**Path2Omics: An AI model integrating FFPE and FF slides to enhance transcriptomic and methylation prediction accuracy from tumor histopathology**

**1. Introduction**

Path2Omics is a deep learning framework that independently predicts gene expression and methylation from histopathology slides across 30 cancer types. Unlike existing approaches that rely solely on Formalin-Fixed and Paraffin-Embedded (FFPE) slides for training, Path2Omics leverages both FFPE and fresh frozen (FF) slides by constructing two separate models and integrating them. Downstream analyses show that the inferred values from Path2Omics are nearly as effective as actual values in predicting patient survival and treatment response.

Path2Omics comprises three main components:

**(i) Image pre-processing: Each** whole slide image is divided into tiles, and Sobel edge detection is applied to select only tiles containing tissue. To minimize staining variation, Macenko’s method is used for color normalization.

**(ii) Feature extraction:** A pre-trained pathology foundation model is utilized to extract image features from the selected tiles.

**(iii) Prediction: A multi-layer perceptron regression model is employed to** predict gene expression (or DNA methylation) from the extracted features.

**2. Installations**

To install Path2Omics, please install the following requirements:

python 3.9.7

numpy 1.20.3

pandas 1.3.4

matplotlib 3.4.3

sklearn 1.2.2

openslide 1.1.2

opencv 4.5.4

torch 1.12.1

**3. Path2Omics computational pipeline**

**- Step 1:** Run “11features/1main\_processing.py” to perform image pre-processing and feature extraction. This code will run on each slide simultaneously.

**- Step 2:** Run “11features/2collect\_mask.py” to collect mask files into a single file “mask.pdf” that will be used to evaluate slide quality and subsequently run “11features/3collect\_features.py” to collect mask files into a single feature file.

**- Step 3:** Run “13train/1main\_train.py” to train and predict gene expression (or methylation) from the extracted features.

**4. License and Terms of use**

This model is permitted solely for academic research purposes. Commercial entities interested in utilizing the model should contact the corresponding authors for authorization.

**5. Reference**

If you find our work useful in your research or if you use parts of this code, please consider citing our paper:

Danh-Tai Hoang et al., *Path2Omics enhances transcriptomic and methylation prediction accuracy from tumor histopathology*, Cancer Research (2025).