



Article

Cancer Stem Cell Markers—Clinical Relevance and Prognostic Value in High-Grade Serous Ovarian Cancer (HGSOC) Based on The Cancer Genome Atlas Analysis

Natalia Iżycka ^{1,*}, Mikołaj Piotr Zaborowski ^{1,2}, Łukasz Ciecierski ², Kamila Jaz ¹, Sebastian Szubert ¹, Cezary Miedziarek ¹, Marta Rezler ¹, Kinga Piątek-Bajan ¹, Aneta Synakiewicz ¹, Anna Jankowska ³, Marek Figlerowicz ², Karolina Sterzyńska ^{4,†} and Ewa Nowak-Markwitz ^{1,†}

- ¹ Department of Gynecology, Obstetrics and Gynecologic Oncology, Division of Gynecologic Oncology, Poznan University of Medical Sciences, Polna 33 St., 60-535 Poznan, Poland; sszubert@ump.edu.pl (S.S.)
² European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; marek.figlerowicz@ibch.poznan.pl (M.F.)
³ Department of Cell Biology, Poznan University of Medical Sciences, Rokietnicka 5D St., 60-806 Poznan, Poland; ajanko@ump.edu.pl
⁴ Department of Histology and Embryology, Poznan University of Medical Sciences, Swieczkiego 6 St., 61-781 Poznan, Poland
* Correspondence: nizycka@ump.edu.pl
† These authors contributed equally to this work.



Citation: Iżycka, N.; Zaborowski, M.P.; Ciecierski, Ł.; Jaz, K.; Szubert, S.; Miedziarek, C.; Rezler, M.; Piątek-Bajan, K.; Synakiewicz, A.; Jankowska, A.; et al. Cancer Stem Cell Markers—Clinical Relevance and Prognostic Value in High-Grade Serous Ovarian Cancer (HGSOC) Based on The Cancer Genome Atlas Analysis. *Int. J. Mol. Sci.* **2023**, *24*, 12746. <https://doi.org/10.3390/ijms241612746>

Academic Editor: Alvaro Galli

Received: 5 July 2023

Revised: 5 August 2023

Accepted: 9 August 2023

Published: 13 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cancer stem cells (CSCs) may contribute to an increased risk of recurrence in ovarian cancer (OC). Further research is needed to identify associations between CSC markers and OC patients' clinical outcomes with greater certainty. If they prove to be correct, in the future, the CSC markers can be used to help predict survival and indicate new therapeutic targets. This study aimed to determine the CSC markers at mRNA and protein levels and their association with clinical presentation, outcome, and risk of recurrence in HGSOC (High-Grade Serous Ovarian Cancer). TCGA (The Cancer Genome Atlas) database with 558 ovarian cancer tumor samples was used for the evaluation of 13 CSC markers (ALDH1A1, CD44, EPCAM, KIT, LGR5, NES, NOTCH3, POU5F1, PROM1, PTTG1, ROR1, SOX9, and THY1). Data on mRNA and protein levels assessed by microarray and mass spectrometry were retrieved from TCGA. Models to predict chemotherapy response and survival were built using multiple variables, including epidemiological data, expression levels, and machine learning methodology. ALDH1A1 and LGR5 mRNA expressions indicated a higher platinum sensitivity ($p = 3.50 \times 10^{-3}$; $p = 0.01$, respectively). POU5F1 mRNA expression marked platinum-resistant tumors ($p = 9.43 \times 10^{-3}$). CD44 and EPCAM mRNA expression correlated with longer overall survival (OS) ($p = 0.043$; $p = 0.039$, respectively). THY1 mRNA and protein levels were associated with worse OS ($p = 0.019$; $p = 0.015$, respectively). Disease-free survival (DFS) was positively affected by EPCAM ($p = 0.004$), LGR5 ($p = 0.018$), and CD44 ($p = 0.012$). In the multivariate model based on CSC marker expression, the high-risk group had 9.1 months longer median overall survival than the low-risk group ($p < 0.001$). ALDH1A1, CD44, EPCAM, LGR5, POU5F1, and THY1 levels in OC may be used as prognostic factors for the primary outcome and help predict the treatment response.

Keywords: epithelial ovarian cancer; HGSOC; TCGA; ovarian cancer prognosis; cancer stem cells; markers; multi-omics; multi-omics data

1. Introduction

High-grade serous ovarian cancer (HGSOC) is a heterogenous malignancy and one of the leading causes of cancer-related deaths in women worldwide [1].

The cancer cell populations derived from a single patient are highly diverse [2,3]. Cancer stem cells (CSCs), the subpopulation of neoplastic cancer cells able to self-renew [4,5],

contribute to chemotherapy resistance [2,5–7]. CSCs have been detected in OC tumors in the early and advanced stages before and after the conventional treatment [8–10]. The presence of CSCs appears to be a feature of all types of OC, including the most heterogeneous type—primary HGSOC— independent of its molecular subtypes (mesenchymal, immunoreactive, differentiated, and proliferative) [9,11–14].

Despite treatment, around 70% of patients experience cancer recurrence [5]. Due to the potential for self-renewal after a long dormancy, CSCs are perceived as the origin of recurrent tumors. Surgery and chemotherapy may eradicate most cancer cells but not the residual population of CSCs [1]. Their self-protection is caused mainly by the slow cell cycle progression, resulting in decelerated metabolism [2,5], which may induce multiple drug resistance mechanisms [3,15]. The persistence of CSCs in tumor tissue and their drug resistance acquired during primary treatment are important factors in tumor relapse after chemotherapy [3,5]. The presence of CSCs has been associated with a worse prognosis in OC [5].

Thus, CSC markers may be valuable prognostic tools in clinical practice [2].

Several CSC markers have been suggested to date. The most often proposed and best-characterized are CD44, CD24, CD117, CD133, and ALDH1A [6]. Their main functions and roles as CSC markers are described in the table below.

In this study, we aimed to determine the relationships between the clinical features of HGSOC and the expression of selected CSC markers. Besides the well-established CSC markers described above (CD44, CD24, CD117, CD133, and ALDH1A), other proteins might belong to this population. To identify them, we searched for the proteins that were previously mentioned in the literature as potentially related to CSCs in various malignancies, including OC (Table 1). We retrieved proteomic and transcriptomic data and clinical characterizations of 551 HGSOC patients from The Cancer Genome Atlas Database (TCGA). We evaluated the expression levels of the above-defined set of CSC markers (Table 1) in HGSOC (Figure 1). Next, we investigated the relationship of the CSC marker levels with patients' clinical and pathological features, such as clinical stage, grade of the disease, size of residual disease, following surgery, and responsiveness to chemotherapy. Finally, we assessed the relationship between the analyzed markers and the time to progression and OS (Overall Survival).

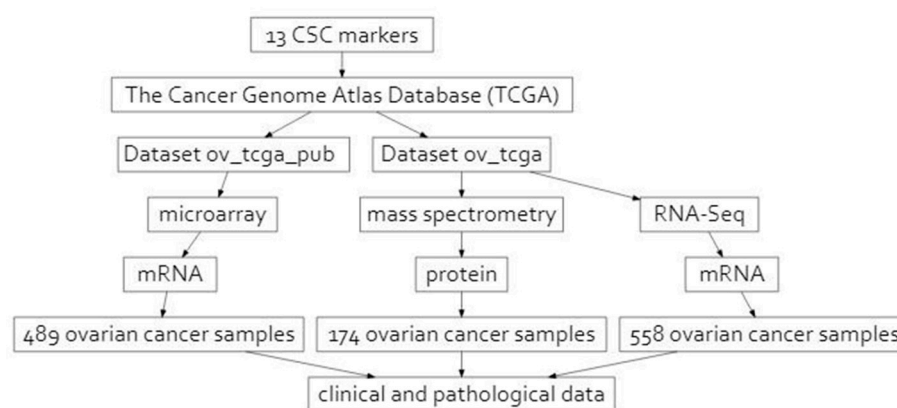


Figure 1. Schematic overview of the study design.

Table 1. CSC markers selected based on a review of the literature. The function and potential contribution to the phenotype of cancer stem cells are described following the given reference.

CSC Marker	Full Name	Function	References
ALDH1A1	Aldehyde Dehydrogenase 1 Family Member A1	Enzyme belonging to the aldehyde dehydrogenase family of proteins participating in the biosynthesis of retinoic acid and allowing for the regulation of proper proliferation and differentiation of cancer stem cells.	[16]

Table 1. *Cont.*

CSC Marker	Full Name	Function	References
CD44	CD44 molecule	A cell-surface glycoprotein that promotes metastasis, stem cell-like phenotypes, and chemoresistance.	[17]
PROM1 (CD133)	Prominin-1	Transmembrane protein that promotes stemness and strengthens adhesion and clearance of mesothelial cells; it causes an increase in peritoneal adhesion. It entails improved adherence to the metastatic niche and infiltration of the peritoneal tissue during metastases.	[18,19]
c-Kit (CD117)	KIT proto-oncogene receptor tyrosine kinase	This cytokine receptor is closely related to the neovascularization of epithelial OC tissue and the formation of vasculogenic mimicry. CSCs are associated with the formation of vasculogenic mimicry, which therefore acts as a molecular CSC marker.	[20]
POU5F1 (OCT4)	POU class 5 homeobox 1 (octamer-binding transcription factor 4)	It encodes embryonic transcription factors that are vital for quiescence, pluripotency, and long-term self-renewal—properties that are characteristic of CSCs.	[21]
NOTCH3	Notch Receptor 3	A protein whose activation increases the adhesion between ovarian tumor cells and collagen-rich peritoneal surfaces.	[22]
SOX9	SRY-box transcription factor 9	A protein that regulates apoptotic and proliferative properties. It allows OC to survive in hypoxic conditions.	[23]
SNORD89	Small nucleolar RNA, C/D box 89	A small nucleolar RNA that promotes stem-cell-like characteristics via the Notch1/c-Myc pathway.	[23–25]
SNORA72	Small nucleolar RNA, H/ACA box 72		
Thy-1 (CD90)	Thy-1 cell surface antigen	GPI-anchored protein located on the cell surface correlated with increased self-renewal and proliferative ability of OC cells. It acts as a tumor suppressor in OC. Thy-1 is overexpressed in CSCs.	[23–25]
EpCAM	Epithelial cell adhesion molecule	Transmembrane glycoprotein mediating Ca ²⁺ -independent homotypic cell–cell adhesion in epithelia. EpCAM regulates chemoresistance by activating the PI3K/Akt/mTOR signaling pathway.	[26]
LGR5	Leucine-rich repeat containing G protein-coupled receptor 5	Protein that may promote epithelial OC development through regulation of the Notch1 signaling pathway, which is associated with CSC self-renewal and drug resistance.	[27]
PTTG1 (Securin)	PTTG1 regulator of sister chromatid separation, securin	Protein with the ability to regulate CSC-associated self-renewal and epithelial–mesenchymal transition pathways.	[28]
ROR1	Receptor tyrosine kinase like orphan receptor 1	A surface antigen playing an important role in the Wnt signaling pathway. ROR1 increases tumor cell proliferation, migration, invasion, and oncogenicity and triggers the formation of spheroids, the invasion of the extracellular matrix, or the development of tumor xenografts, which are functional features associated with CSC.	[29,30]
NES (Nestin)	Neuroepithelial stem cell protein	Intermediate filament protein involved in neovascularization and CSCs and closely related to vasculogenic mimicry formation.	[31]
TMSB4X (Tβ4)	Thymosin β4	A G-actin-sequestering peptide associated with the metastatic potential of tumor cells by stimulating cell migration. Tβ4 expression is strongly associated with CD133 expression and is characteristic of CSCs.	[32]

2. Results

There were 489 patients with complete data for mRNA measured by microarray for 13 CSC markers: ALDH1A1, CD44, EPCAM, KIT, LGR5, NES, NOTCH3, POU5F1, PROM1, PTTG1, ROR1, SOX9, and THY1. Among the analyzed CSC markers, the mass spectrometry data for five proteins (EPCAM, ALDH1A1, CD44, NES, and THY1) were available for 174 patients.

2.1. Clinical and Pathological Features

The characteristics of the studied group correspond well to the mean clinical features observed in OC. All the analyzed samples were retrieved from patients suffering from HGSOC. Most patients (71%) in the database labeled “ov_tcga” were at stage IIIC. The mean age at diagnosis was 59.5–60 (interquartile range (IQR): 51–68 and 51–71 for mRNA and protein samples, respectively). Most OC tissues were characterized as grade 3 (mRNA samples = 82.3%, protein samples = 81.6%). There were no grade 1 tumors.

The expression of two CSC marker genes was associated with the tumor grade. We found that EPCAM and PTTG1 mRNA levels were higher in grade 3 as compared to grade 2 tumors (“ov_tcga” dataset, $n = 537$, $p = 0.02$; $p = 0.03$, respectively) (Figures A1A and A1B, respectively).

The expression of two CSC marker genes was associated with FIGO staging. We found that the CD44 mRNA level was elevated in tumors with FIGO I/II as compared to FIGO III/IV stages (“ov_tcga” dataset, $n = 552$, $p = 0.02$, Figure 2A). At the same time, THY1 mRNA expression was higher in the FIGO III and FIGO IV stages (“ov_tcga” dataset, $n = 552$, $p = 1.43 \times 10^{-4}$) (Figure 2B).

Furthermore, the expression of CSC genes was linked to the tumor’s platinum sensitivity. We found that higher ALDH1A1 and LGR5 mRNA levels have indicated platinum sensitivity (“ov_tcga_pub” dataset, $n = 287$, $p = 3.50 \times 10^{-3}$; $p = 0.01$, respectively) (Figures 2C and 2D, respectively). Conversely, mRNA of POU5F1 showed a higher level in platinum-resistant tumors (“ov_tcga_pub” dataset, $n = 287$, $p = 9.43 \times 10^{-3}$) (Figure 2E).

We also demonstrated that level of one of the CSC markers indicates a chance for complete cytoreduction. The EPCAM level (mRNA) was higher in tumors with no macroscopic residual disease (“ov_tcga_pub” dataset, $n = 179$, $p = 0.03$) (Figure 2F).

2.2. CSC Expression Profiles

Four molecular subtypes of OC are distinguished based on the gene expression pattern [9]. We found that specific CSC markers were linked to OC expression profiles (Figure 3A). High amounts of EPCAM, NES, NOTCH3, ROR1, and LGR5 transcripts were associated with proliferative profile (“ov_tcga_pub” dataset, Figure A2A–D and Figure 3B, respectively). The mesenchymal profile was characterized by upregulation in ALDH1A1, KIT, and THY1 mRNA (Figure 3C, Figure A2E, and Figure 3D, respectively). The immunoreactive profile was marked by high CD44 and PTTG1 expression (Figures 3E and A2F, respectively). POU5F1, PROM1, and SOX9 gene expressions were more prevalent in the fallopian profile (Figure A2G, Figure 3F, and Figure A2H, respectively).

2.3. Overall Survival

We found that the expression of CSC marker genes was associated with OS. In the microarray-assessed dataset, CD44 mRNA levels were associated with longer OS (dataset “ov_tcga_pub”, HR = 0.88, 95%, 0.79–1.00; $p = 0.043$) (Figure 4A). The beneficial effect was also observed in the “ov_tcga” dataset, where EPCAM mRNA expression predicted longer OS (HR = 0.89, 95%, 0.80–0.99; $p = 0.039$, Figure 4B), while THY1 mRNA level was associated with poorer OS (HR = 1.14, 95%, 1.02–1.28; $p = 0.019$) (Figure 4B). Consistently, the level of THY1 protein was also associated with shortened OS (“ov_tcga” protein dataset, HR = 1.26, 95%, 1.05–1.51; $p = 0.015$) (Figure 4C).

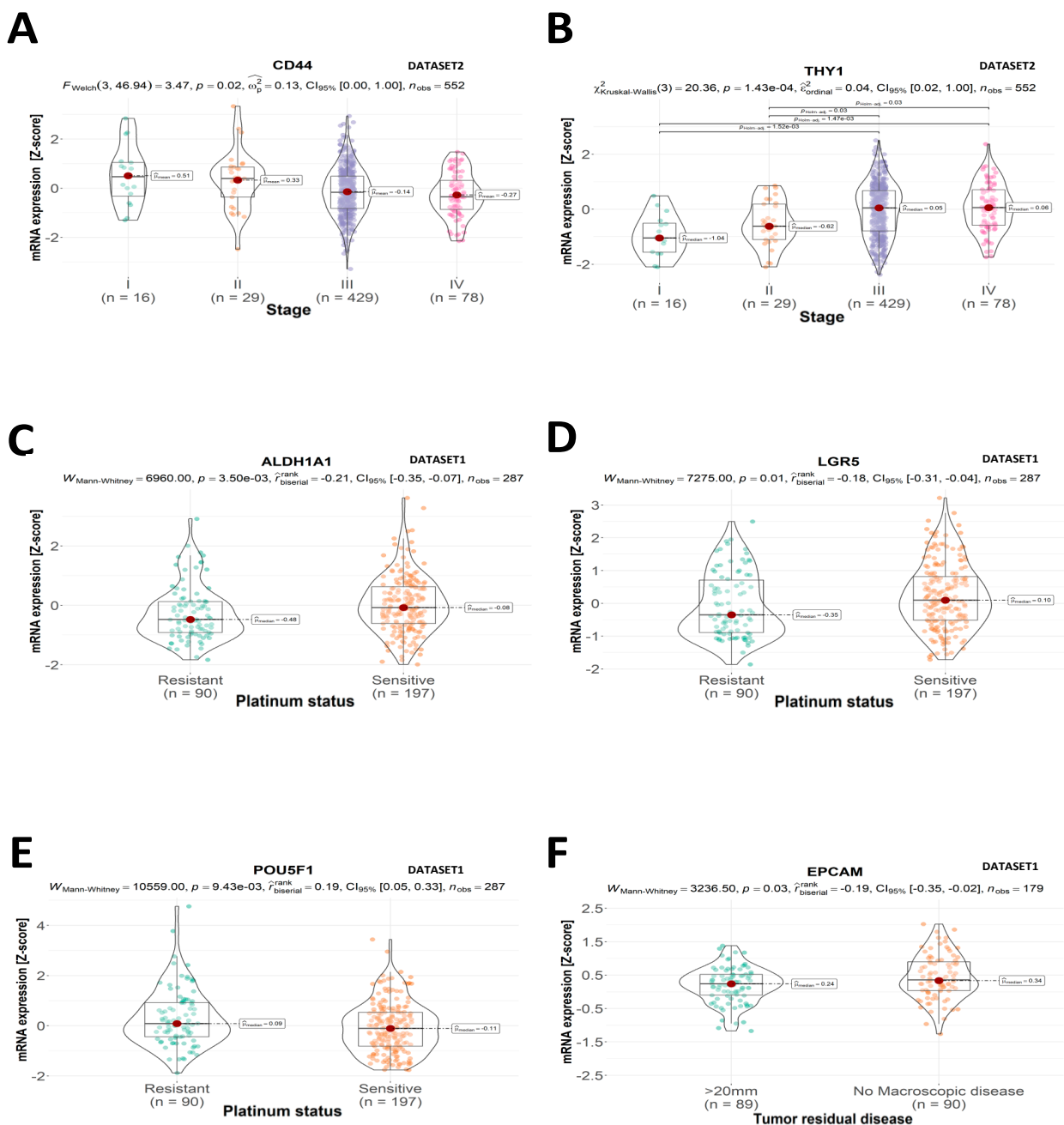


Figure 2. Clinical and pathological features of OC tumors. The expression of CSC markers depending on (A,B) tumor stage ((A) Welch test; (B) Kruskal–Wallis test), responsiveness to platinum compounds ((C–E); Mann–Whitney test), and tumor residual disease following cytoreductive surgery ((F); Mann–Whitney test). Values on the Y axis represent mRNA expression as Z scores. Statistical significance and the of statistics used for analysis are described at the top of each plot. The total number of samples is described at the top of each plot. The number of samples in each subgroup is given as “N” (“ov_tcga_pub”—DATASET1 and “ov_tcga”—DATASET2).

2.4. Disease-Free Survival

The expression of the EPCAM gene was associated with longer DFS (“ov_tcga_pub” dataset, HR = 0.87, 95%, 0.79–0.96; $p = 0.004$), LGR5 (HR = 0.86 95%, 0.77–0.97; $p = 0.018$) and CD44 (HR = 0.86, 95%, 0.77–0.97; $p = 0.012$) (Figure 4D). Consistently, a higher amount of CD44 transcripts was associated with longer DFS (dataset “ov_tcga”, HR = 0.87, 95%, 0.80–1.01; $p = 0.013$) (Figure 4E). In contrast, THY1 expression predicted reduced DFS

(HR = 1.19, 95%, 1.07–1.33; $p = 0.002$) (Figure 4E). In agreement, the levels of THY1 (“ov_tcga” dataset, HR = 1.32, 95%, 1.09–1.58; $p = 0.004$) and NES (“ov_tcga” dataset, HR = 1.19, 95%, 1.01–1.41; $p = 0.039$) proteins were associated with poorer DFS (Figure 4F).

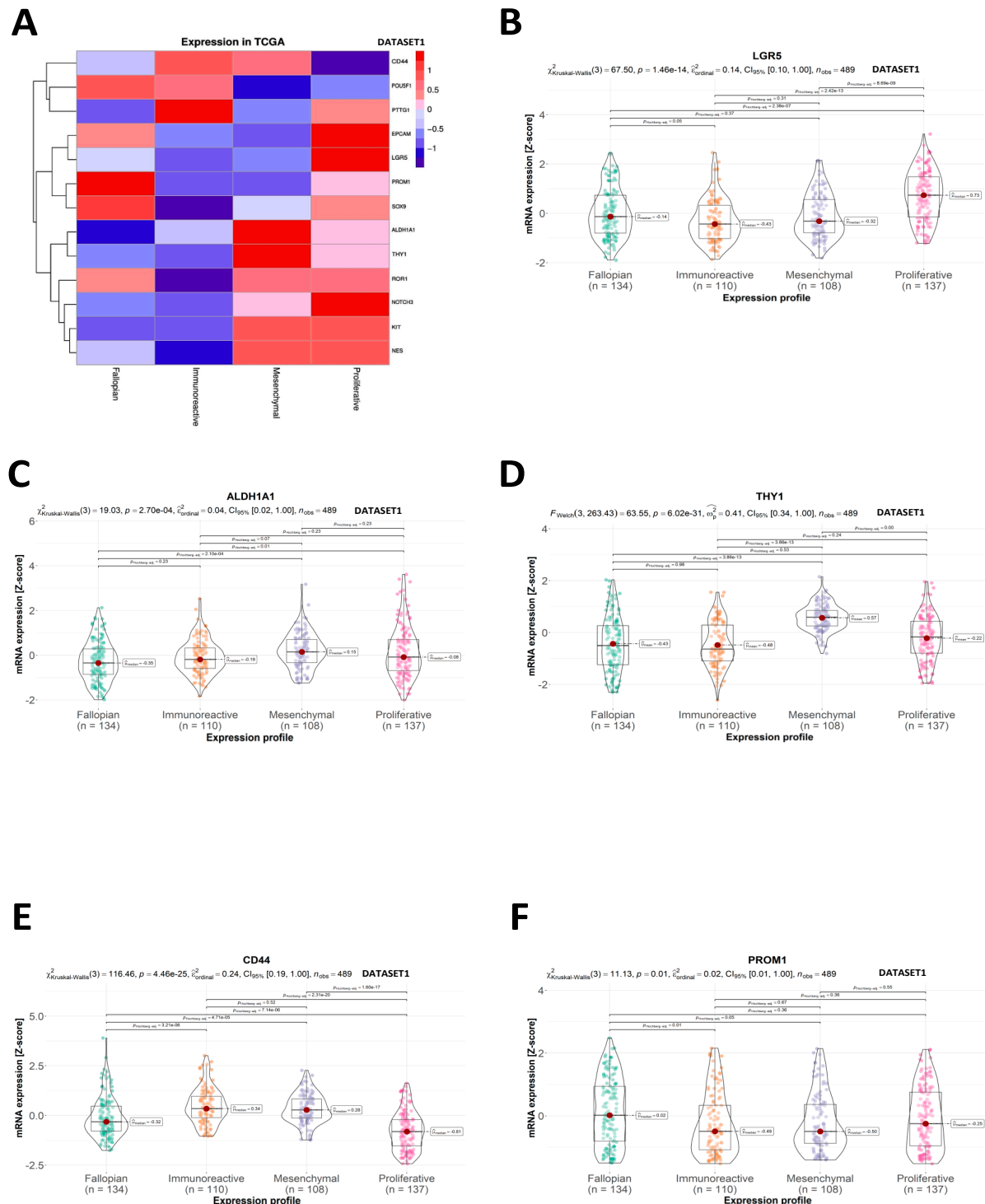


Figure 3. OC expression profiles (“ov_tcga_pub”—DATASET1 and “ov_tcga”—DATASET2). (A) normalized expression of CSC markers depending on molecular subtypes of OC (mesenchymal, proliferative, fallopian, and immunoreactive). The colors represent the intensity of expression according to the scale in the top right of the figure. Markers are clustered in order of similarity across their subtypes. (B–F) mRNA levels represented as Z scores in subgroups according to OC expression profiles. Statistics used: (B,C,E,F)—Kruskal–Wallis test; (D)—Welch test.

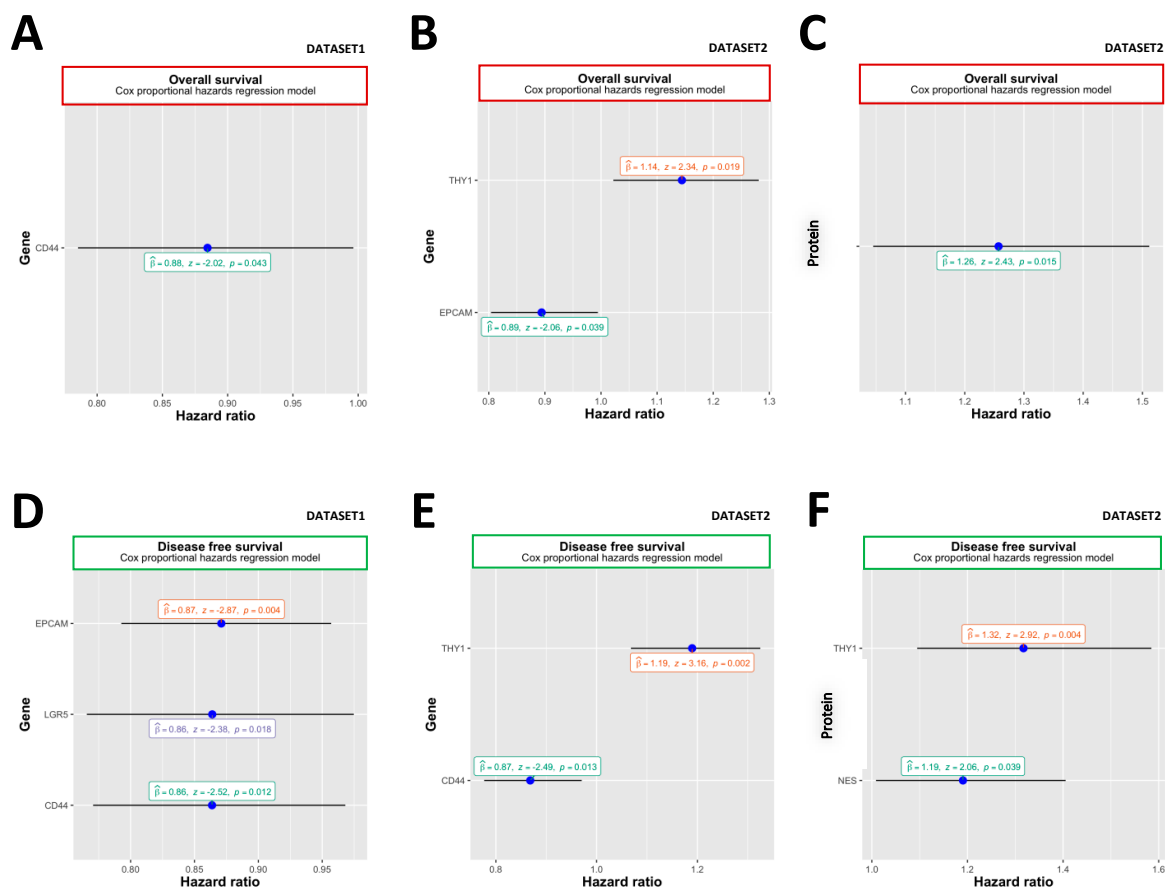


Figure 4. The OS and DFS (“ov_tcga” and “ov_tcga_pub” datasets). Relationship between CSC markers and patient survival. Association between mRNA (miRNA) expression of CSC markers and (A,B) OS, as well as between DFS (D,E) based on the “ov_tcga_pub” and “ov_tcga” datasets. Association between protein expression of CSC markers (mass spectrometry) and (C) OS and (F) DFS based on “ov_tcga” datasets. Panels represent only statistically significant ($p < 0.05$) relationships. Beta coefficient and p values of univariate Cox proportional hazard regression models are given in the panels.

2.5. Correlation Analysis

In both datasets, we found significant positive correlations between ALDH1A1 and KIT mRNA levels ($r = 0.34$; $r = 0.36$, $n = 489$; $n = 558$, $p < 0.05$, “ov_tcga_pub” and “ov_tcga” datasets, respectively; Figures A1C and A1D, respectively) and between ROR1 and SOX9 ($r = 0.31$; $r = 0.31$, $n = 489$; $n = 558$, $p < 0.05$, “ov_tcga_pub” and “ov_tcga” datasets, respectively (Figures A1C and A1D, respectively). Furthermore, in the “ov_tcga_pub” dataset, NOTCH3 expression correlated well with NES (“ov_tcga_pub”, $r = 0.35$, $n = 489$, $p < 0.05$, Figure A1C).

We found a negative correlation between the EPCAM and THY1 proteins and between the EPCAM and CD44 proteins (dataset “ov_tcga”, $r = -0.32$; $r = -0.35$, $n = 174$, $p < 0.05$) (Figure A1E). We also demonstrated that the THY1 protein level was increased in association with NES protein (“ov_tcga”, $r = 0.31$, $n = 174$, $p < 0.05$) (Figure A1E).

2.6. Multivariate Predictive Analysis

To verify the collectively predictive power of 13 CSC markers with respect to the carboplatin treatment outcome and patients survival, we performed multivariate predictive analysis using machine learning (ML) methods. The importance of the CSC markers for the ML models’ prediction was assessed jointly with the clinical and pathological features. We

estimated the performance of each ML model to verify whether they learn from the data that might be useful for prospective studies.

For the predictive analysis of carboplatin treatment outcome, the random forest classifier performance on the “ov_tcga_pub” microarray dataset ($n = 287$) was estimated using the repeated fivefold cross-validation method (CV) and the AUC (area under the ROC curve) score. The expected performance of the model is significantly higher than the efficacy of a random guess ($t\text{-test}(\mu_0 = 0.5) = 22.964$, $p < 0.001$), with a median estimated AUC = 0.61, 95%, 0.601–0.63 (Figure 5A). We also calculated additional model performance scores with the following median values: accuracy = 0.71, 95%, 0.702–0.714, F1 = 0.53, 95%, 0.525–0.545, and MCC = 0.28, 95%, 0.266–0.293 (Figure A3).

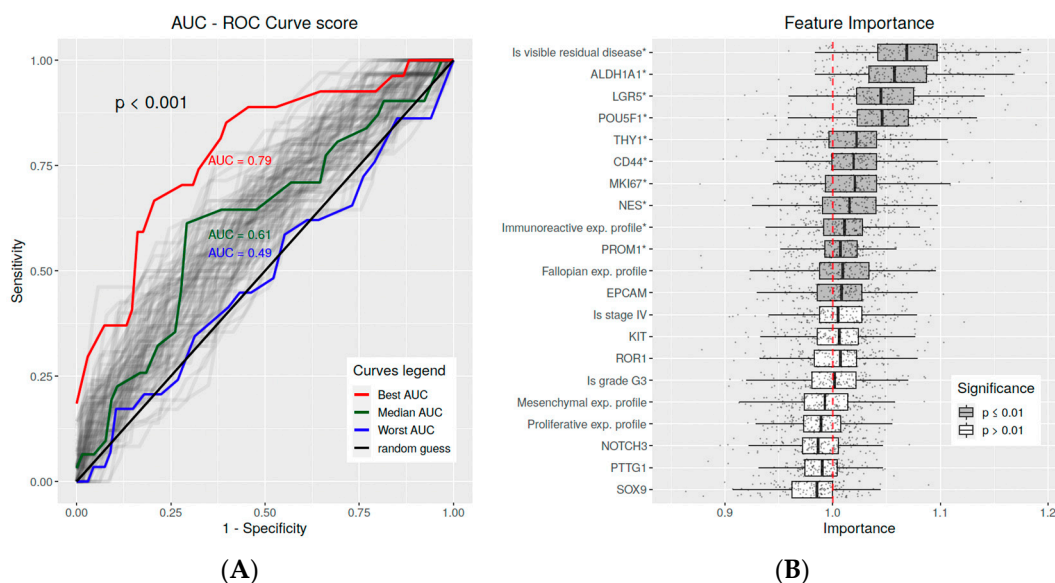


Figure 5. Results of the carboplatin therapy outcome predictive analysis. **(A)** Expected outcome prediction performance of the random forest classifier trained on the “ov_tcga_pub” dataset. The performance was estimated using repeated fivefold cross-validation (CV) and the AUC score. The median AUC = 0.61 95%, 0.601–0.63, which is greater than a random guess with $t\text{-test}(\mu_0 = 0.5) = 22.087$, $p < 0.001$. **(B)** Importance of the “ov_tcga_pub” dataset features for therapy outcome classification was estimated using the permutation feature importance score measured using the AUC score and the classifiers obtained from the CV. (*) The significance of the importance measurements was tested with a one-sample Wilcoxon test under the alternative hypothesis, i.e., that the importance mean > 1.0 , $p \leq 0.001$.

For the DFS and OS analysis, we estimated the performance of Cox proportional hazards models trained on the “ov_tcga_pub” microarray dataset ($n = 487$) with ridge regularization. The performance was assessed using the c-index score measured during repeated fivefold cross validation. The median equated c index = 0.59, 95%, 0.585–0.59 ($t\text{-test}(\mu_0 = 0.5) = 73.94$, $p < 0.001$) for the DFS, and c index = 0.56, 95%, 0.555–0.56 ($t\text{-test}(\mu_0 = 0.5) = 36.76$, $p < 0.001$) for OS (Figure A4). Additionally, the stratification capabilities of the models were estimated by dividing patients into low- and high-risk groups using the calculated median hazard risks of recurrence or death of the patients based on CSC markers. For DFS, the low- and high-risk groups were determined to be significantly divergent ($p < 0.002$), with 4.72 months of difference between the times of the strata median survival probabilities. In the case of OS, the low- and high-risk groups were separated ($p < 0.001$) by 9.12 months in median survival times (Figure 6A).

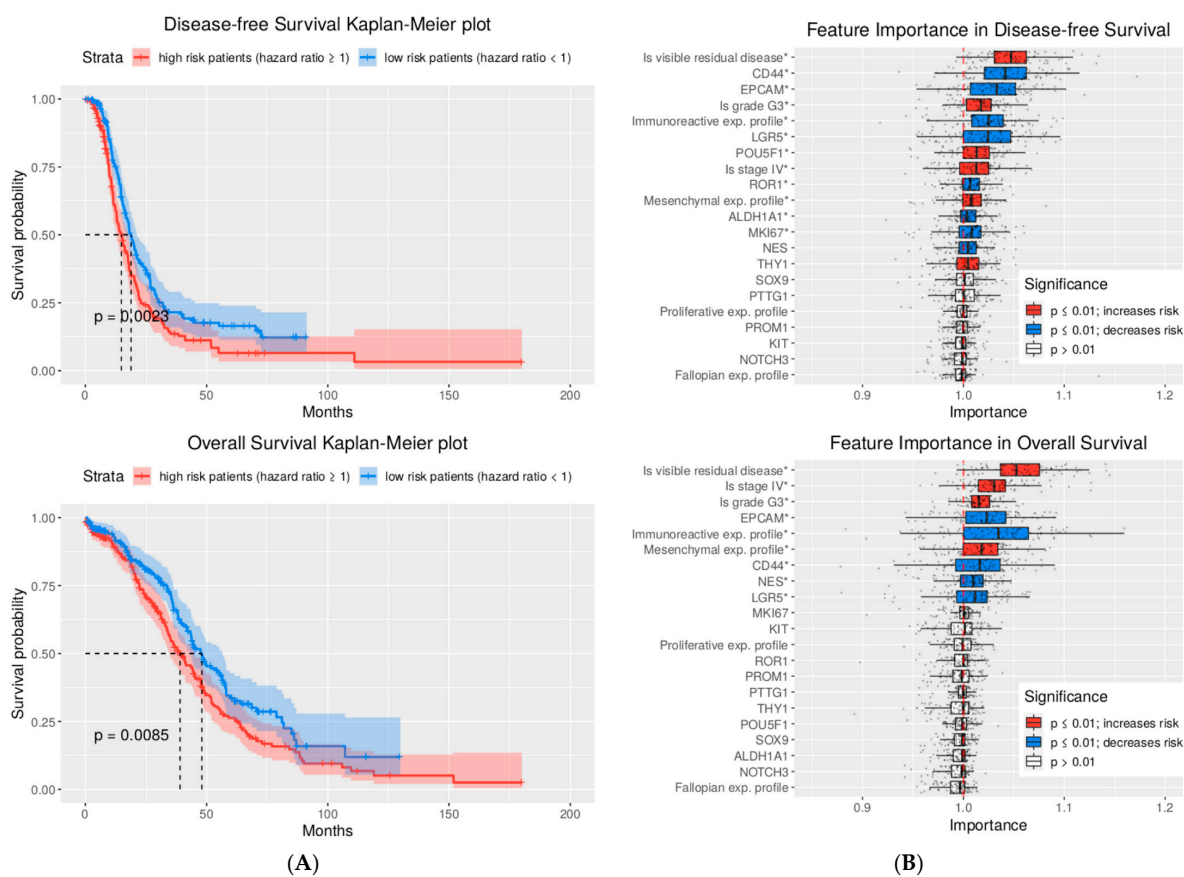


Figure 6. Results of disease-free and overall survival multivariate analyses. (A) Expected stratification capability of the Cox proportional hazards models with ridge regularization fitted to the “ov_tcga_pub” dataset. The stratification performance was estimated using repeated fivefold cross-validation (CV) and c-index score. The null hypothesis of the log-rank test that the low-risk and high-risk strata have identical hazard functions was rejected, with $p < 0.0023$ and $p < 0.0085$ for the disease-free and overall survival data, respectively. (B) Importance of the “ov_tcga_pub” dataset features for disease-free and overall survival hazards risk prediction was estimated using the permutation feature importance score measured using the c-index score and the models obtained from the CV. The significance of the importance measurements was tested with a one-sample Wilcoxon test under the alternative hypothesis, i.e., that the importance mean > 1.0 . (*) The asterisk indicates features for which an estimated importance mean is > 1.0 with $p \leq 0.001$.

Knowing that all the trained models are expected to be relevant, we assessed the importance of the “ov_tcga_pub” dataset features using the permutation feature importance (PFI) score (Figures 5B and 6B). The most important predictive factor of the carboplatin therapy outcome, DFS, and OS was the presence of residual disease, which increases the risk of both disease recurrence and patient death ($p < 0.001$). Another significant pathological feature was the immunoreactive profile of gene expression, which was also meaningful for carboplatin outcome prediction and decreased the risks for both OS and DFS. The pathological features important for predicting the DFS and OS hazard risks included mesenchymal expression profile of gene expression, FIGO stage IV, tumor grade 3, and MKI67 proliferation marker expression ($p < 0.001$).

We subsequently analyzed the PFI scores of the 13 considered CSC markers. We report that the CD44, LGR5, NES, and EPCAM CSC markers were significantly important outcome predictors ($p < 0.01$) in all multivariate analyses. Also, the POU5F1, THY-1, and ALDH1A1 markers were determined as important for both the carboplatin therapy outcome binary classification and hazards risks assessment. Interestingly, only the POU5F1 and THY1

markers were defined as the risk-increasing factors of DFS. The rest of the listed markers were described as risk-decreasing factors.

3. Discussion

Among all the analyzed CSC markers, our study shows that the expression of six genes—CD44, ALDH1A1, EpCAM, THY-1, POU5F1, and LGR5—is significant in OC in terms of clinical outcome, including stage and grade of the disease, as well as the platinum sensitivity and patient survival.

We found that the expression of the CD44 gene was associated with lower FIGO stage and, hence, better OS and DFS of ovarian cancer patients. Moreover, we observed higher levels of the CD44 gene in the immunoreactive subtype of OC. This type is characterized by a more favorable prognosis and extensive intratumoral T-cell infiltration [17]. In fact, CD 44 might play an essential role in regulating the lymphocyte infiltration process in OC tumor tissue, as it is a molecule known for its role in lymphocyte activation and homing [33]. Our results are consistent with those of Sillanpää S et al. [34] and Sosulski et al. [35] and show the association between CD44 gene expression, lower FIGO stage, and better survival times of OC patients. Still, the role played by CD44 in carcinogenesis is more complicated. In some reports, the increase in CD44 gene expression was indicated as a prognostic factor of both shorter OS and DFS of OC patients. Its correlation with higher FIGO stage and grade was also demonstrated [36,37]. The observed differences in CD44 impact on OC biology might be caused by many mRNA splice variants of CD44 and their various effects on the tumor characteristics [35,37]. As a cell-surface glycoprotein, it undergoes numerous post-translational modifications, which may distort its initial abundance and function.

Our results based on the analysis of ALDH1A1 gene expression in OC show that it is highest in the mesenchymal type of OC, which is considered the most aggressive subtype [38]. A negative correlation was demonstrated between an increased number of ALDH1A1-positive cells and patient survival [39,40]. Given many discordant reports, a meta-analysis of ALDH1A function was performed by Ruscito et al., revealing that high levels of ALDH1A1 protein correlated with worse OS and DFS of OC patients [41]. However, our study also revealed that the accumulation of mRNA for ALDH1A1 in tumor tissue is correlated with higher sensitivity to platinum-based chemotherapy, which is usually an indicator of a better prognosis for OC patients. ALDH1A1 may have an impact on chemoresistance. A broad analysis of multiple ovarian cancer cell lines revealed significantly higher ALDH1A1 gene expression in taxane- and platinum-resistant cell lines [42,43]. ALDH1A1 is also active in platinum-resistant cancer cells residing in hypoxic regions [9,40]. In several studies in other cancer types, including breast cancer, stromal ALDH1A1 protein level, as measured by immunohistochemistry, was associated with better clinical outcomes [44].

Our detailed evaluation of TCGA data revealed that a higher level of EPCAM mRNA was associated with a higher percentage of optimal debulking during primary surgery. Furthermore, EPCAM is among the genes whose expression is associated with a positive response to platinum compounds. It is also related to improving DFS and OS survival in OC patients [45]. Thus, the data from the human protein [46], together with the results of our study, indicate that EpCAM gene expression is associated with a favorable prognosis of OC patients. Our univariate and multivariate analyses demonstrated that the EpCAM mRNA level is associated with longer PFS and OS.

Interestingly, we revealed that improved patient survival times are correlated only with increased EpCAM mRNA levels. Such a link was not observed at the protein level. Previous studies showed [27,45] the correlation between EpCAM protein level and decreased OS and higher FIGO stage and grade of the disease [27]. Thus, the EpCAM protein was proposed as a significant factor contributing to the chemoresistance of OC cells. All this suggests that EPCAM is an important marker of OC prognosis. However, further research is needed to completely understand its role in the disease.

Our analysis consistently demonstrated a correlation between THY-1 gene expression and the unfavorable prognosis of OC patients. We have shown that THY-1 mRNA and protein levels are considered an independent factor associated with poor OS, with higher levels in FIGO stage III and IV and the mesenchymal subtype of OC. Moreover, THY-1 mRNA but not protein level was associated with platinum resistance and poor DFS in OC. THY1 transcriptional activity was associated with the highly invasive and metastatic potential of OC cells, and its correlations with significantly shorter median DFS and OS in patients with HGSOC were previously demonstrated [26,47]. However, depending on the cancer type, THY-1 might have ambivalent properties regarding its anti- or protumoral activities. In some studies on OC, THY-1 was suggested to have a tumor-suppressive role [24,48,49]. The mechanism by which the THY-1 gene may inhibit OC cell growth is still unclear. Its ambivalent impact reported in many studies may be the result of its presence in immune and cancer cells [50]. However, among all markers we have studied, THY-1 has a consistently unfavorable impact on the protein and mRNA levels. Therefore, it is an interesting candidate for further studies.

Another cancer stem cell marker with a substantial impact on the platinum sensitivity of OC cells was POU5F1. Our study revealed that its high expression at the mRNA level was associated with platinum resistance and, therefore, worse DFS in OC patients. In a study by Xie W et al. [51], the POU5F1 protein level was significantly correlated with higher tumor grades and lymph node metastases. Therefore, it was suggested to promote proliferation and metastasis in OC. It was also considered an independent predictor of poor prognosis and progression of OC [51]. Moreover, a study by Ruan Z et al. [52] suggested the significant role of POU5F1 in stemness and drug resistance of OC. It was also correlated with a more aggressive phenotype of OC.

LGR5 is another CSC marker presented as a predictor of good response to platinum therapy. Our multivariate and univariate analysis showed that LGR5 was associated with longer DFS. We also observed that the LGR gene was associated with the proliferative type of OC. The expression of LGR was previously shown to be associated with higher stages of the disease, and its upregulation was associated with metastases in OC patients [28]. The increased expression of LGR5 markedly promotes the growth and the EMT of ovarian cancer cells. Thus, it may promote tumorigenesis and the formation of metastases. Though we observed higher expression in the proliferative subtype, samples more responsive to platinum also had higher levels of LGR5 mRNA. Consistently with our results, Kim and colleagues [53] revealed that in HGSOC, high LGR5 expression was associated with improved PFS.

In this study, we analyzed data on the expression of a set of CSC markers in TCGA. Thanks to this approach, we validated specific observations regarding well-known markers and studied proteins of unclear relevance. TCGA is a database containing a single tumor sample that does not provide insight into the heterogeneity of HGSOC. It must be considered as a limitation when interpreting the results. However, it may serve as a valuable tool for screening and narrowing a set of CSC markers in terms of more detailed analyses. Consequently, some proteins we included have well-established CSC marker status, whereas others are still considered candidates.

4. Materials and Methods

4.1. Data Retrieval

Transcriptome (microarray and RNA sequencing) data were obtained from TCGA using <http://www.cbioportal.org/public-portal/> (accessed on 1 September 2022) as CGDS object (mycgds) compatible with downstream analysis with R programming [54].

For analysis of microarray data from OC tumors, values provided in the genetic profile “ov_tcga_pub_mrna_median_Zscores” in the study labeled as “ov_tcga_pub” were selected [55]. We analyzed patients included in the “ov_tcga_pub_all” case list. Patients with missing values or genes defined only in a subset of patients were excluded from the analysis. For comparison of RNA-Seq and microarray input, OC tumor values pro-

vided in the genetic profile “ov_tcga_rna_seq_v2_mrna_median_Zscores” in the study labeled as “ov_tcga” were selected. We used clinical characterization of the patients obtained by the function `cgdsr::getClinicalData()`. The predefined OC expression profiles (fallopian, immunoreactive, mesenchymal, and proliferative) were retrieved with the function `cgdsr::getCaseLists(mycgds, 'ov_tcga_pub')`. The `cgdsr::getCaseLists(mycgds, 'ov_tcga_pub')` function allows the user to fetch a list IDs of the patients from the ov_tcga_pub study. The predefined OC expression profiles (fallopian, immunoreactive, mesenchymal, and proliferative) could be retrieved with the `cgdsr::getProfileData` function.

We are aware that there is an overlap between the “ov_tcga_pub” and “ov_tcga” datasets. Given that the samples have different categories of clinical annotation (such as response to platinum therapy or pathological features of the tumor), we analyzed both sets. All data processing was performed in the R programming language (version 4.2.1) using RStudio (version 2022.12.0 + 353 “Elsbeth Geranium” Release).

4.2. Clinical and Pathological Features

In statistical and predictive analyses, we analyzed mRNA expression assessed by a microarray from the study labeled “ov_tcga_pub” ($n = 489$) for 13 CSC markers: ALDH1A1, CD44, EPCAM, KIT, LGR5, NES, NOTCH3, POU5F1, PROM1, PTTG1, ROR1, SOX9, and THY1. Data for the remaining four out of 16 CSC markers (c-Kit, SNORD89, SNORA72, and TMSB4X) were incomplete and therefore excluded from further analysis. We compared the markers' expression with the clinical and pathological features, including DFS, OS, grade, platinum sensitivity status, primary therapy outcome, tumor residual disease, stage, and expression profile.

For some patients, only data on mRNA expression were available, whereas for others, data on both mRNA and protein expression were available. Additional statistical analyses were performed on the dataset labeled as “ov_tcga”. This included analysis of mRNA ($n = 558$) expression data of the same 13 CSC markers described above and mass spectrometry measurements of the 5 proteins ($n = 174$) produced by the ALDH1A1, CD44, EPCAM, NES, and THY1 genes. Only a part of the dataset records contained information about both mRNA and protein levels; the rest comprised only mRNA data. We statistically compared the expression of the mRNA with clinical and pathological features, including clinical stage, DFS, OS, grade, histological type, treatment outcome, and expression profile. According to TCGA database, after therapy, platinum-resistant cancer recurs within six months, whereas in platinum-sensitive cancer, the recurrence takes place more than six months after completion of platinum-based chemotherapy. In our statistical and predictive analyses, we included information about MKI67 as a proliferation marker.

4.3. Data Preprocessing

For predictive analysis of carboplatin therapy outcome and multivariate survival analyses, we conducted additional data preprocessing steps. First, we excluded records with values of categorical features that were highly under-represented (<10 records). Less under-represented values of ordinal categorical features were grouped with more numerous values in the sequence. For multivariate analysis, the residual disease feature was grouped into categories: “no visible residual disease” and “visible residual disease” (tumor tissue of any size remaining after the surgery). Tumor grade was divided into “grade G3” and “non-grade G3” (G1 and G2) groups, and cancer stage features were split into “stage IV” and “non-stage IV” groups.

To facilitate the use of machine learning methods, we performed numerical encoding of categorical features. The expression profile feature was one-hot-encoded as four separate variables corresponding to either immunoreactive, fallopian, mesenchymal, or proliferative expression profiles. Moreover, we performed binary encoding on categorical variables that were previously divided into two groups. In such a way, we encoded residual disease assessment as “Is visible residual disease” binary feature, tumor grade as “Is grade G3”, and cancer stage as “Is stage IV”.

Lastly, for the purpose of training carboplatin therapy outcome classifiers, we tackled the class imbalance issue with random oversampling. We did not apply any class balancing methods for either of the survival analyses, as they provide no performance gains to the models trained on censored survival data [56].

4.4. Statistics

In our statistical analyses, we tested the normality of the data distribution using a Shapiro–Wilk test. Student’s *t*-test (*t.test*) was used for data with a normal distribution to calculate the statistical significance (*p*-value). A Mann–Whitney *U* test was used for data with a non-Gaussian distribution. Kruskal–Wallis and ANOVA tests were used to compare the data of more than two independent groups for non-normally and normally distributed data, respectively. The R package ‘ggstatsplot’ was used to generate plots [57].

The data were presented as boxplots with an interquartile range (IQR), median, minimum, and maximum values. Jitter plots were used to demonstrate the scattering of the data. The central line in the plots corresponds to the median. The upper and lower “hinges” correspond to the first and third quartiles, respectively. Plot whiskers extend to the most extreme data point, which is within 1.5 times the interquartile range from the box. We visualized data representing individual patients as dots in the boxplots. The correlations were calculated using Pearson’s correlation and presented in a correlation matrix with Pearson correlation coefficients (*r*) and *p* values using the `ggstatsplot::ggcorrmat()` function.

4.5. Univariate Analysis of Overall and Disease-Free Survival

The relationships between OS and DFS and expression data were analyzed based on a study labeled “ov_tcga_pub”, as well as mRNA and mass spectrometry data from a study labeled “ov_tcga”. We applied the survival and survminer libraries. A Cox proportional hazards regression model was used to perform survival analysis. The data were presented as hazard ratios with 95% confidence intervals, beta (β) coefficients, and calculated *p* values. The plots were generated by the `ggstatsplot::ggcoefstats()` function.

4.6. Predictive Analysis of Carboplatin Treatment Outcome

We assessed the predictive power of a random forest model that was trained on the “ov_tcga_pub” microarray dataset to perform binary classification platinum treatment efficacy (*n* = 287). The features considered in analysis are expression levels of the same 13 CSC markers that were used for the univariate analysis. Moreover, the MKI67 proliferation marker, residual disease assessment, tissue expression profile, tumor grade, and cancer stage features were included.

To create random forest models, we used the `randomForest::randomForest` function [58]. We assessed the performance of classifiers using AUC score and a 5-fold cross validation (CV) method repeated 34 times, which resulted in 170 models trained for performance estimation. We optimized each model’s hyperparameters using 5-fold cross validation and a grid search performed on its training set. The hyperparameters were also evaluated using the AUC score. For each model, we chose parameters that showed the best performance on validation sets. Random forest models were optimized for the best values of forest size in a range of 100–300 and tree node size in a range of 5–30. To test whether the estimated performance was significantly better than a random guess, we performed a one-sample Student’s *t*-test (distribution passed the Shapiro–Wilk normality test) and checked an alternative hypothesis, i.e., that the AUC score distribution > 0.5. The AUC score was calculated using the `pracma::trapz` function [59]. The ROC plot was prepared using the `ggplot2` package. Each plotted curve represents a performance assessment of one of the models trained during CV.

4.7. Multivariate Analysis of Overall and Disease-Free Survival

We used the Cox proportional hazards model with ridge regularization [60] to analyze OS and DFS of patients in the “ov_tcga_pub” microarray dataset. We considered only records

with complete mRNA expression levels and complete clinical and pathological data for the 13 CSCs, including residual disease assessment, tissue expression profile, tumor grade, and cancer stage information ($n = 487$). We included the MKI67 proliferation marker in the analysis.

The models were obtained using the `glmnet::cv.glmnet` function [61,62]. For ridge regularization, we set the alpha parameter to zero. Moreover, no baseline function was assumed, as we focused only on hazard risk prediction. We assessed the performance of the hazard models trained on the “ov_tcga_pub” dataset using the c-index score and a 5-fold cross-validation (CV) method repeated 34 times, which resulted in training 170 models for performance estimation. The `cv.glmnet` function incorporates a hyperparameter optimization method by cross validation. The optimization was performed on each model’s training set with 5-folds CV, with the c index as a validation score. The optimized parameter was lambda, which determines the strength of the regularization. Other parameters were left at default values. To test whether assessed performance was significantly better than a random guess, we performed a one-sample Student’s t-test (distribution passed the Shapiro–Wilk normality test) and checked an alternative hypothesis, i.e., that the c index > 0.5 .

To further assess the performance of the hazards model trained on the “ov_tcga_pub” dataset, we stratified patient risk into low- (patient median risk < 1.0) and high-risk (patient median risk ≥ 1.0) strata. The efficacy of the patient division was determined using a log-rank test. We visualized the stratification results with Kaplan–Meier survival curves calculated using the `survival::survfit` function and drawn with the `survminer::ggsurvplot` function [63]. The curves were additionally described with confidence intervals. The dashed lines indicate the time of median survival probability for each stratum.

4.8. Feature Importance Analysis

To uniformly estimate the importance of “ov_tcga_pub” dataset features, in all our machine learning analyses, we used the permutation feature importance (PFI) model-agnostic score [64]. The PFI is measured as an increase in the model’s prediction error after permuting the feature’s values. The PFI scores were obtained using the `iml::FeatureImp$new` function with “ratio” as a comparison method [65].

For PFI measurements, we used the same models that were previously trained for the purpose of performance estimation. PFI measurement was repeated 30 times on each model and its corresponding test set, resulting in 5100 measurements per feature in total. To test whether a feature was significantly important, we performed a one-sample Wilcoxon test (distributions did not pass the Shapiro–Wilk normality test) and checked an alternative hypothesis, i.e., that the obtained PFI score distribution is greater than the value, indicating no prediction importance ($PFI > 1$). A feature was determined as important if its estimated PFI was higher than 1 with $p < 0.01$.

The feature importance is presented as a box plot with interquartile range, median, and minimum and maximum values drawn in the same way as in our statistical analysis using the `ggplot2` package (version 3.4.2).

5. Conclusions

Our study showed that the expression of six CSC markers (CD44, ALDH1A1, EpCAM, THY-1, POU5F1, and LGR5) is significant in OC in terms of clinical outcome, including stage and grade of the disease, as well as the platinum sensitivity and patient survival. The CD44, ALDH1A1, EpCAM, THY-1, POU5F1, and LGR5 levels in OC may be used as prognostic factors for the primary outcome and may be beneficial in predicting the treatment response.

Author Contributions: Conceptualization, N.I., K.S. and E.N.-M.; Methodology, M.P.Z. and Ł.C.; Software, M.P.Z., Ł.C. and K.J.; Validation, M.P.Z., Ł.C. and K.J.; Formal Analysis, K.J., S.S., C.M., M.R., A.S. and K.P.-B.; Investigation, N.I.; Resources, N.I., M.P.Z. and S.S.; Data Curation, N.I. and M.P.Z.; Writing—original draft preparation, N.I., M.P.Z., K.J., C.M., M.R., A.S. and K.P.-B.; Writing—review and editing, K.S., E.N.-M., M.F. and A.J.; Visualization: M.P.Z., Ł.C. and K.J.; Supervision, K.S., M.F. and E.N.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data presented in the study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

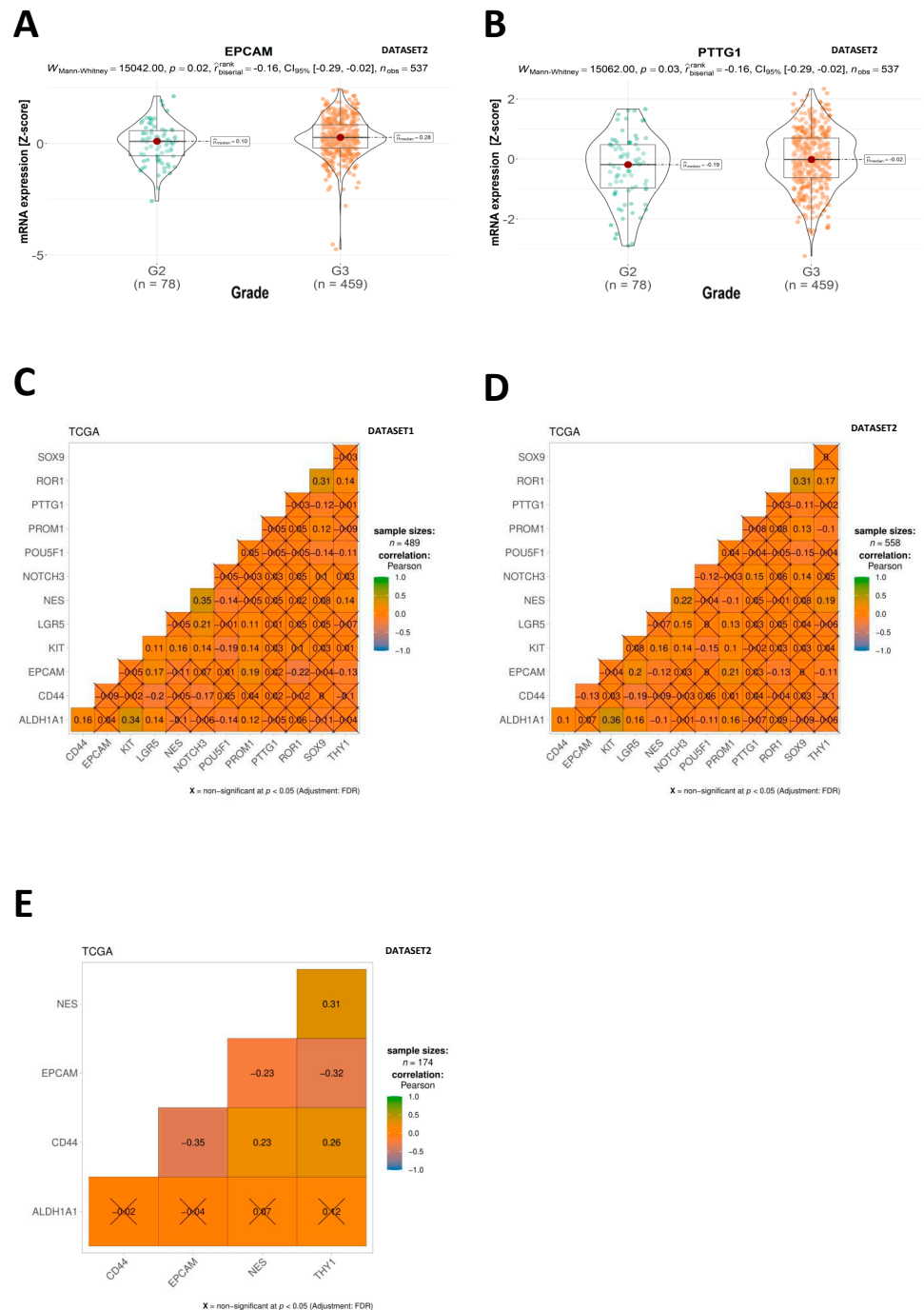


Figure A1. Comparison between clinical and pathological features (A,B) and correlation matrices (C–E) (“ov_tcga” and “ov_tcga_pub” datasets) (in (A), we cut the lower and upper outliers).

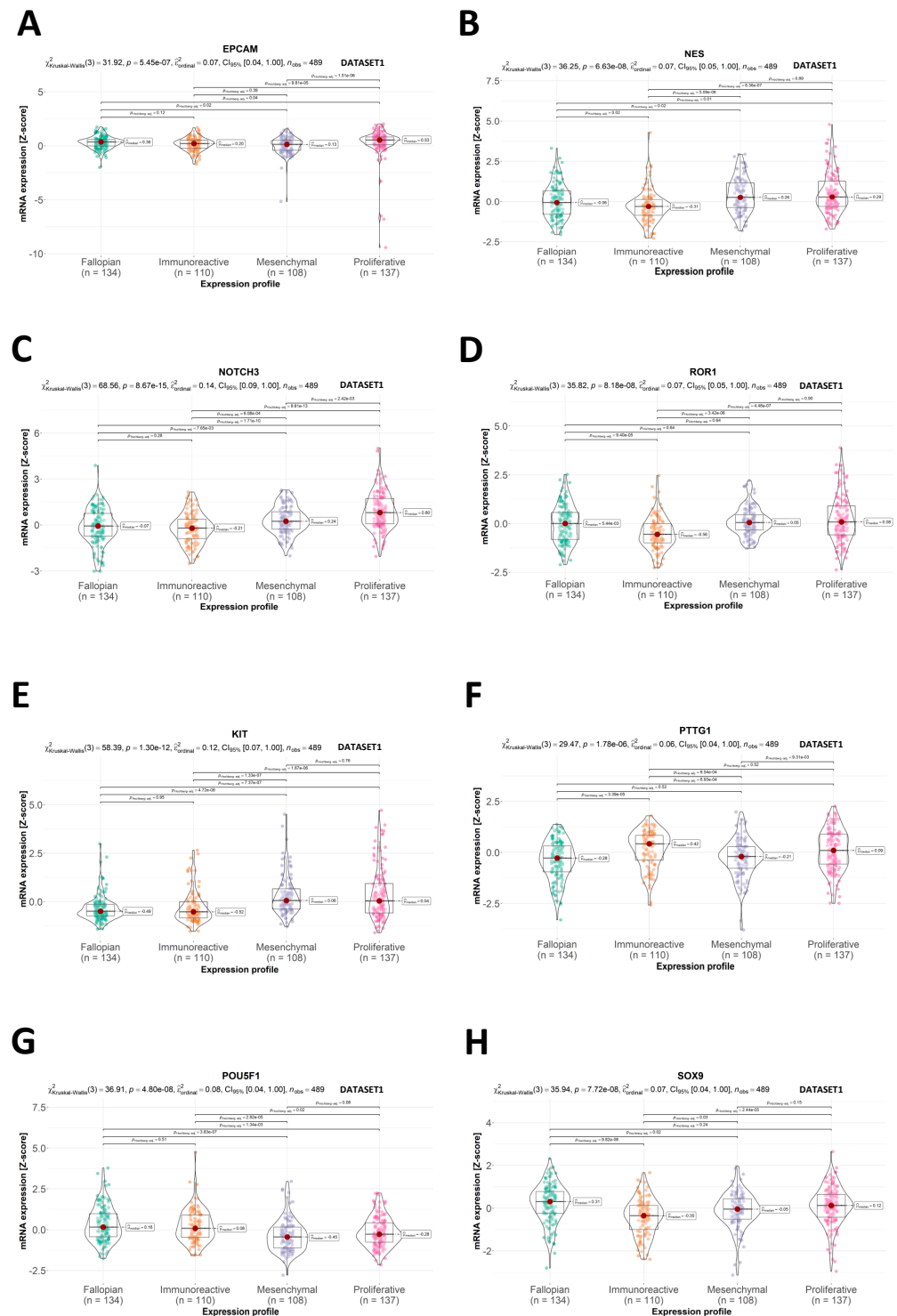


Figure A2. Expression profiles of the “ov_tcga_pub” dataset. mRNA levels are represented as Z scores in subgroups according to OC expression profiles. High amounts of EpCAM (A), NES (B) NOTCH3 (C), ROR 1 (D) transcripts were associated with proliferative profile. The upregulation of KIT (E) was associated with mesenchymal profile. Elevated level of PTTG1 (F) was associated with immunoactive profile. The fallopian profile was characterized by elevation of POU5F1 (G) and SOX9 (H).

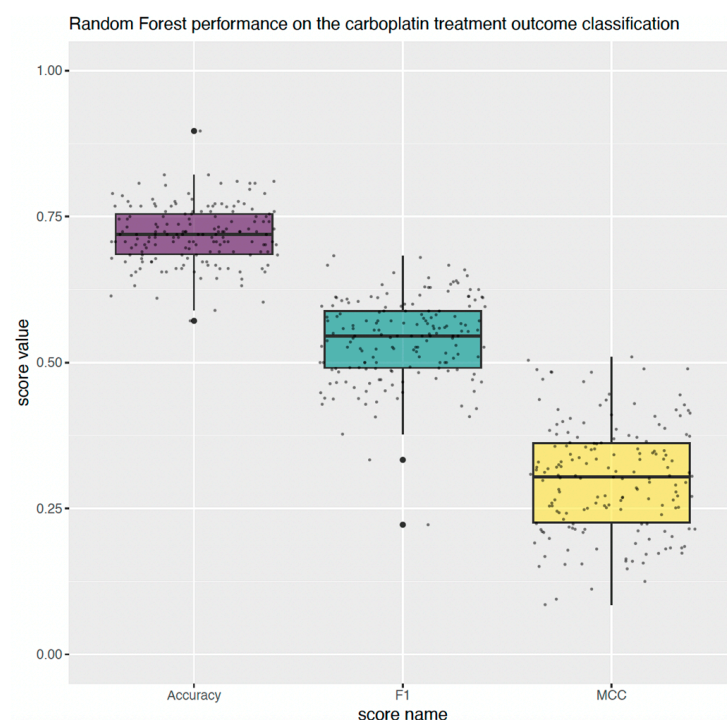


Figure A3. Expected performance of the random forest classifier trained on the “ov_tcga_pub” dataset estimated using repeated fivefold cross validation and different machine learning model assessment scores, including accuracy, F1 score, and MCC (Matthews correlation coefficient).

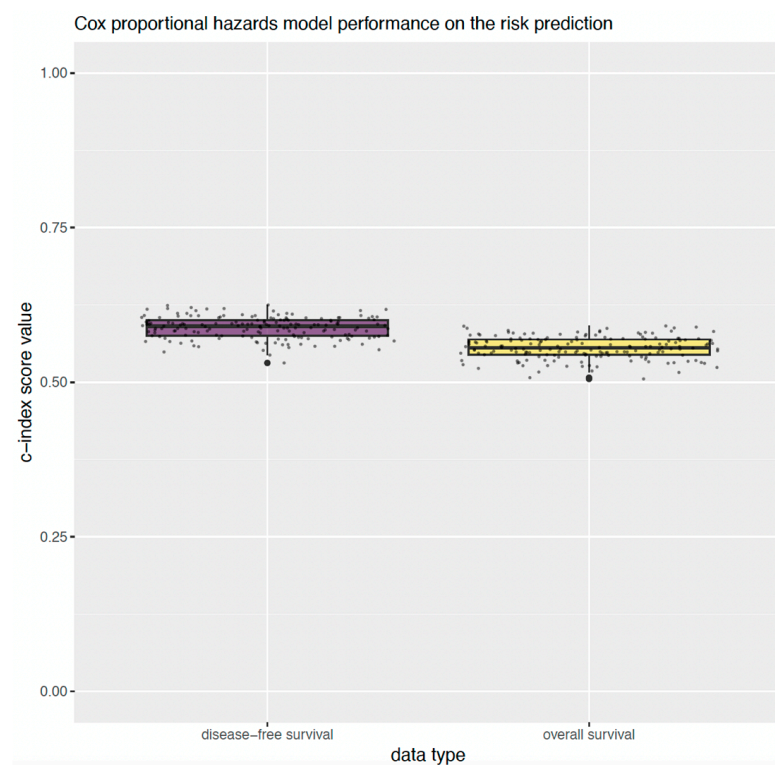


Figure A4. Expected risk prediction performance of the Cox proportional hazards models with ridge regularization fitted to the “ov_tcga_pub” dataset disease-free and overall survival data. The performance was estimated using repeated fivefold cross validation (CV) and the c-index score.

References

- Zhan, Q.; Wang, C.; Ngai, S. Ovarian Cancer Stem Cells: A New Target for Cancer Therapy. *BioMed Res. Int.* **2013**, *2013*, 916819.
- Tomao, F.; Papa, A.; Martina, S.; Rossi, L.; Russo, G.L.; Panici, P.B.; Ciabatta, F.R.; Tomao, S. Investigating Molecular Profiles of Ovarian Cancer: An Update on Cancer Stem Cells. *J. Cancer* **2014**, *5*, 301–310.
- Keyvani, V.; Farshchian, M.; Esmaeili, S.A.; Yari, H.; Moghbeli, M.; Nezhad, S.R.K.; Abbaszadegan, M.R. Ovarian Cancer Stem Cells and Targeted Therapy. *J. Ovarian Res.* **2019**, *12*, 120. [[PubMed](#)]
- Gupta, P.B.; Chaffer, C.L.; Weinberg, R.A. Cancer Stem Cells: Mirage or Reality? *Nat. Med.* **2009**, *15*, 1010–1012.
- Takahashi, A.; Hong, L.; Chefetz, I. Review Open Access Cancer Drug Resistance How to Win the Ovarian Cancer Stem Cell Battle: Destroying the Roots. *Cancer Drug Resist.* **2020**, *3*, 1021–1054.
- Lupia, M.; Cavallaro, U. Ovarian Cancer Stem Cells: Still an Elusive Entity? *Mol. Cancer* **2017**, *16*, 64.
- Markowska, A.; Sajdak, S.; Huczyński, A.; Rehliś, S.; Markowska, J. Ovarian Cancer Stem Cells: A Target for Oncological Therapy. *Adv. Clin. Exp. Med.* **2018**, *27*, 1017–1020. [[PubMed](#)]
- Parte, S.C.; Batra, S.K.; Kakar, S.S. Characterization of Stem Cell and Cancer Stem Cell Populations in Ovary and Ovarian Tumors. *J. Ovarian Res.* **2018**, *11*, 69. [[PubMed](#)]
- The Cancer Genome Atlas Research Network. Integrated Genomic Analyses of Ovarian Carcinoma. *Nature* **2011**, *474*, 609–615, Erratum in *Nature* **2012**, *490*, 292.
- Verhaak, R.G.W.; Tamayo, P.; Yang, J.Y.; Hubbard, D.; Zhang, H.; Creighton, C.J.; Feraday, S.; Lawrence, M.; Carter, S.L.; Mermel, C.H.; et al. Prognostically Relevant Gene Signatures of High-Grade Serous Ovarian Carcinoma. *J. Clin. Invest.* **2013**, *123*, 517–525.
- Riester, M.; Wei, W.; Waldron, L.; Culhane, A.C.; Trippa, L.; Oliva, E.; Kim, S.H.; Michor, F.; Huttenhower, C.; Parmigiani, G.; et al. Risk Prediction for Late-Stage Ovarian Cancer by Meta-Analysis of 1525 Patient Samples. *J. Natl. Cancer Inst.* **2014**, *106*, dju048. [[PubMed](#)]
- Waldron, L.; Riester, M.; Birrer, M. Molecular Subtypes of High-Grade Serous Ovarian Cancer: The Holy Grail? *J. Natl. Cancer Inst.* **2014**, *106*, dju297. [[PubMed](#)]
- Tan, T.Z.; Miow, Q.H.; Huang, R.Y.J.; Wong, M.K.; Ye, J.; Lau, J.A.; Wu, M.C.; Bin Abdul Hadi, L.H.; Soong, R.; Choolani, M.; et al. Functional Genomics Identifies Five Distinct Molecular Subtypes with Clinical Relevance and Pathways for Growth Control in Epithelial Ovarian Cancer. *EMBO Mol. Med.* **2013**, *5*, 1051–1066. [[PubMed](#)]
- Tothill, R.W.; Tinker, A.V.; George, J.; Brown, R.; Fox, S.B.; Lade, S.; Johnson, D.S.; Trivett, M.K.; Etemadmoghadam, D.; Locandro, B.; et al. Novel Molecular Subtypes of Serous and Endometrioid Ovarian Cancer Linked to Clinical Outcome. *Clin. Cancer Res.* **2008**, *14*, 5198–5208.
- Cioffi, M.; Dalterio, C.; Camerlingo, R.; Tirino, V.; Consales, C.; Riccio, A.; Ieranò, C.; Cecere, S.C.; Losito, N.S.; Greggi, S.; et al. Identification of a Distinct Population of CD133⁺ CXCR4⁺ Cancer Stem Cells in Ovarian Cancer. *Sci. Rep.* **2015**, *5*, 10357.
- Napoli, J.L.; Boerman, M.H.E.M.; Chai, X.; Zhai, Y.; Fiorella, P.D. Enzymes and Binding Proteins Affecting Retinoic Acid Concentrations. *J. Steroid Biochem. Mol. Biol.* **1995**, *53*, 497–502.
- Martincuks, A.; Li, P.C.; Zhao, Q.; Zhang, C.; Li, Y.J.; Yu, H.; Rodriguez-Rodriguez, L. CD44 in Ovarian Cancer Progression and Therapy Resistance—A Critical Role for STAT3. *Front. Oncol.* **2020**, *10*, 2551.
- Mizrak, D.; Brittan, M.; Alison, M.R. CD 133: Molecule of the Moment. *J. Pathol.* **2008**, *214*, 3–9.
- Roy, L.; Bobbs, A.; Sattler, R.; Kurkewich, J.L.; Dausinas, P.B.; Nallathamby, P.; Dahl, K.D.C. CD133 Promotes Adhesion to the Ovarian Cancer Metastatic Niche. *Cancer Growth Metastasis* **2018**, *11*, 117906441876788. [[CrossRef](#)]
- Jiang, J.; Chen, Y.; Zhang, M.; Zhou, H.; Wu, H. Relationship between CD177 and the Vasculogenic Mimicry, Clinicopathological Parameters, and Prognosis of Epithelial Ovarian Cancer. *Ann. Palliat. Med.* **2020**, *9*, 3985–3992.
- Robinson, M.; Gilbert, S.F.; Waters, J.A.; Lujano-olazaba, O.; Lara, J.; Alexander, L.J.; Green, S.E.; Burkeen, G.A.; Patrus, O.; Sarwar, Z.; et al. Characterization of SOX2, OCT4 and NANOG in Ovarian Cancer Tumor-Initiating Cells. *Cancers* **2021**, *13*, 262. [[PubMed](#)]
- Price, J.C.; Azizi, E.; Naiche, L.A.; Parvani, J.G.; Shukla, P.; Kim, S.; Slack-Davis, J.K.; Pe'er, D.; Kitajewski, J.K. Notch3 Signaling Promotes Tumor Cell Adhesion and Progression in a Murine Epithelial Ovarian Cancer Model. *PLoS ONE* **2020**, *15*, e0233962.
- Aguilar-Medina, M.; Avendaño-Félix, M.; Lizárraga-Verdugo, E.; Bermúdez, M.; Romero-Quintana, J.G.; Ramos-Payan, R.; Ruíz-García, E.; López-Camarillo, C. SOX9 Stem-Cell Factor: Clinical and Functional Relevance in Cancer. *J. Oncol.* **2019**, *2019*, 6754040. [[PubMed](#)]
- Abeyasinghe, H.; Cao, Q.; Xu, J.; Pollock, S.; Veyberman, Y.; Guckert, N.; Keng, P.; Wang, N. THY1 Expression Is Associated with Tumor Suppression of Human Ovarian Cancer. *Cancer Genet. Cytogenet.* **2003**, *143*, 125–132.
- Abeyasinghe, H.; Pollock, S.; Guckert, N.; Veyberman, Y.; Keng, P.; Halterman, M.; Federoff, H.; Rosenblatt, J.; Wang, N. The Role of the THY1 Gene in Human Ovarian Cancer Suppression Based on Transfection Studies. *Cancer Genet. Cytogenet.* **2004**, *149*, 1–10.
- Connor, E.V.; Saygin, C.; Braley, C.; Wiechert, A.C.; Karunanithi, S.; Crean-Tate, K.; Abdul-Karim, F.W.; Michener, C.M.; Rose, P.G.; Lathia, J.D.; et al. Thy-1 Predicts Poor Prognosis and Is Associated with Self-Renewal in Ovarian Cancer. *J. Ovarian Res.* **2019**, *12*, 112.
- Tayama, S.; Motohara, T.; Narantuya, D.; Li, C.; Fujimoto, K.; Sakaguchi, I.; Tashiro, H.; Saya, H.; Nagano, O.; Katabuchi, H.; et al. The Impact of EpCAM Expression on Response to Chemotherapy and Clinical Outcomes in Patients with Epithelial Ovarian Cancer. *Oncotarget* **2017**, *8*, 44312–44325.

28. Liu, W.; Zhang, J.; Gan, X.; Shen, F.; Yang, X.; Du, N.; Xia, D.; Liu, L.; Qiao, L.; Pan, J.; et al. LGR5 Promotes Epithelial Ovarian Cancer Proliferation, Metastasis, and Epithelial–Mesenchymal Transition through the Notch1 Signaling Pathway. *Cancer Med.* **2018**, *7*, 3132.
29. Parte, S.; Virant-Klun, I.; Patankar, M.; Batra, S.K.; Straughn, A.; Kakar, S.S. PTTG1: A Unique Regulator of Stem/Cancer Stem Cells in the Ovary and Ovarian Cancer. *Stem Cell Rev. Rep.* **2019**, *15*, 866–879.
30. Henry, C.; Llamas, E.; Knipprath-Meszaros, A.; Schoetzau, A.; Obermann, E.; Fuenfschilling, M.; Caduff, R.; Fink, D.; Hacker, N.; Ward, R.; et al. Targeting the ROR1 and ROR2 Receptors in Epithelial Ovarian Cancer Inhibits Cell Migration and Invasion. *Oncotarget* **2015**, *6*, 40310–40326.
31. Osman, W.M.; Shash, L.S.; Ahmed, N.S. Emerging Role of Nestin as an Angiogenesis and Cancer Stem Cell Marker in Epithelial Ovarian Cancer: Immunohistochemical Study. *Appl. Immunohistochem. Mol. Morphol.* **2017**, *25*, 571–580. [PubMed]
32. Yoon, H.J.; Oh, Y.L.; Ko, E.J.; Kang, A.; Eo, W.K.; Kim, K.H.; Lee, J.Y.; Kim, A.; Chun, S.; Kim, H.; et al. Effects of Thymosin B4-Derived Peptides on Migration and Invasion of Ovarian Cancer Cells. *Genes Genom.* **2021**, *43*, 987–993.
33. Abd El-Fattah, G.A.; Ibrahim, E.; Nasif, S.N. Significance of Tumoral and Stromal ALDH 1A1 Expression in Breast Invasive Duct Carcinoma in Egyptian Female Patients. *Egypt. J. Pathol.* **2019**, *39*, 151–158.
34. Woopen, H.; Pietzner, K.; Richter, R.; Fotopoulou, C.; Joens, T.; Braicu, E.I.; Mellstedt, H.; Mahner, S.; Lindhofer, H.; Darb-Esfahani, S.; et al. Overexpression of the Epithelial Cell Adhesion Molecule Is Associated with a More Favorable Prognosis and Response to Platinum-Based Chemotherapy in Ovarian Cancer. *J. Gynecol. Oncol.* **2014**, *25*, 221–228. [PubMed]
35. EPCAM Protein Expression Summary—The Human Protein Atlas. Available online: <https://www.proteinatlas.org/ENSG00000119888-EPCAM> (accessed on 8 June 2023).
36. Sauzay, C.; Voutetakis, K.; Chatziioannou, A.A.; Chevet, E.; Avril, T. CD90/Thy-1, a Cancer-Associated Cell Surface Signaling Molecule. *Front. Cell Dev. Biol.* **2019**, *7*, 66.
37. Zeng, L.; Peng, Z.; Duan, Z. Expression of THY1 Gene in Epithelial Ovarian Cancer. *Chin. J. Oncol.* **2009**, *31*, 118–120.
38. Abeysinghe, H.R.; Li, Q.L.; Guckert, N.L.; Reeder, J.; Wang, N. THY-1 Induction Is Associated with up-Regulation of Fibronectin and Thrombospondin-1 in Human Ovarian Cancer. *Cancer Genet. Cytogenet.* **2005**, *161*, 151–158.
39. Haeryfar, S.M.M.; Hoskin, D.W. Thy-1: More than a Mouse Pan-T Cell Marker. *J. Immunol.* **2004**, *173*, 3581–3588.
40. Xie, W.; Yu, J.; Yin, Y.; Zhang, X.; Zheng, X.; Wang, X. OCT4 Induces EMT and Promotes Ovarian Cancer Progression by Regulating the PI3K/AKT/MTOR Pathway. *Front. Oncol.* **2022**, *12*, 876257.
41. Ruan, Z.; Yang, X.; Cheng, W. OCT4 Accelerates Tumorigenesis through Activating JAK/STAT Signaling in Ovarian Cancer Side Population Cells. *Cancer Manag. Res.* **2019**, *11*, 389.
42. Kim, H.; Lee, D.H.; Park, E.; Myung, J.K.; Park, J.H.; Kim, D.I.; Kim, S.I.; Lee, M.; Kim, Y.; Park, C.M.; et al. Differential Epithelial and Stromal LGR5 Expression in Ovarian Carcinogenesis. *Sci. Rep.* **2022**, *12*, 11200. [PubMed]
43. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2*, 401–404. [PubMed]
44. Van Den Goorbergh, R.; Van Smeden, M.; Timmerman, D.; Van Calster, B. The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation Using Logistic Regression. *J. Am. Med. Inf. Assoc.* **2022**, *29*, 1525–1534.
45. Patil, I. Visualizations with Statistical Details: The “ggstatsplot” Approach. *J. Open Source Softw.* **2021**, *6*, 3167.
46. Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R News* **2002**, *2*, 18–22.
47. Package “pracma” Title Practical Numerical Math Functions Depends R ($\geq 3.1.0$). 2022. Available online: <https://cran.r-project.org/web/packages/pracma/index.html> (accessed on 8 June 2023).
48. Marquardt, D.W.; Snee, R.D. Ridge Regression in Practice. *Am. Stat.* **1975**, *29*, 3–20.
49. Simon, N.; Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* **2011**, *39*, 1–13.
50. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22.
51. Drawing Survival Curves Using “ggplot2” [R Package Survminer Version 0.4.9]. 2021. Available online: <https://rpkgs.datanovia.com/survminer/> (accessed on 8 June 2023).
52. Fisher, A.; Rudin, C.; Dominici, F. All Models Are Wrong, but Many Are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
53. Molnar, C.; Casalicchio, G.; Bischl, B. Iml: An R Package for Interpretable Machine Learning. *J. Open Source Softw.* **2018**, *3*, 786.
54. Stauder, R.; Günthert, U. CD44 Isoforms. Impact on Lymphocyte Activation and Differentiation. *Immunologist* **1995**, *3*, 78–83.
55. CD44 Expression Indicates Favorable Prognosis in Epithelial Ovarian Cancer | Clinical Cancer Research | American Association for Cancer Research. Available online: <https://aacrjournals.org/clincancerres/article/9/14/5318/202149/CD44-Expression-Indicates-Favorable-Prognosis-in> (accessed on 8 June 2023).
56. Sosulski, A.; Horn, H.; Zhang, L.; Coletti, C.; Vathipadiakal, V.; Castro, C.M.; Birrer, M.J.; Nagano, O.; Saya, H.; Lage, K.; et al. CD44 Splice Variant V8-10 as a Marker of Serous Ovarian Cancer Prognosis. *PLoS ONE* **2016**, *11*, e0156595.
57. Mao, M.; Zheng, X.; Jin, B.; Zhang, F.; Zhu, L.; Cui, L. Effects of CD44 and E-Cadherin Overexpression on the Proliferation, Adhesion and Invasion of Ovarian Cancer Cells. *Exp. Ther. Med.* **2017**, *14*, 5557. [PubMed]

58. Zhou, J.; Du, Y.; Lu, Y.; Luan, B.; Xu, C.; Yu, Y.; Zhao, H. CD44 Expression Predicts Prognosis of Ovarian Cancer Patients through Promoting Epithelial-Mesenchymal Transition (EMT) by Regulating Snail, ZEB1, and Caveolin-1. *Front. Oncol.* **2019**, *9*, 802.
59. Hu, Y.; Taylor-Harding, B.; Raz, Y.; Haro, M.; Recouvreux, M.S.; Taylan, E.; Lester, J.; Millstein, J.; Walts, A.E.; Karlan, B.Y.; et al. Are Epithelial Ovarian Cancers of the Mesenchymal Subtype Actually Intraperitoneal Metastases to the Ovary? *Front. Cell Dev. Biol.* **2020**, *8*, 647.
60. Izycka, N.; Rucinski, M.; Andrzejewska, M.; Szubert, S.; Nowak-Markwitz, E.; Sterzynska, K. The Prognostic Value of Cancer Stem Cell Markers (CSCs) Expression-ALDH1A1, CD133, CD44-For Survival and Long-Term Follow-Up of Ovarian Cancer Patients. *Int. J. Mol. Sci.* **2023**, *24*, 2400. [[PubMed](#)]
61. Kaipio, K.; Chen, P.; Roering, P.; Huhtinen, K.; Mikkonen, P.; Östling, P.; Lehtinen, L.; Mansuri, N.; Korpela, T.; Potdar, S.; et al. ALDH1A1-Related Stemness in High-Grade Serous Ovarian Cancer Is a Negative Prognostic Indicator but Potentially Targetable by EGFR/MTOR-PI3K/Aurora Kinase Inhibitors. *J. Pathol.* **2020**, *250*, 159–169. [[PubMed](#)]
62. Ruscito, I.; Darb-Esfahani, S.; Kulbe, H.; Bellati, F.; Zizzari, I.G.; Rahimi Koshkaki, H.; Napoletano, C.; Caserta, D.; Ruggetti, A.; Kessler, M.; et al. The Prognostic Impact of Cancer Stem-like Cell Biomarker Aldehyde Dehydrogenase-1 (ALDH1) in Ovarian Cancer: A Meta-Analysis. *Gynecol. Oncol.* **2018**, *150*, 151–157.
63. Nowacka, M.; Ginter-Matuszewska, B.; Świerczewska, M.; Sterzyńska, K.; Nowicki, M.; Januchowski, R. Effect of ALDH1A1 Gene Knockout on Drug Resistance in Paclitaxel and Topotecan Resistant Human Ovarian Cancer Cell Lines in 2D and 3D Model. *Int. J. Mol. Sci.* **2022**, *23*, 3036.
64. Sterzyńska, K.; Klejewski, A.; Wojtowicz, K.; Świerczewska, M.; Nowacka, M.; Kaźmierczak, D.; Andrzejewska, M.; Rusek, D.; Brażert, M.; Brażert, J.; et al. Mutual Expression of ALDH1A1, LOX, and Collagens in Ovarian Cancer Cell Lines as Combined CSCs- and ECM-Related Models of Drug Resistance Development. *Int. J. Mol. Sci.* **2018**, *20*, 54.
65. Mohyeldin, A.; Garzón-Muvdi, T.; Quiñones-Hinojosa, A. Oxygen in Stem Cell Biology: A Critical Component of the Stem Cell Niche. *Cell Stem Cell* **2010**, *7*, 150–161. [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.