Prognostic gene expression signature for high-grade serous ovarian cancer


PII: S0923-7534(20)39841-0
DOI: https://doi.org/10.1016/j.annonc.2020.05.019
Reference: ANNONC 201

To appear in: Annals of Oncology

Received Date: 4 January 2020
Revised Date: 6 May 2020
Accepted Date: 6 May 2020

Please cite this article as: Millstein J, Budden T, Goode EL, Anglesio MS, Talhouk A, Intermaggio MP, Leong HS, Chen S, Elatre W, Gilks B, Nazeran T, Volchek M, Bentley RC, Wang C, Chiu DS,

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd on behalf of European Society for Medical Oncology.
Original article

Prognostic gene expression signature for high-grade serous ovarian cancer


1. Division of Biostatistics, Department of Preventive Medicine, Keck School of Medicine. University of Southern California. Los Angeles: USA.

2. School of Women’s and Children’s Health, Faculty of Medicine. University of NSW Sydney. Sydney, New South Wales: Australia; 2052.

3. CRUK Manchester Institute. The University of Manchester. Manchester: UK.


10. Department of Pathology, Norris Comprehensive Cancer Center; Keck School of Medicine. University of Southern California. Los Angeles: USA.


14. Department of Women's Health. Tuebingen University Hospital. Tuebingen: Germany; 72076.

15. The Jackson Laboratory for Genomic Medicine. Farmington, CT: USA; CT-06032.


22. Division of Cancer and Ovarian Cancer Action Research Centre, Department Surgery & Cancer. Imperial College London. London: UK; W12 0NN.

23. Institute of Cancer Sciences. University of Glasgow. Glasgow: UK; G61 1QH.


29. Program in Epidemiology, Division of Public Health Sciences. Fred Hutchinson Cancer Research Center. Seattle, WA: USA; 98109.


32. Department of Genetics. Geisel School of Medicine at Dartmouth. Hanover, NH: USA.

33. Division of Cancer Epidemiology. German Cancer Research Center (DKFZ). Heidelberg: Germany; 69120.

34. Department of Pathology, Institute of Pathology. University Hospital Heidelberg. Heidelberg: Germany; 69120.

35. National Center for Tumor Diseases. University Hospital and German Cancer Research Center. Heidelberg: Germany; 69120.

36. David Geffen School of Medicine, Department of Obstetrics and Gynecology. University of California at Los Angeles. Los Angeles, CA: USA; 90095.


38. Department of Genetics and Pathology. Pomeranian Medical University. Szczecin: Poland; 71-252.

39. Department of Gynecological Surgery and Gynecological Oncology of Adults and Adolescents. Pomeranian Medical University. Szczecin: Poland; 70-111.


41. Department of Preventive Medicine, Keck School of Medicine. University of Southern California Norris Comprehensive Cancer Center. Los Angeles, CA: USA; 90033.
42. Department of Epidemiology and Biostatistics. Memorial Sloan-Kettering Cancer Center. New York, NY: USA; 10065.

43. Department of Preventive Medicine, Keck School of Medicine. University of Southern California. Los Angeles, CA: USA; 90033.

44. Centre for Cancer Genetic Epidemiology, Department of Oncology. University of Cambridge. Cambridge: UK; CB1 8RN.

45. Department of Pathology. Addenbrooke's Hospital NHS Trust. Cambridge: UK.

46. MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology. University College London. London: UK; WC1V 6LJ.

47. Department of Women’s Cancer, Institute for Women’s Health. University College London. London: UK; W1T 7DN.

48. Department of Pathology. Barts Health National Health Service Trust. London: UK; E1 8PR.

49. Department of Gynaecological Oncology. Westmead Hospital. Sydney, New South Wales: Australia; 2145.

50. Pathology West ICPMR Westmead, Westmead Hospital. The University of Sydney. Sydney, New South Wales: Australia; 2145.

51. University of Western Sydney at Westmead Hospital. Sydney, New South Wales: Australia; 2145.


53. Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology. Mayo Clinic. Rochester, MN: USA; 55905.


55. Department of Molecular Pharmacology and Experimental Therapeutics. Mayo Clinic. Rochester, MN: USA; 55905.

56. Cancer Epidemiology Program. University of Hawaii Cancer Center. Honolulu, HI: USA; 96813.


58. John A. Burns School of Medicine, Department of Obstetrics and Gynecology. University of Hawaii. Honolulu, HI: USA.

60. Department of Cancer Epidemiology and Prevention. M Sklodowska-Curie Cancer Center, Oncology Institute. Warsaw: Poland; 02-034.


62. Cancer Research UK Clinical Trials Unit, Institute of Cancer Sciences. University of Glasgow. Glasgow: UK; G12 0YN.

63. Department of Medical Oncology. Beatson West of Scotland Cancer Centre and University of Glasgow. Glasgow: UK; G12 0YN.

64. Gynaecology Unit. Royal Marsden Hospital. London: UK; SW3 6JJ.


68. David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology. University of California at Los Angeles. Los Angeles, CA: USA; 90095.


70. Gynecologic Oncology, Laura and Isaac Pearlmutter Cancer Center. NYU Langone Medical Center. New York, NY: USA; 10016.

71. Hollings Cancer Center and Department of Public Health Sciences. Medical University of South Carolina. Charleston, SC: USA; 29425.


74. Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH). University Medical Center Hamburg-Eppendorf. Hamburg: Germany; 20246.

75. Department of Public Health Sciences. University of Virginia. Charlottesville, VA: USA; 22908.
76. Division of Cancer and Ovarian Cancer Action Research Centre, Department of Surgery and Cancer. Imperial College London. London: UK; W12 0NN.


78. Center for Bioinformatics and Functional Genomics and the Cedars Sinai Genomics Core. Cedars-Sinai Medical Center. Los Angeles, CA: USA; 90048.

79. Biomedical Network on Rare Diseases (CIBERER). Madrid: Spain; 28029.


81. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center. Vanderbilt University School of Medicine. Nashville, TN: USA; 37232.

82. Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN. University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg. Erlangen: Germany; 91054.


84. Huntsman Cancer Institute, Department of Population Health Sciences. University of Utah. Salt Lake City, UT: USA; 84112.

85. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care. University of Cambridge. Cambridge: UK; CB1 8RN.


* Corresponding author:

Prof Susan Ramus

School of Women's and Children's Health, Faculty of Medicine

University of NSW Sydney, Sydney, New South Wales, Australia

(02) 9385 1720, s.ramus@unsw.edu.au
Running head: prognostic signature for high grade serous ovarian cancer
Abstract

Background

Median overall survival (OS) for women with high-grade serous ovarian cancer (HGSOC) is approximately four years, yet survival varies widely between patients. There are no well-established, gene expression signatures associated with prognosis. The aim of this study was to develop a robust prognostic signature for overall survival in HGSOC patients.

Patients and methods

Expression of 513 genes, selected from a meta-analysis of 1455 tumours and other candidates, were measured using NanoString technology from formalin-fixed, paraffin-embedded (FFPE) tumour tissue from 3,769 women with HGSOC from multiple studies. Elastic net regularization for survival analysis was applied to develop a prognostic model for 5-year OS, trained on 2702 tumours from fifteen studies and evaluated on an independent set of 1067 tumours from six studies.

Results

Expression levels of 276 genes were associated with OS [false discovery rate (FDR) < 0.05] in covariate-adjusted single gene analyses. The top five genes were TAP1, ZFHX4, CXCL9, FBN1, and PTGER3 ($P < 0.001$). The best performing prognostic signature included 101 genes enriched in pathways with treatment implications. Each gain of one standard deviation in the gene expression score (GES) conferred a greater than two-fold increase in risk of death [$HR = 2.35$ (2.02, 2.71); $P < 0.001$]. Median survival by GES quintile was 9.5 (8.3, --), 5.4 (4.6, 7.0), 3.8 (3.3, 4.6), 3.2 (2.9, 3.7) and 2.3 (2.1, 2.6) years.
Conclusion

The OTTA-SPOT (Ovarian Tumor Tissue Analysis consortium - Stratified Prognosis of Ovarian Tumours) gene expression signature may improve risk stratification in clinical trials by identifying patients who are least likely to achieve 5-year survival. The identified novel genes associated with the outcome may also yield opportunities for the development of targeted therapeutic approaches.

Key words: high grade serous ovarian cancer, gene expression, prognosis, overall survival, formalin fixed paraffin embedded

Highlights

- A gene expression signature for high-grade serous ovarian cancer prognostic for two- and five-year overall survival (OS).
- The 101 gene expression signature performs substantially better than age and stage alone.
- Median survival by quintile was 9.5, 5.4, 3.8, 3.2 and 2.3 years.
- The top five genes associated with OS were TAP1, ZFHX4, CXCL9, FBN1, and PTGER3 ($P \ll 0.001$).
Introduction

Epithelial ovarian cancer (EOC) causes approximately 125,000 deaths globally every year, and long-term survival rates have changed little in the past three decades[1]. Approximately 70% of women with EOC are diagnosed with advanced stage disease (stages III/IV), and fewer than 50% will survive more than 5 years[2]. There are five major EOC histotypes: high-grade serous, low-grade serous, endometrioid, clear cell and mucinous[3]. High-grade serous ovarian cancer (HGSOC) comprises about two-thirds of cases, is responsible for most deaths and is characterized by profound genomic and clinical heterogeneity.

The most informative prognostic factors for HGSOC are International Federation of Gynecology and Obstetrics (FIGO) stage, residual disease following debulking surgery[4], BRCA1 or BRCA2 germline mutation[5, 6] and tumour-infiltrating lymphocyte scores[7, 8]. Patients with HGSOC who carry a loss-of-function germline mutation in BRCA1 or BRCA2 have an increased sensitivity to platinum-based chemotherapy and PARP inhibitor treatment[9, 10] and a medium-term survival advantage[5]. However, the frequent development of drug-resistant disease[6] limits the effectiveness of current therapies.

Gene-expression data have been used to define four tumour molecular subtypes of HGSOC (C1/mesenchymal, C2/immune, C4/differentiated and C5/proliferative)[11, 12]. Using transcriptome-wide data from fresh frozen tissues, The Cancer Genome Atlas (TCGA) project used 215 tumours to identify an overall survival (OS) expression signature of 193 genes that has been validated on three other HGSOC gene expression data sets[12].

Despite these findings, gene expression biomarkers have not been implemented clinically owing to several important shortcomings. The majority of the individual markers comprising the 193 gene signature were not statistically significant across all studies, suggesting that the signature may not be robust. The sample sizes in other discovery efforts have been too small for robust statistical inference[12]. Also, previous studies used
fresh frozen samples, resulting in logistic and cost barriers to examining large clinically relevant data sets, and translation to the clinical setting.

The aim of this study was to identify a robust and clinic-ready prognostic HGSOC profile that can be applied to formalin fixed paraffin embedded (FFPE) tumour tissue.

Patients and methods

Twenty studies provided pre-treatment, FFPE tumour samples from 4,071 women diagnosed with HGSOC (Supplemental Table S1). All HGSOC cases with available tissue were included. During this time period, HGSOC patients were treated with chemotherapy (carboplatin and paclitaxel) after primary debulking surgery. Study protocols were approved by the respective Institutional Review Board / ethics approval committee for each site (Supplemental Table S1).

A schematic of the overall study design is shown in Figure 1. There were four main components: gene selection, gene-expression assay, development of prognostic gene signature in a training set and validation of prognostic signature in an independent validation set.

Gene selection

Candidate prognostic genes were identified by carrying out an individual participant meta-analysis of six transcriptome-wide microarray studies[11-16], which included tumour samples from 1,455 participants. Gene expression association with overall survival was evaluated by Cox proportional hazards regression adjusted for molecular subtype (Supplemental Table S2). In total, 200 genes from the meta-analysis, most achieving a permutation-based FDR[17] of less than 0.05, and an additional 313 candidate genes based on the literature and unpublished data were selected (Supplemental Tables S3 and S4, Figure S1; for more details see Supplemental Material). Five genes, RPL19, ACTB, PGK1, SDHA, and POLR1B, were included as house-keeping genes for normalization.
Gene expression assay in study participants samples

FFPE tumour samples were processed with the NanoString nCounter technology at 3 different locations, Vancouver, Los Angeles and Melbourne. A control set of 48 FFPE tumour samples were run at each location and the average intraclass correlation coefficient (ICC) was 0.987. Approximately 2 percent of the samples were run in duplicate and the average Spearman correlation $r^2$ was 0.995. Single-patient classification methods were used with reference samples to control for batch effects[18]. The data in this publication have been deposited in NCBI's Gene Expression Omnibus[19]; GEO Series accession number GSE132342 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132342). 3,329 samples passed quality control of which 3,769 had survival data and assessable gene expression for 513 genes. Data can be found in NCBI GEO: Accession numbers GSE132342 and GPL26748.

Overall survival analysis of individual genes

Samples that contributed to the meta-analysis data set (n=211) were removed from subsequent selected analyses to enforce independence of study samples between the gene selection and final survival analysis. Time-to-event analyses were carried out for OS with right-censoring at 10 years and left-truncation of prevalent cases. Associations between log-transformed normalized gene expression and survival time were tested using likelihood ratio tests with Cox proportional hazards models adjusted for age, race, and stage, and stratified by study. Patients with missing race or stage information were assigned to ‘unknown’ categories. Age was modelled using a B-spline with a knot at the median age, which yielded a better fit than using knots at quartiles or categorical variables. Stage was dichotomized into early (International Federation of Gynecology and Obstetrics [FIGO] stage I/II) and advanced (FIGO stage III/IV). Genes were scaled to have a standard deviation of one, so hazard ratios correspond to a change of one standard deviation. A Benjamini-Hochberg (BH) false discovery rate (FDR) of less than 0.05 was used to identify notable associations. Since the expression of genes can be correlated, an analysis of correlated genes was performed using data from TCGA. Advanced
stage ovarian cancer usually has disease spread throughout the abdomen, therefore sensitivity analyses were performed to assess effects of the anatomical location of tumour samples included in the study by removing observations corresponding to samples known to be extraovarian (n = 437).

Prognostic signature development and validation

Studies were initially randomized to training set (N = 14) and validation set (N = 6). The TRI study was randomized to the validation set, but, because 107 of the samples were part of the meta-analysis data used for gene selection, the study was split, so those 107 samples were included in the model training data set. Thus 2,702 samples from 15 studies were used for model training and 1,067 samples from 6 studies were used for validation (Supplemental Table S1). In the training set, four modelling approaches (stepwise regression, elastic net regularized regression, boosting and random survival forests) were applied to construct competing gene expression-based biomarkers. Each was evaluated in the training data using 10-fold cross-validation for its prognostic value for OS at two and five years of follow-up using an area under the curve (AUC) measure derived from receiver operator characteristic (ROC) analysis (see Supplemental Material for additional details). The best performing method, elastic net regularized regression, was applied to the full training set to determine the final gene signature and scoring method, which was then evaluated using the independent testing set. All models were constrained to include age and stage, where age was modelled as categorical based on quartiles of the training dataset with groups: less than 53 years old, 53 to 59, 60 to 66, and 67 or greater. Stage was modelled as described above for the OS individual gene analysis.

Results

Association of expression of individual genes with OS in HGSOC.

In a gene-by-gene analysis of the full data set adjusted for age, race, and stage, and stratified by study, 276 of the 513 selected genes were associated with OS (FDR < 0.05). Of these, 138 were selected from the meta-
analysis of six published microarray studies (Supplemental Table S2)[11-16] and 144 from candidate gene approaches (Supplemental Tables S5 and S6). Hazard ratios (HR) for one standard deviation change in gene expression ranged from 0.84–1.19, with multiple genes exhibiting associations at very stringent significance levels (e.g., 19 genes with $P < 1 \times 10^{-8}$; Supplemental Tables S5 and S6). The five most significant genes were TAP1, ZFHX4, CXCL9, FBN1 and PTGER3 (Table 1). We did not find extensive evidence of high co-expression between these five genes and genes measured in TCGA project (Supplemental Table S7). In sensitivity analyses we found that excluding samples from omentum and other extra-ovarian sites did not substantially affect the results (Supplemental Tables S8 and S9).

Development of a novel prognostic gene signature

The four predictive modelling approaches that were evaluated in the training data using 10-fold cross-validation yielded median AUCs that ranged from 0.69 to 0.73 for two-year OS and 0.69 to 0.74 for five-year survival (Supplemental Figure S2) with better prediction of 5-year overall survival than at two years. The elastic net approach yielded the highest median AUC for both two- and five-year OS and was selected for final development of the signature. Using the model on the full training data set resulted in a prognostic signature of 101 genes in addition to age and stage (Supplemental Table S10). Of these, 66 genes were associated with OS (FDR < 0.05) in the single gene models. There was no obvious subset of signature genes that performed as well or nearly as well as the full 101 gene signature (Supplemental Figure S3).

Performance of the signature including age and stage was AUC = 0.69 (95% CI 0.65-0.73) and 0.75 (95% CI 0.72-0.78) for 2-yr and 5-yr OS, respectively (Figure 2, Figure 3, Supplemental Figure S4). This was substantially better than age and stage alone with AUC = 0.61 (95% CI 0.57-0.65) and 0.62 (95% CI 0.59-0.67) for 2-yr and 5-yr OS, respectively), particularly for the 5-yr OS outcome with non-overlapping 95% CI. One standard deviation change in the gene expression score was associated with a hazard ratio of 2.35 [95% CI = (2.02, 2.71); $P = 5.1\times10^{-31}$], and median survival varied substantially across quintiles of the gene expression score [9.5 (8.3, --)].
Table 2].

For a subset of cases, there was clinical and experimental data for known prognostic factors. All samples had molecular subtype classification (Talhouk et al. submitted), residual disease was known for 1,771 cases, primary treatment for 687, germline BRCA mutation status for 904, and nuclear CD8 TIL counts [8] for 1,111 (Supplemental Table S11). When examined by quintile of gene expression score there were differences, as expected, for each of the known prognostic factors, including age and stage that were included in the model (Table 3). However, in sensitivity analyses, applying the signature to specific patient groups, a robustness of stratification was demonstrated, suggesting that the prognostic power of the signature is not explained by the individual factors, residual disease, treatment, BRCA status, or CD8 score (Figure 3, Supplemental Figures S5-S7). The signature score showed modest differences by molecular subtype (Supplemental Figure S8), and adjusting for molecular subtype in the Cox analysis resulted in only minor changes to the HR estimates for signature quintiles (Table 2). The signature was shown to be prognostic within a homogenous group of 316 stage 3C cases with no residual disease, within early stage cases (FIGO 1a and 1b), and within patients whose samples were collected from the omentum (Supplemental Figures S9-S10). Analysis of the signature score for paired ovary and omental tissue from 42 of the cases showed a highly significant Pearson correlation coefficient, $r = 0.79$ ($p = 5.4 \times 10^{-10}$) (Supplemental Figure S11).

A geneset enrichment analysis was performed for the 101 genes in the signature, as well as for genes correlated with signature genes achieving $r^2 > 0.75$ (Supplemental Table S12). For the correlated gene analysis, the three most significant pathways involved the immune system, including the adaptive immune system and cytokine signalling. A further ten immune pathways were significantly enriched and included interferon signalling, innate immune system, and TCR signalling and antigen presentation pathways. Restricting to the signature genes only, there was also enrichment in the immune system, but the top two pathways were PI-3K cascade and GPCR ligand binding. Four other pathways were related to the cell cycle and mitosis, with the
remaining enriched for fibroblast growth factor receptor (FGFR) and epidermal growth factor receptor (ERRB) signalling, and one pathway related to homologous combination repair.

Discussion

In a large-scale study of HGSOC patients, we identified a 101 gene expression signature able to predict clinically relevant differences in OS. Using methods that are both economical and applicable to standard clinical sampling techniques, we showed that the signature performs substantially better than age and stage alone for prognosis of both two- and five-year OS. The number of patients and samples included in this study is an order of magnitude greater than previous comparable studies of gene expression and OS in HGSOC patients[12, 20, 21]. Thus, we have been able to more precisely quantify the prognostic value of gene expression.

We report definitive associations between OS and expression of 276 genes. Of the five most significant genes (TAP1, ZFHX4, CXCL9, FBN1, and PTGER3), four have been previously reported to be associated with survival in HGSOC. The top prognostic gene, TAP1, is involved in the antigen presenting pathway. Expression was reduced in metastatic HGSOC, positively associated with OS[22] as observed here, and linked to tumour regression in response to treatment[23]. Also, hypomethylation of TAP1 was associated with improved time to disease recurrence[24]. CXCL9 is a chemokine that mediates the recruitment of T-cells to solid tumours[25]. High expression of intratumoural CXCL9 was associated with higher OS[26] and higher lymphocytic infiltration, which is also a robust prognostic factor in HGSOC[8, 11, 27] and a feature of the immunoreactive HGSOC molecular subtype[11]. CXCL9 has also been proposed as a therapeutic target due to evidence that it inhibits angiogenesis and promotes antitumour adaptive immunity[28-30]. Strikingly, the signature was able to further refine prognostic groups within patients with high TIL counts suggesting that CXCL9 and TAP1 expression may be strong indicators of immune competency in HGSOC.
*FBN1* is an extracellular matrix (ECM) protein previously found to be a biomarker associated with early recurrence in ovarian cancer patients who are initially sensitive to chemotherapy[31] and strongly correlated with desmoplasia in HGSOC. The prostaglandin E2 receptor *PTGER3* is expressed in ovarian tumour cells and is associated with relapse-free survival[32]. In contrast, *ZFHX4* does not have previous associations with HGSOC.

Associations between the expression of specific genes in tumour tissues and OS in HGSOC patients may suggest new drug targets and lead to insights into biological variation in treatment response. For example, cases in the Q5 quintile with the poorest outcome had increased expression of *IGF2*, *FGFR1*, and *MYC*, a possible argument for the use of *IGFR1*, *FGFR*, Bromodomain (*MYC*), or a combination of PARP and CDK4/6 inhibitors (*MYC*) [33]. More immediately, the signature may help clinicians identify patients most in need of intervention, such as patients that could potentially benefit from neo-adjuvant chemotherapy (NACT).

Alternatively, in clinical trials it could be used to stratify randomization by patients’ risk, thereby reducing heterogeneity within subgroups and increasing heterogeneity between subgroups. The signature will be incorporated into future prospective clinical trials to determine if it can predict response to specific treatments.

Measurement of the signature required standard FFPE tissue used in routine histopathology. Also, data preprocessing and normalization were conducted on an individual level, thus translatable to a general patient population. That is, 5-year OS prognosis of future patients can be evaluated against the patient population reported here by i) following the same steps described here for generating the normalized gene expression data, 2) computing an individual signature score, and 3) assigning an HR based on the score or comparing it to the reported quintiles (Supplemental Material). NanoString gene expression is highly reproducible as seen by our quality control metrics (Supplemental Material) and the FDA approval of the ProSigna test for breast cancer.
The question of heterogeneity by ancestry or ethnicity was beyond the scope of this study but should be pursued in future research. Another important question is whether molecular subtype can improve biomarker performance. A substantial proportion of signature genes were identified by the subtype adjusted meta-analysis, suggesting that the strong performance of the signature is not solely attributable to differences among molecular subtypes. Additionally, all of the individual genes used in the molecular subtype classification were included in development of the signature.

Although the cases received chemotherapy, the FFPE samples used in this study were chemo-naïve, as few patients had NACT during the calendar period in which these samples were collected. Because the signature appears to be prognostic in omentum samples, future studies may assess the value in NACT patients, using pre-treatment omental biopsies or post-treatment tumour samples. Future work will also address if the signature can predict platinum-refractory patients.

We have developed a robust prognostic signature for HGSOC that can be used to stratify patients and identify those in need of alternative treatments. Gene set enrichment analysis applied to the signature indicates an important role for the immune system in overall survival and supports further investigation of immune-therapy in ovarian cancer. More generally, the identification here of high-confidence prognostic genes may lead to new hypotheses for targeted treatments.

**Acknowledgements**

We thank all the study participants who contributed to this study and all the researchers, clinicians and technical and administrative staff who have made possible this work. This project received technical and data management support from OVCARE’s core units, including the Cheryl Brown Ovarian Cancer Outcomes Unit and the Genetic Pathology Evaluation Centre, and statistical analysis support from the Biostatistics Core of the
Norris Comprehensive Cancer Center. The AOV study recognizes the valuable contributions from Mie Konno, Shuhong Liu, Michelle Darago, Faye Chambers and the staff at the Tom Baker Cancer Centre Translational Laboratories. We thank Olivier Tredan and Pierre Heudel as investigators on the TRIO14 study and Sandrine Berge-Montamat as assistant for clinical research. The Australian Ovarian Cancer Study gratefully acknowledges additional support from Ovarian Cancer Australia and the Peter MacCallum Foundation. The AOCS also acknowledges the cooperation of the participating institutions in Australia and acknowledges the contribution of the study nurses, research assistants and all clinical and scientific collaborators to the study. The complete AOCS Study Group can be found at www.aocstudy.org. We would like to thank all of the women who participated in these research programs.

FUNDING

This work was funded by the National Cancer Institute (NCI) Grants R01CA172404 (to SJR) and R01CA168758 (to JAD and MAR), the Canadian Institutes for Health Research (Proof-of-Principle I program) and the United States Department of Defense Ovarian Cancer Research Program (OC110433). M. Milstein and S.J. Ramus received support from NIH/National Cancer Institute award number P30CA014089. M.S. Anglesio receives funding from the Janet D. Cottrelle Foundation Scholar’s program managed by the BC Cancer Foundation. J. George was partially supported by the NIH/National Cancer Institute award number P30CA034196. C.Wang was a Career Enhancement Awardee of the Mayo Clinic SPORE in Ovarian Cancer (P50 CA136393). D.G. Huntsman receives support from the Dr. Chew Wei Memorial Professorship in Gynecologic Oncology, the Canada Research Chairs program (Research Chair in Molecular and Genomic Pathology), and the Janet D. Cottrelle Foundation. M. Widschwendter receives funding from the European Union’s Horizon 2020 European Research Council Programme, H2020 BRCA-ERC under Grant Agreement No. 742432 as well as the charity, The Eve Appeal (https://eveappeal.org.uk/), and support of the National Institute for Health Research (NIHR) and the University College London Hospitals (UCLH) Biomedical Research Centre. G.E. Konecny is supported by the Miriam and Sheldon Adelson Medical Research Foundation. B.Y. Karlan is funded by the American Cancer
Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000124. H.R. Harris is supported by the NIH/National Cancer Institute award number K22 CA193860. OVCARE (including the VAN study) receives core funding through the BC Cancer Foundation and The VGH+UBC Hospital Foundation (authors AT, BG, DGH, and MSA). The AOV study is supported by the Canadian Institutes of Health Research (MOP-86727). The Gynaecological Oncology Biobank at Westmead, a member of the Australasian Biospecimen Network-Oncology group, was funded by the National Health and Medical Research Council Enabling Grants ID 310670 & ID 628903 and the Cancer Institute NSW Grants ID 12/RIG/1-17 & 15/RIG/1-16. The Australian Ovarian Cancer Study Group was supported by the U.S. Army Medical Research and Materiel Command under DAMD17-01-1-0729, The Cancer Council Victoria, Queensland Cancer Fund, The Cancer Council New South Wales, The Cancer Council South Australia, The Cancer Council Tasmania and The Cancer Foundation of Western Australia (Multi-State Applications 191, 211 and 182) and the National Health and Medical Research Council of Australia (NHMRC; ID199600; ID400413 and ID400281). BriTROC-1 was funded by Ovarian Cancer Action (to IAM and JDB, grant number 006) and supported by Cancer Research UK (grant numbers A15973, A15601, A18072, A17197, A19274 and A19694) and the National Institute for Health Research Cambridge and Imperial Biomedical Research Centres. SEARCH was supported by Cancer Research UK (A16561). The University of Cambridge receives salary support for PDPP from the NHS Clinical Academic Reserve (no grant number applicable). Samples from the Mayo Clinic were collected and provided with support of P50 CA136393 (ELG, GLK, SHK, MES).

Disclosure

Beth Y. Karlan served on Invitae Corporation’s Advisory Board from 2017 to 2018. Iain McNeish has acted on Advisory Boards for AstraZeneca, Clovis Oncology, Tesaro, Carrick Therapeutics and Takeda. His institution receives funding from AstraZeneca. Ros Glasspool in on the Advisory Boards for AstraZeneca, Tesaro, Clovis and Immunogen and does consultancy work for SOTIO. She has received support to attend conferences from...
AstraZeneca, Roche and Tesaro. Her institution has received research funding from Boehringer Ingelheim and Lilly/Ignyta and she is the national co-ordinating investigator for the UK for trials sponsored by AstraZeneca and Tesaro and site principal investigator for trials sponsored by AstraZeneca, Tesaro, Immunogen, Pfizer, Lilly and Clovis. Peter Fasching has received grants from Novartis, Biontech and Cepheid as well as personal fees from Novartis, Roche, Pfizer, Celgene, Daiichi-Sankyo, TEVA, Astra Zeneca, Merck Sharp & Dohme, Myelo Therapeutics, Macrogenics, Eisai and Puma during the conduct of the study. Usha Menon has shares in Abcodia Ltd. All remaining authors have declared no conflicts of interest.
References


**Figure Legends**

**Figure 1.** Schematic of study design. * The TRI study was split across the training and validation sets due to 107 samples overlapping with the meta-analysis.

**Figure 2.** ROC curves for prognostic performance of the gene expression signature in independent HGSOC patients (testing data). There was no overlap between studies or patient data used to develop models (training data) and compute ROC curves and AUC values shown here (testing data). All models included age and stage as described in Methods. TP denotes the true positive rate (sensitivity) and FP denotes the false positive rate (1 – specificity).

**Figure 3.** KM curves of overall survival for patients A) in the training and B) testing sets. Patients were assigned to quintiles (Q1-Q5) of the signature score including age and stage. Shaded areas indicate 95 percent confidence regions, only included for plots representing larger sample sizes. Due to limited sample size, the following plots represent all such patients in the entire data set, training or testing, C) no macroscopic residual disease after debulking surgery, D) primary treatment ≥ 4 cycles of IV carboplatin AUC 5 or 6 & paclitaxel 135 or 175 mg/m² every 3 weeks (actual dose known or presumed), E) *BRCA1* or *BRCA2* germline mutation, and F) CD8 > 19.
Appendix

AOCS STUDY GROUP

Management Group: D Bowtell, G Chenevix-Trench, A Green, P Webb, A DeFazio, D Gertig

Project and Data Managers: N Traficante, S Fereday, S Moore, J Hung, K Harrap, T Sadkowsky, N Pandeya

Research Nurses and Assistants:

Clinical and Scientific Collaborators:
1Peter MacCallum Cancer Centre, Melbourne, Victoria, 3000, Australia.

2QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia.

3Department of Pathology, University of Melbourne, Parkville, Victoria, 3052, Australia.

4Sir Peter MacCallum Cancer Centre Department of Oncology, University of Melbourne, Parkville, Victoria, 3052, Australia.

5Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria, 3052, Australia.

6Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, Imperial College London, London, England, W12 0HS, UK.

7Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney and Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, 2145, Australia.

8Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria, 3052, Australia.

9John Hunter Hospital, Lookout Road, New Lambton, New South Wales, 2305, Australia

10Royal Hospital for Women, Barker Street, Randwick, New South Wales, 2031, Australia

11Royal North Shore Hospital, Reserve Road, St Leonards, New South Wales, 2065, Australia

12Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales, 2050, Australia

13Royal Adelaide Hospital, North Terrace, Adelaide, South Australia, 5000, Australia

14Royal Hobart Hospital, 48 Liverpool St, Hobart, Tasmania, 7000, Australia

15Monash Medical Centre, 246 Clayton Rd, Clayton, Victoria, 3168, Australia

16Western Australian Research Tissue Network (WARTN), St John of God Pathology, 23 Walters Drive, Osborne Park, Western Australia, 6017, Australia

17Women and Infant's Research Foundation, King Edward Memorial Hospital, 374 Bagot Road, Subiaco, Western Australia, 6008, Australia
18St John of God Hospital, 12 Salvado Rd, Subiaco, Western Australia, 6008, Australia
19Canberra Hospital, Yamba Drive, Garran, Australian Capitol Territory, 2605, Australia
20Bankstown Cancer Centre, Bankstown Hospital, 70 Eldridge Road, Bankstown, New South Wales, 2200, Australia
21Northern Haematology & Oncology Group, Integrated Cancer Centre, 185 Fox Valley Road, Wahroonga, New South Wales, 2076, Australia
22Illawarra Shoalhaven Local Health District, Wollongong Hospital, Level 4 Lawson House, Wollongong, New South Wales, 2500, Australia
23Nepean Hospital, Derby Street, Kingswood, New South Wales, 2747, Australia
24Newcastle Mater Misericordiae Hospital, Edith Street, Waratah, New South Wales, 2298, Australia
25Port Macquarie Base Hospital, Wrights Road, Port Macquarie, New South Wales, 2444, Australia
26Prince of Wales Clinical School, University of New South Wales, New South Wales, 2031, Australia
27St George Hospital, Gray Street, Kogarah, New South Wales, 2217, Australia
28St Vincent’s Hospital, 390 Victoria Street, Darlinghurst, New South Wales, 2010, Australia
29Wagga Wagga Base Hospital, Docker St, Wagga Wagga, New South Wales, 2650, Australia
30Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, Sydney, New South Wales, 2145, Australia.
31Department of Pathology, Westmead Clinical School, Westmead Hospital, The University of Sydney, New South Wales, 2006, Australia
32Mater Misericordiae Hospital, Raymond Terrace, South Brisbane, Queensland, 4101, Australia
33The Royal Brisbane and Women’s Hospital, Butterfield Street, Herston, Queensland, 4006, Australia
34Wesley Hospital, 451 Coronation Drive, Auchenflower, Queensland, 4066, Australia
35Burnside Hospital, 120 Kensington Road, Toorak Gardens, South Australia, 5065, Australia
36Flinders Medical Centre, Flinders Drive, Bedford Park, South Australia, 5042, Australia
37Queen Elizabeth Hospital, 28 Woodville Road, Woodville South, South Australia, 5011, Australia
38Freemasons Hospital, 20 Victoria Parade, East Melbourne, Victoria, 3002, Australia
The six people underlined are named authors on the manuscript.
Table 1. Hazard ratios and 95% CIs for top 5 prognostic genes in covariate-adjusted single gene analyses.

<table>
<thead>
<tr>
<th>Gene</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>Selection</th>
<th>Correlated gene*</th>
<th>rs</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAP1</td>
<td>0.84 (0.80, 0.87)</td>
<td>8.3x10^{-18}</td>
<td>Meta</td>
<td>PSMB9</td>
<td>0.89</td>
</tr>
<tr>
<td>ZFH4</td>
<td>1.19 (1.14, 1.25)</td>
<td>1.4x10^{-15}</td>
<td>Meta</td>
<td>LOC100192378</td>
<td>0.74</td>
</tr>
<tr>
<td>CXCL9</td>
<td>0.85 (0.82, 0.88)</td>
<td>1.8x10^{-15}</td>
<td>Meta and candidate</td>
<td>CXCR6</td>
<td>0.89</td>
</tr>
<tr>
<td>FBN1</td>
<td>1.18 (1.13, 1.24)</td>
<td>4.2x10^{-14}</td>
<td>Candidate</td>
<td>SPARC^</td>
<td>0.91</td>
</tr>
<tr>
<td>PTGER3</td>
<td>1.18 (1.13, 1.24)</td>
<td>1.2x10^{-13}</td>
<td>Meta</td>
<td>COL8A1</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*Most correlated gene according to Spearman’s rank correlation coefficient, r_s, computed in The Cancer Genome Atlas (TCGA) Ovarian Serous Cystadenocarcinoma RNA-seq data set.

^SPARC was included in this project and was less significant.
Table 2. Hazard ratios and 95% CIs for quintiles of the gene expression signature score in validation data.

<table>
<thead>
<tr>
<th>Quintile</th>
<th>N</th>
<th>Deaths</th>
<th>Median Survival*</th>
<th>HR (95% CI)</th>
<th>Adjusted for Age and Stage HR (95% CI)</th>
<th>Adjusted for M. Subtype Age and Stage HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>214</td>
<td>81</td>
<td>9.47 (8.32, ------)</td>
<td>0.44 (0.33, 0.58)</td>
<td>0.34 (0.22, 0.55)</td>
<td>0.37 (0.23, 0.59)</td>
</tr>
<tr>
<td>Q2</td>
<td>213</td>
<td>117</td>
<td>5.38 (4.63, 6.97)</td>
<td>0.73 (0.57, 0.93)</td>
<td>0.71 (0.55, 0.91)</td>
<td>0.74 (0.58, 0.96)</td>
</tr>
<tr>
<td>Q3</td>
<td>213</td>
<td>145</td>
<td>3.80 (3.34, 4.60)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>213</td>
<td>158</td>
<td>3.23 (2.85, 3.68)</td>
<td>1.56 (1.25, 1.96)</td>
<td>1.56 (1.24, 1.97)</td>
<td>1.56 (1.24, 1.96)</td>
</tr>
<tr>
<td>Q5</td>
<td>214</td>
<td>179</td>
<td>2.27 (2.09, 2.62)</td>
<td>2.23 (1.78, 2.78)</td>
<td>2.11 (1.67, 2.67)</td>
<td>2.07 (1.63, 2.61)</td>
</tr>
</tbody>
</table>

*Median survival (95% CI) in years for patients in the validation set.
Table 3. Clinical data for the 3769 patients that passed quality control and the percentage of patients in each quintile of the gene expression score.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3769</td>
<td>754</td>
<td>754</td>
<td>753</td>
<td>754</td>
<td>754</td>
<td></td>
</tr>
<tr>
<td>median survival (years)</td>
<td>4.1</td>
<td>9.5</td>
<td>5.4</td>
<td>3.8</td>
<td>3.2</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>% 5-year survival</td>
<td>41</td>
<td>75</td>
<td>57</td>
<td>39</td>
<td>25</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Age median</td>
<td>63</td>
<td>58</td>
<td>57</td>
<td>61</td>
<td>64</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Age range</td>
<td>25-89</td>
<td>39-78</td>
<td>25-86</td>
<td>36-82</td>
<td>27-89</td>
<td>39-86</td>
<td></td>
</tr>
<tr>
<td>Age quartile q1</td>
<td>894</td>
<td>30.8</td>
<td>31.3</td>
<td>20.0</td>
<td>13.4</td>
<td>4.5</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>Age quartile q2</td>
<td>838</td>
<td>21.5</td>
<td>20.0</td>
<td>22.9</td>
<td>21.2</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Age quartile q3</td>
<td>961</td>
<td>16.0</td>
<td>20.2</td>
<td>21.4</td>
<td>23.6</td>
<td>18.7</td>
<td></td>
</tr>
<tr>
<td>Age quartile q4</td>
<td>1076</td>
<td>13.5</td>
<td>10.4</td>
<td>16.4</td>
<td>21.3</td>
<td>38.5</td>
<td></td>
</tr>
<tr>
<td>FIGO stage I / II</td>
<td>607</td>
<td>97.4</td>
<td>2.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>FIGO stage III/IV</td>
<td>3067</td>
<td>3.8</td>
<td>23.0</td>
<td>24.1</td>
<td>24.4</td>
<td>24.6</td>
<td></td>
</tr>
<tr>
<td>Primary chemo* 1</td>
<td>136</td>
<td>16.2</td>
<td>22.1</td>
<td>23.5</td>
<td>19.1</td>
<td>19.1</td>
<td>0.163</td>
</tr>
<tr>
<td>Primary chemo* 2</td>
<td>190</td>
<td>16.3</td>
<td>20.0</td>
<td>21.6</td>
<td>22.1</td>
<td>20.0</td>
<td></td>
</tr>
<tr>
<td>Primary chemo* 3</td>
<td>361</td>
<td>11.1</td>
<td>16.9</td>
<td>22.4</td>
<td>20.5</td>
<td>29.1</td>
<td></td>
</tr>
<tr>
<td>Residual disease No</td>
<td>614</td>
<td>32.4</td>
<td>22.1</td>
<td>17.8</td>
<td>15.5</td>
<td>12.2</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>Residual disease Yes</td>
<td>1157</td>
<td>6.0</td>
<td>19.2</td>
<td>24.1</td>
<td>24.5</td>
<td>26.2</td>
<td></td>
</tr>
<tr>
<td>germline BRCA1 mutation</td>
<td>130</td>
<td>23.8</td>
<td>31.5</td>
<td>26.2</td>
<td>11.5</td>
<td>6.9</td>
<td>2.22x10^-7</td>
</tr>
<tr>
<td>germline BRCA2 mutation</td>
<td>71</td>
<td>28.2</td>
<td>26.8</td>
<td>18.3</td>
<td>18.3</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>germline no mutation</td>
<td>663</td>
<td>19.6</td>
<td>16.7</td>
<td>18.7</td>
<td>20.7</td>
<td>24.3</td>
<td></td>
</tr>
<tr>
<td>CD8 TIL score 0</td>
<td>192</td>
<td>19.8</td>
<td>14.6</td>
<td>12.5</td>
<td>21.4</td>
<td>31.8</td>
<td>2.46x10^-14</td>
</tr>
<tr>
<td>CD8 TIL score 1-2</td>
<td>186</td>
<td>18.3</td>
<td>14.0</td>
<td>18.8</td>
<td>21.5</td>
<td>27.4</td>
<td></td>
</tr>
<tr>
<td>CD8 TIL score 3-19</td>
<td>515</td>
<td>19.8</td>
<td>24.1</td>
<td>20.8</td>
<td>17.9</td>
<td>17.5</td>
<td></td>
</tr>
<tr>
<td>CD8 TIL score &gt;20</td>
<td>218</td>
<td>34.4</td>
<td>31.2</td>
<td>16.5</td>
<td>11.5</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>Molecular subtype C1.MES</td>
<td>1105</td>
<td>5.4</td>
<td>10.4</td>
<td>20.7</td>
<td>27.4</td>
<td>36.0</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>Molecular subtype C2.IMM</td>
<td>907</td>
<td>23.2</td>
<td>28.8</td>
<td>21.2</td>
<td>16.2</td>
<td>10.7</td>
<td></td>
</tr>
<tr>
<td>Molecular subtype C4.DIF</td>
<td>1144</td>
<td>32.6</td>
<td>25.5</td>
<td>17.9</td>
<td>12.8</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>Molecular subtype C5.PRO</td>
<td>613</td>
<td>18.1</td>
<td>14.0</td>
<td>20.7</td>
<td>25.8</td>
<td>21.4</td>
<td></td>
</tr>
<tr>
<td>FIGO stage 1A &amp; 1B</td>
<td>111</td>
<td>96.4</td>
<td>3.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>FIGO stage 3C</td>
<td>1979</td>
<td>3.1</td>
<td>23.7</td>
<td>24.6</td>
<td>24.1</td>
<td>24.6</td>
<td>&lt;1x10^-50</td>
</tr>
<tr>
<td>FIGO stage 3C Residual</td>
<td>316</td>
<td>6.3</td>
<td>31.0</td>
<td>24.4</td>
<td>20.9</td>
<td>17.4</td>
<td>6.24x10^-45</td>
</tr>
<tr>
<td>FIGO stage 3C Residual</td>
<td>846</td>
<td>2.6</td>
<td>21.5</td>
<td>25.3</td>
<td>24.6</td>
<td>26.0</td>
<td></td>
</tr>
</tbody>
</table>

Q1 is the quintile with the best survival and Q5 the worst survival. Samples with missing data are reported in Supplementary Table S11. P-values for BRCA1/2 mutation status were calculated for BRCA1 or BRCA2 mutation vs no mutation. * Treatment: 1 = known to have received first line chemotherapy treatment of ≥ 4 cycles of IV carboplatin AUC 5 or 6 & paclitaxel 135 or 175 mg/m² every 3 weeks. 2 = known to have received first line chemotherapy treatment of ≥ 4 cycles of IV carboplatin & paclitaxel 3-weekly but at doses presumed to be
carboplatin AUC 5 or 6 & paclitaxel 135 or 175 mg/m². 3 = all remaining cases with chemo regimens that do not fit criteria 1 or 2 and include unknown or no chemotherapy.