STUDY OF APOE GENE USING BIOINFORMATICS TOOLS

A Thesis Submitted

To

Orissa University of Agriculture and Technology, Bhubaneaswar

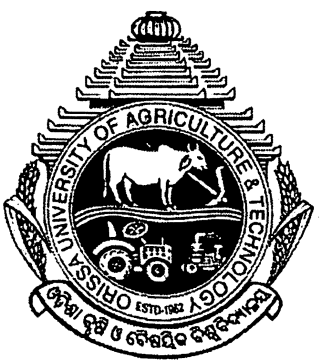
in partial fulfilment of the requirement of the degree of

MASTER OF SCIENCE IN BIOINFORMATICS

BY

ARPEETA ROUT

Adm.No-32BI/12



DEPARTMENT OF BIOINFORMATICS

CENTRE FOR POST GRADUATE STUDIES

ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

BHUBANESWAR-751003

2014

Name of the advisor — Mrs. Sushma Rani Martha

DEDICATED

TO

MY BELOVED PARENTS

National Institute of Mental Health and Neuroscience, Bangalore 560 029, INDIA
(*Deemed University*)

Dr. Meera Purushottam
Senior Scientific Officer
Molecular Genetics Lab
Department of Psychiatry

Fax: ** 91 (80) 2699 5262/3
Tel: ** 91 (80) 2656 4822
Email: meera.purushottam@gmail.com

CERTIFICATE

This is to certify that the dissertation entitled "STUDY OF APOE GENE USING BIOINFORMATICS TOOLS" submitted by ARPEETA ROUT, Reg.no.32/BI /12 to the Orissa University of Agriculture and Technology, for the degree of *Master of Science* in Bioinformatics is her original work, based on the results of the experiments and investigations carried out independently by her during the period (from FEBRUARY 2014 TO MAY 2014) of study under my guidance.

This is also to certify that the above said work has not previously submitted for the award of any degree, diploma, fellowship in any Indian or foreign University.

Meera Purushottam
Signature of the Guide

Place: Bangalore
Date: 30th May 2014

Name and Designation

CERTIFICATE-II

This is to certify that the thesis entitled **"STUDY OF APOE GENE USING BIOINFORMATICS TOOLS"** submitted by **Miss Arpeeta Rout (Adm. No. 32BI/12)** to Orissa University of Agriculture and Technology, Bhubaneswar in partial fulfillment of the requirements for the award of the **degree Master of Science in Bioinformatics** has been approved by the student advisory committee after an oral examination of the same in collaboration with external examiner.

ADVISORY COMMITTEE

1. **Mrs. Sushma Rani Martha**
   Assistant Professor
   Department of Bioinformatics — **Chairperson** .....

2. **Mr. Sukant Kumar Pradhan**
   H.O.D.
   Department of Bioinformatics — **Member** .....

3. **Mr. Abhimanyu Dash**
   H.O.D.
   Department of CSA — **Member** .....

External Examiner .....

ACKNOWLEDGEMENT

I am extremely happy to take this opportunity to acknowledge my debts of gratitude to them who are associated in the successful completion of this dissertation work as well as my M.Sc. course.

First I express my heartfelt thanks to my advisor ***Mrs. Sushma Rani Martha, Asst. Prof., Department of Bioinformatics, Orissa University of Agriculture and Technology, Bhubaneswar*** *for her cooperation.*

I express my sincere obligation to ***Mr. Sukanta Kumar Pradhan, H.O.D., Department of Bioinformatics, Orissa University of Agriculture and Technology*** *for providing an opportunity to complete my dissertation work at National Institute Of Mental Health And Neuroscience*

Words fail to express my profound regards from the inmost recess of my heart to my Guide, ***Dr. Meera Purushottam, Senior Scientific Officer, Molecular Genomic Lab, Department of Phychiatry, NIMHANS*** *for this premier guidance and suggestions throughout the period of my dissertation. He has been the driving force behind this project. I sincerely thank him giving and inspired me and supported to take this innovative project.*

With pleasure I would like to express my sincere thanks to Asst. Prof. ***Mr. Surya Narayan Rath, Asst. Prof. Mrs Sucharita Balabantray****of our department; I am obliged to them forever for their timely help and encouragement.*

A special thanks to my Parents and Family members for their kind support, love, and encouragement.

Arpeeta Rout

CONTENTS

LIST OF FIGURES

LIST OF TABLE

INTRODUCTION

CHAPTER -I

INTRODUCTION

Alzheimer's disease is a neurological disorder in which the death of brain cells causes memory loss and cognitive decline. The prevelance rate of ad in India is about 45%.it was first described by german psychiatrist and neuropathology's alois alzheimer in 1906 and was named after him.[(Flicker et al. 1984; Albert 1996]. There is no cure for the disease, which worsens as it progresses and eventually leads to death. Alzheimer's disease has two different sets of symptoms, which are the cognitive symptoms and the behavioral symptoms.The cognitive symptoms are memory loss, disorientation, confusion, and the difficulty with reasoned thought [Morris et al. 1989] The behavioral symptoms are agitation or anxiety, delusions or hallucinations, depression, insomnia, and wandering.[Storandt and Hill 1989]

Due to the biochemical changes many many structural changes happens in human brain which includes the accumulation of amyloid plaques between nerve cells (neurons) in the brain. Amyloid is a general term for protein fragments that the body produces normally. [Locascio et al. 1995] in a healthy brain, these protein fragments would break down and be eliminated. In Alzheimer's disease, the fragments accumulate to form hard, insoluble plaques.Secondly neurofibrillary tangles are insoluble twisted fibers found inside the brain's nerve cells. They primarily consist of a protein called tau, which forms part of a structure called a microtubule. [Storandt 1991; Welsh et al. 1991]The microtubule helps transport nutrients and other important substances from one part of the nerve cell to another. (Axons are long threadlike extensions that conduct nerve impulses away from the nerve cell; dendrites are short branched threadlike extensions that conduct nerve impulses towards the nerve cell body.)In alzheimer's disease the tau protein is abnormal and the microtubule structures collapse.[j. Biomed. Biotechnol. 2010]. Thirdly as alzheimer's disease progresses, brain tissue shrinks.[Cruchaga C, et al 2011] However, the ventricles, chambers within the brain that contain cerebrospinal fluid, are noticeably enlarged. In the early stages of Alzheimer's disease, short-term memory begins to decline when the cells in the hippocampusdegenerate. Those with the disease lose the ability to perform routine tasks.

There are two forms of AD are

Sporadic AD: Sporadic AD is the more common of the two forms, accounting for 90-95% of cases. The other known risk factor for developing this form of the disease is advancing age. The older you get, the better your chances of developing it. [Jack et al. 1997].

Familial autosomal dominant AD.Familial autosomal dominant AD is clearly passed from one generation to the next. In families who have the disease with an affected parent, each child has a fifty percent chance of developing it., [alberts mj. 1999]. There is a genetic test available for those people who have this form of AD, so those with a family history that can be traced over several generations, where family members who are affected show a similar age of onset and duration of the disease, can be tested. This form of the disease is fairly rare, however, accounting for only five to ten percent of cases of AD.[Singh PP, Singh M, Mastana SS 2009].

Apolipoprotein-e (apoe) is a class of apolipoprotein found in the chylomicron and intermediate-density lipoprotein (idls) that is essential for the normal catabolism of triglyceride-rich lipoprotein constituents.in peripheral tissues, apoe is primarily produced by the liver and macrophages and mediates cholesterol metabolism in an isoform-dependent manner.[Kok E,et al. 2008-06-13].In the central nervous system. Apoe is mainly produced byastrocytes. And transports cholesterol to neurons via apoe receptors, which are members of the low density lipoprotein receptor gene family. [De Toledo-Morrell et al. 2000].

The gene, *apoe*, is mapped to chromosome 19 a cluster with apolipoprotein c1 and the apolipoprotein c2. [Barthel H, 2008] the apoe gene consists of four exons and three introns, totaling 3597 base pairs. *Apoe* is transcriptionally activated by the liver x receptorand peroxisome proliferator-activated receptor γ, nuclear receptors that form heterodimens with retinoid x receptors· In melanocytic cells apoe gene expression may be regulated by mitf. [Michaelson dm, korczyn ad 2009].

OBJECTIVE

- Study of detailed molecular biology of APOE4 gene.
- Study of APOE gene variant.
- In silico anslysis of APOE.
- Analysis of APOE gene interaction with other gene.

REVIEW OF LITERATURE

CHAPTER-II

REVIEW OF LITERATURE

2.1 Biochemistry of Alzheimer's

The biochemistry of Alzheimer's disease (ad), one of the most common causes of adult dementia, is as yet not well understood. Ad has been identified as a protein misfolding disease due to the accumulation of abnormally folded amyloid beta protein in the brains of alzheimer's patients. [Clarke J, Hodges JR 2003]Amyloid beta, also written aβ, is a short peptide that is an abnormal proteolytic byproduct of the transmembrane protein amyloid precursor protein (app), whose function is unclear but thought to be involved in neuronal [singleton ab. 2012] development. The presenilins are components of proteolytic complex involved in app processing and degradation.amyloid beta monomers are soluble and contain short regions of beta sheet and polyproline ii helix secondary structures in solution,though they are largely alpha helical in membranes;however, at sufficiently high concentration, they undergo a dramatic conformational change to form a beta sheet rich tertiary structure that aggregates to form amyloid fibrils these fibrils deposit outside neurons in dense formations known as senile plaques or neuritic plaques, in less dense aggregates as diffuse plaques,[mahley rw april 1988]and sometimes in the walls of small blood vessels in the brain in a process called amyloid angiopathy.

AD is also considered a tauopathy due to abnormal aggregation of the tau protein , a microtubule-associated protein expressed in neurons that normally acts to stabilize microtubules in the cell cytoskeleton[Mckhann g, drachman d et al folstein m].

2.3 Cholinergic hypothesis

The oldest hypothesis is the "cholinergic hypothesis". It states that alzheimer's begins as a deficiency in the production of acetylcholine, a vital neurotransmitter.[Fryer TD, Clarke J 2003] Much early therapeutic research was based on this hypothesis, including restoration of the "cholinergic nuclei". The possibility of cell-replacement therapy was investigated on the basis of this hypothesis. „[Mirra ss, gearing m, mckeel dw, jr, et al1991]all of the first-generation anti-alzheimer's medications are based on this hypothesis and work to preserve acetylcholine by

inhibiting acetylcholinesterases (enzymes that break down acetylcholine). These medications, though sometimes beneficial, have not led to a cure. In all cases, they have served to only treat symptoms of the disease and have neither halted nor reversed it. , [Shoji m, golde t, ghiso j, et al 1995]these results and other research have led to the conclusion that acetylcholine deficiencies may not be directly causal, but are a result of widespread brain tissue damage, damage so widespread that cell-replacement therapies are likely to be impractical.[PMID15236795] more recently, cholinergic effects have been proposed as a potential causative agent for the formation of plaques and tanglesleading to generalized neuroinflammation.

More recent hypotheses center on the effects of the misfolded and aggregated proteins, amyloid beta and tau. The two positions are lightheartedly described as "ba-ptist" and "tau-ist" viewpoints in one scientific publication.[Cai h, wang y, mccarthy d, et al1995] therein, it is suggested that "tau-ists" believe that the tau protein abnormalities initiate the disease cascade, while "ba-ptists" believe that beta amyloid deposits are the causative factor in the disease.

2.4 Tau hypothesis

The hypothesis that tau is the primary causative factor has long been grounded in the observation that deposition of amyloid plaques does not correlate well with neuron loss.[Albert M, et al. 1998] A mechanism for neurotoxicity has been proposed based on the loss of microtubule-stabilizing tau protein that leads to the degradation of the cytoskeleton.[Schroeter eh, ilagan mx, brunkan al, et al1998] however, consensus has not been reached on whether tau hyperphosphorylation precedes or is caused by the formation of the abnormal helical filament aggregates [Irvine gb, el-agnaf om, shankar gm 1998].support for the tau hypothesis also derives from the existence of other diseases known as tauopathies in which the same protein is identifiably misfolded.however, a majority of researchers support the alternative hypothesis that amyloid is the primary causative agent.

2.5Amyloid hypothesis

The amyloid hypothesis is initially compelling because the gene for the amyloid beta precursor (APP) gene is located on chromosome 21, and patients with trisomy 21 better known as down syndrome - who thus have an extra gene copy almost universally exhibit ad-like disorders by 40 years of age.[Lahoz C, et al2001].The traditional formulation of the amyloid hypothesis points to

the cytotoxicity of mature aggregated amyloid fibrils, which are believed to be the toxic form of the protein responsible for disrupting the cell's calcium ion homeostasis and thus inducing apoptosis.[Kosik ks, bakalis sf, selkoe dj1986] this hypothesis is supported by the observation that higher levels of a variant of the beta amyloid protein known to form fibrils faster *in vitro* correlate with earlier onset and greater cognitive impairment in mouse models and with ad diagnosis in humans. However, mechanisms for the induced calcium influx, or proposals for alternative cytotoxic mechanisms, by mature fibrils are not obvious.[Wood jg, mirra ss19986].

2.6 Glucose consumption

The human brain is one of the most metabolically active organs in the body and metabolizes a large amount of glucose to produce cellular energy in the form of adenosine triphosphate (atp). [pmc 3375834]despite its high energy demands, the brain is relatively inflexible in its ability to utilize substrates for energy production and relies almost entirely on circulating glucose for its energy needs.this dependence on glucose puts the brain at risk if the supply of glucose is interrupted, or if its ability to metabolize glucose becomes defective. If the brain is not able to produce ATP, synapses cannot be maintained and cells cannot function, ultimately leading to impaired cognition. [Roses ad, saunders am, alberts ma, et al 1995]

2.7 Genetics of ad

Overall, more than 90% of patients with ad appear to be sporadic and to have a later age at onset of 60 to 65 years of age (load), many carriers of the apoe risk allele (ε4) live into their 90s, which suggests the existence of other load genetic and/or environmental risk factors that have yet to be identified.[. Coon kd, myers aj, craig dw, et al 2007]Unreplicated genetic variants have been reported, and these findings suggest that there may be 5 to 7 major load susceptibility genes.. [Kang j, lemaire hg, unterbeck a, et al 1999]

2.8 AD genes

Genes associated with autosomal dominant ad. [Pub medhardy j. 2001.]_Although several hundred families carry one of the following mutations, they account for less than 1% of cases.

2.9 AD1: amyloid precursor protein (APP)

Amyloid precursor protein is a type-i integral-membrane proteinthat resembles a signal-transduction receptor.[Serretti A, Artioli P, 2005] It is expressed in many tissues and concentrated in the synapses of neurons. Its primary function is not known, though it has been implicated in neural plasticityand as a regulator of synapse formation. [Mann dm, prinja d, davies ca, et al1989] amyloid precursor protein is synthesized in the endoplasmic reticulum, posttranscriptionally modified in the Golgi (n- and o-linked glycosylation, sulfation, and phosphorylation), and transported to the cell surface via the secretory pathway. Amyloid precursor protein is also endocytosed from the cell surface and processed in the endosomal–lysosomal pathway.[Sherrington r, froelich s, sorbi s, et al 1996].amyloid precursor protein and its by-product aβ have been found to be translocated inside mitochondria and implicated in mitochondrial dysfunction.

2.10 AD2: apoe gene function and expression

The mechanisms that govern apoe toxicity in brain tissue are not fully understood. Some proposed mechanisms include isoform-specific toxicity, *apoe* ε4-mediated amyloid aggregation, and *apoe* ε4-mediated tau hyperphosphorylation.[Raux g, guyant-marechal l, martin c, et al 2005] The *apoe* polymorphism is unique to humans and has been proposed to have evolved as a result of adaptive changes to diet. It is known that apoe plays an important role in the distribution and metabolism of cholesterol and triglycerides within many organs and cell types in the human body.individuals carrying *apoe* ε4 have higher total and ldl cholesterol. . [Tanzi re, haines jl, watkins pc, et al 1998] moreover, in vitro neurons have a cholesterol uptake that is lower when lipids are bound to apoe4 compared to apoe2 and apoe3, and apoe4 appears to be less efficient than the other isoforms in promoting cholesterol efflux from both neurons and astrocytes.

2.11 AD3: presenilin 1 (psen1)

Psen1 is a polytopic membrane protein that forms the catalytic core of the γ-secretase complex.γ-secretase is an integral membrane protein typically found at the cell surface, but it may also be found in the Golgi, endoplasmic reticulum, and mitochondria. . [Sastre m, steiner h, fuchs k, et al 2001].Psen1, nicastrin (nct), anterior pharynx defective 1 (aph-1), and presenilin enhancer 2 (psenen) are required for the stability and activity of the γ-secretase.

MATERIALS AND METHODS

CHAPTER-III

MATERIALS AND METHODS

3.1 UCSC GENOME BROWSER-

The ucsc genome browser is an on-line genome browser hosted by the University of California, Santa Cruz it is an interactive website offering access to genome sequence data from a variety of vertebrate and invertebrate species and major model organisms, integrated with a large collection of aligned annotations. The browser is a graphical viewer optimized to support fast interactive performance and is an open-source, web-based tool suite built on top of a mysql database for rapid visualization, examination, and querying of the data at many levels.

- http://genome.ucsc.edu/cgi-bin/hgtracks is the url of ucsc genom browser, and then it will show the home page, to its left there is a track, click the first one i.e genome browser.
- Then comes the gateway consisting of group, genome, assembly, position .select group mammal,genome human, Assembly Feb 2009(GRch 37/ hg 19), position chr19:45,409,033945,412,50 and in the search box type APOE and submit
- Then it will show the genome browser Gateway sample search for Human APOE, on clicking on it, it will show the graph.
- The graph consists of baseposition, methylation in different cells, the GC%, restriction enzyme present, spliced EST, alternative splicing sites, presence of CPG Island and methylation.
- It has different annotation tracks below, on arrangeing it, we can add the required tracks.

3.2 ENSEMBL

It is a joint scientific project between the european bioinformatics institute and the welcome trust sanger institute, which was launched in 1999 in response to the imminent completion of the human genome project.. These are shown as data tracks, and individual tracks can be turned on and off, allowing the user to customise the display to suit their research interests. The interface also enables the user to zoom in to a region or move along the genome in either direction.

- www.ensembl.org/ is the url of ENSEMBL genome browser it is used here to see the variants of APOE, detail regions of APOE, variation image and splice variants
- Then type human apoe, it will show apoe list, from that select APOE human.
- To the left side it contains track. And in the middle it consists of transcript table. APOE has 4 transcripts on clicking on the table it will show each transcript separately, and on clicking on protein families it will show the protein present in the transcript.
- Then to the left there are splice variants, regulation, expression, transcript comparission on going into it it will show the detail of APOE.
- Then comes the gene tree which shows which animal's apoe gene is similar to human apoe, the tree is onstuted useinng maximum likelihood methord.
- Then is shown the sequene of APOE and the reagions containing, exons and variations suh as 3 prime UTR, 5 prime UTR inframe insertion,Non-codeing exons,splice regions, intronic, splie acceptor,stop gained, missene,splice donor are highlighted in different colours.

3.3 EMBL- EBI INTEPRO

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool

- This is used to view the protein domain, family and site present in Apolipo protein
- Go to the google page, search intepro protein sequence analysis and classification.
- It will show the page then paste the apoe protein sequence in the box.In the result page it will show, protein family membership, domains and repeats, and detailed signature match of the APOE gene.

3.4 NCBI-unigen

Unigene is an ncbi database of the transcriptome and thus, despite the name, not primarily a database forgenes. Each entry is a set of transcripts that appear to stem from the same transcription locus (i.e. Gene or expressed pseudogene). Information on protein similarities, gene expression, and

genomic location is included with each entry.descriptions of the unigene transcript based and genome based build procedures are available.

- http://www.ncbi.nlm.nih.gov/Then select unigene and type human APOE human and submit.
- Then EST profile is clicked which shows APOE4.
- It will show APOE4 breakdown by bodysites, by health state and by developmetal stage.
- It will show the protein domain site with family.

3.5 SNAP GENE VIEWER

Snapgene 2.3.3 is a powerful molecular biology program that allows you to document in a rich electronic format any dna construct made in your lab. You can select the dna fragments that you wish to fuse and snapgene will design the primers. It also simplifies the planning of a gibson assembly reaction and automates the primer design, but here it is used only to view the restriction enzymes in the gene apoe,with its position mentioned to its site, here the apoe gene is shown in a circular manner.

- The snapgene 2.3.3 is first downloaded and installed to get started.
- Then to its left side there is a new option ,click on it, then click open new DNA file, paste the fasta format sequence of APOE taken from NCBI
- Then It will show two option i.e view circular , linear we can click on whatever required then it will show the gene in that form with restriction enzymes with its base pair position.

3.6 BIOGRID

The biological general repository for interaction datasets (biogrid) is a curated biological database of protein-protein and genetic interactions. It strives to provide a comprehensive resource of protein–protein and genetic interactions for all major model

- http://thebiogrid.org/ is the url to reach biogrid.
- The biogrid home page has a search option on its top, type the APOE and human.
- Then towards its right there is a graphical viewer on clicking that it will show all the genes related to APOE in a circular form.

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION

4.1 Ensembl

Table-1 Transcript table:

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CDS incomplete	CCDS	GENCODE basic
APOE-001	ENST00000252486	1208	ENSP00000252486	317	Protein coding	-	CCDS12647	Y
APOE-005	ENST00000425718	923	ENSP00000410423	219	Protein coding	3'	-	-
APOE-004	ENST00000434152	864	ENSP00000413653	269	Protein coding	3'	-	-
APOE-002	ENST00000446996	737	ENSP00000413135	216	Protein coding	3'	-	-
APOE-003	ENST00000485628	771	No protein product	-	Retained intron	-	-	-

APOE transcript table

Name	APOE (HGNC Symbol)
Synonyms	AD2 [To view all Ensembl genes linked to the name click here]
CCDS	This gene is a member of the Human CCDS set: CCDS12647
RefSeq	Overlapping RefSeq Gene ID 348 matches and has similar biotype of protein_coding
Ensembl version	ENSG00000130203 5
Gene type	Known protein coding

This is the transcript table of apoe containing five types of transcript.The table shows all splice variants for a gene and includes noncoding transcripts. Each transcript id includes a unique, stable 11 digit number. Transcripts beginning with enst are human transcripts. A three-letter code is inserted for other species;

Name - the hgnc name (for human), or the best match to a known gene name in a public database.

Synonyms - other gene names used for this particular gene.

Ccds - if the gene has transcripts in the consensus coding sequence set, the ccds ids will be listed.

Gene type - the gene type includes both status and biotype.

Prediction method - indicates if automatic annotation and/or manual curation were used to determine transcripts belonging to this gene.

Alternative genes - matching gene ids from the Vega or Havana.

4.2 Protein summary

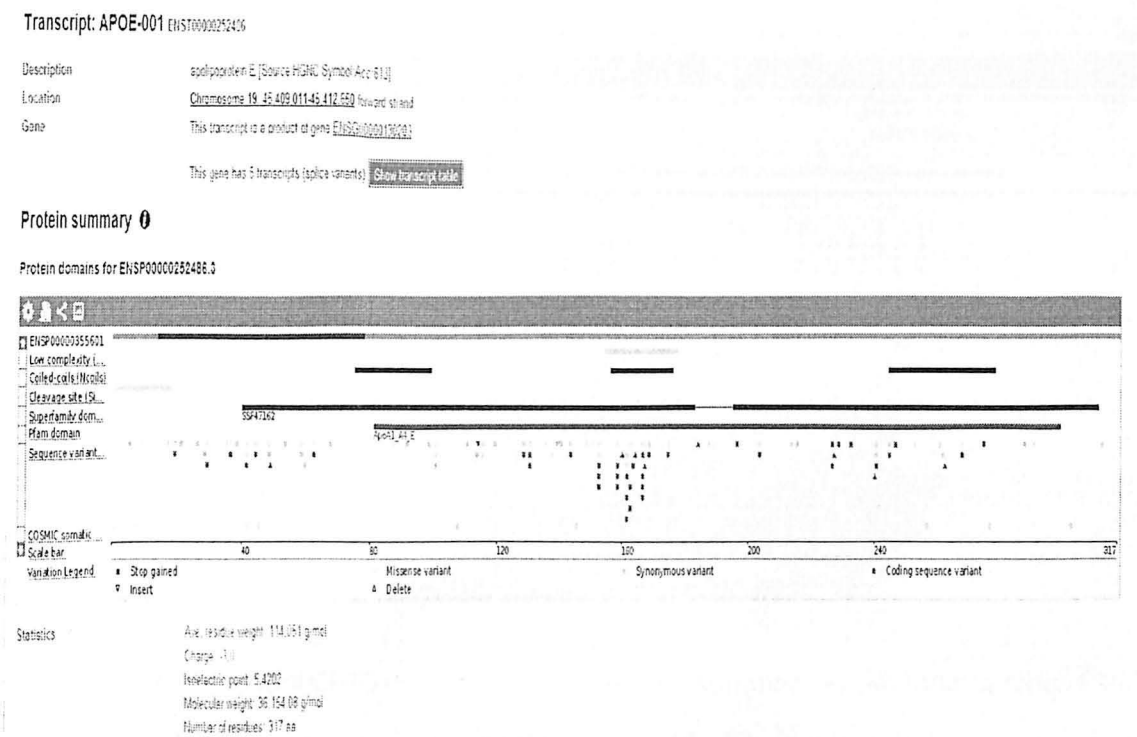


Fig-1 ***Protein summary of APOE -001***

- This is the protein summary of transcript: apoe-001 it consists of ensp00000252486 which is the protein summary id, includes a unique stable 11 digits. The yellow bar is the low complexity seq, the blue bar indicates coiled coils (ncoils), the pink represents cleavage site, the blue represents super family domain , grey represents pfam domain, the dots represenr sequence variant, the variant legend are red for stop gained, yellow for missense variant, light green for synonymous variant and deep green for codeing sequence variant.

- **Statistics**

Ave. Residue weight: 114.051 g/mol

Charge: -3.0

Isoelectric point: 5.4202

Molecular weight: 36,154.08 g/mol

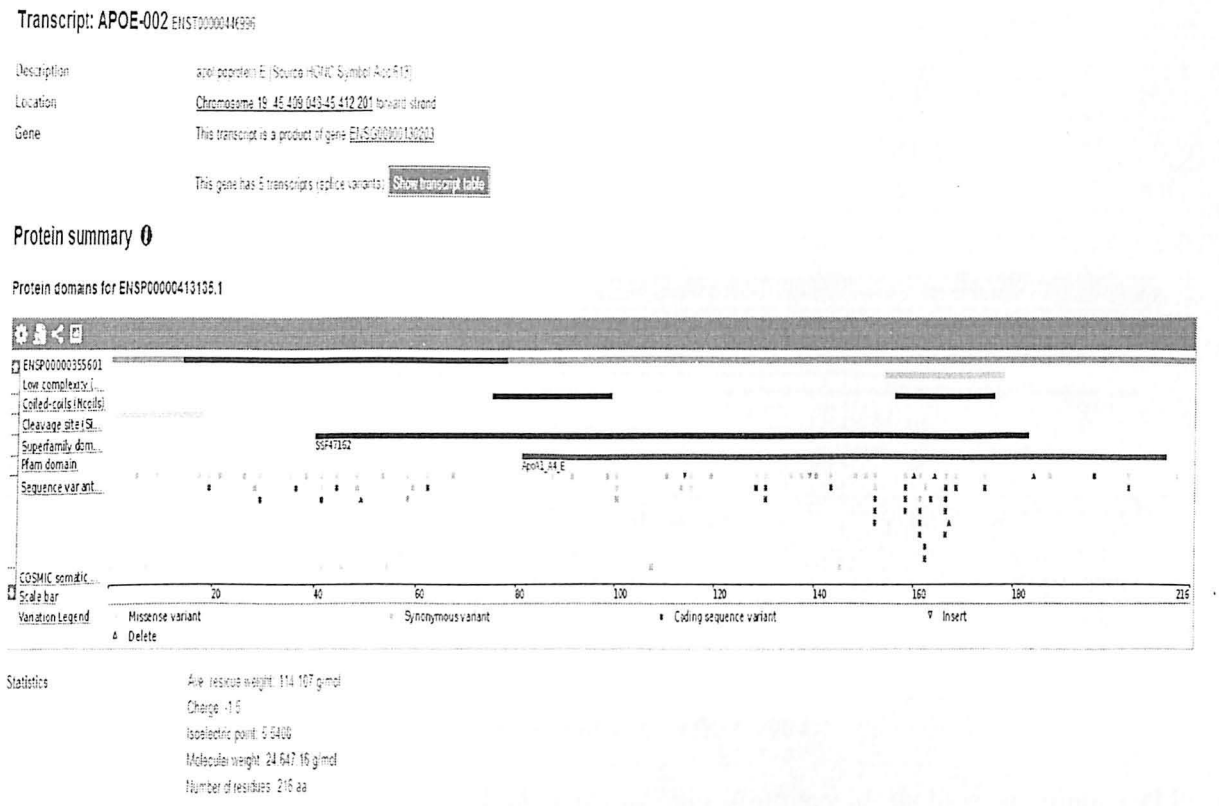


Fig-2 ***Protein summary of APOE -002***

- It consists of ensp00000413135 which is the protein summary id, includes a unique stable 11 digits.The yellow bar is the low complexity seq, the blue bar indicates coiled coils (ncoils), the pink represents cleavage site, the blue represents super family domain , grey represents pfam domain, the dots represents sequence variant, the variant legend are red for stopgained, yellow for missense variant, light green for synonymous variant and deep green for codeing sequence variant.

- **Statistics**

Ave. Residue weight: 114.107 g/mol

Charge: -1.5

Isoelectric point: 5.5408

Molecular weight: 24,647.16 g/mol

Number of residues: 216 aa

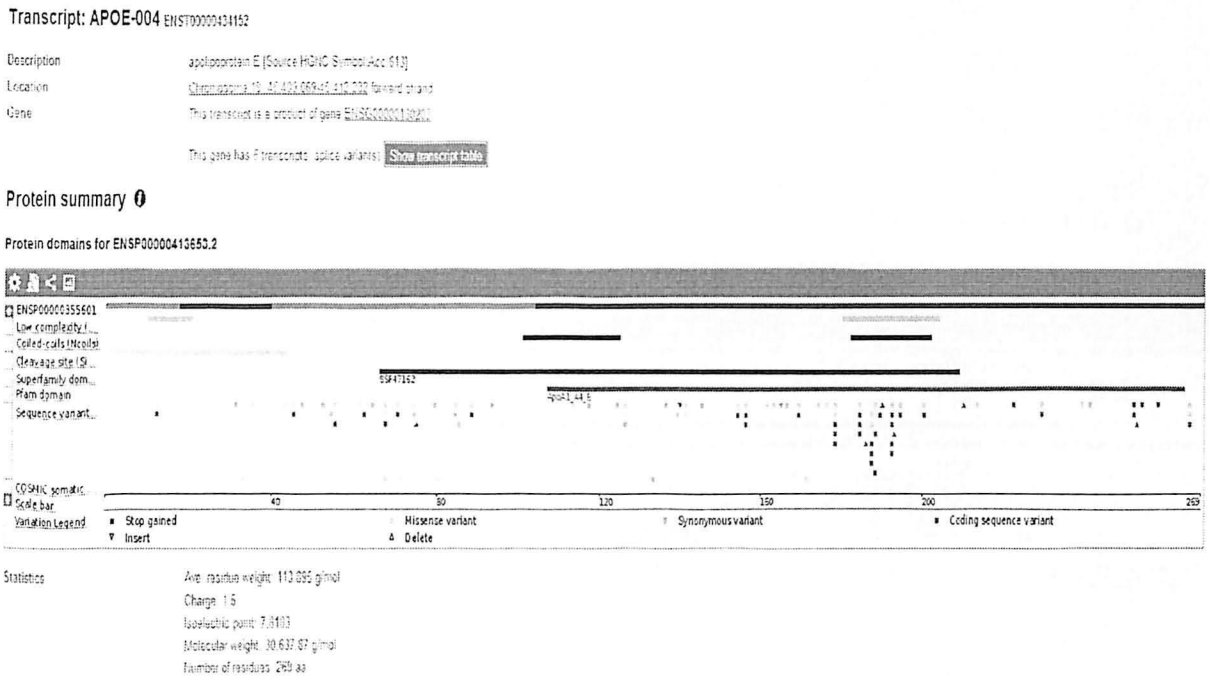


Fig-3 ***Protein summary of APOE -004***

- It consists of ensp00000413653 which is the protein summary id, includes a unique stable 11 digits.The yellow bar is the low complexity seq, the blue bar indicates coiled coils (ncoils), the pink represents cleavage site, the blue represents super family domain , grey represents pfam domain, the dots represents sequence variant, the variant legend are red for stop gained, yellow for missense variant, light green for synonymous variant and deep green for codeing sequence variant.

- **Statistics**

  Ave. Residue weight: 113.895 g/mol

  Charge: 1.5

  Isoelectric point: 7.8183

  Molecular weight: 30,637.87 g/mol

  Number of residues: 269 aa

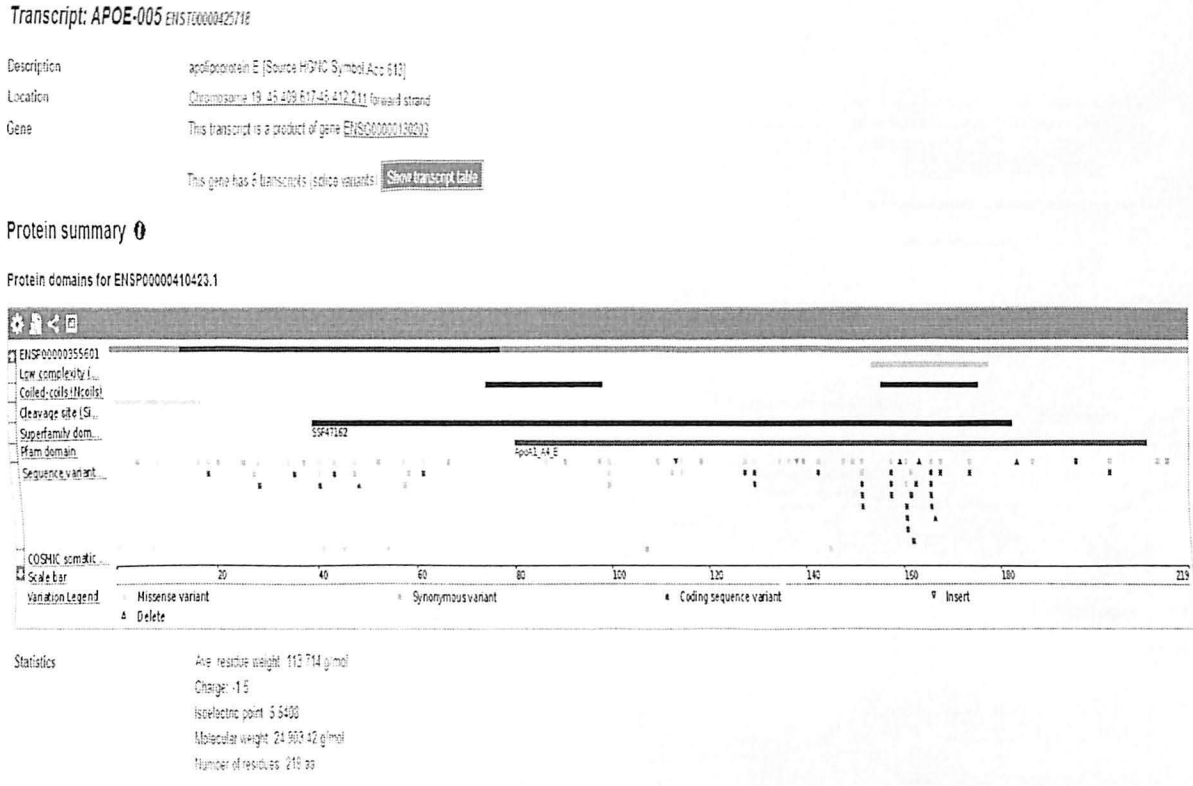


Fig-4 ***Protein summary of APOE -005***

- It consists of ensp00000355601 which is the protein summary id, includes a unique stable 11 digits.The yellow bar is the low complexity seq, the blue bar indicates coiled coils (ncoils), the pink represents cleavage site, the blue represents super family domain , grey represents pfam domain, the dots represents sequence variant, the variant legend are red for stop gained, yellow for missense variant, light green for synonymous variant and deep green for codeing sequence variant.

- **Statistics**

  Ave. Residue weight: 113.714 g/mol

  Charge: -1.5

  Isoelectric point: 5.5408

  mol weight: 24,903.42 g/mol

  Number of residues: 219 aa

4.3 Region in details

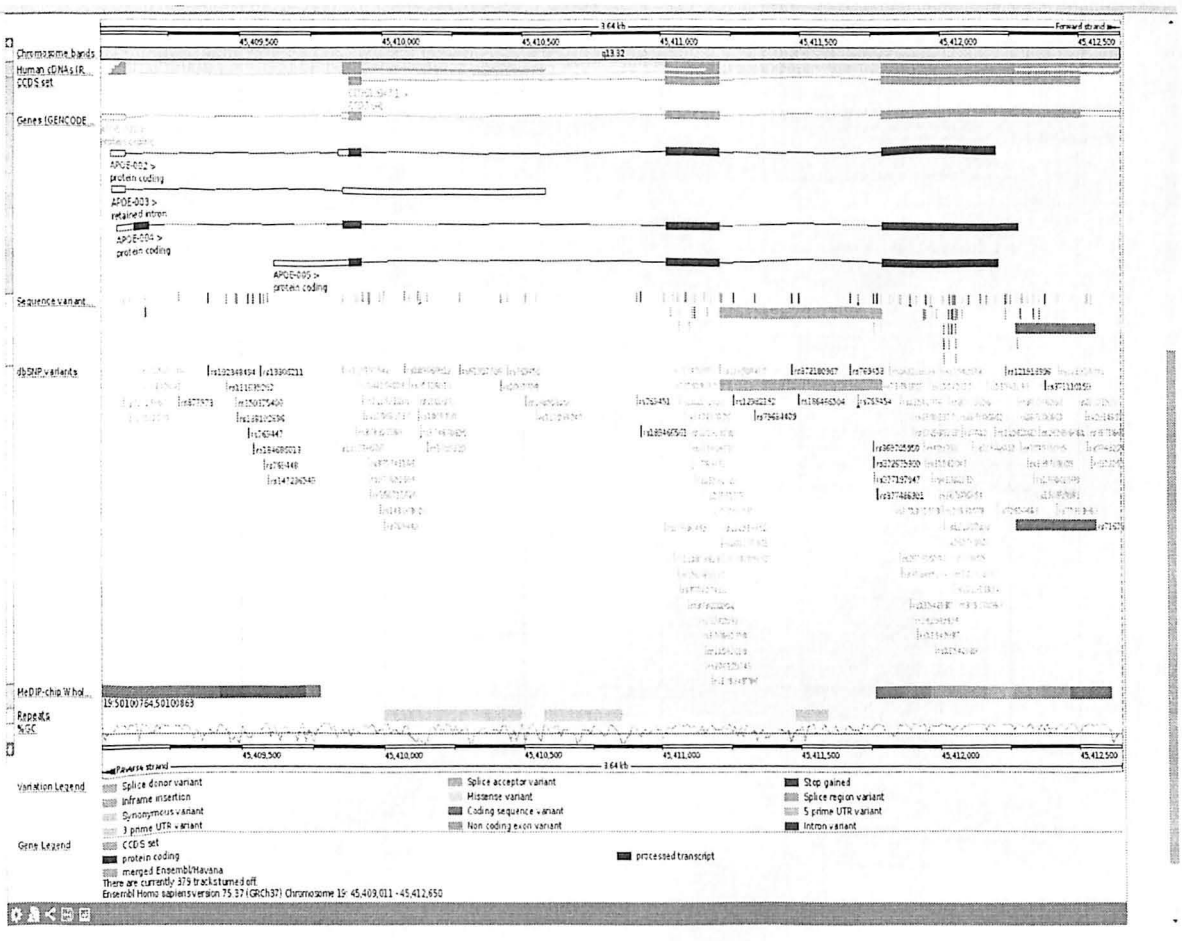
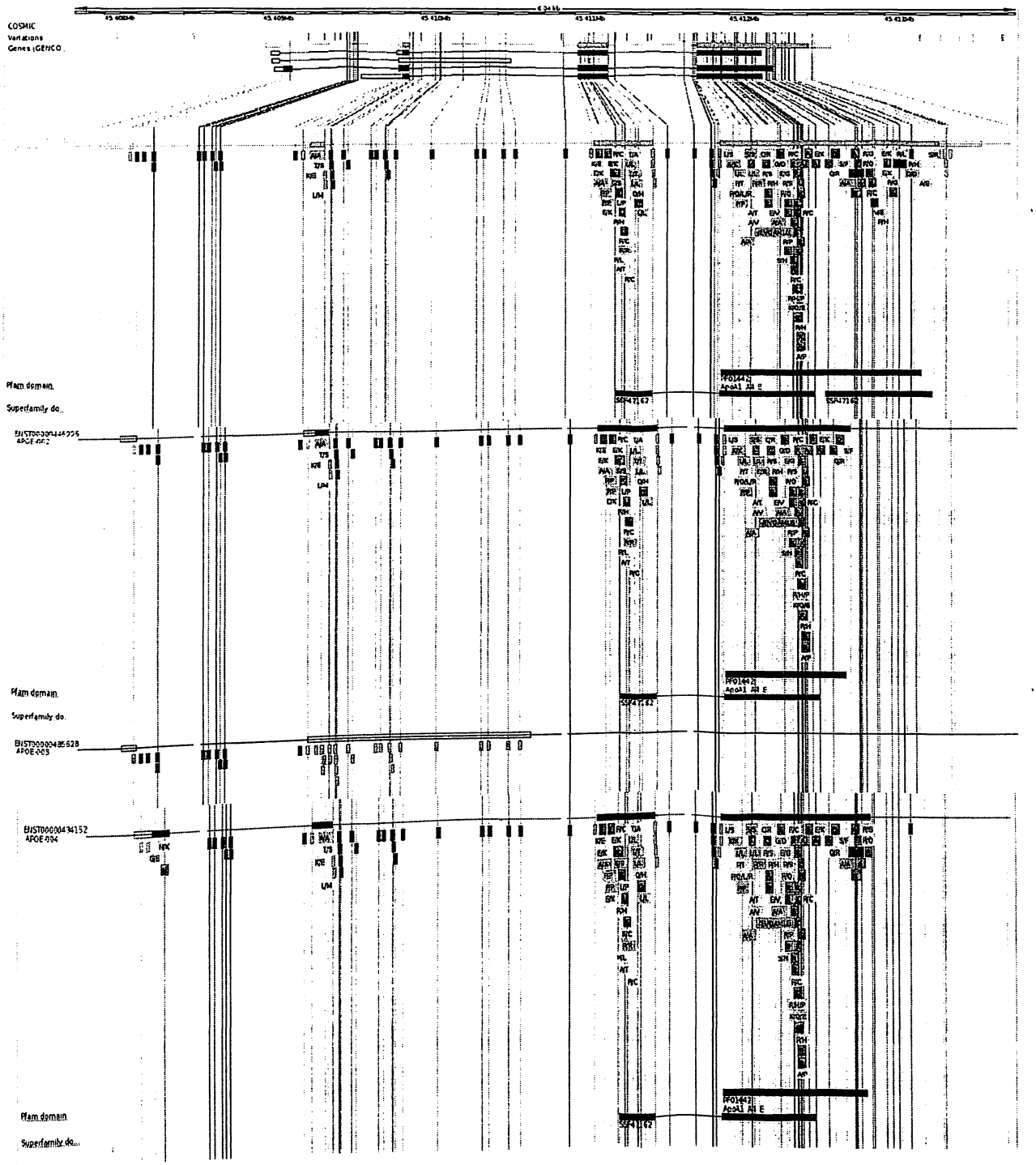


Fig-5 ***Region in details of APOE4 gene***

- The above table shows the gene region in details This is the apoe gene regions shown in details,in the picture all the orange colour codes are splice donor variant, the pink ne are inframe insertion, green , are synonymous variants, light blue are 3 prime utr variants, enlongated orange boxes are splice acceptor variants,yellow shows missence variants, dence green enlongated boxes shows codind sequence variants, green codes for non codeing exon variants, red code for stop gained, peach for splice region variants, navy blue for intron variants,merun box shows exons and light green shows ccds set. It shows in details what are the SNP present, sequence variants, the gc percentage lying in which area of the gene.

4.4 Variation image .



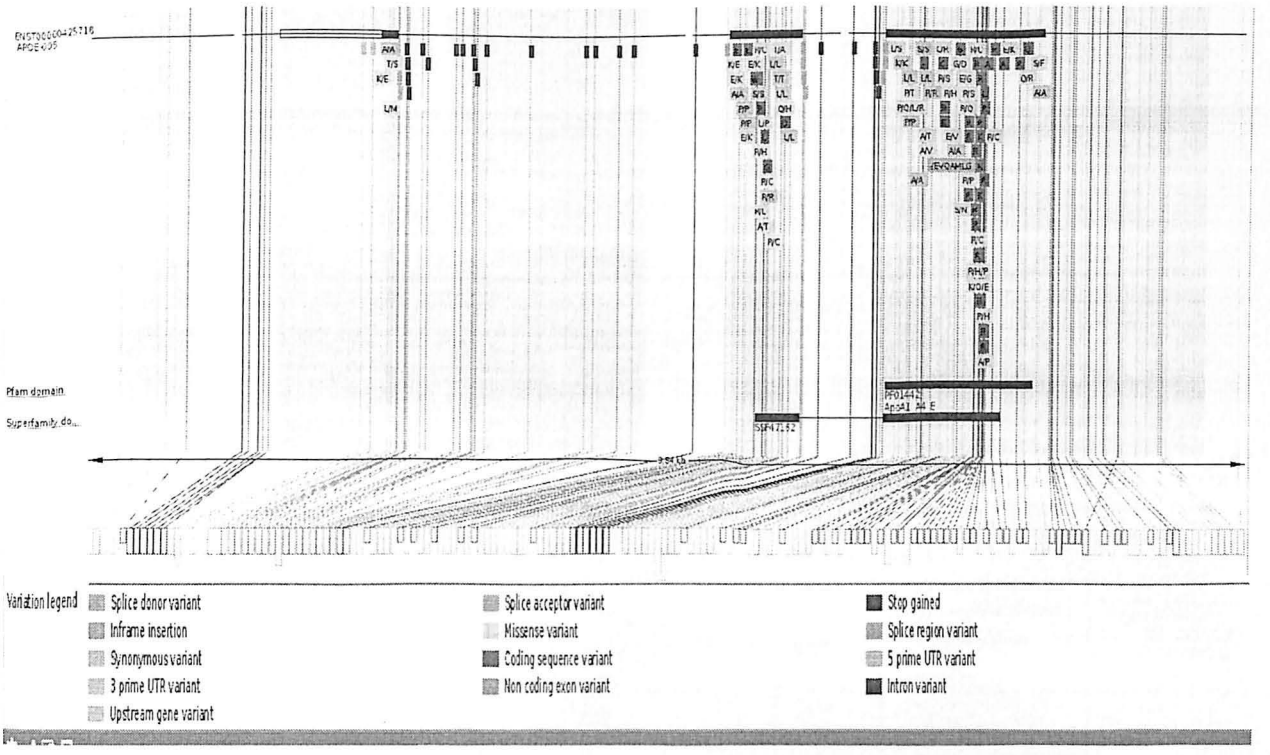


Figure-6 ***Variation region of Apoe 1, Apoe2, Apoe3,Apoe4 region.***

- From the above picture we can find that the Orange shows splice donor variants -a splice variant that changes the 2 base region at the 5' end of an intron, it has 12 number of variant sequence. And splice region variant- a sequence variant in which a change has occurred within the region of the splice site,either within 1-3 bases of the exon or 3-8 bases of the intron .Red shows stop gained- a sequence variant whereby atleast one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript.Pink shows inframe insertion- an inframe non synonymous variant that inserts bases into in the coding sequence.Yellow shows missense variants- a sequence variant that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved. Deep green shows coding sequence variants- a sequence variant that change the codeing sequence.Blue shows 5 prime utr variant – a utr variant of 5 prime utr, and 3 prime utr- a utr variant of 3 prime utr. Blue represents.Intron variant- a transcript variant occurring within an intron.

4.5 splice variants

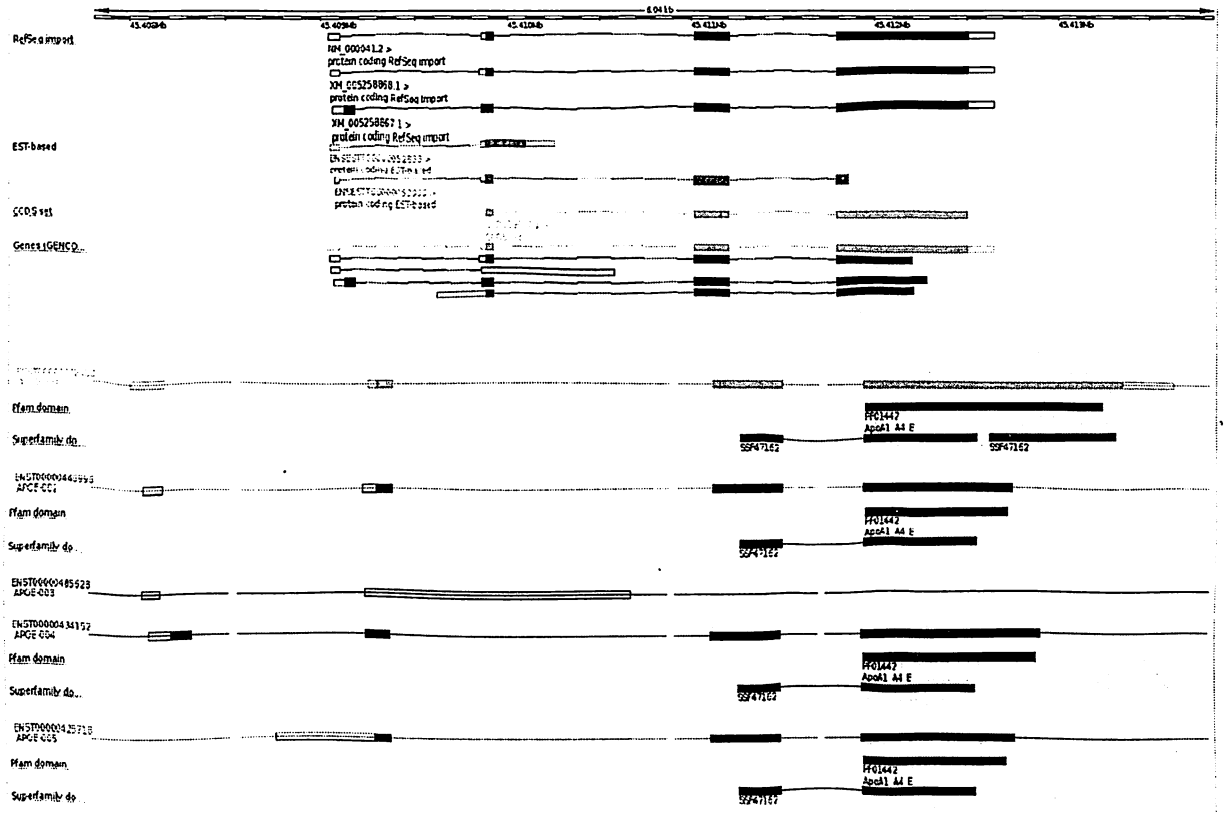


Fig-7 ***Splice variants ofApoe 1, Apoe2, Apoe3,Apoe4 region***

- This view shows all spliced transcripts for a gene, including est transcripts and ncrnas (non-coding rnas).Transcripts are drawn as boxes (exons) and lines connecting the boxes (introns). Filled boxes represent coding sequence and unfilled boxes (or portions of boxes) represent untranslated regions (utr).For coding transcripts (gold or red transcripts), protein motifs and domains are shown in purple. Click on a domain (purple block) to see more information such as amino acid positions and links to individual records. These motifs and domains come from various databases listed at the left of the view, for example prosite and superfamily.verticalbrown highlights show exon positions so that all exons across transcripts may be compared.

4.6 Molecular functions includes

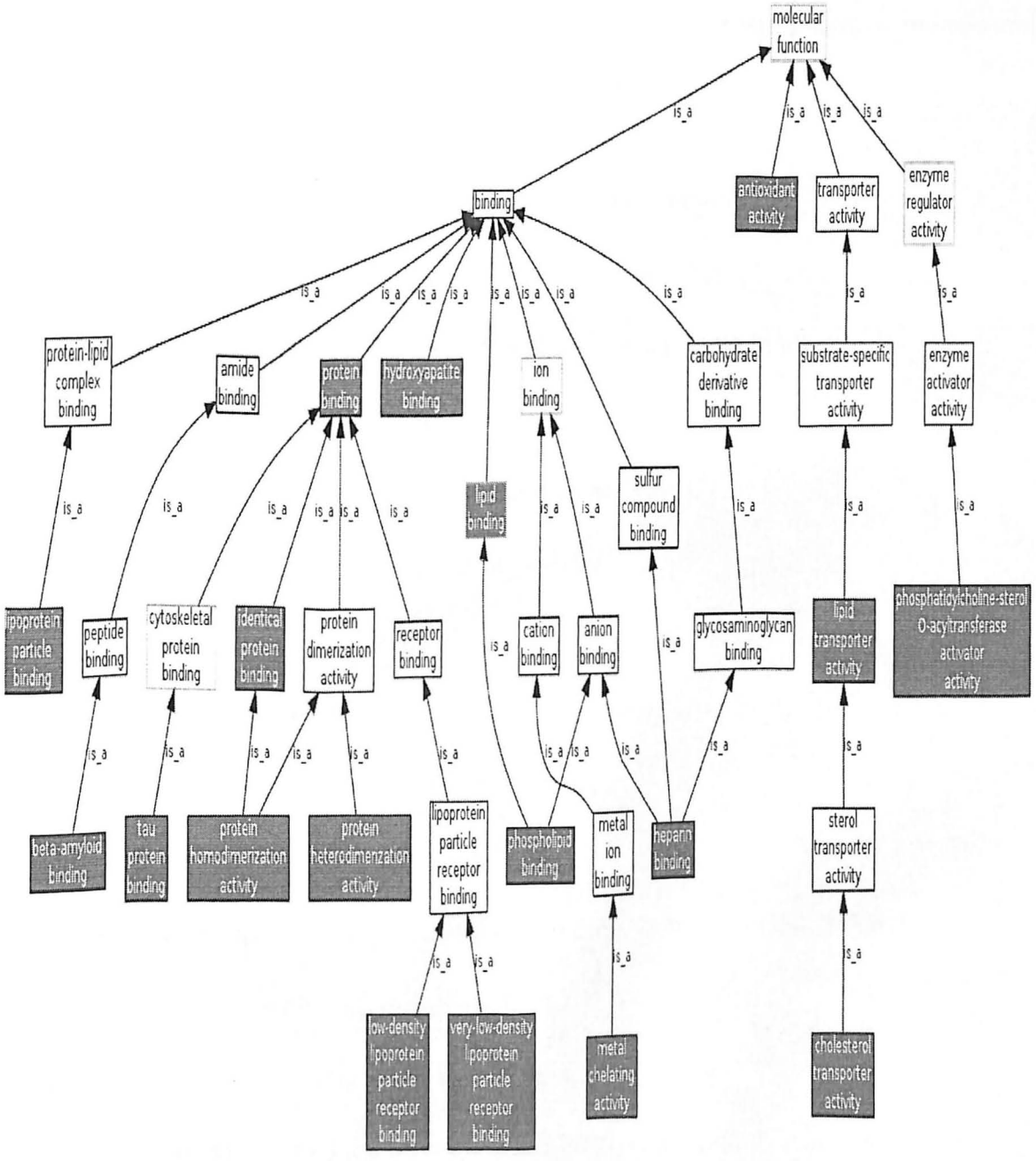


Figure-8 ***GENE ONTOLOGY pathway of APOE***

- Antioxidant activity
- Transporter activity -> substrate specific transporter activity -> lipid transporter activity -> sterol transporter activity -> cholesterol transporter activity
- Enzyme regulator activity -> enzyme activator activity -> phosphatidylcholine-sterol o-acyltransferase activator activity
- Binding

1) Protein –lipid complex binding -> lipoprotein particle binding

2) Amide binding -> peptide binding -> beta amyloid binding

3) Protein binding -> cytoskeletal protein binding -> tau protein binding

->Identical protein

->Protein homodimerization activity

->Protein dimerization activity

->Protein heterodimerization activity

->Protein hetero merisation activity

4) Hydroxyapatite binding

5) Lipid binding -> phospholipid binding

6) Ion binding -> cation binding -> metal ion binding -> metal chelating activity

->Anion binding -> heparin binding

7) Sulphur compound binding -> heparin binding

8) Carbohydrate derivative binding ->glycosaminoglycan binding -> heparin

4.7 Regulation

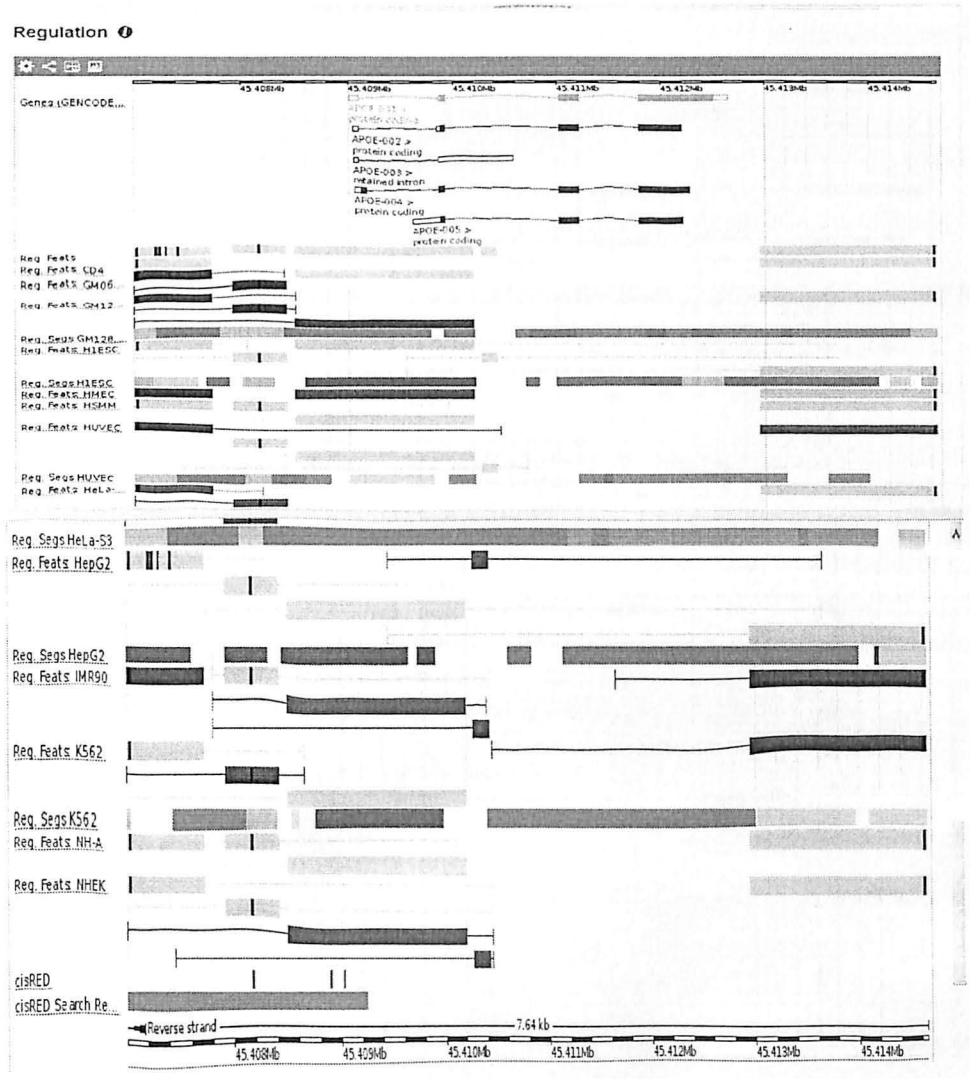


Figure-9 ***Apoe Regulation***

From the above picture we can conclude that the classification of these regions is; promoter associated, gene associated, poliii associated, non-gene associated, unclassifiedEach region is constructed from experimental data that describes chromatin activity or state (e.g. Dnase1, histone modifications, polymerase) and transcription factor binding following a four-step analysis:Genomic regions characterised by dna protein binding or open chromatin across all cell types are defined as regulatory features;Combinations of epigenetic marks and transcription factor binding sites (tfbs) are associated with one of the classification above (e.g. Promoter associated, unclassified, etc) for specific cell lines;For each cell type, any regulatory feature that overlaps with a combination of epigenetic markers and/or tfbs are assigned the appropriate classification;The regulatory features associated to tf binding are annotated with the corresponding tf binding motifs. For example, a regulatory feature that overlaps with a ctcf binding feature is annotated with all the potential ctcf binding sites that it contains.

4.8 APOE expression graph

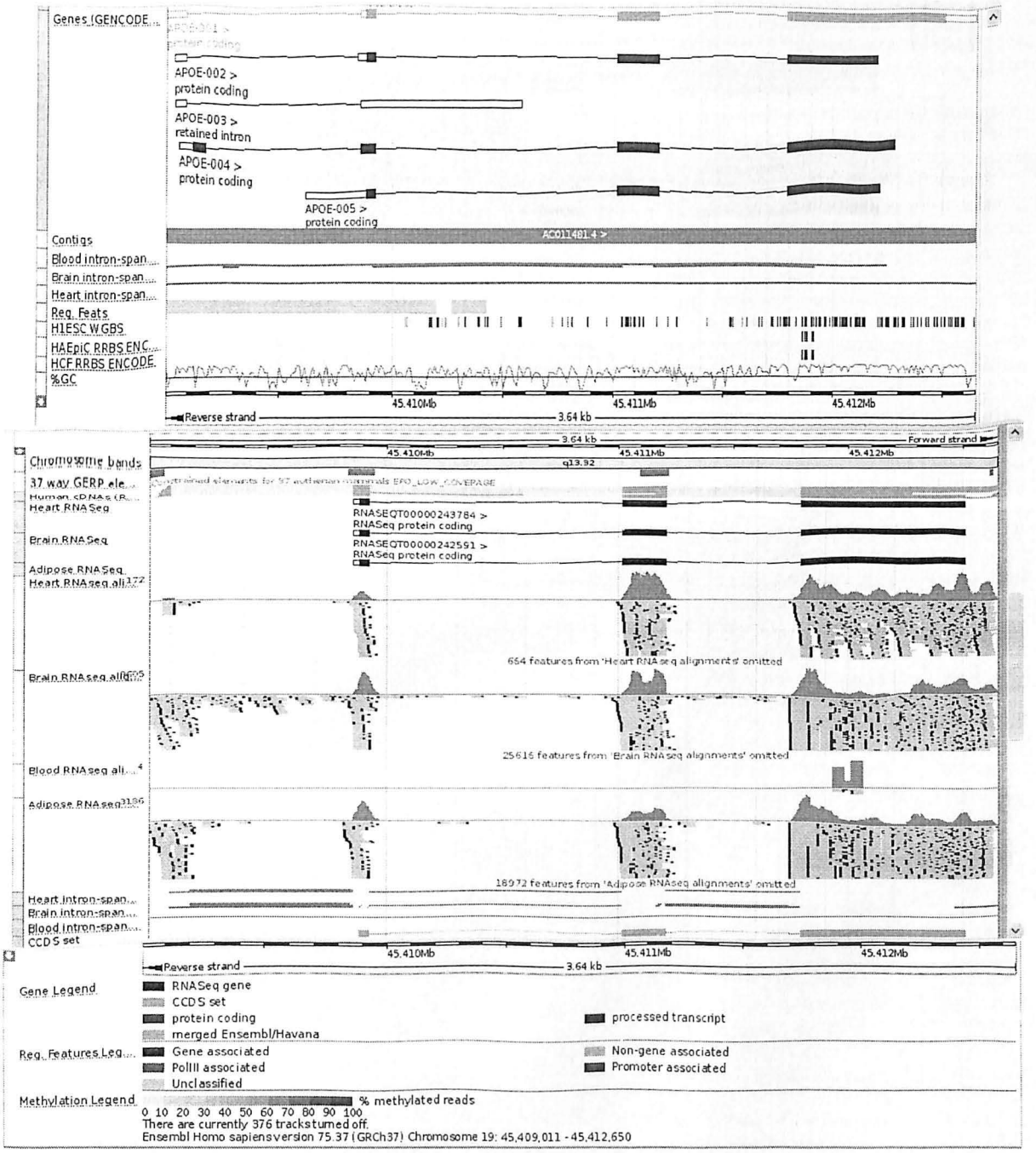


Fig-10*APOE expression graph*

This picture represents the expression data of apoe gene, the blue shows rna seg , green for ccds, red for protein coding, and light blue for processed transcripts, orange for non-gene associator, green for promotor.It also shows methylation in tissues i.e adipose, adrenal, blood, brain, breast,colon,heart.

4.9 Phylogenetic Tree

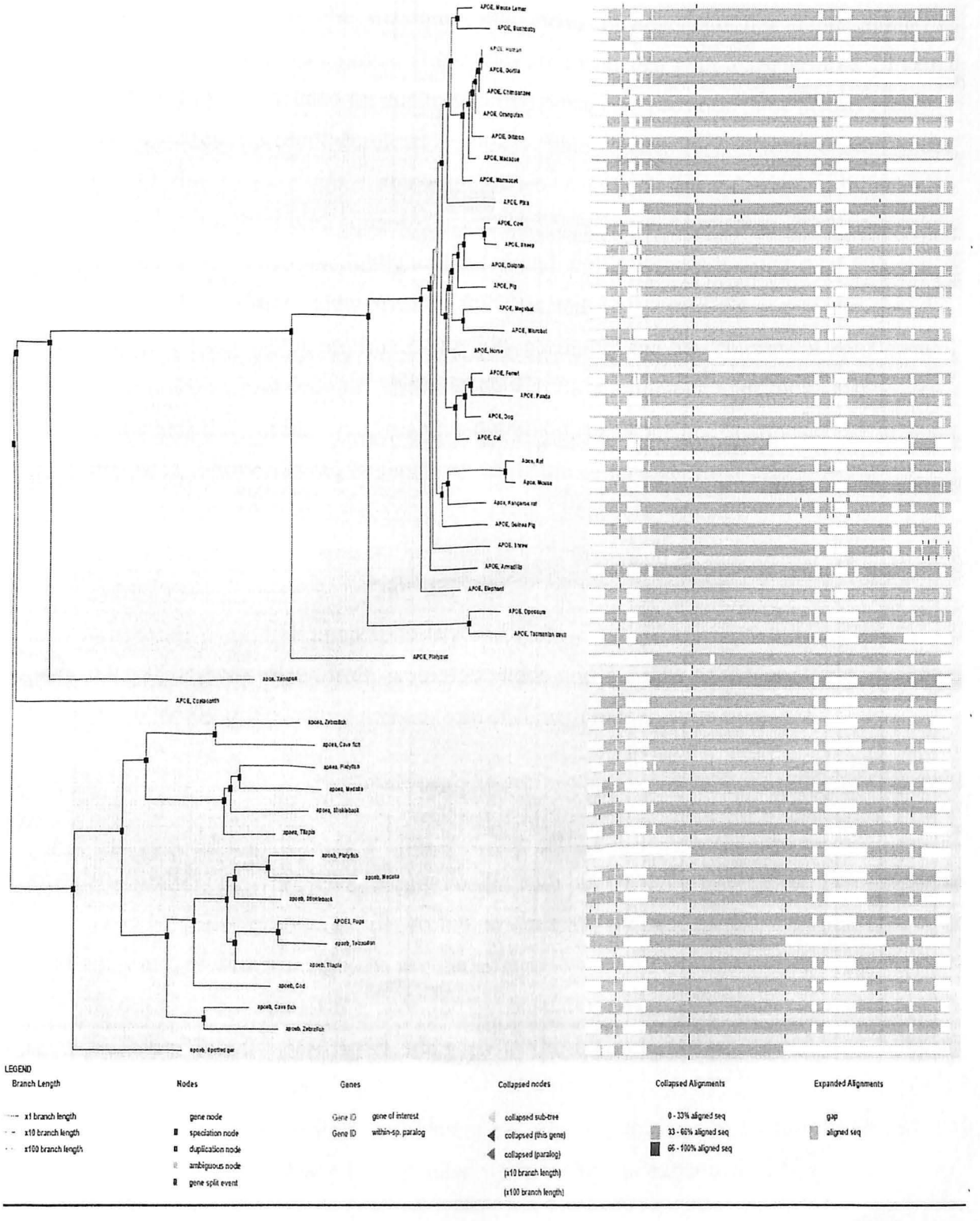
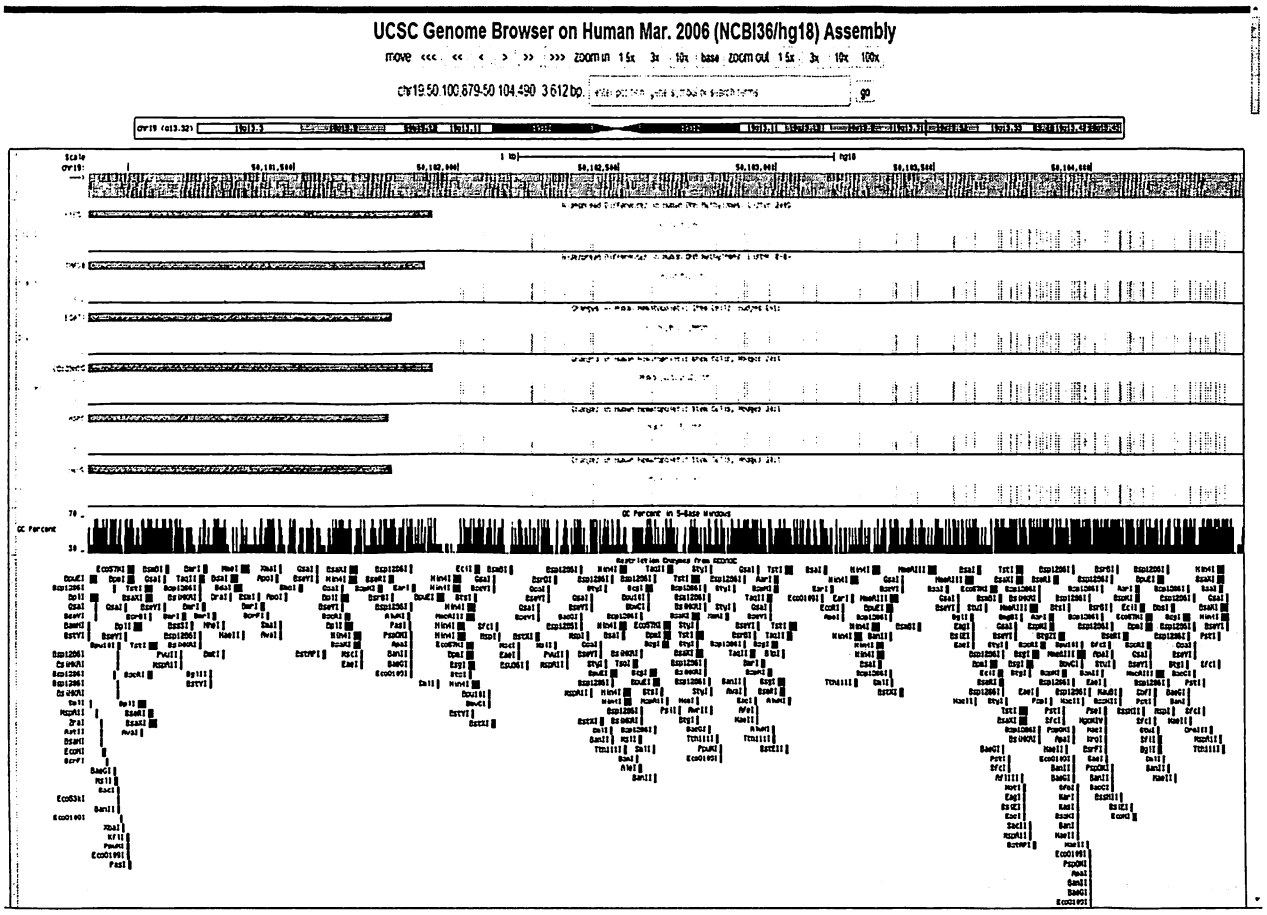


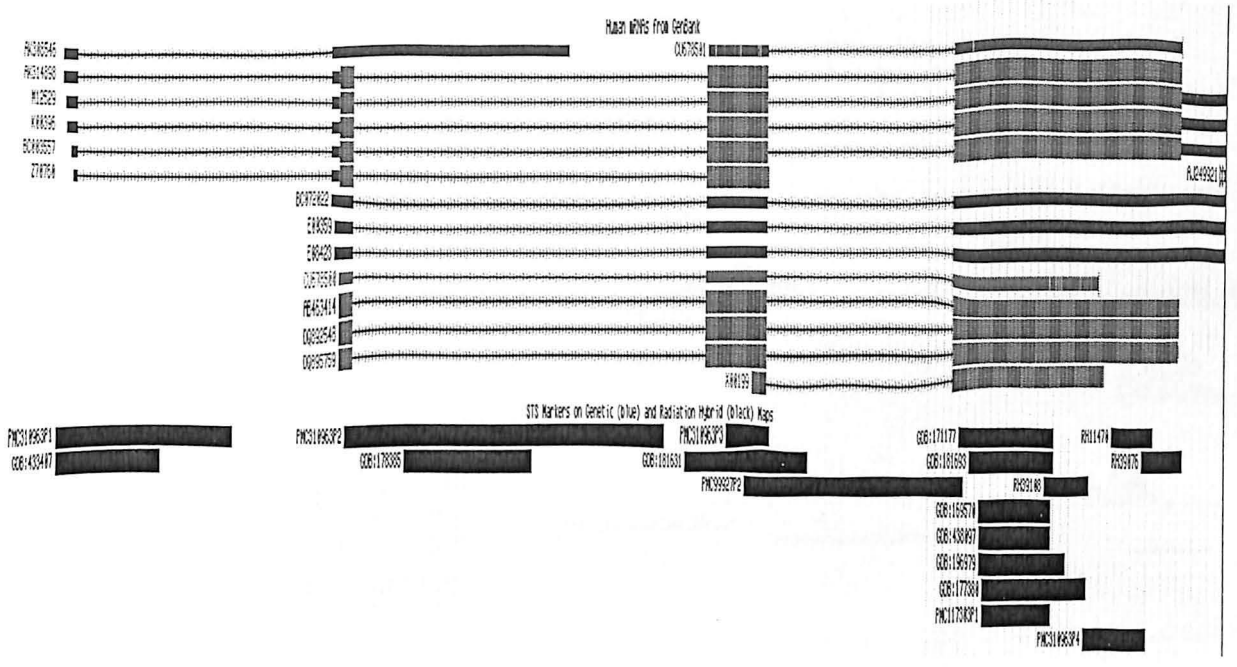
Fig-11 *Phylogenetic Tree*

- Number of genes -49, Number of speciation node- 45, Number of duplicate- 1, Number of ambiguous -2 and Number of gene split events- 0.
- The display shows the maximum likelihood phylogenetic tree representing the evolutionary history of genes. These trees are reconciled with a species tree, . Internal nodes are then annotated for duplication (red boxes) or speciation (blue boxes) events.
- Red squares represent duplications nodes; blue squares represent speciation nodes, giving rise to paralogues and orthologues. Another class of node, ambiguous, is shown as a lighter blue square.
- The gene of interest is highlighted in red and within-species paralogues are shown in blue, if the option to view paralogues is selected.
- The 1st cluster which appears in the grey portion of the tree consists of apoe gene of 2 homologus wet lemurs showing 66-100 % alighnment with human apoe gene, while gorilla, chimpanzee, orangutan,gibbion, macaque,and marmoset, all of these primates, shows 33-66 % similarity with human apoe gene.
- The next gene node consists of pika apoe gene, 33-66% alighnment with human apoe gene.
- The 2nd cluster appearing in the blue portion, conists of apoe gene of 11 homologus laurasiatherian mammals, and they show 66-100% alignment in human sequence when compared to human apoe sequence.
- The 3rd cluster arises from ambigious node and consists of 4 homologus rhodents, showing 66-100 % alingnment in sequence with human apoe sequence.
- The 4th cluster consists of shrew, armadillo, elephant arising from gene node and 2 homologus mursipals; all shows 33-66% alignment sequence with human apoe sequence.
- The next are 2 gene node arising from two different speciation node, i.e xenopus and coelacanth, showing 33-66% alighnment with human apoe gene.
- The 5th cluster consists of 16 homologus ray-finned fishes showing 66-100% alighnment with human apoe sequence.
- Multiple alignment of the peptides (green bars). Green bars shows areas of amino acid alignment, white areas are gaps in the alignment. Dark green bars indicate consensus alignments.
- The consensus amino acid alignment corresponds to the consensus residues in the collapsed node, and will be expanded when the tree is expanded.

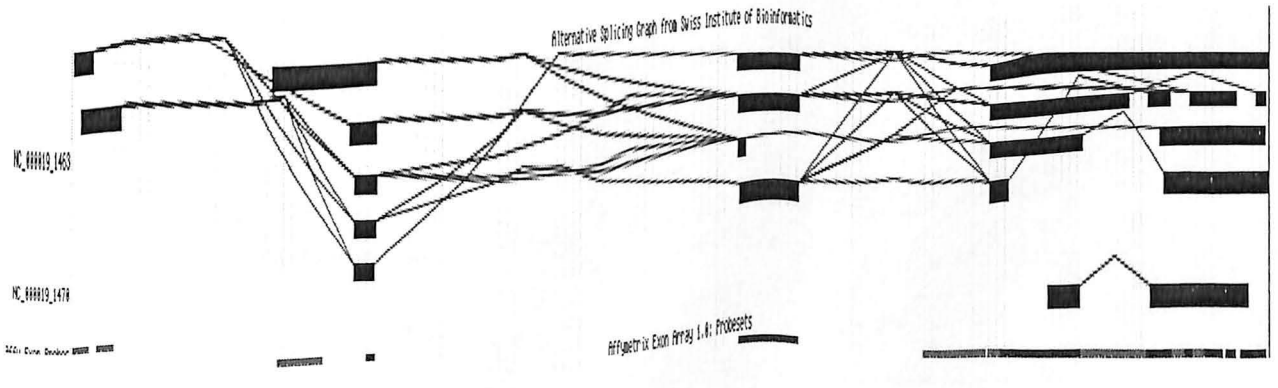
4.10 UCSC Genome Browser



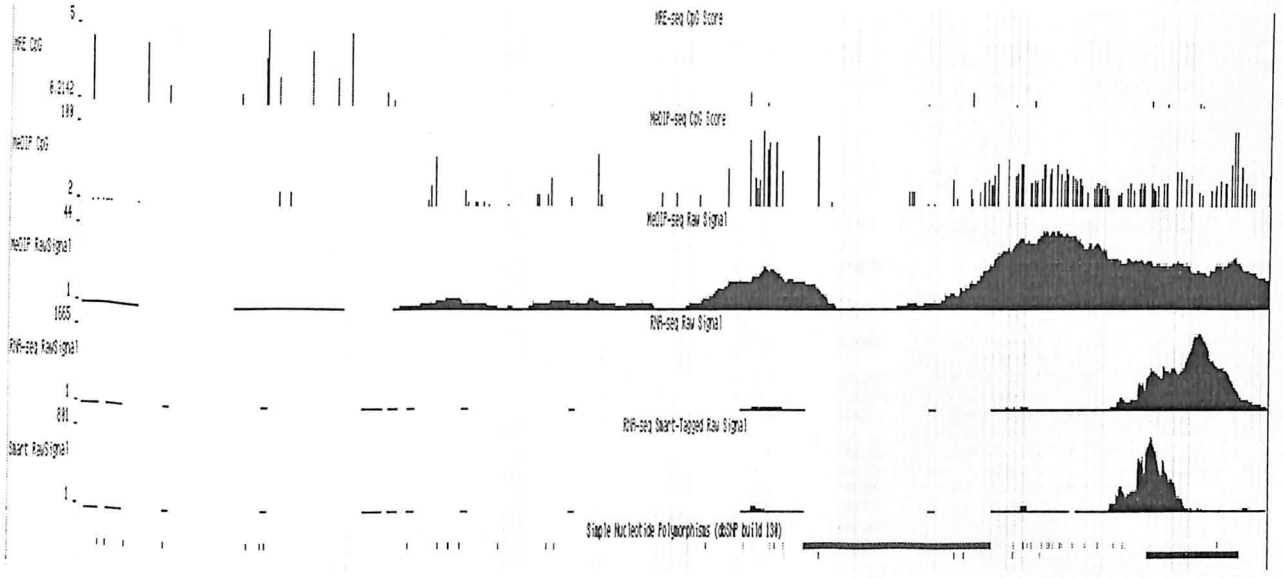
- The 1st bar consisting of red and green lines, on zooming it we find boxes boxes of green in which m is mentioned, so it can be concluded that all the green lines are methionine i.e start codan where as red are stop codans. h1esc is the embriyonic steam cell, imr90 is the fetus lung fibroblast, b cells are human b cells, cd133hcs , hspc,neut are hematopoetic steam cells. The blue intervals represent hypo-methylated region, yellow bars represent methylation levels. Then comes the gc percentage, the pink peak shows the presence of gc i.e if the height is more it shows that more amount of gc is present in that region.



- These are the mrna present, the orange and purple line are unalignable query sequence, orange for middle of a sequence and purple for beginning or end.



- At the top the zig zag lines represent alternative splicing graph, the black box represents exons.



- This represent the cpg presence, the first peak is the mre cpg , medip cpg,medip rawsignal are the brain dna methylation graph,the peaks and density shows the percentage of cpg present in the graph.

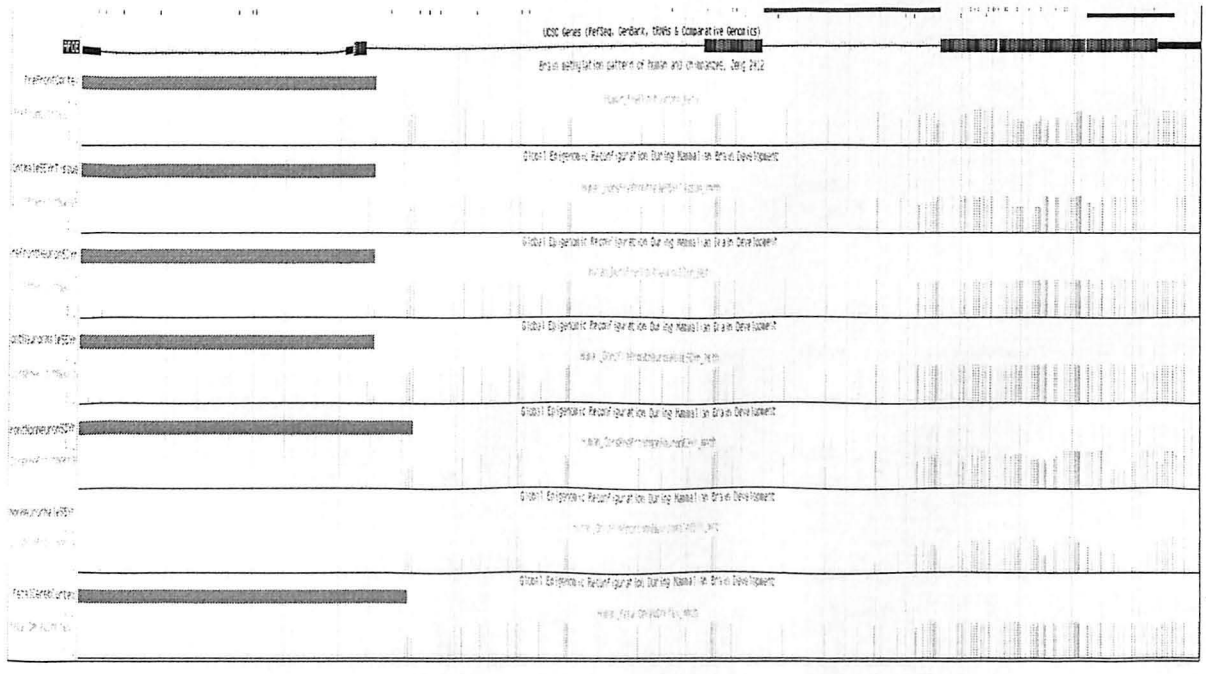


Fig-12 UCSC *GENOME BROWSER view of APOE*

- The above are shown brain methylation pattern in human prefront cortex, human dors prefront, human dors prefront neuron, human fetal cereb cortex, the yellow bars represent methylation in those portions.

4.11EST profile UNIGENE

EST Profile

Hs 654439 - APOE Apolipoprotein E

Breakdown by Body Sites

	Hs 654439	
adipose tissue	0	0/12866
adrenal gland	667	22/32940
ascites	0	0/39834
bladder	0	0/29860
blood	0	0/122252
bone	153	11/71618
bone marrow	0	0/48737
brain	623	681/1092688
cervix	61	3/48486
connective tissue	160	24/149072
ear	0	0/16100
embryonic tissue	295	63/212896
esophagus	49	1/20154
eye	129	27/208840
heart	122	11/89524
intestine	107	25/231981
kidney	555	117/210778
larynx	0	0/23466
liver	414	86/206291
lung	376	126/334815
lymph	0	0/44302
lymph node	55	5/89748
mammary gland	185	28/151230
skin	839	177/210759
spleen	337	18/53397
stomach	41	4/95679
testis	199	87/435204
thymus	12	1/79697
thyroid	42	2/46583
tonsil	0	0/17021
trachea	0	0/51780
umbilical cord	0	0/13764
uterus	56	13/232093
vascular	77	4/51649

Breakdown by Health State

	Hs 654439	
adrenal tumor	316	4/12655
bladder carcinoma	0	0/17584
breast (mammary gland) tumor	193	18/93090
cervical tumor	0	0/34484
chondrosarcoma	374	31/82838
colorectal tumor	44	5/112517
esophageal tumor	57	1/17245
gastrointestinal tumor	25	3/118498
germ cell tumor	220	58/263230
glioma	3199	343/107194
head and neck tumor	37	5/133826
kidney tumor	145	10/68872
leukemia	10	1/94479
liver tumor	510	49/96023
lung tumor	116	12/102765
lymphoma	69	5/72196
non-neoplasia	31	3/96623
normal	161	539/3328811
ovarian tumor	183	14/76185

Breakdown by Developmental Stage

		Hs.654439
embryoid body	457	32/69969
blastocyst	390	24/61448
fetus	143	80/556978
neonate	0	0/31070
infant	0	0/23511
juvenile	17	1/55574
adult	171	329/1921829

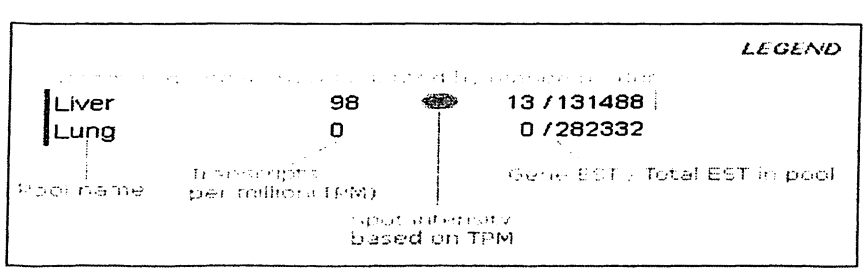


Figure-13 ***EST profile UNIGENE***

- The first table represents est profile of apoe, the 1st,table represents breakdown of apoe in different body sites, below it is the table showing distribution of apoe with respect to health state,and the 3rd, table shows breakdown of apoe as per its developmental stage.The tables consists of poolname, its transcription per million, the spot intensity shows concentration of apoe based on tmp and then the gene est present / the total est in the pool.Adipose tissue, ascites, bladder, blood, bone marrow, ear, larynx, lymph, tonsil, trachea, umbilical cord is the body sites in which apoe are not expressed. Where as in vascular ,thyroid,thymus, stomach,spleen,lymph node,eye,heart, intestine , bone and adrenal glands it is expressed a little , in brain ih shows highest concentration and kidney ,liver ,lung and skin it has moderate expression of apoe.The second table represents breakdown of apoe by health state ,in glioma and in a normal persons its concentration is high, in in bladder carcinoma and cervical tumor apoe is not at all expressed , and it is present moderately in ovarian tumor, lymphoma, bladder carcinoma, cervical tumor, kidney tumor, leukemia,liver tumor ,gastro intestrinal tumor, germ cell tumor,adrenal tumor, brest tumor,condrosarcoma,colorectal tumor,esophageal tumor...The third table shows expression of apoe in developmental stage; in neonate and infants the apoe expression is nil, highly expressed in adults where as moderately found in embriyonic body, blastocyst, fetus and juvenile.

4.13 SNAP GENE VIEWER

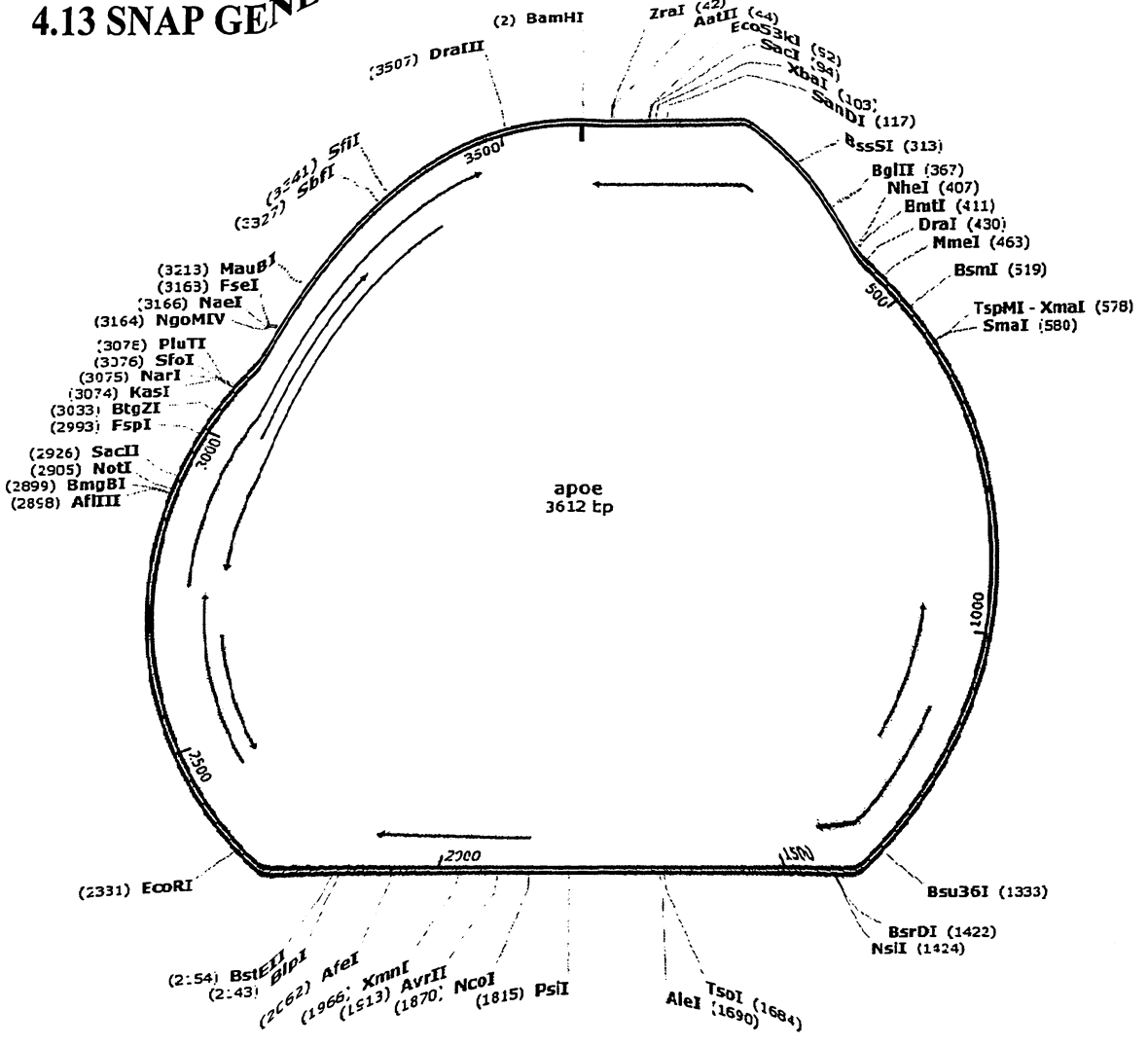


Figure-15 SNAP GENE view of APOE

Snapgene 2.3.3 is a powerful molecular biology program that allows you to document in a rich electronic format any dna construct made in your lab. You can select the dna fragments that you wish to fuse and snapgene will design the primers. It also simplifies the planning of a gibson assembly reaction and automates the primer design, but here it is used only to view the restriction enzymes in the gene apoe, with its position mentioned to its site, here the apoe gene is shown in a circular manner, the is a option shown in this tool in which the linear form of the gene can also be viewed.it consists of nine open reading frame out of which yellow indicates the top three frames where as green indicates bottom two frames.

4.13 SNAP GENE VIEWER

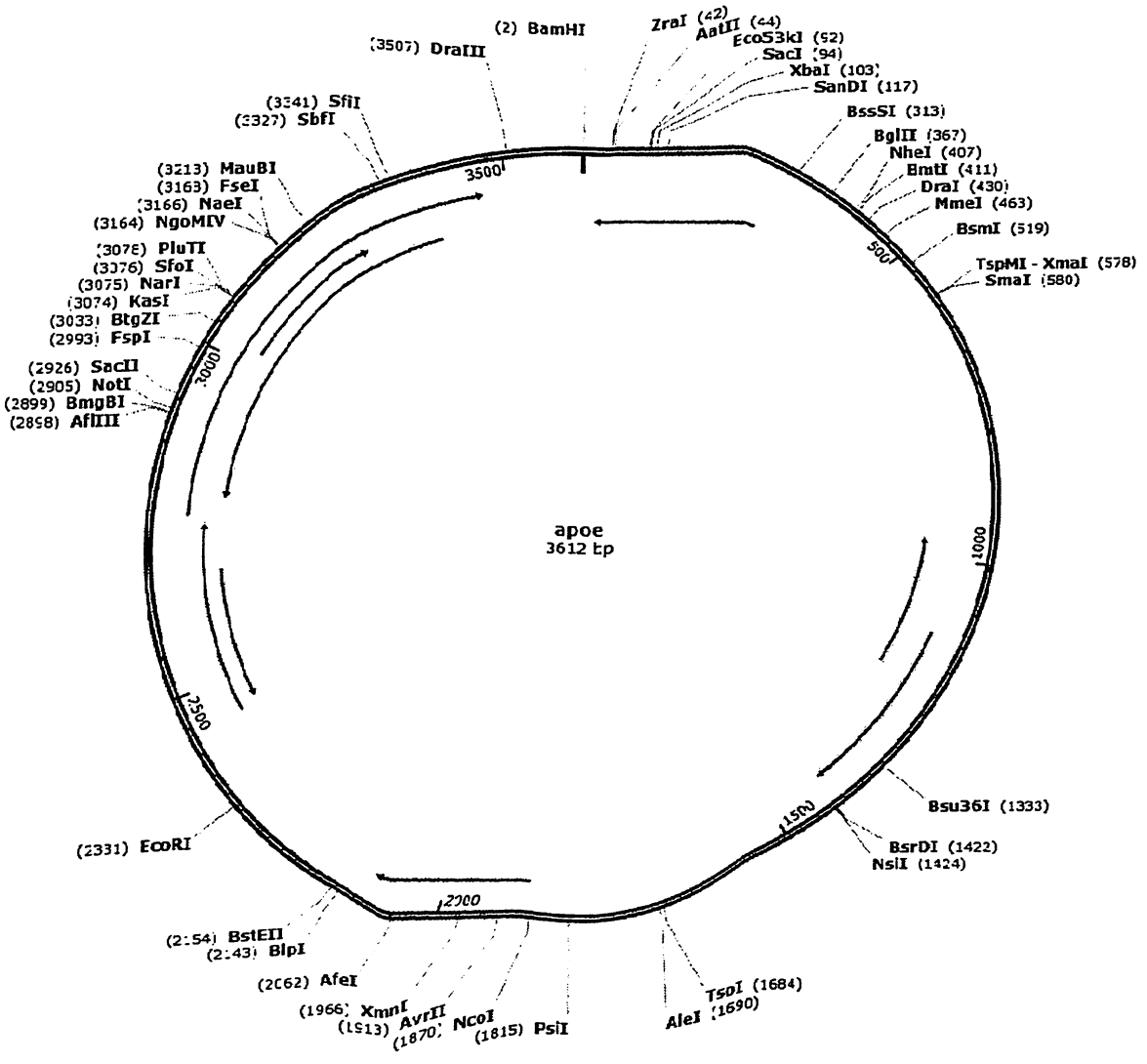


Figure-15 ***SNAP GENE view of APOE***

Snapgene 2.3.3 is a powerful molecular biology program that allows you to document in a rich electronic format any dna construct made in your lab. You can select the dna fragments that you wish to fuse and snapgene will design the primers. It also simplifies the planning of a gibson assembly reaction and automates the primer design, but here it is used only to view the restriction enzymes in the gene apoe, with its position mentioned to its site, here the apoe gene is shown in a circular manner, the is a option shown in this tool in which the linear form of the gene can also be viewed.it consists of nine open reading frame out of which yellow indicates the top three frames where as green indicates bottom two frames.

Table-2 Interactor, Experimental evidence code of APOE

Interactor	Experimental evidence code
A2m	Affinity capture western
A2m	Reconstituted complex
A2m	Two hybrid
Actg1	Two hybrid
Alb	Affinity capture ms
Ankh	Two hybrid
Apoe	Two hybrid
Arfgap1	Two hybrid
C190rf52	Two hybrid
Cdc37	Affinity capture western
Cdc37	Two hybrid
Csnk2a1	Biochemical activity
Cyp2c8	Two hybrid
Cyp2c18	Two hybrid
Ecsit	Affinity capture western
Ecsit	Two hybrid
Elavl1	Two hybrid
Epn2	Two hybrid
Farsa	Two hybrid
Fbxl12	Two hybrid
Foxg1	Two hybrid
Fxyd7	Two hybrid
Gcdh	Two hybrid
Htra1	Two hybrid
Ifit3	Two hybrid
Ifit5	Two hybrid
Iqsec1	Two hybrid
Lcat	Biochemical activity

Lonp1	Two hybrid
Lrp1	Affinity capture western
Lrp2	Reconstituted complex
Lrp8	Reconstituted complex
Lrp8	Reconstituted complex
Mapt	Protein peptide
Mapt	Reconstituted complex
Mast1	Two hybrid
Mid1ip1	Two hybrid
Nefm	Reconstituted complex
Nos3	Reconstituted complex
Nos3	Two hybrid
Pcmt1	Two hybrid
Pdcd4	Affinity capture western
Pdcd4	Two hybrid
Plekha6	Two hybrid
Pltp	Affinity capture western
Pram1	Two hybrid
Prdx2	Two hybrid
Psen1	Two hybrid
Rheb	Two hybrid
Rnf32	Two hybrid
Rpl4	Two hybrid
St13	Affinity capture western
St13	Two hybrid
Set	Reconstituted complex
Tyro3	Two hybrid
Ubc	Affinity capture western
Znf558	Two hybrid

Capture-western

An interaction is inferred when a bait protein is affinity captured from cell extracts by either polyclonal antibody or epitope tag and the associated interaction partner identified by western blot with a specific polyclonal antibody or second epitope tag. This category is also used if an interacting protein is visualized directly by dye stain or radioactivity.

Affinity capture-ms

An interaction is inferred when a bait protein is affinity captured from cell extracts by either polyclonal antibody or epitope tag and the associated interaction partner is identified by mass spectrometric methods.

Reconstituted complex

An interaction is detected between purified proteins in vitro.

Two-hybrid

Bait protein expressed as a dna binding domain (dbd) fusion and prey expressed as a transcriptional activation domain (tad) fusion and interaction measured by reporter gene activation

Protein-peptide

An interaction is detected between a protein and a peptide derived from an interaction partner. This includes phage display experiments.

Biochemical activity

An interaction is inferred from the biochemical effect of one protein upon another, for example, gtp-gdp exchange activity or phosphorylation of a substrate by a kinase. The bait protein executes the activity on the substrate hit protein. A modification value is recorded for interactions of this type with the possible values phosphorylation, ubiquitination, sumoylation, dephosphorylation, methylation, prenylation, acetylation, deubiquitination, proteolytic processing, glucosylation, deacetylation, no modification, demethylation

SUMMARY

CHAPTER-V

SUMMARY

Apoe gene is a complex gene to study as well as an intresting gene ,to work with, at the same time this gene is involved in various pathways and influence many other genes in causing mutations in different part of the body.this gene helps other disease to progress which leads into death of a person. With time this gene is evolving more strongly, no particular medication is there to cure the disease completely, may be to some extent proper diet could help in controlling the disease as APOE is associated with higher level of cholesterol and brain is the richest part of body consisting cholesterol.This gene is present in almost all organisms from fish to human i.e it shows from decades it has been conserved from despite speciation. It has variants which vary from each other structurally as well as functionally, thus causing different forms of diseases.It is not always that a person having APOE4 alleles will definitely be affected by AD, E4 increases the probability of the person to get AD,it may be possible that the persont may not get affected by AD.while a person having no APOE allele could be affected due to the other interaction of APOE.In india the chance of occurance ofalzheimer's is 2%-5% as compared to rest of the world, but still this disease is growing fastly. In USA around 1 in every 9 person is affected by alzheimer's.The APOE plays a very vital part in the functioning of the brain but,its mutation result in mental disturbance and leading to death.

REFERENCES

REFFERENCE

1. Hashimoto M, Rockenstein E, Crews L, masliah e (2003). "role of protein aggregation in mitochondrial dysfunction and neurodegeneration in alzheimer's and parkinson's diseases". *Neuromolecular med.* **4** (1–2): 21–36.doi:10.1385/nmm:4:1-2:21. PMID 14528050.
2. Kerr Ml, Small Dh (2005). "Cytoplasmic domain of the beta-amyloid protein precursor of alzheimer's disease: function, regulation of proteolysis, and implications for drug development". *J. Neurosci. Res.* **80** (2): 151–9. Doi:10.1002/jnr.20408.PMID 15672415.
3. Cai D, Netzer Wj, Zhong M, *et al.*; netzer; zhong; lin; du; frohman; foster; sisodia; xu; gorelick; greengard (2006)."presenilin-1 uses phospholipase d1 as a negative regulator of beta-amyloid formation". *Proc. Natl. Acad. Sci. U.s.a.***103** (6): 1941–6. Bibcode:2006pnas..103.1941c. Doi:10.1073/pnas.0510708103. Pmc 1413665.PMID 16449386.
4. Danielsson J, Andersson A, Jarvet J, Gräslund A (2006). "15n relaxation study of the amyloid beta-peptide: structural propensities and persistence length". *Magnetic resonance in chemistry: mrc.* 44 spec no: s114–21.doi:10.1002/mrc.1814. PMID 16826550.
5. Tomaselli S, Esposito V, Vangone P, *et al.* (2006). "the alpha-to-beta conformational transition of alzheimer's abeta-(1-42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of beta conformation seeding". *Chembiochem* **7** (2): 257–67. Doi:10.1002/cbic.200500223. PMID 16444756.
6. Ohnishi S, Takano K (2004). "Amyloid fibrils from the viewpoint of protein folding". *Cell. Mol. Life sci.* **61** (5): 511–24.doi:10.1007/s00018-003-3264-8. PMID 15004691.
7. Goedert M, Klug A, Crowther Ra (2006). "tau protein, the paired helical filament and alzheimer's disease". *J. Alzheimers dis.* **9** (3 suppl): 195–207. PMID 16914859.
8. Pastorino L, Sun A, Lu Pj, *et al.*; sun; lu; zhou; balastik; finn; wulf; lim; li; li; xia; nicholson; lu (2006). "The prolyl isomerase pin1 regulates amyloid precursor protein processing and amyloid-beta production". *Nature* **440** (7083): 528–34.Bibcode:2006natur.440..528p. Doi:10.1038/nature04543. PMID 16554819.
9. Lim J, Lu Kp (2005). "Pinning down phosphorylated tau and tauopathies". *Biochim. Biophys. Acta* **1739** (2–3): 311–22.doi:10.1016/j.bbadis.2004.10.003. PMID 15615648.

10. Akiyama H,Barger S,Barnum S,Bradt B,Bauer J, Cole Gm, *et al.* (2000). "inflammation and alzheimer's disease".*neurobiol aging* **21** (3): 383– 421. Doi:10.1016/s0197-4580(00)00124-x. PMID 10858586.
11. Swardfager W, Lanctôt K, Rothenburg L, Wong A, Cappell J, Herrmann N (2010). "a meta-analysis of cytokines in alzheimer's disease". *Biol psychiat* **68** (10): 930–941. Doi:10.1016/j.biopsych.2010.06.012. PMID 20692646.
12. Wenk Gl (2003). "neuropathologic changes in alzheimer's disease". *J clin psychiatry.* 64 suppl 9: 7–10.PMID 12934968.
13. Tiraboschi P, Hansen L, Thal L, Corey-Bloom J (2004). "the importance of neuritic plaques and tangles to the development and evolution of ad". *Neurology* **62** (11): 1984–9. Doi:10.1212/01.wnl.0000129697.01779.0a.PMID 15184601.
14. Bouras C, Hof Pr, Giannakopoulos P, Michel Jp, Morrison Jh (1994). "regional distribution of neurofibrillary tangles and senile plaques in the cerebral cortex of elderly patients: a quantitative evaluation of a one-year autopsy population from a geriatric hospital". *Cereb. Cortex* **4** (2): 138–50. Doi:10.1093/cercor/4.2.138. PMID 8038565.
15. Baskys. "receptor found that could lead to better treatments for stroke, alzheimer's disease". Uci medical center. Retrieved 2006-11-04.
16. Shen Zx (2004). "brain cholinesterases: ii. The molecular and cellular basis of alzheimer's disease". *Med. Hypotheses* **63**(2): 308 21. Doi:10.1016/j.mehy.2004.02.031. PMID 15236795.
17. Wenk Gl (2003). "neuropathologic changes in alzheimer's disease". *The journal of clinical psychiatry.* 64 suppl 9: 7–10.PMID 12934968.
18. Mudher A, Lovestone S (2002). "alzheimer's disease-do tauists and baptists finally shake hands?". *Trends neurosci.***25** (1): 22–6. Doi:10.1016/s0166-2236(00)02031-2. PMID 11801334.
19. Schmitz C, Rutten Bp, Pielen A, *et al.* (1 april 2004). "hippocampal neuron loss exceeds amyloid plaque load in a transgenic mouse model of alzheimer's disease". *Am. J. Pathol.* **164** (4): 1495–502. Doi:10.1016/s0002-9440(10)63235-x. Pmc 1615337. PMID 15039236.
20. Gray Eg, paula-barbosa m, roher a (1987). "alzheimer's disease: paired helical filaments and cytomembranes".*neuropathol. Appl. Neurobiol.* **13** (2): 91–110. Doi:10.1111/j.1365-2990.1987.tb00174.x. PMID 3614544.

CURRICULUM VITAE