



# Enhancing Perturbation Modeling in Single-Cell Data through Advanced Deep Learning Approaches

Aspasia Orfanou<sup>1</sup>, Vasileios Vasileiou<sup>1,2</sup>, George Gavriilidis<sup>1</sup>, Theodoros Kantzalis<sup>3</sup>, Myrthe van Baardwijk<sup>4</sup>, Jiedan Xao<sup>5</sup>, Naveed Ishaque<sup>6</sup>, Andrew Stubbs<sup>4</sup>, Pericles A. Mitkas<sup>3</sup>, Fotis Psomopoulos<sup>1</sup>

<sup>1</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece, <sup>2</sup>Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece, <sup>3</sup>Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, <sup>4</sup>Erasmus University Medical Center, Rotterdam, Netherlands, <sup>5</sup>Erasmus Mundus Joint Master in Human Diseases Models Morphological Phenotyping (MorphoPHEN), <sup>6</sup>Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Center of Digital Health, Berlin, Germany

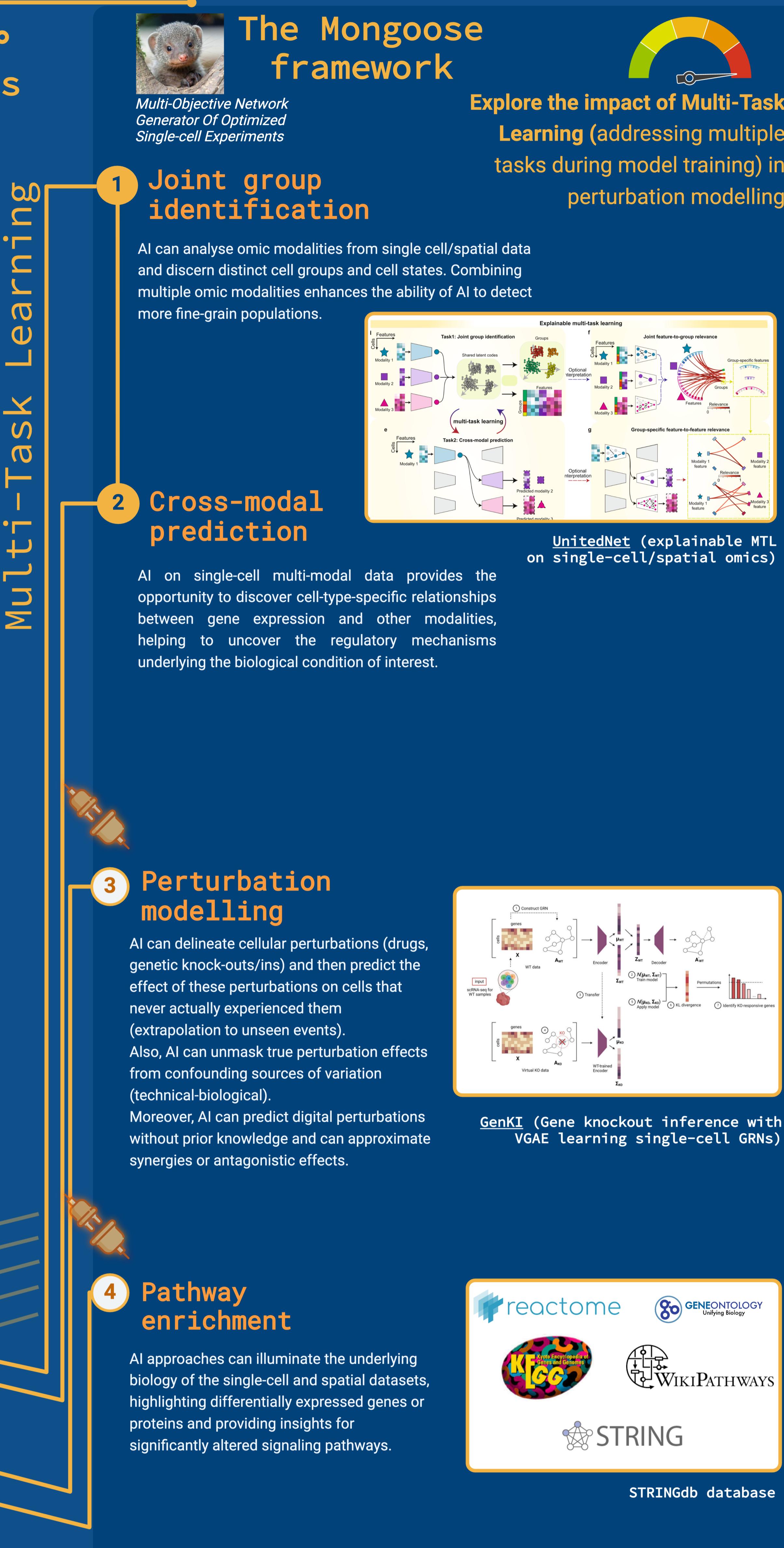
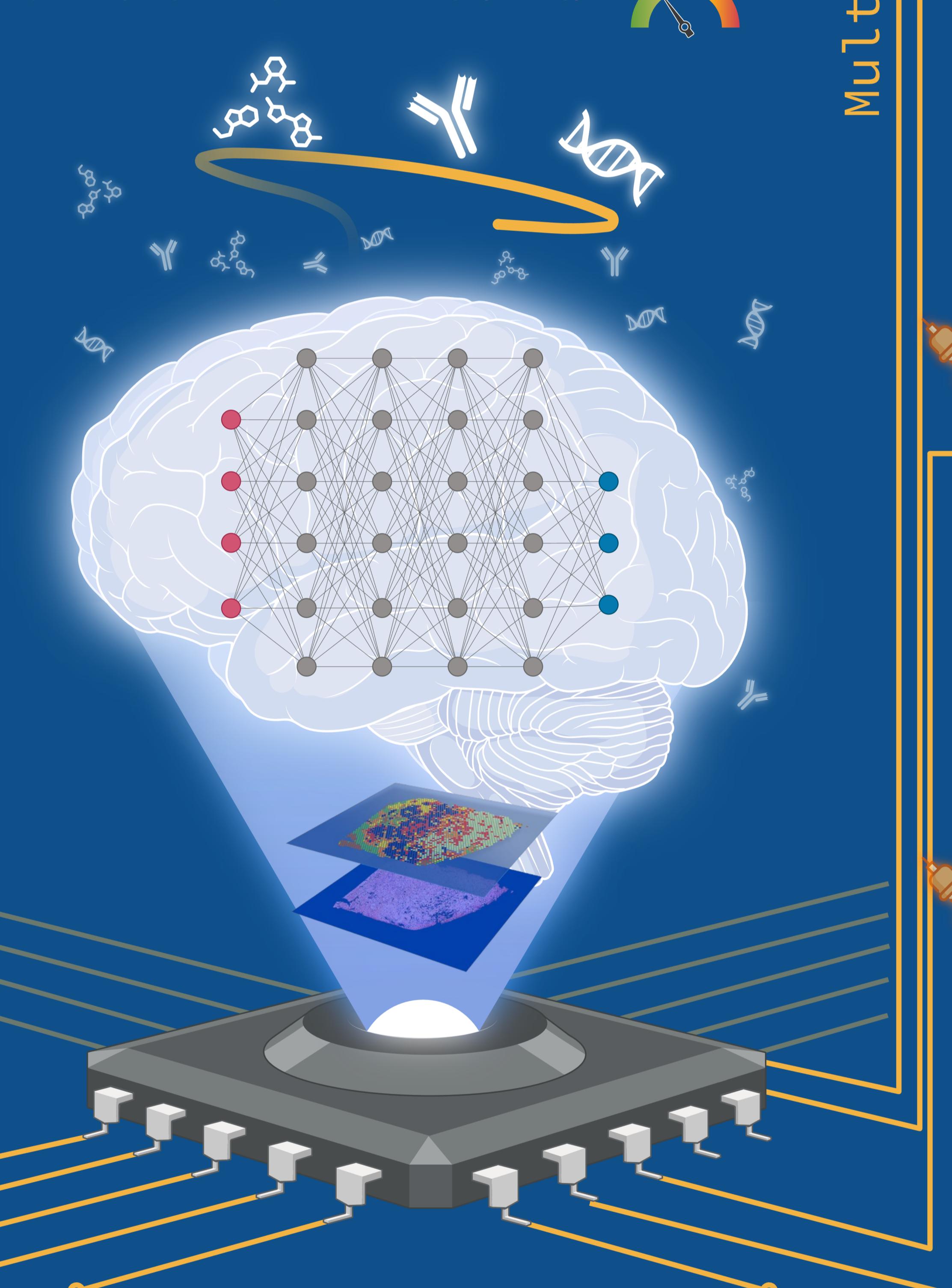
## AI in single-cell spatial omics for Systems Medicine

Artificial intelligence (AI) can perform pattern recognition across high-dimensional single-cell and spatial data, dissecting the innate cellular heterogeneity of various tissues and organs, in health and disease.

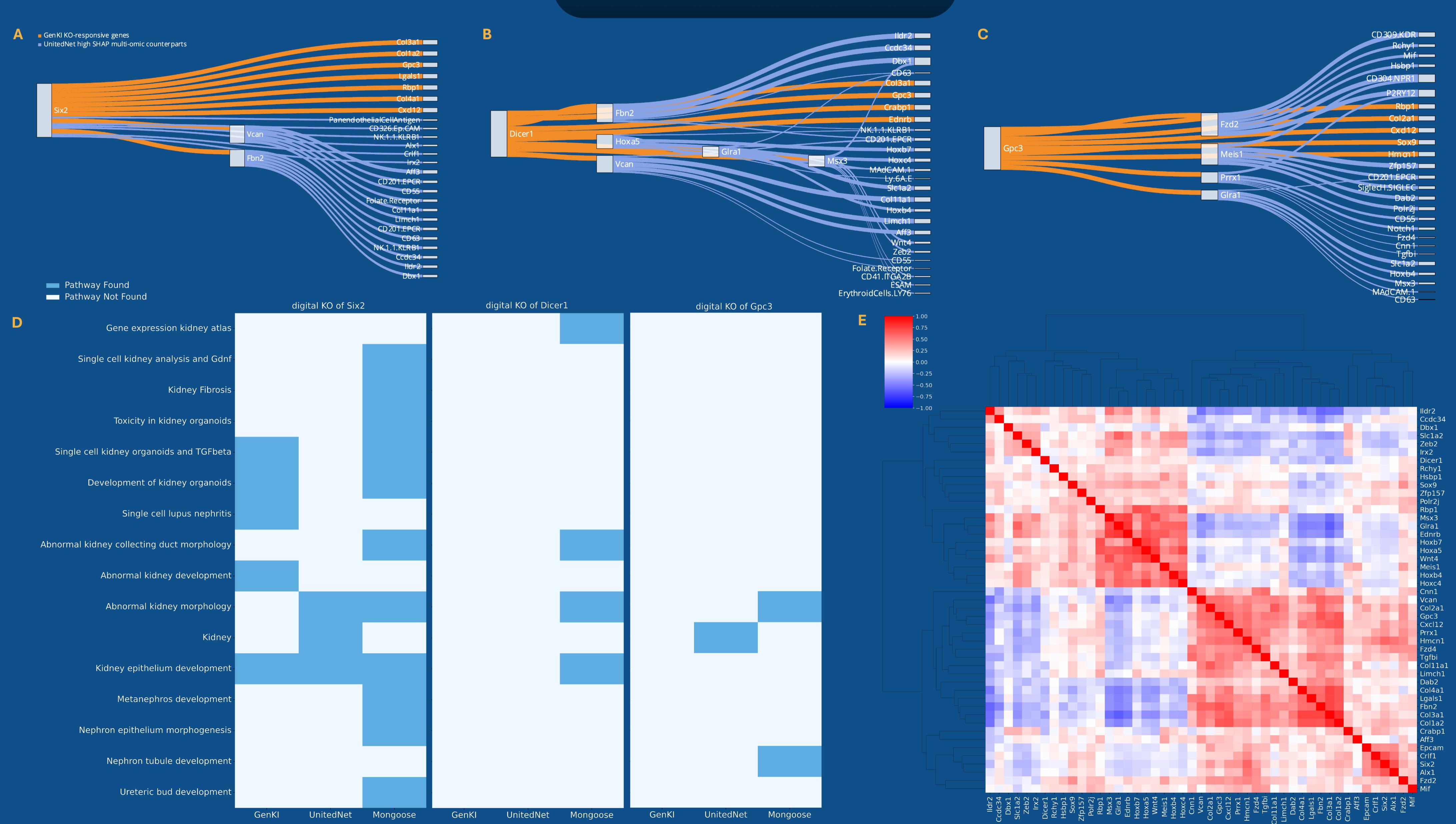
Some of the tasks that AI addresses in single-cell biology are:

- ① Joint group identification
- ② Cross-modal prediction
- ③ Perturbation modelling
- ④ Pathway enrichment

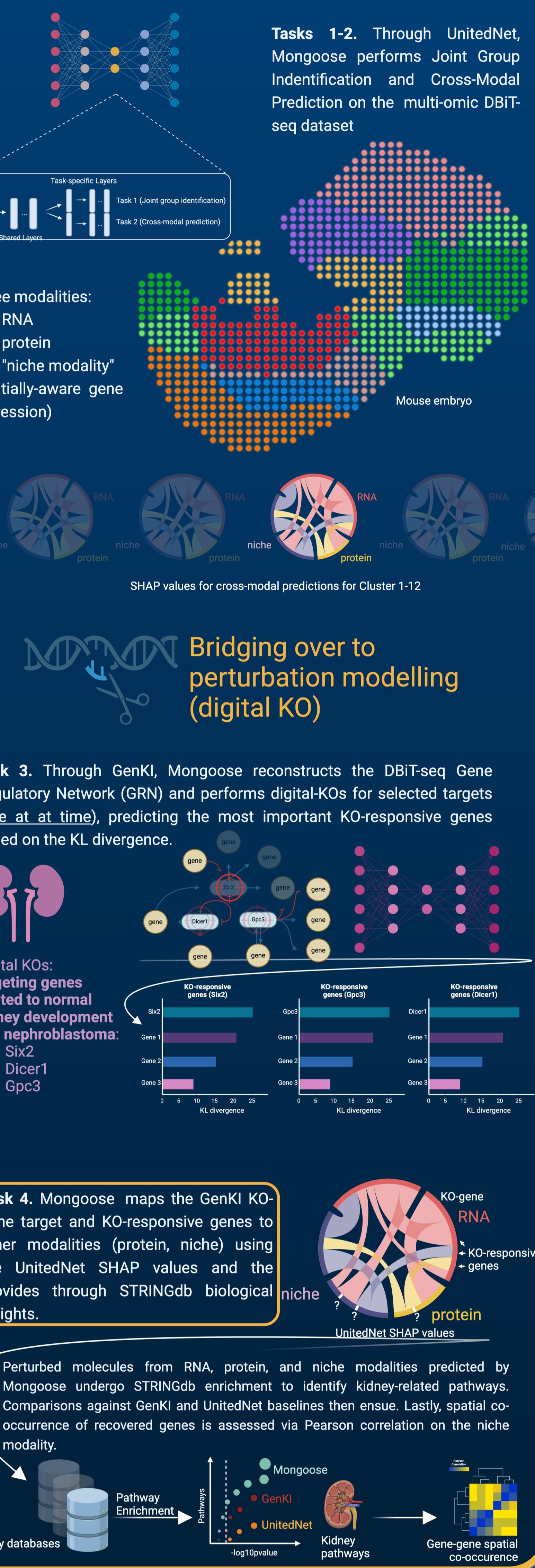
These tasks are usually pursued in isolation, offering partial glimpses of perturbed cell physiology!



## Perturbational predictions



## Deploying Mongoose in healthy DBiT-seq for kidney perturbations



## Highlights

- (A-C): Alluvial plots showing the mappings of digital KO targets Six2, Dicer1 and Gpc3 to several KO-responsive genes through GenKI (orange). KO-targets and KO-responsive genes are further mapped to multi-omic counterparts through high UnitedNet SHAP values (pink) (genes from niche modality and proteins). In this way, the Mongoose framework extends the "signal" of the digital perturbation across multiple omic modalities.
- (D): Heatmaps showing the number of kidney related pathways retrieved by baseline models (GenKI, UnitedNet) and by the Mongoose approach which outperforms them in every scenario.
- (E): Hierarchical clustering heatmap depicting Pearson correlation co-efficients for gene-gene spatial co-occurrence in the DBiT-seq dataset across all 3 scenarios. The prioritised markers that Mongoose is predicting in kidney perturbations exhibit spatial patterning that should be further researched.

## Future Directions

- Future implementations of Mongoose could be extended to multi-omic Perturb-seq and sciplex single-cell datasets which emerged from single-cell CRISPR and drug high-throughput screening experiments respectively.
- Mongoose could also other spatial technologies (e.g. Visium HD) across a variety of disease settings to uncover perturbationally-relevant insights.
- Mongoose could catalyze the creation of advanced foundational models that combine transformers and MTL to dovetail perturbation tasks with other single-cell/spatial tasks (e.g. niche identification, cell-state determination) offering more powerful predictions, closer to the biological ground-truth.

For more info

