

LETTER TO THE EDITOR

Open Access



Large-scale discovery of previously undetected microRNAs specific to human liver

Brenda C. Minatel^{1*†}, Victor D. Martinez^{1†}, Kevin W. Ng¹, Adam P. Sage¹, Tomas Tokar², Erin A. Marshall¹, Christine Anderson¹, Katey S. S. Enfield¹, Greg L. Stewart¹, Patricia P. Reis³, Igor Jurisica² and Wan L. Lam¹

Abstract

MicroRNAs (miRNAs) are crucial regulators of gene expression in normal development and cellular homeostasis. While miRNA repositories contain thousands of unique sequences, they primarily contain molecules that are conserved across several tissues, largely excluding lineage and tissue-specific miRNAs. By analyzing small non-coding RNA sequencing data for abundance and secondary RNA structure, we discovered 103 miRNA candidates previously undescribed in liver tissue. While expression of some of these unannotated sequences is restricted to non-malignant tissue, downregulation of most of the sequences was detected in liver tumors, indicating their importance in the maintenance of liver homeostasis. Furthermore, target prediction revealed the involvement of the unannotated miRNA candidates in fatty-acid metabolism and tissue regeneration, which are key pathways in liver biology. Here, we provide a comprehensive analysis of the undiscovered liver miRNA transcriptome, providing new resources for a deeper exploration of organ-specific biology and disease.

Keywords: Liver, Non-coding RNA, Novel miRNA, Tissue specificity, Liver cancer

MicroRNAs (miRNAs) are known to promote post-transcriptional fine-tuning of gene expression through complementary binding to target mRNA sequences [1]. Their wide-reaching effects are attributed to the fact that a single miRNA can target dozens to hundreds of genes, often affecting multiple nodes of a given signaling pathway [1]. In the liver, miRNAs are believed to orchestrate cell lineage differentiation during organ development, the modulation of homeostatic liver functions such as cholesterol and lipid metabolism, and disease [2, 3]. Clinically, miRNAs hold prognostic and therapeutic value both as biomarkers and therapeutic targets. For example, Miravirsen is a miR-122 antagonist emerging as a promising treatment for hepatitis C infection, which has progressed through Phase 2a clinical trials [4].

Initial attempts to characterize the human miRNA transcriptome were mostly limited to the discovery of

abundant miRNA sequences and/or sequences that are conserved across several tissue types. This restriction may preclude miRNA transcripts with expression patterns that are more specialized to individual tissues or cell lineages [5, 6]. Indeed, recent genome-wide studies using next-generation sequencing have suggested the existence of human-specific previously undetected miRNAs, and they have been shown to exhibit high tissue specificity [5–8]. Therefore, the discovery of such miRNA sequences may uncover novel tissue-specific regulatory mechanisms relevant to developmental biology and disease pathology. In this study, we performed a large-scale discovery of miRNA candidates previously undescribed in liver tissue and showed that these sequences exhibit tissue-specific expression patterns, as well as involvement in liver biology and disease.

Non-malignant liver small RNA sequence data was obtained from The Cancer Genome Atlas (TCGA; $n = 47$). Previously unannotated miRNA sequence discovery was performed using the miRDeep2 algorithm, which scans the transcriptome for novel miRNA candidates and compares them with known miRNA sequences

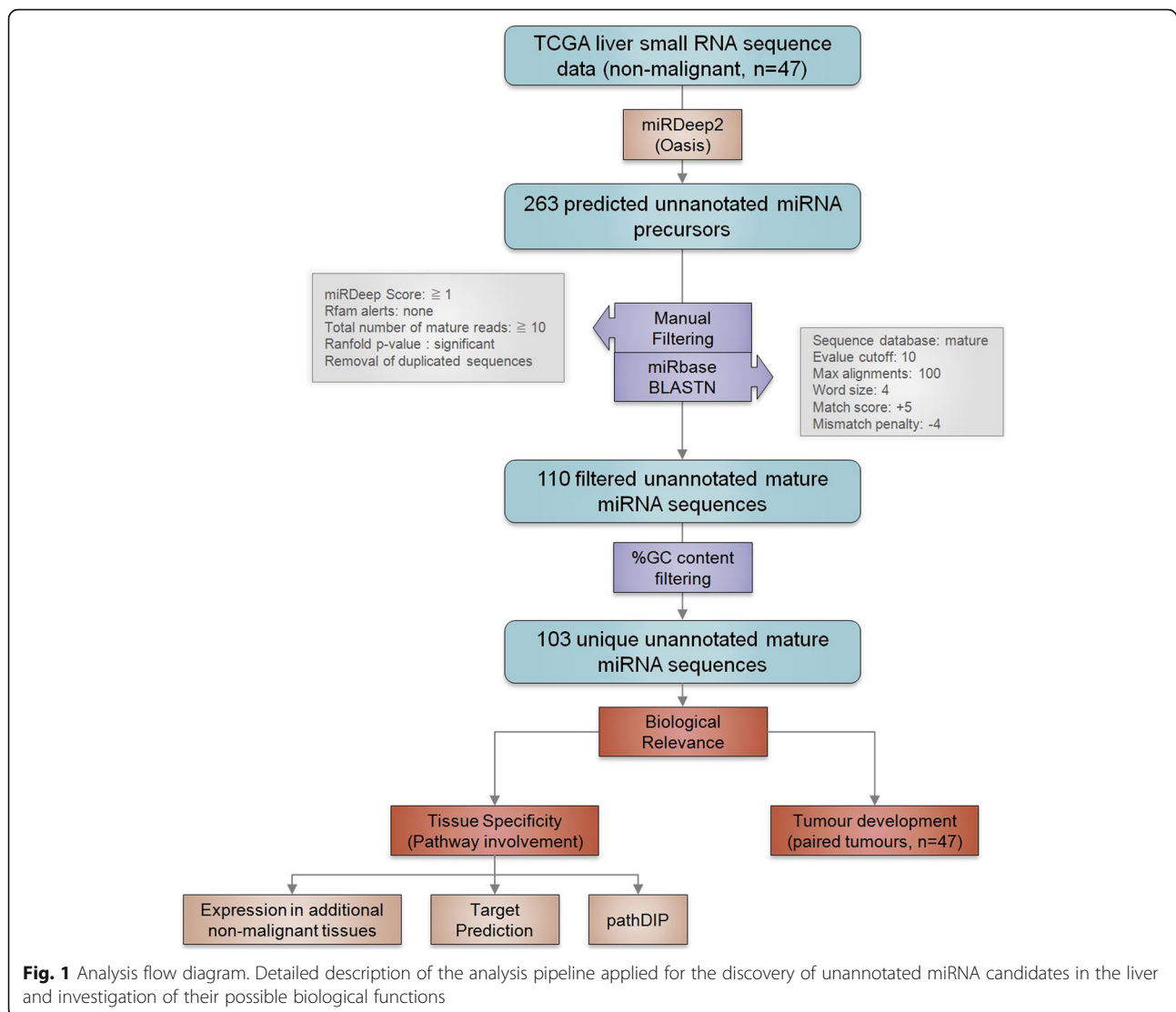
* Correspondence: bminatel@bccrc.ca

†Equal contributors

¹Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, BC, Canada

Full list of author information is available at the end of the article

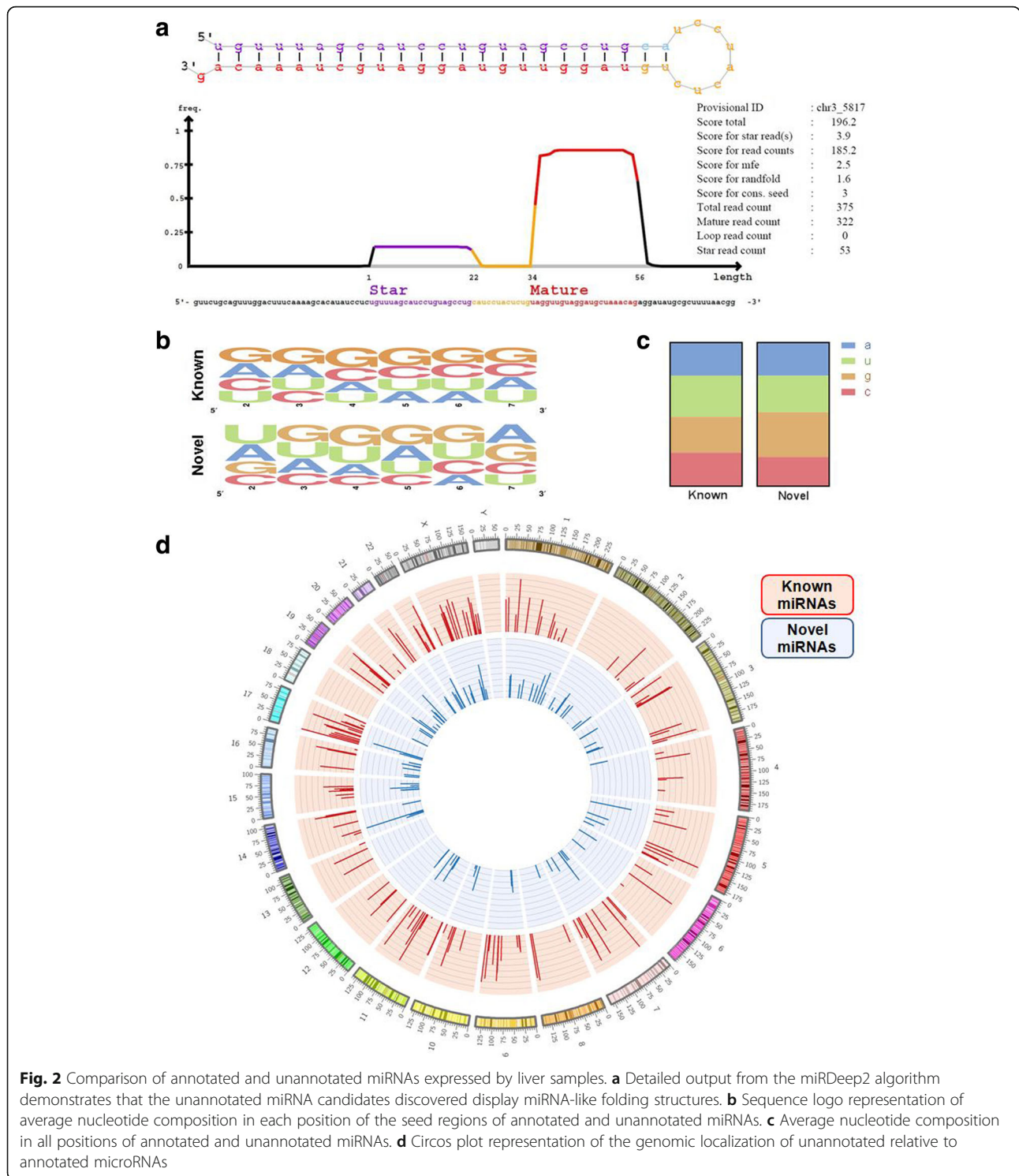




available in public databases, such as miRBase [9]. This established miRNA detection algorithm uses a statistical model to measure the likelihood of a detected small RNA sequence to be a putative novel miRNA. Primarily, this model assesses the hairpin structure of the predicted miRNA precursor and recognizes whether the precursor gives rise to the three products of miRNA processing by DICER, namely (i) mature miRNA, (ii) star sequence, and (iii) hairpin loop [9]. The likelihood of a detected small RNA sequence to be a true positive hit is reflected in the miRDeep2 score [9]. However, the selection of true positives based solely on the provided miRDeep2 score may still yield a large amount of false positive candidates [7]. To overcome these limitations, we applied several additional filtering steps to reduce the rate of false positives.

The initial miRDeep2 analysis discovered 263 unannotated miRNA candidate sequences. First, this output was

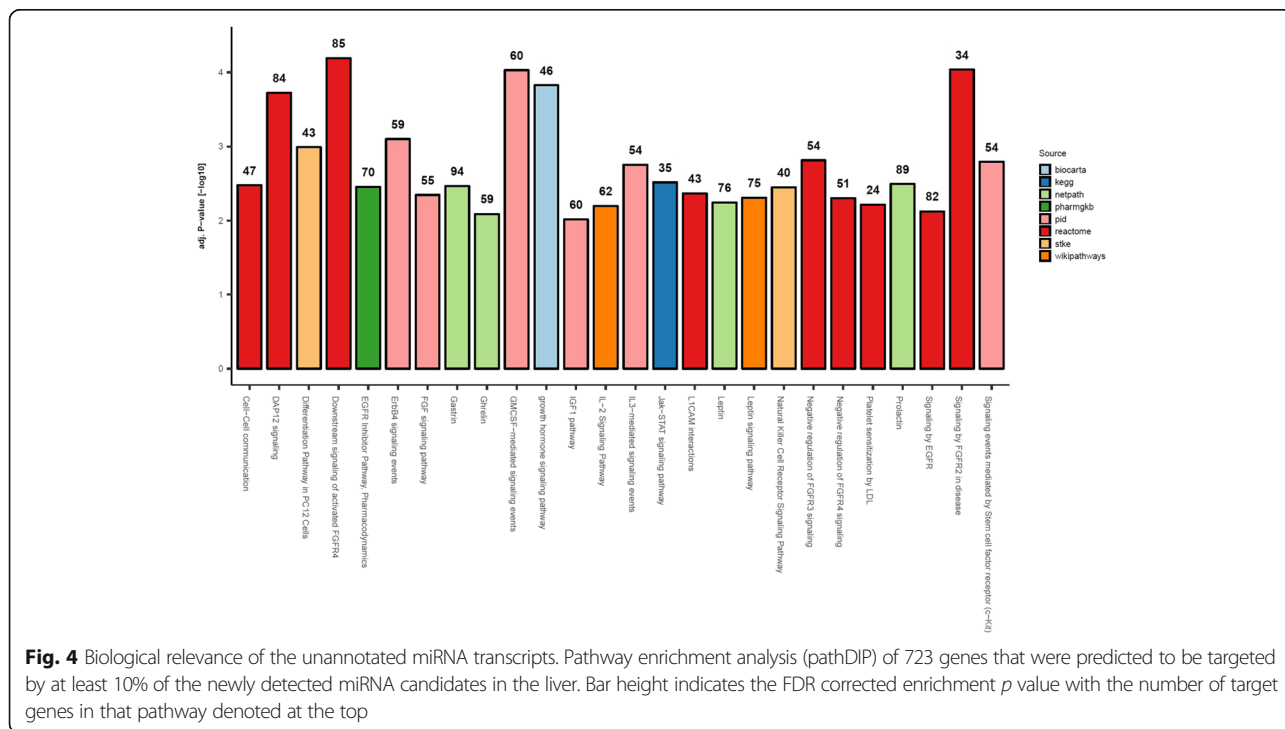
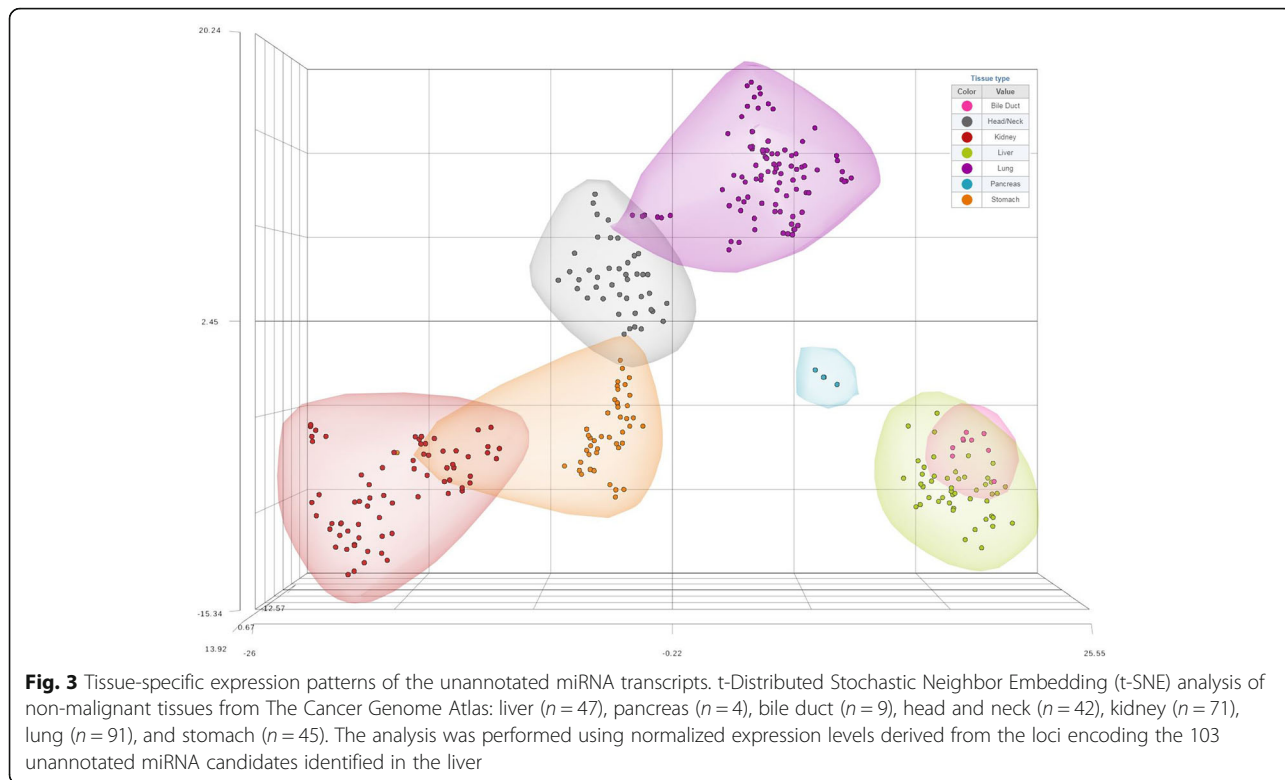
filtered by the number of reads corresponding to the mature sequence (≥ 10), a significant ($p \leq 0.05$) probability of a hairpin-like secondary structure, sequence similarity with annotated miRNAs in the miRBase repository, and a miRDeep2 score ≥ 1 , yielding a set of 110 candidate unannotated miRNA sequences (Fig. 1). We further assessed the similarity of these newly detected miRNAs with annotated miRNAs using the novoMiRank tool [7], which provides z-scores to each sequence based on 24 different features. Briefly, higher z-score numbers indicate less similarity to known miRNAs. Thus, while reads of these sequences may still be detected, miRNAs assigned a z-score ≥ 1 have an increased probability of representing false-positive candidates (Additional file 1: Table S1). Finally, we removed any predicted miRNA sequence with a GC-content ± 2 STD from the mean of currently annotated sequences (Additional file 2: Figure S1). Collectively, our filtering criteria resulted in the identification of 103



unique unannotated miRNA candidates, representing a substantial increase in the total number of miRNAs expressed in human liver (Fig. 1 and Additional file 1: Table S1). Additionally, these miRNA candidates were found to have similar sequence composition, folding structures and genomic distribution relative to annotated

miRNAs, further supporting their identity as true positive miRNA sequences (Fig. 2).

Next, to determine the tissue specificity of these miRNA transcripts, the expression of the 103 previously unannotated miRNA candidates was queried in small RNA sequencing data derived from organ sites representing



distinct anatomical regions and that differ in germ layer derivation (endoderm or mesoderm). The tissues investigated were the pancreas ($n = 4$), bile duct ($n = 9$), head and neck ($n = 42$), stomach ($n = 45$), kidney ($n = 71$), and lung ($n = 91$). We performed non-linear t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction on the normalized expression levels of the 103 unannotated miRNA transcripts against the aforementioned tissues. The expression pattern of these miRNA sequences was similar in both the liver and bile duct, corroborating their shared developmental lineage. In contrast, their expression in the liver is clearly distinct from the head and neck, stomach, kidney, and lung samples (Fig. 3), suggesting that our unannotated miRNA candidates have a unique pattern of expression that relies on cell lineage and that they may be relevant to liver-specific biology.

To identify the pathways regulated by the unannotated miRNAs, we analyzed their predicted targets. We restricted our analysis to protein-coding genes that were identified as targets by at least two of the three algorithms used and were predicted to be targeted by at least 10% of our novel miRNA sequences (Additional file 3: Figure S2). From this, we identified a total of 723 protein-coding gene targets of the newly detected miRNA candidates in the liver.

Strikingly, subsequent pathway enrichment analysis revealed that the 723 predicted targets are enriched ($p < 0.001$) in pathways that are important to normal and diseased liver biology (Fig. 4). These pathways include the following: fibroblast growth factor receptor (FGFR) signaling pathways, epidermal growth factor receptor (EGFR) signaling pathway, DNAX-activating protein of 12 kDa (DAP12) signaling, and granulocyte-macrophages colony-stimulating factor (GM-CSF) mediated signaling. In the liver, the FGFR pathway has been shown to modulate cholesterol and fatty acid metabolism and has been associated with chronic liver diseases and hepatocellular carcinoma (HCC) [10]. Likewise, the EGFR pathway plays a role in liver regeneration and is also associated with HCC aggressiveness through the activation of cells that secrete extracellular matrix components [11]. Lastly, the DAP12 and GM-CSF pathways participate in immune regulation and inflammatory response by modulating the maturation of hepatic dendritic cells and the formation of inflammatory granulomas, respectively [12, 13]. As these newly detected miRNA sequences are predicted to target key pathways in liver biology and disease, their discovery may be a cornerstone for identifying new regulatory mechanisms that may be disrupted in liver pathologies.

In order to further assess the biological relevance of the unannotated miRNA candidates, we sought to evaluate whether these sequences are deregulated in corresponding tumor samples. We compared the expression of the miRNAs between matched non-malignant and tumor tissues.

Strikingly, 83 of the 103 miRNA sequences had lost ($n = 65$) or reduced ($n = 18$, Wilcoxon signed-rank test corrected p value < 0.05) expression in tumor samples (Additional file 4: Figure S3). Thus, the widespread decrease in expression of these unannotated miRNA sequences may contribute to liver tumorigenesis.

In conclusion, we have discovered 103 previously undetected miRNA candidates in the liver. Although further experimental validation is required to confirm these sequences, our results shed light into the existence of unexplored regulatory molecules in liver tissue. Most importantly, these unannotated miRNAs have not only a lineage-specific expression pattern but may also be regulators of key liver processes, including those relevant to pathogenesis. Collectively, our results have substantial implications for liver-specific miRNA biology, emphasizing the need to further explore the undescribed areas of the human transcriptome.

Additional files

Additional file 1: Table S1. Output from miRDeep2 algorithm and novoMiRank scores for the 103 unique unannotated miRNA candidates. (XLS 59 kb)

Additional file 2: Figure S1. Percent GC content of unannotated and annotated miRNAs. Histogram plot of the percent GC content of the 110 filtered unannotated miRNAs predicted from miRDeep2 and all annotated miRNAs from miRBase v21. Dashed red lines indicate the two standard deviation thresholds from the mean of annotated miRNAs and were used as a filtering criteria. (JPEG 308 kb)

Additional file 3: Figure S2. Predicted targets and their overlaps across applied algorithms. A) Resulting number of predicted mRNA targets and their overlaps across the three different algorithms applied during target prediction analysis. B) Total number of predicted mRNA targets per unannotated miRNA. (JPEG 253 kb)

Additional file 4: Figure S3. Expression of the 38 unannotated miRNA transcripts in tumors. The expression of the 103 unannotated miRNAs was evaluated in a cohort of 47 liver tumor samples derived from the same patients in which the original miRNA prediction was performed. The expression of 38 miRNAs (39.1% of all the 103 miRNAs discovered) was detected in these tumor samples. (JPEG 167 kb)

Additional file 5: Detailed materials and methods. (DOC 109 kb)

Abbreviations

DAP12: DNAX-activating protein of 12 kDa; EGFR: Epidermal growth factor receptor; FGFR: Fibroblast growth factor receptor; GM-CSF: Granulocyte-macrophages colony-stimulating factor; HCC: Hepatocellular carcinoma; miRNAs: MicroRNAs; TCGA: The Cancer Genome Atlas; t-SNE: t-Distributed Stochastic Neighbor Embedding

Acknowledgements

The authors thank Heather Sapruff and Kim Lonergan for their expert critique of the manuscript.

Funding

This work was supported by grants from the Canadian Institutes for Health Research (CIHR FRN-143345). BCM, VDM, APS, EAM, and CA are supported by scholarships from the University of British Columbia. EAM and APS are also supported by scholarships from CIHR. IJ is in part supported by Canada Research Chair Program (#225404), Ontario Research Fund (GL2-01-030) and Canada Foundation for Innovation (CFI #225404, #30865).

Availability of data and materials

The datasets analyzed during the current study are available in The Cancer Genome Atlas (TCGA) repository, accessed at <https://gdc.cancer.gov/>. Detailed description of the materials and methods is available in Additional file 5.

Authors' contributions

BCM and VDM designed, performed data analysis, and prepared the manuscript. KWN, APS, and TT contributed to the data analysis and interpretation, as well as manuscript preparation. EAM, CA, KSSE, GLS, PPR, IJ, and WLL contributed to the data interpretation and manuscript preparation. All authors approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, BC, Canada. ²Kremsil Research Institute, University Health Network, Toronto, ON, Canada. ³Faculty of Medicine, São Paulo State University (UNESP), Botucatu, SP, Brazil.

Received: 17 January 2018 Accepted: 19 March 2018

Published online: 27 March 2018

References

- Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat Rev Drug Discov.* 2017;16(3):203–22.
- Wang XW, Heegaard NH, Orum H. MicroRNAs in liver disease. *Gastroenterology.* 2012;142(7):1431–43.
- Ma S, Tang KH, Chan YP, Lee TK, Kwan PS, Castilho A, Ng I, Man K, Wong N, To KF, et al. miR-130b promotes CD133(+) liver tumor-initiating cell growth and self-renewal via tumor protein 53-induced nuclear protein 1. *Cell Stem Cell.* 2010;7(6):694–707.
- Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, van der Meer AJ, Patack AK, Chen A, Zhou Y, et al. Treatment of HCV infection by targeting microRNA. *N Engl J Med.* 2013;368(18):1685–94.
- Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, Hatzimichael E, Kirino Y, Honda S, Lally M, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc Natl Acad Sci U S A.* 2015;112(10):E1106–15.
- McCall MN, Kim MS, Adil M, Patil AH, Lu Y, Mitchell CJ, Leal-Rojas P, Xu J, Kumar M, Dawson VL, et al. Toward the human cellular microRNAome. *Genome Res.* 2017;27(10):1769–81.
- Backes C, Meder B, Hart M, Ludwig N, Leidinger P, Vogel B, Galata V, Roth P, Menegatti J, Grasser F, et al. Prioritizing and selecting likely novel miRNAs from NGS data. *Nucleic Acids Res.* 2016;44(6):e53.
- Wake C, Labadorf A, Dumitriu A, Hoss AG, Bregu J, Albrecht KH, DeStefano AL, Myers RH. Novel microRNA discovery using small RNA sequencing in post-mortem human brain. *BMC Genomics.* 2016;17(1):776.
- Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 2012;40(1):37–52.
- Zheng N, Wei W, Wang Z. Emerging roles of FGF signaling in hepatocellular carcinoma. *Transl Cancer Res.* 2016;5(1):1–6.
- Berasain C, Avila MA. The EGFR signalling system in the liver: from hepatoprotection to hepatocarcinogenesis. *J Gastroenterol.* 2014;49(1):9–23.
- Sumpter TL, Packiam V, Turnquist HR, Castellana A, Yoshida O, Thomson AW. DAP12 promotes IRAK-M expression and IL-10 production by liver myeloid dendritic cells and restrains their T cell allostimulatory ability. *J Immunol.* 2011;186(4):1970–80.
- Wynn AA, Miyakawa K, Miyata E, Dranoff G, Takeya M, Takahashi K. Role of granulocyte/macrophage colony-stimulating factor in zymocel-induced hepatic granuloma formation. *Am J Pathol.* 2001;158(1):131–45.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

