Automated Detection of BK Virus in H&E Whole-Slide Images Using Weakly-Supervised Deep Learning and Interpretable Morphological Biomarkers

Sharifa Sahai[1,4]*, Ana D. Ramos-Guerra[2,3]*, Cristina Almagro-Pérez[1,4,5],
Guillaume Jaume[1,4], Andrew Zhang[1,4,5], Helmut Rennke[4], Astrid Weins[4],
Juan E. Ortuño[2,3], Maria J. Ledesma-Carbayo[2,3], and Faisal Mahmood[1,4]

[1] Harvard Medical School, Boston, MA, USA
[2] Biomedical Image Technologies, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
[3] Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain
[4] Department of Pathology, Mass General Brigham, Boston, MA, USA
[5] Health Sciences and Technology, Harvard-MIT, Cambridge, MA, USA
*Equal contribution and corresponding authors
`sharifasahai@g.harvard.edu, ana.ramos.guerra@upm.es`

Abstract. Detecting BK Virus (BKV) is crucial for managing post-transplant outcomes in kidney patients. While BKV is typically identified using SV40 immunohistochemistry (IHC), this method is time-consuming, limited by tissue availability and resource-intensive, especially in low-resource settings. Recent advances in computational pathology have shown potential for automating disease detection from Hematoxylin and Eosin (H&E)-stained images, though BKV detection remains understudied due to its low prevalence and limited data. We hypothesize that BKV-positive cells exhibit unique morphological patterns in H&E-stained tissue, detectable via computational methods. To address this, we developed BKVision, a weakly-supervised deep learning model for BKV detection in H&E whole-slide images (WSIs). Trained on 3,734 WSIs, BKVision achieves an F1-score of 0.984 ± 0.008 on a test cohort of 936 slides. Additionally, we conducted a morphological analysis on 774 H&E image patches, extracting 37 human interpretable features and validating them against IHC with pathologist guidance. This identified 11 cell attributes, such as nuclear enlargement and chromatin texture changes, that distinguish BKV-positive from negative cases. These findings highlight the potential to enhance BKV diagnostic criteria by integrating these identified morphological features. BKVision demonstrates the potential of computational methods to provide accurate, accessible, and interpretable BKV detection without the need for IHC, offering a cost-effective alternative in low-resource settings while revealing key morphological features of BKV infection.

Keywords: Renal Allograft · Deep Learning · Human-interpretable Biomarkers.

1 Introduction

Transplantation significantly improves survival for patients with end-stage organ failure; however, it is associated with risks such as BK virus (BKV) reactivation due to immunosuppression[6, 19]. BKV, a dormant polyomavirus in the kidney and urinary tract, can cause nephropathy and lead to the loss of the transplant if not promptly detected and managed [6]. The incidence of BKV reactivation within the first year post-kidney transplantation is approximately 5%[26]. Notably, BKV detection is not limited to renal transplant recipients; it also poses a challenge in non-renal transplant patients (e.g., lung, heart), where it can affect native kidneys[2, 8, 9]. The absence of routine BKV-specific staining and reliable biomarkers often results in delayed diagnosis and poorer clinical outcomes[2, 9].

The current standard for diagnosing BKV in renal transplant biopsies relies on immunohistochemistry (IHC) SV40 staining [19]. However, IHC staining is time-consuming, consumes limited diagnostic material, and requires specialized laboratory infrastructure, making it particularly challenging in low-resource settings [19]. Furthermore, the low prevalence of BKV has hindered the exploration of computational approaches for its diagnosis, resulting in a significant research gap. While automated identification of BKV remains underexplored, this gap arises primarily from challenges in data acquisition rather than a lack of clinical significance. BKV detection on H&E-stained slides presents significant diagnostic difficulties, necessitating additional IHC confirmation, which increases turnaround times and costs. Despite its rarity, the identification of BKV is critical for improving transplant efficacy, as delayed diagnosis can lead to graft loss and adverse patient prognoses.

In contrast to SV40 staining, which is costly and susceptible to sampling errors, Hematoxylin and Eosin (H&E) staining is routinely performed on multiple tissue blocks during standard clinical care for renal transplant patients [1]. Therefore, an automated approach for BKV detection directly from H&E could provide a cost-effective and scalable solution for widespread screening. To address this need, we introduce BKVision, the first weakly-supervised deep learning model for BKV detection in H&E whole-slide images (WSIs). Our method not only detects BKV but also emphasizes interpretability by identifying human-interpretable morphological features linked to BKV-positive cells (Figure 1).

In summary, our contributions are (1) we present the first weakly-supervised classification model for BKV detection in H&E whole-slide images; (2) we propose a method for morphological characterization of BKV in H&E; identifying 11 human-interpretable features linked to BKV-positive cells; and (3) we demonstrate that the identified biomarkers align with BKV-positive regions confirmed by IHC and validated by expert clinicians.

2 Related Work

AI-assisted diagnosis and automatic classification of H&E-stained histology images have demonstrated superior performance to pathologists for common diseases when enough data is available [10, 21, 3, 4, 24, 25]. Deep learning (DL) has also shown promise in renal allograft tissue for quantification of relevant structures such as atrophic tubules and disease classification [20, 15]. However, these efforts have predominantly focused on prevalent conditions such as rejection, leaving a critical gap in the automated detection of rare but clinically significant infections like BKV. Despite the advent of new DL technologies, only one prior study has explored BKV detection from histology [20], and its ability to distinguish BKV from other diseases, particularly rejection, remains limited.

The detection of BKV is critical for treatment direction in transplant patients, yet is complicated by the virus's low prevalence and potential limited tissue availability, resulting in notable scarcity of both computational and clinical studies [19, 7, 20]. Conventional strategies for BKV detection have explored avenues outside histopathological examination, relying on serological assays, clinical parameters [22], or single-cell analysis [27]. However, these methods diverge from the direct examination of histological samples, which remains the gold standard for diagnosing pathologies in transplant medicine. These observations underscore the interest in diagnostic tools tailored to BKV detection from the tissue morphology.

Recent research related to foundational models in computational pathology has demonstrated robust performance in downstream vision and visual language tasks [29]. Combining low-dimensional feature representations for gigapixel whole slide images (WSI) and semantic spatial analysis [18] introduces a new generation of diagnostic tools that improve both accuracy and interpretability in transplant pathology. Our study aims to address the gap in BKV detection by combining deep learning enhanced by state of the art foundational models with morphological analysis, providing an effective and accessible approach to identifying BKV in renal transplant recipients.

3 Methods

3.1 Patch encoding

Given an H&E-stained slide as input, we follow the multiple instance learning (MIL) paradigm [16], which consists of segmenting tissue regions, tesselating the tissue into patches, and extracting patch embeddings using a pre-trained feature encoder. Specifically, we extract non-overlapping 256×256 patches, which we encode using four feature encoders: ResNet50 [14] pre-trained on ImageNet, CTransPath [28] pre-tranined on TCGA and PAIP dataset, and CONCH [23] and UNI2-h [5], both pre-trained on proprietary datasets. We denote the resulting patch embeddings as $f_s \in \mathbb{R}^{N_s \times d}$, where N_s is the number of patches in the slide s, and d is the embedding dimension.

3.2 Slide classification

To generate slide predictions from patch embeddings, we use the clustering-constrained attention multiple instance learning (CLAM) architecture [24]. CLAM

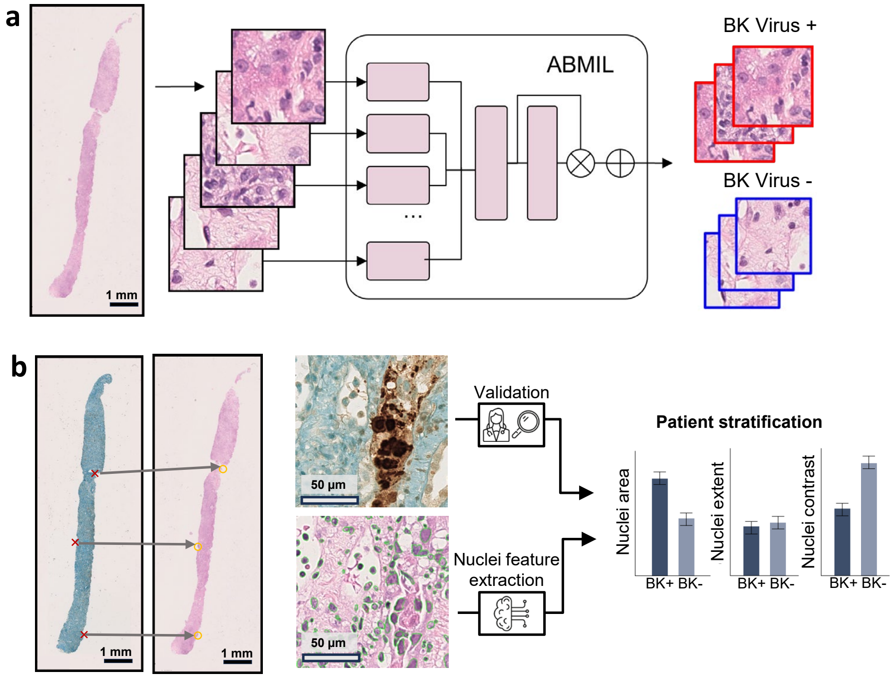

Fig. 1. Overview of BKVision. a. BKV detection from H&E-stained histological sections is accomplished through weakly supervised attention-based multiple instance learning (ABMIL). **b.** Quantitative analysis of cellular morphology to identify human-interpretable features linked to BKV, with validation on SV40 immunohistochemistry staining.

uses a trainable slide-level pooling function ϕ that learns to generate attention weights α for each patch-level feature and takes a weighted sum of the patch features as the slide-level feature $F_s \in \mathbb{R}^{1\times1024}$ for ResNet50, $F_s \in \mathbb{R}^{1\times768}$ for CTransPath, $F_s \in \mathbb{R}^{1\times512}$ for CONCH, and $F_s \in \mathbb{R}^{1\times1536}$ for UNI2-h. Formally, $F_s = \sum_{n=1}^{N_s} \alpha_n f_{s,n}$ where $\alpha_n = \phi(f_{s,n})$ and $f_{s,n}$ represents the nth patch feature in f_s. The slide-level feature F_s is used to predict logits for the positive class via a classification head. In addition to the cross-entropy objective, CLAM uses a patch-level clustering loss to encourage the linear separation of high-attention patches from low-attention patches.

3.3 Post-hoc morphological analysis

The morphological manifestations of BKV infection in renal tissues are often subtle, presenting challenges for traditional histopathological analysis in H&E. To address this, we leveraged ranked patches from our attention-based model to automatically identify regions likely to contain infected cells and characterize the morphological changes driving patient stratification. Fine-grained attention heatmaps were generated using the attention-pooling operation of CLAM, by patching the original slide with 0.5 overlap and computing attention weights for these overlapping patches.

Corresponding SV40 WSIs from BKV-positive cases were used to validate that H&E patches highlighted by the model correlated with BKV-positive regions or exhibited infection-related morphological changes. Adjacent SV40-H&E images were elastically registered [12] to correct for rotations and deformations during slide mounting. An expert pathologist annotated nuclei in highly-attended H&E patches suspected of infection.

Nuclei from the top 30 high-attention patches associated with positive BKV cases in the held-out set were segmented using the HoVerNet model [13] pre-trained on the PanNuke dataset [11]. In consultation with pathologists, we extracted 37 human-interpretable features from the histocartography library [17], selected for their relevance to BKV detection, related to nuclear morphology, size, shape, and topology. These features were compared to those obtained from annotated representative regions of BKV-negative subjects. Features significantly differing between positive and negative cases were identified using unpaired t-tests (p-value < 0.05) and log2 fold change (FC) ratios (log2FC $>|0.6|$, or $\approx 50\%$ increase compared to baseline), with Bonferroni correction for multiple hypothesis testing.

3.4 Dataset description

BKVision was developed on a cohort of 4,679 WSIs from 1,077 renal allograft biopsy cases collected from 2013-2022. Of these cases, 39 were positive, and 1,038 were negative for BKV. The dataset includes a diverse range of pathologies, with 39% of cases involving rejection, the most common differential diagnosis for BKV, ensuring representation of the patient population. Each case includes 1 to 6 H&E slides. For 225 cases, a single SV40 stain was performed in consecutive serial sections to the H&E, prompted by elevated BKV viremia levels in plasma, indicating a potential BKV infection. A selection of these SV40 stains was utilized for post-training evaluation and analysis. All slides were scanned at 20× on a Hamamatsu S210. BKV diagnosis was determined by an expert renal pathologist, initially based on the BKV viral load in plasma. If the viral load suggested a potential infection, further confirmation was sought by examining the presence of infected nuclei in the corresponding SV40 stains of the cases under investigation.

3.5 Split and data augmentation

The dataset was partitioned into train/validation/test splits (70%/10%/20%) stratified by the presence of BKV, with all slides from the same patient biopsy placed into the same set. We performed 5-fold cross-validation and reported the mean and standard error on the test set. The final cohort comprises $3,743$ slides

Table 1. Weakly supervised BKV classification in H&E. Performance metrics at the patient level in the independent held-out set with the minority class upsampled by 0%, 40%, 100%, or 140% (UR, %) for ResNet50 (RN50), CTransPath (CTP), UNI2-h and CONCH encoders.

	UR	**F1 Score**	**Precision**	**Recall**	**Bal ACC**	**AUC**
RN50	0	0.973±0.006	0.972±0.004	0.977±0.009	0.664±0.030	0.925±0.014
	40	0.965±0.012	0.963±0.003	0.972±0.023	0.581±0.072	0.935±0.025
	100	0.973±0.011	0.972±0.002	0.977±0.019	0.664±0.045	0.933±0.016
	140	0.965±0.016	0.963±0.003	0.972±0.025	0.581±0.038	0.931±0.030
CTP	0	0.973±0.006	0.972±0.002	0.977±0.001	0.664±0.030	**0.944±0.062**
	40	0.973±0.005	0.972±0.002	0.977±0.009	0.664±0.044	0.911±0.032
	100	0.979±0.006	0.979±0.001	0.981±0.001	0.748±0.008	0.855±0.033
	140	0.979±0.012	0.979±0.005	0.981±0.018	0.748±0.035	0.891±0.023
UNI2-h	0	0.962±0.012	0.967±0.003	0.958±0.019	0.736±0.013	0.892±0.063
	40	0.968±0.020	0.970±0.005	0.967±0.033	0.740±0.022	0.831±0.080
	100	0.968±0.007	0.970±0.001	0.967±0.011	0.740±0.012	0.849±0.060
	140	0.979±0.007	0.979±0.004	0.981±0.010	0.748±0.016	0.878±0.080
CONCH	0	0.972±0.005	0.972±0.003	0.972±0.009	0.743±0.033	0.926±0.044
	40	0.979±0.006	0.979±0.003	0.981±0.010	0.748±0.027	0.878±0.081
	100	0.959±0.017	0.966±0.004	0.953±0.027	0.733±0.028	0.942±0.048
	140	**0.984±0.008**	**0.986±0.001**	**0.986±0.015**	**0.750±0.044**	0.917±0.059

(864 patients) in the train-val set and 936 slides (213 patients) in the independent test set. Due to the extreme class imbalance in the data (less than 4% positive BKV slides), we applied upsampling of the minority class. Specifically, during training, we upsampled the BKV-positive class at the slide-level (n=130 slides) by adding more positive-case patches into each batch during training. This approach ensured a more balanced representation of BKV-positive cases, improving the model's ability to learn relevant morphological patterns.

3.6 Implementation details

Our model underwent training across 200 epochs, incorporating an early stopping mechanism to prevent overfitting. We used a batch size of 1, optimizing the training process with the Adam optimizer, set at a learning rate of 0.0002 and a weight decay rate of 0.00001. Nvidia RTX 3090 GPUs were used. Recognizing the potential variability in BKV cell distribution across different slides from the same patient, our classification strategy into BKV positive or negative categories was based on the maximum probability observed among all slides per patient, ensuring a comprehensive and sensitive detection approach. The classification threshold was determined by the validation cohort, specifically chosen to maximize the mean F1 score across all validation folds for each model.

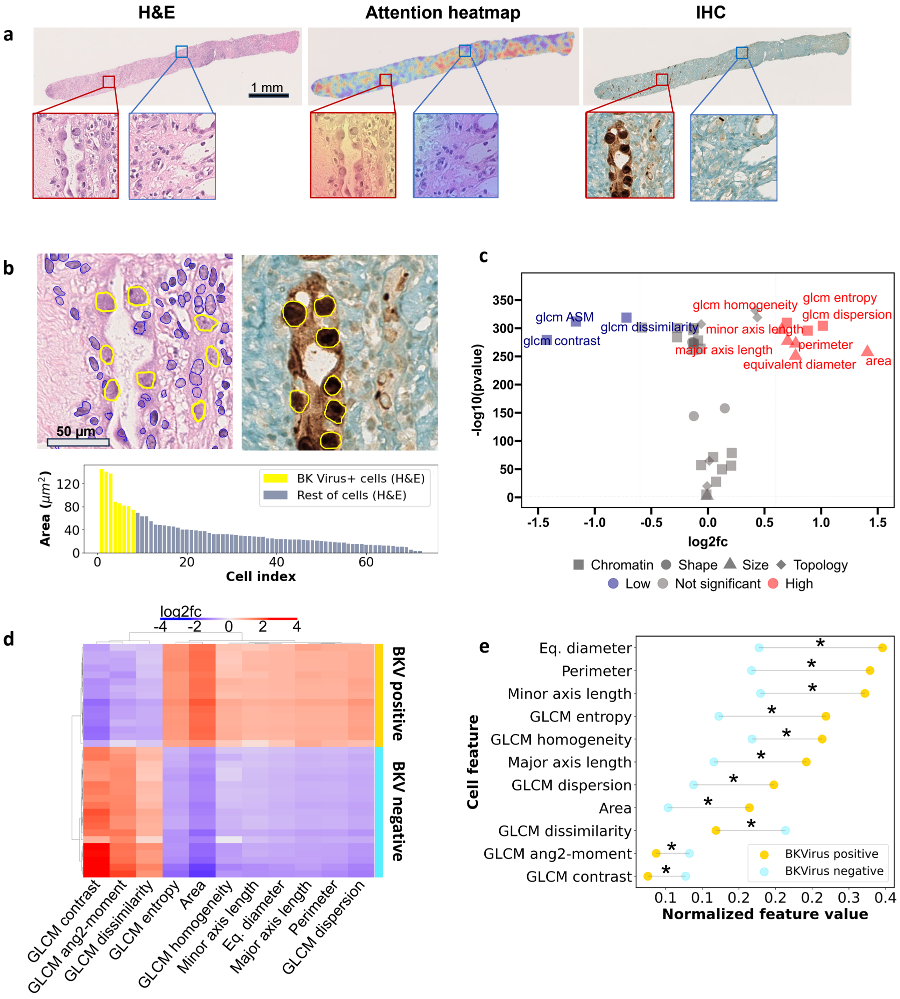

Fig. 2. Morphological characterization of BKV in H&E imaging. a. Visualization of BKV positive H&E slide with corresponding attention heatmap and adjacent IHC slide. **b.** Example of a highly attended H&E patch with registered IHC patch. Nuclear segmentation highlights BKV-positive nuclei (yellow) *vs.* other cells (blue). Nuclear area distribution in that patch. **c.** Log2 fold change of 37 human-interpretable features. Features in blue and red are statistically different than the control (t-test: $p < 0.001$ with log2FC values$>|0.6|$, or 50% increase from control). **d.** Heatmap (log2 Fold Change) summarising BKV positive features relative to BKV negative (control). Rows represent test patients. **e.** Human-interpretable features characterizing BKV morphology (p-value < 0.001 indicated with *).

4 Results

4.1 Classification

To identify an effective experimental setup, we tested various encoders and minority class upsampling rates (0%, 40%, 100%, 140%). The combination of extreme upsampling (140%) with the CONCH patch encoder achieved the highest classification performance across multiple metrics, including F1 score, precision, recall, and balanced accuracy (Table 1). While the CTransPath baseline (0% upsampling) performed best in terms of AUC, this metric is less reliable for imbalanced datasets.

The results indicate that CONCH outperforms ResNet50, CTransPath, and UNI2-h in extracting representative features of BKV infection, likely due to its training on more diverse pathological data. The overrepresentation of BKV-positive cases also improves detection of infection-related morphological patterns in H&E-stained histology, while maintaining accuracy for BKV-negative cases.

4.2 Nuclear characterization

Among the 37 features analyzed, 11 exhibited statistically significant differences between BKV-positive and BKV-negative cases. These included size-related metrics such as nuclear area, perimeter, and equivalent diameter, which were significantly larger in BKV-positive cells (Fig. 2d-e). Additionally, cells within high-attention patches showed a more uniform texture and reduced intensity variance (GLCM homogeneity and contrast), consistent with the hyperchromatic and "ground glass" appearance characteristic of BKV-infected nuclei. These findings align with nuclear enlargement and morphological changes in BKV infection [19].

Interpretable visualization plays a pivotal role in our approach. Heatmaps generated from attention weights identified regions highly predictive of BKV infection, capturing cellular-level pathological changes such as tubular cell injury and lymphoplasmacytic inflammation. Validation through co-registered SV40 IHC images confirmed the alignment of high-attention regions with BKV-positive cells (Fig. 2a). Pathologist annotations of infected cells, guided by SV40 IHC, further corroborated distinct topological features, particularly variations in cell size, distinguishing BKV-infected from healthy cell populations (Fig. 2b). Notably, BKV-infected nuclei were characterized by their enlargement within the tubular epithelium, consistent with prior clinical observations [19].

5 Conclusion

Our study demonstrates high-performance detection of BKV in renal transplant recipients using solely H&E-stained WSIs, providing a preliminary screening tool that could reduce reliance on specialized SV40 IHC tests. This approach accelerates diagnosis and improves accessibility, particularly in resource-limited settings. Through post-hoc analysis, we identified human-interpretable morphological features, such as nuclear area and chromatin density, validated against

IHC-stained sections and expert pathologist assessments. These features serve as reliable markers for BKV infection and could immediately influence clinical decision-making, enabling pathologists to refine diagnostic criteria and guide treatment decisions. To promote reproducibility, we have detailed our methodology, including parameter training details, and will open-source training scripts. Data and model weights can be provided upon institutional approval. As a preliminary study, future work should explore other architectures to enhance performance and generalizability. Formal evaluation of the identified biomarkers is needed to validate their efficacy across diverse populations. Our work not only improves BKV detection but also serves as a blueprint for developing interpretable computational pathology tools for other diagnostic challenges.

6 Acknowledgments

The authors acknowledge the support of Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, under grant PID2022-141493OB-I00 (10.13039/501100011033/MCIN/AEI/ERDF, UE), co-financed by European Regional Development Fund (ERDF),'A way of making Europe'. Additionally, this work has been developed with the financial support of Instituto de Salud Carlos III (ISCIII) project Immune4ALL (PMP22/00054) and the Next Generation EU funds. The authors further acknowledge partial funding from the MAGERIT-CM project (TEC-2024/COM-44), supported by the Research and Development in Technologies program of the Comunidad de Madrid.

7 Disclosure of Interests

The authors declare that they have no competing interests.

References

1. Amann, K., Haas, C.S.: What you should know about the work-up of a renal biopsy. Nephrology Dialysis Transplantation **21**(5), 1157–1161 (2006)
2. Barten, M.J., Zuckermann, A.: BK virus: A cause for concern in thoracic transplantation? Ann Transplant **23**, 310–321 (2018)
3. Bulten, W., Pinckaers, H., van Boven, H., et al.: Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study. The Lancet Oncology **21**(2), 233–241 (2020)
4. Campanella, G., Hanna, M.G., Geneslaw, L., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**(8), 1301–1309 (2019)
5. Chen, R.J., Ding, T., Lu, M.Y., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
6. Cohen-Bucay, A., Ramirez-Andrade, S.E., Gordon, C.E., et al.: Advances in BK virus complications in organ transplantation and beyond. Kidney Med **2**(6), 771–786 (2020)

7. Costigliolo, F., Lombardo, K., Arend, L.J., et al.: BK virus RNA in renal allograft biopsies. J Histochem Cytochem **68**(5), 319–325 (2020)
8. Crowhurst, T., Nolan, J., Faull, R., et al.: BK virus-associated nephropathy in a lung transplant patient: case report and literature review. BMC Infect Dis **20**(1), 600 (2020)
9. Egli, A., Helmersen, D.S., Taub, K., et al.: Renal failure five years after lung transplantation due to polyomavirus BK-associated nephropathy. Am J Transplant **10**(10), 2324–2330 (2010)
10. Esteva, A., Kuprel, B., Novoa, R.A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. nature **542**(7639), 115–118 (2017)
11. Gamper, J., Koohbanani, N.A., Benes, K., et al.: Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778 (2020)
12. Gatenbee, C.D., Baker, A.M., Prabhakaran, S., et al.: Virtual alignment of pathology image series for multi-gigapixel whole slide images. Nature communications **14**(1), 4502 (2023)
13. Graham, S., Vu, Q.D., Raza, S.E.A., et al.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis **58**, 101563 (2019)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
15. Hermsen, M., de Bel, T., den Boer, M., et al.: Deep learning-based histopathologic assessment of kidney tissue. Journal of the American Society of Nephrology : JASN **30**(10), 1968–1979 (2019)
16. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2132–2141 (2018)
17. Jaume, G., Pati, P., Anklin, V., et al.: Histocartography: A toolkit for graph analytics in digital pathology. In: MICCAI Workshop on Computational Pathology. pp. 117–128. PMLR (2021)
18. Jaume, G., Pati, P., Bozorgtabar, B., et al.: Quantifying explainers of graph neural networks in computational pathology. CVPR (2020)
19. Kant, S., Dasgupta, A., Bagnasco, S., et al.: BK virus nephropathy in kidney transplantation: A State-of-the-Art review. Viruses **14**(8) (2022)
20. Kers, J., Bülow, R.D., Klinkhammer, B.M., et al.: Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study. The Lancet Digital Health (2021)
21. Lipkova, J., Chen, T.Y., Lu, M.Y., et al.: Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies. Nature Medicine **28**(3), 575–582 (2022)
22. Liu, Y., Kong, C., Hu, H., et al.: Risk factors for bk virus infection in dcd donor kidney transplant recipients. Frontiers in Medicine **10** (2023)
23. Lu, M.Y., Chen, B., Williamson, D.F.K., et al.: A visual-language foundation model for computational pathology. Nature Medicine **30**(3), 863–874 (2024)
24. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., et al.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)
25. McKinney, S.M., Sieniek, M., Godbole, V., et al.: International evaluation of an AI system for breast cancer screening. Nature **577**(7788), 89–94 (2020)
26. Nickeleit, V., Singh, H.K., Randhawa, P., et al.: The banff working group classification of definitive polyomavirus nephropathy: morphologic definitions and clinical correlations. Journal of the American Society of Nephrology **29**(2), 680–693 (2018)

27. Procario, M.C., Sexton, J.Z., Halligan, B.S., et al.: Single-Cell, High-Content microscopy analysis of BK polyomavirus infection. Microbiol Spectr **11**(3), e0087323 (2023)
28. Wang, X., Yang, S., Zhang, J., et al.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical Image Analysis **81**, 102559 (2022)
29. Xu, H., Wang, M., Shi, D., et al.: When multiple instance learning meets foundation models: Advancing histological whole slide image analysis. Medical Image Analysis **101**, 103456 (2025)