

# Differential Gene Expression in High and Low Grade Head and Neck Carcinomas

Anne Sartori  
BS831 Final Project  
May 7, 2025

## 1 Introduction

Head and neck squamous cell carcinomas (HNSCC) are malignancies of the oral cavity, pharynx, hypopharynx, larynx, nasal cavity, and salivary glands that together constituted the seventh most common cancer diagnosis worldwide in 2023. In less economically developed countries, the use of alcohol and tobacco are thought to be related to the growth of HNSCC, while in more economically developed countries, human papillomavirus (HPV) is thought to be a more-important factor (Barsouk et. al. 2023, p. 1). A 2015 *Nature* article by authors affiliated with the Cancer Genome Atlas provided a comprehensive genomic characterization of HNSCCs. The present report analyzes a subset of those authors' data, gaining insight into the growth of tumors by identifying differences in gene expression between high-grade (G3) and low-grade (G1) tumors. This study has four particular aims: to document differential expression of genes with respect to tumor grade, to annotate this differential expression with previously documented information about genes and their functions, to explore which genes are useful in classifying genes as high- or low-grade tumors, and to explore whether cluster analysis captures the known division of samples into high and low grade tumors.

This report studies a subset of the RNA-seq data presented in the 2015 Cancer Genome Atlas article. The studied subset (henceforth, "the data set") contains information about the expression of information about 34422 genes from 40 samples each of high-grade (G3) and low-grade (G1) tumors, as well as 40 control samples that I do not report on here. The data set has information about 25 sample characteristics, including information about alcohol use and smoking. For 29 of these 80 samples, the data set contains information about the frequency of alcohol consumption, and for 42, it contains information about smoking (patient number of pack-years smoked);

approximately half of each are from G3 tumors, and the other half from G1 tumors. Only one sample is from an individual with HPV, and all samples are from the head and neck. (A description of the full data set is available on page 1 of the referenced *Nature* article.)

## 2 Methods

This study uses four methods to study genetic differences related to tumor growth. First, I conduct differential analysis using DESeq2 to determine which genes are differentially expressed between high-grade (G3) and low-grade (G1) tumors. I pre-process the data, removing zero rows, and keeping only the genes with the top 5000 mean absolute deviations (MADs) to reduce noise and discover biologically important variation. The DESeq2 package normalizes the data and uses an empirical Bayesian method to shrink the gene-wise dispersions towards a trend. To focus on the differences between the G3 and G1 samples that are due to differences in their cancer stage, my differential analysis adjusts for smoking and alcohol use. A disadvantage of adjusting for smoking and alcohol use is that, after filtering for missing on these covariates, the data set contains only six high- (G3) grade tumor samples and twelve low- (G1) grade tumor samples. I conduct all of the analyses in R, and list package versions used for these analyses at the start of the R file.

Second, I conduct gene set enrichment analysis using the 50 gene sets in the Hallmark homo sapiens database (part of the Molecular Signatures database) to discover biological processes or pathways that are linked to the set of differentially expressed genes.. The enrichment analysis uses hper with a bug fix provided by Lucas Weber; I use the Wald statistics from the differential analysis to rank the genes for the enrichment analysis, and the full data set of 34422 genes as the background set. I evaluate enrichment using Kolmogorov-Smirnov (KS) tests. Each KS test tests the null hypothesis that the ranks of the genes from a given gene set are distributed by chance in the ranked genes from the enrichment analysis, or signature.

Since my third and fourth steps do not involve the covariates, I return to the full 80 samples for this part of the study, still limiting the analyses to the genes with the top 5000 MADs. In the third step, I conduct classifica-

tion analysis (a type of supervised learning) to explore which combinations of gene expressions are the most successful at predicting whether the sample is from a high grade (G3) or low grade (G1) tumor. I choose this form of classification analysis based on success of the method in Carey et. al (2015). I train a penalized logistic regression model with Elastic Net regularization. I use the caret and glmnet packages in R, preprocessing the data by centering and scaling them. I choose the model parameters –  $\alpha$  (the mixing parameter between the Lasso and Ridge models) and  $\lambda$  (the regularization strength) – using a grid search with the ROC method of model selection and 5-fold cross-validation performed on the training set. After finding the optimal parameter combination, I evaluate the final model on a reserved test set. Interestingly,  $\alpha = 1$  in the chosen model, which means that the chosen model is fully a Lasso regression model. (The data are split, by genes, into training (75%) and test (25%) data sets (seed=42) at the start of this process. The regularization parameter is approximately 0.144 in the final model.)

Finally, I investigate the extent to which hierarchical clustering (a type of unsupervised learning) can recover the biologically meaningful division of samples that is created by the G3 and G1 labels. I pre-process the data using a variance-stabilizing transformation (using DESeq2) to make it more suitable for clustering analysis and heatmaps. Using the ComplexHeatmap package, I cluster genes and samples based on Euclidean distance and Ward’s D, which minimizes within-cluster variance, search for four clusters, and make a heatmap showing the clustered data.

The next section of the paper reports the results of these analyses. The last section investigates their substantive significance and discusses limitations of this work.

## 3 Results

### 3.1 Differential Expression Analysis and Gene Set Enrichment

The differential analysis identifies 44 genes that are significantly differentially expressed between high-grade (G3) and low-grade (G1) tumor samples

| Gene ID         | Gene Symbol | log2 fold change | Description                              |
|-----------------|-------------|------------------|--|
| ENSG00000118985 | ELL2        | -1.10938         | elongation factor, R NA polymerase II, 2 |
| ENSG00000186395 | KRT10       | -5.23483         | keratin 10, type I                       |
| ENSG00000137747 | TMPRSS13    | -2.20039         | transmembrane                            |
| ENSG00000126233 | SLURP1      | -5.08938         | protease, serine 13 secreted LY6/PLAUR   |
| ENSG00000134760 | DSG1        | -5.42365         | domain containing 1 desmoglein 1         |

Table 1: 5 most significant genes, DESeq 2 differential analysis for high-versus low-grade tumor samples

at the  $\alpha = 0.05$  level (using the adjusted p-values), adjusting for alcohol use and smoking. Table 1 shows the five most highly significant in order of significance. All of these five genes are down-regulated in high-grade tumors compared to low-grade ones. For example, the expression of ELL2, the top gene in the table, in high-grade tumor samples is approximately 46% of that in low-grade tumor samples. The expression of KRT10, the second in the table, in high-grade tumor samples is approximately 2.7% of that in low-grade tumor samples.

In fact, only four of the 44 significantly differentially expressed genes are up-regulated, or more highly expressed, in high grade tumors. Table 2 shows these four. The last 3, FHL1, FLNC, and DSG1, have over four times the expression in the high-grade tumor samples as in the low-grade ones.

Figure 1 presents an MA plot of the differential analysis results, with significant genes shown in red. The plot shows that the significant genes are fairly evenly distributed among mean expression levels (on the log scale) and also confirms visually that most of the significant genes are down-regulated in G3 relative to G1.

The differential expression analysis identifies a number of differences in gene expression between high- (G3) and low- (G1) grade tumors. The gene set enrichment analysis identifies sets of genes with known meaning that are overrepresented in the top ranks of the differential analysis.

| Gene ID         | Gene Symbol | log2 fold change | Description   |
|-----------------|-------------|------------------|---|
| ENSG00000100227 | POLDIP3     | 0.3880056        | encodes an RRM (RNA recognition motif)-containing protein that participates in the regulation of translation elongation factor, R                     |
| ENSG00000186395 | FHL1        | 2.0454521        | encodes a member of the four-and-a-half protein family  |
| ENSG00000128591 | FLNC        | 2.2842634        | encodes gamma filamen   |
| ENSG00000023445 | BIRC3       | 2.0738463        | encodes a member of the IAP family of proteins that inhibit apoptosis by binding to tumor necrosis factor receptor-associated factors TRAF1 and TRAF2 |

Table 2: Significant genes that are upregulated for high-grade versus low-grade tumors

Figure 2 shows the top twenty enriched gene sets, ranked by their false discovery rates (FDR). Of these, seven gene sets, shown in Table 3, are significantly enriched in the differentially expressed genes (FDR<0.05 level). Notably, the five most significantly differentially expressed genes (Table 1) do not appear in any of these top twenty gene sets, suggesting that enrichment captures broader biological patterns not necessarily driven by the most individually significant genes. An appendix to this report lists the matched genes and the brief description from the GSEA website of each process or pathway. I also discuss the findings further below.

### 3.2 Supervised and Unsupervised Learning

The Elastic Net regression analysis identifies a model containing ten genes as the best at classifying. These genes are shown in Table 3; again, they do not include the five most significant genes found in differential expression analysis. However, two of the genes, FABP5 and SERPINB5, appear in gene sets enriched by the differential analysis, the HALLMARK\_ESTROGEN\_RESPONSE and HALLMARK\_P53\_PATHWAY, respectively. (See appendix.)

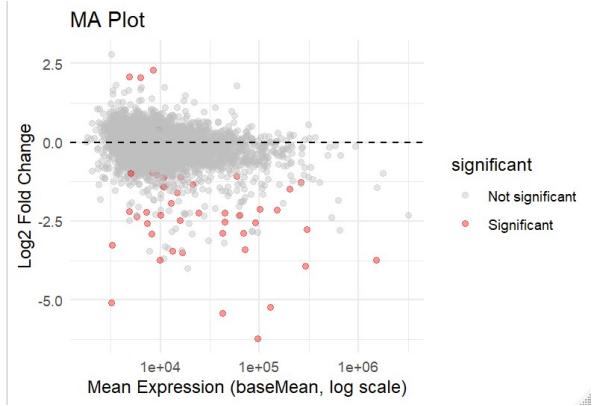


Figure 1: MA plot, differential analysis results

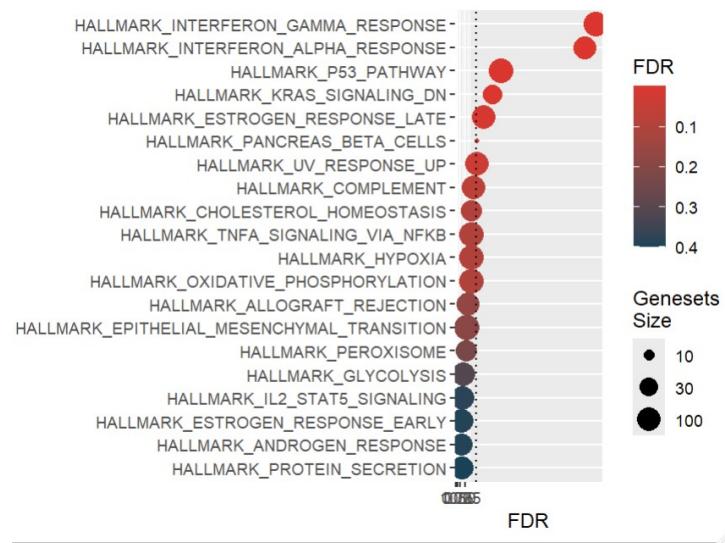


Figure 2: Top gene sets enriching differential analysis of high- vs. low-grade tumors

| label                              | score | fdr    |
|------------------------------------|-------|--------|
|                                    |       | 8.50E- |
| HALLMARK_INTERFERON_GAMMA_RESPONSE | 0.57  | 10     |
|                                    |       | 4.50E- |
| HALLMARK_INTERFERON_ALPHA_RESPONSE | 0.61  | 09     |
| HALLMARK_P53_PATHWAY               | 0.45  | 0.0013 |
| HALLMARK_KRAS_SIGNALING_DN         | 0.58  | 0.0042 |
| HALLMARK_ESTROGEN_RESPONSE_LATE    | 0.42  | 0.015  |
| HALLMARK_UV_RESPONSE_UP            | 0.43  | 0.041  |
| HALLMARK_PANCREAS_BETA_CELLS       | 0.67  | 0.041  |

Figure 3: Gene sets that enriched by the differential analysis, FDR < 0.05

| Gene ID         | Symbol   |
|-----------------|----------|
| ENSG00000164687 | FABP5    |
| ENSG00000189334 | S100A14  |
| ENSG00000158710 | TAGLN2   |
| ENSG00000206075 | SERPINB5 |
| ENSG00000100139 | MICALL1  |
| ENSG00000106278 | PTPRZ1   |
| ENSG00000182240 | BACE2    |
| ENSG00000168961 | LGALS9   |
| ENSG00000146670 | CDCA5    |
| ENSG00000067829 | IDH3G    |

Table 3: Genes selected by Elastic Net

The classification model identified by Elastic Net has a predictive accuracy of 0.88 in the training data and 0.85 in the test data, though the 95% confidence interval for the latter estimate is wide (0.62, 0.97). Table 4 shows the confusion matrix for the classification; 3 samples in the test data are misclassified by the model; two are misclassified as G3 and 1 is one is misclassified as G1.

Important statistics for cancer are the false negative and false positive rates (10% and 20%, respectively, in the test samples), since mischaracterizing the grade of the tumor might be consequential for treatment. While an accuracy of 85% might seem like a reasonable foundation for genetic testing of tumors, ideally doctors and patients would like the false positive and false negative rates to be lower. Further study might be merited; since the confidence interval is so wide, it is possible that studying a larger sample would identify a model similar to this one as the foundation of a useful genetic test.

|            |    | Reference |    |
|------------|----|-----------|----|
|            |    | G3        | G1 |
| Prediction | G3 | 9         | 2  |
|            | G1 | 1         | 8  |

Table 4: Elastic net model, confusion matrix

Thusfar, this report has focused on evaluating the genes and pathways associated with differences between high- and low-grade tumors. I now change focus, examining the ability of hierarchical clustering to identify key differences among samples. In particular, I explore whether clustering can recover the known information about the samples in our data set, particularly whether they come from high- or low-grade tumors.

The heatmap in Figure 4 shows the results of the hierarchical clustering. The samples (shown across the top of the heatmap) are broken into two main clusters. While these clusters roughly mirror the known division of the samples into high- (G3) and low- (G1) grade tumors, the cluster at the left, which primarily consists of grey G3 samples, is striped with pink G1 samples, and the cluster at the right, which primarily consists of pink G1 samples, is striped with grey G3 samples. This exercise suggests that, absent known labels for the samples, hierarchical clustering can be useful in categorizing samples, but also that it is imperfect at doing so.

The heatmap shows two or three blocks of high-expressed genes (in red) that might merit further study. First, genes in the top part of cluster 1 (vertical axis) seem to be highly-expressed mostly for low-(G1) grade tumors

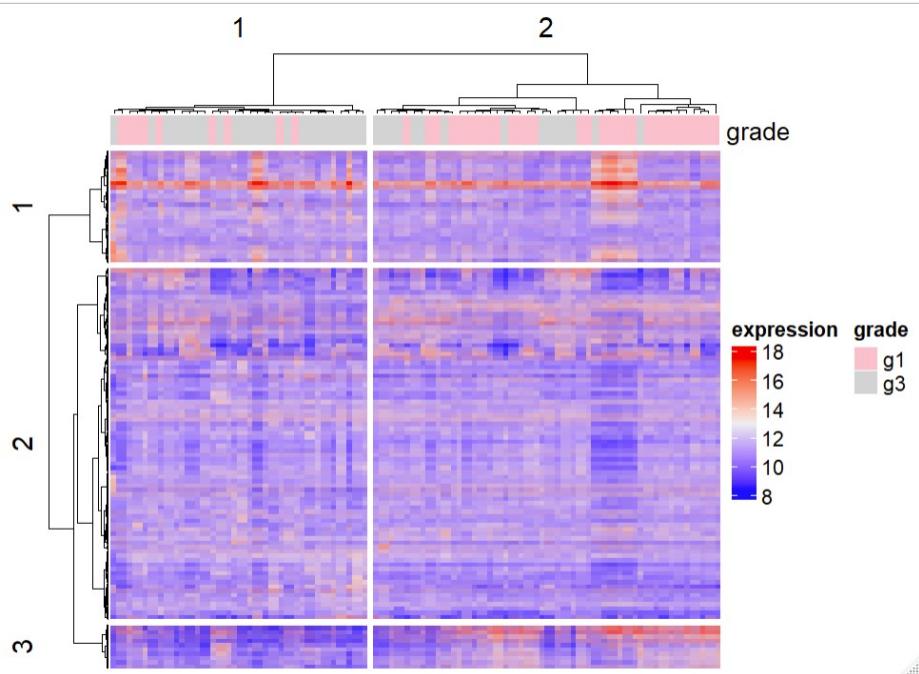


Figure 4: Heatmap with cluster analysis of gene expression in the data set

(shown in pink) in a branch of Cluster 2 (as part of its right-ward branch). Second, genes in Cluster 3 (vertical axis) also appear highly-expressed mostly for low-grade tumors, including the same branch identified previously but also additional samples to right.

## 4 Discussion and Conclusion

By identifying genes that are differentially expressed in high- (G3) and low- (G1) grade tumors, the differential analysis provides useful information for studying the growth of tumors. The gene set enrichment analysis identifies seven gene sets significantly enriched by the differentially expressed genes, and the gene matches between these processes and pathways and the data set do not include any of the top five genes identified in the differential analysis, but themselves provide additional information relevant to tumor growth.

This section explores that some of the clues these analyses provide, and discusses strengths and limitations of the report and directions for future work.

As mentioned earlier, the vast majority (40/44) of the significantly differentially expressed genes found in the DESeq2 analysis were downregulated in the high-grade versus low-grade tumors. Just as the top five genes from the differential analysis are downregulated with regard to high-grade (G3) as opposed to low-grade (G1) tumors, every matched gene in the seven pathways that enrich the analysis is down-regulated in the high- (G3) versus low- (G1) grade tumor samples, as evidenced by their negative log<sub>2</sub> fold changes in the DESeq2 results.

At least three of the five most highly significant differences listed in Table 1 (all downregulated for high-grade versus low-grade tumors) occur in genes already thought to be associated with tumors. ELL2, is used in normal plasma cell function and immunoglobulin production; sequence variation is associated with multiple myeloma.<sup>1</sup> TMPRSS13 encodes a protein that is known to function in tumorigenesis.<sup>2</sup> SLURP1 encodes a protein that is thought to contain antitumor activity.<sup>3</sup> The National Library of Medicine summaries of the last two genes, KRT10 and DSG1, do not discuss tumor growth; if they are not already known to be involved in HNSCC tumor growth, their role might be worth investigating, since both have much lower expression in high- than in low-grade tumors. KRT10 is involved in maintaining the cytoskeleton;<sup>4</sup> its expression in high grade tumors is 2.7% that in low-grade tumors. DSG1 encodes a member of the desmoglein protein subfamily, a component of cell-cell junctions that help resist mechanical stress;<sup>5</sup> its in high-grade tumors is 2.3% that in low-grade tumors.

Since so few differentially expressed genes are upregulated in the high-grade versus low-graded tumors, it may be useful to study the four that,

---

<sup>1</sup>[https://en.wikipedia.org/wiki/Elongation\\_factor\\_for\\_RNA\\_polymerase\\_II](https://en.wikipedia.org/wiki/Elongation_factor_for_RNA_polymerase_II)?utm\_source=chatgpt.com,

<sup>2</sup><https://www.ncbi.nlm.nih.gov/gene/84000>

<sup>3</sup><https://www.ncbi.nlm.nih.gov/gene/57152>.

<sup>4</sup>See also [https://amigo.geneontology.org/amigo/gene\\_product/UniProtKB:P13645](https://amigo.geneontology.org/amigo/gene_product/UniProtKB:P13645)

<sup>5</sup><https://www.ncbi.nlm.nih.gov/gene/3858>,  
<https://www.ncbi.nlm.nih.gov/gene/1828>

contrary to trend, are up-regulated (Table 2). For example, FHL1 has been identified by prior research as downregulated in papillary thyroid carcinoma (PTC) samples as opposed to normal tissues (Zeng et. al., 2024). Further study might identify why this gene is instead upregulated in high-grade HNSCCs. A second gene that is upregulated in the G3 samples, BIRC3, is part of a family of Inhibitor of Apoptosis (IAP) genes that are often overexpressed in tumor samples because they block normal cell death (e.g, Chen 2013).

The gene set enrichment links the differential analysis to previously-documented groups of genes with partially-known functions or characteristics. Here, I explore two of the gene sets enriched in the analysis. First, the Gene Set Enrichment Analysis (GSEA) website identifies Hallmark\_Interferon\_Gamma\_Response as, "genes up-regulated in response to IFNG."<sup>6</sup> The GSEA site also identifies IFNG (IFN- $\gamma$ ), as a gene that encodes a protein secreted by the adaptive and innate immune systems. Second, the GSEA site identifies Hallmark\_Interferon\_Alpha\_Response as "genes up-regulated in response to alpha interferon proteins," also related to the immune system. Since the genes in the Interferon\_Alpha-Response and Interferon-Gamma\_Response gene sets have greater expression in low-grade tumors and are up-regulated with regard to the relevant proteins, it is likely that both IFN- $\gamma$  and IFN- $\alpha$  proteins are more prevalent in low-grade tumors than in high-grade ones.

IFN- $\alpha$  is believed to inhibit tumor growth (Bellardelli et. al. 2002), suggesting that the greater activity of the Alpha\_Response genes in the low-grade (G1) tumor samples indicates an effort to suppress the growth of tumors. The analyses reported here do not shed light on why Alpha\_Response genes are less active in the high-grade tumor samples, but it is possible that by the time it reaches G3 grade, the tumor has harmed the patient in a way that inhibits these genes' activities. While it is established in the literature that IFN- $\gamma$  affects tumor growth, Jorgovanovic et. al. (2020) discuss both pro- and anti-tumorigenic effects of this protein. For this reason, the enrichment of Hallmark\_Interferon\_Gamma\_Response in the differentially expressed genes could suggest that the genes' greater expression in the low-

---

<sup>6</sup>[https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HALLMARK\\_INTERFERON\\_GAMMA\\_RESPONSE.html](https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/HALLMARK_INTERFERON_GAMMA_RESPONSE.html)

grade (G1) tumors relative to the high-grade (G3) ones is either facilitating or depressing the growth of tumors in those samples.

Overall, this report provides a broad investigation of the subset of the HNSCC data set; it investigates the genes related to the growth of HNSCCs using differential expression analysis, gene enrichment analysis, and both supervised and unsupervised learning. In doing so, it casts a wide net to search for these genes, and fills the net with a variety of findings. The top genes identified by the differential expression analysis merit further study, as do the roles of the seven identified gene sets in the growth of HNSCC. Together with genetic testing, models such as the Elastic Net model identified here might be useful in diagnosing high- and low-grade tumors, though further study of larger data sets would be necessary to better understand the rate of false positives and negatives.

In casting a wide methodological net, this study was unable to go very deep with any one method. In particular, the differential expression analysis reported here focuses on the genes with the top 5000 MAD levels in order to find biologically meaningful differential expression. This relatively narrow focus may miss subtler but still meaningful differences between lower expressed genes; a follow-up study of these same data might set the MAD bar lower.

Another extension of this study could compare expression of these particular genes in the low-grade (G1) tumor samples to the control (AE) samples in the data to shed light on the meaning of the differential expression between the high-grade (G3) and G1 samples; doing so would provide further insight into the growth of tumors. Some of the many genes that are upregulated in the low-grade tumor samples relative to the high-grade ones also may be upregulated in low-grade tumors relative to nontumorous tissue, suggesting that their high expression in G1 relative to G3 samples is associated with the body's response to the tumor, possibly as a part of its attempt to fight off cancer, or alternatively as a result of the tumor's coopting the genes to enable its own growth. Other genes that are differentially expressed between the high- and low-grade tumor samples may not be differentially expressed between low-grade tumors and nontumorous tissue, suggesting that their expression in the low-grade tumors is the body's normal expression level, which

is then suppressed in higher-grade tumors. Such an analysis would help to shed light on the role that various genes play in the growth of tumors.

A challenge posed by the DESeq2 approach is that knowing which genes are differentially expressed does not, by itself, answer complicated questions about how these genes affect the genesis and growth of tumors. This challenge is illustrated by the finding discussed earlier in this section that the differential analysis of expression in high- and low-grade tumors is enriched by INTERFERON\_GAMMA\_RESPONSE, where the interferon- $\gamma$  protein has been identified in prior research as *both* promoting and inhibiting tumor growth depending on the tumor microenvironment (Jorgovanovic et. al., 2020). That is, the fact that the genes in INTERFERON\_GAMMA\_RESPONSE are downregulated in high-grade (G3) relative to low-grade (G1) tumors does not tell us what larger purposes they are contributing to in low-grade tumors. The DESeq2 analysis can provide clues about the tumor microenvironment by telling researchers about the simultaneous regulation of other genes along with the upregulation of many of the INTERFERON\_GAMMA\_RESPONSE genes in low- (G1) versus high- (G3) grade tumors.. However, other methods, such as the study of single-cell RNA-seq data, might be better suited to studying the tumor microenvironment, and thus might provide a more complete understanding of the role of interferon- $\gamma$  and of the genes that respond to this protein.

## 5 Appendix

This appendix provides the brief description provided by the Gene Set Enrichment Analysis website (GSEA), maintained by the Broad Institute and UC-San Diego, of each gene set significantly enriched by the DESeq2 analyses. The descriptions were found by searching for the gene set at <https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/> on 5/1/2025. The appendix also lists the matched genes for each pathway from the enrichment analysis. All of these genes were downregulated for G3 relative to G1 in the differential analysis. Genes shown in blue were chosen as part of the Elastic Net model discussed in the report.

**HALLMARK\_INTERFERON\_GAMMA\_RESPONSE:** "Genes up-regulated in response to IFNG [GeneID=3458]" (GSEA)

**matched genes:** UPP1, HERC6, CDKN1A, MX1, CXCL10, ISG15, EPSTI1, IFI27, XAF1, BST2, IFI44, HLA-B, HLA-A, ISG20, IFI44L, IL4R, HIF1A, CASP7, DDX60, VAMP8, MX2, OAS2, MVP, IFIT1, B2M, SP110, GBP6, IRF7, PSMB8, C1S, PARP12, NLRC5, HLA-G, TNFSF10, CXCL9, PNP, CASP1, CMPK2, PSME2, PML, UBE2L6, SECTM1, PSMB9, IFIT3, LAP3, IFITM3, IFIH1, RNF213, STAT2, IFI35, SERPING1, TRIM21, PIM1, LY6E, TAP1, CFH, CASP4, HELZ2, IDO1, SAMD9L, NAMPT, LGALS3BP

**HALLMARK\_INTERFERON\_ALPHA\_RESPONSE:** "Genes up-regulated in response to alpha interferon proteins." (GSEA)

**matched genes:** HERC6, MX1, CXCL10, ISG15, EPSTI1, IFI27, BST2, IFI44, LPAR6, ISG20, IFI44L, IL4R, DDX60, B2M, SP110, IRF7, PSMB8, C1S, PARP12, IFITM1, CASP1, CMPK2, PSME2, PARP9, UBE2L6, PSMB9, IFIT3, HLA-C, LAP3, IFITM3, IFIH1, STAT2, IFI35, GBP2, TRIM21, LY6E, TAP1, HELZ2, SAMD9L, NCOA7, LGALS3BP, LAMP3

**HALLMARK\_P53\_PATHWAY:** "Genes involved in p53 pathways and networks." (GSEA)

**matched genes:** PERP, ST14, PLK2, UPP1, SERPINB5, SFN, KLK8, CDKN1A, FOS, GM2A, CLCA2, HRAS, HBEGF, BTG2, ATF3, SDC1, VAMP8, FAM162A, VDR, RPS12, TPD52L1, EPS8L2, BAK1, JUN, BAX, PHILDA3, TGFB1, TM4SF1, CASP1, RPL36, RAP2B, MKNK2, MAPKAPK3, KRT17, SLC7A11, RPL18, CTSD, SAT1, CD81, TAP1, EPHX1, MXD1, RNF19B, BMP2, S100A4, S100A10, IER5, DNNTIP2, PPP1R15A, SLC3A2, EI24, KLF4

**HALLMARK\_KRAS\_SIGNALING\_DN:** "Genes down-regulated by KRAS activation" (GSEA)

**matched genes:** AKR1B10, KRT1, PKP1, CHST2, CALML5, LYPD3, KLK7, KLK8, MX1, IGFBP2, BTG2, KRT5, IFI44L

**HALLMARK\_ESTROGEN\_RESPONSE\_LATE:** "Genes defining late response to estrogen." (GSEA)

**matched genes:** PERP, FABP5, S100A9, ST14, PKP3, SFN, IDH2, CA2, CXCL14, FOS, PPIF, PDCD4, KLK10, HOMER2, BTG3, SULT2B1, CLIC3, CDH1, ISG20, KLK11, CKB, TFAP2C, TPD52L1, TRIM29, MAPK13, ASS1, FLNB, NBL1, KRT13, SCNN1A, IMPA2, GJB3, CCND1, TPBG, CA12, AMFR

**HALLMARK\_UV\_RESPONSE\_UP:** "Genes up-regulated in response to ultraviolet (UV) radiation" (GSEA)

**matched genes:** AQP3, FOSB, TUBA4A, CLTB, DNAJB1, CA2, AP2S1, FOS, AMD1, PPIF, CTSV, YKT6, GGH, BTG3, IGFBP2, BTG2, ATF3, HLA-F, BSG, FURIN, BAK1, EIF2S3, ALAS1, EIF5, JUNB, ATP6V1F, RHOB, CYB5B, TAP1, CYB5R1, EPHX1, BMP2, KLHDC3, TST, CREG1

**HALLMARK\_PANCREAS\_BETA CELLS:** "Genes specifically up-regulated in pancreatic beta cells." (GSEA)

**matched genes:** MAFB, VDR, SEC11A, SPCS1, FOXO1, SRP14, SRPRB

## 6 References

- Balardelli, F., Ferrantini, M., Proietti, E., & Kirkwood, J.M. (April, 2002), Interferon-alpha in tumor immunity and immunotherapy. *Cytokine & Growth Factor Reviews*, 13 (2) 119-134. <https://www.sciencedirect.com/science/article/abs/>
- Barsouk, A., Aluru, J. S., & Rawla, P. (2023). Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Medical Sciences*, 11(2), 42. <https://doi.org/10.3390/medsci11020042>
- Cancer Genome Atlas, Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*
- Carey, C. D., Gusenleitner, D., Chapuy, B., Kovach, A.E., Kluk, M.J., Sun, H.H., Crossland, R.E., Bacon, C.M. Rand, V., Dal Chin, P., Le, L.P., A., Neuberg, D., Sohani, A.R., Shipp, M.A., Monti, S., & Rodig, S.J. (2015). Molecular classification of MYC-driven B-cell lymphomas by targeted gene expression profiling of fixed biopsy specimens. *The Journal of Molecular Diagnostics*, 17(1), 19–30. <https://doi.org/10.1016/j.jmoldx.2014.08.006>.
- Chen, X., Wang, T., Yang, D., Wang, J., Li, X., He, Z., Chen, F., Che X., & Song, X. (2013) Expression of the IAP protein family acts cooperatively to predict prognosis in human bladder cancer patients. *Oncology Letters*. Published online on: January 23, 2013 <https://doi.org/10.3892/ol.2013.1150>: 1278-1284
- Jorgovanovic, D., Song, M., Wang, L., and Zhang Y. (2020) Roles of IFN- $\gamma$  in tumor progression and regression: a review. *Biomarker Research* 8 (49) <https://pubmed.ncbi.nlm.nih.gov/33005420/>

Referenced websites are listed in footnotes.