Advances in Cancer Biology - Metastasis 7 (2023) 100088

Contents lists available at ScienceDirect

Advances in Cancer Biology - Metastasis

journal homepage: www.journals.elsevier.com/advances-in-cancer-biology-metastasis

Modulation of gene expression associated with copy number variation identifies key regulatory programs in high-grade serous ovarian carcinoma

Martina Vescio [a], Lara Paracchini [b], Luca Beltrame [c], Maurizio D'Incalci [b,c], Sergio Marchini [c,1], Linda Pattini [a,*,1]

[a] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Giuseppe Ponzio 34, 20133, Milan, Italy
[b] Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20090, Pieve Emanuele, Milan, Italy
[c] IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089, Rozzano, Milan, Italy

ARTICLE INFO

Keywords:
High-grade serous ovarian carcinoma
Gene expression
Copy number variation
Systems biology
Cancer informatics

ABSTRACT

High grade serous ovarian carcinoma (HGSOC) is a systemic malignancy characterized by metastatic lesions that spread within the peritoneal cavity. Despite initial sensibility to platinum-based chemotherapy, more than 80% of patients experience a relapse and acquire chemoresistance. From a genomic point of view, HGSOC shows a high level of inter- and intra-tumor heterogeneity. A better understanding of molecular mechanisms of this disease and the identification of driving genetic alterations could provide relevant indications for diagnostic and prognostic evaluation. Here, we accomplished a double-tier omic-analysis by integrating copy number variation data with matched gene expression profiling of multiple lesions in a cohort of 7 patients. We identified potential driver genes contained in amplified regions whose behavior seem to impact on gene expression program. They represent a distinctive signature that can segregate biopsies of different patients. Moreover, a further analysis highlighted *ZNF696*, *ASPSCR1* and *RHPN1* as key drivers, whose regulatory program is confirmed in TCGA cohort. In conclusion, exploration of gene expression program in HSGOC by integrating copy number and transcriptomic data from spatially separated samples obtained from seven patients led to the identification of genes whose amplification is significantly correlated to specific gene expression modules and are related to survival.

1. Introduction

High grade serous ovarian carcinoma (HGSOC) is the most common type of epithelial ovarian cancer, accounting for 60–80% of all cases [1]. It usually manifests as an advanced stage, systemic disease, with multiple metastatic lesions diffused within the abdominal cavity. This type of cancer is characterized by a high degree of genomic instability [2], leading to a progressive accumulation of alterations and to a coexistence of genetically distinct subclones within the same tumor. This mechanism results in an extensive inter- and intra-tumor heterogeneity [3] due to the different evolutionary trajectory followed by metastases located in separated anatomical sites in the abdominal cavity (synchronous lesions) or arising after primary treatment (metachronous lesions). Intra-patient heterogeneity was explored in some recent studies analyzing multiple samples obtained from the same subject. These studies revealed a wide mutational burden with little overlap between biopsies obtained from the same subject and a low correlation between the tumor mass in the ovary and its matched metastatic lesion [4,5]. These results show that the identification of molecular biomarkers is a critical issue in this context. However, genetic lesions identified in the ovary are the only ones currently used for diagnostic and therapeutic purposes. Additionally, the presence of multiple clones makes HGSOC difficult to treat with conventional chemotherapy, as minor resistant subclones can persist in the abdominal cavity as a result of a non-optimal cytoreductive surgery and expand to generate a new chemoresistant malignancy [6–8]. Altogether, the advanced stage at the diagnosis and the acquired chemoresistance result in an overall 5-year survival probability of 31%, a percentage that has not significantly improved over the last years [9]. Therefore, it is of major importance to identify shared molecular features that are really representative of the entire disease to improve therapy and survival. In this context, in a previous study analyzing both synchronous and metachronous lesions, we could

* Corresponding author.
E-mail address: linda.pattini@polimi.it (L. Pattini).
[1] Both authors contributed equally as senior authors.

https://doi.org/10.1016/j.adcanc.2023.100088
Received 30 August 2022; Received in revised form 5 January 2023; Accepted 13 January 2023
Available online 14 January 2023

observe that copy number amplifications in regions 8q24.3 and 3q26.2 seem to overcome heterogeneity and to arise at an early stage of tumor development [10].

In the present study we integrated copy number variations and gene expression data from a total of 28 biopsies obtained from 7 patients to characterize amplified regions and their functional impact. Using a systems biology approach, we identified a selection of genes located in amplified regions that could act as drivers in HGSOC.

2. Materials and methods

2.1. Patient cohort

2.1.1. Cohort A

The analyses were carried out on a cohort composed of 28 biopsies collected at different anatomical sites during cytoreductive surgery from seven patients, with an average of four biopsies per patient. Biopsies were collected at the Division of Gynecology and Oncology Dept., Manzoni Hospital (Lecco, Italy) and stored in Pandora/Lecco frozen tissue bank collection at the Mario Negri Institute Biological Resources Center (BRC).

The study was conducted following the principles of the Declaration of Helsinki; the scientific ethical committee "Brianza" approved the collection and usage of tumor, blood and plasma samples (N° 1065, on November 10th, 2015, emended on February 22nd, 2018). Written informed consent was obtained from all patients.

All patients were diagnosed with HGSOC and did not receive neo-adjuvant chemotherapy before surgery. They were staged accordingly to the International Federation of Gynecology and Obstetrics criteria (FIGO) as stage III/IV. Matched gene expression (Array Express Accession ID: pending) and copy number data (Array Express Accession ID: E-MTAB-6900) were available for all samples in the cohort.

2.1.2. Cohort B

488 additional ovarian biopsies with matched gene expression and copy number data were selected from TCGA to validate the results from cohort A. All samples were taken from stage III/IV HGSOC patients before the administration of chemotherapy. Normalized microarray data were downloaded from the Broad Institute's Firehose (https://firebrowse.org) website. Survival data were retrieved using the R package TCGAbiolinks [11]. Overall survival (OS) data were available for 484 samples, while progression free survival (PFS) data were available for 477 samples.

2.2. RNA microarrays

RNA was extracted using a commercial kit (miRNeasy, QIAGEN, Milan, Italy), according to the manufacturer's instructions. 100 ng of total RNA were reverse-transcribed into Cy3-labeled cRNA using Low-Input QuickAmp labeling kit (Agilent Technologies, Palo Alto, CA, USA), and hybridized onto commercially available array platforms. Acquired images were processed by the scanning software (Agilent Feature Extraction). Then, the calculated intensity ratios were log2-transformed to reduce the variance and further normalized using quantile normalization [12].

2.3. Array comparative genomic hybridization (aCGH)

aCGH data were processed as previously reported in Ref. [10].

Finally, Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm was applied on normalized data to identify the most significant regions with aberrant copy number [13]. Significant regions were selected using a confidence threshold of 75% and a q-value threshold of 0.05. Moreover, minimum log2 values for gain and for loss were set to 0.5 and -0.5, respectively.

2.4. Molecular subtype classification

Each biopsy was assigned to one of the four molecular subtypes of high grade serous ovarian carcinoma identified in Ref. [14] and characterized as mesenchymal, immunoreactive, differentiated, and proliferative in Ref. [15]. Gene signatures distinctive of each subtype included 372 genes for mesenchymal subtype (C1), 289 genes for immunoreactive (C2), 484 genes for differentiated subtype (C4) and 955 genes for proliferative subtype (C5). Gene expression data from cohort A and B were z-transformed to allow a comparison. For each subtype, the median expression of genes belonging to its characteristic signature was calculated on TCGA samples belonging to the specific subtype. The obtained expression profile was used to represent the subtype. The correlation between each biopsy from cohort A and each expression profile was tested by calculating the Spearman's correlation coefficient. The resulting p-values were adjusted for multiple testing and the effects due to different lengths of the gene signatures using a bootstrap procedure. Each test was repeated using 1000 random signatures having the same length of the one associated to the specific subtype obtained by random sampling of genes in the dataset. The bootstrapped p-value was defined as the proportion of signatures obtaining a higher correlation coefficient than the original signature. Each biopsy was then assigned to the most correlated molecular subtype as long as none of the following conditions was verified: the largest correlation coefficient was smaller than 0.1, the smallest bootstrapped p-value was larger than 0.05.

Moreover, if the difference between the two largest correlation coefficients was smaller than 0.01, the biopsy was considered ambiguous and the top two classifications were assigned [16].

2.5. Data integration

The integration of gene expression and copy number data was carried out using the Conexic algorithm [17]. It implements a Bayesian scoring approach to identify driver genes located in regions with altered DNA copy number. It is based on module network, a Bayesian network representation that aims to reduce the complexity of the model space by partitioning the variables into groups that act as placeholders [18]. These groups are called "modules" and contain genes with highly similar expression patterns. Initially, CONEXIC forms the modules and assigns them a driver gene. The driver gene is selected from a list of candidate regulators provided as input to the algorithm. The gene that best explains the expression variation in the module is chosen as its driver gene. In this work, candidate modulators were selected by testing the agreement between copy number and gene expression in genes located in amplified regions. First, samples were ranked based on mRNA levels in descending order. Then, the enrichment of amplified samples in the upper part of the list was tested using the gsea function from the R package phenoTest. To increase the sensitivity of the selection, both cohorts were considered in this analysis. Genes with FDR <0.05 and enrichment score (ES) > 0 in at least one of the cohorts were selected as candidate regulators.

2.6. Regulators' expression analysis

Two-way hierarchical clustering on gene expression data of the complete list of regulators was performed using Python package fastcluster with Pearson's correlation as distance metric. The results were portrayed in a heatmap using the Python package seaborn. Additionally, the first three principal components from principal component analysis were plotted using the Python package matplotlib.

2.7. Target selection

An additional selection was performed on data from cohort B in order to validate the relationships in a larger dataset and reduce module dimensionality. First, the direction of the influence of the driver on the

expression of each module was evaluated in cohort A. For each module, CONEXIC defines an optimal threshold that separates samples in a low-expression and a high-expression group. After the z-transformation of the expression values of the module, the same division was maintained to compare the set of values taken by target genes in low-expression samples to the set of values of the high-expression samples using a Welch *t*-test. The sign of the test statistic was used to determine if the regulator had an activating or an inhibiting effect on the genes in module. Modules with p-value $\geq$0.05 were excluded from further analysis.

Then, an optimal threshold was calculated for each regulator in the independent dataset. This threshold was defined as the value that minimized the p-value of the one-tailed *t*-test comparing samples below threshold and samples above threshold. The smaller group had to contain at least 25 samples (>5%). The optimal partition determined by the threshold was maintained in each target gene and a one-tailed *t*-test was performed between the two groups of samples. The direction of the test was determined by the value of the test statistic computed in cohort A for the examined module. The resulting p-values were adjusted for multiple testing using Bonferroni correction. Target genes with adjusted p-value <0.05 were selected.

A second step of selection was performed using iterative Bayesian Model Averaging (BMA) on survival data from cohort B [19]. Starting from a list of genes, BMA selects the set of genes with higher impact on survival using a multivariate approach. First, genes in the list are ranked in descending order of log likelihood calculated using univariate Cox proportional hazard model. Then, BMA iteratively tests groups of 30 genes starting from the top of the list and removes genes having a posterior probability <1%. BMA was applied to each module for both OS and PFS data using the R package BMA.

2.8. Survival analysis

To remove the influence of the clinical variables, the analysis was carried out separately on patients of grade 3 and stage IIIC and patients of grade 3 and stage IV. Only patients with these clinical characteristics were selected for the analysis to mirror the characteristics of cohort A. A total of 374 patients (313 stage IIIC, 61 stage IV) were tested for OS and 369 patients (310 stage IIIC, 59 stage IV) were tested for PFS. The group including stage IIIC samples was used as the main testing group, while the smaller one containing stage IV patients was used to confirm the results. The groups of genes selected applying BMA on OS data were tested for OS, while genes selected using PFS data were tested for association to PFS. For each module, Cox proportional hazard model was applied to data from the main testing group with the expression of genes in the module as covariates. P-values from the likelihood ratio test were extracted and corrected for multiple testing using Bonferroni correction. Modules with corrected p-value <0.05 in both OS and PFS were selected, and their significance was tested in stage IV patients. Modules obtaining a p-value <0.05 in both OS and PFS in the smaller group were considered significant. All the analyses were carried out using R package survival.

2.9. Network construction

To inspect the function of the modules found significant in the survival analysis, molecular interactions involving selected genes were retrieved using IntAct, a database collecting literature derived interactions [20]. Networks including the molecular interactions were constructed using the software Cytoscape v3.3.0. A functional enrichment analysis was performed on the genes comprised in each network using Gene Ontology (GO) and Reactome gene sets collections retrieved from the Molecular Signatures Database (MSigDB) [21]. To test the enrichment of a specific gene network with terms associated to an annotated gene set, a hypergeometric test was performed. P-values obtained in each collection were corrected for multiple testing using Benjamini-Hochberg procedure. A term was selected as significant if its FDR was lower than 0.05.

3. Results

3.1. Cohorts description

The cohort under study included a total of 28 biopsies obtained from 7 patients at different anatomical sites, with an average of four biopsies per patient. This cohort was already described in a previous work by Ballabio S, Craparotta I, Paracchini L, et al. [10]. All patients enrolled in the study were diagnosed with high grade serous ovarian cancer (HGSOC). Additional clinical information are illustrated in Table 1. For all samples, matched gene expression and copy number variations (CNV) data were available. To validate the results in a larger dataset, we selected 488 HGSOC samples with matched gene expression and CNV data from The Cancer Genome Atlas (TCGA).

3.2. Array CGH analysis

Normalized copy number data from arrayCGH technique were analyzed using GISTIC algorithm [13] to identify the most significant aberrant regions in the cohort. The analysis selected 27 focal amplified regions and 24 focal deleted regions passing the given thresholds (Suppl. Table 1). The most frequently amplified regions were located on chromosome 3 (3q29, 60.71%), chromosome 5 (5p15.33, 64.29%) and chromosome 8 (8q24.22, 89.29%; 8q24.3, 92.86%; 8q24.3, 96.43%). The three regions on chromosome 8 and the region 3q29 also reached the lowest q-values. As for deletions, the most frequently altered regions involved chromosome 16 (16q22.1, 64.29%), chromosome X (Xp11.23, 57.14%), chromosome 5 (5q14.2, 53.57%) and chromosome 13 (13q21.32, 53.57%; 13q21.31, 53.57%). Collectively, amplifications had higher frequencies and higher G-score when compared to deletions. Moreover, they appeared to be more homogeneous over biopsies obtained from the same patient (Fig. 1 A). For these reasons, we chose to focus on amplified regions.

3.3. Biopsies taken at different anatomical sites show distinct expression profiles from the primary tumor

A further characterization was carried out on our cohort to test if biopsies taken from different anatomical sites at the time of surgery exhibited similar molecular characteristics to the biopsy taken at the ovary. To this aim, the biopsies were assigned to one of the HGSOC molecular subtypes defined in Refs. [14,15] according to their expression profile. These groups are called mesenchymal, immunoreactive, differentiated and proliferative and express different molecular signatures, indicating diversified phenotypes. First, we defined a representative expression profile for each molecular subtype. Then, the similarity between subtypes' profiles and biopsies was tested using Spearman's correlation. Three samples did not reach a significant p-value in any tested subtype and were marked as unclassified. Two additional samples are marked as ambiguous and are reported to be associated to two

Table 1
Clinical characteristics of the patients enrolled in the study.

Clinical annotations	No. of patients (%)
Histotype	
Serous	7 (100)
Grade	
3	7 (100)
Stage	
IIIC	4 (57.14)
IV	3 (42.86)
Age at diagnosis	
Mean [min-max]	61.29 [41–81]

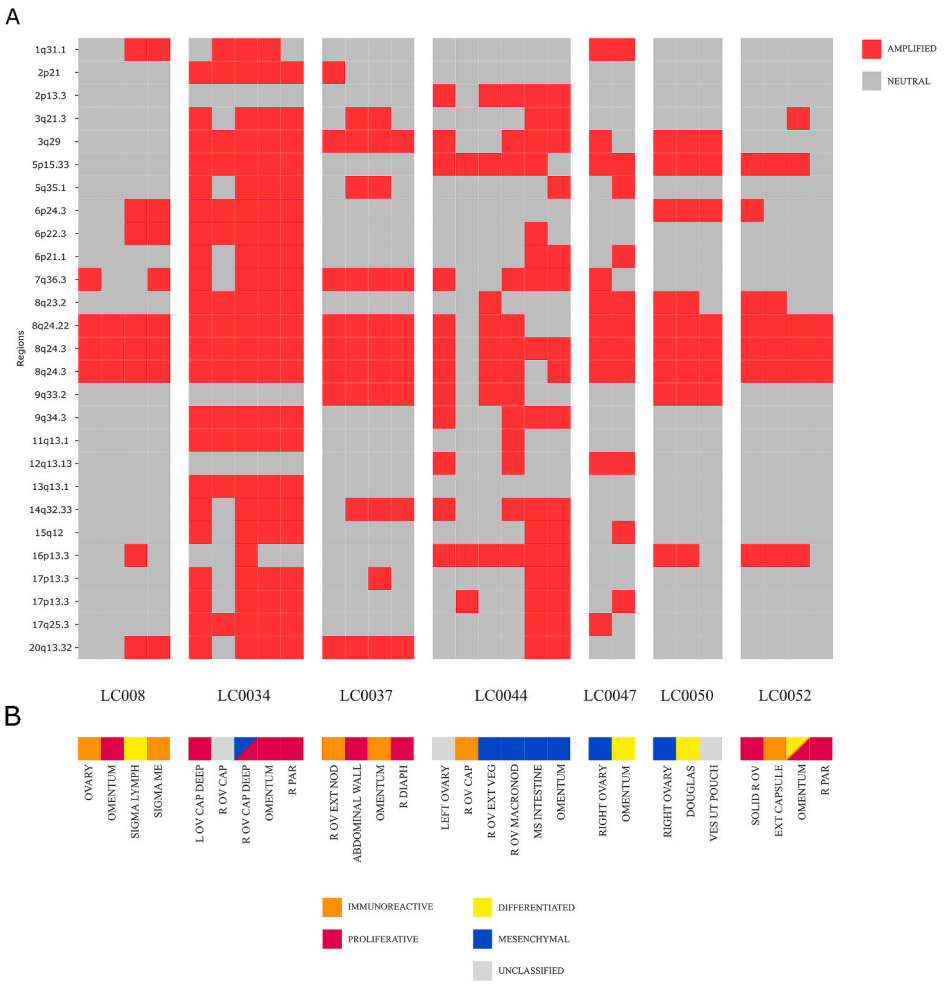

Fig. 1. A) Amplification status of significant regions selected by GISTIC algorithm in each biopsy. Amplified samples are indicated in red, neutral copy number are marked in grey. B) Results of molecular subtype classification. Different colors correspond to the subtype assigned to each biopsy (orange: immunoreactive; magenta: proliferative; yellow: differentiated; blue: mesenchymal). Samples in grey were unclassified, while ambiguous samples are represented by both colors associated to the chosen subtypes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

different subtypes. As shown in Fig. 1B, different molecular subtypes can coexist within the same patient. These results further confirm the high intratumor heterogeneity observed in HGSOC, suggesting that lesions in different anatomical locations follow distinct evolutionary paths.

3.4. Amplified driver genes are able to discriminate biopsies from the same patient

Given the emergence of distinctive amplification patterns, we decided to investigate the impact of these genetic aberrations on tumor phenotype. In particular, we wanted to assess the influence of the expression of amplified genes on the expression levels of other genes. We tested the influence of the amplification state of genes found in the amplified regions identified by GISTIC on their own mRNA level using a GSEA-like approach [22]. We found 70 genes that displayed a significant correlation between transcription level and genetic aberration and we chose them as candidate modulators of gene expression. An algorithm of multi-omic data integration (CONEXIC, [17]) was applied on copy number variation and matched gene expression data to reconstruct the regulatory relationships between the modulators and the other genes in the dataset.

The final output of the algorithm consisted in 43 groups of coexpressed genes, i.e. modules, each with an associated driver gene. The identified regulators comprised 19 genes, listed in Table 2 together with their corresponding cytoband. The most recurrent chromosomal regions in the selection are 2p13.3 and 8q24.3, both accounting for 4 regulators, and 8q24.22, comprising 3 modulators. Interestingly, regions 8q24.3 and 8q24.22 were also the most frequently amplified in our cohort. As

Table 2
List of driver genes selected by CONEXIC algorithm organized by cytoband. In the second column, the number of modules regulated by the gene is shown.

Symbol	Modules	Cytoband
GMCL1	4	2p13.3
PCBP1-AS1	4	
SNRNP27	3	
PCBP1	1	
BDH1	1	3q29
FYTTD1	1	
CLPTM1L	1	5p15.33
SLC35B3	1	6p24.3
ENY2	1	8q23.1
PKHD1L1	1	
EBAG9	1	8q23.2
PHF20L1	2	8q24.22
SLA	2	
WISP1	1	
ZNF696	13	8q24.3
LRRC14	2	
RHPN1	2	
NAPRT	1	
ASPSCR1	1	17q25.3

the same gene can be responsible for more than one module, the number of modules associated to each driver is also reported in Table 2. The gene regulating the highest number of co-expression modules is *ZNF696*, located on 8q24.3 and associated to 13 modules, followed by *GMCL1*, *PCBP1-AS1* and *SNRNP27*, located on 2p13.3 and assigned to 4, 4 and 3 modules, respectively. Gene expression profiles of the identified driver

genes were considered to test their ability to discriminate between different patients. The clustered heatmap and the PCA plot are reported in Fig. 2 and show a high concordance between biopsies from the same patient. Specifically, four patients out of seven had all their biopsies grouped together in the same cluster. Moreover, five out of six biopsies (83%) of an additional patient shared the same cluster.

3.5. Selection of regulatory relationships confirmed in a larger dataset

A two-step selection was carried out on the genes contained in each module by retaining genes whose association to the driver was confirmed in the independent cohort retrieved from TCGA and whose expression modulation was related to survival, subsequently refining the modules to only retain genes with the highest prognostic value. An example of a refined module is shown in Fig. 3, here the target genes included after selection maintain a strong association with the regulator in both datasets.

3.6. Co-expression modules can be associated to prognosis

The prognostic significance of regulatory modules was assessed by applying Cox proportional hazard model separately on stage IIIC and stage IV patients from the TCGA cohort. Three modules were found to be significantly related to both overall survival (OS) and progression free survival (PFS) in stage IIIC and stage IV patients (Table 3). Two modules were driven by *ZNF696* and *RHPN1*, both located on the region 8q24.3, the most frequently amplified in our dataset. The third module was regulated by *ASPSCR1*, located on 17q25.3.

3.7. Gene modules are related to important cellular functions

To investigate the biological function of genes in modules found significant in survival analysis, we retrieved their protein-protein interactions (PPI) from the IntAct database [20]. The interactions were then used to construct a network and an enrichment analysis was carried out on the list of network components to examine their molecular function. Fig. 4 shows the most significant GO terms and Reactome pathways resulting from enrichment analysis.

The networks related to the potential driver *ZNF696* were significantly enriched of genes involved in microtubule organization during cell cycle, immune system, and activation of CREB and Ras proteins. The module driven by *ASPSCR1* was linked to cytoskeleton organization, activation of the immune system and regulation of apoptosis. Finally, the networks regulated by *RHPN1* were significantly associated to intermediate filaments and, more specifically, to keratin filaments.

4. Discussion

In this work, we integrated information from copy number variations and gene expression data to explore molecular heterogeneity between spatially distributed biopsies obtained from seven HGSOC patients and characterize the most significant copy number aberrations.

First, each biopsy was assigned to one of the four molecular subtypes identified in Refs. [14,15] to study the variability of the molecular characteristics among samples from different anatomical sites. The coexistence of multiple subtypes in each patient further confirmed that the heterogeneity of HGSOC is significant not only between patients, but also inside the same subject due to the different evolution of distinct cell populations. A similar result was obtained in a previous study in which the authors observed that biopsies from the same patient belonged to clusters associated to distinct molecular subtypes [4].

The analysis of copy number alterations revealed that amplifications were more frequent and less heterogeneous over biopsies than deletions. This result suggests a crucial role of these aberrations in the development of HGSOC. In particular, the analysis carried out with GISTIC highlighted two separated regions on cytoband 8q24.3 of 619kbp and 75kbp, respectively, that were amplified in more than 90% of samples. A region on the same cytoband was already reported as one of the most significantly amplified in TCGA samples. This region was altered in 74% of samples and it included the 80% of our first region and the totality of our second region. Moreover, in our previous study, 8q24.3 was reported as the most conserved region in both synchronous and metachronous lesions from the same patient [10].

The regulatory relationships involving genes located in aberrant

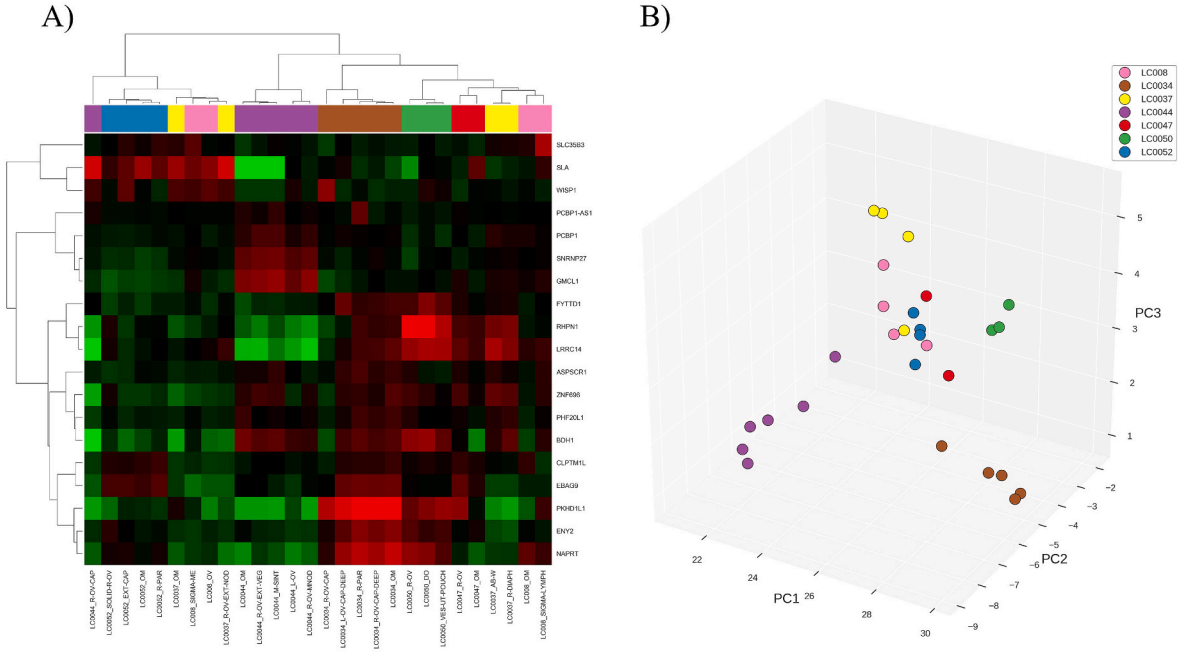

Fig. 2. A) Heatmap of two-way hierarchical clustering on gene expression of putative driver genes identified by CONEXIC algorithm. Colors identify biopsies taken from the same patient. Clustering on the set of genes can discriminate efficiently between different patients. B) Representation of the first three principal components of PCA on drivers' expression. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

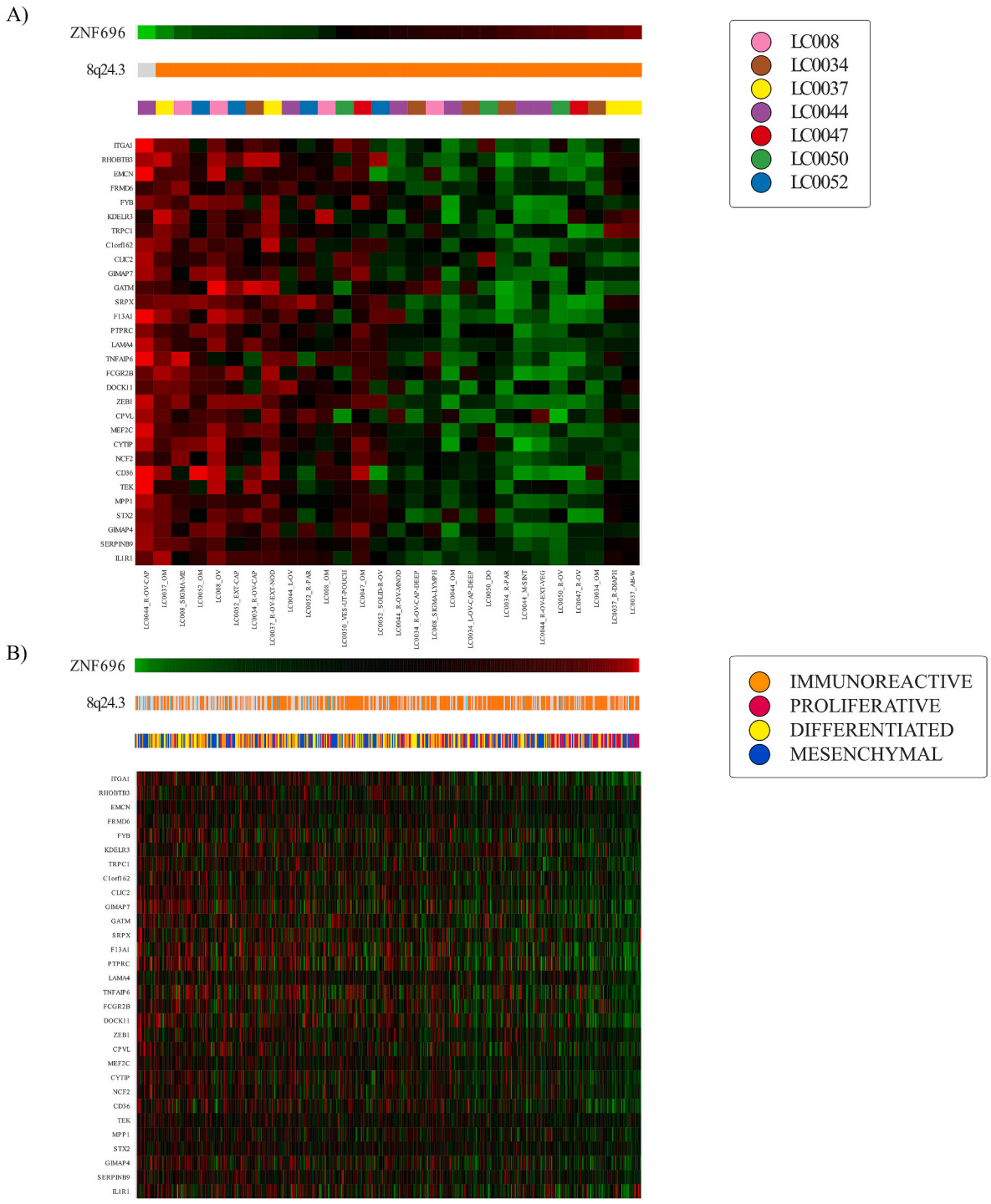

Fig. 3. Expression of a module in our data (panel A) and TCGA dataset (panel B) after target selection. For this specific case, overall survival data were used in BMA selection. Rows represent the genes in the module and columns represent the samples. The upper line portrays the expression of the regulator. The second line outlines the state of amplification of the region the regulator falls in (orange: copy number gain; light blue: copy number loss; grey: copy number neutral). Different colors in the third line indicate patients (panel A) and molecular subtypes (panel B). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

regions were reconstructed using a Bayesian network-based algorithm. This analysis identified 19 driver genes that modulate the expression of 43 groups of co-expressed genes. Notably, 18 out of 43 modules were regulated by genes located on 8q24.3, further confirming the importance of this specific region in the pathology.

Unsupervised analysis on drivers' expression showed their potential in segregating biopsies according to their patient of origin. Specifically, four patients out of seven had all their biopsies grouped together in the same cluster as well as five out of six biopsies of one of the remaining patients. Therefore, these genes could provide information about the intrapatient evolution of the pathology.

Some of the identified driver genes were already associated to various cancer types. For example, *WISP1* (8q24.22) is primarily associated to advanced stage prostate cancer and could have a role in metastasis formation [23]. In a recent study *WISP1* was also found to be related to a worse prognosis in multiple solid cancer types, such as

Table 3
Significant modules for overall survival (OS) and progression free survival (PFS) in TCGA samples of stage IIIC and IV.

Module	Regulator	Cytoband	Stage IIIC		Stage IV	
			p-value (OS)	p-value (PFS)	p-value (OS)	p-value (PFS)
2	ZNF696	8q24.3	2.69E-06	1.43E-04	1.41E-02	1.00E-03
20	ASPSCR1	17q25.3	2.67E-03	4.81E-03	1.25E-02	3.51E-03
7	RHPN1	8q24.3	3.52E-02	2.65E-02	1.24E-05	9.56E-07

low-grade glioma, kidney papillary cell carcinoma, bladder urothelial carcinoma, ovarian serous cystadenocarcinoma, and primary melanoma. Additionally, it is correlated to EMT-related genes in multiple cancer datasets [24]. *EBAG9* is an estrogen-responsive gene located on region 8q23.2 that is widely expressed in breast carcinoma and may contribute to its development [25]. As for epithelial ovarian cancer, its expression was found to be significantly higher in serous histology and advanced disease with respect to other histologies or lower stages. However, *EBAG9* was not correlated to overall survival [26]. *ENY2*, a gene located on region 8q23.1, is involved in mRNA export and transcription activation. Experiments carried out on both breast cancer cell lines and in vivo have demonstrated that *ENY2* expression significantly increases cell migration and invasion [27].

To focus on smaller groups of genes, we reduced the list of targets in each module by selecting only the genes having a robust association

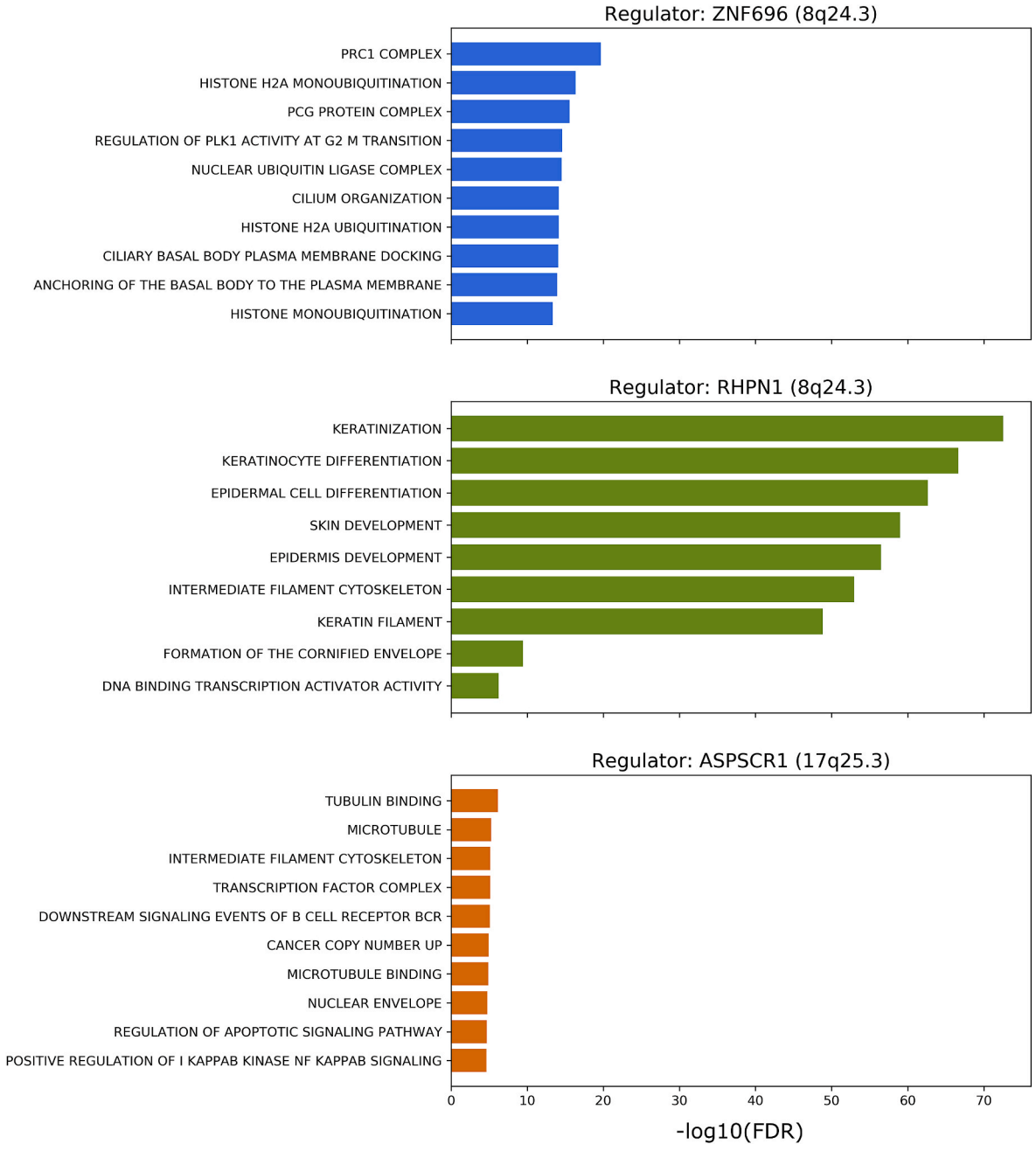

Fig. 4. Barplot illustrating ten of the most significant terms obtained in the enrichment analysis carried out on the interaction networks constructed for the three selected modules.

with the regulator that was maintained in a larger dataset. Additionally, we narrowed down the number of modules under investigation by selecting only the modules associated to both overall survival and progression free survival. This resulted in a final selection of three co-expression modules regulated by *ASPSCR1*, located on 17q25.3, *ZNF696* and *RHPN1*, both located on 8q24.3.

The molecular interactions network comprising the genes in the three selected modules were reconstructed to better characterize functions and pathways related to each module. The protein encoded by *ASPSCR1* is responsible for the redistribution of glucose transporter type 4 (*GLUT4*) in response to insulin. Its fusion with transcription factor E3 (*TFE3*) results in an oncoprotein characteristic of alveolar soft part sarcoma [28]. Its module was associated to regulation of apoptosis through the activation of BH3-only proteins and the activation of NF-kB, a hallmark of inflammatory response essential in the development of inflammation-associated cancer [29].

ZNF696 codes for a member of the zinc finger proteins; various members of this family have been associated to cancer progression and invasion over the last decades [30]. This gene is poorly characterized but it seems to be involved in transcriptional regulation. In our analyses, *ZNF696* was linked to many modules and was encoded on the most frequently amplified region. These findings, together with its supposed function, suggest that this gene could be of major importance in HGSOC and it would be crucial to further investigate its role in future studies. Its most significant module was enriched with pathways related to microtubule organization during cell cycle and immune system. Additionally, the results indicate that this module could contribute to enhance cell growth through the promotion of mitotic spindle formation and the activation of Ras protein. The latter is frequently observed in various types of cancer and it is known to favor malignant transformation [31]. Finally, *RHPN1* codes for a Rho GTPase-interacting protein whose function is mostly unknown. However, it seems to interact with some cytoskeletal components [32], in accordance with the terms found enriched in its module. Specifically, its module was found to be associated to keratin filaments, a cytoskeletal molecule characteristic of epithelial cells, suggesting a role in the control of malignant cell level of differentiation.

In summary, we accomplished a multi-omics characterization of HSGOC by integrating copy number and gene expression data from spatially separated samples obtained from seven patients. The conservative approach, implying selection of signatures confirmed in an independent dataset and provided with prognostic potential, allowed the identification of relevant potential drivers and their co-expressed gene modules. These genes could be crucial to the progression of the pathology and represent triggering elements upstream the development of heterogeneity across biopsies from distinct anatomical sites.

Author contributions

MV performed computational analyses and wrote the manuscript. LP performed gene expression experiments. LB supported bioinformatic analyses. MD and SM conceived the experimental study. LP* conceived computational analyses and modelling approach and revised the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available on public repositories.

Acknowledgments

The results shown here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

This research was funded by the “Alessandra Bono Foundation” for supporting young investigators fellowships, and the genomic infrastructure. The “Nerina and Mario Mattioli” Foundation supported the activity of Pandora tumour tissue collection. SM is supported by The Italian Association for Cancer Research (grant number IG19997).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.adcanc.2023.100088.

References

[1] U.A. Matulonis, A.K. Sood, L. Fallowfield, B.E. Howitt, J. Sehouli, B.Y. Karlan, Ovarian cancer, Nat. Rev. Dis. Prim. 2 (2016) 1–22, https://doi.org/10.1038/nrdp.2016.61.

[2] Z.C. Wang, N.J. Birkbak, A.C. Culhane, R. Drapkin, A. Fatima, R. Tian, M. Schwede, K. Alsop, K.E. Daniels, H. Piao, J. Liu, D. Etemadmoghadam, A. Miron, H. B. Salvesen, G. Mitchell, A. DeFazio, J. Quackenbush, R.S. Berkowitz, J.D. Iglehart, D.D.L. Bowtell, U.A. Matulonis, Profiles of genomic instability in high-grade serous ovarian cancer predict treatment outcome, Clin. Cancer Res. 18 (2012) 5806–5815, https://doi.org/10.1158/1078-0432.CCR-12-0857.

[3] I. Meinhold-Heerlein, S. Hauptmann, The heterogeneity of ovarian cancer, Arch. Gynecol. Obstet. 289 (2014) 237–239, https://doi.org/10.1007/s00404-013-3114-3.

[4] A. Bashashati, G. Ha, A. Tone, J. Ding, L.M. Prentice, A. Roth, J. Rosner, K. Shumansky, S. Kalloger, J. Senz, W. Yang, M. McConechy, N. Melnyk, M. Anglesio, M.T.Y. Luk, K. Tse, T. Zeng, R. Moore, Y. Zhao, M.A. Marra, B. Gilks, S. Yip, D.G. Huntsman, J.N. McAlpine, S.P. Shah, Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling, J. Pathol. 231 (2013) 21–34, https://doi.org/10.1002/path.4230.

[5] L. Paracchini, L. Mannarino, I. Craparotta, C. Romualdi, R. Fruscio, T. Grassi, V. Fotia, G. Caratti, P. Perego, E. Calura, L. Clivio, M. D'Incalci, L. Beltrame, S. Marchini, L. Paracchini, L. Mannarino, I. Craparotta, C. Romualdi, R. Fruscio, T. Grassi, V. Fotia, G. Caratti, P. Perego, E. Calura, L. Clivio, M. D'Incalci, L. Beltrame, S. Marchini, Regional and temporal heterogeneity of epithelial ovarian cancer tumor biopsies: implications for therapeutic strategies, Oncotarget 5 (2016), https://doi.org/10.18632/oncotarget.10505.

[6] A. Salomon-Perzyński, M. Salomon-Perzyńska, B. Michalski, V. Skrzypulec-Plinta, High-grade serous ovarian cancer: the clone wars, Arch. Gynecol. Obstet. (2017), https://doi.org/10.1007/s00404-017-4292-1.

[7] E.L. Christie, D.D.L. Bowtell, Acquired chemotherapy resistance in ovarian cancer, Ann. Oncol. 28 (2017), https://doi.org/10.1093/annonc/mdx446 viii13–viii15.

[8] S. Kim, Y. Han, S.I. Kim, H.-S. Kim, S.J. Kim, Y.S. Song, Tumor evolution and chemoresistance in ovarian cancer, Npj Precis. Oncol. 2 (2018) 1–9, https://doi.org/10.1038/s41698-018-0063-0.

[9] L. Mittempergher, Genomic characterization of high-grade serous ovarian cancer: dissecting its molecular heterogeneity as a road towards effective therapeutic strategies, Curr. Oncol. Rep. 18 (2016), https://doi.org/10.1007/s11912-016-0526-9.

[10] S. Ballabio, I. Craparotta, L. Paracchini, L. Mannarino, S. Corso, M.G. Pezzotta, M. Vescio, R. Fruscio, C. Romualdi, E. Dainese, L. Ceppi, E. Calura, S. Pileggi, G. Siravegna, L. Pattini, P. Martini, M. delle Marchette, C. Mangioni, A. Ardizzoia, A. Pellegrino, F. Landoni, M. D'Incalci, L. Beltrame, S. Marchini, Multisite analysis of high-grade serous epithelial ovarian cancers identifies genomic regions of focal and recurrent copy number alteration in 3q26.2 and 8q24.3, Int. J. Cancer 145 (2019) 2670–2681, https://doi.org/10.1002/ijc.32288.

[11] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T. M. Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (2016) e71, https://doi.org/10.1093/nar/gkv1507.

[12] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193, https://doi.org/10.1093/bioinformatics/19.2.185.

[13] C.H. Mermel, S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (2011) R41, https://doi.org/10.1186/gb-2011-12-4-r41.

[14] R.W. Tothill, A.V. Tinker, J. George, R. Brown, S.B. Fox, S. Lade, D.S. Johnson, M. K. Trivett, D. Etemadmoghadam, B. Locandro, N. Traficante, S. Fereday, J.A. Hung, Y.E. Chiew, I. Haviv, D. Gertig, A. Defazio, D.D.L. Bowtell, Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome, Clin. Cancer Res. 14 (2008) 5198–5208, https://doi.org/10.1158/1078-0432.CCR-08-0196.

[15] Cancer Genome Atlas Research Network, D. Bell, a. Berchuck, M. Birrer, J. Chien, D.W. Cramer, F. Dao, R. Dhir, P. DiSaia, H. Gabra, P. Glenn, a.K. Godwin, J. Gross,

L. Hartmann, M. Huang, D.G. Huntsman, M. Iacocca, M. Imielinski, S. Kalloger, B. Y. Karlan, D.a. Levine, G.B. Mills, C. Morrison, D. Mutch, N. Olvera, S. Orsulic, K. Park, N. Petrelli, B. Rabeno, J.S. Rader, B.I. Sikic, K. Smith-McCune, a.K. Sood, D. Bowtell, R. Penny, J.R. Testa, K. Chang, H.H. Dinh, J.a. Drummond, G. Fowler, P. Gunaratne, a.C. Hawes, C.L. Kovar, L.R. Lewis, M.B. Morgan, I.F. Newsham, J. Santibanez, J.G. Reid, L.R. Trevino, Y.-Q. Wu, M. Wang, D.M. Muzny, D. a. Wheeler, R.a. Gibbs, G. Getz, M.S. Lawrence, K. Cibulskis, a.Y. Sivachenko, C. Sougnez, D. Voet, J. Wilkinson, T. Bloom, K. Ardlie, T. Fennell, J. Baldwin, S. Gabriel, E.S. Lander, L. Ding, R.S. Fulton, D.C. Koboldt, M.D. McLellan, T. Wylie, J. Walker, M. O'Laughlin, D.J. Dooling, L. Fulton, R. Abbott, N.D. Dees, Q. Zhang, C. Kandoth, M. Wendl, W. Schierding, D. Shen, C.C. Harris, H. Schmidt, J. Kalicki, K.D. Delehaunty, C.C. Fronick, R. Demeter, L. Cook, J.W. Wallis, L. Lin, V. J. Magrini, J.S. Hodges, J.M. Eldred, S.M. Smith, C.S. Pohl, F. Vandin, B.J. Raphael, G.M. Weinstock, E.R. Mardis, R.K. Wilson, M. Meyerson, W. Winckler, R.G. W. Verhaak, S.L. Carter, C.H. Mermel, G. Saksena, H. Nguyen, R.C. Onofrio, D. Hubbard, S. Gupta, a. Crenshaw, a.H. Ramos, L. Chin, a. Protopopov, J.J. J. Zhang, T.M. Kim, I. Perna, Y. Xiao, H. Zhang, G. Ren, N. Sathiamoorthy, R. W. Park, E. Lee, P.J. Park, R. Kucherlapati, D.M. Absher, L. Waite, G. Sherlock, J. D. Brooks, J.Z. Li, J. Xu, R.M. Myers, P.W. Laird, L. Cope, J.G. Herman, H. Shen, D. J. Weisenberger, H. Noushmehr, F. Pan, T. Triche Jr., B.P. Berman, D.J. Van Den Berg, J. Buckley, S.B. Baylin, P.T. Spellman, E. Purdom, P. Neuvial, H. Bengtsson, L.R. Jakkula, S. Durinck, J. Han, S. Dorton, H. Marr, Y.G. Choi, V. Wang, N. J. Wang, J. Ngai, J.G. Conboy, B. Parvin, H.S. Feiler, T.P. Speed, J.W. Gray, N. D. Socci, Y. Liang, B.S. Taylor, N. Schultz, L. Borsu, a.E. Lash, C. Brennan, a. Viale, C. Sander, M. Ladanyi, K.a. Hoadley, S. Meng, Y. Du, Y. Shi, L. Li, Y.J. Turman, D. Zang, E.B. Helms, S. Balu, X. Zhou, J. Wu, M.D. Topal, D.N. Hayes, C.M. Perou, C.J. Wu, S. Shukla, R. Jing, Y. Liu, M. Noble, H. Carter, D. Kim, R. Karchin, J. E. Korkola, L.M. Heiser, R.J. Cho, Z. Hu, E. Cerami, a. Olshen, B. Reva, Y. Antipin, R. Shen, P. Mankoo, R. Sheridan, G. Ciriello, W.K. Chang, J.a. Bernanke, D. Haussler, C.C. Benz, J.M. Stuart, S.C. Benz, J.Z. Sanborn, C.J. Vaske, J. Zhu, C. Szeto, G.K. Scott, C. Yau, M.D. Wilkerson, N. Zhang, R. Akbani, K.a. Baggerly, W. K. Yung, J.N. Weinstein, T. Shelton, D. Grimm, M. Hatfield, S. Morris, P. Yena, P. Rhodes, M. Sherman, J. Paulauskis, S. Millis, a. Kahn, J.M. Greene, R. Sfeir, M. a. Jensen, J. Chen, J. Whitmore, S. Alonso, J. Jordan, a. Chu, a. Barker, C. Compton, G. Eley, M. Ferguson, P. Fielding, D.S. Gerhard, R. Myles, C. Schaefer, K.R. Mills Shaw, J. Vaught, J.B. Vockley, P.J. Good, M.S. Guyer, B. Ozenberger, J. Peterson, E. Thomson, Integrated genomic analyses of ovarian carcinoma, Nature 474 (2011) 609–615, https://doi.org/10.1038/nature10166.

[16] X. Chen, J. Li, W.H. Gray, B.D. Lehmann, J.A. Bauer, Y. Shyr, J.A. Pietenpol, TNBCtype: a subtyping tool for triple-negative breast cancer, Cancer Inf. 11 (2012) 147–156, https://doi.org/10.4137/CIN.S9983.

[17] U.D. Akavia, O. Litvin, J. Kim, F. Sanchez-garcia, H.C. Causton, P. Pochanard, E. Mozes, L. a Garraway, An Integrated Approach to Uncover Drivers of Cancer 143 (2011) 1005–1017, https://doi.org/10.1016/j.cell.2010.11.013.An.

[18] E. Segal, D. Pe'er, A. Regev, D. Koller, N. Friedman, Learning module networks, J. Mach. Learn. Res. 6 (2005) 557–588, https://doi.org/10.1016/j.febslet.2004.11.019.

[19] A. Annest, R.E. Bumgarner, A.E. Raftery, K.Y. Yeung, Iterative Bayesian Model Averaging: a method for the application of survival analysis to high-dimensional microarray data, BMC Bioinf. 10 (2009) 72, https://doi.org/10.1186/1471-2105-10-72.

[20] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. Del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. Van Roey, G. Cesareni, H. Hermjakob, The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) 358–363, https://doi.org/10.1093/nar/gkt1115.

[21] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0, Bioinformatics (2011), https://doi.org/10.1093/bioinformatics/btr260.

[22] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. USA 102 (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[23] M. Ono, C.A. Inkson, R. Sonn, T.M. Kilts, L.F. de Castro, A. Maeda, L.W. Fisher, P. G. Robey, A.D. Berendsen, L. Li, N. McCartney-Francis, A.C. Brown, N.P. S. Crawford, A. Molinolo, A. Jain, N.S. Fedarko, M.F. Young, WISP1/CCN4: a potential target for inhibiting prostate cancer growth and spread to bone, PLoS One (2013), https://doi.org/10.1371/journal.pone.0071709.

[24] P.O. Gaudreau, S. Clairefond, C.A. Class, P.L. Boulay, P. Chrobak, B. Allard, F. Azzi, S. Pommey, K.A. Do, F. Saad, D. Trudel, M. Young, J. Stagg, WISP1 is associated to advanced disease, EMT and an inflamed tumor microenvironment in multiple solid tumors, OncoImmunology (2019), https://doi.org/10.1080/2162402X.2019.1581545.

[25] T. Suzuki, S. Inoue, W. Kawabata, J. Akahira, T. Moriya, F. Tsuchiya, S. Ogawa, M. Muramatsu, H. Sasano, EBAG9/RCAS1 in human breast carcinoma: a possible factor in endocrine-immune interactions, Br. J. Cancer (2001), https://doi.org/10.1054/bjoc.2001.2176.

[26] J.I. Akahira, M. Aoki, T. Suzuki, T. Moriya, H. Niikura, K. Ito, S. Inoue, K. Okamura, H. Sasano, M. Yaegashi, Expression of EBAG9/RCASI is associated with advanced disease in human epithelial ovarian cancer, Br. J. Cancer (2004), https://doi.org/10.1038/sj.bjc.6601832.

[27] G. Xie, H. Yang, D. Ma, Y. Sun, H. Chen, X. Hu, Y. Jiang, Z. Shao, Integration of whole-genome sequencing and functional screening identifies a prognostic signature for lung metastasis in triple-negative breast cancer, Breast (2019), https://doi.org/10.1016/s0960-9776(19)30312-1.

[28] M. Ladanyi, M.Y. Lui, C.R. Antonescu, A. Krause-Boehm, A. Meindl, P. Argani, J. H. Healey, T. Ueda, H. Yoshikawa, A. Meloni-Ehrig, P.H.B. Sorensen, F. Mertens, N. Mandahl, H. Van Den Berghe, R. Sciot, P. Dal Cin, J. Bridge, The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL, a novel gene at 17q25, Oncogene 20 (2001) 48–57, https://doi.org/10.1038/sj.onc.1204074.

[29] E. Pikarsky, R.M. Porat, I. Stein, R. Abramovitch, S. Amit, S. Kasem, E. Gutkovich-Pyest, S. Uriell-Shoval, E. Galun, Y. Ben-Neriah, NF-κB functions as a tumour promoter in inflammation-associated cancer, Nature (2004), https://doi.org/10.1038/nature02924.

[30] J. Jen, Y.C. Wang, Zinc finger proteins in cancer progression, J. Biomed. Sci. (2016), https://doi.org/10.1186/s12929-016-0269-9.

[31] K. Rajalingam, R. Schreck, U.R. Rapp, Š. Albert, Ras oncogenes and their downstream targets, Biochim. Biophys. Acta Mol. Cell Res. 1773 (2007) 1177–1195, https://doi.org/10.1016/j.bbamcr.2007.01.012.

[32] J.W. Peck, M. Oberst, K.B. Bouker, E. Bowden, P.D. Burbelo, The RhoA-binding protein, rhophilin-2, regulates actin cytoskeleton organization, J. Biol. Chem. 277 (2002) 43924–43932, https://doi.org/10.1074/jbc.M203569200.