RESEARCH

Evaluating gastroenteropancreatic neuroendocrine tumors through microRNA sequencing

Nicole Panarelli^{1,*†}, Kathrin Tyryshkin^{2,*}, Justin Jong Mun Wong², Adrianna Majewski², Xiaojing Yang², Theresa Scognamiglio¹, Michelle Kang Kim³, Kimberly Bogardus⁴, Thomas Tuschl⁴, Yao-Tseng Chen¹ and Neil Renwick^{2,4}

¹Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York, USA

²Laboratory of Translational RNA Biology, Department of Pathology and Molecular Medicine, Queen's University, Kingston, Ontario, Canada

³Center for Carcinoid and Neuroendocrine Tumors of Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, New York, USA

⁴HHMI, Laboratory of RNA Molecular Biology, The Rockefeller University, New York, New York, USA

Correspondence should be addressed to N Renwick: neil.renwick@queensu.ca

*(N Panarelli and K Tyryshkin contributed equally to this work)

†(N Panarelli is now at Department of Pathology Albert Einstein College of Medicine, Bronx, New York, USA)

Abstract

Gastroenteropancreatic neuroendocrine tumors (GEP-NETs) can be challenging to evaluate histologically. MicroRNAs (miRNAs) are small RNA molecules that often are excellent biomarkers due to their abundance, cell-type and disease stage specificity and stability. To evaluate miRNAs as adjunct tissue markers for classifying and grading well-differentiated GEP-NETs, we generated and compared miRNA expression profiles from four pathological types of GEP-NETs. Using quantitative barcoded small RNA sequencing and state-of-the-art sequence annotation, we generated comprehensive miRNA expression profiles from archived pancreatic, ileal, appendiceal and rectal NETs. Following data preprocessing, we randomly assigned sample profiles to discovery (80%) and validation (20%) sets prior to data mining using machine-learning techniques. High expression analyses indicated that miR-375 was the most abundant individual miRNA and miRNA cistron in all samples. Leveraging prior knowledge that GEP-NET behavior is influenced by embryonic derivation, we developed a dual-layer hierarchical classifier for differentiating GEP-NET types. In the first layer, our classifier discriminated midgut (ileum, appendix) from non-midgut (rectum, pancreas) NETs based on miR-615 and -92b expression. In the second layer, our classifier discriminated ileal from appendiceal NETs based on miR-125b, -192 and -149 expression, and rectal from pancreatic NETs based on miR-429 and -487b expression. Our classifier achieved overall accuracies of 98.5% and 94.4% in discovery and validation sets, respectively. We also found provisional evidence that low- and intermediate-grade pancreatic NETs can be discriminated based on miR-328 expression. GEP-NETs can be reliably classified and potentially graded using a limited panel of miRNA markers, complementing morphological and immunohistochemistry-based approaches to histologic evaluation.

Key Words

- ▶ gastroenteropancreatic neuroendocrine tumors
- ▶ classification
- ▶ biomarkers
- ▶ microRNA
- ▶ small RNA sequencing

Endocrine-Related Cancer
(2019) **26**, 47–57

Introduction

Gastroenteropancreatic neuroendocrine tumors (GEP-NETs) are increasingly common and clinically diverse neoplasms (Yao *et al.* 2008, Lawrence *et al.* 2011) that are challenging to evaluate histologically (Klimstra 2016). Occurring throughout the digestive system, these tumors arise more frequently in the pancreas, ileum, appendix and rectum (Modlin *et al.* 2003, Yao *et al.* 2008, Lawrence *et al.* 2011). Due to non-specific symptomatology, many GEP-NETs are metastatic at diagnosis and the primary site is unknown in up to 20% of cases (Yao *et al.* 2008, Wang *et al.* 2010). Intriguingly, GEP-NET behavior may be linked to site of origin in the embryonic fore-, mid- or hindgut (Williams & Sandler 1963). Pathologic evaluation of NET tissues is a key component of clinical management (Klimstra 2016, Singh *et al.* 2016) because tumor site of origin and grade are linked to treatment and overall survival (Lawrence *et al.* 2011). However, existing immunohistochemical markers (Koo *et al.* 2012, Bellizzi 2013, Koo *et al.* 2013) and time-consuming and subjective mitotic counts or Ki67 immunostaining (Bosman *et al.* 2010, Tang *et al.* 2012, Modlin *et al.* 2016) hamper accurate classification and grading. Novel approaches and tissue markers are needed to assist histologic evaluation.

miRNAs are small (19–24 nucleotide) RNA molecules that are excellent biomarkers due to their abundance, cell type and disease stage specificity and stability in fresh and archived clinical samples (Gustafson *et al.* 2016). These regulatory molecules can also provide valuable insights into tumorigenesis through predictable targeting of mRNAs mediating oncogenesis or tumor suppression (Berindan-Neagoe *et al.* 2014, Acunzo *et al.* 2015). Due to their diagnostic utility in many other cancers (Lu *et al.* 2005), we hypothesized that miRNAs are also valuable tissue markers in GEP-NETs. To date, several groups have studied miRNA expression in GEP-NETs using a variety of study designs, detection methodologies and analytical approaches (Roldo *et al.* 2006, Ruebel *et al.* 2010, Li *et al.* 2013, Thorns *et al.* 2014, Lee *et al.* 2015, Mitsunashi *et al.* 2015, Miller *et al.* 2016, Mandal *et al.* 2017). All studies agree that miRNAs have biomarker potential.

Here, we assessed miRNA-based classification of GEP-NETs through quantitative barcoded small RNA sequencing, state-of-the-art sequence annotation and advanced data mining approaches (Farazi *et al.* 2012, Hafner *et al.* 2012, Brown *et al.* 2013). We also organized our miRNA expression data into transcriptional units, known as cistrons, to gauge data quality; individual

miRNAs from each cistron should be similarly upregulated or downregulated. Through high expression analyses, we identified a common miRNA marker for all tumors in our study. Leveraging prior knowledge that GEP-NET behavior varies by embryonic site of origin, we also constructed a dual-layer hierarchical classifier that accurately discriminates four GEP-NET pathological types. Lastly, we found provisional evidence that miRNAs can be used for tumor grading.

Materials and methods

Clinical materials and study design

GEP-NET cases were identified in the Department of Pathology, Weill Cornell Medicine. Hematoxylin-eosin-stained tissue sections from each case were reviewed and graded by an experienced pathologist (NP) according to the World Health Organization (WHO) Classification of Tumors of the Digestive Tract (Bosman *et al.* 2010). Four additional cases of low-grade rectal carcinoid tumors were obtained (through MK) from the Center for Carcinoid and Neuroendocrine Tumors of Mount Sinai, Icahn School of Medicine at Mount Sinai. Representative formalin-fixed paraffin-embedded (FFPE) tissue blocks from each case were obtained prior to RNA isolation, small RNA sequencing and data preprocessing and mining. Our project for utilizing de-identified archived samples was approved through the Research Ethics Board at Queen's University and the Institutional Review Boards of Weill Cornell Medicine, The Rockefeller University and Mt. Sinai School of Medicine.

Tumor grading

GEP-NETs were graded according to the 2010 WHO classification (Bosman *et al.* 2010). Briefly, tumors with mitotic counts <2 per ten 400X fields and <2% Ki67 proliferation index were classified as low grade (G1), whereas those with either a mitotic count of 2–20 per ten 400X fields or with a Ki67 index between 3 and 20% were classified as intermediate grade (G2).

Total RNA isolation

A representative tumor-bearing block was chosen in each case. The area containing tumor was circled on the corresponding hematoxylin and eosin-stained slide. Tissue cores were obtained from the center of the demarcated area to ensure that RNA was isolated from only neoplastic

tissue. Total RNA was isolated from two 1.5 mm tissue cores, bored from each tissue block, using the Qiagen RNeasy FFPE Kit according to the manufacturer's guidelines. Total RNA concentrations averaged 121 (range: 8–548) ng/ μ L for all samples.

Small RNA sequencing

miRNA expression profiles were generated using an established quantitative barcoded small RNA sequencing method (Hafner *et al.* 2012). Briefly, 100 ng total RNA from each sample was spiked with a set of ten oligoribonucleotide calibration markers prior to barcoded 3' oligonucleotide adapter ligation, sample pooling, 5' adapter ligation and reverse-transcription polymerase chain reaction. Small RNA cDNA libraries were sequenced on an Illumina HiSeq2500 platform in the Genomic Core Facility, The Rockefeller University. FASTQ sequence files were annotated through a purpose-built pipeline yielding individual miRNA and miRNA cistron expression profiles (Farazi *et al.* 2012, Brown *et al.* 2013). Total miRNA concentrations for each sample were calculated by multiplying the ratios of miRNA to calibrator read frequencies by the amount of input calibrator markers per microgram total RNA input. Sequencing data discussed in this manuscript are presented in Supplementary Tables 1, 2 and 3 (see section on [supplementary data](#) given at the end of this article).

Quantitative miRNA reverse-transcription PCR analyses

We measured miR-615 expression in 77 available study samples using TaqMan MicroRNA Assays (Applied Biosystems) according to the manufacturer's guidelines. Briefly, miRNAs were reverse transcribed using miRNA-specific stem-loop RT primers (Applied Biosystems) and 10 ng RNA input for each 15 μ L RT reaction and 1.33 μ L cDNA input for each 20 μ L PCR reaction (Applied Biosystems StepOnePlus System). PCR reactions were incubated in a 96-well plate format at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 1 min. All samples were assayed in triplicate. Mean Ct values were calculated for each sample and normalized against the corresponding U6 Ct values, calculated as $-(Ct_{miR} - Ct_{U6})$ (Schmittgen & Livak 2008); all data are presented as normalized Ct values. To assess the degree of similarity between RT-qPCR and sequencing results, we compared and correlated miR-615 expression data generated through both approaches.

Data preprocessing

Data preprocessing and subsequent analyses were performed using MATLAB (Mathworks, Inc., MA, USA, version R2016b). To identify sample outliers and/or sequencing batch effects, miRNA expression profiles were assessed through visual inspection and correlation analysis prior to data normalization, filtering and mining. Using all miRNA sequence reads, we computed the mean Spearman correlation for each sample to all other samples. Sample outliers were detected using the interquartile range method (Hoaglin & Iglewicz 1987) with $\alpha = 2.2$; three samples were removed from subsequent analyses.

Data normalization and filtering

To normalize miRNA expression profiles, individual miRNA sequence reads were divided by the total number of miRNA sequence reads for each sample; calibrator and miRNA cistron sequence reads were similarly normalized against total calibrator and total miRNA cistron sequence reads. Cases were randomly assigned to discovery (80%) or validation (20%) sets. To filter low expressed miRNAs in both sets, all miRNA STAR sequences and all miRNAs expressed below the 90th quantile of overall expression were removed; only miRNAs expressed above this threshold in more than 5% of samples from each tumor type were included in subsequent analyses. miRNA cistron expression data were similarly filtered.

High expression analyses

To identify candidate miRNA markers that were highly expressed in all GEP-NETs, we selected the top 0.5% expressed individual miRNAs and miRNA cistrons in all samples and ranked candidates in descending order of median expression in discovery and validation sets.

Discovery analyses

To identify individual or combinations of miRNAs that accurately discriminate GEP-NET subgroups and types, we leveraged prior knowledge that GEP-NET behavior is linked to embryonic site of origin in the fore-, mid- or hindgut (Williams & Sandler 1963). First, we looked for miRNAs that discriminated midgut (ileum and appendix) and non-midgut (rectum and pancreas) NET subgroups (comparison A). Subsequently, we investigated miRNA expression differences between ileal and appendiceal NETs (comparison B) and rectal and pancreatic NETs

(comparison C). We also compared miRNA expression differences between NETs from each anatomic site of origin (i.e. pancreas, ileum, appendix or rectum) and the remaining three GEP-NET types (comparisons D, E, F and G). Lastly, where sufficient samples were available, we assessed miRNA expression differences between low- and intermediate-grade NETs (comparison H). To avoid overfitting the classifier model, and thereby improve classifier performance, we reduced feature-space dimensionality (the number of miRNAs) by removing low expressed miRNAs from the discovery and validation sets. To rank individual miRNAs and miRNA cistrons for each comparison, we used an established feature selection (variable reduction) algorithm with 10-fold validation (Ren *et al.* 2017); only the top-ranking 3% individual miRNAs and miRNA cistrons were used in further investigations. To ensure that miRNAs were reliably detectable with our assay, we verified that median miRNA expression in at least one compared subgroup or type was higher than overall miRNA expression.

Hierarchical classifier

To ensure that our model is generalizable to other data sets, we generated our classifier using a discovery set comprising 80% of our data and 'held out' the remaining 20% of our data to serve as a validation set. Because there are no general purpose machine-learning algorithms (No Free Lunch Theorem) (Duda *et al.* 2001), we evaluated 23 different algorithms from the MATLAB Classification Learner App to find the most suitable classifier for the discovery set data. This classifier was subsequently applied in an iterative algorithm with 10-fold validation to identify the smallest subset of individual miRNAs that provided the most discriminatory power for each comparison in our discovery set. Based on these subsets and the selected classifier, we constructed a dual-layer hierarchical classifier in which expression profiles were initially classified as midgut (ileum and appendix) or non-midgut (rectum and pancreas) NETs prior to classification as either ileal or appendiceal NETs or rectal or pancreatic NETs. Lastly, we determined the accuracy of our classifier in discovery and validation sets.

Results

Small RNA sequencing and data preprocessing

Individual miRNA and miRNA cistron expression profiles for all samples were generated through barcoded small

RNA sequencing and sequence annotation. Following data preprocessing, sequencing was of sufficient quality for 81 (96%) of 84 samples; a median of 2,466,486 (range: 268,715–20,439,676) miRNA sequence reads, representing an average of 46% total sequence reads, per sample was obtained (Supplementary Table 4). miRNA content averaged 28.3 and 23.6 fmol per microgram total RNA per sample in discovery and validation sets, respectively; no significant differences in miRNA content were seen between GEP-NET types in either set (Kruskal–Wallis (K–W) test, $P > 0.3$).

Clinicopathologic characteristics of discovery and validation sets

Preprocessed sample profiles were assigned to discovery (80%) and validation (20%) sets. The clinical and pathologic characteristics of both sets were similar. No significant differences in age (Wilcoxon rank-sum test $P = 0.79$), gender (chi-square $P = 0.87$) or tumor grade (chi-square $P = 0.06$) were detected between sets. Similar proportions of GEP-NET types were present in each set; discovery and validation sets comprised 64 (pancreas 21 (33%), ileum 25 (39%), appendix 12 (19%), and rectum 6 (9%)) and 17 (pancreas 6 (35%), ileum 6 (35%), appendix 3 (18%), and rectum 2 (12%)) NETs, respectively. Relevant clinical and pathologic data for each sample are summarized in Table 1.

Data normalization and filtering

Following profile normalization and filtering, 263 and 253 individual miRNAs and 133 and 131 miRNA cistrons remained in our discovery and validation sets, respectively. The distribution of $\log_2$ normalized individual miRNA and miRNA cistron expression data for all samples is presented in Supplementary Fig. 1.

High expression analyses

Candidate miRNA markers for all four GEP-NET types were identified from the top 0.5% expressed individual miRNAs and miRNA cistrons in our discovery set and confirmed in our validation set (Table 2). miR-375 and cluster-mir-375 were the highest expressed individual miRNA and miRNA cistron in all GEP-NETs, with respective median expressions of 16.6% and 20.2% in each set. miRs-143, -21 and -7 were the next most abundant individual miRNAs and miRNA cistrons with median expressions ranging from 3.3 to 6.9%.

Table 1 Clinicopathologic characteristics of study samples.

Features	Tumor site							
	Ileum		Appendix		Rectum		Pancreas	
	Discovery (n = 25)	Validation (n = 6)	Discovery (n = 12)	Validation (n = 3)	Discovery (n = 6)	Validation (n = 2)	Discovery (n = 21)	Validation (n = 6)
Male:female	10:15	2:4	1:11	1:2	1:5	0:2	12:9	3:3
Age avg (min, max)	68 (47, 88)	67 (48, 82)	44 (12, 66)	44 (40, 50)	59 (47, 71)	61 (52, 70)	60 (27, 80)	60 (38, 76)
Low-grade	24 (96%)	6 (100%)	12 (100%)	3 (100%)	6 (100%)	2 (100%)	15 (71%)	2 (33%)
Intermediate-grade	1 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (29%)	4 (66%)

Basic demographic and tumor grading data are presented for the four pathological types of GEP-NET included in discovery and validation sets.

Discovery analyses

Candidate miRNA markers that discriminate GEP-NETs based on site of origin in the embryonic gut (comparisons A, B and C), anatomic site of origin (comparisons D, E, F and G) and tumor grade (comparison H) were identified from the top-ranking 3% individual miRNAs (Supplementary Table 5) and miRNA cistrons (Supplementary Table 6) in our discovery set. Comparisons A, B and C, D, E, F and G and H were respectively used for hierarchical classification, reference and assessing tumor grade below.

Hierarchical classifier

Among all available classifiers in the MATLAB Classification Learner App, a family of linear classifiers showed the highest classification accuracy (data not presented). From this family, we chose the linear Support Vector Machine algorithm as the predictor in our iterative algorithm and final hierarchical classifier (Fig. 1).

Using this algorithm, we constructed and assessed the accuracy of each decision point in our dual-layer classifier (Supplementary Table 7). In the first layer, miR-615 expression was significantly higher in midgut than non-midgut samples (K–W *P*-value<0.01); mir-92b provided additional discrimination (K–W *P*-value<0.01). When combined, these two miRNAs discriminated midgut and non-midgut NETs with an accuracy of 100% in the discovery set and 94.1% in the validation set, with only one sample misclassification (Fig. 2 and Supplementary Table 7). In the second layer, miR-125b expression was significantly lower in ileal than appendiceal NETs; miRs-192 and -149 provided additional discrimination (K–W *P*-value<0.01). In addition, miR-429 expression was significantly higher in rectal vs pancreatic NETs; miR-487b provided additional discrimination (K–W *P*-value<0.01). When combined, miRs-125b, -192 and -149 discriminated ileal and appendiceal NETs with 100% accuracy in the discovery and validation sets (Fig. 2 and Supplementary Table 7), and miRs-429 and -487b discriminated rectal and pancreatic NETs with 96.3% and 100% accuracy in the discovery and validation sets, respectively (Fig. 2 and Supplementary Table 7).

Once established, we determined the overall accuracy of our dual-layer hierarchical classifier on our pathologically verified samples. With one sample misclassification in each set, overall classifier accuracy was 98.5% in the discovery set and 94.4% in the validation set (Table 3). To better understand our classifier, we examined the expression of individual miRNAs used to classify each

Table 2 High expression analyses of preprocessed miRNA profiles.

Discovery set		Validation set	
miRNA	Median% of miRNA in all samples	miRNA	Median% of miRNA in all samples
miR-375	16.6	miR-375	20.2
miR-143	4.9	miR-143	5.3
miR-7	3.9	miR-21	3.9
miR-21	3.3	miR-7	3.2
miR-26a	2.0	miR-192	1.9
miR-125b	1.4	miR-200a	1.4
miR-192	1.4	miR-141	1.4
let-7a	1.3	miR-26a	1.4
miR-29a	1.3	miR-125b	1.2
miR-101	1.2	miR-194	1.2
miR-125a	1.1	miR-27b	1.2
miRNA cistron	Median% of miRNA cistron in all samples	miRNA cistron	Median% of miRNA cistron in all samples
cluster-mir-375(1)	16.6	cluster-mir-375(1)	20.2
cluster-mir-98(13)	6.9	cluster-mir-143(2)	5.7
cluster-mir-143(2)	5.4	cluster-mir-98(13)	5.5
cluster-mir-7-1(3)	3.9	cluster-mir-21(1)	3.9

The median expression of the top 0.5% of individual miRNAs and miRNA cistrons in all preprocessed GEP-NET miRNA profiles is presented in descending order for discovery and validation sets. Cluster information can be used to assess data quality; cluster-mir-375 and mir-21 are monocistronic; cluster-mir-143(2) comprises miRs-143 and -145; cluster-mir-7-1(3) comprises miRs-7-1, -7-2, and -7-3, and cluster-mir-98(13) comprises lets-7a-1, -2, -3, -7b, -7c, -7d, -7f-1, -7f-2, and miRs-98, -99a, -100, -125b-1, -125b-2.

pathological type; miRNA cistrons were also examined to assess data consistency (Supplementary Table 8).

Tumor grading

To evaluate miRNAs as adjunct markers for tumor grading, we performed feature selection using miRNA expression data from pancreatic NETs. Intriguingly, miR-328 expression discriminated low- and intermediate-grade pancreatic NETs (K-W P -value<0.01, Fig. 3) in our discovery set. Although we were unable to confirm our findings in the validation set, we noted one potentially misclassified sample (Fig. 3).

Comparison of RT-qPCR- and sequencing-based miR-615 expression

In a limited assessment of the degree of similarity between miRNA detection approaches, we confirmed that miR-615 qPCR expression is similar to $\log_2$ normalized miR-615 relative frequency in all samples and significantly higher in midgut than non-midgut NETs (Mann-Whitney U test=207, P <0.001, r =0.60, n =77, Supplementary Fig. 2); a moderate degree of correlation between sequencing- and amplification-based approaches was observed (Spearman rank correlation ρ =0.68, n =77, P <0.001).

Figure 1

Hierarchical classifier for discriminating GEP-NETs. Schematic diagram indicating classifier structure and miRNA determinants for discriminating midgut and non-midgut NETs and pathological types at each decision node. The classifier is based on comparisons (A, B and C) in the main text.

Figure 2

Scatter plot assessment of selected individual miRNAs for discriminating GEP-NETs. Selected individual miRNAs effectively discriminate GEP-NET subgroups and types. Midgut and non-midgut NET subgroups were discriminated based on miR-615 and miR-92b expression with no misclassification in the discovery set (A) and one misclassification (indicated by arrow) in the validation set (B). Ileum and appendix NET types were effectively discriminated by miR-192, -149 and -125b expression in discovery (C) and validation (D) sets. Rectum and pancreas NET types were discriminated by miR-429 and -487b expression with one misclassification (indicated by arrow) in the discovery set (E) and no misclassification in the validation set (F). Similar results were generated using relevant miRNA cistron data and are not presented. $\log_2$ RF, $\log_2$ normalized relative frequency. A full color version of this figure is available at <https://doi.org/10.1530/ERC-18-0244>.

Discussion

GEP-NET histologic evaluation is a key component of clinical management. Tumor grade, as determined by evaluation of Ki-67 immunohistochemical stains and mitotic counts, is currently employed to predict prognosis and determine surgical management (Benetatos *et al.* 2018). This scoring system is time-consuming and hampered by poor reproducibility (Reid *et al.* 2015). Furthermore, targeted

therapies are becoming available, but require knowledge of primary tumor site (Raymond *et al.* 2011, Yao *et al.* 2016a,b). Available immunohistochemical markers may identify potential primary sites, but display suboptimal sensitivity and specificity, even when used in panels and in combination with clinical data (Yang *et al.* 2017). To address this issue, we hypothesized that miRNAs are valuable adjunct tissue markers for GEP-NETs (Lu *et al.* 2005). Using a novel approach to biomarker discovery

Table 3 Overall accuracy of hierarchical classifier for discriminating GI-NETs.

	Established discovery set diagnosis				Established validation set diagnosis			
	Ileum	Appendix	Rectum	Pancreas	Ileum	Appendix	Rectum	Pancreas
Hierarchical classifier designation								
Ileum	25	0	0	0	6	0	0	0
Appendix	0	12	0	0	0	3	0	0
Rectum	0	0	6	0	0	0	2	0
Pancreas	0	0	1	20	1	0	0	5
Accuracy	63/64 (98.5%)				(16/17) 94.4%			

Using our hierarchical classifier, samples were assigned to one of four pathological types. Overall classifier accuracy was assessed by comparing these designations to established pathological diagnoses for the same samples in the discovery and validation sets.

Figure 3

Scatter plot assessment of select miRNA for discriminating low- and intermediate-grade pancreatic NETs. Low- and intermediate-grade pancreatic NETs are effectively discriminated based on miR-328 expression (A) and cistron-miR-328(1) expression (B) in the discovery set (left panels) but not in the equivalent individual miRNA (C) or miRNA cistron (D) plots for the validation set (right panels). The 'misclassified' midgut NET (indicated by arrows) was included by random in the validation set. Highly ranked miR-19a and cistron-mir-17(12) expression is presented for comparison only. log₂ RF, log₂ normalized relative frequency. A full color version of this figure is available at <https://doi.org/10.1530/ERC-18-0244>.

and validation, we identified miRNA markers that complement morphologic and immunohistochemistry-based histological evaluation.

The strength of our study stems from addressing analytical and post-analytical variables in miRNA clinical testing (Gustafson *et al.* 2016). For comprehensive miRNA detection, we used a barcoded small RNA sequencing assay that was carefully validated with a pool of 770 synthetic miRNAs and 45 calibrators oligoribonucleotides (Hafner *et al.* 2012). For accurate small RNA sequence annotation, we used a state-of-the-art annotation pipeline that enables comprehensive miRNA expression profiling and quantitation and simplifies analyses through organization of individual miRNAs into miRNA cistrons (Farazi *et al.* 2012, Hafner *et al.* 2012). Following sequencing, we introduced quality control measures to identify and remove outlier profiles. To identify the most powerful predictors and discriminators in our miRNA expression data, we focused on comparisons between GEP-NETs from different anatomic sites and used our novel feature selection algorithm (Ren *et al.* 2017) that combines five different feature-ranking methods (Statistics and Machine Learning Toolbox, MATLAB). Lastly, we validated our discovery markers in an independent sample set using barcoded small RNA sequencing; in our experience, other methods, such as miRNA real-time PCR, are expensive and insufficiently comprehensive to validate multidimensional miRNA profiles (Git *et al.* 2010).

High expression analyses showed that miR-375 and cluster-mir-375 are the most abundant individual miRNA and miRNA cistron in all our samples, indicating its functional importance and potential as a GEP-NET marker. Currently, miR-375 is thought to be an endocrine gland-specific miRNA (Landgraf *et al.* 2007) with regulatory roles, where known, in pancreatic beta cell development and differentiation, proliferation and regulation of insulin secretion (Eliasson 2017). Intriguingly, miR-375 also has a tumor suppressor role in many cancers (Yan *et al.* 2014). Based on our and similar miRNA expression data from small intestinal NETs in which miR-375 is downregulated in metastasis (Arvidsson *et al.* 2018), we propose that this miRNA is an excellent marker and potential tumor suppressor in GEP-NETs; miR-375 is known to target oncogenes such as *YAP1* (Nishikawa *et al.* 2011). miRs-143, -21 and -7 were also highly expressed in our analyses; miR-143 is enriched in smooth-muscle and likely originates from peritumoral tissue, miR-21 is ubiquitously expressed and often raised in cancer, but miR-7 may also be a GEP-NET marker of unknown function (Miller *et al.* 2016). Further profiling, localization and functional studies are required to understand the gene regulatory roles of these miRNAs.

Data mining analyses demonstrated that GEP-NET pathological types can be accurately classified based on miRNA expression. We constructed and successfully validated a dual-layer hierarchical classifier that first

discriminates midgut from non-midgut NETs based on miR-615 and -92b expression prior to discrimination of ileal from appendiceal NETs based on miR-125b, -192 and -149 expression, and rectal from pancreatic NETs based on miR-429 and -487b expression. In the discovery set, one pancreatic NET was misclassified as a rectal NET; classifier accuracy is expected to improve as more samples are sequenced. In the validation set, one non-midgut NET was misclassified as a midgut NET and ultimately an ileal NET. Upon chart review, this tumor was located in the pancreatic head with full thickness invasion of the duodenal wall with liver and lymph node metastases. That the site of origin is pancreatic rather than duodenal or ileal remains an open question in this case.

Data mining analyses also showed that low- and intermediate-grade pancreatic NETs can potentially be discriminated based on miRNA expression. Of particular note, miR-328 is significantly lower expressed in intermediate-grade tumors. Although we were unable to validate our findings, this may be due to the small size of our validation set and the inclusion of the 'misclassified' midgut NET from above. As above, classifier accuracy is expected to improve as more samples are sequenced. Whether miR-328 would be a useful adjunct marker for assessing high-grade pancreatic NETs or for grading NETs in other anatomic sites is being explored.

Direct comparison of our miRNA expression data with those generated through less comprehensive study designs and different detection methodologies is challenging (Malczewska *et al.* 2018). In addition, many candidate miRNA tissue markers, such as miRs-1 and -133, -103 and -107, -10b and -155, -216a, -216b and -217, and -10b and -155, may simply reflect the amount of muscle, fat, pancreatic tissue and/or hematopoietic elements present in the input materials (Szafranska *et al.* 2007). Nonetheless, we agree with existing studies that miRNAs have biomarker potential (Roldo *et al.* 2006) and that some miRNAs, such as miR-196a in pancreatic NETs (Li *et al.* 2013, Lee *et al.* 2015), are likely markers of disease progression. Comparisons of results generated through different miRNA detection methodologies, exemplified here by the moderate degree of correlation between sequencing- and amplification-based miR-615 expression, are now required to move miRNA testing into clinical practice.

We believe that this is the first study in the GEP-NET field to assess miRNA cistron expression data, simplifying analyses and providing insights into important primary transcripts from which multiple miRNAs are cleaved. High expression analyses for miRNA cistrons are the

same as above for individual miRNAs because each of these miRNAs is monocistronic. Discriminant analyses using miRNA cistrons indicate that some cistrons, such as cluster-mir-196a (3) and cluster-mir-134 (43), are abundant and differentially expressed between NET subgroups or types. Cluster-mir-196a comprises three miRNAs (miRs-196a-1, 196a-2 and -615), arising from introns in the *HOXC* cluster on chromosome 12q and is more highly expressed in midgut than non-midgut NETs, whereas cluster-mir-134 comprises 34 miRNAs (not listed) arising from an intergenic region on chromosome 14 and is more highly expressed in pancreatic than other NETs. Teasing apart the functions of individual and combinations of miRNAs from those of their primary transcripts will be challenging. Nonetheless, organizing individual miRNAs into miRNA cistrons provides valuable insights into data quality.

As with most biomarker studies on rare tumors, our study has limitations. Here, we focused on higher prevalence GEP-NET types and did not have the samples to study lower prevalence GEP-NET types, such as duodenal or gastric NETs or high-grade GEP-NETs. Due to low sample numbers, we were unable to evaluate miRNA-based grading for ileal, appendiceal and rectal NETs or to validate miR-328-based grading in pancreatic NETs. Due to the lack of survival data and relevant samples, we were unable to confirm the recent intriguing finding that miR-375 downregulation in small intestinal NETs is associated with shorter patient survival (Arvidsson *et al.* 2018). Despite these limitations, we provide a reliable approach and valuable insights into sequencing-based miRNA marker discovery and validation.

We have developed and validated a dual-layer hierarchical classifier for classifying GEP-NETs based on miRNA expression, identified a candidate miRNA marker for discriminating low- and intermediate-grade pancreatic NETs and provided tissue miRNA profiles to stimulate further research including as reference for liquid biopsy studies. Using advanced miRNA detection, annotation and data mining techniques, we provide a reliable approach for evaluating GEP-NETs. Further investigations will include sequencing-based tissue and plasma miRNA profiling of GEP-NETs of differing grade and from different anatomic sites, and functional characterization of relevant miRNAs in neuroendocrine tumorigenesis.

Supplementary data

This is linked to the online version of the paper at <https://doi.org/10.1530/ERC-18-0244>.

Declaration of interest

M K has received honoraria from Novartis. T T is co-founder of Alnylam Pharmaceuticals and is on the Scientific Advisory Board of Regulus Therapeutics.

Funding

This work was supported through the Southeastern Ontario Academic Medical Organization Innovation Fund, the Canada Foundation for Innovation John R Evans Leaders Fund and the Ontario Research Fund-Research Infrastructure.

Author contribution statement

N P, K T, Y-T C and N R are responsible for study conception and design. N P, A M, X Y, T S, M K, K B and N R contributed to collection and assembly of data. N P, K T, J W, Y-T C, T T and N R contributed to data analysis and interpretation. N P, K T, A M, J W, T S, M K, K B and N R contributed to manuscript writing. All authors contributed to final approval of manuscript. All authors agree to be accountable for all aspects of the work, including ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors wish to thank Ms Sarah Maritan for her valuable comments.

References

- Acunzo M, Romano G, Wernicke D & Croce CM 2015 MicroRNA and cancer – a brief overview. *Advances in Biological Regulation* **57** 1–9. (<https://doi.org/10.1016/j.bior.2014.09.013>)
- Arvidsson Y, Rehammar A, Bergstrom A, Andersson E, Altiparmak G, Sward C, Wangberg B, Kristiansson E & Nilsson O 2018 miRNA profiling of small intestinal neuroendocrine tumors defines novel molecular subtypes and identifies miR-375 as a biomarker of patient survival. *Modern Pathology* [epub]. (<https://doi.org/10.1038/s41379-018-0010-1>)
- Bellizzi AM 2013 Assigning site of origin in metastatic neuroendocrine neoplasms: a clinically significant application of diagnostic immunohistochemistry. *Advances in Anatomic Pathology* **20** 285–314. (<https://doi.org/10.1097/PAP.0b013e3182a2dc67>)
- Benetatos N, Hodson J, Marudanayagam R, Sutcliffe RP, Isaac JR, Ayuk J, Shah T & Roberts KJ 2018 Prognostic factors and survival after surgical resection of pancreatic neuroendocrine tumor with validation of established and modified staging systems. *Hepatobiliary and Pancreatic Diseases International* **17** 169–175. (<https://doi.org/10.1016/j.hbpd.2018.03.002>)
- Berindan-Neagoe I, Monroig Pdel C, Pasculli B & Calin GA 2014 MicroRNAome genome: a treasure for cancer diagnosis and therapy. *CA: A Cancer Journal for Clinicians* **64** 311–336. (<https://doi.org/10.3322/caac.21244>)
- Bosman F, Carneiro F, Hruban R & Theise N 2010 *WHO Classification of Tumours of the Digestive System*. Lyon, France: IARC Press.
- Brown M, Suryawanshi H, Hafner M, Farazi TA & Tuschl T 2013 Mammalian miRNA curation through next-generation sequencing. *Frontiers in Genetics* **4** 145. (<https://doi.org/10.3389/fgene.2013.00145>)
- Duda RO, Hart PE & Stork DG 2001 *Pattern Classification*. New York, NY: Wiley.
- Eliasson L 2017 The small RNA miR-375 – a pancreatic islet abundant miRNA with multiple roles in endocrine beta cell function. *Molecular and Cellular Endocrinology* **456** 95–101. (<https://doi.org/10.1016/j.mce.2017.02.043>)
- Farazi TA, Brown M, Morozov P, Ten Hoeve JJ, Ben-Dov IZ, Hovestadt V, Hafner M, Renwick N, Mihailovic A, Wessels LF, *et al.* 2012 Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. *Methods* **58** 171–187. (<https://doi.org/10.1016/j.ymeth.2012.07.020>)
- Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, Bertone P & Caldas C 2010 Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* **16** 991–1006. (<https://doi.org/10.1261/rna.1947110>)
- Gustafson D, Tyryshkin K & Renwick N 2016 microRNA-guided diagnostics in clinical samples. *Best Practice and Research: Clinical Endocrinology and Metabolism* **30** 563–575. (<https://doi.org/10.1016/j.beem.2016.07.002>)
- Hafner M, Renwick N, Farazi TA, Mihailovic A, Pena JT & Tuschl T 2012 Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* **58** 164–170. (<https://doi.org/10.1016/j.ymeth.2012.07.030>)
- Hoaglin DC & Iglewicz B 1987 Fine-tuning some resistant rules for outlier labeling. *JAMA* **82** 1147–1149. (<https://doi.org/10.2307/2289392>)
- Klimstra DS 2016 Pathologic classification of neuroendocrine neoplasms. *Hematology/Oncology Clinics of North America* **30** 1–19. (<https://doi.org/10.1016/j.hoc.2015.08.005>)
- Koo J, Mertens RB, Mirocha JM, Wang HL & Dhall D 2012 Value of islet 1 and PAX8 in identifying metastatic neuroendocrine tumors of pancreatic origin. *Modern Pathology* **25** 893–901. (<https://doi.org/10.1038/modpathol.2012.34>)
- Koo J, Zhou X, Moschiano E, De Peralta-Venturina M, Mertens RB & Dhall D 2013 The immunohistochemical expression of islet 1 and PAX8 by rectal neuroendocrine tumors should be taken into account in the differential diagnosis of metastatic neuroendocrine tumors of unknown primary origin. *Endocrine Pathology* **24** 184–190. (<https://doi.org/10.1007/s12022-013-9264-9>)
- Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, *et al.* 2007 A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129** 1401–1414. (<https://doi.org/10.1016/j.cell.2007.04.040>)
- Lawrence B, Gustafsson BI, Chan A, Svejda B, Kidd M & Modlin IM 2011 The epidemiology of gastroenteropancreatic neuroendocrine tumors. *Endocrinology and Metabolism Clinics of North America* **40** 1–18. (<https://doi.org/10.1016/j.ecl.2010.12.005>)
- Lee YS, Kim H, Kim HW, Lee JC, Paik KH, Kang J, Kim J, Yoon YS, Han HS, Sohn I, *et al.* 2015 High expression of microRNA-196a indicates poor prognosis in resected pancreatic neuroendocrine tumor. *Medicine* **94** e2224. (<https://doi.org/10.1097/MD.0000000000002224>)
- Li SC, Essaghiri A, Martijn C, Lloyd RV, Demoulin JB, Oberg K & Giandomenico V 2013 Global microRNA profiling of well-differentiated small intestinal neuroendocrine tumors. *Modern Pathology* **26** 685–696. (<https://doi.org/10.1038/modpathol.2012.216>)
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, *et al.* 2005 MicroRNA expression profiles classify human cancers. *Nature* **435** 834–838. (<https://doi.org/10.1038/nature03702>)
- Malczewska A, Kidd M, Matar S, Kos-Kudla B & Modlin IM 2018 A comprehensive assessment of the role of miRNAs as biomarkers in gastroenteropancreatic neuroendocrine tumors. *Neuroendocrinology* **107** 73–90. (<https://doi.org/10.1159/000487326>)
- Mandal R, Hardin H, Baus R, Rehauer W & Lloyd RV 2017 Analysis of miR-96 and miR-133a Expression in Gastrointestinal Neuroendocrine Neoplasms. *Endocrine Pathology* **28** 345–350. (<https://doi.org/10.1007/s12022-017-9504-5>)
- Miller HC, Frampton AE, Malczewska A, Ottaviani S, Stronach EA, Flora R, Kaemmerer D, Schwach G, Pfragner R, Faiz O, *et al.* 2016

- MicroRNAs associated with small bowel neuroendocrine tumours and their metastases. *Endocrine-Related Cancer* **23** 711–726. (<https://doi.org/10.1530/ERC-16-0044>)
- Mitsuhashi K, Yamamoto I, Kurihara H, Kanno S, Ito M, Igarashi H, Ishigami K, Sukawa Y, Tachibana M, Takahashi H, *et al.* 2015 Analysis of the molecular features of rectal carcinoid tumors to identify new biomarkers that predict biological malignancy. *Oncotarget* **6** 22114–22125. (<https://doi.org/10.18632/oncotarget.4294>)
- Modlin IM, Lye KD & Kidd M 2003 A 5-decade analysis of 13,715 carcinoid tumors. *Cancer* **97** 934–959. (<https://doi.org/10.1002/cncr.11105>)
- Modlin IM, Bodei L & Kidd M 2016 Neuroendocrine tumor biomarkers: from monoanalytes to transcripts and algorithms. *Best Practice and Research: Clinical Endocrinology and Metabolism* **30** 59–77. (<https://doi.org/10.1016/j.beem.2016.01.002>)
- Nishikawa E, Osada H, Okazaki Y, Arima C, Tomida S, Tatematsu Y, Taguchi A, Shimada Y, Yanagisawa K, Yatabe Y, *et al.* 2011 miR-375 is activated by ASH1 and inhibits YAP1 in a lineage-dependent manner in lung cancer. *Cancer Research* **71** 6165–6173. (<https://doi.org/10.1158/0008-5472.CAN-11-1020>)
- Raymond E, Dahan L, Raoul JL, Bang YJ, Borbath I, Lombard-Bohas C, Valle J, Metrakos P, Smith D, Vinik A, *et al.* 2011 Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *New England Journal of Medicine* **364** 501–513. (<https://doi.org/10.1056/NEJMoa1003825>)
- Reid MD, Bagci P, Ohike N, Saka B, Erbarut Seven I, Dursun N, Balci S, Gucer H, Jang KT, Tajiri T, *et al.* 2015 Calculation of the Ki67 index in pancreatic neuroendocrine tumors: a comparative analysis of four counting methodologies. *Modern Pathology* **28** 686–694. (<https://doi.org/10.1038/modpathol.2014.156>)
- Ren R, Tyryshkin K, Graham CH, Koti M & Siemens DR 2017 Comprehensive immune transcriptomic analysis in bladder cancer reveals subtype specific immune gene expression patterns of prognostic relevance. *Oncotarget* **8** 70982–71001. (<https://doi.org/10.18632/oncotarget.20237>)
- Roldo C, Missiaglia E, Hagan JP, Falconi M, Capelli P, Bersani S, Calin GA, Volinia S, Liu CG, Scarpa A, *et al.* 2006 MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *Journal of Clinical Oncology* **24** 4677–4684. (<https://doi.org/10.1200/JCO.2005.05.5194>)
- Ruebel K, Leontovich AA, Stilling GA, Zhang SY, Righi A, Jin L & Lloyd RV 2010 MicroRNA expression in ileal carcinoid tumors: downregulation of microRNA-133a with tumor progression. *Modern Pathology* **23** 367–375. (<https://doi.org/10.1038/modpathol.2009.161>)
- Schmittgen TD & Livak KJ 2008 Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols* **3** 1101–1108. (<https://doi.org/10.1038/nprot.2008.73>)
- Singh S, Asa SL, Dey C, Kennecke H, Laidley D, Law C, Asmis T, Chan D, Ezzat S, Goodwin R, *et al.* 2016 Diagnosis and management of gastrointestinal neuroendocrine tumors: an evidence-based Canadian consensus. *Cancer Treatment Reviews* **47** 32–45. (<https://doi.org/10.1016/j.ctrv.2016.05.003>)
- Szafranska AE, Davison TS, John J, Cannon T, Sipos B, Maghnouj A, Labourier E & Hahn SA 2007 MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. *Oncogene* **26** 4442–4452. (<https://doi.org/10.1038/sj.onc.1210228>)
- Tang LH, Gonen M, Hedvat C, Modlin IM & Klimstra DS 2012 Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods. *American Journal of Surgical Pathology* **36** 1761–1770. (<https://doi.org/10.1097/PAS.0b013e318263207c>)
- Thorns C, Schurmann C, Gebauer N, Wallaschofski H, Kumpers C, Bernard V, Feller AC, Keck T, Habermann JK, Begum N, *et al.* 2014 Global microRNA profiling of pancreatic neuroendocrine neoplasias. *Anticancer Research* **34** 2249–2254.
- Wang SC, Parekh JR, Zuraek MB, Venook AP, Bergsland EK, Warren RS & Nakakura EK 2010 Identification of unknown primary tumors in patients with neuroendocrine liver metastases. *Archives of Surgery* **145** 276–280. (<https://doi.org/10.1001/archsurg.2010.10>)
- Williams ED & Sandler M 1963 The classification of carcinoid tumours. *Lancet* **1** 238–239. ([https://doi.org/10.1016/S0140-6736\(63\)90951-6](https://doi.org/10.1016/S0140-6736(63)90951-6))
- Yan JW, Lin JS & He XX 2014 The emerging role of miR-375 in cancer. *International Journal of Cancer* **135** 1011–1018. (<https://doi.org/10.1002/ijc.28563>)
- Yang Z, Klimstra DS, Hruban RH & Tang LH 2017 Immunohistochemical characterization of the origins of metastatic well-differentiated neuroendocrine tumors to the liver. *American Journal of Surgical Pathology* **41** 915–922. (<https://doi.org/10.1097/PAS.0000000000000876>)
- Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, Abdalla EK, Fleming JB, Vauthey JN, Rashid A, *et al.* 2008 One hundred years after ‘carcinoid’: epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *Journal of Clinical Oncology* **26** 3063–3072. (<https://doi.org/10.1200/JCO.2007.15.4377>)
- Yao JC, Fazio N, Singh S, Buzzoni R, Carnaghi C, Wolin E, Tomasek J, Raderer M, Lahner H, Voi M, *et al.* 2016a Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *Lancet* **387** 968–977. ([https://doi.org/10.1016/S0140-6736\(15\)00817-X](https://doi.org/10.1016/S0140-6736(15)00817-X))
- Yao JC, Pavel M, Lombard-Bohas C, Van Cutsem E, Voi M, Brandt U, He W, Chen D, Capdevila J, de Vries EGE, *et al.* 2016b Everolimus for the treatment of advanced pancreatic neuroendocrine tumors: overall survival and circulating biomarkers from the randomized, phase III RADIANT-3 study. *Journal of Clinical Oncology* **34** 3906–3913. (<https://doi.org/10.1200/JCO.2016.68.0702>)

Received in final form 13 July 2018

Accepted 17 July 2018

Accepted Preprint published online 18 July 2018