Analysis of single-cell RNA-seq data from ovarian cancer samples before and after chemotherapy links stress-related transcriptional profile with chemotherapy resistance

Kaiyang Zhang¹, Erdogan Pekcan Erkan¹, Jun Dai¹, Noora Andersson¹, Katja Kaipio², Tarja Lamminen², Naziha Mansuri², Kaisa Huhtinen², Olli Carpén^{1,2,3}, Johanna Hynninen⁴, Sakari Hietanen⁴, Jaana Oikkonen¹, Antti Häkkinen¹, Sampsa Hautaniemi^{1*} & Anna Vähärautio^{1*}

¹*Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland*

²*Cancer Research Unit, Institute of Biomedicine and FICAN West Cancer Centre, University of Turku, Turku, Finland*

³*Department of Pathology, University of Helsinki and HUSLAB, Helsinki University Hospital, Helsinki, Finland*

⁴*Department of Obstetrics and Gynecology, University of Turku and Turku University Hospital, Turku, Finland*

*To whom correspondence should be addressed: Sampsa Hautaniemi (samps.hautaniemi@helsinki.fi) and Anna Vähärautio (anna.vaharautio@helsinki.fi)

Running title: Stress-related profile in chemoresistance in HGSO

Keywords: platinum resistance, transcriptomics, single-cell RNA-seq, high-grade serous ovarian cancer

Competing Interests

The authors declare no potential conflicts of interest.

Translational relevance

We discovered a stress-related transcriptional profile that is significantly enriched in fresh tissue samples after chemotherapy and is significantly associated with poor progression-free survival in an independent patient cohort. The survival association is independent of age, tumor purity or BRCAness. Therefore, this chemotherapy resistance associated profile is intrinsic and could thus be targeted already in treatment-naïve patients. The translation potential of the stress-related transcriptomics profile was further supported by pan-cancer cell line analysis that showed that cell lines with high stress-related transcriptional profile are not affected by platinum, corroborating our results, whereas they were more sensitive to MEK1/2 inhibitors. Taken together, the stress-related transcriptional profile, quantifiable with a set of 35 marker genes, provides a basis for improved prediction of platinum response as well as novel avenues to treat this patient group more effectively.

Abstract

Chemotherapy resistance is the greatest contributor to cancer mortality and the most urgent unmet challenge in oncology. In order to reveal transcriptomics changes due to platinum-based chemotherapy we analyzed single-cell RNA-seq data from fresh tissue samples taken at the time of diagnosis and after neoadjuvant chemotherapy from 11 high-grade serous ovarian cancer (HGSOC) patients. With a novel clustering method accounting for patient-specific variability and technical confounders, we identified 12 clusters. Of these, a stress-related transcriptional profile was enriched during chemotherapy and associated significantly to poor progression-free survival (PFS) and disease-free interval (DFI) in deconvoluted bulk RNA-seq data analysis of treatment-naïve samples in TCGA cohort. Pan-cancer cell line analysis suggests that patients with high stress-related transcriptional profile may benefit from MEK1/2 inhibitors instead of platinum. Further, high proportion of stromal components and high interaction score between tumor and stromal suggest the tumor cells with high-stress profile actively interact with and potentially recruit stromal cells to their microenvironment already prior to chemotherapy, potentially facilitating protection from chemotherapeutic treatments. In summary, we have identified a stress-related transcriptional profile, which is present at the time of diagnosis, enriched during platinum treatment, independent predictor for poor PFS and DFI, and, based on *in vitro* data, targetable with MEK1/2 inhibitors.

Introduction

Chemotherapy resistance is the greatest contributor to cancer mortality and the most urgent unmet challenge in oncology (Holohan et al., 2013; Housman et al., 2014). Platinum-based chemotherapy is prescribed for 10-20% of all cancer patients (NCI, 2014) and presents significant impressive survival advantages, especially in testicular and ovarian cancers (Hanna & Einhorn, 2014; Neijt et al., 1987). Still, the majority of the patients develop resistance to platinum and suffer relapse with limited treatment options. Mechanisms leading to platinum resistance are complex and include platinum uptake and efflux, DNA damage repair, and dysfunctional apoptosis signaling (Galluzzi et al., 2012; Kozłowska et al., 2018). It is, however, poorly understood which of them are active in a patient at relapse, mostly due to paucity of high-quality samples after chemotherapy.

High-grade serous ovarian carcinoma (HGSOC) is the most common and aggressive subtype of ovarian cancer (Koshiyama et al., 2017; Torre et al., 2018). The standard first-line therapy for HGSOC is primary debulking surgery (PDS) followed by platinum-based chemotherapy (Du Bois & Pfisterer, 2005). Platinum-based neoadjuvant chemotherapy (NACT) prior to interval debulking surgery (IDS) is recommended for patients with advanced stage HGSOC when successful PDS is unlikely to be achieved (Wright et al., 2016). Although most HGSOC patients initially respond well to the standard-of-care therapy, the majority develops chemotherapy resistance and relapses leading to a 5-year survival rate of only 43% (Ledermann et al., 2018).

HGSOC is characterized by the universal presence of *TP53* mutations and widespread copy number alterations (CNAs) (Ahmed et al., 2010; Etemadmoghadam et al., 2009). Low frequency of oncogenic mutations and few recurrent CNAs make it difficult to identify targetable genetic lesions (Hoadley et al., 2014; Lheureux et al., 2019). The loss of *TP53* leads to high intratumoral heterogeneity, which is one of the major drivers of therapeutic resistance (Lee et al., 2011; Morris et al., 2016). In addition to genetic heterogeneity, substantial evidence demonstrates that the transcriptional heterogeneity, which is associated with genetic

alterations and other sources of variations, such as epigenetic alterations and tumor microenvironment, plays important roles in cancer progression and treatment resistance (Almendro Navarro et al., 2014; González-Silva et al., 2020). Previous studies have revealed transcriptional subtypes for HGSOC (Chen et al., 2015). However, the analyses using bulk transcriptomes can be severely biased by cell type compositions and have limited ability to capture the heterogeneity within a tumor as multiple subtypes can co-exist in the same tumor. A recent work using single-cell RNA sequencing (scRNA-seq) has characterized six cell subtypes of fallopian tube epithelium and identified a epithelial-mesenchymal transition (EMT)-high subtype of HGSOC with poor prognosis (Hu et al., 2020). However, the transcriptional changes due to chemotherapy have remained elusive.

Here, we performed scRNA-seq on fresh tumor samples collected at the time of diagnosis and after NACT from 11 patients to better understand the tumor cell transcriptional profiles in HGSOC and their changes in response to chemotherapy. We identified 12 subpopulations characterized by 10 distinct gene communities. A pre-existing stress associated subpopulation in pre-NACT samples was found to be expanded after chemotherapy. Further validation using bulk RNA-seq data suggests the stress-related profile independently predicts platinum-based chemotherapy resistance and patient prognosis already in treatment-naïve samples.

Results

Identification of tumor cell clusters in HGSOC

To interrogate the intra-tumor transcriptomic heterogeneity and identify biomarkers related to chemotherapy resistance in HGSOC, we performed scRNA-seq on primary (pre-NACT) - interval (post-NACT) fresh tumor pairs collected from 11 patients (Fig. 1A). Major cell types, i.e., tumor cells, stromal cells and immune cells were identified based on graph-based clustering (Stuart et al., 2019) and acknowledged markers (Fig. S1). As HGSOC is characterized by chromosomal instability and extensive CNAs (Cope et al., 2013; Kuo et al., 2009), we filtered out cells which lack inferred CNAs and have their inferred CNA profiles clustered together with stromal cells. A total of 8,806 tumor cells, with a median of 3677

detected genes per cell, passed our quality controls. The scRNA-seq downstream analysis focused on the tumor cells.

To identify tumor cell clusters in the presence of unwanted sources of variation, including patient-specific variability and technical confounders, e.g., percentages of unique molecular identifier (UMI) counts originating from mitochondrial genes, we separated the expression readouts into nuisance contribution profiles and latent cluster contribution profiles using an iterative Expectation-Maximization (EM) algorithm with multiple random restarts (Fig. 1B, Methods). First we selected the number of clusters to be 12 based on Bayesian information criterion (BIC, Fig. S2) by performing the procedure on up to 25 clusters using 10 restarts. Second, we performed 200 restarts for the selected number of clusters to ascertain a globally good clustering, and the clusters and profiles computed with the highest likelihood were used for subsequent analysis. Three out of 12 clusters (C3, C9, C10) were patient-specific with over 95% of cells from an individual patient, and the majority of cells in C3 were from a primary tumor while C9 and C10 mainly contained cells from interval tumors. All other nine clusters were shared by primary and interval tumors from multiple patients (Fig. 2A-C). To further evaluate the robustness of the cell clusters, we repeated the analysis with another 200 random restarts and the detected clusters were found to be highly consistent with the presented cell clusters (Fig. S3).

Tumor cell subpopulations are characterized by distinct gene communities

To characterize the identified tumor cell clusters and find co-expressed gene communities, we built a gene network using the 4742 genes that were differentially expressed between at least one pair of the identified cell clusters. The significant Pearson correlations ($\rho > 0.8$, $p < 0.01$) of the likelihood-ratio test statistic (LRT) for each gene between all pairs of clusters were defined as the edges of the network. We detected 916 communities in the network via random walks (Pons & Latapy, 2005), and the 10 communities consisting of more than 30 genes per community were kept for further analysis (Fig. 2D).

For each gene community, only the functionally annotated genes with highest coreness centrality were retained. The relations between the identified cell clusters and the gene communities with the corresponding functional enrichment are shown in Fig. S4 and Table S2. Gene communities gc2, gc3 and gc9 were overrepresented in patient-specific clusters, C10, C3, C9, respectively. gc2 was associated with major histocompatibility complex (MHC) class II antigen presentation genes (e.g., *HLA-DPA1*, *HLA-DMA*, *HLA-DOA*, *HLA-DQA1*, *HLA-DRA*), gc3 was enriched in TGF-beta signaling (e.g., *SMAD3*, *ITGA2*, *COL1A2*) and gc9 was related to cytokines and chemokines (e.g., *IL1R2*, *CCL20*, *CXCL1*). Four gene communities (gc5, gc7, gc8, gc10) were shared by multiple cancer cell clusters. They were associated with translation, RNA processing and apoptosis, TCA cycle, and interferon signaling, respectively.

Gene community gc4 was enriched in mucin-type O-linked glycosylation (e.g., *MUC4*, *MUC5B*, *MUC15*, *B3GNT7-8*), which is related to post-translational modifications and protein integrity and has been shown to be involved in tumor growth and progression (Burchell et al., 2018; Gupta et al., 2020).

Gene community gc1 was strongly associated with cell cycle (e.g., *MAD2L1*, *MCM2-7*, *CHEK1*, *HMGB2*) and DNA repair related pathways, including homologous recombination (e.g., *BRCA1/2*, *RAD51*, *POLD1*), Fanconi anemia pathway (e.g., *FANCG*, *FANCI*, *FANCD2*), base excision repair, nucleotide excision repair and mismatch repair pathways. This gene community was overrepresented in cell cluster C5 (Fig. 2A, Fig. S4), which contained on average 14.3% cells from each primary tumor and 3.3% cells from each interval tumor (Fig. 2E). Lower proportion of cycling cells suggests that platinum-taxane chemotherapy has induced cell cycle arrest in the interval samples.

Gene community gc6, consisting of 35 genes, was highly enriched with early response genes (e.g., *CEBPB*, *FOS*, *JUN*, *DDIT3*) and pro-survival genes (e.g., *GADD45B* and *GADD45G*). Cell cluster C7 was characterized by gc6, and we observed increased proportions of interval tumor cells in C7 compared to primary tumor cells ($p = 0.002$, Fig. 2E). We further defined a stress score as the gene set

enrichment (ssGSEA) score (Barbie et al., 2009; Subramanian et al., 2005) for gc6 in individual cells or samples. Eight out of 11 patients showed a higher stress score in the interval tumor than in the primary tumor (Fig. 3A). To investigate whether the increase of stress score after chemotherapy is unique to the scRNA-seq data or conserved in HGSOC bulk RNAseq data, we computed the stress score based on their deconvoluted cancer cell specific profile for 23 HGSOC bulk samples including 18 primary-interval pairs and eight primary-relapse pairs. Consistently, the interval tumors showed significantly higher stress scores compared to their paired primary tumors in the bulk samples ($p = 0.024$, Fig. 3B left). Notably, higher stress scores were also observed in relapse tumors ($p = 0.0078$, Fig. 3B right). Even when accounting for anatomical sites, the stress score varied significantly between primary and interval ($p = 1.45e-07$ for partial rank correlation), primary and relapse ($p = 0.067$ for partial rank correlation) tumors. The results from the bulk RNA-seq show that the relevance of the stress-related transcriptional profile in chemotherapy resistance is not an artifact due to single-cell dissociation.

The stress-related gene community is associated with poor prognosis

We next turned to the bulk HGSOC RNAseq dataset from The Cancer Genome Atlas (TCGA) (Network & The Cancer Genome Atlas Research Network, 2011) and tested the association between the stress scores and patient outcome. The cell type specific proportion and expression profiles were estimated using PRISM, a latent statistical framework to simultaneously extract the sample composition and cell type specific expression profiles (Häkkinen et al., 2019). After deconvolution, the stress scores were computed using cancer cell specific expression profiles. We identified 86 stress-high tumors that were significantly enriched with the stress gene community gc6 ($p < 0.05$) and 145 stress-low tumors ($p > 0.5$). Patients with stress-high tumors had significantly shorter progression free survival (PFS, $p = 0.0037$ in log-rank test) and disease-free interval (DFI, $p = 0.015$ in log-rank test) than the patients with stress-low tumors (Fig. 3C, Fig. S5).

HGSOC patients with COSMIC signature 3 (BRCA signature) are associated with dysfunctional *BRCA1/2* and better response to chemotherapy (Patch et al., 2015). Therefore, we investigated whether the lack of signature 3 could explain the poor

prognosis of the stress-high patients. We found that signature 3 status is not enriched in stress-high or low patients (Fig. 3E, Fisher exact test $p = 0.3133$). Multivariate Cox regression analysis showed that the stress score was significantly associated with PFS ($p = 0.0051$, Fig. 3D) and DFI ($p = 0.0339$, Fig. S5B) independently from the effect of age, tumor purity and signature 3 status. We stratified both stress status and signature 3 status into four groups: (i) stress-high & signature 3 negative (ii) stress-low & signature 3 negative (iii) stress-high & signature 3 positive and (iv) stress-low & signature 3 positive. Short PFS and DFI in stress-high patients were observed in both signature3 positive and signature3 negative groups (Fig. 3F). As TCGA has only primary tumors, these results suggest that the stress-related transcriptional profile pre-exists in the untreated tumors. Our results show that the stress score can identify poor-prognostic HGSOc patients independently of the BRCA signature.

High-stress tumors show an altered microenvironment

Cellular stress has a pivotal role in remodeling the tumor microenvironment (Chang et al., 2017; Koumenis et al., 2013). Accordingly, we tested for an association between the stress scores and the microenvironment compositions in the TCGA cohort. We found significantly higher proportions of stromal ($p = 2.4e-09$) and immune ($p = 3e-04$) components in the stress-high tumors (Fig. 4A).

We further analyzed the cell type specific expression profiles between stress-high and stress-low tumors. Key pro-inflammatory cytokines (e.g., *IL1B*, *IL6* and *TNF*), genes involved in EMT (e.g., *SERPINE1*, *SNAI1* and *SNAI2*), pro-angiogenic cytokines (e.g., *HBEGF*, *CTGF*, *CYR61*) and pro-survival factors (e.g., *GADD45A*, *GADD45B*, *MCL1*) were upregulated in tumor, stromal and immune components of high-stress tumors (Fig. 4B). This suggests that tumor cells with high stress-related profile are more active in recruiting and modulating their surrounding cells to gain survival advantages through cytokine-mediated signaling pathways.

To further characterize the communication between tumor cells and the microenvironment in stress-high tumors, we used an interaction score defined as the product of the ligand expression in one cell type and the corresponding receptor

expression in another cell type to quantify the activity of 2,597 previously collected ligand-receptor interactions (Kumar et al., 2018; Wang et al., 2019). The interactions with the highest increase in stress-high tumors were related to the pro-inflammatory, EMT and pro-survival cytokines. We also observed significantly higher scores for interactions related to laminins (*LAMA2*, *LAMA4*, *LAMB3* and *LAMC2*) that bind to integrin receptors and collagens (*COL5A2*, *COL11A1*) binding to *DDR1* in stress-high tumors (Fig. 4C).

To confirm the occurrence of the stress-related profile at the protein level, we analyzed the TCGA HGSOC reverse-phase protein array (RPPA) data, which is available for 224 proteins in 60 stress-high and 115 stress-low tumors. The phosphorylated c-Jun (CJUN_pS73, $p = 0.0035$) and its upstream kinase phospho-JNK (JNK_pT183Y185, $p = 0.00077$) were increased in the stress-high tumors (Fig. 4D). Moreover, we observed higher levels of phosphorylated p38- α (P38_pT180Y182, $p = 0.017$, Fig. 4D), supporting the induction of JNK and p38 MAPK signaling pathways also on the protein level in stress-high tumors.

Identification of drugs potentially targeting the stress-high tumors

To identify drugs that can effectively target the stress-high tumor cells *ex vivo*, we obtained the gene expression and GDSC2 drug sensitivity datasets from the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2013). Among the 1,018 cell lines with gene expression data available, we identified 233 stress-high cell lines spanning various cancer types (Fig. 5A).

We compared the drug response for 192 compounds between stress-high and stress-low cell lines, consistent with our analysis of patient samples (Fig. 3), significantly increased half-maximal inhibitory concentration (IC50) of cisplatin were observed in stress-high cell lines ($\log_2$ fold-change = 1.09, $p = 4.19\text{e-}11$, Fig. 5B). Notably, the stress-high cell lines were more sensitive to MEK1/2 inhibitors, Trametinib ($\log_2$ fold-change = -0.79, $p = 3.63\text{e-}05$, Fig. 5B) and Selumetinib ($\log_2$ fold-change = -0.68, $p = 8.49\text{e-}04$, Fig. 5B) with significantly lower IC50.

Considering the cancer type specific variability, we next quantified the correlation between stress scores and $\ln\text{IC}_{50}$ across the 12 ovarian cancer cell lines harboring

TP53 mutations. In this relatively small subset, we found a significant negative correlation between the stress score and the InIC_{50} of the IRAK4 inhibitor (IRAK4_4710, $R = -0.66$, $p = 0.02$, Fig. 5C). IRAK4 has been reported to play an important role in IL1-induced JNK and p38 activation (Rhyasen & Starczynowski, 2015; Suzuki et al., 2002).

Discussion

We analyzed fresh tissue samples taken before and after chemotherapy from 11 HGSOC patients and subjected to scRNA-seq to identify molecular processes that drive resistance to platinum-taxane combination therapy. The scRNA-seq results were validated with and expanded to bulk RNA-seq data from 23 HGSOC patients as well as with 272 HGSOC patients available in TCGA.

By using a novel clustering method accounting for patient-specific variability and technical confounders, we identified 10 distinct transcriptional profiles, of which seven were shared across different patients. Our results reveal a stress-related transcriptional profile, which is induced by chemotherapy and related to poor prognosis in treatment-naïve patients. The identified stress-related transcriptional profile can be characterized by a set of only 35 marker genes, providing a basis for identification of poor-prognostic patients as well as developing new treatment strategies. Importantly, our findings imply that the expression pattern is associated with both intrinsic and adaptive resistance, and could be targeted in both treatment-naïve patients predicted to have poor response, and following chemotherapy to minimize residual disease. Analysis of GDSC2 drug sensitivity data sets suggests IRAK4 or MEK1/2 inhibitors as potential drugs to target the cisplatin resistant stress-high subpopulations in tumors. These results open interesting avenues for mechanistic studies.

Increasing evidence shows that stress inducible genes and pathways can be activated in response to the selective microenvironment forces, such as hypoxia, nutrient deprivation and DNA damage, to help cancer cell survival and progression (Avril et al., 2017; Tsuruo et al., 2003). Upregulation of stress induced immediate early genes (IEGs) *FOS* and *ATF3* has been reported to be associated with

increasing resistance to chemotherapeutic agents (Chang et al., 2017; Mukhopadhyay et al., 2001). Baron and colleagues showed that the stress-like cancer cells confer resistance to MEK/BRAF inhibitors in melanoma (Baron et al., 2020). However, most of the studies are based on laboratory models instead of human tumors. The stressed-related profile has been found in other single cell studies in muscle stem cells (van den Brink et al., 2017) and HGSOc (Hu et al., 2020). Hu et al. suggest several genes belonging to the stress-related transcriptional profile to be artifacts due to sample preparation for scRNA-seq experiments. We tested the presence of the stress-related transcriptional profile in bulk RNA-seq samples from HGSOc patients' fresh tumors without single-cell dissociation in sample handling and were able to identify them as predicted from scRNA-seq data. This suggests that, at least in our data set, the stress-related transcriptional profile is not a dissociation artifact.

Tumors with high stress scores were found to have higher proportions of stromal and immune components and importantly, the association between stress score to PFS and DFI was independent of tumor purity. Ligand-receptor analysis showed the genes enriched in high-stress tumors contained a high number of ligands, and to lesser extent receptors, for the tumor microenvironment interactions to mainly stromal cells. This suggests that tumor cells with high stress-related profile actively interact with especially stromal cells and potentially recruit them to their microenvironment already prior to chemotherapy, thus facilitating protection from chemotherapeutic treatments.

Materials and Methods

Human subjects

All patients participating in the study provided written informed consent. The study and the use of all clinical material have been approved by The Ethics Committee of the Hospital District of Southwest Finland (ETMK) under decision number EMTK: 145/1801/2015.

scRNA-seq sample preparation

Fresh HGSOC tumor specimens were collected from 11 patients at the time of laparoscopy and interval debulking surgery (IDS). Detailed clinical information is shown in Table 1. Immediately after surgery, the specimens were incubated overnight in a mixture of collagenase and hyaluronidase (Department of Pathology, University of Turku) to obtain single cell suspensions, that were then processed with the standard Chromium Single Cell 3' Reagent Kit v. 2.0 (10x Genomics) protocol for scRNA-seq with Illumina HiSeq4000 (Jussi Taipale Lab, Karolinska Institute and Institute for Molecular Medicine Finland).

Preprocessing scRNA-seq data

The Cell Ranger Software Suite (Version 3.1.0) was used to perform sample de-multiplexing, alignment, barcode processing and UMI quantification. The reference index was built upon the GRCh38.d1.vd1 reference genome with GENCODE v25 annotation. Shared nearest neighbor (SNN) modularity optimization based clustering from Seurat v3 (Stuart et al., 2019) were used for initial clustering. Three major cell types were revealed based on acknowledged markers: epithelial tumor cells (*WFDC2*, *PAX8*, *EPCAM*), stromal cells (*COL1A2*, *FGFR1*, *DCN*), and immune cells (*CD79A*, *FCER1G*, *PTPRC*). Cells expressing any combinations of *PAX8*, *DCN* and *PTPRC* were excluded to remove potential doublets. We focused on tumor cells in the subsequent analysis.

The quality measures of each tumor cell were quantified using Seurat. We estimated the cutoffs for each quality measure based on its bimodal distribution and then used four criteria for quality control: (1) the number of reads above 8,192; (2) the number of UMI counts above 4,705; (3) the number of detected genes above 1,552; (4) the percentage of UMI counts originating from mitochondrial genes below 12. Next, we filtered out tumor cells that have their CNA profiles clustered together with stromal cells. The CNAs were inferred using inferCNV with default settings (inferCNV of the Trinity CTAT Project. <https://github.com/broadinstitute/inferCNV>). We used an agglomerative hierarchical clustering algorithm to cluster the inferred CNA profiles and the tree was cut based on Bayesian Information Criterion (Häkkinen et al., 2019).

Modeling and clustering scRNA-seq data

We modeled the observed single cell expression profiles as a mixture of latent transcriptional profiles and nuisance expression profiles following a Poisson distribution:

$$Y_{j,i} \sim \text{Poisson} \left(\left(\sum_{l=1}^r (X_{j,l} D_{l,i}) + \sum_{c=1}^k (Z_{j,c} C_{c,i}) \right) G_i \right), \quad \text{where}$$

$l = 1, 2, \dots, r$ runs over r nuisance factors, $c = 1, 2, \dots, k$ runs over k latent transcriptional clusters. $Y_{j,i}$ denotes the observed UMI counts of gene j in the i^{th} cell, $X_{j,l}$ denotes the expression profile centroid of gene j specific to nuisance factor l , $D_{l,i}$ denotes the design coefficient of the l^{th} nuisance factor in the i^{th} cell, $Z_{j,c}$ denotes the cluster c expression profile centroid at gene j , $C_{c,i} \in \{0, 1\}$ is an indicator of whether the i^{th} cell belongs to the cluster c , and G_i is a cell-specific scaling factor.

Given the observations $Y_{j,i}$, known nuisances $D_{l,i}$, known scaling factors G_i and the number of latent clusters k , we can estimate the latent nuisance expression centroids $X_{j,l}$, latent expression centroids $Z_{j,c}$ and the latent cluster memberships $C_{c,i}$ using an Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm is constructed on the latent variables $Z_{X_{j,l,i}} \sim \text{Poisson}(X_{j,l} D_{l,i} G_i)$ and $Z_{Z_{j,c,i}} \sim \text{Poisson}(Z_{j,c} C_{c,i} G_i)$, which are the nuisance and cleaned contributions to the expression, respectively. The parameter set $\theta = (X_{j,l}, Z_{j,c}, C_{c,i})$ was estimated in two stages: first, the expression centroids $X_{j,l}$ and $Z_{j,c}$ can be estimated given $Y_{j,i}$, $D_{l,i}$, $C_{c,i}$ and G_i ; second, the cluster membership $C_{c,i}$ can be updated given $Y_{j,i}$, $X_{j,l}$, $D_{l,i}$, $Z_{j,c}$ and G_i (see Supplementary material for the details).

To select the optimal k , we fitted our model for $k = 1, 2, \dots, 25$ with 10 different random initial parameter sets for each k , and $k = 12$ was selected based on Bayesian Information Criterion (BIC, Fig. S2). We then ran the EM procedure with 200 random initializations for $k = 12$, the maximum likelihood estimates of $X_{j,l}$, $Z_{j,c}$ and the resulting $C_{c,i}$ were used for downstream analysis.

Differential expression (DE) analysis for scRNA-seq data

Let $Y_{j,i}$ denote the observed UMI count of gene j in the i^{th} cell, $\mu_{j,i} = \left(\sum_{l=1}^r (X_{j,l} D_{l,i}) + \sum_{c=1}^k (Z_{j,c} C_{c,i}) \right) G_i$ denote the predicted mean expression rate of gene j for the i^{th} cell based on the estimated model parameters $X_{j,l}$, $Z_{j,c}$, $C_{c,i}$. The differential expression between group g_1 and group g_2 for the gene j can be assessed by testing the alternative hypothesis $H_A : Z_{j,g_1} \neq Z_{j,g_2}$ against the null hypothesis $H_0 : Z_{j,g_1} = Z_{j,g_2}$. For the former, the likelihood is that attained at the maximum-likelihood estimate (MLE) $\hat{\mu}_{j,i}$, while for the latter, the model is refitted giving $\bar{\mu}_{j,i}$, the MLE under H_0 . The logarithmic likelihood-ratio test statistic (LRT) for gene j is

$$LRT_j = \sum_{i \in I_1 \cup I_2} (Y_{j,i} \log \hat{\mu}_{j,i} - \hat{\mu}_{j,i}) - \sum_{i \in I_1 \cup I_2} (Y_{j,i} \log \bar{\mu}_{j,i} - \bar{\mu}_{j,i})$$

Where I_1 and I_2 denote the indices for the samples in g_1 and g_2 , respectively. A p-value for the gene j to be differentially expressed between group g_1 and group g_2 can be computed as the probability to the right of the $-2LRT_j$ for the chi-squared distribution with degrees of freedom equals to the difference in number of parameters, i.e. 1.

Identification of co-expressed gene communities

Co-expressed gene communities were identified as follows: (1) we conducted DE analysis between each pair of cell clusters, resulting in 66 comparisons. The top 1000 high LRT genes with FDR < 0.01 were selected in each comparison, and this resulted in a total of 4742 genes. (2) the Pearson correlations between these 4742 genes were computed using the LRTs from all 66 comparisons. Correlations with $\rho > 0.8$ and p value < 0.01 were used to build a gene network. (3) we detected 916 communities in the network via random walks with step equals to 3 (Pons & Latapy, 2005), and the 10 communities consisting more than 30 genes were retained. (4) let V be the genes in a community, c_j be the coreness of gene j and n_{max} be the number of genes with the maximum coreness ($\max_{j \in V} c_j$, degeneracy) in that community. If $n_{max} > 30$, then the genes with $c = \max_{j \in V} c_j$ were retained, otherwise, we retained the top 30 genes ranked by coreness. (5) gene set

over-representation analysis was performed for the remaining genes in each community using the ConsensusPathDB (Kamburov et al., 2013). We further reduced the redundancy of each gene community with number of genes above 20 by applying the following filters: (i) only genes overlapped with significantly overrepresented gene sets ($FDR < 0.05$, $size < 500$) were kept; (ii) biclustering was applied to the binary matrix of the presence/absence of each gene in each significantly overrepresented gene set using the R package blockcluster (Bhatia et al., 2017), and the gene clusters that have less than 3% presence in any of the gene-set clusters were excluded (Fig. S6). After filtering, the number of genes per community varied between 11 and 106. The genes in each community are listed in Table S1.

Defining stress scores

We defined the stress score as the gene set enrichment score of our identified stress-associated gene community (gc6) in individual cells and samples, which was computed using ssGSEA (Barbie et al., 2009; Subramanian et al., 2005). Samples with permutation test p-value below 0.05 were considered stress-high while samples with p-value above 0.5 were considered stress-low.

Bulk tumor expression data

We acquired 18 primary-interval sample pairs and eight primary-relapse sample pairs from 23 patients in the HERCULES cohort (<http://www.project-hercules.eu/>). The sample collection, data quality control, alignment, quantification were performed as we have previously described (Häkkinen et al., 2019). Briefly, read pairs were trimmed using Trimmomatic (Version 0.33, (Bolger et al., 2014)) and aligned to the GRCh38.d1.vd1 reference genome with GENCODE v25 annotation using STAR (Version 2.5.2b) (Dobin et al., 2013). Gene level effective counts were quantified using eXpress (Version 1.5.1-linux x86 64) (Roberts & Pachter, 2013).

TCGA RNA-Seq data of ovarian serous cystadenocarcinoma (OV, `illuminahtseq_rnaseqv2-RSEM_genes_normalized`) was downloaded from the Broad Firehose (<https://gdac.broadinstitute.org/>), along with the clinical annotations. The

primary tumors from 272 patients with advanced HGSOC (grade: G2-4, stage: IIIA-IV) were included in our analysis.

The proportions of tumor, stromal and immune components and the cell type specific expression profiles for HERCULES and TCGA samples were estimated using PRISM (Häkkinen et al., 2019).

TCGA reverse phase protein array (RPPA) data

The RPPA data (replicates-based normalization) for TCGA ovarian serous cystadenocarcinoma samples (TCGA-OV-L4) was downloaded from the Cancer Proteomics Atlas (TCPA, <https://www.tcportal.org/tcpa/download.html>).

Cell line expression and drug sensitivity data

The GDSC2 drug sensitivity data and the Robust Multichip Average (RMA) normalized expression data were downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC) ftp server (<ftp.sanger.ac.uk/pub4/cancerrxgene/releases/>).

Data and code availability

Raw scRNA-Seq expression data will be available at The European Genome-phenome Archive (EGA). The codes for analyses are available at GitHub (<https://github.com/KaiyangZ/scRNAseq-HGSOC>).

REFERENCES

- Ahmed, A. A., Etemadmoghadam, D., Temple, J., Lynch, A. G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., Defazio, A., Bowtell, D., & Brenton, J. D. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *The Journal of Pathology*, 221(1), 49–56.
- Almendro Navarro, V., Cheng, Y.-K., Randles, A., Itzkovitz, S., Marusyk, A., Ametller, E., Gonzalez-Farre, X., Muñoz, M., Russnes, H. G., Helland, Å., & Others. (2014). Inference of Tumor Evolution during Chemotherapy by Computational Modeling and In Situ Analysis of Genetic and Phenotypic Cellular Diversity. *Cell*

Reports, 2014, Vol. 6, Num. 3, P. 514-527.

<http://diposit.ub.edu/dspace/handle/2445/101862>

Avril, T., Vauléon, E., & Chevet, E. (2017). Endoplasmic reticulum stress signaling and chemotherapy resistance in solid cancers. *Oncogenesis*, 6(8), e373.

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., ... Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462(7269), 108–112.

Baron, M., Tagore, M., Hunter, M. V., Kim, I. S., Moncada, R., Yan, Y., Campbell, N. R., White, R. M., & Yanai, I. (2020). *A population of stress-like cancer cells in melanoma promotes tumorigenesis and confers drug resistance.*

<https://doi.org/10.1101/396622>

Bhatia, P. S., Iovleff, S., & Govaert, G. (2017). blockcluster: An R Package for Model-Based Co-Clustering. In *Journal of Statistical Software* (Vol. 76, Issue 9).

<https://doi.org/10.18637/jss.v076.i09>

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.

Burchell, J. M., Beatson, R., Graham, R., Taylor-Papadimitriou, J., & Tajadura-Ortega, V. (2018). O-linked mucin-type glycosylation in breast cancer. *Biochemical Society Transactions*, 46(4), 779–788.

Chang, Y. S., Jalgaonkar, S. P., Middleton, J. D., & Hai, T. (2017). Stress-inducible gene Atf3 in the noncancer host cells contributes to chemotherapy-exacerbated breast cancer metastasis. In *Proceedings of the National Academy of Sciences*

(Vol. 114, Issue 34, pp. E7159–E7168).

<https://doi.org/10.1073/pnas.1700455114>

Chen, P., Huhtinen, K., Kaipio, K., Mikkonen, P., Aittomäki, V., Lindell, R., Hynninen, J., Auranen, A., Grénman, S., Lehtonen, R., Carpén, O., & Hautaniemi, S. (2015). Identification of Prognostic Groups in High-Grade Serous Ovarian Cancer Treated with Platinum-Taxane Chemotherapy. *Cancer Research*, 75(15), 2987–2998.

Cope, L., Wu, R.-C., Shih, I.-M., & Wang, T.-L. (2013). High level of chromosomal aberration in ovarian cancer genome correlates with poor clinical outcome. *Gynecologic Oncology*, 128(3), 500–505.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. In *Journal of the Royal Statistical Society: Series B (Methodological)* (Vol. 39, Issue 1, pp. 1–22).

<https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.

Du Bois, A., & Pfisterer, J. (2005). Future options for first-line therapy of advanced ovarian cancer. *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society*, 15 Suppl 1, 42–50.

Etemadmoghadam, D., deFazio, A., Beroukhi, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M. B., Harnett, P., Lade, S., Akslen, L. A., Tinker, A. V., Locandro, B., Alsop, K., Chiew, Y.-E., Traficante, N., Fereday, S., Johnson, D., ... AOCs Study Group. (2009). Integrated genome-wide DNA copy

- number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 15(4), 1417–1427.
- Galluzzi, L., Senovilla, L., Vitale, I., Michels, J., Martins, I., Kepp, O., Castedo, M., & Kroemer, G. (2012). Molecular mechanisms of cisplatin resistance. *Oncogene*, 31(15), 1869–1883.
- González-Silva, L., Quevedo, L., & Varela, I. (2020). Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies. *Trends in Cancer Research*, 6(1), 13–19.
- Gupta, R., Leon, F., Rauth, S., Batra, S. K., & Ponnusamy, M. P. (2020). A Systematic Review on the Implications of O-linked Glycan Branching and Truncating Enzymes on Cancer Progression and Metastasis. *Cells*, 9(2). <https://doi.org/10.3390/cells9020446>
- Häkkinen, A., Zhang, K., Alkods, A., Andersson, N., Erkan, E. P., Dai, J., Kaipio, K., Lamminen, T., Mansuri, N., Huhtinen, K., Vähärautio, A., Carpen, O., Hynninen, J., Hietanen, S., Lehtonen, R., & Hautaniemi, S. (2019). *PRISM: Recovering cell type specific expression profiles from composite RNA-seq data*. <https://doi.org/10.1101/854505>
- Hanna, N., & Einhorn, L. H. (2014). Testicular cancer: a reflection on 50 years of discovery. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 32(28), 3085–3092.
- Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., Leiserson, M. D. M., Niu, B., McLellan, M. D., Uzunangelov, V., Zhang, J., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A. A., Van't

- Veer, L. J., Lopez-Bigas, N., ... Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4), 929–944.
- Holohan, C., Van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature Reviews. Cancer*, 13(10), 714–726.
- Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., & Sarkar, S. (2014). Drug resistance in cancer: an overview. *Cancers*, 6(3), 1769–1792.
- Hu, Z., Artibani, M., Alsaadi, A., Wietek, N., Morotti, M., Shi, T., Zhong, Z., Santana Gonzalez, L., El-Sahhar, S., KaramiNejadRanjbar, M., Mallett, G., Feng, Y., Masuda, K., Zheng, Y., Chong, K., Damato, S., Dhar, S., Campo, L., Garruto Campanile, R., ... Ahmed, A. A. (2020). The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells. *Cancer Cell*, 37(2), 226–242.e7.
- Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. In *Nucleic Acids Research* (Vol. 41, Issue D1, pp. D793–D800). <https://doi.org/10.1093/nar/gks1055>
- Koshiyama, M., Matsumura, N., & Konishi, I. (2017). Subtypes of Ovarian Cancer and Ovarian Cancer Screening. In *Diagnostics* (Vol. 7, Issue 1, p. 12). <https://doi.org/10.3390/diagnostics7010012>
- Koumenis, C., Hammond, E., & Giaccia, A. (2013). *Tumor Microenvironment and Cellular Stress: Signaling, Metabolism, Imaging, and Therapeutic Targets*. Springer Science & Business Media.

- Kozłowska, E., Färkkilä, A., Vallius, T., Carpén, O., Kemppainen, J., Grénman, S., Lehtonen, R., Hynninen, J., Hietanen, S., & Hautaniemi, S. (2018). Mathematical Modeling Predicts Response to Chemotherapy and Drug Combinations in Ovarian Cancer. *Cancer Research*, 78(14), 4036–4044.
- Kumar, M. P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D. C., Lauffenburger, D. A., & Raue, A. (2018). Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Reports*, 25(6), 1458–1468.e4.
- Kuo, K.-T., Guan, B., Feng, Y., Mao, T.-L., Chen, X., Jinawath, N., Wang, Y., Kurman, R. J., Shih, I.-M., & Wang, T.-L. (2009). Analysis of DNA copy number alterations in ovarian serous tumors identifies new molecular genetic changes in low-grade and high-grade carcinomas. *Cancer Research*, 69(9), 4036–4042.
- Lee, A. J. X., Endesfelder, D., Rowan, A. J., Walther, A., Birkbak, N. J., Futreal, P. A., Downward, J., Szallasi, Z., Tomlinson, I. P. M., Howell, M., Kschischo, M., & Swanton, C. (2011). Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research*, 71(5), 1858–1870.
- Lheureux, S., Gourley, C., Vergote, I., & Oza, A. M. (2019). Epithelial ovarian cancer. *The Lancet*, 393(10177), 1240–1253.
- Morris, L. G. T., Riaz, N., Desrichard, A., Şenbabaoğlu, Y., Hakimi, A. A., Makarov, V., Reis-Filho, J. S., & Chan, T. A. (2016). Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget*, 7(9), 10051–10063.
- Mukhopadhyay, A., Bueso-Ramos, C., Chatterjee, D., Pantazis, P., & Aggarwal, B. B. (2001). Curcumin downregulates cell survival mechanisms in human prostate

cancer cell lines. *Oncogene*, 20(52), 7597–7609.

NCI (National Cancer Institute): <https://www.cancer.gov/research/progress/discovery/cisplatin>, 2014.

Neijt, J. P., ten Bokkel Huinink, W. W., van der Burg, M. E., van Oosterom, A. T., Willemse, P. H., Heintz, A. P., van Lent, M., Trimpos, J. B., Bouma, J., & Vermorken, J. B. (1987). Randomized trial comparing two combination chemotherapy regimens (CHAP-5 v CP) in advanced ovarian carcinoma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 5(8), 1157–1168.

Network, T. C. G. A. R., & The Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. In *Nature* (Vol. 474, Issue 7353, pp. 609–615). <https://doi.org/10.1038/nature10166>

Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P. J., Kassahn, K. S., Newell, F., Quinn, M. C. J., Kazakoff, S., Quek, K., Wilhelm-Benartzi, C., Curry, E., Leong, H. S., Australian Ovarian Cancer Study Group, ... Bowtell, D. D. L. (2015). Corrigendum: Whole-genome characterization of chemoresistant ovarian cancer. *Nature*, 527(7578), 398.

Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In *Computer and Information Sciences - ISCIS 2005* (pp. 284–293). https://doi.org/10.1007/11569596_31

Rhyasen, G. W., & Starczynowski, D. T. (2015). IRAK signalling in cancer. *British Journal of Cancer*, 112(2), 232–237.

Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time

- analysis of sequencing experiments. *Nature Methods*, 10(1), 71–73.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888–1902.e21.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), 15545–15550.
- Suzuki, N., Suzuki, S., Duncan, G. S., Millar, D. G., Wada, T., Mirtsos, C., Takada, H., Wakeham, A., Itie, A., Li, S., Penninger, J. M., Wesche, H., Ohashi, P. S., Mak, T. W., & Yeh, W.-C. (2002). Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4. *Nature*, 416(6882), 750–756.
- Torre, L. A., Trabert, B., DeSantis, C. E., Miller, K. D., Samimi, G., Runowicz, C. D., Gaudet, M. M., Jemal, A., & Siegel, R. L. (2018). Ovarian cancer statistics, 2018. In *CA: A Cancer Journal for Clinicians* (Vol. 68, Issue 4, pp. 284–296). <https://doi.org/10.3322/caac.21456>
- Tsuruo, T., Naito, M., Tomida, A., Fujita, N., Mashima, T., Sakamoto, H., & Haga, N. (2003). Molecular targeting therapy of cancer: drug resistance, apoptosis and survival signal. In *Cancer Science* (Vol. 94, Issue 1, pp. 15–21). <https://doi.org/10.1111/j.1349-7006.2003.tb01345.x>
- van den Brink, S. C., Sage, F., Vértessy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., Robin, C., & van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature*

Methods, 14(10), 935–936.

- Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., Wang, M., Ajani, J., Futreal, A., & Wang, L. (2019). *iTALK: an R Package to Characterize and Illustrate Intercellular Communication*. <https://doi.org/10.1101/507871>
- Wright, A. A., Bohlke, K., Armstrong, D. K., Bookman, M. A., Cliby, W. A., Coleman, R. L., Dizon, D. S., Kash, J. J., Meyer, L. A., Moore, K. N., Olawaiye, A. B., Oldham, J., Salani, R., Sparacio, D., Tew, W. P., Vergote, I., & Edelson, M. I. (2016). Neoadjuvant chemotherapy for newly diagnosed, advanced ovarian cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology Clinical Practice Guideline. In *Gynecologic Oncology* (Vol. 143, Issue 1, pp. 3–15). <https://doi.org/10.1016/j.ygyno.2016.05.022>
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., Ramaswamy, S., Futreal, P. A., Haber, D. A., Stratton, M. R., Benes, C., McDermott, U., & Garnett, M. J. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(Database issue), D955–D961.

Acknowledgements

This work is supported financially by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 667403 for HERCULES (Comprehensive Characterization and Effective Combinatorial Targeting of High-Grade Serous Ovarian Cancer via Single-Cell Analysis), the Academy of Finland (Projects No. 292402, 325956, 314395, 289059, 322927, 294023 and 319243), the Sigrid Jusélius Foundation, and the Cancer Foundation Finland. We thank CSC-IT Center for Science Ltd. for compute resources. The results published here are in part based upon data generated by TCGA managed by the NCI and NHGRI. Information about TCGA can be found at <https://cancergenome.nih.gov/>. Single-cell RNA sequencing was performed with Illumina HiSeq4000 at Jussi Taipale Lab, Karolinska Institute and Institute for Molecular Medicine Finland.

Table1. Patient and sample information.

Patient ID	Age ¹	Treatment	Stage ²	PFI (days)	Anatomical locations ³				
					scRNA-seq Primary	scRNA-seq Interval	Bulk RNA-seq Primary	Bulk RNA-seq Interval	Bulk RNA-seq Relapse
EOC1	73	NACT	IVA	65	Peritoneum	Tumor	NA	NA	NA
EOC2	64	NACT	IVA	520	Mesentery	Omentum	NA	NA	NA
EOC3	78	NACT	IVA	393	Omentum	Omentum	NA	NA	NA
EOC4	74	NACT	IVA	230	Omentum	Omentum	NA	NA	NA
EOC5	67	NACT	IVA	14	Peritoneum	Omentum	Peritoneum	Omentum	NA
EOC6	67	NACT	IVB	36	Peritoneum	Omentum	NA	NA	NA
EOC7	68	NACT	IIIC	460	Peritoneum	Peritoneum	Peritoneum	Peritoneum	NA
EOC8	54	NACT	IVA	177	Omentum	Omentum	Omentum	Omentum	NA
EOC9	62	NACT	IIIC	126	Omentum	Omentum	NA	NA	NA
EOC10	72	NACT	IVA	83	Peritoneum	Omentum	NA	NA	NA
EOC11	62	NACT	IIIC	30	Peritoneum	Omentum	Omentum	Omentum	NA
EOC12	75	NACT	IIIC	210	NA	NA	Omentum	Mesentery	NA
EOC13	68	PDS	IVB	648	NA	NA	Omentum	NA	Mesentery
EOC14	68	NACT	IIIC	203	NA	NA	Omentum	Omentum	NA
EOC15	71	NACT	IIIC	974	NA	NA	Omentum	Omentum	NA
EOC16	72	NACT	IIIC	0	NA	NA	Omentum	Omentum	Ascites
EOC17	67	NACT	IIIC	280	NA	NA	Omentum	Omentum	NA
EOC18	81	NACT	IIIC	721*	NA	NA	Omentum	Ovary	NA
EOC19	57	NACT	IVA	210	NA	NA	Mesentery	Omentum	NA
EOC20	71	NACT	IVB	27	NA	NA	Peritoneum	Tumor	NA
EOC21	77	NACT	IVB	511	NA	NA	Peritoneum	Omentum	NA
EOC22	68	NACT	IIIC	81	NA	NA	Peritoneum	Peritoneum	Ascites
EOC23	74	NACT	IIIC	91	NA	NA	Adnexa	Ascites	NA
EOC24	71	NACT	IIIC	183	NA	NA	Omentum	NA	Ascites
EOC25	74	NACT	IIIC	632	NA	NA	Peritoneum	Mesentery	NA
EOC26	62	NACT	IIIC	285	NA	NA	Peritoneum	Omentum	Peritoneum
EOC27	66	NACT	IVB	661	NA	NA	Peritoneum	Omentum	NA
EOC28	75	NACT	IIIC	19	NA	NA	Omentum	NA	Ascites
EOC29	66	NACT	IIIC	221	NA	NA	Peritoneum	NA	Ascites
EOC30	64	NACT	IIIC	174	NA	NA	Peritoneum	NA	Ascites

¹ Age at diagnosis, years

² Tumor staging was done according to the International Federation of Gynecology and Obstetrics (FIGO) 2014 guidelines

³ Anatomical locations from which the samples were collected for analyses.

* No progression, PFI at outcome update

Abbreviations: NACT: Neoadjuvant chemotherapy; PDS: Primary debulking surgery; PFI: Platinum-free interval, days; NA: Not available

Figure legends

Figure 1.

Overall workflow of the sample preparation and data analysis. **A**, Diagram shows the sample collection and processing. We collected fresh tumor samples from 11 high-grade serous ovarian cancer (HGSOC) patients before and after neoadjuvant chemotherapy (NACT). scRNA-seq was performed using the 10x Genomics Chromium platform. **B**, Schematic of our approach to cluster cells and detect co-expressed gene communities. The observed single cell expression profiles (Y) is a mixture of latent transcriptional profiles and nuisance contributions. Given the observations Y , the known nuisance factors D , known size factors G and the number of latent clusters k , we can estimate the latent nuisance contributions X , latent transcriptional profiles Z and the latent cluster memberships C using an Expectation-Maximization (EM) algorithm.

Figure 2.

Identification of 10 gene communities characterizing 12 subpopulations of HGSOC tumor cells. **A**, Heatmap of the expression of the 10 distinct gene communities in the 12 identified cell clusters. Rows correspond to genes and columns to cells. The color bar on the top denotes the patients. **B**, The uniform manifold approximation and projection (UMAP) plot shows the clusters within HGSOC tumor cells, colored by the patient. The 4 clusters with no distinguishing gene communities enriched are not shown in this plot. **C**, Stacked bar plot shows the proportion of cells in each identified cluster in each tumor. Colors of the bars denote the clusters. **D**, Force-directed embedding layout of the network with the 4742 differentially expressed genes (DEGs) as vertices and their Pearson correlations of the likelihood-ratio test statistic (LRT) as edges. Colors denote the gene communities to which the cells are assigned, and the unassigned genes are omitted. **E**, Box plots comparing the proportion of cells in cluster C5 (left) and C7 (right) between the primary and interval samples.

Figure 3.

Stress score is associated with chemotherapy response and patient prognosis. A, Violin plots comparing the stress score between cells in primary and interval tumors. Violins are colored by the treatment phase and each dot represents a cell. The statistical significance by Wilcoxon rank-sum test is indicated by asterisks (****, $p < 0.0001$; ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$ or ns, $p > 0.05$). **B,** Box plots comparing the stress scores between primary and interval tumors (left, paired Wilcoxon rank-sum test $p = 0.024$) and primary and relapse tumors (right, paired Wilcoxon rank-sum test $p = 0.0078$) in the HERCULES cohort. **C,** Kaplan–Meier curves on progression-free survival for stress-high and stress-low patients (log-rank test, $p = 0.0037$) from the TCGA cohort. The number of patients at risk is listed below the survival curves for each time point. **D,** Forest plot showing hazard ratios, their confidence intervals and p-values based on a multivariate Cox proportional hazards regression model studying progression-free survival in relation to COSMIC signature 3 status, age at diagnosis, tumor purity and stress score. Significant p values ($P < 0.05$) are shown in bold. **E,** COSMIC signature 3 status is not enriched in stress-high or low patients in the TCGA cohort (Fisher exact test, $p = 0.3133$). **F,** Kaplan–Meier curves on progression-free survival for patients from the TCGA cohort with stress-high & signature 3 negative (ii) stress-low & signature 3 negative (iii) stress-high & signature 3 positive and (iv) stress-low & signature 3 positive. The log-rank test p value is 0.0097. The number of patients at risk is listed below the survival curves for each time point.

Figure 4.

Stress-high tumors show altered microenvironments. A, Box plots comparing the proportion of tumor ($p = 6.1e-11$), stromal ($p = 2.4e-09$) and immune ($p = 3e-04$) components between stress-high and stress-low patients from the TCGA cohort. Each dot represents a sample, the p values were computed using Wilcoxon rank-sum test. **B,** Heatmap shows the tumor cell specific expression of epithelial to mesenchymal transition (EMT), pro-angiogenic, pro-inflammatory and pro-survival marker genes in the 86 stress-high and 145 stress-low samples from TCGA cohorts. Samples are sorted by

the stress scores. **C**, Chord diagram shows interactions between the tumor component and stromal, immune components that are increased in stress-high tumors compared to stress-low tumors ($\log_{10}\text{FDR} < -5$ & fold change > 5). **D**, Box plots comparing the protein expression levels of c-JUN_pS73 ($p = 0.0035$), JNK_pT183Y185 ($p = 0.00077$), P38_pT180Y182 ($p = 0.017$) between stress-high and stress-low patients from the TCGA cohort. P values were computed using Wilcoxon rank-sum test.

Figure 5.

Association between stress scores and drug sensitivity in the GDSC2 cell line dataset. **A**, Dot plot shows the stress-high (233) and stress-low (504) cells lines in the GDSC2 dataset. Colors denote the stress status. **B**, Volcano plot shows the compounds with significantly different half-maximal inhibitory concentration (IC₅₀) between stress-high and stress-low cell lines. The red and green dots correspond to the compounds with significantly different IC₅₀ ($\log_2\text{fold-change} > 0.5$, Wilcoxon rank-sum test $p < 0.05$). **C**, Scatter plot shows the correlation between the stress scores and $\ln\text{IC}_{50}$ across the 12 ovarian cancer cell lines harboring *TP53* mutations ($R = -0.66$, $p = 0.02$).

Figure 1**A** **B**

Figure 2

A

B

D

C

E

Figure 3**A** **B** **C** **D** **E** **F**

Figure 4

A

C

B

D

Figure 5

A

B

C