Oral Lichen Planus and its relation with Oral Squamous Cell Carcinoma: new insights into the potential for malignant transformation

Cristóvão Antunes de Lanna [1*]; Beatriz Nascimento Monteiro da Silva [2*]; Andreia Cristina de Melo [2]; Martín H. Bonamino [3]; Lísia Daltro Borges Alves [2]; Luis Felipe Ribeiro Pinto [4]; Abel Silveira Cardoso [5]; Héliton Spíndola Antunes [2]; Mariana Boroni [1,6]; Daniel Cohen Goldemberg [2, 7].

Affiliations

1 Laboratory of Bioinformatics and Computational Biology, Division of Experimental and Translational Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, 20231-050, Brazil; 2 Division of Clinical Research and Technological Development of the National Cancer Institute José Alencar Gomes da Silva (INCA). Rio de Janeiro (RJ), Brazil; 3 Immunology and Tumor Biology Program - Research Coordination, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil; Presidency of Research and Biological Collections (VPPCB), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil; 4 Research Department, National Cancer Institute, Rio de Janeiro, Rio de Janeiro, Brazil; 5 Emeritus Professor - Department of Oral Pathology and Oral Diagnosis, School of Dentistry, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; 6 Experimental Medicine Research Cluster (EMRC), University of Campinas (UNICAMP), Campinas 13083-970, Brazil; 7 Honorary Oral Medicine Senior Lecturer - University College London (UCL). London, UK. Latin American Cooperative Oncology Group (LACOG) - Head and Neck.

*These authors contributed equally to this work.

Corresponding authors:

Mariana Boroni - mariana.boroni@inca.gov.br

Abstract

Oral Lichen Planus (OLP) is a chronic inflammatory disorder of unknown etiology. However, evidence suggests that it consists of an immunological process that leads to degeneration of the keratinocytes in the basal layer of the oral mucosa. Despite being recognized by WHO as a potentially malignant disorder with progression to oral squamous cell carcinoma (OSCC), the relationship between both pathologies is still controversial. Different studies have investigated factors associated with the potential for malignant transformation of OLP but it remains unclear. Through a bioinformatics approach, we investigated similarities in gene expression profiles of OLP and OSCC in early and advanced stages. Our results revealed gene expression patterns related to processes of keratinization, keratinocyte differentiation, cell proliferation and immune response in common between OLP and early and advanced OSCC, with the cornified envelope formation and antigen processing cross-presentation pathways in common between OLP and early OSCC. Together, these results reveal that key genes such as *PI3*, *SPRR1B* and *KRT17*, in addition to genes associated with different immune processes such as *CXCL-13*, *HIF1A* and *IL1B* may be involved in this oncogenic process. In addition, we performed an analysis of differentially and co-expressed genes and proposed putative therapeutic targets and associated drugs.

Introduction

Oral Lichen Planus (OLP) is a chronic inflammatory disease clinically characterized by six distinct subtypes that can be seen individually or in combination: white reticular striations, papular, plaque-like, erythematous erosions, ulcerative, and bullous forms. Of all the presentations, the reticular form is the most common, exhibiting a delicate white banding, called Wickham's striae (Kurago, 2016;

Warnakulasuriya et al., 2007; Wickham, 1895). Histologically, OLP is characterized by vacuolar degeneration, a band-like dense inflammatory infiltrate of T lymphocytes at the epithelial-stromal junction, and hyperkeratosis or parakeratosis (Cheng et al., 2016).

The origins of this persistent cytotoxic T-cell-mediated damage are currently unknown. However, many authors suggest that the disease is associated with an autoimmune process (Farhi and Dupin, 2010; Ismail et al., 2007; Roopashree et al., 2010; Rutz et al., 2016). In 1910, Hallopeau and colleagues described for the first time a case of oral squamous cell carcinoma (OSCC), a malignant neoplasm that originates in the lining epithelium and is considered the most common malignancy in this region, in a patient with OLP. Since then, different studies have suggested a premalignant potential for OLP injuries over the years (Aghbari et al., 2017; Bardellini et al., 2013; Barnard et al., 1993; Hallopeau, 1910).

The World Health Organization (WHO) defined in 2017 that OLP is an *oral potentially malignant disorder* (OPMD), with a possible progression to OSCC (Müller, 2017; Peng et al., 2017; van der Waal, 2010). Giuliani and colleagues (2019), in a systematic review, demonstrated that 92 of 6559 patients diagnosed with OLP developed OSCC. The authors found a malignancy potential of 1.4% and an annual transformation rate of 0.2% when analyzing retrospective and prospective studies on OLP. Ruokonen et al. (2017) revealed that 17.9% of patients with oral cancer had OLP and 4% had oral lichenoid lesions, suggesting that this dysfunction may be an important etiological factor of OSCC and that smoking and alcohol use was less frequent in patients with OLP and lichenoid lesions, suggesting that OLP may progress to OSCC, even in the absence of well-known etiological agents.

In this study, we investigated gene expression signatures observed in OLP lesions when compared to healthy oral tissues that are also disturbed in the early and advanced stages of OSCC in order to shed light on the potential for malignant transformation of OLP. We demonstrate a repertoire of pathways that could be connected to the potential transition between OLP and OSCC. Our findings suggest that OLP and OSCC share the activation of keratinization- and inflammation-related pathways with a potential role in malignization. We have also explored potential novel therapeutic targets and propose drugs that can interact with them and revert expression alterations.

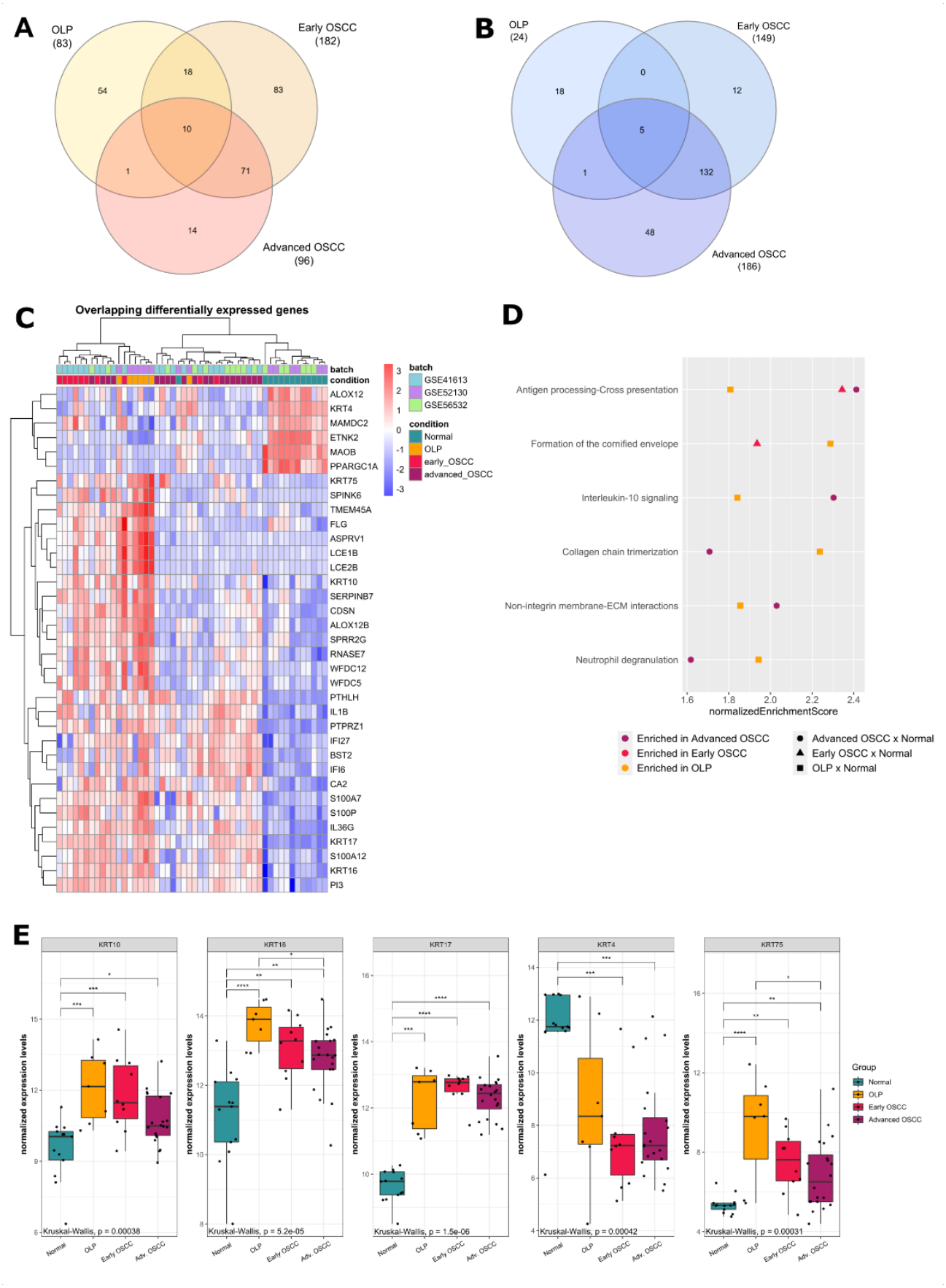
Results

Gene expression profiles among OLP and OSCCs

Gene expression data (16,656 features) comprising 13 normal oral mucosa, 7 OLP, 10 early-stage OSCC, and 20 advanced-stage OSCC samples were integrated and compared. Of note, although OLP samples tend to group closer to normal samples, at least one sample shows high similarity with OSCC samples (Fig. S1).

Differential expression analysis was performed for each group in comparison to normal tissue. A total of 107 DEGs were identified for OLP (83 overexpressed, 24 underexpressed), 331 for early-stage OSCC (182 overexpressed, 149 underexpressed), and 282 for advanced OSCC (96 overexpressed, 186 underexpressed) (Tables S1-S3; Fig. S2A-C). Gene set enrichment analysis (GSEA) was performed for each group. Enriched pathways in OLP were related to keratinization, extracellular matrix (ECM) and its interaction with surrounding tissue, and immunity-related pathways such as complement, interleukin 10 (IL-10) signaling, antimicrobial peptides, antigen presentation, among others (Table S5). In early and

advanced OSCC, enriched pathways included pathways related to keratinization, DNA replication, RNA transcription, DNA repair, cell cycle, multiple interleukin and interferon signaling pathways, among others (Tables S6-S7). To better explore OLP's potential for malignization and its relation with OSCC, the overlapping differentially expressed genes (DEGs) between OLP and OSCC were also investigated, resulting in a total of 35 overlapping DEGs (29 Up- and 6 Down-regulated). Fifteen genes were consistently differentially expressed in all comparisons (10 overexpressed, 5 underexpressed) (Fig. 1A-B; Table S4). Most of the overlapped genes (51.4%) occurred between OLP and early stage OSCC. When using a non-supervised clustering approach based on the expression of the 35 gene signature, all OLP samples were clustered with OSCC samples, mainly in the early stage (Fig. 1C). We also observed overlap among pathways enriched in the three conditions. The enrichment analysis revealed that OLP has six main pathways in common with early or advanced OSCC: antigen processing cross-presentation; formation of the cornified envelope; interleukin-10 signaling; collagen chain trimerization; non-integrin membrane-ECM interactions; and neutrophil degranulation, with the antigen presentation pathway enriched in all conditions. Interestingly, antigen presentation and formation of the cornified envelope were the only common pathways between early OSCC and OLP, with the latter being common only between these two groups (Fig. 1D). Some of the genes influencing the clusterization are those coding for keratins, including the downregulation of *KRT4* and up-regulation of *KRT16*, *KRT17*, *KRT10*, and *KRT75* in comparison to normal samples (Fig. 1E).



118

Figure 1. Differentially Expressed Genes between OLP, early, and advanced OSCC microarray datasets compared to normal oral tissue. (A) Venn diagram of overlapping DEGs of overexpressed genes. Ten of the differentially expressed genes were overexpressed in all three groups, while 18 overexpressed genes were shared by OLP and early OSCC. (B) Venn diagram of overlapping DEGs of underexpressed genes. Five of the differentially expressed genes were underexpressed in all three groups. (C) Clustered heatmap of DEGs shared between OLP, early, and advanced OSCC, as well as overexpressed genes shared by OLP and early OSCC. (D) Gene set enrichment analysis (GSEA) of Reactome pathways in Oral Lichen Planus (OLP), early, and advanced Oral Squamous Cell Carcinoma (OSCC). All pathways are upregulated in comparison with normal tissue. For each pathway, orange squares, red triangles, and purple circles correspond to the normalized Enrichment Score for said pathway in OLP, early OSCC, and advanced OSCC, respectively. Only pathways enriched simultaneously in OLP and at least one of the OSCC stages are represented. (E) Differentially expressed keratin genes between OLP and both OSCC groups compared to normal tissue. Box plots represent the normalized expression distribution in each group. Comparisons among groups were made using the Kruskal-Wallis test followed by Dunn's post-hoc test, with p-values lower than 0.05 considered significant for both tests. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Of note, genes coding for keratinocyte differentiation-related members of the S100 protein family such as *S100A7*, *S100P*, and *S100A12* as well as immunity-related genes such as *IL1B*, *IL36G*, *IFI6*, and *IFI27*, all of them overexpressed in OLP samples compared to normal tissue, were also identified. To validate our findings, independent datasets corresponding to each of the tested groups were used. Similar results were found to the expression of *KRT4* (lower expression) and *KRT75* (higher

144 expression), whereas differences found in *KRT10*, *KRT16* and *KRT17 expression*, all
145 of them higher than normal in our analysis, have not been validated (Fig. S3A and
146 S3B). Data obtained from TCGA for early and advanced OSCC showed lower *KRT4*
147 expression and higher *KRT17* expression when compared with normal samples, also
148 consistent with our analysis (Fig S3C).

149 Overrepresentation analysis (ORA) was performed for all overlapping DEGs
150 between OLP and at least one OSCC stage (29 up and 6 down-regulated, Fig 1A and
151 1B). Pathways enriched in overlapping up-regulated genes included keratinization and
152 formation of the cornified envelope (Fig. S4A). These pathways were also among
153 those enriched in OLP-exclusive overexpressed DEGs, which also included immunity-
154 related pathways, such as interferon alpha and beta and IL-1 family signaling (Fig.
155 S4B). In OLP-exclusive underexpressed DEGs, the keratinization and formation of the
156 cornified envelope pathways were also enriched, along with cell junction- and
157 compound metabolization-related pathways (Fig. S4C). When considering
158 overlapping underexpressed DEGs, no pathway was significantly enriched.

159 ## Immune Microenvironment Composition of OLP and OSCC

160 Given that OLP is a disease with a significant participation of the immune system
161 and both immune evasion and tumor-promoting inflammation are hallmarks of cancer,
162 we have sought to identify the most prominent proportions of infiltrating cells in both
163 conditions. We have used a method that estimates the abundances of cell populations
164 by deconvolution of gene expression data.

165 When clustering samples based on cell's population composition, OLP samples
166 mostly clustered with normal samples, although some samples showed high similarity

with the immune cells composition found in OSCC samples, enriched in plasma cells and memory activated CD4+ T cells (Fig. 2A).

The proportions of activated NK cells in OLP samples were significantly lower than the proportions found in the normal oral mucosa. The same was observed for activated NK cells in early or advanced OSCC. The proportions of CD8+ T lymphocytes, M0 and M1 macrophages in advanced OSCC showed significantly higher values when compared to those observed in OLP. The opposite was observed in resting Mast cells, naïve B cells and monocytes. Reduced proportions for all three populations were observed in both early and advanced OSCC when compared to OLP and healthy oral mucosa. The proportions of activated memory CD4+ T cells, resting NK cells, and M0 macrophages were higher in OSCC than in healthy oral mucosa, regardless of staging (Fig. 2B). Considering the immune infiltrate populations in both validation datasets, significantly reduced values of activated NK cells were also observed in OLP, corroborating our results for this group. Similarly, monocyte proportions were consistent with our analysis for early and advanced OSCC in the microarray and TCGA datasets. The resting mast cell proportions were also consistent with those observed in the analysis, while macrophages M0 showed elevated proportions in advanced OSCC in the microarray dataset and early and advanced OSCC in the TCGA dataset, consistently to the proportions in the discovery dataset (Fig. S5).

Since genes that play an important role in the T helper (Th) 17 cell phenotype such as *IL1B, CCL5* and *CXCL13* were differentially expressed in OLP, early and advanced OSCC (Tables S1-S3), we decided to investigate the genes involved in its differentiation given the role of Th17 cells in maintaining mucosal immunity homeostasis. Samples were clustered using a 33-gene panel built based on

pathogenic and non-pathogenic Th17 phenotypes characterized in the literature (Lee et al., 2012) (Fig. 2C). When clustering samples with regard to pathogenic and non-pathogenic Th17 signature, two groups can be highlighted: one mostly composed of OSCC samples, showing high expression of many genes from the pathogenic signature and another composed mainly of OLP and normal samples. In this group, genes from the non-pathogenic signature are highly expressed (Fig. 2C). Of note, some genes important in the pathogenic signaling were also found differentially modulated in OLP samples, such as *CTSC*, *HIF1A*, *IL1B*, *LTA*, *LTB*, and *TGFB3* in relation to normal oral mucosa. Additionally, *CTSC*, *HIF1A*, and *IL1B* also show higher expression levels in early and late-stage OSCC, with an increasing pattern (Fig. S6A). The validation data also revealed significant high levels of expression of the *CTSC, LTA*, and *TGFB3* genes in OLP (Fig. S6B). Regarding the OSCC groups, however, only *CTSC* and *LTB* exhibited similar expression patterns in comparison to the discovery dataset (Fig S6C-D). *LTA*, while significantly altered compared to normal mucosa in the validation datasets, showed different patterns depending on the validation dataset, with lower expression in the microarray dataset (Fig. S6C) and higher expression in the TCGA validation dataset (Fig. S6D).

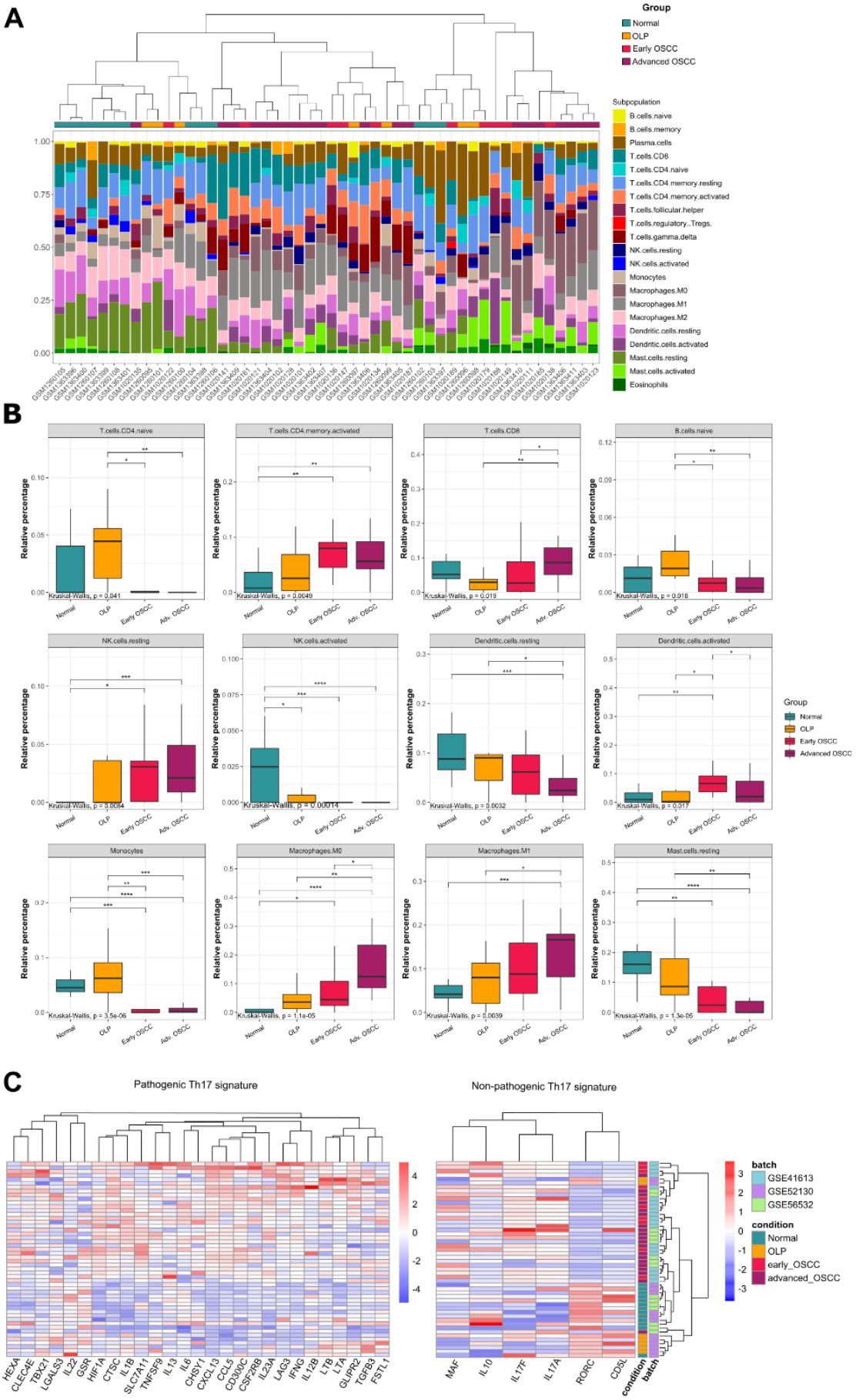


Figure 2. Proportion of cell population signatures identified using CIBERSORTx. (A) Stacked barplot showing cell population proportions in each sample (Oral normal mucosa, OLP, early OSCC, advanced OSCC) based on CIBERSORTx's 22 cell gene expression signatures. Dendrogram represents Ward clustering of the samples. (B) Relative percentage of each immune cell from oral normal mucosa, OLP, early and advanced OSCC. Comparisons among groups were made using the Kruskal-Wallis test followed by Dunn's post-hoc test, with p-values lower than 0.05 considered significant for both tests. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. (C) Clustered heatmap of Th17-related genes. Genes were separated based on non-pathogenic and pathogenic Th17 phenotype-related expression profiles.

Co-expression analysis

In addition to differential expression analysis, we have also constructed co-expression modules using WGCNA to investigate the connection strength between genes with similar expression patterns and identify potentially co-regulated genes associated with the potential malignization process in OLP. By analysing a total of 15,000 genes (highest gene expression variability among samples), 12 co-expression modules were obtained, each identified by a color: magenta, with 311 genes; purple, with 262 genes; blue, with 5059 genes; greenyellow, with 174 genes; tan, with 142 genes; black, with 381 genes; pink, with 370 genes; red, with 525 genes; turquoise, with 5364 genes; yellow, with 628 genes; brown, with 1070 genes; and green, with 624 genes. A 13^{th} module, to which 90 genes with no co-expression patterns were assigned, was also identified and assigned to the color grey (Fig. 3A).

To better understand the relationship between OLP and OSCC, we have investigated which co-expression modules had a similar correlation to both conditions

simultaneously. For each module, the Pearson correlation coefficient of the module eigengene to the sample groups was calculated. None of the identified modules had a simultaneous significant correlation with OLP and early and late stage OSCC. Interestingly, only the magenta module showed a significant and positive correlation with OLP (Pearson's correlation $r = 0.46$; $p < 0.0001$) and early OSCC ($r = 0.33$; $p = 0.02$). Co-expressed genes belonging to this module were mainly associated with keratinization and formation of the cornified envelope (Fig. 3B).

Network analysis

Co-expression modules, while grouping correlated genes, offer only a glimpse of their dynamic in the cells. To understand how genes in the magenta module interacted, we added another layer of information, searching for protein-protein interaction (PPI) data in the STRING database and building a network (Fig. 3C).

By characterizing genes from the network according to their characteristics and connectivity, we identified 11 hubs, 15 transcription factors, 6 clinically actionable genes and 112 members of gene families in the druggable genome, with some of those genes classified in more than one category (Table S8). Additionally, drug-gene interactions were identified using all genes in the module. A total of 89 drugs were identified, which interacted with 3 of the 11 hubs (Table S9).

The hubs' expression levels were compared among conditions, with *PI3* being the only gene significantly up-regulated in OLP and all OSCC stages compared to the normal mucosa. Additionally, *FLG*, *SPRR1B*, and *SPRR2G* were significantly up-regulated in OLP and early OSCC, while *DSP* and *JUP* showed higher expression levels only in early OSCC and *RPTN* was up-regulated only in OLP. The four

259 remaining hubs (*DSC1*, *DSG1*, *IVL*, and *PKP1*) didn't exhibit significant differences in
260 expression compared to the normal mucosa (Fig. 3D).

261 The hubs with known drug-gene interactions are *PI3* (up-regulated in OLP and
262 OSCC), *IVL* (up-regulated in OLP and early OSCC), and *DSP* (up-regulated in OLP
263 and OSCC) (Fig. 3 C-D, Table S9). This result may direct future selections of drug
264 targets that could lead to efficient treatment for both diseases simultaneously.

265

266

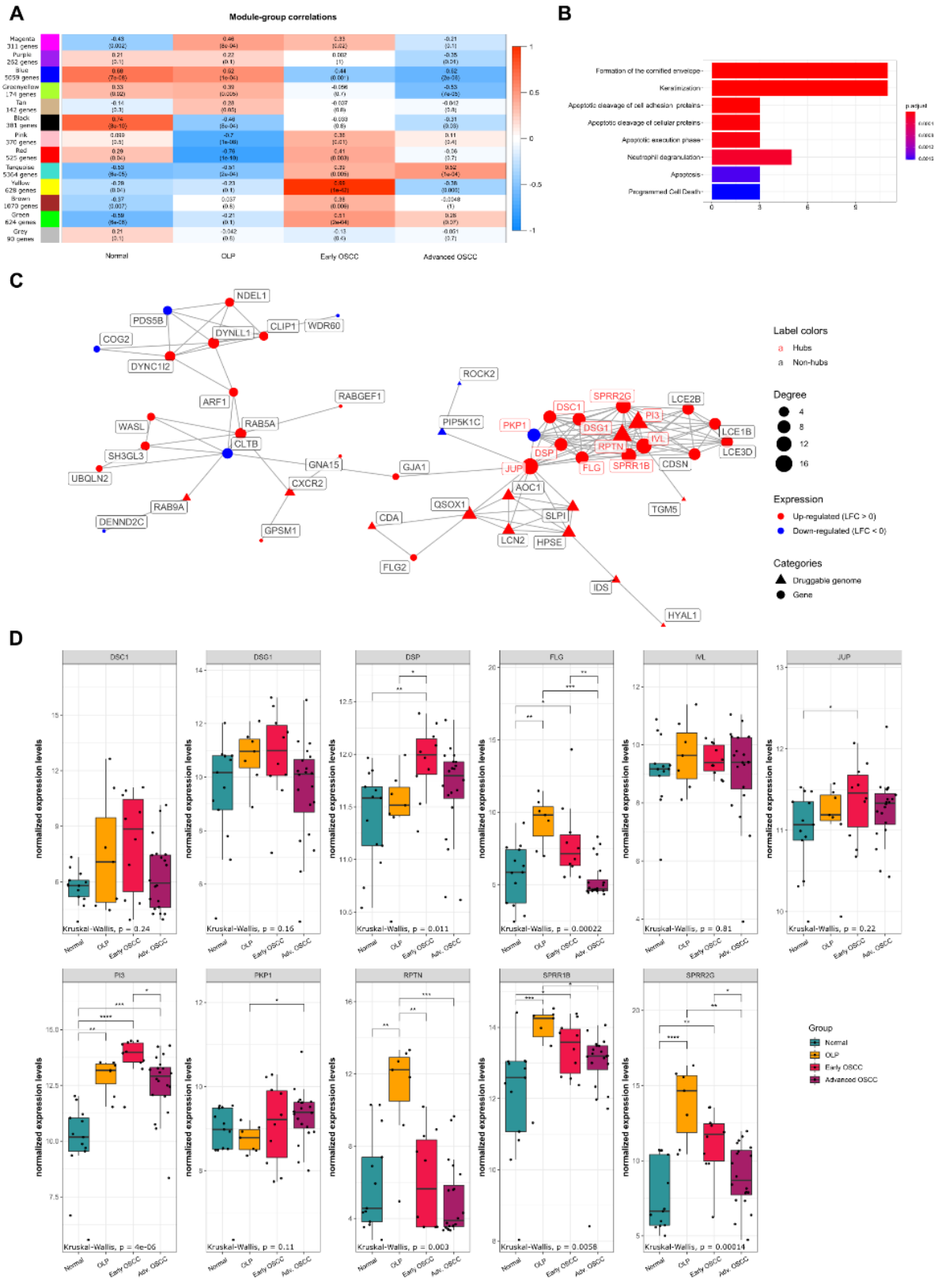


Figure 3. Analysis of co-expressed genes and their regulation. (A) Correlations between the co-expression modules and each of the sample groups (normal tissue,

OLP, early OSCC, and advanced OSCC). Cell colors correspond to Pearson correlation values between each co-expression module (rows) with groups (columns), from blue (100% inverse correlation) to red (100% direct correlation). Numbers in each cell refer to Pearson's correlation r values, with p-values represented below in parentheses. Correlations with $p < 0.05$ were considered statistically significant. (B) Pathway enrichment of genes in the magenta module using ORA. All represented pathways are significantly enriched (FDR > 0.05). Bar length and colors indicate gene counts and BH-adjusted p-values, respectively. (C) Representations of the magenta module as a PPI network (largest connected component). Node size, shape, and color represent degree, category, and LFC in OLP, respectively. Red node labels indicate hubs. (D) Hub genes from the magenta module compared between OLP, both OSCC groups, and normal tissue. Box plots represent the normalized expression distribution in each group. Comparisons among groups were made using the Kruskal-Wallis test followed by Dunn's post-hoc test, with p-values lower than 0.05 considered significant for both tests. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Expression drug-response analysis

Considering the possibility that the 35 DEGs in common between OLP, early and advanced OSCC are involved in the potential for malignant transformation, we investigated pharmacological agents that would best reverse the signature of these genes. A clustergram of the main drugs was obtained according to L1000 CDS2 output, which demonstrates the expected positive or negative regulation of each drug in relation to each DEG by comparing them to the LINCS L1000 small molecule-related expression profiles (Fig. 4). The top fifty matched signatures, corresponding to 42 drugs, were identified based on perturbation data for nineteen of the overlapping

DEGs. Six (14.3%) of these drugs are already in use in the clinic or tested in clinical trials, as indicated in Table S10. Among the predicted drugs the most represented class was that of the PI3K/mTOR pathway inhibitors, which includes INK-128, GSK 1059615, GDC-0980, Torin-2, KU 0060648 Trihydrochloride, AZD-8055, and PI103 Hydrochloride. The signature of some genes has been reversed by a large number of drugs in the LINCS L1000 cells' signatures such as *PI3* (32/42 drugs), *KRT17* (30/42), *KRT10* (29/42), *S100A7* (29/42), and *S100P* (23/42).

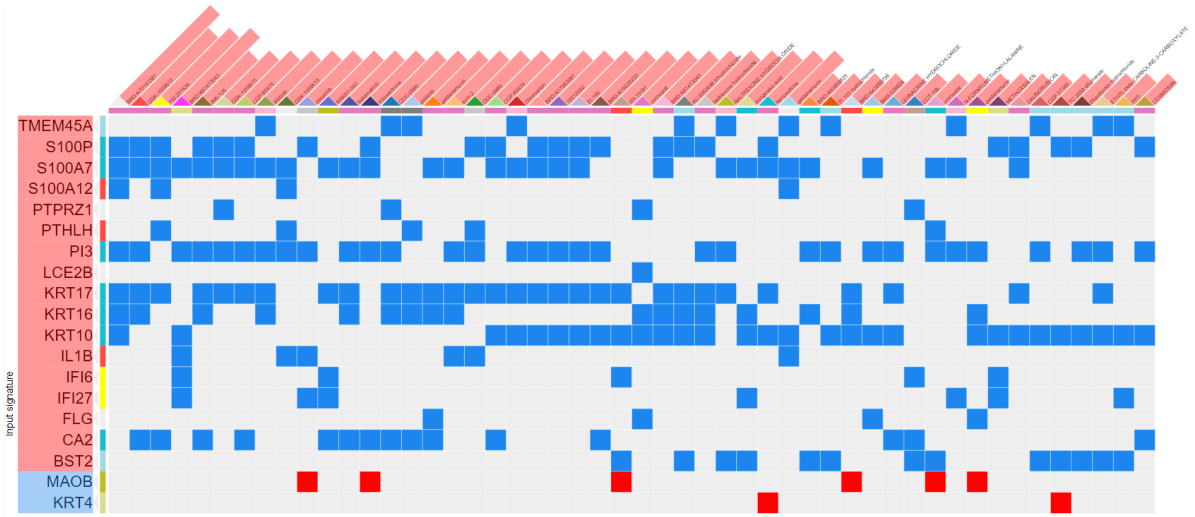


Figure 4. Potential drug target search from shared DEGs using the L1000 CDS^2 tool on the LINCS Program platform. Genes are represented in rows and drugs in columns, with overexpressed genes colored red and underexpressed genes in blue. Colored squares represent drug-gene interactions, with blue cells representing inhibition and red cells representing gene activation.

Drug repositioning evaluation

All identified drugs in this work were evaluated for their repositioning potential for use in OLP and OSCC. Of 89 drugs identified based on the network's hub genes, 70

were tested in cancer and 67 were approved in their respective clinical trials. Of these, only 2 were already tested in OSCC and none was approved for use in this type of cancer, Cisplatin and Sunitinib (Table S9). Considering the drugs identified based on the DEGs using the L1000 CDS2 tool, 6 of the 42 drugs were tested and approved for use in cancer. None of them were tested in OSCC. Additionally, none of the drugs identified in both analyses have been tested in OLP (Table S10).

Discussion

According to the World Health Organization, OLP is categorized as a potentially malignant disorder, with the possibility of progression to OSCC. Despite its well-known status, the relationship between OLP and OSCC remains controversial (Gonzalez-Moles et al., 2008; Peng et al., 2017; Shen et al., 2011). Also, few studies have compared OLP with OSCC using high-throughput data (Giacomelli et al., 2009; J. Liu et al., 2020) and, to the best of our knowledge, no studies have simultaneously analyzed OLP and OSCC progression using high-throughput molecular data analysis. Based on this lack of information, we have sought to uncover which mechanisms are involved in this transformation. OLP, early OSCC, and advanced OSCC mRNA microarray data were analyzed to investigate similarities between gene expression profiles from the three conditions and correlating them to OLP malignant transformation potential.

Differential expression analysis revealed a subset of genes that are consistently modulated in OLP and OSCC, suggesting that there are similar pathways activated in the two conditions. Most of these genes modulated in both OLP and OSCC are related to keratins and keratinocyte differentiation. Shimada et al. (2018) demonstrated that cornified envelope formation proteins would be up-regulated in OLP, contributing to a

hyperkeratosis state commonly observed in this pathology. In accordance with this, OLP presented the enriched and up-regulated cornified envelope formation pathway that was also found in early OSCC, suggesting that changes in proteins belonging to this pathway may be involved in the malignization processes leading to progression from OLP to OSCC. Our co-expression and network analyses reinforce this hypothesis by revealing that genes belonging to the cornified envelope formation and keratinization pathways were consistently co-regulated in OLP and early OSCC. Among the genes belonging to these pathways, *PI3* (peptidase inhibitor 3) stands out, which was also significantly upregulated in the three conditions. Kengkarn et al. (2020), through microarray analysis, identified that *PI3*, along with *KRT17*, were upregulated in 100% of OSCC samples (39 cases), suggesting that these genes can be used as molecular biomarkers for patients with OSCC. This gene was also identified as belonging to the druggable genome, a set of genes that code for proteins belonging to families suitable for drug development, which opens possibilities for developing compounds that interact with its protein (Hopkins and Groom, 2002; Russ and Lampel, 2005). In addition, it is interesting to note that the *SPRR1B*, *SPRR2G* and *FLG* genes were significantly upregulated in OLP and early OSCC, but not in advanced OSCC, when compared to healthy oral mucosa samples. The *SPRR* (Small PRoline Rich) gene family encodes cornified envelope precursor proteins and are closely related to keratinocyte differentiation (Patel et al., 2003). The role of *SPRR2G* in OLP and OSCC has not yet been investigated, however, it has been seen that squamous cell carcinoma of the vulva overexpresses this gene (Micci et al., 2013). As for *SPRR1B*, it was shown that its expression is up-regulated in OSCC-derived stem cells, in addition to having a role in the growth and proliferation of these cells by regulating *RASSF4*, a tumor suppressor gene related with the MAPK pathway,

suggesting that the overexpression of *SPRR1B* may be related to the carcinogenesis of OSCC, as well as the maintenance of stem cells of this carcinoma (Michifuri et al., 2013). Filaggrin, coded by the *FLG* gene, is a protein located in the stratum corneum of the skin, contributing to its integrity and strength (McGrath and Uitto, 2008), and patients with OLP have an altered distribution and overexpression of this protein in the oral mucosa (Larsen et al., 2017). Though filaggrin may be involved in lesions in the oral mucosa (Itoiz et al., 1985), further investigations are needed to show the role of this protein and the malignancy potential of OLP.

KRT17 is overexpressed in OSCC and it may be associated with tumor progression by stimulating multiple signaling pathways (Kitamura et al., 2012; Ohkura et al., 2005). Shen et al. (2006) suggested that Keratin 17 may serve as immunodominant T cell epitopes by stimulating peripheral blood lymphocytes in psoriasis. Liu et al. (2020) demonstrated that high *KRT17* expression levels and tumor differentiation stage were significantly associated with overall survival in 64 patients with esophageal squamous cell carcinoma (ESCC), suggesting that *KRT17* may be a tumor-promoting factor and that an increase in the expression of this keratin may contribute to the malignant progression of the carcinoma. Furthermore, the authors demonstrated that *KRT17* plays a role in proliferation, migration, growth, and metastasis of ESCC cells *in vitro* and *in vivo*. Similar results were observed by Wang et al. (2019) in non-small cell lung cancer (NSCLC). The authors noted that high *KRT17* expression levels correlated with poor prognosis in NSCLC, especially in lung adenocarcinoma. Although not significantly differentially expressed in OLP samples from our validating dataset, which may be related to the low number of samples available, the upregulation of *KRT17* in OLP may also play an important role associated with the malignization process. Interestingly, *KRT17* is one of the main

386 therapeutic targets found in our analyses. Based on these results, it is suggested that
387 *KRT17*, as well as the related pathways, are important for studying the pathology and
388 malignancy potential of OLP to OSCC, since this gene seems to be involved in the
389 transition between both diseases. Besides, our data reveal the elevated and significant
390 expression of *KRT10* in contrast with low levels of *KRT4* in both conditions when
391 compared to the normal oral mucosa. Sakamoto et al. (2011) demonstrated that *KRT4*
392 down-regulation in oral squamous cell carcinoma is associated with changes in the
393 morphology of the epithelium and could be associated with the overexpression of other
394 keratins such as *KRT17*. Therefore, the authors suggest that *KRT4* may serve as a
395 diagnostic biomarker for OSCC. The same was suggested by Schaaij-Visser et al.
396 (2009) when revealing that the low expression of *KRT4* in samples from patients with
397 head and neck squamous cell carcinomas (HNSCC), including OSCC samples, may
398 serve as a screening biomarker for local recurrence risk and allow selection for
399 adjuvant treatment or tertiary prevention studies. Similar results were observed by Liao
400 et al. (2012) in OLP lesions, which generally affect the non-masticatory mucosa (such
401 as the bilateral buccal mucosa), with a shift in keratin expression observed by an
402 increased expression of *KRT10* and reduced expression of *KRT4*. These data were in
403 agreement with our analyses on the reduction of *KRT4* in OLP. However, further
404 investigation is needed regarding *KRT10*, because although the validation analyses
405 suggest an increase in its expression, the difference in expression is not significant, in
406 disagreement with what was observed in our analysis.

407 Our analysis showed that antigen presentation is the only pathway enriched
408 and upregulated in the three evaluated conditions. Antigen cross-presentation seems
409 to be involved in the early processes of OLP pathogenesis.

However, the nature of the antigen responsible for triggering the immune response in OLP has not yet been unveiled. When analyzing the genes present in the antigens cross-presentation pathway in OLP and early and advanced OSCC, we observed genes associated with different tumor responses such as genes belonging to HLA (*HLA-E*; *HLA-F* and *HLA- G)*, among which *HLA-G* which has been identified as an immune evasion-related gene in different tumors through the inhibition of effector cells such as NK, T cells, monocytes, and dendritic cells (Krijgsman et al., 2020). This gene was associated with OSCC prognosis and indicated as a new therapeutic target (Shen et al., 2018). In addition, most of the genes associated with antigen cross-presentation belong to the proteasome subunits, which have been associated with worse prognosis, tumor progression, and metastasis in different tumors (Ding et al., 2020; Kakumu et al., 2017; Munkácsy et al., 2010; Tan et al., 2018). Interestingly, among these *PSMB10*, one of the subunits of the immunoproteasome, was found differentially expressed in the three conditions. Immunoproteasome has been the subject of several studies due to its role in the differentiation of T cells, cytokine regulation, and tumor progression (Chen et al., 2020; Kiuchi et al., 2021; Zerfas et al., 2020). In addition, it is known that the peptides generated by the immunoproteasome for MHC class I are capable of generating a more efficient and accentuated cytotoxic lymphocyte response than those generated by the constitutive proteasome, contributing to CD8+ T cell infiltration (Groettrup et al., 2001; Kloetzel, 2001). We hypothesise that antigen presentation mediated by the immunoproteasome may contribute to the immune infiltrate profile present in OLP and OSCC. Further investigations, however, are needed on the interaction between antigen presentation and cytokines participating in the pathogenesis of OLP and OSCC, mainly on their role in the potential for malignant OLP transformation. Although

previous studies have shown that the immune infiltrate profile in OLP is predominantly composed of CD8+ and CD4+ T cells (Iijima et al., 2003; Wang et al., 2016), our analyses through gene signature deconvolution have demonstrated a significantly reduced proportion of CD8+ T lymphocytes and NK cells in those samples. Still on the predominant microenvironment components in OLP, we demonstrate an up-regulated signature of the chemokine CXCL-13, which has a dual role in tumorigenesis (Kazanietz et al., 2019).

The participation of Th17 cells in OLP and OSCC pathogenesis has also been recently investigated. Th17 is related to the maintenance of chronic inflammation in many conditions, notably autoimmune diseases (Awasthi and Kuchroo, 2009). Pathway enrichment and gene expression analyses performed in this study demonstrated that pathogenic Th17-related pathways were positively correlated to OSCC, which is corroborated by findings by Gaur et al. (2012), that also demonstrated an increase in Th17 cell prevalence in OSCC patients' peripheral blood when compared to a healthy control group. In contrast, it is observed that in the heatmap of Th17 phenotype-related genes, the general expression profile of most OLP samples is closely related to the sample profile of the normal mucosa. However, two samples clustered with early and advanced OSCC samples. This result suggests that different samples of OLP may have different expression profiles, which may be related to lower or higher risk of malignant transformation. Also, individual investigations of gene expressions related to the pathogenic Th17 molecular signature in OLP showed significantly high levels when compared to samples of normal oral mucosa for several genes such as *TGFB3*, *IL1B*, *HIF1A*, *LTA*, and *LTB*. The role of HIF-1α in the potential for malignant transformation of OLP has been the subject of investigation. Wang et al. (2017) showed that *HIF1A* was upregulated in OLP and OSCC samples, contributing

to changes in the expression of genes involved in adaptation to hypoxia and tumor progression. Additionally, Yang et al. (2020) demonstrated that the activation of *HIF1A*, by the accumulation of succinate, plays a fundamental role during the malignant transformation of OLP by stimulating the apoptosis of keratinocytes. Besides, as *HIF1A* is also related to increased transcription of *IL1B* by cells of the immune system (Corcoran and O'Neill, 2016; Ge et al., 2019), it is suggested that the increased expression of both may be correlated. Together, these results suggest that OLP may have elements of a tumor-like microenvironment as proposed by Peng et al. (2017).

Interestingly, in our drug signature analysis, IL-1β was suggested as a target for treatment with the drugs such as GSK-1059615, Torin-2, and GDC-0980, which are PI3K/mTOR inhibitors (Leontieva and Blagosklonny, 2016). Only GSK-1059615 has already been investigated in OSCC, being able to reduce the proliferation of OSCC cell lines (Yang et al., 2020). In OLP lesions, Ma et al. (2019) showed the overexpression of phosphorylated IGF1R and TRB3, which are related to the PI3k/AKT/mTOR signaling pathway, suggesting that this pathway mediates the relationship between T cells and keratinocytes, and influences the imbalanced cytokine networks on the immune microenvironment. According to the data presented in this study, four other mTOR pathway inhibitors were identified as candidate drugs (PI-103; INK-128; KU-0060648 and AZD-8055). Among them, PI-103 and INK-128 treatment demonstrated inhibition of cell growth and proliferation of OSCC (Aggarwal et al., 2019; Liang et al., 2019). Also, AZD-8055, an inhibitor of both mTORC1 and mTORC2, was able to induce autophagy in HNSCC cells (Li et al., 2013). Although corticosteroids are recommended for the treatment of OLP, in this work we

484 demonstrated different pharmacological agents that could assist in the treatment and
485 possibly interfere with the malignancy potential of such lesions.

486 Although our analyses have uncovered some of the genes and pathways
487 potentially involved in the transformation from OLP to OSCC, the limited availability of
488 public OLP data made it difficult to acquire gene expression data. However, even
489 though the low number of available OLP data led us to perform the analysis using
490 batch-corrected datasets from different platforms, we were able to discover important
491 biological similarities between the conditions which may point us to better understand
492 the malignization process.

493 In conclusion, *in silico* analysis revealed that OLP is a pathology that has a
494 proximity to the gene expression profile of OSCC, mainly with early OSCC. This result
495 is compatible with the fact that OLP is a differential diagnosis of epithelial precursor
496 lesions, namely leukoplakia and erythroleukoplakia, which can give rise to initial OSCC
497 if not removed surgically. We clearly reveal signatures in common with the two
498 conditions that can be important targets for drug treatment, as well as in the
499 development of diagnostic and prognostic strategies for the disease. It is considered
500 that OLP and OSCC have multifactorial etiology, and the intersections between the
501 keratinization and differentiation of lymphocytes are interesting potential targets for
502 further investigation.

503 ## Materials and methods

504 ### Datasets

505 Gene expression data from mRNA microarray experiments were obtained from
506 NCBI's Gene Expression Omnibus (GEO) (Barrett et al., 2013) using the GEOquery
507 R package (Davis and Meltzer, 2007). The datasets used corresponded to accession

numbers GSE52130, GSE56532, and GSE41613. From the dataset GSE52130 only oral samples were used, consisting of 7 OLP samples and 7 normal oral tissue samples, with expression values measured using the Illumina HumanHT-12 V4.0 expression BeadChip array platform. GSE56532 consisted of gene expression from 10 advanced OSCC samples and 6 normal oral mucosa samples, measured using the Affymetrix Human Gene 1.0 ST Array platform. GSE41613 is a dataset containing gene expression data of 97 HPV-negative samples from OSCC at varied stages, measured using the Affymetrix Human Genome U133 Plus 2.0 Array platform. Since this dataset contained a much larger sample size compared to the others, we have randomly selected a subset of 20 samples, ten of them at stages I and II (samples GSM1020161, GSM1020136, GSM1020149, GSM1020147, GSM1020122, GSM1020134, GSM1020188, GSM1020189, GSM1020138, and GSM1020179), and ten of them at stages III and IV (samples GSM1020128, GSM1020141, GSM1020185, GSM1020101, GSM1020121, GSM1020135, GSM1020123, GSM1020111, GSM1020102, and GSM1020187), to keep group sizes similar and avoid disproportionately large groups to skew subsequent analyses.

Results were validated using independent microarray expression datasets from GEO as well as RNA-seq expression data from The Cancer Genome Atlas (TCGA). Validation microarray expression datasets from GEO correspond to accession numbers GSE38616 and GSE3524. GSE38616 consisted of 7 OLP and 7 normal oral mucosa, with expression measured on the Affymetrix Human Gene 1.0 ST Array platform. GSE3524 was composed of 16 OSCC samples in stages II and IV and 4 normal tissue samples, with staging information missing for 2 of the tumor samples, which were removed from the analysis. mRNA expression for this dataset was measured on the Affymetrix Human Genome U133A Array platform. Outliers were

identified in each dataset using PCA and removed. GSE38616 had three outliers, GSM946266 (OLP), GSM946263 (OLP), and GSM946254 (Normal); while GSE3524 had one outlier, GSM80467 (advanced OSCC). For TCGA samples, expression data were downloaded in the form of raw counts from the Head and Neck Squamous cell Cancer (TCGA-HNSC) project using the TCGAbiolinks R package (Colaprico et al., 2016). Primary tumor and normal samples belonging to the "Other and unspecified parts of tongue", "Base of tongue", "Lip", "Palate", "Gum", "Floor of mouth", "Other and unspecified parts of mouth", and "Oropharynx" sites were considered OSCC samples and used in this step.

Data integration and cross-platform normalization

Data integration and cross-platform normalization were performed in the discovery datasets according to the methods described in (Binato et al., 2018) and (Walsh et al., 2015) Files containing probe-level intensity data (CEL files for Affymetrix arrays and txt files for Illumina BeadChip arrays) were downloaded using GEOquery. Raw data files were preprocessed using the appropriate package for each platform (*oligo* for Affymetrix and *beadarray* for Illumina). Probe level data was extracted, background corrected, and normalized, with the robust multi-array average (RMA) method used for Affymetrix datasets and neqc for the Illumina BeadChip dataset. Probes were mapped to genes using each platform's annotation, with the resulting matrix containing 16,656 features. When multiple probes corresponded to the same gene, normalized expression was aggregated to gene mean values. Batch effects were corrected using the ComBat method (Johnson et al., 2007) implemented in the sva R package (Leek et al., 2012), considering dataset of origin and sample type as

556 variables. Samples were grouped using principal component analysis (PCA) to
557 validate that the batch effect correction was successful.

558 Differential expression analysis

559 Differential expression analysis was conducted using the limma R package
560 (Ritchie et al., 2015). Gene expression from OLP, early OSCC (stages I and II), and
561 advanced OSCC (stages III and IV) were compared to normal samples. Differentially
562 expressed genes (DEGs) were identified based on the following cutoffs: Benjamini-
563 Hochberg (BH)-adjusted $p < 0.05$ and $\log_2$ fold change (LFC) > 2. DEGs from each
564 comparison were overlapped using the InteractiVenn online tool (Heberle et al., 2015).
565 Heat maps were constructed using the pheatmap R package (Kolde, 2019), using
566 normalized expression values z-scored across samples. Rows representing genes
567 were clustered using Pearson correlation. Columns representing samples were
568 clustered using the *hclust* function, with non-supervised hierarchical clustering
569 performed based on sample distance, measured as 1 - Pearson correlation coefficient.
570 For individual genes, boxplots were plotted using the ggplot2 package (Wickham,
571 2016). Comparisons among groups were made using the Kruskal-Wallis test followed
572 by Dunn's post-hoc test, with p-values lower than 0.05 considered significant for both
573 tests. For RNA-Seq data, expression counts were normalized using variance
574 stabilizing transformation through the DESeq2 R package (Love et al., 2014).
575 Differences in some of the genes' expressions among analyzed groups were also
576 evaluated using the Kruskal-Wallis test followed by pairwise comparisons using
577 Dunn's test. In both cases, $p < 0.05$ indicated a significant difference.

Pathway enrichment analysis

Gene Set Enrichment Analysis (GSEA) was performed on LFC-ranked genes in each condition using the WebgestaltR R package (Liao et al., 2019; Subramanian et al., 2005). The analysis was performed with 1,000 permutations. The false discovery rate (FDR) cutoff was 0.05. Pathways with a minimum of 5 genes were selected. Result lists for each group had redundant pathways reduced using Webgestalt's implementation of the Affinity Propagation algorithm. A subsequent filter kept all gene sets that were enriched in OLP and at least one of the OSCC stage groups. These results were presented in dotplots using the ggplot2 package.

Overrepresentation analysis was performed for overlapping and OLP-exclusive DEGs, as well as co-expression modules' genes using the ReactomePA R package (Yu and He, 2016). Pathways with a minimum of 5 genes and FDR < 0.05 were selected, and the Benjamini-Hochberg (BH) method for multiple testing p-value correction was used (Benjamini and Hochberg, 1995). Pathways from the Reactome database were used for both GSEA and ORA (Jassal et al., 2020).

Immune Infiltration Cells Analysis

Tumor immune infiltration cells composition was estimated using CIBERSORTx (Newman et al., 2019). This tool uses a deconvolution algorithm to estimate immune cell types using gene expression data from samples composed of multiple cells (bulk). Batch-corrected, normalized expression data was used to estimate TIICs using CIBERSORTx's gene signatures for 22 cell types. These cell populations include naïve B cells, memory B cells, plasma cells, 7 T cell types (CD8+ T cells, naïve CD4+ T cells, resting CD4+ memory T cells, activated CD4+ memory T cells, follicular helper T cells, Tregs, γδ T cells), macrophages (M0 macrophages, M1 macrophages, M2

602 macrophages), resting mast cells, activated mast cells, resting NK cells, activated NK
603 cells, resting dendritic cells (resting DC), activated dendritic cells (activated DC),
604 monocytes, eosinophils, and neutrophils.

605 Stacked bar plots were generated from relative cell type populations using
606 ggplot2. Samples were clustered using the hclust function. Distances based on 1 -
607 Pearson correlation coefficient were used for Ward clustering of samples. Population
608 scores were individually compared between groups using Kruskal-Wallis test, followed
609 by pairwise comparisons using Dunn's test. In both cases, significant differences were
610 identified by $p < 0.05$.

611 Additionally, the signatures of genes related to pathogenic and non-pathogenic
612 Th17 cells were investigated using a 33-gene signature panel based on a previous
613 characterization of Th17 phenotypes in the literature (Lee et al., 2012). Gene
614 expression and sample clustering visualizations for pathogenic and non-pathogenic
615 Th17 cell signatures were made using the pheatmap R package (Kolde, 2019).
616 Boxplots for the signature's genes were plotted using the ggplot2 package (Wickham,
617 2016). Comparisons among groups were made using the Kruskal-Wallis test followed
618 by Dunn's post-hoc test, with p-values lower than 0.05 considered significant for both
619 tests.

620 ## Co-expression analysis

621 Gene co-expression modules were constructed using the Weighted Gene Co-
622 expression Network Analysis (WGCNA) R package (Langfelder and Horvath, 2008;
623 Zhang and Horvath, 2005). The 15,000 genes with the highest median absolute
624 deviation (MAD) were selected from the integrated dataset and used as input. A gene
625 pair similarity matrix was generated based on Pearson correlation and converted to a

weighted adjacency matrix by elevating it to a β value of 6. This matrix was used to build a topological overlap (TOM) and a dissimilarity matrix (1 - TOM). The dissimilarity matrix was used to build unsigned co-expression modules with a minimum size of 100 genes. WGCNA's module-trait relationship function was used to calculate correlations between module eigengenes and each of the groups (OLP, early OSCC, advanced OSCC, and normal samples). Correlations were considered significant when $|r| \geq 0.3$, and $p < 0.05$.

Interaction networks construction and drug-gene interactions identification

Genes in the magenta module were used to build a protein-protein interaction (PPI) network using interaction data from the STRING database, v. 11 (Szklarczyk et al., 2019). Interactions with a confidence score < 0.9 and disconnected vertices were discarded. Hub genes were determined by selecting vertices with a degree over the 9th decile of the network's degree distribution and comparisons in the individual hubs' expression levels were performed as described in the differential expression section.

Gene categories and FDA-approved, antineoplastic drug-gene interactions for hubs were identified using the DGIdb online tool (Freshour et al., 2021). Clinically actionable genes, transcription factors, and genes coding for protein families belonging to the druggable genome were identified (Hopkins and Groom, 2002; Russ and Lampel, 2005). Information such as $\log_2$ fold change, identified gene categories, degree, and whether a gene is a hub were also added to the network. The graph's largest connected component was used for visualization. Network manipulation was made using the igraph and tidygraph R packages (Csardi and Nepusz, 2006;

649 Pedersen, 2020). Network plots were constructed using the ggraph R package
650 (Pedersen, 2021).

651 Search for expression drug-response for differentially expressed
652 genes

653 Overlapping DEGs were used to search for drugs able to revert their expression
654 signatures using the L1000 Characteristic Direction Signature Search Engine
655 (L1000CDS^2) tool on the Library of Integrated Network-Based Cellular Signatures
656 (LINCS) Program platform (Duan et al., 2016; Stathias et al., 2020). This tool
657 compares input differentially expressed genes to LINCS-L1000 gene perturbation
658 data. Drug-gene combinations were ranked by search score, calculated based on the
659 overlap between input DEGs and signature DEGs, that is, gene sets that follow the
660 same perturbation patterns (up-regulating underexpressed genes or down-regulating
661 overexpressed genes) when interacting with a small molecule. The top 50 drug
662 signatures are presented as output. Additionally, putative drug combinations among
663 the small molecule signatures were estimated using this tool. Drug combinations were
664 ranked based on their signature overlaps and the top 50 combinations were provided.

665 Drug repositioning opportunities evaluation

666 Drugs identified both by interactions with hub genes and by investigating
667 expression reversion signatures were evaluated for repositioning opportunities using
668 the repoDB database, which compiles information from clinical trials (Brown and Patel,
669 2017). Drugs were searched for in the database and evaluated for whether they were
670 already tested in OSCC, OLP, or other neoplasms. In the case when a drug was used
671 in clinical trials, we have evaluated if it was approved or not for clinical use.

Acknowledgements

The authors would like to thank the Plataforma Multiusuário de Bioinformática of Instituto Nacional de Câncer (INCA) for providing the infrastructure for performing the analyses.

Competing interests

The authors declare no financial or non-financial competing interests.

Funding

This work was supported by CAPES scholarship (CL) and Instituto Nacional de Câncer – Ministério da Saúde.

Bibliography

Aggarwal S, John S, Sapra L, Sharma SC, Das SN. 2019. Targeted disruption of PI3K/Akt/mTOR signaling pathway, via PI3K inhibitors, promotes growth inhibitory effects in oral cancer cells. *Cancer Chemother Pharmacol* **83**:451–461. doi:10.1007/s00280-018-3746-x

Aghbari SMH, Abushouk AI, Attia A, Elmaraezy A, Menshawy A, Ahmed MS, Elsaadany BA, Ahmed EM. 2017. Malignant transformation of oral lichen planus and oral lichenoid lesions: A meta-analysis of 20095 patient data. *Oral Oncol* **68**:92–102. doi:10.1016/j.oraloncology.2017.03.012

Awasthi A, Kuchroo VK. 2009. Th17 cells: from precursors to players in inflammation and infection. *Int Immunol* **21**:489–498. doi:10.1093/intimm/dxp021

Bardellini E, Amadori F, Flocchini P, Bonadeo S, Majorana A. 2013. Clinicopathological features and malignant transformation of oral lichen planus: a 12-years retrospective study. *Acta Odontol Scand* **71**:834–840. doi:10.3109/00016357.2012.734407

Barnard NA, Scully C, Eveson JW, Cunningham S, Porter SR. 1993. Oral cancer development in patients with oral lichen planus. *J Oral Pathol Med* **22**:421–424. doi:10.1111/j.1600-0714.1993.tb00134.x

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**:D991-5. doi:10.1093/nar/gks1193

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**:289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Binato R, Santos EC, Boroni M, Demachki S, Assumpção P, Abdelhay E. 2018. A common molecular signature of intestinal-type gastric carcinoma indicates processes related to gastric carcinogenesis. *Oncotarget* **9**:7359–7371. doi:10.18632/oncotarget.23670

Brown AS, Patel CJ. 2017. A standard database for drug repositioning. *Sci Data* **4**:170029. doi:10.1038/sdata.2017.29

Cheng Y-SL, Gould A, Kurago Z, Fantasia J, Muller S. 2016. Diagnosis of oral lichen planus: a position paper of the American Academy of Oral and Maxillofacial Pathology. *Oral Surg Oral Med Oral Pathol Oral Radiol* **122**:332–354. doi:10.1016/j.oooo.2016.05.004

Chen N-X, Liu K, Liu X, Zhang X-X, Han D-Y. 2020. Induction and regulation of the immunoproteasome subunit β5i (PSMB8) in laryngeal and hypopharyngeal carcinoma cells. *Med Sci Monit* **26**:e923621. doi:10.12659/MSM.923621

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. 2016. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**:e71. doi:10.1093/nar/gkv1507

Corcoran SE, O'Neill LAJ. 2016. HIF1α and metabolic reprogramming in inflammation. *J Clin Invest* **126**:3699–3707. doi:10.1172/JCI84431

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal* **1695**.

Davis S, Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**:1846–1847. doi:10.1093/bioinformatics/btm254

Ding X-Q, Wang Z-Y, Xia D, Wang R-X, Pan X-R, Tong J-H. 2020. Proteomic profiling

of serum exosomes from patients with metastatic gastric cancer. *Front Oncol* **10**:1113. doi:10.3389/fonc.2020.01113

Duan Q, Reid SP, Clark NR, Wang Z, Fernandez NF, Rouillard AD, Readhead B, Tritsch SR, Hodos R, Hafner M, Niepel M, Sorger PK, Dudley JT, Bavari S, Panchal RG, Ma'ayan A. 2016. L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl* **2**. doi:10.1038/npjsba.2016.15

Farhi D, Dupin N. 2010. Pathophysiology, etiologic factors, and clinical management of oral lichen planus, part I: facts and controversies. *Clin Dermatol* **28**:100–108. doi:10.1016/j.clindermatol.2009.03.004

Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH. 2021. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* **49**:D1144–D1151. doi:10.1093/nar/gkaa1084

Gaur P, Qadir GA, Upadhyay S, Singh AK, Shukla NK, Das SN. 2012. Skewed immunological balance between Th17 (CD4(+)IL17A (+)) and Treg (CD4 (+)CD25 (+)FOXP3 (+)) cells in human oral squamous cell carcinoma. *Cell Oncol (Dordr)* **35**:335–343. doi:10.1007/s13402-012-0093-5

Ge X, Wang L, Li M, Xu N, Yu F, Yang F, Li R, Zhang F, Zhao B, Du J. 2019. Vitamin D/VDR signaling inhibits LPS-induced IFNγ and IL-1β in Oral epithelia by regulating hypoxia-inducible factor-1α signaling pathway. *Cell Commun Signal* **17**:18. doi:10.1186/s12964-019-0331-9

Giacomelli L, Oluwadara O, Chiappe G, Barone A, Chiappelli F, Covani U. 2009. Relationship between human oral lichen planus and oral squamous cell carcinoma at a genomic level: a datamining study. *Bioinformation* **4**:258–262. doi:10.6026/97320630004258

Giuliani M, Troiano G, Cordaro M, Corsalini M, Gioco G, Lo Muzio L, Pignatelli P, Lajolo C. 2019. Rate of malignant transformation of oral lichen planus: A systematic review. *Oral Dis* **25**:693–709. doi:10.1111/odi.12885

Gonzalez-Moles MA, Scully C, Gil-Montoya JA. 2008. Oral lichen planus: controversies surrounding malignant transformation. *Oral Dis* **14**:229–243. doi:10.1111/j.1601-0825.2008.01441.x

Groettrup M, van den Broek M, Schwarz K, Macagno A, Khan S, de Giuli R, Schmidtke G. 2001. Structural plasticity of the proteasome and its function in antigen processing. *Crit Rev Immunol* **21**:339–358.

Hallopeau H. 1910. Sur un cas de lichen de Wilson gingival avec néoplasie voisine dans la région maxillaire. . *Bull Soc Fr Dermatol Syphiligr* 32.

Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**:169. doi:10.1186/s12859-015-0611-3

Hopkins AL, Groom CR. 2002. The druggable genome. *Nat Rev Drug Discov* **1**:727–730. doi:10.1038/nrd892

Iijima W, Ohtani H, Nakayama T, Sugawara Y, Sato E, Nagura H, Yoshie O, Sasano T. 2003. Infiltrating CD8+ T cells in oral lichen planus predominantly express CCR5 and CXCR3 and carry respective chemokine ligands RANTES/CCL5 and IP-10/CXCL10 in their cytolytic granules: a potential self-recruiting mechanism. *Am J Pathol* **163**:261–268. doi:10.1016/S0002-9440(10)63649-8

Ismail SB, Kumar SKS, Zain RB. 2007. Oral lichen planus and lichenoid reactions: etiopathogenesis, diagnosis, management and malignant transformation. *J Oral Sci* **49**:89–106. doi:10.2334/josnusd.49.89

Itoiz ME, Conti CJ, Lanfranchi HE, Mamrack M, Klein-Szanto AJ. 1985.

782 Immunohistochemical detection of filaggrin in preneoplastic and neoplastic lesions
783 of the human oral mucosa. *Am J Pathol* **119**:456–461.
784 Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook
785 J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C,
786 Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, D'Eustachio P. 2020. The
787 Reactome Pathway Knowledgebase. *Nucleic Acids Res* **48**:D498–D503.
788 doi:10.1093/nar/gkz1031
789 Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray
790 expression data using empirical Bayes methods. *Biostatistics* **8**:118–127.
791 doi:10.1093/biostatistics/kxj037
792 Kakumu T, Sato M, Goto D, Kato T, Yogo N, Hase T, Morise M, Fukui T, Yokoi K,
793 Sekido Y, Girard L, Minna JD, Byers LA, Heymach JV, Coombes KR, Kondo M,
794 Hasegawa Y. 2017. Identification of proteasomal catalytic subunit PSMA6 as a
795 therapeutic target for lung cancer. *Cancer Sci* **108**:732–743.
796 doi:10.1111/cas.13185
797 Kazanietz MG, Durando M, Cooke M. 2019. CXCL13 and its receptor CXCR5 in
798 cancer: inflammation, immune response, and beyond. *Front Endocrinol*
799 *(Lausanne)* **10**:471. doi:10.3389/fendo.2019.00471
800 Kengkarn S, Petmitr S, Boonyuen U, Reamtong O, Poomsawat S, Sanguansin S.
801 2020. Identification of novel candidate biomarkers for oral squamous cell
802 carcinoma based on whole gene expression profiling. *Pathol Oncol Res* **26**:2315–
803 2325. doi:10.1007/s12253-020-00828-w
804 Kitamura R, Toyoshima T, Tanaka H, Kawano S, Kiyosue T, Matsubara R, Goto Y,
805 Hirano M, Oobu K, Nakamura S. 2012. Association of cytokeratin 17 expression
806 with differentiation in oral squamous cell carcinoma. *J Cancer Res Clin Oncol*

138:1299–1310. doi:10.1007/s00432-012-1202-6

Kiuchi T, Tomaru U, Ishizu A, Imagawa M, Iwasaki S, Suzuki A, Otsuka N, Ohhara Y, Kinoshita I, Matsuno Y, Dosaka-Akita H, Kasahara M. 2021. Expression of the immunoproteasome subunit β5i in non-small cell lung carcinomas. *J Clin Pathol* **74**:300–306. doi:10.1136/jclinpath-2020-206618

Kloetzel PM. 2001. Antigen processing by the proteasome. *Nat Rev Mol Cell Biol* **2**:179–187. doi:10.1038/35056572

Kolde R. 2019. pheatmap: Pretty heatmaps. CRAN.

Krijgsman D, Roelands J, Hendrickx W, Bedognetti D, Kuppen PJK. 2020. HLA-G: A New Immune Checkpoint in Cancer? *Int J Mol Sci* **21**. doi:10.3390/ijms21124528

Kurago ZB. 2016. Etiology and pathogenesis of oral lichen planus: an overview. *Oral Surg Oral Med Oral Pathol Oral Radiol* **122**:72–80. doi:10.1016/j.oooo.2016.03.011

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559. doi:10.1186/1471-2105-9-559

Larsen KR, Johansen JD, Reibel J, Zachariae C, Rosing K, Pedersen AML. 2017. Filaggrin gene mutations and the distribution of filaggrin in oral mucosa of patients with oral lichen planus and healthy controls. *J Eur Acad Dermatol Venereol* **31**:887–893. doi:10.1111/jdv.14098

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**:882–883. doi:10.1093/bioinformatics/bts034

Lee Y, Awasthi A, Yosef N, Quintana FJ, Xiao S, Peters A, Wu C, Kleinewietfeld M, Kunder S, Hafler DA, Sobel RA, Regev A, Kuchroo VK. 2012. Induction and molecular signature of pathogenic TH17 cells. *Nat Immunol* **13**:991–999.

doi:10.1038/ni.2416

Leontieva OV, Blagosklonny MV. 2016. Gerosuppression by pan-mTOR inhibitors. *Aging (Albany NY)* **8**:3535–3551. doi:10.18632/aging.101155

Liang X, Deng M, Zhang C, Ping F, Wang H, Wang Y, Fan Z, Ren X, Tao X, Wu T, Xu J, Cheng B, Xia J. 2019. Combined class I histone deacetylase and mTORC1/C2 inhibition suppresses the initiation and recurrence of oral squamous cell carcinomas by repressing SOX2. *Cancer Lett* **454**:108–119. doi:10.1016/j.canlet.2019.04.010

Liao S-C, Hsieh P-C, Huang J-S, Hsu C-W, Yuan K. 2012. Aberrant keratinization of reticular oral lichen planus is related to elastolysis. *Oral Surg Oral Med Oral Pathol Oral Radiol* **113**:808–816. doi:10.1016/j.oooo.2012.02.007

Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**:W199–W205. doi:10.1093/nar/gkz401

Liu J, Yang Q, Sun H, Wang X, Saiyin H, Zhang H. 2020. The circ-AMOTL1/ENO1 Axis Implicated in the Tumorigenesis of OLP-Associated Oral Squamous Cell Carcinoma. *Cancer Manag Res* **12**:7219–7230. doi:10.2147/CMAR.S251348

Liu Z, Yu S, Ye S, Shen Z, Gao L, Han Z, Zhang P, Luo F, Chen S, Kang M. 2020. Keratin 17 activates AKT signalling and induces epithelial-mesenchymal transition in oesophageal squamous cell carcinoma. *J Proteomics* **211**:103557. doi:10.1016/j.jprot.2019.103557

Li Q, Song X, Ji Y, Jiang H, Xu L. 2013. The dual mTORC1 and mTORC2 inhibitor AZD8055 inhibits head and neck squamous cell carcinoma cell growth in vivo and in vitro. *Biochem Biophys Res Commun* **440**:701–706. doi:10.1016/j.bbrc.2013.09.130

857 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change
858 and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.
859 doi:10.1186/s13059-014-0550-8
860 Ma R-J, Tan Y-Q, Zhou G. 2019. Aberrant IGF1-PI3K/AKT/MTOR signaling pathway
861 regulates the local immunity of oral lichen planus. *Immunobiology* **224**:455–461.
862 doi:10.1016/j.imbio.2019.01.004
863 McGrath JA, Uitto J. 2008. The filaggrin story: novel insights into skin-barrier function
864 and disease. *Trends Mol Med* **14**:20–27. doi:10.1016/j.molmed.2007.10.006
865 Micci F, Panagopoulos I, Haugom L, Dahlback H-SS, Pretorius ME, Davidson B,
866 Abeler VM, Tropé CG, Danielsen HE, Heim S. 2013. Genomic aberration patterns
867 and expression profiles of squamous cell carcinomas of the vulva. *Genes*
868 *Chromosomes Cancer* **52**:551–563. doi:10.1002/gcc.22053
869 Michifuri Y, Hirohashi Y, Torigoe T, Miyazaki A, Fujino J, Tamura Y, Tsukahara T,
870 Kanaseki T, Kobayashi J, Sasaki T, Takahashi A, Nakamori K, Yamaguchi A,
871 Hiratsuka H, Sato N. 2013. Small proline-rich protein-1B is overexpressed in
872 human oral squamous cell cancer stem-like cells and is related to their growth
873 through activation of MAP kinase signal. *Biochem Biophys Res Commun* **439**:96–
874 102. doi:10.1016/j.bbrc.2013.08.021
875 Müller S. 2017. Update from the 4th Edition of the World Health Organization of Head
876 and Neck Tumours: Tumours of the Oral Cavity and Mobile Tongue. *Head Neck*
877 *Pathol* **11**:33–40. doi:10.1007/s12105-017-0792-3
878 Munkácsy G, Abdul-Ghani R, Mihály Z, Tegze B, Tchernitsa O, Surowiak P, Schäfer
879 R, Györffy B. 2010. PSMB7 is associated with anthracycline resistance and is a
880 prognostic biomarker in breast cancer. *Br J Cancer* **102**:361–368.
881 doi:10.1038/sj.bjc.6605478

882 Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust
883 MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. 2019. Determining
884 cell type abundance and expression from bulk tissues with digital cytometry. *Nat*
885 *Biotechnol* **37**:773–782. doi:10.1038/s41587-019-0114-2
886 Ohkura S, Kondoh N, Hada A, Arai M, Yamazaki Y, Sindoh M, Takahashi M,
887 Matsumoto I, Yamamoto M. 2005. Differential expression of the keratin-4, -13, -
888 14, -17 and transglutaminase 3 genes during the development of oral squamous
889 cell carcinoma from leukoplakia. *Oral Oncol* **41**:607–613.
890 doi:10.1016/j.oraloncology.2005.01.011
891 Patel S, Kartasova T, Segre JA. 2003. Mouse Sprr locus: a tandem array of
892 coordinately regulated genes. *Mamm Genome* **14**:140–148. doi:10.1007/s00335-
893 002-2205-4
894 Pedersen TL. 2021. ggraph: An Implementation of Grammar of Graphics for Graphs
895 and Networks. CRAN.
896 Pedersen TL. 2020. tidygraph: A Tidy API for Graph Manipulation. CRAN.
897 Peng Q, Zhang J, Ye X, Zhou G. 2017. Tumor-like microenvironment in oral lichen
898 planus: evidence of malignant transformation? *Expert Rev Clin Immunol* **13**:635–
899 643. doi:10.1080/1744666X.2017.1295852
900 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers
901 differential expression analyses for RNA-sequencing and microarray studies.
902 *Nucleic Acids Res* **43**:e47. doi:10.1093/nar/gkv007
903 Roopashree MR, Gondhalekar RV, Shashikanth MC, George J, Thippeswamy SH,
904 Shukla A. 2010. Pathogenesis of oral lichen planus--a review. *J Oral Pathol Med*
905 **39**:729–734. doi:10.1111/j.1600-0714.2010.00946.x
906 Ruokonen HMA, Juurikivi A, Kauppila T, Heikkinen AM, Seppänen-Kaijansinkko R.

2017. High percentage of oral lichen planus and lichenoid lesion in oral squamous cell carcinomas. *Acta Odontol Scand* **75**:442–445. doi:10.1080/00016357.2017.1332777

Russ AP, Lampel S. 2005. The druggable genome: an update. *Drug Discov Today* **10**:1607–1610. doi:10.1016/S1359-6446(05)03666-4

Rutz S, Eidenschenk C, Kiefer JR, Ouyang W. 2016. Post-translational regulation of RORγt-A therapeutic target for the modulation of interleukin-17-mediated responses in autoimmune diseases. *Cytokine Growth Factor Rev* **30**:1–17. doi:10.1016/j.cytogfr.2016.07.004

Sakamoto K, Aragaki T, Morita K, Kawachi H, Kayamori K, Nakanishi S, Omura K, Miki Y, Okada N, Katsube K-I, Takizawa T, Yamaguchi A. 2011. Down-regulation of keratin 4 and keratin 13 expression in oral squamous cell carcinoma and epithelial dysplasia: a clue for histopathogenesis. *Histopathology* **58**:531–542. doi:10.1111/j.1365-2559.2011.03759.x

Schaaij-Visser TBM, Graveland AP, Gauci S, Braakhuis BJM, Buijze M, Heck AJR, Kuik DJ, Bloemena E, Leemans CR, Slijper M, Brakenhoff RH. 2009. Differential proteomics identifies protein biomarkers that predict local relapse of head and neck squamous cell carcinomas. *Clin Cancer Res* **15**:7666–7675. doi:10.1158/1078-0432.CCR-09-2134

Shen X, Wang P, Dai P, Jin B, Tong Y, Lin H, Shi G. 2018. Correlation between human leukocyte antigen-G expression and clinical parameters in oral squamous cell carcinoma. *Indian J Cancer* **55**:340–343. doi:10.4103/ijc.IJC_602_17

Shen Z, Chen L, Liu Y-F, Gao T-W, Wang G, Fan X-L, Fan J-Y, Fan P-S, Li C-Y, Liu B, Dang Y-P, Li C-X. 2006. Altered keratin 17 peptide ligands inhibit in vitro proliferation of keratinocytes and T cells isolated from patients with psoriasis. *J*

Am Acad Dermatol **54**:992–1002. doi:10.1016/j.jaad.2006.02.033

Shen Z-Y, Liu W, Feng J-Q, Zhou H-W, Zhou Z-T. 2011. Squamous cell carcinoma development in previously diagnosed oral lichen planus: de novo or transformation? *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **112**:592–596. doi:10.1016/j.tripleo.2011.05.029

Shimada K, Ochiai T, Hasegawa H. 2018. Ectopic transglutaminase 1 and 3 expression accelerating keratinization in oral lichen planus. *J Int Med Res* **46**:4722–4730. doi:10.1177/0300060518798261

Stathias V, Turner J, Koleti A, Vidovic D, Cooper D, Fazel-Najafabadi M, Pilarczyk M, Terryn R, Chung C, Umeano A, Clarke DJB, Lachmann A, Evangelista JE, Ma'ayan A, Medvedovic M, Schürer SC. 2020. LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Res* **48**:D431–D439. doi:10.1093/nar/gkz1023

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**:15545–15550. doi:10.1073/pnas.0506580102

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering C von. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**:D607–D613. doi:10.1093/nar/gky1131

Tan S, Li H, Zhang W, Shao Y, Liu Y, Guan H, Wu J, Kang Y, Zhao J, Yu Q, Gu Y, Ding K, Zhang M, Qian W, Zhu Y, Cai H, Chen C, Lobie PE, Zhao X, Sun J, Zhu

957 T. 2018. NUDT21 negatively regulates PSMB2 and CXXC5 by alternative
958 polyadenylation and contributes to hepatocellular carcinoma suppression.
959 *Oncogene* **37**:4887–4900. doi:10.1038/s41388-018-0280-6
960 van der Waal I. 2010. Potentially malignant disorders of the oral and oropharyngeal
961 mucosa; present concepts of management. *Oral Oncol* **46**:423–425.
962 doi:10.1016/j.oraloncology.2010.02.016
963 Walsh CJ, Hu P, Batt J, Santos CCD. 2015. Microarray Meta-Analysis and Cross-
964 Platform Normalization: Integrative Genomics for Robust Biomarker Discovery.
965 *Microarrays (Basel)* **4**:389–406. doi:10.3390/microarrays4030389
966 Wang H, Zhang D, Han Q, Zhao X, Zeng X, Xu Y, Sun Z, Chen Q. 2016. Role of
967 distinct CD4(+) T helper subset in pathogenesis of oral lichen planus. *J Oral Pathol*
968 *Med* **45**:385–393. doi:10.1111/jop.12405
969 Wang X-X, Sun H-Y, Yang Q-Z, Guo B, Sai Y, Zhang J. 2017. Hypoxia-inducible
970 factor-1α and glucose transporter 1 in the malignant transformation of oral lichen
971 planus. *Int J Clin Exp Pathol* **10**:8369–8376.
972 Wang Z, Yang M-Q, Lei L, Fei L-R, Zheng Y-W, Huang W-J, Li Z-H, Liu C-C, Xu H-T.
973 2019. Overexpression of KRT17 promotes proliferation and invasion of non-small
974 cell lung cancer and indicates poor prognosis. *Cancer Manag Res* **11**:7485–7497.
975 doi:10.2147/CMAR.S218926
976 Warnakulasuriya S, Johnson NW, van der Waal I. 2007. Nomenclature and
977 classification of potentially malignant disorders of the oral mucosa. *J Oral Pathol*
978 *Med* **36**:575–580. doi:10.1111/j.1600-0714.2007.00582.x
979 Wickham H. 2016. ggplot2 - Elegant Graphics for Data Analysis, 2nd ed. Cham:
980 Springer International Publishing. doi:10.1007/978-3-319-24277-4
981 Wickham LF. 1895. Sur un signe pathognomonique de lichen de Wilson (lichen plan)

stries et ponctuations grisartres. *Ann Dermatol Syphiligr (Paris)* **6**:517–520.

Yang C-Y, Liu C-R, Chang IY-F, OuYang C-N, Hsieh C-H, Huang Y-L, Wang C-I, Jan F-W, Wang W-L, Tsai T-L, Liu H, Tseng C-P, Chang Y-S, Wu C-C, Chang K-P. 2020. Cotargeting CHK1 and PI3K Synergistically Suppresses Tumor Growth of Oral Cavity Squamous Cell Carcinoma in Patient-Derived Xenografts. *Cancers (Basel)* **12**. doi:10.3390/cancers12071726

Yang Q, Sun H, Wang X, Yu X, Zhang J, Guo B, Hexige S. 2020. Metabolic changes during malignant transformation in primary cells of oral lichen planus: Succinate accumulation and tumour suppression. *J Cell Mol Med* **24**:1179–1188. doi:10.1111/jcmm.14376

Yu G, He Q-Y. 2016. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* **12**:477–479. doi:10.1039/c5mb00663e

Zerfas BL, Maresh ME, Trader DJ. 2020. The immunoproteasome: an emerging target in cancer and autoimmune and neurological disorders. *J Med Chem* **63**:1841–1858. doi:10.1021/acs.jmedchem.9b01226

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**:Article17. doi:10.2202/1544-6115.1128

Supporting information

Supplementary File 1 – Supplementary Figures S1-S6

Supplementary File 2 – Supplementary Tables S1-S10