



Article

Identification of the HNSC88 Molecular Signature for Predicting Subtypes of Head and Neck Cancer

Yi-Hsuan Chuang^{1,†}, Chun-Yu Lin^{1,2,†} , Jih-Chin Lee³, Chia-Hwa Lee^{4,5,6} , Chia-Lin Liu⁷, Sing-Han Huang¹ , Jung-Yu Lee¹, Wen-Sen Lai⁸ and Jinn-Moon Yang^{1,2,9,*}

- ¹ Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan
- ² Center for Intelligent Drug Systems and Smart Bio-Devices, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan
- ³ Department of Otolaryngology—Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan
- ⁴ School of Medical Laboratory Science and Biotechnology, College of Medical Science and Technology, Taipei Medical University, Taipei 110, Taiwan
- ⁵ TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei 110, Taiwan
- ⁶ Ph.D. Program in Medicine Biotechnology, College of Medical Science and Technology, Taipei Medical University, Taipei 110, Taiwan
- ⁷ Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 114, Taiwan
- ⁸ Department of Otolaryngology—Head and Neck Surgery, Taichung Armed Forces General Hospital, Taichung 404, Taiwan
- ⁹ Center of Excellence for Metabolic Associated Fatty Liver Disease, National Sun Yat-Sen University, Kaohsiung 804, Taiwan
- * Correspondence: moon@nycu.edu.tw
- † These authors contributed equally to this work.



Citation: Chuang, Y.-H.; Lin, C.-Y.; Lee, J.-C.; Lee, C.-H.; Liu, C.-L.; Huang, S.-H.; Lee, J.-Y.; Lai, W.-S.; Yang, J.-M. Identification of the HNSC88 Molecular Signature for Predicting Subtypes of Head and Neck Cancer. *Int. J. Mol. Sci.* **2023**, *24*, 13068. <https://doi.org/10.3390/ijms241713068>

Academic Editor: Elisabetta Palazzo

Received: 20 July 2023

Revised: 14 August 2023

Accepted: 17 August 2023

Published: 22 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Head and neck squamous cell carcinoma (HNSC) exhibits genetic heterogeneity in etiologies, tumor sites, and biological processes, which significantly impact therapeutic strategies and prognosis. While the influence of human papillomavirus on clinical outcomes is established, the molecular subtypes determining additional treatment options for HNSC remain unclear and inconsistent. This study aims to identify distinct HNSC molecular subtypes to enhance diagnosis and prognosis accuracy. In this study, we collected three HNSC microarrays ($n = 306$) from the Gene Expression Omnibus (GEO), and HNSC RNA-Seq data ($n = 566$) from The Cancer Genome Atlas (TCGA) to identify differentially expressed genes (DEGs) and validate our results. Two scoring methods, representative score (RS) and perturbative score (PS), were developed for DEGs to summarize their possible activation functions and influence in tumorigenesis. Based on the RS and PS scoring, we selected candidate genes to cluster TCGA samples for the identification of molecular subtypes in HNSC. We have identified 289 up-regulated DEGs and selected 88 genes (called HNSC88) using the RS and PS scoring methods. Based on HNSC88 and TCGA samples, we determined three HNSC subtypes, including one HPV-associated subtype, and two HPV-negative subtypes. One of the HPV-negative subtypes showed a relationship to smoking behavior, while the other exhibited high expression in tumor immune response. The Kaplan–Meier method was used to compare overall survival among the three subtypes. The HPV-associated subtype showed a better prognosis compared to the other two HPV-negative subtypes (log rank, $p = 0.0092$ and 0.0001 ; hazard ratio, 1.36 and 1.39). Additionally, within the HPV-negative group, the smoking-related subgroup exhibited worse prognosis compared to the subgroup with high expression in immune response (log rank, $p = 0.039$; hazard ratio, 1.53). The HNSC88 not only enables the identification of HPV-associated subtypes, but also proposes two potential HPV-negative subtypes with distinct prognoses and molecular signatures. This study provides valuable strategies for summarizing the roles and influences of genes in tumorigenesis for identifying molecular signatures and subtypes of HNSC.

Keywords: head and neck cancer; systems biology; molecular subtypes; molecular signature; human papillomavirus; prognostic biomarker

1. Introduction

Head and neck squamous cell carcinoma (HNSC) is a type of cancer with a high mortality rate worldwide, accounting for over 800,000 new cases and 400,000 deaths annually [1]. The prevalent risk factors contributing to the increased incidence of HNSC include smoking, drinking, human papilloma virus (HPV) infection, and chewing of areca. Among these risk factors, HPV status is a critical prognostic indicator for HNSC. HPV-positive HNSC patients often exhibit more favorable clinical outcomes compared to HPV-negative patients. [2–4]. The survival rates of HNSC patients have not significantly improved in recent decades due to the heterogeneity of the disease in terms of etiologies, tumor sites, and genetic characterization. This heterogeneity has posed challenges in the discovery of effective diagnostic, prognostic, and therapeutic biomarkers for HNSC [2–5]. Due to the heterogeneity of HNSC, some studies have revealed that approximately 60% of HNSC patients with overexpression of epidermal growth factor receptor (EGFR) experienced worse clinical outcomes, but the remaining 40% of patients did not [2,6,7]. Currently, HNSC patients who have failed standard treatments, such as surgery, radiation, and chemotherapy, are lacking alternative treatment options.

An ideal biomarker for cancer includes several characteristics. Firstly, it should exhibit significant expression during the disease state. Secondly, the candidate genes can reveal the underlying mechanisms of the disease. Thirdly, it should be associated with disease prognosis. Additionally, the detection should demonstrate consistency and reliability [8]. Previous research has proposed the benefits of biomarkers and subtypes for cancer diagnosis and prognosis. For example, PSA (prostate-specific antigen) is a well-known biomarker for prostate cancer, used for early detection and monitoring prognosis [9]. In breast cancer, biomarkers such as ER (estrogen receptor), PR (progesterone receptor), and HER2 (human epidermal growth factor receptor 2) are utilized for the identification of intrinsic subtypes [10–13]. Additionally, the breast cancer 70-gene MammaPrint microarray assay is used to predict patient prognosis [14]. These biomarkers are valuable to decide treatment and predict patient clinical outcomes in breast cancer. However, beyond the distinction of the subgroups of HPV-positive and HPV-negative, there are no established molecular subtypes for HNSC clinical application [2,3,15].

In this study, we aim to develop a strategy for the identification of HNSC molecular signatures and subtypes (Figure 1). We first utilized gene expression between tumor and normal tissues to determine 289 differentially up-regulated DEGs of three microarrays. Then, the two scoring methods, representative score (RS) and perturbative score (PS), were developed to group and select 88 genes (HNSC88) by estimating their relative relations of cellular functions and dysregulation in the tumor state. Finally, the hierarchical clustering was performed to discriminate three HNSC subtypes, and Kaplan–Meier analysis was used to identify their prognostic significances.

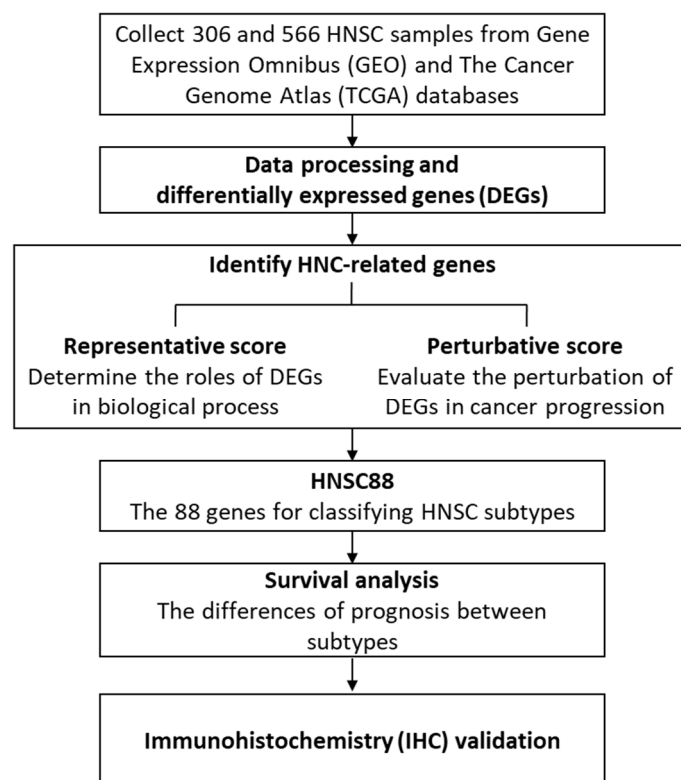


Figure 1. Schematic of identifying HNSC biomarkers.

2. Results

2.1. Identification of Molecular Signature for HNSC Using CS, RS, and PS Scoring

To obtain the molecular signature that reveals cellular functions and subtypes of HNSC, three scoring methods were developed to identify candidate genes; namely, Cluster Score (CS), Representative Score (RS), and Perturbation Score (PS). First, we identified 478 DEGs that significantly changed across three microarrays. Based on the general characteristics of biomarker used in clinical medicine [8], only 289 up-regulated DEGs were further analyzed in this study. The CS value (Equation (1)) was calculated to estimate the relative relation of cellular functions between any two genes among the 289 up-regulated DEGs (Figure 2). Next, the 289 DEGs were clustered into 27 groups by using CS values and unsupervised hierarchical clustering. To identify the cellular functions and pathways of the 27 cluster groups in HNSC, we implemented functional enrichment analysis of GOBP, KEGG pathways, and ten hallmarks of cancer, with the hypergeometric test. The results of the enrichment analysis were ranked by enrichment p -value, with the top three being presented in Table 1.

Of the 27 cluster groups, only 11 groups (G1 to G11, Figure 2, left dendrogram) contained at least one GOBP with an enrichment p -value < 0.05 and had ≥ 3 gene members involved in that GOBP. The cellular functions of these 11 groups could be summarized into five cancer-related functions for HNSC: tumor immune response (G1 to G3), tumor survival (G4), tumor metastasis (G5 to G7, G10), tumor growth (G8, G9), and tumor metabolism (G11).

For these 11 groups, the RS scoring (Equation (2)) was used to calculate the sum of CS values among the gene members, estimating the genes that were representative of cellular functions within a cluster group. The PS scoring (Equation (3)) was applied to each gene to estimate the influence when the gene was dysregulated in HNSC. Finally, the 88 genes (HNSC88) were selected by ranking the geometric means of RS and PS (Equation (4)). In this study, the HNSC88 gene set was considered the molecular signature for classifying and identifying HNSC tumors. The list of 88 genes (HNSC88) and their

corresponding fold change between tumor subgroups and normal tissues are presented in Supplementary Data S1.

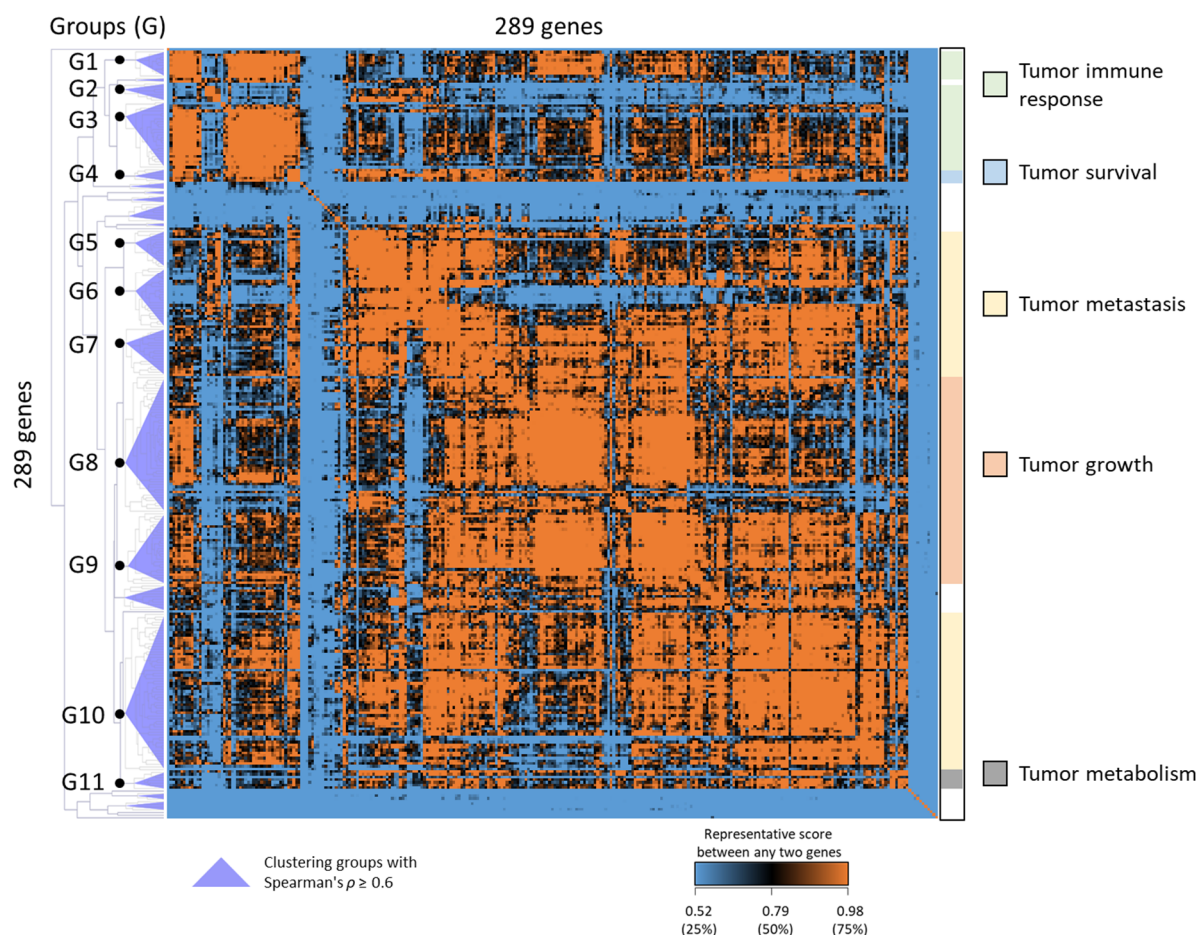


Figure 2. Hierarchical clustering of CS values for 289 significantly up-regulated DEGs. The clustering demonstrates the relative relation of cellular functions between any two genes. The 289 DEGs were classified into 27 cluster groups using a threshold of Spearman's $\rho > 0.6$ (indicated by the left purple triangle). GOBP enrichment analysis was conducted, and only 11 groups (G1 to G11) were identified, which had at least one significantly enriched GOBP with a p -value < 0.05 and ≥ 3 members. The five cancer hallmarks represented by the 11 groups are displayed in the right color bar.

Table 1. GOBP, KEGG and cancer hallmark enrichment analysis.

Cluster Group #	Members	GOBP	KEGG Pathways	Summary of Pathways	Cancer-Related Processes	The Hallmarks of Cancer (Summary of Processes)
G1	APOBEC3A, CMPK2, DTX3L, HERC5, IFI30, STAT1	- Defense response to virus; - Response to virus; - Response to other organism;	-	-	-	-

Table 1. Cont.

Cluster Group #	Members	GOBP	KEGG Pathways	Summary of Pathways	Cancer-Related Processes	The Hallmarks of Cancer (Summary of Processes)
G2	CXCL1, IL1F9, IL8, MMP12, MMP9	- Response to wounding; - Inflammatory response; - Locomotion;	- IL-17 signaling pathway; - Cytokine-cytokine receptor interaction; - Rheumatoid arthritis;	- Signaling molecules and interaction; - Immune disease;	- Cytokines chemokines production; - Tumor growth;	- Activating Invasion and Metastasis; - Tumor-Promoting Inflammation;
G3	DDX60, GBP1, IFI35, IFI6, IFIT1, IFIT3, ISG15, RSAD2, UBE2L6	- Response to virus; - Defense response to virus; - Innate immune response;	- Cytosolic DNA-sensing pathway; - Influenza A; - RIG-I-like receptor signaling pathway;	- Immune system; - Infectious disease: viral;	-	-
G4	CTSL1, NCF2, RGS2, SOD2, TLR2	- Response to lipopolysaccharide; - Response to molecule of bacterial origin; - Cell-type specific apoptotic process;	- Phagosome;	- Transport and catabolism;	-	-
G5	COL4A1, COL4A2, COL5A1, COL5A2, CTHRC1, PXDN	- Collagen catabolic process; - Multicellular organismal catabolic process; - Extracellular matrix disassembly;	- Protein digestion and absorption; - ECM-receptor interaction; - Focal adhesion;	- Digestive system; - Signaling molecules and interaction; - Cellular community—eukaryotes;	- Cell matrix adhesions; - ECM; - EMT regulators;	- Activating Invasion and Metastasis;
G6	INHBA, LAMC2, MMP1, MMP10, MMP3, PLAUI, PTHLH, TGFBI	- Extracellular matrix organization; - Extracellular structure organization; - Cellular component movement;	- TGF-beta signaling pathway; - IL-17 signaling pathway; - Prostate cancer;	- Signal transduction; - Immune system; - Cancer: specific types;	- ECM; - Tumor growth; - Matrix regulation;	- Activating Invasion and Metastasis; - Tumor-Promoting Inflammation;
G7	ACVR1, CLIC4, DSG2, FAT1, FNDC3B, ITGAV, SNAI2	- Vasculature development; - Cardiovascular system development; - Circulatory system development;	-	-	- EMT regulator;	- Activating Invasion and Metastasis;
G8	CDC45, CDC6, CDK4, CHEK1, CKS1B, DTL, FEN1, FOXM1, GINS1, MCM2, RFC4, TPX2, UHRF1	- Cell cycle process; - Cell cycle; - Interphase;	- DNA replication; - Cell cycle; - p53 signaling pathway;	- Replication and repair; - Cell growth and death;	- S-CC phase;	- Evading Growth Suppressors;
G9	AURKA, BIRC5, CDC20, CEP55, ECT2, KIF2C, KIF4A, NUP155, TRIP13	- Nuclear division; - Mitosis; - Organelle fission;	-	-	-	-

Table 1. Cont.

Cluster Group #	Members	GOBP	KEGG Pathways	Summary of Pathways	Cancer-Related Processes	The Hallmarks of Cancer (Summary of Processes)
G10	CDH3, CEBPB, DFNA5, GJA1, HOMER3, JUP, KLF10, KLF7, KRT17, MSN, MYO1B, PANX1, PRNP, RRAS2, SLC16A1	- System development; - Locomotion; - Response to endogenous stimulus;	- Regulation of actin cytoskeleton; - Arrhythmogenic right ventricular cardiomyopathy (ARVC); - Autophagy—animal;	- Cell motility; - Cardiovascular disease; - Transport and catabolism;	- EMT regulator; - Cell-cell adhesions;	- Activating Invasion and Metastasis;
G11	ACOT9, FKBP9, GALNT2, UBE2Q2, VKORC1	- Peptidyl-amino acid modification;	-	-	-	-

The column colors of the Cluster Group correspond to the five cancer hallmarks in Figure 2.

2.2. Unsupervised Clustering of HNSC Subtypes Based on the Expression of HNSC88

To identify HNSC molecular subtypes, we employed HNSC88 and unsupervised hierarchical clustering on 345 HNSC tumors, including full HPV status records (Figure 3). The HPV status was collected by combining the TCGA clinical data with the results of Tang, K. W. et al. [16]. Based on the clustering dendrogram (Spearman's $\rho \geq 0.7$) and the gene expression profile of 345 tumors and HNSC88, three primary subtypes were determined (1: purple, 2: pink, and 3: blue). We observed differential distributions of HPV status, smoking history, and TP53 somatic mutations among the three subtypes, which may reflect in the survival rate.

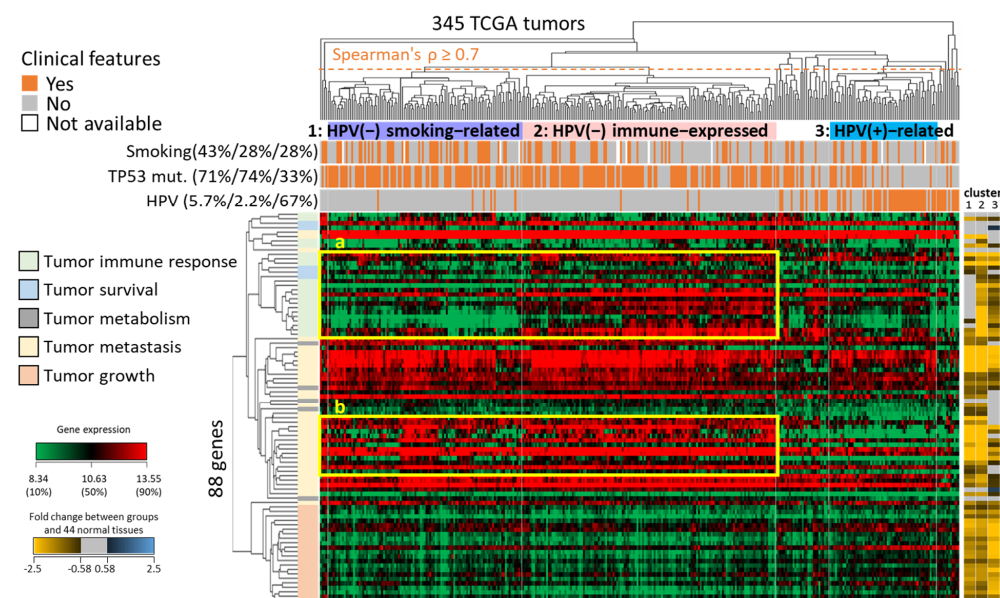


Figure 3. Hierarchical clustering of HNSC88 gene expression for the classification of HNSC subtypes. Unsupervised clustering analysis of gene expression for 88 genes (HNSC88) in 345 HNSC tumors revealed three distinct subgroups (cluster 1, 2, and 3) using Spearman's $\rho \geq 0.7$ as the threshold. Above the heatmap, the distribution of three clinical features (i.e., current smoking, TP53 somatic mutation, and HPV status) is displayed with orange color for the individual tumor samples. To the left of the heatmap, the functions of the 88 genes were identified through enrichment analysis (please see Figure 2 and Table 1). The expression fold changes of the 88 genes between the three subtypes and normal tissues are shown on the right.

For example, cluster 1 was considered an HPV-negative smoking-related subtype, which included the largest percentage of smoking patients (43%) and showed low expression in tumor immune response (indicated by the yellow frame a). Some studies have indicated that smoking may suppress human immune function, thus facilitating immune evasion by cancer cells and reducing the effectiveness of HNSC treatments [17–22]. Cluster 2 was considered an HPV-negative immune-expressed subtype, displaying the highest gene expression in the gene set related to tumor immune response (indicated by the blue frame 1). On the other hand, cluster 3 was classified as an HPV-positive-related subtype, showing the largest percentage of HPV-infected patients compared to cluster 1 and cluster 2, respectively (67% versus 5.7% and 2.2%). The HPV-positive-related subtype also exhibited higher expression in tumor growth, and some studies have indicated that HPV virus may dysregulate the cell cycle to promote cancer cell growth (e.g., CDK4 and MCM2, both members of HNSC88) [2,23]. Furthermore, there was no significant difference between cluster 1 and cluster 2 in the ratio of TP53 somatic mutations (odds ratio was 0.97). However, the frequency of TP53 somatic mutations in both HPV-negative subgroups was higher than that in the HPV-positive-related subgroup (odds ratios were 2.18 and 2.25, respectively). We also observed that the genes and their gene expression involved in tumor metastasis of HPV-negative patients in both subgroups are often up-regulated DEGs or show higher expression, compared to HPV-positive-related patients. This may be one of the reasons why HPV-positive patients have better clinical outcomes (the yellow frame b).

2.3. Prognostic Significance of Three HNSC Subtypes

The genome characterization and clinical features mentioned above are reflected in the survival rates of subtypes (Figure 4). Firstly, the cluster 3 subtype (HPV-positive-related) significantly contains more HPV-infected patients who are often TP53 wild-type and non-smokers, showing more favorable survival outcomes. In addition, compared to patients in clusters 1 and 2, HPV-negative patients in cluster 3 present a similar gene expression profile to HPV-positive patients (lower expression in tumor metastasis), which may contribute to a better survival rate (Figure 4, blue line).

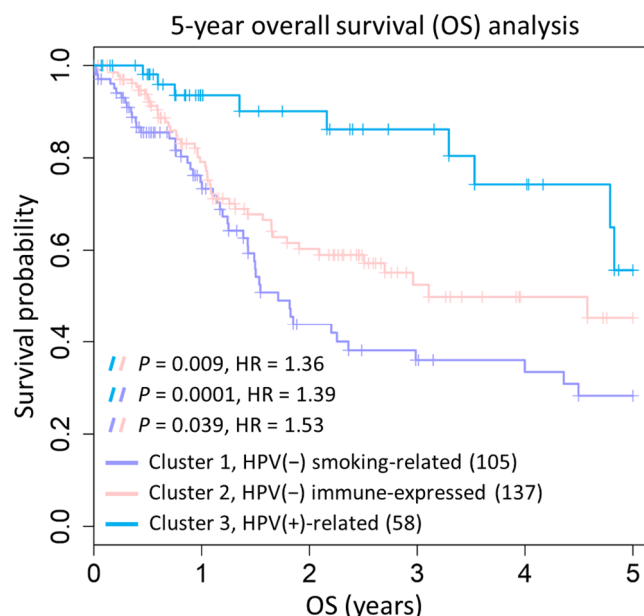


Figure 4. Kaplan–Meier plots of overall survival (OS) for three subtypes identified by HNSC88 in 345 HNSC tumors (blue: HPV-positive; pink: HPV-negative immune-expressed; purple: HPV-negative smoking-related). The p -value of log-rank test and hazard ratio between subtypes were performed and displayed.

We further identified two interesting HPV-negative subtypes (cluster 1 and 2, purple and pink lines) in which the tumor metastasis gene set was significantly up-regulated, leading to adverse prognosis compared to the HPV-positive subtype (log-rank p -value = 0.0001 and 0.009; hazard ratio = 1.39 and 1.36, respectively). Moreover, the survival of the HPV-negative smoking-related subtype (cluster 1) was significantly lower compared to cluster 2, further increasing the risk of death (log-rank p -value = 0.039; HR = 1.53). Based on the clustering and survival analysis, we identified three subtypes with statistically significant differences in overall survival. One subtype was HPV-positive, associated with favorable survival, while the other two subtypes were HPV-negative, exhibiting poor survival and could be distinguished based on gene expression and smoking behavior.

2.4. Immunohistochemistry (IHC) Stain of Clinical HNSC Patients

Based on our domain knowledge, we selected three genes, MMP9, NCF2, and IFI30, to verify their protein expression in tumors and normal tissues using IHC staining (Figure 5). The primary antibodies were purchased from GeneTex (Irvine, CA, USA), with catalog numbers GTX62122, GTX62954, and GTX103967, respectively. The gene MMP9 was used as a positive control and is a well-known biomarker for HNSC [24,25], which showed strong protein expression only in tumor regions (Figure 5A). The other two genes, NCF2 and IFI30, play roles in immune response, antigen presentation, inflammation, cell invasion, and cell survival, and they have been considered as diagnostic and prognostic biomarkers in several cancers [26–31]. However, there are few studies that have provided evidence for their roles in HNSC (Figure 5B,C). According to the literature, the gene IFI30 is a novel biomarker for HNSC, and we may have conducted the first study to verify its protein expression in HNSC slides by IHC staining. The protein expression in tumor and normal regions were presented by – (negative) and + (positive) (Figure 5D).

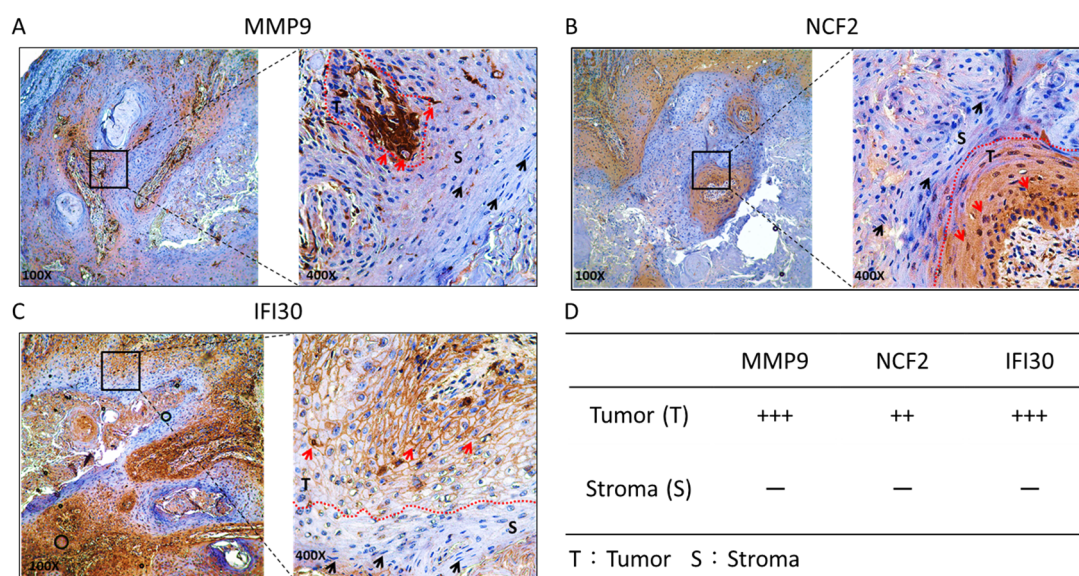


Figure 5. Evaluation of protein expression of (A) MMP9, (B) NCF2, and (C) IFI30 by immunohistochemistry (IHC) staining. The tumor tissue slides were stained with primary antibodies against MMP9, NCF2, and IFI30. Their protein expressions in the cancerous region are indicated by red arrows, whereas those in the normal adjacent region are indicated with black arrows. The red dashed line was used to highlight tumor tissues (i.e., the brown part). (D) The intensity of protein expression is presented as – (negative) and + (positive), respectively.

3. Methods

3.1. Datasets

In this study, we collected 235 HNSC tumor samples and 71 corresponding normal tissues from three GEO microarrays (GSE30784, GSE6791, and GSE9844) for the identifica-

tion of HNSC signatures [32–34]. The platform of the three microarrays is the Affymetrix Human Genome U133 Plus 2.0 Array, which includes 54,675 probes. All raw data (CEL files) were processed using RMA normalization and log2-transformation with the R package “affy” [35]. Next, we mapped the 54,675 probes to 19,213 genes (UniProt, Cambridge, UK) using Affymetrix annotations (version 34). For each gene, we obtained a corresponding expression value from the probes. If a gene was mapped by multiple probes, its expression value was calculated as the average of these probes.

We further collected 520 HNSC tumor samples and 44 normal tissues (level 3 RNA-Seq data with tumor type: 01) from TCGA to validate subtypes [36]. For TCGA samples, we utilized the clinical data, such as HPV status, smoking history, and clinical follow-up information, to annotate HNSC subtypes and conduct survival analysis. Because HPV status is a very important risk factor for HNSC in clinical outcomes, we only selected 345 TCGA patients annotated with HPV status for further subtype analysis. The HPV statuses of 345 patients were obtained and integrated from TCGA clinical data and Tang, K. W. et al. [16]. To evaluate the perturbations of DEGs for signaling pathways, we collected human protein–protein interactions (PPIs) from five databases (BioGRID, Samuel Lunenfeld Research Institute, Toronto, ON, Canada; DIP, University of California, Los Angeles, CA, USA; IntAct, European Molecular Biology Laboratory, Cambridge, UK; MINT, University of Rome Tor Vergata, Rome, Italia; and MIPS, German Research Center for Environmental Health, Neuherberg, Germany) [37–41], and utilized our previous approaches to obtain 267,326 PPIs of 16,596 human proteins [42,43].

3.2. Identification of Differentially Expressed Genes (DEGs)

We used the fold change and *p*-value (R package limma) to identify the differentially expressed genes (DEGs) between normal tissues and tumors in the three microarrays. By using the $|\text{fold change}| \geq 1.5$ and $p\text{-value} < 0.05$, we found 2521, 1814, and 847 DEGs in the GSE30784, GSE6791, and GSE9844 datasets, respectively. Based on the previously mentioned characteristics of an ideal biological marker, we only used 289 DEGs that were significantly up-regulated in all three microarrays for further analysis.

3.3. Functional Enrichment Analysis

The terms of gene ontology biological process (GOBP) and cellular component (GOCC) were obtained from the Gene Ontology database (<http://geneontology.org/>). For these GO terms, we employed five relationships to construct the GO trees: “is_a”, “part_of”, “regulates”, “negatively_regulates”, and “positively_regulates”. These GO data were utilized for developing the Cluster Score (CS) and conducting enrichment analysis for gene cluster groups [44,45]. We additionally collected gene sets of 10 cancer hallmarks to analyze the connections between tumor behaviors and 289 DEGs. The cancer hallmarks data were obtained from the Atlas of Cancer Signalling Network (ACSN) database (<https://acsn.curie.fr/ACSN2/HMC.html>, accessed on 19 July 2023), which offered 93 cancer-related functions and 2652 cancer-related genes [46]. The 330 pathways were also downloaded from the KEGG database and utilized to annotate the potential signaling pathways in which the DEGs are involved. We downloaded the XML pages (KGML) for each KEGG pathway and extracted the gene members of these pathways [47]. The summary of the enrichment analysis is displayed in Table 1.

3.4. The Scoring Methods for Identification of HNSC-Related Genes

We have developed the Cluster Score (CS) to estimate the similarity of cellular functions between any two genes out of the 289 DEGs in HNSC tumors. The CS value, along with hierarchical clustering, was employed to group the 289 DEGs based on the relative relations of their cellular functions. The calculation of CS is as follows:

$$CS_{i,j} = \frac{RSS_{BP_{i,j}} + RSS_{CC_{i,j}}}{2} + \frac{\sum_{n=1}^N PCC_{i,j}^n}{N} \quad (1)$$

where RSS_{BP} and RSS_{CC} represent the relative specificity similarity (RSS) of gene ontology biological process (GOBP) and cellular component (GOCC) between gene i and gene j , respectively (RSS value ranges from 0 to 1) [48]. PCC_{ij} denotes the Pearson's correlation coefficient of gene expression between gene i and gene j . N refers to the number of microarray datasets used (in this case, $N = 3$).

After clustering the 289 DEGs using CS, we developed two scoring methods; namely, the Representative Score (RS) and Perturbation Score (PS), to identify the molecular signatures for the classification of HNSC subtypes. For each gene, the RS value is calculated as the sum of the CS values within each cluster group. This allows us to identify genes that exhibit a high association with every other gene member in terms of cellular functions. On the other hand, to assess the influence of genes (i.e., DEGs) on signaling pathways when dysfunctions occur in HNSC, we utilized the human PPI network to compute the Perturbation Score (PS). The RS and PS are defined as follows:

$$RS_i = \frac{\sum_{j=1}^M CS_{ij}}{2 \times M} \quad (2)$$

where M is the number of gene members in the cluster group, and CS_{ij} is the cluster score between gene i and gene j in the same cluster group.

$$PS = NFC_i + NP_i + NIG_i + NCG_i \quad (3)$$

where NFC_i is the fold change of gene i between normal tissues and tumors; NP_i is the t-statistics p -value of gene i between normal tissues and tumors; NIG_i represents the number of genes that both interact with gene i (i.e., PPIs of gene i) and are co-expressed ($| \text{Pearson's } r | \geq 0.5$) with gene i . For example, the protein NCF2 (UniProt: P19878) has 37 PPIs and is co-expressed with 1 gene among them. Therefore, its NIG_i value in one of the datasets is 1. In addition, we also provided NCG_i values for the genes that lacked PPI data. NCG_i is the number of genes co-expressed ($| \text{Pearson's } r | \geq 0.7$) with gene i in the context of HNSC. The PS values of the three microarrays were ranked, averaged, and then normalized to a range of 0 to 1. Then, the geometric mean of the RS value and PS value was employed to select the candidate genes.

The number of selected candidates (SC) from each cluster group is provided as follows:

$$SC = \frac{G - \min(G)}{\min(G)} + \min(G) \quad (4)$$

where G is the number of genes in each cluster group (i.e., G1 to G11). The $\min(G)$ is the minimum number of members among all cluster groups (here it is 5). Finally, a total of 88 genes (called HNSC88) from 11 groups were selected for classifying HNSC subtypes.

3.5. Survival Analysis

We utilized TCGA clinical data, including clinical follow-up, smoking history, and HPV status, to investigate the prognostic significance of HNSC subtypes. The 5-year overall survival (OS) between HNSC subtypes was compared using Kaplan–Meier analysis (R-2.15.3 package `survplot`) [49]. The log-rank test and Cox proportional hazards regression model (R package `survival`) were employed to estimate their prognostic differences.

4. Discussion

HNSC is a highly heterogeneous cancer, and there are only a few alternative treatment options available for patients who do not respond to surgery or radiotherapy/chemotherapy. Therefore, our aim is to discover molecular signatures that can determine HNSC subtypes with prognostic significance. In this study, we have employed several strategies and made the following findings: (1) We developed two scoring methods to select 88 genes (HNSC88) that were significantly up-regulated in three microarrays; namely, representative score

(RS) and perturbative score (PS). (2) HNSC88 is involved in five major cellular functions, including tumor immune response, tumor metastasis, tumor survival, tumor growth, and tumor metabolism, which contribute to the determination of HNSC subtypes. (3) Three HNSC subtypes were identified and showed different prognostic significance of the one HPV-positive and two HPV-negative subgroups. The most distinct difference in molecular signatures between the two HPV-negative subgroups was related to smoking behavior and tumor immune response. Numerous studies have indicated that smoking can suppress the immune system and hinder the effectiveness of HNSC treatment [17,20,50]. We believe that our results not only provide additional evidence for the relationships between HPV status, smoking, and clinical outcomes, but also highlight the detrimental effects of tobacco consumption on HNSC.

However, some limitations and findings still need to be improved and verified. First, our RS and PS scoring rely on the annotations of cellular functions (i.e., GOBP and GOCC) and protein-interaction records to estimate the relationship between two genes. With more comprehensive annotations, we could potentially cluster genes more accurately and discover additional biological functions by considering surrounding gene members. Second, the percentage of patients with TP53 mutation was not significantly different between the two HPV-negative subtypes (cluster 1 and 2). Therefore, we can only suggest the associations among smoking, immune response, and clinical outcomes without providing specific genomic pattern results (e.g., mutation data).

5. Conclusions

In summary, we have presented a strategy to select the HNSC molecular signature for predicting three subtypes with prognostic significance. The three subtypes include one HPV-positive subtype with a favorable prognosis, which is consistent with current domain knowledge. The other two HPV-negative subtypes showed differences in smoking behavior and the gene expression related to tumor immune response, and they also exhibited distinct survival rates. Additionally, we identified a novel gene, IFI30, that may be associated with favorable prognosis. Compared to the subtype with the worst survival rate, IFI30 was overexpressed by approximately 2-fold in the subtypes with the best and second-best favorable survival. We believe that our scoring methods and HNSC88 provide an opportunity to develop diagnostic and prognostic markers for HNSC in the future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241713068/s1>.

Author Contributions: Y.-H.C. analyzed the omics data and wrote the manuscript; C.-Y.L. provided ideas in this study, managed and arranged staff/resources; J.-C.L. provided the HNSC clinical tissues for IHC staining; C.-H.L. provided the materials of bioassays and knowledge of cancer cell biology; C.-L.L. performed the IHC staining. S.-H.H. prepared the computing environments for R language and program code; J.-Y.L. provided contribution and discussion on biological meaning; W.-S.L. provided the practical clinical comments in HNSC; J.-M.Y. designed and supervised the study, provided the intellectual contribution, discussion and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Center for Intelligent Drug Systems and Smart Bio-devices (IDS2B) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project sponsored by the Ministry of Education (MOE) in Taiwan, the National Science and Technology Council (NSTC) of Taiwan (NSTC110-2634-F-A49-005 and NSTC112-2636-B-A49-007), the Research Center for Epidemic Prevention Science (NSTC111-2321-B-A49-007), the Joint Research Center for AI Technology (NSTC110-2634-F-009-015), the Overseas Project for Post Graduate Research (NSTC112-2917-I-A49-015), Artificial Intelligence to Precision Health: Integrating Dynamic Physiological Signals and EMR to build a Medical Digital Twins Platform (NSTC111-2321-B-A49-010), National Yang Ming Chiao Tung University and National Health Research Institutes (NHRI-EX111-11017BI).

Institutional Review Board Statement: The omics and clinical information in this study were collected from publicly accessible Gene Expression Omnibus (GEO) and TCGA databases, therefore and we did not need the ethical approval from local committee.

Informed Consent Statement: For the clinical IHC assays in Figure 5, we obtained IRB approval provided by Dr. Jih-Chin Lee (one of author) from Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan. The local committee approval number is 1-108-05-104.

Data Availability Statement: The three HNSC microarrays (GSE30784, GSE6791 and GSE9844) were collected from Gene Expression Omnibus (GEO). The RNA-Seq of HNSC tissues and clinical information were collected from TCGA-HNSC cohort of Genomic Data Commons Data Portal database (<https://portal.gdc.cancer.gov/>). Kaplan-Meier analysis was performed by using R tool which could be obtained from R Project (<https://www.r-project.org/>). All relevant materials are provided in the manuscript. The list of 88 genes (HNSC88) and their corresponding fold change between tumor subgroups and normal tissues are presented in Supplementary Data S1.

Acknowledgments: This work was partly supported by the “Center of Excellence for Metabolic Associated Fatty Liver Disease, National Sun Yat-sen University, Kaohsiung” and “Cancer and Immunology Research Center [112W31101], National Yang Ming Chiao Tung University” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Conflicts of Interest: The authors declare that they have no competing interests.

Abbreviations

HNSC: Head and Neck Squamous cell Carcinoma; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; DEG: Differentially Expressed Genes; RS: Representative Score; PS: Perturbative Score; CS: Cluster Score; GO: Gene Ontology; GOBP: Gene Ontology Biological Process; GOCC: Gene Ontology Cellular Component; HPV: Human Papilloma Virus; HPV-neg: HPV-negative; HPV-pos: HPV-positive; PSA: Prostate-Specific Antigen; ER: Estrogen Receptor; PR: Progesterone Receptor; ACSN: Atlas of Cancer Signalling Network database; KEGG: Kyoto Encyclopedia of Genes and Genomes database; RSS: Relative Specificity Similarity; PCC: Pearson's Correlation Coefficient; PPI: Protein-Protein Interaction (PPI); OS: Overall Survival; HR: Hazard Ratio; IHC: Immunohistochemistry.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Leemans, C.R.; Braakhuis, B.J.M.; Brakenhoff, R.H. The molecular biology of head and neck cancer. *Nat. Rev. Cancer* **2011**, *11*, 9–22. [[CrossRef](#)] [[PubMed](#)]
3. Leemans, C.R.; Snijders, P.J.F.; Brakenhoff, R.H. The molecular landscape of head and neck cancer. *Nat. Rev. Cancer* **2018**, *18*, 269–282. [[CrossRef](#)] [[PubMed](#)]
4. Suh, Y.; Amelio, I.; Urbano, T.G.; Tavassoli, M. Clinical update on cancer: Molecular oncology of head and neck cancer. *Cell Death Dis.* **2014**, *5*, e1018. [[CrossRef](#)] [[PubMed](#)]
5. Mroz, E.A.; Tward, A.M.; Hammon, R.J.; Ren, Y.; Rocco, J.W. Intra-tumor Genetic Heterogeneity and Mortality in Head and Neck Cancer: Analysis of Data from the Cancer Genome Atlas. *PLoS Med.* **2015**, *12*, e1001786. [[CrossRef](#)] [[PubMed](#)]
6. Keren, S.; Shoude, Z.; Lu, Z.; Beibei, Y. Role of EGFR as a prognostic factor for survival in head and neck cancer: A meta-analysis. *Tumor Biol.* **2014**, *35*, 2285–2295. [[CrossRef](#)] [[PubMed](#)]
7. Johnson, D.E.; Burtneiss, B.; Leemans, C.R.; Lui, V.W.Y.; Bauman, J.E.; Grandis, J.R. Head and Neck Squamous Cell Carcinoma. *Nat. Rev. Dis. Primers* **2020**, *6*, 92. [[CrossRef](#)] [[PubMed](#)]
8. Madu, C.O.; Lu, Y. Novel diagnostic biomarkers for prostate cancer. *J. Cancer* **2010**, *1*, 150–177. [[CrossRef](#)]
9. D'Amico, A.V.; Chen, M.-H.; Roehl, K.A.; Catalona, W.J. Preoperative PSA Velocity and the Risk of Death from Prostate Cancer after Radical Prostatectomy. *N. Engl. J. Med.* **2004**, *351*, 125–135. [[CrossRef](#)]
10. Sørli, T.; Perou, C.M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10869–10874. [[CrossRef](#)]

11. Hammond, M.E.H.; Hayes, D.F.; Dowsett, M.; Allred, D.C.; Hagerty, K.L.; Badve, S.; Fitzgibbons, P.L.; Francis, G.; Goldstein, N.S.; Hayes, M.; et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* **2010**, *134*, e48–e72. [[CrossRef](#)] [[PubMed](#)]
12. Parker, J.S.; Mullins, M.; Cheang, M.C.; Leung, S.; Voduc, D.; Vickery, T.; Davies, S.; Fauron, C.; He, X.; Hu, Z.; et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* **2009**, *27*, 1160–1167. [[CrossRef](#)] [[PubMed](#)]
13. Bastien, R.R.; Rodriguez-Lescure, A.; Ebbert, M.T.; Prat, A.; Munarriz, B.; Rowe, L.; Miller, P.; Ruiz-Borrego, M.; Anderson, D.; Lyons, B.; et al. Pam50 Breast Cancer Subtyping by Rt-Qpcr and Concordance with Standard Clinical Molecular Markers. *BMC Med. Genom.* **2012**, *5*, 44. [[CrossRef](#)] [[PubMed](#)]
14. Van't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J.; Witteveen, A.T.; et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature* **2002**, *415*, 530–536. [[CrossRef](#)] [[PubMed](#)]
15. Walter, V.; Yin, X.; Wilkerson, M.D.; Cabanski, C.R.; Zhao, N.; Du, Y.; Ang, M.K.; Hayward, M.C.; Salazar, A.H.; Hoadley, K.A.; et al. Molecular Subtypes in Head and Neck Cancer Exhibit Distinct Patterns of Chromosomal Gain and Loss of Canonical Cancer Genes. *PLoS ONE* **2013**, *8*, e56823. [[CrossRef](#)]
16. Tang, K.-W.; Alaei-Mahabadi, B.; Samuelsson, T.; Lindh, M.; Larsson, E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun.* **2013**, *4*, 2513. [[CrossRef](#)] [[PubMed](#)]
17. Sopori, M. Effects of cigarette smoke on the immune system. *Nat. Rev. Immunol.* **2002**, *2*, 372–377. [[CrossRef](#)]
18. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [[CrossRef](#)]
19. Charlesworth, J.C.; Curran, J.E.; Johnson, M.P.; Göring, H.H.; Dyer, T.D.; Diego, V.P.; Kent, J.W.; Mahaney, M.C.; Almasy, L.; MacCluer, J.W.; et al. Transcriptomic epidemiology of smoking: The effect of smoking on gene expression in lymphocytes. *BMC Med. Genom.* **2010**, *3*, 29. [[CrossRef](#)]
20. De la Iglesia, J.V.; Slebos, R.J.C.; Martin-Gomez, L.; Wang, X.; Teer, J.K.; Tan, A.C.; Gerke, T.A.; Aden-Buie, G.; van Veen, T.; Masannat, J.; et al. Effects of Tobacco Smoking on the Tumor Immune Microenvironment in Head and Neck Squamous Cell Carcinoma. *Clin. Cancer Res.* **2020**, *26*, 1474–1485. [[CrossRef](#)]
21. Kumar, B.; Cordell, K.G.; Lee, J.S.; Worden, F.P.; Prince, M.E.; Tran, H.H.; Wolf, G.T.; Urba, S.G.; Chepeha, D.B.; Teknos, T.N.; et al. EGFR, p16, HPV Titer, Bcl-xL and p53, Sex, and Smoking as Indicators of Response to Therapy and Survival in Oropharyngeal Cancer. *J. Clin. Oncol.* **2008**, *26*, 3128–3137. [[CrossRef](#)] [[PubMed](#)]
22. Karam-Hage, M.; Cinciripini, P.M.; Gritz, E.R. Tobacco use and cessation for cancer survivors: An overview for clinicians. *CA Cancer J. Clin.* **2014**, *64*, 272–290. [[CrossRef](#)] [[PubMed](#)]
23. Zheng, J. Diagnostic value of MCM2 immunocytochemical staining in cervical lesions and its relationship with HPV infection. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 875–880. [[PubMed](#)]
24. Pietruszewska, W.; Bojanowska-Poźniak, K.; Kobos, J. Matrix metalloproteinases MMP1, MMP2, MMP9 and their tissue inhibitors TIMP1, TIMP2, TIMP3 in head and neck cancer: An immunohistochemical study. *Otolaryngol. Polska* **2016**, *70*, 29–40. [[CrossRef](#)] [[PubMed](#)]
25. Van Tubergen, E.A.; Banerjee, R.; Liu, M.; Broek, R.V.; Light, E.; Kuo, S.; Feinberg, S.E.; Willis, A.L.; Wolf, G.; Carey, T.; et al. Inactivation or Loss of TTP Promotes Invasion in Head and Neck Cancer via Transcript Stabilization and Secretion of MMP9, MMP2, and IL-6. *Clin. Cancer Res.* **2013**, *19*, 1169–1179. [[CrossRef](#)] [[PubMed](#)]
26. Muthuswamy, R.; Corman, J.M.; Dahl, K.; Chatta, G.S.; Kalinski, P. Functional reprogramming of human prostate cancer to promote local attraction of effector CD8⁺T cells. *Prostate* **2016**, *76*, 1095–1105. [[CrossRef](#)]
27. Lomnytska, M.I.; Becker, S.; Bodin, I.; Olsson, A.; Hellman, K.; Hellström, A.-C.; Mints, M.; Hellman, U.; Auer, G.; Andersson, S. Differential expression of ANXA6, HSP27, PRDX2, NCF2, and TPM4 during uterine cervix carcinogenesis: Diagnostic and prognostic value. *Br. J. Cancer* **2011**, *104*, 110–119. [[CrossRef](#)] [[PubMed](#)]
28. Li, H.; Chiappinelli, K.B.; Guzzetta, A.A.; Easwaran, H.; Yen, R.-W.C.; Vatapalli, R.; Topper, M.J.; Luo, J.; Connolly, R.M.; Azad, N.S.; et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget* **2014**, *5*, 587–598. [[CrossRef](#)]
29. Wenzel, J.; Tomiuk, S.; Zahn, S.; Küsters, D.; Vahsen, A.; Wiechert, A.; Mikus, S.; Birth, M.; Scheler, M.; von Bubnoff, D.; et al. Transcriptional profiling identifies an interferon-associated host immune response in invasive squamous cell carcinoma of the skin. *Int. J. Cancer* **2008**, *123*, 2605–2615. [[CrossRef](#)]
30. Zhu, C.; Chen, X.; Guan, G.; Zou, C.; Guo, Q.; Cheng, P.; Cheng, W.; Wu, A. IFI30 Is a Novel Immune-Related Target with Predicting Value of Prognosis and Treatment Response in Glioblastoma. *Oncotargets Ther.* **2020**, *13*, 1129–1143. [[CrossRef](#)]
31. Italiano, D.; Lena, A.M.; Melino, G.; Candi, E. Identification of Ncf2/P67phox as a Novel P53 Target Gene. *Cell Cycle* **2012**, *11*, 4589–4596. [[CrossRef](#)]
32. Chen, D.-T.; Nasir, A.; Culhane, A.; Venkataramu, C.; Fulp, W.; Rubio, R.; Wang, T.; Agrawal, D.; McCarthy, S.M.; Gruidl, M.; et al. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast Cancer Res. Treat.* **2010**, *119*, 335–346. [[CrossRef](#)] [[PubMed](#)]
33. Pyeon, D.; Newton, M.A.; Lambert, P.F.; den Boon, J.A.; Sengupta, S.; Marsit, C.J.; Woodworth, C.D.; Connor, J.P.; Haugen, T.H.; Smith, E.M.; et al. Fundamental Differences in Cell Cycle Deregulation in Human Papillomavirus-Positive and Human Papillomavirus-Negative Head/Neck and Cervical Cancers. *Cancer Res.* **2007**, *67*, 4605–4619. [[CrossRef](#)]

34. Ye, H.; Yu, T.; Temam, S.; Ziober, B.L.; Wang, J.; Schwartz, J.L.; Mao, L.; Wong, D.T.; Zhou, X. Transcriptomic dissection of tongue squamous cell carcinoma. *BMC Genom.* **2008**, *9*, 69. [[CrossRef](#)] [[PubMed](#)]
35. Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. affy—Analysis of Affymetrix Genechip Data at the Probe Level. *Bioinformatics* **2004**, *20*, 307–315. [[CrossRef](#)]
36. Cancer Genome Atlas Network. Comprehensive Genomic Characterization of Head and Neck Squamous Cell Carcinomas. *Nature* **2015**, *517*, 576–582. [[CrossRef](#)] [[PubMed](#)]
37. Stark, C.; Breitkreutz, B.-J.; Chatr-Aryamontri, A.; Boucher, L.; Oughtred, R.; Livstone, M.S.; Nixon, J.; Van Auken, K.; Wang, X.; Shi, X.; et al. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **2010**, *39*, D698–D704. [[CrossRef](#)]
38. Xenarios, I.; Salwinski, L.; Duan, X.J.; Higney, P.; Kim, S.-M.; Eisenberg, D. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **2002**, *30*, 303–305. [[CrossRef](#)]
39. Mewes, H.W.; Dietmann, S.; Frishman, D.; Gregory, R.; Mannhaupt, G.; Mayer, K.F.X.; Munsterkötter, M.; Ruepp, A.; Spannagl, M.; Stumpflen, V.; et al. MIPS: Analysis and annotation of genome information in 2007. *Nucleic Acids Res.* **2008**, *36*, D196–D201. [[CrossRef](#)]
40. Ceol, A.; Aryamontri, A.C.; Licata, L.; Peluso, D.; Briganti, L.; Perfetto, L.; Castagnoli, L.; Cesareni, G. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.* **2010**, *38*, D532–D539. [[CrossRef](#)]
41. Aranda, B.; Achuthan, P.; Alam-Faruque, Y.; Armean, I.; Bridge, A.; Derow, C.; Feuermann, M.; Ghanbarian, A.T.; Kerrien, S.; Khadake, J.; et al. The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* **2010**, *38*, D525–D531. [[CrossRef](#)] [[PubMed](#)]
42. Chen, Y.-C.; Lo, Y.-S.; Hsu, W.-C.; Yang, J.-M. 3D-partner: A web server to infer interacting partners and binding models. *Nucleic Acids Res.* **2007**, *35*, W561–W567. [[CrossRef](#)] [[PubMed](#)]
43. Chen, C.-C.; Lin, C.-Y.; Lo, Y.-S.; Yang, J.-M. PPISearch: A web server for searching homologous protein–protein interactions across multiple species. *Nucleic Acids Res.* **2009**, *37*, W369–W375. [[CrossRef](#)] [[PubMed](#)]
44. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
45. The Gene Ontology Consortium. The Gene Ontology Resource: 20 Years and Still Going Strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [[CrossRef](#)] [[PubMed](#)]
46. Kuperstein, I.; Bonnet, E.; Nguyen, H.-A.; Cohen, D.; Viara, E.; Grieco, L.; Fourquet, S.; Calzone, L.; Russo, C.; Kondratova, M.; et al. Atlas of Cancer Signalling Network: A systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* **2015**, *4*, e160. [[CrossRef](#)] [[PubMed](#)]
47. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
48. Wu, X.; Zhu, L.; Guo, J.; Zhang, D.-Y.; Lin, K. Prediction of yeast protein-protein interaction network: Insights from the Gene Ontology and annotations. *Nucleic Acids Res.* **2006**, *34*, 2137–2150. [[CrossRef](#)] [[PubMed](#)]
49. Mihály, Z.; Kormos, M.; Lánckzy, A.; Dank, M.; Budczies, J.; Szász, M.A.; Györfy, B. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast Cancer Res. Treat.* **2013**, *140*, 219–232. [[CrossRef](#)]
50. Desrichard, A.; Kuo, F.; Chowell, D.; Lee, K.-W.; Riaz, N.; Wong, R.J.; Chan, T.A.; Morris, L.G.T. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. *J. Natl. Cancer Inst.* **2018**, *110*, 1386–1392. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.