Published in final edited form as:

Stat Biosci. 2017 June ; 9(1): 1–12. doi:10.1007/s12561-016-9144-1.

IPI59: An Actionable Biomarker to Improve Treatment Response in Serous Ovarian Carcinoma Patients

J. Choi¹, S. Ye¹, K. H. Eng¹, K. Korthauer¹, W. H. Bradley², J. S. Rader², and C. Kendzierski¹

¹University of Wisconsin Madison, Madison, WI, USA

²Medical College of Wisconsin, Milwaukee, WI, USA

Abstract

Despite improvements in operative management and therapies, overall survival rates in advanced ovarian cancer have remained largely unchanged over the past three decades. Although it is possible to identify high-risk patients following surgery, the knowledge does not provide information about the genomic aberrations conferring risk, or the implications for treatment. To address these challenges, we developed an integrative pathway-index model and applied it to messenger RNA expression from 458 patients with serous ovarian carcinoma from the Cancer Genome Atlas project. The biomarker derived from this approach, IPI59, contains 59 genes from six pathways. As we demonstrate using independent datasets from six studies, IPI59 is strongly associated with overall and progression-free survival, and also identifies high-risk patients who may benefit from enhanced adjuvant therapy.

1 Introduction

Epithelial ovarian cancer is the leading cause of gynecologic cancer death, accounting for over 14,000 deaths annually in the United States alone [1]. Early diagnosis is difficult, and consequently most patients present at advanced stage (III or IV) where standard treatment is surgical debulking followed by a platinum-based chemotherapy. Most patients with advanced ovarian cancer (AOC) do not achieve a sustainable response under standard care and over 60 % recur within 2 years; only 30 % survive 5 years [1]. Recent clinical trials evaluating the effects of augmenting adjuvant therapy have identified a few promising agents showing remarkable benefit to some patients, but the results overall were not sufficiently uniform or great enough to warrant a change to standard treatment [2]. The discovery of clinically actionable Biomarkers that identify high-risk patients likely to benefit from an enhanced adjuvant therapy would allow for more precise and powerful trials, and ultimately lead to improved AOC patient survival.

A number of methods are available for identifying patients at high-risk for early recurrence or death. Residual disease after surgical cytoreduction is the measure most widely used in clinical practice. It has been associated with both overall and progression-free survival in

* C. Kendzierski, kendzior@biostat.wisc.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s12561-016-9144-1) contains supplementary material, which is available to authorized users.

numerous studies, with age improving predictions slightly. Although it is useful for prognostic purposes, these measures do not provide information to guide treatment. Genomic and proteomic prognostic markers have been developed toward this end, and a few are associated with biological processes that suggest broad classes of candidate therapies [3, 4]. In spite of these advances, validation studies have yet to be conducted, and the ability to identify patients likely to benefit from an enhanced adjuvant therapy remains elusive.

To identify a clinically actionable biomarker of high-risk AOC patients, we developed and applied an integrative pathway-index (IPI) model to mRNA expression and survival data collected on ovarian cancer patients as part of the Cancer Genome Atlas (TCGA) project. In short, the IPI model requires as input a collection of pathways (groups of genes) specified *a priori* and expression measured in a population of patients. Within each pathway, susceptibility (resistance) genes conferring increased (decreased) risk of a time-to-event phenotype such as recurrence or death are selected. Pathways significantly associated with risk are then identified from the full collection. In many cases, we observe sparse effects both within and across significant pathways. To derive a unified biomarker that summarizes salient features across pathways, a second variable selection is conducted that accommodates the pathway information. See Methods for further detail.

An application of the IPI model to the TCGA ovarian cohort identified a biomarker from 59 genes, referred to hereinafter as IPI59. The prognostic performance of IPI59 to identify high-risk AOC patients is evaluated in six independent patient populations, and is shown to be comparable to leading approaches, using far fewer genes. Importantly, IPI59 also identifies a sub-population of patients who may benefit from enhanced adjuvant therapy. Taken together, IPI59 may ultimately serve as a useful prognostic and predictive biomarker [5].

2 Methods

2.1 Expression Data Cohorts

Expression and clinical data for the training data set were obtained from the TCGA Data Portal (<http://cancergenome.nih.gov>) on 08/25/2015. Probe level gene expression files were downloaded before normalization (level1). We considered the 458 TCGA patients with advanced stage serous ovarian carcinoma as a training data set; expression data for the training set were collected using the Affymetrix HT Human Genome U133a platform. Data for validation studies were obtained from the gene expression omnibus, GEO [6]. Specifically, we considered 285 patients from Tothill et al. [3], a study conducted in Australia consisting of patients with ovarian, tubal, and peritoneal cancers (GSE9891); $n = 28$ patients from Konstantinopoulos et al. [7] conducted at Beth Israel Deaconess Medical Center, Boston (GSE19829); $n = 101$ patients from Lisowska et al. [8] conducted in Poland (GSE63885); $n = 58$ patients from Ferriss et al. [9] conducted at the University of Virginia (GSE30161); $n = 185$ patients from Bonome et al. [10] conducted at Massachusetts General Hospital, Boston (GSE26712); and $n = 260$ patients from Yoshihara et al. [11] conducted in Japan (GSE32062). These datasets were collected from three platforms: Affymetrix Human Genome U133 Plus 2.0 (GSE9891, GSE19829, GSE63885, and GSE30161), Affymetrix Human Genome U133a (GSE26712), and Agilent (GSE32062). For the Affymetrix

platforms, the R packages `hthgu133a.db_3.2.2` and `hgu133plus2.db_3.2.2` were used to align probes to genes across all studies. For the Agilent platform, the gene level expression data set was directly obtained from `curatedOvarianData` [12]. Robust multi-array average [13] (RMA) was used for gene expression normalization of CEL files for the TCGA training set as well as the test data sets GSE9891, GSE19829, GSE63885, and GSE30161. For GSE26712, the RMA normalized data provided at GEO were used and for GSE32062, processed data from `curatedOvarianData` [12] were used. Following normalization, each gene was centered and scaled to have unit standard deviation. To minimize the effect of long-term survivors, overall survival time (progression-free survival time) and status were truncated at 60 (48) months; to attenuate outliers, expression values greater (less) than the 97th (3rd) percentile were redefined to be equal to the 97th (3rd) percentile. Patients with low stage (stage I and stage II) and low grade (grade I) were excluded. Further information on patient exclusion is provided in supplementary Sect. 2 (Excluded patients).

2.2 Cytoreduction Status

The traditional definitions of optimal and sub-optimal cytoreduction status that are used for TCGA193 and CLOVAR100 are no residual disease greater than 1cm and residual disease greater than 1cm, respectively. However, recent reports suggest that the distinction between no residual disease (RD^-) and some residual disease (RD^+) is a more accurate metric [14–16], and Supplementary Figures S1–S2 support the new distinction. Consequently, for the TCGA training data set and GSE9891, we followed the recent distinction (RD^- vs. RD^+) of cytoreduction status for each study that provided detailed information regarding residual disease. For other studies (GSE19829, GSE63885, GSE30161, GSE26712, and GSE32062), this information was not available and so we used the status label provided; the patients with optimal cytoreduction are included in the RD^- group and the patients with sub-optimal are included in RD^+ group.

2.3 Integrative Pathway-Index Model

The IPI model requires as input a collection of pathways (groups of genes) specified *a priori* and expression measured in a population of patients. In a first step, within each pathway, genes associated with increased (decreased) risk of a time-to-event phenotype are selected. Significant pathways are then identified, and a second variable selection is conducted that uses genes from significant pathways and accommodates pathway structure. In practice, groups of genes can be identified from any source. While a number of methods exist for selecting genes within a pathway associated with a time-to-event phenotype [17–22], the IPI model uses Cox-proportional hazards lasso regression as in [17, 18, 20]. There are also many methods available for identifying significant pathways [23–27]. The IPI model uses the pathway-index model described in Eng et al. [18]. Briefly, in the pathway-index model, Cox-proportional hazards lasso regression is used to select genes where increased (decreased) expression is associated with increased (decreased) risk, referred to as susceptibility and resistance genes, respectively. The cytoreduction status is included as a covariate in the Cox-proportional hazards lasso regression. An index is calculated for each patient, defined as the difference between average expression of the susceptibility genes and average expression of the resistance genes. A patient is allocated into the high-risk (low-risk) group when her patient-specific index is greater (less) than zero; and pathways are ranked by

p values from log-rank tests between high-risk and low-risk patient groups. The top 15 pathways are selected using the pathway-index model and group-lasso [23] is then used in a second selection stage. Prior to applying group-lasso, the 15 pathways are ranked in ascending order according to the proportion of selected genes (the ratio of the number of selected genes over the pathway size) to avoid over representation by one or a few pathways. Any gene present in more than one pathway is assigned exclusively to the pathway containing the gene that has the smallest proportion of selected genes. Once top pathways of interest are identified and duplicated genes are assigned to a single pathway, the IPI model uses a sparse-group lasso to select a final list of genes [28]. To estimate the tuning parameter, 10-fold cross-validation is used, and the index of the patients is calculated. We used KEGG annotations obtained from the R package hthgu133a.db_3.2.2 available at www.bioconductor.org; 229 KEGG pathways were considered. For training the IPI model, we use the 458 TCGA patients with overall survival (progression-free survival) truncated at 60 (48) months to minimize the effect of long-term survivors.

2.4 Survival Analysis

Comparisons between survival curves were conducted using the log-rank test. Unless otherwise noted, reported *p* values test the null hypothesis that the hazard functions for all groups are equal for the study time shown.

2.5 Paired Sample Analysis

Eighteen TCGA patients are listed as having recurrent solid tumor tissue. Of these, 16 had matched primary tumor tissue samples and we aligned the level 1 TCGA expression data for all of these arrays and RMA normalized them together. Because all of these patients recurred and were re-biopsied, we regressed the time to first recurrence by the relative change in IPI59 signal.

2.6 Assessment of Enhanced Adjuvant Therapies

Treatment information available in the TCGA includes dates bounding the beginning and end of a treatment regimen. As of 08/25/2015, treatment information was available for 504 TCGA ovarian cancer patients. Among these, 411 patients were used with combination of other clinical variables. Further detail is provided in supplementary Sect. 2 (Treatment data definition). With few exceptions, each patient received a platinum-based chemotherapy (cisplatin or carboplatin) combined with paclitaxel for adjuvant therapy. In addition to platinum drugs and paclitaxel, 90 patients also received additional agent(s) for adjuvant treatment. Specifically, we considered 13 patients who received bevacizumab, 39 who received gemcitabine, and 12 who received topotecan. Other treatments were sparsely represented and details can be found in Supplement Table S3. We refer to these 90 patients as receiving enhanced adjuvant therapy.

3 Results

3.1 Identification of IPI59

As detailed in methods, pathways (group of genes) must be identified *a priori*. Of the 229 pathways defined by KEGG [29], the IPI model identified Axon guidance

(KEGG:hsa04360), Neuroactive ligand-receptor interaction (KEGG:hsa04080), Neurotrophin signaling pathway (KEGG:hsa04722), GnRH signaling pathway (KEGG:hsa04912), Wnt signaling pathway (KEGG:hsa04310) and Osteoclast differentiation (KEGG:hsa04380). The model selected a total of 81 probes from these 6 pathways. In order to refine the survival prediction, we applied Cox proportional-hazards sparse-group lasso regression on these 81 probes and identified 62 probes, which correspond to 59 genes (Supplementary table S4). For each patient, we derive an expression index from these 59 genes (30 resistance and 29 susceptibility), defined by the pathway index model as the average expression of susceptibility genes minus the average expression of resistance genes. The index is referred to throughout as IPI59.

3.2 IPI59 Predicts Overall and Progression-Free Survival in Independent Patient Populations

To evaluate the ability of IPI59 to predict overall survival, we compared its predictions of two classes (low-risk vs. high-risk) with the two leading prognostic signatures that were developed using the same TCGA cohort (noting that sample sizes vary slightly due to patient accrual over time). The first marker developed by the TCGA Consortium [30] is based on expression from 193 genes, referred to as TCGA193. The second, CLOVAR [31], uses 100 genes, referred to as CLOVAR100. These two signatures outperform all methods to date with respect to predictive utility, including cytoreduction status. As valid assessment of any prognostic marker requires evaluation in independent datasets, TCGA193, CLOVAR100, and IPI59 were evaluated on 825 patients profiled in six independent studies, summarized in Table 1. As detailed in Methods, these datasets were collected in diverse geographical locations using different array platforms over a time period spanning 20 years (1990–2010). For some datasets, there are slight differences in patient characteristics and clinical definitions (e.g., of cytoreduction status, see Methods). Given this heterogeneity, we expect the results of our validation studies to provide a fairly realistic assessment of response in a clinical setting.

In these independent populations, IPI59 shows predictive performance that is comparable to TCGA193 and CLOVAR100. The left panel of Fig. 1 shows overall survival (OS) stratified by high and low risk as assessed by IPI59 and cytoreduction status ($p = 5.31\text{e-}12$). The middle and the right panels show similar plots, for TCGA193 ($p = 1.87\text{e-}13$) and CLOVAR100 ($p = 2.92\text{e-}13$), respectively. Progression-free survival (PFS) is available for 366 of the 825 patients; results are shown in Fig. 2 ($p = 2.55\text{e-}11$, $2.68\text{e-}07$, $1.26\text{e-}4$, respectively, from the left). Taken together, these results indicate that IPI59 may be a clinically useful prognostic marker for identifying AOC patients at high-risk for early recurrence or death.

3.3 Changes in IPI59 are Associated with Platinum Resistance

Because adjuvant treatments for ovarian cancer are universally platinum-based, patients with short PFS times (<6 months after the end of adjuvant therapy) are referred to as platinum resistant. As noted in the previous section, IPI59 accurately classifies patients into two PFS groups (low-risk vs. high-risk); in other words, IPI59 identifies patients who are at high-risk of developing platinum resistance. However, as the IPI59 values considered thus far are

derived from expression in the original tumor tissue, they provide no information on possible changes in this marker after chemotherapy; and the ability to assess the association between dynamic changes in IPI59 and the development of platinum resistance in patient populations is limited since it is not common to biopsy recurrent tumor tissue. Despite a small cohort, there are thirteen late-stage TCGA patients for which tumor tissue samples are available from both primary and recurrent tumors. Figure 3 shows PFS as a spline function of the percent change in IPI59 for these patients; the middle and right panels show PFS as a function of percent change in the 30 resistance and 29 susceptibility genes that make up the signature. The linear model (not shown) predicts that a 1 % increase in IPI59 decreases time to recurrence by 24 days ($p = 0.098$). These results suggest that platinum has a shorter duration of efficacy for patients with high baseline susceptibility genes and low baseline resistance genes (PFS is shorter), and that increases in IPI59 are associated with accelerated development of platinum resistance (increased tumor growth rate) as assessed by time to first recurrence.

3.4 IPI59 Stratifies Patients into Meaningful Groups that Show Differential Response to Treatment

Since patients with high IPI59 index are at high risk for early recurrence, we investigated whether enhanced adjuvant therapy for these patients provided benefit. Of the 458 TCGA patients considered, we focused on 411 for which treatment information was available. Of these, 321 received standard of care for adjuvant therapy (platinum and taxane drugs only) and 90 received enhanced adjuvant therapy (an agent or combination of agents in addition to platinum and taxane drugs). Between adjuvant therapy and enhanced adjuvant therapy groups, the baseline difference in terms of age was not significant (t test; $p = 0.60$). Since cytoreduction is widely used in clinical practice to identify high-risk patients, it too was considered in our assessment.

The left panel of Fig. 4 shows PFS for RD⁻ patients stratified by IPI59 and treatment (adjuvant vs. enhanced adjuvant). While there is a significant difference in PFS between low and high-risk patients as assessed by IPI59 ($p = 0.001$), there is no difference within risk group between patients who received adjuvant and those who received enhanced adjuvant therapy. The right panel of Fig. 4 shows that there is an effect for RD⁺ patients. Specifically, patients with sub-optimal cytoreduction with low-risk IPI59 show significant benefit from enhanced adjuvant therapy, with median survival increased by 11.7 months ($p = 0.085$). These results suggest that IPI59 can be combined with cytoreduction status to identify a sub-population of patients who may benefit from enhanced adjuvant therapy.

4 Discussion

Ovarian cancer remains one of the most challenging cancers, as diagnosis is often late and prognosis is often poor. Consequently, it is clear that major advances in early detection and treatment are needed. We have focused on the latter, namely, deriving a biomarker that may guide improved treatment/chemotherapy choices. While a number of expression-based biomarkers have been shown to be associated with OS and PFS in ovarian cancer patient populations, most provide little insight into the underlying biological mechanisms that in

part lead to differences in outcome, and none to date provide information to guide treatment. The biomarker derived here, IPI59, is both prognostic and predictive, and in this way has the potential to directly improve patient outcome.

The model was derived using TCGA data from approximately 500 patients, and then validated in over 900 patients from six independent studies. Importantly, the study populations are geographically diverse, with data collected over varying times and array platforms. The heterogeneity inherent to the validation data presents challenges for the approach, but more accurately reflects performance in a clinical setting.

Our results demonstrate that IPI59 provides for accurate ovarian cancer patient prognosis that significantly improves on the clinical variables in use today. Furthermore, we have demonstrated that patients with sub-optimal cytoreduction (RD⁺) with low-risk identified by IPI59 have significantly improved progression-free survival times if treated early on with an enhanced adjuvant therapy. This suggests that IPI59 may prove immediately useful in guiding treatment.

The results presented here also demonstrate that changes in IPI59 levels may serve as a biomarker of response. By analyzing IPI59 in patients for whom tumor tissue was available at both diagnosis and recurrence, we demonstrated that changes in IPI59 are significantly associated with earlier recurrence, which suggests that IPI59 may ultimately be useful in patient monitoring over time.

While IPI59 combines genes from six pathways and proves more accurate than any of the six in isolation, it is useful to consider results from individual pathways (Supplementary Figure S3–S8) as they may provide insight into underlying biological mechanisms, and in some cases provide similar results using fewer genes for at least some of the responses considered. The index from the KEGG: Neuroactive Ligand-Receptor Interaction pathway and that from the Wnt signaling pathway provide examples. With just 18 genes from the Neuroactive Ligand-Receptor Interaction pathway, both OS and PFS are well predicted (Fig. S4). Also, with 21 genes from the Wnt signaling pathway, patients with RD⁺ with low-risk show significant benefit from enhanced adjuvant therapy (median survival increased by 15.2 months; $p = 0.006$). In addition, since these genes are from an individual pathway, mechanistic studies are perhaps more straightforward.

In summary, we have used the IPI model to identify IPI59, a 59-gene mRNA-based prognostic and predictive biomarker to improve outcome in serous ovarian cancer patients. Like other prognostic biomarkers such as TCGA193 and CLOVAR100, IPI59 is useful for identifying patients at high-risk of early recurrence or death following surgery. However, unlike these other markers, by identifying a sub-population of patients who may benefit from augmented adjuvant therapy, IPI59 is also clinically actionable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Drs. Michael Newton and Ning Leng for suggestions that improved the manuscript. This research was supported by NIHGM102756, NIHU54 AI117924, NIHK01LM012100, and the Clinical and Translational Science Institute of Southeastern Wisconsin (NIH UL1RR031973). Results were generated in part using data from the Cancer Genome Atlas (TCGA) pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2013; 63:11–30. [PubMed: 23335087]
2. Tewari D, Java JJ, Salani R, et al. Long-term survival advantage and prognostic factors associated with intraperitoneal chemotherapy treatment in advanced ovarian cancer: a gynecologic oncology group study. *J Clin Oncol*. 2015; 33:1460–1466. [PubMed: 25800756]
3. 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. [PubMed: 18698038]
4. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*. 2006; 439:353–357. [PubMed: 16273092]
5. Mehta S, Shelling A, Muthukaruppan A, et al. Predictive and prognostic molecular markers for cancer medicine. *Ther Adv Med Oncol*. 2010; 2:125–148. [PubMed: 21789130]
6. Services H. GEO?: the gene expression omnibus. *Gene Expr*. 2007; 23:2–3.
7. Konstantinopoulos PA, Spentzos D, Karlan BY, et al. Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. *J Clin Oncol*. 2010; 28:3555–3561. [PubMed: 20547991]
8. Lisowska KM, Olbryt M, Dudaladava V, et al. Gene expression analysis in ovarian cancer: faults and hints from DNA microarray study. *Front Oncol*. 2014; 4:6. [PubMed: 24478986]
9. Ferriss JS, Kim Y, Duska L, et al. Multi-gene expression predictors of single drug responses to adjuvant chemotherapy in ovarian carcinoma: predicting platinum resistance. *PLoS One*. 2012; 7:e30550. [PubMed: 22348014]
10. 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. [PubMed: 18593951]
11. Yoshihara K, Tajima A, Yahata T, et al. Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS One*. 2010; 5:e9615. [PubMed: 20300634]
12. Ganzfried BF, Riester M, Haibe-Kains B, et al. curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. *Database (Oxford)*. 2013
13. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4:249–264. [PubMed: 12925520]
14. Tucker SL, Gharpure K, Herbrich SM, et al. Molecular biomarkers of residual disease after surgical debulking of high-grade serous ovarian cancer. *Clin Cancer Res*. 2014; 20:3280–3288. [PubMed: 24756370]
15. Chang S-J, Bristow RE, Ryu H-S. Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. *Ann Surg Oncol*. 2012; 19:4059–4067. [PubMed: 22766983]
16. 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 Ovarialkarzin. *Cancer*. 2009; 115:1234–1244. [PubMed: 19189349]
17. Wang S, Nan B, Zhu J, Beer DG. Doubly penalized Buckley–James method for survival data with high-dimensional covariates. *Biometrics*. 2008; 64:132–140. [PubMed: 17680828]

18. Eng KH, Wang S, Bradley WH, et al. Pathway index models for construction of patient-specific risk profiles. *Stat Med*. 2013; 32:1524–1535. [PubMed: 23074142]
19. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol*. 2004; 2:E108. [PubMed: 15094809]
20. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*. 2005; 21:3001–3008. [PubMed: 15814556]
21. Ishwaran H, Kogalur UB, Gorodeski EZ, et al. High-dimensional variable selection for survival data. *J Am Stat Assoc*. 2010; 105:205–217.
22. Ma S, Huang J. Clustering threshold gradient descent regularization: With applications to microarray studies. *Bioinformatics*. 2007; 23:466–472. [PubMed: 17182700]
23. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Ser B Stat Methodol*. 2006; 68:49–67.
24. Zhao P, Rocha G, Yu B. The composite absolute penalties family for grouped and hierarchical variable selection. *Ann Stat*. 2009; 37:3468–3497.
25. Luan Y, Li H. Group additive regression models for genomic data analysis. *Biostatistics*. 2008; 9:100–113. [PubMed: 17513311]
26. Gu Z, Wang J. CePa: an R package for finding significant pathways weighted by multiple network centralities. *Bioinformatics*. 2013; 29:658–660. [PubMed: 23314125]
27. Wang S, Nan B, Zhu N, Zhu J. Hierarchically penalized Cox regression with grouped variables. *Biometrika*. 2009; 96:307–322.
28. Simon N, Friedman J. A sparse-group lasso. *J Comput Graph Stat*. 2013; 22:231–245.
29. Kanehisa M, Goto S. KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucl Acids Res*. 2000; 28:27–30. [PubMed: 10592173]
30. Kanehisa M, Goto S. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011; 474:609–615. [PubMed: 21720365]
31. Verhaak R, Tamayo P. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Investig*. 2012; 123:1–9.

Fig. 1.

Kaplan–Meier curves showing overall survival for 825 patients stratified into four groups defined by cytoreduction status (RD⁻ vs. RD⁺) and prognosis (low-risk vs. high-risk) from IPI59 (*left*), TCGA193 (*middle*), and CLOVAR100 (*right*)

Fig. 2.

Kaplan–Meier curves showing progression-free survival for 366 patients stratified into four groups defined by cytoreduction status (RD⁻ vs. RD⁺) and prognosis (low-risk vs. high-risk) from IPI59 (*left*), TCGA193 (*middle*), and CLOVAR100 (*right*)

Fig. 3. Progression-free survival for 13 TCGA patients as a function of percent change in IPI59. For these 13 patients, changes in IPI59 were calculated from expression profiled in tumor tissue at diagnosis and at recurrence

Fig. 4.

Kaplan–Meier curves showing progression-free survival for 72 TCGA RD⁻ (*left*) and 304 RD⁺ (*right*) patients stratified into four groups defined by prognosis by IPI59 (low-risk vs. high-risk) and treatment group (adjuvant therapy vs. enhanced adjuvant therapy)

Table 1

Clinical characteristics of ovarian cancer patients in TCGA and six validation data sets

Characteristics	TCGA	U133plus2				U133a		Agilent
		Tothill	Konstantinopoulos	Lisowska	Ferriss	Bonome	Yoshihara	
No. of patients	520	285	28	101	58	185	260	
Median age	59	59	65	NA	NA	63	NA	
Tumor stage								
Stage I	15	24	0	0	0	0	0	
Stage II	25	18	2	2	0	0	0	
Stage III	396	217	22	64	53	146	204	
Stage IV	80	22	3	11	5	36	56	
Tumor grade								
G1	5	19	0	0	2	NA	0	
G2	64	97	0	10	19	NA	131	
G3	439	164	28	51	33	NA	129	
G4	1	0	0	19	0	NA	0	
Debulking status								
Opt (<1cm)	102	160	4	15	26	90	103	
Sub-opt (≥1cm)	359	70	23	60	30	95	157	
Vital status								
Alive	235	169	11	9	22	56	139	
Dead	284	113	17	66	36	129	121	
Recurrent								
Non-Rec	212	102	7	68	6	32	NA	
Recurrent	308	180	13	7	48	153		

Numbers do not sum to the total due to missing values. Excluded patients are described in supplement