RESEARCH ARTICLE

miTALOS v2: Analyzing Tissue Specific microRNA Function

Martin Preusse[1,2], Fabian J. Theis[1,3], Nikola S. Mueller[1] *

1 Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, **2** Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, **3** Institute for Mathematical Sciences, Technische Universität München, Munich, Germany

* nikola.mueller@helmholtz-muenchen.de

Abstract

MicroRNAs are involved in almost all biological processes and have emerged as regulators of signaling pathways. We show that miRNA target genes and pathway genes are not uniformly expressed across human tissues. To capture tissue specific effects, we developed a novel methodology for tissue specific pathway analysis of miRNAs. We incorporated the most recent and highest quality miRNA targeting data (TargetScan and StarBase), RNA-seq based gene expression data (EBI Expression Atlas) and multiple new pathway data sources to increase the biological relevance of the predicted miRNA-pathway associations. We identified new potential roles of miR-199a-3p, miR-199b-3p and the miR-200 family in hepatocellular carcinoma, involving the regulation of metastasis through MAPK and Wnt signaling. Also, an association of miR-571 and Notch signaling in liver fibrosis was proposed. To facilitate data update and future extensions of our tool, we developed a flexible database backend using the graph database neo4j. The new backend as well as the novel methodology were included in the updated miTALOS v2, a tool that provides insights into tissue specific miRNA regulation of biological pathways. miTALOS v2 is available at http://mips.helmholtz-muenchen.de/mitalos.

Introduction

MicroRNAs (miRNAs) are short, non-coding RNAs that regulate gene expression post transcriptionally through binding to a target mRNA. They are predicted to target hundreds of genes in mammals and most genes are thought to be regulated by miRNAs [1]. Consequently, most biological processes involve miRNAs and miRNA-mediated control of gene expression.

Functional analysis of miRNAs depends on accurate identification of gene targets in a given biological context [2]. Since there is no comprehensive catalogue of tissue and cell type specific miRNA-mRNA interactions, computational target prediction tools are still widely used. Although these prediction tools have improved in accuracy, they still suffer from large numbers of false-positive miRNA-mRNA interactions [2]. Recently, biochemical methods using sequencing of target RNA isolated after UV crosslinking and immunoprecipitation of Ago/miRNA

complexes (CLIP-seq) were developed [3,4]. They produce a map of miRNA binding sites on their target mRNAs. CLIP-seq data is collected in the StarBase database [5], providing a constantly growing resource of experimentally supported interactions. While these experimental methods increase the specificity of miRNA target data, their explanatory power is limited due to differences in experimental procedures and lack of reproducibility [6]. Moreover, all human data sets in StarBase were measured in immortalized cell lines (HEK293, HeLa) and not in primary tissue.

Next to limitations of *in-silico* and experimental gene target identification, miRNA-mediated regulation suggested by *in-vitro* and cell culture experiments is often not supported by *in-vivo* validation studies [7]. This can be partly explained by the fact that most miRNAs show only limited effects on the level of individual target mRNAs under physiological conditions [8]. In addition, target prediction and CLIP-seq studies demonstrated that most mRNAs are regulated by multiple miRNAs [9–11]. Thus, the down-regulation of a target gene depends on the combined effect of multiple miRNAs. And analysis of individual miRNA-mRNA interactions is not sufficient to explain the regulatory role of miRNAs in biological process.

Computational approaches often perform a pathway analysis to increase the explanatory power of target gene sets and to circumvent the shortcomings in targeting data. They use the complete set of miRNA target genes and pathway genes to associate miRNAs to biological pathways as an indication of their biological function. In doing so they do not account for the characteristic tissue expression signature of mammalian genes [12] and thus disregard tissue specific effects of miRNAs. Indeed, miRNAs were shown to facilitate tissue specificity of gene regulation [13]. Moreover, other pathway analysis tools such as DIANA mirPath rely on target prediction only and do not use CLIP-seq based target data [14].

Tissue-specific gene expression data can be obtained using next-generation sequencing of RNA (RNA-seq). The EBI Expression Atlas [15] collects highly curated gene expression data sets and also includes baseline expression data for healthy tissue or untreated cell lines in various organisms. Baseline expression describes the abundance of a gene and is extracted from large-scale expression studies such as ENCODE cell lines.

We developed a novel pathway analysis methodology leveraging this high-quality tissue expression data in order to predict miRNA function. We used our new methodology to first analyze the role of miRNAs in hepatocellular carcinoma and identified the liver-specific effect of miR-199a/b-3p on pathways associated with proliferation and cell migration, a novel function that a recent study proposed. We next dissected the individual functions of the two genomic clusters of the miR-200 family and found hints to new signaling relationships, which were studied in other tissues and cell culture but not yet in liver cancer. We finally extended our analysis to liver fibrosis, which is in general less well studied than liver cancer. miR-571 is known to play a role here and we identified Notch signaling as a putative function. Interestingly, Notch signaling has already been proposed as a drug target for fibrosis in other tissues. With the three case studies we demonstrated the necessity to use tissue-specific target gene information for miRNA function prediction.

To make our novel pathway analysis methodology publicly available, we systematically integrated 1) high-quality miRNA targeting data from TargetScan and CLIP-seq studies from StarBase v2 [5], 2) tissue specific gene expression from the latest version of EBI Expression Atlas [15] with 3) three major pathway databases KEGG [16], WikiPathways [17] and Reactome [18]. A graph database was used to store the data in a flexible manner and increase the query performance compared to relational data stores. The data backend and the corresponding pathway analysis methodology were integrated into miTALOS version 2 (v2), a user-friendly web application to identify pathways regulated by miRNAs in a tissue specific manner. With miTALOS v2 users can analyze multiple miRNAs together to account for combinatorial effects.

MiTALOS v2 is complementary to other functional miRNA analysis tools such as miRGator [19] and ToppMir [20] and adds value with a tissue specific analysis of miRNA impact on signaling pathways. The integration of multiple new state-of-the art data sources increases the biological relevance of the results and a novel tissue filter allows every user to decipher complex miRNA functions.

Results

Tissue specific pathway enrichment

MiRNA target prediction tools and CLIP-seq based methods for target identification yield the full set of potential miRNA-mRNA interactions, i.e. all potential gene targets of a miRNA. However, different tissues and cell types have a characteristic gene expression signature and only a subset of genes are expressed in any cell under physiological conditions [12]. Thus, the function of miRNAs, which is exerted through repression of target genes, is tissue specific.

To learn about the tissue-specificities of miRNAs, we first analyzed the expression of all target genes of hsa-let-7a (TargetScan, see methods) in 42 human tissues from EBI Expression Atlas. The expression of target genes varied greatly between tissues (Fig 1A). To quantify the extent of tissue specificity of a miRNA, we calculated for each of the 42 tissues the fraction of target genes being expressed. The fraction is depicted in Fig 1B (color coded from green = 0 to red = 1). Fig 1C shows the respective distributions for ten representative miRNAs. Thereof, the median of target genes expressed in a tissue was 75%, with many tissues expressing only 60% target genes (Fig 1C). This is in line with studies showing tissue specific functions of miRNAs [13].

Next, we performed the same tissue-specificity analysis now only for genes of the same pathway. The pathway genes in well-described human MAPK signaling (KEGG) showed highly tissue specific expression (Fig 1D). Interestingly pathways showed a characteristic distribution of the fraction of expressed target genes when compared to miRNAs. Ten representative distributions across all 42 tissues are shown in Fig 1E. Some pathways (such as Cell adhesion molecules, Fig 1E) were more tissue specific than others, indicating highly tissue specific functions.

Having established that both miRNA and pathway associated genes have a characteristic gene expression signature across tissues, we next outlined the approach of standard miRNA pathway analysis methods. Typically the set of all miRNA targets are tested for over-representation in the set of all pathway genes (Fig 1F, left). This global analysis of all target and pathway genes will overlook miRNA-pathway associations with a small gene-overlap, while this gene-overlap may in turn be tissue-specific and, thus, functionally highly relevant. Pathway analysis tools that use all target genes to identify miRNA-pathway associations cannot capture tissue specific effects.

We thus propose a novel methodology for miRNA pathway analysis by using a tissue filter in order to increase the relevance of the association. If the target genes or pathway genes outside of the overlap are not expressed in a tissue, the relation of miRNA and pathway is much stronger (Fig 1F, right). Consequently, if the overlapping genes found in a miRNA-pathway association are not expressed in a tissue, the relation is discarded. The novel methodology calculates an enrichment of the target genes of a miRNA in all pathways of different pathway data sources. Significance of the associations is calculated with Fisher's exact test (see methods). Individually for each miRNA-pathway association test, we filtered for expression in a tissue by removing all miRNA target genes and pathway genes that are not expressed in this tissue. We thereby accounted for the highly tissue specific expression of many genes and seek to increase biological relevance of the pathway enrichment.
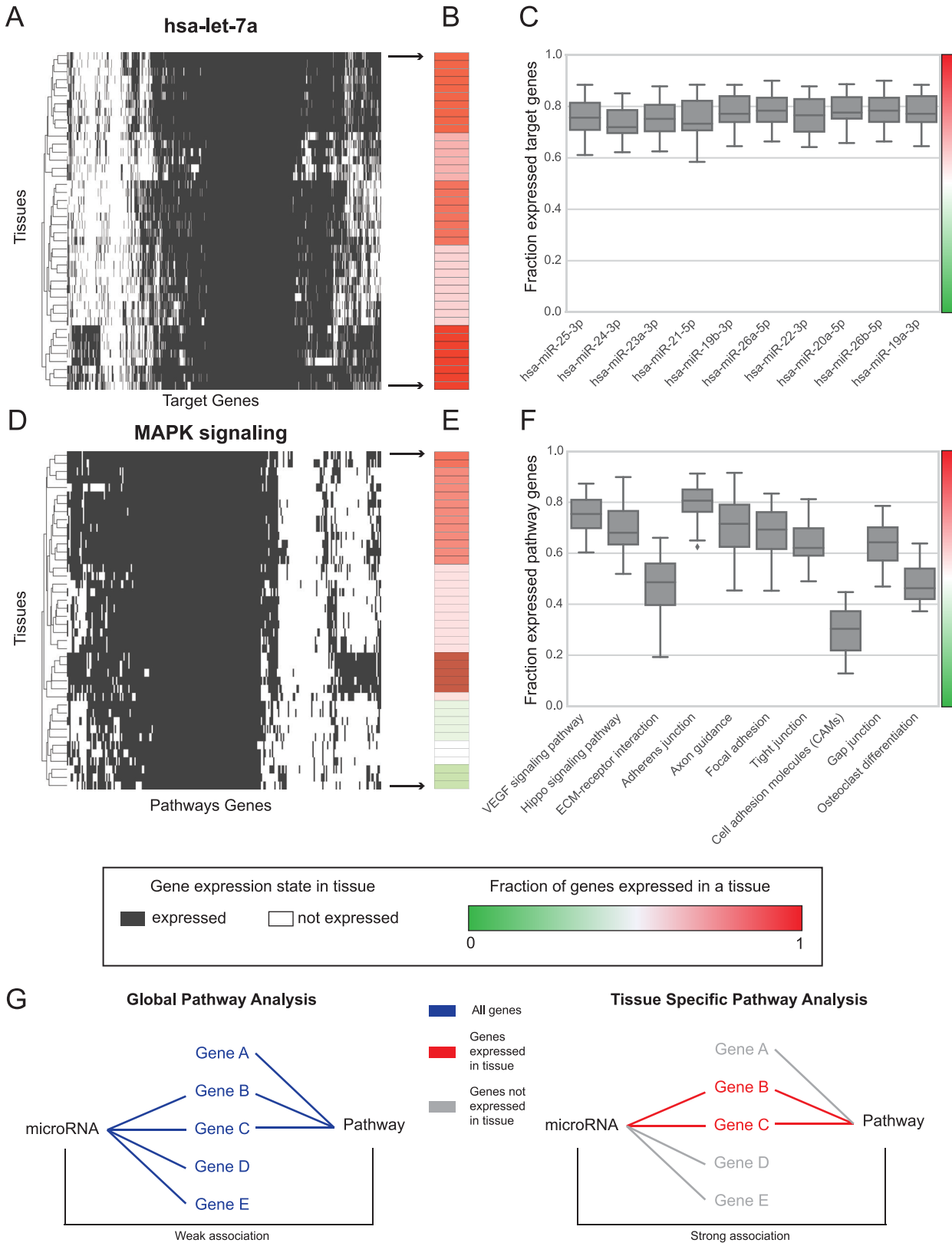
Fig 1. miRNA target genes and pathway genes are tissue specific. (A) Heatmap of all target genes of hsa-let-7a and their expression in 42 human tissues. Tissues are depicted in rows, genes in columns. (B) Fraction of target genes of hsa-let-7a expressed in each tissue, color coded in green (0) to red

(1). (C) Fraction of target genes expressed in all tissues for 10 representative miRNAs. (D)-(F) Corresponding analysis for pathway genes. (G) Pathway analysis with the global set of miRNA targets and pathway genes (left). The miRNA and pathway have only few common genes (gene B, gene C) compared to the other pathway genes (gene A) and miRNA targets (gene D, gene E). When applying a tissue filter (right), genes not in the set of miRNA targets and not in the pathway are discarded. The association derived from the overlap is much stronger, indicating a tissue specific regulation of the pathway by the miRNA.

doi:10.1371/journal.pone.0151771.g001

Case study: microRNAs in liver disease

We analyzed miRNAs known to be involved in liver disease with our novel methodology to evaluate the power of tissue specific pathway analysis. We focused on miRNAs in hepatocellular carcinoma (HCC) and liver fibrosis. Both diseases involve uncontrolled proliferation of liver cells.

First, we analyzed miR-199a-3p and miR-199b-3p. Both miRNAs are up-regulated in some tumor types, such as ovarian cancer and breast cancer [21]. In HCC, conversely, both miRNAs have been shown to be down-regulated [22,23]. While the function of miR-199a-3p and miR-199b-3p is not fully defined, they target members of Raf/MEK/ERK signaling [23]. In general, inhibition of Raf/MEK/ERK signaling will limit proliferation of cells. Thus, downregulation of miR-199a-3p and miR-199b-3p might be a part of the regulatory changes leading to increased proliferation of HCC cells. These miRNAs have consequently been considered as therapeutic targets for treatment of HCC [24].

When performing standard pathway analysis for miR-199a-3p and miR-199b-3p, no cancer-associated pathways were enriched (human, TargetScan). Using our methodology and the Illumina Body Map tissue filter for liver additionally identified two significantly associated pathways: Regulation of actin cytoskeleton (KEGG) and Regulation of Microtubule Cytoskeleton (WikiPathways) (Table 1). The miRNAs were previously not directly associated to regulation of the cytoskeleton, yet both pathways are fundamental for the processes of cell migration, EMT and metastasis. The regulation of actin cytoskeleton (KEGG) pathway overlaps with the MAPK signaling pathway from KEGG and includes several key components of Raf/MEK/ERK signaling (Fig 2A). The liver filter thus identified the known association of miR-199a-3p and miR-199b-3p with Raf/MEK/ERK signaling through associated regulatory pathways. Interestingly, the involvement of miR-199a/b-3p in cell migration and EMT has been described in other tissues [25,26].

Table 1. Pathway enrichment used in the case studies with liver filter.

| source	Pathway	E	Corrected p-value	MP, Mn, Pn, U
	hsa-miR-199a, hsa-miR-199b-3p			
wp	Regulation of Microtubule Cytoskeleton	3,814	0,034	4, 232, 18, 3982
kegg	Regulation of actin cytoskeleton	2,010	0,049	11, 225, 95, 3905
	hsa-miR-200b, hsa-miR-200c, hsa-miR-429			
kegg	Focal adhesion	1,895	0,008	25, 593, 83, 3730
wp	Focal Adhesion	1,791	0,024	22, 596, 77, 3736
wp	Wnt Signaling Pathway	2,765	0,028	8, 610, 18, 3795
	hsa-miR-200a, hsa-miR-141-3p			
wp	MAPK Cascade	3,194	0,047	5, 408, 15, 3910
reactome	NR1D1 (REV-ERBA) represses gene expression	19,095	0,017	2, 411, 1, 3924
	hsa-miR-571-3p			
wp	Notch Signaling Pathway	9,844	0,000	3, 62, 20, 4069
kegg	Notch signaling pathway	8,945	0,001	3, 62, 22, 4067

doi:10.1371/journal.pone.0151771.t001
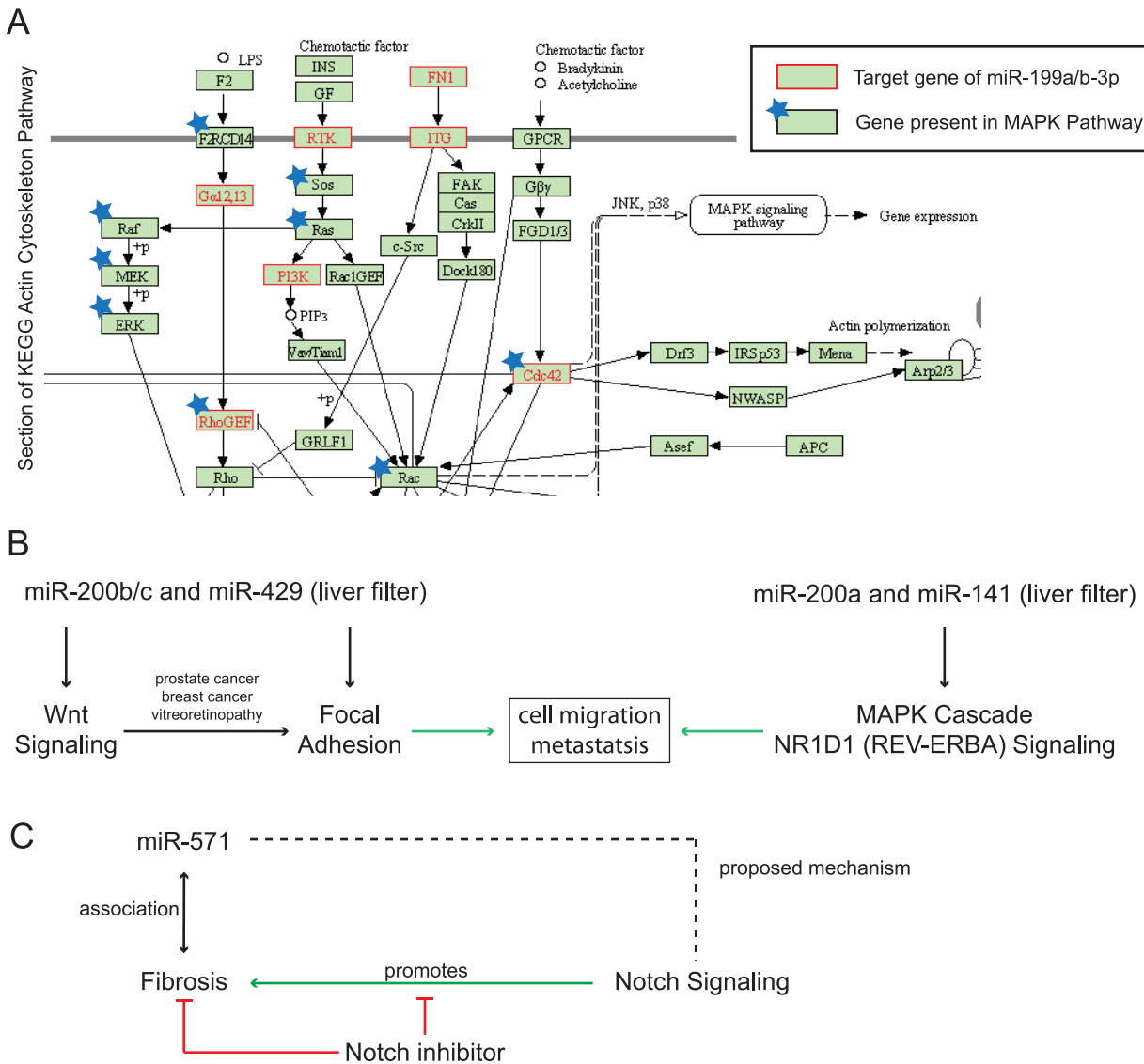
Fig 2. tissue specific enrichment of miRNAs in liver disease. (A) Targets of miR-199a/b-3p in the human KEGG Actin Cytoskeleton pathway (red). Only a section is shown, other parts are not targeted. Blue stars show genes also present in MAPK signaling. (B) Pathway analysis of miR-200 family using the liver filter. MiR-200b/c and miR-429 target Focal adhesion and Wnt signaling, pointing towards a regulatory interdependence in cancer formation. MiR-200a and miR-141 have different associated pathways but also target cancer related signaling. (C) MiR-571 is elevated in fibrosis and associated with notch signaling when using the liver filter. Notch inhibitors are in clinical studies for treatment of early stages of fibrosis.

doi:10.1371/journal.pone.0151771.g002

Our novel methodology with tissue filter suggested a role of miR-199a/b-3p in cell migration, EMT and ultimately metastasis through regulation of cytoskeleton. The decrease of miR-199a/b-3p in HCC might increase metastatic potential in HCC. Indeed, a recent study indicates a role for miR-199a/b-3p in HCC proliferation [27].

Second, we investigated the miR-200 family consisting of two genomic clusters (miR-200b/c/miR-429 and miR-200a/miR-141) that was shown to be involved in EMT and cell migration [28]. The family has been described as a potential cancer therapy target [29]. The miRNAs of the miR-200 family are often analyzed together. Here, we look at specific

functions of the two clusters to show the power of combined pathway analysis of multiple miRNAs. When performing pathway analysis with liver filter for miR-200b/c/miR-429 (Illumina Body Map, TargetScan, human) we identified significant associations with focal adhesion pathways from both KEGG and WikiPathways (Table 1). This finding clearly points towards an involvement in cell migration and EMT (Fig 2B). Interestingly, we also identified Wnt pathway (KEGG) (Table 1). As of today, there was no direct evidence reported for involvement of Wnt signaling in regulation of cell migration, EMT and metastatis in HCC. There was, however, evidence for a connection in other diseases such as breast cancer [30], vitreorenopathy [31] and prostate cancer [32].

Our novel methodology suggested new roles for miR-200b/c/miR-429 in HCC and a functional connection of Wnt signaling with cell migration and EMT (Fig 2B). Pathway analysis with liver filter (Illumina Body Map, TargetScan, human) for the other genomic cluster (miR-200a/141) identifies MAPK signaling and MAPK associated NR1D1-(REV-ERBA) pathway (WikiPathways) (Table 1). MAPK signaling was indeed elevated in HCC [33,34] and has been suggested as target for HCC treatment with success in mouse model [35] (Fig 2B). In summary, our novel methodology found specific HCC related functions for both genomic clusters of the miR-200 family. Analyzing the entire miRNA-200 family did not identify focal adhesion, Wnt or MAPK as significant results.

MiRNAs also play a role in liver fibrosis [36] but are in general less well studied in this disease context. There is only few functional evidence or mechanistic insight into the role of miRNAs in fibrosis. This represents an interesting example for the primary use case of our novel methodology: To generate new hypotheses and filter candidate miRNAs to be tested in the wet lab. The serum levels of miR-571 were found increased in cirrhosis (the final stage of fibrosis) and miR-571 has been suggested as a biomarker [37]. With our pathway analysis, we identified Notch signaling (KEGG) as target of miR-571 with liver filter (Illumina Body Map, human, TargetScan) (Table 1). Interestingly, Notch signaling was shown to be over-active in fibrosis [38] and Notch inhibitors have been discussed as potential drugs for treatment of fibrosis [38,39]. As a result, our novel methodology suggests that miR-571 could potentially inhibit Notch signaling in liver tissue. Thus, miR-571 might be a potential therapeutic target in the context of fibrosis (Fig 2C). In summary, our updated novel methodology supported new functional hypotheses through tissue filtered pathway analysis.

Data sources

In order to make the novel methodology publicly available, we first integrated several data sources on miRNA targeting, biological pathways and gene expression for both mouse and human. We downloaded and integrated computational target prediction data from TargetScan 6.2 [40] and miRanda [41]. We also added miRNA-target interaction data of CLIP-seq studies from StarBase v2 [5]. TargetScan contained the majority of mammalian miRNAs while miRanda and StarBase only represented a small subset (Table 1). Due to the limited availability of CLIP-seq studies, we still rely on target prediction data for many miRNAs. Pathway data was extracted from KEGG [16], Reactome [18] and WikiPathways [17]. Pathways in the Reactome database were structured in top-level pathways with smaller sub pathways. This lead to larger numbers of pathways overall compared to KEGG and WikiPathways (Table 1). To allow for a tissue-specific pathway analysis, we used baseline gene expression data for a total of 68 human and mouse tissues and cell lines from the latest EBI Expression Atlas [15]. Baseline expression data was based on reliable RNA-seq experiments and represents abundance levels in healthy tissue or cell lines. We integrated tissue data sets from 6 different expression studies (Table 2).

Table 2. Overview of data sets.

	Human	Mouse
# miRNAs		
TargetScan v6	1529	1322
Miranda	249	238
StarBase v2	383	296
# Pathways		
KEGG	295	291
Reactome	2224	1882
WikiPathways	293	160
# Tissues		
Mammalian Tissues	8	6
Illumina Body Map	16	-
ENCODE Cell Lines	18	-
Vertebrate Tissues	-	5
6 Mouse Tissues	-	6
Nine Mouse Tissues	-	9

doi:10.1371/journal.pone.0151771.t002

Database backend

Any system that integrates heterogeneous research data has to deal with two major challenges: I) Data has to be stored in a way that it can be queried efficiently and II) the data model must allow for easy updates for new releases of the underlying data sources.

Traditionally, SQL based relational database systems such as MySQL or PostgreSQL were the go-to solution for all data storage needs. In recent years however, new database technologies collectively termed noSQL (short for not-only SQL) were developed to cope with problems arising from big data. Such noSQL technologies have been used successfully in solutions for computational biology, especially in the field of NGS [42]. Among the diverse landscape of new database technologies, graph databases are particularly promising for biological data sets. They enable storing data natively as a property graph, i.e. nodes connected by edges with properties stored on both. Thus, they allow us to directly model biological systems as nodes representing molecular entities connected by edges representing their interaction. This leads to simple queries over multi-step paths through the interaction network and increased performance compared to JOIN operations in relational databases [43,44]. Since queries on biological data are usually centered on relationships between molecular entities (such as genes and miRNAs), graph database have a huge potential to improve data storage solutions. The key advantages are query performance and simple query syntax.

For our study, we used the graph database neo4j and developed a novel graph data model to integrate the data sources described above (Fig 3A). MiRNAs, genes, pathways and tissues were represented as nodes. MiRNAs were connected to genes with 'REGULATES' relationships, genes to tissues with 'EXPRESSED`relationships and genes to pathways with 'MEMBER' relationships. This data structure allowed us to e.g. query the target genes of a miRNA expressed in a tissue (Fig 3B, top) or the pathways in which the target genes are involved (Fig 3B, bottom).

Another challenge in studies based on integration of third party data sources is to keep up with data updates and new releases. Small, specialized data sources publish new versions on their own schedule and changes in one data source are not synchronized with others. Since neo4j is schema-less, changes of parts of the underlying data (e.g. miRNA targeting data for a single data source) and refactoring of the data structure (e.g. renaming of miRNAs) are easier
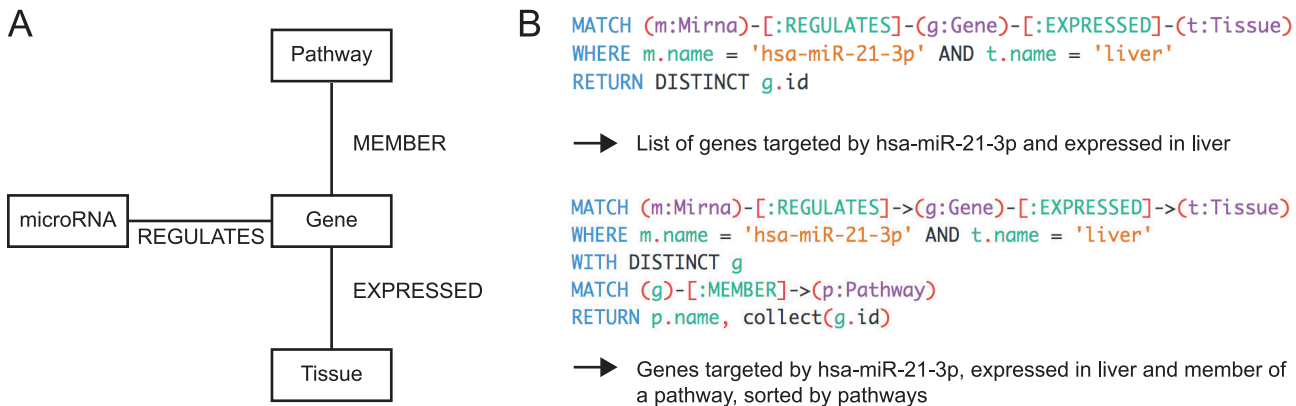
A

B
```
MATCH (m:Mirna)-[:REGULATES]-(g:Gene)-[:EXPRESSED]-(t:Tissue)
WHERE m.name = 'hsa-miR-21-3p' AND t.name = 'liver'
RETURN DISTINCT g.id
```

→ List of genes targeted by hsa-miR-21-3p and expressed in liver

```
MATCH (m:Mirna)-[:REGULATES]->(g:Gene)-[:EXPRESSED]->(t:Tissue)
WHERE m.name = 'hsa-miR-21-3p' AND t.name = 'liver'
WITH DISTINCT g
MATCH (g)-[:MEMBER]->(p:Pathway)
RETURN p.name, collect(g.id)
```

→ Genes targeted by hsa-miR-21-3p, expressed in liver and member of a pathway, sorted by pathways

Fig 3. Database structure of miTALOS v2. (A) The miTALOS v2 dataset is stored in a graph database. The network structure allows for easy extension of the dataset. (B) The Cypher query language allows for simple queries on the network. With one query, the targets of a miRNA in a pathway can be accessed and filtered for tissue expression.

to implement. We thus seek to regularly update our pathway analysis with new data sets especially focusing on NGS based data for miRNA targets and gene expression.

miTALOS v2

In order to make our integrated, tissue specific pathway analysis available to the research community, we included the new analysis methodology and data backend in an update to our miTALOS web application.

MiTALOS v2 is a user-friendly tool to perform tissue specific pathway analysis for a set of miRNAs and tissues of interest (Fig 4). It is available at http://mips.helmholtz-muenchen.de/mitalos. The user can analyze miRNAs from mouse and human. The user begins by selecting the organism and miRNA prediction method (Fig 4A) and then selects one or multiple miR-NAs (Fig 4B). The pathway analysis is carried out dynamically by calculating the pathway enrichment (see Methods) on all pathway data sources. If more than one miRNA is selected, the union of target genes will be used for the analysis. All target genes are counted once and no additional ranking is applied. MiTALOS v2 thereby captures the biological impact of co-targeting by multiple miRNAs. If the user selects a tissue filter, all gene sets (miRNA target genes and genes in pathways) are filtered for this tissue (Fig 4C). All results with a corrected p-value > 0.05 and $E > 1$ are presented in a sortable table and can be accessed with a user specific URL for one week. For KEGG pathways, the user can access a graphical representation of the pathway with highlighted miRNA targets by clicking on a pathway name.

If a tissue filter is used, miTALOS v2 displays the expression score of the selected miRNAs inaddition to the tissue specific pathway enrichment. The user can thereby assess the impact of the selected miRNAs under physiological conditions. The absolute expression score is extended by a rank of the selected miRNA among all miRNAs expressed in this tissue and the miRNA with the overall highest expression value. This allows estimating the relative importance of the selected miRNA in the analyzed tissue.

MiTALOS v2 is geared towards wet-lab researchers working with miRNAs. MiTALOS v2 was designed for scenarios where a set of miRNAs (e.g. from expression studies or literature research) has to be filtered to identify the most promising miRNAs for testing in wet-lab experiments. With the tissue filter, the user can analyze the supposed biological effect of miRNAs in the particular tissue or cell line the user is working on.
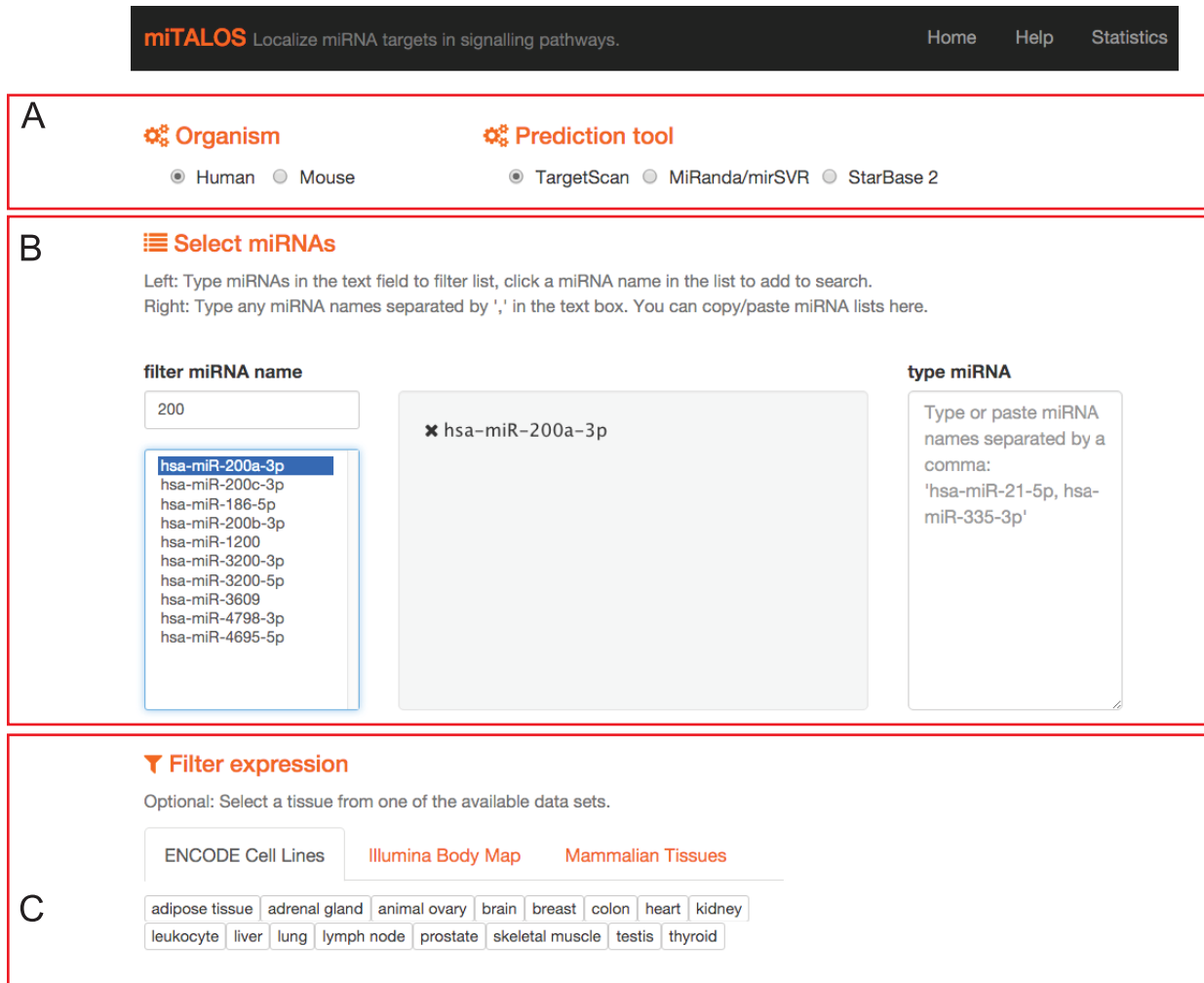
Fig 4. User interface of miTALOS v2. (A) The user starts by selecting the organism and miRNA prediction tool. Next, multiple miRNAs can be selected by filtering the list of available miRNAs (B). Lastly, a tissue filter can be applied by selecting an expression experiment and tissue or cell line (C).

doi:10.1371/journal.pone.0151771.g004

Discussion

It has been established that miRNAs participate in almost all cellular processes but the functional impact of individual miRNAs and the precise mode of target gene regulation remains controversial. Consequently, the dynamic regulatory network of miRNAs and mRNAs under physiological conditions is not fully understood. One of the key issues in miRNA resarch is the identification and quantification of miRNA-mRNA interactions. While computational prediction methods and CLIP-seq approaches yield global sets of gene targets for individual miRNAs, they still suffer from lack of accuracy and fail to predict the regulatory landscape in-vivo.

One way to circumvent shortcomings in miRNA targeting data is to analyze the biological pathways which are incluenced by miRNAs. They can be considered a proxy for the miRNAs effect on biological processes and thus allow to classify miRNAs and generate new hypotheses. While pathway analyses have proven useful, they do not consider that most genes which are targeted by a miRNA or part of a pathway are not uniformly expressed across all cell types. The tissue specifity of miRNAs, which has been demonstrated extensively, is thus not taken into account.

By integrating tissue specific gene expression into our pathway analysis methodology, we seek to close this gap and improve the biological relevance of our miRNA-pathway associations. With our case studies, we recapitulated a common approach to generate new miRNA hypotheses for wet lab research: Based on prior knowledge, i.e. disregulation of several miRNAs in a disease context, the best candidates for experimental testing have to be identified. Our methodology aims at creating functional insight which is as specific as possible for the system studied by the user.

The distinctive feature of miTALOS v2 is the tissue specific pathway enrichment. Other pathway analysis tools, such as DIANA mirPath [45], do not account for this effect. MiTALOS v2 complements other methods for functional miRNA analysis. Tools analyzing the expression of miRNAs, such as MiRGator [19], aid in selecting the best miRNA candidates for a specific biological system. Ranking approaches, such as ToppMir [20], are used to limit the number of miRNAs based on preference for user-defined gene sets. MiTALOS v2 can be used in conjunction with these methods and adds a tissue specific perspective.

MiTALOS v2 includes CLIP-seq based miRNA targeting data from the StarBase database. CLIP-seq experiments generate the full set of target genes based on biochemically identified miRNA-mRNA interactions and likely produce more reliable targeting data than computational prediction. Several public resources, such as miRTarBase [46] and miRecords [47], collect miRNA targets validated in individual experiments. However, since these target sets contain only a potentially small subset of miRNA-mRNA interactions they would introduce a bias to the analysis and are thus not suitable for global pathway enrichment.

Next to TargetScan and miRanda, which were used in this study, there are several other miRNA target prediction tools. However, it is difficult to compare their performance due to the lack of a gold standard of known miRNA targets and systematic comparisons of target prediction tools generated inconsistent results [48–51]. TargetScan and miRanda were chosen based on their widespread use in the miRNA research community. If novel miRNA target data sources arise, the miTALOS v2 data can easily be integrated in miTALOS v2.

In general, the effect of a miRNA on its target genes cannot be quantified cell wide. The complexity of the miRNA-mRNA network was further increased when regulatory effects came into focus [52]. It was demonstrated that the total number of potential binding sites for a miRNA regulates its effect size. If the number of binding sites exceeds the number of miRNA molecules, mRNAs compete for binding to the miRNA and the regulatory impact decreases [53]. This has been subsumed under the concept of competing endogenous RNAs (ceRNAs). Recently, combined computational and experimental studies quantified these effects on a systems level [54]. Including these indirect effects into a pathway analysis presents a future direction for miTALOS v2. Here, using the relative expression levels of miRNAs and their target genes would allow to capture binding competition. However, more data on specific, quantitative effects will be necessary to devise a computational approach that properly describes the biological impact of competing RNAs.

When developing tools for the research community, the underlying data infrastructure is of pivotal importance. The state of the art, especially in research of post-transcriptional regulation, changes quickly and new methods for miRNA target identification might arise. We therefore developed a new database backend using neo4j, the leading graph database. It helps to integrate the numerous datasets used in miTALOS v2 and to keep up with new developments. The flexible backend also allows to integrate new aspects like lncRNAs as regulators of gene expression or disease specific expression profiles to extend tissue specific gene expression. New database technology is therefore instrumental in building tools which can adapt to the rapid generation of new research results.

In summary, our pathway analysis methodology and miTALOS v2 have been developed to generate testable hypotheses and to increase efficiency in experimental miRNA research.

Table 3. 2x2 cross table.

	Pathway *P*	
miRNA *M*	MP	Mn
	Pn	U

doi:10.1371/journal.pone.0151771.t003

Methods

Datasets

We integrated several data sources on miRNA targeting, biological pathways and gene expression in order to analyse tissue specific miRNA functions. For mouse and human, we offer computational target prediction data from the latest releases of TargetScan 6.2 [40] and miRanda [41]. We added miRNA-target interaction data of CLIP-seq studies from StarBase v2 [5] to the miTALOS v2 pathway analysis. Pathway data was extracted from KEGG, Reactome and WikiPathways. In order to analyze tissue specific pathway regulation, miTALOS v2 uses baseline gene expression data for 68 tissues and cell lines from the latest EBI Expression Atlas [15] for both mouse and human.

Pathway analysis

We calculate an enrichment of miRNA target genes in pathways. For a miRNA *M* and Pathway *P* miTALOS v2 calculates a 2x2 cross table, where *MP* is the number of targets of *M* in *P*, *Pn* is the number of not targeted genes in *P*, *Mn* is the number of targets of *M* not in *P* and *U* is the union of all pathway genes and miRNA targets without *MP*, *Pn* and *Mn* (Table 3):

An enrichment score *E* is calculated as the odds ratio of *M* and *P*:

$$E(M, P) = (MP/Pn)/(Mn/U)$$

E describes the dependence of variables *M* and *P*. $E > 1$ indicates an over-representation of targets of miRNA *M* in the pathway *P*. A p-value is calculated using Fisher's exact test and results for multiple pathways are corrected using the Benjamini-Hochberg procedure [55].

To perform a tissue specific pathway enrichment, we remove all genes from *MP*, *Mn*, *Pn* and *U* that are not expressed in the analyzed tissue. We then calculate *E* as described above. A gene is considered expressed if its baseline expression value is > 0.5 (as defined in the EBI Expression Atlas).

When multiple miRNAs are selected, the union of target genes is used for the analysis.

Database and webinterface

The integrated database backend is uses a neo4j graph database (v2.3.1). The miTALOS v2 frontend was developed with AngularJS 1.4.

Author Contributions

Conceived and designed the experiments: MP NSM FJT. Performed the experiments: MP. Analyzed the data: MP. Wrote the paper: MP NSM.

References

1. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19: 92–105. doi: 10.1101/gr.082701.108 PMID: 18955434

2. Thomas M, Lieberman J, Lal A. Desperately seeking microRNA targets. Nat Struct Mol Biol. Nature Publishing Group; 2010; 17: 1169–74. doi: 10.1038/nsmb.1921

3. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. Nature Publishing Group; 2009; 460: 479–86. doi: 10.1038/nature08170

4. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. Elsevier Ltd; 2010; 141: 129–41. doi: 10.1016/j.cell.2010.03.009

5. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2013; 1–6.

6. Jungkamp AC, Stoeckius M, Mecenas D, Grün D, Mastrobuoni G, Kempa S, et al. In vivo and transcriptome-wide identification of RNA binding protein target sites. Mol Cell. 2011; 44: 828–840. doi: 10.1016/j.molcel.2011.11.009 PMID: 22152485

7. Vidigal J a., Ventura A. The biological functions of miRNAs: lessons from in vivo studies. Trends Cell Biol. Elsevier Ltd; 2014; 25: 137–147. doi: 10.1016/j.tcb.2014.11.004

8. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008; 455: 64–71. doi: 10.1038/nature07242 PMID: 18668037

9. Cui Q, Yu Z, Purisima EO, Wang E. Principles of microRNA regulation of a human cellular signaling network. Mol Syst Biol. 2006; 2: 46. doi: 10.1038/msb4100089 PMID: 16969338

10. Inui M, Martello G, Piccolo S. MicroRNA control of signal transduction. Nat Rev Mol Cell Biol. 2010; 11: 252–63. doi: 10.1038/nrm2868 PMID: 20216554

11. Rinck A, Preusse M, Laggerbauer B, Lickert H, Engelhardt S, Theis FJ. The human transcriptome is enriched for miRNA-binding sites located in cooperativity-permitting distance. RNA Biol. 2013; 10: 1125–35. doi: 10.4161/rna.24955 PMID: 23696004

12. Su AI, Wiltshire T, Batalov S, Lapp H, Ching K a, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004; 101: 6062–7. doi: 10.1073/pnas.0400782101 PMID: 15075390

13. Farh KK-H, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, et al. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science. 2005; 310: 1817–21. doi: 10.1126/science.1121158 PMID: 16308420

14. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, et al. DIANA miRPath v.2.0: Investigating the combinatorial effect of microRNAs in pathways. Nucleic Acids Res. 2012; 40: 498–504. doi: 10.1093/nar/gks494

15. Petryszak R, Burdett T, Fiorelli B, Fonseca N a, Gonzalez-Porta M, Hastings E, et al. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic Acids Res. 2014; 42: D926–32. doi: 10.1093/nar/gkt1270 PMID: 24304889

16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27–30. doi: 10.1093/nar/28.1.27 PMID: 10592173

17. Kelder T, Van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: Building research communities on biological pathways. Nucleic Acids Res. 2012; 40: 1301–1307. doi: 10.1093/nar/gkr1074

18. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011; 39: D691–7. doi: 10.1093/nar/gkq1018 PMID: 21067998

19. Cho S, Jang I, Jun Y, Yoon S, Ko M, Kwon Y, et al. MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. Nucleic Acids Res. 2013; 41: D252–7. doi: 10.1093/nar/gks1168 PMID: 23193297

20. Wu C, Bardes EE, Jegga AG, Aronow BJ. ToppMiR: ranking microRNAs and their mRNA targets based on biological functions and context. Nucleic Acids Res. 2014; 42: W107–13. doi: 10.1093/nar/gku409 PMID: 24829448

21. Chen R, Alvero a B, Silasi D a, Kelly MG, Fest S, Visintin I, et al. Regulation of IKKbeta by miR-199a affects NF-kappaB activity in ovarian cancer cells. Oncogene. 2008; 27: 4712–23. doi: 10.1038/onc.2008.112 PMID: 18408758

22. Murakami Y, Yasuda T, Saigo K, Urashima T, Toyoda H, Okanoue T, et al. Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. Oncogene. 2006; 25: 2537–45. doi: 10.1038/sj.onc.1209283 PMID: 16331254

23. Hou J, Lin L, Zhou W, Wang Z, Ding G, Dong Q, et al. Identification of miRNomes in human liver and hepatocellular carcinoma reveals miR-199a/b-3p as therapeutic target for hepatocellular carcinoma. Cancer Cell. Elsevier Inc.; 2011; 19: 232–43. doi: 10.1016/j.ccr.2011.01.001

24. Callegari E, Elamin BK, D'Abundo L, Falzoni S, Donvito G, Moshiri F, et al. Anti-tumor activity of a miR-199-dependent oncolytic adenovirus. PLoS One. 2013; 8: e73964. doi: 10.1371/journal.pone.0073964 PMID: 24069256

25. Duan Z, Choy E, Harmon D, Liu X, Susa M, Mankin H, et al. MicroRNA-199a-3p Is Downregulated in Human Osteosarcoma and Regulates Cell Proliferation and Migration. Mol Cancer Ther. 2011; 10: 1337–1345. doi: 10.1158/1535-7163.MCT-11-0096 PMID: 21666078

26. Bonet F, Dueñas Á, López-Sánchez C, García-Martínez V, Aránega AE, Franco D. MiR-23b and miR-199a impair epithelial-to-mesenchymal transition during atrioventricular endocardial cushion formation. Dev Dyn. 2015; 244: 1259–1275. doi: 10.1002/dvdy.24309 PMID: 26198058

27. Song J, Gao L, Yang G, Tang S, Xie H, Wang Y, et al. MiR-199a regulates cell proliferation and survival by targeting FZD7. PLoS One. 2014; 9: e110074. doi: 10.1371/journal.pone.0110074 PMID: 25313882

28. Gregory P a, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008; 10: 593–601. doi: 10.1038/ncb1722 PMID: 18376396

29. Humphries B, Yang C. The microRNA-200 family: small molecules with novel roles in cancer development, progression and therapy. Oncotarget. 2015; 6: 6472–98. PMID: 25762624

30. Wu Z-Q, Li X-Y, Hu CY, Ford M, Kleer CG, Weiss SJ. Canonical Wnt signaling regulates Slug activity and links epithelial-mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression. Proc Natl Acad Sci U S A. 2012; 109: 16654–9. doi: 10.1073/pnas.1205822109 PMID: 23011797

31. Chen H-C, Zhu Y-T, Chen S-Y, Tseng SCG. Wnt signaling induces epithelial-mesenchymal transition with proliferation in ARPE-19 cells upon loss of contact inhibition. Lab Invest. Nature Publishing Group; 2012; 92: 676–87. doi: 10.1038/labinvest.2011.201

32. Jiang Y-G, Luo Y, He D, Li X, Zhang L, Peng T, et al. Role of Wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha. Int J Urol. 2007; 14: 1034–9. doi: 10.1111/j.1442-2042.2007.01866.x PMID: 17956532

33. Schmidt CM, McKillop IH, Cahill P a, Sitzmann J V. Increased MAPK expression and activity in primary human hepatocellular carcinoma. Biochem Biophys Res Commun. 1997; 236: 54–8. doi: 10.1006/bbrc.1997.6840 PMID: 9223425

34. Huynh H, Nguyen TTT, Chow K-HP, Tan PH, Soo KC, Tran E. Over-expression of the mitogen-activated protein kinase (MAPK) kinase (MEK)-MAPK in hepatocellular carcinoma: its role in tumor progression and apoptosis. BMC Gastroenterol. 2003; 3: 19. doi: 10.1186/1471-230X-3-19 PMID: 12906713

35. Liu L, Cao Y, Chen C, Zhang X, McNabola A, Wilkie D, et al. Sorafenib blocks the RAF/MEK/ERK pathway, inhibits tumor angiogenesis and induces tumor cell apoptosis in hepatocellular carcinoma model PLC/PRF/5. Cancer Res. 2006; 66: 11851–11858. doi: 10.1158/0008-5472.CAN-06-1377 PMID: 17178882

36. Noetel A, Kwiecinski M, Elfimova N, Huang J, Odenthal M. microRNA are Central Players in Anti- and Profibrotic Gene Regulation during Liver Fibrosis. Front Physiol. 2012; 3: 49. doi: 10.3389/fphys.2012.00049 PMID: 22457651

37. Roderburg C, Mollnow T, Bongaerts B, Elfimova N, Vargas Cardenas D, Berger K, et al. Micro-RNA profiling in human serum reveals compartment-specific roles of miR-571 and miR-652 in liver cirrhosis. Lafrenie R, editor. PLoS One. 2012; 7: e32999. doi: 10.1371/journal.pone.0032999 PMID: 22412969

38. Sweetwyne MT, Tao J, Susztak K. Kick it up a notch: Notch signaling and kidney fibrosis. Kidney Int Suppl. 2014; 4: 91–96. doi: 10.1038/kisup.2014.17

39. Morell CM, Strazzabosco M. Notch signaling and new therapeutic options in liver disease. J Hepatol. 2014; 60: 885–90. doi: 10.1016/j.jhep.2013.11.028 PMID: 24308992

40. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol. Nature Publishing Group; 2011; 18: 1139–1146. doi: 10.1038/nsmb.2115

41. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. 2010; 11: R90. doi: 10.1186/gb-2010-11-8-r90 PMID: 20799968

42. de Brevern AG, Meyniel J-P, Fairhead C, Neuvéglise C, Malpertuy A. Trends in IT Innovation to Build a Next Generation Bioinformatics Solution to Manage and Analyse Biological Big Data Produced by NGS Technologies. Biomed Res Int. 2015; 2015: 904541. doi: 10.1155/2015/904541 PMID: 26125026

43. Vicknair C, Macias M, Zhao Z, Nan X, Chen Y, Wilkins D. A comparison of a graph database and a relational database. Proceedings of the 48th Annual Southeast Regional Conference on—ACM SE '10. New York, New York, USA: ACM Press; 2010. p. 1.

44. Jouili S, Vansteenberghe V. An Empirical Comparison of Graph Databases. 2013 International Conference on Social Computing. IEEE; 2013. pp. 708–715.

45. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res. 2015; 43: W460–6. doi: 10.1093/nar/gkv403 PMID: 25977294

46. Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011; 39: D163–9. doi: 10.1093/nar/gkq1107 PMID: 21071411

47. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009; 37: D105–10. doi: 10.1093/nar/gkn851 PMID: 18996891

48. Sethupathy P, Megraw M, Hatzigeorgiou AG. A guide through present computational approaches for the identification of mammalian microRNA targets. Nat Methods. 2006; 3: 881–6. doi: 10.1038/nmeth954 PMID: 17060911

49. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics. 2009; 25: 3049–55. doi: 10.1093/bioinformatics/btp565 PMID: 19789267

50. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008; 455: 58–63. doi: 10.1038/nature07228 PMID: 18668040

51. Fan X, Kurgan L. Comprehensive overview and assessment of computational prediction of microRNA targets in animals. Brief Bioinform. Oxford University Press; 2015; 16: 780–94. doi: 10.1093/bib/bbu044

52. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014; 505: 344–52. doi: 10.1038/nature12986 PMID: 24429633

53. Yang J, Li T, Gao C, Lv X, Liu K, Song H, et al. FOXO1 3'UTR functions as a ceRNA in repressing the metastases of breast cancer cells via regulating miRNA activity. FEBS Lett. Federation of European Biochemical Societies; 2014; 588: 3218–24. doi: 10.1016/j.febslet.2014.07.003

54. Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, et al. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proc Natl Acad Sci U S A. 2015; 112: 3158–63. doi: 10.1073/pnas.1413896112 PMID: 25713348

55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995; 57: 289–300.