Preserving DEG Rankings for Gene Discovery in Histology-Based Spatial Gene Expression Prediction

◯ Kaito Shiku[1], Kazuya Nishimura[2], Yasuhiro Kojima[3], Ryoma Bise[1]

[1] Kyushu University, [2] The University of Osaka, [3] National Cancer Center

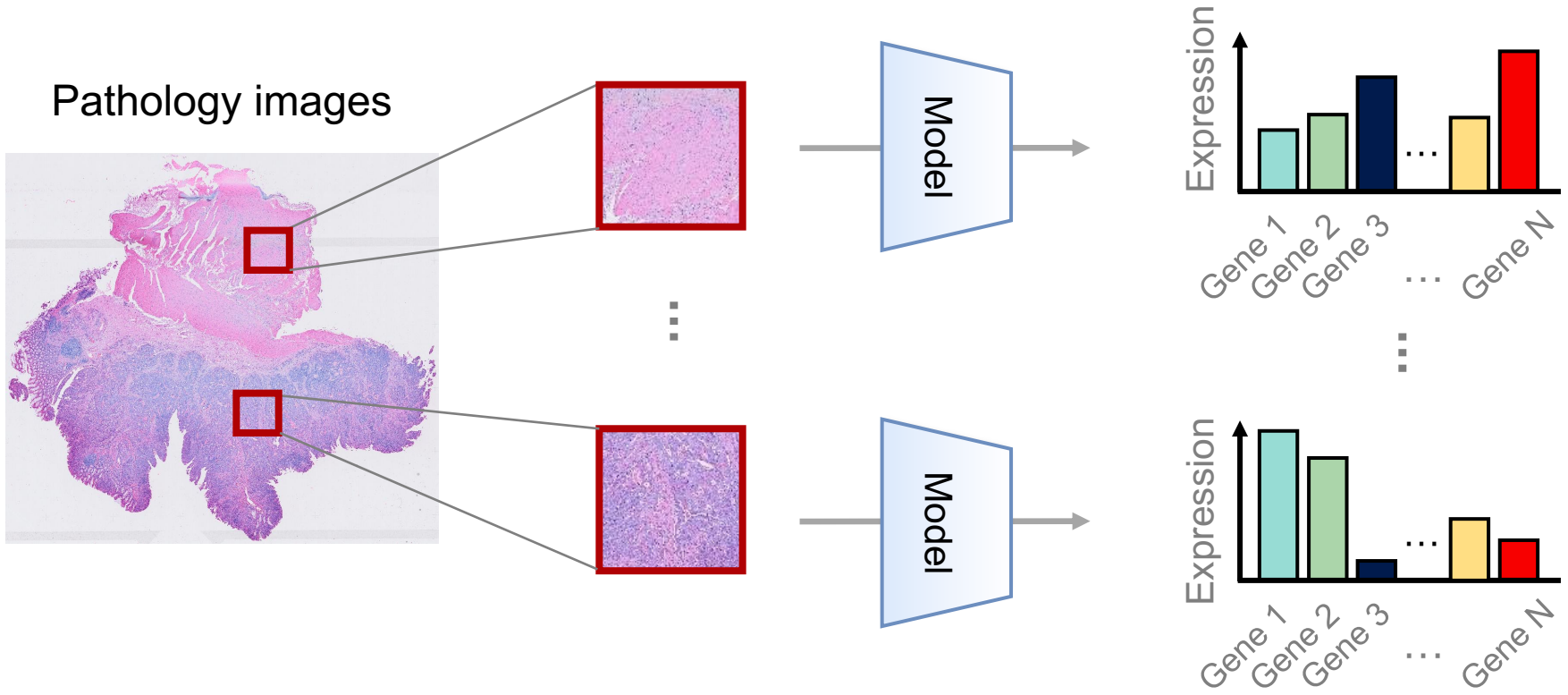
Paper

GitHub

NEURAL INFORMATION PROCESSING SYSTEMS

Pathology image-based spatial gene expression prediction

- Regression task: predict gene expression for each region of pathology images
- Reduce measurement costs



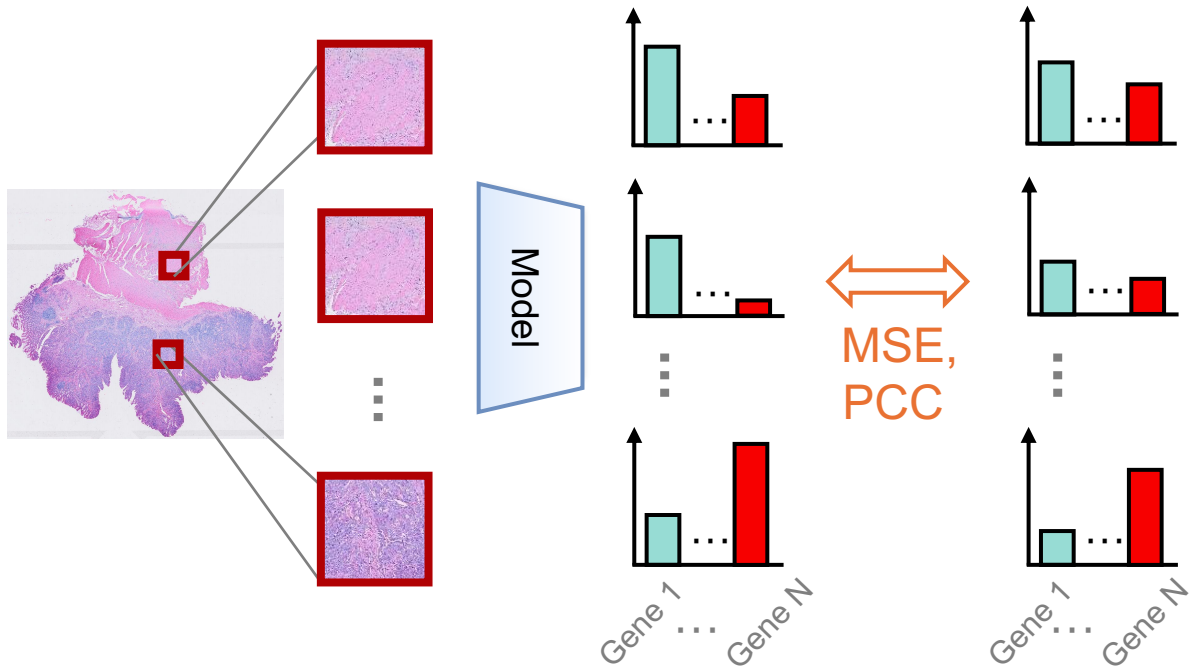
Challenge:

Gap between machine learning objectives and biological analysis

Machine learning

Improve prediction accuracy **for each gene**

(e.g. MSE: Mean Squared Error; PCC: Pearson Correlation)



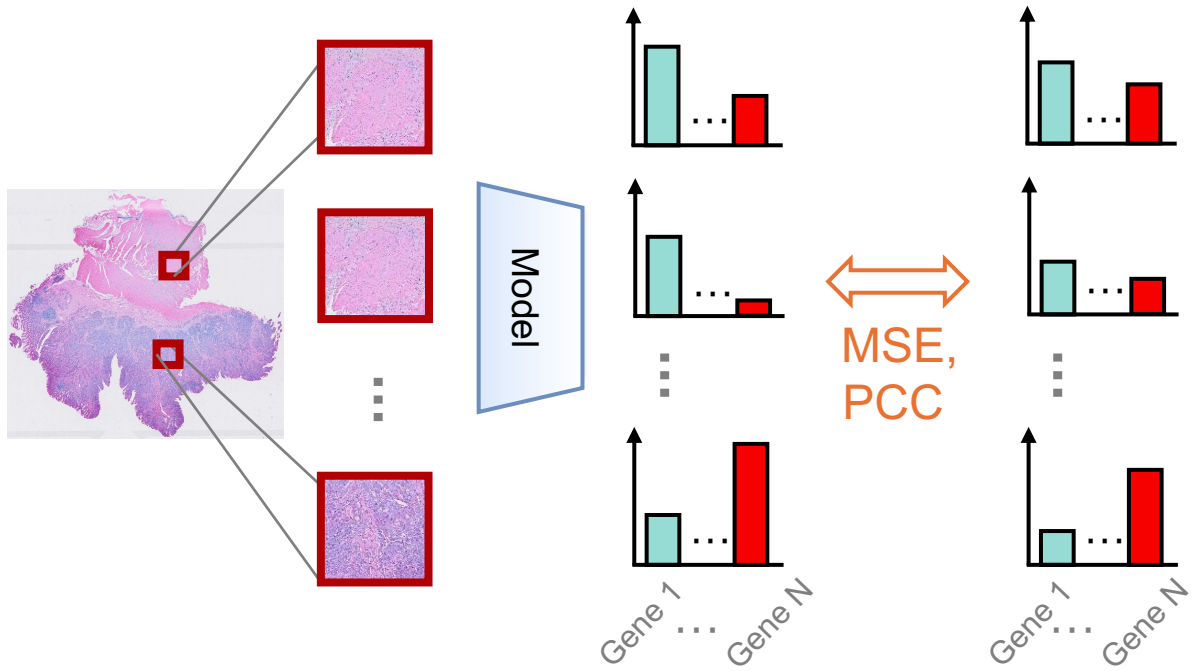
Challenge:

Gap between machine learning objectives and biological analysis

Machine learning

Improve prediction accuracy **for each gene**

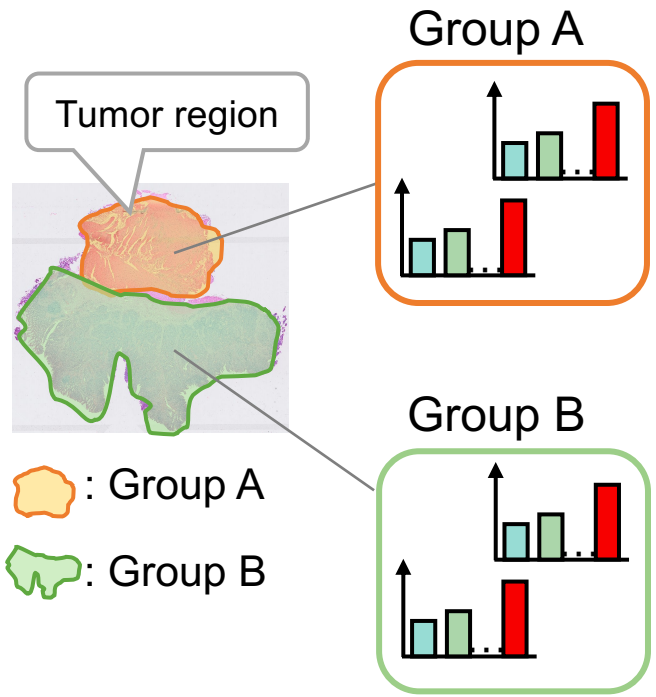
(e.g. MSE: Mean Squared Error; PCC: Pearson Correlation)



Biological discovery

Rank genes to identify specific maker

(Marker gene analysis[1])



[1] Robinson, et al., Bioinformatics, 2010

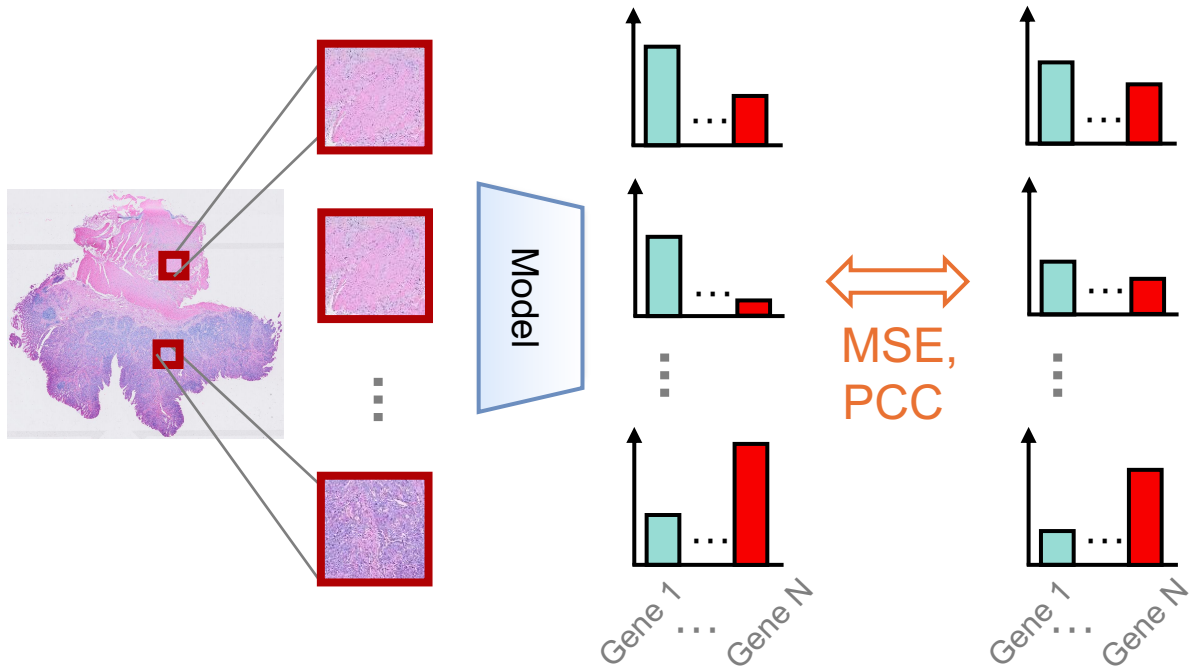
Challenge:

Gap between machine learning objectives and biological analysis

Machine learning

Improve prediction accuracy **for each gene**

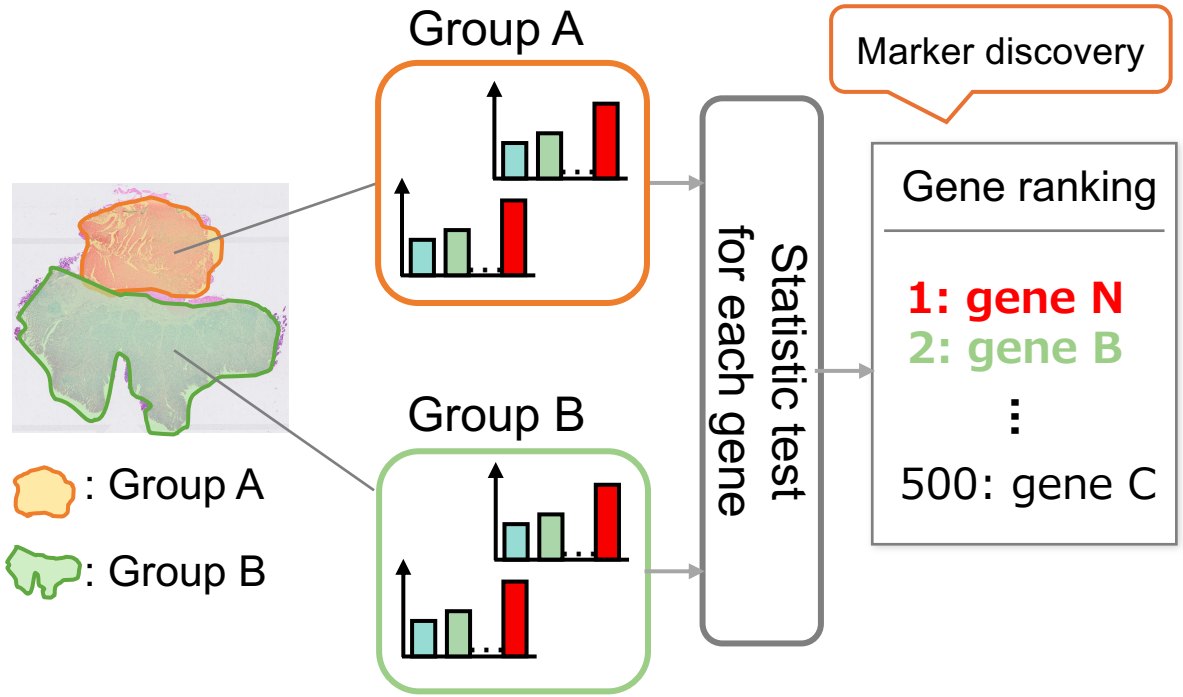
(e.g. MSE: Mean Squared Error; PCC: Pearson Correlation)



Biological discovery

Rank genes to identify specific maker

(Marker gene analysis[1])



[1] Robinson, et al., Bioinformatics, 2010

Challenge:

Gap between machine learning objectives and biological analysis

Previous research:

Ignore rankings across genes
Not aligned with biological needs

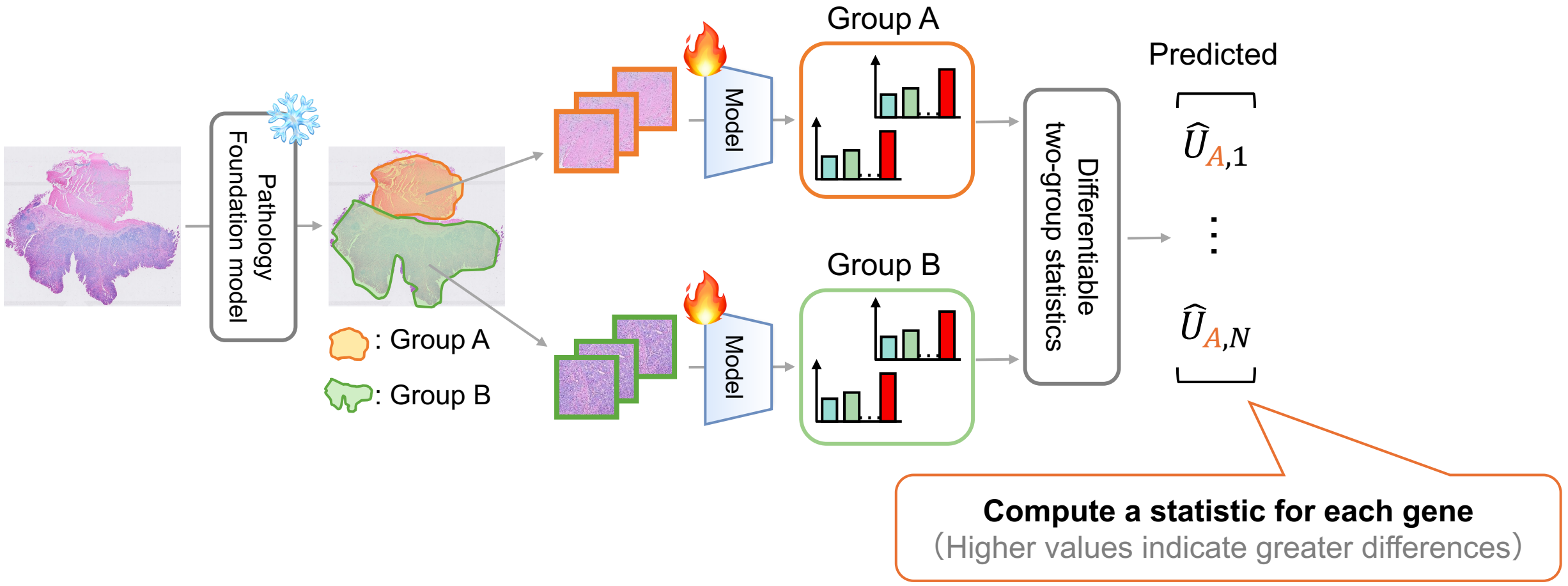
Our proposal:

Learning objective for gene ranking
Evaluation metric for gene ranking agreement

[1] Robinson, et al., Bioinformatics, 2010

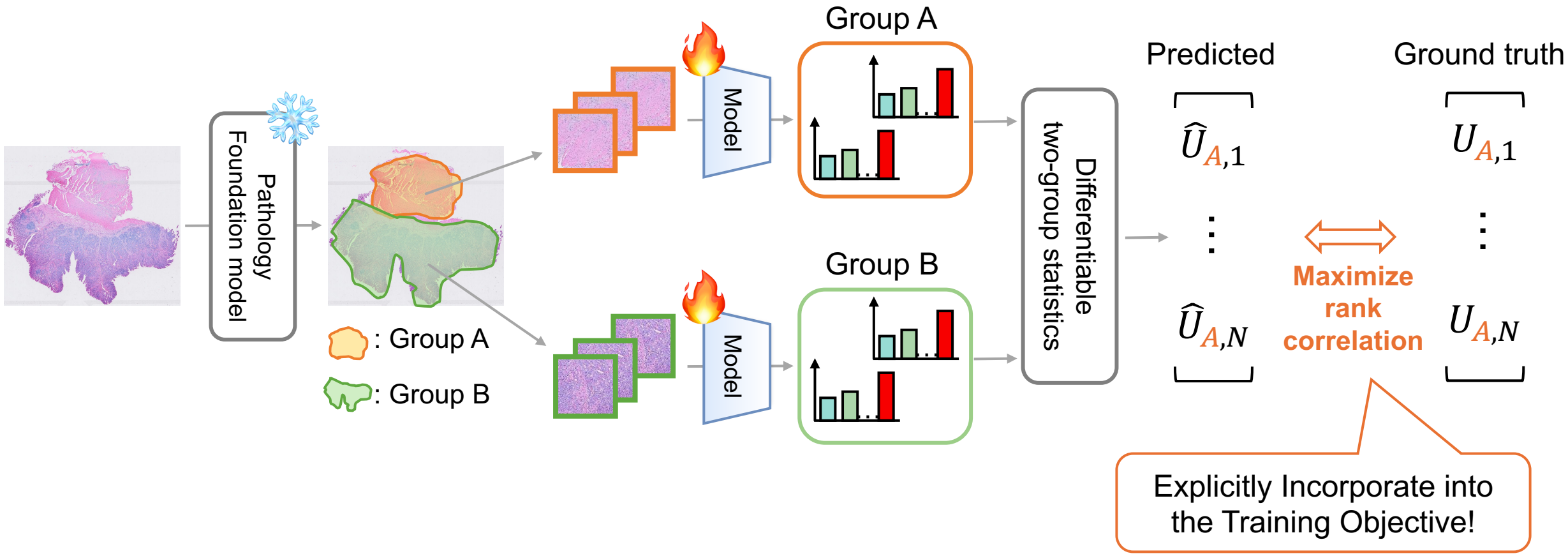
Proposed method:

Image grouping-based gene ranking optimization



Proposed method:

Image grouping-based gene ranking optimization



Performance evaluation on real-world dataset

Dataset (Her2ST): Use pathologist annotations to rank region-specific marker genes.

Proposed evaluation metric

- DEG SCC: Correlation between gene rankings
- DEG nDCG@k: Agreement among the top-k ranked genes

Comparison

Objective	DEG SCC	nDCG@200
MSE	0.239	0.084
Pearson correlation	0.181	0.073
Ours	**0.353**	**0.247**

Plugged into state-of-the-art methods

Method	DEG SCC	nDCG@200
ST-Net	0.181	0.073
ST-Net + **Ours**	**0.353**	**0.247**
TRIPLEX	0.299	0.115
TRIPLEX + **Ours**	**0.457**	**0.242**

Summary

- **Objective:** Establish prediction methods and evaluation metrics for biological knowledge discovery.
- **Method:** Optimize image-based gene ranking through pairwise ranking learning.
- **Evaluation Metrics:** Gene rank correlation (DEG SCC) and top-k ranking agreement (DEG nDCG@k).

