# Differential Gene Expression Study of Oral Squamous Cell Carcinoma

**Oscar Vanhanen (34640481)**

## 1. INTRODUCTION

Oral Squamous Cell Carcinoma (OSCC) is a malignant tumour in the oral mucosa's stratified squamous epithelium, commonly located on the lip, tongue and oral cavity floor (Rivera & Venegas, 2014). OSSC accounts for 90% of all oral cancers and affects 400,000 people annually. The primary risk factors are smoking and alcohol consumption (Oral Cancer, 2022). Differentially expressed genes are known to be contributing factors for cancer initiation and propagation. RNA profiles can be used to determine the underlying transcriptional signatures of specific phenotypes such as a Tumour (Chung et al., 2021). Other research has found 21 genes, involved in antioxidant metabolism and related to carcinogenesis pathways, to be differentially expressed in OSCC tumour samples (Pedro et al., 2018).

Here we use the RNA sequencing data obtained from different patients’ tumours and normal samples to find out which genes are differentially expressed (DE). From here we can determine the pathways affected by these up or downregulated genes to better understand the mechanisms of OCCS proliferation and ultimately determine a panel of genes for clinical decision-making and identify targets for biomarker or treatment research.

## 2. METHODS

Matched normal tissue vs. tumour tissue from 3 oral carcinoma patients were obtained and sequenced to determine the expression profiles.

DE analysis of the data was performed within R using multiple libraries to clean, annotate and visualise the read counts. The libraries and databases used in the R script include:

* **BiocManager** **version 3.15**: Used to install packages for analysis and comprehension of genomic data. Used to install the following packages and databases.
* **EdgeR:** Package for differential expression analysis of RNA-seq expression profiles
* **Org.Hs.eg.db**: Genome-wide annotation database for Human genes, primarily based on mapping using Entrez Gene identifiers.
* **GO.db:** Annotation maps describing the entire Gene Ontology. Data is obtained from GO, a knowledge base containing information on gene functions.

See Appendix 1 for the whole annotated R script (HTML file) and associated outputs.

## 3. RESULTS

### 3a. Data processing and Annotation

There were 5022 duplicated genes (based on Symbol) removed from the gene counts table (Table 1), leaving a total of total of 10,510 genes that the sequences aligned to. The value of reads per sample is found in the library size column of Table 1.

**Table 1.** Normal and Tumour patient samples and the respective library size (read counts).

|  |  |  |  |
| --- | --- | --- | --- |
| Patient sample | Group | Library Size | Norm factors |
| 8N | 1 | 12090121 | 1 |
| 8T | 1 | 10123913 | 1 |
| 33N | 1 | 19890767 | 1 |
| 33T | 1 | 18590376 | 1 |
| 51N | 1 | 33878462 | 1 |
| 51T | 1 | 21832978 | 1 |

### 3b. Distances between gene expression profiles

Multidimensional scaling of the 3 patient samples found there to be a significant separation between Normal samples and Tumour samples on the X-axis, as seen in Figure 1. This indicates the Tissue samples generally have different expression profiles compared to Normal. The plot also shows there to be more variability between Tumour samples than in the Normal samples, which are clustered more tightly together indicating close similarity in their expression profiles.

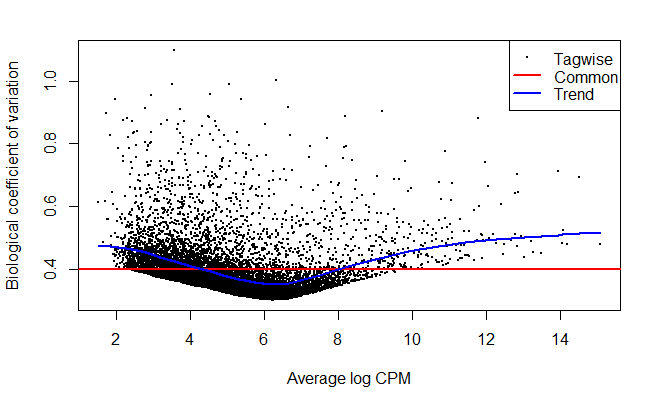
Chart, scatter chart

Description automatically generated

**Figure 1.** Multidimensional scaling plot of distances between gene expression profiles of Tumour/Normal patient samples 8, 33 and 51.

### 3c. Biological variation Coefficient

The coefficient of biological variation was determined to be 0.4 (Figure 2), being within the threshold of 0.2-0.4 indicating there is an acceptable low amount of variation between Tumour and Normal samples (*Some Key Factors for Number of Significant DE Genes - CVR Bioinformatics*, n.d.). Therefore, the amount of Differentially Expressed genes found will not be decreased.

**Figure 2.** Biological Coefficient Variation plot. The common coefficient of biological variation is 0.4.

### 3d. Differentially Expressed Genes

The top 10 DE genes ordered by lowest P-value can be found in Table 2.

The topmost DE gene with the highest significance level is Prostaglandin F Receptor (PTGFR) which was downregulated 5-times more compared to normal (LogFC -5.2). Within the top 10 most DE genes only two genes, Parathyroid Hormone Like Hormone and Collagen Type IV Alpha 6 Chain were upregulated by approximately 4-times compared to Normal.

For all top 10 DE genes, the p-value is significantly below the standard significance threshold of 0.05, therefore we can conclude the results of the DE genes are statistically significant.

The FDR values for all top ten DE genes are significantly below 1 indicating there is a very low chance that these DE genes were falsely discovered, i.e., there is a very low chance that these genes are not differentially expressed.

**Table 2.** Top 10 Differentially Expressed genes in tumour samples. Sorted by P-value.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | RefSeqID | Symbol | Exons | logFC | Log CPM | LR | P-value | FDR |
| 5130 | NM\_001039585 | PTGFR | 4 | -5.182169 | 4.741036 | 98.67802 | 2.970812e-23 | 3.122323e-19 |
| 2235 | NM\_002820 | PTHLH | 4 | 3.969701 | 6.207473 | 92.12607 | 8.132692e-22 | 4.273730e-18 |
| 2317 | NM\_001111283 | IGF1 | 5 | -3.988427 | 5.715198 | 86.50859 | 1.391319e-20 | 4.874254e-17 |
| 3705 | NM\_033641 | COL4A6 | 45 | 3.655768 | 5.715666 | 77.51925 | 1.314354e-18 | 3.453465e-15 |
| 4755 | NM\_007168 | ABCA8 | 38 | -3.983256 | 4.937123 | 75.90535 | 2.975932e-18 | 6.255410e-15 |
| 2762 | NM\_005609 | PYGM | 20 | -5.480931 | 5.986916 | 75.34953 | 3.943421e-18 | 6.907558e-15 |
| 2707 | NM\_004320 | ATP2A1 | 23 | -4.620903 | 5.964161 | 74.79883 | 5.212059e-18 | 7.825535e-15 |
| 4445 | NM\_014440 | IL36A | 4 | -6.166057 | 5.401981 | 72.21203 | 1.932742e-17 | 2.539140e-14 |
| 745 | NM\_173352 | KRT78 | 9 | -4.245563 | 7.610423 | 70.80471 | 3.943953e-17 | 4.605661e-14 |
| 4037 | NM\_031469 | SH3BGRL2 | 4 | -3.934661 | 5.535112 | 67.76600 | 1.840968e-16 | 1.934858e-13 |

### 3e. Quality of Differential Expression Data

### Consistent counts changed between Tumour and Normal

Table 3 indicates that the top 10 genes are consistently and significantly different between Tumour and Normal samples for each patient. Supporting the hypothesis that these genes are either downregulated or upregulated in OCCS.

Differential expression analysis showed there were 3-times as many down-regulated genes as up-regulated genes (Table 4.)

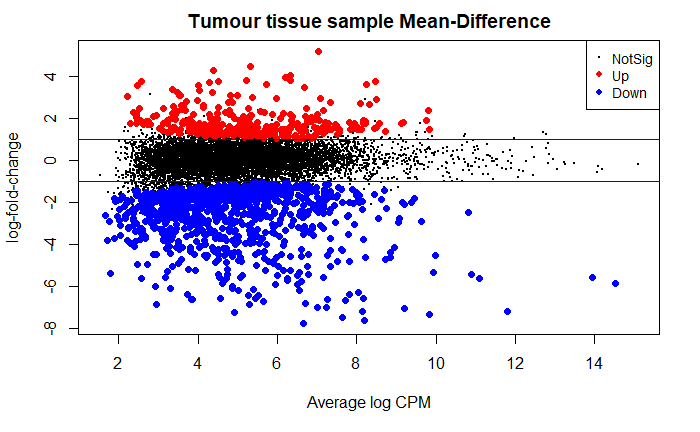
### **Table 3.** Counts-per-million of top 10 genes (by smallest P-value, significance) in each T / N sample.

Text

Description automatically generated

**Table 4.** The number of genes significantly differentially expressed (upregulated or downregulated) in tumour samples compared to normal.

|  |  |
| --- | --- |
|  | Tumour tissue |
| Down-regulated | 938 |
| Not Significant | 9241 |
| Up-regulated | 331 |



**Figure 3.** Mean difference plot comparing the differences (log-fold-change) between up-regulated and down-regulated genes in Tumour samples compared to normal. Genes with a fold change greater than 1 are highlighted in red (Up) and blue (Down).

There is a clear 2 Log-Fold-Change separation between upregulated and downregulated genes, as shown in the Mean Difference Plot (MDP) (Figure 3). There are approximately 3x more downregulated genes compared to Up, consistent with findings in Table 4. Upregulated genes tend to be more clustered, whereas down-regulated genes have the same amount of clustering but with more outliers with higher expression values(CPM). Downregulated genes are also significantly more downregulated with an extreme Log-Fold-Change value of -8 for some genes. The MDP shows that there tends to be higher variability in the higher expressed (>8 log CPM) and lowest differentially expressed (< -4 log-fold-change) genes.

Overall, this indicates the data is of good quality in that there is good separation between the DE genes.

### Gene Ontology – pathways

Table 5 – 8 show the various Biological Processes, Cellular Components, and Molecular Functions that are DE in Tumour samples. All results contained statistically significant P-values.

**Table 5.** Biological Process (BP) Ontology – top five, up-regulated, ordered by most significant

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Ont | N | Up | Down | P.Up | P.Down |
| dendritic spine organization | BP | 49 | 9 | 4 | 1.84E-05 | 0.618643 |
| negative regulation of cell fate commitment | BP | 3 | 3 | 0 | 3.11E-05 | 1 |
| neuron projection organization | BP | 58 | 9 | 5 | 7.44E-05 | 0.566637 |

**Table 6.** Biological Process (BP) Ontology – top five, down-regulated, ordered by most significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Ont | N | Up | Down | P.Up | P.Down |
| endomembrane system organization | BP | 209 | 6 | 37 | 0.6514564 | 1.631034e-05 |
| Golgi organization | BP | 53 | 3 | 13 | 0.2339261 | 4.206414e-04 |
| endoplasmic reticulum organization | BP | 25 | 0 | 8 | 1.0000000 | 8.332371e-04 |
| parturition | BP | 11 | 0 | 5 | 1.0000000 | 1.386215e-03 |

**Table 7.** Cellular Component (CC) ontology - up-regulated most significant

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Ont | N | Up | Down | P.Up | P.Down |
| ruffle | CC | 93 | 10 | 7 | 0.000665 | 0.699654 |
| plasma membrane-bounded cell projection | CC | 1117 | 54 | 99 | 0.000728 | 0.389442 |
| cell projection | CC | 1165 | 55 | 102 | 0.001137 | 0.440875 |

**Table 8.** Molecular Function (MF) ontology - up-regulated most significant – TOP 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Ont | N | Up | Down | P.Up | P.Down |
| kinesin binding | MF | 19 | 5 | 3 | 0.000244 | 0.221106 |
| cytoskeletal protein binding | MF | 476 | 28 | 40 | 0.001025 | 0.590282 |
| scaffold protein binding | MF | 52 | 6 | 4 | 0.005626 | 0.66572 |

## 4. DISCUSSION

In this study, we found many significantly differentially expressed genes in Tumour samples of OSCC patients compared to normal tissue. Here we will discuss a few relevant genes including Interleukin 36 Alpha (IL36A), Collagen Type IV (COL4A6), Parathyroid Hormone Like Hormone (PTHLH) and their relationship to cancer processes, how their pathways may provide assay or therapeutic targets for detection and treatment. Many other genes were differentially expressed with statistically significant P-values, discussing the most significant ones listed previously will provide a more impactful overview of the causes of OSCC tumour proliferation.

The major down and up-regulated Gene Ontology (GO) Biological Process (BP) pathways in tumour samples were Negative regulation of cell fate commitment and Golgi organization. These are all related to cellular proliferation functions which correspond well to the description of cancer being uncontrolled cellular growth. Notability, the top three down-regulated BP pathways were all related to the Golgi Apparatus and Endoplasmic reticulum, both involved in the processing of molecules and thus inter/intracellular signalling. This points to a common theme of the processing of proteins being a determinant of cancer proliferation.

### Interleukin 36 Alpha – IL36A

IL36A was downregulated by a LogFC of -6.1, which was the most LogFC out of the top ten DE genes with a high significance value (Table 1). IL36A encodes for a type of cytokine that activates inflammation activation pathways through different mechanisms and its main expression location is the skin cells. Specifically, this cytokine binds to receptor IL36R which induces an inflammatory response via the NF-kappa-B and MAPK signalling pathways (IL36A UniProt, n.d.). MAPK are kinases that induce apoptosis, inflammation, cell-cycle arrest, and cell differentiation among many other responses. With these cell processes being less activated the cell has less chance of mitigating cancer proliferation processes from occurring. For example, less apoptosis occurring provides a cell build-up, and older cells are more likely to introduce errors into their DNA, therefore, contributing to the cancer proliferation factors. Additionally, with fewer inflammation processes being instigated there is less ability for the immune response to identify cancer signs emanating from cells, therefore (for example) less chance for white blood cells to kill off cancer cells.

Some associations between IL36A and cancer have been explored. Decreased expression of IL36A has been linked to an increase in colorectal cancer tumour size and was determined to be a prognostic factor for survival rate (Bao et al., 2020). Similarly, IL36A is downregulated in Ovarian cancer, and lower levels of IL36A were linked to tumour progression. This indicates that IL36A may have anti-tumour effects through the action of the inflammatory response. There is some importance in the inflammatory response regarding Cancer.

IL36A may be a good treatment target by a therapeutic agent to induce the production of this cytokine, ideally inducing an inflammatory response to help with reducing cancer progression factors. Additionally, this cytokine or downstream proteins could be used as a marker for the prognosis of OSCC, by testing for low levels of this protein or levels of other proteins from affected downstream pathways.

### Collagen Type IV - COL4A6

COL4A6 encodes for Collagen Type IV, which is involved in giving structure to membranes and extra-cellular matrices (COL4A6 UniProt, n.d.). DE analysis showed that this gene was upregulated by a LogFC 3.6 (Table 1). As this molecule is involved with the structural components of cell forming, this corresponds well in the context of cancer; the uncontrollable growth of cells. With the proliferation of one cell type comes identical structural components needed to support the cell growth, and thus the tumour growth, evidently shown here by the high expression of a particular type of collagen.

Downregulation of COL4A6 has been shown to activate the p-FAK/MMP-9 signalling pathway in prostate cancer cells and promote progression and invasion of the tumour(Ma et al., 2020), something that could also be occurring in OSCC. The MMPs are known to be involved with tumour progression through the degradation of collagen IV in membranes and ECMs (Groblewska et al., 2012).

COL4A6 may have the potential for being an assay marker to prognose OSCC if it is easily detected within the blood or by some other method.

### Parathyroid Hormone Like Hormone - PTHLH

PTHLH (Parathyroid Hormone Like Hormone) was upregulated by 4-times and has various functions relating to the regulation of cell growth, development, and differentiation. Most notably PTHLH is a known regulator of tumour growth and proliferation in different cancer types and can either protect against normal cell apoptosis or drive tumour growth depending on the cancer type (Edwards & Johnson, 2021).

Due to this protein’s multifunctional aspect and diverse effects on tumour cell behaviour, there are many considerations to consider when determining ways of targeting therapeutics. This hormone has the potential to be used as a biomarker of OSCC proliferation as hormones are easily tested for in the blood. Targeting PTHLP production by inhibiting enzymes related to production or by inhibitors blocking binding to specific receptors within cellular growth pathways could be therapeutically beneficial. However, understanding the other downstream effects of reduced PTHLH will be essential to the efficacy.

Depending on the stage of cancer, targeting upstream enzymes or inhibiting/promoting direct producers could be beneficial for reducing early-stage proliferation or reducing resurgence from a dormant state.

### Other Differentially Expressed Genes

Several genes that were DE expressed with high significance levels did not pose meaningful and relevant functions in cancer proliferation nor corresponded with other research.

PTGFR was downregulated 5-times with high significance and is involved with G protein-coupled receptor activity and Prostaglandin F receptor activity. Research has indicated that the upregulation of this gene could be a cause of endometrial carcinomas (Sales et al., 2004) and ovarian cancer (Anderson et al., 2011). These are inconclusive with the results found here so there may be some other reason why these genes are highly downregulated.

A common trend among the downregulated genes is their gene ontology of Structural Molecule activity, including COL4A6, KRT78.

There were many other genes determined to be DE in tumour samples that could have significant prognosis or treatment capabilities. Here we have highlighted the most significant and relevant to the context of OSCC proliferation.

### Golgi Organisation

Biological Process ontology analysis showed that the Golgi Organisation pathway was significantly (p-value: 4.2 e-04) downregulated in tumour cells compared to normal (Table 6). There were several DE genes within this pathway, with 3 being upregulated and 13 downregulated. The Golgi apparatus is essential for trafficking, processing, and sorting proteins /lipids contributing to the regulation of cellular migration, mitosis, apoptosis, and DNA repair among many others. It has been shown that cancer cells exhibit dysfunctional Golgi in terms of structure and function, therefore leading to cellular processes being over-activated or dysregulated leading to the hallmarks of cancer proliferation. Ras GTPase (a known oncogene) expression can be abnormally affected leading to malignant transformation and proliferation, due to their function of transmitting cell growth signals (Spano & Colanzi, 2022).

Rab proteins are important proteins within the Golgi, that help regulate membrane transport of molecules via signalling to recruit effector molecules. They are members of the Ras-like GTPase family, and mutations in these genes are the cause of a wide variety of diseases (Liu et al., 2021). Rab proteins have been determined to regulate cell cycle and immune responses, and high expressions were significantly associated with advanced Colorectal cancer tumours (Jiang et al., 2022).

### Negative regulation of cell fate commitment

Biological Process ontology analysis showed that the ‘Negative regulation of cell fate commitment’ pathway was significantly (p-value: 3.11E-05) upregulated in tumour cells compared to normal (Table 6). This pathway is any process that stops, prevents or reduces cell apoptosis or differentiation. Since this process is upregulated then cells are not going to complete their normal cell cycles or differentiation into required cells, therefore contributing to the tumour malignancy.

All three genes within this pathway were upregulated posing a common method for therapeutic agents: inhibition of action. Downstream receptors of this pathway’s reactant molecules could be inhibited to enable proper cell cycles to occur, reducing the tumour progression. Upstream producers of the molecules in these pathways could be inhibited to reduce production to achieve the same outcome. Specific genes within this pathway include Gfi1 (growth factor independent 1 transcription repressor) and TNFSF18 (tumour necrosis factor). TNF is particularly important for the suppression of tumours via the regulation of T-cell response, and monocyte migration to sites of inflammation (TNFSF18 UniProt, n.d.).

### Conclusion

Various genes were identified as Differentially Expressed and are involved in several aspects related to cancer including the structural component of tumours, tumour growth regulation and immune responses. These specific targets include IL36A, COL4A6, and PTHLH hormone and their respective upstream production pathways or the receptors they bind to regulate downstream effects.

To extend the R analysis, extraction of the DE genes within this enriched GO term (Table 6) could determine whether the above genes were the specific ones DE up or down. Bioconductor resource provides methods by which this could be achieved.

Further research could focus on knock-outs to determine whether these genes are having the effect described here. Or over-expression studies to conclude the effect on tumour progression.

These findings provide avenues of for further studies to investigate these Differentially expressed genes as assay markers or targets for therapeutics to help diagnose and treat OSCC.

## 5. REFERENCES

Anderson, K. S., Sibani, S., Wallstrom, G., Mendoza, E. A., Raphael, J., Hainsworth, E., Montor, W. R., Wong, J., Park, J. G., Lokko, N., Logvinenko, T., Ramachandran, N., Godwin, A. K., Marks, J., Engstrom, P., & LaBaer, J. (2011). Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *Journal of Proteome Research*, *10*(1), 85–96. https://doi.org/10.1021/PR100686B/SUPPL\_FILE/PR100686B\_SI\_001.XLS

Bao, S., Hu, R., & Hambly, B. D. (2020). IL-34, IL-36 and IL-38 in colorectal cancer-key immunoregulators of carcinogenesis. *Biophysical Reviews*, *12*(4), 925–930. https://doi.org/10.1007/s12551-020-00726-0

Chung, M., Bruno, V. M., Rasko, D. A., Cuomo, C. A., Muñoz, J. F., Livny, J., Shetty, A. C., Mahurkar, A., & Dunning Hotopp, J. C. (2021). Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biology 2021 22:1*, *22*(1), 1–23. https://doi.org/10.1186/S13059-021-02337-8

*COL4A6 - Collagen alpha-6(IV) chain - Homo sapiens (Human) | UniProtKB | UniProt*. (n.d.). Retrieved October 30, 2022, from https://www.uniprot.org/uniprotkb/Q14031/entry#function

Edwards, C. M., & Johnson, R. W. (2021). From Good to Bad: The Opposing Effects of PTHrP on Tumor Growth, Dormancy, and Metastasis Throughout Cancer Progression. *Frontiers in Oncology*, *11*, 728. https://doi.org/10.3389/FONC.2021.644303/BIBTEX

Groblewska, M., Siewko, M., Mroczko, B., & Szmitkowski, M. (2012). The role of matrix metalloproteinases (MMPs) and their inhibitors (TIMPs) in the development of esophageal cancer. *Folia Histochemica et Cytobiologica*, *50*(1), 12–19. https://doi.org/10.5603/FHC.2012.0002

*IL36A - Interleukin-36 alpha - Homo sapiens (Human) | UniProtKB | UniProt*. (n.d.). Retrieved October 29, 2022, from https://www.uniprot.org/uniprotkb/Q9UHA7/entry#function

Jiang, X., Yang, L., Gao, Q., Liu, Y., Feng, X., Ye, S., & Yang, Z. (2022). The Role of RAB GTPases and Its Potential in Predicting Immunotherapy Response and Prognosis in Colorectal Cancer. *Frontiers in Genetics*, *13*, 60. https://doi.org/10.3389/FGENE.2022.828373/BIBTEX

Liu, J., Huang, Y., Li, T., Jiang, Z., Zeng, L., & Hu, Z. (2021). The role of the Golgi apparatus in disease (Review). *International Journal of Molecular Medicine*, *47*(4). https://doi.org/10.3892/IJMM.2021.4871

Ma, J. bin, Bai, J. Y., Zhang, H. B., Gu, L., He, D., & Guo, P. (2020). Downregulation of Collagen COL4A6 Is Associated with Prostate Cancer Progression and Metastasis. *Genetic Testing and Molecular Biomarkers*, *24*(7), 399–408. https://doi.org/10.1089/GTMB.2020.0009

Oral cancer - the fight must go on against all odds…. (2022). *Evidence-Based Dentistry 2022 23:1*, *23*(1), 4–5. https://doi.org/10.1038/s41432-022-0243-1

Pedro, N. F., Biselli, J. M., Maniglia, J. V., Santi-Neto, D. de, Pavarino, É. C., Goloni-Bertollo, E. M., & Biselli-Chicote, P. M. (2018). Candidate Biomarkers for Oral Squamous Cell Carcinoma: Differential Expression of Oxidative Stress-Related Genes. *Asian Pacific Journal of Cancer Prevention : APJCP*, *19*(5), 1343. https://doi.org/10.22034/APJCP.2018.19.5.1343

Rivera, C., & Venegas, B. (2014). Histological and molecular aspects of oral squamous cell carcinoma. *Oncology Letters*, *8*(1), 7–11. https://doi.org/10.3892/OL.2014.2103/HTML

Sales, K. J., Milne, S. A., Williams, A. R. W., Anderson, R. A., & Jabbour, H. N. (2004). Expression, Localization, and Signaling of Prostaglandin F2α Receptor in Human Endometrial Adenocarcinoma: Regulation of Proliferation by Activation of the Epidermal Growth Factor Receptor and Mitogen-Activated Protein Kinase Signaling Pathways. *The Journal of Clinical Endocrinology & Metabolism*, *89*(2), 986–993. https://doi.org/10.1210/JC.2003-031434

*Some key factors for number of significant DE genes - CVR Bioinformatics*. (n.d.). Retrieved October 31, 2022, from https://bioinformatics.cvr.ac.uk/some-key-factors-for-number-of-significant-de-genes/

Spano, D., & Colanzi, A. (2022). Golgi Complex: A Signaling Hub in Cancer. *Cells*, *11*(13). https://doi.org/10.3390/CELLS11131990

*TNFSF18 - Tumor necrosis factor ligand superfamily member 18 - Homo sapiens (Human) | UniProtKB | UniProt*. (n.d.). Retrieved October 31, 2022, from https://www.uniprot.org/uniprotkb/Q9UNG2/entry#function