## Sequence Analysis

This program is developed for the DNA Resource Core sequencing users to identify their sequencing samples provided that these samples are human or mouse genes. Users can directly upload the sequencing results to the program and blast the RefSeq human or mouse database to find the best matched genes for each sample. For samples that encode genes for other species, or if you want to blast your own sequence library, please contact us.

## Input

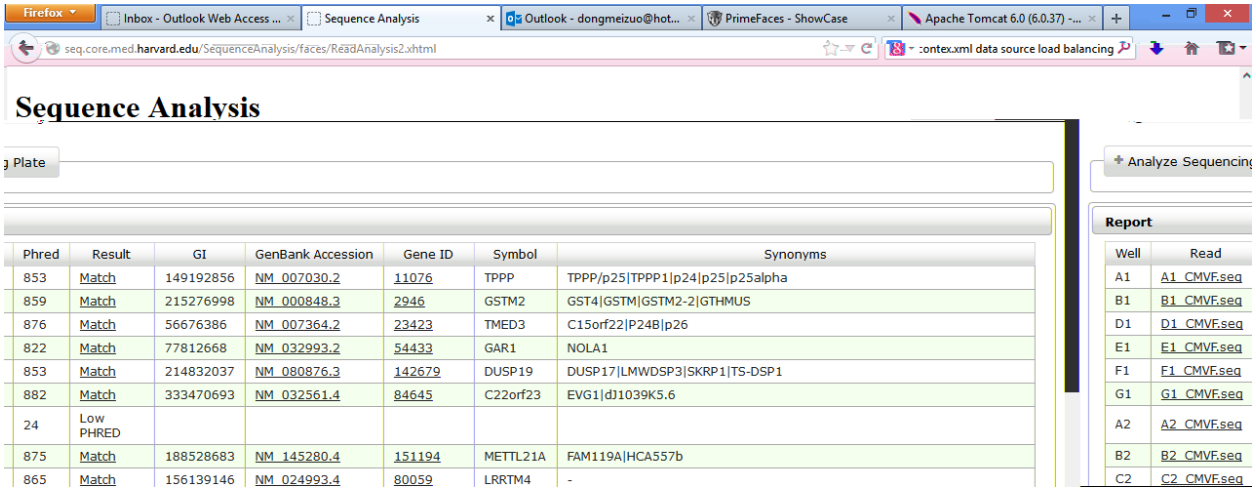
1. Sequence trim on the left: Some users want to trim off the unwanted sequences on the 5' end due to the low quality or being vector sequence. Default to 185nt.
2. Minimum alignment length: The minimum length of the longest blast alignment that would allow a match to be called. Default to 100nt.
3. Minimum percent identity: The minimum percent identity of the longest blast alignment that would allow a match to be called. Default to 90%.
4. Low phred cutoff: The minimum phred score that would allow blast to be performed. Default to 350.
5. High phred cutoff: The minimum phred score that would allow no match to be called. If the sample falls between the low phred cutoff and high phred cutoff, and blast finds no match, the final result will be manual analysis. Default to 650.
6. Sequence database: The database to be blasted. We currently support Human and mouse RefSeq database. Users can choose to blast the entire sequence or the CDS region. This database was last updated in April 2013. It will be updated annually. If you would like to blast other public database or your own databases, please contact us.

On the next page, please upload the sequencing files. The files to be uploaded must be in the same directory. To upload the files, click "Choose" button to browse to the directory that contains the sequencing files that you would like to upload. Once the files are listed, click "Upload" button to finish the upload. Then you can click "Analyze" button to start the analysis. Following is the example of the sequencing file. The first number after the sample name is the phred score.

```
>19321-1_A1-CMVF_A1 851 1449 34 904
NNNGGTCGGTGGCTATACAACTCGTTAGTGACCGTCAGATCGCCTGGAGA
CGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGCGGCCGCAGAT
CTCTCGAGGTTGATCTAATACGACTCACTATAGGGAGACCCAAGCTGGCT
AGTTAAGCTATCAACAAGTTTGTACAAAAAAGTTGGCATGAAGACCCCAT
TCGGAAAGACACCTGGCCAGCGGTCCAGAGCTGATGCAGGCCATGCTGGA
GTATCTGCCAACATGATGAAGAAGAGGACATCCCACAAAAAACATCGGAG
CAGTGTGGGTCCGAGCAAACCTGTTTCCCAGCCCCGGCGGAACATCGTAG
GCTGCAGGATTCAGCATGGGTGGAAAGAGGGGAATGGCCCTGTTACCCAG
TGGAAAGGAACCGTTCTGGACCAGGTGCCTGTAAATCCTTCTTTGTATCT
TATAAAATACGATGGATTTGACTGTGTTTATGGACTAGAACTTAATAAAG
ATGAAAGAGTTTCTGCGCTTGAAGTCCTCCCTGATAGAGTTGCGACATCT
CGAATCAGCGATGCACACTTGGCAGACACAATGATTGGCAAAGCAGTGGA
ACATATGTTTGAGACAGAGGATGGTTCTAAAGATGAGTGGAGGGGAATGG
TCTTAGCACGTGCACCTGTCATGAACACATGGTTTTACATTACCTATGAG
AAAGACCCTGTCTTGTACATGTACCAACTCTTAGATGATTACAAAGAAGG
```

CGACCTTCGCATTATGCCTGATTCCAATGATTCACCTCCAGCAGAAAGGG
AACCAGGAGAAGTTGTGGACAGCCTGGTAGGCAAACAAGTGGAATATGCT
CAAGAAGATGGCTCGAAAAGGACTGGCATGGTCATTCATCAAGTAGAAGC

Result




| Phred | Result | GI | GenBank Accession | Gene ID | Symbol | Synonyms |
|---|---|---|---|---|---|---|
| 853 | Match | 149192856 | NM_007030.2 | 11076 | TPPP | TPPP/p25\|TPPP1\|p24\|p25\|p25alpha |
| 859 | Match | 215276998 | NM_000848.3 | 2946 | GSTM2 | GST4\|GSTM\|GSTM2-2\|GTHMUS |
| 876 | Match | 56676386 | NM_007364.2 | 23423 | TMED3 | C15orf22\|P24B\|p26 |
| 822 | Match | 77812668 | NM_032993.2 | 54433 | GAR1 | NOLA1 |
| 853 | Match | 214832037 | NM_080876.3 | 142679 | DUSP19 | DUSP17\|LMWDSP3\|SKRP1\|TS-DSP1 |
| 882 | Match | 333470693 | NM_032561.4 | 84645 | C22orf23 | EVG1\|dJ1039K5.6 |
| 24 | Low PHRED | | | | | |
| 875 | Match | 188528683 | NM_145280.4 | 151194 | METTL21A | FAM119A\|HCA557b |
| 865 | Match | 156139146 | NM_024993.4 | 80059 | LRRTM4 | - |

| Well | Read |
|---|---|
| A1 | A1_CMVF.seq |
| B1 | B1_CMVF.seq |
| D1 | D1_CMVF.seq |
| E1 | E1_CMVF.seq |
| F1 | F1_CMVF.seq |
| G1 | G1_CMVF.seq |
| A2 | A2_CMVF.seq |
| B2 | B2_CMVF.seq |
| C2 | C2_CMVF.seq |

The result is displayed in table format. It has the following information:

1. Well: Well location of each sample on the plate.
2. Read: The name of the read. It links to the read sequence.
3. Phred: The phred score of each read.
4. Result: The final result of the blast. It links to the blast alignment. There is no blast performed for read with low phred score (score below the low phred cutoff).
5. GI: The GI number of the matched RefSeq sequence.
6. GenBank Accession: The GenBank Accession of the matched RefSeq sequence. It links to GenBank.
7. Gene ID: The Gene ID of the matched gene. It links to Gene database.
8. Symbol: The official symbol of the matched gene.
9. Synonyms: All the other synonyms of the matched gene.