Universidad Europea

GRADUATION PROJECT

Degree in Dentistry

STUDY OF THE TRANSCRIPTOME IN ORAL CANCER TUMOUR CELLS USING A BIOINFORMATIC APPROACH

Madrid, academic year 2024/2025

Identification number: 105

RESUMEN

Introducción: El cáncer oral representa un desafío sanitario a nivel mundial, siendo el carcinoma oral de células escamosas (CCEO) responsable de más del 90 % de los casos. Su desarrollo está asociado a la acumulación de mutaciones somáticas que alteran los mecanismos de regulación celular. Estas alteraciones genéticas afectan a protooncogenes y genes supresores de tumores, dando lugar a una proliferación descontrolada, pérdida de la función normal, invasión de tejidos cercanos y metástasis potencial. El análisis transcriptómico proporciona información valiosa sobre estos cambios, y las herramientas bioinformáticas ofrecen un enfoque eficiente para procesar cantidades tan grandes de datos; **Objetivos:** Identificar genes expresados diferencialmente (DEG) en células tumorales de cáncer oral en comparación con tejido sano; **Metodología:** Se analizaron 8 muestras humanas, de las cuales 5 correspondían a tejido sano (muestras 1, 2, 4, 6 y 8) y 3 a tejido tumoral (muestras 3, 5 y 7). El análisis de RNA-seq se realizó en Galaxy Europe, utilizando FastQC, HISAT2 y FeatureCounts para el control de calidad, alineación y cuantificación. Limma-Voom se empleó para la identificación de DEG; **Resultados:** Se obtuvo una alta calidad de datos (>97 % de alineación), permitiendo identificar 20 genes desregulados significativos. El análisis reveló la implicación de rutas clave como Wnt/β-catenina, apoptosis, reprogramación metabólica y regulación inmunitaria, destacando la utilidad de herramientas bioinformáticas en la investigación del cáncer oral; **Conclusiones:** Este estudio demuestra la eficacia de los enfoques bioinformáticos en la investigación del cáncer oral, identificando con éxito los DEG en las células tumorales y destacando las vías implicadas (Wnt/β-catenina, reprogramación metabólica, homeostasis redox, regulación inmunitaria).

PALABRAS CLAVE

Odontología, cáncer oral, transcriptoma, secuenciación de ARN, bioinformática.

ABSTRACT

Introduction: Oral cancer remains a global health concern, with oral squamous cell carcinoma (OSCC) accounting for over 90% of cases. At the molecular level, oral cancer develops through the accumulation of somatic mutations that interfere with cellular regulatory mechanisms. These genetic alterations affect proto-oncogenes and tumor suppressor genes, resulting in uncontrolled proliferation, loss of normal function, invasion of nearby tissues, and potential metastasis. Transcriptomic analysis provides valuable insight into these changes, and bioinformatic tools offer an efficient approach to process such large quantities of data; **Objectives**: The main objective of this study was to identify differentially expressed genes (DEGs) in oral cancer tumor cells compared to healthy tissue; **Methodology:** The study comprised of 8 samples from Homo sapiens: Samples 1, 2, 4, 6 and 8 were from healthy patients, and 3, 5 and 7 from tumor tissue. RNA sequencing data from oral cancer and healthy tissue samples were analyzed using a bioinformatic pipeline implemented on Galaxy Europe. Quality control, sequence alignment, and count generation were performed using FastQC, HISAT2, and FeatureCounts. Differential expression analysis was conducted using Limma-Voom; **Results**: The RNA-seq pipeline achieved >97 % alignment efficiency and minimal adapter contamination, confirming data quality and suitability for downstream analysis. A table of the 20 most statistically significant dysregulated genes was constructed, revealing enrichment in Wnt/β-catenin signaling and apoptosis pathways; **Conclusions:** This study demonstrates the effectiveness of bioinformatic approaches in oral cancer research, successfully identifying DEGs in tumor cells and highlights pathways involved (Wnt/β-catenin, metabolic reprogramming, redox homeostasis, immune regulation).

KEYWORDS

Odontology, oral cancer, transcriptome, RNA sequencing, bioinformatics.

INDEX

1. INTRODUCTION

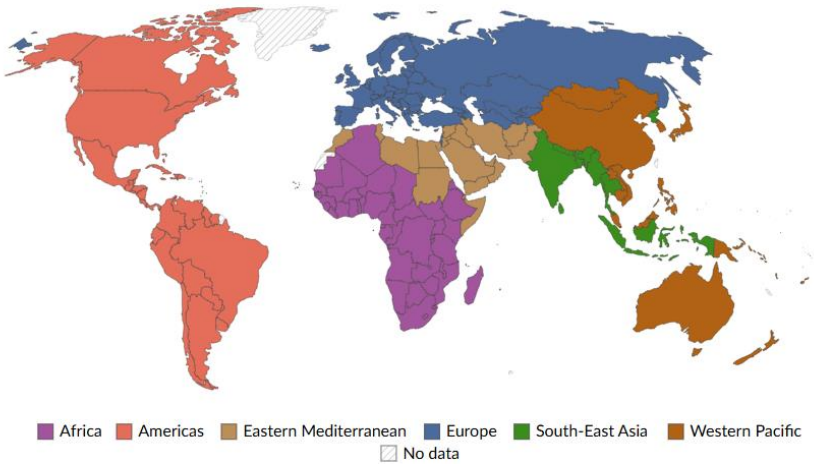
1.1. Clinical Relevance in Global Health

In 2021, the Seventy-fourth World Health Assembly passed a Resolution on oral health. Consequently, in 2022, the Secretariat established the "Global Strategy and Action Plan on Oral Health 2023–2030", which identifies oral cancer as "[one of] those with highest public health relevance"(1). According to World Health Organization (WHO) data, oral cancer ranks as the 13th most frequently occurring cancer globally. Reports combining oral cancer with pharyngeal cancers indicate they together represent the sixth most prevalent cancer worldwide (2).

1.1.1. Epidemiology

Research highlights that higher-risk groups for oral cancer include men and older individuals, with men facing a two to three-fold increased risk compared to women (3). Socio-economic factors significantly impact risk: worldwide, both crude and age-standardized incidence rates are higher in developed areas, whereas mortality rates are elevated in less developed regions, reflecting social disparities (4). Based on the Global Burden of Cancer Study by the United Nations Development Program in 2012, the WHO South-East Asia region (SEARO) and WHO Europe region (EURO) show the most critical figures.

Figure 1. World Regions according to World Health Organization (5)



1.2 Classification

In 2007, the WHO published its first classification of oral potentially malignant disorders (OPMDs). In clinical practice, "malignant" typically refers to disease that causes obvious and often rapid organ injury (6). Oral cancer has been defined as Squamous Cell Carcinoma (OSCC)

given that 90% of said cancer is histologically originated in the squamous cells (3,7). Squamous cells are part of the epithelial tissue, forming protective barriers on body surfaces and internal organs.

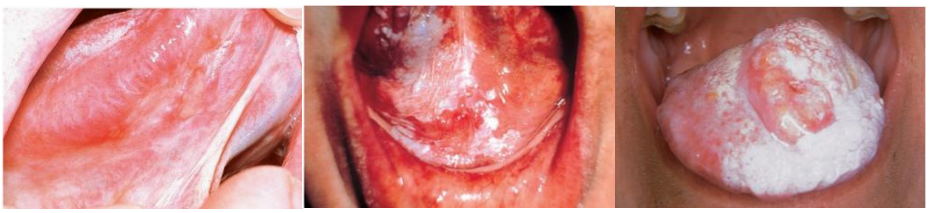
Within the scope of oral cancer are included malignant neoplasms of the lip, oral cavity and oropharynx (8). The tongue, particularly its ventral–lateral aspect, is the primary site for roughly 40% of oral cancers, the floor of the mouth accounts for about 30%, and the lower lip is affected less commonly (9).

1.2.1 Types of Oral Cancer

OSCC is an aggressive cancer affecting the oral epithelium, histologically originated in the squamous cells (3,7). It represents the predominant form of oral cancer, making up over 90% of all malignant growths within the oral cavity. While oral cancer can develop at any age, it is most commonly seen in older adults. Recent research shows that 90% of cases occur in individuals over 45 years old, with a significantly higher occurrence in men compared to women, at a ratio of 2.6 to 1. It is most frequently located intraorally at the lateral border and ventral surface of tongue. Location of the lesion is also dependent on habits, as chronic betel nut users more often present OSCC on the buccal mucosa. The gingiva is the site of about 5% to 10% of all cases of oral SCC (9).

The clinicopathological aspect of OSCC is heterogenous, demonstrated in Figure 2. It can manifest in a range of colors and surface patterns: most often red and white, exophytic, infiltrative or ulcerated (9).

Figure 2. OSCC ulcer on ventral tongue, floor of mouth, tumor of tongue (in order left to right) (10)



There are other types of oral cancers in addition to squamous cell carcinoma, the most common are: salivary gland cancer, basal cell carcinoma, lymphoma, melanoma, and sarcoma (7), for more information refer to Annex 1.

1.3 Phases of Cancer Evolution

Figure 3. Model of Cancer Evolution (11)

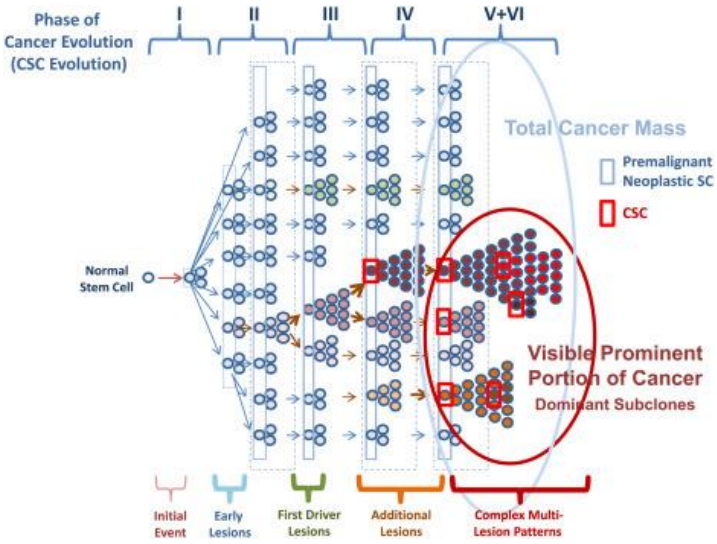


Figure 3, adapted from Valent et al., presents a conceptual model of how cancer develops through progressive genetic changes originating in a normal stem cell. The process is described in six distinct phases, illustrating the gradual accumulation of mutations and the emergence of cellular heterogeneity within the tumor mass.

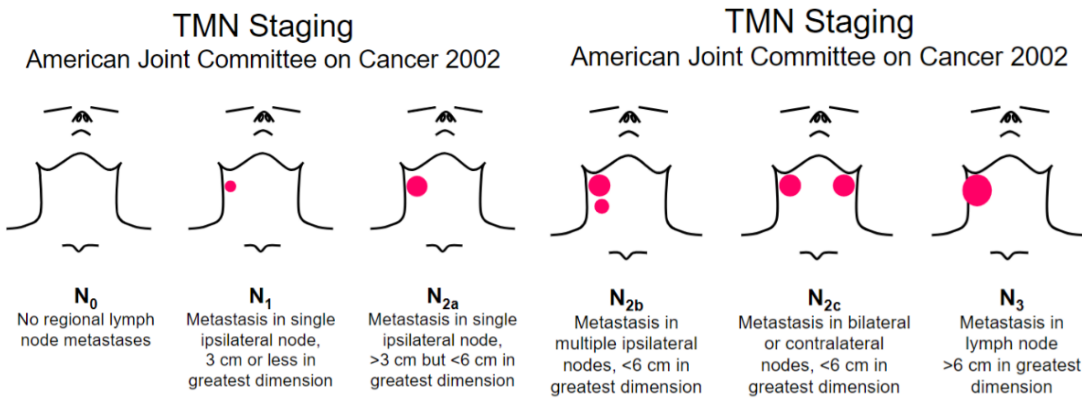
- Phase I: Initiation – A genetic or epigenetic hit alters the stem cell's DNA.
- Phase II: Early Clonal Expansion – The altered stem cell spawns a small, premalignant clone.
- Phase III: Acquisition of Driver Mutations – One or more "driver" mutations emerge, enabling certain clones to proliferate more aggressively than others.
- Phase IV: Progressive Mutation Accumulation – As the clone expands, further mutations create subclonal diversity.
- Phases V and VI: Advanced Clonal Diversification –Multiple genetically distinct subclones coexist within the tumor. One or more subclones become dominant forming the major and clinically visible portion of cancer, and cancer stem cells (shown in red) sustain growth and evolution.

1.3.1 Oral Cancer Metastasis

Metastasis is "the spread of cancer cells from the place where they first formed to another part of the body" (17) and occurs in oral cancer with variable probability. Cervical lymph node metastasis is "universally accepted as the main factor influencing survival in patients with

(OSCC)" (12). The TNM staging system (Tumor, Node, Metastasis) as seen in Figure 4 is crucial for assessing it. The "N" component specifically categorizes the extent of lymph node involvement, indicating the cancer's spread beyond the primary site. This system informs prognosis, guiding the course of treatment in accordance. For instance, the N3 classification is associated with a significantly poorer prognosis compared to N1 or N2b stages.

Figure 4. TMN Staging: Extent of Lymph node involvement: Taken from World Health Organization Classification of Tumors (13).



1.4 Risk Factors

The International Agency for Research on Cancer (IARC) have cited smoked and smokeless tobacco use as carcinogenic to humans (14). Together with alcohol consumption, they are regarded as the primary causes of oral cancer (15). Other risk factors include betel nut chewing, high-risk Human Papillomaviruses and Epstein-Barr Virus presence, chewing habits, diet and nutrition, and chronic inflammation among many others (16–18).

What lacks clarity in scientific literature is the role of genetic factors in oral cancer incidence- also known as Family History of Cancer (FHC). Though many epidemiological studies suggest a possible correlation (19–21), "some researchers believe that there is no evidence of a clear hereditary trait for oral cancers, except for Cowden syndrome and congenital dyskeratosis" (18).

1.5 Genetics and Oral Cancer

Cancer is the result of an accumulation of alterations (known as mutations) in the cellular pathways that regulate excitation and inhibition of cellular processes (22). It generally takes three to six somatic alterations to convert a normal cell into a malignant one (23). As these mutations accumulate, the cell gains independence from the surrounding oral epithelium, overriding normal cellular functions. This eventually leads to unchecked growth, stimulation of

new blood vessel formation, and the capacity to invade nearby tissues or even reach other parts of the body (22).

Genetic harm in oral cancer cells can be categorized into two kinds: dominant changes, often found in proto-oncogenes and occasionally in certain tumor suppressor genes (TSGs), which lead to an increase of function. On the other hand, recessive changes typically seen in genes involved in growth-inhibitory pathways or commonly in TSGs result in a loss of function (22).

1.5.1 Functional Genomics: the role of the Transcriptome in Oral Cancer

Functional genomics examines how genes and their products (proteins, RNAs) function and interact in living systems (30). The transcriptome represents the complete set of transcripts (RNA species) i.e. both coding and non-coding, within a cell, tissue, or organ. Unlike the largely stable nuclear genome, it is highly dynamic since it varies according to factors like cell cycle stage, tissue type, environmental exposure, ageing, disease, and other variables. This adaptability makes the transcriptome a valuable tool for identifying gene functions (22,24).

The transcriptome approach, which involves large-scale measurement of mRNA, quickly became a favored method within the field of functional genomics. It allows analysis of cellular activity on a grand scale by simultaneously analyzing activity of many genes in cells and tissues, known as parallel hybridization methods (22,25) .

1.6 The Bioinformatic Approach

1.6.1 What is Bioinformatics?

As defined by the National Human Genome Research Institute, bioinformatics “is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences” (26). A bioinformatic approach to studying the transcriptome in oral cancer cells involves the use of computational tools and algorithms to extract meaningful biological insights from large-scale sequencing data, such as RNA sequencing (RNA-seq) datasets, which measure gene expression levels in cancer cells.

1.6.2 Advantages of Bioinformatic application in Oral Cancer Research

The large amounts of data that need to be processed make it impractical to analyze manually. Conveniently, bioinformatics provides the tools to manage the grand scale of information

efficiently. Other advantages include the possibility of gene expression profiling, pathway analysis, ability to detect mutations, copy number variations, and the analysis of transcriptomes of individual patients in order to personalize treatment strategies based on specific gene expression profiles and genetic mutations (27–29).

1.6.3 Relevance to this Study

This study applies a bioinformatics workflow to RNA-seq data from healthy and oral cancer tumor samples to pinpoint differentially expressed genes. The entire pipeline is performed on Galaxy Europe, a web-based platform whose accessible, reproducible workflows ensure transparency and are used to process raw reads and quantify expression levels.

1.7 Hypothesis

It is hypothesized that differential gene expression will be observed between tumor and healthy samples from oral mucosa, and bioinformatic tools will effectively detect these changes.

2. OBJETIVE

1. To find genes that are differentially expressed in oral cancer tumors compared to healthy cells.

3. MATERIAL AND METHODS

3.1 The Samples

- Description of the sample: Samples are taken from the study "RNA Sequencing of Oral Cancer Tumor Tissue and Healthy tissue", conducted by Gujarat Biotechnology Research Centre. Comprising of 8 samples from Homo sapiens: Samples 1, 2, 4, 6 and 8 are from healthy patients, and 3, 5 and 7 from tumor tissue.
- Design: RNA sequence analysis
- Instrument: Illumina MiSeq
- Source: Transcriptomic
- Selection: Random

  Referring to RNA-seq library preparation: random selection uses random primers without enriching for specific RNAs (e.g., mRNA), yielding a library containing a mixture of RNA species (mRNA, rRNA, tRNA).
- Layout: Paired

  Specifies single-end versus paired-end sequencing. Paired-end reads both ends of each fragment—forward and reverse—boosting mapping accuracy and enabling structural-variation detection.
- Date of completion: June 29 2022
- Description of data storage

  The RNA-seq data are publicly available in NCBI's SRA under accession: SRP384104. The eight samples include metadata on platform (Illumina MiSeq), file sizes, read counts, and library prep, enabling easy public data download and research reproducibility.

3.2 The Galaxy Platform

The Galaxy platform is an open-source, web-based bioinformatics tool that simplifies large-scale data analysis. It provides an intuitive interface for managing RNA-Seq data workflows, from quality control to functional analysis, using a variety of built-in tools. It is ideal for this thesis as it allows reproducibility and scalability, making it easy to process the RNA data from the eight tumor and healthy tissue samples collected.

3.3 Tools Used in Analysis

Each tool is tailored to a specific step in the pipeline.

A. Quality Control

- **FastQC**: Evaluates raw sequence data quality.
- **MultiQC**: Aggregates FastQC reports for a comprehensive overview.

B. Alignment

- **HISAT2**: Aligns RNA reads to the reference genome.
- **Samtools Flagstat**: Performs manipulations like sorting and indexing on BAM files.

C. Count Generation

- **FeatureCounts**: Assigns aligned reads to genes.

D. Differential Expression Analysis

- **Limma-Voom**: Suitable for large datasets with RNA-seq data.

3.4 Pipeline for RNA-Seq Data Analysis

1. RNA Extraction: Extract RNA from tumor samples using standard lab methods.
2. Library Preparation: Convert RNA to cDNA and prepare for sequencing.
3. Sequencing: Perform RNA-Seq using high-throughput platforms like Illumina.
4. Quality Control (QC): Use tools to assess and improve sequence quality.
5. Alignment: Map reads to the reference genome.
6. Count Generation: Generate read counts per gene.
7. Differential Expression Analysis: Identify differentially expressed genes.

3.5 Authorization

The department has authorized this study under the number OD.021/2425.

4. RESULTS

4.1 Quality Control of Raw Reads

The analysis began with quality control of raw RNA-seq reads using FastQC in Galaxy Europe, applied to all eight samples with default settings. It generated multiple quality metrics (Annex 2 shows results for sample 1) which were summarized with MultiQC to provide a combined view. From these metrics, three key indicators were selected for the Results section based on their relevance to overall read quality and suitability for downstream analysis.

Figure 5. Adapter Content Across Read Positions

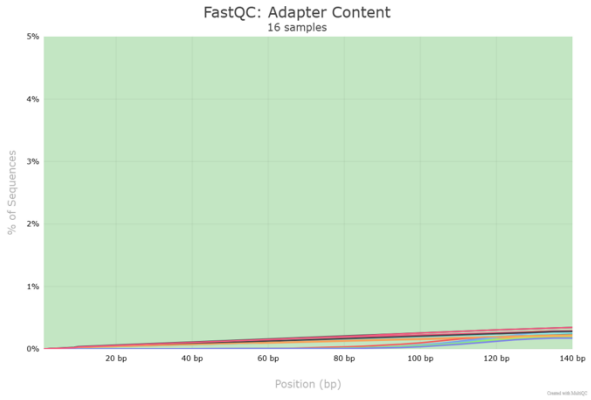


Figure 5 shows the adapter content across all samples. This plot assesses the presence of leftover adapter sequences, which can interfere with accurate alignment and quantification. The graph reveals that adapter contamination was minimal (<1% across all positions), remaining well within the green "pass" threshold, and thus no additional trimming steps were required.

Figure 6. Average Phred quality scores across base positions.

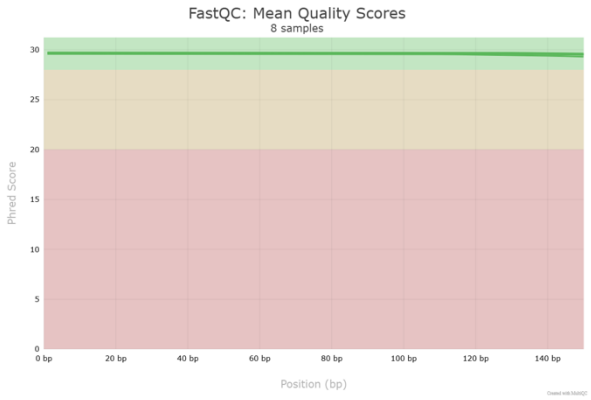


Figure 6 presents the per-base sequence quality, indicating the Phred score at each base position across all reads. The data show that nearly all base positions exhibit Phred scores above 30, falling within the green zone which reflect very high confidence in base calling- the process by

which each nucleotide (A, T, G, or C) is identified during sequencing. High scores indicate reliable read accuracy and low error rates.

Figure 7. Per-Sequence Quality Score Distribution

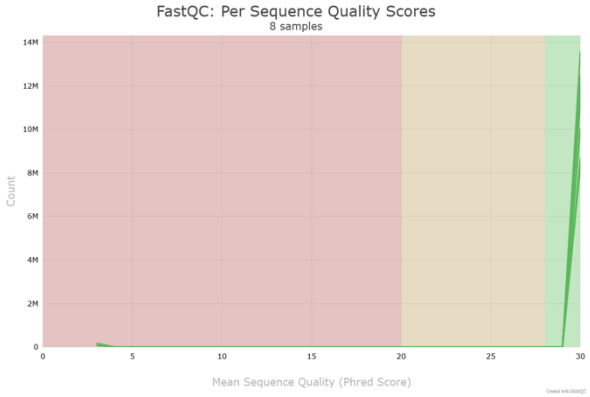


Figure 7 illustrates per-sequence quality score distribution, with most sequences clustered near the maximum score of 30 and very few below the commonly accepted threshold of 20. This confirms that the dataset is composed of consistently high-quality reads.

These three FastQC metrics were prioritized as they best represent sequencing quality. Additional metrics such as GC content, sequence duplication, and length distribution are provided in Annex 3.

HISAT2 then aligned paired-end reads to the *Homo sapiens* reference genome (GRCh38) using default settings, generating BAM files (Annex 4). Samtools Flagstat reported > 97% mapping and strong primary alignment metrics (Annex 5), indicated reliable mapping of reads to the reference genome and gene quantification.

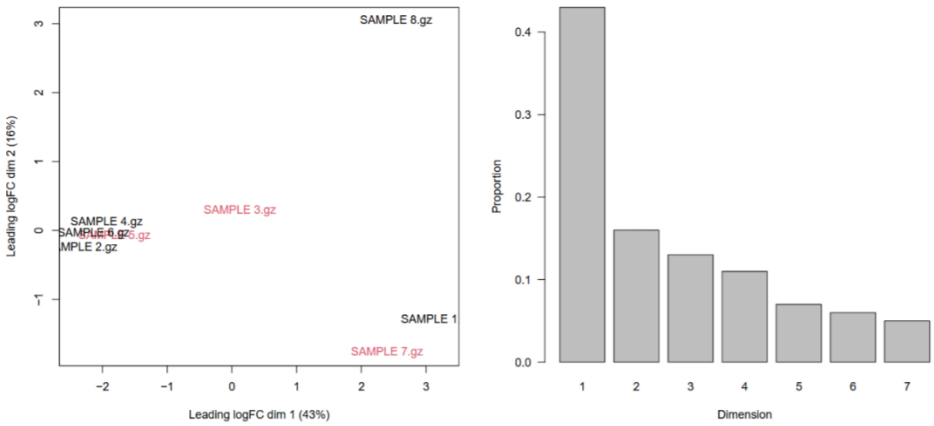
4.2 RNA-seq Reads to Counts

To generate the raw count matrix, FeatureCounts was used to assign aligned reads to genomic features based on the *Homo sapiens* GRCh38 annotation. Key options selected included counting at the gene level and specifying paired-end reads. The resulting count table, covering all eight samples, was validated using MultiQC to summarize mapping and counting statistics (Annex 6), completing the transition from raw reads to a structured count matrix (Annex 7).

A final quality check was then performed on the BAM files using a Galaxy workflow including Infer Experiment, MarkDuplicates, and Samtools IdxStats, to assess strandness, duplication, and read distribution across chromosomes (Annex 8).

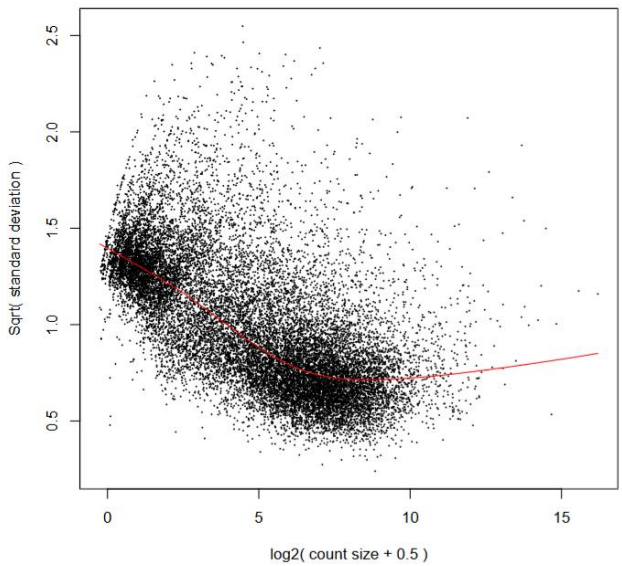
4.3 Differential Expression Analysis

Figure 8. Left: Multidimensional Scaling (MDS) Plot: Dims 1 and 2, Right: Scree Plot: Variance Explained



Differential expression analysis was performed using the RNA-seq counts to genes tutorial pipeline. The process began with data normalization and transformation using the Limma-Voom method- chosen for its robustness in handling small sample sizes. First, a Multi-dimensional Scaling (MDS) plot was generated to assess clustering based on gene expression patterns (Figure 8, left). While some samples cluster closely (e.g., healthy samples 2, 4, and 6), there is no consistent separation between tumor and healthy groups. Notably, Sample 8 (healthy) and Sample 7 (tumor) appear as outliers, and Sample 1 (healthy) overlaps with a tumor sample, suggesting potential within-group heterogeneity or technical variation. The scree plot (Figure 8, right) shows that Dimension 1 captures 43% of the variance, which supports that there are strong global expression differences in the dataset, although these may not align strictly with clinical groupings.

Figure 9. Voom Mean–Variance Trend Plot



The voom transformation was then applied, which models the mean-variance relationship of the log-counts (Figure 9). Each dot represents a gene, with the x-axis showing the log2-transformed count and the y-axis the square root of its standard deviation. The red trend line highlights that genes with lower expression levels exhibit higher variability, as indicated by the wider spread and elevated positions on the left side of the plot. In contrast, highly expressed genes show more consistent behavior, clustering lower along the Y-axis. This trend validates the voom transformation, which stabilizes variance before applying linear modeling in the Limma framework.

Figure 10. Final Model: Mean-variance trend (SA Plot)

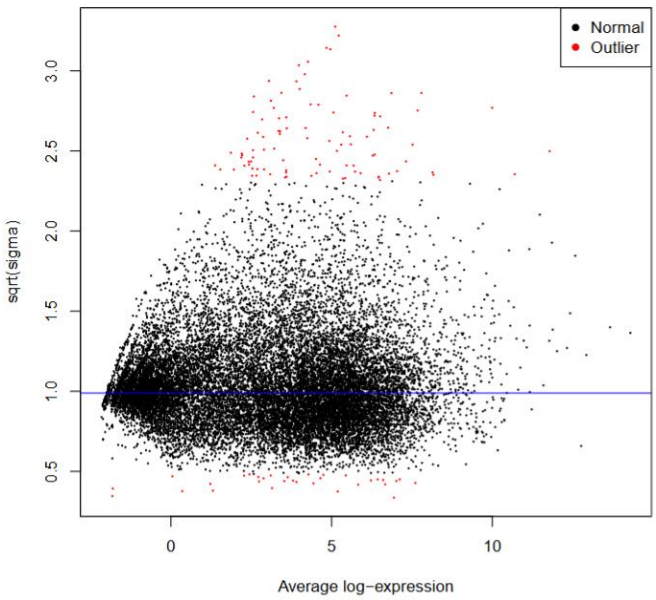
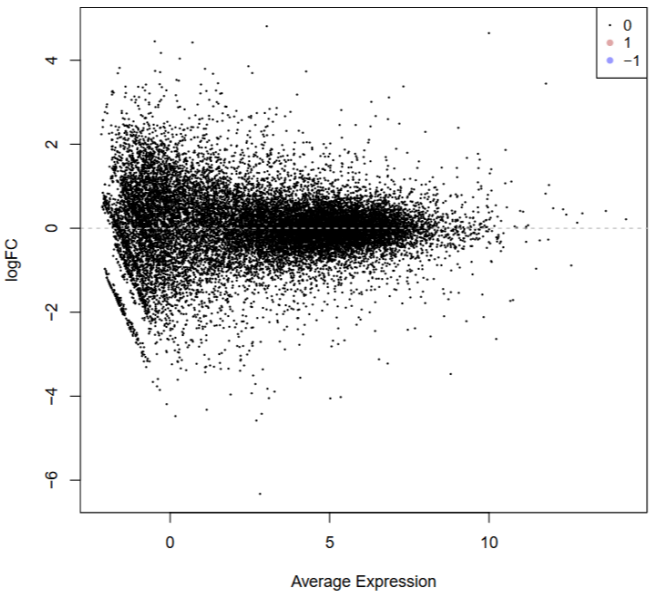
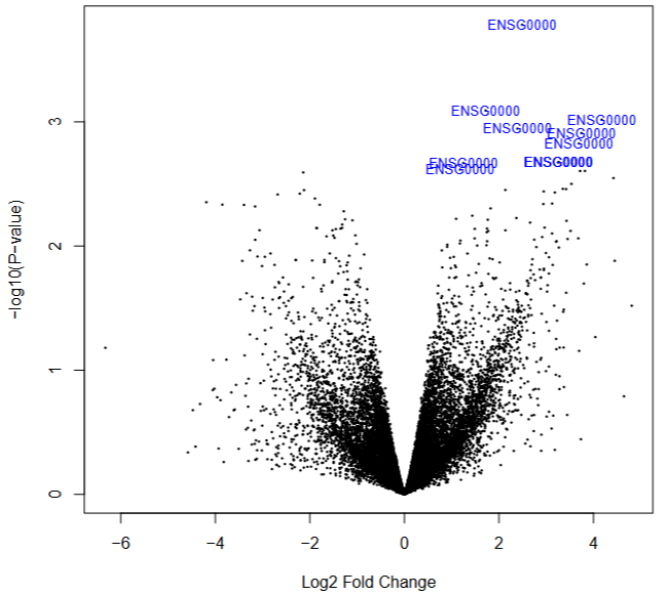


Figure 11. MD Plot: Tumor vs. Healthy



Following voom transformation, linear modeling and empirical Bayes moderation were applied using Limma. Model diagnostics were assessed with two plots. The SA plot (Figure 10) showed residual standard deviation (√σ) versus average log-expression. Most genes followed the fitted trend line, confirming consistent variance and good model fit, while a few outliers (in red) showed elevated variability. The MD plot (Figure 11) displayed logFC against average expression. While most genes clustered near logFC = 0, many deviated above or below. Genes above zero indicated upregulation in tumors, while those below were downregulated, reflecting widespread transcriptomic differences between conditions.

Figure 12. Volcano Plot: Tumor vs. Healthy



To visualize the statistical significance and magnitude of gene expression changes, a volcano plot was generated (Figure 12). This plot displays the log2 fold change on the x-axis and the -log10 p-value on the y-axis. Genes located in the upper left and right corners represent those with both large fold changes and strong statistical significance. These are the most biologically relevant DEGs.

Table 1. The top 20 most significantly dysregulated genes sorted by ascending p-value, with corresponding log2 fold change (logFC), t-statistics, and regulation direction.

Gene Name	logFC	P.Value	t	Regulation
SLC24A3	1.718471992	0.0008104398657	4.652902552	Upregulated
PPP2R2B	4.173101513	0.0009604055031	4.545086547	Upregulated
SPINK5	2.390760091	0.001134244251	4.440419008	Upregulated
CEL	3.745479349	0.001231130074	4.389196259	Upregulated
DHRS4L1	3.249599102	0.002122611731	4.054184972	Upregulated
UGT1A10	3.818255477	0.002484642848	3.958947977	Upregulated
ALAS2	3.716011705	0.002488094745	3.958111372	Upregulated
DNAJB5	-2.139921639	0.002555492604	-3.942015207	Downregulated
HRNR	4.422741465	0.002835787903	3.879515924	Upregulated
RSPO1	3.533321894	0.003159339204	3.814923215	Upregulated
TPSD1	3.421191257	0.003479390792	3.757474097	Upregulated
PLAUR	-2.122550785	0.00354414764	-3.746520222	Downregulated
GUCY2C	2.946596262	0.003633402886	3.73175836	Upregulated
FN1	-2.213959677	0.003785784917	-3.728221169	Downregulated
MEIS3	-2.680145526	0.003846556435	-3.697975783	Downregulated
FKBP10	-1.892531276	0.004141953588	-3.654237926	Downregulated
DOK5	-4.189461772	0.004438562346	-3.613459025	Downregulated
SRRM4	3.125108841	0.004550353602	3.598817247	Upregulated
KCNK2	-3.851426686	0.004642528749	-3.58702193	Downregulated
SPRY4	-1.791328536	0.004652367032	-3.585777294	Downregulated

After model fitting, a table of differentially expressed genes (DEGs) was generated (Annex 9). The table included log2 fold change (logFC), p-values, adjusted p-values, t-statistics, and gene symbols. To focus on the most statistically significant genes, the list was filtered to include genes with p-value < 0.05, and logFC ≥ 1.58 (upregulated) or ≤ -1.58 (downregulated). This logFC cutoff corresponds to at least a threefold difference in expression between tumor and healthy samples, a commonly accepted benchmark in transcriptomics to identify genes with meaningful biological impact. The column 13 of the table, which includes gene names, was used to discard any genes labeled "NA" (not mapped). The resulting DEG list was sorted by ascending p-value, and the 20 most significantly dysregulated genes were extracted (Table 3).

5. DISCUSION

The differential expression analysis highlighted the 20 most significant dysregulated genes in tumors versus healthy tissue, directly supporting the objective of identifying genes altered in oral cancer.

SLC24A3, or NCKX3, encodes a K^+-dependent Na^+/Ca^{2+} exchanger important for calcium-regulated processes like gene expression and apoptosis (30). Yu et al. suggest it may be a marker in OSCC by finding consistent SLC24A3 expression in all of the tested oral cancer cell lines (SCC4, SCC9, SCC15, SCC25, and CAL27), regardless of chemoresistance (31). Functional enrichment links it to DNA repair, mitochondrial organization, ncRNA metabolism, and the cell cycle, whilst another study has also shown the role of SLC24A3 in maintaining cell stability (32). This gene also appears to modulate the tumor-immune microenvironment and pain-related signaling. In cervical and endometrial cancers high expression predicts poorer predicted clinical outcomes (33). Although this evidence comes from other types of cancer, it raises the question of whether a similar association between expression levels and prognosis could exist in OSCC. Overall, SLC24A3 appears to have several roles in cancer development and symptoms, supporting its importance as a differentially expressed gene in OSCC.

PPP2R2B encodes the B55β regulatory subunit of protein phosphatase 2A (PP2A), a tumor suppressor that restrains cell proliferation (34). Multiple gene expression profiling studies found that PPP2R2B is significantly suppressed in OSCC tumors compared to normal oral tissue (35). PPP2R2B mRNA and protein levels also tend to be lower in HNSCC tumor samples and cell lines than in non-tumor controls (36). Whilst this downregulation reflects PPP2R2B's role as a tumor suppressor, the mechanism by which it occurs in oral cancer remains unclear. Though PPP2R2B's promoter is frequently hypermethylated in laryngeal squamous cell carcinoma, the same study found no evidence of PPP2R2B promoter hypermethylation in oral cancers (37). This distinction suggests that promoter methylation status of PPP2R2B varies by tumor site and that other mechanisms (e.g. transcriptional repression or deletions) are responsible for reduced expression of the gene in OSCC.

In OSCC, low PPP2R2B expression correlates with more aggressive disease features and resistance to chemotherapy. A bioinformatic analysis of OSCC cell lines revealed that cell lines with lower PPP2R2B levels were significantly more resistant to cisplatin, a leading chemotherapeutic for oral cancer (35). More specifically, Gouttia et al. demonstrate that PPP2R2B expression had one of the highest predictive values for cisplatin response. Low

PPP2R2B expression in combination with high MASTL kinase (an inhibitor of PP2A-B55) were strong predictors of a higher cisplatin IC50 (poorer drug sensitivity and worse response to chemotherapy) across OSCC cell lines.

In this study, PPP2R2B was significantly upregulated in oral cancer samples compared to healthy tissue, contrasting with its established role as a tumor suppressor in literature. This discrepancy suggests a context-dependent or stage-specific function for PPP2R2B in oral squamous cell carcinoma (OSCC). The observed upregulation may reflect a compensatory cellular response to oncogenic stress or altered signaling aimed at restoring PP2A balance. Alternatively, it may indicate a subtype of OSCC with distinct PP2A complex regulation. These findings highlight the need for further investigation into PPP2R2B expression in distinguishing tumors from healthy tissue in OSCC.

Extensive research identifies SPINK5 as a tumor suppressor frequently downregulated in OSCC and head and neck cancers. It encodes LEKTI, a serine protease inhibitor that regulates kallikrein activity and maintains epithelial barrier integrity (38). Loss of SPINK5 enhances protease-driven invasion, activates Wnt/β-catenin signaling, and promotes epithelial-mesenchymal transition (EMT), contributing to malignancy and chemoresistance (39). Epigenetic silencing through EHMT2 (G9a)-mediated histone modification and promoter methylation is linked to reduced expression (40). It has been proposed as a diagnostic and prognostic biomarker since clinically, low SPINK5 levels correlate with advanced stage, poor differentiation, and worse survival (41). Restoring SPINK5 suppresses tumor growth and enhances chemosensitivity in vitro(39).

Contrastingly, a recent spatial transcriptomics study identified high SPINK5 expression in an OSCC epithelial cell subtype named Epithelial01, particularly in carcinoma in situ and early-stage lesions (42). Such heterogeneity may explain the discrepancy between existing literature and the current findings since SPINK5 was significantly upregulated in tumor samples. Moreover, it implies expression may vary by tumor subpopulation, differentiation state, or disease stage. Further study is required to determine if SPINK5 could have dual or context-dependent roles- as a tumor suppressor in advanced disease and retained in early lesions.

Part of the DHRS4 cluster, little is currently known about DHRS4L1's structure or expression (43). There are no published studies directly linking DHRS4L1 expression to outcomes in OSCC or head and neck cancers. However, DHRS4 is expressed in multiple tissues and cancer cell lines. DHRS4 encodes a NADP(H)-dependent oxidoreductase involved in retinol and steroid metabolism,

contributing to the production of all-trans retinoic acid (RA) which is a key regulator of cell growth and differentiation. Since RA pathways control oral epithelial differentiation and proliferation, changes in DHRS4L1 may hold diagnostic or prognostic value. The DHRS4L1 cluster is involved in the retinol-to-RA and steroid metabolism pathways, which influences tumor behavior by activating nuclear receptors like RAR and RXR, promoting differentiation and apoptosis in epithelial cells (44,45). Disruptions in this pathway can influence cancer development.

Although HRNR is not well-characterized in oral cancer, it has been linked to tumor progression in several epithelial malignancies, including gastric cancer and hepatocellular carcinoma. In gastric cancer, high HRNR expression in stage II and III tumors is associated with significantly worse overall survival and serves as an independent prognostic marker (5-year OS: 53.6% vs. 74.9%; HR = 1.53) (46). In liver cancer, HRNR promotes tumor progression via activation of the AKT signaling pathway (47). As a member of the S100 protein family, HRNR may also influence epithelial differentiation and stress responses through calcium-dependent mechanisms (48). These findings show that HRNR contributes to epithelial tumor biology across tissue types and may hold prognostic value. Its upregulation in OSCC highlights the possible involvement of both epithelial-specific and immune-related pathways in oral cancer development.

RSPO1 potentiates Wnt/β-catenin signaling by blocking the breakdown of key Wnt pathway components (49), playing an important role in cancer by helping tumors reprogram their metabolism via glycolysis, glutamine use, fat production (50). While there is limited research on RSPO1 in OSCC, studies in other cancers show that dysregulated RSPO1 expression is linked to changes in immune cell activity, suggesting it influences the tumor immune environment (51). In head and neck cancers, single-cell RNA sequencing found a group of epithelial cells with high RSPO1 expression and strong tumor-forming ability (52). In gastrointestinal cancers, RSPO1 overexpression showed promotion of cell growth, movement, and survival, mainly by activating the Wnt pathway (53). In the DEG table, RSPO1 was significantly upregulated in tumor samples, which supports the idea that RSPO1 may drive metabolic changes and tumor progression in oral cancer.

DOK5 encodes a docking protein involved in signal transduction, particularly the MAPK and Wnt/β-catenin signaling pathways, and plays a role in cell proliferation and differentiation (54,55). In gastric cancer, high DOK5 expression was linked to increased immune cell infiltration

and poor prognosis (56). The downregulation of DOK5 in oral cancer samples may reflect tissue-specific functions or differences in immune regulation.

KCNK2 (also known as TREK-1) encodes a potassium channel involved in membrane potential regulation, neuronal signaling, and cellular stress response (57). In papillary thyroid carcinoma, KCNK2 was found to be downregulated, and its expression negatively correlated with tumor stage, suggesting a possible tumor-suppressive role (58). In contrast, data from the Human Protein Atlas shows that KCNK2 is upregulated in breast cancer and classified as cancer-enhanced, underlining the context-dependent role of this gene. Despite these observations, KCNK2 currently lacks consistent prognostic value across cancer types, as its expression does not reliably correlate with patient outcomes (59). These findings underscore the complexity of gene regulation in cancer, where genes like DOK5 and KCNK2 may have distinct roles depending on tissue type and tumor context.

Among the DEGs identified in this study, SPRY4 and UGT1A10 are mentioned in previous cancer research, including limited findings in oral squamous cell carcinoma. One recent study is referred to as “the first to confer the potential involvement of SPRY4 protein expression in human oral squamous cell carcinogenesis” (60), while UGT1A10, a detoxification enzyme involved in glucuronidation, has been shown to be dysregulated in several cancers, with its overexpression potentially reflecting metabolic adaptation in tumor cells (61,62). ALAS2, a mitochondrial enzyme that catalyzes the first step in heme biosynthesis, is significantly upregulated in the present dataset which according to studies has been shown to reduce oxidative stress and protect against ferroptosis in non-erythroid cells (63). This suggests that tumors may exploit ALAS2 to enhance metabolic resilience and survival. Conversely, DNAJB5, a member of the HSP40 chaperone family, is downregulated in our dataset. While it has been linked to cell survival and therapy resistance in other cancers, its reduced expression in OSCC may impair stress response pathways and protein stability (64). Whilst these observations reflect alignment with emerging evidence, current available literature concerning these genes remains limited, so it does not yet provide a sufficient basis for a broader discussion.

In contrast, TPSD1, CEL, FKBP10, GUCY2C, MEIS3, and SRRM4 appear to be novel findings in the context of OSCC, as there is little to no existing literature linking them directly. These observations highlight the identification of potentially new molecular players and combination of underexplored pathways that may play roles in OSCC progression and warrant further investigation.

By identifying dysregulated genes, this study contributes to a clearer understanding of the molecular mechanisms involved in tumor progression. While earlier research has focused largely on well-known oncogenes and tumor suppressors, this study highlights both established players (e.g., PPP2R2B, SPINK5) and novel candidates (TPSD1, SRRM4). These findings deepen our understanding of oral-cancer regulation and guide improved profiling as well as future diagnostic and therapeutic efforts.

Despite these insights, several limitations remain. Small sample size limits generalizability and likely omits oral cancer's full variability. Relying on pre-existing RNA-seq data with only species, sex, and provider metadata prevented analysis of race, exposures, habits, and tumor stage. This limited the ability to assess how expression changes relate to oral cancer staging. Additionally, though the DEG list aligns with published findings, many transcripts are still poorly characterized and warrant further study.

Technically, Galaxy workflows use default settings and standard gene filters, which may overlook subtle or novel expression changes. Future work with larger datasets and customizable pipelines could build on these findings and offer a more complete understanding of gene expression in oral cancer.

This study lays a foundation for follow-up research using diverse cohorts, detailed clinical metadata, and experimental validation (in vitro or in vivo). Integrating proteomics or epigenetics could offer a more comprehensive view of gene function. Overall, this research underscores the value of bioinformatics in cancer genomics and supports ongoing transcriptomic analysis in both research and clinical contexts.

6. CONCLUSIONS

1. This study successfully identified differentially expressed genes (DEGs) between oral cancer tumor cells and healthy tissue using RNA-Seq data. 20 statistically significantly dysregulated genes were identified specifically.
2. As a complementary conclusion, this study revealed several biological pathways in which these DEGs are involved, including Wnt/β-catenin signaling, metabolic reprogramming and redox homeostasis, and immune system regulation.
3. As a complementary conclusion, this study also emphasizes the value of bioinformatic workflows in oncological research. The use of Galaxy Europe enabled a reproducible, accessible, and efficient analysis pipeline for transcriptomic data.

7. SUSTAINABILITY

Integration of bioinformatics into biomedical research contributes significantly to sustainability by reducing reliance on physical materials, lab reagents, and animal models, aligning with SDG 12: Responsible Consumption and Production, and SDG 15: Life on Land. Enabling in-silico experimentation allows researchers to pre-screen and prioritize only the most promising targets for in vitro validation, minimizing waste and resource use. This approach reduces the environmental footprint of research activities, contributing to SDG 13: Climate Action.

Economically, bioinformatic pipelines accelerate data analysis, reducing the time and cost of discovery phases and allowing research funds to be allocated more efficiently. This facilitates more agile responses to public health challenges, aligning with SDG 3: Good Health and Well-being.

Socially, free open-source platforms democratize participation in scientific research, promoting equity and inclusion across global research communities by reducing technical and financial barriers, bioinformatics supports broader collaboration and capacity building, particularly in low-resource settings. This contributes to SDG 10: Reduced Inequalities.

8. REFERENCES

1. World Health Organization. Global strategy and action plan on oral health 2023–2030 [Internet]. 2024. Available from: https://iris.who.int/bitstream/handle/10665/376623/9789240090538-eng.pdf?sequence=1

2. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol. 2009 Apr;45(4–5):309–16.

3. Rivera C. Essentials of oral cancer. Int J Clin Exp Pathol. 2015;8(9):11884–94.

4. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015 Mar;65(2):87–108.

5. World Health Organization. World regions according to the World Health Organization [Internet]. 2017. Available from: https://ourworldindata.org/grapher/who-regions

6. Valent P, Akin C, Arock M, Bock C, George TI, Galli SJ, et al. Proposed Terminology and Classification of Pre-Malignant Neoplastic Conditions: A Consensus Proposal. EBioMedicine. 2017 Dec;26:17–24.

7. Cancer Research UK. Types of mouth and oropharyngeal cancers [Internet]. 2024. Available from: https://www.cancerresearchuk.org/about-cancer/mouth-cancer/stages-types-grades/types-grades

8. Lingen MW, Kalmar JR, Karrison T, Speight PM. Critical evaluation of diagnostic aids for the detection of oral cancer. Oral Oncol. 2008 Jan;44(1):10–22.

9. Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. Oral Oncol. 2010 Jun;46(6):414–7.

10. Langlais RP, Miller CS, editors. Color atlas of common oral diseases. Philadelphia: Lea & Febiger; 1992. 167 p.

11. Valent P, Akin C, Arock M, Bock C, George TI, Galli SJ, et al. Proposed Terminology and Classification of Pre-Malignant Neoplastic Conditions: A Consensus Proposal. EBioMedicine. 2017 Dec;26:17–24.

12. Woolgar JA, Rogers SN, Lowe D, Brown JS, Vaughan ED. Cervical lymph node metastasis in oral cancer: the importance of even microscopic extracapsular spread. Oral Oncol. 2003 Feb;39(2):130–7.

13. Thompson LDR. World Health Organization Classification of Tumours: Pathology and Genetics of Head and Neck Tumours. Ear Nose Throat J. 2006 Feb;85(2):74–74.

14. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol. 89. Lyon France; 2007.

15. Rock LD, Rosin MP, Zhang L, Chan B, Shariati B, Laronde DM. Characterization of epithelial oral dysplasia in non-smokers: First steps towards precision medicine. Oral Oncol. 2018 Mar;78:119–25.

16. Al Moustafa AE, editor. Development of Oral Cancer [Internet]. Cham: Springer International Publishing; 2017 [cited 2024 Nov 13]. Available from: http://link.springer.com/10.1007/978-3-319-48054-1

17. Irani S. New insights into oral cancer—Risk factors and prevention: A review of literature. Int J Prev Med. 2020;11(1):202.

18. Aghiorghiesei O, Zanoaga O, Nutu A, Braicu C, Campian RS, Lucaciu O, et al. The World of Oral Cancer and Its Risk Factors Viewed from the Aspect of MicroRNA Expression Patterns. Genes. 2022 Mar 26;13(4):594.

19. Morris Brown L, Gridley G, Diehl SR, Winn DM, Harty LC, Otero EB, et al. Family cancer history and susceptibility to oral carcinoma in Puerto Rico. Cancer. 2001 Oct 15;92(8):2102–8.

20. Garavello W, Foschi R, Talamini R, La Vecchia C, Rossi M, Dal Maso L, et al. Family history and the risk of oral and pharyngeal cancer. Int J Cancer. 2008 Apr 15;122(8):1827–31.

21. Radoï L, Paget-Bailly S, Guida F, Cyr D, Menvielle G, Schmaus A, et al. Family history of cancer, personal history of medical conditions and risk of oral cavity cancer in France: the ICARE study. BMC Cancer. 2013 Dec;13(1):560.

22. Jurel S, Gupta D, Singh RaghuwarD, Singh M, Srivastava S. Genes and oral cancer. Indian J Hum Genet. 2014;20(1):4.

23. Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends Genet TIG. 1993 Apr;9(4):138–41.

24. McGettigan PA. Transcriptomics in the RNA-seq era. Curr Opin Chem Biol. 2013 Feb;17(1):4–11.

25. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. Bioinforma Biol Insights. 2015;9(Suppl 1):29–46.

26. Bioinformatics Definition. In: Talking Glossary of Genomic and Genetic Terms [Internet]. 2024. Available from: https://www.genome.gov/genetics-glossary/Bioinformatics

27. Clark AJ, Lillard JW. A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology. Genes. 2024 Aug 6;15(8):1036.

28. Mount DW. Bioinformatics: Sequence and Genome Analysis. Woodbury: Cold Spring Harbor Laboratory Press; 2001. 1 p.

29. Bayat A. Science, medicine, and the future: Bioinformatics. BMJ. 2002 Apr 27;324(7344):1018–22.

30. National Center for Biotechnology Information. SLC24A3 solute carrier family 24 member 3 [Homo sapiens (human) [Internet]. 2025 [cited 2025 Apr 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/57419?utm_source

31. Yu B, Kruse N, Howard KM, Kingsley K. Downstream Target Analysis for miR-365 among Oral Squamous Cell Carcinomas Reveals Differential Associations with Chemoresistance. Life. 2024 Jun 10;14(6):741.

32. Zhou S, Jin Q, Yao H, Ying J, Tian L, Jiang X, et al. Pain-Related Gene Solute Carrier Family 24 Member 3 Is a Prognostic Biomarker and Correlated with Immune Infiltrates in Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma: A Study via Integrated Bioinformatics Analyses and Experimental Verification. Wang J, editor. Comput Math Methods Med. 2023 Jan;2023(1):4164232.

33. Yu SH, Cai JH, Chen DL, Liao SH, Lin YZ, Chung YT, et al. LASSO and Bioinformatics Analysis in the Identification of Key Genes for Prognostic Genes of Gynecologic Cancer. J Pers Med. 2021 Nov 11;11(11):1177.

34. National Center for Biotechnology Information. PPP2R2B protein phosphatase 2 regulatory subunit Bbeta [Homo sapiens (human)] [Internet]. 2025 [cited 2025 Apr 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/5521

35. Gouttia OG, Zhao J, Li Y, Zwiener MJ, Wang L, Oakley GG, et al. The MASTL-ENSA-PP2A/B55 axis modulates cisplatin resistance in oral squamous cell carcinoma. Front Cell Dev Biol. 2022 Sep 28;10:904719.

36. Suyarnsestakorn C, Thanasupawat T, Leelahavanichkul K, Gutkind JS, Mutirangura A. Ataxia telangiectasia mutated nuclear localization in head and neck cancer cells is PPP2R2B-dependent. Asian Biomed. 2010 Jun 1;4(3):373–83.

37. Paluszczak J, Sarbak J, Kostrzewska-Poczekaj M, Kiwerska K, Jarmuż-Szymczak M, Grenman R, et al. The negative regulators of Wnt pathway—DACH1, DKK1, and WIF1 are methylated in oral and oropharyngeal cancer and WIF1 methylation predicts shorter survival. Tumor Biol. 2015 Apr;36(4):2855–61.

38. National Center for Biotechnology Information. SPINK5 serine peptidase inhibitor Kazal type 5 [Homo sapiens (human)] [Internet]. 2025. Available from: https://www.ncbi.nlm.nih.gov/gene/11005

39. Guo E, Mao X, Wang X, Guo L, An C, Zhang C, et al. Alternatively spliced ANLN isoforms synergistically contribute to the progression of head and neck squamous cell carcinoma. Cell Death Dis. 2021 Aug 3;12(8):764.

40. Sun S, Su G, Zheng X. Inhibition of the Tumor Suppressor Gene SPINK5 via EHMT2 Induces the Oral Squamous Cell Carcinoma Development. Mol Biotechnol. 2024 Feb;66(2):208–21.

41. Zhao C, Zou H, Zhang J, Wang J, Liu H. An integrated methylation and gene expression microarray analysis reveals significant prognostic biomarkers in oral squamous cell carcinoma. Oncol Rep [Internet]. 2018 Sep 12 [cited 2025 Apr 23]; Available from: http://www.spandidos-publications.com/10.3892/or.2018.6702

42. Zhang H, Qian Y, Zhang Y, Zhou X, Shen S, Li J, et al. Multi-omics analysis deciphers intercellular communication regulating oxidative stress to promote oral squamous cell carcinoma progression. Npj Precis Oncol. 2024 Nov 21;8(1):272.

43. Su ZJ, Zhang QX, Liu GF, Song XH, Li Q, Wang RJ, et al. Bioinformatic analysis of the human DHRS4 gene cluster and a proposed mechanism for its transcriptional regulation. BMC Mol Biol. 2010;11(1):43.

44. Wang C, Wang G, Zhang Z, Wang Z, Ren M, Wang X, et al. The downregulated long noncoding RNA DHRS4-AS1 is protumoral and associated with the prognosis of clear cell renal cell carcinoma. OncoTargets Ther. 2018 Sep;Volume 11:5631–46.

45. Gabrielli F, Tofanelli S. Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes. Gene. 2012 Dec;511(2):461–9.

46. Oshima T, Hashimoto I, Hiroshima Y, Kimura Y, Tanabe M, Nakayama Y, et al. Clinical Significance of *HRNR* Expression in Patients With Stage II/III Gastric Cancer After Curative Gastrectomy. Anticancer Res. 2024 Oct;44(10):4579–84.

47. Fu SJ, Shen SL, Li SQ, Hua YP, Hu WJ, Guo B, et al. Hornerin promotes tumor progression and is associated with poor prognosis in hepatocellular carcinoma. BMC Cancer. 2018 Dec;18(1):815.

48. Dessen P. HRNR (hornerin) [Internet]. 2013. Available from: https://atlasgeneticsoncology.org/gene/53156/#section-3

49. Binnerts ME, Kim KA, Bright JM, Patel SM, Tran K, Zhou M, et al. R-Spondin1 regulates Wnt signaling by inhibiting internalization of LRP6. Proc Natl Acad Sci. 2007 Sep 11;104(37):14700–5.

50. Ter Steege EJ, Bakker ERM. The role of R-spondin proteins in cancer biology. Oncogene. 2021 Nov 25;40(47):6469–78.

51. Lou X. Prognostic and immunological roles of RSPO1 in pan-cancer and its correlation with LUAD proliferation and metastasis. Am J Cancer Res. 2024;14(8):3800–15.

52. Liu ZL, Meng XY, Bao RJ, Shen MY, Sun JJ, Chen WD, et al. Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer. Nat Commun. 2024 Mar 22;15(1):2595.

53. He Z, Zhang J, Ma J, Zhao L, Jin X, Li H. R-spondin family biology and emerging linkages to cancer. Ann Med. 2023 Dec 12;55(1):428–46.

54. National Center for Biotechnology Information. DOK5 docking protein 5 [Homo sapiens (human)] (Gene ID 55816) [Internet]. 2025. Available from: https://www.ncbi.nlm.nih.gov/gene/55816

55. Xu L, Wu J, Yu Y, Li H, Sun S, Zhang T, et al. Dok5 regulates proliferation and differentiation of osteoblast via canonical Wnt/β-catenin signaling. J Musculoskelet Neuronal Interact. 2022 Mar 1;22(1):113–22.

56. Luo F, Wang Z, Chen S, Luo Z, Wang G, Yang H, et al. DOK5 as a Prognostic Biomarker of Gastric Cancer Immunoinvasion: A Bioinformatics Analysis. Raju N, editor. BioMed Res Int. 2022 Jan;2022(1):9914778.

57. National Center for Biotechnology Information. KCNK2 potassium two pore domain channel subfamily K member 2 [Homo sapiens (human)] [Internet]. 2025. Available from: https://www.ncbi.nlm.nih.gov/gene/3776

58. Lin X, Wu JF, Wang DM, Zhang J, Zhang WJ, Xue G. The correlation and role analysis of KCNK2/4/5/15 in Human Papillary Thyroid Carcinoma microenvironment. J Cancer. 2020;11(17):5162–76.

59. Uhlén M et al." (or "The Human Protein Atlas Team. The Human Protein Atlas. KCNK2. Available from: https://www.proteinatlas.org/ENSG00000082482-KCNK2/cancer?

60. Liao PH, Chuang FH, Wang YY, Wang WC, Su CW, Hsu CW, et al. Sprouty 4 expression in human oral squamous cell carcinogenesis. J Dent Sci. 2023 Apr;18(2):781–90.

61. Elahi A, Bendaly J, Zheng Z, Muscat JE, Richie JP, Schantz SP, et al. Detection of *UGT1A10* polymorphisms and their association with orolaryngeal carcinoma risk. Cancer. 2003 Aug 15;98(4):872–80.

62. Balliet RM, Chen G, Dellinger RW, Lazarus P. UDP-Glucuronosyltransferase 1A10: Activity against the Tobacco-Specific Nitrosamine, 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanol, and a Potential Role for a Novel UGT1A10 Promoter Deletion Polymorphism in Cancer Susceptibility. Drug Metab Dispos. 2010 Mar;38(3):484–90.

63. He Y, Wang X, Li D, Zhu Q, Xiang Y, He Y, et al. ALAS2 overexpression alleviates oxidative stress-induced ferroptosis in aortic aneurysms via GATA1 activation. J Thorac Dis. 2024 Apr;16(4):2510–27.

64. Kim HY, Hong S. Multi-Faceted Roles of DNAJB Protein in Cancer Metastasis and Clinical Implications. Int J Mol Sci. 2022 Nov 29;23(23):14970.

65. Alsanie I, Rajab S, Cottom H, Adegun O, Agarwal R, Jay A, et al. Distribution and Frequency of Salivary Gland Tumours: An International Multicenter Study. Head Neck Pathol. 2022 May 27;16(4):1043–54.

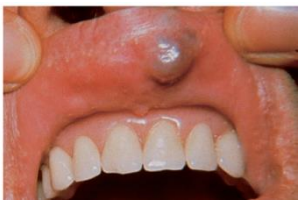
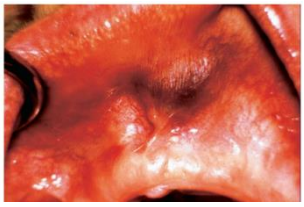
66. Elmas ÖF, Metin MS, Kilitçi A. Dermoscopic Features of Lower Lip Squamous Cell Carcinoma: A Descriptive Study. Indian Dermatol Online J. 2019;10(5):536–41.

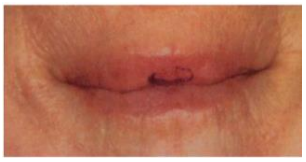
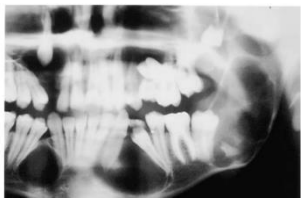
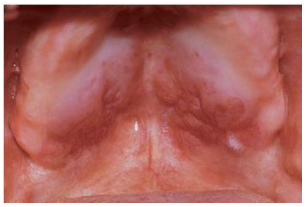
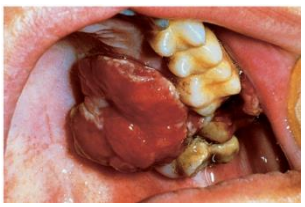
67. Waldt S. Primary and Secondary Bone Tumors. In: Vogl TJ, Reith W, Rummeny EJ, editors. Diagnostic and Interventional Radiology [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016 [cited 2024 Nov 13]. p. 1007–32. Available from: https://link.springer.com/10.1007/978-3-662-44037-7_36

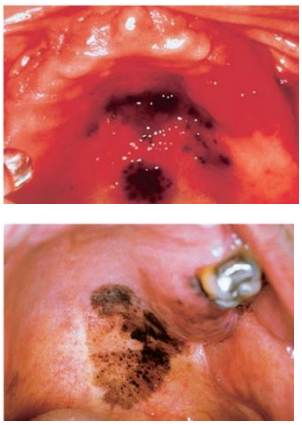
68. Yamaguchi S, Nagasawa H, Suzuki T, Fujii E, Iwaki H, Takagi M, et al. Sarcomas of the oral and maxillofacial region: a review of 32 cases in 25 years. Clin Oral Investig [Internet]. 2004 Jun [cited 2024 Nov 13];8(2). Available from: http://link.springer.com/10.1007/s00784-003-0233-4

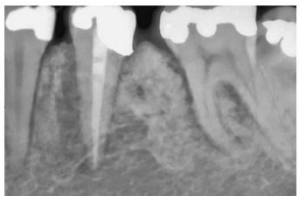
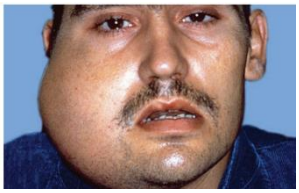
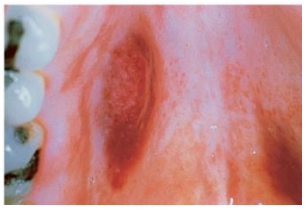
9. ANNEXES

ANNEX 1. Overview of other Oral Cancer Subtypes: Descriptions, Histology, and Clinicopathological Features

Oral Cancer	Description	Histology	Clinicopathological aspect
Salivary gland cancer	The salivary glands comprise three pairs major salivary glands (parotid, submandibular, and sublingual), as well as hundreds of minor salivary glands. The most common type of salivary gland cancer is mucoepidermoid carcinomas- most often starting in the parotid glands (65). However, the incidence of malignancy is higher in the sublingual and submandibular glands since around 70-90% and 45% of tumours in the respective glands are cancer in comparison to 15-35% in parotid glands (7,65).	Salivary gland cancers show various tumor types since a healthy salivary gland contains inner luminal/epithelial or acinar/mucous cells and outer basal/myoepithelial cells in the duct or the secretory part (13).	**Figure 3.** Pleomorphic adenoma: firm bluish nodule, Canalicular adenoma: purplish labial nodule (in order).  

Basal Cell Carcinoma (BCC)	BCC is a type of skin cancer. It usually presents with ulceration or bleeding, therefore caution should be taken when performing the differential diagnosis as not to confuse with Herpes Simplex, Aphthous ulcer, Actinic cheilitis, traumatic lesions to name a few (10,66).	Developing from basal cells found on the lips.	**Figure 4.** Basal cell carcinoma on vermilion lip of 45-year-old woman, Nevus basal cell carcinoma with multiple odontogenic keratocysts, which is a hallmark feature of the syndrome (in order).  
Lymphoma	Lymphomas are the 2nd most common neoplasm of the head and neck, but are relatively rare within the scope of oral malignancies- accounting for around 3.5% (10). They are usually classified into Hodgkin or non-Hodgkin lymphoma and subdivided into nodal and extra-nodal disease. Extra-nodal lymphomas often present in the oral region, usually in the masticatory mucosa. These lymphomas are not always primary, but rather secondary tumours invading from surrounding	Lymphoma is a malignant neoplastic growth of lymphocytes.	**Figure 5.** Lymphoma of palate: nontender with telangiectasia, HIV-associated non-Hodgkin lymphoma (in order)  

	structures such as maxillary sinus or bone marrow (67). Those most prone to primary lymphomas of the palate are young AIDS patients or adults over 60 years of age (10).		
Melanoma	Oral malignant melanoma (OMM) is very rare but highly aggressive, making up 0.5% of all oral malignancies and <1% of all other melanomas, 80% of which occur on the palate or maxillary alveolar ridge (10).	During embryogenesis, melanocytes arise from neural-crest precursors that migrate into and reside in the basal layer of the epithelium. Melanoma is a malignancy of these epidermal melanocytes.	**Figure 6.** Melanoma satellite lesions on palate, Melanoma color variation in soft palate and tuberosity (in order). 

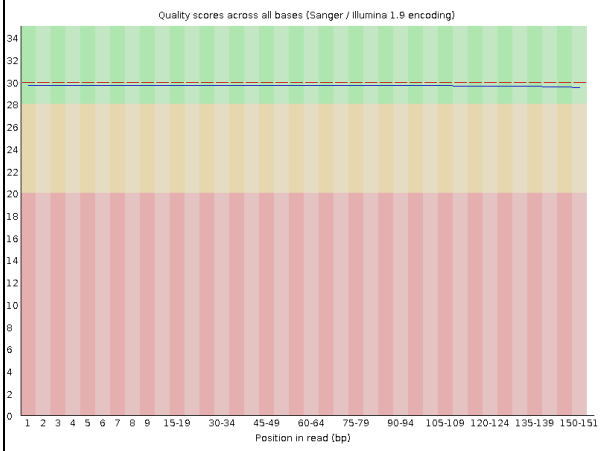
Sarcoma	Oral sarcomas are very rare <1% , as sarcomas account for nearly 1% of all neoplasms in the head and neck region (68). It can include malignant periodontal defects, osteosarcoma, chondrosarcoma, Ewing sarcoma (cell malignancy caused by a chromosomal translocation), Kaposi's sarcoma (associated to HIV/IAIDs infection) (10).	Derived from mesenchymal progenitor cells.	**Figure 7.** Malignant disease: chondrosarcoma with widened PDL, Ewing sarcoma, HIV-associated Kaposi Sarcoma: purplish macules (in order).   

ANNEX 2. FASTQC on sample 1 (webpage)

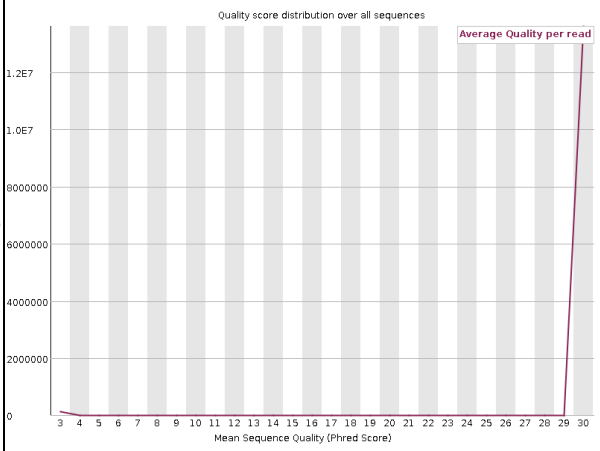
Overview information

Measure	Value
Filename	SAMPLE 1_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	13752018
Total Bases	1.9 Gbp
Sequences flagged as poor quality	0
Sequence length	35-151
%GC	53

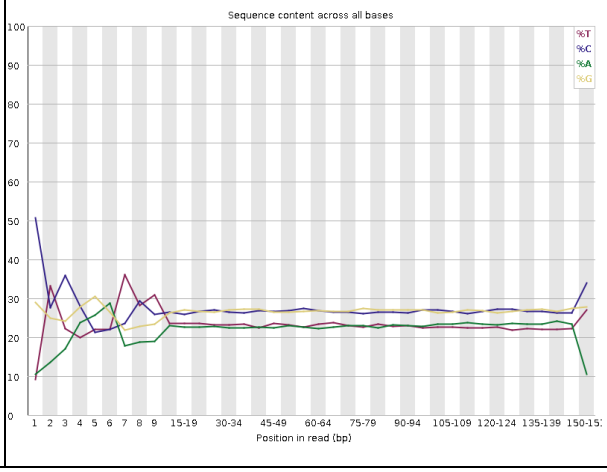
Per Base Sequence quality

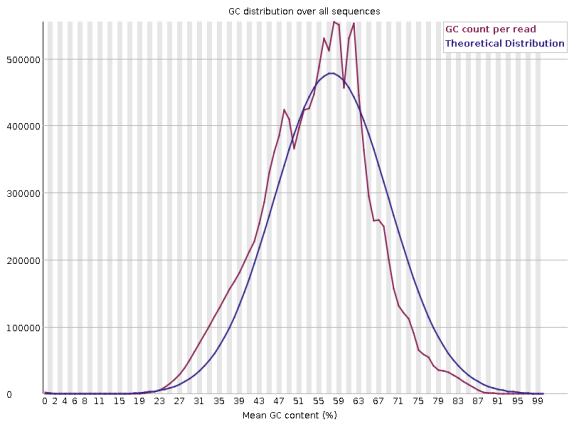
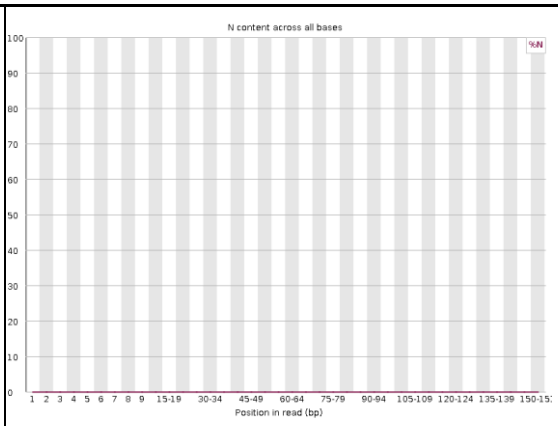
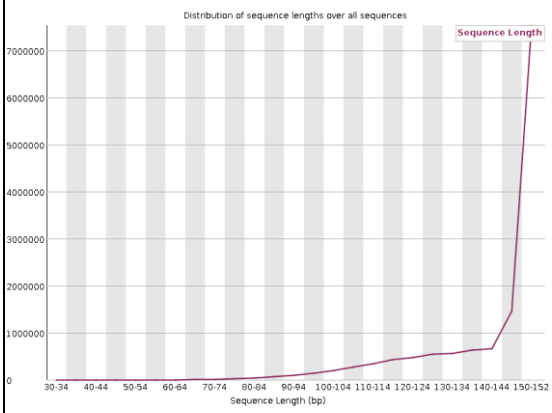
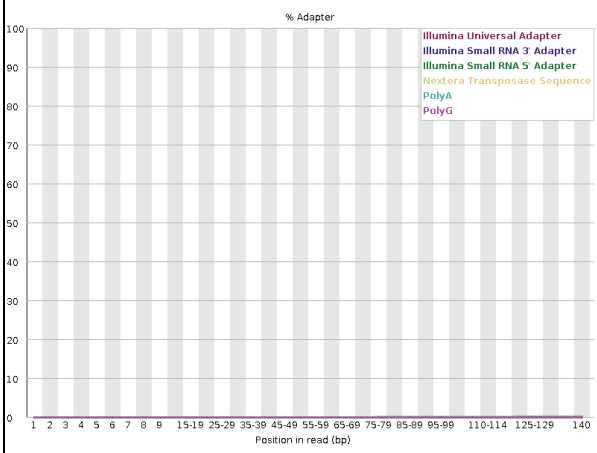


Per Sequence Quality Score

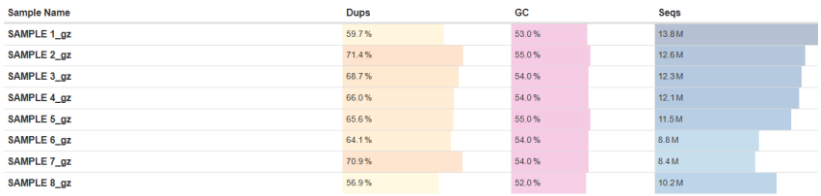
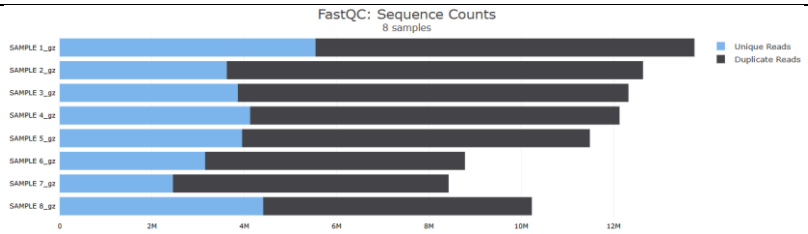
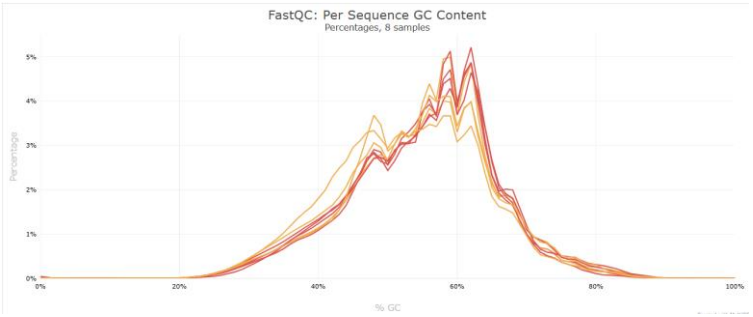
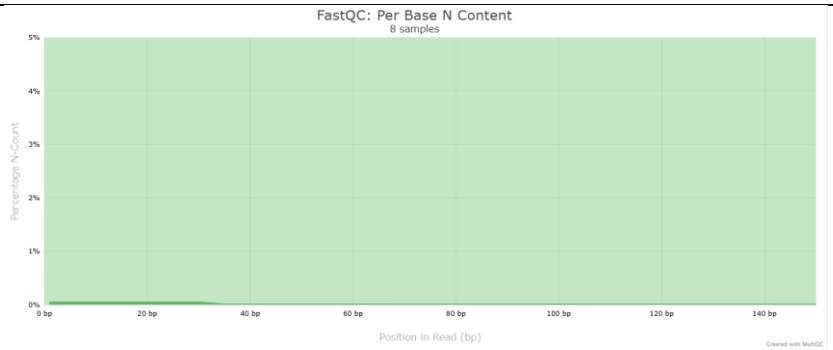


Per Base Sequence content

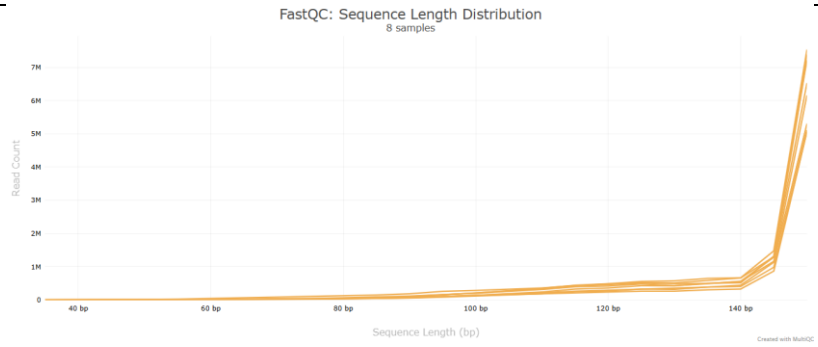


Per Sequence GC content	
Per Base N content	
Sequence length distribution	
Adapter content	

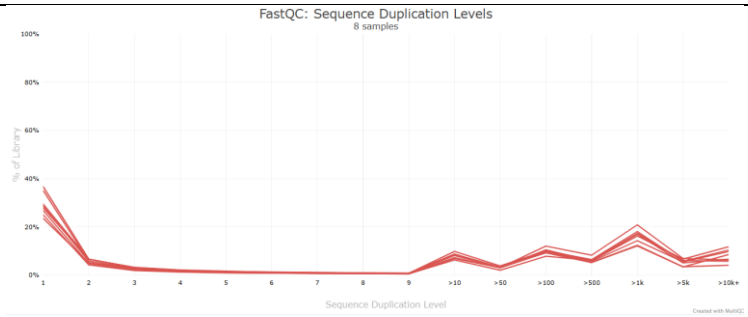
ANNEX 3. MULTIQC data (webpage)

General Statistics	
Sequence counts for each sample: Sequence counts for each sample. Duplicate read counts are an estimate only.	
Per Sequence GC Content: The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content.	
Per Base N Content: The percentage of base calls at each position for which an N was called.	

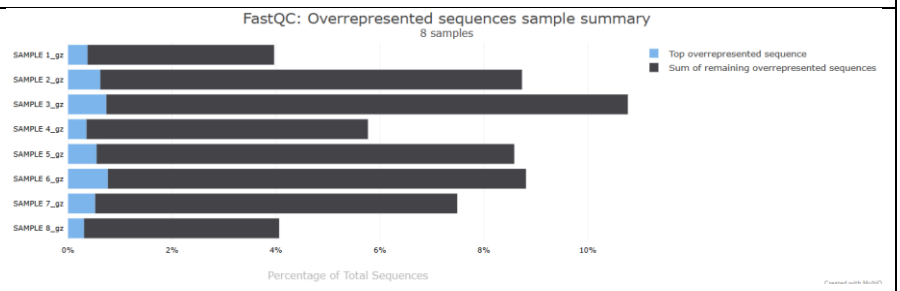
Sequence Length Distribution: The distribution of fragment sizes (read lengths) found.



Sequence Duplication Levels: The relative level of duplication found for every sequence.

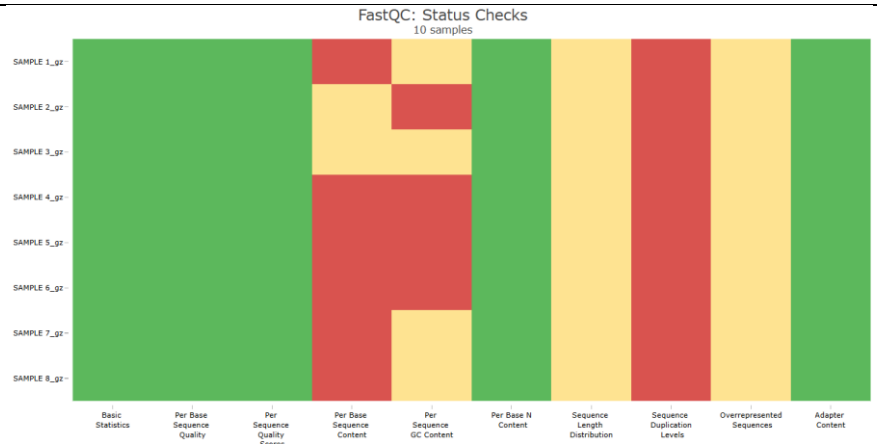


Overrepresented sequences by sample: The total amount of overrepresented sequences found in each library.



Top overrepresented sequences across all samples: The table shows 20 most overrepresented sequences across all

Overrepresented sequence	Reports	Occurrences	% of all reads
CTGTATTGTTATTTTTCGTCACTACCTCCCCGGGTCGGGAGTGGGTAATT	8	452 485	0.5041 %
CCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTGCCTA	8	312 843	0.3485 %
CGACGACCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCG	8	389 428	0.4338 %
CCCGAAGTTACGGATCCGGCTTGCCGACTTCCCTTACCTACATTGTTCCA	8	160 485	0.1788 %
CTTTAAATGGGTAAGAAGCCCGGCTCGCTGGCGTGGAGCCGGGCGTGGAA	8	181 280	0.2020 %
CTCCCGTCCACTCTCGACTGCCGGCGACGGCCGGGTATGGGCCCGACGCT	8	209 170	0.2330 %
CTGGGGTCTGATGAGCGTCGGCATCGGGCGCCTTAACCCGGCGTTCGGTT	8	135 902	0.1514 %
GCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCCG	8	175 668	0.1957 %
CGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCC	8	182 541	0.2034 %
CCCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATAT	7	206 204	0.2297 %
GTCCTGTATTGTTATTTTTCGTCACTACCTCCCCGGGTCGGGAGTGGGTA	7	197 718	0.2203 %
CTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCCGGA	7	132 760	0.1479 %
CAAACTTTAAATGGGTAAGAAGCCCGGCTCGCTGGCGTGGAGCCGGGCGT	7	128 804	0.1435 %
CTTGATTAATGAAAACATTCTTGGCAAATGCTTTCGCTCTGGTCCGTCTT	6	143 558	0.1599 %
CGCAGTTTTATCCGGTAAAGCGAATGATTAGAGGTCTTGGGGCCGAAACG	6	103 611	0.1154 %
ACCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTGCCT	6	142 110	0.1583 %
CCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGATG	5	132 595	0.1477 %
CCCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATT	5	152 480	0.1699 %
CTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTGA	5	128 716	0.1434 %
CCTCCTTAGGCAACCTGGTGGTCCCCCGCTCCCGGGAGGTCACCATATTG	5	90 294	0.1006 %

samples, ranked by the number of samples they occur in.	
Status Checks: Status for each FastQC section showing whether results seem entirely normal (green), slightly abnormal (orange) or very unusual (red).	

ANNEX 4. Screenshot of BAM file generated by HISAT2 tool

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	MRNM	MPOS	ISIZE	SEQ
SRR19894454.1001984	16	chr1	13676	0	151M	*	0	0	GGTGGGTCTTGGCCATCCGTGAGATCTTCCCAGGGCAGCTCCCCTCTG
SRR19894454.1122534	256	chr1	14260	1	151M	*	0	0	CTCCCTCTCATCCCAGAGAAACAGGTCAGCTGGGAGCTTCTGCCCCCA
SRR19894454.910030	256	chr1	14407	1	151M	*	0	0	CTGCTCAGTTCTTTATTGATTGGTGTGCCGTTTTCTCTGGAAGCCTCTTA
SRR19894454.1390530	256	chr1	14407	1	151M	*	0	0	CTGCTCAGTTCTTTATTGATTGGTGTGCCGTTTTCTCTGGAAGCCTCTTA
SRR19894454.3739345	256	chr1	14407	1	151M	*	0	0	CTGCTCAGTTCTTTATTGATTGGTGTGCCGTTTTCTCTGGAAGCCTCTTA
SRR19894454.6001278	256	chr1	14410	1	151M	*	0	0	CTCAGTTCTTTATTGATTGGTGTGCCGTTTTCTCTGGAAGCCTCTTAAGA
SRR19894454.910030	272	chr1	14416	0	150M	*	0	0	TCTTTATTGATTGGTGTGCCGTTTTCTCTGGAAGCCTCTTAAGAACACG

ANNEX 5. Samtools Flagstat results on samples 1-8

Sample 1

```
34320119 + 0 in total (QC-passed reads + QC-failed rea
13752018 + 0 primary
20568101 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
33625182 + 0 mapped (97.98% : N/A)
13057081 + 0 primary mapped (94.95% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 2

```
32822330 + 0 in total (QC-passed reads + QC-failed
reads)
12637758 + 0 primary
20184572 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
32091357 + 0 mapped (97.77% : N/A)
11906785 + 0 primary mapped (94.22% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 3

```
33526846 + 0 in total (QC-passed reads + QC-failed rea
12322538 + 0 primary
21204308 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
32808276 + 0 mapped (97.86% : N/A)
11603968 + 0 primary mapped (94.17% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 4

```
29062583 + 0 in total (QC-passed reads + QC-failed rea
12128842 + 0 primary
16933741 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
28537932 + 0 mapped (98.19% : N/A)
11604191 + 0 primary mapped (95.67% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 5

```
28042000 + 0 in total (QC-passed reads + QC-failed rea
11486096 + 0 primary
16555904 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
27341419 + 0 mapped (97.50% : N/A)
10785515 + 0 primary mapped (93.90% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 6

```
19878547 + 0 in total (QC-passed reads + QC-failed rea
8779642 + 0 primary
11098905 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
18836300 + 0 mapped (94.76% : N/A)
7737395 + 0 primary mapped (88.13% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Sample 7

Sample 8

```
24488620 + 0 in total (QC-passed reads + QC-failed rea
8426238 + 0 primary
16062382 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
24085004 + 0 mapped (98.35% : N/A)
8022622 + 0 primary mapped (95.21% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

```
21164567 + 0 in total (QC-passed reads + QC-failed rea
10229530 + 0 primary
10935037 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
20699704 + 0 mapped (97.80% : N/A)
9764667 + 0 primary mapped (95.46% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Key Metrics and their relevance (Taking sample 1 for example):

```
34320119 + 0 in total (QC-passed reads + QC-failed reads)
13752018 + 0 primary
20568101 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
33625182 + 0 mapped (97.98% : N/A)
13057081 + 0 primary mapped (94.95% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

1. Total reads (34320119): the total number of sequencing reads in the dataset

A high read count ensures sufficient data coverage for accurate gene expression analysis.

2. Mapped reads (33625182, 97.98%): how many reads were successfully aligned to the reference genome

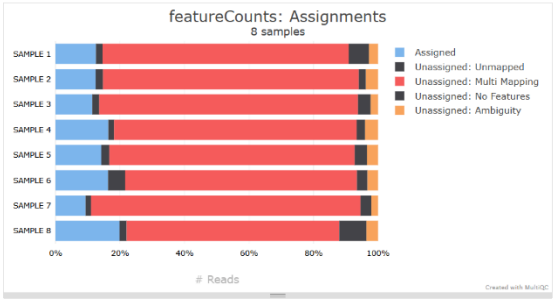
As I said earlier a high mapping rate (>70%) is good, meaning most reads matched known genomic regions.

3. Primary Mapped reads (94.95%): reads that uniquely mapped to a single location in the genome

High primary alignment means reads are specific to genes, improving quantification accuracy.

4. Secondary reads (20568101): reads which map to multiple locations in the genome (common in repetitive regions). High secondary alignments may indicate contamination, duplicated sequences, or issues with reading.

ANNEX 6. MultiQC results from the 8 summary featurecount files

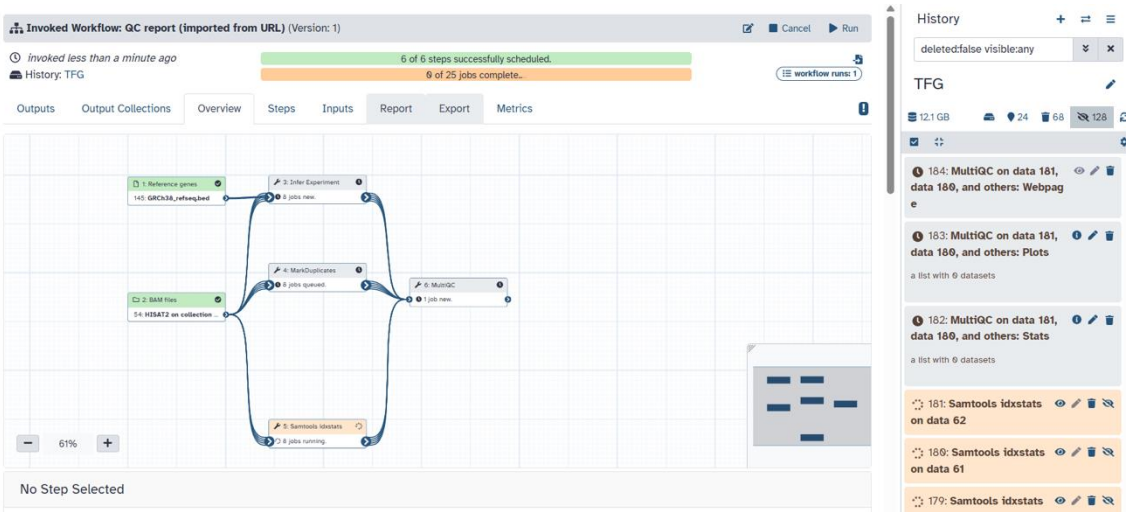
<table>
<tr>
<td>General Statistics</td>
<td>General Statistics
Copy table | Configure columns | Scatter plot | Violin plot | Export as CSV... | Showing 8/8 rows and 1/2 columns. | Summarize table
<table><tr><th>Sample Name</th><th>Assigned</th></tr><tr><td>SAMPLE 1</td><td>12.5%</td></tr><tr><td>SAMPLE 2</td><td>12.5%</td></tr><tr><td>SAMPLE 3</td><td>11.4%</td></tr><tr><td>SAMPLE 4</td><td>16.4%</td></tr><tr><td>SAMPLE 5</td><td>14.2%</td></tr><tr><td>SAMPLE 6</td><td>16.3%</td></tr><tr><td>SAMPLE 7</td><td>9.3%</td></tr><tr><td>SAMPLE 8</td><td>19.8%</td></tr></table></td>
</tr>
<tr>
<td>FeatureCounts:
Counts mapped reads for genomic features such as genes, exons, promoter, gene bodies, genomic bins and chromosomal locations</td>
<td>Assignments
Percentages | Summarize plot | Export...
 
</td>
</tr>
</table>

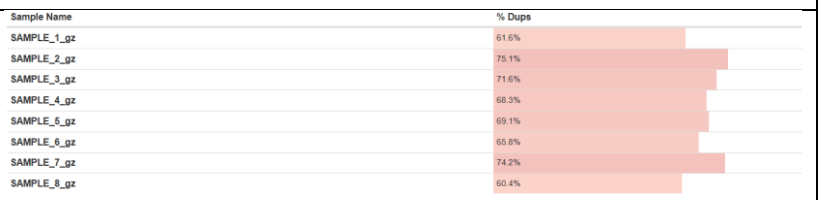
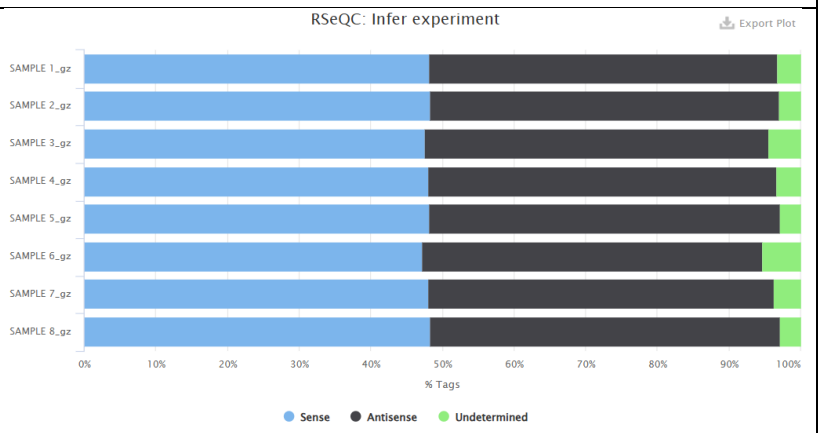
ANNEX 7. Count Matrix (the screenshot is for demonstrative purposes, part of a larger table that cannot be included in its entirety)

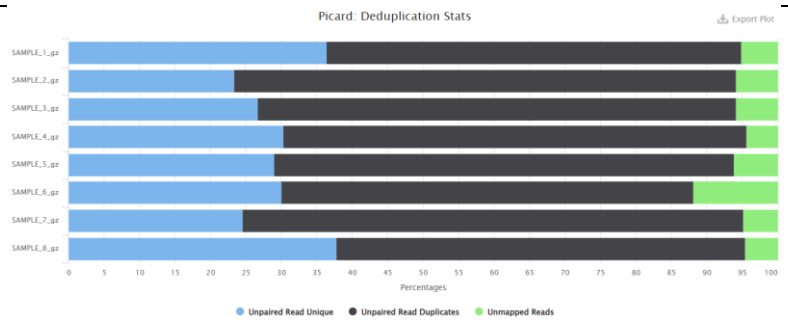
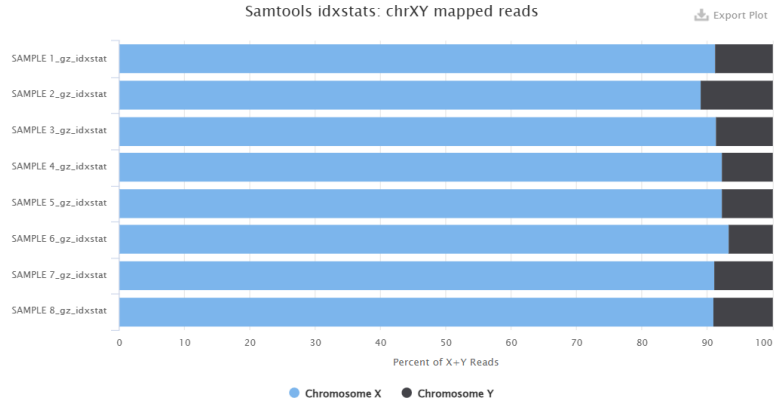
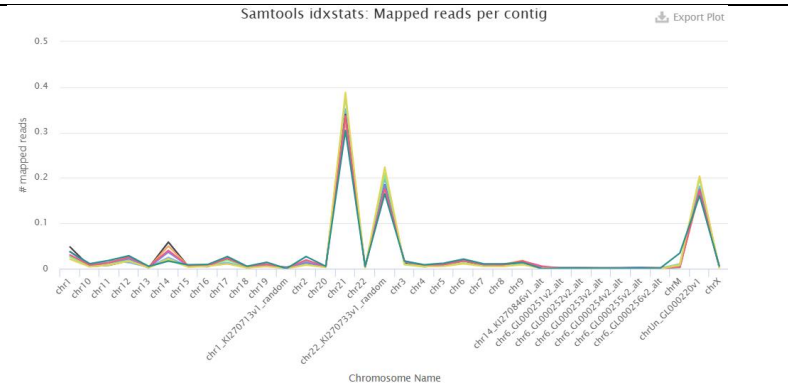
Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
Geneid	SAMPLE 1.gz	SAMPLE 2.gz	SAMPLE 3.gz	SAMPLE 4.gz	SAMPLE 5.gz	SAMPLE 6.gz	SAMPLE 7.gz	SAMPLE 8.gz
ENSG00000000003	367	218	207	183	47	50	263	85
ENSG00000000005	0	0	0	0	0	0	2	2
ENSG00000000419	131	156	87	176	170	80	83	129
ENSG00000000457	66	63	40	28	48	32	36	32
ENSG00000000460	23	44	27	50	18	4	12	14
ENSG00000000938	492	48	137	66	49	27	43	53
ENSG00000000971	149	62	82	57	41	20	57	84
ENSG00000001036	126	123	57	137	135	80	79	83
ENSG00000001084	130	211	224	295	294	151	123	256
ENSG00000001167	120	101	93	127	97	123	59	158
ENSG00000001460	25	47	31	40	40	21	33	46
ENSG00000001461	375	130	192	167	257	116	195	196
ENSG00000001497	111	135	122	294	196	139	60	149
ENSG00000001561	187	27	30	18	22	6	102	95
ENSG00000001617	138	241	349	417	233	179	172	357
ENSG00000001626	71	1	7	0	2	0	75	4
ENSG00000001629	396	362	211	492	271	237	215	236
ENSG00000001630	24	10	27	28	48	2	11	51

ANNEX 8. QC summary report

Using a prepared workflow (first picture below), the following three tools were run: Infer Experiment, MarkDuplicates and IdxStats. Then a MultiQC report was generated which has been copied in continuation.



Output	Graph
General Statistics	
Infer experiment: Counts the percentage of reads and read pairs that match the strandedness of overlapping transcripts. It can be used to infer whether RNA-seq library preps are stranded (sense or antisense).	

Picard	
Samtools: XY counts	
Samtools: Mapped reads per contig	

ANNEX 9. Filtered DEG Table (the screenshot is for demonstrative purposes, part of a larger table that cannot be included in its entirety)

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13
GeneID	logFC	AveExpr	t	P.Value	adj.P.Val	B	NA	NA	NA	NA	NA	NA
ENSG00000281181	2.4062624615849 2	7.1876283165484 6	5.7194716763844 8	0.0001651711957687 16	0.0005483242242 62	-4.3319016465696 7	21	8437628	8438551	-	lncRNA	NA
ENSG00000185652	1.7184719918155 6	4.7151439728359 6	4.6529925516855 1	0.0008164398657696 76	0.0005483242242 62	-4.3948312557622 1	29	19212641	19722926	+	protein_coding	SLC24A3
ENSG00000156475	4.1731915134449 4	-0.2807103919451 52	4.5459865466184 5	0.0006664955631326 1	0.0005483242242 62	-4.5623017626786 9	5	146586741	147984784	-	protein_coding	PPP2R2B
ENSG00000133719	2.3967669984969 9	9.9371691944819	4.4464196684465	0.0011342442513785 7	0.0005483242242 62	-4.3995511259668 6	5	148625682	148137382	+	protein_coding	SPINK5
ENSG00000176835	3.7454793491059 8	-0.4716642978176 28	4.3891962592619 6	0.0012311139974156	0.0005483242242 62	-4.5653972661545 8	9	133961989	133971861	+	protein_coding	CEL
ENSG00000266196	3.6967329651294 7	-1.6496991555143 9	4.2692486999926 2	0.0014937675435242 3	0.0005483242242 62	-4.5725122449634 4	11	17386648	17383531	+	lncRNA	NA
ENSG00000236495	3.2647835948684	-1.8962586316945 5	4.9589619412181 8	0.0021859767523891	0.0005483242242 62	-4.5749626262993	29	17599254	17599897	-	processed_pseudogene	NA
ENSG00000201639	3.2495919233453	-1.6216157259774	4.6541849726262 7	0.0021226117366976	0.0005483242242 62	-4.5759488958326 3	14	24966989	24951377	+	lncRNA	DHRS4L1
ENSG00000242515	3.8182554766671 6	-1.5873976853382 2	3.9589470774012	0.0024846428481827	0.0005483242242 62	-4.5735918917084	2	233636447	233773399	+	protein_coding	UGT1A19
ENSG00000158578	3.7169117659214 2	-0.3137438866967 11	3.9581113719273 1	0.0024889947456537 8	0.0005483242242 62	-4.5664788631921 9	X	55969954	55939977	-	protein_coding	ALAS2
ENSG00000137994	-2.1399216396322 6	4.9163552547275	-3.0426152968896 8	0.0025554926446492	0.0005483242242 62	-4.5878435368986	9	34989649	34998999	+	protein_coding	DNAJB5
ENSG00000197915	4.4227414651539 8	0.7995890545915 83	3.8795159241345 8	0.0028357879628969 4	0.0005483242242 62	-4.5557219967856 9	1	152212675	152224193	-	protein_coding	HRNR
ENSG00000169218	3.5333218944795 2	0.2996338996867 78	3.8149232145366	0.0031593392944116	0.0005483242242 62	-4.5626896918614 2	1	37611349	37634892	-	protein_coding	RSPO1
ENSG00000262993	3.3517045269697 5	-0.8587719992242 92	3.7664656346583	0.0034619181722382 2	0.0005483242242 62	-4.5725179644619 6	17	969631	917961	+	lncRNA	NA
ENSG00000095017	3.4211912572967 2	-0.7503676176519 2	3.7574749972476 7	0.0034703969723919	0.0005483242242 62	-4.5718266634639 9	16	1256958	1259988	+	protein_coding	TPSD1
ENSG00000201946	2.1357448358481 1	1.8942782591 92	3.7478735933597 8	0.0035369704753594 1	0.0005483242242 62	-4.5391469967736 1	13	21298114	21532496	+	lncRNA	NA
ENSG00000091422	-2.1225567849471 1	5.9157333587829 4	-3.7465292219355 9	0.0035441476492678	0.0005483242242 62	-4.4357877241113 6	19	43646994	43679547	-	protein_coding	PLAUR
ENSG00000079919	2.0465962617521 1	0.8624639946456 26	3.7317583598612 4	0.0036334928857464	0.0005483242242 62	-4.5556149969187	12	14612631	14696599	-	protein_coding	GUCY2C
ENSG00000259259	3.1829195166591 9	-1.6216157259774	3.7262639646698 2	0.0037945222352361	0.0005483242242 62	-4.5766411634554 4	15	58587596	58591676	+	lncRNA	NA
ENSG00000115414	-2.2139596766812 6	9.2913589699319 2	-3.7282211692167 7	0.0037857849168471 5	0.0005483242242 62	-4.4195581942674 7	2	215369439	215436973	-	protein_coding	FN1
ENSG00000185419	-2.6891455255521 2	1.8479774222124 7	-3.6970757833974 1	0.0038465564353659 2	0.0005483242242 62	-4.5625694398295	19	47493123	47419527	-	protein_coding	MEIS3
ENSG00000141756	-1.8025312758369 4	5.5989291845696	-3.6542379263532 4	0.0041419535883557 9	0.0005483242242 62	-4.4475426759967 6	17	41812679	41823213	+	protein_coding	FKBP10
ENSG00000191134	-4.1894617715399 5	-0.9185332929293 28	-3.6134509251466 2	0.0044385623457969 7	0.0005483242242 62	-4.5776523882639	29	54475592	54651169	+	protein_coding	DOK5
ENSG00000139767	3.1251688467818	-1.4166824832468 1	3.5988172469383 2	0.0045593536924916 9	0.0005483242242 62	-4.5764194419448 4	12	118981549	119163951	+	protein_coding	SRRM4
ENSG00000295712	2.9388613971933 7	-0.3715685399392 58	3.5938639973233	0.0045886828283917	0.0005483242242 62	-4.5786922616561	1	18999634	18919811	+	lncRNA	NA
ENSG00000182462	-3.8514266862476 1	-0.3259786697138 372	-3.5876219396761 5	0.0046452587499627 6	0.0005483242242 62	-4.5782628643365 3	1	215995774	215237999	+	protein_coding	KCNK2
ENSG00000187678	-1.7913285357229 8	4.1847643756677 9	-3.5857772939548 3	0.0046523676329679 8	0.0005483242242 62	-4.5915138919645	5	142319426	142326455	-	protein_coding	SPRY4
ENSG00000198796	-3.3936239193593 8	2.1946311719639 5	-3.5842631828768 8	0.0046643644696421 1	0.0005483242242 62	-4.5633561594658 4	18	58481246	58629991	-	protein_coding	ALPK2
ENSG00000201929	-3.1588239461596 1	-0.7131544491917 164	-3.5673393853739	0.0048896622831494	0.0005483242242 62	-4.5796717358419 4	7	55678493	55689587	-	lncRNA	NA