

Repertoire of microRNAs in Epithelial Ovarian Cancer as Determined by Next Generation Sequencing of Small RNA cDNA Libraries

Stacia K. Wyman², Rachael K. Parkin², Patrick S. Mitchell², Brian R. Fritz^{2*}, Kathy O'Briant³, Andrew K. Godwin⁴, Nicole Urban³, Charles W. Drescher³, Beatrice S. Knudsen³, Muneesh Tewari^{1,2*}

1 Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **2** Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **3** Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **4** Department of Medical Oncology, Fox Chase Cancer Center, Philadelphia, Pennsylvania, United States of America

Abstract

Background: MicroRNAs (miRNAs) are small regulatory RNAs that are implicated in cancer pathogenesis and have recently shown promise as blood-based biomarkers for cancer detection. Epithelial ovarian cancer is a deadly disease for which improved outcomes could be achieved by successful early detection and enhanced understanding of molecular pathogenesis that leads to improved therapies. A critical step toward these goals is to establish a comprehensive view of miRNAs expressed in epithelial ovarian cancer tissues as well as in normal ovarian surface epithelial cells.

Methodology: We used massively parallel pyrosequencing (i.e., “454 sequencing”) to discover and characterize novel and known miRNAs expressed in primary cultures of normal human ovarian surface epithelium (HOSE) and in tissue from three of the most common histotypes of ovarian cancer. Deep sequencing of small RNA cDNA libraries derived from normal HOSE and ovarian cancer samples yielded a total of 738,710 high-quality sequence reads, generating comprehensive digital profiles of miRNA expression. Expression profiles for 498 previously annotated miRNAs were delineated and we discovered six novel miRNAs and 39 candidate miRNAs. A set of 124 miRNAs was differentially expressed in normal versus cancer samples and 38 miRNAs were differentially expressed across histologic subtypes of ovarian cancer. Taqman qRT-PCR performed on a subset of miRNAs confirmed results of the sequencing-based study.

Conclusions: This report expands the body of miRNAs known to be expressed in epithelial ovarian cancer and provides a useful resource for future studies of the role of miRNAs in the pathogenesis and early detection of ovarian cancer.

Citation: Wyman SK, Parkin RK, Mitchell PS, Fritz BR, O'Briant K, et al. (2009) Repertoire of microRNAs in Epithelial Ovarian Cancer as Determined by Next Generation Sequencing of Small RNA cDNA Libraries. PLoS ONE 4(4): e5311. doi:10.1371/journal.pone.0005311

Editor: Sudhansu Kumar Dey, Cincinnati Children's Research Foundation, United States of America

Received: November 23, 2008; **Accepted:** February 7, 2009; **Published:** April 23, 2009

Copyright: © 2009 Wyman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The cost of sequencing was partially defrayed by Roche. The study was supported in part by the Ovarian Cancer SPORE at Fox Chase Cancer Center (P50 CA083638), the Pacific Ovarian Cancer Research Consortium Ovarian SPORE (P50 CA83636 to N.U. and M.T.), Chromosome Metabolism Training Grant 5 T32 CA09657-16 (to S.K.W.), a fellowship from the Merck Research Laboratories (S.K.W.) and by Fred Hutchinson Cancer Research Center and Canary Foundation funding (New Development funds to M.T.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: M. Tewari is a paid member of the Scientific Advisory Board of Combimatrix, Inc.

* E-mail: mtewari@fhccr.org

‡ Current address: Illumina, Inc., San Diego, California, United States of America

Introduction

Epithelial ovarian cancer is the leading cause of gynecologic cancer-related deaths in the United States [1], with late-stage diagnoses having a <30% five-year survival rate [2]. Survival rates could be improved by a better understanding of molecular pathogenesis, which may lead to development of superior targeted therapies, as well as by earlier detection of disease at a surgically curable stage. When detected at a stage in which disease is confined to the ovary, for example, the five-year survival rate increases to >80%. Clinically effective biomarkers for early detection of ovarian cancer could substantially improve survival rates and ovarian cancer biomarker discovery is an important area of ongoing research [3].

MicroRNAs (miRNAs) are a class of small (~22 nt) non-coding RNA molecules that act post-transcriptionally to regulate gene expression [4]. MicroRNAs originate from hairpin RNA precursors that are processed to generate both the functional mature miRNA and a miRNA “star form” of similar length derived from the opposite strand of the hairpin. MicroRNA-mediated modulation of biological systems has been found to be perturbed in multiple diseases [5], including cancer [6–8]. Expression patterns of miRNAs correlate with tissue of origin [8,9], prognosis [10,11] and with clinical cancer behaviors [12], making miRNAs valuable tissue-based biomarkers. Furthermore, we and others have recently shown that tumor-derived miRNAs enter the bloodstream at measurable levels, indicating that miRNAs released by tumor tissue represent a powerful new class of blood-based, minimally

invasive biomarkers for cancer detection [13–16]. As such, there is a strong impetus for a comprehensive analysis of the miRNA repertoire expressed in epithelial ovarian cancer.

Several recent reports have begun to characterize miRNA expression in ovarian cancer using microarrays spotted with probes for a varying number of known miRNAs [17–20]. Although these are pioneering studies, they also have limitations. The foremost of these is that all arrays reported to date have incomplete coverage of known miRNAs. Of the 959 miRNAs and star forms present in miRBase (release 12.0) [21,22], most microarray platforms applied to date assay only a few hundred miRNAs, with the most comprehensive of the arrays assaying 470 miRNAs and star forms. In addition, microarray approaches only measure previously identified miRNAs and are not equipped to discover and profile novel miRNAs. This is an important distinction to make given that several reports suggest that a significant number of miRNAs, particularly those that may be cell-type specific in expression, remain to be discovered [23,24]. Finally, because of the small size of miRNAs and the challenges of achieving uniform hybridization conditions across an entire miRNA microarray, there is potential for cross-hybridization of miRNAs that are highly related in sequence, making it sometimes impossible to definitively distinguish between miRNAs differing by one or two nucleotides.

To overcome the limitations associated with microarray studies, we used “next generation” sequencing technology (specifically massively parallel pyrosequencing using the 454 Life Sciences platform) of small RNA cDNA libraries to more comprehensively characterize miRNA expression in epithelial ovarian cancer, as well as in primary cultures of normal ovarian surface epithelial cells. This approach, which is based on 5′ and 3′ universal linker ligation/RT-PCR and counting of sequence reads obtained for distinct miRNAs, can in theory capture all miRNAs present in the RNA samples (including ones of novel sequence) and permits precise identification of miRNAs, even when differing by only a single nucleotide. We generated a total of 738,710 sequence reads from four small RNA cDNA libraries representing normal primary human ovarian surface epithelium (HOSE) cultures and three common epithelial ovarian cancer subtypes (serous, clear cell and endometrioid). Here we describe the results of this effort, including the identification and profiling of both previously annotated and novel miRNAs expressed in ovarian cancer.

Methods

Additional details regarding cell culture, RNA isolation, small RNA cDNA library construction, massively parallel sequencing, computational pipeline description and miRNA TaqMan quantitative reverse-transcription polymerase chain reaction (qRT-PCR) are provided in detail in **Supplementary Methods S1** and in [23].

Cell culture and clinical materials

Normal primary human ovarian surface epithelial (HOSE) cells were obtained from the normal ovaries of postmenopausal women using a modification of the technique described previously in [25]. In all cases, specimens were taken from normal-appearing ovarian surface epithelium, which was confirmed following histopathological review. HOSE specimens were obtained under a protocol that was approved by the Research Review Committee and Internal Review Board at the Fox Chase Cancer Center from patients undergoing clinically indicated surgery (either prophylactic risk reduction salpingo-oophorectomy or total abdominal hysterectomy with bilateral salpingo-oophorectomy) and who provided

written informed consent for tissue not required for diagnosis to be used for research. Primary HOSE cells were cultured using media described in [26] at 37°C in 5% CO₂.

Ovarian cancer snap-frozen specimens corresponding to Stage III/IV epithelial ovarian cancer (19 serous, four clear cell and 10 endometrioid) contained >70% malignant epithelial cellularity as assessed by pathologist review of an adjacent stained tissue section. Ovarian cancer specimens were obtained from the Pacific Ovarian Cancer Research Consortium Repository. Specimens were collected under a protocol approved by the Fred Hutchinson Cancer Research Center Institutional Review Board and were obtained from patients who had provided written informed consent for tissue not required for clinical management purposes to be used for research.

RNA extraction and miRNA library preparation

Normal primary HOSE cultures (passage 4–7) were lysed in 600 µl mirVana Lysis/Binding buffer (Ambion, Inc., Austin, TX). Tumor specimens (~100 mg each) were homogenized in a Qjagen TissueLyser with 600 µl mirVana Lysis/Binding buffer. Total RNA was extracted from both sample types using the mirVana™ miRNA isolation kit (Ambion) per the manufacturer’s protocol. RNA from snap-frozen tumors (19 serous, four clear cell and 10 endometrioid tumors, respectively) was pooled according to histologic subtype. MicroRNA cloning was performed as previously described [27], (<http://web.wi.mit.edu/bartel/pub/protocols/miRNACloningUpdate0705.pdf>) with modifications at the amplification steps to prepare the samples for massively parallel sequencing by 454 Life Sciences.

Custom bioinformatic data analysis pipeline

Processing and annotation of sequences based on identity to known transcribed RNAs or as novel miRNAs was performed using a custom bioinformatics pipeline described in detail in **Supplementary Methods S1**.

Measurement of miRNA levels in RNA from clinical samples using TaqMan® qRT-PCR assays

Total RNA was reverse transcribed and prepared for qRT-PCR using the TaqMan® miRNA Reverse Transcription Kit and miRNA-specific stem-loop primers (Applied BioSystems, Inc.) as previously described [23]. Data was analyzed with SDS Relative Quantification Software version 2.2.2 (Applied BioSystems, Inc.), with the automatic Ct setting for assigning baseline and threshold for Ct determination.

Results

Overview of small RNA cDNA libraries and 454 sequencing

Small RNA cDNA libraries were constructed from RNA sample pools isolated from primary cultures of histologically normal primary HOSE ($n = 4$) and epithelial ovarian cancer specimens of serous (OSC, $n = 19$), clear cell (OCC, $n = 4$) and endometrioid (OEC, $n = 10$) histologies. Tumor specimens were selected to contain at least 70% malignant epithelial cells. Total RNA was size-fractionated via polyacrylamide gel electrophoresis and small RNAs corresponding in molecular weight to the mature miRNA population (18–24 nt fraction) were gel extracted and processed for reverse transcription and PCR amplification to create cDNA libraries (**Figure 1A**). Massively parallel pyrosequencing using the 454 Life Sciences’ platform generated 80,501 normal HOSE reads, 273,739 OSC reads, 292,942 OCC reads, and 91,528 OEC

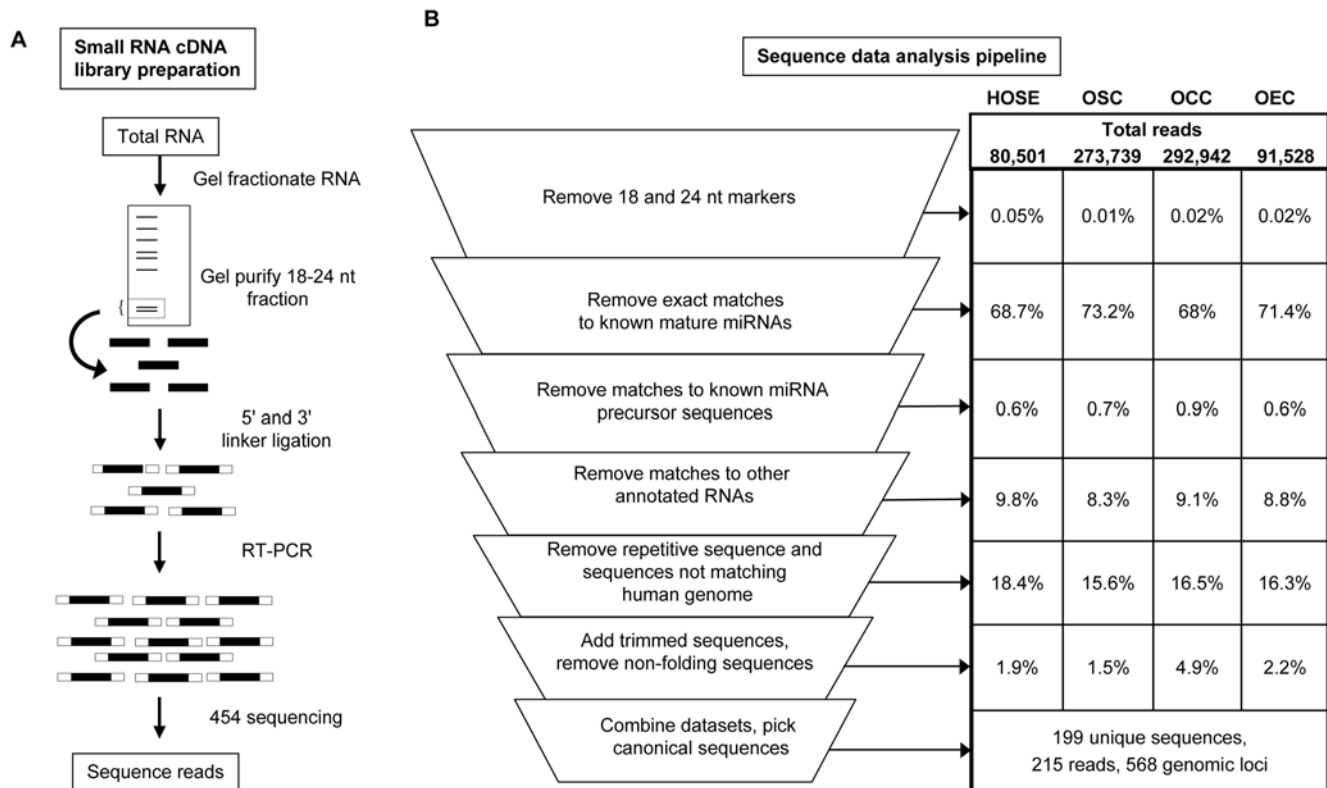


Figure 1. Small RNA cDNA library generation and data analysis pipeline. A. Small RNAs were isolated from normal primary HOSE cultures and from serous (OSC), clear cell (OCC) and endometrioid (OEC) ovarian cancer tissues. Following 5' and 3' linker ligation, RT-PCR was performed to generate four independent cDNA libraries of small RNAs that were then used as templates for massively parallel pyrosequencing (454 sequencing). B. **Description of the steps in the data analysis pipeline.** The initial step of the data analysis was removal of sequences corresponding to 18 nt and 24 nt RNA markers that had been spiked into the total RNA prior to gel electrophoresis. Percentage of total reads from HOSE, OSC, OCC and OEC datasets classified into the designated categories and filtered out at each step are listed in the table on the right. At the bottom of the table, the number of "unique sequences" represents the non-redundant sequences derived after collapsing multiple reads of identical sequence. Some canonical sequences map to more than one locus in the genome. At completion of the analysis, 199 sequences totaling 215 reads and mapping to 568 loci met our criteria for canonical hairpin-derived sequences from the combined datasets. doi:10.1371/journal.pone.0005311.g001

reads (Figure 1B). These corresponded to 10,544 HOSE, 26,849 OSC, 37,792 OCC and 13,134 OEC non-redundant sequences. The greater number of reads corresponding to serous and clear cell samples was due to greater depth of sequencing (i.e., a larger portion of the sequencing plate used) rather than to any particular variations during library preparation.

Sequence data was processed using our custom computational pipeline (Figures 1B, S1 and Table S1). The initial steps of the pipeline identified sequence matches to databases of previously annotated RNAs (e.g., known miRNAs, other noncoding RNAs, messenger RNAs) and to highly repetitive sequence elements. Remaining sequences were processed to identify novel miRNAs (Figure S1; details of the computational pipeline are provided in Supplementary Methods S1). In the next sections, we discuss in detail the sequences matching to known miRNAs and the identification of novel miRNAs.

MicroRNA profiling: previously annotated miRNAs. Matches to known miRNAs in miRBase (release 12.0) [21,22] represented 68–73% of the reads, depending on the dataset (Figure 1B and Table S1) and corresponded to a total of 498 known miRNAs (including star forms). The cloning frequency of individual miRNAs, expressed as a fraction of total reads obtained from a given sample, can be used to compare relative expression of miRNAs between samples [28–30]. The global summary of our 454 sequencing datasets is presented in Figure 2, which depicts

the relative expression of miRNAs between normal primary HOSE and cancer tissue datasets (A), as well as between the three epithelial ovarian subtypes (B). The most differentially expressed miRNAs (i.e., ≥ 4 fold-change) are shown as red dots. It is notable, though not unexpected, that the ovarian cancer histologic subtype miRNA profiles were more different from normal HOSE (A) than they were from each other (B). Tabulated 454 sequencing data for miRNA abundance in all datasets, as well as miRNA names and abundance ratios for all pairwise comparisons between sample types, are provided in their entirety in Table S2.

The overlap of differentially expressed miRNAs between subtypes is visualized in the Venn diagrams in Figure 3, which also incorporate miRNAs excluded from Figure 2 because of zero reads in either normal HOSE or ovarian cancer samples. From the intersection of the three groups in the Venn diagrams, it is evident that in many cases miRNAs differentially expressed in ovarian cancer relative to normal HOSE show the same pattern of differential expression regardless of histologic subtype. Table S3 provides a full list of names of individual miRNAs and expression ratio and *P*-values for the comparisons summarized in Figure 3.

We directly compared miRNA expression between ovarian cancer histologic subtypes by deriving expression ratios for each miRNA based on fractional abundance values. Significance was assessed using Fisher's exact test. The top ten miRNAs differentially expressed in each of the pairwise comparisons of

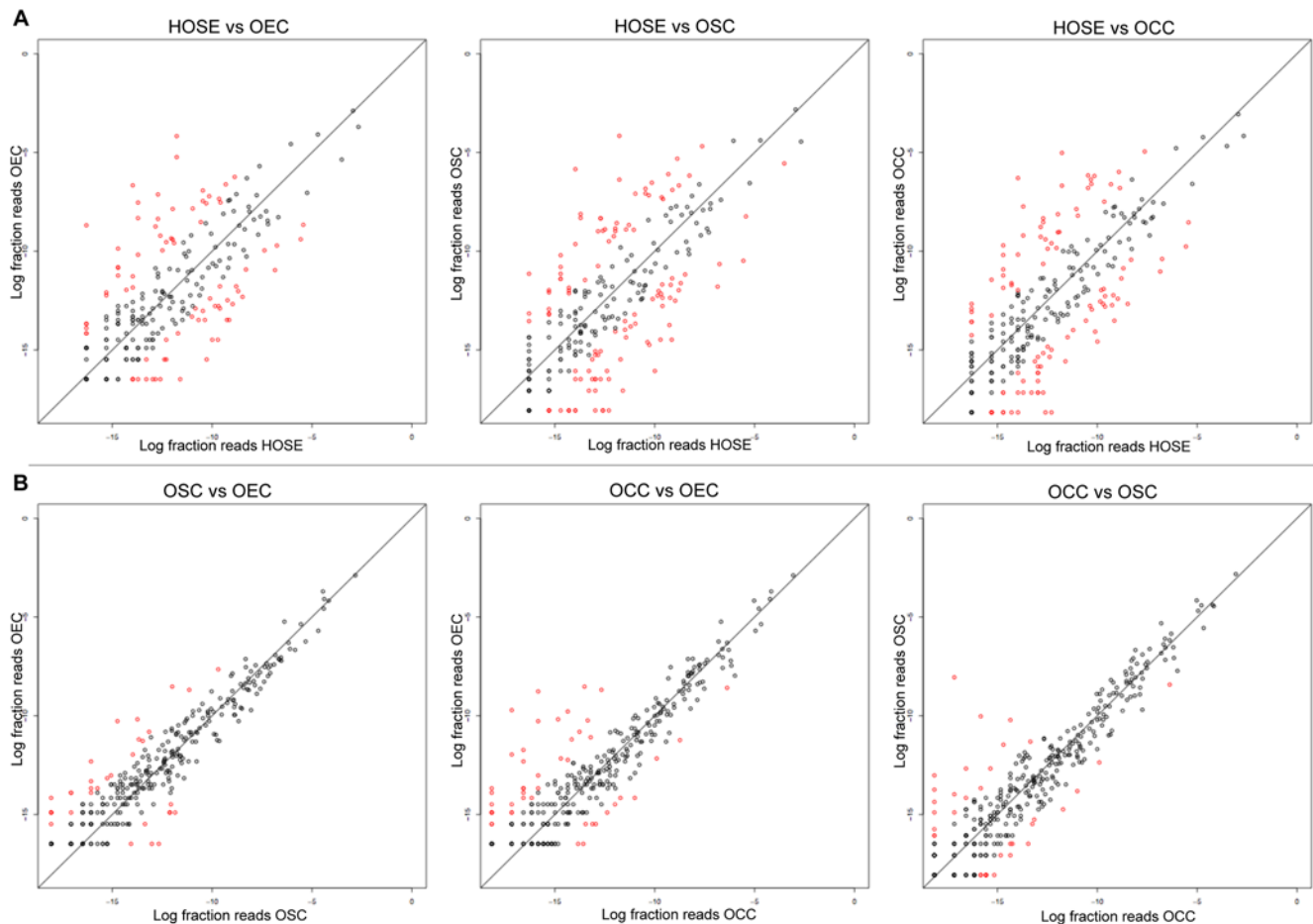


Figure 2. Comparison of abundance of known miRNAs across datasets. Datasets are compared pairwise, plotting the log of the fraction of total reads for a given miRNA in a given dataset against its corresponding value in the second dataset. For each plot, only miRNAs that were sequenced in both datasets are plotted; miRNAs that were sequenced only in one of the two datasets are not shown. The top row (A), compares ovarian cancer datasets (OSC, OCC and OEC) to normal primary HOSE cultures; the bottom row (B), compares ovarian tumor histologic subtypes to each other. Red dots signify miRNAs that had a fold change ≥ 4 . doi:10.1371/journal.pone.0005311.g002

ovarian cancer histologic subtypes are listed in **Table S4** (results for all miRNAs are provided in their entirety in **Table S2**). Among the most histologic subtype-specific miRNAs are miR-449a (serous-specific), miR-499-5p/miR-375/miR196a/miR-196b/miR-182 (endometrioid-specific) and miR-486-5p/miR-144/miR-30a/miR-199a-5p (clear cell-specific).

Validation of 454 sequencing miRNA expression results by qRT-PCR. We used commercially available miRNA TaqMan qRT-PCR assays as an independent confirmation of miRNA differential expression identified by 454 sequencing. MicroRNA TaqMan qRT-PCR assays were performed for a sample of 38 known miRNAs that were selected based on over- or under-expression with respect to primary normal HOSE, or based on differential expression between different subtypes of ovarian cancer (calculated based on 454 sequencing data). Taqman assays were carried out on aliquots of the same RNA pools used for sequencing. Thirty-six of the 38 differentially expressed miRNAs examined by TaqMan qRT-PCR demonstrated consistent results with those obtained by 454 sequencing (**Figures 4 and 5, A, B, C**). Of the eight miRNAs identified by 454 sequencing to be over-expressed in ovarian cancer relative to the normal HOSE group (**Figure 4A**), all eight demonstrated over-expression in serous and clear cell cancer subtypes when measured using TaqMan qRT-

PCR, while six of the eight showed consistent results in the endometrioid cancer subtype (**miRs-126*, -142-3p, 195, 200a, 200b, 200c, 338-3p, 378***). Nine miRNAs were under-expressed in ovarian cancer relative to normal HOSE (**Figure 4B**), with all subtypes showing consistent results when measured by TaqMan qRT-PCR. There were six miRNAs that were undetectable (zero reads) in normal HOSE cultures (**Figure 4C**), but had detectable expression in OEC, OSC and/or OCC in the 454 sequencing data, and they likewise showed increased expression in ovarian histologic subtype samples versus normal primary HOSE when examined by TaqMan qRT-PCR (**Figure 4C**).

Fifteen miRNAs identified by 454 sequencing to be differentially expressed across ovarian cancer subtypes (**Figure 5A, B, C**) showed similar expression patterns when quantitated by qRT-PCR. Eight miRNAs were over-expressed specifically in endometrioid ovarian cancer (**Figure 5A**), four in serous cancer (**Figure 5B**), and three in clear cell cancer (**Figure 5C**) relative to the other subtypes.

Analysis of sequencing data to identify novel hairpin-derived small RNAs. Having validated 454 sequencing results corresponding to differential expression of previously annotated miRNAs, we next turned to the identification of novel miRNA

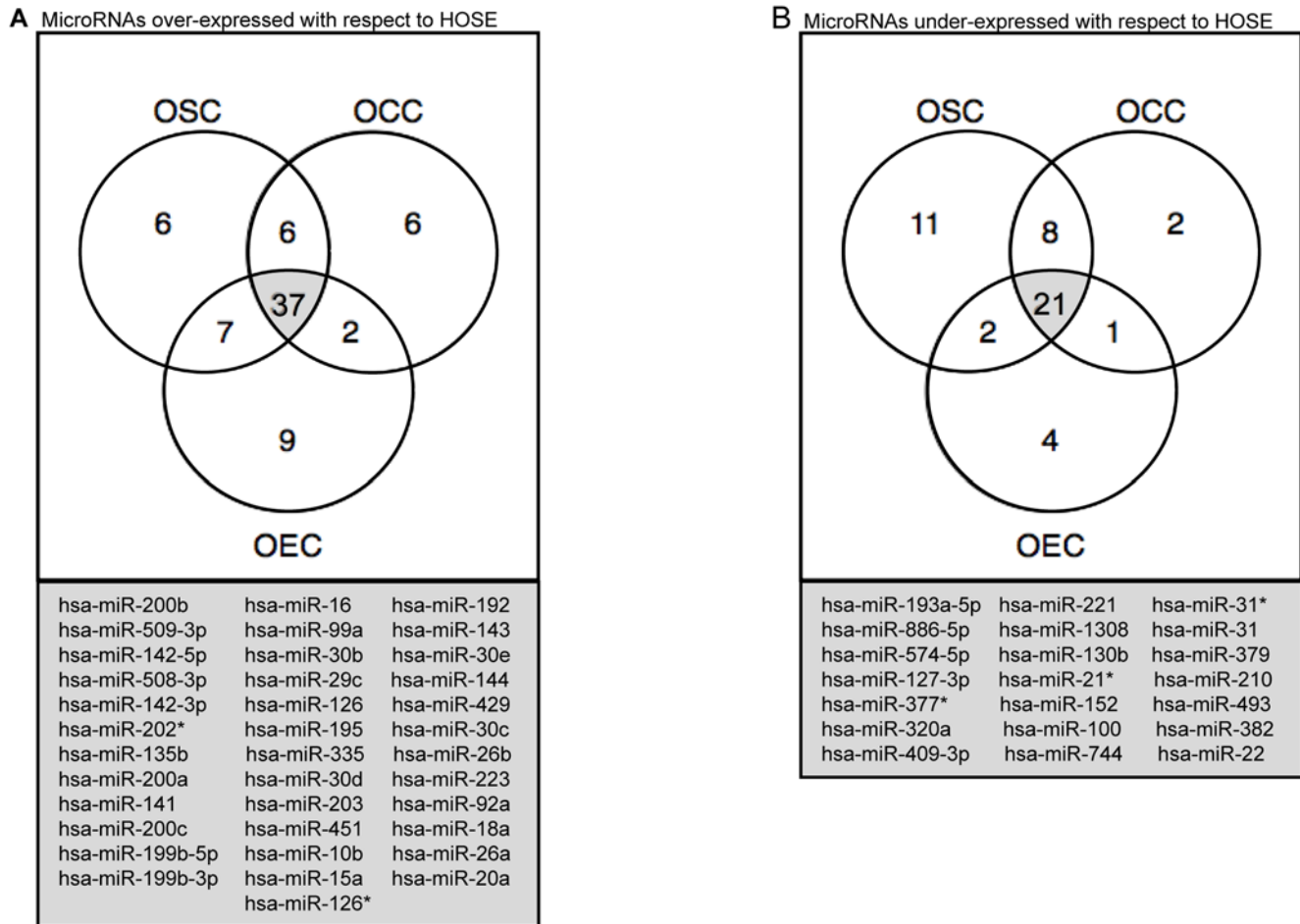


Figure 3. Differential expression of miRNAs in ovarian tumor histologic subtypes relative to normal primary HOSE. Venn diagrams depict numbers of miRNAs differentially expressed in ovarian cancer histologic subtypes relative to normal primary human ovarian surface epithelium (HOSE) cultures. Criteria for a miRNA as differentially expressed was defined by a fold-change ≥ 4 with a Fisher's exact test P -value $< 5.0 \times 10^{-6}$ for miRNAs that had detectable expression in both HOSE and ovarian cancer samples, or by a Fisher's exact test P -value $< 5.0 \times 10^{-6}$ alone in cases where a fold-change value was indefinable due to zero reads for a given miRNA in either the HOSE or ovarian cancer comparison sample. A total of 74 over-expressed miRNAs (Panel A) and 49 under-expressed miRNAs (Panel B) in ovarian cancer with respect to normal HOSE were identified. MicroRNAs that were found to be consistently over-expressed or consistently under-expressed in all three ovarian cancer histologic subtypes are listed by name. doi:10.1371/journal.pone.0005311.g003

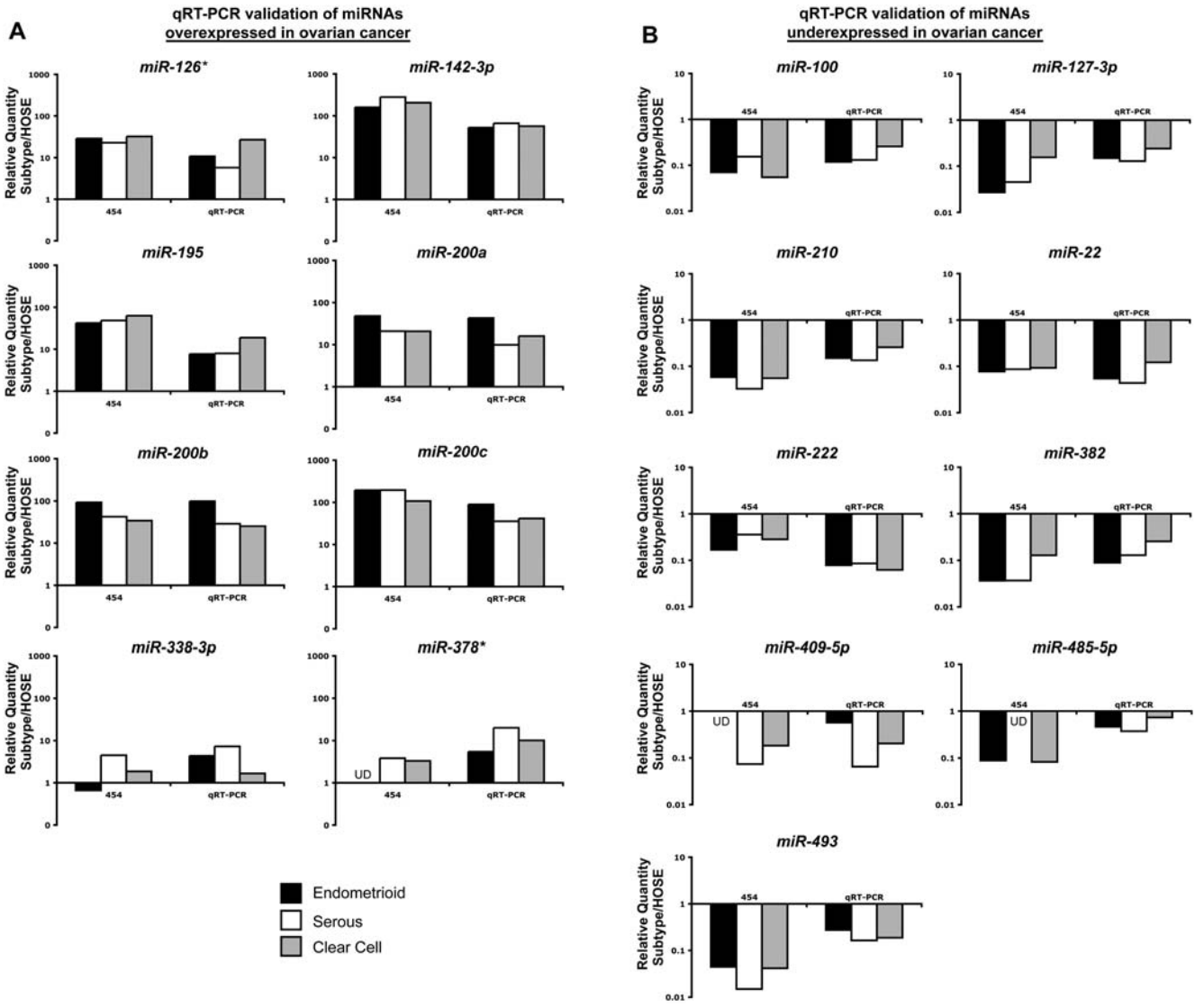
sequences. After filtering out matches to previously annotated features including messenger RNAs, tRNAs and other known noncoding RNAs, the remaining sequences were aligned to the reference human genome sequence (NCBI Build 36.1) [31]. Sequences were required to match the human genome sequence perfectly to be carried further for additional computational analysis, with the only exception to this being sequences that demonstrated addition of 1–3 non-templated nucleotides at the 3' terminus. In these cases, we trimmed non-templated bases from the original sequence and it was then carried further for analysis [32,33].

The data processing steps described thus far yielded 841 normal HOSE, 2,840 OSC, 11,224 OCC, and 1,764 OEC unique sequences corresponding to novel small RNAs. These unique sequences corresponded to 1,766 normal HOSE, 5,795 OSC, 19,464 OCC and 3,301 OEC genomic loci that could potentially generate these small RNAs (Figure 1B, Table S1). The data analysis pipeline next screened these genomic loci (in addition to 100 nucleotides of up- and down-stream flanking sequence) for the presence of a predicted hairpin secondary structure by computing free energy, shape probability (as determined by the RNashapes program [34]) and the Randfold-computed [35] P -value of

predicted secondary structures. A detailed description of hairpin folding criteria is provided in **Supplementary Methods S1** and in [23]. This resulted in 224 normal HOSE, 511 OSC, 500 OCC, and 247 OEC novel small RNAs found to be potentially derived from a precursor hairpin structure.

Sequences passing the above filtering criteria from the four datasets were then combined and sorted with respect to chromosomal coordinates into groups sharing 5' ends. From each of these groups we chose a "canonical" sequence representing that genomic locus. The canonical sequences were chosen based on a common 5' terminus, abundance, and sequence length as described previously [23] (additional details are provided in **Supplementary Methods S1**). This process further refined the data into a combined list of 199 unique sequences (potentially derived from 568 genomic loci) that we designated as "novel hairpin-derived small RNAs" from which we identified novel and candidate miRNAs.

Identifying novel and candidate miRNAs. In order to determine which sequences to designate as novel miRNAs, we screened the set of novel hairpin-derived small RNAs using criteria similar to those used in other recent miRNA discovery studies



C **MicroRNAs undetectable in normal primary HOSE by 454**

miRNA	HOSE		OEC		OSC		OCC	
	454 (percentage of total reads)	qRT-PCR (cycle threshold)	454 (percentage of total reads)	qRT-PCR (cycle threshold)	454 (percentage of total reads)	qRT-PCR (cycle threshold)	454 (percentage of total reads)	qRT-PCR (cycle threshold)
miR-135a	0	36.30	0.1137	25.70	0.0840	25.15	0.0048	25.74
miR-141	0	34.00	0.4491	24.54	0.5155	25.10	0.3100	23.38
miR-142-5p	0	36.73	0.0699	29.69	0.0782	27.95	0.0806	26.52
miR-429	0	36.36	0.0885	28.93	0.0409	28.74	0.0464	26.33
miR-592	0	UD	0.0033	34.18	0.0011	35.69	0.0048	30.91
miR-9*	0	UD	0.0066	31.86	0.0037	31.48	0.0010	31.38

Figure 4. Validation of 454 sequencing-derived miRNA expression results by qRT-PCR: miRNAs differentially expressed in ovarian cancer relative to normal HOSE. The graphs display miRNA Taqman qRT-PCR results for a sampling of miRNAs identified by 454 sequencing to be over-expressed (Panel A) or under-expressed (Panel B) in ovarian cancer relative to normal HOSE. Relative expression values derived from 454 sequencing are displayed on the left segment of each graph for comparison. Relative expression values of miRNAs in endometrioid (OEC, black bars), serous (OSC, white bars) and clear cell (OCC, grey bars) cancers are depicted. MicroRNAs that were not detected in normal HOSE by 454 sequencing but were detected in ovarian cancer are presented in Panel C, where miRNAs are ordered by descending expression (which corresponds to ascending cycle threshold, Ct) in normal HOSE. 454 sequencing data is given in terms of percentage of total reads within each sample. UD, undetectable by miRNA TaqMan qRT-PCR.
doi:10.1371/journal.pone.0005311.g004

[23,33,36]: *i.* pairing characteristics of the hairpin, *ii.* the presence of multiple reads sharing the same 5' terminus, *iii.* absence of annotation indicating non-miRNA biogenesis, *iv.* shared seed region with a known animal miRNA and *v.* presence of

corresponding miRNA star form read(s). We considered including phylogenetic sequence conservation as a criterion; however, this did not add significantly to our analysis as none of the novel hairpin-derived miRNAs were conserved beyond

qRT-PCR validation of miRNAs differentially expressed between ovarian cancer subtypes

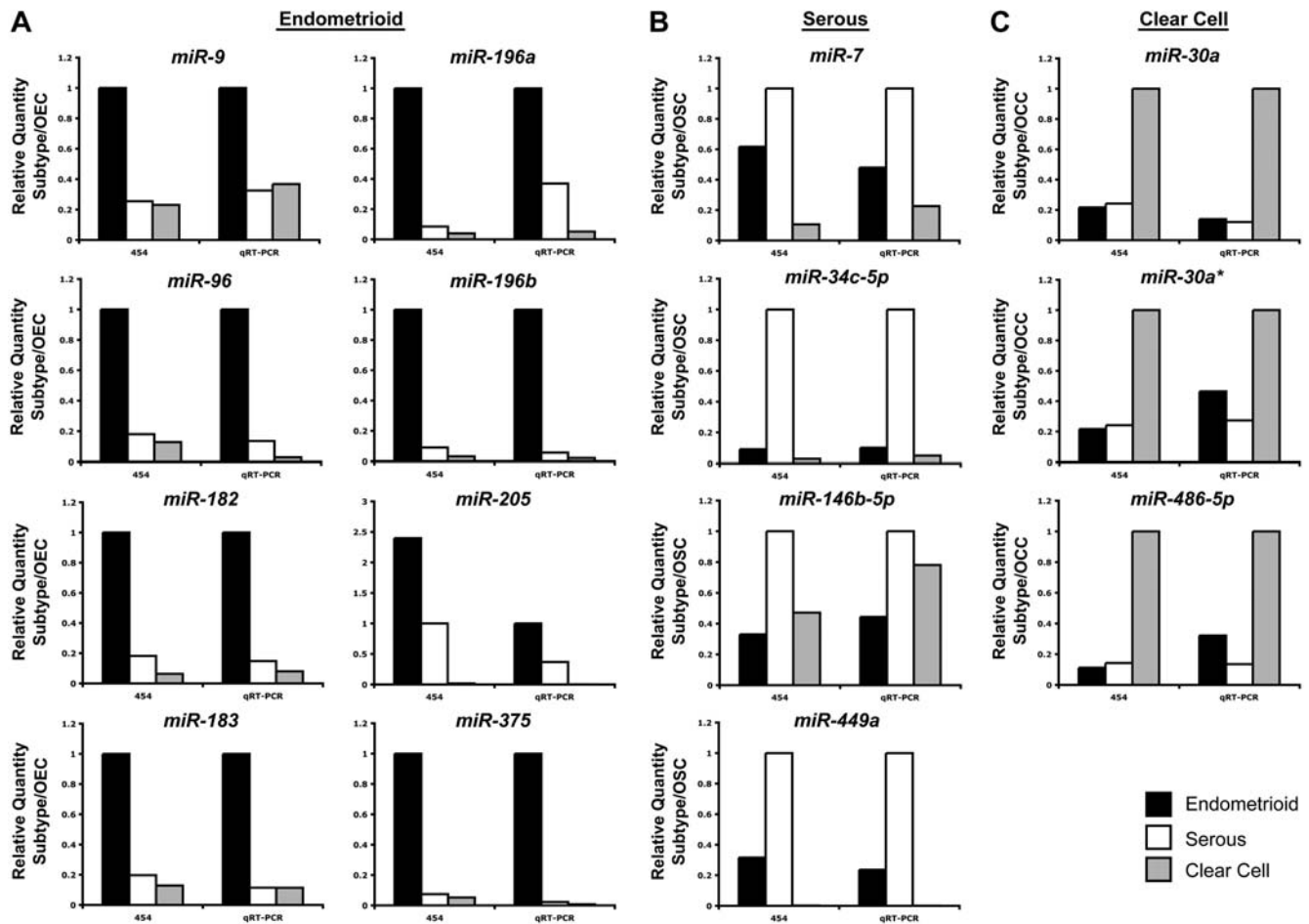


Figure 5. Validation of 454 sequencing-derived miRNA expression results by qRT-PCR: miRNAs differentially expressed between ovarian cancer histologic subtypes. The graphs display Taqman qRT-PCR results for miRNAs differentially expressed between ovarian cancer subtypes, with 454 sequencing-derived data presented for comparison in the left segment of each graph. Relative expression values of miRNAs specifically over-expressed in endometrioid (Panel A, OEC, black bars), serous (Panel B, OSC, white bars) and clear cell (Panel C, OCC, grey bars) cancers are depicted. UD, undetectable by miRNA TaqMan qRT-PCR. doi:10.1371/journal.pone.0005311.g005

primates. This is not unexpected, given that most evolutionarily conserved microRNAs are likely to have already been discovered using computational approaches followed up by directed validation experiments.

Using the criteria above, we identified six novel miRNAs. All of the six met criteria *i-iii* and three of the six met criterion *iv* and three met criterion *v* (Figure 6; more detail is provided in Table S5). In addition, mapping the novel miRNA sequences to the reference human genome sequence revealed that three of the six novel miRNAs are present in introns of other genes (and encoded on the same strand as the respective host genes) (Figure 6), much like many previously annotated miRNAs [37,38]. Two of our novel miRNAs are located in annotated repeats, but have strong enough supporting evidence that they were designated as novel. The context of the predicted precursor structure and the sequence of miRNA star forms corresponding to novel miRNAs is given in Figure 6.

The remaining sequences comprised 181 hairpin-derived small RNAs (corresponding to 550 genomic loci). We sought to select from this list the most promising sequences to designate as “candidate miRNAs” that might be confirmed in the future as *bona*

fide miRNAs as additional evidence accumulates. Candidate miRNAs were required to meet criteria *i-iii* and have some additional form of supporting evidence (i.e., shared seed region with a known miRNA). This allowed us to refine a final list of 39 candidate miRNAs (potentially originating from 50 genomic loci) (Table S5).

Discussion

The work reported here was motivated by the hypothesis that the entire repertoire of miRNAs expressed in ovarian cancer, including potentially tissue- and cancer-specific miRNAs, had not yet been elucidated. The massively parallel sequencing approach allowed us to be comprehensive (i.e., identifying not only known but also novel miRNAs), and the “digital” nature of the data permitted a semi-quantitative estimation of the relative expression level for many miRNAs. Using this approach, we discovered six novel and 39 candidate miRNAs, and delineated the expression pattern of 498 previously annotated miRNAs in normal primary HOSE cultures and epithelial ovarian cancer tissues. It is notable that of the four published miRNA microarray studies of ovarian

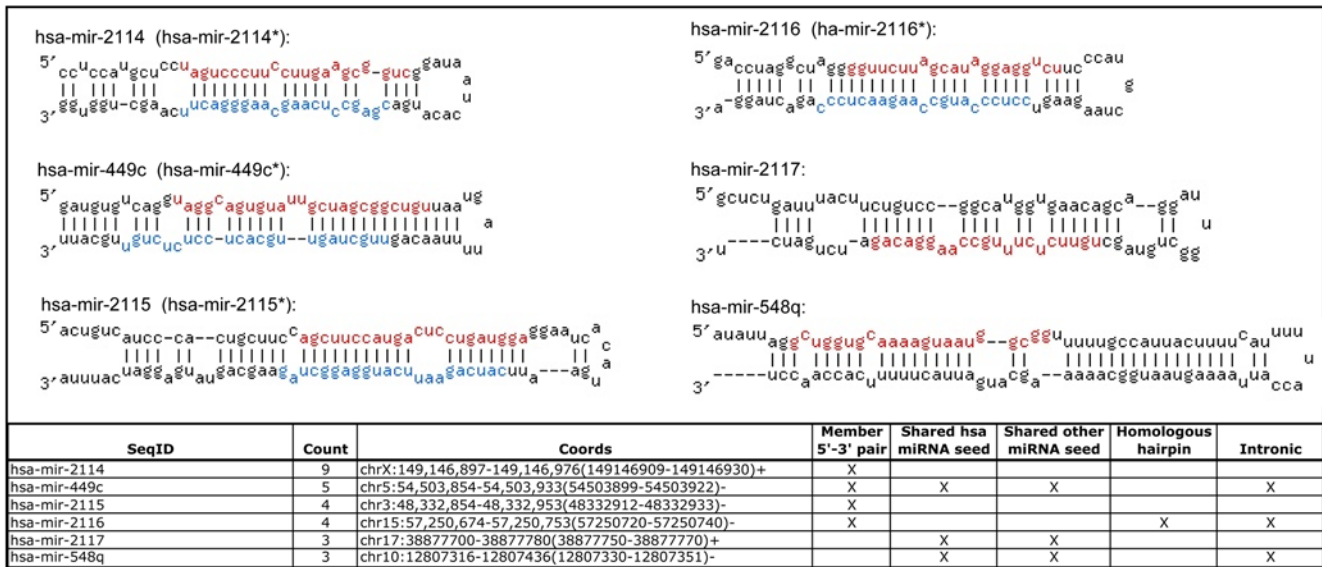


Figure 6. Novel miRNAs discovered by 454 sequencing. Putative secondary structures for the six novel miRNAs discovered in this study are shown. Novel miRNA sequences are shown in red and star forms of novel miRNAs are shown in blue (if identified in our sequencing data). Sequences are novel with respect to miRBase release 12.0 [21,22]. Identifiers in parentheses refer to the miRNA star forms where these were identified in the 454 sequencing data. Corresponding details of the novel miRNAs are given in the table below the hairpin structures. doi:10.1371/journal.pone.0005311.g006

cancer [17–20], even the most comprehensive of the array platforms was limited to 470 human miRNAs and star forms [17], while the current release of miRBase (release 12.0) [21,22] contains 969 human miRNAs and star forms. Furthermore, 223 of the known miRNAs we identified and characterized in our sequencing datasets were not represented on even the most comprehensive microarray. Therefore, the results presented here substantially extend the breadth of knowledge of miRNAs expressed in ovarian cancer.

The identification of novel miRNAs in our data is particularly notable. Given that these novel miRNAs have not been detected in multiple large-scale miRNA sequencing studies of other tissues, we speculate that they are expressed in an ovarian cancer-specific manner and may be of special interest as cancer biomarkers. Future studies will be needed to test this hypothesis, as well as to investigate the biological roles of these novel miRNAs in ovarian cancer pathogenesis.

Taken together, the known and novel miRNAs identified in this study provide a pool of candidate miRNA markers for future analysis as blood-based markers for the detection of ovarian cancer. Differential expression between normal and malignant states may provide one method for prioritization of candidates moving forward. We also envision that the miRNAs that we found to be differentially expressed between histologic cancer subtypes could serve a dual purpose as blood-based markers for both cancer detection and histologic classification.

One of the unique features of our study is the use of primary cultures of normal human ovarian surface epithelium as the normal comparison group. Although the ovarian surface epithelium (along with nearby distal fallopian tube epithelium) is considered the origin of epithelial ovarian cancer [39] most other miRNA profiling studies have typically used whole ovary as the normal comparison. This is problematic for identifying miRNAs that are differentially expressed in ovarian cancer relative to normal because the single cell-thick surface epithelial layer comprises far less than 1% of the cellular content of the whole ovary. Performing comparisons of ovarian cancer tissue to an

appropriate normal control is a challenging problem in ovarian cancer research. Our approach, although it avoids the problems associated with using whole ovary as a normal sample, has the limitation that we do not know whether and which changes in microRNA expression may be attributed to culture conditions used for primary HOSE cell culture. Future studies incorporating parallel analyses of primary cell cultures derived from ovarian cancer tissues may address this question.

Although the use of different types of “normal” ovarian specimens makes comparison of our data with other studies difficult in most cases, in one microarray study, however, the normal control was human ovarian surface epithelial cells that had been immortalized using a retrovirus expressing human papillomavirus-derived viral oncoproteins E6 and E7 [17,40]. Because the normal primary HOSE cultures we studied are not immortalized, comparison of our results to that report could identify miRNAs that may be induced by immortalization by E6 and E7 proteins. A notable finding from this comparison is that whereas we identified the miR-200 family of miRNAs (miR-200a/b/c, miR-141 and miR-429) to be expressed at low levels in normal HOSE with highly increased expression in ovarian cancer, the study using E6/E7-immortalized HOSE did not find increased expression of miR-200 family miRNAs in ovarian cancer. This family of miRNAs has important functions in repressing the proteins ZEB1 and ZEB2 [41–46], which are transcription factors that function to promote epithelial-to-mesenchymal transition [47–51]. One appealing interpretation of the differing results is that primary normal HOSE express low levels of the miR-200 family miRNAs, but the E6 and E7 viral oncoproteins induce miR-200 family expression, which may be an important step in ovarian carcinogenesis given that this miRNA family was the most highly over-expressed miRNA family in all ovarian cancer types we examined. Future work will be needed to address this hypothesis.

In summary, our work expands the current view of the repertoire of miRNAs expressed in normal HOSE and common histologic subtypes of ovarian cancer. Our results represent a

resource for future investigations of the role of miRNAs in ovarian cancer and of their applications as clinically useful biomarkers.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [52] and are accessible through GEO Series accession number GSE15190 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15190>).

Supporting Information

Table S1 Overview of computational pipeline data processing results. Shaded lines represent steps at which sequences are pruned out of the dataset before the next step. The left column for each dataset contains the number of nonredundant unique sequences, and the right column is the sum of all the reads for each sequence. Found at: doi:10.1371/journal.pone.0005311.s001 (0.02 MB XLS)

Table S2 Identity, abundance data and pairwise relative expression data for all known miRNAs in 454 sequencing datasets. Columns B-I display the number of reads and fractional abundance values for miRNAs corresponding to the 4 sample types studied. Columns J-U provide pairwise comparisons of expression (along with Fisher's exact test P-values) of each miRNA between all combinations of the sample types studied. Found at: doi:10.1371/journal.pone.0005311.s002 (0.21 MB XLS)

Table S3 MicroRNAs differentially expressed between ovarian cancer and normal HOSE. Found at: doi:10.1371/journal.pone.0005311.s003 (0.05 MB XLS)

Table S4 MicroRNAs differentially expressed between ovarian cancer histologic subtypes. Bold blocks of numbers show top 10 most differentially expressed known miRNAs between ovarian cancer subtypes. Found at: doi:10.1371/journal.pone.0005311.s004 (0.04 MB XLS)

References

- Berkenblit A, Cannistra SA (2005) Advances in the management of epithelial ovarian cancer. *J Reprod Med* 50: 426–438.
- Cannistra SA (2004) Cancer of the ovary. *N Engl J Med* 351: 2519–2529.
- Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, et al. (2003) The case for early detection. *Nat Rev Cancer* 3: 243–252.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Kloosterman WP, Plasterk RH (2006) The diverse functions of microRNAs in animal development and disease. *Dev Cell* 11: 441–450.
- Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer* 6: 857–866.
- Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 6: 259–269.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. (2005) MicroRNA expression profiles classify human cancers. *Nature* 435: 834–838.
- Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, et al. (2008) MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 26: 462–469.
- Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, et al. (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 353: 1793–1801.
- Cummins JM, Velculescu VE (2006) Implications of micro-RNA profiling for cancer diagnosis. *Oncogene* 25: 6220–6227.
- Yu SL, Chen HY, Chang GC, Chen CY, Chen HW, et al. (2008) MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 13: 48–57.
- Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, et al. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105: 10513–10518.
- Taylor DD, Gerceel-Taylor C (2008) MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* 110: 13–21.
- Chen X, Ba Y, Ma L, Cai X, Yin Y, et al. (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*.
- Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, et al. (2008) Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol* 141: 672–675.
- Dahiya N, Sherman-Baust CA, Wang TL, Davidson B, Shih Ie M, et al. (2008) MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLoS ONE* 3: e2436.
- Iorio MV, Visone R, Di Leva G, Donati V, Petrocca F, et al. (2007) MicroRNA signatures in human ovarian cancer. *Cancer Res* 67: 8699–8707.
- Nam EJ, Yoon H, Kim SW, Kim H, Kim YT, et al. (2008) MicroRNA expression profiles in serous ovarian carcinoma. *Clin Cancer Res* 14: 2690–2695.
- Zhang L, Volinia S, Bonome T, Calin GA, Greshock J, et al. (2008) Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc Natl Acad Sci U S A* 105: 7004–7009.
- Griffiths-Jones S (2004) The microRNA Registry. *Nucleic Acids Res* 32: D109–111.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140–144.
- Bar M, Wyman SK, Fritz BR, Qj J, Garg KS, et al. (2008) MicroRNA Discovery and Profiling in Human Embryonic Stem Cells by Deep Sequencing of Small RNA Libraries. *Stem Cells*.
- Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, et al. (2006) Diversity of microRNAs in human and chimpanzee brain. *Nat Genet* 38: 1375–1377.
- Dyck HG, Hamilton TC, Godwin AK, Lynch HT, Maines-Bandiera S, et al. (1996) Autonomy of the epithelial phenotype in human ovarian surface epithelium: changes with neoplastic progression and with a family history of ovarian cancer. *Int J Cancer* 69: 429–436.
- Li NF, Wilbanks G, Balkwill F, Jacobs IJ, Dafou D, et al. (2004) A modified medium that significantly improves the growth of human normal ovarian surface epithelial (OSE) cells in vitro. *Lab Invest* 84: 923–931.

Table S5 Novel and candidate miRNAs. Alternating yellow and orange miRNAs represent 5'-3' pairs, with the yellow indicating the dominant sequence. Pink rows are the two remaining novel miRNAs. The non-highlighted sequences represent candidate miRNAs.

Found at: doi:10.1371/journal.pone.0005311.s005 (0.04 MB XLS)

Methods S1

Found at: doi:10.1371/journal.pone.0005311.s006 (0.12 MB DOC)

Figure S1 Flow chart of sequence data analysis pipeline. The flow chart shows the steps in the computational analysis of the 454 sequencing data. At each step, sequences may be removed for further analysis, or carried on to the next step in the pipeline. The first steps remove previously annotated features from the pipeline, and then remaining sequences are tested for presence of hairpin secondary structure and other criteria to be designated novel miRNAs. Bolded lower-case letters are referred to in the Supplementary Methods.

Found at: doi:10.1371/journal.pone.0005311.s007 (0.77 MB TIF)

Acknowledgments

We thank Carolyn Slater and Lisa Vanderveer for technical assistance with normal primary HOSE cell culture, April Allen for assistance with miRNA isolation and Angie Lackey, Danielle Blecha and Tim Harkins at Roche Diagnostics for technical advice regarding small RNA library sequencing.

Author Contributions

Conceived and designed the experiments: MT. Performed the experiments: RKP PSM BRF. Analyzed the data: SKW MT. Contributed reagents/materials/analysis tools: KCO AKG NU CD BSK. Wrote the paper: SKW PSM MT.

27. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858–862.
28. Aravin A, Tuschl T (2005) Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett* 579: 5830–5840.
29. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.
30. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, et al. (2006) The colorectal microRNAome. *Proc Natl Acad Sci U S A* 103: 3687–3692.
31. Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
32. Berezikov E, Cuppen E, Plasterk RH (2006) Approaches to microRNA discovery. *Nat Genet* 38 Suppl: S2–7.
33. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193–1207.
34. Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 22: 500–503.
35. Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20: 2911–2917.
36. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, et al. (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* 17: 1850–1864.
37. Kim YK, Kim VN (2007) Processing of intronic microRNAs. *Embo J* 26: 775–783.
38. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res* 14: 1902–1910.
39. Auersperg N, Wong AS, Choi KC, Kang SK, Leung PC (2001) Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr Rev* 22: 255–288.
40. Gregoire L, Munkarah A, Rabah R, Morris RT, Lancaster WD (1998) Organotypic culture of human ovarian surface epithelial cells: a potential model for ovarian carcinogenesis. *In Vitro Cell Dev Biol Anim* 34: 636–639.
41. Christoffersen NR, Silahatoglu A, Orom UA, Kauppinen S, Lund AH (2007) miR-200b mediates post-transcriptional repression of ZFHx1B. *Rna* 13: 1172–1178.
42. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, et al. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 10: 593–601.
43. Hurteau GJ, Carlson JA, Spivack SD, Brock GJ (2007) Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. *Cancer Res* 67: 7972–7976.
44. Korpala M, Lee ES, Hu G, Kang Y (2008) The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* 283: 14910–14914.
45. Park SM, Gaur AB, Lengyel E, Peter ME (2008) The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 22: 894–907.
46. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, et al. (2008) A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* 9: 582–589.
47. Spaderna S, Schmalhofer O, Wahlbuhl M, Dimmler A, Bauer K, et al. (2008) The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. *Cancer Res* 68: 537–544.
48. Shirakihara T, Saitoh M, Miyazono K (2007) Differential regulation of epithelial and mesenchymal markers by deltaEF1 proteins in epithelial mesenchymal transition induced by TGF-beta. *Mol Biol Cell* 18: 3533–3544.
49. Aigner K, Dampier B, Descovich L, Mikula M, Sultan A, et al. (2007) The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene* 26: 6979–6988.
50. Eger A, Aigner K, Sonderegger S, Dampier B, Oehler S, et al. (2005) DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene* 24: 2375–2385.
51. Comijn J, Berx G, Vermassen P, Verschuere K, van Grunsven L, et al. (2001) The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell* 7: 1267–1278.
52. Edgar RDM, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207–210.