The Clinical, Genetic and Therapeutic Implications of Perineural Invasion in Head and Neck Squamous Cell Carcinomas

Hirad Hosseini, University of Southern California

QBIO 490: Multi-Omic Data Analysis, Fall 2022

I. Introduction

Head and neck cancers are primarily composed of squamous cell carcinomas (SCCs) of the oral cavity, pharynx and larynx. They account for nearly 50,000 cases and 11,000 deaths in the US annually, a small portion of the 500,000 forecasted global incidences per year (Vokes, E. et al., 1993). Such figures qualify HNSCC as the sixth most prevalent cancer worldwide (Zhou, G., 2016). Common risk factors for oral cavity and laryngeal SCCs include tobacco consumption, alcohol abuse, occupational exposure during nickel refining or woodworking and more. A distinct risk factor for pharyngeal SCCs is infection with the human papilloma virus (HPV), particularly the HPV-16 strain (Johnson, D. et al., 2020). Epidemiological factors include the dietary availability of fruits and vegetables which are suspected to supply dietary carotenoids inversely associated with HNSCC incidence, age, occupation, sex and more. Men over 50 years old with history of tobacco usage and/or alcohol abuse are the primary demographic affected by HNSCCs, though a genetic component is suspected due to some incidence in young patients and patients with no history of tobacco and alcohol use (Vokes, E. et al., 1993).

Perineural invasion (PNI)–characterized as the invasion of surrounding nerve tissue by tumor cells– is an important pathologic risk factor often observed in HNSCC cases. The presence of PNI is associated with poor morbidity and mortality outcomes due to the extensive involvement of potentially both the central and peripheral nervous systems (Zhang, Z. et al., 2019). However,

the genetics and molecular pathways underlying the presence of PNI in HNSCCs have yet to be fully understood. The Clinical Genome Atlas Data Bank operated by the National Cancer Institute offers a trove of relevant clinical, genomic, transcriptomic and proteomic datasets attributed to over 500 cases of HNSCCs (The Cancer Genome Atlas Network, 2015). Access to pertinent perineural invasion status and transcriptomic profiles for HNSCC patients offers a promising avenue for the identification of genes as potential biomarkers to determine PNI status, and thus prognostic outcomes. Herein, we will utilize the clinical and mutation annotated format (MAF) databases of the TCGA-HNSCC Project to identify commonly mutated genes or mutation types associated with PNIs and their impact on patient survival rates as obtained via co-lollipop and mutational co-occurrence plots. We will also utilize the drug sub-dataframe of the clinical database to observe the effects of PNI on patient survival rates when administered various chemotherapeutics as observed in Kaplan-Meier plots. Overall, exploring the genetic and biomolecular basis of HNSCCs can be fruitful in regards to developing new diagnostic and prognostic biomarkers and determining the patient's therapeutic course. These analyses will corroborate previous findings that PNIs are generally associated with poor prognostic outcomes and that they exhibit a unique composition of mutations that is differentiable from that of non-PNI HNSCCs (Saidak, Z. et al., 2018). Furthermore, our analysis will uncover several candidate genes and biochemical pathways that are overexpressed in PNI development, along with the utility of PNI status as an independent predictor for patient prognostic and chemotherapy outcomes.

II. Methods

All statistical analyses were conducted locally utilizing R statistical software. The ggplot2, BiocManager, TCGAbiolinks, maftools, survival, survminer and dplyr libraries were installed

and loaded prior to any analysis. Clinical, drug and MAF data from NCI's TCGA database was queried, downloaded and processed into a local dataframe utilizing the accession code "TCGA-HNSC ". The clinical data database was preprocessed to exclude tumors with primary sites in the larynx, hypopharynx and lip via boolean masking to include only anatomic subdivisions within the oral cavity. The clinical dataframe was then subsetted into two smaller dataframes (via boolean masking)–peri_inv_clinical and no_peri_inv_clinical–corresponding to the presence and absence of perineural invasion within the tumor samples. The R default function boxplot() was used to construct a boxplot comparing lymph node positivity counts obtained via histological examination and perineural invasion status for each sample. Then, the R library ggplot2 was used to construct histogram plots of the pathologic stage classifications for samples with and without perineural invasion. The clinical dataframe was cleaned of NA and infinity values (via boolean masking) before being inputted into functions from the R libraries survival and survminer to construct a Kaplan-Meier survival plot. The R library maftools was used on the MAF database to an oncoplot of all HNSCC samples. The MAF database was then subsetted into two databases(via boolean masking)–y_inv_maf and n_inv_maf–denoting the presence and absence of perineural invasion respectively. These two dataframes were then inputted into functions from the maftools library to construct a co-oncoplot, several co-lollipop plots and several somaticMutations plots. The drug sub-dataframe from the clinical database was first masked to include only samples treated via chemotherapeutic agents. The drug names were then counted via the R library dplyr and the top 3 most common drug names were saved into a vector. After merging the clinical and clinical.drug dataframes into an all-encompassing total_clinical dataframe, we masked this dataframe to include patients who were treated with one of the top 3 drugs. Following cleaning of NA and blank values, the total_clinical dataframe was

inputted into various functions from the surv and survminer libraries to construct Kaplan Meier plots of the overall HNSCC sample set, the PNI-present sample set and the PNI-absent sample set, each containing three curves corresponding to each of the three drugs.

III. Results

Our analysis demonstrated that PNI status was viable as an independent prognostic marker, as PNI-present patient populations experienced poor long-term survival rate outcomes compared to PNI-absent patient populations (Figure 1).

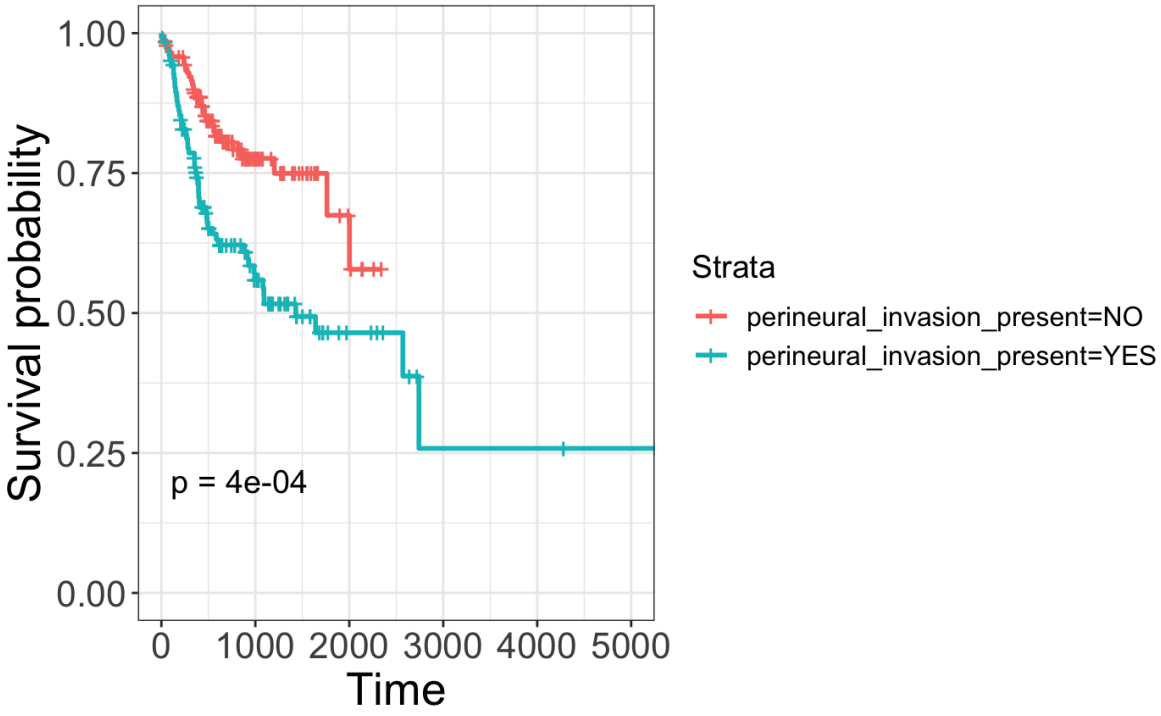

Figure 1. Kaplan-Meier survival plot showing that HNSCC patient populations with PNI experienced poor long-term prognostic outcomes compared to their non-PNI counterparts. P-value of 4 x 10^-4 indicates the statistical significance of the aforementioned result. Time is represented as the difference between the days to death (or days to last follow up if such data is missing) and the day of diagnosis.

When compared to both histologically examined lymph node positivity counts and pathologic staging classifications, PNI status was fairly independent of factors measuring tumor aggression and proliferation (Figures 2 and 3).

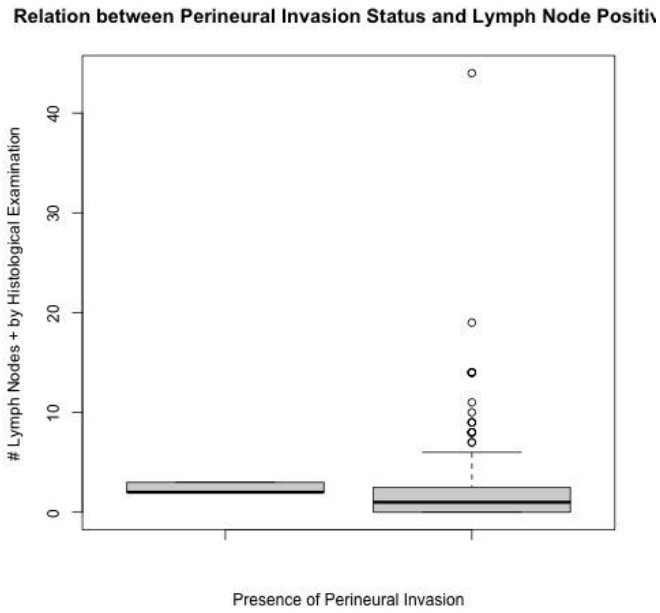

Figure 2. Boxplot showing that histologically examined cancer-positive lymph node counts are not significantly differentiated by PNI status, although significant outliers can be observed for PNI-positive samples. The plot on the left denotes PNI-absent patient samples whereas the right denotes PNI-present patient samples. While abnormally high positive lymph node counts may be indicative of PNI presence, the mean and interquartile values for both plots fall within the 0-5 count range.

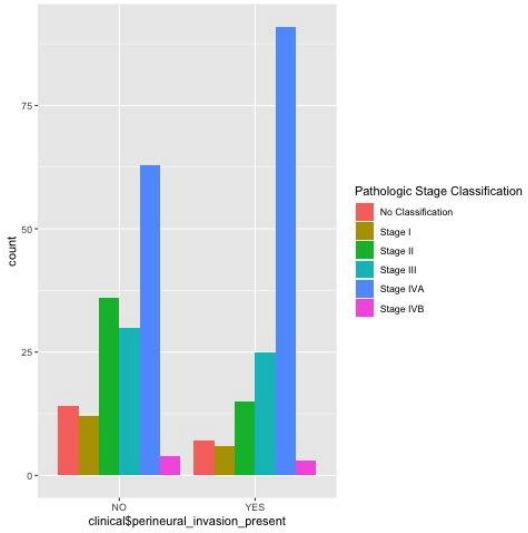

Figure 3. Histogram plot revealing similar distribution of pathologic stage classifications for HNSCC patient samples differentiated by PNI status, demonstrating that PNI is not necessarily a discriminating factor for more advanced HNSCC tumors. The "No Classification" count shows the number of samples for which no staging procedure was performed. The high prevalence of Stage IVA tumors can indicate the poor early diagnostic capabilities for HNSCCs. Both distributions appear to be non-Gaussian and skewed left.

An oncoplot construction of mutation data revealed a combination of both classical tumor suppressor genes such as TP53 and tissue-specific genes such TNN and SYNE1, which encode proteins involved in the musculoskeletal system (Figure 4). The mutation types were majorly composed of missense mutations followed by an even breakdown of multi-hit and nonsense mutations. Frameshift mutations were only significant in the TP53 and FAT1 genes. Only one gene–TP53–was implicated in a majority of HNSCC samples, meaning that there is a notable extent of heterogeneity within mutational compositions of the samples induced potentially by PNI status and other factors.

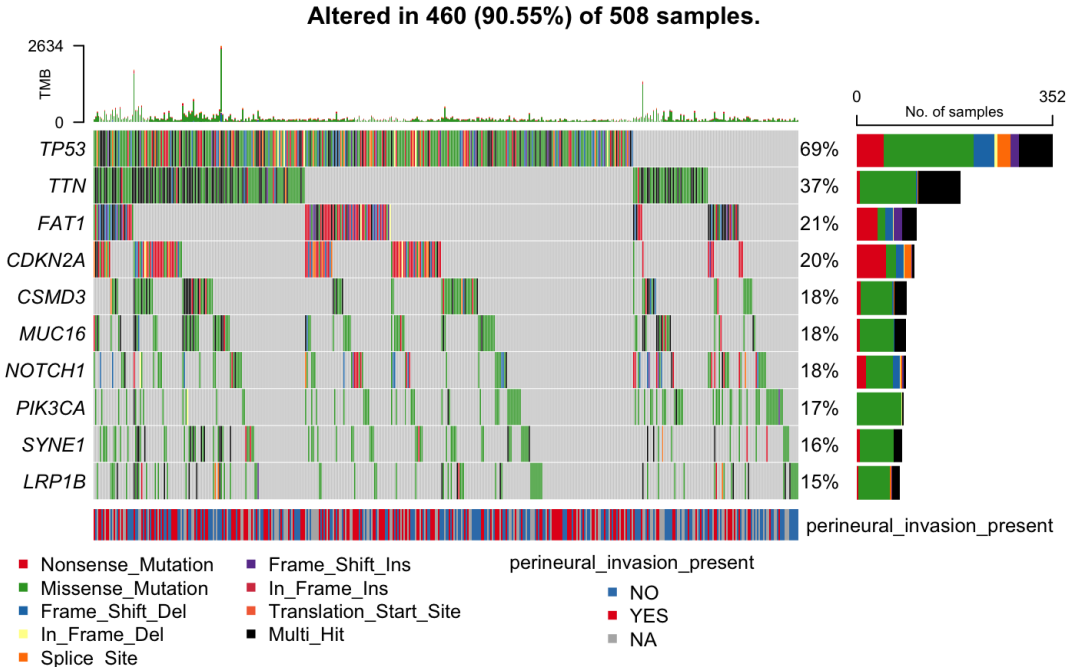

Figure 4. An oncoplot of all tissue samples of HNSCC patients demonstrates mutations affecting both general cell growth/proliferation pathways and tissue-specific pathways in the musculoskeletal system. Gray bars denote samples for which PNI status was not collected. Only the top 10 most prevalent mutated genes are shown on this plot.

When stratifying based on PNI status, there is a variety in prevalence of specific mutations with some mutations equally represented while others potentially serving as discriminators of PNI status (Figure 5). Among such differentiating mutations are TP53, CDKN2A, FAT1, NOTCH1, LRP1B and MUC16. There are no discernable differences in mutation type based on PNI status, as most samples are predominantly impacted by missense mutations.

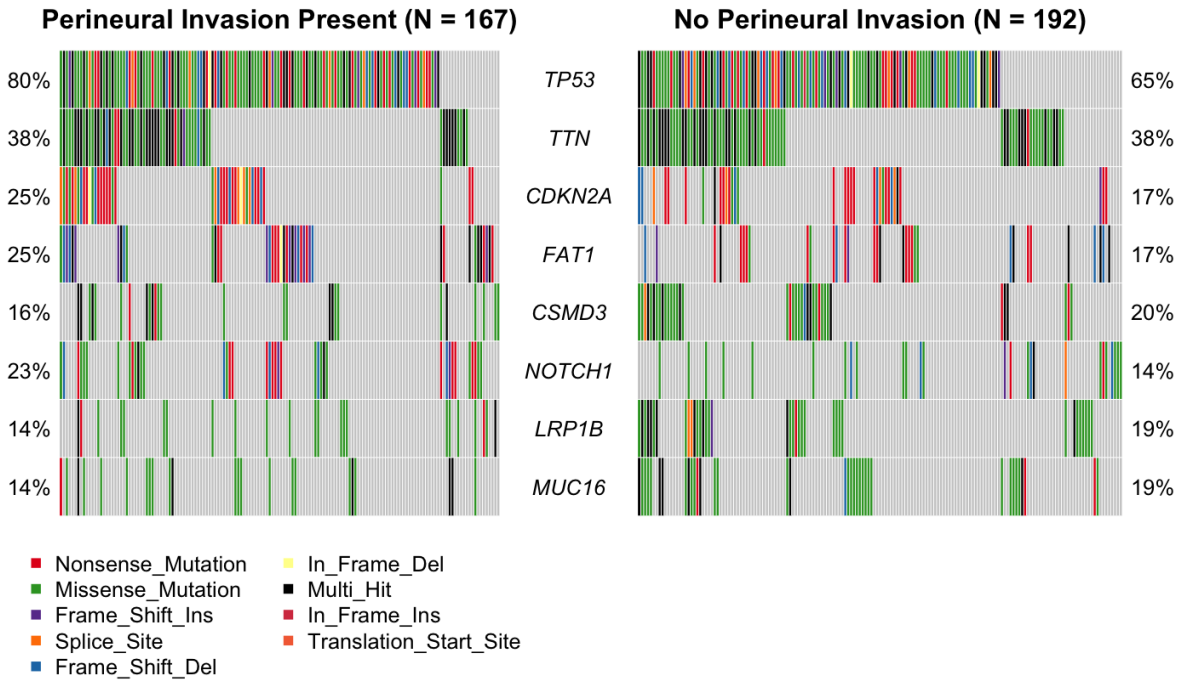

Figure 5. A co-oncoplot stratified by PNI-status indicates that certain gene mutations can be potentially used as determinants of whether an HNSCC sample will exhibit PNI. These numerous candidate genes must be further analyzed to determine if they can be used as biomarkers for PNI in HNSCC cases.

Co-lollipop plots of several variably expressed mutations (as elucidated in Figure 5) show minor variations in mutation site and type along the gene responsible for different prevalences based upon PNI status (Figures 6a-d). TP53 exhibits both a major missense and nonsense mutation along the center of the coding region of the gene in PNI-present samples but only exhibits one major missense mutation in the rightmost portion of the coding region in PNI-absent samples (Figure 6a). Mutations in NOTCH1 appear to be heterogeneous in both site and type regardless of PNI status (Figure 6b). MUC16 is uniformly rich in mutations in PNI-present samples whereas mutations tend to concentrate the rightmost portion of the gene in PNI-absent samples (Figure 6c). Mutations in LRP1B appear to be heterogeneous in both site and type regardless of PNI status (Figure 6d). Overall, minor variances in mutation site and type can help differentiate PNI status of HNSCC samples both by overall expression of the gene and micro-alterations in the sequence of the gene.

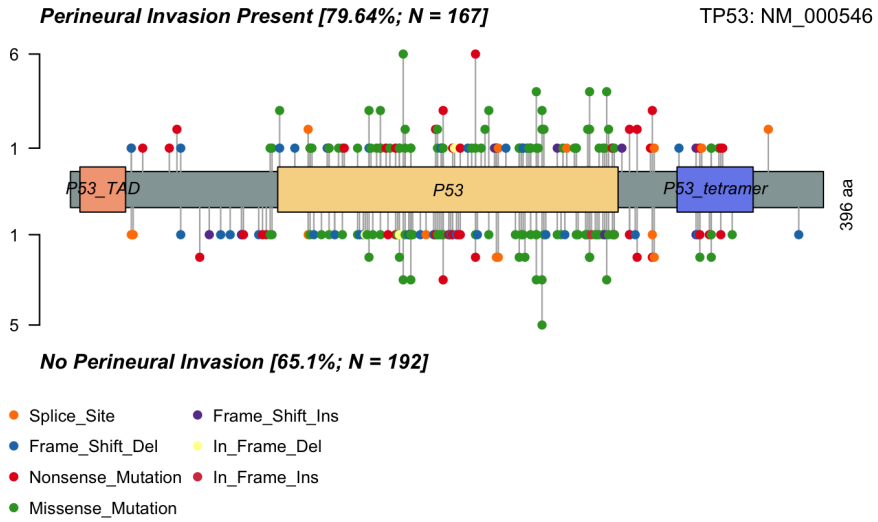

Figure 6a. Co-lollipop plot of TP53 stratified by PNI status shows a marked difference in both mutation site and type for major sites. TP53 exhibits both a major missense and nonsense mutation along the center of the coding region of the gene in PNI-present samples but only exhibits one major missense mutation in the rightmost portion of the coding region in PNI-absent samples

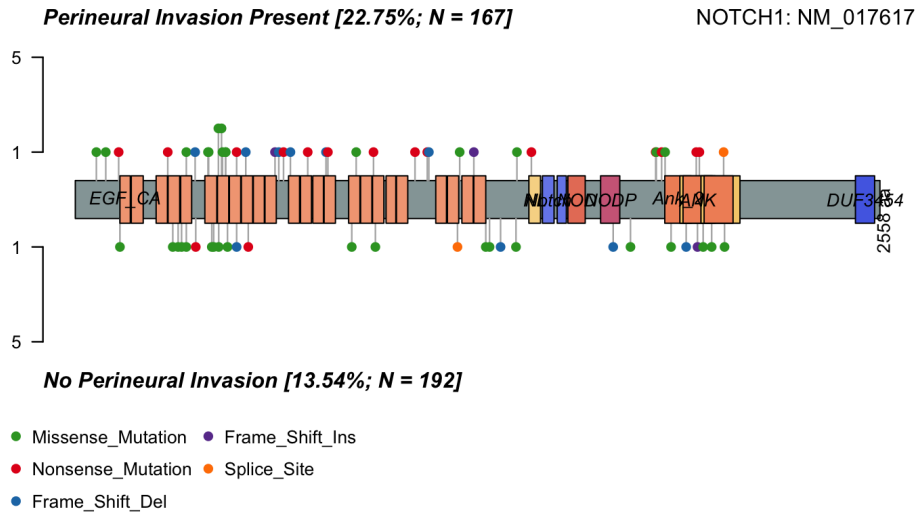

Figure 6b. Co-lollipop plot of NOTCH1 stratified by PNI status shows no marked differences in mutation site or type.

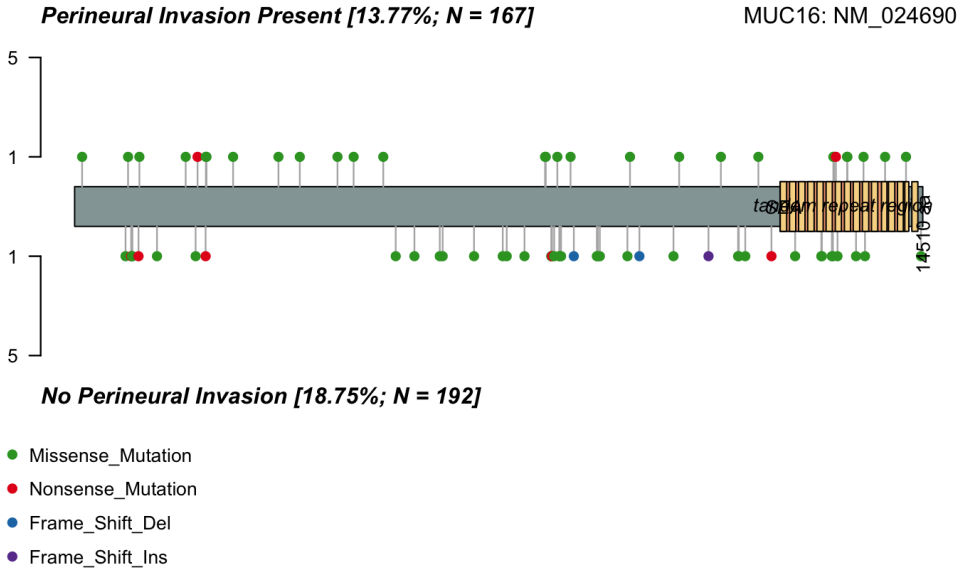

Figure 6c. Co-lollipop plot of MUC16 stratified by PNI status shows slight differences in mutation site as evidenced by the uniform distribution of mutations along PNI-present samples compared to the right-concentrated locality of mutations within PNI-absent samples.

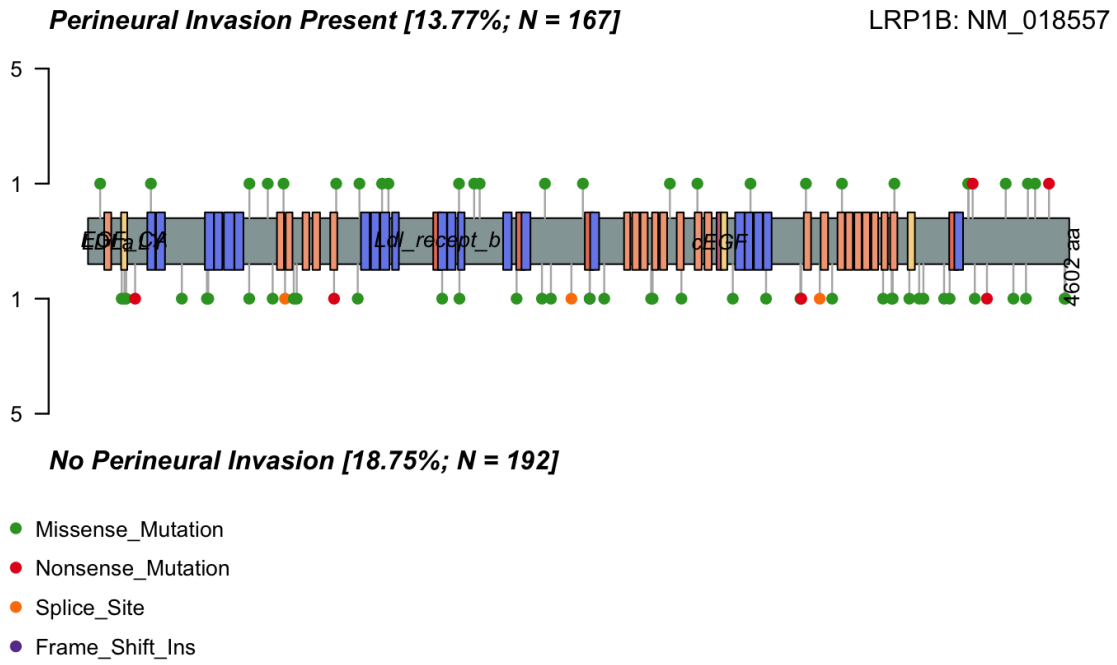

Figure 6d. Co-lollipop plot of LRP1B stratified by PNI status shows no marked differences in mutation site or type.

Construction of somaticMutation plots for PNI-agnostic, PNI-present and PNI-absent patient subsets demonstrates specific mutational co-occurrences that can differentiate PNI status with greater certainty (Figures 7a-c). The PNI-agnostic co-occurrence plot demonstrates a statistically significant mutual exclusivity between PIK3CA and TP53, HUWE1 and TP53 and CASP8 and TP53 (Figure 7a). There are many statistically significant co-incidences among various genes. The PNI-present co-occurrence plot reveals a unique mutually exclusive signature between COL11A1 and TP53 (Figure 7b). All other statistically significant occurrences appear within the PNI-agnostic plot. The PNI-absent sample set plot is identifiable by the absence of any significant mutually exclusive relationships, along with several unique co-incidence relationships (Figure 7c). Overall, PNI status may be determined via mutation data by analyzing the ensemble of genes that mutate together (or the absence thereof).

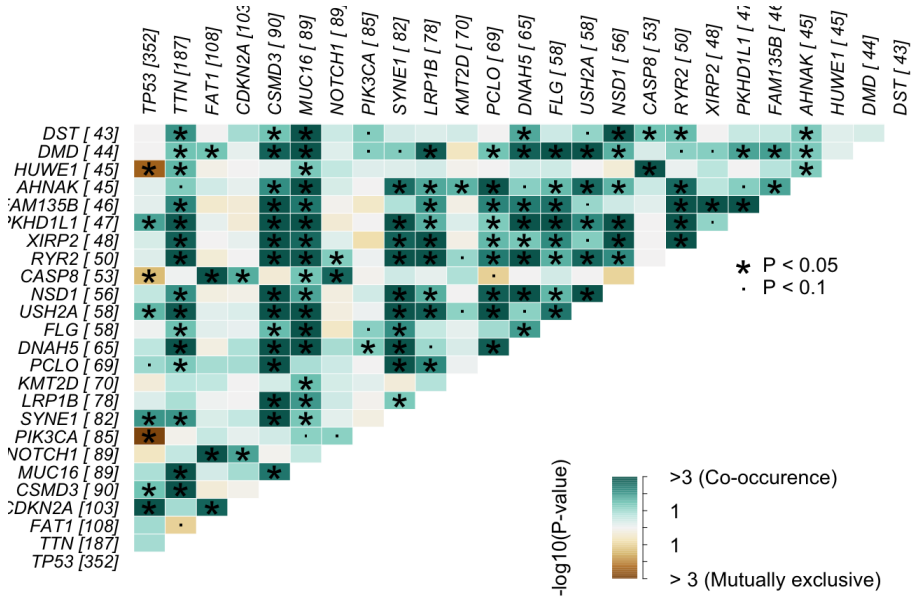

Figure 7a. Co-occurrence plot of HNSCC mutation data NOT stratified by PNI status (generated via somaticMutations) shows a mix of statistically significant co-incidences and mutually exclusive relationships. Analyzing mutual exclusivities can be a strong differentiating factor since they are relatively sparse and variable.

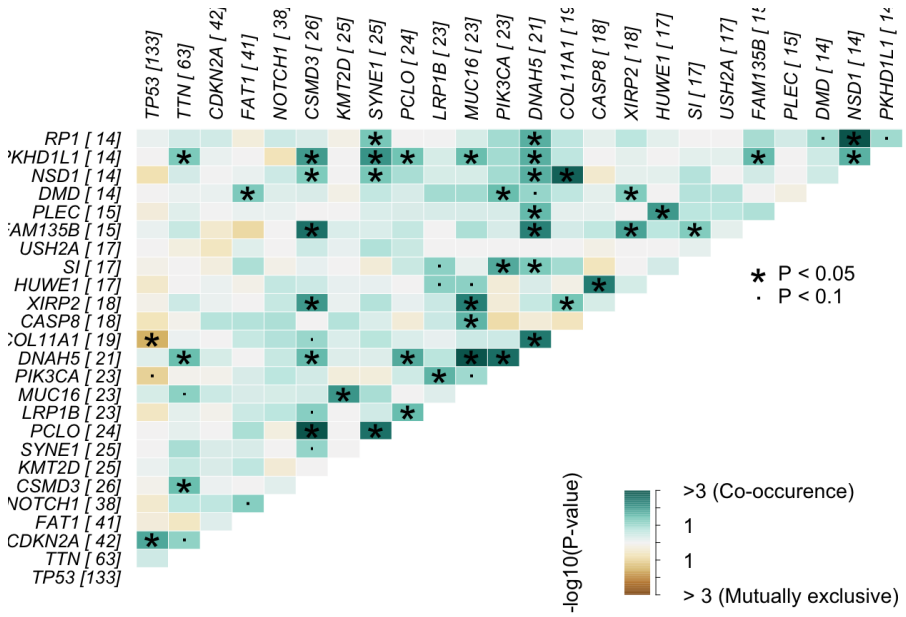

Figure 7b. Co-occurrence plot of HSNCC mutation data stratified by PNI-present status (generated via somaticMutations) shows a signature mutually exclusive relationship between COL11A1 and TP53 not observed in other plots.

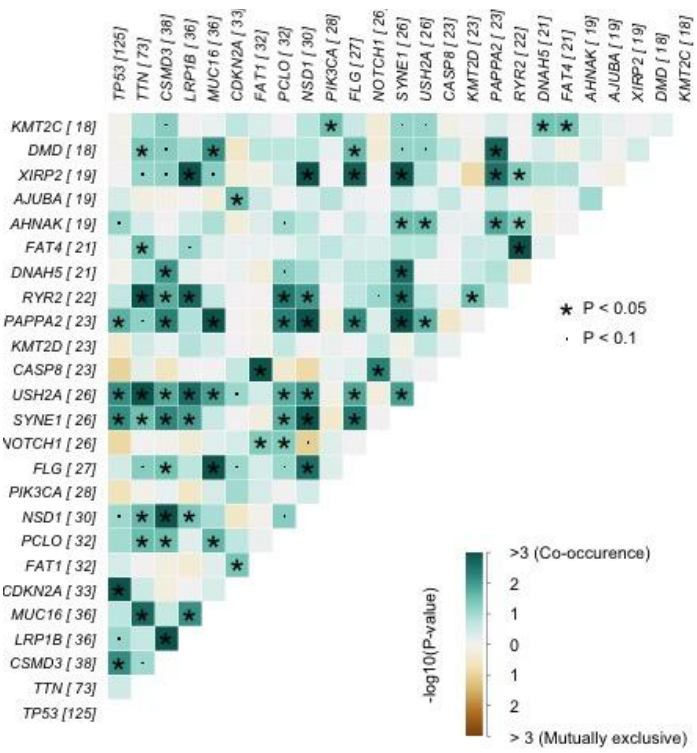

Figure 7c. Co-occurrence plot of HSNCC mutation data stratified by PNI-present status (generated via somaticMutations) shows no signature mutually exclusive relationship among genes, but rather several unique co-incidences along the TP53 column. This absence of mutual exclusivity can potentially serve as a determinant of PNI absence since other plots demonstrate at least one mutually exclusive relationship.

Lastly, comparisons of PNI-agnostic, PNI-present and PNI-absent Kaplan-Meier plots reveals differing responses of HNSCCs to popular chemotherapeutic drugs including Carboplatin, Cisplatin and Paclitaxel (Figures 7a-c). While all plots demonstrate that the three chemotherapy drugs have similar effects on long-term survival, patients with PNI-present HNSCC had relatively better survival rates than PNI-absent HNSCC patients. This is a major finding in terms of elucidating the therapeutic relevance of assessing PNI status of HNSCC samples.

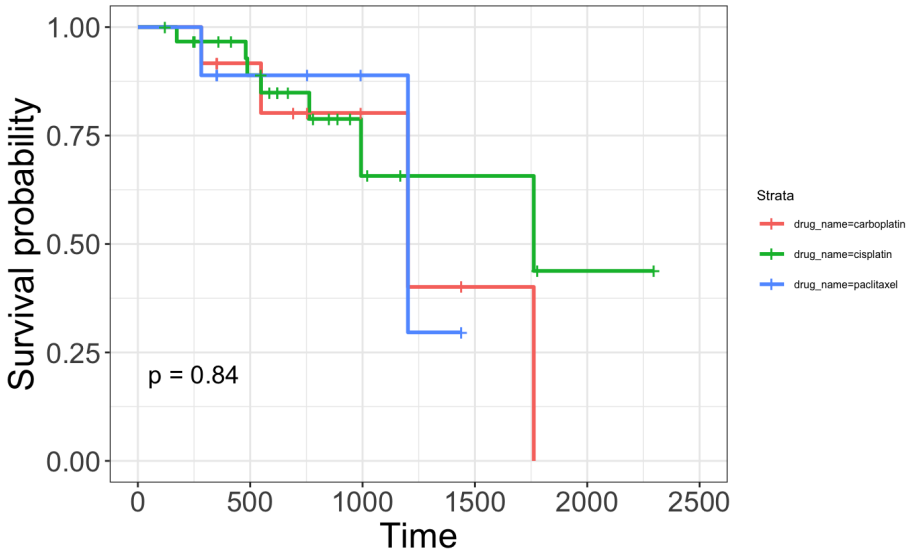

Figure 8a. PNI-agnostic Kaplan-Meier plot demonstrates similar long-term survival rates among HNSCC patients regardless of drug administered. P-value of 0.84 indicates the failure to reject the null hypothesis that all three drugs have a similar impact on patient survival outcomes. Time is represented as the difference between the days to death (or days to last follow up if such data is missing) and the day of diagnosis.

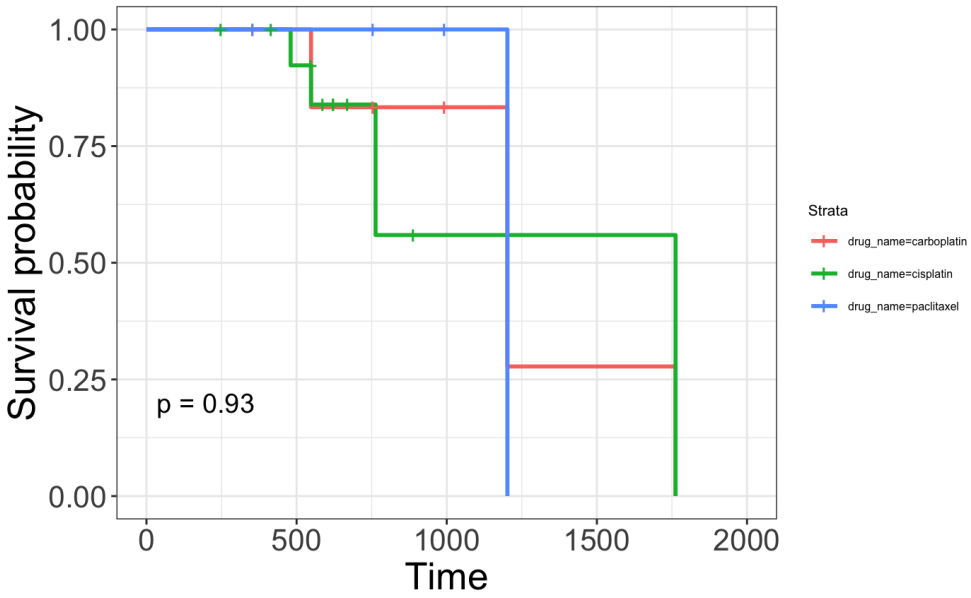

Figure 8b. PNI-absent Kaplan-Meier plot demonstrates relatively poor long-term survival outcomes regardless of drug administered. P-value of 0.93 indicates the failure to reject the null hypothesis that all three

drugs have a similar impact on patient survival outcomes. Time is represented as the difference between the days to death (or days to last follow up if such data is missing) and the day of diagnosis.

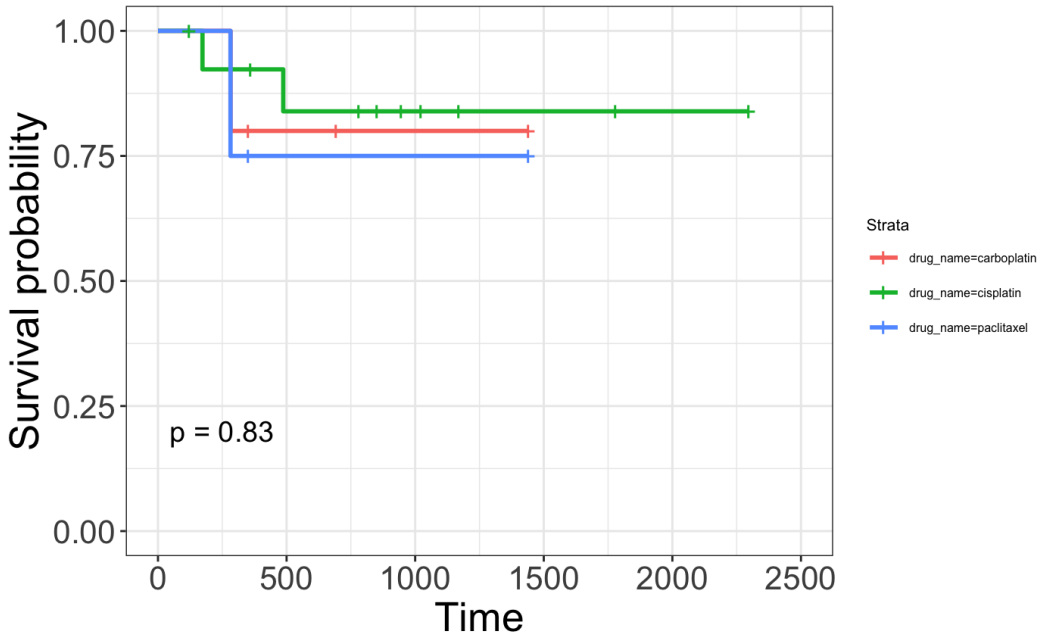

Figure 8c. PNI-present Kaplan-Meier plot shows relatively superior long-term survival outcome regardless of drug administered. P-value of 0.93 indicates the failure to reject the null hypothesis that all three drugs have a similar impact on patient survival outcomes. Time is represented as the difference between the days to death (or days to last follow up if such data is missing) and the day of diagnosis.

IV. Discussion

Our analysis validates perineural invasion status as a clinically relevant pathologic factor for assessing both prognostic and therapeutic outcomes. While PNI has already been distinguished as a factor in determining risk of postoperative recurrence (Saidak, Z. et al., 2018), we can further conclude that it offers an independent indication of prognostic outcome. Our KM survival plot stratified by PNI status shows that PNI-present HSNCC patients experience significantly worse long-term survival rates than PNI-absent patients. This finding corroborates the association of PNI with aggressive forms of other cancers including prostate and colorectal cancers (Saeter, T. et al., 2016). Surprisingly, however, our analysis does not support the general

positive association of PNI with higher grade tumors as observed in prostate, pancreatic and other cancers. In fact, we found that PNI is wholly independent of pathologic stage classification, meaning that the underlying development mechanism of PNI in HNSCC does not rely on the rate at which malignant cells proliferate or what tissue systems have already been occupied. Given the reliance of PNI on continual interactions between malignant cells and a surrounding nerve microenvironment, it may be plausible that PNI can develop even at less advanced stages due to the relative abundance of nerves in the head and neck region (Bakst, R., 2016). There also appeared to be no significant relationship between lymphovascular and perineural invasion as shown in Figure 2, and no such relation has been established in the literature either.

The overall oncoplot of HNSCC compared to the co-oncoplot with patient populations stratified by PNI status demonstrate significant genetic mutation markers that may point us to the biomolecular mechanism behind PNI. The high prevalence of P53 among most cases of HSNCC is supported by Zhou, G. et al., since TP53 serves as a tumor suppressor gene that controls an extensive network of cell proliferation processes. While TP53 doesn't offer much diagnostic or prognostic power given its ubiquity among many cancer types, musculoskeletal-specific genes such as TTN, FAT1 and SYNE1 are better indicators of HNSCCs. In fact, Jiang, A. et al. conducted a similar analysis of HNSCC TCGA data and found that genes such as TTN and FAT1 are somewhat unique to HNSCCs since they encode structural components of the musculoskeletal system. PNI-specific mutational differences shown in the co-oncoplot are also supported by the literature; for instance, the gene NOTCH1 has been shown to be upregulated in salivary adenoid cystic carcinoma cases with PNI due to the gene's promotion of cell invasion and metastasis (Chen, W. et al. 2015).

Comparing the co-occurence plots stratified by PNI status, we observed a peculiar difference in mutually exclusive relationships that may help diagnostically determine PNI status without histological examination. TP53 and COL11A1 are mutually exclusive in PNI-present samples whereas there are no mutually exclusive co-occurrences in PNI-absent samples. Recent work by Suanders, H. et al. shows that COL11A1 may contribute to the immune evasion of malignant cells, allowing for further invasion of surrounding tissues. It is then plausible that the upregulation of COL11A1 (and coincidentally downregulation of TP53) would be an adequate mutually exclusive combination promoting both cell proliferation and invasion. This mutually exclusive relationship has yet to be addressed in literature as of October 2022.

PNI status also serves an important role in the therapeutic realm, as evidenced by our KM plot analyses comparing the three most abundantly utilized chemotherapy drugs. While there was no statistically significant difference in the impact of any of the three drugs on long-term survivorship, we observed a notable discrepancy in chemotherapy response rates between the PNI-present and PNI-absent subsets. It was evident that PNI-present HNSCC patients responded better to chemotherapy treatment and enjoyed enhanced long-term survivorship. This finding is corroborated by literature stating that adjuvant chemotherapy is an effective treatment for only patients with confirmed PNI, with little to no beneficial effects for PNI-absent gastric cancer patients (Tao, Q., 2020).

Overall, our analysis offers strong support in favor of PNI status as both a prognostic determination and a therapeutic efficacy determination tool. We also identified numerous genes implicated in cell adhesion and structure specifically implicated in HNSCCs, which is characteristic of the musculoskeletal properties of the head and neck regions. Our observation of the mutual exclusivity relationships between PNA-present and -absent co-occurrence plots leaves

room for further research into specific genes and biomolecular pathways that ultimately dictate the underlying mechanism behind PNI development. Future investigation of these variably expressed genes can be crucial for honing in on very mechanics of PNI and the growingly important role it plays in morbidity and mortality among many cancer patients.

V. References

1. Bakst, R. L., & Wong, R. J. (2016). Mechanisms of Perineural Invasion. *Journal of neurological surgery. Part B, Skull base*, *77*(2), 96–106. https://doi.org/10.1055/s-0036-1571835
2. Chen, W., Cao, G., Yuan, X., Zhang, X., Zhang, Q., Zhu, Y., Dong, Z., & Zhang, S. (2015). Notch-1 knockdown suppresses proliferation, migration and metastasis of salivary adenoid cystic carcinoma cells. *Journal of translational medicine*, *13*, 167. https://doi.org/10.1186/s12967-015-0520-2
3. Eviston, T.J., Minaei, E., Mueller, S.A. *et al.* Gene expression profiling of perineural invasion in head and neck cutaneous squamous cell carcinoma. *Sci Rep* 11, 13192 (2021). https://doi.org/10.1038/s41598-021-92335-4
4. Jiang, A. M., Ren, M. D., Liu, N., Gao, H., Wang, J. J., Zheng, X. Q., Fu, X., Liang, X., Ruan, Z. P., Tian, T., & Yao, Y. (2021). Tumor Mutation Burden, Immune Cell Infiltration, and Construction of Immune-Related Genes Prognostic Model in Head and Neck Cancer. *International journal of medical sciences*, *18*(1), 226–238. https://doi.org/10.7150/ijms.51064
5. Johnson, D. E., Burtness, B., Leemans, C. R., Lui, V., Bauman, J. E., & Grandis, J. R. (2020). Head and neck squamous cell carcinoma. *Nature reviews. Disease primers*, *6*(1), 92. https://doi.org/10.1038/s41572-020-00224-3
6. Liebig, C., Ayala, G., Wilks, J., Verstovsek, G., Liu, H., Agarwal, N., Berger, D. H., & Albo, D. (2009). Perineural invasion is an independent predictor of outcome in colorectal cancer. *Journal of Clinical Oncology*, *27*(31), 5131–5137. https://doi.org/10.1200/jco.2009.22.4949
7. Saeter, T., Bogaard, M., Vlatkovic, L., Waaler, G., Servoll, E., Nesland, J. M., Axcrona, K., & Axcrona, U. (2015). The relationship between Perineural invasion, tumor grade, reactive stroma and prostate cancer-specific mortality: A clinicopathologic study on a

population-based cohort. *The Prostate*, *76*(2), 207–214. https://doi.org/10.1002/pros.23112

8. Saidak, Z., Clatot, F., Chatelain, D., & Galmiche, A. (2018). A gene expression profile associated with perineural invasion identifies a subset of HNSCC at risk of post-surgical recurrence. *Oral Oncology*, *86*, 53–60. https://doi.org/10.1016/j.oraloncology.2018.09.005
9. Saunders, H. I., Arnold, L., Markiewicz, M., & Thomas, S. M. (2022). Abstract 3844: Col11a1: A driver of tumor progression and immune evasion in head and neck squamous cell carcinoma (HNSCC). *Cancer Research*, *82*(12_Supplement), 3844–3844. https://doi.org/10.1158/1538-7445.am2022-3844
10. Tao, Q., Zhu, W., Zhao, X., Li, M., Shu, Y., Wang, D., & Li, X. (2020). Perineural Invasion and Postoperative Adjuvant Chemotherapy Efficacy in Patients With Gastric Cancer. *Frontiers in oncology*, *10*, 530. https://doi.org/10.3389/fonc.2020.00530
11. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517, 576–582 (2015). https://doi.org/10.1038/nature14129
12. Vokes, E. E., Weichselbaum, R. R., Lippman, S. M., & Hong, W. K. (1993). Head and Neck Cancer. *New England Journal of Medicine*, *328*(3), 184–194. https://doi.org/10.1056/nejm199301213280306
13. Zhang, Z., Liu, R., Jin, R., Fan, Y., Li, T., Shuai, Y., Li, X., Wang, X., & Luo, J. (2019). Integrating clinical and genetic analysis of perineural invasion in head and neck squamous cell carcinoma. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00434

14. Zhou, G., Liu, Z., & Myers, J. N. (2016). TP53 Mutations in Head and Neck Squamous Cell Carcinoma and Their Impact on Disease Progression and Treatment Response. *Journal of cellular biochemistry*, *117*(12), 2682–2692. https://doi.org/10.1002/jcb.25592

Part 3: Review Questions

Attach the answers to the following questions after your paper references.

General Concepts

1. What is TCGA and why is it important?

The Cancer Genome Atlas is a program spearheaded by the National Cancer Institute that hosts many databases containing genomic, transcriptomic, proteomic and clinical data for various cancer types, free of charge and freely accessible to researchers to be downloaded over the web. This is a major step in scientific research because all novice and experienced scientists can obtain high fidelity cancer data and conduct their own analyses to reach new conclusions that may propel cancer biology to new heights.

2. What are some strengths and weaknesses of TCGA?

While TCGA contains highly granular data both in the clinical and molecular realms, it is a significant exercise of trust for both the NCI and patients involved because any breach of the anonymity of the patients who handed over the data will result in disastrous consequences. Patients may become permanently reluctant to hand over their multi-omic data for the sake of science and such high fidelity data may never be available over the internet again. Therefore, it is imperative that TCGA adheres to privacy procedures or it could risk taking a step backwards for all of cancer biology as a field.

Coding Skills

1. What commands are used to save a file to your GitHub repository?

Git add -a
Git commit -m “UPLOADING FILE”
Git push

2. What command(s) must be run in order to use a standard package in R?

Given our standard package is ggplot2 for instance:

```
if (!require("ggplot2", quietly = TRUE))
   install.packages("ggplot2")
   library("ggplot2")
#downloading and importing the package from library
```

3. What command(s) must be run in order to use a Bioconductor package in R?

```
if (!require("BiocManager", quietly = TRUE))
   install.packages("BiocManager") #first install Bioconductor
BiocManager::install(“GenomicRanges”) #install the package of your choice
```

library("GenomicRanges")

4. What is boolean indexing? What are some applications of it?
Boolean indexing is a way that we can segment or subset an existing dataframe to only keep the data that we truly want. For instance, we can boolean index a df in order to remove all NA values in a specific column or obtain rows that meet specific conditions. We can utilize multiple conditional statements within a single boolean index if we wish.

5. Draw a mock up (just a few rows and columns) of a sample dataframe. Show an example of the following and explain what each line of code does. a. an ifelse() statement b. boolean indexing