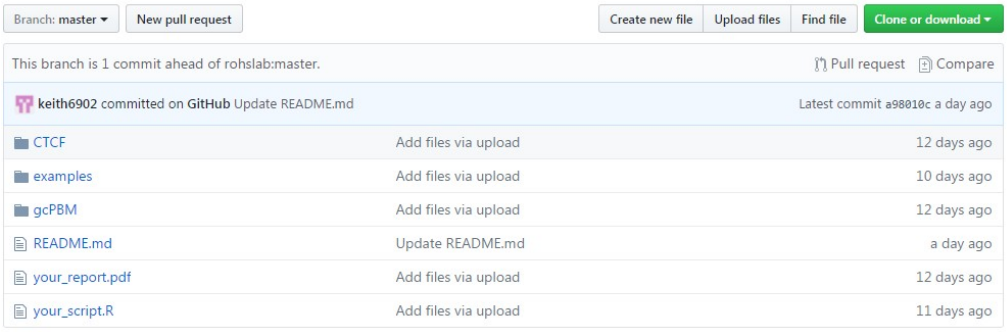
BISC 577a, Spring 2017

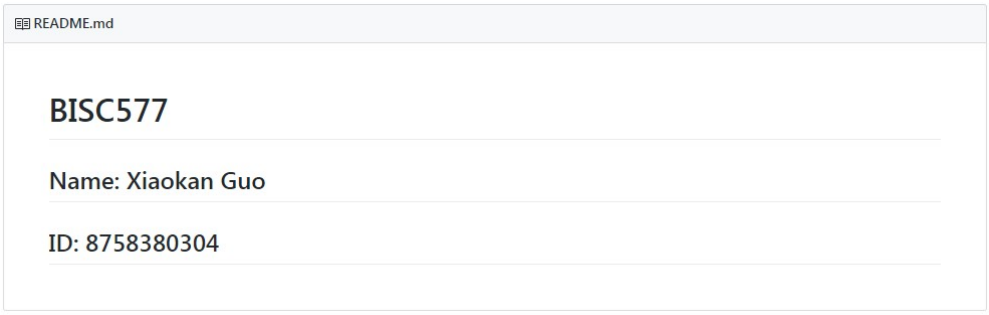
Second Assignment for 3rd section of the course (Rohs)

(1) Application of an open-source and distributed revision control project

(a)



(b)



(c)

(2) High-throughput biding assays

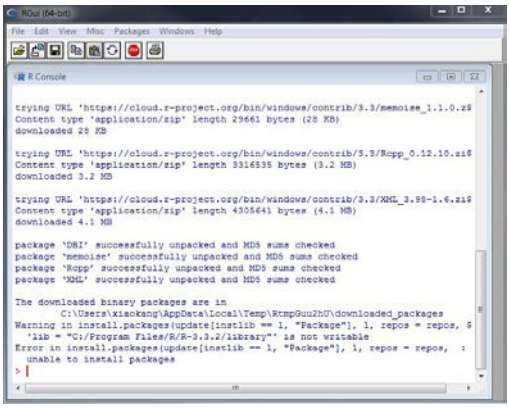
(a) SELEX-seq is an experimental and computational platform used to characterize DNA-binding specificities and determine the relative affinities to any DNA sequence. PBM is a high-throughput technique for characterizing sequence specificity of DNA-binding proteins in vitro.

(b) ChIP-seq is method used to analyze protein interactions with DNA, which combines chromatin immunoprecipitation with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins.

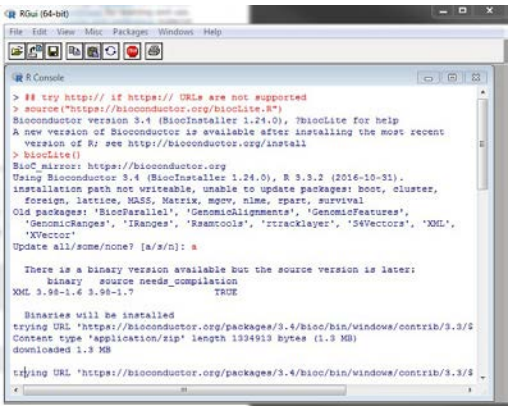
(c) PBM is good for general characterization of binding specificity for multiple proteins. However, ChIP-seq is good for identifying gene targets bound/regulated under a specific condition. SELEX-seq is more accurate than PBM and ChIP-seq models.

(3) Preparation of high-throughput in vitro data analysis

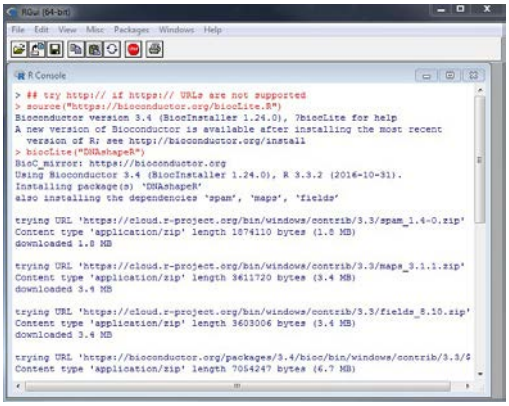
(a)



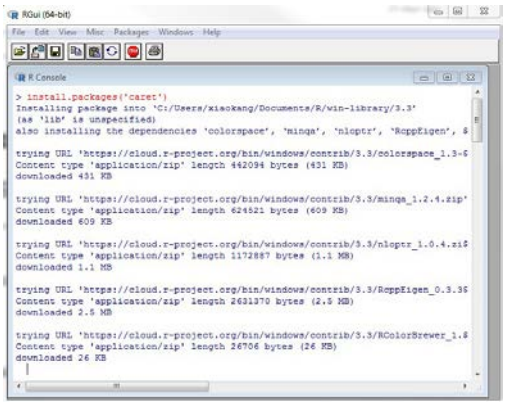
(b)



(c)



(d)



(4) Build prediction models for in vitro data

(a)

Mad 1-mer+shape

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
seq1	0	0	1	0	0	0	1	0	0	0	1	0	0
seq2	0	1	0	0	0	0	1	0	0	0	1	0	0
seq3	0	0	1	0	0	0	1	0	0	1	0	0	0

	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
seq1	1	0	0	1	0	0	0	0	0	0	1
seq2	0	1	0	0	0	1	0	1	0	0	0
seq3	0	1	0	0	0	1	0	0	1	0	0

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]
seq1	0	0	1	0	1	0	0	0	1	0	0
seq2	0	0	1	0	0	0	1	0	1	0	0
seq3	0	1	0	0	0	1	0	0	0	0	1

	[,36]	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]
seq1	0	1	0	0	0	0	0	1	0	0	1
seq2	0	0	0	1	0	0	1	0	0	0	1
seq3	0	0	0	1	0	0	0	1	0	0	1

	[,47]	[,48]	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]
seq1	0	0	0	1	0	0	1	0	0	0	0

```
seq2    0    0    1    0    0    0    0    1    0    0    0
seq3    0    0    1    0    0    0    0    0    0    1    0
    [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65] [,66] [,67] [,68]
seq1    1    0    0    0    1    0    0    0    0    0    1
seq2    1    0    0    0    1    0    0    1    0    0    0
seq3    1    0    0    0    1    0    0    1    0    0    0
    [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79]
seq1    0    1    0    0    0    0    1    0    0    0    0
seq2    0    1    0    0    0    0    1    0    0    1    0
seq3    0    1    0    0    0    0    1    0    0    0    0
    [,80] [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89] [,90]
seq1    1    0    0    1    0    0    0    1    0    0    0
seq2    0    0    0    1    0    0    1    0    0    0    1
seq3    1    0    0    1    0    0    0    1    0    0    0
    [,91] [,92] [,93] [,94] [,95] [,96] [,97] [,98] [,99] [,100] [,101]
seq1    0    1    0    0    0    1    0    0    0    1    0
seq2    0    0    1    0    0    0    0    1    0    0    1
seq3    1    0    0    0    0    1    0    1    0    0    0
    [,102] [,103] [,104] [,105] [,106] [,107] [,108] [,109] [,110] [,111]
seq1    0    0    1    0    0    1    0    0    1    0
seq2    0    0    0    0    0    1    0    0    1    0
seq3    0    1    0    0    0    1    0    0    0    1
    [,112] [,113] [,114] [,115] [,116] [,117] [,118] [,119] [,120] [,121]
seq1    0    1    0    0    0    0    0    1    0    0
seq2    0    1    0    0    0    1    0    0    0    0
seq3    0    0    1    0    0    0    0    0    1    0
    [,122] [,123] [,124] [,125] [,126] [,127] [,128] [,129] [,130] [,131]
seq1    1    0    0    1    0    0    0    1    0    0
seq2    0    1    0    0    0    0    1    0    1    0
seq3    1    0    0    0    1    0    0    0    0    0
    [,132] [,133] [,134] [,135] [,136] [,137] [,138] [,139] [,140] [,141]
seq1    0    0    0    1    0    0    0    0    1    0
seq2    0    0    1    0    0    0    0    0    1    0
seq3    1    0    0    1    0    0    0    1    0    0
    [,142] [,143] [,144]   [,145]   [,146]   [,147]   [,148]   [,149]
seq1    0    1    0 0.6477612 0.7402985 0.7910448 0.7223881 0.7910448
seq2    0    1    0 0.6238806 0.5701493 0.5970149 0.6208955 0.5820896
seq3    0    0    1 0.7283582 0.7522388 0.6298507 0.5731343 0.6268657
      [,150]   [,151]   [,152]   [,153]   [,154]   [,155]   [,156]
seq1 0.7611940 0.4507463 0.3522388 0.3940299 0.5014925 0.7791045 0.8089552
seq2 0.5313433 0.6567164 0.5432836 0.5014925 0.7791045 0.8089552 0.5731343
seq3 0.6985075 0.6985075 0.6268657 0.6477612 0.7402985 0.7552239 0.5940299
      [,157]   [,158]   [,159]   [,160]   [,161]   [,162]   [,163]
seq1 0.4686567 0.4537313 0.6955224 0.6805970 0.7283582 0.8328358 0.7552239
seq2 0.5402985 0.7641791 0.8328358 0.7164179 0.7014925 0.8029851 0.7641791
seq3 0.6238806 0.8149254 0.8328358 0.5970149 0.5970149 0.8328358 0.7641791
      [,164]   [,165]   [,166]   [,167]   [,168]   [,169]   [,170]
seq1 0.4417910 0.3611940 0.3582090 0.5701493 0.7641791 0.8746269 0.8776119
seq2 0.6597015 0.7791045 0.8805970 0.7373134 0.7343284 0.7223881 0.6656716
seq3 0.5402985 0.4776119 0.5880597 0.8029851 0.7074627 0.6268657 0.5253731
      [,171]   [,172]   [,173]   [,174]   [,175]   [,176]   [,177]
seq1 0.8417910 0.7223881 0.6656716 0.7761194 0.4985075 0.3850746 0.8489078
seq2 0.7761194 0.4985075 0.2656716 0.4955224 0.5850746 0.6179104 0.8834951
seq3 0.5432836 0.6567164 0.5313433 0.6179104 0.7611940 0.6985075 0.8118932
      [,178]   [,179]   [,180]   [,181]   [,182]   [,183]   [,184]
seq1 0.8125000 0.4296117 0.4399272 0.6334951 0.3458738 0.2779126 0.3598301
seq2 0.9144417 0.7226942 0.6529126 0.9751214 0.7099515 0.6334951 0.9696602
```

```
seq3 0.7754854 0.8367718 0.8847087 0.8361650 0.8040049 0.8040049 0.8361650
     [,185]    [,186]    [,187]    [,188]    [,189]    [,190]    [,191]
seq1 0.9077670 0.8343447 0.7572816 0.5703883 0.5989078 0.9484223 0.6183252
seq2 0.8343447 0.7572816 0.5703883 0.6444175 0.7967233 0.8234223 0.5242718
seq3 0.8489078 0.8125000 0.5406553 0.5376214 0.6559466 0.7906553 0.5242718
     [,192]    [,193]    [,194]    [,195]    [,196]    [,197]    [,198]
seq1 0.6074029 0.5194175 0.5242718 0.6984223 0.4641990 0.2487864 0.1237864
seq2 0.5175971 0.7148058 0.7457524 0.7961165 0.8209951 0.7572816 0.5570388
seq3 0.4690534 0.4690534 0.5242718 0.8234223 0.7967233 0.6783981 0.5109223
     [,199]    [,200]    [,201]    [,202]    [,203]    [,204]    [,205]
seq1 0.1650485 0.3416262 0.7712379 0.7785194 0.5916262 0.8974515 0.7979369
seq2 0.5091019 0.5564320 0.8974515 0.7979369 0.4010922 0.3822816 0.7548544
seq3 0.5928398 0.8246359 0.8361650 0.8513350 0.9696602 0.6334951 0.7099515
     [,206]    [,207]    [,208]    [,209]    [,210]    [,211]    [,212]
seq1 0.4010922 0.3822816 0.7299757 0.3945381 0.3893085 0.7576990 0.2562464
seq2 0.5564320 0.6747573 0.9472087 0.3951191 0.3794306 0.4334689 0.3881464
seq3 0.9472087 0.6656553 0.7324029 0.3805927 0.6310285 0.4218478 0.3701336
     [,213]    [,214]    [,215]    [,216]    [,217]    [,218]    [,219]
seq1 0.8361418 0.5136549 0.3631610 0.2068565 0.2754213 0.2597327 0.3829169
seq2 0.3201627 0.4195235 0.3759442 0.3381755 0.2969204 0.3829169 0.6589192
seq3 0.3794306 0.4020918 0.6101104 0.4020918 0.3962812 0.3893085 0.7280651
     [,220]    [,221]    [,222]    [,223]    [,224]    [,225]    [,226]
seq1 0.6589192 0.3294596 0.3817548 0.3160953 0.4032539 0.7960488 0.3666473
seq2 0.3370134 0.4253341 0.3858222 0.6653109 0.3649041 0.7658338 0.4241720
seq3 0.2382336 0.4369553 0.4677513 0.6717025 0.3393376 0.8256827 0.3393376
     [,227]    [,228]    [,229]    [,230]    [,231]    [,232]    [,233]
seq1 0.6891342 0.4665892 0.2667054 0.2492737 0.3126089 0.3271354 0.7466589
seq2 0.6717025 0.3933759 0.4055782 0.6629866 0.3829169 0.7902382 0.3829169
seq3 0.6653109 0.3858222 0.4049971 0.2963393 0.3910517 0.6943637 0.3974433
     [,234]    [,235]    [,236]    [,237]    [,238]    [,239]    [,240]
seq1 0.4416037 0.7007554 0.3829169 0.3660662 0.7373620 0.2951772 0.3486345
seq2 0.3660662 0.7373620 0.2951772 0.3131900 0.2277745 0.3625799 0.4398605
seq3 0.3910517 0.3230680 0.3381755 0.3759442 0.4288205 0.3335270 0.6583382
     [,241]    [,242]    [,243]    [,244]    [,245]    [,246]
seq1 0.2359094 0.4050633 0.75105485 0.5344585 0.09845288 0.57383966
seq2 0.3358512 0.3206751 0.37130802 0.3699015 0.65260197 0.07594937
seq3 0.4607786 0.7552743 0.20393812 0.3361463 0.81434599 0.38396624
     [,247]    [,248]    [,249]    [,250]    [,251]    [,252]
seq1 0.5907173 0.69901547 0.75808720 0.2405063 0.9479606 0.32208158
seq2 0.4514768 0.71026723 0.06188467 0.9029536 0.3220816 0.46272855
seq3 0.3375527 0.22362869 0.33755274 0.4022504 0.7510549 0.51758087
     [,253]    [,254]    [,255]    [,256]    [,257]    [,258]    [,259]
seq1 0.4627286 0.4472574 0.5836850 0.07735584 0.6343179 0.4078762 0.3811533
seq2 0.4345992 0.5035162 0.3656821 0.45007032 0.3783404 0.3277075 0.6835443
seq3 0.1490858 0.7398031 0.3909986 0.46694796 0.4008439 0.4978903 0.4008439
     [,260]    [,261]    [,262]    [,263]    [,264]    [,265]
seq1 0.4556962 0.5189873 0.6736990 0.80309423 0.8565401 0.6272855
seq2 0.2025316 0.7524613 0.3220816 0.50070323 0.3473980 0.5879044
seq3 0.4500703 0.3656821 0.5007032 0.54008439 0.6526020 0.2672293
     [,266]    [,267]    [,268]    [,269]    [,270]    [,271]
seq1 0.49789030 0.64978903 0.5007032 0.04500703 0.8241913 0.4852321
seq2 0.04922644 0.82419128 0.4852321 0.55836850 0.3094233 0.5991561
seq3 0.36146273 0.37974684 0.8804501 0.06188467 0.7102672 0.4542897
     [,272]    [,273]    [,274]
seq1 0.55836850 0.3066104 0.55836850
seq2 0.69901547 0.4810127 0.03797468
seq3 0.03516174 0.4486639 0.48523207
```

Mad 1-mer

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
seq1	0	0	1	0	0	0	1	0	0	0	1	0	0
seq2	0	1	0	0	0	0	1	0	0	0	1	0	0
seq3	0	0	1	0	0	0	1	0	0	1	0	0	0
seq4	0	0	0	1	0	0	1	0	0	1	0	0	0
seq5	0	0	0	1	0	0	1	0	0	1	0	0	0
seq6	0	0	1	0	0	1	0	0	1	0	0	0	0

	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
seq1	1	0	0	1	0	0	0	0	0	0	1
seq2	0	1	0	0	0	1	0	1	0	0	0
seq3	0	1	0	0	0	1	0	0	1	0	0
seq4	0	1	0	0	0	1	0	0	1	0	0
seq5	0	0	1	0	1	0	0	0	0	1	0
seq6	0	1	0	0	1	0	0	0	0	1	0

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]
seq1	0	0	1	0	1	0	0	0	1	0	0
seq2	0	0	1	0	0	0	1	0	1	0	0
seq3	0	1	0	0	0	1	0	0	0	0	1
seq4	0	0	0	1	0	0	0	1	0	1	0
seq5	0	1	0	0	0	1	0	0	0	0	0
seq6	0	0	1	0	0	1	0	0	0	1	0

	[,36]	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]
seq1	0	1	0	0	0	0	0	1	0	0	1
seq2	0	0	0	1	0	0	1	0	0	0	1
seq3	0	0	0	1	0	0	0	1	0	0	1
seq4	0	0	0	1	0	0	0	1	0	0	1
seq5	1	0	0	1	0	0	1	0	0	0	1
seq6	0	0	1	0	0	0	1	0	0	1	0

	[,47]	[,48]	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]
seq1	0	0	0	1	0	0	1	0	0	0	0
seq2	0	0	1	0	0	0	0	1	0	0	0
seq3	0	0	1	0	0	0	0	0	0	1	0
seq4	0	0	0	0	0	1	0	0	0	1	0
seq5	0	0	0	0	0	1	0	0	1	0	0
seq6	0	0	0	0	0	1	0	1	0	0	0

	[,58]	[,59]	[,60]	[,61]	[,62]	[,63]	[,64]	[,65]	[,66]	[,67]	[,68]
seq1	1	0	0	0	1	0	0	0	0	0	1
seq2	1	0	0	0	1	0	0	1	0	0	0
seq3	1	0	0	0	1	0	0	1	0	0	0
seq4	1	0	0	0	1	0	0	1	0	0	0
seq5	1	0	0	0	1	0	0	1	0	0	0
seq6	1	0	0	0	1	0	0	1	0	0	0

	[,69]	[,70]	[,71]	[,72]	[,73]	[,74]	[,75]	[,76]	[,77]	[,78]	[,79]
seq1	0	1	0	0	0	0	1	0	0	0	0
seq2	0	1	0	0	0	0	1	0	0	1	0
seq3	0	1	0	0	0	0	1	0	0	0	0
seq4	0	1	0	0	0	0	1	0	0	1	0
seq5	0	0	0	1	0	0	1	0	0	0	0
seq6	0	1	0	0	0	0	1	0	0	0	0

	[,80]	[,81]	[,82]	[,83]	[,84]	[,85]	[,86]	[,87]	[,88]	[,89]	[,90]
seq1	1	0	0	1	0	0	0	1	0	0	0
seq2	0	0	0	1	0	0	1	0	0	0	1
seq3	1	0	0	1	0	0	0	1	0	0	0

seq4	0	0	0	1	0	0	0	1	0	0	1
seq5	1	0	0	1	0	0	1	0	0	1	0
seq6	1	0	0	1	0	0	0	1	0	0	0

	[,91]	[,92]	[,93]	[,94]	[,95]	[,96]	[,97]	[,98]	[,99]	[,100]	[,101]
seq1	0	1	0	0	0	1	0	0	0	1	0
seq2	0	0	1	0	0	0	0	1	0	0	1
seq3	1	0	0	0	0	1	0	1	0	0	0
seq4	0	0	1	0	0	0	0	0	0	1	0
seq5	0	0	0	0	1	0	0	1	0	0	0
seq6	0	1	0	0	0	1	1	0	0	0	1

	[,102]	[,103]	[,104]	[,105]	[,106]	[,107]	[,108]	[,109]	[,110]	[,111]
seq1	0	0	1	0	0	1	0	0	1	0
seq2	0	0	0	0	0	1	0	0	1	0
seq3	0	1	0	0	0	1	0	0	0	1
seq4	1	0	0	0	1	0	0	0	0	0
seq5	0	1	0	0	0	1	0	0	1	0
seq6	0	0	0	0	0	1	0	0	0	0

	[,112]	[,113]	[,114]	[,115]	[,116]	[,117]	[,118]	[,119]	[,120]	[,121]
seq1	0	1	0	0	0	0	0	1	0	0
seq2	0	1	0	0	0	1	0	0	0	0
seq3	0	0	1	0	0	0	0	0	1	0
seq4	1	1	0	0	0	0	0	1	0	0
seq5	0	0	1	0	0	0	1	0	0	0
seq6	1	0	0	1	0	0	0	1	0	0

	[,122]	[,123]	[,124]	[,125]	[,126]	[,127]	[,128]	[,129]	[,130]	[,131]
seq1	1	0	0	1	0	0	0	1	0	0
seq2	0	1	0	0	0	0	1	0	1	0
seq3	1	0	0	0	1	0	0	0	0	0
seq4	0	1	0	0	0	0	1	1	0	0
seq5	1	0	0	0	0	0	1	0	0	0
seq6	0	1	0	0	0	0	1	0	0	1

	[,132]	[,133]	[,134]	[,135]	[,136]	[,137]	[,138]	[,139]	[,140]	[,141]
seq1	0	0	0	1	0	0	0	0	1	0
seq2	0	0	1	0	0	0	0	0	1	0
seq3	1	0	0	1	0	0	0	1	0	0
seq4	0	1	0	0	0	0	0	1	0	0
seq5	1	0	0	0	1	0	0	1	0	0
seq6	0	0	0	1	0	0	1	0	0	0

	[,142]	[,143]	[,144]
seq1	0	1	0
seq2	0	1	0
seq3	0	0	1
seq4	1	0	0
seq5	0	0	1
seq6	1	0	0

Max 1-mer+shape

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
seq1	0	0	1	0	1	0	0	0	1	0	0	0	0
seq2	0	0	0	1	0	1	0	0	0	1	0	0	0
seq3	0	0	1	0	0	0	1	0	0	1	0	0	0

	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
seq1	0	1	0	0	1	0	0	0	1	0	0
seq2	1	0	0	0	0	1	0	0	0	1	0
seq3	0	1	0	0	1	0	0	0	0	0	1

```
     [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
seq1    0    1    0    0    0    0    0    1    0    0    1
seq2    0    0    1    0    0    0    1    0    0    1    0
seq3    0    0    1    0    0    1    0    0    0    1    0
     [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
seq1    0    0    0    1    0    0    1    0    0    0    0
seq2    0    0    0    0    1    1    0    0    0    0    0
seq3    0    0    0    1    0    0    0    1    0    1    0
     [,47] [,48] [,49] [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57]
seq1    1    0    0    0    1    0    0    0    1    0    0
seq2    1    0    1    0    0    0    0    0    1    0    0
seq3    0    0    0    1    0    0    0    0    0    1    0
     [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65] [,66] [,67] [,68]
seq1    0    1    0    0    1    0    0    0    0    1    0
seq2    0    1    0    0    1    0    0    1    0    0    0
seq3    0    1    0    0    1    0    0    0    0    1    0
     [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79]
seq1    0    1    0    0    0    0    1    0    0    0    0
seq2    0    0    0    1    0    0    1    0    0    0    0
seq3    0    1    0    0    0    0    1    0    0    0    0
     [,80] [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89] [,90]
seq1    1    0    0    1    0    0    1    0    0    0    1
seq2    1    0    0    1    0    0    0    1    0    1    0
seq3    1    0    0    1    0    0    0    1    0    1    0
     [,91] [,92] [,93] [,94] [,95] [,96] [,97] [,98] [,99] [,100] [,101]
seq1    0    0    0    1    0    0    0    0    1    0    0
seq2    0    0    0    1    0    0    0    0    0    1    0
seq3    0    0    0    0    1    0    0    0    0    1    0
     [,102] [,103] [,104] [,105] [,106] [,107] [,108] [,109] [,110] [,111]
seq1    1    0    0    0    1    0    0    0    0    1
seq2    1    0    0    1    0    0    0    0    0    1
seq3    0    1    0    0    0    1    0    0    1    0
     [,112] [,113] [,114] [,115] [,116] [,117] [,118] [,119] [,120] [,121]
seq1    0    0    1    0    0    0    1    0    0    0
seq2    0    0    0    1    0    1    0    0    0    0
seq3    0    0    0    1    0    0    1    0    0    0
     [,122] [,123] [,124] [,125] [,126] [,127] [,128] [,129] [,130] [,131]
seq1    0    1    0    0    1    0    0    0    1    0
seq2    0    1    0    0    0    1    0    1    0    0
seq3    0    1    0    0    1    0    0    0    0    0
     [,132] [,133] [,134] [,135] [,136] [,137] [,138] [,139] [,140] [,141]
seq1    0    0    0    1    0    0    1    0    0    0
seq2    0    0    0    0    1    0    0    1    0    1
seq3    1    0    0    1    0    0    1    0    0    0
     [,142] [,143] [,144]   [,145]   [,146]   [,147]   [,148]   [,149]
seq1    1    0    0 0.5880597 0.3940299 0.5253731 0.5611940 0.5970149
seq2    0    0    0 0.6268657 0.6985075 0.6985075 0.6238806 0.5880597
seq3    0    0    1 0.7641791 0.6955224 0.6805970 0.8417910 0.7641791
       [,150]   [,151]   [,152]   [,153]   [,154]   [,155]   [,156]
seq1 0.7611940 0.7701493 0.6597015 0.7283582 0.7313433 0.6238806 0.5880597
seq2 0.5253731 0.6417910 0.8417910 0.8925373 0.7641791 0.6328358 0.5970149
seq3 0.6925373 0.7492537 0.7582090 0.6656716 0.4955224 0.6417910 0.7611940
       [,157]   [,158]   [,159]   [,160]   [,161]   [,162]   [,163]
seq1 0.6298507 0.7641791 0.8029851 0.7014925 0.7164179 0.8626866 0.8388060
seq2 0.5880597 0.7402985 0.7910448 0.6388060 0.6865672 0.8805970 0.8149254
seq3 0.8417910 0.8537313 0.8029851 0.7014925 0.7164179 0.8328358 0.8149254
       [,164]   [,165]   [,166]   [,167]   [,168]   [,169]   [,170]
```

seq1 0.6477612 0.6268657 0.7313433 0.7283582 0.6686567 0.7522388 0.7283582
seq2 0.6985075 0.4955224 0.5014925 0.6089552 0.8089552 0.7820896 0.6179104
seq3 0.6477612 0.5611940 0.5761194 0.7343284 0.7791045 0.6597015 0.7641791
[,171] [,172] [,173] [,174] [,175] [,176] [,177]
seq1 0.6686567 0.7522388 0.7283582 0.6686567 0.7522388 0.7283582 0.5157767
seq2 0.5313433 0.6208955 0.5820896 0.4955224 0.5940299 0.6656716 0.8464806
seq3 0.8029851 0.8029851 0.6955224 0.6805970 0.8417910 0.6895522 0.7961165
[,178] [,179] [,180] [,181] [,182] [,183] [,184]
seq1 0.9077670 0.8513350 0.8877427 0.9933252 0.6656553 0.7991505 0.8209951
seq2 0.8040049 0.8040049 0.8834951 0.8725728 0.8513350 0.9277913 0.6049757
seq3 0.7669903 0.8974515 0.5916262 0.8076456 0.8131068 0.7845874 0.8082524
[,185] [,186] [,187] [,188] [,189] [,190] [,191]
seq1 0.8118932 0.8349515 0.8834951 0.8725728 0.8264563 0.7961165 0.7457524
seq2 0.5782767 0.8064320 0.6146845 0.9593447 0.8574029 0.8125000 0.4296117
seq3 0.6334951 0.5564320 0.7615291 0.5291262 0.7682039 0.7627427 0.7457524
[,192] [,193] [,194] [,195] [,196] [,197] [,198]
seq1 0.7148058 0.5175971 0.4921117 0.7949029 0.8489078 0.8361650 0.8349515
seq2 0.3859223 0.4660194 0.5570388 0.7906553 0.6510922 0.5564320 0.8131068
seq3 0.7148058 0.5175971 0.5242718 0.7906553 0.6711165 0.5922330 0.8009709
[,199] [,200] [,201] [,202] [,203] [,204] [,205]
seq1 0.8118932 0.8094660 0.7754854 0.8118932 0.8094660 0.7754854 0.8118932
seq2 0.5467233 0.6328883 0.6796117 0.9472087 0.7099515 0.6529126 0.9751214
seq3 0.5527913 0.7572816 0.8209951 0.7961165 0.7457524 0.7457524 0.7669903
[,206] [,207] [,208] [,209] [,210] [,211] [,212]
seq1 0.8094660 0.7754854 0.8118932 0.3027310 0.3091226 0.2858803 0.3776874
seq2 0.6808252 0.5376214 0.5546117 0.4450901 0.3957002 0.6101104 0.4003486
seq3 0.8974515 0.5916262 0.7833738 0.3986055 0.6647298 0.3933759 0.3811737
[,213] [,214] [,215] [,216] [,217] [,218] [,219]
seq1 0.3480535 0.3160953 0.6484602 0.4055782 0.3840790 0.6176641 0.3962812
seq2 0.3881464 0.3834980 0.3428239 0.3527019 0.7681580 0.4230099 0.4526438
seq3 0.6990122 0.4189425 0.4230099 0.5967461 0.4654271 0.3765253 0.3137711
[,220] [,221] [,222] [,223] [,224] [,225] [,226]
seq1 0.3881464 0.3939570 0.3997676 0.6717025 0.4241720 0.7658338 0.3672284
seq2 0.3521209 0.3747821 0.3718768 0.7576990 0.2498547 0.8553167 0.3805927
seq3 0.4160372 0.7106334 0.4526438 0.6914585 0.4241720 0.7658338 0.3649041
[,227] [,228] [,229] [,230] [,231] [,232] [,233]
seq1 0.6798373 0.4061592 0.3962812 0.3986055 0.6176641 0.3968623 0.4212667
seq2 0.6531087 0.4729808 0.3660662 0.2701917 0.3463103 0.4212667 0.7274840
seq3 0.6717025 0.4584544 0.3858222 0.3579314 0.3190006 0.6647298 0.4055782
[,234] [,235] [,236] [,237] [,238] [,239] [,240]
seq1 0.6310285 0.3968623 0.4212667 0.6310285 0.3968623 0.4212667 0.6310285
seq2 0.3259733 0.4288205 0.3660662 0.3201627 0.4171993 0.4288205 0.2248693
seq3 0.3933759 0.6717025 0.4311447 0.6722836 0.3933759 0.3811737 0.7100523
[,241] [,242] [,243] [,244] [,245] [,246]
seq1 0.3805927 0.5654008 0.18846695 0.9254571 0.36286920 0.4345992
seq2 0.7315514 0.4472574 0.34599156 0.2236287 0.32208158 0.3727145
seq3 0.3399187 0.7454290 0.17721519 0.8129395 0.04219409 0.4472574
[,247] [,248] [,249] [,250] [,251] [,252]
seq1 0.0281294 0.4528833 0.35302391 0.7580872 0.19127989 0.33333333
seq2 0.3727145 0.8945148 0.05907173 0.5007032 0.07594937 0.62025316
seq3 0.7412096 0.3586498 0.25175809 0.3881857 0.70886076 0.54430380
[,253] [,254] [,255] [,256] [,257] [,258]
seq1 0.37271449 0.39803094 0.7412096 0.2025316 0.6835443 0.3277075
seq2 0.06610408 0.42756681 0.7496484 0.5344585 0.1167370 0.6807314
seq3 0.07594937 0.44303797 0.6526020 0.2348805 0.6835443 0.3277075
[,259] [,260] [,261] [,262] [,263] [,264] [,265]
seq1 0.3952180 0.4978903 0.7524613 0.4022504 0.3488045 0.1912799 0.7510549

seq2 0.3319269 0.5091421 0.3980309 0.7046414 0.5597750 0.1462729 0.5794655
seq3 0.3783404 0.4669480 0.4022504 0.6807314 0.1209564 0.4655415 0.4528833
[,266] [,267] [,268] [,269] [,270] [,271]
seq1 0.3628692 0.20393812 0.7510549 0.362869198 0.20393812 0.75105485
seq2 0.4683544 0.01969058 0.4542897 0.697609001 0.07594937 0.46835443
seq3 0.3220816 0.75246132 0.2025316 0.669479606 0.18565401 0.81293952
[,272] [,273] [,274]
seq1 0.36286920 0.20393812 0.7552743
seq2 0.73417722 0.15330520 0.5203938
seq3 0.04219409 0.45428973 0.8481013

Max 1-mer

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
seq1	0	0	1	0	1	0	0	0	1	0	0	0	0
seq2	0	0	0	1	0	1	0	0	0	1	0	0	0
seq3	0	0	1	0	0	0	1	0	0	1	0	0	0
seq4	0	0	1	0	0	1	0	0	1	0	0	0	0
seq5	0	0	1	0	0	0	1	0	1	0	0	0	1
seq6	0	0	1	0	0	1	0	0	1	0	0	0	0

	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
seq1	0	1	0	0	1	0	0	0	1	0	0
seq2	1	0	0	0	0	1	0	0	0	1	0
seq3	0	1	0	0	1	0	0	0	0	0	1
seq4	0	1	0	0	1	0	0	0	0	0	1
seq5	0	0	0	0	0	0	1	0	0	0	1
seq6	0	1	0	0	1	0	0	1	0	0	0

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]
seq1	0	1	0	0	0	0	0	1	0	0	1
seq2	0	0	1	0	0	0	1	0	0	1	0
seq3	0	0	1	0	0	1	0	0	0	1	0
seq4	0	1	0	0	0	0	1	0	0	1	0
seq5	1	0	0	0	0	1	0	0	0	1	0
seq6	1	0	0	0	0	1	0	0	0	0	1

	[,36]	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]
seq1	0	0	0	1	0	0	1	0	0	0	0
seq2	0	0	0	0	1	1	0	0	0	0	0
seq3	0	0	0	1	0	0	0	1	0	1	0
seq4	0	0	1	0	0	0	1	0	0	0	0
seq5	0	0	0	0	1	0	0	0	1	0	1
seq6	0	0	0	1	0	0	0	1	0	1	0

	[,47]	[,48]	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]
seq1	1	0	0	0	1	0	0	0	1	0	0
seq2	1	0	1	0	0	0	0	0	1	0	0
seq3	0	0	0	1	0	0	0	0	0	1	0
seq4	1	0	0	0	1	0	0	0	1	0	0
seq5	0	0	1	0	0	0	0	0	0	1	0
seq6	0	0	0	0	1	0	1	0	0	0	0

	[,58]	[,59]	[,60]	[,61]	[,62]	[,63]	[,64]	[,65]	[,66]	[,67]	[,68]
seq1	0	1	0	0	1	0	0	0	0	1	0
seq2	0	1	0	0	1	0	0	1	0	0	0
seq3	0	1	0	0	1	0	0	0	0	1	0
seq4	1	0	0	0	1	0	0	1	0	0	0
seq5	1	0	0	0	1	0	0	1	0	0	0
seq6	0	1	0	0	1	0	0	0	0	1	0

[,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79]

seq1	0	1	0	0	0	0	1	0	0	0	0
seq2	0	0	0	1	0	0	1	0	0	0	0
seq3	0	1	0	0	0	0	1	0	0	0	0
seq4	0	1	0	0	0	0	1	0	0	1	0
seq5	0	1	0	0	1	0	0	0	0	0	0
seq6	0	1	0	0	0	0	1	0	0	0	0

	[,80]	[,81]	[,82]	[,83]	[,84]	[,85]	[,86]	[,87]	[,88]	[,89]	[,90]
seq1	1	0	0	1	0	0	1	0	0	0	1
seq2	1	0	0	1	0	0	0	1	0	1	0
seq3	1	0	0	1	0	0	0	1	0	1	0
seq4	0	0	0	1	0	0	0	1	0	0	0
seq5	1	0	0	1	0	0	0	1	0	0	0
seq6	1	0	0	1	0	0	1	0	0	0	0

	[,91]	[,92]	[,93]	[,94]	[,95]	[,96]	[,97]	[,98]	[,99]	[,100]	[,101]
seq1	0	0	0	1	0	0	0	0	1	0	0
seq2	0	0	0	1	0	0	0	0	0	1	0
seq3	0	0	0	0	1	0	0	0	0	1	0
seq4	0	1	0	0	1	0	0	1	0	0	0
seq5	0	1	0	0	0	1	0	0	1	0	1
seq6	1	0	0	1	0	0	0	1	0	0	0

	[,102]	[,103]	[,104]	[,105]	[,106]	[,107]	[,108]	[,109]	[,110]	[,111]
seq1	1	0	0	0	1	0	0	0	0	1
seq2	1	0	0	1	0	0	0	0	0	1
seq3	0	1	0	0	0	1	0	0	1	0
seq4	1	0	0	0	0	1	0	0	0	1
seq5	0	0	0	1	0	0	0	0	0	1
seq6	0	1	0	0	1	0	0	0	1	0

	[,112]	[,113]	[,114]	[,115]	[,116]	[,117]	[,118]	[,119]	[,120]	[,121]
seq1	0	0	1	0	0	0	1	0	0	0
seq2	0	0	0	1	0	1	0	0	0	0
seq3	0	0	0	1	0	0	1	0	0	0
seq4	0	0	1	0	0	0	1	0	0	0
seq5	0	0	1	0	0	0	0	0	1	0
seq6	0	0	1	0	0	0	1	0	0	1

	[,122]	[,123]	[,124]	[,125]	[,126]	[,127]	[,128]	[,129]	[,130]	[,131]
seq1	0	1	0	0	1	0	0	0	1	0
seq2	0	1	0	0	0	1	0	1	0	0
seq3	0	1	0	0	1	0	0	0	0	0
seq4	1	0	0	0	0	0	1	0	1	0
seq5	0	1	0	0	0	1	0	0	1	0
seq6	0	0	0	0	0	1	0	1	0	0

	[,132]	[,133]	[,134]	[,135]	[,136]	[,137]	[,138]	[,139]	[,140]	[,141]
seq1	0	0	0	1	0	0	1	0	0	0
seq2	0	0	0	0	1	0	0	1	0	1
seq3	1	0	0	1	0	0	1	0	0	0
seq4	0	0	0	1	0	0	0	1	0	0
seq5	0	0	0	0	1	0	0	1	0	0
seq6	0	0	1	0	0	0	0	1	0	1

	[,142]	[,143]	[,144]
seq1	1	0	0
seq2	0	0	0
seq3	0	0	1
seq4	1	0	0
seq5	0	1	0
seq6	0	0	0

Myc 1-mer+shape

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]
seq1	1	0	0	0	0	1	0	0	0	1	0	0	0
seq2	1	0	0	0	1	0	0	0	0	1	0	0	1
seq3	0	0	1	0	0	1	0	0	0	0	1	0	0

	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]
seq1	0	1	0	1	0	0	0	0	1	0	0
seq2	0	0	0	0	0	1	0	0	1	0	0
seq3	0	1	0	0	1	0	0	0	1	0	0

	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]	[,33]	[,34]	[,35]
seq1	0	1	0	0	0	0	1	0	0	0	1
seq2	0	0	1	0	0	1	0	0	0	1	0
seq3	0	0	1	0	0	0	1	0	0	0	0

	[,36]	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]	[,43]	[,44]	[,45]	[,46]
seq1	0	0	1	0	0	0	0	1	0	0	1
seq2	0	1	0	0	0	0	1	0	0	0	1
seq3	1	0	1	0	0	0	0	0	1	0	0

	[,47]	[,48]	[,49]	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]	[,57]
seq1	0	0	0	0	1	0	0	0	1	0	0
seq2	0	0	0	0	1	0	0	0	1	0	0
seq3	1	0	0	1	0	0	1	0	0	0	0

	[,58]	[,59]	[,60]	[,61]	[,62]	[,63]	[,64]	[,65]	[,66]	[,67]	[,68]
seq1	0	1	0	0	1	0	0	1	0	0	0
seq2	1	0	0	0	1	0	0	0	0	0	1
seq3	1	0	0	0	1	0	0	1	0	0	0

	[,69]	[,70]	[,71]	[,72]	[,73]	[,74]	[,75]	[,76]	[,77]	[,78]	[,79]
seq1	0	1	0	0	0	0	1	0	1	0	0
seq2	0	1	0	0	0	0	1	0	0	0	0
seq3	0	0	0	1	0	0	1	0	0	1	0

	[,80]	[,81]	[,82]	[,83]	[,84]	[,85]	[,86]	[,87]	[,88]	[,89]	[,90]
seq1	0	0	0	1	0	0	0	1	0	0	1
seq2	1	0	0	1	0	0	1	0	0	1	0
seq3	0	0	0	0	1	0	0	1	0	0	1

	[,91]	[,92]	[,93]	[,94]	[,95]	[,96]	[,97]	[,98]	[,99]	[,100]	[,101]
seq1	0	0	1	0	0	0	1	0	0	0	0
seq2	0	0	0	1	0	0	0	0	0	1	0
seq3	0	0	0	0	1	0	1	0	0	0	1

	[,102]	[,103]	[,104]	[,105]	[,106]	[,107]	[,108]	[,109]	[,110]	[,111]
seq1	0	0	1	0	0	1	0	0	0	1
seq2	0	0	1	0	1	0	0	0	0	0
seq3	0	0	0	0	1	0	0	0	0	1

	[,112]	[,113]	[,114]	[,115]	[,116]	[,117]	[,118]	[,119]	[,120]	[,121]
seq1	0	0	1	0	0	0	0	1	0	0
seq2	1	0	0	0	1	0	1	0	0	0
seq3	0	0	0	0	1	0	1	0	0	0

	[,122]	[,123]	[,124]	[,125]	[,126]	[,127]	[,128]	[,129]	[,130]	[,131]
seq1	0	1	0	0	1	0	0	0	1	0
seq2	1	0	0	1	0	0	0	0	1	0
seq3	1	0	0	0	0	1	0	0	0	0

	[,132]	[,133]	[,134]	[,135]	[,136]	[,137]	[,138]	[,139]	[,140]	[,141]
seq1	0	0	0	1	0	0	0	1	0	0
seq2	0	0	0	0	1	0	0	1	0	0
seq3	1	0	1	0	0	0	1	0	0	0

	[,142]	[,143]	[,144]	[,145]	[,146]	[,147]	[,148]	[,149]
seq1	0	1	0	0.6477612	0.8029851	0.5880597	0.5492537	0.6895522
seq2	0	0	1	0.6268657	0.7343284	0.6805970	0.6955224	0.7641791

```
seq3     0     0     1 0.7522388 0.6298507 0.6298507 0.7492537 0.6895522
      [,150]   [,151]   [,152]   [,153]   [,154]   [,155]   [,156]
seq1 0.7492537 0.6686567 0.7641791 0.8029851 0.7940299 0.7313433 0.6268657
seq2 0.6597015 0.7791045 0.8089552 0.6179104 0.6895522 0.7492537 0.6298507
seq3 0.5492537 0.5611940 0.7074627 0.8656716 0.8776119 0.9223881 0.8089552
      [,157]   [,158]   [,159]   [,160]   [,161]   [,162]   [,163]
seq1 0.6477612 0.8388060 0.8626866 0.7283582 0.6805970 0.6955224 0.5970149
seq2 0.5582090 0.5970149 0.6955224 0.6805970 0.7283582 0.8626866 0.9223881
seq3 0.5940299 0.6328358 0.7432836 0.7910448 0.7104478 0.7223881 0.8417910
      [,164]   [,165]   [,166]   [,167]   [,168]   [,169]   [,170]
seq1 0.5880597 0.7313433 0.6865672 0.5492537 0.6656716 0.7104478 0.6597015
seq2 0.9223881 0.7313433 0.3850746 0.4358209 0.4925373 0.5820896 0.4985075
seq3 0.8417910 0.8447761 0.8000000 0.6716418 0.5641791 0.4059701 0.5522388
      [,171]   [,172]   [,173]   [,174]   [,175]   [,176]   [,177]
seq1 0.7283582 0.7522388 0.6298507 0.6298507 0.7492537 0.6985075 0.7415049
seq2 0.5313433 0.6328358 0.8149254 0.7343284 0.6358209 0.7014925 0.4399272
seq3 0.6477612 0.6656716 0.6597015 0.6716418 0.6477612 0.5850746 0.7754854
      [,178]   [,179]   [,180]   [,181]   [,182]   [,183]   [,184]
seq1 0.5928398 0.5109223 0.6383495 0.6966019 0.7845874 0.8094660 0.7961165
seq2 0.5564320 0.8974515 0.7669903 0.7961165 0.8209951 0.7572816 0.5703883
seq3 0.8367718 0.8367718 0.7845874 0.6966019 0.6383495 0.5139563 0.7609223
      [,185]   [,186]   [,187]   [,188]   [,189]   [,190]   [,191]
seq1 0.7457524 0.7439320 0.8349515 0.8361650 0.8489078 0.7949029 0.4921117
seq2 0.6037621 0.6966019 0.7845874 0.8367718 0.8683252 0.9593447 0.6183252
seq3 0.5958738 0.7785194 0.7748786 0.5230583 0.5770631 0.7390777 0.4550971
      [,192]   [,193]   [,194]   [,195]   [,196]   [,197]   [,198]
seq1 0.5194175 0.6074029 0.6183252 0.9593447 0.8574029 0.8112864 0.3586165
seq2 0.6074029 0.5194175 0.4921117 0.7748786 0.7748786 0.5358010 0.7299757
seq3 0.4296117 0.7785194 0.8974515 0.5916262 0.7682039 0.7706311 0.6328883
      [,199]   [,200]   [,201]   [,202]   [,203]   [,204]   [,205]
seq1 0.3695388 0.4569175 0.8125000 0.8209951 0.8118932 0.7754854 0.8367718
seq2 0.4660194 0.3768204 0.7918689 0.5054612 0.4290049 0.6862864 0.7906553
seq3 0.3658981 0.3307039 0.3713592 0.5212379 0.5097087 0.6334951 0.7979369
      [,206]   [,207]   [,208]   [,209]   [,210]   [,211]   [,212]
seq1 0.8367718 0.7845874 0.8040049 0.4433469 0.6943637 0.3910517 0.3160953
seq2 0.5527913 0.7275485 0.5091019 0.3585125 0.7972109 0.3805927 0.3933759
seq3 0.5679612 0.5097087 0.6747573 0.6595003 0.4218478 0.3817548 0.4218478
      [,213]   [,214]   [,215]   [,216]   [,217]   [,218]   [,219]
seq1 0.4607786 0.6321906 0.4206856 0.4061592 0.6717025 0.4340500 0.6438117
seq2 0.6647298 0.3933759 0.4055782 0.6589192 0.3521209 0.4770482 0.6321906
seq3 0.6321906 0.4607786 0.3259733 0.3986055 0.4288205 0.7164439 0.4555491
      [,220]   [,221]   [,222]   [,223]   [,224]   [,225]   [,226]
seq1 0.3986055 0.3962812 0.4061592 0.6798373 0.3683905 0.7960488 0.4032539
seq2 0.4218478 0.3718768 0.3794306 0.3463103 0.4032539 0.7960488 0.3683905
seq3 0.6670540 0.3375944 0.4427658 0.7303893 0.2684486 0.7675770 0.3904707
      [,227]   [,228]   [,229]   [,230]   [,231]   [,232]   [,233]
seq1 0.3463103 0.3747821 0.3625799 0.7269030 0.3370134 0.2498547 0.7094712
seq2 0.6746078 0.4538059 0.6600813 0.2638001 0.3201627 0.2568274 0.4317257
seq3 0.3829169 0.7065660 0.4456711 0.7292272 0.4840209 0.3166764 0.3131900
      [,234]   [,235]   [,236]   [,237]   [,238]   [,239]   [,240]
seq1 0.3864033 0.3840790 0.6310285 0.4218478 0.3817548 0.4218478 0.6141778
seq2 0.3794306 0.2847182 0.4601976 0.4584544 0.6548518 0.3428239 0.4108077
seq3 0.8082510 0.3364323 0.3951191 0.4218478 0.6624056 0.3538640 0.3817548
      [,241]   [,242]   [,243]    [,244]   [,245]   [,246]
seq1 0.4055782 0.4556962 0.2517581 0.65260197 0.5288326 0.4500703
seq2 0.7635096 0.5541491 0.5668073 0.04641350 0.8129395 0.1772152
seq3 0.4421848 0.2137834 0.3361463 0.83122363 0.3347398 0.2222222
```

```
      [,247]    [,248]    [,249]    [,250]    [,251]    [,252]
seq1 0.22222222 0.3600563 0.74542897 0.20253165 0.6582278 0.19549930
seq2 0.75246132 0.3220816 0.46272855 0.42756681 0.4641350 0.22222222
seq3 0.45007032 0.5288326 0.74964838 0.05063291 0.4641350 0.64838256
      [,253]    [,254]    [,255]    [,256]    [,257]    [,258]
seq1 0.3488045 0.4022504 0.7524613 0.49789030 0.39943741 0.4078762
seq2 0.3347398 0.8171589 0.4036568 0.07032349 0.63431786 0.4078762
seq3 0.4936709 0.4613221 0.5260197 0.53867792 0.08860759 0.5274262
      [,259]    [,260]    [,261]    [,262]    [,263]    [,264]
seq1 0.6343179 0.07032349 0.4275668 0.7538678 0.5007032 0.60056259
seq2 0.3994374 0.52883263 0.6357243 0.4739803 0.5344585 0.28129395
seq3 0.8199719 0.04500703 0.4824191 0.6540084 0.2911392 0.65119550
      [,265]    [,266]    [,267]    [,268]    [,269]    [,270]    [,271]
seq1 0.1870605 0.5569620 0.3431786 0.7580872 0.2039381 0.3361463 0.8312236
seq2 0.6188467 0.8045007 0.1997187 0.5752461 0.6722925 0.3980309 0.4627286
seq3 0.6610408 0.5808720 0.4880450 0.4725738 0.7327707 0.4585091 0.2967651
      [,272]    [,273]    [,274]
seq1 0.3347398 0.22925457 0.32348805
seq2 0.4781997 0.10407876 0.53164557
seq3 0.4374121 0.72292546 0.49929677
```

Myc 1-mer

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
seq1    0    0    1    0    1    0    0    0    1     0     0     0     0
seq2    0    0    0    1    0    1    0    0    0     1     0     0     0
seq3    0    0    1    0    0    0    1    0    0     1     0     0     0
     [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
seq1     0     1     0     0     1     0     0     0     1     0     0
seq2     1     0     0     0     0     1     0     0     0     1     0
seq3     0     1     0     0     1     0     0     0     0     0     1
     [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
seq1     0     1     0     0     0     0     0     1     0     0     1
seq2     0     0     1     0     0     0     1     0     0     1     0
seq3     0     0     1     0     0     1     0     0     0     1     0
     [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
seq1     0     0     0     1     0     0     1     0     0     0     0
seq2     0     0     0     0     1     1     0     0     0     0     0
seq3     0     0     0     1     0     0     0     1     0     1     0
     [,47] [,48] [,49] [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57]
seq1     1     0     0     0     1     0     0     0     1     0     0
seq2     1     0     1     0     0     0     0     0     1     0     0
seq3     0     0     0     1     0     0     0     0     0     1     0
     [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65] [,66] [,67] [,68]
seq1     0     1     0     0     1     0     0     0     0     1     0
seq2     0     1     0     0     1     0     0     1     0     0     0
seq3     0     1     0     0     1     0     0     0     0     1     0
     [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78] [,79]
seq1     0     1     0     0     0     0     1     0     0     0     0
seq2     0     0     0     1     0     0     1     0     0     0     0
seq3     0     1     0     0     0     0     1     0     0     0     0
     [,80] [,81] [,82] [,83] [,84] [,85] [,86] [,87] [,88] [,89] [,90]
seq1     1     0     0     1     0     0     1     0     0     0     1
seq2     1     0     0     1     0     0     0     1     0     1     0
seq3     1     0     0     1     0     0     0     1     0     1     0
     [,91] [,92] [,93] [,94] [,95] [,96] [,97] [,98] [,99] [,100] [,101]
```

```
seq1   0   0   0   1   0   0   0   0   1   0   0
seq2   0   0   0   1   0   0   0   0   0   1   0
seq3   0   0   0   0   1   0   0   0   0   1   0
   [,102] [,103] [,104] [,105] [,106] [,107] [,108] [,109] [,110] [,111]
seq1   1   0   0   0   1   0   0   0   0   1
seq2   1   0   0   1   0   0   0   0   0   1
seq3   0   1   0   0   0   1   0   0   1   0
   [,112] [,113] [,114] [,115] [,116] [,117] [,118] [,119] [,120] [,121]
seq1   0   0   1   0   0   0   1   0   0   0
seq2   0   0   0   1   0   1   0   0   0   0
seq3   0   0   0   1   0   0   1   0   0   0
   [,122] [,123] [,124] [,125] [,126] [,127] [,128] [,129] [,130] [,131]
seq1   0   1   0   0   1   0   0   0   1   0
seq2   0   1   0   0   0   1   0   1   0   0
seq3   0   1   0   0   1   0   0   0   0   0
   [,132] [,133] [,134] [,135] [,136] [,137] [,138] [,139] [,140] [,141]
seq1   0   0   0   1   0   0   1   0   0   0
seq2   0   0   0   0   1   0   0   1   0   1
seq3   1   0   0   1   0   0   1   0   0   0
   [,142] [,143] [,144]   [,145]   [,146]   [,147]   [,148]   [,149]
seq1   1   0   0 0.5880597 0.3940299 0.5253731 0.5611940 0.5970149
seq2   0   0   0 0.6268657 0.6985075 0.6985075 0.6238806 0.5880597
seq3   0   0   1 0.7641791 0.6955224 0.6805970 0.8417910 0.7641791
     [,150]   [,151]   [,152]   [,153]   [,154]   [,155]   [,156]
seq1 0.7611940 0.7701493 0.6597015 0.7283582 0.7313433 0.6238806 0.5880597
seq2 0.5253731 0.6417910 0.8417910 0.8925373 0.7641791 0.6328358 0.5970149
seq3 0.6925373 0.7492537 0.7582090 0.6656716 0.4955224 0.6417910 0.7611940
     [,157]   [,158]   [,159]   [,160]   [,161]   [,162]   [,163]
seq1 0.6298507 0.7641791 0.8029851 0.7014925 0.7164179 0.8626866 0.8388060
seq2 0.5880597 0.7402985 0.7910448 0.6388060 0.6865672 0.8805970 0.8149254
seq3 0.8417910 0.8537313 0.8029851 0.7014925 0.7164179 0.8328358 0.8149254
     [,164]   [,165]   [,166]   [,167]   [,168]   [,169]   [,170]
seq1 0.6477612 0.6268657 0.7313433 0.7283582 0.6686567 0.7522388 0.7283582
seq2 0.6985075 0.4955224 0.5014925 0.6089552 0.8089552 0.7820896 0.6179104
seq3 0.6477612 0.5611940 0.5761194 0.7343284 0.7791045 0.6597015 0.7641791
     [,171]   [,172]   [,173]   [,174]   [,175]   [,176]   [,177]
seq1 0.6686567 0.7522388 0.7283582 0.6686567 0.7522388 0.7283582 0.5157767
seq2 0.5313433 0.6208955 0.5820896 0.4955224 0.5940299 0.6656716 0.8464806
seq3 0.8029851 0.8029851 0.6955224 0.6805970 0.8417910 0.6895522 0.7961165
     [,178]   [,179]   [,180]   [,181]   [,182]   [,183]   [,184]
seq1 0.9077670 0.8513350 0.8877427 0.9933252 0.6656553 0.7991505 0.8209951
seq2 0.8040049 0.8040049 0.8834951 0.8725728 0.8513350 0.9277913 0.6049757
seq3 0.7669903 0.8974515 0.5916262 0.8076456 0.8131068 0.7845874 0.8082524
     [,185]   [,186]   [,187]   [,188]   [,189]   [,190]   [,191]
seq1 0.8118932 0.8349515 0.8834951 0.8725728 0.8264563 0.7961165 0.7457524
seq2 0.5782767 0.8064320 0.6146845 0.9593447 0.8574029 0.8125000 0.4296117
seq3 0.6334951 0.5564320 0.7615291 0.5291262 0.7682039 0.7627427 0.7457524
     [,192]   [,193]   [,194]   [,195]   [,196]   [,197]   [,198]
seq1 0.7148058 0.5175971 0.4921117 0.7949029 0.8489078 0.8361650 0.8349515
seq2 0.3859223 0.4660194 0.5570388 0.7906553 0.6510922 0.5564320 0.8131068
seq3 0.7148058 0.5175971 0.5242718 0.7906553 0.6711165 0.5922330 0.8009709
     [,199]   [,200]   [,201]   [,202]   [,203]   [,204]   [,205]
seq1 0.8118932 0.8094660 0.7754854 0.8118932 0.8094660 0.7754854 0.8118932
seq2 0.5467233 0.6328883 0.6796117 0.9472087 0.7099515 0.6529126 0.9751214
seq3 0.5527913 0.7572816 0.8209951 0.7961165 0.7457524 0.7457524 0.7669903
     [,206]   [,207]   [,208]   [,209]   [,210]   [,211]   [,212]
seq1 0.8094660 0.7754854 0.8118932 0.3027310 0.3091226 0.2858803 0.3776874
```

seq2 0.6808252 0.5376214 0.5546117 0.4450901 0.3957002 0.6101104 0.4003486
seq3 0.8974515 0.5916262 0.7833738 0.3986055 0.6647298 0.3933759 0.3811737
[,213] [,214] [,215] [,216] [,217] [,218] [,219]
seq1 0.3480535 0.3160953 0.6484602 0.4055782 0.3840790 0.6176641 0.3962812
seq2 0.3881464 0.3834980 0.3428239 0.3527019 0.7681580 0.4230099 0.4526438
seq3 0.6990122 0.4189425 0.4230099 0.5967461 0.4654271 0.3765253 0.3137711
[,220] [,221] [,222] [,223] [,224] [,225] [,226]
seq1 0.3881464 0.3939570 0.3997676 0.6717025 0.4241720 0.7658338 0.3672284
seq2 0.3521209 0.3747821 0.3718768 0.7576990 0.2498547 0.8553167 0.3805927
seq3 0.4160372 0.7106334 0.4526438 0.6914585 0.4241720 0.7658338 0.3649041
[,227] [,228] [,229] [,230] [,231] [,232] [,233]
seq1 0.6798373 0.4061592 0.3962812 0.3986055 0.6176641 0.3968623 0.4212667
seq2 0.6531087 0.4729808 0.3660662 0.2701917 0.3463103 0.4212667 0.7274840
seq3 0.6717025 0.4584544 0.3858222 0.3579314 0.3190006 0.6647298 0.4055782
[,234] [,235] [,236] [,237] [,238] [,239] [,240]
seq1 0.6310285 0.3968623 0.4212667 0.6310285 0.3968623 0.4212667 0.6310285
seq2 0.3259733 0.4288205 0.3660662 0.3201627 0.4171993 0.4288205 0.2248693
seq3 0.3933759 0.6717025 0.4311447 0.6722836 0.3933759 0.3811737 0.7100523
[,241] [,242] [,243] [,244] [,245] [,246]
seq1 0.3805927 0.5654008 0.18846695 0.9254571 0.36286920 0.4345992
seq2 0.7315514 0.4472574 0.34599156 0.2236287 0.32208158 0.3727145
seq3 0.3399187 0.7454290 0.17721519 0.8129395 0.04219409 0.4472574
[,247] [,248] [,249] [,250] [,251] [,252]
seq1 0.0281294 0.4528833 0.35302391 0.7580872 0.19127989 0.33333333
seq2 0.3727145 0.8945148 0.05907173 0.5007032 0.07594937 0.62025316
seq3 0.7412096 0.3586498 0.25175809 0.3881857 0.70886076 0.54430380
[,253] [,254] [,255] [,256] [,257] [,258]
seq1 0.37271449 0.39803094 0.7412096 0.2025316 0.6835443 0.3277075
seq2 0.06610408 0.42756681 0.7496484 0.5344585 0.1167370 0.6807314
seq3 0.07594937 0.44303797 0.6526020 0.2348805 0.6835443 0.3277075
[,259] [,260] [,261] [,262] [,263] [,264] [,265]
seq1 0.3952180 0.4978903 0.7524613 0.4022504 0.3488045 0.1912799 0.7510549
seq2 0.3319269 0.5091421 0.3980309 0.7046414 0.5597750 0.1462729 0.5794655
seq3 0.3783404 0.4669480 0.4022504 0.6807314 0.1209564 0.4655415 0.4528833
[,266] [,267] [,268] [,269] [,270] [,271]
seq1 0.3628692 0.20393812 0.7510549 0.362869198 0.20393812 0.75105485
seq2 0.4683544 0.01969058 0.4542897 0.697609001 0.07594937 0.46835443
seq3 0.3220816 0.75246132 0.2025316 0.669479606 0.18565401 0.81293952
[,272] [,273] [,274]
seq1 0.36286920 0.20393812 0.7552743
seq2 0.73417722 0.15330520 0.5203938
seq3 0.04219409 0.45428973 0.8481013

(b)

Max 1mer+shape

```
glmnet

8568 samples
 274 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 7712, 7712, 7712, 7712, 7711, 7709, ...
Resampling results across tuning parameters:
```

lambda	RMSE	Rsquared
3.051758e-05	0.07979423	0.8642443
6.103516e-05	0.07979423	0.8642443
1.220703e-04	0.07979423	0.8642443
2.441406e-04	0.07979423	0.8642443
4.882812e-04	0.07979423	0.8642443
9.765625e-04	0.07979423	0.8642443
1.953125e-03	0.07979423	0.8642443
3.906250e-03	0.07979423	0.8642443
7.812500e-03	0.07979423	0.8642443
1.562500e-02	0.08290638	0.8540335
3.125000e-02	0.08986680	0.8312288
6.250000e-02	0.09769390	0.8074125
1.250000e-01	0.10777214	0.7813023
2.500000e-01	0.12195072	0.7481568
5.000000e-01	0.14043919	0.6996691
1.000000e+00	0.16036402	0.6320349
2.000000e+00	0.17803561	0.5563498
4.000000e+00	0.19170124	0.4906228
8.000000e+00	0.20111800	0.4453291
1.600000e+01	0.20705524	0.4180406
3.200000e+01	0.21046932	0.4031887
6.400000e+01	0.21232865	0.3953357
1.280000e+02	0.21430041	NaN
2.560000e+02	0.21430041	NaN
5.120000e+02	0.21430041	NaN
1.024000e+03	0.21430041	NaN
2.048000e+03	0.21430041	NaN
4.096000e+03	0.21430041	NaN
8.192000e+03	0.21430041	NaN
1.638400e+04	0.21430041	NaN
3.276800e+04	0.21430041	NaN

Max 1mer

```
glmnet

8568 samples
 144 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 7711, 7711, 7711, 7712, 7712, 7711, ...
Resampling results across tuning parameters:
```

lambda	RMSE	Rsquared
3.051758e-05	0.09951122	0.7855663
6.103516e-05	0.09951122	0.7855663
1.220703e-04	0.09951122	0.7855663
2.441406e-04	0.09951122	0.7855663
4.882812e-04	0.09951122	0.7855663
9.765625e-04	0.09951122	0.7855663
1.953125e-03	0.09951122	0.7855663
3.906250e-03	0.09951122	0.7855663
7.812500e-03	0.09952114	0.7855599

```
  1.562500e-02  0.10023432  0.7851275
  3.125000e-02  0.10262741  0.7836586
  6.250000e-02  0.10921830  0.7790695
  1.250000e-01  0.12294778  0.7666436
  2.500000e-01  0.14326284  0.7396933
  5.000000e-01  0.16519530  0.6946632
  1.000000e+00  0.18353022  0.6392797
  2.000000e+00  0.19646395  0.5879465
  4.000000e+00  0.20452072  0.5507817
  8.000000e+00  0.20914209  0.5278996
  1.600000e+01  0.21164418  0.5151166
  3.200000e+01  0.21294796  0.5083854
  6.400000e+01  0.21361675  0.5048975
  1.280000e+02  0.21429409         NaN
  2.560000e+02  0.21429409         NaN
  5.120000e+02  0.21429409         NaN
  1.024000e+03  0.21429409         NaN
  2.048000e+03  0.21429409         NaN
  4.096000e+03  0.21429409         NaN
  8.192000e+03  0.21429409         NaN
  1.638400e+04  0.21429409         NaN
  3.276800e+04  0.21429409         NaN
```

Mad 1mer+shape

```
glmnet

7534 samples
 274 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 6782, 6779, 6779, 6780, 6781, 6782, ...
Resampling results across tuning parameters:

  lambda        RMSE       Rsquared
  3.051758e-05  0.3018130  0.8628934
  6.103516e-05  0.3018130  0.8628934
  1.220703e-04  0.3018130  0.8628934
  2.441406e-04  0.3018130  0.8628934
  4.882812e-04  0.3018130  0.8628934
  9.765625e-04  0.3018130  0.8628934
  1.953125e-03  0.3018130  0.8628934
  3.906250e-03  0.3018130  0.8628934
  7.812500e-03  0.3018130  0.8628934
  1.562500e-02  0.3018130  0.8628934
  3.125000e-02  0.3018130  0.8628934
  6.250000e-02  0.3164769  0.8494464
  1.250000e-01  0.3511398  0.8164869
  2.500000e-01  0.3841694  0.7864097
  5.000000e-01  0.4213171  0.7589079
  1.000000e+00  0.4717558  0.7287667
  2.000000e+00  0.5383830  0.6878240
```

```
  4.000000e+00  0.6111583  0.6315115
  8.000000e+00  0.6760171  0.5668379
  1.600000e+01  0.7249502  0.5098011
  3.200000e+01  0.7578121  0.4691378
  6.400000e+01  0.7776431  0.4447251
  1.280000e+02  0.7888355  0.4311162
  2.560000e+02  0.7947885  0.4239883
  5.120000e+02  0.8010416         NaN
  1.024000e+03  0.8010416         NaN
  2.048000e+03  0.8010416         NaN
  4.096000e+03  0.8010416         NaN
  8.192000e+03  0.8010416         NaN
  1.638400e+04  0.8010416         NaN
  3.276800e+04  0.8010416         NaN
```

Mad 1mer

```
glmnet

7534 samples
 144 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 6782, 6780, 6780, 6782, 6782, 6779, ...
Resampling results across tuning parameters:

  lambda        RMSE       Rsquared
  3.051758e-05  0.3814439  0.7746827
  6.103516e-05  0.3814439  0.7746827
  1.220703e-04  0.3814439  0.7746827
  2.441406e-04  0.3814439  0.7746827
  4.882812e-04  0.3814439  0.7746827
  9.765625e-04  0.3814439  0.7746827
  1.953125e-03  0.3814439  0.7746827
  3.906250e-03  0.3814439  0.7746827
  7.812500e-03  0.3814439  0.7746827
  1.562500e-02  0.3814439  0.7746827
  3.125000e-02  0.3816530  0.7746503
  6.250000e-02  0.3847175  0.7742195
  1.250000e-01  0.3949555  0.7727056
  2.500000e-01  0.4223560  0.7680281
  5.000000e-01  0.4770465  0.7559688
  1.000000e+00  0.5542697  0.7320388
  2.000000e+00  0.6343252  0.6963678
  4.000000e+00  0.6991083  0.6568682
  8.000000e+00  0.7433124  0.6228006
  1.600000e+01  0.7700024  0.5988878
  3.200000e+01  0.7848826  0.5843705
  6.400000e+01  0.7927978  0.5762663
  1.280000e+02  0.7968704  0.5720114
  2.560000e+02  0.7990252  0.5698101
  5.120000e+02  0.8010430         NaN
  1.024000e+03  0.8010430         NaN
  2.048000e+03  0.8010430         NaN
```

```
  4.096000e+03  0.8010430        NaN
  8.192000e+03  0.8010430        NaN
  1.638400e+04  0.8010430        NaN
  3.276800e+04  0.8010430        NaN
```

Myc 1mer+shape

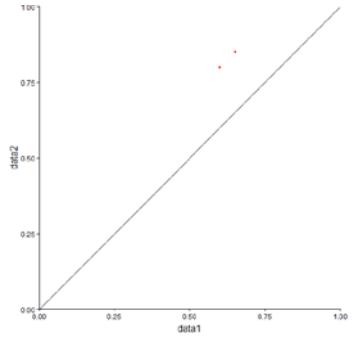
```
glmnet

6926 samples
 274 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 6234, 6234, 6233, 6233, 6233, 6233, ...
Resampling results across tuning parameters:

  lambda        RMSE       Rsquared
  3.051758e-05  0.3047096  0.8549888
  6.103516e-05  0.3047096  0.8549888
  1.220703e-04  0.3047096  0.8549888
  2.441406e-04  0.3047096  0.8549888
  4.882812e-04  0.3047096  0.8549888
  9.765625e-04  0.3047096  0.8549888
  1.953125e-03  0.3047096  0.8549888
  3.906250e-03  0.3047096  0.8549888
  7.812500e-03  0.3047096  0.8549888
  1.562500e-02  0.3047096  0.8549888
  3.125000e-02  0.3047096  0.8549888
  6.250000e-02  0.3241223  0.8370296
  1.250000e-01  0.3480953  0.8153897
  2.500000e-01  0.3757670  0.7936026
  5.000000e-01  0.4138085  0.7695631
  1.000000e+00  0.4694384  0.7376329
  2.000000e+00  0.5405449  0.6912764
  4.000000e+00  0.6142656  0.6296857
  8.000000e+00  0.6770883  0.5636472
  1.600000e+01  0.7232968  0.5073935
  3.200000e+01  0.7535666  0.4679809
  6.400000e+01  0.7715807  0.4441754
  1.280000e+02  0.7816159  0.4309029
  2.560000e+02  0.7869178  0.4239328
  5.120000e+02  0.7924612        NaN
  1.024000e+03  0.7924612        NaN
  2.048000e+03  0.7924612        NaN
  4.096000e+03  0.7924612        NaN
  8.192000e+03  0.7924612        NaN
  1.638400e+04  0.7924612        NaN
  3.276800e+04  0.7924612        NaN
```

Myc 1mer

```
glmnet

6926 samples
 144 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 6234, 6234, 6234, 6234, 6234, 6234, ...
Resampling results across tuning parameters:

  lambda        RMSE       Rsquared
  3.051758e-05  0.3739013  0.7781748
  6.103516e-05  0.3739013  0.7781748
  1.220703e-04  0.3739013  0.7781748
  2.441406e-04  0.3739013  0.7781748
  4.882812e-04  0.3739013  0.7781748
  9.765625e-04  0.3739013  0.7781748
  1.953125e-03  0.3739013  0.7781748
  3.906250e-03  0.3739013  0.7781748
  7.812500e-03  0.3739013  0.7781748
  1.562500e-02  0.3739013  0.7781748
  3.125000e-02  0.3742715  0.7781077
  6.250000e-02  0.3773222  0.7775714
  1.250000e-01  0.3874122  0.7757523
  2.500000e-01  0.4142908  0.7702824
  5.000000e-01  0.4677354  0.7566756
  1.000000e+00  0.5435006  0.7308971
  2.000000e+00  0.6228532  0.6947852
  4.000000e+00  0.6879662  0.6573331
  8.000000e+00  0.7330480  0.6264127
  1.600000e+01  0.7604246  0.6053231
  3.200000e+01  0.7757943  0.5925712
  6.400000e+01  0.7839633  0.5855116
  1.280000e+02  0.7881827  0.5817870
  2.560000e+02  0.7924835  0.5747232
  5.120000e+02  0.7925007        NaN
  1.024000e+03  0.7925007        NaN
  2.048000e+03  0.7925007        NaN
  4.096000e+03  0.7925007        NaN
  8.192000e+03  0.7925007        NaN
  1.638400e+04  0.7925007        NaN
  3.276800e+04  0.7925007        NaN
```

(5) High-throughput in vitro data analysis

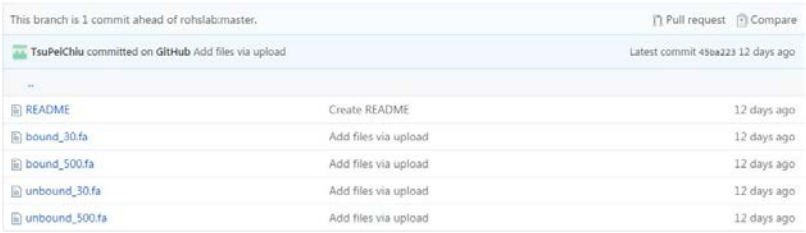
(a)



(b) 1mer+shape model shows larger R squared than 1mer model, which means 1mer+shape model is better.

(6) Preparation of high-throughput in vivo data analysis

(a)

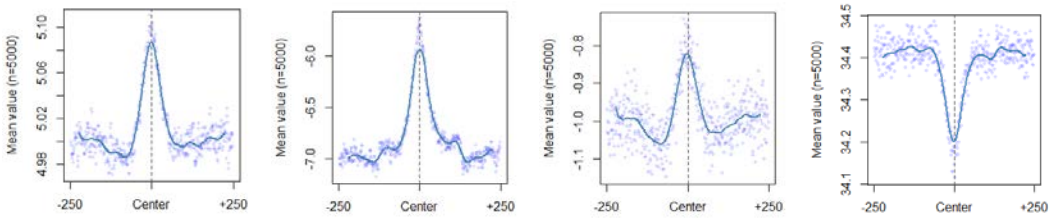


(b) Installed as shown in question (3).

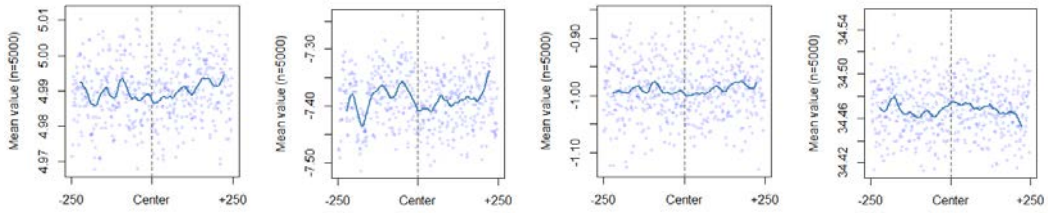
(7) High-throughput in vivo data analysis

(a)

Bound



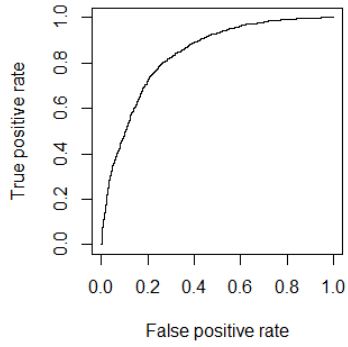
Unbound



(b) Bound data shows high symmetry, where unbound data does not.

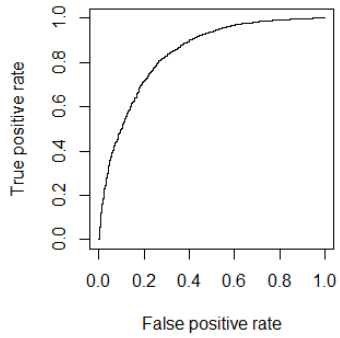
(8) Build prediction models for in vitro data

(a)
ROC 1mer+shape



AUC score 0.8386508

ROC 1mer



AUC score 0.8421386

(b) 1mer feature has slightly larger AUC score than 1mer+shape feature, which means 1mer is a better model for the in vitro data bound_30.fa and unbound_30.fa.