## 体外数据和体内数据的区别？

体外是在受控环境下进行，例如试管或培养皿。这种方法使科学家能够评估特定细胞中的各种生物现象，而不会分散整个生物体的注意力和潜在的混淆变量。但是尽管提供了可控的环境，但是无法复制活体内发生的条件。因此必须谨慎解释体外数据，因为这些结果不一定能预测整个生物的反应。

体内是指在活体生物体内或之上进行的实验。提供了有关特定物质或疾病进展对整个活体的影响的有价值的信息。环境不可控。

## 体内数据中序列成对概率特征的计算中可能会有一些偏差？

学习结合偏好——初级序列

预测结合位点

由于基于体外分析的，RNA序列与体内条件无关，因此在不同生理状态下的预测准确性受到限制。

体内RNA结构数据是准确的建模和在生理学相关的上下文蛋白质-RNA作用。

## 分体内数据集、体外数据集还是每个数据计中都包括体内和体外？

## 体外数据建模用于体内数据？体内数据受噪音和其它蛋白质的干扰？噪音和其它蛋白质是什么？

## 体内数据限制预测的原因是序列数据长度不同（5-186个nt之间），增加了数据处理的不确定性。提取的阳性序列可能包含些假阳性序列，可能会影响我们的体内结合预测。

## 对于体内数据，需要更多的分析以构建不同的模型。

## 序列成对概率特征也是从二级结构中提取的，受二级结构的准确性影响。

## 重复核苷酸之间的有效信息在蛋白质-RNA结合中起重要作用，核苷酸之间的信息怎样获得？

## 为什么改进了这么多，准确性还是这么低0.82？

## 我们无法探讨蛋白质-RNA结合特性如何作用于深度神经网络的训练过程以提供模型的可解释性。我们无法使用可用的主题数据库（例如CIS-BP [38]）和比较工具（例如TOMTOM [39]），因为仅保留和处理了序列基序，并选择了这些序列基序ķ -mer序列不能使用WebLogo [40]进行比对以获得序列基序。

## 序列的有序和无序怎样定义？

## RNA的结构motif怎样获得？

## 结构域、堆叠、疏水是什么？