

[MEME][http://meme-suite.org/doc/meme.html?man_type=web#sp_search]

MEME用于搜索新的, 无间隔(gapped)的motifs(recurring, fixed-length patterns). MEME将多种长度的模型分为两个或多个区分的motifs. motif是一个类似的序列模型, 重复出现在一组相关的序列中.

MEME展示的motifs是一个位置依赖的同时包含字符概率的矩阵, 描述一种模式中每个位置每一种可能字符在每个位置的概率. MEME motifs不包含gaps, 因此具有多种长度gap的模型将被MEME分到2个或多个不同的motif中.

MEME输入一组序列, 输出要求数目的motif. MEME使用统计模型自动选择最佳的宽度, motif出现次数.

网页版MEME可接受第二个输入序列作为control, 然后查询相对于control序列集, 在第一个序列集中富集的motifs, 称为: discriminative motif discovery

MEME algorithm overview

首先, 在输入序列中搜索可能的EM起点; 其次, MEME使用EM(**Expectation Maximization**)将所有这样的起点聚集起来; 再次, EM完成聚集后, MEME更新motif的宽度和深度(位置出现数目); 最终, MEME执行'soft'去除过程(用于搜索额外的motifs), 去除所有发现的潜在的最好的motif位置, 以证实该位置是真实可靠的. MEME重复执行以上4步骤查询motifs直到满足停止搜索条件.

Input

```
<primary sequence file>
```

FASTA文件格式输入, 另外需要指定序列类型, -dna, -rna, -protein或-alph. 同时, MEME支持包含weighting值的FASTA格式输入.

Output

输出至命名为meme_out的目录中. 可通过-o或-oc选项修改输出目录. 该路径下包含:

- meme.html 提供可读的交互式HTML文件
- meme.txt 文本文件
- meme.xml 可用于machine processing的XML文件
- logoN.png MEME分析得到的PNG/EPS图像, 包含序列logos
- logo_rcN.png 包含反向互补序列的PNG/EPS图像(complementable alphabets only)

MEME输出文件中的position-specific probability matrix(PSPM)结果可通过FIMO/MAST搜索与之匹配序列.

Examples

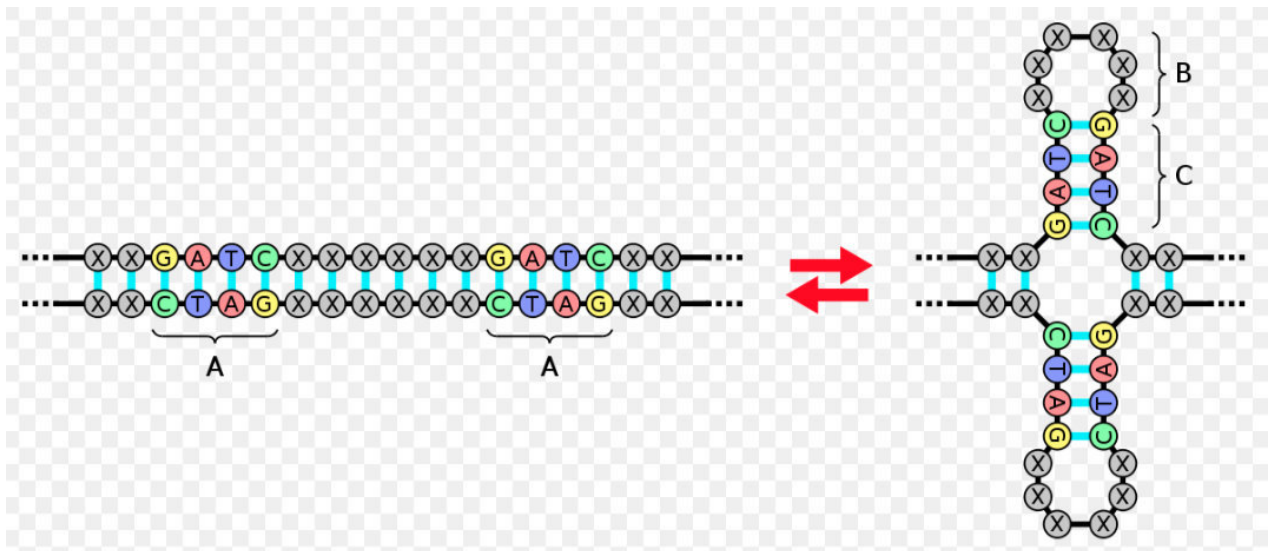
1. 简单DNA搜索

```
meme crp0.s -dna -mod oops -pal
```

在包含FASTA格式的输入的DNA序列中搜索单个motif.

跟据oops位置分布模型, MEME假设每一序列恰好包含一个该motif(该选项最快最敏感, 但是如果输入序列没有将使MEME变得'blurry'; zoops, 对应包含最多一个或没有; anr, 对应包含任意数目非重叠的motif, 可用于检测包含重复多个motif的序列); 默认假设每条序列出现1次或0次

指定pal参数, MEME仅查看存在互补序列的输入数据的回文序列. 在EM过程中, MEME将一同平均letter频率, in complementary columns of the motif(PSPM). 例如, 如果宽度为10, 第一列尾10, 第二列为9, 第三列为8, 将其平均起来; 针对DNA, 该平均过程将会包含所有碱基在一列中的频率; 默认在motif的互补列中不会平均letter频率



由于没有指定宽度width, MEME自动选择motif的最佳宽度width

2. 在双链搜索DNA序列

```
meme crp0.s -dna -mod oops -revcomp
```

选项-revcomp指定MEME在(in a completable datasets)输入链以及反向互补链搜索motif; 默认仅搜索给定链

3. 快速搜索DNA

```
meme crp0.s -dna -mod oops -revcomp -w 2
```

仅考虑宽度为20的motif

4. 使用higher-order 背景模型

```
meme INO_up800.s -dna -mod anr -revcomp -bfile yeast.nc.6.freq
```

参数-mod anr, 指定每个序列中motif可出现任意次数

参数-bfile, Markov model指定使用yeast.nc.6.freq用于背景模型. bfile, the name of Markov background model file, MEME使用背景模型来标准化letters分布偏差, 同时将序列中letters分组. 0-order模型用于调整单个letter偏差; 1-order模型用于调整dimer偏差(例如, DNA序列中的GC含量); 假如-markov_order选项给定了, MEME仅读取bfile中Markov模型中给定的order, 忽略更高order输入.

Example Background models

0-order DNA Markov model

#	order 0
a	0.324
c	0.176
g	0.176
t	0.324

1st-order DNA Markov model

#	order 0
A	2.563e-01
C	2.437e-01
G	2.437e-01
T	2.563e-01

#	order 1
AA	7.020e-02
AC	5.388e-02
AG	8.888e-02

使用higher order背景模型会带来更高的检出敏感性. 因为背景模型为更准确的非motif序列, 使MEME在搜索序列时避免这些moitfs.

5. 简单蛋白例子

```
meme lipocalin.s -mod oops -maxw 20 -nmotifs 2
```

不指定-dna, MEME假设输入为蛋白序列.

-nmotifs指定搜索2个motifs.

-maxw 每个motif的宽度小于或等于20.

-mod oops 每个序列中motif出现一次

6. 另一蛋白例子

```
meme farntrans5.s -mod anr -maxw 40 -maxsites 50
```

-maxw 40 /-maxsites 50 在整个输入序列数据中搜索宽度最多为40, 出现次数最多为50的motifs

7. 快速蛋白搜索

```
meme farntrans5.s -mod anr -w 10 -maxsites 30 -nmotifs 3
```

-nmotifs 指定搜索moitfs数目为3, -w 10宽度为10

8. 将sites分成3部分

```
meme farntrans5.s -mod anr -maxw 12 -nsites 24 -nmotifs 3
```

-nsites 24/-maxw 12 指定每一个moitf出现24次, 且之多12个字符宽

9. 包含E-value阈值的更大蛋白

```
meme adh.s -mod zoops -nmotifs 20 -evt 0.01
```

-nmotifs 20/-evt 0.01 查询至多20个motif, 但是当个motif的E-value大于0.01时, 搜索终止.

[DREME][http://meme-suite.org/doc/dreme.html?man_type=web]

DREME, Discriminative Regular Expression Motif Elicitation, DREME搜索短的, 无gap的 motifs(recurring, fixed-length patterns). 搜索motifs(至多8个位置)相对较短, 对照序列应该和处理序列长度相近, 如果没有提供对照序列, 程序将打乱处理序列作为对照序列. 该软件使用Fisher' Exact Test来判断搜索到的序列的显著性, 可在使用时设置显著性阈值.

Input

包含序列的FASTA格式文件, 且所有输入序列应几乎一样长

Output

如果未通过 `-o` / `-oc` 指定, 那么输出到dreme_out目录

- dreme.html 可读的交互式HTML输出格式
- deme.txt 文本格式
- deme.xml 用于machine processing的XML格式文件
- NAMEnc_CONSENSUS.png DREME搜索到的motifs的序列logo PNG图片
- NAMErc_CONSENSUS.png 为motifs的反向互补序列

Options

`-p` primary sequence file

`-n` control sequence file

`-dna/-rna/-protein/-alph` 输入文件格式

`-norc` 仅搜索指定的primary文件; 默认搜索输入序列和其反向互补序列(when the alphabet is complementable)

`-g` 设置指定数目REs; 默认100REs

`-e/-m/-t` 指定停止搜索时条件, motif's E-value > e/ m motifs/ t seconds; 默认为e为0.05, 其余无

`-mink/-maxk/-k` 设置最小/最大/固定motif core宽度, 默认为3-7

[MEME-ChIP][http://meme-suite.org/doc/meme-chip.html?man_type=web]

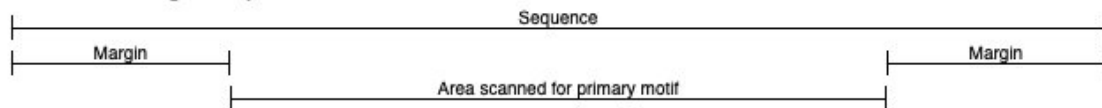
```
meme-chip [options] [-db <motif file>]* <primary sequence file>
```

MEME-ChIP针对一组大的序列, 例如ChIP-seq/CLIP-seq实验获得的, 进行综合性的motif分析(包含motif搜索). 注意, 输入序列应该在中心的100字符区域内包含motifs(the input sequences should be centered on a 100 character region expected to contain motifs).

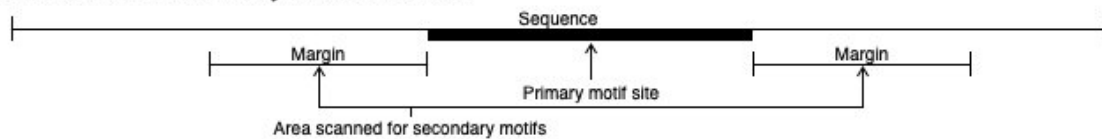
MEME-ChIP:

1. 在输入序列默认中心区域100bp内发现新的DNA结合motifs(with MEME and DREME)
2. 判断哪些motifs是最集中富集的(with CentriMo)
3. 分析相似性以获得结合的motifs(with Tomtom)
4. 同时, 根据相似性自动将显著性motifs聚集
5. 执行motif spacing分析(with SpaMo)

Restrictions on primary motif locations.



Restrictions on secondary motif locations.



6. 构建GFF文件用于基因组范围查看

Input

```
[-db <motif file>]*
```

推荐的可选参数, 一个或多个包含MEME formatd motifs的文件. 支持来自MEME/DREME的输出, 同时还有Minimal MEME Format. 也可以通过[脚本][http://meme-suite.org/doc/overview.html#motif_conversion_utilities]将任何motifs格式转换为MEME Motif格式. 这些motif文件将用于Tomtom和CentriMo

```
<primary sequence file>
```

FASTA格式序列文件. 理想条件下, 所有序列应该都是一样长, 为100-500bp且motifs位于序列中心区域. 理想输入为, 来自转录因子(TF) ChIP-seq实验的单个ChIP-seq 'peak'的临近区域. 建议采用的最小100bp为一般ChIP-seq peaks的典型分辨率情况, 含有跟多的围绕序列有利于CentriMo判断motifs是否中心富集. 推荐, 'repak mask'输入序列, 同时使用'N'代替重复序列

Output

输出为memechip_out文件夹.

- meme-chip.html 提供可读交互式网页HTML文件
- summary.tsv 可用于脚本解析和Excel查看的TSV结果格式
- combined.meme 包含MEME Motif Format的文本文件

Options

`-db` 使用MEME Motif Format格式的DNA motifs文件. 该文件将用于Tomtom和CentriMo分析, 可指定多次传递多个文件; When no files are provided, Tomtom can't suggest similar motifs and CentriMo is limited to the discovered motifs.

`-neg` 查询primary序列相对于该输入文件(fasta 格式)的moitfs. 这些序列将会用做control序列输入到MEME, DREME和CentriMo. MEME使用其'Differential Enrichment' objective函数. 当使用该选项时, primary和控制序列应该是相同长度; 否则CentriMo E-values将不会准确. 如果, primary序列为转录因子ChIP-seq peak区域, 类似的来自knockout细胞系或生物的区域为一可能的control序列选择; 默认, 没有control序列用于MEME或CentriMo, MEME使用 `classic` objective函数. 针对DREME, positive sequences大乱, 保留双核苷酸频率, 用于构建control set.

`-psp-gen` 使[pop-gen][<http://meme-suite.org/doc/psp-gen.html>]构建[position-specific prior][<http://meme-suite.org/doc/psp-format.html>]用于MEME分析. MEME将使用该先验值和'Classic objective'函数, 而不是'Differential Enrichment' objective function; 默认MEME通过'Differential Enrichment' 直接采用control序列, 不构建position-specific prior. 注意, 该选项需要 `-neg` 参数

`-ccut` 针对MEME和DREME输入, 针对其中心区域nbp裁剪序列.(全长序列输入至CentriMo和SpaMo). 0表示不裁剪; 默认为100

`-norc` 仅在给定序列链搜索; 默认给定序列链及其互补链

MEME Specific Options

`-meme-minw/-meme-maxw` 最小最大motif宽度; 默认为6-30

`-meme-nmotifs` MEME搜索motifs数目, 默认为3, 指定为0将不运行MEME

`-meme-minsites/-meme-maxsites` 针对一个motif, MEME需要满足的最小/最大查询到的位置数目; 默认没限制

`-meme-pal` 限制MEME仅搜索回文序列

DREME Specific Options

`-dreme-e/-dreme-m` 停止搜索时的最大E-value和数目

CentriMo Specific Options

`-centrimo-local` CentriMo执行local motif富集分析, 计算每一个可能序列区域的富集情况; 默认仅执行central motif富集分析, 针对中心区域计算富集情况

`-centrimo-score` 接受匹配的最小值; 默认为5

`-centrimo-maxreg` 最大检测区域值; 默认为所有有效区域

`-centrimo-ethresh` 满足富集的中心区域的E值; 默认为10

SpaMo Specific Options

`-spamo-skip` 跳过SpaMo

FIMO Specific Options

`-fimo-skip` 跳过FIMO

Example

根据PacBio三代测序获得的甲基化结果(motifs), 查询对应的methyltransferase基因. 使用程序: `fimo`

FIMO, find individual motif occurrences, 搜索一套序列查询已知的motif, 且独立对待每一个motif. motif必须为[MEME Motif Format][<http://meme-suite.org/doc/meme-format.html>].

FIMO将所有输入的motif转换为log-odds PSSM, 且使用独立使用每一PSSM搜索每一个输入序列. 针对一个motif输出每一条序列满足统计显著性log-odds值的所有匹配位置. 可以通过匹配的p值, 和是否输出正负链的输出.

FIMO支持DNA, RNA和蛋白, 以及定制的序列alphabets搜索. 该alphabets通过motif文件指定, 且搜索的序列必须和该alphabets兼容. (DNA motifs可用于搜索RNA序列, 反之亦然)

MEME Motif Format

存在两种MEME Motif Formats:

MEME and DREME Output Formats

Minimal MEME Motif Format

注意: 针对**DNA minimal motif format**, 参数 **strands:** 将用于指示搜索的链是否为给定链和反向互补链. 如果该行信息未提供, 将默认为搜索双链. 若仅指定给定链为: **strands: +**

minimal DNA motif

```
MEME version 4
ALPHABET= ACGT
strands: + -
Background letter frequencies
A 0.303 C 0.183 G 0.209 T 0.306
MOTIF crp
letter-probability matrix: alength= 4 w= 19 nsites= 17 E= 4.1e-009
0.000000 0.176471 0.000000 0.823529
0.000000 0.058824 0.647059 0.294118
0.000000 0.058824 0.000000 0.941176
0.176471 0.000000 0.764706 0.058824
0.823529 0.058824 0.000000 0.117647
0.294118 0.176471 0.176471 0.352941
0.294118 0.352941 0.235294 0.117647
0.117647 0.235294 0.352941 0.294118
0.529412 0.000000 0.176471 0.294118
0.058824 0.235294 0.588235 0.117647
```

minimal Protein motif

```
MEME version 4
ALPHABET= ACDEFGHIKLMNPQRSTVWY
Background letter frequencies
A 0.071 C 0.029 D 0.069 E 0.077 F 0.043 G 0.057 H 0.026 I 0.048 K 0.085
L 0.087 M 0.018 N 0.053 P 0.032 Q 0.029 R 0.031 S 0.058 T 0.048 V 0.069
W 0.017 Y 0.050
MOTIF lipo_1
letter-probability matrix: alength= 20 w= 26 nsites= 5 E= 5.0e-006
0.0 0.2 0.0 0.0 0.0 0.4 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0
0.6 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.2
0.0 0.4 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.0
0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.2 0.0 0.0 0.0
0.2 0.0 0.2 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.2 0.0 0.0
0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.2 0.2 0.4 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.0 0.2 0.2 0.2 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.4 0.2 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.2 0.0 0.0 0.2 0.0 0.2 0.0 0.0 0.0 0.0 0.0 0.4 0.0 0.0 0.0
```

letter-probability matrix: alength= alphabet length w= motif length nsites= source sites E= source E-value ... (letter-probability matrix goes here) ...

[MEME Suite][http://meme-suite.org/doc/motif_conversion.html]包含多种程序可将其他motif 格式转换为Minimal MEME Motif Format

最简单的采用 `iupac2meme` 将该格式转换为Minimal MEME Motif Format

iupac2meme

Input is string(s) on the command line.

DNA motif

ACGGWNNYCGT

Protein motif

IKLVBZYXXHG

iupac2meme, 将consensus序列转为MEME motif format的motif文件。

consensus序列为IUPAC [alphabets][<http://meme-suite.org/doc/alphabet-format.html>]的核酸或蛋白序列, 后自定义的alphabet. 该程序还支持正则表达式, 多重的letters可包含在方括号内. 同时, 被取消的字符也可以通过使用脱字符表示(^).

-numseqs 指定consensus序列每一位置字符的计数情况(assume frequencies based on count sequence sites)

-bg 指定背景文件([markov backgroud model][<http://meme-suite.org/doc/bfile-format.html>]). 包含了用于指定pseudocounts的字符背景频率. 该背景频率将包含在输出的MEME文件中.

-pseudo Add total pseudocounts times letter background to each frequency

-logodds 输出包含log-odds矩阵

Examples

DNA IUPAC motif:

ACGGWN[ACGT]YCGT

protein IUPAC motif:

IKLVB[[^]ILVM]ZYXXHG

另外可以通过指定 -alph 参数, [alphabet file][<http://meme-suite.org/doc/alphabet-format.html>]来指定比对规则(例如增加对应的甲基化特征等)

1) Standard DNA alphabet

```
ALPHABET "DNA" DNA-LIKE

# Core symbols
A "Adenine" CC0000 ~ T "Thymine" 008000
C "Cytosine" 0000CC ~ G "Guanine" FFB300

# Ambiguous symbols
U = T # alias Uracil to Thymine (permit U in input sequences)
R = AG
Y = CT
K = GT
M = AC
S = CG
W = AT
B = CGT
D = GAT
H = ACT
V = ACG
N = ACGT # wildcard symbol
```

根据实际目的将输入序列和alphabet file文件修改, 例如针对甲基化比对, 将序列中对应的m6A修改为X, 同时在构建MEME Motif Format前也将对应的m6A修改为X


```

1 ALPHABET
2
3 # Core symbols
4 A "Adenine" CC0000
5 T "Thymine" 008000
6 C "Cytosine" 0000CC
7 G "Guanine" FFB300
8 X "Methylated Adenine"
9
10 # Ambiguous symbols
11 U = T # alias Uracil to Thymine (permit U in input sequences)
12 R = AG
13 Y = CT
14 K = GT
15 M = AC
16 S = CG
17 W = AT
18 B = CGT
19 D = GAT
20 H = ACT
21 V = ACG
22 N = ACGT # wildcard symbol

```

```
iupac2meme -alph ../alphabet_methylated_m6a.dna * > meme_m6a_motif_pattern
```

[FIMO Parameters][http://meme-suite.org/doc/fimo.html?man_type=web]

该软件使用动态算法将log-odds值转换为p-values, 针对给定序列或其反向互补序列, FIMO给出对应的统计阈值(p-value). 默认软件报告p-value小于1e-4的motifs, 可通过 `--thresh` 选项设置.

每个motif发生的p-value经过Benjamini and Hochberg转换后得到q-values(q-value, 最小的假阳性率).

`--qv-thresh` 指导软件使用q-values而不是p-values

如果motif的strand特征为 `+/-` (rather than `+`), 软件将使用双链来搜索motif. 这个可在alphabet文件内修改

`--text` 显著输出为tsv(tab-separated values), 该结果未排序且没有q-values, 用于很大的文件搜索

`--norc` 仅输出给定链而不考虑反向互补链的结果

`--prior-dist` 包含priors的binned分布的文件, 该文件可通过[create-priors][<http://meme-suite.org/doc/create-priors.html>]生成

`--psp` 包含位置特异性先验值的文件(position specific priors, PSP), 格式为[MEME PSP format][<http://meme-suite.org/doc/psp-format.html>]或[wiggle format]

[<http://genome.ucsc.edu/goldenPath/help/wiggle.html>]. 该参数会导致具有高的先验值的位置获得更高的motif分值. 同样使用create-priors生成. 同时需要参数 `--prior-dist`

注意: 在使用**minimal motif format**时, 发现软件仅识别**DNA/RNA/PROTEIN alphabets**, 因此需要在此基础上修改**format motif**文件

Strands line

Strands line (optional)

The strands line only has meaning for motifs of complementable alphabets like DNA and indicates if motifs were created from sites on both the given and the reverse complement strands of the sequences. If this line is not supplied then the MEME Suite will assume that motifs of complementable alphabets were created from both strands.

```
strands: which strands
```

The `which strands` can be replaced with `+` to indicate only the given strand and `+` - to indicate both strands.

For example to indicate only the given strand was used:

```
strands: +
```

Background frequencies lines

Background frequencies lines (recommended)

The background frequencies tell the MEME Suite how prevalent each letter of the motif alphabet was in the source sequences that were used to create the motifs. If the background frequencies are not supplied then the MEME Suite will assume uniform background frequencies. The MEME Suite uses this background to automatically create the log-odds matrices for MAST.

```
Background letter frequencies (from source):  
letter 1 frequency 1 letter 2 frequency 2 ... (repeated) ... letter n-1  
frequency n-1 letter n frequency n
```

或者在fimo前手动修改Background letter frequencies信息, 增加搜索特异性; 建议根据patterns模式, 使用fasta-get-markov构建. 但是实际操作中意义不大.

```
1 MEME version 4  
2  
3 ALPHABET ""  
4 A "Adenine" CC0000  
5 C "Cytosine" 0000CC  
6 G "Guanine" FFB300  
7 T "Thymine" 008000  
8 X "Methylated Adenine"  
9 ? = ACGTX  
10 N = ACGT  
11 V = ACG  
12 H = ACT  
13 D = AGT  
14 B = CGT  
15 M = AC  
16 R = AG  
17 W = AT  
18 S = CG  
19 Y = CT  
20 K = GT  
21 U = T  
22 END ALPHABET  
23  
24 Background letter frequencies (from uniform background):  
25 A 0.20000 C 0.20000 G 0.20000 T 0.20000 X 0.20000  
26  
27 MOTIF Aba0088II.txt  
28  
29 letter-probability matrix: alength= 5 w= 6 nsites= 20 E= 0  
30 1 0.000000 0.000000 0.000000 0.000000 0.000000
```

这里尝试使用位置先验概率帮助增加搜索特异性(针对需要搜索的fasta序列构建):

首先根据motif序列情况, 使用create-priors构建 `--psp` 和 `--prior-dist` 文件

根据create-priors说明构建输入序列文件和wiggle文件, 其中文件名称需要对应, wiggle文件中的序列名称需要包含于sequence文件:

Input

<sequence file>

The name of a file the name of a file containing **FASTA formatted sequences**.

<wiggle file>

The name of a file in **wiggle format** containing the scores that will be converted to priors.

这里wiggle文件中甲基化位点值定位10, 其余为1:

```
create-priors Total.seqs all_seqs_wiggle.txt
```

再根据构建的motif pattern搜索

```
fimo --psp create-priors_out/priors.wig --psp create-priors_out/priors.dist  
meme_m6a_motif_pattern Total.seqs
```

Wiggle Track Format(WIG)

bigwig是几乎图形化轨迹所需的格式([wiki page]

[http://genomewiki.ucsc.edu/index.php/Selecting_a_graphing_track_data_format]). wiggle(WIG)格式是旧式的用于展示dense, 例如GC比例, 概率值, 和转录组数据的连续数据. Wiggle数据单元必须相等大小(equally sized). bedGraph格式也是一个用于展示稀疏数据或包含不同大小数量的数据的旧式格式.

为了速度和效率, wiggle数据会压缩并存储在128个独立的bins中. 该压缩意味当把数据从一个wiggle轨迹中导出时其精确度会丢失(i.e. with output format 'data points' or 'bed format'). 若在导出时需要保存精确的原始数据, 需要使用bedGraph格式.

Wiggle format is line-oriented. 针对wiggle的定制轨迹, 第一行必须为[track definition line]

[<http://genome.ucsc.edu/goldenPath/help/customTrack.html#TRACK>](i.e. track type=wiggle_0), 该行指定track为wiggle track, 同时增加一系列控制默认显示的选型.

Wiggle格式由声明行和数据行组成, 同时需要一个单独的wiggle track定义行. 存在两种选择供用户构建wiggle数据: **variableStep**和**fixedStep**.

variableStep format 该格式用于非规则的intervals数据(between new data points), 且时更常用的wiggle格式. 经过wiggle track定义行后, variableStep开始于一个声明行, 随后就是两列包含染色体位置和数据值:

```
variableStep chrom=chrN  
[span=windowSize]  
chromStartA dataValueA  
chromStartB dataValueB  
... etc ... .. etc ...
```

声明行起始于**variableStep**, 伴随着为染色体的指定信息. 可选参数**span**(默认: span=1)允许数据由连续的碱基runs组成. span开始于染色体指定的位置和该位置应该包含的碱基数目:

```
variableStep chrom=chr2
300701 12.5
300702 12.5
300703 12.5
300704 12.5
300705 12.5
```

is equivalent to:

```
variableStep chrom=chr2 span=5
300701 12.5
```

注意: wiggle格式用于快速展示比较dense的数据, 如果数据比较稀疏, 建议使用[BedGraph]
[<http://genome.ucsc.edu/goldenPath/help/bedgraph.html>]格式.

fixedStep format

该格式用于数据具有规则的interval(between new data values), 是更加紧凑的wiggle格式. 首先定义wiggle track行, fixedStep行开始于一个声明行, 随后为单列的数据点:

```
fixedStep chrom=chrN  
start=position step=stepInterval  
[span=windowSize]  
  dataValue1  
  dataValue2  
  ... etc ...
```

声明行起始于fixedStep, 随后为指定的染色体, 染色体起点, 步长. 同时, span具有和variableStep格式一样的作用:

```
fixedStep chrom=chr3 start=400601 step=100
11
22
33
```

显示值为11/22/33作为chromosome 3单碱基位置400601/400701/400801. 加入span=5:

```
fixedStep chrom=chr3 start=400601 step=100 span=5
11
22
33
```

显示值11/22/33为chromosome 3以5碱基为一区间: 400601~400605, 400701~400705, 400801~499805.

注意: 针对variableStep和fixedStep格式, 相同的span必须贯穿整个数据集. 如果没有指定span, 使用默认span: 1. 同时, 针对fixedStep, wiggles需要相同的step贯穿整个数据集, 若没指定step, 使用默认step: 1.

Data values Wiggle track数据值为整数或实际数值, 正数或负数. 且仅指定的位置罕有数值. 没有指定的位置不含有数值将不会展示再图形中. 所有指定在输入数据中的数据必须是numerical order. 不支持NaN值.

BigWig文件使用"1-start coordinate"系统; 而bedGraph支持"0-start, half-open" coordinates.

Parameters for custom wiggle track definition lines 所有定义行都放置于一行且相互直降使用空格分开, 可选内容如下:

```
track type=wiggle_0 name=track_label
description=center_label
visibility=display_mode color=r,g,b
altColor=r,g,b priority=priority
autoScale=on/off alwaysZero=on/off
gridDefault=on/off
maxHeightPixels=max:default:min
graphType=bar/points
viewLimits=lower:upper
yLineMark=real-value yLineOnOff=on/off
windowingFunction=mean+whiskers/maximum/mean/minimum
smoothingWindow=off|2-16
```

具有版本号的track type是必须的, 当前必须为"wiggle_0":

```
type wiggle_0
```

剩余值可选:

name	<trackLabel>	# default is "User Track"
description	<centerLabel>	# default is "User Supplied Track"
visibility	<full/dense/hide>	# default is hide
color	<RRR,GGG,BBB>	# default is 255,255,255
altColor	<RRR,GGG,BBB>	# default is 128,128,128
priority	<N>	# default is 100

Examples 该例子中指定了19个分散数据点, 两个tracks(name="variableStep", name="fixedStep", "priority=" 进行排序)

```
browser position chr19:49304200-49310700
browser hide all
# 150 base wide bar graph at arbitrarily spaced positions,
# threshold line drawn at y=11.76
# autoScale off viewing range set to [0:25]
# priority = 10 positions this as the first graph
# Note, one-relative coordinate system in use for this format
track type=wiggle_0 name="variableStep" description="variableStep format" visibility=full autoScale=off
variableStep chrom=chr19 span=150
49304701 10.0
49304901 12.5
49305401 15.0
49305601 17.5
49305901 20.0
49306081 17.5
49306301 15.0
49306691 12.5
49307871 10.0
# 200 base wide points graph at every 300 bases, 50 pixel high graph
# autoScale off and viewing range set to [0:1000]
# priority = 20 positions this as the second graph
# Note, one-relative coordinate system in use for this format
track type=wiggle_0 name="fixedStep" description="fixedStep format" visibility=full autoScale=off
fixedStep chrom=chr19 start=49307401 step=300 span=200
1000
900
800
700
600
500
400
300
200
100
```

