv058 (2015).
51. Goos, J. A. *et al.* Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis. *Br J Cancer* **109**, 2445–2452, doi: 10.1038/bjc.2013.608 (2013).
52. Siggelkow, W. *et al.* Expression of aurora kinase A is associated with metastasis-free survival in node-negative breast cancer patients. *BMC Cancer* **12**, 562, doi: 10.1186/1471-2407-12-562 (2012).
53. Xu, H. T. *et al.* Expression of serine threonine kinase 15 is associated with poor differentiation in lung squamous cell carcinoma and adenocarcinoma. *Pathol Int* **56**, 375–380, doi: 10.1111/j.1440-1827.2006.01974.x (2006).
54. Lambie, H. *et al.* Prognostic significance of BRCA1 expression in sporadic breast carcinomas. *J Pathol* **200**, 207–213, doi: 10.1002/path.1348 (2003).
55. Lesnock, J. L. *et al.* BRCA1 expression and improved survival in ovarian cancer patients treated with intraperitoneal cisplatin and paclitaxel: a Gynecologic Oncology Group Study. *Br J Cancer* **108**, 1231–1237, doi: 10.1038/bjc.2013.70 (2013).
56. Takimoto, M. *et al.* Frequent expression of new cancer/testis gene D40/AF15q14 in lung cancers of smokers. *Br J Cancer* **86**, 1757–1762, doi: 10.1038/sj.bjc.6600328 (2002).
57. Sun, C. *et al.* Hsa-miR-139-5p inhibits proliferation and causes apoptosis associated with down-regulation of c-Met. *Oncotarget* **6**, 39756–39792, doi: 10.18632/oncotarget.5476 (2015).
58. Lv, Y. G., Yu, F., Yao, Q., Chen, J. H. & Wang, L. The role of survivin in diagnosis, prognosis and treatment of breast cancer. *J Thorac Dis* **2**, 100–110 (2010).
59. Akhtar, M., Gallagher, L. & Rohan, S. Survivin: role in diagnosis, prognosis, and treatment of bladder cancer. *Adv Anat Pathol* **13**, 122–126 (2006).
60. de Jong, Y. *et al.* Targeting survivin as a potential new treatment for chondrosarcoma of bone. *Oncogenesis* **5**, e222, doi: 10.1038/oncsis.2016.33 (2016).
61. Taubert, H. *et al.* Elevated expression of survivin-splice variants predicts a poor outcome for soft-tissue sarcomas patients. *Oncogene* **24**, 5258–5261, doi: 10.1038/sj.onc.1208702 (2005).
62. Zhang, Y. *et al.* Downregulation of miR-145-5p correlates with poor prognosis in gastric cancer. *Eur Rev Med Pharmacol Sci* **20**, 3026–3030 (2016).
63. Slaby, O. *et al.* Identification of MicroRNAs associated with early relapse after nephrectomy in renal cell carcinoma patients. *Genes Chromosomes Cancer* **51**, 707–716, doi: 10.1002/gcc.21957 (2012).
64. Wu, W. J. *et al.* CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer. *J Transl Med* **11**, 142, doi: 10.1186/1479-5876-11-142 (2013).
65. Karra, H. *et al.* Cdc20 and securin overexpression predict short-term breast cancer survival. *Br J Cancer* **110**, 2905–2913, doi: 10.1038/bjc.2014.252 (2014).
66. Liang, M. L. *et al.* Downregulation of miR-137 and miR-6500-3p promotes cell proliferation in pediatric high-grade gliomas. *Oncotarget* **7**, 19723–19737, doi: 10.18632/oncotarget.7736 (2016).
67. Duan, J., Huang, W. & Shi, H. Positive expression of KIF20A indicates poor prognosis of glioma patients. *Onco Targets Ther* **9**, 6741–6749, doi: 10.2147/OTT.S115974 (2016).
68. Lee, Y. C., Huang, C. C., Lin, D. Y., Chang, W. C. & Lee, K. H. Overexpression of centromere protein K (CENPK) in ovarian cancer is correlated with poor patient survival and associated with predictive and prognostic relevance. *PeerJ* **3**, e1386, doi: 10.7717/peerj.1386 (2015).
69. Jusufovic, E. *et al.* let-7b and miR-126 are down-regulated in tumor tissue and correlate with microvessel density and survival outcomes in non-small-cell lung cancer. *PLoS One* **7**, e45577, doi: 10.1371/journal.pone.0045577 (2012).
70. Ma, L., Li, G. Z., Wu, Z. S. & Meng, G. Prognostic significance of let-7b expression in breast cancer and correlation to its target gene of BSG expression. *Med Oncol* **31**, 773, doi: 10.1007/s12032-013-0773-7 (2014).
71. Wang, Z. *et al.* miRNA let-7b modulates macrophage polarization and enhances tumor-associated macrophages to promote angiogenesis and mobility in prostate cancer. *Sci Rep* **6**, 25602, doi: 10.1038/srep25602 (2016).
72. Kang, W. *et al.* let-7b/g silencing activates AKT signaling to promote gastric carcinogenesis. *J Transl Med* **12**, 281, doi: 10.1186/s12967-014-0281-3 (2014).
73. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* **44**, D481–487, doi: 10.1093/nar/gkv1351 (2016).
74. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457–462, doi: 10.1093/nar/gkv1070 (2016).
75. Arcas, A., Fernandez-Capetillo, O., Cases, I. & Rojas, A. M. Emergence and evolutionary analysis of the human DDR network: implications in comparative genomics and downstream analyses. *Mol Biol Evol* **31**, 940–961, doi: 10.1093/molbev/msu046 (2014).
76. Andres-Leon, E., Cases, I., Arcas, A. & Rojas, A. M. DDRpro: a database of DNA damage response-related proteins. *Database (Oxford)* **2016**, doi: 10.1093/database/baw123 (2016).
77. Creixell, P. *et al.* Pathway and network analysis of cancer genomes. *Nat Methods* **12**, 615–621, doi: 10.1038/nmeth.3440 (2015).
78. Andres-Leon, E., Nunez-Torres, R. & Rojas, A. M. miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Sci Rep* **6**, 25749, doi: 10.1038/srep25749 (2016).
79. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68–73, doi: 10.1093/nar/gkt1181 (2014).
80. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res* **44**, D710–716, doi: 10.1093/nar/gkv1157 (2016).
81. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, doi: 10.1093/bioinformatics/btp616 (2010).
82. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251, doi: 10.1016/j.cell.2013.02.014 (2013).
83. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 12 (1995).
84. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi: 10.1186/gb-2011-12-4-r41 (2011).
85. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi: 10.1126/scisignal.2004088 (2013).

86. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404, doi: 10.1158/2159-8290.CD-12-0095 (2012).
87. Diez-Villanueva, A., Mallona, I. & Peinado, M. A. Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. *Epigenetics Chromatin* **8**, 22, doi: 10.1186/s13072-015-0014-8 (2015).
88. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587, doi: 10.1186/1471-2105-11-587 (2010).

Acknowledgements

The results published here are based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. Consequently, we thank all the donors that provided samples to that consortium. We are very grateful to Dr Izaskun Mallona at the PMPPC-IGTP for facilitating the DNA methylation information directly from the Wanderer database. E.A.-L. was partially funded by the European Commission grant [FP7-REGPOT-2012-2013-1].

Author Contributions

E.A.-L. contributed to the data acquisition and data analysis, interpreting and summarizing the results, and the writing of the manuscript. S.A. contributed to the data acquisition and data analysis, as well as interpreting and summarising the results, and he participated in the manuscript revisions. I.C. contributed to interpreting and summarising the results and manuscript revisions. A.M.R. contributed to the experimental design, the interpretation and summarising of the results, and to writing the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Andrés-León, E. *et al.* Novel miRNA-mRNA interactions conserved in essential cancer pathways. *Sci. Rep.* **7**, 46101; doi: 10.1038/srep46101 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017