Statistical Methods Applied to Gene Expression Data to Explore Cancer Features

HANNAH SIMPSON-CLANCY*

Project advisor: A. Smith †

Abstract. A small number of epithelial ovarian cancer cases are deemed preventable and its overall survival rates are low. The developments in omics data analysis paves a way for biomarker discovery for epithelial ovarian cancer in order to improve survival rates and prevent its development. This report provides analysis of gene expression data for epithelial ovarian cancer to compare the gene expression of 99 epithelial ovarian cancer samples and 4 non-cancerous ovary samples. Serous and endometrioid epithelial ovarian cancer subtypes were most similar based on hierarchical clustering. Serous was the subtype with the most differentially expressed genes when compared with normal ovary samples whereas mucinous had the least by the Wilcoxon Rank-Sum test (Benjamini Hochberg, $p < 0.05$). The number of down-regulated genes exceeded the number of up-regulated genes when comparing each cancer subtype with normal ovary samples. In this case, the clear cell subtype had the greatest number of dysregulated genes when compared to normal ovaries whereas endometrioid had the least. The dysregulated genes were found by fold change analysis ($FC > 2$ or $FC < 0.5$). Differences in gene expression levels between epithelial ovarian cancer subtypes were suggested due to $11,181$ differentially expressed genes identified when comparing expression levels in all sample groups by the Kruskal-Wallis test (Benjamini Hochberg, $p < 0.05$). Genes proposed as biomarkers for (1) epithelial ovarian cancer and (2) individual epithelial ovarian cancer subtypes, when compared with normal ovaries included MUC1, SCNN1A, CD24, ITM2A, AGR2 and WFDC2.

Key words. gene expression, epithelial ovarian cancer, statistics

1. Introduction.

Ovarian cancer represents 2% of all cancer cases in the UK with approximately $7,500$ new cases diagnosed annually between 2017 and 2019. It is estimated that from 2013 to 2017 in England and Wales the survival of 35% of patients diagnosed with ovarian cancer exceeded 10 years. On estimation, in the UK only 11% of cases in 2015 were deemed preventable [6]. In the US, $19,880$ new cases were predicted for 2022 along with $12,810$ deaths [58]. For all cases of ovarian cancer, approximately 90% are recorded as epithelial ovarian cancer; of which there are four types including clear cell, endometrioid, mucinous and serous, with serous being most prevalent [25].

Omics is defined as the study of biomolecules (examples of which include genomics, proteomics and metabolomics) [29]. Studies have provided steps to developing personalised drug treatments [11, 67], allowed detection of biomolecules (biomarkers) that indicate presence or absence of a disease [8, 12, 44] and a means to compare diseases [28]. Statistical analysis of omics data [1, 13, 27] and multi-omics data [36, 37, 69] aids these developments. For example, statistical analysis of gene expression data has led to research developments in precision medicine [20, 61, 34], prediction of survival [2] and responses to treatment [7, 49].

Studies on gene expression for ovarian cancer date back as far as the 1990s [4, 19, 42]. Aims of

* (h.sc@hotmail.co.uk, University of Huddersfield, UK).
† (a.smith@hud.ac.uk, University of Huddersfield, UK)

analysing gene expression in ovarian cancers are to aid the development of treatments, earlier diagnosis and improvement of survival chances for patients. Genetic differential expression in ovarian cancer is a topic researched abundantly in the 2000s; a multitude of statistical analyses have taken place to distinguish between normal and cancerous ovarian samples. Often multiple forms of statistical analysis are performed on data. A small number of analytical methods include hierarchical clustering [15], fold change analysis [63, 48, 22], hypothesis testing [63, 22, 60, 57, 38, 47, 46], Pearson's correlation [23] and meta-analysis [31, 52, 53].

This study will explore statistical analysis of epithelial ovarian cancer gene expression data through the application of hierarchical clustering, fold change analysis and hypothesis testing. The detection of both differentially expressed and dysregulated genes when comparing epithelial ovarian cancer samples with normal ovary samples will allow for biomarker suggestion. Furthermore, this paper will explore the similarities between subtypes of epithelial ovarian cancer to determine which subtypes are most similar. Analysis will be performed on the supplementary data set made available by [66]. The expression measures of $22,283$ genes are provided for 103 human samples consisting of 4 normal ovarian cell samples and 99 epithelial ovarian cancer cell samples separated into 4 groups. These 4 groups consist of 8 clear cell, 37 endometrioid, 13 mucinous and 41 serous samples. Quantile normalization with trimmed means and a log_2 transformation have previously been performed on the raw data. Statistical methods such as ANOVA and fold change analysis have also been performed. The raw data has previously been included in a large-scale meta-analysis to identify core genes in ovarian cancer [35]. This report will provide a more in depth analysis of the gene expressions provided in the supplementary data set [66].

2. Methods.

The data analysed in this study describes the gene expression levels of $22,283$ genes for 103 samples. The samples are made up of 4 normal ovary samples and 8 clear cell, 37 endometrioid, 13 mucinous and 41 serous epithelial ovarian cancer samples.

Hierarchical clustering with a Euclidean proximity measure and Ward's linkage [9, 16] was performed to determine which samples were most similar based on their gene expression levels.

The Wilcoxon Rank-Sum Test [54] was applied to determine whether differentially expressed genes existed when comparing normal ovary samples with samples from each subtype of epithelial ovarian cancer. The Kruskal-Wallis Test [55] was applied to the data to compare the gene expression levels of all five groups to determine whether differentially expressed genes existed between at least one pair of groups. Both hypothesis tests included a Benjamini-Hochberg (BH) test correction [3] in which a p-value<0.05 was applied. The 20 most differentially expressed genes detected by the Kruskal-Wallis test were hierarchically clustered using Pearson's proximity measure and average-linkage.

Fold change analysis [33] was performed to determine differences in expression levels of genes between normal ovaries (control) and the epithelial ovarian cancer (condition) groups based on the ratio of the mean expression levels of each gene. The log_2 fold change of a gene calculated by

$$log_2(FC) = log_2(\bar{a}_i) - log_2(\bar{a}_c) \tag{2.1}$$

In which $\bar{a}_i$ is the mean expression level of a gene for all samples in a condition group, $\bar{a}_c$ is the mean expression level of a gene for all samples in the control group, FC is the fold change value.

The log_2 fold change value can be transformed in order to calculate the fold change value of a gene as follows.

$$FC = 2^{(log_2(\bar{a}_i) - log_2(\bar{a}_c))} \tag{2.2}$$

Genes are described as up-regulated if they have a fold change value greater than a cut-off C

$$FC > C \tag{2.3}$$

Genes are described as down-regulated if the fold change value is less than the cut off $\frac{1}{C}$.

$$FC < \frac{1}{C} \tag{2.4}$$

Up-regulation indicates a condition group has greater expression level of a gene than the control group and is described as up-regulated by C-fold. Down-regulation implies a control group has greater expression level of a gene than the condition group and is described to be down-regulated by C-fold.

3. Results.

3.1. Hierarchical Clustering.

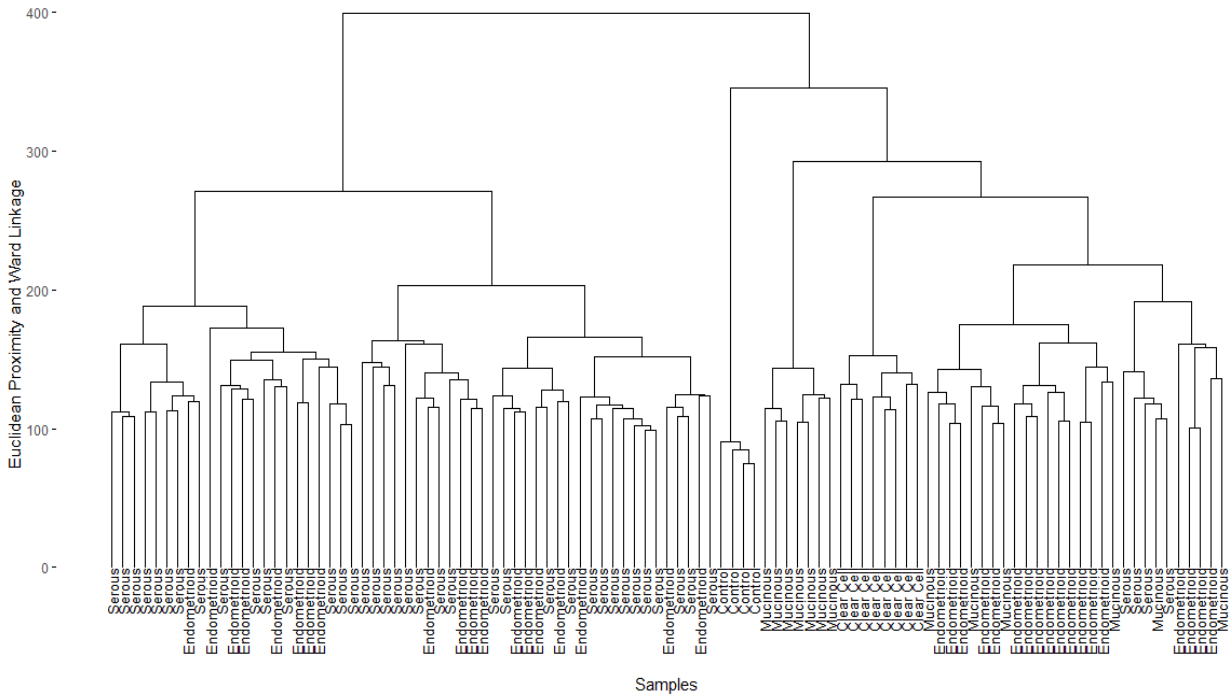

Figure 1. *Hierarchical Clustering Dendrogram of 103 Samples Using Euclidean Proximity Measure and Ward-Linkage Method.*

Figure 1 presents the dendrogram produced when hierarchically clustering the 103 samples. The control samples formed one branch and were the last group to join a cluster. This suggests the normal ovary samples differed the most in expression levels when comparing gene expression levels in all five groups. All clear cell samples also formed one branch implying their gene expression levels were most unique when compared to other sample types. Clear cell could be easier to distinguish in comparison to other cancer subtypes based on its expression of genes. Endometrioid and serous samples formed one large cluster which may suggest they are the two most similar subtypes of epithelial ovarian cancer based on their gene expression levels.

3.2. Wilcoxon-Rank Sum Test. The aim of the Wilcoxon Rank-Sum test was to determine whether any of the $22,283$ genes varied significantly in expression level between epithelial ovarian cancer subtypes and normal ovaries. A BH adjusted p-value with a significance level of 5% indicated that the distribution of expression levels in a cancer group and the control group were not equal for the gene tested; the gene is differentially expressed. All gene symbols and titles for gene tables were taken directly from the supplementary data set [66].

3.2.1. Clear Cell vs Control Samples. A total of $8,764$ genes were found with $p < 0.05$ indicating the rejection of the null hypothesis and differential expression of these genes. When

the BH test correction was applied, 6,329 statistically significant p-values were detected. Table SM1 presents 20 of the most differentially expressed genes and corresponding BH adjusted p-values for this comparison. The smallest p-value presented in table SM1 is 0.036943; 5,107 genes had this p-value. Based on the BH p-value adjustments, all 20 genes in this table were equally significant.

3.2.2. Endometrioid vs Control Samples. The test detected 9,770 genes for $p < 0.05$ indicating statistical significance and null hypothesis rejection prior to the BH test correction. The BH test correction detected 6,988 statistically significant p-values; 20 of the most differentially expressed genes and corresponding BH adjusted p-values are presented in table SM2. Of the 6,988 differentially expressed genes, 2,696 were observed with the smallest p-value of 0.018677 and all 20 genes in table SM2 were considered equally significant.

3.2.3. Mucinous vs Control Samples. A comparison of mucinous epithelial ovarian cancer and normal ovary samples found 9,669 genes had statistically significant p-values prior to BH test correction. The BH test correction detected 5,546 statistically significant p-values. Table SM3 presents 20 of the most differentially expressed genes and corresponding BH adjusted p-values. There were 3,054 differentially expressed genes with a p-value of 0.028331 including the genes in table SM3.

3.2.4. Serous vs Control Samples. The comparison of serous epithelial ovarian cancer and normal ovary samples detected 9,906 genes with statistically significant p-values prior to BH corrections. The BH test correction detected 7,260 statistically significant p-values. Table SM4 presents 20 of the most differentially expressed genes with corresponding BH adjusted p-values. The two most differentially expressed genes found when comparing serous and control samples were MCPH1 and GLP1R with p-values of 0.008 and 0.010688, respectively. All other genes in table SM4 had the p-value 0.017239; 2,576 genes in total had this p-value.

3.3. Fold Change Analysis.

3.3.1. Clear Cell vs Control Samples. When comparing clear cell and control group expression levels, 2,317 genes were up-regulated and 2,752 genes were down-regulated. Tables SM5 and SM6 present the top 20 up-regulated and down-regulated genes, respectively. MUC1 was the most up-regulated gene by 345-fold in clear cell samples compared with control samples. The genes CLDN3 and CD24 were observed multiple times within the 20 most up-regulated genes. MAOB was the most down-regulated gene by 50-fold. The fold change values were more significant for the 20 most up-regulated genes than the 20 most down-regulated genes.

3.3.2. Endometrioid vs Control Samples. Performing fold change analysis to compare control and endometrioid samples found 1,724 up-regulated genes and 2,021 down-regulated genes. Tables SM7 and SM8 present the top 20 up-regulated and down-regulated genes, respectively. MUC1 was the most up-regulated gene and was up-regulated by 128-fold. Both CLDN3 and CD24 were observed multiple times in the 20 most up-regulated genes list. STAR was the most down-regulated gene and was down-regulated by 37-fold. For this comparison, the up-regulation of the top 20 genes were more significant than the top 20 down-regulated genes.

158 **3.3.3. Mucinous vs Control Samples.** The fold change analysis on mucinous and control
159 sample groups detected 2,008 up-regulated genes and 2,397 down-regulated genes. Tables
160 SM9 and SM10 present the top 20 up-regulated and down-regulated genes for this case, re-
161 spectively. MUC1 was the gene with the greatest fold change value; it was up-regulated by
162 278-fold with greater expression in mucinous samples. CD24 was observed multiple times
163 within the top 20 up-regulated gene table. ADAMTS1 was the most down-regulated gene at
164 20-fold.

165 **3.3.4. Serous vs Control Samples.** Fold change analysis for the groups serous and control
166 detected 1,753 up-regulated and 2,063 down-regulated genes. Tables SM11 and SM12 present
167 the top 20 up-regulated and down-regulated genes, respectively. MUC1 was the most up-
168 regulated gene in the comparison of serous and control sample groups and was up-regulated
169 by 224-fold. CLDN3 and CD24 were observed multiple times within the top 20 gene table.
170 The most down-regulated gene found between serous and control groups was STAR which
171 was down-regulated by 113-fold.

172 **3.4. Kruskal-Wallis.** The Kruskal-Wallis test compared the gene expression in normal
173 ovary, clear cell, endometrioid, mucinous and serous epithelial ovarian cancer samples to de-
174 termine whether any genes were differentially expressed between at least one pair of sample
175 types. The total number of statistically significant p-values found when performing Kruskal-
176 Wallis on the 22,283 genes was 13,667; the BH test correction reduced this number to 11,181.
177 Table SM13 presents the top 20 differentially expressed genes and corresponding BH adjusted
178 p-values. EPS8 was the most differentially expressed gene when comparing all five groups
179 followed by AGR2, TSPAN1 and PEA15. Both observations of PEA15 were found in the top
180 20 differentially expressed gene list and one was more significant than the other.
181 The top 20 differentially expressed genes detected by the Kruskal-Wallis test were hierarchi-
182 cally clustered based on their mean expression level for each group (see SM14).

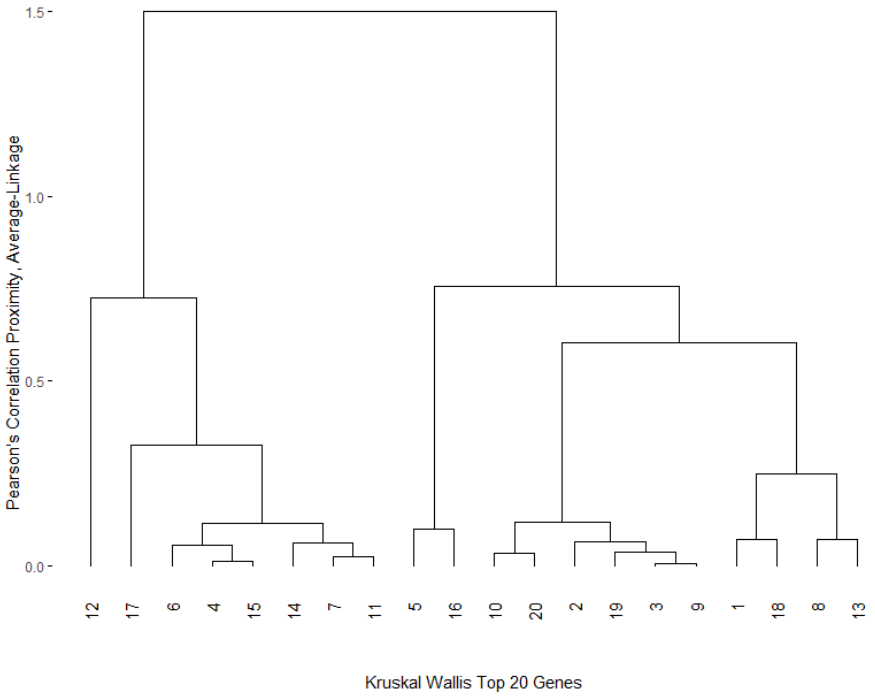

Figure 2. *Hierarchical clustering of Kruskal-Wallis top 20 genes; Pearson proximity and average linkage.*

Figure 2 presents the dendrogram for the 20 most differentially expressed genes determined by the Kruskal-Wallis test.

When considering the mean expression levels of gene 12 (GAS1) displayed in table SM14, clear cell epithelial ovarian cancer presented the lowest expression of GAS1 when compared with all other cancer types and normal ovary samples. Endometrioid and mucinous cancer groups had similar mean expression levels of GAS1 (8.837596 vs 8.518631). In addition, serous and the control groups had similar mean expression levels of GAS1 (11.273543 vs 12.106772). This suggests GAS1 was differentially expressed in clear cell epithelial ovarian cancer compared to all other groups. Also that GAS1 was differentially expressed when comparing endometrioid and mucinous epithelial ovarian cancer with serous epithelial ovarian cancer and normal ovaries.

Clear cell, mucinous and control groups had similar mean expression levels of genes 6 (PEA15), 4 (PEA15), 15 (CRABP2), 14 (PNOC), 7 (CLDN16) and 11 (KLK5), whereas serous displayed greater mean expression levels as indicated by table SM14. This suggests differential expression of these genes when comparing clear cell, mucinous and normal ovaries with serous epithelial ovarian cancer.

Mucinous had the highest mean expression levels of genes 5 (TFF3) and 16 (IQGAP2) when

compared with all other sample types. Clear cell epithelial ovarian cancer mean expression levels for these genes differed from normal ovaries. These mean expression levels can be found in table SM14. Differential expression of these genes in mucinous epithelial ovarian cancer compared with the other ovary types was indicated. In addition, these genes may have been some of the differentially expressed genes between clear cell epithelial ovarian cancer and normal ovaries not included in table SM1.

Clear cell and mucinous epithelial ovarian cancer displayed a higher mean expression level of genes 10 (SLC3A1), 20 (CYP2C9), 2 (AGR2), 19 (USH1C), 3 (TSPAN1) and 9 (ACADS) than other ovary types (see table SM14), suggesting their differential expression in this case.

Endometrioid and serous epithelial ovarian cancer had lower mean expression levels of genes 1 (EPS8), 18 (ARL1), 8 (TTC38) and 13 (PPAP2A) than other ovary types (see table SM14), implying the differential expression of these genes in endometrioid and serous epithelial ovarian cancer compared with clear cell, mucinous and normal ovary types.

3.5. Test Result Comparisons. Providing an overall comparison of the Wilcoxon Rank-Sum tests, serous epithelial ovarian cancer had the greatest number of differentially expressed genes $(7,260)$ when compared with control samples. Mucinous had the least number of differentially expressed genes compared to control samples $(5,546)$. However, when considering the fold change analysis results, clear cell samples had the greatest overall number of dysregulated genes when compared with control samples $(5,067)$. Endometrioid samples displayed the least dysregulated genes when compared with control samples $(3,745)$.

Comparisons of differentially expressed genes in Wilcoxon Rank-Sum tests found endometrioid and serous groups when compared with the control group shared the most differentially expressed genes (approximately 84% and 80.9%, respectively). Whereas, approximately 69% of clear cell differentially expressed genes and 77% of mucinous differentially expressed genes were also detected as differentially expressed in endometrioid samples when all three groups were compared with the control group.

When comparing the four fold change analyses results with the Wilcoxon Rank-Sum test results, approximately 61% of differentially expressed genes between clear cell and control groups, 46% of differentially expressed genes between endometrioid and control groups, 57% of differentially expressed genes between mucinous and the control group and 47% of differentially expressed genes between serous and the control group were also found to be dysregulated when fold change analysis was performed to compare these same groups. There were $1,700$ genes both differentially expressed and up-regulated compared to $2,176$ genes both differentially expressed and down-regulated in clear cell samples compared to control samples. A total of $1,398$ genes were both differentially expressed and up-regulated compared to $1,819$ genes differentially expressed and down-regulated in endometrioid samples compared with control samples. The total number of genes differentially expressed and up-regulated between mucinous and control samples was $1,370$ compared with $1,818$ genes differentially expressed and down-regulated. There were $1,503$ genes observed as differentially expressed and up-regulated

between serous and control samples compared to the 1,928 differentially expressed and down-regulated.

MUC1 was the gene most up-regulated for each fold change analysis; it was up-regulated by 345-fold, 128-fold, 278-fold and 224-fold for clear cell, endometrioid, mucinous and serous epithelial ovarian cancer when compared with normal ovaries, respectively. Five observations of CD24 were within the 20 most up-regulated genes for all four cancer groups compared with the control groups. SCNN1A was one of the 20 most up-regulated genes by 194-fold, 85-fold, 95-fold and 124-fold for clear cell, endometrioid, mucinous and serous epithelial ovarian cancer, respectively. ITM2A is the only gene from this analysis that was one of the 20 most down-regulated genes by 38-fold, 20-fold, 16-fold and 21-fold in clear cell, endometrioid, mucinous and serous epithelial ovarian cancer when compared with normal ovaries. These four genes were also some of the most differentially expressed genes when comparing each cancer group with the control group by the Wilcoxon Rank-Sum test.

Many genes were both differentially expressed and dysregulated for a single cancer group compared with the control group. For example, AGR2, CEACAM6, ST14, SLC44A4 and S100P were within the most differentially expressed and up-regulated genes for mucinous compared with control samples. They were up-regulated by 199-fold, 69-fold, 38-fold, 37-fold and 31-fold, respectively. EFEMP1, ADAMTS1, TRO, PRELP, WT1, SLC4A3 and AOX1 were within the most differentially expressed and down-regulated genes for mucinous compared with control groups. These genes were down-regulated by approximately 17-fold, 19-fold, 19-fold, 16-fold, 15-fold, 13-fold and 13-fold, respectively.

Genes indicating differential expression and dysregulation in clear cell samples compared with control samples include LBP, HGD, GAS1, NR4A2, GULP1, RGS2, ANG and PELI2. LBP and HGD were up-regulated in clear cell samples by 78-fold and 66-fold, respectively. Whereas GAS1, NR4A2, GULP1, RGS2, ANG and PELI2 were down-regulated in clear cell samples by approximately 40-fold, 22-fold, 35-fold, 23-fold, 21-fold and 20-fold, respectively.

SCGB1D2 and WFDC2 were both most differentially expressed for endometrioid cancer compared with control samples and were up-regulated by approximately 36-fold and 29-fold, respectively. WISP2 was one of the most differentially expressed genes for endometrioid cancer samples compared with control samples and was down-regulated by approximately 15-fold.

C7, ALDH1A1 and GATM are three genes originally detected as three of the most differentially expressed in clear cell, endometrioid and serous cancers when compared with control samples. However, they were only found to be most down-regulated in serous samples. Examples of genes that were within the top 20 up-regulated genes in serous cancers and were some of the most differentially expressed genes for this group were MSLN, KLK8, FOLR1, CHI3L1 and MUC16. PEG3 and PTPRN2 were down-regulated.

All 20 genes with the most significant differential expression when tested by Kruskal-Wallis, excluding CLDN16 and PPAP2A, were also differentially expressed when comparing at least

one cancer subtype with control samples. TSPAN1 was the only gene differentially expressed in all four cancer subtypes compared with the control group. ACADS was differentially expressed when clear cell, endometrioid and mucinous groups were compared with the control group. AGR2, TFF3, GAS1, USH1C and CYP2C9 were differentially expressed when comparing clear cell and mucinous individually with the control group. One observation of PEA15 and BCAM were differentially expressed when clear cell and serous were compared with the control group. EPS8 and ARL1 were both differentially expressed when endometrioid and serous groups were compared with the control group. One observation of PEA15, TTC38, KLK5, PNOC and CRABP2 were differentially expressed when comparing the serous group with the control group. Both SLC3A1 were IQGAP2 are both differentially expressed when comparing clear cell and control groups.

4. Conclusions. Hierarchical clustering of the 103 samples suggests that gene expression in normal ovaries and clear cell ovarian cancer are distinct. Implications include clear cell being more easily identifiable than other ovarian cancer subtypes based on its gene expression patterns. Furthermore, gene expression patterns can distinguish between cancerous and non-cancerous ovaries. Serous and endometrioid samples clustered together suggesting similarities in their expression levels in each gene; this study suggests they are the two most similar epithelial ovarian cancer subtypes based on gene expression.

Genes presented as both differentially expressed and dysregulated for each cancer subtype comparison with normal ovaries were detected. Based on the presence of these genes, a two step criteria was introduced to reduce the number of significant genes suggested for further study. Therefore, a gene suggested for further study had to be detected as significant in both the Wilcoxon Rank-Sum test and fold change analysis.

Considering the above discussion, the first set of genes suggested for further analysis as biomarkers was based on epithelial ovarian cancer overall, with no distinction between cancer subtypes. The genes selected met both of the following criteria. (1) The gene was one of the most significantly differentially expressed when comparing each of the four cancer subtypes with normal ovaries. (2) The gene was one of the top 40 dysregulated genes for each of the four cancer subtypes when compared with normal ovaries. The genes meeting both of these criteria were MUC1, SCNN1A, CD24 and ITM2A. Similarly to this study, ovarian cancer has previously been found to display greater expression levels of MUC1, SCNN1A and CD24 than normal ovaries [40, 65, 26, 32, 59, 64, 62, 63]. Furthermore, studies have found down-regulation of ITM2A in ovarian cancer compared with normal ovaries providing agreement with this current study [45, 17]. The differential expression and dysregulation of these genes in both this study and other studies is indicative of a significance of these genes in ovarian cancer.

Still considering the use of a two step criteria for gene selection, genes were then suggested for further analysis as biomarkers for specific epithelial ovarian cancer subtypes. For a specific subtype, a gene was suggested for further analysis if it met both of the following criteria for only the specific subtype. (1) The gene was one of the most significantly differentially expressed

when comparing the specific cancer subtype with normal ovaries. (2) The gene was one of the top 40 dysregulated genes for the specific cancer subtype when compared with normal ovaries.

Genes suggested for further analysis as biomarkers for mucinous epithelial ovarian cancer based on the two step criteria were AGR2, EFEMP1, CEACAM6, ST14, SLC44A4, S100P, ADAMTS1, TRO, PRELP, WT1, SLC4A3 and AOX1. Genes AGR2, CEACAM6, ST14, SLC44A4 and S100P all presented in significantly greater expression levels in mucinous epithelial ovarian cancer compared with normal ovaries. The converse was true for genes EFEMP1, ADAMTS1, TRO, PRELP, WT1, SLC4A3 and AOX1. AGR2, S100P and CEACAM6 have all previously displayed significance of their differential expression in mucinous epithelial ovarian cancer compared with normal ovaries [43, 50, 5] providing further feasibility of these genes being biomarkers for mucinous epithelial ovarian cancer. Down-regulation of ADAMTS1 and PRELP in ovarian cancer has been identified [68, 64], indicating tumour suppressive roles. This study recommends further analysis to determine possible prognostic associations of the selected genes in mucinous ovarian cancer.

Genes suggested for further analysis as biomarkers in endometrioid epithelial ovarian cancer based on the two step criteria were SCGB1D2, WFDC2 and WISP2. The increased expression of SCGB1D2 and WFDC2 in endometrioid ovarian cancer compared with normal ovaries in this study has also been identified in other studies [43, 14], with WFDC2 being described as a biomarker for ovarian cancer [21]. Whereas, the lower expression levels of WISP2 in endometrioid epithelial ovarian cancer observed in this study has also been demonstrated in ovarian cancer by [64]. Associations of these genes with ovarian cancer in other studies implies further study is needed to determine their role in ovarian cancer. This study suggests these genes could have a significant role in endometrioid ovarian cancer development, specifically.

Genes suggested for further analysis as biomarkers in clear cell epithelial ovarian cancer based on the two step criteria were LBP, HGD, GAS1, NR4A2, GULP1, RGS2, ANG and PELI2. LBP and HGD presented in significantly greater expression levels in clear cell ovarian cancer compared with normal ovaries. Whereas GAS1, NR4A2, GULP1, RGS2, ANG and PELI2 presented in significantly lower expression levels in clear cell ovarian cancer compared with normal ovaries. The expression level of GULP1 has been associated with ovarian cancer development and tumour suppression [39, 41]. Down-regulation of GAS1, RGS2 and PELI2 have been identified in ovarian cancer previously [24, 68]. Whereas, HGD was previously associated with up-regulation in ovarian cancer [71]. A prior study has also proposed LBP as an ovarian cancer biomarker [70]. Associations of these genes with ovarian cancer in both this study and other studies indicates their possible role in ovarian cancer. Analysis performed in this study further suggests a role in clear cell ovarian cancer that should be considered.

Genes suggested for further analysis as biomarkers for serous epithelial ovarian cancer based on the two step criteria were C7, ALDH1A1, GATM, MSLN, KLK8, FOLR1, CHI3L1, MUC16, PEG3 and PTPRN2. The genes MSLN, KLK8, FOLR1, CHI3L1 and MUC16 presented in significantly greater expression levels for serous epithelial ovarian cancer compared with normal ovaries. The converse was true for genes C7, ALDH1A1, GATM, PEG3 and PTPRN2.

ALDH1A1 has been proposed as a biomarker for serous ovarian cancer previously [51]. Furthermore, the down-regulation of ALDH1A1, C7, GATM and PEG3 in ovarian cancer has been determined [68, 56, 18]. Up-regulation of CHI3L1 was identified in serous epithelial ovarian cancer compared with normal ovaries previously [43]. The up-regulation of MUC16, FOLR1 and KLK8 has been indicated in ovarian cancer [10, 64, 30], with MUC16 being proposed as a biomarker for ovarian cancer [21]. Both this study and other studies imply an importance in the selected genes in ovarian cancer. However, based on analysis performed in this study the recommendation is to consider the importance of these genes specifically in serous ovarian cancer.

Comparing differentially expressed genes detected by the Kruskal-Wallis test and each individual Wilcoxon Rank-Sum test provided evidence that differential expression of genes was also present between ovarian cancer subtypes. Approximately 49.9%, 52.9%, 44.1% and 55.3% of Kruskal-Wallis significant genes were also significant for the Wilcoxon Rank-sum tests associated with clear cell, endometrioid, mucinous and serous ovarian cancers, respectively. Further evidence to this statement was provided by hierarchical clustering of the top 20 Kruskal-Wallis differentially expressed genes. The clusters formed using mean expression levels of these genes indicated varying expression level patterns between ovarian cancer subtypes. In particular, differing expression levels of genes in clear cell and mucinous ovarian cancer were displayed when compared with other subtypes.

Further analysis of genes suggested as biomarkers could provide ways of detecting individual subtypes of epithelial ovarian cancer, thus improving presence detection of subtypes. This will provide more efficient and effective diagnosis of epithelial ovarian cancer. Earlier diagnosis can also lead to earlier treatment and improved chance of survival. Similarities found in gene expression between epithelial ovarian cancer subtypes could also allow for development of treatments that are effective for multiple subtypes.

4.1. Limitations of Research. The small number of normal ovary samples may have affected the reliability of the Wilcoxon Rank-Sum test and Kruskal-Wallis test results; a larger sample size may be needed to increase this reliability.

The Wilcoxon Rank-Sum test and fold change analysis were performed independently of one another and then results were compared. It may have been more efficient to combine the analysis to originally find the genes that satisfied $p < 0.05$ and $FC > 2$ or $FC < 0.5$. This would have provided an easier comparison between the genes that were both dysregulated and differentially expressed in each cancer subtype in comparison to the normal ovary samples. The number of differentially expressed genes detected also led to complications. Only a subset of genes to discuss were selected based on certain criteria. This criteria may mean important biomarker candidates were not selected for further analysis or discussion.

As the Kruskal-Wallis test only suggested differential expression of genes between groups but gave no indication of which groups were different, it meant biomarkers that distinguish between specific epithelial ovarian cancer subtypes could not be determined or suggested.

4.2. Further Proposed Analysis. To build upon the hierarchical clustering of samples, with a view to establishing classification rules, heatmaps could be useful in visually identifying patterns in expression levels for both samples and genes simultaneously.

As there are many genes indicated to be differentially expressed by the Kruskal-Wallis test, the Wilcoxon Rank-Sum test could be applied to make comparisons between gene expressions of epithelial ovarian cancer subtypes. It is unclear which of the groups vary in distribution based on the Kruskal-Wallis test alone. Therefore, applying the Wilcoxon Rank-sum test to compare cancer subtypes could detect possible biomarkers to distinguish between epithelial ovarian cancer subtypes.

As meta-analysis is increasingly common in determining genes that could be biomarkers for disease, meta-analysis could also be performed to determine whether the genes suggested as biomarkers in this report are reliable choices.

It may also be of interest to analyse gene expression data for another cancer type and compare differentially expressed genes between conditions. This could allow detection of genes that may be specific biomarkers for a particular cancer or biomarkers for multiple cancers.

Collaboration with an expert in genomics would also be valuable to understand the functions of genes detected as differentially expressed. Furthermore, it could provide more information on the co-expression of genes; whether one gene affects the regulation of another gene could also provide insight into which genes should be studied based on their expressions together.

REFERENCES

[1] J. A, J. Trygg, J. Gullberg, A. I. Johansson, P. Jonsson, H. Antti, S. L. Marklund, and T. Moritz, *Extraction and gc/ms analysis of the human blood plasma metabolome*, Analytical Chemistry, 77 (2005), pp. 8086–8094, https://doi.org/10.1021/ac051211v.

[2] D. G. Beer, S. L. Kardia, C.-C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, M. L. Lizyness, R. Kuick, S. Hayasaka, J. M. Taylor, M. D. Iannettoni, M. B. Orringer, and S. Hanash, *Gene-expression profiles predict survival of patients with lung adenocarcinoma*, Nature Medicine, 8 (2002), pp. 816–824, https://doi.org/10.1038/nm733.

[3] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing*, Journal of the Royal Statistical Society. Series B (Methodological), 57 (1995), pp. 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[4] A. Berchuck, A. Kamel, R. Whitaker, B. Kerns, G. Olt, R. Kinney, J. T. Soper, R. Dodge, D. L. Clarke-Pearson, P. Marks, S. McKenzie, S. in, and R. C. Bast, *Overexpression of her-2/neu is associated with poor survival in advanced epithelial ovarian cancer*, Cancer Research, 50 (1990), p. 4087, https://pubmed.ncbi.nlm.nih.gov/1972347/.

[5] R. D. Blumenthal, E. Leon, H. J. Hansen, and D. M. Goldenberg, *Expression patterns of ceacam5 and ceacam6 in primary and metastatic cancers*, BMC Cancer, 7 (2007), p. 2, https://doi.org/10.1186/1471-2407-7-2.

[6] Cancer Research UK, *Ovarian cancer statistics.* www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer- type/ovarian-cancer, n.d. Accessed: 23-11-21.

[7] F. Cappuzzo, F. R. Hirsch, E. Rossi, S. Bartolini, G. L. Ceresoli, L. Bemis, J. Haney, S. Witta, K. Danenberg, I. Domenichini, V. Ludovini, E. Magrini, V. Gregorc, C. Doglioni,

A. Sidoni, M. Tonato, W. A. Franklin, L. Crino, J. Bunn, Paul A., and M. Varella-Garcia, *Epidermal Growth Factor Receptor Gene and Protein and Gefitinib Sensitivity in Non–Small-Cell Lung Cancer*, JNCI: Journal of the National Cancer Institute, 97 (2005), pp. 643–655, https://doi.org/10.1093/jnci/dji112.

[8] L. J. Castro-Vega, E. Letouzé, N. Burnichon, A. Buffet, P.-H. Disderot, E. Khalifa, C. Loriot, N. Elarouci, A. Morin, M. Menara, C. Lepoutre-Lussey, C. Badoual, M. Sibony, B. Dousset, R. Libé, F. Zinzindohoue, P. F. Plouin, J. Bertherat, L. Amar, A. de Reyniès, J. Favier, and A.-P. Gimenez-Roqueplo, *Multi-omics analysis defines core genomic alterations in pheochromocytomas and paragangliomas*, Nature Communications, 6 (2015), p. 6044, https://doi.org/10.1038/ncomms7044.

[9] C. Chatfield and A. J. Collins, *Introduction to multivariate analysis*, Science paperbacks, Chapman & Hall, 1st ed. ed., 1980.

[10] X. Chen, X. Li, X. Wang, Q. Zhu, X. Wu, and X. Wang, *Muc16 impacts tumor proliferation and migration through cytoplasmic translocation of p120-catenin in epithelial ovarian cancer cells: an original research*, BMC Cancer, 19 (2019), p. 171, https://doi.org/10.1186/s12885-019-5371-4.

[11] S.-H. Cho, J. Jeon, and S. I. Kim, *Personalized medicine in breast cancer: A systematic review*, J Breast Cancer, 15 (2012), pp. 265–272, https://doi.org/10.4048/jbc.2012.15.3.265.

[12] C. Denkert, J. Budczies, T. Kind, W. Weichert, P. Tablack, J. Sehouli, S. Niesporek, D. Könsgen, M. Dietel, and O. Fiehn, *Mass Spectrometry–Based Metabolic Profiling Reveals Different Metabolite Patterns in Invasive Ovarian Carcinomas and Ovarian Borderline Tumors*, Cancer Research, 66 (2006), pp. 10795–10804, https://doi.org/10.1158/0008-5472.CAN-06-0755.

[13] E. Domenici, D. R. Willé, F. Tozzi, I. Prokopenko, S. Miller, A. McKeown, C. Brittain, D. Rujescu, I. Giegling, C. W. Turck, F. Holsboer, E. T. Bullmore, L. Middleton, E. Merlo-Pich, R. C. Alexander, and P. Muglia, *Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections*, PLOS ONE, 5 (2010), pp. 1–12, https://doi.org/10.1371/journal.pone.0009166.

[14] R. Drapkin, H. H. von Horsten, Y. Lin, S. C. Mok, C. P. Crum, W. R. Welch, and J. L. Hecht, *Human Epididymis Protein 4 (HE4) Is a Secreted Glycoprotein that Is Overexpressed by Serous and Endometrioid Ovarian Carcinomas*, Cancer Research, 65 (2005), pp. 2162–2169, https://doi.org/10.1158/0008-5472.CAN-04-3924.

[15] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, *Cluster analysis and display of genome-wide expression patterns*, Proceedings of the National Academy of Sciences, 95 (1998), pp. 14863–14868, https://www.pnas.org/content/95/25/14863.

[16] B. Everitt, *Cluster analysis*, Arnold, London, 3rd ed. ed., 1993, p. 40.

[17] H. Feng, Z.-Y. Gu, Q. Li, Q.-H. Liu, X.-Y. Yang, and J.-J. Zhang, *Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis*, Journal of Ovarian Research, 12 (2019), p. 35, https://doi.org/10.1186/s13048-019-0508-2.

[18] W. Feng, R. T. Marquez, Z. Lu, J. Liu, K. H. Lu, J.-P. J. Issa, D. M. Fishman, Y. Yu, and R. C. Bast Jr, *Imprinted tumor suppressor genes arhi and peg3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation*, Cancer, 112 (2008), pp. 1489–1502, https://doi.org/10.1002/cncr.23323.

[19] J. S. Haldane, V. Hird, C. M. Hughes, and W. J. Gullick, *c-erbb-2 oncogene expression in ovarian cancer*, The Journal of Pathology, 162 (1990), pp. 231–237, https://doi.org/https://doi.org/10.1002/path.1711620309.

[20] Y. Hasin, M. Seldin, and A. Lusis, *Multi-omics approaches to disease*, Genome Biology, 18 (2017), p. 83, https://doi.org/10.1186/s13059-017-1215-1.

[21] I. Hellström, J. Raycraft, M. Hayden-Ledbetter, J. A. Ledbetter, M. Schummer, M. McIntosh, C. Drescher, N. Urban, and K. E. Hellström, *The HE4 (WFDC2) Protein Is a Biomarker for Ovarian Carcinoma1*, Cancer Research, 63 (2003), pp. 3695–3700, https://arxiv.org/abs/https://aacrjournals.org/cancerres/article-\pdf/63/13/3695/2504765/ch1303003695.pdf.

[22] K. Hibbs, K. M. Skubitz, S. E. Pambuccian, R. C. Casey, K. M. Burleson, T. R. Oegema Jr., J. J. Thiele, S. M. Grindle, R. L. Bliss, and A. P. Skubitz, *Differential gene expression in ovarian carcinoma: Identification of potential biomarkers*, The American Journal of Pathology, 165 (2004), pp. 397–414, https://doi.org/10.1016/S0002-9440(10)63306-8.

[23] C. D. Hough, C. A. Sherman-Baust, E. S. Pizer, F. J. Montz, D. D. Im, N. B. Rosenshein, K. R. Cho, G. J. Riggins, and P. J. Morin, *Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer*, Cancer Research, 60 (2000), p. 6281, https://pubmed.ncbi.nlm.nih.gov/11103784/.

[24] J. H. Hurst, N. Mendpara, and S. B. Hooks, *Regulator of g-protein signalling expression and function in ovarian cancer cell lines*, Cellular & Molecular Biology Letters, 14 (2009), pp. 153–174, https://doi.org/10.2478/s11658-008-0040-7.

[25] D. Jelovac and D. K. Armstrong, *Recent progress in the diagnosis and treatment of ovarian cancer*, CA: A Cancer Journal for Clinicians, 61 (2011), pp. 183–203, https://doi.org/https://doi.org/10.3322/caac.20113.

[26] L. W. Jiayan Lou and H. Wang, *Scnn1a overexpression correlates with poor prognosis and immune infiltrates in ovarian cancer*, International Journal of General Medicine, 15 (2022), pp. 1743–1763, https://doi.org/10.2147/IJGM.S351976.

[27] N. L. Johnston-Wilson, C. D. Sims, J.-P. Hofmann, L. Anderson, A. D. Shore, E. F. Torrey, R. H. Yolken, and the Stanley Neuropathology Consortium, *Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder*, Molecular Psychiatry, 5 (2000), pp. 142–149, https://doi.org/10.1038/sj.mp.4000696.

[28] J. Kelly, R. Moyeed, C. Carroll, D. Albani, and X. Li, *Gene expression meta-analysis of parkinson's disease and its relationship with alzheimer's disease*, Molecular Brain, 12 (2019), p. 16, https://doi.org/10.1186/s13041-019-0436-5.

[29] M. Kim and I. Tagkopoulos, *Data integration and predictive modeling methods for multi-omics datasets*, Mol. Omics, 14 (2018), pp. 8–25, https://doi.org/10.1039/C7MO00051K.

[30] T. Kishi, L. Grass, A. Soosaipillai, A. Scorilas, N. Harbeck, B. Schmalfeldt, J. Dorn, M. Mysliwiec, M. Schmitt, and E. P. Diamandis, *Human Kallikrein 8, a Novel Biomarker for Ovarian Carcinoma1*, Cancer Research, 63 (2003), pp. 2771–2774, https://arxiv.org/abs/https://aacrjournals.org/cancerres/article-pdf/63/11/2771/2503673/ch1103002771.pdf.

[31] K. B. Kuchenbaecker, S. J. Ramus, J. Tyrer, A. Lee, H. C. Shen, J. Beesley, K. Lawrenson, L. McGuffog, S. Healey, J. M. Lee, T. J. Spindler, Y. G. Lin, T. Pejovic, Y. Bean, Q. Li, S. Coetzee, D. Hazelett, A. Miron, M. Southey, M. B. Terry, D. E. Goldgar, S. S. Buys, R. Janavicius, C. M. Dorfling, E. J. van Rensburg, S. L. Neuhausen, Y. C. Ding, T. V. O. Hansen, L. Jønson, A.-M. Gerdes, B. Ejlertsen, D. Barrowdale, J. Dennis, J. Benitez, A. Osorio, M. J. Garcia, I. Komenaka, J. N. Weitzel, P. Ganschow, P. Peterlongo, L. Bernard, A. Viel, B. Bonanni, B. Peissel, S. Manoukian, P. Radice, L. Papi, L. Ottini, F. Fostira, I. Konstantopoulou, J. Garber, D. Frost, J. Perkins, R. Platte, S. Ellis, A. K. Godwin, R. K. Schmutzler, A. Meindl, C. Engel, C. Sutter, O. M. Sinilnikova, F. Damiola, S. Mazoyer, D. Stoppa-Lyonnet, K. Claes, K. De Leeneer, J. Kirk, G. C. Rodriguez, M. Piedmonte, D. M. O'Malley, M. de la Hoya, T. Caldes, K. Aittomäki, H. Nevanlinna, J. M. Collée, M. A. Rookus, J. C. Oosterwijk, L. Tihomirova, N. Tung, U. Hamann, C. Isaccs, M. Tischkowitz, E. N. Imyanitov, M. A. Caligo, I. G. Campbell, F. B. L. Hogervorst, E. Olah, O. Diez, I. Blanco, J. Brunet, C. Lazaro, M. A. Pujana, A. Jakubowska, J. Gronwald, J. Lubinski, G. Sukiennicki, R. B. Barkardottir, M. Plante, J. Simard, P. Soucy, M. Montagna, S. Tognazzo, M. R. Teixeira, V. S. Pankratz, X. Wang, N. Lindor, C. I. Szabo, N. Kauff, J. Vijai, C. A. Aghajanian, G. Pfeiler, A. Berger, C. F. Singer, M.-K. Tea, C. M. Phelan, M. H. Greene, P. L. Mai, G. Rennert, A. M. Mulligan, S. Tchatchou, I. L. Andrulis, G. Glendon, A. E. Toland, U. B. Jensen, T. A. Kruse, M. Thomassen, A. Bojesen, J. Zidan, E. Friedman, Y. Laitman, M. Soller, A. Liljegren, B. Arver, Z. Einbeigi, M. Stenmark-Askmalm, O. I. Olopade, R. L. Nussbaum, T. R. Rebbeck, K. L. Nathanson, S. M. Domchek, K. H. Lu, B. Y. Karlan, C. Walsh, J. Lester, A. Hein, A. B. Ekici, M. W. Beckmann, P. A. Fasching, D. Lambrechts, E. Van Nieuwenhuysen, I. Vergote, S. Lambrechts, E. Dicks, J. A. Doherty, K. G. Wicklund, M. A. Rossing, A. Rudolph, J. Chang-Claude, S. Wang-Gohrke, U. Eilber, K. B. Moysich, K. Odunsi, L. Sucheston, S. Lele, L. R. Wilkens, M. T. Goodman, P. J. Thompson, Y. B. Shvetsov, I. B. Runnebaum, M. Dürst, P. Hillemanns, T. Dörk, N. Antonenkova, N. Bogdanova, A. Leminen, L. M. Pelttari, R. Butzow, F. Modugno, J. L. Kelley, R. P. Edwards, R. B.

Ness, A. du Bois, F. Heitz, I. Schwaab, P. Harter, K. Matsuo, S. Hosono, S. Orsulic, A. Jensen, S. K. Kjaer, E. Hogdall, H. N. Hasmad, M. A. N. Azmi, S.-H. Teo, Y.-L. Woo, B. L. Fridley, E. L. Goode, J. M. Cunningham, R. A. Vierkant, F. Bruinsma, G. G. Giles, D. Liang, M. A. T. Hildebrandt, X. Wu, D. A. Levine, M. Bisogna, A. Berchuck, E. S. Iversen, J. M. Schildkraut, P. Concannon, R. P. Weber, D. W. Cramer, K. L. Terry, E. M. Poole, S. S. Tworoger, E. V. Bandera, I. Orlow, S. H. Olson, C. Krakstad, H. B. Salvesen, I. L. Tangen, L. Bjorge, A. M. van Altena, K. K. H. Aben, L. A. Kiemeney, L. F. A. G. Massuger, M. Kellar, A. Brooks-Wilson, L. E. Kelemen, L. S. Cook, N. D. Le, C. Cybulski, H. Yang, J. Lissowska, L. A. Brinton, N. Wentzensen, C. Hogdall, L. Lundvall, L. Nedergaard, H. Baker, H. Song, D. Eccles, I. McNeish, J. Paul, K. Carty, N. Siddiqui, R. Glasspool, A. S. Whittemore, J. H. Rothstein, V. McGuire, W. Sieh, B.-T. Ji, W. Zheng, X.-O. Shu, Y.-T. Gao, B. Rosen, H. A. Risch, J. R. McLaughlin, S. A. Narod, A. N. Monteiro, A. Chen, H.-Y. Lin, J. Permuth-Wey, T. A. Sellers, Y.-Y. Tsai, Z. Chen, A. Ziogas, H. Anton-Culver, A. Gentry-Maharaj, U. Menon, P. Harrington, A. W. Lee, A. H. Wu, C. L. Pearce, G. Coetzee, M. C. Pike, A. Dansonka-Mieszkowska, A. Timorek, I. K. Rzepecka, J. Kupryjanczyk, M. Freedman, H. Noushmehr, D. F. Easton, K. Offit, F. J. Couch, S. Gayther, P. P. Pharoah, A. C. Antoniou, G. Chenevix-Trench, EMBRACE, G. S. Collaborators, B. C. F. Registry, HEBON, K. Investigators, A. C. S. O. C. Investigators), A. O. C. S. Group, t. C. o. I. o. M. of BRCA1, and BRCA2, *Identification of six new susceptibility loci for invasive epithelial ovarian cancer*, Nature Genetics, 47 (2015), pp. 164–171, https://doi.org/10.1038/ng.3185.

[32] J.-H. Lee, S.-H. Kim, E.-S. Lee, and Y.-S. Kim, *Cd24 overexpression in cancer development and progression: A meta-analysis*, Oncol Rep, 22 (2009), pp. 1149–1156, https://doi.org/10.3892/or_00000548.

[33] J. K. Lee, *Statistical Bioinformatics: For Biomedical and Life Science Researchers*, John Wiley & Sons, Incorporated, 2014.

[34] S. Lheureux, M. Braunstein, and A. M. Oza, *Epithelial ovarian cancer: evolution of management in the era of precision medicine*, CA: a cancer journal for clinicians, 69 (2019), pp. 280–304, https://doi.org/10.3322/caac.21559.

[35] W. Li, Z. Liu, B. Liang, S. Chen, X. Zhang, X. Tong, W. Lou, L. Le, X. Tang, and F. Fu, *Identification of core genes in ovarian cancer by an integrative meta-analysis*, Journal of Ovarian Research, 11 (2018), p. 94, https://doi.org/10.1186/s13048-018-0467-z.

[36] Z. Li and C. Chan, *Integrating gene expression and metabolic profiles**, Journal of Biological Chemistry, 279 (2004), pp. 27124–27137, https://doi.org/10.1074/jbc.M403494200.

[37] J. Lloyd-Price, C. Arze, A. N. Ananthakrishnan, M. Schirmer, J. Avila-Pacheco, T. W. Poon, E. Andrews, N. J. Ajami, K. S. Bonham, C. J. Brislawn, D. Casero, H. Courtney, A. Gonzalez, T. G. Graeber, A. B. Hall, K. Lake, C. J. Landers, H. Mallick, D. R. Plichta, M. Prasad, G. Rahnavard, J. Sauk, D. Shungin, Y. Vázquez-Baeza, R. A. White, J. Bishai, K. Bullock, A. Deik, C. Dennis, J. L. Kaplan, H. Khalili, L. J. McIver, C. J. Moran, L. Nguyen, K. A. Pierce, R. Schwager, A. Sirota-Madi, B. W. Stevens, W. Tan, J. J. ten Hoeve, G. Weingart, R. G. Wilson, V. Yajnik, J. Braun, L. A. Denson, J. K. Jansson, R. Knight, S. Kugathasan, D. P. B. McGovern, J. F. Petrosino, T. S. Stappenbeck, H. S. Winter, C. B. Clish, E. A. Franzosa, H. Vlamakis, R. J. Xavier, C. Huttenhower, and I. B. D. M. D. B. Investigators, *Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases*, Nature, 569 (2019), pp. 655–662, https://doi.org/10.1038/s41586-019-1237-9.

[38] M. A. lzquierdo, A. G. J. d. Zee, J. B. Vermorken, P. v. d. Valk, J. A. M. Beliën, G. Giaccone, G. L. Scheffer, M. J. Flens, H. M. Pinedo, P. Kenemans, C. J. L. M. Meijer, E. G. E. Vries, and R. J. Scheper, *Drug Resistance-Associated Marker Lrp for Prediction of Response to Chemotherapy and Prognoses in Advanced Ovarian Carcinoma*, JNCI: Journal of the National Cancer Institute, 87 (1995), pp. 1230–1237, https://doi.org/10.1093/jnci/87.16.1230.

[39] C.-I. J. Ma, C. Martin, Z. Ma, A. Hafiane, M. Dai, J.-J. Lebrun, and R. S. Kiss, *Engulfment protein gulp is regulator of transforming growth factor-β response in ovarian cells*, Journal of Biological Chemistry, 287 (2012), pp. 20636–20651, https://doi.org/10.1074/jbc.M111.314997.

[40] Q. Ma, J. Song, S. Wang, and N. He, *Muc1 regulates akt signaling pathway by upregulating egfr*

expression in ovarian cancer cells, Pathology - Research and Practice, 224 (2021), p. 153509, https://doi.org/https://doi.org/10.1016/j.prp.2021.153509.

[41] L. Maldonado, M. Brait, E. Izumchenko, S. Begum, A. Chatterjee, T. Sen, M. Loyo, A. Barbosa, M. L. Poeta, E. Makarev, A. Zhavoronkov, V. M. Fazio, R. Angioli, C. Rabitti, M. Ongenaert, W. Van Criekinge, M. G. Noordhuis, P. de Graeff, G. B. A. Wisman, A. G. van der Zee, and M. O. Hoque, *Integrated transcriptomic and epigenomic analysis of ovarian cancer reveals epigenetically silenced gulp1*, Cancer Letters, 433 (2018), pp. 242–251, https://doi.org/https://doi.org/10.1016/j.canlet.2018.06.030.

[42] J. R. Marks, A. M. Davidoff, B. J. Kerns, P. A. Humphrey, J. C. Pence, R. K. Dodge, D. L. Clarke-Pearson, J. D. Iglehart, R. C. Bast, and A. Berchuck, *Overexpression and mutation of p53 in epithelial ovarian cancer*, Cancer Research, 51 (1991), p. 2979, https://pubmed.ncbi.nlm.nih.gov/2032235/.

[43] R. T. Marquez, K. A. Baggerly, A. P. Patterson, J. Liu, R. Broaddus, M. Frumovitz, E. N. Atkinson, D. I. Smith, L. Hartmann, D. Fishman, A. Berchuck, R. Whitaker, D. M. Gershenson, G. B. Mills, J. Bast, Robert C., and K. H. Lu, *Patterns of Gene Expression in Different Histotypes of Epithelial Ovarian Cancer Correlate with Those in Normal Fallopian Tube, Endometrium, and Colon*, Clinical Cancer Research, 11 (2005), pp. 6116–6126, https://doi.org/10.1158/1078-0432.CCR-04-2509.

[44] B. Meder, J. Haas, F. Sedaghat-Hamedani, E. Kayvanpour, K. Frese, A. Lai, R. Nietsch, C. Scheiner, S. Mester, D. M. Bordalo, A. Amr, C. Dietrich, D. Pils, D. Siede, H. Hund, A. Bauer, D. B. Holzer, A. Ruhparwar, M. Mueller-Hennessen, D. Weichenhan, C. Plass, T. Weis, J. Backs, M. Wuerstle, A. Keller, H. A. Katus, and A. E. Posch, *Epigenome-wide association study identifies cardiac gene patterning and a novel class of biomarkers for heart failure*, Circulation, 136 (2017), pp. 1528–1544, https://doi.org/10.1161/CIRCULATIONAHA.117.027355.

[45] T. M. H. Nguyen, I.-W. Shin, T. J. Lee, J. Park, J. H. Kim, M. S. Park, and E.-J. Lee, *Loss of itm2a, a novel tumor suppressor of ovarian cancer through g2/m cell cycle arrest, is a poor prognostic factor of epithelial ovarian cancer*, Gynecologic Oncology, 140 (2016), pp. 545–553, https://doi.org/10.1016/j.ygyno.2015.12.006.

[46] L. Ning, B. Long, W. Zhang, M. Yu, S. Wang, D. Cao, J. Yang, K. Shen, Y. Huang, and J. Lang, *Circular rna profiling reveals circexoc6b and circn4bp2l2 as novel prognostic biomarkers in epithelial ovarian cancer*, Int J Oncol, 53 (2018), pp. 2637–2646, https://doi.org/10.3892/ijo.2018.4566.

[47] B. Norquist, K. A. Wurz, C. C. Pennil, R. Garcia, J. Gross, W. Sakai, B. Y. Karlan, T. Taniguchi, and E. M. Swisher, *Secondary somatic mutations restoring brca1/2 predict chemotherapy resistance in hereditary ovarian carcinomas*, Journal of Clinical Oncology, 29 (2011), pp. 3008–3015, https://doi.org/10.1200/JCO.2010.34.2980.

[48] K. Ono, T. Tanaka, T. Tsunoda, O. Kitahara, C. Kihara, A. Okamoto, K. Ochiai, T. Takagi, and Y. Nakamura, *Identification by cDNA Microarray of Genes Involved in Ovarian Carcinogenesis1*, Cancer Research, 60 (2000), pp. 5007–5011, https://aacrjournals.org/cancerres/article/60/18/5007/506699/Identification-by-cDNA-Microarray-of-Genes.

[49] S. Paik, G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. P. Costantino, C. E. Geyer, D. L. Wickerham, and N. Wolmark, *Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor–positive breast cancer*, Journal of Clinical Oncology, 24 (2006), pp. 3726–3734, https://doi.org/10.1200/JCO.2005.04.7985.

[50] K. Park, Y. J. Chung, H. So, K. Kim, J. Park, M. Oh, M. Jo, K. Choi, E.-J. Lee, Y.-L. Choi, S. Y. Song, D.-S. Bae, B.-G. Kim, and J.-H. Lee, *Agr2, a mucinous ovarian cancer marker, promotes cell proliferation and migration*, Experimental & molecular medicine, 43 (2011), pp. 91–100, https://doi.org/10.3858/emm.2011.43.2.011.

[51] K. Penumatsa, S. L. Edassery, A. Barua, M. J. Bradaric, and J. L. Luborsky, *Differential expression of aldehyde dehydrogenase 1a1 (aldh1) in normal ovary and serous ovarian tumors*, Journal of Ovarian Research, 3 (2010), p. 28, https://doi.org/10.1186/1757-2215-3-28.

[52] P. D. P. Pharoah, Y.-Y. Tsai, S. J. Ramus, C. M. Phelan, E. L. Goode, K. Lawrenson, M. Buckley, B. L. Fridley, J. P. Tyrer, H. Shen, R. Weber, R. Karevan, M. C. Larson, H. Song, D. C. Tessier, F. Bacot, D. Vincent, J. M. Cunningham, J. Dennis, E. Dicks,

682 K. K. Aben, H. Anton-Culver, N. Antonenkova, S. M. Armasu, L. Baglietto, E. V. Ban-
683 dera, M. W. Beckmann, M. J. Birrer, G. Bloom, N. Bogdanova, J. D. Brenton, L. A.
684 Brinton, A. Brooks-Wilson, R. Brown, R. Butzow, I. Campbell, M. E. Carney, R. S.
685 Carvalho, J. Chang-Claude, Y. A. Chen, Z. Chen, W.-H. Chow, M. S. Cicek, G. Coet-
686 zee, L. S. Cook, D. W. Cramer, C. Cybulski, A. Dansonka-Mieszkowska, E. Despierre,
687 J. A. Doherty, T. Dörk, A. du Bois, M. Dürst, D. Eccles, R. Edwards, A. B. Ekici, P. A.
688 Fasching, D. Fenstermacher, J. Flanagan, Y.-T. Gao, M. Garcia-Closas, A. Gentry-
689 Maharaj, G. Giles, A. Gjyshi, M. Gore, J. Gronwald, Q. Guo, M. K. Halle, P. Harter,
690 A. Hein, F. Heitz, P. Hillemanns, M. Hoatlin, E. Høgdall, C. K. Høgdall, S. Hosono,
691 A. Jakubowska, A. Jensen, K. R. Kalli, B. Y. Karlan, L. E. Kelemen, L. A. Kiemeney,
692 S. K. Kjaer, G. E. Konecny, C. Krakstad, J. Kupryjanczyk, D. Lambrechts, S. Lam-
693 brechts, N. D. Le, N. Lee, J. Lee, A. Leminen, B. K. Lim, J. Lissowska, J. Lubiński, L. Lund-
694 vall, L. F. A. G. Lurie, Galinaand Massuger, K. Matsuo, V. McGuire, J. R. McLaughlin,
695 U. Menon, F. Modugno, K. B. Moysich, T. Nakanishi, S. A. Narod, R. B. Ness, H. Nevan-
696 linna, S. Nickels, H. Noushmehr, K. Odunsi, S. Olson, I. Orlow, J. Paul, T. Pejovic, L. M.
697 Pelttari, J. Permuth-Wey, M. C. Pike, E. M. Poole, X. Qu, H. A. Risch, L. Rodriguez-
698 Rodriguez, M. A. Rossing, A. Rudolph, I. Runnebaum, I. K. Rzepecka, H. B. Salvesen,
699 I. Schwaab, G. Severi, H. Shen, V. Shridhar, X.-O. Shu, W. Sieh, M. C. Southey, P. Spell-
700 man, K. Tajima, S.-H. Teo, K. L. Terry, P. J. Thompson, A. Timorek, S. S. Tworoger,
701 A. M. van Altena, D. van den Berg, I. Vergote, R. A. Vierkant, A. F. Vitonis, S. Wang-
702 Gohrke, N. Wentzensen, A. S. Whittemore, E. Wik, B. Winterhoff, Y. L. Woo, A. H.
703 Wu, H. P. Yang, W. Zheng, A. Ziogas, F. Zulkifli, M. T. Goodman, P. Hall, D. F. Eas-
704 ton, C. L. Pearce, A. Berchuck, G. Chenevix-Trench, E. Iversen, A. N. A. Monteiro,
705 S. A. Gayther, J. M. Schildkraut, T. A. Sellers, A. C. Study, and A. O. C. S. Group,
706 *Gwas meta-analysis and replication identifies three new susceptibility loci for ovarian cancer*, Nature
707 Genetics, 45 (2013), pp. 362–370, https://doi.org/10.1038/ng.2564.
708 [53] C. M. Phelan, K. B. Kuchenbaecker, J. P. Tyrer, S. P. Kar, K. Lawrenson, S. J. Winham,
709 J. Dennis, A. Pirie, M. J. Riggan, G. Chornokur, M. A. Earp, P. C. Lyra, J. M. Lee, S. Co-
710 etzee, J. Beesley, L. McGuffog, P. Soucy, E. Dicks, A. Lee, D. Barrowdale, J. Lecar-
711 pentier, G. Leslie, C. M. Aalfs, K. K. H. Aben, M. Adams, J. Adlard, I. L. Andrulis,
712 H. Anton-Culver, N. Antonenkova, G. Aravantinos, N. Arnold, B. K. Arun, B. Arver,
713 J. Azzollini, J. Balmaña, S. N. Banerjee, L. Barjhoux, R. B. Barkardottir, Y. Bean,
714 M. W. Beckmann, A. Beeghly-Fadiel, J. Benitez, M. Bermisheva, M. Q. Bernardini, M. J.
715 Birrer, L. Bjorge, A. Black, K. Blankstein, M. J. Blok, C. Bodelon, N. Bogdanova,
716 A. Bojesen, B. Bonanni, Å. Borg, A. R. Bradbury, J. D. Brenton, C. Brewer, L. Brin-
717 ton, P. Broberg, A. Brooks-Wilson, F. Bruinsma, J. Brunet, B. Buecher, R. Butzow,
718 S. S. Buys, T. Caldes, M. A. Caligo, I. Campbell, R. Cannioto, M. E. Carney, T. Cescon,
719 S. B. Chan, J. Chang-Claude, S. Chanock, X. Q. Chen, Y.-E. Chiew, J. Chiquette, W. K.
720 Chung, K. B. M. Claes, T. Conner, L. S. Cook, J. Cook, D. W. Cramer, J. M. Cunningham,
721 A. A. D'Aloisio, M. B. Daly, F. Damiola, S. D. Damirovna, A. Dansonka-Mieszkowska,
722 F. Dao, R. Davidson, A. DeFazio, C. Delnatte, K. F. Doheny, O. Diez, Y. C. Ding, J. A.
723 Doherty, S. M. Domchek, C. M. Dorfling, T. Dörk, L. Dossus, M. Duran, M. Dürst,
724 B. Dworniczak, D. Eccles, T. Edwards, R. Eeles, U. Eilber, B. Ejlertsen, A. B. Ekici,
725 S. Ellis, M. Elvira, K. H. Eng, C. Engel, D. G. Evans, P. A. Fasching, S. Ferguson,
726 S. F. Ferrer, J. M. Flanagan, Z. C. Fogarty, R. T. Fortner, F. Fostira, W. D. Foulkes,
727 G. Fountzilas, B. L. Fridley, T. M. Friebel, E. Friedman, D. Frost, P. A. Ganz, J. Gar-
728 ber, M. J. García, V. Garcia-Barberan, A. Gehrig, A. Gentry-Maharaj, A.-M. Gerdes,
729 G. G. Giles, R. Glasspool, G. Glendon, A. K. Godwin, D. E. Goldgar, T. Goranova,
730 M. Gore, M. H. Greene, J. Gronwald, S. Gruber, E. Hahnen, C. A. Haiman, N. Håkansson,
731 U. Hamann, T. V. O. Hansen, P. A. Harrington, H. R. Harris, J. Hauke, A. Hein, A. Hen-
732 derson, M. A. T. Hildebrandt, P. Hillemanns, S. Hodgson, C. K. Høgdall, E. Høgdall,
733 F. B. L. Hogervorst, H. Holland, M. J. Hooning, K. Hosking, R.-Y. Huang, P. J. Hulick,
734 J. Hung, D. J. Hunter, D. G. Huntsman, T. Huzarski, E. N. Imyanitov, C. Isaacs, E. S.
735 Iversen, L. Izatt, A. Izquierdo, A. Jakubowska, P. James, R. Janavicius, M. Jernetz,

A. Jensen, U. B. Jensen, E. M. John, S. Johnatty, M. E. Jones, P. Kannisto, B. Y. Karlan, A. Karnezis, K. Kast, C. J. Kennedy, E. Khusnutdinova, L. A. Kiemeney, J. I. Kiiski, S.-W. Kim, S. K. Kjaer, M. Köbel, R. K. Kopperud, T. A. Kruse, J. Kupryjanczyk, A. Kwong, Y. Laitman, D. Lambrechts, N. Larrañaga, M. C. Larson, C. Lazaro, N. D. Le, L. Le Marchand, J. W. Lee, S. B. Lele, A. Leminen, D. Leroux, J. Lester, F. Lesueur, D. A. Levine, D. Liang, C. Liebrich, J. Lilyquist, L. Lipworth, J. Lissowska, K. H. Lu, J. Lubiński, C. Luccarini, L. Lundvall, P. L. Mai, G. Mendoza-Fandiño, S. Manoukian, L. F. A. G. Massuger, T. May, S. Mazoyer, J. N. McAlpine, V. McGuire, J. R. McLaughlin, I. McNeish, H. Meijers-Heijboer, A. Meindl, U. Menon, A. R. Mensenkamp, M. A. Merritt, R. L. Milne, G. Mitchell, F. Modugno, J. Moes-Sosnowska, M. Moffitt, M. Montagna, K. B. Moysich, A. M. Mulligan, J. Musinsky, K. L. Nathanson, L. Nedergaard, R. B. Ness, S. L. Neuhausen, H. Nevanlinna, D. Niederacher, R. L. Nussbaum, K. Odunsi, E. Olah, O. I. Olopade, H. Olsson, C. Olswold, D. M. O'Malley, K.-r. Ong, N. C. Onland-Moret, N. Orr, S. Orsulic, A. Osorio, D. Palli, L. Papi, T.-W. Park-Simon, J. Paul, C. L. Pearce, I. S. Pedersen, P. H. M. Peeters, B. Peissel, A. Peixoto, T. Pejovic, L. M. Peltari, J. B. Permuth, P. Peterlongo, L. Pezzani, G. Pfeiler, K.-A. Phillips, M. Piedmonte, M. C. Pike, A. M. Piskorz, S. R. Poblete, T. Pocza, E. M. Poole, B. Poppe, M. E. Porteous, F. Prieur, D. Prokofyeva, E. Pugh, M. A. Pujana, P. Pujol, P. Radice, A. s. group, E. M. B. R. A. C. E. Study, G. S. Collaborators, H. E. B. O. N. Study, K. Investigators, and O. s. group, *Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer*, Nature Genetics, 49 (2017), pp. 680–691, https://doi.org/10.1038/ng.3826.

[54] S. M. Ross, *Introductory statistics*, Academic Press, Jan 2017, pp. 538–540.

[55] S. M. Ross, *Introductory statistics*, Academic Press, Jan 2017, pp. 550–553.

[56] A. D. Santin, F. Zhan, S. Bellone, M. Palmieri, S. Cane, E. Bignotti, S. Anfossi, M. Gokden, D. Dunn, J. J. Roman, T. J. O'Brien, E. Tian, M. J. Cannon, J. Shaughnessy Jr., and S. Pecorelli, *Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: Identification of candidate molecular markers for ovarian cancer diagnosis and therapy*, International Journal of Cancer, 112 (2004), pp. 14–25, https://doi.org/https://doi.org/10.1002/ijc.20408.

[57] N. L. Sieben, J. Oosting, A. M. Flanagan, J. Prat, G. M. Roemen, S. M. Kolkman-Uljee, R. van Eijk, C. J. Cornelisse, G. J. Fleuren, and M. van Engeland, *Differential gene expression in ovarian tumors reveals dusp 4 and serpina 5 as key regulators for benign behavior of serous borderline tumors*, Journal of Clinical Oncology, 23 (2005), pp. 7257–7264, https://doi.org/10.1200/JCO.2005.02.2541.

[58] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, *Cancer statistics, 2022.*, CA: a cancer journal for clinicians, 72 (2022), https://doi.org/10.3322/caac.21708.

[59] B. Soltész, J. Lukács, E. Szilágyi, Éva Márton, M. Szilágyi Bónizs, A. Penyige, R. Póka, and B. Nagy, *Expression of cd24 in plasma, exosome and ovarian tissue samples of serous ovarian cancer patients*, Journal of Biotechnology, 298 (2019), pp. 16–20, https://doi.org/https://doi.org/10.1016/j.jbiotec.2019.03.018.

[60] V. G. Tusher, R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response*, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001), pp. 5116–5121, https://doi.org/10.1073/pnas.091062498.

[61] A. J. Vargas and C. C. Harris, *Biomarker development in the precision medicine era: lung cancer as a case study*, Nature Reviews Cancer, 16 (2016), pp. 525–537, https://doi.org/10.1038/nrc.2016.56.

[62] L. Wang, J. Ma, F. Liu, Q. Yu, G. Chu, A. C. Perkins, and Y. Li, *Expression of muc1 in primary and metastatic human epithelial ovarian cancer and its therapeutic significance*, Gynecologic Oncology, 105 (2007), pp. 695–702, https://doi.org/10.1016/j.ygyno.2007.02.004.

[63] J. B. Welsh, P. P. Zarrinkar, L. M. Sapinoso, S. G. Kern, C. A. Behling, B. J. Monk, D. J. Lockhart, R. A. Burger, and G. M. Hampton, *Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer*, Proceedings of the National Academy of Sciences, 98 (2001), pp. 1176–1181, https://doi.org/10.1073/pnas.98.3.1176.

[64] Y.-L. Wong, A. Z. H. M. Dali, I. Mohamed Rose, R. Jamal, and N. M. Mokhtar, *Potential*

molecular signatures in epithelial ovarian cancer by genome wide expression profiling, Asia-Pacific Journal of Clinical Oncology, 12 (2016), pp. e259–e268, https://doi.org/https://doi.org/10.1111/ajco.12182.

[65] L. Wu, Z.-H. Ling, H. Wang, X.-Y. Wang, and J. Gui, *Upregulation of scnn1a promotes cell proliferation, migration, and predicts poor prognosis in ovarian cancer through regulating epithelial–mesenchymal transformation*, Cancer Biotherapy and Radiopharmaceuticals, 34 (2019), pp. 642–649, https://doi.org/10.1089/cbr.2019.2824. PMID: 31549859.

[66] R. Wu, Y. Zhai, R. Kuick, P. Garcia, A. Naseeem, T. C. Hu, E. R. Fearon, K. R. Cho, and A. N. Karnezis, *Oviductal vs. ovarian epithelial transformation yields very different tumor phenotypes in apc fl/fl;pten fl/fl mice*. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80085, 2016. Accessed: 23-11-21.

[67] L. Xie, X. Ge, H. Tan, L. Xie, Y. Zhang, T. Hart, X. Yang, and P. E. Bourne, *Towards structural systems pharmacology to study complex diseases and personalized medicine*, PLOS Computational Biology, 10 (2014), pp. 1–15, https://doi.org/10.1371/journal.pcbi.1003554.

[68] Y. Xu, G. Wu, J. Li, J. Li, N. Ruan, L. Ma, X. Han, Y. Wei, L. Li, H. Zhang, Y. Chen, and Q. Xia, *Screening and identification of key biomarkers for bladder cancer: A study based on tcga and geo data*, BioMed Research International, 2020 (2020), p. 8283401, https://doi.org/https://doi.org/10.1155/2020/8283401.

[69] S. Yachida, S. Mizutani, H. Shiroma, S. Shiba, T. Nakajima, T. Sakamoto, H. Watanabe, K. Masuda, Y. Nishimoto, M. Kubo, F. Hosoda, H. Rokutan, M. Matsumoto, H. Takamaru, M. Yamada, T. Matsuda, M. Iwasaki, T. Yamaji, T. Yachida, T. Soga, K. Kurokawa, A. Toyoda, Y. Ogura, T. Hayashi, M. Hatakeyama, H. Nakagama, Y. Saito, S. Fukuda, T. Shibata, and T. Yamada, *Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer*, Nature Medicine, 25 (2019), pp. 968–976, https://doi.org/10.1038/s41591-019-0458-7.

[70] W. Zhang, X. Ou, and X. Wu, *Proteomics profiling of plasma exosomes in epithelial ovarian cancer: A potential role in the coagulation cascade, diagnosis and prognosis*, Int J Oncol, 54 (2019), pp. 1719–1733, https://doi.org/10.3892/ijo.2019.4742.

[71] K. K. Zorn, T. Bonome, L. Gangi, G. V. Chandramouli, C. S. Awtrey, G. J. Gardner, J. C. Barrett, J. Boyd, and M. J. Birrer, *Gene Expression Profiles of Serous, Endometrioid, and Clear Cell Subtypes of Ovarian and Endometrial Cancer*, Clinical Cancer Research, 11 (2005), pp. 6422–6430, https://doi.org/10.1158/1078-0432.CCR-05-0508.