Single Nucleotide Polymorphism rs6942067 Is a Risk Factor in Young and in Non-Smoking Patients with HPV Negative Head and Neck Squamous Cell Carcinoma

Guillaume B. Cardin [1,2] , **Monique Bernard** [1,2], **Houda Bahig** [1,3], **Phuc Felix Nguyen-Tan** [3], **Olivier Ballivy** [3], **Edith Filion** [3], **Denis Soulieres** [4], **Pierre Philouze** [5], **Tareck Ayad** [1,5], **Louis Guertin** [5], **Eric Bissada** [5], **Francis Rodier** [1,2,6] **and Apostolos Christopoulos** [1,2,5,*]

1   Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, QC H2X 0A9, Canada; cardinguill@gmail.com (G.B.C.); bernardmonique07@gmail.com (M.B.); houdabahig@gmail.com (H.B.); tareck.ayad@umontreal.ca (T.A.); rodierf@mac.com (F.R.)
2   Institut du Cancer de Montréal, Montréal, QC H2X 0A9, Canada
3   Department of Radiation Oncology, Centre hospitalier de l'Université de Montréal, Montréal, QC H2X 3E4, Canada; felixnguyentan@gmail.com (P.F.N.-T.); olivier.ballivy.chum@ssss.gouv.qc.ca (O.B.); edith.filion@gmail.com (E.F.)
4   Department of Medicine, Service of Hemato-Oncology, Centre Hospitalier de l'Université de Montréal, Montréal, QC H2X 3E4, Canada; denis.soulieres@umontreal.ca
5   Otolaryngology-Head and Neck Surgery Service, Centre Hospitalier de l'Université de Montréal, Montréal, QC H2X 3E4, Canada; pierre.philouze@chu-lyon.fr (P.P.); guertinorl@gmail.com (L.G.); ericbissada@gmail.com (E.B.)
6   Département de Radiologie, Radio-Oncologie et Médecine Nucléaire, Université de Montréal, Montréal, QC H2X 3E4, Canada
*   Correspondence: a.christopoulos@umontreal.ca; Tel.: +514-890-8000 (ex. 31292)

Abstract: Genetic factors behind the increasing incidence of human papillomavirus (HPV) negative head and neck squamous cell carcinoma (HNSCC) in young non-smokers are suspected, but have not been identified. Recently, rs6942067, a single nucleotide polymorphism (SNP) located upstream of the DCBLD1 gene, was found associated with non-smoking lung adenocarcinoma. To validate if this SNP is also implicated in HNSCC, participants of The Cancer Genome Atlas HNSCC cohort were investigated for rs6942067 status, associated DCBLD1 expression, and clinical characteristics. Occurrence of the rs6942067 GG genotype is significantly higher in young and in HPV negative non-smoking HNSCC than in other HNSCC. Additionally, rs6942067 GG is associated with higher DCBLD1 expression in HNSCC and patients with high DCBLD1 expression have a worse overall survival at three years, both in univariate and multivariate analysis. Furthermore, high DCBLD1 expression is associated with activation of the integrin signaling pathway and its phosphorylation with EGFR and MET. Collectively, these findings suggest that DCBLD1 plays a critical role in HNSCC and demonstrate an association between rs6942067 and clinical characteristics of young age and HPV negative non-smoking status in HNSCC patients.

Keywords: head and neck squamous cell carcinoma; rs6942067; DCBLD1; cancer susceptibility genes; tobacco; human papillomavirus

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is a group of cancers occurring in the oral cavity, oropharynx, hypopharynx, and larynx, which accounts for over 650,000 new cases and 330,000 deaths annually worldwide [1]. HNSCC patients usually present classic risk factors, which are tobacco and alcohol use and/or human papillomavirus (HPV) infection [2–4]. While increased occurrence of HPV has been implicated in young patients developing HNSCC, there has also been a parallel increasing incidence of young patients with HNSCC with no known risk factors [5–8]. Most typically, these latter patients are characterized as young non-smokers, most often females, presenting with HPV negative disease, with the most frequent subsite being the oral tongue [5–7,9–14]. Over the last 45 years, many studies reported that those young patients developing HNSCC, especially in the oral tongue, had a more aggressive disease and a worse prognosis than older patients, although some studies also reported no significant difference [15–22]. Family history of cancer in young relatives has been associated with head and neck cancer; therefore, suggesting the presence of a genetic factor in these patients [9]. Such genetic factor is yet to be identified.

Whole exome analysis of young non-smoking HNSCC patients compared to older smoking patients has been studied but exomes were similar in both groups [23]. Whole genome analysis looking at single nucleotide polymorphisms (SNPs) was never studied in HNSCC, partly because of the difficulty to build a large enough cohort to obtain sufficient statistical power. On the other hand, it has been analyzed in non-smoking lung adenocarcinoma on a cohort of 6609 never-smoking and 7457 control patients, finding a SNP-heavy region upstream of the DCBLD1 gene, although no relationship with age was reported [24,25]. It was later shown in vitro that it is one specific SNP from this region, rs6942067, which causes an upregulation of the DCBLD1 gene, and the association between rs6942067 and DCBLD1 overexpression was corroborated in vivo in lung and thyroid tissue [26]. Whether rs6942067 or DCBLD1 would be implied in non-smoking or other HNSCC was never studied.

DCBLD1 is a transmembrane protein with domain structures similar to that of neuropilins [27]. DCBLD1 has been associated with lung adenocarcinoma and glioma [24–27]. DCBLD1 remains largely uncharacterized and not many other human studies have been published. A functional study identified YxxP motifs in DCBLD1, which are phosphorylated by Src family kinases and Abl, and promote the binding of CRK adaptor proteins, although the implications are unknown [28]. Although DCBLD1 was never studied in HNSCC, DCBLD2 (paralog with DCBLD1) has been shown to be a bad prognostic marker in HNSCC when co-activated with EGFR [29].

While DCBLD1's function remains unclear, it seems relevant in specific cancer settings. The purpose of this study is to evaluate the presence and clinical relevance of rs6942067 and DCBLD1 specifically in the young and HPV negative non-smoking patients with HNSCC.

2. Results

2.1. rs6942067 Status in HPV Negative Non-Smoker HNSCC and Young HNSCC

To evaluate if rs6942067 could explain in part the occurrence of HPV negative non-smoker HNSCC, the Cancer Genome Atlas (TCGA) HNSCC cohort (154 participants with whole genome sequencing data, rs6942067 AA = 61, AG = 70, GG = 23) was used as the investigational population. Patients from the TCGA HNSCC cohort were stratified in an HPV positive and/or smoker group (rs6942067 AA = 50, AG = 59, GG = 15) and an HPV negative non-smoker group (rs6942067 AA = 11, AG = 11, GG = 8). The HPV negative non-smoker group showed higher occurrence of the homozygous GG genotype in comparison with other HNSCC (hazard ratio (HR) = 2.20, 95% confidence interval (CI) = 1.03 to 4.71), suggesting a recessive effect (Figure 1A).

To test if rs6942067 could also explain the occurrence of HNSCC in a younger population, rs6942067 status was determined for HNSCC in a defined young population (<40 years old, rs6942067 AA = 0, AG = 6, GG = 4) versus older HNSCC (≥40 years old, rs6942067 AA = 61, AG = 64, GG = 20) in the

TCGA HNSCC cohort (Figure 1B). The occurrence of young HNSCC cases was small but their genotype distribution was significantly different from older HNSCC cases (HR = 3.03, 95% CI = 1.27 to 7.21).
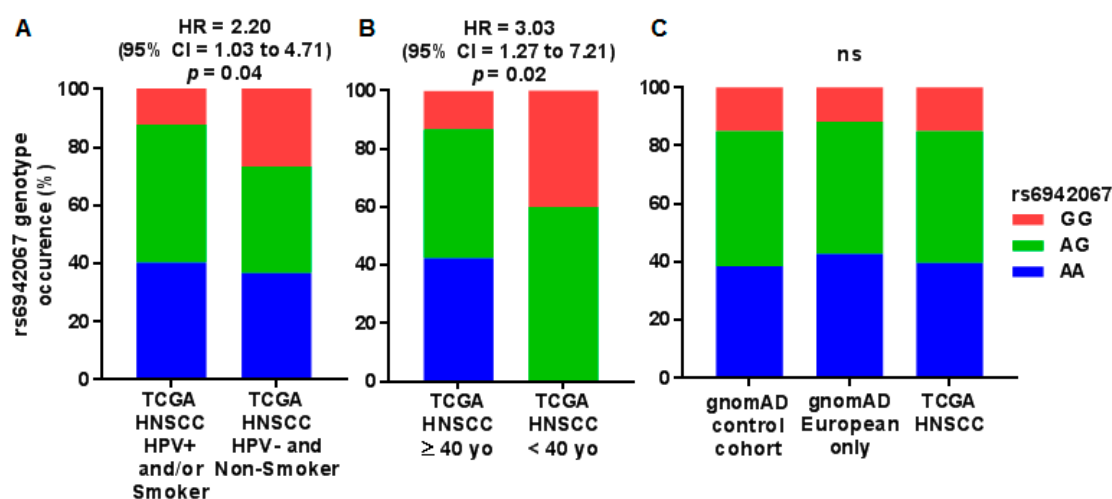

Figure 1. rs6942067 genotype status in the Cancer TCGA HNSCC cohort and the gnomAD cohort. Comparison of rs6942067 genotype for (**A**) HPV positive and/or smoking patients (*n* = 124) and non-smoker and HPV negative patients (*n* = 30) of the TCGA HNSCC cohort; (**B**) ≥40 years old (*n* = 144) and <40 years old (*n* = 10) patients of the TCGA HNSCC cohort; and (**C**) all participants of the gnomAD cohort (*n* = 15,654), European participant of the gnomAD cohort (*n* = 9431), and patients of the TCGA HNSCC cohort. Only 154 patients had whole genome sequencing data for the TCGA HNSCC cohort. Relative risk and *p* values are for a Pearson's chi-square test and a recessive model of penetrance.

Distribution of rs6942067 genotype was also investigated regarding alcohol usage for the 67 patients from the TCGA HNSCC cohort which had both alcohol history and whole genome sequencing data. They were stratified in an HPV positive and/or drinker group (rs6942067 AA = 20, AG = 23, GG = 4) and an HPV negative non-drinker group (rs6942067 AA = 5, AG = 11, GG =4). The HPV negative non-drinker group showed non-significant higher occurrence of the homozygous GG genotype in comparison with other HNSCC (HR = 2.35, 95% CI = 0.65 to 8.48) (Figure S1).

It was then validated that the HNSCC TCGA cohort had a similar rs6942067 genotype distribution then the healthy population. The gnomAD cohort was used as the healthy control group (mostly from case-control studies, rs6942067 AA = 6031, AG = 7282, GG = 2341). Since the HNSCC TCGA cohort is mostly Caucasians, European participants of the gnomAD cohort (rs6942067 AA = 4037, AG = 4290, GG = 1104) were also looked at, specifically, as a generally Caucasian cohort. No significant difference was observed between the three groups (Figure 1C) and the HNSCC TCGA cohort showed a phenotype distribution much alike the gnomAD cohort.

Clinical features were different according to rs6942067 genotype in the HNSCC TCGA cohort (Table 1). There was no significant difference for age group, gender, anatomical subsite, smoking status, HPV status, tumor size, and nodal status. *p* values were calculated by a two-sided Pearson's chi-square test and corrected for false discovery rate (FDR) using the Benjamini–Hochberg procedure. The absence of independent significance for HPV and smoking status suggests that it is the complete absence of HNSCC risk factor which is associated with the rs6942067 GG genotype.

Table 1. Clinical features of the TCGA HNSCC cohort stratified by rs6942067 status.

Characteristics	rs6942067 (AA/AG)	rs6942067 (GG)	*p*
Total	131	23	
Age group, *n* (%)			0.12
≥40 years old	125 (95.4)	19 (82.6)	
<40 years old	6 (4.6)	4 (17.4)	
Gender, *n* (%)			0.43
Female	38 (29.0)	4 (17.4)	
Male	93 (71.0)	19 (82.6)	
Anatomical subsite, *n* (%)			0.88
Oral tongue	47 (35.9)	8 (34.8)	
Other oral cavity	42 (32.1)	6 (26.1)	
Oropharynx	23 (17.6)	4 (17.4)	
Larynx and hypopharynx	19 (14.5)	5 (21.7)	
Smoking status, *n* (%)			0.14
Current/previous ≥ 10 PY [1]	99 (75.6)	13 (56.5)	
Never/previous < 10 PY [1]	32 (24.4)	10 (43.5)	
HPV, *n* (%)			0.59
Positive	33 (25.2)	4 (17.4)	
Negative	98 (74.8)	19 (82.6)	
Tumor size, *n* (%)			0.88
T1-2	53 (40.5)	9 (39.1)	
T3+	77 (58.8)	14 (60.9)	
Tx, unknown	1 (0.8)	0 (0)	
Nodal status, *n* (%)			0.12
N0	64 (48.9)	6 (26.1)	
N+	64 (48.9)	17 (73.9)	
Nx, unknown	2 (1.5)	0 (0)	

[1] Pack-year.

2.2. DCBLD1 Association with rs6942067

To evaluate the global effect of rs6942067 on DCBLD1 gene expression, we used multi-tissue expression quantitative trait locus analysis from the Genotype-Tissue Expression (GTEx) project. Meta-analysis using the Han and Eskin's random effects model (RE2) [30] shows a significant human multi-tissue association (RE2: $p = 2.1 \times 10^{-54}$) between rs6942067 and DCBLD1 expression (Figure 2A). This analysis shows that in most normal human tissues, the rs6942067 G allele is associated with an upregulation of DCBLD1 expression.

More specifically in the TCGA HNSCC cohort, DCBLD1 expression is upregulated in the rs6942067 GG patients of that cohort in comparison to other genotype ($p = 0.004$) (Figure 2B). Additionally, DCBLD1 expression is upregulated in the HPV negative non-smoker patients subgroup, in comparison to HPV positive ($p < 0.001$) and HPV negative smoker patients ($p = 0.001$) (Figure 2C), thus strengthening the hypothesis that it is the complete absence of HNSCC risk factor which is associated with the rs6942067 GG genotype.
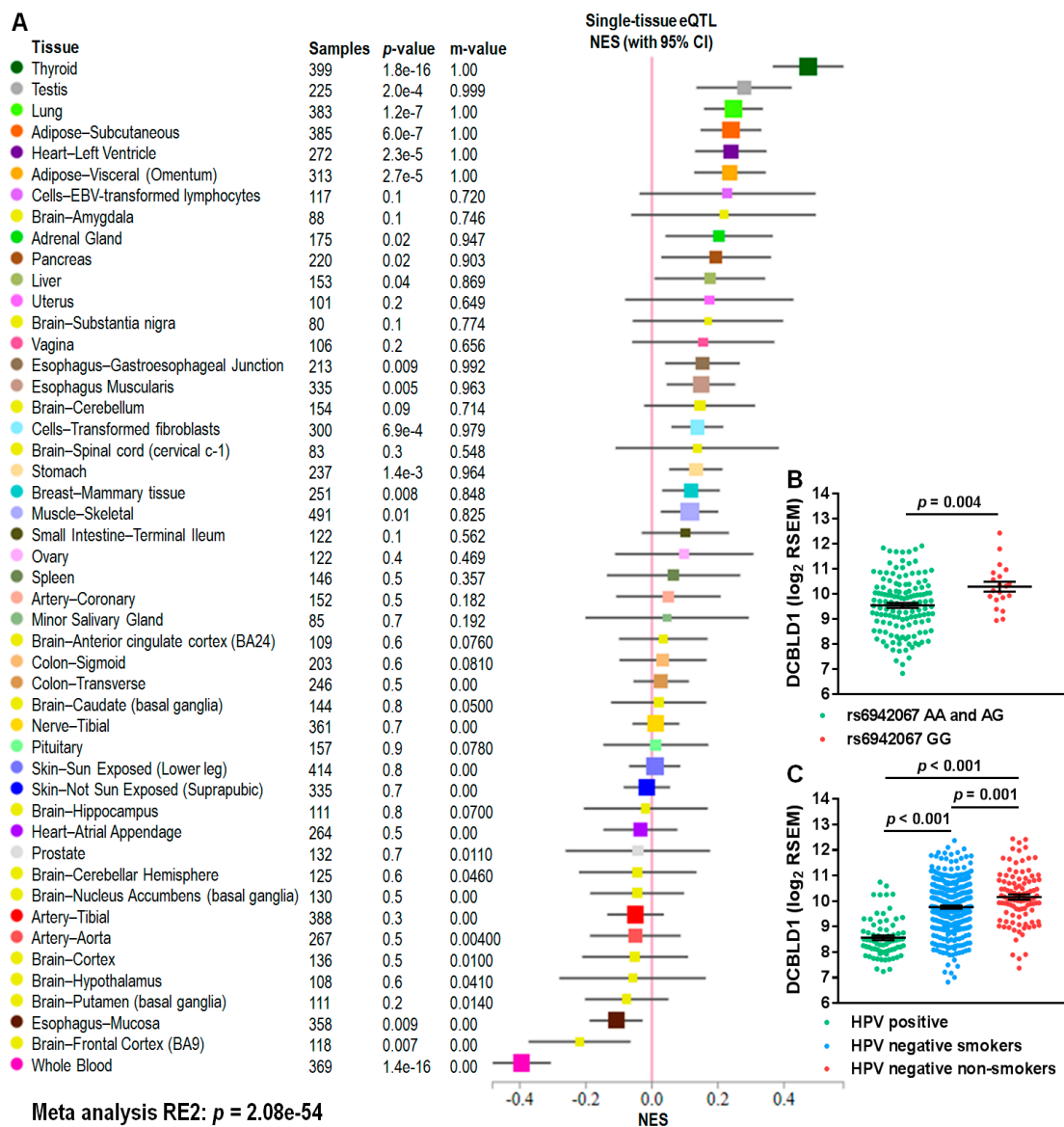
Figure 2. DCBLD1 gene expression, rs6942067, and HNSCC risk factors. (**A**) Multi-tissue expression quantitative trait locus analysis of rs6942067 and DCBLD1 using the GTEx project and the Han and Eskin's random effects model (RE2). (**B**,**C**) Comparison of DCBLD1 gene expression for (**B**) rs6942067 AA and AG (n = 135) or rs6940267 GG (n = 20) and (**C**) HPV positive (green, n = 71), HPV negative smokers (blue, n = 334), or HPV negative non-smokers (red, n = 101) patients of the TCGA HNSCC cohort. Mean and standard error are shown.

2.3. DCBLD1 Association with Clinical Features and Overall Survival

HNSCC patients of the TCGA cohort were stratified in a DCBLD1-low (n = 258) and a DCBLD1-high (n = 259) group using the expression median. The impact of DCBLD1 expression and rs6942067 genotype on patients' overall survival was then evaluated on the TCGA HNSCC cohort (Figure 3).

Univariate Kaplan–Meier analysis showed that high DCBLD1 gene expression was significantly associated to a worse overall survival at three years (HR = 1.69, 95% CI = 1.26 to 2.29), but no difference in overall survival was observed for the rs6942067 GG genotype.

Clinical characteristics were compared between the DCBLD1-high and DCBLD1-low groups (Table 2). The DCBLD1-high patients had lower occurrence of HPV (p < 0.001), were more often female (p = 0.02), and had a different anatomical subsite distribution (p < 0.001), with 2.4 times more

occurrence of oral tongue cancer. There were no significant difference for age groups, smoking status, tumor size, and nodal status.
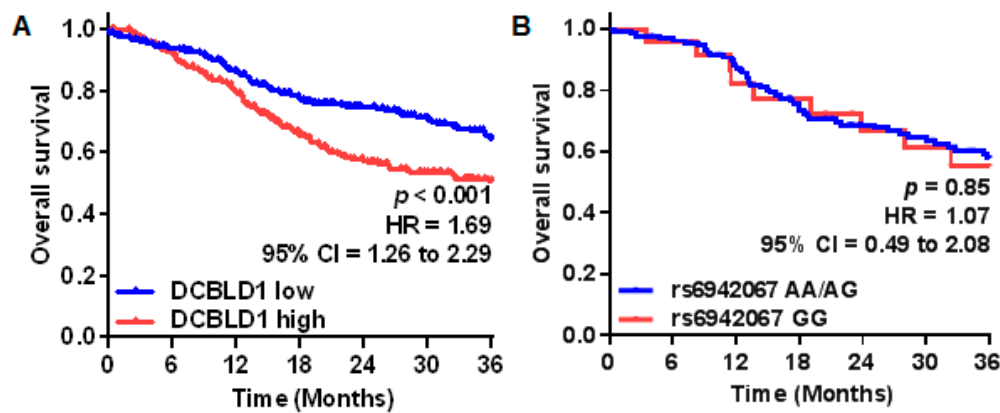

Figure 3. DCBLD1 gene expression, rs6942067, and HNSCC overall survival comparison for HNSCC patients of the TCGA cohort using the Kaplan–Meier curve estimation for (**A**) tumor low (blue, *n* = 258) and high (red, *n* = 259) DCBLD1 gene expression; and for (**B**) rs6942067 AA and AG (blue, *n* = 131) versus rs6942067 GG (red, *n* = 23). *p* values were calculated using a two-sided log-rank test.

Table 2. Clinical features of the TCGA HNSCC cohort stratified by DCBLD1 gene expression.

Characteristics	DCBLD1-High	DCBLD1-Low	*p*
Total	258	259	
Age group, *n* (%)			0.74
≥40 years old	250 (96.9)	248 (95.8)	
<40 years old	8 (3.1)	10 (3.9)	
unknown	0 (0)	1 (0.4)	
Gender, *n* (%)			0.02
Female	55 (21.3)	81 (31.3)	
Male	203 (78.7)	178 (68.7)	
Anatomical subsite, *n* (%)			<0.001
Oral tongue	38 (14.7)	92 (35.5)	
Other oral cavity	85 (32.9)	98 (37.8)	
Oropharynx	60 (23.3)	19 (7.3)	
Larynx and hypopharynx	75 (29.1)	50 (19.3)	
Smoking status, *n* (%)			0.15
Current/previous ≥ 10 PY [1]	198 (76.7)	182 (70.3)	
Never/previous < 10 PY [1]	54 (20.9)	71 (27.4)	
Unknown	6 (2.3)	6 (2.3)	
HPV, *n* (%)			<0.001
Positive	64 (24.8)	7 (2.7)	
Negative	194 (75.2)	252 (97.3)	
Tumor size, *n* (%)			0.67
T1-2	97 (37.6)	89 (34.3)	
T3+	154 (59.7)	161 (62.2)	
Tx, unknown	7 (2.7)	9 (3.5)	
Nodal status, *n* (%)			0.75
N0	119 (46.1)	122 (47.1)	
N+	129 (50.0)	125 (48.3)	
Nx, unknown	10 (3.9)	12 (4.6)	

[1] Pack-year.

Since DCBLD1 expression is strongly associated to HPV status and HPV is an important prognostic factor for oropharyngeal SCC [31], we evaluated the independence of DCBLD1 as a prognostic marker.

Univariate and multivariate Cox proportional hazards analysis of overall survival were performed on the HNSCC TCGA cohort (Table 3). Multivariate Cox proportional hazards analysis of overall survival showed that DCBLD1 expression is an independent prognostic marker (HR = 2.69, 95% CI = 1.09 to 6.58) along with HPV status, nodal status, and age. Those factors were also significantly predictive of overall survival in the univariate analysis. Gender and anatomical subsite were significantly predictive of overall survival in the univariate analysis but not in the multivariate analysis. DCBLD1 expression and age were analyzed as numerical variables. Assumption of proportionality was verified by testing that the interactions between survival time and covariates were statistically not significant.

Table 3. Multivariate Cox proportional hazards analysis of overall survival in the TCGA HNSCC cohort.

Variable	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	p	HR (95% CI)	p
HPV		<0.001		0.001
Negative	Reference		Reference	
Positive	0.27 (0.13 to 0.50)		0.28 (0.11 to 0.62)	
Nodal status		0.04		0.006
N0	Reference		Reference	
N+	1.35 (1.01 to 1.82)		1.56 (1.13 to 2.15)	
Age, years old (range) [1]	3.48 (1.38 to 8.86)	0.008	3.44 (1.22 to 9.91)	0.02
DCBLD1, $\log_2$ RSEM (range) [1]	3.97 (1.87 to 8.43)	<0.001	2.69 (1.09 to 6.58)	0.03
Smoking status		0.30		0.24
Never/previous < 10 PY [2]	Reference		Reference	
Current/previous ≥ 10 PY [2]	1.21 (0.86 to 1.74)		1.26 (0.86 to 1.89)	
Tumor size		0.09		0.29
T1-2	Reference		Reference	
T3+	1.31 (0.96 to 1.79)		1.21 (0.85 to 1.73)	
Gender		0.02		0.43
Female	Reference		Reference	
Male	0.70 (0.51 to 0.95)		0.87 (0.62 to 1.24)	
Anatomical subsite		0.01		0.61
Oropharynx	Reference		Reference	
Oral tongue	2.00 (1.16 to 3.64)		0.83 (0.44 to 1.67)	
Other oral cavity	2.23 (1.35 to 3.97)		0.92 (0.50 to 1.78)	
Larynx and hypopharynx	1.70 (0.98 to 3.12)		0.71 (0.37 to 1.44)	

[1] HR > 1 for numerical variables means that higher expression is associated with a worse overall survival. [2] Pack-year.

2.4. Bioinformatical Analysis of DCBLD1 Function

To get insight into DCBLD1 function in HNSCC, we analyzed RNAseq data from the 25 highest and 25 lowest DCBLD1-expressing patients from the HNSCC TCGA cohort, to identify genes that were significantly differentially expressed in the two groups using Bonferroni correction. Those two groups represent the top and bottom 5% of the 517 participants with RNAseq data. Upregulated and downregulated genes were then independently tested for gene ontology using the PANTHER pathways (Figure 4). For significantly enriched pathways in the DCBLD1-high group, the integrin signaling pathway especially stood out with an FDR of 6.08×10^{-16}. Other significantly enriched pathways were the plasminogen activating cascade, the Alzheimer disease–presenilin pathway, blood coagulation, the cadherin signaling pathway, the transforming growth factor beta (TGF–beta) signaling pathway and the Wnt signaling pathway. There was no significantly enriched pathway in the DCBLD1-low group. We also performed this analysis using the 50 highest and 50 lowest DCBLD1 expressing patients (Figure S2). Results were similar to the previous analysis, although significance was higher in all pathways and four other pathways related to cell function became significant.
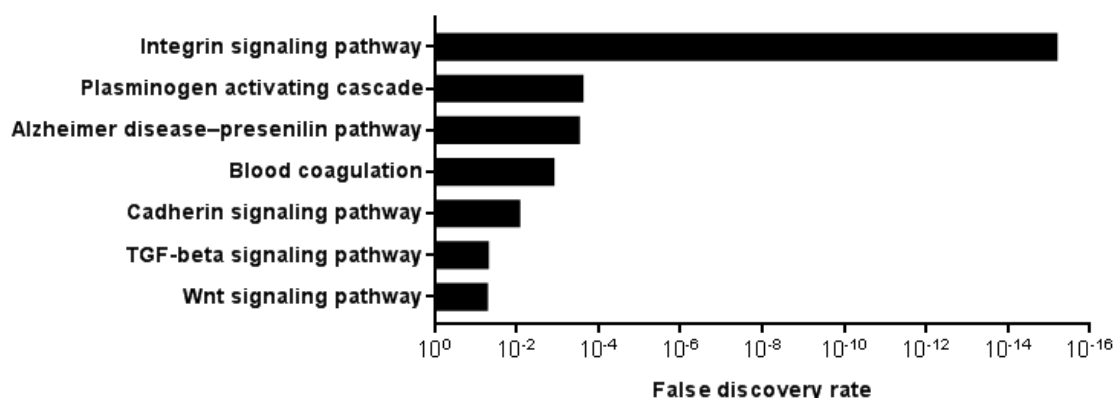
Figure 4. Significantly enriched PANTHER pathways in the higher (*n* = 25) versus lower (*n* = 25) DCBLD1 expressing participants from the HNSCC TCGA cohort. *p* values for significance cut-off was calculated using Bonferroni correction. No pathway was significantly enriched in the lower DCBLD1 participants.

Protein modifications and their regulation are linked to function. PhosphoSitePlus® analysis on the DCBLD1 protein reveals two types of modifications on DCBLD1: An S513 modification, which is highly present but not associated with any specific condition; and six YxxP sites (Figure 5). Those YxxP sites are the same ones which were previously associated with the Src family kinases, Abl, and CRK adaptor proteins [28]. Our analysis showed that two important receptor tyrosine kinases are also linked to these YxxP sites—EGFR and MET—which are, respectively, ligands for EGF and HGF.
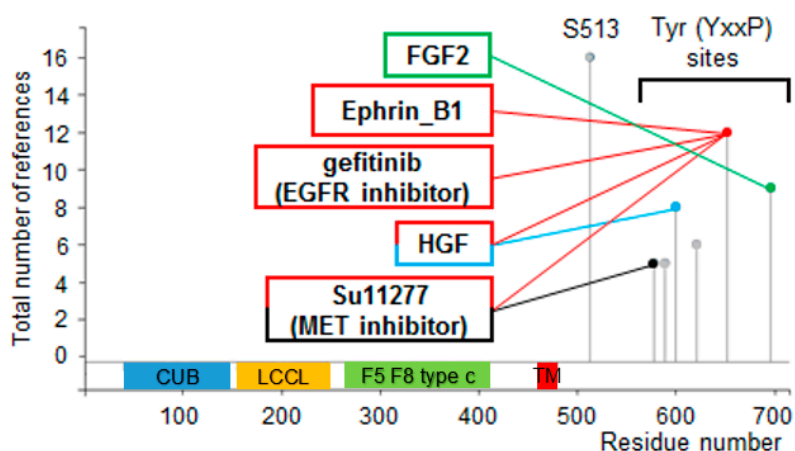
Figure 5. Phosphoproteomic bioinformatical analysis of the DCBLD1 protein. PhosphoSitePlus® analysis shows one frequently detected S513 phosphorylation site of unknown function and six YxxP sites. Four of the six sites are linked with three different receptor tyrosine kinases.

3. Discussion

The reason why we observe an increasing incidence of young patients with HNSCC with no known risk factors has been elusive for decades. In this study, we report a SNP (rs6942067) which is associated with HNSCC in non-smoker HPV negative patients and also in young (<40 years old) patients. It is important to state that a large corroboration cohort will be needed before drawing definitive conclusions on the associations presented here in HNSCC. This SNP had been previously identified in a susceptibility loci at 6q22.2 for patients with higher risk of developing never-smoking lung adenocarcinoma [24]. It was shown in vitro that rs6942067 is the SNP causing an upregulation of the DCBLD1 gene [26]. We validated this association between rs6942067 and DCBLD1 in HNSCC, but, while we show that DCBLD1 is a prognosis factor in HNSCC, interestingly rs6942067 itself is not

associated with HNSCC overall survival. This is somewhat expected given that other unknown factors probably affect DCBLD1 gene expression, as there are many patients with high DCBLD1 expression in the rs6942067 AA and AG group in Figure 2b. Additionally, rs6942067 GG occurrence in the healthy population is too high to explain on its own why some young patients with no classical HNSCC risk factors develop HNSCC. It is possible that additional unknown factors are synergizing with rs6942067 in young and non-smokers with HPV negative HNSCC, and future studies will be necessary to answer this question.

It has been suggested in many studies that young patients with no classical HNSCC risk factors have a worse prognosis than other patients. Whether DCBLD1 overexpression would be the explanation for that reported worse prognosis is unknown. Interestingly, DCBLD1-high patients have a higher occurrence of oral tongue cancer and are more often female, all characteristics which were previously described in young patients with HNSCC with no known risk factors. Our gene ontology analysis suggests many pathways are associated with DCBLD1 expression but the integrin signaling pathway especially stood out with an FDR of 6.08×10^{-16}. This pathway is triggered when integrins in the cell membrane bind to extracellular matrix components, and downstream events include actin reorganization and activation of the MAPK signaling cascade [32–37]. In HNSCC, the integrin signaling pathway has been associated with progression, stemness, and resistance to radiotherapy [38–40].

Our PhosphoSitePlus® analysis showed that many receptor tyrosine kinases are linked to DCBLD1 phosphorylation. Of special interest are MET and its ligand (HGF) reported for Y578, Y600, and Y652 regulation, and EGFR reported for Y652 regulation. MET has been reported as a poor prognosis marker in HNSCC in many studies [41–45]. EGFR is an important protein in HNSCC as it is associated with a worse outcome in the context of overexpression and is the target of Cetuximab [46–48]. The fact that DCBLD1 is a strong prognosis factor in HNSCC and experimentally associated with EGFR and MET suggests that DCBLD1 may play a critical role in HNSCC. Interestingly, DCBLD2, which is paralog to DCBLD1, has also been shown to be interacting with receptor tyrosine kinases such as EGFR, VEGFR, and INSR [28,49,50]. In the never-smoking lung adenocarcinoma study, the DCBLD1 susceptibility locus was associated with EGFR mutations in exons 19 and 21 (in the tyrosine kinase encoding region) [24,25].

The current study has limitations that must be stated. First, there is a need for a larger confirmation cohort to corroborate our results before drawing definitive conclusions, as this study was limited by the number of participants, especially for the analysis of the impact of alcohol use on rs6942067 genotype distribution, as only 67 participants of the TCGA HNSCC cohort had documented alcohol history and whole genome sequencing data. Second, it would be interesting to evaluate the occurrence of HNSCC patients that are both under 40 years old and HPV negative non-smokers. Sadly, there were only three such patients with whole genome sequencing data available in the TCGA HNSCC cohort and that analysis was; therefore, impossible.

It is important to acknowledge that the prognostic value of DCBLD1 detailed here is made without evaluation of treatment applied to the studied populations, which is itself dependent on the disease state. Since treatment is standard for most (surgery, radiation therapy, and chemoradiation therapy), predictive value of DCBLD1 could be evaluated by studying an interaction in statistical terms. Such evaluation would be highly interesting, but will have to be conducted from samples of patients treated in randomized trials to determine the predictive value of this gene to define therapy.

4. Materials and Methods

4.1. Study Cohorts

This study involved 525 participants of the TGCA HNSCC cohort (TCGA, Provisional) [51]. Three participants from this cohort with DCBLD1 gene mutations were excluded from this study, as DCBLD1 mutations are too rare in this cohort (3 cases) to reach meaningful conclusions. As a reference cohort for rs6942067 status, we used the 15,654 participants of the Genome Aggregation Database

(gnomAD) from the GTEx project [52]. Data obtained from the GTEx Portal and dbGaP Accession phs000424.v7.p2 on 22 May 2019 were used to evaluate rs6942067 global effect on DCBLD1 expression. The TCGA overall survival and gene expression data were extracted using cBioPortal for Cancer Genomics (www.cbioportal.org) [53,54]. Clinical characteristics of the TCGA cohort were extracted from FirehoseR (gdac.broadinstitute.org) [55]. Raw data used for these analysis is available in Table S1 (TCGA) and Table S2 (gnomAD). Access to controlled access data of whole genome sequencing was done though project #19174 (OMB 0925-0670).

4.2. Clinical Characteristics

HPV status of the TCGA HNSCC cohort was reported in a previous study [56]. Lifelong non-smokers and current previous smokers for over 15 years with less than 10 pack-year (PY) smoked were classified as non-smokers. Current smokers and other previous smokers were considered as smokers. For young patient age cut-off, <40 years old was used for rs6942067 occurrence, based on previous studies for young HNSCC patients with no classical risk factors [11,21,57–59]. For anatomical subsites, oral tongue was separated from other oral cavity cancers to evaluate previously reported higher occurrence of oral tongue cancer for young non-smoking HPV negative HNSCC. Tumor size (pT) was divided between small (≤T2, ≤4.0 cm) and large (≥T3, >4.0 cm) tumors. Nodal status was divided between negative (N0) and positive lymph node status (N+). For alcohol use, patients with an average of ≥1 drink per day were considered drinkers, while patients with an average of <1 drink per day were considered non-drinkers. The rs6942067A/G genotype of 154 participants of the TCGA HNSCC cohort was determined using whole genome sequencing data. Other participants of this cohort had no available information regarding their rs6942067A/G genotype. This SNP is located on chr6:117,464,533 of GRCh38.p12.

4.3. Bioinformatical Analysis

Multi-tissue expression quantitative trait locus analysis was done using the GTEx Portal [52]. Gene ontology analysis of genes upregulated for the 25 participants of the TCGA HNSCC cohort with the higher versus lower DCBLD1 gene expression was performed using the PANTHER pathway database to evaluate pathway enrichment [60–62]. PhosphoSitePlus® is a bioinformatical tool that finds reports of protein modifications in low and high-throughput phosphoproteomic data sources, and then searches for experimentally verified regulation by specific treatments or other binomial stratifications [63]. We used a cut-off of five reports of protein modification for this analysis.

4.4. Statistics

Genotype frequencies and variations in clinical characteristics were analyzed using a Pearson's chi-square test. Tukey–Kramer was used for gene expression analysis. Kaplan–Meier curves were assessed using log-rank *p* values. Analysis of clinical characteristics was corrected for FDR using the Benjamini–Hochberg procedure. A multivariate Cox proportional hazards analysis was used to evaluate hazard ratios and to perform the multivariate survival prediction model. Assumption of proportionality for this analysis was verified by testing that the interactions between survival time and covariates were statistically not significant.

Throughout the analyses, tests of statistical significance are two-sided and *p* values less than 0.05 corrected for multiple testing were considered statistically significant. Statistical analysis were performed using JMP 12.0.1 statistical software (SAS Institute Inc., Cary, NC, United States).

5. Conclusions

In summary, our study demonstrates an association between rs6942067 and clinical characteristics of young age and HPV negative non-smoking status in HNSCC patients. This study also associates high DCBLD1 expression with a worse prognosis and the integrin signaling pathways in HNSCC. We also showed an association of DCBLD1 phosphorylation with EGFR and MET.

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. 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]

3. Ragin, C.C.; 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] [PubMed]

4. Goldenberg, D.; Lee, J.; Koch, W.M.; Kim, M.M.; Trink, B.; Sidransky, D.; Moon, C.S. Habitual risk factors for head and neck cancer. *Otolaryngol. Head Neck Surg.* **2004**, *131*, 986–993. [CrossRef]

5. Jin, Y.T.; Myers, J.; Tsai, S.T.; Goepfert, H.; Batsakis, J.G.; el-Naggar, A.K. Genetic alterations in oral squamous cell carcinoma of young adults. *Oral Oncol.* **1999**, *35*, 251–256. [CrossRef]

6. Shiboski, C.H.; Schmidt, B.L.; Jordan, R.C. Tongue and tonsil carcinoma: Increasing trends in the U.S. population ages 20–44 years. *Cancer* **2005**, *103*, 1843–1849. [CrossRef]

7. Sturgis, E.M.; Cinciripini, P.M. Trends in head and neck cancer incidence in relation to smoking prevalence: An emerging epidemic of human papillomavirus-associated cancers? *Cancer* **2007**, *110*, 1429–1435. [CrossRef]

8. Golas, S.M. Trends in palatine tonsillar cancer incidence and mortality rates in the United States. *Community Dent. Oral Epidemiol.* **2007**, *35*, 98–108. [CrossRef]

9. Toporcov, T.N.; Znaor, A.; Zhang, Z.F.; Yu, G.P.; Winn, D.M.; Wei, Q.; Vilensky, M.; Vaughan, T.; Thomson, P.; Talamini, R.; et al. Risk factors for head and neck cancer in young adults: A pooled analysis in the INHANCE consortium. *Int. J. Epidemiol.* **2015**, *44*, 169–185. [CrossRef]

10. Patel, S.C.; Carpenter, W.R.; Tyree, S.; Couch, M.E.; Weissler, M.; Hackman, T.; Hayes, D.N.; Shores, C.; Chera, B.S. Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years. *J. Clin. Oncol.* **2011**, *29*, 1488–1494. [CrossRef]

11. Harris, S.L.; Kimple, R.J.; Hayes, D.N.; Couch, M.E.; Rosenman, J.G. Never-smokers, never-drinkers: Unique clinical subgroup of young patients with head and neck squamous cell cancers. *Head Neck* **2010**, *32*, 499–503. [CrossRef] [PubMed]

12. Annertz, K.; Anderson, H.; Biorklund, A.; Moller, T.; Kantola, S.; Mork, J.; Olsen, J.H.; Wennerberg, J. Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. *Int. J. Cancer* **2002**, *101*, 95–99. [CrossRef] [PubMed]

13. Liu, X.; Gao, X.L.; Liang, X.H.; Tang, Y.L. The etiologic spectrum of head and neck squamous cell carcinoma in young patients. *Oncotarget* **2016**, *7*, 66226–66238. [CrossRef] [PubMed]

14. Schantz, S.P.; Yu, G.P. Head and neck cancer incidence trends in young Americans, 1973-1997, with a special analysis for tongue cancer. *Arch. Otolaryngol. Head Neck Surg.* **2002**, *128*, 268–274. [CrossRef]

15. Kappert, H.F.; Ulbrich, J.; Glass, P.; Huber, H.; Neumuller, H. Determining the threshold level for sensibility responses to galvanic potentials in the oral cavity. *Dtsch. Zahnarztl. Z.* **1989**, *44*, 50–52.

16. Amsterdam, J.T.; Strawitz, J.G. Squamous cell carcinoma of the oral cavity in young adults. *J. Surg. Oncol.* **1982**, *19*, 65–68. [CrossRef]

17. Sarkaria, J.N.; Harari, P.M. Oral tongue cancer in young adults less than 40 years of age: Rationale for aggressive therapy. *Head Neck* **1994**, *16*, 107–111. [CrossRef]

18. Jeon, J.H.; Kim, M.G.; Park, J.Y.; Lee, J.H.; Kim, M.J.; Myoung, H.; Choi, S.W. Analysis of the outcome of young age tongue squamous cell carcinoma. *Maxillofac. Plast Reconstr. Surg.* **2017**, *39*, 41. [CrossRef]

19. Byers, R.M. Squamous cell carcinoma of the oral tongue in patients less than thirty years of age. *Am. J. Surg.* **1975**, *130*, 475–478. [CrossRef]

20. Kim, B.G.; Kim, J.H.; Kim, M.I.; Han, J.J.; Jung, S.; Kook, M.S.; Park, H.J.; Ryu, S.Y.; Oh, H.K. Retrospective study on factors affecting the prognosis in oral cancer patients who underwent surgical treatment only. *Maxillofac. Plast Reconstr. Surg.* **2016**, *38*, 3. [CrossRef]

21. Pytynia, K.B.; Grant, J.R.; Etzel, C.J.; Roberts, D.; Wei, Q.; Sturgis, E.M. Matched analysis of survival in patients with squamous cell carcinoma of the head and neck diagnosed before and after 40 years of age. *Arch. Otolaryngol. Head Neck Surg.* **2004**, *130*, 869–873. [CrossRef] [PubMed]

22. Bachar, G.; Hod, R.; Goldstein, D.P.; Irish, J.C.; Gullane, P.J.; Brown, D.; Gilbert, R.W.; Hadar, T.; Feinmesser, R.; Shpitzer, T. Outcome of oral tongue squamous cell carcinoma in patients with and without known risk factors. *Oral Oncol.* **2011**, *47*, 45–50. [CrossRef] [PubMed]

23. Pickering, C.R.; Zhang, J.; Neskey, D.M.; Zhao, M.; Jasser, S.A.; Wang, J.; Ward, A.; Tsai, C.J.; Ortega Alves, M.V.; Zhou, J.H.; et al. Squamous cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in older smokers. *Clin. Cancer Res.* **2014**, *20*, 3842–3848. [CrossRef] [PubMed]

24. Lan, Q.; Hsiung, C.A.; Matsuo, K.; Hong, Y.C.; Seow, A.; Wang, Z.; Hosgood, H.D., 3rd; Chen, K.; Wang, J.C.; Chatterjee, N.; et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **2012**, *44*, 1330–1335. [CrossRef]

25. Seow, W.J.; Matsuo, K.; Hsiung, C.A.; Shiraishi, K.; Song, M.; Kim, H.N.; Wong, M.P.; Hong, Y.C.; Hosgood, H.D., 3rd; Wang, Z.; et al. Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Hum. Mol. Genet.* **2017**, *26*, 454–465.

26. Yang, C.; Stueve, T.R.; Yan, C.; Rhie, S.K.; Mullen, D.J.; Luo, J.; Zhou, B.; Borok, Z.; Marconett, C.N.; Offringa, I.A. Positional integration of lung adenocarcinoma susceptibility loci with primary human alveolar epithelial cell epigenomes. *Epigenomics* **2018**, *10*, 1167–1187. [CrossRef]

27. Schmoker, A.M.; Ebert, A.M.; Ballif, B.A. The DCBLD receptor family: Emerging signaling roles in development, homeostasis and disease. *Biochem. J.* **2019**, *476*, 931–950. [CrossRef]

28. Schmoker, A.M.; Weinert, J.L.; Kellett, K.J.; Johnson, H.E.; Joy, R.M.; Weir, M.E.; Ebert, A.M.; Ballif, B.A. Dynamic multi-site phosphorylation by Fyn and Abl drives the interaction between CRKL and the novel scaffolding receptors DCBLD1 and DCBLD2. *Biochem. J.* **2017**, *474*, 3963–3984. [CrossRef]

29. Feng, H.; Lopez, G.Y.; Kim, C.K.; Alvarez, A.; Duncan, C.G.; Nishikawa, R.; Nagane, M.; Su, A.J.; Auron, P.E.; Hedberg, M.L.; et al. EGFR phosphorylation of DCBLD2 recruits TRAF6 and stimulates AKT-promoted tumorigenesis. *J. Clin. Investig.* **2014**, *124*, 3741–3756. [CrossRef]

30. Han, B.; Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **2011**, *88*, 586–598. [CrossRef]

31. Gillison, M.L. HPV and prognosis for patients with oropharynx cancer. *Eur. J. Cancer* **2009**, *45* (Suppl. 1), 383–385. [CrossRef]

32. Giancotti, F.G.; Ruoslahti, E. Integrin signaling. *Science* **1999**, *285*, 1028–1032. [CrossRef] [PubMed]

33. Liu, S.; Calderwood, D.A.; Ginsberg, M.H. Integrin cytoplasmic domain-binding proteins. *J. Cell Sci.* **2000**, *113*, 3563–3571.

34. Zamir, E.; Geiger, B. Molecular complexity and dynamics of cell-matrix adhesions. *J. Cell Sci.* **2001**, *114*, 3583–3590. [PubMed]

35. Juliano, R.L. Signal transduction by cell adhesion receptors and the cytoskeleton: Functions of integrins, cadherins, selectins, and immunoglobulin-superfamily members. *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 283–323. [CrossRef]

36. Parsons, J.T. Focal adhesion kinase: The first ten years. *J. Cell Sci.* **2003**, *116*, 1409–1416. [CrossRef]

37. Brakebusch, C.; Fassler, R. The integrin-actin connection, an eternal love affair. *EMBO J.* **2003**, *22*, 2324–2333. [CrossRef]

38. Eke, I.; Deuse, Y.; Hehlgans, S.; Gurtner, K.; Krause, M.; Baumann, M.; Shevchenko, A.; Sandfort, V.; Cordes, N. Beta (1) Integrin/FAK/cortactin signaling is essential for human head and neck cancer resistance to radiotherapy. *J. Clin. Investig.* **2012**, *122*, 1529–1540. [CrossRef]

39. Bornstein, S.; Schmidt, M.; Choonoo, G.; Levin, T.; Gray, J.; Thomas, C.R., Jr.; Wong, M.; McWeeney, S. IL-10 and integrin signaling pathways are associated with head and neck cancer progression. *BMC Genom.* **2016**, *17*, 38. [CrossRef]

40. Moon, J.H.; Rho, Y.S.; Lee, S.H.; Koo, B.S.; Lee, H.J.; Do, S.I.; Cho, J.H.; Eun, Y.G.; Park, M.W.; Shin, H.A.; et al. Role of integrin beta1 as a biomarker of stemness in head and neck squamous cell carcinoma. *Oral Oncol.* **2019**, *96*, 34–41. [CrossRef]

41. Aebersold, D.M.; Kollar, A.; Beer, K.T.; Laissue, J.; Greiner, R.H.; Djonov, V. Involvement of the hepatocyte growth factor/scatter factor receptor c-met and of Bcl-xL in the resistance of oropharyngeal cancer to ionizing radiation. *Int. J. Cancer* **2001**, *96*, 41–54. [CrossRef]

42. Lo Muzio, L.; Leonardi, R.; Mignogna, M.D.; Pannone, G.; Rubini, C.; Pieramici, T.; Trevisiol, L.; Ferrari, F.; Serpico, R.; Testa, N.; et al. Scatter factor receptor (c-Met) as possible prognostic factor in patients with oral squamous cell carcinoma. *Anticancer Res.* **2004**, *24*, 1063–1069. [PubMed]

43. Lo Muzio, L.; Farina, A.; Rubini, C.; Coccia, E.; Capogreco, M.; Colella, G.; Leonardi, R.; Campisi, G.; Carinci, F. Effect of c-Met expression on survival in head and neck squamous cell carcinoma. *Tumour Biol.* **2006**, *27*, 115–121. [CrossRef] [PubMed]

44. Rosko, A.J.; Birkeland, A.C.; Wilson, K.F.; Muenz, D.G.; Bellile, E.; Bradford, C.R.; McHugh, J.B.; Spector, M.E. Tumor Biomarkers in Spindle Cell Variant Squamous Cell Carcinoma of the Head and Neck. *Otolaryngol. Head Neck Surg.* **2016**, *155*, 106–112. [CrossRef] [PubMed]

45. Baschnagel, A.M.; Tonlaar, N.; Eskandari, M.; Kumar, T.; Williams, L.; Hanna, A.; Pruetz, B.L.; Wilson, G.D. Combined CD44, c-MET, and EGFR expression in p16-positive and p16-negative head and neck squamous cell carcinomas. *J. Oral Pathol. Med.* **2017**, *46*, 208–213. [CrossRef] [PubMed]

46. Ang, K.K.; Berkey, B.A.; Tu, X.; Zhang, H.Z.; Katz, R.; Hammond, E.H.; Fu, K.K.; Milas, L. Impact of epidermal growth factor receptor expression on survival and pattern of relapse in patients with advanced head and neck carcinoma. *Cancer Res.* **2002**, *62*, 7350–7356.

47. Grandis, J.R.; Tweardy, D.J. Elevated levels of transforming growth factor alpha and epidermal growth factor receptor messenger RNA are early markers of carcinogenesis in head and neck cancer. *Cancer Res.* **1993**, *53*, 3579–3584.

48. Rubin Grandis, J.; Melhem, M.F.; Barnes, E.L.; Tweardy, D.J. Quantitative immunohistochemical analysis of transforming growth factor-alpha and epidermal growth factor receptor in patients with squamous cell carcinoma of the head and neck. *Cancer* **1996**, *78*, 1284–1292. [CrossRef]

49. Li, X.; Jung, J.J.; Nie, L.; Razavian, M.; Zhang, J.; Samuel, V.; Sadeghi, M.M. The neuropilin-like protein ESDN regulates insulin signaling and sensitivity. *Am. J. Physiol. Heart Circ. Physiol.* **2016**, *310*, H1184–H1193. [CrossRef]

50. Nie, L.; Guo, X.; Esmailzadeh, L.; Zhang, J.; Asadi, A.; Collinge, M.; Li, X.; Kim, J.D.; Woolls, M.; Jin, S.W.; et al. Transmembrane protein ESDN promotes endothelial VEGF signaling and regulates angiogenesis. *J. Clin. Investig.* **2013**, *123*, 5082–5097. [CrossRef]

51. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [CrossRef] [PubMed]

52. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alföldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* **2019**, *2019*, 531210.

53. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef] [PubMed]

54. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [CrossRef] [PubMed]

55. Deng, M.; Bragelmann, J.; Kryukov, I.; Saraiva-Agostinho, N.; Perner, S. FirebrowseR: An R client to the Broad Institute's Firehose Pipeline. *Database (Oxford)* **2017**, *2017*, baw160. [CrossRef]

56. Nulton, T.J.; Olex, A.L.; Dozmorov, M.; Morgan, I.M.; Windle, B. Analysis of The Cancer Genome Atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. *Oncotarget* **2017**, *8*, 17684–17699. [CrossRef]

57. Verschuur, H.P.; Irish, J.C.; O'Sullivan, B.; Goh, C.; Gullane, P.J.; Pintilie, M. A matched control study of treatment outcome in young patients with squamous cell carcinoma of the head and neck. *Laryngoscope* **1999**, *109*, 249–258. [CrossRef]

58. Gilroy, J.S.; Morris, C.G.; Amdur, R.J.; Mendenhall, W.M. Impact of young age on prognosis for head and neck cancer: A matched-pair analysis. *Head Neck* **2005**, *27*, 269–273. [CrossRef]

59. O'Regan, E.M.; Timon, C.; Sheils, O.; Codd, M.; O'Leary, J.J.; Toner, M. Squamous cell carcinoma of the head and neck in young Irish adults. *Br. J. Oral Maxillofac. Surg.* **2006**, *44*, 203–206. [CrossRef]

60. 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. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]

61. The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338.

62. Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **2019**, *47*, D419–D426. [CrossRef] [PubMed]

63. Hornbeck, P.V.; Zhang, B.; Murray, B.; Kornhauser, J.M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512–D520. [CrossRef] [PubMed]