

# 转录调控网络基因表达与转录因子 活性

专业：生物化学与分子生物学

学号：2017304110110

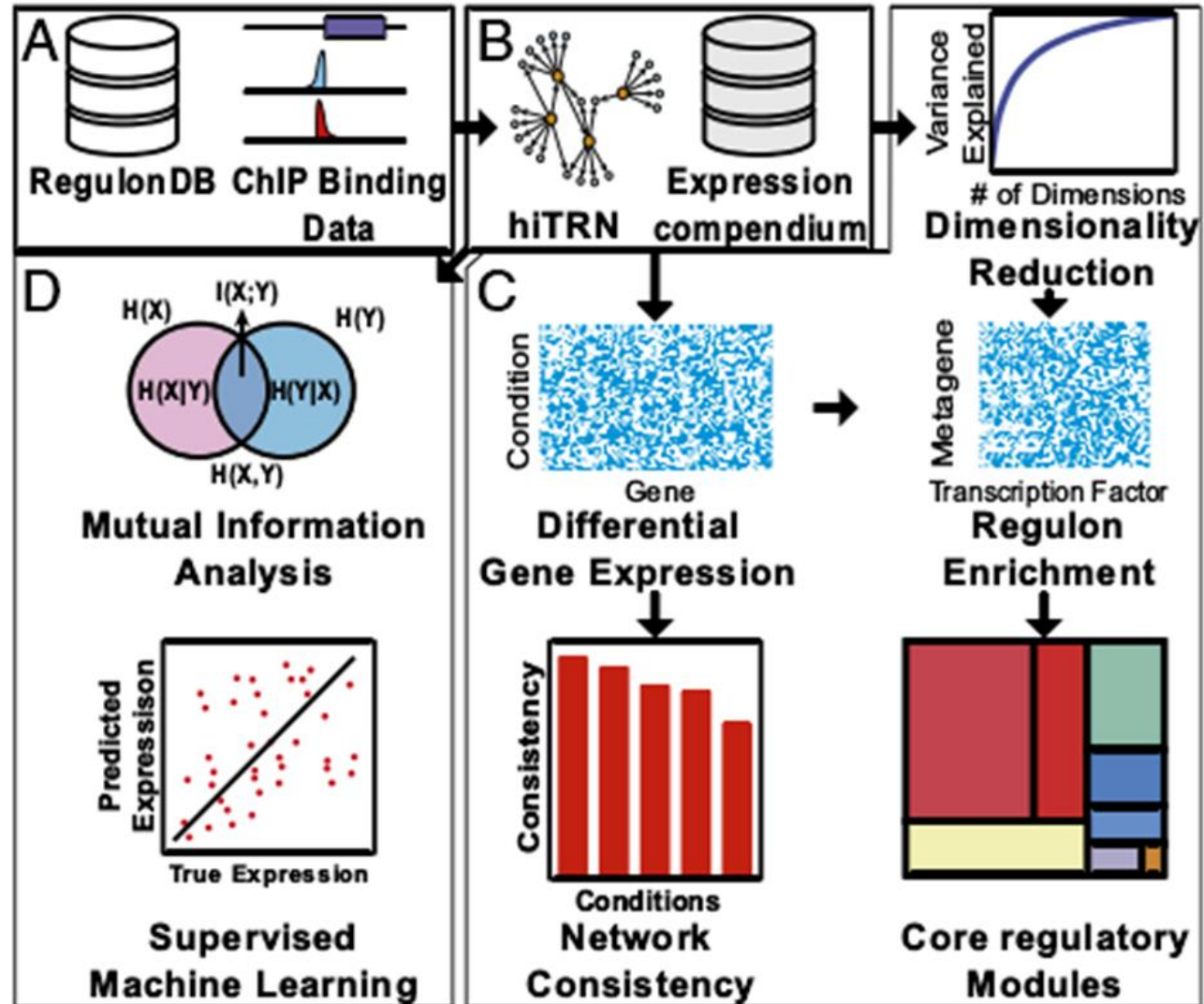
姓名：张昕冉

导师：周道绣

# Background

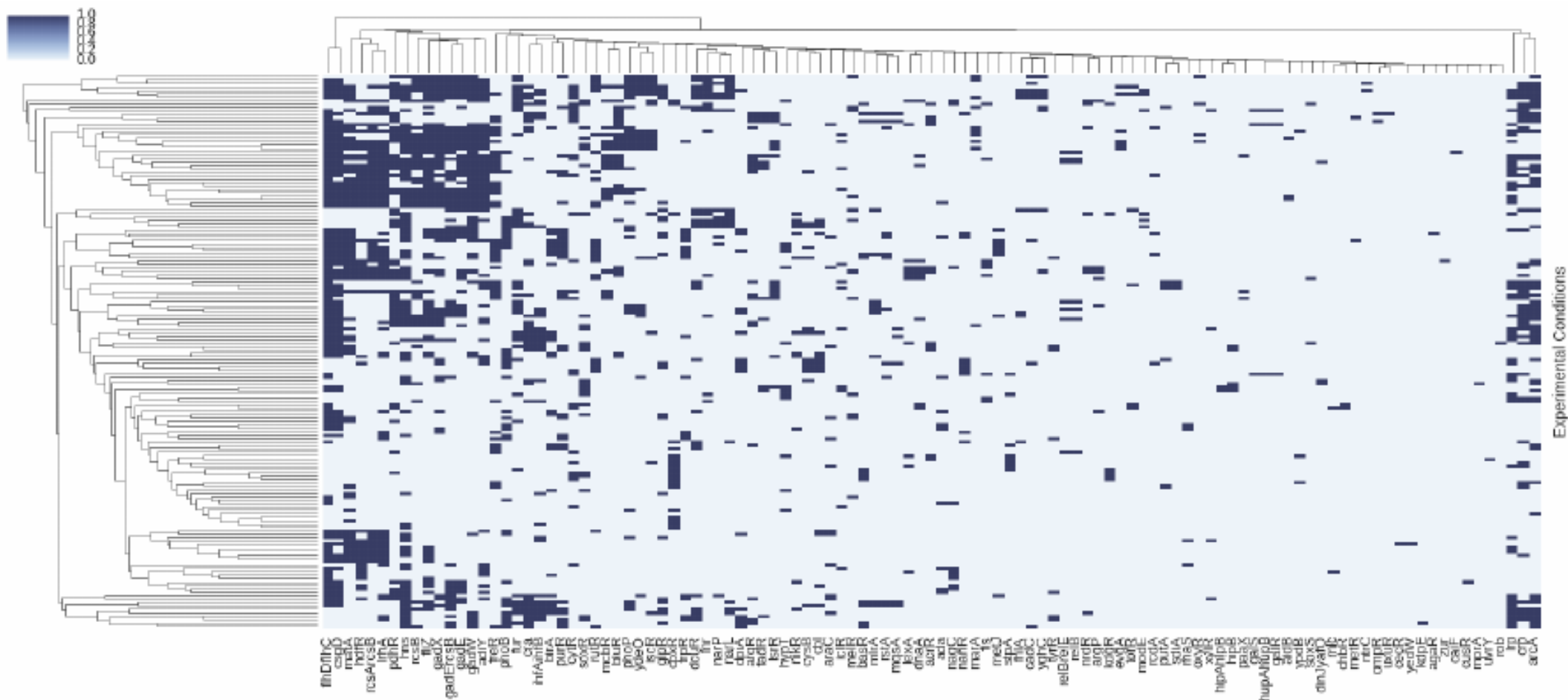
- 转录调控网络（TRN）在生物体内能够调节数千个基因的表达，从而在应对环境变化和遗传物质方面发挥重要作用。在所有生物体中大肠杆菌的TRN可能是最需要广泛研究的。然而，这种TRN的结构仍然具有许多不确定性，严重限制了其用于预测基因表达或解释不同数据集的作用。事实上，十多年前，大肠杆菌的代谢和调控网络模型的组合在缺氧环境变化中只能解释基因差异性表达的15%。从那以后，虽然取得了很大进展，但预测基因表达依然存在根本挑战。
- TRN可以由从数据源中确定的调节交互作用组成。这些包括直接和间接的实验证据或计算预测。对于后者，减少假阳性相互作用仍然是挑战性，而现有技术达到60%的精度。近年来，改进的染色质免疫沉淀（ChIP）方法已经能够精确表示转录因子（TF）结合位点。对TF KO菌株使用ChIP，从而产生了高置信度的调节相互作用。目前已经有十多个主要的TFs已经开始研究这种知识产权研究，每项研究都将已知的TF结合位点数量增加了74-400%。

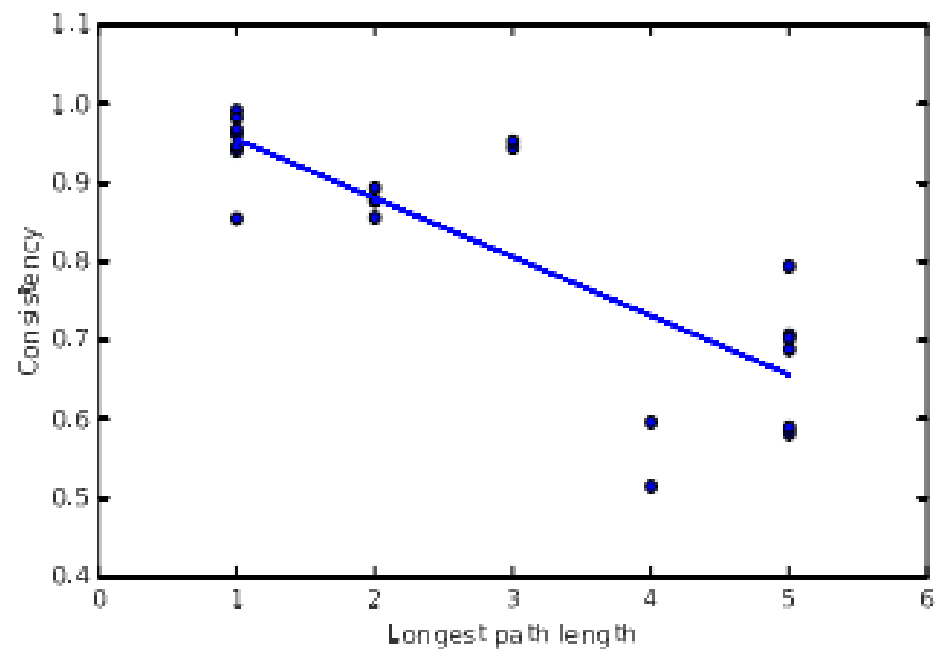
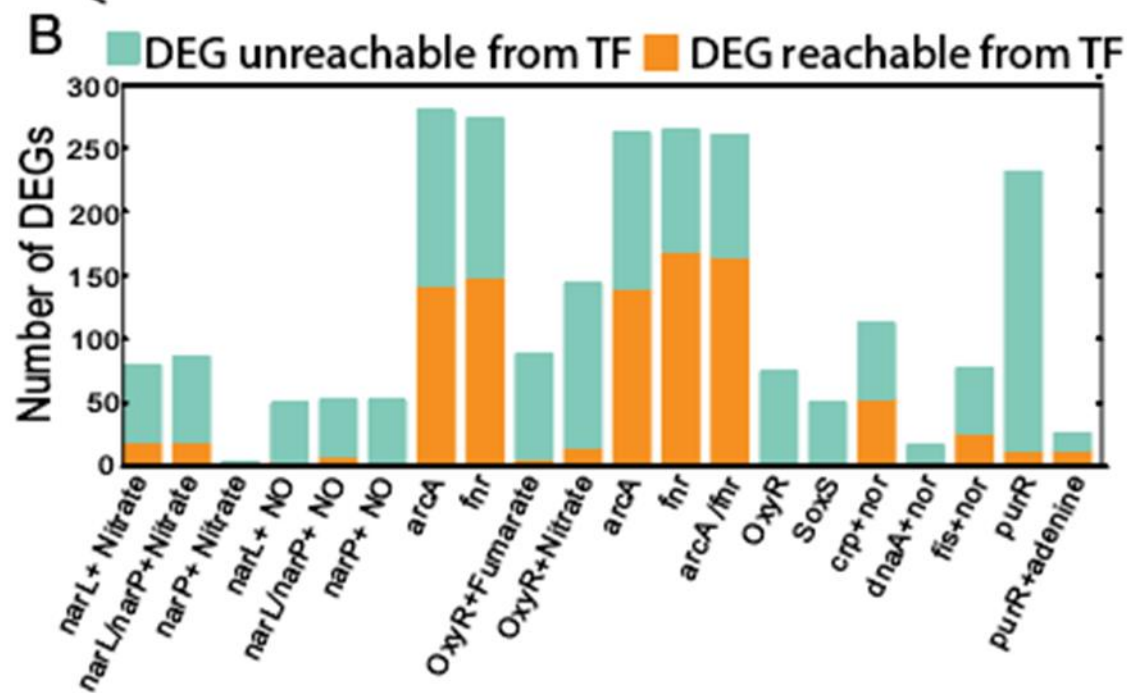
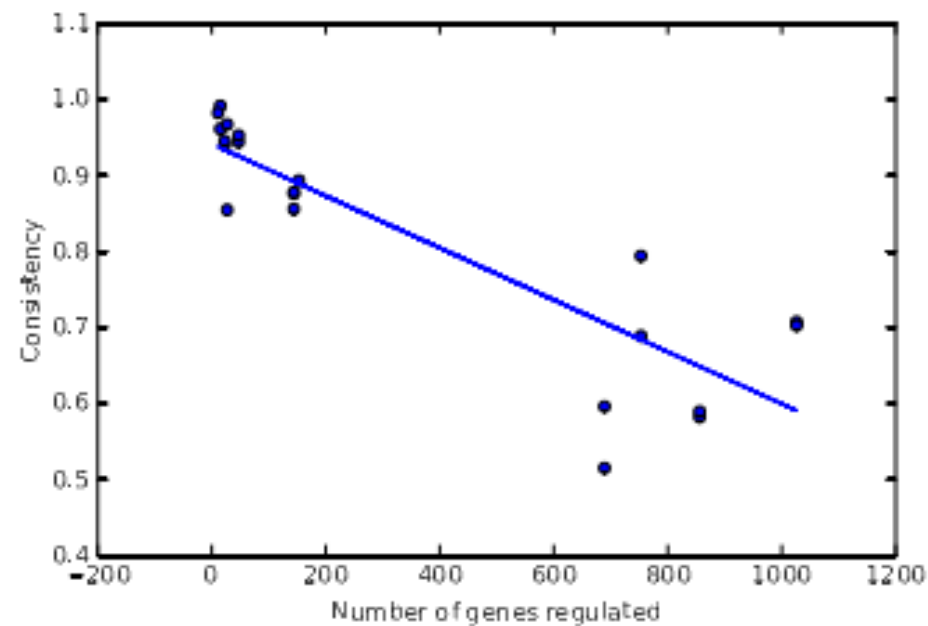
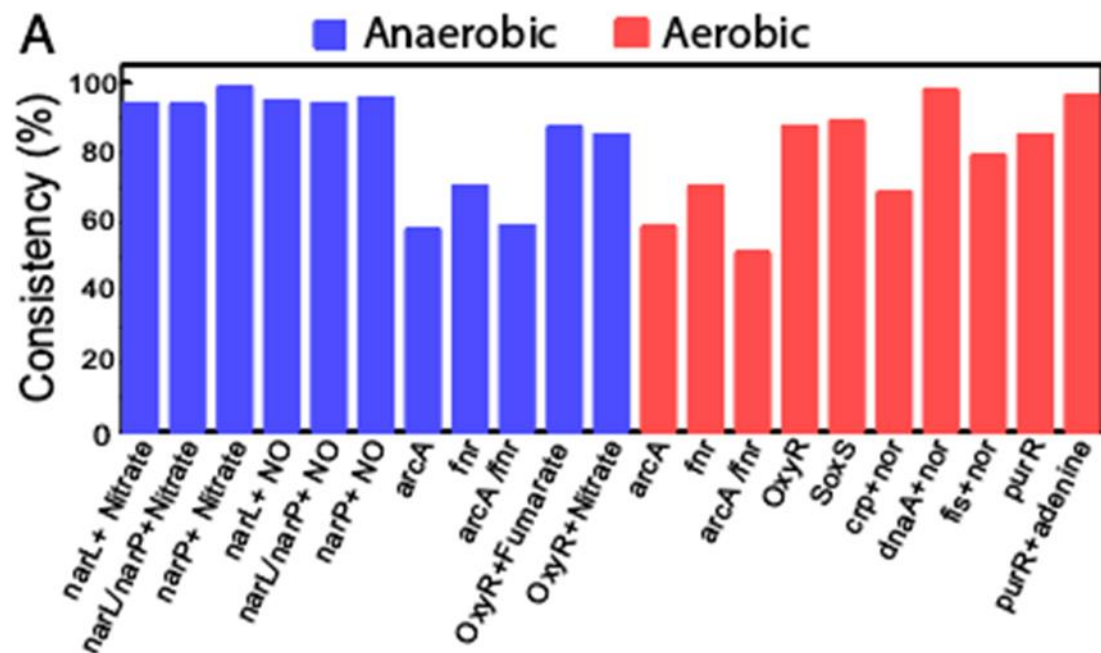
# Overview of workflow



# TRN的覆盖范围已经扩大，但仍然不完整。

154个实验条件下，1764个基因表达的变化





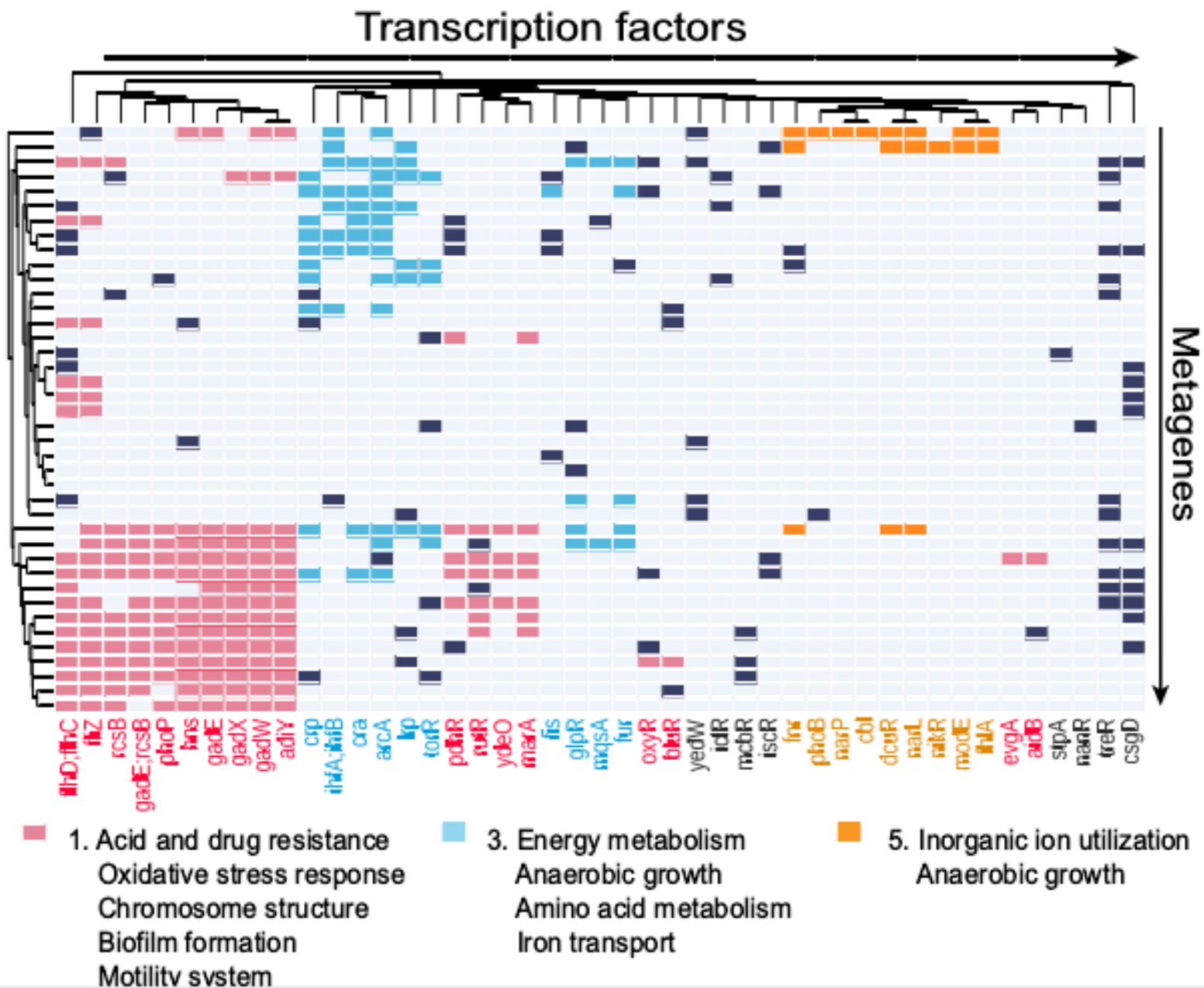
总体而言，对DEG的hi TRN的覆盖范围因实验条件而异，平均覆盖率为26%，这是显著的。单个TF KO实验的低DEG覆盖率反映了许多调控子是高度互作的。因此，实现100%DEG覆盖可能需要精确重建单个调控子，包括超越TRN拓扑结构和现有的调控偏差机制。

# hi TRN与整个转录组的主要变化模式一致的。

在由4,189个基因×441个样品组成的转录组学数据的背景下评估了hi TRN。转录组学数据难以解释，部分原因是它们是高维和嘈杂的。因此，通过使用非负矩阵因子分解（NMF）来识别跨条件转录组变化的主要模式。NMF通过将数千个基因减少到几十个元素，从复杂表达数据集中识别出粘连子系统。

**元基因**是其表达变化在条件之间相关的基因的线性组合。使用NMF将表达数据从4,189个基因×441个样本降低到40个元素×441个样本。

**调控子**：每个元基因中被确定为主要贡献者的基因，这些基因被富集成调控子。



他们发现所有的元基因都被富集至少一个调控子



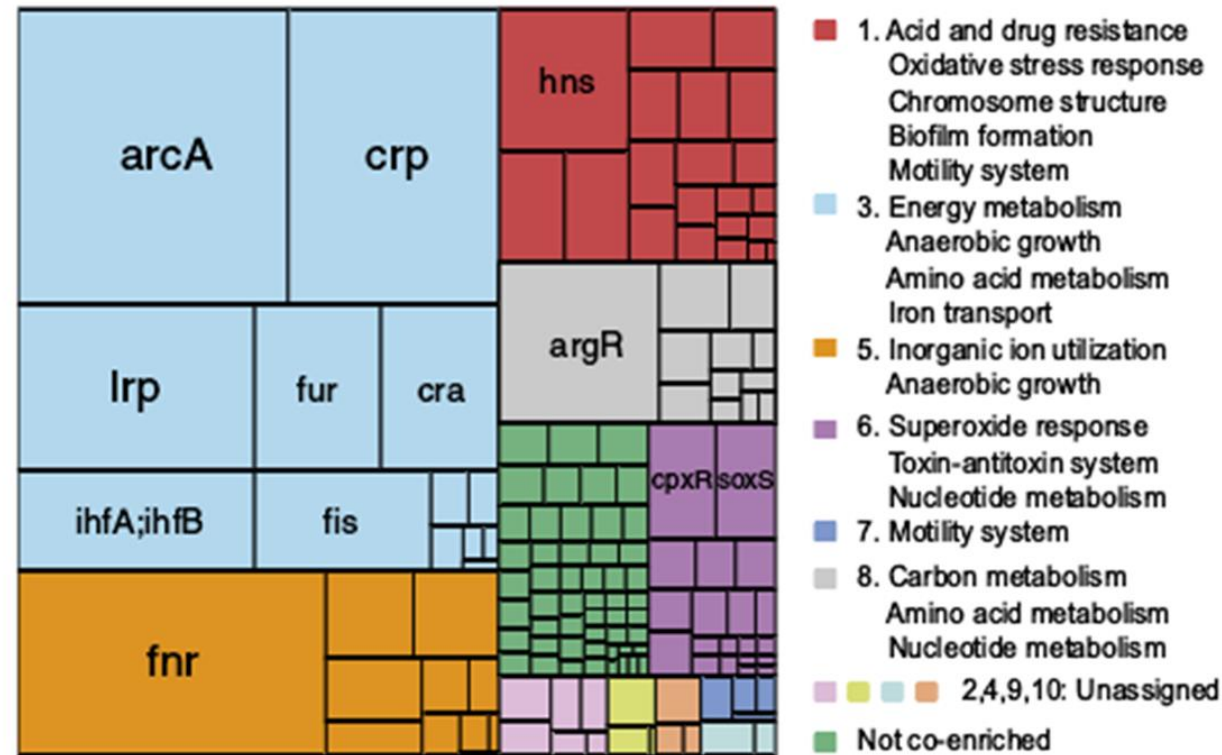
元基因倾向于共享相关功能的hi TRN调控子

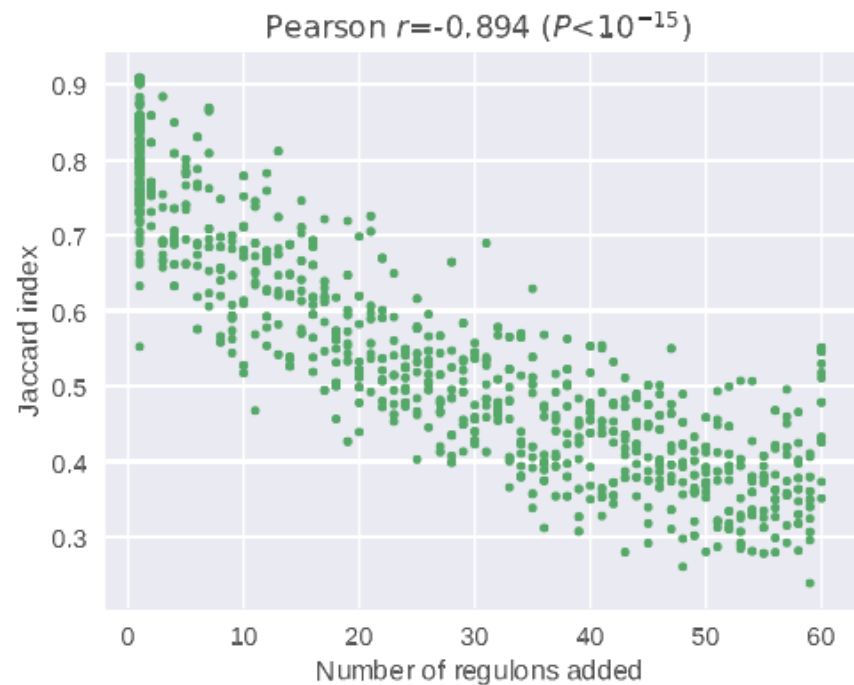
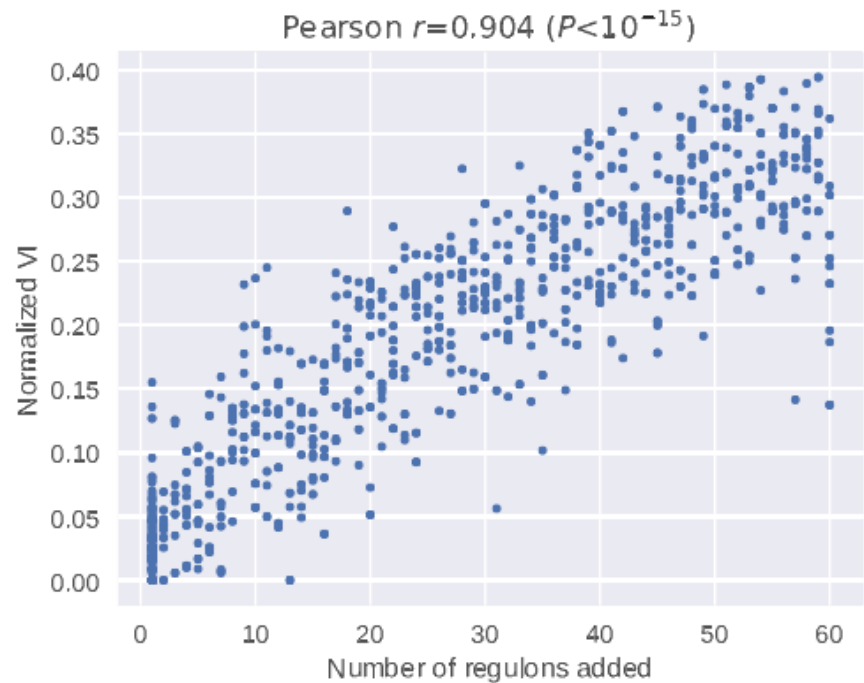


这个结果表明了hi TRN的覆盖率与一致性以及高维度转录数据的一致性



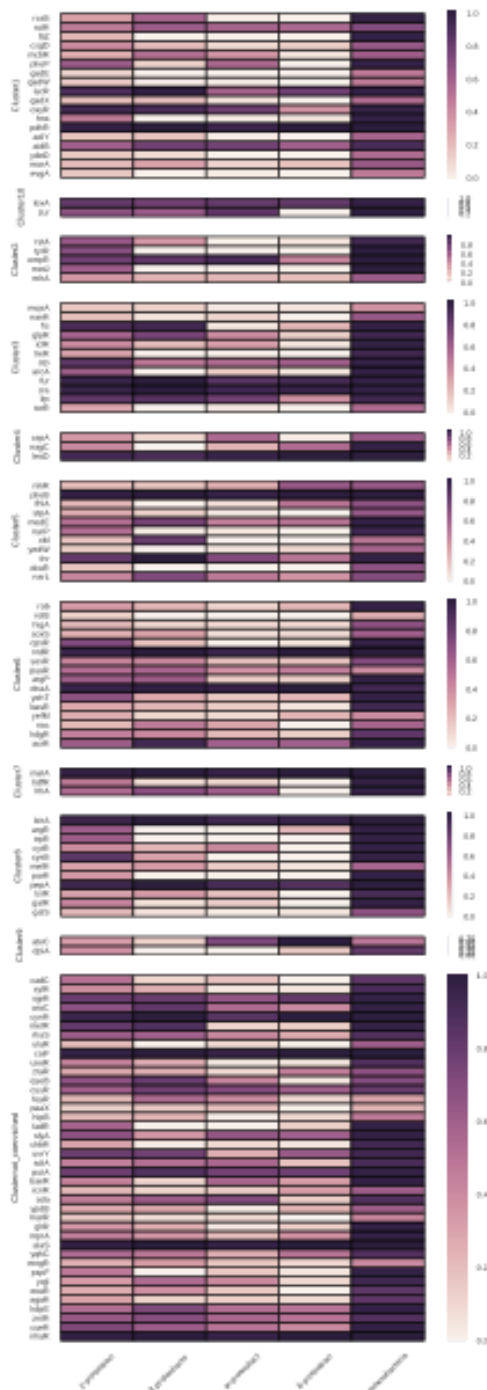
强大的调节模式被识别。接下来，他们评估了在hi TRN中TFs的组织结构是否与从转录组得到的调控子的功能组织一致。因此，他们开发了一个计算流程，以识别跨条件显著和强烈地融合的TF簇。他们确定了**10个集中模块**。六个模块特别代表核心生物功能。例如，模块6包括与多种应激反应相关的TF和毒素 - 抗毒素对。



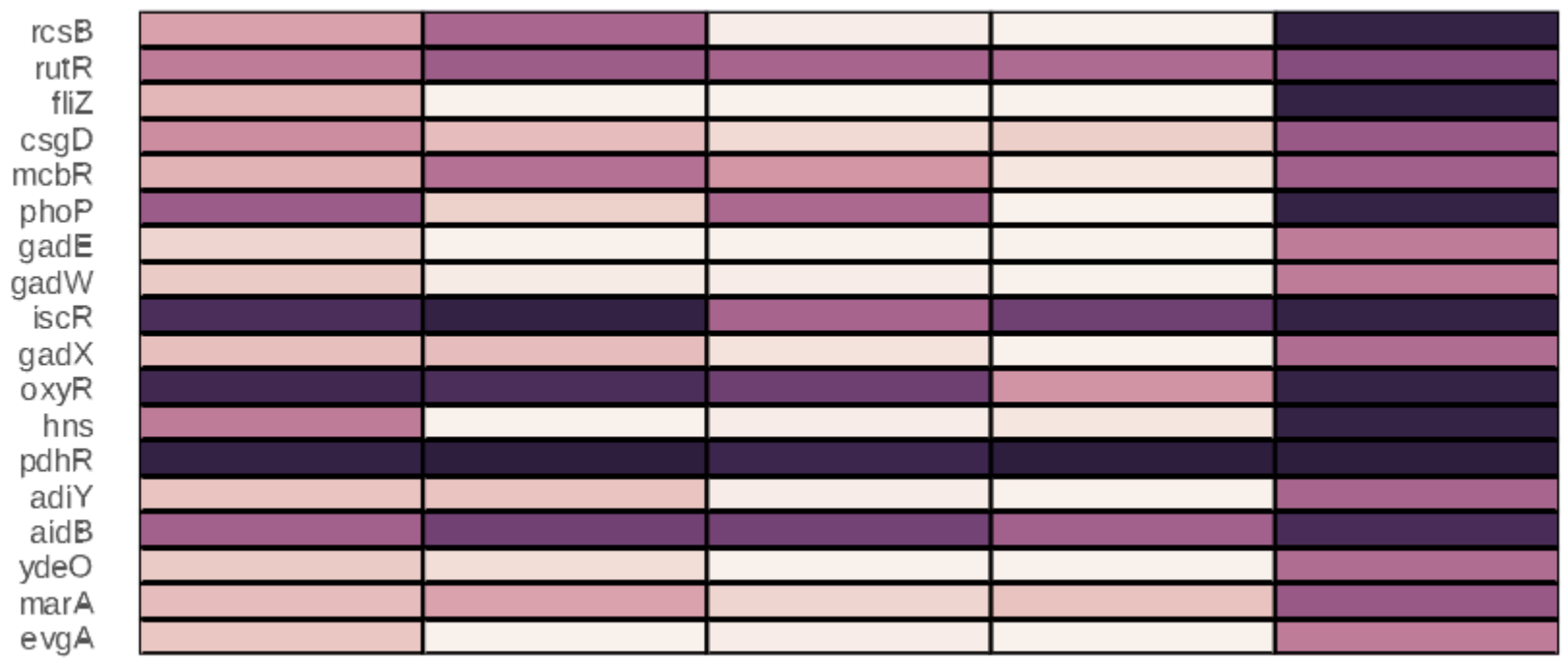


由于大肠杆菌TRN仍在扩大，他们评估了这些模块是否保持稳定，因为新的调节互作被纳入TRN。随机添加了60个调控子，并使用他们的方式来识别新模块，与原始模块进行了比较。模块之间的平均Jaccard指数范围为0.34至0.81，归一化变化范围为0.025至0.34。

接下来，他们检查了在DNA序列和蛋白质结构水平上调节模块的进化特征。假设与生命功能相关的模块将在物种间保守，而生物特异性反应则不会。接着计算了在肠杆菌科和 $\gamma$ -， $\beta$ -， $\alpha$ -和 $\delta$ -变形杆菌中hi TRN的147个TFs的保守程度。

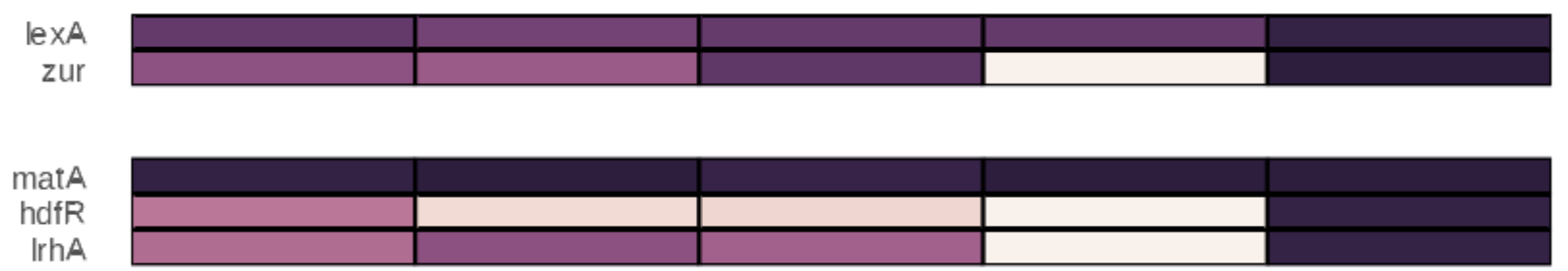


Cluster1

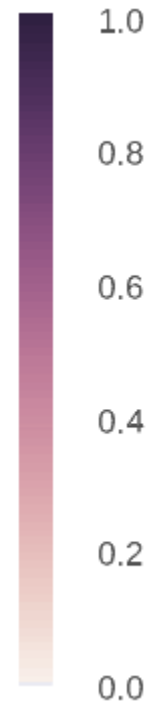


涉及不同变化的应激反应，如酸胁迫反应

Cluster7 Cluster10



涉及动力，金属离子摄取和DNA损伤响应



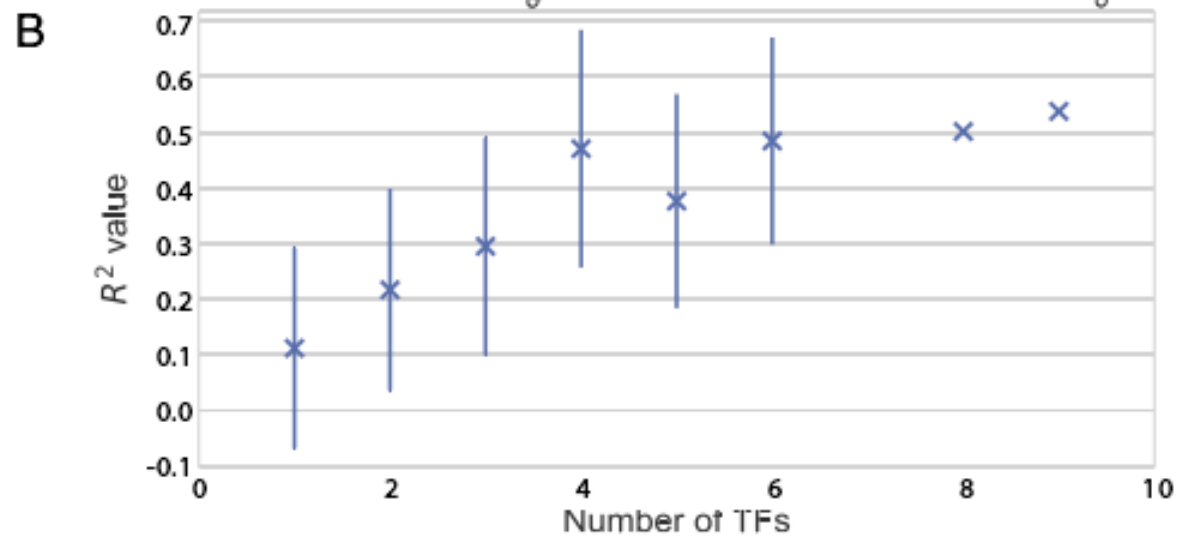
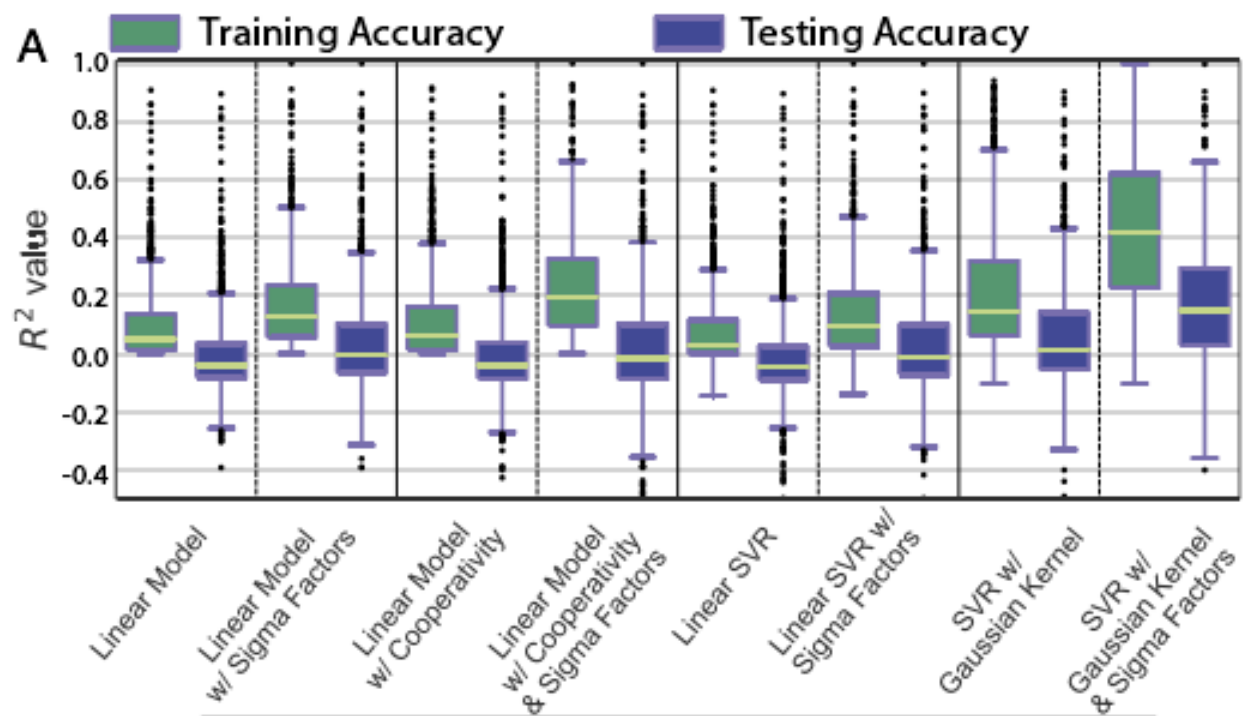
他们调查了147 TFs的DNA结合基序的序列同源性。他们发现模块1,2,3,8和9中的TF与其他模块中的模块相比具有更相似的结合基序。

在蛋白质结构上，他们探讨了每个模块中的TF之间的结构相似性，以确定模块之间是否存在结构 - 功能关系。他们使用FATCAT结构对准工具对DNA结合结构域和TF对的完整结构进行了对齐注释。研究发现，模块4,5和8中的TF在两种情况下显示出显着更高的结构相似性。例如，模块8集中在周质结合蛋白样I结构域。

大多数TU的表达可以从TF表达定量预测。

他们查看了基因表达是否可以定量地预测在不同条件下TF表达水平的功能和遗传扰动。

探索了八个模型结构，以确定最佳建模过程：四个多元线性回归模型和四个支持向量回归（SVR）



多变量线性回归。我们使用多元线性回归来预测来自Regulon DB的TU而组成的基因的平均表达。



非线性作用。为了更好地考虑非线性调节相互作用，无论是否具有 $\sigma$ 因子，他们接下来探寻了具有线性和高斯内核的SVR。



当我们更多地了解TRN的结构时，TRN的预测能力应该会增加。

# 结论

- 1.结论： 转录调控网络说明了基因差异性表达与转录因子的活性之间存在许多的因果关系，我们可以利用这些预测86%的转录单位的表达。
- 2.启发： 本文是在针对大肠杆菌等微生物进行实验，这种转录调控网络在水稻中是否也存在？
- 3.存在的问题： 还有许多TRN的功能没有研究清楚,DEG覆盖率还没达到100%。



谢谢