

Домашнее задание №9 'ChIP-seq'

Цель: Проанализировать данные *ChIP-seq*

Задачи: 1- Обработать и картировать прочтения

2 - Найти пики и гены, к которым эти пики относятся

3 - Предположить, какие функции выполняет белок, с которым мы работаем

4 - Найти мотив для выбранного белка

Результаты:

1-Выберите набор данных

Были выбраны данные для фактора транскрипции *TF1* :

TF1_input_R1_001.fastq.gz, TF1_input_R2_001.fastq.gz

TF1_rep1_R1_001.fastq.gz, TF1_rep1_R2_001.fastq.gz

TF1_rep2_R1_001.fastq.gz, TF1_rep2_R2_001.fastq.gz

2-Экстрагируйте риды и обрежьте их с помощью *Trimmomatic*

Обработка ридов осуществлялась с помощью *Trimmomatic*, код в файле [Chip-seq.ipynb](#)

3-Наложите на геном *E. coli K-12 MG1655 genome* с помощью *Bowtie*.

Оцените, хорошо ли картировались ваши риды.

Были скачаны геном и аннотация *E. coli K-12 MG1655*. Картирование ридов осуществлялось с помощью *Bowtie*, код в файле [Chip-seq.ipynb](#)

Результаты картирования : риды картировались достаточно хорошо

TF1_input

```
# reads processed: 18245035
# reads with at least one alignment: 8969059 (49.16%)
# reads that failed to align: 9275976 (50.84%)
# reads with alignments suppressed due to -m: 201565 (1.10%)
Reported 8767494 paired-end alignments
```

TF1_rep1

```
# reads processed: 8184718
# reads with at least one alignment: 2651670 (32.40%)
# reads that failed to align: 5533048 (67.60%)
# reads with alignments suppressed due to -m: 49703 (0.61%)
Reported 2601967 paired-end alignments
```

TF1_rep2

```
# reads processed: 4820396
# reads with at least one alignment: 1554406 (32.25%)
# reads that failed to align: 3265990 (67.75%)
# reads with alignments suppressed due to -m: 29843 (0.62%)
Reported 1524563 paired-end alignments
```

4-Найдите пики с помощью *MACS2* и сделайте табличку с координатами пиков и генами, к которым они относятся.

Запустили *MACS2* для образцов *TF1_rep1* и *TF1_rep2*, код в файле *Chip-seq.ipynb*, результат в папке *macs2*.

Для сопоставления координат пиков с генами использовалась аннотация *E. coli* и *bedtools intersect*. Таблица с координатами пиков и генами, к которым они относятся, в Приложении 1 в таблице 1.

5-На основании полученного списка генов предположите, какие функции выполняет белок, с которым вы работаете.

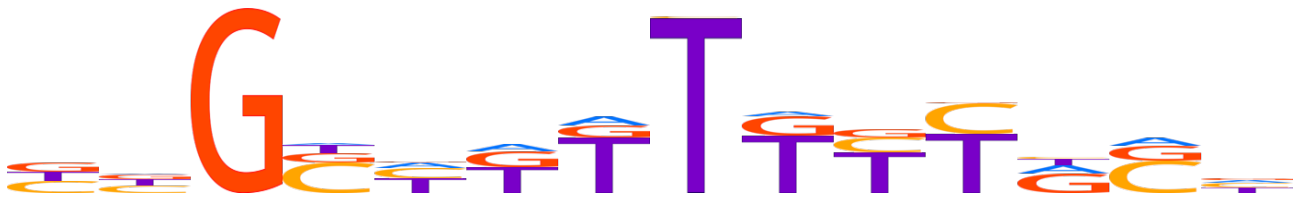
Для выявления функции белка я взяла базу *GeneOntology*. При использовании датасета *PANTHER GO-Slim Biological Process* для аннотации (параметры в Приложении 4) были найдены 2 процесса, в котором задействован наш белок, : процесс биосинтеза олигосахаридов и процесс биосинтеза липополисахаридов. Также я использовала для аннотации датасет *GO biological process complete* (параметры и результаты в Приложении 5).

6-Используя *ChIPMunk*, найдите мотив для выбранного белка.

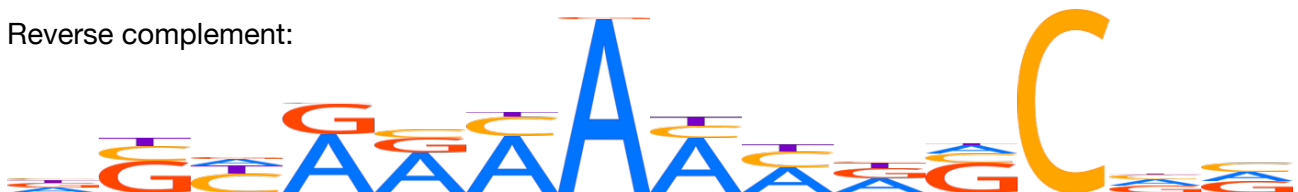
Для запуска *ChIPMunk* необходим файл с последовательностями, которые находятся вблизи наших пиков (в рекомендациях для *MEME* я нашла, что оптимальный размер входных последовательностей ~300 bp), поэтому с помощью *bedtools slop* координаты пиков были расширены на ± 150 , а с помощью *bedtools getfasta* получены последовательности для наших координат. Код в файле *Chip-seq.ipynb*, результат в папке *motif*.

Мотив для нашего белка :

Primary motif:



Reverse complement:



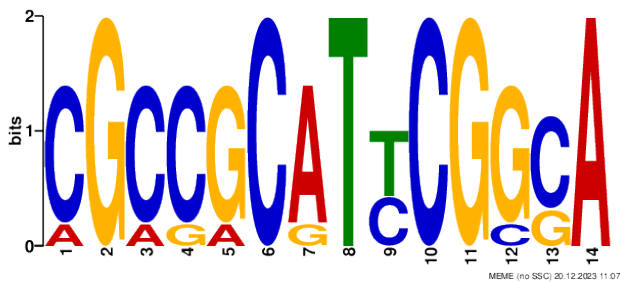
Threshold 4.28022 ?
P-value 0.0 ?
KDIC 0.39156
Weight 239.0 Input 277 ? Words 239 ? Sequences 239 ?
Motif length 14
Diagnosis success; strong motif found within the given lengths interval
IUPAC consensus dShRRMAMMdSCvv
Background ? 0.28601, 0.21399, 0.21399, 0.28601

7-Загрузите данные в *MEME*, сравните полученные мотивы и обоснуйте, какой мотив лучше с вашей точки зрения и почему.

Параметры для запуска *MEME-ChiP* в Приложении 3.

Мотивы для нашего белка :

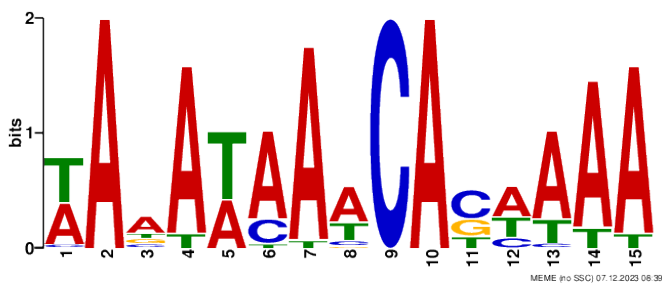
Primary motif:



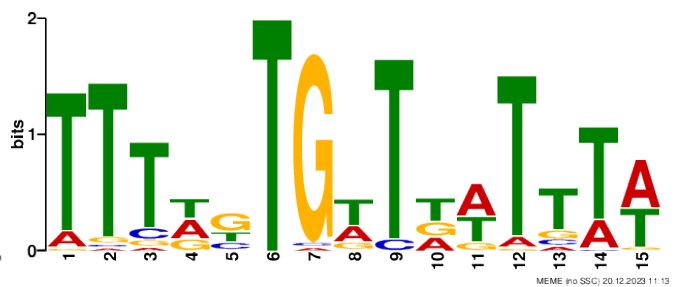
Reverse complement:



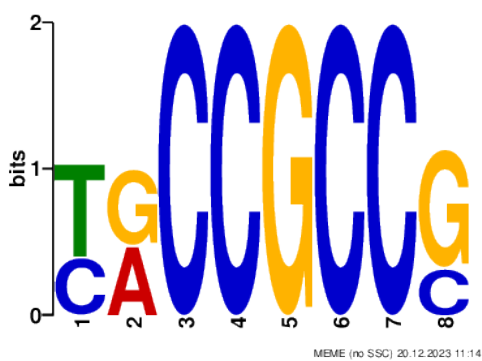
Primary motif:



Reverse complement:



Primary motif:



Reverse complement:

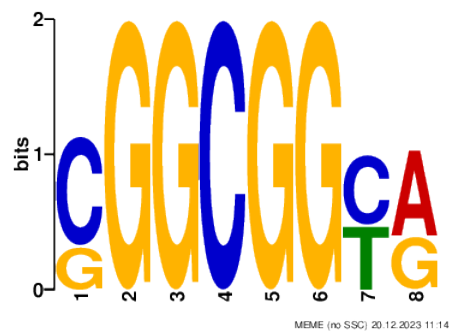
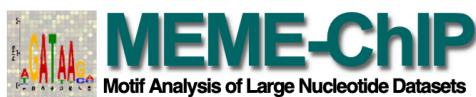


Таблица 1

| | start | end | peak | | | start | end | strand | gene |
|----------|---------|---------|-------------------|---------|------|---------|---------|--------|-----------|
| U00096.3 | 18590 | 18591 | TF1_rep1_peak_1 | 2.79905 | gene | 17489 | 18655 | + | gene=nhaA |
| U00096.3 | 79800 | 79801 | TF1_rep1_peak_3 | 3.80691 | gene | 79464 | 80864 | - | gene=leuC |
| U00096.3 | 113542 | 113543 | TF1_rep1_peak_6 | 2.67335 | gene | 113444 | 114487 | + | gene=guaC |
| U00096.3 | 150113 | 150114 | TF1_rep1_peak_7 | 6.36211 | gene | 149715 | 150953 | - | gene=yadC |
| U00096.3 | 156124 | 156125 | TF1_rep1_peak_8 | 3.99762 | gene | 155461 | 156201 | - | gene=yadV |
| U00096.3 | 434816 | 434817 | TF1_rep1_peak_10 | 1.33603 | gene | 434647 | 435117 | + | gene=ribE |
| U00096.3 | 454706 | 454707 | TF1_rep1_peak_11 | 2.67917 | gene | 454472 | 454789 | + | gene=bolA |
| U00096.3 | 503553 | 503554 | TF1_rep1_peak_13 | 4.1384 | gene | 503476 | 504696 | - | gene=fsr |
| U00096.3 | 504591 | 504592 | TF1_rep1_peak_14 | 4.2693 | gene | 503476 | 504696 | - | gene=fsr |
| U00096.3 | 527096 | 527097 | TF1_rep1_peak_16 | 2.33733 | gene | 523261 | 527541 | + | gene=rhsD |
| U00096.3 | 527764 | 527765 | TF1_rep1_peak_17 | 5.85897 | gene | 527581 | 527949 | + | gene=ybbC |
| U00096.3 | 541736 | 541737 | TF1_rep1_peak_20 | 2.56254 | gene | 540565 | 541866 | + | gene=ybbY |
| U00096.3 | 776972 | 776973 | TF1_rep1_peak_29 | 3.04753 | gene | 776342 | 777607 | + | gene=tolA |
| U00096.3 | 803706 | 803707 | TF1_rep1_peak_32 | 3.99762 | gene | 803503 | 805764 | + | gene=ybhJ |
| U00096.3 | 3121169 | 3121170 | TF1_rep1_peak_39 | 2.41045 | gene | 3119597 | 3121279 | - | gene=glcA |
| U00096.3 | 3134219 | 3134220 | TF1_rep1_peak_40 | 3.57666 | gene | 3134131 | 3134823 | + | gene=yghT |
| U00096.3 | 3268998 | 3268999 | TF1_rep1_peak_42 | 2.67917 | gene | 3268415 | 3269602 | + | gene=yhaC |
| U00096.3 | 3320709 | 3320710 | TF1_rep1_peak_45 | 3.07546 | gene | 3319988 | 3321613 | - | gene=yhbX |
| U00096.3 | 3332088 | 3332089 | TF1_rep1_peak_48 | 3.00606 | gene | 3331770 | 3332735 | - | gene=yhbE |
| U00096.3 | 3336788 | 3336789 | TF1_rep1_peak_50 | 3.00577 | gene | 3336549 | 3336803 | - | gene=ibaG |
| U00096.3 | 3349950 | 3349951 | TF1_rep1_peak_51 | 4.67315 | gene | 3349806 | 3350459 | - | gene=elbB |
| U00096.3 | 3363343 | 3363344 | TF1_rep1_peak_53 | 7.08602 | gene | 3362807 | 3365188 | + | gene=yhcD |
| U00096.3 | 3364112 | 3364113 | TF1_rep1_peak_54 | 3.04753 | gene | 3362807 | 3365188 | + | gene=yhcD |
| U00096.3 | 3367557 | 3367558 | TF1_rep1_peak_55 | 6.71495 | gene | 3366926 | 3367642 | + | gene=yhcF |
| U00096.3 | 3369095 | 3369096 | TF1_rep1_peak_56 | 5.53201 | gene | 3369014 | 3369478 | - | gene=nanQ |
| U00096.3 | 3377299 | 3377300 | TF1_rep1_peak_58 | 3.66484 | gene | 3376782 | 3377420 | - | gene=sspA |
| U00096.3 | 3453935 | 3453936 | TF1_rep1_peak_66 | 3.3085 | gene | 3453929 | 3455398 | - | gene=gspA |
| U00096.3 | 3454827 | 3454828 | TF1_rep1_peak_67 | 3.0738 | gene | 3453929 | 3455398 | - | gene=gspA |
| U00096.3 | 3455921 | 3455922 | TF1_rep1_peak_68 | 2.32522 | gene | 3455578 | 3456393 | + | gene=gspC |
| U00096.3 | 3463693 | 3463694 | TF1_rep1_peak_69 | 4.1384 | gene | 3462924 | 3463907 | + | gene=gspK |
| U00096.3 | 3464372 | 3464373 | TF1_rep1_peak_70 | 2.89249 | gene | 3463922 | 3465085 | + | gene=gspL |
| U00096.3 | 3469520 | 3469521 | TF1_rep1_peak_71 | 3.85404 | gene | 3467160 | 3469853 | - | gene=chiA |
| U00096.3 | 3471711 | 3471712 | TF1_rep1_peak_72 | 3.66484 | gene | 3471400 | 3473514 | - | gene=fusA |
| U00096.3 | 3478654 | 3478655 | TF1_rep1_peak_74 | 2.22338 | gene | 3478592 | 3478792 | - | gene=yheV |
| U00096.3 | 3487495 | 3487496 | TF1_rep1_peak_75 | 4.85009 | gene | 3486791 | 3488893 | + | gene=yhfK |
| U00096.3 | 3488289 | 3488290 | TF1_rep1_peak_76 | 5.40518 | gene | 3486791 | 3488893 | + | gene=yhfK |
| U00096.3 | 3491591 | 3491592 | TF1_rep1_peak_78 | 4.04305 | gene | 3491453 | 3491620 | - | gene=yhfG |
| U00096.3 | 3496833 | 3496834 | TF1_rep1_peak_80 | 1.30887 | gene | 3496551 | 3496877 | + | gene=nirD |
| U00096.3 | 3504355 | 3504356 | TF1_rep1_peak_82 | 2.25395 | gene | 3504052 | 3504783 | + | gene=frlR |
| U00096.3 | 3507437 | 3507438 | TF1_rep1_peak_83 | 2.01288 | gene | 3507348 | 3507701 | - | gene=yhfU |
| U00096.3 | 3511814 | 3511815 | TF1_rep1_peak_84 | 3.76548 | gene | 3511439 | 3512344 | - | gene=yhfZ |
| U00096.3 | 3515319 | 3515320 | TF1_rep1_peak_86 | 2.99063 | gene | 3515077 | 3515913 | - | gene=dam |
| U00096.3 | 3519039 | 3519040 | TF1_rep1_peak_88 | 5.9865 | gene | 3518543 | 3519064 | - | gene=aroK |
| U00096.3 | 3532054 | 3532055 | TF1_rep1_peak_90 | 3.91429 | gene | 3530715 | 3532439 | - | gene=yhgE |
| U00096.3 | 3536497 | 3536498 | TF1_rep1_peak_92 | 4.0281 | gene | 3535865 | 3536584 | - | gene=ompR |
| U00096.3 | 3555256 | 3555257 | TF1_rep1_peak_94 | 4.48767 | gene | 3553085 | 3555790 | + | gene=malT |
| U00096.3 | 3557349 | 3557350 | TF1_rep1_peak_95 | 2.90196 | gene | 3556853 | 3558079 | - | gene=rtcB |
| U00096.3 | 3558400 | 3558401 | TF1_rep1_peak_96 | 4.55192 | gene | 3558268 | 3559866 | + | gene=rtcR |
| U00096.3 | 3562017 | 3562018 | TF1_rep1_peak_98 | 2.80818 | gene | 3562013 | 3563518 | + | gene=glpD |
| U00096.3 | 3569304 | 3569305 | TF1_rep1_peak_99 | 6.45743 | gene | 3568033 | 3569328 | - | gene=glgC |
| U00096.3 | 3571866 | 3571867 | TF1_rep1_peak_100 | 2.18114 | gene | 3571316 | 3573502 | - | gene=glgB |
| U00096.3 | 3573399 | 3573400 | TF1_rep1_peak_101 | 3.24145 | gene | 3571316 | 3573502 | - | gene=glgB |
| U00096.3 | 3574637 | 3574638 | TF1_rep1_peak_102 | 4.05223 | gene | 3573775 | 3574878 | - | gene=asd |
| U00096.3 | 3577195 | 3577196 | TF1_rep1_peak_103 | 5.53201 | gene | 3577065 | 3577592 | - | gene=gntK |
| U00096.3 | 3580343 | 3580344 | TF1_rep1_peak_105 | 4.906 | gene | 3579768 | 3580805 | - | gene=yhhX |
| U00096.3 | 3585849 | 3585850 | TF1_rep1_peak_108 | 11.7148 | gene | 3585081 | 3586823 | - | gene=ggt |
| U00096.3 | 3586797 | 3586798 | TF1_rep1_peak_109 | 3.85404 | gene | 3585081 | 3586823 | - | gene=ggt |
| U00096.3 | 3587407 | 3587408 | TF1_rep1_peak_110 | 1.55419 | gene | 3587370 | 3588113 | - | gene=ugpQ |
| U00096.3 | 3596728 | 3596729 | TF1_rep1_peak_111 | 2.92078 | gene | 3596451 | 3597560 | - | gene=livK |

| | start | end | peak | | | start | end | strand | gene |
|----------|---------|---------|-------------------|---------|------|---------|---------|--------|-----------|
| U00096.3 | 3603942 | 3603943 | TF1_rep1_peak_115 | 1.66876 | gene | 3602750 | 3604243 | - | gene=ftsY |
| U00096.3 | 3604794 | 3604795 | TF1_rep1_peak_116 | 3.04753 | gene | 3604393 | 3604989 | + | gene=rsmD |
| U00096.3 | 3623410 | 3623411 | TF1_rep1_peak_118 | 18.4444 | gene | 3623399 | 3623782 | + | gene=yhhH |
| U00096.3 | 3623410 | 3623411 | TF1_rep1_peak_118 | 18.4444 | gene | 3619192 | 3623427 | + | gene=rhsB |
| U00096.3 | 3635743 | 3635744 | TF1_rep1_peak_121 | 4.58159 | gene | 3634841 | 3635893 | + | gene=yhiM |
| U00096.3 | 3637750 | 3637751 | TF1_rep1_peak_122 | 8.04676 | gene | 3637642 | 3639141 | + | gene=pitA |
| U00096.3 | 3644367 | 3644368 | TF1_rep1_peak_124 | 2.16948 | gene | 3643140 | 3645182 | - | gene=prlC |
| U00096.3 | 3648671 | 3648672 | TF1_rep1_peak_126 | 4.58159 | gene | 3648528 | 3648881 | + | gene=arsR |
| U00096.3 | 3649946 | 3649947 | TF1_rep1_peak_127 | 1.69818 | gene | 3648935 | 3650224 | + | gene=arsB |
| U00096.3 | 3654813 | 3654814 | TF1_rep1_peak_130 | 3.71502 | gene | 3654683 | 3655213 | + | gene=dctR |
| U00096.3 | 3658051 | 3658052 | TF1_rep1_peak_133 | 6.5352 | gene | 3657985 | 3658054 | - | gene=arrS |
| U00096.3 | 3658051 | 3658052 | TF1_rep1_peak_133 | 6.5352 | exon | 3657985 | 3658054 | - | gene=arrS |
| U00096.3 | 3664307 | 3664308 | TF1_rep1_peak_134 | 2.67917 | gene | 3663890 | 3664618 | - | gene=gadW |
| U00096.3 | 3674039 | 3674040 | TF1_rep1_peak_140 | 3.34557 | gene | 3673362 | 3674375 | + | gene=yhjD |
| U00096.3 | 3677865 | 3677866 | TF1_rep1_peak_142 | 5.04036 | gene | 3676290 | 3678350 | - | gene=yhjG |
| U00096.3 | 3697305 | 3697306 | TF1_rep1_peak_145 | 1.57407 | gene | 3696458 | 3698029 | + | gene=bcsE |
| U00096.3 | 3700172 | 3700173 | TF1_rep1_peak_146 | 3.76435 | gene | 3700136 | 3700199 | + | gene=rdlD |
| U00096.3 | 3700172 | 3700173 | TF1_rep1_peak_146 | 3.76435 | exon | 3700136 | 3700199 | + | gene=rdlD |
| U00096.3 | 3706195 | 3706196 | TF1_rep1_peak_147 | 3.30249 | gene | 3706098 | 3707705 | - | gene=dppA |
| U00096.3 | 3710415 | 3710416 | TF1_rep1_peak_149 | 11.9663 | gene | 3708784 | 3710475 | - | gene=eptB |
| U00096.3 | 3712623 | 3712624 | TF1_rep1_peak_151 | 3.30643 | gene | 3712236 | 3712934 | - | gene=yhjY |
| U00096.3 | 3715029 | 3715030 | TF1_rep1_peak_152 | 2.60345 | gene | 3714061 | 3716394 | - | gene=bisC |
| U00096.3 | 3720933 | 3720934 | TF1_rep1_peak_154 | 3.22424 | gene | 3720680 | 3721201 | + | gene=insJ |
| U00096.3 | 3731321 | 3731322 | TF1_rep1_peak_157 | 4.1384 | gene | 3731131 | 3732123 | + | gene=xylF |
| U00096.3 | 3735710 | 3735711 | TF1_rep1_peak_159 | 2.88897 | gene | 3734979 | 3736157 | + | gene=xylR |
| U00096.3 | 3754055 | 3754056 | TF1_rep1_peak_160 | 4.90217 | gene | 3752963 | 3754099 | - | gene=yiaV |
| U00096.3 | 3766041 | 3766042 | TF1_rep1_peak_162 | 5.21087 | gene | 3762183 | 3766316 | + | gene=rhsA |
| U00096.3 | 3766745 | 3766746 | TF1_rep1_peak_163 | 3.57666 | gene | 3766337 | 3767179 | + | gene=yibA |
| U00096.3 | 3768452 | 3768453 | TF1_rep1_peak_164 | 8.67426 | gene | 3768177 | 3768638 | + | gene=yibG |
| U00096.3 | 3776711 | 3776712 | TF1_rep1_peak_166 | 2.56146 | gene | 3776665 | 3777027 | + | gene=yibL |
| U00096.3 | 3789675 | 3789676 | TF1_rep1_peak_167 | 6.11742 | gene | 3789047 | 3790081 | - | gene=waaH |
| U00096.3 | 3790394 | 3790395 | TF1_rep1_peak_168 | 3.17991 | gene | 3790320 | 3791345 | - | gene=tdh |
| U00096.3 | 3793003 | 3793004 | TF1_rep1_peak_169 | 4.1384 | gene | 3792826 | 3793683 | - | gene=htrL |
| U00096.3 | 3797063 | 3797064 | TF1_rep1_peak_171 | 6.8948 | gene | 3796948 | 3798207 | + | gene=waaL |
| U00096.3 | 3800909 | 3800910 | TF1_rep1_peak_173 | 5.37125 | gene | 3800267 | 3800965 | - | gene=waaY |
| U00096.3 | 3802393 | 3802394 | TF1_rep1_peak_174 | 3.85404 | gene | 3802039 | 3803058 | - | gene=waaO |
| U00096.3 | 3802850 | 3802851 | TF1_rep1_peak_175 | 2.5624 | gene | 3802039 | 3803058 | - | gene=waaO |
| U00096.3 | 3806245 | 3806246 | TF1_rep1_peak_177 | 2.5624 | gene | 3805943 | 3807067 | - | gene=waaG |
| U00096.3 | 3811050 | 3811051 | TF1_rep1_peak_179 | 6.17619 | gene | 3810343 | 3811152 | - | gene=mutM |
| U00096.3 | 3811927 | 3811928 | TF1_rep1_peak_180 | 3.45464 | gene | 3811891 | 3812559 | - | gene=yicR |
| U00096.3 | 3819542 | 3819543 | TF1_rep1_peak_182 | 4.32618 | gene | 3819488 | 3821170 | - | gene=ligB |
| U00096.3 | 3822027 | 3822028 | TF1_rep1_peak_184 | 6.48814 | gene | 3821428 | 3822051 | + | gene=gmk |
| U00096.3 | 3833621 | 3833622 | TF1_rep1_peak_185 | 4.29546 | gene | 3832219 | 3834537 | - | gene=yicI |
| U00096.3 | 3838281 | 3838282 | TF1_rep1_peak_188 | 3.30643 | gene | 3838248 | 3839171 | + | gene=yicL |
| U00096.3 | 3840410 | 3840411 | TF1_rep1_peak_189 | 8.04676 | gene | 3840215 | 3840508 | + | gene=yicS |
| U00096.3 | 3843274 | 3843275 | TF1_rep1_peak_190 | 5.19737 | gene | 3842455 | 3843789 | - | gene=adeQ |
| U00096.3 | 3863941 | 3863942 | TF1_rep1_peak_194 | 5.68994 | gene | 3863899 | 3864615 | + | gene=yidP |
| U00096.3 | 3864635 | 3864636 | TF1_rep1_peak_195 | 6.76459 | gene | 3864612 | 3866273 | - | gene=yidE |
| U00096.3 | 3866734 | 3866735 | TF1_rep1_peak_196 | 6.56962 | gene | 3866469 | 3866897 | - | gene=ibpB |
| U00096.3 | 3869809 | 3869810 | TF1_rep1_peak_198 | 2.92261 | gene | 3869377 | 3870441 | + | gene=cbrA |
| U00096.3 | 3871003 | 3871004 | TF1_rep1_peak_199 | 1.71841 | gene | 3870438 | 3871730 | - | gene=dgoT |
| U00096.3 | 3872045 | 3872046 | TF1_rep1_peak_200 | 4.3121 | gene | 3871850 | 3872998 | - | gene=dgoD |
| U00096.3 | 3875888 | 3875889 | TF1_rep1_peak_201 | 2.7059 | gene | 3875438 | 3876094 | + | gene=yidX |
| U00096.3 | 3888132 | 3888133 | TF1_rep1_peak_202 | 5.53201 | gene | 3886828 | 3888192 | + | gene=mnmE |
| U00096.3 | 3891136 | 3891137 | TF1_rep1_peak_203 | 2.4714 | gene | 3890236 | 3891483 | + | gene=tnaB |
| U00096.3 | 3893721 | 3893722 | TF1_rep1_peak_204 | 4.32508 | gene | 3892765 | 3893724 | + | gene=yidZ |
| U00096.3 | 3896858 | 3896859 | TF1_rep1_peak_205 | 5.99455 | gene | 3896774 | 3897439 | + | gene=yieH |
| U00096.3 | 3897641 | 3897642 | TF1_rep1_peak_206 | 7.19605 | gene | 3897506 | 3897973 | + | gene=cbrB |
| U00096.3 | 3898362 | 3898363 | TF1_rep1_peak_207 | 3.71502 | gene | 3898022 | 3898609 | + | gene=cbrC |
| U00096.3 | 3899314 | 3899315 | TF1_rep1_peak_208 | 4.73241 | gene | 3898671 | 3899393 | - | gene=yieK |

| | start | end | peak | | | start | end | strand | gene |
|----------|---------|---------|-------------------|---------|------|---------|---------|--------|------------|
| U00096.3 | 3900876 | 3900877 | TF1_rep1_peak_209 | 1.91842 | gene | 3900604 | 3902220 | - | gene=bgIH |
| U00096.3 | 3901291 | 3901292 | TF1_rep1_peak_210 | 2.44841 | gene | 3900604 | 3902220 | - | gene=bgIH |
| U00096.3 | 3902942 | 3902943 | TF1_rep1_peak_211 | 3.44151 | gene | 3902289 | 3903701 | - | gene=bgIB |
| U00096.3 | 3926386 | 3926387 | TF1_rep1_peak_215 | 2.6931 | gene | 3926012 | 3926455 | - | gene=mioC |
| U00096.3 | 3927168 | 3927169 | TF1_rep1_peak_216 | 3.84422 | gene | 3927155 | 3928147 | + | gene=asnA |
| U00096.3 | 3939683 | 3939684 | TF1_rep1_peak_217 | 5.6941 | gene | 3939185 | 3940612 | - | gene=hsrA |
| U00096.3 | 3947166 | 3947167 | TF1_rep1_peak_219 | 5.21087 | gene | 3947128 | 3947967 | - | gene=hdfR |
| U00096.3 | 3947538 | 3947539 | TF1_rep1_peak_220 | 3.85404 | gene | 3947128 | 3947967 | - | gene=hdfR |
| U00096.3 | 3950350 | 3950351 | TF1_rep1_peak_221 | 3.57666 | gene | 3950322 | 3950420 | + | gene=ilvL |
| U00096.3 | 3966328 | 3966329 | TF1_rep1_peak_223 | 2.80291 | gene | 3966231 | 3966332 | + | gene=rhoL |
| U00096.3 | 4051045 | 4051046 | TF1_rep1_peak_230 | 3.32607 | gene | 4051036 | 4051281 | + | gene=csrC |
| U00096.3 | 4051045 | 4051046 | TF1_rep1_peak_230 | 3.32607 | exon | 4051036 | 4051281 | + | gene=csrC |
| U00096.3 | 4282976 | 4282977 | TF1_rep1_peak_233 | 5.13172 | gene | 4281783 | 4283075 | - | gene=yjcF |
| U00096.3 | 4304638 | 4304639 | TF1_rep1_peak_234 | 3.13842 | gene | 4304612 | 4306597 | - | gene=yjcS |
| U00096.3 | 4312023 | 4312024 | TF1_rep1_peak_236 | 1.91842 | gene | 4311107 | 4312042 | - | gene=alsB |
| U00096.3 | 4331831 | 4331832 | TF1_rep1_peak_240 | 2.4934 | gene | 4330502 | 4332004 | + | gene=proP |
| U00096.3 | 4337558 | 4337559 | TF1_rep1_peak_241 | 3.59156 | gene | 4337168 | 4337929 | - | gene=adiY |
| U00096.3 | 4358494 | 4358495 | TF1_rep1_peak_247 | 3.17581 | gene | 4356470 | 4358617 | - | gene=cadA |
| U00096.3 | 4361931 | 4361932 | TF1_rep1_peak_249 | 5.04794 | gene | 4360396 | 4361934 | - | gene=cadC |
| U00096.3 | 4369700 | 4369701 | TF1_rep1_peak_250 | 2.30807 | gene | 4369156 | 4370412 | - | gene=yjeH |
| U00096.3 | 4373990 | 4373991 | TF1_rep1_peak_251 | 3.99762 | gene | 4373365 | 4374234 | - | gene=yjeJ |
| U00096.3 | 4377682 | 4377683 | TF1_rep1_peak_253 | 4.36395 | gene | 4377189 | 4377722 | - | gene=blc |
| U00096.3 | 4378296 | 4378297 | TF1_rep1_peak_254 | 3.31756 | gene | 4377811 | 4378944 | - | gene=ampC |
| U00096.3 | 4410294 | 4410295 | TF1_rep1_peak_256 | 3.9641 | gene | 4410133 | 4410534 | + | gene=yjfl |
| U00096.3 | 4412129 | 4412130 | TF1_rep1_peak_257 | 2.53993 | gene | 4411979 | 4412377 | + | gene=yjfl |
| U00096.3 | 4438633 | 4438634 | TF1_rep1_peak_259 | 3.44151 | gene | 4437707 | 4438645 | + | gene=ytfl |
| U00096.3 | 4439781 | 4439782 | TF1_rep1_peak_260 | 2.67917 | gene | 4439587 | 4439793 | + | gene=ytflK |
| U00096.3 | 4442755 | 4442756 | TF1_rep1_peak_261 | 3.61791 | gene | 4442382 | 4444115 | + | gene=tamA |
| U00096.3 | 4474506 | 4474507 | TF1_rep1_peak_262 | 2.33733 | gene | 4474124 | 4474717 | + | gene=bdcR |
| U00096.3 | 4476369 | 4476370 | TF1_rep1_peak_263 | 2.92261 | gene | 4475437 | 4477251 | + | gene=yjgL |
| U00096.3 | 4479743 | 4479744 | TF1_rep1_peak_264 | 2.22786 | gene | 4479730 | 4480926 | + | gene=yjgN |
| U00096.3 | 4504722 | 4504723 | TF1_rep1_peak_266 | 4.43082 | gene | 4504058 | 4505275 | + | gene=nanX |
| U00096.3 | 464020 | 464021 | TF1_rep2_peak_1 | 2.09282 | gene | 463937 | 464308 | + | gene=ybaV |
| U00096.3 | 3586282 | 3586283 | TF1_rep2_peak_2 | 2.09282 | gene | 3585081 | 3586823 | - | gene=ggT |
| U00096.3 | 3623363 | 3623364 | TF1_rep2_peak_3 | 2.00051 | gene | 3619192 | 3623427 | + | gene=rhsB |
| U00096.3 | 3635801 | 3635802 | TF1_rep2_peak_5 | 2.79423 | gene | 3634841 | 3635893 | + | gene=yhiM |
| U00096.3 | 3897022 | 3897023 | TF1_rep2_peak_9 | 2.35818 | gene | 3896774 | 3897439 | + | gene=yjeH |
| U00096.3 | 4308108 | 4308109 | TF1_rep2_peak_10 | 2.47106 | gene | 4307783 | 4308478 | - | gene=alsE |



Version 5.5.4

MEME-ChIP performs **comprehensive motif analysis** (including motif discovery) on sequences where the motif sites tend to be **centrally** located, such as ChIP-seq peaks (sample output from [sequences](#)). The input sequences should be **centered** on a **100 character region** expected to contain motifs, and each sequence should ideally be around **500 letters** long. See this [Manual](#) for more information.

Data Submission Form

Perform motif discovery, motif enrichment analysis and clustering on large nucleotide datasets.

Select the motif discovery and enrichment mode [?](#)

☒ Classic mode ☐ Discriminative mode ☐ Differential Enrichment mode

Select the sequence alphabet

Use sequences with a standard alphabet or specify a custom alphabet. [?](#)

☒ DNA, RNA or Protein ☐ Custom Выбрать файл файл не выбран

Input the primary sequences

Enter the (equal-length) nucleotide sequences to be analyzed. [?](#)

Upload sequences Выбрать файл TF1_rep1.fasta DNA [?](#)

Convert DNA sequences to RNA?

☐ Convert DNA to RNA [?](#)

Input the motifs

Select, upload or enter a set of known motifs. [?](#)

ECOLI (Escherichia coli) DNA DNA [?](#)
Swiss Regulon [?](#)

Input job details

(Optional) Enter your email address. [?](#)

(Optional) Enter a job description. [?](#)

- Universal options NEW OPTIONS
- MEME options
- STREME options
- CentriMo options



The mission of the PANTHER knowledgebase is to support biomedical and other research by providing **comprehensive information about the evolution of protein-coding gene families**, particularly protein phylogeny, function and genetic variation impacting that function. [Learn more](#)

PANTHER18.0 Released. [Click](#) for more details.

AI Go

[Home](#) [About](#) [Data Version](#) [Tools](#) [API/Services](#) [Publications](#) [Workspace](#) [Downloads](#) [FAQ/Help/Tutorial](#) [Login](#) [Register](#) [Contact us](#)

Current Release: [PANTHER 18.0](#) | [15,693](#) family phylogenetic trees | [143](#) species | [News](#)
[Whole genome function views](#)

Analysis Summary: Please report in publication ?

Analysis Type: PANTHER Overrepresentation Test (Released 20231017)

Annotation Version and Release Date: PANTHER version 18.0 Released 2023-08-01

Analzyed List:

upload_1 (Escherichia coli)

Change

Reference List:

Escherichia coli (all genes in database)

Change

Annotation Data Set:

PANTHER GO-Slim Biological Process

+ ?

Test Type:

☒ Fisher's Exact

☐ Binomial

Correction:

☐ Calculate False Discovery Rate

☐ Use the Bonferroni correction for multiple testing ?

☒ No correction

Results ?

| | | |
|-------------------------------|----------------------------------|--------------------------------|
| | Reference list | upload_1 |
| Uniquely Mapped IDs: | 4401 out of 4401 | 132 out of 135 |
| Unmapped IDs: | 0 | 11 |
| Multiple mapping information: | 0 | 3 |

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#) View:

-- Please select a chart to display --

Displaying only results for uncorrected P < 0.05, [click here to display all results](#)

| | | | | | | |
|---|--|---|--------------------------|---------------------------------|---------------------|-----------------------------|
| | Escherichia coli (REF) | upload_1 (▼ Hierarchy NEW! ?) | | | | |
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | raw P value |
| oligosaccharide biosynthetic process | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| lipopolysaccharide biosynthetic process | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |



The mission of the PANTHER knowledgebase is to support biomedical and other research by providing comprehensive information about the evolution of protein-coding gene families, particularly protein phylogeny, function and genetic variation impacting that function. [Learn more](#)

PANTHER18.0 Released. [Click](#) for more details.

search keyword

AI Go

Home

About

Data Version

Tools

API/Services

Publications

Workspace

Downloads

FAQ/Help/Tutorial

Login

Register

Contact us

Current Release: [PANTHER 18.0](#) | [15,693](#) family phylogenetic trees | [143](#) species | [News](#)
[Whole genome function views](#)

Analysis Summary: Please report in publication ?

Analysis Type:

PANTHER Overrepresentation Test (Released 20231017)

Annotation Version and Release Date:

GO Ontology database DOI: 10.5281/zenodo.7942786 Released 2023-01-05

Analyzed List:

upload_1 (Escherichia coli)

Change

Reference List:

Escherichia coli (all genes in database)

Change

Annotation Data Set:

GO biological process complete

?

Test Type:

☒ Fisher's Exact

☐ Binomial

Correction:

☐ Calculate False Discovery Rate

☐ Use the Bonferroni correction for multiple testing

☒ No correction

?

Results ?

| | | |
|-------------------------------|----------------------------------|--------------------------------|
| | Reference list | upload_1 |
| Uniquely Mapped IDs: | 4401 out of 4401 | 132 out of 135 |
| Unmapped IDs: | 0 | 11 |
| Multiple mapping information: | 0 | 3 |

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)

Displaying only results for uncorrected P < 0.05, [click here to display all results](#)

| | Escherichia coli (REF) | upload_1 (Hierarchy NEW! ?) | | | | |
|--|------------------------|-----------------------------|----------|-----------------|----|-------------|
| GO biological process complete | # | # | expected | Fold Enrichment | +- | raw P value |
| bacteriocin immunity | 2 | 2 | .06 | 32.60 | + | 5.07E-03 |
| ↳ toxin metabolic process | 6 | 2 | .18 | 10.87 | + | 2.19E-02 |
| glycogen biosynthetic process | 3 | 2 | .09 | 21.73 | + | 8.29E-03 |
| ↳ glycogen metabolic process | 7 | 2 | .21 | 9.31 | + | 2.76E-02 |
| ↳ macromolecule metabolic process | 1037 | 22 | 31.81 | .69 | - | 4.98E-02 |
| ↳ energy reserve metabolic process | 7 | 2 | .21 | 9.31 | + | 2.76E-02 |
| ↳ glucan biosynthetic process | 18 | 3 | .55 | 5.43 | + | 2.32E-02 |
| D-galactonate catabolic process | 4 | 2 | .12 | 16.30 | + | 1.22E-02 |
| ↳ galactonate catabolic process | 6 | 2 | .18 | 10.87 | + | 2.19E-02 |
| ↳ galactonate metabolic process | 6 | 2 | .18 | 10.87 | + | 2.19E-02 |
| ↳ aldonic acid metabolic process | 23 | 3 | .71 | 4.25 | + | 4.07E-02 |
| ↳ aldonic acid catabolic process | 21 | 3 | .64 | 4.66 | + | 3.30E-02 |
| ↳ D-galactonate metabolic process | 4 | 2 | .12 | 16.30 | + | 1.22E-02 |
| stringent response | 4 | 2 | .12 | 16.30 | + | 1.22E-02 |
| ↳ cellular response to stimulus | 559 | 28 | 17.15 | 1.63 | + | 8.97E-03 |
| ↳ cellular response to stress | 405 | 21 | 12.42 | 1.69 | + | 2.30E-02 |

| | | | | | | |
|--|------|----|-------|-------|---|----------|
| cellular response to organonitrogen compound | 5 | 2 | .15 | 13.04 | + | 1.67E-02 |
| ↳cellular response to nitrogen compound | 8 | 2 | .25 | 8.15 | + | 3.38E-02 |
| ↳response to organonitrogen compound | 20 | 3 | .61 | 4.89 | + | 2.95E-02 |
| ↳cellular response to organic substance | 8 | 2 | .25 | 8.15 | + | 3.38E-02 |
| ↳cellular response to endogenous stimulus | 5 | 2 | .15 | 13.04 | + | 1.67E-02 |
| ↳response to endogenous stimulus | 5 | 2 | .15 | 13.04 | + | 1.67E-02 |
| cellular response to acid chemical | 5 | 2 | .15 | 13.04 | + | 1.67E-02 |
| ↳response to acid chemical | 12 | 3 | .37 | 8.15 | + | 9.02E-03 |
| response to arsenic-containing substance | 5 | 2 | .15 | 13.04 | + | 1.67E-02 |
| leucine transport | 6 | 2 | .18 | 10.87 | + | 2.19E-02 |
| glucosamine-containing compound catabolic process | 6 | 2 | .18 | 10.87 | + | 2.19E-02 |
| ↳glucosamine-containing compound metabolic process | 7 | 2 | .21 | 9.31 | + | 2.76E-02 |
| ↳amino sugar catabolic process | 15 | 3 | .46 | 6.52 | + | 1.52E-02 |
| response to salt stress | 7 | 2 | .21 | 9.31 | + | 2.76E-02 |
| purine nucleobase transport | 8 | 2 | .25 | 8.15 | + | 3.38E-02 |
| transcriptional attenuation | 9 | 2 | .28 | 7.24 | + | 4.06E-02 |
| ↳positive regulation of termination of DNA-templated transcription | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| ↳positive regulation of protein-containing complex disassembly | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| ↳positive regulation of cellular process | 141 | 10 | 4.33 | 2.31 | + | 1.38E-02 |
| ↳regulation of RNA biosynthetic process | 376 | 19 | 11.53 | 1.65 | + | 3.00E-02 |
| ↳regulation of RNA metabolic process | 387 | 19 | 11.87 | 1.60 | + | 4.49E-02 |
| ↳positive regulation of DNA-templated transcription | 99 | 9 | 3.04 | 2.96 | + | 4.52E-03 |
| ↳regulation of DNA-templated transcription | 376 | 19 | 11.53 | 1.65 | + | 3.00E-02 |
| ↳positive regulation of RNA biosynthetic process | 99 | 9 | 3.04 | 2.96 | + | 4.52E-03 |
| ↳positive regulation of RNA metabolic process | 103 | 9 | 3.16 | 2.85 | + | 5.75E-03 |
| ↳positive regulation of nucleobase-containing compound metabolic process | 108 | 9 | 3.31 | 2.72 | + | 7.62E-03 |
| ↳positive regulation of nitrogen compound metabolic process | 123 | 9 | 3.77 | 2.39 | + | 1.61E-02 |
| ↳positive regulation of metabolic process | 150 | 10 | 4.60 | 2.17 | + | 2.73E-02 |
| ↳positive regulation of cellular metabolic process | 121 | 10 | 3.71 | 2.69 | + | 5.26E-03 |
| ↳positive regulation of macromolecule metabolic process | 145 | 9 | 4.45 | 2.02 | + | 4.77E-02 |
| response to amino acid | 9 | 2 | .28 | 7.24 | + | 4.06E-02 |
| riboflavin biosynthetic process | 9 | 2 | .28 | 7.24 | + | 4.06E-02 |
| ↳riboflavin metabolic process | 9 | 2 | .28 | 7.24 | + | 4.06E-02 |
| ↳flavin-containing compound metabolic process | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| ↳heterocycle metabolic process | 861 | 17 | 26.41 | .64 | - | 4.60E-02 |
| ↳nitrogen compound metabolic process | 1511 | 33 | 46.35 | .71 | - | 1.65E-02 |
| ↳organic cyclic compound metabolic process | 909 | 18 | 27.88 | .65 | - | 3.91E-02 |
| ↳flavin-containing compound biosynthetic process | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| N-acetylneuraminate catabolic process | 10 | 2 | .31 | 6.52 | + | 4.77E-02 |
| lipopolysaccharide core region biosynthetic process | 25 | 4 | .77 | 5.22 | + | 1.00E-02 |
| ↳lipopolysaccharide core region metabolic process | 25 | 4 | .77 | 5.22 | + | 1.00E-02 |
| ↳oligosaccharide metabolic process | 53 | 5 | 1.63 | 3.08 | + | 2.81E-02 |
| ↳oligosaccharide biosynthetic process | 29 | 4 | .89 | 4.50 | + | 1.57E-02 |
| protein secretion | 33 | 4 | 1.01 | 3.95 | + | 2.32E-02 |
| ↳establishment of protein localization to extracellular region | 33 | 4 | 1.01 | 3.95 | + | 2.32E-02 |
| ↳establishment of protein localization | 93 | 7 | 2.85 | 2.45 | + | 2.85E-02 |
| ↳protein localization | 96 | 7 | 2.94 | 2.38 | + | 3.28E-02 |
| ↳cellular macromolecule localization | 96 | 7 | 2.94 | 2.38 | + | 3.28E-02 |
| ↳macromolecule localization | 143 | 9 | 4.39 | 2.05 | + | 4.56E-02 |
| ↳protein localization to extracellular region | 33 | 4 | 1.01 | 3.95 | + | 2.32E-02 |
| ↳protein transport | 86 | 7 | 2.64 | 2.65 | + | 2.00E-02 |
| ↳secretion by cell | 35 | 4 | 1.07 | 3.73 | + | 2.76E-02 |
| ↳secretion | 35 | 4 | 1.07 | 3.73 | + | 2.76E-02 |
| branched-chain amino acid biosynthetic process | 25 | 3 | .77 | 3.91 | + | 4.91E-02 |
| ↳branched-chain amino acid metabolic process | 25 | 3 | .77 | 3.91 | + | 4.91E-02 |
| response to pH | 42 | 4 | 1.29 | 3.10 | + | 4.67E-02 |
| DNA damage response | 277 | 15 | 8.50 | 1.77 | + | 3.18E-02 |
| cellular aromatic compound metabolic process | 874 | 16 | 26.81 | .60 | - | 2.05E-02 |
| protein metabolic process | 385 | 4 | 11.81 | .34 | - | 1.23E-02 |
| protein-containing complex organization | 180 | 1 | 5.52 | .18 | - | 4.37E-02 |
| ↳cellular component organization | 475 | 7 | 14.57 | .48 | - | 3.34E-02 |