

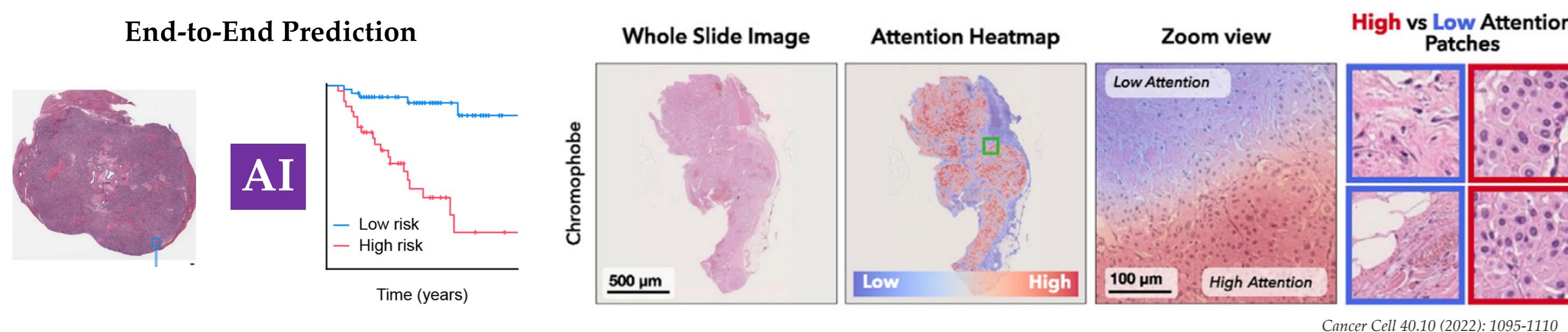
# AI Inspired Discovery of New Pathological Biomarkers for Cancer Prognosis

Junhao Liang<sup>1</sup> @Junhao\_Liang, Weisheng Zhang<sup>1</sup>, Jianghui Yang<sup>2</sup>, Meilong Wu<sup>3</sup>, Qionghai Dai<sup>4\*</sup>, Hongfang Yin<sup>2\*</sup>, Ying Xiao<sup>2\*</sup>, and Lingjie Kong<sup>1,5</sup>



## Background

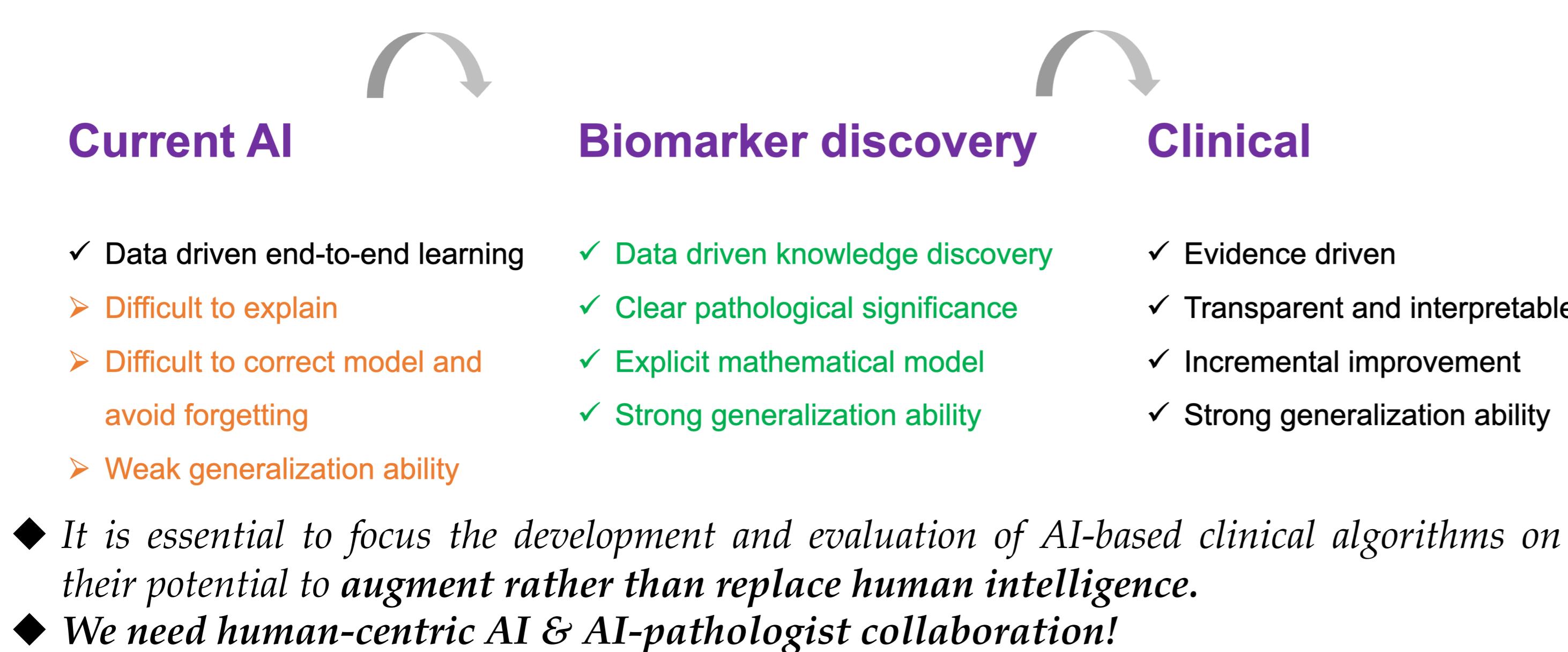
- Current deep learning methods can achieve accurate survival prediction based on HE biopsy whole slide images, showing great potential of learning knowledge from vast medical data.



- However, due to limited interpretability of deep learning and rich information of pathological images, current attribution methods can only tell us where the model focuses on, but not exactly what it is, letting alone characterization and verification of potential pathological biomarkers.
- In such situation, many works only use attribution methods to prove the trained high-performance deep neural network is reasonable by linking high attention area to already known prognosis-related pathology morphologies, fail in finding new understandable pathological biomarkers.

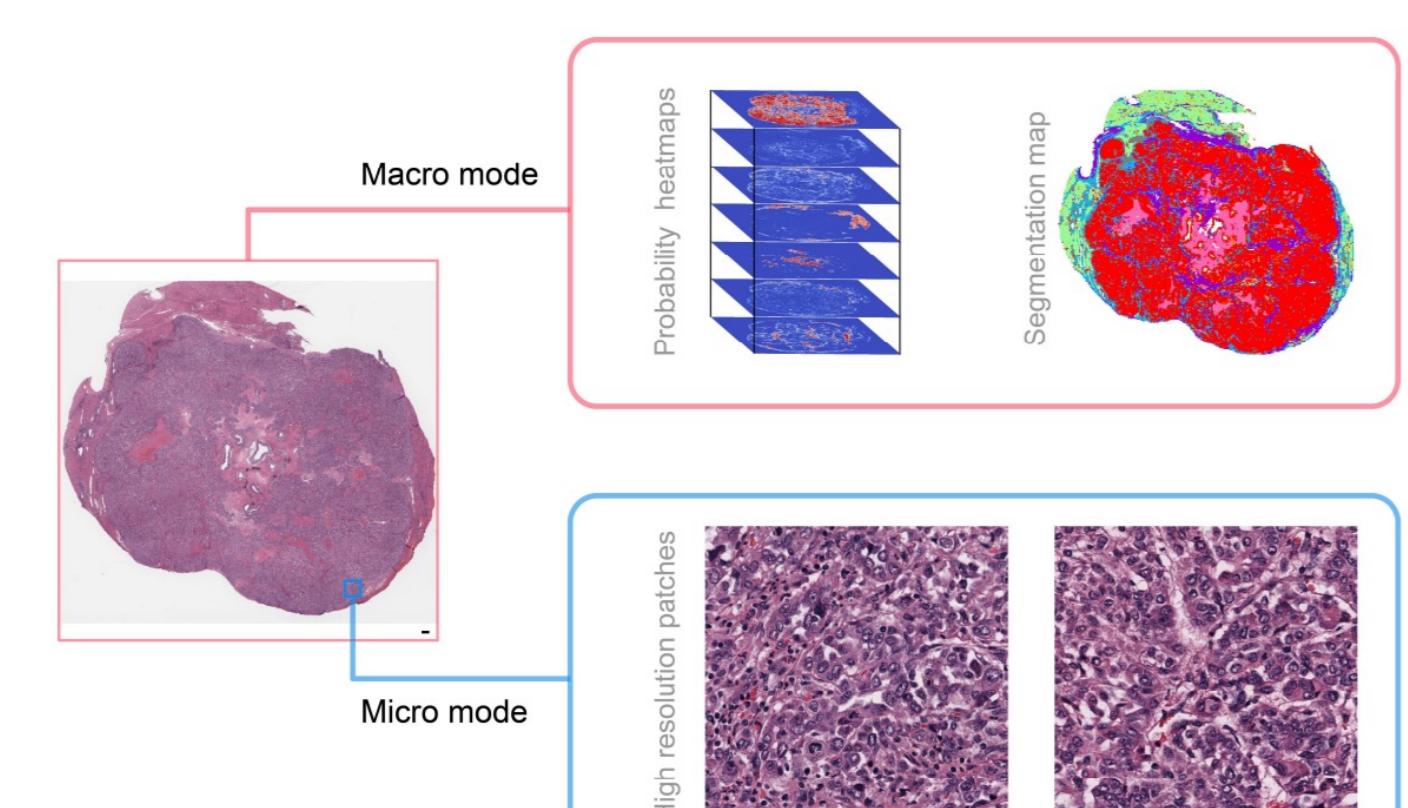
## Motivation

- Finding new pathological biomarkers is not only urgently needed to achieve accurate clinical prognosis, but also necessary to understand the mechanisms of cancer evolution, design new treatment options and drugs, and expand the boundaries of cancer knowledge.
- The potential of end-to-end deep learning in pathological prognosis has been hampered by limited interpretability in clinical applications.
- Using AI as a knowledge discovery tool to find unknown or dominant pathological biomarkers might bridge the gap between AI and clinical practice.



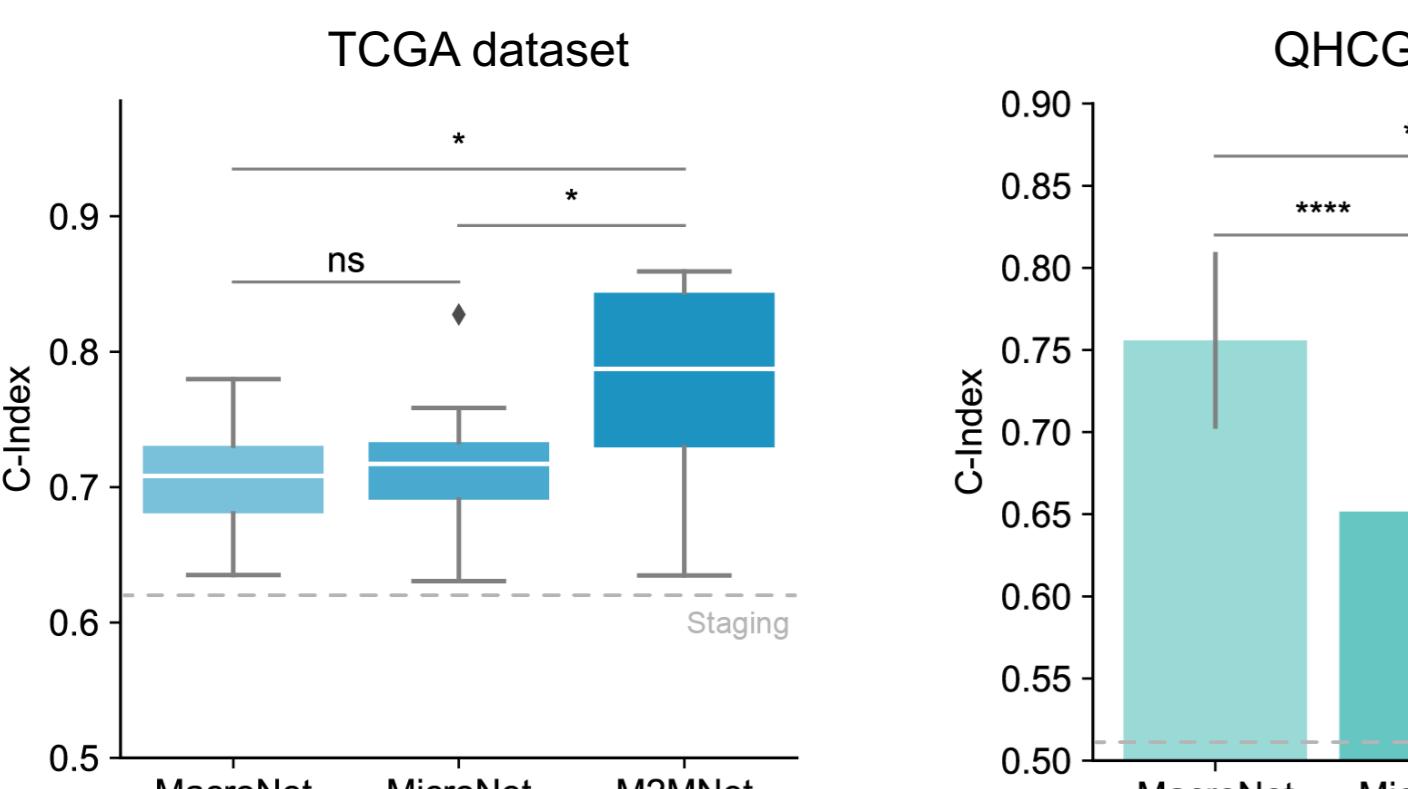
## Pathological Biomarker Finder

### 1. WSI decoupling with pathological prior knowledge



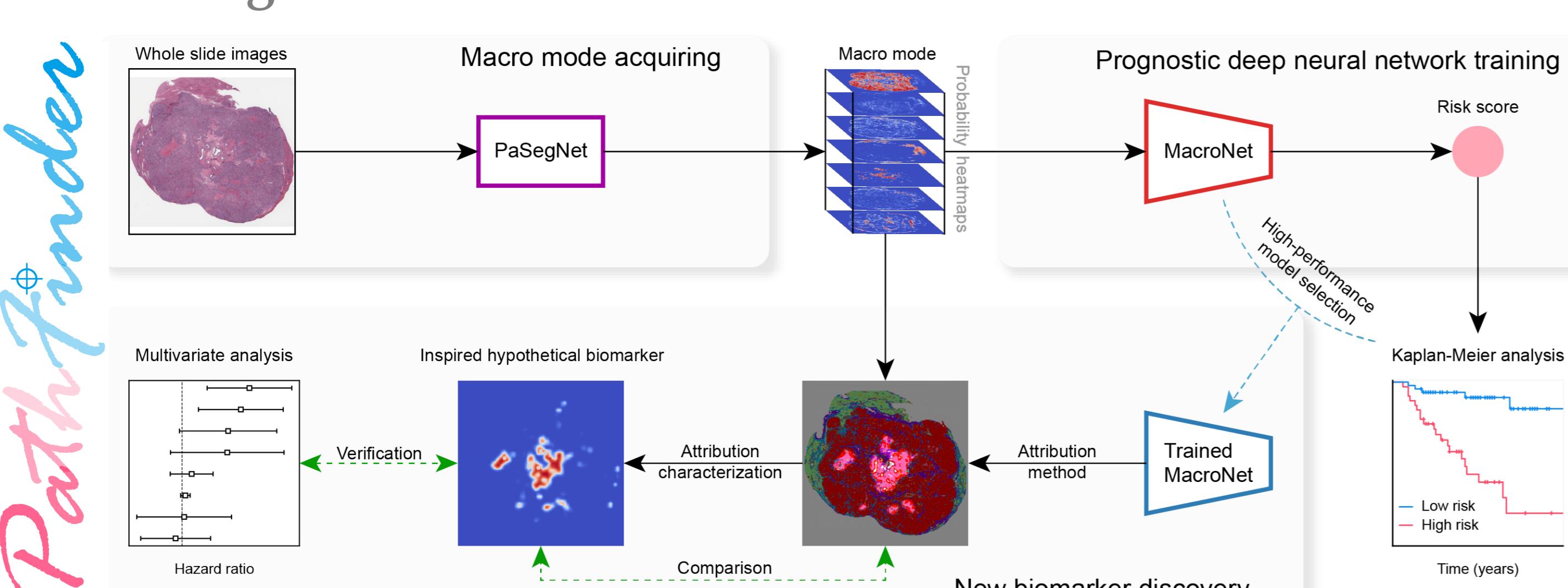
- To overcome the problem of the high information density of WSIs and make prognostic DL model more suitable for current attribution methods, we decoupled the input WSI into macro mode and micro mode.
- The former contains multi-class tissue spatial distribution and interaction information, while the latter contains cell texture and structure information.

### 2. Macro mode can achieve SOTA prognostic performance



- We trained different modes in a ten-fold cross-validation on TCGA dataset, and tested on the QHCG dataset.
- Macro mode get SOTA performance as micro mode does, and has better generalizability on external dataset.

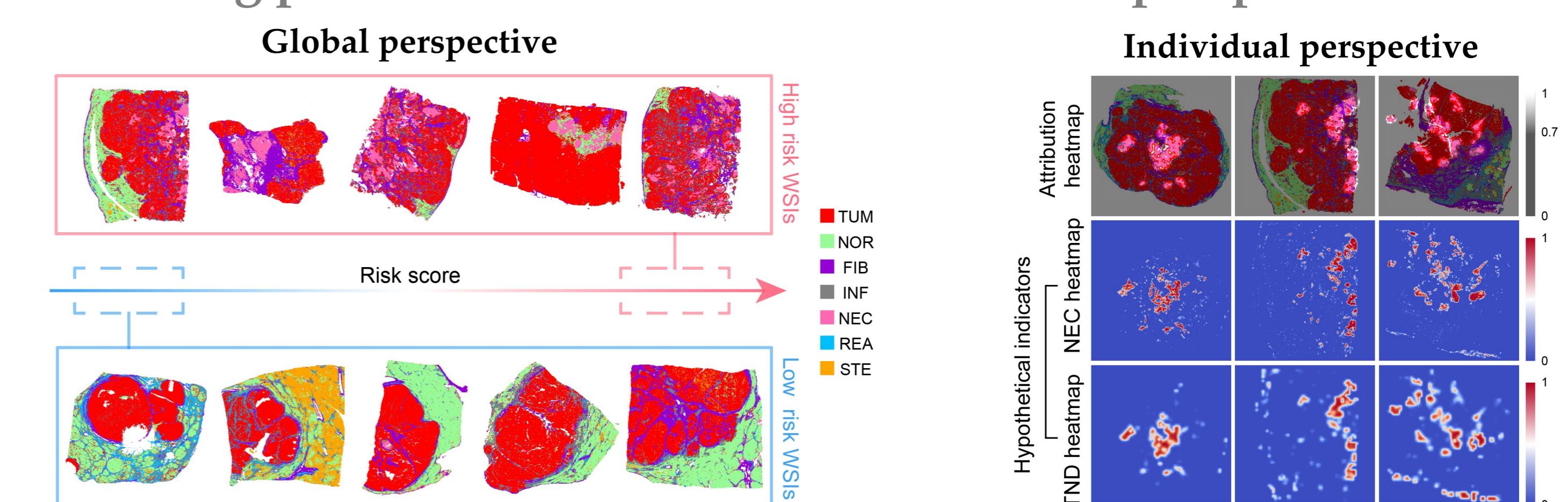
### 3. Pathological biomarker finder ---- PathFinder



Discovering prognostic biomarkers by AI-pathologist collaboration

## Discovery & Verification of New Biomarkers

### 1. Finding potential biomarkers from different perspectives



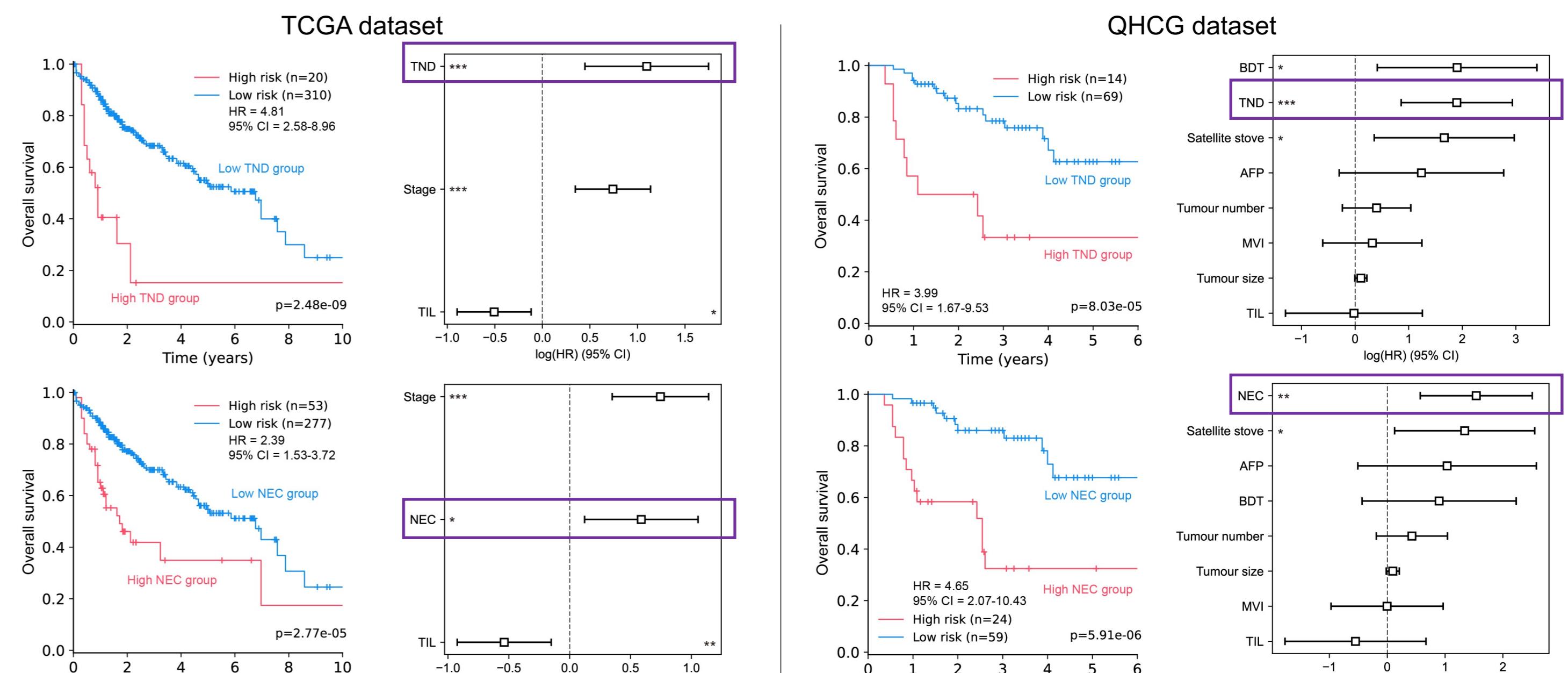
- Pink area spatial distribution (Necrosis) may closely related to patient prognosis.

### 2. Building mathematical models of AI inspired biomarkers

$$TND = \begin{cases} \frac{M'}{2} \times \frac{\sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{NEC})}{\sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{TUM})}, \sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{TUM}) > 0 & M' = \frac{2 \sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{NEC} \times p_{ij}^{TUM})}{\sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{NEC}) + \sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{TUM})} \\ 1, \sum_{i=1}^m \sum_{j=1}^n (p_{ij}^{TUM}) \leq 0 & NEC = Fraction_{NEC} = \frac{N_{NEC}}{N - N_{empty}} \end{cases}$$

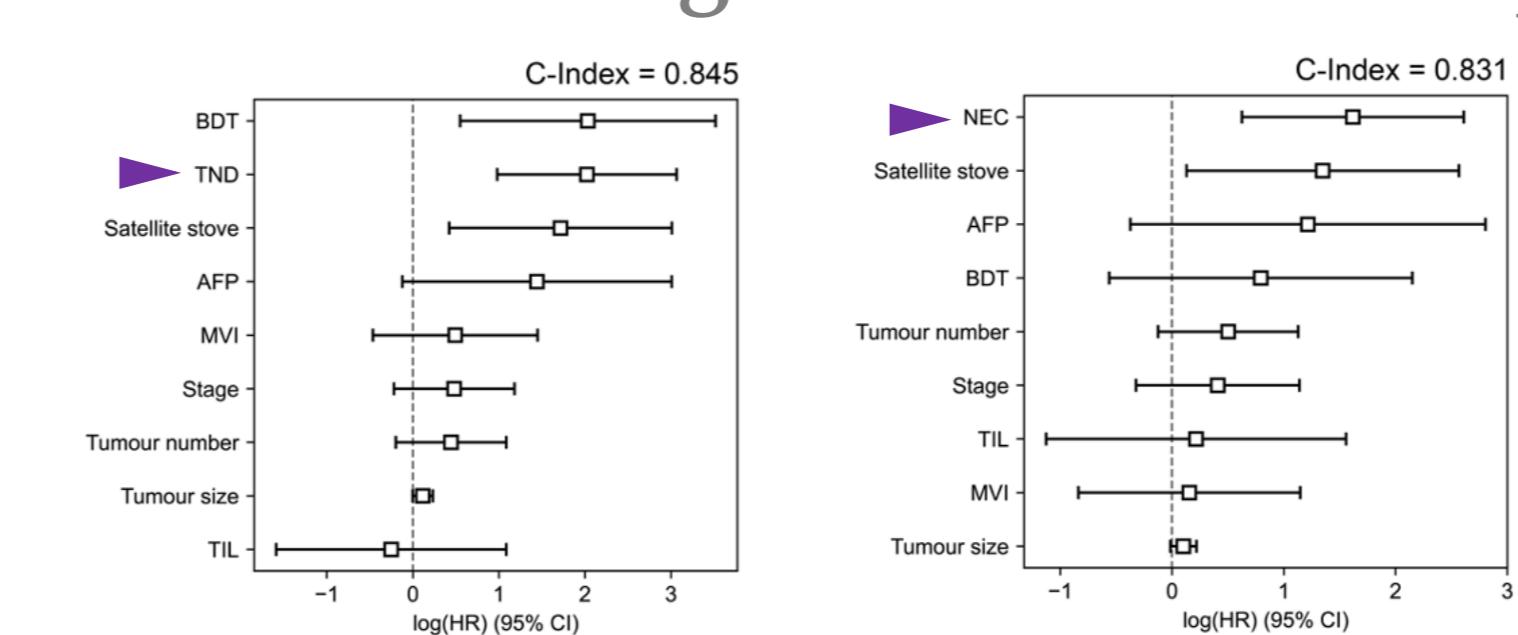
- We proposed two hypotheses of new biomarkers, namely necrosis area fraction in WSIs (NEC) and tumour necrosis distribution (TND) based on digitalized macro mode.

### 3. Verification of hypothetical biomarkers



- The univariate and multivariable analyses revealed that the dependences of overall survival on NEC and TND were more significant than most clinical indicators including TILs.

### 4. Constructing multimodal transparent prognostic models

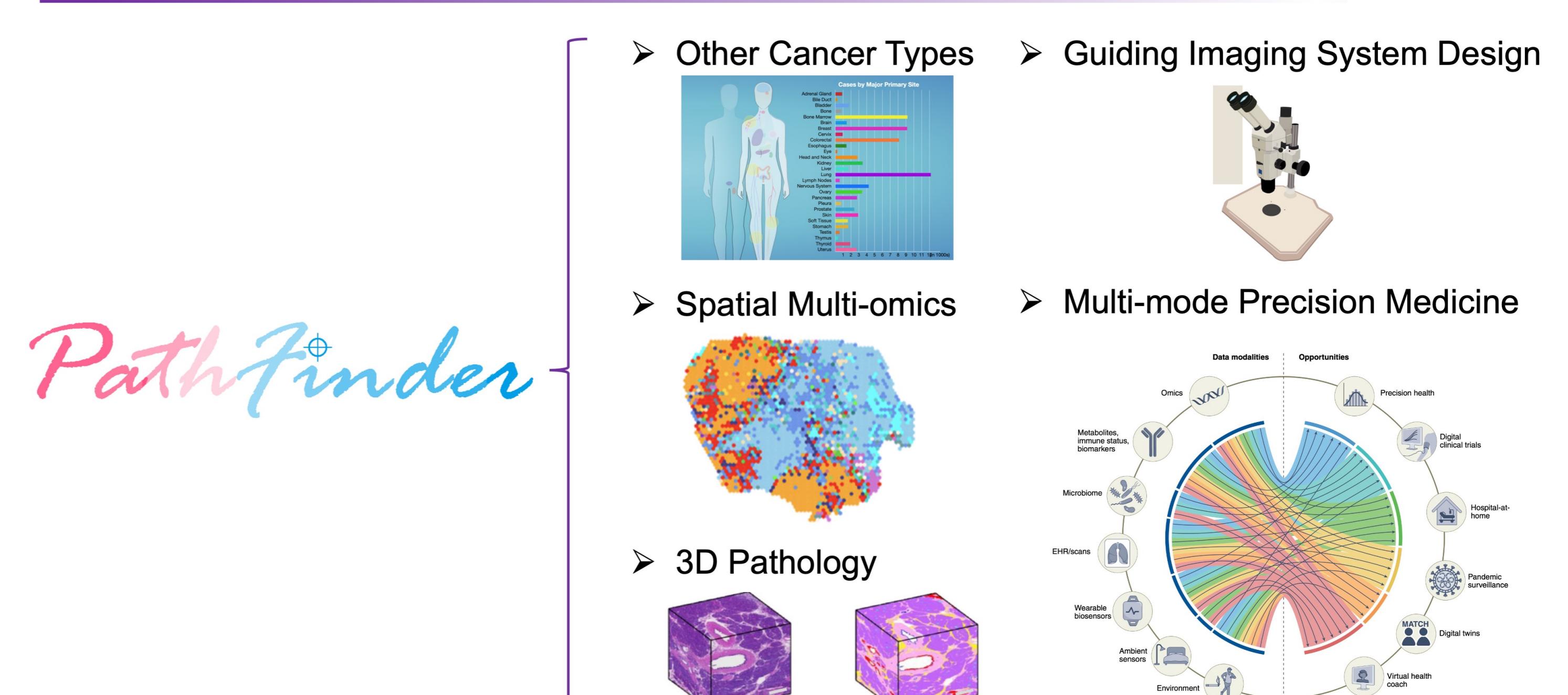


- Taking new discovered biomarkers and other clinical factors together into multimodal COX regression model, the C-Indices of NEC and TND can be further improved to 0.831 and 0.845.

## Conclusion

- We demonstrate a method of bridging the interpretation gap between AI and clinical prognosis, and prove the potential of AI in learning and exploring new prognostic biomarkers based on large datasets and objective survival information.
- We find the sparse spatial distribution can make accurate prognosis. The prognostic performance is still good even when the input is reduced from WSIs of several gigabytes to macro mode of several megabytes.
- Inspired by PathFinder, we defined two necrosis-related clinical prognostic indicators, NEC and TND, and demonstrated their feasibility in HCC prognosis.

## Potential



<sup>1</sup>State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China.

<sup>2</sup>Department of Pathology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing 102218, China.

<sup>3</sup>School of Clinical Medicine, Tsinghua University, Beijing 100084, China.

<sup>4</sup>Department of Automation, Tsinghua University, Beijing 102218, China.

<sup>5</sup>IDG/McGovern Institute for Brain Research, Tsinghua University, Beijing 100084, China.

Published in



Highlighted in

