



# 从机理到应用：几何深度学习解密光控反硝化微生物组

廖洋<sup>1</sup>, 赵婧<sup>1</sup>, 边继踊<sup>1</sup>, 张子威, 徐斯琪<sup>1</sup>, 秦一鉴<sup>2</sup>, 苗时雨<sup>1</sup>, 李瑞<sup>1</sup>, 刘锐平<sup>1</sup>, 张梦<sup>3</sup>, 朱文武<sup>2</sup>, 刘会娟<sup>1</sup>, 曲久辉<sup>1</sup>

<sup>1</sup> 水质与水生态研究中心, 环境模拟与污染控制国家重点实验室, 环境学院, 清华大学, 北京, 中国

<sup>2</sup> 计算机科学与技术系, 清华大学, 北京, 中国

<sup>3</sup> 电子信息工程, 北京航空航天大学, 北京, 中国

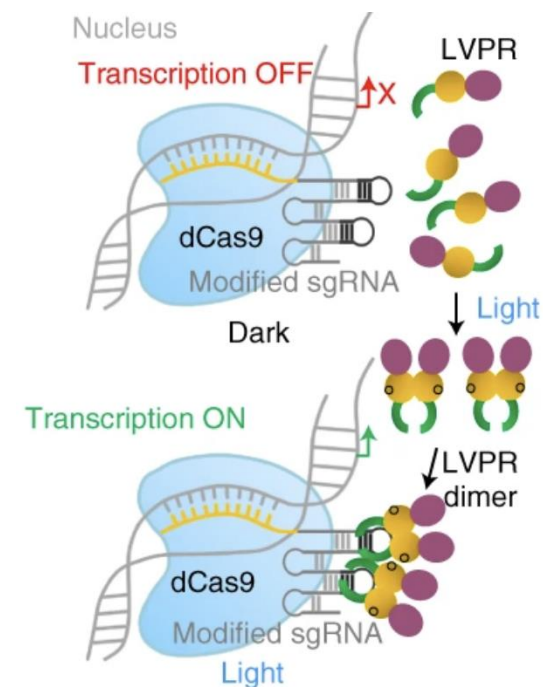


Yang Liao<sup>1</sup>, Jing Zhao<sup>1</sup>, Jiyong Bian<sup>1</sup>, Ziwei Zhang<sup>2</sup>, Siqi Xu<sup>1</sup>, Yijian Qin<sup>2</sup>, Shiyu Miao<sup>1</sup>, Rui Li<sup>1</sup>, Ruiping Liu<sup>1</sup>, Meng Zhang<sup>3</sup>, Wenwu Zhu<sup>2</sup>, Huijuan Liu<sup>1</sup>, Jiuhui Qu<sup>1</sup>.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

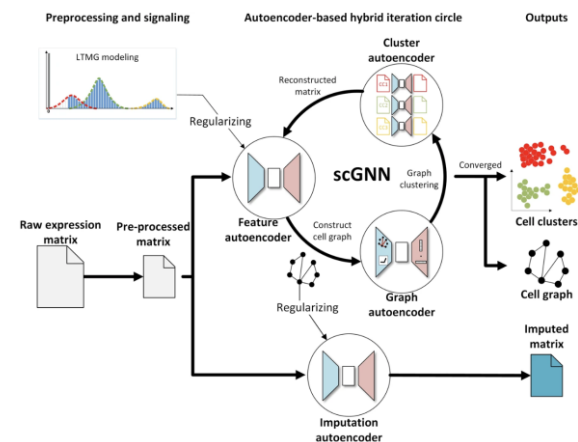


# 简介

- **反硝化微生物组**对维持生态系统中的氮循环至关重要，并在社会经济发展中发挥着关键作用，如粮食生产、能源、废水处理和资源回收。为了调节氮代谢，**光遗传学**是一种很有前途的微生物组**调控策略**。
- 解密**生物大数据**需要复杂的技能和高度专业的生物知识，蕴含环境微生物大数据的**宏组学**由于包括复杂的跨物种相互作用信息，其解析尤其具有挑战性。
- 深度学习在表征生物大数据、学习遗传拓扑和共表达原理方面表现出优越的性能，其中**图神经网络（GNN）**是利用异构信息和复杂拓扑关系最具代表性和最成功的网络。



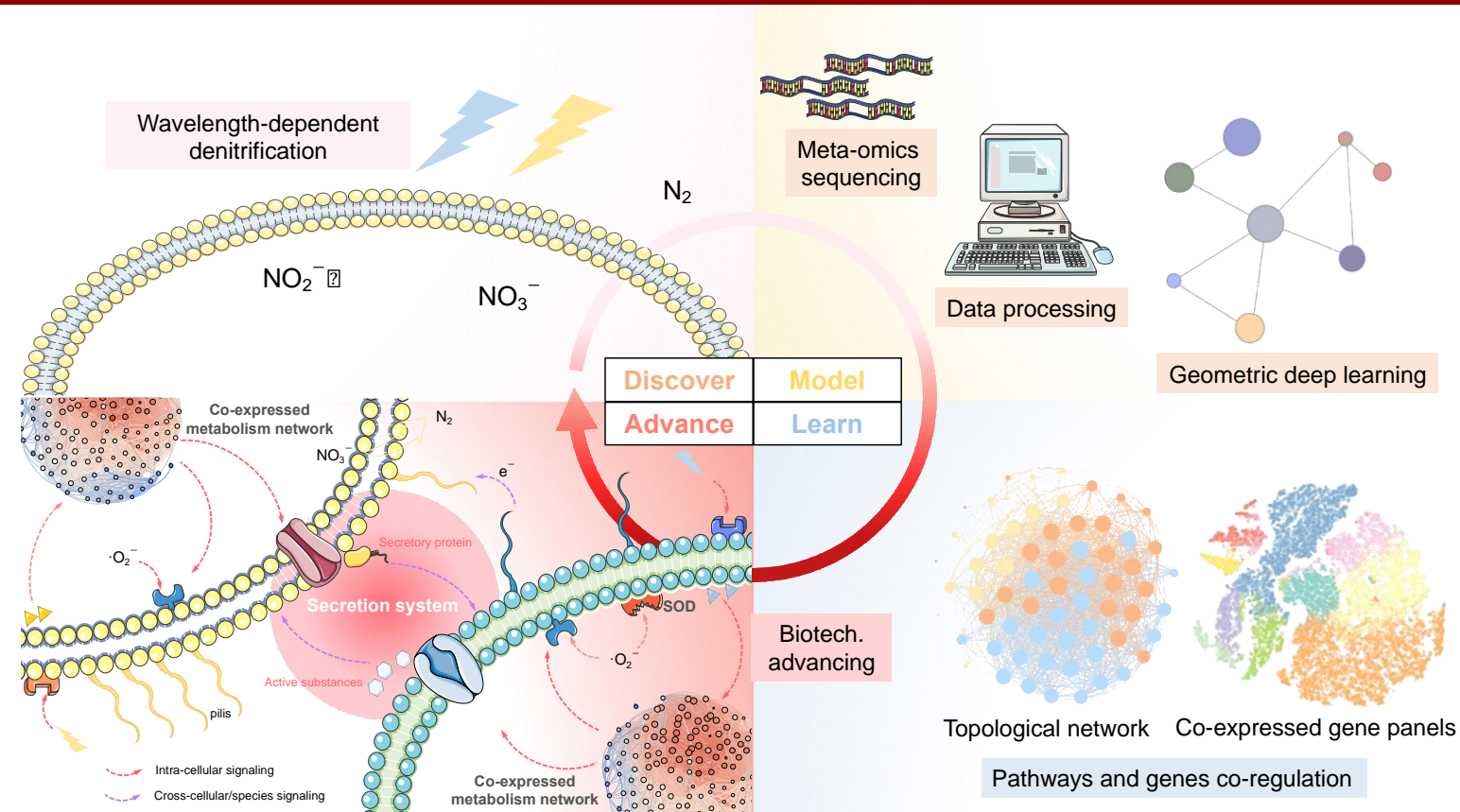
(Liu et al., *Nature Biotechnology*, 2022)



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



# 亮点

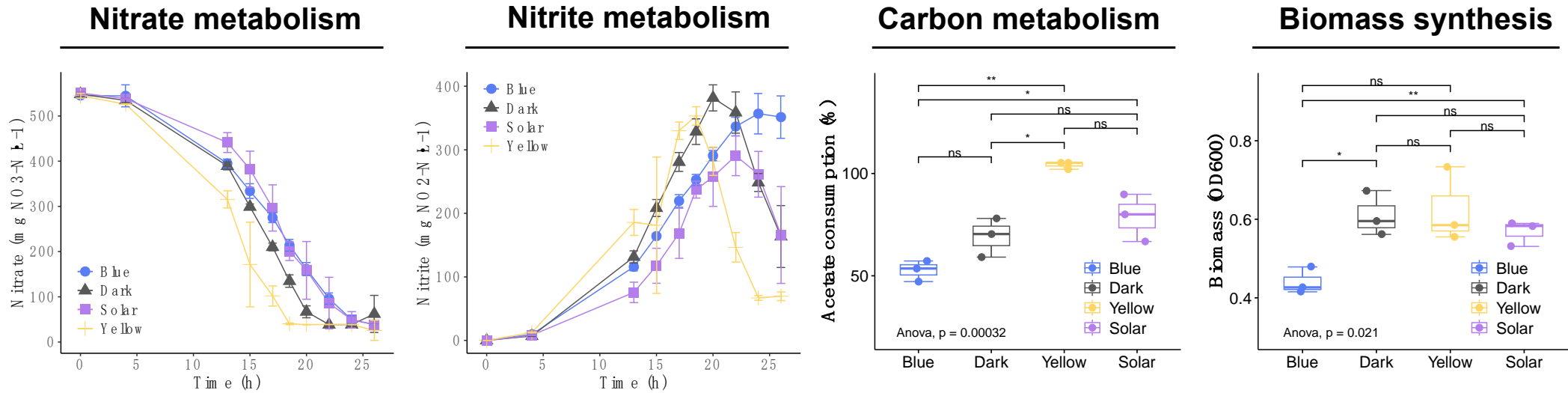


- 基于**图神经网络**（GNNs）且**融入生物领域知识**的计算框架，在识别共表达基因面板、解密波长依赖性反硝化上表现出优异的性能。
- 湿实验室验证了GNNs揭秘的知识，包括波长特异的**分泌系统**和**硝酸盐-超氧自由基**共调控线下，并可用于硝酸盐去除和资源回收。
- 本研究开发了**共表达基因面板**和**拓扑网络**工具包，从而实现加速新科学知识的发现和多功能生物技术的开发。



# 结果

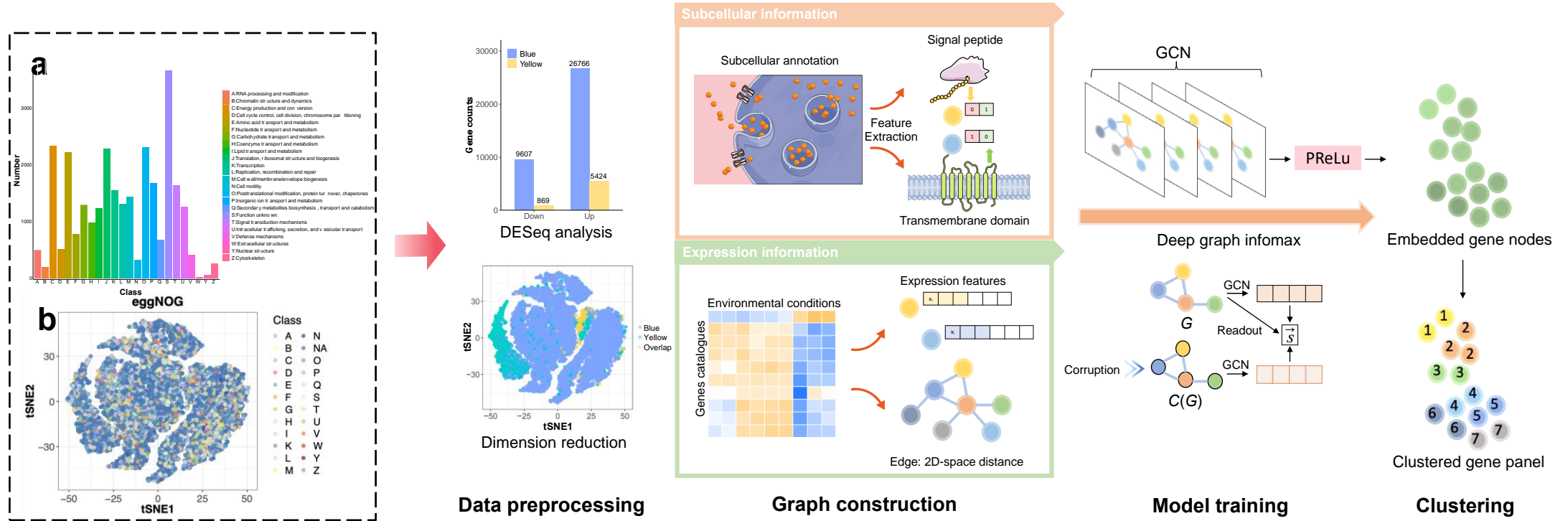
## 光控生物代谢的发现与图神经网络建模宏组学



- 蓝光和黄光**分解**了太阳光的抑制作用。
- 蓝光对微生物代谢具有**抑制作用**，在26 h时实现了部分反硝化（PD），亚硝酸盐积累率（NAR）为**69.4%**，远高于暗态（**33.6%**），有利于部分反硝化-厌氧氨氧化（PD/a）。
- 黄光**促进**了硝酸盐和亚硝酸盐的去除，与暗光相比，黄光还促进了**49.5%**的碳源摄入，但生物量合成没有显著增加，表明微生物群利用乙酸盐进行其他代谢。
- 与黑暗相比，蓝光**减少**了**27.8%**的生物量积累，但碳源摄入**没有显著减少**，这也意味着**代谢通量**从生物量合成中**转移**。

# 结果

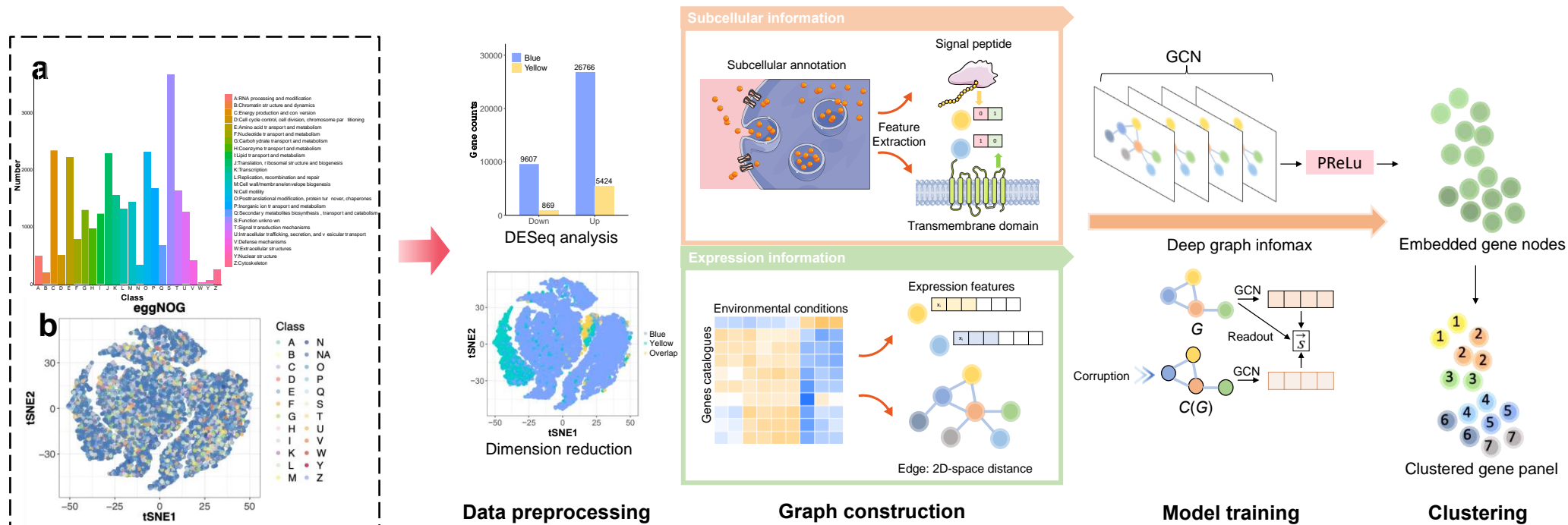
## 光控生物代谢的发现与图神经网络建模宏组学



➤ 基于先验知识的数据库分类未能获得有效的共表达基因面板（左）。宏组学数据集具有**微生物相互作用**和频繁的**细胞内外物质交换**的特点。因此，我们通过**图卷积网络（GCN）**整合了这些生物学知识。

# 结果

## 光控生物代谢的发现与图神经网络建模宏组学



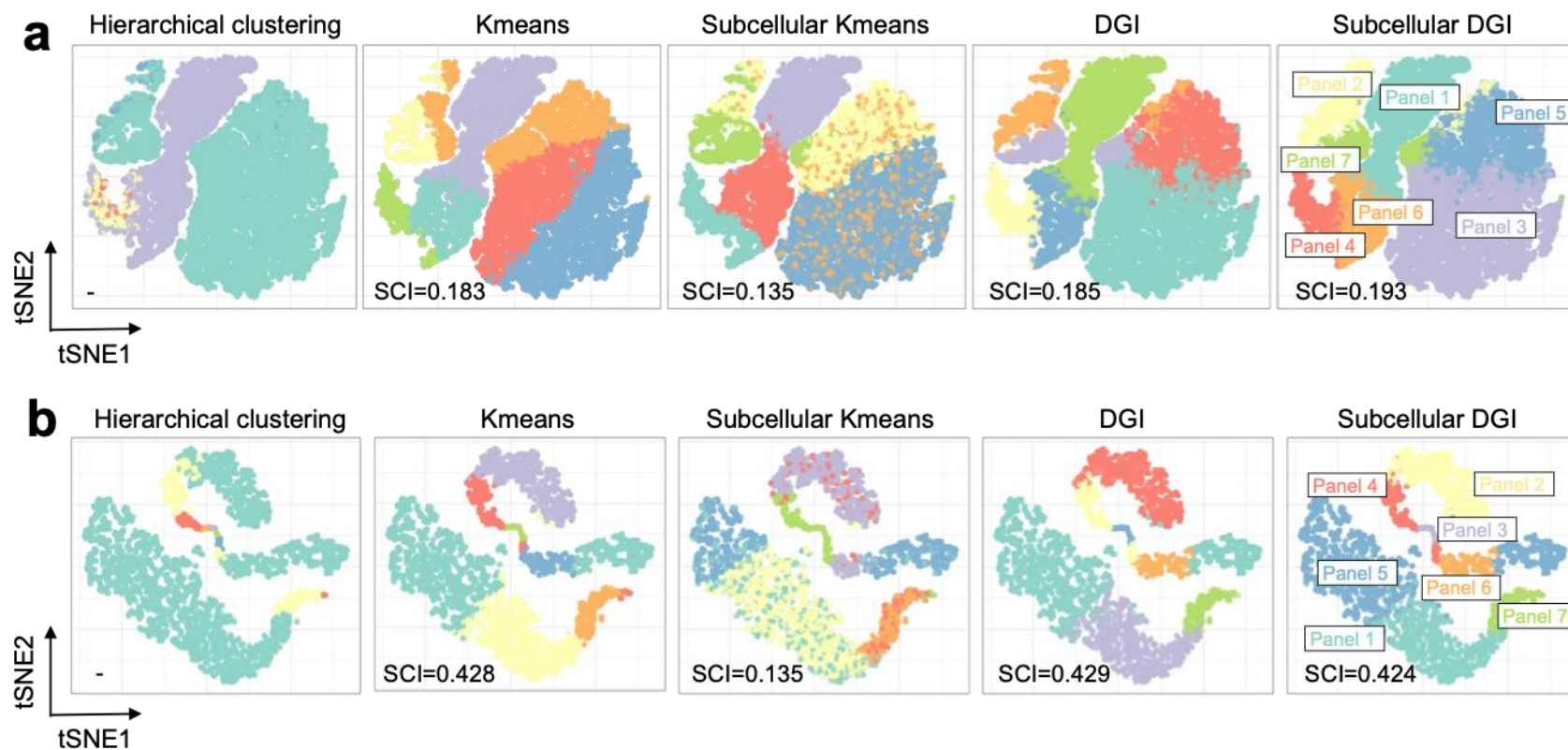
- 数据预处理后，我们获得了有效的**差异表达基因 (DEGs)**，其亚细胞定位信息和基因表达量信息通过特征工程建立为图结构的数据集。基因表达信息表征了细胞内调控网络，而亚细胞信息，包括信号肽和跨膜结构域，代表了细胞间的相互作用。
- 我们使用**Deep Graph Infomax (DGI)** 算法来进行无监督学习获得节点嵌入，然后对其进行聚类以获得**基因面板**。



# 结果

➤ 几何深度学习在基因面板识别中取得了卓越的效果

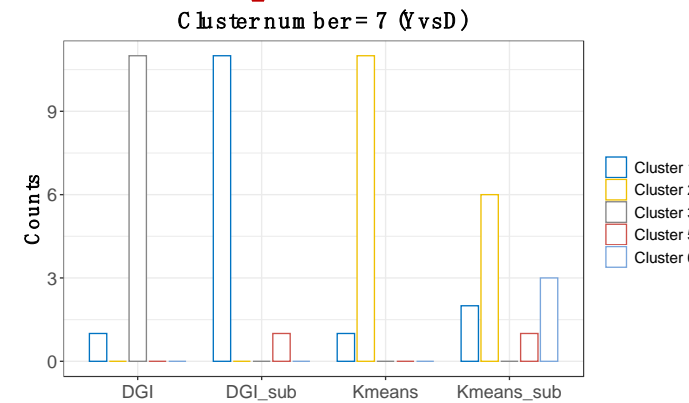
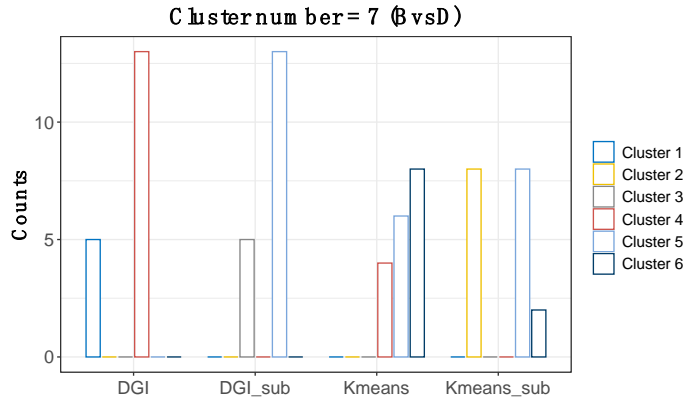
—— 数学验证：定性和定量



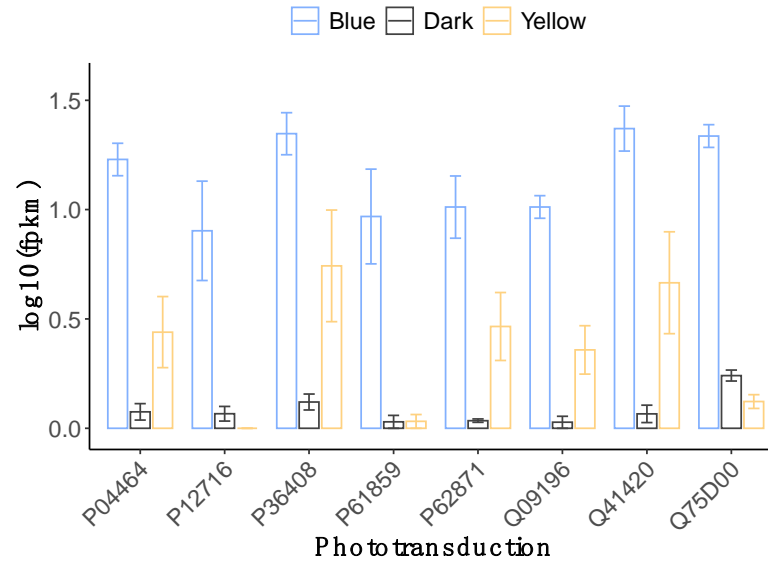
- 常用的方法，如层次聚类，表现出强烈的**偏差**，倾向于将大多数基因聚类为2个聚类。
- 通过DGI的无监督学习获得了基于基因表达和亚细胞信息的更清晰的基因面板，显示出其**整合异质信息的优越能力**。

# 结果

➤ 几何深度学习在基因面板识别中取得了卓越的效果

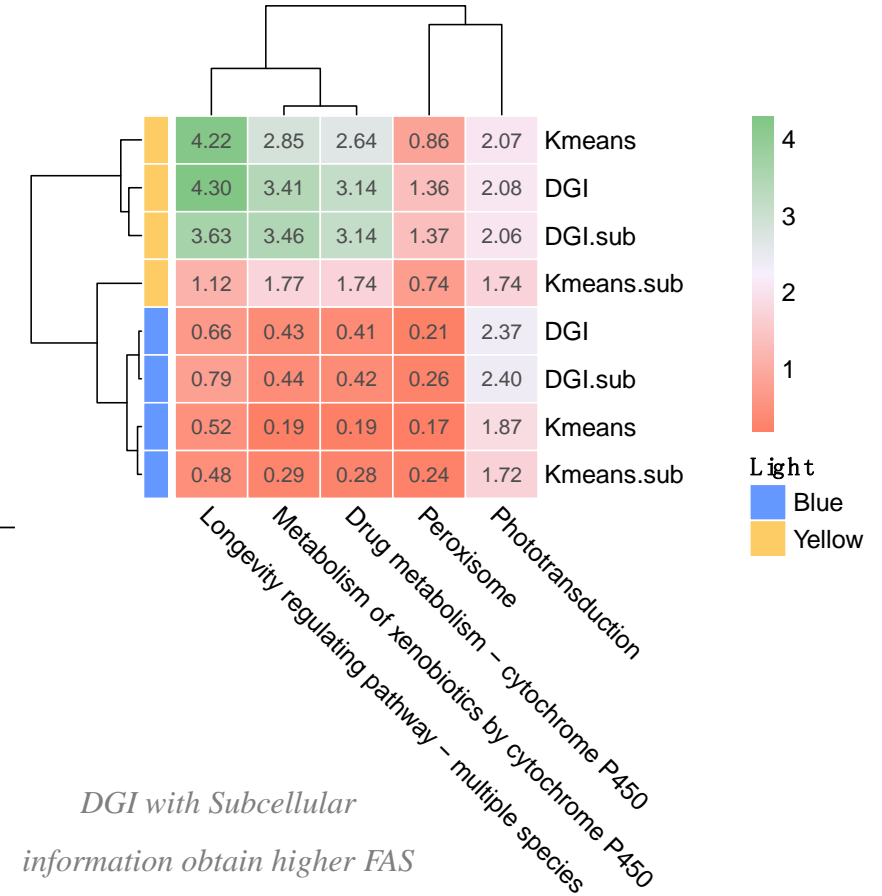


Assign to the same cluster



Co-expression pattern of phototransduction genes

—— 生物验证：生物功能一致性



➤ 功能分配得分 (FAS) 表明几何深度学习具有出色的生物功能匹配性能，这意味着它可以更好地识别某些功能的基因。



# 结果

## 通过中枢和信号基因面板中的不同代谢途径预测表型

### —— 工具软件1: 共调控基因面板用于新发现

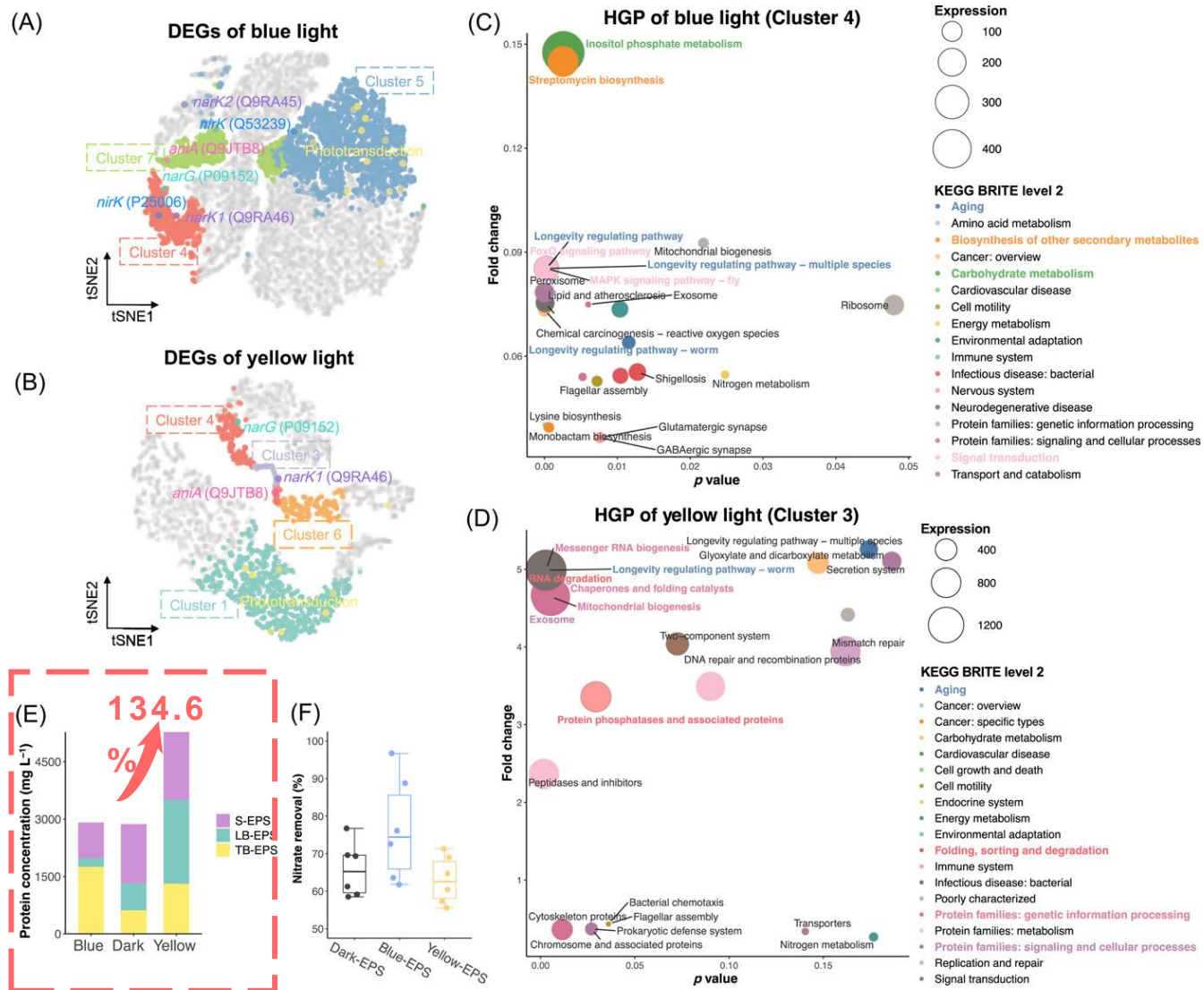
➤ **中枢基因面板 (HGP<sub>s</sub>)**: 以高表达水平和低倍数变化为特征的代谢途径, 对应于微生物群的集体行为 (表型)

➤ 对于黄光, 与**活性氧 (ROS)**、**肽和蛋白质合成**相关的途径被富集

➤ 对于蓝光, 与**信号转导**、**氧化应激**和**第二代代谢产物**相关的途径得到了丰富

➤ **分泌系统在蓝光和黄光下均上调**

✓ **湿实验验证: 蛋白质生产**





# 结果

## 通过中枢和信号基因面板中的不同代谢途径预测表型

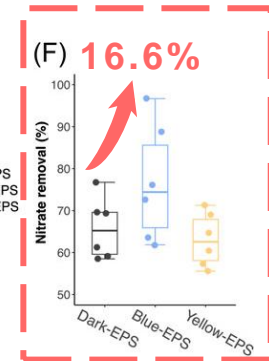
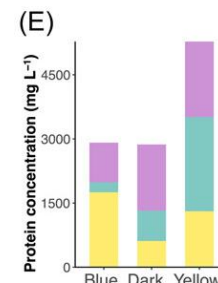
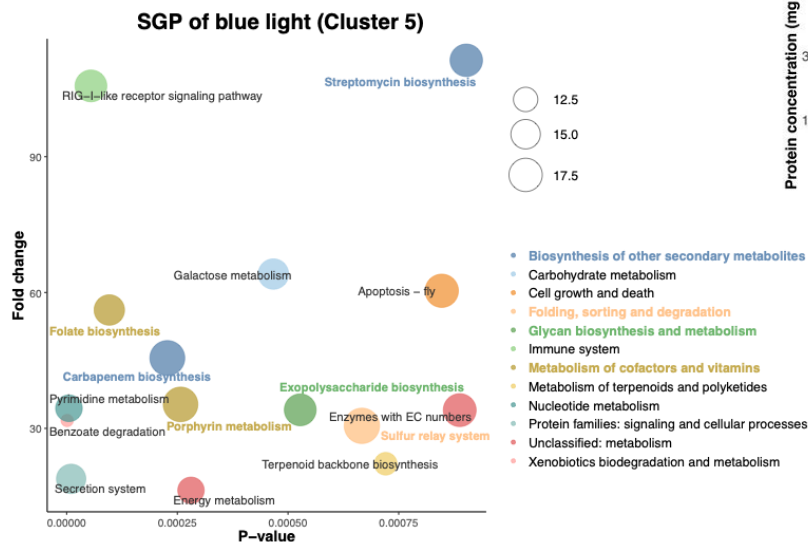
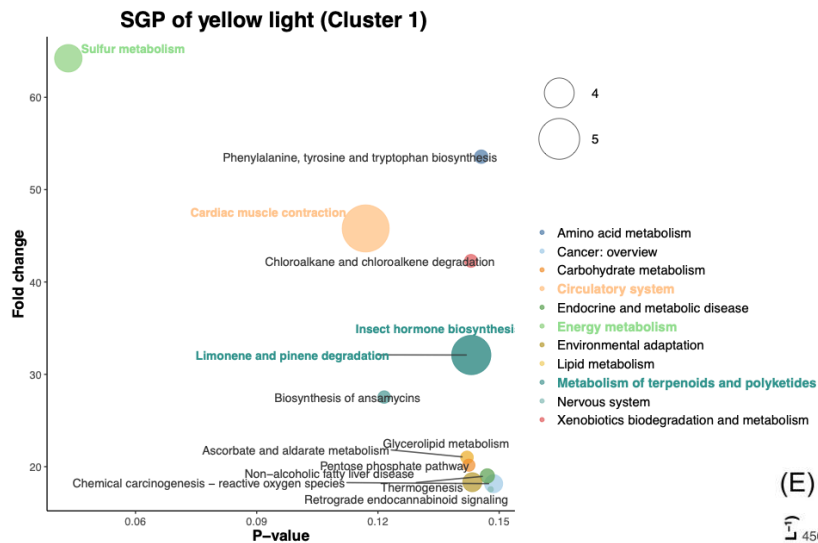
### —— 工具软件1: 共调控基因面板用于新发现

➢ 信号基因面板 (SGPs): 以低表达水平但高倍数变化为特征的代谢途径, 属于或类似于生物活性物质。

➢ 对于黄光, SGPs表现出**萜类和聚酮类**的高表达, 而**硫代谢**显著富集。

➢ 对于蓝光, SGPs表现出高表达的**辅因子、维生素**和其他次级代谢产物。这些都是多种生物活性物质。

✓ 湿实验验证: 胞外聚合物的生物催化效果





# 结果

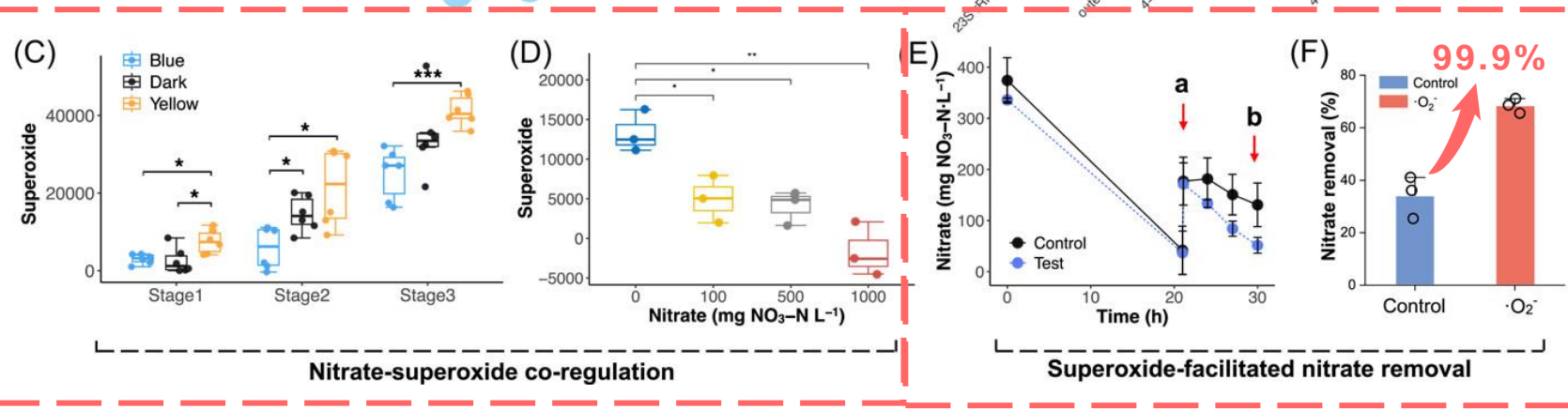
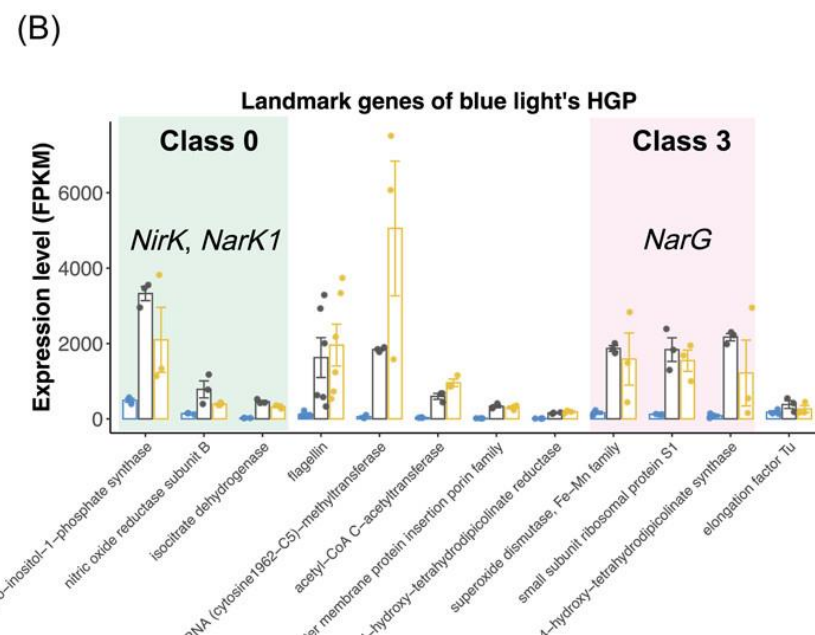
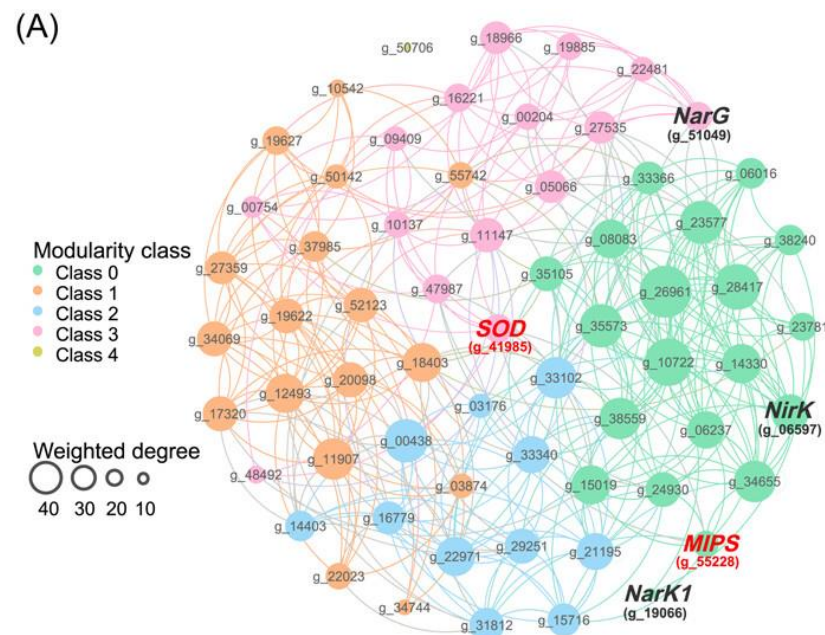
## 通过拓扑网络的代表性基因调节微生物组

### —— 工具软件2: 基于代表性基因的拓扑网络调控策略

案例1 - 蓝光: 超氧化物歧化酶 (SOD) 是微生物群落硝酸还原酶 (NarG) 的第一代表性基因。

湿实验验证: 新机理 & 新调控策略 —— 硝酸盐-超氧自由基共调控

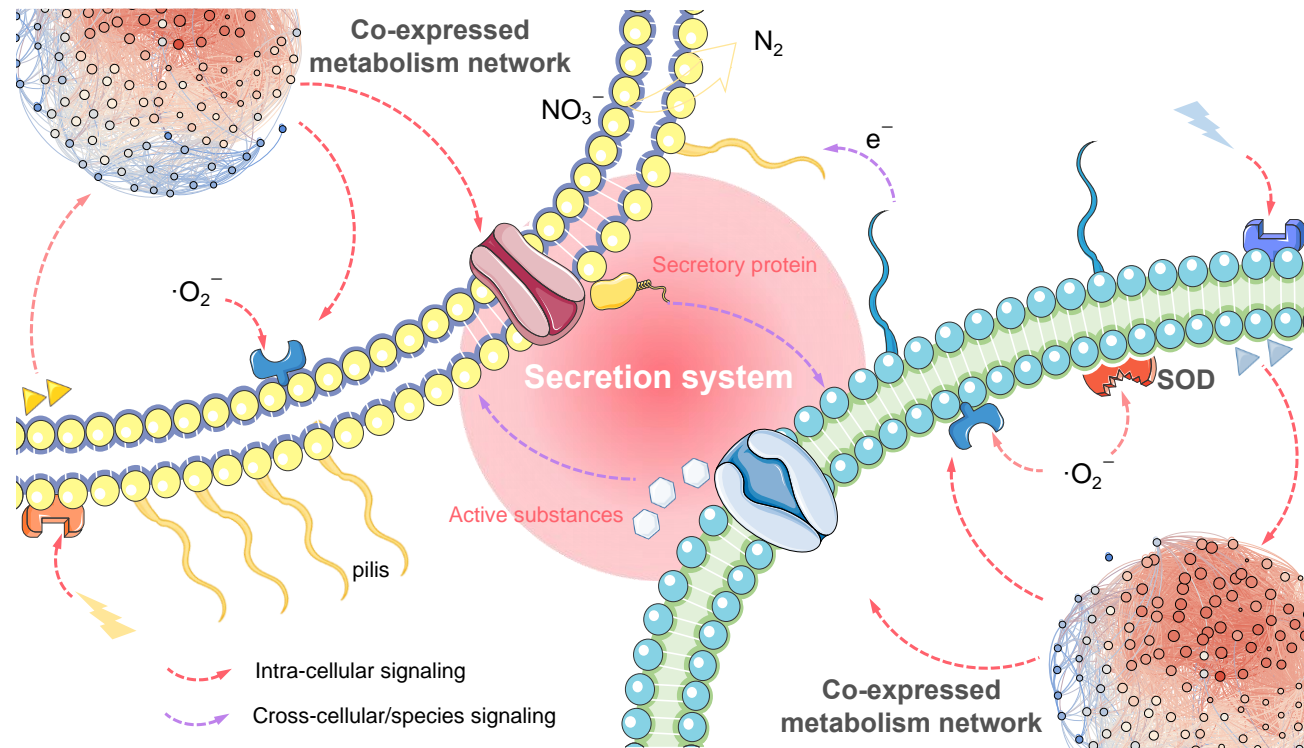
湿实验验证: 新型生物技术 —— 利用功能酶创造微刺激环境去除氮污染物





# 结果

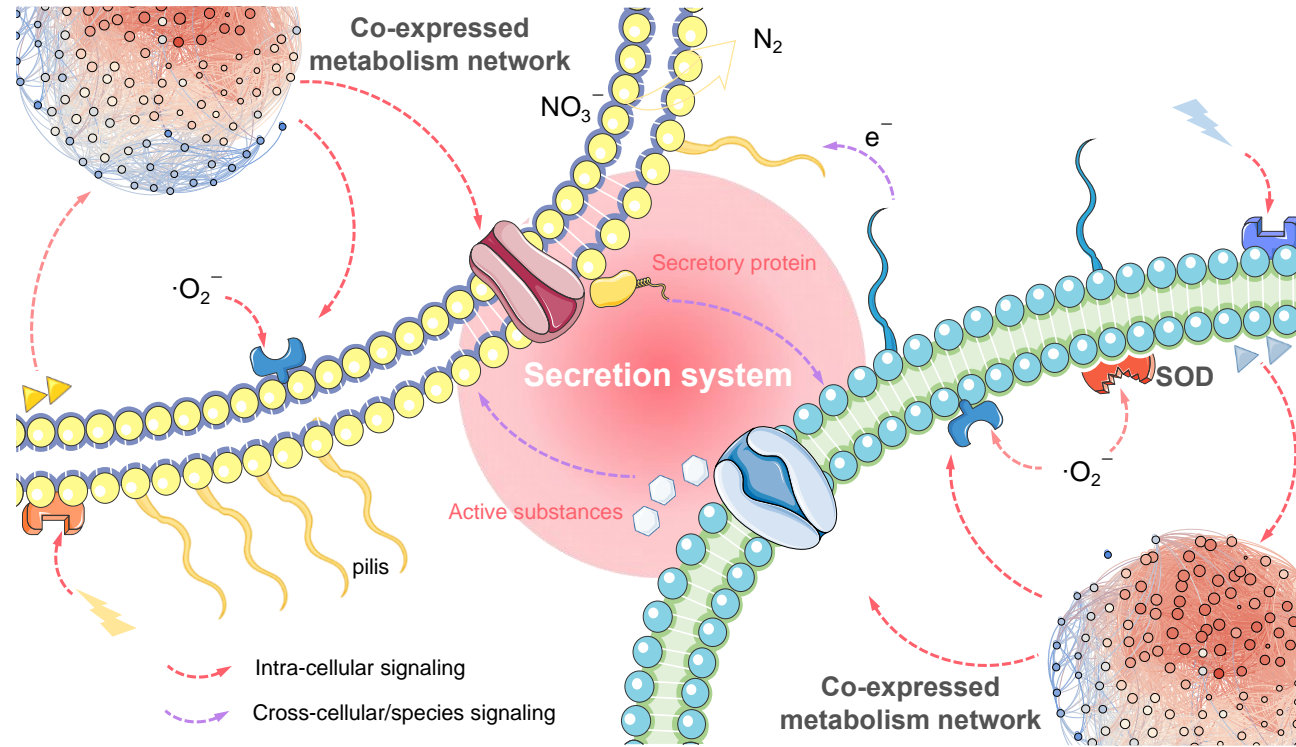
## 光控反硝化的机理与潜力



- 总体而言，**分泌系统**是跨物种相互作用的核心之一。
- 蓝光光感受器无处不在，并已在广泛的生物平台中实现，解释了蓝光在光反硝化中引发的**代谢通量分散**。合成了更多样化的代谢产物，通常是活性物质，如**辅助因子和维生素**。其中一些分泌分子对微生物群在光化学胁迫下维持体内平衡至关重要。这些代谢物是潜在的高价值资源，这为**利用硝酸盐作为底物**来节省高价值化学品生产成本的生物过程提供了线索。

# 结果

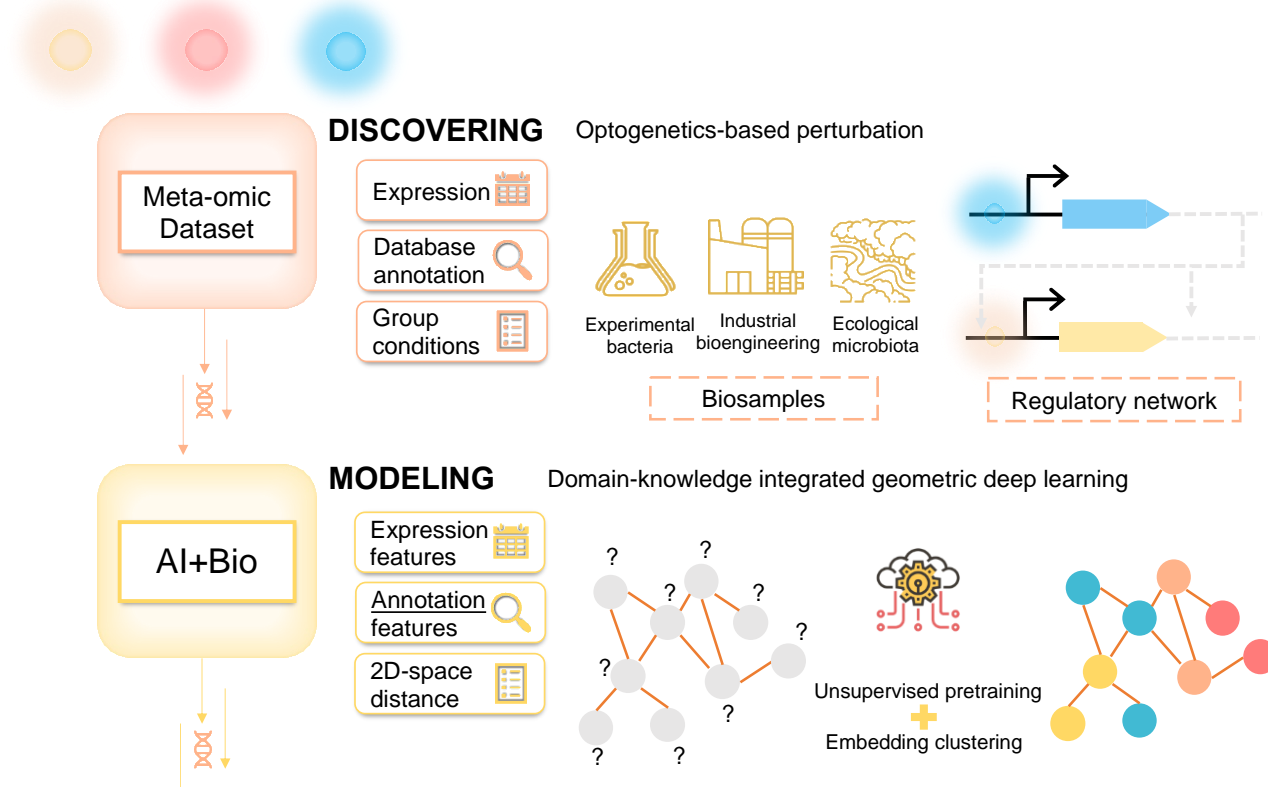
## ► 光控反硝化的机理与潜力



- 有趣的是，黄光具有**显著的激活效果**，而基于黄光的光遗传学开关鲜有报道，这可能是由于**跨物种的相互作用**。黄光触发的中心代谢通量主要用于蛋白质合成，尤其是**菌毛相关蛋白质**，这暗示着菌毛在加速种间电子转移以进行的集体功能代谢中起到了作用，例如黄光中枢基因面板中富集的**萜类和聚酮**代谢。



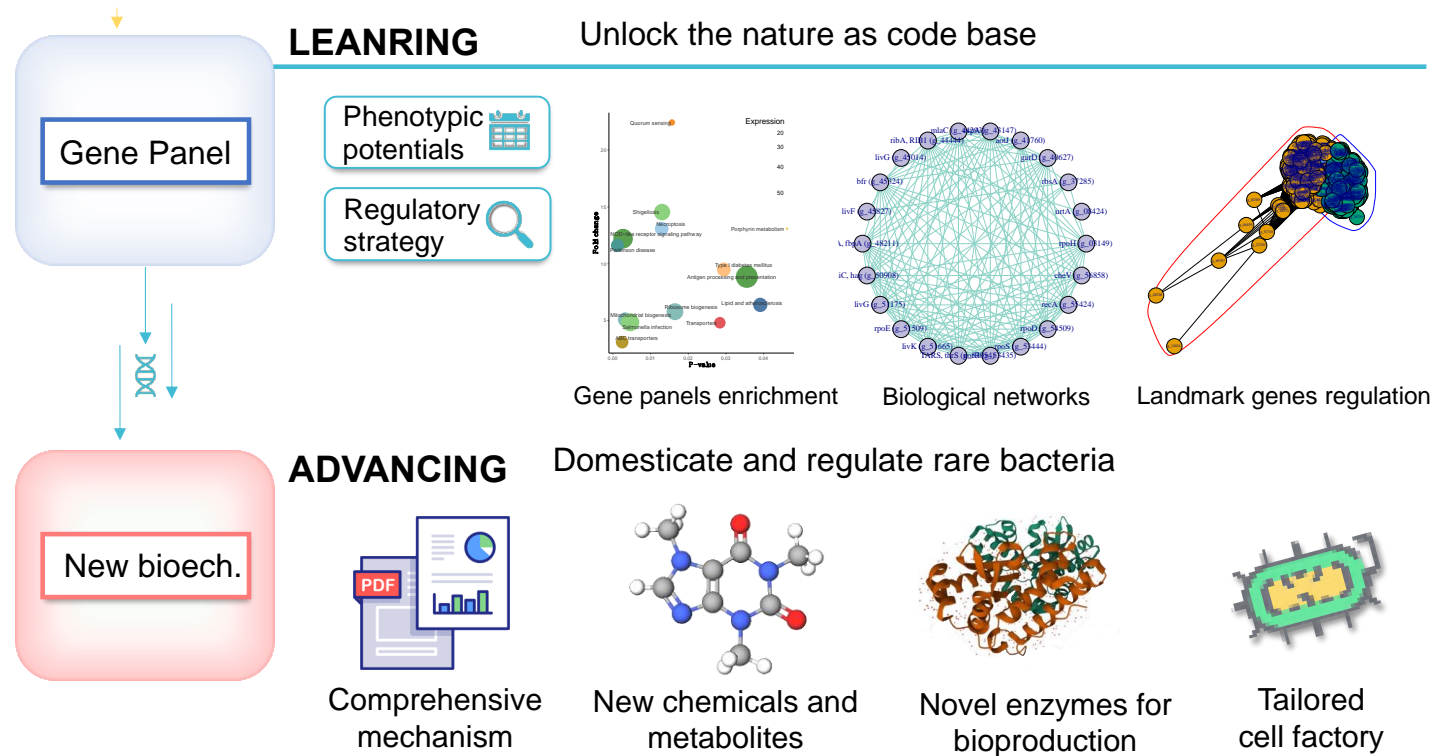
# 讨论



- 发现 (Discovering)：环境扰动是激活遗传过程以捕获动态调控网络所必需的。在这里，我们使用**基于光遗传学的遗传扰动**来触发通过光遗传学开关的信号传输。光遗传学诱导的动态调控网络变化包含在输入数据集中，包括表达矩阵、数据库注释和组条件。
- 建模 (Modeling)：生物拓扑原理由模型自动捕获。包括**表达水平**和**亚细胞注释**的生物大数据，通过特征工程构建为节点特征，并将二维隐空间中的基因表达距离作为边缘。异构知识通过无监督预训练获得嵌入后，再通过聚类获得共表达基因面板。



# 讨论



- 学习 (Learning)：应用软件套装释放自然代码库的潜力。基于基因面板，富集的途径可以揭示未知的表型。以标志性基因为特征的生物网络是阐明生物学机制和制定调控策略的有力工具。
- 迭代 (Advancing)：基于上述过程，新的生物技术可以用于驯化和调控稀有微生物，提供多方面的应用，包括综合机理报告、新化学物质和代谢产物的发现、开发用于生物生产的新型酶、合成生物学个性化细胞工厂等。



“iMeta”是由威立、肠菌分会和本领域数百位华人科学家合作出版的开放获取期刊，主编由中科院微生物所刘双江研究员和荷兰格罗宁根大学傅静远教授共同担任。目的是发表原创研究、方法和综述以促进宏基因组学、微生物组和生物信息学发展。目标是发表前10%(IF > 15)的高影响力论文。期刊特色包括视频投稿、可重复分析、图片打磨、青年编委、前3年免出版费、50万用户的社交媒体宣传等。2022年的三月、六月和九月期已正式在线出版发行!



主页: <http://www.imeta.science>

出版社: <https://wileyonlinelibrary.com/journal/imeta>



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



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



[iMeta](#)

[宣传片](#)

