BARCELONA 2024 ESMO congress

1942O

Application of GigaPath: An open-weight billion-parameter AI foundation model based on a novel vision transformer architecture for cancer mutation prediction and TME analysis

C. Bifulco[1], H. Poon[2], N. Usuyama[3], H. Xu[4], S. Wang[4], B. Piening[5], R. Leidner[6]

[1] Pathology, Providence Cancer Institute, Southfield, MI, USA, [2] Research Division, Microsoft Research, Redmond, WA, USA, [3] Research Division, Microsoft, Redmond, WA, USA, [4] Computer Sciences, University of Washington, Seattle, WA, USA, [5] Providence Genomics, Earle A. Chiles Research Institute at the Robert W. Franz Cancer Center, Portland, OR, USA[6] Earle A Chiles Research Institute, Providence Cancer Institute Franz Clinic, Portland, OR, USA

Background

Computational pathology has the potential to transform cancer diagnostics by empowering diverse clinical applications. GigaPath is an open-weight billion-parameter AI foundation model pretrained on a large digital pathology dataset from 28 cancer centers (Providence), containing 1,384,860,229 image tiles from 171,189 H&E slides of biopsies and resections in more than 30,000 patients, covering 31 major tissue types. Predicting tumor mutations from pathology images may help increase personalized medicine utilization.

Methods

GigaPath excels in long-context modelling of gigapixel pathology slides, by distilling varied local pathological structures and integrating global signatures across the whole slide. We compared GigaPath H&E molecular prediction with competing methods HIPT, CtransPath, REMEDIS, across three tasks: lung adeno 5-gene (*EGFR*, *FAT1*, *KRAS*, *TP53, LRP1B*), pan-cancer 5-gene, and TMB prediction.

Results

For lung adenocarcinoma, GigaPath achieved an average macro area under the receiver operator characteristic (AUROC) of 0.626 surpassing all competing approaches (*P* value < 0.01). For pan-cancer, GigaPath also outperformed the best competing methods on these 5 genes with 6.5% macro-AUROC improvement and 18.7% improvement in macro-area under the precision-recall curve (AUPRC, *P* value < 0.01). For TMB prediction GigaPath achieved the best performance with an average AUPRC of 0.35, with significant improvement (P value < 0.01) over the second-best method.

Conclusions

GigaPath can potentially be applicable to broader biomedical domains for efficient self-supervised learning from high-resolution images, including applications leveraging the long-context modelling features of the model to deconvolute emerging spatial biology datasets to drive a personalized and comprehensive characterization the tumor microenvironment (to be presented). To accelerate research progress in digital pathology, we made GigaPath fully open-weight, including source code and pretrained model weights. Ethics Approval: PSJH IRB# 2018000188.

Legal entity responsible for the study

The authors.

Funding

Providence, Microsoft.

Disclosure

C. Bifulco: Financial Interests, Personal, Stocks/Shares, Scientific Advisory Board: PrimeVax, BioAI; Financial Interests, Personal, Advisory Board: SeronaDx, Lunaphore, Sanofi, Roche, Agilent, Incendia; Financial Interests, Personal and Institutional, Research Funding: Illumina. B. Piening: Financial Interests, Personal and Institutional, Research Funding: Illumina. R. Leidner: Financial Interests, Personal, Advisory Board: Vir, RAPT, Merck, CDR-Life, AstraZeneca, Incyte, BMS; Financial Interests, Personal and Institutional, Research Grant: BMS; Financial Interests, Institutional, Research Grant: AstraZeneca. All other authors have declared no conflicts of interest.