



## Insilico Study of Human Lung Cancer Oncogenes and their Proteins using Different Software Tools

G. Anitha Mary

Asst. Prof, Research scholar (Ph.D. Part - time)  
Dept. of MCA, Loyola Academy Degree & PG College  
(Sri Venkateswara University, Tirupati, A.P, India)  
Old Alwal, Secunderabad-10, Telangana, India

G. Anjan Babu

Professor & Head  
Dept. of Computer Science  
Sri Venkateswara University,  
Tirupati, Chittoor(Dt), A.P, India

**Abstract:** Early stage lung cancer detection is the first step toward successful clinical therapy, the methods outlined and studied in the present research gives a solution to find the basic reasons for causing lung cancer. Clinicians analyze cancer progression by extracting tumor cell proteins in the blood plasma of the patients. Lung cancer is the leading cause of cancer death in the world. The patient survival rates for 5 years and 10 years remain very low at 14% and 8%, respectively. Though we have many advanced methods like diagnostic imaging and therapy, improvements over the past decade have shown a considerable good results. We report here an approach to study human genome lung cancer oncogenic proteins by Insilico methods.

The present work aims to study the proteins of oncogenes of lung cancer patients of the human genome. This is the first attempt made to analyze the genes of lung cancer. The NCBI oncogenic protein transcripts were obtained. They were analyzed by BLASTP method, the dendograms are compared and interpreted. There are 7 oncogenic proteins of Human genome identified and analyzed by us using BLASTP. The interpreted results show that they have maximum similarity with other oncogenic proteins of human cancers and other organisms.

**Keywords:** BLASTP; Cancer; Lungs; Oncogenes; Proteins

### I. INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide. The 5- and 10-year patient survival rates remain very low at 14 and 8%, respectively [1], despite diagnostic imaging and therapy improvements over the past decade. The best therapeutic opportunity for surgical resection is in the early stage patients, but disease diagnosis is usually late, and prognosis is accordingly poor [2]. Detection of lung cancer at an early disease stage is critical for successful clinical therapy, an improved prognosis, and increased survival rate.

Scientists have been seeking tumor-specific biomarkers for decades. Neuron-specific enolase, carcinoembryonic antigens, cytokeratin 19 fragments, and some other proteins are the most commonly used lung cancer markers [3]. However, few of these markers are useful in a routine clinical setting, thus underscoring the need for new clinically relevant sources [4]. Liotta et al. [5] state "it is time to rethink our approach to tumor biomarker discovery. The blood contains a treasure trove of previously unstudied biomarkers that could reflect the ongoing physiological state of all tissues" [5] including tumor tissues. Protein expression level meta-analysis of patients' blood serum holds promise for diagnostic tumor signature detection.

Each year millions of blood samples are collected for medical diagnosis, forming a vast clinical data resource. However, blood plasma is difficult to analyze because it comprises large amounts of albumin and a wide dynamic range of other heterogenic proteins [6]. The human plasma proteome provides a tremendous opportunity for detecting, therapeutic monitoring, and deciphering basic cancer mechanisms. Mass spectrometry is a powerful high throughput analytical approach to cancer identification that

yields comprehensive peptide and protein serum profiles. Mass spectrometry is efficient because it does not require large samples or arduous protein separations [7], [8]. However, its clinical effectiveness has garnered doubt from the scientific community [8, 9]. In the current study the purpose is to find the reason of causing lung cancer and the summarized results show the accuracy information.

### II. METHODOLOGY

#### A. NCBI:

- 1) In NCBI home page <https://www.ncbi.nlm.nih.gov/> select Protein in All databases
  - 2) Search for Homo Sapiens Oncogenes of Lung cancer
  - 3) Several results were obtained
  - 4) Select only Oncogenes of lung cancer 1,2,3,4,5,6 and 7
  - 5) Open the files and select the Protein sequences
- Copy each protein sequences in the BLASTP program

#### B. BLAST P:

- 1) Go to <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and choose one of the "Basic BLAST" search options:
- 2) Select "Blast 2 sequences" option and "Enter Subject Sequence" section appears:
- 3) Enter query and subject sequences and select "BLAST":
- 4) View standard BLAST report. "Dot Matrix view" is available if there was only one subject sequence

#### C. CLUSTAL W:

CLUSTALW uses the progressive algorithm, by adding the sequence one by one until all the sequences are completely aligned.

Steps for CLUSTAL algorithm

- 1) Calculate all possible pairwise alignments, record the score for each pair.
- 2) Calculate a guide tree based on the pairwise distances (algorithm: Neighbor Joining).
- 3) Find the two most closely related sequences
- 4) Align the sequences by progressive method
  - i. Calculate a consensus of this alignment
  - ii. Replace the two sequences with the consensus
  - iii. Find the two next-most closely related sequences (one of these could be a previously determined consensus sequence).
  - iv. Iterate until all sequences have been aligned
- 5) Expand the consensus sequences with the (gapped) original sequences
- 6) Report the multiple sequence alignment

**D. T-COFFEE:**

(Tree based Consistency Objective Function For alignment Evaluation) is a multiple sequence alignment software using a progressive approach. It generates a library of pairwise alignments to guide the multiple sequence alignment. It can also combine multiple sequences alignments obtained previously and in the latest versions can use structural information from PDB files (3D-Coffee). It has advanced features to evaluate the quality of the alignments and some capacity for identifying occurrence of motifs (Mocca). It produces alignment in the aln format (Clustal) by default, but can also produce PIR, MSF, and FASTA format. The most common input formats are supported (FASTA, PIR).

**E. MAFT:**

Multiple alignment program for amino acid or nucleotide sequences. It offers a range of multiple alignment methods, L-INS-I (accurate; for alignment of <~200 sequences), FFT-NS-2 (fast; for alignment of <~30,000 sequences), etc.

Usage:

```
% mafft [arguments] input > output
```

An alias for an accurate option (L-INS-I) for an alignment of up to ~200 sequences x ~2,000 sites:

```
% mafft-linsi input > output
```

A fast option (FFT-NS-2) for a larger sequence alignment:

```
% mafft input > output
```

If not sure which option to use,

```
% mafft --auto input > output
```

**F. MAFT-EBI:**

MAFFT (Multiple Alignment using Fast Fourier Transform) is a high speed multiple sequence alignment program.

MAFFT is a multiple sequence alignment program for Unix-like operating systems. It offers a range of multiple alignment methods, L-INS-I (accurate; for alignment of

<~200 sequences), FFT-NS-2 (fast; for alignment of <~30,000 sequences), etc.

**G. MAFT-CBRC:**

Multiple alignment program for amino acid or nucleotide sequences

MAFFT is a multiple sequence alignment program for unix-like operating systems. It offers a range of multiple alignment methods, L-INS-i (accurate; for alignment of <~200 sequences), FFT-NS-2 (fast; for alignment of <~30,000 sequences), etc.

**III. RESULTS**

**A. NCBI RESULTTS**

The seven Oncogenes of lung cancer are as follows



Figure 1: lung cancer oncogene 1 [Homo sapiens]. It can be accessed through <https://www.ncbi.nlm.nih.gov/protein/AAM21294.1>



Figure 2: Lung cancer oncogene 2 [Homo sapiens] It can be accessed through <https://www.ncbi.nlm.nih.gov/protein/20385456>



Figure 3: Lung cancer oncogene 3 [Homo sapiens]. It can be accessed through <https://www.ncbi.nlm.nih.gov/protein/AAN12269.1>



Figure 4: Lung cancer oncogene 4 [Homo sapiens]  
It can be accessed through  
<https://www.ncbi.nlm.nih.gov/protein/AAN12270.1>



Figure 5: Lung cancer oncogene 5 [Homo sapiens]  
It can be accessed through  
<https://www.ncbi.nlm.nih.gov/protein/AAN12271.1>



Figure 6: Lung cancer oncogene 6 [Homo sapiens]  
It can be accessed through  
<https://www.ncbi.nlm.nih.gov/protein/33332354>

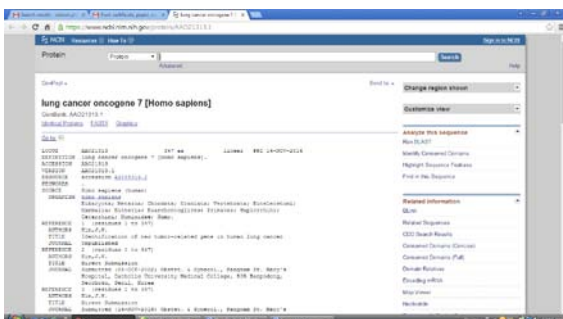


Figure 7: Lung cancer oncogene 7 [Homo sapiens]  
It can be accessed through  
<https://www.ncbi.nlm.nih.gov/protein/AAO21313.1>

**1. The Sequences of the above proteins are given as follows.**

The FASTA format of the sequences are shown.

a) lung cancer oncogene 1 [Homo sapiens]  
>AAM21294.1 lung cancer oncogene 1 [Homo sapiens]  
MIARRNPEPLRFLPDEARSLPPPKLTDPRLLYIGFL  
GYCSGLIDNVIRRRPIATAGLHRQLLYITAFFAGY

YLVKREDYLYAVRDREMFGYMKLHPEDFPEEDK  
KTYGEIFEKHFPIR

b) Lung cancer oncogene 2 [Homo sapiens]  
>AAM21295.1 HLC-2 [Homo sapiens]  
MIARRNPEPLRFLPDEARSLPPPKLTDPRLLYIGFL  
GYCSGLIDNVIRRRPIATAGLHRQLLYITAFFFCWI  
LCKT

c) Lung cancer oncogene 3 [Homo sapiens]  
>AAN12269.1 lung cancer oncogene 3 [Homo sapiens]  
MLGEDSDEEEEMDTSERKINAGSQDDEMGC  
WGMGEDAVEDDAEENPIVLEFQQEREAFYIKDPK  
KALQGFDDREGEELEYEFDEQGHSTWLCRVRLPV  
DDSTGKQLVAEAIHSGKKKEAMIQCSLEACRILD  
LGLIRQEA VSRKRKAKNWEDEDFYDSDDDTFLDR  
TGLIEKRLNRMKKAGKIDEPETFESLVAKLND  
ERELSEISERLKASSQVLSSEPSQDSLDAFMSEMKS  
GSTLDGVSRRKLLHRTFELRKEQRLKGLIKIVKP  
AEIPELKKTTETQTTGAENKAKKLTPLPLFGAMKGG  
KFKLKTGTVGLPPELPPPTLMRMKDEPEVEE  
EEEEEEEEKEKEHEKKLEDGSLSRPQPEIEPEAA  
VQEMRPPTDLTHFKETQTHENMSQLSEEEQNKDY  
QDCSKTTS LCAGPSASKNEYEKSRGELKKKTPGP  
GKLPPTLSSKYPEDDPDYCVVWPPEGQSGDGRTH  
LNDKYG

d) Lung cancer oncogene 4 [Homo sapiens]  
>AAN12270.1 lung cancer oncogene 4 [Homo sapiens]  
MLGTGKLGFSFVRITALMVSCNRLWVGTGNGVIIS  
IPLTETVILHQGRLLGLRANKTSGVPGNRPGSVIRV  
YGDENSDK VTPGTFIPYCSMAHAQLCFHGHDRDAV  
KFFVAVPGQVISPQSSSGTDLTGDKAGPSAQEPGS  
QTPLKSMLVISGGEGYIDFRMGDEGGESELLGEDL  
PLEPSVTKAERSHLIVWQVMYGN

e) Lung cancer oncogene 4 [Homo sapiens]  
>AAN12271.1 lung cancer oncogene 5 [Homo sapiens]  
MLLTRFEQKNHGLAQVEVDEVFQFMCHITTEVPP  
HDAMPGGIVLLVKFLLDMSRNVLLYVFLQRLSS  
ALHRVLLHLFRHVRIFDHGLSVAHGYRGGGWPT  
ATVSWGRC

f) Lung cancer oncogene 6 [Homo sapiens]  
>AAQ11383.1 HLC-6 [Homo sapiens]  
MTSNSSAETDSSLGGITVVGCSAEGVTGAATSP  
STNGASPVMKPPMEMEAENSEVDENVPTAEEATE  
ATEGNAGSAEDTVDISQTVGYTEHVFTDPLGVQIP  
EDLSPVYQSSNDSDAYKQDQISVLPNEQDLVREEAQ  
KMSSLLPTMWLGAQNGCLYVHSSVAQWRKCLHS  
IKLKDSILSIVHVKGIVLVALADGTLAIFHRGVDCQ  
WDLSNYHLLDLGRPHHSIRCMTVVHDKVWCGYR  
NKIYVVPKAMKIEKSFDAHPRKESQVRQLAWVG  
DGVVWSIRLDSTLRLYHAHTYQHLQVDIEPYVS  
KMLGTGKLGFSFVRITALMVSCNRLWVGTGNGVI  
ISIPLTETVILHQGRLLGLRANKTSGVPGNRPGSVIR  
VYGDENSDK VTPGTFIPYCSMAHAQLCFHGHDRDA  
VKFFVAVPGQVISPQSSSGTDLTGDKAGPSAQEP  
GSQTPLKSMLVISGGEGYIDFRMGDEGGESELLGE  
DLPLEPSVTKAERSHLIVWQVMYGN

g) Lung cancer oncogene 7 [Homo sapiens]  
>AAO21313.1 lung cancer oncogene 7 [Homo sapiens]

MGRRGCGASFSKPSAILVAAATHALAAAMTEQ  
 MTLRGTGKGNHGWVTQIATTPQFPDMILSARDK  
 TIIMWKLTRDETNYGIPQRALRGHSHFVSDVVISSD  
 GQFALSGSWDGLRLWDLTTGTTTRRFVGHTKDV  
 LSVAFSSDNQIVSGSRDKTIKLWNTLGVCKYTVQ  
 DESHSEWVSCVRFSPNSSNPIIVSCGWDKLVKVVN  
 LANCKLKTNHIGHTGYLNTVTVSPDGLCASGGK  
 DGQAMLWDLNEGKHLTYLDGGDIINALCFSPNRY  
 WLCAATGPSIKIWDLEGKIIIVDELKQEVISTSSKAE  
 PPQCTSLAWSADGQTLFAGYTDNLVRVWQVTIGT  
 R

**B. BLASTP RESULTS**

Then the sequences are selected to perform the BLASTP program.



Figure 8 : lung cancer oncogene 1 [Homo sapiens]



Figure 9: Lung cancer oncogene 2 [Homo sapiens]



Figure 10: Lung cancer oncogene 3 [Homo sapiens]

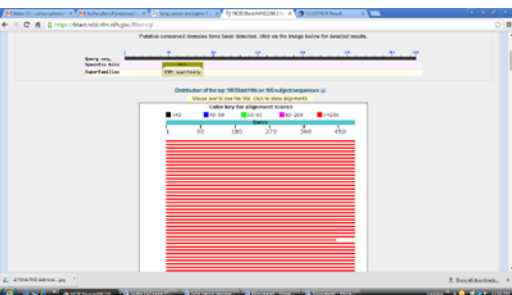


Figure 11: Lung cancer oncogene 4 [Homo sapiens]

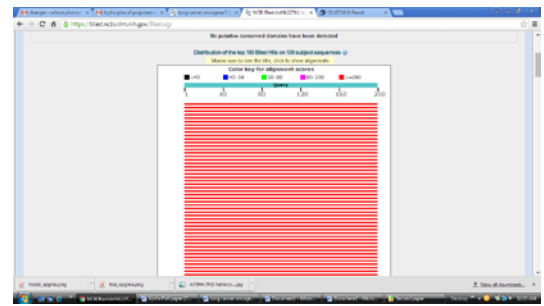


Figure 12: Lung cancer oncogene 5 [Homo sapiens]



Figure 13: Lung cancer oncogene 6 [Homo sapiens]



Figure 14: Lung cancer oncogene 7 [Homo sapiens]

**C. CLUSTAL W RESULTS**

The 7 sequences were performed with Multiple sequence alignment to study their interrelationship along with their phylogenetic trees.

**1. CLUSTAL A.2.a Multiple Sequence Alignments**

Sequence type explicitly set to Protein  
 Sequence format is Pearson

- Sequence 1: AAM21294.1 119 aa
- Sequence 2: AAM21295.1 77 aa
- Sequence 3: AAN12269.1 490 aa
- Sequence 4: AAN12270.1 200 aa
- Sequence 5: AAN12271.1 110 aa
- Sequence 6: AAQ11383.1 508 aa
- Sequence 7: AAO21313.1 347 aa

Start of Pairwise alignments

Aligning...

- Sequences (1:2) Aligned. Score: 90.9091
- Sequences (1:3) Aligned. Score: 13.4454
- Sequences (1:4) Aligned. Score: 14.2857
- Sequences (1:5) Aligned. Score: 10.9091
- Sequences (1:6) Aligned. Score: 15.9664
- Sequences (1:7) Aligned. Score: 10.9244
- Sequences (2:3) Aligned. Score: 14.2857
- Sequences (2:4) Aligned. Score: 14.2857
- Sequences (2:5) Aligned. Score: 15.5844
- Sequences (2:6) Aligned. Score: 15.5844
- Sequences (2:7) Aligned. Score: 12.987
- Sequences (3:4) Aligned. Score: 12

Sequences (3:5) Aligned. Score: 14.5455  
 Sequences (3:6) Aligned. Score: 9.59184  
 Sequences (3:7) Aligned. Score: 10.0865  
 Sequences (4:5) Aligned. Score: 12.7273  
 Sequences (4:6) Aligned. Score: 100  
 Sequences (4:7) Aligned. Score: 12  
 Sequences (5:6) Aligned. Score: 19.0909  
 Sequences (5:7) Aligned. Score: 16.3636  
 Sequences (6:7) Aligned. Score: 13.5447  
 Guide tree file created: [\[clustalw.dnd\]](#)

There are 6 groups  
 Start of Multiple Alignments

2. Aligning...

Group 1: Delayed  
 Group 2: Sequences: 2 Score: 3304  
 Group 3: Delayed  
 Group 4: Sequences: 2 Score: 1203  
 Group 5: Delayed  
 Group 6: Sequences: 4 Score: 707  
 Alignment Score -416

CLUSTAL-Alignment file created [\[clustalw.aln\]](#)

3. [Clustalw.aln](#)

CLUSTAL A.2.a multiple sequence alignment

```

AAN12270.1 -----
-----
AAQ11383.1
MTSNSSAETDSSLGGITVVGCSAEGVTGAATSPST
NGASPVMDKPPMEMAENSEVDENV
AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAO21313.1 -----
-----
AAN12269.1 -----
MLGEDSDEEEEMDTSERKINAGSQDDEMGC

AAN12270.1 -----
-----
AAQ11383.1
TAEATEATEGNAGSAEDTVDISQTGVYTEHVFTD
PLGVQIPEDLSPVYQSSNSDAYKD
AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAO21313.1 -----
MGRRGCGASFSKPSAILVAAATHALAAAMTEQ
MTRLRGTGKHNGWVT
AAN12269.1
TWGMGEDAVEDDAEENPIVLEFQQEREAIFYIKDP
KKALQGFFDREGEELEYEFDEQGHST
    
```

```

AAN12270.1 -----
-----
AAQ11383.1
QISVLPNEQDLVREEAQKMSSLLPTMWLGAQNGC
LYVHSSVAQWRKCLHSIKLKDSILSI
AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAO21313.1
QIATTPQFPDMILSASRDKTIMWKLTRDETNYGIP
QRALRGHSHFVSDVVISSDQGFAL
AAN12269.1
WLCRVRLPVDDSTGKQLVAEAIHSGKKKEAMIQC
SLEACRILDTLGLIRQEAVSRKRKAK

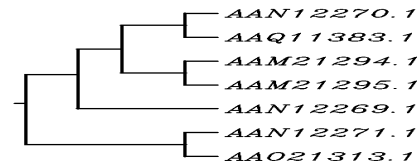
AAN12270.1 -----
-----
AAQ11383.1
VHVKGIVLVALADGTLAIFHRGVDGQWDLSNYHL
LDLGRPHHSIRCMTVVHDKVWCGYRN
AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAO21313.1
SGSWDGTLLRLWDLTTGTTTRRFVG-----
-----
AAN12269.1
NWEDEDYFSDDDTFLDRTGLIEKKRLNRMKKAG
KIDEKPEFESLVAKLNDARELSEI

AAN12270.1 -----
-----
AAQ11383.1
KIYVVQPKAMKIEKSFDAHPRKESQVRQLAWVGD
GVVVSIRLDSTLRLYHAHTYQHLQDV
AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAO21313.1 -----
HTKDVLSVAFSSDNRQIVSGS
AAN12269.1
SERLKASSQVLSESPSQDSLDAFMSEMKSGSTLDG
VSRKKLHLRTFELRKEQQRLKGLIK

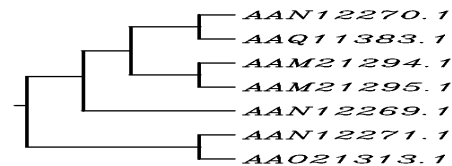
AAN12270.1 -----
MLGTGKLGFSFVRITALMVSCNRLWVGTGNGVVIS
IP----LTETVILHQG
AAQ11383.1
DIEPYVSKMLGTGKLGFSFVRITALMVSCNRLWV
GTGNGVVISIP----LTETVILHQG
AAM21294.1 -----
-----
    
```

AAM21295.1 -----  
 -----  
 AAN12271.1 -----  
 -----  
 AAO21313.1  
 RDKTIKLWNTLGVCKYTVQDESHSEWVSCVRFSP  
 NSSNPIIVSCG-----WDKLVKVVNL  
 AAN12269.1  
 IVKPAEIPKKTETQTTGAENKAKKLTLPFGAM  
 KGGSKFKLKTGTVGKLPKRPELPP  
  
 AAN12270.1  
 RLLGLRANKTSGVPGNRPGSVIRVYGDENS DKVTP  
 GTFIPYCSMAHAQLCFHGHRDAVKF  
 AAQ11383.1  
 RLLGLRANKTSGVPGNRPGSVIRVYGDENS DKVTP  
 GTFIPYCSMAHAQLCFHGHRDAVKF  
 AAM21294.1 --MIARRNPE-----  
 PLRFLPDEARSLPPPCLTDPRLLYIG  
 AAM21295.1 --MIARRNPE-----  
 PLRFLPDEARSLPPPCLTDPRLLYIG  
 AAN12271.1 -MLLTRFEQKNG-----  
 HLAQVEVDEVFGFMCHITTEVPPHDAMPG  
 AAO21313.1 ANCKLKTNHIGHTG-----  
 YLNTVTVSPDGLCASGGKDGQAMLWDLNEG  
 AAN12269.1  
 TLMRMKDEPEVEEEEEEEEEEEKEKEEHEKKKLE  
 DGSLSRPQPEIEPEAAVQEMRPPTDL  
  
 AAN12270.1  
 FVAVPGQVISPQSSSGTDLTGDKAGPSAQEPGSQ  
 TPLKSMVISGGEGYIDFRMGDEGG  
 AAQ11383.1  
 FVAVPGQVISPQSSSGTDLTGDKAGPSAQEPGSQ  
 TPLKSMVISGGEGYIDFRMGDEGG  
 AAM21294.1  
 FLGYCSGLIDNVIRRRPIATAGLHR---  
 QLLYITAFFFAGYYLVKREDYLYAVRDREMFQ  
 AAM21295.1  
 FLGYCSGLIDNVIRRRPIATAGLHR---  
 QLLYITAFFFCWILSCKT-----  
 AAN12271.1  
 GIVLLVKFLLDMSRNVLLYVFLQRLSSALHRVLL  
 HLFRRHVRIFDHGLSVAHG YRGE GGGW  
 AAO21313.1  
 KHL YTL DGGDIINALCFSPNRYWLCAATGPSIKIW  
 DLEGKIIVDELKQEVISTSSKA EPP  
 AAN12269.1  
 THFKETQTHENMSQLSEEEQNKD YQDCSKT TSLC  
 AGPSASKNEYEKSRGELKKKKT PGP G  
  
 AAN12270.1  
 ESELLGEDLPLEPSVTKAERSHLIVWQVMYGNE----  
 ---  
 AAQ11383.1  
 ESELLGEDLPLEPSVTKAERSHLIVWQVMYGNE----  
 ---  
 AAM21294.1  
 YMKLHPEDFPEEDKKT YGEIFEK FHP IR-----  
 AAM21295.1 -----  
 AAN12271.1 PTATVSWGRC-----  
 --

AAO21313.1  
 QCTSLAWSADGQTLFAGYTDNLVVRVWQVTIGTR--  
 ----  
 AAN12269.1  
 KLPPTLSSKYPEDDPDYCVWVWVPEPQSGDGRTHL  
 NDKYGY



4. Rooted phylogenetic Tree (UPGMA) Method


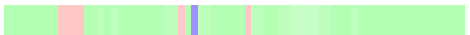
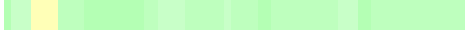



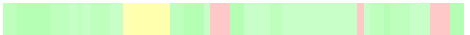
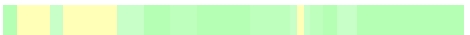

5. Rooted phylogenetic Tree with branch length (UPGMA) Method

**D. T-COFFEE RESULTS**


PSI/TM-Coffee alignment result  
 MSAThe multiple sequence alignment result as produced by T-coffee.  
 T-COFFEE, Version\_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)  
 Cedric Notredame  
 SCORE=537  
 \*  
 BAD AVG GOOD  
 \*  
 AAM21294.1 : 67  
 AAM21295.1 : 67  
 AAN12269.1 : 55  
 AAN12270.1 : 67  
 AAN12271.1 : 60  
 AAQ11383.1 : 55  
 AAO21313.1 : 55  
 cons : 53

AAM21294.1 MIARRNPEPLRFLPDEARSLPPP-----  
 -----  
 AAM21295.1 MIARRNPEPLRFLPDEARSLPPP-----  
 -----  
 AAN12269.1 MLGEDSDEEEEMDTSER-  
 KINAGSQD-----  
 DEMGCTWGMGEDA---  
 AAN12270.1 MLGTGKLG-----  
 -----  
 AAN12271.1 MLLTRFEQKNGHLA QV-----  
 -----  
 AAQ11383.1 MTSNSSAETDSLGGIT-  
 VVGCSAEGVTGAATSPSTNGASPVM D K P P E M E A E  
 NSEVDENVPTAEEATEA  
 AAO21313.1 MGRRRGCGASFSPSSAILVAAATH

ALAAAMT-----  
 cons \*   
 AAM21294.1 -----  
 AAM21295.1 -----  
 AAN12269.1 -----  
 VEDDAEENPIVLEFQQEREA  
 AAN12270.1 -----  
 AAN12271.1 -----  
 AAQ11383.1 TEGNAGSAEDTVDISQTGVYTEHVF  
 TDPLGVQIPEDLSPVYQSSNDSDAYK-----  
 DQISVLPN---  
 AAO21313.1 -----  
 -----  
 cons   
 AAM21294.1 -KLTDPRL-----  
 AAM21295.1 -KLTDPRL-----  
 AAN12269.1 FYIKDPKKALQGFFDREGEELEYEFD  
 EQGHSTWLCRVR--LPVDDSTGKQLVA--  
 EAIHSGKKKEAMIQ  
 AAN12270.1 -----  
 -----  
 AAN12271.1 -----EVDEV--  
 FGFMCCHITTEVPPHDAMPGGIVLLVKFLLDMSR---  
 --  
 AAQ11383.1 -----EQDLVREEA-  
 QKMSSLLPTMWLGAQN--GCLYVHSSVAQWR--  
 KCLHSIKLKDSILS  
 AAO21313.1 -----  
 EQMTLRGTLKGHNGWVTQIA--TTP-----QFP--  
 DMILSASRDKTIM  
 -----  
 cons   
 AAM21294.1 -----  
 AAM21295.1 -----  
 AAN12269.1 CS-----  
 LEACRILDTLGLIRQEAVSRKRKAKNWEDEDFYDS  
 DDD-----  
 AAN12270.1 -----  
 -----  
 AAN12271.1 -----  
 -----  
 AAQ11383.1 IV-----HVKGIVLVALAD-  
 GTLAIFHRGVDGQWDLSNYHLLDLGRPHHSIRCM  
 T  
 AAO21313.1 WKLTRDETNYGIPQRALRGHSHFVS  
 DVV-ISSD-GQFAL----SGSWDG-  
 TLRWLWDLTTGTTTRRFVG  
 -----  
 cons 

AAM21294.1 -----  
 -----LYI-----  
 AAM21295.1 -----  
 -----LYI-----  
 AAN12269.1 -----TFLDRTGLIEKK-----  
 RLNRMKKAGKIDEKPFESLVAKLNDAERELS  
 AAN12270.1 -----  
 -----  
 AAN12271.1 -----  
 -----NVLLEYV-----  
 AAQ11383.1 VVHDK-----  
 VWCGRNKIYVVQPK---AMK---IEKSFDAHPRK-  
 ESQVRQLAWV-----  
 AAO21313.1 HTKDVLSVAFSSDNRQIVSGSRDKTI  
 K--LWNTLGVCK---YTVQDESHSEW-  
 VSCVRFSPNS-----  
 -----  
 cons   
 AAM21294.1 --GFLGYCSGLIDNVIRRRPI-----  
 -----ATA-----  
 AAM21295.1 --GFLGYCSGLIDNVIRRRPI-----  
 -----ATA-----  
 AAN12269.1 EIS-----ERLKASSQVLSE-----  
 -----SPSQDSLDAFMSEMKS GST  
 AAN12270.1 -----  
 -----  
 AAN12271.1 --VFLQRLS-----  
 -----S-----  
 AAQ11383.1 --GDGVVWSIRLDSTLRLYHAHTYQ--  
 -----HLQDV DIEPYVSKMLGTG-----  
 AAO21313.1 --  
 SNPIIVSCGWDKLVKVNLANCKLKTNHIGHTGY  
 LNTVTVSPDGSLCASGG-----  
 -----  
 cons   
 AAM21294.1 --GLHRQL-----YITAF--FF--  
 -----  
 AAM21295.1 --GLHRQL-----YITAF--FF--  
 -----  
 AAN12269.1 LDGVSRRKLL-----HLRTF--EL-  
 -----  
 AAN12270.1 ---FSFV-----RITAL--  
 MVSCNRLWVGTGNGVVIISIPLTETVI-  
 LHQGRLLGLR  
 AAN12271.1 --ALHRVLL-----HLFRHVRIF-  
 -----D-----  
 AAQ11383.1 --KLGFSFV-----RITAL--  
 MVSCNRLWVGTGNGVVIISIPLTETVI-  
 LHQGRLLGLR  
 AAO21313.1 --  
 KDQAMLWDLNEGKHLTYLDGGDIINAL--  
 CFSPNRYWLCAATGPSIKIWDLEGKIIVDELKQE--  
 -----  
 cons   
 AAM21294.1 -----AGYYLV-  
 KREDYLYAVRDREM-----  
 AAM21295.1 -----CWILSC-KT-  
 -----  
 AAN12269.1 -----  
 RKEQQRKGLIKIV--K-----PAE


AAN12270.1 ANKTSVPGNRPGSVIRVYGDENSD  
 KVTPGTFIPYCSMAHAQLCFHGHRDAVKFFVAVP  
 GQVISPQSSS  
 AAN12271.1 -----  
 HGLSVAHG YRGEG-----  
 AAQ11383.1 ANKTSVPGNRPGSVIRVYGDENSD  
 KVTPGTFIPYCSMAHAQLCFHGHRDAVKFFVAVP  
 GQVISPQSSS  
 AAO21313.1 -----  
 VISTSSKAEPQCTSLAWSADGQTLFAGYTD-----  
 -----

cons 

AAM21294.1 -----F-----GY-----  
 -----MKLHPEDFPEEDKKT  
 AAM21295.1 -----  
 -----  
 AAN12269.1 IPELKKTETQTTGAENKAKKLTLPF  
 GAMKGGSKFKLKTGTVGKLPKRPELPPTLMRMK  
 DEPEVEEEE  
 AAN12270.1 SGTDLTGDKAGPSAQEPGSQTPLKS  
 MLVISGGEGY-----IDFRMGDEGGESELL  
 AAN12271.1 -----GWPT-----  
 -----  
 AAQ11383.1 SGTDLTGDKAGPSAQEPGSQTPLKS  
 MLVISGGEGY-----IDFRMGDEGGESELL  
 AAO21313.1 -----  
 -----


cons 

AAM21294.1 -----YGEIFE-----  
 -----  
 AAM21295.1 -----  
 -----  
 AAN12269.1 EEEEEEEKEKEEHEKKLEDGSLSRP  
 QPEIEPEAAVQEMRPPTDLTHFKETQTHENMSQLS  
 EEEQNKDY  
 AAN12270.1 GEDL----PLEPSVTKAE-----  
 -----  
 AAN12271.1 -----  
 -----  
 AAQ11383.1 GEDL----PLEPSVTKAE-----  
 -----  
 AAO21313.1 -----  
 -----

cons 

AAM21294.1 -----KFH-----  
 -----PIR-----  
 AAM21295.1 -----  
 -----  
 AAN12269.1 QDCSKTTS LCAGPSASKNEYEKSRG  
 ELKKKKTGPGKLPPTLSSKYPEDDPDYCVWVPE  
 GQSGDGRTH  
 AAN12270.1 -----RSH-----  
 -----LIVWQVMYGNE-----  
 AAN12271.1 -----A-----  
 -----TVSWGRC-----  
 AAQ11383.1 -----RSH-----  
 -----LIVWQVMYGNE-----

AAO21313.1 -----NL-----  
 -----VRVWQVTIGTR-----

cons 

AAM21294.1 -----  
 AAM21295.1 -----  
 AAN12269.1 LNDKYG  
 AAN12270.1 -----  
 AAN12271.1 -----  
 AAQ11383.1 -----  
 AAO21313.1 -----

cons 

**E. MAFT-EBI RESULTS**

>AAM21294.1 lung cancer oncogene 1 [Homo sapiens]  
 -----MIAR-----  
 -----  
 -----RNPEPLRF-----  
 -----LPDEARSL-----  
 ---PPPKL-TD---PRLLYIGFL-GYC-----  
 SGLIDNV--  
 -----IRRRPIATAG-----  
 LHRQLLYITAFFFAGYLVKR-----  
 -----EDY-LYAVRDREMF-----  
 -----  
 -----GYM-----KLHPEDFP-----EEDKKTYGE-----  
 ---IFEKHFPIR  
 >AAM21295.1 HLC-2 [Homo sapiens]  
 -----MIAR-----  
 -----  
 -----RNPEPLRF-----  
 -----LPDEARSL-----  
 ---PPPKL-TD---PRLLYIGFL-GYC-----  
 SGLIDNV--  
 -----IRRRPIATAG-----  
 LHRQLLYITAFFFCWILSCKT-----  
 -----  
 -----  
 >AAN12271.1 lung cancer oncogene 5 [Homo sapiens]  
 M-----LLTR-----  
 -----FEQ-----  
 -----  
 -----KNGHLAQV-----  
 -----EVDEVFGF-----  
 MCHI  
 TTEVPPHDA-MPG--GIVLLVKFLLDMS-----  
 -----RNVLLYV--  
 -----VFLQRLSSA-----  
 LHRVLLHLFRHVRIF-----DHG  
 LLSVAHG-----YRGEFG-----  
 -----  
 -----WPTA-----  
 -----TVSWGRC-----  
 >AAN12269.1 lung cancer oncogene 3 [Homo sapiens]  
 -----  
 MLGEDSDEEEEM

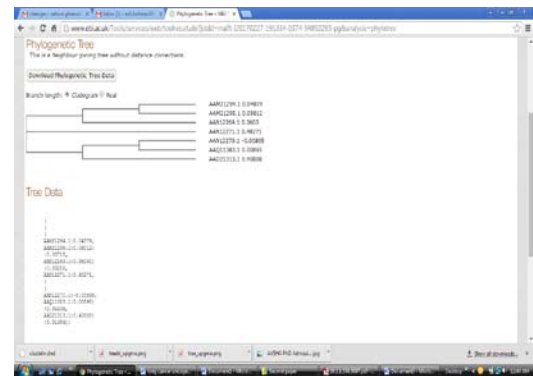


DTSERKI----NAGSQDDEMGCT-  
 WGMGEDAVEDD---AEENPIVLEFQQEREAFYIK  
 DPKKALQGFFDREGEELEYEFDEQGHSTWL-----  
 CRVRLPVDDSTGKQLVAEAIHSGKK  
 KEAMIQCS----  
 LEACRILDTLGLIRQEA VSRKRKAKNWEDEDFYDS  
 DDDTFLDRTGLIE  
 KKRLNRMKKAGKIDEKP----  
 ETFESLVAKLNDAERELSEISE-----RLKAS---SQV  
 LSESPSQDS-  
 LDAFMSEMKSGSTLDG VSRKKLHLRTFELRKEQQ  
 RL----KGLIKIV--  
 -  
 KP AEIPELKKTETQTTGAENKAKKLTLPFGAMKG  
 GSKFKLKTGTGKLPKRPPELPT  
 LMRMKDEPEVEEEEEEEEEEEKEKEEHEKK---  
 KLEDGSLSRPQPEIEPEAAVQEMRPPT  
 DLTHFKETQTHENMSQLSEEEQNKDYQDCSKTTS  
 LC---AGPSASKNEYEKS--RGELKK  
 K--KTPGPGKLPPTL-----  
 SSKYPEDDPDYCVWVPPEGQSGDGRTHL----  
 NDKYGY-

>AAN12270.1 lung cancer oncogene 4 [Homo sapiens]

-----MLGTGKL-GFS---  
 FVRITALMVSCNRLWVGTGNGVSI--  
 -----PLTETVI-----  
 LHQGRLLGLRANKTSGVPGNR---PGS  
 VIRVYGD-----ENS--DKVTPGTFI---  
 PYCSMAHAQLCFHGHR  
 DAVKFFVAVPGQVISP-----  
 QSSSGTDLTGDKAGPSA---QEPGS--QTPLKS  
 MLVISGGEYIDFRMGDEGGESELLGEDLP----  
 LEPSVTKAE-RSHLI-VWQVMYGN  
 >AAQ11383.1 HLC-6 [Homo sapiens]  
 MTSNSSAETDSLGGITVVGCSAEGVTGAATSPST  
 NGASPVMKPPPE-MEAENSEVDENV  
 PTAEATEATEGNAGSAEDTVDISQTGVYTEHVFT  
 DPLGVQIPEDLSPVYQSSNDSDAYK  
 DQISVLPNEQDLVREEAQ-  
 KMSSLLPTMWLGAQNGC--LYVHSSVAQW--  
 RKCLHSIKL  
 KDSILSIVHVKGIVLVALADGTLAIFHRGVDGQWD  
 LSNYHLLDL-GRPHHSIRCMTVVHD  
 KVVCGYRNKIYVVPKAMKIEKSFDA-  
 HPRKESQVRQLAWVDG VVWSIRLDSTLRLYHA  
 HTYQHLQDV DIEPYVSKMLGTGKL-GFS---  
 FVRITALMVSCNRLWVGTGNGVSI--  
 -----PLTETVI-----  
 LHQGRLLGLRANKTSGVPGNR---PGS  
 VIRVYGD-----ENS--DKVTPGTFI---  
 PYCSMAHAQLCFHGHR  
 DAVKFFVAVPGQVISP-----  
 QSSSGTDLTGDKAGPSA---QEPGS--QTPLKS  
 MLVISGGEYIDFRMGDEGGESELLGEDLP----  
 LEPSVTKAE-RSHLI-VWQVMYGN  
 >AAO21313.1 lung cancer oncogene 7 [Homo sapiens]  
 MGGRR-----GC-----GAS--  
 FSKPSSAILVAAATHALAA

AMTEQMT-----LRGTLKGHNGWV-  
 TQIATTPQFPDMILSASRDKTIIMWKLTRDETNYG  
 IPQRALRG-----  
 HSHFV  
 SDVVISS-----DG----QFALSGSWD-  
 GTLRLWDL-----TTGTTT  
 RRFVGH TK-----DVLSVAFSSDNRQI-----  
 -----  
 --VSGSRD-----KTIKLWNTL-  
 GVCKYTVQDESHSEWVSCVRFSPNSSNPIIVSCGW  
 DKLVKVVNLANCKLKT-----  
 NHIGHTGYLNTVTVS-----PDG  
 SLCASGG-----  
 KDGQAMLWDLNEGKHLTYLDGGDII-----  
 NALCFSPNR  
 YW-----LCA-  
 ATGPSIKIWDLEGKIIVDELKQ  
 EVISTSSKAEPQCT-----SLAWSADGQ-----  
 TLFAGY-TDNLVRVWQVTIGTR



Figure

15: Phylogenetic Tree-MAFT-EBI

**F. MAFT-CBRC RESULTS**

**MAFFT-L-INS-i Result**

CLUSTAL format alignment by MAFFT (v7.309)

```

AAM21294.1  MIARR-----
-----NPEP-----
AAM21295.1  MIARR-----
-----NPEP-----
AAN12271.1      MLLTRFEQKNGH-LAQVEV----
DEVFGFMCHITT-----EVPPHDAMPGGIV----
AAN12269.1
MLGEDSDEEEEMDTSERKINAGSQDDEMGCTWG
MGE-----DAVEDDAEENPIV----
AAO21313.1      MGRR-----
GCGASF SK-----PSSAI-----
AAN12270.1  M-----

AAQ11383.1      MTSNSSAETDSL-
LGGITVVGCSAEGVTGAATSPSTNGASPVMKPPPE
MEAENSEVDENV
*
AAM21294.1      -----
LRFLPDEARSLPPPCLTDPRL
AAM21295.1      -----
LRFLPDEARSLPPPCLTDPRL
    
```

```

AAN12271.1 -----
LLVKFLLDMSRNV-----L

AAN12269.1 -----
LEFQQEREAFY---IKDPKK
AAO21313.1 -----
LVAAATHALAAAM-----TEQM
AAN12270.1 -----
-----
AAQ11383.1
PTAEEATEATEGNAGSAEDTVDISQTGVYTE
HVFTDPLGVQIPEDLSPVYQSSNDSDAYK

AAM21294.1  LYIGFL-----
-----
AAM21295.1  LYIGFL-----
-----
AAN12271.1  LYVVFL-----
-----
AAN12269.1
ALQGFFDREGEELEYEFDEQGHSTWLCRVRL
PVD-----DSTGKQLVAEAIHSGKKKEA
AAO21313.1          TLRGTL-----
KGHNGWVTQI-----ATTPQF-
PDMILSASRDKT
AAN12270.1 -----
-----
AAQ11383.1          DQISVLPNEQDLVREE--
AQKMSSLLPTMWLGAQNGCLYVHSSVAQW-
RKCLHSIKLKDS

AAM21294.1 -----
-----
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAN12269.1  MIQCSLEACRILDTLGLIRQE--
----AVSRKRKAKNWE---DEFYDSDD-----
AAO21313.1  IIM-----
WKLTRDETNYG-----
AAN12270.1 -----
-----
AAQ11383.1          ILS-----
IVHVKGIVLVALADGTLAIFHRGVDGQW----
DLSNYHLLDLGRPHHSIR

AAM21294.1 -----
-----GYCS---
AAM21295.1 -----
-----GYCS---
AAN12271.1 -----
-----QRLS---
AAN12269.1          -----DT---
FLDRTGLIEKKRLNRMKKAGKIDEKETFESL
VAKLNDAERELSEISE
AAO21313.1          -----IPQRALRGHSH-
-----FVSDVVISSDGQFALS---
    
```

```

AAN12270.1 -----
AAQ11383.1
CMTVVHDKVWCGYRNKIYVVQPKAMKIEKS
---FDAHPR-KESQV-----RQLAWVGD

AAM21294.1 -----GLIDNVIRR-----
-----RPIATA-----GLHR
AAM21295.1 -----GLIDNVIRR-----
-----RPIATA-----GLHR
AAN12271.1 -----
-----S-----ALHR
AAN12269.1
RLKASSQVLSESPSQDSLDAFMS-----
----EMKSGST---LDGVSr
AAO21313.1 -----GSWDGTLRL-----
-----WDLTTGTTTRRFVGHTK
AAN12270.1 -----
---LGTG-----KLGf
AAQ11383.1          GVVVS-----
IRLDSTLRLYHAHTYQHLQDVDIEPYVSKML
GTG-----KLGf

AAM21294.1  QLLYITAF-----
-----
AAM21295.1  QLLYITAF-----
-----
AAN12271.1  VLLHL-----
-----
AAN12269.1          KKLHLRTFELRKEQQRl--
KGLI-----KIVKPAEIPeLkKtET-----
AAO21313.1  DVLSV-AFSSDNrQ-----
----IVSGSRDKTIKLWNTLGV-----CK
AAN12270.1
SFVRITAlMVScNRLWVGtGNGViiSIPLtET
VilHQGRllGLRANKtSGVpGNrPGSVI
AAQ11383.1
SFVRITAlMVScNRLWVGtGNGViiSIPLtET
VilHQGRllGLRANKtSGVpGNrPGSVI
        ::

AAM21294.1          -----FFAGYY-----
-----LVKR-----
AAM21295.1          -----FFCWIL-----
-----SCKT-----
AAN12271.1 -----
-----
AAN12269.1
QTTGAENKAKKLTLPLFGAMKGG-----
----SKFKLKTGTVGKLPPKRp
AAO21313.1
YTVQDESHSEWVScVrFSPNSSNPIIVSc--
GWDKLVKVWNLANCKLkTNHIGHTG----
AAN12270.1          RvYGDEN-
SDKVTPGTfIPYCSMAHAQLCFHG-----
HRDAVKFFVAVPGQV----
AAQ11383.1          RvYGDEN-
SDKVTPGTfIPYCSMAHAQLCFHG-----
HRDAVKFFVAVPGQV----
    
```

```

AAM21294.1 -----
-----EDYLYAVR-
AAM21295.1 -----
-----
AAN12271.1 -----
-----
AAN12269.1
ELPPTLMRMKDEPEVEEEEEEEEEEEKEKEEH
EKKKLEDGSLSRPQPEIEPEAAVQEMRP
AAO21313.1 ---YLNTVTVSP-----
----DGSLCASGGK-DGQAMLW----
AAN12270.1 -----ISPQS-----
---SSSGTDLTGDKAGPSAQE
AAQ11383.1 -----ISPQS-----
---SSSGTDLTGDKAGPSAQE

AAM21294.1 --DREMFGYMKLHPE-----
-----DFPEEDKKT
AAM21295.1 -----
-----
AAN12271.1 ----FRHVRI-----
-----FDHGLSVA
AAN12269.1 PTDLTHFKETQTHEN--
MSQLSEEEQNKDYQDCSKTTS--LCA--GPS--
ASKNEYEKS
AAO21313.1 --
DLNEGKHLTYLDGGDIINAL-----
CFSPNRYWLCAATGPSIKIWDLEGKII
AAN12270.1 PGSQTPLKSMVISG-----
-----GEGYIDFRMGDEGG
AAQ11383.1 PGSQTPLKSMVISG-----
-----GEGYIDFRMGDEGG

AAM21294.1 YGEIFEKFHPIR-----
-----
AAM21295.1 -----
-----
AAN12271.1 HGYRGECCWPTAT-----
---VSW-----
AAN12269.1 RGELKKKTPGPGKL-
PPTLSSKYPEDDPDYCVWVPEGQSGDGRTH
LND-----K
AAO21313.1
VDELKQEVISTSSKAEPQCTS-----LAW----
-SADGQTLFAGYTDNLVRVWQ
AAN12270.1 ESELLGEDLPLE----PSVTK--
AERSHLIVW-----Q
AAQ11383.1 ESELLGEDLPLE----PSVTK--
AERSHLIVW-----Q

AAM21294.1 -----
AAM21295.1 -----
AAN12271.1 --GRC
AAN12269.1 YGY--
AAO21313.1 VTIGTR
AAN12270.1 VMYGNE
AAQ11383.1 VMYGNE
    
```

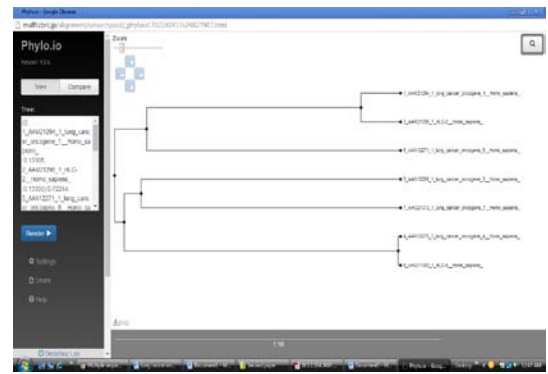


Figure 16: Phylogenetic Tree-MAFT-CBRC

#### IV. CONCLUSION

In conclusion, the study of Human lung cancer oncogenes and their proteins using different software tools were analyzed. The gene sequences of the human genome responsible for causing lung cancer in humans which is most prevalent in many countries. The results shown here are useful for further analysis of these sequences for many researches to work on their protein expression studies both in genomics and proteomics. The data can also be utilized for development of many applications like app development in android, iOS and other operating systems. The results also interpret that the lung cancer gene analysis can also be used by oncologists besides bioinformaticians in many aspects.

#### V. REFERENCES

- [1] Chen, G., Gharib, T. G., Wang, H., Huang, C. C., Kuick, R., Thomas, D. G., Shedden, K. A., Misek, D. E., Taylor, J. M., Giordano, T. J., Kardias, S. L., Iannettoni, M. D., Yee, J., Hogg, P. J., Orringer, M. B., Hanash, S. M., and Beer, D. G. (2003) Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13537–13542.
- [2] Hoffman, P. C., Mauer, A. M., and Vokes, E. E. (2000) Lung cancer. *Lancet* 355, 479–485.
- [3] Stieber, P., Aronsson, A. C., Bialk, P., Kulpa, J., Lamerz, R., Molina, R., Molina, R., and van Dalen, H. (1999) Tumor markers in lung cancer: EGTM recommendations. *Anticancer Res.* 19, 2817–2819.
- [4] Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M., Ambartsumian, N., Sandelin, K., Rank, F., and Gromova, I. (2004) Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol. Cell. Proteomics* 3, 327–344.
- [5] Liotta, L. A., Ferrari, M., and Petricion, E. (2003) Clinical proteomics: written in blood. *Nature* 425, 905.
- [6] Anderson, N. L., and Anderson, N. G. (2002) the human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867.
- [7] Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S., Canto, M. I., Yeo, C. J., Chan, D. W., and Goggins, M. (2004) Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin. Cancer Res.* 10, 860–868.
- [8] Hanash, S. (2003) Disease proteomics. *Nature* 422, 226–232.
- [9] Check, E. (2004) Proteomics and cancer: running before we can walk? *Nature* 429, 496–497.