Detection by Single-Cell RNA Sequencing of Virally Mediated Skin Diseases

Linda Zhou[1] and Thomas H. Leung[1,2]

JID Innovations (2025);**5:**100348 doi:10.1016/j.xjidi.2025.100348

Viruses are well-documented agents of specific skin diseases. However, their role and precise mechanism of action in other skin diseases remain unknown. We describe a single-cell RNA-sequencing–based strategy to interrogate human skin biopsies for viral transcripts, permitting detection of viral infection within a sample, single-cell resolution of virally infected cells and identification of subsequent transcriptomic perturbations. We validate our pipeline with 100% sensitivity and specificity by (i) detecting Merkel cell polyomavirus in Merkel cell carcinoma samples, (ii) detecting specific human papillomavirus strains in known human papillomavirus–positive tumors, and (iii) detecting rubella virus transcripts in patients with known rubella-associated granulomas. We identify infection of known and previously unreported cell types and elucidate viral-mediated transcriptional perturbations. In rubella virus–infected cells, we discover macrophage-specific evolution of the rubella virus E1 capsid protein. Finally, we interrogate skin biopsies from many established nonvirally mediated inflammatory skin diseases and do not find consistent evidence of viral infection in any condition. Combining single-cell RNA-sequencing data with virome detection strategies represents a potentially powerful approach to investigate and elucidate virus-mediated gene regulation in health and disease.

Keywords: Human papillomavirus, Merkel cell carcinoma, Rubella granuloma, scRNA-seq, Virome

INTRODUCTION

Viruses established to cause human skin disease include herpesviruses, poxviruses, human papillomaviruses (HPVs), and Merkel cell polyomavirus (MCPyV). However, viral triggers are often postulated for many other skin diseases, including lichen planus, granuloma annulare, and sarcoidosis, but evidence supporting these associations has been lacking.

Next-generation sequencing–based methods to detect viral transcripts have increased our understanding of the skin virome. Viral genetic material can be detected from DNA or RNA-sequencing datasets. Methods to detect viral infection in single-cell RNA-sequencing (scRNA-seq) datasets are particularly of interest because they would also permit cell-specific resolution of infection. Previous studies have used scRNA-seq to interrogate viral infection in cell culture and mouse models, including COVID-19 and influenza. Many of these studies are designed to detect specific viruses in scRNA-seq data rather than as a comprehensive virome screen. Comprehensive virome detection using scRNA-seq has been demonstrated on primate peripheral blood samples and several human tissue samples; however, this technique has not yet been published using human skin. In this study, we assemble a pipeline to detect viral transcripts in scRNA-seq datasets of human skin biopsies and demonstrate that 4-mm skin biopsies are sufficient to detect viral transcripts in neoplastic, infectious, and granulomatous skin lesions. We further show how viral detection at single-cell resolution of can lead to mechanistic insights seen in these disorders.

RESULTS

To validate and demonstrate our method, we apply the viral detection pipeline to previously published scRNA-seq datasets of Merkel cell carcinoma (MCC), HPV-associated cervical cancer, as well as our own data generated from patients with granulomatous skin diseases. In the next paragraphs, we discuss our findings for each case study.

Transcriptomic analysis of MCPyV-positive MCC reveals coinfection with Epstein–Barr virus and distinct E2F4-associated viral transcription patterns

Raw FASTQ data from Das et al (2023) were downloaded from Gene Expression Omnibus (GEO) and processed as described in the Materials and Methods. The dataset included scRNA-seq from 11 samples of MCC tumors biopsied from the skin, lymph nodes, and parotid glands, including 5 that are known MCPyV negative and 6 that are MCPyV positive by PCR (Das et al, 2023). Using our pipeline, we detected hundreds to thousands of MCPyV transcripts in the 6 MCPyV-positive samples and no transcripts in the MCPyV-negative samples. Unexpectedly, we also detected low numbers of reads that align to the genome of human herpes virus 4

[1]Department of Dermatology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; and [2]Corporal Michael Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania, USA

Correspondence: Thomas H. Leung, Department of Dermatology, University of Pennsylvania School of Medicine, 421 Curie Boulevard, 1006 BRB, Philadelphia, Pennsylvania 19104, USA. E-mail: thl@pennmedicine.upenn.edu

Abbreviations: GEO, Gene Expression Omnibus; HPV, human papillomavirus; MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus; scRNA-seq, single-cell RNA sequencing

Received 1 August 2024; revised 4 December 2024; accepted 23 December 2024; accepted manuscript published online XXX; corrected proof published online XXX

Cite this article as: JID Innovations 2025;5:100348

(Epstein–Barr virus) in 4 of the 6 MCPyV-positive samples (Figure 1a). Each of these reads aligned completely end to end and are on average >99% identical to the corresponding Epstein–Barr sequence. No other human viruses in our queried database (Supplementary Table S2) were detected in these samples.

We next focused on the 3 skin tumor samples. We found that MCPyV transcripts originated almost exclusively from tumor cells, with very few reads found in immune cells (Figure 1b). In comparison, when present, Epstein–Barr virus transcripts originated exclusively from immune cells (data not shown). The 3 skin biopsy samples demonstrate significant heterogeneity with respect to cell-type composition and MCPyV positivity (Figure 1c).

We hypothesized that tumor cells containing MCPyV may harbor distinct transcriptional perturbations compared with MCPyV-negative tumor cells. We performed differential gene expression between MCPyV-positive and MCPyV-negative tumor cells separately in samples P7 and P9. We then took the intersection of the top 10% of differentially expressed genes by fold change in each dataset to find common transcriptional perturbations (Figure 1d). We used Enrichr (Xie et al, 2021) to perform gene set enrichment analyses on this set of 19 genes. Notably, we found that 14 of 19 genes were targets of the E2F4 transcription factor, and they shared gene ontogeny pathways related to DNA replication (Figure 1e). These results suggested that MCPyV infection may facilitate cell proliferation through mechanisms that involve the E2F4 transcription factor.

Differential gene expression in HPV16/18-positive cervical cancer reveals consistent HS3ST4 upregulation linked to increased viral transcripts

As a second example, raw FASTQ data from Guo et al (2023) were downloaded from GEO and processed as described in the Materials and Methods. The dataset included 2 HPV-negative cervical biopsy samples, 2 HPV-positive high-grade squamous intraepithelial lesion cervical biopsy samples, and 3 HPV-positive cervical cancer samples, where all HPV-positive samples were positive for either HPV16 or HPV18. Our pipeline detected no HPV transcripts in the HPV-negative normal patients and hundreds to thousands of transcripts corresponding to HPV16 and HPV18 in HPV-positive high-grade squamous intraepithelial lesions and cervical cancer samples, in agreement with Guo et al (2023) (Figure 2a). No other human viruses in our queried database (Supplementary Table S2) were detected. As expected, HPV-infected cells were largely epithelial cells, although up to 20% of HPV+ cells were immune cells in the cervical cancer samples (Figures 2b and 3a and b).

We next performed differential gene expression analysis in the 3 cervical cancer samples comparing epithelial cells with and without HPV16/18 transcripts. We then intersected differentially expressed genes from each sample and surprisingly found just 1 gene, *HS3ST4*, that is consistently expressed in all HPV16/18-positive samples (Figures 2c and d and 3b). Approximately, 50% of all virally infected epithelial cells expressed *HS3ST4* across all cervical cancer samples, and HPV16/18-negative epithelial cells exhibited decreased expression of *HS3ST4*, between 2.5 and 100-fold less depending on the sample. In contrast, epithelial cells in high-grade squamous intraepithelial lesions do not express *HS3ST4* regardless of HPV16/18 presence (Figure 2c). Taken together, this pipeline suggested a previously unreported role for *HS3ST4* in cervical cancer progression.

Transcriptomic profiling of rubella virus in granulomatous lesions reveals strain-specific cellular tropism and distinct host transcriptional responses

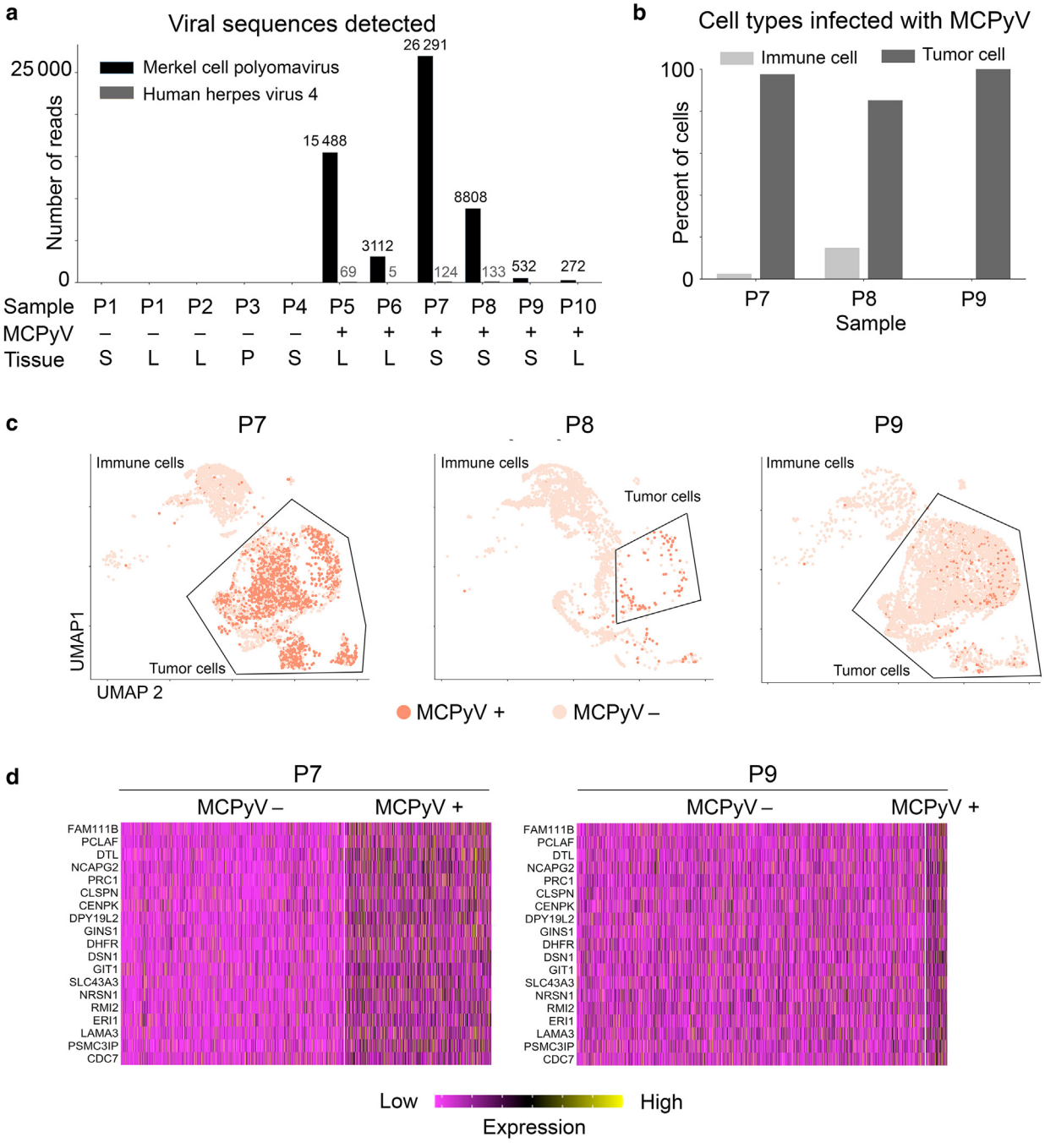
We applied our viral transcript detection pipeline to scRNA-seq datasets generated from 4-mm skin punch biopsies in 60 different patients with inflammatory skin diseases, including but not limited to nonsarcoid granulomatous disease, sarcoidosis, atopic dermatitis, psoriasis, bullous pemphigoid, and lichen planus (Liu et al, 2022; Sati et al, 2024).

We initially focused on 2 patients diagnosed with rubella virus–associated granulomas (patients S1 and S2), a granulomatous skin disease known to be virally mediated. As a negative control, we interrogated 14 patients diagnosed with granulomatous skin diseases, including granuloma annulare or xanthogranulomas (patients S3–S16). The pipeline identified rubella virus transcripts in the lesional skin samples of S1 and S2 but did not detect any virus-related transcripts in S3–S16, although the nonlesional skin sample from S13 did have several reads aligning to human erythrovirus V9 (Figure 4a). These results demonstrate the high sensitivity and specificity of our method when applied to skin punch biopsies.

We next investigated the cell types that contained the rubella virus transcripts. The majority of rubella virus transcripts primarily originated from macrophages in S1, whereas in S2, rubella reads originated from macrophages and T cells equally (Figures 4b and 5a).

We wondered whether this cell-type difference could be related to differences in the rubella virus strain affecting S1 and S2. S1 was taken from a man aged 70 years with disseminated rubella lesions who has been previously described in the literature (Shields et al, 2021). Notably, in this previous work, full rubella genome sequencing revealed that this patient's strain was likely a wild-type strain and not vaccine derived. In comparison, S2 was taken from a female in her 50s who developed a single granulomatous lesion in the location of a measles, mumps, and rubella booster several years prior. A skin biopsy from this patient was sent to the Centers for Disease Control and Prevention and was found to be positive for a vaccine strain rubella virus through immunohistochemistry.

To investigate whether our pipeline could detect these differences, we extracted the reads aligning to rubella virus genome and aligned them to different rubella virus strains. In both patients S1 and S2, the detected rubella transcripts aligned to the E1 glycoprotein of the rubella virus. The sequences from S2 were an exact match to a known rubella vaccine strain (RA 27/3 strain), and the sequence from S1 predominantly aligned with the sequence of the Philadelphia strain of rubella virus, except for a single sequence variant. Notably, utilizing the single-cell resolution afforded by our method, we determined that this sequence variant was present exclusively in the macrophages from the affected tissue and not in other cell types (Figure 4c). The sequence variant



e

Ontology term	Source	Adjusted p-val	OR
E2F4 ENCODE	ChIP-X	7.917e-12	47.36
Cell cycle	Reactome	3.039e-7	26.98
Mitotic Sister Chromatid Segregation	GO Biological Process	2.946e-4	49.53
Mitotic nuclear division	GO Biological Process	4.269e-4	89.01

Figure 1. Detection of Merkel cell polyomavirus in tumor cells of skin biopsies. (**a**) Barplot showing results from viral RNA detection pipeline run on scRNA-seq datasets of MCC samples from Das et al (2023). The number of reads aligning to MCPyV and human herpesvirus 4 are depicted. Tissues include skin (denoted as S), lymph node (denoted as L), and parotid glands (denoted as P). **(b)** Barplot showing the cell types found to contain reads aligning to MCPyV in skin MCC samples. **(c)** UMAP representation scRNA-seq of skin MCC samples, highlighted by MCPyV status. Cell types were identified using SingleR (Aran et al, 2019). **(d)** Heatmap showing differentially expressed genes in MCC tumor cells, segregated by the presence of MCPyV transcripts. **(e)** Results of gene set enrichment analysis using Enrichr (Xie et al, 2021) on 19 genes shown in **d**. MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus; scRNA-seq, single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and Projection.

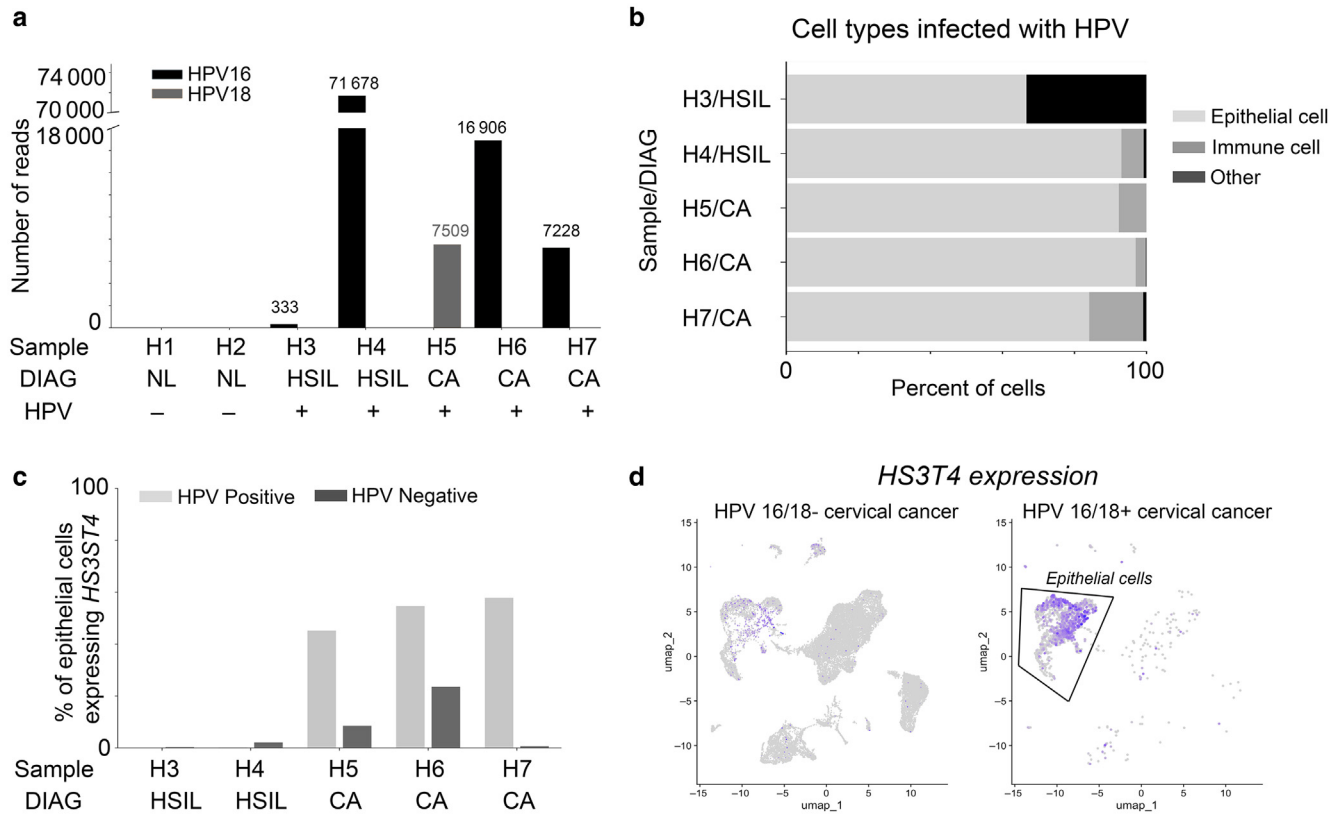


Figure 2. Detection of HPV in primarily epithelial cells of cervical tissue biopsies. (**a**) Barplot showing results from viral RNA detection pipeline run on scRNA-seq datasets of cervical tissue from Guo et al (2023). Numbers of reads aligning to HPV16 or HPV18 are depicted. Tissues include normal cervix (denoted as NL), HSILs, and CA. **(b)** Barplot showing the cell types found to contain reads aligning to HPV16/18 in HSIL and CA samples. **(c)** Barplot showing the percentage of epithelial cells expressing *HS3ST4* transcript in HSIL and CA samples, depending on HPV presence. **(d)** UMAP plot comparing the expression of *HS3T4* in HPV16/18-positive versus HPV16/18-negative cells in CA biopsies (H5, H6, H7). CA, cervical cancer; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; scRNA-seq, single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and Projection.

resulted in a threonine to isoleucine change, a polar to hydrophobic change. Given that the E1 glycoprotein plays a crucial role in viral attachment to host cells and entry, variants in this protein could have implications for the virus's infectivity and immune evasion, including allowing macrophage-specific presence. This specific sequence variant landscape provides valuable insights into the adaptive evolution of the rubella virus within different cellular environments.

We next performed differential gene expression analysis on macrophages infected with rubella compared with that on macrophages not infected with rubella. In patient S1 who was infected with the wild-type rubella strain, we identified a set of 19 genes that were expressed by 100% of macrophages at levels up to 11-fold higher compared with that by macrophages not infected with rubella. These genes included those related to IL-2/signal transducer and activator of transcription 5 pathway (*SPP1*, *SPRED2*), mTOR signaling (*HSPA9*, *EGLN3*), and chromatin remodeling (*KDM6A, H2AFY, DEK)* (Figure 4d). In contrast, in patient S2 who was infected with the rubella vaccine strain, infected macrophages did not exhibit these transcriptional perturbations. Instead, these macrophages expressed genes related to cellular trafficking and lipid metabolism (Figure 4d).

Finally, we applied our pipeline to 44 additional skin biopsy–derived scRNA-seq datasets from patients with lichen planus, sarcoidosis, atopic dermatitis, psoriasis, and bullous pemphigoid (Liu et al, 2022). We did not find evidence of infection with any human virus within our database (Supplementary Table S2) in any of the samples mentioned earlier (Figure 6).

DISCUSSION

Overall, we demonstrate that comprehensive virome detection may be performed using scRNA-seq data from 4-mm human skin biopsies in a highly specific manner and with cell-specific resolution. By applying this method to datasets from MCC, HPV-associated cervical cancer, and rubella-associated granulomas, we demonstrate its robustness and specificity in detecting viral transcripts and potential for uncovering previously unreported biological insights.

In MCPyV-positive MCC, we identified specific transcriptional patterns linked to the E2F4 transcription factor. MCPyV is known to directly bind to the retinoblastoma gene (*RB1*) and inhibits its tumor suppressor functions (Hesbacher et al, 2016). The retinoblastoma protein normally forms complexes with the E2F transcription factor family members to downregulate target genes involved in DNA replication and

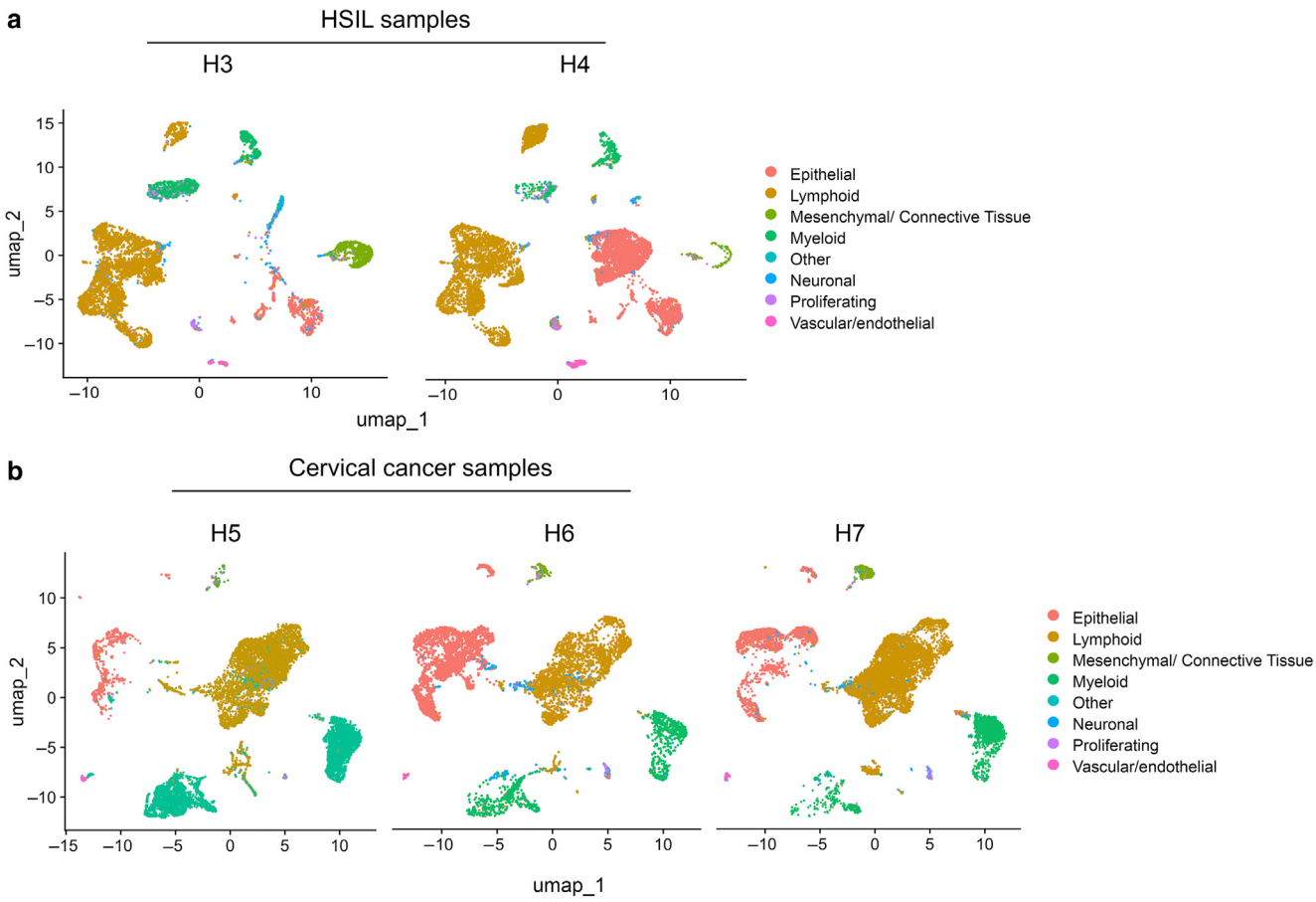


Figure 3. UMAPs showing cell-type clustering of HSIL and cervical cancer samples. (**a**) UMAP showing cell-type clustering of scRNA-seq performed on cervical biopsies diagnosed as HSIL or **(b)** cervical cancer. Clustering was performed using Seurat 'FindClusters' function, and clusters were annotated using SingleR (details are presented in Materials and Methods). HSIL, high-grade squamous intraepithelial lesion; scRNA-seq, single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and Projection.

cell cycle control. Although retinoblastoma preferentially binds E2F1, E2F2, and E2F3, it is also capable to bind E2F4 (Engeland, 2022). Prior work has shown that E2F4 has a noncanonical role in regulating cell proliferation separate from retinoblastoma (Hsu et al, 2019). Finally, we find evidence of Epstein–Barr virus transcripts in immune cells within MCC tumors (4 of 6 samples). This leads to an intriguing hypothesis that immune dysregulation secondary to Epstein–Barr virus infection may regulate the progression from MCPyV infection to MCC tumor formation. Future work validating this finding is necessary. We noted that one of the MCC tumor samples (S9) contained few virus-positive cells. MCPyV is known to be clonally integrated in MCC tumors, with integration preceding the expansion of tumor cells (Feng et al, 2008). Thus, given the limitations of scRNA-seq previously discussed, it is most likely that the virus-negative cells in S9 are false negatives rather than representing a situation where there is a subpopulation of tumor cells that are MCPyV positive and a population that are negative. However, other possibilities may exist, including lack of clonal integration or loss of even viral integration (Feng et al, 2008), and additional work is needed to explore these mechanisms.

Taken together, virome detection in human skin biopsies can recapitulate known viral carcinogenesis mechanisms and also identify new mechanisms, thereby enriching our understanding of the disease pathophysiology.

With regard to cervical cancer samples, we observed a consistent upregulation of *HS3ST4* in HPV16/18-positive epithelial cancer cells and an absence in HPV-negative epithelial cells and epithelial cells in HPV-positive high-grade squamous intraepithelial lesion samples. HS3ST4 is an enzyme that generates 3-O-sulfated glucosaminyl residues on heparan sulphate, a cell surface and extracellular matrix protein. These modified residues have been shown to increase infectivity of herpes simplex virus 1 in neurons and enhance fusogenic activity of varicella-zoster virus 1 but have not previously been linked to HPV infection (Ohka et al, 2021; Tiwari et al, 2005). This previously unreported association with HPV infection suggests that *HS3ST4* may play a crucial role in facilitating HPV entry and replication, potentially offering a target for therapeutic intervention. Of note, our viral detection workflow detected HPV transcripts largely in epithelial cells, as expected. However, a small number of transcripts were detected in lymphoid immune cells. HPV is not known to infected lymphoid cells. Thus, these results may be due to technical issues inherent in the scRNA-seq technique, for example, smaller lytic cells may be inappropriately captured with nearby immune cells in a single droplet

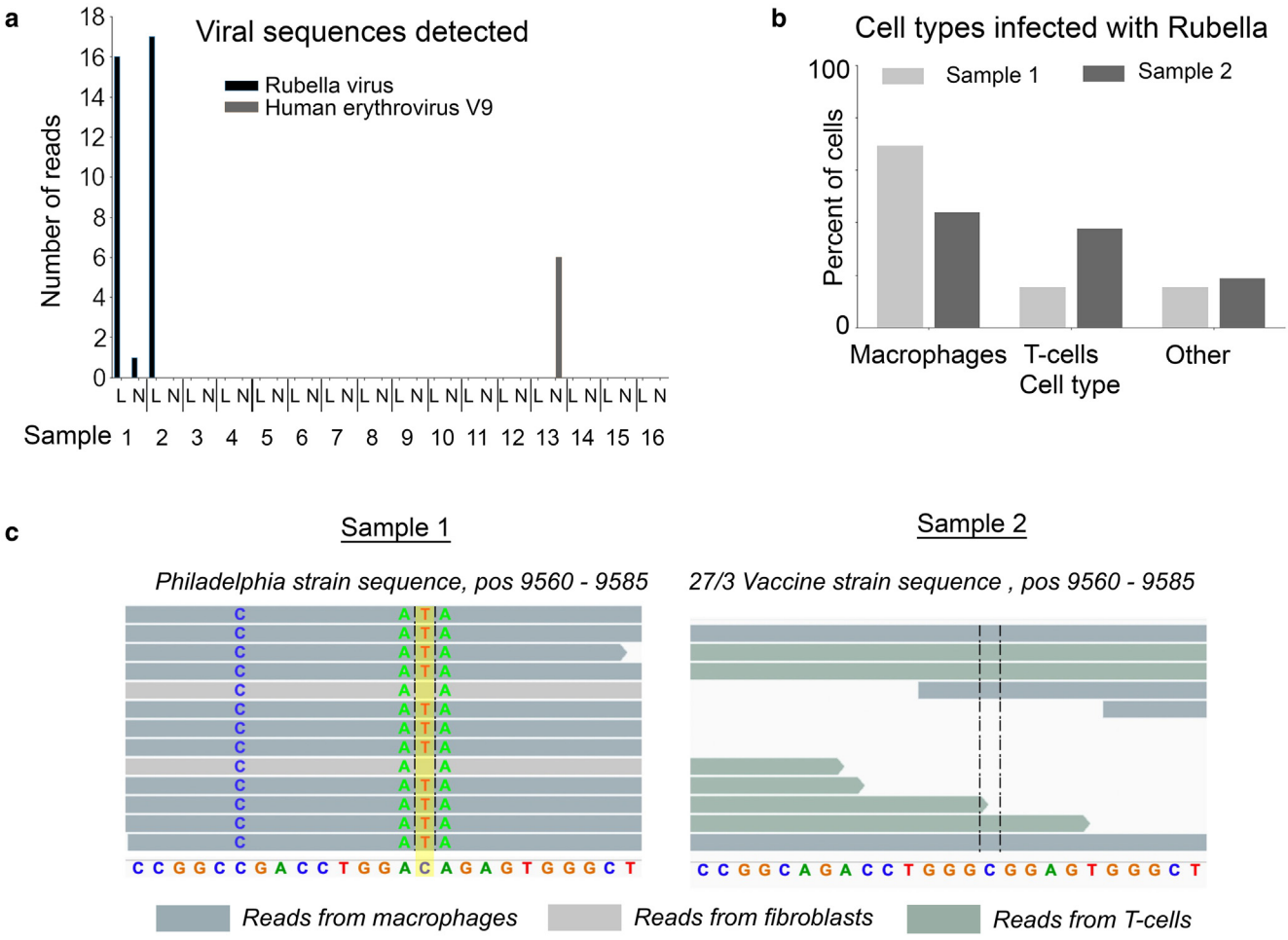


Figure 4. Detection of rubella virus in immune cells of skin biopsies. (**a**) Bar plot showing results from viral RNA detection pipeline run on lesional (denoted as L) and nonlesional (denoted as NL) skin punch biopsy samples from rubella-associated granulomas (sample 1, sample 2) and 14 patients with granuloma annulare (Sati et al, 2024). **(b)** Cell types where rubella transcripts were detected in sample 1 (disseminated lesions, nonvaccine strain) compared with those in sample 2 (single lesion, vaccine strain). "Other" includes NK cells, dendritic cells, and fibroblasts. **(c)** scRNA-seq reads aligning to the rubella E1 capsid sequence in sample 1 and sample 2 highlighted by cell type of origin. Nucleotide variants that distinguish the Philadelphia strain are bolded. Highlighted in yellow in sample 1 is a C -> U sequence variant isolated to macrophages. This variant leads to a threonine -> isoleucine change, a polar to hydrophobic change. **(d)** Heatmap of differentially expressed genes in macrophages ± rubella infection in sample 1 versus sample 2. scRNA-seq, single-cell RNA sequencing.

without triggering doublet detection. Further work would be necessary to establish whether HPV transcripts can truly be detected within immune cells.

The recent discovery of rubella virus–associated granulomatous skin lesions suggests that viruses can remain asymptomatic for years before resurfacing owing to yet

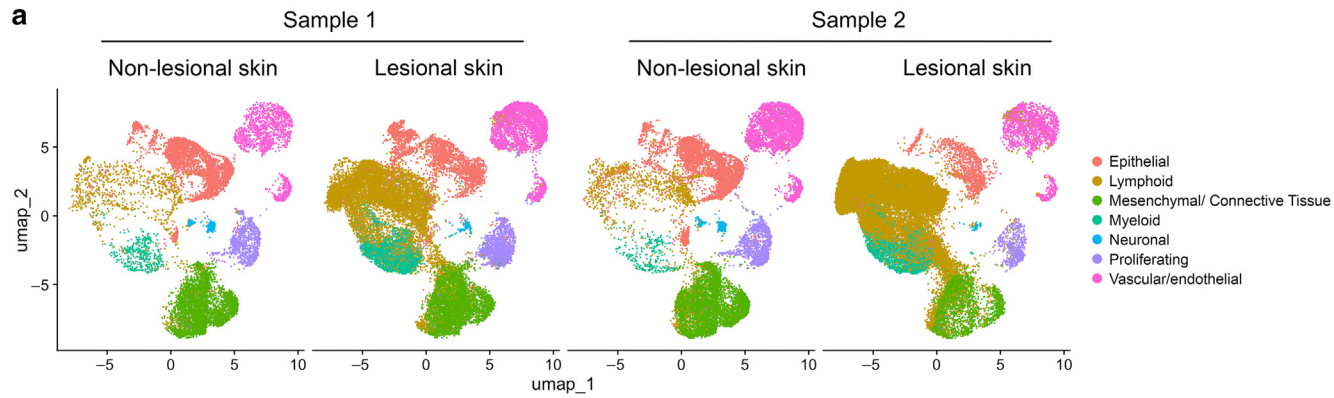


Figure 5. UMAPs showing cell-type clustering of rubella granuloma samples. (**a**) UMAP showing cell-type clustering of scRNA-seq performed on normal skin (nonlesional) and rubella-associated cutaneous granulomas (lesional skin) in 2 different individuals (sample 1, sample 2). Clustering was performed using Seurat 'FindClusters' function, and clusters were annotated using SingleR (Materials and Methods). scRNA-seq, single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and Projection.

unknown triggers and may contribute to more complex skin conditions than previously recognized. Because rubella virus–associated granulomatous infections are increasingly being reported, further work to understand the mechanisms of rubella virus persistence and granuloma formation are needed to develop targeted treatments. Our analysis revealed strain- and sequence variant–specific cellular tropism. Our identification of a previously unreported sequence variant in the E1 glycoprotein of the Philadelphia strain that was exclusive to macrophages underscores the virus's adaptive evolution within different cellular environments. Furthermore, differential gene expression analysis revealed distinct transcriptional perturbations in macrophages infected with the Philadelphia strain, involving pathways such as IL-2/signal transducer and activator of transcription 5 and mTOR signaling, which were absent in vaccine strain infections. We hypothesize that this sequence variant, resulting in a threonine to isoleucine substitution, may play a role in the virus's infectivity and immune evasion capabilities. These findings advance our understanding of rubella virus pathogenesis and illustrate the power of our detection pipeline in identifying critical viral–host interactions.

In contrast to the samples of Merkel cell cancer and cervical cancer, which contained high numbers of viral transcripts, we found lower numbers of rubella-associated transcripts (<20 reads). This discrepancy could be partially due different viral etiologies or lower sequencing depth of the rubella-associated granuloma samples (Supplementary Table S1) or may reflect that skin biopsies, in comparison with cancer tissue samples where presumably more tissue was obtained, give a lower yield. That said, we also did not find any viral transcripts in scRNA-seq datasets from skin biopsies from other diagnoses such as atopic dermatitis, psoriasis, bullous pemphigoid, lichen planus, or sarcoidosis. These conditions are not associated with viral infections and serve as a negative control.

Limitations of the study

We recognize that single-cell methods have known issues of dropout, particularly affecting lowly expressed transcripts

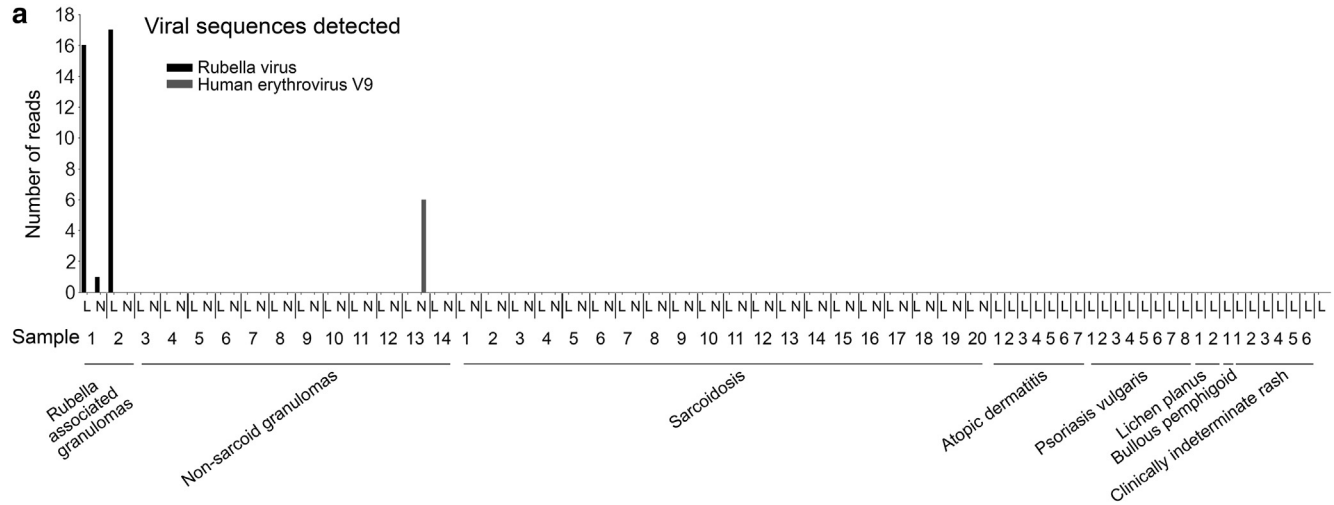


Figure 6. Virome detection in inflammatory skin conditions. **(a)** Bar plot showing results from viral RNA detection pipeline run on lesional (denoted as L) and nonlesional (denoted as NL) skin punch biopsy samples from rubella-associated granulomas, granuloma annulare, sarcoidosis (11), atopic dermatitis, psoriasis vulgaris, lichen planus, bullous pemphigoid, and clinically indeterminate rashes (Liu et al, 2022).

during library generation and sequencing. Thus, viral transcripts that are expressed at very low levels may not be detected in scRNA-seq data. In addition, rare cell populations may not be captured. For these reasons, it is possible that negative virally infected cells of a particular cell type are false negatives. This limitation may be addressed by purifying cell types known to be targets of viral infection, taking larger skin biopsies, and/or sequencing more cells. Overall, skin biopsies have yielded detection of rare immune cell populations in different diseases (Sati et al, 2024). That said, this pipeline represents a convenient method for investigators to screen for viral transcripts in skin biopsy samples to evaluate the possibility of viral infection, instead of definitive proof of absence of virus given these caveats. scRNA-seq analysis remains too expensive for routine clinical diagnosis application. A similar approach has been taken for bulk RNA-sequencing datasets, but this approach will not provide cell-type–specific information (Selitsky et al, 2020).

In conclusion, our study is a proof of concept that demonstrates that viral transcript detection is feasible in human skin biopsies and can provide high-resolution insights into viral-mediated pathology. The ability to detect and analyze viral transcripts with such specificity and sensitivity across diverse clinical samples underscores the potential of this method in diagnosing viral-associated skin disease and advancing the understanding of the complex interactions between viruses and their hosts.

MATERIALS AND METHODS

Human scRNA-seq datasets

Previously published data in this study can be found at GEO accession GSE226438 (MCC) (Das et al, 2023), GEO accession GSE208653 (HPV-associated cervical cancer) (Guo et al, 2023), GEO accession GSE226896 (granulomatous skin disease) (Sati et al, 2024), and European Genome Archive accession EGAS00001005271 (other inflammatory skin conditions) (Liu et al, 2022).

Human tissue processing

The 4-mm skin punch biopsies obtained from lesional and nonlesional skin were first incubated in 200 μl of Dispase solution (2 mg/ml, D4818, Sigma-Aldrich) for 30 minutes at 37 °C. The dermis was then separated from the epidermis, washed in PBS, and minced into smaller pieces. Dispase digestion was terminated with 100 μl fetal bovine serum and 3 μl of 0.5 M EDTA. Cells were then filtered through a 70-mm cell strainer (22-363-548, Fisher Scientific), pelleted, and washed twice with PBS containing 1% BSA before being counted and prepared for sequencing. Samples were sequenced using 10X Chromium 3 kit, version 3.1 (1000268, 10X Genomics). The sequencing libraries were prepared per the manufacturer's instructions and sequenced with 2 × 100 bp paired-end sequencing on the Illumina HiSeq2000/HiSeq2500 platforms using commercial services from BGI America.

Single-cell RNA data processing

scRNA-seq data were aligned to the GRCh38 reference genome to produce gene count and cell barcode matrices utilizing the "cellranger count" tool from the CellRanger pipeline (version 7.0.0, 10X Genomics). All subsequent analysis steps were conducted using the Seurat R package (version 5.0.3, https://github.com/satijalab/seurat) unless stated otherwise. Briefly, we used the Seurat functions 'Read10X' and 'CreateSeuratObject' to import and merge Seurat objects from all filtered feature barcode matrices generated by the Cell Ranger pipeline. Cells were filtered out if they had <250 genes, <500 unique molecular identifiers, a log10 gene count per unique molecular identifier below 0.80, or >20% mitochondrial reads (Supplementary Table S1). In addition, genes present in <10 cells were removed. To account for cell cycle effects, each cell was assigned a phase score using Seurat's 'CellCycleScoring' function. The data were then log normalized and scaled by performing linear regression against the number of reads. We identified variable genes using the 'FindVariableFeatures' function followed by the 'SelectIntegrationFeatures' function with nfeatures set to 3000. For batch correction and cross-tissue data integration, the 'FindIntegrationAnchors' and 'IntegrateData' functions were applied to the merged Seurat object. Dimensionality reduction was carried out using the 'RunPCA' and 'RunUMAP' functions to produce Uniform Manifold Approximation and Projection plots. Louvain clustering was then performed using the 'FindClusters' function with the first 40 principal components and at a resolution of 1.4. To determine the optimal number of dimensions and resolution, we utilized the ElbowPlot function in Seurat and visually inspected DimHeatmap plots at various dimensions. Differential gene expression analysis was performed with "FindConservedMarkers." Annotation of cell types was completed using SingleR (version 2.4.1 [Aran et al, 2019]) with the Human Primary Cell Atlas as the reference database (Mabbott et al, 2013).

Viral detection pipeline

After alignment of scRNA-seq data to the GRCh38 reference genome using "cellranger count" as described earlier, all unmapped reads were extracted and filtered for those that originated from cells that pass the filtering steps, as described earlier. This was to ensure that viral reads selected for further analysis originated from within a specific cell. These reads were then aligned to a previously published FASTA file containing vertebrate viral genomes downloaded from GenBank (Selitsky et al, 2020). The viral genomes are masked such that low complexity sequences and sequences that could map to the human genome are hidden to increase specificity (Selitsky et al, 2022). The full list of all viral genomes queried is reported in Supplementary Table S2. Alignment was completed using the National Center for Biotechnology Information blastn tool (Camacho et al, 2009). Custom scripts were written to quantify the number of unique reads originating from each viral genome. Virus was considered "detected" if at least 5 reads aligning to the genome were detected, where each read had to map with >95% specificity with no read trimming. All scripts used in the viral detection pipeline are available on Github at https://github.com/skinviromedetection/scRNA-seq-virome.

ETHICS STATEMENT

The Institutional Review Board of the University of Pennsylvania School of Medicine granted approval for the study's protocol. Informed consent in writing was obtained from all study participants prior to their inclusion. The informed consent documents clearly stated that the study was exploratory in nature and that participants should not anticipate any specific medical interventions, treatments, or changes in their healthcare management. Participants provided samples of skin for the study.

DATA AVAILABILITY STATEMENT

This paper reports original code. Previously published data that were reanalyzed in this study are available on Gene Expression Omnibus (GSE226438, GSE208653, GSE226896) and European Genome Archive (EGAS00001005271). All code used in the viral detection pipeline is available on GitHub at https://github.com/skinviromedetection/scRNA-seq-virome. Any

additional information required to reanalyze the data reported in this paper is available from the corresponding author upon request.

ORCIDs
Thomas H. Leung: http://orcid.org/0000-0003-1086-0824
Linda Zhou: http://orcid.org/0000-0002-8218-2973

CONFLICT OF INTEREST
The authors state no conflict of interest.

ACKNOWLEDGMENTS
We thank Misha Rosenbach and Satish Sati for their helpful discussions and insights. This project received support from the National Institutes of Health (R01-AR079483), Department of Veteran's Affairs (I01-RX002701), Berstein Foundation, and H.T. Leung Foundation.

AUTHOR CONTRIBUTIONS
Conceptualization: LZ, THL; Investigation: LZ, THL; Methodology: LZ, THL; Writing − Original Draft Preparation: LZ; Writing − Review and Editing: THL

DECLARATION OF GENERATIVE ARTIFICIAL INTELLIGENCE (AI) OR LARGE LANGUAGE MODELS (LLMS)
The authors did not use AI/LLMS in any part of the research process and/or manuscript preparation

SUPPLEMENTARY MATERIAL
Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.xjidi.2025.100348.

REFERENCES
Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20:163−72.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. Blast+: architecture and applications. BMC Bioinformatics 2009;10:421.

Das BK, Kannan A, Velasco GJ, Kunika MD, Lambrecht N, Nguyen Q, et al. Single-cell dissection of Merkel cell carcinoma heterogeneity unveils transcriptomic plasticity and therapeutic vulnerabilities. Cell Rep Med 2023;4:101101.

Engeland K. Cell cycle regulation: p53-p21-RB signaling. Cell Death Differ 2022;29:946−60.

Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008;319:1096−100.

Guo C, Qu X, Tang X, Song Y, Wang J, Hua K, et al. Spatiotemporally deciphering the mysterious mechanism of persistent HPV-induced malignant transition and immune remodelling from HPV-infected normal cervix, precancer to cervical cancer: integrating single-cell RNA-sequencing and spatial transcriptome. Clin Transl Med 2023;13:e1219.

Hesbacher S, Pfitzer L, Wiedorfer K, Angermeyer S, Borst A, Haferkamp S, et al. RB1 is the crucial target of the Merkel cell polyomavirus Large T antigen in Merkel cell carcinoma cells. Oncotarget 2016;7:32956−68.

Hsu J, Arand J, Chaikovsky A, Mooney NA, Demeter J, Brison CM, et al. E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family. Nat Commun 2019;10:2939.

Liu Y, Wang H, Taylor M, Cook C, Martínez-Berdeja A, North JP, et al. Classification of human chronic inflammatory skin disease based on single-cell immune profiling. Sci Immunol 2022;7:eabl9165.

Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics 2013;14:632.

Ohka S, Yamada S, Nishizawa D, Fukui Y, Arita H, Hanaoka K, et al. Heparan sulfate 3-O-sulfotransferase 4 is genetically associated with herpes zoster and enhances varicella-zoster virus-mediated fusogenic activity. Mol Pain 2021;17:17448069211052171.

Sati S, Huang J, Kersh AE, Jones P, Ahart O, Murphy C, et al. Recruitment of CXCR4+ type 1 innate lymphoid cells distinguishes sarcoidosis from other skin granulomatous diseases. J Clin Invest 2024;134:e178711.

Selitsky SR, Marron D, Hollern D, Mose LE, Hoadley KA, Jones C, et al. Virus expression detection reveals RNA-sequencing contamination in TCGA. BMC Genomics 2020;21:79.

Shields BE, Perelygina L, Samimi S, Haun P, Leung T, Abernathy E, et al. Granulomatous dermatitis associated with rubella virus infection in an adult with immunodeficiency. JAMA Dermatol 2021;157:842−7.

Tiwari V, O'Donnell CD, Oh MJ, Valyi-Nagy T, Shukla D. A role for 3-O-sulfotransferase isoform-4 in assisting HSV-1 entry and spread. Biochem Biophys Res Commun 2005;338:930−7.

Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. Curr Protoc 2021;1:e90.