Integrative Molecular and Cellular Bioengineering, SPRING '07
16:125:584:01

Assignment 2: Due February 13, 2007

1. In an accompanying Excel file is the data from the Baker et al. paper.
a) By performing a search on "fibronectin", find the profiles of all genes encoding fibronectin. Comment on how similar or dissimilar they are.

There are 6 hits on fibronectin, summarized in the file "baker_fibonectin_hots.xls". As can be seen in the graph of their profiles, some of them have similar shapes but overall they are not so similar. Even when they represent 2 probes to the same clone (see series 2 and series 5), the results do not match as closely as we would like.

b) Examine the expression of the "immediate-early genes" in Table 6c of the paper. Specifically, compute the Euclidean distance between each pair of these 13 gene profiles (use Excel of Matlab or whatever you are comfortable with to somewhat automate this task) using the fold-change numbers in the paper. For the three genes that are replicated (c-myc, ERK-1, and c-fos), determine whether their closest distance is to their replicate or to another gene.

See "hw2_baker_immed_early_analysis.xls". In general, the replicates are similar, but probably not as similar as you would expect or help for. With c-fos, the expression changes are small, so the differences are amplified by the negative number notation utilized by Baler et al. But even with the others, the repeats are usually closer to another gene than to its replicate.

c) Compute the log2 changes in expression for these same 13 genes (think about what the negatives mean and how to deal with them) and then repeat the calculations in b), and see whether the interpretation regarding replicates changes.

Unfortunately, this did not change the interpretation markedly, though it was a useful exercise. See the same spreadsheet.

2. Using the Primer3 site (or some other if you prefer), design a set of real-time PCR primers that will detect mouse STAT3 cDNA while not detecting human STAT3 cDNA.

Searching for mouse STAT3 mRNA complete, there were 3 possibilities that turned up. I chose BC003806 because it appeared to be full-length (the second hit was not). When I blast this (BL2SEQ) against this human stat3 (variant 1), I get a lot of similarity. With respect to the mouse sequence, the homologous regions include [128..2507], [2548..2678], and [2804..2837]. So, I ran Primer 3 excluding the region from [128..2678] from consideration, and I got the following. I would not pick the top choice to use, since runs of 4 Gs are not good. Instead, I would use alternate oligos #2.

WARNING: Numbers in input sequence were deleted.

```
PRIMER PICKING RESULTS FOR mouse stat3

No mispriming library specified
Using 1-based sequence positions
WARNING: Unrecognized base in input sequence

OLIGO            start  len      tm     gc%   any    3' seq
```

```
LEFT PRIMER        2793   20   59.45   55.00  2.00   2.00 ctcggggttggttgttagac
RIGHT PRIMER       2895   20   59.60   55.00  4.00   2.00 aagtgcagagccaggagttc
SEQUENCE SIZE: 2972
INCLUDED REGION SIZE: 2972

PRODUCT SIZE: 103, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00
EXCLUDED REGIONS (start, len)*: 128,2550

    1 ORIGINcccacgcgtccgcgctgaggtacaaccccgctcggtgtcgcctgaccgcgtcgg


   61 ctaggagaggccaggcggccctcgggagcccagcagctcgcgcctggagtcagcgcaggc


  121 cggccagtcgggcctcagccccggagacagtcgagacccctgactgcagcaggatggctc
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  181 agtggaaccagctgcagcagctggacacacgctacctggagcagctgcaccagctgtaca
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  241 gcgacagcttccccatggagctgcggcagttcctggcaccttggattgagagtcaagact
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  301 gggcatatgcagccagcaaagagtcacatgccacgttggtgtttcataatctcttgggtg
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  361 aaattgaccagcaatatagccgattcctgcaagagtccaatgtcctctatcagcacaacc
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  421 ttcgaagaatcaagcagtttctgcagagcaggtatcttgagaagccaatggaaattgccc
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  481 ggatcgtggcccgatgcctgtgggaagagtctcgcctcctccagacggcagccacggcag
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  541 cccagcaaggggccaggccaaccacccaacagccgccgtagtgacagagaagcagcaga
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  601 tgttggagcagcatcttcaggatgtccggaagcgagtgcaggatctagaacagaaaatga
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  661 aggtggtggagaacctccaggacgactttgatttcaactacaaaaccctcaagagccaag
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  721 gagacatgcaggatctgaatggaaacaaccagtctgtgaccagacagaagatgcagcagc
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  781 tggaacagatgctcacagccctggaccagatgcggagaagcattgtgagtgagctggcgg
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  841 ggctcttgtcagcaatggagtacgtgcagaagacactgactgatgaagagctggctgact
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  901 ggaagaggcggcagcagatcgcgtgcatcggaggccctcccaacatctgcctggaccgtc
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

```
 961 tggaaaactggataacttcattagcagaatctcaacttcagacccgccaacaaattaaga
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1021 aactggaggagctgcagcagaaagtgtcctacaagggcgaccctatcgtgcagcaccggc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1081 ccatgctggaggagaggatcgtggagctgttcagaaacttaatgaagagtgccttcgtgg
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1141 tggagcggcagccctgcatgcccatgcacccggaccggcccttagtcatcaagactggtg
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1201 tccagtttaccacgaaagtcaggttgctggtcaaatttcctgagttgaattatcagctta
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1261 aaattaaagtgtgcattgataaagactctggggatgttgctgccctcagagggtctcgga
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1321 aatttaacattctgggcacgaacacaaaagtgatgaacatggaggagtctaacaacggca
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1381 gcctgtctgcagagttcaagcacctgacccttagggagcagagatgtgggaatggaggcc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1441 gtgccaattgtgatgcctccttgatcgtgactgaggagctgcacctgatcaccttcgaga
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1501 ctgaggtgtaccaccaaggcctcaagattgacctagagacccactccttgccagttgtgg
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1561 tgatctccaacatctgtcagatgccaaatgcttgggcatcaatcctgtggtataacatgc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1621 tgaccaataaccccaagaacgtgaacttcttcactaagccgccaattggaacctgggacc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1681 aagtggccgaggtgctcagctggcagttctcgtccaccaccaagcgggggctgagcatcg
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1741 agcagctgacaacgctggctgagaagctcctagggcctggtgtgaactactcagggtgtc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1801 agatcacatgggctaaattctgcaaagaaaacatggctggcaagggcttctccttctggg
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1861 tctggctagacaatatcatcgaccttgtgaaaaagtatatcttggccctttggaatgaag
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1921 ggtacatcatgggtttcatcagcaaggagcgggagcggccatcctaagcacaaagcccc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

1981 cgggcaccttcctactgcgcttcagcgagagcagcaaagaaggaggggtcactttcactt
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

2041 gggtggaaaaggacatcagtggcaagacccagatccagtctgtagagccatacaccaagc
     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

```
 2101 agcagctgaacaacatgtcatttgctgaaatcatcatgggctataagatcatggatgcga
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2161 ccaacatcctggtgtctccacttgtctacctctaccccgacattcccaaggaggaggcat
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2221 ttggaaagtactgtaggcccgagagccaggagcaccccgaagccgacccaggtagtgctg
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2281 ccccgtacctgaagaccaagttcatctgtgtgacaccaacgacctgcagcaataccattg
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2341 acctgccgatgtccccccgcactttagattcattgatgcagtttggaaataacggtgaag
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2401 gtgctgagccctcagcaggagggcagtttgagtcgctcacgtttgacatggatctgacct
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2461 cggagtgtgctacctcccccatgtgaggagctgaaaccagaagctgcagagacgtgactt
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2521 gagacacctgccccgtgctccacccctaagcagccgaaccccatatcgtctgaaactcct
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2581 aactttgtggttccagatttttttttttaatttcctacttctgctatctttgggcaatct
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2641 gggcacttttaaaatagagaaatgagtgagtgtgggtgataaactgttatgtaaagagg
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

 2701 agagcacctctgagtctggggatggggctgagagcagaagggagcaaggggaacacctcc


 2761 tgtcctgcccgcctgccctcctttttcagcagctcggggttggttgttagacaagtgcct
                                      >>>>>>>>>>>>>>>>>>>>

 2821 cctggtgcccatggcatcctgttgccccactctgtgagctgataccccaggctgggaact
                                                             <<<<<

 2881 cctggctctgcactttcaaccttgctaatatccacatagaagctaggactaagcccagag
      <<<<<<<<<<<<<<<

 2941 gttcctctttaaattaaaaaaaaaaaaaaa//
```

```
KEYS (in order of precedence):
XXXXXX excluded region
>>>>>> left primer
<<<<<< right primer
```

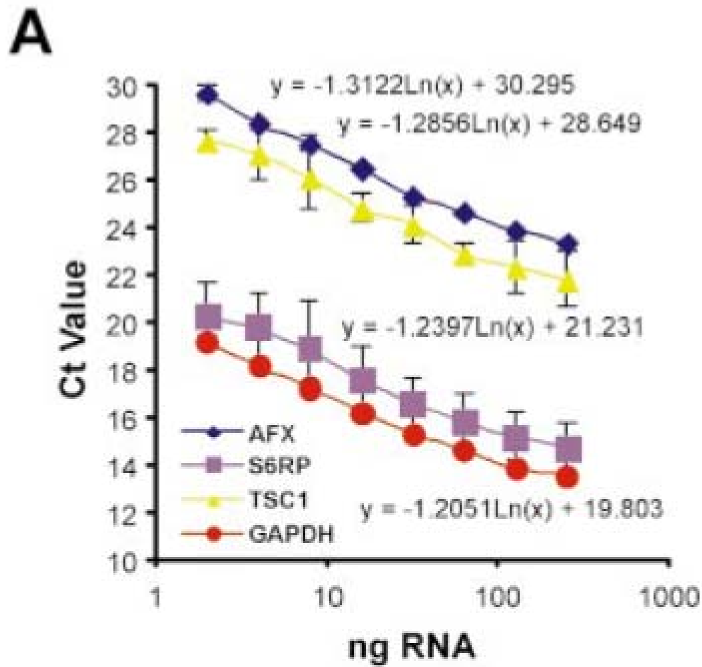
ADDITIONAL OLIGOS

	start	len	tm	gc%	any	3'	seq
1 LEFT PRIMER	2793	20	59.45	55.00	2.00	2.00	

ctcggggttggttgttagac

```
   RIGHT PRIMER       2897   20   59.60   55.00  4.00  1.00
gaaagtgcagagccaggagt
   PRODUCT SIZE: 105, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

 2 LEFT PRIMER        2729   20   60.81   55.00  2.00  0.00
tgagagcagaagggagcaag
   RIGHT PRIMER       2897   20   59.60   55.00  4.00  1.00
gaaagtgcagagccaggagt
   PRODUCT SIZE: 169, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00

 3 LEFT PRIMER        2728   20   60.81   55.00  3.00  0.00
ctgagagcagaagggagcaa
   RIGHT PRIMER       2897   20   59.60   55.00  4.00  1.00
gaaagtgcagagccaggagt
   PRODUCT SIZE: 170, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

 4 LEFT PRIMER        2729   20   60.81   55.00  2.00  0.00
tgagagcagaagggagcaag
   RIGHT PRIMER       2895   20   59.60   55.00  4.00  2.00
aagtgcagagccaggagttc
   PRODUCT SIZE: 167, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

Statistics
         con   too    in    in          no    tm    tm  high  high
high
         sid  many   tar  excl   bad    GC   too   too   any    3'  poly
end
        ered    Ns   get   reg   GC% clamp   low  high compl compl     X
stab    ok
Left    4287    11     0  2572  1271     0   118   173     0     2     0
23   117
Right   4269     7     0  2572  1042     0   230   190     0     1     0
30   197
Pair Stats:
considered 16, unacceptable product size 7, high end compl 1, ok 8
primer3 release 1.0


(primer3_www_results.cgi v 0.4)
```

3. The following plot gives real-time PCR data in terms of threshold cycle number for dilutions of each of 4 genes. For each of the four genes in this figure, estimate the PCR efficiency, averaged over all the runs in the dilution series.



Again, we can replace the log by ln in our equation from the notes:

$$n_t = \frac{\ln\left(F_t/\alpha\right)}{\ln(1+\varepsilon)} - \frac{1}{\ln(1+\varepsilon)}\ln(N_0)$$

Thus, each slope is equal to $-1/\ln(1+\varepsilon)$

Thus,

Gene	-slope	ε
AFX	1.3122	1.14
S6RP	1.2397	1.24
TSC1	1.2856	1.18
GAPDH	1.2051	1.29

The fact that all the values are significantly greater than 1 suggests that there is some flaw in the way that the values were generated.