

长非编码RNA的功能注释

The function annotation of long noncoding RNA

报告人：G07杨德昌

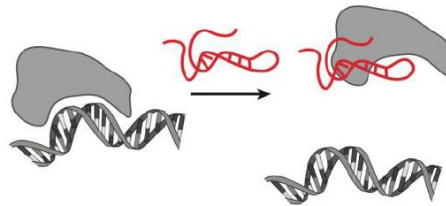
组员：王宇、聂维、尚文浩

lncRNA的功能

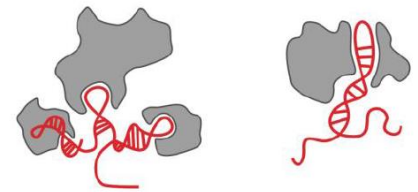
- lncRNA 在剂量补偿效应、表观遗传调控、细胞周期调控和细胞分化调控等众多生命活动中发挥重要作用

- 四种lncRNA的经典作用机制：

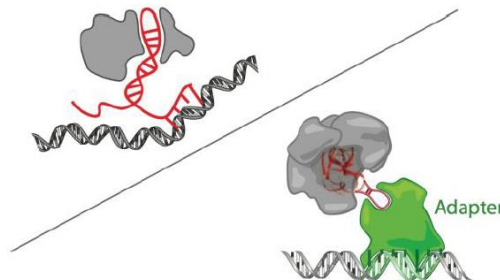
a Decoy



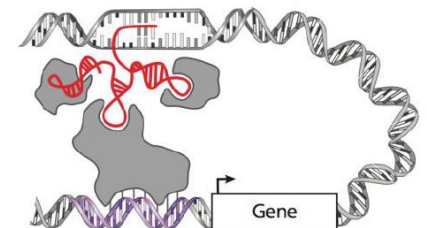
b Scaffold



c Guide



d Enhancer



为什么要进行注释

- 虽然lncRNA大量存在，但目前为止其功能研究清楚的lncRNA还不到百分之一
- 为了研究清楚lncRNA的功能和作用机制，便需要对现有的lncRNA进行归纳整理和对未知的lncRNA进行功能预测——注释
- 已有的工作：
lncRNADB——人工文献阅读
Noncode——旨在建立最完整的lncRNA集合和基本注释

我课题是对lncRNA进行全面而准确的注释，甚至尚未发现的新lncRNA

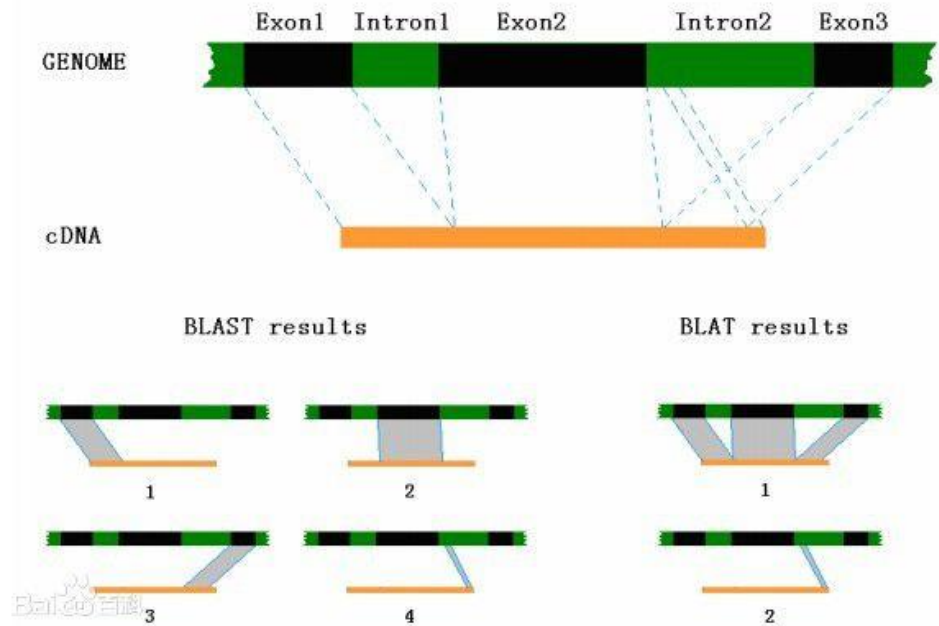
今天我主要介绍几个和课堂知识紧密联系的方面

- 1、利用blast类似工具——blat来进行lncRNA的基因组定位
- 2、通过GO数据库以及lncRNA的共表达基因来研究lncRNA的功能通路
- 3、注释lncRNA在演化方面的保守性
- 4、利用RNAfold对lncRNA进行二级结构预测
- 5、探究lncRNA的三维结构预测以及三维结构的可视化

1、利用blast类似工具——blat来进行lncRNA的基因组定位

- 序列比对 → ncbi的blast工具 → blat

- gfServer
- gfClient



psLayout version 3

match	mis-match	rep. match	N's	Q gap count	Q gap bases	T gap count	T gap bases	strand	Q name	Q size	Q start	Q end	T name	T size	T start	T end
2308	0	0	0	0	0	4	352	-	H19	2322	0	2308	chr11	135006516	2016405	2019065

H19_transl chr11 - 2016406 2019065 2016406-2017024,2017106-2017228,2017309-2017421,2017517-2017651,2017748-2019065

2、通过GO数据库以及lncRNA的共表达基因来研究lncRNA的功能通路

- 共表达：皮尔森相关系数
- 我们可以得到和某个lncRNA共表达的许多基因，一般分为两部分，正相关和负相关的基因集合，然后对这些共表达的基因进行GO富集分析

Molecular Function [Top](#)

Positively Correlated

Normal Samples

p value cutoff:

Search:

GO Term	Description	p Value
GO:0008307	structural constituent of muscle	9.55e-5
GO:0003779	actin binding	3.81e-3

Showing 1 to 2 of 2 entries

Previous **1** Next

Biological Process [Top](#)

Positively Correlated

Normal Samples

p value cutoff:

Search:

GO Term	Description	p Value
GO:0003012	muscle system process	1.34e-3
GO:0006936	muscle contraction	3.48e-3
GO:0033275	actin-myosin filament sliding	3.81e-3
GO:0030049	muscle filament sliding	3.81e-3
GO:0048747	muscle fiber development	3.81e-3
GO:0061061	muscle structure development	4.39e-3

Showing 1 to 6 of 6 entries

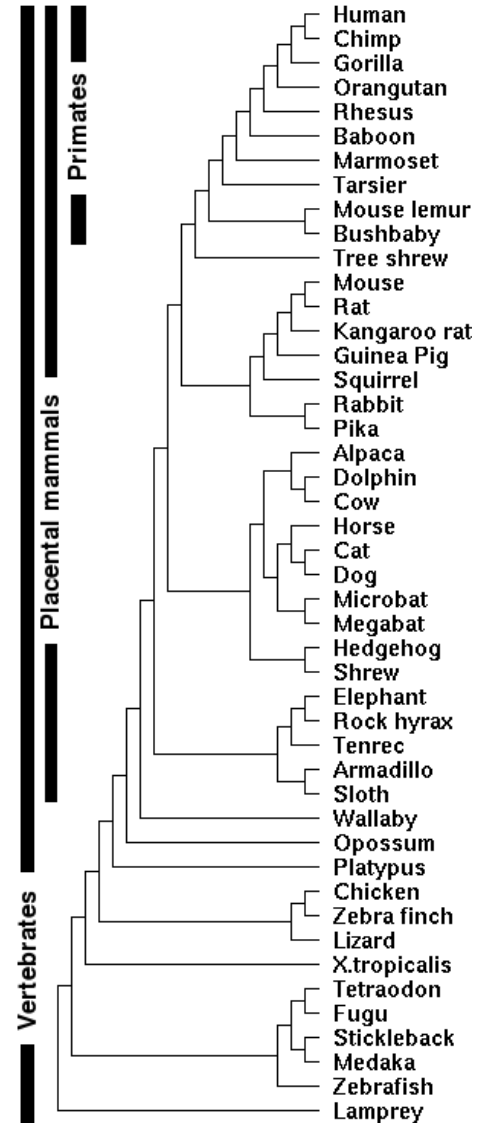
Previous **1** Next

3、注释lncRNA在演化方面的保守性

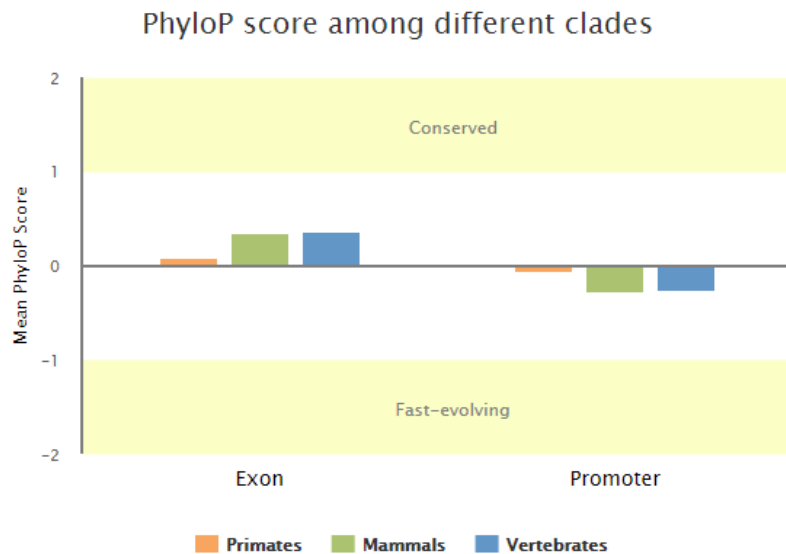
- 多物种多序列比对中Way的概念——46-way 的maf格式文件
- 系统发生树
- PHAST软件包分析——phyloP和phastCons

1. all 46 species:

```
((((((((((((((((((hg19,panTro2),gorGor1),ponAbe2),rheMac2,papHam1),calJac1,tarSyr1),(micMur1,otoGar1),tupBel1),((((mm9,rrn4),dipOrd1),cavPor3),speTri1),(oryCur2,ochPri2))),((vicPac1,(turTru1,bosTau4)),(equCab2,(felCat3,canFam2)),(myoLucl,pteVam1)),(eriEur1,sorAra1)),((loxAfr3,proCap1),echTel1),(dasNov2,choHof1)),(monDom5,macEug1),ornAna1),((galGal3,tasGut1),anoCar1),xenTro2),((tetNig2,fr2),(gasAcul,oryLat2)),danRer6)),petMar1)
```



- 便可以得到人类全基因组位置上每个碱基的 phyloP 的得分和保守元件的位置
- 由于第一步中我们已经得到 lncRNA 的基因组坐标的位置，只需查询该坐标内的得分和元件信息



Conserved Elements ?

Primates

Region	Length	Score ↓
chr11:2018648-2018783	136	38
chr11:2017898-2018073	176	20
chr11:2019020-2019069	50	20
chr11:2019085-2019138	54	16

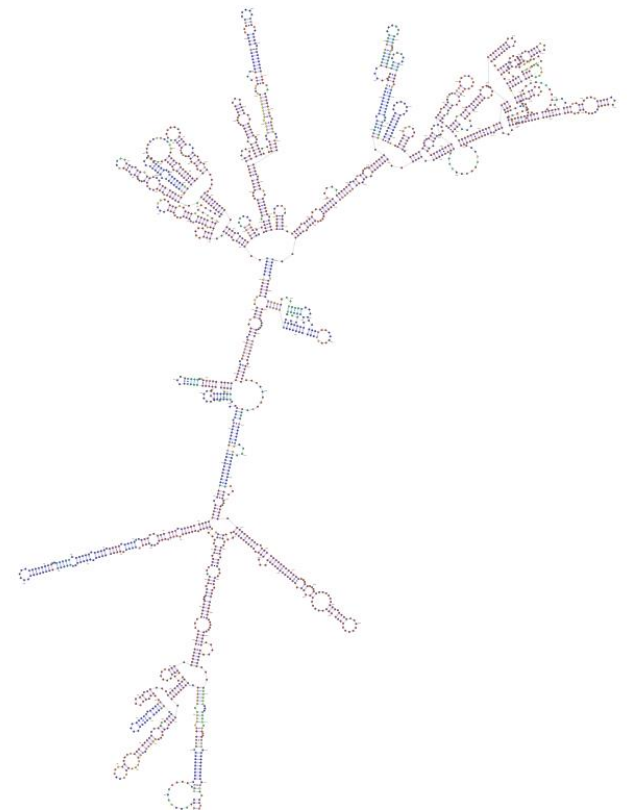
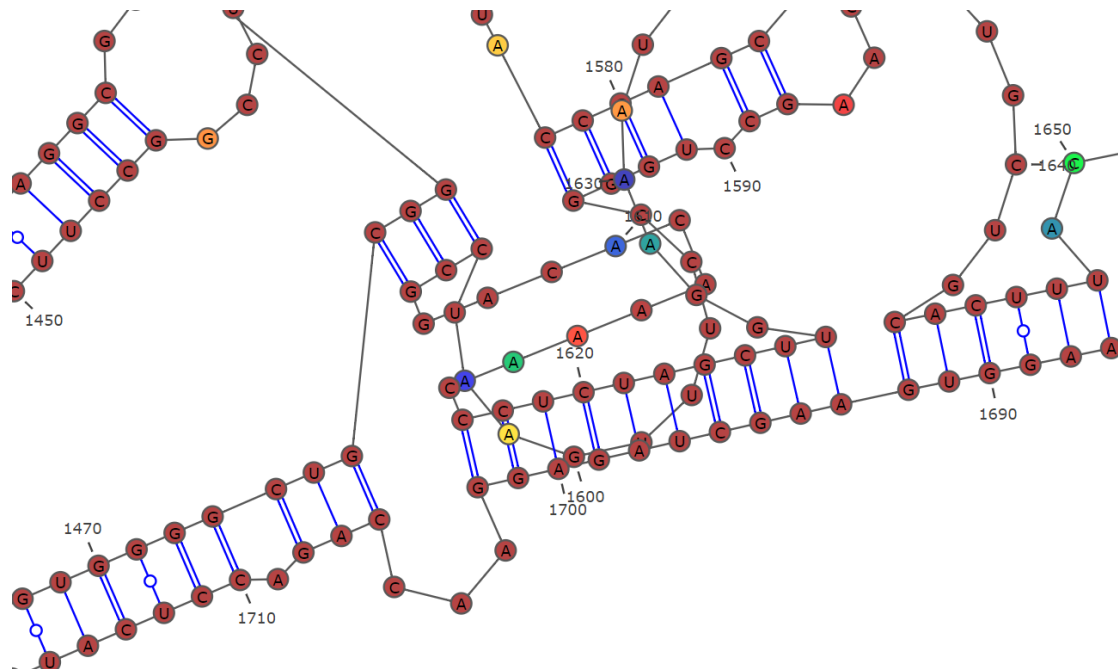
Showing 1 to 4 of 4 entries

Previous

Next

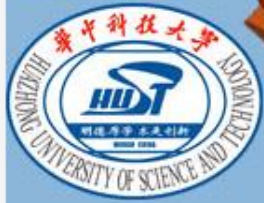
4、利用RNAfold对lncRNA进行二级结构预测

- RNAfold是RNA二级结构预测中非常常用的软件，属于著名的ViennaRNA软件包中的一个工具，基于最小自由能的方法进行预测



5、探究lncRNA的三维结构预测以及三维结构的可视化

- 二级结构和三级结构同样都和lncRNA的功能有着很大的关系，但是三级结构的预测难度却要大很多，以前蛋白质的三级结构预测的时候，罗老师曾讲过，有三种方法，1、基于序列2、比较基因组学（从别的物种的同源已知结构来做）3、二者结合。但是lncRNA的研究，并不具备这种条件，大家只能基于序列和预测的二级结构来做三级结构的预测，辅以质谱,X射线的实验数据来验证



3dRNA v2.0: Automatic building of ncRNA 3D structures

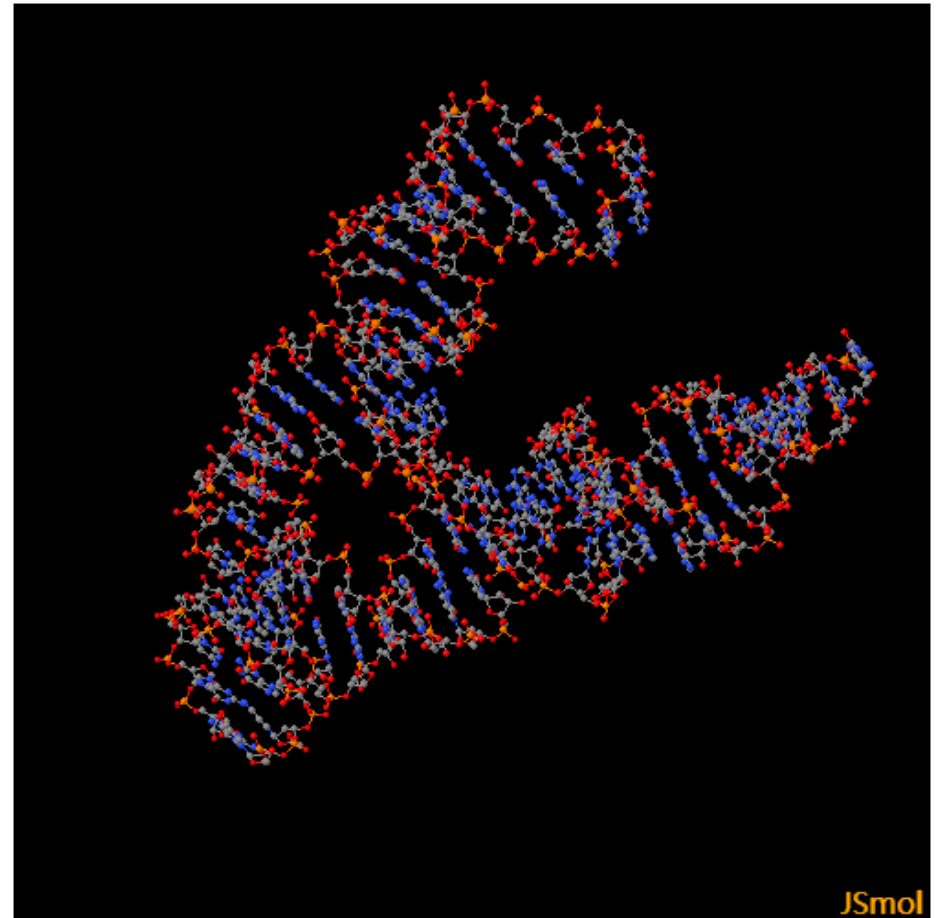
Model Name 64179633-1

Sequence

```
UCGGCGGUGGGGGAGCAUCUCCUGUAGGGGAGAUGUAACCC  
CCUUUACCUGCCGAACCCCGCCAGGCCCGGAAGGGAGCAAC  
GGUAGGCAGGACGUCGG
```

2D Structure

```
(((((.....(((((((.....)))))).....)))))).....((((.....(((.....  
(((.....))).....))).....)))))).....))))))
```



JSmol



谢谢



谢谢

