2022

COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF CANCER TESTIS GENES IN OVARIAN CANCER

Zayne Knuth
Michigan Technological University, ztknuth@mtu.edu

COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF CANCER TESTIS GENES IN OVARIAN CANCER

By

Zayne T. Knuth

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Biological Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2022

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Biological Sciences.

Department of Biological Sciences

Thesis Advisor:	*Paul Goetsch*
Committee Member:	*Guiliang Tang*
Committee Member:	*Hairong Wei*
Department Chair:	*Chandrashekhar Joshi*

Table of Contents

List of Figures

List of Tables

Author Contribution Statement

Let it be known that Caleb Hiltunen and Paul Goetsch were instrumental in the standardization and renaming of the data files used. Both were also responsible for the development of the metafile used for the cancer testis gene analysis.

Acknowledgements

I would first like to thank my family for always being supportive of me and my endeavors. It is because of you that I have made it this far and have earned this great achievement. Knowing that you always have my back has been a lifesaver. I love you all. I would next like to thank my advisor, Dr. Paul Goetsch, for guiding me through the dense web that is bioinformatic analysis, and for giving me the opportunity to pursue this topic. I have learned so much and have come much farther than I originally intended, and for that I am grateful. Finally, I would like to thank my friends for always being there during this stressful time of my life. You all mean more to me than I can ever say, and I am forever grateful for your unconditional love and support.

Abstract

Cancer testis genes are common targets for the development of immunotherapy for cancer treatment. Ovarian cancer is one of the leading causes of death in women cancer patients. Cancer testis genes play a role in tumorigenesis, but it is not clear how these genes are activated. This study utilized differential expression analysis between The Cancer Genome Atlas (TCGA) ovarian cancer data, Genotype-Tissue Expression (GTEx) non-cancerous ovary and testis data, and cell line data to identify a list of cancer testis genes that have a novel expression profile. To identify ovarian cancer testis genes, we obtained normal ovary tissue data and normal testis germline data from GTEx and cancerous ovarian tissue samples from TCGA. We defined a cancer testis gene if it matched an expression pattern of being differentially expressed in cancer and germline samples, as compared to normal tissue samples. The analysis discovered 4,578 genes that satisfied the condition of cancer testis genes. To facilitate downstream mechanistic analyses aimed to evaluate how these genes are misregulated, we identified 87 genes that satisfied the conditions in common ovarian cancer cell lines, as compared to human foreskin fibroblast normal cells. There were 28 genes found in common between both these lists that met our target expression profile whose misexpression will be evaluated in future studies.

1 Introduction

This experiment aims to further investigate how cancer testis (CT) genes are misregulated in cancer. A key feature of oncogenesis is the switch of a cell from its differentiated type to a germline-like state, a process where CT genes are known to be involved in [1]. Previous analyses using presence/absence methods reveal a stochastic model by which germline genes are activated in some cancers [1]. We hypothesize that differential analysis will reveal novel information to the CT genes involved in this process. In addition to testing tissue level data, we are testing our novel pipeline using cell line data from the BJ/HFF, SKOV3, and OVCAR3 lines as part of the filtering process for CT genes. SKOV3 and OVCAR3 are known models for ovarian cancer study, making our analysis consistent with other analyses.

1.1 Cancer testis genes

Cancer testis (CT) genes were first discovered in 1991 when van der Bruggen noticed that cytolytic T lymphocytes were able to target the tumor cells [2]. Since then, it has been noted that up to 40% of cancers express CT genes [3]. CT genes are designated as genes that are expressed in testis and cancerous cells but are not expressed in normal cells [4]. As of 2017, there have been nearly 800 CT genes identified [5,6]. These genes have been classified into two groups: CT-X and non-x CT genes, standing for CT genes found on the X chromosome and CT genes found on autosomal chromosomes [5,7]. Of the nearly 900 genes on the X chromosome, 10% of them have been designated as CT genes [5].

Despite the identification of many CT genes, the function of many is unknown [8]. The difficulty arises from many factors but primarily that CT genes are not well

conserved or that they lack known motifs to identify function. With this being said, CT genes have been proven to contribute to tumorigenesis in *Drosophila melanogaster* [9]. Many studies have shown that individual CT genes are oncogenic in certain cancers [10 – 52]. Knowing that there are CT genes that play a role in tumorigenesis, the discovery of novel CT genes could provide a breakthrough in immunotherapy development.

1.2 Cancer testis antigens and immunotherapy

While there have been numerous CT genes identified, not every CT gene is useful in immunotherapy. A CT gene is useful in immunotherapy if that gene encodes for an antigen that is expressed by a tumor cell [3]. If a cancer testis antigen (CTA) is expressed by the tumor cell, a vaccine or genetically engineered cytolytic-T cell could be used to treat the cancer [2, 7, 53 – 56]. Without an antigen to target, it is very difficult to perform treatment specific to the cancerous cells, which is why when a CT gene is discovered, it is imperative that the gene is studied to determine if it encodes for an antigen.

There have been a couple of molecular methods that have shown how CT antigens become expressed in cancer cells. Among these are the DNA hypomethylation of promoters, and histone marker modification and modulation that results in the loss of repressive histone marks and/or gain of activating histone marks [57]. Specifically for DNA hypomethylation, it is not the methylation of the CTA promoter that occurs, but rather the inhibition of a promoter of a regulatory element that suppresses the CTA gene. It is not known which regulatory elements are responsible for these gene suppressions and is likely to be specific to each CTA gene.

1.3 Ovarian cancer and CT genes

As of 2022, ovarian cancer is the fifth leading cause of cancer deaths in women and is the most lethal cancer in women [58]. It is estimated that nearly 20,000 new women will be diagnosed with ovarian cancer in 2022, of which 12,000 are expected to die. Ovarian cancer is defined as a group of tumors that grow in the ovary but are not necessarily originated in the ovary [59]. Nearly 90% of all ovarian cancer diagnoses are epithelial ovarian cancers, which can be subdivided into five histological types: high-grade serous, low-grade serous, endometrioid, clear cell carcinoma, and mucinous carcinomas. Cancer testis genes are abundantly found in various ovarian cancer types, with most of them being expressed in epithelial ovarian cancer [60]. Xie et al identified 38 CT genes to be expressed in ovarian cancer. A second study elaborated to indicate that in ovarian cancer the CT genes MAGEB1, MAGEB2, GAGE1, NY-ESO1 have an increased expression compared to other cancer types while the MAGEA4, GAGE3, GAGE4, XAGE3, SSX2, SCP1, and PRAME1 genes saw decreased expression [61]. While the number of ovarian cancer diagnoses are declining over the years, any new revelations in treatment targets is welcome to attempt to increase survivability.

1.4 New resources available

With the advancement of next generation sequencing technology, the need for publicly available data has become paramount to advance the field in terms of accessibility and reproducibility. The Cancer Genome Atlas (TCGA) has emerged as a public source for genomic, epigenomic, transcriptomic, and proteomic data for 20,000 primary cancer and matched normal samples of 33 cancer types [62]. Similarly, the Genotype-Tissue Expression (GTEx) project is an ongoing effort to

build a public resource of tissue-specific gene expression patterns [63]. It is a collection of 54 non-diseased tissue sites from almost 1000 individuals of various molecular assays. Both resources were founded on the principal of creating publicly available data so that researchers could pursue their interests with data that has already been collected.

There have been many publications that utilize TCGA and GTEx for their sample gathering, including genome-wide association studies and single cancer analysis [5, 6, 60, 64]. For two of these papers, da Silva 2017 and Wang 2016, TCGA and GTEx data were used to identify cancer testis genes. These are our basis for which we developed the pipeline to pursue our study.

1.5 Differential analysis technique

In this study, differential expression analysis of TCGA and GTEx data will be used to detect CT genes, primarily utilizing the DESeq2 R package [65]. As with previously mentioned genome-wide identification studies [4, 5, 6], we will be using the DESeq2 and R to perform a large data sample analysis of ovarian cancer, ovarian normal, and testis normal genomic data to identify CT genes. In addition, our study will be intersecting the CT genes discovered with a list of genes obtained from differential expression analysis of three cell lines: BJ/HFF, SKOV3, and OVCAR3.

The three cell lines are being utilized to identify genes that are highly expressed in SKOV3 and OVCAR3 but not in the BJ/HFF line. BJ/HFF is being used as a control for normal cell gene expression [66], SKOV3 as a low-grade serous/non-serous epithelial line [59, 67], and OVCAR3 as a high-grade serous line [59, 68]. Each cell line will be compared to each other, and genes will be selected based on the criteria of taking genes that are more highly expressed in SKOV3 in the SKOV3-BJ/HFF

comparison, in OVCAR3 in the OVCAR3-BJ/HFF comparison, and OVCAR3 in the SKOV3-OVCAR3 comparison. The reasoning for this selection is based on the findings of Hallas-Potts et al in 2019 where they stated that the OVCAR3 line is clinically more aggressive than the SKOV3 line [59]. We hypothesize that the more aggressive cancer cell line will have a higher expression level of CT genes compared to the less aggressive cancer cell line.

Once the CT genes have been identified and the cell line genes exhibiting the described expression pattern, the lists will be intersected to determine the CT genes that also exhibit the desired expression pattern. Together, this analysis will reveal what CT genes are misexpressed in ovarian cancer cell line models, which will enable future studies into what drives the misexpression of these genes.

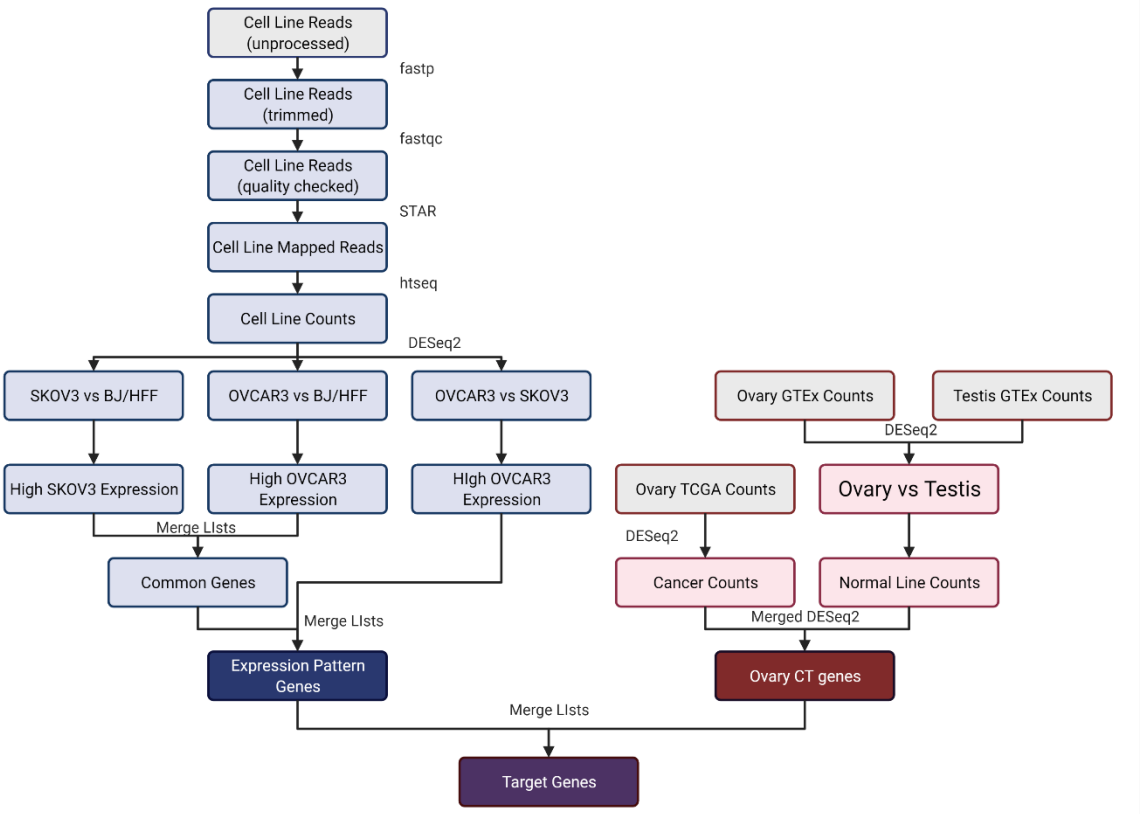

Figure 1-1 Pipeline of analysis

Depiction of the pipeline utilized for the generation of the final gene list. The right side depicts the process of the differential expression analysis of the ovary and testis data for the identification of the CT genes. The left pathway depicts the process to convert the raw reads for the three cell lines (BJ/HFF, SKOV3, OVCAR3) to the count data that was used in the differential expression analysis.

1.6 Goals and hypotheses

The goal of this study is to utilize a novel bioinformatic pipeline to discover new CT genes for ovarian cancer. While this study will not investigate further whether the discovered CT genes produce antigens, the hypothesis is that we will obtain a set of genes matching are combined expression pattern. The discovery of a novel set of genes will provide the scientific community with possible genes to pursue to be analyzed as potential immunotherapy targets, as well as additional genomic data that could aid in the early detection of ovarian cancer development.

2 Materials and Methods

2.1 Data utilized

For the CT gene identification analysis, 170 samples of normal ovary data from GTEx were utilized (Appendix A.1), 451 samples of ovary cancer data from TCGA (Appendix A.2), and 351 samples of normal testis data from GTEx (Appendix A.3).

The materials used for the cell line analysis include 3 samples of SKOV3 cell line data and 3 samples of OVCAR3 cell line data from GSE134375 in GEO [68]. The replicates for the SKOV3 lines were SRR9694244, SRR9694245, and SRR9694246. The replicates for the OVCAR3 line were SRR96942450, SRR9694251, and SRR9694252. The data for the BJ/HFF data was accessed from GEO with accession number GSE117808. It should be noted that the contributors of the data were Hemmerich P, Marthandan S, Huhne R, Groth M, and Platzer M in 2018, but there is no citation given for the data. The samples used were SRR7513003, SRR7513004, and SRR7513005. These data were RNAseq data and needed to be converted to HTSEQ data. This was done using the Ubuntu command line and in a Bioconda environment. The data files were downloaded and trimmed using the fastp command package [69]. The quality of each trimmed sample was assessed by running the fastqc command package[70]. The trimmed reads were mapped to the Gencode_GRch38.p13 reference genome using the STAR package[71]. The mapping files were then counted using the htseq-count command to generate the HTSEQ data files [72].

2.2 Differential expression analysis of the ovarian and testis tissue

The tissue sample data for the testis and ovaries were compared using the DESeq2 R package in a similar fashion to the previous section [73]. First the GTEx ovary data was compared to the GTEx testis data to establish a list of genes that were up-regulated in the testis compared to the ovary using a cut-off of a $\log_2$ score greater than 1. Next, the TCGA ovary data was compared to the GTEx ovary data to obtain a list of genes up-regulated in the cancerous ovary line compared to the normal line. The gene names of both sets were altered to remove gene identification data of different variations of the same genes. This was done to allow expression data of the same genes but different mutations to be overlapped. The two gene list datasets had their counts extracted from their sources and were compared on a differential expression basis to obtain a list of genes that met the criteria of being up-regulated in testis and cancerous ovarian lines compared to the normal ovarian tissue.

2.3 Differential expression analysis of ovarian cancer cell lines

The HTSEQ count files for the BJ/HFF, SKOV3, and OVCAR3 cell lines were analyzed for differential expression between the tissue samples using the DESeq2 R package. The control group used was the BJ/HFF cell line data since this cell line is not cancerous. Three analysis were conducted: one comparing the SKOV3 line to the BJ/HFF line, OVCAR3 line compared to BJ/HFF line, and the OVCAR3 line compared to the SKOV3 line. The results of the DESeq runs were extracted and the genes that

were differentially expressed by a $\log_2$ fold of 2 or greater were recorded. For the BJ/HFF to SKOV3 comparison the genes expressed higher in SKOV3 were recorded, For BJ/HFF and OVCAR3, the higher OVCAR3 genes expressed were recorded. Likewise, the OVCAR3 genes that were more highly differentially expressed compared to the SKOV3 genes were recorded. The intersection of the three gene sets were recorded to obtain a list of genes that were up-regulated in SKOV3 against BJ/HFF, up-regulated in OVCAR3 against BJ/HFF, and up-regulated in OVCAR3 compared to SKOV3.

2.4 Obtaining a list of CT genes

After both gene lists from the previous differential analysis were obtained, the intersection of the two gene lists was taken to create a final gene list that contained the CT genes in the ovarian samples that also had the expression pattern desired from the OVCAR3, SKOV3, and BJ/HFF analysis.

3 Results

3.1 CT gene analysis pipeline

We first performed a principle component analysis (PCA) of the testis, normal ovary, and cancerous ovary tissues (Figure 3-1). This analysis confirmed the quality of the data sets. Our analysis indicates that the data is of sufficient quality to continue with the rest of the pipeline without worrying about poor quality data biasing the analysis.

The first differential expression analysis was performed on the testis tissue and normal ovary tissue GTEx data. We plotted the $-\log_{10}$ p-adjusted values against the $\log_2$ fold change of the results of this analysis (Figure 3-2). There were 16,807 genes up-regulated with a $\log_2 > 2$ in the testis compared to the normal ovary tissues, as indicated by the teal dots. The quality of the analysis was further assessed with an MA plot to compare the $\log_2$ fold change of the samples to their base mean reads (Figure 3-3). The overall distribution confirmed the results in Figure 3-2 and indicated a clear distinction of the up-regulated genes in the testis. With the quality of the analysis confirmed, the 16,807 up-regulated testis genes were selected and advanced for further analysis.

To determine the genes up-regulated in the cancerous ovarian cells, ovary tumor samples were compared to the normal ovary tissue samples using differential expression analysis. The $-\log_{10}$ p-adjusted values were plotted against the $\log_2$ fold change values of each gene (Figure 3-4). There were 11,850 genes found to be up-regulated by a $\log_2$ fold factor > 2, depicted on Figure 3-4 as the purple dots. As done previously, an MA plot of the results was created to determine the quality and

confidence in the gene expressions (Figure 3-5). The distribution along the y-axis indicated that the genes expressed had a high confidence interval travelling up the side. The 11,580 genes were selected and advanced to the next analysis.

The final step of the analysis was identifying the genes in common that were up in the testis and the cancerous ovary tissues (Figure 3-6). The 16,807 up-regulated testis genes were intersected with the 11,580 up-regulated tumor genes. A total of 5,478 genes are shared between the two groups, indicated in the purple region. We defined these genes as ovarian CT genes, which were advanced to the next step of the analysis.

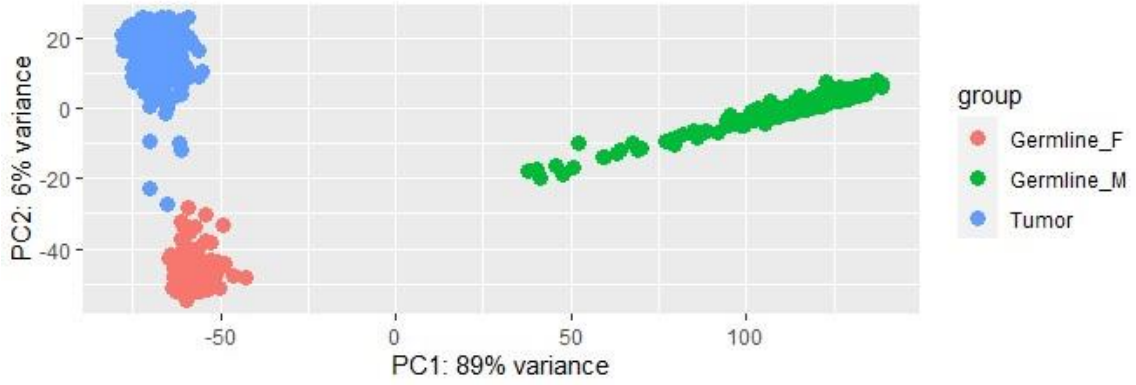

Figure 3-2 PCA plot of testis and ovary data

This figure depicts the variance between the tissue data for normal ovary (red), normal testis (green), and cancerous ovary (blue).

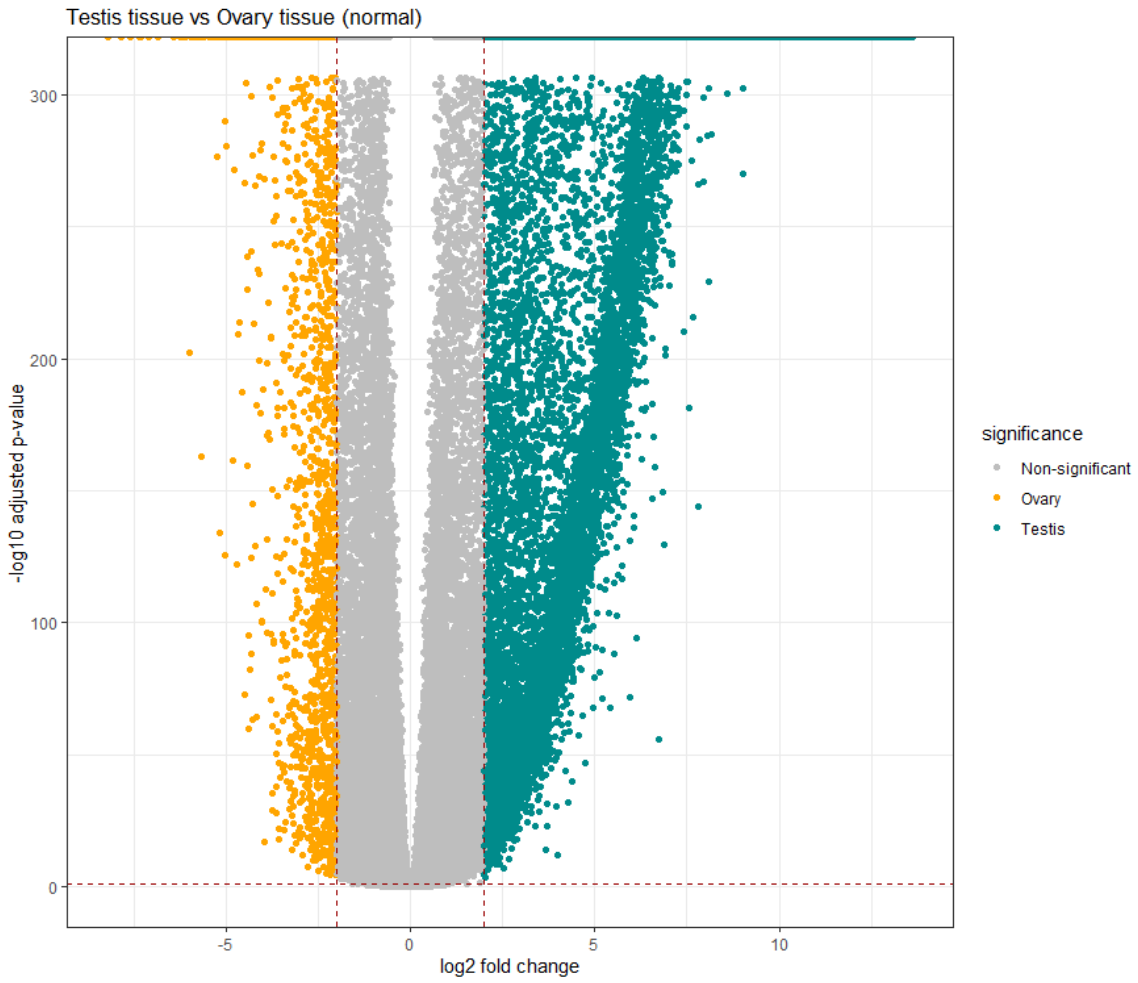

Figure 3-3 Volcano plot of testis tissue vs ovary tissue

This figure depicts the -$\log_{10}$ adjusted p-value score plotted against the $\log_2$ fold change in expression pattern between the testis tissue (teal) and the normal ovary tissue (orange). The grey dots show the genes that were not statistically differentially expressed by a factor of 2 on the $\log_2$ fold change scale.

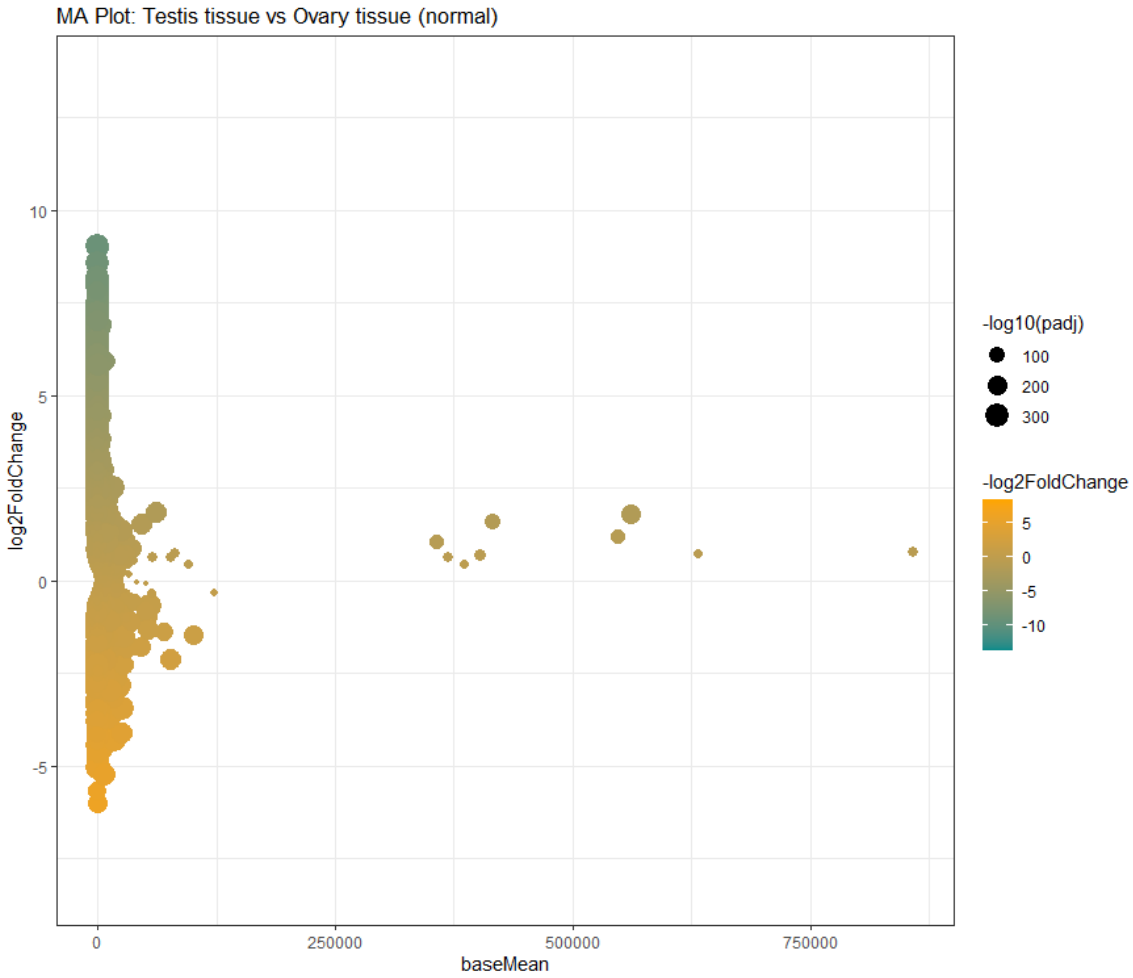

Figure 3-4 MA plot of testis tissue vs normal ovary tissue

This figure shows the gene expressions' $\log_2$ fold change plotted against the base mean counts of the data sets for the testis against ovary tissues. The orange dots represent the genes that were up--regulated in the ovary tissues, and the teal dots are the genes that were up-regulated in the testis tissues. The brighter the color is, the more differentially expressed the gene is. The size of the dots corresponds to the adjusted p-value of the differential expression analysis. Larger dots correspond to more confidence in the value plotted.

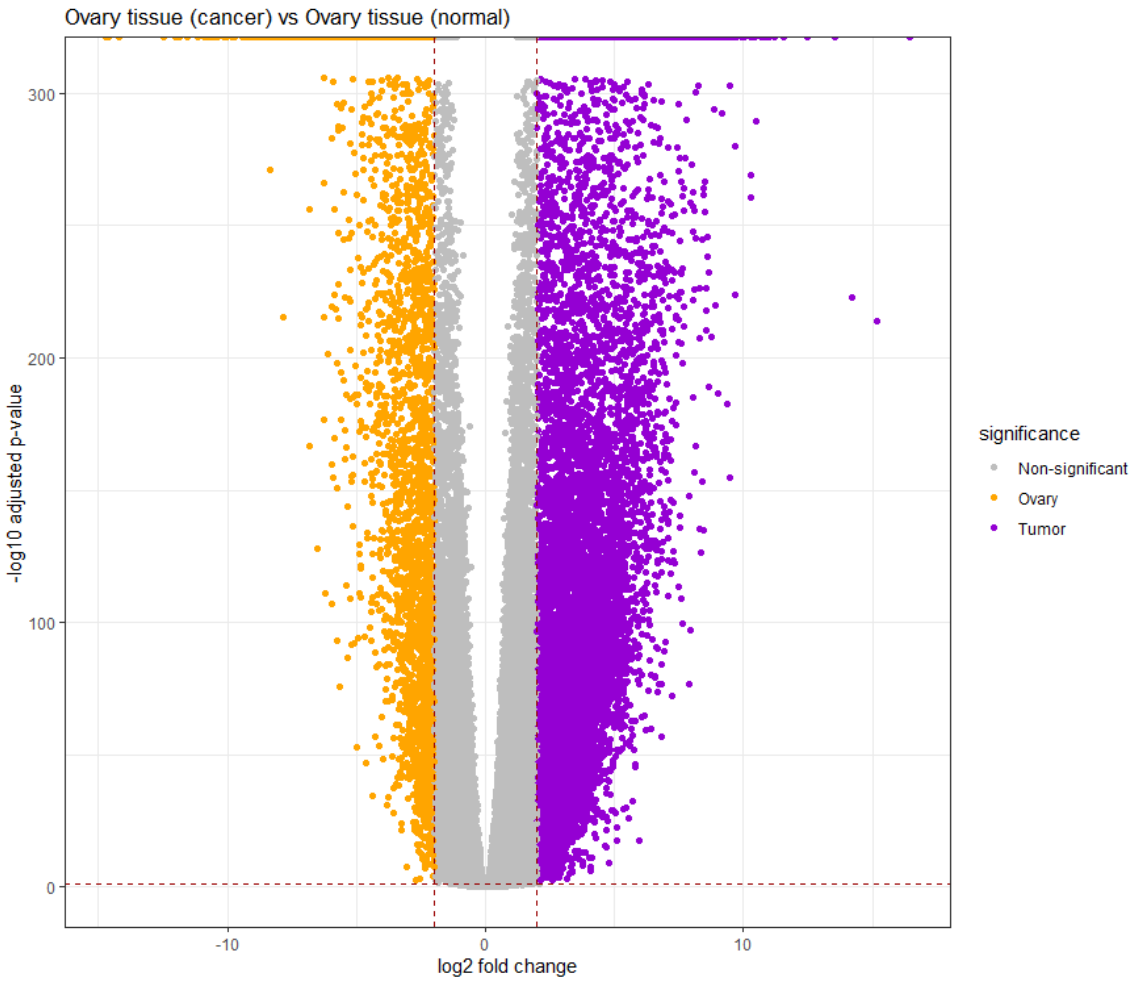

Figure 3-5 Volcano plot of cancerous ovary tissue vs normal ovary tissue

This figure depicts the $-\log_{10}$ adjusted p-value score plotted against the $\log_2$ fold change in expression pattern between the cancerous ovary tissue (purple) and the normal ovary tissue (orange). The grey dots show the genes that were not statistically differentially expressed by a factor of 2 on the $\log_2$ fold change scale.

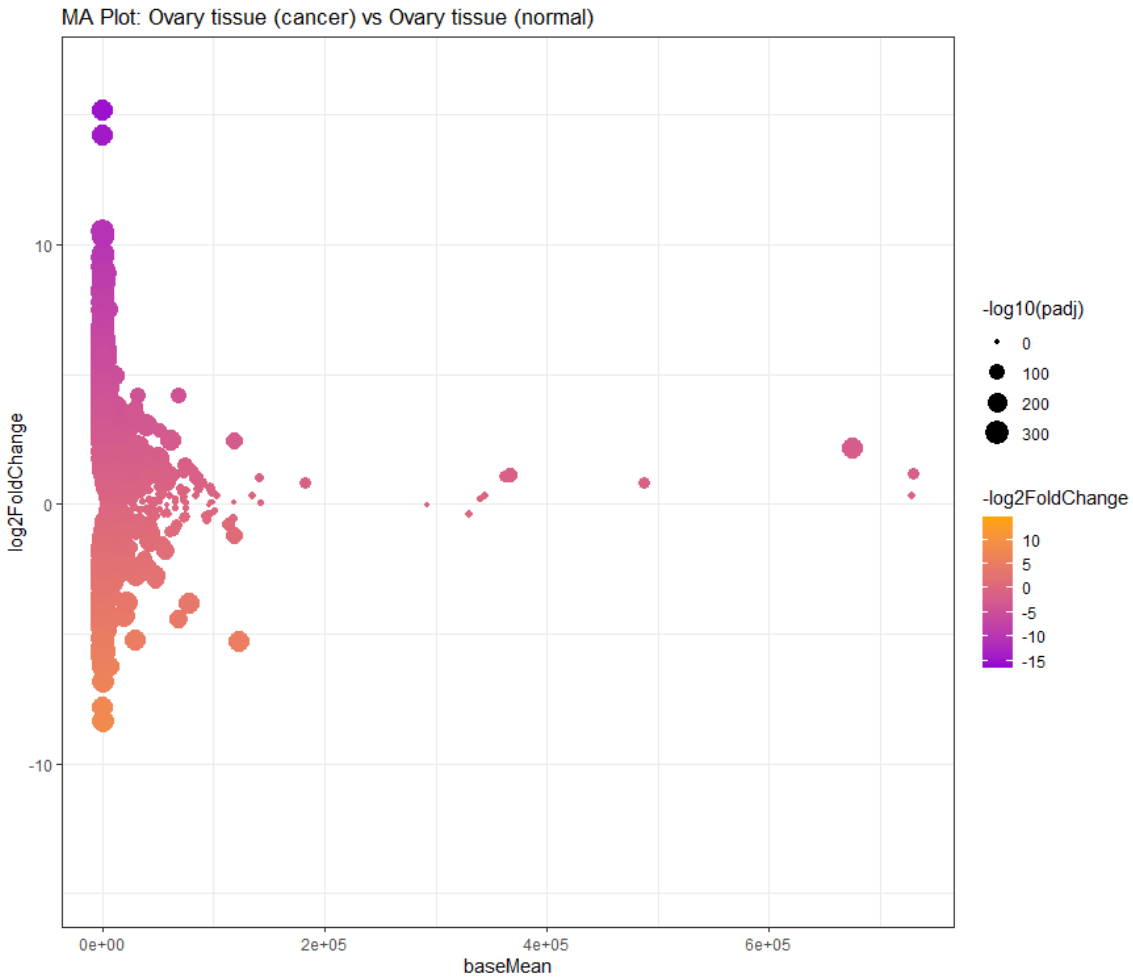

Figure 3-6 MA plot of cancerous ovary tissue vs normal ovary tissue

This figure shows the gene expressions' $\log_2$ fold change plotted against the base mean counts of the data sets for the cancerous ovary against normal ovary tissues. The orange dots represent the genes that were up--regulated in the normal ovary tissues, and the purple dots are the genes that were up-regulated in the cancerous ovary tissues. The brighter the color is, the more differentially expressed the gene is. The size of the dots corresponds to the adjusted p-value of the differential expression analysis. Larger dots correspond to more confidence in the value plotted.

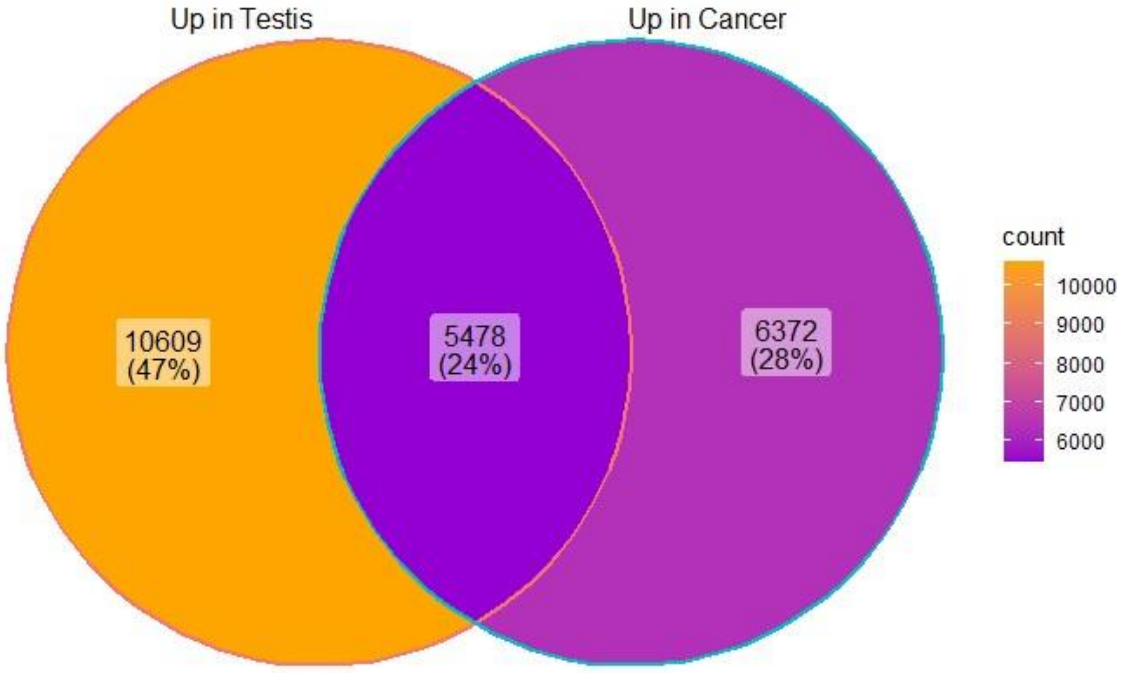

Figure 3-7 Venn Diagram of cancer testis genes

This figure depicts the number of genes found to be up-regulated in the following comparisons: 1) the left (orange) region shows the number of genes up-regulated exclusively in testis compared to normal ovary 2) the right (magenta) region shows the number of genes up-regulated exclusively in cancerous ovary tissue compared to normal tissue 3) the middle region (purple) shows the number of genes up-regulated in both comparisons and are referred to as cancer testis genes.

3.2 Cell line gene analysis pipeline

To identify what ovarian CT genes are activated in ovarian cancer cell lines, we evaluated differential gene expression in OVCAR3 and SKOV3 ovarian cancer cell lines compared to BJ/HFF normal cell lines. We performed principle component analysis (PCA) of the BJ/HFF, OVCAR3, and SKOV3 cell line transcriptome data (Figure 3-7). We observed very little variation between samples of the same cell line. We next evaluated our identified ovarian CT genes with a qualitative heatmap (Figure 3-8). There are distinct gene expression levels between cell lines and very little variation between samples of the same cell line, confirming with the Figure 3-7. With this knowledge, the analysis was continued.

The next step in the cell line analysis was the comparison of the SKOV3 and BJ/HFF cell lines. The results had the genes' $-\log_{10}$ p-adjusted value plotted against $\log_2$ fold change (Figure 3-9). There were 2,415 genes found to be up-regulated by a $\log_2$ fold change > 2 in the SKOV3 line, depicted by the teal dots. The quality was assessed with the MA plot to determine if the genes were differentially expressed enough to continue with the analysis (Figure 3-10). The distribution of the dots along the y-axis confirmed the confidence in the results. The 2,415 genes were forwarded to the next part of the analysis.

The second cell line comparison was between the OVCAR3 and BJ/HFF lines. The results were plotted with their $-\log_{10}$ p-adjusted against the $\log_2$ fold change (Figure 3-11). There were 2,414 genes that were up-regulated by a $\log_2$ fold change > 2 in the OVCAR3 line, depicted by the purple dots. An MA plot was created to verify the quality of the expressions (Figure 3-12). The points plotted traveling up

the y-axis indicated a high confidence in the gene expression levels. The 2,414 genes were selected and advanced to the next step in the analysis.

The final cell line comparison was between the OVCAR3 and SKOV3 lines. This comparison is testing a theory that there are genes in the OVCAR3 cell line that are more highly expressed in OVCAR3 than SKOV3. We base this in that OVCAR3 is a "worse" cancer in that it is more aggressive [59]. The results were similarly plotted with their $-\log_{10}$ p-adjusted values against their $\log_2$ fold change (Figure 3-13). There were 1,604 genes with a $\log_2$ fold change > 2 in the OVCAR3 cell line, depicted by the purple dots. The confidence in the expression patterns were confirmed by the MA plot (Figure 3-14). The 1,604 genes were advanced to the next step of the analysis.

The final step in the cell line analysis required determining if there were any genes with the expression pattern of being up-regulated in SKOV3 vs BJ/HFF, up-regulated in OVCAR3 vs BJ/HFF, and up-regulated in OVCAR3 vs SKOV3. The intersection of the three gene sets found in this analysis was conducted (Figure 3-15). There were 87 genes that were found to have the desired expression pattern (Table 3-1). These genes were passed along to the final step of our analysis.

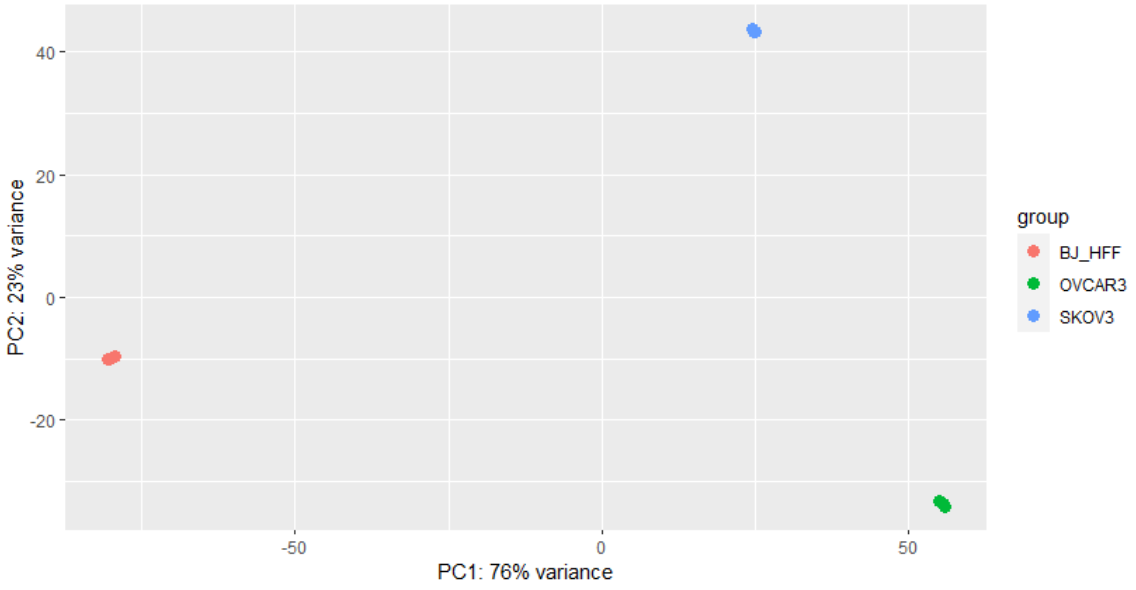

Figure 3-8 PCA plot of the BJ/HFF, OVCAR3, and SKOV3 cell lines

Pictured above are the 9 samples of the BJ/HFF, SKOV3, and OVCAR3 plotted against each other to show similarity and difference between the three different tissues, with the three BJ/HFF samples in orange, the three SKOV3 samples in blue, and the three OVCAR3 samples in green.

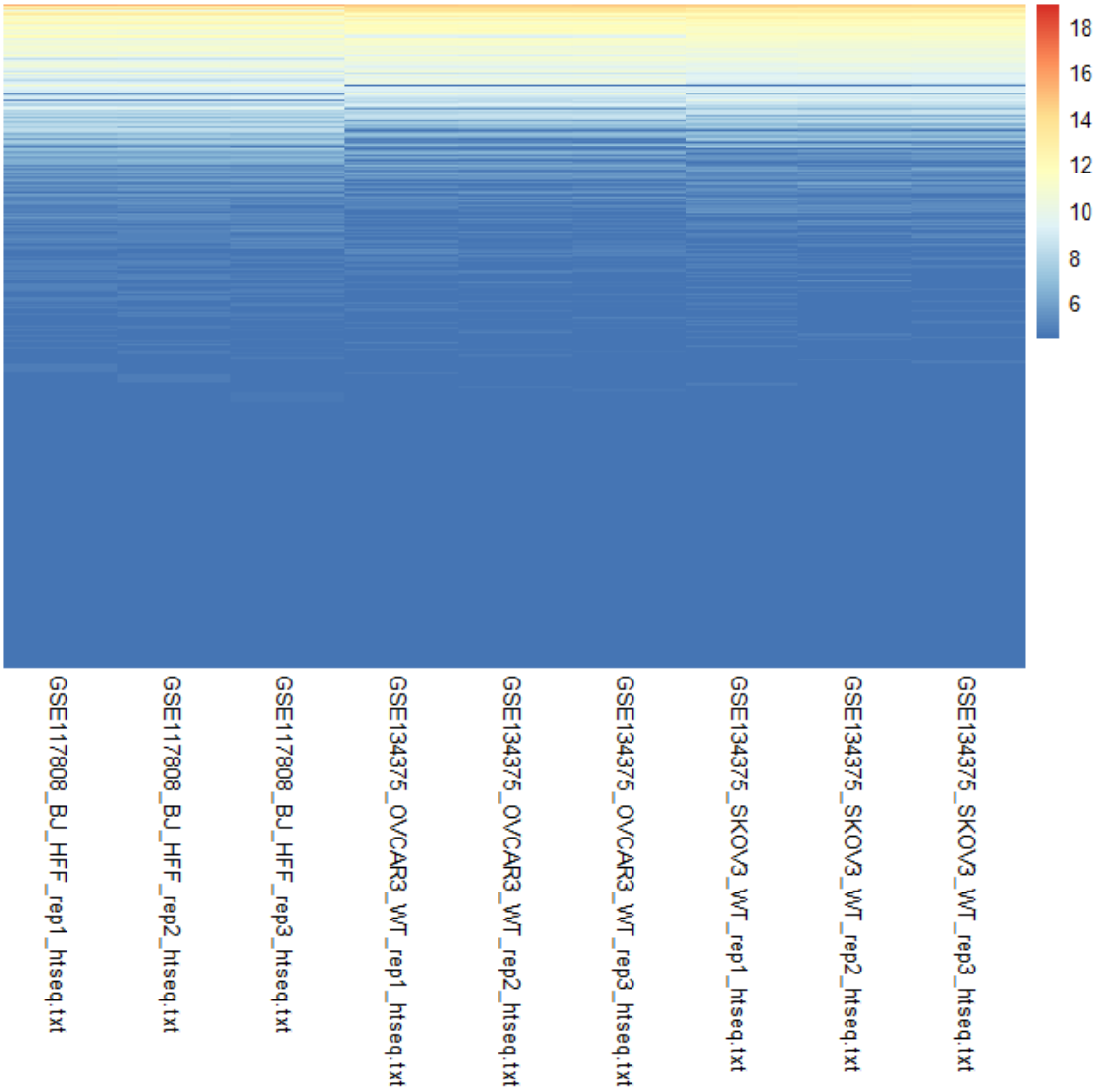

Figure 3-9 Heatmap of CT genes

This figure depicts the relative expression of the genes identified as cancer testis genes relative to the tissues of normal ovary, normal testis, and cancerous ovary. The first three samples on the left correspond to the BJ/HFF cell line, the middle three samples correspond to the OVCAR3 cell line, and the right three samples represent the SKOV3 cell line. A red color indicates a gene with a high expression level where blue represents a gene with a low expression level.

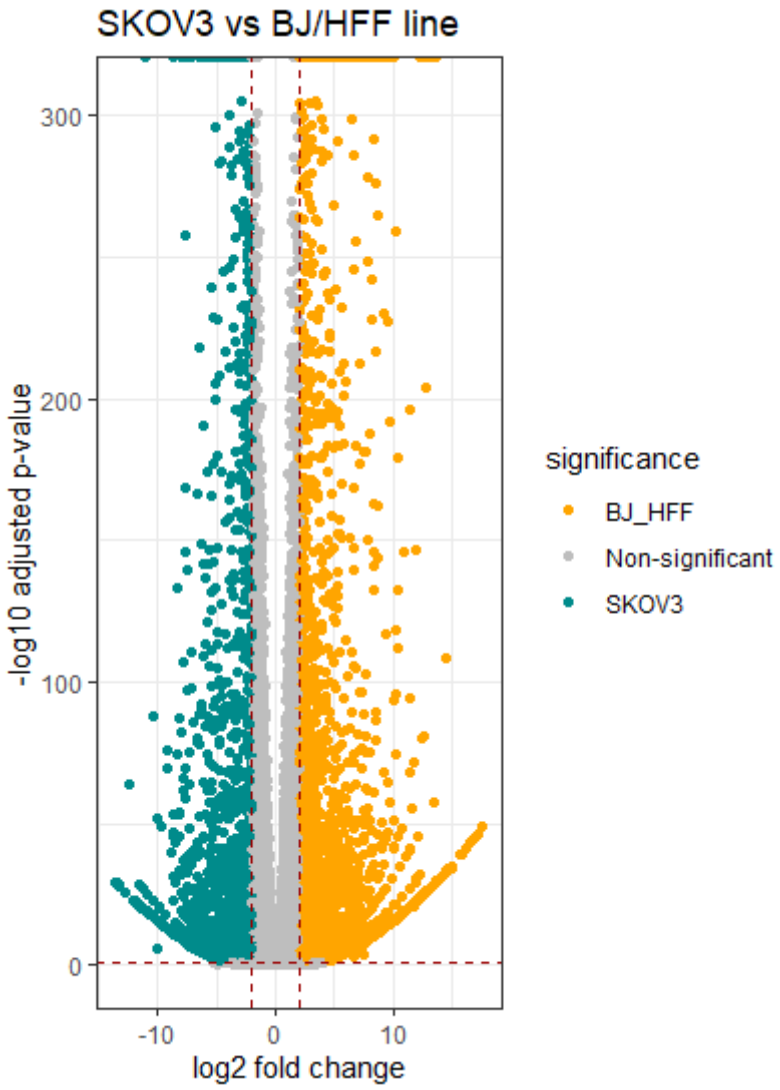

Figure 3-10. Volcano plot of differential expression of BJ/HFF line vs SKOV3 line

This figure depicts the $-\log_{10}$ adjusted p-value score plotted against the $\log_2$ fold change in expression pattern between the BJ/HFF cell line(orange) and the SKOV3 cell line(teal). The grey dots show the genes that were not statistically differentially expressed by a factor of 2 on the $\log_2$ fold change scale.

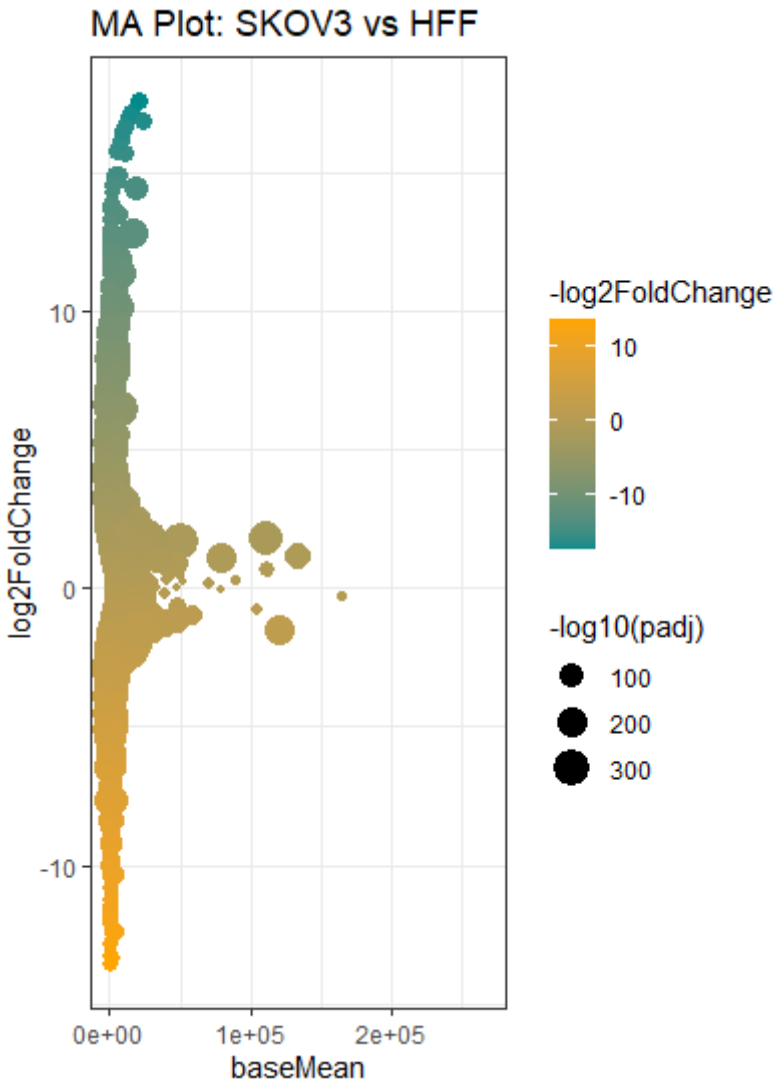

Figure 3-11. MA plot of the differential expression of the BJ/HFF and SKOV3 cell lines

This figure shows the gene expressions' $\log_2$ fold change plotted against the base mean counts of the data sets for the BJ/HFF against SKOV3 lines. The orange dots represent the genes that were up--regulated in the BJ/HFF lines, and the teal dots are the genes that were up-regulated in the SKOV3 line. The brighter the color is, the more differentially expressed the gene is. The size of the dots corresponds to the adjusted p-value of the differential expression analysis. Larger dots correspond to more confidence in the value plotted.

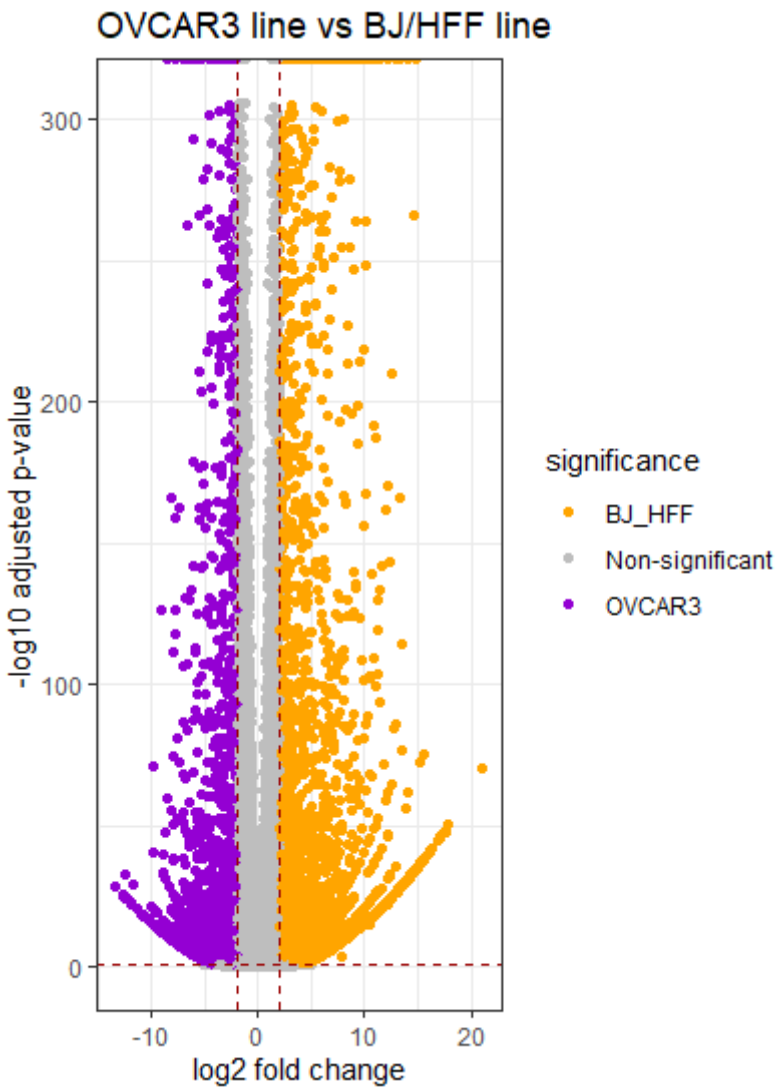

Figure 3-12. Volcano plot of the differential expression of the BJ/HFF line vs OVCAR3 line

This figure depicts the $-\log_{10}$ adjusted p-value score plotted against the $\log_2$ fold change in expression pattern between the BJ/HFF cell line(orange) and the OVCAR3 cell line(purple). The grey dots show the genes that were not statistically differentially expressed by a factor of 2 on the $\log_2$ fold change scale.

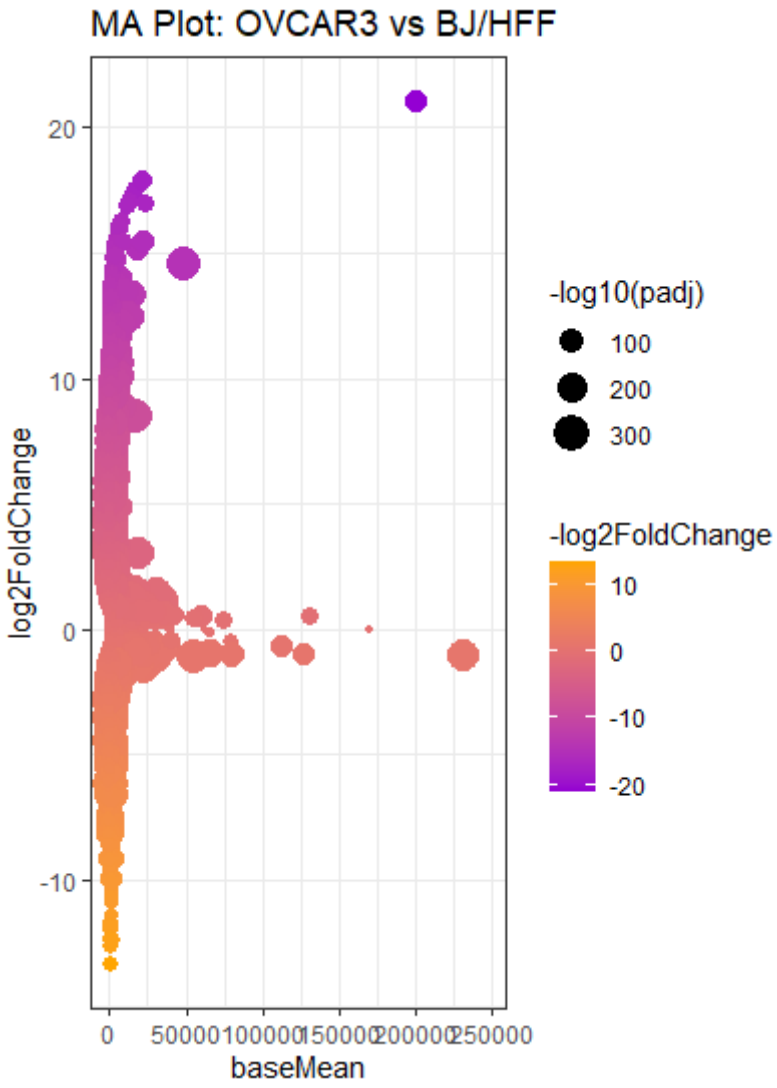

Figure 3-13. MA plot of the differential analysis of the BJ/HFF line vs the OVCAR3 line

This figure shows the gene expressions' $\log_2$ fold change plotted against the base mean counts of the data sets for the BJ/HFF against OVAR3 lines. The orange dots represent the genes that were up-regulated in the BJ/HFF line, and the purple dots are the genes that were up-regulated in the OVCAR3 line. The brighter the color is, the more differentially expressed the gene is. The size of the dots corresponds to the adjusted p-value of the differential expression analysis. Larger dots correspond to more confidence in the value plotted.

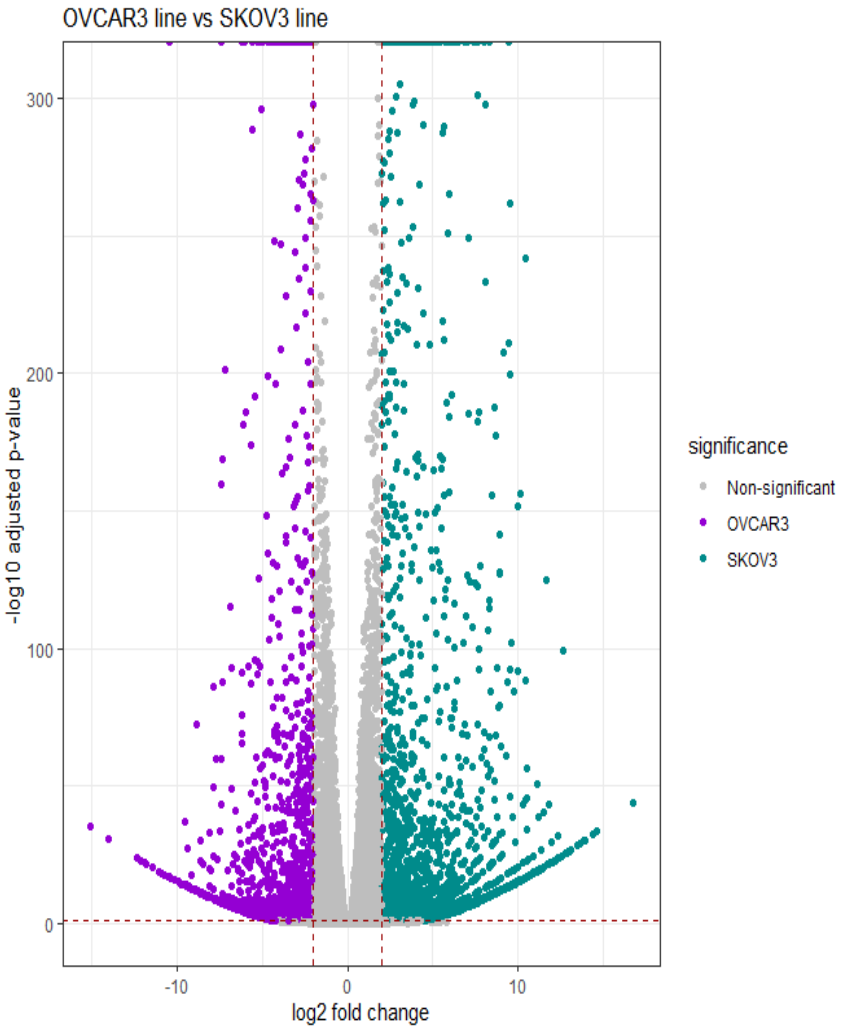

Figure 3-14 Volcano plot of the differential expression of the OVCAR3 line vs SKOV3 line

This figure depicts the $-\log_{10}$ adjusted p-value score plotted against the $\log_2$ fold change in expression pattern between the SKOV3 cell line(purple) and the OVCAR3 cell line(teal). The grey dots show the genes that were not statistically differentially expressed by a factor of 2 on the $\log_2$ fold change scale.

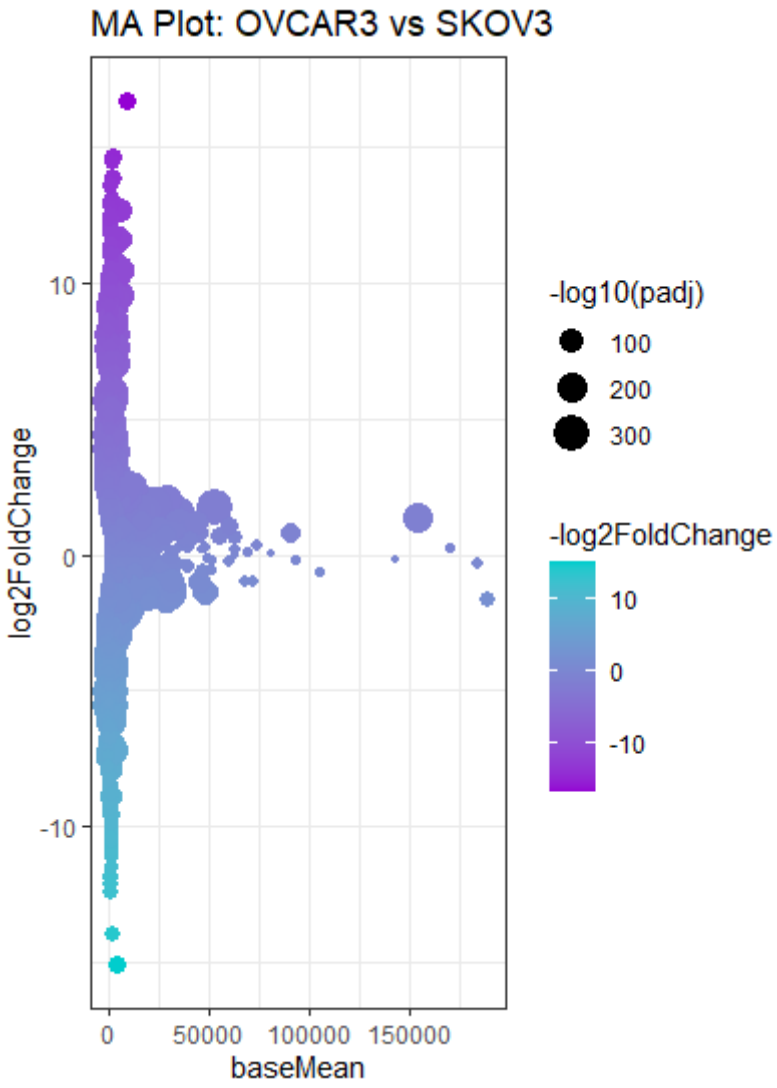

Figure 3-15 MA plot of the differential expression of the OVCAR3 and SKOV3 lines

This figure shows the gene expressions' log2 fold change plotted against the base mean counts of the data sets for the SKOV3 against OVAR3 lines. The teal dots represent the genes that were up-regulated in the SKOV3 line and the purple dots are the genes that were up-regulated in the OVCAR3 line. The brighter the color is, the more differentially expressed the gene is. The size of the dots corresponds to the adjusted p-value of the differential expression analysis. Larger dots correspond to more confidence in the value plotted.

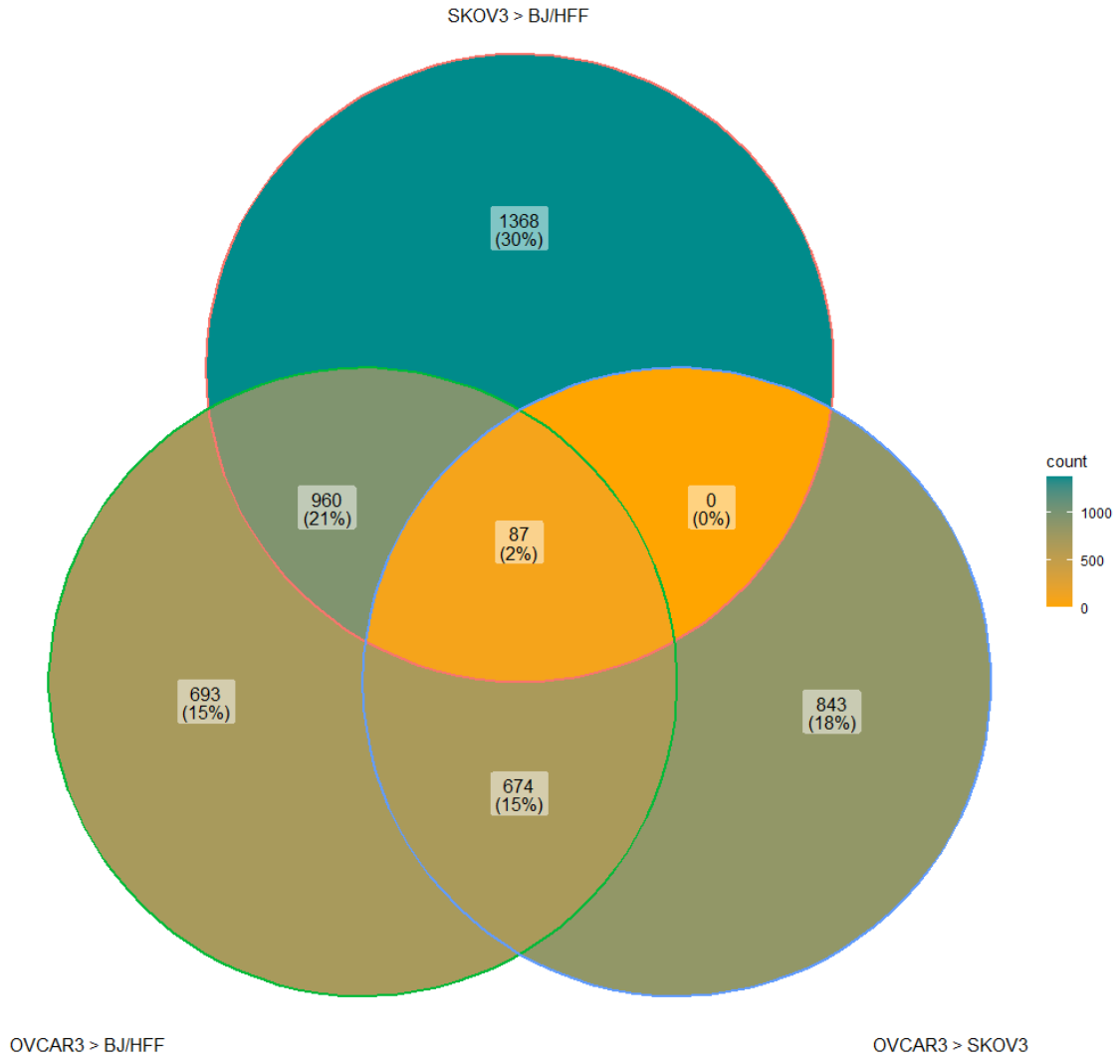

Figure 3-16 Venn diagram of the genes shared between the cell lines

This figure depicts the number of genes up-regulated in the three comparisons. The bottom left circle shows the number of genes up-regulated in OVCAR3 compared to BJ/HFF lines. The bottom right circle is the number of genes up-regulated in the OVCAR3 compared to SKOV3 lines. The top circle is the number of genes up-regulated in the SKOV3 compared to BJ/HFF lines. The middle intersection is the number of genes matching the target expression profile.

Table 3-1 Genes identified in the cell lines analysis

ENSG00000005187	ENSG00000047597	ENSG00000058866
ENSG00000076826	ENSG00000077274	ENSG00000092929
ENSG00000100473	ENSG00000100505	ENSG00000101276
ENSG00000104267	ENSG00000104413	ENSG00000104833
ENSG00000105357	ENSG00000105383	ENSG00000106852
ENSG00000111344	ENSG00000113763	ENSG00000117148
ENSG00000124920	ENSG00000126878	ENSG00000131187
ENSG00000134709	ENSG00000136237	ENSG00000136574
ENSG00000136944	ENSG00000136982	ENSG00000140470
ENSG00000141524	ENSG00000143851	ENSG00000147684
ENSG00000149328	ENSG00000151025	ENSG00000153012
ENSG00000153208	ENSG00000154080	ENSG00000160191
ENSG00000160867	ENSG00000164626	ENSG00000164867
ENSG00000165238	ENSG00000165810	ENSG00000167644
ENSG00000168505	ENSG00000168675	ENSG00000169242
ENSG00000169896	ENSG00000171126	ENSG00000171766
ENSG00000172137	ENSG00000173432	ENSG00000173557
ENSG00000178826	ENSG00000179674	ENSG00000180066
ENSG00000181418	ENSG00000184613	ENSG00000186481
ENSG00000188322	ENSG00000188488	ENSG00000189068
ENSG00000197444	ENSG00000197587	ENSG00000198435
ENSG00000203867	ENSG00000204282	ENSG00000214049
ENSG00000224420	ENSG00000225210	ENSG00000225383
ENSG00000225756	ENSG00000228742	ENSG00000229372
ENSG00000232133	ENSG00000236081	ENSG00000238178
ENSG00000239830	ENSG00000240086	ENSG00000243069
ENSG00000244306	ENSG00000246898	ENSG00000258733
ENSG00000262160	ENSG00000266401	ENSG00000270540
ENSG00000271361	ENSG00000277268	ENSG00000283172

3.3 Comparison of CT genes with cell line pattern genes

The 5,478 genes from the CT analysis were intersected with the 87 genes from the cell line analysis (Figure 3-16). There were 28 genes that were found to be present in both data sets (Table 3-2). These genes will be the target of a future analysis evaluating the mechanism of their misexpression, being both CT genes and matching the cell line expression pattern of being up-regulated in OVCAR3 and SKOV3.

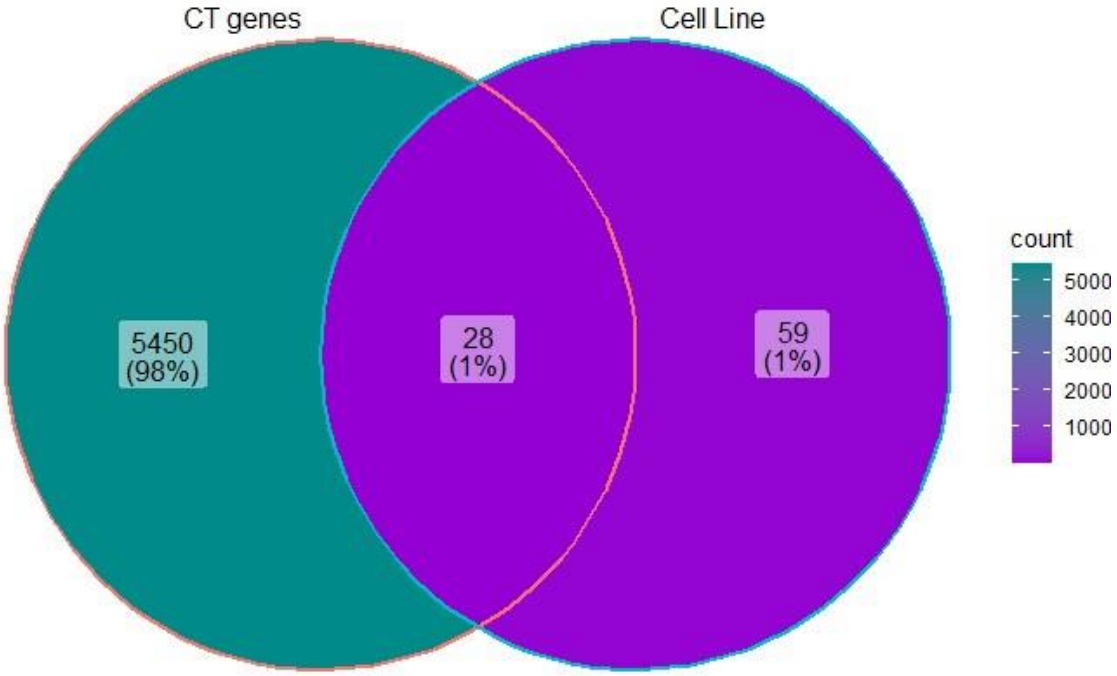

Figure 3-17 Venn diagram of CT genes and cell line pattern genes

The venn diagram depicts the number of genes shared between the CT gene analysis and the cell line analysis. The left circle shows the number of genes expressed explicitly in the CT analysis. The right circle indicates the number of genes exclusively expressed in the cell line analysis. The middle intersection indicates the number of genes that were shared between the CT and cell line analysis.

Table 3-2 Common genes between cell line and ovary CT gene analyses

ENSG00000076826	ENSG00000101276	ENSG00000104267	ENSG00000104833
ENSG00000105357	ENSG00000111344	ENSG00000113763	ENSG00000117148
ENSG00000131187	ENSG00000134709	ENSG00000136944	ENSG00000143851
ENSG00000151025	ENSG00000154080	ENSG00000165238	ENSG00000168505
ENSG00000173557	ENSG00000178826	ENSG00000180066	ENSG00000188322
ENSG00000189068	ENSG00000197587	ENSG00000198435	ENSG00000214049
ENSG00000228742	ENSG00000238178	ENSG00000240086	ENSG00000244306

4 Discussion

The existence of a gene list at the end of the analyses confirms that this method of detecting CT genes is viable. This is reinforced by six of the 28 genes having been confirmed to be involved in cancer development and progression: ENSG00000111344 (RASAL1), ENSG00000136944 (LMX1B), ENSG00000165238 (WNK2), ENSG00000168505 (GBX2), ENSG00000188322 (SBK1), and ENSG00000197587 (DMBX1) [74 – 79]. It is intriguing to note that RASAL1, a tumor suppressor gene, is highly expressed in cancerous cells, which would seem counterintuitive to a tumor cell. Genes that promote proliferation are expected to be found in cancerous cells.

Continuation of this research could have more samples of each of the cell lines incorporated, and more data for the ovary and testis from TCGA and GTEx. Utilizing more samples would increase the confidence in the genes discovered and provide potential targets to explore the process by which they are activated in ovarian cancer. It could be beneficial to run this analysis targeting specific ovarian cancer types to determine if there are different genes expressed in different ovarian cancers to determine how each ovarian cancer develops.

5 Conclusion

Cancer testis genes are useful in the development of cancer therapy vaccines. Understanding how cancer testis genes are activated in tumorigenesis is imperative to determine how cancer develops. There were 4,578 genes that were detected as ovarian CT genes. Comparing normal cell lines like the BJ/HFF lines to cancerous SKOV3 and OVACAR3, there were 87 genes being matching the expression profile of being highly expressed in SKOV3 compared to BJ/HFF, OVCAR3 compared to BJ/HFF, and OVCAR3 compared to SKOV3. There were 28 genes found to match the expression pattern of a CT gene and of the expression pattern set forth for our cell line data. Our analysis sheds new light of ovarian cancer testis genes that become active and creates new targets to pursue the understanding of the mechanisms of ovarian cancer development. Once these mechanisms are understood, another step will be taken in the treatment of ovarian cancer.

6 Reference List

1. McFarlane RJ, Wakeman JA. Meiosis-like functions in oncogenesis: A new view of cancer. Cancer Research. 2017;77(21):5712–6.

2. van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science. 1991;254(5038):1643–7.

3. Scanlan MJ, Simpson AJG, Old LJ. The cancer/testis genes: Review, standardization, and commentary. Cancer Immunology Research. 2004Jan1;4(1).

4. Hofmann O, Caballero OL, Stevenson BJ, Chen Y-T, Cohen T, Chua R, et al. Genome-wide analysis of cancer/testis gene expression. Proceedings of the National Academy of Sciences. 2008;105(51):20422–7.

5. da Silva VL, Fonseca AF, Fonseca M, da Silva TE, Coelho AC, Kroll JE, et al. Genome-wide identification of cancer/testis genes and their association with prognosis in a pan-cancer analysis. Oncotarget. 2017;8(54):92966–77.

6. Wang C, Gu Y, Zhang K, Xie K, Zhu M, Dai N, et al. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. Nature Communications. 2016Jan27;7(1).

7. Scanlan. MJ, Gure. AO, Jungbluth. AA, Old. LJ, Chen. Y-T. Cancer/testis antigens: An expanding family of targets for cancer immunotherapy. Immunological Reviews. 2002;188(1):22–32.

8. Gibbs ZA, Whitehurst AW. Emerging contributions of cancer/testis antigens to neoplastic behaviors. Trends in Cancer. 2018Sep20;4(10):701–12.

9. Janic A, Mendizabal L, Llamazares S, Rossell D, Gonzalez C. Ectopic expression of germline genes drives malignant brain tumor growth in drosophila. Science. 2010;330(6012):1824–7.

10. Chen C, Gao D, Huo J, Qu R, Guo Y, Hu X, et al. Multiomics analysis reveals CT83 is the most specific gene for triple negative breast cancer and its hypomethylation is oncogenic in breast cancer. Scientific Reports. 2021;11(1).

11. Zhang W, Barger CJ, Link PA, Mhawech-Fauceglia P, Miller A, Akers SN, et al. DNA hypomethylation-mediated activation ofCancer/testis antigen 45(CT45) genes is associated with disease progression and reduced survival in epithelial ovarian cancer. Epigenetics. 2015;10(8):736–48.

12. Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, et al. Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature. 2007;446(7137):815–9.

13. Whitehurst AW, Xie Y, Purinton SC, Cappell KM, Swanik JT, Larson B, et al. Tumor antigen acrosin binding protein normalizes mitotic spindle function to promote cancer cell proliferation. Cancer Research. 2010;70(19):7652–61.

14. Kumar V, Jagadish N, Suri A. Role of a-kinase anchor protein (AKAP4) in growth and survival of ovarian cancer cells. Oncotarget. 2017;8(32):53124–36.

15. Koo SJ, Fernández-Montalván AE, Badock V, Ott CJ, Holton SJ, von Ahsen O, et al. ATAD2 is an epigenetic reader of newly synthesized histone marks during DNA replication. Oncotarget. 2016;7(43):70323–35.

16. Ciró M, Prosperini E, Quarto M, Grazini U, Walfridsson J, McBlane F, et al. ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. Cancer Research. 2009;69(21):8491–8.

17. Cheeseman IM, Hori T, Fukagawa T, Desai A. KNL1 and the CENP-H/i/K complex coordinately direct kinetochore assembly in Vertebrates. Molecular Biology of the Cell. 2008;19(2):587–94.17. Zhao W-meng, Seki A, Fang G. CEP55, a microtubule-bundling protein, associates with Centralspindlin to control the midbody integrity and cell abscission during cytokinesis. Molecular Biology of the Cell. 2006;17(9):3881–96.

18. Chen C-H, Lu P-J, Chen Y-C, Fu S-L, Wu K-J, Tsou A-P, et al. FLJ10540-elicited cell transformation is through the activation of PI3-kinase/akt pathway. Oncogene. 2007;26(29):4272–83.

19. Maine EA, Westcott JM, Prechtl AM, Dang TT, Whitehurst AW, Pearson GW. The cancer-testis antigens spanx-A/c/D and CTAG2 promote breast cancer invasion. Oncotarget. 2016;7(12):14708–26.

20. Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, et al. Conditional expression of the CTCF-paralogous transcriptional factor Boris in normal cells results in demethylation and derepression of Mage-A1 and reactivation of other cancer-testis genes. Cancer Research. 2005;65(17):7751–62.

21. Singh S, Narayanan SP, Biswas K, Gupta A, Ahuja N, Yadav S, et al. Intragenic DNA methylation and Boris-mediated cancer-specific splicing contribute to the Warburg effect. Proceedings of the National Academy of Sciences. 2017;114(43):11440–5.

22. Fanjul-Fernández M, Quesada V, Cabanillas R, Cadiñanos J, Fontanil T, Obaya Á, et al. Cell–cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. Nature Communications. 2013;4(1).

23. Wegiel B, Bjartell A, Ekberg J, Gadaleanu V, Brunhoff C, Persson JL. A role for cyclin A1 in mediating the autocrine expression of vascular endothelial growth factor in prostate cancer. Oncogene. 2005;24(42):6385–93.

24. Miftakhova R, Hedblom A, Semenas J, Robinson B, Simoulis A, Malm J, et al. Cyclin A1 and P450 aromatase promote metastatic homing and growth of stem-like prostate cancer cells in the bone marrow. Cancer Research. 2016;76(8):2453–64.

25. Mathieu MG, Miles AK, Ahmad M, Buczek ME, Pockley AG, Rees RC, et al. The helicase hage prevents interferon-α-induced PML expression in ABCB5+ malignant melanoma-initiating cells by promoting the expression of SOCS1. Cell Death & Disease. 2014;5(2).

26. Tung P-Y, Varlakhanova NV, Knoepfler PS. Identification of DPPA4 and DPPA2 as a novel family of pluripotency-related oncogenes. Stem Cells. 2013;31(11):2330–42.

27. Cappell KM, Sinnott R, Taus P, Maxfield K, Scarbrough M, Whitehurst AW. Multiple cancer testis antigens function to support tumor cell mitotic fidelity. Molecular and Cellular Biology. 2012;32(20):4131–40.

28. Watkins J, Weekes D, Shah V, Gazinska P, Joshi S, Sidhu B, et al. Genomic complexity profiling reveals that hormad1 overexpression contributes to homologous recombination deficiency in triple-negative breast cancers. Cancer Discovery. 2015;5(5):488–505.

29. Suvasini R, Shruti B, Thota B, Shinde SV, Friedmann-Morvinski D, Nawaz Z, et al. Insulin growth factor-2 binding protein 3 (IGF2BP3) is a glioblastoma-specific marker that activates phosphatidylinositol 3-kinase/mitogen-activated protein kinase (PI3K/MAPK) pathways by modulating IGF-2. Journal of Biological Chemistry. 2011;286(29):25882–90.

30. Ennajdaoui H, Howard JM, Sterne-Weiler T, Jahanbani F, Coyne DJ, Uren PJ, et al. IGF2BP3 modulates the interaction of invasion-associated transcripts with RISC. Cell Reports. 2016;15(9):1876–83.

31. Viphakone N, Cumberbatch MG, Livingstone MJ, Heath PR, Dickman MJ, Catto JW, et al. LUZP4 defines a new mrna export pathway in cancer cells. Nucleic Acids Research. 2015;43(4):2353–66.

32. AlHossiny M, Luo L, Frazier WR, Steiner N, Gusev Y, Kallakury B, et al. LY6E/K signaling to TGFB promotes breast cancer progression, immune escape, and drug resistance. Cancer Research. 2016;76(11):3376–86.

33. Zhang X, Ning Y, Xiao Y, Duan H, Qu G, Liu X, et al. Mael contributes to gastric cancer progression by promoting ILKAP degradation. Oncotarget. 2017;8(69):113331–44.

34. Doyle JM, Gao J, Wang J, Yang M, Potts PR. Mage-ring protein complexes comprise a family of E3 ubiquitin ligases. Molecular Cell. 2010;39(6):963–74.

35. Kanehira M, Katagiri T, Shimo A, Takata R, Shuin T, Miki T, et al. Oncogenic role of MPHOSPH1, a cancer-testis antigen specific to human bladder cancer. Cancer Research. 2007;67(7):3276–85.

36. Liu D, Ding X, Du J, Cai X, Huang Y, Ward T, et al. Human NUF2 interacts with centromere-associated protein E and is essential for a stable spindle microtubule-kinetochore attachment. Journal of Biological Chemistry. 2007;282(29):21415–24.

37. DeLuca JG, Moree B, Hickey JM, Kilmartin JV, Salmon ED. HNUF2 inhibition blocks stable kinetochore–microtubule attachment and induces mitotic cell death in Hela cells. Journal of Cell Biology. 2002;159(4):549–55.

38. Hayama S, Daigo Y, Kato T, Ishikawa N, Yamabuki T, Miyamoto M, et al. Activation of CDCA1-KNTC2, members of centromere protein complex, involved in pulmonary carcinogenesis. Cancer Research. 2006;66(21):10339–48.

39. Michael AK, Harvey SL, Sammons PJ, Anderson AP, Kopalle HM, Banham AH, et al. Cancer/testis antigen PASD1 silences the circadian clock. Molecular Cell. 2015;58(5):743–54.

40. Oh S-M, Zhu F, Cho Y-Y, Lee KW, Kang BS, Kim H-G, et al. T-lymphokine–activated killer cell–originated protein kinase functions as a positive regulator of c-jun-NH2-kinase 1 signaling and H-Ras–induced cell transformation. Cancer Research. 2007;67(11):5186–94.

41. Lu Y, Zheng X, Hu W, Bian S, Zhang Z, Tao D, et al. Cancer/testis antigen PIWIL2 suppresses circadian rhythms by regulating the stability and activity of BMAL1 and clock. Oncotarget. 2017;8(33):54913–24.

42. Epping MT, Wang L, Edel MJ, Carlée L, Hernandez M, Bernards R. The human tumor antigen prame is a dominant repressor of retinoic acid receptor signaling. Cell. 2005;122(6):835–47.

43. Ramkumar P, Lee CM, Moradian A, Sweredoski MJ, Hess S, Sharrocks AD, et al. JNK-associated leucine zipper protein functions as a docking platform for Polo-like kinase 1 and regulation of the associating transcription factor forkhead box protein K1. Journal of Biological Chemistry. 2015;290(49):29617–28.

44. Banito A, Li X, Laporte AN, Roe J-S, Sanchez-Vega F, Huang C-H, et al. The SS18-SSX Oncoprotein hijacks KDM2B-PRC1.1 to drive synovial sarcoma. Cancer Cell. 2018;34(2):346–8.

45. McBride MJ, Pulice JL, Beird HC, Ingram DR, D'Avino AR, Shern JF, et al. The SS18-SSX fusion oncoprotein hijacks BAF complex targeting and function to drive synovial sarcoma. Cancer Cell. 2018;33(6).

46. Hosoya N, Okajima M, Kinomura A, Fujii Y, Hiyama T, Sun J, et al. Synaptonemal complex protein SYCP3 impairs mitotic recombination by interfering with BRCA2. EMBO reports. 2011;13(1):44–51.

47. Mondal G, Ohashi A, Yang L, Rowley M, Couch FJ. Tex14, a PLK1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the Spindle Assembly checkpoint. Molecular Cell. 2012;45(5):680–95.

48. Qiao H, Di Stefano L, Tian C, Li Y-Y, Yin Y-H, Qian X-P, et al. Human TFDP3, a novel DP protein, inhibits DNA binding and transactivation by E2F. Journal of Biological Chemistry. 2007;282(1):454–66.

49. Song Z-B, Wu P, Ni J-S, Liu T, Fan C, Bao Y-L, et al. Testes-specific protease 50 promotes cell proliferation via inhibiting activin signaling. Oncogene. 2017;36(43):5948–57.

50. Abrieu A, Magnaghi-Jaulin L, Kahana JA, Peter M, Castro A, Vigneron S, et al. MPS1 is a kinetochore-associated kinase essential for the vertebrate mitotic checkpoint. Cell. 2001;106(1):83–93.

51. Dou Z, Liu X, Wang W, Zhu T, Wang X, Xu L, et al. Dynamic localization of MPS1 kinase to kinetochores is essential for accurate spindle microtubule attachment. Proceedings of the National Academy of Sciences. 2015;112(33).

52. von Schubert C, Cubizolles F, Bracher JM, Sliedrecht T, Kops GJPL, Nigg EA. PLK1 and MPS1 cooperatively regulate the spindle assembly checkpoint in human cells. Cell Reports. 2015;12(1):66–78.

53. Simpson AJ, Caballero OL, Jungbluth A, Chen Y-T, Old LJ. Cancer/testis antigens, gametogenesis and cancer. Nature Reviews Cancer. 2005Jul20;5(8):615–25.

54. Caballero OL, Chen Y-T. Cancer/testis (CT) antigens: Potential targets for immunotherapy. Cancer Science. 2009Aug27;100(11):2014–21.

55. Gjerstorff MF, Andersen MH, Ditzel HJ. Oncogenic cancer/testis antigens: Prime candidates for immunotherapy. Oncotarget. 2015Jun30;6(18):15772–87.

56. Jakobsen MK, Gjerstorff MF. Car T-cell cancer therapy targeting surface cancer/testis antigens. Frontiers in Immunology. 2020;11.

57. Xie K, Fu C, Wang S, Xu H, Liu S, Shao Y, et al. Cancer-testis antigens in ovarian cancer: Implication for biomarkers and therapeutic targets. Journal of Ovarian Research. 2019;12(1).

58. Ovarian cancer statistics: How common is ovarian cancer [Internet]. American Cancer Society. American Cancer Society; 2022 [cited 2022Mar16]. Available from: https://www.cancer.org/cancer/ovarian-cancer/about/key-statistics.html

59. Hallas-Potts A, Dawson JC, Herrington CS. Ovarian cancer cell lines derived from non-serous carcinomas migrate and invade more aggressively than those derived from high-grade serous carcinomas. Scientific Reports. 2019;9(1).

60. Soldatova KI, Kit OI, Tolmakh RE, Vladimirova LY, Kutilin DS. Cancer-testis gene-expression features in various tumors. Journal of Clinical Oncology. 2019;37(15_suppl).

61. The cancer genome atlas program [Internet]. National Cancer Institute. NIH; 2020 [cited 2022Mar14]. Available from: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

62. Resource Overview [Internet]. GTEX portal. NIH Common Fund; 2021 [cited 2022Mar14]. Available from: https://gtexportal.org/home/

63. Shih I-M, Nakayama K, Wu G, Nakayama N, Zhang J, Wang T-L. Amplification of the CH19P13.2 NACC1 locus in ovarian high-grade serous carcinoma. Modern Pathology. 2011;24(5):638–45.

64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESEQ2 - Genome Biology [Internet]. BioMed Central. BioMed Central; 2014 [cited 2022Mar16]. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8

65. BJ [Internet]. ATCC. American Tissue Culture Collection; 2000 [cited 2022Mar16]. Available from: https://www.atcc.org/products/crl-2522

66. SK-OV-3 [Skov-3; skov3] | ATCC [Internet]. ATCC(a). American Type Culture Collection; 2021 [cited 2021Dec10]. Available from: https://www.atcc.org/products/htb-77

67. NIH:ovcar-3 [OVCAR3] [Internet]. ATCC. American Type Culture Collection; 221AD [cited 2021Dec10]. Available from: https://www.atcc.org/products/htb-161

68. Lu T, Bankhead A, Ljungman M, Neamati N. Multi-omics profiling reveals key signaling pathways in ovarian cancer controlled by STAT3. Theranostics. 2019;9(19):5478–96.

69. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018Sep1;34(17):i884–i890.

70. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Internet]. Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data. 2010 [cited 2022Mar22]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

71. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. Star: Ultrafast universal RNA-seq aligner. Bioinformatics. 2012Oct25;29(1):15–21.

72. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics. 2022Mar21;

73. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with deseq2. Genome Biology. 2014;15(12).

74. Liu D, Yang C, Bojdani E, Murugan AK, Xing M. Identification of RASAL1 as a major tumor suppressor gene in thyroid cancer. JNCI Journal of the National Cancer Institute. 2013Dec17;105(21):1617–27.

75. He L, Guo L, Vathipadiekal V, Sergent PA, Growdon WB, Engler DA, et al. Identification of LMX1B as a novel oncogene in human ovarian cancer. Oncogene. 2013Sep23;33(33):4226–35.

76. Zhou S-L, Zhou Z-J, Hu Z-Q, Song C-L, Luo Y-J, Luo C-B, et al. Genomic sequencing identifies WNK2 as a driver in hepatocellular carcinoma and a risk factor for early recurrence. Journal of Hepatology. 2019Jul23;71(6):1152–63.

77. Fang Y, Yuan Y, Zhang L-L, Lu J-W, Feng J-F, Hu S-N. Downregulated GBX2 gene suppresses proliferation, invasion and angiogenesis of breast cancer cells through inhibiting the WNT/β-catenin signaling pathway. Cancer Biomarkers. 2018;23(3):405–18.

78. Wang P, Guo J, Wang F, Shi T, Ma D. Human SBK1 is dysregulated in multiple cancers and promotes survival of ovary cancer SK-OV-3 cells. Cell Biology International. 2010Nov1;34(8).

79. Luo J, Liu K, Yao Y, Sun Q, Zheng X, Zhu B, et al. DMBX1 promotes tumor proliferation and regulates cell cycle progression via repressing otx2-mediated transcription of p21 in lung adenocarcinoma cell. Cancer Letters. 2019Mar27;453:45–56

A GTEx and TCGA data files

A.1 Ovarian GTEx samples

Table A.1 Sample IDs for ovary GTEx data

11DXX-1426-SM-5GIDU	11EM3-1726-SM-5N9D1	11EMC-1726-SM-5H11P
11GSP-0226-SM-5A5KV	11I78-1926-SM-59878	11P81-1526-SM-5P9GS
11VI4-1026-SM-5EQM1	11XUK-1626-SM-5GU5O	11ZTS-1926-SM-5CVLA
11ZTT-1826-SM-5CVLN	11ZVC-1426-SM-5EGGA	1269C-1826-SM-5N9E1
12WSD-2726-SM-59HLR	12WSJ-1226-SM-5LU91	12WSK-1926-SM-5LZVK
12ZZX-2026-SM-5LZV9	1313W-2726-SM-5P9IX	131XG-1926-SM-5LZVG
131YS-2226-SM-5P9G8	132AR-1426-SM-5IFF1	1399S-2126-SM-5J2MH
139D8-2426-SM-5KM3A	13D11-1126-SM-5KLYS	13FTX-0926-SM-5IFF7
13N11-0526-SM-5IJFI	13O3O-2726-SM-5KM35	13OVI-0726-SM-5L3DD
13OVJ-2426-SM-5KM3I	13PL7-2326-SM-5L3FY	13PVR-1226-SM-5RQJ2
13QBU-1526-SM-5LU3W	13QIC-1526-SM-5IJFL	13SLX-2426-SM-664OM
13VXT-1526-SM-5LU3J	13W3W-1526-SM-5LU5I	13X6H-1026-SM-5SIBE
145LS-2326-SM-5TDCY	145ME-1226-SM-5SIB6	145MI-2626-SM-5QGQ8
146FH-2526-SM-5Q5BP	14AS3-1326-SM-5RQJE	14BIN-1426-SM-664NI
14BMU-1726-SM-5TDE8	14DAQ-2226-SM-5YYAD	14JG6-0626-SM-68716
14PJM-2426-SM-686ZC	14PKU-1226-SM-686ZM	14PKV-1226-SM-6871T
14PN3-1526-SM-6ETZM	15DDE-2326-SM-6AJA8	15DYW-2626-SM-6LPK7
15EOM-1526-SM-68723	15ER7-2326-SM-7KUN3	15FZZ-0826-SM-6LLJO
15SDE-0926-SM-6LPIA	15UF6-2026-SM-6LPI3	16MT8-2226-SM-6LPK2
16NGA-0426-SM-7LG6K	17JCI-2326-SM-7IGPC	183WM-1926-SM-72D5F
18A7A-2626-SM-7LG6L	18D9A-0826-SM-718C1	1A32A-2726-SM-73KVU
1A8FM-2226-SM-7PC1C	1AMEY-0926-SM-72D6G	1AMFI-1426-SM-731EL
1B933-2126-SM-7MGXR	1C475-0426-SM-7P8OU	1CAMS-2126-SM-7MXTF
1EWIQ-2226-SM-7MGXJ	1F48J-2326-SM-7MKG3	1F52S-2226-SM-7MXVE
1F5PL-1626-SM-7MXTY	1F6RS-2726-SM-7MKHA	1F88E-2026-SM-7MXVP
1F88F-2126-SM-9KNUK	1GF9X-2426-SM-7SB7W	1GN1U-2326-SM-9MQK4
1GN1W-1426-SM-9JGHP	1GZHY-2426-SM-9JGGK	1H1CY-1926-SM-9WG7X
1H2FU-2326-SM-9WG7R	1HBPH-2726-SM-A96TW	1HCUA-1326-SM-9WPP6
1HSGN-2426-SM-C1YRI	1I4MK-1526-SM-CE6T1	1ICG6-1526-SM-C1YQT
1IDJE-1126-SM-A96SO	1J1OQ-1626-SM-CE6TB	1J8JJ-0826-SM-AHZ3F
1JN1M-2126-SM-D4P2X	1JN76-1426-SM-AHZ4F	1K9T9-2426-SM-DHXKT
1LGOU-1026-SM-DHXL4	1LGRB-2326-SM-CNNPU	1LKK1-1626-SM-EVR4Y
1LNCM-0926-SM-DHXK5	1LVAM-1826-SM-DHXKR	1LVAN-2726-SM-CNPQG

1MA7W-1426-SM-EV7A1	1MCC2-1026-SM-EVR5I	1OJC4-2026-SM-E6CQW
1PBJI-2026-SM-DPRYF	1PIEJ-1626-SM-E6CP9	1QEPI-2726-SM-DPRZ6
1QP28-1326-SM-DPRXZ	1QP67-1526-SM-E76P4	1QPFJ-1726-SM-EXOJU
1R9PN-0626-SM-E6CR9	1RDX4-0826-SM-E9TK7	1S5ZU-2326-SM-EAZ5B
OHPK-2426-SM-3MJGH	OHPL-2426-SM-48TDN	P4PP-2426-SM-3P61L
P4QT-2426-SM-3NMCL	P78B-1926-SM-3P616	PLZ4-2826-SM-3P617
POMQ-1426-SM-3P61D	PWCY-1326-SM-48TCU	PWN1-2426-SM-48TDD
PX3G-2426-SM-48TZZ	QCQG-1426-SM-48U22	QDT8-2226-SM-EZ6LT
QVJO-3126-SM-COH22	R55G-1526-SM-48FER	RTLS-2326-SM-46MUH
RU1J-0826-SM-46MUU	RU72-2226-SM-46MUE	RWS6-1026-SM-47JXD
S32W-1226-SM-4AD67	S341-0726-SM-4AD5R	S4P3-0926-SM-4AD54
S7SF-1326-SM-4AD4P	SN8G-2426-SM-EVR31	T2IS-2026-SM-4DM6W
T5JW-0426-SM-4DM7M	T6MO-1126-SM-4DM5D	TML8-0926-SM-4DXSJ
TMMY-1726-SM-4DXTD	TSE9-2226-SM-4DXUR	U3ZN-1026-SM-4DXTC
UJHI-1226-SM-4IHLR	W5WG-0926-SM-4RTX9	WEY5-0626-SM-4LMIA
WI4N-2526-SM-4OOSE	WWYW-2726-SM-4MVOP	WXYG-1426-SM-4ONCK
WYBS-2426-SM-4ONDI	WYVS-1526-SM-4OND4	X15G-1726-SM-4PQZN
X4EO-2426-SM-4QASD	X8HC-2726-SM-46MUA	XMD3-2326-SM-4AT5H
XUW1-0126-SM-4BOOQ	XUZC-1026-SM-4BOPY	XV7Q-1426-SM-4BRWA
XYKS-1726-SM-4E3IO	Y114-1726-SM-4TT7U	Y3IK-2026-SM-4YCDG
YFC4-2726-SM-5IFJW	YJ8O-1026-SM-5CVM5	ZAJG-2726-SM-5S2MU
ZC5H-2026-SM-5CVN2	ZLWG-0926-SM-4WWC3	ZV6S-1426-SM-59HKX
ZVT2-0326-SM-5E44G	ZVT3-2626-SM-5GU5L	ZYFG-1726-SM-5GZZB
ZYY3-2726-SM-5EGH4	ZZPU-2126-SM-5EGIU	

A.2 Ovarian TCGA samples

Table A.2 Sample IDs for ovary TCGA data

0045a228-1dad-4ced-9bc6-5067aeee0518	01770ca7-0c13-4078-b956-a35dc68fbe28
01a7573b-1ced-444d-9cb8-44212a30f878	01eac123-1e21-440d-9495-8b7a4166249f
0315439e-694c-4b14-a27c-29dfa117c764	03c8a50d-0484-42e6-ade8-d214684f62c9
04098317-63b1-43eb-9b7f-e8e619e52781	043993e1-370b-4bdc-8783-be65b97ed886
0495ff8c-8878-47b4-b10e-660d7252e5de	04adbacc-4d08-42ff-b261-5262344b08f3
053f5fb2-8b5f-4252-82fd-ba167a58c23f	054352e9-a6fb-4f61-b524-2741017c2bbf
06199259-c8da-4b44-b957-8148e31b38cf	062b6d0c-06b0-4d27-8702-32501278cd9c
065393c7-9928-4fdc-9dbf-68ba9c708561	06884e47-0650-4748-9f4e-b8cd9f24795c
06affed3-2a28-4fe6-8c16-b9432050bff6	076b33da-f743-41b3-816c-cce371a527f7
08dd298e-9a5b-495f-8f5f-42998ae6dabf	094dbc70-aa59-40b1-8ceb-1ce6e5e6bf01
0991973f-cd4d-4fcc-bf95-ee735422abb6	09ca93e6-6ff3-4e50-ba6a-34e5eb7c39b1
09f70243-303e-4312-9913-ecc76f2cee3e	0a398afb-4549-40f7-83e7-ba33739965ad
0ab0f22f-bfeb-4a5e-bd8f-f0fb30eac992	0d11c451-2dfa-40ed-93fc-261d94f299f3
0e52e958-9450-4c3e-b169-5ed4a63f044d	0f878485-4f73-45ec-8d2d-b89850813aa5
0fc26774-1494-4578-97df-7b4fee07483b	0ffb5d54-6522-4074-9c41-8371f284d1dc
111d4d64-e1cc-404d-8fa9-ca0a545041cb	11f80ba0-162f-417b-96b8-4d21ddaef88d
12c8b289-b9d0-4697-b3a6-db8eed617937	1475a657-0f90-4308-a59f-f6c51f60216c
15241763-4cf7-452e-bc24-b3d877348b2d	16417d82-1ae0-4097-85c1-5e2df313fd35
16b391ee-1f65-4665-bf34-5447eed6800b	16ce29ef-a16f-49df-9ee1-630f229208f7
16e1bba3-6db0-429c-bfc8-82f27c8a3252	178c423a-3ae9-4343-aa81-02292024e421
1865f7a2-31da-4e7a-8cd5-a4c15372ffaf	1902f5f0-8474-4f4c-a17b-f813fe02ca58
195a13be-31a4-47ce-bc3d-6aba8451e304	1ac930e3-e67d-410d-9182-44d85799fd18
1ade6fa3-de71-4760-9760-8abc10963255	1be906cc-a22b-4c94-975b-d0d68257f4da
1c314a2a-c268-4c89-a6ee-c679aff9052e	1d6a0834-0561-4d45-813d-1ea499b613c0
1db081ff-e475-412c-9f55-fc5face40558	1e0f7f1f-e4aa-43e6-a6c1-bc4a0439d3af
1e10d520-3bcf-4edd-93d0-03be32f7e8d7	1e918517-3a1f-45ae-953b-b65ecfaffd2d
1f2dbb8e-0713-4952-a9e0-70fefa263116	1fbaa35c-4ff8-49e4-8aed-4a20b408f840
20c77d79-6934-40ce-b490-54221196a090	211b9b7e-5a5b-4da7-b75c-4a0a738cdaed
2179de4b-5872-4129-82dd-ef19dc540898	21fccc2c-bc60-4ed4-ba78-5dfbb0d5f498
24a15ca8-0811-4131-baa1-94f904973ecd	24a828c5-596e-4a8a-ae44-9787841c6978
24b7176d-483b-416e-8ecd-d3beae934fc3	253d376e-81e8-446f-ace4-985184741bc3
25577d95-744c-45a8-b84a-a9c1b51c2baf	25d6dd26-51c7-40dd-962a-c00abb70c5cb
273c58d9-1f2a-491c-8487-408f751a7a3b	276e88b0-deb7-452f-9499-bd770c72f0cc
297978f4-9907-482f-89db-3a0a2a0f19a8	298026a5-291e-4376-bb5b-e1f063843235
29d09b58-3cc2-47c0-b78e-18743eed2c2f	2a07ad59-da28-44da-8da9-d948186c99ff

2a4a2670-4ad4-49ac-8571-477c44201219	2ab2d4f1-4171-4326-91e1-e2c795004468
2ad2adf3-212d-42e5-87df-a4025e900584	2ad549b2-56d8-4047-a398-cbe1f6295722
2df4d9b2-5c9f-4952-bd5b-62f1a244f5f1	2e01d725-ae04-49e8-9a45-52634cde5662
2f635d92-f453-41c4-b5e8-d7c3256877dd	30b1f3cd-604d-4a13-a196-1216658c87fa
30bc60d4-7bf1-474e-a2f9-3c1108820cd1	30cff944-ebc0-425e-8cd2-e509ff218471
3146c2ae-ff5e-4097-8005-f4d261af032e	31e45387-5c86-4d8e-9b54-da954ee4634f
320a6126-c3fc-483f-b79d-b15700a51997	32c77c11-20f5-4719-b668-b2da53c914c2
3344d45a-7ec5-4d77-8d77-b95aa55b7f4c	347a8eeb-204f-41df-98d3-61394d2b7cd7
34e382cd-6fef-4386-8a30-cd1c8f889f30	353ff7d9-113b-4802-8cc4-635cd44b8706
36a15497-1f35-446d-886f-ea34196a5fdd	374fe6df-3752-47d9-8060-f9e5d05750e5
3769209b-a12f-4a84-9d68-0e1c934e5aea	3788684a-b405-44fe-8d09-4e40a46d5214
38721491-84e5-4eee-92d4-fa8f8cf463f1	39ae7d32-a9c6-4599-a860-5c25e05899fb
39cf924a-ed8c-4a0e-a7a6-beea001be23b	3a34e792-461f-45f5-9c4e-a29deeef9ce2
3ad93b22-6d06-4c58-b365-2a0558bfb5f7	3c052e47-a4d4-4e13-a67b-b50fd1b7cbe5
3db0253d-7bc6-4d7c-8c8f-fbfefc3556a6	3e00e8aa-31bc-454e-b558-7bcbad5f47ad
3e3fc76f-8ff8-4679-9bb4-37d0509a92f3	3ef6bcf6-93a8-4c7c-b2c1-9c48fba76b32
3f04b67c-ac10-459d-a661-dc1ee3963311	3fe6e5df-486e-4887-b79a-89a1591f4a29
4013973f-bab5-4844-951c-c80f34fcb828	40b41560-b08b-47a7-8149-b73c9d00d25c
41060841-8b20-48b9-9bfc-fa669d36d635	41c1021b-1086-470b-9907-afe11444ed3f
42d8c04e-0355-455c-b982-581727f75665	43622957-7bed-4d24-a31e-5fadf41216e1
43b0eb35-8278-440a-a743-fd91ca544b1b	44c2a2eb-6671-4951-9088-cdf4c23d60d8
45197059-d0fc-49bb-819b-bd7621427236	451aa144-4d34-4858-b81e-3458db34892b
4524f1ca-c763-47af-8f80-331c64704d7d	45b64a61-30d8-4ecc-909a-803be9c3ff2f
45c99019-ad30-473d-af3b-2fa0ccb819b9	4610ddcb-2924-4d3f-9d2a-b6416f7c3f03
462e7d42-19b7-4e95-bf2d-d2a66852f6cd	4722ba18-27f6-480d-975b-59742f9f0f89
47ecef21-f50f-4da1-a267-4466eca80b28	48297930-d22a-463d-80ac-52214964f067
4a67d3fe-88a5-4255-89d2-7814321cf6ab	4b449f26-ee8f-47d1-9d93-de9ceb61ffba
4b4d6cbb-92b0-4c40-8338-5acdf5f3f2a9	4b63696f-c8f8-4601-8751-d8efda9224a3
4c6f3410-7bb6-4be3-a4d1-b3717279b212	4c8d97ab-39ad-41b9-b7f9-3aa56f373c2a
4d337bb6-8e53-4142-8c22-432bce3e754d	4e1c9b43-bfd3-4242-9d55-66ad4b66d84f
4e4b6800-a128-4daf-a9f9-803b24e04742	4e618fff-4f1a-403f-bfb7-3a613758383c
4e63ba56-ce7b-4f2a-9b87-aee011ad13e8	4f715b36-328a-4d62-8cf8-28cf65f73841
4fc22ec0-632d-4491-9d3e-99878010a689	5112a69c-c9ec-4c72-b93d-31f110d74eb1
511ebfdb-5384-4dac-9ea3-8916b541a7e9	517d7f3f-2ea0-4452-a743-8d0e3083a1bb
51b8dd7e-1686-46bb-9a29-b0a368eaadcc	52292cad-9758-4d1a-8162-1f883ff327c0
528fc2fd-fded-426d-a62c-af4ba50e9645	52a3d27b-f415-4e7c-8000-957574724cfd
5327762f-5add-4008-ac5f-beca30ee5b4d	538a1fe4-8e69-4a29-a789-da52c075e75f
53ff71a4-9487-4ae6-9d25-f55cfb70cf81	56427511-2167-465f-bdde-689b895b3d4b
568de3f5-ff1b-4dc2-9a8b-1a3034e58ee0	5784a937-f090-4866-8d8f-e7cbbb3e327d
5817ee66-fd3b-413a-a479-9a89a5a28a99	58d3df78-134d-43aa-a7cb-0e07db7e325e

58f06031-8e8c-4865-8cea-ba65d28317a9	590b183e-6bd5-4e23-b41c-7301575c4aa6
59633988-fe8b-4f38-a6ef-06910fef0846	59bebce4-6006-42a6-8d73-c6ef8fb3c9ce
5a8d910c-004f-4368-ad4d-0e8deec45384	5a928267-356d-47e9-b8b2-f477eaa261fa
5b23f813-9970-485b-8577-ab2f3c029c26	5cf0bf40-6b9a-4bcb-a88f-7256a298c5ef
5d0bb715-68ef-4688-9d58-0f18478e6a5c	5dbeb062-22ba-4057-afd6-8627f1ac9500
5ee01776-d2f8-433e-8dca-761151099491	5fa4dd02-be02-480e-a8ad-ae89bddc17b9
608f7a0f-75a7-425d-9a82-6bf7e2023138	60d09a9b-add7-498a-a4a3-3e926a23dc1d
60da3a1a-65ee-48fa-a08f-88b6d52a18c2	60deafcb-603b-4eb8-92b1-28656a674581
60fdab7c-849a-440f-815a-c67741e1106c	61b6a3de-8bb2-4fc1-85d9-849c823d8fd5
623e04e4-491d-4381-8c11-b64ece9a4c31	64430665-49d5-49d4-9ef8-7559cbfcabf8
654acb8d-c1bb-42b1-91af-4f0b37abf02c	655c5a8b-a87a-482e-95cf-1519b3146534
659a82b6-9aae-4aed-8df7-964697ad1640	65c32518-eda4-4a4b-8d39-00a28936664d
65ce7190-f1ef-496a-8e47-bcedd8dce223	65d87c44-cb1f-4889-bdfa-4788f7183ae1
660a745a-f0a2-4439-b3f3-51a60132b97d	662e7355-e7b4-4d12-8429-f53b7b70bbe6
678a3fad-2c4a-450f-ae1e-8258fc8a618a	67bee95c-2a85-443b-b662-28103dfefdd4
68339551-7f92-4000-b0ca-d33669bb2c39	68975d5d-407b-4ceb-96d3-8e2a37ca3c1e
68ebef00-5c43-475a-9117-33ea920cab69	69134f9d-2738-4fc7-820d-0fb80b699480
6aa7225b-3d10-4b19-a472-0adeb21c26cf	6b9a94e7-144b-4a17-97ed-75f7cd311baa
6c962389-a73a-4627-8a2c-41e4a1cc8df3	6e40a1e7-9ff6-455f-8f70-77611a66088e
6e6ab126-1ab3-428c-b780-a1a98c475154	701b8c71-6c05-4e5b-ac10-396d245d62ea
7024891a-6a74-45d1-813d-7199707c45c5	7093a0f5-822d-4f61-9adb-fe9adbbff769
715d1493-6c47-4f20-bd62-c08d4f3870de	7163f82e-33ef-4a92-946d-7dd07a9657af
72251745-4686-4ce0-b70e-5a0444b36b59	726b125e-47ab-43da-9737-f8cba68fbeae
72bb9902-d411-494f-98ca-0e273967b96a	72dad4d7-0138-4de3-924c-119658aa9084
7394d96b-2719-4eb4-b776-439171584f2a	73bc4ef3-2d63-4829-a28e-f400abf4290a
73ed989a-7f6a-4cff-a7f7-f99157535c7b	74318679-8eae-42ee-8721-fd38de88938a
75303476-cdec-4ae4-aaf5-01abdc3213ab	756efdf3-3d00-42dd-b854-0054444a7de5
7570289f-debf-40f9-9e2e-c5acf19626f4	75d8a8c7-bc64-48f4-8fe9-a0e7640e0b44
75fdcd7e-f701-4c93-b3cc-22ff09b770ad	760e59a8-b3bb-4d63-9e0e-23b5023546a4
765c9c00-fbd8-4dac-89da-a4cee971e8bd	7707599c-51ca-4801-adf9-1c9642254fdb
773a7fb2-61e4-4a6a-9eab-1cddcb8b83cd	78362a69-6a41-4c14-8e08-4fd7773ce03a
7849325e-0e51-4696-bd8e-14e0b0cf40e8	78c27c9c-843b-4e6a-a36a-1aa4dcb8decf
790570e9-099f-4458-88c9-1ebe12205a67	7a181264-18d3-4a81-b8c1-43e08ddab953
7a7e4b6b-de2d-4e49-bcbf-c705164477b9	7ac80350-d086-4aff-bbb9-a38a57d4cf28
7b632a5d-f69e-4fc4-a33c-73c3ff85a42f	7b6e2a12-c537-4f1f-911f-d430bd864d2d
7b707153-0a5f-4923-a3d3-a53a798d4259	7c4769c1-1907-47ea-98f7-26cd438d022e
7d1a066f-8791-45ff-98e5-f4eb3d1e1d8d	7d5d98d6-91b7-470f-879c-8e83ab4d3fc6
7d9443ff-9ff5-415e-b327-bc71d6cbc4b3	7dfc4c0f-bb68-4ac1-95fc-e2a3fee16704
7ee8ec85-e646-40c4-beef-5b8e8ab35168	7f0f3d94-1376-470a-a959-cd56039ae10a
7f3810b3-4ff8-4a86-944e-7c87cf5593a2	7fb4f0f7-027f-4cf8-b8bc-43ca398c46e8

8166ba38-a099-4a07-8271-0587fe8d10d7	81830d92-01c0-4607-888e-be947ccae6e2
825ebcdf-3e59-4d89-9f6d-b27c7dc896d1	829fbe64-5c02-4a8f-96e6-9e251c776107
84156df8-23f4-41e0-b305-e4812fded4da	84e6af70-0c66-410c-9fad-6188151ef356
85c382ef-b9c1-4c1a-8ef1-5a2fbbb5f72a	85ca6ac2-6cec-4249-9166-54a2f0e10404
85fe5749-1fef-49a2-bbac-9200c3c5a1cb	864f36c5-41fa-4f90-b737-e164ce1eb845
86cb28d4-5dcd-4e7e-a4c9-c5a5b16a9739	871da002-8649-42bc-8ce4-9d2a8ff66213
875c2a27-9732-4b8a-affc-8ea591595a43	87ccd166-7fbd-498c-b03f-d4c369ff30ef
87fd11ca-f797-4336-8e23-c8c0ccef6b0c	8822cd66-d592-4601-a81a-caf939cb8882
882d61db-f3ae-4c4c-8d2d-32c18b597624	88f9c446-a966-4131-b379-63e3cc1b7aa9
89758e34-511b-465a-81db-2af07f047d5f	8bb54b5a-d23c-46a2-839e-32355236477b
8beaeffb-5e9b-4589-8999-52c3d736aea7	8ca881ca-da4d-43f3-9d6d-1ec5289927b8
8cfe4a4f-7ca0-49fa-8820-99321c22176d	8e0d147d-16de-4d8e-b9f6-5e493917b7b5
8eb0f5ac-1196-4cc5-b25b-f9e76301772e	8ed16638-71b1-47bf-9fbf-1e9efee46e07
8fc333b1-b2ee-4eff-afa9-8606b282a341	904f8983-9011-4b3c-a5f6-94fca0a0573a
90e896dd-f619-4a9f-b386-d060c02bb876	92a8c68d-23ec-467f-84c0-4876b71359f8
94823ae2-8fb8-4be7-b2ca-88dff791b29b	95595a36-1582-4fb6-bf79-f0111b0abfc9
95ce6307-6006-4b0e-9fa7-942551eaf05a	968544ba-7990-41e8-a0ee-5206ceeb47ca
97d62176-1e34-4508-886b-df643f8a7e75	98e7898a-bd2e-4c49-9f76-aea65d7960a3
995b85f1-d46f-4989-a4a0-2a73b5df69bf	9a111391-cea9-47b0-bb5e-2bd19ff09a5a
9a420e2d-15c8-41f3-859f-25872765f75e	9ac17699-409b-4750-9317-aacc8c04da46
9bca614a-922a-4bef-8a96-c2b71dcaf17f	9c330249-5ea7-4795-98c1-9c8f2cea6ac3
9c6f2d76-7701-434e-a876-c04ab14cccea	9d2d7bcb-8af4-4143-bd1c-b7b5a239795f
9d80d1a6-3a9c-42ca-8139-c88659088638	9d904376-a13a-4827-a3f5-3772f579e5dd
9e198dfe-9fc6-48e2-ba06-90c49ddf48aa	9e51cba1-f13e-46a1-b37a-99e6db2d6134
9e623149-b279-41fd-9be0-5d20d8cc024b	9f223240-1020-4450-a6ec-a6168a9211e5
9f8c8f79-2581-48fb-90bb-5d3e898689d6	9facd5d5-33cc-485f-b80a-b613f936f2d5
a085cea5-b7ca-449c-b218-07c7715ca736	a090ec54-edf8-4188-9bfc-add752bdaefa
a1c4f19e-079e-47e7-8939-3122c56bbb98	a1ca5d5d-d468-476a-99be-0d6ea90f6533
a2082ad4-279e-422f-b5a7-cb7fbeb7a6df	a26fe084-f208-423b-ab61-6d34598c9cad
a46fc22b-4074-4184-b471-c372139f9486	a528840a-20b9-40ed-acfa-46b8b52ffef6
a5929721-d8ba-4b73-aa2f-c4903fb380bb	a5d9ce27-ad5c-4d0a-9494-b0d75c00f80b
a6542a39-6092-4a62-aa6d-51bfba48ef10	a65ae799-abca-4512-a65f-04f775818ab2
a6bbe2d8-2f21-4eaa-b2ed-4da1ff48054a	a6c5d247-364e-41c4-8a89-dda2410dd3af
a72d4647-2034-4bc0-9a58-db0f97b7e76a	a7ce619c-20e3-4a20-adee-1b1d0c367e76
aa77a3af-80ce-44d2-b227-faec0f734468	aac2baeb-b9e2-4e94-82c5-8d591d90cf91
ab8603dd-2f94-4c83-9927-455958be0007	ab89687d-2019-406c-95e8-a189c13c9fc0
ab9f8d1d-d981-4139-8191-ad8db91c187f	ac6dbd3b-d3fc-4126-b637-e6be2e7454db
ac6e38b7-143c-491e-9892-4c28f51ddce5	acaec383-b91d-46a9-a1d1-6eb4764cd000
ad9a29de-7c46-4caa-a802-b18053b4e32d	ade137ec-8b83-48d0-9162-4a2f52db5b0f
ae0eb407-7bf5-4023-a297-470c93e15d3d	aea568b9-f9d8-4458-89fd-894c9e9ecb57

af11caa0-a20b-4c17-8207-0a1142edd9a9	af6e5654-e755-4c15-b3e5-807da2642e25
b00351d7-1785-4157-9955-aadc2f195181	b065b395-20b6-4417-b4c8-d8161b5a9db2
b14d1cf5-7063-4d0c-aa27-bf2802102751	b1808878-10a4-4024-9359-70d7ef17a439
b1d5c444-da0d-4360-bd45-31c94217adfc	b244d2f7-122c-4529-8fcc-c19e2c719cf0
b2552f6f-dd15-410f-a621-195d90f44831	b269c35d-7f91-4c66-8bef-59906ec87745
b2838a0f-fdea-4d0e-b908-d09e866035f1	b28d6dd2-923a-4160-8876-7b0fbb7a7a26
b2ccd224-e50e-4dac-898f-83bda68866e8	b322d383-64cb-4033-9b44-872ad37acc4a
b3b1dbca-bce2-40a3-92a4-848f616b98e3	b4760c72-6405-4b0c-b13b-dbef127b7ffb
b4f37e7b-e458-4395-a204-8a1dad62a28b	b5381bca-ac72-41f2-840a-f7f23382fdd3
b64ef80a-d41c-4f92-a3ed-e43d55abb2c2	b90ceb09-1097-4c53-ae4d-0e7acd0ea8df
b9aa8d09-5536-4b37-8259-8018bfa6d886	ba8ce34a-28bc-485c-af2e-1c835c351997
bb2408ba-58c3-4ee3-8ddc-1af4e1394b94	bb94288a-d4c4-4811-bcd5-7306891779fd
bd0abda0-f330-40f3-b06b-1f6c8f5c667a	bd727430-2539-4897-a754-2c8d8de59f9e
bde7dae9-5985-4195-b22b-cc0babb99c75	be626971-cdea-495d-88cd-1323e7c23fe7
bea1858a-5de7-48ee-9fc3-f9abdd4b0ded	bea2f6be-177f-4454-8831-0d287cac265b
bec22c57-3929-4fd0-8cf8-314cf4666719	bf4058ee-ed0b-43e4-bc4e-3343239f187b
c255d022-a659-42bd-9099-6853b41b64c7	c29d7649-364a-4fe5-9dd2-9f8937a565e4
c2de32e3-b9bc-49e1-9dbe-8c09b48ff637	c5e8e5e5-91e0-4e3c-97ad-e1cb63aaf70a
c6287ac6-570d-4431-87b3-290db2dbe58b	c62a3590-7d1d-4a46-b021-279e7a36b77e
c65d79cb-b639-4030-97d9-ddb84bc39195	c773cdc3-7702-4833-9b12-ba954ac4b357
c9689d9f-6138-42a8-a58e-1b44dc4b193f	c96b872f-d842-4e6a-8845-59308a28c279
c9e12ef8-5bf6-496d-8c8a-e0596c2f1d7c	caa50267-cc68-4ccc-b506-49a89b163510
cbe087ce-2924-4c5a-b307-51c80b41d6d8	cc7bfa3f-7301-44d4-b29a-35905c59eae6
ce511378-d8f8-494e-a07d-2e0dbed68bf8	ce515bac-2166-4e01-a3f7-fa822fe8b7d5
ce9deae0-fdae-4c71-baee-d9054fdd0f77	cebba8db-616b-49a5-ac4c-c3bee701e48a
cf3d8158-3e8d-47ba-8569-e976c8e8e1a6	cf7f32fa-03aa-441e-92fb-e2fd5df5bbc0
cf926727-8d8a-4dea-87cb-18aeb4ab216d	d04892ff-0ddc-45d1-93a7-abe21c67eb94
d0744791-f4c0-45c7-b233-c6512f3c766b	d0e4aa5d-e918-4156-99ba-14b14a2a9975
d1245eee-b6b7-42aa-856c-79148869cb2b	d12e86a0-97a4-4260-9c8a-0122931addf3
d143ba0b-3878-428d-9432-2a179a9abf8c	d1488c77-de3d-4bfe-aee9-7c398f62d8ec
d15fbebd-39c7-4714-b6f3-fe920cdd8bef	d26ea6e6-8770-486f-9d79-249e5b039606
d2e8bf06-abd8-4703-a4a0-29cf6fe01d18	d3d0227f-bf96-43be-aee0-d11518748e11
d4653d1b-2cfd-4d30-938d-7b12b66b8aaf	d490312c-c941-4336-8b94-163881b72be6
d5159274-e487-430f-8204-02653f41690c	d522e111-8a5f-48de-b894-1bf7505bda98
d687d704-1868-4120-9445-186e15cf6216	d7191f19-150b-4175-83a7-21052fefc488
d73c0f69-ab9b-4408-bf89-aa34bf829351	d7be2882-fa19-4a0f-937e-286066dd0642
d7e6e15d-d26c-43f3-b0a9-d5a4dc861e9d	d85fe825-5437-4717-8848-45e1290dd928
d94e4aa8-366d-4330-a08c-b543fc0bd18e	da436927-a3f5-4f12-b63d-d6f13f436ab7
dabdc0ac-e436-4140-ba66-52a7d7c50ee0	db2f54d8-82cb-48b2-a3d4-1fb7daa8e75c
de855db4-12d8-47bc-855b-96f61d306541	df06bca2-4481-455f-98b2-4c1981698f56

e04f45cb-e226-47b6-8e15-5bfcdec1a078	e0b32572-2ff9-4929-9917-83624e63aa2a
e28ba1c3-344d-45d3-904b-cace53ca78d0	e2cf2389-07bc-49d0-9426-82a98728a685
e360132f-9b30-4d44-ba1c-68941706d9ee	e377a906-e48f-4f6c-998e-7f914b8cd712
e47d5da5-4752-46d0-846f-f626be1793ed	e4a4badf-b4f9-4729-bec7-191867c7c229
e5c1e1c2-d80c-47a1-9f18-8125b186b1cd	e6993299-9429-4689-acb9-1fff67ce6642
e6cbc9a5-b8a2-485f-abc5-e66767520680	e78cab9d-1b43-4416-8dbc-78d51558d895
e78e6e23-183e-4492-8b69-28cf7c2623f4	e8b90f89-c047-404e-9e04-5589ff43495a
e8bd5099-de91-4c44-8573-f5eaba611dfe	e956031b-931a-4ccc-b0d5-51cb1ab9baed
e9841cd8-c4a9-49ac-9d53-6b808d82ca8a	ea70df70-0135-4ac7-8f65-f2b7839ebc23
eb329194-22d5-4675-a982-0c2dd8e5a324	eb4572b5-43a0-4d92-ab9c-1c5853710297
ec380f9e-879a-44cb-b16c-2a47417a09d0	ec69e15f-b3cf-4f42-94ed-72d80565a1b6
edbf38b1-342c-44a6-8d6f-b5bc871848d8	ee8019df-d2d2-4a32-9ef4-6da186fa2866
eeb18b62-a0f1-4af8-8115-cdaeaeef0f32	eeddee84-15ec-401f-95ee-de2bb7015168
effcab37-0ee4-422d-9d51-5cdbb78d2db9	f00793a7-8ba8-4486-95d1-3ea0beae2508
f0586e76-31db-472c-9ccb-4c63fdf5cd15	f10f8b8e-c6e7-46e1-a97d-08f90c76b539
f12b9d7b-1b32-43e9-a496-0794db75f69e	f1b52b3a-9748-4862-ba4a-a8f8e8a0166c
f27920c4-1504-45ef-8933-65f42ec9ab98	f3c8c91a-71e0-4b1d-a03f-06ed0123ce9a
f562a9d0-39fb-4062-b535-87a96ad3212a	f6a67523-d202-4ce7-af2e-b2cc4d8db20a
f72d0be0-de49-402e-8c92-4968ae079973	f76c1066-d6e9-4d27-a0ac-0a493d8e85f4
f7722eae-ade1-4087-a6d7-aff561972b73	f7d12919-2133-403d-9492-155e4666fb94
f8397fad-30b3-41a1-8d44-52d939e77a06	f84a525a-7fab-4c5c-b37a-6027787ee817
f89131d8-8ee1-4733-b97a-0447f7ce601b	f8cef5ad-7841-4f50-ba42-55e11ed1fc36
f920e615-9b0a-4767-821b-201c407730c4	f9bf79a2-9d3a-4df1-a03d-5a9dad5b5fbf
f9dc7237-e186-40e8-ab4c-d142dba7cf3d	fa2df153-c852-4df1-8499-fe2517dee301
fa3553e7-0255-4d60-b8c9-3caba0d945cd	faf78582-3ccd-4ee7-8193-92ed0945fd6c
fb4b490a-de40-4f75-919d-4fb180647eca	fb9c5a20-75cd-49f4-b75f-09a054e72267
fd15971c-b78e-4bba-bf87-59cd758b7cc6	fd638406-d933-47da-ba38-8ffc5046d49e
fda19653-d2d3-41f2-9122-72e892fd2853	fea3c4d0-7b9f-4279-989e-535aaefbdfeb
fed23e32-490a-4849-b335-d6cc8f0187fa	fedd52be-18a8-423f-ba8a-4f9416f11ff5
ff2264ee-19ac-4f74-b3fe-eaca826ea493	ff54b7b3-1622-43f7-a9b3-96a99384999b
ffa465fc-7af9-401f-af3c-bca394a1ab25	

A.3 Testis GTEx samples

Table A.3 Sample IDs of testis GTEx data

111CU-1726-SM-5EGHM	111FC-1926-SM-5GZYC	111VG-1926-SM-5GIDO
111YS-2026-SM-5EGGL	117XS-2026-SM-5GID1	117YW-1526-SM-5EGGP
117YX-2026-SM-5GIEF	11DXY-0226-SM-5H123	11DXZ-2126-SM-59881
11EI6-2226-SM-5EGJM	11EQ8-1426-SM-5EGJR	11EQ9-1926-SM-5PNVV
11GS4-2026-SM-5N9CP	11LCK-2326-SM-5HL53	11NSD-1026-SM-5N9BE
11NUK-2626-SM-5A5MB	11NV4-1726-SM-5N9FC	11O72-0726-SM-5P9GO
11OF3-1826-SM-5987N	11ONC-2226-SM-5HL6D	11P82-1526-SM-5BC5M
11TT1-2226-SM-5GU6B	11TUW-2226-SM-5EQL9	11WQC-2326-SM-5EQKE
11WQK-2826-SM-5EQKH	11ZUS-2726-SM-5FQUA	1212Z-0326-SM-5FQSJ
12696-0226-SM-5EGL3	12BJ1-1326-SM-5BC5P	12C56-1426-SM-5FQSW
12WSH-0326-SM-5GCNH	12WSI-2126-SM-5GCMV	12WSL-2326-SM-5DUXQ
12WSM-1326-SM-5GCP9	12ZZY-0126-SM-5LZV2	13111-1526-SM-5EGJX
13112-0226-SM-5P9IV	131XE-0426-SM-5IJF4	132QS-1226-SM-5P9GD
1339X-1926-SM-5PNVP	1399Q-2826-SM-5IJEZ	1399R-1626-SM-5P9GG
1399T-1526-SM-5P9J6	139T6-1226-SM-5IFFC	139TS-1726-SM-5IJG5
139TT-2226-SM-5LZWO	13FHP-2826-SM-5IJFW	13FLW-2126-SM-5N9FD
13FTW-1326-SM-5LZZD	13N1W-2626-SM-5IJEP	13N2G-0126-SM-5N9DV
13NYB-2226-SM-5MR58	13NZA-2526-SM-5IJFX	13NZB-2026-SM-5MR4M
13O1R-0726-SM-5IJEI	13O21-1226-SM-5J2MK	13O61-2026-SM-5J2M6
13OVH-0726-SM-5N9BU	13OVK-2226-SM-6LPJY	13OVL-0426-SM-5IFG6
13OW5-2526-SM-5L3I1	13OW6-0126-SM-5IJGM	13OW8-0526-SM-5KM24
13QJ3-0226-SM-5S2PU	13VXU-0726-SM-5J2O7	144GL-0726-SM-5LU4P
144GM-0426-SM-5Q5C8	144GN-1626-SM-5Q5BU	145LT-0426-SM-5LUAP
145MF-1726-SM-5LU9H	145MH-2326-SM-5O9AW	145MO-0126-SM-5S2QU
14753-0626-SM-5Q5CY	147F4-0626-SM-5LUAK	147GR-0626-SM-5S2PK
147JS-0126-SM-5S2TW	14A5H-0626-SM-5TDCO	14A6H-2326-SM-5Q5B5
14ABY-0626-SM-5Q5C9	14B4R-1026-SM-5TDDS	14BIL-2226-SM-73KWF
14C38-0126-SM-5YY9V	14C39-0526-SM-664OF	14C5O-2326-SM-73KYU
14DAR-1126-SM-793AT	14E1K-1926-SM-73KWS	14E6E-1026-SM-664N9
14E7W-0726-SM-664OK	14PHX-1426-SM-69LPP	14PJ2-2026-SM-6AJAQ
14PJ3-1626-SM-664O5	14PJ4-1726-SM-664OC	14PJN-1626-SM-68727
15G19-1326-SM-6LPIR	15RIE-0626-SM-6M47G	15RIF-1826-SM-6M469
15RJ7-1826-SM-7KUMJ	15SKB-0326-SM-6M477	16AAH-1626-SM-7EWE2
16MTA-1126-SM-6LPJA	16XZZ-0626-SM-6M47T	16YQH-1826-SM-6LPJW
16Z82-2726-SM-7KULV	17EVP-2426-SM-7IGNK	17F9E-1626-SM-7EWDB

17GQL-1726-SM-718B8	17HGU-0726-SM-7DUFN	17HHE-1226-SM-793C6
17HII-0126-SM-7KFSS	17KNJ-1126-SM-7KFT6	17MF6-2726-SM-7IGMZ
17MFQ-0826-SM-793C8	183FY-1926-SM-7KFRI	18465-1526-SM-7KFTV
18A66-0726-SM-72D77	18A67-1826-SM-7KFT7	18A6Q-2626-SM-7KFT3
18A7B-1526-SM-7KFTH	18D9B-2726-SM-72D73	18QFQ-1926-SM-72D5H
1A3MV-0726-SM-72D5M	1A3MX-1026-SM-731F8	1AX8Z-2826-SM-73KTZ
1AX9I-1826-SM-72D5I	1AX9J-1226-SM-72D6P	1AX9K-1826-SM-731CY
1AYCT-1026-SM-79ONR	1B8KE-1426-SM-7EWEL	1B8KZ-1826-SM-73KVJ
1B8L1-2326-SM-9JGGA	1B8SF-2926-SM-731DL	1B996-2326-SM-731EC
1BAJH-0726-SM-7IGMF	1C4CL-1926-SM-731DY	1C64N-2826-SM-7IGP5
1C64O-2126-SM-7IGPB	1C6VR-1526-SM-7MKG9	1C6VS-1126-SM-7EWEW
1CAMR-1526-SM-79OLP	1CB4F-2126-SM-7MXUL	1CB4G-2426-SM-7DUGM
1CB4I-0926-SM-7MKFY	1E1VI-1026-SM-7MKGQ	1E2YA-2826-SM-7EPIN
1EH9U-2626-SM-7IGQN	1EKGG-2726-SM-79OOD	1EMGI-0626-SM-7MXUS
1EN7A-2226-SM-7MXU6	1EU9M-1626-SM-7EWF2	1F6I4-1426-SM-7MXU3
1F6IF-0526-SM-7MKHD	1F75A-2826-SM-7MXTS	1GF9V-0626-SM-7MXU1
1GF9W-2126-SM-7MXUX	1GL5R-0826-SM-9KNTS	1GMR2-0726-SM-7MXVK
1GMR3-1726-SM-9JGGQ	1GMRU-0626-SM-7MKH2	1GN1V-1826-SM-9JGGD
1GN2E-0126-SM-9OSW8	1GN73-1726-SM-9JGFR	1GPI7-1426-SM-9JGGR
1GTWX-2226-SM-9JGHD	1GZ4I-2726-SM-9JGG9	1H11D-1626-SM-9JGHM
1H1DG-2726-SM-9JGI1	1H1E6-0626-SM-9OSWF	1H1ZS-2826-SM-9OSXS
1H4P4-1926-SM-9WPOO	1HB9E-1926-SM-D4P34	1HBPN-2226-SM-9WPNR
1HCVE-2826-SM-9WPPI	1HKZK-1926-SM-9WPPK	1HSEH-0726-SM-B2LXD
1HSKV-1826-SM-CNPPM	1HSMO-2426-SM-A9G24	1HSMQ-1726-SM-B2LXZ
1HUB1-2026-SM-A96S8	1I1GP-2426-SM-B2LXK	1I1GS-1826-SM-C1YS8
1I1GU-0726-SM-ARU7H	1I1GV-2526-SM-B2LXL	1I6K7-2426-SM-AHZ2I
1ICLY-2926-SM-C1YS5	1IDJF-2226-SM-AHZ2T	1IDJH-2326-SM-D4P2K
1IKJJ-2126-SM-C1YQU	1IL2U-0126-SM-CNPQ5	1IOXB-1526-SM-A96T6
1IY9M-2526-SM-C1YQF	1J1R8-1326-SM-CE6TC	1J8Q3-0726-SM-AHZ3W
1JJ6O-0526-SM-AHZ3G	1JJEA-1926-SM-AHZ41	1JK1U-0726-SM-C1YPZ
1JKYN-1026-SM-CGQG4	1JMLX-0726-SM-AHZ3M	1JMQJ-0726-SM-CXZJJ
1JMQK-2726-SM-CNNPI	1JN6P-2826-SM-CXZJK	1K2DA-1326-SM-CGQGP
1K2DU-2226-SM-CXZJY	1KANA-1626-SM-CXZL2	1KANB-2126-SM-DHXJV
1KD4Q-2126-SM-EV7A6	1KD5A-1826-SM-DHXJI	1KWVE-1126-SM-DHXJY
1KXAM-1826-SM-D3LAF	1L5NE-1626-SM-DHXKW	1LB8K-2326-SM-DHXJL
1LBAC-1826-SM-D3L9X	1LG7Y-2326-SM-EVR56	1LSNL-2426-SM-DHXKO
1LSVX-2126-SM-EV7A9	1LVA9-2826-SM-EAZ5D	1M5QR-2626-SM-EV7AV
1MA7X-1526-SM-DHXJF	1MCQQ-2326-SM-DLHBX	1MJK2-1326-SM-EV7A4
1N2EE-0826-SM-E6CRI	1NHNU-1526-SM-E6CP2	1O97I-2426-SM-DPRZ2
1OKEX-1926-SM-E6CP7	1PDJ9-1426-SM-DPRYW	1PIGE-1626-SM-DPRZE

1PIIG-0526-SM-E6CQM	1PPGY-0726-SM-EXOJ7	1QAET-1926-SM-DPRZG
1QMI2-2826-SM-DPRXR	1QP29-0426-SM-EVR3Y	1QP2A-1626-SM-E6CPO
1QP6S-0826-SM-E6CPP	1QP9N-1326-SM-DPRZN	1R7EU-1726-SM-DTX9M
1R9K5-2326-SM-DPRZC	1RAZA-2126-SM-EVR4A	1RAZQ-1126-SM-EAZ4S
1RAZR-2226-SM-E6CRG	1RB15-1926-SM-EV79J	1RQED-1626-SM-EAZ55
1S5VW-2626-SM-EVR4W	1S831-2026-SM-EXOJP	N7MS-0126-SM-3TW8O
NFK9-0126-SM-3LK5H	O5YT-2126-SM-3MJGD	OHPM-2126-SM-3LK75
OIZF-2126-SM-7P8R2	OIZG-0126-SM-E9TI2	OIZH-2126-SM-3NB1P
OIZI-0126-SM-3NB13	OOBJ-2126-SM-3NB1N	OOBK-2126-SM-3LK5T
OXRL-2126-SM-3NM98	P4QS-2126-SM-3NMCF	PLZ6-1226-SM-3P5ZS
PVOW-2126-SM-EZ6LN	PW2O-1426-SM-48TCD	Q2AH-1526-SM-48TZG
Q2AI-1226-SM-48U14	QEG4-0126-SM-48TZE	QEG5-0126-SM-CMKFD
QLQW-1026-SM-447A9	QMRM-1526-SM-4R1K6	QV31-1126-SM-CKZNA
QV44-1726-SM-4R1KG	R55C-1426-SM-48FED	R55D-0126-SM-48FEL
R55E-0726-SM-48FCZ	REY6-0126-SM-48FDT	RM2N-1326-SM-48FCW
RN64-2326-SM-48FDW	RUSQ-2126-SM-47JXK	RVPV-1126-SM-EAZAV
RWSA-2426-SM-47JXR	S33H-0126-SM-4AD62	S3XE-1526-SM-4AD5A
S4Q7-1226-SM-4AD5I	S4Z8-2126-SM-4AD5H	S7PM-0626-SM-4AD4Q
S7SE-0326-SM-4AT5Q	S95S-1126-SM-4B64E	SNMC-1026-SM-4DM7K
SNOS-1126-SM-4DM67	SUCS-1326-SM-4DM5T	T5JC-0726-SM-4DM55
T6MN-2026-SM-4DM7L	TKQ1-0926-SM-4DXU2	TKQ2-1526-SM-4DXUN
U3ZH-1526-SM-4DXV1	U3ZM-1626-SM-4DXSK	U4B1-1526-SM-4DXSL
U8T8-1126-SM-4DXUE	U8XE-0126-SM-4E3I3	UPJH-0126-SM-4IHLL
V1D1-2126-SM-4JBH4	VJYA-1426-SM-4KL1Y	WFG8-1926-SM-4LVM1
WFON-2026-SM-4LVMW	WH7G-1926-SM-4LVMM	WHSB-2126-SM-4M1XF
WHSE-0426-SM-4M1XO	WK11-0326-SM-4OOS6	WOFM-1126-SM-4OOSB
WQUQ-0326-SM-EWRM7	WVLH-2626-SM-4MVNV	WY7C-2226-SM-4ONCS
WYJK-1826-SM-4ONDM	WZTO-0326-SM-4PQYZ	X261-2326-SM-4PQYU
X3Y1-2626-SM-4PQZI	X5EB-2026-SM-4E3KA	XAJ8-1326-SM-47JYT
XBEC-0126-SM-4GIDT	XBED-2026-SM-4AT5D	XGQ4-2026-SM-4AT6G
XK95-2726-SM-4V6G6	XLM4-1526-SM-4AT6D	XMK1-2026-SM-4B65K
XPT6-1626-SM-4B655	XPVG-2226-SM-4B65U	XQ3S-2726-SM-4BOP2
Y111-2426-SM-4TT23	Y3I4-2026-SM-4TT6Z	Y5V6-1726-SM-4VDSZ
Y8E4-2226-SM-5LU94	Y9LG-1726-SM-4VBQE	YB5E-1926-SM-5IFIG
YEC3-1726-SM-5IFIK	YEC4-1526-SM-4W1YU	YF7O-2026-SM-4W1YE
YFCO-1726-SM-4W21S	YJ89-0626-SM-4TT3Z	Z93S-1726-SM-5HL8G
ZA64-1626-SM-5CVME	ZAB4-0126-SM-5CVMG	ZAB5-2426-SM-5CVMW
ZDYS-1326-SM-5IJFF	ZLFU-2026-SM-4WWG2	ZPU1-2126-SM-57WED
ZQUD-2026-SM-51MSM	ZT9W-2226-SM-57WFU	ZT9X-1426-SM-5DUX1
ZTSS-1526-SM-51MTC	ZTX8-1126-SM-51MRM	ZUA1-2726-SM-59HLJ

ZV7C-2026-SM-5NQ8F	ZVTK-0126-SM-57WDG	ZVZP-2226-SM-57WBF
ZYFC-0126-SM-5GIEH	ZYT6-2726-SM-5GICP	ZZ64-1126-SM-5GZXY

B Executed Code

B.1 Ubuntu command line for cell line sample preparation

```
#BJ_ctrl_SRR7613003_rep1, BJ_ctrl_SRR7613004_rep1, BJ_ctrl_SRR7613005_rep1
fastq-dump --split-files --origfmt SRR7613003 SRR7613004 SRR7613005

#Trim reads
fastp -i SRR7613003_1.fastq -o SRR7613003_1.fastp_trimmed.fq -j
SRR7513003.fastp_trimmed.json -h SRR7613003_1.fastp_trimmed.html
fastp -i SRR7613004_1.fastq -o SRR7613004_1.fastp_trimmed.fq -j
SRR7513004.fastp_trimmed.json -h SRR7613004_1.fastp_trimmed.html
fastp -i SRR7613005_1.fastq -o SRR7613005_1.fastp_trimmed.fq -j
SRR7513005.fastp_trimmed.json -h SRR7613005_1.fastp_trimmed.html

#Run fastqc on trimmed files
fastqc *.fq

#Perform alignment/mapping with STAR
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR7613003_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE117808_BJ_HFF_rep1_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR7613004_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE117808_BJ_HFF_rep2_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR7613005_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE117808_BJ_HFF_rep3_

#Run htseq-count to count mapped reads
htseq-count -f bam -r pos -s no
GSE117808_BJ_HFF_rep1_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE117808_BJ_HFF_rep1_htseq.txt
htseq-count -f bam -r pos -s no
GSE117808_BJ_HFF_rep2_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE117808_BJ_HFF_rep2_htseq.txt
htseq-count -f bam -r pos -s no
GSE117808_BJ_HFF_rep3_Aligned.sortedByCoord.out.bam
```

```
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE117808_BJ_HFF_rep3_htseq.txt

#SKOV3_ctrl_SRR9694244_rep1, SKOV3_ctrl_SRR9694245_rep1,
SKOV3_ctrl_SRR9694246_rep1
fastq-dump --split-files --origfmt SRR7613003 SRR7613004 SRR7613005

#Trim reads
fastp -i SRR9694244_1.fastq -o SRR9694244_1.fastp_trimmed.fq -j
SRR9694244.fastp_trimmed.json -h SRR9694244_1.fastp_trimmed.html
fastp -i SRR9694245_1.fastq -o SRR9694245_1.fastp_trimmed.fq -j
SRR9694245.fastp_trimmed.json -h SRR9694245_1.fastp_trimmed.html
fastp -i SRR9694246_1.fastq -o SRR9694246_1.fastp_trimmed.fq -j
SRR9694246.fastp_trimmed.json -h SRR9694246_1.fastp_trimmed.html

#Run fastqc on trimmed files
fastqc *.fq

#Perform alignment/mapping with STAR
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694244_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_SKOV3_WT_rep1_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694245_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_SKOV3_WT_rep2_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694246_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_SKOV3_WT_rep3_

#Run htseq-count to count mapped reads
htseq-count -f bam -r pos -s no
GSE134375_SKOV3_WT_rep1_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_SKOV3_WT_rep1_htseq.txt
htseq-count -f bam -r pos -s no
GSE134375_SKOV3_WT_rep2_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_SKOV3_WT_rep2_htseq.txt
htseq-count -f bam -r pos -s no
GSE134375_SKOV3_WT_rep3_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_SKOV3_WT_rep3_htseq.txt
```

```
#OVCAR3_ctrl_SRR9694250_rep1, OVCAR3_ctrl_SRR9694251_rep1,
OVCAR3_ctrl_SRR9694252_rep1

fastq-dump --split-files --origfmt SRR7613003 SRR7613004 SRR7613005

#Trim reads
fastp -i SRR9694250_1.fastq -o SRR9694250_1.fastp_trimmed.fq -j
SRR9694250.fastp_trimmed.json -h SRR9694250_1.fastp_trimmed.html
fastp -i SRR9694251_1.fastq -o SRR9694251_1.fastp_trimmed.fq -j
SRR9694251.fastp_trimmed.json -h SRR9694251_1.fastp_trimmed.html
fastp -i SRR9694252_1.fastq -o SRR9694252_1.fastp_trimmed.fq -j
SRR9694252.fastp_trimmed.json -h SRR9694252_1.fastp_trimmed.html

#Run fastqc on trimmed files
fastqc *.fq

#Perform alignment/mapping with STAR
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694250_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_OVCAR3_WT_rep1_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694251_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_OVCAR3_WT_rep2_
STAR --runThreadN 12 --genomeDir
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/STAR_Index/ --readFilesIn
SRR9694252_1.fastp_trimmed.fq --outSAMtype BAM SortedByCoordinate --
outFileNamePrefix GSE134375_OVCAR3_WT_rep3_
#Run htseq-count to count mapped reads
htseq-count -f bam -r pos -s no
GSE134375_OVCAR3_WT_rep1_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_OVCAR3_WT_rep1_htseq.txt
htseq-count -f bam -r pos -s no
GSE134375_OVCAR3_WT_rep2_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_OVCAR3_WT_rep2_htseq.txt
htseq-count -f bam -r pos -s no
GSE134375_OVCAR3_WT_rep3_Aligned.sortedByCoord.out.bam
/mnt/d/Indexes/Homo_sapiens/Gencode_GRCh38.p13/gencode.v33.annotation.gtf >
GSE134375_OVCAR3_WT_rep3_htseq.txt
```

B.2 R pipeline for cell line and CT gene analysis

```
## VERSION 3 UPDATE NOTICE
# 3/21/22
# Removed the code that couldn't write plots as csv
# Updated figure and plot colors to be more consistent and help with colorblindness
# Added MA and Volcano plots to part 2 and increased log2FoldChange to be >2 from
>1

#Install Packages
install.packages("tidyverse")
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.14")
BiocManager::install("DESeq2")

#Load Libraries
library(tidyverse)
library(DESeq2)
library(readr)
library(ggVennDiagram)
library(pheatmap)
#
# Part 1: Cell line comparison
# The following code compares the expression of HFF, OVCAR3, and SKOV3 cell lines
#

## Compare all three together
#Get current directory path, where cell line files are stored
directory <- getwd()

#Get names of cell line files
cell_line_Files <- grep("rep._htseq",list.files(directory),value=TRUE)

#Assign conditions based on order they appear in sampleFiles
cell_line_Condition <-
data.frame(condition=rep(c("BJ_HFF","OVCAR3","SKOV3"),each=3))

#Create sampleTable for DESeq input using sampleFiles and sampleCondition
cell_line_Table <- data.frame(sampleName = cell_line_Files,
                              fileName = cell_line_Files,
                              condition = cell_line_Condition)

#Set the condition column to be a factor
```

```
cell_line_Table$condition <- factor(cell_line_Table$condition)

#Prep data for DESeq running using HTseq data
dds.Cell_Lines <- DESeqDataSetFromHTSeqCount(sampleTable = cell_line_Table,
                                directory = directory,
                                design=~condition)

#Run DESeq
ddsCL <- DESeq(dds.Cell_Lines)

#Create PCA plot of the cell line data
vsdCL <- vst(ddsCL)
plotPCA(vsdCL, intgroup = 'condition')
write.csv(plotPCA(vsdCL, intgroup = 'condition', returnData=T),
file="HFF_SKOV3_OVCAR3.PCA.csv")

# Create heatmap of cell lines
cell_line_ordered_counts <- order(rowMeans(counts(ddsCL,normalized=TRUE)),
decreasing=TRUE)
CL_ordered_counts_df <- as.data.frame(colData(ddsCL)[,c("condition","sizeFactor")])
CL_select <- subset(CL_ordered_counts_df,select=-c(sizeFactor))
pheatmap(assay(vsdCL)[cell_line_ordered_counts,], cluster_rows=FALSE,
show_rownames=FALSE, cluster_cols=FALSE, annotaton_col=CL_select)


## Compare HFF to SKOV3
# Get HFF file names
hffFiles <- grep(pattern="_BJ_HFF_rep",list.files(directory), value=T)

# Get SKOV3 file names
skov3Files <- grep(pattern="_SKOV3_WT",list.files(directory), value=T)

# Combine the two lists
hsFiles <- append(hffFiles, skov3Files)

# Create condition table
hsCondition <- data.frame(condition=rep(c("BJ_HFF","SKOV3"),each=3))

# Create table for DESeq2
hsTable <- data.frame(sampleName = hsFiles,
                      fileName = hsFiles,
                      condition = hsCondition)

#Prep for DESeq2
dds.hs <- DESeqDataSetFromHTSeqCount(sampleTable=hsTable,
```

```
                        directory=directory,
                        design=~condition)

#Run DESeq2
ddsHS<- DESeq(dds.hs)

# Extract Results
hs.res <- results(ddsHS, contrast=c("condition","BJ_HFF","SKOV3"))

# Convert to data frame
hs.res.df <- as.data.frame(hs.res)

# Extract baseMean, log2FoldChange, padj columns
hs.res.df2 <- hs.res.df%>%select(baseMean,log2FoldChange,padj)

# Remove NA rows
hs.res.df2 <- na.omit(hs.res.df2)

# Add significance coloumn
hs.res.df.labeled <- hs.res.df2%>%mutate(significance=ifelse(padj<=0.05 &
log2FoldChange>=2, "BJ_HFF",
                                        ifelse(padj<=0.05 & log2FoldChange<=-2,
"SKOV3",
                                               "Non-significant")))
# Calculate median adjusted p-value by significance
hs.res.df.labeled%>%group_by(significance)%>%summarize(median(padj))

# Extract highly expressed SKOV3 rows
highSKOV.hs <- hs.res.df.labeled%>%filter(padj<=0.05 & log2FoldChange <= -2)

# Create Volcano plot
#Orange=HFF, Cyan=SKOV3
ggplot(hs.res.df.labeled) +
  geom_point(aes(x=log2FoldChange, y=-log10(padj), colour=significance)) +
  ggtitle("SKOV3 vs BJ/HFF line") +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_colour_manual(values=c("orange", "grey", "darkcyan"))+
  geom_hline(yintercept = 1.3, colour="#990000", linetype="dashed") +
geom_vline(xintercept = 2, colour="#990000", linetype="dashed") +
geom_vline(xintercept = -2, colour="#990000", linetype="dashed")+
  theme_bw()

# Create MA plot
#Orange=HFF, Cyan=SKOV3
```

```
ggplot(hs.res.df.labeled)+
  geom_point(aes(x=baseMean, y=log2FoldChange, size=-log10(padj), color=-
log2FoldChange)) +
  ggtitle("MA Plot: SKOV3 vs HFF")+
  xlab("baseMean")+
  ylab("log2FoldChange")+
  theme_bw()+
  scale_color_gradient(low="darkcyan", high="orange")

## Compare HFF to OVCAR3
# Get OVCAR3 file names
ovcar3Files <- grep(pattern="_OVCAR3_WT",list.files(directory), value=T)

# Combine the two lists
hoFiles <- append(hffFiles, ovcar3Files)

# Create condition table
hoCondition <- data.frame(condition=rep(c("BJ_HFF","OVCAR3"),each=3))

# Create table for DESeq2
hoTable <- data.frame(sampleName = hoFiles,
                      fileName = hoFiles,
                      condition = hoCondition)

#Prep for DESeq2
dds.ho <- DESeqDataSetFromHTSeqCount(sampleTable=hoTable,
                                     directory=directory,
                                     design=~condition)

#Run DESeq2
ddsHO<- DESeq(dds.ho)

# Extract Results
ho.res <- results(ddsHO, contrast=c("condition","BJ_HFF","OVCAR3"))

# Convert to data frame
ho.res.df <- as.data.frame(ho.res)

# Extract baseMean, log2FoldChange, padj columns
ho.res.df2 <- ho.res.df%>%select(baseMean,log2FoldChange,padj)

# Remove NA rows
ho.res.df2 <- na.omit(ho.res.df2)

# Add significance coloumn
```

```
ho.res.df.labeled <- ho.res.df2%>%mutate(significance=ifelse(padj<=0.05 &
log2FoldChange>=2, "BJ_HFF",
                                        ifelse(padj<=0.05 & log2FoldChange<=-2,
"OVCAR3",
                                               "Non-significant")))
# Calculate median adjusted p-value by significance
ho.res.df.labeled%>%group_by(significance)%>%summarize(median(padj))

# Extract highly expressed OVCAR3 rows
highOVCAR.ho <- ho.res.df.labeled%>%filter(padj<=0.05 & log2FoldChange <= -2)

# Create Volcano plot
# Orange=HFF Violet=OVCAR
ggplot(ho.res.df.labeled) +
  geom_point(aes(x=log2FoldChange, y=-log10(padj), colour=significance)) +
  ggtitle("OVCAR3 line vs BJ/HFF line") +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_colour_manual(values=c("orange", "grey", "darkviolet"))+
  geom_hline(yintercept = 1.3, colour="#990000", linetype="dashed") +
geom_vline(xintercept = 2, colour="#990000", linetype="dashed") +
geom_vline(xintercept = -2, colour="#990000", linetype="dashed")+
  theme_bw()

# Create MA plot
# Orange=HFF Violet=OVCAR
ggplot(ho.res.df.labeled)+
  geom_point(aes(x=baseMean, y=log2FoldChange, size=-log10(padj), color=-
log2FoldChange)) +
  ggtitle("MA Plot: OVCAR3 vs BJ/HFF")+
  xlab("baseMean")+
  ylab("log2FoldChange")+
  theme_bw()+
  scale_color_gradient(low="darkviolet", high="orange")

## Compare SKOV3 to OVCAR3
# Combine the two lists
soFiles <- append(skov3Files, ovcar3Files)

# Create condition table
soCondition <- data.frame(condition=rep(c("SKOV3","OVCAR3"),each=3))

# Create table for DESeq2
soTable <- data.frame(sampleName = soFiles,
                      fileName = soFiles,
```

```
                condition = soCondition)

#Prep for DESeq2
dds.so <- DESeqDataSetFromHTSeqCount(sampleTable=soTable,
                            directory=directory,
                            design=~condition)

#Run DESeq2
ddsSO<- DESeq(dds.so)

# Extract Results
so.res <- results(ddsSO, contrast=c("condition","SKOV3","OVCAR3"))

# Convert to data frame
so.res.df <- as.data.frame(so.res)

# Extract baseMean, log2FoldChange, padj columns
so.res.df2 <- so.res.df%>%select(baseMean,log2FoldChange,padj)

# Remove NA rows
so.res.df2 <- na.omit(so.res.df2)

# Add significance coloumn
so.res.df.labeled <- so.res.df2%>%mutate(significance=ifelse(padj<=0.05 &
log2FoldChange>=2, "SKOV3",
                                        ifelse(padj<=0.05 & log2FoldChange<=-2,
"OVCAR3",
                                               "Non-significant")))
# Calculate median adjusted p-value by significance
so.res.df.labeled%>%group_by(significance)%>%summarize(median(padj))

# Extract highly expressed OVCAR3 rows
highOVCAR.so <- so.res.df.labeled%>%filter(padj<=0.05 & log2FoldChange <= -2)

# Create Volcano plot
# cyan=SKOV Violet=OVCAR
ggplot(so.res.df.labeled) +
  geom_point(aes(x=log2FoldChange, y=-log10(padj), colour=significance)) +
  ggtitle("OVCAR3 line vs SKOV3 line") +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_colour_manual(values=c("grey", "darkviolet", "darkcyan"))+
  geom_hline(yintercept = 1.3, colour="#990000", linetype="dashed") +
geom_vline(xintercept = 2, colour="#990000", linetype="dashed") +
geom_vline(xintercept = -2, colour="#990000", linetype="dashed")+
```

```
  theme_bw()

# Create MA plot
# Cyan=SKOV, Vilet=OVCAR
ggplot(so.res.df.labeled)+
  geom_point(aes(x=baseMean, y=log2FoldChange, size=-log10(padj), color=-
log2FoldChange)) +
  ggtitle("MA Plot: OVCAR3 vs SKOV3")+
  xlab("baseMean")+
  ylab("log2FoldChange")+
  theme_bw()+
  scale_color_gradient(low="darkviolet", high="cyan3")

#
# Part 2: DESeq of Ovary and Testis
# The following part compares expression of normal ovarian tissue to cancerous
# ovarian tissue, and normal ovarian tissue to normal testis tissue

# Get necessary files
OVARY_GTEX <- list.files(pattern="OVARY_GTEX-")
OVARY_TCGA <- list.files(pattern="TCGA_OVARY_")
TESTIS_GTEX <- list.files(pattern="TESTIS_GTEX-")

#Combine GTEX data and create TCGA data
GTEXData <- c(OVARY_GTEX, TESTIS_GTEX)
TCGAData <- c(OVARY_TCGA)

# Pull in metadata
metaData <- read.csv('meta_file_v3.csv',header = TRUE)

# Searching metadata for the given tissue types
h <- c()
for(i in 1:length(GTEXData)){
  h <- append(h, which(metaData$SampleName == GTEXData[i]))
}
sampGTEX <- metaData[h,]

j <- c()
for(i in 1:length(TCGAData)){
  j <- append(j, which(metaData$SampleName == TCGAData[i]))
}
sampTCGA <- metaData[j,]

# Condense metadata sets
metaSamp <- rbind(sampTCGA, sampGTEX)
```

```
## DESeq2 performed on the normal tissues
NormalTable <- data.frame(sampleName = sampGTEX$SampleName, fileName =
sampGTEX$SampleName, condition = sampGTEX$Condition, primary_diagnosis =
sampGTEX$primary_diagnosis, gender = sampGTEX$gender, stage =
sampGTEX$tumor_stage)
ddsNormal <- DESeqDataSetFromHTSeqCount(sampleTable = NormalTable, directory
= getwd(), design =~condition)
ddsN <- DESeq(ddsNormal)
metaSamp <- droplevels(metaSamp)

# Extract results
ddsN.res <- results(ddsN, contrast=c("condition", "Germline_M", "Germline_F"))

# Convert to data frame
ddsN.res.df <- as.data.frame(ddsN.res)

# Extract baseMean, log2FoldChange, padj columns
ddsN.res.df2 <- ddsN.res.df%>%select(baseMean,log2FoldChange,padj)

# Remove NA rows
ddsN.res.df2 <- na.omit(ddsN.res.df2)

# Add significance column
ddsN.res.df.labeled <- ddsN.res.df2%>%mutate(significance=ifelse(padj<=0.01 &
log2FoldChange>=2, "Testis",
                                                 ifelse(padj<=0.01 & log2FoldChange<=-
2,"Ovary",
                                                        "Non-significant")))

# Calculate median adjusted p-value by significance
ddsN.res.df.labeled%>%group_by(significance)%>%summarize(median(padj))

# Create Volcano plot
# Teal=Testis Orange=Ovary
ggplot(ddsN.res.df.labeled) +
  geom_point(aes(x=log2FoldChange, y=-log10(padj), colour=significance)) +
  ggtitle("Testis tissue vs Ovary tissue (normal)") +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_colour_manual(values=c("grey", "orange", "darkcyan"))+
  geom_hline(yintercept = 1.3, colour="#990000", linetype="dashed") +
geom_vline(xintercept = 2, colour="#990000", linetype="dashed") +
geom_vline(xintercept = -2, colour="#990000", linetype="dashed")+
  theme_bw()
```

```
# Create MA plot
# Teal=Testis Orange=Ovary
ggplot(ddsN.res.df.labeled)+
  geom_point(aes(x=baseMean, y=log2FoldChange, size=-log10(padj), color=-
log2FoldChange)) +
  ggtitle("MA Plot: Testis tissue vs Ovary tissue (normal)")+
  xlab("baseMean")+
  ylab("log2FoldChange")+
  theme_bw()+
  scale_color_gradient(low="darkcyan", high="orange")

#
# PCA plot of the DESeq2 of normal tissues
vsdN <- vst(ddsN)
plotPCA(vsdN, intgroup = 'condition') ##Export as PNG and PDF
write.csv(plotPCA(vsdN, intgroup = 'condition', returnData =
T),file="Ovary.normal.PCA.csv")

## DESeq2 performed on cancer tissues
CancerTable <- data.frame(sampleName = sampTCGA$SampleName, fileName =
sampTCGA$SampleName, condition = sampTCGA$Condition, primary_diagnosis =
sampTCGA$primary_diagnosis, gender = sampTCGA$gender, stage =
sampTCGA$tumor_stage)
ddsCancer <- DESeqDataSetFromHTSeqCount(sampleTable = CancerTable, directory =
getwd(), design = ~1)
ddsC <- DESeq(ddsCancer)

# PCA FOR CANCER DATA
vsdC <- vst(ddsC)
plotPCA(vsdC, intgroup = 'condition') ##Export as PNG and PDF
write.csv(plotPCA(vsdC, intgroup = 'condition', returnData =
T),file="Ovary.tumor.PCA.csv")

# Getting rid of extra decimals, letters on the end of gene names, and combining normal +
germline data sets with cancer data sets
Normal1 <- counts(ddsN, normalized = FALSE)
x <- row.names(Normal1)
x <- gsub("\\..*","",x)
row.names(Normal1) <- x
Cancer1 <- counts(ddsC, normalized = FALSE)
y <- row.names(Cancer1)
y <- gsub("\\..*","",y)
row.names(Cancer1) <- y
Merge <- merge(Cancer1, Normal1, by = "row.names")
```

```
write.csv(Merge, file="Ovary_unnormalized.csv")

# DESeq2 performed on cancer and normal data together
ddsMerge <- DESeqDataSetFromMatrix(countData = Merge, colData = metaSamp,
design = ~Condition, tidy = TRUE)
ddsM <- DESeq(ddsMerge)
write.csv(counts(ddsM, normalized=TRUE), file="Ovary_normalized.csv")

# Extract Results and convert to dataframe
ddsM.res <- results(ddsM, contrast = c("Condition","Tumor","Germline_F"))
ddsM.res.df <- as.data.frame(ddsM.res)

# Extract baseMean, log2FoldChange, padj columns
ddsM.res.df2 <- ddsM.res.df%>%select(baseMean,log2FoldChange,padj)

# Remove NA rows
ddsM.res.df2 <- na.omit(ddsM.res.df2)

# Add significance column
ddsM.res.df.labeled <- ddsM.res.df2%>%mutate(significance=ifelse(padj<=0.01 &
log2FoldChange>=2, "Tumor",
                                                 ifelse(padj<=0.01 & log2FoldChange<=-
2,"Ovary",
                                                        "Non-significant")))

# Calculate median adjusted p-value by significance
ddsM.res.df.labeled%>%group_by(significance)%>%summarize(median(padj))

# Create Volcano plot
# Purple=Tumor Orange=Ovary
ggplot(ddsM.res.df.labeled) +
  geom_point(aes(x=log2FoldChange, y=-log10(padj), colour=significance)) +
  ggtitle("Ovary tissue (cancer) vs Ovary tissue (normal)") +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  scale_colour_manual(values=c("grey", "orange", "darkviolet"))+
  geom_hline(yintercept = 1.3, colour="#990000", linetype="dashed") +
geom_vline(xintercept = 2, colour="#990000", linetype="dashed") +
geom_vline(xintercept = -2, colour="#990000", linetype="dashed")+
  theme_bw()

# Create MA plot
# Purple=Tumor Orange=Ovary
ggplot(ddsM.res.df.labeled)+
```

```
  geom_point(aes(x=baseMean, y=log2FoldChange, size=-log10(padj), color=-
log2FoldChange)) +
  ggtitle("MA Plot: Ovary tissue (cancer) vs Ovary tissue (normal)")+
  xlab("baseMean")+
  ylab("log2FoldChange")+
  theme_bw()+
  scale_color_gradient(low="darkviolet", high="orange")

# PCA of merged data
vsd <- vst(ddsM, blind = FALSE)
plotPCA(vsd, intgroup = c("Condition"))
write.csv(plotPCA(vsd, intgroup = 'Condition', returnData = T),file="Ovary.PCA.csv")

# Extracting the gene list and number of genes
ddsM$Condition <- as.factor(ddsM$Condition)
ddsMerge$Condition <- as.factor(ddsMerge$Condition)

# Number of genes UP in testis VS normal (not cancers!!)
TvN <- results(ddsM, contrast = c('Condition', 'Germline_M', 'Germline_F'))
which1 <- which(TvN$padj < 0.01 & TvN$log2FoldChange > 2)
TvN_count <- length(which1)
write.csv(as.data.frame(TvN), file="Ovary_TestisvNormal2.csv")

# Number of genes UP in cancer VS normal
CvN <- results(ddsM, contrast = c('Condition', 'Tumor', 'Germline_F'))
which2 <- which(CvN$padj < 0.01 & CvN$log2FoldChange > 2)
CvN_count <- length(which2)
write.csv(as.data.frame(CvN), file="Ovary_TumorvNormal2.csv")

## Create vennDiagram
TvN_genes <- row.names(TvN[which1,])
CvN_genes <- row.names(CvN[which2,])
CT.venn <- list(A=TvN_genes,B=CvN_genes)
ggVennDiagram(CT.venn, category.names=c("Up in Testis", "Up in Cancer")) +
scale_fill_gradient(low="darkviolet", high="orange")

## Gene list for OVARY: Cancer+germlineTissue
geneNames <- CvN[intersect(which1, which2),]
genes <- row.names(geneNames)
write.table(genes, file = 'OVARY_genes_Cancer+germlineTissue2.txt', sep = ",",
row.names = FALSE)

## Generate heatmap
select <- order(rowMeans(counts(ddsM,normalized=TRUE)), decreasing=TRUE)
df <- as.data.frame(colData(ddsM)[,c("Condition","Tissue")])
```

```
pheatmap(assay(vsd)[select,], cluster_rows=FALSE, show_rownames=FALSE,
cluster_cols=FALSE, annotation_col=df)

#
# Part 3: Compare CT genes to Cell Lines
# Using the genes found in part 2, run DESeq2 with the ovary cancer tissues on the
# HFF, OVCAR3, and SKOV3 tissues to get gene list with the following criteria:
# 1. Down regulated in HFF
# 2. Up regulated in SKOV3
# 3. Higher up regulation in OVCAR3
#
# v2 Update: gene are normalized so they can be compared to CT gene list

# Extract genes from the Cell Line lists
highOVCAR.ho.genes <- rownames(highOVCAR.ho)
highOVCAR.so.genes <- rownames(highOVCAR.so)
highSKOV.hs.genes <- rownames(highSKOV.hs)

# Normalize gene names to compare against CT gene list
highOVCAR.ho.genes.norm <- gsub("\\..*","",highOVCAR.ho.genes)
highOVCAR.so.genes.norm <- gsub("\\..*","",highOVCAR.so.genes)
highSKOV.hs.genes.norm <- gsub("\\..*","",highSKOV.hs.genes)

## Create vennDiagram of the Cell Line genes
#create list of number of genes in each comparison
CL.venn <-
list(A=highSKOV.hs.genes.norm,B=highOVCAR.ho.genes.norm,C=highOVCAR.so.gen
es.norm)

#plot vennDiagram
ggVennDiagram(CL.venn, category.names = c("SKOV3 > BJ/HFF","OVCAR3 >
BJ/HFF", "OVCAR3 > SKOV3"))+
  scale_fill_gradient(low="orange",high="darkcyan")

# In order to get a list of genes meeting the criteria, the intersect of these three gene lists
is taken
patternGenes <- intersect(intersect(highSKOV.hs.genes.norm,
highOVCAR.ho.genes.norm), highOVCAR.so.genes.norm)
write.csv(patternGenes, "PatternGenes.csv")

# Check if any ovarian ct genes are in this set
checkGenes <- intersect(genes, patternGenes)
write.csv(checkGenes, "ResultsGenes.csv")

## Create vennDiagram of overlapping genes
```

```
final.venn <- list(A=genes,B=patternGenes)
ggVennDiagram(final.venn,category.names = c("CT genes","Cell Line"))+
  scale_fill_gradient(low="darkviolet", high="cyan4")
```