目录：

1.脚本的目的

2.联系方式

3.脚本列表和说明

4.安装和设置

5.用法

6.变更日志

1.脚本的目的

核小体亲和力改变的天然或合成启动子的计算重新设计。

2.联系方式

哈尔阿尔珀

德克萨斯大学奥斯汀分校

东经200度。迪恩基顿街，C0400站

德克萨斯州奥斯汀78712-1589

CPE 5.408/电话：（512）471-4417/传真：（512）471-7060

电子邮件：halper@che.utexas.edu

3.脚本列表和说明

亲和力.m

获取DNA序列并计算每个核苷酸的核小体亲和力值。

容器废弃.m

这个脚本在DNA序列中寻找用户定义的DNA基序实例。基序可以包括退化基。

gc含量.m

计算序列的GC内容

GCM剖面图

计算输入DNA序列每个100bp滑动窗口的GC含量。

maxprom公司

这个程序将使用一个启动子，并以用户定义的碱基对增量（轮数）迭代地减少预测的核小体占有率，直到占有率不再减少为止。

最小值m

Nucleomin以一个序列作为输入，搜索起始序列的所有n-核苷酸变体，以找到具有最小预测核小体亲和力的候选序列，要求该序列也是可合成的，并且不包含额外或更少的转录因子结合位点。n是用户定义的。

问题等级.m

注意包含特定DNA基序的输入DNA序列的位置，并将它们从最低核苷酸排列到最高核苷酸。

兰德普罗姆

初始化合成启动子的随机DNA序列，并生成启动子内的序列列表，这些序列必须在基于用户规范的设计过程中保存。

随机序列m

生成指定长度和GC含量的随机DNA序列

雷姆

尝试从输入序列中删除尽可能多的与一组DNA基序匹配的序列。用户还可以指定在此过程中不能更改的底座位置。

面积m

计算DNA序列的累积亲和力分数。

顺序检查.m

这个程序的唯一目的是确保一个序列可以由IDT的gblocks合成。在撰写本文时，它已经足够了，但随着合成技术的改进，它的某些特性可能不再是必需的。

合成PROM.m

这个功能为启动子提供了一个大致的轮廓，并使合成的核小体优化启动子。

4.安装和设置

为Windows系统提供的安装说明。

1） 获取安装了生物信息学工具箱（在r2013b上测试）的MATLAB副本（在r2011b上测试）

2） 将上面列出的脚本复制到MATLAB工作目录中

3） 下载NuPoP的FORTRAN代码副本（截至2013年6月28日，位于http://nucleosome.stats.northwestern.edu/ 作为“NuPoP\u F”）

插入下载PERL和MinGW for Gfortran编译器

4） 编辑numpop\u F的FORTRAN代码，如下所示：

替换npred.f90中的以下内容：

替换：

隐式无

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

整数i，lfn，mlL，rep，species，order；字符\*80文件名；字符\*3 tpc

实\*8频率1（4），事务处理1（4,4），事务处理2（16,4），事务处理3（64,4），事务处理4（256,4），Pd（500,11）

实\*8 freqN4（64,4），tranN4（（147-4）\*256,4），freqN1（147,4），tranN1（584,4）

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

打开（1，file=&apos;yourpath/NuPoP\u F/profile/freqL.txt&apos;）

读取（1，\*）频率1；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL.txt&apos;）

do i=1,4；读取（1，\*）事务处理1（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL2.txt&apos;）

do i=1,16；读（1，\*）tranL2（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL3.txt&apos;）

do i=1,64；读取（1，\*）事务处理3（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL4.txt&apos;）

do i=1256；读（1，\*）tranL4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/147freqN.txt&apos;）

do i=1147；读取（1，\*）频率1（i，：）；结束do

关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/147tranN.txt&apos;）

do i=1584；读取（1，\*）事务1（i，：）；结束do

关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/146-149freqN4.txt&apos;）

do i=1,64；读（1，\*）频4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/146-149tranN4.txt&apos;）

do i=1，（147-4）\*256；读取（1，\*）事务处理4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/Pd.txt&apos;）

do i=1500；读取（1，\*）Pd（i，1:11）；结束do；关闭（1）

写（\*，&apos;（a）&apos;请输入&apos;

写入（\*，&apos;（a）&apos;，advance=&apos;no&apos;）&apos;DNA序列文件名（FASTA）：&apos;；读取\*，文件名

mlL=500

写（\*，&apos;（a）&apos;，advance=&apos;no&apos;）&apos;马尔可夫模型（1或4）的阶：&apos;；读取\*，顺序

如果（order/=1和order/=4），那么；打印\*，&apos;1或4应输入！“住手。”；停止；结束if

代表=1

打印\*，“”

write（\*，&apos;（a）&apos;从以下列表中选择物种：&apos;

打印\*，&apos;1=人类2=老鼠3=老鼠&apos;

打印\*，&apos;4=斑马鱼5=D。黑腹果蝇6=C。象棋

打印\*，&apos;7=S。酿酒酵母8=C。白色念珠菌9=S。庞贝

打印\*，&apos;10=A。拟南芥11=玉米0=其他&apos;

打印\*，“”

write（\*，&apos;（a）&apos;，advance=&apos;no&apos;）&apos;输入所选物种的标签：&apos;；阅读\*，物种

打印\*，“”

写（\*，&apos;（a）&apos;预测……&apos;

lfn=len\u trim（文件名）

如果（顺序==1），则

调用vtbfb（lfn，trim（fileName），freqL1，tranL1，freqN1，tranN1，mlL，rep，species，Pd）

如果（顺序==4），则

调用vtbfbNL4（lfn，trim（fileName），freqL1，tranL1，tranL2，tranL3，tranL4，freqN4，tranN4，mlL，rep，species，Pd）

结束if

写入（\*，&apos;（a）&apos;完成&apos;

结束

使用：

隐式无

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

整数i，lfn，mlL，rep，species，order；character\*80文件名，stringorder，stringspecies；字符\*3 tpc

实\*8频率1（4），事务处理1（4,4），事务处理2（16,4），事务处理3（64,4），事务处理4（256,4），Pd（500,11）

实\*8 freqN4（64,4），tranN4（（147-4）\*256,4），freqN1（147,4），tranN1（584,4）

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

打开（1，file=&apos;yourpath/NuPoP\u F/profile/freqL.txt&apos;）

读取（1，\*）频率1；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL.txt&apos;）

do i=1,4；读取（1，\*）事务处理1（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL2.txt&apos;）

do i=1,16；读（1，\*）tranL2（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL3.txt&apos;）

do i=1,64；读取（1，\*）事务处理3（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/tranL4.txt&apos;）

do i=1256；读（1，\*）tranL4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/147freqN.txt&apos;）

do i=1147；读取（1，\*）频率1（i，：）；结束do

关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/147tranN.txt&apos;）

do i=1584；读取（1，\*）事务1（i，：）；结束do

关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/146-149freqN4.txt&apos;）

do i=1,64；读（1，\*）频4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/146-149tranN4.txt&apos;）

do i=1，（147-4）\*256；读取（1，\*）事务处理4（i，：）；结束do；关闭（1）

打开（1，file=&apos;yourpath/NuPoP\u F/profile/Pd.txt&apos;）

do i=1500；读取（1，\*）Pd（i，1:11）；结束do；关闭（1）

调用GETARG（1，文件名）

调用GETARG（2，stringorder）

调用GETARG（3，stringspecies）

读取（stringorder，\*）顺序

读取（stringspecies，\*）物种

mlL=500

如果（order/=1和order/=4），那么；打印\*，&apos;1或4应输入！“住手。”；停止；结束if

代表=1

lfn=len\u trim（文件名）

如果（顺序==1），则

调用vtbfb（lfn，trim（fileName），freqL1，tranL1，freqN1，tranN1，mlL，rep，species，Pd）

如果（顺序==4），则

调用vtbfbNL4（lfn，trim（fileName），freqL1，tranL1，tranL2，tranL3，tranL4，freqN4，tranN4，mlL，rep，species，Pd）

结束if

结束

5） 将字符串“yourpath”替换为NuPoP\u F所在的目录（使用完整路径，例如“C:/Users/…”

6） 将文件重命名为“Npred2.f90”，并使用NuPoP\u F附带的手册中提供的说明将Npred2.f90编译为Npred2.exe。用包含的安装文件替换NuPop提供的安装文件。有关NuPoP安装的详细信息，请参阅NuPoP手册。

7） 将包含Npred2.exe的目录添加到系统路径。单击“开始”，右键单击“计算机”，选择“属性”，选择“高级系统设置”，在“高级”选项卡上，单击“环境变量”，在“系统变量”类别中编辑“路径”变量，方法是在最后一个目录后加上分号并添加要添加的目录。

您现在可以开始设计促销员了！

5.用法

1） 挑选一个促销员。启动子的设计必须包括其基因组或质粒背景的上游200bp和下游100bp。这将确保为启动子变体计算的核小体亲和力值彼此可比。注意启动子起始端和末端的核苷酸位置。

2） 注释转录因子结合位点，并注意结合位点所覆盖的核苷酸。一个特别用户友好的存储库是酵母启动子图谱http://ypa.ee.ncku.edu.tw/

3） 注释任何你不想引入到设计的启动子中的序列。这些序列，如果存在于野生型启动子，将不会改变。

4） 生成输入文件。对于TEF启动子，我们输入启动子本身的DNA序列加上上游200bp和下游100bp，如下所示：

TEF=&apos;GGAAAGGGGCAGGCAACGCAATATATGTGAGCATACTCATGCATGCATGCATGCATGCATGCATGCATTCATGCATTCATTCATGCATTCATGCATTCATGCATTCATGCATGCATTCATGCATGCATGCATGCATGCATGCATGCATGCATGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATGCATGCATGCATTCATGCATGCATGCATTCATGCATGCATTCATTCATGCATTCATGCATTCATGCATTCATGCATTCATGCATGCATTCTCATGCATTCATGCATGCATGCATTCATTCATGCATTCATTCATTCTCATTCTCATGCATGCATGCATGCATGCATGCATGCATTCATTCATTCATTCATTC塔加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加加TTCTGTCT&apos;；

对于被转录因子结合位点覆盖的核苷酸，我们进入：

TEF=[281:291 334:343 377:383 443:484]；

对于不能从设计的启动子中引入或移除的序列，制作包含相关基序的细胞阵列。我们使用了在yeastract.com上找到的TF共识列表，以及我们研究的起始密码子和TATA框。

这些输入文件包含在Sample Data.mat中

MATLAB命令示例：

按1bp步骤优化TEF：

[TEFproms，tefareaces，TEFcurves]=maxprom（TEF，TEFstart，TEFend，1，tefporbiddenseqs）；

按1bp步骤设计Psynth1：

[psynth1proms，psynth1reas，psynth1curves]=synthprom（psynth1params，psynth1start，psynth1end，1，forbiddenseqs）；

对于每个命令，第一个输出是核小体优化启动子列表，从野生型（或种子）序列开始，以1bp的步骤向预测核小体亲和力降低的变体前进。第二个输出是每个启动子对应的累积亲和分数，第三个输出是用于计算每个启动子累积亲和分数的核小体亲和曲线。

当程序运行时，它们会周期性地显示一个进度指示器，它描述了程序在计算当前变异方面的进展。

6.变更日志

2014年4月18日：更新安装信息。多亏了郑约瑟