



Data in Brief

Profiling miRNAs in nasopharyngeal carcinoma FFPE tissue by microarray and Next Generation Sequencing



Jin Peng^{a,b}, Yanjun Feng^{a,b}, Gabriel Rinaldi^{a,b}, Paul Levine^c, Samantha Easley^d, Elizabeth Martinez^d, Salman Hashmi^c, Nader Sadeghi^e, Paul J. Brindley^{a,b}, Jason P. Mulvenna^{f,g}, Jeffrey M. Bethony^{a,b}, Jordan L. Plieskatt^{a,b,*}

^a Department of Microbiology, Immunology and Tropical Medicine, School of Medicine and Health Science, George Washington University, Washington DC, USA

^b Research Center for Neglected Diseases of Poverty, School of Medicine and Health Science, George Washington University, Washington DC, USA

^c Department of Epidemiology and Biostatistics, The George Washington University School of Public Health and Health Services, Washington DC 20037, USA

^d Department of Pathology, School of Medicine and Health Science, George Washington University, Washington DC, USA

^e Medical Faculty Associates, The George Washington University, Washington DC 20037, USA

^f School of Biomedical Sciences, Faculty of Medicine and Biomedical Sciences, University of Queensland, Brisbane, Australia

^g Infectious Diseases Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

ARTICLE INFO

Article history:

Received 28 July 2014

Received in revised form 18 August 2014

Accepted 18 August 2014

Available online 27 August 2014

Keywords:

miRNA

Biomarker

Microarray

Nasopharyngeal carcinoma

RNA-Seq

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a non-lymphomatous, squamous-cell carcinoma that occurs in the epithelial lining of the nasopharynx. Nasopharyngeal carcinoma has a geographically well-defined distribution worldwide, with the highest prevalence in China, Southeast Asia, and Northern Africa. Symptoms of nascent NPC may be unapparent or trivial, with diagnosis based on the histopathology of biopsied tissue following endoscopy of the nasopharynx. The tumor node metastasis (TNM) staging system is the benchmark for the prognosis of NPC and guides treatment strategy. However, there is a consensus that the TNM system is not sufficiently specific for the prognosis of NPC, as it does not reflect the biological heterogeneity of this tumor, making another biomarker for the detection of NPC a priority. We have previously reported on different approaches for microRNA (miRNA) biomarker discovery for Formalin Fixed Paraffin Embedded (FFPE) NPC tissue samples by both a targeted (microarray) and an untargeted (small RNA-Seq) discovery platform. Both miRNA discovery platforms produced similar results, narrowing the miRNA signature to 1–5% of the known mature human miRNAs, with untargeted (small RNA-Seq approach) having the advantage of indicating “unknown” miRNAs associated with NPC. Both miRNA profiles strongly associated with NPC, providing two potential discovery platforms for biomarker signatures for NPC. Herein, we provide a detailed description of the methods that we used to interrogate FFPE samples to discover biomarkers for NPC.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

Specifications

Organism/cell line/tissue	<i>Homo sapiens</i>
Sex	5 males 3 females
Sequencer or array type	Agilent human miRNA microarray Illumina Genome Analyzer IIx
Data format	Raw and processed
Experimental factors	Tumor vs. adjacent tissue in FFPE
Experimental features	Biomarker analysis in NPC FFPE tissue between tumor and control nasopharynx.
Consent	IRB approved
Sample source location	Washington DC, United States

Direct link to deposited data

Microarray deposited data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46172>

RNA-Seq deposited data can be found here: <http://www.ncbi.nlm.nih.gov/sra/?term=SRP029599>

Experimental design, materials and methods

Experimental cases

Case and control tissue including sample characteristics are presented in Table 1 (and in detail in [1]). In brief, four formalin fixed paraffin-embedded (FFPE) tissues from cases of histologically confirmed non-keratinizing NPC and four FFPE cases of normal nasopharyngeal tissue

* Corresponding author at: Department of Microbiology, Immunology and Tropical Medicine, School of Medicine & Health Sciences, The George Washington University, 2300 Eye Street, NW, Washington DC, 20037, USA. Fax: +1 202 994 2913.

E-mail address: jordanp@gwu.edu (J.L. Plieskatt).

were obtained from the biological repository in the Department of Pathology of The George Washington University Hospital, Washington, DC. Tissue sections from FFPE were reviewed by two independent pathologists (E.M. and S.E.) to confirm the diagnosis as shown in [1]. FFPE preparation, hematoxylin and eosin (H&E) staining, and representative images have also been previously reported [1]. It should also be noted that the SRA project submission contains four additional samples (Accession: SRX345915, SRX345913, SRX345913 and SRX345909). These samples reference a survey of serum pools from NPC positive and control individuals discussed in [1] but not further referenced herein.

RNA isolation

Total RNA was isolated from $2 \times 10 \mu\text{m}$ sections from each FFPE case using the miRNeasy FFPE kit (Qiagen) [1]. RNA concentration, purity, and integrity (RIN) were determined by spectrophotometry (Nanodrop 1000) and the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano and small RNA kits. Purified RNA was stored at $<-50^\circ\text{C}$.

Yields of total RNA derived from FFPE were approximately $100 \text{ ng}/\mu\text{m}$ with 260/280 and 260/230 ratios of ~ 2.0 and ~ 1.9 , respectively. Analysis on the Agilent Bioanalyzer indicated that the samples were enriched for small RNA species with integrity (RNA Integrity Number or RIN) values of two to three. Though typically indicative of RNA degradation, the robustness of miRNAs in these FFPE tissue [2] and reports from other groups [3] that RIN values have negligible effect on miRNA results enabled us to consider this purified RNA suitable for further analysis by microarray and RNA sequencing.

Microarray, data normalization and analysis

All eight samples underwent analysis via microarray (Table 1). Total RNA isolated from each FFPE case was labeled and hybridized to an Agilent human miRNA microarray (miRBase Release 16.0) and scanned [1]. The intensities of each sample were transferred to digital data and \log_2 transformed using Agilent Feature Extraction (V.10.7). Raw data files in text (.txt) format were analyzed with Agilent GeneSpring

software (GX 12.6) [4]. A total of 1205 human and 144 human viral microRNAs were used from miRBase v16.0.

To analyze the differentially expressed miRNAs, quantile normalization was performed to standardize these data across the samples. Raw data (thresholded and log base 2 transformed) were filtered by expression values (20.0–336133.0) with at least two out of the eight samples having values within the cut-off range to remove very low signal values and background influence. The four tumor samples were grouped and analyzed against the four control samples by unpaired Student's *t*-test with a *p*-value cut-off of 0.05 (*p*-value obtained by Asymptotic analysis) and a fold-change cut-off of 2.0. Hierarchical clustering was then performed [1] using the Euclidean distance metric and Centroid linkage rule. We identified 35 significantly dysregulated miRNAs, including four Epstein–Barr Virus (EBV) miRNAs and 31 human miRNAs (13 down-regulated and 18 up-regulated) [1]. These analyses were conducted again for this manuscript to verify their reproducibility. In addition, the miRNA signatures were compared to the recently released miRBase (v 19.0) with its up-date the miRNA nomenclature (Table 2) than in the original publication of these data, which used miRBase (v 16.0) [1].

Significance analysis was completed using GeneSpring [4] as detailed below:

- 1) A new *project* was created, followed by a new *experiment*, and *miRNA* was selected for analysis type, followed by the *data import wizard* for workflow type.
- 2) In *New miRNA Experiment Steps*, the raw intensity files were uploaded. The selected technology was set to 31181_v16_0 and no baseline transformation was performed. The threshold raw signals were set to 1.0 and *quantile* was chosen as the normalization algorithm along.
- 3) In the *Experiment Setup*, the samples were grouped into four tumor and four control cases under the *Experiment Grouping* option. While further interpretations may be created depending on analysis requirements, in this case experimental parameters “tumor/control” (categorical) were set up.

The condition tumor and control were selected and *Non-Averaged for the Average Over Replicates* in Conditions. *Detected* and *Not Detected* were selected and *Compromised in Use Measurements Flagged*.

Table 1
List of the raw data files deposited to NCBI GEO and SRA with accession numbers. Further details on the FFPE sample set in [1] with histological type, TNM staging [9] and WHO classification [10].

Sample ID	Type	Sex	Age	Histological type	Clinical Stage (TNM)	Discovery Techniques Performed	Yes	Accession Number	File
08-S-6658A ^a	Tumor	M	46	Differentiated	T4N2M0	Microarray	Yes	NCBI: GSM1125481	GSM1125481_US12302349_253118112871_S0_1_miRNA_107_Sep09_1_1.txt.gz
						Small RNA-Seq	Yes	SRA: SRX345905	O8S-6658A_L1.LB1.fastq
04-S-341B ^b	Tumor	F	80	Differentiated	T3N0M0	Microarray	Yes	NCBI: GSM1125482	GSM1125482_US12302349_253118112871_S0_1_miRNA_107_Sep09_1_2.txt.gz
						Small RNA-Seq	No	//////	//////
04-S-6103A	Tumor	M	47	Undifferentiated	T4N2M0	Microarray	Yes	NCBI: GSM1125483	GSM1125483_US12302349_253118112871_S0_1_miRNA_107_Sep09_1_3.txt.gz
						Small RNA-Seq	Yes	SRA: SRX345917	O4S-6103A_L1.LB3.fastq
12-S-432A	Tumor	M	49	Undifferentiated	T3N1M0	Microarray	Yes	NCBI: GSM1125484	GSM1125484_US12302349_253118112871_S0_1_miRNA_107_Sep09_1_4.txt.gz
						Small RNA-Seq	Yes	SRA: SRX345916	12S-432A_L1.LB4.fastq
08-S-6658C ^a	Control: Adjacent tissue	M	46	Differentiated	T4N2M0	Microarray	Yes	NCBI: GSM1125485	GSM1125485_US12302349_253118112871_S0_1_miRNA_107_Sep09_2_1.txt.gz
						Small RNA-Seq	Yes	SRA: SRX345908	O8S-6658C_L1.LB5.fastq
04-S-341E ^b	Control: Adjacent tissue	F	80	Differentiated	T3N0M0	Microarray	Yes	NCBI: GSM1125486	GSM1125486_US12302349_253118112871_S0_1_miRNA_107_Sep09_2_2.txt.gz
						Small RNA-Seq	No	//////	//////
05-S-5264-F	Control: Adjacent tissue	F	43	Undifferentiated	T4N0M0	Microarray	Yes	NCBI: GSM1125487	GSM1125487_US12302349_253118112871_S0_1_miRNA_107_Sep09_2_3.txt.gz
						Small RNA-Seq	Yes	SRA: SRX345906	O5S-5264F_L1.LB6.fastq
11-S-11311-E	Control: Adjacent tissue	M	65	Undifferentiated	T1N0M0	Microarray	Yes	NCBI: GSM1125488	GSM1125488_US12302349_253118112871_S0_1_miRNA_107_Sep09_2_4.txt.gz
						Small RNA-Seq	No	//////	//////

^{a,b}Denotes those from the same patient (i.e. paired NPC/Control tissue samples).

Table 2

Microarray miRNA expression analysis between tumor and control NPC FFPE tissue using unpaired Student's *t*-test (*p*-value <0.05 and fold-change >2.0). In this repeated analysis by GeneSpring updated nomenclature found in miRBase v19.0 was utilized to update the sample set found in [1]. Thirty-five miRNAs were dysregulated comprising four EBV specific miRNAs.

Systematic name	FC	Log FC	Regulation	Active sequence	Chromosome	miRBase accession no.
ebv-miR-BART4-3p	99.86	6.64	Up	ACACCTGGTGCCTAC	-	MIMAT0009204
ebv-miR-BART5-5p	69.01	6.11	Up	CGATGGGCAGCTATA	-	MIMAT0003413
ebv-miR-BART6-3p	92.81	6.54	Up	TCTAAGGCTAGTCCGAT	-	MIMAT0003415
ebv-miR-BART6-5p	99.86	6.64	Up	CCTATGGATTGGACCAA	-	MIMAT0003414
hsa-let-7b-5p	-2.09	-1.06	Down	AACCACACAACCTACTACC	chr22	MIMAT0000063
hsa-miR-100-5p	-2.92	-1.55	Down	CACAAGTTCGGATCTACGG	chr11	MIMAT0000098
hsa-miR-106b-5p	2.13	1.09	Up	ATCTGCACTGTCCAGC	chr7	MIMAT0000680
hsa-miR-125b-5p	-2.20	-1.14	Down	TCACAAGTTAGGGTCTC	chr11	MIMAT0000423
hsa-miR-1260a	2.27	1.18	Up	TGGTGGCAGAGGTGG	chr14	MIMAT0005911
hsa-miR-1274a_v16.0	2.94	1.56	Up	TGGCGCTGAAACAG	chr5	MIMAT0005927
hsa-miR-1274b_v16.0	2.44	1.29	Up	TGGCGCCCGAACAA	chr19	MIMAT0005938
hsa-miR-1275	-3.38	-1.76	Down	GACAGCCTCTCCCC	chr6	MIMAT0005929
hsa-miR-130b-3p	2.18	1.12	Up	ATGCCCTTTCATCATTGC	chr22	MIMAT0000691
hsa-miR-133b	-688.84	-9.43	Down	TAGCTGGTTGAAGGGGACC	chr6	MIMAT0000770
hsa-miR-141-3p	4.93	2.30	Up	CCATCTTTACCAGACAG	chr12	MIMAT0000432
hsa-miR-149-5p	7.11	2.83	Up	GGGAGTGAAGACACGGAG	chr2	MIMAT0000450
hsa-miR-15b-5p	2.05	1.03	Up	TGTAACCATGATGTGCTGC	chr3	MIMAT0000417
hsa-miR-17-3p	8.90	3.15	Up	CTACAAGTCCCTTAC	chr13	MIMAT0000071
hsa-miR-17-5p	2.44	1.29	Up	CTACCTGCACTGTAAGC	chr13	MIMAT0000070
hsa-miR-18a-5p	13.99	3.81	Up	CTATCTGCACTAGATGCA	chr13	MIMAT0000072
hsa-miR-195-5p	-6.00	-2.59	Down	GCCAATATTTCTGTGCTGC	chr17	MIMAT0000461
hsa-miR-196b-5p	47.26	5.56	Up	CCCAACAACAGGAAACTACC	chr7	MIMAT0001080
hsa-miR-199a-3p	-2.65	-1.41	Down	TAACCAATGTGCACTACT	chr1	MIMAT0000232
hsa-miR-199b-5p	-3.96	-1.99	Down	GAACAGATAGTCTAACACTGG	chr9	MIMAT0000263
hsa-miR-203a	41.64	5.38	Up	CTAGTGGTCTTAAACATT	chr14	MIMAT0000264
hsa-miR-20a-5p	2.18	1.13	Up	CTACCTGCACTATAAGCAC	chr13	MIMAT0000075
hsa-miR-221-3p	4.19	2.07	Up	GAAACCAGCAGACAATGT	chrX	MIMAT0000278
hsa-miR-25-3p	2.28	1.19	Up	TCAGACCGAGACAAGTGC	chr7	MIMAT0000081
hsa-miR-3138	-2.31	-1.21	Down	ACTCCCTCACTCACT	chr4	MIMAT0015006
hsa-miR-3651	3.06	1.62	Up	TCATGTACCAGCGACC	chr9	MIMAT0018071
hsa-miR-3663-3p	-2.55	-1.35	Down	GCGCCCGGCT	chr10	MIMAT0018085
hsa-miR-451a	-6.12	-2.61	Down	AACTCAGTAATGGTAAACGTTT	chr17	MIMAT0001631
hsa-miR-486-5p	-3.90	-1.96	Down	CTCGGGCAGCTCA	chr8	MIMAT0002177
hsa-miR-497-5p	-5.17	-2.37	Down	ACAAACCACAGTGTGCTG	chr17	MIMAT0002820
hsa-miR-93-5p	3.28	1.71	Up	CTACCTGCACGACAG	chr7	MIMAT0000093

- Quality control: The correlation coefficient value of all samples was >0.7 and therefore all the samples were used in further analysis. Further, 3D Principle Components Analysis (PCA) scores and plotting were used to determine any association among the samples (Fig. 1). It was noted that paired samples did not exhibit more significant clustering than non-paired (NPC/Control tissue) in the analysis (Fig. 1 and Hierarchical clustering [1]). In *Filter by Expression*, the right entity and interpretation were selected and filtered by raw data value. The lower cut-off value of the interest range was set to 20 and at least two out of eight samples had values within this range.
- In *Analysis*, the condition was set as tumor versus control, tested by *t*-test unpaired, and an asymptotic *p*-value was computed without correction. The fold change cut-off was >2.0 and analyzed under pairs of conditions with tumor compared to control. Hierarchical clustering analysis of differentially expressed genes from all samples was conducted on both entities and the conditions by normalized intensity values using Euclidean distance metric and Centroid linkage rule.

Small RNA sequencing

Small RNA sequencing was performed on five of the same samples used in microarray analysis (Table 1) of the three samples used in the previous analysis (control samples 341E and 11311E and tumor sample 341B) omitted due to the exhaustion of total RNA purified from the small tissue areas available for the study. Total RNA derived from the FFPE was subjected to Ribo-Zero Pretreatment using Ribo-Zero rRNA Removal Kit (Epicentre) as described by the manufacturer and in [1]. Library preparation and sequencing have been described in further details in [1]. Briefly, Illumina libraries were constructed from 1 µg of total RNA

using the TruSeq Small RNA Sample Kit (Illumina). Libraries were subjected to quality control prior to sequencing using an Agilent 2100 BioAnalyzer and concentration determination using PicoGreen (Invitrogen). The Illumina Genome Analyzer Iix was used to perform the sequencing by Expression Analysis, A Quintiles Company (Durham, NC).

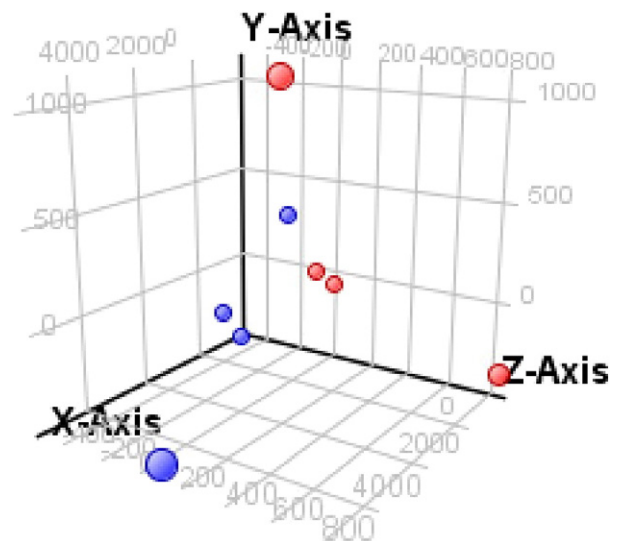


Fig. 1. 3D principal components analysis (PCA). PCA analysis of FFPE samples analyzed via microarray. Control FFPE tissue is denoted by red circles and NPC FFPE tissue is denoted by blue. No significant clustering was observed.

Table 3
Total reads derived from small RNA-Seq of five FFPE samples. Total numbers of reads mapped to miRNA are reported.

Sample ID	Type	Total reads	miRNA mapped	Unmapped reads	%mapped
08-S-6658A ^a	Tumor	5,609,415.00	2,569,276.00	3,040,139.00	46%
04-S-6103A	Tumor	4,338,399.00	2,258,744.00	2,079,655.00	52%
12-S-432A	Tumor	5,354,631.00	3,236,216.00	2,118,415.00	60%
08-S-6658C ^a	Control	6,265,846.00	3,755,819.00	2,510,027.00	60%
05-S-5264-F	Control	6,533,346.00	4,262,886.00	2,270,460.00	65%
	Total	28,101,637.00	16,082,941.00	12,018,696.00	57%

^a Denotes from the same patient (i.e. paired NPC/Control tissue samples).

Sequencing processing: alignment, mapping and annotation

Initial processing was performed using both FastqMcf and FastQC both of which can be accessed at <http://code.google.com/p/ea-utils/wiki/FastqMcf> and <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

After adaptor removal and quality filtering, ~28 million reads were aligned to the human (UCSC hg19) and Human herpes virus 4 (Epstein–Barr virus or EBV) genome (NCBI NC_007605.1) and miRNA counts generated for each sample [1]. Both miRDeep 2.0.0.5 [5] and miRExpress 2.0 [6] were used to generate counts, and each provided comparable results, with over 50% of the reads mapping to miRNAs in either the human or EBV genomes (Table 3). Identification of known miRNAs was based on miRBase Release 19 [7], with an alignment identity of 1%, a tolerance range of 4, and a similarity threshold of 0.8 [1]. In total, using miRDeep and miRExpress, 984 and 847 human and EBV miRNAs were identified, respectively, with a count per million greater than one in at least two of the samples.

Using EdgeR [8], a binomial distribution was used to compare the independent analyses from miRDeep and miRExpress [1]. The biological coefficient of variation (BCV) was used to estimate the variability across the dataset and plotted via the *plotBCV* function (Fig. 2A), with a common dispersion of 67% indicating a relatively high dispersion of gene expression levels. Given that this was an observational study on independent NPC cases using NPC tumors of different histological grades, such a value would not be considered atypical. Using the function *plotsmear* in EdgeR, log-fold changes were plotted against log-cpm (Fig. 2B). Using EdgeR, 99 dysregulated miRNAs were identified in NPC tumor tissue versus control tissue samples.

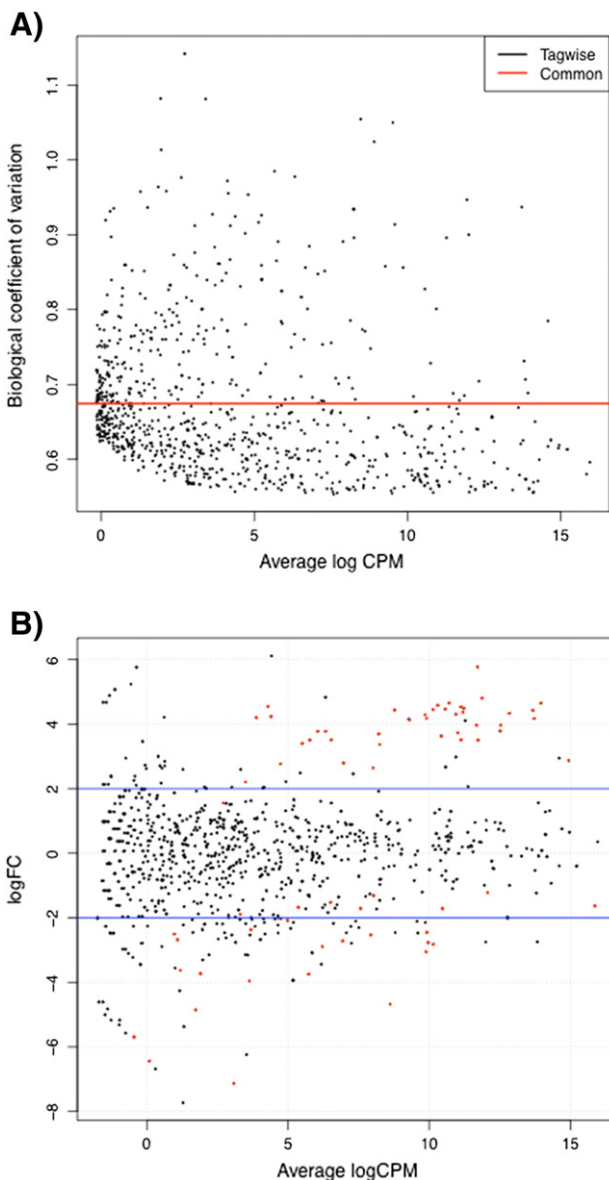


Fig. 2. RNA-sequencing output of five NPC FFPE samples. (A) Biological coefficient of variation reported against average log CPM. Red common trend line indicates the BCV of 67%. (B) logFC reported versus average log CPM. Log fold change of two is indicated by bracketed blue lines. Red dots indicate human miRNAs identified as significant (p value < 0.05).

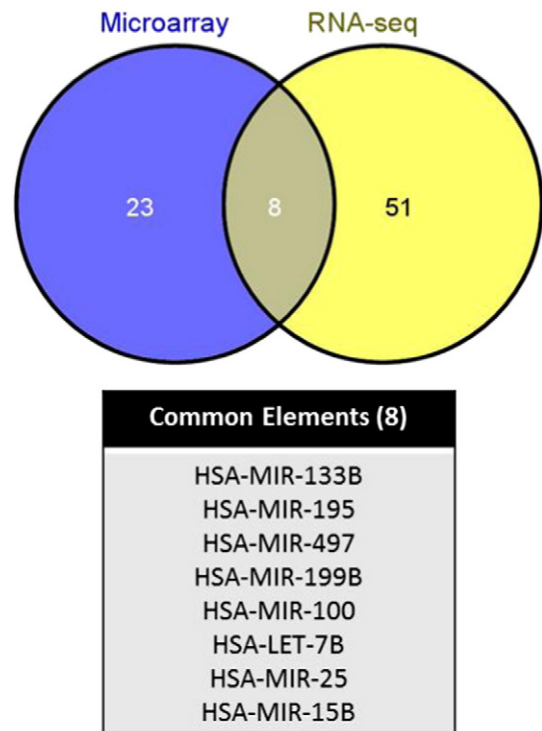


Fig. 3. Top common human miRNAs illustrated [11] as detected in corresponding independent analyses from both microarray and RNA-Seq. A total of eight common miRNAs were highlighted across both methods under the statistical cut-offs previously described [1].

Comparison of datasets

Both targeted (microarray) and untargeted (small RNA-Seq) approaches were extensively compared in our previously published manuscript [1]. While only eight dysregulated human miRNAs were identified in both the microarray and RNA-Seq analysis (Fig. 3) as statistically significant, the overall datasets were comparable. All but three miRNAs identified by microarray as significantly dysregulated were also identified by RNA-Seq, albeit not as significant [1]. In addition, these miRNAs also showed a similar dysregulation: i.e. if identified as up-regulated by microarray, they were also identified as up-regulated by small RNA-Seq [1].

Acknowledgments

The Agilent microarrays were hybridized and scanned at the Genomics and Epigenomics Shared Resource, Georgetown University Medical Center with the assistance of David Goerlitz.

This research was partially supported by awards R01CA155297 (JMB, JPM, and PJB) from the National Cancer Institute, P50AI098639 (BS, JMB, and PJB) from the National Institute of Allergy and Infectious Disease, fellowship support (JPM) and research support (JMB and JPM) under GNT1051627 from the National Health and Medical Research Council of Australia, and research support from the Dr. Cyrus And Myrtle Katzen Cancer Research Center at the George Washington University (PJB and JMB) (RG00982). The contents are solely the responsibility of the authors and do not represent the official views of

NIAID, NCI, the Katzen Cancer Research Center of the George Washington University, or the NHMRC of Australia.

References

- [1] J.L. Plieskatt, G. Rinaldi, Y. Feng, P.H. Levine, S. Easley, E. Martinez, S. Hashmi, N. Sadeghi, P.J. Brindley, J.M. Bethony, J.P. Mulvenna, Methods and matrices: approaches to identifying miRNAs for Nasopharyngeal carcinoma. *J. Transl. Med.* 12 (2014) 3.
- [2] Y. Xi, G. Nakajima, E. Gavin, C.G. Morris, K. Kudo, K. Hayashi, J. Ju, Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* 13 (2007) 1668–1674.
- [3] M. Jung, A. Schaefer, I. Steiner, C. Kempkensteffen, C. Stephan, A. Erbersdobler, K. Jung, Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. *Clin. Chem.* 56 (2010) 998–1006.
- [4] Agilent GeneSpring user manual, edition: May 2013 http://genespring-support.com/files/gs_12_6/GeneSpring-manual.pdf.
- [5] M.R. Friedlander, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, N. Rajewsky, Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26 (2008) 407–415.
- [6] W.C. Wang, F.M. Lin, W.C. Chang, K.Y. Lin, H.D. Huang, Lin NS: miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinforma.* 10 (2009) 328.
- [7] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39 (2011) D152–D157.
- [8] M.D. Robinson, G.K. Smyth, Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23 (2007) 2881–2887.
- [9] S. Edge, D. Byrd, C. Compton, A. Fritz, F. Greene, A. Trotti, *AJCC Cancer Staging Manual*. 7th edition Springer, New York, NY, 2010.
- [10] J. Chan, B. Pilch, T. Kuo, B. Wenig, A. Lee, *Nasopharyngeal Carcinoma*. IARC Press, Lyon, France, 2005.
- [11] J.C. Oliveros, VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>2007.