





Article

# Gene Expression and DNA Methylation in Human Papillomavirus Positive and Negative Head and Neck Squamous Cell Carcinomas

Snežana Hinić <sup>1,2</sup>, April Rich <sup>1,3</sup>, Nicole V. Anayannis <sup>1</sup>, Stephanie Cabarcas-Petroski <sup>4</sup> , Laura Schramm <sup>5</sup>   
and Patricio I. Meneses <sup>1,\*</sup>

<sup>1</sup> Department of Biological Sciences, Fordham University, Bronx, NY 10458, USA

<sup>2</sup> Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525GA Nijmegen, The Netherlands

<sup>3</sup> Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15213, USA

<sup>4</sup> Biology Department, Pennsylvania State University, Beaver Campus, Monaca, PA 15061, USA

<sup>5</sup> Department of Biological Sciences, St. John's University, Queens, NY 11439, USA

\* Correspondence: pmeneses@fordham.edu

**Abstract:** High-risk human papillomaviruses (HPV) are important agents, responsible for a large percentage of the 745,000 cases of head and neck squamous cell carcinomas (HNSCC), which were identified worldwide in 2020. In addition to being virally induced, tobacco and heavy alcohol consumption are believed to cause DNA damage contributing to the high number of HNSCC cases. Gene expression and DNA methylation differ between HNSCC based on HPV status. We used publicly available gene expression and DNA methylation profiles from the Cancer Genome Atlas and compared HPV positive and HPV negative HNSCC groups. We used differential gene expression analysis, differential methylation analysis, and a combination of these two analyses to identify the differences. Differential expression analysis identified 1854 differentially expressed genes, including *PCNA*, *TNFRSF14*, *TRAF1*, *TRAF2*, *BCL2*, and *BIRC3*. *SYCP2* was identified as one of the top deregulated genes in the differential methylation analysis and in the combined differential expression and methylation analyses. Additionally, pathway and ontology analyses identified the extracellular matrix and receptor interaction pathway as the most altered between HPV negative and HPV positive HNSCC groups. Combining gene expression and DNA methylation can help in elucidating the genes involved in HPV positive HNSCC tumorigenesis, such as *SYCP2* and *TAF7L*.

**Keywords:** cancer; head and neck cancer; gene expression; methylation; HPV



**Citation:** Hinić, S.; Rich, A.; Anayannis, N.V.; Cabarcas-Petroski, S.; Schramm, L.; Meneses, P.I. Gene Expression and DNA Methylation in Human Papillomavirus Positive and Negative Head and Neck Squamous Cell Carcinomas. *Int. J. Mol. Sci.* **2022**, *23*, 10967. <https://doi.org/10.3390/ijms231810967>

Academic Editor: Valentina Gatta

Received: 4 August 2022

Accepted: 14 September 2022

Published: 19 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Background

Head and neck squamous cell carcinomas (HNSCC) are a group of cancers from anatomically distinct areas: Oropharynx, larynx, hypopharynx, oral cavity, and tongue. HNSCC accounted for approximately 745,000 cancer cases worldwide in 2020, which is an alarming number [1,2]. Etiologic agents identified as causes of HNSCC include alcohol consumption and tobacco use, and high-risk human papillomavirus (HPV) infection [3,4]. Changes in sexual behavior have been found to be associated with higher HPV oral and oropharyngeal incidence and HPV is becoming increasingly indicated as one of the major HNSCC etiologic agents [5–9]. The predominant HPV genotype identified in HNSCC is HPV 16 (in as many as 90% of cases).

HPVs are non-enveloped, double-stranded DNA viruses with a genome that includes six early expressed genes, and two late expressed genes [10]. Two of the early genes, *E6* and *E7*, are characterized as oncogenes in cervical, oral, anal, and penile cases [11–13]. *E6* interferes with cell survival pathways by targeting p53 for proteasome degradation, and *E7* promotes cell proliferation interfering with the function of the Retinoblastoma protein (pRb) [11,12].

Genes involved in cancer development and progression can affect cell proliferation, metastasis, and invasion [14]. As a result of HPV infection, pathways that control cytoskeletal rearrangement, immune response, extracellular matrix formation, and receptor activation are differentially altered [14–16]. These genetic changes sustained during carcinogenesis and viral oncogenesis are a result of changes in gene expression and transcriptome profile. HNSCC onset, progression, and outcome differ depending on the presence or absence of HPV. In HPV negative (HPVN) HNSCC patients, the tumor suppressor genes *TP53* and *p16*, along with *CCND1* oncogene are the most frequently identified mutated genes [17,18]. In HPV positive (HPVP) HNSCC patients, the viral oncogenes *E6* and *E7* initiate deregulation by targeting p53 and pRb, respectively [11,12]. Studies have started to describe epigenetic profile changes, specifically on the level of DNA methylation, and it has been reported that the methylation status in HNSCC patients is associated with HPV infection (i.e., positive versus negative) [19–22].

Recently, there has been a growing interest and need for understanding the biological significance of HPVP and HPVN HNSCC. Many of these studies have utilized tools of the rapidly expanding field of bioinformatics [23–25].

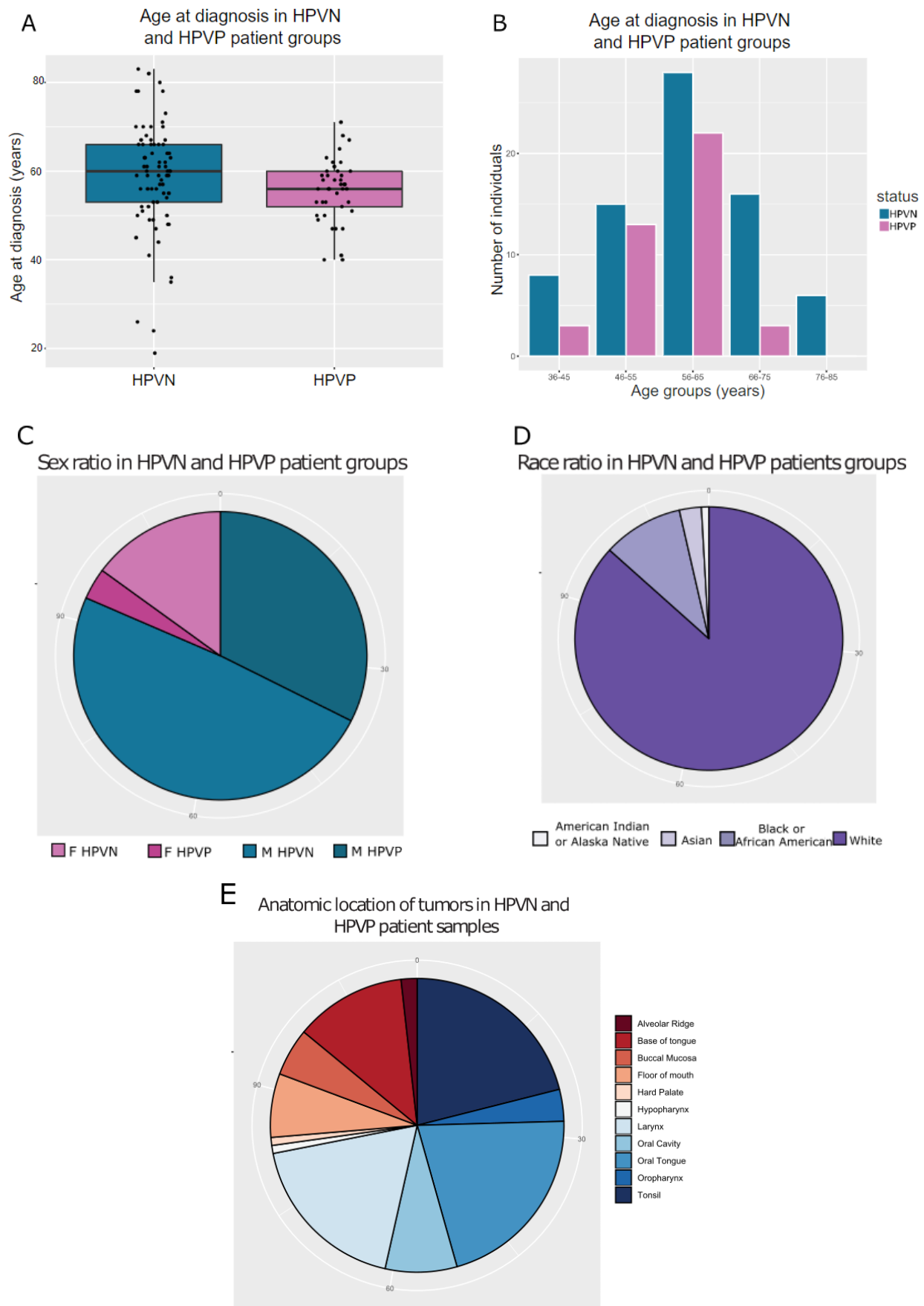
We performed a meta-analysis of The Cancer Genome Atlas (TCGA) HPVP and HPVN HNSCC transcriptome and DNA methylome data [26,27]. To our knowledge, this is the first time a study bridges these two datasets and compares groups based on the HPV status in HNSCC patient samples. Our findings show that pathways involved in viral invasion and trafficking, as well as immune system activation are differentially expressed in HPVP HNSCC. We identified that the differential expression of these pathways positively correlates with the differential methylation analysis.

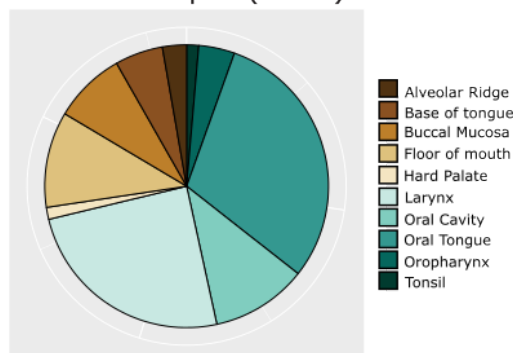
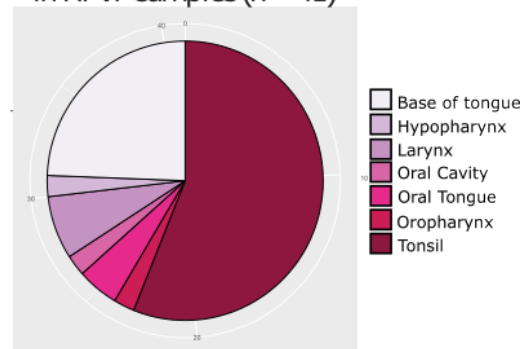
This study demonstrates the ability of computational methods to identify biomarkers of potential clinical significance from a centralized resource of available datasets, such as TCGA.

## 2. Results

### 2.1. TCGA HNSCC HPVP and HPVN Patients Have Comparable Clinical History

To ensure that the data from the TCGA database were comparable, we first examined the clinical profile of the patients in both HPVP and HPVN patient groups. Patient age distribution showed that the median values were comparable in both HPVN and HPVP groups (HPVN = 59, HPVP = 58 years) (Figure 1A). Most patients were grouped into 56–65 years of age range, and the collective (both for HPVP and HPVN groups) median age was in the same age range, as well (median = 58.5 years) (Figure 1B). We observed that there were no HPVP patients in the oldest patient category of 76–85 years of age (Figure 1B). Sex distribution in our sample groups revealed that females ( $n = 21$ ) were underrepresented in comparison to males ( $n = 93$ ) (Figure 1C). According to TCGA's classification in different race categories, race distribution showed that the white race was significantly the most represented one ( $n = 98$ ) (Figure 1D). Anatomical site analysis of these HNSCC showed that there were apparent differences in the location of tumor depending on the HPV status (Figure 1E–G). HPVP cancers were primarily found in the tonsil region and the base of the tongue, and HPVN cancers were primarily in the oral tongue and the larynx (Figure 1F,G).

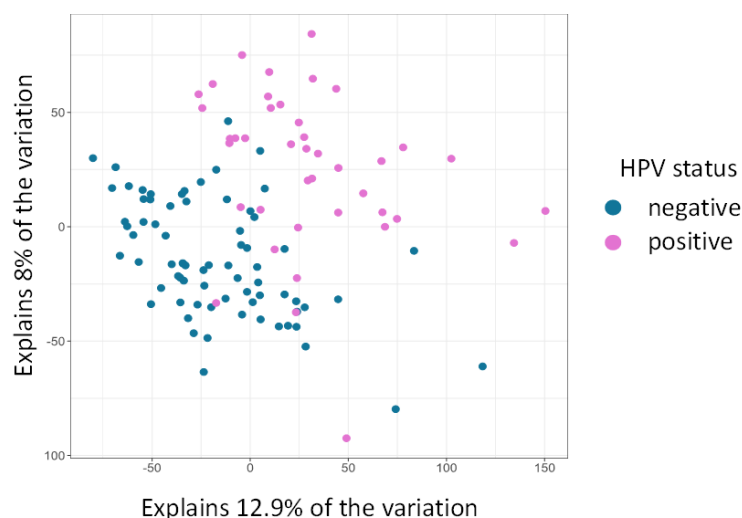


**F** Anatomic location of tumors in HPV<sup>+</sup> samples (n = 73)**G** Anatomic location of tumors in HPV<sup>-</sup> samples (n = 41)

**Figure 1.** Clinical data of the TCGA HNSCC HPV- positive and negative patients. Patients were filtered according to the HPV status using the information regarding *p16* expression and in situ hybridization information (only patients with information present were included in the study;  $n = 114$ , HPVP = 41, HPV<sup>+</sup> = 73). (A) Distribution of age at cancer diagnosis between two groups of patients, HPVP and HPV<sup>+</sup>; (B) distribution of patients in different age groups, and sidewise comparison of age groups and HPV status; (C) representation of male and female patients, HPVP and HPV<sup>+</sup>; (D) representation of different races, independent of sex or HPV status; (E) distribution of different anatomical sites where cancer originated; (F,G) a closer look at the specific location of HPV<sup>+</sup> patients (F) and HPVP (G).

## 2.2. Clustering of Samples Confirms That HPV<sup>+</sup> and HPVP Are Two Separate Comparable Groups

To explore clustering and similarity of samples, we analyzed the two experimental groups, HPV<sup>+</sup> and HPVP, by PCA and clustering on heatmap. PCA revealed that the two groups (detailed explanation in Material and Methods) clustered mostly as two separate and distinct groups with an overlapping middle area (Figure 2A, HPV<sup>+</sup> in blue, HPVP in pink).

**A** Clustering of the HPVP and HPV<sup>+</sup> HNSCC**Figure 2.** Cont.

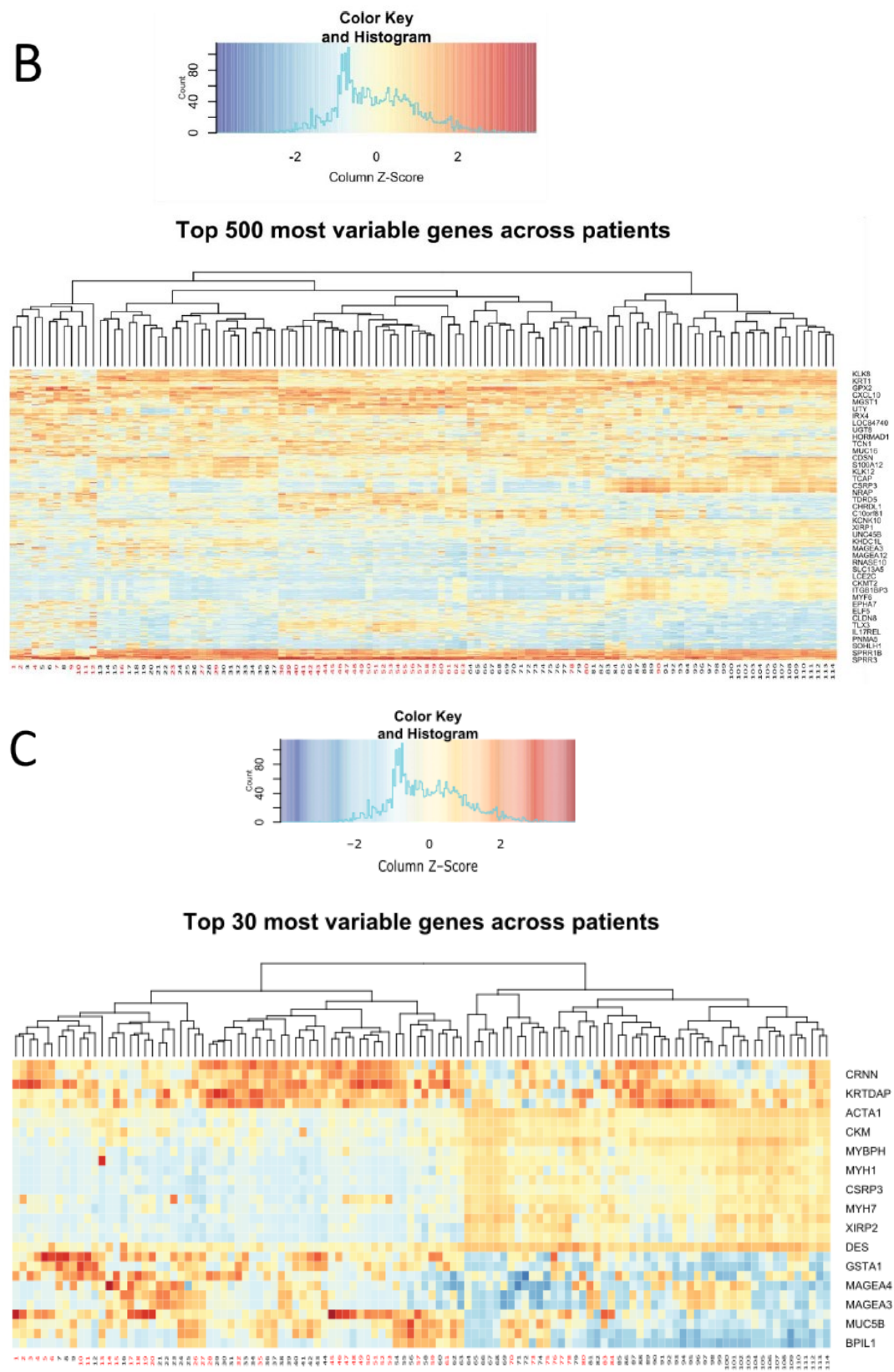
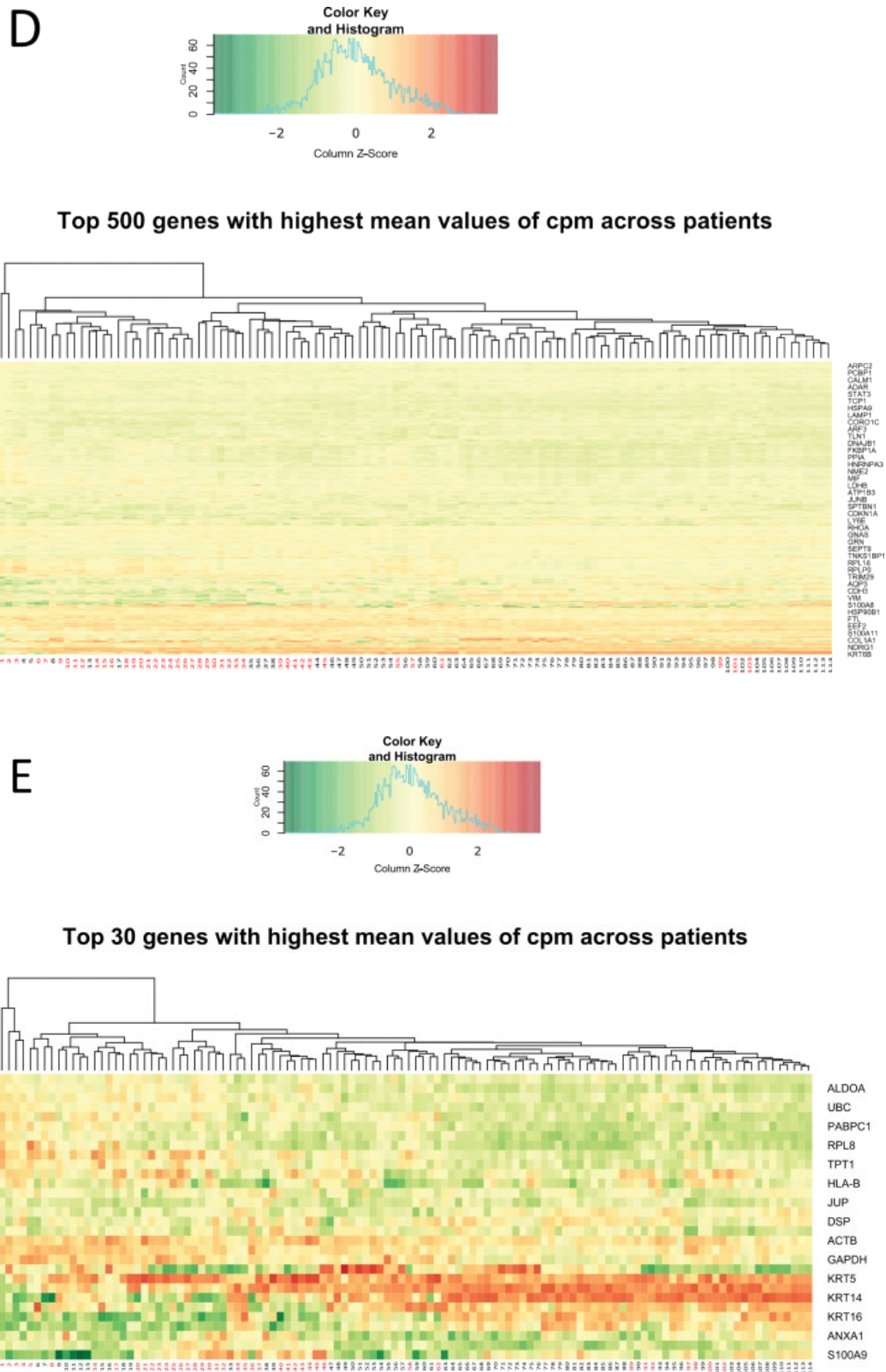


Figure 2. Cont.





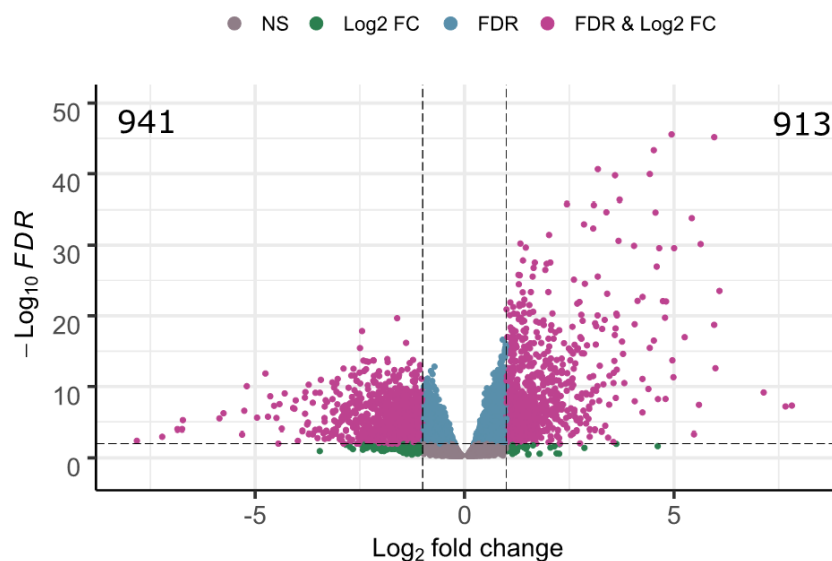
**Figure 2.** Clustering of the TCGA HNSCC HPVP and HPVn samples and genes. PCA and heatmap clustering shows distinct patient groups as we classify them. (A) PCA shows that patients classify in two separate groups for the most part, confirming that separation in HPVP and HPVn groups by *p16* expression and in situ hybridization was a valid parameter; (B–E) are heatmaps of the most variable

genes (B,C) and most abundant transcripts (D,E) among  $n = 114$  HNSCC samples; (B) shows the top 500 most variable genes, while a closer look at the top 30 most variable genes is shown in (C); top 500 transcripts with the highest mean values are depicted in (D) with a zoomed-in perspective to the top 30 in (E) HPV samples (labeled in black) and HPV (in red colored numbers).

Heatmaps (Figure 2B–E) revealed specific patterns:

1. The pattern of specific groups of genes in a larger scale analysis of the top 500 most variable genes (Figure 2B) and of the top 30 most variable genes (Figure 2C), remained consistent. Genes including kallikreins family genes (serine proteases) remained highly variable between the patients [28,29]. Keratin, a structural component was found as variable, as well as oxidative damage protection proteins (GPX2), cytokines, inflammatory response genes, immune response genes, and cell cycle controlling genes. Figure 2C depicts a gene responsible for stratification of the skin (*KRTDAP*), that was highly variably expressed, as well as an epithelial immune response and differentiation gene (*CRNN*). At a larger scale, genes from two groups of patients seemed to cluster mostly separately, significantly resembling the clustering observed in PCA (Figure 2A), with an intermediate overlapping cluster of samples (Figure 2B).
2. Top 500 and top 30 most abundant transcripts clustered mostly in two different groups (Figures 2D and 3E, respectively). Notably, some of the genes with the highest numbers of transcripts were cell cycle checkpoint genes, cytoskeletal regulatory genes, and immune response genes.

#### Differentially expressed genes between HPV and HPV HNSCC



**Figure 3.** Volcano plot depicting differentially expressed genes. Differential expression analysis identified 1854 DEGs, 941 were downregulated and 913 were upregulated between HPV and HPV HNSCC patient groups (using HPV as a baseline for comparison). Purple represents DEGs, blue is statistically significant according to the  $p$ -value, green is statistically significant according to the  $\log_{2}FC$ , while grey is not statistically significant.

#### 2.3. Transcriptome Analysis Identified 1854 Differentially Expressed Genes among HPV and HPV HNSCC Groups

To explore the impact of HPV on gene expression in HNSCC, we performed differential gene expression analysis (DGE) using TCGA Bioconductor package for R (Material and Methods). Using  $FDR \leq 0.01$  and  $|\log_{2}FC| \geq 1$ , with HPV as baseline, DGE identified 1854 differentially expressed genes (DEG), 941 downregulated and 913 upregulated in HPV samples (Figure 3). Significant DEGs are in purple, and genes that are non-significant

or significant by one of the parameters are in grey, blue, and green. Some of the key representative DEGs are: *PCNA*, *TNFRSF14*, *TRAF1*, *TRAF2*, *BCL2*, and *BIRC3*.

To functionally explain the up- and downregulated genes, we performed KEGG analysis [30]. Table 1 presents ten of the significantly enriched pathways (a full list of DEGs and KEGG pathways can be found in Supplementary Tables S1 and S3, respectively) [30,31]. The KEGG enriched pathways included those involved in ECM-interaction, cytokine production, cell cycle regulators, apoptosis, and genes identified as part of an HPV infection.

**Table 1.** Representative enriched KEGG pathways with some of the top enriched respective genes from the DEG pool.

| Representative KEGG Pathways                                  | Mapped Differentially Expressed Genes in HPVP vs. HPV N                        |
|---|--|
| ECM-receptor interaction                                      | CD36, ITGA6, ITGA5, ITGB3  |
| Focal adhesion  | BCL2, EGF, EGFR, ERBB2, IGF1, VEGFC  |
| Viral protein interaction with cytokine and cytokine receptor | IL18, IL18RAP, IL19, LTA, TNFRSF14, IL6  |
| Proteoglycans in cancer                                       | TP53, EGFR, ERBB2, IGF1  |
| Transcriptional misregulation in cancer                       | TP53, BCL2A1, CCNA1, CDKN2C, CSF2, GADD45G, ID2, IL6, MYCN, MEF2C, TLX3, TRAF1 |
| Human papillomavirus infection                                | CCNE2, CDK6, E2F1, PDGFRB, EGF, EGFR, TP53                                     |
| TNF signaling pathway   | BIRC3, CCL20, CSF2, IL15, IL6, TRAF1, TRAF2, VEGFC                             |
| Cell cycle  | BIRC3, BCL2, BCL2A1, NGF, TRAF1, TRAF2, PCNA, TP53, GADD45G                    |
| TGF-beta signaling pathway                                    | AMH, DCN, ID2, IFNG, INHBA, INHBB, LTBP1, NOG, THBS1                           |
| Apoptosis   | BCL2, BCL2A1, BIRC3, GADD45G, NGF, TP53, TRAF1, TRAF2                          |

Pathway and ontology analyses were performed using the Enrichr and PANTHER classification systems (Tables 2 and 3) [32–34]. These tools identified similar pathways and patterns as KEGG (Tables 2–4, Supplementary Table S2). The top KEGG significantly enriched pathways (i.e., enrichment of genes) were consistent with HPV infection (Table 4). Notably, transcription factors and genes involved in cell cycle progression were identified as upregulated. In contrast, genes involved in cellular response to stimulus, including chemotherapeutic agents and radiation were downregulated [35].

**Table 2.** Representative enriched pathways, ontologies, and transcription factors filtered by A. Enrichr and B. PANTHER (from November 2019).

| Category                                       | Regulation Level | q-Value               | Database                          |
|--|------------------|-----------------------|-----------------------------------|
| <b>Transcription Factors</b>                   |                  |                       |                                   |
| NFKB   | upregulated      | $3.29 \times 10^{-2}$ | TRANSFAC and JASPAR PWMs          |
| SP1 human                                      | downregulated    | $2.37 \times 10^{-6}$ | TRRUST Transcription Factors 2019 |
| <b>Pathways</b>                                |                  |                       |                                   |
| Retinoblastoma gene in cancer WP2446           | upregulated      | $4.58 \times 10^{-6}$ | WikiPathways 2019 Human           |
| DNA strand elongation Homo Sapiens R-HAS-69190 | upregulated      | $2.02 \times 10^{-4}$ | Reactome 2016                     |



Table 2. Cont.

| Category   | Regulation Level | q-Value                | Database                     |
|--|------------------|------------------------|------------------------------|
| Beta1 integrin cell surface interactions<br>Homo Sapiens | downregulated    | $1.25 \times 10^{-15}$ | NCI-Nature 2016              |
| ITGB1  | downregulated    | $5.64 \times 10^{-5}$  | PPI Hub Proteins             |
| Integrin signaling pathway<br>Homo Sapiens               | downregulated    | $2.02 \times 10^{-6}$  | PANTHER 2016                 |
| <b>Ontologies</b>  |                  |                        |                              |
| G1/S transition in mitotic cell cycle<br>(GO:0000082)    | upregulated      | $3.04 \times 10^{-2}$  | GO Biological Processes 2018 |
| T cell receptor complex (GO:0042101)                     | upregulated      | $7.57 \times 10^{-5}$  | GO Cellular Component 2018   |
| Collagen binding (GO:0005518)                            | downregulated    | $1.70 \times 10^{-8}$  | GO Molecular Function 2018   |

Table 3. Representative enriched pathways, ontologies, and transcription factors filtered by PANTHER (from November 2019).

| Upregulated Processes                                   | Downregulated Processes                                  |
|---|--|
| Category: Biological Processes                          | Category: Biological Processes                           |
| Cellular Process was top hit (GO:0009987) 311/913 genes | Cellular Process was top hit (GO:0009987) 344/941 genes  |
| subcategory   | subcategory  |
| cell cycle (GO:0007049) 44/311 genes                    | cellular response to stimulus (GO:0051716) 110/344 genes |

Table 4. Top 9 most statistically significant KEGG pathways obtained with the DEG from the differential expression analysis.

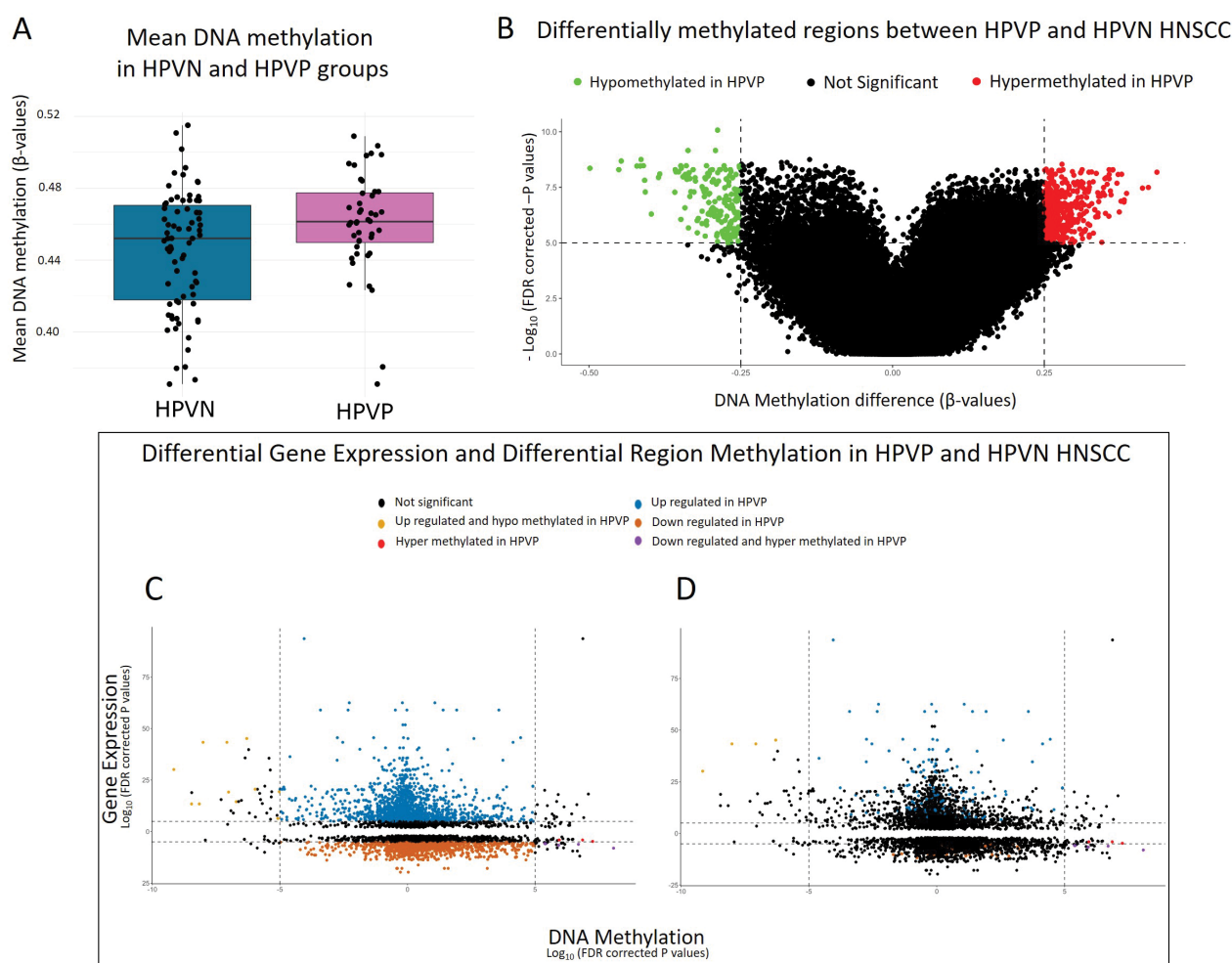
| ID       | Pathway   | Gene Ratio | q-Value                |
|----------|---|------------|------------------------|
| hsa04512 | ECM-receptor interaction                                      | 32/714     | $3.47 \times 10^{-10}$ |
| hsa04060 | Cytokine-cytokine receptor interaction                        | 61/714     | $3.38 \times 10^{-8}$  |
| hsa04640 | Hematopoietic cell lineage                                    | 30/714     | $5.86 \times 10^{-8}$  |
| hsa04974 | Protein digestion and absorption                              | 29/714     | $1.19 \times 10^{-7}$  |
| hsa04510 | Focal adhesion  | 44/714     | $6.22 \times 10^{-7}$  |
| hsa05410 | Hypertrophic cardiomyopathy (HCM)                             | 25/714     | $8.20 \times 10^{-6}$  |
| hsa04151 | PI3K-Akt signaling pathway                                    | 61/714     | $1.42 \times 10^{-5}$  |
| hsa04061 | Viral protein interaction with cytokine and cytokine receptor | 26/714     | $1.47 \times 10^{-5}$  |
| hsa05150 | Staphylococcus aureus infection                               | 24/714     | $6.90 \times 10^{-5}$  |

#### 2.4. DNA Methylome Analysis Showed HPVP and HPV N HNSCC Methylation Levels Were Comparable

To explore epigenetic changes in HNSCC due to HPV, we focused on DNA methylation. We performed a differential methylation analysis (DMA) using the following parameters:  $\bar{\beta} \geq 0.25$  and  $p \leq 10^{-5}$  that identified top hypo- and hypermethylated regions of the genome and genes involved (Table 5, Supplementary Table S3). We compared the overall median methylation levels of our two groups of patient samples, HPV N and HPVP, and observed that their median values were comparable ( $\bar{\beta} \sim 0.46$ ) (Figure 4A). DMA results are represented on a volcano plot comparing hypomethylated and hypermethylated regions in HPVP (HPV N samples used as baseline comparison) (Figure 4B).

**Table 5.** Top 5 hypo- and hypermethylated genes among HPV and HPVP HNSCC patients. HPV is used as a baseline.

| Probe ID   | Gene Symbol    | Adjusted <i>p</i> -Value | Status in HPVP  |
|------------|----------------|--------------------------|-----------------|
| cg11456145 | CDC42EP5       | $8.40 \times 10^{-11}$   | Hypomethylated  |
| cg07915849 | ABCA17P;ABCA3  | $6.97 \times 10^{-10}$   | Hypomethylated  |
| cg10504436 | DERL3          | $6.97 \times 10^{-10}$   | Hypomethylated  |
| cg12181372 | SYCP2          | $1.77 \times 10^{-9}$    | Hypomethylated  |
| cg22220310 | SDF4;B3GALT6   | $2.08 \times 10^{-9}$    | Hypomethylated  |
| cg07907859 | FAM133A;NAP1L3 | $2.90 \times 10^{-9}$    | Hypermethylated |
| cg00757182 | ZNF773         | $4.81 \times 10^{-9}$    | Hypermethylated |
| cg12387713 | MSX2           | $5.08 \times 10^{-9}$    | Hypermethylated |
| cg13458645 | PITX2          | $5.08 \times 10^{-9}$    | Hypermethylated |
| cg11876013 | SCHIP1         | $5.29 \times 10^{-9}$    | Hypermethylated |

**Figure 4.** Differential methylation in HNSCC patients. Represented above is the methylation profile in HNSCC. (A) Mean methylation between HPV and HPVP HNSCC patient samples; (B) volcano plot showing the hypomethylated genes in green and hypermethylated genes in red. HPV samples are used as a baseline. We used  $\beta \geq 0.25$  and  $p \leq 10^{-5}$ ; (C,D) show a Starburst plot that combined differential gene expression data with differential methylation data. HPV is used as a baseline. We used  $\beta \geq 0.25$ ,  $\text{FDR}_{\text{expression}} \leq 10^{-5}$ ,  $\text{FDR}_{\text{DNA methylation}} \leq 10^{-5}$   $|\log_{2}\text{FC}| \geq 1$  in (C) and more stringent parameters  $\beta \geq 0.25$ ,  $\text{FDR}_{\text{expression}} \leq 10^{-5}$ ,  $\text{FDR}_{\text{DNA methylation}} \leq 10^{-5}$   $|\log_{2}\text{FC}| \geq 3$  in (D).

### 2.5. Starburst: An Analysis That Bridges Differentially Expressed and Methylated Genes Revealed Similar Patterns to DNA Methylome Analysis and Potential Biomarker Gene for HPVP HNSCC

To identify common DEG and DMA genes, we performed a Starburst analysis [36]. This analysis identifies genes with similar DEG and DMR patterns (i.e., hypomethylated and upregulated and hypermethylated and downregulated), using the following parameters:  $\beta \geq 0.25$ ,  $FDR_{expression} \leq 10^{-5}$ ,  $FDR_{DNAmethylation} \leq 10^{-5}$ . Our analysis showed that a similar pattern was observed with  $|\log FC|$ , which is set to  $\geq 1$ , and more stringent  $|\log FC|$  set to be  $\geq 3$ . The pattern of DEG and DMR expression remained comparable with both parameters used, and the top statistically significant DEG and DMR identified in both analyses were consistent (Figure 4C,D). We decided to proceed with  $|\log FC| \geq 1$  and depict some of the representative results (Table 6), and a complete list can be found in Supplementary Table S5.

**Table 6.** Representative top differentially expressed genes and differentially methylated regions filtered by Starburst (FDR cutoff = 1; HPV group is used as a baseline).

| Gene Symbol     | Status in HPVP  | Gene Name   |
|-----------------|-----------------|---|
| TAF7L           | Hypomethylated  | TATA-Box Binding Protein Associated Factor 7 Like |
| SYCP2           | Hypomethylated  | Synaptonemal Complex Protein 2                    |
| LOC285954;INHBA | Hypermethylated | Inhibin Subunit Beta A                            |
| SULF1           | Hypermethylated | Sulfatase 1                                       |
| CCNA1           | Hypermethylated | Cyclin A1   |

### 3. Discussion

HPV has been recognized as an important driver of HNSCC [23,37,38]. The patient treatment varies depending on HPV positive (HPVP) versus negative (HPVN) HNSCC; therefore, it is important to gain further knowledge of the genetic profile of HNSCC. Our study showed that HPVP HNSCC patients exhibit gene deregulation at gene transcription and methylation levels different from HPV HNSCC patients. When analyzed, both independently and collectively, gene expression and methylation deregulation patterns specifically point out changes in gene pathways including those involved in controlling invasion, immune response, differentiation, and cell division.

In total, the cohort of patient's samples analyzed was 114 (HPVN = 73 and HPVP = 41). There was a disparity in the male/female self-described sample ratio, where male samples accounted for 93, and female samples the remaining 21 (Figure 1C). A possible explanation for this disparity might be that HNSCC cases are sex biased and more prevalent in males, but a larger cohort needs to be analyzed to address this disparity. Moreover, there was an overrepresentation of white race ( $n = 98$ ) in this cohort for HNSCC (Figure 1D). This lack of racial representation is unfortunately not uncommon in clinical studies. We have since identified studies that report HNSCC incidence in non-white population, and a similar analysis will be conducted in the future to include more equally distributed races [39–42]. There was an apparent absence of HPVP HNSCC patients in the oldest patient category (76–85 years of age—Figure 1B), and we theorize that might be due to the fact that HPVP HNSCC are significantly more rare than HPV patients, thus causing this age groups' underrepresentation. Alternatively, the HPVP HNSCC patients do not survive for a long period to be included in the data (76–85 years of age) [43,44]. We observed differences in anatomical sites of HNSCC that were dependent on the HPV status (Figure 1E–G). Tonsil was the predominant location in HPVP patients, while the oral tongue had the most cases in HPV patients (Figure 1F,G). In the US, regardless of HPV status, the oral tongue is the most common site for HNSCC [39].

In our analysis, genes that play a role in all HNSCC development belonged to four main functional pathways: Cell survival, cellular proliferation, squamous epithelial differentiation, and invasion/metastasis. We identified differentially expressed and methylated genes in HPVP versus HPV HNSCC. Of the 1854 DEG, 16 genes were the top hits identi-

fied in the transcriptome and methylome analyses. The functions of these genes range from cell cycle, immune response, to cell death regulation. Specifically, we found that *SYCP2* and *TAF7L* were the two most deregulated genes in both analyses. Synaptonemal complex protein 2 (*SYCP2*) was the top hypomethylated and upregulated gene in HPVP HNSCC. This gene is the testis-specific human gene and has been associated with impaired meiosis [45]. It is known that *SYCP2* aberrant expression in HPVP cancers may contribute to the genomic instability induced by high-risk HPVs and subsequent oncogenic change [46]. In 2015, a paper by Masterson et al. reported that deregulation of *SYCP2* predicts early-stage human papillomavirus-positive oropharyngeal carcinoma. The same authors concluded their study by proposing *SYCP2* as a potential biomarker [47]. In addition, an independent study showed that *SYCP2* was hypomethylated in HPVP HNSCC, which is in concordance with what we have discovered [19]. This might imply that the previously proposed biomarker function for *SYCP2* is not unlikely. In addition to these reports, the elevated expression of *SYCP2* in HPV-associated tumors has previously been observed in three additional gene expression analysis studies [48–50]. The human protein atlas reports the highest expression of *SYCP2* in male tissues, while this protein is also expressed in female tissues, although less (<https://www.proteinatlas.org/ENSG00000196074-SYCP2/tissue>, accessed on 13 September 2022). All of this suggests that *SYCP2* is involved in more than its primary function as the synaptonemal complex protein when deregulated. Additional research is needed to determine the significance of *SYCP2* levels in male and female samples. Similarly, the second highlighted gene that was hypomethylated and upregulated in HPVP HNSCC is TATA-box binding protein associated factor 7-like (*TAF7L*), a gene involved in spermatogenesis [51]. According to a study by Mobasher et al., *TAF7L* is upregulated in breast cancer; therefore, it is possible that it is not an exclusive feature, which is observed only in breast cancer tissue [52].

DEG analysis identified that *PCNA*, *TNFRSF14*, *TRAF1*, *TRAF2*, *BIRC3*, and *BCL2* were significantly altered in HPVP HNSCC.

Proliferating cell nuclear antigen (*PCNA*) is a gene that was significantly overexpressed in HPVP versus HPV N HNSCC patient samples. It has been shown that *PCNA* expression levels change during cell cycle, as *PCNA* is associated with proliferation and cell transformation in cancer [53,54]. *PCNA* is one of the crucial regulators in cell cycle as it forms complexes with cell cycle activators (cyclins and cyclin dependent kinases) and inhibitors (*p21*) [53]. Post-translational modifications are crucial for the *PCNA* function, significantly, that *PCNA* exists in an alternative methylated form in cancers [55].

Tumor necrosis factor receptor superfamily member 14 (*TNFRSF14*) is known to be a herpesvirus entry mediator by being a part of signal transduction pathways that activate inflammatory and inhibitory T-cell immune response [56]. It is not surprising to observe that it was upregulated in HPVP HNSCC, although it is interesting that a herpesvirus-related gene has been upregulated upon HPV infection in this cancer type. *TNFRSF14* is known to interact with TNF receptor associated factor 2 (*TRAF2*), which is also upregulated in HPVP HNSCC. This protein directly interacts with the TNF receptors, and forms a complex with another TRAF family member, *TRAF1* which is also upregulated in HPVP HNSCC. This is all necessary for *TNF $\alpha$* -mediated activation of *MAPK8/JNK* and *NF- $\kappa$ B*, which are known to be involved in cell survival. The protein complex formed by *TRAF2* and *TRAF1* interacts with the inhibitor-of-apoptosis proteins (IAPs), and functions as a mediator of the anti-apoptotic and pro-survival signals from TNF receptors. One of those IAPs that is upregulated in HPVP HNSCC is *BIRC3*-apoptosis inhibitor [57–59]. According to The Human Protein Atlas (THPA), *TRAF2* has the highest expression in HNSCC, followed by cervical cancer among all sampled cancer types (17 cancer types) [60]. *BIRC3* shows similar observations, implying that this pattern may be specific for HPV-related HNSCC [60]. Another role of *TRAF1* is a negative regulation of Toll-like receptor (TLR) and Nod-like receptor (NLR) signaling. *TRAF1* can also, independently from *TRAF2*, contribute to *NF- $\kappa$ B* activation; conversely, during TLR and NLR signaling, *TRAF1* can also negatively regulate *NF- $\kappa$ B* activation. According to THPA, *TRAF1* has been found to be

overexpressed in HNSCC. Additionally, *TRAF1* can contribute to chronic viral infection and limit inflammation, contributing to the survival of Epstein-Barr virus dependent cancers [57,60]. TRAF family genes (*TRAF1* and *TRAF2*, specifically) have been found to be differentially expressed in a couple of HPV-related studies, including one in our lab [61,62]. An interesting question follows: Does *TRAF1* have a similar role in HPV-dependent cancers, as well? To investigate this, more research is required.

In addition to *BIRC3*-apoptosis inhibitor which is upregulated in HPVP HNSCC, *BCL-2*, an anti-apoptotic gene has been observed to be upregulated in HPVP HNSCC, as well. An existing model explains the observed picture in our data. Similarly to oncogene addiction, some tumor cells may be dependent on *BCL-2* for survival [63]. As tumor environment may induce higher stress signal production that is pro-apoptotic in nature, a proportion of cancer cells manage to overexpress *BCL-2* and survive the production of this anti-apoptotic signal. In this way, *BCL-2* helps cancer progression by promoting the survival of altered cells [64,65]. Moreover, *BCL-2* is known to be overexpressed in non-hematologic tumors as ovarian, neuroblastoma, colorectal, and HNSCC [66–69].

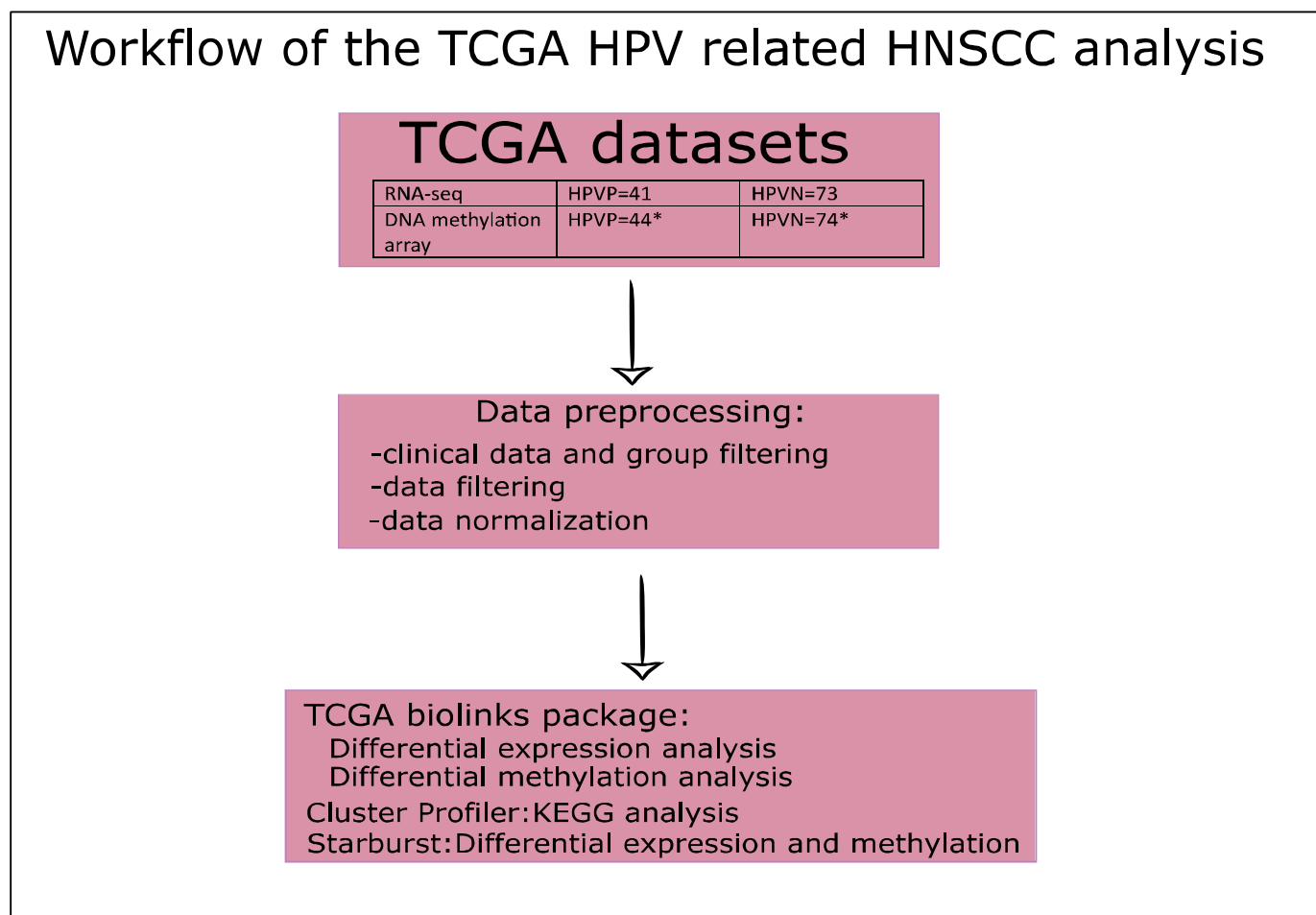
Starburst analysis combined DEG and DMR results and highlighted genes that were the most hypomethylated and upregulated and the most hypermethylated and downregulated. We performed Starburst with FDR cutoff = 1 and a more stringent parameter FDR cutoff = 3 and maintained the top highlighted gene profile (Supplementary Table S4, and Figure 4C,D), specifically *SYCP2* and *TAF7L*. Considered together, some of the DEG identified as top hits may be used as potential biomarkers for early identification of HPVP HNSCC, including *SYCP2*, *TAF7L*, and *ZFR2*. The analysis of DEG of tonsil HPVP HNSCC and oral tongue HPV HNSCC (predominant anatomical locations of samples), identified unique genes that were downregulated in HPVP tonsil HNSCC (Supplementary Table S5). Of these genes, *RBM24*, is shown to mediate repression of p53/TP53 mRNA translation and *INHBA*, a member of the transforming growth factor-beta (*TGF-β*) superfamily of proteins. According to THPA, the highest expression of *RBM24* is observed in HNSCC, followed by cervical cancer, although we have not seen its use as a diagnostic tool [60,70]. This implies that when these genes are downregulated, this might specifically indicate HPVP HNSCC site specific (tonsil) cancer development.

## 4. Methods

### 4.1. Study Design, Patient Samples, and Analysis Workflow

In this study, data were acquired through the publicly available database TCGA and NCI Genomic Data Commons (GDC) [26,27]. We focused on HNSCC tumor data, and all data used in this study were open access (downloaded in 2019). We grouped the HNSCC patient samples in two experimental groups: (1) HPVP HNSCC, and (2) HPV HNSCC. We were interested in comparing gene expression and methylation state of tumors in the absence or presence of HPV. The TCGA gene expression and DNA methylome data were extracted from RNA-seq studies of HNSCC, and from DNA methylation arrays, respectively. Moreover, we requested corresponding clinical data [27]. We used the clinical information to filter the samples into HPVP or HPV HNSCC. We used two criteria to determine the presence of HPV: (1) The expression levels of *p16* gene, a well-known tumor-suppressor gene indicative of high-risk HPV-related cancers [71]; and (2) we used the in situ hybridization information for *p16* gene if the expression information of *p16* was not available. Using these criteria, we were able to acquire the information from 73 HPV patients and 41 HPVP patients from the transcriptome studies, and 74 HPV and 44 HPVP patients from the DNA methylome studies (detailed list of patients in Supplementary Table S6). For the patients that we had RNA-seq data available, we performed the analysis on clinical status, as well. To visualize clinical data, we used *gplots*, *ggplot2*, *RColorBrewer*, and *colorRamps* Bioconductor packages [72–75]. TCGA data consisted of already mapped reads that were downloaded using Bioconductor's package *TCGAbiolinks* for TCGA data handling. R (version 3.6.1) and RStudio software were used for all data analyses [36,76–80]. Figure 5 shows our overall workflow, with each part described in detail in the following sections.





**Figure 5.** Workflow of the TCGA HPV-related HNSCC data. A schematic representation of the stepwise workflow of TCGA data analysis. \* Clinical data analysis was performed only on RNA-seq patients' data, and not on DNA methylation data.

#### 4.2. Data Preprocessing to Normalize Data

We preprocessed and filtered the data according to the parameters of HPV status. Preprocessing makes the data as uniform as possible, rearranges, and enables it for the analysis software to handle it. Moreover, we normalized the data to be able to perform subsequent clustering steps. Data were filtered using TCGAbiolinks and xlsx packages, and used embedded functions TCGAanalyze\_Preprocessing, TCGAanalyze\_Normalization, and TCGAanalyze\_Filtering [36,81].

#### 4.3. Data Clustering Analyses

To investigate whether clustering was as expected (HPVP versus HPVN HNSCC), principal component analysis (PCA) and hierarchical clustering with heatmaps using edgeR and gplots packages, and heatmap.2 function in R were performed [75,82,83]. For the PCA analysis, we used prcomp function already existent in R, and for the hierarchical clustering with heatmaps, we used edgeR package for R and gplots, ggplot2, and RColorBrewer libraries for data visualization throughout the analyses [72,73,75,82]. For heatmap clustering, we followed a recommended online tutorial [84]. Using heatmap clustering, we investigated the most variable transcripts, as well as the genes that have the highest mean values across 114 patients, using it as a proxy for the most abundant transcripts.

#### 4.4. Transcriptome Analysis: Differential Gene Expression Analysis (DGE) and Pathway Analysis

To understand differential gene expression of the filtered data, a DGE analysis was performed using TCGAbiolinks TCGAanalyze\_DEA function. We used a false discovery rate (FDR) cutoff of 0.01, which represents a threshold to filter DEGs according to their corrected  $p$ -value. Moreover, a probe expression fold change (logFC) cutoff of 1 was used. To understand the nature of the extracted deregulated genes, we performed a pathway analysis using clusterProfiler Bioconductor package, and the function enrichKEGG, along with packages SummarizedExperiment, MultiAssayExperiment, and genefilter [85–88]. To visualize the identified pathways, we used pathview Bioconductor package, and to visualize DEG in a volcano plot we used EnhancedVolcano Bioconductor package [31,89]. We used PANTHER (Protein ANalysis THrough Evolutionary Relationships) and Enrichr, two comprehensive gene set enrichment analysis tools, to investigate the enriched pathways in the DEG dataset [32–34].

#### 4.5. DNA Methylation Analysis: Differential Methylation Analysis (DMA)

To analyze the DNA methylation patterns, we used TCGAbiolinks function TCGAanalyze\_DMR, and used  $p$ -value cutoff =  $10^{-5}$  and  $\beta \geq 0.25$ . “ $\beta$ ” is a parameter for differential methylation levels that ranges between 0 and 1, 0 being unmethylated and 1 being fully methylated.

#### 4.6. Starburst Analysis: Integrative Analysis of DEG and Differentially Methylated Regions (DMR)

To observe common patterns of gene silencing or overexpression, we combined the two datasets (DEG and DMR) using TCGAbiolinks TCGAvisualize\_starburst function [36]. We used  $\beta \geq 0.25$ ,  $\text{FDR}_{\text{expression}} \leq 10^{-5}$ ,  $\text{FDR}_{\text{DNAmethylation}} \leq 10^{-5}$ , and  $|\log\text{FC}| \geq 1$ . Moreover, we tested the data with a more stringent parameter of  $|\log\text{FC}| \geq 3$ , and decided to work with the former parameter, as the analysis demonstrated that the most prominent genes were filtered under both parameters.

### 5. Conclusions

In conclusion, using TCGA transcriptome data enabled us to identify 1854 DEG, and these DEG belong to a wide range of pathways, including cell cycle, papillomavirus infection, transcriptional misregulation, TNF signaling, cytoskeletal rearrangement, and apoptosis. Combining the knowledge gained, both by transcriptome and DNA methylome data analyses, we identified potential players that might contribute to cancer development in HPVP HNSCC. In particular, SYCP2 and TAF7L, which have been shown in the past to be deregulated in cancer development [46,47,52]. SYCP2 specifically attracts our attention, as it has been shown that deregulation of SYCP2 predicts early stage HPVP oropharyngeal carcinoma and it has been proposed to serve as a biomarker by other authors [47]. Moreover, we propose a potential panel of genes to serve for HPVP HNSCC detection and possible anatomical characterization. Screening for circulating tumor DNA from peripheral blood is low invasive and provides fast results, and we suggest screening for HPVP HNSCC using a panel, including RBM24, INHBA, SYCP2, TAFL7, and ZFR2. This may serve as an informative tool for HNSCC HPVP screening, and even for the detection of the specific anatomical location.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231810967/s1>.

**Author Contributions:** Conceptualization, L.S. and P.I.M.; data curation, S.H., A.R., S.C.-P. and L.S.; formal analysis, S.H., A.R., N.V.A., S.C.-P. and P.I.M.; investigation, S.H., N.V.A., L.S. and P.I.M.; methodology, S.H., A.R., L.S. and P.I.M.; project administration, P.I.M.; resources, P.I.M.; supervision, P.I.M.; validation, S.C.-P. and L.S.; writing—original draft, S.H., A.R. and P.I.M.; writing—review and editing, S.H., A.R., N.V.A., S.C.-P., L.S. and P.I.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the university provided funding to P.I.M., no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** All data used in this study are open access and obtained from the publicly available NIH GDC legacy archive.

**Data Availability Statement:** Data used in this study are open access and can be found on NIH GDC Legacy Archive under the following link: <https://portal.gdc.cancer.gov/legacy-archive/search/f>, accessed on 30 August 2019.

**Acknowledgments:** We thank the NIH/GDC.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

|                |  |
|----------------|--|
| HPV            | Human Papillomavirus   |
| HNSCC          | head and neck squamous cell carcinoma                          |
| pRb            | Retinoblastoma protein   |
| PCNA           | proliferating cell nuclear antigen                             |
| TNFRSF14       | Tumor Necrosis Factor (TNF) Receptor Superfamily Member 14     |
| TRAF1          | TNF Receptor Associated Factor 1                               |
| TRAF2          | TNF Receptor Associated Factor 2                               |
| BCL2           | B-cell lymphoma 2 apoptosis regulator                          |
| BIRC3          | Baculoviral Inhibitor of Apoptosis (IAP) Repeat Containing 3   |
| SYCP2          | Synaptonemal Complex Protein 2                                 |
| TAF7L          | TATA-Box Binding Protein Associated Factor 7-Like              |
| E6             | HPV early gene 6   |
| E7             | HPV early gene 7   |
| TP53/p53       | tumor suppressor gene/protein p53                              |
| HPVN           | HPV negative HNSCC   |
| HPVP           | HPV positive HNSCC   |
| p16            | cyclin-dependent kinase inhibitor                              |
| CCND1          | cyclin D1  |
| TCGA           | The Cancer Genome Atlas  |
| GDC            | NCI Genomic Data Commons                                       |
| PCA            | Principal Component Analysis                                   |
| DGE            | differential gene expression                                   |
| DMA            | differential methylation analysis                              |
| DMR            | differential methylated regions                                |
| GPX2           | Glutathione peroxidase 2                                       |
| KRTDAP         | Keratinocyte Differentiation Associated Protein                |
| CRNN           | Cornulin   |
| PCA            | Principal component analysis                                   |
| KEGG           | Kyoto Encyclopedia of Genes and Genomes                        |
| ECM            | extracellular matrix   |
| PANTHER        | protein annotation through evolutionary relationship           |
| THPA           | The Human Protein Atlas  |
| TLR            | Toll-like receptor   |
| NLR            | Nod-like receptor  |
| NF- $\kappa$ B | nuclear factor kappa-light-chain-enhancer of activated B cells |
| ZFR2           | Zinc Finger RNA Binding Protein 2                              |
| INHBA          | Inhibin Subunit Beta A   |

## References

1. Global Cancer Observatory. Available online: <https://gco.iarc.fr/> (accessed on 30 August 2022).
2. Worldwide Cancer Data | World Cancer Research Fund International. Available online: <https://www.wcrf.org/cancer-trends/worldwide-cancer-data/> (accessed on 30 August 2022).
3. Jethwa, A.R.; Khariwala, S.S. Tobacco-related carcinogenesis in head and neck cancer. *Cancer Metastasis Rev.* **2017**, *36*, 411–423. [CrossRef] [PubMed]
4. Maier, H.; Dietz, A.; Gewelke, U.; Heller, W.D.; Weidauer, H. Tobacco and alcohol and the risk of head and neck cancer. *Clin. Investig.* **1992**, *70*, 320–327. [CrossRef] [PubMed]
5. Kreimer, A.R. Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide: A Systematic Review. *Cancer Epidemiol. Biomark. Prev.* **2005**, *14*, 467–475. [CrossRef] [PubMed]
6. Ragin, C.C.R.; Modugno, F.; Gollin, S.M. The epidemiology and risk factors of head and neck cancer: A focus on human papillomavirus. *J. Dent. Res.* **2007**, *86*, 104–114. [CrossRef]
7. Sano, D.; Oridate, N. The molecular mechanism of human papillomavirus-induced carcinogenesis in head and neck squamous cell carcinoma. *Int. J. Clin. Oncol.* **2016**, *21*, 819–826. [CrossRef]
8. Kobayashi, K.; Hisamatsu, K.; Suzui, N.; Hara, A.; Tomita, H.; Miyazaki, T. A Review of HPV-Related Head and Neck Cancer. *J. Clin. Med.* **2018**, *7*, 241. [CrossRef]
9. Bajos, N.; Bozon, M.; Beltzer, N.; Laborde, C.; Andro, A.; Ferrand, M.; Goulet, V.; Laporte, A.; Le Van, C.; Leridon, H.; et al. Changes in sexual behaviours: From secular trends to public health policies. *AIDS Lond. Engl.* **2010**, *24*, 1185–1191. [CrossRef]
10. Zheng, Z.-M.; Baker, C.C. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front. Biosci. J. Virtual Libr.* **2006**, *11*, 2286–2302. [CrossRef]
11. Munger, K.; Werness, B.A.; Dyson, N.; Phelps, W.C.; Harlow, E.; Howley, P.M. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J.* **1989**, *8*, 4099–4105. [CrossRef]
12. Mantovani, F.; Banks, L. The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene* **2001**, *20*, 7874–7887. [CrossRef]
13. Rumfield, C.S.; Roller, N.; Pellom, S.T.; Schlom, J.; Jochems, C. Therapeutic Vaccines for HPV-Associated Malignancies. *Immuno-Targets Ther.* **2020**, *9*, 167–200. [CrossRef] [PubMed]
14. Sever, R.; Brugge, J.S. Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* **2015**, *5*, a006098. [CrossRef] [PubMed]
15. Mogensen, T.H.; Paludan, S.R. Molecular Pathways in Virus-Induced Cytokine Production. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 131–150. [CrossRef] [PubMed]
16. Kotwal, G.J.; Hatch, S.; Marshall, W.L. Viral infection: An evolving insight into the signal transduction pathways responsible for the innate immune response. *Adv. Virol.* **2012**, *2012*, 131457. [CrossRef]
17. Seiwert, T.Y.; Zuo, Z.; Keck, M.K.; Khattri, A.; Pedamallu, C.S.; Stricker, T.; Brown, C.; Pugh, T.J.; Stojanov, P.; Cho, J.; et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2015**, *21*, 632–641. [CrossRef]
18. Cancer Genome Atlas Network Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [CrossRef]
19. Esposti, D.D.; Sklias, A.; Lima, S.C.; Beghelli-de la Forest Divonne, S.; Cahais, V.; Fernandez-Jimenez, N.; Cros, M.-P.; Ecsedi, S.; Cuenin, C.; Bouaoun, L.; et al. Unique DNA methylation signature in HPV-positive head and neck squamous cell carcinomas. *Genome Med.* **2017**, *9*, 33. [CrossRef]
20. Lim, Y.; Wan, Y.; Vagenas, D.; Ovchinnikov, D.A.; Perry, C.F.L.; Davis, M.J.; Punyadeera, C. Salivary DNA methylation panel to diagnose HPV-positive and HPV-negative head and neck cancers. *BMC Cancer* **2016**, *16*, 749. [CrossRef]
21. Boscolo-Rizzo, P.; Furlan, C.; Lupato, V.; Polesel, J.; Fratta, E. Novel insights into epigenetic drivers of oropharyngeal squamous cell carcinoma: Role of HPV and lifestyle factors. *Clin. Epigenetics* **2017**, *9*, 124. [CrossRef]
22. Khanal, S.; Shumway, B.S.; Zahin, M.; Redman, R.A.; Strickley, J.D.; Trainor, P.J.; Rai, S.N.; Ghim, S.-J.; Jenson, A.B.; Joh, J. Viral DNA integration and methylation of human papillomavirus type 16 in high-grade oral epithelial dysplasia and head and neck squamous cell carcinoma. *Oncotarget* **2018**, *9*, 30419–30433. [CrossRef]
23. Koneva, L.A.; Zhang, Y.; Virani, S.; Hall, P.B.; McHugh, J.B.; Chepeha, D.B.; Wolf, G.T.; Carey, T.E.; Rozek, L.S.; Sartor, M.A. HPV Integration in HNSCC Correlates with Survival Outcomes, Immune Response Signatures, and Candidate Drivers. *Mol. Cancer Res. MCR* **2018**, *16*, 90–102. [CrossRef] [PubMed]
24. Shen, Y.; Liu, J.; Zhang, L.; Dong, S.; Zhang, J.; Liu, Y.; Zhou, H.; Dong, W. Identification of Potential Biomarkers and Survival Analysis for Head and Neck Squamous Cell Carcinoma Using Bioinformatics Strategy: A Study Based on TCGA and GEO Datasets. *BioMed Res. Int.* **2019**, *2019*, e7376034. [CrossRef] [PubMed]
25. Cheng, H.; Yang, X.; Si, H.; Saleh, A.D.; Xiao, W.; Coupar, J.; Gollin, S.M.; Ferris, R.L.; Issaeva, N.; Yarbrough, W.G.; et al. Genomic and Transcriptomic Characterization Links Cell Lines with Aggressive Head and Neck Cancers. *Cell Rep.* **2018**, *25*, 1332–1345.e5. [CrossRef] [PubMed]
26. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [CrossRef]
27. Lee, H.; Palm, J.; Grimes, S.M.; Ji, H.P. The Cancer Genome Atlas Clinical Explorer: A web and mobile interface for identifying clinical-genomic driver associations. *Genome Med.* **2015**, *7*, 112. [CrossRef]

28. Diamandis, E.P.; Yousef, G.M. Human tissue kallikreins: A family of new cancer biomarkers. *Clin. Chem.* **2002**, *48*, 1198–1205. [\[CrossRef\]](#)
29. Fuhrman-Luck, R.A.; Loessner, D.; Clements, J.A. Kallikrein-Related Peptidases in Prostate Cancer: From Molecular Function to Clinical Application. *EJIFCC* **2014**, *25*, 269–281.
30. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [\[CrossRef\]](#)
31. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **2013**, *29*, 1830–1831. [\[CrossRef\]](#)
32. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **2003**, *13*, 2129–2141. [\[CrossRef\]](#)
33. Mi, H.; Dong, Q.; Muruganujan, A.; Gaudet, P.; Lewis, S.; Thomas, P.D. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **2010**, *38*, D204–D210. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Howe, G.A.; Addison, C.L.  $\beta 1$  integrin: An emerging player in the modulation of tumorigenesis and response to therapy. *Cell Adhes. Migr.* **2012**, *6*, 71–77. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [\[CrossRef\]](#)
37. Husain, N.; Neyaz, A. Human papillomavirus associated head and neck squamous cell carcinoma: Controversies and new concepts. *J. Oral Biol. Craniofacial Res.* **2017**, *7*, 198–205. [\[CrossRef\]](#)
38. Dok, R.; Nuyts, S. HPV Positive Head and Neck Cancers: Molecular Pathogenesis and Evolving Treatment Strategies. *Cancers* **2016**, *8*, 41. [\[CrossRef\]](#)
39. Vigneswaran, N.; Williams, M.D. Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral Maxillofac. Surg. Clin. N. Am.* **2014**, *26*, 123–141. [\[CrossRef\]](#)
40. Özdemir, B.C.; Dotto, G.-P. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? *Trends Cancer* **2017**, *3*, 181–197. [\[CrossRef\]](#)
41. Joshi, P.; Dutta, S.; Chaturvedi, P.; Nair, S. Head and Neck Cancers in Developing Countries. *Rambam Maimonides Med. J.* **2014**, *5*, e0009. [\[CrossRef\]](#)
42. Gourin, C.G.; Podolsky, R.H. Racial disparities in patients with head and neck squamous cell carcinoma. *Laryngoscope* **2006**, *116*, 1093–1106. [\[CrossRef\]](#)
43. Marur, S.; D'Souza, G.; Westra, W.H.; Forastiere, A.A. HPV-associated head and neck cancer: A virus-related cancer epidemic. *Lancet Oncol.* **2010**, *11*, 781–789. [\[CrossRef\]](#)
44. Mahal, B.A.; Catalano, P.J.; Haddad, R.I.; Hanna, G.J.; Kass, J.I.; Schoenfeld, J.D.; Tishler, R.B.; Margalit, D.N. Incidence and Demographic Burden of HPV-Associated Oropharyngeal Head and Neck Cancers in the United States. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 1660–1667. [\[CrossRef\]](#)
45. Offenberger, H.H.; Schalk, J.A.; Meuwissen, R.L.; van Aalderen, M.; Kester, H.A.; Dietrich, A.J.; Heyting, C. SCP2: A major protein component of the axial elements of synaptonemal complexes of the rat. *Nucleic Acids Res.* **1998**, *26*, 2572–2579. [\[CrossRef\]](#)
46. Pannone, G.; Santoro, A.; Papagerakis, S.; Lo Muzio, L.; De Rosa, G.; Bufo, P. The role of human papillomavirus in the pathogenesis of head & neck squamous cell carcinoma: An overview. *Infect. Agent. Cancer* **2011**, *6*, 4. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Masterson, L.; Sorgeloos, F.; Winder, D.; Lechner, M.; Marker, A.; Malhotra, S.; Sudhoff, H.; Jani, P.; Goon, P.; Sterling, J. Deregulation of SYCP2 predicts early stage human papillomavirus-positive oropharyngeal carcinoma: A prospective whole transcriptome analysis. *Cancer Sci.* **2015**, *106*, 1568–1575. [\[CrossRef\]](#) [\[PubMed\]](#)
48. 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. [\[CrossRef\]](#)
49. Martinez, I.; Wang, J.; Hobson, K.F.; Ferris, R.L.; Khan, S.A. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. *Eur. J. Cancer* **2007**, *43*, 415–432. [\[CrossRef\]](#)
50. Slebos, R.J.C.; Yi, Y.; Ely, K.; Carter, J.; Evjen, A.; Zhang, X.; Shyr, Y.; Murphy, B.M.; Cmelak, A.J.; Burkey, B.B.; et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2006**, *12*, 701–709. [\[CrossRef\]](#)
51. Zhou, H.; Grubisic, I.; Zheng, K.; He, Y.; Wang, P.J.; Kaplan, T.; Tjian, R. Taf7l cooperates with Trf2 to regulate spermiogenesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16886–16891. [\[CrossRef\]](#)
52. Mobasher, M.B.; Shirkoobi, R.; Modarressi, M.H. Cancer/Testis OIP5 and TAF7L Genes are Up-Regulated in Breast Cancer. *Asian Pac. J. Cancer Prev. APJCP* **2015**, *16*, 4623–4628. [\[CrossRef\]](#)
53. Stoimenov, I.; Helleday, T. PCNA on the crossroad of cancer. *Biochem. Soc. Trans.* **2009**, *37*, 605–613. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Bravo, R.; Fey, S.J.; Bellatin, J.; Larsen, P.M.; Celis, J.E. Identification of a nuclear polypeptide (“cyclin”) whose relative proportion is sensitive to changes in the rate of cell proliferation and to transformation. *Prog. Clin. Biol. Res.* **1982**, *85*, 235–248.



55. Hoelz, D.J.; Arnold, R.J.; Dobrolecki, L.E.; Abdel-Aziz, W.; Loehrer, A.P.; Novotny, M.V.; Schnaper, L.; Hickey, R.J.; Malkas, L.H. The discovery of labile methyl esters on proliferating cell nuclear antigen by MS/MS. *Proteomics* **2006**, *6*, 4808–4816. [\[CrossRef\]](#)
56. Montgomery, R.I.; Warner, M.S.; Lum, B.J.; Spear, P.G. Herpes simplex virus-1 entry into cells mediated by a novel member of the TNF/NGF receptor family. *Cell* **1996**, *87*, 427–436. [\[CrossRef\]](#)
57. Edilova, M.I.; Abdul-Sater, A.A.; Watts, T.H. TRAF1 Signaling in Human Health and Disease. *Front. Immunol.* **2018**, *9*, 2969. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Marsters, S.A.; Ayres, T.M.; Skubatch, M.; Gray, C.L.; Rothe, M.; Ashkenazi, A. Herpesvirus entry mediator, a member of the tumor necrosis factor receptor (TNFR) family, interacts with members of the TNFR-associated factor family and activates the transcription factors NF-kappaB and AP-1. *J. Biol. Chem.* **1997**, *272*, 14029–14032. [\[CrossRef\]](#)
59. Rothe, M.; Pan, M.G.; Henzel, W.J.; Ayres, T.M.; Goeddel, D.V. The TNFR2-TRAF signaling complex contains two novel proteins related to baculoviral inhibitor of apoptosis proteins. *Cell* **1995**, *83*, 1243–1252. [\[CrossRef\]](#)
60. The Human Protein Atlas. Available online: <https://www.proteinatlas.org/> (accessed on 18 September 2021).
61. Evans, M.R.; James, C.D.; Loughran, O.; Nulton, T.J.; Wang, X.; Bristol, M.L.; Windle, B.; Morgan, I.M. An oral keratinocyte life cycle model identifies novel host genome regulation by human papillomavirus 16 relevant to HPV positive head and neck cancer. *Oncotarget* **2017**, *8*, 81892–81909. [\[CrossRef\]](#)
62. An, X.; Hao, Y.; Meneses, P.I. Host cell transcriptome modification upon exogenous HPV16 L2 protein expression. *Oncotarget* **2017**, *8*, 90730–90747. [\[CrossRef\]](#)
63. Certo, M.; Del Gaizo Moore, V.; Nishino, M.; Wei, G.; Korsmeyer, S.; Armstrong, S.A.; Letai, A. Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* **2006**, *9*, 351–365. [\[CrossRef\]](#)
64. Akl, H.; Vervloessem, T.; Kiviluoto, S.; Bittremieux, M.; Parys, J.B.; De Smedt, H.; Bultynck, G. A dual role for the anti-apoptotic Bcl-2 protein in cancer: Mitochondria versus endoplasmic reticulum. *Biochim. Biophys. Acta* **2014**, *1843*, 2240–2252. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Letai, A.G. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. *Nat. Rev. Cancer* **2008**, *8*, 121–132. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Henriksen, R.; Wilander, E.; Oberg, K. Expression and prognostic significance of Bcl-2 in ovarian tumours. *Br. J. Cancer* **1995**, *72*, 1324–1329. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Lamers, F.; Schild, L.; den Hartog, I.J.M.; Ebus, M.E.; Westerhout, E.M.; Ora, I.; Koster, J.; Versteeg, R.; Caron, H.N.; Molenaar, J.J. Targeted BCL2 inhibition effectively inhibits neuroblastoma tumour growth. *Eur. J. Cancer* **2012**, *48*, 3093–3103. [\[CrossRef\]](#)
68. Zhao, D.; Ding, X.; Peng, J.; Zheng, Y.; Zhang, S. Prognostic significance of bcl-2 and p53 expression in colorectal carcinoma. *J. Zhejiang Univ. Sci. B* **2005**, *6*, 1163–1169. [\[CrossRef\]](#)
69. Pena, J.C.; Thompson, C.B.; Recant, W.; Vokes, E.E.; Rudin, C.M. Bcl-xL and Bcl-2 expression in squamous cell carcinoma of the head and neck. *Cancer* **1999**, *85*, 164–170. [\[CrossRef\]](#)
70. Zhang, M.; Zhang, Y.; Xu, E.; Mohibi, S.; de Anda, D.M.; Jiang, Y.; Zhang, J.; Chen, X. Rbm24, a target of p53, is necessary for proper expression of p53 and heart development. *Cell Death Differ.* **2018**, *25*, 1118–1130. [\[CrossRef\]](#)
71. Munger, K.; Gwin, T.K.; McLaughlin-Drubin, M.E. p16 in HPV-associated cancers. *Oncotarget* **2013**, *4*, 1864–1865. [\[CrossRef\]](#)
72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Use R! Springer: New York, NY, USA, 2009; ISBN 978-0-387-98141-3.
73. Neuwirth, E. *RColorBrewer: ColorBrewer Palettes*. 2014.
74. Keitt, T. *colorRamps: Builds Color Tables*. 2012.
75. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; et al. *gplots: Various R Programming Tools for Plotting Data*. 2020.
76. Team, R. *RStudio: Integrated Development for R*; Team R: Boston, MA, USA, 2018.
77. R Core Team. European Environment Agency. 2019. Available online: <https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006> (accessed on 18 September 2021).
78. Silva, T.C.; Colaprico, A.; Olsen, C.; D'Angelo, F.; Bontempi, G.; Ceccarelli, M.; Noushmehr, H. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Research* **2016**, *5*, 1542. [\[CrossRef\]](#)
79. Mounir, M.; Lucchetta, M.; Silva, T.C.; Olsen, C.; Bontempi, G.; Chen, X.; Noushmehr, H.; Colaprico, A.; Papaleo, E. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* **2019**, *15*, e1006701. [\[CrossRef\]](#)
80. Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5*, R80. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Dragulescu, A.; Arendt, C. *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. 2020. Available online: <https://github.com/colearendt/xlsx> (accessed on 30 August 2022).
82. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [\[CrossRef\]](#) [\[PubMed\]](#)
83. McCarthy, D.J.; Chen, Y.; Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **2012**, *40*, 4288–4297. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Phipson, B.; Trigos, A.; Ritchie, M.; Su, S.; Doyle, M.; Dashnow, H.; Law, C. *RNA-seq Analysis in R; Differential Expression Analysis* 2016. Available online: <https://combine-australia.github.io/RNAseq-R/06-mnaseq-day1.html> (accessed on 30 August 2019).

- 
85. Morgan, M.; Obenchain, V.; Hester, J.; Pagès, H. SummarizedExperiment: SummarizedExperiment container, Bioconductor version: Release (3.13). 2021.
  86. Gentleman, R.; Carey, V.J.; Huber, W.; Hahne, F. Genefilter: Genefilter: Methods for Filtering Genes from High-Throughput Experiments, Bioconductor Version: Release (3.13). 2021.
  87. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)]
  88. Ramos, M.; Schiffer, L.; Re, A.; Azhar, R.; Basunia, A.; Rodriguez, C.; Chan, T.; Chapman, P.; Davis, S.R.; Gomez-Cabrero, D.; et al. Software for the Integration of Multiomics Experiments in Bioconductor. *Cancer Res.* **2017**, *77*, e39–e42. [[CrossRef](#)]
  89. Blighe, K.; Rana, S.; Turkes, E.; Ostendorf, B.; Grioni, A.; Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling, Bioconductor Version: Release (3.13). 2021.