Title:

A complex network of tumor microenvironment in human high grade serous ovarian cancer

Authors and affiliations:

Caroline Kreuzinger[1], Angelika Geroldinger[2], Dominiek Smeets[3,4], Elena Ioana Braicu[5,6], Jalid Sehouli[5,6], Julia Koller[1], Andrea Wolf[1], Silvia Darb-Esfahani[5,7], Korinna Joehrens[7], Ignace Vergote[8], Adriaan Vanderstichele[8], Bram Boeckx[3,4], Diether Lambrechts[3,4], Hani Gabra[9,10], G. Bea A. Wisman[11], Fabian Trillsch[12,13], Georg Heinze[2], Reinhard Horvat[14], Stephan Polterauer[15], Els Berns[16], Charles Theillet[17,18], Dan Cacsire Castillo-Tong[1*]

[1]Translational Gynecology Group, Department of Obstetrics and Gynecology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria;

[2]Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria;

[3]KU Leuven, Department of Human Genetics, Laboratory for Translational Genetics, 3000 Leuven, Belgium;

[4]VIB, VIB Center for Cancer Biology, Laboratory for Translational Genetics, 3000 Leuven, Belgium;

[5]Tumor Bank Ovarian Cancer Network, Department of Gynecology, Charité Universitätsmedizin Berlin, Germany;

[6]Department of Gynecology, Charité Universitätsmedizin Berlin, Germany;

[7]Institute of Pathology, Charité Universitätsmedizin Berlin, Germany;

[8]Departments of Gynecologic Oncology, Leuven Cancer Institute, University Hospitals Leuven, KU Leuven, Leuven, Belgium;

[9]Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, Imperial College London, London, United Kingdom;

[10]Clinical Discovery Unit, Early Clinical Development, AstraZeneca, Cambridge, UK;

[11]Department of Gynecologic Oncology, Cancer Research Center Groningen, University of Groningen, University Medical Center Groningen, The Netherlands;

[12]Department of Gynecology and Obstetrics, University of Munich, Munich, Germany;

[13]Department of Gynecology and Gynecologic Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;

[14]Department of Clinical Pathology, Medical University of Vienna, Vienna, Austria;

[15]Department of Gynecology and Gynecologic Oncology, Gynecologic Cancer Unit, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria;

[16]Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands;

[17]Institut de Recherche en Cancérologie de Montpellier, Université Montpellier, Montpellier, France;

[18]INSERM U1194, Montpellier, France;

Running title: tumor microenvironment in high grade serous ovarian cancer

Key words: matched primary and recurrent tumors, high grade serous ovarian cancer, molecular profiling, network of microenvironment, immune activation

Additional information:

- Financial support: This work was supported by European Community's Seventh Framework Programme under grant agreement No. 279113-2 (OCTIPS; www.octips.eu).
- Corresponding author: Dan Cacsire Castillo-Tong, Ebene 5Q, AKH; Waehringer Guertel 18-20, Translational Gynecology Group, Department of Obstetrics and Gynecology, Medical University of Vienna, 1090 Vienna, Austria
  Tel: 0043-1-4040078330
  Email: dan.cacsire-castillo@meduniwien.ac.at

- Conflict of interest statement:

  The authors have declared that no conflict of interest exists.

Statement of translational relevance (146 words)

By comparing gene expression profiles of matched primary and recurrent fresh frozen tissue samples from the same patients with HGSOC, we discovered that the predominant gene expression difference is their tumor microenvironment, presented by a panel of genes covering all major pathways of immune activation together with a number of genes involved in the remodeling of extracellular matrix and adipose tissues. The findings have important translational relevance, especially on the treatment of recurrent ovarian cancer. Not only the molecular nearness of recurrent HGSOC to their primary counterparts can guide the treatment by recurrence, but also the whole tumor microenvironment network can be taken into consideration by developing immune therapies or therapies targeting the tumor microenvironment. Finally, the immune suppression mechanisms in HGSOC could be different as in other types of tumors. The functions of the B7-CD28 molecules should be carefully investigated before therapies can be developed.

Abstract (250 words)

Purpose: Most high grade serous ovarian cancer (HGSOC) patients develop recurrent disease after first line treatment, frequently with fatal outcome. This work aims at studying the molecular biology of both primary and recurrent HGSOC.

Experimental design: Gene expression profiles of matched primary and recurrent fresh frozen tumor tissues from 66 HGSOC patients were obtained by RNA sequencing. Clustering analyses and pairwise comparison of the profiles between matched samples and subsequent functional alignment were used for the identification of molecular characteristics of HGSOC.

Results: Both primary and recurrent HGSOC samples presented predominant gene expression differences in their microenvironment, determined by a panel of genes covering all major pathways of immune activation together with a number of genes involved in the remodeling of extracellular matrix and adipose tissues. Stratifying tumor tissues into immune active and silent groups, we further discovered that while some recurrent tumors shared the same immune status as their primary counterparts, others switched the immune status, either from silent to active or active to silent. Interestingly, genes belonging to the B7-CD28 immune checkpoint family, known for their major role as negative regulators of the immune response, were overexpressed in the immune active tumors. Searching for potential tumor antigens, *CEACAM21*, a member of the carcinoembryonic antigen family, was found to be significantly overexpressed in immune active tissues in comparison to the silent ones.

Conclusion: The results illustrate the complexity of the tumor microenvironment in HGSOC and reveal the molecular relationship between primary and recurrent tumors, which have multiple therapeutic implications.

Introduction

High grade serous ovarian cancers (HGSOC) account for about 75% of all ovarian carcinomas (1). Standard treatment includes debulking surgery and platinum-based chemotherapy. Although most patients show primary response, they frequently relapse and eventually develop fatal resistant disease. The five year survival rate of HGSOC is between 35-40% (2). Cancer recurrence and therapy resistance are considered the major causes of death.

Recent next generation sequencing (NGS) data on primary ovarian cancer demonstrate that nearly all HGSOCs harbor a mutation in the *TP53* gene (2). Other major molecular alterations are found in DNA damage and repair pathways, in which *BRCA1* and *BRCA2* genes are frequently mutated (3), and in the RB1/CCNE1 pathway (4). In addition, HGSOC is characterized with a high genetic instability and frequent promoter hypermethylation. Primary HGSOC is considered to be heterogeneous, comprising different tumor clones with distinct genetic characteristics (4-8).

Beside specific molecular characteristics of tumor cells, tumor microenvironment also plays an important role in the progression of HGSOC. Various molecules, pathways, or specific cell types involved in the immune activation, extracellular matrix remodeling and other stromal alterations were reported to affect the clinical outcome of the patients (9-15). Molecular subclassification of the primary HGSOC also suggested different microenvironments around the tumors (16).

We hypothesize that gene expression changes of both tumors and their microenvironment by recurrence might be determinants for uncontrolled proliferation and resistance of tumors to therapies. By defining the molecular characteristics of recurrent tumors, new therapeutic strategies might be developed against cancer progression. So far, recurrent HGSOC tumor samples have not been systematically collected and analyzed in depth (1) and hence remain poorly understood. In order to investigate the molecular characteristics of recurrent HGSOC and their microenvironment in comparison with their primary counterparts, we have collected

matched primary and recurrent fresh frozen tumor tissues from the same patients and analyzed expression profiles by RNA sequencing within the frame of an international project OCTIPS (Ovarian Cancer Therapy – Innovative Models Prolong Survival), funded by the European Commission's 7th Framework Programme.

Materials and methods

- Patients and clinical materials

66 patients with high grade serous epithelial ovarian cancer were included at the Department of Gynecology, European Competence Center for Ovarian Cancer; Campus Virchow Klinikum, Charité - Universitätsmedizin Berlin, Berlin, Germany; Division of Gynecological Oncology, Department of Obstetrics and Gynecology, Universitaire Ziekenhuizen Leuven, Katholieke Universiteit Leuven, Leuven, Belgium; Ovarian Cancer Action Research Centre and Imperial College, London, UK; Department of Gynecology and Gynecologic Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Department of Obstetrics and Gynecology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria; and Department of Gynecologic Oncology, Cancer Research Center Groningen, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. Informed consents were obtained from all patients prior to sample collection. All processes were approved by the local ethical committee (EK207/2003, ML2524, 05/Q0406/178, EK130113, EK366/2003, EK260/2003).

This retrospective study population consisted of matched primary and recurrent fresh frozen tumors from HGSOC patients diagnosed between 1993 and 2013. All patients were operable and most of them were platinum sensitive.

- RNA isolation and quality control

RNA extraction was performed using the “OCTIPS (#279113) Standard Sample Processing Protocol V2”, section “RNA extraction from tumor tissue” (17).

Quality control of RNA was performed using Agilent RNA 6000 Nano Kit (Agilent Technologies, CA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions. RNAs with an RNA Integrity Number >4 were further processed.

- RNA sequencing and annotation

RNA libraries were created using the Illumina TruSeq RNA sample preparation kit V2 according to the manufacturer's instructions and sequenced on a HiSeq2000 (Illumina, California, USA) using a V3 flowcell generating 1 x 50 bp reads. Raw sequencing reads were mapped to the transcriptome and the human reference genome (NCBI37/hg19) using TopHat 2.0 (18) and Bowtie 2.0 (19).

- Data processing and pre-filtering of genes

Genes with a mean read count ≤ 0.5 or with ten or fewer observations different from zero, across all 132 samples, were excluded from the analysis, resulting in 28,235 genes eligible for analysis. Data was normalized to equal sequencing depth using the sequencing depth estimation method implemented in the "samr" package (20). For numerical reasons, 0.5 was added to all normalized expression read counts.

- Unsupervised clustering analyses

The recurrent samples were clustered by Euclidean distance and the complete linkage clustering method. 369 genes with interquartile ratios (ratio of third and first quartiles) greater than or equal to ten and a third quartile read counts greater than or equal to 50 were considered for clustering. Clustering of primary samples was performed in the same way, taking 338 genes that passed the same criteria into account. In order to avoid disproportional

influence of high gene expression values in the clustering, the normalized data were logarithmized. For graphical displays in heatmaps, values were truncated at the 99.9th percentile in order to increase the color intensity.

- Fold change (FC) and the "area under the receiver operating characteristic curve (AUROC)" for the comparison of gene expressions between two clusters

The FC with respect to two clusters of samples was calculated as the ratio of the group-wise medians of normalized read counts. AUROCs were computed to assess the concordance of gene expression between pairs of recurrent and primary samples, or to assess separation of the gene expression distributions of two groups of samples. For comparability across genes, AUROCs lower than 0.5 were flipped at 0.5 such that 0.5 and 1 were the lowest and highest achievable values, indicating no overlap (no concordance) or complete separation (full concordance) of gene expressions between the groups, respectively.

- Functional annotation of differentially expressed genes

Genes, which had the highest expression difference in the two clusters identified by unsupervised clustering (FC>5 or <0.2, AUROC>0.75 and q3>50 read counts), were first assigned to their specific cell expression according to Angelova et al (21) and examined for their specific functions in databases such as Wikipedia (https://en.wikipedia.org), WikiGenes (https://www.wikigenes.org/app/WikiGene), GeneCards (http://www.genecards.org), Ensembl (22) (www.ensembl.org), NCBI Resources (www.nibi.nlm.hih.gov), and KEGG PATHWAY Database (www.genome.jp/kegg/pathway.html).

- Supervised clustering analyses

We used the 126 immune related genes found by unsupervised clustering and the subsequent functional analysis to define the immune status of primary and recurrent samples by clustering similarly as described above. This clustering is referred as “supervised” clustering.

- Heat map presentation of expression of selected genes with known functions

In order to illustrate the within-gene differences in the expression of selected genes, logarithmized normalized read counts were centered and scaled to unit variance, separately performed for the primary and the recurrent samples. The transformed values were truncated at two standard deviations to increase the color intensity of the heatmaps. Samples were grouped by immune-status and were ordered alphabetically within these groups. In particular, no clustering was performed on either samples or genes. Because of its equivalence to the AUROC, the Mann-Whitney U test was used to compare gene expressions between active and silent samples. P values were corrected for multiple testing within the two groups of genes using the Bonferroni-Holm method (23). Adjusted $p<0.05$ was considered to indicate statistical significance.

- Pairwise fold change and selection of differentially expressed genes in paired samples

For each gene and each pair of samples, the pairwise FC was calculated as the ratio of normalized read counts. Genes were considered to be differentially expressed in a pair of samples, if the FC>5 or <0.2 and at least one of the two samples had read count>50. In this way, we generated two gene lists for each pair of samples. One included genes overexpressed in primary samples and the other included genes overexpressed in recurrent samples. For each subgroup, overexpressed genes presented in more than 50% of the samples were extracted and presented for primary and recurrent samples.

- Comparison of clinical outcome of patients with different immune status

Cumulative probabilities of survival were estimated using the product-limit method, censoring for end of follow-up. Crude and adjusted hazard ratios with 95% confidence intervals were estimated by Cox regression. For the OCTIPS study cohort, operable recurrent tumor was the inclusion criterion and therefore, cumulative survival probabilities could not be unbiasedly estimated if the date of primary diagnosis served as baseline (24). Therefore, we compared the mortality hazard after cancer diagnosis between patients with active and silent tumors using a delayed-entry (entry at recurrence) Cox regression model (25). Time from primary diagnosis to recurrence was compared between patients with active and silent primary tumors using the Mann-Whitney U test. Two-sided p-values <0.05 or 95% confidence intervals excluding parity were considered as statistically significant.

- Analyses of TCGA gene expression data

488 patients with ovarian cancer were included in the TCGA data. We included all HGSOC and excluded cases missing information on residual tumor and on survival, obtaining a total of 421 patients. TCGA_489_UE.txt was used to access the gene expression data and TCGA Table S1 for the clinical information. We aligned the 126 genes to the TCGA annotation and could find 59 overlapping genes. Supervised clustering was performed using these 59 genes and in the same way as described above to define the immune active and silent cluster. Overall and relapse-free survival were compared between the two groups.

- Analyses of RNA sequencing data of Patch et al.(4)

In a recent study (4), RNA sequencing data from 80 primary HGSOC using TruSeq RNA Sample Preparation v2 kit and HiSeq2000 instrument (Illumina) were provided. RNAseq data were downloaded from the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/), which is hosted at the EBI, under accession number EGAS00001000397: RNAseq data (exp_seq.OV-AU.tsv). The corresponding clinical data were filed donor.OV-AU.tsv and specimen.OV-AU.tsv. The Access to datasets was approved by the specified Data Access

Committee (DAC). We defined the immune activation status of these 80 primary tumors by supervised clustering using the 126 immune relevant genes in the same way as for our primary and recurrent samples. Overall and relapse-free survival were compared between the two groups similarly as described above. The information on residual tumor after debulking surgery was not available.

- Data Availability

Transcriptome sequencing data will be deposited in the European Genome-phenome Archive (EGA) (https://www.ebi.ac.uk/ega/home).

- Immunohistochemistry (IHC)

Formalin fixed paraffin embedded (FFPE) tissues were sectioned at 3 µm. The IHC staining of EpCAM, CD45, CD8, and CD20 were performed with the Dako LSAB+ System-HRP kit (Dako, CA, USA) including the second antibodies following the manufacturer's instructions. Primary antibodies were diluted with Dako REAL Antibody diluent (Agilent Technologies, St. Clara, California) and incubated overnight at 4°C. FLEX Negative Control Mouse Cocktail (Agilent Technologies) and Negative Control Rabbit IgG (Biocare Medical, Concord, USA) were used as isotype controls. For the staining of the immunoglobulins, DAB- substrate was applied directly after primary antibody incubation. Nuclei were stained with Hematoxylin solution modified acc. to Gill III (Merck Millipore Darmstadt, Germany) before mounting the slide with Kaisers Glyceringelatine (Merck Millipore).

PD-L1 was stained with Leica BOND staining device (Leica Biosystems, Nussloch, Germany), PD-1 and CD1a with Ventana Benchmark staining device (Roche Diagnostics, Basel, Switzerland), both following the company’s instructions.

Antibodies and dilutions: anti-EpCAM (IgG, rabbit, clone E144; abcam, Cambridge, UK), 1:300; anti-CD45 (IgG, rabbit, clone E19-G; DB Biotech, Kosice, Slovakia), 1:800; anti-CD8

(IgG1, mouse, clone 144B; abcam), 1:25; anti-CD20 (IgG2a, mouse, clone L26; Thermo Fisher Scientific, CA, USA), 1:200; anti-CD1a (IgG1, mouse, clone 010; Dako), 1:50; anti-PD-1 ready to use (IgG1, mouse, clone NAT105; Ventana Medical Systems, Inc., AZ, USA); anti-PD-L1 (rabbit, clone E1L3N; Cell signaling, Beverly, MA, USA), 1:200; anti-Ig (Goat anti-Human IgG, IgM, IgA (H+L), HRP-conjugated, Thermo Fisher Scientific), 1:100.

Results

- Study cohort

The study cohort consisted of 66 patients (aged from 21-74 years, median 56) with recurrent HGSOC. Matched fresh frozen tumor samples from both primary and recurrent tumors were collected. All patients underwent cytoreductive surgery at first diagnosis as well as at recurrence and received platinum-based first-line chemotherapy. Median interval between primary operation and diagnosis of recurrence was 20 (interquartile range: 14; 28) months, which is representative of a normal HGSOC population (26). Patients with previous malignant disease were only included if they had at least five years disease free interval at the first diagnosis of HGSOC (Supplementary Table S1).

- Identification of immune active and immune silent HGSOC samples

Since recurrent HGSOCs present very different clinical manifestations that most possibly reflect different molecular subtypes, we first examined whether there would be strong expression differences within the 66 recurrent tumors. Unsupervised clustering analysis of gene expression profiles revealed the existence of two distinct clusters of tumor tissues C1 and C2, containing 36 and 30 samples, respectively (Figure 1A). These two clusters were virtually unchanged, when we performed the analyses with varied numbers of genes that were subjected to clustering, indicating the stability of the clusters.

In order to define the major biological differences between these two clusters, we assigned the 142 most differentially expressed to their biological functions. While only 2 genes (*PCP4 and PHOX2A*) were overexpressed in cluster C2, the remaining 140 had overexpression in C1 (Supplementary Table S2). Of these 140 genes (Figure 1B), 126 (89%) were directly related to immune response, corresponding to genes specifically expressed in B (8%) or T

cells (23%), monocytes (1%) and genes coding for proteins involved in immunoglobulin construction (42%) or for proteins with other direct functions in the immune activation (15%). These results clearly indicated that recurrent HGSOC samples presented significant gene expression differences of tumor microenvironment.

In the same way, unsupervised clustering of the 66 primary tumors generated two clusters, which also presented outweighing expression differences in immune related genes (Supplementary Table S3). Of the 113 most differentially expressed genes, 90 were specifically expressed in various types of B and T cells, or encoded immunoglobulins and proteins directly related to immune activation.

Our data thus showed that HGSOC could be stratified into two subgroups according to the expression of immune specific genes. Hence, we re-clustered the recurrent and primary tumors separately using the 126 immune related genes identified in the recurrent samples. By doing so, primary and recurrent tumors were classified as immune-active (higher expression of the immune genes) and immune-silent (lower expression) (Supplementary Figure S1A, S1B).

We therefore defined 4 subgroups of matched samples according to the immune status (Supplementary Table S4) and named them active-active (16 pairs), silent-active (21 pairs), active-silent (9 pairs), and silent-silent (20 pairs). Supplementary Figure S1C shows examples of expression difference of two immune related genes (*IGKV4-1* and *CD79A*) in these subgroups.

- Matched primary and recurrent samples presented various immune activation status

In order to obtain an overview of important immune related molecular processes in paired samples, we compared the expression of selected genes covering T- and B cell response,

natural immunity, as well as immune checkpoint modulation, irrespectively if they were included in the 126 genes. As shown in Figure 2, not only the genes specific for the immune cells (*CD8, CD4, FOXP3, CD79A, CD20*, all *p<0.001*; and *CD1a, p=0.08* and *p=0.06*, in primary and recurrent samples, respectively), but also some genes encoding their effectors such as interferon gamma (*IFNG, p<0.01*), granzyme B (*GZMB, p<0.01*) and perforin (*PRF1, p<0.01*) were significantly overexpressed in the immune active tissues, thus strengthening the notion of an overall activation of the immune response. Interestingly, lymphotoxin alpha (*LTA*) and lymphotoxin beta (*LTB*) presented significant overexpression in immune active recurrent tumor samples (*p<0.01*) but did not differ in their expression in primary tumor samples significantly. Of note, macrophage migration inhibitory factor (*MIF*), which is considered to regulate innate immunity, did not show significant difference in expression between immune active and silent samples.

Remarkably, the major ligands and receptors (*PD-L1, PD-L2* with their receptor *PD-1*; *CD80, CD86* with their receptors *CTLA4* and *CD28*) of the CD28/B7 costimulatory pathway, known to negatively regulate the immune checkpoint, were all significantly overexpressed in the immune active samples (all *p<0.01*). Interestingly, while the ligand *ICOSLG* did not show any difference in active versus silent, its receptor *ICOS* showed significantly overexpression in active samples (*p<0.001*). *CD276 (B7-H3)* and *VTCN1 (B7-H4)* were not differentially expressed in immune active versus silent samples (Figure 2).

Further, we performed immunohistochemistry staining of CD45, CD8, CD20, CD1A, PD-L1, PD-1 and immunoglobulins together with EpCAM in 6 active and 6 silent tissue samples. Figure 3 showed examples of the IHC staining results in immune active tissues. Abundant CD8+ T cells as well as CD20+ B cells were found in immune active tumor tissues (Figure 3B, 3C, 3F, 3G). This was accompanied by a strong immunoglobulin staining (Figure 3D, 3H), which was colocalized with the CD20+ cells. The antigen presenting CD1a protein was expressed in dendritic cells, most of which were found interspersed within tumor cells (Figure

3I). PD-L1 staining was mainly found in tumor cells (Figure 3J) whereas the receptor PD-1 was clearly expressed in lymphocytes (Figure 3K). This staining pattern was found in other 4 active tissues but in none of the 6 silent tissues (Supplementary Figure S2). The results were in concordance with those generated from RNA sequencing, confirming that the strong expression of the immune genes in HGSOC tumor tissues was accompanied by strong lymphocytic infiltrations in tumors.

- Matched primary and recurrent HGSOCs presented varied patterns of microenvironment changes

The pairwise comparison of gene expression profiles of tumor pairs in each of the four subgroups presented diverse pattern of gene expression differences (Figure 4). In the active-active and the silent-silent subgroups, no differentially expressed genes could be identified, with the sole exception of *CCL19*, which was overexpressed in 9/16 recurrent samples of the active-active subgroup (Figure 4A and 4D). This indicated that primary and recurrent tumors showed little if any expression difference in these two subgroups. In the active-silent as well as in the silent-active subgroups, the active samples showed overexpression of immune related genes as expected and there were no notable overexpressed genes in the silent samples (Figure 4B and 4C). Furthermore, in the divergent subgroups, immune active samples presented elevated expression of a number of genes involved in remodeling of extracellular matrix (ECM). These did not only include genes responsible for the degradation of ECM, such as *ADAMTS2* and *FAP*, but also genes, such as *COL5A2*, *COL3A1*, or *HAS1* (hyaluronan synthase), which code for the components of ECM. In addition, numerous genes involved in adipose tissue remodeling were overexpressed in most of the immune active primary tumor samples (*FABP4, GPD1, PLIN*, *ADIPOQ*, *and TUSC5*) in the active-silent subgroup (Figure 4C). We examined these adipose related genes in the silent-active sample pairs to verify whether we may have missed them by using stringent selection criteria.

Indeed, a number of active recurrent samples showed overexpression of the genes too (e.g. 10/21 had FC>5 higher expression of *FABP4*).

We also compared expression of genes known to play important roles in epithelial ovarian cancer formation and progression, such as *TP53, CDKN1A (p21), BRCA1, BRCA2, PARP1, CCNE1, ERBB2* (*HER-2), PAX8, VEGFA,* and *MUC1* (1,3), but did not observe any notable expression difference between active and silent tumor tissues in both primary and recurrent samples (Figure 5A). In addition, we examined the expression of *ABCB1 (MDR1)*, which has been associated with acquired resistance of HGSOC (4). Interestingly, we observed a significant increase of *MDR1* expression in immune active tumors (Figure 5C; *p<0.001* in primary tissues; *p=0.04* in recurrent tissues).

To investigate what could be the trigger of the strong immune activation in HGSOC, we searched for potential tumor antigens which could be overexpressed in immune active samples. Alpha-fetoprotein (*AFP*), carcinoembryonic antigens *CEACAM1*, *CEACAM3* (ENSG00000170956), *CEACAM4* and *CEACAM19*, *CA-125*, trophoblast glycoprotein (*TPBG*), tyrosinase (*TYR*), and melanoma-associated antigens *MAGEA1*, *MAGEB1* and *MAGEB2* showed either no or very low expression levels (median≤3, the third quartile≤13 read counts) or did not present any expression difference in immune active versus silent samples (Figure 5B). Notably, *CEACAM21* (ENSG00000007129) showed an overexpression in the immune active samples (Figure 5D; primary samples: *p=0.056*; recurrent samples: *p<0.001*).

- Immune activation status and patient outcome

No difference could be found for the interval from primary diagnosis to recurrence between patients with active versus silent primary tumors (*p=0.517*), for mortality after recurrence between patients with active and silent recurrent tumors (*p=0.454*), and for mortality after

primary diagnosis between patients with active and silent primary samples (*p=0.089*). Adjusting the presence of residual tumors in the analyses of mortality gave similar, non-significant results.

When performing supervised clustering of the TCGA samples, we could align 59/126 genes and define a cluster of 45 samples as immune active and 376 samples as immune silent (Supplementary Figure S3). Patients with immune active status showed a statistically significant better disease free survival as the silent ones, whereas no difference in overall survival could be found in the two groups (Figure 6A and 6B).

In addition, we used the 126 immune related genes to perform the supervised clustering of the 80 primary HGSOCs in the cohort of Patch et al (4). Two clusters were defined consisting of 54 immune active and 26 silent samples (Supplementary Figure S4). Even though Kaplan-Meier curves showed a slightly better outcome of the active group over the silent one, there was no statistically significant difference in overall survival or relapse-free survival between the two groups of patients (Figure 6C and 6D).

Discussion

This study was designed to characterize recurrent HGSOC and their microenvironment in comparison to their primary counterparts at the transcriptomic level. To reach this aim, matched primary and recurrent HGSOC samples were collected from 66 patients that underwent debulking surgery at both primary diagnosis and upon recurrence. The unique design of the study allowed pairwise analyses based on a real “case-control” comparison, reducing the perturbation of individual gene expression difference to a minimum. The relatively large sample size for this type of collection further enabled the acquisition of new knowledge on HGSOC.

We first discovered that both primary and recurrent tumors, presented predominant molecular differences in their microenvironment, determined by a panel of genes covering all major pathways of immune activation accompanied by numerous genes involved in ECM and adipose tissue remodeling. By defining tumors as immune active and silent ones, we further discovered that while some paired primary and recurrent tumors shared the same immune status, being either both active or both silent, other pairs switched the status, either from silent to active or active to silent. Using the comparison analyses, we obtained results, which were based on expression difference of numerous functional related genes involved in a certain biological process and thus represented genuine and reliable molecular difference between individual samples. Our data strongly suggest that the appearance of recurrent tumors was not driven by large scale changes of gene expression regulation in tumor cells, but could rather be determined by the host immune reactions.

Immune status of the tumor samples was defined by 126 genes, which were identified by unsupervised clustering. The active samples had an overexpression of the major genes responsible for immune activation, some of which were immune cell specific, suggesting a lymphocyte infiltration. This substantiated that the overexpression of the 126 genes signed

for a strong immune reaction of the host against the tumor. Along with T and B cell specific markers, elevated levels of interferon gamma (*IFNG*), granzyme B (*GZMB*) and perforin (*PRF1*) were also measured in immune active tissues, indicating a *bona fide* immune response. The presence of infiltrating immune cells in ovarian cancer was first reported nearly 40 years ago (15). Numerous studies elicited the favorable prognostic significance of increased numbers of various types of immune cells or overexpression of individual immune cell markers (9-11,27-29). Our works revealed that the immune activation in HGSOC was an overall reaction of the host rather than a response of a single cell type or the expression alteration of a single gene.

Additionally, our work revealed the complex network of tumor environment in HGSOC, presented by the co-overexpression of genes involved in the remodeling of ECM and adipose tissue along with the immune genes. HGSOC tumor tissues present intricate admixtures of tumors, infiltrating immune cells and stromal cells, making it difficult to isolate specific cell contingents by means of microdissection. RNA sequencing of the total tumor enabled us to identified differentially expressed genes, most of which are cell specific. Tumor cells are usually embedded within the ECM and separated in space from immune cells. To execute their immune functions, some immune cells like CD8+ T cells, must come into contact with epithelial tumor cells. In lung cancer, active T cell motility was observed in loose fibronectin and collagen regions. Furthermore, reducing ECM constructive components can increase the number of T cells in contact with tumor cells (30). As a consequence, the destruction of the ECM induced by immune cells could trigger a process resembling wound healing (31,32), whereas ECM components could be produced by other stromal cells or tumor cells. Therefore, we found an overexpression of genes not only responsible for the degradation (32,33) but also for the production of ECM components in immune active samples, demonstrating that the remodeling of the ECM could be a dynamic process, in which different molecules and cells interact in the microenvironment in response to tumors. Besides, we also found numerous genes responsible for the remodeling of adipose tissues,

which were predominantly overexpressed in most of the immune active tumor samples. Advanced ovarian cancer is often embedded in fat tissue (34,35). In the same sense that active immune cells have to traverse the ECM, they also have to travel across the adipose tissue to come into contact with tumor cells. For the development of immune therapies against HGSOC, it is very essential to understand the network of tumor microenvironment. If we could destroy the adipose tissue and the ECM around tumors, the accessibility of immune cells and other therapeutics might be increased and thus the therapy efficacy might be improved, an issue that has already drawn the attention of scientists (36).

In particular, overexpression of the B7-CD28 gene family members (*PD-1/PD-L1*, *CD80/CTLA4* among others), which are known to negatively regulate T cell activation (37-40), was also found in immune active samples. It is in line with a recent report, showing that PD-1+ T infiltrating lymphocytes and PD-L1+ tumor cells were both favorable prognostic factors in ovarian cancer (41). The inverse function of the immune checkpoint molecules in controlling the T cell activation in other types of cancers was also suggested by several other independent studies (42-44). As a matter of fact, some of the B7-CD28 members, such as CTLA-4, have been reported to brake T regulatory cell proliferation driven by CD28 (45) or to inhibit CD28 costimulation by depleting the ligands (46). Similar findings suggest that the immune system can also facilitate tumor progression (47). Our data support the complex regulation of the immune system by the B7-CD28 family genes.

The identification of tumor associated antigens (TAA) is of great importance because they represent potential targets of tumor vaccines (48). We found a significant overexpression of *CEACAM21* in immune active samples. *CEACAM21* is a member of the carcinoembryonic antigen family, which is usually expressed in gastrointestinal tissues during embryonal development and is silenced before birth. Very little is known concerning this gene, making it a very interesting target for further investigations. However, this observation must be validated on an independent HGSOC cohort and the real function of CEACAM21 needs to be

unraveled before being conclusive on its true status as a tumor antigen. We have established a series of HGSOC cell lines (49), which could serve as models to study the impact of *CEACAM21* on immunogenicity.

We found overexpression of *ABCB1 (MDR1),* which was correlated with platinum-resistance in HGSOC in a recent study (4). It is of note that *MDR1* was also reported to be expressed in CD34+ stem cells, dendritic cells, NK cells and cytotoxic T lymphocytes suggesting that it may protect immune cells against stress-induced or bystander lysis (50). Thus the upregulation of *MDR1* in the immune active samples might also be part of the immune response in these tumors.

The findings reported here lead to some considerations for ovarian cancer research and therapy development. First, the B7-CD86 family members showed a concordant overexpression with the host immune activation. This suggests that immune regulation in HGSOC might be different than in melanomas. We could show a disease-free survival difference between immune active and silent tumors in the TCGA data, implying that these genes might have other functions in immune modulation. However, this difference could neither be confirmed by our own data nor by the data from Patch et al., indicating that further studies are needed to verify the results. Studies with a larger cohort of samples and a thorough investigation on the functions of the gene family must be performed in order to design innovative immune therapy strategies. Second, the co-expression of immune genes with that of ECM and adipose tissue remodeling genes is intriguing, because enzymes for both ECM construction and destruction have been found. It could be of important therapeutic consequences, since loosening the structure of the ECM or adipose tissue would not only facilitate contact of immune cells with tumor cells, but might also increase exposure of cancer cells to drugs. Hence, it will be important to fine tune immune therapy strategies in HGSOC. Finally, some recurrent tumors are quite similar as their primary counterparts while others are not. If recurrent tumors could be analyzed at molecular levels and compared with their

primary counterparts, the scheme of treatment could be adjusted according to the primary response of the patients.

It is very important to examine if the immune activation status would have an effect on the patients' outcome. Our cohort was rather small for survival analyses, when stratifying the patients by residual tumor status, one of the most important factors affecting both disease-free and overall survival. In the cohort from a public dataset with 80 primary tumors, the residual tumor status was missing and the results of survival advantage of patients with immune activation were not significant. Studies with larger sample cohort with available information on rest tumor will be needed to clarify the impact of immune activation on clinical outcome of the patients.

The study produced a rather large dataset. Based on the variance in tumor cell contents of the samples, we decided to use stringent conditions to analyze differentially expressed genes in all analytical procedures in order to obtain reliable results. This might lead to the nonobservance of some genes, which had less expression changes or/and lower expression. Alternative analyses using lower stringency could be performed to investigate such genes or their related biological processes. However, results from such analyses must be validated using bigger sample cohorts.

In conclusion, we observed a difference of a complex microenvironment network including immune activation and remodeling of tissues around tumors across all tumor samples and defined tumors as immune active and immune silent. By subgrouping paired tumor lesions into active-active, active-silent, silent-active, and silent-silent ones, we found that there was no notable difference between primary and recurrent tumors in the active-active and silent-silent subgroups, suggesting that these recurrent tumors resemble their primary counterparts. Other tumors presented major molecular difference by switching their immune status, either from silent to active or active to silent, accompanied by the corresponding

changes in the microenvironment. Interestingly, major genes of the B7-CD28 immune checkpoint family, which are considered to negatively regulate the immune activation, as well as *CEACAM21*, which is a member of the carcinoembryonic antigen family, and *MDR1*, were also overexpressed in immune active tissues in comparison to the silent ones.

Author contributions

Experiments were performed by CK, DS, JK, AW, KJ and SDE. IB, JS, EB, IV, DL, HG, GH, CT, DCCT conceived the study design and coordination. AG, GH, CK, BB, CT and DCCT contributed to the data processing, data analyses, data interpretation and presentation. CK, IB, AW, JS, FT, AV, HG, BW, SP provided samples and clinical information. SDE and RH approved the pathological results; DCCT, CK, CT drafted the manuscript.

Acknowledgements

This work was supported by European Community's Seventh Framework Programme under grant agreement No. 279113-2 (OCTIPS; www.octips.eu). The documentation of clinical and patient's data was managed with "AlcedisTRIAL the web based documentation system" of Alcedis GmbH, Winchesterstr. 3, 35394 Giessen, Germany. We thank Ms. Elisabeth Maritschnegg and Ms. Jasmin Karacs for their technical support.

References

1. Bowtell DD, Bohm S, Ahmed AA, Aspuria PJ, Bast RC, Jr., Beral V, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer 2015;15(11):668-79.
2. Berns EM, Bowtell DD. The changing view of high-grade serous ovarian cancer. Cancer Res 2012;72(11):2701-4.
3. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature 2011;474(7353):609-15.
4. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. Nature 2015;521(7553):489-94.
5. Cooke SL, Brenton JD. Evolution of platinum resistance in high-grade serous ovarian cancer. Lancet Oncol 2011;12(12):1169-74.
6. Cooke SL, Ng CK, Melnyk N, Garcia MJ, Hardcastle T, Temple J, et al. Genomic analysis of genetic heterogeneity and evolution in high-grade serous ovarian carcinoma. Oncogene 2010;29(35):4905-13.
7. Schwarz RF, Ng CK, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med 2015;12(2):e1001789.
8. Castellarin M, Milne K, Zeng T, Tse K, Mayo M, Zhao Y, et al. Clonal evolution of high-grade serous ovarian carcinoma from primary to recurrent disease. J Pathol 2013;229(4):515-24.
9. Milne K, Kobel M, Kalloger SE, Barnes RO, Gao D, Gilks CB, et al. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. PLoS One 2009;4(7):e6412.
10. Bachmayr-Heyda A, Aust S, Heinze G, Polterauer S, Grimm C, Braicu EI, et al. Prognostic impact of tumor infiltrating CD8+ T cells in association with cell proliferation in ovarian cancer patients--a study of the OVCAD consortium. BMC Cancer 2013;13:422.
11. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003;348(3):203-13.
12. Ryner L, Guan Y, Firestein R, Xiao Y, Choi Y, Rabe C, et al. Upregulation of Periostin and Reactive Stroma Is Associated with Primary Chemoresistance and Predicts Clinical Outcomes in Epithelial Ovarian Cancer. Clin Cancer Res 2015;21(13):2941-51.
13. Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, et al. A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res 2014;20(3):711-23.
14. Schauer IG, Sood AK, Mok S, Liu J. Cancer-associated fibroblasts and their putative role in potentiating the initiation and development of epithelial ovarian cancer. Neoplasia 2011;13(5):393-405.
15. Haskill S, Becker S, Fowler W, Walton L. Mononuclear-cell infiltration in ovarian cancer. I. Inflammatory-cell infiltrates from tumour and ascites material. Br J Cancer 1982;45(5):728-36.
16. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 2008;14(16):5198-208.

17. Berns E, Gourley C, Lambrechts S, Schmitt W, Darb-Esfahani S, Cacsire Castillo-Tong D. Standard Sample Processing Protocol. Protoc exch 2016.
18. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013;14(4):R36.
19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357-9.
20. Tibshirani R., Chu G., Balasubramanian Narasimhan, Li J. SAM: Significance Analysis of Microarrays. 2011.
21. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol 2015;16:64.
22. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res 2016;44(D1):D710-6.
23. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 1979;Vol. 6, No. 2 65-70.
24. Wolkewitz M, Allignol A, Harbarth S, de Angelis G, Schumacher M, Beyersmann J. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. J Clin Epidemiol 2012;65(11):1171-80.
25. Therneau TM, Grambsch, Patricia M. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health 2000.
26. Chekerov R, Braicu I, Castillo-Tong DC, Richter R, Cadron I, Mahner S, et al. Outcome and clinical management of 275 patients with advanced ovarian cancer International Federation of Obstetrics and Gynecology II to IV inside the European Ovarian Cancer Translational Research Consortium-OVCAD. Int J Gynecol Cancer 2013;23(2):268-75.
27. Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, Qian F, et al. Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. Proc Natl Acad Sci U S A 2005;102(51):18538-43.
28. Dong HP, Elstrand MB, Holth A, Silins I, Berner A, Trope CG, et al. NK- and B-cell infiltration correlates with worse outcome in metastatic ovarian carcinoma. Am J Clin Pathol 2006;125(3):451-8.
29. Wouters MC, Komdeur FL, Workel HH, Klip HG, Plat A, Kooi NM, et al. Treatment Regimen, Surgical Outcome, and T-cell Differentiation Influence Prognostic Benefit of Tumor-Infiltrating Lymphocytes in High-Grade Serous Ovarian Cancer. Clin Cancer Res 2016;22(3):714-24.
30. Salmon H, Donnadieu E. Within tumors, interactions between T cells and tumor cells are impeded by the extracellular matrix. Oncoimmunology 2012;1(6):992-94.
31. Arwert EN, Hoste E, Watt FM. Epithelial stem cells, wound healing and cancer. Nat Rev Cancer 2012;12(3):170-80.
32. Fan MH, Zhu Q, Li HH, Ra HJ, Majumdar S, Gulick DL, et al. Fibroblast Activation Protein (FAP) Accelerates Collagen Degradation and Clearance from Lungs in Mice. J Biol Chem 2016;291(15):8070-89.
33. Bekhouche M, Leduc C, Dupont L, Janssen L, Delolme F, Vadon-Le Goff S, et al. Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-beta signaling as primary targets. FASEB J 2016;30(5):1741-56.

34. Nieman KM, Kenny HA, Penicka CV, Ladanyi A, Buell-Gutbrod R, Zillhardt MR, et al. Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. Nat Med 2011;17(11):1498-503.
35. Villanueva MT. Gynecological cancer: Home is where the fat is. Nature reviews Clinical oncology 2012;9(1):6.
36. Au JL, Yeung BZ, Wientjes MG, Lu Z, Wientjes MG. Delivery of cancer therapeutics to extracellular and intracellular targets: Determinants, barriers, challenges and opportunities. Adv Drug Deliv Rev 2016;97:280-301.
37. Chambers CA, Kuhns MS, Egen JG, Allison JP. CTLA-4-mediated inhibition in regulation of T cell responses: mechanisms and manipulation in tumor immunotherapy. Annual review of immunology 2001;19:565-94.
38. Cimino-Mathews A, Thompson E, Taube JM, Ye X, Lu Y, Meeker A, et al. PD-L1 (B7-H1) expression and the immune tumor microenvironment in primary and metastatic breast carcinomas. Hum Pathol 2016;47(1):52-63.
39. Lyford-Pike S, Peng S, Young GD, Taube JM, Westra WH, Akpeng B, et al. Evidence for a role of the PD-1:PD-L1 pathway in immune resistance of HPV-associated head and neck squamous cell carcinoma. Cancer Res 2013;73(6):1733-41.
40. Taube JM, Young GD, McMiller TL, Chen S, Salas JT, Pritchard TS, et al. Differential Expression of Immune-Regulatory Genes Associated with PD-L1 Display in Melanoma: Implications for PD-1 Pathway Blockade. Clin Cancer Res 2015;21(17):3969-76.
41. Darb-Esfahani S, Kunze CA, Kulbe H, Sehouli J, Wienert S, Lindner J, et al. Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high grade serous carcinoma. Oncotarget 2016;7(2):1486-99.
42. Kluger HM, Zito CR, Barr ML, Baine MK, Chiang VL, Sznol M, et al. Characterization of PD-L1 Expression and Associated T-cell Infiltrates in Metastatic Melanoma Samples from Variable Anatomic Sites. Clin Cancer Res 2015;21(13):3052-60.
43. Schmidt LH, Kummel A, Gorlich D, Mohr M, Brockling S, Mikesch JH, et al. PD-1 and PD-L1 Expression in NSCLC Indicate a Favorable Prognosis in Defined Subgroups. PLoS One 2015;10(8):e0136023.
44. Tjin EP, Krebbers G, Meijlink KJ, van de Kasteele W, Rosenberg EH, Sanders J, et al. Immune-escape markers in relation to clinical outcome of advanced melanoma patients following immunotherapy. Cancer Immunol Res 2014;2(6):538-46.
45. Holt MP, Punkosdy GA, Glass DD, Shevach EM. TCR Signaling and CD28/CTLA-4 Signaling Cooperatively Modulate T Regulatory Cell Homeostasis. Journal of immunology 2017;198(4):1503-11.
46. Qureshi OS, Zheng Y, Nakamura K, Attridge K, Manzotti C, Schmidt EM, et al. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. Science 2011;332(6029):600-3.
47. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annual review of immunology 2004;22:329-60.
48. Trajanoski Z, Maccalli C, Mennonna D, Casorati G, Parmiani G, Dellabona P. Somatically mutated tumor antigens in the quest for a more efficacious patient-oriented immunotherapy of cancer. Cancer Immunol Immunother 2015;64(1):99-104.
49. Kreuzinger C, Gamperl M, Wolf A, Heinze G, Geroldinger A, Lambrechts D, et al. Molecular characterization of 7 new established cell lines from high grade serous ovarian cancer. Cancer Lett 2015;362(2):218-28.
50. Johnstone RW, Ruefli AA, Tainton KM, Smyth MJ. A role for P-glycoprotein in regulating cell death. Leuk Lymphoma 2000;38(1-2):1-11.

Figure legends

Figure 1. Unsupervised clustering of recurrent samples (n=66) and presentation of genes with the highest expression difference between the two clusters. A. The coding numbers of the samples are indicated at the right side. Numbers under the color key indicate the normalized read counts. C1 and C2 indicate cluster 1 and cluster 2, respectively. B. Cell specific expression and major functions of the 140 genes with overexpression in C1 compared with C2. While 126 genes have direct functions in immune response (12, 33 and 1 genes expressed specifically in B cells, T cells, and monocytes, respectively; 59 genes code for immunoglobulins and 21 have other functions directly involved in immune activation), no certain immune relevant functions could be found for 14 genes. CTL: cytotoxic T cells; CTL / NK: genes expressed specifically in cytolytic T cells and natural killer cells.

Figure 2. Expression of key genes involved in immune activation in each subgroup. Immune activation status is indicated on the top and bottom of the Figure for primary and recurrent tumor tissues, respectively. The patient coding is shown in the middle lane. Gene expression levels for primary and recurrent samples are given separately at the upper and lower blocks of the Figure. ** and * indicate genes with significant overexpression in active versus silent samples with $p<0.001$ and $p<0.01$, respectively. Color key on the right bottom indicates gene expression.

Figure 3. Representative IHC staining of 2 selected immune active tumor samples. The antibodies and the tumor tissue codes were indicated on top of each figure. A-D are from a series of sections, E-H are from another series of sections.

Figure 4. Genes with the highest expression difference in primary versus recurrent samples in each subgroup. Differentially expressed genes in primary versus recurrent samples were extracted for each subgroup, presented in each pyramid. Over-expressed genes in primary

and recurrent tumors are presented at the left and right side of the pyramids, respectively. The number in the middle of the pyramids indicates the number of samples (For example: B: 15 in the middle of the pyramid with 11 immune gene and *POSTN* at the right side indicates that in 15/21 patients in this subgroup, 11 immune related genes and the gene *POSTN* are over-expressed in recurrent samples). Immune relevant genes are indicated in red; extracellular matrix related genes are labeled green; genes related to the remodeling of adipose tissues are labeled pink. Additional genes without related functions are labeled blue.

Figure 5. Expression of genes with important roles in HGSOC in each subgroup. 5A and 5B: The sample coding is given at the bottom of 5B. Multiplicity adjusted *p*-values indicating the results of the comparison of gene expressions between active and silent tissues were indicated for *ABCB1* and *CEACAM21*. For other genes, they are >0.2. A: genes relevant to HGSOC; B: genes coding for possible tumor antigens. 5C and 5D: Expression of *ABCB1* and *CEACAM21*, respectively. Numbers below the line-plots indicate median gene expression for primary and recurrent tumors in each subgroup. D and R at the bottom of each plot indicate the samples collected at diagnosis and recurrence, respectively.

Figure 6. Outcome of the patients from the cohort of TCGA (n=421) and Patch et al. (n=80). A: overall survival (TCGA data); B: relapse-free survival (TCGA data); C: overall survival (data from Patch et al.); D: relapse-free survival (data from Patch et al.). The adjusted p-values were indicated for each analysis.

Titles and notes of supplementary materials:

Supplementary Table S1. Clinical and pathological parameters of the patients and the tumors. Primary tumor samples are indicated with “D” and a number (e.g. TB001D1) and the recurrent samples are indicated with “R”. pN is the regional lymph node status with “1” and

“0” indicating the presence and the absence of tumor cells in lymph node, respectively. pT indicates the extension of the tumor spread and the size of tumors.

Supplementary Table S2. Functional assignment of the 142 genes with the highest expression difference in the two clusters revealed by unsupervised clustering of recurrent samples. FC indicates the fold change of gene expression in C1/C2 shown in figure 1. Only two genes (*PCP4* and *PHOX2A*, last two lines) are overexpressed in C2. The assignment of genes specifically expressed in immune cell is according to Angelova et al. Other immune related genes are indicated with * and are assigned according to literature.

Supplementary Table S3. Functional assignment of the 113 genes with the highest expression difference in the two clusters revealed by unsupervised clustering of primary samples. FC indicates the fold change of gene expression in the cluster with 25 samples vs. 41 samples. Only three genes (*NME2P1*, *RPS13P2,* and *APOA1*, last three lines) were overexpressed in the cluster with 41 samples. The assignment of immune cell specific gene expression is according to Angelova et al. Other immune related genes were indicated with *, which were assigned according to literature.

Supplementary Table S4. Subgrouping of paired tumor tissues defined by the immune activation status.

Supplementary Figure S1. Supervised clustering of primary and recurrent tumors. The primary (A) and recurrent tumors (B) were clustered using the 126 immune relevant genes. The coding numbers of the samples are indicated on the right side of each heatmap. The numbers under the color keys indicate the normalized read counts. Samples with overexpression of the 126 genes were indicated as active and the other samples as silent. The ordering of the genes in Figure B was set in the same way as in Figure A. C: Expression of IGKV4-1 and CD79A. Line-plots indicate gene expression for primary and recurrent

tumors in each subgroup. D and R at the bottom of each plot indicate the samples at diagnosis and recurrence, respectively.

Supplementary Figure S2. IHC staining of CD8 and CD20 in 4 active and 6 silent tissues. The antibody and the tumor tissue coding are shown on top of each figure. The immune activation status of the tissues is indicated at the left side. D at the end of the sample coding indicates a sample at diagnosis and R at recurrence.

Supplementary Figure S3. Supervised clustering of primary HGSOC (n=421) from the TCGA data using 59/126 immune related genes. Only 59 out of 126 genes could be found in the TCGA data set (see methods). Tumor samples from 45 patients were classified as immune active and 376 as immune silent. Numbers under the color key indicate the normalized gene expression values. The ordering of the genes was set the same as the ordering in Supplementary Figure S1A.

Supplementary Figure S4. Supervised clustering of primary HGSOC (n=80) from the data set of Patch et al using the 126 immune related genes. 54 samples are defined as immune active and 26 as immune silent. Numbers under the color key indicate the normalized read counts. The ordering of the genes was set the same as the ordering in Supplementary Figure S1A.

Figure 1.

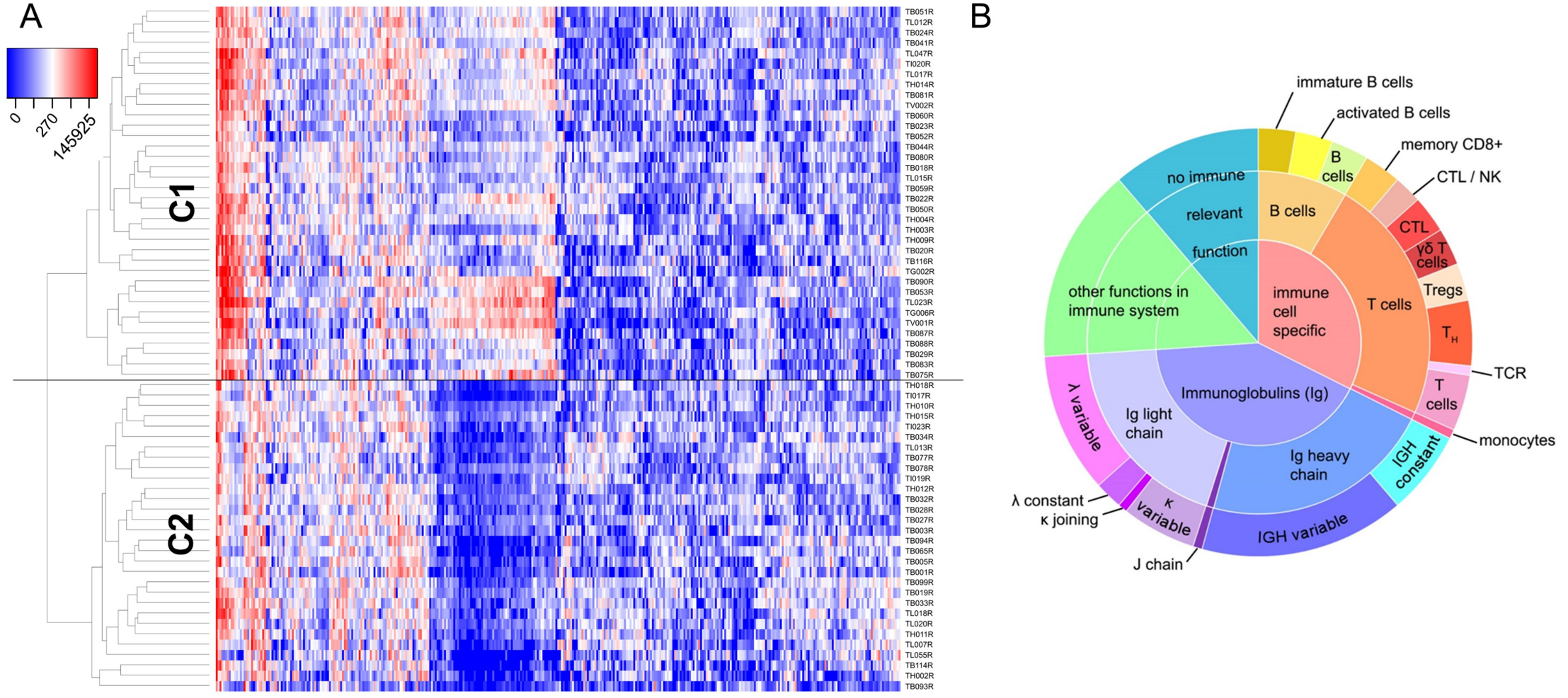

A
0
270
145925
C1
C2
TB051R
TL012R
TB024R
TB041R
TL047R
TI020R
TL017R
TH014R
TB081R
TV002R
TB060R
TB023R
TB052R
TB044R
TB080R
TB018R
TL015R
TB059R
TB022R
TB050R
TH004R
TH003R
TH009R
TB020R
TB116R
TG002R
TB090R
TB053R
TL023R
TG006R
TV001R
TB087R
TB088R
TB029R
TB083R
TB075R
TH018R
TI017R
TH010R
TH015R
TI023R
TB034R
TL013R
TB077R
TB078R
TI019R
TH012R
TB032R
TB028R
TB027R
TB003R
TB094R
TB065R
TB005R
TB001R
TB099R
TB019R
TB033R
TL018R
TL020R
TH011R
TL007R
TL055R
TB114R
TH002R
TB093R
B
immature B cells
activated B cells
memory CD8+
CTL / NK
B cells
no immune relevant function
B cells
CTL
γδ T cells
Tregs
T cells
other functions in immune system
immune cell specific
T cells
TCR
monocytes
Immunoglobulins (Ig)
λ variable
Ig light chain
Ig heavy chain
IGH constant
λ constant
κ joining
κ variable
IGH variable
J chain

Figure 2.

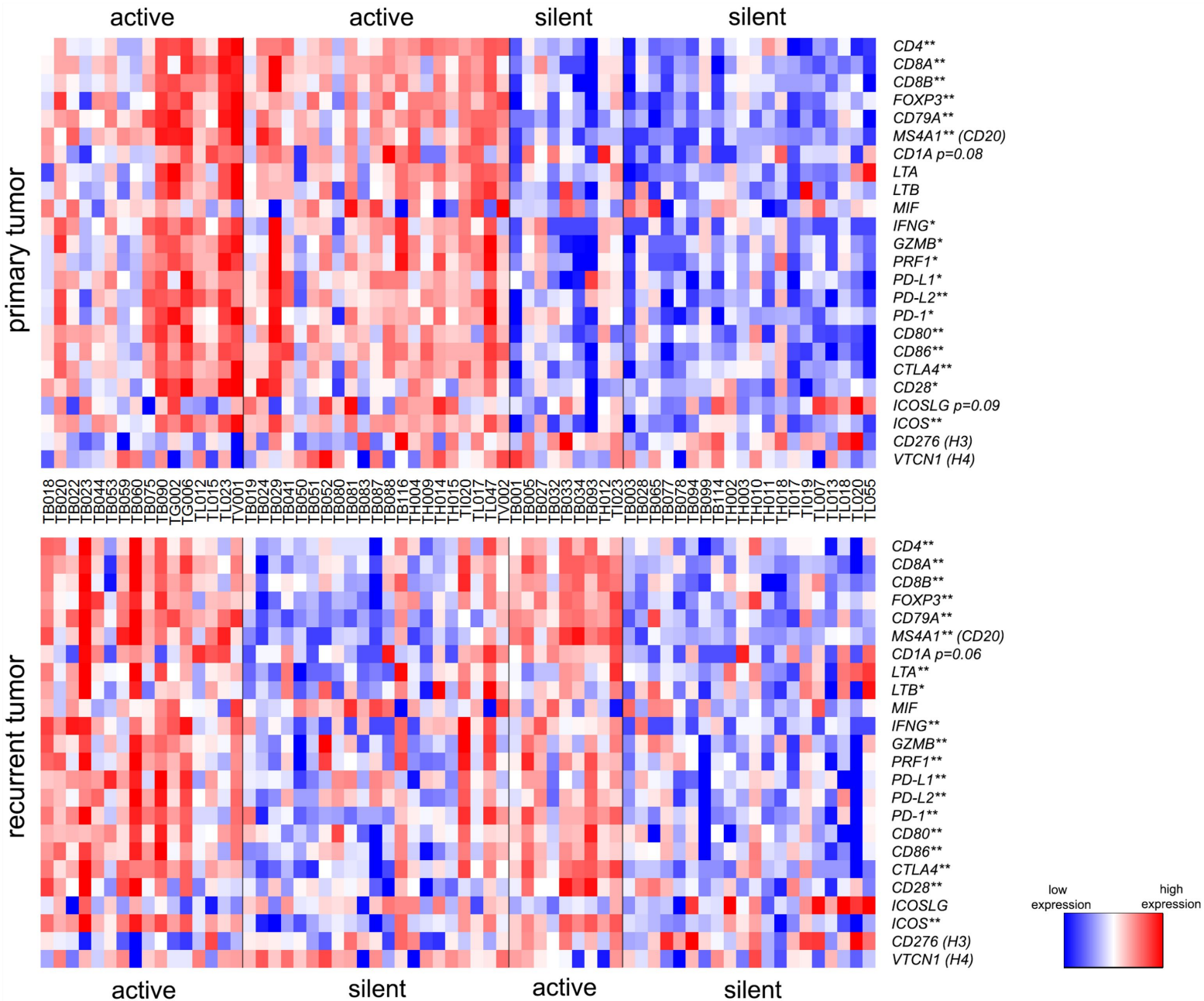

active
active
silent
silent
primary tumor
CD4**
CD8A**
CD8B**
FOXP3**
CD79A**
MS4A1** (CD20)
CD1A p=0.08
LTA
LTB
MIF
IFNG*
GZMB*
PRF1*
PD-L1*
PD-L2**
PD-1*
CD80**
CD86**
CTLA4**
CD28*
ICOSLG p=0.09
ICOS**
CD276 (H3)
VTCN1 (H4)
TB018 TB020 TB022 TB023 TB044 TB053 TB059 TB060 TB075 TB090 TG002 TG006 TL012 TL015 TL023 TV001 TB019 TB024 TB029 TB041 TB050 TB051 TB052 TB080 TB081 TB083 TB087 TB088 TB116 TH004 TH009 TH014 TH015 TI020 TL017 TL047 TV002 TB001 TB005 TB027 TB032 TB033 TB034 TB093 TH012 TI023 TB003 TB028 TB065 TB077 TB078 TB094 TB099 TB114 TH002 TH003 TH010 TH011 TH018 TI017 TI019 TL007 TL013 TL018 TL020 TL055
recurrent tumor
CD4**
CD8A**
CD8B**
FOXP3**
CD79A**
MS4A1** (CD20)
CD1A p=0.06
LTA**
LTB*
MIF
IFNG**
GZMB**
PRF1**
PD-L1**
PD-L2**
PD-1**
CD80**
CD86**
CTLA4**
CD28**
ICOSLG
ICOS**
CD276 (H3)
VTCN1 (H4)
active
silent
active
silent
low expression
high expression

Figure 3

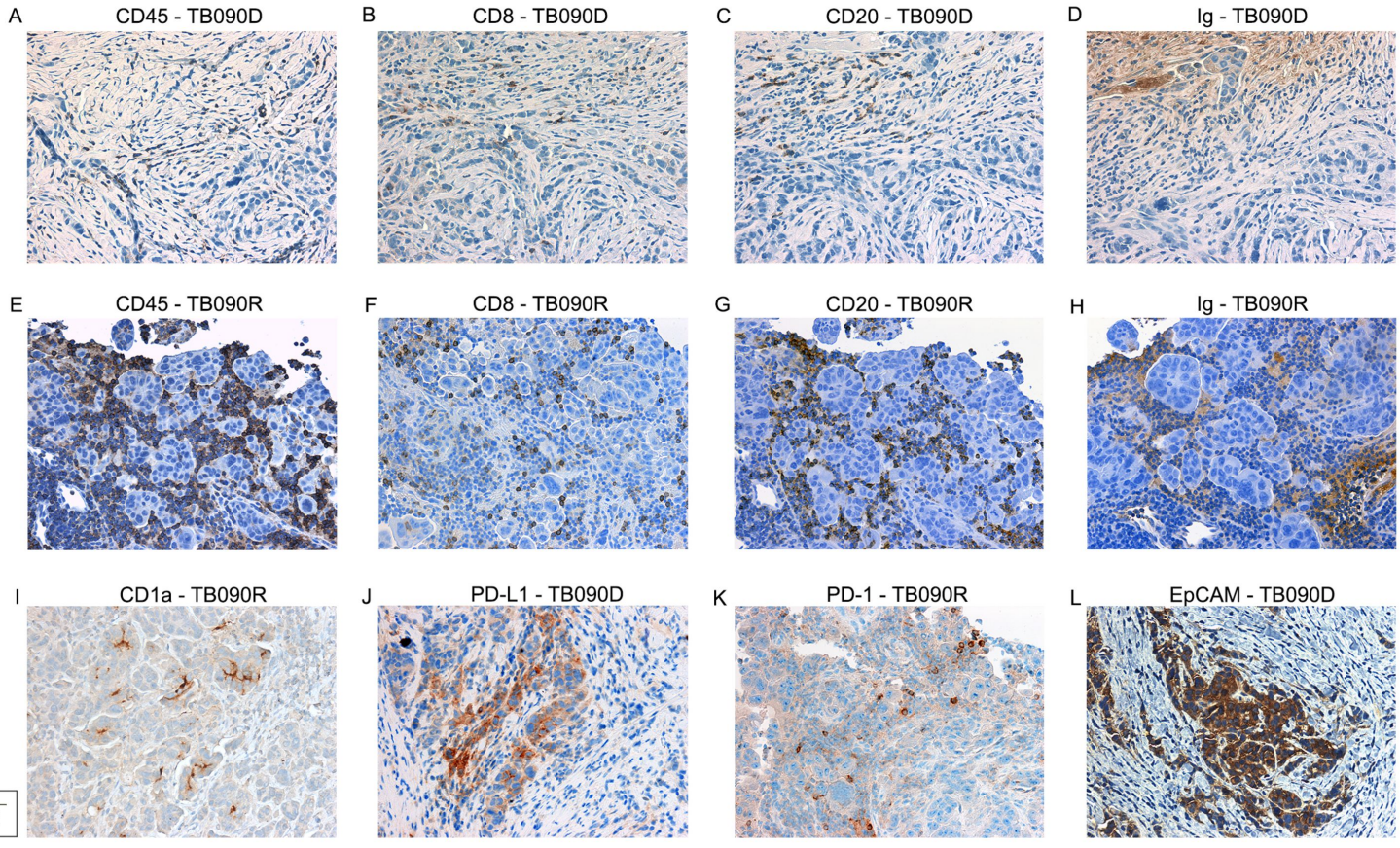

A
CD45 - TB090D
B
CD8 - TB090D
C
CD20 - TB090D
D
Ig - TB090D
E
CD45 - TB090R
F
CD8 - TB090R
G
CD20 - TB090R
H
Ig - TB090R
I
CD1a - TB090R
J
PD-L1 - TB090D
K
PD-1 - TB090R
L
EpCAM - TB090D
100µm

Figure 4.

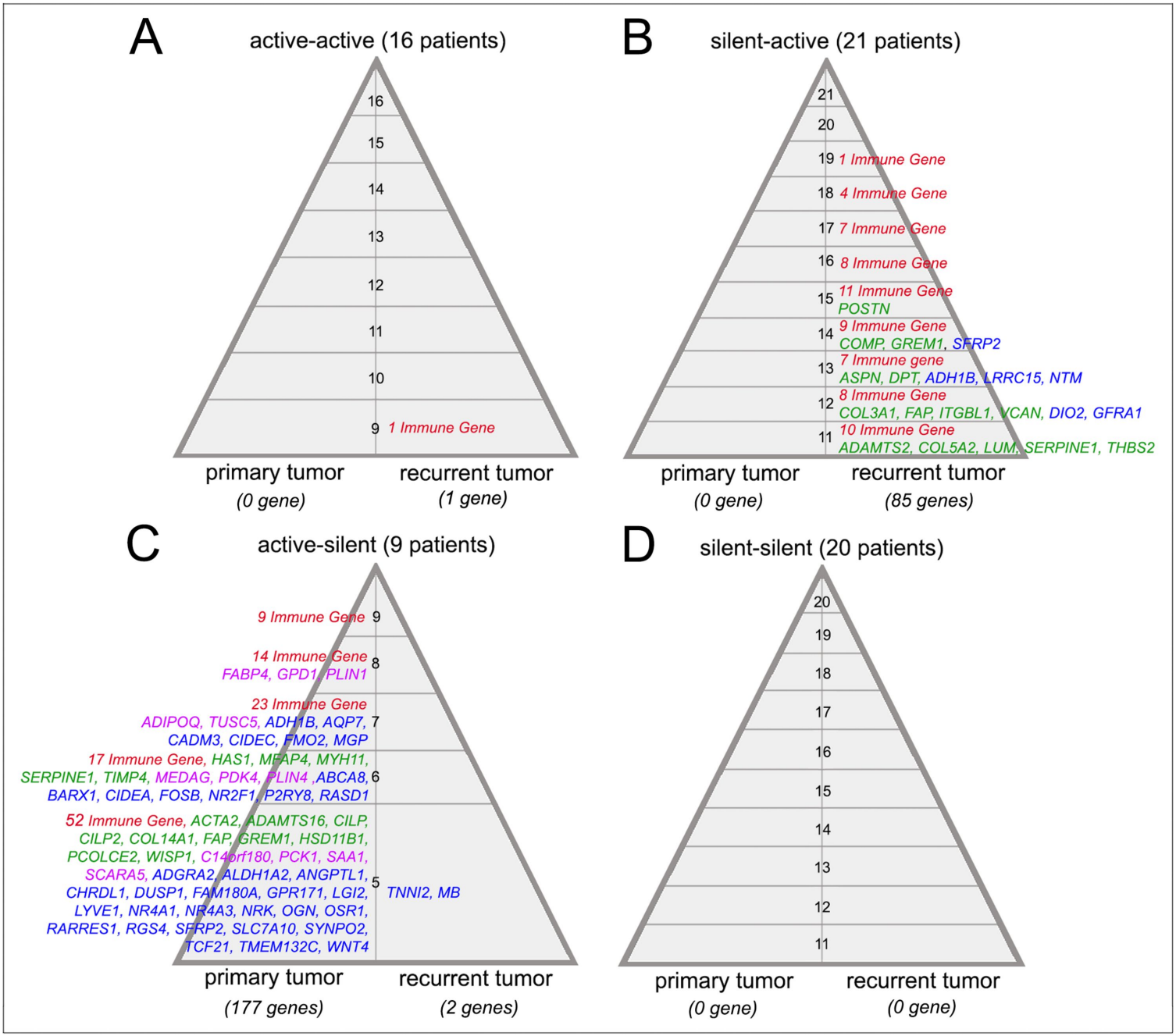
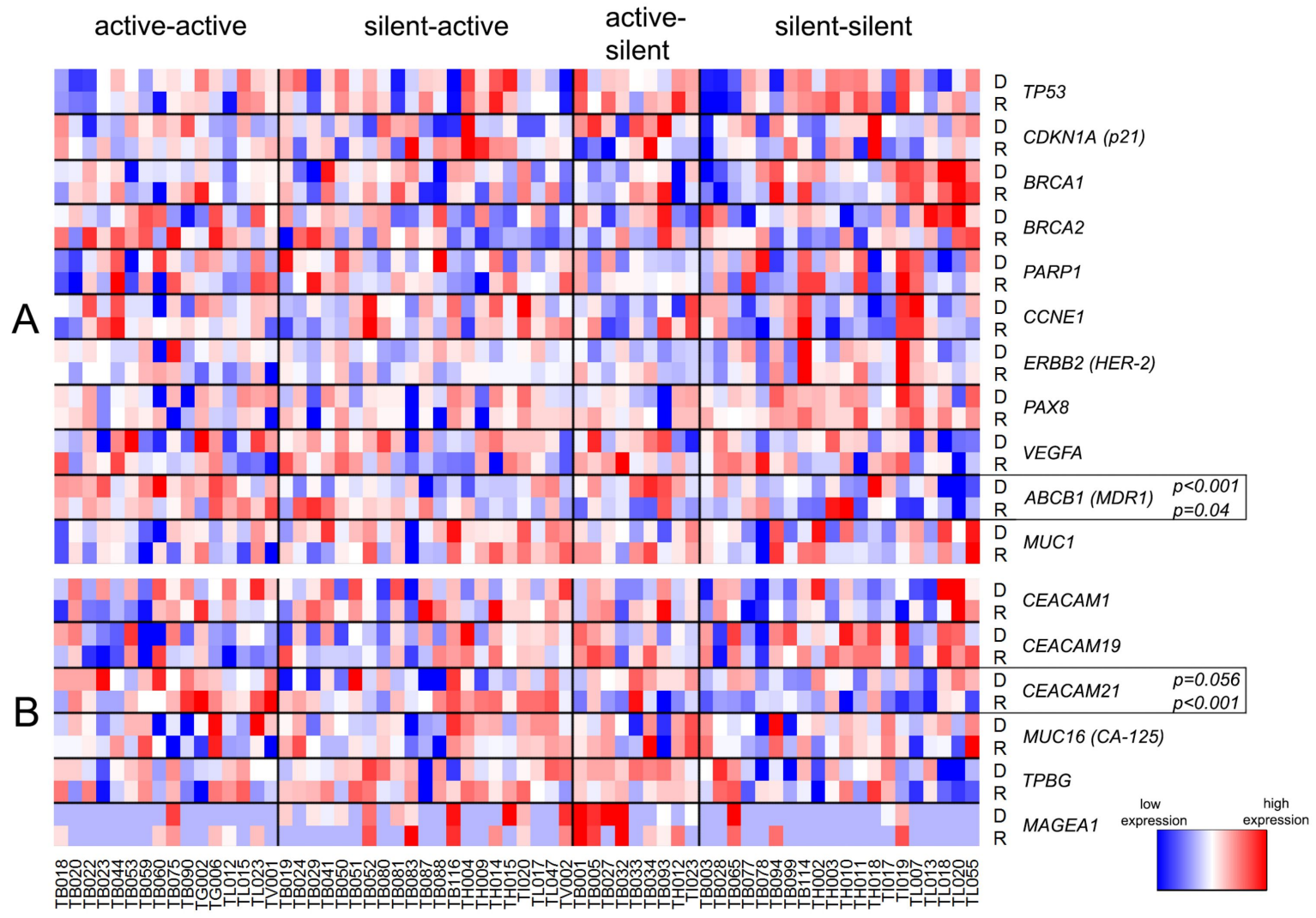

A
active-active (16 patients)
16
15
14
13
12
11
10
9 1 Immune Gene
primary tumor
(0 gene)
recurrent tumor
(1 gene)
B
silent-active (21 patients)
21
20
19 1 Immune Gene
18 4 Immune Gene
17 7 Immune Gene
16 8 Immune Gene
15 11 Immune Gene
POSTN
14 9 Immune Gene
COMP, GREM1, SFRP2
13 7 Immune gene
ASPN, DPT, ADH1B, LRRC15, NTM
12 8 Immune Gene
COL3A1, FAP, ITGBL1, VCAN, DIO2, GFRA1
11 10 Immune Gene
ADAMTS2, COL5A2, LUM, SERPINE1, THBS2
primary tumor
(0 gene)
recurrent tumor
(85 genes)
C
active-silent (9 patients)
9 Immune Gene 9
14 Immune Gene
FABP4, GPD1, PLIN1 8
23 Immune Gene
ADIPOQ, TUSC5, ADH1B, AQP7, 7
CADM3, CIDEC, FMO2, MGP
17 Immune Gene, HAS1, MFAP4, MYH11,
SERPINE1, TIMP4, MEDAG, PDK4, PLIN4 ,ABCA8, 6
BARX1, CIDEA, FOSB, NR2F1, P2RY8, RASD1
52 Immune Gene, ACTA2, ADAMTS16, CILP,
CILP2, COL14A1, FAP, GREM1, HSD11B1,
PCOLCE2, WISP1, C14orf180, PCK1, SAA1,
SCARA5, ADGRA2, ALDH1A2, ANGPTL1,
CHRDL1, DUSP1, FAM180A, GPR171, LGI2, 5 TNNI2, MB
LYVE1, NR4A1, NR4A3, NRK, OGN, OSR1,
RARRES1, RGS4, SFRP2, SLC7A10, SYNPO2,
TCF21, TMEM132C, WNT4
primary tumor
(177 genes)
recurrent tumor
(2 genes)
D
silent-silent (20 patients)
20
19
18
17
16
15
14
13
12
11
primary tumor
(0 gene)
recurrent tumor
(0 gene)

Figure 5.


active-active
silent-active
active-silent
silent-silent
A
B
D
R
TP53
CDKN1A (p21)
BRCA1
BRCA2
PARP1
CCNE1
ERBB2 (HER-2)
PAX8
VEGFA
ABCB1 (MDR1)
p<0.001
p=0.04
MUC1
CEACAM1
CEACAM19
CEACAM21
p=0.056
p<0.001
MUC16 (CA-125)
TPBG
MAGEA1
low expression
high expression

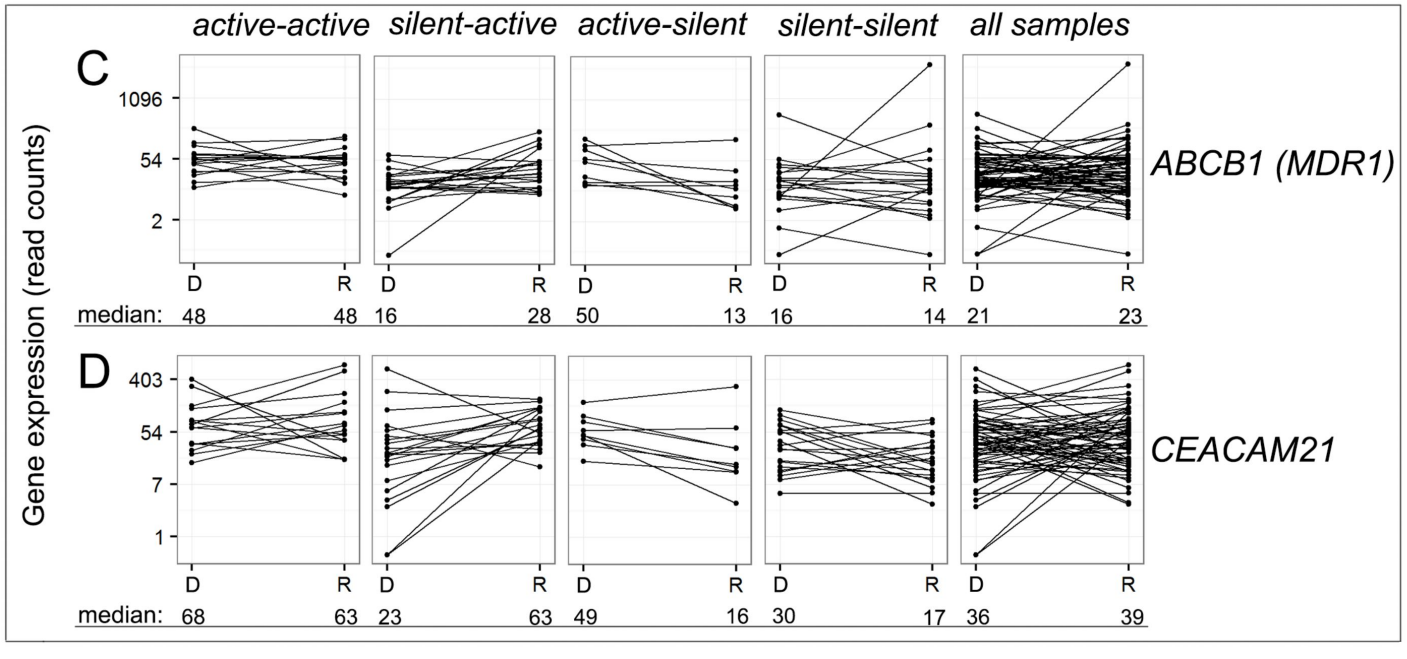
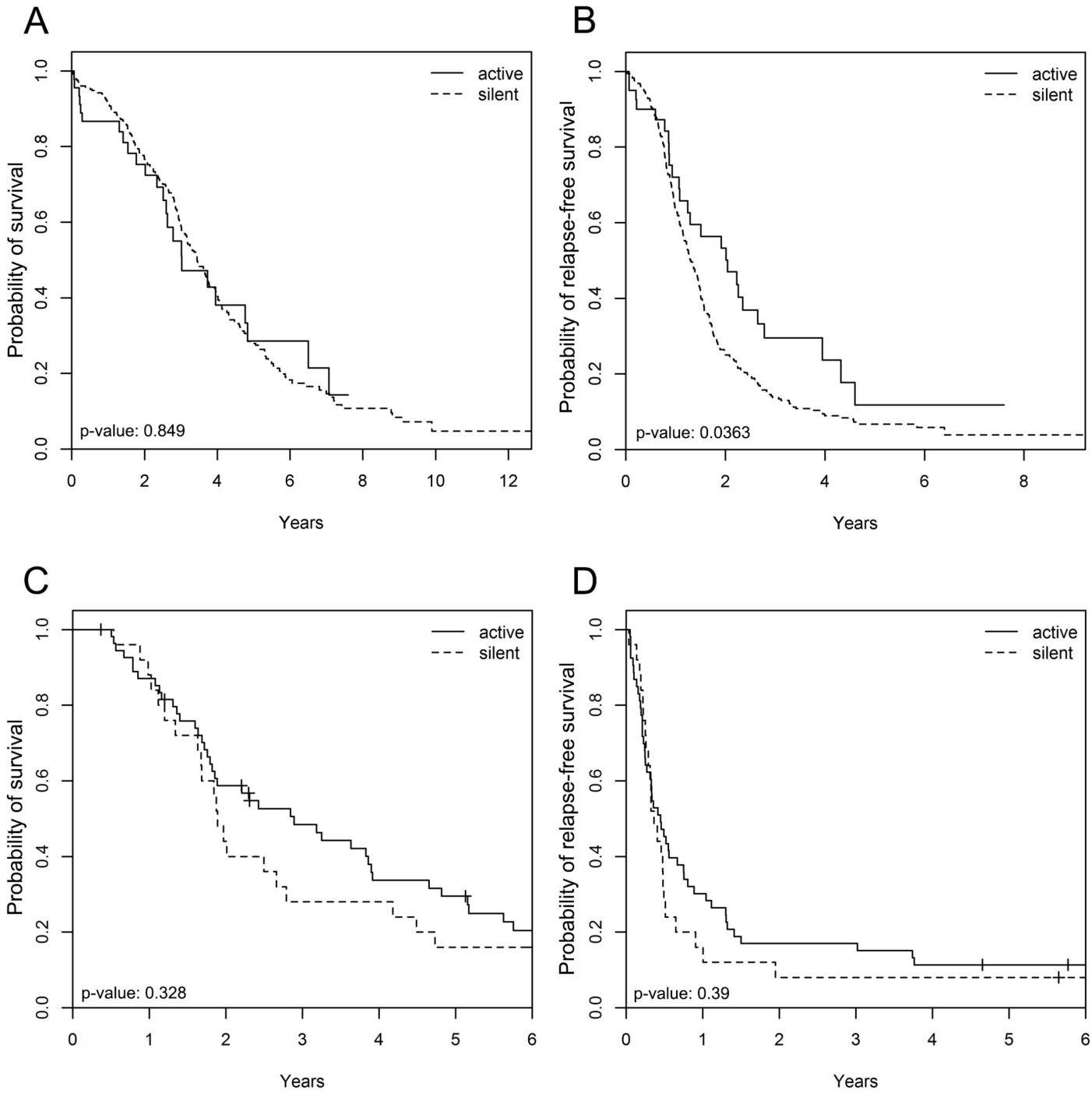

active-active
silent-active
active-silent
silent-silent
all samples
C
D
Gene expression (read counts)
1096
54
2
403
7
1
D
R
median:
48 48 16 28 50 13 16 14 21 23
68 63 23 63 49 16 30 17 36 39
ABCB1 (MDR1)
CEACAM21

Figure 6.


A
B
C
D
active
silent
Probability of survival
Probability of relapse-free survival
p-value: 0.849
p-value: 0.0363
p-value: 0.328
p-value: 0.39
Years
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
1
3
5