**Intro to Bioinformatics – 236523**

**Final Project – TODO add a cool title**

Tomer Goldstein – 3rd year CS student

Eden Doron - 3rd year CS student

Maybe put an image?

**Abstract**

**Introduction**

Epithelial ovarian cancer represents one of the major causes of cancer death among women and certainly the most lethal gynecological cancer. Epithelial ovarian cancer tumor is generally diagnosed at late stage when the tumor is disseminated throughout the peritoneal cavity, limiting the potential benefit of debulking surgery (Damia and Broggini, 2019). Ovarian cancer is the fifth leading cause of cancer deaths in women in the United States, with an incidence of 23,000 new cases and 14,000 deaths annually (Schaner *et al.*, 2003). Contributing to the poor prognosis is the lack of symptoms in the early stages of the disease. More than 75% of diagnoses are made in stage III and IV, after distant metastasis has occurred. The 5-year survival rate for women diagnosed with late-stage disease is 25%, compared to more than 90% for women diagnosed with stage I of the disease (Hibbs *et al.*, 2004). Carcinomas of the surface epithelium of the ovary comprise the large majority (80–90%) of ovarian cancers. Among these epithelial cancers, the most common morphological subtype is serous papillary, with less common subtypes including clear cell, mucinous, endometrioid, transitional, and undifferentiated (Schaner *et al.*, 2003).

The application of DNA microarray technology has enabled the study of gene expression profiles of large numbers of tumor samples and has provided an opportunity to classify different neoplasms based on characteristic expression patterns (Schaner et al., 2003). In recent years, large-scale gene expression analyses have been performed to identify differentially expressed genes in ovarian carcinoma. A common goal of these studies was to identify potential tumor markers for the diagnosis of early-stage ovarian cancer, as well as to use these markers as targets for improved therapy and treatment of the disease during all stages. These earlier studies compared the gene expression profiles of tissues or cell lines derived from ovarian cancer samples, normal ovaries, other normal samples, and other types of tumors. A major problem in identifying genes up-regulated in ovarian carcinoma is that normal ovary epithelial cells are very difficult to obtain in large enough numbers to perform gene microarray experiments. Although some groups have analyzed gene expression of the cells that are on the surface of normal ovaries, it is still controversial whether these cells truly serve as the normal counterpart for ovarian epithelial tumors. The cumulative results of these gene expression studies reveal more than 150 potentially up-regulated genes that are associated with ovarian cancer. However, only a small portion of the genes reported as up-regulated in ovarian carcinoma were further validated by a second technique such as immunohistochemical analysis or reverse transcriptase polymerase chain reaction. (Hibbs *et al.*, 2004)

Therapeutic potentials of metal-based compounds date back to ancient time.1 During this period, the ancient Assyrians, Egyptians and Chinese knew about the importance of using metal-based compounds in the treatment of diseases,1 such as the use of cinnabar (mercury sulfide) in the treatment of ailments.1 The advent of “theoretical science”, by Greek philosophers (Empedocles and Aristotle) in the 5th and 4th century BC, boosted the knowledge of metal-based compounds as therapeutic agents. This was supported by the information handed down by Pliny and Aulus Cornelius Celsus (Roman physicians) on the use of cinnabar in the treatment of trachoma and venereal diseases. In the 9th and 11th century BC, the contributions of ancient scientists such as Rhazes (Al-Razi) and Avicenna (Ibn Sina) were applauded, sequel to the discovery of toxicological effects of mercury in the animals and the use of mercury (quicksilver ointment) for skin diseases respectively. Arsenic trioxide (ATO) was used as an antiseptic and in the treatment of rheumatoid diseases, syphilis and psoriasis by traditional Chinese medical practitioners. Certainly, ATO was among the first compounds suggested for use in the treatment of leukemia during 18th and 19th centuries, until in the early 20th century when its use was replaced by radiation and cytotoxic chemotherapy. Therapeutic use of gold and copper can be traced to the history of civilization, where the Egyptians and Chinese were famous users in the treatment of certain disease conditions, such as syphilis. The discovery of platinum compound (cisplatin) by Barnett Rosenberg in 1960s was a milestone in the history of metal-based compounds used in the treatment of cancer. This forms the foundation for the modern era of the metal-based anticancer drug. Despite the wide use of the metal-based compounds, the lack of clear distinction between the therapeutic and toxic doses was a major challenge. This was so because practitioners of ancient time lack adequate knowledge of dose-related biological response. The advent of molecular biology and combinatorial chemistry paves the way for the rational design of chemical compounds to target specific molecules. (Ndagi, Mhlongo and Soliman, 2017)

Recently, there has been a growing demand for metal-based compounds in the treatment of cancer. Platinum compounds, particularly cisplatin, are the heart-beat of the metal-based compounds in cancer therapy. Clinical use of platinum complexes as an adjuvant in cancer therapy is based on the desire to achieve tumor cell death and the spectrum of activity of the candidate drug. Such complexes are mostly indicated for the treatment of cervical, ovarian, testicular, head and neck, breast, bladder, stomach, prostate and lung cancers. (Ndagi, Mhlongo and Soliman, 2017). For the last three decades, platinum-based chemotherapy has been the cornerstone of systemic treatment for EOC. Standard front-line treatment for advanced EOC consists of cytoreductive surgery, with the goal of no residual disease (R0), and platinum-based chemotherapy. At the time of relapse, if the time interval since the last dose of platinum chemotherapy (Treatment Free Interval-Platinum; TFIp) is more than six months, the disease is considered platinum sensitive, and standard treatment consists of re-challenge with platinum. The majority of high-grade serous ovarian cancers (HGSC) are initially platinum sensitive. However, even with optimal treatment, HGSC will typically follow a frequent relapse-response pattern, before eventually becoming platinum resistant (McMullen *et al.*, 2020).

Gene expression profiles reflect distinct biological states with associated features, and therefore one hypothesis is that gene expression subtypes might also associate with drug response. A stratification of ovarian cancer using gene expression profiles could therefore be utilized for prediction of an individual's response to platinum treatment. In short, such stratification could help us to predict those patients who might exhibit a genuine benefit from platinum, and conversely, those who are likely to exhibit resistance (Murakami *et al.*, 2016). In this project, we aim to study a potent organo-osmium compound with improved activity over cisplatin (platinum) and no cross-resistance in platinum-resistant cancers. This compound disrupts metabolism in A2780 human ovarian cancer cells, generating reactive oxygen species and damaging DNA (Hearn *et al.*, 2015). We will attempt to learn more about the behavior of this compound by using the different analysis methods we learned and implemented during the course. By doing so, we might get a glimpse of future possibilities of improved therapeutic strategies regarding treatment of ovarian cancer.

During our search, we have found a very large number of datasets relevant to our study, therefore we have decided to limit our table to 10 datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Title** | **No. of Samples** | **No. of Patients** | **No. of Controls** | **Cell Type** | **Findings** |
| Genome wide DNA methylation profiling of United Kingdom Ovarian Cancer Population Study | 540 | 274 | 266 | Ovarian cells | - |
| Survival Related Profile, Pathways and Transcription Factors in Ovarian Cancer | 157 | 157 | 0 | epithelial ovarian cancer cell | identified an 86-gene overall survival gene expression profile that seems to predict overall survival for women with advanced serous ovarian cancer. |
| Neoadjuvant chemotherapy modulates T cell responses in high-grade serous Ovarian Cancer metastasess | 35 | - | - | Omental tissue samples | neoadjuvant chemotherapy enhances anti-cancer responses of T cells in peritoneal metastases of patients with high-grade serous Ovarian Cancer but does not decrease levels of immune checkpoint molecules, providing a rationale for sequential chemo-immunotherapy. |
| RNA-seq of coding RNA in human Ovarian Cancer cell lines exposed to a new osmium-based anticancer agent | 30 | 12 | 18 | epithelial ovarian cancer cells | Identified mutations in complex I of the electron transport chain in A2780 cells and suggest that the osmium compound may exploit these mutations to exert a potent mechanism of action. |
| Resistance of primary Ovarian Cancer cells to oncolytic adenoviruses | 124 | - | - | epithelial Ovarian Cancer cell line |  |
| Activation of phosphatidylcholine-cycle enzymes in human epithelial Ovarian Cancer cells | 28 | 28 | 0 | epithelial ovarian cancer cells & Ovarian Surface Epithelial | demonstrated that the elevated PCho pool detected in EOC cells primarily resulted from the upregulation/activation of ChoK and PC-plc involved in the biosynthetic and in a degradative pathway of the PC-cycle, respectively. |
| CHAC1 mRNA expression is a strong prognostic biomarker in breast and Ovarian Cancer | 12 | 12 | 0 | breast cancer cell lines & epithelial Ovarian Cancer cell line | Univariate analysis in Ovarian Cancer showed that CHAC1 mRNA expression above the median was associated with a poor relapse free survival (p=0.03). In younger Ovarian Cancer patients (age < median age), a high CHAC1 mRNA expression was associated with overall survival (p=0.007) and relapse free survival (p=0.015). |
| Effect of a dual specificity PI3K/mTOR inhibitor, NVP-BEZ235, on human Ovarian Cancer cell lines grown in 3D culture | 24 | 12 | 12 | ovarian cell lines | demonstrate that acute adaptive response to PI3K/mTOR inhibition resembles well-conserved adaptive response to nutrient and growth factor deprivation and how development of rational drug combinations can bypass resistance mechanisms. |
| Wild Type Tumor Repressor Protein 53 (TRP53) Promotes Ovarian Cancer Cell survival | 3 | 3 | 0 | ovarian surface epithelial (OSE) cells | suggest that activation of TP53 may provide a promising new therapy for managing type I Ovarian Cancer and other cancers in humans where wild-type TP53 is expressed. |

Table 1 - Ovarian Cancer Datasets

**Results**

Differential Expression Analysis

We decided to perform this specific analysis because we wanted to test the connection between the up\down-regulated genes of the ovarian epithelial cell lines and the treatment of the osmium-based compound according to the duration of the experiment.

In the tutorials in class we have learned that the best way to achieve these insights is by performing the differential expression analysis, this stems from the fact that the analysis allows us to differentiate the genes according to chosen values. which in our case are the control\treatment and the timepoints.

To perform the analysis, first we had to clean and organize the data from the dataset (Hearn *et al.*, 2015) run the DESeq method on the data, normalized it via vst and visualized it using a heatmap.

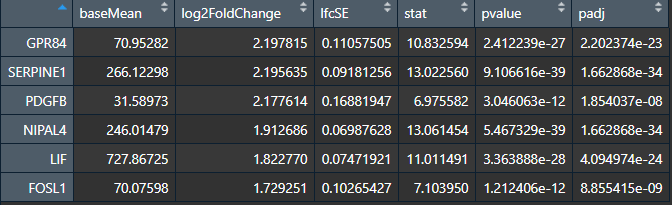
We found the following results (Table 2) that portray the top up-regulated genes (GPR84, SERPINE1, PDGFB, NIPAL4, LIF, FOSL1) in the cancerous cells:

Table 2 – Summary Up-Regulated Genes

Chart

Description automatically generated

Figure 1 – Heatmap of up-regulated genes by control vs treatment and time sequenced

From figure 1, we can see that after 48 hours, the osmium-based compound affected the regulation of the treated subjects’ cells when compared to the control patients, where these same genes were down-regulated in the same time.

Moreover, it is noticeable that the osmium-based compound affected the above genes in an extreme way (compared to the control) as soon as 4 hours after administration, this result validates the findings of the Hearn et al paper from which we took the dataset.

We can therefore deduce that the osmium-based compound treatment does indeed influences the patients’ cancerous cells since the genes we found have a much greater presence in the treated cells than in those of the control patients.

After checking the original paper’s data analysis results we have found that our up-regulated genes conform with the ones found in the original results.

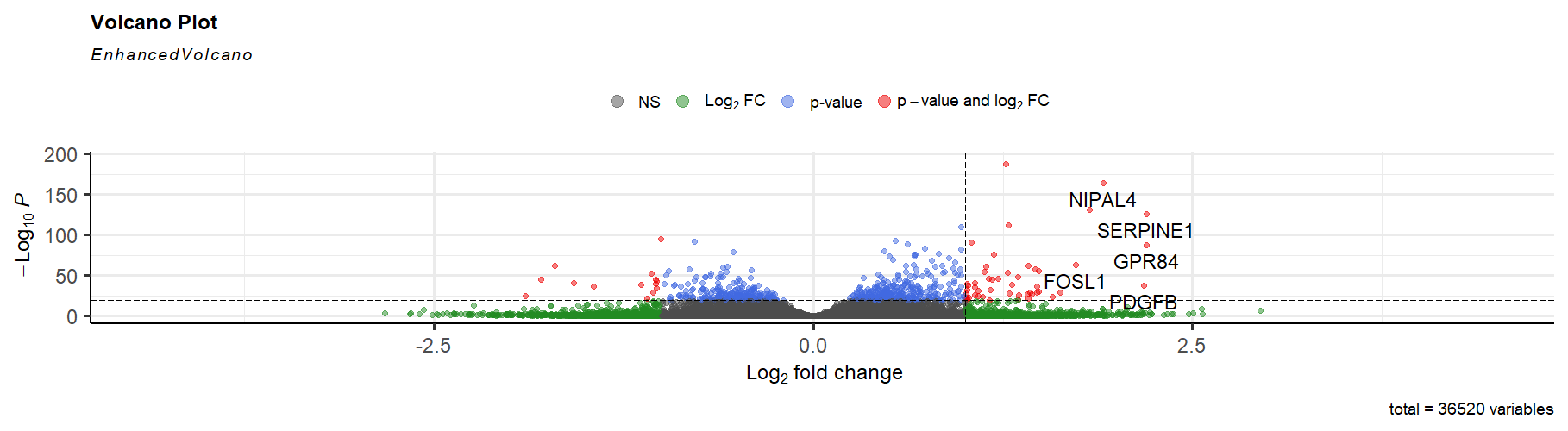
Additionally, visualizing the data using the enhancedVolcano library (figure 2), we can clearly see the same up-regulated genes appear in the cutoffs of both p-value= (which implies the high significance of these genes) and <1.

Figure 2 - Volcano Plot with Top Up-regulated Genes

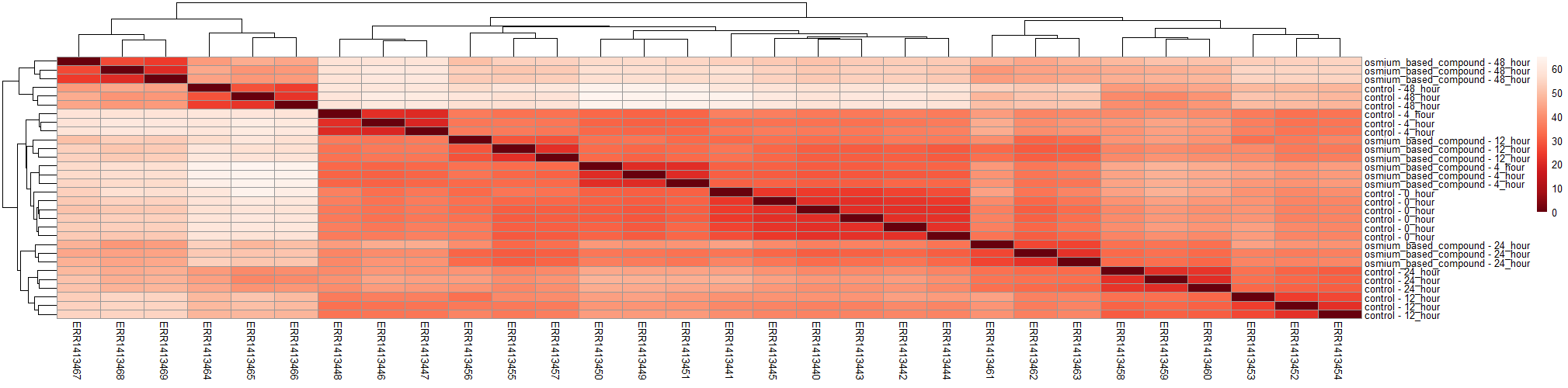
While working & researching the dataset we have also created a dendrogram (figure 3) of the data in regards of the time sequenced samples of both the control and treated patients. This figure further validates the paper’s & our results regarding the effect the osmium-based compound had on the cancerous cells w.r.t time elapsed.

Figure 3 - Dendrogram of Time-Sequenced Samples by control/treated samples

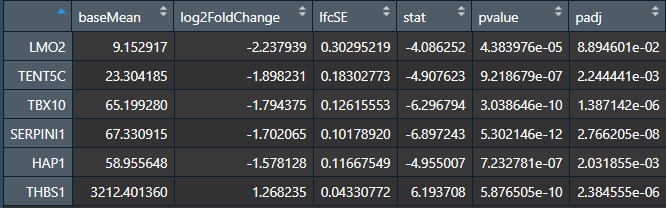
During the differential expression analysis we have also examined the top down-regulated genes as seen in table 3, which supports the original paper’s analysis report.

Table 3 - Summary Down-Regulated Genes

Even though these genes’ p-values are low enough to further investigate them, we have decided to stick with up-regulated genes for further analysis since they have much lower p-values which makes them more significant for our research.

**Discussion**

**Methods**

Data Processing (cleaning & organizing)

In order to work with the dataset found in (<https://github.com/edendoron/bioinformatics-project>), we first had to load the data to the RStudio environment using the read.csv function. Additionally, we created a dictionary that maps between a gene’s symbol to it’s name in order to make our data visualization figures clearer.

We also filtered the data by removing the samples ids, leaving us with a clean, integer values matrix.

Furthermore, we downloaded and loaded the experiment design provided in order to map the samples ids to control or treatment types and their correct time point. After loading the two files, we need to sort the raw data matrix so the samples ids will be in the same order as in the experiment design (this was needed for performing differential expression).

Differential expression analysis

By using the DESeq2 library we performed differential gene expression analysis like we saw in the tutorials. We used the DESeqDataSetFromMatrix library function in order to create a DESeqDataSet object from the matrix we built previously. In order to use the DESeq function we had to remove rows with less than 1 count.

Next, we normalized the DESeqDataSet object by using variance stabilizing transformation built-in function to stabilize the variance across the mean. Now we could finally perform the analysis itself by calling the DESeq function on our DESeqDataSet object.

With the results from the analysis, we found both the top up-regulated genes and the top down-regulated genes across the different samples by filtering the results by applying a threshold of 1. Additionally we took a subset of those genes with a padj value smaller than 0.1 to get the most significant genes.

Heatmap

To get a better visuzlation of our most significant genes in regards to control vs treatment samples with respect to their time points, we decided to create 2 heatmaps using the “pheatmap” & “RColorBrewer” libraries. In the first heatmap we used the clean data by calculating the distances between the genes across the samples as a metric. Finally we used the colorRampPalette to give the heatmap better colors to help observe the data.

In the second heatmap we used the scaled data and added annotation to better see the differences between the samples against the top up-regulated genes with respect the time elapsed and whether or not the sample was treated with osmium-based compound or not.

Volcano Plot

We used the “EnhancedVolcano” library to create a volcano plot visualization of the differential expression analysis results which displays the cell’s genes with the top up-regulated genes (GPR84, SERPINE1, PDGFB, NIPAL4, FOSL1) marked for clarity and also added a p-value cutoff of and a FC cutoff of 1.

**References**

Damia, G. and Broggini, M. (2019) “Platinum Resistance in Ovarian Cancer: Role of DNA Repair,” *Cancers*, 11(1). doi: 10.3390/cancers11010119.

Hearn, J. M. *et al.* (2015) “Potent organo-osmium compound shifts metabolism in epithelial ovarian cancer cells,” *Proceedings of the National Academy of Sciences*, 112(29), p. E3800. doi: 10.1073/pnas.1500925112.

Hibbs, K. *et al.* (2004) “Differential gene expression in ovarian carcinoma: identification of potential biomarkers,” *The American journal of pathology*, 165(2), pp. 397–414. doi: 10.1016/S0002-9440(10)63306-8.

McMullen, M. *et al.* (2020) “Overcoming Platinum and PARP-Inhibitor Resistance in Ovarian Cancer,” *Cancers*, 12(6). doi: 10.3390/cancers12061607.

Murakami, R. *et al.* (2016) “Prediction of taxane and platinum sensitivity in ovarian cancer based on gene expression profiles,” *Gynecologic Oncology*, 141(1). doi: 10.1016/j.ygyno.2016.02.027.

Ndagi, U., Mhlongo, N. and Soliman, M. (2017) “Metal complexes in cancer therapy &ndash; an update from drug design perspective,” *Drug Design, Development and Therapy*, Volume11. doi: 10.2147/DDDT.S119488.

Schaner, M. E. *et al.* (2003) “Gene expression patterns in ovarian carcinomas,” *Molecular biology of the cell*. 2003/09/05, 14(11), pp. 4376–4386. doi: 10.1091/mbc.e03-05-0279.