SCIENTIFIC REPORTS
natureresearch

OPEN

The integrative knowledge base for miRNA-mRNA expression in colorectal cancer

Daša Jevšinek Skok[1,2,3*], Nina Hauptman[1,3], Emanuela Boštjančič[1] & Nina Zidar[1]

"miRNA colorectal cancer" (https://mirna-coadread.omics.si/) is a freely available web application for studying microRNA and mRNA expression and their correlation in colorectal cancer. To the best of our knowledge, "miRNA colorectal cancer" has the largest knowledge base of miRNA-target gene expressions and correlations in colorectal cancer, based on the largest available sample size from the same source of data. Data from high-throughput molecular profiling of 295 colon and rectum adenocarcinoma samples from The Cancer Genome Atlas was analyzed and integrated into our knowledge base. The objective of developing this web application was to help researchers to discover the behavior and role of miRNA-target gene interactions in colorectal cancer. For this purpose, results of differential expression and correlation analyses of miRNA and mRNA data collected in our knowledge base are available through web forms. To validate our knowledge base experimentally, we selected genes *FN1*, *TGFB2*, *RND3*, *ZEB1* and *ZEB2* and miRNAs *hsa-miR-200a/b/c-3p*, *hsa-miR-141-3p* and *hsa-miR-429*. Both approaches revealed a negative correlation between miRNA *hsa-miR-200b/c-3p* and its target gene *FN1* and between *hsa-miR-200a-3p* and its target *TGFB2*, thus supporting the usefulness of the developed knowledge base.

microRNAs (miRNAs) are approximately 21-nucleotide long RNA molecules involved in post-transcriptional regulation by binding to their mRNA targets[1]. RNA interference (RNAi) is one of the processes by which miRNAs regulate gene expression and is one of the most promising biological processes for the development of therapeutic by gene silencing. Each small RNA forms a sequence-specific, gene-silencing ribonucleoprotein, with specificity conferred by base-pairing between the small (guide) RNA and its target mRNA. There are many combinations of miRNAs and mRNAs, since miRNAs can bind completely to a complementary sequence of mRNAs, but can also bind incompletely (some nucleotides can have a mismatch complement base). The number of miRNA and mRNA pairs is increased because one miRNA can target multiple genes, and one gene can also be targeted by multiple miRNAs.

miRNAs are involved in the regulation of physiological processes, such as development, epithelial-mesenchymal transition (EMT), regulation of homeostasis and metabolism, etc. However, many miRNAs have been proven to be deregulated in many diseases, including cancer. They are involved in cell transformation from normal to malignant, including in colorectal cancer (CRC), which is one of the most common cancers and one of the leading causes of death worldwide[2]. There are many relationships among miRNAs and genes known to be involved in the initiation and progression of CRC[3]. Some of the known frequently inactivated genes in CRC are *APC*, *TGFBR2*, *TP53*, *SMAD4*, *PTEN*, constitutively activated *KRAS* or overexpressed *MYC* in CRC[3,4]. In addition to protein-coding genes and mRNAs, there are also miRNAs that are known to regulate CRC tumor-initiating cells, such as *hsa-miR-34a*[5,6], *hsa-miR-106b*[7], *hsa-miR-140*[8], *hsa-miR-146a*[9], *hsa-miR-183*[10], *hsa-miR-200*[10], *hsa-miR-203*[10], *hsa-miR-215*[11], *hsa-miR-302b*[12], *hsa-miR-328*[13], *hsa-miR-363*[14], *hsa-miR-371*[15] and *hsa-miR-451*[16]. In CRC, miRNA-mediated modulation thus regulates features of cellular transformation. However, many miRNAs also function downstream of these factors[3]. Furthermore, miRNA deregulation is also correlated to angiogenesis, proliferation and migration of cancer cells in CRC, thus contributing to cancerogenesis and invasion[17–20]. It is therefore crucial to increase our understanding of the role of miRNAs in CRC, and to discover novel therapeutic targets to improve diagnostic, prognostic and treatment options.

[1]Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. [2]Agricultural Institute of Slovenia, Ljubljana, Slovenia. [3]These authors contributed equally: Daša Jevšinek Skok and Nina Hauptman. *email: dasa.jevsinekskok@kis.si

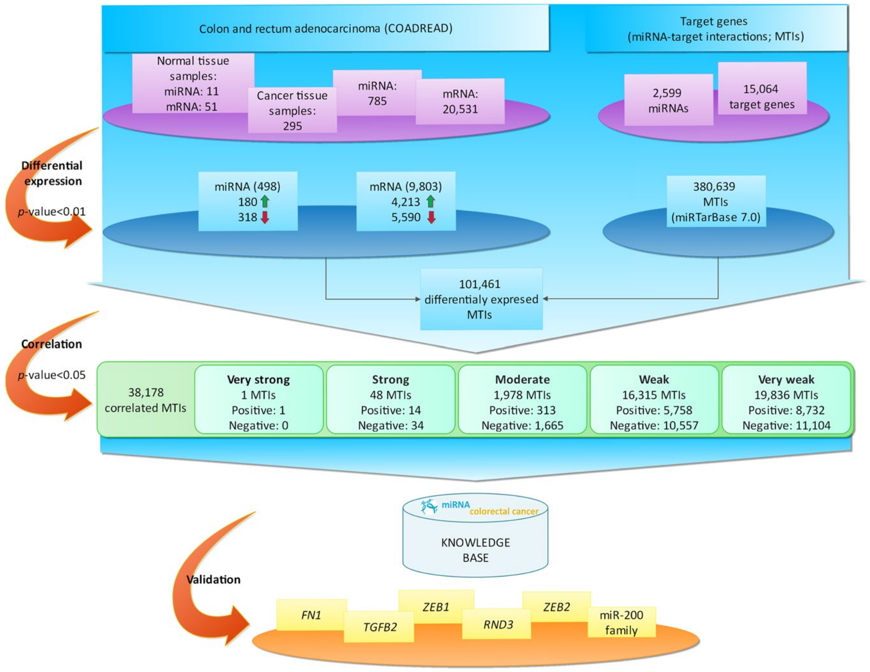

Colon and rectum adenocarcinoma (COADREAD)
Target genes (miRNA-target interactions; MTIs)
Normal tissue samples: miRNA: 11 mRNA: 51
Cancer tissue samples: 295
miRNA: 785
mRNA: 20,531
2,599 miRNAs
15,064 target genes
Differential expression
p-value<0.01
miRNA (498) 180 ↑ 318 ↓
mRNA (9,803) 4,213 ↑ 5,590 ↓
380,639 MTIs (miRTarBase 7.0)
101,461 differentialy expresed MTIs
Correlation
p-value<0.05
38,178 correlated MTIs
Very strong 1 MTIs Positive: 1 Negative: 0
Strong 48 MTIs Positive: 14 Negative: 34
Moderate 1,978 MTIs Positive: 313 Negative: 1,665
Weak 16,315 MTIs Positive: 5,758 Negative: 10,557
Very weak 19,836 MTIs Positive: 8,732 Negative: 11,104
miRNA colorectal cancer
KNOWLEDGE BASE
Validation
FN1
TGFB2
ZEB1
RND3
ZEB2
miR-200 family

Figure 1. The workflow of the study. "miRNA colorectal cancer" was constructed from The Cancer Genome Atlas (TCGA) and from the miRTarBase databases. From TCGA, samples of Colon adenocarcinoma (COAD) and Rectum adenocarcinoma (READ) projects were used. Data from miRNA experiments obtained 11 normal samples, while data from mRNA obtained 51 normal samples. Number of cancer tissue samples with data from both miRNA and mRNA experiments was 295. Total number of miRNA in data from miRNA experiments was 785 and total number of genes from mRNA experiments was 20,531. Analysis of differential expression gave 498 differentially expressed miRNAs of which 180 were upregulated and 318 were down regulated. Differential expression analysis of mRNA gave 9,803 deregulated mRNA of which 4,213 were upregulated and 5,590 were down regulated. miRTarBase contains 2,599 miRNAs and 15,064 target genes from which there are 101,461 validated miRNA-target interactions (MTIs). Deregulated miRNAs and mRNAs were combined according to validated MTIs downloaded from miRTarBase and correlations were calculated between the pairs of miRNA and mRNA involved in one MTI. Correlation analysis revealed 38,178 correlated MTIs of which one MTI had very strong correlation, 48 MTIs had strong correlation, 1,978 MTIs had moderate correlation, 16,315 MTIs had weak correlation and 19,836 MTIs had very weak correlation. Results from differential expression and correlation analyses are combined in web application "miRNA colorectal cancer". The knowledge base was further validated with wet-lab experiment, where we analyzed correlation between genes *FN1*, *SOX2*, *TGFB2* and *ZEB1* and *miR-200* family.

Bioinformatics and data-mining are being increasingly used to study a multi-omics approach to understanding the biology behind diseases. New technologies that enable large-scale data are being used to create databases. One of them is The Cancer Genome Atlas (TCGA), which is the largest cancer database with clinical and experimental genomics data. Another database is miRTarBase, which summarizes all low-scale and large-scale experimentally confirmed miRNA-mRNA pairs in various diseases[21].

To identify mRNA-miRNA pairs that might serve as starting point to validate new potential prognostic and therapeutic targets, we performed an integrative analysis of miRNA-target gene interactions (MTIs) on high-throughput molecular profiling of 295 CRC samples from TCGA. We also experimentally validated 10 MTIs on our clinical samples. Figure 1 illustrates the workflow of the study. It consists of the following steps: data collection, statistics and data visualization, *in silico* screening for miRNA-mRNA interactions in CRC, and experimental validation.

The objective of this study was to help researchers in the field of miRNA-target gene interactions to decide which is the best miRNA or target gene to be validated in an experimental setting, whether or not they have previous experience in the field. For this purpose, we developed a freely available tool entitled "miRNA colorectal cancer". The researchers can select a miRNA or gene they want to investigate further or they can use our tool even if they have no pre-selected miRNA or target gene.

Results

Our knowledge base is based on high-throughput molecular profiling of colon and rectum adenocarcinoma samples from TCGA. We analyzed mRNA and miRNA expression in CRC samples and calculated the correlation coefficient for any of the known validated MTIs. The results of this meta-analysis are freely available through the web application "miRNA colorectal cancer", which provides simple searching and reviewing of expression profiles of mRNA and miRNA and their correlation in CRC samples.

Analysis of mRNA and miRNA expression in colorectal adenocarcinoma. To identify clinically relevant genes and miRNAs involved in CRC, the differential expression profiles of mRNA and miRNA were analyzed for CRC and normal samples. The analysis contained 20,531 protein coding genes, of which 9,803 were found to be differentially expressed genes (DEG) in CRC samples. A total of 4,213 DEGs were shown to be upregulated and 5,590 were shown to be downregulated in CRC. In the case of miRNA analysis, we found 498 differentially expressed miRNAs (DE-miRNAs) in CRC samples compared to normal tissue, 180 of miRNA were upregulated and 318 were downregulated in CRC.

Target genes. The set of validated miRNA-target genes list obtained from miRTarBase contained 380,639 MTIs. These MTI pairs were obtained by combination of 2,599 miRNAs and 15,064 target genes. All of the validated 380,639 MTIs were combined with the set of analyzed DE-miRNA/DEG. The intersect between those sets of data revealed 101,461 MTIs in which both miRNA and mRNA were found to be significantly expressed in more than ten CRC samples (p-value < 0.01).

Correlations between miRNA and their target genes. The correlation coefficient (R) between DE-miRNAs and their target genes was calculated using the Pearson correlation test. We analyzed 101,461 MTIs, with the condition that the miRNA-target gene had to be expressed in at least ten samples. This analysis provided the correlation coefficient for 38,178 MTIs that were significant, with a p-value of less than 0.05 (Supplementary Table 1).

Using the "miRNA colorectal cancer" knowledge base. In this section, we will present examples of obtaining data from our "miRNA colorectal cancer" knowledge base through the web application. We will describe the usefulness of miRNA-mRNA expression and their interactions to understand their behavior of CRC better and to select appropriate candidate miRNAs or/and mRNAs for further experimental validation.

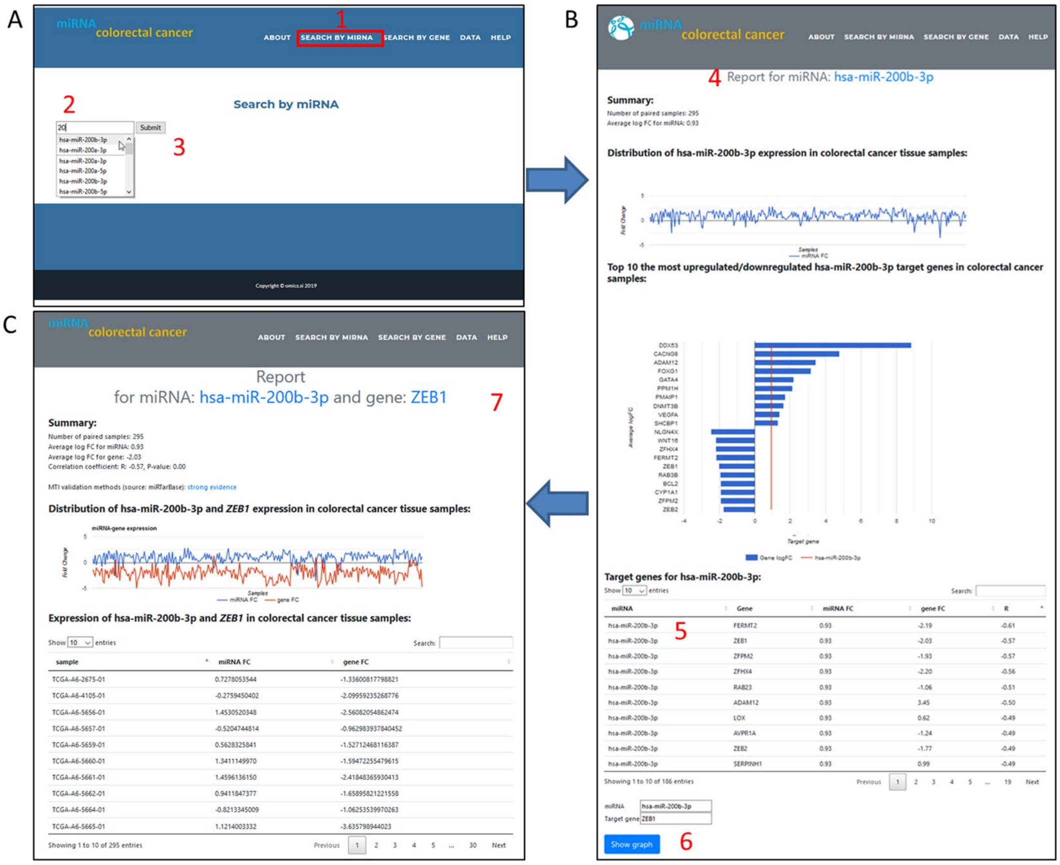
We can use two query interfaces to obtain data from the "miRNA colorectal cancer" web application: search by miRNA and search by gene. We present here search by miRNA option (1) (Fig. 2A). The user inserts an arbitrary miRNA (2) and presses the submit button (3). The result is opened in a new tab (4) and contains four parts (Fig. 2B). The first part of the report is a summary of the analyzed miRNA, which contains the number of paired samples and average logFC for this miRNA (it depends on user's search input). The second part of the report is a distribution of the miRNA expression in the analyzed CRC tissue samples. The third part of this report shows the average logFC values for the ten (if there are ten or more target genes for this miRNA) most upregulated and ten most downregulated target genes for the searched miRNA. The fourth part of the report is a table view with all known miRNA-mRNA interactions for the searched miRNA. The table report enables a table search, as well as sorting individual columns by arrows at the end of each column. By clicking on the specific row, the form on the bottom of the page is auto-filled with the miRNA and gene names (5). A selection is submitted using the "Show graph" button (6). A new tab shows a new report for the specific MTI (previously selected) (Fig. 2C). The report contains: a) a summary of the analyzed miRNA and gene, which contains the number of paired samples, average logFC for the miRNA and gene, correlation coefficient and calculated p-value for selected MTI and MTI validation methods obtained from miRTarBase. The second part of the final report for the selected MTI is the expression in CRC tissue by samples shown as b) the distribution on a chart and c) a tabular view.

Experimental validation. To validate our bioinformatics approach experimentally, we selected five miRNAs involved in EMT and their target genes based on literature and our previous and ongoing research in CRC[22,23]. We used our database to search for candidate MTIs which are among other functions involved in EMT (Table 1). For further experimental validation MTIs with p-value <0.01 were selected, including *FN1*/*hsa-miR-200b/c-3p*, *TGFB2*/*hsa-miR-200a-3p*, *TGFB2*/*hsa-miR-141-3p*, *RND3*/*hsa-miR-200b-3p*, *ZEB1*/*hsa-miR-200* family members and *ZEB2*/*hsa-miR-200* family members. For the selected MTIs, our bioinformatics and experimental results, including average logFC for miRNA and gene as well as correlation coefficient with its p-value are presented in Table 1 while the distributions of miRNA-target gene expressions are presented in Supplementary Figure 1.

As an example of MTIs involved in the regulation of cellular processes, there are several MTIs involved in EMT regulation. Genes *ZEB1* and *ZEB2*, the transcriptional factors of EMT, have a significant strong and/or moderate negative correlation with all five *hsa-miR-200* family miRNAs, which are the most studied EMT related miRNAs. In our *in silico* study, the correlation between *ZEB1* and *hsa-miR-200a/b/c-3p* was shown to be moderate (R = −0.57). Gene *ZEB1* (logFC = −1.33) is targeted by other two *hsa-miR-200* family members: *hsa-miR-141-3p* (logFC = 4.25) and *hsa-miR-429* (logFC = 5.03), with strong negative correlations. Another transcription factor of EMT, *ZEB2* (logFC = −1.77) has moderate negative correlations with *hsa-miR-200* family miRNAs, ranging from −0.45 to −0.54.

Gene *FN1* is a target gene for two *hsa-miR-200* family members (*hsa-miR-200b/c-3p*), with a moderate negative correlation (−0.49, −0.45, respectively), based on our bioinformatics study. Three other MTIs: *TGFB2*/*hsa-miR-200a-3p*, *TGFB2*/*hsa-miR-141-3p* and *RND3*/*hsa-miR-200b-3p* were selected with a statistically significant although weak negative correlation coefficient.

Our experimental data show the average logFC for gene *FN1* was −1.01 while the average logFC for *hsa-miR-200b-3p* and *hsa-miR-200c-3p* was 3.26 and 0.90, respectively. However, the correlation coefficient for MTI *FN1*/*hsa-miR-200b-3p* was very weak negative (R = −0.13) and for MTI *FN1*/*hsa-miR-200c-3p* we observed moderate negative correlation coefficient (R = −0.45). Gene *TGFB2* amplified in 15 clinical samples, with average logFC −0.94, while the average logFC for *hsa-miR-200a-3p* was −1.17 and for *hsa-miR-141-3p* it was 0.78. We observed a negative weak correlation for MTI *TGFB2*/*hsa-miR-200a-3p* (R = −0.19) and positive very weak correlation for MTI *TGFB2*/*hsa-miR-141-3p* (R = 0.10). Our experimental validation included also *ZEB1* and *ZEB2*



A

Search by miRNA

20

hsa-miR-200b-3p
hsa-miR-200a-3p
hsa-miR-200a-3p
hsa-miR-200a-5p
hsa-miR-200b-3p
hsa-miR-200b-5p

Copyright © omics.si 2019

B

Report for miRNA: hsa-miR-200b-3p

Summary:
Number of paired samples: 295
Average log FC for miRNA: 0.93

Distribution of hsa-miR-200b-3p expression in colorectal cancer tissue samples:

Top 10 the most upregulated/downregulated hsa-miR-200b-3p target genes in colorectal cancer samples:

Target genes for hsa-miR-200b-3p:

Show 10 entries Search:

miRNA	Gene	miRNA FC	gene FC	R
hsa-miR-200b-3p	FERMT2	0.93	-2.19	-0.61
hsa-miR-200b-3p	ZEB1	0.93	-2.03	-0.57
hsa-miR-200b-3p	ZFPM2	0.93	-1.93	-0.57
hsa-miR-200b-3p	ZFHX4	0.93	-2.20	-0.56
hsa-miR-200b-3p	RAB23	0.93	-1.06	-0.51
hsa-miR-200b-3p	ADAM12	0.93	3.45	-0.50
hsa-miR-200b-3p	LOX	0.93	0.62	-0.49
hsa-miR-200b-3p	AVPR1A	0.93	-1.24	-0.49
hsa-miR-200b-3p	ZEB2	0.93	-1.77	-0.49
hsa-miR-200b-3p	SERPINH1	0.93	0.99	-0.49

Showing 1 to 10 of 186 entries Previous 1 2 3 4 5 ... 19 Next

miRNA hsa-miR-200b-3p
Target gene ZEB1

C

Report
for miRNA: hsa-miR-200b-3p and gene: ZEB1

Summary:
Number of paired samples: 295
Average log FC for miRNA: 0.93
Average log FC for gene: -2.03
Correlation coefficient: R: -0.57, P-value: 0.00

MTI validation methods (source: miRTarBase): strong evidence

Distribution of hsa-miR-200b-3p and *ZEB1* expression in colorectal cancer tissue samples:

Expression of hsa-miR-200b-3p and *ZEB1* in colorectal cancer tissue samples:

Show 10 entries Search:

sample	miRNA FC	gene FC
TCGA-A6-2675-01	0.7278053544	-1.33600817798821
TCGA-A6-4105-01	-0.2759450402	-2.09959235268776
TCGA-A6-5656-01	1.4530520348	-2.56082054862474
TCGA-A6-5657-01	-0.5204744814	-0.962983937840452
TCGA-A6-5659-01	0.5628325841	-1.52712468116387
TCGA-A6-5660-01	1.3411149970	-1.59472255479615
TCGA-A6-5661-01	1.4596136150	-2.41848365930413
TCGA-A6-5662-01	0.9411847377	-1.65895821221558
TCGA-A6-5664-01	-0.8213345009	-1.06253539970263
TCGA-A6-5665-01	1.1214003332	-3.635798944023

Showing 1 to 10 of 295 entries Previous 1 2 3 4 5 ... 30 Next

Figure 2. Example of "miRNA colorectal cancer" use. (**A**) Search by miRNA tab (1) allows users to start typing (2) the name of miRNA of interest or select it from a dropdown list (3). (**B**) Submit button opens a new browser tab with report for selected miRNA (4) in this case for *hsa-miR-200b-3p*. The report contains: summary of number of paired samples and average log fold change for selected miRNA (*hsa-miR-200b-3p*), distribution of selected miRNA (*hsa-miR-200b-3p*) in colorectal cancer tissue samples, top 10 most upregulated/downregulated miRNA (*hsa-miR-200b-3p*) target genes in colorectal cancer samples and list of target genes for selected miRNA (*hsa-miR-200b-3p*) (5). There are several options in table with target genes for selected miRNA (5): user can sort the table with arrows placed at the end of each column or can perform a search in the search form. By clicking miRNA-target gene in table (5) the queries for miRNA and target gene below are auto filled and user can get more information about this MTI by clicking on "Show graph" button (6), which opens another tab (**C**), the report for selected MTI – in this case *hsa-miR-200b-3p* and *ZEB1* gene) (7). MTI report contains: summary of number of paired samples, average log FC for miRNA, average log FC for gene, correlation coefficient with its *p*-value and MTI validation method (obtained from miRTarBase). The report also contains distribution of selected MTI (*hsa-miR-200b-3p* and *ZEB1*) expression in colorectal cancer tissue samples and expression of selected MTI (*hsa-miR-200b-3p* and *ZEB1*) in colorectal cancer tissue samples.

of which *ZEB1* was under the limit of detection, while *ZEB2* was expressed with logFC of 0.91 and had a very weak positive correlation with *hsa-miR-200a-3p*, a weak positive correlation with *hsa-miR-200b-3p* and negative week/very weak correlations with *hsa-miR-200c-3p*, *hsa-miR-141-3p* and *hsa-miR-429*.

Discussion

miRNAs are important post-transcriptional regulators of gene expression during development and physiology of human tissues. In general, miRNAs have an inhibitory effect on their targets. A prediction of active miRNAs is based on enrichment of differentially downregulated target genes of miRNAs[24,25]. Some bioinformatics tools have integrated miRNA-mRNA expression and correlation coefficients into their databases[21] because the expression of RNA molecules is different in different tissues. Regarding CRC, we developed "miRNA colorectal cancer" web application, the largest knowledge base of the microRNA-target gene expressions and correlations in CRC, based on the largest available sample size from the same source of data. Therefore, the pre-calculated correlation coefficient between differentially expressed miRNAs and their targets is more reliable than that of other databases available.

Although our knowledge base is a standalone application for searching the information about miRNAs, mRNAs or MTIs, we attempted to test the application in experimental setting. There might be discrepancies between the application and experimental results, the most significant factor being the sample size, followed by the experimental method used and by the sample quality. We believe that with the same conditions for all

	Data from knowledge base				Experimental data		
MTI	Average logFC miRNA	Average logFC gene	R	*p*-value	Average logFC miRNA	Average logFC gene	R
FN1/hsa-miR-200b-3p	0.93	−0.19	−0.49	<0.01	3.26	−1.01	−0.13
FN1/hsa-miR-200c-3p	−1.83	−0.19	−0.45	<0.01	0.90	−1.01	−0.45
TGFB2/hsa-miR-200a-3p	2.83	0.14	−0.32	<0.01	−1.17	−0.94	−0.19
TGFB2/hsa-miR-141-3p	4.16	0.14	−0.35	<0.01	0.78	−0.94	0.10
ZEB2/hsa-miR-200a-3p	2.83	−1.77	−0.48	<0.01	−1.17	0.91	0.15
ZEB2/hsa-miR-200b-3p	0.93	−1.77	−0.49	<0.01	3.26	0.91	0.31
ZEB2/hsa-miR-200c-3p	−1.83	−1.77	−0.45	<0.01	0.90	0.91	−0.06
ZEB2/hsa-miR-141-3p	4.16	−1.77	−0.54	<0.01	0.78	0.91	−0.04
ZEB2/hsa-miR-429	4.71	−1.77	−0.54	<0.01	1.73	0.91	−0.28
RND3/hsa-miR200b-3p	0.93	−0.45	−0.26	<0.01	3.26	−0.23	0.25
ZEB1/hsa-miR-200a-3p	2.83	−2.03	−0.57	<0.01	*ZEB1* was under limit of detection.		
ZEB1/hsa-miR-200b-3p	0.93	−2.03	−0.57	<0.01			
ZEB1/hsa-miR-200c-3p	−1.83	−2.03	−0.57	<0.01			
ZEB1/hsa-miR-141-3p	4.16	−2.03	−0.65	<0.01			
ZEB1/hsa-miR-429	4.71	−2.03	−0.62	<0.01			
SOX2/hsa-miR-429	4.71	−1.96	−0.11	0.08	Not experimentally validated MTIs.		
CDKN1B/hsa-miR-200b-3p	0.93	0.11	−0.03	0.56			
CDKN1B/hsa-miR-200c-3p	−1.83	0.11	0.10	0.09			
CDKN1B/hsa-miR-429	4.71	0.11	−0.09	0.14			
ONECUT2/hsa-miR-429	4.71	1.80	0.06	0.34			
PTPN13/hsa-miR-200c-3p	−1.83	0.49	−0.12	0.03			

Table 1. Candidate miRNA-target gene pair data from knowledge base and data from experimental validation. logFC: logarithm of fold change, R: correlation coefficient.

samples, we could obtain the same results as in bioinformatics study if a large sample cohort was used. Moreover, when investigating miRNA and their target genes, the location of the sample used in experimental setting should be considered, for example different sections of tumor tissue could have different expression patterns.

For experimental validation, *miR-200* family and their target genes were used. miRNAs of the *miR-200* family are involved in several physiologic and pathologic processes, such as wound healing[26], endometrial development during early pregnancy[27], neurogenesis[28] and inflammation[29]. The most documented process where *hsa-miR-200* family is involved is EMT, which is present also in CRC.

Many previous studies have revealed an association between gene *ZEB1*, a transcription factor of EMT, and CRC[10,30–35]. In humans, gene *ZEB1* is a target gene for 46 miRNAs[21] and numerous studies indicate involvement of *ZEB1* in CRC. The complex formed by telomerase (hTERT) and ZEB1 binds directly to the E-cadherin promoter, inhibits E-cadherin expression and promotes EMT in CRC cells[32]. Gene *ZEB1* has numerous potential regulators beside miRNAs. MEF2D directly regulates transcription of *ZEB1* and facilitates histone acetylation at the *ZEB1* promoter in CRC cells[30,34]. Gene *ZEB1* not only promotes tumor cell dissemination but is also necessary for the tumor-initiating capacity of pancreatic and CRC cells[10]. In line with that, Quan *et al.* identified *ZEB1* as a target of *GRHL2* and suggested a reciprocal GRHL2-ZEB1 repressive relationship, providing a novel mechanism through which proliferation may be modulated in CRC cells[33]. Activation of *ZEB1* in CRC is also performed by overexpression of gene *SIX1*, which partly represses *hsa-miR-200* family expression[31].

The *miR-200* family comprises *hsa-miR-200a*, *−200b*, *−200c*, *−141* and *−429*, and has been shown to play a role in cancer initiation and metastasis[22]. The majority of studies on the *miR-200* family are based on the assumption that they induce EMT in various diseases[36–40]. *miR-200* family members maintain epithelial cell integrity by suppressing EMT through direct inhibition of mesenchymal transcription factors such as ZEB1, ZEB2 and TGF-β[41]. Accordingly, we observed a correlation between upregulated miRNAs *hsa-miR-200a-3p*, *−200b-3p*, *−141-3p* and *-429* (Table 1, Supplementary Table 1) and their downregulated target gene *ZEB1*. Moreover, *hsa-miR-429* expression was upregulated in human CRC tissues, and high *hsa-miR-429* expression was significantly associated with tumor size, lymph node metastasis and poor prognosis[42]. Since the discovery of *miR-200* family functions in cancers, more detailed knowledge of this miRNA family may have therapeutic implications for the treatment of metastatic and drug-resistant tumors[37].

However, the biological functions of intestinal miRNAs, not only the pathological aspect, are far more complex than initially suggested. The intestinal epithelium requires rapid changes in gene expression patterns to adapt to stress, and maintain epithelial homeostasis. miRNAs have emerged as fundamental regulators involved in many aspects of intestinal epithelial differentiation, architecture, and barrier function, as well as cellular self-renewal and inflammation, with specific miRNAs that promote intestinal homeostasis. Impact of miRNAs on the gut microbiota also plays an important role in intestinal homeostasis[43,44].

miRNAs are also involved in regulating immune response and influence the expression of immune-related genes in intestine. They play important role in modulating the balance of resolution of inflammation, preventing tissue damage[43]. The most studied miRNAs whose targets are related to immune pathways are *hsa-miR-155*,

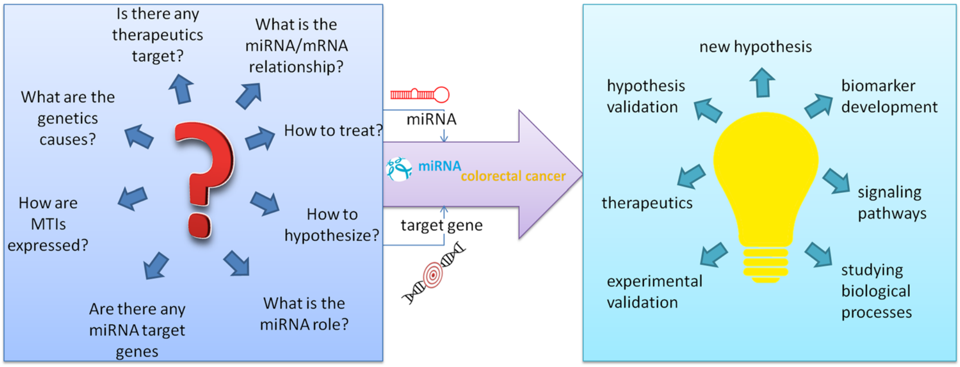


Figure 3. "miRNA colorectal cancer" as an integral part of bioinformatics process. How "miRNA colorectal cancer" can serve as a starting point in testing more targeted hypotheses and designing experiments using optimal miRNA, gene or both loci for further researches and analyses.

hsa-miR-146a, cluster *hsa-miRs-17∼92*, and *hsa-miR-181a*. Many of these miRNAs target genes in Jak/Stat NF-κB and Akt pathways, which are immune response pathways[45,46].

We believe that our knowledge base can assists researchers by providing step-by-step guidance on the formulation of a new research questions in the field of miRNA and mRNA expression profiles, their interactions and role in CRC (Fig. 3). Moreover, "miRNA colorectal cancer" knowledge base can also serve researchers as a starting point in testing more targeted hypotheses and designing experiments using optimal miRNA, gene or both loci for further researches and analyses.

Conclusion

We performed *in silico* screening for miRNA-target gene interactions in CRC and presented them through an integrative knowledge base. We experimentally validated five genes and five miRNAs. These results can be applied not only for the experimental validation of MTIs involved in CRC, but also to contribute to our understanding of the biological processes and cellular mechanisms of organogenesis and cell differentiation. Our finding of a negative correlation between miRNA *hsa-miR-200b/c-3p* and *hsa-miR-200a-3p* and their target genes *FN1* and *TGFB2*, respectively, by both bioinformatics and experimental approaches, strongly supports the usefulness of the developed knowledge base. Collectively, our approach is efficient for choosing the appropriate candidate miRNA target genes in CRC and may provide an opportunity to discover and validate novel genes and cellular mechanisms as potential prognosis and/or diagnosis markers and therapeutic targets in CRC cancerogenesis.

Methods

The "miRNA colorectal cancer" web application contains results of differential expression and correlation analyses of miRNA and mRNA data available from TCGA. The results are available through web forms within the newly developed tool.

Data pre-processing.

Data were obtained from the colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) project of TCGA from the Broad GDAC Firehose portal (https://gdac.broadinstitute.org/). We selected available RNAseqV2 and miRNAseq data, both performed on an Illumina HiSeq platform. A total of 295 samples that had both RNAseqV2 and miRNAseq data were selected. Additionally, 11 normal samples for the miRNAseq experiment and 51 normal samples for the RNAseqV2 experiment were selected. In both mRNA and miRNA experiments, level 3 data were used, which contain normalized gene counts and read per million mapped miRNA isoforms, respectively (Supplementary Table 2). For miRNAseq, we used data of calculated expression for each individual miRNA sequence isoform observed. For mRNAseq, we used RNASeqV2, which uses a combination of MapSplice and "scaled estimate" (RSEM) to determine expression levels. RNAseqV2 data contains a normalized read count, which represents the upper quartile normalized RSEM count estimates.

In tumor samples, values of normalized read counts were transformed to logarithmic values. In normal samples, normalized read counts were averaged through the samples and the averages were then transformed to logarithmic values. The logarithmic values of the average of normal samples were subtracted from each tumor sample logarithmic value, to obtain the logarithmic value of fold change for each tumor sample.

In miRNA level 3 data, we removed the stem-loop, precursor and unannotated reads, and summarized all reads belonging to a single isoform. The counts for each isoform in each sample were transformed to logarithmic values. For normal tissue samples, we also removed stem-loop, precursor and unannotated reads, summarized the reads for each isoform and averaged the isoforms before making the transformation to logarithmic values. The logarithmic values of the average of normal samples were then subtracted from logarithmic values of each tumor sample to obtain logarithmic values of fold change.

Probes and genes.

Target genes for miRNA were obtained from miRTarBase Release 7.0 (http://mirtarbase.mbc.nctu.edu.tw/php/index.php) (Sept. 13, 2019)[21], and overlapped with mRNAseq data. The set of miRNA-target gene pairs contained 2,599 miRNAs and 15,064 targets. Further analysis was applied to 380,639 MTIs, downloaded from the miRTarBase database. The coordinates of protein-coding genes were obtained from

Ensembl, release 98[47], while the mature miRNA coordinates were obtained from miRBase Release 21[48]. The nomenclature of genes was unified according to the HUGO Gene Nomenclature Committee (HGNC; http://www.genenames.org/)[49].

Statistical analysis. The analysis was performed in R/Bioconductor software packages (http://www.bioconductor.org). A linear model fit was used for determining the significantly differentially expressed miRNA and mRNA in cancer compared to normal tissue, constructed in the "limma" package separately for mRNA and miRNA[50]. The threshold for the adjusted p-value was set to 0.01.

Pearson correlations were calculated using the mRNA and miRNA logarithm of fold change (logFC). The function was constructed using the "sigr" package in R language[51]. The Pearson correlation test was used to estimate the correlations between miRNAs and their target genes, with the p-value set to 0.05.

Knowledge base development. The data from our study are stored in a relational database management system, MySQL (http://www.mysql.com). The data from the database is retrieved through the web interface and developed using HTML, CSS, JavaScript and Apache web server. Interactive charts were created by Google chart tools (https://developers.google.com/chart/) using a combination of programming languages PHP Hypertext Preprocessor (PHP) and MySQL.

Experimental validation. *Clinical samples.* Our validation study was comprised of 30 patients with CRC. Tumor samples were collected during surgical colectomy of patients diagnosed with primary CRC, as well as samples from macroscopically normal colon mucosa from the same patient (at least 20 cm away from the tumor). Tissue samples were formalin-fixed, paraffin-embedded (FFPE) and stained with hematoxylin and eosin (HE). All the samples were analyzed as paired tumor samples, CRC sample and corresponding normal mucosa.

RNA isolation. RNA was isolated with the AllPrep DNA/RNA FFPE Kit (Qiagen), according to the manufacturer's recommendations. RNA quantity and quality were determined spectrophotometrically by NanoDrop ND-1000 (Thermo Fisher Scientific).

Reverse transcription and qPCR experiment. Gene expression levels were determined using TaqMan based qPCR (Thermo Fisher Scientific), which was performed on the RotorGene system (Qiagen). Total RNA was reverse transcribed, 60 ng using the One Taq RT PCR Kit (NEB) according to the manufacturer's instructions for mRNA quantification and 10 ng using the TaqMan microRNA reverse transcription kit (Thermo Fisher Scientific) according to the manufacturer's instructions for miRNA quantification. For mRNAs quantification, the resulting cDNA was pre-amplified using a TaqMan PreAmp Master Mix Kit (Thermo Fisher Scientific). Relative levels of miRNAs and mRNAs were determined from diluted cDNA and pre-amplified cDNA, respectively, using qPCR and appropriate TaqMan probes. The expression levels were calculated after efficiency correction (based on the results of serial dilution) relative to four endogenous controls *hsa-miR-1247b* and *RNU6B* for miRNAs and *IPO8* and *B2M* for mRNAs.

Ethics approval and consent to participate. The responsible ethics committee (Slovenian National Medical Ethics Committee) approved the reported investigation, which was carried out in accordance with the principles of the Helsinki Declaration.

Data availability

The knowledge base is freely available through a web interface at https://mirna-coadread.omics.si.

Received: 8 March 2019; Accepted: 13 November 2019;
Published online: 02 December 2019

References

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. McGuire, S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv. Nutr.* **7**, 418–419 (2016).
3. Strubberg, A. M. & Madison, B. B. MicroRNAs in the etiology of colorectal cancer: pathways and clinical implications. *Dis. Model. Mech.* **10**, 197–214 (2017).
4. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
5. Bu, P. *et al.* A microRNA miR-34a-regulated bimodal switch targets Notch in colon cancer stem cells. *Cell Stem Cell* **12**, 602–615 (2013).
6. Bu, P. *et al.* A miR-34a-numb feedforward loop triggered by inflammation regulates asymmetric stem cell division in intestine and colon cancer. *Cell Stem Cell* **18**, 189–202 (2016).
7. Zheng, L. *et al.* Down-regualtion of miR-106b induces epithelial-mesenchymal transition but suppresses metastatic colonization by targeting Prrx1 in colorectal cancer. *Int. J. Clin. Exp. Pathol.* **8**, 10534–10544 (2015).
8. Zhai, H., Fesler, A., Ba, Y., Wu, S. & Ju, J. Inhibition of colorectal cancer stem cell survival and invasive potential by hsa-miR-140-5p mediated suppression of Smad2 and autophagy. *Oncotarget* **6**, 19735–19746 (2015).
9. Hwang, W. L. *et al.* MicroRNA-146a directs the symmetric division of Snail-dominant colorectal cancer stem cells. *Nat. Cell Biol.* **16**, 268–280 (2014).
10. Wellner, U. *et al.* The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* **11**, 1487–1495 (2009).
11. Jones, M. F. *et al.* The CDX1-microRNA-215 axis regulates colorectal cancer stem cell differentiation. *Proc. Natl. Acad. Sci. USA* **112**, E1550–1558 (2015).
12. Zhu, R. *et al.* Ascl2 knockdown results in tumor growth arrest by miRNA-302b-related inhibition of colon cancer progenitor cells. *PLoS ONE* **7**, e32170 (2012).

13. Xu, X. T. *et al*. MicroRNA expression profiling identifies miR-328 regulates cancer stem cell-like SP cells in colorectal cancer. *Br. J. Cancer* **106**, 1320–1330 (2012).
14. Tsuji, S. *et al*. The miR-363-GATA6-Lgr5 pathway is critical for colorectal tumourigenesis. *Nat. Commun.* **5**, 3150 (2014).
15. Li, Y. *et al*. The SOX17/miR-371-5p/SOX2 axis inhibits EMT, stem cell properties and metastasis in colorectal cancer. *Oncotarget* **6**, 9099–9112 (2015).
16. Bitarte, N. *et al*. MicroRNA-451 is involved in the self-renewal, tumorigenicity, and chemoresistance of colorectal cancer stem cells. *Stem Cells* **29**, 1661–1671 (2011).
17. Garzon, R., Calin, G. A. & Croce, C. M. MicroRNAs in cancer. *Annu. Rev. Med.* **60**, 167–179 (2009).
18. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* **6**, 259–269 (2006).
19. Meng, F. *et al*. Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines. *Gastroenterology* **130**, 2113–2129 (2006).
20. Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. microRNAs as oncogenes and tumor suppressors. *Dev. Biol.* **302**, 1–12 (2007).
21. Chou, C. H. *et al*. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302 (2018).
22. Humphries, B. & Yang, C. The microRNA-200 family: small molecules with novel roles in cancer development, progression and therapy. *Oncotarget* **6**, 6472–6498 (2015).
23. Zidar, N. *et al*. Down-regulation of microRNAs of the miR-200 family and up-regulation of Snail and Slug in inflammatory bowel diseases - hallmark of epithelial-mesenchymal transition. *J. Cell. Mol. Med.* **20**, 1813–1820 (2016).
24. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
25. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
26. Aunin, E., Broadley, D., Ahmed, M. I., Mardaryev, A. N. & Botchkareva, N. V. Exploring a role for regulatory miRNAs in wound healing during ageing: involvement of miR-200c in wound repair. *Sci. Rep.* **7**, 3257 (2017).
27. Jimenez, P. T., Mainigi, M. A., Word, R. A., Kraus, W. L. & Mendelson, C. R. miR-200 regulates endometrial development during early pregnancy. *Mol. Endocrinol.* **30**, 977–987 (2016).
28. Trumbach, D. & Prakash, N. The conserved miR-8/miR-200 microRNA family and their role in invertebrate and vertebrate neurogenesis. *Cell Tissue Res.* **359**, 161–177 (2015).
29. Huang, Z. *et al*. miR-141 Regulates colonic leukocytic trafficking by targeting CXCL12beta during murine colitis and human Crohn's disease. *Gut* **63**, 1247–1257 (2014).
30. Jagle, S., Dertmann, A., Schrempp, M. & Hecht, A. ZEB1 is neither sufficient nor required for epithelial-mesenchymal transition in LS174T colorectal cancer cells. *Biochem. Biophys. Res. Commun.* **482**, 1226–1232 (2017).
31. Ono, H. *et al*. SIX1 promotes epithelial-mesenchymal transition in colorectal cancer through ZEB1 activation. *Oncogene* **31**, 4923–4934 (2012).
32. Qin, Y. *et al*. An hTERT/ZEB1 complex directly regulates E-cadherin to promote epithelial-to-mesenchymal transition (EMT) in colorectal cancer. *Oncotarget* **7**, 351–361 (2016).
33. Quan, Y. *et al*. Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol. Ther.* **15**, 878–887 (2014).
34. Su, L. *et al*. MEF2D transduces microenvironment stimuli to ZEB1 to promote epithelial-mesenchymal transition and metastasis in colorectal cancer. *Cancer Res.* **76**, 5054–5067 (2016).
35. Gong, H. *et al*. High expression of long non-coding RNA ZEB1-AS1 promotes colorectal cancer cell proliferation partially by suppressing p15 expression. *Tumour Biol.* **39**, 1010428317705336 (2017).
36. Korpal, M. & Kang, Y. The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. *RNA Biol.* **5**, 115–119 (2008).
37. Mongroo, P. S. & Rustgi, A. K. The role of the miR-200 family in epithelial-mesenchymal transition. *Cancer Biol. Ther.* **10**, 219–222 (2010).
38. Gregory, P. A. *et al*. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10**, 593–601 (2008).
39. Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* **22**, 894–907 (2008).
40. Jing, J. *et al*. Sex-biased miRNAs in gonad and their potential roles for testis development in yellow catfish. *PLoS ONE* **9**, e107946 (2014).
41. Choi, P. W. & Ng, S. W. The functions of microRNA-200 family in ovarian cancer: beyond epithelial-mesenchymal transition. Int. J. Mol. Sci., 18 (2017).
42. Li, J. *et al*. MiR-429 is an independent prognostic factor in colorectal cancer and exerts its anti-apoptotic function by targeting SOX2. *Cancer Lett.* **329**, 84–90 (2013).
43. Runtsch, M. C., Round, J. L. & O'Connell, R. M. MicroRNAs and the regulation of intestinal homeostasis. *Front Genet* **5**, 347 (2014).
44. Zou, L., Xiong, X., Wang, K. & Yin, Y. MicroRNAs in the intestine: role in renewal, homeostasis, and inflammation. *Curr. Mol. Med.* **18**, 190–198 (2018).
45. Baltimore, D., Boldin, M. P., O'Connell, R. M., Rao, D. S. & Taganov, K. D. MicroRNAs: new regulators of immune cell development and function. *Nat. Immunol.* **9**, 839–845 (2008).
46. O'Connell, R. M., Rao, D. S. & Baltimore, D. microRNA regulation of inflammatory responses. *Annu. Rev. Immunol.* **30**, 295–312 (2012).
47. Aken, B. L. *et al*. Ensembl 2017. *Nucleic Acids Res.* **45**, D635–D642 (2017).
48. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–73 (2014).
49. Yates, B. *et al*. Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.* **45**, D619–D625 (2017).
50. Ritchie, M. E. *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
51. Mount, J. & Zumel, N. sigr: succinct and correct statistical summaries for reports (2017).

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency through research core funding No. P3-0054. NH acknowledges the financial support from the Slovenian Research Agency through project Z3-6797.

Author contributions

D.J.S. and N.H. designed the study, participated in bioinformatics analysis, performed database, developed web page and writing. E.B. and N.Z. participated in the experimental validation, study concept and writing and revised the manuscript critically. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-54358-w.

Correspondence and requests for materials should be addressed to D.J.S.

Reprints and permissions information is available at www.nature.com/reprints.

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