TLPath: Leveraging the Digital Pathology Foundational model for Telomere Length Prediction

Anamika Yadav ANAMIKAY@SBPDISCOVERY.ORG
Kyle Alvarez KALVAREZ@SBPDISCOVERY.ORG
Akanimoh Adeleye SADELEYE@SBPDISCOVERY.ORG
Sanju Sinha SSINHA@SBPDISCOVERY.ORG
Sanford Burnham Prebys Medical Discovery Institute

Editors: Under Review for MIDL 2025

Abstract

Telomere dysfunction drives aging and age-related diseases, yet its large-scale study is hindered by reliance on specialized molecular assays, limiting clinical and research advancements. Here we present TLPath, a deep learning framework that leverages digital pathology foundational model, UNI, to predict bulk-tissue telomere length from routine H&E-stained images. The pipeline extracts morphological features from image patches and aggregates them into a whole slide-level representations, which are then used in a supervised model to accurately predict telomere length. These extracted features can predict bulk-telomere length with significant accuracy (> 0.51 in well-represented tissues), outperforming chronological age as a predictor (correlation $= 0.20$) and identifying age-discordant cases – detecting both accelerated telomeres shortening in young individuals and preserved telomeres in older individuals. Moreover, the mechanistic interpretation of TLPath reveals that its predictions are grounded in established cellular senescence markers such as the nuclear to cytoplasmic ratio and nuclear shape variation.
Keywords: Digital Patholgy, Foundational Model, Machine Learning

1. Introduction

Telomeres are protective structures at chromosome ends that shorten with cell division, triggering visible cellular changes when critically shortened(Blackburn, 1991) (Rossiello et al., 2022). Telomere shortening has been recognized as one of the aging hallmarks (López-Otín et al., 2023). While specialized methods exist to measure telomere length, they require complex molecular techniques, limiting large-scale tissue studies (Rossiello et al., 2022) (Kimura et al., 2010) (Cawthon, 2002) (Norris et al., 2021). Research has shown that telomere shortening triggers cellular senescence with distinct morphological changes including increased cell size, irregular shape, and enhanced granularity. Computational pathology now enables prediction of molecular properties from tissue images. This suggests cellular morphology in routine histology slides could predict telomere length, offering scalable measurement without specialized techniques.

To address this need, we present **TLPath**, a novel computational framework that predicts telomere length directly from standard histopathology (H&E) images by analyzing cellular morphology patterns. Leveraging 7.3 million patch images from the GTEx project across 18 tissue types from 919 individuals, our model employs the UNI foundational model for feature extraction, building on recent advancements in digital pathology foundation

models that enable transfer learning across diverse histopathological tasks. TLPath enables, for the first time, the prediction of bulk-tissue telomere length directly from standard histopathology images, potentially transforming our ability to study telomere biology at scale.

2. Methods

Whole-slide images from the GTEx project (digitized at 20X) were preprocessed via background removal, color normalization, and segmented into 512×512 patches to extract tissue-rich regions. We selected 18 tissue types (each with ≥ 70 samples, totaling 5,263 samples) and utilized the UNI foundational model (Chen et al., 2024) to extract 1,024-dimensional embeddings for each patch. These embeddings were mean-aggregated into whole-slide representations. A Random Forest regression model was then trained to predict relative telomere length using nested cross-validation. Model performance was evaluated via mean squared error (MSE), coefficient of determination (R^2), and Pearson correlation coefficient. To interpret TLPath predictions, we used SHAP values from the Random Forest model to identify key predictive features. High-activation patches were further analyzed using QuPath (Bankhead et al., 2017) to quantify cellular morphology and CONCH (Lu et al., 2024) to map high-dimensional features to histological terms from GTEx pathology reports, elucidating the morphological basis of telomere prediction.

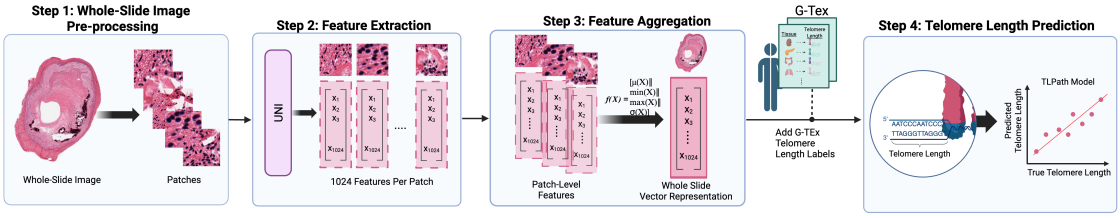

Figure 1: The TLPath pipeline

3. Results

3.1. TLPath Enables Age-Independent Telomere Assessment

Using UNI-extracted H&E features, TLPath achieved an average correlation of 0.32 across tissues, with > 0.4 in 11 of 18 tissue with $r = 0.66$ in Pancreas, a well represented tissue(n = 540). Few tissues show negative correlation likely due to insufficient sample sizes, as prior studies (Ozturk, 2024), (Rossiello et al., 2022) have consistently demonstrated that telomere length shortens with increasing age. Critically, TLPath maintained significant predictive power in age-matched cohorts—accurately distinguishing telomere length differences within 5-year brackets, including atypical cases—demonstrating its ability to capture age-independent morphological signatures. (see Figure 2)

3.2. Interpretation of Morphological Features

Focusing on the pancreas, SHAP analysis identified five key UNI features driving telomere predictions. High-activation patches for these features were analyzed using QuPath to quantify 14 cellular and nuclear properties. Results revealed a higher nucleus-to-cell ratio was linked to shorter telomeres, while increased nuclear circularity and intensity variation were associated with longer telomeres. Additionally, the CONCH model linked these features to higher-order pathological terms (e.g., necrosis, fascia, hypertrophy), underscoring TLPath's computational interpretability . (see Figure 3)

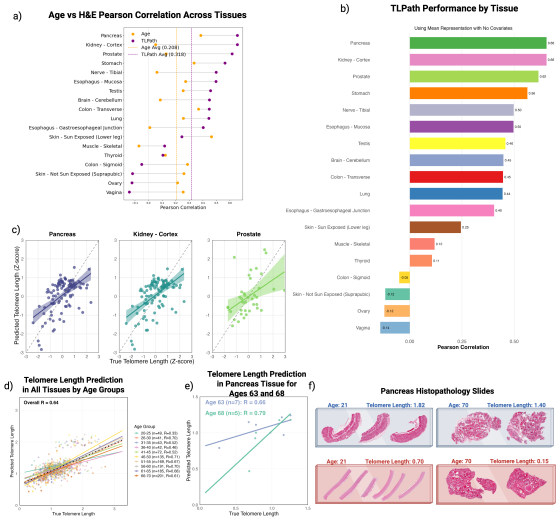

Figure 2 : (a) Comparison of Pearson correlations between chronological age and HE features (b) TLPath's performance in predicting telomere length across tissue with significance (c) The strongest correlations were observed in pancreas (r = 0.66, n = 540), kidney – cortex (r = 0.66, n = 201), and prostate (r = 0.62, n = 186) and the scatter plots show each predicted vs. true telomere length (Z-scored) values for these tissues. (d) Telomere length prediction by age group in all tissues, with number of samples shown. (e) Telomere length prediction in pancreas tissue in age-matched samples. (f) Pancreatic histopathology comparing samples from a young and old individual

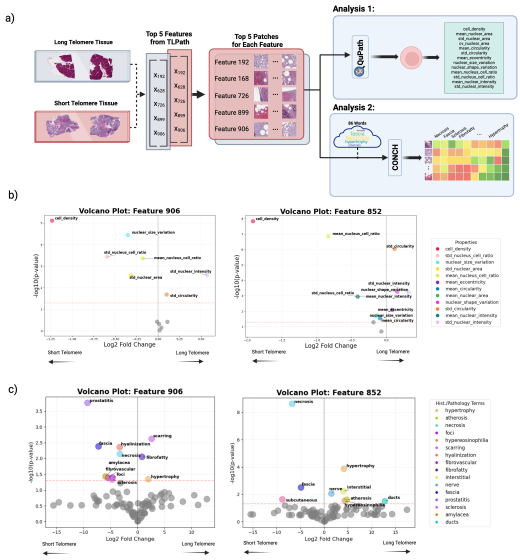

Figure 3: (a) High-activation patches of TLPath's most important features are extracted from long and short telomere samples. This was fed into QuPath (analysis 1) to identify cells and measure 14 distinct cellular and nuclear morphological properties and CONCH model (analysis 2). The average confidence score is calculated for all terms, across all patches. (b) The plot compares differential cellular morphology characteristics in the long versus short telomere patches for features 906 and 852 of UNI. (c) The plots highlight the association of our terms from GTEx pathology reports to features patches 906 and 825, with long and short telomere from WSI.

3.3. Benchmarking and Architectural Optimization

We systematically benchmarked TLPath's components across models (TITAN, UNI), aggregation methods (mean, min-max, concatenated), and regression approaches. Mean aggregation produced robust representations, while Random Forest yielded consistent predictions across 18 tissues. Although Elastic Net had slightly higher correlations in some common tissues, TLPath demonstrated generalizable performance across the board (Table 1).

WSI Aggregation	TLPath	ElasticNet	SVR	Ridge	Lasso	LR
Concatenated	0.520	0.526	0.495	0.478	0.494	0.450
Mean	0.521	0.526	0.495	0.478	0.494	0.449
Min-Max	0.508	0.526	0.494	0.477	0.494	0.444
TITAN	0.520	0.526	0.495	0.478	0.494	0.172

Table 1: Mean test Pearson correlation across selected tissues (Brain - Cerebellum, Colon - Transverse, Esophagus - Mucosa, Kidney - Cortex, Lung, Pancreas, Stomach, Testis) for different model types and WSI representation strategies.

References

P. Bankhead, M. B. Loughrey, J. A. Fernández, et al. Qupath: Open source software for digital pathology image analysis. *Scientific Reports*, 7:16878, 2017. doi: 10.1038/s41598-017-17204-5.

E. H. Blackburn. Structure and function of telomeres. *Nature*, 350(6319):569–573, 1991.

R. M. Cawthon. Telomere measurement by quantitative pcr. *Nucleic Acids Research*, 30 (10):e47–e47, 2002.

R. J. Chen, T. Ding, M. Y. Lu, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30:850–862, 2024. doi: 10.1038/s41591-024-02857-3.

Masayuki Kimura et al. Measurement of telomere length by the southern blot analysis of terminal restriction fragment lengths. *Nature Protocols*, 5(9):1596–1607, 2010.

M. Y. Lu, B. Chen, D. F. K. Williamson, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30:863–874, 2024. doi: 10.1038/s41591-024-02856-4.

C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer. Hallmarks of aging: An expanding universe. *Cell*, 186(2):243–278, January 2023. doi: 10.1016/j.cell.2022.11.001. Epub 2023 Jan 3.

K. Norris, A. J. Walne, M. J. Ponsford, K. Cleal, J. W. Grimstead, A. Ellison, and D. M. Baird. High-throughput stela provides a rapid test for the diagnosis of telomere biology disorders. *Human Genetics*, 140:945–955, 2021.

S. Ozturk. The close relationship between oocyte aging and telomere shortening, and possible interventions for telomere protection. *Mechanisms of Ageing and Development*, 218:111913, April 2024. doi: 10.1016/j.mad.2024.111913. Epub 2024 Feb 1.

F. Rossiello, D. Jurk, J. F. Passos, and F. d'Adda di Fagagna. Telomere dysfunction in ageing and age-related diseases. *Nature Cell Biology*, 24(2):135–147, 2022.