Integrative analysis in head and neck cancer reveals distinct role of miRNome and methylome as tumour epigenetic drivers

Katarina Mandić[1]*, Nina Milutin Gašperov[2]*, Ksenija Božinović[2], Emil Dediol[3], Jure Krasić[4], Nino Sinčić[4,5,6], Magdalena Grce[2], Ivan Sabol[2]**, Anja Barešić[1]**

*/**Equally contributing authors

1 Division of Electronics, Ruđer Bošković Institute, Zagreb, Croatia

2 Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia

3 Clinical Hospital Dubrava, Department of Maxillofacial Surgery, Zagreb, Croatia

4 Department of Medical Biology, University of Zagreb School of Medicine, Zagreb, Croatia

5 Centre of Excellence in Reproductive and Regenerative Medicine, University of Zagreb School of Medicine, Zagreb, Croatia

6 Biomedical Research Center Šalata, University of Zagreb School of Medicine, Zagreb, Croatia

Running title: Integrative HNSCC omics analysis

Conflict of interests:

The authors declare no potential conflicts of interest

Corresponding author: Ivan Sabol, PhD; Division of Molecular Medicine, Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia; E-mail address: isabol@irb.hr; Phone: +38514561110; Fax: +38514561010;

Keywords: head and neck cancer, HPV, miRNome, methylome, transcriptome, integrative analysis

Abstract

Head and neck cancer (HNC) is the sixth most common malignancy worldwide with a relatively low 5-year survival rate, mainly because it is diagnosed at a late stage. Current literature suggests that epigenetics is the main driving force for this type of cancer and that HPV infection plays a role in a subset of HNC. In this study, we aimed to find potential epigenetic biomarkers by integrating miRNome, methylome and transcriptome analyses.

From the fresh tissue samples of HNC tumours, we selected a group for miRNome, methylome and transcriptome profiling in comparison with appropriate control samples. Bioinformatic analyses were performed in R v4.2.2. Count normalization and group differential expression for mRNA and previously obtained miRNA count data were performed (DESeq2_v1.36). Gene set enrichment analysis was performed and visualised (gProfiler2_v0.2.1). Identification of miRNA targets was performed by querying miRTarBase (multiMiR_v1.18.0). For the integrative analysis, we performed kmeans clustering using 8-12 clusters and nstart 100 (stats_v4.2.2).

Transcriptome analysis divided the samples into cancer' and control' clusters without reference to HPV status or anatomic location. The differentially expressed genes (n=2781) were associated with signaling pathways of tumour progression. We identified a cluster of genes under the control of the transcription factor E2F that were significantly underexpressed in cancer tissue along with T-cell immunity and transcriptional regulation related genes. Among tumour overexpressed genes, we found genes related to cell cycle regulation and vasculature.

In this Integrative study we showed distinct role of miRNome and methylome as epigenetic drivers in HNC.

Introduction

Head and neck cancer is the sixth most common malignancy worldwide, with head and neck squamous cell carcinoma (HNSCC) accounting for more than 90% of cases. Unfortunately, HNSCC often gets diagnosed in a late phase when it is challenging to treat the tumour, which contributes to poor 5-year survival rate, currently at 66% (1). Even though this type of cancer arises from a single cell type, i.e. the squamous cell, HNSCC is surprisingly of very heterogeneous nature (2). Primary site of tumour origin contributes to tumour heterogeneity, with most common sites being oral cavity, oropharynx, larynx, and sinonasal tract (3). Besides tumour location, the infection with high-risk human papilloma virus (HPV) types, especially in oropharyngeal area predominately affects HNSCC nature and patients' survival (4). HPV positive and HPV negative HNSCC can substantially differ in terms of etiology, genetics and epigenetics (1,5). HPV-positive HNSCC mostly affect younger population, with no or low levels of smoking and/or alcohol use (6). Moreover, unlike the HPV-negative group, preferable site of tumour origin for HPV-positive HNSCC is oropharynx and it has been reported in many Western countries that this group of patients is associated with better overall survival and better response to therapy (7). There has been significant shift in the incidence of HPV-positive oropharyngeal squamous cancer in Western countries, where HPV is associated to HNSCC in almost 70% of cases (8), and in fact, increase in 5-year survival of overall HNSCC cases could be contributed to higher prevalence of HPV-associated HNSCC (1).

Up to date, many studies reported better outcomes for HPV-positive HNSCC (9–11), however the consensus on specific biomarkers, which could improve diagnostic, prognostic and/or therapeutic approaches is still missing. Lack of consensus biomarkers could partially be owed to differences in studied populations sometimes inadequately registered tumour

site of origin and possibly to the lack of stratification by the HPV status. Therefore, we consider the cohort presented below to be a specific case of tumours in terms of lifestyle, i.e. higher smoking and alcohol intake rates.

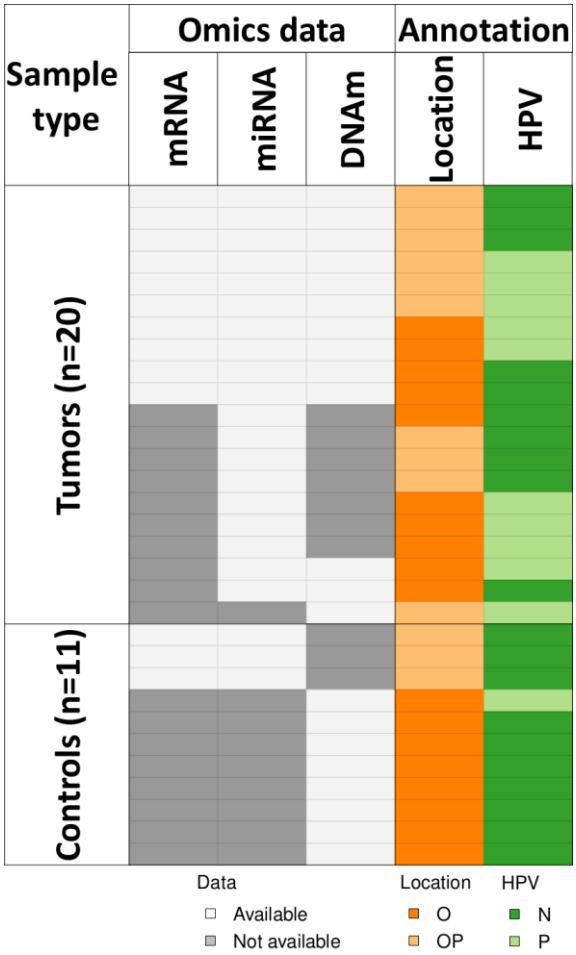
In this study, we aimed to identify epigenetic patterns in HNSCC since the main driving force for the HNSCC development is considered to be epigenetics, rather than genetics. We have previously analyzed the content and transcriptional levels of whole genome microRNA (miRNA) profile (miRNome) in HPV-positive and HPV-negative oral and oropharyngeal HNSCC samples using next generation sequencing and validated with qRT-PCR. We have also analyzed the whole genome methylation profile (methylome) in the same set of samples using the whole genome methylation array and validated by pyrosequencing. In this study, we assessed the whole genome transcriptome analysis on the same sample set and combined the mRNA, miRNA and methylation information to elucidate relevant epigenetic mechanisms and their interactions. Motivated by the limited literature regarding integrative studies on epigenetic changes, we set out to elucidate whether miRNAs or DNA methylation have more prominent impact on HNSCC development. Moreover, we aimed to identify key mechanisms and/or cellular signaling pathways of genes that have been differentially expressed in HNSCC compared to healthy controls. Consequently, we provide a better understanding of the epigenetic influence on the development and progression of HPV-positive and HPV-negative HNSCC, shedding light on mechanisms underlying prognostic, diagnostic and therapeutic strategies. In addition, integrative analysis points out key target genes and signaling pathways deregulated in HNSCC in Croatian and similar populations.

97 ## Material and Methods

98 ### Patient material

99 A total of 61 fresh oropharyngeal (OP) and oral (O) HNSCC primary tumour samples have
100 been obtained from Dubrava Clinical Hospital in the 2015-2018 time period. Twenty-two of
101 them and three non-malignant tonsil control samples were included in the miRNA profiling
102 (12), column 2 in **Table 1**. Of 22 sequenced samples, two were excluded, due to one sample
103 being a recurrent tumour and one outlier. miRNome analysis and detailed patients'
104 socioepidemiologic and clinical data have been described in our previous study . From the
105 previously published analysis of HNSCC methylome (13), of 22 samples included in miRNA
106 study, 12 HNSCC samples were included in the whole genome methylation profiling (column
107 3, **Table 1**), as well as one additional sample that had methylome but no transcriptome and
108 miRNome profile. Eight healthy swabs of oral mucosa were included in methylome profiling,
109 which served as control samples .
110 For the transcriptome profiling, we selected 10 fresh HNSCC samples (6 oral and 4
111 oropharyngeal) with previous miRNome and methylome data (column 1, **Table 1**). Three
112 fresh tonsil tissues from the miRNome dataset were included in transcriptome profiling as
113 control tissue. HPV status was assessed by PCR in all samples as described previously .
114

Table 1. Samples used in integrative analysis. Sequenced samples are shown in white. O=oral, OP=oropharyngeal, P= HPV-positive, N=HPV-negative.



Transcriptome profiling

As stated above, 13 samples were selected for transcriptome analysis in the current study to complement previously obtained omics results. Sample annotation for samples used in transcriptome analysis and subsequent integrative epigenetic study is presented in **Table 1**. Transcriptome was assessed using TrueSeq Stranded mRNA (Illumina) kit for library preparation, followed by sequencing on Nextseq500 instrument using NextSeq HIGH 150 cycles (Illumina) reagents on high output flow cell in paired end mode. Debarcoding was performed using Basespace (Illumina) platform. Quality control of sequencing data was performed with FastQC v0.11.9. Sequences were aligned to the human reference genome

GRCh37 downloaded from GENCODE (released in 2013) and quantified using salmon v1.5.0, (14) reporting estimated fragment counts per transcript.

Bioinformatics preprocessing

All bioinformatics analyses listed hereafter were performed in R v4.2.2, and refer to R language packages, unless otherwise stated. The preprocessing consisted of transcriptome gene count determination, reprocessing of miRNA counts (for a subset of samples from Božinović 2019) and reprocessing of differentially methylated CpGs (for a subset of samples from Milutin Gašperov 2020), here expanded to CpG islands.

Count normalization and group differential expression for mRNA and the previously obtained miRNA count data was performed with DESeq2 v1.36. Pre-filtering of the data set was performed to reduce genes that have nearly no information about gene expression by filtering out genes and miRNAs with total counts across all samples less than 10. Count variance stabilization transformation was performed with regularized logarithm transformation (DESeq2 rlog function) and used as input to calculate principal components with plotPCA function from DESeq2 and Euclidean sample distances with dist function from the stats v4.2.2 package. Counts normalization was performed with DESeq function with log fold change shrinkage. We calculated differential expression in tumour *versus* control and HPV positive *versus* negative samples (**Table 1**). Differentially expressed genes (DEGs) and miRNA (DEmiR) were selected with BH-adjusted $p<0.05$ set as significant.

Gene set enrichment analysis (GSEA) was performed and visualized using gProfiler2 v0.2.1 on DEGs using annotated human genes as the background set for the hypergeometric test. We applied the tailor-made multiple test correction method Set Counts and Sizes (SCS) (15) and a p-value threshold of 0.05. GSEA was performed across multiple databases: Gene Ontology: Biological Processes (GO:BP) (16), Kyoto Encyclopedia of Genes and Genomes

151 (KEGG) (17), Reactome (18), miRTarBase (19) and TRANSFAC (20). We used these
152 parameters for all subsequent GSEA.
153 Identification of miRNA targets was performed by querying significant DEGs against DEmiRs
154 using multiMiR v1.18.0 and the miRTarBase database (21), selecting only experimentally
155 validated miRNA-gene interactions.
156 Previously published DNA methylation (DNAm) data (13) was loaded with RnBeads v2.14.0.
157 Differential methylation (DM) was calculated comparing tumour *versus* control groups on
158 CpG sites and islands (CGIs). Annotation of CpG sites merging into islands was obtained from
159 RnBeads.hg19 v1.28.0 package. All results are presented on CGIs unless otherwise stated.
160 CGI DM was defined as the quotient of methylation between groups across all CpG sites in a
161 CGI, with significance level set to FDR-adjusted $p<0.05$. Promoter region annotations were
162 obtained from the RnBeads package, defined as 1500 bases upstream and 500 bases
163 downstream of the transcription start site. CGI targets were assigned where CGI overlapped
164 a promoter region.

165 **Integrative analysis**

166 Workflow for comparing all omics data and identifying specific mechanisms in HNSCC is
167 summarized in **Figure 1**.

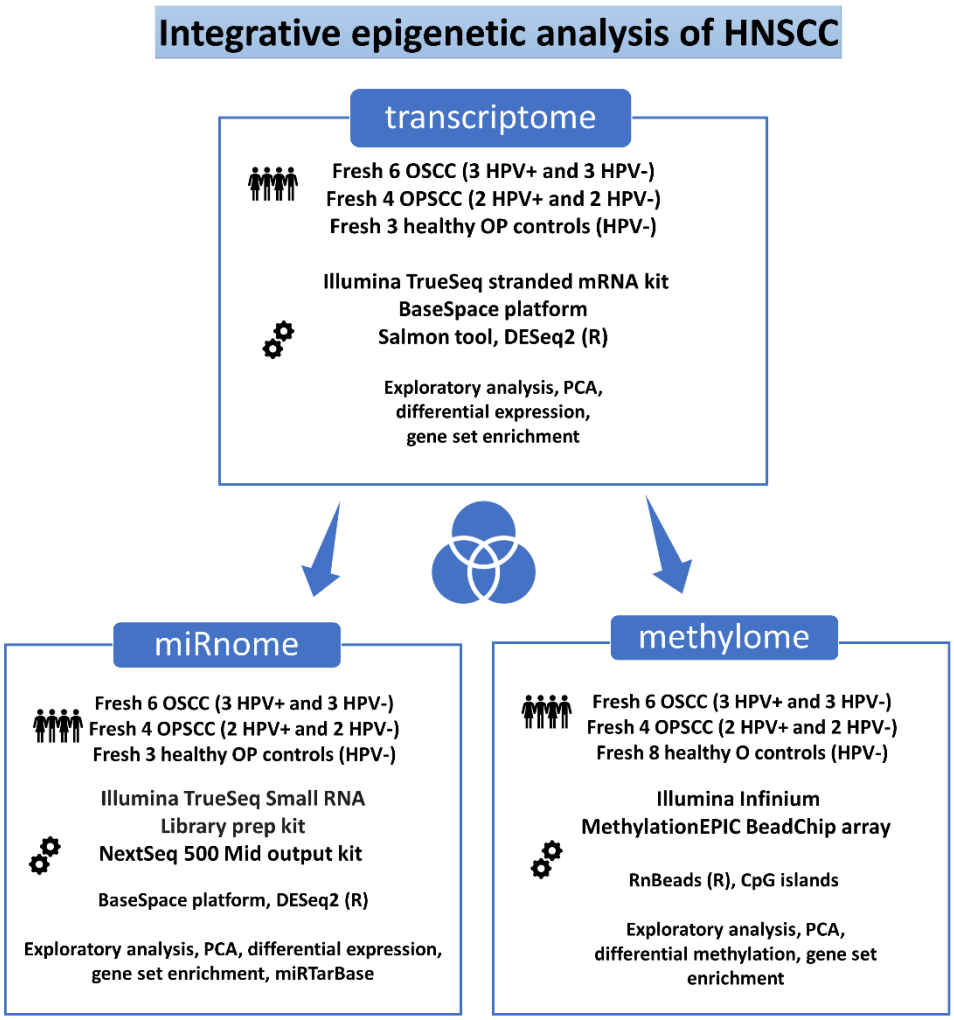


Figure 1. Schematic representation of integrative epigenetic HNSCC study.

Three -omics datasets were compared, centered on identified DEGs (**Supplemental table S1**). Multiple miRNA-gene pairs were reduced by selecting only the miRNA with highest absolute value of Spearman correlation of counts (mirRNA_Rho) for the given DEG. Multiple gene–CGI pairs were not as common, and the CGI with the lowest p-value was selected for every DEG. Resulting table contained DEGs and their associated miRNA and/or CGI. We assessed the interplay of DEmiRs, DM CGIs and DEGs in two ways: by obtaining simple overlap of DEGs targeted by DEmiRS or DM CGIs and kmeans clustering. The simple overlap was based on whether DEG's expression levels were associated with DEmiR's activity and/or CGI

179 methylation levels. Kmeans clustering was performed on log-transformed, centered and
180 scaled values of fold change for DEGs and DEmiRs, and DNAm quotient. We performed
181 kmeans clustering using stats v4.2.2 package, using 8-12 clusters and nstart 100, with the
182 best performing result for 11 clusters.
183

184 ## Results

185 ### Transcriptome profiling

186 The whole transcriptome profiling was performed on tissue samples from 13 individuals with
187 different anatomical locations and HPV infection status (**Table 1**). The top two principal
188 components of the transcriptome (**Figure 2a**), which explain 55% of the variance in the
189 samples, splits samples into two clusters of tumour and control samples, with no obvious
190 contribution of HPV status or tumour anatomical location. We observe the same pattern in
191 miRNome and methylome data (**Figure 2b** and **c**).

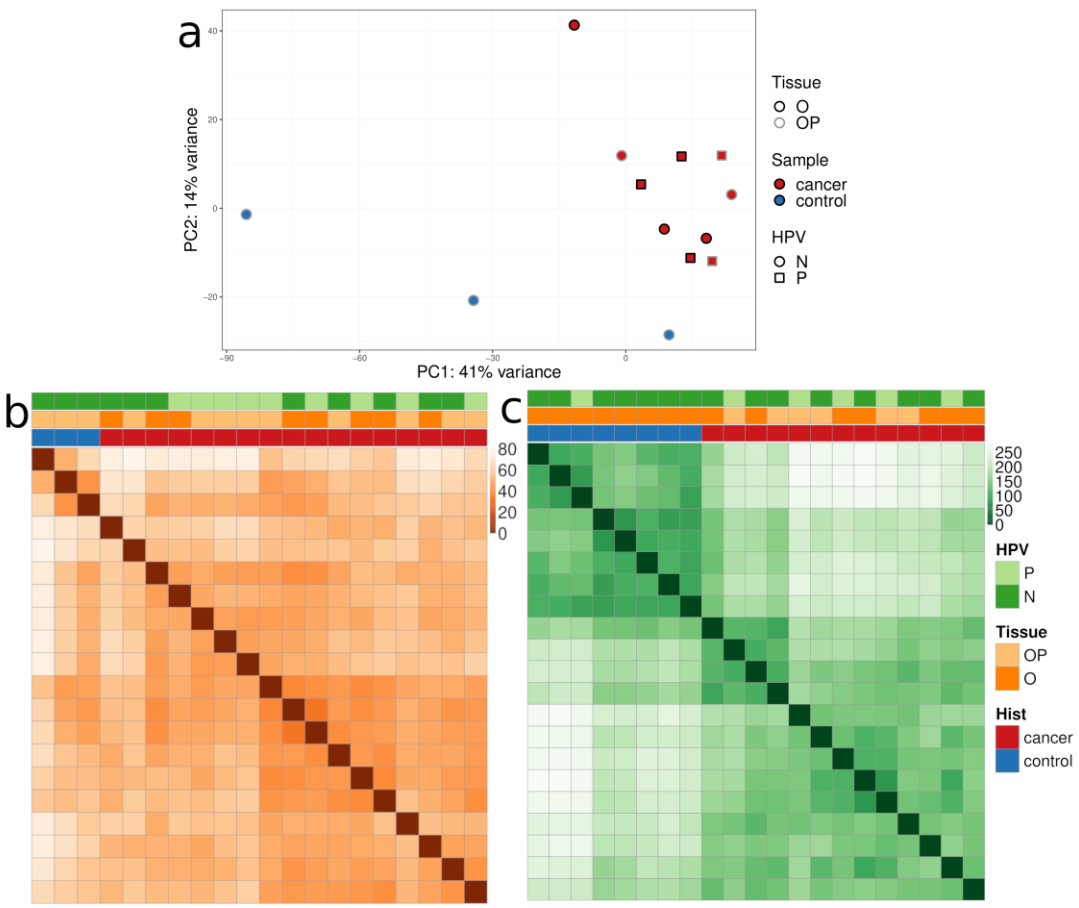
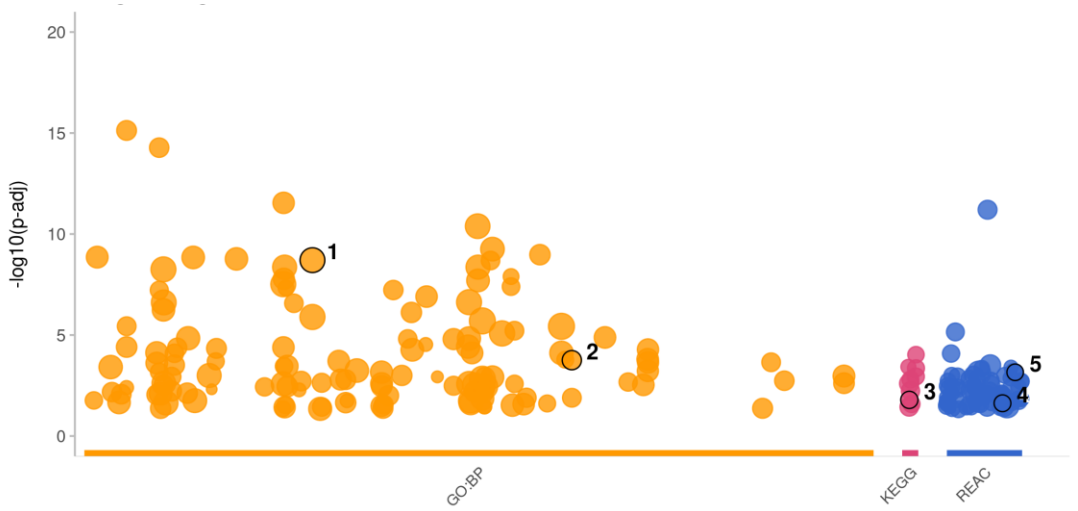


Figure 2. Exploratory analysis of three omics data sets: a) PCA plot for transcriptome data; b) sample distance heatmap for miRNome data; c) sample distance heatmap for methylome. N - HPV-negative, P - HPV-positive O - oral tumours, OP - oropharyngeal tumours.

Differential expression and GSEA

Exactly 18104 genes had no less than 10 detected reads across samples. Differential expression analysis revealed that 1248 (6.9%) of genes were significantly overexpressed and 1533 (8.5%) were significantly underexpressed in tumour (BH-adjusted $p<0.05$), when compared to transcriptomes of control samples. GSEA on DEGs showed enrichment of signaling pathways associated with tumour progression, such as cell cycle regulation and oncogenic signaling pathways such as NF-kB (**Figure 3**), for a full list of terms see **Supplemental table S2**.



id	source	term id	term name	term size		p value
1	GO:BP	GO:0032502	developmental process	4582		2.0e-09
2	GO:BP	GO:0070661	leukocyte proliferation	235		1.8e-04
3	KEGG	KEGG:04064	NF-kappa B signaling pathway	102		1.6e-02
4	REAC	REAC:R-HSA-453276	Regulation of mitotic cell cycle	88		2.4e-02
5	REAC	REAC:R-HSA-8852276	The role of GTSE1 in G2/M progression after G2 checkpoint	79		7.0e-04

Figure 3. Gene enrichment analysis of tumour *vs.* control samples. Databases shown are GO:BP (orange), KEGG (pink), and REACTOME (REAC, blue). The highlighted terms correspond to key concepts associated with cancer progression and regulation.

Differential miRNA and DNA methylation assessment

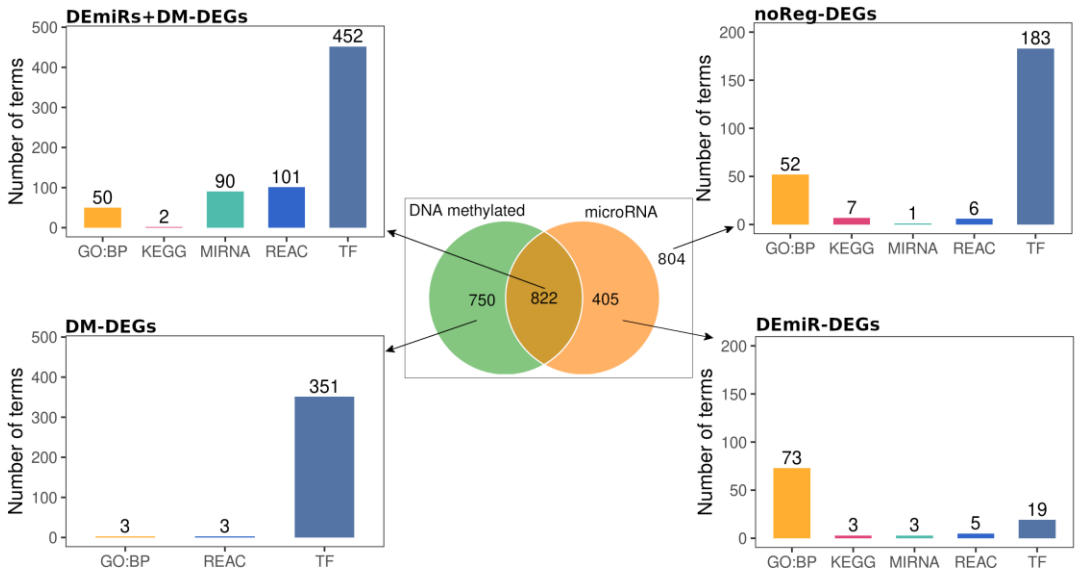
Integration of 514 significant DEmiRs based on MiRNome data (12) with the database containing experimentally supported miRNA-target gene interactions yielded 3042 interactions among 90 different miRNAs and 1227 target genes, hereafter termed DEmiR-DEGs. In our analysis Most of these genes were regulated by a single miRNA, however some were associated with multiple miRNAs. Notably, two genes with the highest number of associated DEmiRs, *BCL2* and *MYC*, which play a role in regulating cell proliferation and apoptosis, and were associated with 18 and 16 miRNAs, respectively. In addition, some miRNAs were associated with many target genes, harboring potential therapeutic value by potentially affecting many cancer-related signaling pathways. An extreme example is a frequently reported tumour suppressor hsa-miR-26b-5p putatively forming interactions with

221 255 unique DEGs. Within our dataset hsa-miR-26b-5p showed a high negative expression
222 correlation with known oncogenes such as *MMP10* (Rho= -0.81, p<0.005), *IGSF3* (Rho= -0.79,
223 p<0.005) and *ARNTL2* (Rho= -0.79, p<0.005) (**Supplemental table S1**).
224 Overall methylation levels in our samples showed that more than 300,000 CpG sites were
225 hypomethylated in tumour samples compared to controls (FDR-adjusted p<0.05), while
226 72,000 were hypermethylated (**Supplemental table S3**). We also noted a larger number of
227 hypomethylated CGIs (6557) than hypermethylated (5457), although the difference is not as
228 pronounced as in CpG sites (**Supplemental table S4**). The DM CpGs predominantly fell into
229 gene bodies, but also some fell in intergenic regions corresponding to repeated elements.
230 These regions consistently showed a larger number of hypomethylated CpGs compared to
231 hypermethylated ones (**Supplemental figure S1**).
232 To identify deregulated tumour suppressors in our samples, we used the TSGene database
233 2.0 (22) containing tumour suppressors curated from several thousand publications
234 combined with expression and mutational profiles. We pinpointed 10 genes that were
235 underexpressed and contained hypermethylated CGIs in the promoter region in our samples
236 (**Supplemental table S5**).
237 To assess the effects of transcription factors on DEGs that had differentially methylated CGIs
238 (DM-DEGS) within our dataset we performed GSEA using transcription factor database
239 TRANSFAC. We found associations with several groups of transcription factors including E2F,
240 SP/KLF and AP-2 families (**Supplemental table S6**).
241 Of note we found the enrichment of genes regulated by the E2F family of TFs (SCS-adjusted
242 $p<10^{-43}$), which play a critical role in cell cycle control. Furthermore, GSEA on the same
243 dataset using the Reactome database showed enrichment of genes involved in cell cycle
244 regulation such as G1/S Transition and G2/M Checkpoints (SCS-adjusted $p<10^{-4}$), supporting
245 the above findings.

246 **Integrative analysis: Influence of miRNome and methylome on transcription on**

247 **transcriptional activity in HNSCC**

248 Each DEG was grouped based on the association with DEmiR or DM CGI, shown in **Figure 4**.
249 We obtained four groups: DEGs associated only with DEmiR (DEmiR-DEGs), DEGs associated
250 only with DM CGIs (DM-DEGs), DEGs associated with both DEmiR and DM (DEmiR+DM-
251 DEGs), DEGs not associated with either DEmiR or DM CGIs (noReg-DEGs). The groups had
252 different proportion of differentially represented GSEA terms, i.e. GO:BP, KEGG, miRTarBase
253 (MIRNA), Reactome and TRANSFAC (TF) (**Figure 4**, **Supplemental table S7**). The DM-DEGs
254 group were enriched by genes regulated by TFs but no significant results of GO biological
255 pathways was detected. The DEmiR-DEGs group were enriched with pathways related to
256 signaling and immunological systems. The noReg-DEGs group was enriched with immunological
257 processes and muscle system processes. DEmiR+DM-DEGs were enriched with general regulatory
258 processes: biological, metabolic and cellular processes. These terms are high level terms in GO
259 biological processes, suggesting that genes in the intersection have broad regulatory roles.



260

261 **Figure 4.** DEGs between tumour and control samples, in association with DEmiR or DM
262 promoters. Four panels show number of terms from GSEA for each group.

263 Clustering of DEGs with accompanying DEmiR and DM levels revealed a critical distinction in
264 the epigenetic factors driving HNSCC. The identified clusters predominantly demonstrated
265 either significant DEmiRs or DM. This suggests that in HNSCC, gene expression is
266 predominantly influenced by either miRNA regulation or methylation, but not both
267 concurrently. This key finding illuminates the mutually exclusive roles of miRNA and
268 methylation in shaping the gene expression landscape in HNSCC.
269 It is generally believed that miRNA and DNAm have an inhibitory effect on gene expression,
270 and this pattern was evident in clusters 3, 4, 7 and 10. In addition, we observed miRNA
271 expression and methylation patterns in clusters 1, 5, 6, and 9 that were indicative of
272 noncanonical regulatory activity. It is important to mention that due to multiple target genes
273 of miRNAs there was a significant overlap of underexpressed miRNAs in clusters 3 and 9
274 (73%) and overexpressed miRNAs in clusters 5 and 7 (77%).
275 GSEA analysis of each cluster linked expression and epigenetic patterns to biological function.
276 Representative terms that best described each cluster are shown in **Figure 5** and the full
277 GSEA results for each cluster are shown in **Supplemental table S7**. DM-DEGs in clusters 1, 4,
278 6 and 10 were highly enriched with genes regulated by the E2F family of TFs. Most of the
279 genes from the E2F family were not differentially expressed in our samples (**Supplemental**
280 **figure S2**), although it has been previously shown that the methylation status of CpG sites of
281 E2F binding motif influences the binding of the transcription factor (23,24). In addition,
282 cluster 6 contained underexpressed genes with hypomethylated CGIs and was enriched with
283 genes regulated by the ZF5 transcription factor, which can function as a repressor (25).
284 Clusters 4 and 5 were associated with similar terms related to various cell cycle processes
285 but have different patterns of miRNA expression and DNAm.
286

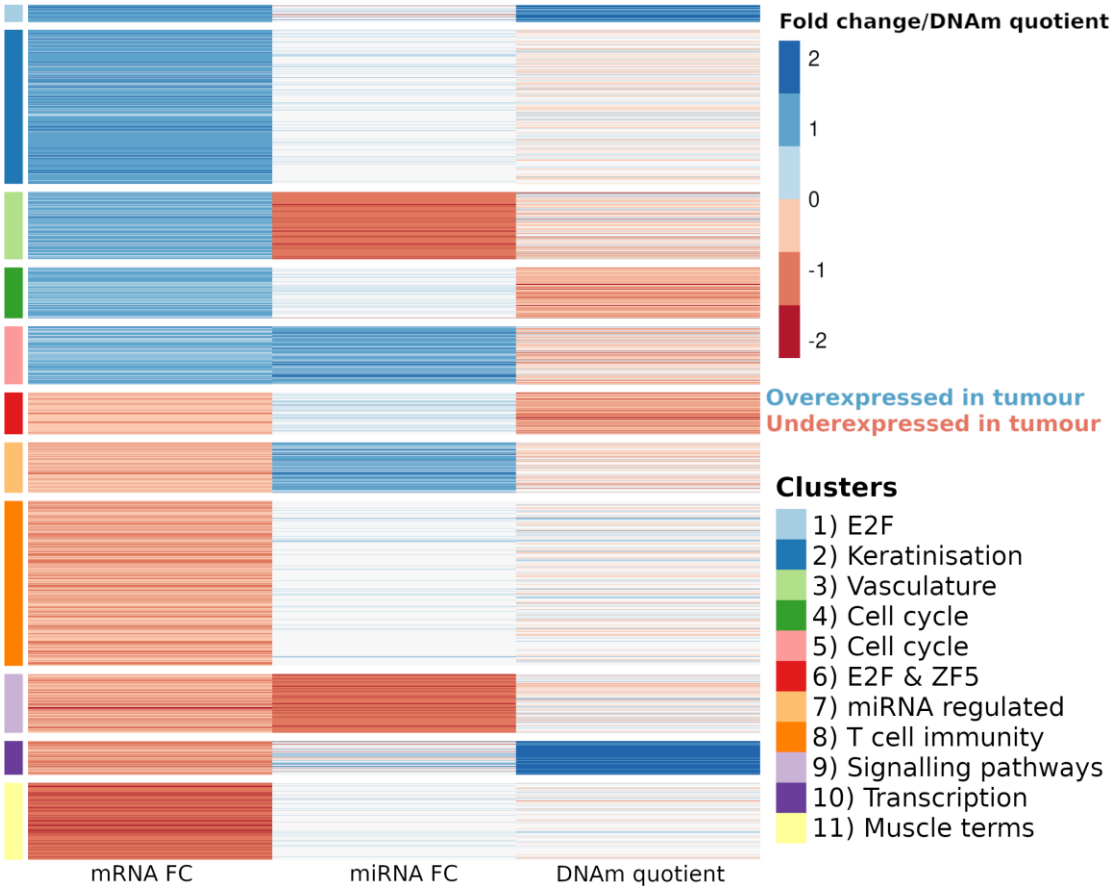


Figure 5. Kmeans clustering on integrated transcriptome, miRNome and methylome data on DEGs. Fold change values for transcriptome and miRNome and DNAm quotient were scaled and normalized prior to clustering. Eleven clusters are functionally described by applying GSEA over several databases (GO:BP, KEGG, REACTOME, MIRNA and TRANSFAC).

Influence of HPV

HPV influence on gene expression was found to be inconclusive (**Figure 2**). We observed a small number of genes and miRNAs, and no DM CGIs or sites that were significantly changed while comparing HPV positive and negative tumour sample groups (FDR-adjusted $p<0.1$). Number of DEGs associated with HPV status was 44, of which 25 were overexpressed (**Supplemental table S9** and **S10**). Among these, we focused on the selection of DEGs found to be functionally relevant and/or previously found to be differentially expressed in HPV-associated cancers (**Figure 6**).

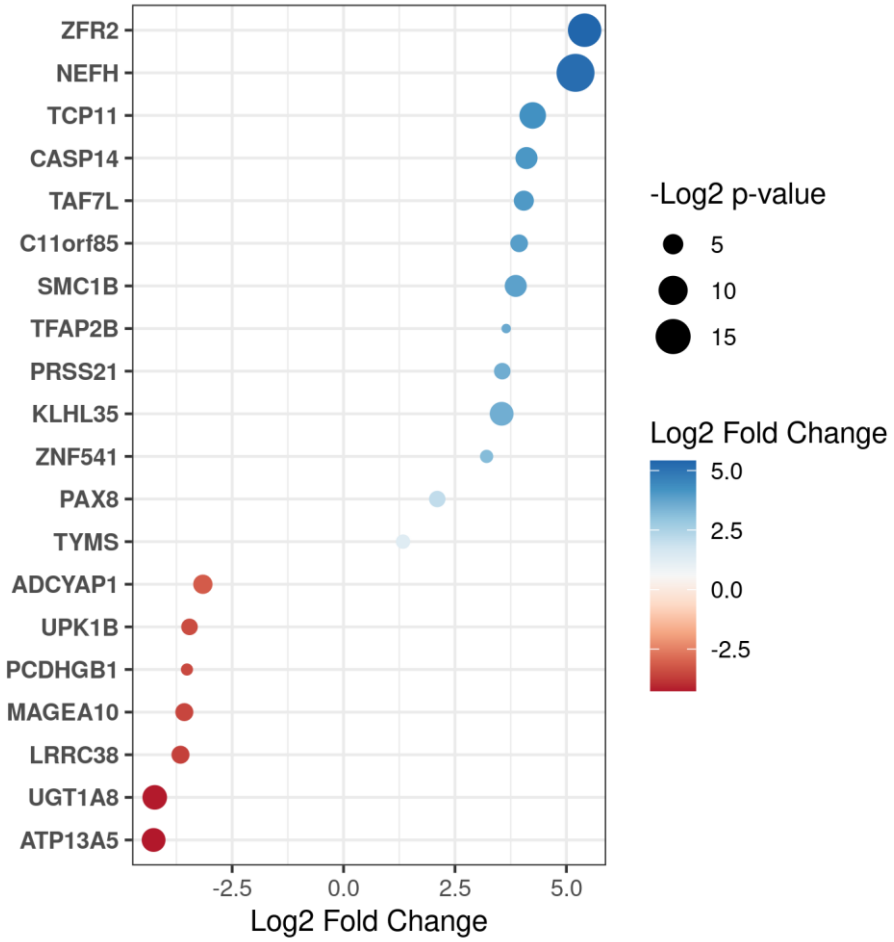


Figure 6. Top 20 DEGs in HPV positive compared to HPV negative tumour patients.

Discussion

In this study, we aimed to integrate miRNome and methylome analysis with transcriptome analysis on the same samples to elucidate whether miRNAs or DNAm have more prominent impact on HNSCC development. The main goal was to identify key mechanisms and/or cellular signaling pathways of DEGs in HNSCC compared to healthy control samples.

Herein, we observed that transcriptome analysis divided samples into two major clusters: tumour and control samples, with no obvious contribution of HPV status or tumour anatomical location. Similar results were seen in the miRNA expression profile study and

311 DNAm study in an overlapping set of samples (**Figure 2 b** and **c**) as previously published
312 (12,13). This suggests that these tumours may share similar cells of origin or common
313 pathways during tumourigenesis.
314 The emergence of tumour progression pathways among DEGs (n=2781, **Figure 3**) resembles
315 transcriptomes of radioresistant HNSCCs, where hypoxia response, p53 pathway, NF-kB
316 pathway and inflammatory response were found abnormally activated. Interestingly, some
317 authors found that NF-kB and other transcription factors signatures differentiate HPV-
318 positive and HPV-negative HNSCCs.
319 Using unsupervised clustering, we found a cluster of genes under control of the transcription
320 factor E2F, that were significantly underexpressed in tumour samples in comparison to
321 controls (**Figure 5**, cluster 6). This is in line with overall hypomethylation in tumours and
322 points to exclusive control of overexpressed miRNAs. The genome-wide hypomethylation of
323 CpG sites is a known phenomenon that occurs in most cancers (26,27). E2F family members
324 expression was also found to be negatively regulated by miRNAs in other cancers, like lung
325 cancers, pediatric brain tumours, and chronic myeloid leukemia. In addition, genes that were
326 significantly associated with the E2F were found in clusters 1, 5 and 10 (**Supplemental table**
327 **S6**).
328 In cluster 8, we found genes related to T cell immunity underexpressed in tumour samples in
329 comparison to control samples (**Figure 5**), mostly found to be targets of overexpressed
330 miRNAs, and in less extent associated with hypermethylation. This finding is in line with
331 trends detected in several previous methylation studies of head and neck cancers (28,29),
332 and in cervical cancers (30,31).
333 Among underexpressed genes in tumours, we observed a cluster of DM-DEGs related to
334 regulation of transcription (**Figure 5**, cluster 10). Our result showed that the cluster mostly
335 contains genes with hypermethylation of CGIs in promoter regions and some genes that

336 were targeted by overexpressed miRNAs. This implies that the hypermethylation of CGIs in
337 promoter region of genes in this cluster may have disrupted binding of transcription factors
338 required for their expression, that is a well-studied phenomena (32–34).
339 Among overexpressed genes in tumours were those that belong to cell cycle regulation
340 (**Figure 5**, clusters 4 and 5) and vasculature (**Figure 5**, cluster 3). Herein, we found cell cycle
341 genes have mostly hypomethylated promoter regions. On the other hand, our result showed
342 that genes involved in vascularization in tumours are DEmiR+DM-DEGs, likely influenced by
343 both epigenetic drivers. Moreover, other authors found a link with DNAm and cardiovascular
344 diseases directly rather than changes driven by miRNA expression (35), or that both
345 epigenetic changes are involved in vascular diseases, claiming that, the most prominent
346 gene cluster is activated via hypomethylation (36). Interestingly, the same authors found
347 keratin mRNAs upregulated in atherosclerosis and concluded that DNA hypomethylation
348 may play a role in keratin gene activation and ectopic keratin expression (36). Keratins are
349 characteristic to epithelial cells and are not expressed in normal arteries and induction of
350 keratin expression in atherosclerosis has been detected even earlier (37). Our study
351 detected similar findings in HNSCC; keratin mRNAs were found upregulated, that was
352 partially under control of DNAm and partially under control of miRNome (**Figure 5**, cluster 2).
353 Herein, we found that cluster 5 (**Figure 5**) contained upregulated miRNA and their
354 corresponding upregulated target genes. Conversely, cluster 9 displayed downregulated
355 miRNA and corresponding downregulated target genes. This positive correlation of miRNAs
356 and target genes has been previously observed in multiple studies (38,39). While our results
357 could imply either a direct or indirect effect of miRNA on DEmiR-DEGs potentially through
358 negating the action of repressive miRNA, protein complexes (38), it is important to note that
359 our findings represent correlation, not causation. Further studies are required to establish
360 the causal mechanisms involved. In addition, it is worth considering that miRNAs often have

361 multiple targets, and the interactions in existing databases may be derived from high-
362 throughput methods that may not reflect the full complexity of in vivo biological interactions.
363 A small number of genes were found significantly differentially expressed in HPV positive
364 *versus* HPV negative tumours (**Figure 6**) in our study, a subset of which was previously found
365 differentially expressed in head and neck cancer and related to HPV status, i.e. *NEFH*, *ZRF2,*
366 *TAF7L, ZNF541* and *TYMS*. Notably, some authors have already found hypermethylation of
367 *NEFH* (protein kinase binding and microtubule binding), in pharyngeal squamous cell
368 carcinoma, which significantly correlated with HPV positivity (40). *ZFR2* gene codes for RNA-
369 binding protein characterized by its domain associated with zinc fingers and zinc ions. It was
370 found overexpressed in HPV-positive HNSCC patients when compared to HPV-negative
371 patients and has been flagged as prognostic for the cervical cancer cases (41). The
372 expression of the transcription factor *TAF7L* (TAF7-like RNA polymerase II) has been already
373 found increased, even 220-fold, in HPV positive HNSCCs (42). *ZNF541* encodes a zinc finger
374 protein supposed to be a component of chromatin remodeling complexes. The
375 hypomethylation and higher expression of this gene was also observed in HPV-positive
376 oropharyngeal squamous cell carcinoma (OPSCC) and significantly associated with a better
377 overall survival, independent of HPV status (43). *TYMS* gene codes for thymidylate
378 synthetase that catalyzes the methylation of deoxyuridylate to deoxythymidylate. In HPV-
379 positive cancers they have found it overexpressed, indicating that HPV-positive
380 oropharyngeal cancers may be more resistant to 5-fluorouracil chemotherapy than HPV-
381 negative (44).
382 To our knowledge, we presented the most comprehensive study on an overlapping set of
383 samples where the integration of miRNome, methylome and transcriptome analyses in
384 HNSCC was conducted. We acknowledge certain limitations of our study, including the
385 relatively small sample size, which may affect the validity of our findings. In addition,

386 obtaining control samples that are genuinely representative of healthy tissue poses its own
387 set of challenges. Associating miRNome and methylome with transcriptome profiles, we
388 found many DEGs in HNSCC in comparison to control samples, that were mainly associated
389 with cell cycle regulation, transcription, immunity and tumour development. We
390 demonstrated the majority of genes belong to clusters associated exclusively with miRNome
391 or methylome and, to a lesser extent, under control of both epigenetic mechanisms.
392

393 **Acknowledgements**

394 The authors are thankful to Prof. Marinka Mravak-Stipetić for oral normal tissue samples
395 collection, and to Dr. Ruth Tachezy for providing reference non-malignant tonsil control
396 tissues. The authors are also grateful to Jasminka Golubić Talić for her technical assistance.
397

398 **References**

399 1. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck
400 squamous cell carcinoma. Nat Rev Dis Primers. 2020;6:1–22.
401 2. Canning M, Guo G, Yu M, Myint C, Groves MW, Byrd JK, et al. Heterogeneity of the
402 Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on
403 Immunotherapy. Front Cell Dev Biol. 2019;7:52.
404 3. Yan W, Wistuba II, Emmert-Buck MR, Erickson HS. Squamous Cell Carcinoma -
405 Similarities and Differences among Anatomical Sites. Am J Cancer Res. 2011;1:275–300.
406 4. Kobayashi K, Hisamatsu K, Suzui N, Hara A, Tomita H, Miyazaki T. A Review of HPV-
407 Related Head and Neck Cancer. J Clin Med. 2018;7:241.
408 5. Joseph AW, D'Souza G. Epidemiology of human papillomavirus-related head and neck
409 cancer. Otolaryngol Clin North Am. 2012;45:739–64.
410 6. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human
411 Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. J Clin Oncol.
412 2015;33:3235–42.
413 7. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated Head and Neck Cancer:
414 A Virus-related Cancer Epidemic – A Review of Epidemiology, Biology, Virus Detection
415 and Issues in Management. Lancet Oncol [Internet]. 2010 [cited 2018 Sep 12];11:781–9.
416 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5242182/
417 8. Marur S, Forastiere AA. Head and Neck Squamous Cell Carcinoma: Update on
418 Epidemiology, Diagnosis, and Treatment. Mayo Clin Proc. 2016;91:386–96.
419 9. Perri F, Longo F, Caponigro F, Sandomenico F, Guida A, Della Vittoria Scarpati G, et al.

Management of HPV-Related Squamous Cell Carcinoma of the Head and Neck: Pitfalls and Caveat. Cancers (Basel). 2020;12:975.
10. Kofler B, Laban S, Busch CJ, Lörincz B, Knecht R. New treatment strategies for HPV-positive head and neck cancer. Eur Arch Otorhinolaryngol. 2014;271:1861–7.
11. Sun Y, Wang Z, Qiu S, Wang R. Therapeutic strategies of different HPV status in Head and Neck Squamous Cell Carcinoma. Int J Biol Sci. 2021;17:1104–18.
12. Božinović K, Sabol I, Dediol E, Gašperov NM, Manojlović S, Vojtechova Z, et al. Genome-wide miRNA profiling reinforces the importance of miR-9 in human papillomavirus associated oral and oropharyngeal head and neck cancer. Sci Rep [Internet]. 2019 [cited 2019 Feb 21];9:2306. Available from: https://www.nature.com/articles/s41598-019-38797-z
13. Milutin Gašperov N, Sabol I, Božinović K, Dediol E, Mravak-Stipetić M, Licastro D, et al. DNA Methylome Distinguishes Head and Neck Cancer from Potentially Malignant Oral Lesions and Healthy Oral Mucosa. Int J Mol Sci. 2020;21.
14. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–9.
15. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47:W191–8.
16. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
17. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47:D590–5.
18. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018;46:D649–55.
19. Chou C-H, Shrestha S, Yang C-D, Chang N-W, Lin Y-L, Liao K-W, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. 2018;46:D296–302.
20. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006;34:D108-110.
21. Huang H-Y, Lin Y-C-D, Cui S, Huang Y, Tang Y, Xu J, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2022;50:D222–30.
22. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2016;44:D1023-1031.
23. Campanero MR, Armstrong MI, Flemington EK. CpG methylation as a mechanism for the regulation of E2F activity. Proceedings of the National Academy of Sciences [Internet]. 2000 [cited 2023 Nov 28];97:6481–6. Available from: https://www.pnas.org/doi/10.1073/pnas.100340697
24. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science [Internet]. 2017 [cited 2023 Nov 28];356:eaaj2239. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8009048/
25. Numoto M, Yokoro K, Koshi J. ZF5, which is a Kruppel-type transcriptional repressor, requires the zinc finger domain for self-association. Biochem Biophys Res Commun. 1999;256:573–8.
26. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, et al. The

5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 1983;11:6883–94.

27. Debernardi C, Libera L, Berrino E, Sahnane N, Chiaravalli AM, Laudi C, et al. Evaluation of global and intragenic hypomethylation in colorectal adenomas improves patient stratification and colorectal cancer risk prediction. Clin Epigenetics. 2021;13:154.
28. Moody L, Crowder SL, Fruge AD, Locher JL, Demark-Wahnefried W, Rogers LQ, et al. Epigenetic stratification of head and neck cancer survivors reveals differences in lycopene levels, alcohol consumption, and methylation of immune regulatory genes. Clin Epigenetics [Internet]. 2020 [cited 2023 Sep 12];12:138. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7488769/
29. Anić P, Golubić Talić J, Božinović K, Dediol E, Mravak-Stipetić M, Grce M, et al. Methylation of Immune Gene Promoters in Oral and Oropharyngeal Cancer. Int J Mol Sci. 2023;24:7698.
30. Farkas SA, Milutin-Gašperov N, Grce M, Nilsson TK. Genome-wide DNA methylation assay reveals novel candidate biomarker genes in cervical cancer. Epigenetics [Internet]. 2013 [cited 2018 Mar 27];8:1213–25. Available from: http://www.tandfonline.com/doi/abs/10.4161/epi.26346
31. Milutin Gašperov N, Farkas SA, Nilsson TK, Grce M. Epigenetic activation of immune genes in cervical cancer. Immunol Lett. 2014;162:256–7.
32. Eden S, Cedar H. Role of DNA methylation in the regulation of transcription. Current Opinion in Genetics & Development [Internet]. 1994 [cited 2023 Sep 12];4:255–9. Available from: https://www.sciencedirect.com/science/article/pii/S0959437X05800528
33. Pikaart MJ, Recillas-Targa F, Felsenfeld G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. Genes Dev [Internet]. 1998 [cited 2023 Sep 12];12:2852–62. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC317165/
34. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev [Internet]. 2002 [cited 2023 Sep 12];16:6–21. Available from: http://genesdev.cshlp.org/content/16/1/6
35. Goossens EAC, de Vries MR, Simons KH, Putter H, Quax PHA, Nossent AY. miRMap: Profiling 14q32 microRNA Expression and DNA Methylation Throughout the Human Vasculature. Front Cardiovasc Med [Internet]. 2019 [cited 2023 Sep 19];6:113. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6694280/
36. Aavik E, Lumivuori H, Leppänen O, Wirth T, Häkkinen S-K, Bräsen J-H, et al. Global DNA methylation analysis of human atherosclerotic plaques reveals extensive genomic hypomethylation and reactivation at imprinted locus 14q32 involving induction of a miRNA cluster. Eur Heart J. 2015;36:993–1000.
37. Slomp J, Gittenberger-de Groot AC, Koteliansky VE, Glukhova MA, Bogers AJ, Poelmann RE. Cytokeratin expression in human arteries pertinent to intimal thickening formation in the ductus arteriosus. Differentiation. 1997;61:305–11.
38. Tan H, Huang S, Zhang Z, Qian X, Sun P, Zhou X. Pan-cancer analysis on microRNA-associated gene activation. EBioMedicine. 2019;43:82–97.
39. Xu P, Wu Q, Yu J, Rao Y, Kou Z, Fang G, et al. A Systematic Way to Infer the Regulation Relations of miRNAs on Target Genes and Critical miRNAs in Cancers. Frontiers in Genetics [Internet]. 2020 [cited 2023 Nov 28];11. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2020.00278
40. Nakagawa T, Matsusaka K, Misawa K, Ota S, Takane K, Fukuyo M, et al. Frequent promoter hypermethylation associated with human papillomavirus infection in

pharyngeal cancer. Cancer Lett. 2017;407:21–31.
41. Tripathi N, Keshari S, Shahi P, Maurya P, Bhattacharjee A, Gupta K, et al. Human papillomavirus elevated genetic biomarker signature by statistical algorithm. J Cell Physiol. 2020;235:9922–32.
42. Slebos RJC, Yi Y, Ely K, Carter J, Evjen A, Zhang X, et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. Clin Cancer Res. 2006;12:701–9.
43. Camuzi D, Buexm LA, Lourenço S de QC, Esposti DD, Cuenin C, Lopes M de SA, et al. HPV Infection Leaves a DNA Methylation Signature in Oropharyngeal Cancer Affecting Both Coding Genes and Transposable Elements. Cancers (Basel). 2021;13:3621.
44. Lohavanichbutr P, Houck J, Fan W, Yueh B, Mendez E, Futran N, et al. Genomewide gene expression profiles of HPV-positive and HPV-negative oropharyngeal cancer: potential implications for treatment choices. Arch Otolaryngol Head Neck Surg. 2009;135:180–8.

Data availability statement

Code used is available at https://github.com/PKatarina/HNSCC_Grce

Raw datasets analyzed during the current study are available in the Figshare repository under the following DOI identfiers:

Transcriptome https://doi.org/10.6084/m9.figshare.24721746

miRNome https://doi.org/10.6084/m9.figshare.24721731

Methylome https://doi.org/10.6084/m9.figshare.24721728

All datasets and raw data will be deposited in ArrayExpress repository as well. Upon publication links will be replaced by accession numbers

Due to the file size restrictions Supplementary table S3 is deposited at Figshare repository https://doi.org/10.6084/m9.figshare.24722268

Ethics approval and consent to participate

All patients provided informed consent to participate at the time of primary cancer treatment. The collection of samples was approved by Bioethical Bord of the Ruđer Bošković Institute (BEP-3748/2-2014) and the Ethical Board of the Clinical Hospital Dubrava (EP- KBD-10.06.2014).

Funding

This study was funded by the Croatian Science Foundation grants HRZZ-IP-2013-11-4758 and UIP-2020-02-1623. KM is supported by Croatian Science Foundation doctoral student scholarship DOK-2021-02-7831

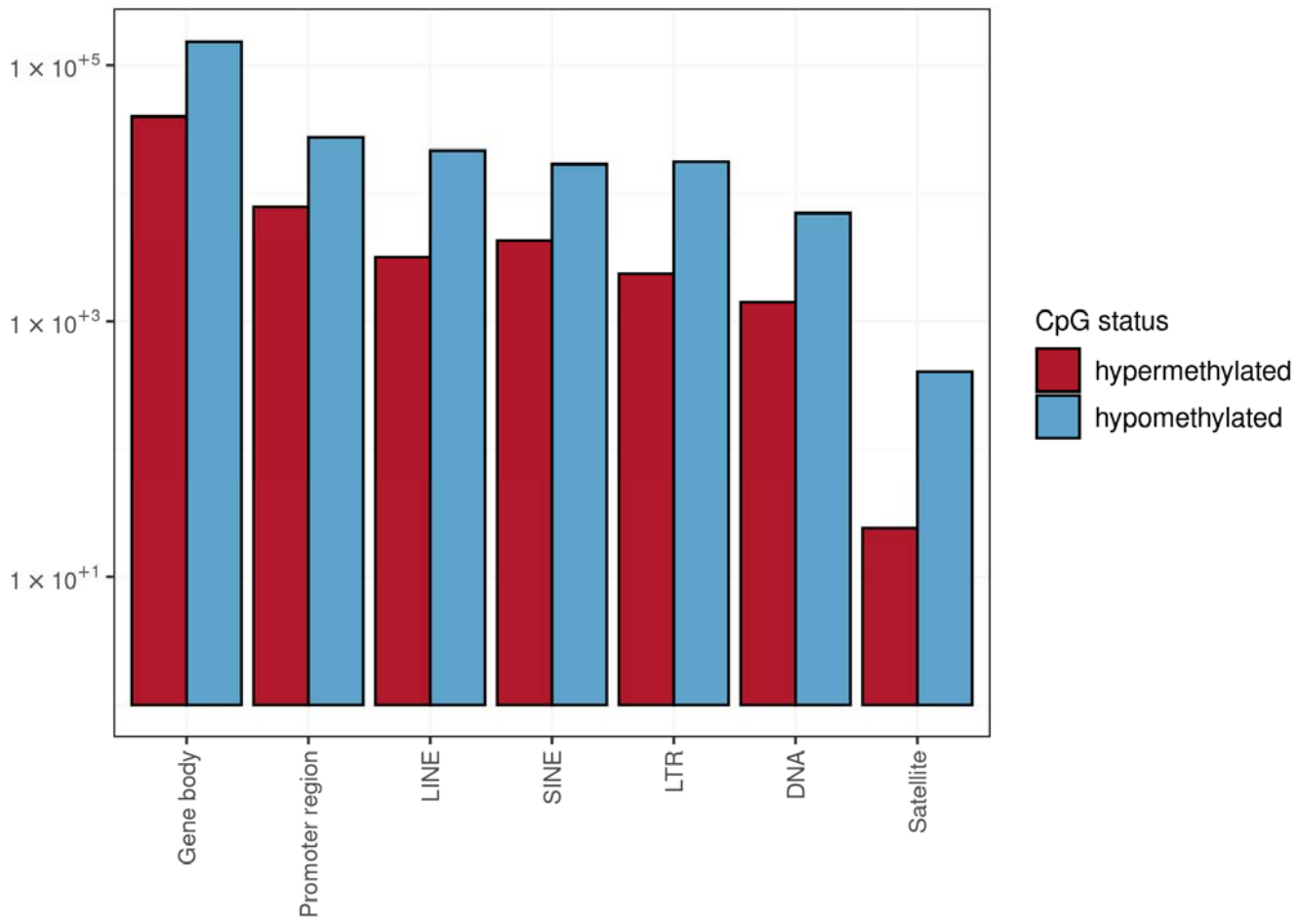
Authors' contributions

NMG, IS, MG and AB designed the study. KB, ED, JK, NS, IS and KM collected and processed the data. KM performed the analyses. KM, KB and NMG wrote the first draft of the manuscript. MG, IS and AB supervised and acquired funding for the study. All authors contributed to the writing of the final manuscript. All authors read and approved the final manuscript.

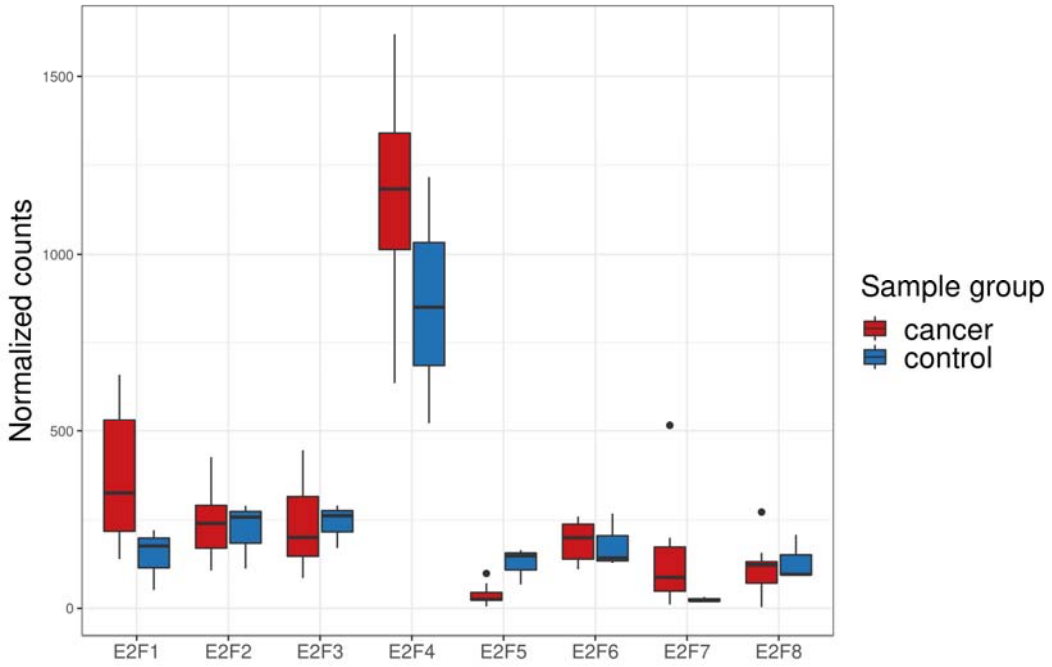
Supplemental figures

Integrative analysis in head and neck cancer reveals distinct role of miRNome and methylome as tumour epigenetic drivers

Katarina Mandić[1*], Nina Milutin Gašperov[2*], Ksenija Božinović[2], Emil Dediol[3], Jure Krasić[4], Nino Sinčić[4,5], Magdalena Grce[2], Anja Barešić[1**], Ivan Sabol[2**]



Suplemental figure 1. Significant number of hypermethylated and hypomethylated CpGs in cancer vs. control groups across all genomic elements (Chi squared p < 2x10^-16). Labels: Gene body – across exons and introns; Promoter region – from 500 to 2000 distance from canonical transcription start site; LINE – Long interspersed nuclear elements; SINE – Short interspersed nuclear elements; LTR – Long terminal repeat elements which include retroposons; DNA – DNA repeat elements; Satellite – Satellite repeats. Annotation for genomic elements were used from Jin et al., 2015.



Supplemental figure 2. Expression of E2F family genes in cancer vs control samples.