Thursday, April 30, 2026

TReC 2026 Project Proposal Submission Form

Submit your project proposal for the 7th TRAIL Research Camp (August 24th - September 4th, 2026, Lausanne, Switzerland). Please complete all required sections and submit your proposal before April 30th, 01:00 PM (CET).

Administrative Data

Full Name of Team Leader Isinsu Katircioglu

Contact Email isinsu.katircioglu@chuv.ch

Project Information

Project Title Beyond the Patch: Context Aggregation Strategies for Foundation Model-Driven Tertiary Lymphoid Structure Segmentation

Profile of the Team Leader(s) & Expected Team Composition

Isinsu Katircioglu: Senior Data Scientist with expertise in computer vision, including object segmentation and tracking, self-supervised learning, and vision-language foundation models, currently focused on morpho-molecular representation learning for clinical decision support.

Caner Ercan: Postdoctoral researcher in computational pathology and translational molecular pathology, focusing on developing models capturing the spatial dynamics of the tumor microenvironment to identify outcome-relevant characteristics from pathology images.

Have you already identified potential team members for your project? No

Domain of Application Healthcare

Scientific Theme Active & Weakly-supervised Learning

Proposal Content

Abstract

Tertiary lymphoid structures (TLSs) and their germinal centers (GCs) are key biomarkers for cancer prognosis and immunotherapy success. While specialized deep-learning models can automate TLS quantification on H&E slides, their reliance on extensive manual annotations limits their portability to new clinical datasets.

This project investigates whether pathology foundation models (FMs), such as UNI [2], Hibou [9], and CONCH [8], can overcome this bottleneck by exploring two context aggregation strategies. The first strategy uses FM embeddings integrated into a multi-resolution segmentation framework with cross-attention mechanisms, where topology is learned implicitly through dense connectivity and can capture global features. The second strategy uses explicit graph neural networks with predefined topology, where FM embeddings serve as node features and learned neighborhood aggregation models long-range and fine-grained context. The project will benchmark both context aggregation strategies against a specialist

dual-branch U-Net baseline [16,17] trained on TCGA data, assess label efficiency to determine if FMs require fewer annotations to reach high performance, and evaluate cross-cohort generalization by testing both strategies trained on TCGA data against an independent internal dataset.

A key objective is to understand the relative advantages of explicit graph-based context modeling with predefined topology versus implicit cross-attention-based approaches for TLS segmentation. The neutral hypothesis, that FM-based variants reach baseline performance with substantially fewer labels and generalize better across cohorts, is treated as a question to be measured, not assumed. Both positive and negative findings are scientifically informative and inform clinical translation.

Background Information & Problem Statement

Tertiary lymphoid structures (TLSs) are organised aggregates of immune cells that form in chronically inflamed tissues, including solid tumours. Their density and maturity, in particular the presence of germinal centres (GCs), have been associated with improved prognosis and response [3,14] to immunotherapy across several cancer types. Reproducible, automated TLS / GC quantification on routine H&E whole-slide images (WSIs) is therefore an active translational goal.

In this context, the current strong methodological baseline for automated TLS / GC quantification is a dual-branch specialist U-Net segmentation model [16,17]. Trained from scratch on approximately one thousand manually annotated TCGA slides across three cancer types, the specialist model demonstrates strong agreement with expert pathologists in TLS detection. Two practical limitations of this baseline motivate the present project: strong label dependence and limited cross-cohort generalisation.

More recently, self-supervised pathology foundation models (FMs) [1,2,5,6,8,9,13,19,20,21] have reshaped the field. However, their use for multi-resolution TLS/GC segmentation is not well studied. Recent benchmarks [10,18] show that (i) patch-level performance does not reliably transfer to dense segmentation, and larger FMs are not always better; (ii) optimal adapters/decoders and the role of multi-resolution context in FM-based segmentation remain open questions.

In parallel, an emerging alternative to cross-attention is graph neural networks with explicit topology for modelling spatial context. In this setting, WSI patches or regions are represented as nodes with FM embeddings, and context is aggregated via neighbourhood connections, capturing long-range dependencies without multi-resolution feature fusion [15]. However, it remains unclear whether this approach matches or exceeds cross-attention methods in terms of performance and efficiency.

Problem Statement:
The dual-branch specialist model is a strong but data-intensive baseline and does not leverage modern self-supervised foundation models (FMs). This project investigates how FMs can be adapted for TLS/GC segmentation and whether explicit graph-based topology offers advantages over implicit cross-attention. Key questions are: (i) do FMs improve data efficiency in dense segmentation; (ii) does implicit (cross-attention, multi-resolution) or explicit (graph-based) context aggregation better capture spatial relationships; and (iii) do FM-based models generalise better than the specialist baseline across cohorts. To address this, we compare two strategies: (1) cross-attention-based aggregation using frozen FM encoders with implicit multi-resolution context, and (2) graph-based aggregation using predefined topology with FM node embeddings. Both are evaluated for label efficiency and cross-cohort generalisation.

Project Objectives & Concrete Implementation

The project pursues three objectives, ordered by priority. Throughout this section, *specialist model* refers to the existing dual-branch U-Net segmentation model used as the baseline.

O1 - Build and benchmark FM-based context aggregation strategies

Strategy 1: Cross-attention-based context aggregation with FM encoders explores FM encoders as alternatives to the specialist model's encoders. This strategy compares three foundation models with

complementary properties: UNI [2] (vision-only, DINOv2 [11], ViT-H), Hibou [9] (vision-only, DINOv2, ViT-L), and CONCH [8] (vision–language [12], smaller backbone). Two variants are considered: **(A)** a single-branch model operating at target resolution to isolate the effect of multi-resolution context in FM encoders; and **(B)** a multi-resolution dual-branch model integrating the FM encoder with CNN decoding [4] and cross-branch context fusion. Both strategies can rely on fine-tuning via LoRA [7].

Strategy 2: Graph-based context aggregation with FM embeddings models spatial context through explicit graph topology, where WSI patches become nodes carrying FM embeddings as features. Graph-based aggregation mechanisms (e.g., graph attention networks) capture long-range dependencies at the same magnification through learned neighborhood relationships. This strategy contrasts with the hard-wired multi-resolution fusion of the cross-attention approach, exploring whether explicit topology can match the context modeling of multi-resolution feature splicing.

Expected outcome: A benchmark comparing cross-attention (Variants A & B) and graph-based context aggregation on the TCGA test set, reporting per-class Dice score (TLS, GC, Rest) and object-level F1 at slide level. The study identifies the best-performing strategy, providing a quantitative comparison of modelling approaches; even negative results are informative and publishable in this comparative setting.

O2 - Quantify label efficiency

Taking the best FM-based variant from O1, label efficiency is assessed by retraining both that variant and the specialist baseline across three label fractions: 25 %, 50 %, and 100 % of the available training slides.

Expected outcome: A label-efficiency curve comparing the best FM-based variant with the specialist baseline, plotting per-class Dice against labelled-slide fraction. The result quantifies whether and by how much FM-based models match baseline performance with fewer labels.

O3 - Evaluate cross-cohort domain shift

The TCGA-trained models from O1 (both cross-attention-based and graph-based variants) are evaluated zero-shot on the in-house cohort (10 expert-annotated H&E slides), reporting per-class Dice and object-level F1. If a held-out fine-tuning split of the in-house cohort is available, short fine-tuning on a small subset of slides is explored to characterise recovery curves.

Expected outcome: A TCGA vs in-house comparison of all models (cross-attention Variants A & B, graph-based, and specialist baseline), with optional fine-tuning recovery curves. This directly quantifies cross-cohort robustness and determines whether FM-based strategies improve domain generalisation for clinical use.

Do you plan to deliver, as an outcome of your project, a reusable “brick” for the TRAIL Factory (https://factory.trail.ac/en/home_page) that could later be transferred and converted into a company process?

No

Project Dataset

The dataset comprises 1,019 manually annotated H&E whole-slide images (WSIs) with Tertiary Lymphoid Structure (TLS) and Germinal Center (GC) annotations. The images are sourced from The Cancer Genome Atlas (TCGA) cohorts (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) for Lung Squamous Cell Carcinoma (LUSC), Kidney Renal Clear Cell Carcinoma (KIRC), and Bladder Urothelial

Carcinoma (BLCA). All corresponding TCGA SVS files for the H&E WSIs have been downloaded and are available for the workshop. The annotations (https://zenodo.org/records/10614928, https://zenodo.org/records/10635034) are provided as XML files, in which each TLS and GC region is represented as a polygon in WSI pixel coordinates. GC polygons are spatially nested within their corresponding TLS polygons.

The in-house cohort consists of 10 expert-annotated H&E slides used for evaluating cross-cohort domain shift and generalization in Objective O3.

Detailed Work Plan

The team leaders will prepare the training, validation, and test datasets, as well as a baseline specialist model based on dual U-Net branches, trained specifically for the TLS/GC semantic segmentation task for comparison, before the start of the workshop.

Week 1

Cross-attention-based context aggregation strategy: Explore FM-based segmentation variants (Variants A & B) with FM encoders (UNI, Hibou, or CONCH). Compare the dual-branch multi-resolution approach against single-branch target-resolution-only, using frozen encoder vs LoRA as the adaptation strategy. Design decoders for FM features that progressively upsample spatially coarse ViT tokens to pixel resolution.

Graph-based context aggregation strategy: Explore graph-based context aggregation where patch graphs carry FM embeddings as node features. Implement graph-based neighborhood aggregation using attention mechanisms to model long-range dependencies. Begin with single-magnification graphs as the primary variant in Week 1.

Deliverables by end of Week 1: Benchmark both cross-attention variants and the graph-based single-magnification approach against the specialist baseline on the TCGA test split.

Week 2

Label-efficiency curves and cross-cohort evaluation: Select the best performing variant from each context aggregation strategy (cross-attention-based and graph-based) identified in Week 1. Train both with progressively smaller fractions of annotated TCGA slides (25%, 50%, 100%) and compare against the specialist baseline. Evaluate all variants zero-shot on the in-house cohort (10 expert-annotated H&E slides), reporting per-class Dice and object-level F1 for cross-cohort robustness.

Graph-based stretch goal (if on schedule): Extend the graph-based approach to multi-scale graph edges linking target (0.5 µm/px) and context (2.0 µm/px) magnifications using learned attention-based fusion.

Deliverables by end of Week 2: Final benchmark comparison table (cross-attention Variants A & B, graph-based strategy, specialist baseline) with label-efficiency curves and cross-cohort evaluation.

Bibliographic References

[1] Campanella, G., Chen, S., Singh, M., et al. A clinical benchmark of public self-supervised pathology foundation models. *Nature Communications*. 2025;16(1):412.

[2] Chen, R. J., Ding, T., Lu, M. Y., et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*. 2024;30(2):317-326. (UNI / UNI2)

[3] Chen, Z., Wang, X., Jin, Z., et al. Deep learning on tertiary lymphoid structures in hematoxylin-eosin predicts cancer prognosis and immunotherapy response. *npj Precision Oncology*. 2024;8(1):52.

[4] Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R. Masked-attention mask transformer for universal image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022:1290-1299. (Mask2Former)

[5] Ding, T., Chen, R. J., Lu, M. Y., et al. Multimodal whole-slide foundation model for pathology. *Nature Medicine*. 2025;31(1):152-164. (TITAN)

[6] Filiot, A., Gidaris, S., Raille, Z., et al. Scaling self-supervised learning for histopathology with masked image modeling. *MedRxiv (2023): 2023-07*.

[7] Hu, E. J., Shen, Y., Wallis, P., et al. LoRA: Low-rank adaptation of large language models. *International Conference on Learning Representations (ICLR)*. 2022.

[8] Lu, M. Y., Chen, R. J., Chen, T. Y., et al. A visual–language foundation model for computational pathology. *Nature Medicine*. 2024;30(3):863-874. (CONCH)

[9] Nechaev, D., Pchelnikov, A., and Ivanova, E. Hibou: A Family of Foundational Vision Transformers for Pathology. *arXiv preprint arXiv:2406.05074*. 2024.

[10] Neidlinger, P., El Nahhas, O.S.M., Muti, H.S. et al. Benchmarking foundation models as feature extractors for weakly supervised computational pathology. Nat. Biomed. Eng. 2025. https://doi.org/10.1038/s41551-025-01516-3.

[11] Oquab, M., Darcet, T., Moutakanni, T., et al. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*. 2023.

[12] Radford, A., Kim, J. W., Hallacy, C., et al. Learning transferable visual models from natural language supervision. *International Conference on Machine Learning (ICML)*. 2021:8748-8763. (CLIP)

[13] Saillard, C., Jenatton, R., Llinares-López, F., Mariet, Z., Cahané, D., Durand, E., Vert, J.P.: H-optimus-0 (2024), https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0.

[14] Schumacher, T. N. and Thommen, D. S. Tertiary lymphoid structures in cancer. *Science*. 2022;375(6576):eadab943.

[15] Su, L., Liu, Z., Wu, J., et al. GCUNet: A GNN-Based Contextual Learning Network for Tertiary Lymphoid Structure Semantic Segmentation in Whole Slide Image. *arXiv preprint arXiv:2412.06129*. 2024.

[16] van Rijthoven, M., Obahor, S., Pagliarulo, F., et al. Multi-resolution deep learning characterizes tertiary lymphoid structures and their prognostic relevance in solid tumors. *Communications Medicine*. 2024;4(1):5.

[17] van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., and Ciompi, F. HookNet: multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. *Medical Image Analysis*. 2021;68:101890.

[18] Vitória, F., van Rijthoven, M., van der Laak, J., and Ciompi, F. Benchmarking computational pathology foundation models for semantic segmentation. *arXiv preprint arXiv:2602.18747*. 2026.

[19] Vorontsov, E., Bozkurt, A., Casson, A. et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med 30, 2924–2935 (2024). https://doi.org/10.1038/s41591-024-03141-0.

[20] Xu, H., Usuyama, N., Bagga, J. et al. A whole-slide foundation model for digital pathology from real-world data. Nature 630, 181–188 (2024). https://doi.org/10.1038/s41586-024-07441-w.

[21] Zimmermann, E., Shaikovski, G., Adam, K., et al. Virchow2: Scaling self-supervised mixed magnification models in pathology. *arXiv preprint arXiv:2408.00738*. 2024.

Eligibility & Evaluation

Does the project include multidisciplinarity between STEM & SSH?

No

We confirm that the Team Leader will be present for the full duration of TReC'26 if the project is selected (August 24th - September 4th, 2026, Lausanne, Switzerland)

I/We agree and confirm