# Polygenic Risk Score Prediction of Head and Neck Squamous Cell Carcinoma

Brian Schwab, MD[1]; Mridu Middha, MS[2]; Robert Klein, PhD[2]

[1]New York Eye and Ear Infirmary of Mount Sinai, [2]Icahn School of Medicine at Mount Sinai

## Abstract

Objectives: New research on the search for susceptibility alleles in complex disease processes provides support for the possibility of a polygenic approach in the prevention and treatment of common diseases. Despite the current multi-modality therapies available, approximately 50 percent of all patients will die of the disease. We sought to create a novel polygenic risk score to determine a patient's genetic risk of developing HPV-negative head and neck squamous cell carcinoma.

Study Design: Case-control study design using publicly available genotypes and controls.

Methods: Our study sample included 376 patients from the Cancer Genome Atlas and 10,842 controls. We created a polygenic risk score based on 15 known head and neck cancer-associated single nucleotide polymorphisms for risk stratification. These analyses were based on multiple genome-wide association studies. We then used multivariate logistic regression to determine the statistical significance of our model in cases versus controls. We then used a combination of Mann-Whitney U test and Kruskal-Wallis test to perform within-group analyses of the model.

Results: In a case-control study of 376 cancer patients vs. 10,842 control genotypes, our score was statistically significant (p = 0.0234). Our polygenic risk score was shown to be an independent risk factor for developing head and neck squamous cell carcinoma. This effect was independent of smoking or drinking history.

Conclusions: We have successfully created a novel polygenic risk score for predicting head and neck squamous cell carcinoma using 15 single nucleotide polymorphisms. This was shown to be an independent risk factor and not implicitly due to smoking or alcohol status. Further work is needed to determine the clinical utility of this approach.

## Introduction

Over 50,000 new cases of head and neck squamous cell carcinoma (HNSCC) are diagnosed in the United States annually. The WHO / GLOBOCAN estimates over 600,000 patients per year are affected. Despite the current multi-modality therapies available, approximately 50% of all patients will die of the disease. Of note, the focus of this investigation is on HPV(-) HNSCC only.

The search for genetic susceptibility has led to a recent innovation called genome-wide association studies (GWAS). GWAS focus on single-base pair changes called single-nucleotide polymorphisms (SNPs). GWAS are used to identify the SNPs significantly associated with disease risk. GWAS give you statistically significant SNPs, showing a normal base that is mutated with an odds ratio for cancer risk and p-value. However, the odds ratio of a single base-pair change can remain low.

A polygenic risk score (PRS) combines multiple SNPs into one single score. The goal of this project was to create a novel PRS for HPV(-) HNSCC genetic susceptibility.

## Methods and Materials

A systematic review was conducted to compile a list of significant SNPs as found in prior GWAS. The summation of individual SNP risk alleles multiplied by the log odds ratio for each of these 15 SNPs were used to create a PRS. Where $OR_i$ is the allelic odds ratio as estimated in the discovery dataset and $x_i$ is the number of risk alleles present at a single bi-allelic locus within a subject, the PRS formula is governed by the following: Polygenic Risk Score = $\sum [x_i \times \log(OR_i)]$

Samples were from the cancer genome atlas (TCGA), which is an NIH collaborative effort to collect genetic data from patients along with normal tissue. TCGA data included the genomes and clinical data of 376 patients with HPV-negative HNSCC. Shared controls were used with genotyped data from three studies totaling over 10,000 controls.

Entire genomes of both cases and controls were scanned using the Mount Sinai supercomputer Minerva. A multivariate logistic regression model was created to determine cancer as a function of PRS and 10 levels of genetic control. Either the Mann-Whitney U tests or Kruskal-Wallis test was used where appropriate for a number of within-group analyses such as the difference in PRS score based on age, gender, smoking status, subsite, and side.

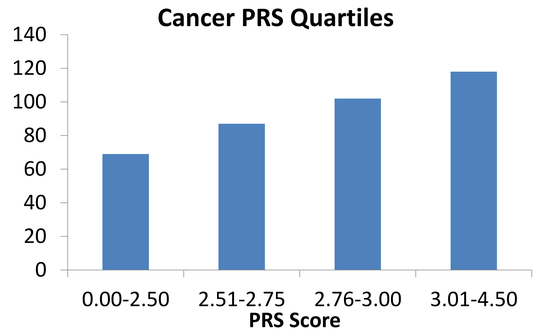
## Results

The demographic information for the 376 TCGA samples included a median age of 62 years (range 19-90) with 55.8% being over 60 years. There were 263 males and 113 females. The racial composition was 82.7% white, 10.6% black, and 6.6% other. The site of cancer was most often in oral cavity (67.8%), followed by larynx (26.6%), followed by oropharynx (5.6%). The tobacco use was as follows: 23.4% never smoked, 41.2% were former smokers, and 35.4% were current smokers. The average alcohol consumption was 54.5% 0-2 U/day, 21.7% 2-5 U/day, and 23.8% >5 U/day.

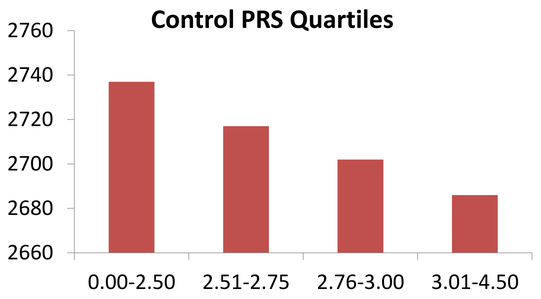
PRS quartiles were calculated for both TCGA and control groups, respectively. The TCGA PRS median was 2.860 and the control PRS median was 2.772. The multivariate logistic regression of TCGA samples versus 10,842 controls with 10 levels of principal component control showed a p-value of 0.0234 (odds ratio 1.324).

A number of within-group analyses were performed on the TCGA data and are shown in the following table:

| TCGA vs. Controls | P-Value | Odds Ratio |
|---|---|---|
| Cancer vs. Control | 0.0234 | 1.324 |
| **Within-Group Variable** | **P-Value** | |
| Male vs. Female | 0.917 | |
| Age >64 | 0.837 | |
| Subsite (oral tongue, larynx, etc.) | 0.994 | |
| Side (left, right, midline) | 0.710 | |
| Smoking Never vs. Ever | 0.680 | |
| Smoking Never vs. Former vs. Current | 0.827 | |
| Smoking By Status | 0.923 | |
| Alcohol None vs. Current | 0.466 | |
| Alcohol Less vs. More | 0.485 | |
| Alcohol By Status | 0.602 | |




| PRS | TCGA |
|---|---|
| Median | 2.860 |
| Std. dev. | 0.433 |
| Min | 1.681 |
| Max | 3.921 |




| PRS | Controls |
|---|---|
| Median | 2.772 |
| Std. dev. | 0.466 |
| Min | 0.898 |
| Max | 4.579 |

| RS ID | Chromosome | Gene | Description | Odds Ratio | P value |
|---|---|---|---|---|---|
| rs1494961 | 4q21 | HEL308 | DNA repair helicase | 1.11 | 1.00E-08 |
| rs1573496 | chr 4 | ADH7 | Alcohol dehydrogenase | 0.75 | 2.00E-08 |
| rs4767364 | 12q24 | ALDH2 | Aldehyde dehydrogenase | 1.1 | 5.00E-08 |
| rs698 | chr 4 | ADH1C | Alcohol dehydrogenase | 1.12 | 2.00E-02 |
| rs1229984 | 4q23 | ADH1B | Alcohol dehydrogenase | 1.29 | 0.031 |
| rs671 | 12q24 | ALDH2 | Aldehyde dehydrogenase | 0.46 | 0.015 |
| rs2274223 | 10q23 | PLCE1 | Phospholipase C | 1.85 | 0.022 |
| rs10492336 | 12q24 | TBX5 | T-box transcription factor | 0.71 | 4.50E-14 |
| rs174549 | 11q12 | FADS1 | Fatty acid desaturase | 0.41 | 1.00E-20 |
| rs2857595 | 6p21 | AIF1 | Allograft inflammatory factor | 0.5 | 2.40E-15 |
| rs2494938 | 6p21.1 | LRFN2 | Leucine repeat/fibronectin | 1.22 | 0.036 |
| rs1047840 | chr 1 | EXO1 | Exonuclease | 0.92 | 0.001 |
| rs1800734 | chr 3 | MLH1 | MutL homolog | 1.16 | 0.03 |
| rs2303426 | chr 2 | MSH2 | MutS homolog | 1.12 | 0.03 |
| rs26279 | chr 5 | MSH3 | MutS homolog | 1.38 | 0.04 |

## Discussion

Here we present evidence that our novel PRS is an independent risk factor for developing HNSCC. In a case-control study of 376 cancer patients versus 10,842 controls, our score was statistically significant (p = 0.0234). Only within-group analyses could be performed with the demographic and clinical data due to the usage of shared controls. A larger dataset with clinical information would help elucidate the strength of our PRS. The future direction of our project is to use a large clinical cohort to test treatment outcomes, survival, and other clinical variables. Our specific intent is to use PRS to determine risk for second primaries or recurrence for clinical usage.

## Conclusions

We have successfully created a novel polygenic risk score for predicting head and neck squamous cell carcinoma using 15 single nucleotide polymorphisms. This was shown to be an independent risk factor and not implicitly due to smoking or alcohol status. Further work is needed to determine the clinical utility of this approach.

## Contact

Brian Schwab, MD
New York Eye and Ear Infirmary of Mount Sinai
Email: bschwab@nyee.edu
Website: www.nyee.edu
Phone: 212-979-4000

## References

1. Rettig, Eleni M., and Gypsyamber D'Souza. "Epidemiology of Head and Neck Cancer." Surgical oncology clinics of North America 24.3 (2015): 379-396.
2. Cancer Genome Atlas Network. "Comprehensive genomic characterization of head and neck squamous cell carcinomas." Nature 517.7536 (2015): 576-582.
3. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress. A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014
4. Stadler, Zsofia K., et al. "Genome-wide association studies of cancer." Journal of Clinical Oncology 28.27 (2010): 4255-4267.
5. Pharoah, Paul DP, et al. "Polygenic susceptibility to breast cancer and implications for prevention." Nature genetics 31.1 (2002): 33-36.
6. Purcell, Shaun M., et al. "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder." Nature 460.7256 (2009): 748-752.
7. Garcia-Closas, Montserrat, et al. "Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer." Cancer research 73.7 (2013): 2211-2220.
8. Machiela, Mitchell J., et al. "Evaluation of polygenic risk scores for predicting breast and prostate cancer risk." Genetic epidemiology 35.6 (2011): 506-514.
9. Cross-Disorder Group of the Psychiatric Genomics Consortium. "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis." The Lancet 381.9875 (2013): 1371-1379.