



# From mechanism to application: Decrypting light-regulated denitrifying microbiome through geometric deep learning

Yang Liao<sup>1</sup>, Jing Zhao<sup>1</sup>, Jiyong Bian<sup>1</sup>, Ziwei Zhang<sup>2</sup>, Siqi Xu<sup>1</sup>, Yijian Qin<sup>2</sup>, Shiyu Miao<sup>1</sup>, Rui Li<sup>1</sup>, Ruiping Liu<sup>1</sup>, Meng Zhang<sup>3</sup>, Wenwu Zhu<sup>2</sup>, Huijuan Liu<sup>1</sup>, Jiuhui Qu<sup>1</sup>

- <sup>1</sup> Center for Water and Ecology, State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China  
<sup>2</sup> Department of Computer Science and Technology, Tsinghua University, Beijing, China;  
<sup>3</sup> School of Electronic and Information Engineering, Beihang University, Beijing, China

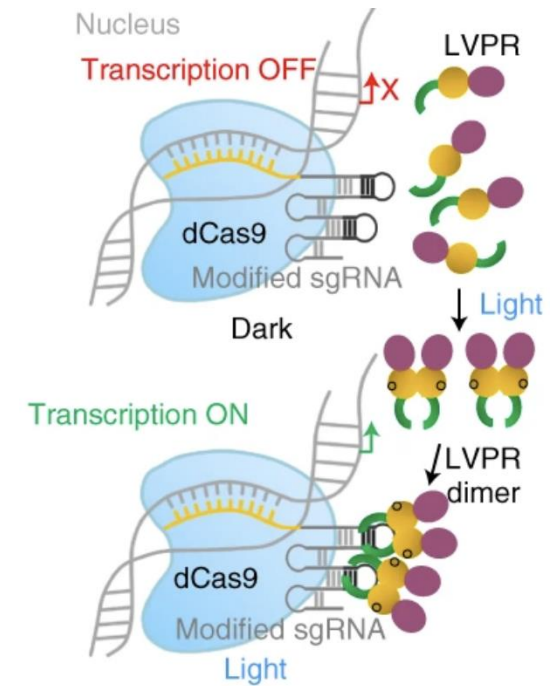


Yang Liao<sup>1</sup>, Jing Zhao<sup>1</sup>, Jiyong Bian<sup>1</sup>, Ziwei Zhang<sup>2</sup>, Siqi Xu<sup>1</sup>, Yijian Qin<sup>2</sup>, Shiyu Miao<sup>1</sup>, Rui Li<sup>1</sup>, Ruiping Liu<sup>1</sup>, Meng Zhang<sup>3</sup>, Wenwu Zhu<sup>2</sup>, Huijuan Liu<sup>1</sup>, Jiuhui Qu<sup>1</sup>. 2024. From mechanism to application: Decrypting light-regulated denitrifying microbiome through geometric deep learning. *iMeta* e162. <https://doi.org/10.1002/imt2.162>

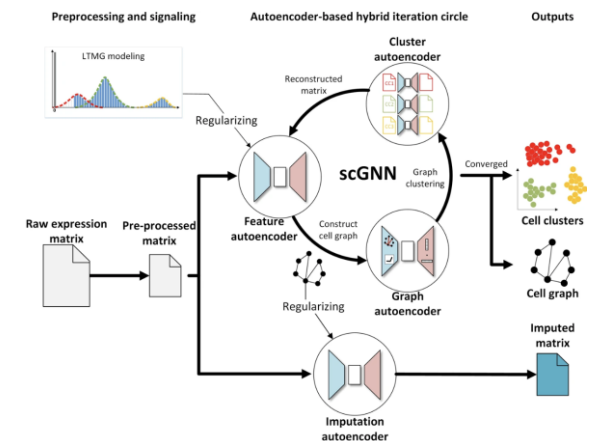


# Introduction

- **Denitrifying microbiome** is essential in maintaining nitrogen cycling in the ecosystem, as well as plays a pivotal role in socio-economic development, such as food production, energy, wastewater treatment, and resource recovery. To regulate nitrogen metabolism, **optogenetics** is a promising **regulatory strategy** on microbiome.
- Decrypting **biological big data** requires sophisticated skills and highly professional biological knowledge, which is especially challenging for **environmental microbiota**, corresponding to **meta-omics**, given its complexity and cross-species interaction.
- Deep learning exhibits superior performance in characterizing biological big data and learning genetic topological and co-expression principles, among which **graph neural network (GNN)** is the most representative and successful one in exploiting **heterogeneous information** and complex **topological relationship**.



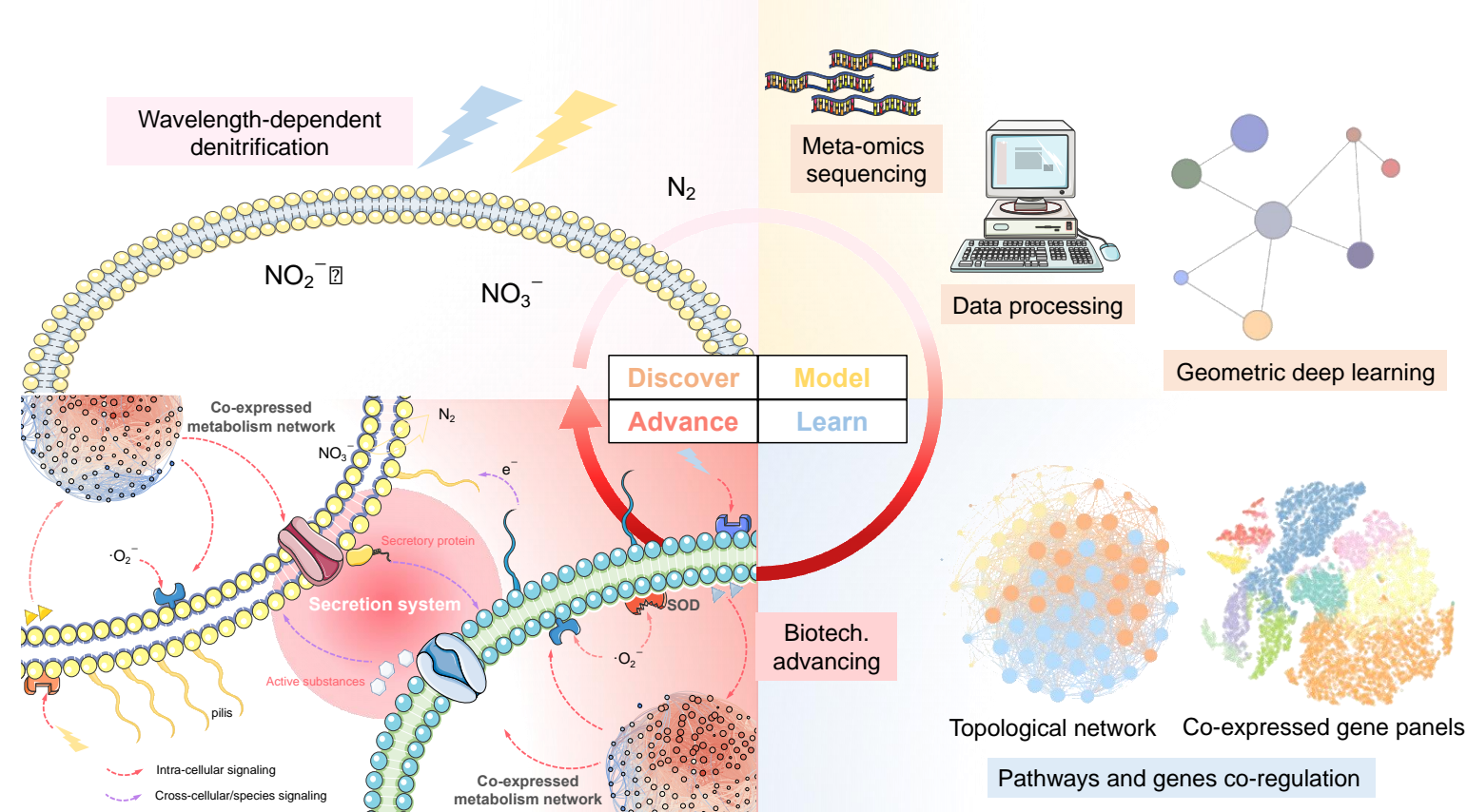
(Liu et al., *Nature Biotechnology*, 2022)



(Wang et al., *Nature Communication*, 2021)



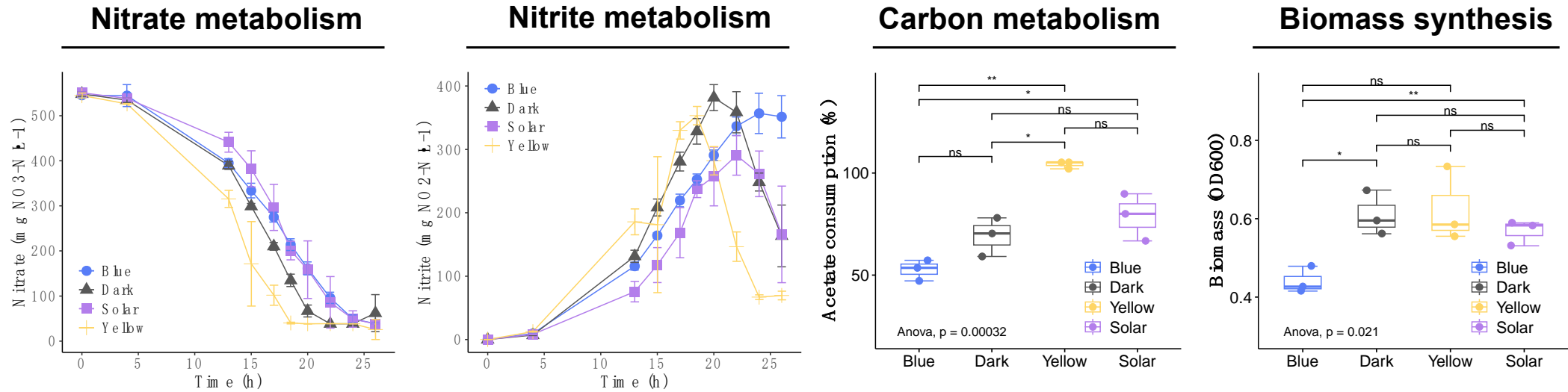
# Highlights



- **GNNs-based biology-contextualized** computational framework exhibited superior performance in identifying co-expressed gene panels and decrypting wavelength-dependent denitrification.
- Wet-lab demonstrations validated the wavelength-divergent **secretion system** and **nitrate-superoxide co-regulation** as unveiled by GNNs, which could be utilized for nitrate removal and resource recovery.
- The **co-expressed gene panels** and **topological network toolkits** were developed to guide scientific discovery and versatile biotechnology development.

# Results

## ➤ Discovery of light-regulated microbial metabolism and modeling meta-omics through GNNs

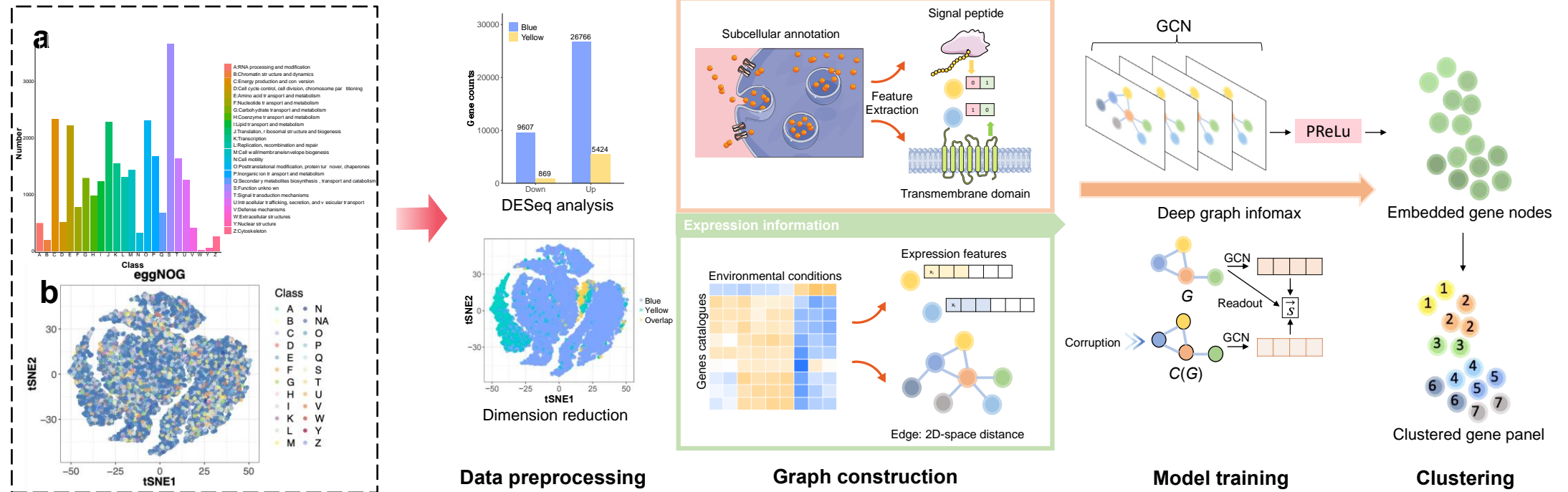


- Blue and yellow light **decomposed** the inhibitory effects of solar light.
- Blue light exhibited **inhibitory effects** on microbial metabolism, realizing partial denitrification (PD) with a 69.4% nitrite accumulation ratio (NAR) at 26 h, much higher than dark (33.6% NAR), which is favorable for partial denitrification coupled anaerobic ammonium oxidation (PD/A).
- Yellow light **boosted** both nitrate and nitrite removal, it also promoted carbon source intake by 49.5% compared to the dark, but no significant increase in biomass synthesis, indicating microbiota utilized acetate for other metabolism.
- In comparison, blue light reduced biomass accumulation by 27.8% compared to the dark, but **no significant** reduction in carbon source intake, which also implied the **metabolism fluxes diversion** from biomass synthesis.



# Results

## ➤ Discovery of light-regulated microbial metabolism and modeling meta-omics through GNNs

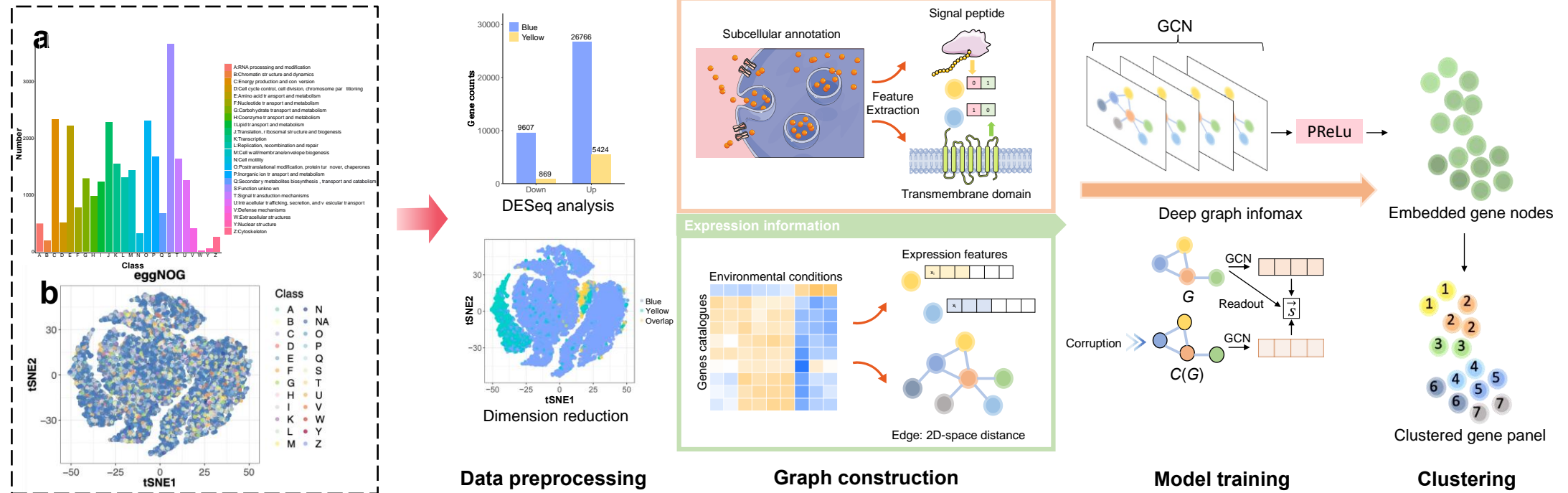


➤ Prior-knowledge-based database classification failed to obtain co-expressed gene panels (left). Meta-omics datasets are characterized by **microbial interactions** and frequent **extra- and intra-cellular** substance changes, thus, we integrated biological knowledge through **graph convolutional networks** (GCNs).



# Results

## ➤ Discovery of light-regulated microbial metabolism and modeling meta-omics through GNNs



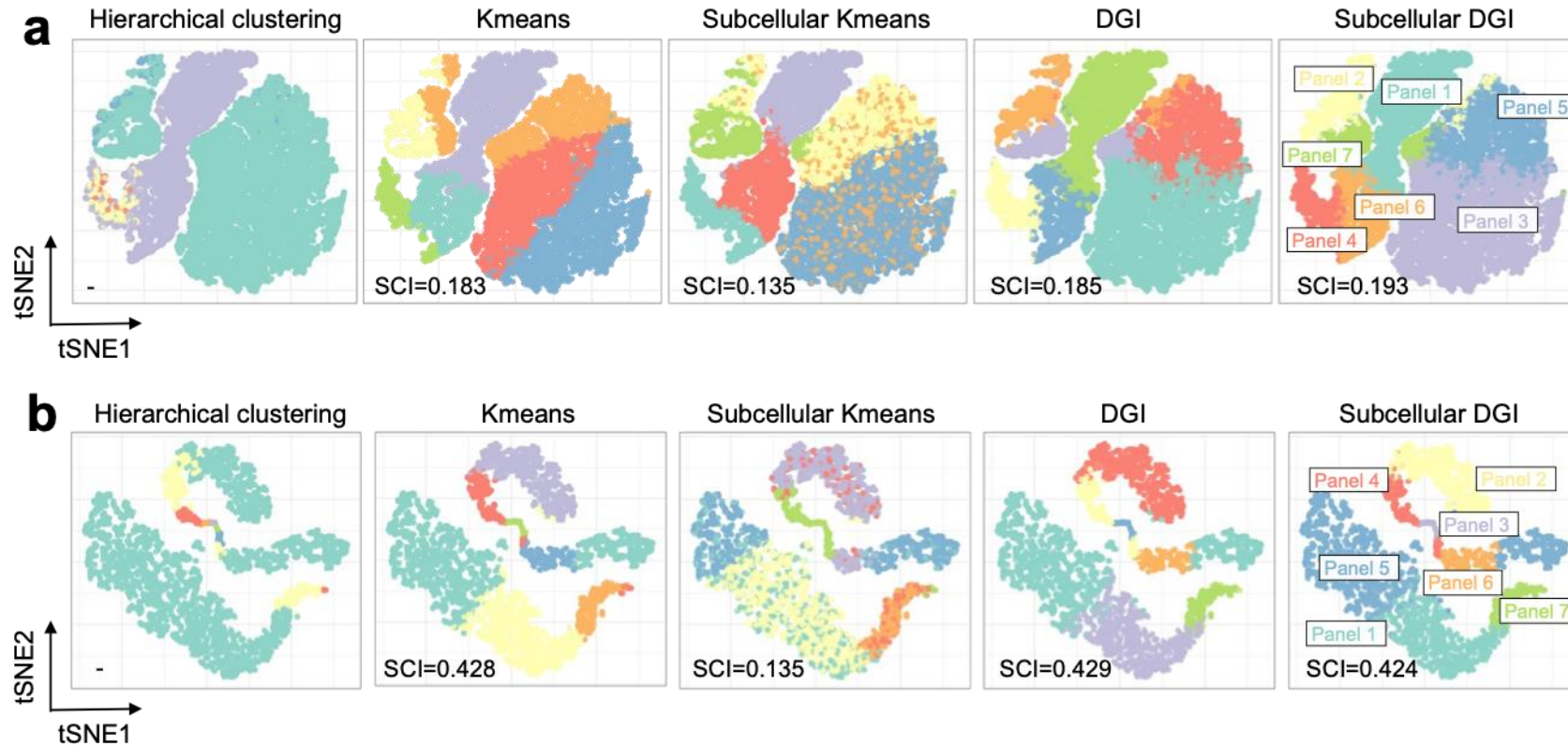
- After data processing, we obtained valid **differentially expressed genes (DEGs)**, whose subcellular information and expression information were engineered into graph-structured datasets.
- **Expression information** characterized the intra-cellular regulatory networks, whereas **subcellular information**, including signal peptides and transmembrane domains, represented the inter-cellular interaction.
- We employed **Deep Graph Infomax (DGI)** algorithm to unsupervised learn the node embedding, which were then clustered to obtain the **gene panels**.



# Results

## ➤ Geometric deep learning achieves superior performance in gene panel identification

—— Mathematical validation: qualitative and quantitative



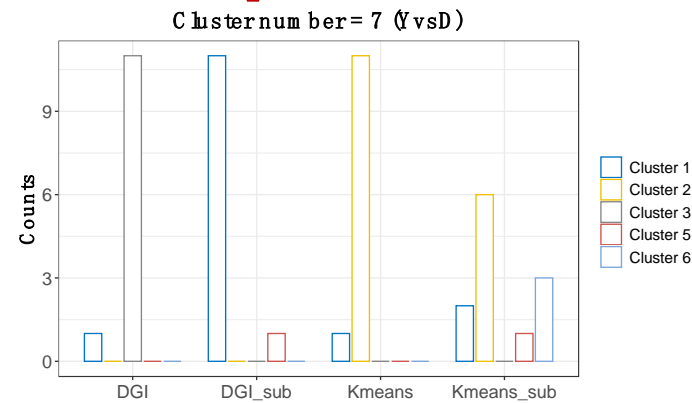
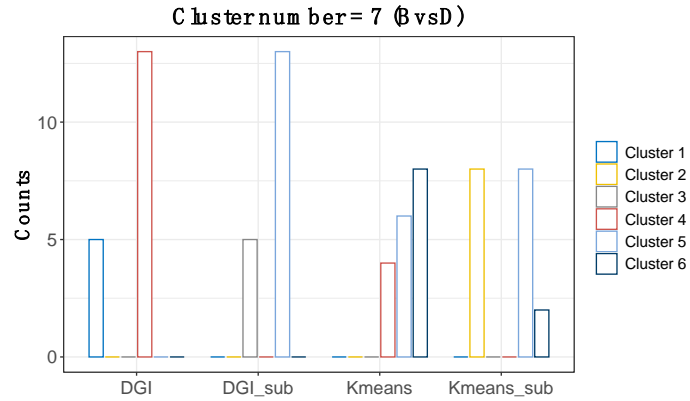
- Commonly used method, e.g. Hierarchical clustering, presented strong **biases** that tended to cluster most genes into 2 clusters.
- Unsupervised learning through DGI obtained clearer gene panels based on gene expression and subcellular information, which exhibited its **superior capability in integrating heterogeneous information**.



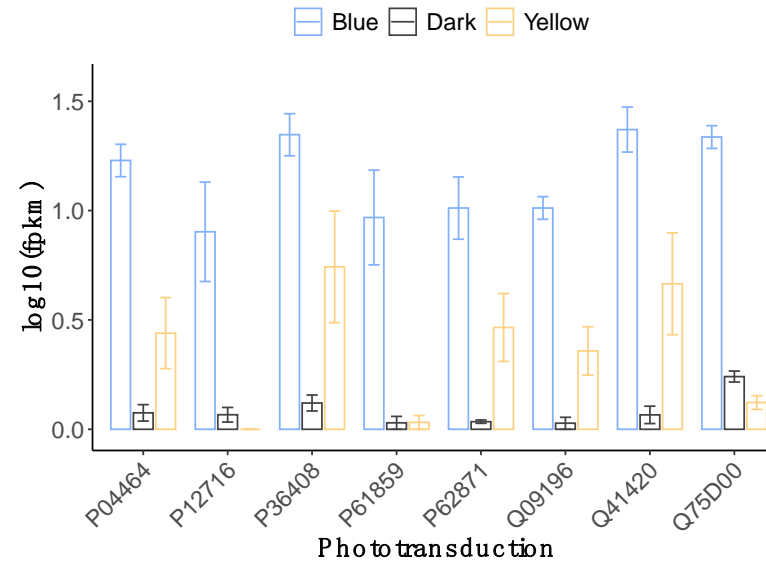
# Results

## ➤ Geometric deep learning achieves superior performance in gene panel identification

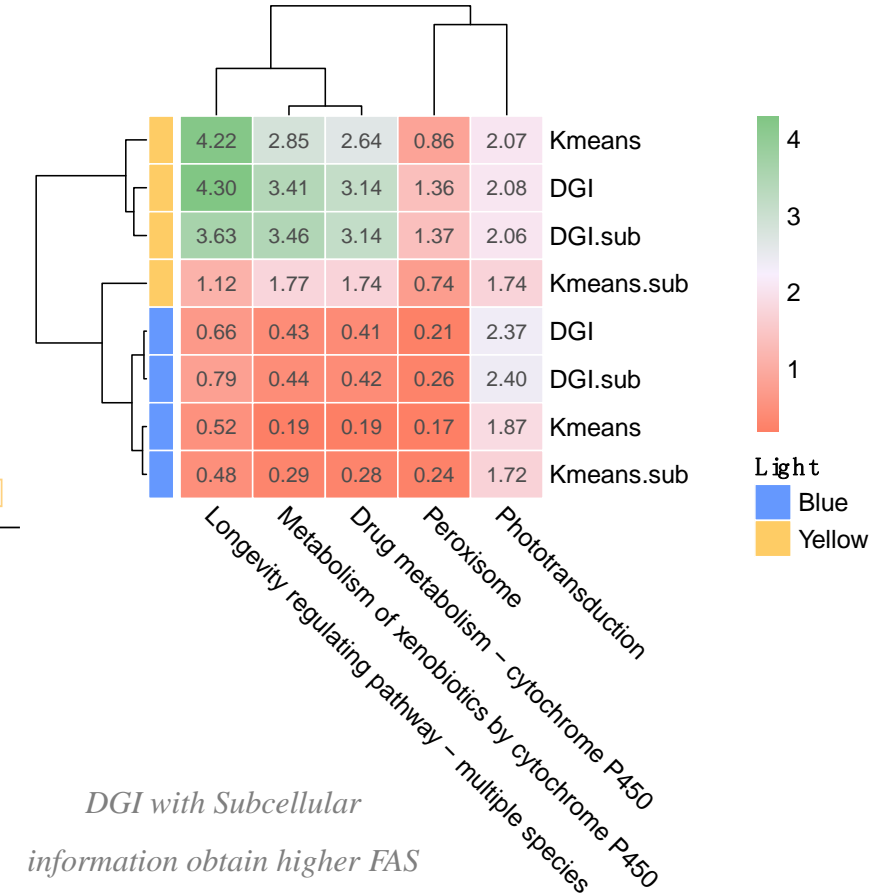
—— Biological validation: biological function consistency



Assign to the same cluster



Co-expression pattern of phototransduction genes



DGI with Subcellular information obtain higher FAS

- Function assignment score (FAS) indicated the outstanding biological functions matching performance of geometric deep learning, which implied it could **better identify genes for certain functions.**



# Results

## ➤ Predicting phenotype through differential pathways in the hub and signaling gene panels

➤ **Hub gene panels (HGP):** metabolism pathways that characterized with high expression levels and low fold changes, corresponding to the collective behaviors of microbiota (phenotype).

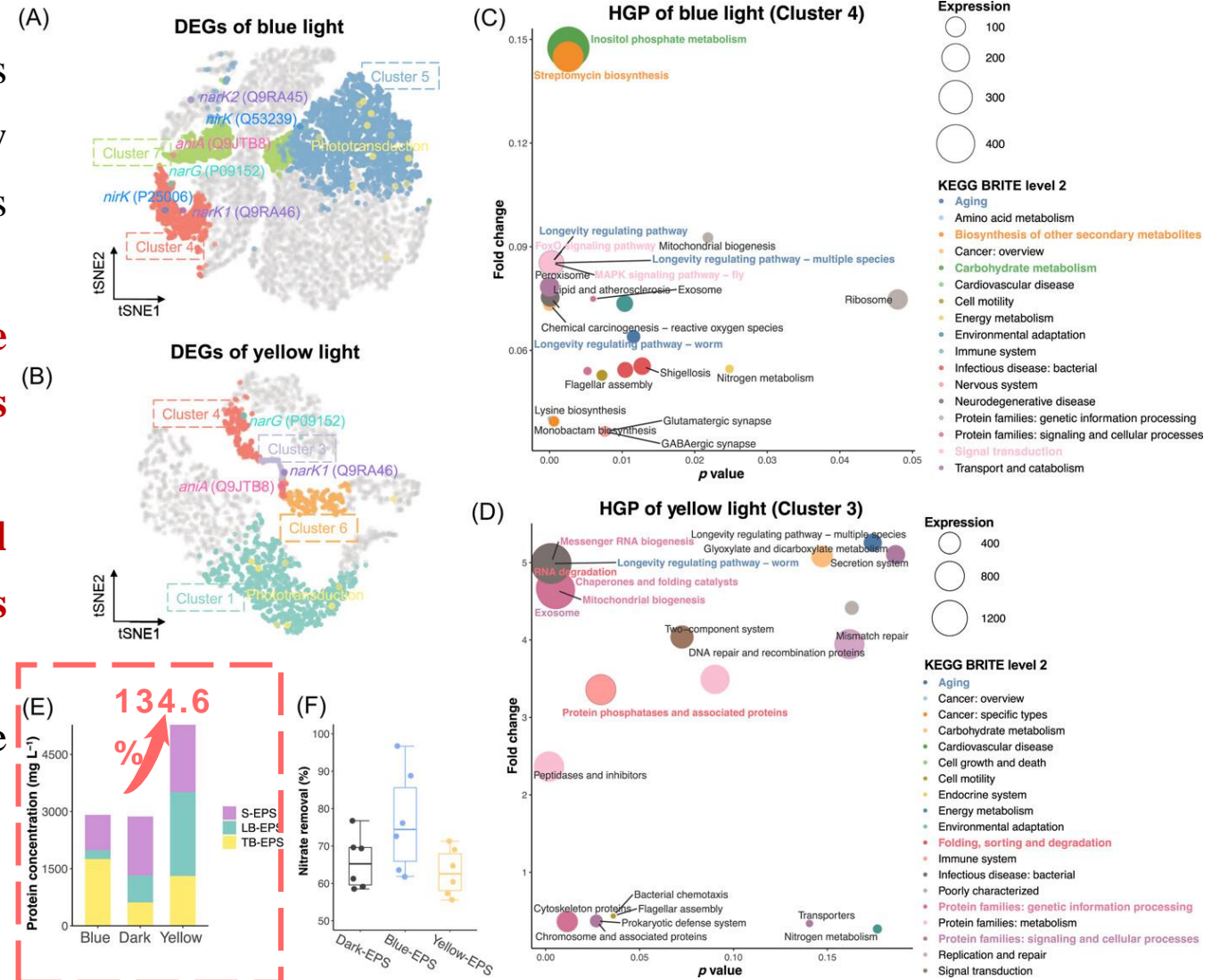
➤ For yellow light, pathways related to **reactive oxygen species (ROS), peptide, and protein synthesis** were enriched.

➤ For blue light, pathways related to **signal transduction, oxidative stress and second metabolites** were enriched.

➤ **Secretion system** was up-regulated under both blue and yellow light.

✓ **Wet-lab validation: protein production**

## — Toolkit 1: co-regulation panels for new discoveries





# Results

## ➤ Predicting phenotype through differential pathways in the hub and signaling gene panels

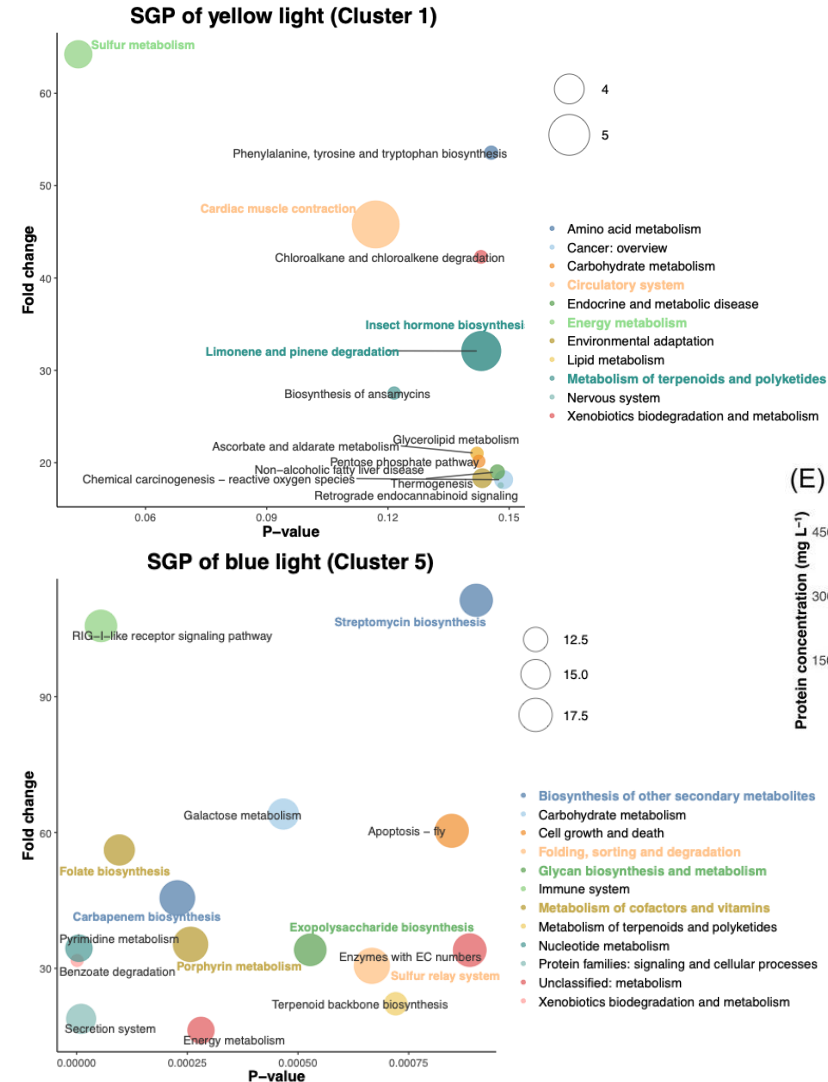
➤ **Signaling gene panels (SGPs):** metabolism pathways that characterized with low expression levels but high fold changes, which are belong to or similar to bioactive substances.

➤ For yellow light, SGPs exhibited high expression of **terpenoids and polyketides**, while **sulfur metabolism** was significantly enriched.

➤ For blue light, SGPs exhibited high expression **cofactors, vitamins** and other **secondary metabolites**. These were diverse bioactive substances.

✓ **Wet-lab validation: Biocatalytic effects of EPS**

## —— Toolkit 1: co-regulation panels for new discoveries





# Results

## ➤ Regulating microbiome through landmark genes of topological networks

### —— Toolkit 2: topology network with landmark genes for regulation strategy

➤ Case 1 - Blue light: **superoxide**

**dismutase (SOD)** was the top 1

landmark genes of the community

**nitrate reductase (NarG)** subjected

to.

✓ **Wet-lab validation:** New

mechanism & regulatory

strategy —— nitrate-superoxide

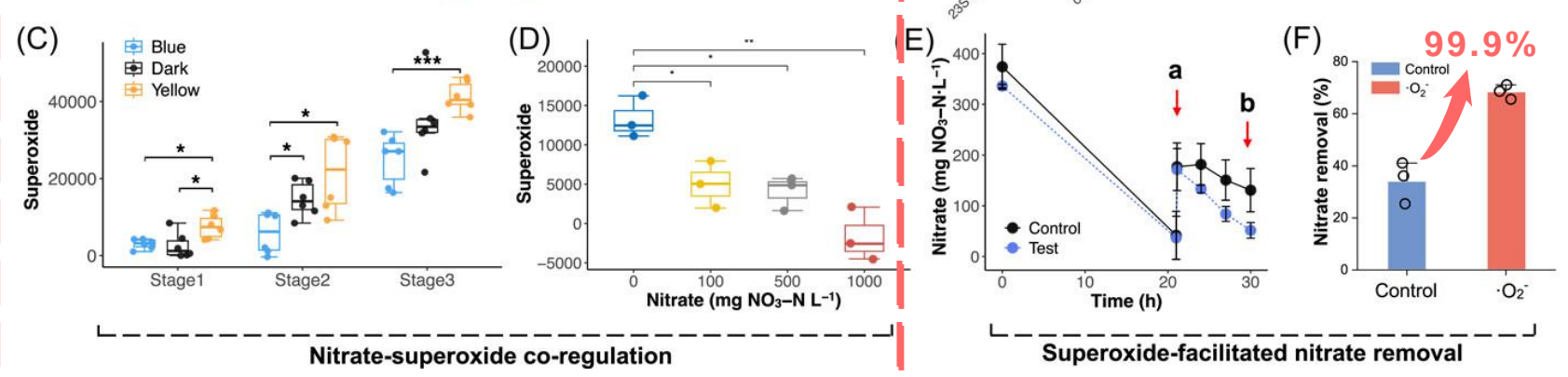
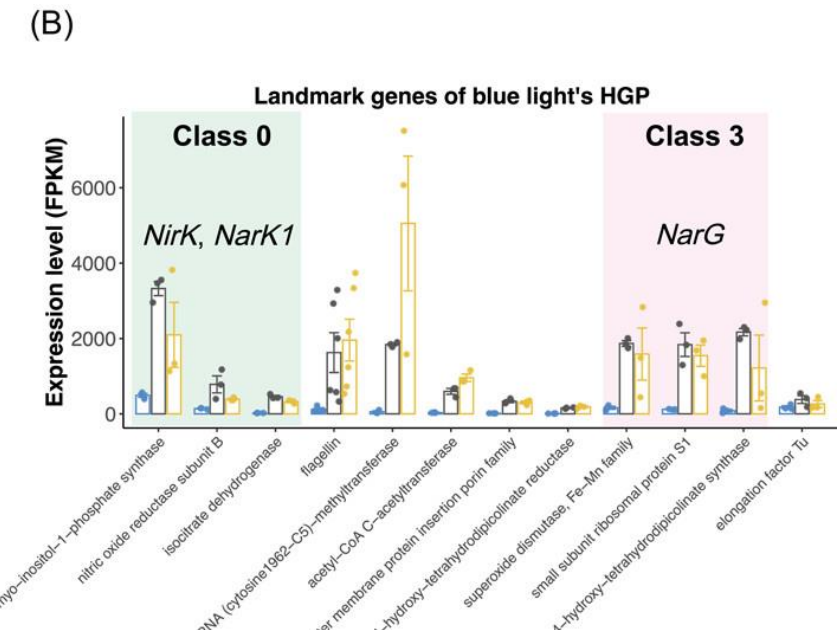
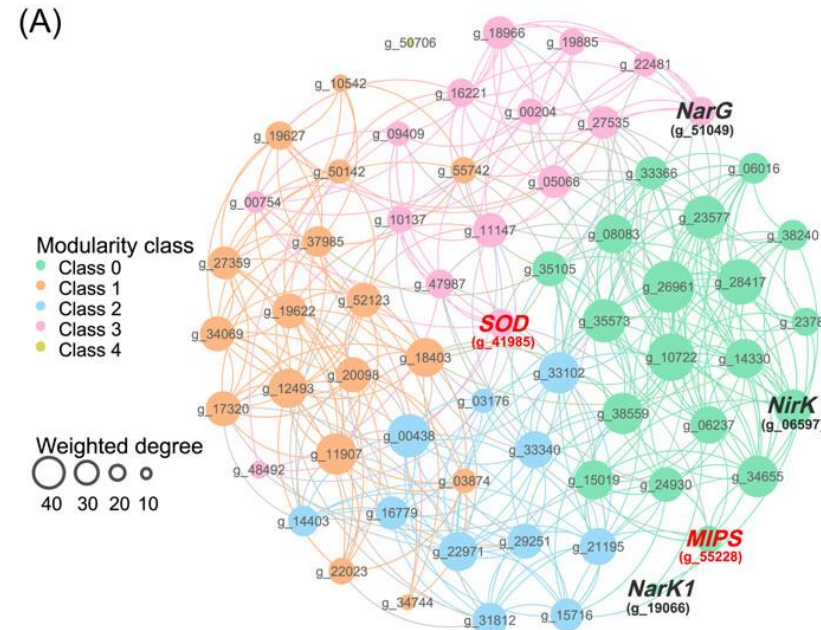
co-regulation

✓ **Wet-lab validation:** New

biotechnology —— create a micro-

stimulate environment with functional

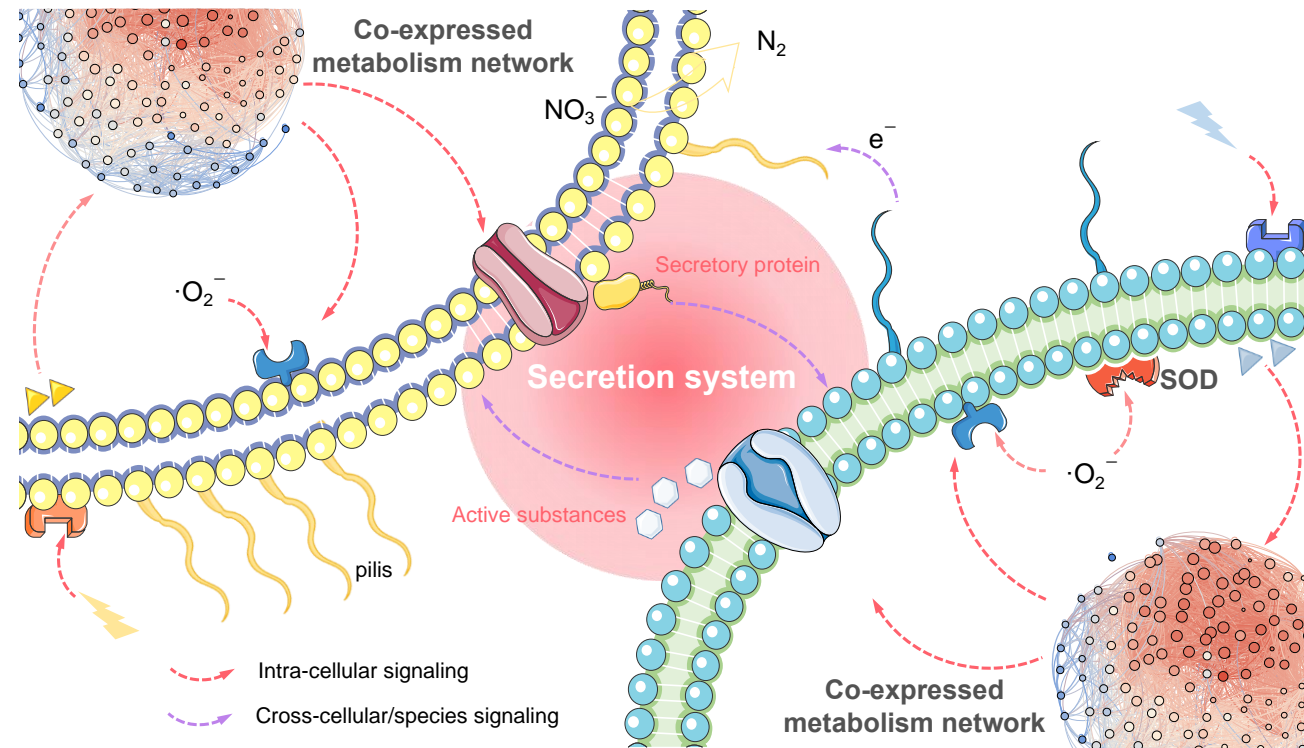
enzyme to removal nitrogen pollutant





# Results

## ➤ The mechanism and potentials of light-regulated photo-denitrification

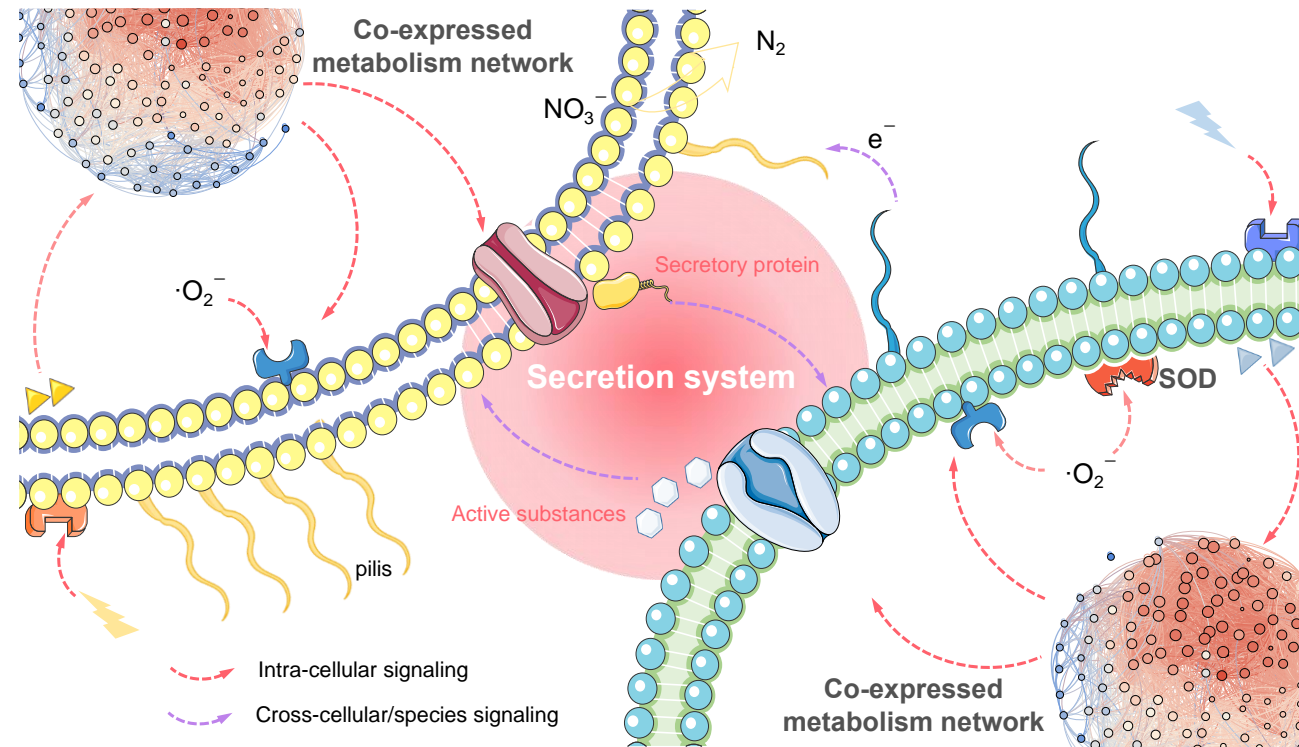


- Overall, **secretion system** was the core of cross-species interaction.
- Blue light photoreceptors are ubiquitous and have been implemented in a broad spectrum of biological platforms, explaining the **decentralized metabolism fluxes** triggered by blue light in photo-denitrification. More diverse metabolites are synthesized, typically active substances like **co-factors and vitamins**. Some of those secreted molecules were crucial for microbiomes to maintain homeostasis under photochemical stress. These metabolites are potential high-value resources, which shed light on bioprocess that could **utilize nitrate as substrates** to save the costs of high-value chemical production.



# Results

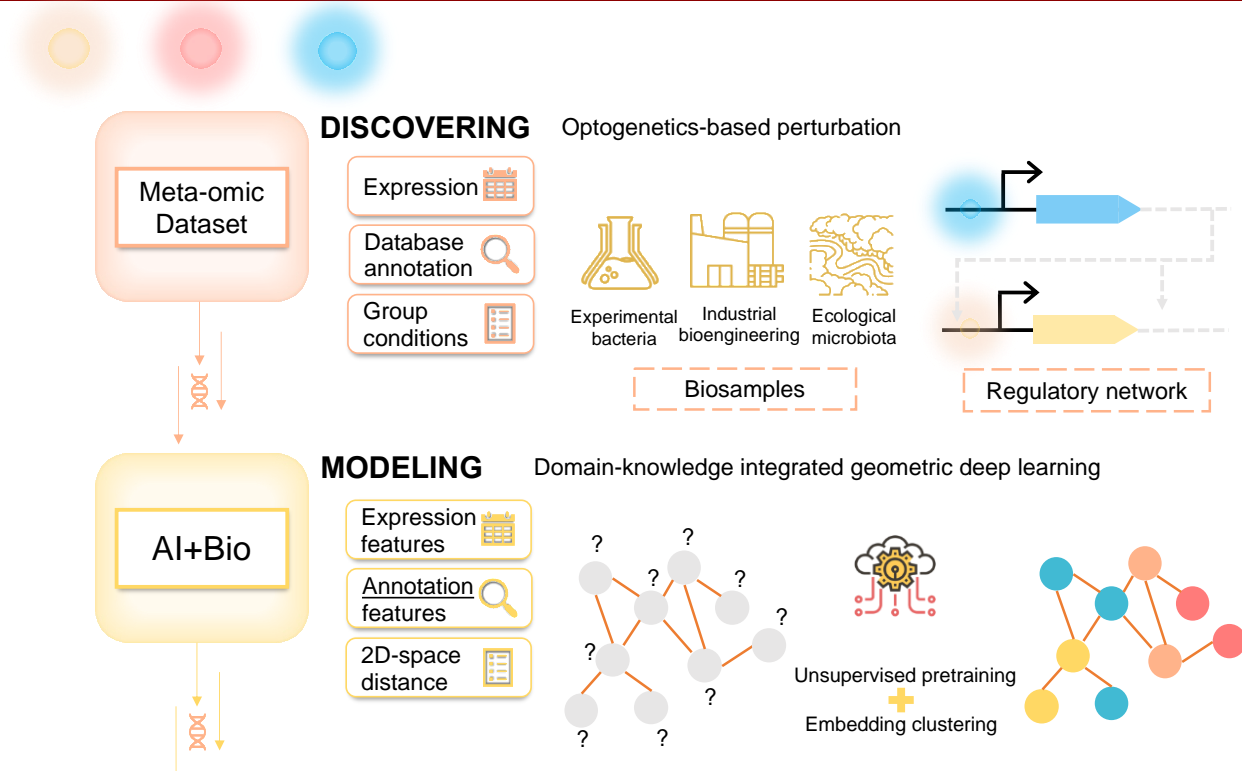
## ➤ The mechanism and potentials of light-regulated photo-denitrification



- What's interesting is the **superior activation effects of yellow light**, which has few reports on yellow light optogenetic switches. This might contribute to the **cross-species interactions**. The centralized metabolism fluxes triggered by yellow light were mostly used for protein synthesis, especially **pilus-related proteins**, suggesting the role of pilis in accelerating interspecies electron transfer for collective functional metabolisms, such as the enriched **terpenoids and polyketides** in SGPs of yellow light.



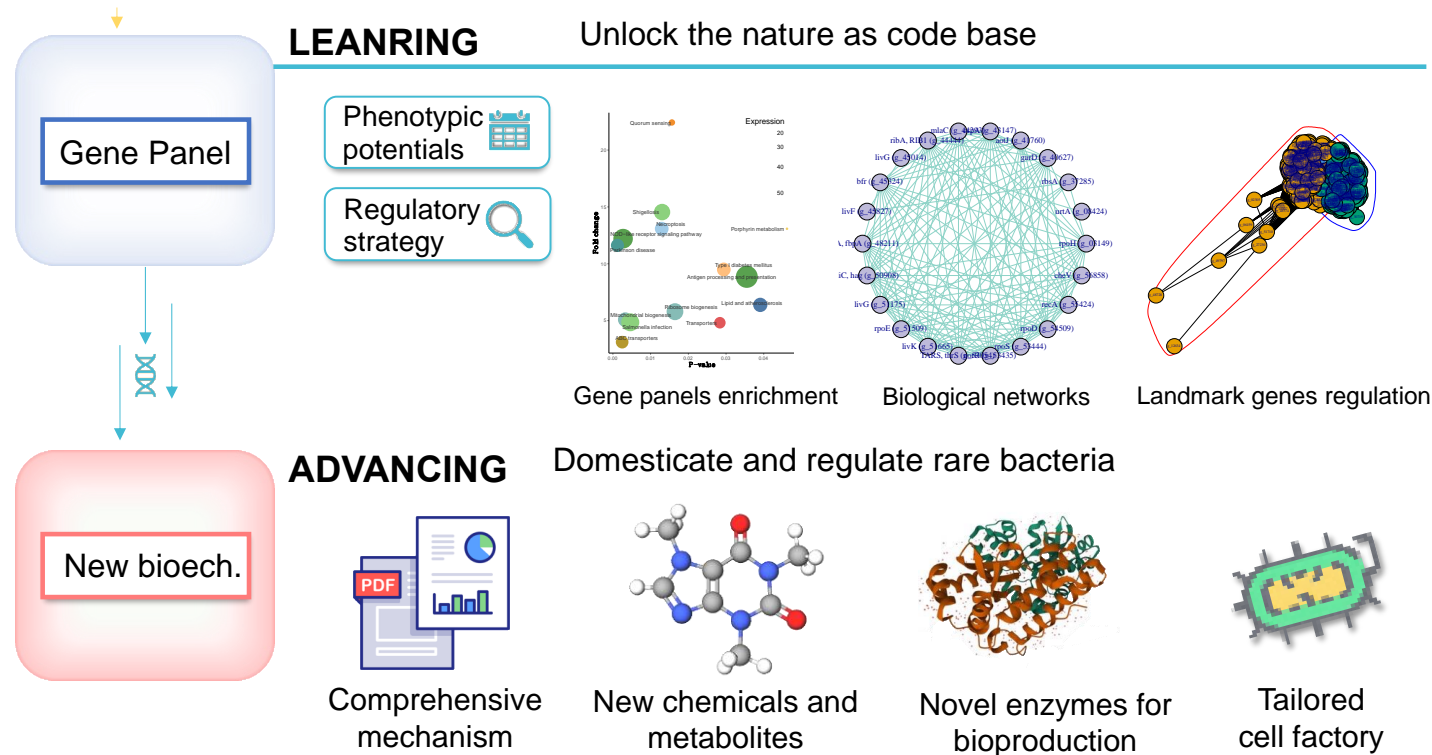
# Discussion



- **Discovering:** Environmental perturbation is necessary to activate genetic processes to capture the dynamic regulatory network. Here, we employ **optogenetic-based genetic perturbation** to trigger signal transfers through optogenetic switches. The optogenetic-induced dynamic regulatory network changes are contained in the input datasets, including expression matrix, database annotation, and group conditions.
- **Modeling:** Biological topology principles are captured by models automatically. The biological datasets, including **expression levels** and **subcellular annotations**, are engineered as node features and gene expression distance in the 2D latent space as edges. The heterogeneous knowledge is integrated through unsupervised pretraining to obtain embeddings, which are then clustered to obtain co-expressed gene panels.



# Discussion



- **Learning:** The app suites unleash the potential of nature code base. Based on the gene panels, enriched pathways can **unveil unknown phenotypes**. The biological network characterized by landmark genes is a powerful tool for elucidating **biological mechanisms** and developing **regulatory strategies**.
- **Advancing:** Based on above process, the new biotechnology could be utilized to **domesticate and regulate rare bacteria**, providing versatile applications, including comprehensive mechanism reports, new chemicals and metabolites discovery, developing novel enzymes for bioproduction, tailored cell factory for synthetic biology, etc.

**iMeta:** Integrated meta-omics to change the understanding of the biology and environment

**WILEY**



“*iMeta*” is an open-access Wiley partner journal launched by scientists of the Chinese Academy of Sciences. *iMeta* aims to promote metagenomics, microbiome, and bioinformatics research by publishing original research, methods, or protocols, and reviews. The goal is to publish high-quality papers (Top 10%, IF > 15) targeting a broad audience. Unique features include video submission, reproducible analysis, figure polishing, APC waiver, and promotion by social media with 500,000 followers. Three issues were released in [March](#), [June](#), and [September](#) 2022.



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

Submission: <https://mc.manuscriptcentral.com/imeta>



[office@imeta.science](mailto:office@imeta.science)



[iMeta](#)



[iMetaScience](#)



[iMetaScience](#)