BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 21, Issue 7

DOI: 10.58240/1829006X-2025.21.7-423

ORIGINAL REPORT

# ACE2 EXPRESSION AS A PREDICTIVE BIOMARKER IN ORAL SQUAMOUS CELL CARCINOMA: A NON-LINEAR DIMENSIONALITY REDUCTION APPROACH

**Shivani Sivasakthivel[1], Pratibha Ramani[2], Pradeep Kumar Yadalam[3]**

[1]Research scholar, ORCHID No.:0000-0003-2021-4906, Department of Oral and Maxillofacial pathology, Saveetha Dental college and Hospitals, Saveetha institute of Medical and technical sciences, jsshivaninarayanan@gmail.com

[2]Professor and Head, Department of Oral and Maxillofacial pathology, Saveetha Dental college and Hospitals, Saveetha institute of Medical and technical sciences, dr_pratibha@rediffmail.com

[3]Assisstant Professor, Department of Periodontics, Saveetha Dental college and Hospitals, Saveetha institute of Medical and technical sciences, yadalam1980@gmail.com

***Corresponding Author**: **Shivani Sivasakthivel** Research scholar, Department of Oral and Maxillofacial pathology, Saveetha Dental college and Hospitals, Saveetha institute of Medical and technical sciences, jsshivaninarayanan@gmail.com

***Received:*** *Jun 7. 2025;* ***Accepted:*** *Jul 28, 2025;* ***Published:*** *Aug 27, 2025*

## ABSTRACT

**Purpose:** Angiotensin-converting enzyme 2 (ACE2), a critical regulator of the Circulating endocrine renin-angiotensin system, regulates cardio vascular and pulmonary homeostasis. Emerging studies gives solid evidence linking ACE2 to various cancer types. However, the expression pattern and functional impacts of ACE2 in cancer is little known. This study systematically analysed ACE2 expression across various cancer forms, with an emphasis on head and neck squamous cell carcinoma (HNSC), using advanced non-linear dimensionality reduction techniques.

**Methods:** The study analysed pan-cancer dataset from The Cancer Genome Atlas (TCGA) that included 5,141 tumour samples from 10 distinct cancer types. A radial basis function (RBF) kernel was employed in Kernel Principal Component Analysis (Kernel PCA) to identify non-linear patterns in ACE2 expression. K-means clustering and t-distributed Stochastic Neighbor Embedding (t-SNE) were employed to demonstrate the high-dimensional data and to identify distinct subgroups. The Mann-Whitney U test was used to assess the statistical significance in ACE2 expression between HNSC and other cancer types.

**Results**: Kernel PCA revealed three distinct ACE2 expression patterns across cancer types: high (KIRC, COADREAD), moderate (HNSC), and low (BRCA, PRAD). Comparative analysis using the Mann–Whitney U test showed a statistically significant difference in ACE2 expression between HNSC and other malignancies ($p = 2.169 \times 10^{-5}$). Clustering with K-means in the Kernel PCA-transformed space achieved optimal separation of expression profiles, with a silhouette score of 0.68. Compared to PCA and t-SNE, Kernel PCA demonstrated superior performance with a lower reconstruction error (0.2760) and improved interpretability of tumour-associated heterogeneity.

**Conclusion:** This study demonstrates the effectiveness of non-linear dimensionality reduction using Kernel Principal Component Analysis (Kernel PCA) in delineating the heterogeneity of ACE2 expression across multiple cancer types. Head and Neck Squamous Cell Carcinoma (HNSC) was identified as a distinct molecular subtype, exhibiting unique ACE2 expression profiles. These findings support the potential utility of ACE2 profiling in cancer stratification, highlighting its relevance in precision oncology.

***Keywords****:* ACE2 Expression, Kernel Principal Component Analysis (Kernel PCA). Head and Neck Squamous Cell Carcinoma (HNSC), Pan-Cancer Analysis, Non-Linear Dimensionality Reduction, Machine learning

## INTRODUCTION

Angiotensin-converting enzyme 2, a transmembrane metalloenzyme of the Renin Angiotensin System, gained significant interest for its role as the SARS-CoV-2 [1] entry receptor and also for its expanding importance in cancer biology. Beyond its classical function of regulating cardiovascular and pulmonary homeostasis, ACE2 also plays vital role in the regulation of oxidative stress, inflammation, tissue remodelling, along with infiltrating tumours, interferon signalling, and the epithelial-mesenchymal transition, all of which are associated with tumour microenvironment[2.] Hence understanding ACE2's expression pattern and its functional impact across various range of cancers, can help us identify new biomarkers and therapeutic targets in the era of oncology intersections and tailored cancer immunotherapy. In the world of cancer research, integrating Machine Learning (ML) algorithm and gene expression data has been proven to learn about cancer

patients. Conventional ML techniques could not process the data with precision. In order to attain précised output, genes are selected and the data size is reduced before using ML analysis. Dimensionality reduction technique is one of the conventionally used ways to reduce the data size. Dimensionality reduction techniques reduce the dimension and complexity of the data, but maintaining the relevance between genes by generating a representation of the original data in a lower dimension. Linear dimensionality reduction techniques such as Principal Component Analysis (PCA) have been extensively used for gene expression data analysis [3,4]. However, gene regulatory systems, cellular signalling pathways, and gene-environment interactions involves complex non-linear dynamics that are difficult for linear models technique. Principal Component Analysis [5,6] and other standard linear dimensionality reduction approaches are commonly used and they depend on Euclidean distance principle, but their inherent disadvantage is that they are incapable of precisely representing the complex, non-linear interactions with high dimension that constitute biological systems. Kernel PCA (KPCA) surpasses this limitation and effectively captures non-linear patterns from the original data space by projecting data into a higher-dimensional feature space where linear links can be discovered. Clustering analysis helps in identifying comparable expression patterns, but it is challenging due to small sample numbers and high dimensionality [7,8.] A study demonstrates that when preprocessing five cancer expression datasets prior to grouping, nonlinear dimensionality reduction (NDR) works better than PCA (*9*).

Previous studies have analysed distribution of ACE2 expression using the GTEx, CCLE, and TCGA pan-cancer databases and moreover majority of the earlier analyses used linear dimensionality reduction approaches [10]. Hence by employing non-linear dimensionality reduction technique we analysed ACE 2 expression patterns and their intricate correlations across various cancers in high-dimensional expression. Studies have evaluated the effect of ACE2 on clinical prognosis using the Kaplan-Meier survival plot and COX regression analysis [11]. Further, the relationship between ACE2 and immune infiltration levels, ACE2's interactions with immunological neoantigens, DNA Methyltransferase (DNMT), TMB, microsatellite instability, Mismatch Repair Genes (MMRs), and HLA gene members were also analysed. Analysis was performed on the frequency of ACE2 gene mutations in different tumour types [12,13]. ACE2 is widely expressed in normal tissues but shows low to medium levels in cancer cell lines. Its expression correlates with immune cell infiltration in various cancers, indicating a potential role in the tumour environment [14].

With a particular emphasis on head and neck squamous cell carcinoma (HNSCC), this study uses transcriptomic data from The Cancer Genome Atlas (TCGA) to systematically analyse ACE2 expression profiles using Kernel Principal Component Analysis (Kernel PCA). Comparative analyses of Kernel PCA with conventional techniques like PCA in determining the prognostic variables and tumour-associated heterogeneity was done to check the reliability and accuracy of Kernel PCA. Furthermore, its capacity to distinguish ACE2 expression patterns across subtypes of cancers was also assessed. Thus, our study aims to promote the use of non-linear dimensionality reduction techniques in oncogenomic and offers novel insights on the ACE2 expression patterns across various cancers.

## MATERIALS AND METHODS

### Data Source and Preprocessing

The study utilized data from the pan-cancer ACE2 expression dataset (Dataset ID: TCGA-PANCAN-ACE2-2023) from The Cancer Genome Atlas (TCGA) for the analysis. The dataset includes 5,141 tumour samples from 10 different cancer types. Each dataset record contains a unique sample identifier ID, their associated cancer type, normalized ACE2 gene expression levels, and an extra binary indicator variable designating particular cancer type Preprocessing procedures like data cleaning, creating one-hot encoded label using feature engineering depending on the cancer type were carried out to guarantee data consistency and quality before employing dimensionality reduction techniques. Followed by standardization of ACE2 expression values was done by normalizing to zero mean and unit variance using the StandardScaler function from scikit-learn which is paramount for distance-based algorithms such as Kernel PCA.

### Dimensionality Reduction with Kernel PCA

Dimensionality reduction of the ACE2 expression dataset was performed using Kernel Principal Component Analysis (Kernel PCA) to detect the non-linear ACE2 expression patterns across cancer types. Scikit-learn library's Kernel PCA class with two set of components was used to facilitate visualization. The Radial Basis Function (RBF) kernel was chosen due to its superior ability to distinguish between different types of cancer within the reduced feature space, following a comparison of several kernel functions, including linear, polynomial, sigmoid, and cosine. The mathematical definition of the RBF kernel isK(x,y)=exp⁡(−γ∥x−y∥2)K(x, y) = \exp(-\gamma \|x – y\|^2), where γ\gamma (gamma) is a crucial parameter governing the influence radius of every data point. While a higher gamma highlights local variations by reducing the radius of influence, a lower gamma captures larger, global structures. After much empirical testing, this study determined that a gamma value 0.1 would effectively balance local and global patterns. This number ensured the reduced representation maintained significant biological variance by offering the best possible separation between cancer types without overfitting. A fixed random state of 42 was used for reproducibility.

**Statistical Analysis**

**Mann-Whitney U Test for HNSC**

Mann-Whitney U test was employed to evaluate and compare the ACE2 expression in head and neck squamous cell carcinoma (HNSC) with other cancer types. Mann-Whitney U test, a non -parametric test is used because gene expression data is inconsistent and deviates from the assumptions of a normal distribution. The analysis revealed statistically significant difference, with a p-value of 2.169×10−52.169 \times 10^ {-5} signifying the differences in ACE2 expression levels between HNSC and the pooled group of other cancer types. These findings infer a substantial variation in ACE2 expression in HNSC. Degree of uncertainty surrounding the observed differences can be measured by using confidence intervals (CIs) values to improve the results' interpretability. By doing this, the results would be more reliable as these intervals would furnish the actual range where the true variation in ACE 2 expression exists. Similarly, in depth analysis on the variation of ACE2 expression across the cancer field could be evaluated by employing Mann-Whitney U tests in pairwise comparisons among cancer types.

**Clustering and Comparative Analysis**

K-means clustering analysis was applied to identify underlying intrinsic patterns of ACE2 expression from the data extracted from Kernel PCA. The optimal number of clusters was determined by calculating silhouette scores for cluster numbers ranging from 2 to 10 and selected the number that maximized the score. Three clusters were used in the clustering configuration. K-means++ initialization method with a maximum of 300 iterations, a fixed random state of 42 and K-means algorithm was applied to increase convergence efficiency and reproducibility. Thus, the clustering analysis approach facilitated to stratify significant subgroups of varying ACE2 expression patterns within the cancer types.

Comparative analysis with t-Distributed Stochastic Neighbor Embedding (t-SNE) was performed using two visualization components with a perplexity of 30, at a learning rate of 200, 1,000 iterations and with random state set at 42. Various qualitative and quantitative techniques were employed to detect the efficacy of dimensionality reduction and clustering techniques. Silhouette scoring method was employed to measure the cluster cohesion and separation range. Visual analysis provided the new aspects into clustering efficacy. Furthermore, biological relevance of ACE2 expression levels across clusters, analysis of cluster composition by cancer type and the composition of each cluster was examined for cancer types.

## RESULTS

The analysis found significant variation in ACE2 expression among the 10 cancer types examined (**Table 1, Fig 2,3,4,5**). Head and neck squamous cell carcinoma (HNSC) produced a distinct clear cluster demonstrating moderate ACE2 expression, when dimensionality was decreased using Kernel PCA.


**Figure 1. Study workflow chart**

**Table 1.The summary statistics of ACE2 expression by cancer type, sorted by mean expression level**

| | Cancer Type | Sample Count | Mean ACE2 Expression | Std Dev | Min | Max |
|---|---|---|---|---|---|---|
| 3 | KIRC | 510 | 1910.06 | 2003.24 | 0.0 | 17099.9 |
| 1 | COADREAD | 592 | 1029.76 | 1093.35 | -0.08 | 8560.71 |
| 7 | PAAD | 177 | 398.52 | 768.05 | 0.53 | 4924.19 |
| 9 | STAD | 412 | 369.4 | 745.18 | -0.09 | 7216.32 |
| 4 | LIHC | 366 | 264.84 | 731.08 | 0.0 | 7809.05 |
| 5 | LUAD | 510 | 222.64 | 678.08 | 0.0 | 12898.6 |
| 2 | HNSC | 515 | 121.12 | 134.88 | 0.0 | 1256.01 |
| 6 | LUSC | 484 | 120.6 | 202.16 | 0.0 | 1997.15 |
| 0 | BRCA | 1082 | 88.82 | 780.82 | 0.0 | 13683.8 |

This distribution of ACE2 expression demonstrates highly right-skewed pattern, with majority of samples exhibiting relatively low expression levels and minority of samples exhibit extremely high expression with a median value of 52.35 and the maximum value was 17099.9. This pattern is consistent with the known biology of ACE2, which is highly expressed in specific tissues such as the kidney and intestinal epithelium. (**Table 1**).

The Mann-Whitney U test compared ACE2 expression in head and neck squamous cell carcinoma (HNSCC) to other cancer types. The findings showed a statistically significant difference with a p-value of 2.169000428741905e-05 and a U-statistic of 1326881.5. Mann-Whitney U test demonstrates statistically significant difference in ACE2 expression between HNSC and other cancer types ($p < 0.05$). Kidney Renal Clear Cell Carcinoma (KIRC) has the highest median ACE2 expression of any cancer type at 1430.04, followed by Colorectal Adenocarcinoma (COADREAD) at 717.84. Additional noteworthy mentions include Head and Neck Squamous Cell Carcinoma (HNSC) at 79.37, Pancreatic Adenocarcinoma (PAAD) at 182.96, and Stomach Adenocarcinoma (STAD) at 108.15. HNSC ranks fifth in ACE 2 expression among all the ten cancer types. And on the list expressing ACE2 contains Lung Adenocarcinoma (LUAD) at 72.37, Lung Squamous Cell Carcinoma (LUSC) at 61.35, Liver Hepatocellular Carcinoma (LIHC) at 18.79, Prostate Adenocarcinoma (PRAD) at 6.37, and Breast Carcinoma (BRCA) at 2.64.

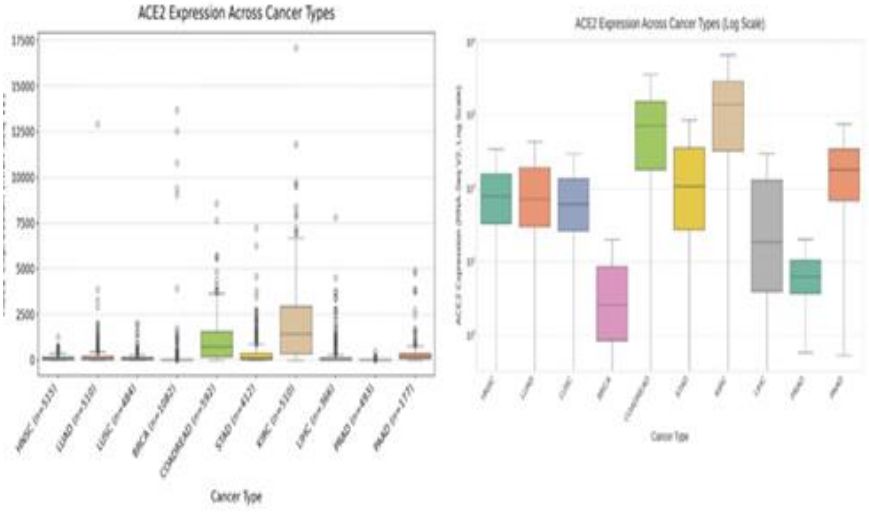


**Figure 2** The boxplot of ACE2 expression levels across different cancer types, with sample sizes indicated for each. There is considerable variation in ACE2 expression among different types of cancer

The bar graph exhibits marked differences in the median ACE2 expression between kidney (KIRC) and colorectal (COADREAD) cancers and other cancer types (**Fig 2**).

Out of all the cancer types, Breast cancer (BRCA) exhibits the lowest expression levels, whereas KIRC and COADREAD exhibit the highest (**Fig 3)** and HNSC (oral cancer) shows moderate ACE2 e xpression (**Fig 4,5).**

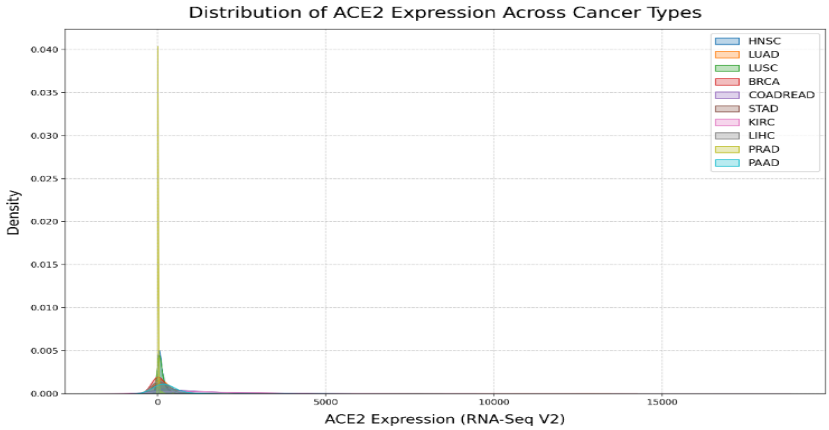


**Figure 3.** shows that the density plot reveals different distribution patterns of ACE2 expr ession across cancer types. Some cancer types, such as KIRC and COADREAD, exhibit higher overall expression levels with wider distributions.

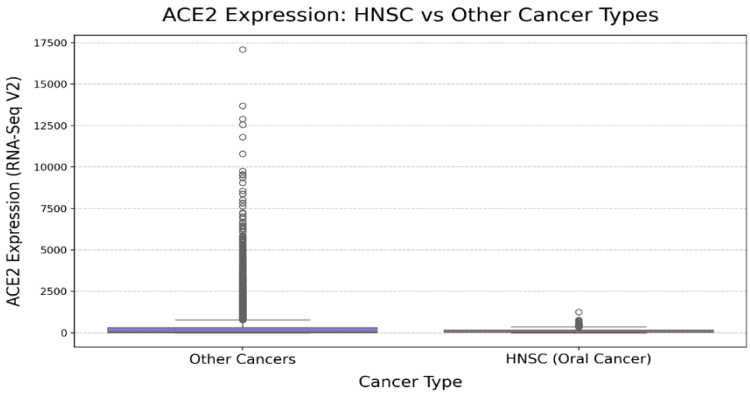


**Figure 4.** compares Head and neck squamous cell carcinoma (HNSC) with other cancer types.

The study identified ACE2 expression levels varies substantially across most cancer types, with some exceptions where cancers share similar regulatory environments and tissue origin. The strong significance in pairwise differences involving KIRC, COADREAD, and PRAD reinforces their distinctive ACE2 expression profiles.

This visualization supports the study's conclusion that ACE2 expression is highly heterogeneous across cancers and supports further exploration of its implications for cancer biology (**Fig 6**).

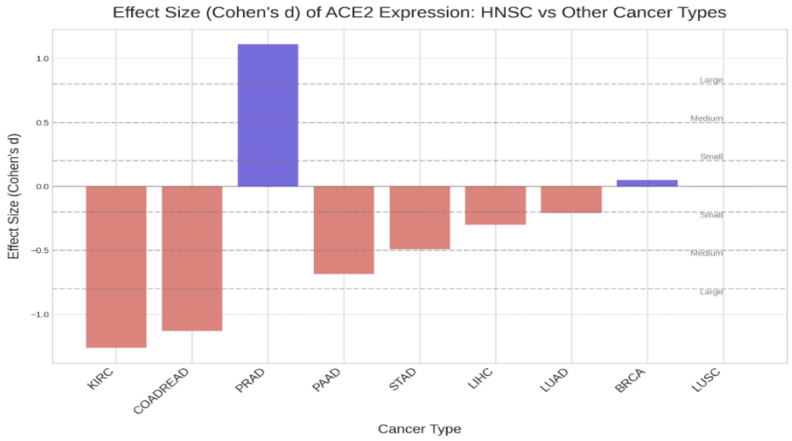


**Figure 5.** Effect size (Cohen's d) of ACE 2 expression: HNSC vs other cancer types

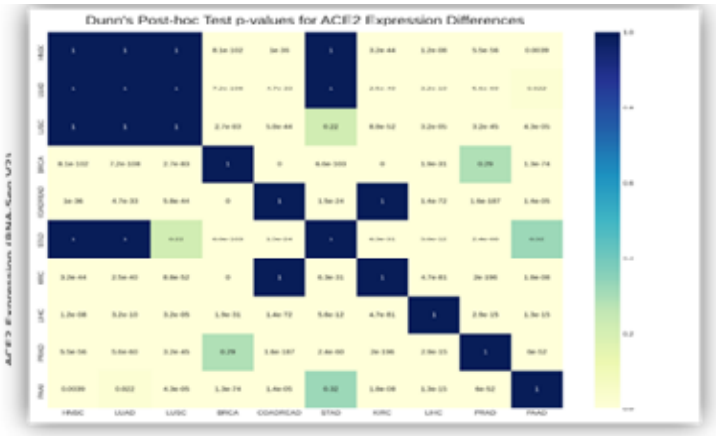


**Figure 6.** 10 cancer types (HNSC, LUAD, LUSC, BRCA, COADREAD, STAD, KIRC, LIHC, PRAD, PAAD), which found that most cancer types show significant differences in ACE2 expression compared to KIRC, COADREAD, and PRAD

**Clustering Analysis:** Kernel PCA transformation using Radial Basis Function (RBF) kernel successfully captured non-linear relationships in the data and demonstrated distinct clustering patterns between BRCA and HNSC groups (**Fig 7**). The clear separation of BRCA and HNSC indicates distinct ACE2 expression patterns, while the mixed cluster shows some internal structure, with subclusters for specific types like KIRC and COADREAD. These clustering patterns exhibits their biological significance blending with their tissue-specific ACE2 expression profiles.

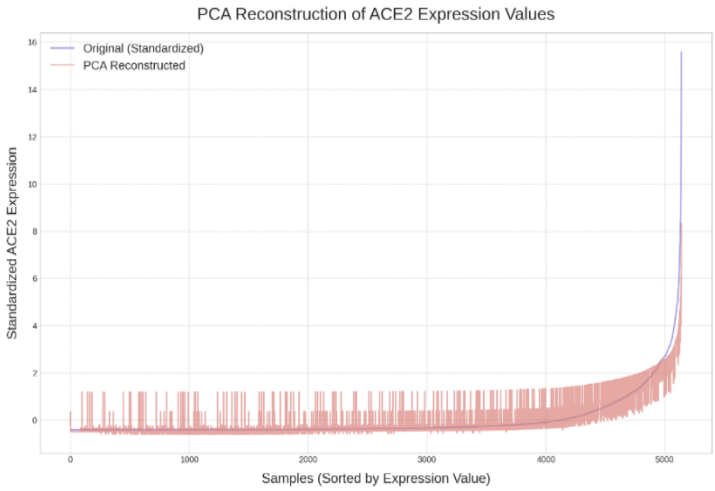


**Figure 7** shows PCA reconstruction of ACE 2 expression values

**Cluster Statistics:** The study identifies three distinct clusters of samples based on ACE2 expression, produced by K-means clustering from the data extracted from Kernel PCA, and the silhouette score of those clusters were 0.68 (**Tab 2**).

**Table 2** Percentage of Each Cancer Type in Each Cluster

| | **0** | **1** | **2** |
|---|---|---|---|
| BRCA | 100.0 | 0.17 | 0.0 |
| COADREAD | 0.0 | 16.68 | 0.0 |
| HNSC | 0.0 | 0.0 | 100.0 |
| KIRC | 0.0 | 14.37 | 0.0 |
| LIHC | 0.0 | 10.31 | 0.0 |
| LUAD | 0.0 | 14.37 | 0.0 |
| LUSC | 0.0 | 13.63 | 0.0 |
| PAAD | 0.0 | 4.99 | 0.0 |
| PRAD | 0.0 | 13.89 | 0.0 |

Cluster 1, which has the highest average ACE2 expression (602.96), Cluster 2, which has moderate expression with some variability, and Cluster 0, which has very low ACE2 expression (34.16). Cluster 0 consisted only BRCA samples, Cluster 2 was made up of HNSC samples, and Cluster 1 contains of all other cancer types. (**Fig 8,9)**

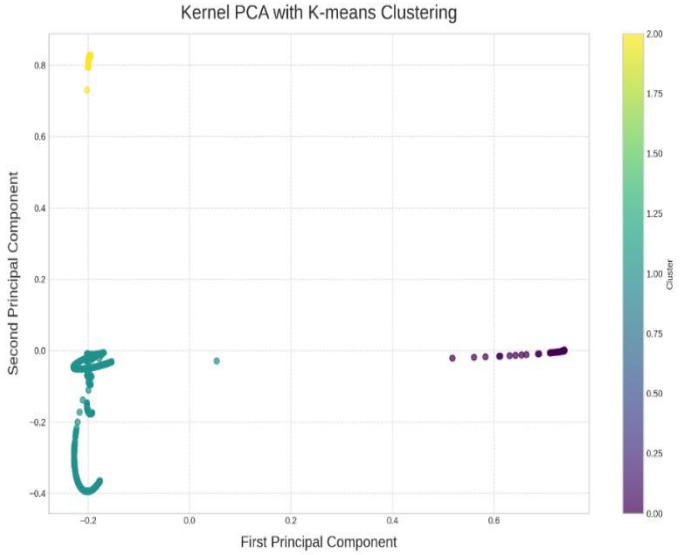


**Figure 8.** Kernel PCA plot reveals three clusters: one for BRCA samples, one for HNSC samples, and a mixed cluster of other cancer types.

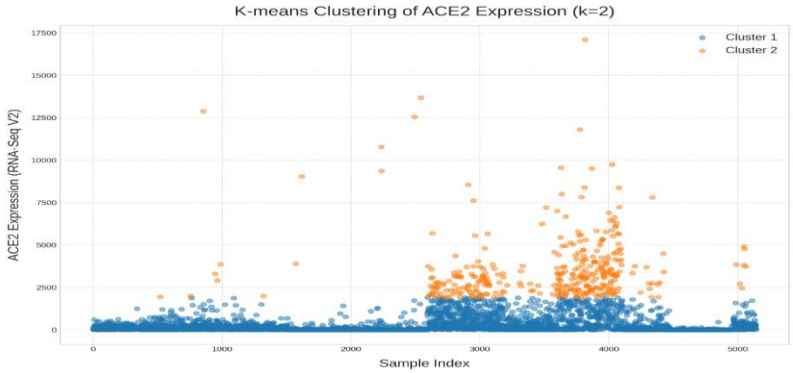


**Figure 9.** Bimodal distribution of ACE2 expression in cancers. Cluster 1 (Blue) includes most samples with low expression, while Cluster 2 captures a smaller group with high or outlier levels.

The clustering analysis based on ACE2 expression, reveals potential distinctions in cancer biology between the clusters (**Tab 3**, **Fig 10)**

**Table.3 Results presented in the clusters and provides valuable insights into the ACE2 expression levels across different groupings of samples.**

| | Cluster | Sample Count | Mean ACE2 Expression | Std Dev | Min | Max |
|---|---|---|---|---|---|---|
| 1 | 1 | 3550 | 602.96 | 1246.76 | -0.09 | 17099.9 |
| 2 | 2 | 515 | 121.12 | 134.88 | 0.0 | 1256.01 |
| 0 | 0 | 1076 | 34.16 | 137.43 | 0.0 | 1735.27 |

**Figure 10.** Mean ACE2 Expression by cluster

***Silhouette Scores:*** The highest silhouette score occurs at 3 clusters, indicating the best-defined structure. The score


decreases as the number of clusters increases, with a slight rebound at 9 clusters. Hence, the optimal number of clusters is 3, aligning with previous results. (**Fig11)**

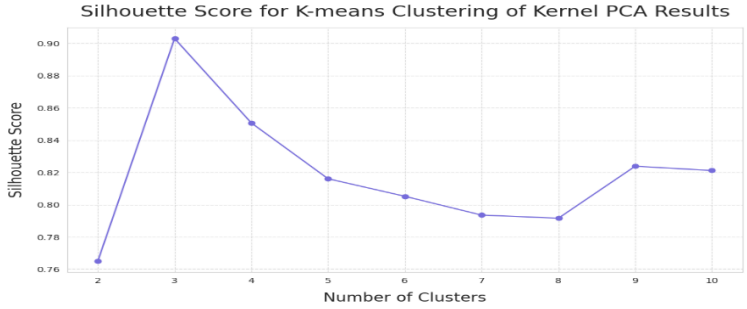


**Figure 11.** shows that the study uses the K-means algorithm to cluster ACE2 expression profiles across 10 cancer types into three biologically meaningful patterns.

**Cluster Cancer Composition:** Kernel Principal Component Analysis with an RBF kernel effectively processed the high-dimensional data into a two-dimensional capturing the non-linear relationships with an error of 0.2760 whereas KNN Imputation results were with higher error of 1.3901 and Iterative Imputation using MICE has an error of 0.5984. Final visualization by Kernel PCA showed distinct patterns and clusters that corresponded well with different types of cancer, Principal Component Analysis (PCA) and Truncated Singular Value Decomposition (SVD) report an error of 0.4369 (**Fig 8**). At the same time, non-negative matrix factorization (NMF) applied to the ACE2 expression records an error of 0.0000.

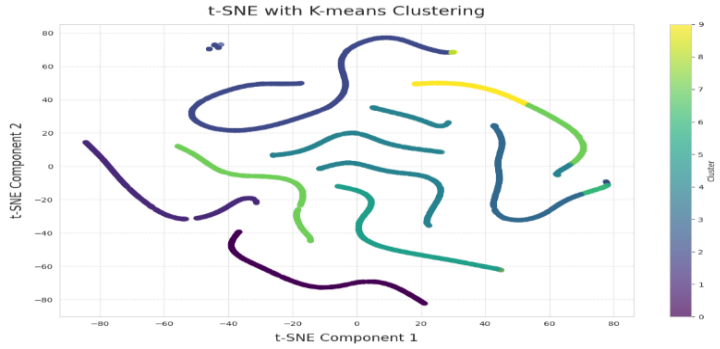


**Figure 12.** t-SNE with K-means Clustering

**Comparison with t-SNE:** The t-SNE visualization of the data showed analogous patterns as noted in Kernel PCA findings, particularly with the t-SNE clusters for BRCA and HNSC samples. In terms of interpretability and cluster separation, t-SNE offered a nonlinear projection method to contrast with Kernel PCA (**Fig 12**). While t-SNE offers a visually convincing data, Kernel PCA delivered greater comprehension mapping in terms of original feature space and the reduced dimensions. The difference between t-SNE and Kernel PCA indicates need to balance between the impact of visualization and the interpretability of the data in dimension reduction techniques.

**Kernel Comparison:** The Radial Basis Function (RBF) kernel showed the best separation of cancer types, resulting in compact and well-defined clusters while preserving both local and global structure. The sigmoid kernel had moderate clustering performance and was less effective in capturing underlying biological variability. The linear kernel, while basic and interpretable, and could not capture the non-linear correlation data. Though polynomial kernel could account higher variance, it was compact and less interpretable. The cosine kernel captured angular relationships well, but couldn't cope up with variations in expression magnitude. Further kernels demonstrated modest separation but unable clear or biologically relevant clusters. (**Fig 13**).

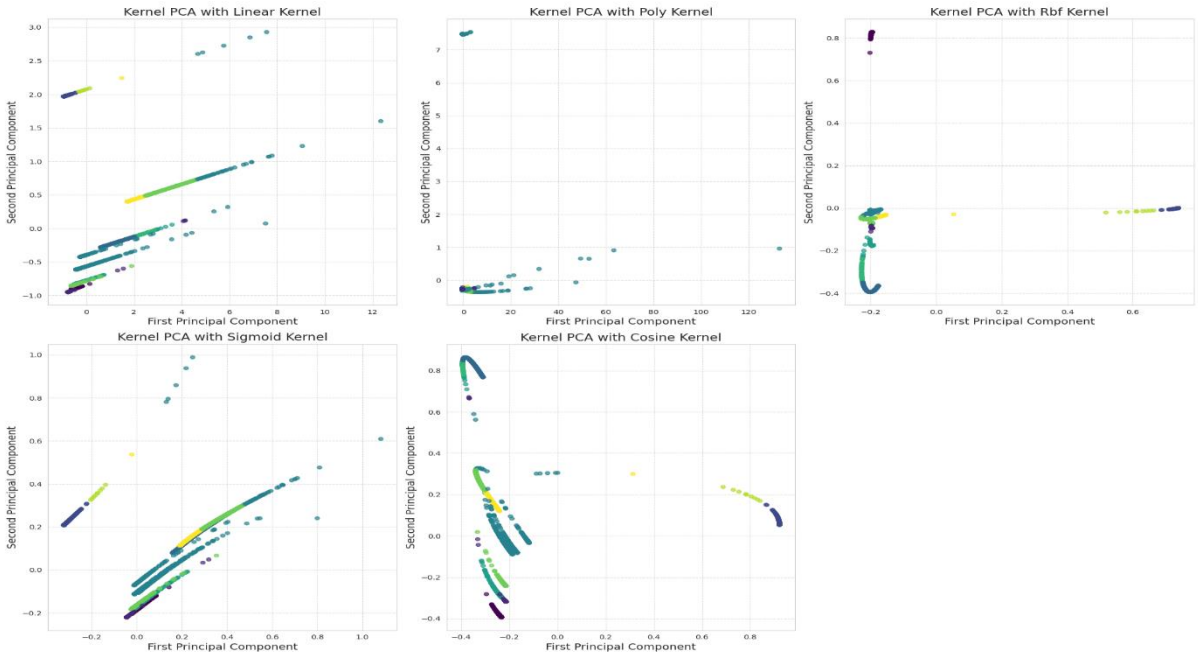


**Figure 13.** illustrates various kernel functions used in Kernel PCA for dimensionality reduction across data on cancer expression.

As per these findings, Kernel PCA has still the best overall reconstruction performance across all the dimensionality reduction techniques with the least reconstruction error of 0.2760. Non-negative Factorization (NMF) showed the best reconstruction for ACE2 expression values, which proves its importance for investigation of gene expression as it retains non-negativity, a crucial characteristic of biological data.The analysis of ACE2 expression in different cancer types uncovered a unique pattern in oral cancer, especially in head and neck squamous cell carcinoma (HNSC). HNSC exhibited a moderate level of ACE2 expression, creating a distinct cluster in the Kernel PCA analysis. This discrete ACE2 expression pattern of oral, head and neck cancers differentiate HNSC from other cancer types. The relatively stable ACE2 expression pattern that is observed in HNSC during the analysis suggests particular regulatory mechanisms that are modulated by environmental, molecular traits and epigenetic factors of HNSC.

## DISCUSSION

Angiotensin-Converting Enzyme 2 plays predominant role in maintaining cardiovascular, pulmonary homeostasis. The biological significance of ACE2 drastically increased, during COVID-19 pandemic, as ACE2 remained the primary entry receptor for the SARS-CoV-2, which facilitated viral invasion into host cells. In addition to its importance in virion penetration, ACE2 related pathways have been linked to numerous human cancers. Over the past decades, several studies have explored the role of ACE 2 in various malignancies [2,10,12,14,15]. Accordingly, studies have shown that, ACE2 is upregulated in epithelial-derived cancers, particularly those that involve mucosal linings such as the respiratory and gastrointestinal systems [13,14,16]. Experimental studies have demonstrated that down-regulation of ACE 2 promotes cancer metastasis [17]. By modulating significant molecular pathways linked to angiogenesis, the epithelial–mesenchymal transition (EMT), cell proliferation, and immune regulation, ACE2 may play a role in tumor growth, invasion, and metastasis. Hence, it is necessary to understand the pattern of ACE2 expression in different cancer types.

An earlier ELISA-based study revealed that ACE2 protein expression was significantly lower in OSCC tissues compared to normal oral tissues [18].

However, contrastingly, RT-PCR analysis shows mRNA expression levels of ACE2 were found to be elevated in OSCC samples. These results suggest possible post-transcriptional regulation, protein degradation, or translational inhibition mechanisms affecting ACE2 in oral squamous cell carcinoma. Wherefore in the present study we emphasized on ACE2 expression in Head and Neck Squamous Cell Carcinoma, and to analyse, compare the expression among HNSC and different cancers.

In this study, we analysed the ACE2 expression levels in ten different cancer types by using the data from high-throughput genomic repositories like Cancer Genome Atlas (TCGA), GEPIA2, and Oncomine. Analysis shows that ACE2 expression varies among the tumor types that includes lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), kidney renal papillary cell carcinoma (KIRP), colon adenocarcinoma (COAD), and oral squamous cell carcinoma (OSCC). The distribution of ACE2 expression is represented in highly right-skewed pattern, with a median value of 52.35 and the maximum value was 17099.9. The Dunn post hoc p value for ACE2 expression (***Fig 6***) shows that ACE2 differs greatly amongst cancer types, with colorectal adenocarcinoma (COADREAD) and kidney renal clear cell carcinoma (KIRC) showing the highest levels. On the other hand, breast cancer (BRCA) and prostate adenocarcinoma (PRAD) displayed the lowest. ACE2 expression was found to be moderate in head and neck squamous cell carcinoma (HNSC).

Cancer prediction models better performed on dimensionally reduced data. Nonlinear dimensionality reduction techniques provide greater performance than linear and sublinear techniques on high dimensional continuous data like gene expression data [19,20]. This study evaluated various non-linear dimensional reduction approaches to determine appropriate dimensional reduction methods for high-dimensional gene and protein expression profiling. Methods like kernel-based PCA, and principal component analysis (PCA) were used for the dimensional reduction of data. Kernel PCA employing an RBF kernel technique performed better than other kernels and conventional PCA in maintaining non-linear relationships, with a lower reconstruction error (0.2760) and a more distinct visual separation of ACE2 expression clusters (***Fig. 2, 3, 4, 5, 6***). Though t-SNE revealed consistent patterns, Kernel PCA offered superior global interpretability (***Fig. 7, 8, 9, 10, 11***). Cluster analysis of ACE2 in HNSC created a unique cluster pattern in the Kernel PCA space and Mann-Whitney U test with the HNSC data demonstrated statistically significant difference from other cancers ($p = 2.169 \times 10^{5}$). The ACE2 expression distribution of cancer types were supported by a peak silhouette score at k=3, and K-means clustering on Kernel PCA-transformed data revealed three distinct clusters that corresponded to low (Cluster 0), moderate (Cluster 2), and high (Cluster 1) ACE2 expression. The study suggests Kernel Principal Component Analysis (Kernel PCA) with a Radial Basis Function (RBF) kernel technique for identifying non-linear ACE2 expression patterns in various cancer types [12,14,19]. The study findings indicate ACE2's potential role in tumor stratification, viral vulnerability profiling, and biomarker-driven cancer grouping.

**Limitations**

The study has few limitations. The study utilized bulk tumor samples, which has limited generalizability. The study has considered only 10 cancers, unstudied cancer types to be included in future for profiling. Furthermore, we took only three non-linear dimensional reduction algorithms. Hence, studies including more cancer types, larger cohorts, multi-omics data, single-cell RNA sequencing methods, clinical information integration, functional studies and other non-linear dimensional reduction algorithms to be carried out to validate our findings.

## CONCLUSION

The study used Kernel Principal Component Analysis (Kernel PCA) to examine ACE2 gene expression data from various cancer types, pinpointing unique patterns that distinguish them, particularly between breast and head and neck squamous cell carcinoma. RBF kernel algorithm shows greater capacity in capturing non-linear relationships demonstrating the intricate regulation of gene expression throughout the cancer profiling. Our findings highlight the significant variability in levels and distribution pattern of ACE2 expression among cancer types. Head and Neck Squamous Cell Carcinoma (HNSC) was identified as unique molecular subtype, exhibiting distinct ACE2 expression profile. Future investigations using non-linear dimensionality reduction technique could deepen our understanding of ACE2 in cancer profiling that aids in precision oncology, risk-stratification, management strategies for cancers and associated infectious diseases.

## DECLARATION

**Funding**

This study received no particular grants from public, commercial, or non-profit funding entities.

**Declaration of competing interest**

The authors declare that they have no known competing interests.

## ABBREVATION LIST

ACE2: Angiotensin-converting enzyme 2
HNSC : Head And Neck Squamous Cell Carcinoma
TCGA: The Cancer Genome Atlas
RBF: Radial Basis Function
PCA: Principal Component Analysis
t-SNE : t-distributed Stochastic Neighbour Embedding
KIRC: Kidney Renal Clear Cell Carcinoma
COADREAD: Colorectal Adenocarcinoma
BRCA: Breast Carcinoma
PRAD: Prostate Adenocarcinoma
SARS-CoV-2:Severe acute respiratory syndrome coronavirus 2
ML: Machine Learning
NDR: Nonlinear Dimensionality Reduction
GTEx:Genotype-Tissue Expression database
CCLE:Cancer Cell Line Encyclopedia
GEPIA2:Gene Expression Profiling Interactive Analysis
DNMT: DNA Methyltransferase
MMR: Mismatch Repair Genes
PAAD: Pancreatic Adenocarcinoma
STAD: Stomach Adenocarcinoma
LUAD: Lung Adenocarcinoma
LUSU: Lung Squamous Cell Carcinoma
LIHC: Liver Hepatocellular Carcinoma
SVD: Truncated Singular Value Decomposition
NMF: Non-Negative Matrix Factorization
EMT : Epithelial–Mesenchymal Transition

## REFERENCES

1. Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J. C., Turner, A. J., Raizada, M. K., Grant, M. B., & Oudit, G. Y. (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th Anniversary of the Discovery of ACE2. *Circulation research*, *126*(10),1456–1474. https://doi.org/10.1161/CIRCRESAHA.120.317015
2. Jaiswal, A., Shrivastav, S., Kushwaha, H.R. *et al*. Oncogenic potential of SARS-CoV-2—targeting hallmarks of cancer pathways. *Cell Commun Signal* 22,447 (2024). https://doi.org/10.1186/s12964-024-01818-0
3. Bhola, A., & Tiwari, A. K. (2015). Machine learning based approaches for cancer classification using gene expression data. *Machine Learning and Applications: An International Journal*, *2*(3/4), 01-12.
4. Adiwijaya, W. U., Lisnawati, E., Aditsania, A., & Kusumo, D. S. (2018). Dimensionality reduction using principal component analysis for cancer detection based on microarray data classification. *Journal of Computer Science*, *14*(11), 1521-1530.
5. Malepathirana, T., Senanayake, D., Vidanaarachchi, R., Gautam, V., & Halgamuge, S. (2022). Dimensionality reduction for visualizing high-dimensional biological data. *Bio Systems*, *220*, 104749.https://doi.org/10.1016/j.biosystems.2022.104749
6. Manthena, V., Jarquín, D., Varshney, R. K., Roorkiwal, M., Dixit, G. P., Bharadwaj, C., & Howard, R. (2022). Evaluating dimensionality reduction for genomic prediction. *Frontiers in genetics*, *13*,958780. https://doi.org/10.3389/fgene.2022.958780
7. Reetu Sharma,Cluster analysis to identify prominent patterns of anti-hypertensives: A three-tiered unsupervised learning approach,Informatics in Medicine Unlocked,Volume 19,2020,100303,ISSN 23529148,https://doi.org/10.1016/j.imu.2020.100303.
8. Du, X. (2023). A Robust and High-Dimensional Clustering Algorithm Based on Feature Weight and

9. Entropy. *Entropy*, *25*(3),510. https://doi.org/10.3390/e25030510
10. Shi, J., & Luo, Z. (2010). Nonlinear dimensionality reduction of gene expression data for visualization and clustering analysis of cancer tissue samples. *Computers in biology and medicine*, *40*(8),723–732. https://doi.org/10.1016/j.compbiomed.2010.06.007
11. Song, J., Han, J., Liu, F., Chen, X., Qian, S., Wang, Y., Jia, Z., Duan, X., Zhang, X., & Zhu, J. (2020). Systematic Analysis of Coronavirus Disease 2019 (COVID-19) Receptor ACE2 in Malignant Tumors: Pan-Cancer Analysis. *Frontiers in molecular biosciences*, *7*, 569414. https://doi.org/10.3389/fmolb.2020.569414
12. Kim, Kihun, Ko, Yeji, Ko, Dai Sik, Kim, Yun Hak, Prognostic Significance of COVID-19 Receptor ACE2 and Recommendation for Antihypertensive Drug in Renal Cell Carcinoma, *BioMed Research International*, 2020, 2054376, 7 pages,2020 **https://doi.org/10.1155/2020/2054376**
13. Tang, Q., Wang, Y., Ou, L., Li, J., Zheng, K., Zhan, H., Gu, J., Zhou, G., Xie, S., Zhang, J., Huang, W., Wang, S., Wang, X. (2021). Downregulation of ACE2 expression by SARS-CoV-2 worsens the prognosis of KIRC and KIRP patients via metabolism and immunoregulation. *International Journal of Biological Sciences*, 17(8), 1925-1939. https://doi.org/10.7150/ijbs.57802.
14. Errarte, P., Beitia, M., Perez, I., Manterola, L., Lawrie, C. H., Solano-Iturri, J. D., Calvete-Candenas, J., Unda, M., López, J. I., & Larrinaga, G. (2017). Expression and activity of angiotensin-regulating enzymes is associated with prognostic outcome in clear cell renal cell carcinoma patients. *PloS one*, *12*(8), e0181711. https://doi.org/10.1371/journal.pone.0181711
15. Yang, J., Li, H., Hu, S., & Zhou, Y. (2020). ACE2 correlated with immune infiltration serves as a prognostic biomarker in endometrial carcinoma and renal papillary cell carcinoma: implication for COVID-19. *Aging*, *12*(8), 6518–6535. https://doi.org/10.18632/aging.103100
16. Bhonde, S. B., & Prasad, D. J. R. (2021). Performance Analysis of Dimensionality Reduction Techniques in Cancer Detection using Microarray Data. *Asian Journal For Convergence In Technology (AJCT) ISSN -2350-1146*, *7*(1), 53-57. https://doi.org/10.33130/AJCT.2021v07i01.012
17. Gheware, A., Ray, A., Rana, D. *et al.* ACE2 protein expression in lung tissues of severe COVID-19 infection. *Sci Rep* **12**, 4058 (2022). https://doi.org/10.1038/s41598-022-07918-6
18. Yu, C., Tang, W., Wang, Y., Shen, Q., Wang, B., Cai, C., Meng, X., & Zou, F. (2016). Downregulation of ACE2/Ang-(1-7)/Mas axis promotes breast cancer metastasis by enhancing store-operated calcium entry. *Cancer letters*, *376*(2), 268–277. https://doi.org/10.1016/j.canlet.2016.04.006
19. Shivani Sivasakthivel, Pratibha Ramani, Selvaraj Jayaraman,Expression of angiotensin converting enzyme 2 in patients with oral squamous cell carcinoma,Oral Oncology Reports,Volume 10,2024,100476,ISSN2772-9060, https://doi.org/10.1016/j.oor.2024.100476.
20. 18.Md Faisal Kabir, Tianjie Chen, Simone A. Ludwig,A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction,HealthcareAnalytics,Volume 3,2023,100125,ISSN2772-4425, https://doi.org/10.1016/j.health.2022.100125.
21. 19.Vogelstein, J. T., Bridgeford, E. W., Tang, M., Zheng, D., Douville, C., Burns, R., & Maggioni, M. (2021). Supervised dimensionality reduction for big data. *Nature communications*, *12*(1), 2872. https://doi.org/10.1038/s41467-021-23102-2