TP53 study overview

Data acquisition

Genomic Data Commons (GDC)

Whole Slide Images (WSIs) across 32 cancer types

Labelled samples via TCGA barcode matching

The Cancer Genome Atlas (TCGA) Pan-Cancer (PANCAN)

- Cancer types
- Mutation status
- TP53 gene expression level
- Survivial outcomes

Model development and validation

Model development

- 6,055 WSIs

TCGA-UT dataset

- Six resolution (0.5-1.0 µm/pixel) patches per slide

Feature encoding

- DINOv2 (768-dim)

Multi-instance aggregation

- 6-layer encoder, 12 heads per layer

Multi-task heads

- Cancer type (32)
- TP53 mutation status
- TP53 gene expression
- Overall survival event and time
- Progression free interval event and time

Trained model

WSI pre-processing

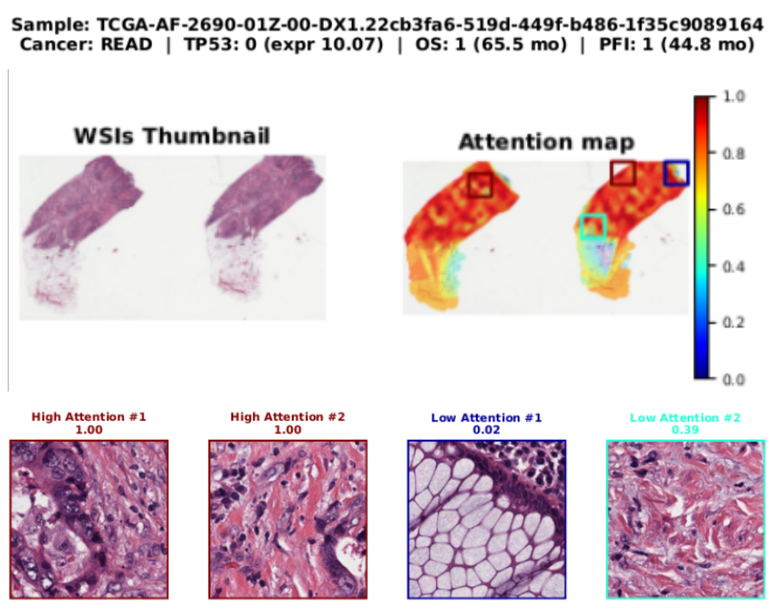
- Tissue mask
- Patch extraction
- Quality control
- Stain normalisation

- Cancerous and non-cancerous patches at a single resolution, w/ 6x downsampled patch

Fine-tuned model on WSIs

External validation

- 1,729 WSIs

Result

Prediction with attention maps at slide level