

Platform-Independent Classification System to Predict Molecular Subtypes of High-Grade Serous Ovarian Carcinoma

Arunima Shilpi, PhD¹; Manoj Kandpal, PhD¹; Yanrong Ji¹; Brandon L. Seagle, MD, MS¹; Shohreh Shahabi, MD¹; and Ramana V. Davuluri, PhD¹

PURPOSE Molecular cancer subtyping is an important tool in predicting prognosis and developing novel precision medicine approaches. We developed a novel platform-independent gene expression–based classification system for molecular subtyping of patients with high-grade serous ovarian carcinoma (HGSOC).

METHODS Unprocessed exon array (569 tumor and nine normal) and RNA sequencing (RNA-seq; 376 tumor) HGSOC data sets, with clinical annotations, were downloaded from the Genomic Data Commons portal. Sample clustering was performed by non-negative matrix factorization by using isoform-level expression estimates. The association between the subtypes and overall survival was evaluated by Cox proportional hazards regression model after adjusting for the covariates. A novel classification system was developed for HGSOC molecular subtyping. Robustness and generalizability of the gene signatures were validated using independent microarray and RNA-seq data sets.

RESULTS Sample clustering recaptured the four known The Cancer Genome Atlas molecular subtypes but switched the subtype for 22% of the cases, which resulted in significant ($P = .006$) survival differences among the refined subgroups. After adjusting for covariate effects, the mesenchymal subgroup was found to be at an increased hazard for death compared with the immunoreactive subgroup. Both gene- and isoform-level signatures achieved more than 92% prediction accuracy when tested on independent samples profiled on the exon array platform. When the classifier was applied to RNA-seq data, the subtyping calls agreed with the predictions made from exon array data for 95% of the 279 samples profiled by both platforms.

CONCLUSION Isoform-level expression analysis successfully stratifies patients with HGSOC into groups with differing prognosis and has led to the development of robust, platform-independent gene signatures for HGSOC molecular subtyping. The association of the refined The Cancer Genome Atlas HGSOC subtypes with overall survival, independent of covariates, enhances the clinical annotation of the HGSOC cohort.

Clin Cancer Inform. © 2019 by American Society of Clinical Oncology

INTRODUCTION

High-grade serous ovarian carcinoma (HGSOC) accounts for 70% to 80% of ovarian cancer deaths, with little improvement in overall survival (OS) in recent years.¹ The standard therapy for HGSOC includes maximal cytoreductive surgery followed by platinum and taxane chemotherapy. Although the majority of patients with HGSOC respond to initial treatment, most tumors recur and become increasingly resistant to chemotherapy, with a 5-year OS rate of approximately 30%.² As a heterogeneous disease, HGSOC molecular subtyping can serve as a useful clinical tool to predict response to therapy and inspire novel personalized medicine treatment plans. Indeed, genomic and transcriptome profiling by The Cancer Genome Atlas (TCGA) consortium and others revealed few recurrent somatic mutations but a highly complex genomic terrain marked by copy number alterations and intertumor

heterogeneity.³ Four molecular subtypes—mesenchymal (M), immunoreactive (I), differentiated (D), and proliferative (P)—were independently identified, first by the Australian Ovarian Cancer Study (AOCS)⁴ and then by the TCGA consortium.³ However, TCGA subtypes did not show statistically significant survival differences. Therefore, an important question is whether the unsupervised clustering of samples on the basis of gene-level annotations is still the best approach to identify clinically relevant molecular subtypes.

The majority of human genes produce multiple functional products, or isoforms, through alternative transcription and alternative splicing.⁵⁻⁷ Different protein isoforms of a gene participate in different functional pathways,^{8,9} and cancer-associated aberrant alternative splicing events have been reported.¹⁰⁻¹³ In ovarian cancer, specific splice variants have been identified as prognostic markers and predictors of resistance to

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on March 4, 2019 and published at ascopubs.org/journal/cci on April 19, 2019; DOI <https://doi.org/10.1200/CCI.18.00096>

CONTEXT

Key Objective

To develop a novel platform-independent molecular subtyping assay for high-grade serous ovarian carcinoma classification.

Knowledge Generated

Refined molecular subtyping on the basis of the isoform-level transcriptome of The Cancer Genome Atlas high-grade serous ovarian carcinoma cohort switched the subtype calls for 22% of the samples and led to improved prognostic stratification among the subtypes. The derived gene/isoform-level signatures are platform independent such that more than 90% of the patients were correctly classified into one of the four molecular subtypes, irrespective of whether the gene expression data were generated by either a microarray or a next-generation sequencing platform.

Relevance

We anticipate that the molecular subtyping assay can be used in a next-generation sequencing clinical diagnostics laboratory in combination with pathologic information, advanced imaging analytics, and radiomics approaches to enable improved precision diagnostics and treatment planning.

therapy, such as p53 δ of *TP53* and CD44v8-10 of *CD44* as prognostic markers^{14,15}; *MRP1* splice variant resistance to doxorubicin¹⁶; the role of *EVI1* transcript variant (*EVI1*Del190-515) in tumorigenesis¹⁷; and the role of osteopontin-c isoform in active proliferation, migration, and tumor growth.¹⁸ Therefore, specific transcript variants could be more effective as diagnostic and prognostic markers than corresponding genes^{19,20} and suggest that biomarker and molecular subtyping studies should explore the isoform-level transcriptome. We hypothesized that isoform-based clustering of HGSOC tumors will lead to more clinically relevant subgrouping than gene-level subgrouping. Moreover, isoform-level gene classifiers can identify specific isoforms as biomarkers and generate robust and clinically translatable assays for HGSOC stratification. We have adopted our previously developed classification system, PIGExClass (platform-independent isoform-level gene-expression based classification-system) to robustly cluster HGSOC tumors on the basis of isoform-level transcriptome profiles, develop platform-independent classification models for HGSOC molecular subtyping, and validate the classifiers on independent data sets from different platforms.

METHODS

Preprocessing of TCGA HGSOC Exon Array Data

TCGA unprocessed exon array data for 569 HGSOC and eight normal samples were downloaded from the Genomic Data Commons portal.²¹ Gene-level and isoform-level expression estimates were obtained using multimapping Bayesian gene expression for whole-transcript arrays,²² using Ensembl database version 56 as the reference genome. Expression estimates were normalized across the samples using the LOWESS algorithm.²³

Data Filtration

Two-step filtration was applied to obtain highly variable isoforms for clustering. The first filter retains only one isoform among highly correlated isoforms of the same gene. Two

isoforms of a gene are considered highly correlated if the Pearson's correlation coefficient of isoform-level expressions across the samples is higher than 0.8. The isoform with the highest coefficient of variation was retained among the correlated isoforms of a gene. The second filter selects, using coefficient of variation, the isoforms that are most variable across patients.

Identification of HGSOC Subtypes on the Basis of Isoform-Level Expression

Unsupervised non-negative matrix factorization consensus clustering was applied using the NMF package of R.²⁴⁻²⁶ Consensus matrices with different factorization ranks (2 to 7) were obtained by taking the average of 50 connectivities. Clustering quality was evaluated using the cophenetic correlation coefficient and heat map plots (Data Supplement). The factorization was repeated for 100 runs, and the one with the lowest approximation error was retained. Samples that were not true representatives of the subclasses were filtered out by silhouette width procedure.²⁷

Identification of OS Differences Among the Subtypes

Log-rank test was applied to determine the prognostic relevance of the four subtypes. Kaplan-Meier survival curves were plotted using the R function *survival*.²⁸ Pairwise comparison of survival among subtypes was adjusted for multiple comparisons using the Benjamini-Hochberg

TABLE 1. Number of Up- and Downregulated Genes or Transcript Variants Identified in the TCGA Ovarian Cohort Exon Array Data

Differential Expression	Level, No.	
	Genes	Isoform (transcript variant)
Upregulated	639	1,841
Downregulated	995	2,882
Total	1,634	4,723

NOTE. $q \leq .01$; fold-change ≥ 2.0 .

Abbreviation: TCGA, The Cancer Genome Atlas.

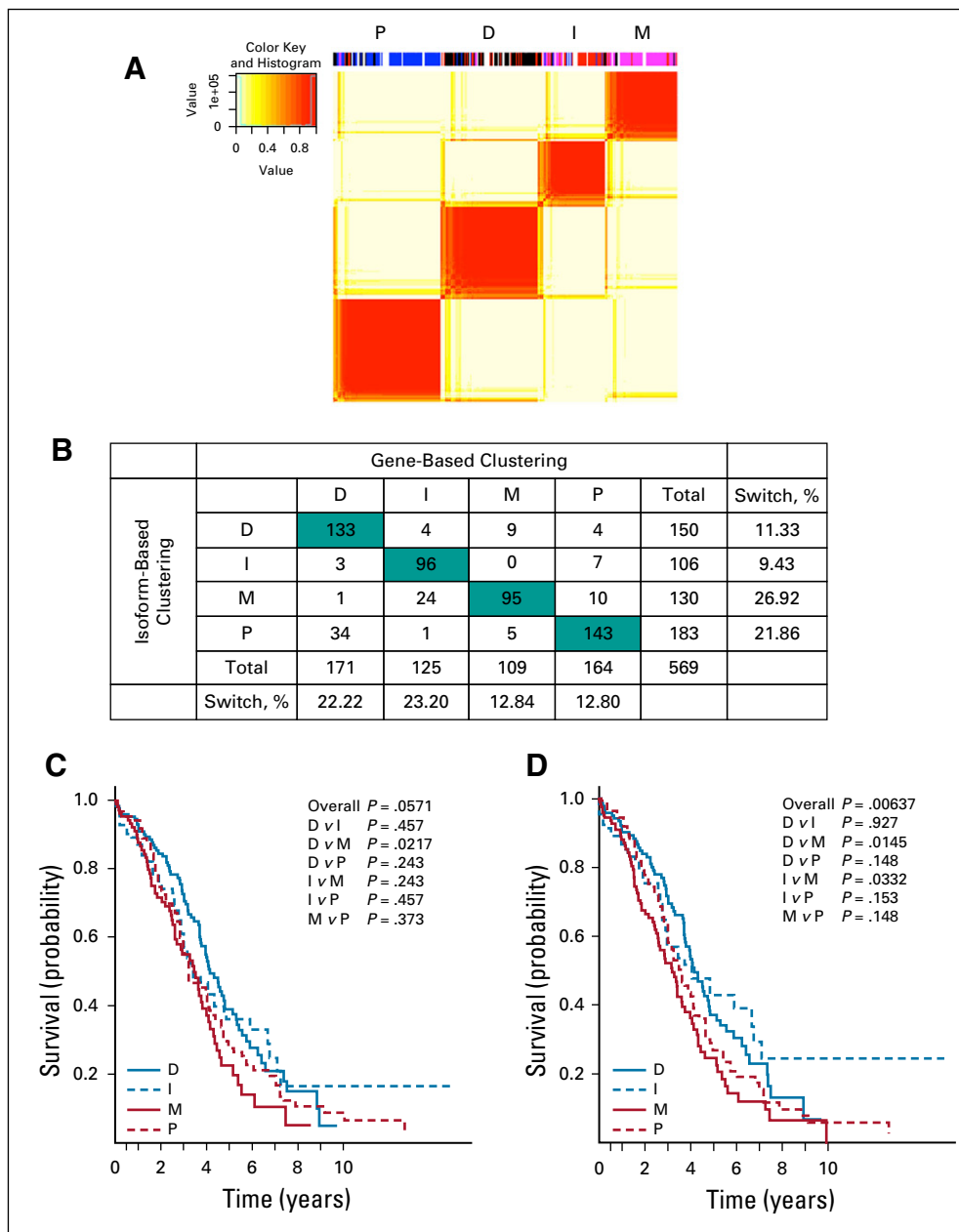


FIG 1. (A) Unsupervised non-negative matrix factorization clustering of 569 patients with high-grade serous ovarian carcinoma (HGSOC) on the basis of highly variable (930) isoform-level signatures. The clusters were identified and assigned into four subgroups on the basis of original The Cancer Genome Atlas core grouping. The color code for samples in each cluster are as follows: differentiated (D), black; immunoreactive (I), red; mesenchymal (M), pink; and proliferative (P), blue. (B) The concordance table shows the agreement of sample assignment with The Cancer Genome Atlas subgroups (gene-based subtypes) and our isoform-based subgroups. Although the P subgroup showed good agreement, the I subgroup showed the worst agreement followed by the M subgroup. (C and D) Kaplan-Meier survival curves plotted to determine the prognostic difference among four the subtypes identified using gene- and isoform-level expression-based clustering, respectively. The statistical significance in overall survival was determined at a threshold $P = .05$.

procedure.²⁹ Imputation of missing values was performed by the R package mice.³⁰ The association of the subgroups with OS was modeled by fitting a multivariable Cox proportional hazards (PH) regression model adjusted for age, tumor stage, cytoreduction, and chemotherapy.

Processing of the RNA Sequencing Data Set

TCGA RNA sequencing (RNA-seq) data for 376 HGSOC samples were downloaded from the Genomic Data Commons portal.²¹ A subset of 279 samples were profiled by both exon array and RNA-seq platforms. For control samples, RNA-seq

data for six normal fallopian-tissue samples were downloaded from the Genotype-Tissue Expression database. RNA-seq data were analyzed using Picard tools³¹ and the RNA-Seq by Expectation-Maximization³² program.

Variable Selection and Building Classification Model

We adopted the PIGExClass algorithm, which combines data discretization and random forest-based variable selection procedures, to build gene/isoform-level classifiers.³³ Equal-frequency binning data discretization was applied on fold-change values (cancer over normal).³⁴ Variable selection was performed using fold-change estimates to select a small set of nonredundant genes/isoforms that were used to build random forest-based classification models. Tenfold cross-validation was applied followed by training the models on 75% of the samples and testing on the remaining 25% of samples.

Validation of Classification Model on Independent Data Sets

We evaluated the gene-level classification model on data from two independent platforms. The misclassification rate was computed for 279 RNA-seq samples that were profiled by both exon array and RNA-seq platforms. In addition, we built gene-level classification models on the TCGA exon array data set and validated the models using two independent microarray data sets—GSE9891⁴ and GSE26712³⁵—downloaded from the Gene Expression Omnibus database.³⁶ This study followed the recommendations in the Transparent Reporting of a Multivariable Prediction Model for Individual

Prognosis or Diagnosis statement (Data Supplement). Additional details about the validation methods are provided in the Data Supplement.

RESULTS

Numerous Genes Were Differentially Expressed at the Isoform Level Between HGSOC and Normal Samples

We obtained expression estimates for 35,612 gene-level and 114,930 transcript-level features by analyzing the TCGA exon array data set of 569 tumor and eight normal samples. Although only 1,634 genes were obtained as differentially expressed at the gene level, isoform-level analysis resulted in 4,723 transcript variants corresponding to 2,245 genes as differentially expressed between HGSOC and normal samples (fold-change ≥ 2 ; $q \leq .001$; Table 1). Because we found more than double the number of genes differentially expressed at the isoform level than at the gene level, we investigated whether the isoform-level transcriptome can provide a better molecular subgrouping of patients with HGSOC in terms of overall prognosis.

Unsupervised Clustering of TCGA HGSOC Samples Using Isoform-Level Transcriptome Recaptured the Molecular Subtypes With Improved Prognostic Stratification

The TCGA Research Network reported four gene-based molecular subtypes with no significant survival differences ($P = .117$; Data Supplement). Therefore, we performed clustering of 569 TCGA HGSOC samples using isoform-level expression estimates of 930 highly variable isoforms.

TABLE 2. Cox Proportional Hazards Regression Model by Gene- and Isoform-Based Subgroups

Variable	Gene-Based Subgroup		Isoform-Based Subgroup	
	HR (95% CI)	P	HR (95% CI)	P
Subtypes				
Mesenchymal	1			
Immunoreactive	0.81 (0.56 to 1.18)	.278	0.66 (0.45 to 0.97)	.034
Differentiated	0.85 (0.60 to 1.19)	.341	0.81 (0.58 to 1.14)	.234
Proliferative	0.82 (0.59 to 1.14)	.244	0.84 (0.62 to 1.13)	.247
Stage				
II	1			
III	1.87 (0.96 to 3.64)	.064	1.79 (0.92 to 3.48)	.086
IV	2.56 (1.26 to 5.20)	.009	2.41 (1.19 to 4.89)	.015
Age	1.02 (1.01 to 1.03)	.001	1.02 (1.01 to 1.03)	.001
Cytoreduction				
RD = 0 cm	1			
RD ≤ 1 cm	1.95 (1.29 to 2.93)	.001	1.96 (1.30 to 2.94)	.001
RD > 1 cm	1.81 (1.16 to 2.81)	.001	1.86 (1.19 to 2.90)	.006
Chemotherapy				
Intravenous	1			
Intraperitoneal	0.62 (0.38 to 1.00)	.049	0.61 (0.38 to 0.98)	.042

NOTE. Boldface indicates significance at $P < .05$.

Abbreviations: HR, hazard ratio; RD, residual disease.

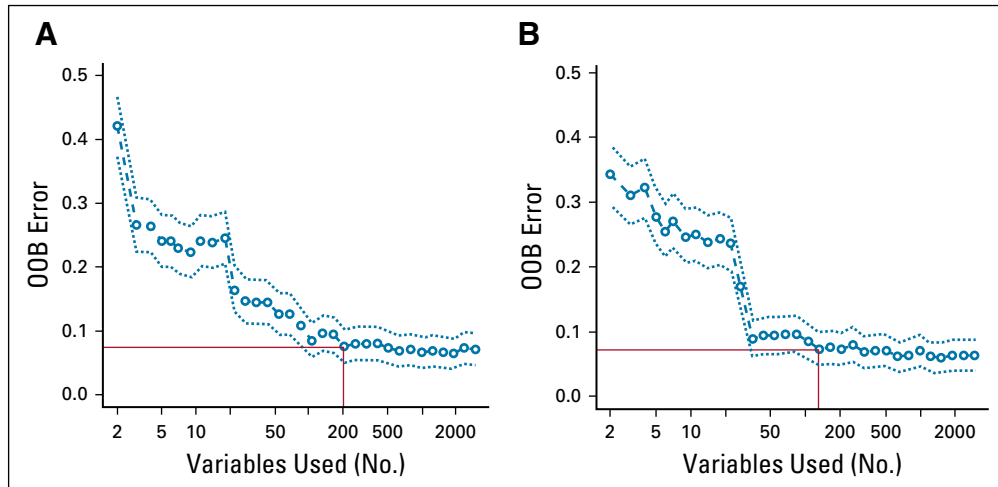


FIG 2. The classifier was generated on the basis of the selection subset of (A) isoform/transcript variants and (B) genes. Out-of-bag (OOB) error rate was plotted, where the x-axis denotes the selection of features/variables and the y-axis represents the error rate.

The samples clustered into four distinct subgroups, which largely overlapped with the TCGA subgroups and, as such, refined the TCGA subtypes. Therefore, we retained TCGA subgroup nomenclature of I, D, P, and M on the basis of the concordance in cluster membership calls between our isoform-based groups and the TCGA core sample subgrouping (Fig 1A).

To find homogenous clusters, we filtered out 134 samples with negative silhouette width values, which resulted in the final set of 435 samples (isoform-based core samples) grouped into 128 as I, 106 as D, 97 as M, and 104 as P. Among the 371 tumor samples that are common between TCGA and our isoform-based subgroups, on the basis of the concordance table (Fig 1B), 80 samples were clustered into a different subgroup by our isoform-based clustering. The switching of 22% of the samples into a different subgroup resulted in a statistically significant difference in OS among the four isoform-based subtypes ($P = .006$; Fig 1D), whereas, the gene-based subtypes, derived by using the gene-level expression profiles, did not show significant OS differences ($P = .057$; Fig 1C), similar to the TCGA subtypes ($P = .117$; Data Supplement). In pairwise comparisons among the four subtypes, after controlling for

false discovery rate, we found that isoform-based subtype pairs D and M ($P = .0145$) and I and M ($P = .0332$) were significantly different in OS (Fig 1D). However, none of the adjusted pairwise comparisons among TCGA subtypes were statistically significant at the $\alpha = .05$ level (Data Supplement). Similarly, for the gene-level subtyping of 569 samples (Fig 1C), although the OS P value was smaller compared with that of TCGA subgrouping (Data Supplement), it was still not significant at $\alpha = .05$. For pairwise comparisons, only one comparison (D v M) was found to be statistically significant. For the isoform-based subgroups, the median OS for the M subtype was 3.2 years (95% CI, 2.6 to 4 years); for the D subtype, 4.2 years (95% CI, 3.8 to 4.8 years); for the I subtype, 4 years (95% CI, 3.0 to 6.8 years); and for the P subtype, 3.6 years (95% CI, 3.2 to 4.1 years).

Multivariable Cox PH regression analysis revealed that isoform-based, but not gene-based, molecular subtypes have a significant association with OS after accounting for clinical variables age, tumor stage, residual tumor size after cytoreductive surgery, and chemotherapy (intravenous or intraperitoneal). More specifically, the I subtype hazard ratio was 0.66 (95% CI, 0.45 to 0.97; $P = .034$) compared with that of the M subtype and after accounting for the four

TABLE 3. Concordance Between the True-Class Labels and Predicted Calls by the Isoform-Level Classifier That Was Trained on the Exon Array Data Set and Applied on RNA-Seq Data Set

Predicted Class RNA-Seq	True Class				Total
	Differentiated	Immunoreactive	Mesenchymal	Proliferative	
Differentiated	71	2	4	2	79
Immunoreactive	0	64	0	1	65
Mesenchymal	2	0	63	0	65
Proliferative	2	0	2	66	70
Total	75	66	69	69	279

Abbreviation: RNA-seq, RNA sequencing.

clinical variables. However, although the D and P subtypes showed better OS compared with the M subtype, the Cox PH *P* values were not statistically significant at the $\alpha = .05$ level (Table 2). Although factors such as increased age at the time of diagnosis, tumor stage III or IV, and suboptimal debulking surgery (residual disease > 0 cm) are associated with an increased risk of death, intraperitoneal chemotherapy was associated with longer survival than intravenous chemotherapy (Table 2). In summary, both univariable and multivariable survival analyses demonstrated that the isoform-level subgrouping shows improved association with OS compared with the gene-based subgrouping.

Gene- and Isoform-Based Classification Models for HGSOC Subtype Prediction

Using the isoform-based subtypes as the four classes, we built both gene-based and isoform-based classifiers on the basis of the 435 isoform-based core samples. Although the gene-based classifier can be applied to data from either microarray or RNA-seq platforms, the isoform-based classifier is mainly for exon array and RNA-seq platforms. The random forest feature selection step selected 206 isoforms as feature variables that are most discriminating among the four subgroups (Fig 2A; Data Supplement). The isoform-based random forest classifier achieved 92% accuracy with 206 isoforms as feature variables when trained and tested by 10-fold cross-validation on the TCGA exon array data set. The classifier was further tested by dividing the isoform-based core samples into two groups: 75% of samples to be used as a training set and 25% to be used as a test set. The classification model generated from the training set was applied to the test set. The results of this additional testing agreed with those of the 10-fold cross-validation approach in 99% of the sample calls in the test set, which confirmed that the algorithm effectively distinguishes the four subgroups. Similarly, the gene-based classifier achieved 91% accuracy, with 132 genes as feature variables (Fig 2B; Data Supplement). Although the feature selection step of the random forest algorithm selected a different number of features for gene- and isoform-based

classifiers, both classifiers achieved more than 90% accuracy with as few as 100 feature variables (Fig 2).

Platform Transition of the Classification Models to RNA-Seq and Other Microarray Platforms

Because the gene- and isoform-based classifiers (both of which were derived on data from the exon array platform) achieved a prediction accuracy of more than 90%, we tested the robustness and generalizability of the classification models by testing on data from independent platforms, such as RNA-seq and other microarray platforms. First, we evaluated the transition of the isoform-based classifier from the exon array to the RNA-seq platform by applying 279 RNA-seq TCGA samples that overlapped with the isoform-level core samples and were profiled by both exon array and RNA-seq methods. Therefore, the class labels for these 279 samples are known from the isoform-level clustering. By comparing the concordance among the prediction calls by applying the classifier on RNA-seq data and true-class labels, we found that the classifier made 95% (Table 3; Data Supplement) similar subtype calls between the two platforms and achieved 91% prediction accuracy compared with the true-class labels (Data Supplement).

Next, we tested the gene-based classification models derived from the TCGA exon array data set on data from two independent studies.^{4,35} After filtering outlier genes (not well-behaved genes) between the exon array and other microarray platforms by fitting regression models between the mean gene expression vectors from the exon array and microarray platforms, we retained 11,319 and 7,280 genes from the GSE9891 and GSE26712 data sets, respectively. Two separate classifiers that consisted of 106 and 43 variables were derived for the GSE9891 and GS26712 data sets, respectively (Data Supplement). Because the true accuracy of the classifiers cannot be assessed as a result of the nonavailability of the class labels in these two data sets, we evaluated the degree of concordance between the previously defined subgroups and the subtypes predicted by our classifiers. We compared the predicted molecular subtypes with those of previously predicted molecular

TABLE 4. Overlap in Cluster Membership of 245 Ovarian Serous Samples Between Our Predicted Isoform-Based Subgroups and AOCs Clusters (Gene Expression Omnibus GSE9891)

AOCs Cluster	Predicted Isoform-Based Subgroup			
	Differentiated	Immunoreactive	Mesenchymal	Proliferative
C1	12	11	57	1
C2	3	36	1	0
C3	13	0	0	5
C4	23	8	0	9
C5	1	1	0	31
C6	0	0	0	1
NC	18	7	0	7

NOTE. From Tothill et al.⁴

Abbreviations: AOCs, Australian Ovarian Cancer Study; NC, no cluster assigned.

subgroupings (C1 [high stromal response], C2 [high immune signature], C4 [low stromal response], and C5 [low immune signature] for HGSOc) by AOCs.⁴ Overall, we found that the isoform-level subtyping was significantly associated with the AOCs study subtyping (χ^2 test $P < .001$), where the C1, C2, C4, and C5 subgroups mapped to the M, I, D, and P isoform-based subgroups, respectively (Table 4). Within the combined data set of 430 patients, the predicted subgroups showed a significant difference in OS ($P < .001$). Moreover, the OS pattern of the predicted subgroups mirrored that of our isoform-based subgrouping, and the patients grouped under the M class were found to have the highest risk of death, with a median survival of 2.25 years (Fig 3).

DISCUSSION

Molecular classification of cancers is essential for developing personalized therapies.³⁷ Although gene expression-based patient stratification strategies have been published for numerous cancer studies, the reproduction and validation of the derived gene signatures across different laboratories or profiling platforms have proven to be a complex and difficult informatics problem. In classification, the main goal is to derive a probabilistic model for predicting the class membership of a new observation (patient with ovarian cancer) on the basis of a training set of data (tumor samples) that contains observations (eg, gene expression measures) for which class membership is already known. A considerable overlap of the molecular subtypes exists among various HGSOc publications,^{3,4,38,39} but these studies were solely based on gene-level expression estimates and ignored the variability associated with splice/transcript variants. Although aberrant expression of splice variants in ovarian cancer has been reported,⁴⁰⁻⁴² the current study is the first to our knowledge that has explored the use of isoform-level transcriptome in the molecular subtyping of ovarian cancer. Our isoform-level subtypes provide more prognostic information than the gene-level subtypes that were previously reported.^{3,4,35,43} Similar overall prognostic significance among the four molecular subtypes have been reported,³⁹ but we noticed some inconsistencies between independent data sets. For example, we observed similar overall prognostic significance and pairwise (D v M, I v M) survival differences in both TCGA (Fig 1D) and independent data sets (Fig 3). However, the survival difference between the I and M subtypes was inconsistent between the TCGA and the other two cohorts, which could be due to differences in the patient population and errors in the subtype calls on data from different platforms. Although the univariable Kaplan-Meier survival analysis showed both D and I isoform-based subtypes as significantly different in their OS from the M subtype, multivariable Cox PH regression analysis showed only the I subtype as significantly different from the M subtype after adjusting for the clinical variables.

Having established improved association of the refined subgroups with the OS, we translated the exon array-based

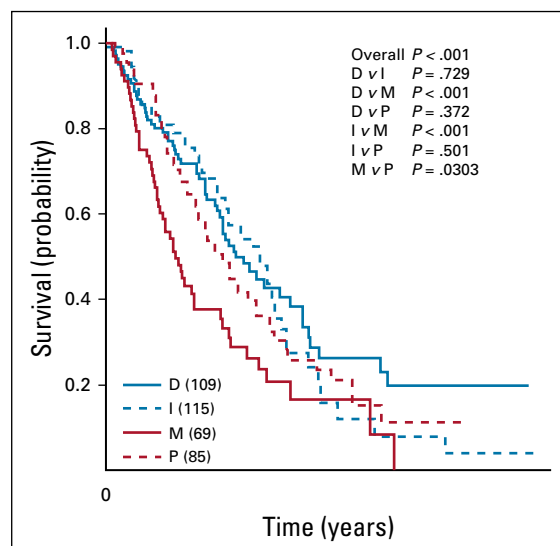


FIG 3. The prediction of four classes in 378 high-grade serous ovarian carcinoma samples obtained from Gene Expression Omnibus (GSE9891 and GSE26712) data set was clinically evaluated for prognostic determinates. Kaplan-Meier survival plot of the four subtypes shows significant difference in overall survival rate ($P < .001$).

classification system to independent platforms. The testing of the derived models on data from two independent platforms showed a high level of cross-platform (exon array to RNA-seq) consistency and accuracy ($> 90\%$) without loss of analytic precision. In addition, the classification system simplified the interplatform translation with the selection of a small subset of genes and/or transcript variants as feature variables. The small set of genes/transcript variants for molecular classification will be clinically useful and cost effective for patient subgroupings and a key resource to develop precision medicine strategies further.⁴³ For example, molecular subtypes with poor prognoses (P and M subgroups) have been reported to benefit from treatment with the vascular endothelial growth factor inhibitor bevacizumab.^{44,45} Furthermore, patients in the P subgroup are sensitive to poly (ADP-ribose) polymerase inhibition (veliparib).^{43,46} These data provide a rational basis for selecting specific treatments for histologic and molecular subtypes of ovarian cancer. A major limitation of the classification system is the requirement of the tumor transcriptome profile from either a microarray or a next-generation sequencing platform. To translate the assay to a low-dimensional platform, such as quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) or NanoString (NanoString Technologies, Seattle, WA), additional experiments are required on an independent patient cohort.³³ In addition, the combining of the molecular subtype information with pathologic subtypes, advanced imaging analytics, and radiomics approaches would enable improved precision diagnostics and treatment planning.^{47,48}

In conclusion, we developed a new platform-independent isoform-level classification system for efficient and accurate stratification of patients with HGSOc with prognostic

significance. The classifiers derived here have the potential to develop into prognostic biomarkers for stratification of patients with HGSOc.

AFFILIATION

¹Northwestern University, Chicago, IL

CORRESPONDING AUTHOR

Ramana V. Davuluri, PhD, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Dr, 11-168, Chicago, IL 60611; e-mail: ramana.davuluri@northwestern.edu.

EQUAL CONTRIBUTION

S.S. and R.V.D. contributed equally to this work.

PRIOR PRESENTATION

Presented at the American Association for Cancer Research Addressing Critical Questions in Ovarian Cancer Research and Treatment, Pittsburgh, PA, October 1-4, 2017.

SUPPORT

Supported by the National Library of Medicine of the National Institutes of Health (R01LM011297; R.V.D.) and partially supported by the Phebe Novakovic Fund and the John and Ruth Brewer Endowment.

AUTHOR CONTRIBUTIONS

Conception and design: Shohreh Shahabi, Ramana V. Davuluri

Financial support: Shohreh Shahabi, Ramana V. Davuluri

Administrative support: Shohreh Shahabi, Ramana V. Davuluri

Collection and assembly of data: Arunima Shilpi, Manoj Kandpal

Data analysis and interpretation: Arunima Shilpi, Manoj Kandpal, Brandon L. Seagle, Yanrong Ji, Shohreh Shahabi, Ramana V. Davuluri

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated.

Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Shohreh Shahabi

Research Funding: AbbVie (Inst)

Ramana V. Davuluri

Patents, Royalties, Other Intellectual Property: Patent US10113201B2: Methods and Compositions for Diagnosis of Glioblastoma or a Subtype Thereof (Inst)

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank W.S. Dhiman, PhD, and A.K. Grace, PhD, for their scientific review of the manuscript.

REFERENCES

1. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2016. *CA Cancer J Clin* 66:7-30, 2016
2. Reid BM, Permuth JB, Sellers TA: Epidemiology of ovarian cancer: A review. *Cancer Biol Med* 14:9-32, 2017
3. Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. *Nature* 474:609-615, 2011 [Erratum: *Nature* 490:298, 2012]
4. Tothill RW, Tinker AV, George J, et al: Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* 14:5198-5208, 2008
5. Pal S, Gupta R, Kim H, et al: Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res* 21:1260-1272, 2011
6. Pan Q, Shai O, Lee LJ, et al: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413-1415, 2008 [Erratum: *Nat Genet* 41:762, 2009]
7. Wang ET, Sandberg R, Luo S, et al: Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470-476, 2008
8. Khoury MP, Bourdon JC: p53 isoforms: An intracellular microprocessor? *Genes Cancer* 2:453-465, 2011
9. Grabowski P: Alternative splicing takes shape during neuronal development. *Curr Opin Genet Dev* 21:388-394, 2011
10. Pal S, Gupta R, Davuluri RV: Alternative transcription and alternative splicing in cancer. *Pharmacol Ther* 136:283-294, 2012
11. Lapuk A, Marr H, Jakkula L, et al: Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res* 8:961-974, 2010
12. Misquitta-Ali CM, Cheng E, O'Hanlon D, et al: Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer. *Mol Cell Biol* 31:138-150, 2011
13. Ebert B, Bernard OA: Mutations in RNA splicing machinery in human cancers. *N Engl J Med* 365:2534-2535, 2011
14. Hofstetter G, Berger A, Fiegl H, et al: Alternative splicing of p53 and p73: The novel p53 splice variant p53delta is an independent prognostic marker in ovarian cancer. *Oncogene* 29:1997-2004, 2010
15. Sosulski A, Horn H, Zhang L, et al: CD44 splice variant v8-10 as a marker of serous ovarian cancer prognosis. *PLoS One* 11:e0156595, 2016
16. He X, Ee PL, Coon JS, et al: Alternative splicing of the multidrug resistance protein 1/ATP binding cassette transporter subfamily gene in ovarian cancer creates functional splice variants and is associated with increased expression of the splicing factors PTB and SRp20. *Clin Cancer Res* 10:4652-4660, 2004
17. Bagchi S, Wise LS, Brown ML, et al: Structure and expression of murine malic enzyme mRNA. Differentiation-dependent accumulation of two forms of malic enzyme mRNA in 3T3-L1 cells. *J Biol Chem* 262:1558-1565, 1987
18. Tilli TM, Franco VF, Robbs BK, et al: Osteopontin-c splicing isoform contributes to ovarian cancer progression. *Mol Cancer Res* 9:280-293, 2011
19. Climente-González H, Porta-Pardo E, Godzik A, et al: The functional impact of alternative splicing in cancer. *Cell Reports* 20:2215-2226, 2017

20. Zhu J, Chen Z, Yong L: Systematic profiling of alternative splicing signature reveals prognostic predictor for ovarian cancer. *Gynecol Oncol*, 148:368-374, 2018
21. Grossman RL, Heath AP, Ferretti V, et al: Toward a shared vision for cancer genomic data. *N Engl J Med* 375:1109-1112, 2016
22. Turro E, Lewin A, Rose A, et al: MMBGX: A method for estimating expression at the isoform level and detecting differential splicing using whole-transcript Affymetrix arrays. *Nucleic Acids Res* 38:e4, 2010
23. Workman C, Jensen LJ, Jarmer H, et al: A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 3: research0048, 2002
24. Brunet JP, Tamayo P, Golub TR, et al: Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 101:4164-4169, 2004
25. Devarajan K: Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLOS Comput Biol* 4:e1000029, 2008
26. Gaujoux R, Seoighe C: A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11:367, 2010
27. Lovmar L, Ahlford A, Jonsson M, et al: Silhouette scores for assessment of SNP genotype clusters. *BMC Genomics* 6:35, 2005
28. Therneau T: A package for survival analysis in S. version 2.38, 2015. <https://CRAN.R-project.org/package=survival>
29. Benjamini Y, Hochberg Y: Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289-300, 1995
30. van Buuren S, Groothuis-Oudshoorn K: mice: Multivariate imputation by chained equations in R. *J Stat Softw* 45:1-67, 2011
31. Broad Institute: Picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data such as SAM/BAM/CRAM and VCF, 2018. <http://broadinstitute.github.io/picard>
32. Li B, Dewey CN: RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323, 2011
33. Pal S, Bi Y, Macyszyn L, et al: Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes. *Nucleic Acids Res* 42:e64, 2014
34. Jung S, Bi Y, Davuluri RV: Evaluation of data discretization methods to derive platform independent isoform expression signatures for multi-class tumor subtyping. *BMC Genomics* 16:S3, 2015
35. Bonome T, Levine DA, Shih J, et al: A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res* 68:5478-5486, 2008
36. Clough E, Barrett T: The Gene Expression Omnibus database. *Methods Mol Biol* 1418:93-110, 2016
37. Symeonides S, Gourley C: Ovarian cancer molecular stratification and tumor heterogeneity: A necessity and a challenge. *Front Oncol* 5:229, 2015
38. Verhaak RG, Tamayo P, Yang JY, et al: Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest* 123:517-525, 2013
39. Konecny GE, Wang C, Hamidi H, et al: Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J Natl Cancer Inst* 106:dju249, 2014
40. Davy G, Rousselin A, Goardon N, et al: Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur J Hum Genet* 25:1147-1154, 2017
41. Klinck R, Bramard A, Inkel L, et al: Multiple alternative splicing markers for ovarian cancer. *Cancer Res* 68:657-663, 2008
42. Brosseau JP, Lucier JF, Nwilati H, et al: Tumor microenvironment-associated modifications of alternative splicing. *RNA* 20:189-201, 2014
43. Tan TZ, Miow QH, Huang RY, et al: Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol Med* 5:1051-1066, 2013
44. Kommos S, Winterhoff B, Oberg AL, et al: Bevacizumab may differentially improve ovarian cancer outcome in patients with proliferative and mesenchymal molecular subtypes. *Clin Cancer Res* 23:3794-3801, 2017
45. Gourley C, McCavigan A, Perren T, et al: Molecular subgroup of high-grade serous ovarian cancer (HGSOC) as a predictor of outcome following bevacizumab. *J Clin Oncol* 32, 2014 (suppl; abstr 5502)
46. Coleman RL, Sill MW, Bell-McGuinn K, et al: A phase II evaluation of the potent, highly selective PARP inhibitor veliparib in the treatment of persistent or recurrent epithelial ovarian, fallopian tube, or primary peritoneal cancer in patients who carry a germline BRCA1 or BRCA2 mutation - An NRG Oncology/ Gynecologic Oncology Group study. *Gynecol Oncol* 137:386-391, 2015
47. Ohsuga T, Yamaguchi K, Kido A, et al: Distinct preoperative clinical features predict four histopathological subtypes of high-grade serous carcinoma of the ovary, fallopian tube, and peritoneum. *BMC Cancer* 17:580, 2017
48. Rathore S, Akbari H, Rozycki M, et al: Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* 8:5087, 2018

