

RESEARCH

Open Access

OvMark: a user-friendly system for the identification of prognostic biomarkers in publically available ovarian cancer gene expression datasets

Stephen F Madden^{1*†}, Colin Clarke^{1†}, Britta Stordal³, Mark S Carey⁷, Russell Broaddus⁶, William M Gallagher², John Crown¹, Gordon B Mills⁵ and Bryan T Hennessy⁴

Abstract

Background: Ovarian cancer has the lowest survival rate of all gynaecologic cancers and is characterised by a lack of early symptoms and frequent late stage diagnosis. There is a paucity of robust molecular markers that are independent of and complementary to clinical parameters such as disease stage and tumour grade.

Methods: We have developed a user-friendly, web-based system to evaluate the association of genes/miRNAs with outcome in ovarian cancer. The *OvMark* algorithm combines data from multiple microarray platforms (including probesets targeting miRNAs) and correlates them with clinical parameters (e.g. tumour grade, stage) and outcomes (disease free survival (DFS), overall survival). In total, *OvMark* combines 14 datasets from 7 different array platforms measuring the expression of ~17,000 genes and 341 miRNAs across 2,129 ovarian cancer samples.

Results: To demonstrate the utility of the system we confirmed the prognostic ability of 14 genes and 2 miRNAs known to play a role in ovarian cancer. Of these genes, CXCL12 was the most significant predictor of DFS (HR = 1.42, p-value = 2.42×10^{-6}). Surprisingly, those genes found to have the greatest correlation with outcome have not been heavily studied in ovarian cancer, or in some cases in any cancer. For instance, the three genes with the greatest association with survival are SNAI3, VWA3A and DNAH12.

Conclusions/Impact: *OvMark* is a powerful tool for examining putative gene/miRNA prognostic biomarkers in ovarian cancer (available at <http://glados.ucd.ie/OvMark/index.html>). The impact of this tool will be in the preliminary assessment of putative biomarkers in ovarian cancer, particularly for research groups with limited bioinformatics facilities.

Background

Ovarian cancer is the most lethal gynecological malignancy. Due to its lack of early symptoms, this disease is usually diagnosed at an advanced stage when the cancer has already spread to secondary sites [1]. While initial rates of response to first treatment are >80%, the long-term survival rate of patients is very low, mainly due to development of drug resistance [1]. Clinical parameters such as disease stage and tumor grade are important considerations in the management of patients after their initial surgery to establish the necessity for chemotherapy [2]. The amount of residual tumour after surgery is another very important

determinant of prognosis. However, reliable markers that are independent of and complementary to clinical parameters are needed for better prognostic stratification of patients and for individualisation of therapy.

For several years now, efforts to identify novel prognostic and predictive factors have focused on molecular markers, with a large number having been identified and investigated [3]. However, while there is evidence that BRCA mutations correlating with the selection of patients for treatment with PARP inhibitors, known ovarian cancer biomarkers are not sufficiently prognostic or predictive, at least for clinical use [3,4]. It is clear that the selection of molecular markers could benefit greatly from the integration of datasets across multiple studies to increase confidence in the selected markers, by substantially improving the statistical power and robustness of the analysis. To this end, we have developed an easy-to-use algorithm (called *OvMark*) which

* Correspondence: stephen.madden@dcu.ie

†Equal contributors

¹Molecular Therapeutics for Cancer Ireland, National Institute for Cellular Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland
Full list of author information is available at the end of the article

allows the association of gene(s) of interest with patient outcomes in ovarian cancer. This algorithm integrates gene expression data from 2,129 patients in 14 DNA microarray studies and corresponding clinical (stage, grade, amount of residual disease after surgery, outcomes etc.) and treatment (chemotherapy) data. Among all ovarian cancers included in *OvMark*, the vast majority are grade 2/3 (high-grade) serous cancers, which compose more than 80% of ovarian cancers that occur in women. The other ovarian cancers in *OvMark* represent less common epithelial ovarian cancer subtypes that are biologically distinct from high-grade serous ovarian cancers, mainly grade 1 (low-grade) serous cancers and low and high-grade endometrioid ovarian cancers. The user can stratify samples based on subtype and these clinical data for investigation of prognostic markers in the context of disease-free survival (DFS) and overall survival (OS). A similar approach has been “previously used” in integrating gene expression and detailed clinical data including survival information in breast cancer [5,6].

The association of small non-coding RNAs known as microRNAs (miRNAs) with ovarian cancer has been well established [7]. Over the last decade, our understanding of the function that miRNAs play in ovarian and other cancers as well as an array of fundamental biological processes in both plants and animals has increased dramatically [8]. These short endogenous non-coding RNAs act primarily by negatively regulating the expression of target mRNAs through translational inhibition and/or mRNA degradation [8]. Approximately 50% of known human miRNAs are intronic (miRBase release 20, June 2013). Of these, 341 or roughly one third of human miRNA host genes are hybridized by probes on the U133plus2 affymetrix gene chip. A number of studies have reported that many intronic miRNAs show significantly correlated expression profiles with their host genes [9,10]. Estimates of the number of miRNAs whose expression profiles are significantly correlated with their host gene are as high as 70% [11]. The expression of miRNAs can, in some instances, be inferred from the expression of their host genes, and we have therefore included these miRNAs in our algorithm *OvMark* to allow their evaluation as putative prognostic markers in ovarian cancer from gene expression data.

We confirm the utility of this approach following comparisons of outputs with a panel of 14 previously identified markers of prognosis in ovarian cancer. In addition, we also conducted an unbiased analysis of all genes present in the dataset to assess their prognostic potential and found that the list of the most significant genes is dominated by novel markers of prognosis in ovarian cancer. The feasibility of using miRNA host gene expression as a surrogate for miRNA levels was confirmed using the known miRNA prognostic markers,

let-7f-2 [12] and miR-16-1 [13]. Although these markers were previously examined in smaller patient cohorts, *OvMark* was able to confirm the robustness of these prognostic marker across a large and diverse patient dataset.

Our novel user-friendly algorithm *OvMark* is a powerful tool for examining putative gene/miRNA prognostic biomarkers in ovarian cancer. The value of this tool will be in the preliminary assessment of biomarkers in ovarian cancer, particularly for research groups with limited bioinformatics facilities.

Methods

Gene expression data

Gene expression data sets were downloaded from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) in the form of raw data files, where possible. Only ovarian cancer datasets with survival information and at least 50 patients were included. In total, 2,129 samples across 14 datasets incorporating 7 different array platforms were utilised to develop the *OvMark* system. Table 1 contains a breakdown of the platforms used and the clinical information available with each dataset. Clinical data was manually checked to ensure that clinical factors are defined equally across studies. Clinical information where ambiguity occurred were excluded. In the case of the raw data for the Affymetrix datasets (.cel files), gene expression values were called using the GC robust multichip average method [14] and data were quantile normalised using the Bioconductor package, *affy* (www.bioconductor.org). For the dual-channel platforms, data were loess normalised [15] using the Bioconductor package *limma*. Where raw data was not available, the normalised data as published by the original authors was used. Hybridisation probes were mapped to Entrez gene IDs to gene centre the data to allow the comparison of the expression across disparate platforms [16]. The Entrez gene IDs corresponding to the array probes were obtained using Biomart [17] (www.biomart.org/) and the Bioconductor annotation libraries. Probes that mapped to multiple genes were filtered out. If there were multiple probes for the same gene, the probe values were averaged for that gene. This resulted in expression data for a total of 20,017 Entrez gene IDs across 2,129 samples. These 20,017 Entrez gene IDs corresponding to approximately 17,000 genes.

MiRNAs are frequently located within the introns of protein coding genes and in exons of non-coding transcripts. miRNA expression can be detected using conventional microarrays through host gene expression for intragenic miRNAs or by direct probe matching for intergenic miRNAs. 827 samples were processed on U133A Affymetrix arrays, while 531 were processed on U133plus2 Affymetrix arrays (1358 in total). U133A and U133plus2 microarrays have 22,277 probe sets in

common. Using this information, it is possible to infer the expression of 341 miRNAs across 1,358 samples based on a previously published mapping of Affymetrix probesets to miRBase [32]. As with the gene centred data, this information was also combined with the available clinical data for survival analysis. This approach does not measure the expression of the miRNA directly but rather uses its host gene as a surrogate. The cancer genome atlas (TCGA) [33] provides matched Affymetrix data and miRNA data. These datasets were overlapped using probe information from [32] and correlation between the two datasets was used to identify potential promising surrogates. This approach was obviously restricted to the miRNAs/genes available on the platforms used by TCGA and the limitations of these technologies. ~60% of the miRNAs significantly correlated with their host gene expression data. This information is available in Additional file 1. As this is not a definitive assessment of the correlation between miRNAs and their host genes the *OvMark* user is not restricted from searching all 341 miRNAs but is advised to use caution.

Survival analysis

The first stage in *OvMark* survival analysis dichotomises the expression of the gene of interest based on a median, high (within the 75% quartile) or low expression (within 25% quartile) cutoff. For example, if median expression is chosen, the expression of the gene of interest in a particular dataset is calculated. Those samples where the expression of the gene is greater than the median expression of that gene for that dataset are placed in the high expression group and those with less than median expression are placed in the low group. To account for study-to-study variation this phase is conducted separately for each of the 14 datasets. Once stratification is complete the individual datasets are combined and a global pooled survival analysis is performed to determine if the gene is associated with either OS and DFS. It is important to treat each dataset separately when determining if a sample belongs to the high or low expression groups, as the expression of the gene of interest will vary greatly across the different experiments/platforms. This approach is robust enough to detect the expression changes at low levels. A gene that goes from not detected to low expression, will “change quartiles”. In this presence or absence scenario, those samples where the gene is absent would be in the lower quartile, the “low expression cohort” and those samples where the gene is present would be in the higher quartile, the “high expression cohort” even though the absolute expression of the gene is very low. This is because though the interquartile range is low, it is still sufficient to distinguish between the two groups.

Survival curves are based on Kaplan-Meier estimates and the log-rank p-value is shown for difference in survival.

Cox regression analysis is used to calculate hazard ratios. The R package survival is used to calculate and plot the Kaplan-Meier survival curve. All calculations are carried out in the R statistical environment (<http://cran.r-project.org/>). For further details see Madden *et al.* [5].

Web server

The interface (that we have named *OvMark*) is available on a publically accessible a web server at <http://glados.ucd.ie/OvMark/index.html> and will be updated on a regular basis. The software uses Common Gateway interface (CGI) to link the web server with the R/PERL based algorithm. All calculations are carried out in real-time. All data/scripts are available upon request from the authors.

User input options

The software incorporates the clinical data made available by the original authors. This allows the gene expression data to be analysed based on one or more common clinical parameters including patient age, residual tumour, histological type, chemotherapy, neo-adjuvant chemotherapy, taxane treatment, platinum treatment, tumour grade, Federation of Gynaecologists and Obstetricians (FIGO) stage and histology subtype. The software also allows for the median expression and the upper or lower quartiles of the expression of the gene of interest to be used to determine high and low groups within each of the 14 individual datasets.

Testing *OvMark* using known markers of prognosis in Ovarian cancer

OvMark was run using 14 previously identified prognostic markers, LPR [33], PRL [34], SPP1 [34], IGF2 [35], MIF [36], CA125 [37], BRCA1 [3], BRCA2 [3], CDKN1B [38], MLH1 [39], ApoA1 [40], SNAI2 [41], CXCL12 [42] and IFNG [43]. Each of these 14 genes was queried in our novel *OvMark* database using the median expression option to dichotomise the data and DFS and OS as the survival endpoints.

Screen of all genes for their prognostic potential

All ~20,000 Entrez gene IDs corresponding to approximately 17,000 genes were queried in the *OvMark* database. The database was dichotomised using median gene expression and overall survival was chosen as the survival endpoint. No other software parameters were used. The resultant p-values are adjusted for multiple testing using the Benjamini-Hochberg method [44]. The significant results after adjustment for multiple testing are ranked by their hazard ratio.

To illustrate the utility of the *OvMark* system we identified the 10 genes most closely associated with

survival (i.e. highest and lowest hazard ratios from Cox regression). To further demonstrate their potential as biomarkers for ovarian cancer prognosis the concordance index (C-index) was calculated for each of these genes, using the bioconductor pack *survcomp* [45]. The C-index is a commonly used metric for the assessment of prognostic biomarker performance and has been utilised in two recent studies focussed on ovarian cancer [46,47]. Briefly, the C-Index, measures the ability of a particular gene's expression levels to classify patient survival times; a C-Index = 0.5 represents random classification while a C-Index = 1 represents perfect discrimination. We also calculated the probability of observing the C-Index value for each gene at random by shuffling the survival information 10,000 times and determining the C-Index for each iteration. We determined the number of times a random C-Index was greater than or equal to the "true" C-Index to generate an empirical p-value.

Results

In order to test the robustness of our gene-centred survival meta-analysis we used a panel of 14 known markers of prognosis in ovarian cancer. In addition we chose to screen all human genes for their prognostic potential in ovarian cancer. As there is currently no large-scale robust signature for miRNAs in ovarian cancer, we tested our approach on known individual miRNAs which have previously been shown to be prognostic markers, namely *let-7f-2* [12] and *mir-16-1* [13]. These miRNAs were chosen as proof of concept examples, to demonstrate the robustness of *OvMark*. Many other miRNAs are significantly associated with survival depending on the combination of clinical parameters chosen.

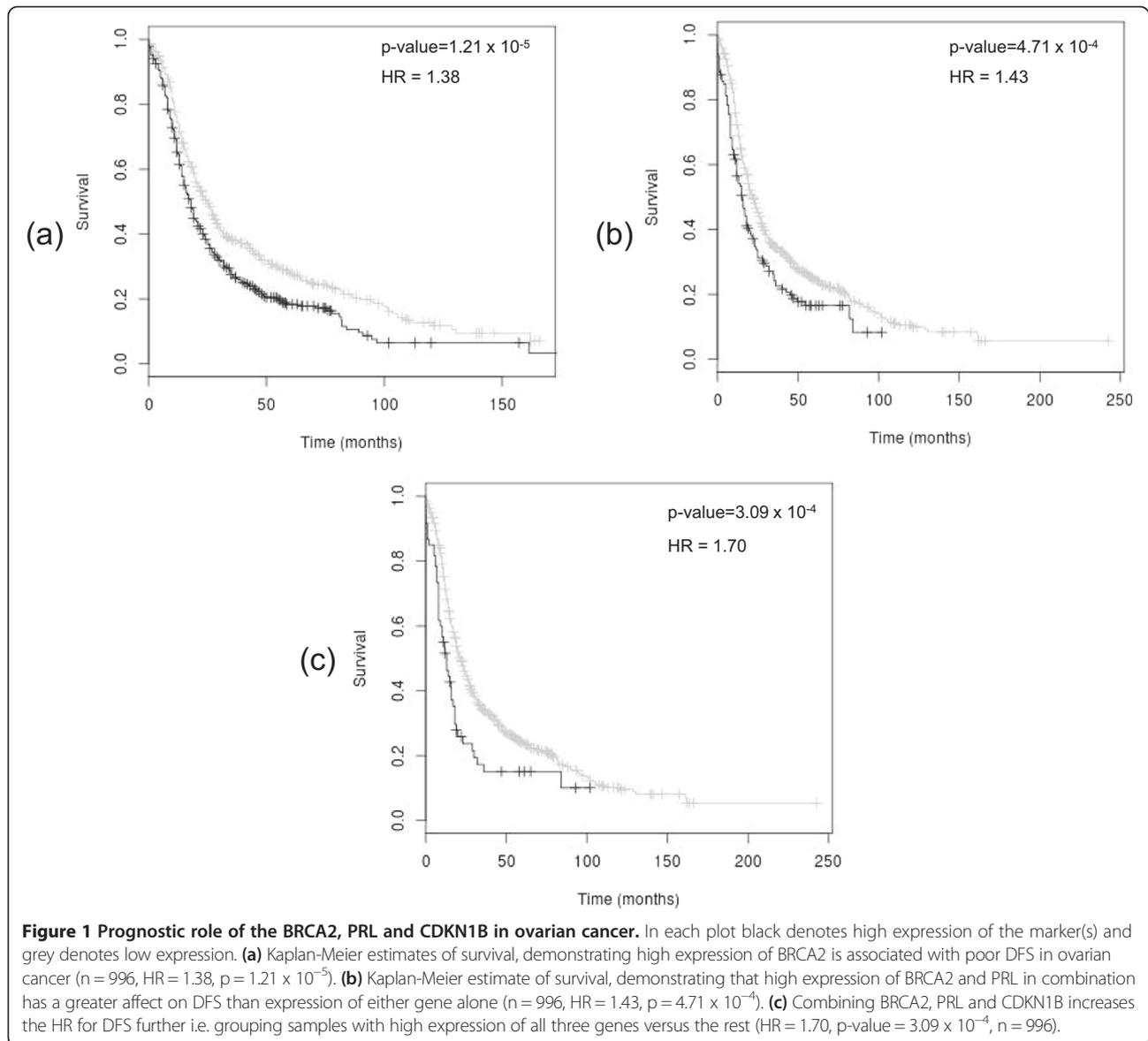
Ovmark results correlate with previously identified mRNA and miRNA based biomarkers for ovarian cancer

Each of the 14 ovarian cancer gene expression markers identified above were analysed separately within *OvMark* using median expression to dichotomise the data and DFS and OS as the survival end points. All patients in the *OvMark* database were chosen for this analysis without sub-selection based on any clinical parameters. This information is summarised in Table 2. A hazard ratio (HR) of greater than 1 indicates a negative effect on survival and a HR of less than one has a positive effect. For HRs greater than 1, the higher the HR the greater the effect the gene has on survival. For HRs less than 1, the lower the HR the greater the effect the gene has on survival. As can be seen from Table 2, several of the markers were significantly or borderline significantly associated with ovarian cancer patient outcomes, with the direction of those associations all consistent with what would be expected based on prior ovarian cancer studies. BRCA2 is one of the most significant individual marker of prognosis as can be seen in Figure 1(a) (HR for DFS = 1.38, $p = 1.21 \times 10^{-5}$, $n = 996$). Combining the markers can improve HRs in comparison with single markers alone. For example, the Kaplan-Meier DFS plot for BRCA2 and PRL in combination (i.e. comparing the OS of patients with greater than median expression of both BRCA2 and PRL, against the rest) is shown in Figure 1(b) (HR = 1.43, $p = 4.71 \times 10^{-4}$, $n = 996$). Patients with high-level expression of both BRCA2 and PRL in their ovarian cancers have a worse prognosis than those with high expression of BRCA2 or PRL alone in their ovarian cancers (with a HR of 1.43 versus a HR of 1.38 or 1.25 respectively). When greater than median expression of CDKN1B is combined

Table 2 *OvMark* results for known markers of ovarian cancer using the median cut-off

Gene symbol	Survival endpoint	Hazard ratio	P-value	Sample number
LEPR [33]	OS/DFS	1.14/1.08	0.04/0.28	1833/996
PRL [34]	OS/DFS	1.05/1.25	$0.49/6.78 \times 10^{-3}$	1990/996
SPP1 [34]	OS/DFS	1.13/1.17	0.05/0.03	1990/996
IGF2 [35]	OS/DFS	1.22/0.99	0.06/0.99	709/422
SNAI2 [41]	OS/DFS	1.28/1.30	$5.89 \times 10^{-5}/3.37 \times 10^{-4}$	1990/996
CXCL12 [42]	OS/DFS	1.27/1.42	$2.42 \times 10^{-4}/1.52 \times 10^{-6}$	1833/996
ApoA1 [40]	OS/DFS	1.02/1.17	0.77/0.03	1990/996
MIF [36]	OS/DFS	0.89/0.89	0.09/0.13	1833/996
CA125 [37]	OS/DFS	0.89/0.97	0.09/0.69	1651/924
BRCA1 [3]	OS/DFS	1.13/1.01	0.04/0.26	1990/996
BRCA2 [3]	OS/DFS	1.18/1.38	$0.01/1.21 \times 10^{-5}$	1833/996
CDKN1B [38]	OS/DFS	1.18/1.14	$9.13 \times 10^{-3}/0.07$	1990/996
MLH1 [39]	OS/DFS	0.98/0.81	$0.78/5.18 \times 10^{-3}$	1833/996
IFNG [43]	OS/DFS	0.80/1.06	$4.29 \times 10^{-3}/0.43$	1833/996

DFS = Disease Free survival, OS = Overall Survival.



with greater than median expression of BRCA2 and PRL (Figure 1(c)) the HR for OS is further increased ($HR = 1.70$, $p = 3.09 \times 10^{-4}$, $n = 996$).

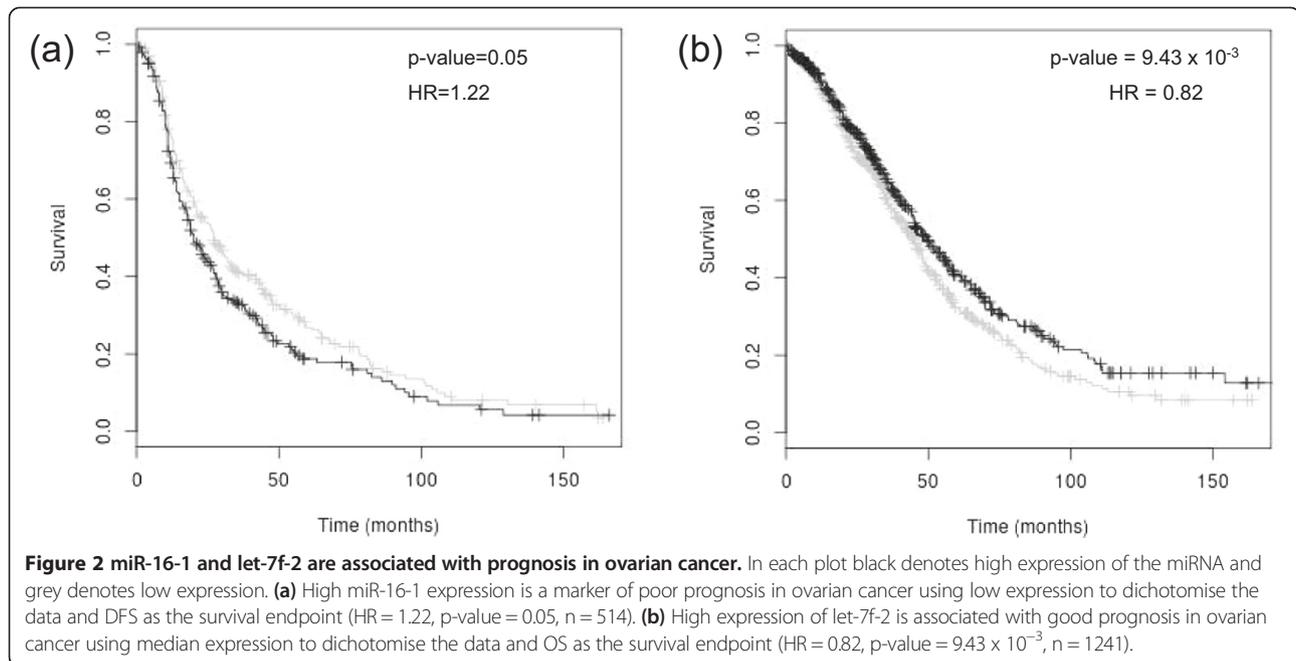
The miRNA MiR-16-1 has previously been associated with poor prognosis in ovarian cancer [13] and let-7f-2 is a member of the let-7 family of miRNAs which are frequently down regulated in cancer and are associated with a good prognosis in ovarian and other cancers [48]. The expression of both these markers is significantly correlated with their host gene expression in the TCGA data. To test our approach and to demonstrate the robustness of these markers in ovarian cancer, we examined the association of the host genes of these miRNAs with prognosis in ovarian cancer using our *OvMark* database. Our results for miR-16-1 and let-7f-2 are shown in Figure 2 (a) and (b), respectively. We confirmed high expression of

the host gene of miR-16-1 to be associated with a poor prognosis ($HR = 1.22$, $p\text{-value} = 0.05$, $n = 514$) and high expression of the host gene of let-7f-2 to be associated with a good prognosis ($HR = 0.82$, $p\text{-value} = 9.43 \times 10^{-3}$, $n = 1241$).

In summary, outcome predictions by the *OvMark* database for known ovarian cancer biomarkers are consistent with previously published data, thus validating the potential utility of this database for the study of the clinical and outcome implications of the expression of other genes and miRNAs in ovarian cancer.

A transcriptome-wide screen using OVMARK identifies of potential novel biomarkers

Although all of the prognostic markers above chosen for testing have been well studied in ovarian cancer



(in some cases with mixed results), only BRCA2, SNAI2 and CXCL12 were a particularly convincing marker of outcome in *OvMark*. We then queried all human genes in *OvMark* so as to identify which genes had the greatest association with OS of ovarian cancer patients. The results for the top ten genes are summarised in Table 3, along with their C-index. Additional file 2, shows a forrest plot for each of the 10 genes to illustrate the C-index and the upper and lower bounds. The sample numbers vary depending on the number of platforms with probes for the gene of interest. The genes are ranked in Table 3 based on the strength of the OS hazard ratio. Figure 3 shows the Kaplan-Meier plot for three of genes, snail homolog 3 (SNAI3), primary ciliary dyskinesia protein 1 (PCDP1) and

serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 2 (SERPINA2). Figure 3(a) shows the Kaplan-Meier plot for SNAI3 on its own (HR = 0.61, $p = 5.73 \times 10^{-5}$) and Figure 3(b) and (c) show the plots for SNAI3 in combination with PCDP1 and in combination with PCDP1 and SERPINA2, respectively. What is most striking about these results is how few of the genes have been previously linked to ovarian cancer, with only follicle stimulating hormone receptor (FSHR) having been previously well studied in ovarian cancer [49].

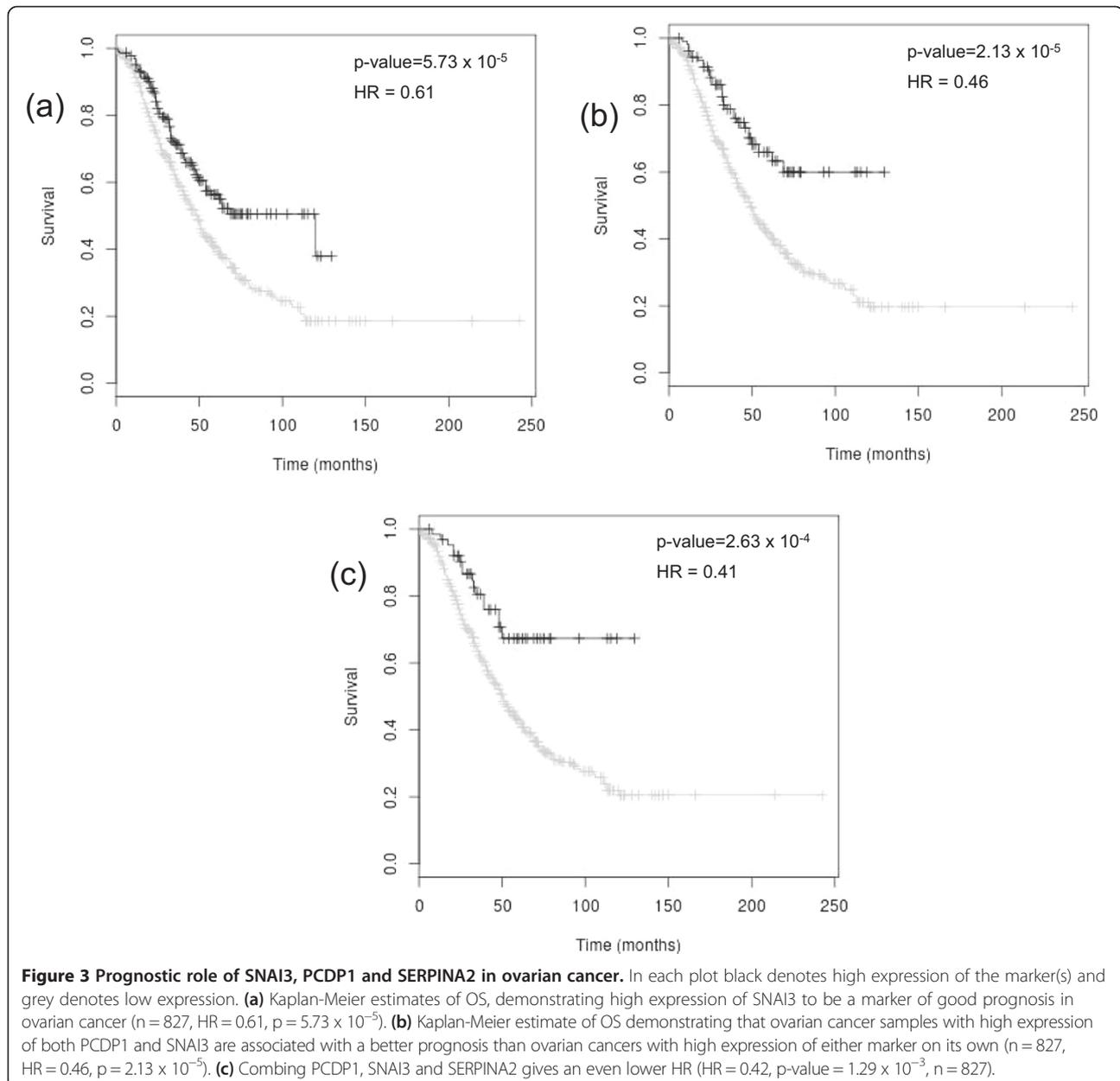
Discussion

OvMark provides a user-friendly tool for examining putative prognostic biomarkers in ovarian cancer. It builds on

Table 3 The top 10 highest/lowest hazard ratios from *OvMark* screen of all Entrez Gene IDs (n = 20,016) using OS as the survival end point and the median cut-off

Gene symbol	Entrez gene ID	Previous association with ovarian cancer	Previous association with cancer	Hazard ratio	P-value*	Sample number	C-Index	P-value
SNAI3	333929	Yes [50]	Yes [51]	0.61	0.03	827	0.62	2.0×10^{-4}
VWA3A	146177	No	No	0.63	0.02	610	0.59	2.1×10^{-3}
DNAH12	201625	No	No	0.64	0.01	610	0.61	2.0×10^{-4}
SERPINA2	390502	No	No	0.64	0.01	1293	0.64	< 0.0001
TMEM181	57583	No	No	1.56	0.03	527	0.60	8.0×10^{-4}
PCDP1	200373	No	No	0.65	8.96×10^{-3}	827	0.62	< 0.0001
ANKIB1	54467	No	No	1.53	0.04	527	0.60	3.2×10^{-3}
C11ORF88	399949	No	No	0.66	0.01	910	0.62	< 0.0001
FSHR	2492	Yes [52]	Yes [52]	0.68	9.14×10^{-3}	1833	0.60	1.0×10^{-3}
TBC1D23	55773	No	Yes [53]	1.46	0.05	610	0.60	1.2×10^{-3}

*P-value adjusted for multiple testing using the Benjamini-Hochberg method [44].



our previous work in breast cancer where we successfully combined multiple datasets to perform cross-dataset survival analysis [5,6]. The value of the approach used here is based on its simplicity of operation, and the statistical power gained through the combination of a large cohort of patients as compared to single microarray experiments. Unlike previous approaches [54], the *OvMark* system allows users to assess prognostic markers across multiple microarray platforms by utilising study by study dicotimisation to reduce batch effects. We are therefore not reliant on complex dataset transformations. Also, as the database is gene-centred, rather than probe-centred, we are not limited to the gene coverage of a particular platform. In summary, *OvMark* allows the analysis of ~20,000 unique

Entrez gene IDs in 2,129 ovarian cancer samples. It is our intention to increase this number as more ovarian cancer datasets become available.

Surprisingly, those genes found to have the greatest correlation with outcome across the dataset have not been studied in ovarian cancer, or in some cases in any cancer. For example, of the three genes with the greatest association with overall survival SNAI3, VWA3A and DNAH12, only SNAI3 has been previously associated with carcinogenesis where it is involved in epithelial to mesenchymal transition [55].

After confirming the robustness of our algorithm using genes already identified as biomarkers in ovarian cancer we proceeded to examine its potential for inferring the

prognostic ability of miRNAs from the gene expression data. The attraction of miRNA biology to cancer researchers arises from the potential of miRNAs to alter an entire pathway or indeed, pathways. miRNAs have been heavily studied in ovarian cancer; however, their role as prognostic markers are not well characterised. There are only a few large-scale studies which incorporate miRNA profiling integrated with detailed clinical data [56,57]. Despite the huge efforts required to compile these datasets, their sample numbers are only in the hundreds and therefore they have limited statistical power. However, as shown here in our study, there is a wealth of gene expression data available with detailed clinical information which can be exploited by inferring miRNA activity from host gene expression.

Again, our algorithm gene centres the data, and allows us to examine miRNAs as prognostic markers in ovarian cancer. We were able to confirm the results of other studies [12,13,48], which demonstrated that reduced expression of let-7f-2 and increased expression of miR-16-1 are associated with poor prognosis in ovarian cancer. It should be noted, however, that not all miRNAs and host genes are co-expressed [9] and care needs to be taken when interpreting miRNA results from *OvMark*. This issue cannot be fully resolved until such time as there is a better understanding of which miRNAs are co-expressed with their host genes (using the TCGA data we estimate this number to be ~60%) and, further, if those that are not significantly co-expressed do so in a disease/tissue-specific manner. It also needs to be determined whether the miRNAs themselves are subject to some level of post-transcriptional regulation.

Our new algorithm *OvMark* has some limitations. It can not overcome the inherent problems associated with transcriptomic analysis of ovarian cancer, in that often samples are taken at a late stage. The identification of biomarkers through the *OvMark* system should only be considered as part of the discovery phase. In order to confirm the utility of genes identified by *OvMark* further validation is required to assess biomarker candidates in an independent replication cohort. The *OvMark* system will allow researchers to easily assess the prognostic performance of their targets of interest within a large scale dataset and reduce false discovery rates when prioritising putative biomarkers for subsequent validation in their laboratories.

Conclusions

In this study, we have developed a simple user-friendly tool for examining putative gene/miRNA prognostic markers in ovarian cancer. The value of this tool is both in the simplicity of its design and the robustness of its approach. It is designed with non-bioinformatic research groups in mind and will be of great value in the preliminary assessment of putative biomarkers in ovarian cancer.

Additional files

Additional file 1: miRNAs significantly correlated with their host gene expression data.

Additional file 2: Forest plot illustrating the concordance index (CI) for each of the 10 genes most closely associated with patient outcome. The statistical significance of each CI was demonstrated by shuffling the survival information 10,000 times and calculating an empirical p-value.

Competing interests

The authors declare they have no competing interests.

Authors' contributions

Conception and design: SFM, CC, BTH. Development of methodology: SFM, CC. Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): SFM, MSC, RB, GBM. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): SFM, CC, BS, BTH. Writing, review, and/or revision of the manuscript: SFM, CC, BS, MSC, GBM, BTH. Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): CC. Study supervision: WMG, JC, BTH. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Martin Clynes and all of the other members of the NICB for their support during this project.

Grant support

Funding is acknowledged from the Science Foundation Ireland Strategic Research Cluster, Molecular Therapeutics for Cancer Ireland (www.mtci.ie) (award 08/SRC/B1410), Irish Cancer Society Postdoctoral Fellowship (to BS), Marie Curie Reintegration Grant from the European Union FP7 programme (to BS), a Science Foundation Ireland (SFI)/Health Research Board (HRB) Translational Research Award (TRA/2010/8) to BTH, a Career Development Award (CDA) from the Conquer Cancer Foundation of the American Society of Clinical Oncology (ASCO) to BTH and Science Foundation Ireland Grant Number (award 13/SIRG/2084) to CC.

Author details

¹Molecular Therapeutics for Cancer Ireland, National Institute for Cellular Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland. ²UCD School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Dublin 4, Ireland. ³Department of Histopathology, St James' Hospital and Trinity College Dublin, Dublin 8, Ireland. ⁴Department of Medical Oncology, Beaumont Hospital, P.O. Box 1297, Beaumont Road, Dublin 9, Ireland. ⁵Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Tx, USA. ⁶Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Tx, USA. ⁷Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada.

Received: 11 April 2014 Accepted: 26 September 2014

Published: 24 October 2014

References

1. Hennessy BT, Coleman RL, Markman M: **Ovarian cancer.** *Lancet* 2009, **374**(9698):1371–1382.
2. Al Rawahi T, Lopes AD, Bristow RE, Bryant A, Elattar A, Galaal K: **Surgical cytoreduction for recurrent epithelial ovarian cancer.** *Cochrane Database Syst Rev* 2013, **2**:CD008765.
3. Le Page C, Huntsman DG, Provencher DM, Mes-Masson A-M: **Predictive and prognostic protein biomarkers in epithelial ovarian cancer: recommendation for future studies.** *Cancers* 2010, **2**(2):913–954.
4. Hennessy BT, Timms KM, Carey MS, Gutin A, Meyer LA, Flake DD, Abkevich V, Potter J, Pruss D, Glenn P, Li Y, Li J, Gonzalez-Angulo AM, McCune KS, Markman M, Broaddus RR, Lanchbury JS, Lu KH, Mills GB: **Somatic mutations in BRCA1 and BRCA2 could expand the number of patients that benefit from poly (ADP ribose) polymerase inhibitors in ovarian cancer.** *J Clin Oncol* 2010, **28**(22):3570–3576.

5. Madden SF, Clarke C, Gaule P, Aherne ST, O'Donovan N, Clynes M, Crown J, Gallagher WM: **BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome.** *Breast Cancer Res* 2013, **15**(4):R52.
6. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S, Clynes M: **Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis.** *Carcinogenesis* 2013, **34**(10):2300–2308.
7. Iorio MV, Visone R, Leva GD, Donati V, Petrocca F, Casalini P, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S, Clynes M: **MicroRNA signatures in human ovarian cancer.** *Cancer Res* 2007, **67**(18):8699–8707.
8. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**(2):215–233.
9. Baskerville S, Bartel DP: **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes.** *RNA* 2005, **11**(3):241–247.
10. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A: **Identification of mammalian microRNA host genes and transcription units.** *Genome Res* 2004, **14**(10A):1902–1910.
11. Liang Y, Ridzon D, Wong L, Chen C: **Characterization of microRNA expression profiles in normal human tissues.** *BMC Genomics* 2007, **8**(1):166.
12. Helland Å, Anglesio MS, George J, Cowin PA, Johnstone CN, House CM, Sheppard KE, Etemadmoghadam D, Melnyk N, Rustgi AK, Phillips WA, Johnsen H, Holm R, Kristensen GB, Birrer MJ, Australian Ovarian Cancer Study Group, Pearson RB, Børresen-Dale AL, Huntsman DG, deFazio A, Creighton CJ, Smyth GK, Bowtell DD: **Deregulation of MYCN, LIN28B and LET7 in a molecular subtype of aggressive high-grade serous ovarian cancers.** *PLoS One* 2011, **6**(4):e18064.
13. Delfino KR, Rodriguez-Zas SL: **Transcription factor-MicroRNA-target gene networks associated with ovarian cancer survival and recurrence.** *PLoS One* 2013, **8**(3):e58608.
14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**(2):249–264.
15. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucl Acids Res* 2002, **30**(4):e15–e15.
16. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005, **33**(Database issue):D54–D58.
17. Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, Baldock R, Chelala C, Croft D, Cros A, Cutts RJ, Di Génova A, Forbes S, Fujisawa T, Gadaleta E, Goodstein DM, Gundem G, Haggarty B, Haider S, Hall M, Harris T, Haw R, Hu S, Hubbard S, Hsu J, Iyer V, Jones P, Katayama T, Kinsella R, Kong L, Lawson D, et al: **BioMart central portal: an open database network for the biological community.** *Database (Oxford)* 2011, **2011**:bar041.
18. Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, Bogomolny F, Ozbun L, Brady J, Barrett JC, Boyd J, Birrer MJ: **A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer.** *Cancer Res* 2008, **68**(13):5478–5486.
19. Crijns APG, Fehrmann RSN, de Jong S, Gerbens F, Meersma GJ, Klip HG, Hollema H, Hofstra RM, te Meerman GJ, de Vries EG, van der Zee AG: **Survival-related profile, pathways, and transcription factors in ovarian cancer.** *PLoS Med* 2009, **6**(2):e1000024.
20. Denkert C, Budczies J, Darb-Esfahani S, Györfy B, Sehoul J, Könsgen D, Zeillinger R, Weichert W, Noske A, Buckendahl AC, Müller BM, Dietel M, Lage H: **A prognostic gene expression index in ovarian cancer - validation across different independent data sets.** *J Pathol* 2009, **218**(2):273–280.
21. Ferriss JS, Kim Y, Duska L, Birrer M, Levine DA, Moskaluk C, Theodorescu D, Lee JK: **Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance.** *PLoS One* 2012, **7**(2):e30550.
22. Konstantinopoulos PA, Cannistra SA, Fountzilias H, Culhane A, Pillay K, Rueda B, Cramer D, Seiden M, Birrer M, Coukos G, Zhang L, Quackenbush J, Spentzos D: **Integrated analysis of multiple microarray datasets identifies a reproducible survival predictor in ovarian cancer.** *PLoS One* 2011, **6**(3):e18202.
23. Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilias E, Francoeur N, Levine DA: **Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer.** *J Clin Oncol* 2010, **28**(22):3555–3561.
24. Mateescu B, Batista L, Cardon M, Gruosso T, de Feraudy Y, Mariani O, Nicolas A, Meyniel JP, Cottu P, Sastre-Garau X, Mechta-Grigoriou F: **miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response.** *Nat Med* 2011, **17**(12):1627–1635.
25. Home - The Cancer Genome Atlas - Cancer Genome - TCGA [Internet]; 2014. Available from: <http://cancergenome.nih.gov/>.
26. Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**(7216):1061–1068.
27. Mok SC, Bonome T, Vathipadiekal V, Bell A, Johnson ME, Wong K, Park DC, Hao K, Yip DK, Donninger H, Ozbun L, Samimi G, Brady J, Randonovich M, Pise-Masison CA, Barrett JC, Wong WH, Welch WR, Berkowitz RS, Birrer MJ: **A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2.** *Cancer Cell* 2009, **16**(6):521–532.
28. Spentzos D, Levine DA, Kolia S, Otu H, Boyd J, Libermann TA, Cannistra SA: **Unique gene expression profile based on pathologic response in epithelial ovarian cancer.** *J Clin Oncol* 2005, **23**(31):7911–7918.
29. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B, Traficante N, Fereday S, Hung JA, Chiew YE, Haviv I, Australian Ovarian Cancer Study Group, Gertig D, DeFazio A, Bowtell DD: **Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome.** *Clin Cancer Res* 2008, **14**(16):5198–5208.
30. Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H, Suzuki M, Onishi Y, Hatae M, Sueyoshi K, Fujiwara H, Kudo Y, Kotera K, Masuzaki H, Tashiro H, Katabuchi H, Inoue I, Tanaka K: **Gene Expression Profile for Predicting Survival in Advanced-Stage Serous Ovarian Cancer Across Two Independent Datasets.** *PLoS One* 2010, **5**(3):e9615.
31. Yoshihara K, Tsunoda T, Shigemizu D, Fujiwara H, Hatae M, Fujiwara H, Masuzaki H, Katabuchi H, Kawakami Y, Okamoto A, Nogawa T, Matsumura N, Udagawa Y, Saito T, Itamochi H, Takano M, Miyagi E, Sudo T, Ushijima K, Iwase H, Seki H, Terao Y, Enomoto T, Mikami M, Akazawa K, Tsuda H, Moriya T, Tajima A, Inoue I, Tanaka K, et al: **High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway.** *Clin Cancer Res* 2012, **18**(5):1374–1385.
32. Rainer J, Ploner C, Jesacher S, Ploner A, Eduardoff M, Mansha M, Wasim M, Panzer-Grümayer R, Trajanoski Z, Niederegger H, Kofler R: **Glucocorticoid-regulated microRNAs and mirtrons in acute lymphoblastic leukemia.** *Leukemia* 2009, **23**(4):746–752.
33. Uddin S, Bu R, Ahmed M, Abubaker J, Al-Dayel F, Bavi P, Al-Kuraya KS: **Overexpression of leptin receptor predicts an unfavorable outcome in Middle Eastern ovarian cancer.** *Mol Cancer* 2009, **8**(1):74.
34. Palmer C, Duan X, Hawley S, Scholler N, Thorpe JD, Sahota RA, Wong MQ, Wray A, Bergan LA, Drescher CW, McIntosh MW, Brown PO, Nelson BH, Urban N: **Systematic evaluation of candidate blood markers for detecting ovarian cancer.** *PLoS One* 2008, **3**(7):e2633.
35. Huang GS, Brouwer-Visser J, Ramirez MJ, Kim CH, Hebert TM, Lin J, Arias-Pulido H, Qualls CR, Prossnitz ER, Goldberg GL, Smith HO, Horwitz SB: **Insulin-like growth factor 2 expression modulates Taxol resistance and is a candidate biomarker for reduced disease-free survival in ovarian cancer.** *Clin Cancer Res* 2010, **16**(11):2999–3010.
36. Krockenberger M, Kranke P, Häusler S, Engel JB, Horn E, Nürnberg K, Wischhusen J, Dietl J, Höning A: **Macrophage migration-inhibitory factor levels in serum of patients with ovarian cancer correlates with poor prognosis.** *Anticancer Res* 2012, **32**(12):5233–5238.
37. Skates SJ, Mai P, Horick NK, Piedmonte M, Drescher CW, Isaacs C, Armstrong DK, Buys SS, Rodriguez GC, Horowitz IR, Berchuck A, Daly MB, Domchek S, Cohn DE, Van Le L, Schorge JO, Newland W, Davidson SA, Barnes M, Brewster W, Azodi M, Nerenstone S, Kauff ND, Fabian CJ, Sluss PM, Nayfield SG, Kasten CH, Finkelstein DM, Greene MH, Lu K: **Large prospective study of ovarian cancer screening in high-risk women: CA125 cut-point defined by menopausal status.** *Cancer Prev Res (Phila)* 2011, **4**(9):1401–1408.
38. Newcomb EW, Sosnow M, Demopoulos RI, Zeleniuch-Jacquotte A, Sorich J, Speyer JL: **Expression of the Cell Cycle Inhibitor p27KIP1 Is a New Prognostic Marker Associated with Survival in Epithelial Ovarian Tumors.** *Am J Pathol* 1999, **154**(1):119–125.
39. Samimi G, Fink D, Varki NM, Husain A, Hoskins WJ, Alberts DS, Howell SB: **Analysis of MLH1 and MSH2 expression in ovarian cancer before and after platinum drug-based chemotherapy.** *Clin Cancer Res* 2000, **6**(4):1415–1421.
40. Su F, Kozak KR, Imaizumi S, Gao F, Amneus MW, Grijalva V, Ng C, Wagner A, Hough G, Farias-Eisner G, Anantharamaiah GM, Van Lenten BJ, Navab M,

- Fogelman AM, Reddy ST, Farias-Eisner R: **Apolipoprotein A-I (apoA-I) and apoA-I mimetic peptides inhibit tumor development in a mouse model of ovarian cancer.** *Proc Natl Acad Sci U S A* 2010, **107**(46):19997–20002.
41. Haslehurst AM, Koti M, Dharsee M, Nuin P, Evans K, Geraci J, Childs T, Chen J, Li J, Weberpals J, Davey S, Squire J, Park PC, Feilolter H: **EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer.** *BMC Cancer* 2012, **12**(1):91.
42. Poppo A, Durrant LG, Spendlove I, Rolland P, Scott IV, Deen S, Ramage JM: **The chemokine, CXCL12, is an independent predictor of poor survival in ovarian cancer.** *Br J Cancer* 2012, **106**(7):1306–1313.
43. Chen YL, Cheng WF, Chang MC, Lin HW, Huang CT, Chien CL, Chen CA: **Interferon-gamma in ascites could be a predictive biomarker of outcome in ovarian carcinoma.** *Gynecol Oncol* 2013, **131**(1):63–68.
44. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, **57**(1):289–300.
45. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B: **survcomp: an R/Bioconductor package for performance assessment and comparison of survival models.** *Bioinformatics* 2011, **27**(22):3206–3208.
46. Waldron L, Haibe-Kains B, Culhane AC, Riestler M, Ding J, Wang XV, Ahmadifar M, Tyekucheva S, Bernau C, Risch T, Ganzfried BF, Huttenhower C, Birrer M, Parmigiani G: **Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer.** *J Natl Cancer Inst* 2014, **106**(5):dju049.
47. Riestler M, Wei W, Waldron L, Culhane AC, Trippa L, Oliva E, Kim SH, Michor F, Huttenhower C, Parmigiani G, Birrer MJ: **Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples.** *JNCI J Natl Cancer Inst* 2014, **106**(5):dju048.
48. Boyerinas B, Park S-M, Hau A, Murmann AE, Peter ME: **The role of let-7 in cell differentiation and cancer.** *Endocr Relat Cancer* 2010, **17**(1):F19–F36.
49. Bose CK: **Follicle stimulating hormone receptor in ovarian surface epithelium and epithelial ovarian cancer.** *Oncol Res* 2008, **17**(5):231–238.
50. Chen X, Jiang W, Wang Q, Huang T, Wang P, Li Y: **Systematically characterizing and prioritizing chemosensitivity related gene based on Gene Ontology and protein interaction network.** *BMC Med Genomics* 2012, **5**(1):43.
51. Katoh M, Katoh M: **Identification and characterization of human SNAIL3 (SNAI3) gene in silico.** *Int J Mol Med* 2003, **11**(3):383–388.
52. Mertens-Walker I, Baxter RC, Marsh DJ: **Gonadotropin signalling in epithelial ovarian cancer.** *Cancer Lett* 2012, **324**(2):152–159.
53. Shin N, You KT, Lee H, Kim WK, Song M, Choi HJ, Rhee H, Nam SW, Kim H: **Identification of frequently mutated genes with relevance to nonsense mediated mRNA decay in the high microsatellite instability cancers.** *Int J Cancer* 2011, **128**(12):2872–2880.
54. Györfy B, Lánckzy A, Szállási Z: **Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients.** *Endocr Relat Cancer* 2012, **19**(2):197–208.
55. Katoh Y, Katoh M: **Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review).** *Int J Mol Med* 2008, **22**(3):271–275.
56. Creighton CJ, Hernandez-Herrera A, Jacobsen A, Levine DA, Mankoo P, Schultz N, Du Y, Zhang Y, Larsson E, Sheridan R, Xiao W, Spellman PT, Getz G, Wheeler DA, Perou CM, Gibbs RA, Sander C, Hayes DN, Gunaratne PH, Network Cancer Genome Atlas Research: **Integrated analyses of microRNAs demonstrate their widespread influence on gene expression in high-grade serous ovarian carcinoma.** *PLoS One* 2012, **7**(3):e34546.
57. Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, Wendschlag A, Giese N, Tjaden C, Ott K, Werner J, Hackert T, Ruprecht K, Huwer H, Huebers J, Jacobs G, Rosenstiel P, Dommisch H, Schaefer A, Müller-Quernheim J, Wullich B, Keck B, Graf N, Reichrath J, Vogel B, Nebel A, Jäger SU, Staehler P, Amarantos I, Boisguerin V, *et al.*: **Toward the blood-borne miRNome of human diseases.** *Nat Methods* 2011, **8**(10):841–843.

doi:10.1186/1476-4598-13-241

Cite this article as: Madden *et al.*: *OvMark*: a user-friendly system for the identification of prognostic biomarkers in publically available ovarian cancer gene expression datasets. *Molecular Cancer* 2014 **13**:241.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

