

Single-sample expression-based chemo-sensitivity score improves survival associations independently from genomic mutations for ovarian cancer patients

Michael T. Zimmermann, PhD¹, Guoqian Jiang, MD, PhD¹, Chen Wang, PhD^{1*}

¹Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

*Corresponding author electronic address: wang.chen@mayo.edu

Abstract

Platinum-based chemotherapies are first-line treatments for ovarian cancer (OC) patients. Although chemotherapy has a high initial response rate, some patients exhibit inherent chemo-resistance. With advancements of molecular and genomic profiling, it is of high interest to identify molecular and genomic signatures predictive of chemo-sensitivity priori to treatment initiation in order to better personalize care decisions. Previous efforts have made use of mRNA expression levels of selected genes responsible for repairing DNA damage, under the hypothesis that chemo efficacy is associated with their proficiency. However, the resulting scores have been difficult to interpret. In this study, we designed a single-sample based approach known as eCARD to investigate chemo-sensitivity in ovarian cancer patients from The Cancer Genome Atlas. We demonstrated that the proposed single-sample based approach can lead to a molecular-based chemo-sensitivity score predictive of prognosis, which validates in 5 independent cohorts, and associates with increasing mutation burden and likelihood of BRCA1/2 mutation.

Introduction

Ovarian cancer (OC) is one of leading causes of cancer mortality among women in the United States. About 70% of patients present at diagnosis with advanced-stage and high-grade serous ovarian cancer (1). Platinum-based chemotherapy is a standard treatment following a cytoreductive surgery. Approximately 25% of patients develop platinum-resistance within six months and almost all patients with recurrent disease ultimately develop platinum resistance (2, 3). In addition, partly due to the lack of successful treatment strategies, the overall five-year survival rate for high-grade serous ovarian cancer is only 31%. Although several mechanisms have been revealed to contribute to chemotherapy response (4-6), there are no valid clinical or molecular markers that effectively predict chemotherapy response. The cancer research community is compiling cancer genomic information, with the goal that new therapeutic options will be indicated, leading to tailored treatment for individual patients according to their personal tumor genome. A notable example is The Cancer Genome Atlas (TCGA) research network (1, 7, 8). TCGA has released an ovarian cancer dataset containing a large (for genomics) sample size with comprehensive genomic profiles and clinical outcome information (1). The dataset has been utilized to analyze chemotherapeutic response in ovarian cancers in several previous studies (8-12).

Since the major antineoplastic mechanism of platinum-based chemotherapy is to induce DNA damage in tumor cells, leading to apoptosis, chemo-resistance is believed to be associated with proficiency of DNA-damage repair pathways, such as ataxia-telangiectasia-mutated (ATM), Fanconi anemia (FA) and Nucleotide excision repair (NER) (13-15). Such prior biological knowledge of DNA repair pathways and the genes they contain, has led to hypothesis-driven molecular analysis to develop molecular scores associated with OC patient's prognosis (9). However, we have found that the resulting scores lacked interpretability of their genomic basis (high scores indicate greater repair activation being associated with improved survival) (12, 16) and did not exhibit ideal reproducibility for different expression platforms applied to the same OC patient subjects.

In the present study, our goal was to develop a simple yet robust molecular score known as eCARD (expression CDF transform of Rank Distribution) that is able to predict a single patient's chemo-sensitivity, and link the score to an interpretable genomic basis. We first developed our single-sample based expression scoring scheme by restricting our attentions to expression of a set of DNA repair genes. Next, robustness of our score with respect to assay platform was analyzed. We then performed associations between molecular scores vs. BRCA1/2 mutation status and global mutation burdens in OC patients with whole Exome sequencing measurements. We found that eCARD exhibits only a modest association with BRCA1/2 mutation status, as well as global mutation burden. Thus, expression-based scores are not proxies for these known prognostic features. Further, using Cox regression, we conducted univariate and multivariate survival analysis to investigate if our proposed molecular score predicts

patient outcomes. Finally, we confirmed associations of our chemo-sensitivity score within five independent cohorts where expression, clinical annotation, and follow-up data were provided.

Methods

TCGA expression

Three expression datasets from TCGA were downloaded through the curatedOvarianData version 1.3.5 (17) package for R (18): Unified microarray Expression (UE; where each gene’s expression is a median across three individual microarray platforms), individual microarray platforms (Affymetrix U133A), and RNA-Seq. Nearly half (46%) of OC patients’ tumor gene expression is assayed using all three datasets. Affymetrix-based expression of TCGA OC patients was used to train molecular features used in our score. Comparison to the other expression platforms (e.g. RNA-seq) for the same subjects was used to quantify technical reproducibility and robustness.

TCGA mutations (BRCA1/2 and global mutation status)

We downloaded the confident somatic mutation calls of the OC patients from pan-TCGA mutation study (<https://www.synapse.org/#!Synapse:syn1461171>) (19). BRCA1/2 mutation status of OC patients were extracted and organized from previous related TCGA OC studies (1, 20).

Selection of prognosis-relevant genes

Cox proportional hazards models were used to quantify the relationship between expression of individual DNA damage repair genes and overall survival. The details of DNA damage pathways and related gene enumeration was described in (9) and its online supplementary materials. Candidate genes were selected from known DNA damage repair pathways. Any gene which exhibited a marginally significant trend with survival ($P < 0.15$) was included in our final score (9). The sign of the cox-regression coefficient was used to classify genes as indicating “chemo-sensitive” if higher expression values associated with better outcome and indicating “chemo-resistant” if higher expression values associated with worse outcome. This classification is based on the assumption that gene expression leading to chemo-sensitivity imparts a direct survival benefit, and vice versa.

Construction of molecular scores for chemo-sensitivity score by eCARD

Previous studies (9) used binary representations of gene expression, indicating if a gene’s expression is greater than the median level across patients, or not. A median-based dichotomized score (MDS) may be sensitive to biologic and technical noise, especially for samples exhibiting expression close to the median level. Similar to previous studies, the MDS for the k^{th} subject is the per-subject sum of DNA Repair Genes (G) that are above the median expression level: $MDS_k = \sum_i G_k^i \geq \bar{G}^i$.

We developed a single-sample expression scheme named eCARD for quantifying the prognosis value of “chemo-sensitive” and “chemo-resistant” genes which are learned from a training cohort (see above). Given an expression sample of N genes, eCARD computes “chemo-sensitive” and “chemo-resistant” scores, per sample, by summarizing rank-transformed gene expression:

$$\text{sensitive_score}(\mathbf{x}) = \sum_{i \in \text{sensitive-Set}} CDF^{-1}\left(\frac{\text{rank}_{\mathbf{x}}(x_i)}{N + \sigma}\right); \text{resistant_score}(\mathbf{x}) = \sum_{j \in \text{resistant-Set}} CDF^{-1}\left(\frac{\text{rank}_{\mathbf{x}}(x_j)}{N + \sigma}\right).$$

In which, $\text{rank}_{\mathbf{x}}(x_i)$ is the relative rank of i -th gene expression value inform among all N measured genes in the give sample; the higher expression value of i -th gene x_i , the higher rank position it has (e.g. the gene with largest expression value will be ranked at 1st). $CDF^{-1}(\cdot)$ is an inverse cumulative density function (CDF) transformation function according to the normal distribution, so that an input will be transformed to a normally distributed signal. The purpose of rank-transformation is to improve robustness of expression measurements, so that the scoring scheme can be easily generalized across different platforms. For instance, if many genes are expressed at similar levels, small amounts of noise or technical variation could significantly change a given gene’s rank, but noise will not significantly change the normalized rank. The inverse CDF transformation, $CDF^{-1}(\cdot)$, further prioritizes the expression changes close to the top or bottom of the ranking list (the highest and lowest genes spread out along the tails of the normal distribution), and puts less emphasis for the genes with intermediate ranks. In another word, the $CDF^{-1}(\cdot)$ function enforces a nonlinear weighting scheme to highlight genes with large expression changes and down-weights moderate changes in expression, within a single sample. The final chemo-sensitivity score for the k^{th}

subject is defined as the difference between these two scores; multiplying by an empirically determined constant for scale:

$$eCARD_k = 10 * (\text{sensitive_score}(\mathbf{x}_k) - \text{resistant_score}(\mathbf{x}_k))$$

Application of eCARD Score in Independent Cohorts

Independent cohorts of patient samples were from GEO (21) and used to evaluate and validate our scoring scheme's generalizability: GSE49997 (22), GSE32062 (23), GSE9891 (24), GSE26712 (25), and Bentink *et al.* (26).

Survival Analysis

In order to focus our survival analysis on the most homogeneous and relevant OC patient population, we limited all cohorts to primary high-grade serous OC receiving platinum and taxane therapies, and excluded patients treated in the neo-adjuvant setting. We used overall survival (OS) time as the clinical endpoint in all cohorts. Each cohort is evaluated independently using both univariate (eCARD score alone) and multivariate (score and patient age, tumor stage, grade, and surgical de-bulking status, when available from each study) Cox proportional hazards models. The outcome of Cox regression is reported in terms of a Hazard Ratio (HR). Values greater than 1 indicate increased hazard (probability of cancer-related mortality), while values less than 1 indicate decreased relative risk, compared to a reference group. In order to summarize across the cohorts, we generate a mixed effects Cox model, adding a random intercept variable by cohort membership which models a different baseline hazard for each study. The HR reported for each score is the association of the score with OS, independent of baseline hazard and other clinical covariates. All analyses were performed in the R programming language version 3.2.0 (18), leveraging the packages survival v2.38.3 and coxme v2.2.5.

Results

In total, we identified 24 genes for which higher expression levels were indicative of chemo-sensitivity, and nine genes for which higher expression levels indicated chemo-resistance. Table 1 lists these genes by which DNA repair pathway they contribute. BER and MMR pathways contain genes exclusively associated with better survival outcomes, while genes within NER and HR show mixed associations.

First, we investigated the technical reproducibility of our eCARD score using gene expression across platforms. For example, we quantify the extent of agreement between microarray-based expression signatures and RNA-seq measurements. The MDS DNA-Damage score (similar to Kang *et al.* (9)) and our eCARD score were calculated for each platform. All datasets correlate well with each other but our eCARD score exhibits consistent gains in cross-platform reproducibility, increasing the correlation between U133A and RNA-Seq to 0.88 and between UE and RNA-Seq to 0.93. Overall, our method achieves better concordance between independent platforms of the same patients.

Table 1. Genes and associated pathways chosen in the eCARD model to predict outcome

pathway abbreviation	pathway full name	genes chosen in the eCARD model (blue: associated with better survival; red: associated with worse survival)
ATM	ataxia telangiectasia mutated	<i>CHEK1, H2AFX, RAD1, RAD9A, RNF8, TP53BP1</i>
BER	base excision repair	<i>APEX1, APEX2, MUTYH, PARP2, SMUG1, TDPI, UNG</i>
FA/HR	Fanconi Anemia /homologous recombination	<i>BLM, C17orf70, FANCF, FANCI, PALB2, RAD54L, NBN, RAD52</i>
MMR	mismatch repair	<i>MLH3, POLE</i>
NER	nucleotide excision repair	<i>CETN2, DDB1, DDB2, ERCC1, ERCC2, XAB2</i>
NHEJ	non-homologous end joining	<i>XRCC4</i>
OTHER	other	<i>PER1</i>
RECQ	recQ helicase pathway	<i>RECQL</i>
XLR	cross-link repair	<i>DCLRE1A</i>

Table 2. Across-platform concordance of median-based and eCARD chemo-sensitivity scores. Gain is the gain in performance for our eCARD score over the Median-Dichotomized Score (see Methods). Affy signifies Affymetrix U133A; RNA, mRNA-Seq-V2; UE, Unified Expression. These comparisons were made using all patient samples common to a given pair of expression datasets.

	MDS			eCARD			Gain		
	Affy vs RNA	UE vs RNA	UE vs Affy	Affy vs RNA	UE vs RNA	UE vs Affy	Affy vs RNA	UE vs RNA	UE vs Affy
Cramer's V	0.62	0.62	0.65	0.69	0.72	0.80	0.07	0.10	0.15
Spearman's CC	0.82	0.80	0.83	0.88	0.92	0.92	0.06	0.12	0.09
Linear R ²	0.69	0.62	0.68	0.80	0.86	0.84	0.11	0.24	0.16

While our scoring method is based on only gene expression levels, we further incorporated other genomic measurements, such as the number of single-nucleotide variants per sample, into the considerations of chemo-sensitivity/resistance prediction. Multiple previous studies have demonstrated that mutation burden and constitutional mutations, such as BRCA1/2, are predictive of chemotherapy response for OC patients. It is critical to investigate if expression-based scores introduce independent predictive information, or if they simply serve as surrogates for underlying genomic aberrations. In Fig. 1 (c), we showed that BRCA1/2 mutation status in 316 subjects with exome-seq measurements, is highly correlated with increasing mutation burden (t-test p-value = 1.2e-04); in contrast, BRCA1/2 mutation status only has a marginally significant associations with eCARD score.

In the same TCGA-OC cohort (316 subjects, 13 removed from missing data, 172 with cancer-related mortality), we performed multivariate survival analysis, incorporating expression signatures with genomic information and important clinical factors (age, grade, stage). The result is shown in Table 3, from which chemo-sensitivity score indeed provides additional independent prediction in multivariate models, where increase of each unit of eCARD score is indicative of 20% of risk reduction (HR = 0.78, 95% CI = [0.71, 0.86]).

In an expanded effort to validate if proposed scoring provided a generalized predictive model for prognosis of OC patients. We chose 5 OC cohorts, each with at least 100 patient subjects and expression measurements (22-26). A univariate mixed effects model shows a strong relationship between eCARD score across these independent cohorts (HR = 0.72; P = 2.1E-3). Similar effects were observed in multivariate mixed-effects models (HR = 0.67; P = 2.4E-3; see Figure 2).

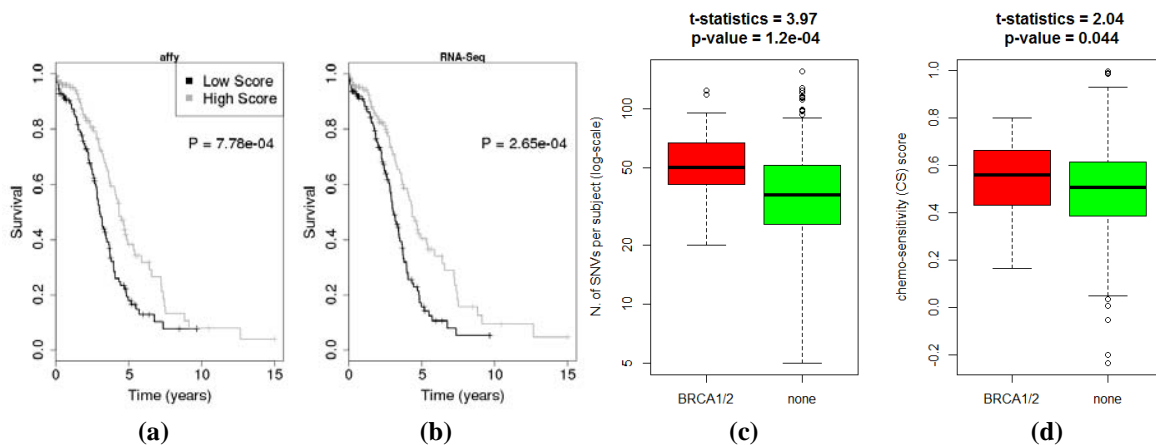


Figure 1. KM-plot of TCGA OC overall survival showing high eCARD scores, computed from expression measured by a) Affymetrix U133a microarray platform and b) mRNA sequencing, are associated with improved survival. To contrast whether eCARD scores are mainly determined by BRCA1/2 mutation status, (c) and (d) display mutation burden (measured by number of SNVs per sample) vs. BRCA1/2 mutation status, and chemo-sensitivity score vs. BRCA1/2 mutation status, respectively.

We compare our score to other previously published chemotherapy sensitivity scores. CLOVAR (10) is an expression-based score that utilizes a small set of representative genes from multiple DNA damage repair pathways. A subset of samples meeting our selection criteria (N = 167) is labeled as “Good” or “Poor” prognosis based on their

score (1, 10). We have used this designation to compare the dichotomized CLOVAR score to the dichotomized eCARD score. Both show significant association with OS, with eCARD (HR = 0.50, CI = [0.33, 0.77], $p = 1.8 \times 10^{-3}$) exhibiting a more consistent association compared to CLOVAR (HR = 0.58, CI = [0.38, 0.90], $p = 1.4 \times 10^{-2}$). The same trend holds after adjusting for patient age and surgical de-bulking status: eCARD (HR = 0.54, CI = [0.34, 0.85], $p = 8.4 \times 10^{-3}$) and CLOVAR (HR = 0.61, CI = [0.38, 0.98], $p = 4.1 \times 10^{-2}$). PROVAR scores were shown to be of greater prognostic power than CLOVAR (11). We calculated PROVAR scores according to the published formula (a weighted sum of 9 protein levels) and using the Revers-Phase Protein Array (RPPA) data provided in their supplemental data tables (11). For comparisons to PROVAR, only the subset of patients with both Unified Expression and RPPA data was used. Within this subset and adjusting for covariates, eCARD is more consistently associated with OS (HR = 0.94, CI = [0.91, 0.97], $p = 2.6 \times 10^{-4}$) than PROVAR (HR = 0.26, CI = [0.09, 0.78], $p = 1.7 \times 10^{-2}$). Univariate analyses yield similar estimates (not shown). Our eCARD score also exhibits a more consistent association with OS among high-grade OC patients than previously established prognostic scores.

Table 3: The multivariate cox-regression analysis results of chemo-sensitivity scores, along with age, tumor-grade, stage, BRCA1/2 mutation status and global mutation burden, which is measured by total number of somatic SNVs in each tumor sample.

	multivariate p-value	HR	95% Conf. Int.	
diagnosis age	1.99E-02	1.017	1.003	1.031
tumor-grade (grade-3 vs. -2)	2.31E-01	1.383	0.813	2.352
tumor-stage (stage-III vs. -II)	6.00E-01	1.280	0.509	3.218
(stage-IV vs. -II)	3.35E-01	1.612	0.611	4.255
eCARD score	5.87E-07	0.784	0.712	0.862
BRCA status (BRCA1/2 vs. none)	2.99E-02	0.629	0.414	0.956
global mutation burden (N. SNVs per sample)	6.48E-03	0.991	0.985	0.998

Table 4: Comparison of eCARD and MDS scores across three high-throughput data platforms. Hazard ratios (HRs) are presented from multivariate (Multi) and univariate (Uni) Cox regression, and either a single microarray platform (U133A), the per-gene median across three microarray platforms (Unified Expression), or mRNA sequencing from TCGA.

	Affymetrix U133A		mRNA-Seq-V2		Unified Expression	
	Multi	Uni	Multi	Uni	Multi	Uni
N	356	386	156	174	319	349
eCARD						
HR	0.80	0.79	0.89	0.87	0.94	0.93
p-value	1.9×10^{-7}	2.4×10^{-9}	0.4×10^{-1}	0.4×10^{-2}	1.7×10^{-5}	2.2×10^{-7}
95% CI	0.73, 0.87	0.73, 0.85	0.80, 1.00	0.79, 0.96	0.92, 0.97	0.91, 0.96
MDS						
HR	0.97	0.96	0.95	0.96	0.97	0.95
p-value	2.1×10^{-1}	0.6×10^{-1}	0.9×10^{-1}	1.8×10^{-1}	2.7×10^{-1}	2.1×10^{-2}
95% CI	0.92, 1.02	0.91, 1.00	0.89, 1.01	0.90, 1.02	0.92, 1.02	0.90, 0.99
eCARD (High versus Low)						
HR	0.50	0.47	0.78	0.64	0.58	0.56
p-value	1.6×10^{-5}	0.6×10^{-6}	3×10^{-1}	0.4×10^{-1}	0.7×10^{-3}	2.1×10^{-4}
95% CI	0.37, 0.69	0.35, 0.63	0.49, 1.24	0.42, 0.98	0.43, 0.80	0.42, 0.76

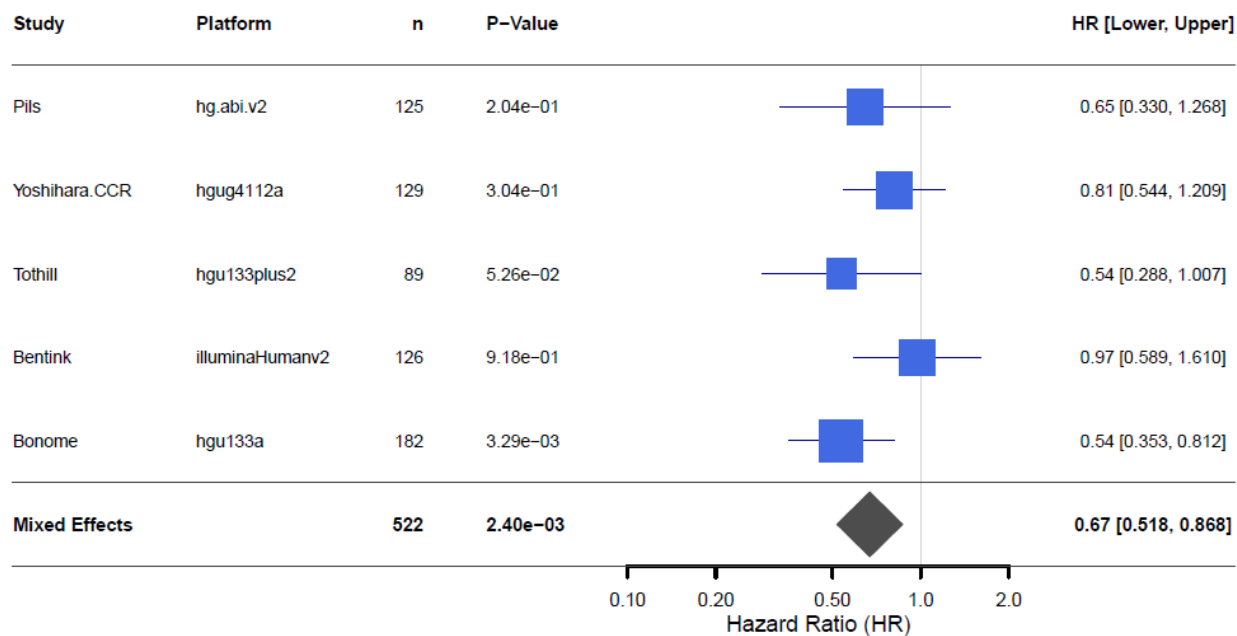


Fig. 2. Forest plot summarizing the eCARD scores from multivariate Cox regression of 5 independent validation cohorts. A mixed effects model summarizes across the cohorts.

Discussion

DNA-Damage scores have been met with mixed reactions (16), partly due to the difficulty of interpreting the results given that the observed survival associations show a protective effect of higher scores. Higher scores, indicative of greater DNA-Damage pathway activity or capacity (expression of the genes contained), were hypothesized to increase the likelihood that tumor cells could repair chemotherapy-induced lesions. If this hypothesis were true, higher scores would be associated with poorer survival. This may be due to tumor cellular mechanisms compensating homologous-repairing deficiency (12), which is one of the key characteristics of ovarian tumors (27, 28). Also, it may be that the DNA-Damage repair pathway activity is more informative of the state of the tumor cells and the extent of pre-existing genomic instability, rather than indicative of direct cellular response to cytotoxic agents. We hypothesize that tumors exhibiting higher DNA-Damage pathway activity scores have more stable genomes, perhaps due to more intact damage monitoring and repair systems. This state makes chemotherapy-induced DNA damage a more disruptive change for these tumors, compared to tumors which already exhibit high instability.

High reproducibility between technical platforms is a necessity for modern precision initiatives (29). We have shown our version of the DNA-Damage score, eCARD, to be more reproducible between two common technology platforms. Further, we have not required thorough normalization procedures, favoring a simple data transformation that naturally emphasizes the most prominent expression features and de-emphasizes small differences that are more likely to be affected by biologic and technical noise.

A feature of our method is that we have taken into account the sign of each gene; if its expression level is positively or negatively associated with survival. This is both a positive feature and a potential limitation. It is positive because we are not confounding opposing effects. However, this procedure could potentially limit the generalizability of our score. One could argue that because we have chosen the sign of each gene based on its independent association with survival, that we are highly biased to a positive result. To address this concern, we have tested our scoring method on five independent cohorts which validate its utility.

Hazard ratios (HRs) are calculated either for one group versus another, or for a unit-increase in a continuous score. Thus, when considering the impact of a score, one must also consider the range of values that the score attains across a given dataset. For example, the HR of eCARD is 0.94 and of PROVAR is 0.26, for the cohort with both Unified Expression and RPPA data. The Inter-Quartile Range (IQR; 75th and 25th percentiles) of eCARD score for this cohort is [-3.8, 4.6]. Thus, moving from the first to the third quartile is a difference of more than 8 units, indicating a relative risk of 0.61. However, the IQR of PROVAR is [-0.13, 0.08]; a difference of 0.22 units,

indicating a relative risk of 0.75. Thus, eCARD has a more consistent survival association, which is more interpretable after accounting for their different ranges.

Previously published prognostic scores were generated after thorough data normalization. While thorough data normalization is appropriate for approaching the true expression level, it can be a barrier to reproducibility and application to single samples. We have not attempted to fully reproduce all normalization steps for each score. Instead, we aim to identify if there is an approach that could be more readily adapted and applied to diverse datasets and potentially single samples. Our eCARD score shows significantly more stable cross-platform reproducibility and robust survival association. Future work could expand the eCARD score to include representatives from additional repair mechanisms and further refine the interplay between these features.

Conclusion

In summary, we have proposed a single sample-based scoring method known as eCARD for evaluating OC tumors. We have shown the prognostic value of our eCARD score, independent of major genomic features such as BRCA mutation, across five validation cohorts. Further, our score exhibits less sensitivity to data assay platform (array versus NGS) and normalization method. Finally, it exhibits more consistent association with survival than three previously established prognostic scores.

Acknowledgements: The study is supported in part by a NCI U01 Project – caCDE-QA (1U01CA180940-01A1), and Mayo Clinic Ovarian Cancer SPORE (P50 CA136393).

References

1. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011 Jun 30;474(7353):609-15.
2. Peters GJ, Avan A, Ruiz MG, Orsini V, Giovannetti E, Smit EF. Predictive role of repair enzymes in the efficacy of Cisplatin combinations in pancreatic and lung cancer. *Anticancer Res*. 2014 Jan;34(1):435-42.
3. Zhang YY, Gu KS. DNA repair capacity in peripheral blood lymphocytes predicts efficacy of platinum-based chemotherapy in patients with gastric cancer. *Asian Pac J Cancer Prev*. 2013;14(9):5507-12.
4. Dasari S, Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol*. 2014 Oct 5;740:364-78.
5. Bowden NA. Nucleotide excision repair: why is it not used to predict response to platinum-based chemotherapy? *Cancer Lett*. 2014 May 1;346(2):163-71.
6. Lange SS, Takata K, Wood RD. DNA polymerases and cancer. *Nat Rev Cancer*. 2011 Feb;11(2):96-110.
7. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013 Oct;45(10):1113-20.
8. Hsu FH, Serpedin E, Hsiao TH, Bishop AJ, Dougherty ER, Chen Y. Reducing confounding and suppression effects in TCGA data: an integrated analysis of chemotherapy response in ovarian cancer. *BMC Genomics*. 2012;13 Suppl 6:S13.
9. Kang J, D'Andrea AD, Kozono D. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J Natl Cancer Inst*. 2012 May 2;104(9):670-81.
10. Verhaak RG, Tamayo P, Yang JY, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest*. 2013 Jan;123(1):517-25.
11. Yang JY, Yoshihara K, Tanaka K, et al. Predicting time to ovarian carcinoma recurrence using protein markers. *J Clin Invest*. 2013 Sep;123(9):3740-50.
12. Pitroda SP, Pashtan IM, Logan HL, et al. DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. *Sci Transl Med*. 2014 Mar 26;6(229):229ra42.
13. Ghosal G, Chen J. DNA damage tolerance: a double-edged sword guarding the genome. *Transl Cancer Res*. 2013;2(3):107-29.
14. Kim H, D'Andrea AD. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev*. 2012 Jul 1;26(13):1393-408.
15. Postel-Vinay S, Vanhecke E, Olausson KA, Lord CJ, Ashworth A, Soria JC. The potential of exploiting DNA-repair defects for optimizing lung cancer treatment. *Nat Rev Clin Oncol*. 2012 Mar;9(3):144-55.
16. Swisher EM, Taniguchi T, Karlan BY. Molecular scores to predict ovarian cancer outcomes: a worthy goal, but not ready for prime time. *J Natl Cancer Inst*. 2012 May 2;104(9):642-5.
17. Ganzfried BF, Riester M, Haibe-Kains B, et al. curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database (Oxford)*. 2013;2013:bat013.

18. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (<http://www.R-project.org>). 2014.
19. Kandath C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013 Oct 17;502(7471):333-9.
20. Birkbak NJ, Kochupurakkal B, Izarzugaza JM, et al. Tumor mutation burden forecasts outcome in ovarian cancer with BRCA1 or BRCA2 mutations. *PLoS One*. 2013;8(11):e80023.
21. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2007 Jan;35(Database issue):D5-12.
22. Pils D, Hager G, Tong D, et al. Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. *Cancer Sci*. 2012 Jul;103(7):1334-41.
23. Yoshihara K, Tsunoda T, Shigemizu D, et al. High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clin Cancer Res*. 2012 Mar 1;18(5):1374-85.
24. 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 Aug 15;14(16):5198-208.
25. Bonome T, Levine DA, Shih J, et al. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res*. 2008 Jul 1;68(13):5478-86.
26. Bentink S, Haibe-Kains B, Risch T, et al. Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PLoS One*. 2012;7(2):e30269.
27. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov*. 2015 Nov;5(11):1137-54.
28. Walsh CS. Two decades beyond BRCA1/2: Homologous recombination, hereditary cancer risk and a target for ovarian cancer therapy. *Gynecol Oncol*. 2015 May;137(2):343-50.
29. Weigelt B, Mackay A, A'Hern R, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *The Lancet Oncology*. 2010 Apr;11(4):339-49.