**Ovarian Cancer**

Ovarian cancer is a prominent contributor to cancer-related mortality in women, occupying the fifth position across all age groups [1]. Based on the data provided by [2], it is projected that the number of new cases and deaths attributed to ovarian cancer in 2024 will be 19,680 and 12,740, respectively. In general, these carcinoma metastasis through two basic mechanisms: direct extension from the ovarian or fallopian tumor to adjacent organs such as the bladder or colon, or detachment of cancer cells from the primary tumor by undergoing an epithelial-to-mesenchymal transition named passive metastasis. Upon reaching the designated location, ovarian cancer cells undergo a transformation from mesenchymal to epithelial cells, resulting in the formation of an epithelial phenotype. This epithelial phenotype enables the cells to respond effectively to paracrine growth stimuli. Once tumors attain a specific size, they necessitate the formation of additional blood vessels, as the process of diffusion alone becomes insufficient in supplying the necessary nutrients for tumor growth and initiating their subsequent cycle [3]. Despite the absence of a widely recognized explanation for the development of ovarian cancer, a genetic factor plays an important role. Furthermore, one of the most difficult aspects of understanding ovarian cancer's pathogenesis is its heterogeneity, comprising diverse behaviors and characteristics including 1) high-grade serous carcinoma 2) high-grade endometrioid carcinoma 3) low-grade serous carcinoma 4) Low-grade endometrioid cancer 5) clear cell carcinoma 6) mucinous carcinoma [4, 5].

One of the hallmarks of cancer is the presence of somatic mutation. Mutations usually lead to cancer progression and aggressiveness, and possibly link to the resistance to treatments. Based on the analysis of exomes and whole genomes, the Catalogue Of Somatic Mutations In Cancer (COSMIC) reported that ovarian cancer is associated with mutation signature 3 which strongly associated with germline and somatic BRCA1 and BRCA2 mutations [6]. The deleterious mutations in the BRCA1/2 gene are linked to the homologous recombinant (HR), a type of DNA strand break repair, deficient signature. However, not all HR deficiency is associated with BRCA1/2 mutation. The specific signatures of base substitutions and structural chromosomal variation can, also, cause the HR deficient ovarian cancer [7]. Somatic mutation in TP53 is also a signature reported in epithelial ovarian cancer (EOC) and linked to resistance to platinum-based therapy [8]. As mutations play an important role in ovarian cancer, previous research used the somatic mutation profile to classify EOC tumor subtypes. One with the mutations in the MAPK pathway—KRAS, BRAF, PTEN, CTNNB1, etc.; the other with TP53, BRCA1, BRCA2, KIT, and EGFR mutations [9]. For the more aggressive type of ovarian cancer, high-grade serous ovarian carcinomas (HGSOC), the elderly (age equal or over 75) showed more substitution mutations in “Signature 1”, which is the spontaneous deamination of 5-methylcytosine due to age [10]. Moreover, analysis of the genetic alterations across different ovarian cancer tissues showed that BRCA1/2 alterations, PTEN loss, and gain of PIK3CA and CCND1 were characteristic of HGSOC [11]. Therefore, identifying these mutations and mechanisms offers crucial insights into the development of ovarian cancer.

* 1. **Ovarian cancer treatments and targeted therapy**

Ovarian cancer treatment strategies can be classified into local and systemic therapy. Local therapies, including radiotherapy and surgery, are designed to treat the site of the tumor with the intention of minimizing any adverse effects. Systemic treatments, such as chemotherapy, hormone therapy, targeted drugs, and immunotherapy, are able to target cancer cells by circulating throughout the body. For the current standard ovarian cancer treatments, surgery followed by platinum-based chemotherapy has been adopted depending on the patient's situation and disease stages. However, chemotherapeutic drugs are formulated on toxic compounds which leads to the undesirable side effects observed in cancer therapy [12]. This indiscriminate destruction of cells and the adverse effects of chemotherapy have prompted the search for alternative methods to treat cancer.

Targeted therapy is the novel methods that aim to interfere with specific pathways related to carcinogenesis and tumor growth by inducing apoptosis, and blocking specific enzymes and growth factor receptors involved in cancer cell proliferation. Unlike chemotherapy, targeted drugs is able to modify the function of proteins that regulate gene expression, and cellular activities, and interfere with specific tumor pathways [13].

The DNA repair mechanism of cancer cells is one target of targeted treatments used to treat ovarian cancer. Specific DNA repair mechanisms, including single-stranded break repair (SSBR), homologous recombination repair (HR), and non-homologous end joining repair (NHEJ) are capable of eliminating DNA damage during the DNA damage response [14]. The NHEJ and HR pathways are the two primary mechanisms used in DNA double-strand breaks (DSBs). NHEJ mechanism is prone to errors, leading to the accumulation of genetic abnormalities and eventually apoptosis while HR adeptly restores DSBs with a notable degree of precision. To respond to DNA single-strand breaks (SSBs), cells initiate the base excision repair (BER) pathway, which is dependent on Poly (ADP-ribose) polymerases (PARPs), specifically PARP1, PARP2, and PARP3 [14].

Thereby employing the rational idea of synthetic lethality, in which mutations in two genes are synthetically fatal, but a mutation in one gene alone preserves cell viability, leading to the creation of PARP inhibitors (PARPi). PARPi specifically targets the BRCA-associated subtype of ovarian cancer by disrupting the BER pathway since individuals who possess heterozygous germline of BRCA1 or BRCA2 mutations encounter a higher risk of developing ovarian cancer, with a lifetime risk ranging from 10% to 40% [15]. Recognizing the functions of the BRCA1 and BRCA2 proteins is crucial in preserving the integrity of the genome and ensuring the effective repair of DSBs through the HR pathway. As a result, when cancer cells with a BRCA mutation are exposed to PARPi, they undergo the accumulation of SSBs. This accumulation then leads to the formation of DSBs, which are unable to be repaired by the HR pathway due to alteration in the BRCA gene. Consequently, this process eventually leads to cell cycle arrest or cell death [16].

PARP inhibitors (PARPi), such as olaparib, niraparib, talazoparib, and rucaparib, have demonstrated potential in the treatment of ovarian, breast, and pancreatic cancers. They can be used as standalone treatments, in conjunction with other therapies, or as maintenance therapy after a positive response to platinum-based chemotherapy. These medicines have been praised for their ability to provide long-lasting survival advantages to specific patients, both in controlled experiments and in real-life situations. Nevertheless, the initial enthusiasm for PARPi has been dampened by the establishment of both innate and acquired resistances. It has been demonstrated that the mechanisms producing treatment resistance in tumors are complicated and involve several mechanisms, rather than being caused by a single, isolated cause [17].

* 1. **Resistance to therapy and Mechanism**

The successiveness of utilizing chemotherapeutic and targeted drugs for ovarian cancer treatment might be limited by the resistance mechanism. Intrinsic resistance, which the resistance exists before the patient is administered with drugs. This can be caused by pre-existing genetic mutations, and the presence of insensitive subpopulations within heterogeneous tumors, leading to recurrence and the activation of intrinsic mechanisms that serve as a defense mechanism against environmental poisons, such as anticancer drugs [18]. In addition to intrinsic resistance, cancer cells have the potential to develop resistance phenotypes throughout treatment. Cisplatin, platinum-based chemotherapy, which is used after standard surgical intervention can induce resistance phenotypes by upregulating the colony-stimulating-factor-1 receptor (CSF-1R). This upregulation causes the evasion of apoptosis, a downstream effect of the activation of the PI3K/AKT and ERK1/2 pathways which play an important role in mediating the anti-apoptotic activity that contributes to chemo-resistance [19].

Targeted therapies, like PARPi, encounter the same issue as chemotherapy in which cancer cells develop resistance to the treatment. Like chemotherapy, resistant phenotype can arise either from inherent resistance or acquired resistance induced by therapy. One of the possible ways in the development of resistance to PARP inhibitors is the genomic reversal of BRCA1/2 gene mutations. Typically, people who have impaired DNA repair processes caused by defects in the BRCA1 or BRCA2 genes experience benefit from PARPi therapy. Nevertheless, certain types of somatic reversion mutations can regain the normal functioning of these genes, allowing the cells to efficiently repair DNA damage once more and thereby avoid the effects of PARPi [20]. Aside from mutations, acquired resistance in cancer cells can also occur through the overexpression of P-glycoprotein. This mechanism, known as multi-drug resistance, leads to a decrease in the concentration of drugs within the cancer cells due to the upregulation of ATP-binding cassette (ABC) transporters. Nevertheless, because the nature of cancer cells is heterogeneous and can differ even within the same tumor site, the resistance mechanism might vary. Therefore, comprehensive studies are necessary to gain a better understanding and improve therapeutic effectiveness.

* 1. **Single cell-RNA sequencing**

Conventional approaches to dissecting the mechanism of drug resistance in cancer might be low throughput, with specific possible pathways being investigated. However, these strategies do not encompass all potential resistance mechanisms. Transcriptomics is an approach that uses RNA sequencing to analyze the whole expression of cellular mRNAs providing insights into the changes in biological processes caused by resistance. In addition, RNA sequencing offers a high throughput efficiency that aids in expediting the study procedure. However, because ovarian cancer is complex, and individual cells may reflect different phenotypes, greater resolution RNA-sequencing is required.

The complexity of ovarian cancer is increased by the presence of tumor heterogeneity, which can be classified as intra- and inter-tumoral heterogeneity. Inter-tumoral heterogeneity refers to the genetic and physical variations observed among numerous tumors of the same type within an individual. Intra-tumoral heterogeneity is the presence of various cell groups within a single tumor. Therefore, bulk transcriptomics is not ideal for conducting cell-level research on tumor tissue that is characterized by a high degree of heterogeneity and complicated components. This restriction makes it more difficult to have a precise knowledge of the clinical and biological mechanisms that underlie these conditions. Therefore, single-cell RNA sequencing (scRNA-seq) is employed in the analysis to investigate the RNA transcript within individual cells, thereby revealing the variability within and across tumors [21].

A prior study utilizing scRNA-sequencing to analyze ovarian cancer heterogeneity demonstrated that HGSOC may be classified into six different populations: epithelial cells, fibroblasts, T cells, macrophages, endothelial cells, and B cells. More precisely, among groups of epithelial cells (EC), different subtypes of these cells exhibited varied pathways that were more abundant. As an example, EC5 exhibited an increased presence of chemoresistance markers, such as TOP2A, NEK2, and FEN1. In addition, these subtypes were linked to an increase in the presence of homologous recombinant signature genes such as RAD51, BARD1, and MND1. Through trajectory analysis of HGSOC, it was observed that the early stage of the disease progressed towards metastases along the pseudo-time. The analysis of epithelial subtype trajectories revealed an increase in the presence of GO and KEGG related to cell cycle and cell division. Conversely, there was a decrease in the presence of GO and KEGG related to apoptosis.

The previous report indicates that [22] EOC that may be resistant to PARP inhibitors showed large and notable enrichment for pathways related to cell cycle, DNA replication, DNA repair, and drug metabolism. Furthermore, this subgroup had a significant differential expression of other genes related to homologous recombination, including BRIP1, BARD1, RAD51, MND1, TTK, RAD51AP1, BIRC5, RPA1, and HJURP [23]. Apart from the homologous recombination-related gene expression, the overall composition of this subgroup consists of S-phase and G2/M-phase indicating a state of hyperproliferation [23].

scRNA-seq in the context of ovarian cancer, provides valuable information that may be used to monitor lineages and make predictions. Lineages can be used to determine differences in cell differentiation and developmental stages. Additionally, in the context of acquired drug resistance, the emergence of resistance characteristics during and after treatment can be regarded as distinct lineages. Pseudo-time trajectory interference is a method used to examine the relationship between lineages and demonstrate the emergence of particular cell subgroups as time progresses [24]. An example of this is the segregation of the trajectory based on cancer progression, with the starting point being the normal epithelial cells. Within the context of the aforementioned study, it was observed that a certain subgroup of ovarian cancer exhibited a correlation with metastasis and a poorer prognosis, as indicated by the pseudo-time trajectories [25]. Therefore, utilizing trajectory analysis to elucidate PAPRi resistance would offer a comprehensive understanding of differences in gene expression at the level of individual cells.

Regarding all the mentioned evidence, this research aims to investigate the underlying mechanism of resistance in epithelial cancer cells through the investigation of drug-induced resistance in target therapy, specifically PARPi. Although significant progress has been made in studying various strategies to address and minimize resistance pathways at the molecular level, there is still yet much more to explore. For instance, combining PARPi with other targeted medicines demonstrated potential in making patients responsive again to PARPi or improving the efficacy of treatment to reach synthetic lethality [26]. In this study, we employ the A2780 endometroid adenocarcinoma cell line and subject it to treatment with Olaparib until it develops a resistance phenotype, replicating the therapeutic approach used for ovarian cancer, subsequently, through the process of integrating single-cell analysis and understanding ovarian cancer heterogeneity to reveal the mechanisms of PARPi resistance and discovering biomarkers for predicting the outcomes of treatment. Therefore, identifying individuals who are susceptible to PARPi and minimize the development of drug resistance.

**3. Material and Methods:**

**Single-cell RNA seq data processing:**

The Seurat object is used to structure data into various slots that store specific information, such as cell-level metadata, a list of assays within the object, nearest-neighbor graphs, and more [27]. The “merge()” function is used to combine the data from the sensitive (A2780S\_S) and resistance (A2780S\_R) groups within a Seurat object creating a new object called A2780S.

**Quality Control, Filtering, and Normalization:**

Following the creation of the A2780S object, quality control, and filtering were performed by identification of mitochondrial genes identified using the pattern matching function "grep()", with the pattern parameter set to "^MT-". The proportion of reads that mapped to the mitochondrial genome was computed using "PercentageFeatureSet()". The result was divided by 100 and saved as “percent.mito" to A2780S. The upper limit for nFeature\_RNA and “percent.mito” was determined by adding the mean value to two standard deviations. The minimum threshold for the number of RNA features was set at 200, while no specific criterion was applied for the "percent.mito" parameter. Following conducting quality control, the next step is the normalization process. In this procedure, we use the “LogNormalize” argument in parameter normalization.method inside NormalizeData() function.

**Identification of Highly Variable Features (Feature Selection):**

For the identification of genes that display significant variability among cells, "mean.var.plot" argument was selected for the selection method in the FindVariableFeature() function to calculate the mean expression and variance of genes. In this study, we retain genes that have average expression levels ranging from 0.0125 to 3 and a dispersion value starting from 0.5 to infinity for further investigation.

**Scaling the data:**

Linear Transformation is performed using ScaleData() function to scale and center features in the dataset using “vars.to.regress” method to regress out the effect of “nCount\_RNA” and “percent.mito”.

**Linear Dimensionality Reduction (PCA):**

We compute PCA using the RunPCA() function on the standardized data and store 200 principle components by setting the argument “npcs=200” in A2780S. The number of principal components (PCs) that will be utilized will be determined using the elbow plot and JackStraw technique. JackStraw uses each PC to plot quantile plots comparing the distribution of p-values for all genes across PCs and determine the p-value for overall significance for each PC.

**Cluster the cells:**

To generate clusters, we utilize the FindNeighbors() method to create a K-nearest neighbor graph based on the Euclidean distance in the PCA space. This function accepts input from a predetermined dimensionality in the prior step. Next, we employ the Louvain method, a technique designed to optimize modularity. This procedure iteratively clusters cells together using the FindCluster() function, with the resolution as an argument. In this case, we utilize the clustree() method to iterate over a range of resolutions to select the appropriate resolution to pass into the FindCluster() function. Following running the FindCluster() function, we generate a CSV file and heatmap using GraphPad to determine the cluster to which the sensitive and resistant cells belong.

**Non-linear dimensional reduction (UMAP):**

The data visualization was conducted using the RunUMAP() function with arguments "reduction=pca" and “dims=1:50”.

**Differential gene expression analysis:**

The Differential Expression (DE) process is performed to determine the marker genes within each cluster. This is achieved by comparing the positive and negative markers of a single cluster and compare to all other clusters using the FindAllMarkers() method. The argument "logfc.threshold = 0.25" and "min.pct = 0.25" in this analysis. The default of "logfc.threshold = 0.25" is 0.1.

**Trajectory analysis:**

After clustering data in the Seurat object with pre-defined resolutions of 0.1 and 0.15, the trajectory analysis was performed using Monocle 3 (R-based), which made use of the reduced dimension data from Seurat by converting to cell data object using “as.cell\_data\_set(A2780S)” and store in “cds”. The trajectory of cells was performed using “learn\_graph(cds, use\_partition=FALSE)”. The partition argument is set to "FALSE" to learn a single graph across all partitions and determine their cellular lineage connections. The UMAP algorithm built into Monocle 3 was used to visualize these trajectories using “FeaturePlot()”. Following that, a head-to-head comparison of trajectory analysis was performed with Slingshot (R-based). The Slingshot package allows for the monitoring of developmental progressions within cell populations. This method constructs a minimum spanning tree using Seurat's cluster labels and reduced dimensionality data, sorting cells along developmental paths. The Seurat object was converted to SingleCellExperiment using “as.SingleCellExperiment(A2780S) and store in “sce”. The trajectory analysis for Slingshot was run using “slingshot(sce, clusterLabels = 'seurat\_clusters', reducedDim = 'UMAP')” command. The output of pseudotime Slingshot was extracted using “slingPseudotime(SlingshotDataset(sce), na = FALSE)” and visualize using “plot()” function.

**4. Result:**

**Single‐cell RNA sequence data merging and Quality Control:**

By merging the data from the A2780S\_S and A2780S\_R groups, we obtain a total of 3,870 cells in A2780S. Next, we conduct quality control on the cells within A2780S, resulting in 3,449 cells that meet the criteria. The threshold values for nFeature\_RNA and "percent.mito" are indicated by red lines as shown in Figures 1 and 2, respectively. Normalization was performed to ensure that each variable contributes equally to the downstream analysis. Following the normalization process, we were able to identify 8,256 different genes, as seen in Figure 3 that display significant variability among cells.

|  |  |
| --- | --- |
| **A diagram of a graph  Description automatically generated with medium confidence**  Figure 1. The threshold cut-off value for nFeature is 4,710.25 which represents the red line. | **A diagram of a graph  Description automatically generated**  Figure 2. The threshold cut-off value for nFeature is 4,710.25 which represents the red line. |
| **A graph showing the average expression  Description automatically generated**  Figure 3. The number of feature subsets that demonstrate significant cell-to-cell variance in the dataset. The value of the variable count is 8257 | |

**Linear Transformation and Linear Dimensionality Reduction:**

The next step is to do a linear transformation which is a common pre-processing step before using dimensional reduction methods like Principal Component Analysis (PCA). This step is crucial to guarantee that both genes with high and low expression levels are included in the analysis based on their variance, rather than just on their average expression. It is important to consider both high and low-expression genes to be important if they show considerable variation between cells. Subsequently, we perform PCA on the standardized data and create an elbow plot to identify the most suitable number of PCs to keep, as illustrated in Figure 4. The elbow plot allows us to identify the principal components by analyzing the percentage of variance explained by each PC. Furthermore, by utilizing the JackStraw method to determine the significant principal components, Figure 5 demonstrates that even with 50 PCs, there is a strong enrichment of features with low p-values. Therefore, opting for 50 PCs appears to be the most suitable decision.

|  |
| --- |
| A line graph with numbers and symbols  Description automatically generated  Figure 4 Elbow plot used to determine the appropriate number of principal components to retain. The elbow point for this plot is around 25 PCs. |
| A rainbow colored lines on a black background  Description automatically generated  Figure 5. JackStraw plot for analysis of PCA significance. The calculated p-value for each is shown on the right, where they are ordered from 1 to 50 PCs. This indicates that even at 50 PCs, there is still some level of significance in the PCs, however, the p-value is not significant for certain PCs, and below the black dotted line are insignificant PCs. |

**Clustering and Non-Linear Dimensionality Reduction:**

The clustree visualizations between a range of resolutions are illustrated in Figure 6. In this case, we have used a resolution of 0.1 and 0.15 in order to group the data into three and five clusters accordingly. The UMAP visualization of resolution 0.1 (three clustered data) is presented in Figure 7, while the UMAP visualization of resolution 0.15 (five clustered data) is presented in Figure 8. The UMAP of visualization highlighting the sensitive and resistant cells is displayed in Figure 9.

The heatmap visualization of clustering resolutions 0.1 and 0.15 is shown in Figures 10 and 11 respectively. From Figure 10, we can infer that at resolution 0.1, cluster 0 predominantly consists of the sensitive group, cluster 1 primarily represents the resistance group, and cluster 2 is composed of a few cells. Conversely, while observing at a resolution of 0.15, it can be observed that group 0 predominantly consists of sensitive cells, while group 2 consists largely of resistant cells. On the other hand, groups 1 and 3 are a combination of both sensitive and resistant cells while group 4 is composed of very few cells.

|  |
| --- |
| Figure 6. Clustree for resolution ranging from 0 to 1. Each node represents the number of samples in each cluster. The color of each node indicates the clustering resolution. The arrow connecting nodes represents the proportion of sample from incoming nodes. |

|  |  |
| --- | --- |
| **A diagram of a red green and blue dot  Description automatically generated**  Figure 7. Uniform Manifold Approximation and Projection (UMAP) of cells at 0.1 resolution. | **A colorful dots on a white background  Description automatically generated**  Figure 8 Uniform Manifold Approximation and Projection (UMAP) of cells at 0.15 resolution. |
| **A diagram of a number of dots  Description automatically generated**  Figure 9. UMAP visualization of the sensitive and resistance groups. A2780S\_R is a group that shows resistance. A2780S\_S is a group that shows sensitive. | |

|  |  |
| --- | --- |
| Figure 10. Heatmap clustering at 0.1 resolution Cluster 0 is mainly consisting of sensitive cells. Cluster 1 is mainly consisting of resistance cells. Cluster 2 consist of few cells | **A blue and white squares  Description automatically generated**  Figure 11. Heatmap clustering at 0.15 resolution. Cluster 0 is mainly consisting of sensitive cells. Cluster 2 is mainly consisting of resistance cells. Cluster 1 and 3 are combination of sensitive and resistance cells. Cluster 4 consist of few cells. |

**Differential gene expression analysis:**

The DE is performed to determine the marker genes within each cluster by choosing genes that have a minimum 1.18-fold difference to be considered significantly differentially expressed between the two groups of cells. By increasing the logfc.threshold value to 0.25, the analysis is expedited as it decreases the number of genes requiring testing for differential expression, hence enhancing efficiency. However, there is a potential drawback of potentially overlooking genes that exhibit more subtle changes in expression, which could still be relevant. Additionally, we indicate that at least 25% of the gene had to be present in each of the two groups being compared to be considered for further analysis. This made sure that the analysis focused on genes that are broadly expressed in one or both conditions instead of genes that are only occasionally found. The dot plot for resolution 0.1 shown in Figure 12 reveals the top 5 variations in gene expression among cluster 0 (mostly sensitive cells) consisting of PGF, CNN3, MFAP4, BST2, and UCHL1 while cluster 1 (mostly resistance cell) consists of **GTSE1, FDXR, DSC2, SLC10A4, and KCMA1**. Cluster 2 consists of mostly mitochondrial genes, MT-ND3, MT-ATP6, MT-CO2, MT-CYB and MT-ND2. The dot plot for resolution 0.1 shown in Figure 13 reveals the top 5 variations in gene expression among cluster 0 (mostly sensitive cells) consisting of PGF, HEBP2, MT1X, SOX4, and UCHL1 while cluster 2 (mostly resistance cells) consisting of **GTSE1, CDKN1A, DSC2, PTTG1 and KCNMA1**. Cluster 4 consists of mostly mitochondrial genes, MT-ND3, MT-ATP6, MT-CO2, MT-CYB, and MT-ND2.

|  |
| --- |
| **A group of dots in different colors  Description automatically generated**  Figure 12. Top 5 differential gene expression at 0.1 resolution in each cluster. Top 5 gene expression in cluster 0 consist of PGF, CNN3, MFAP4, BST2 and UCHL. Top 5 gene expression in cluster 1 consist of GTSE1, FDXR, DSC2, SLC10A4 and KCMA1. Top 5 gene expression in cluster 2 consist of MT-ND3, MT-ATP6, MT-CO2, MT-CYB and MT-ND2. The differential gene expression analysis was conducted using the non-parametric Wilcoxon rank sum test, and the resulting p-values were adjusted using Bonferroni correction. |
| **A screenshot of a computer screen  Description automatically generated**  Figure 13. Top 5 differential gene expression at 0.15 resolution in each cluster. Top 5 gene expression in cluster 0 consist of PGF, HEBP2, MT1X, SOX4 and UCHL1. Top 5 gene expression in cluster 1 consist of C1orf56, TMEM251, MARCH5, KLHDC3 and PCCB. Top 5 gene expression in cluster 2 consist ofGTSE1, CDKN1A, DSC2, PTTG1 and KCNMA1. Top 5 gene expression in cluster 3 consist of FAM184B, DSC2, LXOL2, SLC10A4 and NAA25. Top 5 gene expression in cluster 4 consist of MT-ND3, MT-ATP6, MT-CO2, MT-CYB and MT-ND2. Wilcoxon rank sum test, and the resulting p-values were adjusted using Bonferroni correction. |

**Trajectory analysis:**

In addition to identifying cell identities and analyzing differential gene expression among sub-populations, we perform trajectory analysis to reveal the dynamics of cellular lineages over time. When examining the single-cell RNA sequencing data from the A2780S cell line, significant variations in cluster formation and trajectory analysis were detected when comparing two different resolutions (0.1 and 0.15) and two computational methods (Monocle3 and Slingshot).

The UMAP plots produced by Monocle3 and Slingshot reveal the cellular diversity and the underlying chronological paths that represent the progression of diverse lineages. When using a clustering resolution of 0.1, the initial population was assigned as the A2780S sensitive phenotype (A2780S\_S, cluster 0). The data visualization provided by Monocle3 showed a non-uniform distribution of clusters, suggesting that there may be less distinct differences between cell states related to drug sensitivity and resistance. In contrast, when the resolution was increased to 0.15, clusters were shown to be more widely spaced out. This indicates that this greater resolution is capable of capturing greater differences in the phenotypes. The pseudotime UMAP analysis revealed that the lineage originated from cluster 0, which consisted of PARPi-sensitive A2780S cells. The lineage then progressed to cluster 1, which represented most of the sensitive population (as shown in Figure 11), and cluster 2, which represented the population with the highest resistance to PARPi (also shown in Figure 11).

Furthermore, Slingshot aided in the identification of potential developmental trajectories between the sensitive and resistant states of the cells. The trajectory plots indicate a clear and direct progression from sensitivity to resistance on both clustering resolutions, which is especially noticeable in the plots generated using the 0.15 resolution. This suggests that using a higher resolution may allow for a more accurate depiction of the subtle transitions between these phenotypes. However, the Slingshot gave the un-well boundary between each subpopulation over pseudotime (Figure 15)These results emphasize the importance of choosing the right tools and setting the appropriate resolution when analyzing scRNA-seq data. These factors can have a major impact on how cellular states and their developmental trajectories are interpreted.

|  |  |
| --- | --- |
| Resolution 0.1 | Resolution 0.15 |
| **A diagram of a map  Description automatically generated with medium confidence** |  |
| *Figure 14. UMAP map comparing two trajectory analyses with different resolutions using Monocle3. The graph displays the cluster identification as a number and expresses pseudotime using different colors.* | |

|  |  |
| --- | --- |
| Resolution 0.1 | Resolution 0.15 |
| **A red green and blue blotches  Description automatically generated** | **A colorful blot of paint  Description automatically generated with medium confidence** |
| **A chart with red and green dots  Description automatically generated** | **A chart with many colored dots  Description automatically generated with medium confidence** |
| *Figure 15. UMAPs and pseudotime displays of Slingshot trajectory analysis. The connecting line in the UMAP plot represents the commitment of distinct cell subpopulations to specific lineages, as indicated by the different colors observed along the pseudotime progression. Two distinct clustering resolutions, 0.1 and 0.15, are demonstrated.* | |

**5. Discussion:**

The application of single-cell RNA sequencing to evaluate the mechanism of induced drug resistance in ovarian cancer provides promising insight into the role of heterogeneity in inducing certain pathways that may be associated with resistance. Our framework could elaborate on drug sensitivity in real-world circumstances by analyzing the two separate phenotypes of A2780 cell lines, sensitive and resistant. Our single-cell transcriptomics analysis workflow used R language with numerous packages, including Seurat and Slingshot, and modified methods from earlier studies to match well with our data.

Regarding the first step of the pipeline, QC and filtering, we successfully reduced the amount of data originating from cells of low quality and those that were dying by selectively filtering gene and mitochondria expression. Cells with high gene expression may result from doublets, while low gene expression may result from empty droplets or low-quality cells. Furthermore, death cells typically exhibit a substantial abundance of mitochondria, therefore necessitating the filtration of mitochondria. The normalization step ensures that each gene expression contributes equally to later analyses. To identify a highly variable feature, cells that display significant variation between each other are determined by modeling the relationship between the mean and variance, this information is then utilized in PCA. Moreover, a linear transformation (scaling) was used to ensure that genes with both high and low expression levels contribute equally based on their variation, rather than just their mean expression.

Next, we conducted PCA to find the optimal number of PCs to retain. This was accomplished by employing an elbow plot and JackStraw analysis, as the number of PCs to be kept varies depending on the dataset. According to Seurat documentation, selecting higher PC parameters is recommended as insufficient PCs might have a significant and negative impact on the outcomes. From the elbow plot, we can observe an ‘elbow’ around 20 – 30 PCs suggesting the majority of the true signal (not from technical artifacts) is captured in the first 30 PCs. Apart from the elbow plot, which is a heuristic method, JackStraw showed the p-value distribution for every PC [28], identifying a total of 50 PCs having curves above the black dashed line, indicating significant PCs. Since the JackStraw approach provides for a more accurate estimation of the control variables used to estimate the surrogate variables [28], 50 PCs were retained for further analysis.

Once the PCA has been performed, the following step is to determine the resolution for clustering the cells together. The clustree visualization indicates that the number of clusters can vary based on the resolution selected. In this report, we utilized our prior understanding of this study, which included only the sensitive and resistant groups of the A2780 cell line. We selected a resolution of 0.1 to generate three distinct clusters. It is important to note that it is possible to obtain two clusters with a resolution of 0.05. However, the algorithm groups sensitive and resistant together in a single group, resulting in another group with very few cells. Furthermore, we employ a resolution of 0.15 to identify subpopulations within each cluster. By utilizing a heatmap in Figures 10 and 11, we can determine that at a resolution of 0.1, cluster 0 corresponds to the sensitive group, while cluster 1 represents the resistance group. At a resolution of 0.15, cluster 0 represents the sensitive group and cluster 2 represents the resistant group.

The results of our investigation using single-cell transcriptomics on the ovarian cancer cell lines A2780, which are susceptible and resistant to PARPi, demonstrate the presence of heterogeneity and provide insights into the mechanisms of resistance to this treatment. The A2780 cell line exhibited a wild-type (non-mutated) genotype for the BRCA1/2 genes [29], which are crucial genes involved in PARPi therapy. This genotype does not result in HRD phenotypes and is associated with a high likelihood of resistance to PARPi. These are not exclusive to particular cancer types or cancer cell lines. Prior studies have indicated that gastric cancer with a BRCA1/2 wild-type genetic background also develop HRD [30], which makes it susceptible to synthetic lethality medicines. Although the genetic background of the BRCA genes in the A2780S cell does not indicate that it is sensitive to PARPi, our findings demonstrate that A2780 exhibits both PARPi sensitive and resistant phenotypes, indicating heterogeneity. The examination of single-cell RNA sequencing revealed distinct clusters at different clustering resolutions (0.1 and 0.15) showing differential expression signature genes of each cluster and could potentially serve as the basis for the drug sensitivity profiles.

At a resolution of 0.1, there were three distinct clusters. The majority of cells within the cluster 0 are sensitive, while the majority of cells within the cluster 1 are resistant. The G2 and S-phase expressed 1 (GTSE1) gene is one of the interesting genes identified in the PARPi resistance cluster. Prior research has indicated that GTSE1 may be linked to a poor prognosis in various types of cancers [31]. In hepatocellular carcinoma, GTSE1 exhibited a positive correlation with both cell proliferation markers and epithelial to mesenchymal transition. This correlation was linked to the stemness of cancer cells [32]. GTSE1 expression in gastric and bone cancer was associated with resistance to cisplatin [33, 34]. Therefore, GTSE1 could potentially function as the indicator for resistance to PARP inhibitors. Moreover, it could illustrate the adaptive resistance phenotype to the PARPi of ovarian cancer by enhancing cell proliferation and decreasing apoptosis. However, when using a clustering resolution of 0.1, certain important genes that are characteristic of distinct clusters do not appear to completely differentiate between PARPi sensitivity and resistance phenotypes. For instance, Ubiquitin C-terminal hydrolase L1 (UCHL1), which is increased in cluster 0 (comprising mostly of PARPi sensitive cells), has previously been associated with Doxorubicin resistance [35] and assisting in the proliferation of cancer cells [36]. These observations may suggest that the setup of the study is limited by the use of cell lines with the same genetic background, which may lack intrinsic heterogeneity.

In contrast to the clustering with a lesser resolution, the clustering with a threshold of 0.15 resulted in the formation of 5 differentiated clusters. The majority of cells were distributed among three clusters: cluster 0 (comprising the most PARPi sensitive cells), cluster 1, and cluster 2 (comprising the most PARPi resistant cells). The heatmap illustrating the correlation between PARPi resistance traits and cluster number indicated that cluster 1 potentially contains both PARPi-sensitive and PARPi-resistant populations, therefore serving as the transitional cluster. The trajectory analysis conducted by Monocle3 has the potential to provide evidence of the lineage history during the transitional stage. Nevertheless, the trajectory analysis conducted by Slingshot did not fully align with this notion. However, as previously stated, the Slingshot did not provide a completely distinct trajectory. Further inquiry is required to examine the transition between groups that are sensitive to PARPi, using both laboratory experiments and analytical evidence.

In terms of signature gene expression, increasing clustering resolution revealed some distinct signature genes. Cluster 0, which comprised the majority of PARPi-sensitive cells, exhibited PGF and UCHL as the genes that served as intermediaries between the two resolutions. The genes CNN3, MFAP4, and BST2 were identified as the newly discovered signature genes. Nevertheless, these genes had not been previously documented to be associated with PARPi sensitivity, suggesting a poor prognosis and resistance to chemotherapy in ovarian and other forms of cancer [37-39]. The gene PTTG1, which encodes a protein involved in regulating sister chromatid separation, was shown to be increased in cluster 2. Prior studies have indicated that the increased expression of this gene is associated with the development of ovarian cancer stemness [40] and resistance to drugs [41].

To summarize, our extensive investigation of the A2780 ovarian cancer cell line by scRNA-seq research highlights the correlation between genetic background and drug resistance pathways, while also illustrating the complex response to PARPi. Despite the absence of mutations in the BRCA1/2 genes, which are usually associated with sensitivity to PARPi, the identification of varied transcriptional profiles and distinct clusters at various resolutions indicates a better comprehension of resistance. The identification of GTSE1 as a possible marker for resistance to PARPi, together with other relevant signature genes, presents novel prospects for investigating the adaptive mechanisms that contribute to ovarian cancer's resistance to PARPi. These findings highlight the necessity for further investigation into various forms and mechanisms of resistance at the cellular level, which may ultimately lead to more accurate and effective therapeutic strategies. This study not only improves our understanding of the efforts to combat ovarian cancer but also highlights the critical importance of single-cell analysis in uncovering the hidden characteristics of cancer biology that are necessary for creative treatment approaches.

**Reference:**

1. Siegel, R.L., et al., *Cancer statistics, 2023.* CA Cancer J Clin, 2023. **73**(1): p. 17-48.

2. Siegel, R.L., A.N. Giaquinto, and A. Jemal, *Cancer statistics, 2024.* CA Cancer J Clin, 2024. **74**(1): p. 12-49.

3. Lengyel, E., *Ovarian cancer development and metastasis.* Am J Pathol, 2010. **177**(3): p. 1053-64.

4. Budiana, I.N.G., M. Angelina, and T.G.A. Pemayun, *Ovarian cancer: Pathogenesis and current recommendations for prophylactic surgery.* J Turk Ger Gynecol Assoc, 2019. **20**(1): p. 47-54.

5. Matulonis, U.A., et al., *Ovarian cancer.* Nat Rev Dis Primers, 2016. **2**: p. 16061.

6. Sondka, Z., et al., *COSMIC: a curated database of somatic variants and clinical data for cancer.* Nucleic Acids Res, 2024. **52**(D1): p. D1210-D1217.

7. Vanderstichele, A., et al., *Genomic signatures as predictive biomarkers of homologous recombination deficiency in ovarian cancer.* Eur J Cancer, 2017. **86**: p. 5-14.

8. Hlavac, V., et al., *Whole-exome sequencing of epithelial ovarian carcinomas differing in resistance to platinum therapy.* Life Sci Alliance, 2022. **5**(12).

9. Rojas, V., et al., *Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment.* Int J Mol Sci, 2016. **17**(12).

10. Serio, P., et al., *Somatic Mutational Profile of High-Grade Serous Ovarian Carcinoma and Triple-Negative Breast Carcinoma in Young and Elderly Patients: Similarities and Divergences.* Cells, 2021. **10**(12).

11. Lapke, N., et al., *Genetic alterations and their therapeutic implications in epithelial ovarian cancer.* BMC Cancer, 2021. **21**(1): p. 499.

12. Pérez-Herrero, E. and A. Fernández-Medarde, *Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy.* Eur J Pharm Biopharm, 2015. **93**: p. 52-79.

13. Joo, W.D., I. Visintin, and G. Mor, *Targeted cancer therapy--are the days of systemic chemotherapy numbered?* Maturitas, 2013. **76**(4): p. 308-14.

14. Li, H., et al., *PARP inhibitor resistance: the underlying mechanisms and clinical implications.* Mol Cancer, 2020. **19**(1): p. 107.

15. Banerjee, S. and M. Gore, *The future of targeted therapies in ovarian cancer.* Oncologist, 2009. **14**(7): p. 706-16.

16. Madariaga, A., S. Lheureux, and A.M. Oza, *Tailoring Ovarian Cancer Treatment: Implications of BRCA1/2 Mutations.* Cancers (Basel), 2019. **11**(3).

17. Giudice, E., et al., *PARP Inhibitors Resistance: Mechanisms and Perspectives.* Cancers (Basel), 2022. **14**(6).

18. Wang, X., H. Zhang, and X. Chen, *Drug resistance and combating drug resistance in cancer.* Cancer Drug Resist, 2019. **2**(2): p. 141-160.

19. *<Chemotherapy Resistance in Advanced Ovarian Cancer Patients.pdf>.*

20. Jackson, L.M. and G.L. Moldovan, *Mechanisms of PARP1 inhibitor resistance and their implications for cancer treatment.* NAR Cancer, 2022. **4**(4): p. zcac042.

21. Roberts, C.M., C. Cardenas, and R. Tedja, *The Role of Intra-Tumoral Heterogeneity and Its Clinical Relevance in Epithelial Ovarian Cancer Recurrence and Metastasis.* Cancers (Basel), 2019. **11**(8).

22. Hao, Q., et al., *Single-cell transcriptomes reveal heterogeneity of high-grade serous ovarian carcinoma.* Clin Transl Med, 2021. **11**(8): p. e500.

23. Izar, B., et al., *A single-cell landscape of high-grade serous ovarian cancer.* Nat Med, 2020. **26**(8): p. 1271-1279.

24. Van den Berge, K., et al., *Trajectory-based differential expression analysis for single-cell sequencing data.* Nat Commun, 2020. **11**(1): p. 1201.

25. Guo, J., et al., *Single-cell transcriptomics in ovarian cancer identify a metastasis-associated cell cluster overexpressed RAB13.* J Transl Med, 2023. **21**(1): p. 254.

26. Fu, X., et al., *Mechanism of PARP inhibitor resistance and potential overcoming strategies.* Genes Dis, 2024. **11**(1): p. 306-320.

27. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587.e29.

28. Chung, N.C. and J.D. Storey, *Statistical significance of variables driving systematic variation in high-dimensional data.* Bioinformatics, 2015. **31**(4): p. 545-54.

29. Stordal, B., et al., *BRCA1/2 mutation analysis in 41 ovarian cell lines reveals only one functionally deleterious BRCA1 mutation.* Mol Oncol, 2013. **7**(3): p. 567-79.

30. Alexandrov, L.B., et al., *A mutational signature in gastric cancer suggests therapeutic strategies.* Nat Commun, 2015. **6**: p. 8683.

31. Tan, K., et al., *Pan-cancer analyses reveal GTSE1 as a biomarker for the immunosuppressive tumor microenvironment.* Medicine (Baltimore), 2023. **102**(34): p. e34996.

32. Wu, X., et al., *GTSE1 promotes cell migration and invasion by regulating EMT in hepatocellular carcinoma and is associated with poor prognosis.* Sci Rep, 2017. **7**(1): p. 5129.

33. Subhash, V.V., et al., *GTSE1 expression represses apoptotic signaling and confers cisplatin resistance in gastric cancer cells.* BMC Cancer, 2015. **15**: p. 550.

34. Xie, C., et al., *GTSE1 is possibly involved in the DNA damage repair and cisplatin resistance in osteosarcoma.* J Orthop Surg Res, 2021. **16**(1): p. 713.

35. Lu, G., et al., *The Deubiquitinating Enzyme UCHL1 Induces Resistance to Doxorubicin in HER2+ Breast Cancer by Promoting Free Fatty Acid Synthesis.* Front Oncol, 2021. **11**: p. 629640.

36. Yao, J., et al., *UCHL1 acts as a potential oncogene and affects sensitivity of common anti-tumor drugs in lung adenocarcinoma.* World J Surg Oncol, 2022. **20**(1): p. 153.

37. Xia, L., et al., *CNN3 acts as a potential oncogene in cervical cancer by affecting RPLP1 mRNA expression.* Sci Rep, 2020. **10**(1): p. 2427.

38. Zhao, H., et al., *High Expression Levels of AGGF1 and MFAP4 Predict Primary Platinum-Based Chemoresistance and are Associated with Adverse Prognosis in Patients with Serous Ovarian Cancer.* J Cancer, 2019. **10**(2): p. 397-407.

39. Mahauad-Fernandez, W.D., et al., *Bone marrow stromal antigen 2 expressed in cancer cells promotes mammary tumor growth and metastasis.* Breast Cancer Res, 2014. **16**(6): p. 493.

40. Parte, S., et al., *PTTG1: a Unique Regulator of Stem/Cancer Stem Cells in the Ovary and Ovarian Cancer.* Stem Cell Rev Rep, 2019. **15**(6): p. 866-879.

41. Nakachi, I., et al., *PTTG1 Levels Are Predictive of Saracatinib Sensitivity in Ovarian Cancer Cell Lines.* Clin Transl Sci, 2016. **9**(6): p. 293-301.