ELUCIDATING MOLECULAR MECHANISMS OF HPV-POSITIVE HEAD AND NECK SQUAMOUS CELL CARCINOMA

by

Mitzi Boardman

A CAPSTONE

Presented to the Division of Bioinformatics and Computational Biology, Department of Medical Informatics and Clinical Epidemiology, and the Oregon Health and Sciences University School of Medicine in partial fulfillment of the requirements of the degree of

Master of Biomedical Informatics

2018

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Mitzi L. Boardman

"ELUCIDATING MOLECULAR MECHANISMS OF HPV-POSITIVE HEAD AND NECK SQUAMOUS CELL CARCINOMA"

Has been approved

Guanming Wu

TABLE OF CONTENTS

	Page
Abstract	1
Introduction	3
Methods	9
Results	10
Discussion	18
References	21

TABLES AND FIGURES

Page

Figure 1. Gene Set Enrichment Analysis Pipeline 7

Table 1. HPV-positive HNSCC vs HPV-negative HNSCC GSEA 10

Table 2. HPV-positive CSCC vs HPV-negative HNSCC GSEA 12

Figure 2. Overlap between HPV-positive CSCC vs HPV-negative HNSCC and HPV-positive HNSCC vs HPV-negative HNSCC appearing at least one time 13

Figure 3. Overlap between HPV-positive CSCC vs HPV-negative HNSCC and HPV-positive HNSCC vs HPV-negative HNSCC appearing 75% of the time 14

Figure 4. Heatmaps produced by GSEA leading edge analyses 16

Figure 5. Close ups of large clusters within the heatmaps produced by GSEA leading edge analyses 17

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Guanming Wu, Dr. Shannon McWeeney, Dr. Molly Kulesz-Martin, Dr. Eilis Boudreau, Dr. Karen Eden, Gabrielle Choonoo, Fred Loney, Virginia Lankes and Diane Doctor along with all of the DMICE faculty, staff and fellows

ABSTRACT

Head and neck squamous cell carcinoma (HNSCC), the sixth most common malignancy worldwide, is a disease with a profoundly high incidence of long-term morbidity as a result of current treatment protocols. Given the prevalence of this cancer type, reduction in debilitating treatment-related morbidities is of critical importance. HNSCC is comprised of two biologically distinct tumor types: Those caused by environmental factors such as smoking and alcohol consumption, and those caused by human papilloma virus (HPV). The presence of HPV in HNSCC is known to improve patient prognosis as a result of increased sensitivity to treatment; the significance of this lies in the potential to develop less aggressive treatment protocols for HPV-driven tumors, and thus, fewer long-term morbidities. In order to develop novel treatment protocols specific for HPV-positive HNSCC, it is critical that we can reliably differentiate between HPV-positive and HPV-negative tumors and understand the mechanistic differences. Ninety-five percent of all cases of cervical cancer are driven by HPV, and it has been suggested that a molecular comparison between cervical squamous cell carcinoma (CSCC) and HPV-positive HNSCC may provide the most substantial evidence in support of HPV-positive status in HNSCC. The purpose of this study was to conduct a comparison pathway enrichment analysis using the Gene Set Enrichment Analysis (GSEA) approach between three cancer cohorts: HPV-positive HNSCC, HPV-negative HNSCC, and HPV-positive cervical squamous cell carcinoma (CSCC). Comparing to HPV-negative HNSCC tumors, we found that 26 and 24 pathways were enriched with upregulated genes in the HPV-positive HNSCC and CSCC, respectively. Nineteen of these enriched pathways were overlapping and the majority were annotated for DNA repair. Intriguingly, no significant pathway was enriched when we compared HPV-positive HNSCC and

CSCC. These results suggest that we may be able to distinguish HPV-positive HNSCC from HPV-negative HNSCC at a cellular pathway level based on gene expression data, and that HPV-positive SCCs are more similar to each other than they are to HPV-negative HNSCC despite the distant anatomical location. We also conducted a leading edge analysis based on our GSEA experiments and found that each of the groups contained a similar cluster of genes that contributed to the enrichment of "DNA Repair" pathways. These genes may be of significance when we are thinking about targeted therapies specifically for HPV-related tumors.

Keywords: Head and Neck Squamous Cell Carcinoma (HNSCC), Human Papilloma Virus (HPV), Cervical Squamous Cell Carcinoma (CSCC), The Cancer Genome Atlas (TCGA), Gene Set Enrichment Analysis (GSEA)

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is a common malignancy with more than 500,000 new cases diagnosed worldwide each year; there are 40,000 new cases with 7890 deaths within the United States each year (Siegel, Miller, & Jemal, 2015; Torre et al., 2015). HNSCC often results in life-long morbidities such as osteoradionecrosis, trismus, xerostomia, swallowing difficulties, and radiation caries as result of current treatment protocols (Lalla et al., 2017). As well as exhibiting tremendous heterogeneity in terms of location within the upper aerodigestive tract, HNSCC is comprised of two distinct tumor types: those caused by environmental factors such as smoking and alcohol consumption, and those driven by human papilloma virus (HPV) (Mork et al., 2001; Slebos et al., 2006).

HPV was presented as a potential etiology in HNSCC by Gillison et al., 1999 (Gillison, Koch, & Shah, 1999). There were many factors supporting this theory. A 1996 study by Snijders et al. reported that 21% of 63 HNSCCs they examined contained HPV DNA, and the most common viral subtype found was HPV-16 which had previously been identified as a high-risk subtype for the development of cervical cancer (Snijders et al., 1996). It was also noted that HPV-associated HNSCC was site specific with viral DNA frequently found in the tonsillar region (Paz, Cook, Odom-Maryon, Xie, & Wilczynski, 1997). HPV DNA positive tonsillar tumors (DNA *in situ* hybridization) were found to express HPV viral oncoproteins E6 and E7 and were associated with poorly keratinized tumors whereas tumors that lacked HPV DNA were well keratinized

(Wilczynski, Lin, Xie, & Paz, 1998). Further evidence in support of HPV as an etiologic factor in a subgroup of HNSCC was the presence of E6 and E7 viral oncoprotein transcripts. E6 and E7 had been shown to inactivate the tumor suppressor gene products, p53 and pRb, respectively, which led to disruption in cell cycle control and genomic instability (Galloway & McDougall, 1996). They were also found to be consistently expressed and required for malignant transformation and maintenance of the transformed phenotype in cervical cancer specimens, and were shown to be mutagenic in normal oral keratinocytes (Crook, Morgenstern, Crawford, & Banks, 1989; Liu, Han, Baluda, & Park, 1997). Further evidence to support a mechanistic difference between HPV-related and non-HPV-related HNSCC was the finding that cyclin D1, a cell cycle regulatory protein that interacts with pRb, was overexpressed tumors that were free of HPV but was not present in tumors in which HPV was detected (Wilczynski et al., 1998).

The presence of HPV in HNSCC improves patient prognosis and survival as a result of increased sensitivity to treatment (Ang et al., 2010). The importance of this distinction is amplified by the rapidly changing landscape of HNSCC. From 1988 to 2004 there was a 225% increase in tumors containing HPV, a 50% decrease in HPV-negative tumor, as well as a demographic shift toward younger, non-smoking men who have a history of multiple sexual partners and higher socioeconomic status (Chaturvedi et al., 2011; Marur & Forastiere, 2016). Historically, HNSCC was a disease found mostly in men over 60 with a history of alcohol and tobacco abuse. Concordant with the increase in HPV-positive tumors, there has been a shift in tumor location from predominantly hypopharynx and larynx to oropharynx (Marur & Forastiere, 2016).

Understanding the differences and being able to distinguish between HPV-positive and HPV-negative tumors is essential while investigators are racing to develop new treatments which will

be more effective and result in fewer devastating long-term effects. Work has been done to distinguish HPV-positive tumors from HPV-negative tumors clinically and biologically, as well as through genomic analysis, transcriptomic analysis, and epigenomic analysis (Lechner & Fenton, 2016; Marur & Forastiere, 2016). However, there has not been a definitive molecular description of what actually constitutes a tumor that is HPV-positive vs one that is HPV-negative. In addition, it has been suggested that a molecular comparison between cervical squamous cell carcinoma (CSCC) and HPV-positive HNSCC may provide the most substantial evidence in support of HPV-positive status in HNSCC (Holzinger et al., 2017).

The Cancer Genome Atlas (TCGA) has provided a platform by which the research community can rapidly answer molecular questions related to cancer (TCGA, 2017 [cited 5 September 2017]). Through its genomic data analysis pipeline, it can "effectively collect, select, and analyze human tissues for genomic alterations on a very large scale" . The Head and Neck Squamous Cell Carcinoma project within TCGA contains 528 individual cases, with most having extensive mutation data, RNA-Seq data, and clinical data. We combined TCGA's provisional HPV annotation and its more detailed 2015 annotation of 279 cases, along with a more rigorous annotation using TCGA's WGS and RNA-Seq data provided by Nulton et al. (Nulton, Olex, Dozmorov, Morgan, & Windle, 2017), to create a list of cases which we consider to be "very high confidence" for HPV as the tumorigenic driving force (HPV-driven) and a group of "very high confidence" HPV-negative tumors. We have also created a list of 84 HPV-positive cases from the TCGA HPV-positive cervical squamous cell cohort (CSCC) (TCGA, 2017). Because HPV16 is by far the most prevalent type of HPV found in the TCGA cohorts, we selected only HPV16-positive tumors in order to increase homogeneity within our populations.

In our effort to understand the molecular differences between HPV-positive and HPV-negative tumors, we performed comparison pathway enrichment analysis using the Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) approach (Figure 1) between our three well-defined cancer cohorts: HPV-positive HNSCC, HPV-negative HNSCC, and HPV-positive cervical squamous cell carcinoma (CSCC). GSEA can be used to differentiate phenotypes on a molecular level, and focuses on groups of genes that share a common biological function (gene sets). It provides a powerful approach to analyzing gene expression data, which can provide insight into common biological functions, in contrast to single gene analysis, which offers little functional information. We used gene sets from the Reactome database (Croft et al., 2014; Fabregat et al., 2018), and RNA-Seq data from the TCGA cohorts described above. With this analysis we can compare and contrast expression level differences across gene sets between HPV-positive HNSCC and HPV-negative HNSCC, which may elucidate mechanistic differences between the two tumor types. We can also make a comparison between HPV-positive CSCC and HPV-negative HNSCC in order to understand whether or not HPV-viral-related tumors of the same tissue type (stratified squamous epithelium), but a vastly different location, will yield the same mechanistic result. Finally, we can use GSEA to compare both HPV-positive tumor types (HNSCC and CSCC) which may provide insight into how similar these two tumor types are given their disparate anatomical locations. With this information we hope to gain knowledge regarding the mechanistic differences between HPV-positive HNSCC and HPV-negative HNSCC, which may eventually be applied to developing novel treatments for HPV-positive HNSCC, decreasing the devastating morbidities related to the treatment of these tumors.

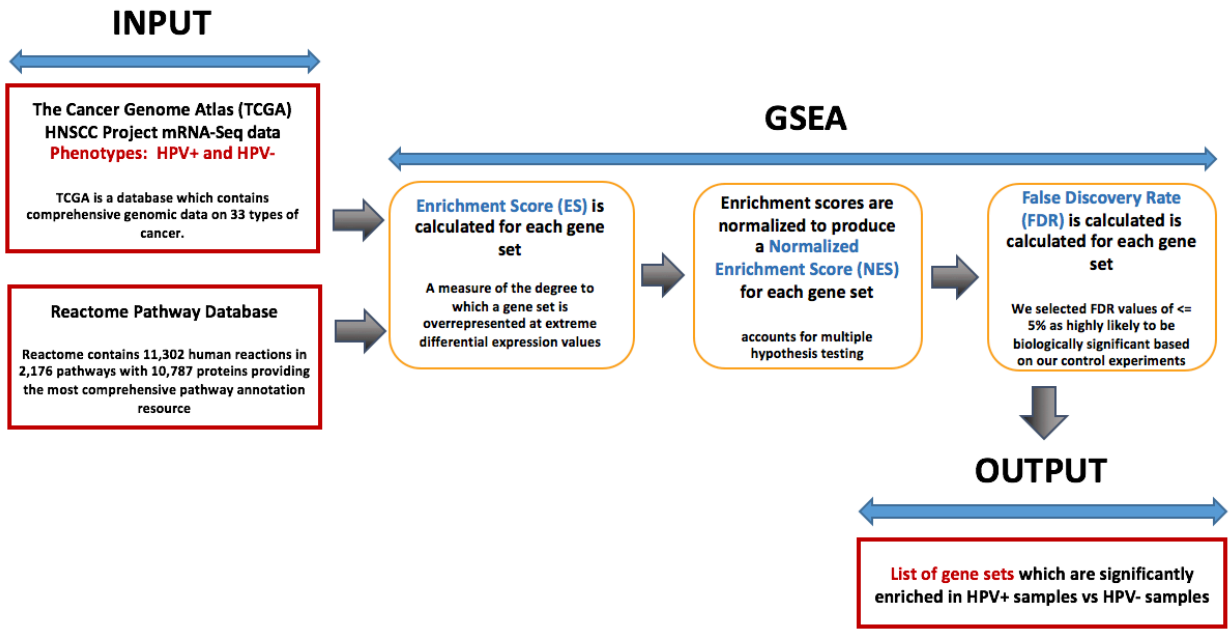


Figure 1. Gene Set Enrichment Analysis Pipeline. GSEA requires expression data representing two phenotypes, and a gene set database as input. In this example the input is RNA-Seq data from TCGA HNSCC HPV+ and HPV- phenotypes. The gene set database is a comprehensive list of Reactome pathways with their corresponding genes. The first step of GSEA is calculation of an enrichment score for each gene set, which is a measure of the degree to which a gene set is overrepresented at extreme differential expression values. Second, the enrichment scores are normalized for each gene set to account for multiple hypothesis testing. Finally, a false discovery rate is calculated for each gene set. We set a cutoff of FDR <= 0.05 to represent those gene sets which are highly likely to be biologically significant based on our

control experiments. The output of GSEA is a list of gene sets which are significantly enriched with differentially expressed genes in our chosen phenotype, HPV+ in this example.

METHODS

DATA ACQUISITION

HPV annotation data for TCGA HNSCC cases was collected from three sources: 1) TCGA clinical data from the HNSCC project (downloaded from cBioportal Head and Neck Provisional) (Gao et al., 2013; TCGA, 2015), 2) TCGA HNSCC HPV annotation data from TCGA 2015 Nature publication (TCGA, 2015), and 3) TCGA HNSCC HPV annotation from Nulton et al., 2017 (Nulton et al., 2017). HPV annotation data for TCGA cervical cancer cases was collected from two sources: 1) TCGA clinical data from the cervical cancer project (downloaded from cBioportal) (Gao et al., 2013), and 2) HPV annotation data from supplemental table 3, TCGA cervical cancer publication 2017 (TCGA, 2017). TCGA RNA-Seq data for both HNSCC and cervical cancer cohorts was downloaded from Broad GDAC Firehose (doi:10.7908/C11G0KM9). The gene set database we selected was the "Reactome Pathway Gene Set" (https://reactome.org/download/current/ReactomePathways.gmt.zip, downloaded September 20, 2017) from Reactome.org for use in GSEA.

HPV ANNOTATION

For the HNSCC cohort, cases were annotated as HPV-positive if they were found in the Nulton et al. (Nulton et al., 2017) study to contain HPV E6 and E7 transcripts, and had an HPV-positive annotation in either the TCGA provisional data or the TCGA, 2015, Nature study (TCGA, 2015). Cases were annotated as HPV-negative if whole genome sequencing data was available and found to be HPV-negative in the Nulton et al. study. For the cervical cohort, the cases annotated as HPV- CSCC were selected from the TCGA clinical annotation and the TCGA,

2017, Nature comprehensive evaluation of cervical cancer (TCGA, 2017). In order to increase homogeneity in our study, we selected only HPV-positive HNSCC and HPV-positive CSCC cases annotated as HPV type 16 because it is, by far, the most common variant in both cohorts. This left us with 52 HPV-positive HNSCC, 114 HPV-negative HNSCC, and 84 HPV-positive CSCC.

GENE SET ENRICHMENT ANALYSIS

Gene set enrichment analysis (GSEA) was performed using the Python module GSEApy (https://github.com/BioNinja/GSEApy) (Subramanian et al., 2005). GSEA formatted Phenotype files and RNA-Seq files were created using the R statistical programming language. Control GSEA experiments using HPV-positive HNSCC vs HPV-positive HNSCC resulted in no pathways with an FDR <= 0.05, therefore we used FDR <= 0.05 as our cutoff for significance. We ran three experiments: 1) HPV-positive HNSCC vs HPV-negative HNSCC, 2) HPV-positive CSCC vs HPV-negative HNSCC, and 3) HPV-positive CSCC vs HPV-positive HNSCC. For each experiment, GSEA was run 100 times with 1000 permutations per run. We considered enriched pathways that appeared in 75% of the runs with FDR <= 0.05 to be significantly enriched pathways.

LEADING EDGE ANALYSIS

The GSEA Java desktop application allows "leading edge analysis" to be performed on any GSEA results that have been obtained. The leading edge analysis outputs information about the differential expression levels and contributions to the resulting enriched pathways. We performed leading edge analysis on the HPV-positive HNSCC vs HPV-negative HNSCC and

HPV-positive CSCC vs HPV-negative HSNCC GSEA experiments from one run through the desktop application

RESULTS

HPV-POSITIVE HNSCC VS HPV-NEGATIVE HNSCC GSEA

When we ran our GSEA pipeline using the HPV-positive HNSCC cohort against the HPV-negative HNSCC cohort, 46 pathways with an FDR <= 0.05 appeared at least one time. When we selected for pathways that showed up in 75% of the runs at an FDR <= 0.05, we found 26 Reactome pathways enriched with differentially expressed genes (Table 1). Twenty of these pathways fell under the Reactome topic, "DNA Repair". The remaining 6 fell under "Cell Cycle", "Cellular Response to External Stimuli", and/or "DNA Replication".

Table 1. HPV-positive HNSCC vs HPV-negative HNSCC GSEA. Pathways enriched with differentially expressed genes and their corresponding Reactome pathway topics. Seventeen of these pathways fell under the Reactome category "DNA Repair". The remaining 7 fell under "Cell Cycle", "Cellular Response to External Stimuli", and/or "Reproduction".

PATHWAY NAME	REACTOME TOPIC	% APPEARED
1. Fanconi Anemia Pathway	DNA Repair	100
2. Recognition of DNA damage by PCNA-containing replication complex	DNA Repair	100
3. Translesion synthesis by POLK	DNA Repair	100
4. DNA Damage Bypass	DNA Repair	100
5. Resolution of Abasic Sites (AP sites)	DNA Repair	100
6. Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template	DNA Repair	100
7. Base Excision Repair	DNA Repair	100
8. Gap-filling DNA repair synthesis and ligation in GG-NER	DNA Repair	100
9. Translesion synthesis by REV1	DNA Repair	100
10. Oncogene Induced Senescence	Cellular Response to External Stimuli	100

11. PCNA-Dependent Long Patch Base Excision Repair	DNA Repair	99
12. Translesion synthesis by POLI	DNA Repair	99
13. Resolution of AP sites via the multiple-nucleotide patch replacement pathway	DNA Repair	99
14. Translesion Synthesis by POLH	DNA Repair	99
15. HDR through Homologous Recombination (HRR)	DNA Repair	99
16. Termination of translesion DNA synthesis	DNA Repair	99
17. Mismatch Repair	DNA Repair	98
18. Dual Incision in GG-NER	DNA Repair	98
19. Telomere C-strand (Lagging Strand) Synthesis	Cell Cycle	95
20. HDR through Single Strand Annealing (SSA)	DNA Repair	92
21. Lagging Strand Synthesis	Cell Cycle DNA Replication	92
22. Extension of Telomeres	Cell Cycle	84
23. E2F mediated regulation of DNA replication	Cell Cycle	82
24. Homology Directed Repair	DNA Repair	81
25. DNA strand elongation	Cell Cycle DNA Replication	79
26. DNA Double-Strand Break Repair	DNA Repair	78

HPV-POSITIVE CSCC VS HPV-NEGATIVE HNSCC GSEA

When we ran our GSEA pipeline using the HPV-positive CSCC cohort against the HPV-negative HNSCC cohort, 58 pathways with an FDR <= 0.05 appeared at least one time. Significantly, 40 of these pathways overlapped with pathways that appeared in the HPV-positive HNSCC vs HPV-negative experiment with an FDR <= 0.05 at least one time (Figure 2, Fisher's Exact Test, $p < 0.01$). When we selected for pathways that showed up in 75% of the runs at an FDR <= 0.05, we found 24 pathways enriched with differentially expressed genes (table 2). Seventeen of these pathways fell under the Reactome category "DNA Repair". The remaining 7 fell under "Cell Cycle", "Cellular Response to External Stimuli", and/or "Reproduction". A total of 58 pathways showed up at least one time at an FDR <= 0.05. Also of significance, 19 of these pathways overlapped with pathways that appeared in the HPV-positive HNSCC vs HPV-negative experiment 75% of the time with an FDR <= 0.05 (Figure 3, Fisher's Exact Test, $p < 0.01$).

Table 2. HPV-positive CSCC vs HPV-negative HNSCC GSEA. Pathways enriched with differentially expressed genes and their corresponding Reactome pathway categories. Twenty of these pathways fell under the Reactome category "DNA Repair". The remaining 6 fell under "Cell Cycle", "Cellular Response to External Stimuli", and/or "DNA Replication".

PATHWAY NAME	REACTOME CATEGORY	% APPEARED
1. Fanconi Anemia Pathway	DNA Repair	85
2. Base Excision Repair	DNA Repair	85
3. Resolution of Abasic Sites (AP sites)	DNA Repair	85
4. Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template	DNA Repair	85
5. Meiosis	Cell Cycle Reproduction	84
6. Recognition of DNA damage by PCNA-containing replication complex	DNA Repair	83
7. Resolution of AP sites via the multiple-nucleotide patch replacement pathway	DNA Repair	83
8. DNA Damage Bypass	DNA Repair	82
9. Homology Directed Repair	DNA Repair	81
10. PCNA-Dependent Long Patch Base Excision Repair	DNA Repair	80
11. DNA Repair	DNA Repair	80
12. HDR through Homologous Recombination (HRR)	DNA Repair	80
13. Extension of Telomeres	Cell Cycle	79
14. Termination of translesion DNA synthesis	DNA Repair	79
15. HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)	DNA Repair	77
16. Chromosome Maintenance	Cell Cycle	77
17. Meiotic synapsis	Cell Cycle Reproduction	77
18. Lagging Strand Synthesis	Cell Cycle DNA Replication	77
19. DNA Double-Strand Break Repair	DNA Repair	76
20. Telomere C-strand (Lagging Strand) Synthesis	Cell Cycle	76
21. Gap-filling DNA repair synthesis and ligation in GG-NER	DNA Repair	75
22. E2F mediated regulation of DNA replication	Cell Cycle	75
23. Translesion synthesis by POLK	DNA Repair	75
24. Mismatch Repair	DNA Repair	75

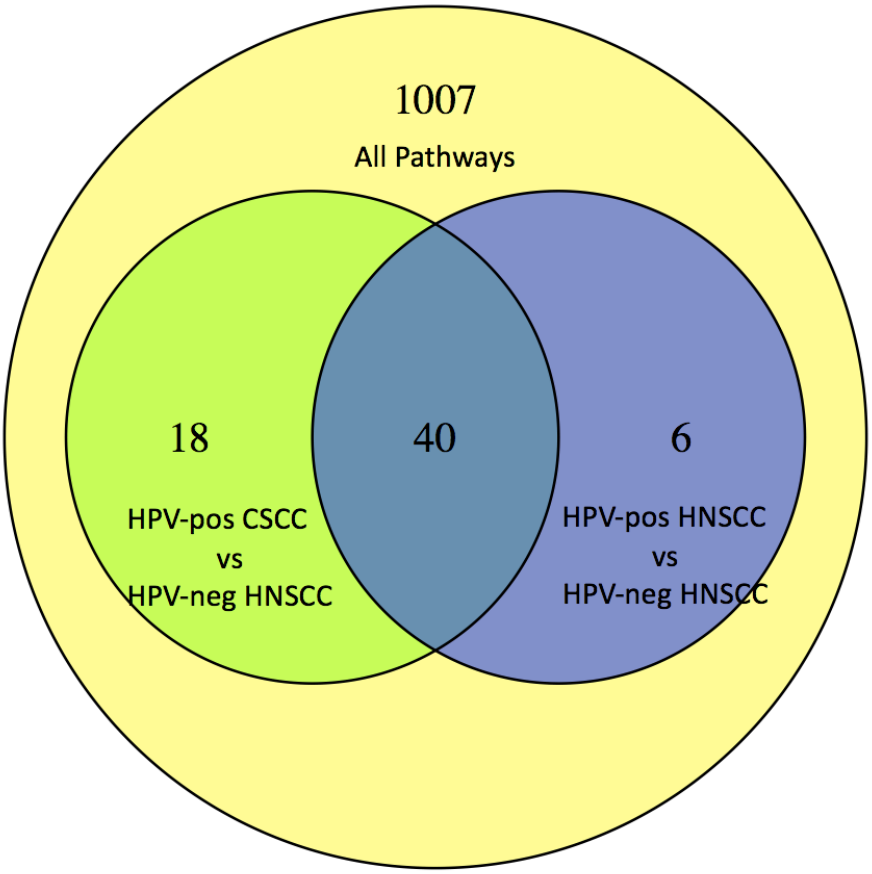


Figure 2. Overlap between HPV-positive CSCC vs HPV-negative HNSCC and HPV-positive HNSCC vs HPV-negative HNSCC appearing at least one time. Of the 1071 Reactome pathways included in the GSEA experiments, when we ran our GSEA pipeline using the HPV-positive CSCC cohort against the HPV-negative HNSCC cohort, 58 pathways with an FDR <= 0.05 appeared at least one time. Significantly, 40 of these pathways overlapped with pathways that appeared in the HPV-positive HNSCC vs HPV-negative experiment with an FDR <= 0.05 at least one time (Figure 2, Fisher's Exact Test, $p < 0.01$).

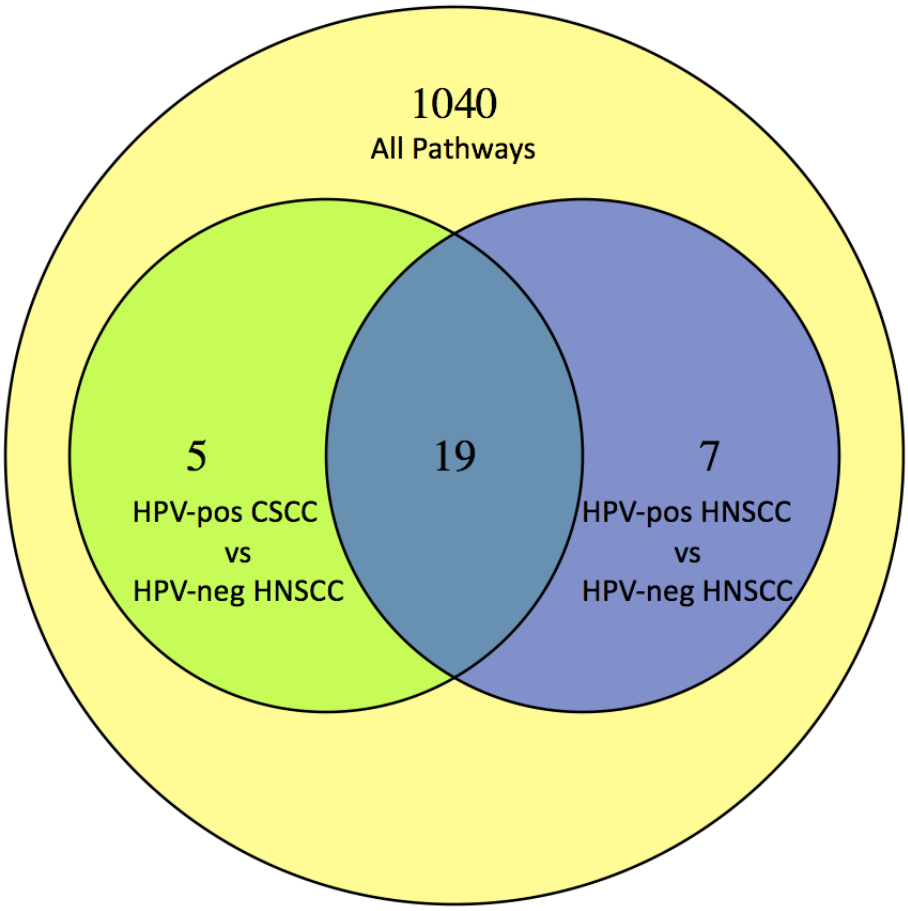


Figure 3. Overlap between HPV-positive CSCC vs HPV-negative HNSCC and HPV-positive HNSCC vs HPV-negative HNSCC appearing 75% of the time. Of the 1071 Reactome pathways included in the GSEA experiments, when we ran our GSEA pipeline using the HPV-positive CSCC cohort against the HPV-negative HNSCC cohort, of significance, 19 of these pathways overlapped with pathways that appeared in the HPV-positive HNSCC vs HPV-negative experiment 75% of the time with an FDR <= 0.05 (Figure 3, Fisher's Exact Test, $p < 0.01$).

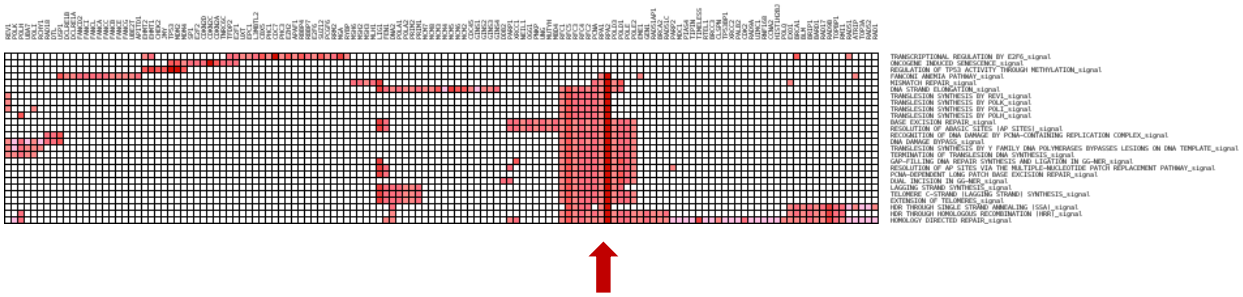
HPV-POSITIVE CSCC VS HPV-POSITIVE HNSCC GSEA

Notably, when we ran our GSEA pipeline using the HPV-positive HNSCC cohort against the HPV-negative HNSCC cohort, we found no Reactome pathways were enriched with differentially expressed genes in any of the 100 runs.

LEADING EDGE ANALYSES

The leading edge analysis on the HPV-positive HNSCC vs HPV-negative HNSCC and HPV-positive CSCC vs HPV-negative HSNCC GSEA experiments from one run through the desktop application output clustered heat maps displaying genes that contributed to the enriched pathways in the HPV-positive cohorts. For both HPV-positive HNSCC and HPV-positive CSCC vs HPV-negative HNSCC all contributing genes were upregulated, suggesting that the pathways that were enriched in these cohorts were enriched with upregulated genes rather than downregulated. Because most of the resulting enriched pathways in both cohorts were under the Reactome topic "DNA Repair", we can deduce that these "DNA Repair" pathways are upregulated in the HPV-positive groups. We also noted that each leading edge heatmap contained a distinct large cluster (Figure 4). Looking closely at the large clusters from the two different experiments, we noticed that many of the same genes are contained within both clusters and are significant contributors to the "DNA Repair" pathway enrichment that we found in our GSEA experiments. These genes include RPA2 (the most upregulated gene in both groups), RFC5, RFC4, RFC3, RFC2, PCNA, POLD3, POLD1, POLE, and POLE2. A close up view of these clusters and their related genes can be seen in Figure 5.

(a)



(b)

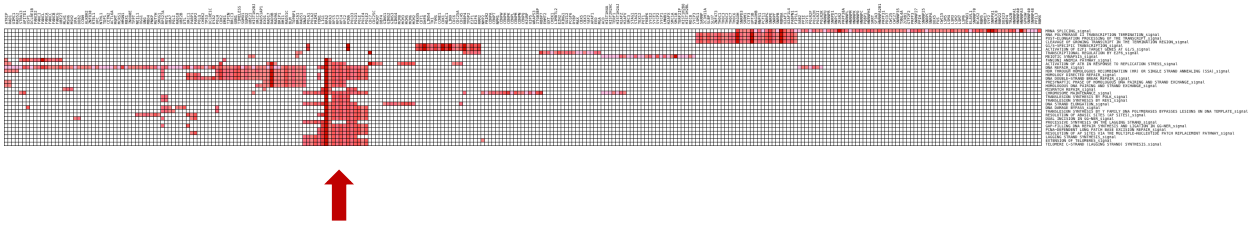


Figure 4. Heatmaps produced by GSEA leading edge analyses. Heatmaps produced from leading edge analyses based on one GSEA run each of HPV-positive HNSCC vs HPV-negative HNSCC (a) and HPV- positive CSCC vs HPV-negative HSNCC (b). Differentially expressed genes that contribute to enriched pathways are along the X-axis. Pathways are listed along the Y-axis. Each heatmap contains a large, distinct cluster (red arrows).

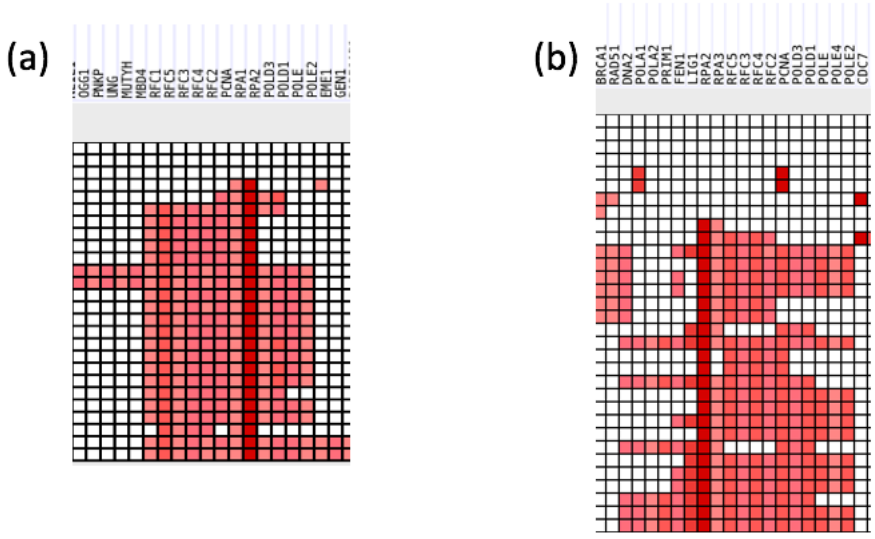


Figure 5. Close ups of large clusters within the heatmaps produced by GSEA leading edge analyses. Close up views of large distinct clusters from the heatmaps produced by leading edge analyses based on one GSEA run each of HPV-positive HNSCC vs HPV-negative HNSCC (a) and HPV- positive CSCC vs HPV-negative HSNCC (b). Overlapping genes between these two clusters include RPA2 (the most upregulated gene in both groups), RFC5, RFC4, RFC3, RFC2, PCNA, POLD3, POLD1, POLE, and POLE2.

DISCUSSION

Notable findings from these GSEA experiments are that HPV-positive HNSCC and HPV-positive CSCC show no differential pathway enrichment between them, that the enriched pathways in the HPV-positive HNSCC vs HPV-negative HNSCC and HPV-positive CSCC vs HPV-negative HNSCC show significant overlap, and that those enriched pathways mainly fall under the Reactome pathway topic "DNA Repair". This upregulation in DNA repair pathways is concordant with the literature which suggests that HPV replication requires the employment of host DNA repair molecules (Bristol, Das, & Morgan, 2017). These findings are significant because they suggest that the HPV-positive tumors are molecularly similar to each other while being distinct from HPV-negative HNSCC by virtue of upregulation of DNA repair pathways. This points to the possibility of classifying and treating tumors based on tissue type and etiology rather than location. It also suggests the possibility that dysregulation of DNA repair pathways may contribute to HPV-positive HNSCC increased sensitivity to treatment (i.e. if DNA repair molecules are tied up in viral replication, they may not be available to repair DNA damage introduced during radiation therapy).

Further notable findings came out of our leading edge analysis. For both HPV-positive HNSCC and HPV-positive CSCC vs HPV-negative HNSCC we found that all genes contributing to our enriched pathways were upregulated. Again, confirming that the DNA repair pathways were upregulated in our HPV-positive cohorts. We also noted above that each leading edge heatmap contained a distinct large cluster, and the clusters from the two different experiments contained

many of the same genes and are significant contributors to the "DNA Repair" pathway enrichment that we found in our GSEA experiments. These genes, as stated in our results, include RPA2 (the most upregulated gene in both groups), RFC5, RFC4, RFC3, RFC2, PCNA, POLD3, POLD1, POLE, and POLE2. Because these genes are upregulated specifically in HPV-positive SCCs vs HPV-negative HNSCC, it is tempting to think about targeted therapy. Mutations in one of these genes, POLE, have been linked to effectiveness of Pembrolizumab which is FDA approved for the treatment of head and neck cancer (Mehnert et al., 2016). An indicator for effectiveness of Pembrolizumab is HPV-positive status (Saleh, Eid, Haddad, Khalife-Saleh, & Kourie, 2018). In addition, RPA2 has been identified as a potential cancer drug target (Byrne & Oakley, 2018). This is encouraging with respect to using the protocol described in this capstone to identify potential drug targets for laboratory study of HPV-positive tumors.

This work has demonstrated that HPV-positive and HPV-negative HNSCC are molecularly distinct. It is paramount to be able to differentiate between these two tumor types to understand their molecular distinction when making choices to enlist new or less aggressive treatment protocols. In addition, our use of GSEA to explore whether HPV-positive HNSCC is more similar to HPV-positive CSCC than to HPV-negative HNSCC suggested that this is the case, given that we saw statistically significant overlap of enriched pathways with HPV-positive HSNCC and HPV-positive CSCC when compared to HPV-negative HNSCC, and no enriched pathways when the HPV-positive tumors were compared with each other. This implies that tumor etiology might be more important than tumor location when thinking about tailored treatments for cancer. We also identified a set of genes that contribute to the enrichment of DNA

repair pathways in our cohorts; this information can be used to further understand the mechanism by which HPV enlists the host cell DNA repair molecules as well as providing insight into potential drug targets for treatment.

REFERENCES

Ang, K. K., Harris, J., Wheeler, R., Weber, R., Rosenthal, D. I., Nguyen-Tan, P. F., . . . Gillison, M. L. (2010). Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med, 363*(1), 24-35. doi:10.1056/NEJMoa0912217

Bristol, M. L., Das, D., & Morgan, I. M. (2017). Why Human Papillomaviruses Activate the DNA Damage Response (DDR) and How Cellular and Viral Replication Persists in the Presence of DDR Signaling. *Viruses, 9*(10). doi:10.3390/v9100268

Byrne, B. M., & Oakley, G. G. (2018). Replication protein A, the laxative that keeps DNA regular: The importance of RPA phosphorylation in maintaining genome stability. *Semin Cell Dev Biol*. doi:10.1016/j.semcdb.2018.04.005

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., . . . Gillison, M. L. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol, 29*(32), 4294-4301. doi:10.1200/jco.2011.36.4596

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., . . . D'Eustachio, P. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res, 42*(Database issue), D472-477. doi:10.1093/nar/gkt1102

Crook, T., Morgenstern, J. P., Crawford, L., & Banks, L. (1989). Continued expression of HPV-16 E7 protein is required for maintenance of the transformed phenotype of cells co-transformed by HPV-16 plus EJ-ras. *Embo j, 8*(2), 513-519.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., . . . D'Eustachio, P. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res, 46*(D1), D649-d655. doi:10.1093/nar/gkx1132

Galloway, D. A., & McDougall, J. K. (1996). The disruption of cell cycle checkpoints by papillomavirus oncoproteins contributes to anogenital neoplasia. *Semin Cancer Biol, 7*(6), 309-315. doi:10.1006/scbi.1996.0040

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal, 6*(269), pl1. doi:10.1126/scisignal.2004088

Gillison, M. L., Koch, W. M., & Shah, K. V. (1999). Human papillomavirus in head and neck squamous cell carcinoma: are some head and neck cancers a sexually transmitted disease? *Curr Opin Oncol, 11*(3), 191-199.

Holzinger, D., Wichmann, G., Baboci, L., Michel, A., Hofler, D., Wiesenfarth, M., . . . Waterboer, T. (2017). Sensitivity and specificity of antibodies against HPV16 E6 and other early proteins for the detection of HPV16-driven oropharyngeal squamous cell carcinoma. *Int J Cancer, 140*(12), 2748-2757. doi:10.1002/ijc.30697

Lalla, R. V., Treister, N., Sollecito, T., Schmidt, B., Patton, L. L., Mohammadi, K., . . . Brennan, M. T. (2017). Oral Complications at Six Months after Radiation Therapy for Head and Neck Cancer. *Oral Dis*. doi:10.1111/odi.12710

Lechner, M., & Fenton, T. R. (2016). The Genomics, Epigenomics, and Transcriptomics of HPV-Associated Oropharyngeal Cancer--Understanding the Basis of a Rapidly Evolving Disease. *Adv Genet, 93*, 1-56. doi:10.1016/bs.adgen.2015.12.001

Liu, X., Han, S., Baluda, M. A., & Park, N. H. (1997). HPV-16 oncogenes E6 and E7 are mutagenic in normal human oral keratinocytes. *Oncogene, 14*(19), 2347-2353. doi:10.1038/sj.onc.1201078

Marur, S., & Forastiere, A. A. (2016). Head and Neck Squamous Cell Carcinoma: Update on Epidemiology, Diagnosis, and Treatment. *Mayo Clin Proc, 91*(3), 386-396. doi:10.1016/j.mayocp.2015.12.017

Mehnert, J. M., Panda, A., Zhong, H., Hirshfield, K., Damare, S., Lane, K., . . . Ganesan, S. (2016). Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. *J Clin Invest, 126*(6), 2334-2340. doi:10.1172/jci84940

Mork, J., Lie, A. K., Glattre, E., Hallmans, G., Jellum, E., Koskela, P., . . . Dillner, J. (2001). Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med, 344*(15), 1125-1131. doi:10.1056/nejm200104123441503

Nulton, T. J., Olex, A. L., Dozmorov, M., Morgan, I. M., & Windle, B. (2017). Analysis of The Cancer Genome Atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. *Oncotarget, 8*(11), 17684-17699. doi:10.18632/oncotarget.15179

Paz, I. B., Cook, N., Odom-Maryon, T., Xie, Y., & Wilczynski, S. P. (1997). Human papillomavirus (HPV) in head and neck cancer. An association of HPV 16 with squamous cell carcinoma of Waldeyer's tonsillar ring. *Cancer, 79*(3), 595-604.

Saleh, K., Eid, R., Haddad, F. G., Khalife-Saleh, N., & Kourie, H. R. (2018). New developments in the management of head and neck cancer - impact of pembrolizumab. *Ther Clin Risk Manag, 14*, 295-303. doi:10.2147/tcrm.s125059

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA Cancer J Clin, 65*(1), 5-29. doi:10.3322/caac.21254

Slebos, R. J., Yi, Y., Ely, K., Carter, J., Evjen, A., Zhang, X., . . . Chung, C. H. (2006). Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res, 12*(3 Pt 1), 701-709. doi:10.1158/1078-0432.ccr-05-2017

Snijders, P. J., Scholes, A. G., Hart, C. A., Jones, A. S., Vaughan, E. D., Woolgar, J. A., . . . Field, J. K. (1996). Prevalence of mucosotropic human papillomaviruses in squamous-cell carcinoma of the head and neck. *Int J Cancer, 66*(4), 464-469. doi:10.1002/(sici)1097-0215(19960516)66:4<464::aid-ijc9>3.0.co;2-u

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A, 102*(43), 15545-15550. doi:10.1073/pnas.0506580102

TCGA. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature, 517*(7536), 576-582. doi:10.1038/nature14129

TCGA. (2017). Integrated genomic and molecular characterization of cervical cancer. *Nature, 543*(7645), 378-384. doi:10.1038/nature21386

TCGA. (2017 [cited 5 September 2017]). The Cancer Genome Atlas Home Page. *National Cancer Institute,* http://cancergenome.nih.gov/.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin, 65*(2), 87-108. doi:10.3322/caac.21262

Wilczynski, S. P., Lin, B. T., Xie, Y., & Paz, I. B. (1998). Detection of human papillomavirus DNA and oncoprotein overexpression are associated with distinct morphological patterns of tonsillar squamous cell carcinoma. *Am J Pathol, 152*(1), 145-156.