



# A system-level model for the microbial regulatory genome

Aaron N Brooks<sup>1,2,†</sup>, David J Reiss<sup>1,\*†</sup>, Antoine Allard<sup>3</sup>, Wei-Ju Wu<sup>1</sup>, Diego M Salvanha<sup>1,4</sup>, Christopher L Plaisier<sup>1</sup>, Sriram Chandrasekaran<sup>1</sup>, Min Pan<sup>1</sup>, Amardeep Kaur<sup>1</sup> & Nitin S Baliga<sup>1,2,5,6,\*\*</sup>

一个系统水平上的微生物调控基因组模型

黄志韬 孙霄鹏

# Abstract

## 作者感兴趣的问题

与真核生物不同，微生物所拥有的遗传信息十分“简练”，只有很小的基因组和比较少的调控元件。那么如何在不同的环境下调控自身的生理状态？

## 作者所做的工作

通过已有的数据构建一个模型来捕捉动态的环境与基因组编码调节程序间的互作。并通过一些已有的数据和自己的一部分实验来验证模型的准确性

# 作者取得的一些成就

- 1、顺式作用元件对全基因组范围内的基因调控的贡献，以及受到特定环境影响的转录因子对那些环境特异性应答的效用。
- 2、部分抛弃了原有的调节子和操纵子的概念，定义“核”的概念并归类了不同环境下基因与基因间的关联。
- 3、利用模型捕获不同的环境条件下对细菌适应性影响协同的基因，并研究它们受整个基因组的调控机制
- 4、提供了**EGRIN2.0**模型的网络服务，帮助促进合成生物学的发展。

# 介绍词汇简写

**GRNs:**基因调控网络

**GREs:**基因调控元件

**TF:**转录因子

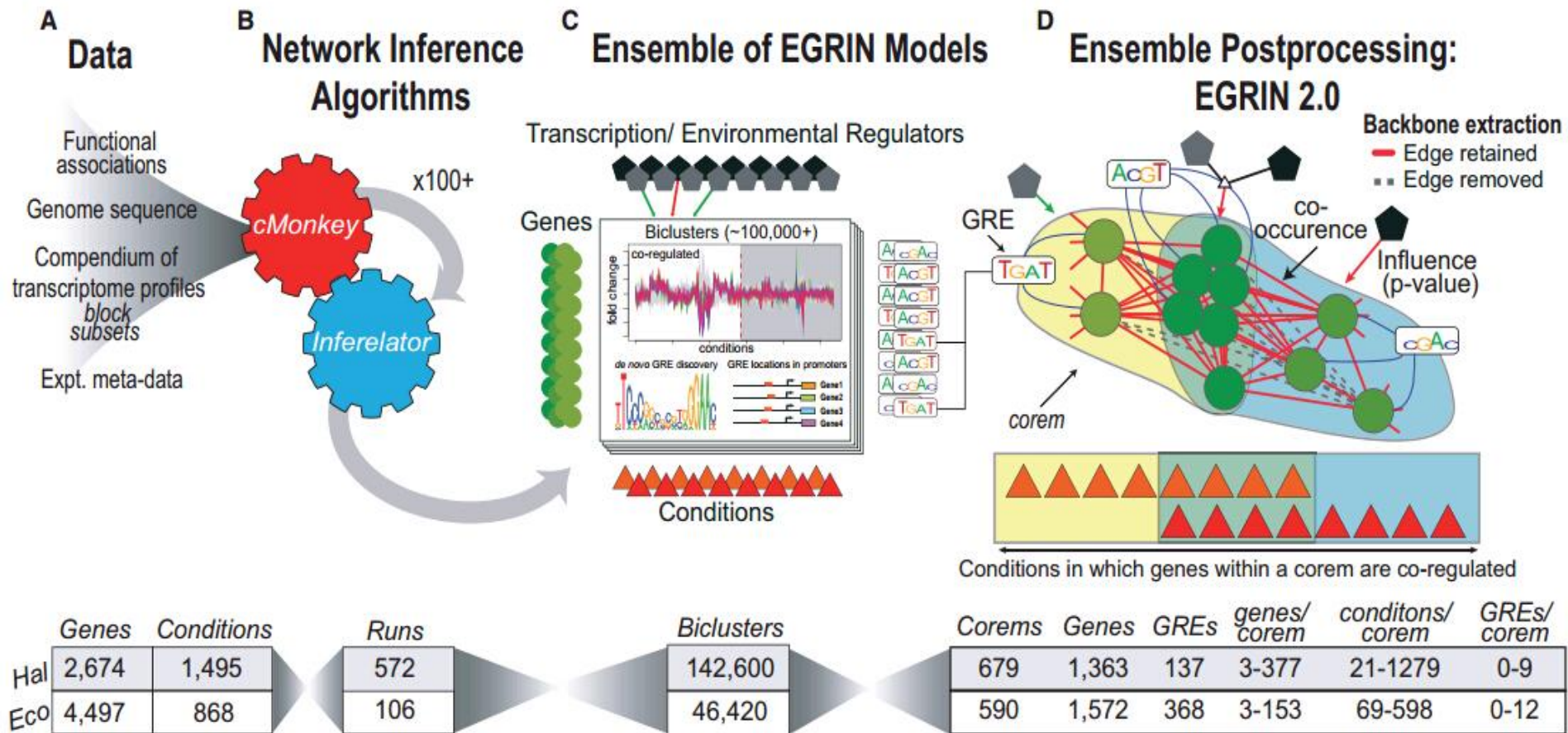
**EGRIN:**环境和基因调控影响网络

**Corems:**共调控基因重叠位点

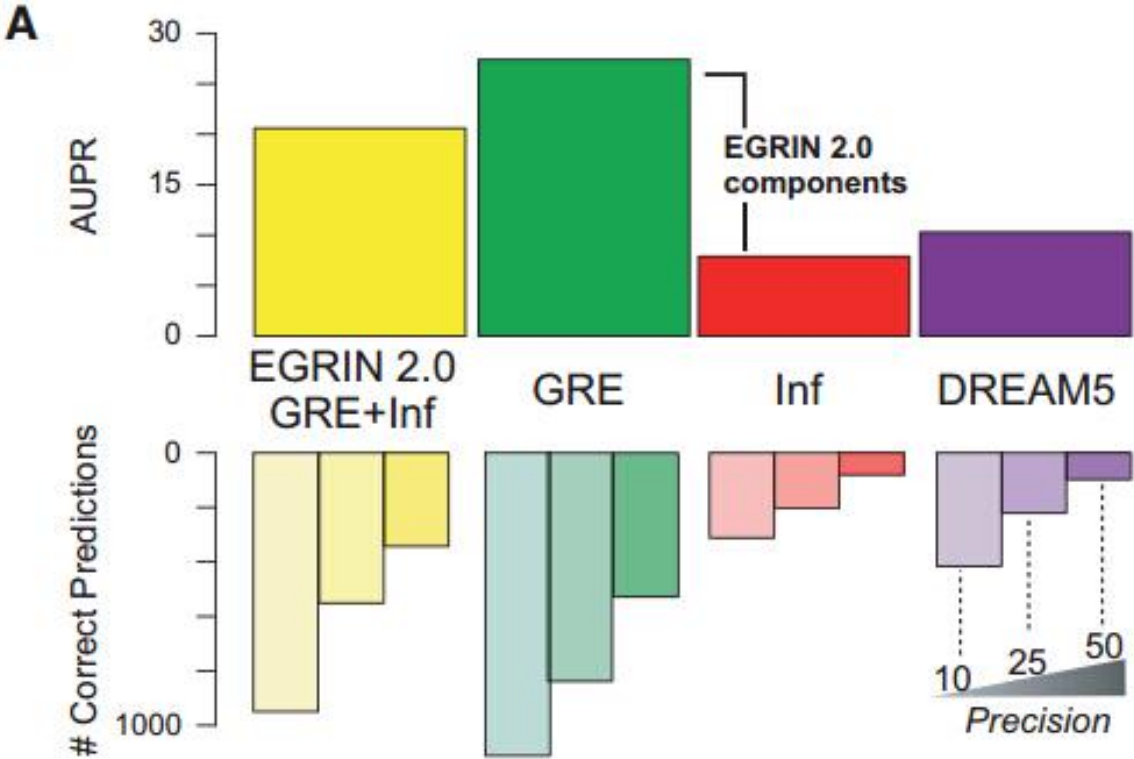
**RegulonDB:**大肠杆菌转录调控数据库

# Results

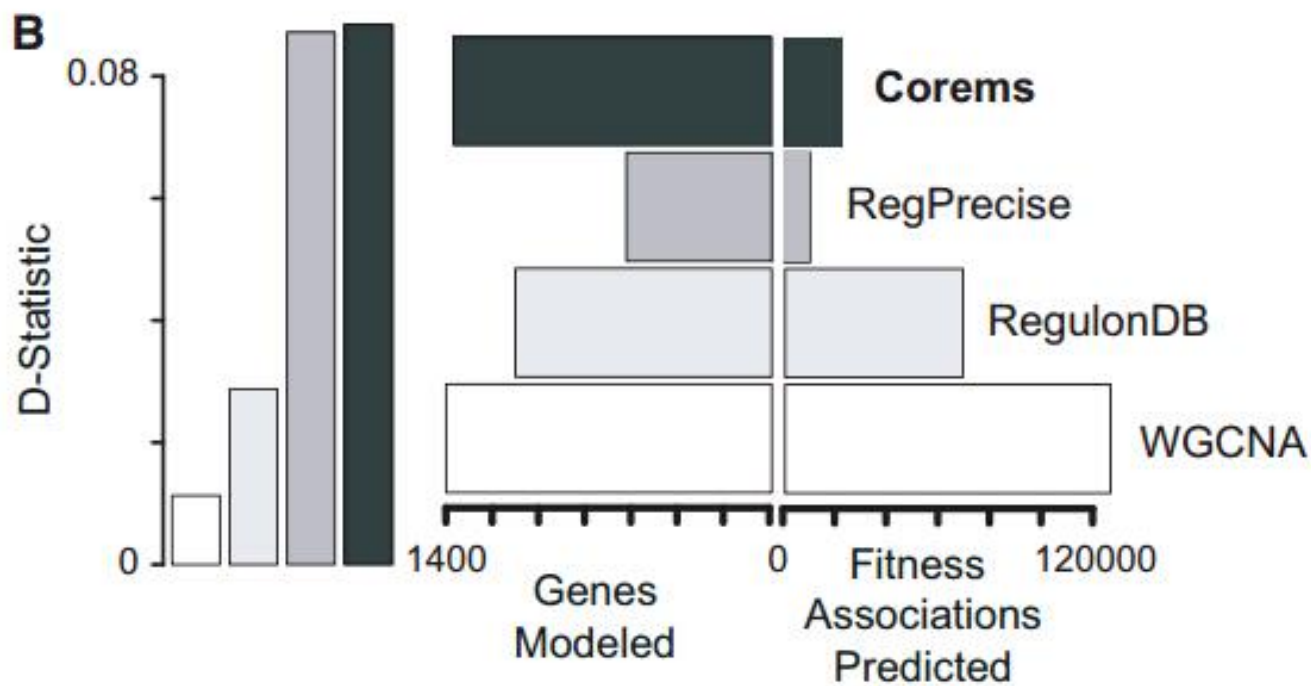
构建EGRIN2.0的模型



# EGRIN2.0可以预测经过试验验证的调控机制



利用“核”的定义来研究对细菌适应性影响相似的基因



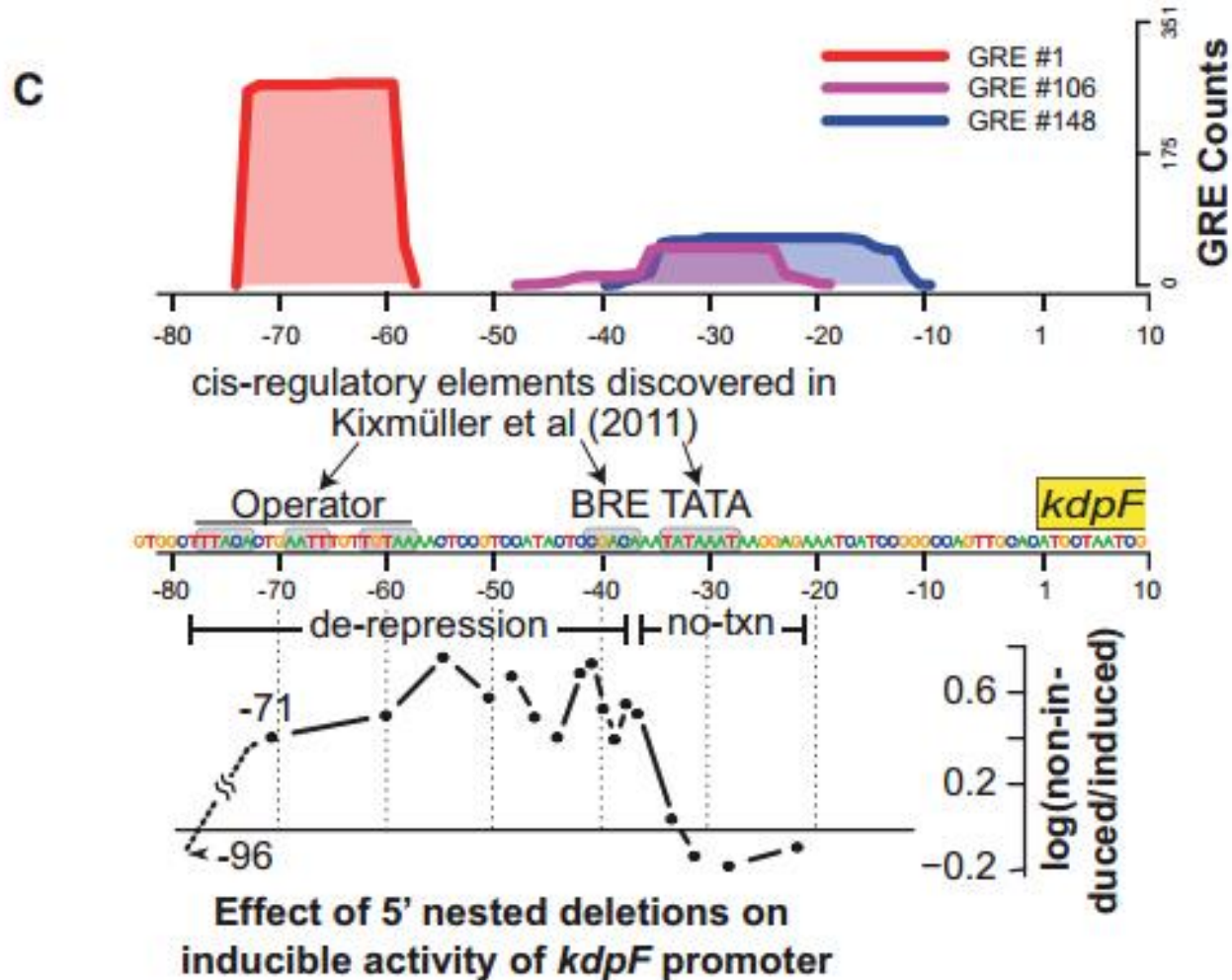


Algorithm	Fitness Correlation Cutoff	# Modeled Correlations Above Threshold	Total # Predictions	Total # Similar Fitness (Predicted Subset)	Total # Similar Fitnesses	Recall	Precision
<b>Corems</b>	0.9	134	22160	185	265	<b>0.72</b>	<b>0.006</b>
	0.75	467	22160	917	1335	<b>0.51</b>	<b>0.021</b>
	0.5	752	22160	2280	5251	0.329	<b>0.033</b>
<b>Corem (-operons/-regulons)</b>	0.9	95	16793	184	265	0.516	0.005
	0.75	317	16793	893	1335	0.354	0.018
	0.5	415	16793	2219	5251	0.187	0.024
<b>RegPrecise</b>	0.9	23	9357	63	265	0.365	0.002
	0.75	124	9357	290	1335	0.427	0.013
	0.5	271	9357	689	5251	<b>0.39</b>	0.028
<b>RegulonDB</b>	0.9	27	68891	98	265	0.276	0.0003
	0.75	131	68891	396	1335	0.331	0.002
	0.5	226	68891	1115	5251	0.301	0.005
<b>WGCNA</b>	0.9	24	126166	246	265	0.097	0.0002
	0.75	171	126166	1049	1335	0.163	0.001
	0.5	507	126166	2540	5251	0.199	0.004

核群体聚集5倍于RegPrecise、RegulonDB或WGCNA的高相关适应性的基因对。最重要的是，即便去除掉所有调节子和操纵子的影响，核依旧可以保持对适应度影响相似基因的高度富集。这显示基因间的核模型调控关联可以不用存在的调节子和操纵子来解释。

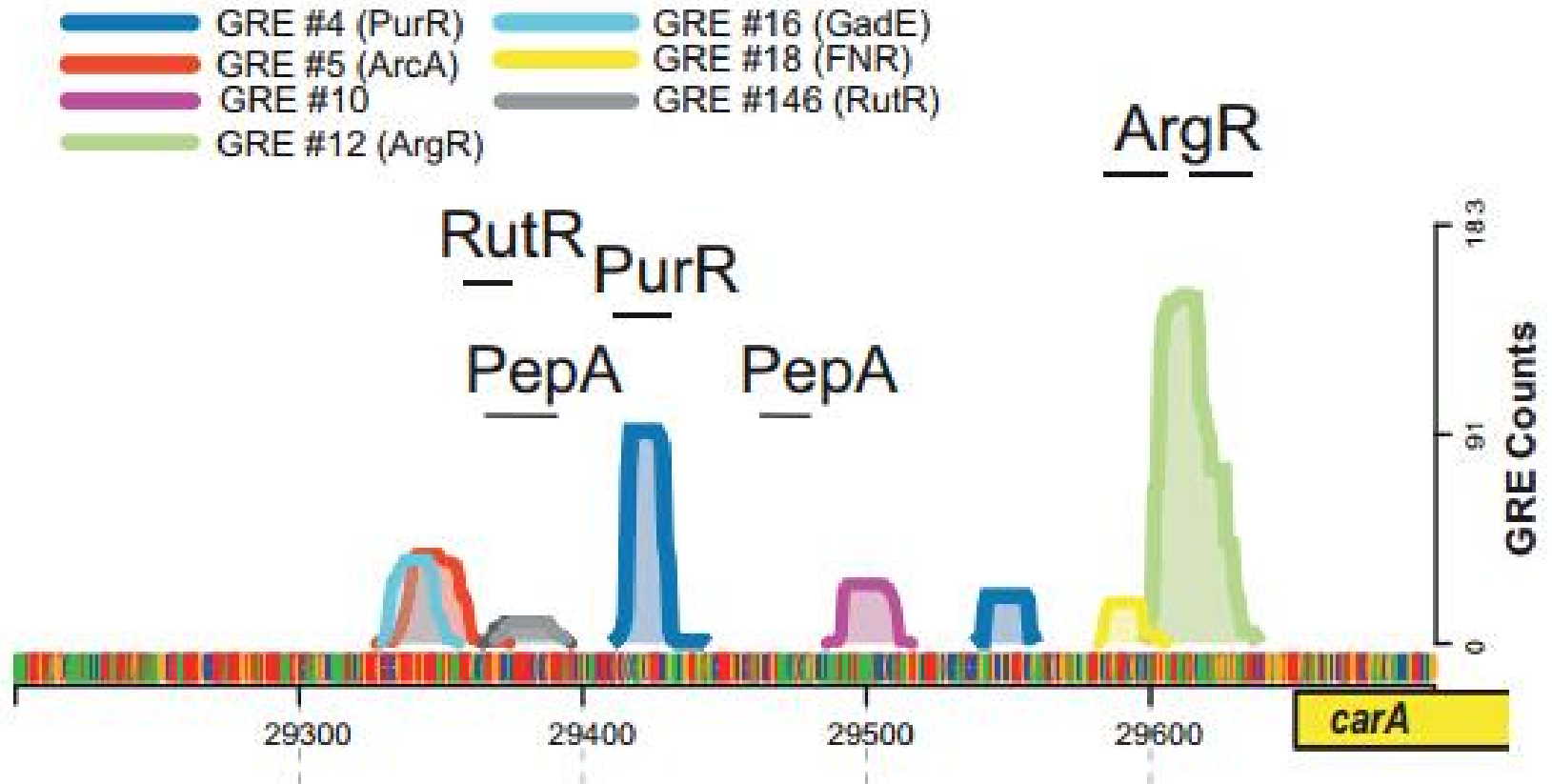


EGRIN2.0可以预测基因启动子内GREs详细的组成和其在特定环境下的重要作用

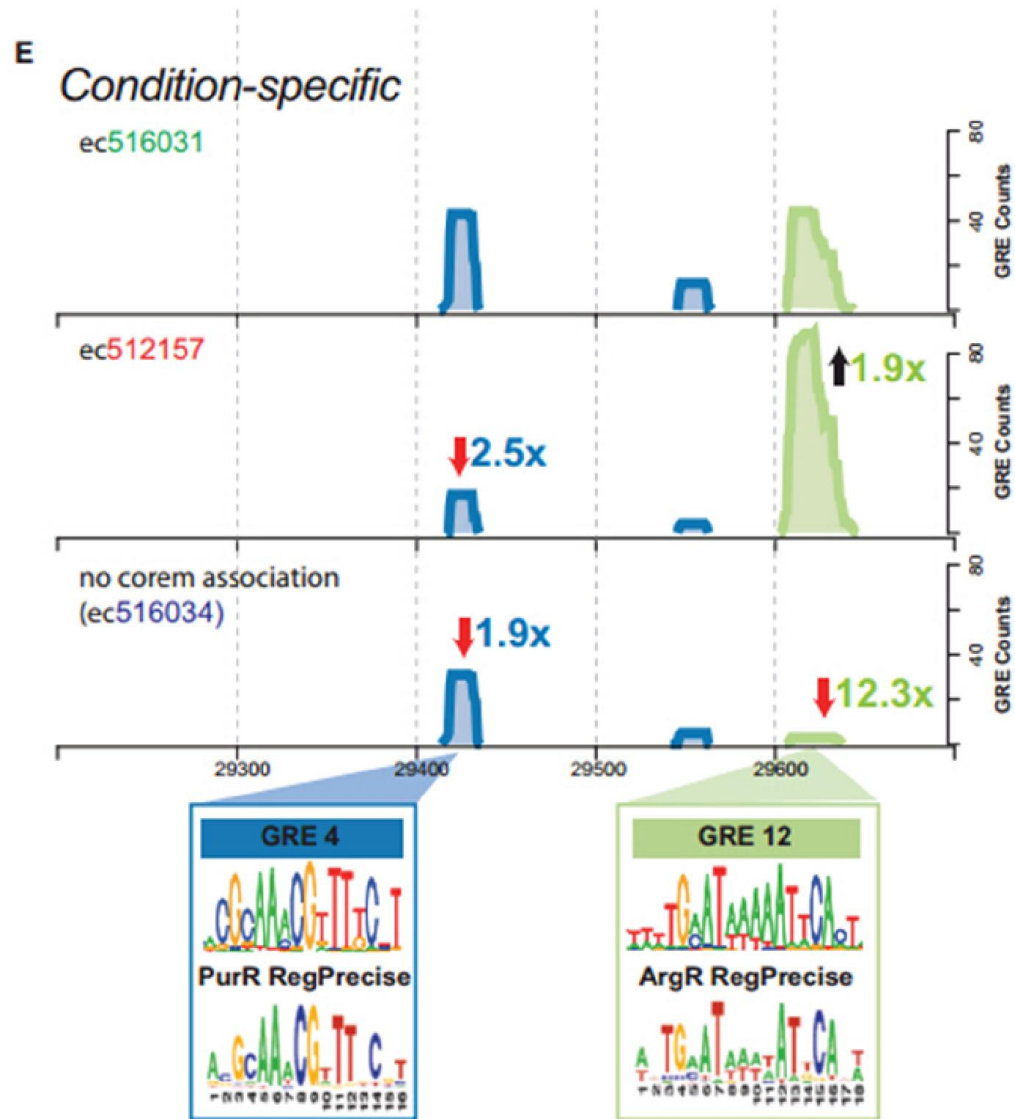


每个启动子中的特定环境下激活的GREs将基因归类为不同“核”

### D *All predictions*

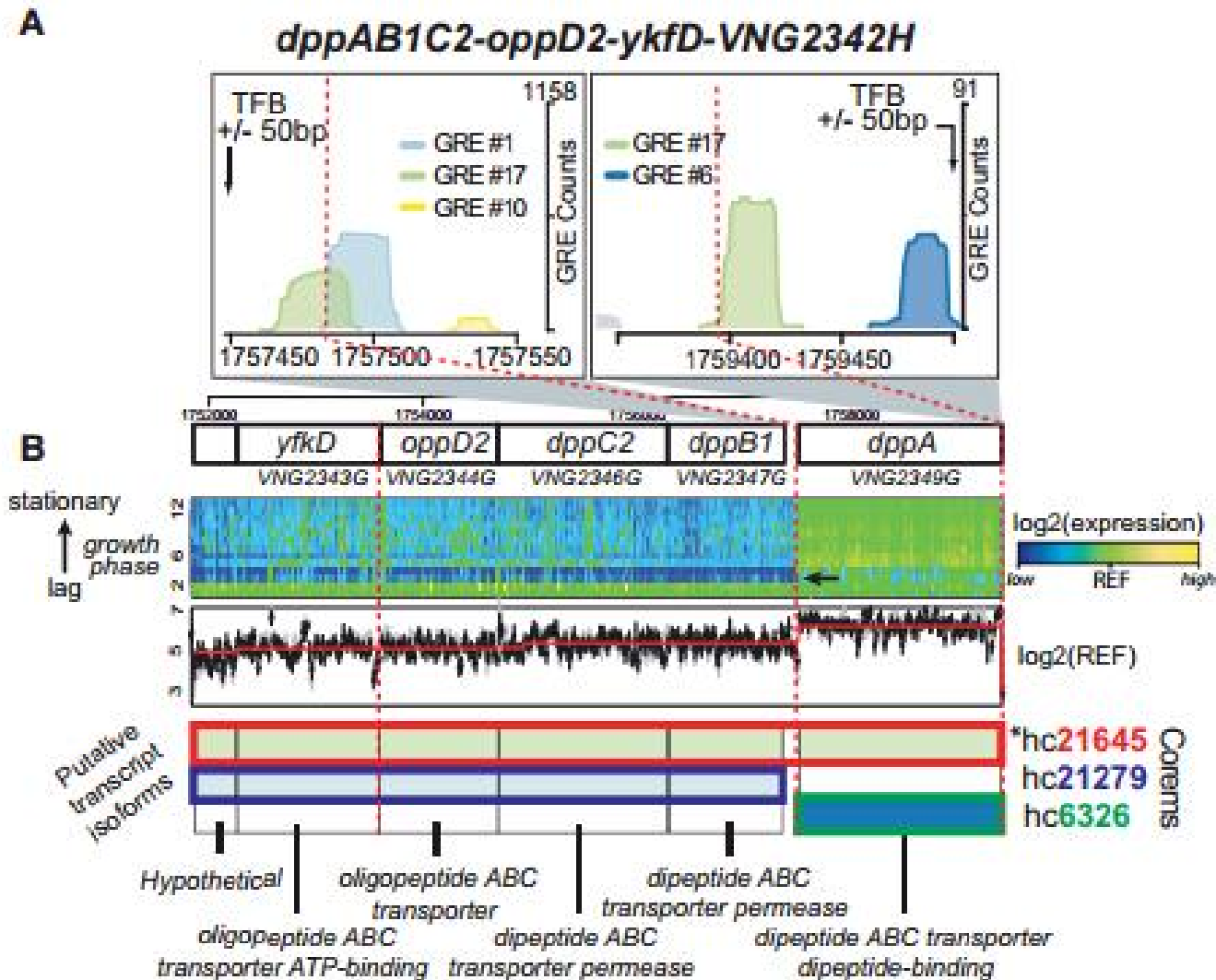


在不同的环境（ec516013, ec512751, ec516034）下，尽管PurR和ArgR处在同一个启动子区，它们的活性并不协同（这里的倍数是与D图相比）

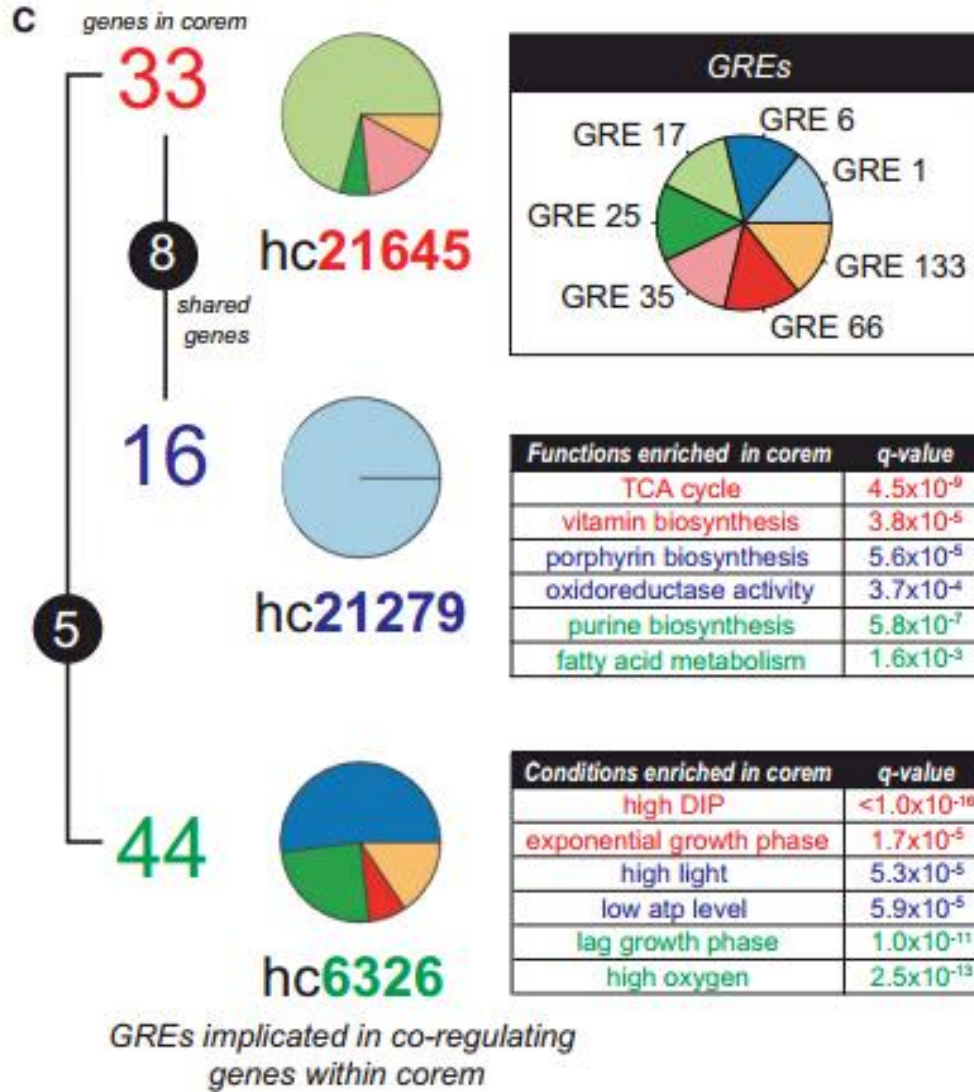


这种现象并不能用操纵子的概念很好的解释

操纵子内特定环境下激活的GREs会产生并调控多种部分重叠的转录异构体



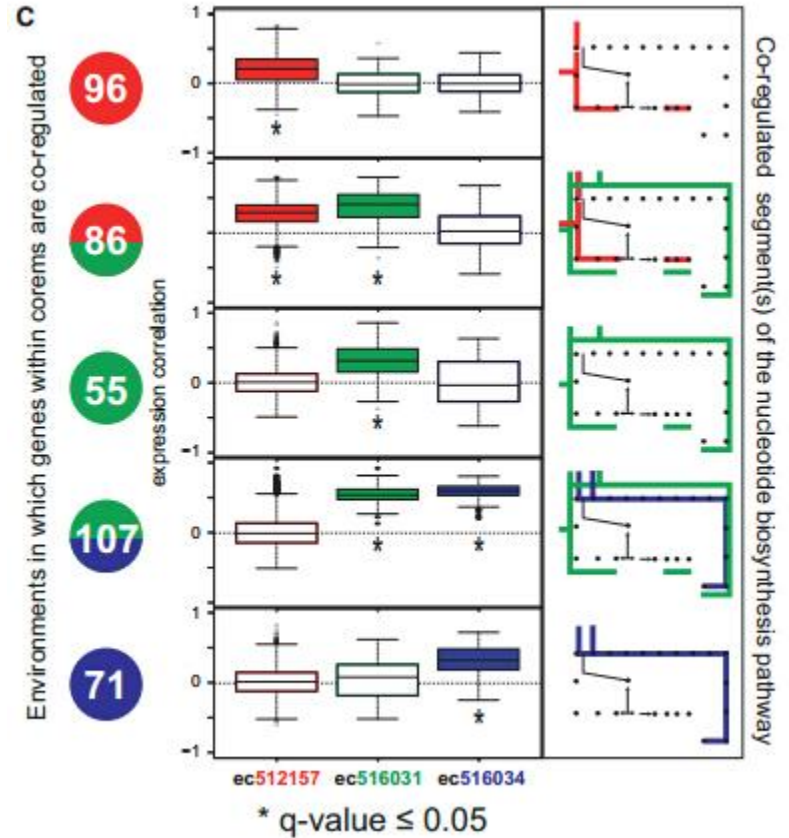
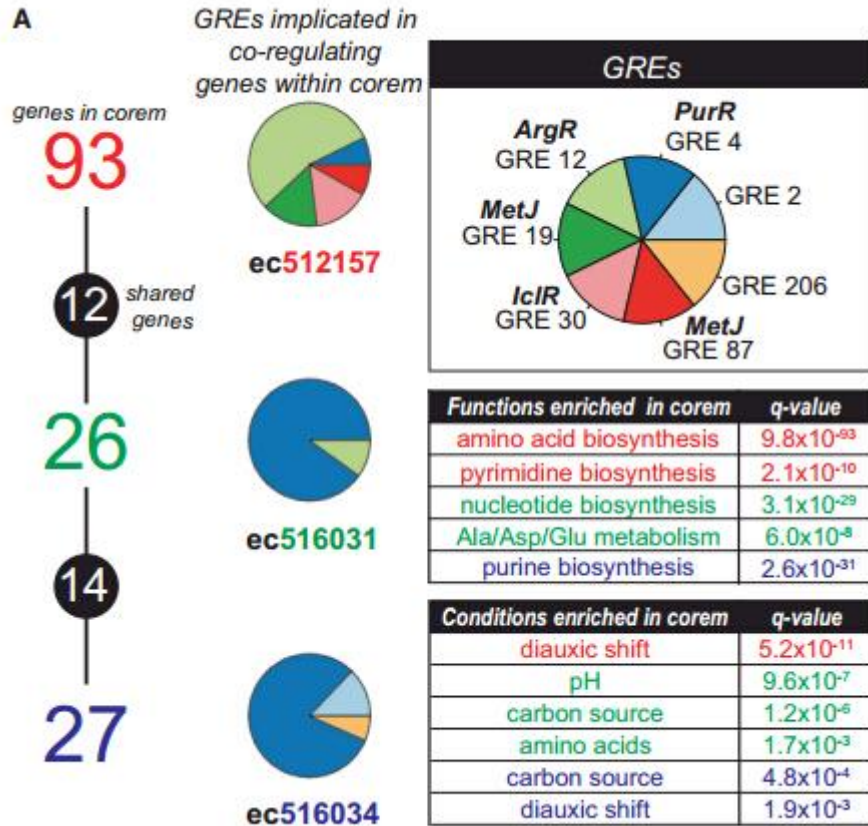
同一个操纵子内的三个不同的“核”，在特定的环境下受到不同的GRE调节



不同的颜色代表不同的GRE以及“核”模型与环境和功能的相关性

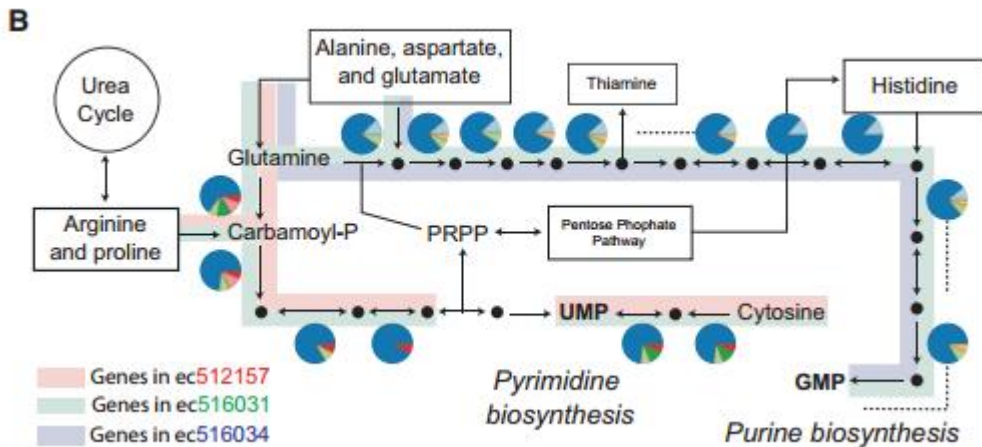
“核”模型中的基因数量

在特定环境下，一些转录因子的表现相似，并以此调控它们对应的调节子内相关一系列基因

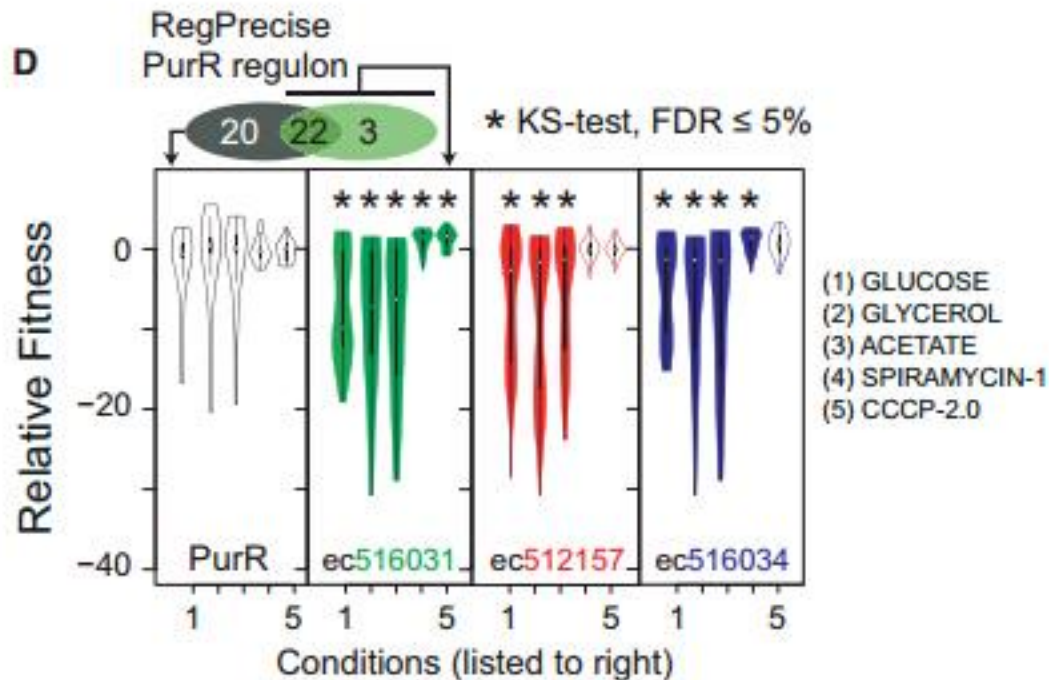


不同的环境（分别以红、绿、蓝表示）导致三个不同的“核”模型基因受到的调控不同





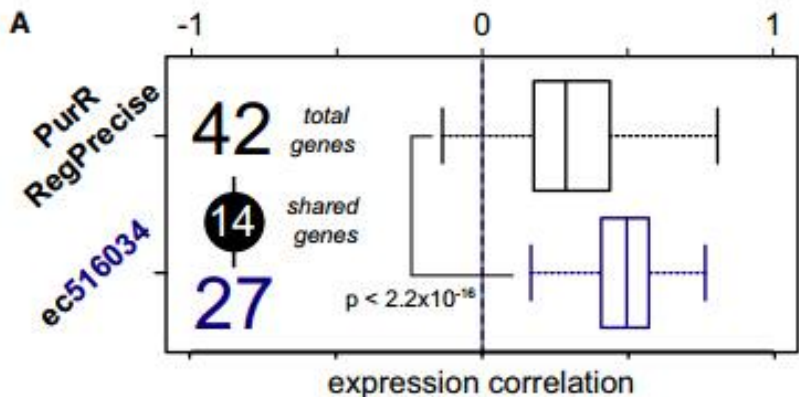
不同的“核”所对应的基因在代谢通路中的分布情况



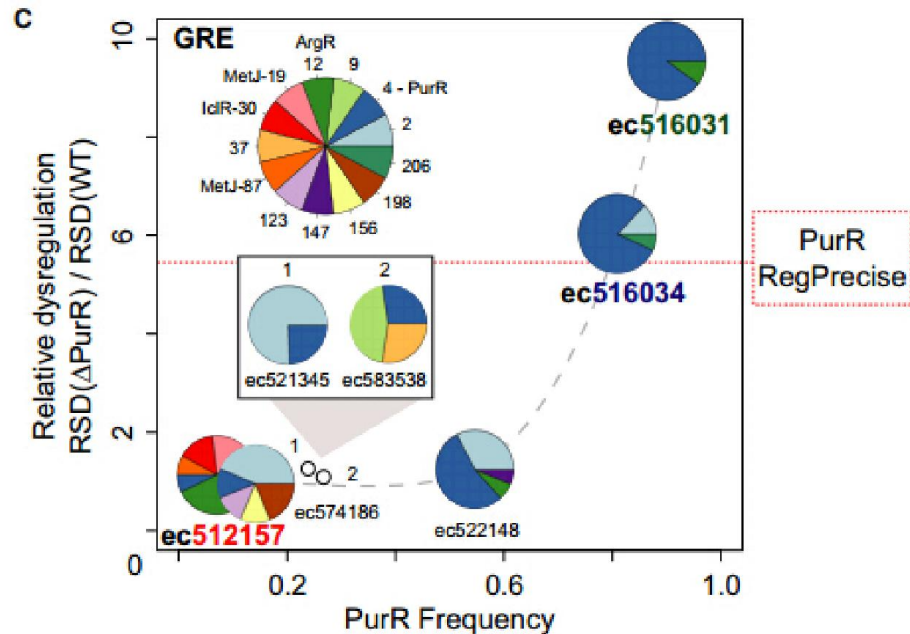
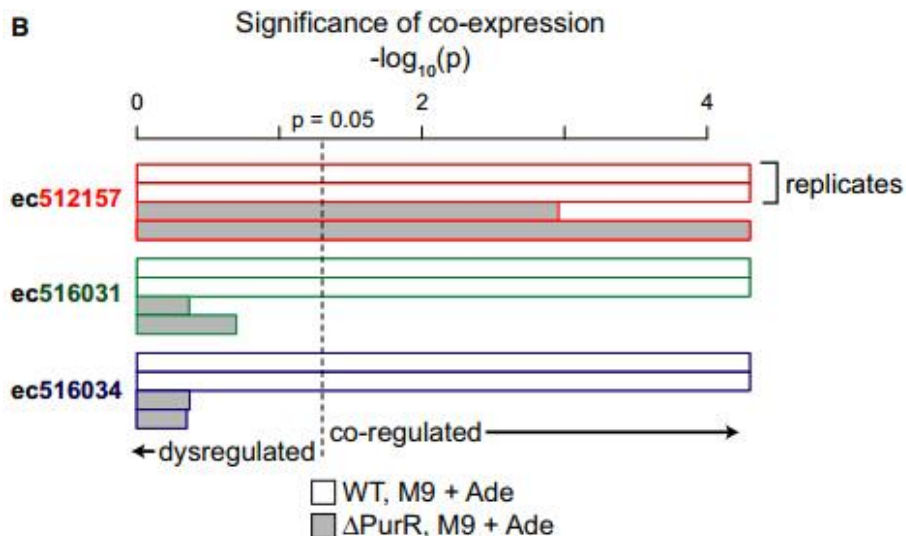
不同的环境下每个“核”的缺失株的环境适应性的协同。  
\*代表测试的显著性



EGRIN2.0预测不同环境对基因组内的TF结合位点的影响。



所有的PurR调控的基因以及其中一个“核”内的基因表达的差异，“核”模型可以显著的找出以操纵子的概念无法找出的基因集合。



# 总结

**EGRIN2.0**解释了微生物通过在全基因组范围内通过组织和对不同启动子特异性活性起作用的**GREs**调控转录来应答多变的环境。这个综合的模型阐明了微生物如何在多变的组合中重复利用基因来适应调节各种代谢途径的机制。提供了通过**EGRIN2.0**对一个细菌和一个古菌的预测。另外，作者同样利用实验来验证预测的模型的准确性。

对合成生物学的帮助：

- 1) 它提供了可培养物种基因组尺度范围的精确而广泛的模型。
- 2) 它明确的建构了整个基因组内环境依赖的调节机制，包括非常规的位点。

# 评论

优点：（1）这篇文章建模用来作为训练集合的数据量很庞大，提出的“核”的概念可以很好的解决一些原有定义无法解释的问题。

（2）从文章中可以看出，这个模型做出的预测灵敏性很高，可以预测出比原先多出数倍的可能目标，

不足：精确度并没有很明显的改观，所以可能使用过程中假阳性的概率会高一些

启发：（1）在基因的研究中不拘泥于原有的概念，可以根据实际提出新的概念。

（2）理论需要实验的验证才能更有说服力。

Thank you !