

第9回バイオインフォマティクス研究会

Web tool FIMOを用いた既知モチーフ検索

2014年2月24日

中林潤

FIMO

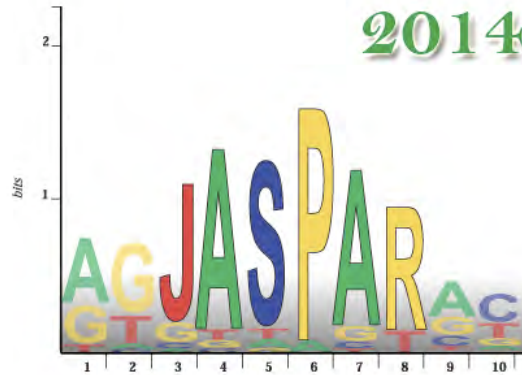
- Find Individual Motif Occurrences
- Position specific scoring matrixとして記述された既知のモチーフを、配列中から検索するツール
- <http://meme.sdsc.edu>
MEME suite の一部として利用可能

position specific scoring matrix

MEME version 4			
ALPHABET=ACGT			
strands:+ -			
MOTIF motif.name.1			
0	0	1	0
1	0	0	0
1	0	0	0
1	0	0	0
0	0	1	0
0	0	0	1

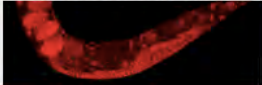
JASPAR database

Version: 5.0_ALPHA You are using the JASPAR development server: jaspardev.genereg.net Visit the stable JASPAR server: jaspar.genereg.net



The high-quality transcription factor binding profile database

Browse the JASPAR CORE database directly:

 JASPAR CORE Vertebrata	 JASPAR CORE Nematoda	 JASPAR CORE Insecta
 JASPAR CORE Plantae	 JASPAR CORE Fungi	 JASPAR CORE by Structural Class

[DOCUMENTATION](#) [DOWNLOAD](#) [TOOLS](#) [CONTACT](#)

<http://jaspardev.genereg.net/>

BRCA1 motif

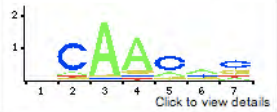
http://jaspar.bmr.ku.dk/cgi-bin/jaspar_do.pl

アップル Yahoo! Japan Google マップ YouTube Wikipedia ニュース (1171) お役立ち

http://bioinformatics.oxfordjourn... OS X: スクリーンショットを撮る方法 The JASPAR database

SEARCH Name AND Species AND Class SEARCH ?

JASPAR matrix models:

TOGGLE	ID	name	species	class	family	Sequence logo
☐	MA0133.1	BRCA1	9606	Other	Other	

Click to view details

ANALYZE selected matrix models:

CLUSTER ? selected models using STAMP

Create RANDOM matrix models based on selected models

Number of matrices: 200 Format: Raw

RANDOMIZE ?

Create models with PERMUTED columns from selected:

Type: Within each matrix Format: Raw

PERMUTE ?

SCAN this (fasta-formatted) sequence with selected matrix models

Relative profile score threshold 80 %

BRCA1

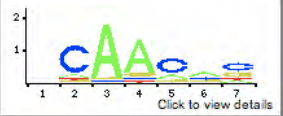
http://jaspar.binf.ku.dk/cgi-bin/jaspar_db.pl

アップル Yahoo! Japan Google マップ YouTube Wikipedia ニュース (1171) お役立ち

http://bioinformatics.oxfordjourn... OS X : スクリーンショットを撮る方法 The JASPAR database

SEARCH Name AND Species AND Class

JASPAR matrix models:

TOGGLE	ID	name	species	class	family	Sequence logo
☐	MA0133.1	BRCA1	9606	Other	Other	

Click to view details

CLUSTER

Create RAND

Number of ma

Create models

Type: Within ea

SCAN this (fas

models

Relative profile score threshold 80 %

Summary page for ID: MA0133.1 NAME: BRCA1 from the JASPAR CORE database

DATA

Name	BRCA1
Class	Other
Family	Other
Species	Homo sapiens
Taxon	vertebrates
ACC	P38398
Type	SELEX
MEDLINE	14502648
Pazar ID	TF0000825
TFBSshape ID	128
TFencyclopedia ID	-
Comment	Multi-protein complex

VERSION INFORMATION

There is only one version of the model.

SITES

Show me all the binding sites

...as web page

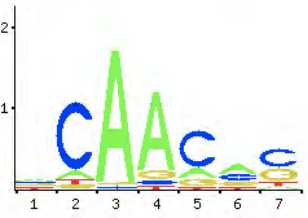
..as fasta file

..as bed file

ChIP-seq Centrality

There is no centrality information of the model.

SEQUENCE LOGO



Make a SVG logo

FREQUENCY MATRIX

A	[15	4	41	36	7	19	3]
C	[11	35	1	2	29	14	22]
G	[10	2	1	4	6	7	15]
T	[7	2	0	1	1	3	3]

Reverse complement

TF INFORMATION

TFencyclopedia

PDB Structure

TFBS PROFILES

TFBSshape

BRCA1 Matrix

A	15	4	41	36	7	19	3
C	11	35	1	2	29	14	22
G	10	2	1	4	6	7	15
T	7	2	0	1	1	3	3

BRCA1 Matrix

A	G	C	T
0.348837209	0.255813953	0.23255814	0.162790698
0.093023256	0.813953488	0.046511628	0.046511628
0.931818182	0.022727273	0.022727273	0.022727273
0.837209302	0.046511628	0.093023256	0.023255814
0.162790698	0.674418605	0.139534884	0.023255814
0.441860465	0.325581395	0.162790698	0.069767442
0.069767442	0.47826087	0.348837209	0.069767442

BRCA1 Matrix

MEME version 4			
ALPHABET=ACGT			
strands:+ -			
MOTIF BRCA1			
letter-probability matrix:	alength=4 w=7		
0.348837209	0.255813953	0.23255814	0.162790698
0.093023256	0.813953488	0.046511628	0.046511628
0.931818182	0.022727273	0.022727273	0.022727273
0.837209302	0.046511628	0.093023256	0.023255814
0.162790698	0.674418605	0.139534884	0.023255814
0.441860465	0.325581395	0.162790698	0.069767442
0.069767442	0.47826087	0.348837209	0.069767442

FAST sequence

>mm9_ct_UserTrack_3545_Ccnyl1

AATTTTAAATTGAAGTAACAAGCCATTAATTTTCAGTACTAGCTGCTGAT
TTCCCATCATATAGGTAAAAGTTGTTTCTACTGTGACCATCATTTTCCTTC
TTCCATATCCTTTTTTTTTTTTTTAAATCATGTCAGTGTTTCATTATGACTATG
CACATAGTTAGTTGCTGTCTTTTGTCTTCAGTGGTTTTGAATCTGAATTC
TTCATGCCCTCTGGATGGATAACTTTTCCTGGCCCTATGAGC

>Sequence_Name

配列

FIMO

http://meme.nbcr.net/meme/

アップル Yahoo! Japan Google マップ YouTube Wikipedia ニュース (1171) お役立ち

http://bioinformatics.oxfordjourn... OS X: スクリーンショットを撮る方法 MEME Suite - Introduction

MEME Suite Menu

- Submit A Job
- Documentation
- Downloads
- User Support
- Alternate Servers
- Authors
- Citing

The MEME Suite

Motif-based sequence analysis tools

Unaligned sequences

MEME
DREME
MEME-ChIP
GLAM2

Motifs

Motif database

TOMTOM → Aligned motifs

CentriMo
AME
SpaMo → Motif enrichment

Sequence database

FIMO
MAST
MCAST
GLAM2SCAN → Annotated sequences

GO database

GOMO → Annotated motifs

GO function
GO compartment
GO process

MEME De-novo Motif Discovery
DREME Discriminative DNA Motif Discovery
MEME-ChIP Motif Analysis of Large DNA Datasets
GLAM2 Gapped Local Alignment of Motifs
TOMTOM Motif Comparison
CentriMo Central Motif Enrichment Analysis
AME Analysis of Motif Enrichment
SpaMo Spaced Motif Analysis
FIMO Find Individual Motif Occurrences
MAST Motif Alignment and Scan Tool
MCAST Motif Cluster Alignment and Search Tool
GLAM2SCAN Scanning with Gapped Motifs
GOMO Gene Ontology for Motifs

The MEME Suite allows you to:

- discover motifs using **MEME**, **DREME** (DNA only) or **GLAM2** on groups of related DNA or protein sequences,
- search sequence databases with motifs using **MAST**, **FIMO**, **MCAST** or **GLAM2SCAN**,
- compare a motif to all motifs in a database of motifs,
- associate motifs with Gene Ontology terms via their putative target genes, and
- analyse motif enrichment using **SpaMo** or **CentriMo**.

To submit a query, click on one of the logos below or select "Submit A Job" from the menu at the left.

MEME Suite Menu (intro.html)

FIMO

http://meme.nbcr.net/meme/

アップル Yahoo! Japan Google マップ YouTube Wikipedia ニュース (1171) お役立ち

http://bioinformatics.oxfordjourn... OS X: スクリーンショットを撮る方法 MEME Suite - Introduction

AME Analysis of Motif Enrichment
SpaMo Spaced Motif Analysis
FIMO Find Individual Motif Occurrences
MAST Motif Alignment and Scan Tool
MCAST Motif Cluster Alignment and Search Tool
GLAM2SCAN Scanning with Gapped Motifs
GOMO Gene Ontology for Motifs

GO database

GOMO → Annotated motifs

GO function
GO compartment
GO process

The MEME Suite allows you to:

- discover motifs using [MEME](#), [DREME](#) (DNA only) or [GLAM2](#) on groups of related DNA or protein sequences,
- search sequence databases with motifs using [MAST](#), [FIMO](#), [MCAST](#) or [GLAM2SCAN](#),
- compare a motif to all motifs in a database of motifs,
- associate motifs with Gene Ontology terms via their putative target genes, and
- analyse motif enrichment using [SpaMo](#) or [CentriMo](#).

To submit a query, click on one of the logos below or select "Submit A Job" from the menu at the left.

MEME
Multiple Em for Motif Elicitation

MAST
Motif Alignment & Search Tool

TOMTOM
Motif Comparison Tool

GOMO
Gene Ontology for Motifs

GLAM2
Gapped Local Alignment of Motifs

GLAM2SCAN
Scanning with Gapped Motifs

FIMO
Find Individual Motif Occurrences

MCAST
Motif Cluster Alignment and Search Tool

MEME-ChIP
Motif Analysis of Large DNA Datasets

SPAMO
Spaced Motif Analysis Tool

DREME
Discriminative DNA Motif Discovery

CentriMo
Central Motif Enrichment Analysis

Maintenance and development of the MEME Suite is funded by the National Center for Research Resources grant NIH/NCRR R01 RR021692. The MEME Suite web server is funded by the [National Biomedical Computation Resource](#).

Developed and maintained by:

 THE UNIVERSITY OF QUEENSLAND

 UNIVERSITY OF WASHINGTON

 UCSD

 SDSC

MEME Suite Menu (intro.html)

FIMO

http://meme.nbc.net/meme/cgi-bin/fimo.cgi

アップル Yahoo! Japan Google マップ YouTube Wikipedia ニュース (1171) お役立ち

http://bioinformatics.oxfordjourn... OS X: スクリーンショットを撮る方法 FIMO - Submission form

MEME Suite Menu

- Submit A Job
- Documentation
- Downloads
- User Support
- Alternate Servers
- Authors
- Citing

FIMO
Find Individual Motif Occurrences

Version 4.9.1

Use this form to submit motifs to FIMO to be used in searching sequence database.

Data Submission Form

Required

Your e-mail address:

Re-enter e-mail address:

Your motif file:
 no file selected
Sample DNA minimal motif format file.

Sequence database to search--select one of the following:
A supported database:
Category:

Database:

or
Your FASTA sequence file (1000000 sequence characters maximum):
 no file selected

Sample DNA database.

Optional

p-value output threshold

☐ Scan given strand only

Description of your motifs:

FIMO Custom Tracks For the UCSC Genome Browser

The links below display custom tracks for FIMO in the UCSC Genome Browser. The tracks display

MEME Suite Menu ../intro.html

メールアドレス

position specific scoring matrix file

sequence data set

motif & data set

- BRCA1 motif: “ACAACAC”
- 130 配列 31460塩基

result

http://nbc-222.ucsd.edu/opal-jobs/appFIMO_4.9.11393208066262-1913314707/nmo.html

http://bioinformatics.oxfordjourn... 「若者の献血離れ」 こうすれば食い止... FIMO Results

For further information on how to interpret these results or to get a copy of the FIMO software please access <http://meme.nbc.net>

If you use FIMO in your research, please cite the following paper:
Charles E. Grant, Timothy L. Bailey, and William Stafford Noble, "FIMO: Scanning for occurrences of a given motif", *Bioinformatics*, **27**(7):1017-1018, 2011.

DATABASE AND MOTIFS

DATABASE test_sequence.txt
Database contains 130 sequences, 31460 residues

MOTIFS PSS_BRCA1.txt (nucleotide)

MOTIF	WIDTH	BEST POSSIBLE MATCH
BRCA1	7	ACAACAC

Random model letter frequencies (from non-redundant database):
A 0.275 C 0.225 G 0.225 T 0.275

SECTION I: HIGH-SCORING MOTIF OCCURENCES

- There were 2 motif occurrences with a p-value less than 0.0001.
- The p-value of a motif occurrence is defined as the probability of a random sequence of the same length as the motif matching that position of the sequence with as good or better a score.
- The score for the match of a position in a sequence to a motif is computed by summing the appropriate entries from each column of the position-dependent scoring matrix that represents the motif.
- The q-value of a motif occurrence is defined as the false discovery rate if the occurrence is accepted as significant.
- The table is sorted by increasing p-value.

Motif	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
BRCA1	mm9_ct_UserTrack_3545_Rc3h1	+	134	140	6.51e-05	1	ACAACAC
BRCA1	mm9_ct_UserTrack_3545_Slc40a1	-	92	98	6.51e-05	1	ACAACAC

DEBUGGING INFORMATION

Command line:

```
/opt/meme_4.9.1/bin/fimo --oc . --verbosity 1 --thresh 0.0001 PSS_BRCA1.txt test_sequence.txt
```

Settings:

result

Motif	Sequence Name	Strand	Start	End	p-value	q-value	Matched Sequence
BRCA1	mm9_ct_UserTrack_3545_Rc3h1	+	134	140	6.51e-05	1	ACAACAC
BRCA1	mm9_ct_UserTrack_3545_Slc40a1	-	92	98	6.51e-05	1	ACAACAC

result

>mm9_ct_UserTrack_3545_Rc3h1

TAAATGAGTTTGATGAATATTTTCTGAAGAGCCCCAGTGTCATGTGA
TCATCTGCTGTTACCATCTGTTTGCACAAAGAGATTTTGCAGTTTTGC
AGGAAGTGCATCTAATTGGTATATTTTCACTTCCTTAA**ACAACACCCA**
TTATCACTTGCAAGCTGATTATTCCTAGTCTTTGTTTTGTATTTTCAAC
ATGTAGACTACCGCAGGGAGAGCTTCTTCTCTGGTCATGTCTAAGGT
TCT

>mm9_ct_UserTrack_3545_Slc40a1

CAGAGCCTTTCTGTTTCTTCTACAGATTACATCCTGAAATGAAAGCCC
CGGAAAACACAAGGAAGTGTGAGCAAGAACCAGGACTTATTAT**GTG**
TTGTTAATTTTATCTAACAAAAGAGAACTGCTTTTTGTTTTGAATTCCT
CTATTCAGAGTTAGTTCAGGTAGCAAATACTCAGGGGGATATGAAA
ACACATGACAGTCATATGGCAGAGGAAGTAAGCAAAGATTGCAAAA
TATCCT