bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OMCD: OncoMir Cancer Database

Aaron Sarver*[1,4], Anne Sarver[2], Ce Yuan[2,3], Subbaya Subramanian*[2,4]

[1]Institute of Health Informatics, University of Minnesota, Minneapolis, MN

[2]Department of Surgery, University of Minnesota Medical School, Minneapolis, MN

[3]Bioinformatics and Computational Biology, University of Minnesota, Minneapolis, MN

[4]Masonic Cancer Center, University of Minnesota, Minneapolis, MN

***Correspondence**

Subbaya Subramanian, MS, PhD
Department of Surgery
University of Minnesota Medical School
11-212 Moos Tower
Mayo Mail Code 195
420 Delaware Street SE
Minneapolis, MN 55455
Tel: 612-626-4330
Email: subree@umn.edu

Aaron Sarver, PhD
Institute of Health Informatics
University of Minnesota Medical School
Email: sarver@umn.edu

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

microRNAs (miRNAs) are small non-coding RNAs crucially important in cancer. Dysregulation of miRNAs is commonly observed in cancers and it largely cancer dependent. It is thus important to study pan-cancer miRNA expression in order to develop accurate and sensitive biomarkers. Here, we developed the OMCD - OncoMir Cancer Database, a web server that allows easy and systematic comparative genomic analyses of miRNAs sequencing data derived from over 9,500 cancer patients with associated clinical information and organ-specific controls present in the Cancer Genome Atlas (TCGA) database. The OMCD webserver provides 1). simple visualization of TCGA miRNA sequencing data, 2). statistical analysis of differentially expressed miRNAs for each cancer type and 3). allows for exploring miRNA clusters across cancer types. The OMCD web server is freely available to all users and can be accessed at www.oncomir.umn.edu/omcd/.

Database URL: www.oncomir.umn.edu/omcd/

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Background

microRNAs are small non-coding RNAs that regulate gene expressions through post-transcriptional modifications, by binding to the 3' UTR region of the target messenger RNAs (1). Dysregulation of miRNAs has been associated with cancers, such as colorectal cancer, lung cancer, lymphoma, glioblastoma, and osteosarcoma (2). The dysregulation of miRNAs in cancer makes them candidate biomarkers for diagnosis, classification, prognosis and potential therapeutic targets (2). Currently, using miRNAs as diagnostic and classification biomarkers has been approved by the United States Food and Drug Administration (FDA) for cancers of the lung, thyroid, and kidney, as well as for identifying cancer origin. However, because miRNA dysregulation is largely cancer dependent, it is important to study pan-cancer miRNA expression in order to develop accurate and sensitive biomarkers.

The Cancer Genome Atlas (TCGA) contains miRNA expression data for over 10,000 patients from 33 different types of cancer (3). There are currently two major web-based repositories of analyzed TCGA data - cBioPortal and GDAC firebrowse (4). However, both platforms mainly focus on the analysis and visualization of genomic and mRNA data and neither provides in-depth analysis and comparative visualization of miRNAs data. One database, ONCOMIR, provides analysis of TCGA miRNA data and calculates miRNA markers as survival signatures (5). However, this database lacks simple visualization of the TCGA miRNA expression data and the ability to explore miRNA clusters. Here we developed OncoMir Cancer Database (OMCD) that provides 1). simple visualization of TCGA miRNA data, 2). statistical analysis of differentially expressed miRNAs for each cancer type and 3). allows for exploring miRNA clusters across cancer types.

Construction and content

Method

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The OMCD repository was created using the LAMP software bundle (Linux, Apache2, MySQL 5.0 and PHP) and HTML, as described previously (6). Briefly, the application accessible to researchers across the globe. An Apache web server is used to host the web application. PHP is used to generate the user interface and used to communicate with the MySQL database at the backend. Normalized expression data, statistical results, and annotation data are stored within the database. A user-friendly graphical interface is provided to assist the users for ease of data retrieval and for selection of different criteria for data analysis. PHP is used to generate the HTML content through a database-driven architecture that was designed for incorporation of additional information.

Data Analyses

The miRNA expression data of 9,656 cases (8,993 Tumor and 663 Control samples) spanning 33 cancer types were downloaded from TCGA data repository (https://gdc.nci.nih.gov; **Table 1**). Differential expression was analyzed using a two-group t-test to determine miRNAs that were differentially expressed between 1) normal and tumor tissue for a given tumor type, 2) normal tissue compared to all other normal tissues and 3) tumor tissue compared to all other tumor tissues based on the availability of samples. It is important to mention that each of the experiments has a different statistical power available and that the absence of a miRNA from a specific dataset may be due to the variable power available.

Results

The OMCD repository is available at https://www.oncomir.umn.edu/omcd. OMCD provides 4 different types of search functions (**Figure 1A**). Here we use the miR-21 expression in colon cancer (COAD) as an example. In the current version of OMCD repository, we have 8 control samples and 272 tumor samples for COAD. Searching for miR-21 in COAD samples (**Figure 1A, B**) returns a heat map showing the absolute expression level of miR-21 in all COAD samples (**Figure 1C**). Users can also obtain the numerical expression data (**Figure 1D**; not

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

completely shown due to limited space) and the relative expression data (**Figure 1E**). When users click on the miRNA, they will be taken to a page showing links to additional analysis (**Figure 1F**). Here the website provides detailed information about the location of the miRNA and the name of colocalized miRNAs (miRNA clusters), as well as internal links to the expression data of miR-21 in other cancers and further statistical analysis (**Figure 1H**). Additionally, we provide external links to the miRDB website for target prediction and literature search on Google scholar (7). From this page, we can then visualize the colocalized miRNA expressions in a heat map showing absolute expression (**Figure 1G**). This analysis will display the expression levels for colocalized miRNAs in all cancer types (not shown due to limited space) and can be visualized in both absolute and relative heat map and numerical data form. Finally, we show the statistical analyses of miR-21 comparing Control vs. Tumor, Tumor vs. Tumor and Control vs. Control across cancer types (**Figure 1H**). The 3 different comparisons allow for simple visualization of the expression patterns of miR-21 across different cancer groups.

To further demonstrate the utility of our database we identified miRNA which was recurrently significantly differentially expressed between tumor and normal control samples with a highly significant p-value < 0.000001 and an average fold change greater than the absolute value of 2 which were recurrently present in 5 or more tumor normal comparisons (**Figure 2**). Many of these miRNAs are well known in cancer and have been reported to differentially expressed between tumor and normal in a wide range of tumor types. For example, miR-21 is consistently upregulated in most cancer types, consistent with previous reports (8). Suggesting miR-21 could potentially be a general cancer biomarker, however not a suitable biomarker for specific cancer types.

The miRNA expression in cancers from similar anatomical regions shows more similar patterns. This indicates tissue-specific miRNA expressions. For example, the colon cancer (COAD) and rectal cancer (READ) clusters have very similar miRNA expression pattern

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared to other cancers. Although COAD and READ have overall similar miRNA expression pattern, miR-101-1 is significantly different between COAD tumor and control samples while not for READ **(Figure 2)**. Additionally, because miR-101 do not show significant high expression in other cancers, it is reasonable to hypothesize that this miRNA is a biomarker for COAD. Similarly, miR-10b shows significant high expression level in hepatocellular carcinoma (LIHC) but not in other cancer types. These are two examples of the potentially testable hypothesis that OMCD is able to generate. Further experimental validation is thus warranted to investigate the function of miR-101 in COAD and miR-10b in LIHC.

Discussion

Evidence from the past decade indicates miRNAs play crucial roles in cancers. With the development of high-throughput sequencing technology, more high-throughput miRNAs data is publicly available. Here we developed the OncoMir Cancer Database (OMCD), a simple web-based repository that allows easy and systematic comparative genomic analyses of miRNA expression data. Using OMCD, we were able to identify miR-101 as a biomarker candidate specifically for COAD. We found that the expression level of miR-101 is significantly higher in COAD tumors but not in other tumors, compared to respective normal controls. Previous studies, however, show different expression levels of miR-101 in colorectal cancer (9,10). Contrary to the TCGA-COAD data which shows high miR-101 in tumors, previous studies suggest miR-101 is downregulated in colorectal tumors and this miRNA act as a tumor-suppressing miRNA where overexpression of miR-101 can inhibit the tumor invasion and growth (9,10). Using the OncomiR database, we were able to verify that miR-101 is indeed overexpressed in COAD tumors based on the TCGA data. We think the mixed results for miR-101 in COAD between TCGA and other cohorts warrant a further investigation into the function of miR-101 in COAD to ascertain if miR-101 is a suitable biomarker for COAD. We also observed from OMCD that miR-10b could be a potential biomarker for LIHC (11). Previous

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

studies confirmed that miR-10b is indeed highly expressed in LIHC and it is involved in neoplastic transformation of liver cancer stem cells and promotes metastasis (12–14). Additionally, other studies also show an oncogenic role of miR-10b in breast cancer, gastric cancer, and glioblastoma (15–18). These findings suggest miR-10b has a multifaceted function in many cancers, this further warrant studies to confirm if it is a suitable biomarker candidate specifically for LIHC.

In the current version, OMCD contains data derived from 9,656 cancer patients with associated clinical information and organ-specific controls present in TCGA. To our knowledge, OncomiR database (www.oncomir.org) is the only other online resource for analyzing miRNA expression data (5). The conflicting results we found between the TCGA and other cohorts demonstrate a limitation of the current version of OMCD as well as the OncomiR, which lack miRNA datasets from other cancer patient cohorts. In addition, the OncomiR database lacks the option to analyze miRNA clusters. It is important to consider miRNA cluster members when studying miRNAs in cancers, especially to generate hypothesis from high-throughput data. This is because miRNA cluster members usually have similar expression levels, however, potentially vast different biological functions. The ability to visualize and explore miRNA clusters in OMCD is crucial to develop a defendable hypothesis. In the future, we plan to expand OMCD by incorporating additional miRNA expression data sets from public data repositories such as Gene Expression Omnibus (GEO), Genomic Data Commons (GDC), and European Bioinformatics Institute (EBI). We believe this will significantly improve the ability to use OMCD to develop a defendable hypothesis.

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Number of cases included in the OMCD for each cancer type.

Figure legends

Figure 1: Screenshots of sample analysis of miR-21 in COAD. **A, B)** Advanced Search options in OMCD allows searching by miRNA, tissue types, and statistical results. **C)** Heatmap, **D)** numerical view showing absolute expressions of miR-21 in COAD normal and tumor tissues. **E)** Heatmap showing relative expression of miR-21 in COAD. **F)** Information and external links and **G)** heat map showing miR-21 cluster members allows exploring the expression patterns of colocalized miRNAs. **H)** Statistical results of group-based comparisons for miR-21 for different tissue types.

Figure 2: Heatmap of significantly differentially expressed miRNAs between tumor and normal samples. MiRNAs with p-value < 0.000001 and an average fold change greater than the absolute value of 2 which were recurrently present in 5 or more tumor-normal comparisons are shown. Red indicates upregulation and green indicate downregulation between tumor and normal tissues.

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

SS is supported by research grants funded by the NIH/NCI grant R03CA219129; and CY, by a research fellowship from the Bioinformatics and Computaional Biology graduate program, University of Minnesota. We thank Dr. Mary Knatterud for assisting in manuscript preparation.

Conflict of interest

The authors declare no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

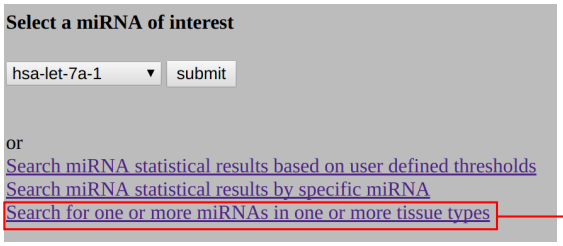
1. Ambros, V. (2004) *Nature*, **431**, 350–355, The functions of animal microRNAs.
2. Calin, G. A. and Croce, C. M. (2006) *Nat Rev Cancer*, **6**, 857–866, MicroRNA signatures in human cancers.
3. Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., et al. (2013) *Nat Genet*, **45**, 1113–1120, The Cancer Genome Atlas Pan-Cancer analysis project.
4. Gao, J., Aksoy, B. A., Dogrusoz, U., et al. (2013) *Sci Signal*, **6**, pl1, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.
5. Wong, N. W., Chen, Y., Chen, S., et al. (2017) *Bioinformatics*, OncomiR: An online resource for exploring pan-cancer microRNA dysregulation.
6. Sarver, A. L., Phalak, R., Thayanithy, V., et al. (2010) *Lab Invest*, **90**, 753–761, S-MED: sarcoma microRNA expression database.
7. Wang, X. (2008) *RNA*, **14**, 1012–1017, miRDB: a microRNA target prediction and functional annotation database with a wiki interface.
8. Krichevsky, A. M. and Gabriely, G. (2009) *J Cell Mol Med*, **13**, 39–53, miR-21: a small multi-faceted RNA.
9. Chen, M.-B., Yang, L., Lu, P.-H., et al. (2015) *Biochem Biophys Res Commun*, **463**, 954–960, MicroRNA-101 down-regulates sphingosine kinase 1 in colorectal cancer cells.
10. Strillacci, A., Griffoni, C., Sansone, P., et al. (2009) *Exp Cell Res*, **315**, 1439–1447, MiR-101 downregulation is involved in cyclooxygenase-2 overexpression in human colon cancer cells.
11. Zhang, Y., Liao, R.-B., Hu, L.-L., et al. (2017) *Oncotarget*, **8**, 104543–104551, The microRNA miR-10b as a potentially promising biomarker to predict the prognosis of cancer patients: a meta-analysis.
12. Li, R., Qian, N., Tao, K., et al. (2010) *J Exp Clin Cancer Res*, **29**, 169, MicroRNAs involved in neoplastic transformation of liver cancer stem cells.
13. Ladeiro, Y., Couchy, G., Balabaud, C., et al. (2008) *Hepatology*, **47**, 1955–1963, MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations.
14. Li, D., Zhang, Y., Zhang, H., et al. (2018) *J Exp Clin Cancer Res*, **37**, 46, CADM2, as a new target of miR-10b, promotes tumor metastasis through FAK/AKT pathway in hepatocellular carcinoma.
15. Guessous, F., Alvarado-Velez, M., Marcinkiewicz, L., et al. (2013) *J Neurooncol*, **112**, 153–163, Oncogenic effects of miR-10b in glioblastoma stem cells.

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

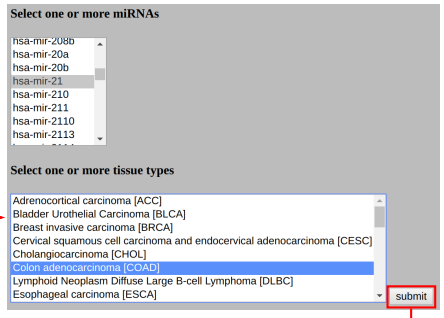
16. Liu, Z., Zhu, J., Cao, H., et al. (2012) *Int J Oncol*, **40**, 1553–1560, miR-10b promotes cell invasion through RhoC-AKT signaling pathway by targeting HOXD10 in gastric cancer.
17. Ma, L., Teruya-Feldstein, J. and Weinberg, R. A. (2007) *Nature*, **449**, 682–688, Tumour invasion and metastasis initiated by microRNA-10b in breast cancer.
18. Zhang, J., Yang, J., Zhang, X., et al. (2018) *PLoS ONE*, **13**, e0192509, MicroRNA-10b expression in breast cancer and its clinical association.

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

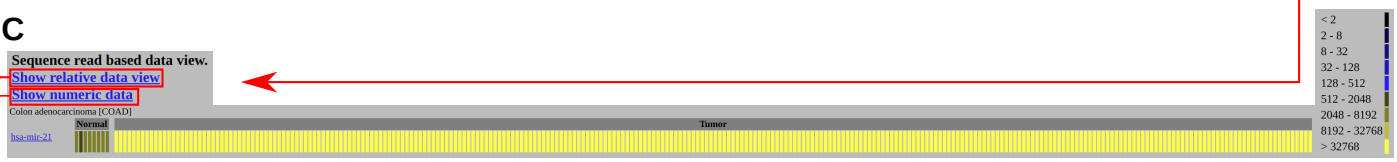
A



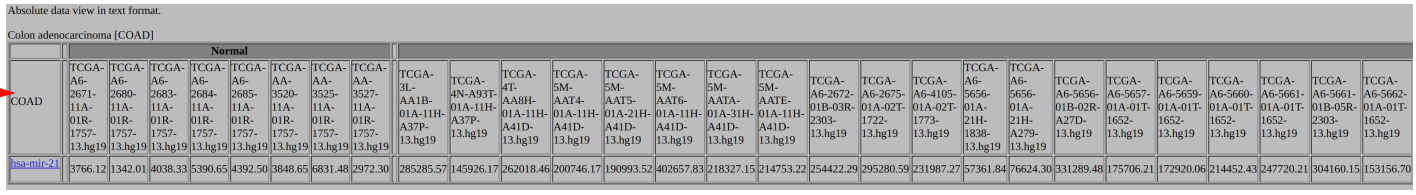
B



C



D

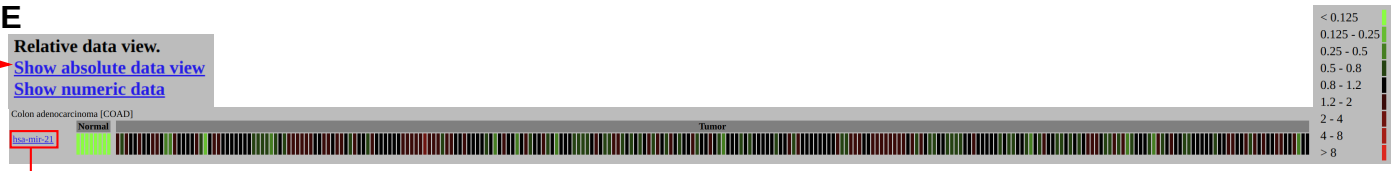


Absolute data view in text format.

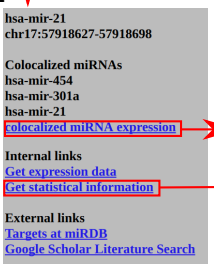
Colon adenocarcinoma [COAD]

	Normal								Tumor																				
	TCGA-A6-2671-11A-01R-1757-13.hg19	TCGA-A6-2680-11A-01R-1757-13.hg19	TCGA-A6-2683-11A-01R-1757-13.hg19	TCGA-A6-2684-11A-01R-1757-13.hg19	TCGA-A6-2685-11A-01R-1757-13.hg19	TCGA-AA-3520-11A-01R-1757-13.hg19	TCGA-AA-3525-11A-01R-1757-13.hg19	TCGA-AA-3527-11A-01R-1757-13.hg19	TCGA-3L-AA1B-01A-11H-A37P-13.hg19	TCGA-4N-A93T-01A-11H-A37P-13.hg19	TCGA-4T-AA8H-01A-11H-A41D-13.hg19	TCGA-5M-AAT4-01A-11H-A41D-13.hg19	TCGA-5M-AAT5-01A-21H-A41D-13.hg19	TCGA-5M-AAT6-01A-11H-A41D-13.hg19	TCGA-5M-AATA-01A-31H-A41D-13.hg19	TCGA-5M-AATE-01A-11H-A41D-13.hg19	TCGA-A6-2672-01B-03R-2303-13.hg19	TCGA-A6-2675-01A-02T-1722-13.hg19	TCGA-A6-4105-01A-02T-1773-13.hg19	TCGA-A6-5656-01A-21H-1838-13.hg19	TCGA-A6-5656-01A-21H-A279-13.hg19	TCGA-A6-5656-01B-02R-A27D-13.hg19	TCGA-A6-5657-01A-01T-1652-13.hg19	TCGA-A6-5659-01A-01T-1652-13.hg19	TCGA-A6-5660-01A-01T-1652-13.hg19	TCGA-A6-5661-01A-01T-1652-13.hg19	TCGA-A6-5661-01B-05R-2303-13.hg19	TCGA-A6-5662-01A-01T-1652-13.hg19
COAD																													
hsa-mir-21	3766.12	1342.01	4038.33	5390.65	4392.50	3848.65	6831.48	2972.30	285285.57	145926.17	262018.46	200746.17	190993.52	402657.83	218327.15	214753.22	254422.29	295280.59	231987.27	57361.84	76624.30	331289.48	175706.21	172920.06	214452.43	247720.21	304160.15	153156.70

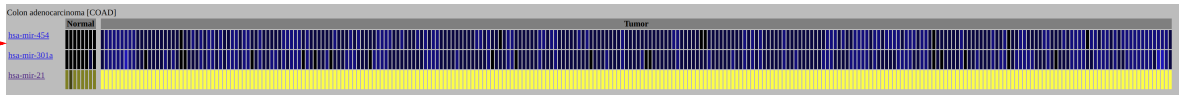
E



F



G

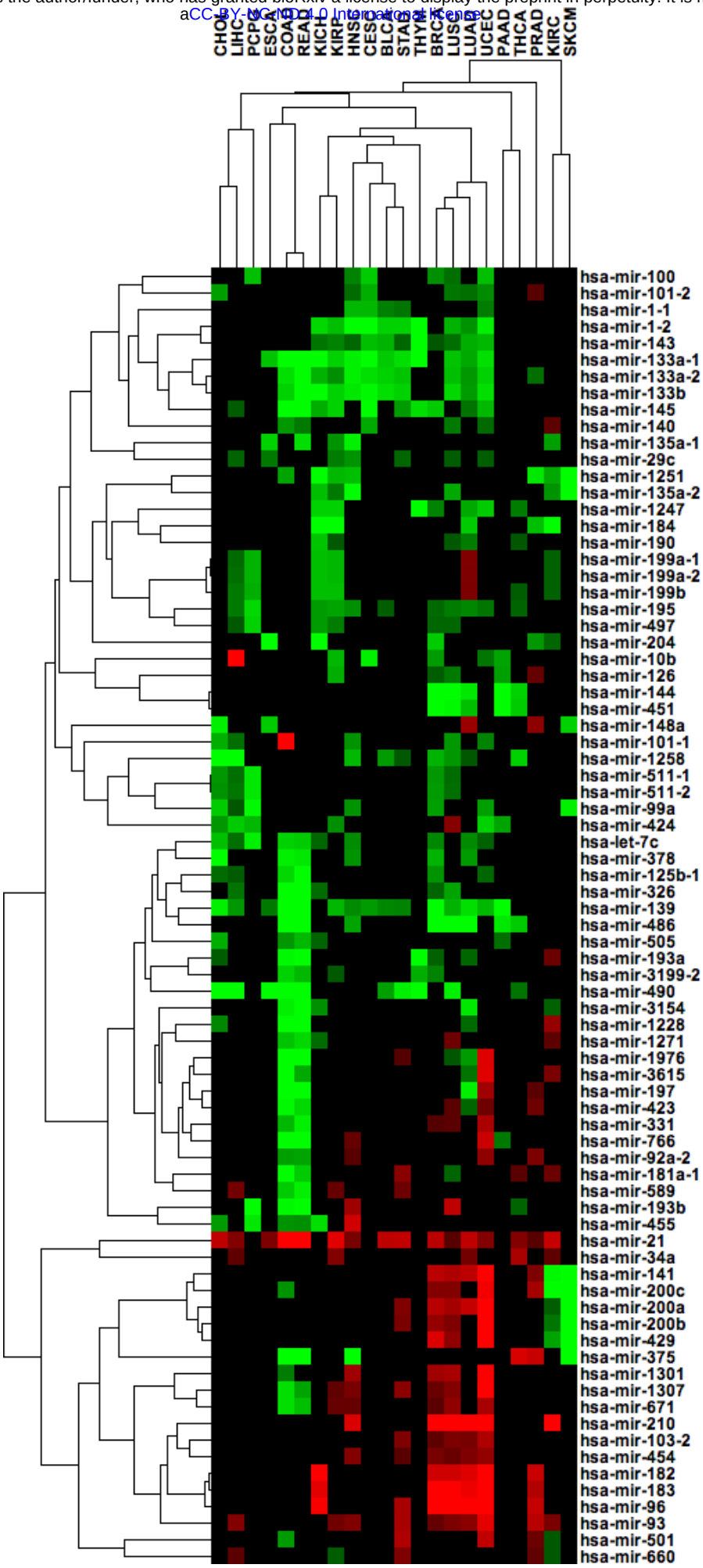


H

Group based statistics for each comparison for the chosen miRNA are provided in the table below.
hsa-mir-21 expression data

Dataset	Comparison	*P* Value	Ratio	AVG group1	AVG group2	Comparison Description
ACC	TT	0.00000000000001488	0.41	90048.39	219664.73	Adrenocortical carcinoma [ACC]: Tumor Tissue compared to all Tumor Tissues
BLCA	NT	2.431e-16	4.73	64282.18	304372.55	Bladder Urothelial Carcinoma [BLCA]: Normal Tissue compared to Tumor Tissue
BLCA	NN	0.5913	0.89	64282.18	72217.76	Bladder Urothelial Carcinoma [BLCA]: Normal Tissue compared to all Normal Tissues
BLCA	TT	2.2e-16	1.39	304372.60	219664.70	Bladder Urothelial Carcinoma [BLCA]: Tumor Tissue compared to all Tumor Tissues
BRCA	NT	2.2e-16	4.6	54095.55	249091.46	Breast invasive carcinoma [BRCA]: Normal Tissue compared to Tumor Tissue
BRCA	NN	0.009038	0.75	54095.55	72217.76	Breast invasive carcinoma [BRCA]: Normal Tissue compared to all Normal Tissues
BRCA	TT	0.00000006455	1.13	249091.50	219664.70	Breast invasive carcinoma [BRCA]: Tumor Tissue compared to all Tumor Tissues
CESC	NT	0.000005659	6.27	59372.14	372010.30	Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]: Normal Tissue compared to Tumor Tissue
CESC	NN	0.7263	0.82	59372.14	72217.76	Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]: Normal Tissue compared to all Normal Tissues
CESC	TT	2.2e-16	1.69	372010.30	219664.70	Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]: Tumor Tissue compared to all Tumor Tissues
CHOL	NT	0.00000000008587	4.8	92729.64	444823.37	Cholangiocarcinoma [CHOL]: Normal Tissue compared to Tumor Tissue
CHOL	NN	0.3338	1.28	92729.64	72217.76	Cholangiocarcinoma [CHOL]: Normal Tissue compared to all Normal Tissues
CHOL	TT	2.2e-16	2.03	444823.40	219664.70	Cholangiocarcinoma [CHOL]: Tumor Tissue compared to all Tumor Tissues
COAD	NT	8.031e-16	52.62	4072.76	214324.28	Colon adenocarcinoma [COAD]: Normal Tissue compared to Tumor Tissue
COAD	NN	0.002503	0.06	4072.76	72217.76	Colon adenocarcinoma [COAD]: Normal Tissue compared to all Normal Tissues
COAD	TT	0.5567	0.98	214324.30	219664.70	Colon adenocarcinoma [COAD]: Tumor Tissue compared to all Tumor Tissues

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cancer Type (TCGA Code)	Total number of samples	Tumor	Normal
Breast invasive carcinoma [BRCA]	869	782	87
Brain Lower Grade Glioma [LGG]	530	530	0
Thyroid carcinoma [THCA]	573	514	59
Prostate adenocarcinoma [PRAD]	551	499	52
Ovarian serous cystadenocarcinoma [OV]	495	495	0
Head and Neck squamous cell carcinoma [HNSC]	532	488	44
Lung adenocarcinoma [LUAD]	504	458	46
Skin Cutaneous Melanoma [SKCM]	453	451	2
Uterine Carcinosarcoma [UCS]	450	418	32
Bladder Urothelial Carcinoma [BLCA]	436	417	19
Stomach adenocarcinoma [STAD]	450	404	46
Liver hepatocellular carcinoma [LIHC]	426	375	51
Lung squamous cell carcinoma [LUSC]	388	343	45
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	313	310	3
Kidney renal papillary cell carcinoma [KIRP]	326	292	34
Colon adenocarcinoma [COAD]	280	272	8
Sarcoma [SARC]	263	263	0
Kidney renal clear cell carcinoma [KIRC]	332	261	71
Esophageal carcinoma [ESCA]	200	187	13
Pheochromocytoma and Paraganglioma [PCPG]	187	184	3
Pancreatic adenocarcinoma [PAAD]	183	179	4
Testicular Germ Cell Tumors [TGCT]	156	156	0
Thymoma [THYM]	126	124	2
Rectum adenocarcinoma [READ]	97	94	3
Mesothelioma [MESO]	87	87	0
Uveal Melanoma [UVM]	80	80	0
Adrenocortical carcinoma [ACC]	79	79	0
Kidney Chromophobe [KICH]	91	66	25
Uterine Corpus Endometrial Carcinoma [UCEC]	57	57	0
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma [DLBC]	47	47	0
FFPE Pilot Phase II [FPPP]	45	45	0
Cholangiocarcinoma [CHOL]	45	36	9
Glioblastoma multiforme [GBM]	5	0	5
Total	**9,656**	**8,993**	**663**

bioRxiv preprint doi: https://doi.org/10.1101/279125; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.