OPEN ACCESS

Biomedical Engineering LETTER

The Innovation Informatics

CAR-Tumor-WSI: A benchmark dataset for automated tumor segmentation of tumor-bearing mice in CAR-T-based immunotherapy studies using whole-slide imaging

Xiangyu Zhao,[1,3,7] Yifan Yang,[2,4,5,7] Zixuan Lai,[2,4,5,7] Ying Huang,[2,4] Jiaxin Hou,[1,3] Ranran Zhang,[1,3] Zhi Lin,[2,4] Zhe Qu,[2,4,*] and Wenjian Qin[1,3,6,*]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[2]State Key Laboratory of Drug Regulatory Science, Beijing 102629, China

[3]University of Chinese Academy of Sciences, Beijing 101400, China

[4]National Institutes for Drug Control, National Center for Safety Evaluation of Drugs, Beijing 100176, China

[5]State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China

[6]State Key Laboratory of Biomedical Imaging Science and System, Shenzhen 518055, China

[7]These authors contributed equally

*Correspondence: quzhe@nifdc.org.cn (Z.Q.); wj.qin@siat.ac.cn (W.Q.)

Received: June 24, 2026; Accepted: October 9, 2026; Published Online: October 19, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100065

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Zhao X., Yang Y., Lai Z., et al. (2026). CAR-Tumor-WSI: A benchmark dataset for automated tumor segmentation of tumor-bearing mice in CAR-T-based immunotherapy studies using whole-slide imaging. *The Innovation Informatics* **2**:100065.

Dear Editor,

Histopathological evaluation of tumor-bearing animal models is essential for preclinical assessment of CAR-T immunotherapy, yet manual tumor delineation in whole-slide images (WSIs) remains labor-intensive and observer-dependent. To address the lack of dedicated benchmarks, we introduce CAR-Tumor-WSI, an expert-annotated dataset derived from preclinical CAR-T studies, and develop CLOPS, a closed-loop human–AI annotation pipeline combining semi-automatic labeling, model-assisted inference, expert review, and iterative refinement. Experiments using SAM-Path show that CLOPS improves annotation quality and segmentation performance while enabling the refinement of ambiguous lesion boundaries. CAR-Tumor-WSI provides a practical benchmark for automated tumor segmentation and reproducible pathological assessment in CAR-T immunotherapy studies.

INTRODUCTION

Whole-slide imaging (WSI) enables high-resolution digitization and large-scale pathological assessment of tumor-bearing animal models, supporting preclinical evaluation of precision oncology therapies, including CAR-T immunotherapy.[1] However, manual tumor identification in WSIs remains labor-intensive, time-consuming, and prone to inter-observer variability. Although deep learning has shown strong potential in computational pathology tasks such as breast cancer lymph node metastasis prediction,[2,3] lung cancer subtype classification,[4,5] and colon adenocarcinoma prognosis assessment,[6] automated tumor segmentation in histopathological WSIs from tumor-bearing animal models remains insufficiently explored. To fill this gap, we constructed CAR-Tumor-WSI, a high-quality expert-annotated dataset derived from tumor-bearing mouse models, and benchmarked advanced deep learning models for automatic tumor lesion segmentation, providing a practical foundation for efficient and reproducible pathological assessment in preclinical CAR-T and precision oncology studies.

IMAGE DATA SETS

Sample Source: The tumor-bearing mouse liver transplanted tumor lesion samples were obtained from 12 preclinical studies of CAR-T cell therapy products, conducted at the National Center for Safety Evaluation of Drugs, National Institutes for Drug Control (NIDC). Immunodeficient mice were used as hosts to establish a B-cell lymphoma disease model. Therapeutic products were administered via tail-vein injection. Five parallel experimental groups were included for comparative analysis: a negative control group, a solvent control group, a T cell control group, a low-dose group, and a high-dose group. Animals were monitored longitudinally, and tissue samples were collected at predefined time points for histopathological examination.

The entire workflow, from tissue collection to whole-slide imaging, was conducted under stringent quality control protocols shown in Figure 1A. First, organ tissues were collected and preserved using appropriate fixation methods before undergoing tissue processing. The tissues were dehydrated with gradient ethanol to ensure uniform dehydration and then embedded in paraffin after thorough wax infiltration. The paraffin blocks were sectioned into thin slices approximately 3 μm thick using a microtome, ensuring that the slices were of uniform thickness, smooth, and wrinkle-free. Subsequently, the sections were stained with hematoxylin and eosin (HE) and scanned using a digital slide scanner (NanoZoomer-S60 scanner C13210-01; Hamamatsu Photonics). The scanning parameters were set to a resolution of 0.23 μm/pixel at 40× magnification. All lesion areas were ensured to be within the effective field of view, with no imaging of artifacts.

Annotation Process: The annotation of tumor tissues was completed under the guidance of senior pathologists, following a closed-loop annotation strategy of "semi-automatic assistance–expert review–fully automated iteration." The process includes three main stages: (1) Preliminary annotation stage (Figure 1B): After task initiation, two junior annotators independently performed image identification and qualitative judgment of lesion categories, with 100% cross-validation of their classification results. Discrepancies were submitted to an arbitration expert for review and final classification. Further, annotators used the semi-automatic segmentation tool Qupath to delineate tumor regions,[7] followed by another 100% cross-revision of the contour results. The data then proceeded to model training, inference, and fine-tuning stages. (2) Model iteration stage: As shown in Figure 1C, a deep learning segmentation model was trained on the preliminary annotated data to predict across whole slides, identifying small, overlooked tumor foci and necrotic areas at liver tissue edges. (3) Expert review stage: As shown in Figure 1D, reviewers assessed the generalized prediction masks across the sample space, performing a 30% random sampling review of samples without disputes from the preliminary cross-validation. In case of discrepancies, samples were sent to arbitration experts for reclassification, while consistent samples underwent final verification. Additionally, arbitration experts conducted a 30% random review of lesion boundary annotations. This closed-loop annotation pipeline (CLOPS, Closed-Loop Annotation Pipeline for Pathology Slides) is detailed in Section 4.

CLOSED-LOOP ANNOTATION PIPELINE FOR PATHOLOGY SLIDES

To efficiently annotate the dataset for constructing a CAR-Tumor-WSI segmentation benchmark, we developed a semi-automated annotation pipeline, named CLOPS (Closed-Loop Annotation Pipeline for Pathology Slides), as illustrated in Figure 1C. Based on regional annotations and tumor tissue labels, regions of interest (ROIs) were segmented and tiled into 1024 × 1024 patches, each with a corresponding binary mask. For cold-start initialization of the annotation pipeline, 98 tumor-containing samples were manually annotated. Of these, 19 samples (19.4%) were randomly assigned to an independent test set. To prevent data leakage, all patches derived from the same whole-slide image (WSI) were restricted to a single fold during dataset partitioning. The final dataset for model training and evaluation comprised 36,623 patches, with strict fold independence and no inter-fold overlap, ensuring experimental reliability. Model generalization was subsequently

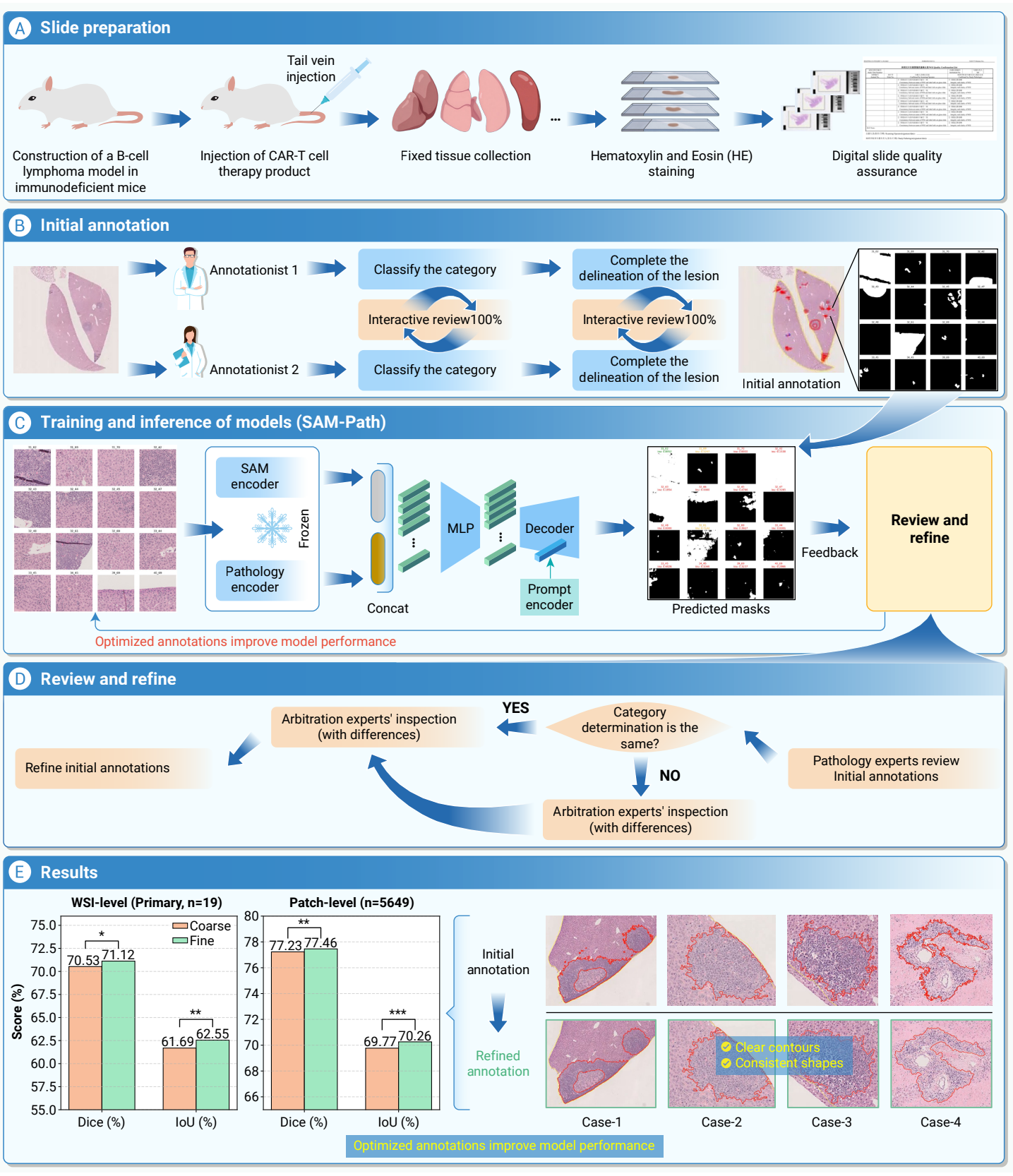


Figure 1. Comprehensive workflow and quality control for slide collection and annotation in CAR-T cell therapy (A) Digital pathology collection workflow for CAR-T cell therapy products; (B) Manually label for initial slide annotation; (C) Training and inference based on segmentation models: closed-loop annotation pipeline for pathology slides; (D) Expert review and annotation refinement; (E) Comparison results of quantitative and qualitative performance of segmentation with closed-loop annotation pipeline. Statistical significance was evaluated using paired t-tests (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

assessed on the held-out test set (19 samples; 5,649 patches). As the core segmentation model, SAM-Path was adopted in this study.[8] Built upon the Segment Anything Model (SAM) framework, SAM-Path integrates a HIPT pathology encoder and performs automatic semantic segmentation via trainable class prompts. In the practical implementation of CLOPS, single-fold experiments were conducted to enable rapid model training. Inference was

carried out on all WSIs, and the resulting segmentation outputs were used as annotation proposals to assist pathologist review. After 2–3 iterations of expert verification and correction, high-quality annotations were obtained.

TASKS, METRICS AND RESULTS

This study aims to construct a benchmark dataset for tumor segmentation in CAR-Tumor-WSIs. Baseline performance on the held-out test set was evaluated using two standard pixel-level segmentation metrics: Dice coefficient and Intersection over Union (IoU). As shown in Figure 1E, the annotation closed-loop optimization strategy substantially improved segmentation performance. When evaluated at the WSI level, the fine-label SAM-Path model significantly outperformed the coarse-label baseline, with the Dice coefficient increasing from 70.53% to 71.12% (+0.59%, $p = 0.0204$) and IoU from 61.69% to 62.55% (+0.86%, $p = 0.0022$). Consistent statistical significance was also observed at the patch level (Dice: +0.23%, $p = 0.0011$; IoU: +0.49%, $p < 0.0001$).

Whole slide inference results in Figure 1E demonstrate the ability to detect minute lesions and generate smooth boundaries that conform to biological morphology. These advantages rely on the model's generalized feature representation capabilities learned from the dataset. In summary, the CLOPS system, through a "human-machine collaborative" iterative mechanism, improved annotation efficiency while ensuring data quality, correcting annotation consistency in boundary-ambiguous regions.

CONCLUSION

This study constructed CAR-Tumor-WSI benchmark dataset, the first whole slide image dataset of tumor-bearing animal tumors for CAR-T cell therapy toxicology assessment. We developed the CLOPS pipeline, achieving efficient and refined annotation of tumor ROIs in whole slide images through iterative optimization. Independent test set validation demonstrated that models trained on CLOPS-optimized data significantly outperformed those trained on initial semi-automated annotations across multiple key metrics, validating the effectiveness of this closed-loop annotation strategy.

REFERENCES

1. Madabhushi A. and Lee G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **33**:170–175. DOI:10.1016/j.media.2016.06.037
2. Bándi P., Balkenhol M., Van Dijk M., et al. (2023). Continual learning strategies for cancer-independent detection of lymph node metastases. *Med. Image Anal.* **85**:102755. DOI:10.1016/j.media.2023.102755
3. Yu J.G., Wu Z., Ming Y., et al. (2023). Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images. *Med. Image Anal.* **85**:102748. DOI:10.1016/j.media.2023.102748
4. Zheng Y., Gindra R.H., Green E.J., et al. (2022). A graph-transformer for whole slide image classification. *IEEE Trans. Med. Imaging* **41**:3003–3015. DOI:10.1109/TMI.2022.3176598
5. Hu Y., Sirinukunwattana K., Li B., et al. (2025). Self-interactive learning: Fusion and evolution of multi-scale histomorphology features for molecular traits prediction in computational pathology. *Med. Image Anal.* **101**:103437. DOI:10.1016/j.media.2024.103437
6. Hou J., Zhang R., Xie Y., et al. (2025). Multimodal deep learning for cancer prognosis prediction with clinical information prompts integration. *npj Digit. Med.* **9**:76. DOI:10.1038/s41746-025-02257-y
7. Bankhead P., Loughrey M.B., Fernández J.A., et al. (2017). QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**:16878. DOI:10.1038/s41598-017-17204-5
8. Zhang J., Ma K., Kapse S., et al. (2023). SAM-path: A segment anything model for semantic segmentation in digital pathology. *MICCAI 2023.* **14393**:161-170. DOI:10.1007/978-3-031-47401-9_16

FUNDING AND ACKNOWLEDGMENTS

This work was supported in part by the State Key Laboratory of Drug Regulatory Science (2025SKLDRS0364), National Natural Science Foundation of China (No. 62271475), Guangdong Youth Talent Program (2024TQ08A386), and the Youth Innovation Promotion Association CAS (2022365). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: Xiangyu Zhao, Zhe Qu, Wenjian Qin; Methodology: Xiangyu Zhao, Wenjian Qin, Jiaxin Hou; Investigation: Yifan Yang, Zixuan Lai, Ying Huang; Software: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Ying Huang; Validation: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Ying Huang, Ranran Zhang; Formal Analysis: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Ying Huang, Wenjian Qin; Data Curation: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Ying Huang; Visualization: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Ying Huang; Resources: Zhi Lin, Zhe Qu, Wenjian Qin; Funding Acquisition: Zhe Qu, Wenjian Qin; Project Administration: Zhi Lin, Zhe Qu, Wenjian Qin; Writing – Original Draft: Xiangyu Zhao, Yifan Yang, Zixuan Lai, Wenjian Qin; Writing – Review & Editing: Ranran Zhang, Zhi Lin, Zhe Qu, Wenjian Qin; Supervision: Zhe Qu, Wenjian Qin. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

This study was approved by the Institutional Animal Care and Use Committee of the Institute for National Center for Safety Evaluation of Drugs, National Institutes for Drug Control (IACUC-2019-048) and was conducted in strict accordance with relevant guidelines and regulations for the care and use of laboratory animals.

DATA AND CODE AVAILABILITY

The data that support the findings of this study are openly available in Science Data Bank at https://cstr.cn/31253.11.sciencedb.016p3.