

Do Multiple Instance Learning Models Transfer?

Daniel Shao, Richard J. Chen, Andrew H. Song, Joel Runevic, Ming Y. Lu, Tong Ding, Faisal Mahmood
dshao@mit.edu, faisal.mahmood@bwh.harvard.edu

Motivation

Computational Pathology

- Slide-level analysis of **digitized tissue sections** (whole-slide images, WSIs), of up to **100,000 x 100,000** pixels
- Multiple Instance Learning** (MIL) (1) divides WSI into a set of patches encoded with a pretrained vision encoder and (2) aggregates patch embeddings into a slide embedding

Learning to aggregate is challenging

- MIL aims to identify relevant patches from ~10,000 patches per slide
- Poor generalization is common due to data scarcity

Why not leverage transfer learning?

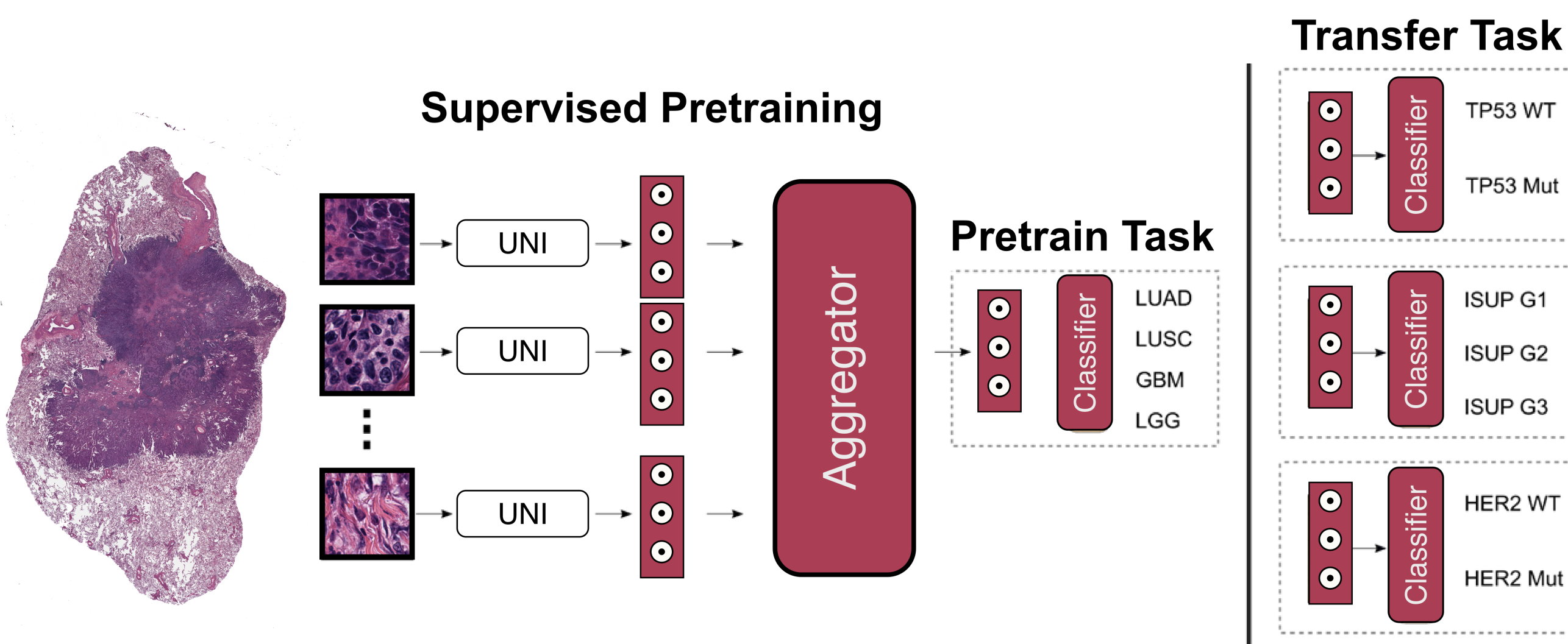
MIL Transfer Setup

Slide-level transfer

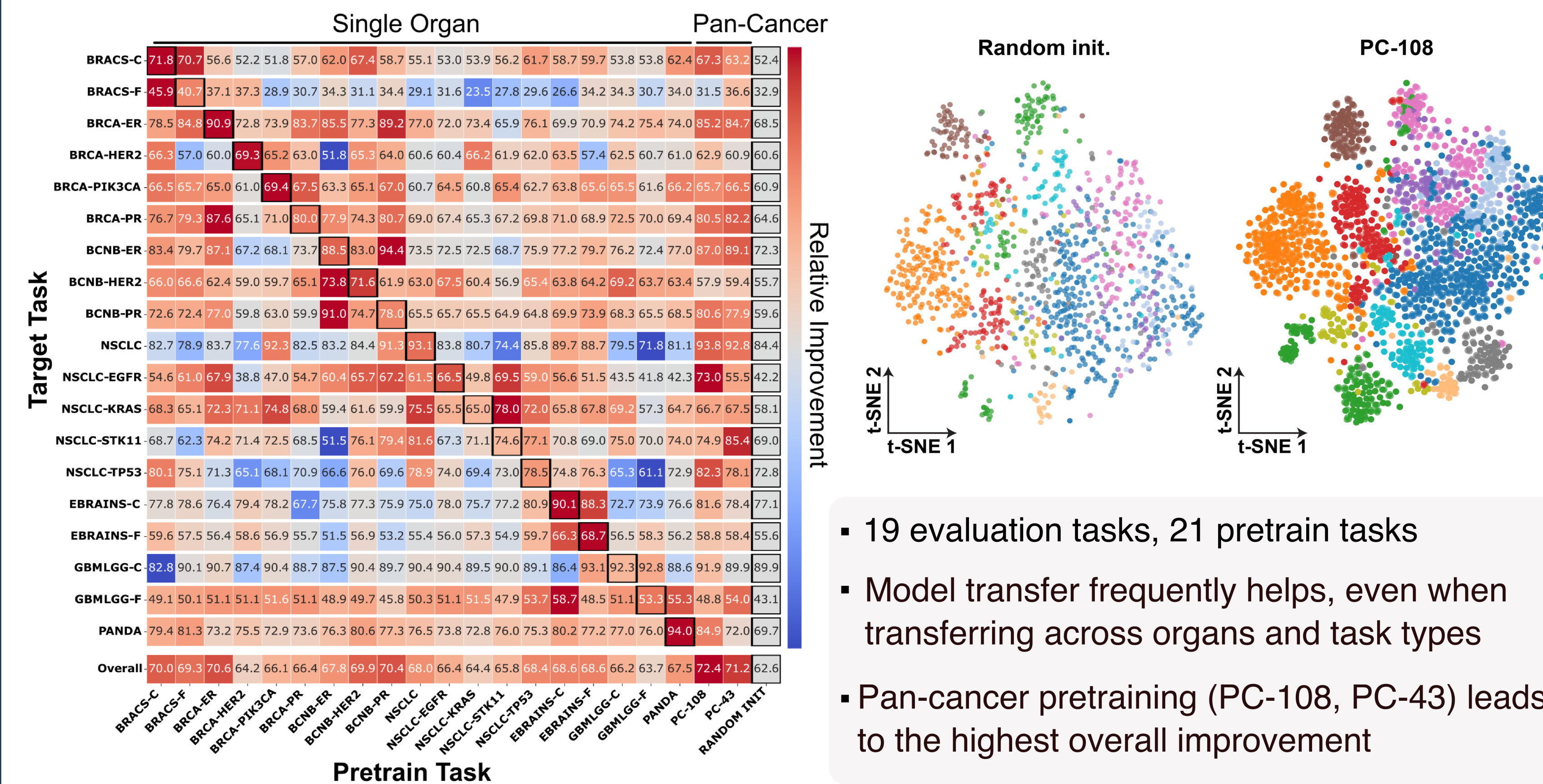
- Self-supervised* foundation models (CHIEF, GigaPath, Titan, THREADS)
- Data expensive (n=60,000 - 300,000 WSIs), difficult to finetune

Supervised pretraining for MIL models

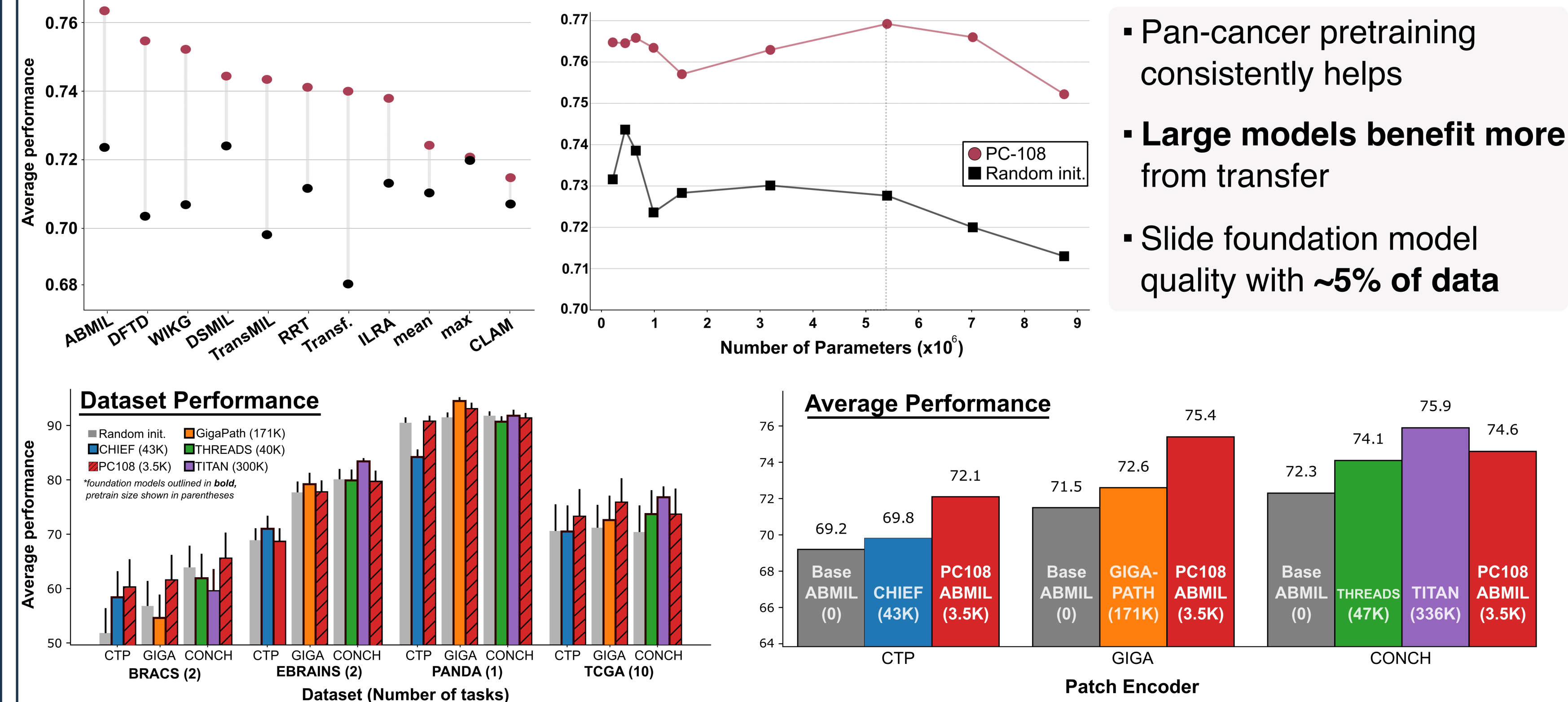
- Can we leverage higher-quality labels to train data-efficient models?*
- Motifs across tasks:** Discard background, identify tumor regions
- Flexible alternative for training lightweight slide foundation models



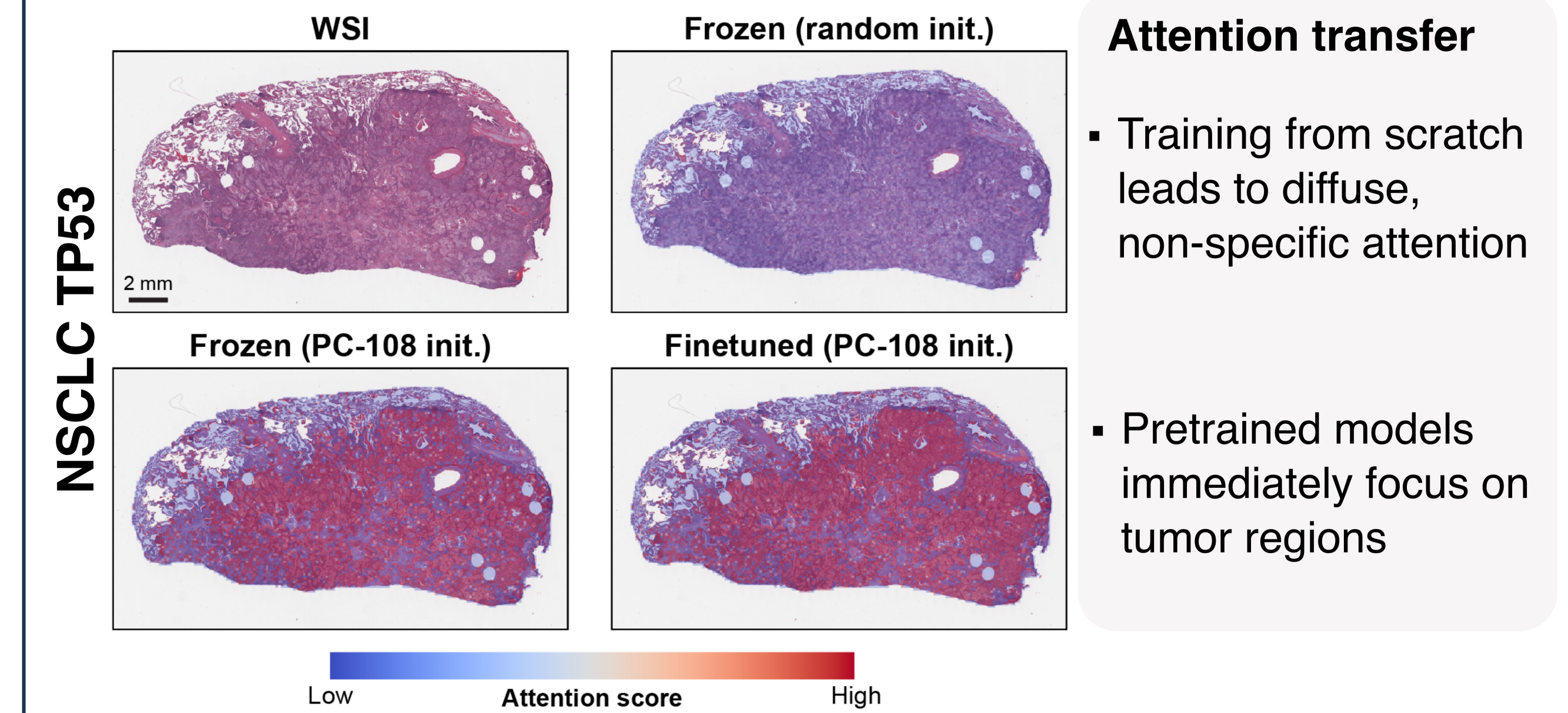
Frozen MIL Transfer



Finetuned Pan-cancer Transfer



What is being transferred?



How transferrable is each layer?

