Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

HTGAnalyzer: An accessible R package with a web interface for enhanced transcriptomic analysis in precision medicine

Laia Díez-Ahijado [a,b], Aarón Marcén del Rincón [c], Lorena Marimón [a,b,d], Adela Saco [c,d], Marta del Pino [b,c,e], Aureli Torné [b,c,e], Katarzyna Darecka [d], Lia Sisuashvili [a,b,d], Núria Peñuelas [a,b], Pau Pascual-Mas [a], Núria Carreras-Dieguez [c,e], Oriol Ordi [a,b,f], Natalia Rakislova [a,b,d,f,1], Robert Albero [c,1,*]

[a] ISGlobal, Barcelona, Spain
[b] Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona (UB), c. Casanova, 143, 08036, Barcelona, Spain
[c] Institut d'investigacions biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[d] Department of Pathology, Hospital Clínic of Barcelona-University of Barcelona, Barcelona, Spain
[e] Department of Obstetrics and Gynecology, Hospital Clínic - Universitat de Barcelona, Barcelona, Spain
[f] CIBER de Epidemiología y Salud Pública, Instituto de Salud Carlos III, Spain

ARTICLE INFO

Keywords:
Transcriptomics
HTG
RNA-seq
Molecular pathology
Precision medicine
Clinical analysis

ABSTRACT

Transcriptomics generates promising data for personalized medicine, but its clinical utility is hindered by technical barriers associated with formalin-fixed, paraffin-embedded samples. Advances in automation and simplified workflows are key for integrating these technologies into routine care. We have developed HTGAnalyzer, an R package that enables bulk transcriptomics analysis from multiple data sources, such as HTG EdgeSeq and RNA-seq. Designed with clinical and molecular diagnostics in mind, HTGAnalyzer includes a fully automatic pipeline that performs sample quality control, data normalization, and a set of transcriptomic and clinical analyses. Importantly, HTGAnalyzer provides quality control and analysis tailored for HTG data, filling the gap left by HTG closure with no alternative solutions. Designed for ease of use without knowledge of bioinformatics, it features a user-friendly Shiny app, standardized functions for easy Docker setup, and enhanced connectivity to clinical data management systems. It delivers a complete analytical workflow in less than 20 min on a standard PC (Intel i7-7500U, 16 GB RAM). We applied HTGAnalyzer to diverse datasets, including a new vulvar cancer cohort, and generated interpretable reports with rich visual outputs. In summary, HTGAnalyzer offers robust quality control and advanced analytics in a single, unified tool, supporting the clinical use of transcriptomics in precision medicine.

1. Introduction

In recent decades, studies have shown that gene expression profiles are associated with specific cancer subtypes and that RNA-seq analysis provides more precise diagnoses and targeted therapies [1]. Currently, transcriptomics is being increasingly explored as a complementary method to genomic testing for precision-based treatments and patient stratification [2–4]. Gene expression profiling, in combination with computational algorithms, can also be effectively used to characterize the complex cell composition of tissues. Understanding and quantifying the tumor microenvironment (TME) is crucial in immunotherapy, as it significantly influences tumor progression and response to treatment [5,6]. For example, in melanoma, transcriptomic identification of different immune cell populations within the TME provides relevant prognostic and therapeutic data, including actionable immunotherapy targets [7]. To date, a few tools, which are primarily based on *in-silico* deconvolution, have been developed for this purpose within the context of clinical investigation. While the *in-silico* approach of characterizing the TME can address some limitations of the traditional methods, such as immunohistochemistry and flow cytometry, these methods are restricted to pre-defined cell populations and may not identify unknown or closely related phenotypes [8].

[*] Corresponding author.
E-mail address: ralbero@recerca.clinic.cat (R. Albero).

[1] These authors contributed equally.

https://doi.org/10.1016/j.compbiomed.2025.110772
Received 26 November 2024; Received in revised form 9 July 2025; Accepted 11 July 2025
Available online 23 July 2025

Despite the promising potential benefits of incorporating transcriptomic data into clinical practice, there continues to be a significant gap due to the challenges of integrating precision medicine technologies [9]. A major limitation lies in the widespread use of formalin fixed and paraffin-embedded (FFPE) samples, which complicates RNA extraction and analysis due to nucleic acid fragmentation, covalent modifications, and cross-linking with proteins [10]. These chemical alterations, along with factors like time to fixation, fixation duration, and storage conditions, negatively affect RNA quality and reproducibility. RNA degradation in FFPE samples tends to impact genes with short transcripts and high free energy, and introduces biases such as increased intronic reads and altered gene expression rankings [11]. While FFPE samples are valuable for retrospective studies, important questions remain about which extraction protocols yield high-quality RNA and how reliably these samples perform in downstream analyses [12]. Moreover, despite the low RNA concentrations measured in FFPE samples, successful polymerase chain reaction (PCR)-based sequencing is still feasible, highlighting the importance of evaluating RNA integrity beyond concentration thresholds [13]. Nevertheless, recent developments suggest that FFPE tissue can also support robust single-cell transcriptomic profiling, revealing clinically relevant traits and cell diversity, despite some technical variability [14]. Additionally, the limited amount of tumor tissue available for sequencing, especially in lung or lymphoid malignancies, further complicates the utility of these samples.

In this context, HTG Molecular Diagnostics developed the HTG EdgeSeq system, a platform that allowed successful gene expression profiling of FFPE-stored tumor samples without RNA extraction, offering high sensitivity and specificity, reduced sample requirements, a simplified and automated workflow, and integration of data across molecular assays [15]. This platform has been successfully utilized in several studies profiling immune-related gene expression in bladder [16] and oral squamous cell carcinomas [17]. However, the recent bankruptcy of HTG Molecular Diagnostics has created a critical gap in the field. With the dissolution of the company, many analytical tools, including HTG EdgeSeq Reveal software, a web-based platform for sample quality control (QC) and bioinformatic analysis are no longer available [18]. Consequently, valuable transcriptomic data worldwide is at risk of remaining unanalyzed and unreported, hampering progress in research and clinical applications. Moreover, previous efforts to develop pipelines for HTG analysis have been hindered by the lack of transparency in the analytical tools and methods, jeopardizing proper data interpretation and validation [19].

Transcriptomic pipelines are often designed for standard RNA-seq data and are not optimized for chemistries such as those used in targeted sequencing platforms, which lack specific normalization strategies and quality control metrics [20,21]. This mismatch can compromise data comparability and downstream analyses, especially when repurposing RNA-seq tools for platforms with fundamentally different workflows. Nonetheless, emerging artificial intelligence (AI)-driven optimization frameworks offer promising solutions for analyzing complex, non-standard transcriptomic data. Nature-inspired algorithms like the Greylag Goose Optimization model [22] and various machine learning and deep learning approaches have shown strong performance in domains such as energy forecasting [23], disease surveillance [24], food consumption prediction [25], and agricultural disease modeling [26]. Despite their potential, these AI models face challenges in clinical implementation due to issues including interpretability, validation, and integration with existing healthcare systems, which still rely on expert-driven genomic and transcriptomic frameworks.

To address this critical need, we have developed the HTGAnalyzer R package, a novel open-source analytical tool that allows users to perform comprehensive analysis and visualization of various types of transcriptomic data (HTG Transcriptome Panel and bulk RNA-seq). This software expands the functionalities previously included in the HTG EdgeSeq Reveal software, including oncogenic pathway analysis, various TME methods, and survival analysis. We herein present the main features of the package and describe its application in a series of squamous cell carcinomas of the vulva, which includes examples of the two pathological types of this tumor: human papillomavirus (HPV)-associated and HPV-independent. Additionally, validation with The Cancer Genome Atlas (TCGA) RNA-seq data confirmed the cross-platform robustness and utility of HTGAnalyzer.

2. Materials and methods

2.1. Patient data and tumor classification

This study used HTG sequencing results from FFPE samples of primary vulvar squamous cell carcinoma (VSCC) from patients treated at the Hospital Clínic of Barcelona between 2010 and 2022. The main characteristics of the series have recently been reported elsewhere [27]. The study was approved by the investigational review board and Ethics Committee of the Hospital Clínic of Barcelona (reference HCB/2020/1198). Written informed consent was obtained from patients whenever possible, while consent was waived for others according to Spanish regulations.

Tumors were classified following the World Health Organization 2020 classification [28] as HPV-associated or HPV-independent based on p16 immunohistochemistry (CINtec Histology anti-p16INK4a Kit, clone E6H4, Roche Diagnostics, Rotkreuz, Switzerland) and HPV-PCR testing (SPF10 PCR and LiPA25 system, Labo Biomedical Products, Rijswijk, The Netherlands), as previously described [27]. Tumors with positive staining for p16 and/or a positive testing result for high-risk HPV were classified as HPV-associated, whereas the categorization of a tumor as HPV-independent required a negative result by both tests.

2.2. HTG transcriptome panel

A total of 58 FFPE VSCC samples from 58 patients were analyzed using the HTG Transcriptome Panel. The panel contains 19,398 probes specific to human RNA transcripts and 218 control probes used for sample QC (4 positive control probes, 100 negative control probes, 22 genomic DNA probes, and 92 external RNA control consortium (ERCC) probes). Briefly, lysates from samples were run on the HTG EdgeSeq Processor (HTG Molecular Diagnostics, Tucson, AZ, USA) following the manufacturer's guidelines. Samples were individually barcoded using a 16-cycle PCR reaction to add adapters and molecular barcodes, purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA), and quantified using a KAPA Library Quantification kit (KAPA Biosystems, Wilmington, MA, USA). Libraries were sequenced on the Illumina sequencer platform (Illumina, San Diego, CA, USA). HTG EdgeSeq Reveal software was used to assess the QC status of 50 of the 58 samples, while 8 samples could not undergo QC due to discontinuation of the software. The QC results of these 50 samples are shown in Table S1.

2.3. Datasets

To benchmark QC parameters and define relevant filters for the *HTG_QC* function, we used the QC results from the HTG EdgeSeq Reveal of 50 VSCC samples, referred to as the "QC training dataset". To evaluate the functionalities of the HTGAnalyzer, we prepared a test dataset of 11 samples, which included the 8 samples without QC information (due to HTG EdgeSeq Reveal unavailability) and 3 HPV-independent VSCC samples from the QC training dataset that had previously passed QC. These 3 QC-passed samples served as positive controls to compare the QC results of HTG EdgeSeq Reveal with those established in our pipeline. Tables S2 and S3 contain the raw counts and clinical data of these 11 samples, respectively. The 11 cases in the test dataset included 4 HPV-associated and 7 HPV-independent tumors. Additionally, we incorporated external data from the TCGA Head and Neck Squamous Cell Carcinoma (TCGA-HNSC) cohort, accessed through the cBioPortal for Cancer Genomics (https://www.cbioportal.org/study/summary?

id=hnsc_tcga_gdc). The dataset, originally released in July 2024 with source data from the Genomic Data Commons, included a total of 502 primary tumor samples, from 133 female and 369 male patients. Both gene expression data—initially available as raw read counts—and clinical annotations were harmonized to align with our analytical pipeline. This included gene identifier standardization, such as the conversion of Entrez IDs to gene symbols using the org.Hs.eg.db database, to ensure consistency across datasets. Among the available clinical variables, we prioritized "SEX" as a contrast factor in downstream analyses and utilized "OS_months" and "OS_status" to define overall survival outcomes (Table S13).

2.4. Overview of the HTGAnalyzer package

The schematic workflow for HTGAnalyzer is shown in Fig. 1. Our pipeline comprises three functions: a) input of HTG/bulk RNA-seq counts file (*HTG_import_counts*), b) QC for HTG datasets (*HTG_QC*), and c) the transcriptomic analysis function (*HTG_analysis*), divided into four sub-functions. This modular design enhances flexibility, allowing users to adjust the workflow according to their bioinformatics expertise.

In addition, we included a global function called *HTG_auto*, which has been designed to automate the entire analysis workflow (Fig. 1). This function does not introduce novel analytical options, but sequentially executes the *HTG_import_counts*, *HTG_QC* (for HTG datasets only) and *HTG_analysis* with default parameters. This automation not only streamlines the process but also reduces the potential for user error, ensuring consistent and reliable results across analyses. For ease of installation and improved reproducibility, the package can be installed using standard R installation procedures or through Docker/Singularity containers, which help avoid compatibility issues across environments (guidelines at https://github.com/ISGLOBAL-Rakislova-Lab/HTGAnalyzer/).

Finally, the package includes complementary tools (See "Complementary tools" section below) designed for specific purposes, but which are not necessary for the default pipeline of analysis. Charts and tables derived from HTGAnalyzer meet the quality requirements of clinical reports and/or publications and can be modified locally.

a) Data importing

The *HTG_import_counts* function imports and preprocesses raw transcriptomic data in R with minimal modifications. This function is fully compatible with Excel files, preserving metadata rows like "Sample ID", "Well", "Date Parsed", and "Total Counts". This tool is designed to import data from both the HTG EdgeSeq transcriptomic panel and from bulk RNA-seq experiments. The function produces a data frame ready for QC processing and transcriptomic analysis.

In addition, annotation data can be imported into R by various methods, including integration with platforms like REDCap. Using the REDCapR package (Beasley W, 2025. REDCapR: Interaction Between R and REDCap. R package version 1.4.0.9000), clinical data collected at hospitals can be retrieved directly into R, enabling seamless incorporation into the HTGAnalyzer workflow.

b) QC assessment

The *HTG_QC* function evaluates six metrics, including control probes and sequencing depth (QC0 to QC5), specifically designed to assess the quality of HTG Transcriptome Panels [19]. These QC parameters are described in Table 1, along with the default thresholds set in the HTG

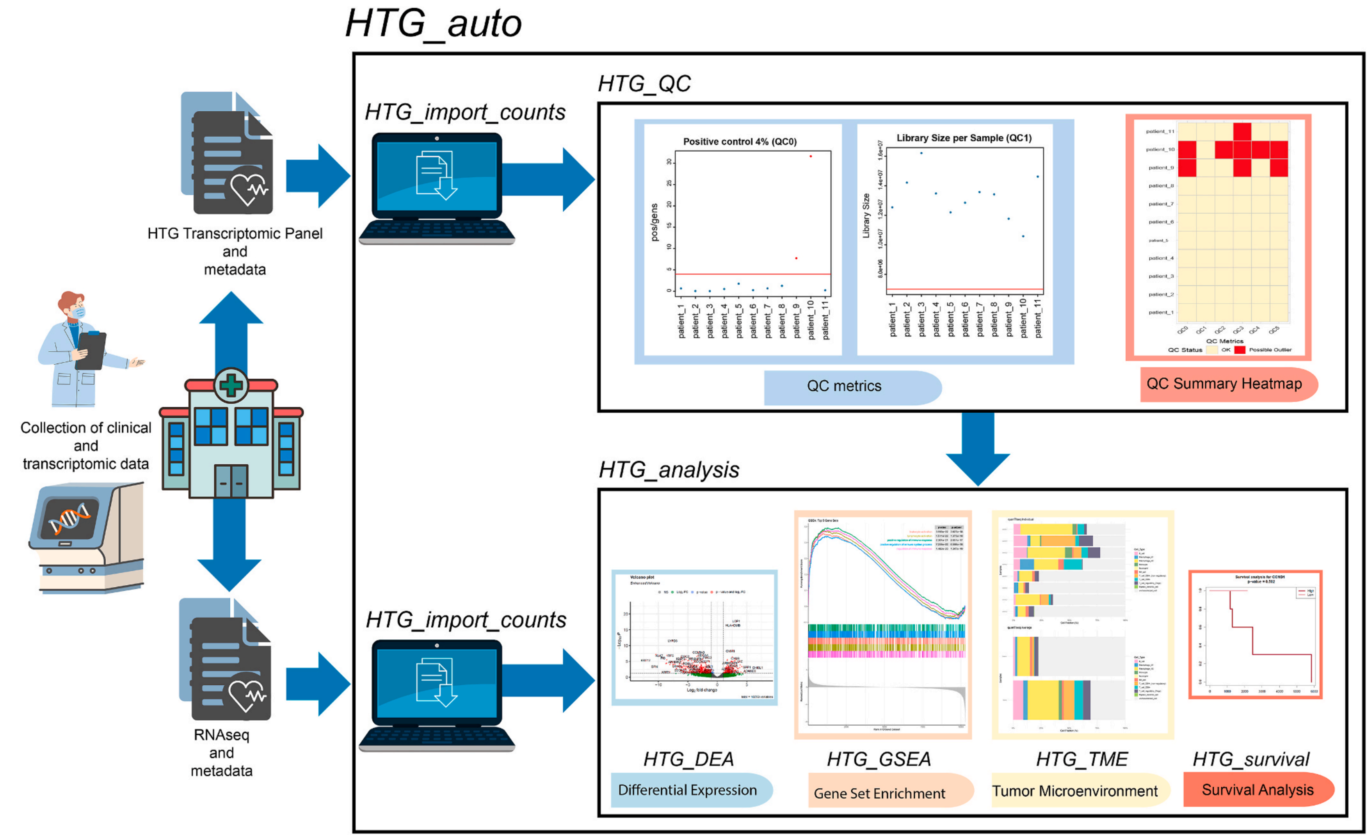


Fig. 1. Schematic representation of the HTGAnalyzer pipeline. HTGAnalyzer includes three main functions: (a) *HTG_import_counts* for importing HTG/bulk RNA-seq counts, (b) *HTG_QC* for quality control assessment, and (c) *HTG_analysis* for comprehensive data analysis through four subfunctions. This modular design allows users to adapt the workflow to their bioinformatics expertise. Additionally, the *HTG_auto* function automates the entire process and launches the whole HTGAnalyzer pipeline with default parameters.

Table 1
Overview of quality control (QC) parameters for the *HTG_QC* function.

QC Metric	QC criteria	Default threshold	Potential Cause of QC Failure	Criteria of failure	Failure Mode Detected
QC0	Percentage of counts in positive control probes	Higher than 4 %	Poor RNA quality	Proportion of positive control genes >4 %	Poor RNA quality
QC1	Total library size	7 million reads	Insufficient library size	Total library size <7 million reads	Insufficient library size
QC2	Percentage of counts in negative control probes	Higher than 0.045 %	High background noise from negative controls	Negative control value > 0.045	High background noise from negative controls
QC3	Percentage of counts in genomic DNA probes	Higher than 0.02 %	Genomic DNA contamination	Genomic DNA contamination value > 0.02	Genomic DNA contamination
QC4	Percentage of counts in ERCC spike-in probes	Higher than 0.025 %	ERCC spike-in control failure	ERCC spike-in control value > 0.025	ERCC spike-in control failure
QC5	Median number of counts per gene	Lower than 5 counts	Inadequate transcriptomic representation	Median of sample counts <5	Inadequate transcriptomic representation

ERCC: External RNA Controls Consortium.

analysis. Default parameters were established using results from the HTG Reveal analysis of 50 cases in the QC training dataset, for which QC status was already available (Table S2), together with previously defined thresholds. Table 1 outlines the QC parameters implemented by default in the *HTG_QC* function.

Each QC analysis (QC0 to QC5) yields an individual QC plot. In these individual QC plots, samples that fall outside these thresholds are flagged as failed (shown in red). However, *HTG_QC* flags samples close to these thresholds as borderline samples (shown in yellow). Specifically, the thresholds for yellow coloring are set at 5 % for positive controls (QC0), 5 million reads for library size (QC1), 0.05 for negative controls (QC2), 0.025 for genomic DNA (QC3), 0.03 for ERCC (QC4), and 3 for the median value (QC5). The samples with these borderline values will still be detected as outliers. A summary heatmap provides an overview of the six QC results.

c) Transcriptomic analysis function (HTG_analysis)

The *HTG_QC* function sends its results directly to the *HTG_analysis* function, creating a smooth pipeline that improves efficiency and simplifies the workflow. *HTG_analysis* offers four analytical subfunctions: 1) differential expression analysis (*HTG_DEA*), 2) gene set enrichment analysis (*HTG_GSEA*), 3) TME profiling (*HTG_TME*), and 4) survival analysis (*HTG_survival*). Each of these four subfunctions can be performed independently.

The *HTG_DEA* subfunction is designed to support data exploration through unsupervised analysis and perform differential expression analysis. First, the subfunction uses the DESeq2 package [29] for normalization, applying variance stabilization transformation to handle low or zero counts. Unsupervised analysis using sample clustering and principal component analysis (PCA) are conducted, to visualize the data structure and identify potential outliers. For this analysis, genes with low expression in at least 20 % of the samples are excluded. By default, “low-expressed” genes are defined using a threshold of 200 counts, although this threshold can be adjusted. Then, the *HTG_DEA* subfunction applies DESeq2 algorithms for differential analysis, offering the option to apply log-fold change (logFC) shrinkage using the statistical “apeglm” method if desired [30]. Finally, the *HTG_DEA* subfunction generates volcano plots and other complementary plots used in expression analysis, such as the MA plot, which shows the mean of normalized counts versus the logFC values, providing further insight into the contribution of these differentially expressed genes.

The *HTG_GSEA* subfunction performs comprehensive gene set enrichment analysis (GSEA) using the clusterProfiler and enrichplot packages to explore functional biological pathways and processes derived from gene expression data. By default, *HTG_GSEA* performs functional analysis using two libraries: Gene Ontology (GO) categories and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. First, the gseGO function from the clusterProfiler package is used to identify enriched GO terms, specifically focusing on the Biological Process category. This subfunction prepares a ranked gene list based on log2 fold change values and uses the gseGO utilities to detect pathway enrichment. This is followed by multiple visualizations, including dot plots, enrichment maps, ridge plots, heatmaps, and tree plots, which provide insights into significant pathways. Furthermore, the subfunction includes GO enrichment analysis, specifically focusing on biological processes affected by significantly differentially expressed genes, represented through bar plots of significant GO terms. Next, a gseKEGG analysis is performed by converting gene symbols to Entrez IDs to assess pathway enrichment in the KEGG database. The KEGG analysis is similarly visualized through dot plots, enrichment maps, ridge plots, and heatmaps. Results are stored in both PDF and CSV formats to ensure that graphical and tabular representations of the enrichment analyses are accessible for further interpretation.

The *HTG_TME* subfunction calculates transcripts per million (TPM)-normalized counts to assess the infiltrating cell populations within the TME using methods including EPIC, quanTIseq, and xCell [31–33], all of which are included in the IOBR package [34]. The normality of distributions is verified using the Shapiro-Wilk test, and statistical analyses, including t-tests and ANOVA, which are performed to compare cell fractions across different experimental conditions.

The *HTG_survival* subfunction allows performing two types of analysis. By default, if DEA results are available, the application shows the survival analysis, based on the top ten differentially expressed genes with the lowest adjusted p-values (padj). Alternatively, users can specify a gene or a set of genes to interrogate their association with survival. In all cases, Kaplan-Meier curves are generated using the *survfit* function from the Survival R package [35], and log-rank tests are used to assess differences. Furthermore, *HTG_survival* also performs this analysis for the TME populations identified via the *HTG_TME* subfunction. To generate the clinical groups of samples, the survival of which is compared by this subfunction, we implemented different categorization algorithms, such as median analysis, quartile analysis, or max-stat analysis, the latter being an algorithm that allows the selection of an unbiased statistical cut-off point for a continuous variable.

d) Complementary tools

Our package also includes three auxiliary functions outside the default workflow, designed to support specific analyses and extend the capabilities of HTGAnalyzer. These functions -*HTG_subset*, *HTG_quant_to_qual*, and *HTG_calculate_summary_stats*-enable researchers to perform targeted data manipulations and additional analyses as needed.

The *HTG_subset* function subsets a data frame or matrix by gene prefix and optionally normalizes the data to TPM. It also retrieves detailed gene-specific information from differential gene expression results, such as base mean counts and p-values. The *HTG_quant_to_qual* function converts quantitative data into qualitative labels based on a threshold, categorizing values as above or below the specified limit. Finally, we developed the *HTG_calculate_summary_stats* function to compute the same statistical tests as the *HTG_QC* function. This function is particularly useful for analyzing samples that are not processed by *HTG_QC*, such as RNA-seq raw counts matrices.

2.5. HTGAnalyzer Shiny Application

We have developed an interactive web application using the Shiny application to execute the *HTG_auto* global function, the core function of the HTGAnalyzer package that encompasses all its functionalities (Fig. 2). The HTGAnalyzer Shiny App enables researchers who are not familiar with the R programming language to perform comprehensive analyses of both HTG panels and bulk RNA-seq datasets. As indicated previously, samples from HTG panels can undergo QC using the *HTG_QC* function, which provides interactive QC plots, while bulk RNA-seq panels are directly fed into the *HTG_analysis* function. Users simply need to upload their input data files to access the complete suite of analyses provided by HTGAnalyzer, including *HTG_QC* and *HTG_analysis* functions. The application features an intuitive interface that facilitates the visualization and downloading of results. Importantly, the Shiny application also allows users to modify default parameters for the differential analysis functions, providing flexibility for customized analyses.

3. Results

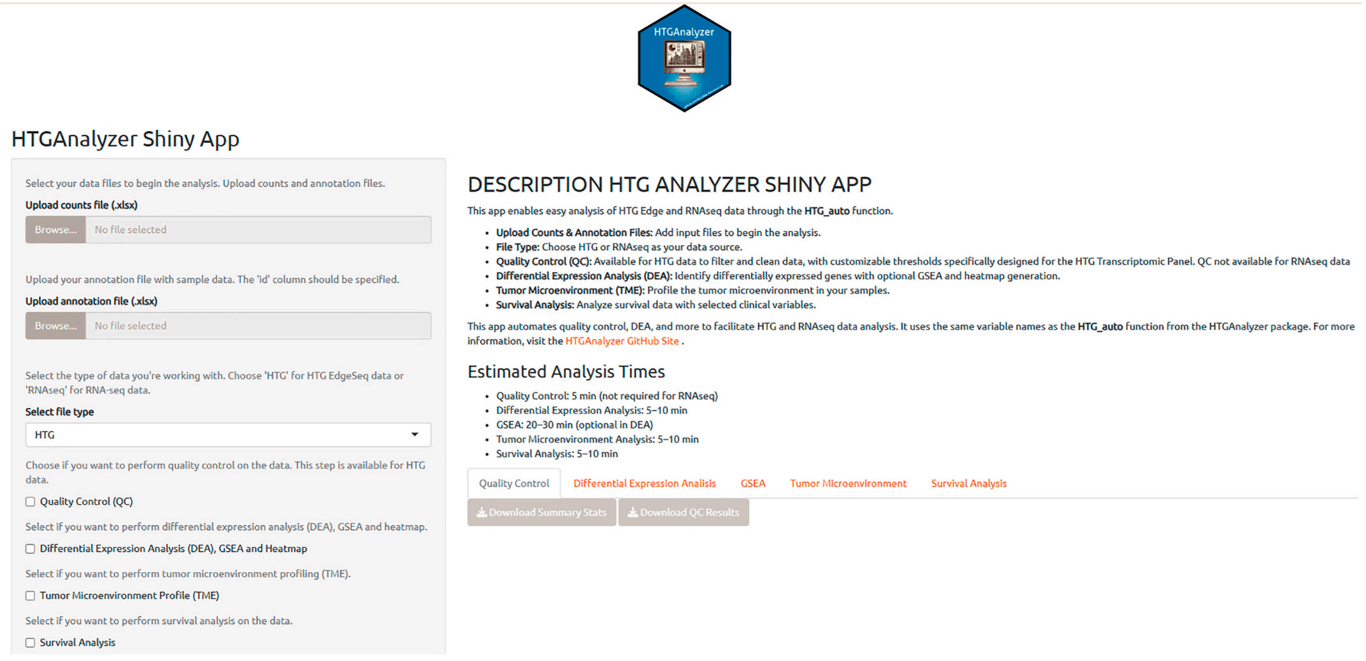
3.1. HTGAnalyzer allows easy data importing and QC assessment

The datasets of the 11 FFPE samples (Tables S2 and S3) were imported into R using the function *HTG_import_counts*. Next, the *HTG_QC* function calculated the individual QC values for each patient. Fig. 3A–F, show the six graphical summaries provided by the function (QC0 to QC5), highlighting samples that did not meet the QC criteria in red. The *HTG_QC* function also generates a heatmap with a graphical summary of the results (Fig. 3 G) together with two additional tables: one showing the general statistical measures for the dataset (Table S4), and a second table that focuses on QC metrics specific to control and genomic counts (Table S5). Three out of 11 VSCC samples (3/11) did not meet the QC criteria and were excluded. Specifically, patient 9 failed QC0, QC3, and QC5; patient 10 failed QC0, QC2, QC3, QC4, and QC5; and patient 11 failed QC3. Thus, a total of eight samples were deemed suitable to follow the analysis; four tumors were HPV-associated and four HPV-independent.

3.2. The HTG_Analysis function: HTG_DEA and HTG_GSEA subfunctions identify critical pathways and differential gene expression in the two subtypes of VSCC

HTG_DEA subfunction analysis, performed using PCA in an unsupervised manner, revealed that HPV-associated tumors clustered closely together, and were significantly different from HPV-independent tumors along the first and second PCA axes (Fig. 4 A). Since the test dataset included samples from both tumor groups, we could conduct differential analysis to further investigate their differences. Of the total of 10,293 genes that passed our filters, DEA identified 79 up-regulated and 181 down-regulated genes in four HPV-associated samples ($p < 0.05$). Notably, down-regulated genes included proteins associated with stroma, such as keratins and collagen *KRT5, KRT17*, and *COL5A3*, with *KRT5* showing a log2 FC of −8.77 (padj = 0.027), *KRT17* of −12.17 (padj = 5.08E-06), and *COL5A3* of −3.18 (padj = 1.28E-08) (Fig. 4 B; Table S6). On the other hand, upregulated genes were associated with increased immune infiltrate populations, such as *LCP* with a log2 FC of 3.22 (padj = 2.07E-18), *HLA-DMB* with log2 FC of 2.64 (padj = 4.12E-17), and *TNFAIP2* with log2 FC of 1.70 (padj = 2.70E-06). Other complementary plots generated by *HTG_DEA* include the MA plot, which showed a symmetric distribution of differentially expressed genes, highlighted in blue, in the two study groups (HPV-associated and HPV-independent) (Fig. S1).

In agreement with the DEA analysis, the *HTG_GSEA* subfunction showed significant enrichment in pathways related to host immune response in HPV-associated tumors (Fig. 5 A and B). The tree plot illustrates the close relationship among several gene expression signatures related to immune-related pathways, encompassing both the adaptive and innate immune systems, signaling pathways, and cell activation (Fig. 5 C). The application also generated a comprehensive list of oncogenic pathways provided as data tables, including significant GO terms with descriptions and associated statistics (Table S7).



HTGAnalyzer Shiny App

Select your data files to begin the analysis. Upload counts and annotation files.

Upload counts file (.xlsx)

Upload your annotation file with sample data. The 'id' column should be specified.

Upload annotation file (.xlsx)

Select the type of data you're working with. Choose 'HTG' for HTG EdgeSeq data or 'RNAseq' for RNA-seq data.

Select file type

HTG

Choose if you want to perform quality control on the data. This step is available for HTG data.

☐ Quality Control (QC)

Select if you want to perform differential expression analysis (DEA), GSEA and heatmap.

☐ Differential Expression Analysis (DEA), GSEA and Heatmap

Select if you want to perform tumor microenvironment profiling (TME).

☐ Tumor Microenvironment Profile (TME)

Select if you want to perform survival analysis on the data.

☐ Survival Analysis

DESCRIPTION HTG ANALYZER SHINY APP

This app enables easy analysis of HTG Edge and RNAseq data through the **HTG_auto** function.

- **Upload Counts & Annotation Files:** Add input files to begin the analysis.
- **File Type:** Choose HTG or RNAseq as your data source.
- **Quality Control (QC):** Available for HTG data to filter and clean data, with customizable thresholds specifically designed for the HTG Transcriptomic Panel. QC not available for RNAseq data
- **Differential Expression Analysis (DEA):** Identify differentially expressed genes with optional GSEA and heatmap generation.
- **Tumor Microenvironment (TME):** Profile the tumor microenvironment in your samples.
- **Survival Analysis:** Analyze survival data with selected clinical variables.

This app automates quality control, DEA, and more to facilitate HTG and RNAseq data analysis. It uses the same variable names as the **HTG_auto** function from the HTGAnalyzer package. For more information, visit the HTGAnalyzer GitHub Site .

Estimated Analysis Times

- Quality Control: 5 min (not required for RNAseq)
- Differential Expression Analysis: 5–10 min
- GSEA: 20–30 min (optional in DEA)
- Tumor Microenvironment Analysis: 5–10 min
- Survival Analysis: 5–10 min

Fig. 2. Screenshot of the web-based interface of the HTGAnalyzer Shiny App.

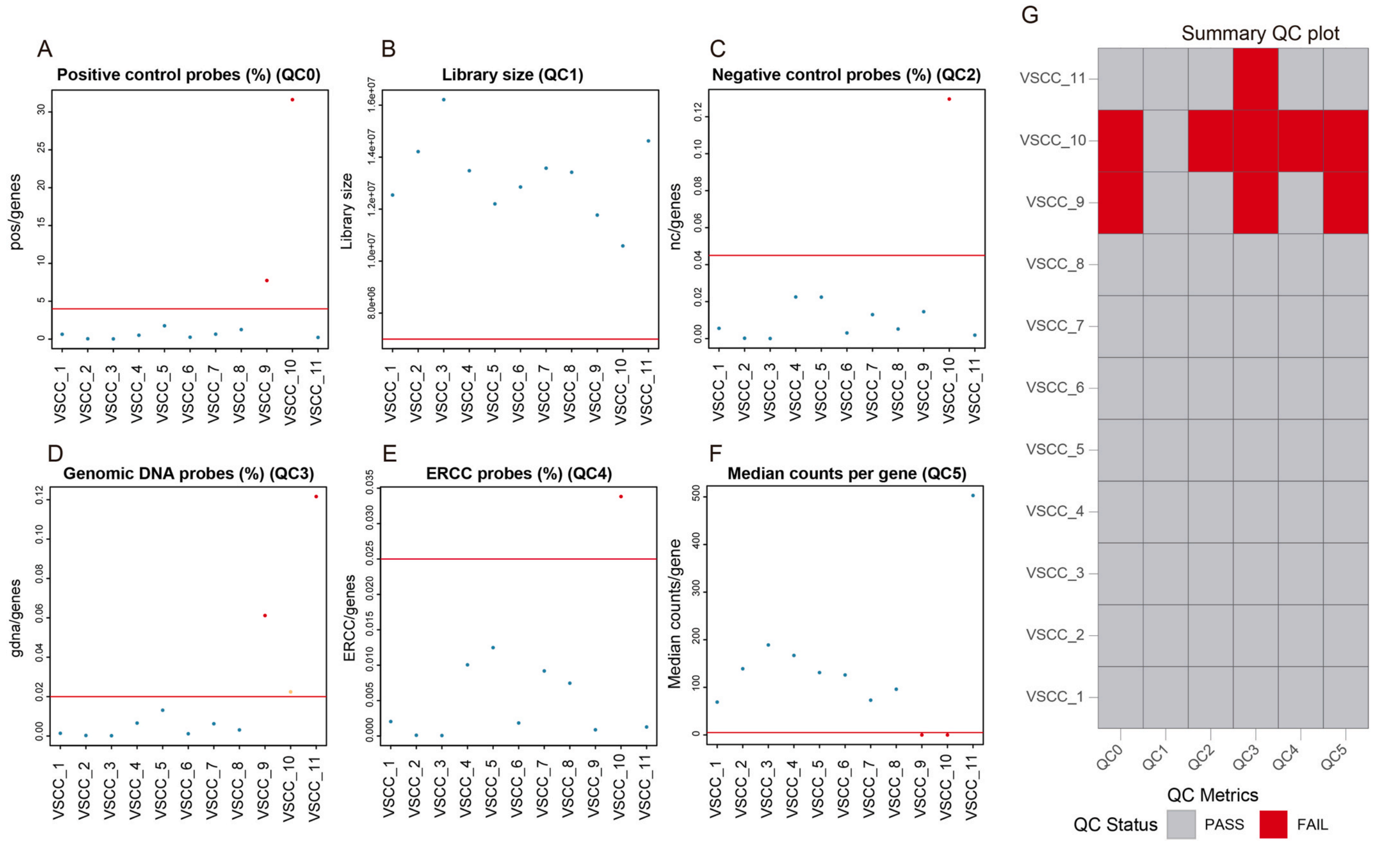


Fig. 3. Summary of the *HTG_QC* function output. A–F: Individual quality control (QC) analysis. Samples that do not pass QC thresholds are flagged in red or yellow (borderline samples). QC tests include: A: QC0: Positive control probes; B: QC1: Library size; C: QC2: Negative control probes (%); D: QC3: Genomic DNA probes (%); E: QC4: ERCC probes (%); F: QC5: Median counts per gene. G: Summary QC plot, showing samples that do not meet QC standards in red. QC: Quality Control. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

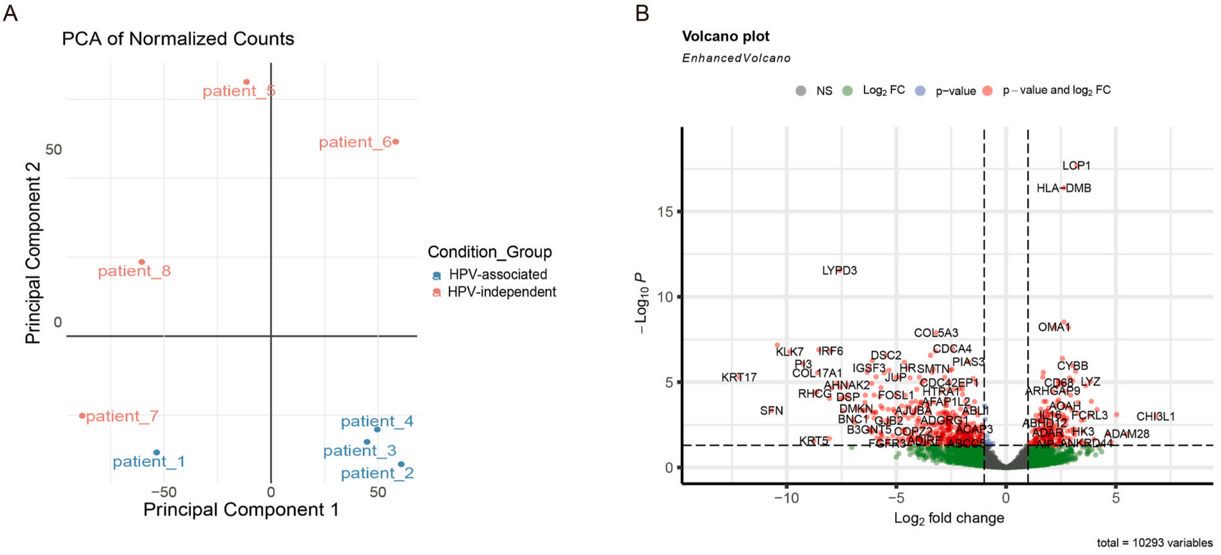


Fig. 4. Differential gene expression analysis using the subfunction *HTG_DEA* of HTGAnalyzer. A: Unsupervised Principal components analysis of vulvar squamous cell carcinoma (VSCC) samples that passed the quality control (QC) (n = 8); B: Volcano plot representation of the differential gene expression in human papillomavirus (HPV)-associated versus HPV-independent tumors. Differentially expressed genes are highlighted in red, with dashed lines indicating adjusted p-value and log fold change (LogFC) thresholds. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

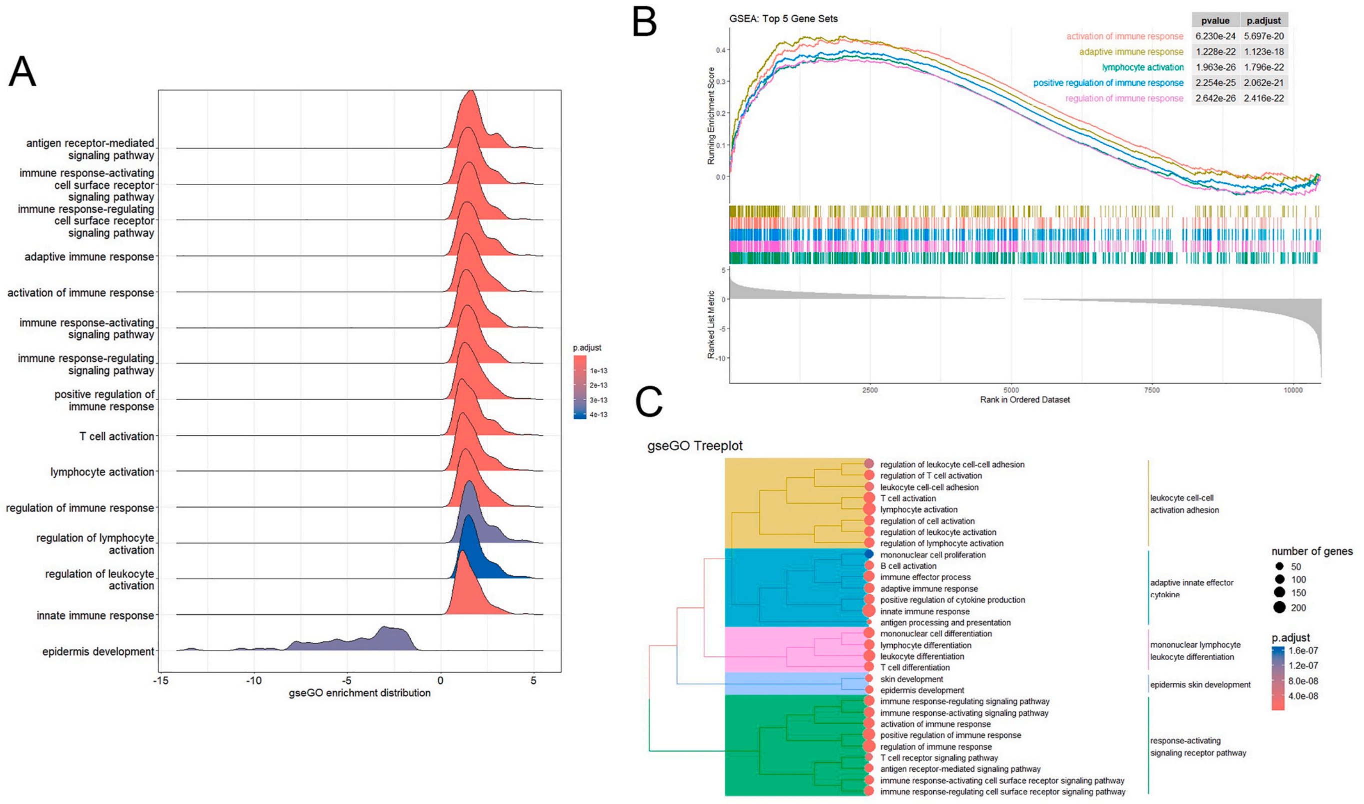


Fig. 5. Pathway analysis using HTGAnalyzer. A–C: Gene set enrichment analysis (GSEA) analysis plots, including Ridge plot of top 15 significant Ontology signatures (A), the top 5 enriched gene sets (B), and TreePlot gene set clustering (C).

3.3. The HTG_Analysis function: HTG_TME and HTG_survival subfunctions integrate transcriptomic profiling and tumor microenvironment and clinical analysis

Characterization of the TME with the *HTG_TME* subfunction revealed distinct profiles for HPV-associated and HPV-independent tumors (Fig. 6 A). Our findings highlight significant differences in the TME related to HPV status in VSCC. HPV-associated tumors exhibited a more complex TME, with a higher percentage of macrophages and lymphoid populations (B cells and $CD8^+$ T cells). Unsupervised clustering of TME populations using quanTIseq confirmed these findings, revealing two distinct clusters: one composed exclusively of HPV-associated cases and another containing all the HPV-independent tumors (Fig. 6 B). Additionally, an excel file containing the TPM-normalized counts, needed for TME, was generated (Table S8) together with the deconvolution reports from the three methods (Tables S9, S10, and S11).

Finally, the *HTG_survival* subfunction reported survival analysis of the top differentially expressed genes. Interestingly, we identified *LCP1* as a gene overexpressed in HPV-associated tumors and associated with a higher risk of recurrence (Fig. 6 C; Table S12). Finally, we applied *HTG_survival* to the test dataset to explore patient survival in relation to specific TME populations. Although we did not find any significant correlations between specific populations and time to recurrence, we observed a trend showing that patients with a higher proportion of cancer-associated fibroblasts in their tumors tended to have longer times to recurrence (Fig. 6 D).

3.4. Application of HTGAnalyzer to TCGA RNA-seq data

To study the standardization and scalability of our tool, we applied HTGAnalyzer to RNA-seq data from 502 tumor samples in the TCGA-HNSC (PanCancer Atlas) cohort obtained via cBioPortal. Using the integrated *HTG_auto* function, we streamlined data import, visualization with annotated heatmaps (e.g., sex, survival), and downstream analyses. DEA identified 3175 sex-associated genes (padj <0.05; Fig. S2A), including strong upregulation of Y-linked genes in males: *NLGN4Y* ($\log_2$FC ≈ 7.85), *TXLNGY* ($\log_2$FC ≈ 9.15), *PRKY* ($\log_2$FC ≈ 8.10), and *RPS4Y1* ($\log_2$FC ≈ 9.10), consistent with expected genomic differences (Table S14). GSEA further revealed sex-linked biological processes, immune responses, and lipoprotein metabolism pathways (Table S15; Fig. S2B) [36]. Immune deconvolution using EPIC, QTI, and xCell in TPM-normalized data provided complementary TME profiles (Table S16–S18; Fig. S2C). Survival analysis identified *CDKN2A* overexpression as being significantly associated with poor prognosis ($p = 0.0009$), in alignment with previous reports [37,38] (Table S19; Fig. S2D).

3.5. Performance and resource usage

To evaluate the computational performance of HTGAnalyzer, we assessed runtime and memory consumption during execution of the complete analysis pipeline via the HTG_auto function, using two datasets: the HTG tutorial dataset and the TCGA HNSC dataset. Both analyses were conducted on a standard laptop (Intel i7-7500U, 16 GB RAM), thereby demonstrating the scalability and efficiency of the tool across datasets of varying size and complexity.

Execution of the HTG tutorial dataset was completed in 1020.72 s, with 879.87 s attributed to user processing time and 60.94 s to system operations, and a peak memory usage of 1.75 GB (Table S20). In contrast, analysis of the larger TCGA dataset required 4859.12 s (4199.11 user seconds, 132.76 system seconds), with a maximum memory usage of 2.60 GB (Table S21).

For reference, the HTG dataset consisted of a 5.5 KB clinical annotation file and a 1.07 MB gene expression matrix, while the TCGA dataset comprised a 101 KB annotation file and a 73.7 MB expression matrix. These benchmarks underscore robust performance and practical

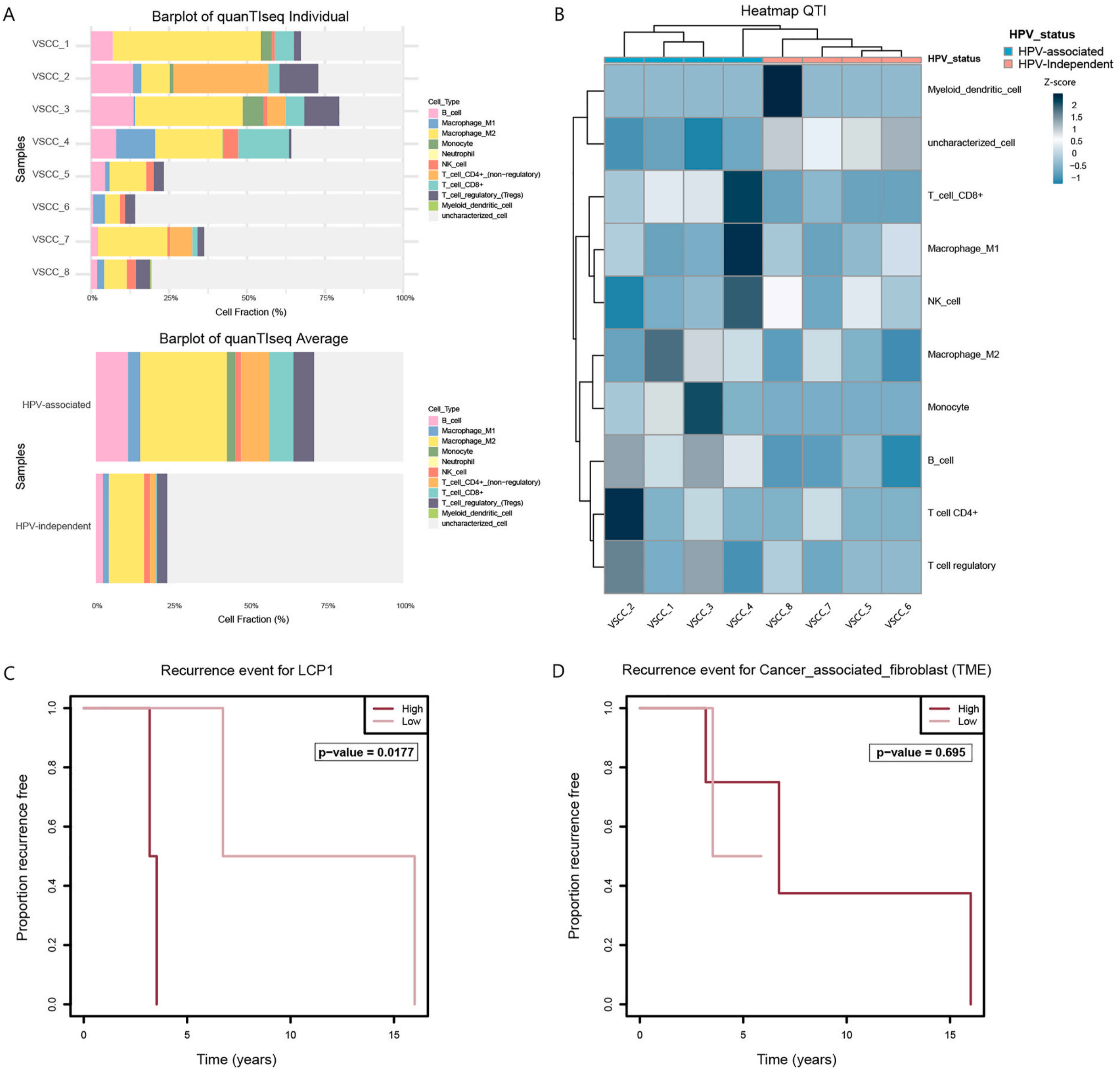


Fig. 6. Tumor microenvironment (TME) and survival analysis using HTGAnalyzer. A: TME deconvolution results using the quanTIseq algorithm, indicating the percentage of each immuno-population detected per sample (Upper panel) or grouped by condition (Bottom panel). B: Unsupervised clustering based on TME deconvoluted proportions predicted by quanTIseq. C–D: Kaplan-Meir analysis for Time to Recurrence for LCP1 (C) and Cancer associated fibroblasts (D). P-values are calculated using the log-rank test with the associated chi-squared statistic. Sample groups (High and Low) were determined using Max-Stat.

utility of HTGAnalyzer for both small-scale and high-dimensional transcriptomic datasets.

4. Discussion

HTGAnalyzer is a publicly available R package designed to provide a user-friendly pipeline for integrating QC and gene expression analysis of bulk transcriptomic data, including data generated by HTG and RNA-seq. This tool replicates the analyses implemented in HTG EdgeSeq Reveal-QC, unsupervised analysis, DEA, and GSEA-while introducing novel functionalities, such as TME and survival analysis focused on inflammatory cell populations essential for understanding tumor biology and prognostic significance.

Technically, HTGAnalyzer performed efficiently in both tutorial and real RNA-seq datasets, completing full pipelines with low memory usage (<2 GB) and short runtimes (~17 min for HTG-seq data with <20 samples; Table S20). In larger datasets (e.g., TCGA, >500 samples), it remained scalable (<3 GB, ~60 min; Table S21), without high-performance computing. The tool can be deployed locally or on servers using Docker or Singularity and is compatible with platforms like REDCapR and potentially with electronic health record systems, supporting clinical decision-making workflows.

A key innovation of HTGAnalyzer is its response to the unmet need for a dedicated open-source pipeline for HTG transcriptomic panels. After the discontinuation of HTG Molecular Diagnostics, no tools specifically addressed these datasets. Existing alternatives, such as Lodovico (2022) [19], lack specific compatibility with HTG transcriptome panels, compromising reproducibility. Although several tools are available for specific tasks - such as GSEA, CIBERSORTx, or IOBR [34,39,40] - these are typically designed for general RNA-seq or microarray data in different steps. In contrast, HTGAnalyzer integrates QC, DEA, GSEA, TME, and survival analysis into a single workflow for HTG technology and RNA-seq.

Another major strength of this tool is its support for multiple data formats, enabling use in both research and clinical workflows across oncology, immunology, and infectious diseases. Application in a small

VSCC dataset produced results consistent with known HPV-related tumor inflammation [41].

HTGAnalyzer identified *LCP1* overexpression in HPV-associated VSCC, matching prior links between *LCP1* and immune infiltration in HPV-related cancers [42–44]. Increased *LCP1* levels were also associated with disease recurrence, showing the promising application of HTGAnalyzer for identifying prognostic biomarkers. Additional validation using TCGA datasets showed results in concordance with existing literature [45,46], further supporting its reliability. Although primarily intended for HTG whole transcriptome panels, it can also be adapted for other HTG panels and transcriptomic assays, including NanoString.

Nonetheless, some limitations should be noted. First, the effectiveness of the tool depends on input data quality and careful selection of analytical parameters, which may require adjustment for specific datasets. Secondly, as it has only been tested in human datasets, further validation in non-human models would be beneficial to evaluate its broader applicability in translational research. Moreover, as HTGAnalyzer relies on several R and Bioconductor dependencies, maintaining compatibility may require periodic updates. However, this can be mitigated through the use of reproducible environments, such as Docker, which preserves package versions and ensures long-term stability.

While HTGAnalyzer does not currently incorporate multi-omics integration or AI-based optimization, this is an intentional design choice. AI models often require large, homogeneous datasets and complex training pipelines, which can be challenging to implement in clinical workflows. In contrast, HTGAnalyzer employs classical statistical methods that emphasize transparency, reproducibility, and interpretability, all of which are essential features for clinical adoption. The tool has been developed in response to a growing need for an accessible, routine transcriptomic analysis in medical settings, where diagnostic questions are often constrained by limited bioinformatic resources and expertise. Although it does not directly implement machine learning, the structured and modular design of HTGAnalyzer is aligned with the goals of AI-driven clinical decision support systems [25,26] making it a strong foundation for future AI integration.

In conclusion, HTGAnalyzer complemented with a HTG Shiny App, simplifies data importation and QC assessment, making the process more efficient while ensuring data integrity. Our QC method successfully replicated HTG EdgeSeq Reveal results, in alignment with previous studies on HTG-derived data analysis and further demonstrating the accuracy of the package. By providing easy-to-use analysis, comprehensive reports, and visualizations, HTGAnalyzer serves as a valuable alternative for transcriptomic data analysis, and is particularly beneficial for the clinical community and researchers with limited bioinformatics experience.

CRediT authorship contribution statement

Laia Díez-Ahijado: Writing – original draft, Conceptualization. **Aarón Marcén del Rincón:** Formal analysis. **Lorena Marimón:** Formal analysis, Data curation. **Adela Saco:** Writing – review & editing. **Marta del Pino:** Writing – review & editing. **Aureli Torné:** Writing – review & editing, Supervision. **Katarzyna Darecka:** Writing – review & editing, Data curation. **Lia Sisuashvili:** Writing – review & editing. **Núria Peñuelas:** Supervision, Data curation. **Pau Pascual-Mas:** Writing – review & editing. **Núria Carreras-Dieguez:** Writing – review & editing, Data curation. **Oriol Ordi:** Writing – review & editing. **Natalia Rakislova:** Writing – original draft, Supervision, Conceptualization. **Robert Albero:** Writing – original draft, Supervision, Conceptualization.

Data availability

Detailed tutorials and data to reproduce the analyses, as well as documentation on the HTGAnalyzer pipeline are available at https://github.com/ISGLOBAL-Rakislova-Lab/HTGAnalyzer/. HTGAnalyzer Shiny App (installation and basic usage instructions) is available at: https://github.com/ISGLOBAL-Rakislova-Lab/HTGAnalyzer_shiny. Additionally, a folder containing the full report (tables and plots) generated by HTGAnalyzer (*HTG_auto* function) is included here: https://github.com/ISGLOBAL-Rakislova-Lab/HTGAnalyzer/tree/main/vignettes All detailed results and complete examples of the output generated by our pipeline is available in the Supplementary Output folder on GitHub (https://github.com/ISGLOBAL-Rakislova-Lab/HTGAnalyzer/tree/main/SUPLEMENTARY_OUTPUT).

Ethics compliance statement

To the Editors of Computers in Biology and Medicine, We hereby confirm that our manuscript entitled: "HTGAnalyzer: An accessible R package with a web interface for enhanced transcriptomic analysis in precision medicine" complies with the ethical standards required by Computers in Biology and Medicine.

All procedures involving human participants were performed in accordance with relevant laws and institutional guidelines and have been approved by the appropriate institutional committee. Specifically, the study was approved by the Institutional Review Board and Ethics Committee of the Hospital Clínic of Barcelona (approval reference HCB/2020/1198). Written informed consent was obtained from patients where possible, and for cases where this was not feasible, consent was waived in accordance with Spanish regulations. Patient privacy rights have been fully respected and protected.

The following paragraph, included in the Methods section of the manuscript, details this compliance:

This study used HTG sequencing results from FFPE samples of primary vulvar squamous cell carcinoma (VSCC) from patients treated at Hospital Clínic of Barcelona between 2010 and 2022. The main characteristics of the series have been recently reported elsewhere. The study was approved by the investigational review board and Ethics Committee of the Hospital Clínic of Barcelona (reference HCB/2020/1198). Written informed consent was obtained from patients where possible; for others, consent was waived according to Spanish regulations.

Funding sources

This work was supported by the Spanish Instituto de Salud Carlos III (FIS, PI23/00494; NR and JO).

Declaration of competing interest

The authors declare no conflict of interests.

Acknowledgments

The authors thank the Barcelona Institute for Global Health (ISGlobal), The Hospital Clínic of Barcelona and the Institut d'Investigacions Biomèdiques August Pi i Sunyer. ISGlobal projects include “PI23/00494; Perfiles de expresión génica de los tres subtipos principales de los carcinomas escamosos de vulva”, funded by Instituto de Salud Carlos III and cofunded by the European Union (ERDF) “A way to make Europe”.

We acknowledge support from the grant CEX2023-0001290-S funded by MCIN/AEI/10.13039/501100011033, and support from the Generalitat de Catalunya through the CERCA Program”. RA is supported by the Spanish Ministry of Science through the *Ramon y Cajal* Program (RYC2021-031306-I). This work was partially developed at the Center Esther Koplowitz (Barcelona, Spain) and ISGlobal (Barcelona, Spain).

Abbreviations

AI: artificial intelligence; DEA: Differential Expression Analysis; FFPE: Formalin-Fixed Paraffin-Embedded; ERCC: External RNA Controls Consortium; GSEA: Gene Set Enrichment Analysis; HPV: Human

Papillomavirus; KEGG: Kyoto Encyclopedia of Genes and GenomeslogFC: Log-fold change; PCA: Principal Component Analysis; PCR: Polymerase Chain Reaction; QC: Quality Control; TME: Tumor Microenvironment; TPM: Transcripts Per Million; VSCC: Vulvar Squamous Cell Carcinoma; GO: Gene Ontology; TCGA: The Cancer Genome Atlas; HNSC: Head and Neck Squamous Cell Carcinoma.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.110772.

References

[1] J.N. Weinstein, et al., The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (10) (Oct. 2013) 1113–1120.

[2] A.M. Tsimberidou, E. Fountzilas, L. Bleris, R. Kurzrock, Transcriptomics and solid tumors: the next frontier in precision cancer medicine, Semin. Cancer Biol. 84 (Sep. 2022) 50–59.

[3] C.M. van Tilburg, et al., The pediatric precision oncology INFORM registry: clinical outcome and benefit for patients with very high-evidence targets, Cancer Discov. 11 (11) (Nov. 2021) 2764–2779.

[4] H. Lee, et al., Diagnostic utility of transcriptome sequencing for rare Mendelian diseases, Genet. Med. 22 (3) (Mar. 2020) 490–499.

[5] A.J. Gentles, et al., The prognostic landscape of genes and infiltrating immune cells across human cancers, Nat. Med. 21 (8) (Aug. 2015) 938–945.

[6] A. Tiwari, R. Trivedi, S.-Y. Lin, Tumor microenvironment: barrier or opportunity towards effective cancer therapy, J. Biomed. Sci. 29 (1) (Oct. 2022) 83.

[7] R. He, et al., Advancing immunotherapy for melanoma: the critical role of single-cell analysis in identifying predictive biomarkers, Front. Immunol. 15 (Jul. 2024).

[8] K. Rangamuwa, et al., Methods for assessment of the tumour microenvironment and immune interactions in non-small cell lung cancer. A narrative review, Front. Oncol. 13 (Apr. 2023).

[9] D. Stefanicka-Wojtas, D. Kurpas, Personalised medicine—implementation to the healthcare system in Europe (Focus Group discussions), J. Personalized Med. 13 (3) (Feb. 2023) 380.

[10] J. Hedegaard, et al., Next-Generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue, PLoS One 9 (5) (May 2014) e98187.

[11] Y. Lin, et al., Optimization of FFPE preparation and identification of gene attributes associated with RNA degradation, NAR Genomics Bioinforma. 6 (Mar. 2024).

[12] N.M. Odogwu, et al., Optimizing RNA extraction methods for high-throughput transcriptome sequencing of formalin-fixed paraffin-embedded cardiac tissue specimens, PLoS One 19 (Dec. 2024).

[13] K. Masago, et al., Comparison between fluorimetry (Qubit) and spectrophotometry (NanoDrop) in the quantification of DNA and RNA extracted from frozen and FFPE tissues from lung cancer patients: a real-world use of genomic tests, Méd. Sur 57 (Dec. 2021).

[14] A. Trinks, et al., Robust detection of clinically relevant features in single-cell RNA profiles of patient-matched fresh and formalin-fixed paraffin-embedded (FFPE) lung cancer tissue, Cell. Oncol. 47 (Aug. 2024) 1221–1231.

[15] Y. Shi, et al., Evaluation of the EdgeSeq precision immuno-oncology panel for gene expression profiling from clinical formalin-fixed paraffin-embedded tumor specimens, Front. Cell Dev. Biol. 10 (May 2022).

[16] F.J. Koll, et al., Optimizing identification of consensus molecular subtypes in muscle-invasive bladder cancer: a comparison of two sequencing methods and gene sets using FFPE specimens, BMC Cancer 23 (1) (Jun. 2023) 504.

[17] K. Horny, et al., Mesenchymal–epithelial transition in lymph node metastases of oral squamous cell carcinoma is accompanied by ZEB1 expression, J. Transl. Med. 21 (1) (Apr. 2023) 267.

[18] D. Ran, et al., Platform comparison of HTG EdgeSeq and RNA-Seq for gene expression profiling of tumor tissue specimens, J. Clin. Oncol. 38 (15_suppl) (May 2020) 3566.

[19] L. Terzi di Bergamo, F. Guidetti, D. Rossi, F. Bertoni, L. Cascione, HTGQC and shinyHTGQC: an R package and shinyR application for quality controls of HTG EDGE-seq protocols, Gigabyte 2022 (Dec. 2022) 1–5.

[20] A. Conesa, et al., A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (Jan-2016). BioMed Central Ltd.

[21] R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: the teenage years, Nat. Rev. Genet. 20 (Nov-2019) 631–656. Nature Publishing Group.

[22] E.S.M. El-kenawy, N. Khodadadi, S. Mirjalili, A.A. Abdelhamid, M.M. Eid, A. Ibrahim, Greylag goose optimization: nature-inspired optimization algorithm, Expert Syst. Appl. 238 (Mar) (2024).

[23] M.A. Yassen, et al., An AI-Based system for predicting renewable energy power output using advanced optimization algorithms, J. Artif. Intell. Metaheuristics 8 (2024) 1–8.

[24] E. El-Sayed, M.M. Eid, L. Abualigah, Machine learning in public health forecasting and monitoring the zika virus, Metaheuristic Optim. Rev. 1 (2024) 1–11.

[25] M. Eed, A.A. Alhussan, A.S.T. Qenawy, A.M. Osman, A.M. Elshewey, R. Arnous, Potato consumption forecasting based on a hybrid stacked deep learning model, Potato Res 68 (2024) 809–833.

[26] M. Radwan, A.A. Alhussan, A. Ibrahim, S.M. Tawfeek, Potato leaf disease classification using optimized machine learning models and feature selection techniques, Potato Res 68 (2024) 897–921.

[27] O. Ordi, et al., Whole-Exome sequencing of vulvar squamous cell carcinomas reveals an impaired prognosis in patients with TP53 mutations and concurrent CCND1 gains, Mod. Pathol. 37 (10) (Oct. 2024) 100574.

[28] A.K. Höhn, C.E. Brambs, G.G.R. Hiller, D. May, E. Schmoeckel, L.C. Horn, 2020 WHO classification of female genital tumors. WHO-Klassifikation 2020 für Tumoren des unteren weiblichen Genitales, Geburtshilfe Frauenheilkd. 81 (10) (Oct. 2021) 1145–1153.

[29] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (Dec. 2014) 550.

[30] A. Zhu, J.G. Ibrahim, M.I. Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences, Bioinformatics 35 (12) (Jun. 2019) 2084–2092.

[31] J. Racle, K. de Jonge, P. Baumgaertner, D.E. Speiser, D. Gfeller, Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, eLife 6 (Nov) (2017).

[32] G. Sturm, et al., Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, Bioinformatics 35 (14) (Jul. 2019) i436–i445.

[33] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, Genome Biol. 18 (1) (Dec. 2017) 220.

[34] D. Zeng, et al., IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures, Front. Immunol. 12 (Jul. 2021).

[35] A.H. Shakur, S. Huang, X. Qian, X. Chang, SURVFIT: doubly sparse rule learning for survival data, J. Biomed. Inf. 117 (May 2021) 103691.

[36] X. Zeng, et al., M2 macrophage-derived TGF-β induces age-associated loss of adipogenesis through progenitor cell senescence, Mol. Metabol. 84 (Jun) (2024).

[37] Y. Dong, M. Zheng, X. Wang, C. Yu, T. Qin, X. Shen, High expression of CDKN2A is associated with poor prognosis in colorectal cancer and may guide PD-1-mediated immunotherapy, BMC Cancer 23 (1) (Dec. 2023) 1097.

[38] J.P. Luo, J. Wang, J.H. Huang, CDKN2A is a prognostic biomarker and correlated with immune infiltrates in hepatocellular carcinoma, Biosci. Rep. 41 (10) (Oct. 2021) 20211103.

[39] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (Oct. 2005) 15545–15550.

[40] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259. Humana Press Inc.

[41] K.E. Kortekaas, et al., High numbers of activated helper T cells are associated with better clinical outcome in early stage vulvar cancer, irrespective of HPV or p53 status, J. Immunother. Cancer 7 (1) (Dec. 2019) 236.

[42] S. Pan, M. Wan, H. Jin, R. Ning, J. Zhang, X. Han, LCP1 correlates with immune infiltration: a prognostic marker for triple-negative breast cancer, BMC Immunol. 25 (1) (Jul. 2024) 42.

[43] C. Ji, et al., Identification of immune infiltrating cell-related biomarkers in early gastric cancer progression, Technol. Cancer Res. Treat. 23 (Jan. 2024).

[44] R. Yang, et al., Combined transcriptome and proteome analysis of immortalized human keratinocytes expressing human papillomavirus 16 (HPV16) oncogenes reveals novel key factors and networks in HPV-Induced carcinogenesis, mSphere 4 (2) (Apr. 2019).

[45] Y. Wang, C. Zhou, T. Li, J. Luo, Prognostic value of CDKN2A in head and neck squamous cell carcinoma via pathomics and machine learning, J. Cell Mol. Med. 28 (May 2024).

[46] L. Xue, et al., Next-generation sequencing identifies CDKN2A alterations as prognostic biomarkers in recurrent or metastatic head and neck squamous cell carcinoma predominantly receiving immune checkpoint inhibitors, Front. Oncol. 13 (2023) 1276009.