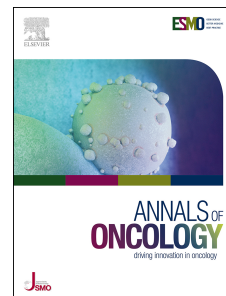# Journal Pre-proof

Prognostic gene expression signature for high-grade serous ovarian cancer

J. Millstein, T. Budden, E.L. Goode, M.S. Anglesio, A. Talhouk, M.P. Intermaggio, H.S. Leong, S. Chen, W. Elatre, B. Gilks, T. Nazeran, M. Volchek, R.C. Bentley, C. Wang, D.S. Chiu, S. Kommoss, S.C.Y. Leung, J. Senz, A. Lum, V. Chow, H. Sudderuddin, R. Mackenzie, J. George, AOCS Group, S. Fereday, J. Hendley, N. Traficante, H. Steed, J.M. Koziak, M. Köbel, I.A. McNeish, T. Goranova, D. Ennis, G. Macintyre, D. Silva, T. Ramón y Cajal, J. García-Donas, S. Hernando Polo, G.C. Rodriguez, K.L. Cushing-Haugen, H.R. Harris, C.S. Greene, R.A. Zelaya, S. Behrens, R.T. Fortner, P. Sinn, E. Herpel, J. Lester, J. Lubiński, O. Oszurek, A. Tołoczko, C. Cybulski, J. Menkiszak, C.L. Pearce, M.C. Pike, C. Tseng, J. Alsop, V. Rhenius, H. Song, M. Jimenez-Linan, A. Piskorz, A. Gentry-Maharaj, C. Karpinskyj, M. Widschwendter, N. Singh, C.J. Kennedy, R. Sharma, P.R. Harnett, B. Gao, S.E. Johnatty, R. Sayer, J. Boros, S.J. Winham, G.L. Keeney, S.H. Kaufmann, M.C. Larson, H. Luk, B.Y. Hernandez, P.J. Thompson, L.R. Wilkens, M.E. Carney, B. Trabert, J. Lissowska, L. Brinton, M.E. Sherman, C. Bodelon, S. Hinsley, L.A. Lewsley, R. Glasspool, S.N. Banerjee, E.A. Stronach, P. Haluska, I. Ray-Coquard, S. Mahner, B. Winterhoff, D. Slamon, D.A. Levine, L.E. Kelemen, J. Benitez, J. Chang-Claude, J. Gronwald, A.H. Wu, U. Menon, M.T. Goodman, J.M. Schildkraut, N. Wentzensen, R. Brown, A. Berchuck, G. Chenevix-Trench, A. deFazio, S.A. Gayther, M.J. García, M. Henderson, M.A. Rossing, A. Beeghly-Fadiel, P.A. Fasching, S. Orsulic, B.Y. Karlan, G.E. Konecny, D.G. Huntsman, D.D. Bowtell, J.D. Brenton, J.A. Doherty, P.D.P. Pharoah, S.J. Ramus

Original article

Prognostic gene expression signature for high-grade serous ovarian cancer

J. Millstein[1], T. Budden[2,3], E. L. Goode[4], M. S Anglesio[5-7], A. Talhouk[5,7], M. P. Intermaggio[2], H. S. Leong[8], S. Chen[9], W. Elatre[10], B. Gilks[5,6], T. Nazeran[5], M. Volchek[11], R. C. Bentley[12], C. Wang[13], D. S. Chiu[5], S. Kommoss[14], S. C. Y. Leung[5], J. Senz[5,6], A. Lum[5], V. Chow[5], H. Sudderuddin[5], R. Mackenzie[5], J. George[15], AOCS Group[8,16,17], S. Fereday[8,18], J. Hendley[8,18], N. Traficante[8,18], H. Steed[19], J. M. Koziak[20], M. Köbel[21], I. A. McNeish[22,23], T. Goranova[24], D. Ennis[22,23], G. Macintyre[24], D. Silva[24], T. Ramón y Cajal[25], J. García-Donas[26], S. Hernando Polo[27], G. C. Rodriguez[28], K. L. Cushing-Haugen[29], H. R. Harris[29,30], C. S. Greene[31], R. A. Zelaya[32], S. Behrens[33], R. T. Fortner[33], P. Sinn[34], E. Herpel[35], J. Lester[36,37], J. Lubiński[38], O. Oszurek[38], A. Tołoczko[38], C. Cybulski[38], J. Menkiszak[39], C. L. Pearce[40,41], M. C. Pike[41,42], C. Tseng[43], J. Alsop[44], V. Rhenius[44], H. Song[44], M. Jimenez-Linan[45], A. Piskorz[24], A. Gentry-Maharaj[46], C. Karpinskyj[46], M. Widschwendter[47], N. Singh[48], C. J. Kennedy[17,49], R. Sharma[50,51], P. R. Harnett[17,52], B. Gao[17,52], S. E. Johnatty[16], R. Sayer[49], J. Boros[17,49], S. J. Winham[13], G. L. Keeney[53], S. H. Kaufmann[54,55], M. C. Larson[13], H. Luk[56], B. Y. Hernandez[56], P. J. Thompson[57], L. R. Wilkens[56], M. E. Carney[58], B. Trabert[59], J. Lissowska[60], L. Brinton[59], M. E. Sherman[61], C. Bodelon[59], S. Hinsley[62], L. A. Lewsley[62], R. Glasspool[63], S. N. Banerjee[64], E. A. Stronach[22], P. Haluska[54], I. Ray-Coquard[65], S. Mahner[66], B. Winterhoff[67], D. Slamon[68], D. A. Levine[69,70], L. E. Kelemen[71], J. Benitez[72,73], J. Chang-Claude[33,74], J. Gronwald[38], A. H. Wu[43], U. Menon[46], M. T. Goodman[57], J. M. Schildkraut[75], N. Wentzensen[59], R. Brown[76], A. Berchuck[77], G. Chenevix-Trench[16], A. deFazio[17,49], S. A. Gayther[78], M. J. García[73,79], M. Henderson[80], M. A. Rossing[29,30], A. Beeghly-Fadiel[81], P. A. Fasching[68,82], S. Orsulic[37], B. Y. Karlan[36,37], G. E. Konecny[68], D. G. Huntsman[5-7,83], D. D. Bowtell[8,18], J. D. Brenton[24], J. A. Doherty[84], P. D. P. Pharoah[44,85], S. J. Ramus[2,86] *

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.

4.   Department of Health Science Research, Division of Epidemiology. Mayo Clinic. Rochester, MN: USA; 55905.

5.   British Columbia's Ovarian Cancer Research (OVCARE) Program. BC Cancer, Vancouver General Hospital, and University of British Columbia. Vancouver, BC: Canada; V5Z 4E6.

6.   Department of Pathology and Laboratory Medicine. University of British Columbia. Vancouver, BC: Canada; V5Z 4E6.

7.   Department of Obstetrics and Gynecology. University of British Columbia. Vancouver, BC: Canada.

8.   Peter MacCallum Cancer Center. Melbourne, Victoria: Australia; 3000.

9.   Center for Cancer Prevention and Translational Genomics, Samuel Oschin Comprehensive Cancer Institute. Cedars-Sinai Medical Center. Los Angeles, CA: USA; 90048.

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

11.   Anatomical Pathology. Royal Women's Hospital. Parkville, Victoria: Australia.

12.   Department of Pathology. Duke University Hospital. Durham, NC: USA; 27710.

13.   Department of Health Science Research, Division of Biomedical Statistics and Informatics. Mayo Clinic. Rochester, MN: USA; 55905.

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

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

16.   Department of Genetics and Computational Biology. QIMR Berghofer Medical Research Institute. Brisbane, Queensland: Australia; 4006.

17.   Centre for Cancer Research, The Westmead Institute for Medical Research. The University of Sydney. Sydney, New South Wales: Australia; 2145.

18.   Sir Peter MacCallum Department of Oncology. The University of Melbourne. Parkville, Victoria: Australia; 3000.

19.   Department of Obstetrics and Gynecology, Division of Gynecologic Oncology. Royal Alexandra Hospital. Edmonton, Alberta: Canada; T5H 3V9.

20.   Alberta Health Services-Cancer Care. Calgary, AB: Canada.

21.   Department of Pathology and Laboratory Medicine. University of Calgary, Foothills Medical Center. Calgary, AB: Canada; T2N 2T9.

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.

24.   Cancer Research UK Cambridge Institute. University of Cambridge. Cambridge: UK; CB2 0RE.

25.   Medical Oncology Service. Hospital Sant Pau. Barcelona: Spain.

26.   HM Hospitales Đ Centro Integral Oncol—gico HM Clara Campal. Madrid: Spain.

27.   Medical Oncology Service. Hospital Universitario Funcacion Alcorcon. Alcorc—n: Spain.

28.   Division of Gynecologic Oncology. NorthShore University HealthSystem, University of Chicago. Evanston, IL: USA; 60201.

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

30.   Department of Epidemiology. University of Washington. Seattle, WA: USA; 98195.

31.   Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine. University of Pennsylvania. Philadelphia: USA; PA 19103.

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.

37.   Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute. Cedars-Sinai Medical Center. Los Angeles, CA: USA; 90048.

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.

40.   Department of Epidemiology. University of Michigan School of Public Health. Ann Arbor, MI: USA; 48109.

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.

52.   The Crown Princess Mary Cancer Centre Westmead, Sydney-West Cancer Network. 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.

54.   Department of Oncology. 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.

57.   Samuel Oschin Comprehensive Cancer Institute, Cancer Prevention and Genetics Program. Cedars-Sinai Medical Center. Los Angeles, CA: USA; 90048.

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

59.   Division of Cancer Epidemiology and Genetics. National Cancer Institute. Bethesda, MD: USA; 20892.

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

61.   Department of Health Sciences Research. Mayo Clinic College of Medicine. Jacksonville, FL: USA; 32224.

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.

65.   Centre Leon Berard and University Claude Bernard Lyon 1. Lyon: France; 69373.

66.   Department of Gynecology and Obstetrics. Ludwig Maximilian University of Munich. Munich: Germany; 80336.

67.   Department of Obstetrics, Gynecology and Women's Health. University of Minnesota. Minneapolis, MN: USA; 55455.

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.

69.   Gynecology Service, Department of Surgery. Memorial Sloan Kettering Cancer Center. New York, NY: USA; 10065.

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.

72.   Centro de Investigaci—n en Red de Enfermedades Raras (CIBERER). Madrid: Spain; 28029.

73.   Human Cancer Genetics Programme. Spanish National Cancer Research Centre (CNIO). Madrid: Spain; 28029.

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.

5

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

77.   Department of Gynecologic Oncology. Duke University Hospital. Durham, NC: USA; 27710.

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.

80.   Children's Cancer Institute, Lowy Cancer Research Centre. University of NSW Sydney. Sydney, New South Wales: Australia; 2052.

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.

83.   Department of Molecular Oncology. BC Cancer Research Centre. Vancouver, BC: Canada; V5Z 4E6.

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.

86.   Adult Cancer Program, Lowy Cancer Research Centre. University of NSW Sydney. Sydney, New South Wales: Australia; 2052.

* **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

1   **Abstract**

2   **Background**

3   Median overall survival (OS) for women with high-grade serous ovarian cancer (HGSOC) is approximately four

4   years, yet survival varies widely between patients. There are no well-established, gene expression signatures

5   associated with prognosis. The aim of this study was to develop a robust prognostic signature for overall

6   survival in HGSOC patients.

7   **Patients and methods**

8   Expression of 513 genes, selected from a meta-analysis of 1455 tumours and other candidates, were measured

9   using NanoString technology from formalin-fixed, paraffin-embedded (FFPE) tumour tissue from 3,769 women

10   with HGSOC from multiple studies. Elastic net regularization for survival analysis was applied to develop a

11   prognostic model for 5-year OS, trained on 2702 tumours from fifteen studies and evaluated on an

12   independent set of 1067 tumours from six studies.

13   **Results**

14   Expression levels of 276 genes were associated with OS [false discovery rate (FDR) < 0.05] in covariate-adjusted

15   single gene analyses. The top five genes were *TAP1, ZFHX4, CXCL9, FBN1,* and *PTGER3* ($P < 0.001$). The best

16   performing prognostic signature included 101 genes enriched in pathways with treatment implications. Each

17   gain of one standard deviation in the gene expression score (GES) conferred a greater than two-fold increase in

18   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),

19   3.8 (3.3, 4.6), 3.2 (2.9, 3.7) and 2.3 (2.1, 2.6) years.

20

21

22   **Conclusion**

23   The OTTA-SPOT (Ovarian Tumor Tissue Analysis consortium - Stratified Prognosis of Ovarian Tumours) gene

24   expression signature may improve risk stratification in clinical trials by identifying patients who are least likely

25   to achieve 5-year survival. The identified novel genes associated with the outcome may also yield

26   opportunities for the development of targeted therapeutic approaches.

27

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

29   paraffin embedded

30

31   # Highlights

32   - A gene expression signature for high-grade serous ovarian cancer prognostic for two- and five-year

33   overall survival (OS).

34   - The 101 gene expression signature performs substantially better than age and stage alone.

35   - Median survival by quintile was 9.5, 5.4, 3.8, 3.2 and 2.3 years.

36   - The top five genes associated with OS were *TAP1, ZFHX4, CXCL9, FBN1,* and *PTGER3* ($P \ll 0.001$).

Introduction

37

38 Epithelial ovarian cancer (EOC) causes approximately 125,000 deaths globally every year, and long-term

39 survival rates have changed little in the past three decades[1]. Approximately 70% of women with EOC are

40 diagnosed with advanced stage disease (stages III/IV), and fewer than 50% will survive more than 5 years[2].

41 There are five major EOC histotypes: high-grade serous, low-grade serous, endometrioid, clear cell and

42 mucinous[3]. High-grade serous ovarian cancer (HGSOC) comprises about two-thirds of cases, is responsible

43 for most deaths and is characterized by profound genomic and clinical heterogeneity.

44 The most informative prognostic factors for HGSOC are International Federation of Gynecology and Obstetrics

45 (FIGO) stage, residual disease following debulking surgery[4], *BRCA1* or *BRCA2* germline mutation[5, 6] and

46 tumour-infiltrating lymphocyte scores[7, 8]. Patients with HGSOC who carry a loss-of-function germline

47 mutation in *BRCA1* or *BRCA2* have an increased sensitivity to platinum-based chemotherapy and PARP

48 inhibitor treatment[9, 10] and a medium-term survival advantage[5]. However, the frequent development of

49 drug-resistant disease[6] limits the effectiveness of current therapies.

50 Gene-expression data have been used to define four tumour molecular subtypes of HGSOC (C1/mesenchymal,

51 C2/immune, C4/differentiated and C5/proliferative)[11, 12]. Using transcriptome-wide data from fresh frozen

52 tissues, The Cancer Genome Atlas (TCGA) project used 215 tumours to identify an overall survival (OS)

53 expression signature of 193 genes that has been validated on three other HGSOC gene expression data

54 sets[12].

55 Despite these findings, gene expression biomarkers have not been implemented clinically owing to several

56 important shortcomings. The majority of the individual markers comprising the 193 gene signature were not

57 statistically significant across all studies, suggesting that the signature may not be robust. The sample sizes in

58 other discovery efforts have been too small for robust statistical inference [12]. Also, previous studies used

10

59 fresh frozen samples, resulting in logistic and cost barriers to examining large clinically relevant data sets, and

60 translation to the clinical setting.

61 The aim of this study was to identify a robust and clinic-ready prognostic HGSOC profile that can be applied to

62 formalin fixed paraffin embedded (FFPE) tumour tissue.

Patients and methods

64 Twenty studies provided pre-treatment, FFPE tumour samples from 4,071 women diagnosed with HGSOC

65 (Supplemental Table S1). All HGSOC cases with available tissue were included. During this time period, HGSOC

66 patients were treated with chemotherapy (carboplatin and paclitaxel) after primary debulking surgery. Study

67 protocols were approved by the respective Institutional Review Board / ethics approval committee for each

68 site (Supplemental Table S1).

69 A schematic of the overall study design is shown in Figure 1. There were four main components: gene

70 selection, gene-expression assay, development of prognostic gene signature in a training set and validation of

71 prognostic signature in an independent validation set.

Gene selection

73 Candidate prognostic genes were identified by carrying out an individual participant meta-analysis of six

74 transcriptome-wide microarray studies[11-16], which included tumour samples from 1,455 participants. Gene

75 expression association with overall survival was evaluated by Cox proportional hazards regression adjusted for

76 molecular subtype (Supplemental Table S2). In total, 200 genes from the meta-analysis, most achieving a

77 permutation-based FDR[17] of less than 0.05, and an additional 313 candidate genes based on the literature

78 and unpublished data were selected (Supplemental Tables S3 and S4, Figure S1; for more details see

79 Supplemental Material). Five genes, RPL19, ACTB, PGK1, SDHA, and POLR1B, were included as house-keeping

80 genes for normalization.

81 **Gene expression assay in study participants samples**

82 FFPE tumour samples were processed with the NanoString nCounter technology at 3 different locations,

83 Vancouver, Los Angeles and Melbourne. A control set of 48 FFPE tumour samples were run at each location

84 and the average intraclass correlation coefficient (ICC) was 0.987. Approximately 2 percent of the samples

85 were run in duplicate and the average Spearman correlation r^2 was 0.995. Single-patient classification methods

86 were used with reference samples to control for batch effects[18]. The data in this publication have been

87 deposited in NCBI's Gene Expression Omnibus[19]; GEO Series accession number GSE132342

88 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132342). 3,329 samples passed quality control of

89 which 3,769 had survival data and assessable gene expression for 513 genes. Data can be found in NCBI GEO:

90 Accession numbers GSE132342 and GPL26748.

91 **Overall survival analysis of individual genes**

92 Samples that contributed to the meta-analysis data set (n=211) were removed from subsequent selected

93 analyses to enforce independence of study samples between the gene selection and final survival analysis.

94 Time-to-event analyses were carried out for OS with right-censoring at 10 years and left-truncation of

95 prevalent cases. Associations between log-transformed normalized gene expression and survival time were

96 tested using likelihood ratio tests with Cox proportional hazards models adjusted for age, race, and stage, and

97 stratified by study. Patients with missing race or stage information were assigned to 'unknown' categories. Age

98 was modelled using a B-spline with a knot at the median age, which yielded a better fit than using knots at

99 quartiles or categorical variables. Stage was dichotomized into early (International Federation of Gynecology

100 and Obstetrics [FIGO] stage I/II) and advanced (FIGO stage III/IV). Genes were scaled to have a standard

101 deviation of one, so hazard ratios correspond to a change of one standard deviation. A Benjamini-Hochberg

102 (BH) false discovery rate (FDR) of less than 0.05 was used to identify notable associations. Since the expression

103 of genes can be correlated, an analysis of correlated genes was performed using data from TCGA. Advanced

104   stage ovarian cancer usually has disease spread throughout the abdomen, therefore sensitivity analyses were

105   performed to assess effects of the anatomical location of tumour samples included in the study by removing

106   observations corresponding to samples known to be extraovarian (n = 437).

107   **Prognostic signature development and validation**

108   Studies were initially randomized to training set (N = 14) and validation set (N = 6). The TRI study was

109   randomized to the validation set, but, because 107 of the samples were part of the meta-analysis data used for

110   gene selection, the study was split, so those 107 samples were included in the model training data set. Thus

111   2,702 samples from 15 studies were used for model training and 1,067 samples from 6 studies were used for

112   validation (Supplemental Table S1). In the training set, four modelling approaches (stepwise regression, elastic

113   net regularized regression, boosting and random survival forests) were applied to construct competing gene

114   expression-based biomarkers. Each was evaluated in the training data using 10-fold cross-validation for its

115   prognostic value for OS at two and five years of follow-up using an area under the curve (AUC) measure

116   derived from receiver operator characteristic (ROC) analysis (see Supplemental Material for additional details).

117   The best performing method, elastic net regularized regression, was applied to the full training set to

118   determine the final gene signature and scoring method, which was then evaluated using the independent

119   testing set. All models were constrained to include age and stage, where age was modelled as categorical

120   based on quartiles of the training dataset with groups: less than 53 years old, 53 to 59, 60 to 66, and 67 or

121   greater. Stage was modelled as described above for the OS individual gene analysis.

122   ## Results

123   **Association of expression of individual genes with OS in HGSOC.**

124   In a gene-by-gene analysis of the full data set adjusted for age, race, and stage, and stratified by study, 276 of

125   the 513 selected genes were associated with OS (FDR < 0.05). Of these, 138 were selected from the meta-

126   analysis of six published microarray studies (Supplemental Table S2)[11-16] and 144 from candidate gene

127   approaches (Supplemental Tables S5 and S6). Hazard ratios (HR) for one standard deviation change in gene

128   expression ranged from 0.84–1.19, with multiple genes exhibiting associations at very stringent significance

129   levels (e.g., 19 genes with $P < 1 \times 10^{-8}$; Supplemental Tables S5 and S6). The five most significant genes were

130   *TAP1*, *ZFHX4*, *CXCL9*, *FBN1* and *PTGER3* (Table 1). We did not find extensive evidence of high co-expression

131   between these five genes and genes measured in TCGA project (Supplemental Table S7). In sensitivity analyses

132   we found that excluding samples from omentum and other extra-ovarian sites did not substantially affect the

133   results (Supplemental Tables S8 and S9).

134   **Development of a novel prognostic gene signature**

135   The four predictive modelling approaches that were evaluated in the training data using 10-fold cross-

136   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

137   survival (Supplemental Figure S2) with better prediction of 5-year overall survival than at two years. The

138   elastic net approach yielded the highest median AUC for both two- and five-year OS and was selected for final

139   development of the signature. Using the model on the full training data set resulted in a prognostic signature

140   of 101 genes in addition to age and stage (Supplemental Table S10). Of these, 66 genes were associated with

141   OS (FDR < 0.05) in the single gene models. There was no obvious subset of signature genes that performed as

142   well or nearly as well as the full 101 gene signature (Supplemental Figure S3).

143   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-

144   0.78) for 2-yr and 5-yr OS, respectively (Figure 2, Figure 3, Supplemental Figure S4). This was substantially

145   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-

146   yr OS, respectively), particularly for the 5-yr OS outcome with non-overlapping 95% CI. One standard deviation

147   change in the gene expression score was associated with a hazard ratio of 2.35 [95% CI = (2.02, 2.71); $P =$

148   5.1×10^{-31}], and median survival varied substantially across quintiles of the gene expression score [9.5 (8.3, ---),

14

149   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, respectively, from smallest to largest quintile;

150   Table 2].

151   For a subset of cases, there was clinical and experimental data for known prognostic factors. All samples had

152   molecular subtype classification (Talhouk et al. submitted), residual disease was known for 1,771 cases,

153   primary treatment for 687, germline *BRCA* mutation status for 904, and nuclear CD8 TIL counts[8] for 1,111

154   (Supplemental Table S11). When examined by quintile of gene expression score there were differences, as

155   expected, for each of the known prognostic factors, including age and stage that were included in the model

156   (Table 3). However, in sensitivity analyses, applying the signature to specific patient groups, a robustness of

157   stratification was demonstrated, suggesting that the prognostic power of the signature is not explained by the

158   individual factors, residual disease, treatment, *BRCA* status, or CD8 score (Figure 3, Supplemental Figures S5-

159   S7). The signature score showed modest differences by molecular subtype (Supplemental Figure S8), and

160   adjusting for molecular subtype in the Cox analysis resulted in only minor changes to the HR estimates for

161   signature quintiles (Table 2). The signature was shown to be prognostic within a homogenous group of 316

162   stage 3C cases with no residual disease, within early stage cases (FIGO 1a and 1b), and within patients whose

163   samples were collected from the omentum (Supplemental Figures S9-S10). Analysis of the signature score for

164   paired ovary and omental tissue from 42 of the cases showed a highly significant Pearson correlation

165   coefficient, r = 0.79 (p = 5.4 x 10^{-10}) (Supplemental Figure S11).

166   A geneset enrichment analysis was performed for the 101 genes in the signature, as well as for genes

167   correlated with signature genes achieving r2 > 0.75 (Supplemental Table S12). For the correlated gene analysis,

168   the three most significant pathways involved the immune system, including the adaptive immune system and

169   cytokine signalling. A further ten immune pathways were significantly enriched and included interferon

170   signalling, innate immune system, and TCR signalling and antigen presentation pathways. Restricting to the

171   signature genes only, there was also enrichment in the immune system, but the top two pathways were PI-3K

172   cascade and GPCR ligand binding. Four other pathways were related to the cell cycle and mitosis, with the

15

173   remaining enriched for fibroblast growth factor receptor (FGFR) and epidermal growth factor receptor (ERRB)

174   signalling, and one pathway related to homologous combination repair.

175   **Discussion**

176   In a large-scale study of HGSOC patients, we identified a 101 gene expression signature able to predict

177   clinically relevant differences in OS. Using methods that are both economical and applicable to standard

178   clinical sampling techniques, we showed that the signature performs substantially better than age and stage

179   alone for prognosis of both two- and five-year OS. The number of patients and samples included in this study

180   is an order of magnitude greater than previous comparable studies of gene expression and OS in HGSOC

181   patients[12, 20, 21]. Thus, we have been able to more precisely quantify the prognostic value of gene

182   expression.

183   We report definitive associations between OS and expression of 276 genes. Of the five most significant genes

184   (*TAP1*, *ZFHX4*, *CXCL9*, *FBN1*, and *PTGER3*), four have been previously reported to be associated with survival in

185   HGSOC. The top prognostic gene, *TAP1*, is involved in the antigen presenting pathway. Expression was reduced

186   in metastatic HGSOC, positively associated with OS[22] as observed here, and linked to tumour regression in

187   response to treatment[23]. Also, hypomethylation of *TAP1* was associated with improved time to disease

188   recurrence[24]. *CXCL9* is a chemokine that mediates the recruitment of T-cells to solid tumours[25]. High

189   expression of intratumoural *CXCL9* was associated with higher OS[26] and higher lymphocytic infiltration,

190   which is also a robust prognostic factor in HGSOC[8, 11, 27] and a feature of the immunoreactive HGSOC

191   molecular subtype[11]. *CXCL9* has also been proposed as a therapeutic target due to evidence that it inhibits

192   angiogenesis and promotes antitumour adaptive immunity[28-30]. Strikingly, the signature was able to further

193   refine prognostic groups within patients with high TIL counts suggesting that *CXCL9* and *TAP1* expression may

194   be strong indicators of immune competency in HGSOC.

195   *FBN1* is an extracellular matrix (ECM) protein previously found to be a biomarker associated with early

196   recurrence in ovarian cancer patients who are initially sensitive to chemotherapy[31] and strongly correlated

197   with desmoplasia in HGSOC. The prostaglandin E2 receptor *PTGER3* is expressed in ovarian tumour cells and is

198   associated with relapse-free survival[32]. In contrast, *ZFHX4* does not have previous associations with HGSOC.

199   Associations between the expression of specific genes in tumour tissues and OS in HGSOC patients may

200   suggest new drug targets and lead to insights into biological variation in treatment response. For example,

201   cases in the Q5 quintile with the poorest outcome had increased expression of *IGF2*, *FGFR1*, and *MYC*, a

202   possible argument for the use of *IGFR1*, *FGFR*, Bromodomain (*MYC*), or a combination of PARP and CDK4/6

203   inhibitors (*MYC*) [33]. More immediately, the signature may help clinicians identify patients most in need of

204   intervention, such as patients that could potentially benefit from neo-adjuvant chemotherapy (NACT).

205   Alternatively, in clinical trials it could be used to stratify randomization by patients' risk, thereby reducing

206   heterogeneity within subgroups and increasing heterogeneity between subgroups. The signature will be

207   incorporated into future prospective clinical trials to determine if it can predict response to specific

208   treatments.

209   Measurement of the signature required standard FFPE tissue used in routine histopathology. Also, data

210   preprocessing and normalization were conducted on an individual level, thus translatable to a general patient

211   population. That is, 5-year OS prognosis of future patients can be evaluated against the patient population

212   reported here by i) following the same steps described here for generating the normalized gene expression

213   data, 2) computing an individual signature score, and 3) assigning an HR based on the score or comparing it to

214   the reported quintiles (Supplemental Material). NanoString gene expression is highly reproducible as seen by

215   our quality control metrics (Supplemental Material) and the FDA approval of the ProSigna test for breast

216   cancer.

217   The question of heterogeneity by ancestry or ethnicity was beyond the scope of this study but should be

218   pursued in future research. Another important question is whether molecular subtype can improve biomarker

219   performance. A substantial proportion of signature genes were identified by the subtype adjusted meta-

220   analysis, suggesting that the strong performance of the signature is not solely attributable to differences

221   among molecular subtypes. Additionally, all of the individual genes used in the molecular subtype classification

222   were included in development of the signature.

223   Although the cases received chemotherapy, the FFPE samples used in this study were chemo-naïve, as few

224   patients had NACT during the calendar period in which these samples were collected. Because the signature

225   appears to be prognostic in omentum samples, future studies may assess the value in NACT patients, using

226   pre-treatment omental biopsies or post-treatment tumour samples. Future work will also address if the

227   signature can predict platinum-refractory patients.

228   We have developed a robust prognostic signature for HGSOC that can be used to stratify patients and identify

229   those in need of alternative treatments. Gene set enrichment analysis applied to the signature indicates an

230   important role for the immune system in overall survival and supports further investigation of immune-therapy

231   in ovarian cancer. More generally, the identification here of high-confidence prognostic genes may lead to new

232   hypotheses for targeted treatments.

233

234   **Acknowledgements**

239   Norris Comprehensive Cancer Center. The AOV study recognizes the valuable contributions from Mie Konno,

240   Shuhong Liu, Michelle Darago, Faye Chambers and the staff at the Tom Baker Cancer Centre Translational

241   Laboratories. We thank Olivier Tredan and Pierre Heudel as investigators on the TRIO14 study and Sandrine

242   Berge-Montamat as assistant for clinical research. The Australian Ovarian Cancer Study gratefully

243   acknowledges additional support from Ovarian Cancer Australia and the Peter MacCallum Foundation. The

244   AOCS also acknowledges the cooperation of the participating institutions in Australia and acknowledges the

245   contribution of the study nurses, research assistants and all clinical and scientific collaborators to the study.

246   The complete AOCS Study Group can be found at www.aocstudy.org. We would like to thank all of the women

247   who participated in these research programs.

248   **FUNDING**

282 **Disclosure**

283 Beth Y. Karlan served on Invitae Corporation's Advisory Board from 2017 to 2018. Iain McNeish has acted on

284 Advisory Boards for AstraZeneca, Clovis Oncology, Tesaro, Carrick Therapeutics and Takeda. His institution

285 receives funding from AstraZeneca. Ros Glasspool in on the Advisory Boards for AstraZeneca, Tesaro, Clovis

286 and Immunogen and does consultancy work for SOTIO. She has received support to attend conferences from

References

1.   Vaughan S, Coward JI, Bast RC, Jr. et al. Rethinking ovarian cancer: recommendations for improving outcomes. Nat Rev Cancer 2011; 11: 719-725.

2.   Torre LA, Trabert B, DeSantis CE et al. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018; 68: 284-296.

3.   Bowtell DD, Bohm S, Ahmed AA et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer 2015; 15: 668-679.

4.   du Bois A, Reuss A, Pujade-Lauraine E et al. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). Cancer 2009; 115: 1234-1244.

5.   Bolton KL, Chenevix-Trench G, Goh C et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA 2012; 307: 382-390.

6.   Candido-dos-Reis FJ, Song H, Goode EL et al. Germline mutation in BRCA1 or BRCA2 and ten-year survival for women diagnosed with epithelial ovarian cancer. Clin Cancer Res 2015; 21: 652-657.

7.   Goode EL, Block MS, Kalli KR et al. Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. JAMA Oncol 2017; 3: e173290.

8.   Zhang L, Conejo-Garcia JR, Katsaros D et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003; 348: 203-213.

9.   Pujade-Lauraine E, Ledermann JA, Selle F et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol 2017; 18: 1274-1284.

10.   Moore K, Colombo N, Scambia G et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N Engl J Med 2018; 379: 2495-2505.

11.   Tothill RW, Tinker AV, George J et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 2008; 14: 5198-5208.

12.   Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature 2011; 474: 609-615.

13.   Bonome T, Levine DA, Shih J et al. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. Cancer Res 2008; 68: 5478-5486.

330   14.   Karlan BY, Dering J, Walsh C et al. POSTN/TGFBI-associated stromal signature predicts
331   poor prognosis in serous epithelial ovarian cancer. Gynecol Oncol 2014; 132: 334-342.

332   15.   Konecny GE, Haluska P, Janicke F et al. A phase II, multicenter, randomized, double-blind,
333   placebo-controlled trial of ganitumab or placebo in combination with carboplatin/paclitaxel as
334   front-line therapy for optimally debulked primary ovarian cancer: The TRIO14 trial. Journal of
335   Clinical Oncology 2014; 32: 5529.

336   16.   Konecny GE, Wang C, Hamidi H et al. Prognostic and therapeutic relevance of molecular
337   subtypes in high-grade serous ovarian cancer. J Natl Cancer Inst 2014; 106.

338   17.   Millstein J, Volfson D. Computationally efficient permutation-based confidence interval
339   estimation for tail-area FDR. Front Genet 2013; 4: 179.

340   18.   Talhouk A, Kommoss S, Mackenzie R et al. Single-Patient Molecular Testing with
341   NanoString nCounter Data Using a Reference-Based Strategy for Batch Effect Correction. PLoS
342   One 2016; 11: e0153844.

343   19.   Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and
344   hybridization array data repository. Nucleic Acids Res 2002; 30: 207-210.

345   20.   Jin C, Xue Y, Li Y et al. A 2-Protein Signature Predicting Clinical Outcome in High-Grade
346   Serous Ovarian Cancer. Int J Gynecol Cancer 2018; 28: 51-58.

347   21.   Mankoo PK, Shen R, Schultz N et al. Time to recurrence and survival in serous ovarian
348   tumors predicted from integrated genomic profiles. PLoS One 2011; 6: e24709.

349   22.   Nymoen DA, Hetland Falkenthal TE, Holth A et al. Expression and clinical role of
350   chemoresponse-associated genes in ovarian serous carcinoma. Gynecol Oncol 2015; 139: 30-39.

351   23.   Jimenez-Sanchez A, Memon D, Pourpe S et al. Heterogeneous Tumor-Immune
352   Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. Cell
353   2017; 170: 927-938.e920.

354   24.   Wang C, Cicek MS, Charbonneau B et al. Tumor hypomethylation at 6p21.3 associates with
355   longer time to recurrence of high-grade serous epithelial ovarian cancer. Cancer Res 2014; 74:
356   3084-3091.

357   25.   Gorbachev AV, Kobayashi H, Kudo D et al. CXC chemokine ligand 9/monokine induced by
358   IFN-gamma production by tumor cells is critical for T cell-mediated suppression of cutaneous
359   tumors. J Immunol 2007; 178: 2278-2286.

360   26.   Bronger H, Singer J, Windmuller C et al. CXCL9 and CXCL10 predict survival and are
361   regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. Br J Cancer 2016;
362   115: 553-563.

363   27.   Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in
364   High-Grade Serous Ovarian Cancer. JAMA Oncol 2017; 3: e173290.

365   28.   Tokunaga R, Zhang W, Naseem M et al. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune
366   activation - A target for novel cancer therapy. Cancer Treat Rev 2018; 63: 40-47.

367   29.   Xiao P, Guo Y, Zhang H et al. Myeloid-restricted ablation of Shp2 restrains melanoma
368   growth by amplifying the reciprocal promotion of CXCL9 and IFN-gamma production in tumor
369   microenvironment. Oncogene 2018.

370   30.   Zhang R, Tian L, Chen LJ et al. Combination of MIG (CXCL9) chemokine gene therapy with
371   low-dose cisplatin improves therapeutic efficacy against murine carcinoma. Gene Ther 2006; 13:
372   1263-1271.

373   31.   Zhang W, Ota T, Shridhar V et al. Network-based survival analysis reveals subnetwork
374   signatures for predicting outcomes of ovarian cancer treatment. PLoS Comput Biol 2013; 9:
375   e1002975.

376   32.   Reinartz S, Finkernagel F, Adhikary T et al. A transcriptome-based global map of signaling
377   pathways in the ovarian cancer microenvironment associated with clinical outcome. Genome
378   Biol 2016; 17: 108.

379   33.   Konecny GE. Combining PARP and CDK4/6 inhibitors in MYC driven ovarian cancer.
380   EBioMedicine 2019; 43:9-10.
381

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

426

427

[1]Peter MacCallum Cancer Centre, Melbourne, Victoria, 3000, Australia.

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

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

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

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

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

[7]Centre 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.

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

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

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

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

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

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

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

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

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

[17]Women and Infant's Research Foundation, King Edward Memorial Hospital, 374 Bagot Road, Subiaco, Western Australia, 6008, Australia

453   [18]St John of God Hospital, 12 Salvado Rd, Subiaco, Western Australia, 6008, Australia

454   [19]Canberra Hospital, Yamba Drive, Garran, Australian Capitol Territory, 2605, Australia

455   [20]Bankstown Cancer Centre, Bankstown Hospital, 70 Eldridge Road, Bankstown, New South
456   Wales, 2200, Australia

457   [21]Northern Haematology & Oncology Group, Integrated Cancer Centre, 185 Fox Valley Road,
458   Wahroonga, New South Wales, 2076, Australia

459   [22]Illawarra Shoalhaven Local Health District, Wollongong Hospital, Level 4 Lawson House,
460   Wollongong, New South Wales, 2500, Australia

461   [23]Nepean Hospital, Derby Street, Kingswood, New South Wales, 2747, Australia

462   [24]Newcastle Mater Misericordiae Hospital, Edith Street, Waratah, New South Wales, 2298, Australia

463   [25]Port Macquarie Base Hospital, Wrights Road, Port Macquarie, New South Wales, 2444, Australia

464   [26]Prince of Wales Clinical School, University of New South Wales, New South Wales, 2031, Australia

465   [27]St George Hospital, Gray Street, Kogarah, New South Wales, 2217, Australia

466   [28]St Vincent's Hospital, 390 Victoria Street, Darlinghurst, New South Wales, 2010, Australia

467   [29]Wagga Wagga Base Hospital, Docker St, Wagga Wagga, New South Wales, 2650, Australia

468   [30]Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, Sydney, New South Wales,
469   2145, Australia.

470   [31]Department of Pathology, Westmead Clinical School, Westmead Hospital, The University of Sydney,
471   New South Wales, 2006, Australia

472   [32]Mater Misericordiae Hospital, Raymond Terrace, South Brisbane, Queensland, 4101, Australia

473   [33]The Royal Brisbane and Women's Hospital, Butterfield Street, Herston, Queensland, 4006, Australia

474   [34] Wesley Hospital, 451 Coronation Drive, Auchenflower, Queensland, 4066, Australia

475   [35]Burnside Hospital, 120 Kensington Road, Toorak Gardens, South Australia, 5065, Australia

476   [36]Flinders Medical Centre, Flinders Drive, Bedford Park, South Australia, 5042, Australia

477   [37]Queen Elizabeth Hospital, 28 Woodville Road, Woodville South, South Australia, 5011, Australia

478   [38]Freemasons Hospital, 20 Victoria Parade, East Melbourne, Victoria, 3002, Australia

479 [39]Mercy Hospital for Women, 163 Studley Road, Heidelberg, Victoria, 3084, Australia

480 [40]The Royal Women's Hospital, Parkville, Victoria, 3052, Australia

481 [41]Border Medical Oncology, Wodonga, Victoria, 3690, Australia

482 [42]Andrew Love Cancer Centre, 70 Swanston Street, Geelong, Victoria, 3220, Australia

483 [43]Ballarat Base Hospital, Drummond Street North, Ballarat, Victoria, 3350, Australia

484 [44]Bendigo Health Care Group, 62 Lucan Street, Bendigo, Victoria, 3550, Australia

485 [45]Peninsula Health, 2 Hastings Road, Frankston, Victoria, 3199, Australia

486 [46]Mount Hospital, 150 Mounts Bay Road, Perth, Western Australia 6000, Australia

487 [47]Faculty of Medicine, Bar-Ilan University, 8 Henrietta Szold St, Safed, Israel

488

489 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.

Gene	HR (95% CI)	P	Selection	Correlated gene*	r_s
TAP1	0.84 (0.80, 0.87)	8.3×10^{-18}	Meta	PSMB9	0.89
ZFHX4	1.19 (1.14, 1.25)	1.4×10^{-15}	Meta	LOC100192378	0.74
CXCL9	0.85 (0.82, 0.88)	1.8×10^{-15}	Meta and candidate	CXCR6	0.89
FBN1	1.18 (1.13, 1.24)	4.2×10^{-14}	Candidate	SPARC^	0.91
PTGER3	1.18 (1.13, 1.24)	1.2×10^{-13}	Meta	COL8A1	0.67

*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.

1

Table 2. Hazard ratios and 95% CIs for quintiles of the gene expression signature score in validation data.

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

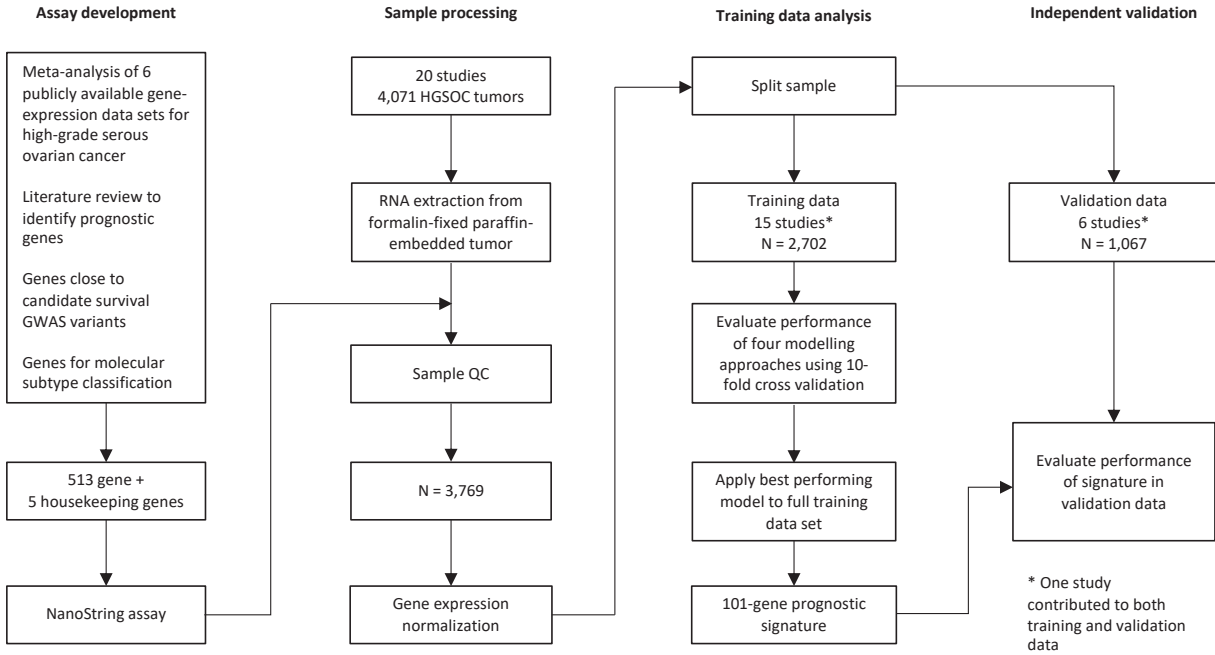*Median survival (95% CI) in years for patients in the validation set.

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.
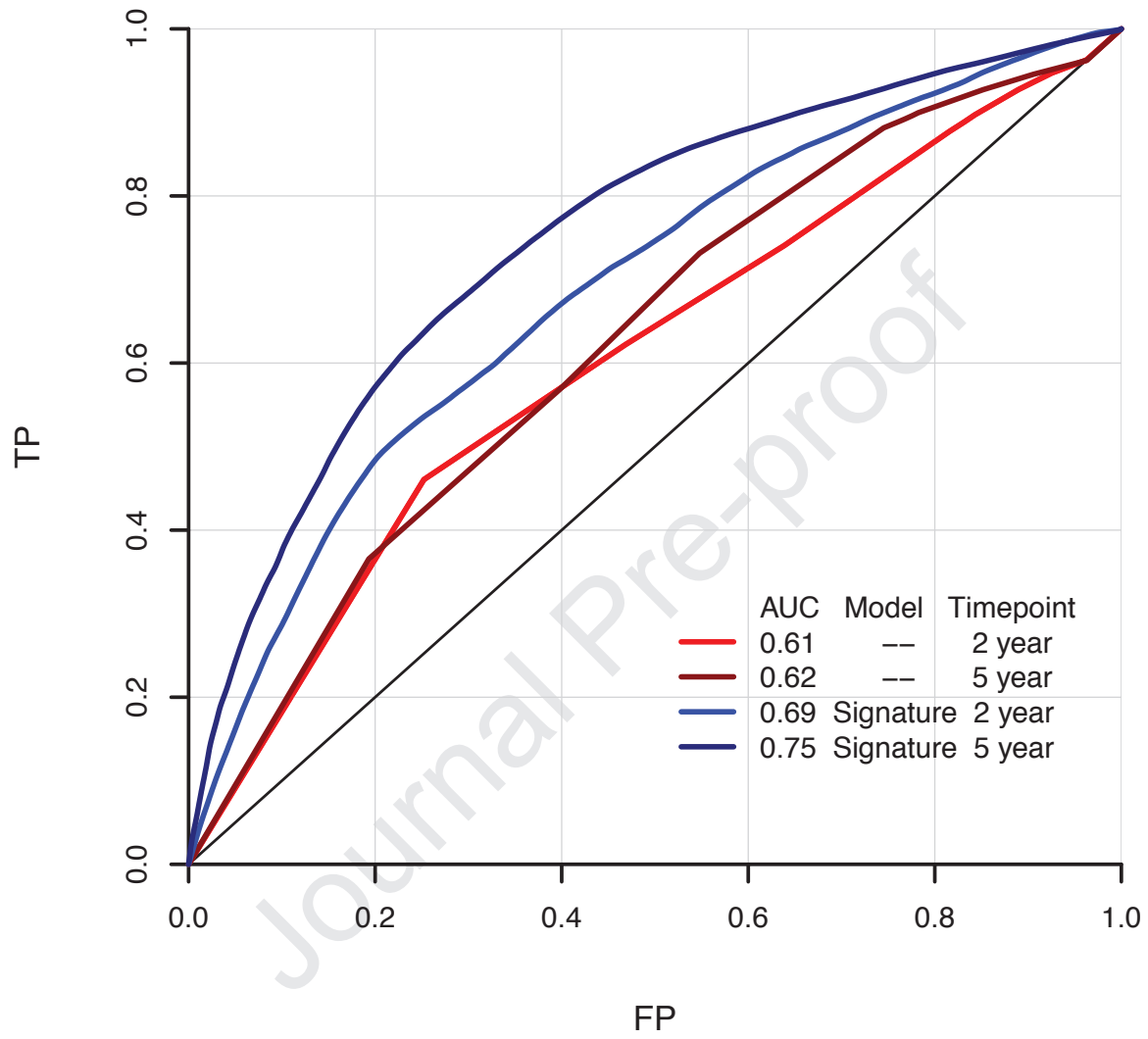
	Total	Q1	Q2	Q3	Q4	Q5	p-value
N	3769	754	754	753	754	754	
median survival (years)	4.1	9.5	5.4	3.8	3.2	2.3	
% 5-year survival	41	75	57	39	25	10	
Age median	63	58	57	61	64	70	
Age range	25-89	39-78	25-86	36-82	27-89	39-86	
Age quartile q1	894	30.8	31.3	20.0	13.4	4.5	$<1\times10^{-50}$
Age quartile q2	838	21.5	20.0	22.9	21.2	14.3	
Age quartile q3	961	16.0	20.2	21.4	23.6	18.7	
Age quartile q4	1076	13.5	10.4	16.4	21.3	38.5	
FIGO stage I / II	607	97.4	2.6	0.0	0.0	0.0	$<1\times10^{-50}$
FIGO stage III/IV	3067	3.8	23.0	24.1	24.4	24.6	
Primary chemo* 1	136	16.2	22.1	23.5	19.1	19.1	0.163
Primary chemo* 2	190	16.3	20.0	21.6	22.1	20.0	
Primary chemo* 3	361	11.1	16.9	22.4	20.5	29.1	
Residual disease No	614	32.4	22.1	17.8	15.5	12.2	$<1\times10^{-50}$
Residual disease Yes	1157	6.0	19.2	24.1	24.5	26.2	
germline BRCA1 mutation	130	23.8	31.5	26.2	11.5	6.9	2.22×10^{-7}
germline BRCA2 mutation	71	28.2	26.8	18.3	18.3	8.5	
germline no mutation	663	19.6	16.7	18.7	20.7	24.3	
CD8 TIL score 0	192	19.8	14.6	12.5	21.4	31.8	2.46×10^{-14}
CD8 TIL score 1-2	186	18.3	14.0	18.8	21.5	27.4	
CD8 TIL score 3-19	515	19.8	24.1	20.8	17.9	17.5	
CD8 TIL score >20	218	34.4	31.2	16.5	11.5	6.4	
Molecular subtype C1.MES	1105	5.4	10.4	20.7	27.4	36.0	$<1\times10^{-50}$
Molecular subtype C2.IMM	907	23.2	28.8	21.2	16.2	10.7	
Molecular subtype C4.DIF	1144	32.6	25.5	17.9	12.8	11.2	
Molecular subtype C5.PRO	613	18.1	14.0	20.7	25.8	21.4	
FIGO stage 1A & 1B	111	96.4	3.6	0.0	0.0	0.0	$<1\times10^{-50}$
FIGO stage 3C	1979	3.1	23.7	24.6	24.1	24.6	$<1\times10^{-50}$
FIGO stage 3C Residual	316	6.3	31.0	24.4	20.9	17.4	6.24×10^{-45}
FIGO stage 3C Residual	846	2.6	21.5	25.3	24.6	26.0	

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.

Assay development
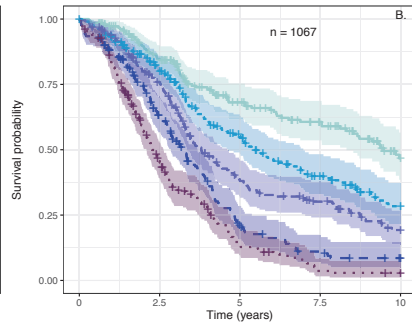
Meta-analysis of 6 publicly available gene-expression data sets for high-grade serous ovarian cancer

Literature review to identify prognostic genes

Genes close to candidate survival GWAS variants

Genes for molecular subtype classification

↓

513 gene + 5 housekeeping genes

↓

NanoString assay

Sample processing

20 studies 4,071 HGSOC tumors

↓

RNA extraction from formalin-fixed paraffin-embedded tumor

↓

Sample QC

↓

N = 3,769

↓

Gene expression normalization

Training data analysis

Split sample

↓

Training data 15 studies* N = 2,702

↓

Evaluate performance of four modelling approaches using 10-fold cross validation

↓

Apply best performing model to full training data set

↓

101-gene prognostic signature

Independent validation

Validation data 6 studies* N = 1,067

↓

Evaluate performance of signature in validation data

* One study contributed to both training and validation data

Number at risk (A.)

541	459	339	221	113
540	412	249	127	65
540	365	171	66	31
540	296	115	42	10
541	202	43	17	10

Number at risk (B.)

214	161	99	72	33
213	150	87	50	26
213	146	64	44	22
213	113	31	14	7
214	85	22	5	3
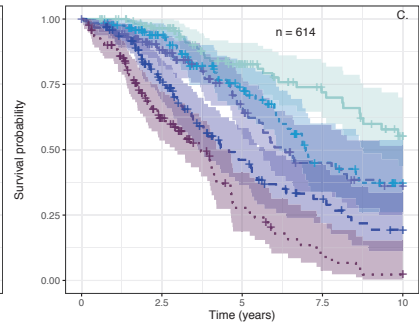
Number at risk (C.)

123	100	56	36	20
123	93	49	19	9
122	77	39	21	14
123	74	31	15	7
123	53	15	4	1
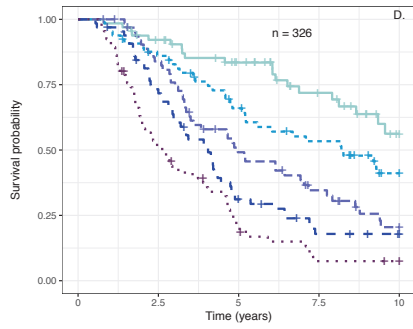
Number at risk (D.)

65	56	43	28	13
65	54	38	29	16
65	52	28	17	8
65	45	18	9	6
66	33	13	4	3
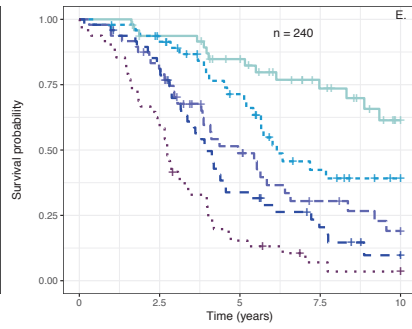
Number at risk (E.)

48	43	35	22	10
48	41	27	13	7
48	38	18	9	5
48	38	15	7	2
48	29	7	2	1
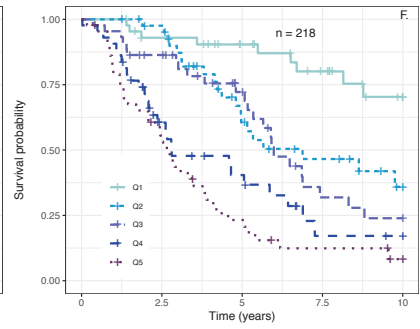
Number at risk (F.)

44	38	29	21	11
43	39	19	12	5
44	35	22	8	6
43	19	11	3	2
44	22	9	4	1