

# 转录因子结合位点(TFBS)

作者邮箱: 2740881706@qq.com



# 转录因子

能够结合在某基因上游特异核苷酸序列上的蛋白质，活化后从胞质转位至胞核，通过识别和结合基因启动子区的顺式作用元件，启动和调控基因表达。

# 转录因子

- 常见：FOXC1, SOX9, CEBPB, CTCF
- BRCA1 蛋白具有锌指结构、酸性氨基酸区域等, 与许多DNA 结合蛋白结构类似,提示其可能为一种转录因子。体外实验证实BRCA1 蛋白羧基末端的酸性区域与酵母 Ga14 基因DNA 结合位点结合,可充当转录起始因子。BRCA1 蛋白能刺激P21 的表达,这些为其是一种转录因子提供直接证据。

## reference:

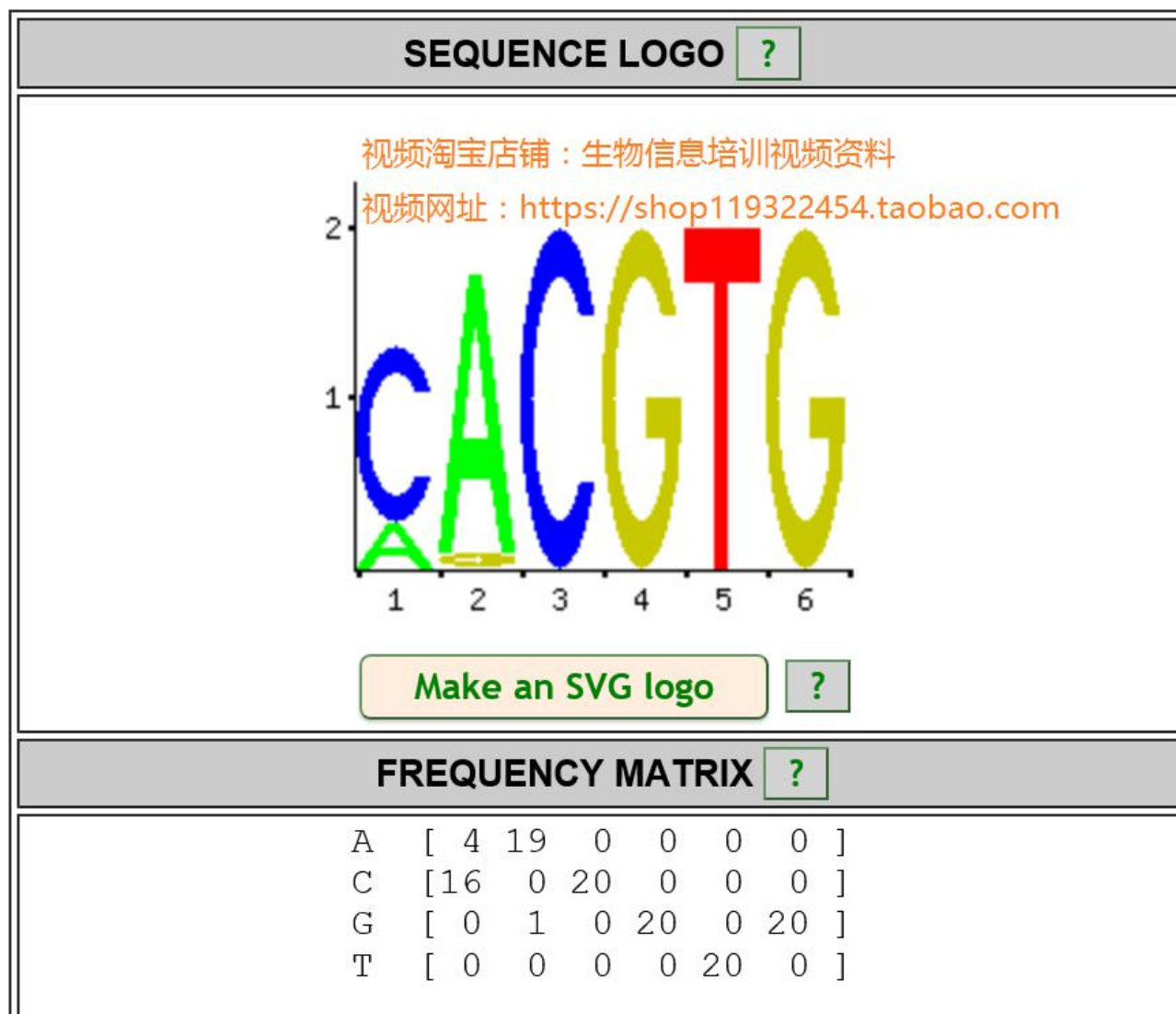
BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells

Activation of Transcription in Vitro by the BRCA1 Carboxyl-terminal Domain

# 转录因子结合位点

转录因子结合位点(Transcription factor binding site, TFBS)是与转录因子结合的 DNA 片段, 长度通常在 5~20 bp 范围内, 一个转录因子往往同时调控若干个基因, 而它在不同基因上的结合位点具有一定的保守性, 又不完全相同。

# 位置权重矩阵模型 (Position Weight Matrix, PWM)



# 位置权重矩阵模型 (Position Weight Matrix, PWM)

如果TFBS 长度为  $L$ , PWM 就是一个 4 行  $L$  列的矩阵, 这个矩阵中每行对应着一种核苷酸, 每列对应着TFBS 中的一个位置, 第  $i$  行第  $j$  列的元素是 TFBS 中第  $j$  位上出现核苷酸  $i$  的概率。一个长度为  $L$  的序列与该转录因子结合的概率即为各个位置上核苷酸对应概率的乘积。某段序列与转录因子结合的概率越大, 就说明它与转录因子相结合的结合能力越强。

# 预测方法

```
data(MA0003.2)      视频淘宝店铺：生物信息培训视频资料
data(MA0004.1)      视频网址：https://shop119322454.taobao.com
pwmList <- PWMMatrixList(MA0003.2=toPWM(MA0003.2), MA0004.1=toPWM(MA0004.1),
                          use.names=TRUE)
subject <- DNASTring("GAATTCTCTCTTGTTGTAGTCTCTTGACAAAATG")
siteset <- searchSeq(pwm, subject, seqname="seq1", min.score="60%", strand="*")

sitesetList <- searchSeq(pwmList, subject, seqname="seq1",
                         min.score="60%", strand="*")
```

# 数据

- 基因上游序列
- PWM矩阵



# 结果

Gene	start	end	score	strand	attributes
gene1	17	30	0.8300765	+	TF=NR2F1;class=Zinc-coordinating;sequence=TAGTCTCTTGACAA
gene1	18	31	0.8077556	+	TF=NR2F1;class=Zinc-coordinating;sequence=AGTCTCTTGACAAA
gene1	10	23	0.8181152	-	TF=NR2F1;class=Zinc-coordinating;sequence=GAGACTACAACAAG
gene1	9	19	0.8642018	+	TF=NFIL3;class=Zipper-Type;sequence=TCTTGTTGTAG
gene1	12	22	0.8642018	+	TF=NFIL3;class=Zipper-Type;sequence=TGTTGTAGTCT



# Thanks!!!

作者邮箱: 2740881706@qq.com