Association between miR-200c and the survival of patients with stage I epithelial ovarian cancer: a retrospective study of two independent tumour tissue collections

Sergio Marchini*, Duccio Cavalieri*, Robert Fruscio, Enrica Calura, Daniela Garavaglia, Ilaria Fuso Nerini, Costantino Mangioni, Giorgio Cattoretti, Luca Clivio, Luca Beltrame, Dionyssios Katsaros, Luca Scarampi, Guido Menato, Patrizia Perego, Giovanna Chiorino, Alessandro Buda, Chiara Romualdi, Maurizio D'Incalci

Summary

Background International Federation of Gynecology and Obstetrics stage I epithelial ovarian cancer (EOC) has a significantly better prognosis than stage III/IV EOC, with about 80% of patients surviving at 5 years (compared with about 20% of those with stage III/IV EOC). However, 20% of patients with stage I EOC relapse within 5 years. It is therefore crucial that the biological properties of stage I EOCs are further elucidated. MicroRNAs (miRNAs) have shown diagnostic and prognostic potential in stage III and IV EOCs, but the small number of patients diagnosed with stage I EOC has so far prevented an investigation of its molecular features. We profiled miRNA expression in stage I EOC tumours to assess whether there is a miRNA signature associated with overall and progression-free survival (PFS) in stage I EOC.

Methods We analysed tumour samples from 144 patients (29 of whom relapsed) with stage I EOC gathered from two independent tumour tissue collections (A and B), both with a median follow-up of 9 years. 89 samples from tumour tissue collection A were stratified into a training set (51 samples, 15 of which were from patients who relapsed) for miRNA signature generation, and into a validation set (38 samples, seven of which were from patients who relapsed) for signature validation. Tumour tissue collection B (55 samples, seven of which were from patients who relapsed) was used as an independent test set. The Cox proportional hazards model and the log-rank test were used to assess the correlation of quantitative reverse transcription PCR (qRT-PCR)-validated miRNAs with overall survival and PFS.

Findings A signature of 34 miRNAs associated with survival was generated by microarray analysis in the training set. In both the training set and validation set, qRT-PCR analysis confirmed that 11 miRNAs (miR-214, miR-199a-3p, miR-199a-5p, miR-145, miR-200b, miR-30a, miR-30a*, miR-30d, miR-200c, miR-20a, and miR-143) were expressed differently in relapsers compared with non-relapsers. Three of these miRNAs (miR-200c, miR-199a-3p, miR-199a-5p) were associated with PFS, overall survival, or both in multivariate analysis. qRT-PCR analysis in the test set confirmed the downregulation of miR-200c in relapsers compared with non-relapsers, but not the upregulation of miR-199a-3p and miR-199a-5p. Multivariate analysis confirmed that downregulation of miR-200c in the test set was associated with overall survival (HR 0·094, 95% CI 0·012–0·766, p=0·0272) and PFS (0·035, 0·004–0·311; p=0·0026), independent of clinical covariates.

Interpretation miR-200c has potential as a predictor of survival, and is a biomarker of relapse, in stage I EOC.

Introduction

The 5-year survival of patients with epithelial ovarian cancer (EOC) depends on the spread of the disease at diagnosis. In patients with disease limited to the ovary (International Federation of Gynecology and Obstetrics [FIGO] stage I), survival is close to 80%. However, in cases in which the disease involves the upper abdomen or beyond (FIGO stages III and IV), only about 20% of patients survive at 5 years.[1] Fewer than 20% of patients with stage I EOC have aggressive disease and relapse within 5 years of primary surgery, but an unresolved clinical issue is how to discriminate at an early stage between patients with curable stage I disease and those who will relapse after adjuvant chemotherapy. Prediction

of relapse on the basis of currently known clinical and pathological features is difficult.[1] Predicting relapse might be possible with a better knowledge of the molecular and genetic mechanisms that are associated with each tumour stage. Knowledge of the molecular pathways that are altered during neoplastic transformation might help expedite the discovery of biomarkers for early disease detection, prediction of clinical response, and guidance of treatment.

Fewer than 10% of all patients with EOC diagnosed every year have stage I disease, which makes it difficult to recruit patient cohorts of a size adequate to produce statistical power. Knowledge of the biological differences between stage I disease, which is curable in most cases,

Department of Oncology, Mario Negri Institute for Pharmacological Research, Milan, Italy (S Marchini PhD, I Fuso Nerini BSc, D Garavaglia BSc, L Clivio Eng, M D'Incalci MD); **Department of Obstetrics and Gynecology** (R Fruscio MD, A Buda MD, Prof C Mangioni MD), **and the Unit of Anatomo-pathology** (Prof G Cattoretti MD, P Perego MD), **University of Milano–Bicocca, San Gerardo Hospital, Monza, Italy; Department of Pharmacology, University of Firenze, Firenze, Italy** (Prof D Cavalieri PhD, L Beltrame PhD); **Department of Biology, University of Padova, Padova, Italy** (C Romualdi PhD, E Calura BSc); **Department of Gynecology-Oncology, Azienda Ospedaliera OIRM S Anna, University of Torino, Torino** (D Katsaros MD, L Scarampi MD, G Menato MD); **Cancer Genomic Laboratory, Edo Tempia Foundation, Biella, Italy** (G Chiorino PhD); **and MaNGO Group, Milan, Italy** (M D'Incalci, R Fruscio, D Katsaros, C Mangioni)

Correspondence to:
Dr Maurizio D'Incalci, Department of Oncology, Mario Negri Institute for Pharmacological Research, Via La Masa 19, 20145 Milan, Italy
maurizio.dincalci@ marionegri.it

and stage III, which is generally incurable, is limited, although important biological differences between stage I and stage III disease have been suggested. All patients with stage I disease express wild-type P53, whereas a large proportion of patients with stage III disease have mutated P53[2] or abnormal levels of other members of the P53 family (eg, increased expression of DNp63).[3] In a collection of tumour samples from a cohort of 68 stage I tumours and 15 tumours with low malignant potential (also known as borderline) selected from a collection of fresh-frozen samples available at the Mario Negri Institute, we recently defined a gene-expression signature that distinguished patients with stage I disease who relapsed from those who did not.[4]

MicroRNAs (miRNAs) are small 21–23 nucleotide long non-coding RNAs.[5] During the past 5 years the importance of miRNAs in cell biology has been confirmed, and the case for their role in disease, particularly cancer, is compelling. The use of array technologies to characterise miRNAs in cancer tissue banks has discovered unique miRNA-expression patterns related to clinical features of several cancers. The present hypothesis is that miRNAs can be used to generate discriminators with diagnostic and prognostic value in situations where histology and mRNA-based diagnostic tools are deficient.

Studies of stage III and IV EOC and of ovarian cancer cell lines suggest that miRNAs are downregulated in tumour samples compared with normal tissue, and that their altered expression affects response to chemotherapy. Genes involved in the biogenesis of miRNAs were also altered in EOC. Dicer and Drosha mRNA and protein concentrations have been associated with outcome in a cohort of 108 stage III and stage IV ovarian tumours.[6] There appear to be no published studies on stage I ovarian cancers, perhaps because of the relatively small proportion of patients who present with stage I disease. We aimed to investigate the association between miRNAs and prognosis in stage I EOC. To that end, we studied a genome-wide miRNA-expression profile in tumour samples from a cohort of 144 patients (29 of whom relapsed) with stage I EOC.

Methods
Sample collection
Tumour samples from a cohort of 144 patients staged according to the FIGO criteria as stage I EOC[7] were gathered from two independent tumour tissue collections (collections A and B). For collection A, 89 samples were taken from a frozen tissue bank containing 1300 samples that were collected between September, 1992, and March, 2005, and available at the Department of Oncology, Mario Negri Institute, Milan, Italy. Tumour tissue had been collected from patients who underwent surgery for EOC at the Obstetrics and Gynecology Department, San Gerardo Hospital, Monza, Italy, as described previously.[4] For collection B, 55 samples were taken from the collection of 600 frozen samples that were collected between January, 1992, and December, 2005, and available at the Department of Gynecology-Oncology, University of Torino, Torino, Italy. The study was done on samples from patients with stage I EOC and at least 5 years follow-up data. Samples had been collected in the operating theatre from patients undergoing cytoreductive surgery, frozen within 15 min in liquid nitrogen and stored at –80°C.

The tumour content of the specimens in collections A and B was assessed by haematoxylin and eosin stain in the pathology units. Only specimens containing more than 70% of tumour tissue were used. Patients' clinical and anatomo-pathological information was registered, and follow-up data was obtained from periodic gynecological oncological check-ups. The collection and use of tumour samples was approved by the local scientific ethical committees and written consent was obtained from the patients.

Samples from collection A were randomly allocated with a ratio of 4:3, stratified by substage and grade, into a training set and a validation set. Samples from collection B were defined as the test set.

	Tissue collection A			Tissue collection B
	Overall (n=89)	Training set (n=51)	Validation set (n=38)	Test set (n=55)
Median age (range) years	52 (21–82)	51 (23–82)	55 (21–75)	57 (48–69)
Chemotherapy				
Yes	61 (69%)	38 (75%)	23 (61%)	33 (60%)
No	28 (31%)	13 (25%)	15 (39%)	22 (40%)
Histotype*				
Clear cell	14 (16·1%)	8 (16·00%)	6 (16·22%)	6 (11%)
Endometroid	23 (26·4%)	15 (30·00%)	8 (21·62%)	19 (35%)
Mucinous	20 (22·99%)	14 (28·00%)	6 (16·22%)	9 (16%)
Mullerian	0 (0·00%)	0 (0·00%)	0 (0·00%)	2 (4%)
Serous	29 (33·33%)	12 (24·00%)	17 (45·95%)	13 (24%)
Undifferentiated	1 (1·15%)	1 (2·00%)	0 (0·00%)	6 (11%)
Figo I substage				
A	28 (31%)	17 (33%)	11 (29%)	22 (40%)
B	5 (6%)	3 (6%)	2 (5%)	8 (15%)
C	56 (63%)	31 (61%)	25 (66%)	25 (45%)
Grade				
1	23 (26%)	13 (25%)	10 (26%)	23 (42%)
2	28 (31%)	15 (29%)	13 (34%)	10 (18%)
3	38 (43%)	23 (45%)	15 (39%)	22 (40%)
Figo I substage (chemotherapy)				
A	13 (21%)	9 (24%)	4 (17%)	8 (24%)
B	5 (8%)	3 (8%)	2 (9%)	5 (15%)
C	43 (70%)	26 (68%)	17 (74%)	20 (61%)
Figo I substage (no chemotherapy)				
A	15 (54%)	8 (62%)	7 (47%)	14 (64%)
B	0 (0%)	0 (0%)	0 (0%)	3 (14%)
C	13 (46%)	5 (38%)	8 (53%)	5 (23%)

All data are n (%) unless stated otherwise. *There was one missing value each in the training and validation set.

Table 1: **Main characteristics of patients and tissue samples**

Procedures

Frozen samples (30 mg) were homogenised in an Ultraturrax at 4°C, and total RNA enriched in miRNAs fraction was purified using a mirVana isolation kit according to the manufacturer's instructions (Ambion-ABI, Milan, Italy). RNA was measured by Nanodrop, and the presence of small RNAs was checked with a 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA). Aliquots were stored at –80°C until use. 100 ng of total RNA enriched in miRNA fraction was Cy5-labelled and hybridised with a miRNA labelling and hybridisation kit according to the manufacturer's instructions (Agilent Technologies). We used the commercially available G4470B human miRNA Microarray kit (Agilent Technologies), which consists of 15K features printed in an 8-plex format (8×15 array), and can detect all known human miRNAs (723 human and 76 human-viral miRNAs) sourced from the Sanger miRBASE public database, release 10.1. The arrays were washed and scanned with a laser confocal scanner (G2565BA, Agilent Technologies) according to the manufacturer's instructions. miRNA microarrays underwent standard post hybridisation processing and the intensities of fluorescence were calculated by Feature Extraction software version 11 (Agilent Technologies).

Analysis of microarray data

Raw data were pre-processed to filter out those probes with more than 40% of measurements classed as not available. The filter on the coverage of the probes was done to reduce biases in the normalisation step. Pre-processed data were normalised using cyclic lowess.[8] To identify miRNAs that were expressed differently in relapsers compared with non-relapsers, we did a two-class unpaired significance analysis of microarrays (SAM) test,[9] with 100 random permutations of the results. p values and their corresponding q values (the adjusted p value for false discovery rate; FDR) were used to control test multiplicity, with 25% chosen the cutoff. q values for each miRNA were defined as $q=(p*n)/i$, where p is the raw p value of the gene, n the total number of genes, and i is the number of genes at or better than p. Statistical analyses were done using the R programming language (version 2.11), and the Bioconductor software suite (version 2.6). Microarray raw data have been submitted to ArrayExpress (experiment number E-MTAB-297), in accordance with MIAME (minimum information about a microarray experiment) guidelines. Clustering analysis was done on normalised data. Single miRNA signals for each patient were divided by a baseline signal defined as the mean of the signal of that specific miRNA in all patients. A subset of the previously mentioned matrix was made using only 34 miRNAs coming from the analysis of differential expression. Hierarchical dendrograms were calculated with Pearson's correlation as clustering metric and by bootstrapping data over 1000 iterations. The calculation was done with the TIGR Multi-experiment Viewer feature of the TM4 software suite (version 4.4).[10]

	Overall survival				Progression-free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p
Age (5-year increment)	1·324 (1·051–1·670)	0·0175	1·236 (0·939–1·628)	0·1309	1·225 (1·028–1·459)	0·0231	1·249 (1·011–1·543)	0·0394
Chemotherapy (yes *vs* no)	8·230 (1·085–62·435)	0·0415	3·230 (0·367–28·448)	0·2909	14·543 (1·955–108·184)	0·0089	11·226 (1·376–91·560)	0·0239
Histotype								
Clear cell (reference)	..	..	..	..	..	..	..	..
Endometroid	0·152 (0·03–0·754)	0·0212	0·328 (0·036–3·001)	0·3240	0·321 (0·095–1·086)	0·0676	0·619 (0·112–3·417)	0·5825
Undifferentiated	1·983 (0·235–16·718)	0·5291	1·532 (0·177–13·246)	0·6981	4·391 (0·503–38·373)	0·1810	3·791 (0·427–33·628)	0·2314
Mucinous	0·170 (0·034–0·846)	0·0305	1·563 (0·163–14·958)	0·6983	0·134 (0·026–0·701)	0·0173	0·973 (0·141–6·741)	0·9782
Serous	0·285 (0·086–0·941)	0·0394	0·777 (0·189–3·202)	0·7271	0·650 (0·236–1·791)	0·4043	2·698 (0·838–8·679)	0·0960
Figo I substage								
A (reference)	..	..	..	..	..	..	..	..
B	6·014 (0·376–96·265)	0·2048	3·093 (0·141–67·908)	0·4737	3·692 (0·326–41·832)	0·2917	0·750 (0·060–9·352)	0·8234
C	7·951 (1·045–60·491)	0·0452	5·659 (0·638–50·187)	0·1196	7·513 (1·709–33·025)	0·0076	4·715 (1·028–21·627)	0·0460
Grade								
1 (reference)	..	..	..	..	..	..	..	..
2	3·415 (0·382–30·568)	0·2720	2·987 (0·262–34·126)	0·3785	1·760 (0·438–7·068)	0·4253	1·172 (0·222–6·179)	0·8517
3	7·652 (0·987–59·308)	0·0514	3·546 (0·309–40·663)	0·3092	3·674 (1·063–12·699)	0·0398	1·251 (0·292–5·365)	0·7633
Cluster All (2 *vs* 1)	0·798 (0·240–2·651)	0·7122	0·569 (0·088–3·662)	0·5526	0·770 (0·273–2·177)	0·6224	0·285 (0·039–2·069)	0·2148
Cluster DEG (2 *vs* 1)	3·894 (0·852–17·786)	0·0795	4·207 (0·417–42·476)	0·2233	3·744 (1·065–13·163)	0·0396	5·499 (0·618–48·930)	0·1264

Cox proportional hazards regression analysis of overall and progression-free survival for the 89 patients with stage I EOC from tumour tissue collection A. Cluster analyses performed only on the 51 patients of the training set.

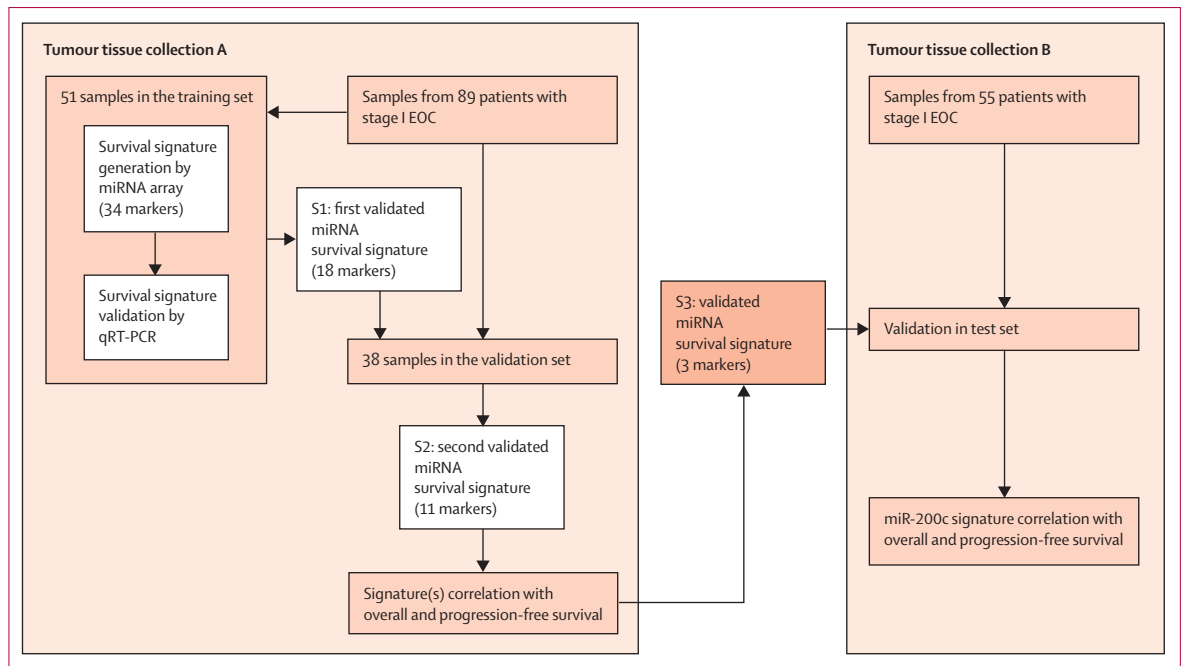*Table 2*: Univariate and multivariate analysis of tumour tissue collection A

Figure 1: **Developing a prognostic miRNA signature in stage I EOC**
A subset of 51 samples from tumour tissue collection A (89 samples of stage I EOC; training set) was used to develop a prognostic micro RNA (miRNA) signature. Microarray analysis identified 18 miRNAs as being differentially expressed in tumour samples from patients who relapsed versus samples from non-relapsing patients. These data were validated by qRT-PCR (survival signature 1, S1) in the training set. In an independent subset made up of the remaining samples from collection A (validation set; n=38), 11 out of 18 miRNAs were confirmed as being differentially expressed in samples from patients who relapsed versus samples from non-relapsing patients (survival signature 2, S2). Of these, only three miRNAs correlated with overall survival (OS) or progression-free survival (PFS) in multivariate analysis (survival signature 3, S3), and therefore their expression was assessed in tumour tissue collection B (55 samples of stage I EOC; test set). Only miR-200c was differentially expressed in samples from patients who relapsed versus samples from non-relapsing patients, and correlated with overall survival and PFS.

miRNA real-time RT-PCR

Mature miRNA expression levels were examined by real-time reverse transcription PCR (RT-PCR) using dedicated sets of commercial primers (Qiagen, Milan, Italy). Briefly, specific cDNA was generated in a single-step reaction from 1 µg of mirVana-purified total RNA (Ambion-ABI, Milan, Italy), using the the miScript Reverse Transcription Kit, following the manufacturer's instructions. miRNA expression was quantified by Sybr Green chemistry using the commercial miScript Universal Primer, together with the miScript Primer Assay (Qiagen). Experiments were run in triplicate for each case to allow for technical variability, using 384-well reaction plates in an automatic liquid handling station (epMotion 5075LH; Eppendorf, Milan, Italy). Real-time PCR was done on an Applied Biosystems 7900HT (Ambion-ABI). Raw data was generated with SDS Relative Quantification software (version 2.3; Ambion-ABI) and analysed using the R software package (version 2.11), as described[3] using *RU-6b* as a housekeeping gene. For miRNA expression experiments, median values were compared using the non parametric Mann-Whitney *t* test. Differences were considered statistically significant with a two sided p-value of less than 0·05. All tests and data plots were done using GraphPad Prism (version 5.01).

qRT-PCR

Quantitative reverse transcription real-time-PCR (qRT-PCR) was used to validate the differential expression of selected genes in EOC biopsies. The automatic liquid-handling station (epMotion 5075LH; Eppendorf), real-time PCR (ABI-7900; Applied Biosystems), and primer pairs sequences for the four housekeeping genes used for data normalisation (actin, *28S*, *GAPDH*, and cyclophilin A) were as previously described.[3] Primer pair sequences for *VEGFA* (NM_001171630): Fw, GGAGGAGGGCAGAATCATCA and Rv, TCCGCATAATCTGCATGGTG; for *TUBB3* (NM_006086): Fw, CGAAGCCAGCAGTGTCTAAA and Rv, GGAGGACGAGGCCATAAATA; for *DICER* (NM_177438): Fw, CGTGCTGCAGTAAGCTGTG and Rv, TCTTCCCTGAGCCAGTGTTT; for *DROSHA* (NM_013235): Fw, AGATCTGGAAGGAGTTACGGTTT and Rv, GGTTGTCACTCCAACGGTCT. For *ZEB1* (NM_030751): Fw, CGGCGCAATAACGTTACAA and Rv, CATCCTCTGGTACACCTTCACA; for *TGFBR2* (NM_003242): Fw, GGGGAAACAATACTGGCTGA and Rv, GAGCTCTTGAGGTCCCTGTG. The annealing temperature was 60°C. Absolute copy numbers were determined by qRT-PCR with the Sybr Green protocol, as previously described.[3] For gene-expression experiments, medians were compared using the non-parametric
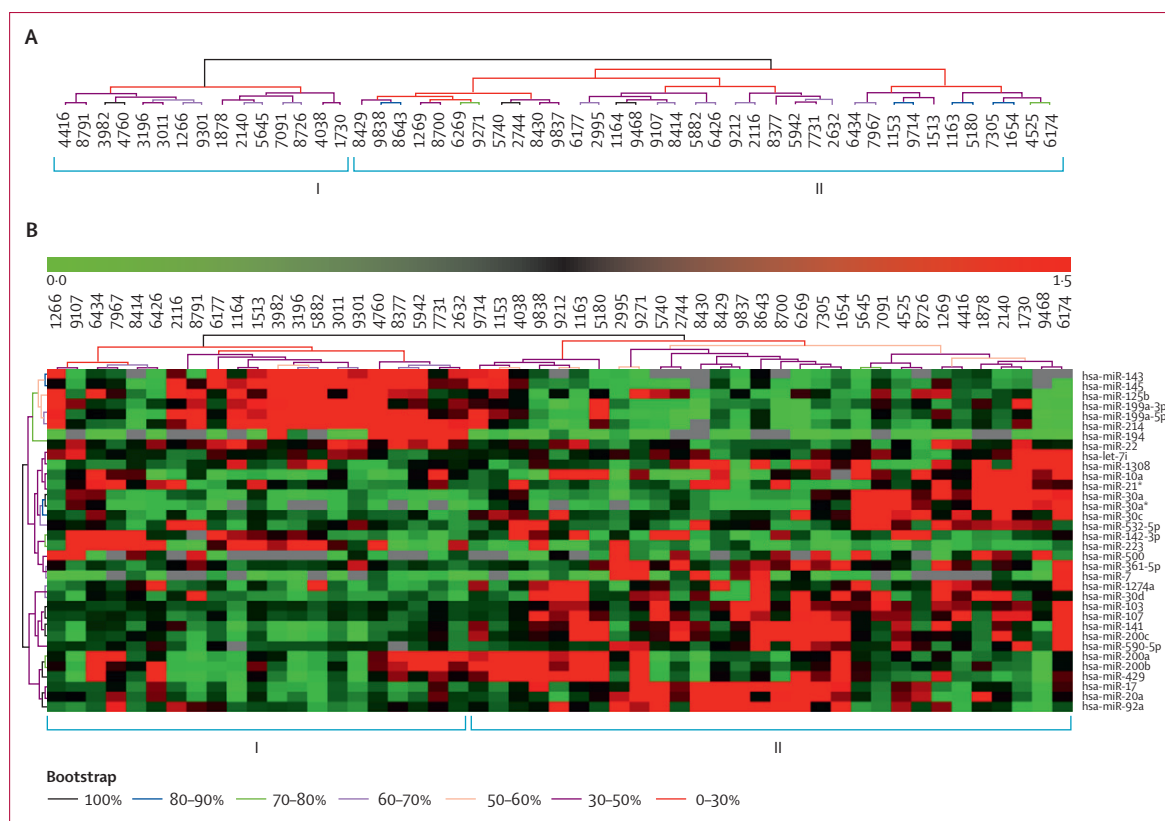
Figure 2: Unsupervised two-dimensional hierarchical cluster analysis of all miRNAs (A) and the miRNA signatures (B)

Each row represents a single miRNA and each column a tumour. The colours of the legend bar indicate the size of gene-expression changes, green for values between 0 and 1, and red for those larger than 1 (numbers beside the bar). The colours of branches, as shown in the legend, represent the bootstrap support of the division in clusters. Bootstrapping is a measure indicating how well the resulting tree is supported by the source data. The black branches divide the set of patients into two clusters for each tree. In panel A, the patients between 4416 and 1730 make up the first cluster, and those patients between 8429 and 6174 form the second cluster. In panel B, the two clusters are made up of patients from 1266 to 2632 and from 9714 to 6174.

Mann Whitney *t* test. Differences were considered statistically significant with a two sided p-value less than 0·05. All tests and data plots were done using GraphPad Prism (version 5.01).

miRNA target gene predictions

To assess possible biological mechanisms involving any identified miRNAs, we used the results of a previously published study on gene-expression profiles (obtained from the same biological samples as for miRNAs expression profiles) in ovarian cancer.[4] We expected that putative targets of identified miRNAs would have been differentially expressed (under the assumption that miRNA alters targets expression). Thus, we first identified putative targets with TargetScan version 5.1,[11] and then retained only those targets previously identified as differentially expressed. The resulting genes were reported according to the length of the alignment (seven or eight base pairs) with the miRNA seed and its level of evolutionary conservation.[12]

Statistical analysis

We assessed the association between clinical features and miRNA expression with overall survival and progression-free survival (PFS). Overall survival was defined as the time from the diagnosis to the date of death from any cause. Patients known to be alive at the time of analysis were censored at their last available contact date. PFS was defined as the time from the diagnosis to the first appearance of relapse or progressive disease. Patients were monitored during follow-up with a physical examination, including pelvic examination, and serum CA-125 measurement. In instances where there were abnormal findings a CT scan was done, and relapse was defined according to RECIST (Response Evaluation Criteria In Solid Tumours)[13] criteria or death from any cause; patients known to be alive and free of progressive disease at the time of analysis were censored at their last available follow-up assessment. The Contal and O'Quigley method[14] was applied (SAS macro was provided by Mandrekar and colleagues) to choose a cutoff value for the expression of each miRNA and mRNA that best discriminates between patients with or without an event during follow-up. Patients were classified according to their miRNA or mRNA expression as under or over the cutoff points computed to discriminate for death (in overall survival) or for progression or death (in PFS). Survival curves were estimated using the

	qRT-PCR				Array	
	Non relapsers (n=36)	Relapsers (n=15)	R	p	Ra	q (%)
Upregulated miRNAs						
miR-223	398·4 (142·8–1182)	1820 (617·9–3730)	4·56	0·0039	2·72	0
miR-125b	29·17 (22·26–63·97)	77·46 (29·83–197·3)	2·65	0·014	1·52	0
miR-200a	148·1 (88·57–211·2)	300·1 (170·4–898·4)	2·02	0·0005	1·70	0
miR-200b	74·90 (41·57–134·1)	217·8 (99·01–367)	2·9	0·014	1·57	0
miR-22	81·82 (59·03–312·1)	145·3 (89·69–250·4)	1·77	0·0023	1·36	0
miR-143	140·8 (47·87–116·3)	8323 (4493–24977)	101·7	<0·0001	1·93	0
miR-214	213·3 (109·8–344·2)	309·6 (221·7–593·5)	1·61	0·031	1·67	24·77
miR-199a-3p	147 (54·97–435·2)	381·2 (203·8–839·5)	2·59	0·012	1·64	24·77
miR-199a-5p	124·5 (60·42–343·8)	207·4 (162·9–476·5)	1·66	0·049	1·68	24·77
miR-145	20·59 (10·29–50·08)	32·05 (20·16–109·4)	1·55	0·024	1·73	24·77
miR-194	6161 (569·9–10574)	8323 (4493–24977)	1·35	0·037	12·12	24·77
let-7i	82·57 (40·51–131·9)	76·40 (48·29–150·9)	0·92	0·693	0·72	24·77
Downregulated miRNAs						
miR-30a	190 (104–243·1)	75·96 (41·99–145·6)	0·39	0·0027	0·48	24·77
miR-30a*	674·3 (426–1709)	431·2 (253·5–715·4)	0·64	0·018	0·43	24·77
miR-30d	250·2 (184·2–385·9)	116 (75·31–263)	0·46	0·026	0·69	24·77
miR-92a	42·14 (14·77–61·57)	17·97 (7·06–38·70)	0·42	0·041	0·73	24·77
miR-141	501·6 (290·6–708·6)	271 (132·8–539·7)	0·54	0·029	0·72	24·77
miR-17	803·9 (416·8–1295)	526·6 (244·5–1283)	0·65	0·35	0·69	24·77
miR-200c	20·22 (10·79–32·80)	6·96 (5·72–17·26)	0·34	0·0071	0·74	24·77
miR-20a	6·81 (5·22–13·01)	4·29 (1·63–7·83)	0·62	0·019	0·72	24·77
miR-361-5p	3754 (1640–6514)	2920 (1090–4838)	0·77	0·25	0·68	24·77

Expression analysis was done in the training set in patients who relapsed compared with those who did not. All data are median distribution (IQR) of fluorescence intensity, normalised, unless otherwise stated. R=the ratio of the median distribution of relapsers to non-relapsers measured by real-time qRT-PCR, as described in the Methods section. Ra=the ratio of relapsers to non-relapsers measured by array technology. p=the level of significance according to the Mann-Whitney t test. q (%) is as defined in the Methods section.

Table 3: RT-PCR expression analysis of selected miRNA from the initial survival signature

See **Online** for webappendix

Kaplan-Meier method, and the results of the log-rank test were added. The Cox proportional hazards model was used to test demographic characteristics, clinical features, and miRNA or mRNA expression for their association with PFS and overall survival. The assumption of proportional hazards was verified with Schoenfeld's residuals. Hazard ratios (HR) with 95% CI were calculated for each factor. When no events occurred in a variable level, the Cox model fails to estimate HR. In this case, "not determined" is reported. In the multivariate model each miRNA (S2 and S3 signature) or mRNA (VEGFA and TUBB3) was inserted together with demographic characteristics and clinical features so as not to overlook the potential confounding role of all the available variables.

A likelihood test was done to assess whether the addition of each miRNA in the Cox proportional hazards multivariate model gave a better fit than the multivariate model with only demographic characteristics and clinical features. The receiver-operating-characteristic (ROC) curve was plotted, and the non-parametric estimate of the area under the ROC curve, the C-statistic, was computed for each model. Because standard methods do not exist for deriving a ROC curve for time-to-event data, we used

occurrence and non-occurrence of events within 10 years as the outcome of these analyses. Differences in the C-statistic were tested to compare the ability to classify risk of models with demographic characteristics, clinical features, and miRNA expression level versus models with only demographic characteristics and clinical features.[15] Statistical significance was set at a p value of less than 0·05. Analysis was done using SAS software (version 9.1).

Role of the funding source

No sponsor was involved in the study design, data collection, data analysis, data interpretation, writing the report, or in the decision to submit for publication. The corresponding author had full access to all the data in the study, and had final responsibility for the decision to submit for publication.

Results

Table 1 shows histopathology of the tumour samples and the number of patients who received post-surgical chemotherapy. Briefly, for tumour tissue collection A, the median age at diagnosis was 52 years (IQR 21–82 years), and the median follow-up time was 9·2 years (IQR 6·9–11·6 years), with 87·5% of patients surviving and 78·5% of patients were progression free at 5 years. Platinum-based adjuvant chemotherapy was given to 61 of 89 patients. For tumour collection B, the median age at diagnosis was 57 years (IQR 48–69 years), and the median follow-up time was 8·9 years (IQR 6·1–11·9 years), with 92·2% of patients surviving and 86·1% of patients were progression free at 5 years. Platinum-based adjuvant chemotherapy was given to 33 of 55 patients. The slight difference in terms of PFS and overall survival between tumour tissue collections A and B (webappendix p 1) could be explained by the fact that collection A includes more patients with stage IC disease, which has a worse prognosis than stage IA and IB (table 1). PFS and overall survival were similar to those reported in the literature with respect to age, substage, grading, and histotype.[1] None of the patients presented with any P53 mutation. 22 of 89 samples in tissue collection A and seven of 55 in tissue collection B came from patients who relapsed, and are referred to from this point onwards as relapsers. Univariate Cox proportional hazards model indicated that age, chemotherapy, and histotype were associated with both overall survival and PFS (table 2). Substage was only significantly associated with PFS (table 2). In the multivariate model, no variable was significantly associated with overall survival, whereas age, chemotherapy, and substage were significantly associated with PFS (table 2).

To identify the entire repertoire of miRNAs in patients with stage I EOC, we did miRNA microarray experiments (figure 1). We profiled the training set (51 samples from tissue collection A, 15 of which were from patients who relapsed), and found 34 miRNAs that were expressed differently in relapsed and non-relapsed patients (webappendix p 2). To investigate similarity across

samples, an unsupervised cluster analysis was done on expression values of all miRNAs (figure 2A) and on differentially expressed miRNAs (figure 2B). The absolute expression values of each miRNA in each patient were divided by the mean of expression of that specific miRNA in all patients. Bootstrap analysis then enabled us to quantify the robustness of the clustering. Using a resampling approach on differentially expressed genes, the similarity structure of the data was perturbed, and the support of the cluster is defined as the number of times the cluster occurs in the perturbed data. The cluster analysis enabled us to divide samples into two groups (characterised by 100% bootstrap support). These groups were further investigated with a Cox proportional hazards model to assess their association with overall survival and PFS (table 2). Univariate analysis of PFS suggests that the signature of 34 miRNAs reported in table 2 as cluster DEG (differentially expressed genes) enabled us to separate patients on the basis of prognosis: the two clusters reported as cluster 1 and 2 (figure 2B) had significantly different prognosis in terms of PFS (table 2). No other association with clinical parameters was detected. By contrast, cluster analysis of all miRNAs (figure 2A) did not discriminate between patients on the global scale (shown in table 2 as Cluster All), suggesting that only a small subset of miRNAs is responsible for differences between relapsers and non-relapsers. This hypothesis was strengthened by association analysis on the two clusters (figure 2B), where there was an indication that one or more miRNAs included in the list could potentially be markers of relapse.

To assess the reproducibility and robustness of the miRNA signature identified in the training set, we measured by qRT-PCR the expression of all 34 miRNAs; however, only the six top ranking miRNAs (q=0) and 50% of the miRNAs with a q value less than 25% were analysable. Experiments were done on both the training and validation set of tumour collection A (figure 1). To correct for a potential batch effect, new batches of snap-frozen material for the entire cohort of samples from tissue collection A were used. qRT-PCR results for the training set indicate the median distribution levels of normalised fluorescence intensity and the IQR of upregulated or downregulated miRNAs in relapsers (n=15) compared with non-relapsers (n=36); Mann-Whitney *t* test confirmed the difference in expression level between relapsers and non-relapsers for 18 of the 21 selected miRNAs (table 3). There were significant differences in the expression levels of the top six ranked miRNAs between relapsers and non-relapsers. By contrast, miR-17, miR-361, and let-7i were not expressed differently in relapsers and non-relapsers, and were excluded from further analysis. The 18 miRNA with confirmed differences in expression between relapsers and non-relapsers were thus selected for further analysis (S1 signature; figure 1). The Pearson correlation index between array and qRT-PCR data was 0·898, suggesting substantial overlap. Box

	Non-relapsers (n=31)	Relapsers (n=7)	R	p
Upregulated miRNAs				
miR-223	699·1 (368–2029)	798·3 (165·7–3165)	1·44	0·73
miR-214	275·8 (127·4–401·7)	1105 (274–4610)	4	0·029
miR-199a-3p	157·4 (64·29–540·7)	593·9 (164·7–1519)	3·77	0·033
miR-199a-5p	252·3 (109·5–450·6)	1025 (574·8–1637)	4·06	0·0084
miR-145	39·38 (19·18–84·62)	130·4 (39·18–347·3)	3·31	0·031
miR-200b	50·65 (38·60–79·44)	120·5 (86·60–127·1)	2·37	0·024
miR-125b	51·67 (24·92–108·9)	68·40 (21·98–356)	1·02	0·66
miR-200a	162 (79·71–259)	149·7 (88·47–259·6)	0·92	0·8
miR-143	200·1 (118·7–313·9)	589·2 (245·3–922·1)	2·94	0·0148
miR-194	5935 (1262–10 601)	4973 (3670–9895)	0·83	0·985
Downregulated miRNAs				
miR-30a	157·7 (82·83–225)	73·23 (53·89–97·84)	0·46	0·014
miR-30a*	896·5 (580·6–1211)	281·6 (197·1–577)	0·31	0·028
miR-30d	288·1 (197·9–376)	145·9 (87·72–185·2)	0·50	0·032
miR-92a	31·92 (22·28–49·95)	37·92 (8·76–51·44)	0·60	0·67
miR-141	370·9 (225·1–664·1)	226·3 (167·8–536)	0·61	0·17
miR-200c	14·95 (11·76–27·06)	7·47 (5·41–15·33)	0·49	0·019
miR-20a	10·97 (8·56–15·52)	6·94 (3·44–10·63)	0·63	0·043
miR-22	136·6 (73·49–211·5)	91·51 (72·23–137·5)	0·67	0·54

All data are median distribution (IQR) of fluorescence intensity, normalised, unless otherwise stated. R=the ratio of the median distribution of relapsers compared with non-relapsers. p=the level of significance according to the Mann-Whitney *t* test.

Table 4: RT-PCR expression analysis of the S2 signature in the validation set

plots showing the distribution of expression levels of the 21 selected miRNAs are shown (webappendix p3–4).

Table 4 shows qRT-PCR results from the validation set (31 samples from non-relapsing and seven samples from relapsed patients). Although the differences in the expression levels mirrored those reported in table 3, only 11 miRNAs from the S1 signature were significantly differently expressed between patients who relapsed and patients who did not. miR-214, miR-199a-3p, miR-199a-5p, miR-145, miR-200b, and miR-143 were confirmed upregulated in tumour tissue samples from relapsers compared with non-relapsers. miR-30a, miR-30a*, miR-30d, miR-200c, and miR-20a were downregulated in samples from relapsers relative to non-relapsers (table 4). We defined this signature of 11 validated miRNAs as S2 (figure 1).

We investigated the association of the S2 miRNAs with survival in collection A (table 5). In a univariate model, patients who expressed high levels of miR-199a-3p and miR-199a-5p had a significantly lower overall survival and PFS than did patients with low expression levels (table 5). Figure 3 shows Kaplan-Meier curves by expression levels of miR-199a-3p and miR-199a-5p, confirming these findings. These associations were confirmed in the multivariate model, where demographic characteristics and clinical features were taken into account: both miR-199a-3p and miR-199a-5p maintained their significant association with PFS, but only miR-199a-3p was significantly associated with overall

	Overall survival				Progression-free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p
miR-214 (high vs low)	2·973 (0·675–13·087)	0·1496	2·028 (0·352–11·684)	0·4287	3·410 (1·009–11·531)	0·0484	1·629 (0·449–5·915)	0·4579
miR-199a-3p (high vs low)	7·832 (1·033–59·382)	0·0464	10·356 (1·056–101·569)	0·0448	8·271 (1·939–35·273)	0·0043	6·427 (1·460–28·287)	0·0139
miR-199a-5p (high vs low)	8·689 (1·144–65·979)	0·0366	8·199 (0·880–76·415)	0·0647	11·798 (1·582–87·963)	0·0161	8·823 (1·136–68·559)	0·0374
miR-145 (high vs low)	0·288 (0·065–1·270)	0·1002	0·140 (0·026–0·752)	0·0219	4·947 (1·151–21·255)	0·0316	2·241 (0·474–10·589)	0·3085
miR-200b (high vs low)	2·137 (0·801–5·701)	0·1292	2·051 (0·640–6·570)	0·2264	3·197 (1·417–7·213)	0·0051	2·335 (0·857–6·363)	0·0974
miR-30 (high vs low)	nd	nd	nd	nd	0·322 (0·140–0·736)	0·0073	0·694 (0·239–2·015)	0·5022
miR-30a* (high vs low)	0·286 (0·081–1·004)	0·0506	0·500 (0·112–2·228)	0·3635	0·362 (0·160–0·817)	0·0144	0·931 (0·310–2·797)	0·8981
miR-30d (high vs low)	0·444 (0·165–1·195)	0·1080	0·640 (0·187–2·184)	0·4758	0·332 (0·142–0·779)	0·0113	0·595 (0·200–1·775)	0·3521
miR-200c (high vs low)	0·309 (0·112–0·850)	0·0230	0·244 (0·076–0·785)	0·0180	0·392 (0·174–0·885)	0·0241	0·419 (0·146–1·204)	0·1063
miR-20 (high vs low)	0·376 (0·141–1·006)	0·0513	0·367 (0·115–1·172)	0·0906	0·356 (0·159–0·801)	0·0125	0·392 (0·142–1·080)	0·0700
miR-143 (high vs low)	3·823 (1·089–13·425)	0·0364	2·811 (0·633–12·488)	0·1743	3·945 (1·455–10·701)	0·0070	2·234 (0·728–6·859)	0·1600

nd=not determined.

Table 5: Cox proportional hazards regression analysis of the relation between the S2 miRNAs and overall and progression-free survival for the 89 patients with stage I EOC from tumour tissue collection A

survival (table 5). The Cox proportional hazards model explained the survival data significantly better with the inclusion of miR-199a-3p and miR-199a-5p than without these miRNAs, both in analysis of overall survival (for miR-199a-3p the likelihood ratio [LR] was 7·12, p=0·0076; and for miR-199a-5p LR was 5·74, p=0·0165) and PFS (for miR-199a-3p LR was 9·47, p=0·0021; for miR-199a-5p LR was 7·94, p=0·0048). Adding miR-199a-3p to the multivariate model improved the C-statistic by 5·80% (from 0·82, 95% CI 0·70–0·95, to 0·87, 0·75–0·99) for overall survival and by 4·70% (from 0·83, 0·73–0·92, to 0·87, 0·77–0·96) for PFS. Adding miR-199a-5p improved the C-statistic by 4·73% (from 0·82, 0·70–0·95, to 0·86, 0·74–0·98) for overall survival and by 3·67% (from 0·83, 0·73–0·92, to 0·86, 0·76–0·95) for PFS. These improvements were not found to be statistically significant.

The expression level of miR-200c was found to be significantly associated with both overall survival and PFS in univariate analysis (table 5). Patients with high expression levels of miR-200c survived for longer than did those with low levels; similarly, patients with high expression levels of miR-200c had a longer PFS than did those with low levels (table 5). The Kaplan-Meier curves in figure 3 confirm these results: patients with high expression levels of miR-200c had a better prognosis than those with low levels, both in terms of overall survival and PFS. A multivariate Cox proportional hazards model was fitted to test the effect of expression level on survival, taking into account the effect of demographic characteristics and clinical features. miR-200c maintained its significance as an independent prognostic factor for overall survival, but not for PFS (table 5). Including miR-200c significantly improved the fit of the Cox proportional hazards model in an analysis of overall survival (LR 5·79, p=0·0161), but not in an analysis of

PFS (LR 2·57, p=0·1088). Adding miR-200c to the multivariate model improved the C-statistic by 2·79% (from 0·82, 95% CI 0·70–0·95, to 0·84, 0·72–0·97) for overall survival, and by 1·17% (from 0·83, 0·73–0·92, to 0·84, 0·74–0·93) for PFS. However, these improvements were not significant.

RT-PCR was applied to miR-199a-3p, miR-199a-5p, and miR-200c (S3 signature) in tissue tumour collection B (table 6). The Cox proportional hazards model was then used to assess the effect on PFS and overall survival of demographic characteristics (age), clinical features (chemotherapy-treated or untreated, histotype, grade and substage) and the levels of miRNA expression in miR-200c, which was the only miRNA validated by qRT-PCR. miR-200c had a significant effect on survival (table 7). In the univariate model, patients with high expression levels of miR-200c had a higher probability of overall survival and PFS than did patients with low expression levels (table 7). Kaplan-Meier survival estimates are shown in figure 4: patients with high expression levels of miR-200c had a better prognosis than did those with low levels, both in terms of overall survival and PFS. miR-200c was an independent prognostic factor in the multivariate model in terms of both overall survival and PFS (HR 0·094; 95%CI 0·012–0·766, p=0·0272 for overall survival; and HR 0·035; 0·004–0·311, p=0·0026 for PFS; table 7). The likelihood ratio test assessed the fit of the Cox proportional hazards model considering miR-200c in the analysis both of overall survival (LR 6·09, p=0·0136) and PFS (LR 12·61, p=0·0004); adding miR-200c to the multivariate model improved the C-statistic by 6·87% (from 0·78, 95% CI 0·62–0·95, to 0·84, 0·68–0·99) for overall survival, and by 10·03% (from 0·79, 0·64–0·94, to 0·87, 0·71–1·00) for PFS. These improvements were not statistically significant.

To investigate events downstream of miR-200c downregulation in relapsing patients in more detail, we did in-silico screening to obtain a list of putative targets of miR-200c among differentially expressed genes that were detected in the same group of patients previously.[4] Because the computational approaches currently used to predict miRNA–mRNA interactions have low precision and sensitivity, we compared in-silico predictions with high-throughput measurements of gene expression.[16] We focused on a subset of 188 genes identified previously as being differentially expressed in relapsers versus non-relapsers in the same cohort of patients.[4] We used the TargetScan algorithm to search for genes with the right features to be putative downstream targets of miR-200c, and we identified 11 genes that had at least one conserved response element in their 3′ untranslated regions
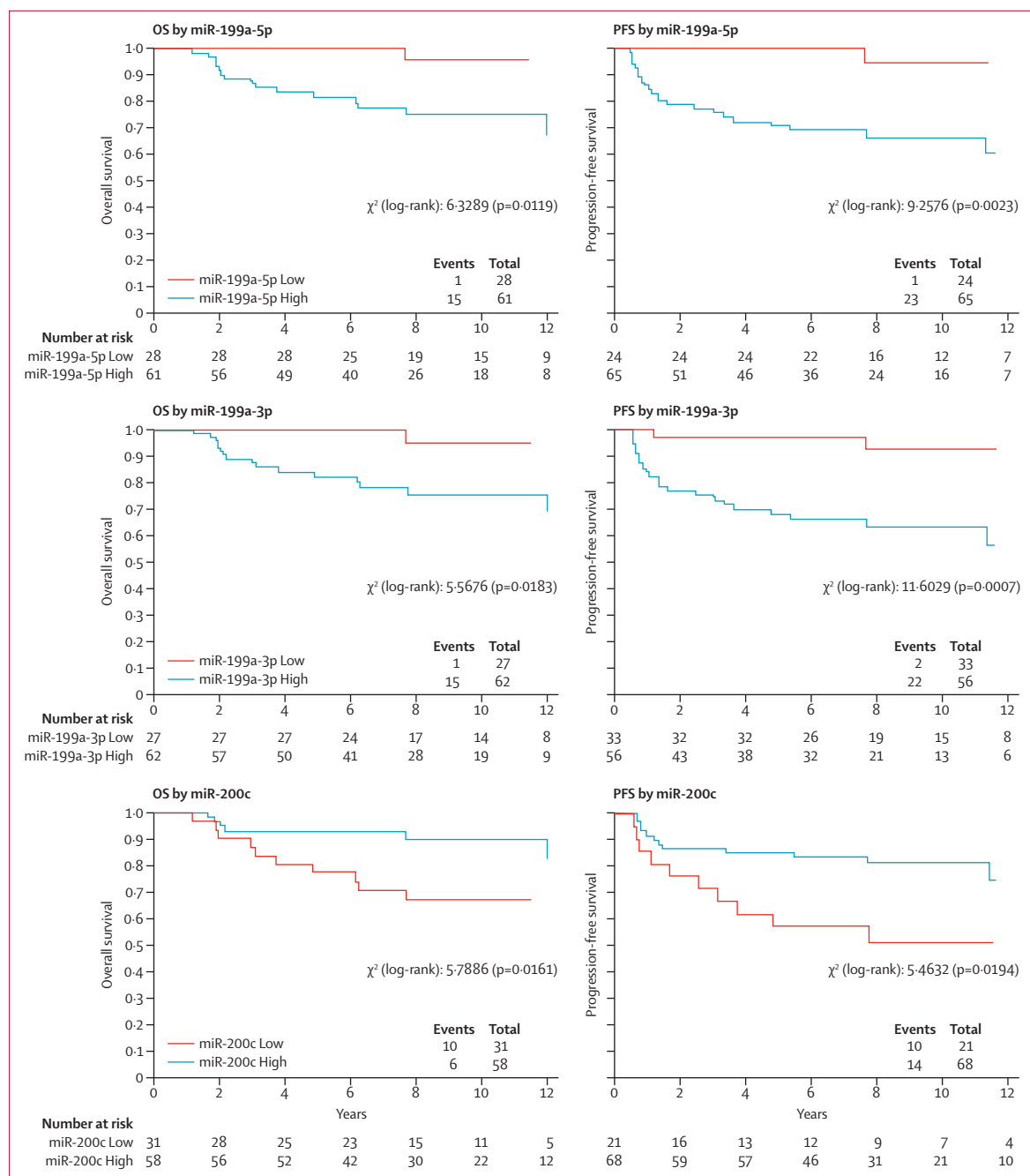

Figure 3: Kaplan-Meier curves of independent prognostic factors for overall and progression-free survival for the S3 signature in tumour collection A
miRNAs expression levels were converted into discrete variables by dividing the samples into two classes (high and low), under or over the cutoff points. Results of the log-rank test are shown.

	Non-relapsers (n=48)	Relapsers (n=7)	R	p
miR-199a-3p	212·1 (45·55–848·6)	543·9 (271–952·4)	2·56	0·33
miR-199a-5p	63·38 (28·31–211·9)	197·2 (59·47–239·3)	3·11	0·19
miR-200c	9·34 (4·17–18·56)	3·82 (3·38–5·93)	0·40	0·047

All data are median distribution (IQR) of fluorescence intensity, normalised, unless otherwise stated. R=the ratio of the median distribution of relapsers compared with non-relapsers. p=the level of significance according to the Mann-Whitney t test.

Table 6: RT-PCR expression analysis of the S3 signature

(webappendix p 5). We focused on *VEGFA* and tubulin III b isotype (*TUBB3*), the expression levels of which have been validated independently by qRT-PCR (webappendix p 6). In agreement with previous data on tumour tissue collection A,[4] the expression of both *VEGFA* and *TUBB3* in relapsing patients is about double that in non-relapsers. Median expression levels for *VEGFA* were 0·0029 and 0·0071 molecules of mRNA normalised for non-relapsers and relapsers, respectively (p=0·003, webappendix p 6); median expression levels for *TUBB3*

	Overall survival				Progression free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p	HR (95% CI)	p	HR (95%CI)	p	HR (95%CI)	p
Age (5-year increment)	1·281 (0·954–1·719)	0·0996	1·198 (0·828–1·733)	0·3370	1·179 (0·906–1·533)	0·2197	1·029 (0·755–1·403)	0·8541
Chemotherapy (yes vs no)	0·527 (0·159–1·746)	0·2948	0·539 (0·107–2·713)	0·4538	0·706 (0·225–2·209)	0·5491	1·197 (0·265–5·395)	0·8152
Histotype								
Clear cell	Reference				Reference			
Endometroid	0·663 (0·068–6·476)	0·7237	0·412 (0·037–4·590)	0·4710	0·674 (0·069–6·582)	0·7342	0·477 (0·043–5·306)	0·5466
Mucinous	1·185 (0·117–11·951)	0·8857	1·030 (0·081–13·028)	0·9817	1·209 (0·120–12·196)	0·8719	1·562 (0·135–18·100)	0·7211
Mullerian	nd	nd	nd	nd	nd	nd	nd	nd
Serous	2·406 (0·262–22·072)	0·4374	1·497 (0·138–16·192)	0·7399	2·793 (0·320–24·407)	0·3531	2·547 (0·254–25·524)	0·4265
Undifferentiated	nd	nd	nd	nd	nd	nd	nd	nd
Figo I substage								
A	Reference				Reference			
B	0·427 (0·051–3·586)	0·4333	0·246 (0·017–3·552)	0·3032	0·361 (0·044–2·965)	0·3431	0·240 (0·021–2·763)	0·2524
C	0·689 (0·194–2·450)	0·5652	0·516 (0·073–3·667)	0·5082	0·550 (0·161–1·888)	0·3424	0·274 (0·045–1·656)	0·1582
Grade								
1	Reference				Reference			
2	1·260 (0·221–7·186)	0·7950	4·177 (0·475–36·737)	0·1974	0·951 (0·179–5·048)	0·9527	1·931 (0·304–12·262)	0·4854
3	1·872 (0·490–7·152)	0·3593	10·650 (1·269–89·377)	0·0293	1·408 (0·399–4·969)	0·5950	5·041 (0·762–33·350)	0·0934
miR-200c (high vs low)	0·209 (0·045–0·970)	0·0456	0·094 (0·012–0·766)	0·0272	0·174 (0·038–0·793)	0·0239	0·035 (0·004–0·311)	0·0026

Cox proportional hazards regression analysis of overall survival and PFS for 55 patients with stage I EOC from tumour tissue collection B. nd=not determined.

Table 7: Univariate and multivariate analysis of tumour tissue collection B
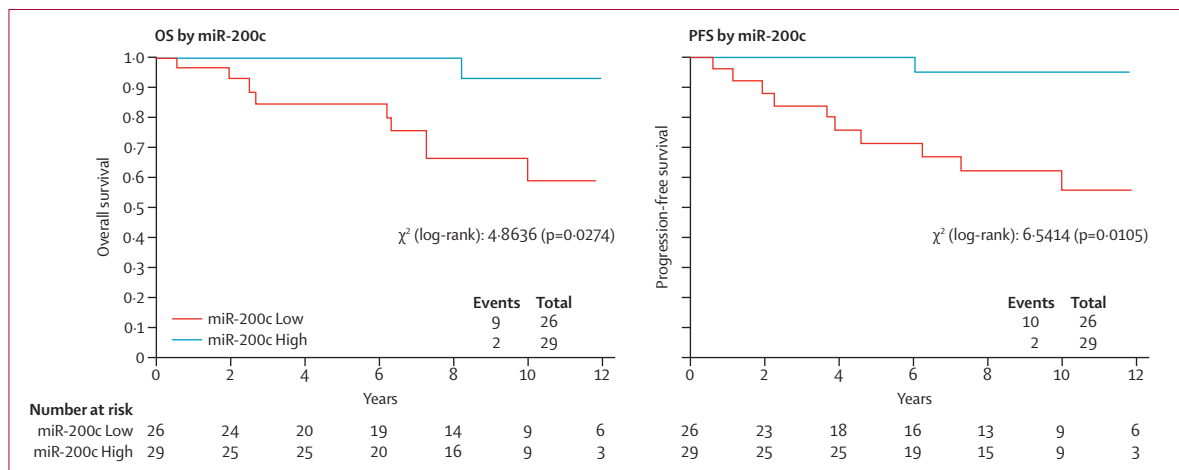

Figure 4: Kaplan-Meier curves of independent prognostic factors for overall and progression-free survival for miR-200c in tumour collection B
miRNAs expression levels were converted into discrete variables by dividing the samples into two classes (high and low), under or over the cutoff points. Results of the log-rank test are shown.

	Overall survival				Progression-free survival			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p	HR (95% CI)	p
Tissue collection A								
VEGFA (high *vs* low)	5·033 (1·827–13·865)	0·0018	5·560 (1·538–20·103)	0·0089	5·615 (1·668–18·902)	0·0053	5·843 (1·556–21·950)	0·0089
TUBB3 (high *vs* low)	1·998 (0·725–5·500)	0·1806	2·312 (0·722–7·402)	0·1582	2·081 (0·930–4·655)	0·0744	2·735 (1·039–7·198)	0·0416
Tissue collection B								
VEGFA (high *vs* low)	3·683 (0·793–17·101)	0·0960	2·900 (0·554–15·180)	0·2074	4·590 (1·005–20·964)	0·0492	4·842 (0·941–24·913)	0·0591

Data for TUBB3 in tumour tissue collection B is not reported, because TUBB3 did not have a significant association with overall and pregression-free survival in univariate analysis of tumor tissue collection A.

Table 8: Univariate and multivariate analysis of association between VEGFA and TUBB3 expression and survival in the two tissue collections

were 0·0025 and 0·0068 for non-relapsers and relapsers, respectively (p=0·00097, webappendix p 6). Results relating to *VEGFA* were confirmed when analysed by association with survival in both tissue collections A and B. In tissue collection A, univariate analysis suggested that patients with a high level of *VEGFA* expression had significantly higher hazards ratios both for overall survival and progression-free survival (table 8); Kaplan-Meier analyses suggested that patients with low levels of expression of *VEGFA* had a better prognosis than those with higher levels, both in terms of overall survival and PFS (webappendix p 7), and multivariate analysis including demographic and clinical factors showed that *VEGFA* expression was independently associated with overall survival and progression-free survival (table 8). In tissue collection B, differences in *VEGFA* expression levels were confirmed between relapsers and non-relapsers (median expression levels for *VEGFA* were 0·00018 and 0·0011 for non-relapsers and relapsers, respectively, p=0·0035; webappendix p 6). Univariate analysis and Kaplan-Meier analyses suggested that patients with higher expression levels of *VEGFA* had a worse prognosis in terms of PFS than those with lower levels (table 8; webappendix p 8). However, *VEGFA* expression was not a significant prognostic factor in the multivariate analyses (table 8).

Discussion

This study suggests that the expression of miR-200c in patients with stage I EOC correlates with disease outcome. FIGO stage is used to define the spread of the disease at diagnosis, and it is currently one of the most important prognostic factors for survival in EOC. Patients with stage I disease generally have good prognosis, whereas stages III–IV are associated with poor overall survival, and shorter PFS within 5 years of diagnosis. The risk of recurrence is higher in patients with certain clinical and histopathological features, such as grade 3, substage C, or clear-cell histotype. Adjuvant chemotherapy can prolong survival in these subgroups, although relapses in theoretically low-risk patients are common.[1] The challenge at present is to identify new molecular features with prognostic significance to distinguish which patients are at high risk and would benefit most from additional medical intervention. A problem is that stage I EOC is not common: less than 10% of patients with EOC are diagnosed as stage I every year, which makes statistical and clinical interpretation difficult. An additional level of complexity is generated by the controversial relationship between stages. Mirroring models from other solid tumours (eg, colon cancer), stage I EOC has long been considered an early phase in tumour progression. However, recent data on the possible pathogenesis of EOC strongly suggests that stage I disease, especially non-relapsing, might be a completely different disease, not just an early step in the progression of EOC toward malignancy. However, this theory has not yet been generally accepted.[17]

This heterogeneous clinical and pathological scenario mirrors, at the molecular level, the complex genomic rearrangements and structural variations observed in the ovarian cancer genome, which has until now made it difficult to exploit genome-wide information to stratify patients more precisely for diagnosis and prognosis. miRNAs now offer a new regulatory model of gene expression, and miRNA expression signatures correlate well with specific clinical characteristics of cancer, so that they can be used to classify normal and cancerous tissues, as well as for prognosis.[18–20] Despite partial overlap between different studies, data on EOC thus far indicates that the miRNA landscape is very important to understand ovarian cancer biology and therapy.[21] miRNAs have been found to be downregulated in EOC cells compared with normal surface epithelial cells of the ovary.[22] The clinical and pathological significance has been documented, mainly in stage III–IV disease with sensitivity or resistance

Panel: Research in context

Systematic review

A systematic review using search terms such as "stage I EOC AND miRNA" returned no results. To our knowledge, no data were available on miRNA expression profile in stage I EOC

Interpretation

If our findings are confirmed, miR-200c expression levels could be used to stratify patients with stage I ovarian cancer more accurately, which would enable the most appropriate therapies to be selected according to risk of relapse. Our findings also provide the rationale to investigate new therapies in selected cohorts of patients with stage I ovarian cancer.

to platinum-based chemotherapy. miRNA loci are targeted by genetic and epigenetic defects,[23] and altered expression of proteins involved in miRNA maturation, such as DICER and DROSHA, are associated with most miRNAs alterations reported in stage III–IV disease.[6]

To the best of our knowledge, no data are available on miRNAs in stage I EOC, particularly in relation to prognosis. To identify miRNAs that are differentially expressed in relapsing and non-relapsing patients, our study made use of tumour samples from a cohort of 144 patients (29 of whom relapsed) with stage I EOC, gathered from two independent tumour tissue collections. All samples were taken from patients who were naive to chemotherapy presurgery, and who had a minimum of 5-years follow-up (median follow-up was 9 years). The patients from tumour tissue collection A had already been characterised by our group as lacking any mutation in the *TP53* gene, and a gene-expression signature of 188 markers was associated with relapse.[4] We found that the loss of miR-200c is a hallmark of poor prognosis in stage I EOC, and the loss of miR-200c parallels increased expression levels of *VEGFA*, which is a putative functional downstream miR-200c target that correlates with PFS and overall survival in multivariate analysis.

By contrast with stages III–IV EOC, the small proportion of patients who are diagnosed with stage I disease has made it difficult to recruit cohorts of patients that are large enough to achieve suitable statistical power. In this context, the miRNAs that we found to be significantly associated with overall survival and PFS might be an important part of a larger set of predictive factors. Further studies stemming from this first attempt are likely to add to the list of miRNAs with prognostic value.

Our data suggest that a small subset of miRNAs correlate with prognosis, although marked differences in miRNA expression profiles were not noted. This conclusion is consistent with the idea that a small set of tumour features can cause differences in aggressiveness and outcome. The fact that gross miRNA alterations were not detected is in accordance with the finding that DICER and DROSHA levels were comparable in the relapsing and non-relapsing patients (webappendix p 6).

Of the 34 top-ranked differentially expressed miRNAs defined by the microarray analysis as a signature of relapse, we focused our attention on 21 miRNAs for which qRT-PCR amplification parameters enabled us to draw conclusions. miR-200c was correlated independently with survival in the two independent datasets, and in line with previous data[24] the miR-199 family was found upregulated in tumour tissue collection A. We cannot draw any conclusion at the moment about the remaining 13 miRNAs that have yet to be analysed; this point will be addressed in future studies.

miR-200c expression levels were independently associated with survival in the multivariate model. Thus, despite patients' different genetic backgrounds and histologically complex characteristics, miR-200c is linked to general mechanisms of relapse. The miR-200 family consists of five members localised on two genomic clusters (miR-200a/b and miR-429 on chromosome 1, and miR-200c and miR-141 on chromosome 12).[25] Interestingly, high levels of expression of all miR-200 family members have been linked with ovarian cancer in several studies.[21,26] Moreover, high expression levels of the miR-200 family are associated with shorter PFS and overall survival of patients with ovarian cancer.[26] Although the expression levels of miR-200a and miR-200b might be lower in stage III than in stage I tumours,[24] a high level of miR-200a expression in stage III ovarian cancer correlates with poor outcome.[27]

From a functional point of view, several targets for miR-200c have already been identified. In many cancers, including EOC stages III/IV, miR-200c has been identified as a marker of epithelial cells, and a powerful master regulator of epithelial-to-mesenchymal transition, through a negative-feedback loop with the transcription factors ZEB1 and TGFBR2. In our study, loss of miR-200c did not affect the expression of either of these (webappendix p 6), suggesting that this miRNA functions differently in stage I disease. Current bioinformatic methods to predict miRNAs targets have low sensitivity and precision, providing a large number of potential targets, many of which are probably false-positive.[28] To narrow down the list of potential genes of interest, we used the intersection of miR-200c predicted targets with a list of genes previously identified and differently expressed in relapsing patients.[4] Among the downstream targets of miR-200c, the *VEGFA* and *TUBB3* genes were significantly upregulated in patients who relapsed. High *VEGFA* mRNA expression levels were also associated in both univariate and multivariate analyses with reduced overall survival and PFS. Although we have not provided a mechanistic link between reduced miR-200c expression levels and increased VEGFA mRNA, our preliminary results are in line with recently published data.[29] Liu and colleagues[29] showed in two independent data sets of clear-cell renal-cell carcinoma, and in normal tissues, a strong anti-correlation between miR-200c expression levels and VEGFA mRNA levels. This point merits further study, particularly considering the potential relevance of this target for new therapeutic approaches in stage I EOC.

Although our findings need to be validated in a larger cohort, this study shows in unprecedented resolution and detail the transcriptional regulation mechanisms that drive stage I EOC characteristics. Assessing miR-200c or VEGFA levels might serve to stratify patients for alternative treatments such as antibodies against VEGF, which has recently been shown to be useful in the treatment of advanced ovarian cancer.[30] The other potential miR-200c target that was downregulated was *TUBB3*. This gene encodes class III b tubulin, which is a direct target of miR-200c, and the expression of which has been shown to be associated with aggressive and drug-resistant cancers.[31] In advanced stages of EOC, levels of *TUBB3* expression correlate with poor outcome.[32] Consistent with recent results that suggest an inverse relationship between

miR-200c and *TUBB3* expression in advanced EOC,[33] our study suggests that *TUBB3* overexpression might be a negative prognostic marker in stage I EOC. Further studies using different techniques (eg, in-situ hybridisation or RT-PCR coupled with laser capture microdissection) are clearly needed to establish whether miR-200c has the potential to be a prognostic biomarker in stage I EOC. As soon as the technology has been validated, it would be interesting to measure miR-200c levels in peripheral blood as a marker for screening and prognostic purposes.[34] If our findings can be confirmed in other studies, miR-200c might serve as a molecular marker for the prediction of patient response to standard treatment (platinum-based or paclitaxel-based chemotherapy), which might reduce cost and minimise drug toxicity for those who do not benefit from the treatment. Additionally, new investigational and anti-angiogenic agent combinations could be proposed earlier in the treatment algorithm for those patients.

References
1 Cannistra SA. Cancer of the ovary. *N Engl J Med* 2004; **351:** 2519–29.
2 Marabese M, Marchini S, Marrazzo E, et al. Expression levels of p53 and p73 isoforms in stage I and stage III ovarian cancer. *Eur J Cancer* 2008; **44:** 131–41.
3 Marchini S, Marabese M, Marrazzo E, et al. DeltaNp63 expression is associated with poor survival in ovarian cancer. *Ann Oncol* 2008; **19:** 501–07.
4 Marchini S, Mariani P, Chiorino G, et al. Analysis of gene expression in early-stage ovarian cancer. *Clin Cancer Res* 2008; **14:** 7850–60.
5 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; **116:** 281–97.
6 Merritt WM, Lin YG, Han LY, et al. Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med* 2008; **359:** 2641–50.
7 Trimbos JB, Parmar M, Vergote I, et al. International Collaborative Ovarian Neoplasm trial 1 and Adjuvant ChemoTherapy In Ovarian Neoplasm trial: two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma. *J Natl Cancer Inst* 2003; **95:** 105–12.
8 Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19:** 185–93.
9 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100:** 9440–45.
10 Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003; **34:** 374–78.
11 Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; **120:** 15–20.

12 Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; **19:** 92–105.
13 Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000; **92:** 205–16.
14 Williams BA, Mandrekar JN, Mandrekar SJ, Cha SS, Furth AF. Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes. Rochester, MN: Mayo Clinic; 2006.
15 Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148:** 839–43.
16 Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005; **433:** 769–73.
17 Shih Ie M, Kurman RJ. Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am J Pathol* 2004; **164:** 1511–18.
18 Cummins JM, Velculescu VE. Implications of micro-RNA profiling for cancer diagnosis. *Oncogene* 2006; **25:** 6220–27.
19 Dalmay T, Edwards DR. MicroRNAs and the hallmarks of cancer. *Oncogene* 2006; **25:** 6170–75.
20 Tricoli JV, Jacobson JW. MicroRNA: potential for cancer detection, diagnosis, and prognosis. *Cancer Res* 2007; **67:** 4553–55.
21 van Jaarsveld MT, Helleman J, Berns EM, Wiemer EA. MicroRNAs in ovarian cancer biology and therapy resistance. *Int J Biochem Cell Biol* 2011; **42:** 1282–90.
22 Iorio MV, Visone R, Di Leva G, et al. MicroRNA signatures in human ovarian cancer. *Cancer Res* 2007; **67:** 8699–707.
23 Zhang L, Huang J, Yang N, et al. microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci USA* 2006; **103:** 9136–41.
24 Eitan R, Kushnir M, Lithwick-Yanai G, et al. Tumor microRNA expression patterns associated with resistance to platinum based chemotherapy and survival in ovarian cancer patients. *Gynecol Oncol* 2009; **114:** 253–59.
25 Korpal M, Kang Y. The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. *RNA Biol* 2008; **5:** 115–19.
26 Nam EJ, Yoon H, Kim SW, et al. MicroRNA expression profiles in serous ovarian carcinoma. *Clin Cancer Res* 2008; **14:** 2690–95.
27 Hu X, Macdonald DM, Huettner PC, et al. A miR-200 microRNA cluster as prognostic marker in advanced ovarian cancer. *Gynecol Oncol* 2009; **114:** 457–64.
28 Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* 2009; **25:** 3049–55.
29 Liu H, Brannon AR, Reddy AR, et al. Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. *BMC Syst Biol* 2010; **4:** 51.
30 Burger RA, Brady MF, Bookman MA, et al. Phase III trial of bevacizumab (BEV) in the primary treatment of advanced epithelial ovarian cancer (EOC), primary peritoneal cavity (PPC), or fallopian tube cancer (FTC): a Gynecologic Oncology Group study. *Proc Am Soc Clin Oncol* 2010; **28:** abstr 5015.
31 Cochrane DR, Spoelstra NS, Howe EN, Nordeen SK, Richer JK. MicroRNA-200c mitigates invasiveness and restores sensitivity to microtubule-targeting chemotherapeutic agents. *Mol Cancer Ther* 2009; **8:** 1055.
32 Ferrandina G, Zannoni GF, Martinelli E, et al. Class III beta-tubulin overexpression is a marker of poor clinical outcome in advanced ovarian cancer patients. *Clin Cancer Res* 2006; **12:** 2774–79.
33 Leskela S, Leandro-Garcia LJ, Mendiola M, et al. miR-200 family controls {beta}-tubulin III expression and is associated with paclitaxel-based treatment response and progression-free survival in ovarian cancer patients. *Endocr Relat Cancer* 2010; **18:** 85–95.
34 Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* 2008; **110:** 13–21.