Università degli Studi di Padova

Head Office: Università degli Studi di Padova

Department of Biology

Ph.D. COURSE IN: Biosciences

CURRICULUM: Genetics, Genomics and Bioinformatics

SERIES XXXII

A bioinformatic approach to define transcriptome alterations in platinum resistance ovarian cancers

Thesis written with the financial contribution of Associazione Italiana per la Ricerca sul Cancro (AIRC)

Coordinator: Prof. Ildikò Szabò

Supervisor: Prof. Chiara Romualdi

Ph.D. student: Giuseppe Benvenuto

INDEX

ABSTRACT

Epithelial ovarian cancer (EOC) is the most lethal gynecological malignancy due to its diagnosis at advanced stages, when the disease has already spread beyond the ovaries. EOC is generally sensitive to first line chemotherapy, and the vast majority of patients respond to platinum (Pt)-based therapy after debulking surgery.

Unfortunately, more than 80% of Pt-responsive patients relapse with a disease that progressively becomes Pt-resistant. Based mainly on clinical evidence, the process by which disease relapses is still poorly understood. The aim is to identify biomarkers of sensitivity to chemotherapy and therapeutic targets in HGS-EOC by integrating transcriptomic data, coding and non-coding RNAs. The bioinformatic analysis was applied on microarray data and RNA-seq data, embracing different classes of patients (resistant, sensitive, partially sensitive and normal).

Two complementary approaches have been adopted to identify biomarkers of therapy response in microarray data: i) a classic approach and ii) a network-based approach using micrographite. The results obtained with both procedures have then been used to reconstruct a regulatory circuit involved in therapy response. The final outcome is a regulatory cell signal pathway composed of genes and miRNAs mainly involved in the therapy response. Circuit has been validated using two external and independent cohorts by quantitative real-time PCR (qRT-PCR). However, in order to complete the characterization of network as prognostic factor we decided to consider in survival analysis defect of the Homologous Recombination (HR). Approaching in survival analysis, a signature of three genes (*SDF2L1*, *PPP1R12A* and *PRKG1*) found to be independent prognostic biomarkers, was able to predict, at the time of diagnosis, resistance to Pt-based chemotherapy.

Also, a new approach has been evaluated in order to characterize new mechanisms of chemotherapy resistance in ovarian cancers. On microarray data, we tried to stratify patients for the immunotherapy, with recent improved understanding of the immune recognition and regulation of cancer cells. In addition, using RNA-seq data and somatic DNA mutations, we went deeper in immunogenicity of ovarian cancer trying to find new elements as therapy targets, neoantigens, not associated to this tumor till now.

At last, in addition, the small amount of molecular differences between Pt-r and Pt-s patients suggested the presence of potential new transcripts involved in therapy response maybe due to aberrant splicing events. To investigate this hypothesis, we used a set of RNA-seq experiments, to identify new aberrant splicing such as circular RNAs. We reported 5 circRNAs differentially expressed between tumour resistance types, and a large number of class-specific circRNAs. In particular, circ_BARD1 showed a

character as prognostic factor significative in OS and PFS, in multivariate analysis with residual tumour and age as covariates. The consistency of circular RNA expression, in conjunction with the regulatory circuit, may offer new candidates for cancer treatment and prognosis, revealing that the integration of coding and non-coding RNAs data may shed light on chemotherapy resistance mechanisms in ovarian cancer.

1. INTRODUCTION

Ovarian cancer ranks 7th in both incidence and mortality among women worldwide. In 2018, approximately 295,000 women were diagnosed with ovarian cancer and 184,000 women died, representing 2% of all new cancer diagnoses and 4% of cancer deaths among women [1].

Amongst all gynecological cancers, ovarian cancer is the most difficult in diagnosis and treatment. In early stages it is almost symptomless, consequentially is often diagnosed by chance with screening exams (e.g. echography or laparoscopy). Moreover, when it is already in an advanced stage gives rise to common symptoms, such as persistent abdominal pain, bloating or decreased appetite, making this type of gynecological malignancy hard to be diagnosed also in advanced stages [2]. This phenomenon is also due to the fact that our current early diagnostic tools (biomarkers) are extremely limited. Its high mortality rate has made it one of the most investigated fields in gynecological oncology. Even though the evolution of modern medicine, new diagnostic tools and the use of next generation sequence technologies, our arsenal of weapons against diagnosing ovarian cancer is extremely poor.

One reason may be founded in ovarian cancer histology, it has been recognized as not a single disease but composed of various histologically different tumor types [3]. Ovarian cancers have generally been divided into epithelial and non-epithelial groups for many years. Epithelial ovarian cancer (EOC) is the most common type of ovarian tumor, making up 90% of all primary ovarian cancer. Recent studies have contributed to establish this subdivision of epithelial cancers into different groups according to a combination of morphological, molecular and clinical characteristics. The histological subtypes of epithelial tumors include: serous, endometrioid, clear cell, mucinous, transitional cell, squamous and undifferentiated carcinoma. Each histological subtype has specific chemosensitivity, pattern of metastasis and probability of survival, influenced by distinct molecular pathways.

Serous carcinoma is the most common histological subtype and respectively around 50-60% and 66-90% of ovarian cancer and fallopian tube carcinomas are serous tumors [4]. Non-serous peritoneal carcinoma is extremely uncommon. The tumors originated in ovary, tubes and peritoneum are divided into low-grade and high-grade serous carcinoma depending by the degree of differentiation. About tumor grade, except for serous and undifferentiated carcinoma, around 85% of the other histotypes are stage I or II at diagnosis [5].

Table 1. *Principal characteristics and differences between Low-grade and High-grade serous ovarian cancer.*

	Low-grade serous ovarian carcinoma	**High-grade serous ovarian carcinoma**
Clinico-pathologic category	Type I	Type II
Mean age at diagnosis	55.5 years	62.6 years
Proportion of serous epithelial cancers	10%	90%
Possible origin	Salpingeal epithelium, Cortical inclusion cysts, Müllerian metaplasia of ovarian surface epithelium	Salpingeal epithelium, Cortical inclusion cysts, Müllerian metaplasia of ovarian surface epithelium
Presumed precursor	Serous borderline tumor	Serous tubular intraepithelial carcinoma (STIC) of the fallopian tube
Major risk factors	Advanced borderline tumor, Ovulation-inducing drugs	Low parity, BRCA mutations
Molecular features	KRAS/BRAF mutations	TP53 mutations
Clinical course	Slow-growing	Evolves rapidly

The stage of disease is important for an accurate treatment of ovarian cancer. When examining cancer survival, stage at diagnosis is a crucial aspect to be considered, and the staging system should not allow for changes in stage after biopsy or initial treatment. The stages describe the extent of the spread of the cancer and is based on the location and size of the primary tumor, on the involvement of lymph node and on the presence of metastasis at diagnosis. Unless the disease is advanced at diagnosis, ovarian cancer is generally staged through surgery and pathological analysis of tissue samples of the tumor. Three main staging systems for ovarian cancer are used: the Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) system [6], the Union for International Cancer Control's (UICC) Tumor Node Metastasis (TNM) system and the Surveillance [7], Epidemiology and End Results (SEER) program's

Summary Stage 2000 [8]. The most common classification system is the FIGO system, that is applicable to all histological subtypes of ovarian cancer and is based on the degree of dissemination of the epithelial ovarian cancer at diagnosis. In order to obtain a precise diagnosis, laparoscopic surgery usually is performed to get a tumor sample for biopsy and to aid in the staging of the disease.

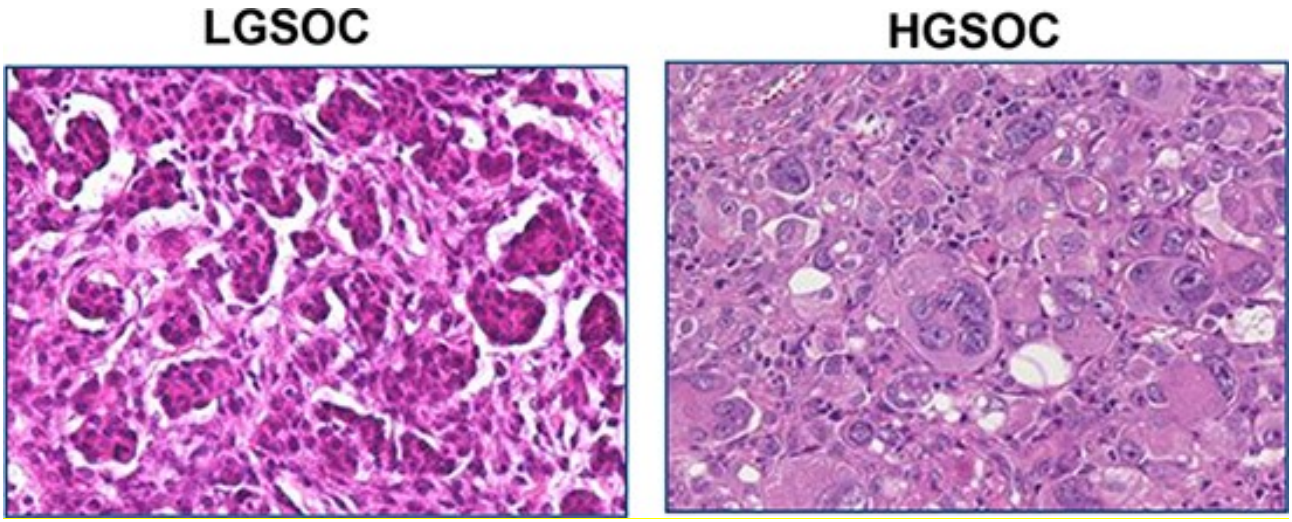
At stage I, the cancer still is confined to the ovaries or fallopian tubes. In the stage II, the disease already has invaded other pelvic areas and organs such as the uterus. Stage III tumor has spread beyond the pelvic area to organs or tissues on the inside of the peritoneal cavity or invaded the lymph nodes. In the stage IV tumor is grown beyond the peritoneal cavity, toward the lungs and involving inguinal and other extra-abdominal lymph nodes [6].

Instead, the grade is a description of how the cancer cells look compared to normal cells and their growth rate. To find out the grade of ovarian cancer, the pathologist looks at a tissue sample from the tumor under a microscope. Serous tumors of ovarian, tubal and peritoneal origin are divided into low-grade and high-grade serous carcinoma depending upon the degree of differentiation (Table 1) [3].

In low grade, cells are well differentiated, and, from a histological point of view, they look almost like normal cells. Lower grade cancer cells tend to be slow-growing and are less likely to spread. Instead, in high grade commonly cancer cells are poorly differentiated or undifferentiated. They look less normal, or even more abnormal, and tend to grow more quickly and are more likely to spread than low-grade cancer cells [9].

Also, the progression from low-grade to high-grade occurs rarely. Low-grade serous ovarian carcinomas (LGSC) are quite rare (5% of serous ovarian carcinomas). Low-grade serous fallopian tube carcinoma is also very rare. High-grade serous ovarian carcinoma (HGSC) is the most common subtype of epithelial ovarian cancer, typically present in women at older ages compared to women with low-grade tumors. Most serous peritoneal tumors are high-grade and are similar to high-grade serous ovarian tumors (**Figure 1**).

Figure 1. *Histological features of Low-grade serous ovarian cancer (LGSOC) and High-grade serous ovarian cancer (HGSOC)* [10]



The aspects that have to be considered are: i) most of diagnosed ovarian cancer are serous types, ii) the high grade is the most common in serous ovarian cancer, iii) women with serous low-grade tumors have higher survival than women diagnosed with serous high-grade tumors, iv) HGSC commonly relapses and v) after relapse HGSC becomes resistant to chemotherapy. All of these reasons make the serous histotype the main focus in the field of precision medicine for ovarian cancer and the target of my work.

2. HIGH-GRADE SEROUS OVARIAN CANCER

2.1 EPIDEMIOLOGY AND RISK FACTORS

The distribution of ovarian cancer incidence worldwide is not even, with a substantial variation based on geography, ethnicity and the level of economic development. The highest incidence is in Northern and Central/Eastern Europe and lowest in Asia and Africa [11]. EOC, like most other epithelial cancers, tends to be diagnosed more frequently in menopause age. In fact, the median age at diagnosis is 63 years [12]. During their lifetime, women have about 1.3% of risk to develop ovarian cancer. However, a number of well-known risk-factors may modify the percentage of risk in individuals [11]. For example,

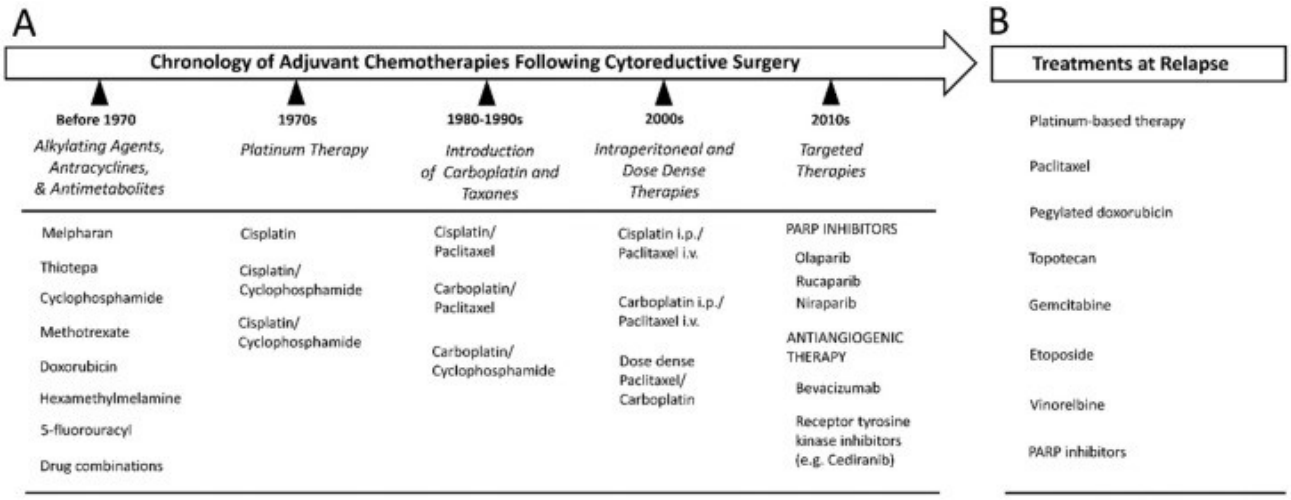
there is a relevant heritable component of risk due to genetic factors. Familiar cases usually are found to be due to germline mutations in the tumor-suppressor genes *BRCA1* and *BRCA2*, which also contribute to increase the risk of developing breast cancer in these subjects [11]. Recent studies found that 3.6% of ovarian cancer patients have germline mutations in BRCA1 while 3.3% have germline mutations in *BRCA2* and all of these are high-grade serous carcinomas [13]. The contribution of high-penetrance alleles of *BRCA1/2* is only a small part of the heritable component of ovarian cancer [14]. Many other genes with low penetrance mutations can have a role in heritable ovarian cancer. For instance, other gene alterations that have been linked with greater risk include *BARD1, CHEK2, MRE11A, RAD50, PALB2* and *ATM* [15, 16]. The common link between all these genes is their role in the homologous recombination (HR)-mediated pathway of DNA repair, which is known to play an important role in the pathophysiology of HGSOC.

Furthermore, individual's risk of developing ovarian cancer can be influenced by many modifiable or lifestyle factors. For example, women who have given birth have a risk reduction of 10–20% than other women, associated with each additional birth [17]. Studies have also found that women who breastfeed have lower risk, such as the women using hormone containing oral contraceptives, they have up to 30% lower risk compared to never-users. Other potential risk factors include obesity, diabetes and smoking [18].

2.3 CHEMOTHERAPY

Historically, ovarian cancer was one of the first malignancies to be successfully treated with cytotoxic chemotherapy (**Figure 2**). The first class of chemotherapeutic drugs to be developed were the alkylating agents, which were introduced in the 1950s [19]. The effect is usually sufficient to inhibit proper DNA synthesis. Although many of these drugs demonstrated good single-agent activities in the treatment of ovarian cancer, it was ascertained promptly that the most effective strategy would be to employ these agents in combination, based on the theory that multiple drugs, each with different mechanisms of action, would behave synergistically and reduce the risk of the disease acquiring chemoresistance. Following a temporal progression, in the 1970s, many combinations were in use for the treatment of ovarian cancer, with the most popular protocol consisting of the use of cyclophosphamide and doxorubicin, in addition to methotrexate and 5-fluorouracil [19].

Figure 2. *A) Timeline of chemotherapy agents in ovarian cancer. B) Chemotherapy drugs used in ovarian cancer second line treatments* [19].



Hereafter, cisplatin drug has been approved for clinical use. It was incorporated into primary chemotherapeutic regimens either singly or in combination with cyclophosphamide, doxorubicin or hexamethyl melamine.

The late 1980s saw the introduction of a new platinating agent in the form of carboplatin. A set of trials results, completed in 1992, convinced clinicians to begin replacing cisplatin with carboplatin [19]. In the late 1980s also has been introduced a new class of drugs, the taxanes, of which paclitaxel is the prototype. As a result of taxanes applications in ovarian cancer, a landmark clinical trial evaluated the effectiveness of cisplatin-paclitaxel combination therapy versus cisplatin-cyclophosphamide. Results indicated that cisplatin-paclitaxel combination therapy was able to significantly improve objective response rates, progression free survival and overall survival. This combination has been, for the last 20 years, the standard of care for the treatment of ovarian cancer [19].

Despite recent efforts to improve survival by combining standard chemotherapy with novel targeted agents such as anti-angiogenic drugs and PARP inhibitors, the overall survival (OS) at five years is about 30% [20] and it has remained almost unchanged for over three decades [21]. The clinical response to first line chemotherapy is heterogeneous and not predictable at the time of diagnosis. Nevertheless, a significant fraction of HGSOC patients at advanced stage are initially sensitive to Pt-based therapy and generally relapse after 12 months from the end of chemotherapy (Platinum-sensitive, Pt-s cases) [22]. Almost 20% of HGSOC patients are intrinsically resistant to Pt-based chemotherapy, which means they do not respond at all or relapse during first line treatment or within 6 months beyond the end of Pt-based therapy (Pt-resistant, Pt-r). For these individual efficacious alternative therapies do not exist, and

death occurs rapidly. Prognostic clinical factors, including age, FIGO stage, tumor grade, subtype, and residual tumor (RT), are insufficient to capture individual variations predicting response to chemotherapy.

For the subset of patients whose disease is judged to be refractory to the front-line chemotherapy, alternative or second-line drug combinations may be utilized in an attempt to elicit an objective response. Typically, "no-refractory" patients are re-treated with the standard platinum-based chemotherapeutic regimen [23]. The decision to re-use platinum is complicated by the presence of persistent side-effects from previous treatment, such as neuropathy and pancytopenia, as well as the potential for life-threatening platinum hypersensitivity reactions. Approximately 50% of patients possess recurrent disease that is still responsive to re-treatment with platinum, although the progression-free survival decreases with each successive platinum therapy [23]. Platinum resistant patients, relapsing with a tumor that is platinum-resistant, can be curated with an alternative treatment modality, including the use of doxorubicin, topotecan, gemcitabine, etoposide and vinorelbine. The average response rate to this kind of salvage therapy is only about 10–15% with a median progression-free survival of 3–4 months [23].

2.3 TARGETED THERAPIES

The aim of targeted therapy is targeting only those pathways known to be activated in the context of cancer cells, avoiding the cytotoxic effect on normal cells. Unfortunately, in ovarian cancer this approach has not been incisive yet, with only a few new treatments reaching the clinic with marginal improvements in outcome. This is partly related to the molecular biology of HGSOC which does not often present with many oncogenic alterations that can be targeted easily with small-molecular inhibitors [24]. One of the more efficient targeted approaches relates to a class of drugs known as *PARP* inhibitors. HGSOC is characterized by an important genomic instability and the majority of patients has some deficiency in DNA repair pathways, *BRCA1* and *BRCA2*. Cancer cells are over-reliant on the poly (ADP-Ribose) polymerase (*PARP*) mediated base excision repair (BER) of single-strand breaks to resolve spontaneous DNA damage, in patients with a deficiency in homologous recombination. Therefore, drugs targeting *PARP* would be expected to have significant anti-tumor activity in these patients [24, 25]. The first *PARP* inhibitor to be tested in patients with HGSOC was olaparib. Its use has primarily been considered as a treatment for recurrent disease or as a maintenance therapy to prolong progression-free survival and showed an impressive clinical response in patients with recurrent HGSOC

with *BRCA* mutations [26]. Recently, the FDA approved the use of two other PARP inhibitors, rucaparib and niraparib, to treat patients with relapsed ovarian cancer, regardless of BRCA-mutation status or platinum-sensitivity [27].

Another therapy being investigated involves targeting the tumor microenvironment through the use of anti-angiogenic agents such as bevacizumab. This humanized monoclonal antibody targets the cytokine *VEGF-A*, which influences the recruitment of blood vessels to the tumor, something required for its growth [26]. This agent showed an increase in progression-free survival with the addition of bevacizumab to the standard carboplatin-paclitaxel regimen as maintenance therapy. Two other trials in the context of platinum-sensitive and platinum-resistant [28] disease have shown that the addition of bevacizumab to cytotoxic chemotherapy improves progression-free survival. The use of bevacizumab in the context of advanced EOC carries the risk of significant adverse effects [23]. Other anti-angiogenic agents currently are being tested and the most promising among these is probably cediranib, which has been shown to carry out significant single-agent activity in the context of both platinum-sensitive and resistant-relapsed disease [23].

Another potential targeted therapy is the inhibition of *AKT* signaling. One recent study reported promising clinical activity of an oral *AKT* inhibitor (afuresertib) in combination with carboplatin and paclitaxel in platinum-resistant ovarian cancer [29].

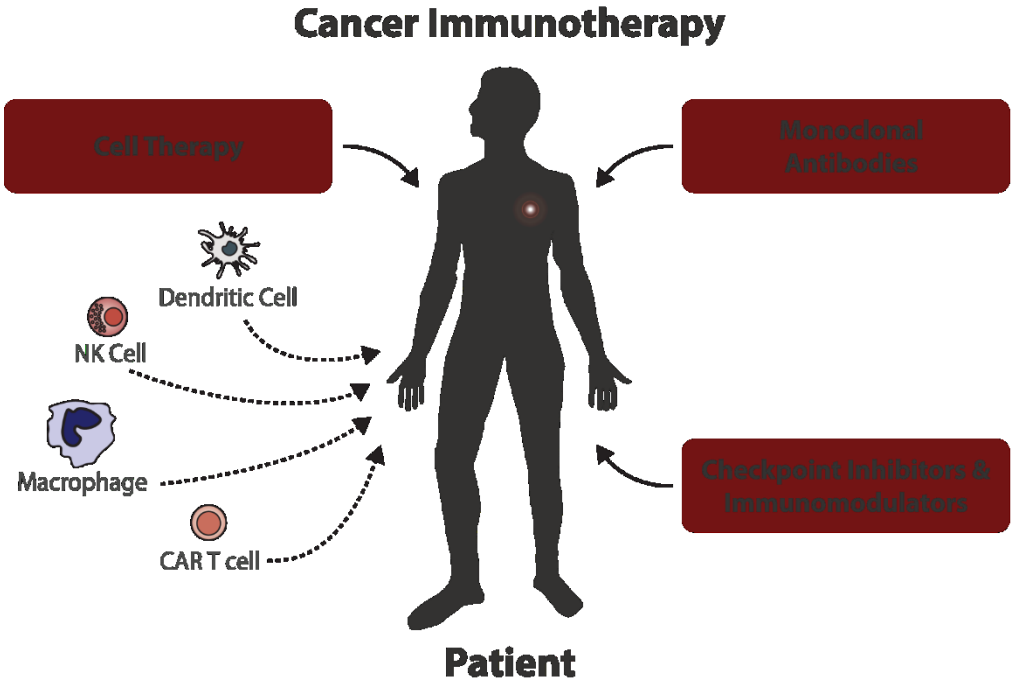
3. IMMUNOTHERAPY

3.1 INTRODUCTION OF IMMUNOTHERAPY

The results of recent clinical trials suggest that a plateau has been reached for conventional therapies as there is no definitive increase in overall survival, pushing towards new treatment strategies. Immunotherapy, with recent improved understanding of the immune recognition and regulation of cancer cells, has attracted significant interest (**Figure 3**). These interventions include cancer vaccines, immune checkpoint blockade, neoantigens and adoptive cell therapy. Indeed, several immune checkpoint inhibitors were recently approved by the Food and Drug Administration for a variety of cancers including melanoma, non-small-cell lung cancer, bladder cancer and classical Hodgkin lymphoma. Unfortunately, there are currently no approved immune therapies for ovarian cancer [30]. Several studies, in order to understand how the immune system interact with ovarian cancer cells,

demonstrated that the presence of tumor-infiltrating lymphocytes (TILs) is associated with a better clinical outcome in ovarian cancer patients [31]. Also, they confirmed that a lack of TILs is significantly associated with a worse survival [31].

Figure 3. *Overview of cancer immunotherapy, TILs and putative therapy applications.*



Together, these studies support the hypothesis that tumor infiltration by lymphocytes is related to the tumor-related immune response. The immunosuppressive tumor microenvironment is the key to obtain for a successful deployment of cancer immunotherapy for ovarian cancer patients. Tumor-specific T cells, even if in large number, may not able to destroy tumor cells in vivo. This immunosuppressive network facilitates tumor progression, serving as an important obstacle to the implementation of efficacious immunotherapeutic strategies. In fact, the major direction is to develop biomarkers able to predict responsiveness to different types of immunotherapy and to allow right combination minimizing toxicities.

3.1.1 CHECKPOINT BLOCKERS

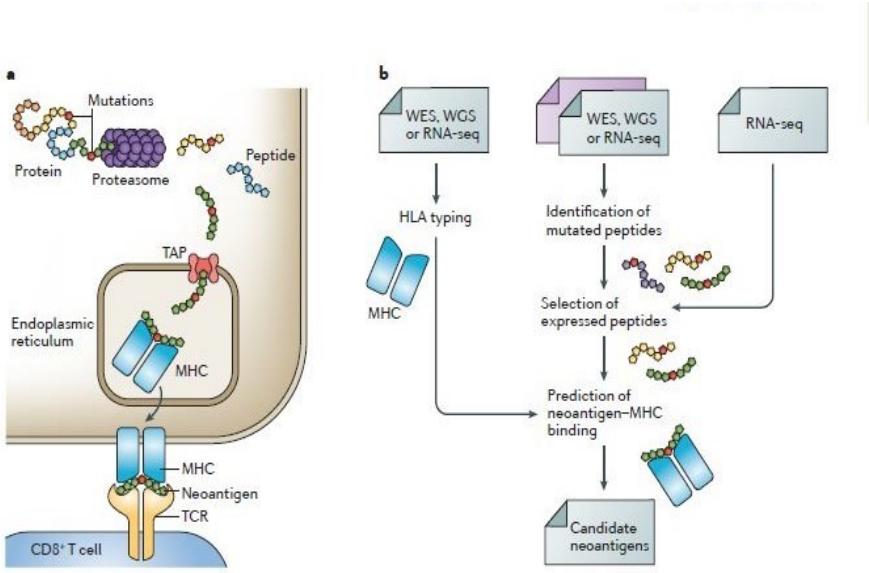
Several works have demonstrated that T-cell expression of inhibitory immune checkpoint receptors is one mechanism by which tumors evade host immune response. These receptors negatively regulate T-cell function and include *CTLA-4, PD-1, LAG-3, TIM-3, TIGIT*, and others. Therapies with a combination of *CTLA-4* and *PD-1* have demonstrated clinical benefit in some human cancers [30]. *CTLA-4* regulates T-cell priming and activation, including tumor-specific T cells and *PD-1* is a cell surface receptor that interacts with two principal ligands, *PD-L1* and *PD-L2*, resulting in the inhibition of T-cell signaling and cytokine production. Consequently, anti-*CTLA-4* therapies are associated with more significant immune-related toxicities compared with *PD-1* blockade. Using specific antibodies is possible to block these inhibitory receptors to reinstate an existing antitumor response. The general strategy provides for the inhibition of both the immunosuppressive receptors expressed by activated T lymphocytes, and the principal ligands of these receptors, such as *PD-L1* [32]. Several antibodies directed against *PD-1, PD-L1,* and *CTLA-4* have been developed and are being tested clinically in patients with ovarian cancer. Moreover, these antibodies have highlighted significant results in mediating tumor regression in melanoma and other solid tumors, in ovarian cancer the response rates have been modest.

3.2 TUMOR NEOANTIGENS AND CANCER VACCINES

Cancer vaccines have been studied in order to generate effector T cells with the ability to stimulate an immune response that will recognize and eliminate cancer cells early enough to prevent tumor progression. One of the major questions is identifying the most effective and safe vaccine targets in ovarian cancer. In detail, T cells can recognize tumor antigens expressed by cancer cells. A class of tumor-associated antigens, is expressed at low levels in some normal tissues but in malignant cells is over-expressed [33].

The second class of antigens is tumor-specific neoantigens, which are generated from somatic mutations that alter amino acid coding sequences (**Figure 4**). This process produces mutated peptides that can be expressed and presented on cell surface and recognized by T cells. Normal tissues do not possess these mutations, so the neoantigen-specific T cells are not able to induce normal tissue destruction. As a result, neoantigens appear to represent ideal targets for T cell-based cancer immunotherapy [30].

Figure 4. *a) Pathway of neoantigens genesis arising from protein mutated. b) Generic bioinformatic pipeline to identify candidate neoantigens.*



There are three types of tumor antigens that can be recognized as immune targets by T-cells: cancer testis antigens (CT), tumor-associated antigens (TA), and tumor-associated antigens (TAA), including viral antigens. These antigens have to go through a process of protein cleavage and binding to MHC molecules in order to be displayed on cells. Then, they have to be recognized by T-cell receptor (TCR) capable of binding the peptide/MHC complex. Next-generation sequencing technologies and algorithms for epitope prediction now allow the identification of mutant tumor neoantigens. Although, in ovarian cancer clinical trials the use of neoantigen-based vaccines has not yet been reported [34].

3.3 BIOINFORMATIC IMMUNE-RELATED CELL TYPE QUANTIFICATION

The approach to the immunotherapy became established due to the launch of new bioinformatic tools able to identify components of immune response. For example, Charoentong et al. [35, 36], using RNA-sequencing data and GSEA, reconstructed the TIL landscape and scored the patients using a prediction score. The score was proposed by Charoentong et al. to increase the understanding of tumor-immune cell interactions on a panel of 20 solid cancers.

This score is a predictor of response to anti-*CTLA-4* and anti-*PD-1* antibodies. They mined the TCGA data to define immunophenotypes, which were defined using the quantification tumor infiltrating cells (TILs) types namely MHC molecules, immunostimulators and immunoinhibitors.

To estimate TILs, they first built a compendium of genes related to specific immune cells using gene expression profiles from 37 studies comprising 366 microarrays. Then, they used gene set enrichment analysis (GSEA) on this gene sets to decompose cellular profiles from RNA sequencing data into immune-related cell type composition.

GSEA is a computational method that determines whether a defined set of genes shows statistically significant differences between two biological states, a.k.a phenotype. This method starts from calculation of an enrichment score (ES) that reflects the degree to which a set is overrepresented at the extremes (top or bottom) of the entire ranked list of genes, and the ranking is based on the correlation of the gene with the phenotype. After that, the significance of ES is tested using an empirical phenotype-based permutation procedure. When an entire database of gene sets is evaluated, ES is normalized for each gene set to account for the size of the set, yielding a normalized enrichment score (NES), then controlling the proportion of false positives by calculating the false discovery rate (FDR) corresponding to each NES.

Using GSEA strategy, Charoentong et al. estimated 28 significantly enriched subpopulations of TILs including major types related to adaptive immunity. Then they employed a machine learning approach (random forest) on the entire list of genes belonging to the 28 subpopulations of TILs to identify those genes predictive of the response to anti-*CTLA-4* and anti-*PD-1* antibodies. For individual cancer types the most predictive features were identified using the mean decrease of accuracy over all cross-validated predictions. The random forest approach identified 163 genes as the most discriminating features of immune-therapy response. These 163 immune-related factors were then classified into four categories: (1) antigen processing (MHC); (2) checkpoints (CP); (3) effector cells (EF); and (4) suppressor cells (SC). To visualize the results, they generated an immunophenogram that includes these four categories along with their specific subpopulation cell types. For each of the four classes an aggregated z-score has been calculated, using the expression of the genes belonging to the categories. Finally, the combination of these z-scores gives the global immunophenoscore (IPS): higher the score (from 0 to 10) more immune-reactive is the patients.

To validate the method, they analyzed two genomic and transcriptomic data sets from patients with melanoma treated with anti-*CTLA-4* and anti-*PD-1* antibodies. They reconstructed the TIL landscape and scored the patients using the IPS being able to stratify patients into responders and non-responders with high accuracy. Similar observation was made also for the patients treated with anti-*PD-1* antibody.

4. NONCODING RNAs

Non-coding RNAs (ncRNAs), unlike messenger RNAs, are not involved in the gene transcription. Instead, their function is to regulate the post transcriptional phase of DNA expression [37]. The ncRNAs can be divided into the following categories: miRNAs, piRNAs, snoRNAs, lncRNAs and other types of ncRNAs, like circular RNAs [38]. Most of the studies have focused on short RNAs, such as miRNAs. This class of non-coding RNAs is characterized by small dimension (max 25nt) and its function as post-transcriptional silencer makes miRNAs object of many studies in a large number of diseases.

In the case of lncRNAs, they are classified as those ncRNAs that are longer than 200 nucleotides on the basis of RNA purification protocols. Their functions are not defined such as miRNAs', but lncRNAs are confirmed as regulators in mechanisms of cell cycle or cancer progression.

One of the most recent class of non-coding RNAs, circular RNAs, has been introduced with the advent of bioinformatics. circRNAs are defined as products of alternative splicing and a fusion of exons, introns or both, in circular structure. Their principal role is miRNA sponge, gaining the ability to regulate translation through miRNAs.

Globally, ncRNAs are already known to have a range of functions, including RNA editing, splicing, the control of chromosome dynamics, mRNA destruction and translational inhibition. The true extent of RNA regulation of these processes is only partially explored. Also, it appears that RNA may play a role in theoretically all levels of gene regulation in eukaryotes [39].

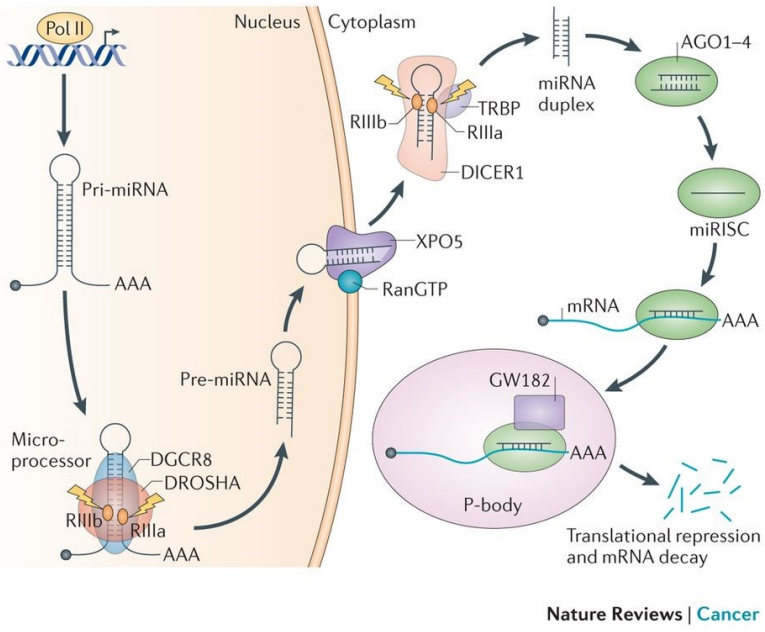
In the following a detailed description of their bio-genesis and functions.

4.1 MICRORNA

The most widely studied class of ncRNAs are miRNAs, which are small ncRNAs of 22-25 nucleotides (nt) that, in animals, mediate post-transcriptional gene silencing by controlling the translation of mRNA into proteins [40].The first report of a regulatory microRNA (miRNA) was made in 1993 in C. elegans, since then, more than 2,500 miRNAs have been annotated in humans. Biogenesis of miRNAs is a multi-step process that involves the RNase III enzymes Drosha and Dicer and results at the end in the production of mature miRNAs. These molecules are loaded by the Dicer–TARBP2 (TAR RNA-binding protein 2) complex into a member of the Argonaute protein family to form the RNA-induced silencing complex (RISC), of which Argonaute proteins are the principal components as catalytic endonuclease

[38]. They are involved in regulating many processes, including proliferation, differentiation, apoptosis and development. miRNAs are estimated to regulate the translation of more than 50% of protein-coding genes. Some miRNAs regulate individual targets, others can operate as regulators of an entire process, so a small number of miRNAs regulate the expression levels of hundreds of genes simultaneously, and also, many other miRNAs regulate their targets cooperatively. Translation of mRNA into proteins is repressed by miRNAs by two main mechanisms: mRNA degradation and the inhibition of translation initiation. They serve as the target-recognition element of an RNA–protein complex known as the RNA-induced silencing complex (RISC), which contains an Argonaute (AGO) family protein that binds the miRNA, along with a range of accessory components [41] (**Figure 5**). The 5′ end of the miRNA, called seed region, forms the target recognition element and it is necessary to group miRNAs into families on the basis of shared seed sequences. When bound to a target mRNA, the RISC complex reduces the rate of translation of the mRNA and accelerates the shortening of the poly(A) tail, resulting in faster mRNA degradation [42].

Figure 5. *Biogenesis of microRNAs*. [41]



miRNAs have important roles regulating gene-expression that underlies normal and pathologic cellular processes, including cancer. Some miRNAs are tumor suppressors, whereas others, when overexpressed, can promote tumor outbreak, growth and progression to metastasis. Probably because of their small size, point mutations are rare in miRNAs; by contrast, their dysregulation is common in many cancers. In addition to regulation, miRNAs can be depleted in tumors compared to normal tissue,

and tumor growth is accelerated in models in which miRNA biogenesis is interrupted [43]. Multiple reports now also suggest that miRNA expression signatures derived from either tumor tissue or liquid biopsies enable more accurate diagnosis and prognosis to be made in patients with cancer and that miRNAs could even represent therapeutic targets in their own right [44]. In order to understand miRNA functions in a network and to map the results of their dysregulation in cancer, methods have been introduced such as genome-wide identification of miRNA–target interactions, RNA sequencing to detect consequences of miRNA overexpression or inhibition, gene and miRNA expression data from both cell lines and patients [40]. In high grade serous ovarian cancer there is a molecular predictor, called MiROvaR, able to stratify patients according to their risk of relapse or progression, identifying groups of patients with significantly different progression-free survival. They found 35 miRNAs having key roles as central nodes in biological processes, associated with a favorable prognosis and a poor prognosis [45].

4.2 LONG NONCODING RNA

Currently there is an inaccurate classification for these transcripts, generally are considered long ncRNAs those transcripts longer than 200 nucleotides, on the basis of a practical cut-off in RNA purification protocols that excludes small RNAs [46]. Indeed, many lncRNAs are like mRNAs: they are transcribed by RNA polymerase II from genomic loci with similar chromatin states to mRNAs. They are often polyadenylated, spliced and 5′-capped and, in most cases, they have the same biochemical characteristic of mRNAs except for the absence of a translated ORF. But, there are also general trends that discriminate lncRNAs from mRNAs: lncRNAs tend i) to have fewer but longer exons, ii) to be shorter, iii) to be expressed at low levels and iv) to be characterized by poorer primary sequence conservation [47].

Anyway, these noncoding transcripts can be partially classified into intronic, antisense or intergenic (‘lincRNA’), even if there is no great evidence of any intrinsic difference between them. Genomic studies based on expressed sequence tag (EST) and full-length cDNA sequencing, microarrays and RNA sequencing (RNA-seq) identified thousands of lncRNAs in diverse genomes species. Future studies will plausibly increase this number, as lncRNAs are more tissue-specific and expressed at lower levels than mRNAs and many cell types have not yet been thoroughly interrogated by RNA-seq [48]. Some lncRNAs seem to play a role in general or differentiation-specific cell biological processes such as regulation of apoptosis and metastatic processes, breast development, and epidermal differentiation, among others. It should also be noted that some functionally and experimentally validated lncRNAs can

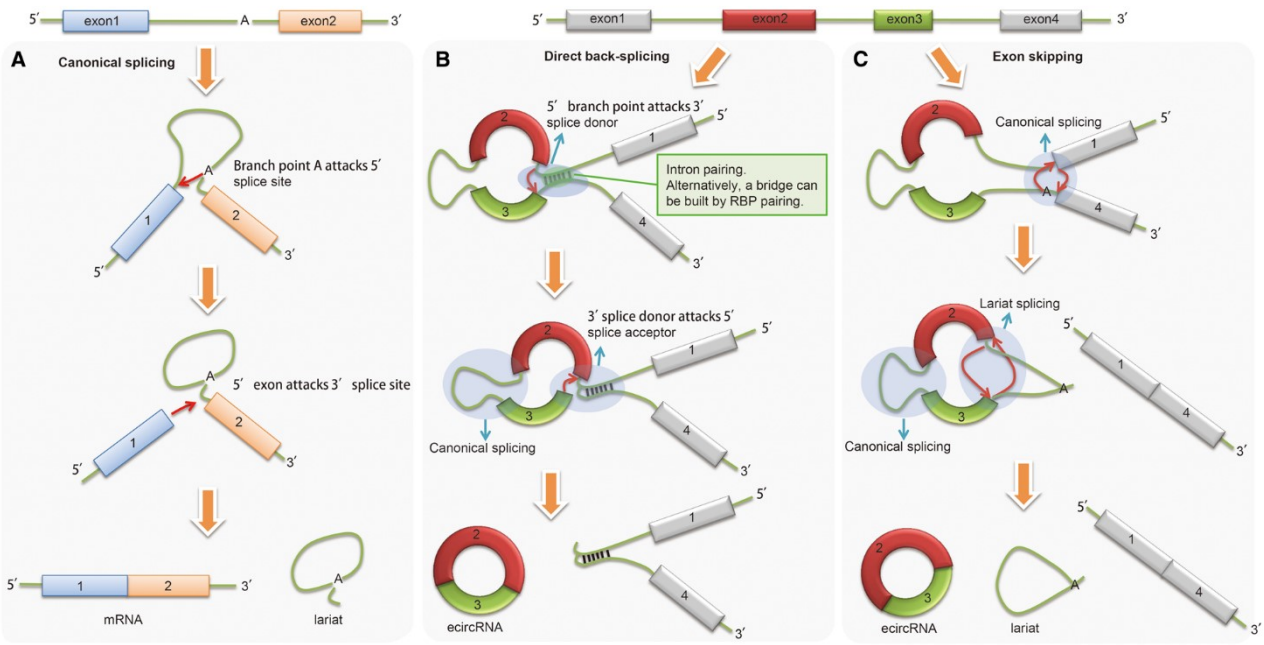
have isoforms that encode proteins [39]. Considering the increasing number of lncRNAs that are involved in changing expression or losing their sequence integrity in many different human diseases, the scientific community focused their studies on lncRNAs' role in cancer.
In high grade serous ovarian cancer, several lncRNAs have been shown to have specific roles in this kind of tumor. These lncRNAs, like H19 or MALAT-1, have been shown to be associated with various biological processes in ovarian cancer, including metastasis, cell growth, cell apoptosis, cell senescence and multidrug resistance [49].

4.3 CIRCULAR RNA

4.3.1 CIRCULAR RNA INTRODUCTION

Thousands of endogenous circular RNAs (circRNAs) have recently been identified and characterized in eukaryotic cells. The introduction of next-generation sequencing technologies with bioinformatics approaches has allowed a comprehensive exploration of circRNAs. Global studies revealed that circRNAs can be generated from intergenic, intronic, coding region and even the tRNA introns could form stable circular RNAs (tricRNA) [50]. Most circRNAs arise from the back splicing event, in which generally a downstream 5′ splice donor joins an upstream 3′ splice acceptor and have no U2 or U12 spliceosome preference (**Figure 6**) [51]. It has been noticed that the circularized regions are flanked by large introns.

Figure 6. *A) Canonical splicing. B) Direct back-splicing. Intron pairing or RBP pairing bridges two flanking introns close together, and then the released two exon ends are connected after the removal of the introns. C) Exon skipping* [51].



The repetitive sequences in the flanking intron facilitate the creation of a loop structure that dissolves the problems of opposite direction and discontinuity between the 5′ and 3′ splice sites, where Alu sequences achieve an efficient pairing [52]. Indeed, circularization improves circRNAs stability, making them resistant to RNase R that degrades linear RNAs. Genome-wide analyses indicated that many circRNAs are conserved across species and they have cell-type or tissue-specific expression, suggesting potential regulatory roles. However, mature circRNAs appear to be noncoding transcript because of start and/or stop codons absence, but engineered circRNAs have been shown to be translatable, although in cytoplasm they are not associated with ribosomes [53].

About cellular localization, most circRNAs tend to be cytoplasmic, but the transport mechanism is currently not clear. At the moment, there are two hypotheses: during the generation of circRNA the spliceosomal machinery participates to the transport by nuclear export system or circRNAs may be delivered during mitosis [51]. Also, it must be considered that many circRNAs are predominantly localized in the nucleus, suggesting that they may regulate gene transcription. Indeed, they become a research hotspot because of their discovered functions in diverse cellular processes. Above all, the majority of circRNAs could serve as miRNA sponges by sequestering and preventing miRNAs from

binding genes target [51]. In addition to miRNA regulation, circRNAs may regulate the intracellular transport of RNAs binding proteins (RBPs) and participate in transcriptional regulations, expanding the complexity of the transcriptomic regulation. Also, circRNAs are involved in disease-related pathways, becoming a promising biomarker for disease diagnosis [54].

Recent studies in ovarian cancer highlighted their potential to alter expression of oncogenic or tumor suppressor genes or miRNAs, leading to cancer development, progression and metastases. The results can lead to networks of circRNAs, miRNAs and mRNAs in which a single circRNA could regulate many genes through a common microRNA target. This ability to communicate via miRNA prompts for new opportunities in research to uncover the role of circRNAs in carcinogenesis and to identify new biomarkers for ovarian cancer diagnosis and prognosis [55].

4.3.2 BIOINFORMATIC IDENTIFICATION OF CIRCULAR RNA

In recent studies, several pipelines have been proposed to specifically identify circRNAs based on deep sequencing datasets. Based on the results of [56], a benchmark study on the circRNA identification tools, five software (circRNA_finder, find_circ, CIRCexplorer, CIRI, and MapSplice) can be considered the most reliable [56]. They found dramatic differences among the algorithms specifically regarding the highly expressed circRNAs and the circRNAs derived from splice sites.

Basically, the number of the predicted circRNAs ranges from 1532 to 4067, out of which only 854 are predicted by all 5 algorithms. The false positive rate measured by the RNase R resistance of circRNA candidates ranges from 12% to 28%.

Therefore, low map-quality reads and backsplice-junction sequences with a homology to linear exon junctions should in general be discarded. Also, a notable difference has been observed between the algorithms concerning splice site distance requirements. Here, CIRI and circRNA_finder have the capacity to identify circRNAs with very proximal splice sites, however most circRNAs below 200 bp are here classified as mis-annotated and probably false positives.

Finally, considering the computational cost, in terms of hardware resources and time needed, CIRCexplorer and MapSplice needed 2–3 days to complete individual dataset predictions, both require gene annotation lists, while CIRCexplorer requires indexed genomes from both Bowtie1 and Bowtie2, and relies on tophat, bedtools, and samtools making this tool one of the more complex pipelines to implement.

At the end of the study the Authors suggested for a fast and almost equally reliable output the intersection of circRNA_finder and find_circ tool. Moreover, find_circ, circRNA_finder, and CIRI all work de novo without knowledge of gene annotations and exon-intron structures. Thus, these algorithms are insightful when conducting unbiased circRNA analyses or in poorly annotated organisms. In conclusion, this study emphasizes that circRNA annotation should be handled with care and that several algorithms should ideally be combined to achieve reliable predictions.

5. PROJECT RATIONALE AND AIMS

The project of my thesis grounds on the extreme heterogeneity of high-grade serous ovarian carcinoma (HGSOC). A significant fraction of HGSOC patients at advanced stage is initially sensitive to Pt-based, but became resistant after relapse beyond the end of Pt-based therapy. For these patients efficacious alternative therapies do not exist, and death occurs rapidly. Prognostic clinical factors, as age, FIGO stage, tumor grade, and residual tumor (RT), are insufficient to capture individual variations predicting response to chemotherapy. In addition, the clinical response to first line chemotherapy is heterogeneous and not predictable at diagnosis. To our knowledge clinical and pathological differences between Pt-r and Pt-s HGSOC patients have not been completely elucidated. In this project, unknown genomic and molecular characteristics have been considered in order to investigate the mechanisms involved in the lack of response. Taking advantage of bioinformatic approach, we integrated a large amount of coding and non-coding data in order to shed new light on the complex mechanisms driving innate platinum resistance in HGSOC, finding putative novel therapeutic targets. We evaluated a wide cohort of patients, with comparable clinic annotations, to explore expression alterations in coding and non-coding regions. Indeed, we considered genes and miRNAs interaction, involving also alternative transcript variant as circRNAs, with the aim to elucidate platinum resistance mechanisms. Moreover, in harmony with some elements of network, we considered to investigate the immunogenicity of HGSOC, due to the presence of some key elements related to immune response. From a clinical point of view this would represent a step forward in the field of precision medicine since at date there is no clinical parameter for HGS-EOC able to predict resistance to chemotherapy. In this perspective, this study tried to resolve the urgent need to identify early predictive biomarkers that might be used to optimize therapeutic strategies for resistant patients.

6. MATERIALS, METHODS AND TECHNOLOGIES

6.1 TISSUE SAMPLE COLLECTION

A retrospective cohort of 1080 EOC patients was gathered together from three independent tumor tissue collections, named cohort A, B and C. All patients were staged according to the International Federation of Gynecologic and Obstetrics (FIGO) criteria and were classified as HGSOC stage III/IV [6]. Based on the time lagging between the end of first line platinum-based therapy and relapse, patients from cohort A and B were classified into three groups: i) Pt-s platinum-sensitive (patients relapsing after more than 12 months from first line chemotherapy), ii) Pt-ps partially-sensitive (patients experiencing relapse within 6-12 months from the last round of chemotherapy), iii) Pt-r platinum-resistant (patients experiencing relapse within 6 months from the end of therapy) [57]. In cohort C the time between the end of first line and relapse is not available, thus, the stratification among Pt-r, Pt-s and Pt-ps is not possible.

Cohort A is composed by 99 snap-frozen HGSOC biopsies obtained at the Division of Obstetrics & Gynecology, ASST Spedali Civili, University of Brescia, between 2003 and 2013. Biopsies are stored in the biobank located at the "A. Nocivelli" Institute, ASST Spedali Civili of Brescia. Samples of cohort A were randomly selected in order to have sample size i) balanced between Pt-r and Pt-s and ii) sufficient to reach a good statistical power for the identification of a robust network-based signature [58].

Cohort B is composed by 143 snap-frozen tumor biopsies derived from HGSOC patients undergoing debulking surgery at the Obstetrics and Gynecology Dept., San Gerardo Hospital (Monza, Italy). Biopsies are stored in the Pandora tumor tissue collection located at Mario Negri Institute for Pharmacological Research (Milan).

For both cohorts A and B clinical and anatomopathological information were registered and follow-up data were obtained from periodical gynecologic and oncological check-ups. Tumor samples were collected during cytoreductive surgery and frozen within 15 minutes in liquid nitrogen and stored long term at -80°C. The tumor content of the biopsies collected was evaluated with hematoxylin and eosin staining and only samples containing more than 70% of epithelial tumor cells were used for downstream analyses.

Cohort C is composed by 838 samples obtained through the curated ovarian cancer dataset publicly available through the Bioconductor platform gathering together the TGCA data plus other ovarian

datasets [59]. The curated ovarian database as reported in the "curatedOvarianData" Bioconductor package [59] was used as a first external independent validation set (Cohort C). The curatedOvarianData database contain several normalized and batched-corrected ovarian datasets. Among these, we selected studies with platinum treated patients and complete follow-up for either progression free (PFS) and overall survival (OS). Unfortunately, the progression free interval (PFI) defined as the time between the relapse and the end of platinum-based therapy is not available. For this reason, samples of the curatedOvarianData cannot be stratified according to Pt-s and Pt-r. Here we will use PFS and OS as a proxy of platinum resistance. The selection leads to 5 datasets (GSE30161, GSE9891, GSE49997, TCGA microarray and also TCGA.RNASeqV2) with a total of 838 samples.

6.2 GENE AND MIRNA EXPRESSION EXPERIMENTS AND DATA ANALYSIS

Snap-frozen tumor biopsies were homogenized in a TissueLyser LT (Qiagen) and RNA was extracted and purified using RNeasy mini kit (Qiagen) with a semi-automated extraction system (QIAcube, Qiagen). The obtained RNA enriched in miRNA fraction was then quantified using a Qubit (Thermo Fisher Scientific) and its quality was evaluated on a commercial automated electrophoresis system (Tape station 4200, Agilent Technologies). Microarray mRNA and miRNA expression experiments were performed on 99 HGSOC specimens (cohort A1+A2) using commercially available kits (G4851B human whole GE kit and G4470B Human miRNA kit, Agilent technologies). According to manufacturer's instructions, RNA was labeled and hybridized as previously published [60] and the fluorescence intensity generated was measured with Agilent Feature Extraction software, version 11 (Agilent Technologies). In accordance to the MIAME guidelines, the array data files have been uploaded to the Array Express database (IDs E-MTAB-7083 and E-MTAB-7084).

6.3 RNA-SEQ EXPERIMENTS AND DATA ANALYSIS

RNA sequencing has been performed by Personal Genomics, in University of Verona, using Illumina HiSeq1000 platform and as sequencing standard TruSeq Stranded Total RNA LTwithRibo-Zero TM Gold, generating 2x100 bp paired end lane. Next, reads have been mapped with STAR using the two-pass alignment strategy with genome sequence (hg38.p2 version) and annotations (Ensembl 79). The alignments were then passed to StringTie (PMID:25690850) for transcript assemblies. The sample-specific transcriptome assemblies were used to create a consensus transcriptome (cuffmerge) and

compare the reconstructed transcripts with the existing annotations (cuffcompare) (PMID:20436464). We used RSEM to estimate transcript abundances (RSEM version 1.2.21) (PMID:21816040).

6.3.1 DNA AND RNA EXTRACTION

Following the manufacturer's instruction, from respectively 25 mg and 30 mg of snap-frozen tumor biopsies, DNA and RNA were extracted and purified (QIAamp DNA Mini kit and RNeasy Mini kit, Qiagen). The integrity of DNA and total RNA was checked using 4200 TapeStation System (Agilent Technologies), and the amount of nucleic acids extracted quantified using fluorometric assay (Qubit, Life Technologies).

6.3.2 DNA SEQUENCING

For each patient, 200 ng of gDNA in a reaction volume of 100 ul of Tris-HCl ph 8.5, was mechanically fragmented using 30'' On/ 90'' Off for 7 cycles as shearing conditions (Bioruptor, Diagenode). After purification with 1,8X AMPure XP Beads (Beckman Coulter), the fragmented DNA was eluted in 50 ul of purified water, quantified using Qubit HS dsDNA High sensitivity assay kit (Life Technologies) and the length of fragments evaluated by 4200 TapeStation System (Agilent Technologies).
The libraries for the sequencing of 6000 disease-related genes and backbone (SureSelect XT, Agilent Technologies) were automatically prepared using Bravo Automated Liquid Handling Platform (Agilent Technologies) following the manufacturer's instructions.
After barcoding procedure 14 samples for run, were loaded on NextSeq 500 Sequencing System (Illumina).

6.3.3 MUTATION ANALYSIS

Raw reads from the instrument were aligned to the reference genome (hg19) with BWA. Germline variant calling was performed with GATK default setting.

6.4 MICROARRAY ANALYSIS

6.4.1 DATA PROCESSING

Raw mRNA, lncRNA and miRNA expression signal was quantified using Agilent FeatureExtraction software. gProcessedSignal and gTotalGeneSignal were used as expression measures respectively for mRNAs and miRNAs. A first step of filtering was set up using the Agilent flag gIsPosAndSignif for mRNAs and lncRNAs and gIsGeneDetected for miRNAs. Probes with more than 60% of low-quality values were removed. A second step of filtering was performed on samples using Relative Log Expression (RLE) plot [61]. Samples with RLE distribution markedly different from the general trend were removed from the following analyses. Probes mapping on the same gene annotation were averaged. Finally, data have been log transformed and quantile normalized. Coefficient of variation (CV) was used to select the most variable genes and miRNAs.

6.4.2 RUV NORMALIZATION

In data with large scale gene expression studies, the observations are commonly contaminated by sources of unwanted variation such as platforms or batch effects. These unwanted variations during analysis can lead to weak associations and to missing important signals.

Using RUV normalization [62] negative control genes and replicate samples can be used to estimate unwanted variation in gene expression based on RNA-Seq data. The proposed method has been translated on microarray data and implemented in the bioconductor package RUVnormalize [63]. They generally manage to remove unwanted variation without losing the signal of interest and compare favorably to state-of-the-art corrections. RUVnormalize, using positive and negative control probes (n=48), was applied (with number of normalizing factor k=2) to remove unwanted variation and batch effect. Hereafter, differentially expressed (DEs) coding and non-coding elements between Pt-r and Pt-s were identified using permutational moderated t-test, as implemented in samr R package (permutational number set to 1000).

6.4.3 NETWORK ANALYSIS: MICROGRAPHITE

In order to integrate miRNAs in pathway analysis, micrographite [58, 64] gave us a lot of advantages: (i) the possibility to integrate and (ii) analyze miRNA and mRNA expression profiles using pathway information and (iii) to biologically contextualize miRNA–mRNA validated and predicted interactions. In our study, micrographite was used to identify integrated circuits of mRNAs and miRNAs associated to the therapy response (Pt-r vs Pt-s).

Pathway topologies derived from *graphite [65]*, a Bioconductor package developed to store, manage and convert pathway annotations into gene–gene networks. *graphite* is a pathway data interpreter that, following biologically driven rules, is able to solve the complexity of the pathway modules to generate interaction networks suitable for topological pathway analyses. KEGG pathways as available through *graphite* [65] and miRNA-target gene interactions available on "TargetScanHuman" database release 7.1 [66] were considered for the analysis.

Only miRNA-target interactions characterized by a Pearson correlation coefficient $|r| \geq 0.4$ and q-value ≤ 0.05 were included into pathway annotation. Then, the topological pathway analyses used in micrographite are a modified version of CliPPER [58, 64].

CliPPER is a Bioconductor package that implements a topological pathway analysis based on Gaussian graphical model theory and then it is able to deal with data deriving from different sources with possibly different measurement scales. This approach is based on two steps. In the first step, pathway graphs were compared in terms of means and variance between groups. On these selected pathways, it identifies portions of the pathway mostly associated with the phenotype. Into details, path identification is based on the graph decomposition into small-connected components, called cliques. Each clique is tested independently (according to the test on the means and/or concentration matrices) and then a significant level (P-value) for each clique is obtained, creating a path as a list of adjacent significant cliques.

Clipper is able to identify and score all of these paths in the graph. Paths obtained have a score that is a function of all the P-values of the cliques contributing to the path and to the higher score correspond the better path. Then, the upper-scored 10th percentile of the portion of significant pathways (in mean and variance with $p<0.1$ and 10 000 permutation) are combined into a non-redundant meta-pathway. Finally, the portion of the meta-pathway mostly associated to the phenotype is revealed.

6.4.4 NETWORK ANALYSIS: META-PATHWAY EXTENSION

A significant portion of genes is not annotated into KEGG pathways, and then is not taken into account by micrographite pipeline. In absence of this information, we considered the construction of networks having all the protein-protein interactions using computational methods for signaling pathways and protein complex identification in specific diseases. Studies have also shown that proteins with larger number of interactions can include families of enzymes, transcription factors, and intrinsically disordered proteins, among others.

Thus, to extend our network taking into consideration the excluded genes, we used STRING [67] and BioGRID [68] databases: differentially expressed genes without KEGG annotation were added to the network if at least one of their interactors (as reported by STRING and BioGRID) were present in the network. Cytoscape [69] was used to visualize the integrated regulatory network.

6.4.5 VALIDATION USING qRT-PCR

Gene expression levels were validated on a total of 242 HGSOC samples (Cohorts A and B) by quantitative real-time PCR (qRT-PCR) using QuantiFast SYBR Green (Qiagen), as previously published [60, 70]. Reactions were prepared following the manufacturer's instructions in 384-well reaction plates with an automatic liquid handling system (EpiMotion 5075LH, Eppendorf) and experiments were run on an Applied Biosystems 7900 instrument. The primer sequences used for the qRT-PCR validation are reported in the table below. Analysis was conducted as beforehand described, using three independent reference genes, *GAPDH, HPRT1* and *PPIA* which were selected among those most stable based on our previous publication [71].

6.4.6 HOMOLOGOUS RECOMBINATION (HR) SIGNATURE IN TCGA SAMPLES

In order to complete clinical information of cohort C, in particular for TCGA dataset, ovarian cancer data were extracted from FireBrowse database (http://www.firebrowse.org), after matching samples ID of cohort C and samples ID of clinical data. Germinal and somatic mutation data were obtained from Mutation_Package_Calls and Mutation_Package_Oncotated_Calls files respectively. Methylation levels (beta values) were obtained from humanmethylation27-within_bioassay_data_set_function.txt file and transformed into M-value [72]. M values greater or equal to 2 have been used as threshold to define methylated genes. Copy number variations were retrieved from all_thresholded.by_genes.txt file.

Defect of the Homologous Recombination (HR) pathway was defined by the presence of least one of the following alterations: amplification of *EMSY*, mutation or methylations of *BRCA1/2*, mutations of *PTEN*, DNA repair genes (*RAD*) and *RAD3*-related (*ATR*), Fanconi Anemia genes (FA) and serine/threonine-protein kinase (*CHEK1/2*) genes. In case of mutational events, both somatic and germinal events have been taken into account [73, 74].

6.5 IMMUNOPHENOTYPE CHARACTERIZATION

In order to identify determinants of tumor immunogenicity in ovarian cancer we followed two different approaches: the first proposed by Charoentong et al. based on a scoring scheme for the quantification of immune-related cell types called immunophenoscore [35] (IPS) and a second based on neoantigens detection strategy for the identification of immunotherapy targets [75].

6.5.1 IMMUNOPHENOSCORE

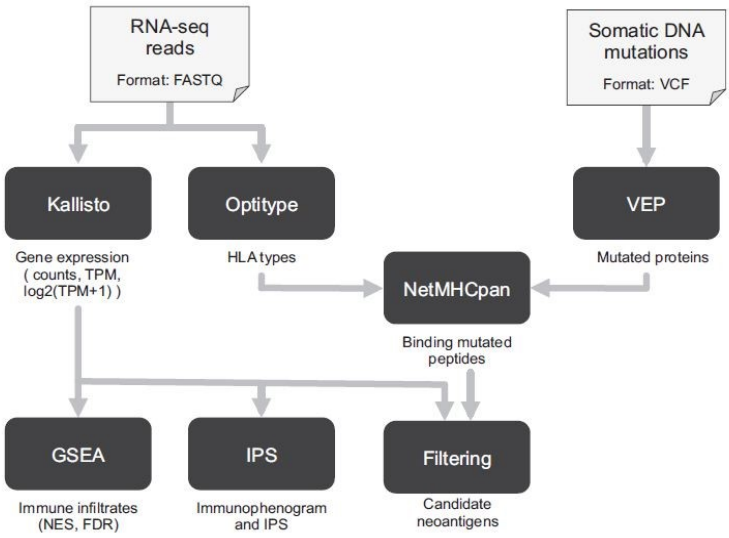
Immunophenoscore [35] (IPS) was applied to our ovarian cancer data in order to characterize the immunological background, to discriminate resistant versus sensitive patients, and to predict response to immunotherapy. The analysis was performed with R script of IPS using as inputs gene expression matrix obtained from previous analysis (4.4) and a list of 163 genes representing the four classes of immune cells, identified during the IPS development. Each of these selected genes has characterized by a weight value based on their role in the immune response and the class affinity. As results, for each sample we have a matrix with: i) z-score class specific, ii) aggregated z-score and iii) IPS score, in a range included between 0 and 10, respectively bad and good response to immunotherapy. In addition, a graphical representation of the results, called immunophenogram, has been generated in pdf file format.

6.5.2 TIMINER

TIminer [75] is an analytical pipeline developed to perform integrative immunogenomic analyses using NGS data. This pipeline integrates different bioinformatics tools to perform immunogenomic analyses starting from RNA-seq reads and somatic DNA mutations. RNA-seq reads must be provided as FASTQ files and somatic DNA mutations should be as Variant Call Format (VCF) files. In first place, RNA-seq FASTQ files are used to quantify gene expression with Kallisto [76] as: gene-specific counts, transcripts per millions (TPM), and normalized log2(TPM+1). Moreover, a cut-off value of

log2(TPM+1)>2 has been used, to filter only candidate neoantigens in consideration of gene expression results.

Figure 7. *The scheme illustrates the different computational tools integrated in TIminer, the input/output data, and the data flow between the tools* [75].



At the same time, Optitype [77] analyzes FASTQ files to predict class-I *HLA* types, generating the output reports with predictions of the *HLA-A, HLA-B* and *HLA-C* alleles for each sample.

Somatic DNA mutations are annotated by VEP [78] considering only the coding regions in order to allow the predictions of affected proteins and their sequences. In the next step, the mutated proteins arising from missense mutations, together with the predicted class-I *HLA* types, obtained with Optitype, are subjected to NetMHCpan [79] to predict mutated peptides binding specifically to *HLA* molecules (**Figure 7**). Binding peptides are filtered considering gene expression results to select only candidate neoantigens associated with expressed genes, avoiding as much as possible false positives.

After that, only the candidate neoantigens present in at least 50% of samples have been selected from TIminer output. This selection of final neoantigens allowed us to intersect resistant and sensitive patients and to analyze differences within the two classes.

6.6 IDENTIFICATION OF CIRCULAR RNAs

6.6.1 CIRI2

CIRI2 [80, 81], is a powerful bioinformatic tool optimized for circRNAs detection and is a recent upgrade of the old version CIRI [80, 81]. It requires as input FASTA formatted reference sequences and the SAM alignment generated by BWA-MEM, for de novo detection of circRNAs based on backsplice junctions (BSJ). Anyway, an optional GTF input can also be used by CIRI2 for an extra annotation. In detail, CIRI2 first divides the segment into n seeds with specific length and attempts to find the location of each seed in a given genomic region with the same length. Then, the key segment should be from either the putative downstream region (Region 1) with back-splice genomic region, or the putative upstream region (Region 2) consistent with forward splice. For the two candidate regions, multiple seed-matching steps are processed individually, with a comparison of the matched seeds in the next step. Indeed, the two possible results of such comparison are i) Region 1 > Region 2 and ii) Region 1 < = Region 2. For i) CIRI2 determines the corresponding read to be a BSJ read. For ii), the key segment will have equal or larger likelihood to be from Region 2 compared with Region 1, and CIRI2 thereby determines the corresponding read to be a FSJ read. Also, CIRI2 on this comparison estimates FDR. The above algorithm is also used in paired-end mapping filtration for all candidate BSJ reads.

6.6.2 FIND_CIRC

One of the most used computational tools for predicting circRNAs is find_circ [82], starting from RNA-seq data. In a first step, sequenced reads that aligned contiguously and full length to the genome were discarded. From the remaining reads were extracted sequences 20 nucleotides long from both ends and aligned them individually to the reference, to find unique anchor positions within spliced exons. Anchors that aligned in the reversed orientation indicated possible circRNA splicing. After that, the anchor alignments were extended in order to complete the sequence and to flank the read aligns, also the breakpoints, by GU/AG splice sites. Obviously, ambiguous breakpoints were discarded. The tool used Bowtie as short-read mapper.

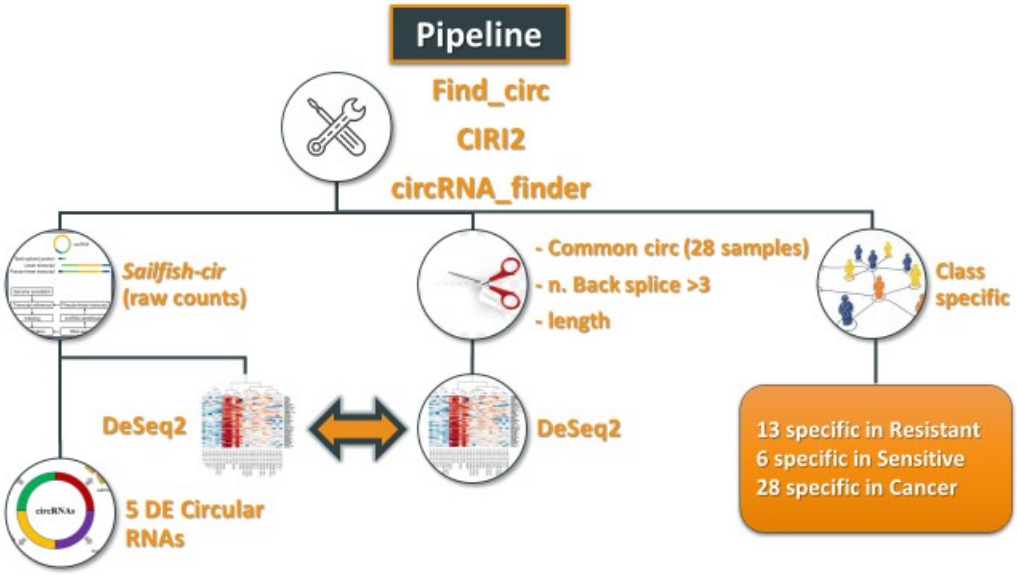
6.6.3 CIRCRNA_FINDER

CircRNA_finder [83] is a computational pipeline that uses the STAR read aligner [84]. Reads were aligned using the different parameters to identify chimeric transcripts such as minimum transcript length or number of mismatches. Thus, at most 3 mismatches were tolerated for each read pair and were used only unique mappers. The candidate chimeric junction reads were then filtered to include cases where one read in a pair spanned a junction with the splice acceptor on the same chromosome and strand as the splice donor. Then, the resulting set of junction-spanning reads were collapsed into a set of possible circularization junctions. In the next analysis only circular junctions matching GT-AG splice sites and supported by at least 10 reads were considered.

6.6.4 SELECTION OF PUTATIVE PROGNOSTIC CIRCRNAS

We kept circRNAs identified by all the above-mentioned methods and we investigated their expression differences between sensible and resistant patients. Expression differences were evaluated using three approaches: 1) differential expression based on the reads mapped on the backsplice junction, 2) differential expression using reads count quantified considering both the circular and the linear transcript, 3) identification of class-specific circRNA (present at detectable level in one class and absent in the other).

Figure 8. *Pipeline overview in circRNAs. We choose an output combination of find_circ, circRNA_finder and CIRI2.*



In the first approach, we selected only those circRNAs expressed in all patients independently from the pathologic classes, with at least 3 reads on the back-splicing events and being shorter than 10kb (**Figure 8**).

The second approach, consolidated our results, quantifying the expression of both circRNAs and its own equivalent linear transcript since they both compete for the reads alignment. To this aim, we used sailfish-cir [85], a computational tool to estimate the relative abundance of circular RNA transcripts from high-throughput RNA-seq data. It accepts output of circRNA identification tools or a BED-format, then it transforms all circular transcripts to pseudo-linear transcripts. Finally, it estimates the expression of both linear and circular transcripts using Sailfish framework. After this filter, we performed a differential analysis on selected circRNAs using a Bioconductor package, called DESeq2 [86], able to estimate variance-mean dependence in count data from RNA-Seq data and test for differential expression based on negative binomial distribution.

At last, to obtain class-specific circRNAs, we started from the output of the first approach and we selected those circRNAs present in all samples of one class but absent in the other classes. For both approaches, differential analysis and class specific selection, the results obtained by the three tools used were intersected. After converting circRNAs coordinates from hg38 to hg19, circRNAs have been annotated using circBase [87].

6.6.5 CIRCULAR RNA VALIDATION IN qRT-PCR

CircRNAs expression levels were validated by quantitative real-time PCR (qRT-PCR) using CFX384 Real-Time and iTaq Universal SYBR Green Supermix (BIO-RAD Laboratories, Hercules, CA, USA). Primers used for circRNA_BARD1 amplification have been prepared starting from the spliced sequence obtained by linear transcript variant 1 of gene BARD1, as reported in circBase (ID: hsa_circ_0001098). The primer sequences used for the qRT-PCR validation are reported here.

Primers circRNA_BARD1: Forward GTGAACACCACCGGGTATCA; Reverse CCAGAAATGCTGTATTTGAAAGAAG.

To estimate the expression levels of the full-length transcript variant 1 of BARD1 (FL BARD1), following primers have been used:

Primer FL BARD1: Forward GACAACTGGACAGCATGATTCAA; Primer Reverse TTGTTTCCTGCATCATTAAACAAAC.

Results have been normalized using geometric mean of two independent reference genes, *HPRT1* and *PPIA*, which were selected among those most stable based on our previous publication [71].

6.7 STATISTICAL ANALYSIS

Kaplan-Meyer curves were used to visualized patients' survival. Univariate analysis was performed with the log rank test, multivariate analysis was performed using Cox proportional hazard model as implemented in survival R package. Overall survival (OS) was calculated considering the time lagging between the diagnosis and the death for any cause or the last follow-up. Instead, progression free survival (PFS) is defined as the time from diagnosis to disease progression or last follow-up. The optimal cutoffs of gene expression values were estimated using the maximally selected rank statistics as implemented in the survMisc R package. Residual tumor and age were used as covariates in the multivariate analysis. Samples with missing survival data were excluded from the analysis. Results were reported as p-value, hazard ratio (HR) and 95% confidence intervals (CI). T-test was used to test the mean difference between Pt-r and Pt-s samples on log transformed qRT-PCR expression values. Unsupervised cluster analyses were performed using pheatmap BioC package with Euclidean distance and average linkage. All the analyses were performed using R platform.

7. RESULTS

7.1 REGULATORY NETWORK ASSOCIATED TO THERAPY RESPONSE

In the current study, using an innovative pathway-based analysis, we identified a complex biological network, composed of coding and non-coding elements, that represents the putative mechanisms of platinum resistance. The network includes five macro areas based of pathways elements function and shows how they interact each other.
This network was validated with an experimental technique in two independent cohorts of HGSOC and three genes were confirmed as independent prognostic factors of survival.

7.2 COHORT DESCRIPTION

Clinico-pathological features of HGSOC patients belonging to cohort A, B and C are shown in **Table 2**. All biopsies enrolled into this study were collected at primary surgery, before chemotherapy, from patients with diagnosis of HGSOC, stage III-IV, according to FIGO guidelines (Federation International of Gynecology and Obstetrics) [6]. The median age at diagnosis was 63 (cohort A), 57 (cohort B), and 59 (cohort C) and the median follow-up time was 3 years for cohorts A and B and 2 years for cohort C. The clinical parameters reported for cohorts A and B are consistent with those suggested in the existing literature as characterizing HGSOC [88]. Most patients presented with a suboptimal residual tumor (RT) after surgery (80% cohort A, 82% cohort B and 34% cohort C). Within cohort A, 36% of patients were defined as Pt-s, 42% as Pt-r and 19% as partially sensitive (Pt-ps), while within cohort B, 48% were defined as Pt-s, 34% as Pt-r and 17% as Pt-ps. Unfortunately, information on progression free interval (PFI) is not available for cohort C, thus samples have been classified into Pt-s and Pt-r, exclusively for differential tests, using PFS.

Table 2. *Clinical and demographic description of cohort A, B and C.*

	Patient characteristics		
	Cohort A	**Cohort B**	**Cohort C**
Clinical Annotations	**No. of Patients, N(%)**	**No. of Patients, N(%)**	**No. of Patients, N(%)**
Total No. of patients	99	143	838
Median (range) years	63 (36-85)	55(33-83)	59 (22-89)
Histological type			
Serous	99 (100%)	143 (100%)	838 (100%)
FIGO Classification			
III X			5(0.5%)
III A	2 (2%)	0 (0%)	12 (1.5%)
III B	1 (1%)	5 (3%)	45 (5%)
III C	74 (75%)	118 (83%)	670 (79%)
IV	22 (22%)	20 (14%)	106 (14%)
Platinum Status$			
Sensitive	36 (36%)	69 (48 %)	-
Partially Sensitive	19 (19%)	25 (17 %)	-
Resistant	41 (42%)	48 (34 %)	-
NA	3 (3%)	1 (1 %)	-
Median (range) years	63 (36-85)	57 (33-83)	59 (22-89)
Median follow up (range) years	3 (0-13)	3 (0-19)	2 (0-17)
Median PFS (range) months*	14.4 (1.2-142)*	31 (0.8-239)	14.3 (0.3-166)
Median OS (range) months	36.2 (1.15-161)	53 (2-239)	28.1 (0.3-214)
Residual Tumor			
0	20 (20%)	26 (18%)	-
0<RT<1	20 (20%)	31 (22%)	-
RT>1	59 (60%)	86 (60%)	-
Optimal	-	-	557 (66%)
Suboptimal	-	-	281 (34%)

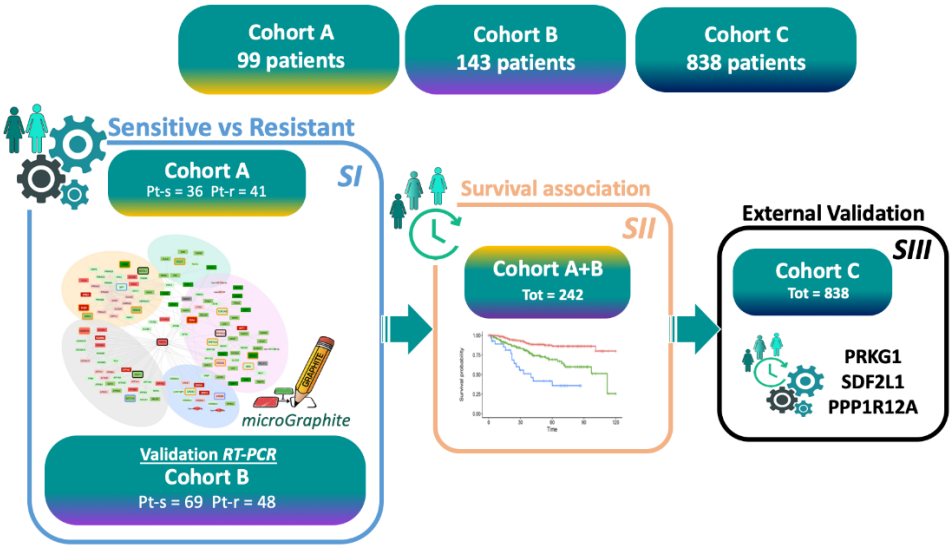
* *Calculated on 95 patients (4 patients do not have PFS follow-up)*

$ *In Cohort C platinum free interval is not available*

7.3 PATHWAY ANALYSIS AND NETWORK GENERATION

To generate novel hypotheses on the underlying molecular mechanisms driving resistance to first line Pt-based chemotherapy in HGSOC, a compendium of coding and non-coding genes (i.e., miRNA) was generated. Pt-s (n= 36) and Pt-r (n=41) cases were selected from cohort A and mRNA and miRNA profiles analyzed using micrographite algorithm (**Figure 10**) [64]. The aim was to move from a list of single genes towards a comprehensive map of functionally related networks. Micrographite analysis resulted in a network composed of 131 mRNAs and five miRNAs (**Figure 11**, Supplementary **Table 1-2**). The network wired five major functional processes: transcription regulation, transmembrane ion transport (such as calcium ion transport and sodium ion export) cell cycle regulation and response to damage, fatty acid metabolism and antigen presentation. The above five functional processes are interconnected by the *PRKG1* gene, which acts as a hub of the network allowing crosstalk among the five different functional areas (**Figure 11**).

Figure 10. *Schematic flow of the study. We applied a pathway-based approach to identify a significantly modulated network between Pt-s and Pt-r samples in cohort A. A selection of the element of the network have been validated in cohort B as differentially expressed (signature SI) and in cohort A+B as associated to prognosis (signature SII). Then, using a meta-analysis on a large and external dataset (cohort C) we confirm the prognostic relevance of the element obtained in the first round of validation (signature SIII). We end up with 3 genes associated to the platinum response with prognostic characteristic (either in OS and PFS).*



7.4 NETWORK VALIDATION IN COHORT B

To demonstrate the reproducibility and robustness of the micrographite analysis, we assessed the expression of 23 out of 136 elements of the circuit by an independent technique, qRT-PCR on an independent cohort (cohort B Pt-s= 69 and Pt-r= 48) (**Figure 10**). The selection was made accordingly to one or more of the following criteria: i) significance level lower than 0.05; ii) log fold change greater than |0.5|; iii) literature evidence.

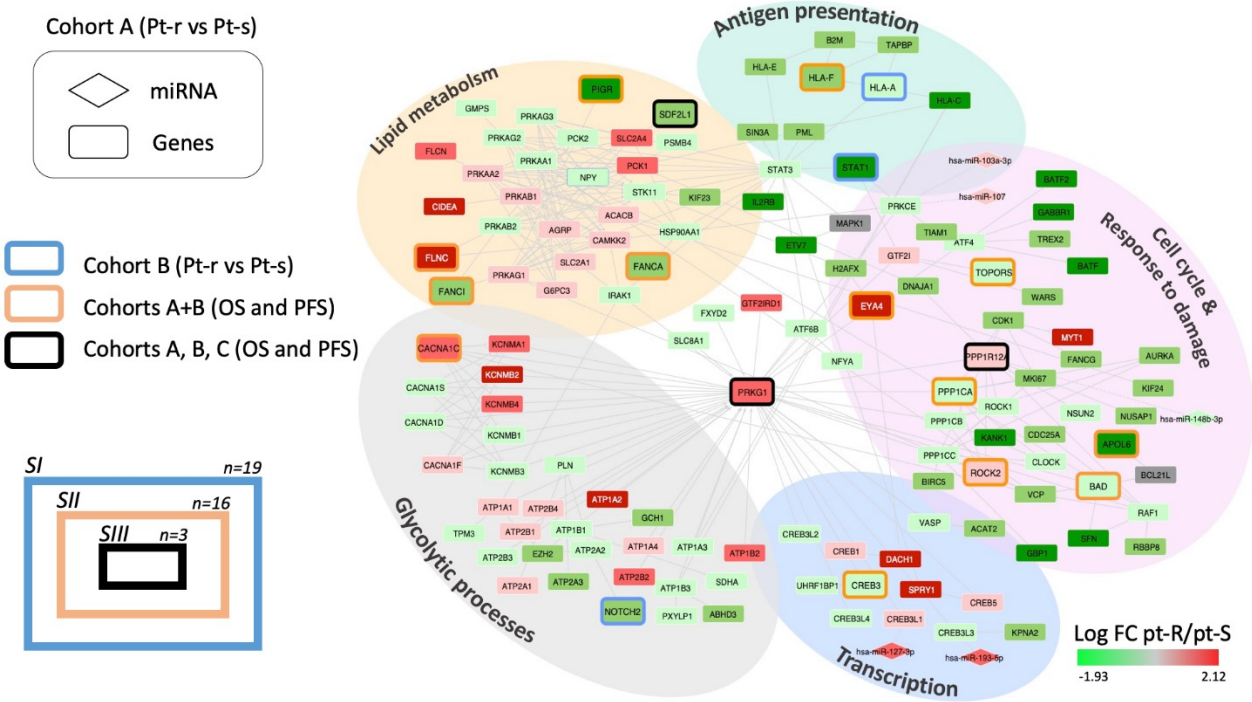
Table 3. *Differential expression evaluated by qRT-PCR between Pt-r and Pt-s samples of the 23 selected genes in cohort A+B. Average expression of Pt-s and Pt-r samples and t-test p-value is reported.*

	Pt-r mean	Pt-s mean	t-test p-value
FANCA	2.25	3.58	0.00
PIGR	1.51	3.23	0.00
BAD1	4.81	5.92	0.00
HLA-A	5.19	6.60	0.00
STAT1	5.96	5.51	0.05
HLA-F	3.48	5.18	0.00
FANCI	3.63	5.32	0.00
TOPORS	1.54	3.99	0.00
ETV7	1.09	1.25	0.56
NOTCH2	4.73	5.30	0.02
CREB3	3.40	4.81	0.00
APOL6	3.30	5.27	0.00
PRKG1	4.80	2.18	0.00
FLNC	4.90	2.23	0.00
PPP1CA	5.11	6.34	0.00
CACNA1C	3.23	1.00	0.00
ROCK2	6.81	5.37	0.00
EYA4	5.52	3.87	0.00
CREB5	3.93	3.26	0.14
PPPR12A	4.68	3.22	0.00
GCH1	1.21	1.50	0.18
NPY	0.73	0.83	0.74
SDF2L1	3.58	4.64	0.00

The results of the qRT-PCR analysis shown in **Table 3** show that 18 out of 23 selected genes (82%) were differentially expressed between Pt-r and Pr-s patients. In cohort A we confirmed as differentially expressed 20 over 23 elements (87%); *GCH1, NPY* and *PPP1CA* did not show significantly different expression between Pt-s and Pt-r. Instead, on cohort B, the same three genes were not confirmed, with the addition of *ETV7* and *CREB5*.

In order to quantify the prediction obtained from these 18 genes we identified the expression threshold (defined as the cut-off maximizing the sum of specificity and sensitivity) in cohort A and evaluated the accuracy of the prediction in cohort B using the same thresholds (Supplementary Table S3). As expected, in cohort A all the 18 genes show excellent levels of specificity (min 75% - max 100%) and sensitivity (min 75% - max 95%), while in cohort B we observed high levels of accuracy (>70%) for *CACNA1C, FLNC, PPPR12A, PRKG1, APOL6* genes. In the following, this selected list of genes is referred to SI signature.

Figure 11. *Integrated (mRNA and miRNA) network obtained through the comparison of Pt-s and Pt-r samples (cohort A) using pathway-based approach. The colors of the nodes represent the log fold change (Pt-r/Pt-s), the border color indicates the cohorts in which the node has been validated. In the network different functional areas are highlighted according to the biological and molecular function of the nodes.*



7.5 PROGNOSTIC PERFORMANCE OF SI SIGNATURE

The presence of an association between SI signature expression and primary resistance to Pt-based chemotherapy does not necessarily imply a prognostic value for the SI signature. Thus, qRT-PCR expression values were used to test SI association with survival parameters. To increase the sample size, analysis was performed including also Pt-ps cases enrolled in both cohorts, A and B (n= 242 for OS and n=238 for PFS, see Material and Methods section). We identified age and RT as clinical factors in univariate approach (**Table 4**) and then we included them in univariate and multivariate Cox proportional hazards models to assess the prognostic significance of each gene.

Table 4. *Univariate Cox proportional hazards model of clinical features.*

	OS				PFS			
	HR	**CI 95% lower**	**CI 95% upper**	**P**	**HR**	**CI 95% lower**	**CI 95% upper**	**P**
Age	1.02	1.01	1.03	<0.001	1.02	1.01	1.03	<0.001
RT	1.9	1.53	2.37	<0.001	1.63	1.35	1.97	<0.001
Stage	0.91	0.62	1.36	0.65	1.1	0.77	1.56	0.62

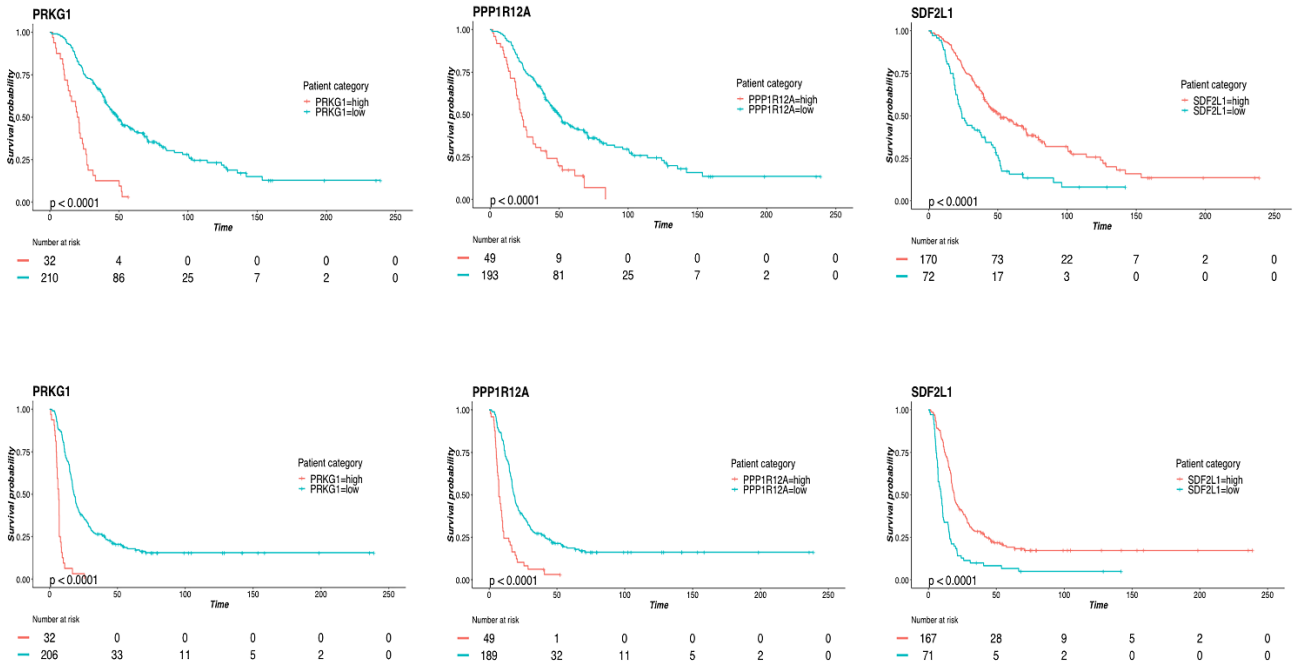
Results selected 15 out of the 19 genes of signature SI (84%) as significantly associated with both OS and PFS in uni- and multivariate analyses (**Table 5** and Supplementary Figures 1 and 2). In the following, this list of genes is called signature SII.

Table 5. *Univariate and multivariate survival models of the 19 DEG elements of the network on cohort A+B. For each gene HR with 95% confidence interval and p-value is reported for either OS and PFS. In grey significant genes both in uni- and multivariate analysis (OS and PFS).*

	OS								PFS							
	UNIVARIATE				MULTIVARIATE				UNIVARIATE				MULTIVARIATE			
GENE	HR	CI 95% lower	CI 95% upper	P	HR	CI 95% lower	CI 95% upper	P	HR	CI 95% lower	CI 95% upper	P	HR	CI 95% lower	CI 95% upper	P
APOL6	0.52	0.38	0.71	<0.001	0.54	0.39	0.74	<0.001	0.44	0.32	0.60	<0.001	0.47	0.34	0.63	<0.001
BAD1	0.49	0.36	0.67	<0.001	0.49	0.36	0.67	<0.001	0.61	0.46	0.81	<0.001	0.57	0.43	0.76	<0.001
CACNA1C	2.01	1.46	2.77	<0.001	2.08	1.50	2.88	<0.001	2.05	1.53	2.75	<0.001	2.23	1.65	3.00	<0.001
CREB3	0.50	0.36	0.68	<0.001	0.53	0.39	0.74	<0.001	0.50	0.37	0.68	<0.001	0.51	0.38	0.70	<0.001
EYA4	1.84	1.17	2.88	0.01	1.79	1.14	2.81	0.01	1.73	1.09	2.74	0.02	1.74	1.09	2.76	0.02
FANCA	0.54	0.39	0.75	<0.001	0.53	0.38	0.73	<0.001	0.55	0.41	0.74	<0.001	0.53	0.40	0.72	<0.001
FANCI	0.45	0.32	0.64	<0.001	0.51	0.36	0.73	<0.001	0.56	0.41	0.76	<0.001	0.60	0.44	0.82	<0.001
FLNC	2.45	1.77	3.38	<0.001	2.49	1.80	3.45	<0.001	2.77	2.02	3.80	<0.001	2.93	2.12	4.04	<0.001
HLA-A	0.60	0.42	0.86	0.01	0.71	0.49	1.02	0.07	0.53	0.37	0.76	<0.001	0.62	0.43	0.89	0.01
HLA-F	0.45	0.33	0.61	<0.001	0.50	0.37	0.69	<0.001	0.44	0.33	0.59	<0.001	0.47	0.35	0.62	<0.001
NOTCH2	0.62	0.40	0.96	0.03	0.65	0.42	1.01	0.06	0.65	0.43	0.97	0.04	0.65	0.43	0.98	0.04
PIGR	0.68	0.49	0.93	0.02	0.61	0.44	0.84	<0.001	0.65	0.49	0.88	<0.001	0.58	0.43	0.79	<0.001
PRKG1	4.13	2.75	6.21	<0.001	3.52	2.31	5.36	<0.001	6.66	4.38	10.13	<0.001	5.66	3.69	8.66	<0.001
ROCK2	2.04	1.43	2.93	<0.001	2.12	1.48	3.05	<0.001	2.28	1.62	3.21	<0.001	2.29	1.62	3.24	<0.001
STAT1	0.83	0.60	1.14	0.24	0.99	0.72	1.36	0.94	0.84	0.60	1.16	0.29	0.98	0.70	1.37	0.90
TOPORS	0.51	0.36	0.72	<0.001	0.49	0.34	0.70	<0.001	0.48	0.35	0.67	<0.001	0.46	0.33	0.63	<0.001
PPP1R12A	2.59	1.82	3.70	<0.001	2.69	1.88	3.85	<0.001	3.04	2.17	4.26	<0.001	3.32	2.36	4.67	<0.001
SDF2L1	0.47	0.34	0.65	<0.001	0.49	0.35	0.67	<0.001	0.43	0.32	0.58	<0.001	0.43	0.32	0.58	<0.001

We focused on the three genes with the strongest HR, namely *PPP1R12A, SDF2L1* and the hub of the network *PRKG1*. Patients with high levels of *PRKG1, PPP1R12A* genes and low levels of *SDF2L1* had a lower probability of survival. When analyzed with respect to PFS and OS, *PRKG1, PPP1R12A*, and *SDF2L1* maintained their statistically significant association with survival (**Figure 12** and **Table 5**). In conclusion, signature SII is made up of 16 genes that are both prognostic and predictive of response to therapy.

Figure 12. *Kaplan Mayer curves and log-rank test p-values of the validated 3 genes either using OS (panel A) and PFS (Panel B) as survival measures. At the bottom of each curve the patients risk table is reported.*



7.6 SIGNATURE SII VALIDATION ACROSS AN EXTERNAL DATASET

To further strength the prognostic independence of signature SII, an external cohort of HGSOC gene expression profiles (cohort C, Curated Ovarian Cancer Dataset [59]) was used. From the whole database of 2970 cases across five databases, expression profiles of 838 platinum-treated patients with complete follow-up including therapeutic outcome (PFS and OS) were selected.

In cohort C we decided to use PFS as an approximation of PFI to classify patients as Pt-s and Pt-r. Using this approach, seven genes (47%) of SII were confirmed as differentially expressed between Pt-s and Pt-r, namely *PRKG1, SDF2L1, PPP1R12A, TOPORS, CREB3, EYA4, CACNA1C*.

Finally, of these seven genes, we found that three genes of SII, namely *PRKG1, SDF2L1* and *PPP1R12A* are significantly associated in uni- and multivariate analyses with OS and PFS (**Table 6**). Patients with high expression levels of PRKG1 and PPP1R12A and low expression levels of SDF2L1 had a shorter survival than those with low levels of *PRKG1* and *PPP1R12A* and high levels of *SDF2L1*. Specifically, in the multivariate model where expression levels are adjusted for age and TR, *PRKG1, PPP1R12A*, and *SDF2L1* maintained their statistically significant association with survival (**Table 6**).

Table 6. *Univariate and multivariate survival models of the significant network elements in cohort C. For each gene HR with 95% confidence interval and p-value is reported for either OS and PFS. In grey significant genes both in uni- and multivariate analysis (OS and PFS).*

	Pt-r* vs Pt-s	OS								PFS UNIVARIATE							
		UNIVARIATE				MULTIIVARIATE				UNIVARIATE				MULTIIVARIATE			
GENE	**Combined p**	**HR**	**CI 95% lower**	**CI 95% upper**	**P**	**HR**	**CI 95% lower**	**CI 95% upper**	**P**	**HR**	**CI 95% lower**	**CI 95% upper**	**P**	**HR**	**CI 95% lower**	**CI 95% upper**	**P**
PRKG1	**0,001187**	1.18	1.08	1.30	**<0.001**	1.15	1.05	1.27	**<0.001**	1.14	1.05	1.24	**<0.001**	1.14	1.05	1.24	**<0.001**
SDF2L1	**0,044751**	0.84	0.77	0.92	**<0.001**	0.85	0.78	0.93	**<0.001**	0.87	0.8	0.94	**<0.001**	0.88	0.82	0.96	**<0.001**
PPP1R12A	**0,011071**	1.15	1.05	1.26	**<0.001**	1.12	1.02	1.23	**0.01**	1.11	1.02	1.2	**0.01**	1.10	1.02	1.19	**0.02**
FANCI	**0,001197**	0.98	0.90	1.07	0.70	1.00	0.92	1.09	0.98	0.94	0.87	1.02	0.12	0.93	0.86	1.01	0.07
TOPORS	0,170422	0.90	0.83	0.98	**0.02**	0.93	0.85	1.01	0.10	0.93	0.86	1.01	0.09	0.93	0.86	1.01	0.07
CREB3	0,136748	0.87	0.79	0.95	**<0.001**	0.89	0.81	0.97	**0.01**	0.9	0.84	0.98	**0.01**	0.93	0.86	1.10	0.08
FLNC	**3,06E-05**	1.11	1.02	1.21	**0.02**	1.10	1.01	1.20	**0.04**	1.07	0.99	1.15	0.08	1.06	0.98	1.15	0.12
FANCA	**0,003142**	0.99	0.90	1.07	0.73	1.01	0.92	1.09	0.90	0.95	0.88	1.03	0.23	0.95	0.88	1.03	0.26
EYA4	**6,74E-06**	1.13	1.04	1.23	**<0.001**	1.08	0.99	1.18	0.09	1.06	0.98	1.14	0.16	1.04	0.96	1.13	0.30
PIGR	**0,030124**	0.92	0.84	1.02	0.10	0.95	0.87	1.05	0.32	1	0.93	1.08	0.93	1.03	0.96	1.11	0.46
HLA-F	**0,023082**	0.95	0.87	1.03	0.24	0.99	0.90	1.07	0.75	0.95	0.88	1.03	0.24	0.97	0.90	1.05	0.48
CACNA1C	0,051102	1.02	0.94	1.12	0.59	1.01	0.93	1.11	0.76	1.02	0.95	1.11	0.55	1.02	0.94	1.10	0.69
ROCK2	**0,001099**	1.08	0.99	1.17	0.10	1.08	0.99	1.18	0.07	1.02	0.95	1.11	0.56	1.01	0.94	1.10	0.72
APOL6	**0,011968**	0.92	0.85	1.00	**0.05**	0.96	0.88	1.05	0.38	0.99	0.92	1.07	0.78	1.01	0.94	1.09	0.78
BAD1	0,216293	1.01	0.93	1.10	0.86	1.00	0.92	1.10	0.94	0.99	0.92	1.07	0.82	1.01	0.93	1.09	0.88

7.7 ASSOCIATION BETWEEN SIII AND HOMOLOGOUS RECOMBINATION DEFICIENCY

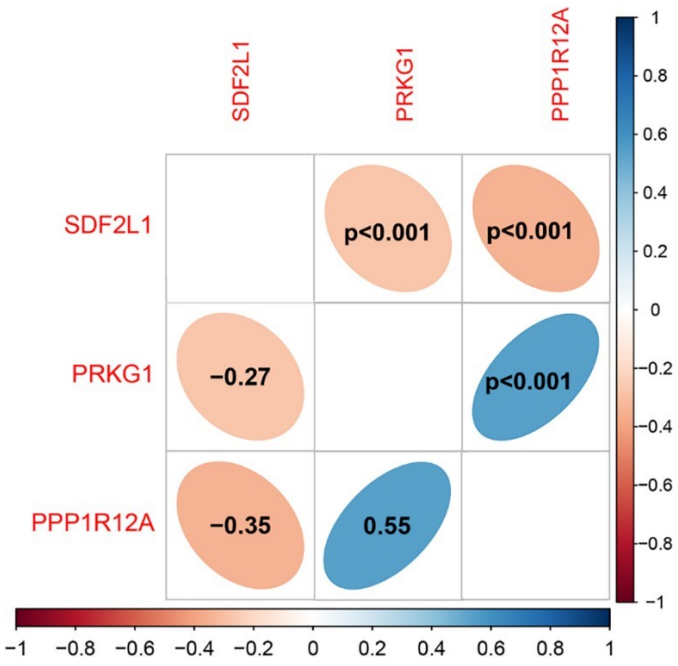
Results from many different clinical trials [89] indicate that platinum sensitivity in HGSOC could be partially explained by deficiency in HR status. Thus, we tested whether the prognostic value of SIII was

associated with HRD status using TCGA data in the cohort C. From the entire TGCA ovarian dataset 314 tumor samples were selected for expression, methylation, mutation and CNV data matched with clinical information regarding age and residual tumor. Of these, 89 samples (28,3%) have HR defects. First, we performed a multivariate analysis to tested if HDR status was prognostic in the TGCA data without considering SIII signature. We found that HDR is not prognostic in TGCA dataset.
Second, using the multivariate Cox proportional hazard model adjusted for age, RT and HRD on each element of SIII, we confirmed that *PRKG1, SDF2L1, PPP1R12A* were significantly associated to both OS and PFS. These results suggest that the prognostic value of the SIII signature is independent of HR status.

7.8 PROGNOSTIC SIGNATURE COMBINATION

The expression levels of *PRKG1, SDF2L1 and PPP1R12A* genes show a variable degree of correlation (**Figure 13**) suggesting a possible synergic effect among them. To identify the combined prognostic contribution of these genes in cohorts A and B, we used a global multivariate survival model where *PRKG1, SDF2L1* and *PPP1R12A* were all included in the model together with RT and age covariates.

Figure 13. *Graphical representation of the correlation among the expression profiles (RT-PCR values in cohort A+B) of the validated 3 genes. On the lower triangles are reported the Spearman correlation index, in the higher triangle the correspondent p-values. The shape of the ellipse within each cell represent the degree and the direction of the correlation, thinner the ellipse higher the correlation, degree of blue means positive correlation, degree of orange negative correlation. We found a significant positive correlation of PPP1R12A with PRKG1 (r=0.55 p-value < 0.001), and a negative correlation with SDF2L1 (r= - 0.35 p-value < 0.001). On the other hand, SDF2L1 shows a significant negative correlation with PRKG1 (r= -0.27 p-value < 0.001).*



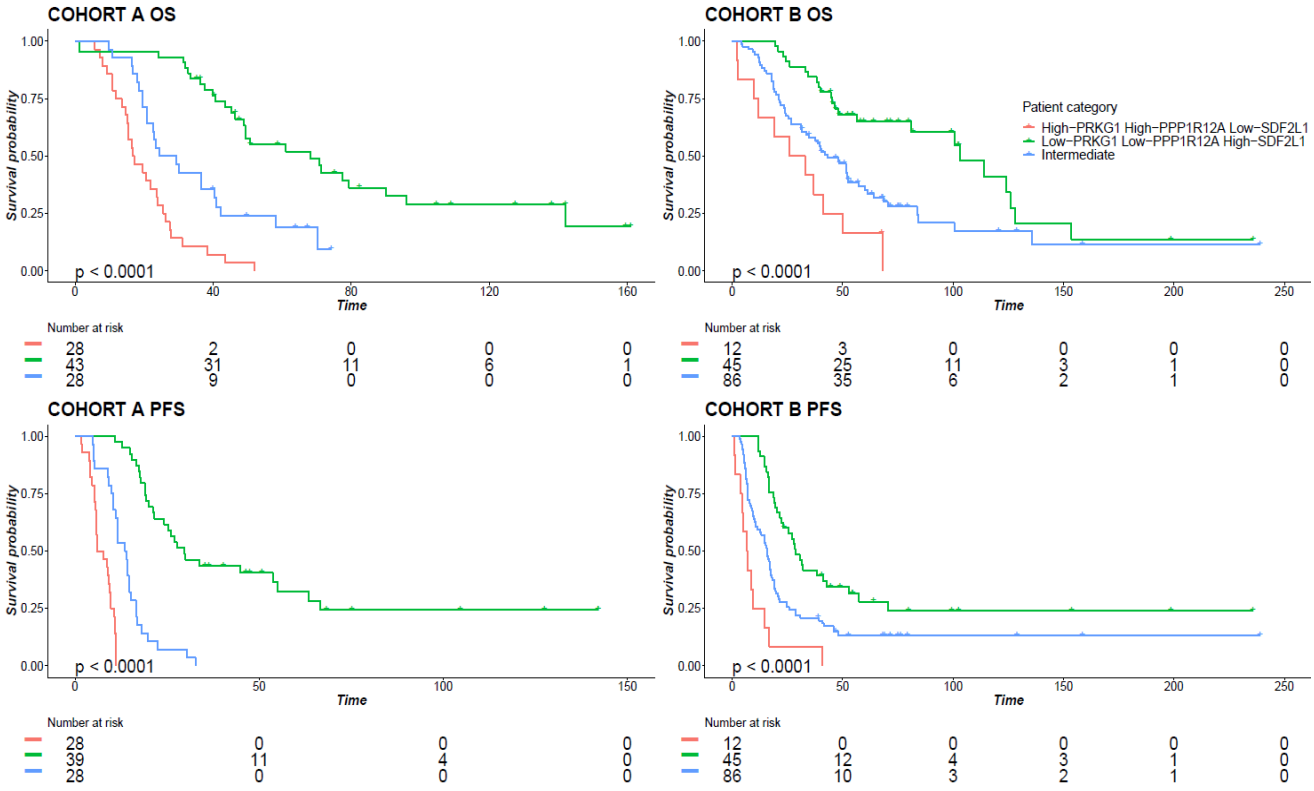
We found that SIII signature genes were all significantly associated with OS or PFS together with RT and age at diagnosis, characterized by the following HRs: *SDF2L1* (PFS HR=0.58 CI95%=0.42-0.80; OS HR=0.68 CI95%=0.47-0.99), *PPPR12A* (PFS HR=1.95 CI95%=1.29-2.97; OS HR=1.70 CI95%=1.08-2.68) and *PRKG1* (PFS HR=2.92 CI95%=1.74-4.89; OS HR=1.97 CI95%=1.15-3.37) (**Table 7**).

Table 7. *Multivariate survival models of the validated 3 genes of the network corrected for residual tumour and age in cohort A+B. For each gene, HR with 95% confidence interval and p-value is reported for either OS and PFS.*

	OS MULTIVARIATE				PFS MULTIVARIATE			
GENE	**HR**	**CI 95% upper**	**CI 95% lower**	**P-VALUE**	**HR**	**CI 95% upper**	**CI 95% lower**	**P-VALUE**
PRKG1	1.97	1.15	3.37	**0.01**	2.92	1.74	4.89	**<0.001**
PPP1R12A	1.70	1.08	2.68	**0.02**	1.95	1.29	2.97	**<0.001**
SDF2L1	0.68	0.47	0.99	**0.04**	0.58	0.42	0.80	**<0.001**
Age	1.02	1.01	1.04	**<0.001**	1.02	1.00	1.03	**0.01**
RT	1.73	1.38	2.16	**<0.001**	1.53	1.26	1.86	**<0.001**

To highlight this synergic effect, we defined for each biomarker a threshold dividing samples into high/low levels of expression (see Methods for details) based on cohort A, then we combined these levels into three classes (high risk: high levels of *PPPR12A* and *PRKG1*, and low levels of *SDF2L1*; low risk: low levels of *PPPR12A* and *PRKG1* and high levels of *SDF2L1*; intermediate: all the other combinations of expression values) testing for survival differences (**Figure 14**), both in cohort A and B using the threshold obtained in cohort A. As expected, we found that the high-risk patients have a significantly poorer prognosis compared to the low-risk ones, and that the intermediate class lays in between the two extremes classes. The results in the cohort A and cohort B confirm that the combination of expression levels of *PPPR12A, SDF2L1* and *PRKG1* is able to significantly stratify patients with good/intermediate/worse prognosis both in OS and PFS.

Figure 14. *Kaplan Mayer curves and log-rank test p-values of the combination of the expression values of PRKG1, SDF2L1 and PPPR12A genes either using OS (panel A) and PFS (Panel B) as survival measures.*



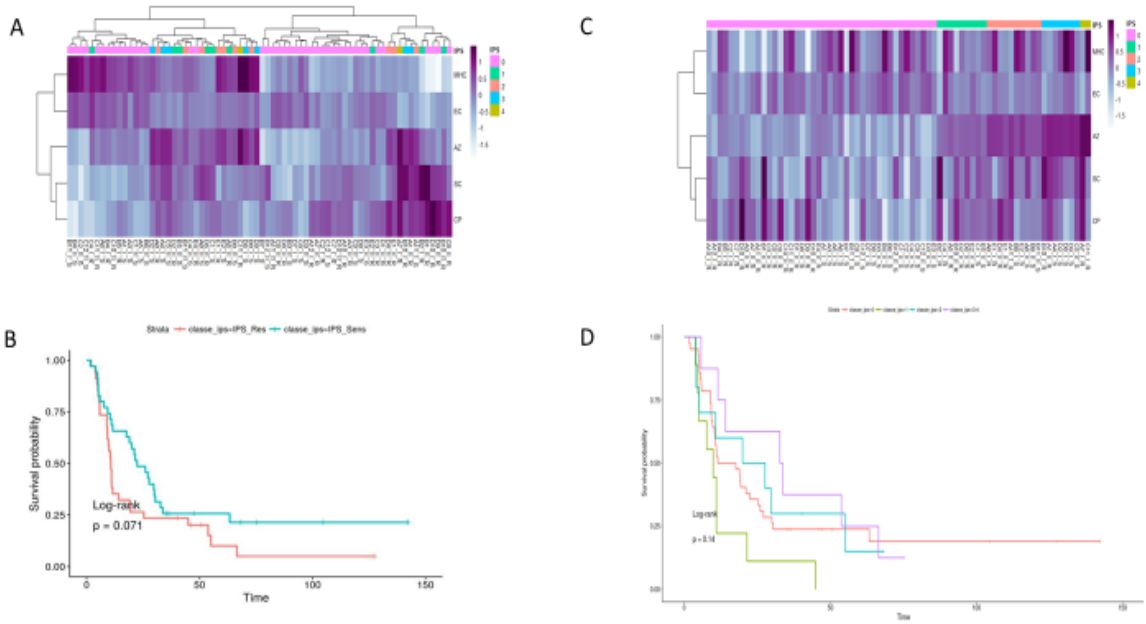
7.9 CANCER IMMUNOLOGY

Some elements of the network are involved as key elements in immune response, for this reason we considered to investigate the immunogenicity of HGSOC. We expected to find out notable differences in genes related to immune response and in tumor infiltrating cells, within resistant and sensitive patients. Also, we take advantage of bioinformatic tools to identify candidate biomarkers for immunotherapy.

7.9.1 TUMOUR INFILTRATING LEUCOCYTES (TIL) COMPOSITION ESTIMATION

Although immunotherapy is transforming the treatment of cancers, the majority of patients still do not respond to immunotherapy [90], making the identification of predictive immune-biomarkers and, the comprehension of the mechanisms involved in therapy resistance an area of intense research.
Using genomic data and bioinformatics tools, it is now possible to computationally provide information on tumour microenvironment, in particular on the composition of the infiltrated immune cells. In order to predict if our ovarian cancer patients could be recruited for the immunotherapy, we used the approach of [35] to characterize samples by class specific z-score or IPS score (see methods section 4.4.1). Although our results indicate that the many of sensitive patients have high expression of genes typical of MHC and effector cells (EC), in contrast to resistant that are characterized by high levels of genes related to suppressor cells (SC) and checkpoint blockers (CP), but these differences are not statistically significant. From the heatmap obtained using the immune z-score it is clear that sensitive and resistant patients are not correctly grouped confirming a high heterogeneity among patients. We also tried to stratify patients, using the IPS score, looking for survival differences, but we didn't obtain significant results (**Figure 15**). These results confirm that HGS-EOC therapy response seems not to be driven by different immunophenotypes.

Figure 15. *A) Heatmap of samples classificated by z-score. B) Kaplan-Meier obtained using classes derived from the heatmap (A) principal branches. C) Heatmap of samples classificated by IPS score. D) Kaplan-Meier obtained using IPS score as classification factor.*



7.9.2 IDENTIFICATION OF NEW NEOANTIGENS

7.9.2.1 COHORT DESCRIPTION (RNA-seq)

Starting from the microarray cohort, the training set for RNA-Seq was composed by 28 HGSOC patients, 14 resistant and 14 sensitive, selected as the extreme of each group. In particular, sensitive have been selected by OS > 102 months and average PFI of 60 months, resistant by OS < 19 months and average PFI of 2 months. Also, we collected 8 normal samples, patients characterized by uterine prolapse and benign uterine pathologies.

Table 8. *Clinical and demographic description of cohort used in RNA-seq experiment.*

Clinical and pathological characteristics		HGSOC Patients (n=28)		Normal Patients (n=8)	
		Pt-r patients (n=14)	Pt-s patients (n=14)	HOSE (n=2)	Tube (n=6)
Average age (range)		67 (54-79)	58 (44-80)	52 (43-65)	
Residual tumor	TR = 0	1 (7 %)	2 (14 %)	-	
	TR > 0	13 (93%)	12 (86%)	-	
FIGO stage	III	12 (86%)	9 (64%)	-	
	IV	2 (14%)	5 (36%)	-	
OS months (range)		19 (8-31)	102 (33-184)	-	
PFI months (range)		2 (0-6)	60 (14-180)	-	
PFS months (range)		8 (5-12)	66 (18-184)	-	

All samples enrolled into RNA-Seq analysis were collected at primary surgery, before chemotherapy, from patients with stage III-IV, according to FIGO guidelines [6].
The median age at diagnosis was 67 for resistant and 58 for sensitive and the median follow-up time was 3 years in both classes (**Table 8**).

7.9.2.2 ANTIGENOME IDENTIFICATION

Studies have reported that the probability of immunotherapy success is reliant on immune response to neoantigens, like *PD-1/PD-L1* for example. This increasing attention in identifying neoantigen sequences has led to an increase in the number of new bioinformatic tools, like TIminer [75]. This tool performs integrative immunogenomic analyses using NGS data, integrating different bioinformatics tools starting from RNA-seq reads and somatic DNA mutations (see methods for details).
Applying TIMINER approach on our cohort we found an average of 103 and 84 neoantigens for respectively resistant and sensitive patients. Considering the mutations, we have 31 and 22 genes involved. This final selection of neoantigens allowed us to compare resistant and sensitive samples and to analyze the differences. Only putative neoantigens expressed at least on 50% of patients have been selected. Summarizing by genes, we obtained 13 genes specific of the resistant class and 14 genes of the sensitive class, within which 9 were in common between the two classes. Moreover, collapsing the results of these 9 genes by their specific mutations, we had 5 mutations in common, 4 resistant specific

and none specific mutations for sensitive samples. Taking together our results, we concluded that the neoantigens identified are not clearly associated to each class. These findings highlighted the similarity within classes in terms of genes and mutations characterizing neoantigens. Indeed, we need to expand our cohort in order to select neoantigens expressed in at least on 70% of patients and, then, to amplify differences between resistant and sensitive samples.

7.10 CIRCULAR RNA IN HGS-EOC

In recent years, products of aberrant splicing events have been associated to specific diseases, defining different roles and functions. Our regulatory network includes expression data of genes, and alternative transcripts may have a correlation with it. In particular, we focused on circRNAs and their function in cancers, as potentially novel biomarkers or therapeutic agents in ovarian cancer.

7.10.1 CIRCULAR RNA INDENTIFICATION

To investigate the presence of circular RNAs, we used RNA-seq experiments described in previous section (6.9.2.1). The unaligned paired reads were independently aligned to the genome, to find unique anchor positions. Only the reads that completely aligned to the genome, with a reverse orientation breakpoint flanked by splice sites, were considered as an indication of a circular splicing events. The result obtained, in terms of number, between find_circ, CIRI2 and circRNA_finder is comparable, with an average of 5000 unfiltered circular RNAs.

Putative circular splicing events were discarded if they did not meet the following thresholds: i) less than 3 back splicing events, ii) detected in all of 28 cancer samples and iii) longer less than 10kb.

After the filtering step we obtained 129 circRNAs using find_circ, 228 with CIRI2 and 96 with circRNA_finder. To guarantee a real estimation of circRNAs abundance we used sailfish-cir [85], generating output files with i) length, ii) effective length, iii) TPM (transcript per million) and iv) number of reads, that we considered a normalized value of back splicing events. Evaluating these data, all of the tools confirmed the correspondence within back splicing events and number of reads, revealing CIRI2 as most reliable tool in de novo detection because of the better proportion.

7.10.2 CIRCULAR RNA DIFFERENTIAL ANALYSIS

Filtered circular RNAs were stored in R matrix format in order to be processed by Bioconductor package, DESeq2 [86]. In find_circ we found 12 circRNAs differentially expressed (DE) with pvalue < 0.05 and a max logFC (fold change Resistant/Sensitive) of |1.29|. Instead, with circRNA_finder 15 circRNAs were selected as differentially expressed with pvalue <0.05 and a max logFC of |1.08|. Finally, with CIRI2 we found significant differences in 22 circRNAs, with a max logFC of |1.28|. To achieve a solid result, we performed the intersection of the three tools results obtaining 5 DE circRNAs, originated from transcript belonging to *MAN1A2, ZBTB44, SHKBP1, PTK2, SLC8A1* genes (**Table 9**).

Table 9. *5 CircRNAs differentially expressed obtained after DeSeq2 analysis.*

CIRCRNA	Chromosome	logFC	P-VALUE
MAN1A2	1	0.87	0.01
ZBTB44	11	1.21	0.02
SHKBP1	19	-0.85	0.04
PTK2	8	0.51	0.00
SLC8A1	2	0.64	0.00

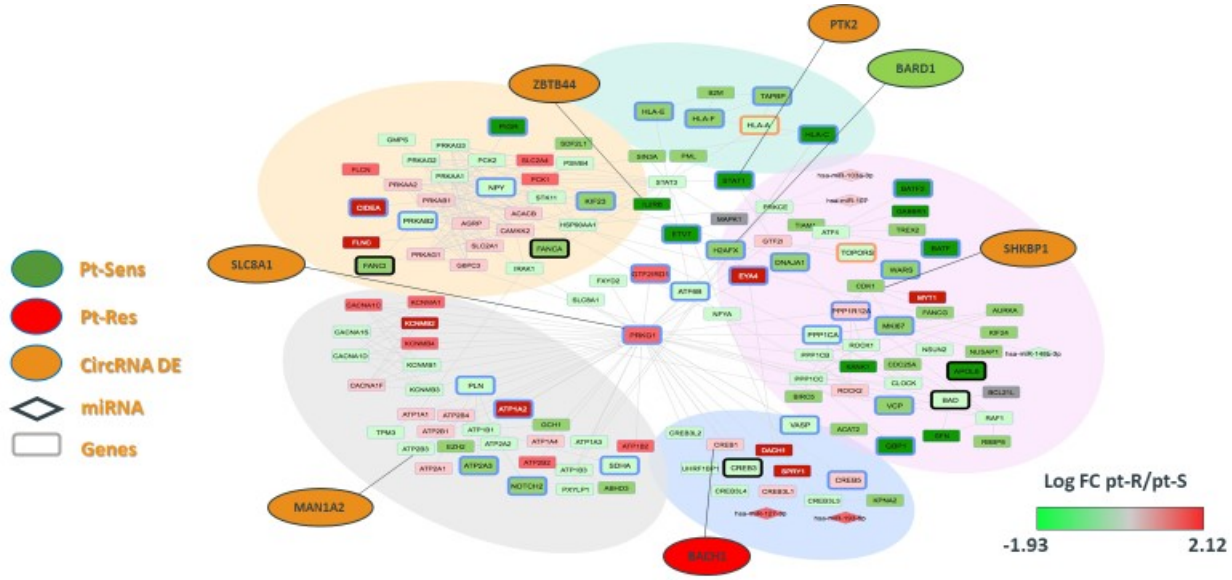
7.10.3 SELECTION OF CLASS SPECIFIC CIRCULAR RNAS

Using CIRI2 we found 111 circRNAs specific for resistant samples and 140 for sensitive class, using find_circ we found 49 circRNAs specific for resistant and 61 for sensitive and, at last, 52 resistant specific and 50 sensitive using circRNA_finder. Although we considered CIRI2 the most reliable tool, we applied the same approach used in differential analysis to obtain a robust result. The intersection of the results of the three tools are selected as reliable, giving us a final selection of 13 circular for resistant and 6 for sensitive samples.

Apart from the expression of circRNAs in tumor samples, it is important to check the expression of these molecules also in healthy patients with the idea that a good biomarker if expressed in tumors (or differentially expressed between resistant and sensible patients) have to be absent in healthy samples to decrease false positive results. Thus, we check the expression of circRNAs in healthy samples obtaining

55 circRNAs expressed in healthy and absent in tumor samples (intersection of 308 CIRI2, 156 find_circ, 118 circRNA_finder).

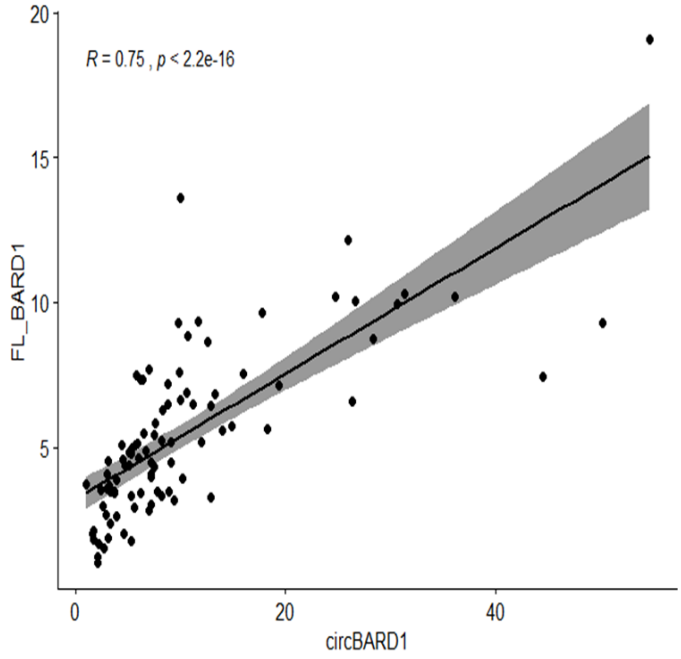
Figure 16. *In the network circRNAs have been linked using protein-protein interaction. 5 DE circRNA (orange), one sensitive specific (green) and one resistant specific (red).*



7.10.4 CIRCULAR RNAS AND THEIR BIOLOGICAL CONTEXT

Recent studies, and experimental validations, confirmed the involvement of circular RNAs in gene regulation. We took advantage of this aspect to connect circRNAs to the genes of our network (**Figure 16**). Some evidences have been reported showing a good correlation between the expression of a circRNA and its linear transcript. For example, we observed this good correlation for *BARD1* (**Figure 17**). Considering the miRNA sponge activity of circRNAs, this correlation may be due to the ability by the circRNA of sequestering those miRNAs that in physiological conditions act as negative regulator of its linear transcript (as in general circRNA and its linear RNA shares miRNA binding batches).

Figure 17. *Pearson's correlation of qRT-PCR normalized values between linear transcript of BARD1 and circBARD1.*



Under this hypothesis, we linked DE circRNA in our regulatory network whenever its linear belong to a gene that interacts with an element of the network. It is clear that we are speculating a sponge-like action of the circRNAs, functions that deserve to be validated with specific assays. However, these speculations can give us a hypothetical idea of the influence of the identified circRNAs on the validated network we found in the previous sections.

Using this approach, we found that 7 circRNAs (5 DE and 2 class specific, one for each class) might influence the network through their genes binding respectively *STAT1, CDK1, ATP2B3, IL2RB, H2AFX, CREB1*. It is of note that one of these interactors is the hub of the network *PRKG1*. The other target genes belong to all areas of the network, except for the lipid metabolism area. This may be representing an important clue of strong connection between circRNAs and network in platinum resistance.

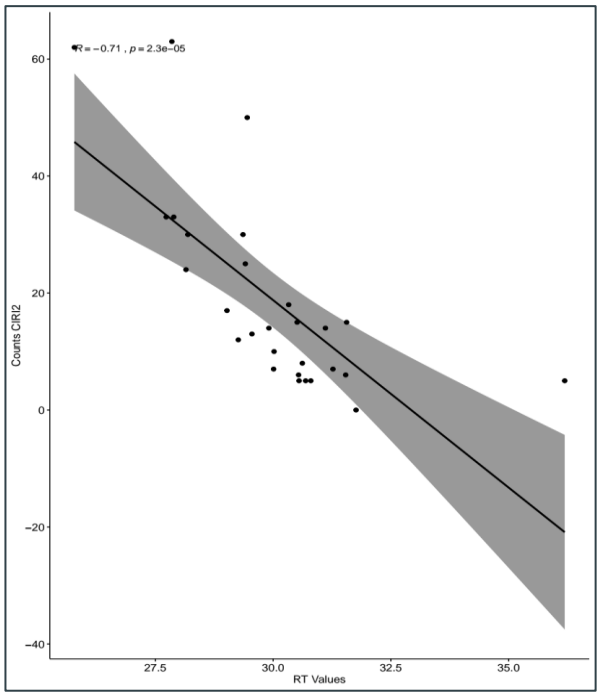
7.10.5 CIRCULAR RNA VALIDATION

Moreover, circBARD1 has several characteristics that makes it a good candidate: i) it is sensitive specific, ii) it is involved in pathway of DNA repair and more importantly iii) it has a strong association with BRCA1, a gene strictly related to ovarian cancers.

7.10.5.1 CIRCBARD1 DIFFERENTIAL ANALYSIS by qRT-PCR

In order to proceed toward the validation, we quantified circBARD1 expression through qRT-PCR experiments in the same samples used for RNA-Seq data. We found a good correlation (Pearson correlation index -0.71 pvalue <0.05) between RNA-Seq counts and values (Ct) obtained in qRT-PCR, that is high expression in sensitive patients and almost absent in resistant and normal samples (**Figure 18**).

Figure 18. *Pearson's correlation between qRT-PCR values of circBARD1 and RNA-seq counts of circBARD1.*



Then, we quantified using qRT-PCR the expression of its linear transcript variant i) to confirm the presence of the linear product from which circBARD1 originates and ii) to check their correlation. We confirmed the detectable expression of linear transcript *BARD1* (ENST00000260947.8) and found a significant correlation (r=0.75, pvalue <0.05) with circ-BARD1 (**Figure 17**). Despite the good correlation between RNA-seq counts and qRT-PCR values we did not confirm circBARD1 differential

expression (using normalized 2^-ΔΔCt qRT-PCR measures) between sensitive and resistant patients either using a parametric and a non-parametric test (**Table 10**). However, it is worth to note that this differential expression is confirmed when using raw Ct values. This negative result could be due to many factors ranging from the large individual variability, the small sample size to the lack of stable control genes.

Table 10. *T-test within Pt-s and Pt-r in qRT-PCR and RNA-Seq counts detected in CIRI2.*

	Mean Res	Mean Sens	pvalue
RT 2^-ΔΔCt	8	11	0.3359
CIRI2	11.357	26.642	0.01542

7.10.5.1 CIRCBARD1 SURVIVAL ANALYSIS

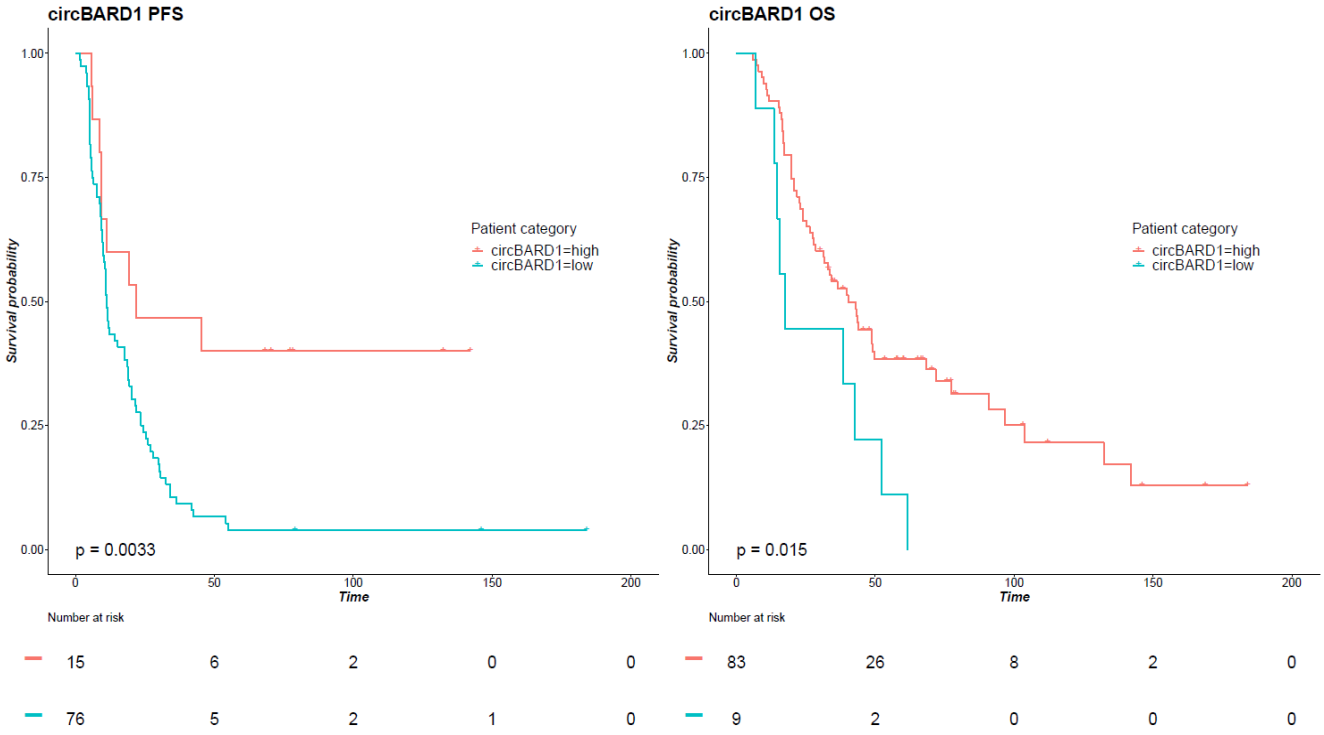
The lack of significant differences of circBARD1 in class comparison, pushed us toward the characterization of circular RNAs as prognostic factors. In order to perform survival analysis, we expanded our cohort from 28 to 92 samples, because this kind of test needs higher number of patients to be statistical reliable. The amplified cohort is also equally distributed within two classes, with 52 resistant and 40 sensitive samples. In addition, we opted for multivariate survival analysis including residual tumor (TR), age and molecular drug (Bevacizumab) as covariates, because of the importance of their role in survival outcome.

Table 11. *Survival analysis of circBARD1 and linear transcript BARD1, in overall survival (OS) and progression free survival (PFS), multivariate approach.*

OS	HR	LOWER	UPPER	P-VALUE		OS	HR	LOWER	UPPER	P-VALUE
2^-ΔΔCt circBARD1	0.97	0.95	1.00	0.05		**2^-ΔΔCt transFL BARD1**	0.95	0.87	1.04	0.33
TR	2.53	1.24	5.17	0.01		**TR**	2.46	1.20	5.02	0.01
Age	1.03	1.01	1.06	0.00		**Age**	1.02	1.00	1.05	0.01
Bevacizumab	0.76	0.33	1.71	0.50		**Bevacizumab**	0.78	0.34	1.77	0.56
PFS	**HR**	**LOWER**	**UPPER**	**P-VALUE**		**PFS**	**HR**	**LOWER**	**UPPER**	**P-VALUE**
2^-ΔΔCt circBARD1	0.95	0.93	0.98	0.00		**2^-ΔΔCt transFL BARD1**	0.93	0.85	1.01	0.10
TR	2.21	1.24	3.93	0.00		**TR**	2.20	1.23	3.94	0.00
Age	1.02	1.00	1.04	0.03		**Age**	1.01	0.99	1.03	0.17
Bevacizumab	1.04	0.57	1.90	0.88		**Bevacizumab**	1.22	0.66	2.24	0.51

We found that circBARD1 was significantly associated with OS or PFS together with RT and age at diagnosis, bevacizumab excluded, characterized by the following HRs: PFS HR=0.95 CI95%=0.93-0.98; OS HR=0.97 CI95%=0.95-1.00 (**Table 11**). Instead, even though covariates showed comparable results, the linear transcript was not significant in survival analysis, consolidating the exclusive prognostic role of circBARD1. In addition, results were also confirmed applying a different method offering wide set of statistical and graphical methods for cutpoint optimization, enabling stratification of population into two groups (**Figure 19**).

Figure 19. *Kaplan-Meyer of circBARD1 obtained determining the optimal cut point for a continuous variable in a coxph model in survival analysis.*



8. DISCUSSION

In the present study we evaluated different bioinformatic approaches to find out prognostic biomarkers able to predict platinum resistance at the time of diagnosis. The first aim was to investigate the mechanisms characterizing the biology of primary resistance against ab-initio Pt-based chemotherapy by network analysis. For this purpose, we integrated transcriptomic profiles to compare patients that did not respond to therapy with those with good and durable responses. This approach allowed us to identify a restricted number of mRNAs and miRNAs the expression of which discriminates between sensitive and resistant cases. The confidence of these findings was strengthened by their validation in a different cohort of cases from an independent tissue biorepository.

The core of our pathways wired 131 mRNAs and five miRNAs, with the *PRKG1* gene playing a role as a hub of the entire network. Our network highlights a very complex scenario, involving several pathways such as transcription and cell cycle regulation, response to DNA damage, lipid and glucose metabolism and antigen presentation (immune system involvement). The biology of tumor therapy resistance is still poorly characterized, and our results might enrich our knowledge of such a complex process. In fact, if alterations in transcription, cell cycle regulation and response to DNA damage are somehow expected, the involvement of lipid and glucose metabolism is intriguing. It is known that cancer cells depend on *de novo* lipid synthesis for the generation of fatty acids to meet the needs of tumor growth, and as a consequence abnormal lipid metabolism play an important role in the pathogenesis of malignancies, including ovarian cancer [91]. At the same time aerobic glycolysis has now been generally accepted as a metabolic hallmark of cancer [92]. However, their causal relationship with cancer progression and Pt-response is still unclear. It is worth noting that some elements of the network have been previously identified as prognostic hints of acquired Pt-based resistance. For example, miR-193-5p -highly expressed in Pt-r patients- is an element of the MIROVAR signature [93] and seven genes of the network (*CLOCK, EZH2, CREB1, TIAM1, CDC25A, ROCK2* and *ROCK1*) are validated targets of some elements of the MIROVAR signature. We chose to validate and confirm the reproducibility of the network analysis, obtaining SI signature, composed by 18 out of 23 selected genes (78%) differentially expressed between Pt-r and Pr-s patients. The presence of an association between SI signature expression and primary resistance to Pt-based chemotherapy, pushed us to test SI association with survival parameters. Results showed a significant association with both OS and PFS in uni- and multivariate analyses, in signature SII, confirming a prognostic role of this selection of genes. Since primary resistance against Pt-based chemotherapy is associated with poor prognosis, we investigated the prognostic role of the SII signature in one of the largest available databases of EOC expression profiles (the Curated Ovarian Database - cohort C). It is important to highlight that these

retrospective databases were not originally intended to be used for studies of chemo-resistance as the primary outcome. Although several clinical parameters have been recorded, the use of different gene expression technologies as well as metrics for optimal treatment response might introduce potential bias in the data and in the analyses. To minimize these biases and to reduce the impact of intra-patient tumor heterogeneity, we used very stringent statistical criteria to limit false positive predictive genes. This strategy identified three genes, *PRKG1, SDF2L1* and *PPP1R12A*, alias signature SIII-as prognostic biomarkers in HGSOC. Moreover, the combination of the expression profiles of signature SIII significantly predicts patient prognosis with a synergic effect. Homologous recombination deficiency (HRD) status is currently one of the most popular genetic paradigms to predict platinum sensitivity in HGSOC [73]. Using the TCGA data, we showed that *PRKG1, SDF2L1* and *PPP1R12A* are associated with patients' survival, independently of HR deficiency status. There are few informations on the biological features of the elements of signature SIII in ovarian cancer, and to our knowledge this is the one of the first study demonstrating their direct involvement in both Pt resistance and prognosis in HGSOC. *PRKG1*, is a cyclic GMP (cGMP)–dependent protein kinase (*PKG*), poorly characterized in terms of mechanism of action, in particular in HGSOC. Recently the cGMP/PKG signaling pathway was found to play an important role as an antiapoptotic mechanism in ovarian cancer cell lines, by promoting cell survival, through interaction with the *SRC* gene [94]. Moreover, the NO/cGMP/PKG signaling pathway has been reported to protect human ovarian cancer cells against both spontaneous and cisplatin-induced apoptosis [95]. The *SDF2L1* gene encodes for a member of the stromal cell derived factors family (SDF) secreted by stromal cells, including fibroblasts. Our results are consistent with the evidence that low levels of *SDF2L1* are associated to poor prognosis, relapse and metastasis in breast [96], colon [97] and ovarian cancer [98]. *PPP1R12A* (protein phosphatase 1 regulatory subunit 12A) belongs to the myosin phosphatase targeting protein (MYPT) family. It is also known as myosin phosphatase target subunit 1 (*MYPT1*). In cancer cells, *PPP1R12A* plays a critical role in major regulatory pathways such those pertaining to Wnt/β-catenin signaling [99] and *PI3K/AKT* [100]. It has also been demonstrated that *PPP1R12A* is a key regulator of the Hippo pathway, and that triggered by external stimuli like *EGF* or *TGβ1* it modulates the nuclear localization and stability of *YAP/TAZ*, thus controlling cell growth, proliferation and *EMT* [101]. Despite this important mechanistic role, few studies address the potential prognostic role of *PPP1R12A*. Genome instability in the *PPP1R12A* locus gene has been found as independent predictor of recurrence and overall survival in colorectal cancer patients receiving oxaliplatin-based adjuvant chemotherapy [102]. It is plausible to hypothesize that *PRKG1, SDF2L1, PPP1R12A* genes are part of a larger prognostic signature, the elements of which have not been identified here due to the stringent statistical thresholds applied and to the use of non-uninform technical platforms for data generation. Future studies could be aimed at expanding the

findings presented here to include other genes, and to improve the prognostic value of the signature. Antigen presentation was one of the biological functions highlighted by our network as a molecular mechanism involved in therapy response. To deeper investigate this issue, we studied, using bioinformatic approaches, the immunogenicity of our HGSOC samples. Currently, new immunotherapeutic approaches based on immune checkpoint inhibitors are changing the landscape in melanoma treatment, however, in ovarian cancer potential success of this therapy relies on better understanding of tumor microenvironment and dominant immunosuppressive pathways [103]. Indeed, immunotherapy in ovarian cancer will have to consider the immune suppressive networks belonging the ovarian tumor microenvironment. For this propose, we used genomic data and bioinformatics tools to computationally provide information on tumour microenvironment, in particular on the composition of the infiltrated immune cells. Immunophenoscore [35] (IPS), a score based on infiltrated immune cells compositions, has been used on our cohort as predictor of response to anti-*CTLA-4* and anti-*PD-1* antibodies. Although we expected sensitive patients more active in immune response and consequently resistant patients valid candidates for immunotherapy our results do not indicate a clear stratification of patients by their estimated level of immune response. This result confirmed the current state of the art, according to which immune-therapies, such as anti-*CTLA-4* and anti-*PD-1* antibodies, do not show significant improvement in ovarian cancer treatment. However, not only tumor microenvironment composition, but also the presence of neoantigens, could determine the response to immune-therapy. This is an important aspect to be considered for HGSOC, given the large number of genomic rearrangements and transcriptome aberrations resulting from homologous recombination defects [104]. In our cohort, sensitive had a higher number of totals neoantigens than resistant patients, also if we gather the specific point mutations and the genes involved. As a consequence, we intersected the selection of neoantigens associated to resistant and sensitive samples in order to analyze differences. Summarizing we obtained 9 genes in common between the two classes of patients, with *EPPK1, MYO9B, MAP2K3, TTN, TSC2* as most represented genes. Moreover, collapsing the results of these 9 genes by their specific mutations, we obtained only 4 resistant specific mutations and none for sensitive samples. Thus, due to the small and heterogeneous results on the differences between resistant and sensitive patients, we are not able to provide solid conclusions on this issue. Thus, although our network proposes immune system as involved in therapy response, we were not able to confirm this involvement in our cohort using bioinformatic approaches. Clearly our results cannot exclude the involvement of the immune system since bioinformatic analyses can be inefficient and gene expression data can be noisy. We need further experimental assays (such as Mass Spectrometry-Based Immunopeptidomics or immunological screening methods [105]) on tumor samples to confirm our in silico results. Regarding regulatory roles of alternative transcript products in ovarian cancer, the correlation and relation between

them are attractive for researchers. Our regulatory network was identified using expression data of genes. However, a gene may have different transcripts with many different functions and many genes may have "aberrant" splicing in specific diseases. In particular, in recent years, numerous studies focused on the expression of circRNAs and their function in cancers, suggesting circRNAs as potentially novel biomarkers or therapeutic agents [106]. Initially, circRNAs were considered as errors during RNA splicing. However, thanks to the development of NGS technology and bioinformatic tools accumulating evidences revealed the critical regulatory roles of them in tumorigenesis, with cancer-type specific distinguishable expression levels. CircRNAs biological functions, such as interactions with specific mRNA, protein and specifically miRNAs could play profound roles as regulator of cancer development, and in chemotherapy resistance. Given their emerging key role in cancer, to complement our identified regulatory network with the information on this novel class of transcript, we investigate the expression of circular RNAs in ovarian cancer. Our approach of searching for back spliced junctions has been built ex novo, using known bioinformatic tools specific for circRNAs de novo identification. The combined use of these tools and the application of restrictive thresholds conferred additional statistical power to the discovery and quantification of candidate circRNAs. CircRNAs identified as differential expressed between sensitive and resistant patients derived from *PTK2, SLC8A1, ZBTB44, TMEM165* and *ZNF124* genes. We found a significantly lower abundance of circRNA isoforms in ovarian cancer compared to linear RNA. Furthermore, the genes encoding differentially expressed circRNAs are connected to some of genes belonging to our network, indicating an involvement of circRNA in chemotherapy resistance in ovarian cancer. We were able to link to the network two circRNAs class specific, *BACH1* for resistant and *BARD1* for sensitive. Regarding circular RNA function, they have been shown to harbor multiple binding sites for microRNAs, and a miRNA in turn can alter the expression of hundreds of genes. For example, circBARD1 has been recently studied as regulator of cell apoptosis binding miR-3942 [107]. In order to consolidate our results, we performed experimental validation on circBARD1, because of its involvement in HRD, network connection and in preliminary studies as miRNA sponge interactor. Results achieved by qRT-PCR confirmed the correlation with back splicing events detected in RNA-seq, also for linear transcript variant of *BARD1*. As we expected, circBARD1 has been confirmed as a circRNA sensitive specific and with no significant differential expression compared to resistant, moving our attention to a possible role in survival analysis. It worth to notice that only the circular variant of *BARD1* was significant in univariate and multivariate approach in survival analysis, giving strength to the hypothesis of circBARD1 as prognostic factor. Our results show a complex network wiring multiple mRNAs, miRNAs and circRNAs involved in several biological processes. These emerging roles of circRNAs to communicate via miRNA prompts for new exciting opportunities in research to uncover the complex biological cross-talk, their role in

carcinogenesis and efficacies as biomarkers for ovarian cancer diagnosis and prognosis. Without doubt, this work contributes to the knowledge of the molecular mechanisms altered in ovarian cancer therapy resistance. The advantage of data integration and network construction gave us a promising opportunity to identify putative biomarkers to translate into clinic or new potential drug target. In particular, deeper analysis on non-coding data may give new hints toward new strategies, supported by bioinformatics, in order to improve the poor outcome of ovarian cancer patients. Subsequently novel approaches can be identified to early diagnosis, prognosis, follow up and therapeutics of ovarian cancers.

BIBLIOGRAPHY

1. Ferlay, J., et al., *Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods.* Int. J. Cancer, 2019. **144**(8): p. 1941-1953.
2. Maringe, C., et al., *Stage at diagnosis and ovarian cancer survival: evidence from the International Cancer Benchmarking Partnership.* Gynecol. Oncol., 2012. **127**(1): p. 75-82.
3. Lalwani, N., et al., *Histologic, molecular, and cytogenetic features of ovarian cancers: implications for diagnosis and treatment.* Radiographics, 2011. **31**(3): p. 625-646.
4. Lisio, M.-A., et al., *High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints.* Int. J. Mol. Sci., 2019. **20**(4).
5. Kurman, R.J. and c. Centre international de recherche sur le, *WHO Classification of Tumours of Female Reproductive Organs*. 2014: World Health Organization. 307.
6. Prat, J. and F.C.o.G. Oncology, *Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum: Abridged Republication of Guidelines From the International Federation of Gynecology and Obstetrics (FIGO).* Obstet. Gynecol., 2015. **126**(1): p. 171-174.
7. Edge, S.B. and C.C. Compton, *The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM.* Ann. Surg. Oncol., 2010. **17**(6): p. 1471-1474.
8. National Cancer Institute . Cancer Statistics, B. and J.L. Young, *SEER summary staging manual 2000: codes and coding instructions*. 2001. 293.
9. Kaldawy, A., et al., *Low-grade serous ovarian cancer: A review.* Gynecol Oncol, 2016. **143**(2): p. 433-438.
10. Ciucci, A., et al., *Ovarian low and high grade serous carcinomas: hidden divergent features in the tumor microenvironment.* Oncotarget, 2016. **7**(42): p. 68033-68043.
11. Webb, P.M. and S.J. Jordan, *Epidemiology of epithelial ovarian cancer.* Best Pract. Res. Clin. Obstet. Gynaecol., 2017. **41**: p. 3-14.
12. *Website.*
13. Cunningham, J.M., et al., *Clinical characteristics of ovarian cancer classified by BRCA1, BRCA2, and RAD51C status.* Sci. Rep., 2014. **4**: p. 4026.
14. Testa, U., et al., *Ovarian Cancers: Genetic Abnormalities, Tumor Heterogeneity and Progression, Clonal Evolution and Cancer Stem Cells.* Medicines (Basel), 2018. **5**(1).
15. Ramus, S.J., et al., *Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer.* J. Natl. Cancer Inst., 2015. **107**(11).
16. Norquist, B.M., et al., *Inherited Mutations in Women With Ovarian Carcinoma.* JAMA Oncology, 2016. **2**(4): p. 482.
17. Wentzensen, N., et al., *Ovarian Cancer Risk Factors by Histologic Subtype: An Analysis From the Ovarian Cancer Cohort Consortium.* J. Clin. Oncol., 2016. **34**(24): p. 2888-2898.
18. Collaborative Group on Epidemiological Studies of Ovarian, C., et al., *Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls.* Lancet, 2008. **371**(9609): p. 303-314.
19. Markman, M., *Optimizing primary chemotherapy in ovarian cancer.* Hematol. Oncol. Clin. North Am., 2003. **17**(4): p. 957-68, viii.
20. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA: A Cancer Journal for Clinicians, 2015. **65**(1): p. 5-29.
21. Vaughan, S., et al., *Rethinking ovarian cancer: recommendations for improving outcomes.* Nat. Rev. Cancer, 2011. **11**(10): p. 719-725.
22. Yap, T.A., C.P. Carden, and S.B. Kaye, *Beyond chemotherapy: targeted therapies in ovarian cancer.* Nature Reviews Cancer, 2009. **9**(3): p. 167-181.
23. Matulonis, U.A., et al., *Ovarian cancer.* Nat Rev Dis Primers, 2016. **2**: p. 16061.

24. Bowtell, D.D., et al., *Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer.* Nat. Rev. Cancer, 2015. **15**(11): p. 668-679.
25. Papa, A., et al., *Update on Poly-ADP-ribose polymerase inhibition for ovarian cancer treatment.* J. Transl. Med., 2016. **14**: p. 267.
26. Jayson, G.C., et al., *Ovarian cancer.* Lancet, 2014. **384**(9951): p. 1376-1388.
27. Pignata, S., et al., *Treatment of recurrent ovarian cancer.* Annals of Oncology, 2017. **28**(suppl_8): p. viii51-viii56.
28. Thigpen, J.T., *Incorporation of Bevacizumab in the Primary Treatment of Ovarian Cancer.* Yearbook of Oncology, 2012. **2012**: p. 110-111.
29. Blagden, S., et al., *3 Afuresertib (GSK2110183), an oral AKT kinase inhibitor, in combination with carboplatin and paclitaxel in recurrent ovarian cancer.* European Journal of Cancer, 2014. **50**: p. 7.
30. Odunsi, K., *Immunotherapy in ovarian cancer.* Ann. Oncol., 2017. **28**(suppl_8): p. viii1-viii7.
31. Hwang, W.-T., et al., *Prognostic significance of tumor-infiltrating T cells in ovarian cancer: a meta-analysis.* Gynecol. Oncol., 2012. **124**(2): p. 192-198.
32. Brahmer, J.R., et al., *Safety and activity of anti-PD-L1 antibody in patients with advanced cancer.* N. Engl. J. Med., 2012. **366**(26): p. 2455-2465.
33. Lu, Y.-C. and P.F. Robbins, *Cancer immunotherapy targeting neoantigens.* Seminars in Immunology, 2016. **28**(1): p. 22-27.
34. Hutchison, S. and A.L. Pritchard, *Identifying neoantigens for use in immunotherapy.* Mamm. Genome, 2018. **29**(11-12): p. 714-730.
35. Charoentong, P., et al., *Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade.* Cell Rep, 2017. **18**(1): p. 248-262.
36. Chen, B., et al., *Profiling Tumor Infiltrating Immune Cells with CIBERSORT.* Methods Mol Biol, 2018. **1711**: p. 243-259.
37. Mattick, J.S. and I.V. Makunin, *Non-coding RNA.* Hum. Mol. Genet., 2006. **15 Spec No 1**: p. R17-29.
38. Esteller, M., *Non-coding RNAs in human disease.* Nat. Rev. Genet., 2011. **12**(12): p. 861-874.
39. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA.* Nature Reviews Genetics, 2014. **15**(6): p. 423-437.
40. Bracken, C.P., H.S. Scott, and G.J. Goodall, *A network-biology perspective of microRNA function and dysfunction in cancer.* Nat. Rev. Genet., 2016. **17**(12): p. 719-732.
41. Lai, E.C., *Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation.* Nature Genetics, 2002. **30**(4): p. 363-364.
42. Jonas, S. and E. Izaurralde, *Towards a molecular understanding of microRNA-mediated gene silencing.* Nat. Rev. Genet., 2015. **16**(7): p. 421-433.
43. Lu, J., et al., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-838.
44. Hayes, J., P.P. Peruzzi, and S. Lawler, *MicroRNAs in cancer: biomarkers, functions and therapy.* Trends in Molecular Medicine, 2014. **20**(8): p. 460-469.
45. Bagnoli, M., et al., *Development and validation of a microRNA-based signature (MiROvaR) to predict early relapse or progression of epithelial ovarian cancer: a cohort study.* The Lancet Oncology, 2016. **17**(8): p. 1137-1146.
46. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions.* Nature Reviews Genetics, 2009. **10**(3): p. 155-159.
47. Quinn, J.J. and H.Y. Chang, *Unique features of long non-coding RNA biogenesis and function.* Nat. Rev. Genet., 2016. **17**(1): p. 47-62.
48. Ulitsky, I., *Evolution to the rescue: using comparative genomics to understand long non-coding RNAs.* Nat. Rev. Genet., 2016. **17**(10): p. 601-614.
49. Zhang, L.-Q., et al., *Long Noncoding RNA MIR4697HG Promotes Cell Growth and Metastasis in Human Ovarian Cancer.* Anal. Cell. Pathol., 2017. **2017**: p. 8267863.
50. Noto, J.J., C.A. Schmidt, and A.G. Matera, *Engineering and expressing circular RNAs via tRNA splicing.* RNA Biol, 2017. **14**(8): p. 978-984.
51. Meng, X., et al., *Circular RNA: an emerging key player in RNA world.* Brief Bioinform, 2017. **18**(4): p. 547-557.

52. Liang, D. and J.E. Wilusz, *Short intronic repeat sequences facilitate circular RNA production.* Genes Dev, 2014. **28**(20): p. 2233-47.
53. Abe, N., et al., *Rolling Circle Translation of Circular RNA in Living Human Cells.* Sci Rep, 2015. **5**: p. 16435.
54. Li, Z., et al., *Exon-intron circular RNAs regulate transcription in the nucleus.* Nat Struct Mol Biol, 2015. **22**(3): p. 256-64.
55. Ahmed, I., et al., *Altered expression pattern of circular RNAs in primary and metastatic sites of epithelial ovarian carcinoma.* Oncotarget, 2016. **7**(24): p. 36366-36381.
56. Hansen, T.B., et al., *Comparison of circular RNA prediction tools.* Nucleic Acids Res, 2016. **44**(6): p. e58.
57. Pfisterer, J. and J.A. Ledermann, *Management of platinum-sensitive recurrent ovarian cancer.* Semin Oncol, 2006. **33**(2 Suppl 6): p. S12-6.
58. Martini, P., et al., *Along signal paths: an empirical gene set approach exploiting pathway topology.* Nucleic Acids Res, 2013. **41**(1): p. e19.
59. Ganzfried, B.F., et al., *curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome.* Database (Oxford), 2013. **2013**: p. bat013.
60. Calura, E., et al., *A prognostic regulatory pathway in stage I epithelial ovarian cancer: new hints for the poor prognosis assessment.* Ann Oncol, 2016. **27**(8): p. 1511-9.
61. Gandolfo, L.C. and T.P. Speed, *RLE plots: Visualizing unwanted variation in high dimensional data.* PLoS One, 2018. **13**(2): p. e0191629.
62. Risso, D., et al., *Normalization of RNA-seq data using factor analysis of control genes or samples.* Nat Biotechnol, 2014. **32**(9): p. 896-902.
63. Jacob, L., J.A. Gagnon-Bartsch, and T.P. Speed, *Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed.* Biostatistics, 2016. **17**(1): p. 16-28.
64. Calura, E., et al., *Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles.* Nucleic Acids Res, 2014. **42**(11): p. e96.
65. Sales, G., et al., *graphite - a Bioconductor package to convert pathway topology to gene network.* BMC Bioinformatics, 2012. **13**: p. 20.
66. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.
67. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
68. Stark, C., et al., *BioGRID: a general repository for interaction datasets.* Nucleic Acids Res, 2006. **34**(Database issue): p. D535-9.
69. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
70. Calura, E., et al., *MiRNA landscape in stage I epithelial ovarian cancer defines the histotype specificities.* Clin Cancer Res, 2013. **19**(15): p. 4114-23.
71. Romani, C., et al., *Identification of optimal reference genes for gene expression normalization in a wide cohort of endometrioid endometrial carcinoma tissues.* PLoS One, 2014. **9**(12): p. e113781.
72. Irizarry, R.A., et al., *Comprehensive high-throughput arrays for relative methylation (CHARM).* Genome Res, 2008. **18**(5): p. 780-90.
73. Konstantinopoulos, P.A., et al., *Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer.* Cancer Discov, 2015. **5**(11): p. 1137-54.
74. Lu, J., et al., *Correlation between gene expression and mutator phenotype predicts homologous recombination deficiency and outcome in ovarian cancer.* J Mol Med (Berl), 2014. **92**(11): p. 1159-68.
75. Tappeiner, E., et al., *TIminer: NGS data mining pipeline for cancer immunology and immunotherapy.* Bioinformatics, 2017. **33**(19): p. 3140-3141.
76. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.
77. Szolek, A., et al., *OptiType: precision HLA typing from next-generation sequencing data.* Bioinformatics, 2014. **30**(23): p. 3310-6.
78. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

79. Nielsen, M. and M. Andreatta, *NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets.* Genome Med, 2016. **8**(1): p. 33.
80. Gao, Y., J. Wang, and F. Zhao, *CIRI: an efficient and unbiased algorithm for de novo circular RNA identification.* Genome Biol, 2015. **16**: p. 4.
81. Gao, Y., J. Zhang, and F. Zhao, *Circular RNA identification based on multiple seed matching.* Brief Bioinform, 2018. **19**(5): p. 803-810.
82. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency.* Nature, 2013. **495**(7441): p. 333-8.
83. Westholm, J.O., et al., *Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation.* Cell Rep, 2014. **9**(5): p. 1966-1980.
84. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
85. Li, M., et al., *Quantifying circular RNA expression from RNA-seq data using model-based framework.* Bioinformatics, 2017. **33**(14): p. 2131-2139.
86. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
87. Glazar, P., P. Papavasileiou, and N. Rajewsky, *circBase: a database for circular RNAs.* RNA, 2014. **20**(11): p. 1666-70.
88. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. **474**(7353): p. 609-15.
89. Moore, K., et al., *Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer.* N Engl J Med, 2018. **379**(26): p. 2495-2505.
90. Coukos, G., J. Tanyi, and L.E. Kandalaft, *Opportunities in immunotherapy of ovarian cancer.* Ann Oncol, 2016. **27 Suppl 1**: p. i11-i15.
91. Pyragius, C.E., et al., *Aberrant lipid metabolism: an emerging diagnostic and therapeutic target in ovarian cancer.* Int J Mol Sci, 2013. **14**(4): p. 7742-56.
92. Hsu, P.P. and D.M. Sabatini, *Cancer cell metabolism: Warburg and beyond.* Cell, 2008. **134**(5): p. 703-7.
93. Bagnoli, M., et al., *Development and validation of a microRNA-based signature (MiROvaR) to predict early relapse or progression of epithelial ovarian cancer: a cohort study.* Lancet Oncol, 2016. **17**(8): p. 1137-1146.
94. Leung, E.L., et al., *Protein kinase G type Ialpha activity in human ovarian cancer cells significantly contributes to enhanced Src activation and DNA synthesis/cell proliferation.* Mol Cancer Res, 2010. **8**(4): p. 578-91.
95. Leung, E.L., et al., *Cisplatin alters nitric oxide synthase levels in human ovarian cancer cells: involvement in p53 regulation and cisplatin resistance.* Br J Cancer, 2008. **98**(11): p. 1803-9.
96. Kang, H., et al., *Transcript analyses of stromal cell derived factors (SDFs): SDF-2, SDF-4 and SDF-5 reveal a different pattern of expression and prognostic association in human breast cancer.* Int J Oncol, 2009. **35**(1): p. 205-11.
97. Vendrell, E., et al., *Genomic and transcriptomic prognostic factors in R0 Dukes B and C colorectal cancer patients.* Int J Oncol, 2007. **30**(5): p. 1099-107.
98. Willis, S., et al., *Single Gene Prognostic Biomarkers in Ovarian Cancer: A Meta-Analysis.* PLoS One, 2016. **11**(2): p. e0149183.
99. Zhou, L., et al., *Merlin-deficient human tumors show loss of contact inhibition and activation of Wnt/beta-catenin signaling linked to the PDGFR/Src and Rac/PAK pathways.* Neoplasia, 2011. **13**(12): p. 1101-12.
100. Rong, R., et al., *Neurofibromatosis 2 (NF2) tumor suppressor merlin inhibits phosphatidylinositol 3-kinase through binding to PIKE-L.* Proc Natl Acad Sci U S A, 2004. **101**(52): p. 18200-5.
101. Serrano, I., et al., *Inactivation of the Hippo tumour suppressor pathway by integrin-linked kinase.* Nat Commun, 2013. **4**: p. 2976.
102. Zhang, C., et al., *PPP1R12A Copy Number Is Associated with Clinical Outcomes of Stage III CRC Receiving Oxaliplatin-Based Chemotherapy.* Mediators Inflamm, 2015. **2015**: p. 417184.

103. Cortez, A.J., et al., *Advances in ovarian cancer therapy.* Cancer Chemother Pharmacol, 2018. **81**(1): p. 17-38.
104. Strickland, K.C., et al., *Association and prognostic significance of BRCA1/2-mutation status with neoantigen load, number of tumor-infiltrating lymphocytes and expression of PD-1/PD-L1 in high grade serous ovarian cancer.* Oncotarget, 2016. **7**(12): p. 13587-98.
105. Garcia-Garijo, A., C.A. Fajardo, and A. Gros, *Determinants for Neoantigen Identification.* Front Immunol, 2019. **10**: p. 1392.
106. Shabaninejad, Z., et al., *Circular RNAs in cancer: new insights into functions and implications in ovarian cancer.* J Ovarian Res, 2019. **12**(1): p. 84.
107. Zhao, J., et al., *Circlular RNA BARD1 (Hsa_circ_0001098) overexpression in breast cancer cells with TCDD treatment could promote cell apoptosis via miR-3942/BARD1 axis.* Cell Cycle, 2018. **17**(24): p. 2731-2744.

SUPPLEMENTARY MATERIAL

Supplementary Table S1. *Micrographite results. Pathway: paths of the significant pathways selected to compose the Metapathway. Genes and miRNAs of the path: genes (Entrez ID) and micro RNAs (miRbase ID) involved in selected pathways. Max Score: details available in micrographite documentation* [64].

Pathway	Max Score	Genes and miRNAs of the path
cGMP-PKG signaling pathway	121,01	10672; 2768; 5592; 6093; 9475; 4659; 5592; 6093; 9475; hsa-miR-148b-3p; 4659; 5499; 5500; 5501; 5592; 6093; 9475; 5592; 90993; hsa-miR-127-3p; 10242; 27094; 27345; 3778; 3779; 5592; 776; 10242; 27094; 27345; 3778; 3779; 5592; 775; 10242; 27094; 27345; 3778; 3779; 5592; 778; 10242; 27094; 27345; 3778; 3779; 5592; 779; 5581; 5592; hsa-miR-103a-3p; hsa-miR-107; 5592; 84699; hsa-miR-193a-5p; 5350; 5592; 5592; 5894; 10488; 5592; 1385; 5592; 1388; 5592; 148327; 5592; 2969; 5592; 468; 5592; 476; 5592; 477; 5592; 478; 5592; 480; 5592; 481; 5592; 482; 5592; 483; 5592; 486; 5592; 490; 5592; 491; 5592; 492; 5592; 493; 5592; 5592; 572; 5592; 64764; 5592; 6546; 5592; 7408; 5592; 9569; 5592; 9586
Adipocytokine signaling pathway	45,44	10645; 32; 5465; 6774; 6794; 9021; 10645; 32; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 6794; 181; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 4852; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 5105; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 5106; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6513; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6517; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 92579
cGMP-PKG signaling pathway	31,38	10672; 2768; 5592; 6093; 9475; 4659; 5592; 6093; 9475; hsa-miR-148b-3p; 4659; 5499; 5500; 5501; 5592; 6093; 9475; 5592; 90993; hsa-miR-127-3p; 10242; 27094; 27345; 3778; 3779; 5592; 776; 10242; 27094; 27345; 3778; 3779; 5592; 775; 10242; 27094; 27345; 3778; 3779; 5592; 778; 10242; 27094; 27345; 3778; 3779; 5592; 779; 5581; 5592; hsa-miR-103a-3p; hsa-miR-107; 5592; 84699; hsa-miR-193a-5p; 5350; 5592; 489; 5350; 487; 5350; 488; 5350
Adipocytokine signaling pathway	25,31	10645; 32; 5465; 6774; 6794; 9021; 126129; 1374; 1375; 32; 5465; 6774; 126129; 1374; 1375; 2180; 126129; 1374; 1375; 2181; 126129; 1374; 1375; 2182; 126129; 1374; 1375; 23205; 126129; 1374; 1375; 23305; 126129; 1374; 1375; 51703; 126129; 1374; 1375; 81616
cGMP-PKG signaling pathway	22,42	10672; 2768; 5592; 6093; 9475; 4659; 5592; 6093; 9475; hsa-miR-148b-3p; 4659; 5499; 5500; 5501; 5592; 6093; 9475; 5592; 90993; hsa-miR-127-3p; 10242; 27094; 27345; 3778; 3779; 5592; 776; 10242; 27094; 27345; 3778; 3779; 5592; 775; 10242; 27094; 27345; 3778; 3779; 5592; 778; 10242; 27094; 27345; 3778; 3779; 5592; 779; 5581; 5592; hsa-miR-103a-3p; hsa-miR-107; 5592; 84699; hsa-miR-193a-5p; 5350; 5592; 5592; 5894
cGMP-PKG signaling pathway	14,24	10000; 207; 208; 2770; 2771; 2773; 4846; 5592; 10000; 146850; 207; 208; 23533; 2770; 2771; 2773; 5294; 146850; 23533; 2770; 2771; 2773; 3667; 5294; 8660; 23533; 2770; 2771; 2773; 3667; 8660; hsa-miR-5001-5p; 134; 140; 153; 154; 155; 2770; 2771; 2773; 4985; 624
Shigellosis	10,99	7414; 7454; 8976; hsa-miR-664a-5p; 10092; 10093; 10094; 10095; 10109; 10163; 10552; 7454; 81873; 8936; 8976; 10092; 10093; 10094; 10095; 10109; 10552; 7454; 81873; hsa-miR-146a-5p

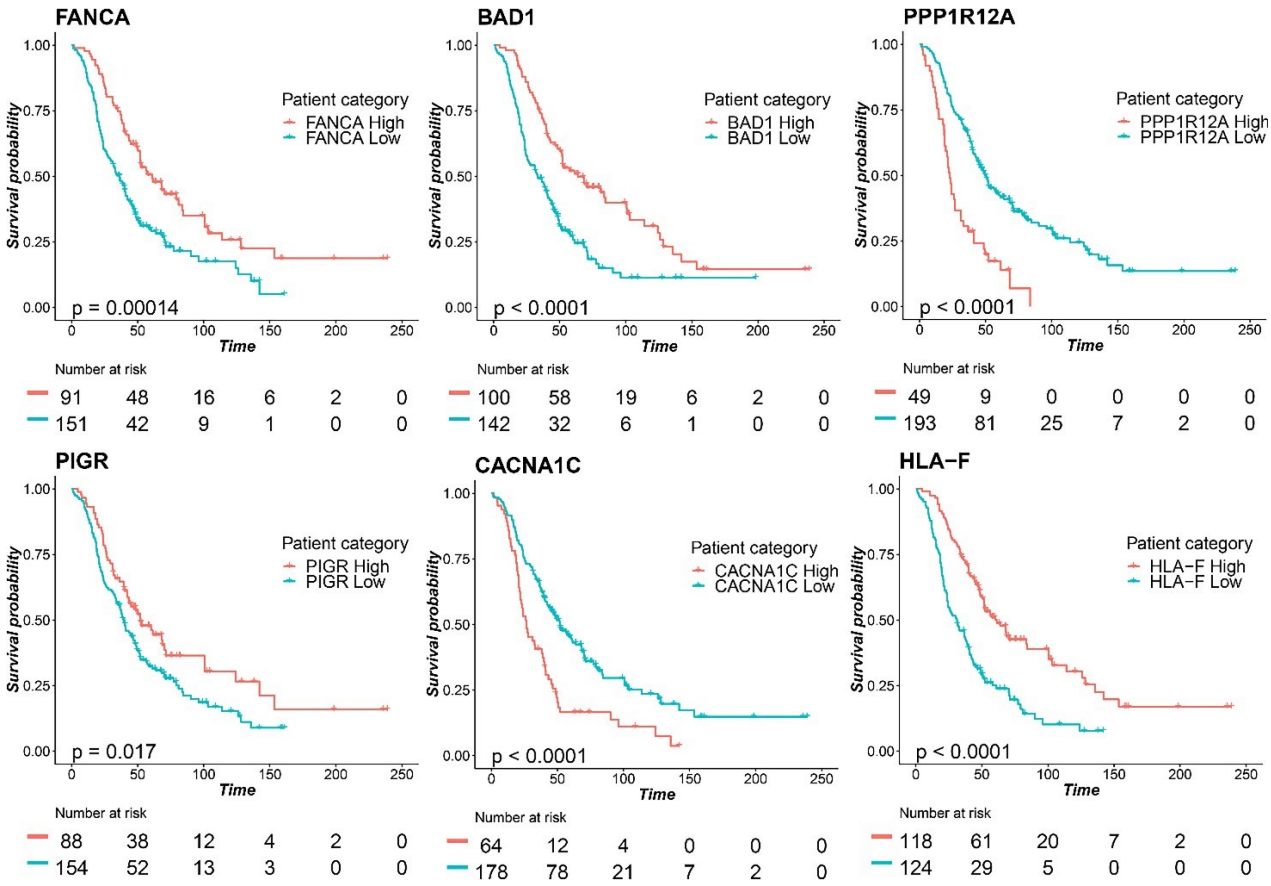
Supplementary Table S2. *Micrographite results. Pathway: Metapathway obtained. Genes and miRNAs of the path: genes (Entrez ID) and micro RNAs (miRbase ID) involved in selected meta-pathways. Max Score: details available in micrographite documentation* [64].

Pathway	Max Score	Genes and miRNAs of the path
Meta-pathway	156,67	4659; 5592; 6093; 9475; hsa-miR-148b-3p; 5499; 5592; 6093; 9475; 5500; 5592; 6093; 9475; 5501; 5592; 6093; 9475; 10242; 27094; 27345; 3778; 3779; 5592; 775; 10242; 27094; 27345; 3778; 3779; 5592; 776; 10242; 27094; 27345; 3778; 3779; 5592; 778; 10242; 27094; 27345; 3778; 3779; 5592; 779; 5581; 5592; hsa-miR-103a-3p; hsa-miR-107; 5592; 84699; hsa-miR-193a-5p; 5592; 90993; hsa-miR-127-3p; 5350; 5592; 10488; 5592; 1385; 5592; 1388; 5592; 148327; 5592; 2969; 5592; 468; 5592; 476; 5592; 477; 5592; 478; 5592; 480; 5592; 481; 5592; 482; 5592; 483; 5592; 486; 5592; 490; 5592; 491; 5592; 492; 5592; 493; 5592; 5592; 572; 5592; 5894; 5592; 64764; 5592; 6546; 5592; 7408; 5592; 9569; 5592; 9586
Meta-pathway	55,31	4659; 5592; 6093; 9475; hsa-miR-148b-3p; 5499; 5592; 6093; 9475; 5500; 5592; 6093; 9475; 5501; 5592; 6093; 9475; 10242; 27094; 27345; 3778; 3779; 5592; 775; 10242; 27094; 27345; 3778; 3779; 5592; 776; 10242; 27094; 27345; 3778; 3779; 5592; 778; 10242; 27094; 27345; 3778; 3779; 5592; 779; 5581; 5592; hsa-miR-103a-3p; hsa-miR-107; 5592; 84699; hsa-miR-193a-5p; 5592; 90993; hsa-miR-127-3p; 5350; 5592; 487; 5350; 488; 5350; 489; 5350
Meta-pathway	12,11	10092; 10093; 10094; 10095; 10109; 10552; 7454; 81873; hsa-miR-146a-5p; 10092; 10093; 10094; 10095; 10109; 10552; 7454; 81873; 8976; 7414; 7454; 8976; hsa-miR-664a-5p
Meta-pathway	10,63	3932; 7535; 915; 916; 917; 919; 3932; 920
Meta-pathway	9,93	3932; 7535; 915; 916; 917; 919; 27040; 7535; 27040; 5335
Meta-pathway	53,87	10645; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 6794; 32; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 181; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 4852; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 5105; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 5106; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6513; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6517; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 92579
Meta-pathway	19,59	10645; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 6794; 32; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 126129; 1374; 1375; 32; 5465; 6774; 126129; 1374; 1375; 2180; 126129; 1374; 1375; 2181; 126129; 1374; 1375; 2182; 126129; 1374; 1375; 23205
Meta-pathway	12,07	10645; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 6794; 32; 51422; 53632; 5562; 5563; 5564; 5565; 5571; 6774; 126129; 1374; 1375; 32; 5465; 6774
Meta-pathway	19,66	23533; 2770; 2771; 2773; 3667; 8660; hsa-miR-5001-5p; 146850; 23533; 2770; 2771; 2773; 3667; 5294; 8660; 10000; 146850; 207; 208; 23533; 2770; 2771; 2773; 5294; 10000; 207; 208; 2770; 2771; 2773; 4846; 134; 140; 153; 154; 155; 2770; 2771; 2773; 4985; 624
Meta-pathway	13,32	2209; 6688; 6688; 929; 3684; 6688

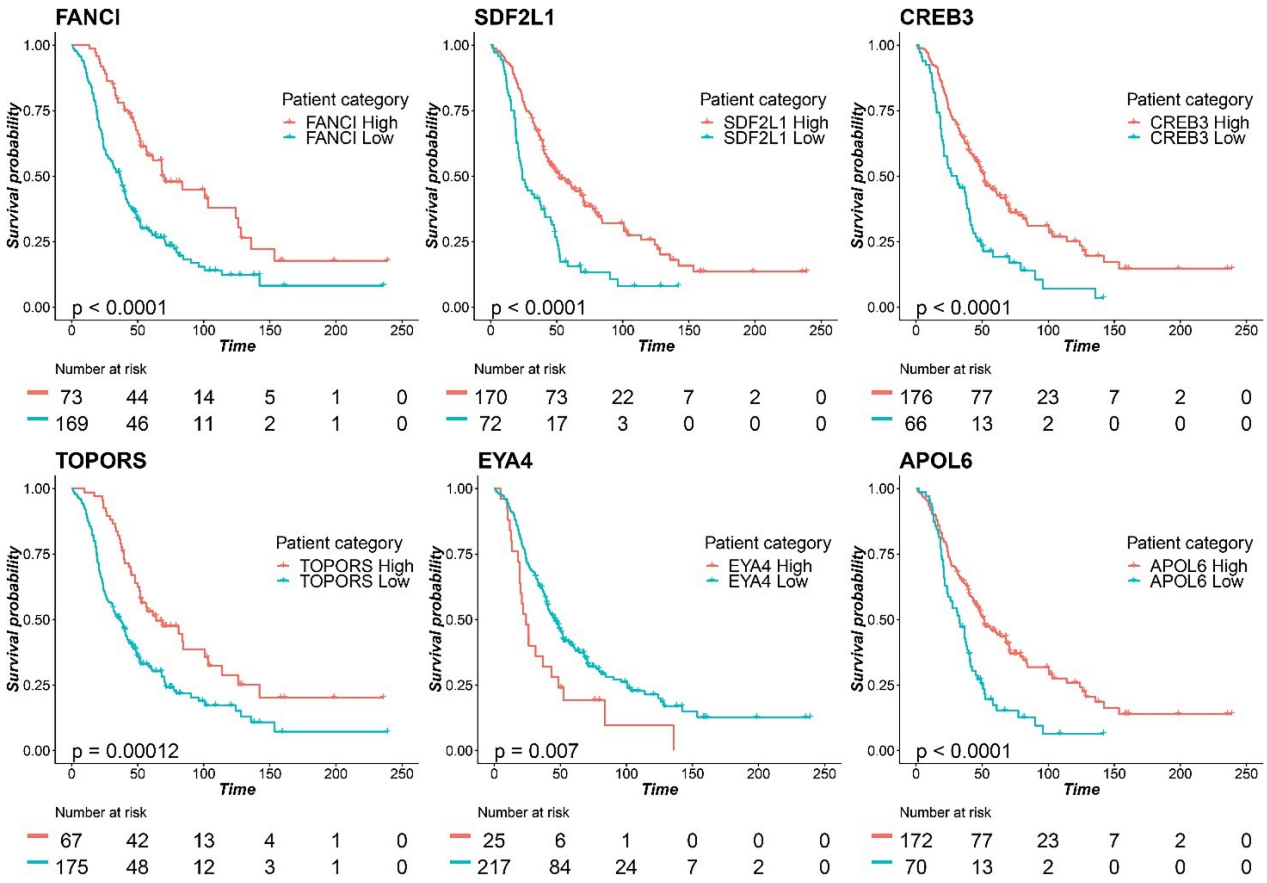
Supplementary Table 3. *Primer sequences used for qRT-PCR validation. For each gene symbol, gene bank ID, forward and reverse sequences, annealing temperature and amplicon length is reported.*

Gene	Accession Bank	FW primer sequence	RV primer sequence	TA (°C)	Amplicon Lenght
GAPDH	NM_002046	AGGTCGGAGTCAACG GATTT	TTAAAAGCAGCCCT GGTGAC	60	58
HPRT1	NM_000194	TGAATACTTCAGGGAT TTGAATCAT	CTCATCTTAGGCTTT GTATTTTGC	60	76
PPIA	NM_021130	GCGTCTCCTTTGAGCT GTTT	CCTTTCTCTCCAGTG CTCAGA	60	79
NOTCH2	NM_024408	GAAGGCAGGTCTCCT GTGTC	ATCTTCTGTGCAGTC AGCCC	60	145
HLA-A	NM_001242758.1	ATGAAGGCCCACTCA CAGAC	GTGAGAACCGTCCTC GCTC	60	81
TOPORS	NM_001195622.1	GAAGAAATAGGGCCT TTCCG	TGCCATTATCATGAA GCCAGT	60	128
EYA4	NM_004100	GTAAACCAGTTTGAA AAATGTTCTGT	AATAGCCGAAAACC CACTTT	60	50
NPY	NM_004535	GGGGATTTTCCCTTG	AAAACCAAAATGTC TTTCTCTCCA	60	50
FANCI	NM_018193	TACGGGTAACGGAAG TGTGG	TCACAGAACTCCGCC ACAAA	60	70
FANCA	NM_000135	GAAGAGGCCTTCCTGC ATGT	GGTTGCCCTGACCCT TGAG	60	127
PPP1CA	NM_002708	GACCGTGGCGTCTCTT TTAC	TCTTCTACCACCTGG TGTGCT	60	101
PIGR	NM_002644	AGAGGCAGGGGTTAC CAACT	TCCTGTGCAATGTTT TAGCCAC	60	89
HLA-F	NM_018950	GCTGCAGTGTGAGAC AGCTT	TGTATGTTCGTGAGG CACAA	60	87
CACNA1C	NM_000719	CGTGGCTGCTCCTCCT ATTA	CATAGTTGGAACCTT GGTGGTT	60	97
CREB3	NM_006368	CTTTCTGAGGTACCGA GCGA	GAGAATGTTCAACG ACGCTG	60	81
STAT1	NM_007315	AACCTCGACAGTCTTG GCAC	GAGACATCCTGCCA CCTTGT	60	96

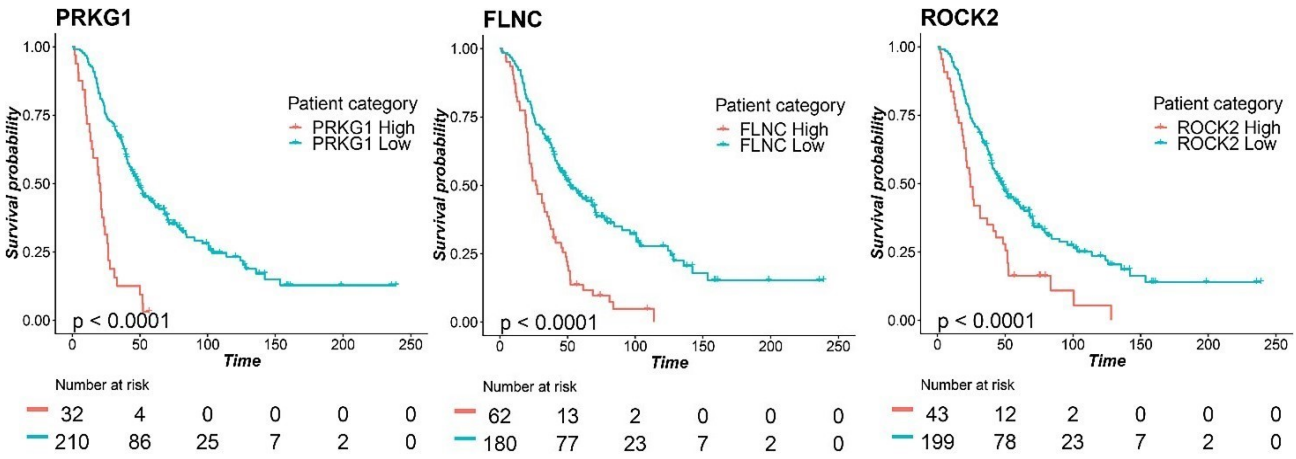
Supplementary Figure S1. *Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using OS. At the bottom of each curve the patients risk table is reported.*



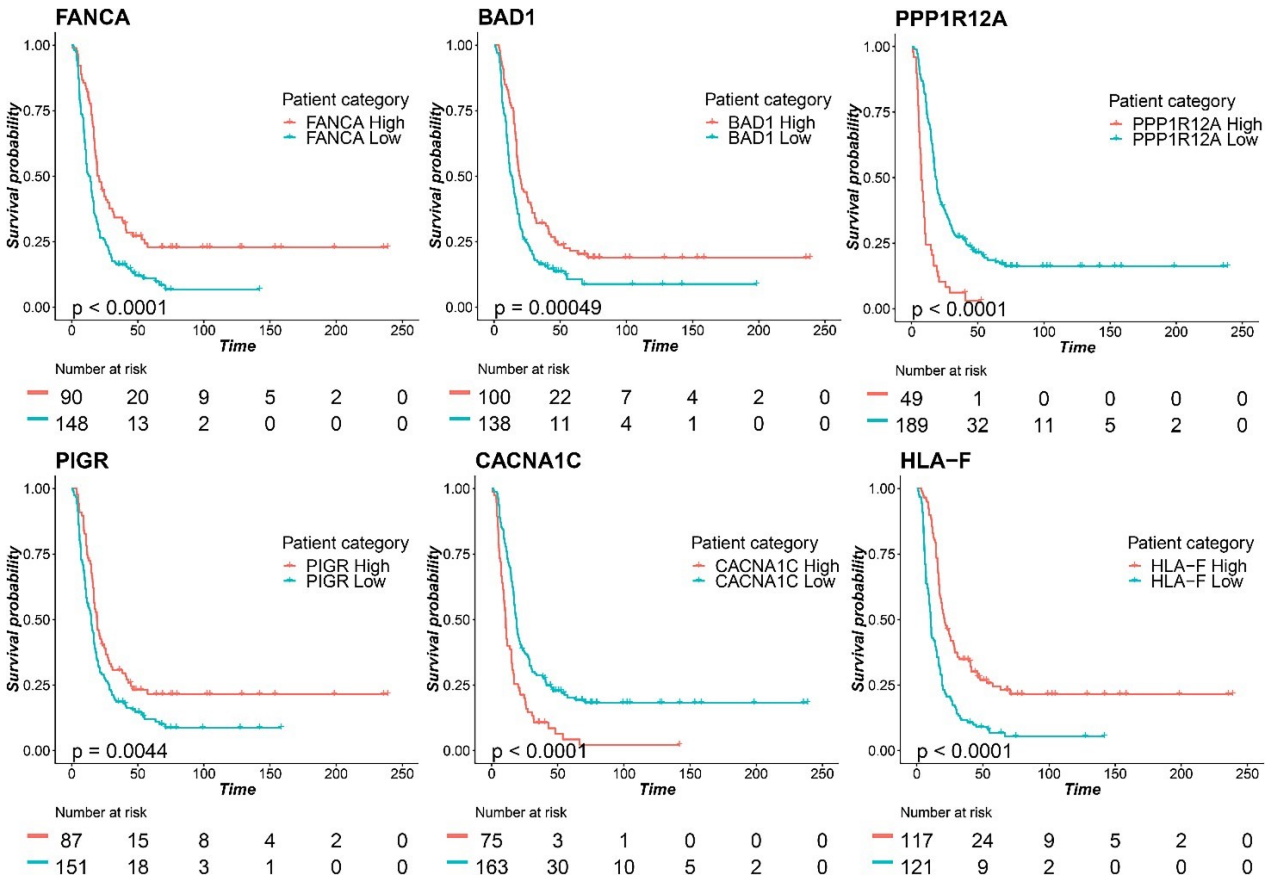
***Supplementary Figure S1 (continued)**. Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using OS. At the bottom of each curve the patients risk table is reported.*



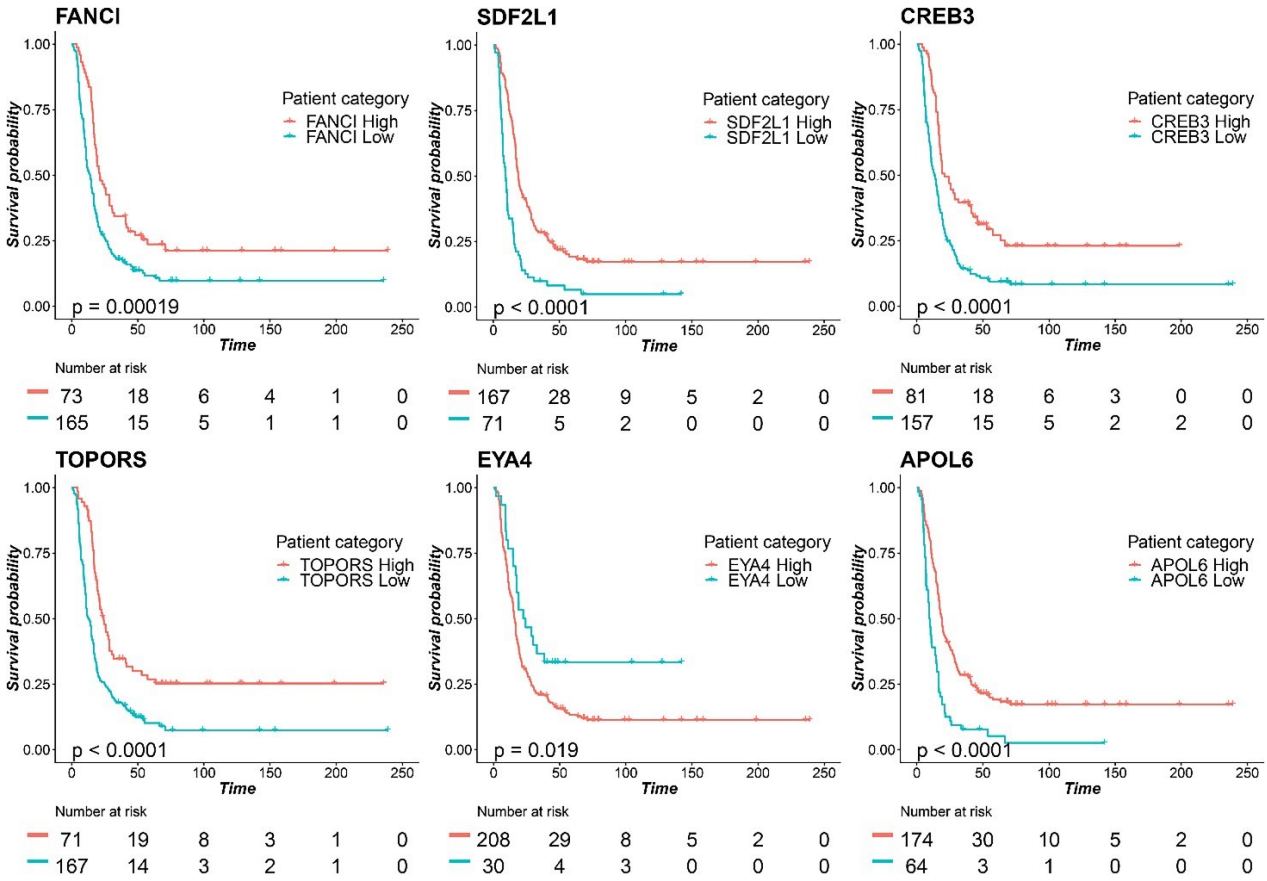
***Supplementary Figure S1 (continued)**. Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using OS. At the bottom of each curve the patients risk table is reported.*



Supplementary Figure S2. *Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using PFS. At the bottom of each curve the patients risk table is reported.*



***Supplementary Figure S2 (continued)**. Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using PFS. At the bottom of each curve the patients risk table is reported.*



***Supplementary Figure S2 (continued)**. Kaplan Mayer curves and log-rank test p-values of the selected 16 genes of the network using PFS. At the bottom of each curve the patients risk table is reported.*