IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

HPV E6 May Affect PACSIN2 Expression in Cervical Cancer Independently from Known Transcription Factors

Allison Moores

The Bishop's School, 7607 La Jolla Blvd, La Jolla, CA 92037, USA; allison.moores.rsf@gmail.com
Mentor: David Laub

ABSTRACT: Cervical cancer is the 4th most common cancer in women. While the roles of E6 and E7 human papillomavirus (HPV) oncogenes in disrupting cell cycle control are known, the specific impact of HPV on immune gene regulation in cervical cancer is not fully understood. We hypothesized that if HPV genes regulate PACSIN2 expression via mechanisms independent of PACSIN2 transcription factors (TFs), a predictive model of PACSIN2 expression would have substantially better performance when using both TFs and HPV genes as input than either alone. Decision trees and XGBoost were used to understand the regulatory mechanisms of HPV genes and TFs on immune response genes. The XGBoost model, the best performer for TF-only input, yielded a Pearson correlation coefficient of 0.52; when both TF and HPV genes were included, a correlation of 0.69 in the XGBoost model was achieved, indicating a notably enhanced predictive accuracy due to the inclusion of HPV genes. The HPV gene E6 may exert a substantial regulatory influence on PACSIN2 expression, encouraging a path for further investigation. This influence was independent of the regulation by known PACSIN2 TFs. In addition, interactions between E6 and PACSIN2 TFs are likely mediated via E6 downregulating p53.

KEYWORDS: Computational Biology and Bioinformatics; Genomics; HPV; Gene Expression; Machine Learning.

Introduction

Cervical cancer is a significant global health concern. It currently affects over 604,000 women worldwide and resulted in 342,000 deaths in 2020.[1] Currently, over 90% of cervical cancers are due to infection from the human papillomavirus (HPV).[2] HPV is transmitted sexually and can cause various conditions ranging from lesions to the development of microinvasion[3] and may eventually result in the development of cervical cancer. There are over 200 distinct HPV strains that infect epithelial cells,[4] and about 40 of these show a preference for mucosal tissues. Depending on their ability to cause cancer, they are further classified as low-risk and high-risk HPV (lr-HPV and hr-HPV, respectively).[5] Worldwide, around 70% of cervical cancer cases are caused by hr-HPV, primarily HPV-16 and HPV-18.[6,7] While HPV vaccination is available, its effectiveness is limited to specific strains, and its broader impact has not yet been confirmed.[8]

Despite substantial research, the exact molecular mechanisms through which HPV initiates and advances cervical cancer remain unclear. Current understanding focuses on HPV viral oncogenes E6 and E7. It has been found that when the viral genome integrates into the host DNA genome, E6 and E7 are upregulated.[9] This, in turn, deregulates critical proteins in the cell signaling pathways, such as inhibition of tumor suppressor proteins, p53, and pRb,[10] causing cell cycle control impairment and genomic instability. However, this view is limited as it does not fully encompass the intricate interactions between HPV and host gene regulation.

The complexity of HPV's role in cervical carcinogenesis necessitates further exploration, which is pivotal for enhancing prevention and treatment strategies. Identifying markers that predict cervical cancer development in HPV-infected individuals could revolutionize preventive measures, potentially through specific biomarker development. Moreover, a deeper insight into HPV-host molecular interactions could lead to the design of targeted therapeutics, thus advancing personalized medicine approaches in cancer treatment.

This study hypothesizes that HPV manipulates the transcriptional environment in cervical cells, fostering a carcinogenic milieu. It is speculated that HPV selectively regulates human genes in these cells, leading to changes conducive to cancer development and progression. To investigate this hypothesis, the study employs computational techniques, including machine learning and bioinformatics, aiming to elucidate the specific gene regulation mechanisms at play in HPV-induced cervical cancer.

The current research seeks to fill the knowledge gap in understanding HPV's regulation of human gene expression, potentially unveiling new insights and strategies for combating cervical cancer.

Methods

Data Collection and Standardization:

The study used the whole gene expression identified by RNA-sequencing (RNA-Seq) of The Cancer Genome Atlas Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA-CESC)[11] dataset, which includes a substantial number of patient samples for cervical cancer human gene expression. The dataset was accessible through Recount3,[12] which provides access to a large amount of uniformly processed RNA-seq data from human and mice

DOI: 10.36838/v6i10.5

samples, which go beyond gene counts. Overall, our dataset included 244 cervical cancer patients.

Matching HPV gene expression data was sourced from a study by Ren *et al.* (2020)[13] for HPV strains 16, 33, and 35 and genes E1, E2, E4, E5, E6, E7, L1, and L2. During the development of cervical cancer, several abnormal events are triggered, including perturbation of antitumor immune response, alteration of gene expression, disruption of cell cycle control, and deregulation of microRNA.[14] Therefore, we also included 2,155 immune-related (IR) genes identified in a 2023 study by Li *et al.*,[15] along with their transcription factors from the Gene Regulatory Network database GRNdb.[16] After conducting deduplication and filtering for active HPV infections, we arrived at a dataset of 136 patients, which was split into the train (80%), validation (10%), and test (10%) groups.

Gene Expression Normalization:

We removed duplicate patient entries and lowly expressed genes as a quality control method. Genes with very low counts may be measurement artifacts and add noise to downstream analyses. We examined expression levels of early expressed HPV genes (E1, E2, E4, E5, E6, E7, L1, and L2) across the entire patient cohort and cases with minimal HPV gene expression, resembling background noise (threshold set to 0.5), were excluded (Figure 1). We used gene expression histograms to establish criteria for active HPV infection and eliminate background noise (Figure 1).

We harmonized the datasets by matching patient identifiers using TCGA barcodes. Furthermore, we used Log Transcripts Per Million (TPM) to perform within-sample normalization of the original raw gene counts. TPM corrects for gene length and library size by dividing gene counts by gene length (in kb) to get transcript counts and then dividing by the total number of transcripts in the sample, resulting in each normalized sample having the same number of total counts[17] (equation 1). Log TPM normalizes counts and stabilizes the variance in the data, which is essential for downstream visualization, analysis, and modeling. We sourced gene length data from Ensembl[18] and The Papillomavirus Episteme (PaVE) for an accurate comparison.[19]

$$TPM_g = \frac{r_g/l_g}{\Sigma_i^n r_i/l_i} \times 10^6$$

For gene g, rg is the number of reads, lg is the transcript, and *n* is the number of genes. We used log TPM >1 in 90% of the patients as a cutoff for selecting IR genes.

Selection of Relevant Genes:

Tumor purity refers to the proportion of tumor cells in the tumor microenvironment (TME). TME comprises different cell types, including endothelial cells, immune cells, and fibroblasts, along with extracellular components such as hormones, cytokines, extracellular matrix, and growth factors.[20] The TME not only plays a pivotal role during tumor initiation and metastasis, but it also affects therapeutic efficacy. Studies have shown that low-risk patients are more sensitive to immunotherapy because of the correlation between tumor purity and immunity.[21] HPV is involved in the recruitment of immune cells within tumors, with HPV-positive tumors showing a stronger activation of immune signaling pathways.[22] If tumor purity affects IR gene expression, we would expect HPV- and HPV+ cases to have differential IR gene expression. Thus, to rule out cases where tumor purity may be confounding IR gene expression, we found genes that exhibited no significant differential expression between HPV-negative and HPV-positive tumors, applying the Mann-Whitney U test with a threshold of p-value greater than 0.5.

Subsequently, we excluded lowly expressed genes that could have a differential expression due to noise and focused on well-expressed genes, characterized by a median log2 TPM (Transcripts Per Million) value exceeding 1. Next, we selected genes showing the effect of the HPV E6 protein on IR gene expression, identified by an absolute slope value greater than 0.2. After applying those filters, we considered the top and bottom 10 IR genes by Pearson r between HPV E6 and IR gene expression (Figure 2). We removed genes specific to immune cell types to account for noise from tumor purity based on annotations from the Human Protein Atlas.[23] Of the remaining genes, we prioritized genes with some known impact on immune function, leaving us with PACSIN2, NDUFS1, APOE, and WNT5B.

Finally, looking at the correlation matrix, we excluded HPV genes with similar gene expression to reduce redundancy. We excluded E7 while retaining E6 since studies have shown that the E7 is translated from spliced E6*I fusion transcripts.[24] Similarly, we excluded E4 and retained E2 since both transcripts are up-regulated by keratinocyte differentiation and give rise to two fusion proteins *in vitro*, E2^E4-S and E2^E4-L.[25] Our final HPV gene selection included E1, E2, E5, and E6.

Predictive Modeling – Decision Tree:

The initial modeling approach was based on decision trees (DT).[26] DTs are a non-parametric supervised learning method used for classification and regression. They form branches based on criteria within nodes and decisions that can be based on categorical or continuous data.[26] The process involves building trees from bootstrapped training data such that each split is based on a random sample of m variables from a full set of n attributes.[26] However, the model runs the inherent risk of overfitting when it fits to noise or misleading points in the input distribution, leading to poor performance and generalization. To handle this, DTs are pruned to reduce the tree size without losing predictive accuracy. We used a cost complexity pruning algorithm parameterized by a complexity parameter $\alpha \geq 0$. The complexity parameter is used to define the cost-complexity measure, $R_a(T)$ of a given tree T as:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|$$

Where $|\tilde{T}|$ is the number of terminal nodes in T and R (T) is the total misclassification rate of the terminal nodes. The cost complexity measure of a single node is $R_\alpha(T) = R(T) + \alpha$. The inputs to the DT have selected HPV genes together/singularly with TF Factors used to predict the immune response

(output IR genes). We used 80% of our dataset for training, 10% for validation, and 10% for testing.

The eXtreme Gradient Boosting (XGBoost) Algorithm:

The XGBoost model iteratively builds competent models from a collection of distinct weak learners–initially, all samples in the dataset are given the same weight. The first model was trained by arbitrarily selecting samples from the dataset to give each sample an equivalent chance to contribute to training.[27] All models were then tested, and the weight of the incorrectly classified cases was updated such that they were chosen for the training of the subsequent model. The XGBoost algorithm has the advantage of strong generalizability and fast operation speed compared to other classifiers.[28] This advanced model underwent fine-tuning through a random search methodology, incorporating cross-validation across over 100 iterations. We fine-tuned the XGBoost algorithm to search randomly across different distributions of tree depth, weights, and loss reduction. Our maximum tree depth range was between 2 and 6 for base learners, the minimum sum of instance weight (hessian) between 0 and 15, and minimum loss reduction between a uniform distribution of 0.9 and 0.2 (required to make a further partition on a leaf node of the tree) as parameters to the model.

We used the XGBoost algorithm trained on HPV gene expression with and without TF Factor expression to predict IR gene expression. We used SHAP (SHapley Additive exPlanations) values to determine feature importance and visualize how each feature affects final model predictions.

Statistical Analysis:

Our analysis used the SciPy version 3.10 software package to conduct statistical evaluations. We used the Pearson r to assess correlation and the Mann-Whitney U test to compare differences between two independent groups. The p-value less than 0.05 was regarded as statistically significant.

Results

Out of the 244 patients available in The Cancer Genome Atlas (TCGA) Cervical Squamous Cell Carcinoma (CESC) cohort, we utilized data from 136 patients. After deduplication, we used the distribution of HPV gene expression to establish criteria for active HPV infection and to eliminate background noise (Figure 1). A cutoff of 0.5 log TPM was chosen, yielding 136 patients. Because of the low level of systematic expression, we excluded L1 and L2 from further analysis (Figure 1).

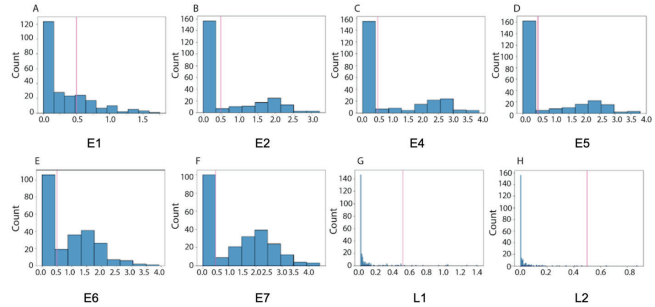

Figure 1: Distribution of HPV gene expression in log TPM across TCGA CESC cohort. Vertical red lines indicate cutoffs for determining whether patients had active HPV infections. Panel A shows gene expression for HPV genes E1, E2, E4, E5, E6, E7, L1, and L2. Most L1 (Panel G) and L2 (Panel H) gene expressions are lower than 0.5 log TPM, so they were excluded from further analysis.

To identify immune-related (IR) genes most likely to be regulated by HPV genes, we first considered the top 10 IR genes most and least co-expressed with HPV genes. We also selected genes not differentially expressed between HPV- and HPV+ patients and were sufficiently expressed across the cohort (Methods: Selection of Relevant Genes). Finally, we excluded IR genes exclusively expressed in immune cells, leaving PACSIN2 as the top candidate IR gene.

Next, to define the input features for models of PACSIN2 gene expression, we extracted known transcription factors (TF) of PACSIN2 in the context of CESC from GRNdb.[16] This yielded SREBF2, XBP1, CTNNB1, and HNF1A. To define HPV-related input, we identified a non-redundant set of HPV genes, E1, E2, E5, and E6, based on their co-expression patterns with IR genes (Figure 2). The log TPM expression of these eight genes served as the final input for models of PACSIN2 expression.

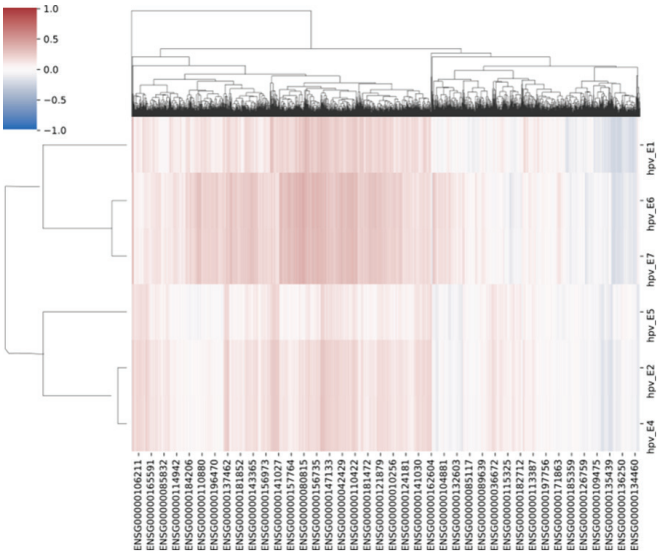

Figure 2: Pearson correlation between IR genes (x-axis) and HPV genes (y-axis), denoted by color. Rows and columns are hierarchically clustered, shown in the dendrograms.

DOI: 10.36838/v6i10.5

Modeling:

After identifying PACSIN2 as the most likely IR gene regulated by HPV, we trained a decision tree (DT) and an extreme gradient boosted tree (XGBoost) to predict PACSIN2 expression using the expression of its TFs and HPV genes as input. We hypothesized that if complex interactions occur between TFs and HPV genes to regulate PACSIN2 expression, the XGBoost model would outperform the DT. In addition, we trained three versions of each model class, using either TFs, HPV genes, or both as input. We hypothesized that if HPV genes regulate PACSIN2 expression via mechanisms independent of PACSIN2 TFs, the models would have substantially better performance when using both TFs and HPV genes as input than either alone.

First, we optimized the DT for maximal performance by measuring the coefficient of determination as a function of the DT's cost-complexity pruning parameter, α (Figure 3). As α increases, more of the tree is pruned until only a single leaf remains. The optimal α for the DTs using TFs only, HPV genes only, and both as input were 0.006, 0.013, and 0.004, respectively.

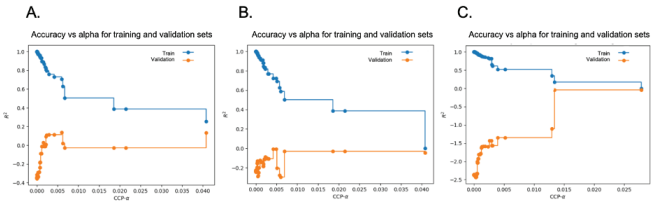

Figure 3: Coefficient of determinant versus cost-complexity pruning α for models using (A) TFs, (B) HPV genes, and (C) both TF and HPV genes as input.

Figure 4 shows the performance of the DT model prediction. The r-value represents Pearson's correlation coefficient between actual and predicted data. We find the biggest correlation (best performance) between predicted and actual (measured) data for combined TF and HPV input DTs (Figure 4C, r = 0.4847).

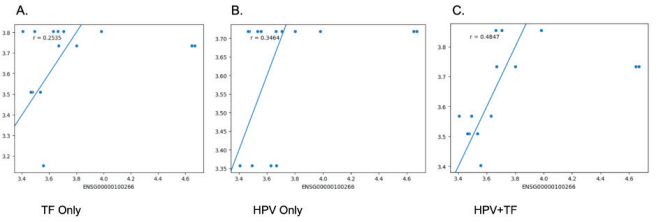

Figure 4: DT predicted vs. measured expression (log TPM) with Pearson's correlation coefficient shown. (A) performance for TF-only input, (B) performance for HPV-only input, and (C) performance using both HPV and TF genes as input.

The XGBoost model trained only on TF has a Pearson correlation coefficient of 0.52 for predicted and actual data for the best estimator (Figure 5A). For the HPV-only training set, R = 0.35 (Figure 5B), and for HPV and TF, the R = 0.69 (Figure 5C).

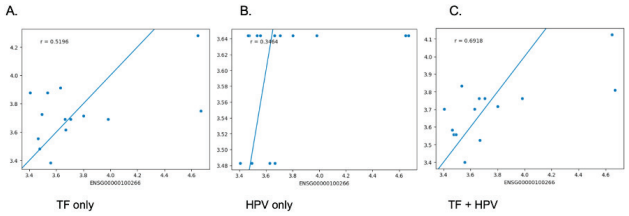

Figure 5: Pearson's correlation coefficient for predicted and actual data for the XGBoost model. Panels A, B, and C show Pearson's r as the best estimator for TF-only, HPV-only, and TF and HPV inputs, respectively.

We computed SHAP values for all patients in the test set to interpret the best-performing model. A plot of feature importance (Figure 6A) shows that the most important features for the XGBoost model with both HPV and TF genes as input are SREBF2, CTNNB1, E6, and XBP1. A wider spread of the dots for SREBF2, CTNNB1, and E6 indicates that these features substantially interact with others. So, we computed SHAP interaction values (Figure 6B) to identify important interactions. The strongest interactions between HPV genes and TFs are E6 and CTNNB1, E6 and SREBF2, and E6 and XBP1.

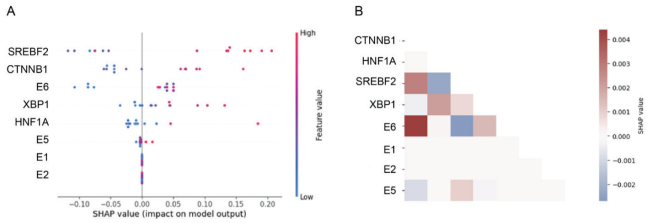

Figure 6: Panel A shows feature importance. Each dot represents the XGBoost model's prediction for a patient in the test data. Panel B shows the SHAP value for each feature. Each box represents the median SHAP interaction value across patients for all pairwise combinations of genes.

Discussion

We hypothesized that, besides its canonical role in down-regulating TP53 and Rb, HPV manipulates the transcriptional environment in cervical cells by regulating immune-related (IR) genes, resulting in immune evasion and enabling cancer progression. An understanding of how HPV contributes to immune evasion can provide new insights and strategies to combat cervical cancer. We found that E6, independently and with transcription factors (TF) SREBF2, CTNNB1, and XBP1, likely regulate PACSIN2 gene expression (Figure 6), suggesting future *in vitro* and *in vivo* studies into this relationship.

PACSIN2 is a protein kinase family member involved in endocytosis and links the actin cytoskeleton with vesicle formation. It has been previously shown that PACSIN2 is involved in cancer's collective cell migration, which affects cancer metastasis.[29] Additionally, PACSIN2 is a negative epidermal growth factor (EGF) regulator whose increased expression has been associated with cancer.[30] This suggests that regulating PACSIN2 by HPV may manipulate the host cell environment in favor of viral persistence and cancer progression.

SREBF2 is a key regulator of lipid metabolism that controls the transcription of sterol-regulated genes. Lipid metabolism disorders such as fatty acid disorder and cholesterol metabo-

lism disorder are the most prominent metabolic changes in cancer.[31] SREBF2 has also been found to be upregulated by mutant p53 in human breast cancer.[32] Given the interaction between E6 and SREBF2 and that E6 effectively downregulates p53, the data suggest that p53 may also regulate SREBF2 in cervical cancer. Future *in vitro* and *in vivo* studies must be conducted to test that hypothesis.

CTNNB1 encodes for beta-catenin, a protein involved in WNT signaling. WNT proteins form a signaling network that regulates tissue morphogenesis and homeostasis.[33] Activation of the WNT signaling pathway has been shown to play a critical role in colorectal carcinogenesis and is considered a therapeutic target to affect tumor progression.[34] It has also been shown that p53 activates MicroRNA-34 to inhibit WNT Signaling, which is crucial for p53 tumor suppression function.[35]

The X-box binding protein 1 XBP1 is a unique basic-region leucine zipper transcription factor. XBP1 plays an important role in regulating cell proliferation, apoptosis, metastasis, drug resistance, and thus, tumorigenesis and tumor progression.[36] It has been found that loss of p53 function activates the XBP1 pathway to enhance protein folding and activation of XBP1.[37] Thus, the XBP1 pathway is a target for treating chemoresistant tumors expressing mutant p53.

These findings indicate that E6 may regulate PACSIN2 expression primarily through mechanisms independent of known TFs. In addition, E6 may influence PACSIN2 TF expression via downregulation of TP53.

Limitations And Future Work:

We utilized GRNs to define PACSIN2 TFs, which have certain inherent limitations. The GRN may be incomplete, and other TFs may also regulate PACSIN2. In addition, the interactions predicted can be indirect (not reflecting physical interaction), and regulators that themselves do not change or are expressed at low levels will be missed.[38]

It would also be beneficial to validate our results in additional cervical cancer patient cohorts from databases such as the International Cancer Genome Consortium (ICGC), Hartwig Medical Foundation, Foundation Medicine, and Tempus. Additionally, a more precise analysis of HPV's mechanisms of immune-related gene regulation can be enabled via single-cell RNA sequencing. This would enable highly accurate disentanglement of differential expression of immune-related genes due to changes in the tumor immune microenvironment versus intracellular processes. Existing single-cell RNA-seq data from cervical cancer biopsies are an attractive next step to address this.

Furthermore, integrating other omics data, such as proteomics or metabolomics, could offer a more comprehensive view of the regulatory mechanisms at play. Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) studies have shown that upstream HPV-associated histone enrichment sites play central roles in cancer-associated pathways.[39] We can correlate these findings with our RNA-seq conclusions to further our research.

In addition, using artificial intelligence tools for structural prediction and advanced data analytics opens new avenues for research. For example, tools for protein structure prediction, such as AlphaFold, could be used to predict the structure of HPV proteins, especially in the presence of mutations. Prior studies have shown that E5 has a drastically different 3D structure between low-risk and high-risk genotypes[40] Protein docking simulations, especially between E6 and putative interacting partners, could yield valuable insights. For example, it is known that E6 can bind to PDZ-domain-containing proteins like hScrib, hDlg, MAGI-family proteins, MUPP1, PATJ, and PTPN3.[41]

■ Conclusion

Our study suggests that HPV genes E6 and TFs may exert a substantial regulatory influence on the IR gene PACSIN2. This finding substantiates our initial hypothesis, suggesting a potential intersection between HPV gene regulation, TF genes, and the pathogenesis of cervical cancer. Specifically, our results suggest that E6 regulates PACSIN2 expression independently of known TFs. We also find that E6, SREBF2, CTNNB1, and XBP1 all interact with p53 and that complex interplay likely contributes to regulating PACSIN2 expression and, hence, tumor development and progression. Future *in vitro* and *in vivo* studies must be conducted to test these findings.

Our findings suggest that HPV can independently influence host gene expression, separate from human transcription factors. The pathways identified in the study may play an important role in the molecular dynamics of HPV-associated cervical carcinogenesis and are worthy of further investigation. This research supports our initial hypothesis and also lays the groundwork for future investigations into targeted therapies and the development of more effective strategies for managing HPV-related cervical cancer.

■ Acknowledgments

We would like to thank Dr. Francis V. Chisari for his insightful comments.

■ References

1. *World Health Organization (2020) Cervical Cancer*. https://www.who.int/news-room/fact-sheets/detail/cervical-cancer?gclid=CjwKCAiA1-6sBhAoEiwArqlGPv3nlaCAL-7j3TfhGONtryShK-IZ0R- qkkTOWXp2LYlXKm6OfzOuAihoCjgsQAvD_BwE [Accessed 8 January 2024 (accessed 2020)
2. Guan, P.; Howell-Jones, R.;, Li, N.; Bruni, L.; de Sanjosé, S.; Franceschi, S.; *et al*. Human Papillomavirus Types in 115,789 HPV-Positive Women: A Meta-Analysis from Cervical Infection to Cancer. *Int. J. Cancer*. **2012**, *131* (10), 2349–2359. DOI: 10.1002/ijc.27485
3. Jung, W. W.; Chun, T.; Sul, D.; Hwang, K. W.; Kang, H. S.; Lee, D. J.; *et al*. Strategies Against Human Papillomavirus Infection and Cervical Cancer. *J. Microbiol*. **2004**, *42* (4), 255–266.
4. Doorbar, J. Molecular Biology of Human Papillomavirus Infection and Cervical Cancer. *Clin Sci*. **2006**, *110*, 525–541.
5. De Villiers, E. M.; Fauquet, C.; Broker, T. R.; Bernard, H. U.; Zur Hausen, H. Classification of Papillomaviruses. *Virol*. **2004**, *324*, 17–27.

DOI: 10.36838/v6i10.5

6. Clifford, G. M.; Smith, J. S.; Aguado, T.; Franceschi, S. Comparison of HPV Type Distribution in High-Grade Cervical Lesions and Cervical Cancer: A Meta-Analysis. *Br. J. Cancer*. **2003**, *89*, 101–105.
7. Braaten, P.; Laufer, M. R. Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. *Rev. Obstet. Gynecol*. **2008**, *1*, 2–10.
8. Centers for Disease Control and Prevention. HPV Vaccine Information for Young Women, 2022. https://www.cdc.gov/std/hpv/std-fact-hpv-vaccine-young-women.htm (Accessed 8 January 2024).
9. Pal, A.; Kundu, R. Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy. *Front. Microbiol*. **2020**, *10*, 3116. DOI: 10.3389/fmicb.2019.03116
10. Oyervides-Munoz, M. A.; Pérez-Maya, A. A.; Rodríguez-Gutiérrez, H. F.; Gómez-Macias, G. S.; Fajardo-Ramírez, O. R.; Treviño, V.; *et al*. Understanding the HPV Integration and its Progression to Cervical Cancer. Infect. Genet. Evol. **2018**, *61*, 134–144.
11. National Cancer Institute. https://portal.gdc.cancer.gov/projects/TCGA-CESC (Accessed 9 January 2024)
12. Wilks, C.; Zheng, S. C.; Chen, F. Y.; Charles, R.; Solomon, B.; Ling, J.P.; *et al*. recount3: Summaries and Queries for Large-Scale RNA-seq Expression and Splicing. *Genome Biol*. **2021**, *22*, 323.
13. Ren, S.; Gaykalova, D. A.; Guo, T.; Favorov, A. V.; Fertig, E. J.; Tamayo, P.; *et al*. HPV E2, E4, E5 Drive Alternative Carcinogenic Pathways in HPV Positive Cancers. *Oncogene*. **2020**, *39* (40), 6327–6339.
14. Gupta, S.; Kumar, P.; Das, B. C. HPV: Molecular Pathways and Targets. *Curr. Probl. Cancer*. **2018**, *42* (2), 161–174. DOI: 10.1016/j.currproblcancer.2018.03.003
15. Li, X.; Meng, X.; Chen, H.; Fu, X.; Wang, P.; Chen, X.; *et al*. Integration of Single Sample and Population Analysis for Understanding Immune Evasion Mechanisms of Lung Cancer. NPJ Syst. *Biol. Appl*. **2023**, *9*, 4.
16. Gene Regulatory Networks Database. http://www.grndb.com (Accessed 9 January 2024).
17. Johnson, K. A.; Krishnan, A. Robust Normalization and Transformation Techniques for Constructing Gene Coexpression Networks from RNA-seq Data. *Genome Biol*. **2022**, *23*, 1. DOI: https://doi.org/10.1186/s13059-021-02568-9
18. Zerbino, D. R.; Achuthan, P.; Akanni, W.; Amode, M. R.; Barrell, D.; Bhai, J.; *et al*. Ensembl 2018. *Nucleic Acids Res*. **2018**, *46*, D754–D761.
19. Van Doorslaer, K.; Tan, Q.; Xirasagar, S.; Bandaru, S.; Gopalan, V.; Mohamoud, Y.; *et al*. The Papillomavirus Episteme: A Central Resource for Papillomavirus Sequence Data and Analysis. *Nucleic Acids Res*. **2012**, *41*: D571–D578.
20. Gu, M.; He, T.; Yuan, Y.; Duan, S.; Li, X.; Shen, C. Single-Cell RNA Sequencing Reveals Multiple Pathways and the Tumor Microenvironment Could Lead to Chemotherapy Resistance in Cervical Cancer. *Front. Oncol*. **2021**, *11*, 753386. DOI:10.3389/fonc.2021.753386.
21. Deng, Y.; Song, Z.; Huang, L.; Guo, Z.; Tong, B.; Sun, M.; *et al*. Tumor Purity as a Prognosis and Immunotherapy Relevant Feature in Cervical Cancer. *Aging (Albany, NY)*. **2021**, *13* (22), 24768–24785. DOI:10.18632/aging.203714
22. Tosi, A.; Parisatto, B.; Menegaldo, A.; Spinato, G.; Guido M, Del Mistro A.; *et al*. The Immune Microenvironment of HPV-Positive and HPV-Negative Oropharyngeal Squamous Cell Carcinoma: A Multiparametric Quantitative and Spatial Analysis Unveils a Rationale to Target Treatment-Naïve Tumors with Immune Checkpoint Inhibitors. *J. Exp. Clin. Cancer. Res*. **2022**, *41* (1): 279. DOI:10.1186/s13046-022-02481-4
23. Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; *et al*. Towards a Knowledge-Based Human Protein Atlas. *Nat. Biotechnol*. **2010**, *28*, 1248–1250. https://doi.org/10.1038/nbt1210-1248
24. Tang, S.; Tao, M.; McCoy, J. P. Jr.; Zheng, Z. M. The E7 Oncoprotein is Translated from Spliced E6*I Transcripts in High-Risk Human Papillomavirus Type 16- or Type 18-Positive Cervical Cancer Cell Lines via Translation Reinitiation. *J. Virol.* **2006**, *80* (9), 4249–4263. DOI:10.1128/JVI.80.9.4249-4263.2006
25. Tan, C. L.; Gunaratne, J.; Lai, D.; Carthagena, L.; Wang, Q.; Xue, Y. Z.; *et al*. HPV-18 E2^E4 Chimera: 2 New Spliced Transcripts and Proteins Induced by Keratinocyte Differentiation. *Virology*. **2012**, *429* (1), 47–56. DOI:10.1016/j.virol.2012.03.023
26. Myles, A. J.; Feudale, R. N.; Liu, Y.; Woody, N. A.; Brown, S. D. An Introduction to Decision Tree Modeling. *J. Chemomet*. **2004**, *18* (6), 275–285.
27. Rahaman, M.; Li, C.; Wu, X.; Yao, Y.; Hu, Z.; Jiang, T. A. O. A Survey for Cervical Cytopathology Image Analysis Using Deep Learning. *IEEE Access*. **2020**, *8*, 61687–61710.
28. Hou, N.; Li, M.; He, L.; Xie, B.; Wang, L.; Zhang, R.; *et al*. Predicting 30-Days Mortality for MIMIC-III Patients with Sepsis-3: A Machine Learning Approach Using XGboost. *J. Transl. Med*. **2020**; *18*, 462.
29. Wint, H.; Li, J.; Abe, T.; Yamada, H.; Higaki, T.; Nasu, Y.; *et al*. Pacsin 2-Dependent N-Cadherin Internalization Regulates the Migration Behaviour of Malignant Cancer Cells. *J. Cell Sci.* **2023**, *136* (10), jcs260827. DOI: 10.1242/jcs.260827
30. de Kreuk, B. J.; Anthony, E. C.; Geerts, D.; Hordijk, P. L. The F-BAR Protein PACSIN2 Regulates Epidermal Growth Factor Receptor Internalization. *J. Biol. Chem*. **2012**, *287* (52), 43438–43453. DOI:10.1074/jbc.M112.391078
31. Ma, Y.; Wang, Z.; Sun, J.; Tang, J.; Zhou, J.; Dong, M. Investigating the Diagnostic and Therapeutic Potential of SREBF2-Related Lipid Metabolism Genes in Colon Cancer. *Onco. Targets Ther*. **2023**, *16*, 1027–1042. DOI: 10.2147/OTT.S428150
32. Freed-Pastor, W.A.; Mizuno, H.; Zhao, X.; Langerød, A.; Moon, S. H.; Rodriguez-Barrueco, R.; *et al*. Mutant p53 Disrupts Mammary Tissue Architecture via the Mevalonate Pathway. *Cell*. **2012**, *148* (2012), 244–258.
33. Wang, B.; Li, X.; Liu, L. β-Catenin: Oncogenic Role and Therapeutic Target in Cervical Cancer. *Biol. Res.* **2020**, *53*, 33. https://doi.org/10.1186/s40659-020-00301-7
34. Klaus, A.; Birchmeier, W. Wnt Signalling and its Impact on Development and Cancer. *Nat. Rev. Cancer*. **2008**, *8* (5), 387–398. DOI:10.1038/nrc2389
35. Kim, N. H.; Kim, H. S.; Kim, N. G.; Lee, I.; Choi, H. S.; Li, X. Y.; *et al*. p53 and MicroRNA-34 are Suppressors of Canonical Wnt Signaling. *Sci. Signal*. **2011**, *4* (197), ra71.
36. Shi, W.; Chen, Z.; Li, L.; Liu, H.; Zhang, R.; Cheng, Q.; *et al*. Unravel the Molecular Mechanism of XBP1 in Regulating the Biology of Cancer Cells. *J. Cancer*. **2019**, *10* (9), 2035–2046. DOI:10.7150/jca.29421
37. Namba, T.; Chu, K.; Kodama, R.; Byun, S.; Yoon, K. W.; Hiraki, M.; *et al*. Loss of p53 Enhances the Function of the Endoplasmic Reticulum Through Activation of the IRE1α/XBP1 Pathway. *Oncotarget*. **2015**, *6* (24), 19990–20001. DOI:10.18632/oncotarget.4598
38. Walhout, A. J. What Does Biologically Meaningful Mean? A Perspective on Gene Regulatory Network Validation. *Genome Biol*. **2011**, *12* (4), 109. DOI:10.1186/gb-2011-12-4-109
39. Kelley, D. Z.; Flam, E. L.; Izumchenko, E.; Danilova, L. V.; Wulf, H. A.; Guo, T.; *et al*. Integrated Analysis of Whole-Genome ChIP-Seq and RNA-Seq Data of Primary Head and Neck Tumor

Samples Associates HPV Integration Sites with Open Chromatin Marks. *Cancer Res*. **2017**, *77* (23), 6538–6550. DOI:10.1158/0008-5472.CAN-17-0833

40. Jones, R. N.; Miyauchi, S.; Roy, S.; Boutros, N.; Mayadev, J. S.; Mell, L. K.; *et al*. Computational and AI-Driven 3D Structural Analysis of Human Papillomavirus (HPV) Oncoproteins E5, E6, and E7 Reveal Significant Divergence of HPV E5 Between Low-Risk and High-Risk Genotypes. *Virology*. **2024**, *590*, 109946. DOI:10.1016/j.virol.2023.109946

41. Tungteakkhun, S. S.; Duerksen-Hughes, P. J. Cellular Binding Partners of the Human Papillomavirus E6 Protein. *Arch. Virol*. **2008**, *153* (3), 397–408. DOI:10.1007/s00705-007-0022-5

Author

Allison Moores is a senior at The Bishop's School. She is focused on virology, immunology, and bioinformatics and is dedicated to international youth cancer prevention, research, and advocacy efforts.

DOI: 10.36838/v6i10.5