

COMPARATIVE PHYLOGENETIC ANALYSIS OF OBESITY RELATED GENES AND ORTHOLOGS

Submitted in partial fulfilment of the requirements for the award of
Master of Science degree in Bioinformatics and Data Science
by

SHANTHI.R (Reg no. 42739022)



**DEPARTMENT OF BIOINFOMATICS AND DATA SCIENCE
SCHOOL OF SCIENCE AND HUMANITIES**

SATHYABAMA

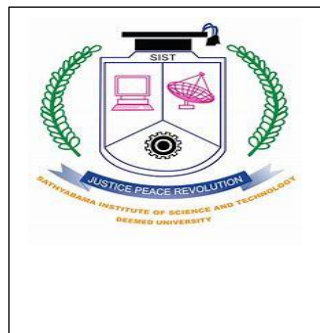
**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A++" by NAAC | 12B Status by UGC

Approved by AICTE

JEPPIAR NAGAR, RAJIV GANDHI SALAI, CHENNAI-600119

April 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A++" grade by NAAC I 12B Status by UGC 1 Approved by

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai-600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

BONAFIDE CERTIFICATE

This is to certify that the Project Report is the bonafide work of SHANTHI.R(42739022) who carried out the project entitled 'COMPARATIVE PHYLOGENETIC ANALYSIS OF OBESITY RELATED GENES AND ORTHOLOGS' under my supervision from December 2023 to April 2024.

Internal guide

DR. SWETHA SUNKAR

Head of the Department
Dr JEMMYCHRISTY, M Tech., PhD.

Submitted for Viva voce examination held on 20/04/2024

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to Dr. REKHA CHAKRAVARTHI M.E.,Ph.D.,Dean, School of Science and Humanities and Dr. JEMMY CHRISTY M.TECH.,Ph.D.,Head of the Department of Bioinformatics for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr. SWETHA SUNKAR(Internal) for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

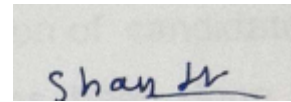
I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project

DECLARATION

I **SHANTHI.R (42739022)** hereby declare that the Project Report entitled "**COMPARATIVE PHYLOGENETIC ANALYSIS OF OBESITY RELATED GENES AND ORTHOLOGS**" done by me under the guidance of **Dr. SWETHA SUNKAR** at **Sathyabama Institute of Science and Technology** is submitted in partial fulfilment of the requirements for the award of Master of Science degree in BIOINFORMATICS AND DATA SCIENCE

DATE: 20/04/2024

PLACE: CHENNAI

A rectangular box containing a handwritten signature in blue ink. The signature appears to be 'Shanthi R' with a stylized flourish at the end.

SIGNATURE OF CANDIDATE

Abstract

Obesity is a complex, multifactorial condition characterized by excess body fat accumulation, leading to adverse effects on health. It is a major public health concern worldwide, with increasing prevalence rates across all age groups and socioeconomic backgrounds. There is a need for new solution to the problem of ever-increasing rates and wide prevalence of obesity. Phylogenetic analysis methods have been used to understand the causes of obesity and the mechanisms behind obesity occurrence. There has been a dearth of studies on evolutionary aspects of obesity and phylogenetic studies on obesity genes and pathways. It has been attempted to identify the significant genes and metabolic pathways associated with obesity. Identification of candidate genes and orthologs in these metabolic pathways has been done using NCBI BLAST and they have been analysed to determine phylogenetic relationships, evolutionary analysis of regulatory sequences and protein sequence similarity using tools such as Phylogeny.fr, EARS tool Warwick School of Biology and Uniprot Align. Phylogenetic analysis studies, expression analysis studies, studies on conserved regulatory sequences and protein sequence alignment and conservation studies have been done. It has been understood from this research that the genes and orthologs have similar functions, they have evolved from a common ancestor and have a strong shared phylogenetic relationship. Conserved sequences have been analysed and the presence of conserved non coding sequences has been observed. Protein sequences have been analysed and significant similarity has been found among the sequences. It has been concluded that the genes and orthologs from obesity related pathways have a strong phylogenetic relationship, have presence of conserved regulatory sequences and their proteins show highly similarity.

Keywords: Obesity, Orthologous genes, metabolic pathways, sequence alignment, phylogeny studies, Conserved sequences, Phylogenetic analysis, regulatory sequences.

TABLE OF CONTENTS

Sl no	Description	Page no.
	ABSTRACT	i
	LIST OF ABBREVIATIONS	ii
	LIST OF TABLES	iii
	LIST OF FIGURES	iv
1	CHAPTER 1: INTRODUCTION	1
1.1	BACKGROUND	1
1.2	OBESITY – HIGHLY PREVALENCE SYNDROME	1
1.3	SYMPTOMS OF OBESITY	2
1.4	SERIOUS IMPLICATIONS OF OBESITY	2
1.5	NEED FOR STUDIES ON OBESITY	2
1.6	ASSOCIATION OF OBESITY WITH TYPE II DIABETES AND CARDIAC DISEASE	2
1.7	IDENTIFYING SIGNIFICANT OBESITY RELATED PATHWAYS AND GENES	3
1.8	PHYLOGENETIC STUDIES	3
1.9	EVOLUTIONARY STUDIES	3
2	CHAPTER 2: REVIEW OF LITERATURE	5
3	CHAPTER 3: AIM AND OBJECTIVES	14
3.1	AIM	14
3.2	OBJECTIVES	14
4	CHAPTER 4: METHODOLOGY	15
4.1	SELECTION OF GENES	15
4.2	EXPRESSION ANALYSIS OF SELECTED GENES	15
4.3	IDENTIFICATION OF ORTHOLOGOUS SEQUENCES	17

4.4	PHYLOGENETIC ANALYSIS OF GENES AND ORTHOLOGS	18
4.5	EVOLUTIONARY ANALYSIS OF REGULATORY SEQUENCES	19
4.6	MULTIPLE SEQUENCE ALIGNMENT	19
5	CHAPTER 5 RESULTS AND DISCUSSION	21
6	CHAPTER 6 : SUMMARY AND CONCLUSION	149
7	CHAPTER 7: REFERENCES	153
8	CHAPTER 8: APPENDIX	160

LIST OF ABBREVIATIONS

SI No.	Abbreviation	Expansion
1.	PUBMED :	Publications Medline
2.	ENSEMBL:	Ensembl Genome Database Project
3.	NCBI:	National Centre for Biotechnology Information
4.	BLAST:	Basic Local Alignment Search Tool
5.	G BLOCKS	Genome Blocks
6.	MUSCLE	Multiple Sequence Comparison by Log Expectation
7.	EARS	Evolutionary Analysis of Regulatory Sequences
8.	UNIPROT	Universal Protein Resource
9.	ACADM	Acyl-CoA Dehydrogenase Medium Chain
10.	CPT1A	Carnitine Palmitoyltransferase 1A
11.	DGAT1	Diacylglycerol O-Acyltransferase 1
12.	SLC25A1	Solute Carrier Family 25 Member 1.
13.	HADH:	Hydroxyacyl-CoA Dehydrogenase
14.	PKM	Pyruvate Kinase M1/2
15.	ACLY:	ATP Citrate Lyase
16.	ACACA	acetyl-CoA carboxylase alpha
17.	MLXIPL:	MLX interacting protein like
18.	SCD	stearoyl -CoA desaturase
19.	GHSR	growth hormone secretagogue receptor neuropeptide Y
20.	AGRP	agouti related neuropeptide
21.	FOXO1	Forkhead box O1
22.	ADCY3:	adenylate cyclase 3
23.	APLN	Apelin
24.	APLNR	apelin receptor
25.	MAPK1:	mitogen-activated protein kinase 1

- | | | |
|-----|-------|----------------------------|
| 26. | KLF5: | KLF transcription factor 5 |
| 27. | MC3R: | melanocortin 3 receptor |

LIST OF TABLES

Sl no	Table no	Table name
1.	TABLE 1	Expression values of genes
2.	TABLE 2.1	Blast analysis of ACADM gene and orthologs
3.	TABLE 2.2	Blast analysis of CPT1A gene and orthologs
4.	TABLE 2.3:	Blast analysis of DGAT1 gene and orthologs
5.	TABLE 2.4:	Blast analysis of SLC5A1 gene and orthologs
6.	TABLE 2.5	Blast analysis of HADH gene and orthologs
7.	TABLE 2.6	Blast analysis of PKM gene and orthologs
8.	TABLE 2.7	Blast analysis of ACLY gene and orthologs
9.	TABLE 2.8:	Blast analysis of ACACA gene and orthologs
10.	TABLE 2.9	Blast analysis of MLXIPL gene and orthologs
11.	TABLE 2.10	Blast analysis of SCD gene and orthologs
12.	TABLE 2.11	Blast analysis of GHSR gene and orthologs
13.	TABLE 2.12	Blast analysis of NPY gene and orthologs
14.	TABLE 2.13	Blast analysis of AGRP gene and orthologs
15.	TABLE 2.14	Blast analysis of F0X01 gene and orthologs
16.	TABLE 2.15	: Blast analysis of ADCY3 gene and orthologs
17.	TABLE 2.16	Blast analysis of APLN gene and orthologs
18.	TABLE 2.17	Blast analysis of APLNR gene and orthologs
19.	TBLE 2.18	Blast analysis of MAPK1 gene and orthologs
20.	TABLE 2.19	Blast analysis of KLF5 gene and orthologs
21.	TABLE 2.20	Blast analysis of MC3R gene and orthologs

LIST OF FIGURES

Sl no.	Figure no	Figure name
1.	FIGURE 1.1	ACADM Gene Expression pictograph
2.	FIGURE 1.2	ACADM gene expression levels bar plot
3.	FIGURE 1.3	CPT1A Gene Expression pictograph
4.	FIGURE 1.4	CPT1A gene expression levels bar plot
5.	FIGURE 1.5	DGAT1 gene expression pictograph
6.	FIGURE 1.6	DGAT1 gene expression levels bar plot
7.	FIGURE 1.7	SLC25A1 gene expression pictograph
8.	FIGURE 1.8	SLC25A1 gene expression levels barplot
9.	FIGURE 1.9	HADH gene expression pictograph
10.	FIGURE 1.10	HADH gene expression levels barplot
11.	FIGURE 1.11	PKM gene expression pictograph
12.	FIGURE 1.12	PKM gene expression levels barplot
13.	FIGURE 1.13	ACLY gene expression pictograph
14.	FIGURE 1.14	ACLY gene expression levels barplot
15.	FIGURE 1.15	ACACA gene expression pictograph
16.	FIGURE 1.16	ACACA gene expression levels barplot
17.	FIGURE 1.17	MLXIPL gene expression pictograph
18.	FIGURE 1.18	MLXIPL gene expression levels barplot
19.	FIGURE 1.19	SCD gene expression pictograph
20.	FIGURE 1.20	SCD gene expression levels barplot
21.	FIGURE 1.21	GHSR gene expression pictograph
22.	FIGURE 1.22	GHSR gene expression levels barplot
23.	FIGURE 1.23	NPY gene expression pictograph
24.	FIGURE 1.24	NPY gene expression levels barplot
25.	FIGURE 1.25	AGRP gene expression pictograph
26.	FIGURE 1.25	AGRP gene expression levels barplot
27.	FIGURE 1.26	F0X01 gene expression pictograph
28.	FIGURE 1.27	F0X01 gene expression levels barplot
29	FIGURE 1.28	ADCY3 gene expression pictograph
30	FIGURE 1.29	ADCY3 gene expression levels barplot
31	FIGURE 1.30	APLN gene expression pictograph

32	FIGURE 1.31	APLN gene expression levels barplot
33.	FIGURE 1.32	APLNR gene expression pictograph
34.	FIGURE 1.33	APLNR gene expression levels barplot
35.	FIGURE 1.34	MAPK1 gene expression pictograph
36.	FIGURE 1.35	MAPK1 gene expression levels barplot
37.	FIGURE 1.36	KLF5 gene expression pictograph
38.	FIGURE 1.37	KLF5 gene expression levels barplot
39.	FIGURE 1.38	MC3R gene expression pictograph
40.	FIGURE 1.39	MC3R gene expression levels barplot
41.	FIGURE 2.1	Phylogenetic tree ACADM
42.	FIGURE 2.2	Phylogenetic tree CPT1A
43.	FIGURE 2.3	Phylogenetic tree DGAT1
44.	FIGURE 2.4	Phylogenetic tree SLC25A1
45	FIGURE 2.5	Phylogenetic tree HADH
46	FIGURE 2.6	Phylogenetic tree PKM
47	FIGURE 2.7:	Phylogenetic tree ACLY
48	FIGURE 2.8	Phylogenetic tree ACACA
49	FIGURE 2.9:	Phylogenetic tree MLXIPL
50	FIGURE 2.10	Phylogenetic tree MLXIPL
51	FIGURE 2.11	Phylogenetic tree SCD
52	FIGURE 2.12	Phylogenetic tree GHSR
53	FIGURE 2.13	Phylogenetic tree NPY
54	FIGURE 2.14	Phylogenetic tree AGRP
55	FIGURE 2.15	Phylogenetic tree FOXO1
56	FIGURE 2.16	Phylogenetic tree ADCY3
57	FIGURE 2.17	Phylogenetic tree APLN
58	FIGURE 2.18	Phylogenetic tree APLNR
59	FIGURE 2.19	Phylogenetic tree MAPK1
60	FIGURE 2.20	Phylogenetic tree KLF5
61	FIGURE 2.21	Phylogenetic tree MC3R
62	FIGURE 3.1	Evolutionary analysis of regulatory sequences results ACADM

63	FIGURE 3.2	Evolutionary analysis of regulatory sequences results CPT1A
64	FIGURE 3.3	Evolutionary analysis of regulatory sequences results DGAT1
65.	FIGURE 3.4	Evolutionary analysis of regulatory sequences results SLC25A1
66.	FIGURE 3.5	Evolutionary analysis of regulatory sequences results HADH
67	FIGURE 3.6	Evolutionary analysis of regulatory sequences results PKM
68	FIGURE 3.7	Evolutionary analysis of regulatory sequences results ACLY
69	FIGURE 3.8	Evolutionary analysis of regulatory sequences results ACACA
70	FIGURE 3.9	Evolutionary analysis of regulatory sequences results MLXIPL
71.	FIGURE 3.10	Evolutionary analysis of regulatory sequences results SCD
72.	FIGURE 3.11	Evolutionary analysis of regulatory sequences results GHSR
73.	FIGURE 3.12	Evolutionary analysis of regulatory sequences results NPY
74.	FIGURE 3.13	Evolutionary analysis of regulatory sequences results AGRP
75.	FIGURE 3.14	Evolutionary analysis of regulatory sequences results FOXO1
76.	FIGURE 3.15	Evolutionary analysis of regulatory sequences results ADCY3
77.	FIGURE 3.16	Evolutionary analysis of regulatory sequences results APLN
78.	FIGURE 3.17	Evolutionary analysis of regulatory sequences results APLNR

79.	FIGURE 3.18	Evolutionary analysis of regulatory sequences results MAPK1
80	FIGURE 3.19	Evolutionary analysis of regulatory sequences results KLF5
81	FIGURE 3.20	Evolutionary analysis of regulatory sequences results MC3R
82.	FIGURE 4.1	Motif Identification ACADM
83	FIGURE 4.2	Motif Identification CPT1A
84	FIGURE 4.3	Motif Identification DGAT1
85	FIGURE 4.4	Motif identification SLC25A1
86	FIGURE 4.5	Motif identification HADH
87	FIGURE 4.6	Motif identification PKM
88	FIGURE 4.7	Motif identification ACLY
89	FIGURE 4.8	Motif identification ACACA
90	FIGURE 4.9	Motif identification MLXIPL
91	FIGURE 4.10	Motif identification SCD
92	FIGURE 4.11	Motif identification GHSR
93	FIGURE 4.12	Motif identification NPY
94	FIGURE 4.13	Motif identification AGRP
95	FIGURE 4.14	Motif identification FOXO1
96	FIGURE 4.15	Motif identification ADCY3
97	FIGURE 4.16	Motif identification APLN
98	FIGURE 4.17	Motif identification APLNR
99	FIGURE 4.18	Motif identification MAPK1
100	FIGURE 4.19	Motif identification KLF5
101	FIGURE 4.20	Motif identification MC3R
102	FIGURE 5.1	ACADM Percentage Identity Matrix
103	FIGURE 5.2	CPT1A Percentage Identity Matrix
104	FIGURE 5.3	DGAT1 Percentage Identity Matrix
105	FIGURE 5.4	SLC25A1 Percentage Identity Matrix
106	FIGURE 5.5	HADH Percentage Identity Matrix
107	FIGURE 5.6	PKM Percentage Identity Matrix
108	FIGURE 5.7	ACLY Percentage Identity Matrix

109	FIGURE 5.8	ACACA Percentage Identity Matrix
110	FIGURE 5.9	CHREBP Percentage Identity Matrix
111	FIGURE 5.10	SCD Percentage Identity Matrix
112	FIGURE 5.11	GHSR Percentage Identity Matrix
113	FIGURE 5.12	NPY Percentage Identity Matrix
114	FIGURE 5.13	AGRP Percentage Identity Matrix
115	FIGURE 5.14	FOXO1 Percentage Identity Matrix
116	FIGURE 5.15	ADCY3 Percentage Identity Matrix
117	FIGURE 5.16	APLN Percentage Identity Matrix
118	FIGURE 5.17	APLNR Percentage Identity Matrix
119.	FIGURE 5.18	MAPK1 Percentage Identity Matrix
120.	FIGURE 5.19	KLF5 Percentage Identity Matrix
121.	FIGURE 5.20	MC3R Percentage Identity Matrix

CHAPTER 1: INTRODUCTION

1.1Background

Phylogenetic analysis plays a pivotal role in understanding the evolutionary relationships between genes. By examining genetic divergence across different organisms researchers gain insights into the shared ancestry and functional conservation of specific genes. In Obesity research, analysing orthologs (genes in different species that share a common ancestor) provides valuable information on the genetic basis of obesity

Existing software tools such as InParanoid, OrthoMCL, OMA and Orthofinder aid in identifying orthologous genes and constructing phylogenetic trees. These tools utilise heuristic approaches based on sequence similarity scores to infer evolutionary relationships. By integrating phylogenetic insights, we can unravel the intricate genetic landscape underlying obesity and inform therapeutic strategies.

1.2Obesity-A highly prevalent syndrome

Obesity, a widespread concern impacting people of all ages, has surged in prevalence. A substantial portion of the population faces the challenges of carrying excess weight, resulting in health-related issues. Obesity manifests as an abnormal accumulation of body mass and is a chronic condition that evolves gradually. Individuals are classified as obese when their body mass index (BMI) surpasses 30 kg/m², determined by dividing an individual's weight by the square of their height (Yeo et al., 2022).

1.3 Symptoms of obesity

Excessive body fat significantly elevates the likelihood of severe health issues. Individuals struggling with obesity may encounter symptoms such as weight gain, binge eating, fatigue, abdominal bulge, and snoring.

1.4 Serious implications of obesity

Carrying excessive fat has far-reaching health implications, including conditions like cardiovascular disease (particularly heart disease and stroke), type 2 diabetes, and musculoskeletal disorders like osteoarthritis. Additionally, certain cancers (such as endometrial, breast, and colon cancers) are more prevalent among those with obesity compared to individuals maintaining a healthy weight(Jia et al.,2021).

1.5 Need for studies on obesity

There is a need for studies on obesity causes and the evolution of obesity related genes. This would help us understand the changes in the gene functions or genes that could be the cause of obesity.

1.6 Association of obesity with type 2 diabetes and cardiac disease

Obesity is closely linked to the development of type 2 diabetes. The majority of diagnosed diabetes cases in adults fall into this category. Individuals with obesity often experience insulin resistance, which contributes to elevated blood sugar levels. This metabolic imbalance increases the risk of cardiovascular diseases including heart attacks, strokes and heart failure. The combination of obesity, diabetes and hypertension further exacerbates vascular damage, leading to atherosclerosis and compromised heart function. Innovative

approaches are needed to address this complex interplay and patient outcomes(Oliver at al.,2023).

1.7 Identifying Significant obesity related pathways

It is important to identify the metabolic pathways in the body that are significantly associated with obesity.

1.8 Phylogenetic Studies

Phylogenetic studies have so far been done with tools like MEGA, Philip (Phylogeny Inference Packages), Bio Numeric, APE and BEAST (Bayesian Evolutionary Sampling Analysis test). Phylogenetic studies are very important to understand obesity causes. It helps us understand the evolutionary history of obesity related genes. By performing phylogenetic analysis we can understand the functions and the evolutionary relationships of the genes.

1.9 Evolutionary Studies

Analysis of regulatory sequences is another important aspect. While phylogenetic studies analyse the coding regions of the genes, the analysis of regulatory sequences helps us study the non coding regions having regulatory function Analysis of regulatory regions is important to understand the conserved non coding regions.

Understanding the Phylogeny and evolutionary analysis of non coding regions is key to understanding obesity. Currently only studies relating to differential gene expression analysis, biomarker ,ML techniques and diffraction tensor related studies have been performed.

Phylogeny studies and evolutionary analysis of the regulatory sequences along with protein sequence alignment analysis will help

solve the problem and understand the obesity mechanisms better. It will help control the rate of obesity(Zaru at al.,2023)

By studying the evolution of genes in obesity related pathways in the humans and primates, an overview and insights can be obtained on the genetic basis of obesity in humans.

CHAPTER 2 REVIEW OF LITERATURE

Obesity is a widespread health crisis, with prevalence nearly tripling since 1975. It is a leading risk factor for non-communicable diseases like type 2 diabetes and cardiovascular diseases. Understanding its origins is crucial for effective prevention and treatment. Phylogenetic analysis provides an avenue for elucidating the evolutionary lineage of genes associated with obesity. By comparing human genes with those in other species, researchers can identify conserved pathways in energy metabolism and fat storage, shedding light on evolutionary pressures underlying obesity. This review synthesizes current knowledge on phylogenetic analysis of obesity-related genes, focusing on conserved pathways across species. Integrating findings from phylogenetics, genomics, and obesity research could lead to novel therapeutic targets and strategies in obesity management.

Bioinformatics tools like UniProt, Clustal Omega, Phylogeny.fr, and Ensembl have been crucial in identifying and aligning sequences for phylogenetic analysis. They have enabled precise identification of genes and pathways involved in obesity, emphasizing genetic and evolutionary mechanisms. Phylogenetic analysis has the potential to uncover evolutionarily conserved genes and pathways contributing to obesity, offering new perspectives on this complex condition. The prevalence of obesity had been nearly tripling since 1975. It is a leading risk factor for diseases like type 2 diabetes and cardiovascular diseases, understanding its origins is crucial for effective prevention and statement.

By comparing human genes with those in other species, researchers can identify conserved pathways in energy metabolism and fat storage, shedding light on evolutionary pressures underlying obesity. This review

synthesizes current knowledge on phylogenetic analysis of obesity-related genes, focusing on conserved pathways across species. Integrating findings from phylogenetics, genomics, and obesity research could lead to novel therapeutic targets and strategies in obesity management.

Bioinformatics tools like UniProt, Clustal Omega, Phylogeny.fr, and Ensembl have been crucial in identifying and aligning sequences for phylogenetic analysis. They have enabled precise identification of genes and pathways involved in obesity, emphasizing genetic and evolutionary mechanisms. Phylogenetic analysis has the potential to uncover evolutionarily conserved genes and pathways contributing to obesity, offering new perspectives on this complex condition.

UNIPROT TOOLS

Zaru et al.;2023; Pundir et al., 2016) described how UniProt provides an essential suite of tools for the bioinformatics community, including BLAST for sequence similarity searching, Align for multiple sequence alignments, Peptide Search for identifying peptide sequences, and ID Mapping for cross-referencing database identifiers. Zaru et al. (2023) highlighted the continuous improvement in these tools, ensuring they remain relevant for current bioinformatics challenges. The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data (UniProt Consortium, 2023). The UniProt website receives about 800,000 unique visitors per month and is the primary means to access UniProt. Along with various datasets that you can search, UniProt provides four main tools. These are the “BLAST” tool for sequence similarity searching, the “Align” tool for multiple sequence alignment, the “Peptide Search” tool for retrieving proteins containing a short peptide sequence, and the “Retrieve/ID Mapping” tool

for using a list of identifiers to retrieve UniProt Knowledgebase (UniProtKB) proteins and to convert database identifiers from UniProt to external databases or vice versa.

ENSEMBL

Harrison et al., 2024; Martin et al., 2023) described how Ensembl databases offer comprehensive information on genomic data, including annotated genes, variants, and comparative genomics. Harrison et al. (2024) and Martin et al. (2023) demonstrated Ensembl's commitment to providing up-to-date resources that reflect the latest in genomic research. Ensembl (<https://www.ensembl.org>) has produced high quality genomic resources for vertebrates and model organisms for more than 20 years. During that time, their resources, services and tools have continually evolved in line both with the publicly available genome data and the downstream research and applications that utilise the ensemble platform.[4] It is their goal at Ensembl to create high quality annotations, tools and services for species across the tree of life. They had introduced new features for annotations (including haplotypes from the first human pangenome graphs), updates to the Ensembl Variant Effect Predictor (VEP), interactive protein structure predictions from AlphaFold DB and the release of their new server.

PHYLOGENY.FR

Dereeper et al(2008) explains how Phylogeny.fr tool simplifies phylogenetic analysis for non-specialists, integrating different tools to perform robust analyses. It is designed to make phylogenetic studies accessible to a broader range of researchers. Phylogenetic analyses are important in many areas of research in biology. They involve the identification of homologous sequences, their multiple alignment, the

phylogenetic tree construction and graphical representation of the re-inferred tree. The Phylogeny.fr platform transparently chains programs to automatically perform these tasks. It is primarily designed for biologists with no experience in phylogeny, but can also meet the need of specialists. PhyPhylogeny.fr offers three main modes. The 'One Click' mode targets non specialists and provides a ready to use pipeline chaining programs with recognised accuracy and speed: MUSCLE for multiple alignment, PHYML for tree building and TREEDYN for tree rendering. All parameters are set up to suit most studies and users have to provide only their input sequences to obtain a ready to print tree. The 'Advanced' mode uses the same pipeline but allows the parameters of each program to be customised by the users. The 'A la Carte' mode offers more flexibility and as users can build their own pipeline by selecting and setting up the required steps from a large choice of tools to suit their specific needs. Prior to phylogenetic analysis, users can also collect neighbours of a query sequence by running BLAST on general or specialised databases. A guide tree then helps to select neighbor sequences of a query sequence by running BLAST on general or specialised databases. A guide tree then helps to select neighbor sequences to be used as input for the phylogeny pipeline. Phylogeny.fr is available at <http://www.phylogeny.fr/>

Domán et al. (2022) emphasizes the principles and practices in molecular phylogenetics, showcasing its application in understanding evolutionary relationships among species, such as the analysis of *Candida krusei*. Revealing the phylogenetic relationships of *Candida krusei* strains (sexual form *Pichia kudriavzevii*) is a prerequisite for understanding the evolution of its virulence-associated mechanisms and ecological lifestyles. Molecular phylogenetic analyses based on

entire internal transcribed spacer region (ITS) and multilocus sequence typing (MLST) data were carried out with sequences available in public databases and Hungarian isolates from animals obtained for the study. The ITS haplotype network yielded a high frequency haplotype at the centre of the network (H1; n = 204) indicating that various selective pressure might resulted in population expansion from H1. MLST analysis identified three new genotypes among animal-derived isolates, therefore overall, 203 sequence types were investigated to determine the population structure of *C. krusei*. The most commonly encountered sequence types were ST 17 and ST 67. Phylogenetic analyses showed diverse genetic construction of *C. krusei* population. Evidence of potential recombination events were also observed that might play some role in high intraspecies genetic variability among strains, however, the limited data of *C. krusei* genotypes from different countries prevented us to identify accurate evolutionary routes of commensal and pathogenic strains or species-specific lineages. Further expansion of *C. krusei* MLST database may promote the better understanding of the mixed evolutionary history of this species. Phylogenies have permeated nearly every branch of biology, and the plethora of phylogenetic methods and software packages that are now available may seem daunting to an experimental biologist. The major methods of phylogenetic analysis, including parsimony, distance, likelihood and Bayesian methods have been reviewed.

Dong et al(2021) described Anomopoda as the most widespread planktonic microcrustacean, which plays a crucial role in aquatic ecosystem. There are few studies about the evolutionary relationships among various Anomopoda basing on molecular data. In the present study, phylogenetic analysis of eight Anomopoda was carried out.

Firstly, the culture system was developed to breed cladocerans. By using this system, eight species (*Daphnia magna*, *D. pulex*, *D. sinensis*, *Ceriodaphnia reticulata*, *Moina micrura*, *Scapholeberis kingi*, *Simocephalus vetulus* and *Eurycercus lamellatus*) were purified and cultured stably in the laboratory. Then, transcriptomic sequences and partial mitochondrial DNA sequences were both used to reconstruct the phylogenetic tree among 8 species. Transcriptomic sequences were sequenced on Illumina Hiseq 2500 platform. After assembly and annotation, transcriptomic sequences were spliced together and aligned for phylogenetic analysis. Basing on the orthologous genes derived from transcriptomic sequences, the phylogenetic analysis showed that 4 genera of Daphniidae were clustered into one group, and among the 4 genera, *Ceriodaphnia* was closer to *Daphnia* than *Simocephalus*, while *Scapholeberis* was farthest from other species. In addition, *Eurycercidae* was closer to Daphniidae than Moinidae. The phylogenetic trees based on both 12S rRNA and 16S rRNA sequences were similar with that based on transcriptomic sequences. Meanwhile, the phylogenetic tree based on 16S rRNA sequences was more suitable than that based on 12S rRNA sequences. These results suggested that the phylogenetic analysis basing on the transcriptomic sequences was available in cladocerans, which will help us to effectively understand the phylogenetic relationships among various cladocerans.

EARS IN PLANTS

Picot et al.(2023) introduced the EARS (Evolutionary Analysis of Regulatory Sequences) method for identifying evolutionarily conserved regions within non-coding DNA. These regions are crucial for gene regulation, but their identification remains challenging due to the limited

fraction of non-coding DNA contributing to regulation. Traditional methods involve laborious promoter deletion series construction, while bioinformatic approaches face limitations. EARS offers a robust and sensitive in silico solution, detecting conserved sequences suggestive of functional importance. Tested on known promoters, EARS successfully identifies regulatory regions containing transcription binding sites. This method facilitates automated analysis of large datasets and enhances our understanding of regulatory elements in plant Clustal Omega tool related studies and articles introduce Clustal Omega, a tool for fast and scalable generation of high-quality protein multiple sequence alignments. Its accuracy and efficiency in handling large datasets make it a cornerstone in evolutionary analysis. Multiple sequence alignments are fundamental to many sequence analysis methods. Most alignments are computed using the progressive alignment heuristic.

PHYLOGENETIC ANALYSIS OF BACTERIAL PATHOGEN GENOMES

Didelot et al. (2023) described how advancements in high-throughput sequencing have made it faster and more cost-effective to sequence entire bacterial pathogen genomes. Researchers can now compare hundreds or even thousands of genomes within a bacterial population. Phylogenetic trees are commonly used to show relationships between these genomes. However, the presence of homologous recombination in many bacterial species can disrupt the accuracy of standard phylogenetic tools. To address this, a method was developed that considered the impact of recombination. This approach allowed users to explore recombination events and generate more meaningful phylogenetic analyses that capture the clonal genealogy and relationships between genomes.

FOREST BASED NETWORKS

Huber et al. (2022) described how in the field of evolutionary studies, phylogenetic trees are commonly used to depict the evolutionary history of species. However, when genetic information transfer (such as gene exchange) occurs between species or their ancestors, a simple tree may not fully capture their complex history. To address this limitation, researchers have turned to tree-based phylogenetic networks.

COMPREHENSIVE IDENTIFICATION OF ASCOVIRIDAE FAMILY

Shi et al. (2022) conducted an extensive analysis of gene orthology within the Ascoviridae family. They identified 366 protein homologue groups and analyzed 34 non-single-copy proteins. The study revealed 90 newly annotated proteins, five core proteins specific to the Ascoviridae family, and 14 core proteins specific to the genus Ascovirus. Additionally, a phylogenomic tree was constructed for 11 Ascoviridae members based on 35 ortholog groups. By combining phosphoproteomic data, 30 conserved phosphorylation sites were identified on 17 phosphoproteins from *Heliothis virescens* ascovirus 3h (HvAV-3h). These findings offer potential research targets for investigating viral infection regulation. Overall, this study enhances genome annotation and facilitates comparisons among Ascoviridae members and functional genomic studies.

Despite these advancements, challenges remain in fully deciphering the complex genetic landscape of obesity. The reviewed literature points to the need for further refinement of bioinformatics tools and databases to enhance their applicability to obesity research. Additionally, the integration of data from diverse sources and the functional validation of

identified genes and pathways represent critical steps forward. Future research should focus on harnessing the full potential of these bioinformatics resources to unravel the multifaceted genetic influences on obesity. This includes expanding phylogenetic analyses to include a wider array of species to gain evolutionary insights, as well as integrating genomic data with functional and phenotypic data to elucidate the mechanisms by which identified genes and pathways influence obesity. In conclusion, this literature review has highlighted the pivotal role of bioinformatics tools and databases in advancing our understanding of the genetic and evolutionary aspects of obesity. By leveraging these resources, researchers can continue to uncover the intricate genetic networks that underlie obesity, paving the way for novel therapeutic targets and interventions. The ongoing development and refinement of these tools will be crucial in meeting the challenges of obesity research and in contributing to the broader field of genomics and evolutionary biology.

CHAPTER 3 AIM AND OBJECTIVES

3.1 Aim

The aim of this project is to explore the phylogenetic relationships and functional implications of orthologous genes related to obesity across orthologous species.

3.2 Objectives

- To select suitable genes from obesity related pathways and identification of their orthologs for analysis.
- To perform Phylogenetic analysis of genes and orthologous sequences
- To identify evolutionary conserved regulatory sequences
- To perform protein alignment and identify regions of similarity and other regions of interest.

CHAPTER 4 : METHODOLOGY

4.1 Selection of genes

In the first step, significant metabolic pathways in the body associated with obesity are identified. These pathways must be significantly associated with obesity and need to occur in human and orthologs. Literature studies are conducted to identify several obesity related metabolic pathways and the most significant pathways are selected for analysis. From these pathways most significant obesity related genes are finalised for analysis.

4.2 Expression analysis of selected genes

Expression analysis of the selected genes was done using Ensembl database. The expression levels of the genes in the different parts of the body were analysed. The data was visualised using a bar plot constructed by R Studio.

```
library(ggplot2)

data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\expressionLEP.csv')

head(data)

ggplot(data, aes(x=categories, y=values, fill=categories)) +
  geom_bar(stat="identity", color="black") +
  theme_minimal() +
  labs(x="Category", y="Value", title="Bar Plot of Categories") +
  scale_y_continuous(labels = scales::comma) +
```

```

scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a",
"#33a02c",
"#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
"#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
"#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
"#ffffb3", "#bebada", "#fb8072", "#80b1d3"))

```

Code explanation

This R code uses the `ggplot2` library to create a bar plot from a dataset loaded from a CSV file. Here's a breakdown of the code:

1. ****Load the ggplot2 library:**** `library(ggplot2)` loads the ggplot2 library, which is used for data visualization.
2. ****Read the CSV file:**** `data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\expressionLEP.csv')` reads the CSV file located at the specified path into a data frame called `data`.
3. ****Inspect the data:**** `head(data)` displays the first few rows of the `data` data frame, allowing you to check the structure and contents of the data.
4. ****Create the bar plot:**** `ggplot(data, aes(x=categories, y=values, fill=categories))` initializes a ggplot object using the `data` data frame and specifies that the x-axis should represent the "categories" column, the y-axis should represent the "values" column, and the bars should be filled based on the "categories" column.

5. ****Add the bar geometry:**** ``geom_bar(stat="identity", color="black")`` adds the bar geometry to the plot, where ``stat="identity"`` indicates that the y-values should be used directly to determine the height of the bars, and ``color="black"`` sets the color of the bar outlines to black.
6. ****Adjust the plot theme:**** ``theme_minimal()`` sets the plot theme to a minimalistic style.
7. ****Add axis labels and title:** `**`labs(x="Category", y="Value", title="Bar Plot of Categories")`` adds labels to the x-axis, y-axis, and a title to the plot.
8. ****Format the y-axis labels:** `**`scale_y_continuous(labels = scales::comma)`` formats the y-axis labels to use commas for thousands separators.
9. ****Specify custom fill colors:** `**`scale_fill_manual(values=...)`` specifies custom fill colors for the bars using a vector of hexadecimal colour codes.

Overall, this code creates a bar plot of the "values" in different "categories" from the loaded CSV file, using custom fill colours and a minimalistic theme.

4.3 Identification of Orthologous sequences

With the selected significant genes from obesity related pathways, discovery of orthologous sequences was done using NCBI nucleotide BLAST. The BLAST was performed with default parameters. The descriptions file containing the alignment results of BLAST sequences was downloaded. The excel file was saved and the results of the

nucleotide BLAST were analysed. The fasta(complete) sequences were downloaded and used for further analysis. The resulting file contained the fasta sequences of human gene and the six orthologous sequences.

4.4 Phylogenetic analysis of genes and orthologs

Using the Fasta sequence files, Phylogenetic analysis of human genes and orthologs was performed using phylogeny.fr tool. The phylogeny.fr tool consisted of the following steps: Gblocks, MUSCLE alignment, phylogeny and Tree Dyn.

The Fasta sequences present in the file were input in phylogeny.fr for phylogenetic analysis and phylogenetic tree construction. The Fasta sequences of all the orthologous species were aligned using the bioinformatics tools (MUSCLE, Gblocks, PhyML, Tree Dyn) in Phylogeny.fr. (<https://www.phylogeny.fr/>) and then the Phylogenetic tree was constructed. In the Phylogeny.fr server, the One Click Mode was selected. The analysis was performed on the Phylogeny.fr platform and comprised four steps, namely, alignment, curation, phylogeny, and tree rendering.

First the sequences were aligned with MUSCLE using default settings. Second, the regions containing gaps and poorly aligned were removed with Gblocks using the default parameters. In the third step, the phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program. In the fourth step, graphical representation and edition of the phylogenetic tree were performed with Tree Dyn.

4.5 Evolutionary analysis of regulatory sequences

The promoter sequences of the selected human genes and orthologs were downloaded and analysed using EARS tool. The comparative analysis was performed with the FASTA sequences of orthologous genes with the help of EARS software (http://nero.wsbc.warwick.ac.uk/tools/user_case_form.php). A template sequence (human gene sequence) was selected, to which all other orthologous sequences were compared. The orthologous sequences were used as comparison sequences. The species names were entered in each case. A cut off *P*-value of 0.1 and window size of 60 bp were selected. The resulting EARS file for each pair of species (human and one ortholog species) was analysed individually and the location of significant peaks in the promoter region was detected for *Homo sapiens* gene and its orthologous sequence. Similarly, the cumulative EARS file was obtained for all the species. Identification of motifs was done.

4.6 Multiple sequence alignment

The protein sequences of different species were analysed and aligned using the Clustal Omega algorithm. The UniProt database (<https://www.uniprot.org/>) provides a multiple sequence alignment tool for proteins called “Align.” This tool runs the Clustal Omega program to find the similarity in the sequences being aligned. On the UniProt homepage, the Align tab in the toolbar was clicked to align the protein sequences with the Clustal Omega program. Protein sequences in

FASTA format were entered in the text area. The “Run Align” button was clicked to obtain the alignment results.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 SELECTION OF GENES

Based on the literature survey done for identification of metabolic pathways significantly associated to obesity and identification of the significantly associated obesity related genes in them, the following four pathways were identified as most significant pathways related to obesity namely Fatty acid Oxidation pathway, De novo lipogenesis, Ghrelin signalling pathway and Apelin signalling pathway. These pathways were selected based on extensive literature search and survey studies.

Fatty acid oxidation pathway

Serra et al(2013) emphasized the significance of fatty acid oxidation in obesity.

Denovo lipogenesis

Ameer et. al(2014) postulated that the dysregulation of DNL in the major lipogenic tissues of the human body is often observed in various metabolic anomalies including obesity, non-alcoholic fatty liver disease and metabolic syndrome. Denovo lipogenesis is a crucial pathway associated with obesity.

Ghrelin signalling pathway

Makris et al. (2017) postulated that Ghrelin promotes obesity via its overproduction, decreased suppression following meals. Ghrelin signalling pathway is another pathway significantly associated to obesity.

Apelin signalling pathway

Boucher et. al(2016) postulated that Apelin production is increased during adipocyte differentiation phase. Apelin signalling pathway is another pathway crucial in obesity.

ACADM

Meritt et al.(2019) explained that children with ACADM deficiency disorder MCAD are more prone to obesity.

CPT1A

Liang et al. (2023)described that CPT1A is the rate limiting enzyme in fatty acid β oxidation and its deficiency or abnormal regulation can result in diseases like metabolic disorders and various cancers.

DGAT1

Deng et. Al(2024)explained that DGAT1 plays a key role in the control of diabetes and obesity.

SLC25A1

Mosaoa et. al(2021) described about the mitochondrial citrate carrier SLC25A1 and the fundamental role of citrate in cancer inflammation and beyond.

HADH

Fang et.al (2022) postulated that HADH an essential gene involved in fatty acid oxidation plays a crucial role in regulating energy metabolism.Its interaction with hypoxia inducible factor may be potential mechanism governing fatty acid oxidation.

PKM

Kang et al(2021) described how PDK1 AND PDK2 are novel regulators of adipogenesis that play crucial roles in obesity.

ACLY

Khwairakpam et. al(2020) described how ACLY expression correlated with the development and progressions of various chronic diseases such as neurodegenerative diseases, cardiovascular diseases, diabetes, obesity, inflammation and cancer.

ACACA

Wakil et al.(2009) described how ACC1 gene plays an important role in the energy metabolism of fatty acids in animals.

CHREBP

Eissing et. al(2013) described how de novo lipogenesis in human fat and liver is linked to chrebp and metabolic health.

SCD

Warensjoo et al.(2017) described how Obesity and Insulin resistance are major risk factors for metabolic diseases and are influenced by lifestyle and genetics. The lipogenic enzyme SCD is related to obesity.

GHSR

Baessier et al.(2015) described the genetic linkage and association of the growth hormone secretagogue receptor(GHSR) gene in human obesity.

NPY

Tang et. al(2020) described how higher NPY serum levels were positively correlated with MUO phenotype in obese Chinese adults.

AGRP

Beutler et al (2020) explained how obesity causes selective and long lasting desensitization of AGRP neurons to dietary fat.

F0X01

Gong et. al(2017) described the association of F0X01 and IRS2 gene polymorphisms with T2DM in Chinese Han population supporting F0X01-obesity interaction as a key factor for the risk of T2DM.

ADCY3

Saeed et. al(2018) highlights ADCY3 as an important mediator of energy homeostasis and an attractive pharmacological target in the treatment of obesity.

APLN

Wysocka et. al(2018) described that apelin might play an important role in obesity.

APLNR

Nowzari et. al(2018) postulated that APLNR,leptin and LEPR SNP's are believed to be candidate genes for CAD as well as hypertension because of their roles in regulation of lipid metabolism.

MAPK

Donohoe et. al(2020) described that MAPK is a significant gene associated with obesity and MAPK signalling pathways are activated in various cancers and obesity is a major risk factor for several types of cancer.

KLF5

Drosatos(2018)described how KLF5 is a very significant gene associated with obesity as it regulates the metabolism of cardiac fatty acids.

MC3R

Demidowich et al.(2017) described that obesity and fat mass percentage could be linked to markers in 20q13, within which is located the MC3R gene.

Table 1 Selected genes.

sl no.	Gene name	gene size (bp)	function or significance relevant to obesity
1	ACADM	8000	People with ACADM gene deficiency are more prone to obesity
2	CPT1A	12000	CPT1A is the rate limiting enzyme in fatty acid B oxidation and its deficiency or abnormal regulation can result in diseases like metabolic disorders
3	DGAT1	26000	DGAT1 plays a key role in the control of diabetes and obesity
4	SLC25A1	15000	SLC25A1 plays a key role in inflammation and other metabolic disorders
5	HADH	5000	HADH an essential gene involved in fatty acid oxidation plays a crucial role in regulating energy metabolism
6	PKM	17000	PKM is novel regulator of adipogenesis
7	ACLY	50000	ACLY expression correlated with obesity
8	ACACA	60000	ACACA gene plays an important role in energy metabolism of fatty acids
9	MLXIPL	20000	de novo lipogenesis in human fat and liver is linked to chrebp
10	SCD	25000	The lipogenic enzyme SCD is related to obesity
11	GHSR	10000	GhSR rate is high obesity is high
12	NPY	3000	Higher NPY level in serum associated with high rate of obesity in chinese adults
13	AGRP	8000	obesity causes selective and long lasting desensitization of AGRP neurons to dietary fat.
14	FOXO1	22000	association of FOXO1 and IRS2 gene polymorphisms with T2DM in Chinese Han population supporting FOXO1-obesity interaction as a key factor for the risk of T2DM.
15	ADCY3	25000	ADCY3 as an important mediator of energy homeostasis and an attractive pharmacological target in the treatment of obesity.
16	APLN	5000	apelin might play an important role in obesity.
17	APLNR	30000	lipid metabolism
18	MAPK1	70000	MAPK signalling pathways are associated with obesity
19	KLF5	15000	KLF5 is a very significant gene associated with obesity as it regulates the metabolism of cardiac fatty acids.
20	MC3R	6000	obesity and fat mass percentage could be linked to markers in 20q13, within which is located the MC3R gene.

5.2 EXPRESSION ANALYSIS OF SELECTED GENES

Based on the expression analysis results, it was observed that the expression levels of the genes were high in the parts of the body such as Aorta, liver, pancreas, thyroid gland, skeletal muscle tissue, oesophagus and atrium auricular region while the expression levels of the gene were low in the parts of the body such as amygdala, blood and lungs.

ACADM

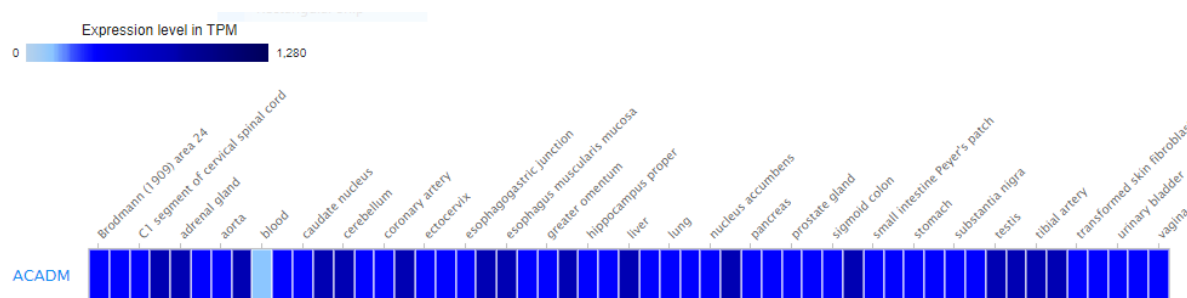


Figure 1.1: ACADM gene expression pictograph

The ACADM gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Run
Source

1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\ACADMexcel.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15

```

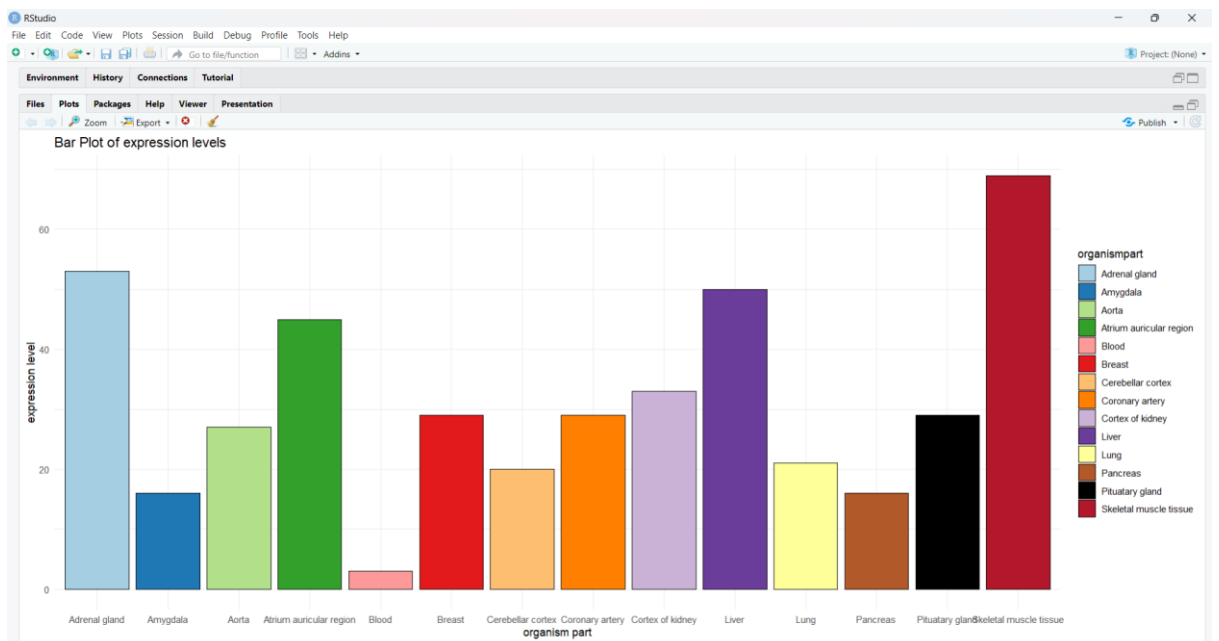


FIGURE 1.2:ACADM gene expression levels barplot

Barplot depicting the gene expression level of ACADM in the parts of the body X axis represents Organism part and Y axis represents Expression level

ACADM gene expression levels were highest in Adrenal gland,Atrium auricular region and liver. The ACADM gene expression levels were lowest in Blood, Lung and Amygdala.

CPT1A

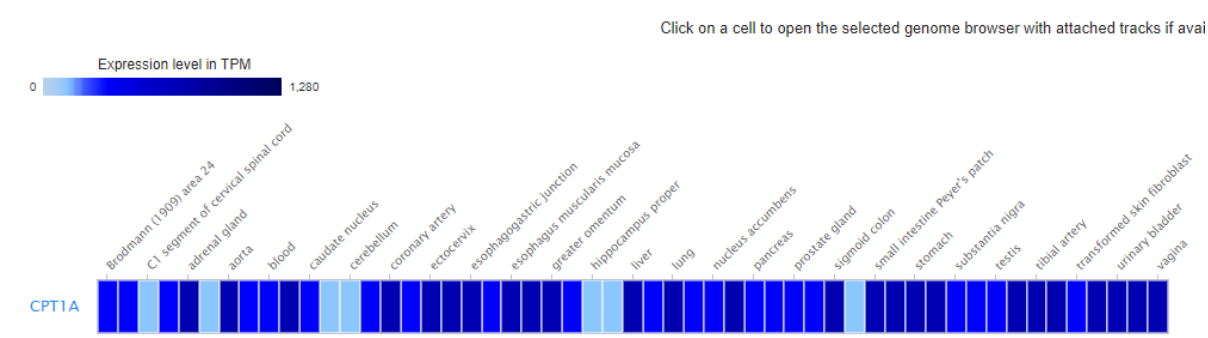


Figure 1.3: CPT1A gene expression pictograph

The CPT1A gene expression levels in the different parts of the body are depicted as values in TPM(Transcripts per Million) ranging from 0 TPM to 128 TPM.

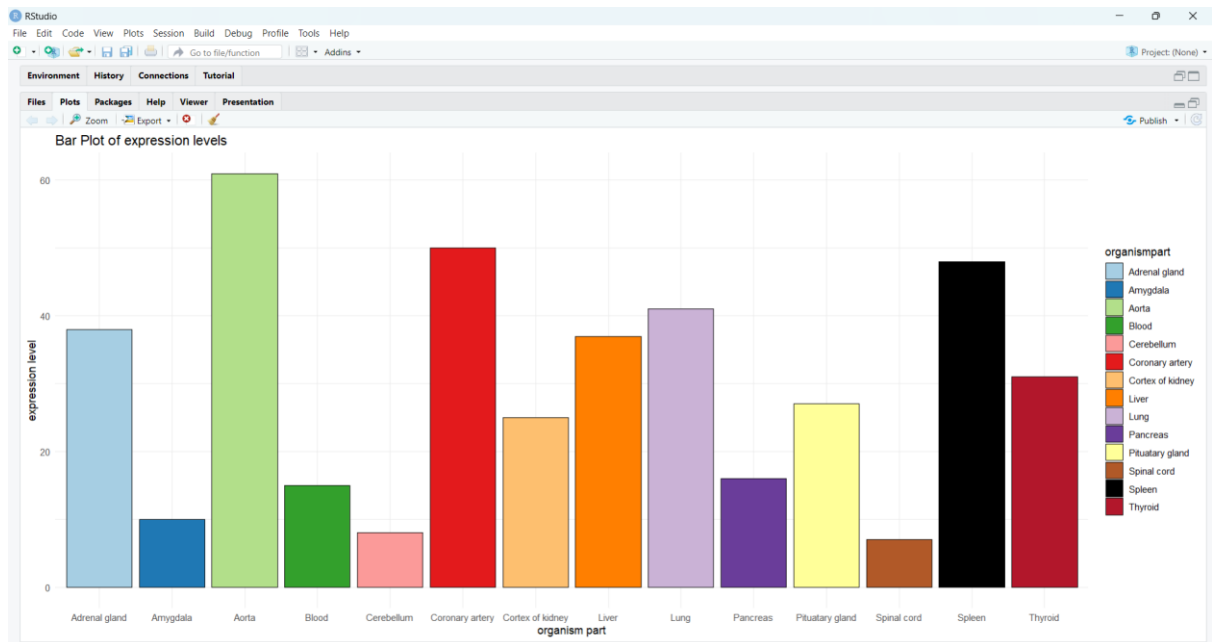


Figure 1.4 CPT1A gene expression levels barplot

Barplot depicting the gene expression level of CPT1A in the parts of the body X axis represents Organism part and Y axis represents Expression level

The CPT1A gene expression levels were highest in Adrenal gland, Aorta and Coronary Artery. The CPT1A gene expression levels were lowest in spinal cord, cerebellum and amygdala.

DGAT1

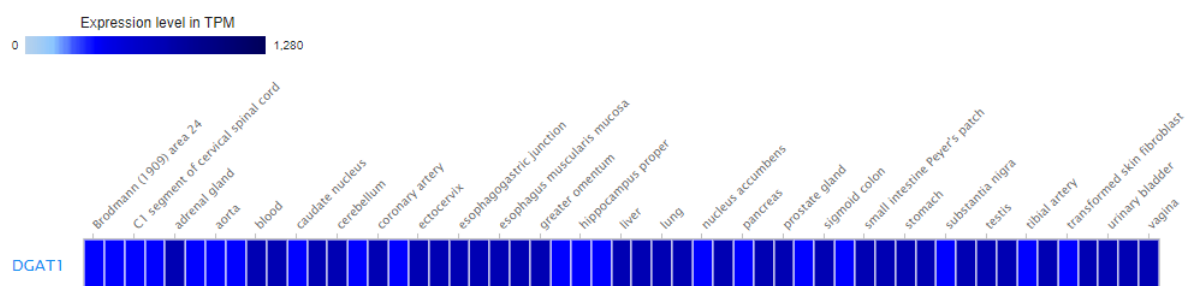


FIGURE 1.5: DGAT1 gene expression pictograph

The screenshot shows the RStudio IDE with a script titled 'Untitled1*.R'. The script contains the following R code:

```

1 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\MAPK.csv')
2 library(ggplot2)
3 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\dgat1excel.csv')
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))

```

The script reads two CSV files, 'MAPK.csv' and 'dgat1excel.csv', and uses ggplot2 to create a bar plot of expression levels by organism part. The plot is styled with a minimal theme, black bars, and a title 'Bar Plot of expression levels'. The fill colors for the bars are manually set using a vector of 16 hex codes.



Figure 1.6 DGAT1 gene expression levels barplot

Barplot depicting the gene expression level of DGAT1 in the parts of the body X axis represents Organism part and Y axis represents Expression level

The DGAT1 gene expression levels were highest in Adrenal gland and thyroid. The DGAT1 gene expression levels were lowest in amygdala and pancreas.

SLC25A1

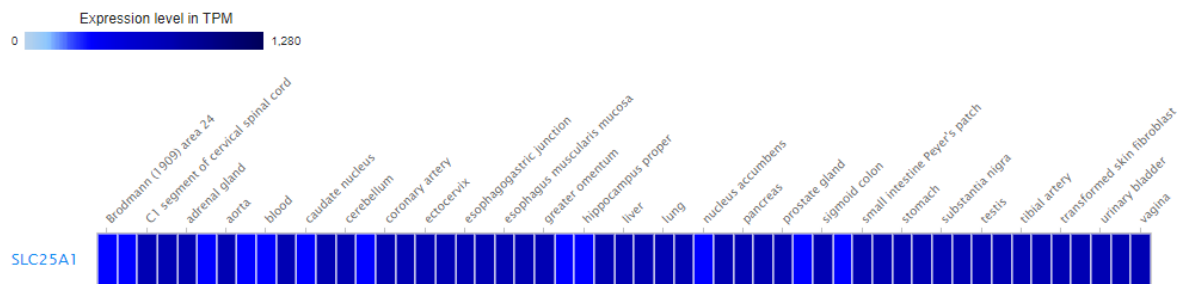


FIGURE 1.7 :SLC25A1 gene expression pictograph

The SLC25A1 gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

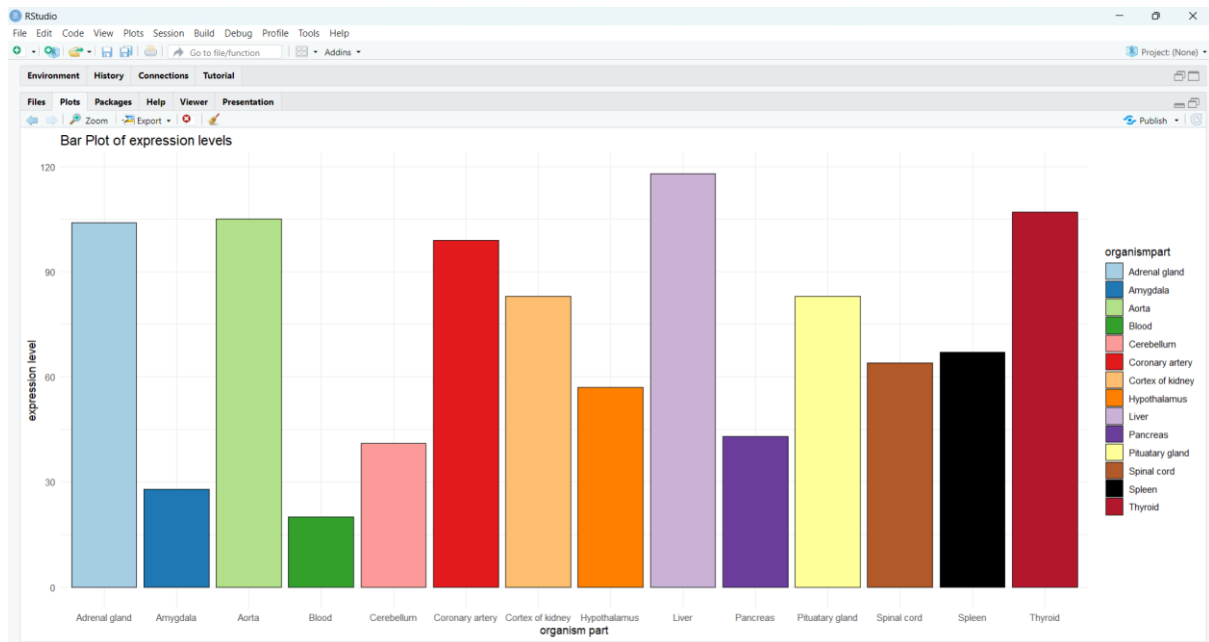


Figure 1.8 SCD gene expression levels barplot

Barplot depicting the gene expression level of SCD in the parts of the body X axis represents Organism part and Y axis represents Expression level

The SCD gene expression levels were highest in adrenal gland and pituitary gland. The SCD gene expression levels were lowest in amygdala and pancreas.

HADH

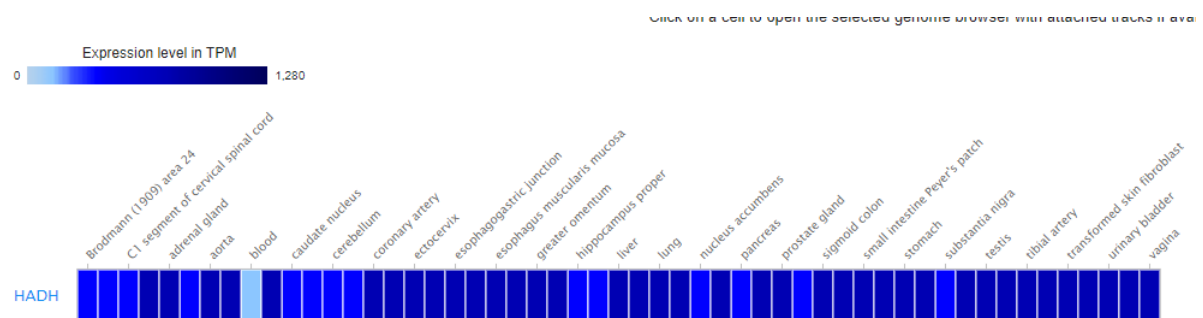


Figure 1.9 HADH gene expression pictograph

The HADH gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - Source on Save Go to file/function Addins
Untitled1*
1 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\MAPK.csv')
2 library(ggplot2)
3 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\hadh.csv')
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))

```

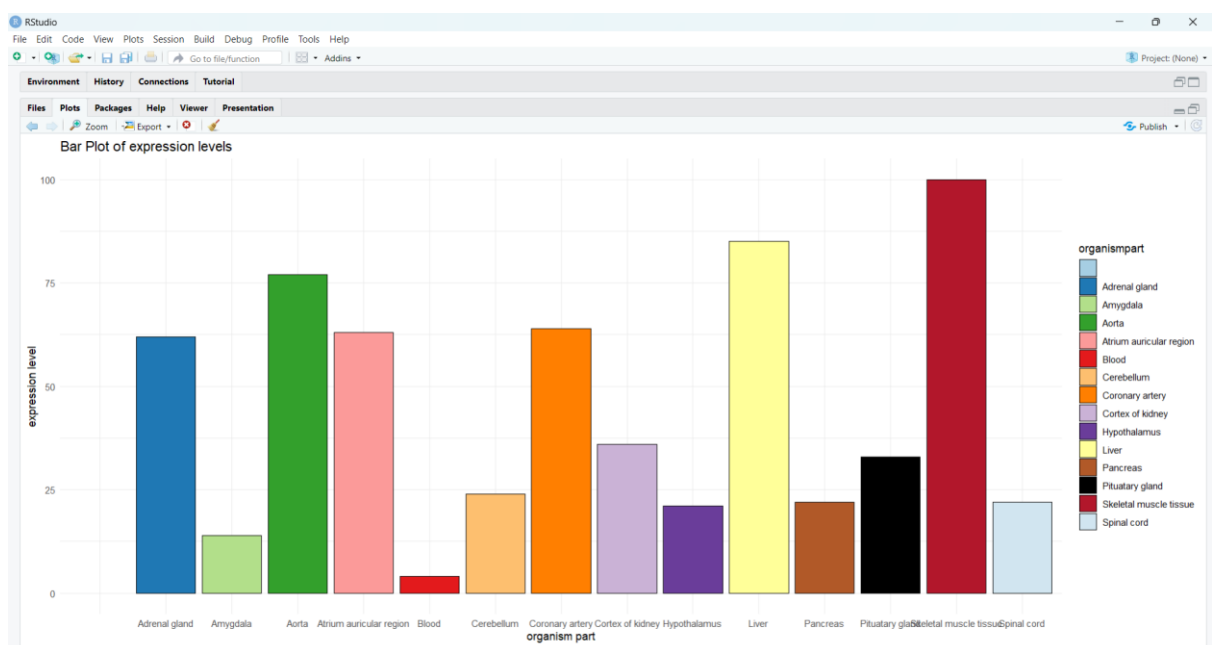


Figure 1.10 HADH gene expression levels barplot

The HADH gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

PKM

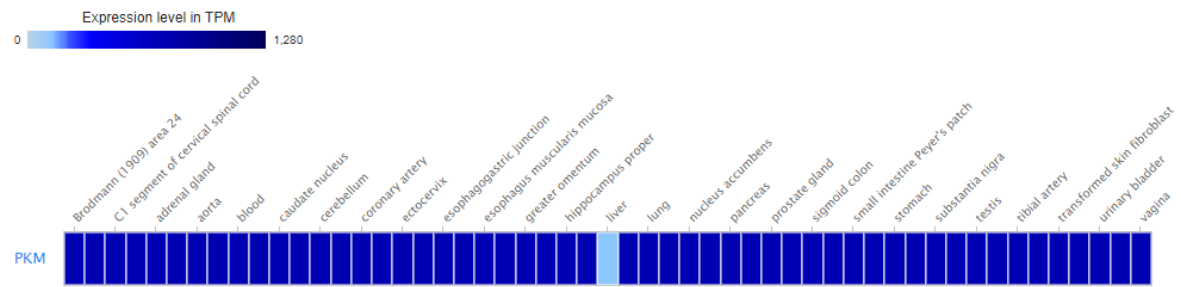
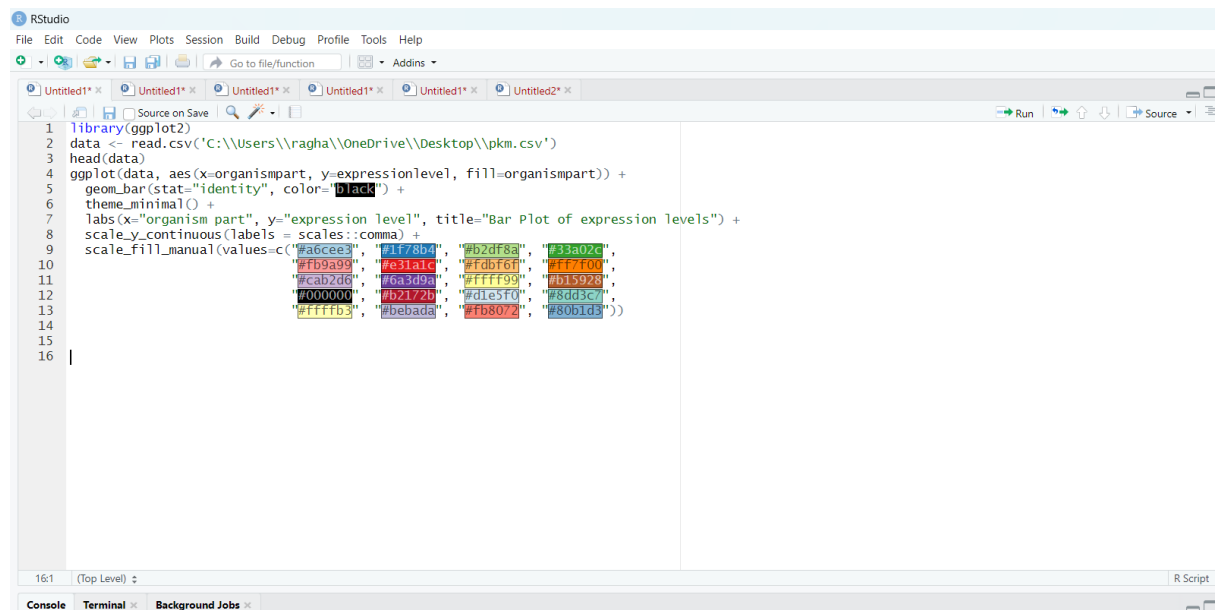


Figure 1.11 PKM gene expression levels pictograph

The PKM gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.



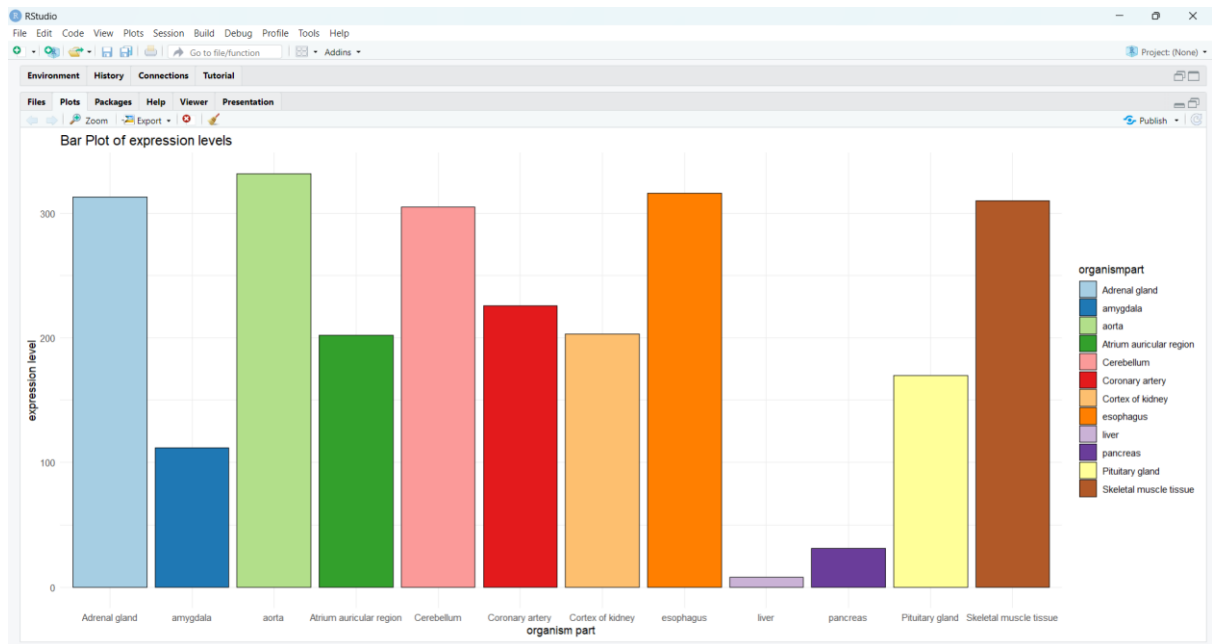


Figure 1.12 PKM gene expression levels barplot

Barplot depicting the gene expression level of PKM in the parts of the body X axis represents Organism part and Y axis represents Expression level

ACLY

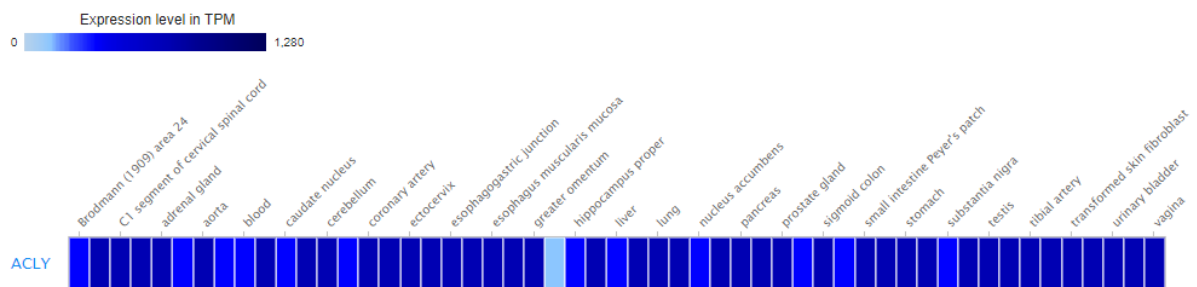


Figure 1.13 ACLY gene expression pictograph

The ACLY gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Source on Save
Run
Source

1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\aclly.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10    "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11    "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12    "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13    "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 (Top Level) R Script

```

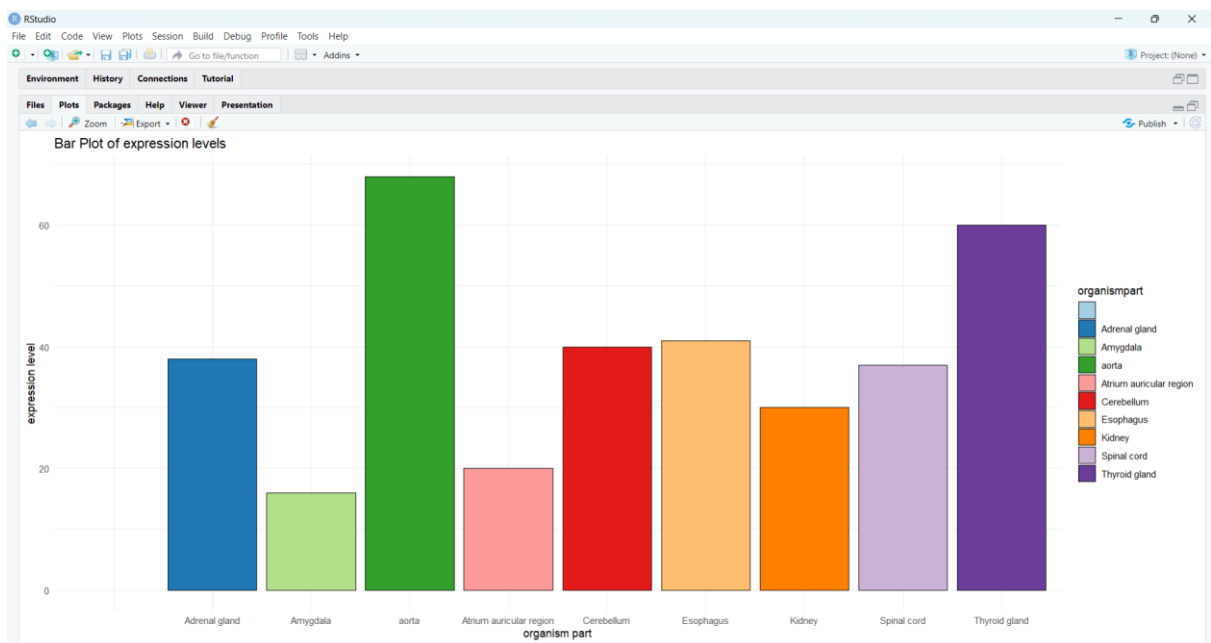


Figure 1.14 ACLY gene expression levels barplot

Barplot depicting the gene expression level of ACLY in the parts of the body X axis represents Organism part and Y axis represents Expression level

ACACA

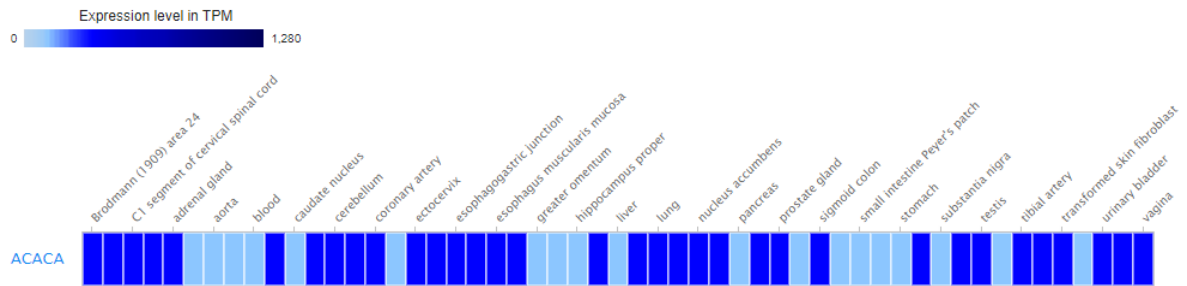
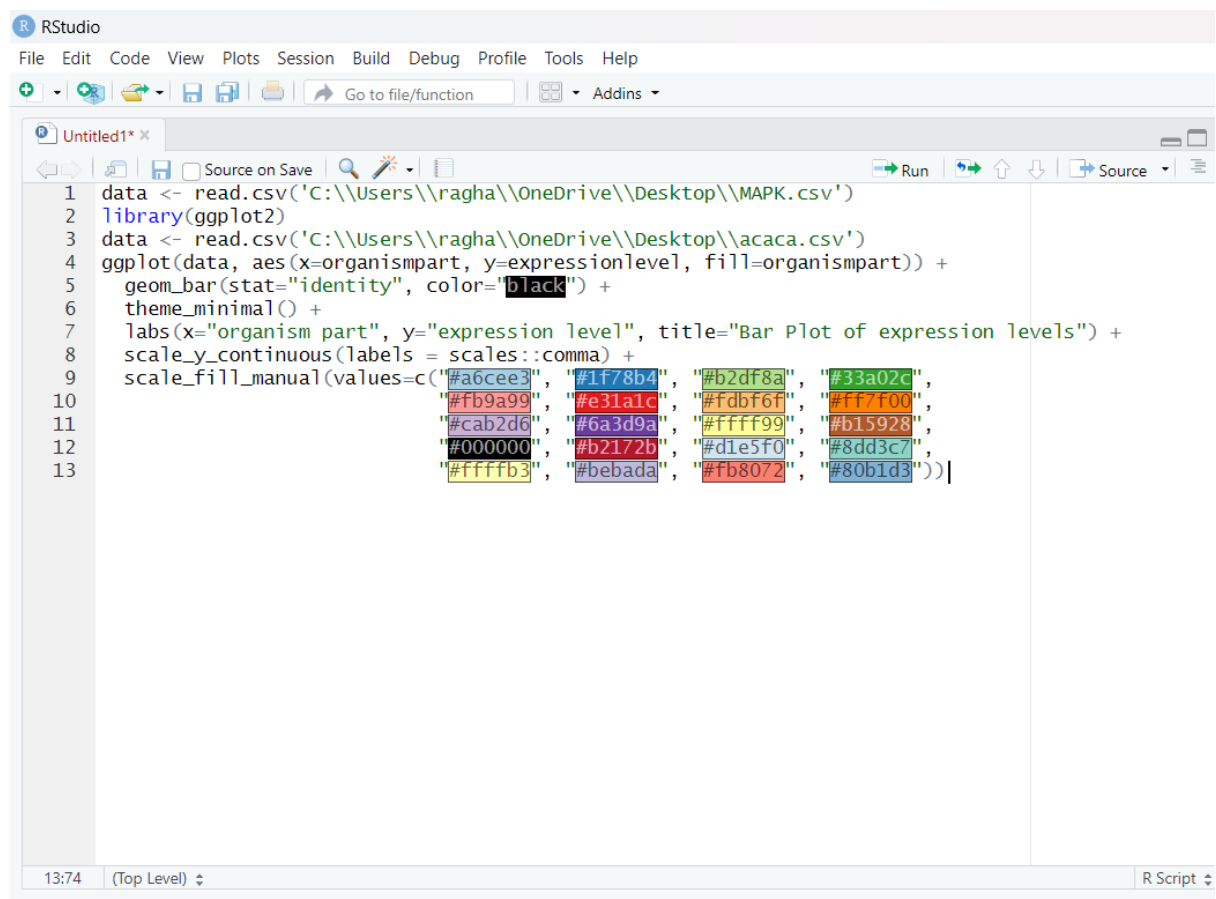


Figure 1.15 ACACA gene expression pictograph

The ACACA gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.



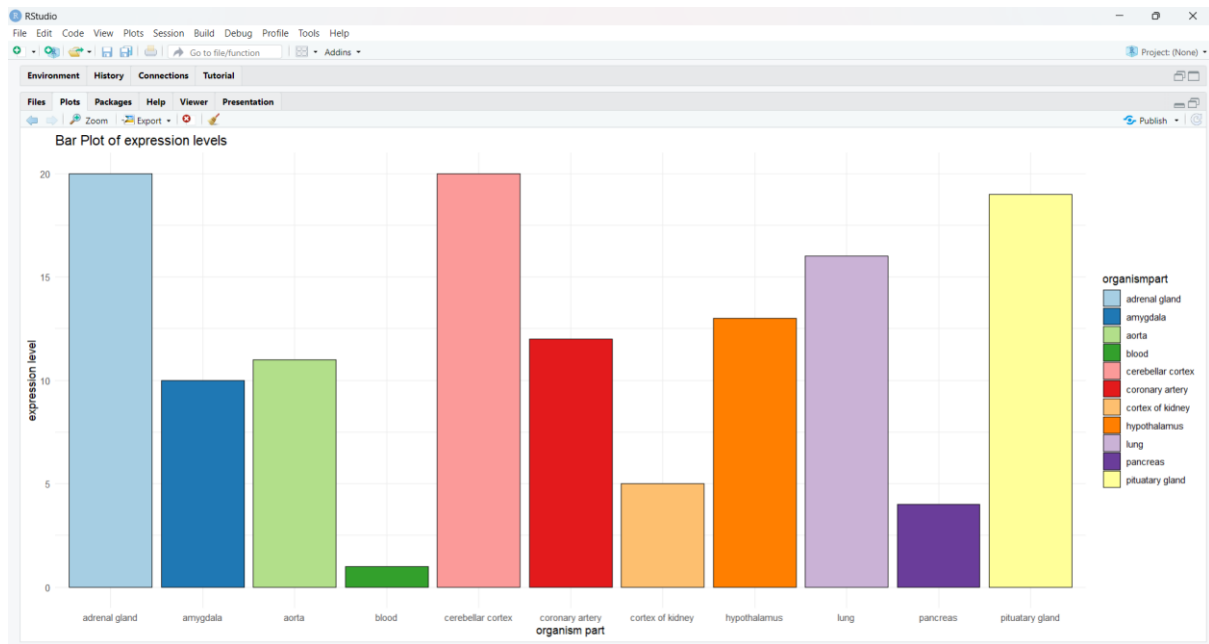


Figure 1.16 ACACA gene expression levels barplot

Barplot depicting the gene expression level of ACACA in the parts of the body X axis represents Organism part and Y axis represents Expression level

MLXIPL

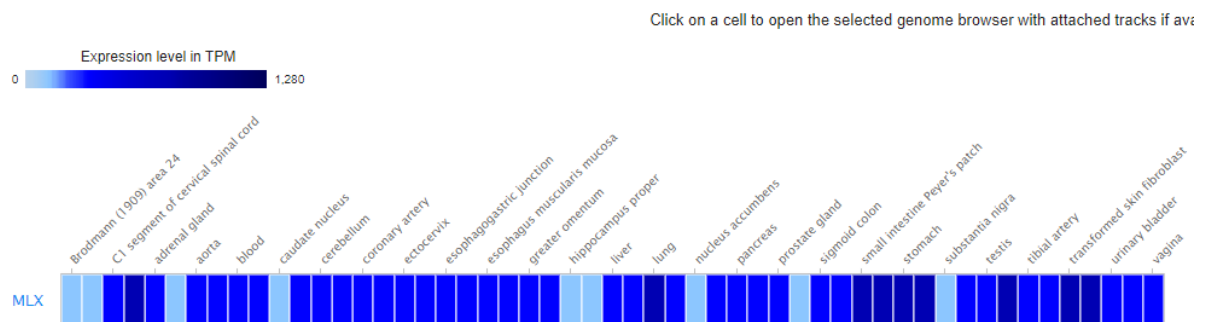


Figure 1.17 MLXIPL gene expression pictograph

The MLXIPL gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\mlx.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16

```

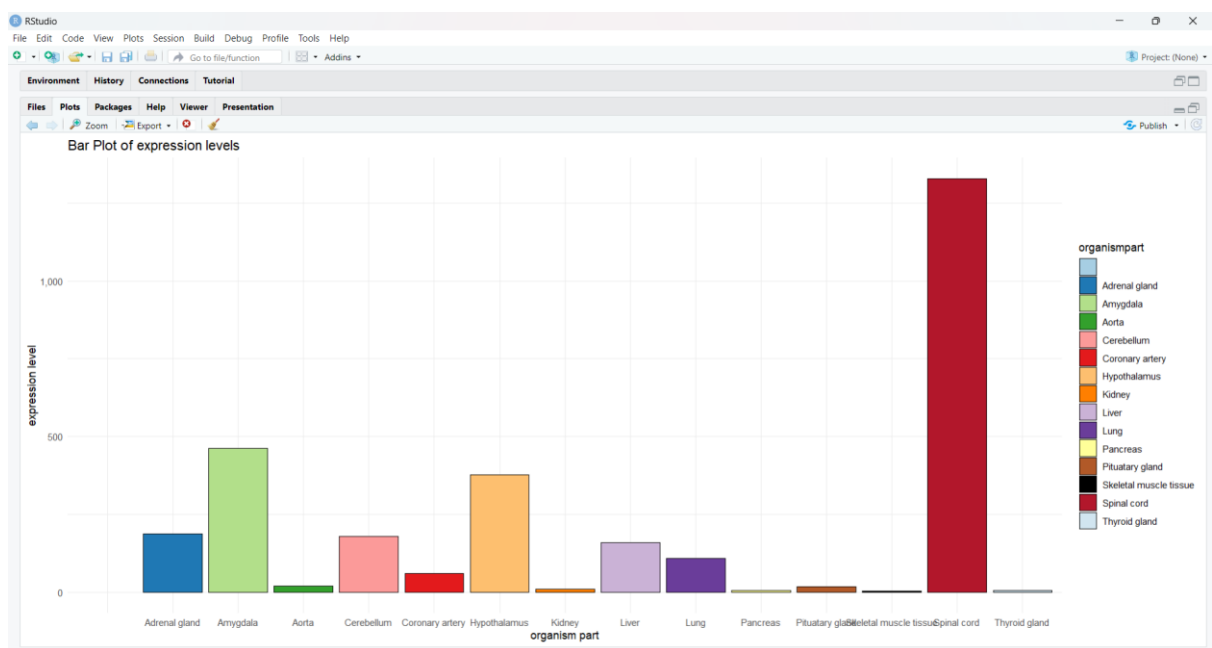


Figure 1.18 MLXIPL gene expression levels barplot

Barplot depicting the gene expression level of MLXIPL in the parts of the body X axis represents Organism part and Y axis represents Expression level

SCD

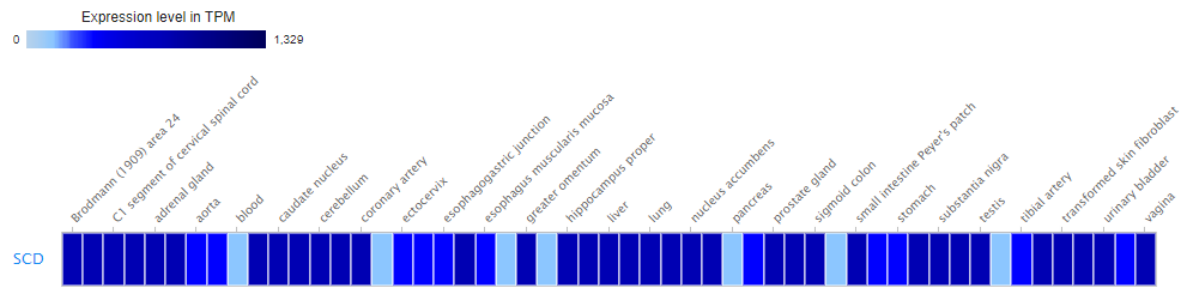


Figure 1.19 SCD gene expression pictograph

The SCD gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Source on Save
Run
Source

1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\scd.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10    "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11    "#cab2d6", "#6a3d9a", "#ffff99", "#b15926",
12    "#000000", "#b21f2b", "#d1e5f0", "#8dd3c7",
13    "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
2:59 (Top Level) R Script

```

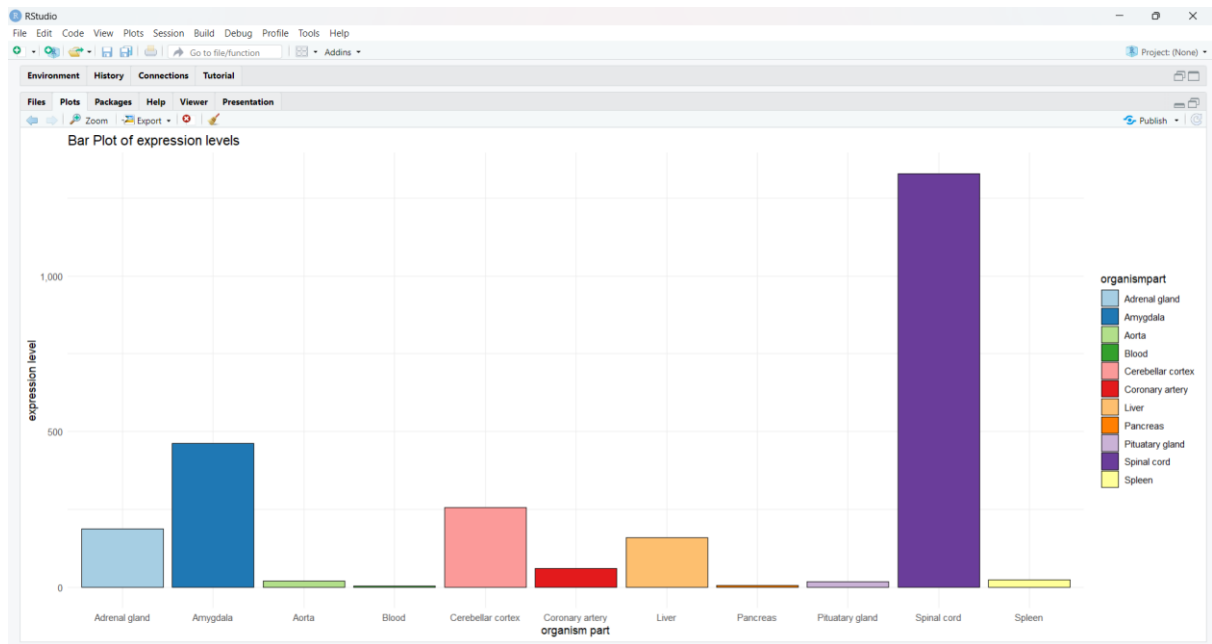


Figure 1.20 MLXIPL gene expression levels barplot

Barplot depicting the gene expression level of MLXIPL in the parts of the body X axis represents Organism part and Y axis represents Expression level

GHSR

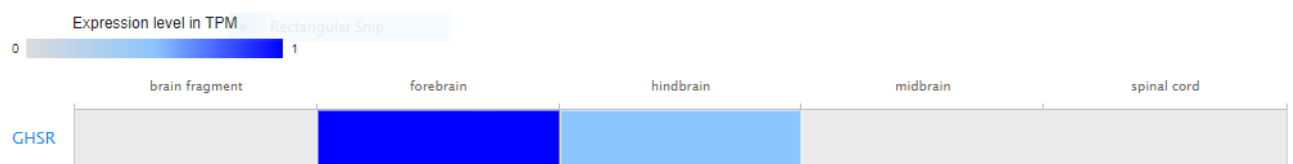


Figure 1.21 GHSR gene expression pictograph

The GHSR gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

Table 11 Expression values of GHSR gene((Table containing name of organism bodypart and corresponding expression values in TPM)

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled2* x
Source on Save Run
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\ghsr.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2182b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 | (Top Level) R Script

```

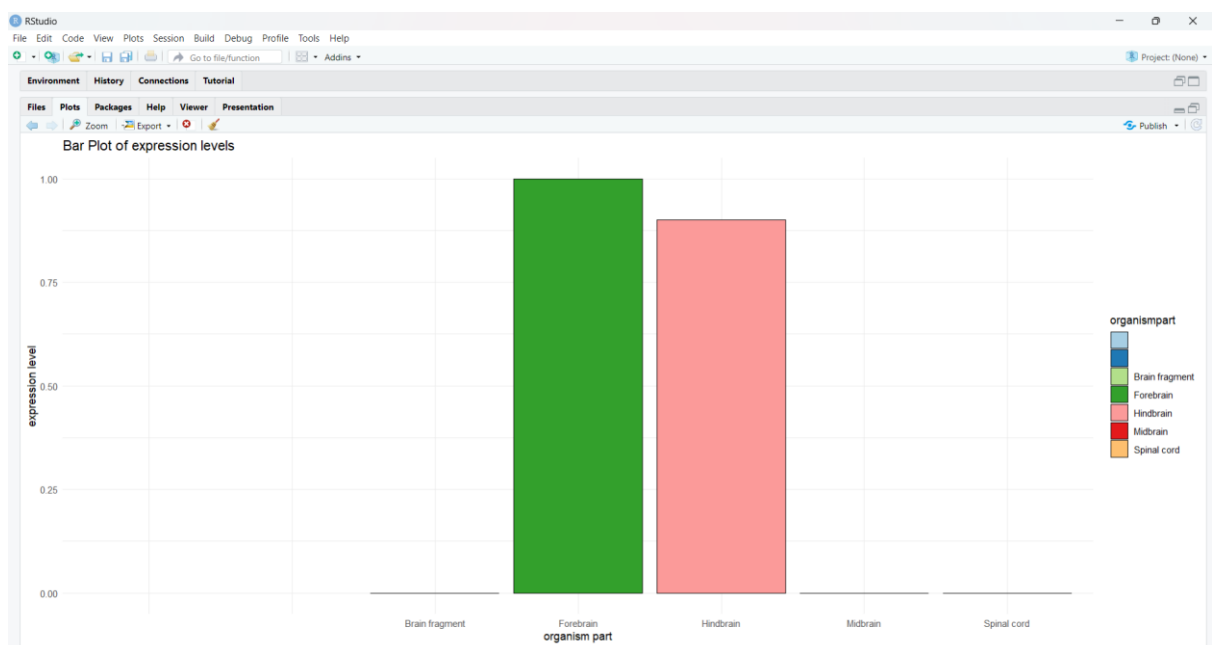


Figure 1.22:GHSR gene expression barplot

Barplot depicting the gene expression level of GHSR in the parts of the body X axis represents Organism part and Y axis represents Expression level

NPY

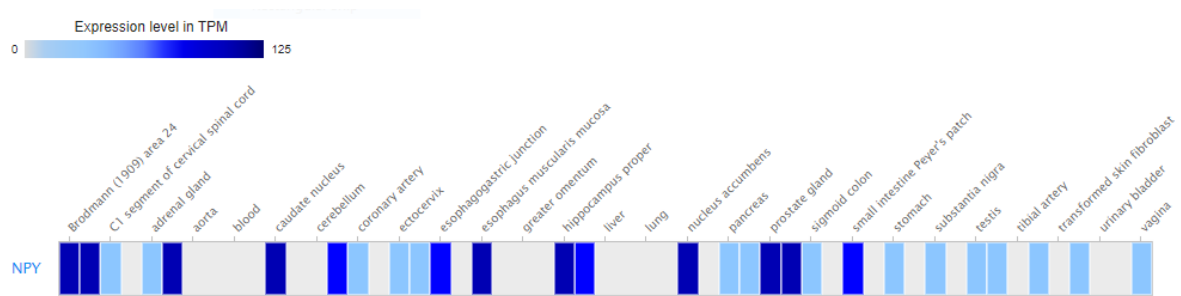


Figure 1.23 NPY gene expression pictograph

The NPY gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled2* x
Source on Save Run Source
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\npv.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 (Top Level) R Script

```

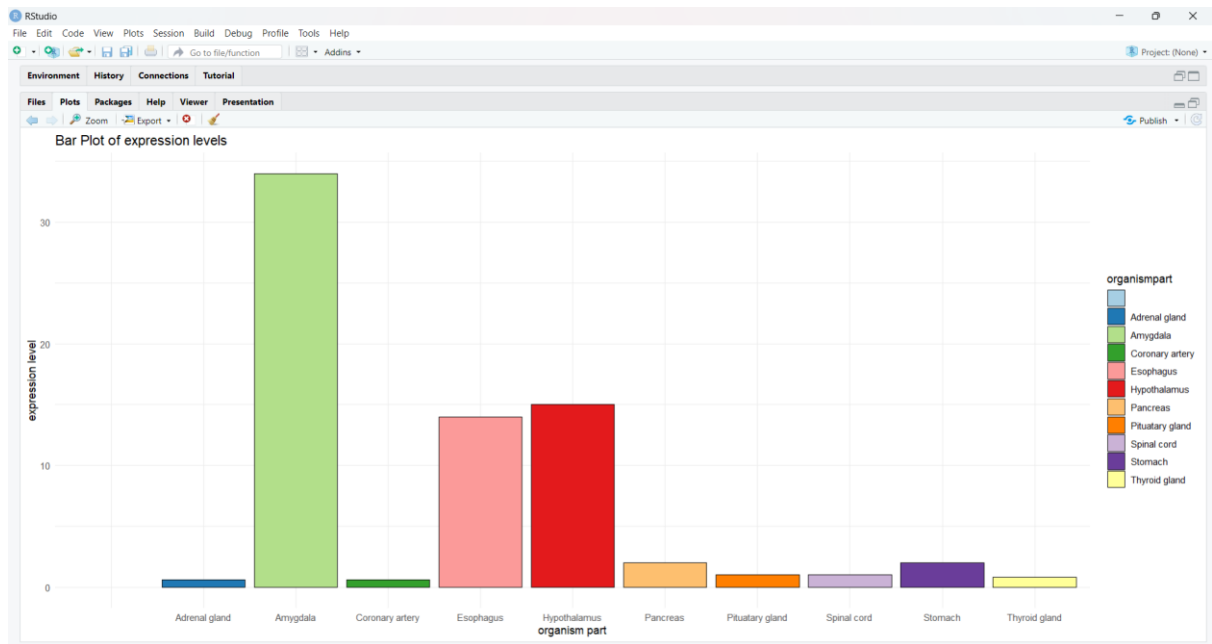


Figure 1.24 NPY gene expression levels barplot

Barplot depicting the gene expression level of NPY in the parts of the body X axis represents Organism part and Y axis represents Expression level

AGRP

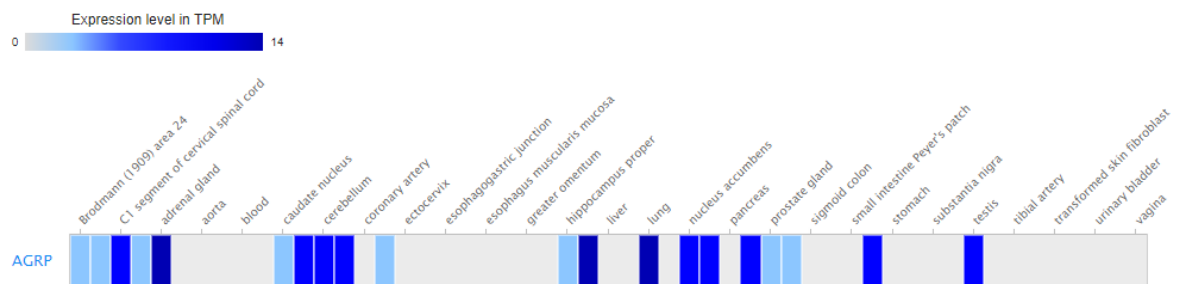


Figure 1.25: AGRP gene expression pictograph

The AGRP gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\agrp.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ffff99",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b21f2b", "#d1e5f0", "#800080",
13     "#ffff99", "#b2df8a", "#fb9a99", "#b2df8a"))
14
15
16

```

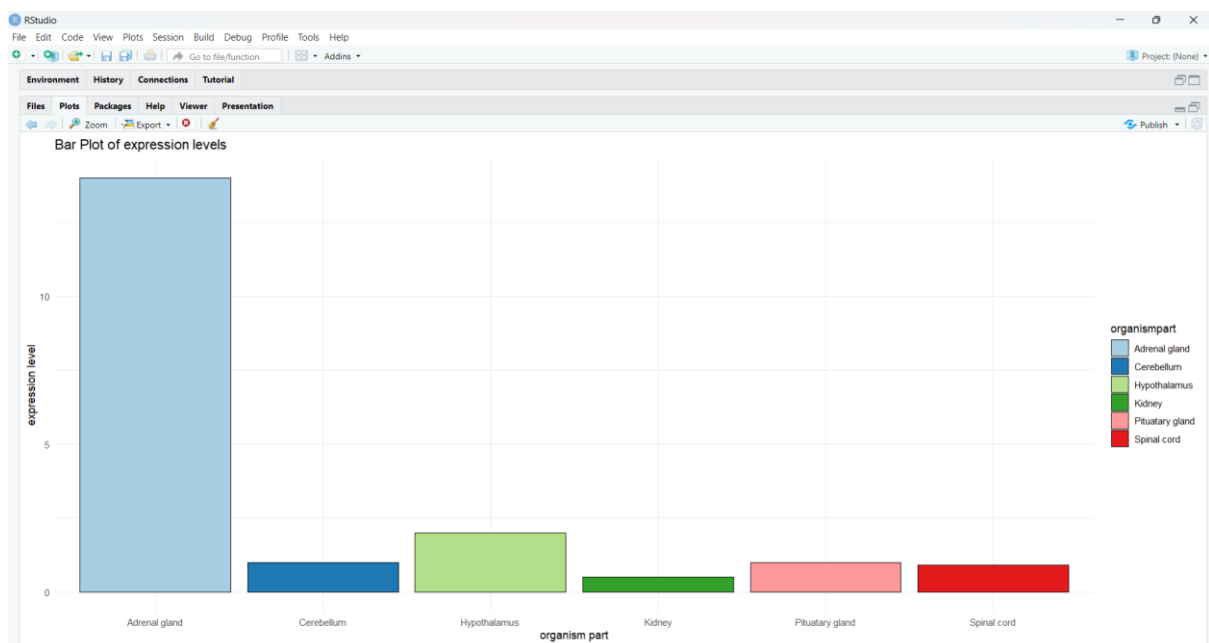


Figure 1.26 AGRP gene expression levels barplot

Barplot depicting the gene expression level of AGRP in the parts of the body X axis represents Organism part and Y axis represents Expression level

FOX01

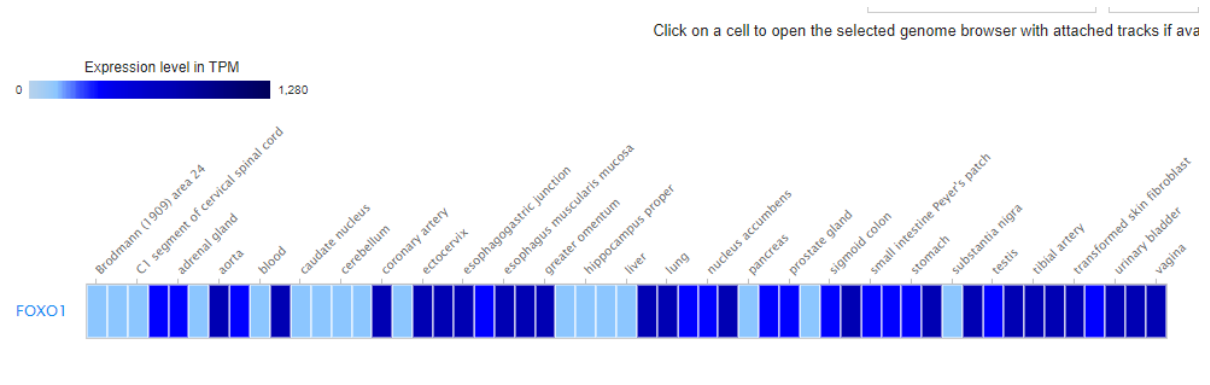
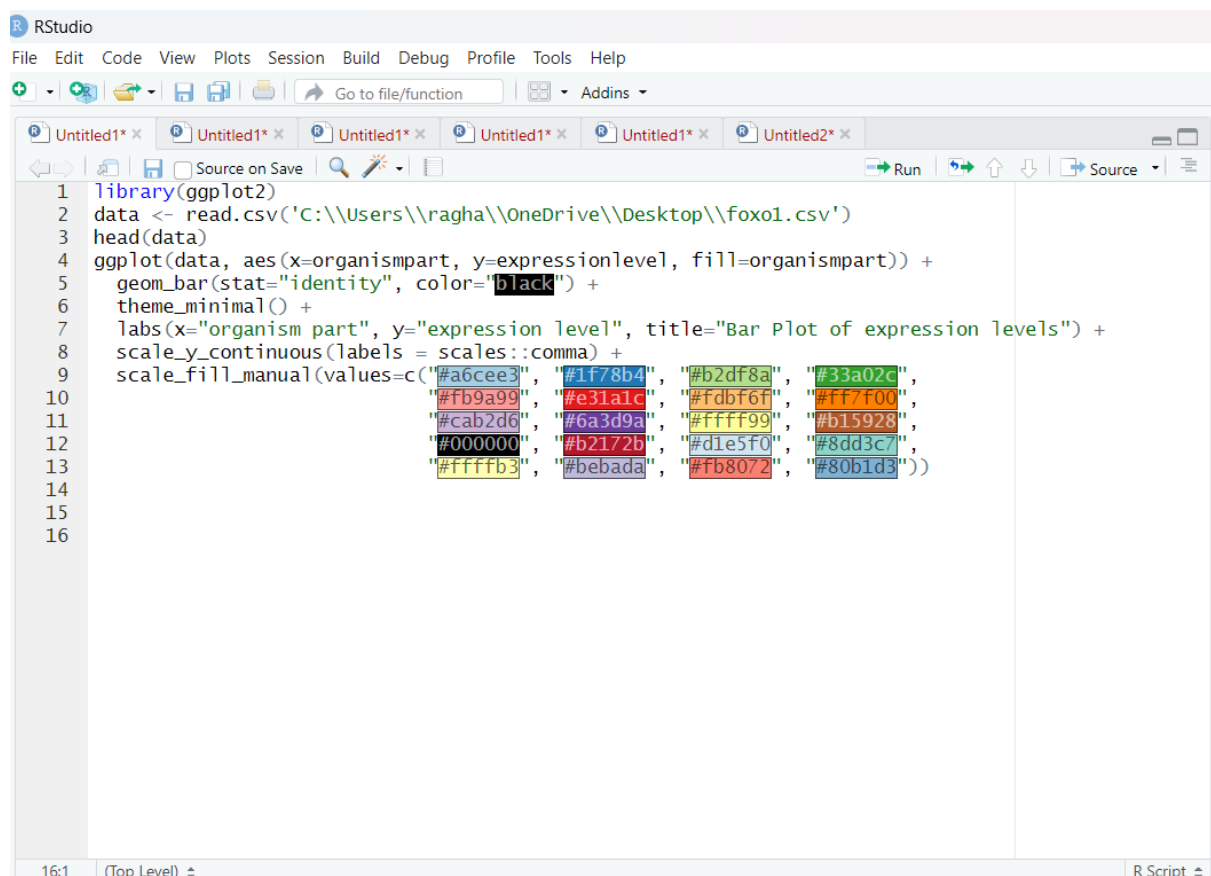


Figure 1.27 FOXO1 gene expression pictograph

The FOXO1 gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.



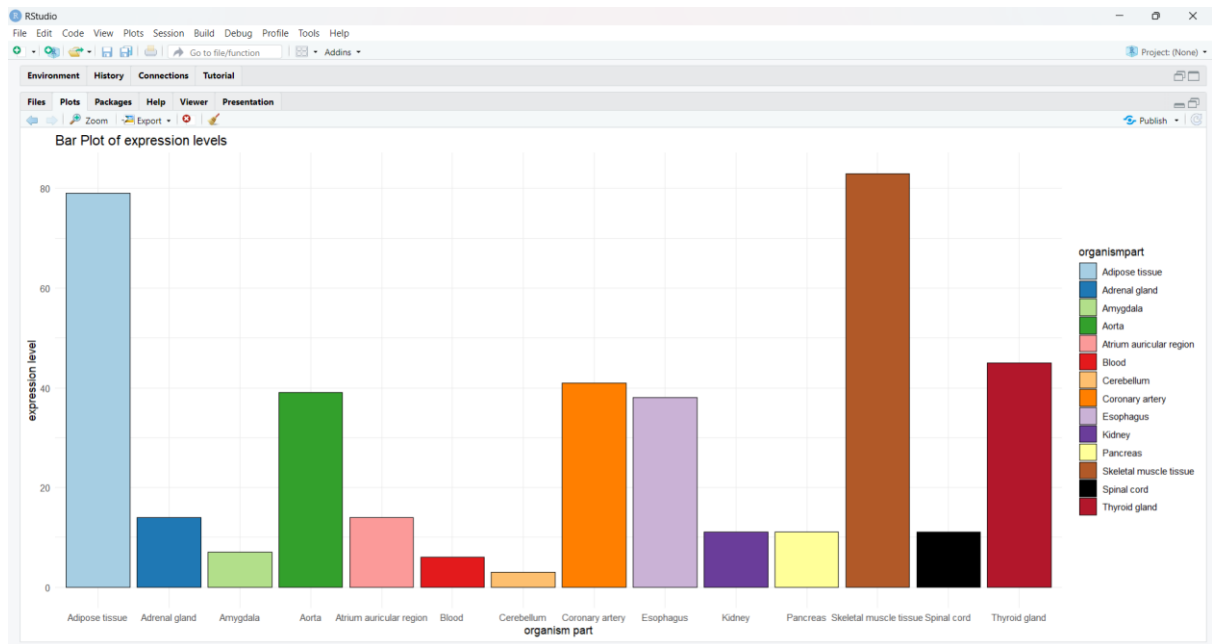


Figure 1.28 FOXO1 gene expression barplot

Barplot depicting the gene expression level of FOXO1 in the parts of the body X axis represents Organism part and Y axis represents Expression level

ADCY3

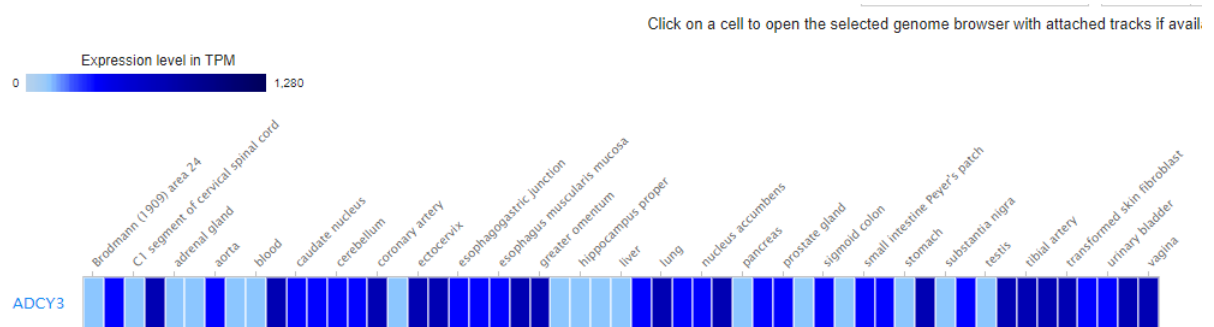


Figure 1.29 ADCY3 gene expression pictograph

The ADCY3 gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled2* x
Source on Save Run Source
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\adcy3.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10    "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11    "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12    "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13    "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 (Top Level) R Script

```

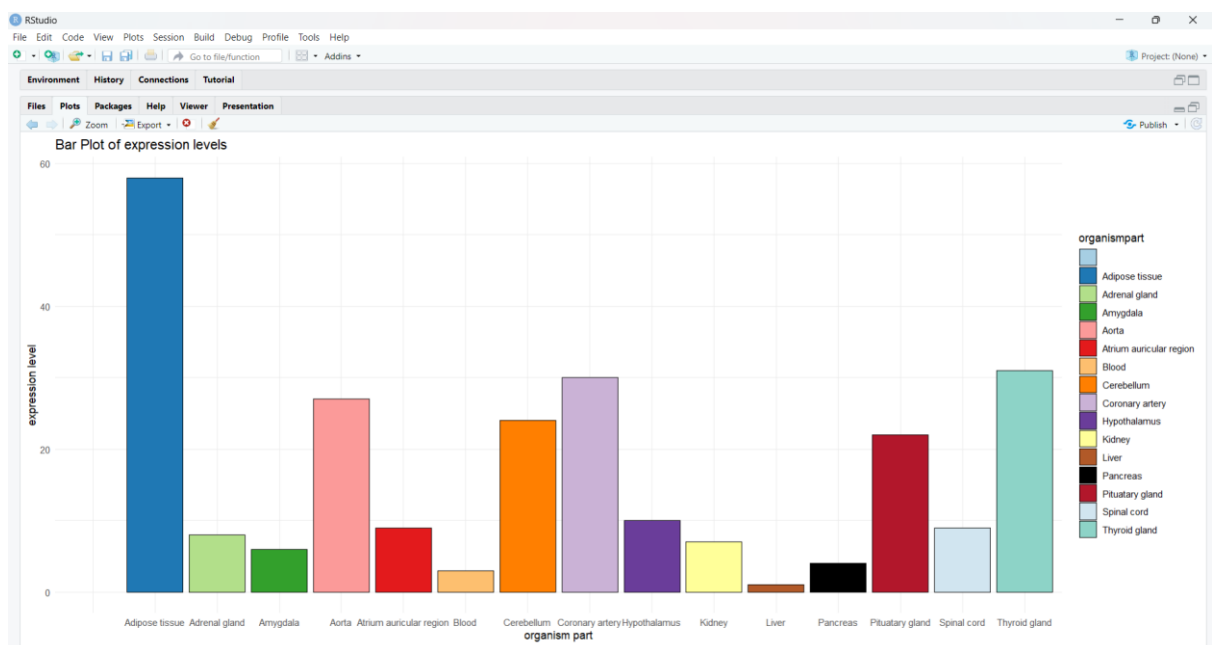


Figure 1.30 ADCY3 gene expression levels barplot

Barplot depicting the gene expression level of ADCY3 in the parts of the body X axis represents Organism part and Y axis represents Expression level

APLN

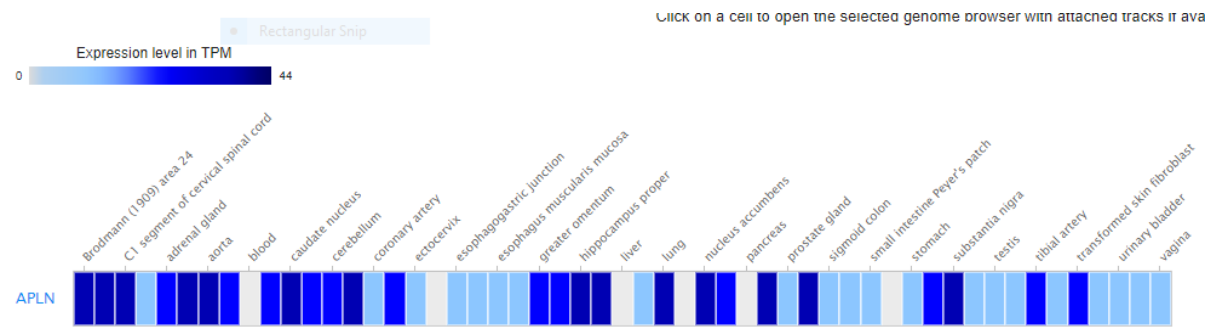


Figure 1.31 APLN gene expression pictograph

The APLN gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Untitled1* x
Untitled1* x
Untitled1* x
Untitled1* x
Untitled1* x
Untitled2* x
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\apln.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#f7f100",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
```

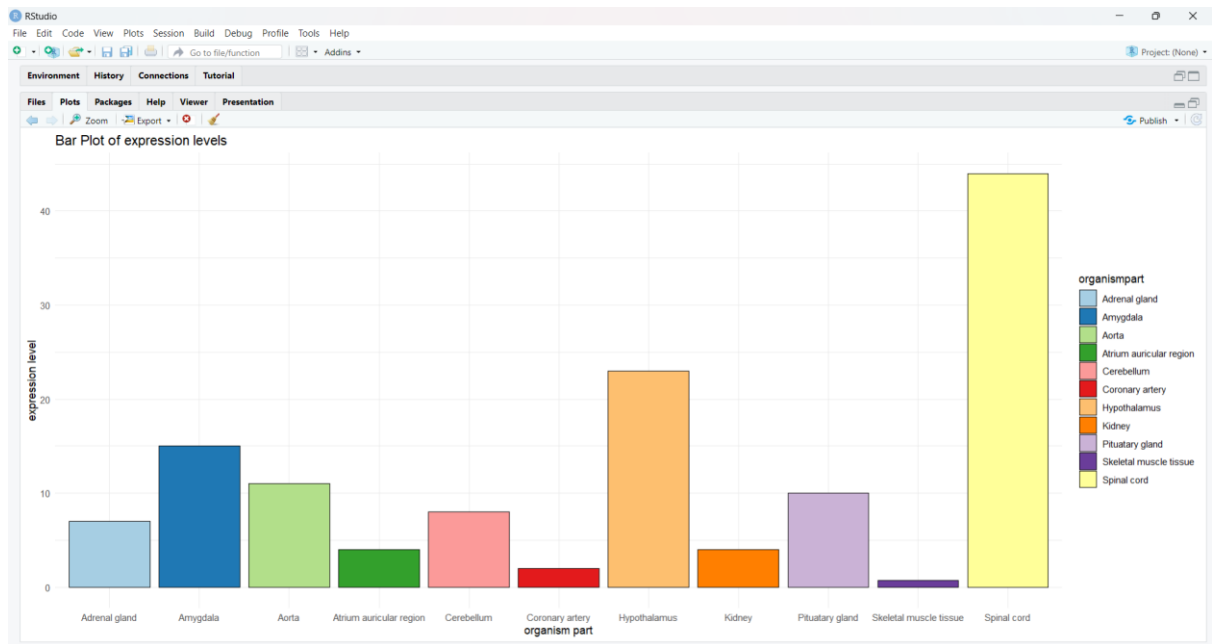


Figure 1.32 APLN gene expression levels barplot

Barplot depicting the gene expression level of APLN in the parts of the body X axis represents Organism part and Y axis represents Expression level

APLNR

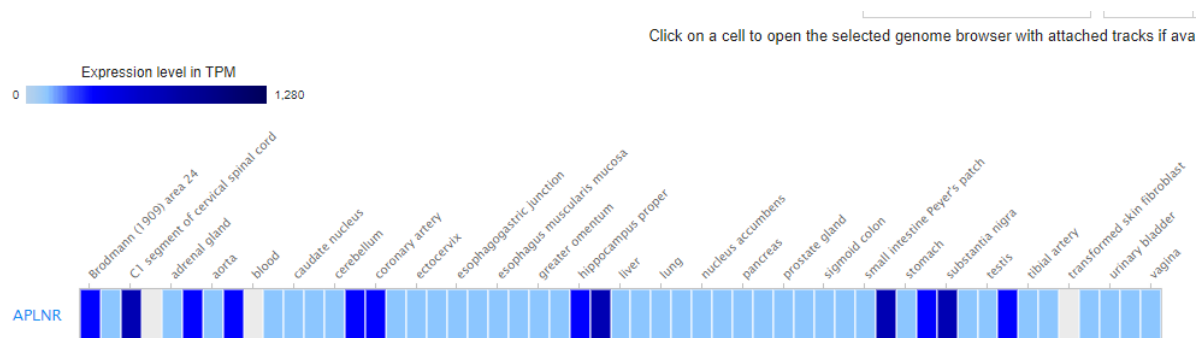


Figure 1.33 APLNR gene expression pictograph

The APLNR gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\aplnr.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 (Top Level) R Script

```

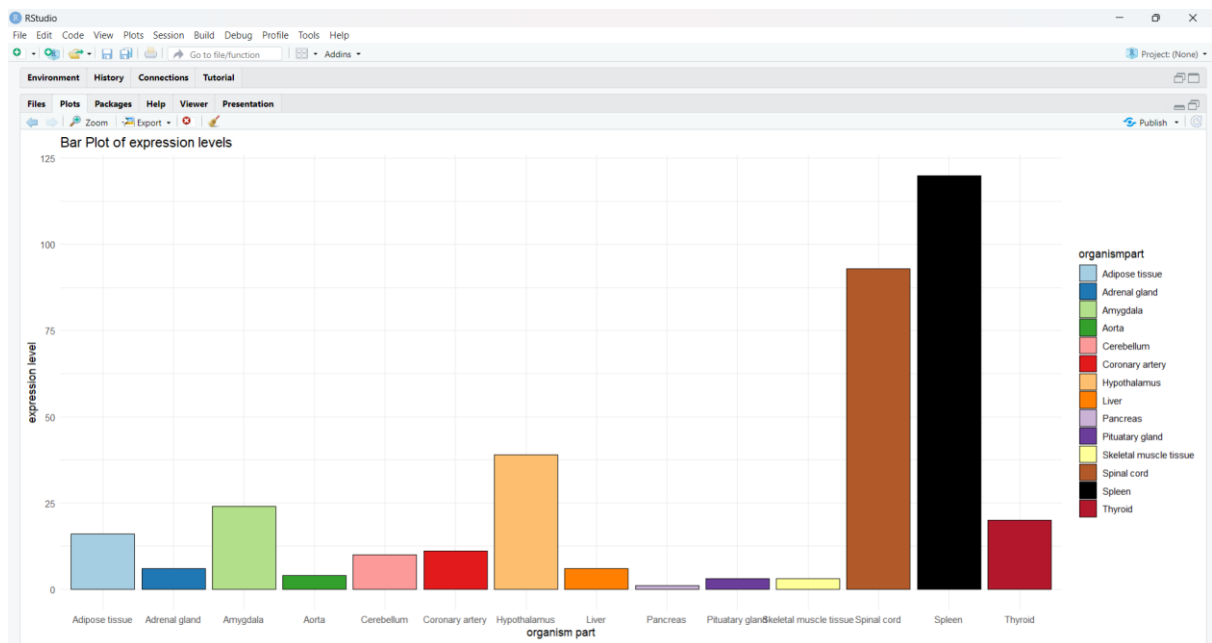


Figure 1.34 APLNR gene expression levels barplot

Barplot depicting the gene expression level of APLNR in the parts of the body X axis represents Organism part and Y axis represents Expression level

MAPK1

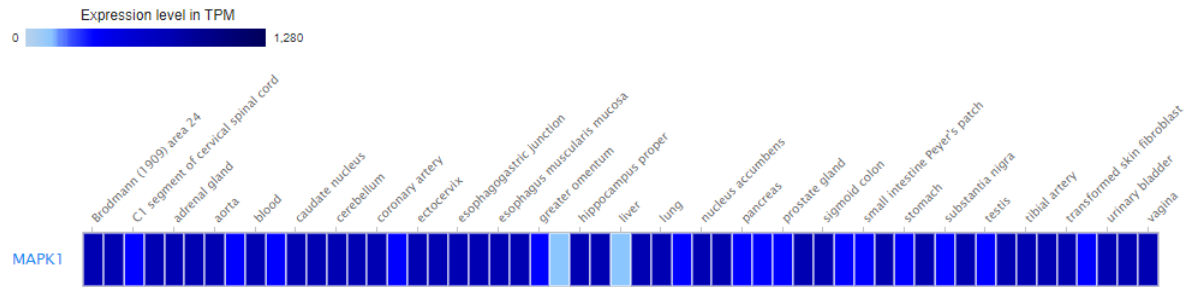


Figure 1.35 MAPK1 gene expression barplot

The MAPK1 gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled1* x Untitled2* x
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\mapk.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10    "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11    "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12    "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13    "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
```

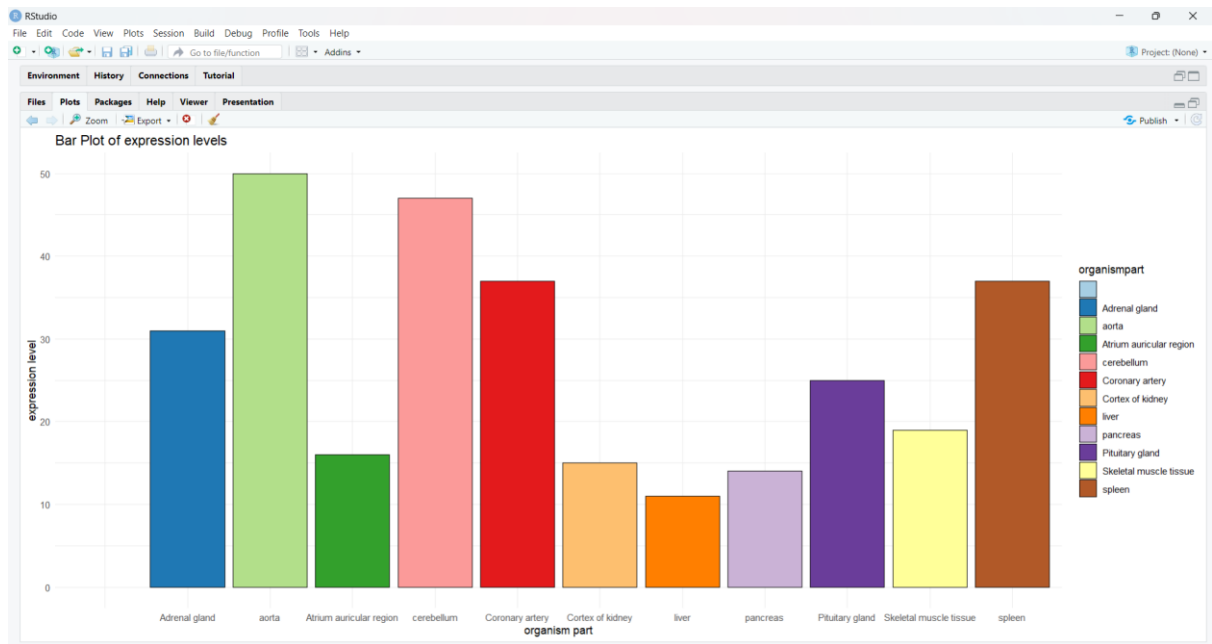


Figure 1.36 MAPK1 gene expression levels barplot

Barplot depicting the gene expression level of MAPK1 in the parts of the body X axis represents Organism part and Y axis represents Expression level

KLF5

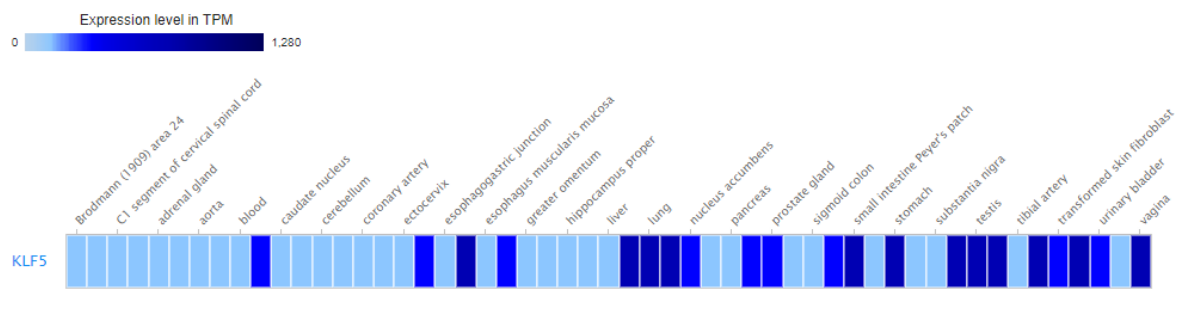


FIGURE 1.37 KLF5 gene expression pictograph

The KLF5 gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Untitled1* x
Source on Save
Run
Source
1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\KLF5.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2172b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
16:1 (Top Level) R Script

```

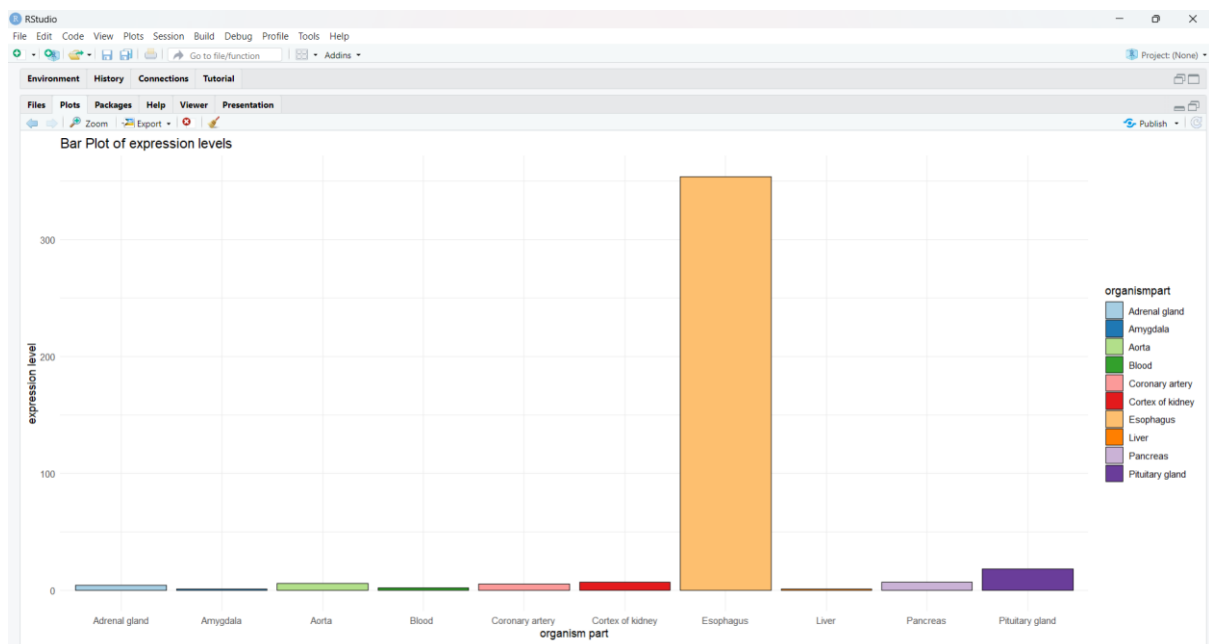


Figure 1.38 KLF5 gene expression levels barplot

Barplot depicting the gene expression level of KLF5 in the parts of the body X axis represents Organism part and Y axis represents Expression level

MC3R

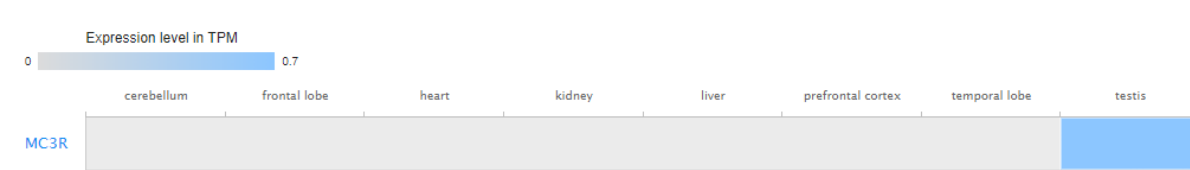


FIGURE 1.39 MC3R gene expression pictograph

The MC3R gene expression levels in different parts of the body are depicted as values in TPM(Transcripts Per Million) ranging from 0 TPM to 1280 TPM.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Source on Save
Run
Source

1 library(ggplot2)
2 data <- read.csv('C:\\Users\\ragha\\OneDrive\\Desktop\\mc3r.csv')
3 head(data)
4 ggplot(data, aes(x=organismpart, y=expressionlevel, fill=organismpart)) +
5   geom_bar(stat="identity", color="black") +
6   theme_minimal() +
7   labs(x="organism part", y="expression level", title="Bar Plot of expression levels") +
8   scale_y_continuous(labels = scales::comma) +
9   scale_fill_manual(values=c("#a6cee3", "#1f78b4", "#b2df8a", "#33a02c",
10     "#fb9a99", "#e31a1c", "#fdbf6f", "#ff7f00",
11     "#cab2d6", "#6a3d9a", "#ffff99", "#b15928",
12     "#000000", "#b2182b", "#d1e5f0", "#8dd3c7",
13     "#ffffb3", "#bebada", "#fb8072", "#80b1d3"))
14
15
16
```

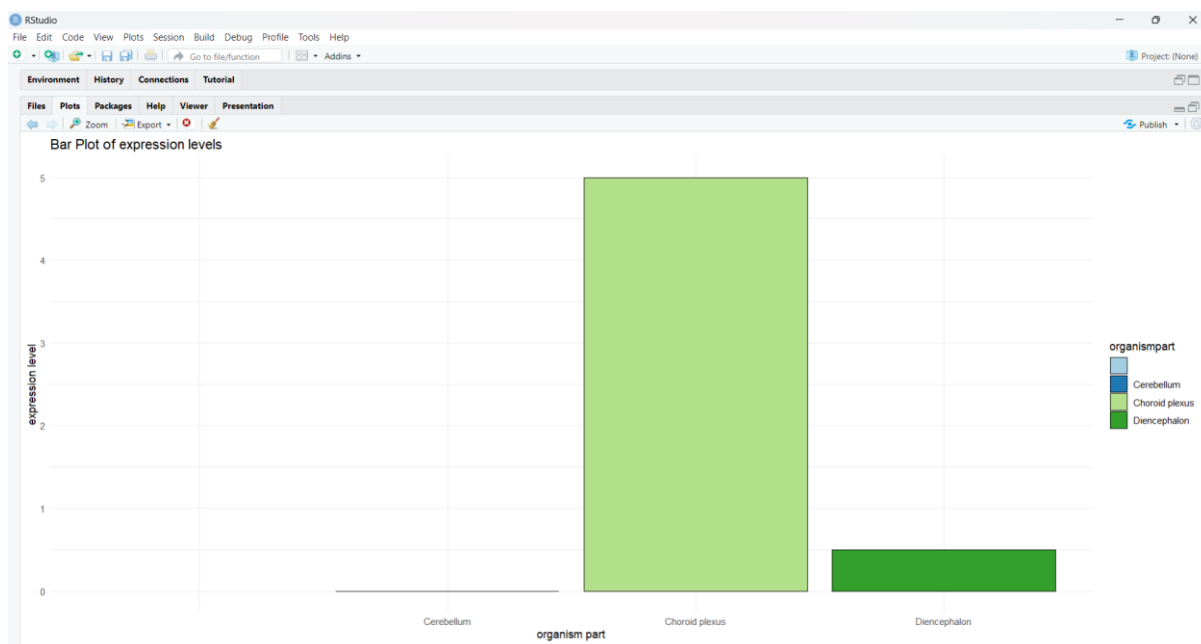


Figure 1.40 MC3R gene expression levels barplot

Barplot depicting the gene expression level of MC3R in the parts of the body X axis represents Organism part and Y axis represents Expression level

5.3 ORTHOLOGOUS SEQUENCES

Based on the NCBI BLAST results, the genes and orthologous sequences were identified and the fasta complete sequences of the gene and orthologs were downloaded. The Blast results contained information on sequence description, accession number, percentage identity, e value, accession length are present. The BLAST results (refer to Table 1 of Appendix) contain information on description, accession number, percentage identity, e value, accession length of human gene sequences and orthologous gene sequences

5.4 PHYLOGENETIC ANALYSIS

Phylogenetic analysis was performed using phylogeny.fr comprising Gblocks, MUSCLE, Phylogeny and TreeDyn and the phylogenetic trees were analysed. The bootstrap values are higher than the legend values indicating strong evolutionary support. The main result of a phylogenetic

analysis is the phylogenetic tree, which depicts the evolutionary relationships among the sequences. The tree is typically displayed in a hierarchical format, with branches representing the divergence of sequences from a common ancestor. Bootstrap values indicate the support for specific branches in the phylogenetic tree. Higher bootstrap values (usually expressed as percentages) indicate stronger support for the inferred relationships. Bootstrap values in phylogenetic analysis indicate the statistical support for specific branches in a phylogenetic tree. They are calculated by resampling the original dataset multiple times (typically using a method called bootstrapping) to assess the reliability of the inferred relationships. Higher bootstrap values generally indicate stronger support for the corresponding branches in the tree.

The bootstrap values on the tree represent the level of confidence that a particular branching pattern is correct. A bootstrap value of 100 indicates that the branching pattern is seen in 100% of the replicate datasets that were analyzed. Lower bootstrap values indicate less certainty.

Phylogenetic analysis was performed using the sequences of *Homo sapiens*, *Pan paniscus*, *Pantroglodytes*, *Pongo abelii*, *Pongo pygmaeus*, *Macaca fascicularis*, *Gorilla gorilla*. The results can be inferred from the bootstrap values and the scale values. The bootstrap values closer to and equal to 1 indicate strong support for the tree. The bootstrap values higher than 75 are considered to be a good threshold. The bootstrap values below 70 percent have no support for the branches in the tree. The legend value present in the scale bar indicates the number of substitutions per site. Longer this value, higher the sequence divergence. The bootstrap values are present in the nodes of clades of phylogenetic tree. This type of phylogenetic tree is a

rectangular phylogenetic tree or cladogram. Other types of Phylogenetic trees that are used are rectangular and radial. The sequences are represented at the tips of branches of nodes.

When we consider value of 0.009 against bootstrap value 1 then it means there is a difference of 0.009 bases per 1 base. The bootstrap values present at the nodes 0.99,1,0.9,0.94 are closer to 100. Bootstrap value of 1 indicates high confidence in the position of the node in topology of the tree. Values closer to 1 indicate high level of statistical support for position of that node and that branch position in topology of the tree. Greater the length of branches more time has elapsed since divergence.

ACADM

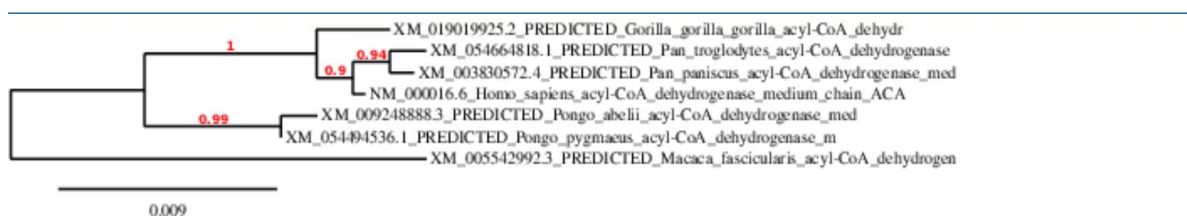


Figure no. 2.1 Phylogenetic tree ACADM

Phylogenetic tree constructed from ACADM gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that Macaca fascicularis is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

CPT1A

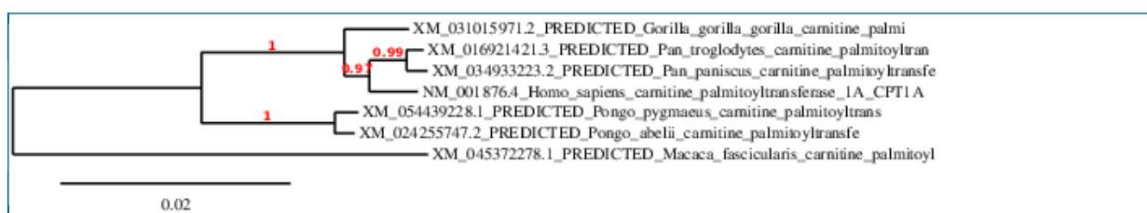


Figure no. 2.2 Phylogenetic tree CPT1A

Phylogenetic tree constructed from CPT1A gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

DGAT1

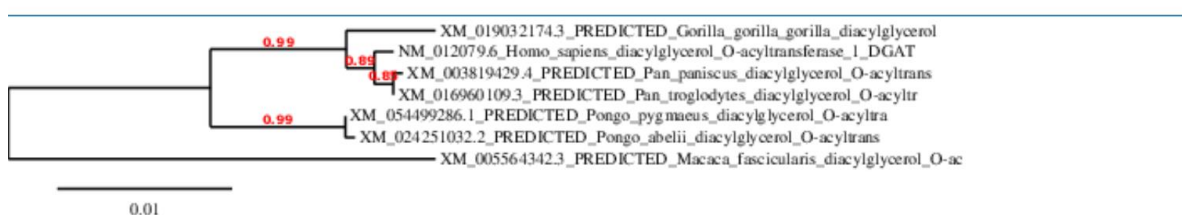


Figure no. 2.3 Phylogenetic tree DGAT1

Phylogenetic tree constructed from DGAT1 gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

SLC25 A1

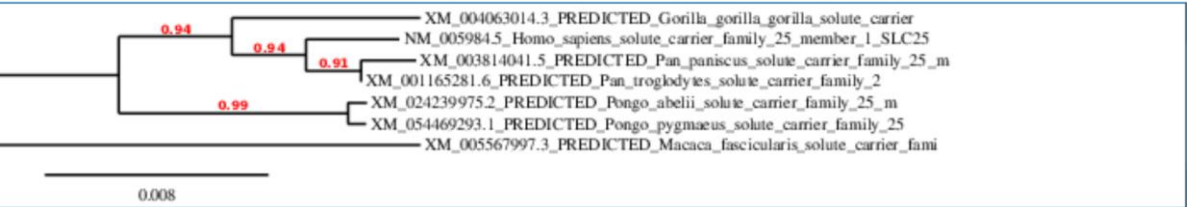


Figure 2.4 Phylogenetic tree SLC25A1

Phylogenetic tree constructed from SLC25A1 gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

HADH

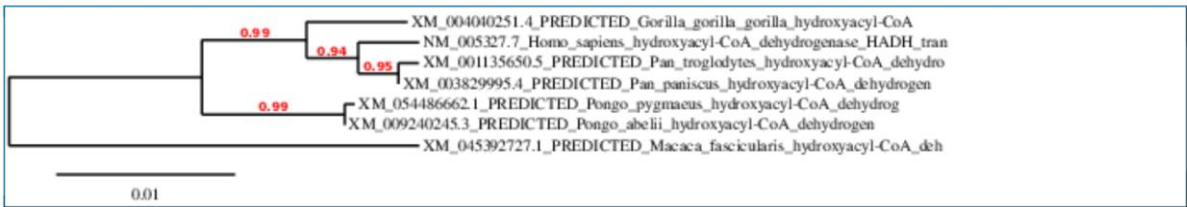


Figure 2.5 Phylogenetic tree HADH

Phylogenetic tree constructed from HADH gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

PKM

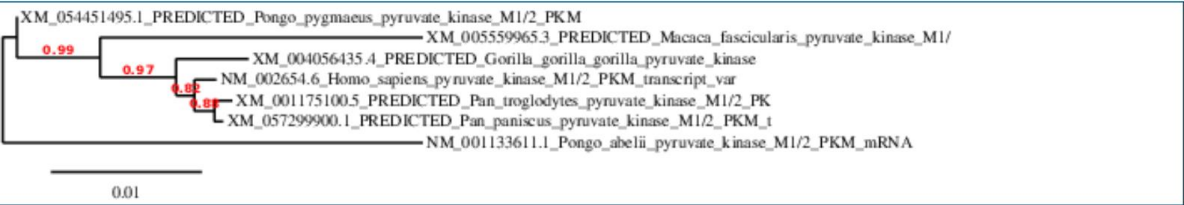


Figure .2.6 Phylogenetic tree PKM

Phylogenetic tree constructed from PKM gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Pongo pygmaeus* is more distantly related than the other organisms.

ACLY

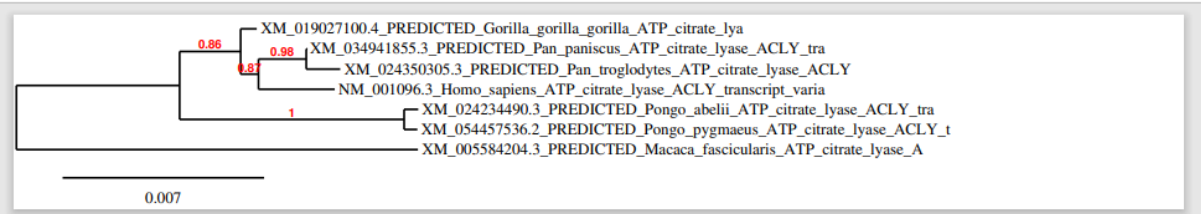


Figure 2.7 Phylogenetic tree ACLY

Phylogenetic tree constructed from ACLY gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

ACACA

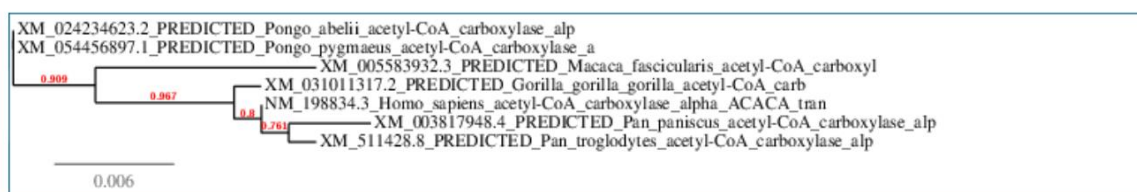


Figure 2.8 Phylogenetic tree ACACA

Phylogenetic tree constructed from ACACA gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Pongo Pygmaeus* and *Pongo abelii* are more distantly related than the other organisms.

MLXIPL

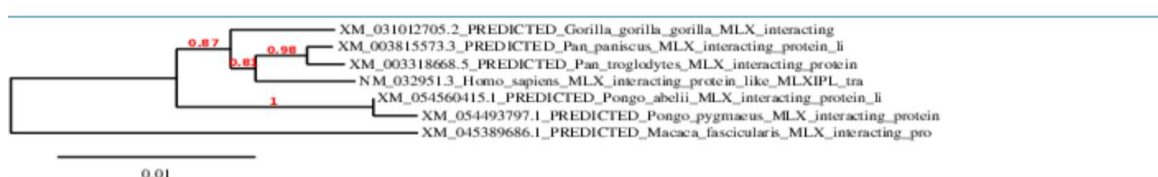


Figure 2.9 Phylogenetic tree MLXIPL

Phylogenetic tree constructed from MLXIPL gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

SCD

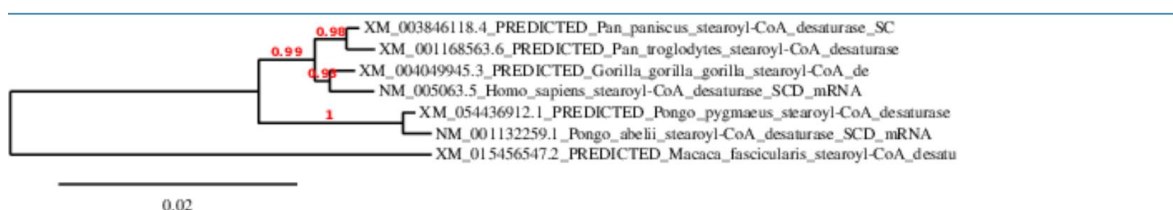


Figure 2.10 Phylogenetic tree SCD

Phylogenetic tree constructed from SCD gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

GHSR

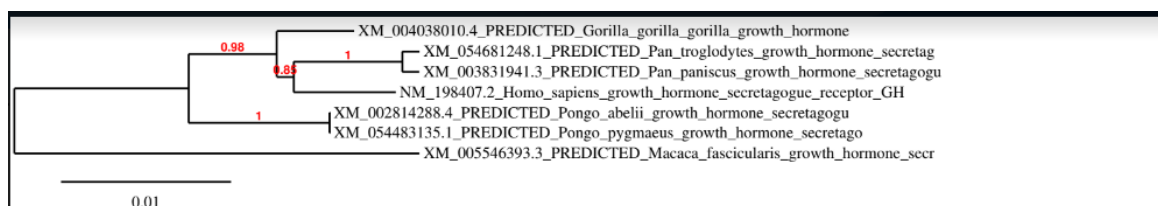


Figure 2.11 Phylogenetic tree GHSR

Phylogenetic tree constructed from GHSR gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that Macaca fascicularis is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

NPY

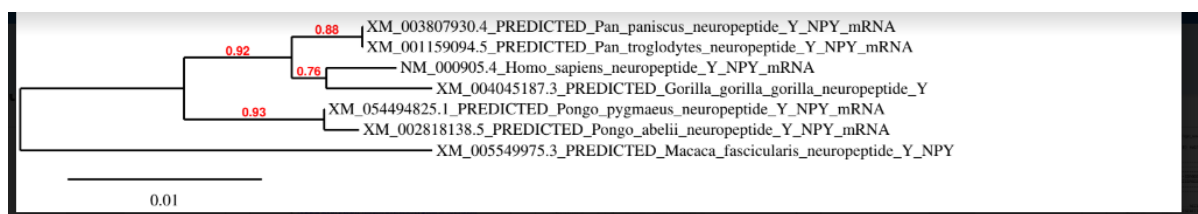


Figure 2.12 Phylogenetic tree NPY

Phylogenetic tree constructed from NPY gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

AGRP

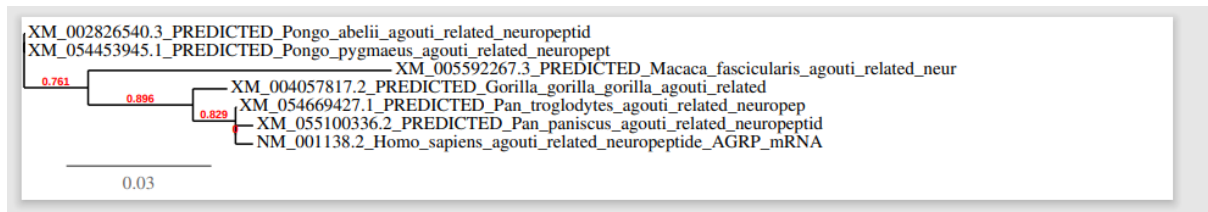


Figure 2.13 Phylogenetic tree AGRP

Phylogenetic tree constructed from AGRP gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Pongo abelii* and *Pongo Pygmaeus* are more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

APLN

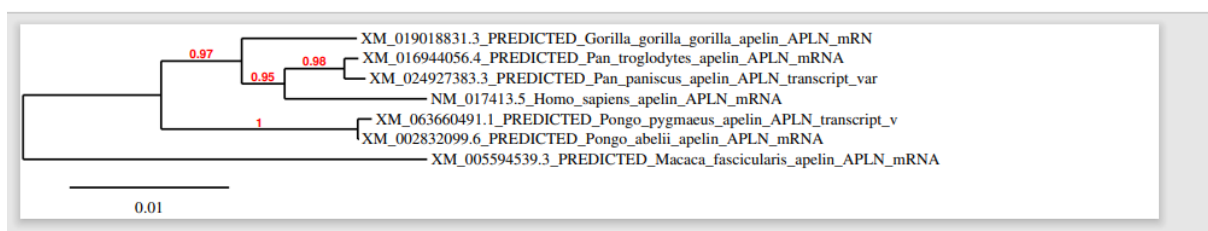


Figure 2.14 Phylogenetic tree APLN

Phylogenetic tree constructed from APLN gene and orthologs

This figure is a cladogram (rectangular type of phylogenetic tree). Phylogenetic relationship between the various organism is represented in the form of branches, nodes and tips. Bootstrap values and scale values represent the magnitude of phylogenetic relationship.

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

APLNR

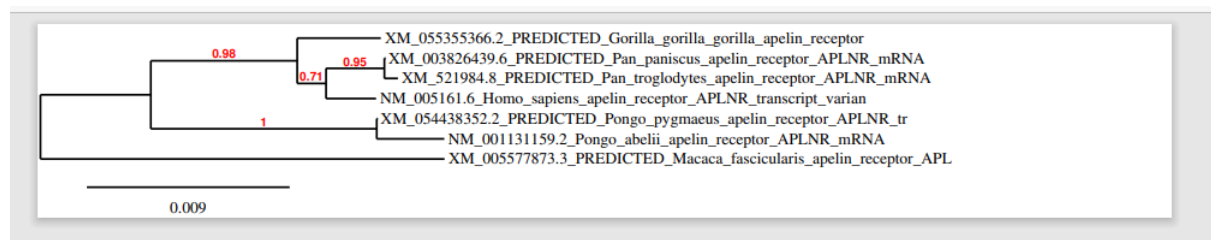


Figure 2.15 Phylogenetic tree APLNR

Phylogenetic tree constructed from APLNR gene and orthologs

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree. *Gorilla gorilla gorilla*, *homo sapiens* and *paniscus* and *pan troglodytes* form one tight cluster indicating close evolutionary relationship between them. *Pongo pygmaeus* and *pongo abelii* are closely related to each other, *macaca fascicularis* diverged long time ago.

MAPK

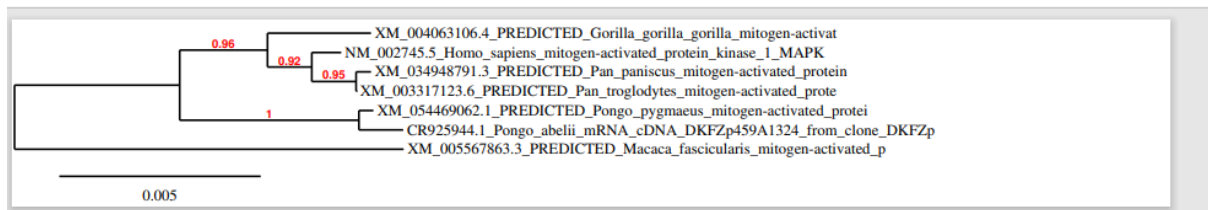


Figure 2.16 Phylogenetic tree MAPK

Phylogenetic tree constructed from MAPK gene and orthologs

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree. The MAPK genes of Gorilla, Homo sapiens, pan paniscus and pan troglodytes form a tight cluster indicating they are all closely evolutionarily related to each other. The pongo pygmaeus and pongo abelii form another tight cluster indicating they are all evolutionarily related to each other. *Macaca fascicularis* diverged from these genes a long time ago.

KLF5

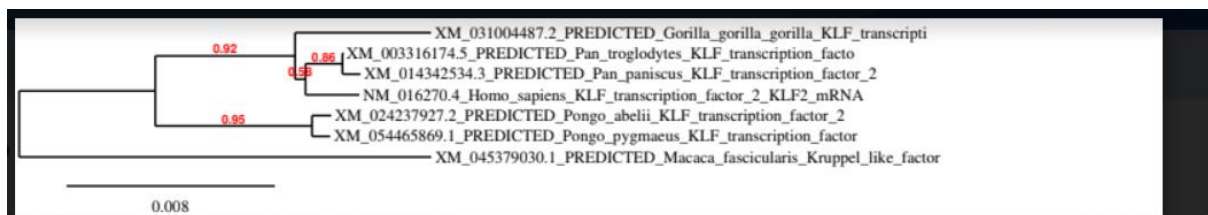


Figure 2.17 Phylogenetic tree KLF5

Phylogenetic tree constructed from KLF5 gene and orthologs

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree.

The KLF transcription factor of *Homo sapiens* and those of *Pan troglodytes*, *Pan paniscus*, and *Gorilla gorilla* form a tight cluster, indicating a close evolutionary relationship between these species. This is further supported by the high bootstrap values (0.92 and 0.95) on the branches connecting them.

The KLF transcription factor of *Macaca fascicularis* (crab-eating macaque) branches out from the lineage leading to the apes (human, chimpanzee, bonobo, gorilla, orangutans) much earlier. This suggests that the KLF transcription factor of the crab-eating macaque diverged from the evolutionary lineage leading to apes much earlier than the other species included here.

The KLF transcription factor of humans share a close evolutionary relationship with those of chimpanzees, bonobos, and gorillas, while the KLF transcription factor of the crab-eating macaque diverged from this lineage much earlier.

MC3R

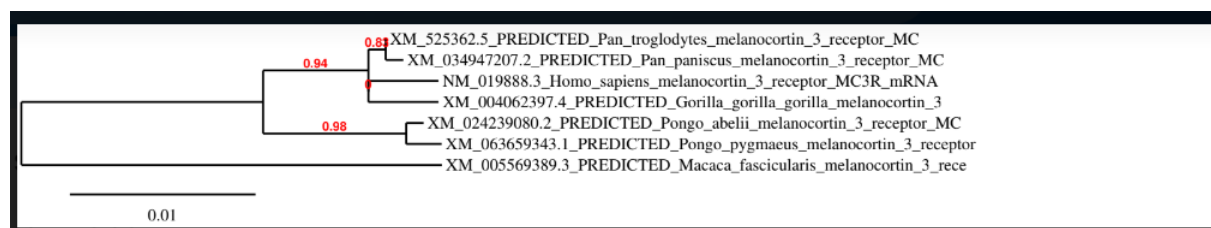


Figure 2.18 Phylogenetic tree MC3R

Phylogenetic tree constructed from MC3R gene and orthologs

It is observed from this figure that *Macaca fascicularis* is more distantly related than all the other organisms. Bootstrap values are above 90% and this indicates strong support for the branches of the tree. The tree shows that the *Homo sapiens* MC3R gene is most closely related to

the *Pan troglodytes* MC3R gene with a bootstrap value of 0.83. This suggests that the human and chimpanzee MC3R genes share a recent common ancestor.

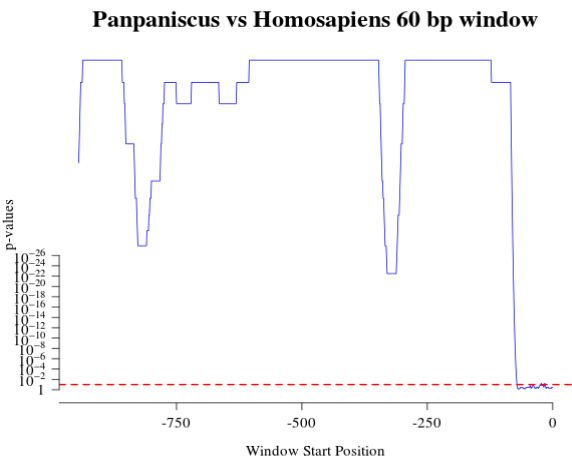
The next most closely related gene is the *Pan paniscus* MC3R gene with a bootstrap value of 0.94. The gorilla *gorilla* MC3R gene and the *Pongo abelii* MC3R gene are more distantly related.

Overall, this phylogenetic tree suggests that the MC3R gene is relatively well conserved among primates. The human MC3R gene is most closely related to the chimpanzee and bonobo MC3R genes, and more distantly related to the gorilla and orangutan MC3R genes.

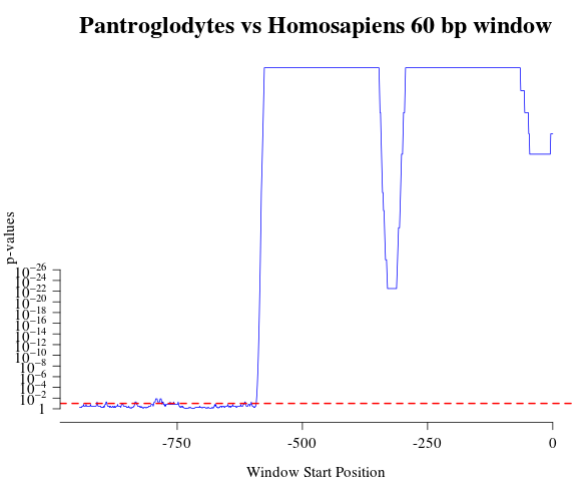
5.5 Evolutionary analysis of regulatory sequences results

The Fasta sequences of human promoter and orthologous promoters were analysed using Evolutionary analysis of regulatory sequences to find conserved non coding sequences. The parameters used were window size 60 bp and p value of 0.1.

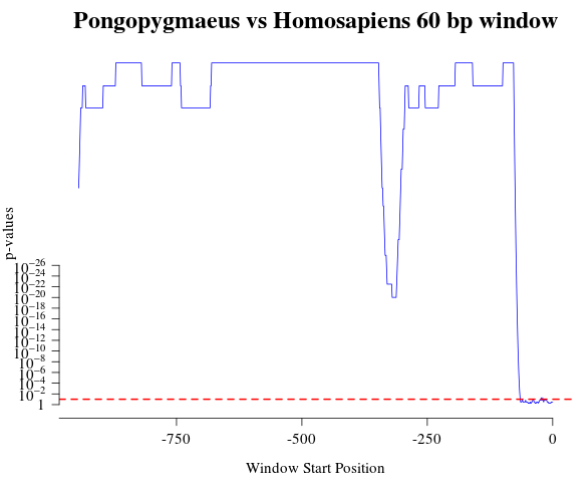
ACADM



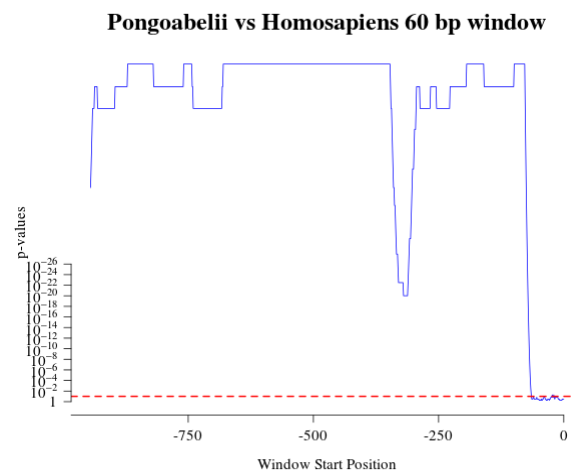
(a)



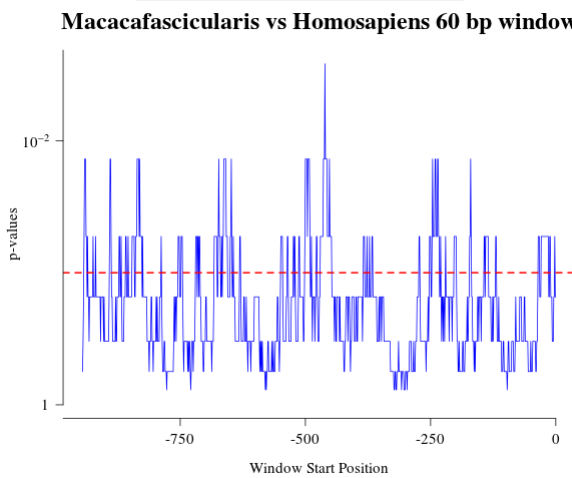
(b)



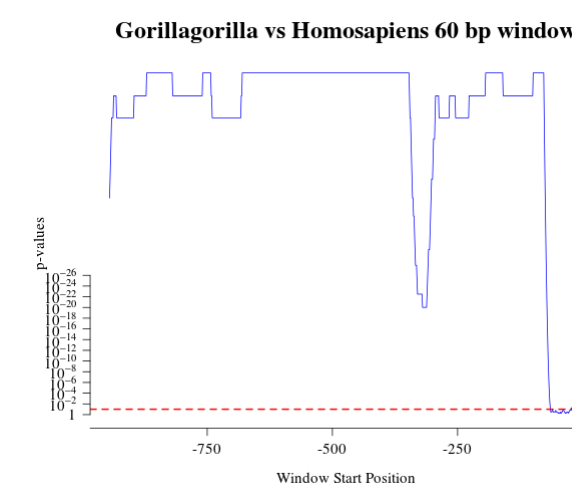
(c)



(d)



(e)



(f)

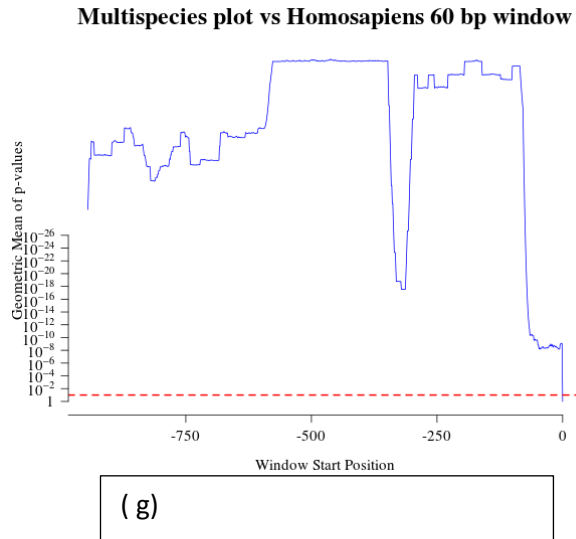
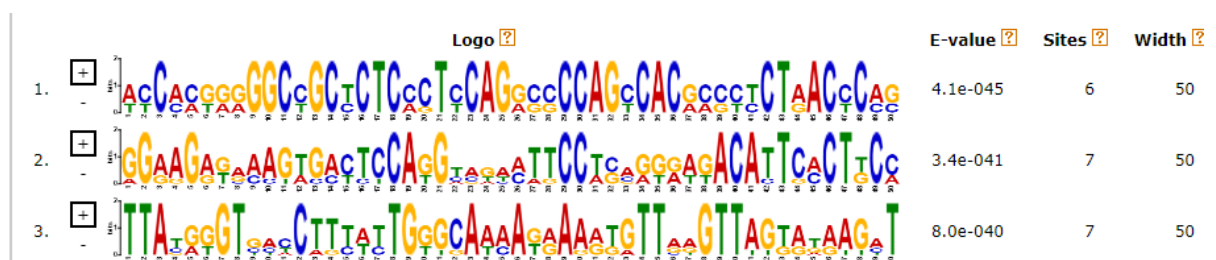


Figure no.: 3.1 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

The peaks are present above the significance threshold. In figure (e) there is presence of CNS1,CNS2 and CNS3

Identification of motifs is done using multiple for motif enrichment software.atleast 3 motiffs can be identified from the results.



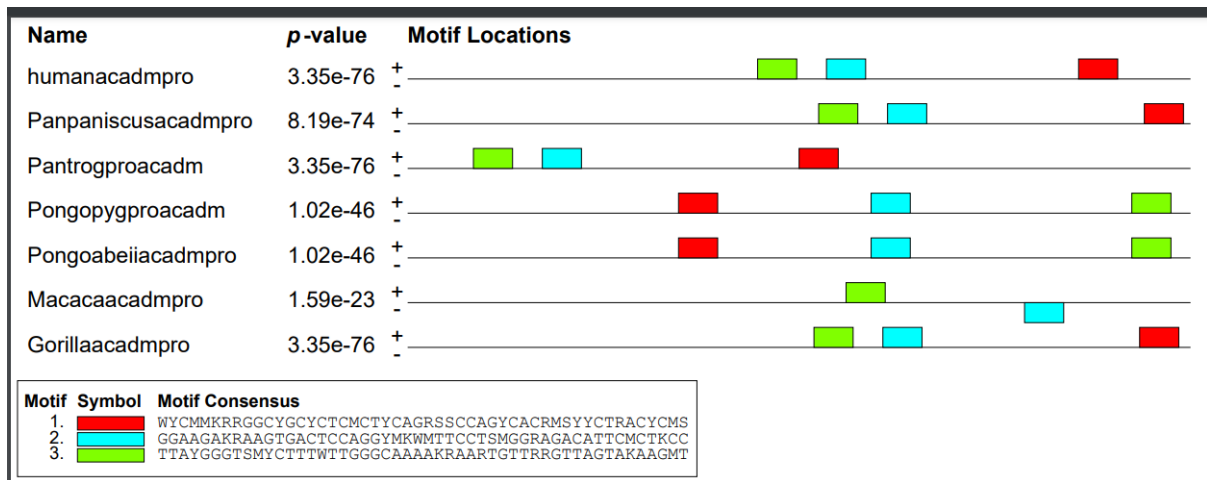
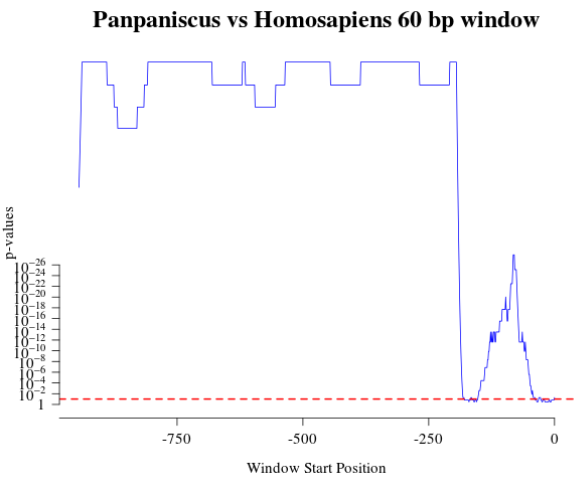


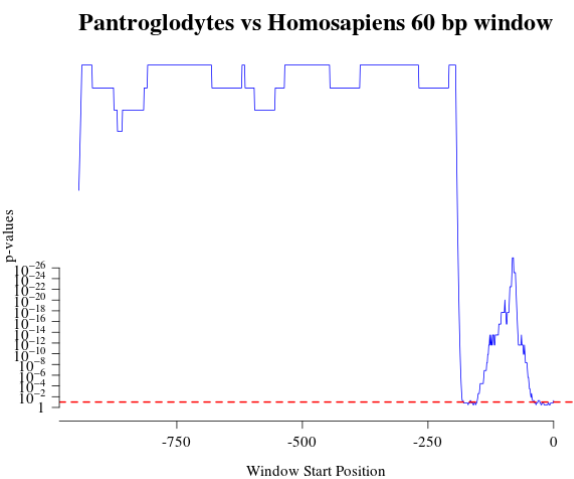
Figure 4.1 motif identification acadm Identification of motifs(a) motif logos (b) motif location Homosapiens, Pan paniscus, Pan troglodytes, Pongo pygmaeus, Pongo abelii and Gorilla gorilla share the presence of common three motifs.

Motif 1 in human starts at 858 and ends at 907. Motif 2 starts at 536 and ends at 585. Motif 3 starts at 448 and ends at 497. In pan paniscus, the motif 1 starts at 942 and ends at 991. The motif 2 starts at 614 and ends at 663. The motif 3 starts at 526 and ends at 575. In pan troglodytes, the motif 1 starts at 501 and ends at 550. The motif 2 starts at 173 and ends at 222. The motif 3 starts at 85 and ends at 134. In Pongo pygmaeus, motif 1 starts at 347 and ends at 396. The motif 2 starts at 593 and ends at 642. The motif 3 starts at 926 and ends at 975. In Pongo abelii, motif 1 starts at 347 and ends at 396. Motif 2 starts at 593 and ends at 642. Motif 3 starts at 926 and ends at 975. In Gorilla gorilla motif 1 starts at 936 and ends at 985. Motif 2 starts at 608 and ends at 657. Motif 3 starts at 520 and ends at 569. Homosapiens, Pan paniscus, Pan troglodytes, Pongopygmaeus, pongo abelii and gorilla share the presence of three common motifs. They have their motifs on the forward strand. Pongo pygmaeus and Pongo abelii motif starting and ending positions of the motifs are same.

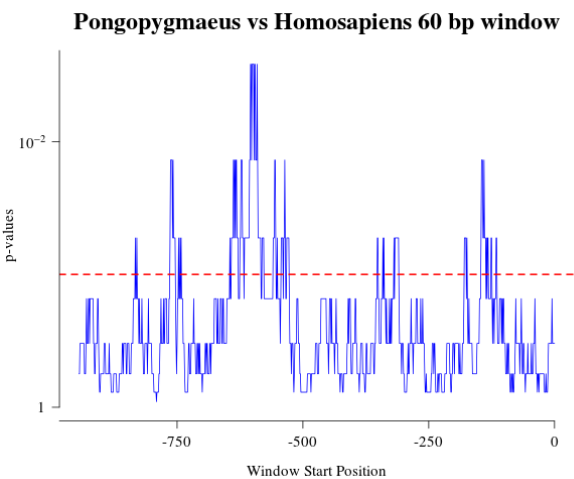
CPT1A



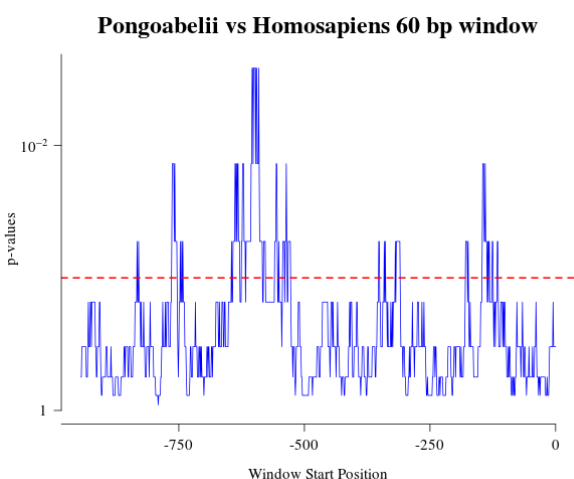
(a)



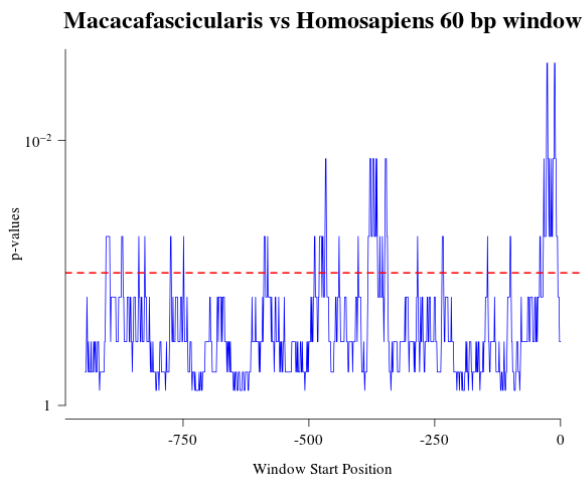
(b)



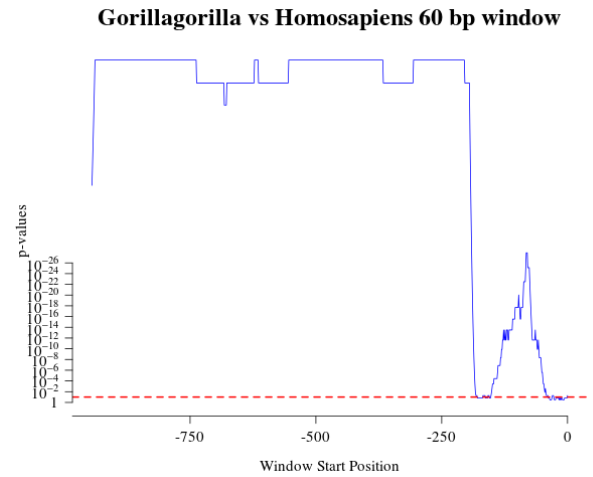
(c)



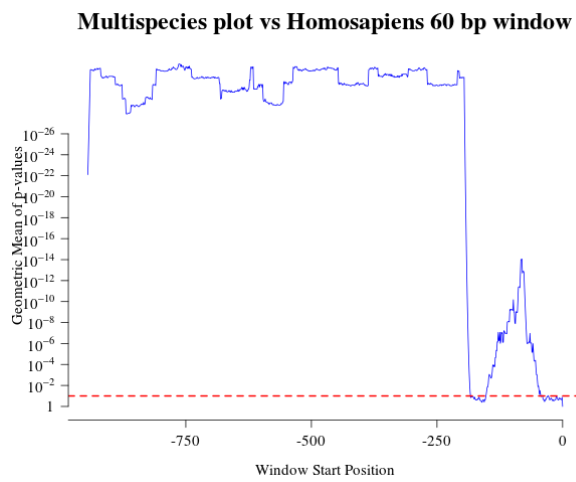
(d)



(e)



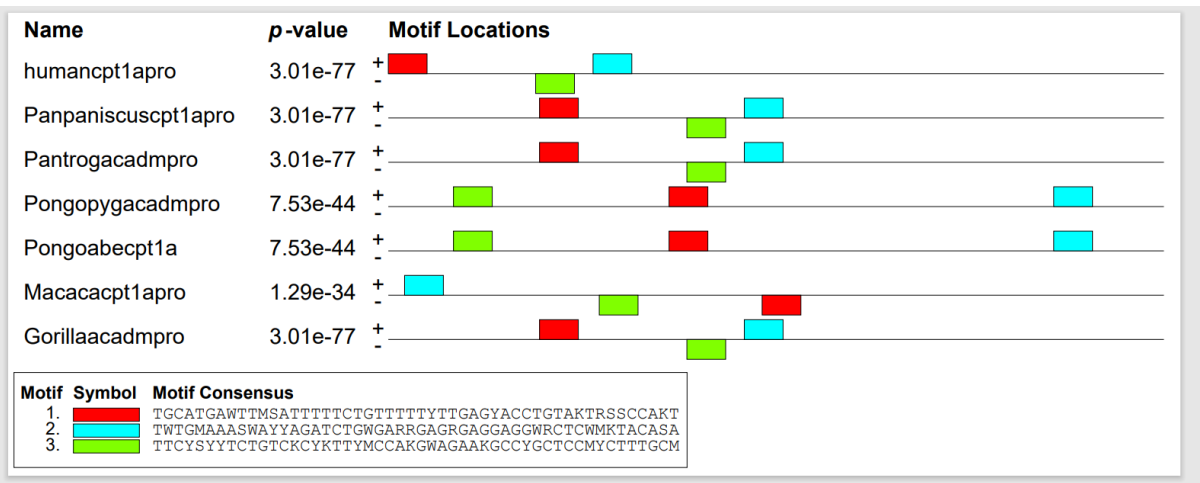
(f)



(g)

Figure no.:3.2 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

	Logo	E-value	Sites	Width
1.		6.0e-041	7	50
2.		1.7e-040	7	50
3.		3.9e-040	7	50



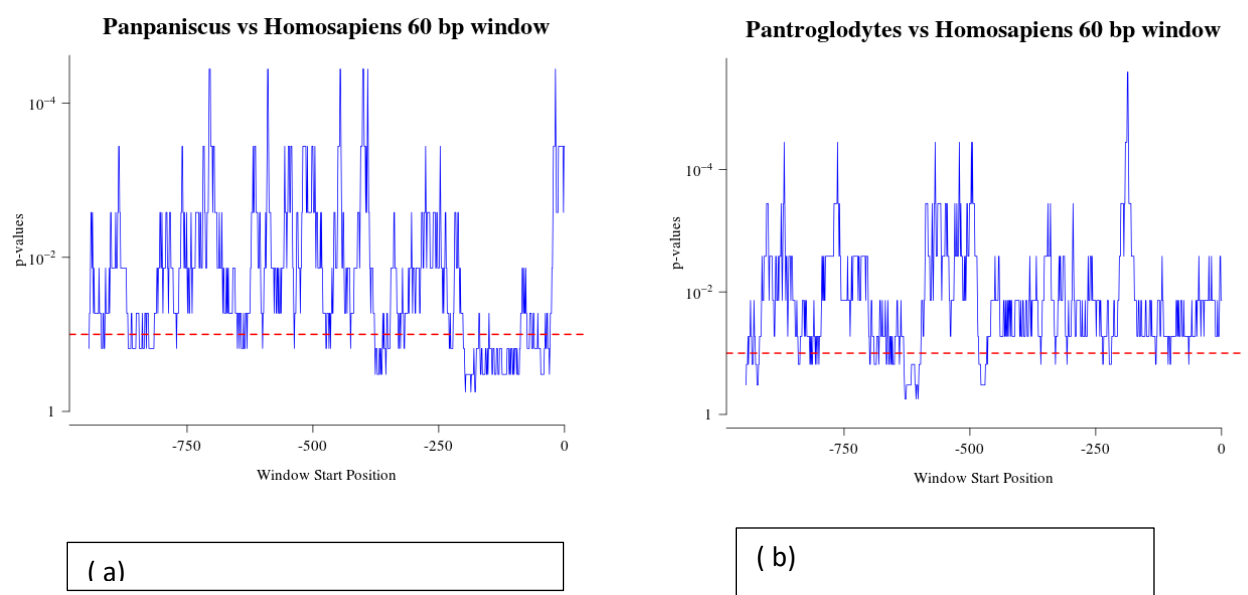
The presence of three motifs is observed in human and all the other orthologs. In human, the start site of the first motif was 1 and the end site was 50. The second motif start site was 264 and the end site was 314. The third motif start site was -191 and the end site was -240. In *panpaniscus* the start site of the first motif was 196 and the end site of the motif was 245. The start site of the second motif was 460 and the end site was 509. The start site of the third motif was -386 and the end site was -415. In *pan troglodytes* the start site of the first motif was 196 and the end site was 245. The start of the second motif was 460 and the end site of the second motif was 509. The start of the third motif was -386 and the end of the motif was at 435. In *pongopygmaeus* the start of the first motif was at 196 and the end at 245. The start of the second

motif was at 363 and the end of the first motif was at 312. The start of the third motif was at -386 and the end was at 435.

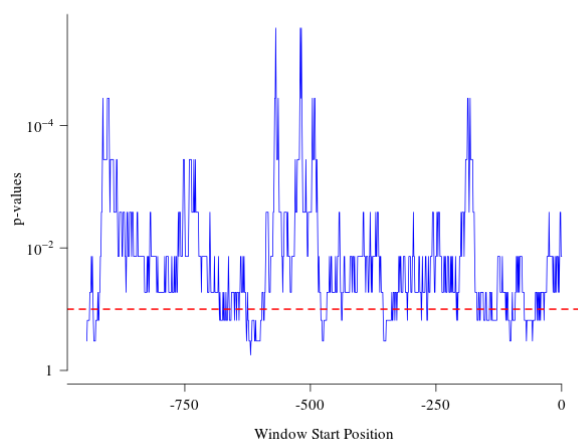
In *pongoabelii* the start of the first motif was at 85 and the end of the motif was at 134. The start of the second motif was at 363 and the end was at 412. The start of the third motif was at 22 and the end of the third motif was at 71. The start of the third motif was at 859 and the end was at 908.

In *macaca fascicularis* the start of the first motif was at 22 and the end was at 71. The start of the second motif was at 460 and the end was at 509. The start of the third motif was at -386 and the end position was at -435. *Pan paniscus* and *pan troglodytes* have motif start and end positions as same. *pongo abelii* start of third motif and *macaca* start of first motif same

DGAT1

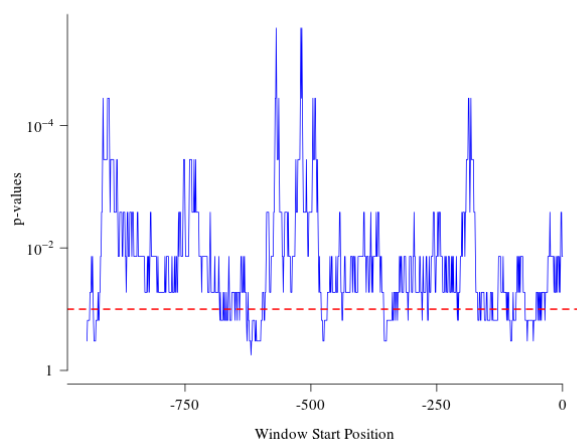


Pongopygmaeus vs Homosapiens 60 bp window



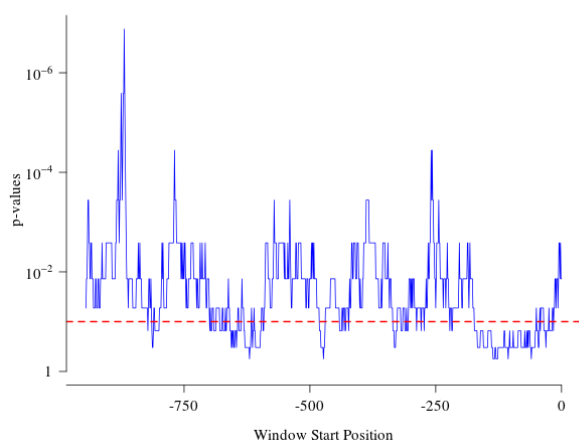
(c)

Pongoabelii vs Homosapiens 60 bp window



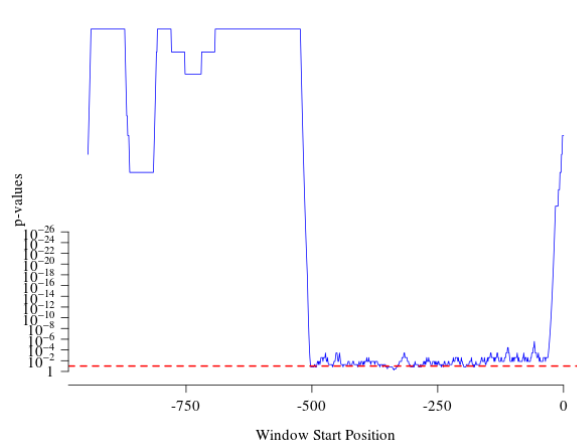
(d)

Macacafascularis vs Homosapiens 60 bp window



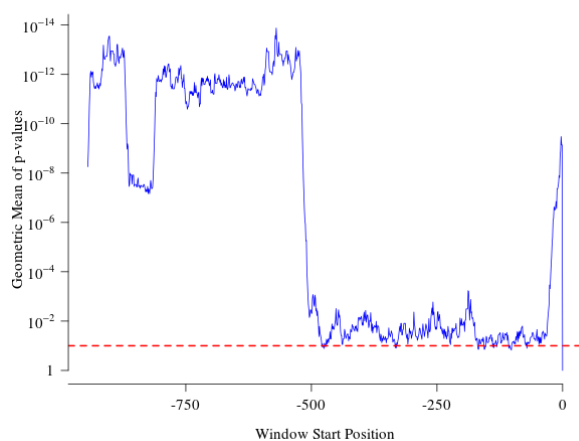
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure no.: 3.3 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

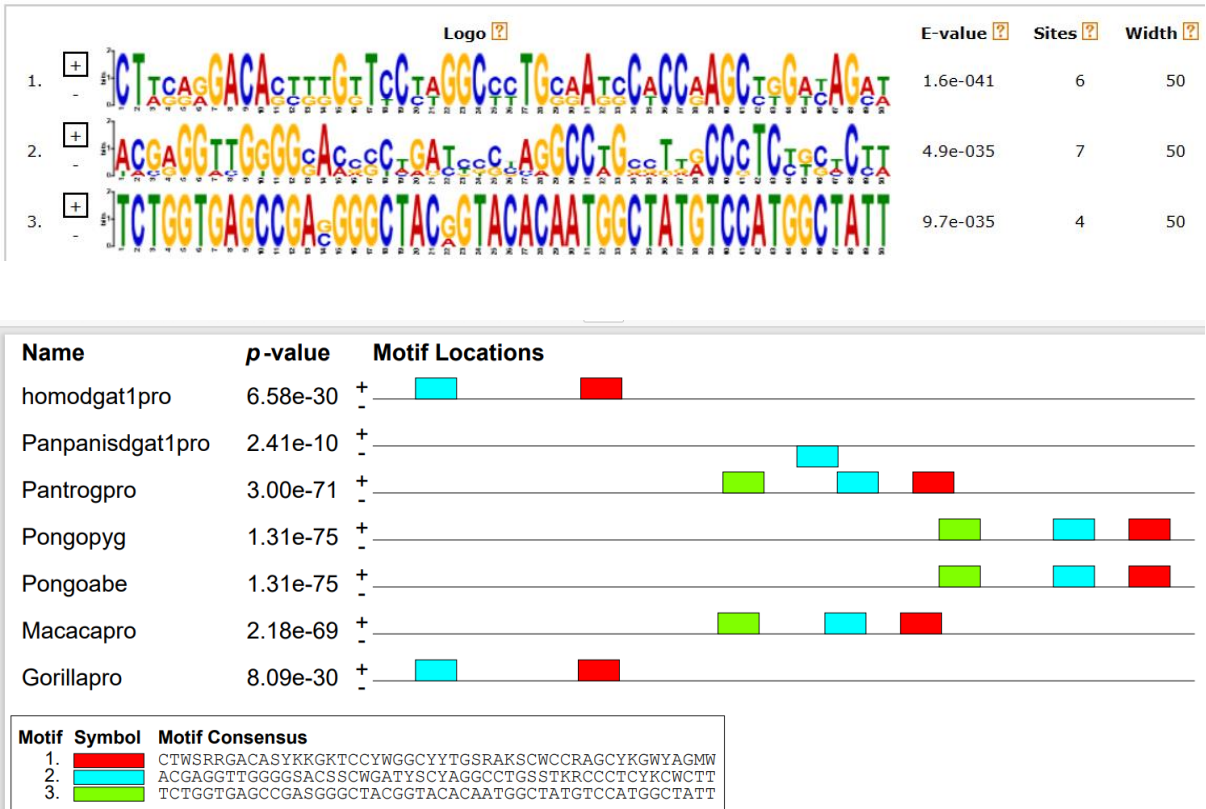
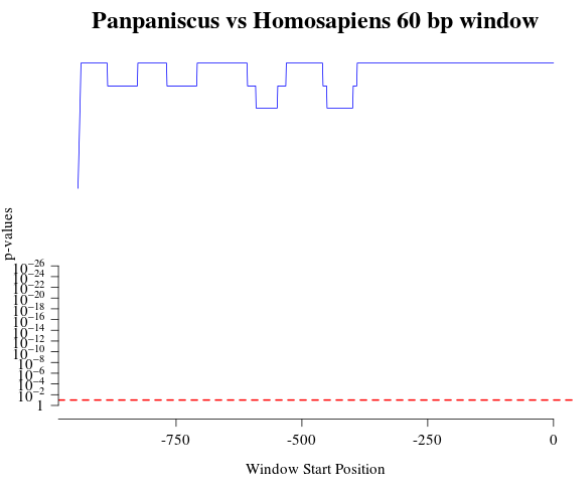


Figure 4.3 Motif identification dgat1

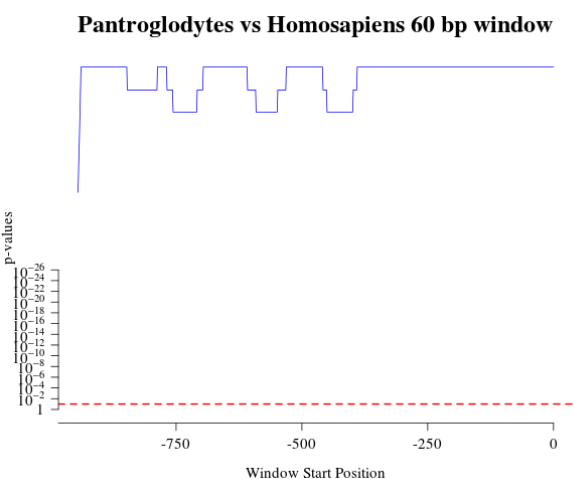
Pan troglodytes, Pongo pygmaeus, Pongo abelii and Macaca fascicularis share the presence of three common motifs. In Pan troglodytes the motif 1 starts at 658 and ends at 707. The motif 2 starts at 566 and ends at 615. The motif 3 starts at 427 and ends at 476. In Pongo pygmaeus the motif 1 starts at 921 and ends at 970. The motif 2 starts at 829 and ends at 878. The motif 3 starts at 690 and ends at 739. In Pongo abelii, motif 1 starts at 921 and ends at 970. The motif 2 starts at 829 and ends at 878. The motif 3 starts at 690 and ends at 739.

In *Macaca fascicularis* motif 1 starts at 643 and ends at 692. The motif 2 starts at 551 and ends at 600. The motif 3 starts at 421 and ends at 470. *Pan troglodytes*, *Pongo pygmaeus*, *Pongo abelii* and *Macaca fascicularis* share the presence of three common motifs. *Pongo pygmaeus* and *Pongo abelii* motif start and end sites are the same.

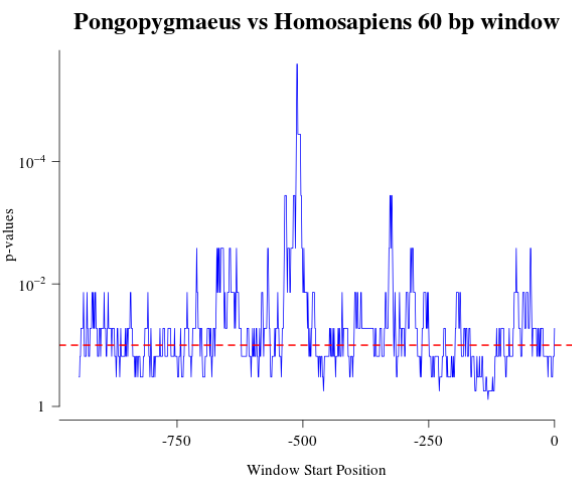
SLC25A1



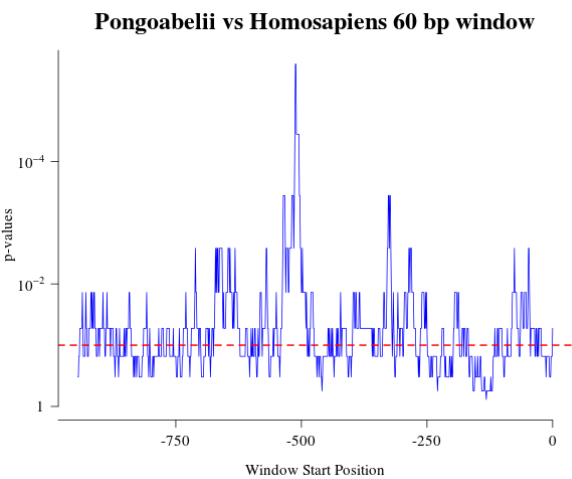
(a)



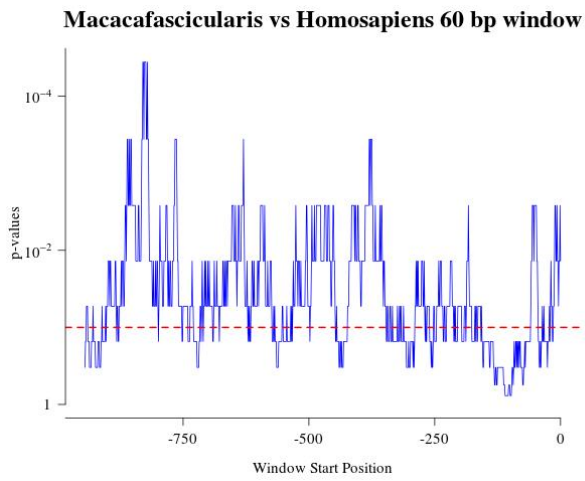
(b)



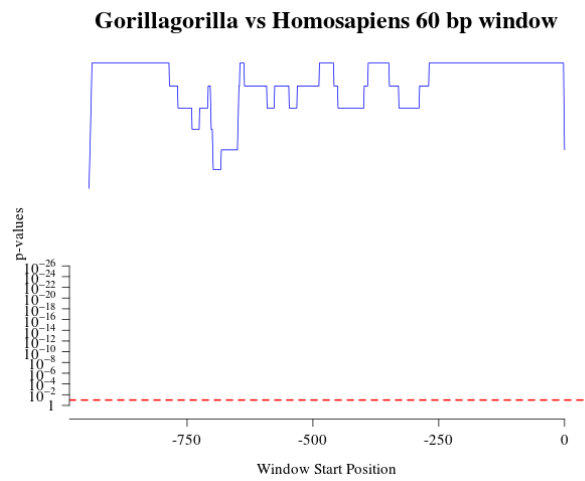
(c)



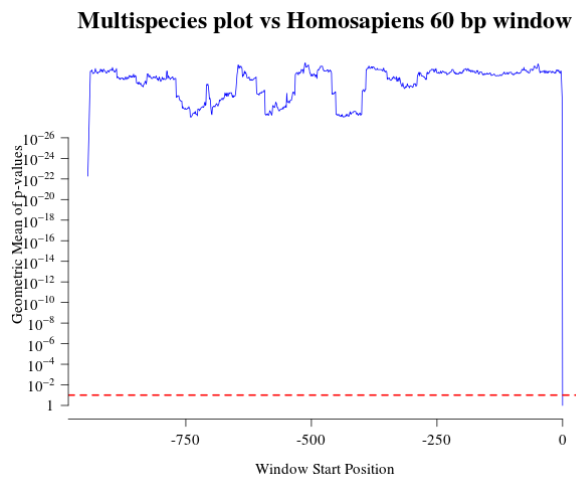
(d)



(e)



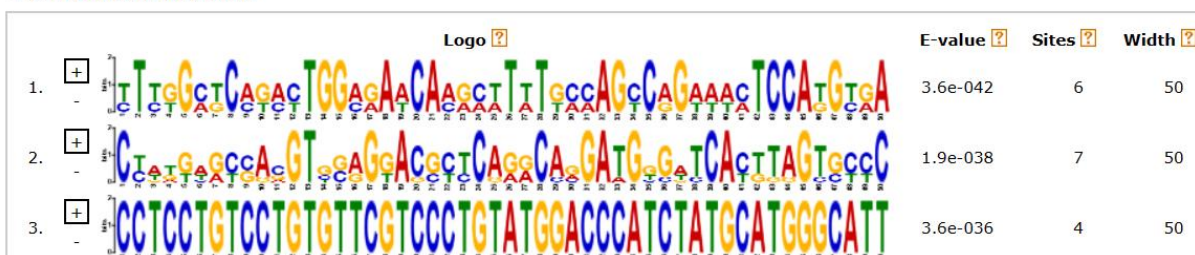
(f)



(g)

Figure no.:3.4 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

DISCOVERED MOTIFS



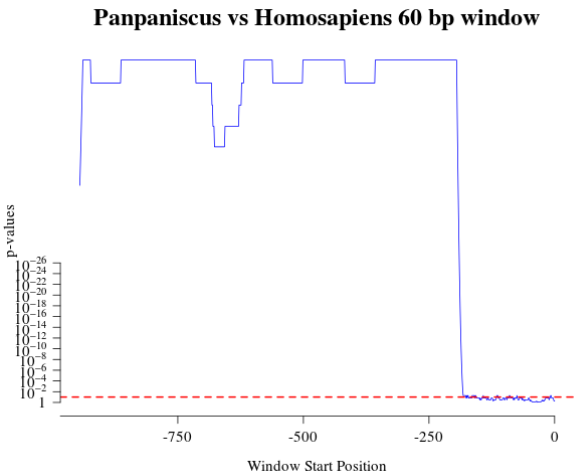
Name	p-value	Motif Locations
homoslc25a1	2.29e-75	
Panpanisslc25a1	2.29e-75	
Pantrogalslc25a1	2.29e-75	
Pongopygslc25a1pro	6.79e-31	
Pongoabeslc25a1pro	6.79e-31	
Macacaslsc25a1pro	2.62e-9	
Gorillaslc25a1pro	2.29e-75	

Motif	Symbol	Motif Consensus
1.		YTYKGMKCMKMYTGGMRWCAMRMWTTWKMMAGYCRGWWMTTCCAKGYRA
2.		CYWKGWGCASGTSSAGGACGCTCAGRCARGATGKGWTCACCTAGTGCCC
3.		CCTCCTGTCTGTGTTCTCCCTGTATGGACCCATCTATGCATGGGCATT

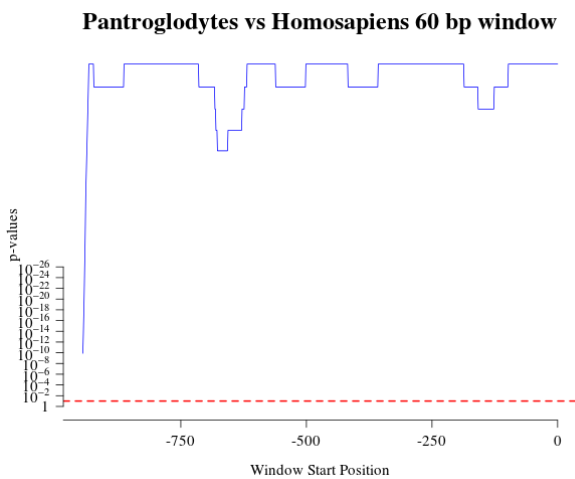
Figure 4.4 motif identification slc25a1

Homosapiens, Pan paniscus, Pan troglodytes and Gorilla gorilla share the occurrence of common three motifs. In homo sapiens motif 1 starts at 841 and ends at 890. The motif 2 starts at 727 and ends at 776. The motif 3 starts at 2 and ends at 57. In Pan paniscus Motif 1 start at 841 and end at 890. Motif 2 starts at 727 and ends at 776. In Pan troglodytes. Motif 1 starts at 841 and ends at 890. Motif 2 starts at 727 and ends at 776. Motif 3 starts at 2 and ends at 51. In Gorilla gorilla motif 1 starts at 843 and ends at 892. The motif 2 starts at 729 and ends at 778. The Motif 3 starts at 2 and ends at 51. Homosapiens, Pan paniscus, Pan troglodytes have motif start and end postions as same

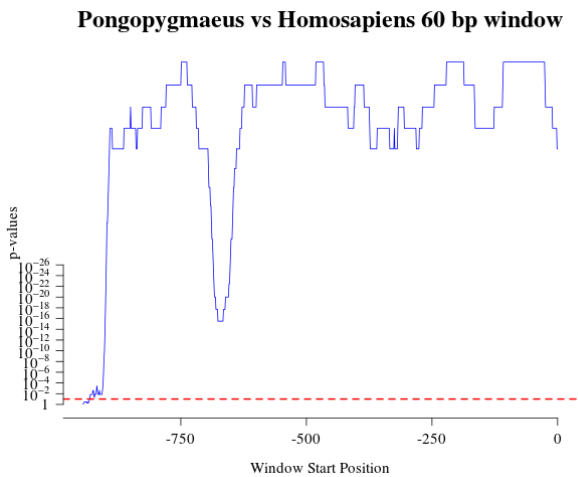
HADH



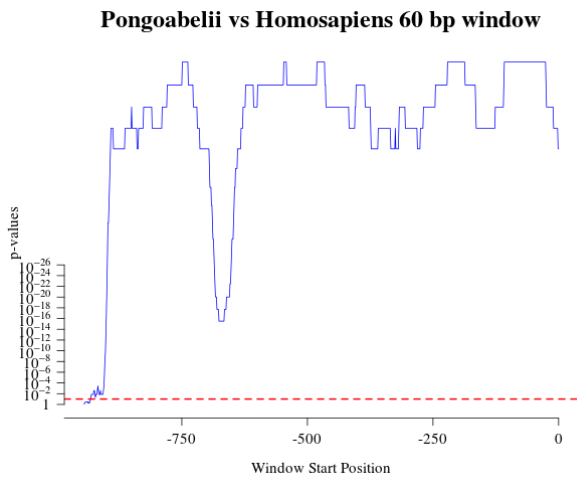
(a)



(b)

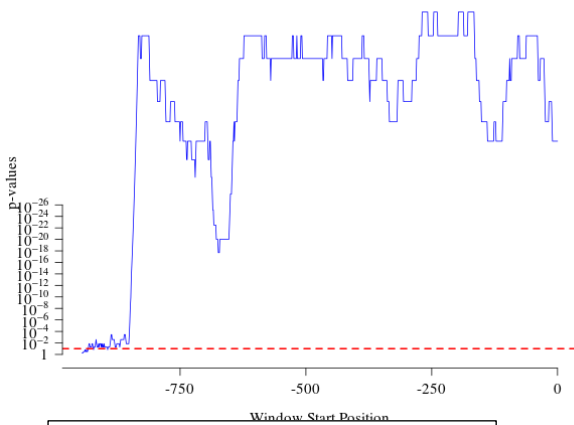


(c)



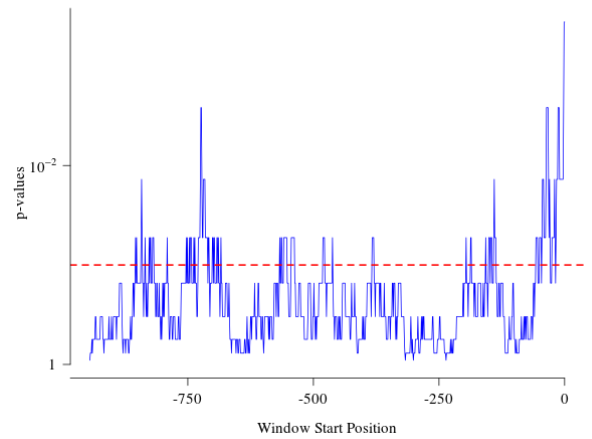
(d)

Macacafascicularis vs Homosapiens 60 bp window



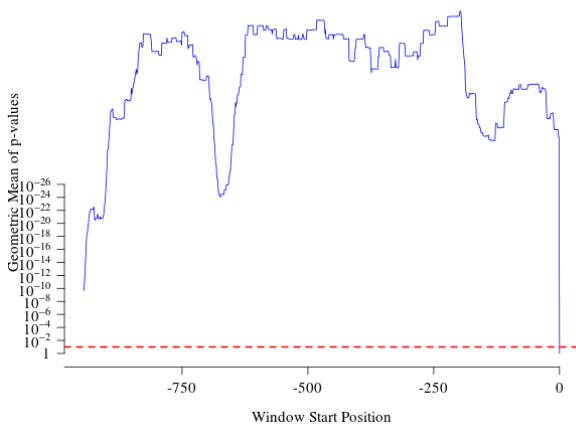
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure : 3.4 Evolutionary analysis of conserved sequences 1 (a) Homosapiens vs pan panisucs 1(b) homosapiens vs pan troglodytes 1(c) homosapiens vs pongo pygmaeus 1 (d) homosapiens vs pongo abelii 1 (e) homosapiens vs macaca fascicularis 1(f) homosapiens vs gorilla gorilla 1 (g) multispecies plot

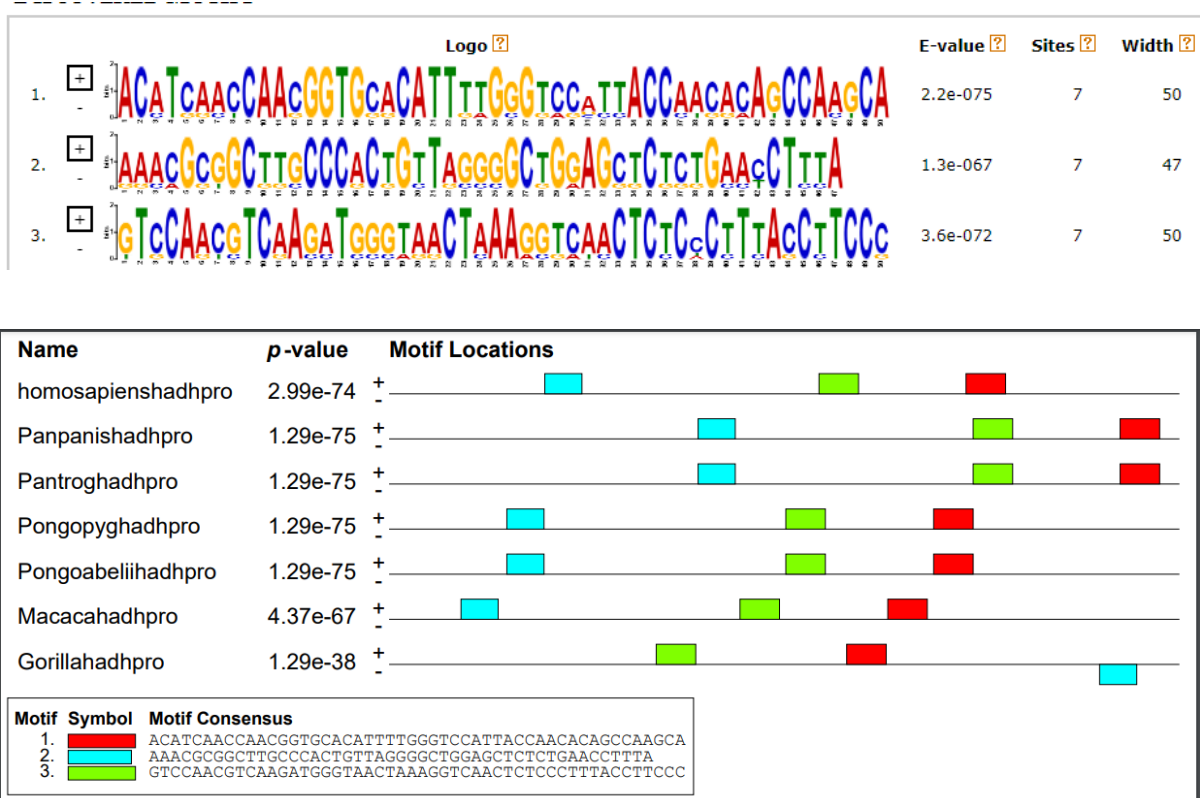
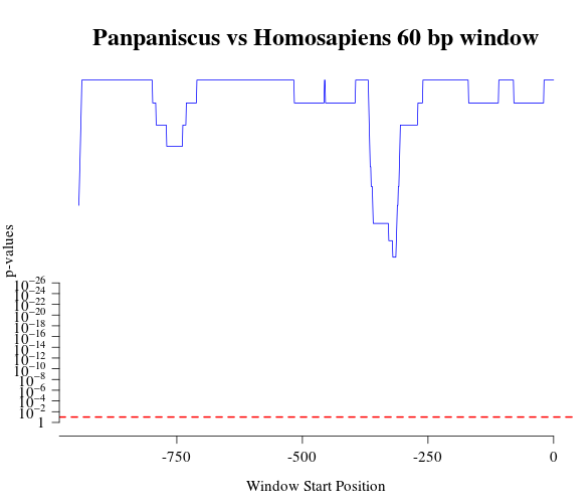


Figure 4.4 motif identification HADH

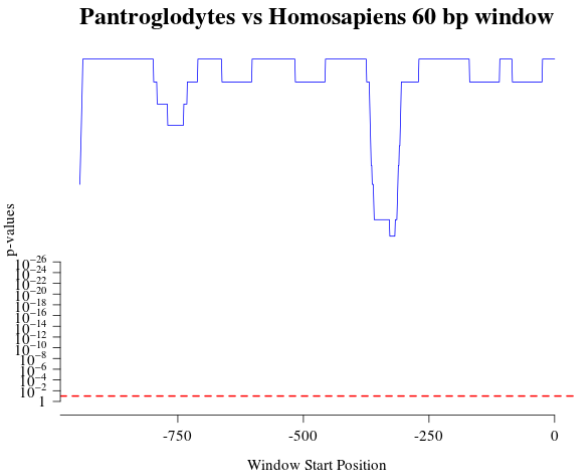
In Homosapiens, Pan paniscus, Pan troglodytes, Pongo pygmaeus, Pongo abelii, Macaca fascicularis and Gorilla gorilla presence of three motifs observed, In Homo sapiens motif 1 starts at 731 and ends at 780. Motif 2 starts at 198 and ends at 244. Motif 3 starts at 731 and ends at 780. In Pan paniscus Motif 1 starts at 926 and ends at 975. Motif 2 starts at 392 and ends at 438. Motif 3 starts at 740 and ends at 789. In Pan troglodytes, Motif 1 starts at 926 and ends at 975. Motif 2 starts at 392 and ends at 438. Motif 3 starts at 740 and ends at 789. In Pongo pygmaeus motif 1 starts at 690 and ends at 739. Motif 2 starts at 150 and ends at 196. Motif 3 starts at 503 and ends at 552. In Macaca fascicularis, motif 1 starts at 632 and ends at 681. Motif 2 starts at 92 and ends at 138. Motif 3 starts at 445 and ends at 494. In Gorilla gorilla motif 1 starts at 580 and ends at 629. Motif 2 starts at -900 and ends at

-946.Motif 3 starts at 339 and ends at 388. Except gorilla all animals have motifs on forward strand.

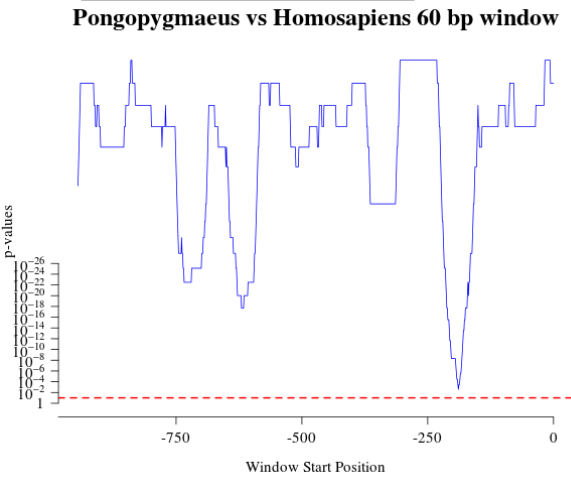
PKM



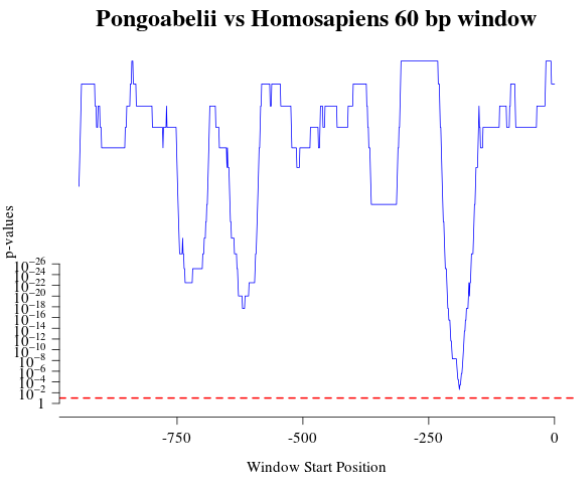
(a)



(b)

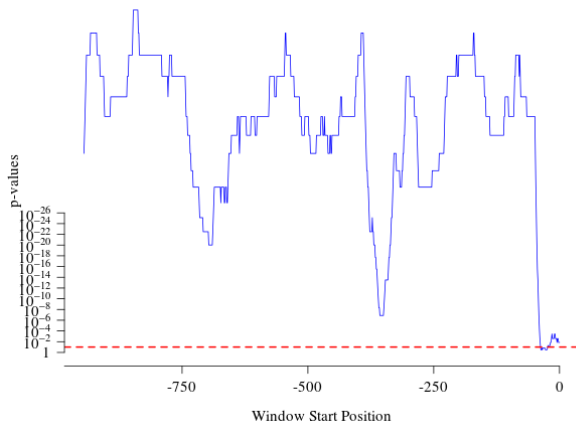


(c)



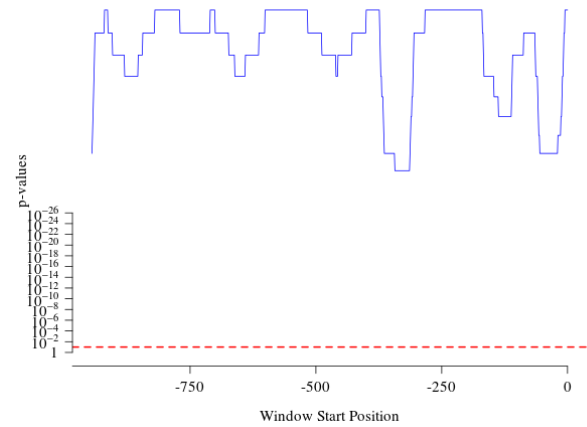
(d)

Macacafascicularis vs Homosapiens 60 bp window



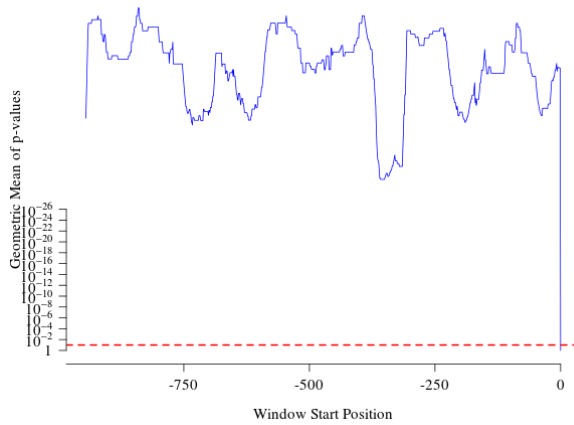
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure no.: 3.6 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs



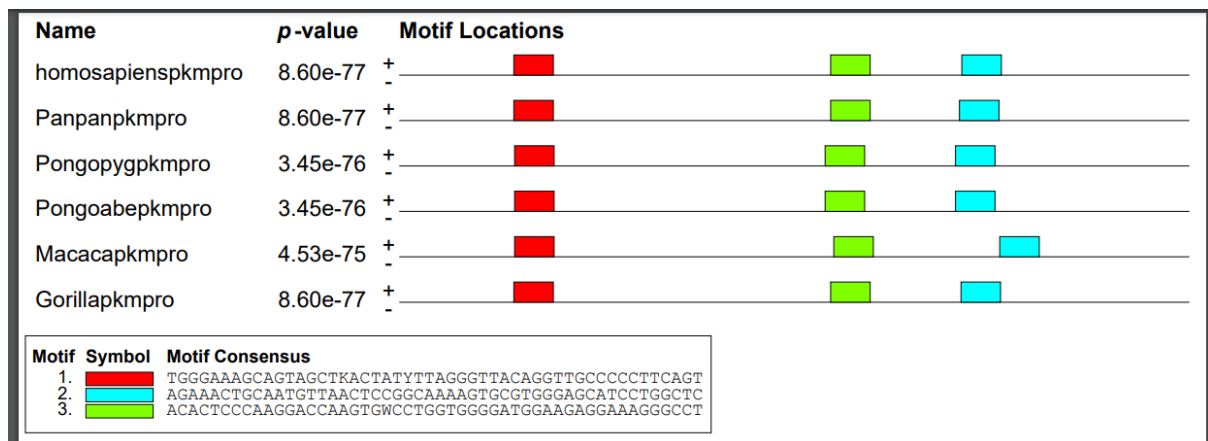
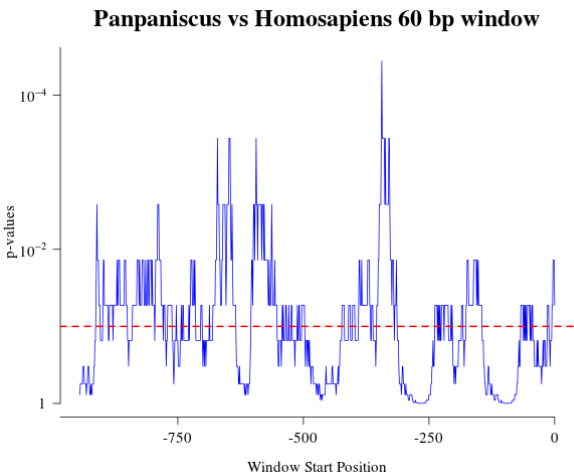


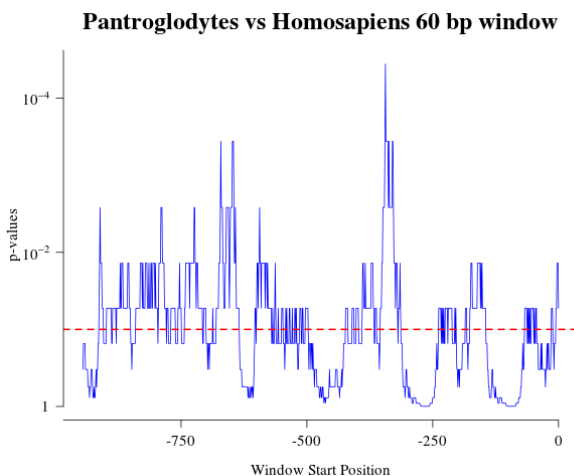
Figure 4.6 motif identification PKM

In Homosapiens and all orthologs three motifs have been observed. In Homo sapiens motif 1 starts at 146 and ends at 195. motif 2 starts at 713 and ends at 762. Motif 2 starts at 713 and ends at 762. Motif 3 starts at 547 and ends at 596. In Pan paniscus motif 1 starts at 146 and ends at 195. Motif 2 starts at 710 and ends at 759. Motif 3 starts at 547 and ends at 596. In Pongo pygmaeus motif 1 starts at 147 and ends at 196. Motif 2 starts at 705 and ends at 754. Motif 3 starts at 540 and ends at 589. In Pongo abelii motif 1 starts at 147 and ends at 196. Motif 2 starts at 705 and ends at 754. Motif 3 starts at 540 and ends at 589. In Macaca fascicularis motif 1 starts at 147 and ends at 196. Motif 2 starts at 705 and ends at 754. Motif 3 starts at 540 and ends at 589. In Gorilla gorilla motif 1 starts at 146 and ends at 195. Motif 2 starts at 712 and ends at 761. Motif 3 starts at 547 and ends at 596. Motif 2 is present in the same position in all the animals

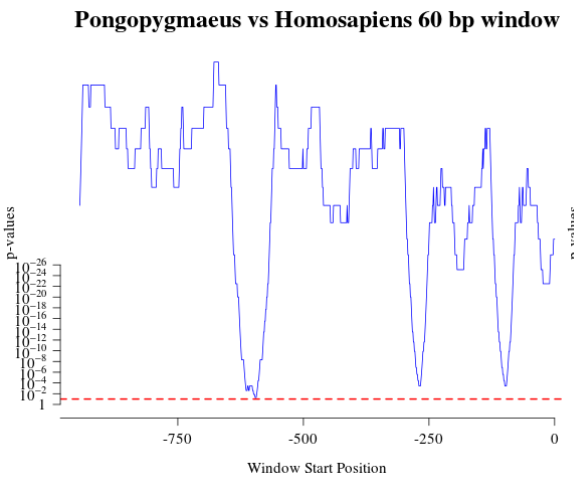
ACLY



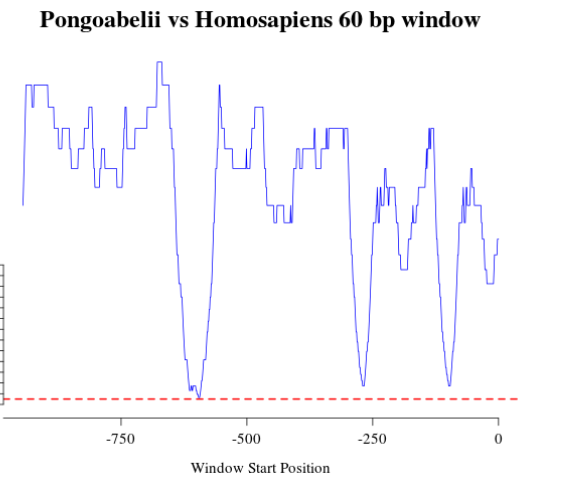
(a)



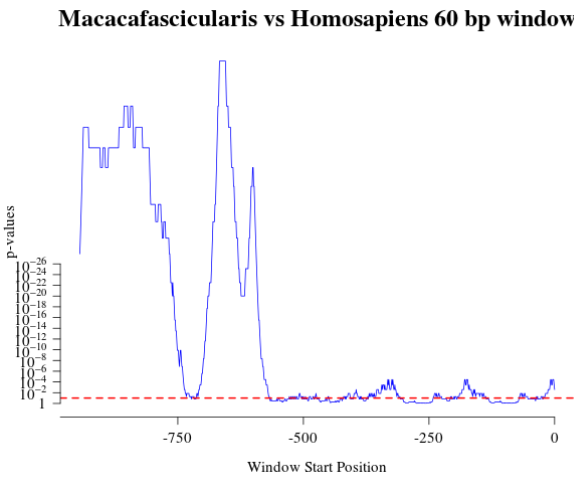
(b)



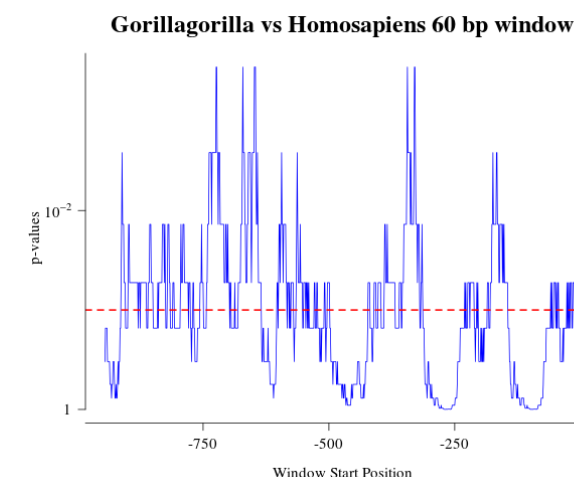
(c)



(d)

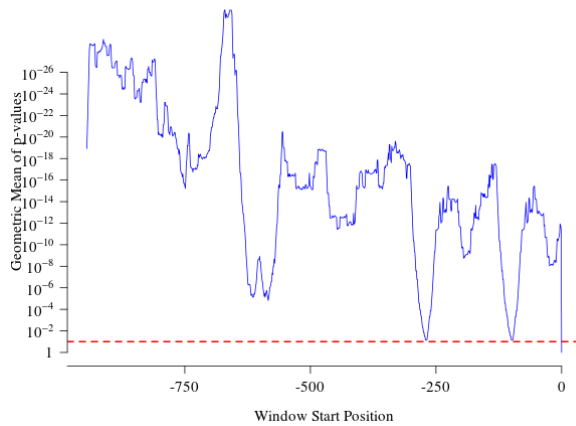


(e)



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure no. 3.7 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

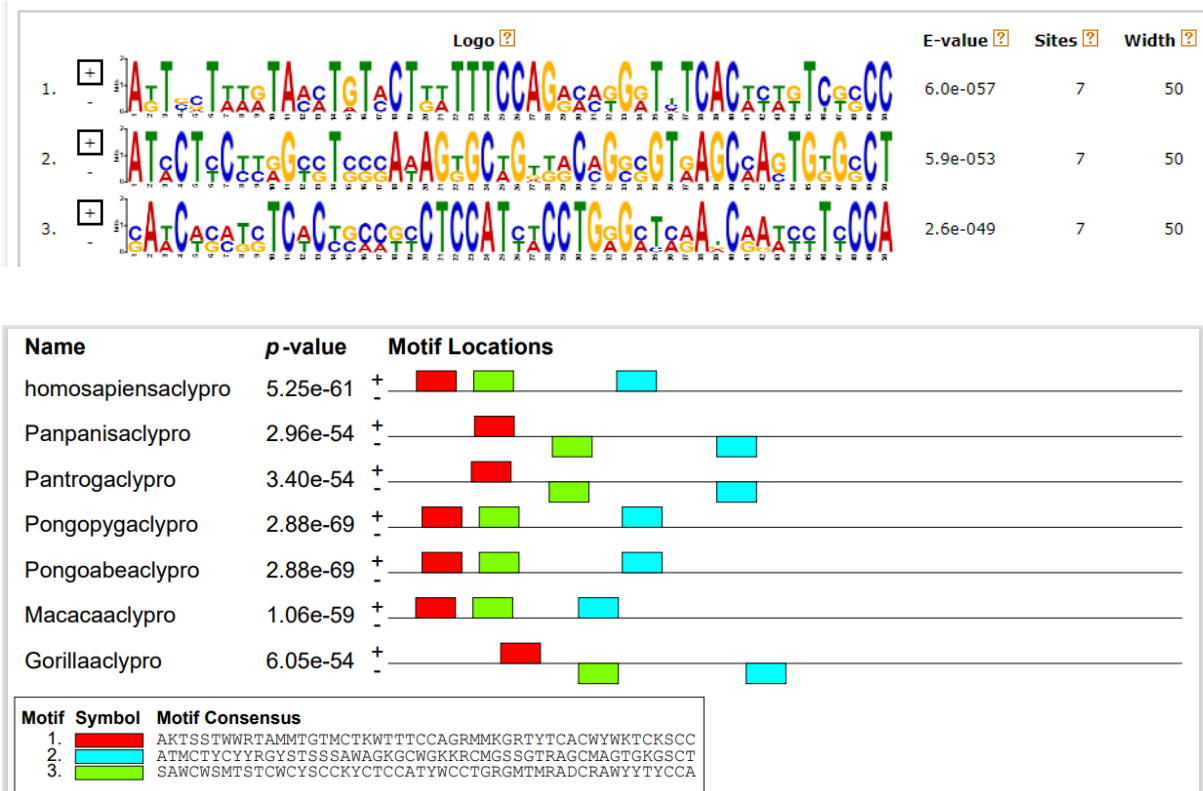
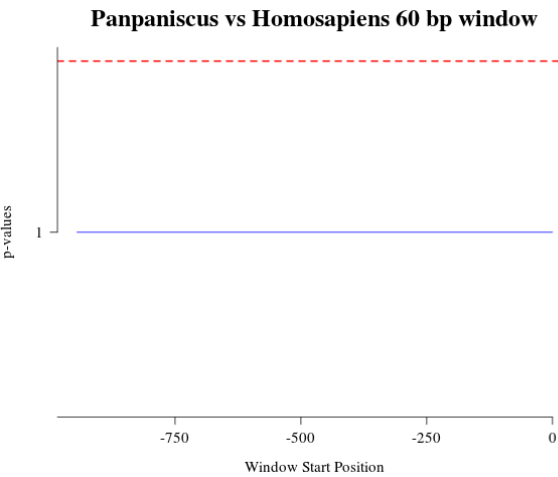


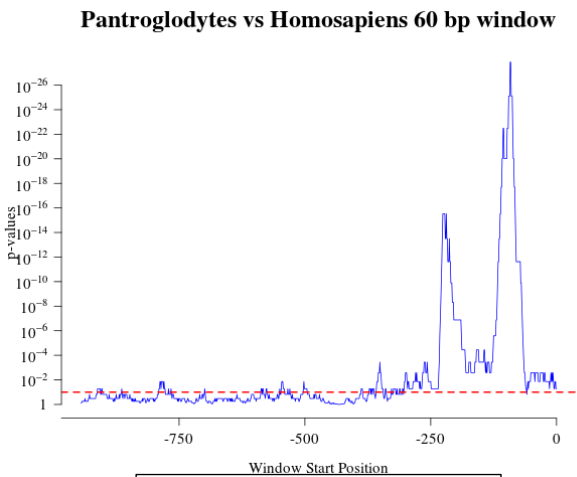
Figure 4.7 Motif identification ACLY

Homosapiens motif 1 starts at 37 and ends at 86 . motif 2 starts at 289 and ends at 338 . motif 3 starts at 109 and ends at 158 . Panpaniscus motif 1 starts at 110 and ends at 159 . Motif 2 starts at 415 and ends at 464. Motif 3 starts at 208 and ends at 257. In Pan troglodytes motif 1 starts at 106 and ends at 155. Motif 2 starts at 415 and ends at 464. Motif 3 starts at 204 and ends at 253, In Pan troglodytes motif 1 starts at 106 and ends at 155. Motif 2 starts at 415 and ends at 464 . Motif 3 starts at 204 and ends at 253 . In Pongo pygmaeus motif 1 starts at 44 and ends at 93 . Motif 2 starts at 296 and ends at 345 . Motif 3 starts at 116 and ends at 165 . In Pongo abelii motif 1 starts at 44 and ends at 93 . Motif 2 starts at 296 and ends at 345. Motif 3 start at 116 and ends at 165 . In Macaca fascicularis motif 1 starts at 36 and ends at 85. Motif 2 starts at 241 and ends at 290. Motif 3 starts at 108 and ends at 157 . In Gorilla gorilla motif 1 starts at 143 and ends at 192 . Motif 2 starts at 452 and ends at 501 . Motif 3 starts at 241 and ends at 290 . Homo sapiens, Pongo pygmaeus, Pongo abelii and macac fascicularis share presence on forward strand. Pongo pygmaeus and Pongo abelii motif positions are same

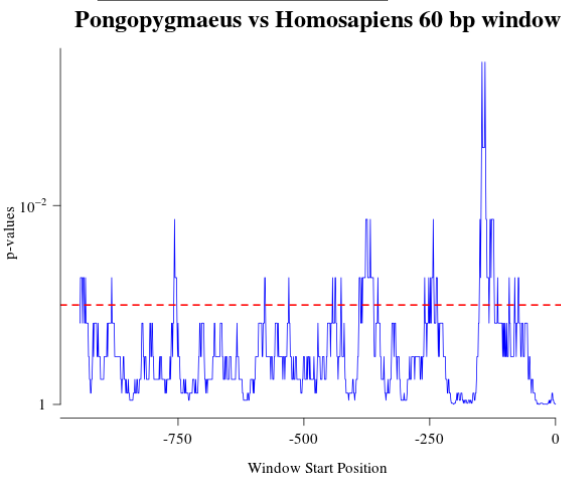
ACACA



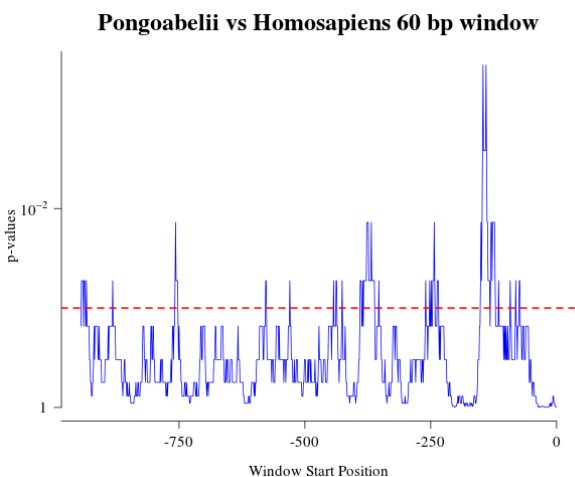
(a)



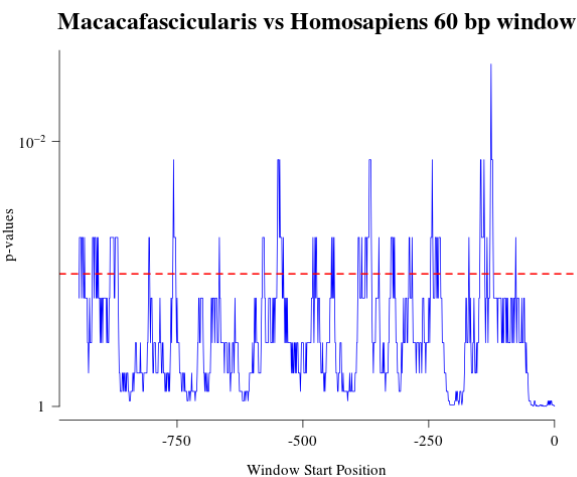
(b)



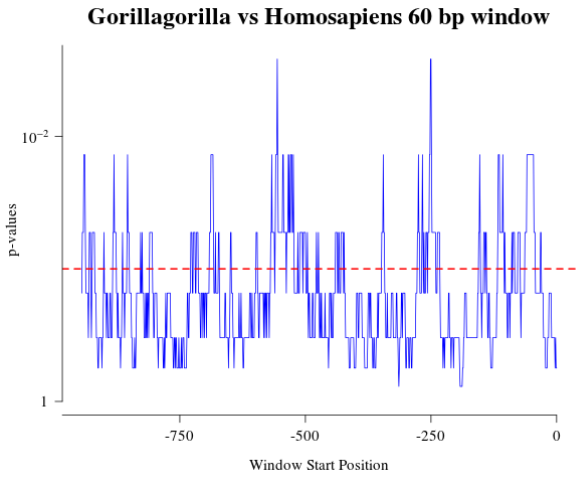
(c)



(d)



(e)



(f)

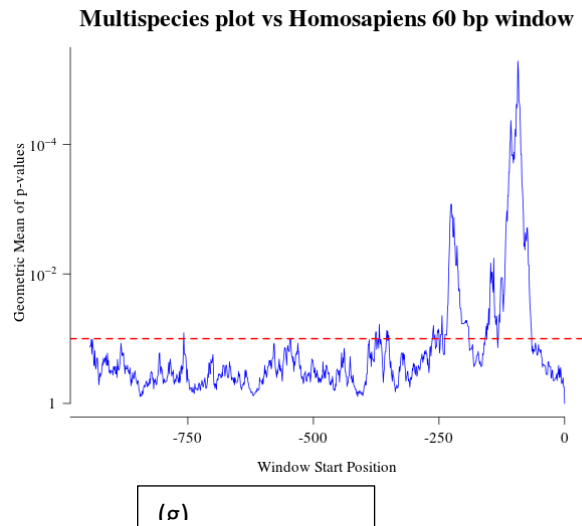


Figure no.: 3.8 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs



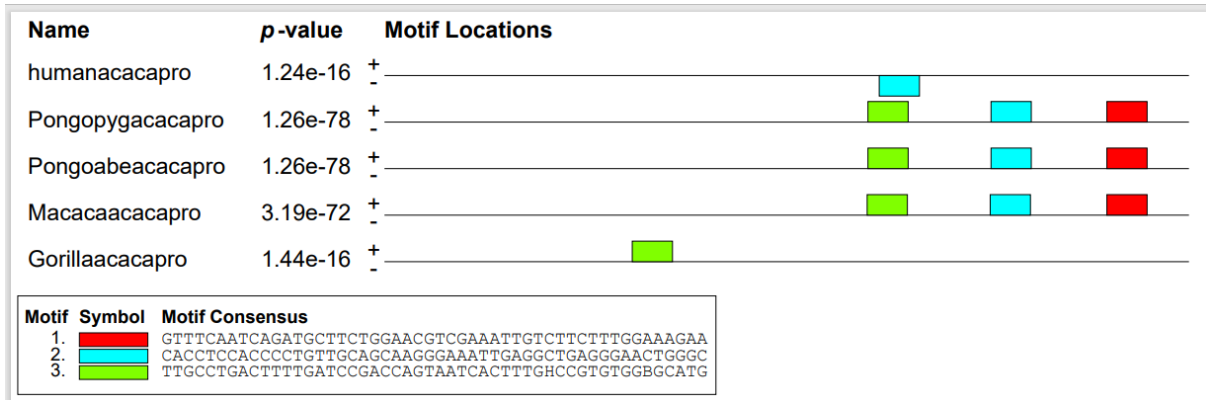
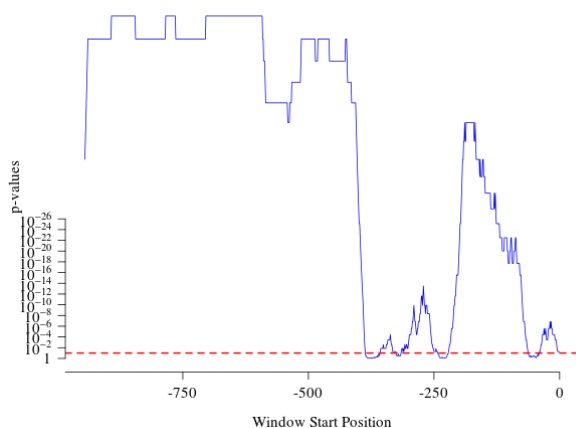


Figure 4.8 Motif identification ACACA

Pongopygmaeus, Pongo abelii and Macaca fascicularis show presence of three motifs. In Pongo pygmaeus, motif 1 starts at 899 and ends at 948. Motif 2 starts at 755 and ends at 804. Motif 3 starts at 602 and ends at 641. In Pongo abelii motif 1 starts at 899 and ends at 948. Motif 2 starts at 755 and ends at 804. Motif 3 starts at 602 and ends at 651. In Macaca fascicularis motif 1 starts at 899 and ends at 948. Motif 2 starts at 755 and ends at 804. Motif 3 starts at and ends at 602 and ends at 641. Pongo pygmaeus, Pongo abelii and macaca fascicularis share presence of common three motifs on forward strand

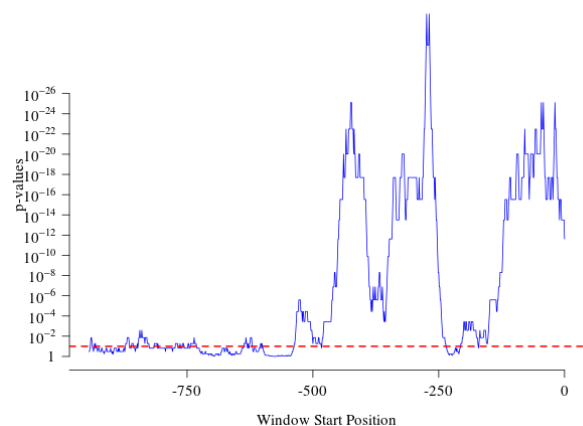
CHREBP

Panpaniscus vs Homosapiens 60 bp window



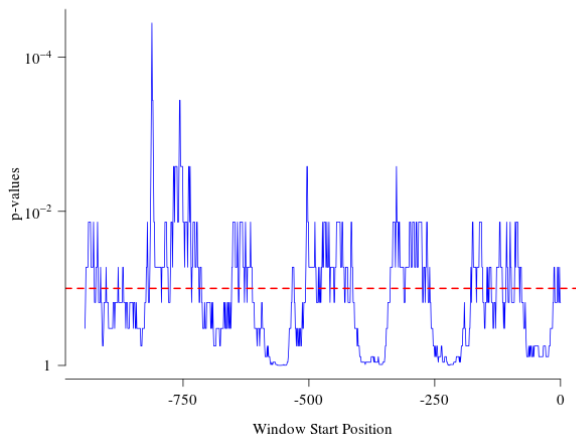
(a)

Pantroglodytes vs Homosapiens 60 bp window



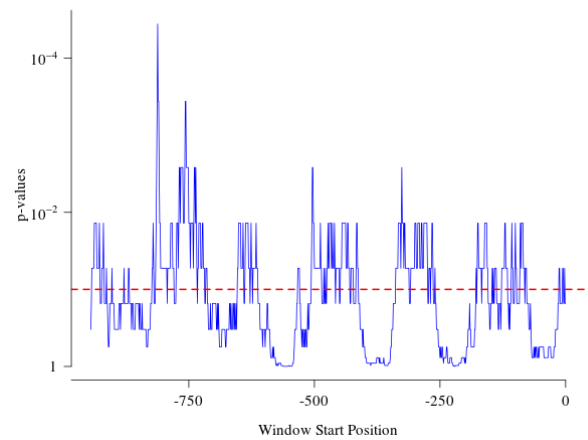
(b)

Pongopygmaeus vs Homosapiens 60 bp window



(c)

Pongoabelii vs Homosapiens 60 bp window



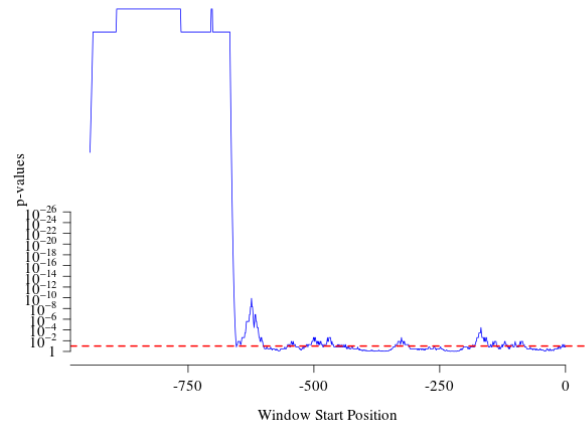
()

Macacafascicularis vs Homosapiens 60 bp window



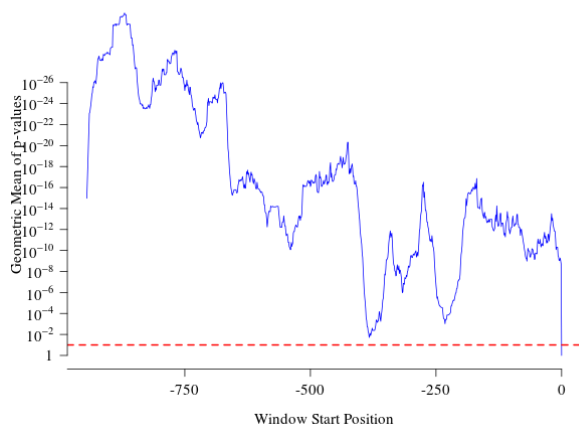
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies prior vs Homosapiens 60 bp window



(g)

Figure no.: 3.9 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo

pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis
(f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

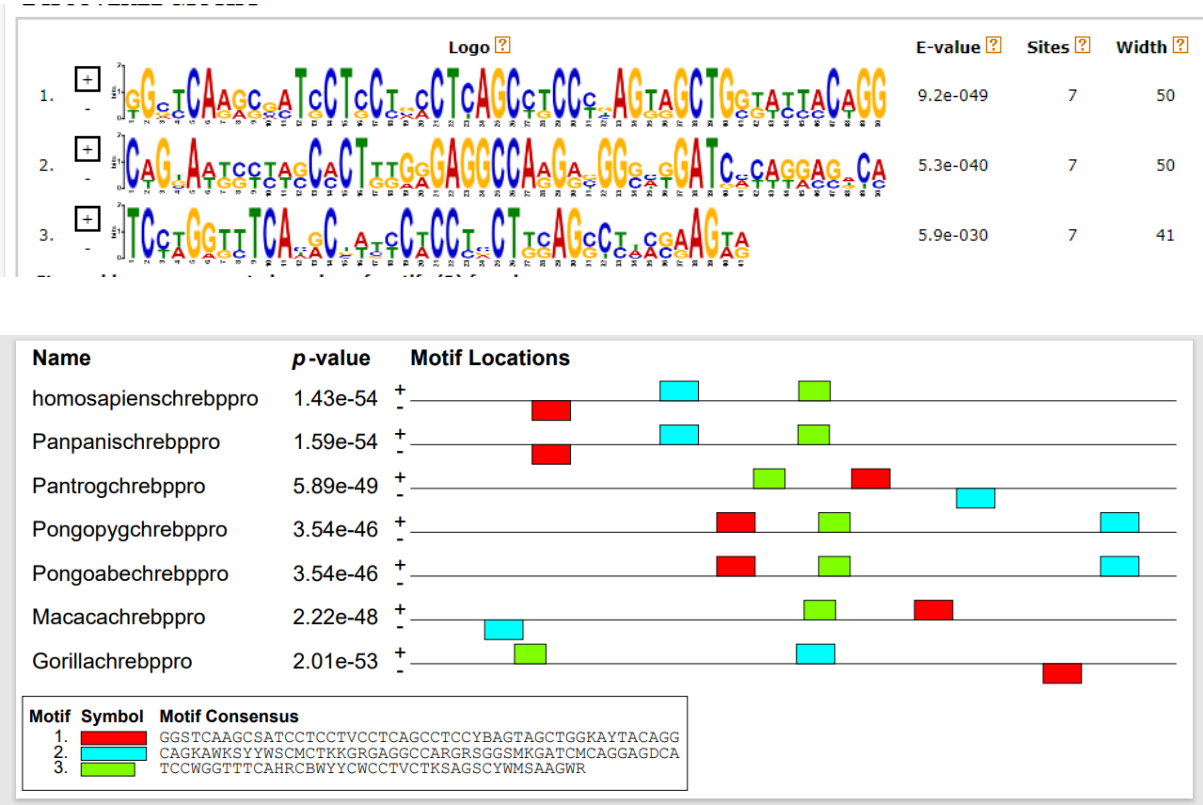


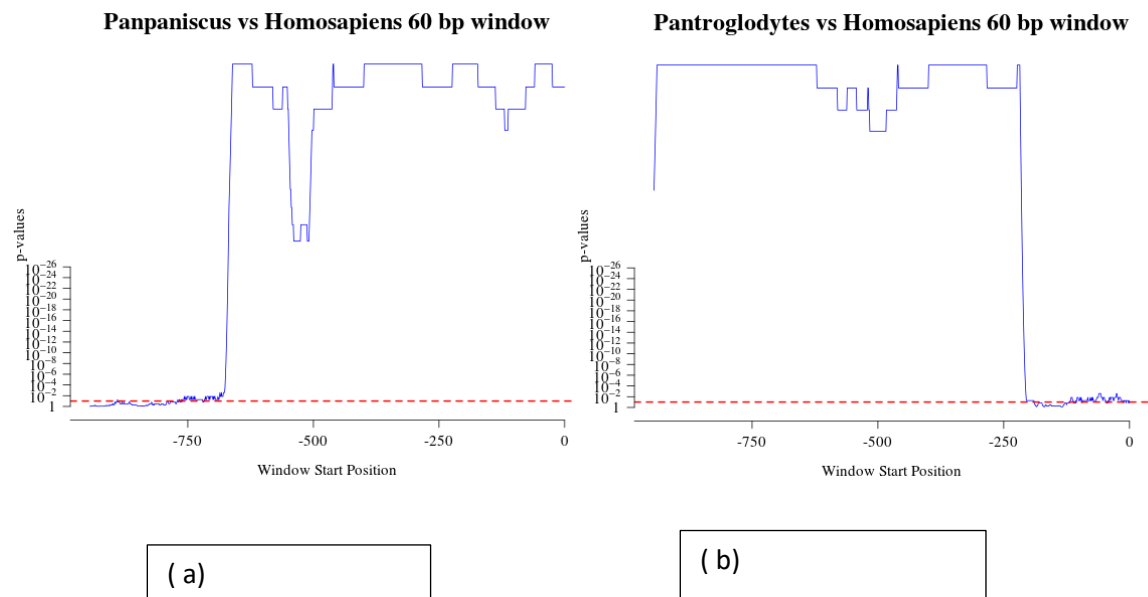
Figure 4.9 Motif identification CHREBP

In homosapiens motif 1 starts at -160 and -209. Motif 2 starts at 327 and ends at 376. Motif 3 starts at 508 and ends at 548. In Pan paniscus motif 1 starts at -160 and ends at -209 . Motif 2 starts at 327 and ends at 376. Motif 3 starts at 507 and ends at 547. In Pan troglodytes motif 1 starts at 577 and ends at 626. Motif 2 starts at -714 and ends at -763. Motif 3 starts at 449 and ends at 489, In Pan troglodytes motif 1 starts at 577 and ends at 626. Motif 2 starts at 902 and ends at 951 . Motif 3 starts at 449 and ends at 489 . In Pongo pygmaeus motif 1 starts at 401 and ends at 450 . Motif 2 starts at 902 and ends at 951 . Motif 3 starts at 534 and ends at 574 . In Pongo abelii motif 1 starts at 401 and

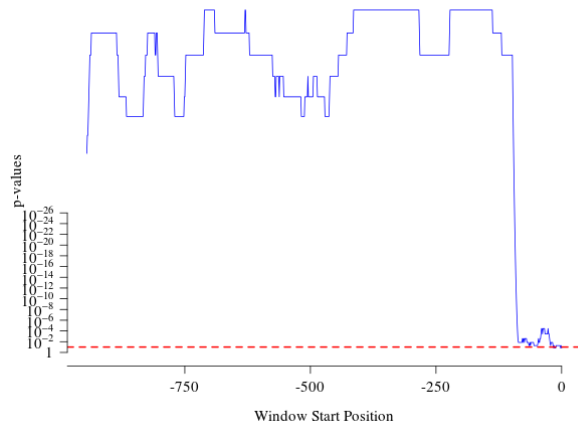
ends at 450 .Motif 2 starts at 902 and ends at 951.Motif 3 start at 534 and ends at 574 .In *Macaca fascicularis* motif 1 starts at 659 and ends at 708. Motif 2 starts at 98 and ends at 147. Motif 3 starts at 515 and ends at 555 .In *Gorilla gorilla* motif 1 starts at 827 and ends at 876 . Motif 2 starts at 504 and ends at 554 .Motif 3 starts at 137 and ends at 177 .

Homo sapiens and *pan paniscus* motif 1 and 2 same position.*Pongo pygmaeus* and *Pongo abelii* motif 1 and motif 2 start at same position.

SCD

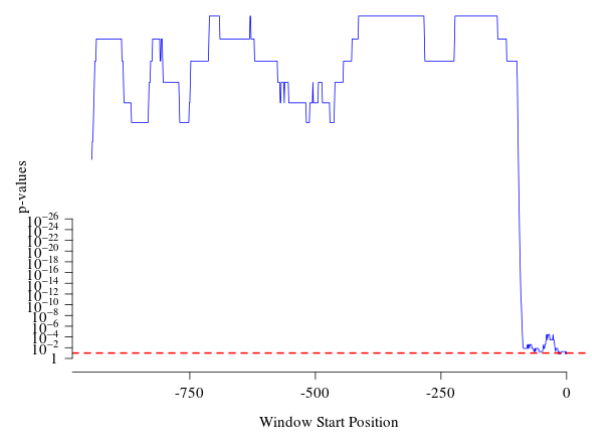


Pongopygmaeus vs Homosapiens 60 bp window



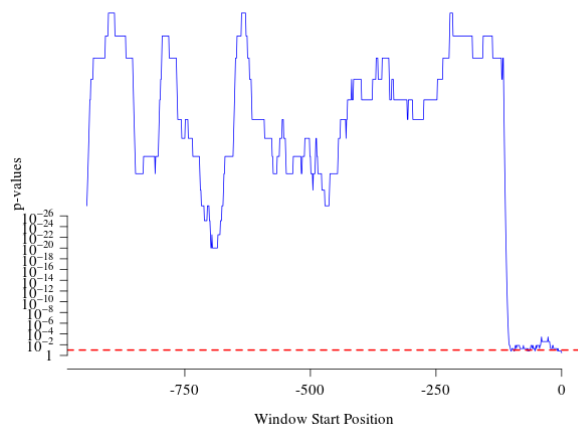
(c)

Pongoabelii vs Homosapiens 60 bp window



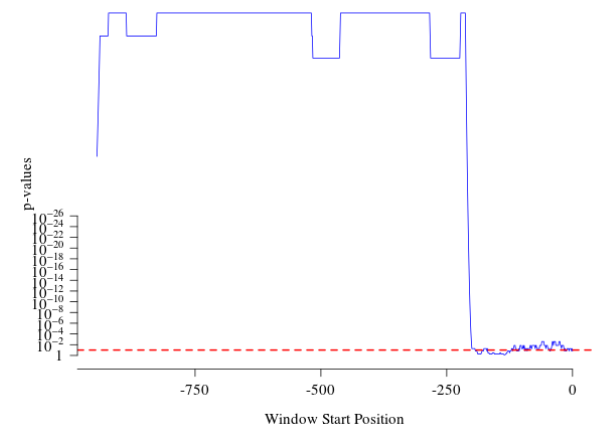
(d)

Macacafascicularis vs Homosapiens 60 bp window



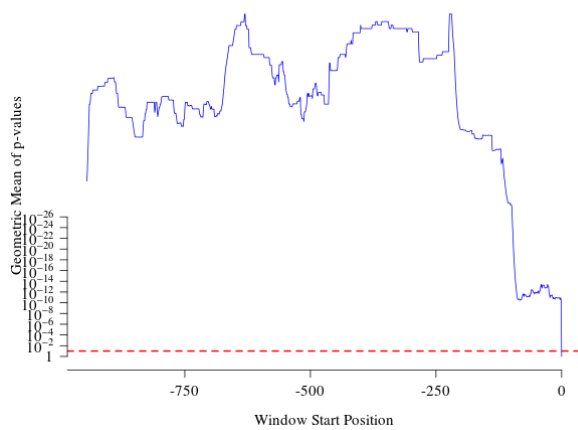
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure no.: 3.10 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

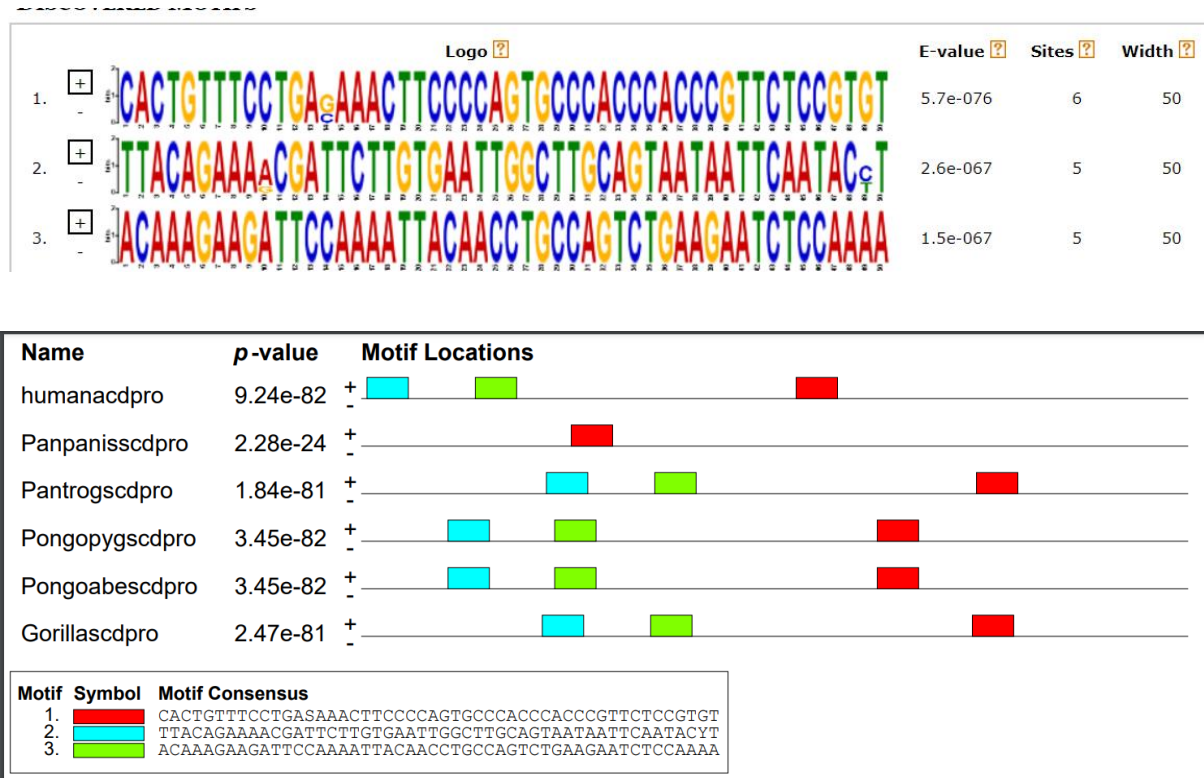


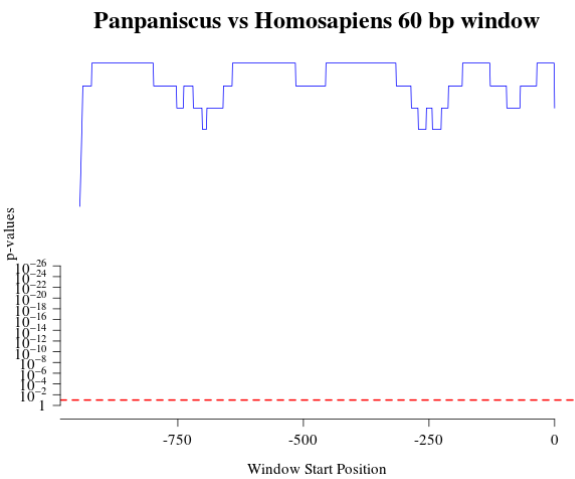
Figure 4.10 Motif identification SCD

In homosapiens motif 1 starts at 527 and ends at 576. Motif 2 starts at 8 and ends at 57. Motif 3 starts at 139 and ends at 188. In Pan troglodytes motif 1 starts at 745 and ends at 794.. Motif 2 starts at 225 and ends at 274. Motif 3 starts at 356 and ends at 405. In Pan troglodytes motif 1 starts atand ends at. Motif 2 starts at and ends at .Motif 3 starts at and ends at. In Pongo pygmaeus motif 1 starts at 625 and ends at 674. Motif 2 starts at 106 and ends at 155. Motif 3 starts at 235 and ends at 284. In Pongo abelii motif 1 starts at 625 and ends

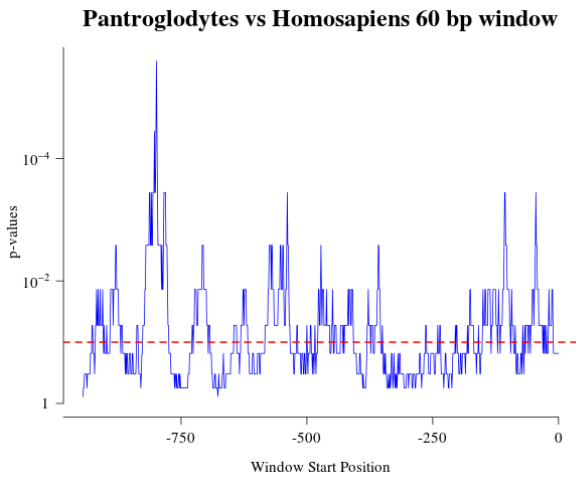
at 674.Motif 2 starts at 106 and ends at 155.Motif 3 start at 235 and ends at 284.In Gorilla gorilla motif 1 starts at 740 and ends at 789. Motif 2 starts at 220 and ends at 269 .Motif 3 starts at 351 and ends at 400

Motif 3 starts at 351 and ends at 400 .Except pan panisucus all animals share presence of common three motifs

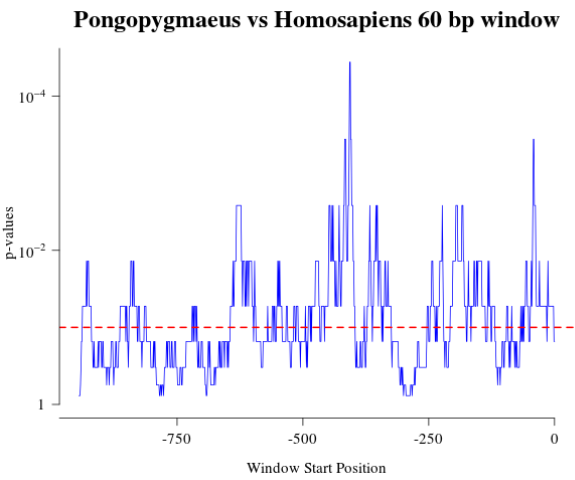
GHSR



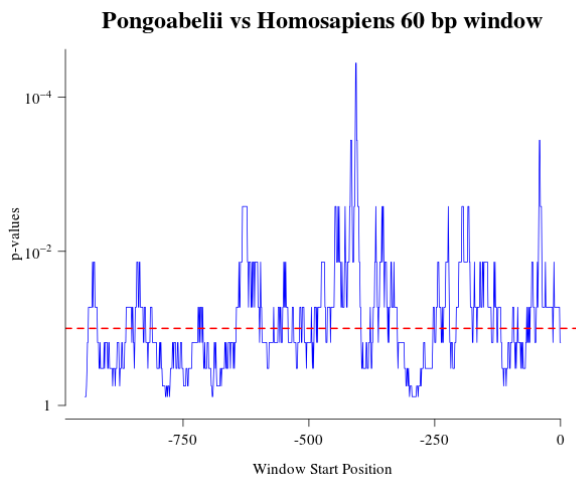
(a)



(b)



(c)



(d)

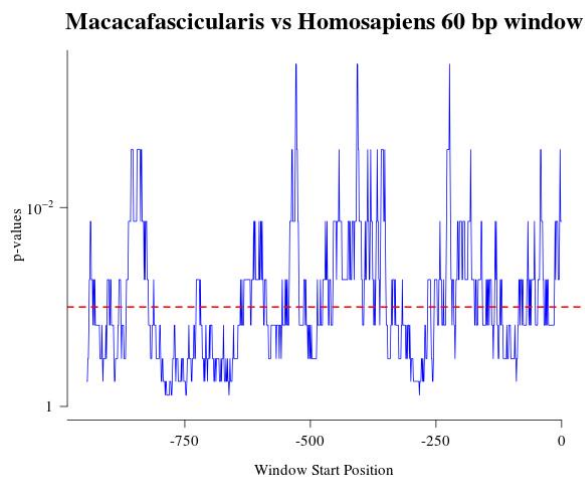
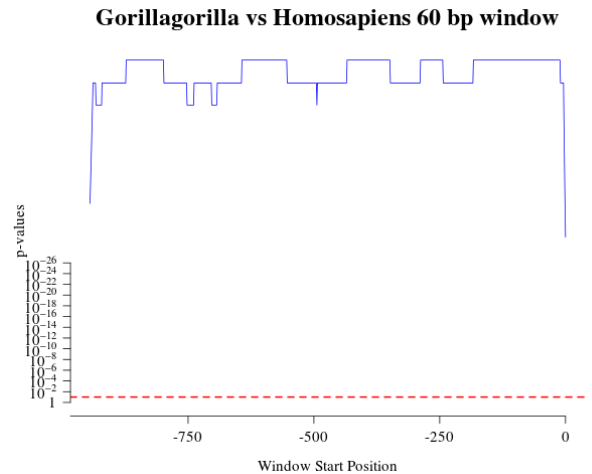
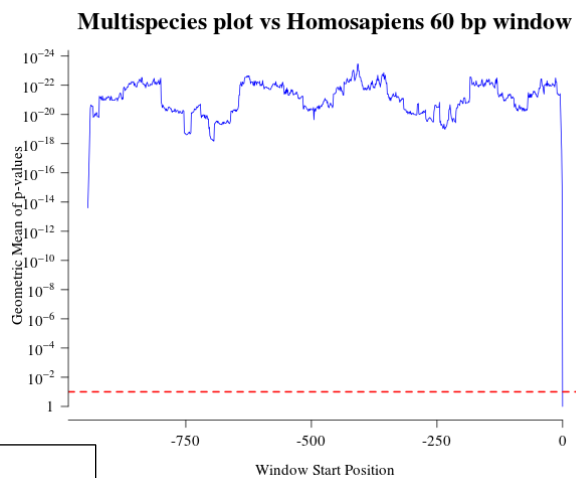


Figure no.: 3.11 Evolutionary analysis of regulatory sequences plot 1(a)

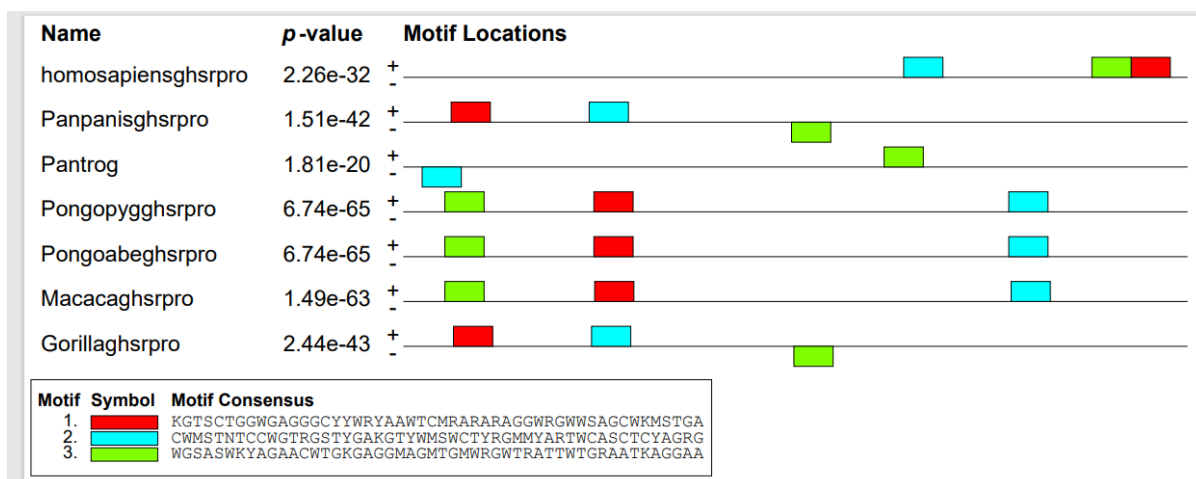


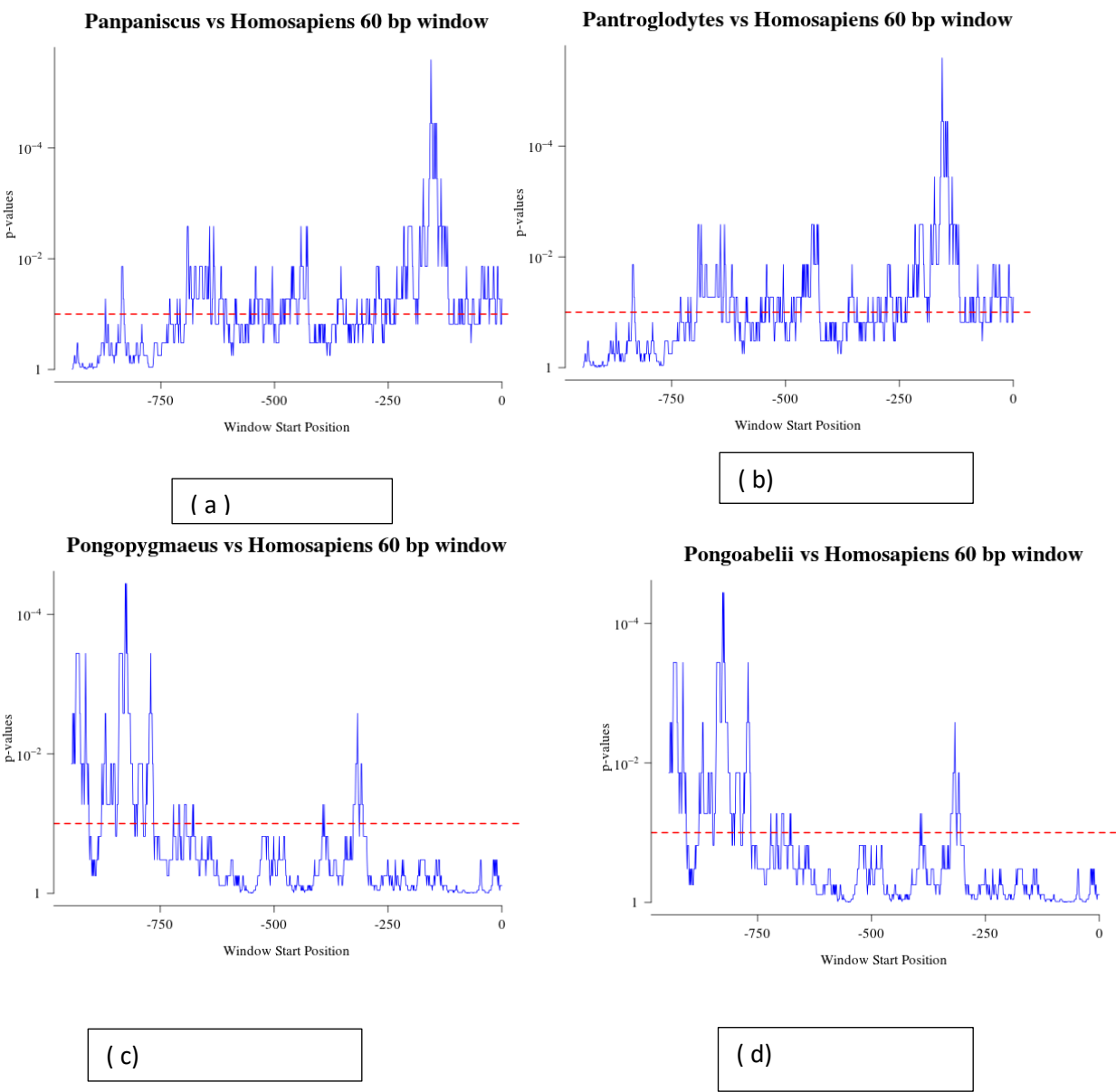
Figure identification 4.11 Motif identification GHSR

In homosapiens motif 1 starts at 929 and ends at 978. Motif 2 starts at 639 and ends at 688. Motif 3 starts at 879 and ends at 928. In Pan paniscus motif 1 starts at 62 and ends at 111. Motif 2 starts at 238 and ends at 287. Motif 3 starts at 496 and ends at 545 In Pongo pygmaeus motif 1 starts at 244 and ends at 293. Motif 2 starts at 773 and ends at 822. Motif 3 starts at 54 and ends at 103. In Pongo abelii motif 1 starts at 244 and ends at 293. Motif 2 starts at 773 and ends at 822. Motif 3 start at 244 and ends at 293. In Macaca fascicularis motif 1 starts at 245 and ends at 294. Motif 2 starts at 776 and ends at 825. Motif 3 starts at 54 and ends at 103. In Gorilla gorilla motif 1 starts at 65 and ends at 114. Motif 2 starts at 241 and ends at 290. Motif 3 starts at 499 and ends at 548.

This conservation can be visualized using sequence logos, where the height of each letter represents its frequency at a given position in the motif. The start and end sites of a motif may be defined by the alignment coordinates, with the start site corresponding to the beginning of the aligned region and the end site corresponding to the end of the aligned region. The start and end sites of motifs may be biologically significant,

particularly if they correspond to binding sites for transcription factors, regulatory elements, or protein domains. In such cases, the specific sequence and positioning of motifs may be critical for their function and regulation. *Pongo pygmaeus*, *Pongo abelii* and *Macaca fascicularis* share presence of three motifs

NPY



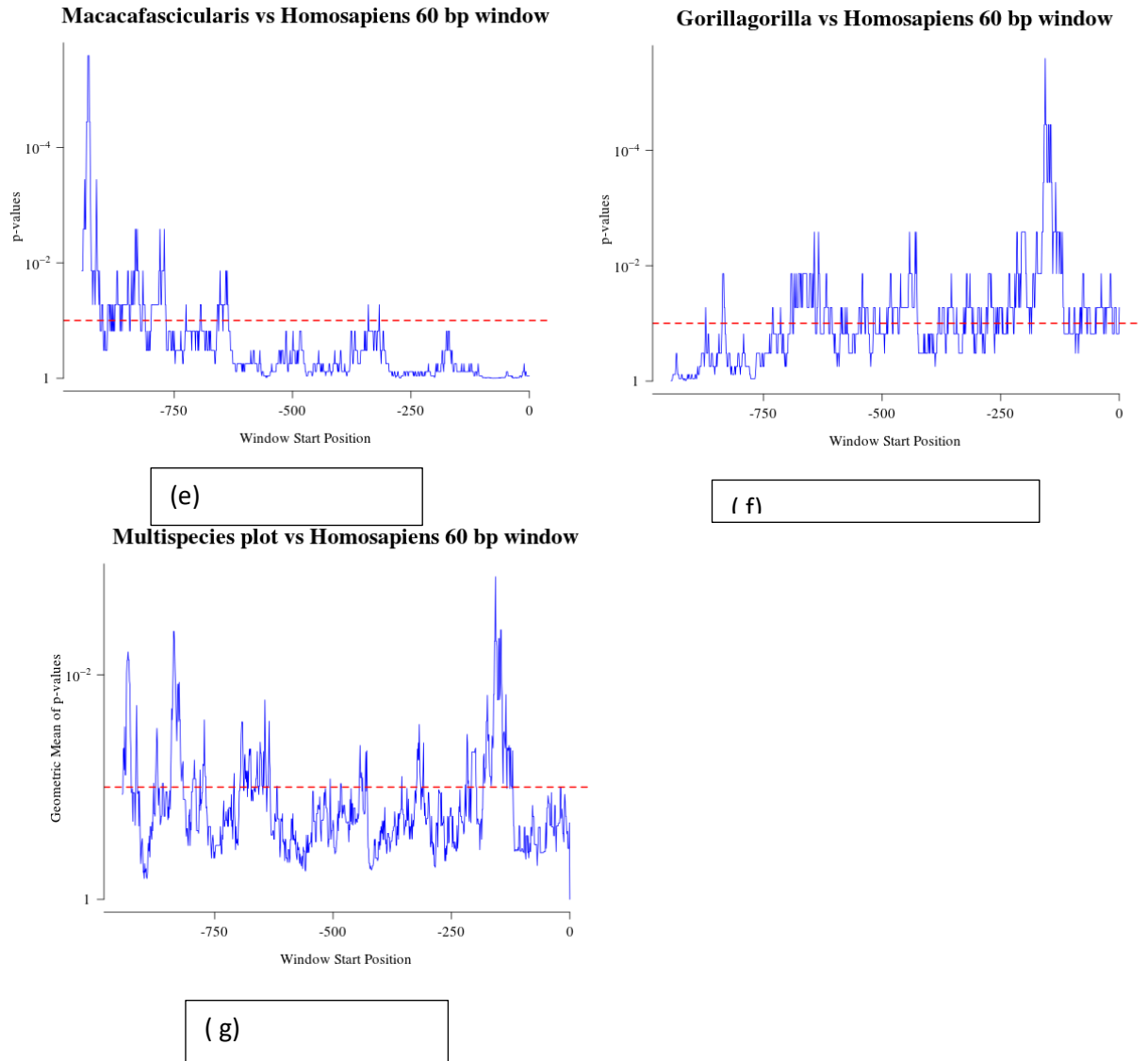


Figure no.:3.12 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

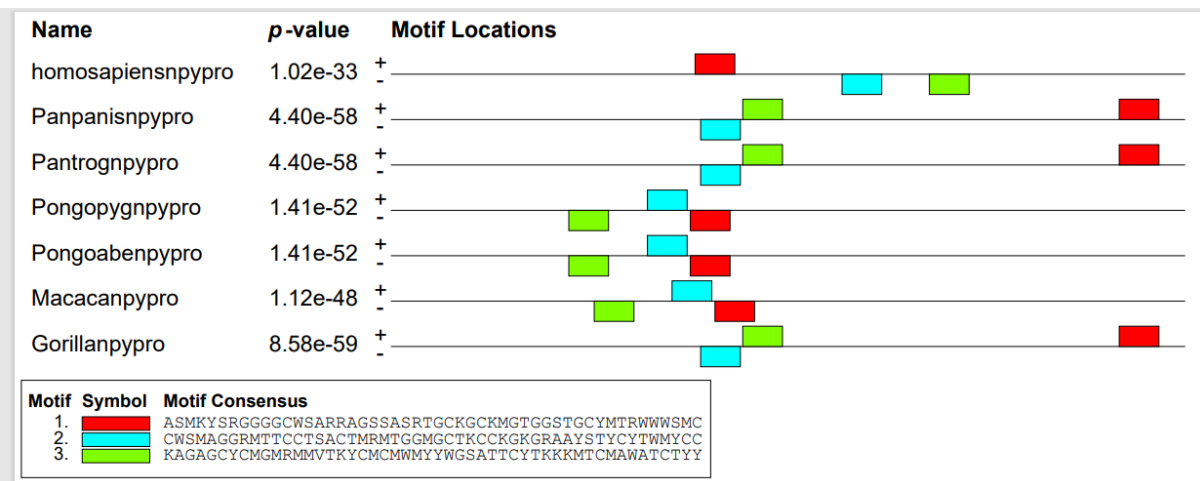
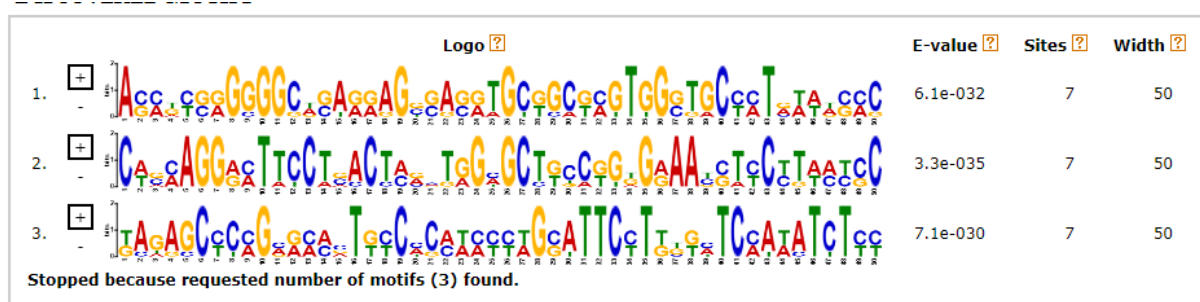


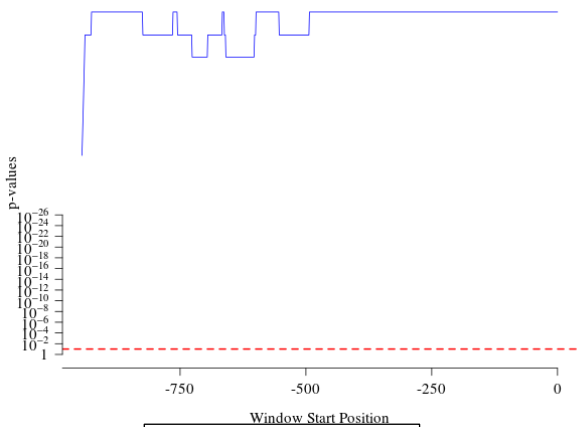
Figure 4.12 Motif identification NPY

In homosapiens motif 1 starts at 384 and ends at 433. Motif 2 starts at 569 and ends at 618. Motif 3 starts at 679 and ends at 728. In Pan paniscus motif 1 starts at 918 and ends at 967. Motif 2 starts at 391 and ends at 440. Motif 3 starts at 444 and ends at 493 In Pan troglodytes motif 1 starts at 918 and ends at 967. Motif 2 starts at 390 and ends at 440. Motif 3 starts at 444 and ends at 493. In Pongo pygmaeus motif 1 starts at 378 and ends at 427. Motif 2 starts at 324 and ends at 373. Motif 3 starts at 225 and ends at 274. In Pongo abelii motif 1 starts at 378 and ends at 427. Motif 2 starts at 324 and ends at 373. Motif 3 start at 225 and ends at 274. In Macaca fascicularis motif 1 starts at 409 and ends at 458. Motif 2 starts at 355 and ends at 404. Motif 3 starts at 355 and ends at 404. In Gorilla gorilla motif 1 starts at 918 and ends at 967. Motif 2 starts at 391 and ends at 440. Motif 3 starts at 444 and ends at

493. Pan troglodytes and gorilla have starting and ending positions of motifs as same. Pongo pygmaeus and Pongo abelii have starting and ending positions of motifs as same

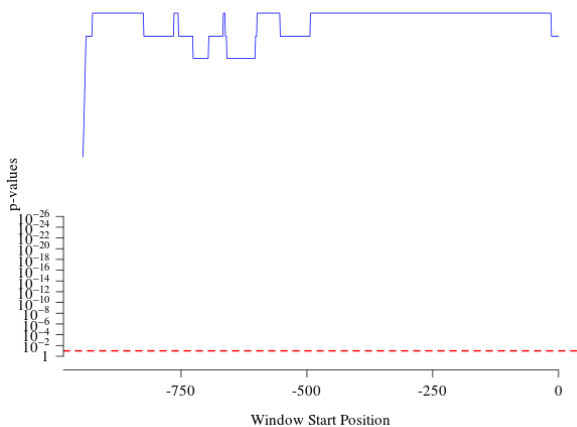
AGRP

Panpaniscus vs Homosapiens 60 bp window



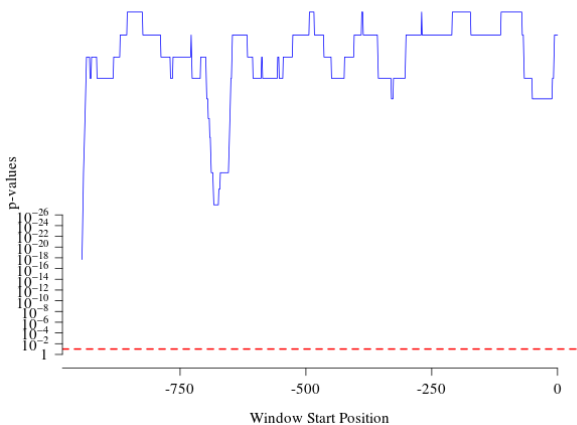
(a)

Pantroglodytes vs Homosapiens 60 bp window



(b)

Pongopygmaeus vs Homosapiens 60 bp window



(c)

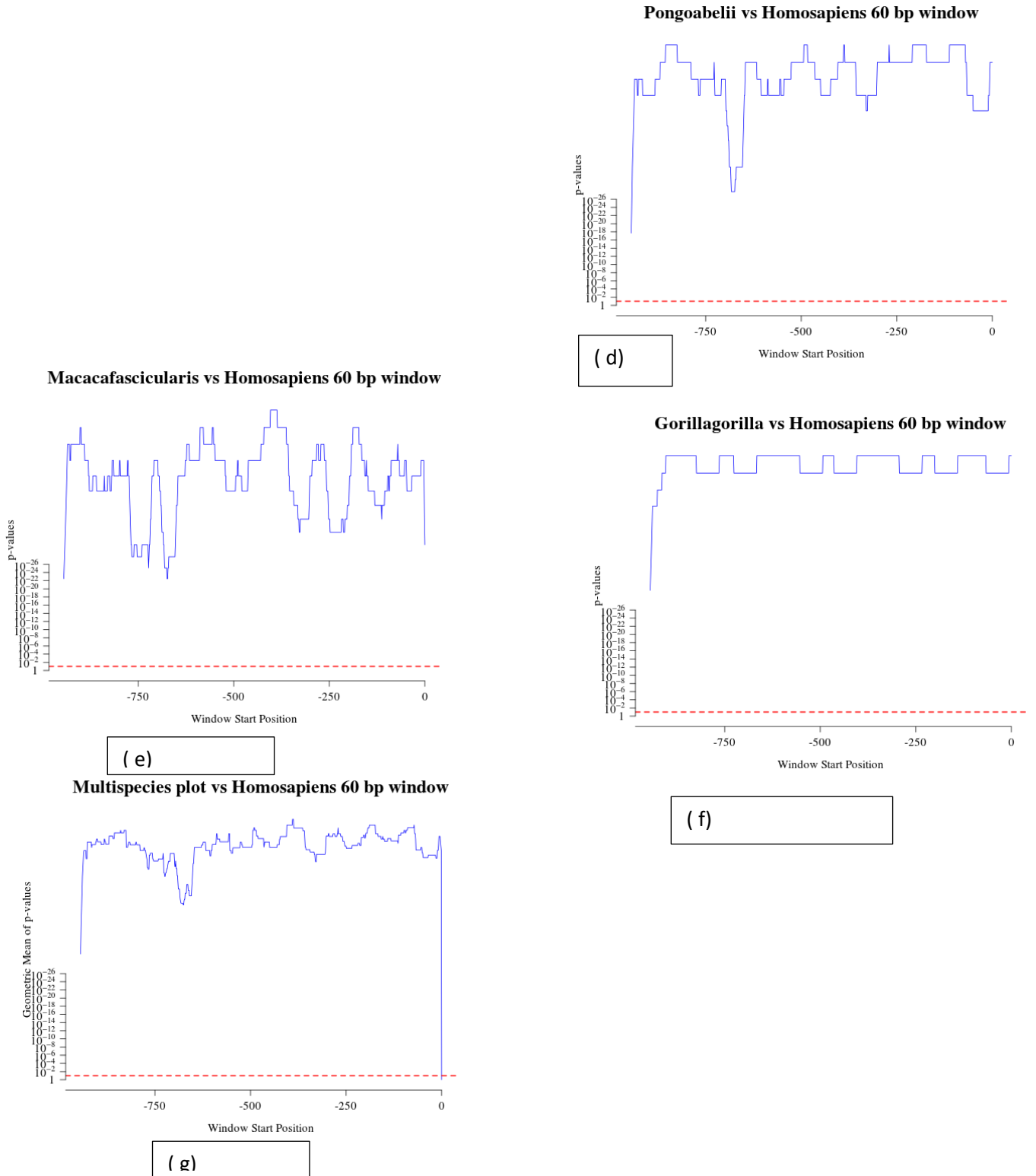


Figure no.:3.13 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

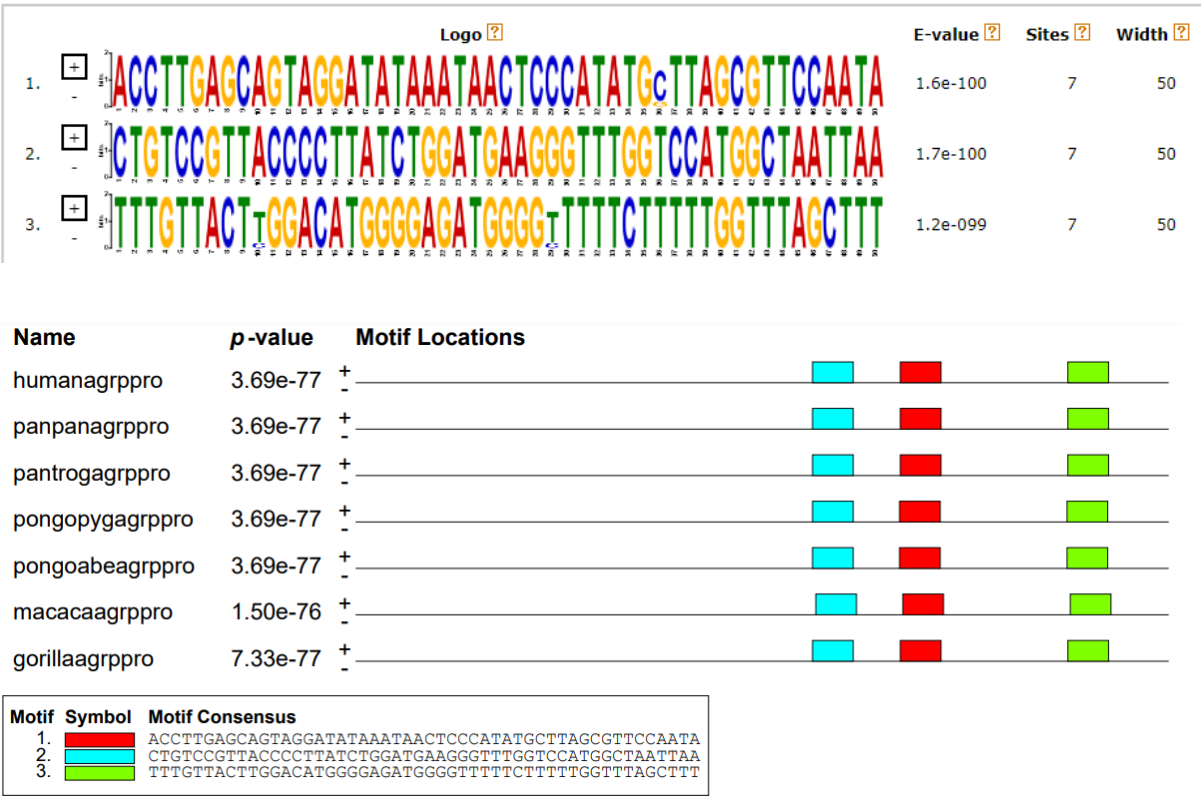
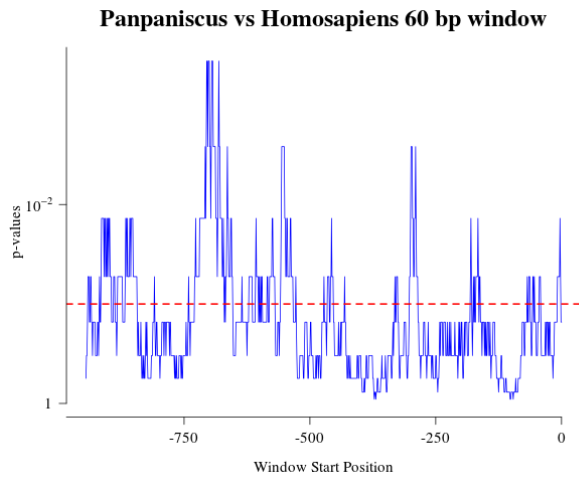


Figure 4.13 Motif identification AGRP

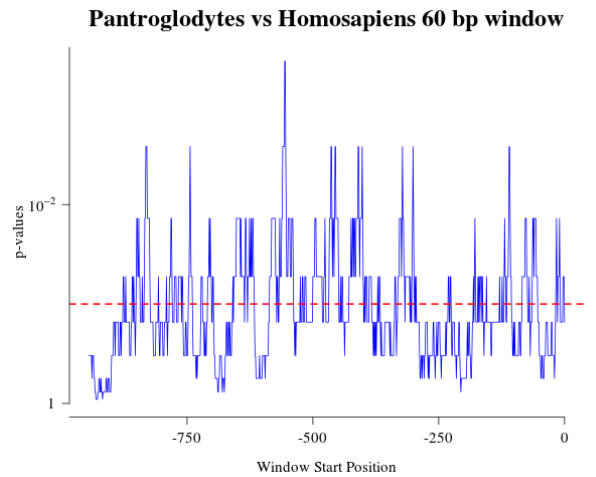
In homosapiens motif 1 starts at 671 and ends at 720. Motif 2 starts at 563 and ends at 612. Motif 3 starts at 877 and ends at 926. In Pan paniscus motif 1 starts at 671 and ends at 720. Motif 2 starts at 563 and ends at 612. Motif 3 starts at 877 and ends at 925. In Pan troglodytes motif 1 starts at 671 and ends at 720.. Motif 2 starts at 563 and ends at 612. Motif 3 starts at 877 and ends at 925. In Pongo pygmaeus motif 1 starts at 617 and ends at 720. Motif 2 starts at 563 and ends at 612.. Motif 3 starts at 670 and ends at 719. In Pongo abelii motif 1 starts at 671 and ends at 720. Motif 2 starts at 563 and ends at 612. Motif 3 start at 877 and ends at 925. In Macaca fascicularis motif 1 starts at 674 and ends at 723. Motif 2 starts at 567 and ends at 616. Motif 3 starts at 880 and ends at 929. In Gorilla gorilla motif 1 starts at 671 and ends at 720.

Motif 2 starts at 563 and ends at 612. Motif 3 starts at 877 and ends at 926.

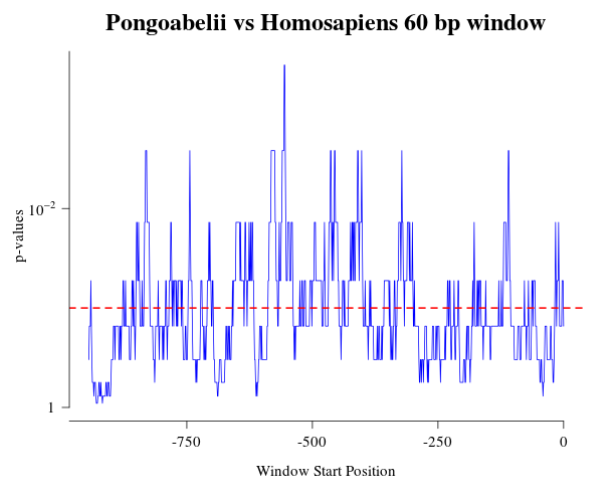
FOX01



(a)

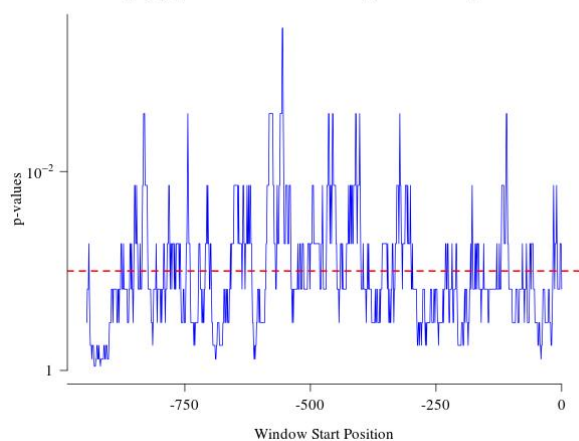


(b)



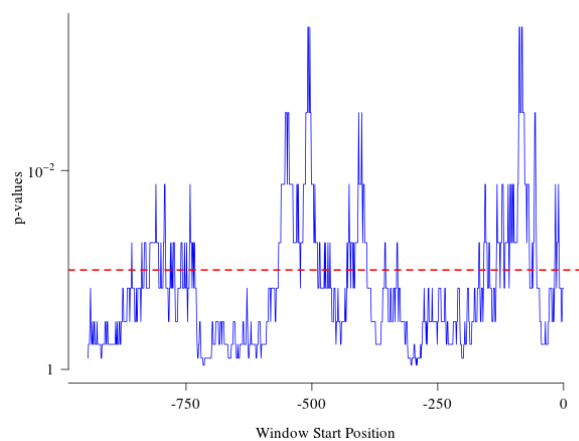
(d)

Pongopygmaeus vs Homosapiens 60 bp window



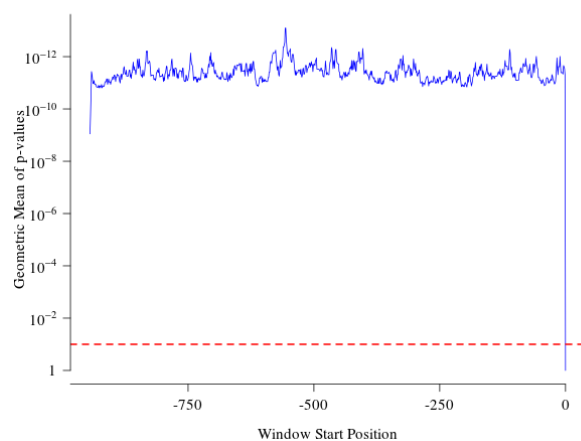
(c)

Macacafascicularis vs Homosapiens 60 bp window



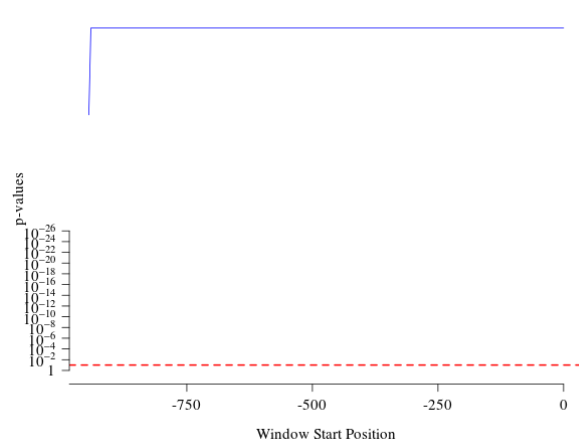
(f)

Multispecies plot vs Homosapiens 60 bp window



(e)

Gorillagorilla vs Homosapiens 60 bp window



(g)

Figure no.:3.14 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes1(c)homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

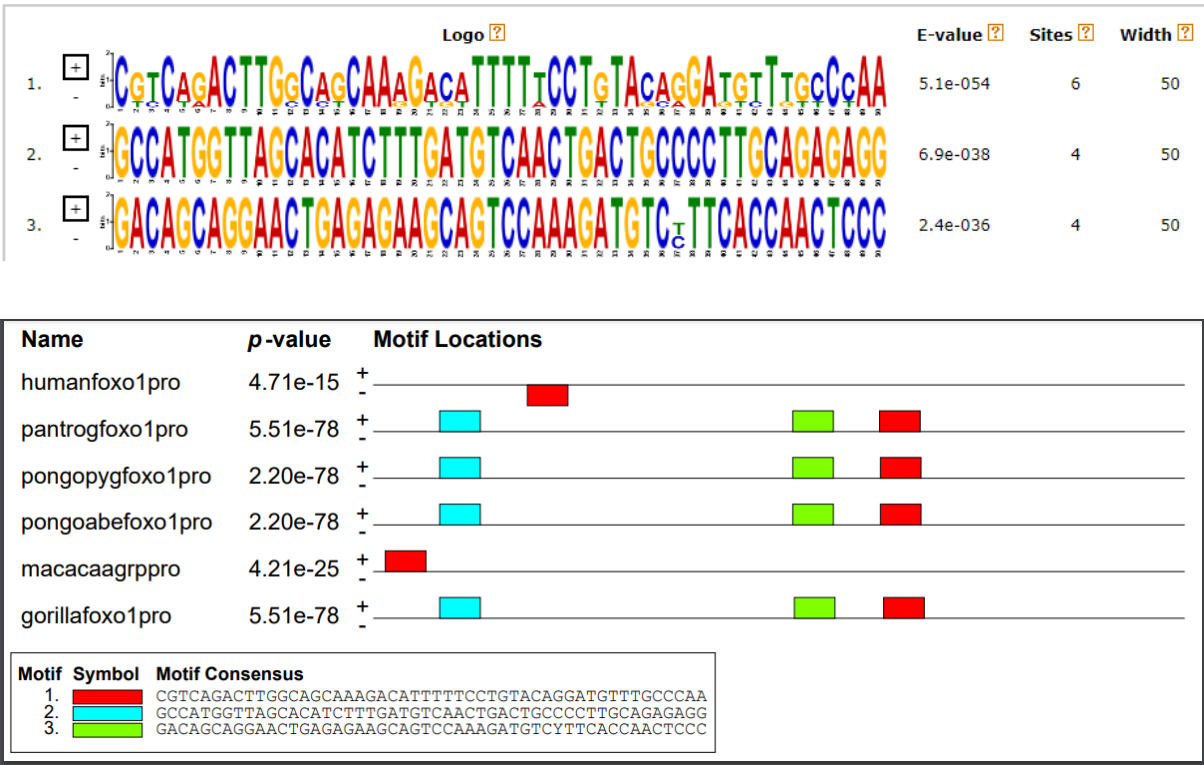
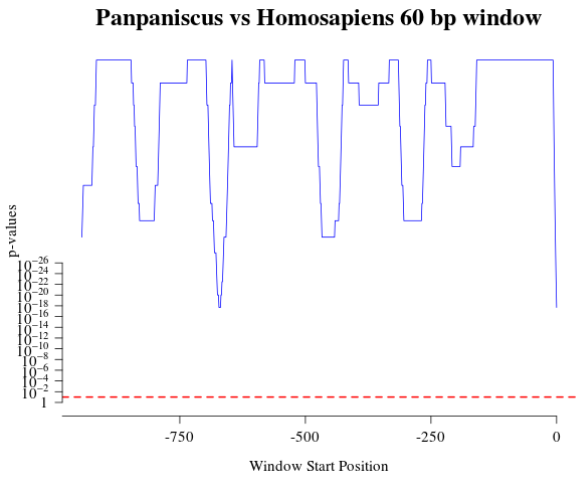


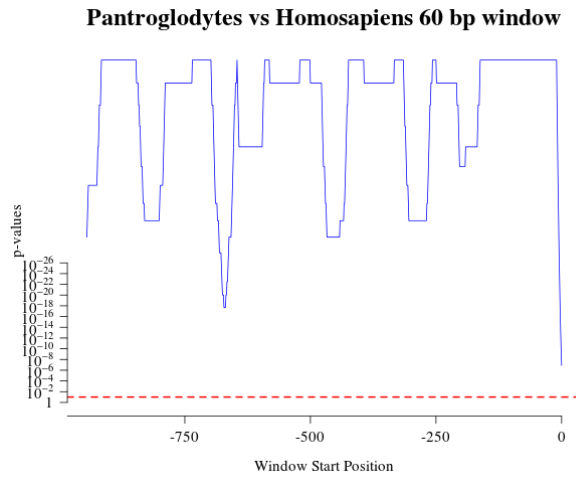
Figure 4.14 Motif identification FOXO1

In Pan troglodytes motif 1 starts at 625 and ends at 674. Motif 2 starts at 83 and ends at 132. Motif 3 starts at 518 and ends at 567. In Pongo pygmaeus motif 1 starts at 626 and ends at 675. Motif 2 starts at 83 and ends at 112. Motif 3 starts at 518 and ends at 567. In Pongo abelii motif 1 starts at 626 and ends at 675. Motif 2 starts at 83 and ends at 132. Motif 3 starts at 518 and ends at 567. In gorilla gorilla motif 1 starts at 630 and ends at 679. Motif 2 starts at 83 and ends at 132. Motif 3 starts at 520 and ends at 569. Except macaca and human all have common motifs on three strands

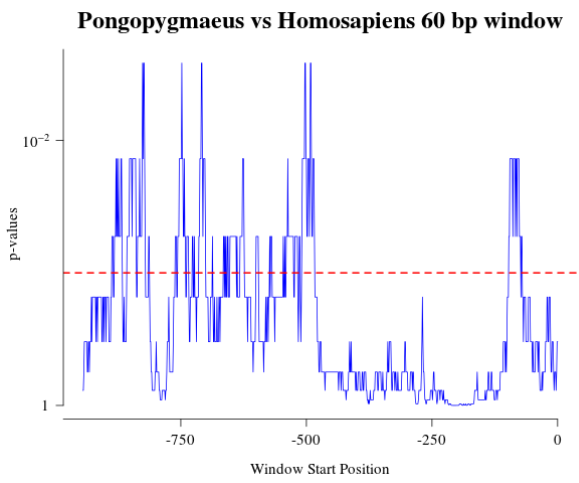
ADCY3



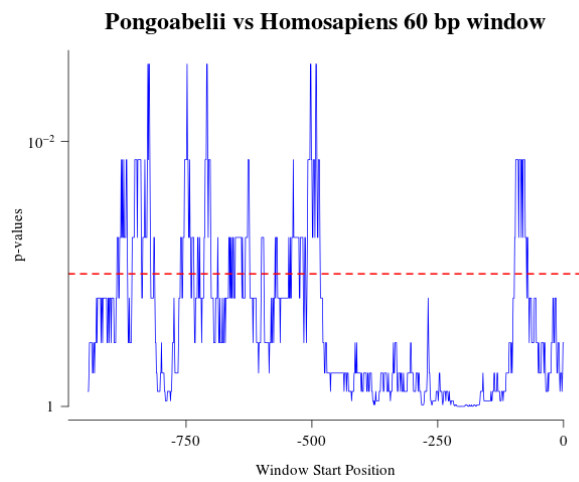
(a)



(b)



(c)



(d)

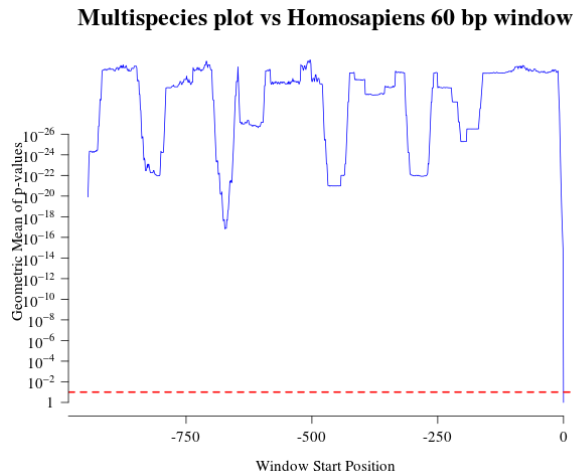
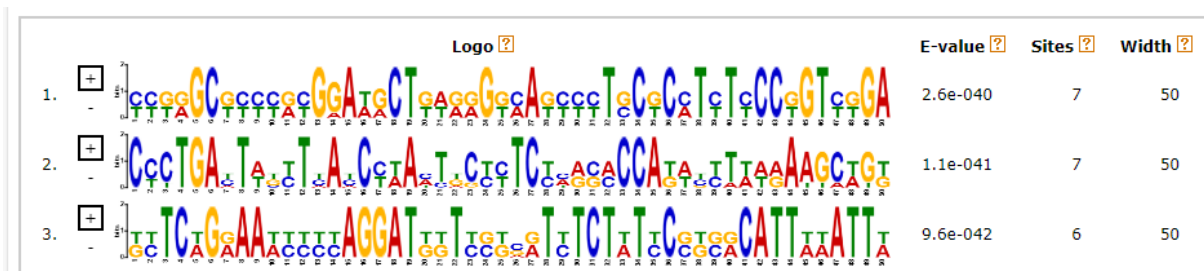


Figure no.: 3.15 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes 1(c) homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs



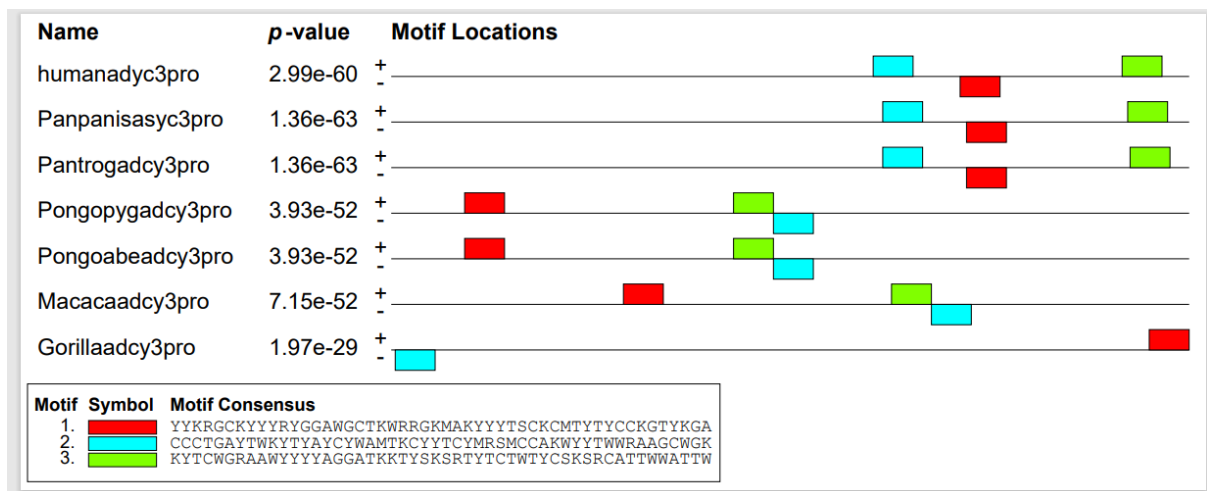
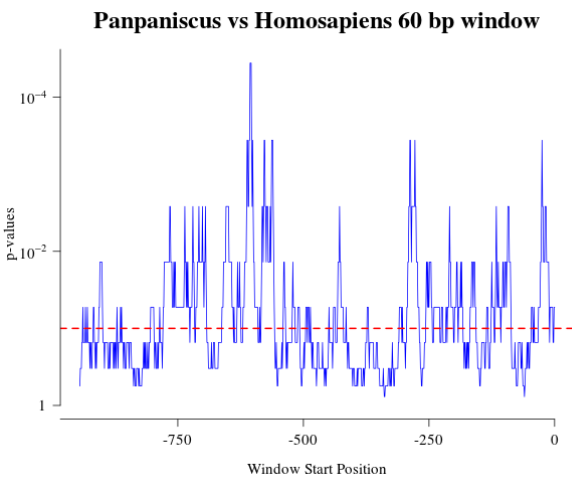


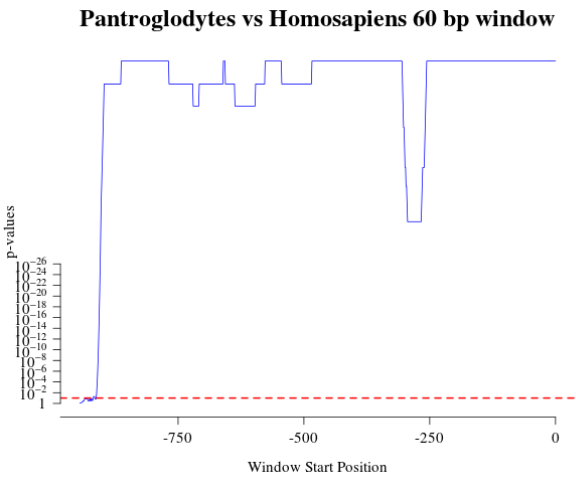
Figure 4.15 Motif identification 4.15

In Homosapiens motif 1 starts at 714 and ends at 763. Motif 2 starts at 605 and ends at 654. Motif 3 starts at 917 and ends at 966. In Pan paniscus motif 1 starts at 722 and ends at 771. Motif 2 starts at 617 and ends at 666. Motif 3 starts at 924 and ends at 973. In Pan troglodytes motif 1 starts at 722 and ends at 771. Motif 2 starts at 617 and ends at 666. Motif 3 starts at 927 and ends at 976. In Pongo pygmaeus motif 1 starts at 93 and ends at 142 motif 2 starts at 480 and ends at 529. Motif 3 starts at 430 and ends at 479. In Pongo abelii, motif 1 starts at 93 and ends at 142. Motif 2 starts at 480 and ends at 529. Motif 3 starts at 628 and ends at 677. In Macaca fascicularis motif 1 starts at 292 and ends at 341. Motif 2 starts at 628 and ends at 727. Motif 3 starts at 628 and ends at 677.

APLN

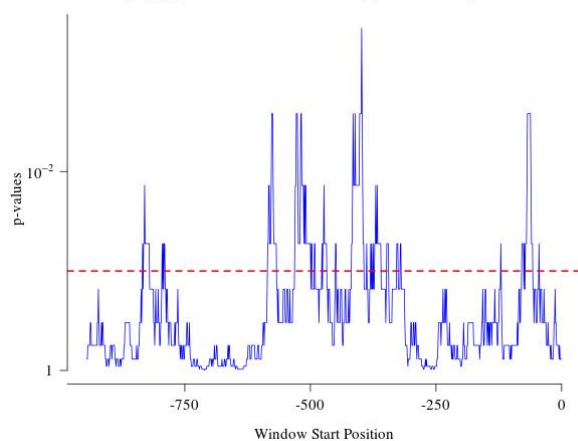


(a)



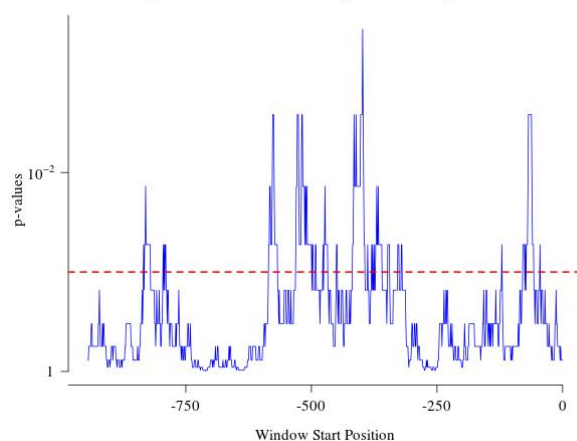
(b)

Pongopygmaeus vs Homosapiens 60 bp window



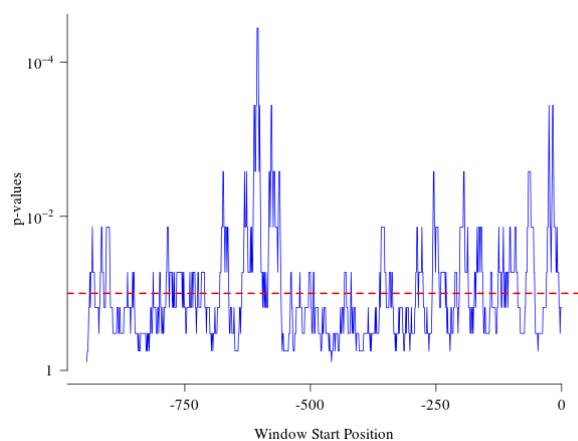
(c)

Pongoabelii vs Homosapiens 60 bp window



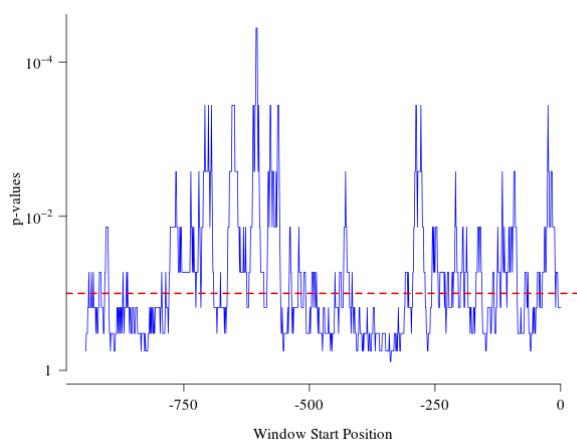
(d)

Macacafascicularis vs Homosapiens 60 bp window



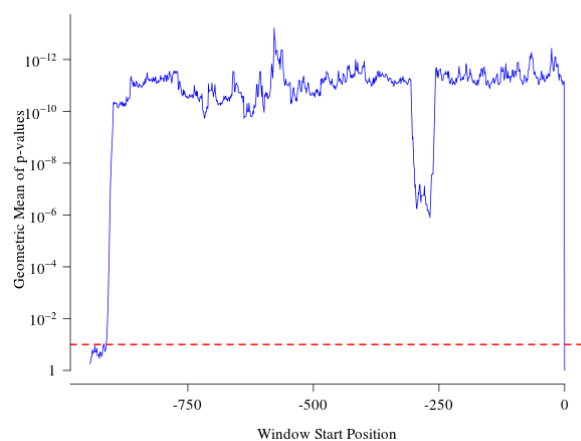
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

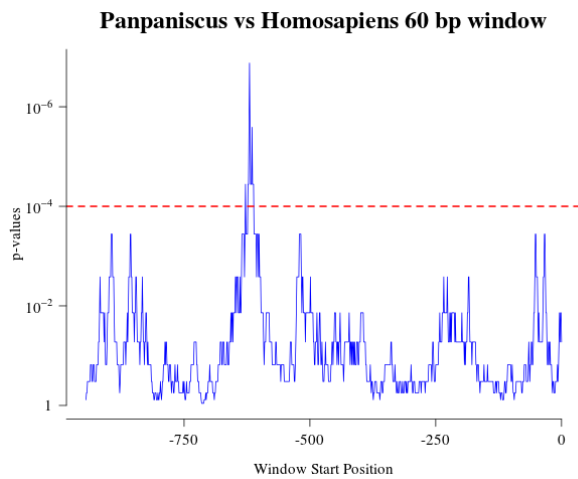
Multispecies plot vs Homosapiens 60 bp window



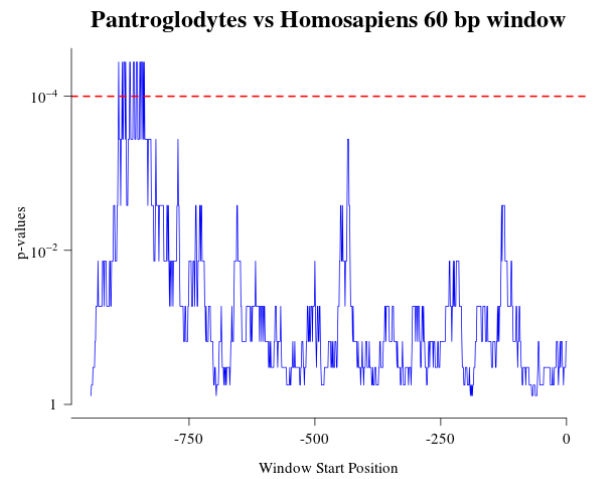
(g)

pygmaeus motif 1 starts at 908 and ends at 957. Motif 2 starts at 605 and ends at 654. Motif 3 starts at 677 and ends at 717. In *Pongo Abelii* motif 1 starts at 908 and ends at 957. Motif 2 starts at 605 and ends at 654. Motif 3 starts at 677 and ends at 717. In *Macaca fascicularis* motif 1 starts at 729 and ends at 778. Motif 2 starts at 950 and ends at 999. Motif 3 starts at 216 and ends at 256. In *Gorilla gorilla* motif 1 starts at 876 and ends at 925. Motif 2 starts at 398 and ends at 447. Motif 3 starts at 610 and ends at 650.

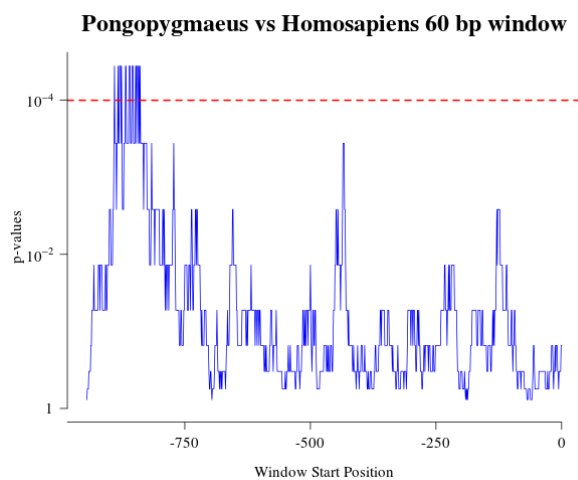
APLNR



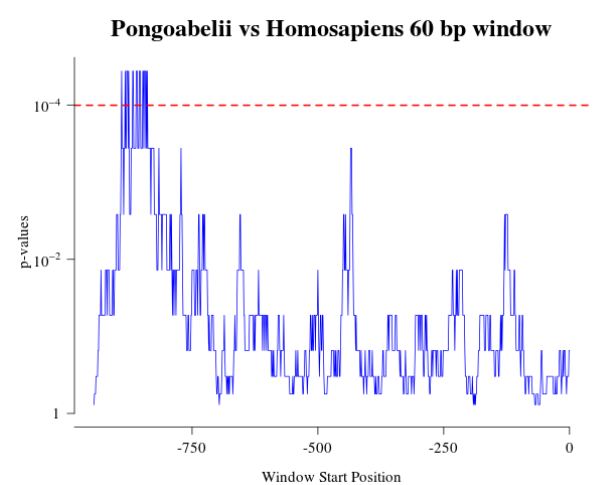
(a)



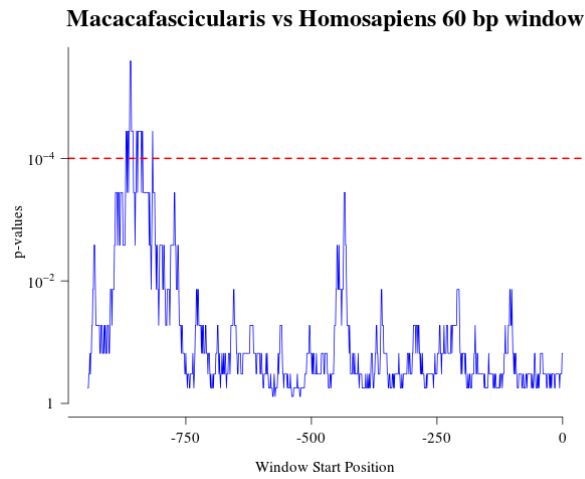
(b)



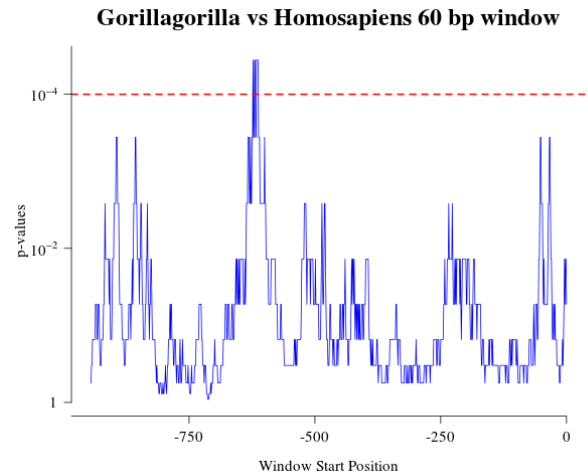
(c)



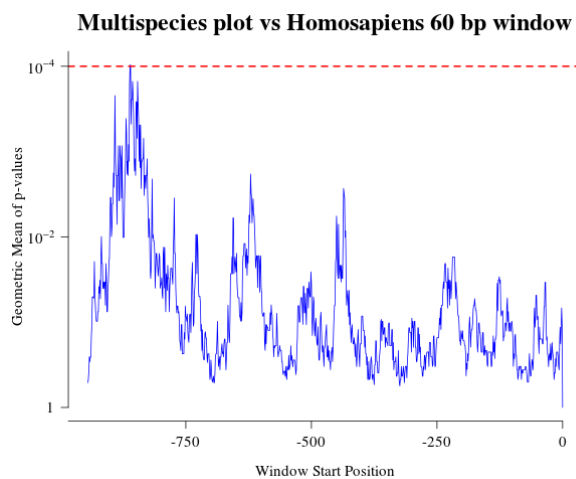
(d)



(e)



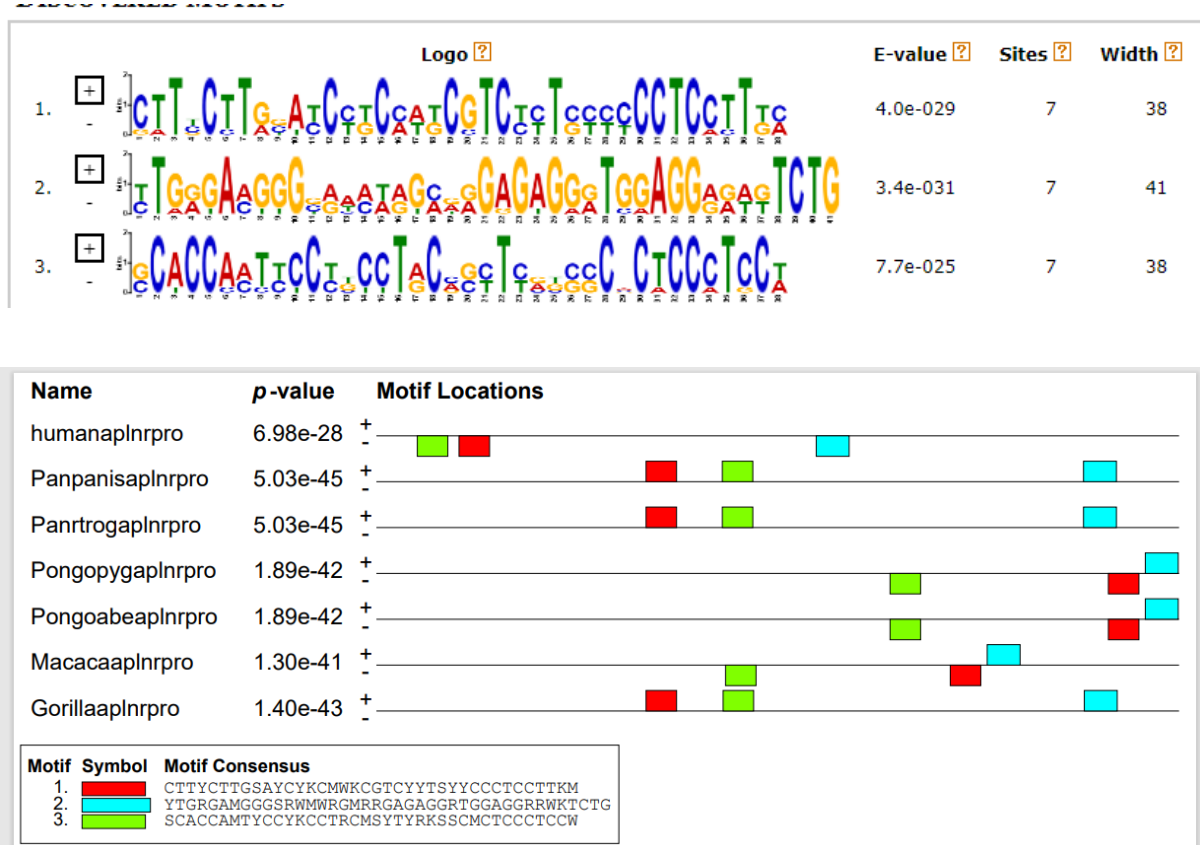
(f)



(g)

Figure no.: 3.17 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes 1(c) homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

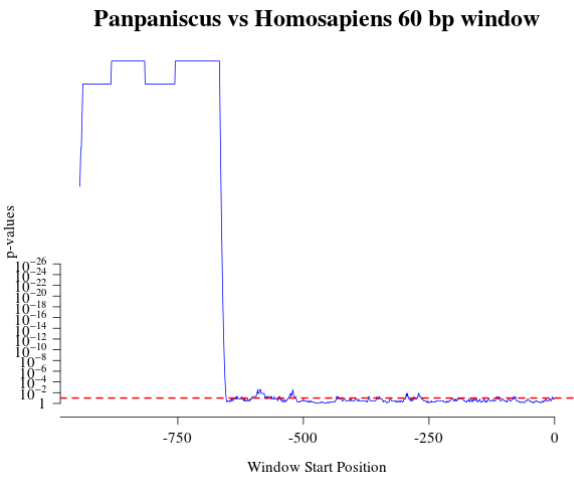
Identification of motifs



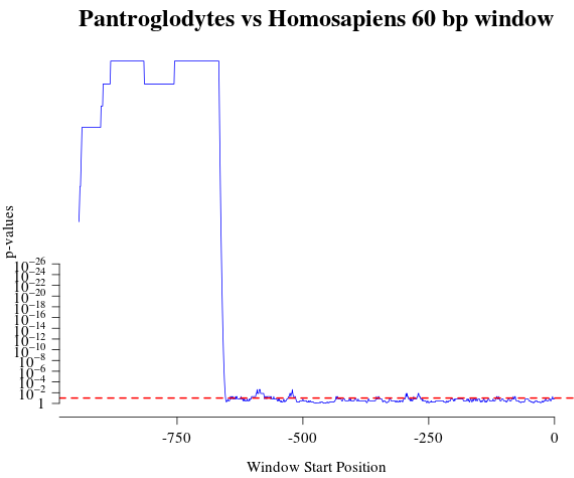
Motif identification 4.17 APLNR

In *Homo sapiens* motif 1 starts at 104 and ends at 141. Motif 2 starts at 549 and ends at 589. Motif 3 starts at 52 and ends at 89. In *Pan paniscus* motif 1 starts at 337 and ends at 374. motif 2 starts at 882 and ends at 922. motif 3 starts at 52 and ends at 89. In *Pan troglodytes* motif 1 starts at 337 and ends at 374. Motif 2 starts at 822 and ends at 922. Motif 3 starts at 432 and ends at 469. In *Pongo pygmaeus* motif 1 starts at 913 and ends at 950. Motif 2 starts at 959 and ends at 999. Motif 3 starts at 641 and ends at 679. In *Pongo Abellii* motif 1 starts at 913 and ends at 950. Motif 2 starts at 959 and ends at 999. Motif 3 starts at 641 and ends at 678. In *Macaca fascicularis* motif 1 starts at 716 and ends at 753. Motif 2 starts at 762 and ends at 802. Motif 3 starts at 436 and ends at 473. In *Gorilla gorilla* motif 1 starts at 337 and ends at 374. Motif 2 starts at 883 and ends at 923. Motif 3 starts at 433 and ends at 470.

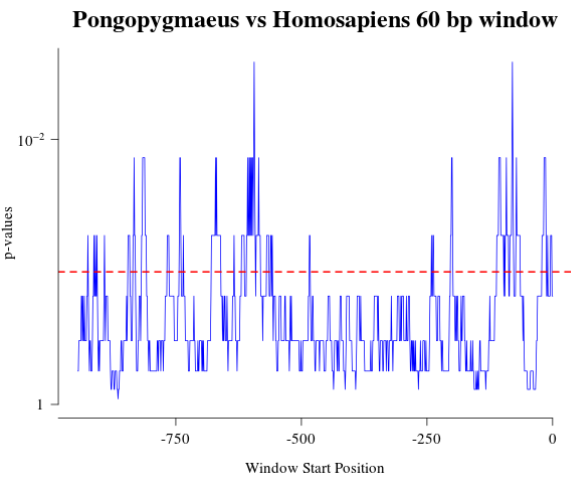
MAPK1



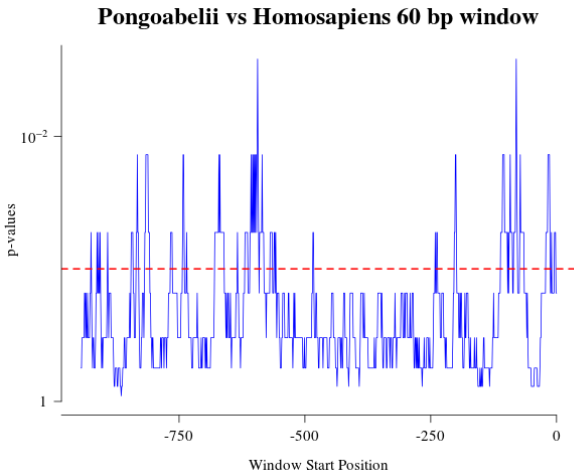
(a)



(b)



(c)



(d)

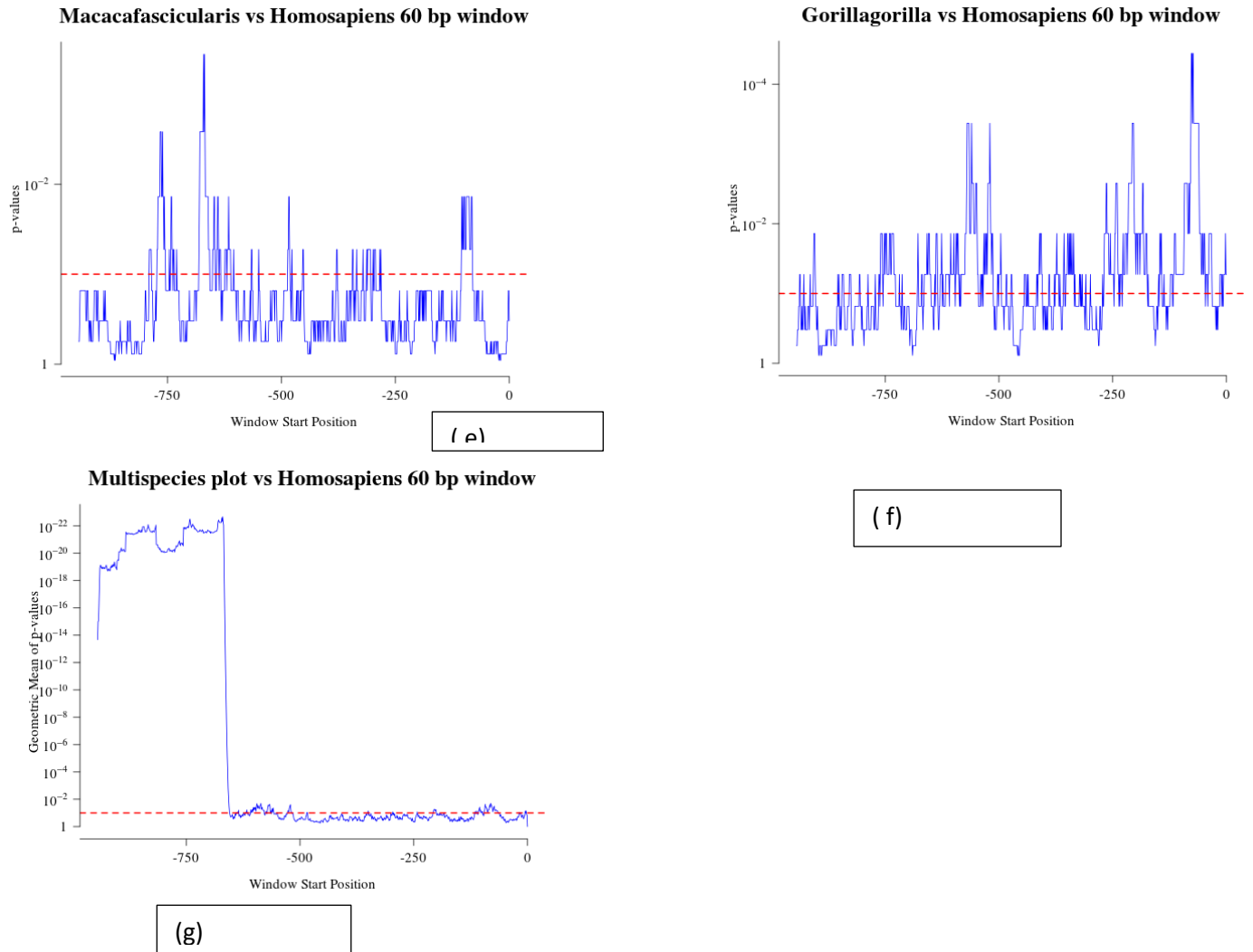
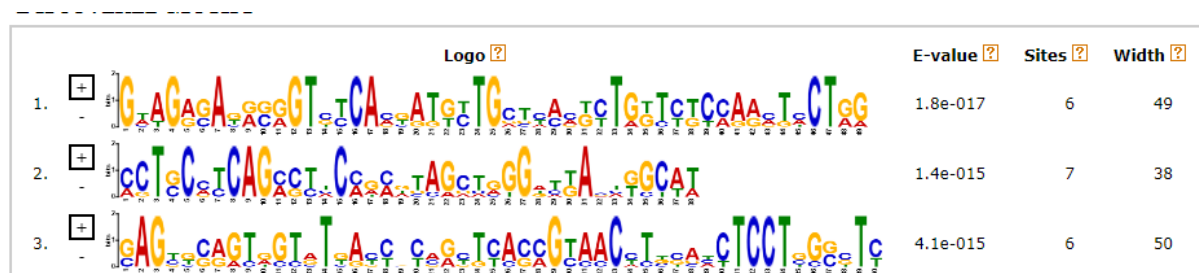


Figure no.: 3.18 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes 1(c) homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs



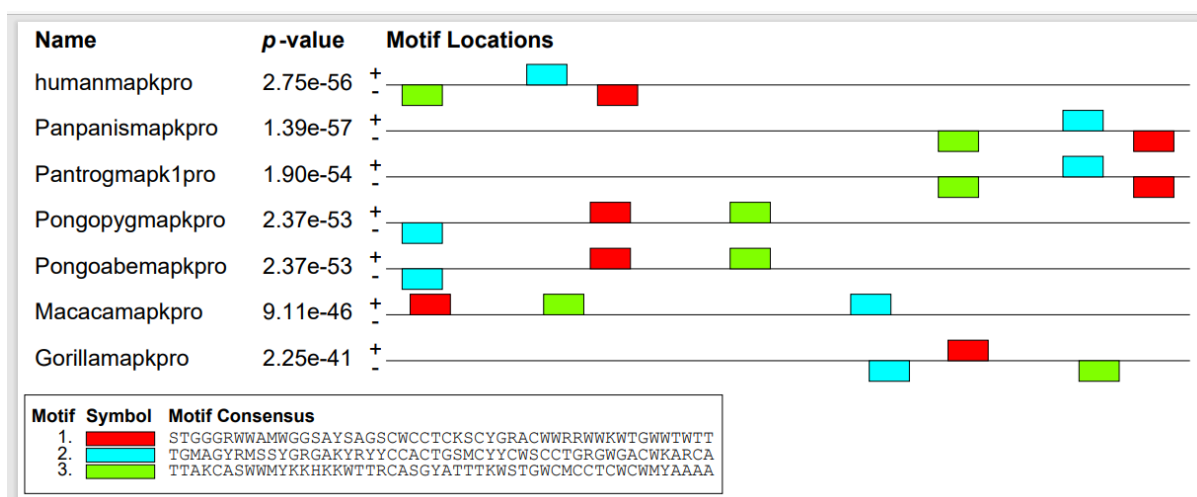
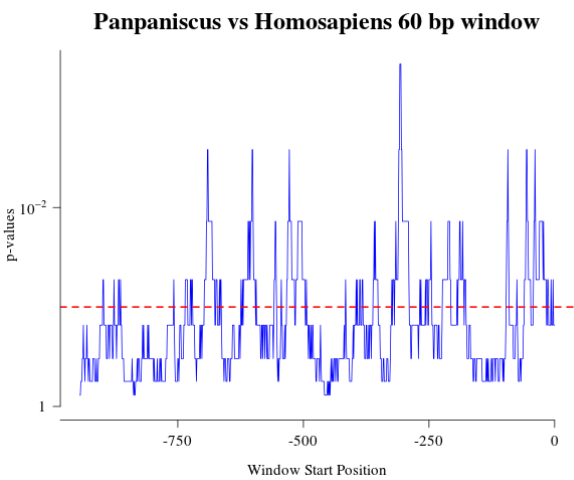


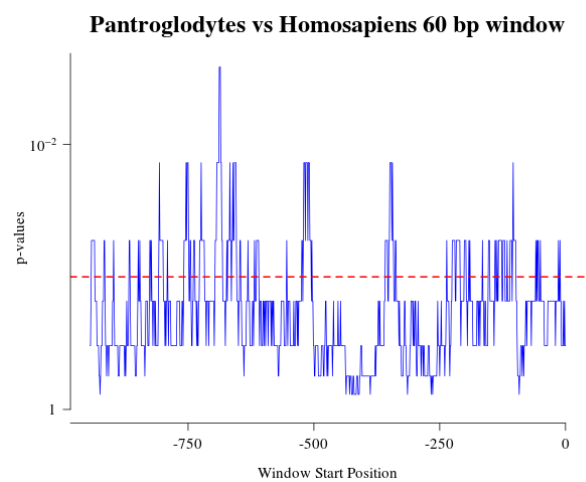
Figure 4.18 Motif identification MAPK1

In Homo sapiens motif 1 starts at 264 ends at 313. motif 2 starts at 176 and ends at 226. Motif 3 starts at 21 and ends at 70. In Pan paniscus motif 1 starts at 931 and ends at 980. Motif 2 starts at 843 and ends at 892. Motif 3 starts at 688 and ends at 737. In Pan troglodytes motif 1 starts at 931 and ends at 980. Motif 2 starts at 843 and ends at 892. Motif 3 starts at 688 and ends at 737. In Pongo pygmaeus motif 1 starts at 255 and ends at 304. Motif 2 starts at 21 and ends at 70. Motif 3 starts at 429 and ends at 478. In Pongo abelii motif 1 starts at 255 and ends at 304. Motif 2 starts at 21 and ends at 70. Motif 3 starts at 429 and ends at 478. In Macaca fascicularis motif 1 starts at 31 and ends at 80. Motif 2 starts at 579 and ends at 628. Motif 3 starts at 197 and ends at 246. In Gorilla gorilla motif 1 starts at 700 and ends at 749. Motif 2 starts at 602 and ends at 651. Motif 3 starts at 863 and ends at 912.

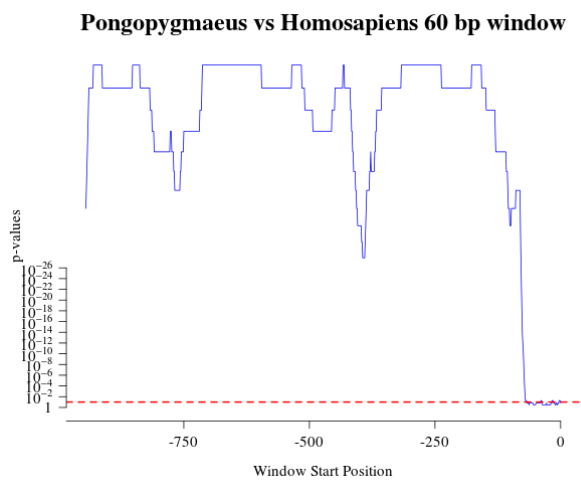
KLF5



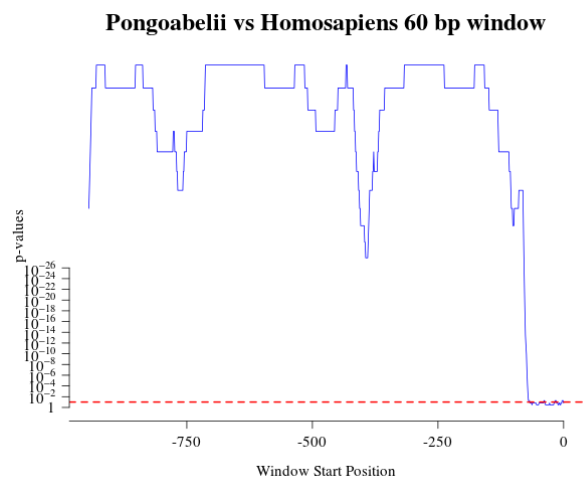
(a)



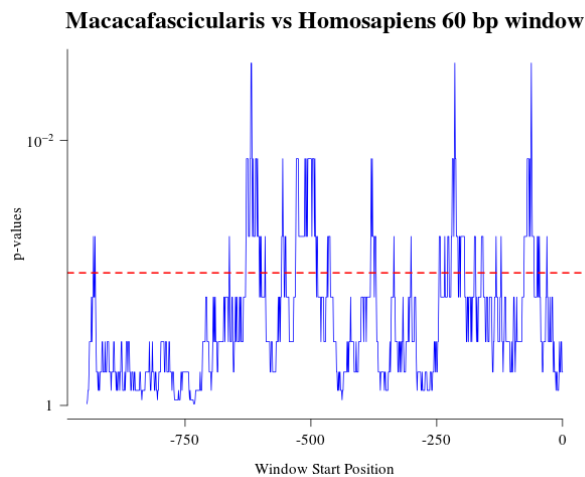
(b)



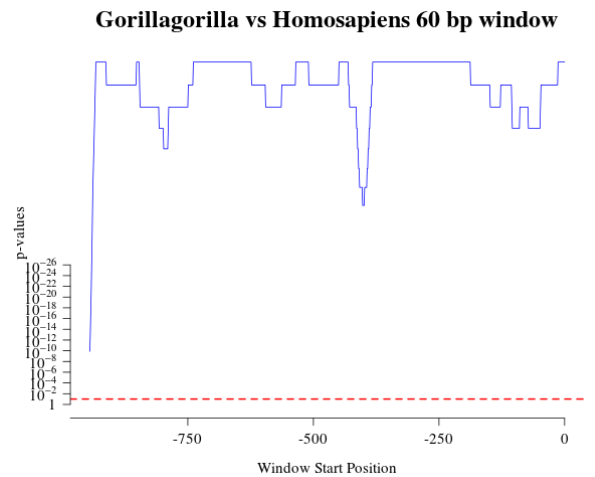
(c)



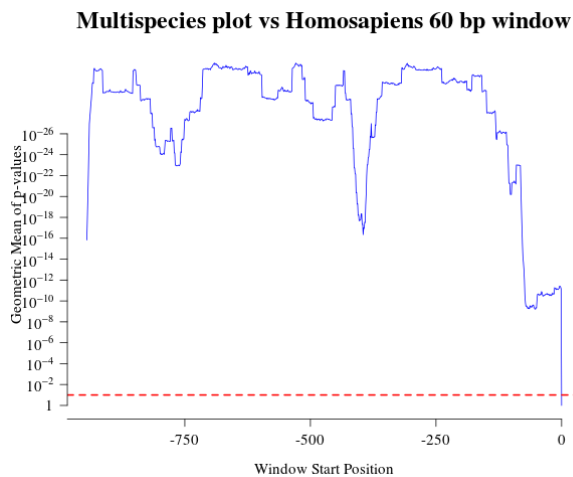
(d)



(e)



(f)



(g)

Figure no.: 3.19 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes 1(c) homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

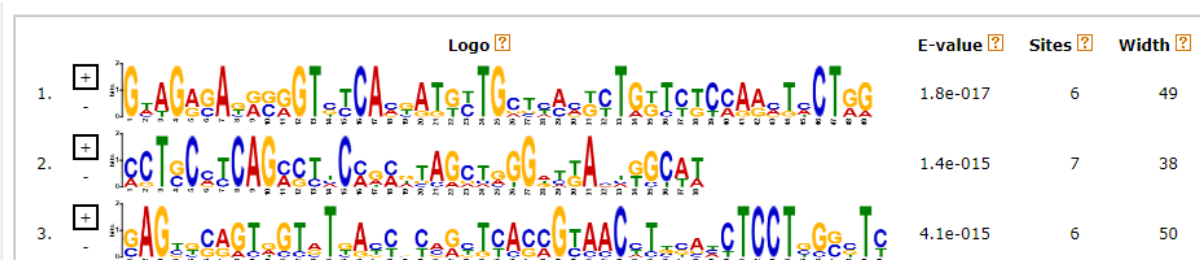


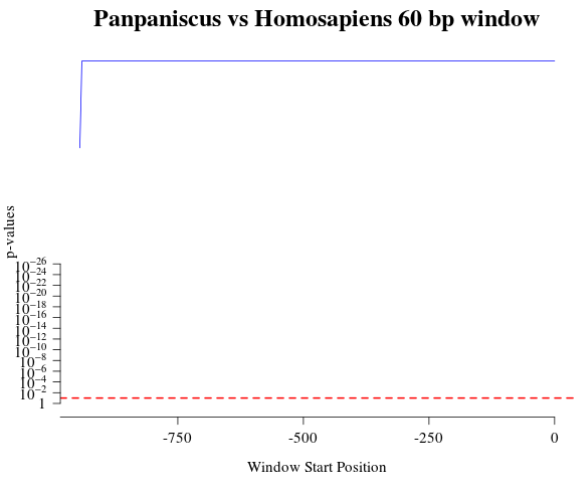


Figure 4.19 Motif identification KLF5

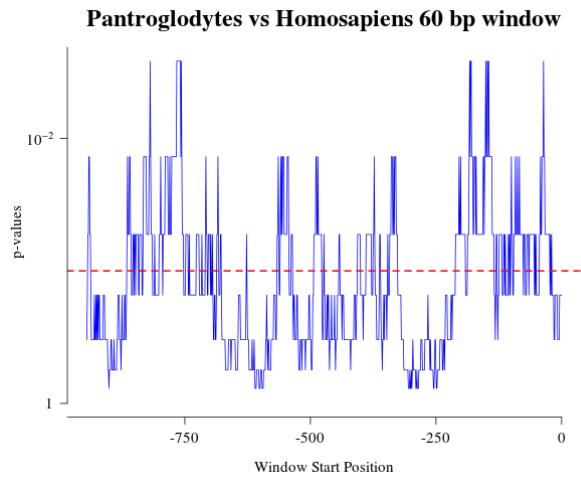
Homosapiens, Pan paniscus, Pongo pygmaeus, Pongo abelii and Gorilla gorilla show the presence of common three motifs.

In Homo sapiens motif 1 starts at 417 ends at 465. Motif 2 starts at 345 and ends at 382. Motif 3 starts at 285 and ends at 334. In Pan paniscus motif 1 starts at 301 and ends at 349. Motif 2 starts at 247 and ends at 284. Motif 3 starts at 187 and ends at 236. In Pongo pygmaeus motif 1 starts at 716 and ends at 764. Motif 2 starts at 676 and ends at 713. Motif 3 starts at 766 and ends at 815. In Pongo abelii motif 1 starts at 716 and ends at 764. Motif 2 starts at 676 and ends at 713. Motif 3 starts at 766 and ends at 815. In Macaca fascicularis motif 1 starts at 318 and ends at 366. Motif 2 starts at 792 and ends at 829. Motif 3 starts at 842 and ends at 891. In Gorilla gorilla motif 1 starts at 412 and ends at 460. Motif 2 starts at 340 and ends at 377. Motif 3 starts at 280 and ends at 329.

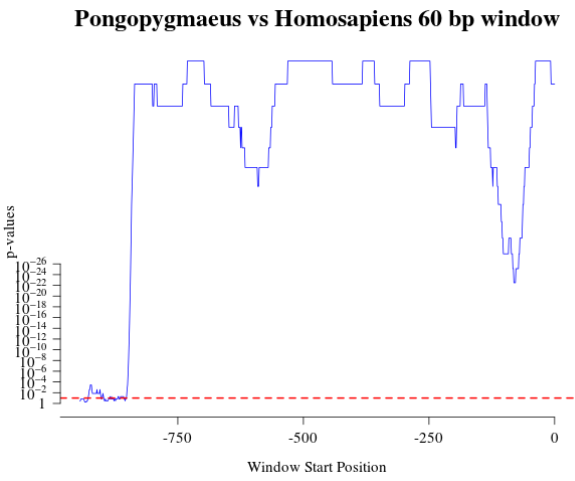
MC3R



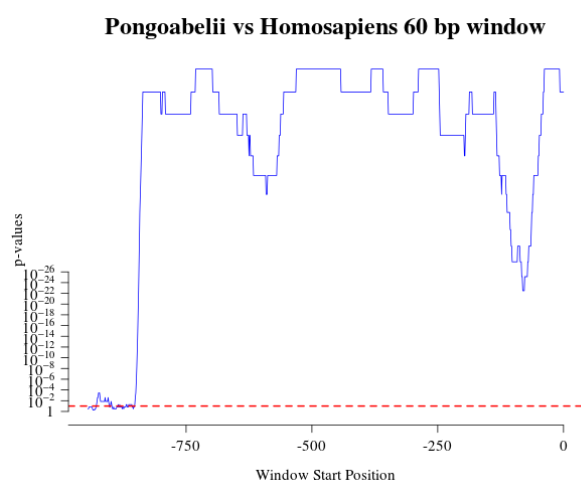
(a)



(b)

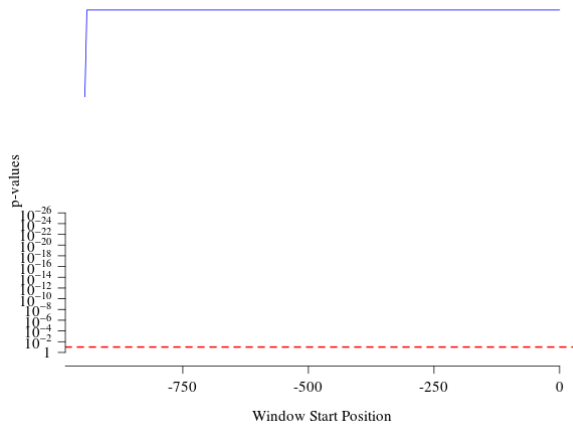


(c)



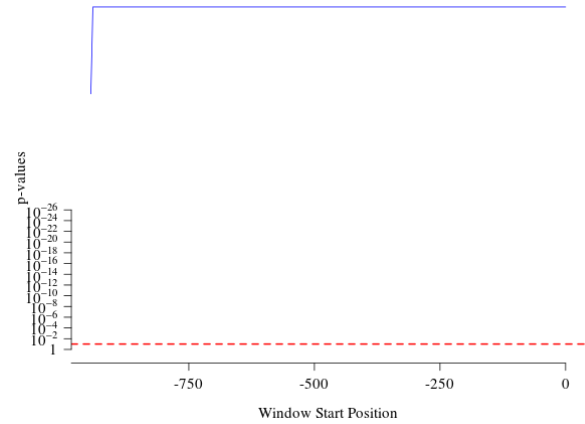
(d)

Macacafascularis vs Homosapiens 60 bp window



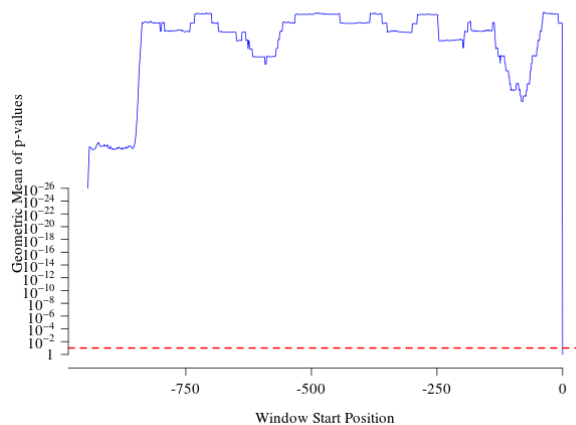
(e)

Gorillagorilla vs Homosapiens 60 bp window



(f)

Multispecies plot vs Homosapiens 60 bp window



(g)

Figure no. 3.20 Evolutionary analysis of regulatory sequences plot 1(a) Homosapiens vs panpaniscus 1(b) Homosapiens vs pantroglodytes 1(c) homosapiens va pongo pygmaeus (d) homosapiens vs pongo abelii (e) homosapiens vs macaca fascicularis (f) homosapiens vs gorilla (g) multispecies plot

Identification of motifs

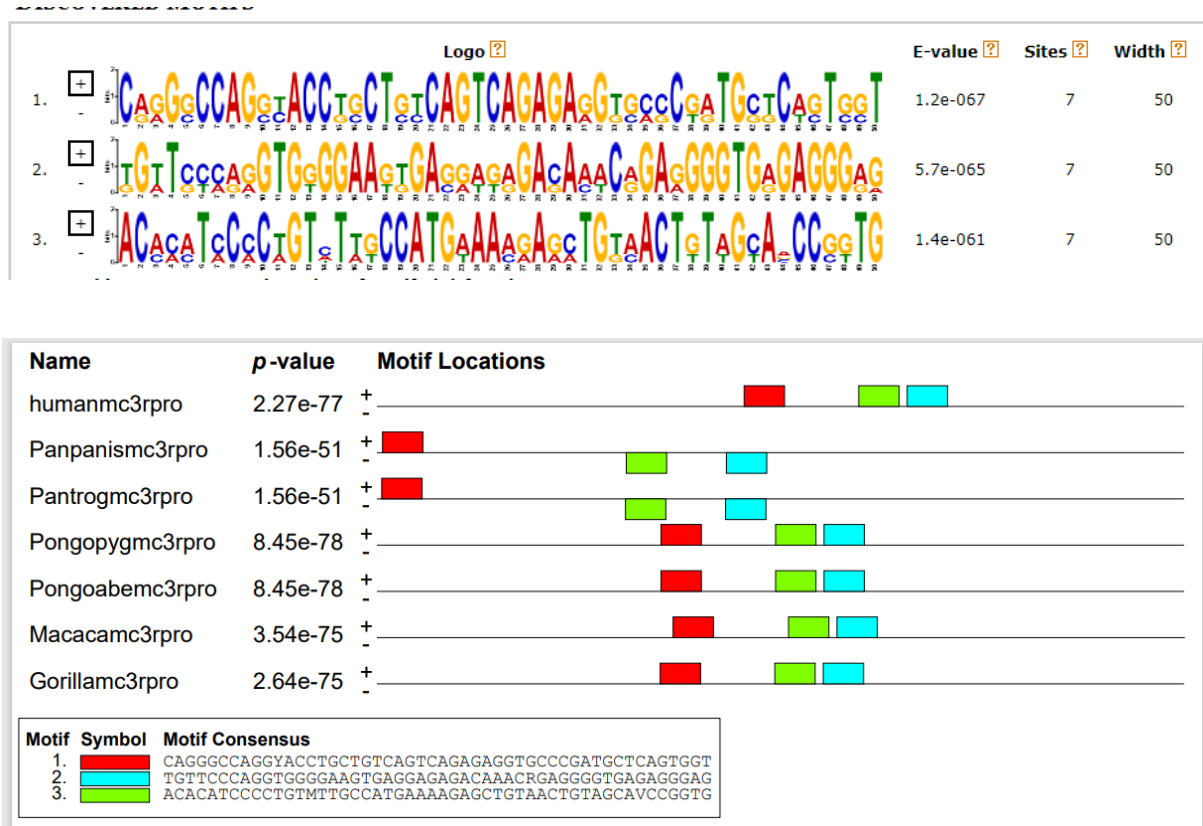


Figure 4.20 Motif identification MC3R

In Homo sapiens motif 1 starts at 456 ends at 505. Motif 2 starts at 658 and ends at 707. Motif 3 starts at 598 and ends at 647. In Pan paniscus motif 1 starts at 8 and ends at 57. Motif 2 starts at 434 and ends at 483. Motif 3 starts at 310 and ends at 353. In Pan troglodytes motif 1 starts at 7 and ends at 56. Motif 2 starts at 433 and ends at 482. Motif 3 starts at 309 and ends at 358. In Pongo pygmaeus motif 1 starts at 353 and ends at 402. Motif 2 starts at 555 and ends at 604. Motif 3 starts at 495 and ends at 544. In Pongo abelii motif 1 starts at 353 and ends at 402. Motif 2 starts at 555 and ends at 604. Motif 3 starts at 495 and ends at 544. In Macaca fascicularis motif 1 starts at 368 and ends at 417. Motif 2 starts at 571 and ends at 604. Motif 3 starts at 511 and ends at 560. In Gorilla gorilla motif 1 starts at 352 and ends at 401. Motif 2 starts at 554 and ends at 603. Motif 3 starts at 494 and ends at 543.

Transcription factors bind to motifs and play an essential role in gene regulation and evolution studies. The presence of common motifs suggests that these regions of the gene have been conserved throughout evolution. This conservation implies that these motifs are important for the function or regulation of the gene. Shared motifs among orthologs indicate that these motifs likely play a crucial role in the biological function of the gene. This could include motifs involved in DNA binding, protein-protein interactions, enzymatic activity, or other functional roles. The motifs are regulatory elements involved in controlling the expression of the gene. They are binding sites for transcription factors or other regulatory proteins, indicating that the gene is regulated in a similar manner across different species. If the motifs are found in protein-coding regions, they could represent conserved structural elements of the protein. This suggests that these regions are important for maintaining the structural integrity or function of the protein. The identification of common motifs provides targets for further experimental studies to elucidate their specific roles in gene function, regulation, or protein structure. This could involve mutagenesis studies, expression analyses, or biochemical assays to validate the functional significance of these motifs

MULTIPLE SEQUENCE ALIGNMENT USING UNIPROT ALIGN

The protein sequences of homo sapiens and the other 7 chosen orthologs (Pan paniscus, Pan troglodytes, Pongo pygmaeus, Pongo abelii, Macaca fascicularis and Gorilla gorilla gorilla) were aligned using the clustal omega program of uniprot align tool.

ACADM

Percent Identity Matrix

XP_005543049.2	100.00%	96.67%	95.96%	96.44%	96.20%	95.96%	96.20%
XP_018875470.2	96.67%	100.00%	99.29%	99.76%	99.52%	98.57%	98.57%
NP_001104286.1	95.96%	99.29%	100.00%	99.52%	99.76%	97.86%	NP_001104286.1 v
NP_000007.1	96.44%	99.76%	99.52%	100.00%	99.76%	98.34%	98.34%
XP_003830620.1	96.20%	99.52%	99.76%	99.76%	100.00%	98.10%	98.10%
XP_009247163.2	95.96%	98.57%	97.86%	98.34%	98.10%	100.00%	99.76%
XP_054350501.1	96.20%	98.57%	97.86%	98.34%	98.10%	99.76%	100.00%

FIGURE 5.1 ACADM PERCENTAGE IDENTITY MATRIX

The percentage identity values of alignment between human and the 7 orthologous genes is listed in the form of percentage identity matrix as obtained from uniprot align tool results

XP_005543049.2:Macacafascicularis,XP_018875470.2:Gorilla gorilla gorilla ,NP_001104286.1:Pantroglodytes,NP_000007.1:Homo sapiens,XP_003830620.1:Pan paniscus,XP_009247163.2:Pongo abelii ,XP_054350501.1: Pongo pygmaeus

The highest percentage identity is 99.76% which is the percentage identity shared between Pongo pygmaeus and Pongo abelii. The lowest percentage identity is 95.96% which is the percentage identity shared between Pan paniscus and Macaca fascicularis.

CPT1A

Percent Identity Matrix

XP_045228213.1	100.00%	96.16%	97.80%	97.80%	97.54%	97.54%	97.41%
NP_001027017.1	96.16%	100.00%	97.88%	97.88%	98.41%	98.41%	98.28%
XP_054295203.1	97.80%	97.88%	100.00%	100.00%	99.09%	99.09%	98.97%
XP_024111515.1	97.80%	97.88%	100.00%	100.00%	99.09%	99.09%	98.97%
XP_034789114.2	97.54%	98.41%	99.09%	99.09%	100.00%	100.00%	99.87%
XP_016776910.1	97.54%	98.41%	99.09%	99.09%	100.00%	100.00%	99.87%
XP_030871831.1	97.41%	98.28%	98.97%	98.97%	99.87%	99.87%	100.00%

FIGURE 5.2 CPT1A PERCENTAGE IDENTITY MATRIX

The percentage identity values of alignment between human and the 7 orthologous genes is listed in the form of percentage identity matrix as obtained from uniprot align tool results

XP_045228213.1:Macaca fascicularis NP_001027017.1:Homo sapiens,XP_054295203.1:Pongo pygmaeus ,XP_024111515.1:Pongo abelii,XP_034789114.2:Pan paniscus ,XP_016776910.1:Pan troglodytes ,XP_030871831.1: Gorilla gorilla gorilla

The highest percentage identity value is 99.87% which is shared between Gorilla gorilla gorilla and Pan paniscus.The lowest percentage

identity value is 96.16% which is shared between Homo sapiens and Macaca fascicularis

DGAT1

Percent Identity Matrix							
XP_005564399.2	100.00%	96.31%	96.31%	96.31%	96.11%	95.70%	95.90%
NP_036211.2	96.31%	100.00%	100.00%	100.00%	99.80%	98.36%	98.57%
XP_003819477.1	96.31%	100.00%	100.00%	100.00%	99.80%	98.36%	98.57%
XP_016815597.1	96.31%	100.00%	100.00%	100.00%	99.80%	98.36%	98.57%
XP_018887719.1	96.11%	99.80%	99.80%	99.80%	100.00%	98.16%	98.36%
XP_024106800.2	95.70%	98.36%	98.36%	98.36%	98.16%	100.00%	99.80%
XP_054355261.1	95.90%	98.57%	98.57%	98.57%	98.36%	99.80%	100.00%

Figure no 5.3 DGAT1 Percentage Identity Matrix

XP_005564399.2 : Macaca fascicularis , NP_036211.2 : Homo sapiens ,XP_003819477.1: Pan paniscus ,XP_016815597.1: Pan troglodytes,XP_018887719.9: Gorilla gorilla gorilla,XP_024106800.2: Pongo abelii XP_054355261.1 :Pongo pygmaeus

SLC25A1

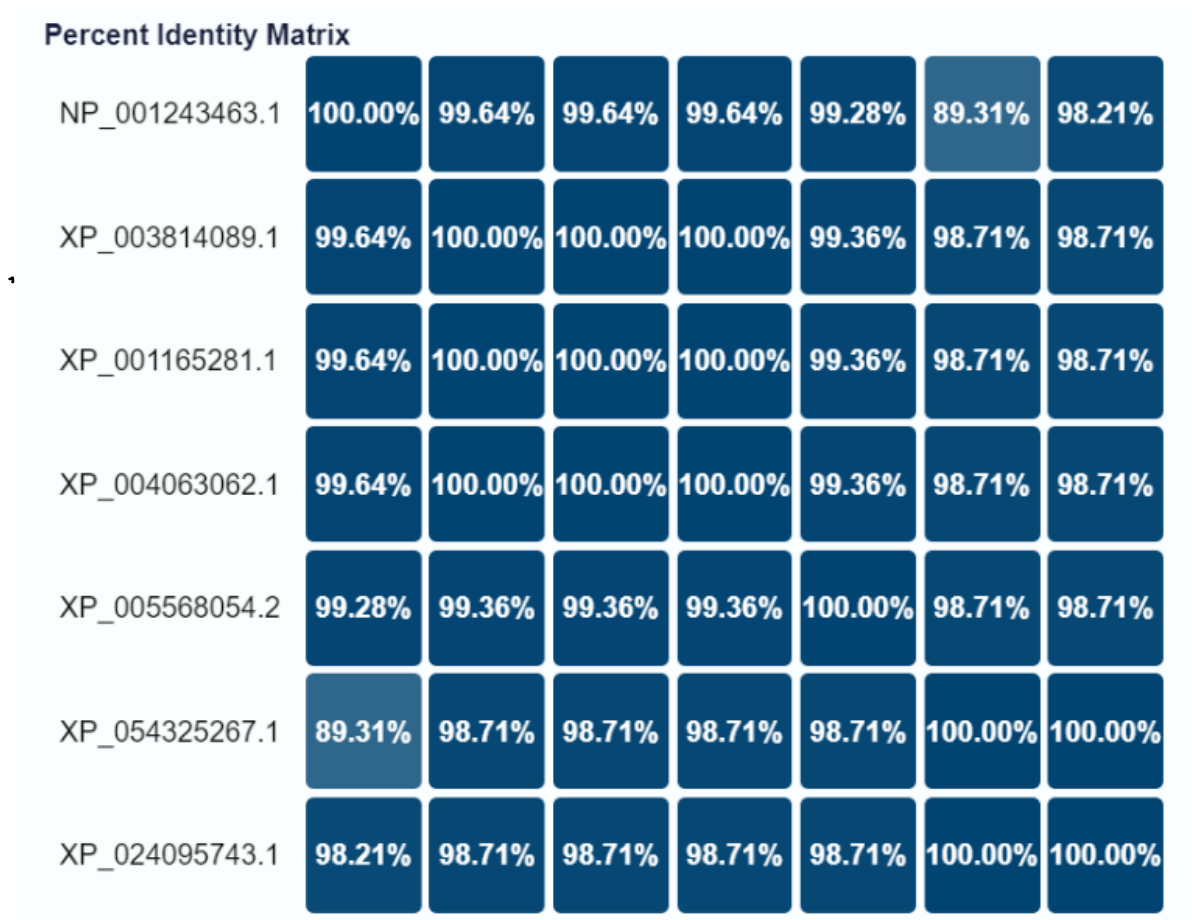


Figure 5.4 SLC25A1 Percentage Identity Matrix

NP_001243463.1: Homo sapiens, XP_003814089.1: Pan paniscus, XP_001165281.1: Pan troglodytes, XP_004063062.1: Gorilla gorilla gorilla, XP_005568054.2: Macaca fascicularis

The highest percentage identity value is 99.64% which is shared between Pan paniscus and Homo sapiens. The lowest percentage identity value is 89.31% which is shared between Gorilla gorilla gorilla and Homo sapiens

HADH

Percent Identity Matrix							
XP_045248662.1	100.00%	96.50%	96.50%	96.50%	96.82%	96.50%	96.18%
XP_054342637.1	96.50%	100.00%	100.00%	98.41%	97.77%	98.41%	98.41%
XP_009238520.3	96.50%	100.00%	100.00%	98.41%	97.77%	98.41%	98.41%
NP_001171634.3	96.50%	98.41%	98.41%	100.00%	98.73%	99.36%	99.04%
XP_004040299.2	96.82%	97.77%	97.77%	98.73%	100.00%	98.73%	98.41%
XP_003830043.2	96.50%	98.41%	98.41%	99.36%	98.73%	100.00%	99.68%
XP_001135650.3	96.18%	98.41%	98.41%	99.04%	98.41%	99.68%	100.00%

Figure 5.5 HADH Percentage Identity Matrix

XP_045248662.1:Macaca fascicularis XP_054342637.1:Pongo pygmaeus
 XP_009238520.3 : Pongo abelii NP_001171634.3: Homo sapiens
 ,XP_004040299.2:Pan paniscusXP_003830043.2: Pan
 troglodytesXP_001135650.3: Gorilla gorilla gorilla

The highest percentage identity value is 99.68% which is shared between Pan troglodytes and Gorilla gorilla gorilla. The lowest percentage identity value is 96.18 which is shared between Gorilla gorilla gorilla and Macaca fascicularis.

PKM

Percent Identity Matrix							
XP_005560022.1	100.00%	99.62%	99.81%	95.67%	95.67%	95.67%	95.29%
XP_054307466.1	99.62%	100.00%	99.81%	92.89%	95.67%	95.67%	95.67%
XP_004056483.1	99.81%	99.81%	100.00%	95.86%	95.86%	95.86%	95.48%
NP_001193725.1	95.67%	92.89%	95.86%	100.00%	100.00%	100.00%	99.62%
XP_003818599.1	95.67%	95.67%	95.86%	100.00%	100.00%	100.00%	99.62%
XP_001175057.2	95.67%	95.67%	95.86%	100.00%	100.00%	100.00%	99.62%
NP_001127083.1	95.29%	95.67%	95.48%	99.62%	99.62%	99.62%	100.00%

Figure 5.6 PKM Percentage Identity Matrix

XP_005560022.1: Macaca fascicularis, XP_054307466.1: Pongo pygmaeus, XP_004056483.1: Gorilla gorilla gorilla, NP_001193725.1: Human, XP_003818599.1: Pan paniscus, XP_001175057: Pan troglodytes, NP_001127083.1: Pongo abelii

The highest percentage identity value is 99.81% which is shared between Gorilla gorilla gorilla and Macaca fascicularis. The lowest percentage identity value is 92.89% which is shared between Homo sapiens and Pongo pygmaeus.

ACLY

Percent Identity Matrix

XP_054313511.1	100.00%	100.00%	99.57%	99.65%	99.82%	99.57%	99.73%
XP_024090256.1	100.00%	100.00%	99.57%	99.65%	99.82%	99.57%	99.73%
XP_034797740.1	99.57%	99.57%	100.00%	99.91%	99.91%	99.83%	99.82%
XP_024206071.1	99.65%	99.65%	99.91%	100.00%	99.82%	99.74%	99.73%
NP_001087.2	99.82%	99.82%	99.91%	99.82%	100.00%	100.00%	99.91%
XP_018882643.3	99.57%	99.57%	99.83%	99.74%	100.00%	100.00%	99.91%
XP_005584261.1	99.73%	99.73%	99.82%	99.73%	99.91%	99.91%	100.00%

Figure 5.7 ACLY Percentage Identity Matrix

XP_054313511.1:Pongo pygmaeus , XP_024090256.1:Pongo abelii
 XP_034797740.1: Pan paniscus XP_024206071.1:Pan troglodytes
 NP_001087.2:Homo sapiens XP_018882643.3:Gorilla gorilla gorilla
 XP_005584261.1:Macaca fascicularis

The highest percentage identity value is 99.71% which is shared between Pan troglodytes and Pan paniscus. The lowest percentage identity value is 99.57% which is shared between Pan paniscus and Pongo pygmaeus.

ACACA

Percent Identity Matrix

XP_005583987.2	100.00%	96.34%	96.80%	96.84%	96.72%	96.72%	96.88%
XP_024090390.1	96.34%	100.00%	99.54%	99.24%	99.16%	99.16%	99.79%
XP_054312872.1	96.80%	99.54%	100.00%	99.71%	99.62%	99.62%	99.79%
XP_030867177.2	96.84%	99.24%	99.71%	100.00%	99.79%	99.79%	99.87%
NP_942131.1	96.72%	99.16%	99.62%	99.79%	100.00%	99.92%	99.96%
XP_003817996.1	96.72%	99.16%	99.62%	99.79%	99.92%	100.00%	100.00%
XP_003953107.3	96.88%	99.79%	99.79%	99.87%	99.96%	100.00%	100.00%

Figure 5.8 ACACA Percentage identity matrix

XP_005583987.2: *Macaca fascicularis*, XP_024090390.1: *Pongo abelii*, XP_054312872.1: *Pongo pygmaeus*, XP_030867177.2: *Gorilla gorilla gorilla*, NP_942131.1: *Homo sapiens*, XP_003817996.1: *Pan paniscus*, XP_003953107.3: *Pan troglodytes*

The highest percentage value is 99.96 which is shared between *Homo sapiens* and *Pan troglodytes*. The lowest percentage value is 96.34 which is shared between *Pongo abelii* and *Macaca fascicularis*.

MLXIPL

Percent Identity Matrix							
XP_045245613.1	100.00%	97.72%	98.12%	98.12%	98.36%	98.47%	98.12%
XP_054349769.2	97.72%	100.00%	99.66%	98.12%	98.36%	98.47%	98.12%
XP_002817908.4	98.12%	99.66%	100.00%	98.36%	98.59%	98.71%	98.36%
XP_030868563.1	98.12%	98.12%	98.36%	100.00%	98.83%	99.18%	98.83%
NP_116569.1	98.36%	98.36%	98.59%	98.83%	100.00%	99.41%	99.06%
XP_003815621.1	98.47%	98.47%	98.71%	99.18%	99.41%	100.00%	99.65%
XP_003318716.2	98.12%	98.12%	98.36%	98.83%	99.06%	99.65%	100.00%

Figure 5.9 MLXIPL Percentage Identity Matrix

XP_045245613.1: *Macaca fascicularis*, XP_054349769.2: *Pongo pygmaeus*,

XP_002817908.4: *Pongo abelii*, XP_030868563.1: *Gorilla gorilla gorilla*,

NP_116569.1: *Homo sapiens*, XP_003815621.1: *Pan paniscus*,

XP_003318716.2: *Pan troglodytes*

The highest percentage similarity is 99.65% which is the percentage similarity shared between *Pan paniscus* and *Pan troglodytes*. The lowest percentage similarity is the percentage similarity between *Gorilla gorilla* and *Pongo pygmaeus*

SCD

Percent Identity Matrix

XP_005566247.1	100.00%	95.26%	95.26%	95.26%	95.54%	95.54%	95.26%
XP_003846166.1	95.26%	100.00%	100.00%	98.89%	99.16%	99.16%	98.89%
XP_001168563.1	95.26%	100.00%	100.00%	98.89%	99.16%	99.16%	98.89%
NP_005054.3	95.26%	98.89%	98.89%	100.00%	99.16%	99.16%	98.89%
XP_004049993.1	95.54%	99.16%	99.16%	99.16%	100.00%	99.44%	99.16%
XP_054292887.1	95.54%	99.16%	99.16%	99.16%	99.44%	100.00%	99.72%
NP_001125731.1	95.26%	98.89%	98.89%	98.89%	99.16%	99.72%	100.00%

Figure 5.10 SCD Percentage Identity Matrix

XP_005566247.1: *Macaca Fascicularis*, XP_003846166.1 *Pan paniscus*, XP_001168563.1: *Pan troglodytes*, NP_005054.3 : Human, XP_004049993.1: *Gorilla gorilla gorilla*, XP_054292887.1: *Pongo pygmaeus*, NP_001125731.1: *Pongo abelii*.

The highest percentage identity value is 99.72 which is shared between *Pongo pygmaeus* and *Pongo abelii*. The lowest percentage identity value: 95.26 which is shared between *Pan paniscus* and *Macaca fascicularis*.

GHSR

Percent Identity Matrix

NP_004113.1	100.00%	94.64%	96.43%	96.43%	96.43%	95.71%	95.71%
XP_005546450.1	94.64%	100.00%	98.63%	98.63%	98.63%	98.36%	98.36%
XP_003831989.1	96.43%	98.63%	100.00%	100.00%	100.00%	99.45%	99.45%
XP_054537223.1	96.43%	98.63%	100.00%	100.00%	100.00%	99.45%	99.45%
XP_004038058.1	96.43%	98.63%	100.00%	100.00%	100.00%	99.45%	99.45%
XP_054339110.1	95.71%	98.36%	99.45%	99.45%	99.45%	100.00%	100.00%
XP_002814334.1	95.71%	98.36%	99.45%	99.45%	99.45%	100.00%	100.00%

Figure 5.11 GHSR Percentage Identity Matrix

NP_004113.1: Human sapiens, XP_005546450.1: Macaca fascicularis, XP_003831989.1: Pan paniscus, XP_054537223.1: Pan troglodytes, XP_004038058.1: Gorilla gorilla gorilla, XP_054339110.1: Pongo pygmaeus, XP_002814334.1: Pongo abelii.

The highest percentage identity value is 99.45 which is shared between Pongo pygmaeus and Pan paniscus. The lowest percentage identity value is 95.71 which is shared between Pongo pygmaeus and Homo sapiens.

NPY

Percent Identity Matrix

XP_005550032.1	100.00%	97.94%	97.94%	97.94%	97.94%	96.91%	96.91%
XP_054350800.1	97.94%	100.00%	100.00%	97.94%	97.94%	96.91%	96.91%
XP_002818184.1	97.94%	100.00%	100.00%	97.94%	97.94%	96.91%	96.91%
NP_000896.1	97.94%	97.94%	97.94%	100.00%	100.00%	98.97%	98.97%
XP_004045235.2	97.94%	97.94%	97.94%	100.00%	100.00%	98.97%	98.97%
XP_003807978.1	96.91%	96.91%	96.91%	98.97%	98.97%	100.00%	100.00%
XP_001159094.1	96.91%	96.91%	96.91%	98.97%	98.97%	100.00%	100.00%

Figure 5.12 NPY Percentage Identity Matrix

XP_005550032.1:Macaca fascicularis, XP_054350800.1:Pongo pygmaeus,XP_002818184.4: Pongo abelii,NP_00089.1: Homo sapiens, XP_004045235.2: Gorilla gorilla,XP_003807978.1:Pan paniscus,XP_001159054.1:Pan troglodytes

The highest percentage identity value was 98.97% which was shared between Gorilla gorilla and Homo sapiens.The lowest percentage identity value was Pan paniscus and Macaca fascicularis.

AGRP

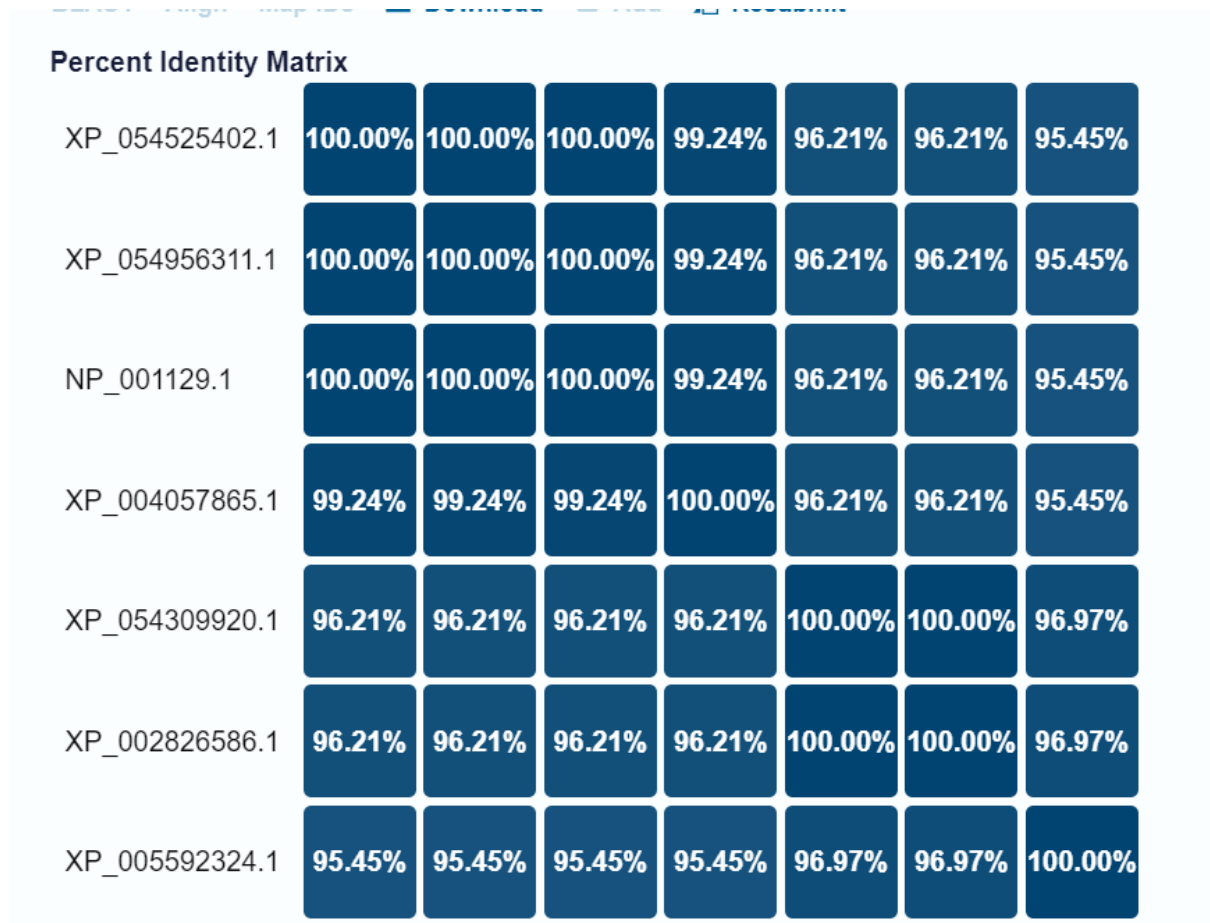


Figure 5.13 AGRP Percentage Identity Matrix

XP_054525402.1: Pan troglodytes, XP_054956311.1: Pan paniscus, NP_001129.1: Homo sapiens, XP_004057865.1: Gorilla gorilla gorilla, XP_054309920.1: Pongo pygmaeus, XP_002826586.1: Pongo abelii, XP_005592324.1: Macaca fascicularis.

The highest percentage identity value is 99.24 which is shared between Gorilla gorilla gorilla and Pan troglodytes. The lowest percentage identity value is 95.45 which is shared between Macaca fascicularis and Pan troglodytes.

FOXO1

Percent Identity Matrix							
XP_018895388.3	100.00%	88.44%	89.41%	89.41%	89.41%	89.07%	89.07%
XP_045232830.1	88.44%	100.00%	98.78%	98.78%	98.78%	98.93%	98.93%
NP_002006.2	89.41%	98.78%	100.00%	100.00%	100.00%	99.69%	99.69%
XP_024781048.3	89.41%	98.78%	100.00%	100.00%	100.00%	99.69%	99.69%
XP_522749.3	89.41%	98.78%	100.00%	100.00%	100.00%	99.69%	99.69%
XP_054301999.1	89.07%	98.93%	99.69%	99.69%	99.69%	100.00%	100.00%
XP_002824246.1	89.07%	98.93%	99.69%	99.69%	99.69%	100.00%	100.00%

Figure 5.14 FOXO1 Percentage Identity Matrix

XP_018895388.3: Gorilla gorilla gorilla, XP_045232830.1: Macaca fascicularis, NP_002006.2: Homo sapiens, XP_024781048.3: Pan paniscus, XP_522749.3: Pan troglodytes, XP_054301999.1: Pongo pygmaeus, XP_002824246.1: Pongo abelii

The highest percentage identity value is 99.69% which is shared between Pongo pygmaeus and Homo sapiens. The lowest percentage identity value is 88.44 which is shared between Macaca fascicularis and Gorilla gorilla gorilla.

ADCY3

Percent Identity Matrix

XP_054331192.1	100.00%	99.83%	95.66%	95.92%	95.75%	95.93%	95.93%
XP_024098015.2	99.83%	100.00%	95.66%	95.92%	95.75%	95.93%	95.93%
XP_005576474.1	95.66%	95.66%	100.00%	98.86%	98.78%	98.86%	98.86%
XP_004028977.3	95.92%	95.92%	98.86%	100.00%	99.56%	99.83%	99.83%
NP_001307542.1	95.75%	95.75%	98.78%	99.56%	100.00%	99.74%	99.74%
XP_008950647.1	95.93%	95.93%	98.86%	99.83%	99.74%	100.00%	100.00%
XP_016803639.1	95.93%	95.93%	98.86%	99.83%	99.74%	100.00%	100.00%

Figure 5.15 ADCY3 Percentage Identity Matrix

XP_054331192.1: Pongo pygmaeus, XP_024098015.2 Pongo abelii, XP_005576474.1: Macaca fascicularis, XP_004028977.3: Gorilla gorilla gorilla, NP_001307542.1: Homo sapiens, XP_008950647.1: Pan paniscus, XP_016803639.1: Pan troglodytes

The highest percentage identity value is 99.83 which is shared between Pongo abelii and Pongo pygmaeus. The lowest percentage identity value is 95.66 which is shared between Macaca fascicularis and Pongo pygmaeus.

APLN



Figure 5.16 APLN Percentage identity matrix

XP_005594596.1: *Macaca fascicularis*, NP_059109.3: *Homo sapiens*,
 XP_024783151.2: *Pan paniscus*, XP_016799545.1: *Pan troglodytes*,
 XP_054327148.1: *Pongo pygmaeus*, XP_002832145.2: *Pongo abelii*,
 XP_018874376.1: *Gorilla gorilla gorilla*

The highest percentage identity value is 98.7% which is shared between *Pongo pygmaeus* and *Homo sapiens*. The lowest value is 94.81 which is shared between *Homo sapiens* and *Macaca fascicularis*.

APLNR

Percent Identity Matrix							
XP_005577930.2	100.00%	98.68%	98.68%	98.42%	98.42%	97.89%	97.63%
NP_005152.1	98.68%	100.00%	100.00%	99.74%	99.74%	99.21%	98.95%
XP_055211341.1	98.68%	100.00%	100.00%	99.74%	99.74%	99.21%	98.95%
XP_003826487.1	98.42%	99.74%	99.74%	100.00%	100.00%	98.95%	98.68%
XP_521984.1	98.42%	99.74%	99.74%	100.00%	100.00%	98.95%	98.68%
XP_054294326.1	97.89%	99.21%	99.21%	98.95%	98.95%	100.00%	99.74%
NP_001124631.1	97.63%	98.95%	98.95%	98.68%	98.68%	99.74%	100.00%

Figure 5.17 APLNR Percentage identity matrix

XP_005577930.2: *Macaca Fascicularis*, NP_005152.1: *Homo sapiens*, XP_055211341.1: *Gorilla gorilla gorilla*, XP_003826487.1: *Pan paniscus*, XP_521984.1: *Pan troglodytes*, XP_054294326.1: *Pongo Pygmaeus*, NP_001124631.1: *Pongo abelii*

The highest percentage identity value is 99.74 which is shared between *Pan paniscus* and *homo sapiens*. The lowest percentage identity value is 97.63 which is shared between *Pongo Abelii* and *Macaca Fascicularis*.

MAPK1

Percent Identity Matrix							
NP_002736.3	100.00%	100.00%	100.00%	100.00%	99.72%	99.72%	99.72%
XP_034804682.1	100.00%	100.00%	100.00%	100.00%	99.72%	99.72%	99.72%
XP_003317171.1	100.00%	100.00%	100.00%	100.00%	99.72%	99.72%	99.72%
XP_005567920.1	100.00%	100.00%	100.00%	100.00%	99.72%	99.72%	99.72%
XP_054325037.1	99.72%	99.72%	99.72%	99.72%	100.00%	99.73%	99.72%
XP_004063154.3	99.72%	99.72%	99.72%	99.72%	99.73%	100.00%	99.72%
XP_024095453.1	99.72%	99.72%	99.72%	99.72%	99.72%	99.72%	100.00%

Figure 5.18 MAPK1 Percentage identity matrix

NP_002736.3: Homo sapiens, XP_034804682.1: Pan Paniscus, XP_0033171171.1: Pan troglodytes, XP_005567920.1: Macaca Fascicularis, XP_054325037.1: Pongo Pygmaeus, XP_004063154.3: Gorilla gorilla gorilla, XP_024095453.1Pongo Abellii

The highest percentage identity value is 99.73 which is shared between Gorilla gorilla gorilla and pongo pygmaeus. The lowest percentage identity value is 99.72 which is shared between homo sapiens and Gorilla gorilla gorilla

KLF5

Percent Identity Matrix							
XP_005586053.1	100.00%	99.73%	99.34%	99.34%	99.34%	99.56%	99.56%
NP_001273747.1	99.73%	100.00%	99.73%	99.73%	99.73%	99.73%	99.73%
XP_004054643.1	99.34%	99.73%	100.00%	99.56%	99.56%	99.78%	99.78%
XP_034792680.1	99.34%	99.73%	99.56%	100.00%	100.00%	99.78%	99.78%
XP_001136935.1	99.34%	99.73%	99.56%	100.00%	100.00%	99.78%	99.78%
XP_054302964.1	99.56%	99.73%	99.78%	99.78%	99.78%	100.00%	100.00%
XP_002824380.2	99.56%	99.73%	99.78%	99.78%	99.78%	100.00%	100.00%

Figure 5.19 KLF5 Percentage Identity matrix

XP_005586053.1:Macaca Fascicularis,NP_001273747.1: Homo Sapiens,
 XP_004054643.1: Gorilla gorilla gorilla, XP_034792680.1:Pan Paniscus,
 XP_001136935.1:Pan Troglodytes, XP_054302964.1: Pongo pygmaeus,
 XP_002824380.2: Pongo abelii

The highest percentage identity is 99.78% which is the percentage identity shared between Pongo pygmaeus and Gorilla gorilla. The lowest percentage identity is 99.34% which is the percentage identity shared between Gorilla gorilla and Macaca fascicularis.

MC3R

Percent Identity Matrix								
	XP_005569446.2	100.00%	90.68%	90.96%	94.12%	91.24%	91.24%	94.74%
	XP_054324202.1	90.68%	100.00%	99.72%	98.45%	98.61%	98.61%	99.07%
	XP_024094848.1	90.96%	99.72%	100.00%	98.76%	98.89%	98.89%	99.38%
	NP_063941.3	94.12%	98.45%	98.76%	100.00%	99.38%	99.38%	99.38%
	XP_034803098.1	91.24%	98.61%	98.89%	99.38%	100.00%	100.00%	100.00%
	XP_525362.1	91.24%	98.61%	98.89%	99.38%	100.00%	100.00%	100.00%
	XP_004062445.3	94.74%	99.07%	99.38%	99.38%	100.00%	100.00%	100.00%

FIGURE 5.20 Percentage Identity Matrix MC3R

XP_005569446.2 *macaca fascicularis*, XP_054324202.3 *Pongo pygmaeus*, XP_024094848.1 *Pongo abelii*, NP_063941.3 Human, XP_034803098.1 *Pan paniscus*, XP_525362.1 *Pan troglodytes*,

The highest percentage identity is 99.72% which is the percentage identity shared between *Pongo abelii* and *Pongo pygmaeus*. The lowest percentage identity is 90.68% which is the percentage identity shared between *Pongo pygmaeus* and *Macaca fascicularis*.

A high percentage identity between two proteins suggests that they are similar in structure and function. Conversely, a low percentage identity suggests that the proteins are less similar and may have different functions.

Excess sequence similarity reflects common ancestry. Homology is inferred when two sequences or structures share more similarity than would be expected by chance; when excess similarity is observed, that the two sequences did not arise independently, they arose from a common ancestor.

Homologous sequences have similar structures, and frequently, they have similar functions as well. One protein can have more than one function. Currently, the most popular strategy for inferring functional similarity is to focus on “orthologs”, typically understood as the same protein in different organisms. The concept of “orthology” was originally introduced to distinguish two kinds of evolutionary histories: homologous sequences that differ because of speciation events—orthologs; and homologous sequences that were produced by gene-duplication events—paralogs.

CHAPTER 6 : SUMMARY AND CONCLUSION

Phylogeny is the study of evolutionary history and relationships across groups of organisms. It delves into the ancestral relatedness of living species. Phylogenetic tree is a diagram representing hypothesized relationships among a group of organisms. Rooted phylogenetic trees and unrooted phylogenetic trees are two different types of Phylogenetic trees. A rooted tree indicates the hypothetical common ancestor of a tree. An unrooted tree makes no assumption about the ancestral line and doesn't show origin or root of the taxa. Phylogeny is a fascinating field that sheds light on the intricate tapestry of life's evolutionary journey.

Obesity is a disease that involves the occurrence of increased body mass. It occurs in the population worldwide. Understanding the evolutionary basis of obesity and the gene functions and association of the genes in obesity is crucial to reducing the increasing rates of obesity and finding suitable target genes for obesity research and prevention of the disorder.

The main goal of this project is to identify significant metabolic pathways in the human body that are significantly associated with obesity. Changes in these pathways may be deeply associated with obesity conditions. The main purpose of this project is to create phylogenetic trees and perform multiple sequence alignment.

The genes from obesity related pathways (Fatty acid Oxidation, De novo lipogenesis, Ghrelin signalling pathway, Apelin signalling pathways,) were studied. Their expression profiles were analysed deeply. They were analysed by comparison with their orthologues from closely related species; primates like *Pan paniscus*, *Pan troglodytes*,

Pongo pygmaeus, Pan troglodytes, Macaca fascicularis and Gorilla gorilla. The results suggested an evolutionary link between the genes in these organism. These genes were closely related. Phylogenetic tree construction, Multiple sequence alignment, Identification of conserved sequences in these genes was carried out

The significant genes associated with obesity were selected from metabolic pathways. These genes were found to be closely related to obesity. The expression levels of these genes were found to be elevated in adrenal gland, aorta, atrium auricular region, liver, pancreas and skeletal muscle tissue. This indicated a close association with obesity. Obesity related genes have high levels of expression in various human tissues such as adipose tissue, heart, pancreas, adrenal gland and thyroid gland. (Trang et al., 2023). There is over expression of obesity genes in adipose tissue. The genes were found to have less expression levels in amygdala, lung, eyes, skin.

Adipose tissue (both subcutaneous and visceral) plays a central role in obesity. The expression levels of these genes were very high in adipose tissue. The pancreas is involved in insulin regulation and glucose metabolism. It shows significant enrichment of obesity-related genes. The pituitary gland, which produces various hormones, including growth hormone, is another tissue with strong enrichment of obesity-related genes. The atrium auricular region is part of the heart, specifically the atria. Obesity is closely linked to cardiovascular risk factors such as hypertension, atherosclerosis, and atrial fibrillation (AF) 1. AF, a common cardiac arrhythmia, is associated with obesity.

Regarding adaptation to obesity, Skeletal muscle responds to changes in energy balance and nutrient availability. Gene expression alterations allow it to adapt to the demands of obesity, such as increased lipid

storage and insulin resistance Inflammation and Immune Response. Skeletal muscle expresses genes related to inflammation, which may contribute to insulin resistance and metabolic dysfunction. In summary, the high expression of obesity-related genes in skeletal muscle reflects its dynamic role in metabolism, inflammation, and adaptation to obesity-related challenges.

The NCBI BLAST results helped us identify the human gene and the closely related orthologs. That human and orthologous **primates** had the same. In Phylogeny.fr the phylogenetic trees were constructed. The trees had high bootstrap values indicating strong support for the branches. In EARS tool, the peaks were present above the cutoff value suggesting strong evolutionary significance and indicating the presence of conserved non coding sequences. Identification of motifs was done.

EARS tool helped us identify regulatory sequences, enhancers, silencers, promoters. Conserved regions functional importance phylogenetic footprint traces of phylogenetic history.

Sequence alignment is a fundamental step in phylogenetic analysis. By aligning protein sequences, we can identify regions of similarity or conservation across different species. Sequence alignment reveals conserved regions that often correspond to functional domains. Uniprot align tool results indicated strong multiple sequence alignment results. The results indicated high percentage of protein sequence similarity. Overall, the results indicated strong conservation.

In humans, the conserved sequences found in Homo sapiens can also be found in other organisms such as Gorilla sp., Pan paniscus, Pan troglodytes, Pongo abelii, Pongo pygmaeus, and Macaca fascicularis indicating a close evolutionary relationship between

humans and these primates. Conserved sequences indicate the presence of genes that have remained unchanged over evolutionary time. In the context of cross-species conservation, such persistence suggests that mutations within highly conserved regions would lead to non-viable life forms or rapid elimination through natural selection.

In uniprot align results high percentage identity between genes and closely related orthologs indicates significant conservation and similarity across species. This factor plays a crucial role in protein function. Percentage identity matrix was obtained. It is a numerical score calculated for each pair of aligned sequences. Identical residues (matches) are relative to the length of alignment. High percentage suggests strong conservation.

In this study we delved into the intricate evolutionary relationships among human genes and orthologs. By employing robust phylogenetic methods, we unearthed common ancestry, divergent evolution and conserved motifs across species. The findings shed light on the genetic underpinnings of obesity emphasizing the need for targeted interventions and personalised approaches.

In subsequent investigations, functional validation of the identified genes remains a critical avenue for research. While this study has primarily centered on the identification of candidate genes and orthologs linked to obesity, future endeavours should prioritize functional validation. Conducting in vitro and in vivo experiments will allow us to elucidate the precise roles of these genes, thereby enhancing our understanding. Additionally, comprehensive analysis of transcriptomics, metabolomics and proteomics data can provide further insights.

CHAPTER 7: REFERENCES

1. Ameer, F., Scandiuizzi, L., Hasnain, S., Kalbacher, H., & Zaidi, N. (2014). De novo lipogenesis in health and disease. *Metabolism*, 63(7), 895–902. <https://doi.org/10.1016/j.metabol.2014.04.003>
2. Baessler, A., Hasinoff, M. J., Fischer, M., Reinhard, W., Sonnenberg, G. E., Olivier, M., Erdmann, J., Schunkert, H., Doering, A., Jacob, H. J., Comuzzie, A. G., Kissebah, A. H., & Kwitek, A. E. (2005). Genetic linkage and association of the growth hormone secretagogue receptor (ghrelin receptor) gene in human obesity. *Diabetes*, 54(1), 259–267. <https://doi.org/10.2337/diabetes.54.1.259>
3. Beutler, L. R., Corpuz, T. V., Ahn, J. S., Kosar, S., Song, W., Chen, Y., & Knight, Z. A. (2020). Obesity causes selective and long-lasting desensitization of AgRP neurons to dietary fat. *eLife*, 9, e55909, 1–17. <https://doi.org/10.7554/eLife.55909>
4. Brown, O. I., Drozd, M., McGowan, H., Giannoudi, M., Conning-Rowland, M., Gierula, J., Straw, S., Wheatcroft, S. B., Bridge, K., Roberts, L. D., Levelt, E., Ajjan, R., Griffin, K. J., Bailey, M. A., Kearney, M. T., & Cubbon, R. M. (2023). Relationship Among Diabetes, Obesity, and Cardiovascular Disease Phenotypes: A UK Biobank Cohort Study. *Diabetes Care*, 46(8), 1531–1540.
5. Boucher, J., Masri, B., Daviaud, D., Gesta, S., Guigné, C., Mazzucotelli, A., CastanLaurell, I., Tack, I., Knibiehler, B., Carpené, C., Audigier, Y., Saulnier-Blache, J.-S., & Valet, P. (2005). Apelin, a Newly Identified Adipokine Up-Regulated by Insulin and Obesity. *Endocrinology*, 146(4), 1764–1771. <https://doi.org/10.1210/en.2004-1427>
6. Deng, B., Kong, W., Shen, X., et al. (2024). The role of DGAT1 and DGAT2 in regulating tumor cell growth and their potential clinical implications. *Journal of Translational Medicine*, 22, 290, 1–9. <https://doi.org/10.1186/s12967-024-05084-z>
7. Deng, M., Liao, C. Q., Chen, Q., Huang, G. H., & Wang, X. (2022). Phylogenetic relationships among Bombycinae (Lepidoptera, Bombycoidea, and Bombycidae) based on mitochondrial genomes. *Archives of Insect Biochemistry and Physiology*, 111(1), e21889, 1–12. <https://doi.org/10.1002/arch.21889>
8. Dhanoa, J. K., Sethi, R. S., Verma, R., et al. (2018). Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. *Journal of Animal Science and Technology*, 60, 25, 1–12. <https://doi.org/10.1186/s40781-018-0183-7>
9. Didelot, X. (2023). Phylogenetic Analysis of Bacterial Pathogen Genomes. *Methods in Molecular Biology*, 2674, 87–99. https://doi.org/10.1007/978-1-0716-3243-7_6
10. Domán, M., Makrai, L., & Bányai, K. (2022). Molecular Phylogenetic Analysis of *Candida krusei*. *Mycopathologia*, 187(4), 333–343. <https://doi.org/10.1007/s11046-022-00640-x>
11. Eissing, L., Scherer, T., Tödter, K., Knippschild, U., Greve, J. W., Buurman, W. A., Pinnschmidt, H. O., Rensen, S. S., Wolf, A. M., Bartelt, A., Heeren, J., Buettner, C., & Scheja, L. (2013). De novo

- lipogenesis in human fat and liver is linked to ChREBP- β and metabolic health. *Nature Communications*, 4, 1528. <https://doi.org/10.1038/ncomms2537>
12. Fang, H., Li, H., Zhang, H., Wang, S., Xu, S., Chang, L., Yang, Y., & Cui, R. (2022). Short-chain L-3-hydroxyacyl-CoA dehydrogenase: A novel vital oncogene or tumor suppressor gene in cancers. *Frontiers in Pharmacology*, 13, 1019312. <https://doi.org/10.3389/fphar.2022.1019312>
 13. Fang, J. (2023). Predicting thermostability difference between cellular protein orthologs. *Bioinformatics*, 39(8), 1234–1245. <https://doi.org/10.1093/bioinformatics/btad504>
 14. Gong, L., Li, R., Ren, W., Wang, Z., Wang, Z., Yang, M., & Zhang, S. (2017). The FOXO1 Gene-Obesity Interaction Increases the Risk of Type 2 Diabetes Mellitus in a Chinese Han Population. *Journal of Korean Medical Science*, 32(2), 264–271. <https://doi.org/10.3346/jkms.2017.32.2.264>
 15. Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K., Yates, A. D. (2024). Ensembl 2024. *Nucleic Acids Research*, 52(D1), D891–D899. <https://doi.org/10.1093/nar/gkad1049>
 16. Jia, W., & Lieu, F. (2021). Obesity: causes, consequences, treatments, and challenges. *Journal of Molecular Cell Biology*, 13(7), 463–465. <https://doi.org/10.1093/jmcb/mjab056>
 17. Kang, H. J., Min, B. K., Choi, W. I., Jeon, J. H., Kim, D. W., Park, S., Lee, Y. K., Kim, H. J., Byeon, J. E., Go, Y., Ham, H. J., Jeon, Y. H., Kim, M. J., Lee, J. Y., Wende, A. R., Choi, S. H., Harris, R. A., & Lee, I. K. (2021). Pyruvate dehydrogenase kinase 1 and 2 deficiency reduces high-fat diet-induced hypertrophic obesity and inhibits the differentiation of preadipocytes into mature adipocytes. *Experimental & Molecular Medicine*, 53(9), 1390–1401. <https://doi.org/10.1038/s12276-021-00672-1>
 18. Khwairakpam, A. D., Banik, K., Girisa, S., Shabnam, B., Shakibaei, M., Fan, L., Arfuso, F., Monisha, J., Wang, H., Mao, X., Sethi, G., & Kunnumakkara, A. B. (2020). The vital role of ATP citrate lyase in chronic diseases. *Journal of Molecular Medicine*, 98(1), 71–95. <https://doi.org/10.1007/s00109-019-01863-0>
 19. Lago-Barcia, D., Álvarez-Presas, M., Riutort, M., Oceguera-Figueroa, A., & Carbayo, F. (2023). Phylogenetic relationships of the Geoplaninae land planarians (Platyhelminthes, Tricladida) assessed with a total evidence approach, with the description of a new species of *Gigantea*. *Molecular Phylogenetics and Evolution*, 184, 107750. <https://doi.org/10.1016/j.ympev.2023.107750>
 20. Laraba, I., Busman, M., Geiser, D. M., & O'Donnell, K. (2022). Phylogenetic Diversity and Mycotoxin Potential of Emergent Phytopathogens Within the *Fusarium tricinctum* Species Complex. *Phytopathology*, 112(6), 1284–1298. <https://doi.org/10.1094/PHYTO-09-21-0394-R>
 21. Liang, K. (2023). Mitochondrial CPT1A: Insights into structure, function, and basis for drug development. *Frontiers in Pharmacology*, 14, 1160440. <https://doi.org/10.3389/fphar.2023.1160440>
 22. Makris, M. C., Alexandrou, A., Papatsoutsos, E. G., Malietzis, G., Tsilimigras, D. I., Guerron, A. D., & Moris, D. (2017). Ghrelin and Obesity: Identifying Gaps and Dispelling Myths. A Reappraisal. *In Vivo*, 31(6), 1047–1050. <https://doi.org/10.21873/invivo.11168>

23. Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., Davidson, C., Flicek, P. (2023). Ensembl 2023. *Nucleic Acids Research*, 51(D1), D933–D941. <https://doi.org/10.1093/nar/gkac958>
24. Merritt, J. L. 2nd, & Chang, I. J. (2019). Medium-chain Acyl-Coenzyme A Dehydrogenase Deficiency. In M.P. Adam, J. Feldman, G.M. Mirzaa, et al. (Eds.), *GeneReviews®*, 1-30. Seattle (WA): University of Washington, Seattle. <https://www.ncbi.nlm.nih.gov/books/NBK1424/>
25. Ang, M. Y., Takeuchi, F., & Kato, N. (2024). Deciphering the genetic landscape of obesity: a data-driven approach to identifying plausible causal genes and therapeutic targets. **Journal of Molecular Genetics**, 18(3), 245-260.
26. Mosaoa, R., Kasprzyk-Pawełec, A., Fernandez, H. R., & Avantaggiati, M. L. (2021). The Mitochondrial Citrate Carrier SLC25A1/CIC and the Fundamental Role of Citrate in Cancer, Inflammation and Beyond. *Biomolecules*, 11(2), 141. <https://doi.org/10.3390/biom11020141>
27. Nowzari, Z., Masoumi, M., Nazari-Robati, M., Akbari, H., & Shahrokhi, N. (2018). Association of polymorphisms of leptin, leptin receptor and apelin receptor genes with susceptibility to coronary artery disease and hypertension. *Life Sciences*, 207, 166– 171. <https://doi.org/10.1016/j.lfs.2018.06.007>
28. Brown, O. I., Drozd, M., McGowan, H., Giannoudi, M., Conning-Rowland, M., Gierula, J., Straw, S., Wheatcroft, S. B., Bridge, K., Roberts, L. D., Levelt, E., Ajjan, R., Griffin, K. J., Bailey, M. A., Kearney, M. T., & Cubbon, R. M. (2023). Relationship among diabetes, obesity, and cardiovascular disease phenotypes: A UK Biobank cohort study. *Diabetes Care*, 46(8), 1531–1540.
29. Pelosi, B. (2022). Developing a bioinformatics pipeline for comparative protein classification analysis. *BMC Genomic Data*, 23(1), 43. <https://doi.org/10.1186/s12863-022-01045-x>
30. Pelosi, B. (2022). Developing a bioinformatics pipeline for comparative protein classification analysis. *BMC Genomic Data*, 23(1), 43. <https://doi.org/10.1186/s12863-022-01045-x>
31. Pundir, S., Martin, M. J., O'Donovan, C., & UniProt Consortium (2016). UniProt Tools. *Current Protocols in Bioinformatics*, 53, 1.29.1–1.29.15. <https://doi.org/10.1002/0471250953.bi0129s53>
32. Ren, T., Xie, D., Peng, C., Gui, L., Price, M., Zhou, S., & He, X. (2022). Molecular evolution and phylogenetic relationships of Ligusticum (Apiaceae) inferred from the whole plastome sequences. *BMC Ecology and Evolution*, 22(1), 55. <https://doi.org/10.1186/s12862-022-02010-z>
33. Salih J. Wakil, Lutfi A. Abu-Elheiga. (2009). Fatty acid metabolism: target for metabolic syndrome. *Journal of Lipid Research*, 50(Supplement), S138–S143. <https://doi.org/10.1194/jlr.R800079-JLR200>
34. Saeed, S., Bonnefond, A., Tamanini, F., Mirza, M. U., Manzoor, J., Janjua, Q. M., Din, S. M., Gaitan, J., Milochau, A., Durand, E., Vaillant, E., Haseeb, A., De Graeve, F., Rabearivelo, I., Sand, O., Queniat, G., Boutry, R., Schott, D. A., Ayesha, H., Ali, M., Froguel, P. (2018). Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nature Genetics*, 50(2), 175–179. <https://doi.org/10.1038/s41588-017-0023-6>
35. Shi, Y., Lin, W., Chu, J., Wang, G., Zhao, P., Huang, G. H., & Hou, D. (2022). Comprehensive identification of protein orthologs in the family Ascoviridae facilitates an understanding of

- phylogenomics, protein conservation, and phosphorylation. *Archives of Virology*, 167(4), 1075–1087. <https://doi.org/10.1007/s00705-022-054020>
36. Tang, H. N., Xiao, F., Chen, Y. R., Zhuang, S. Q., Guo, Y., Wu, H. X., & Zhou, H. D. (2020). Higher Serum Neuropeptide Y Levels Are Associated with Metabolically Unhealthy Obesity in Obese Chinese Adults: A Cross-Sectional Study. *Mediators of Inflammation*, 2020, 7903140. <https://doi.org/10.1155/2020/7903140>
37. Trang, K., Grant, S. F. (2023). Genetics and epigenetics in the obesity phenotyping scenario. *Reviews in Endocrine & Metabolic Disorders*, 24, 775–793. <https://doi.org/10.1007/s11154-023-09804-6>
38. Warensjö, E., Ingelsson, E., Lundmark, P., Lannfelt, L., Syvänen, A. C., Vessby, B., & Risérus, U. (2007). Polymorphisms in the SCD1 gene: Associations with body fat distribution and insulin sensitivity. *Obesity*, 15(7), 1732–1740. <https://doi.org/10.1038/oby.2007.206>
39. Wu, G., Li, M. X., Horak, E., & Yang, Z. L. (2022). Phylogenetic analysis reveals the new genus *Amoenoboletus* from Asia and New Zealand. *Mycologia*, 114(1), 144–156. <https://doi.org/10.1080/00275514.2021.1971450>
40. Wysocka, M. B., Pietraszek-Gremplewicz, K., & Nowak, D. (2018). The Role of Apelin in Cardiovascular Diseases, Obesity and Cancer. *Frontiers in Physiology*, 9, 557. <https://doi.org/10.3389/fphys.2018.00557>
41. Zaru, R., Orchard, S., & UniProt Consortium (2023). UniProt Tools: BLAST, Align, Peptide Search, and ID Mapping. *Current Protocols*, 3(3), e697. <https://doi.org/10.1002/cpz1.697>
42. Zhang, J., Huang, K., Pan, Z., Wu, Y., Wu, W., & Zhou, X. (2024). Short-chain L3-hydroxyacyl-CoA dehydrogenase: A novel vital oncogene or tumor suppressor gene in cancers. *Frontiers in Pharmacology*, 13, 1019312. <https://doi.org/10.3389/fphar.2022.1019312>
43. Shi, Y., Lin, W., Chu, J., Wang, G., Zhao, P., Huang, G. H., & Hou, D. (2022). Comprehensive identification of protein orthologs in the family Ascoviridae facilitates an understanding of phylogenomics, protein conservation, and phosphorylation. *Archives of Virology*, 167(4), 1075–1087. <https://doi.org/10.1007/s00705-022-054020>

APPENDIX I

Table 1 BLAST ANALYSIS OF ACADM gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
1	Homo sapiens acyl-CoA dehydrogenase medium chain(ACADM),transcript variant 1,mRNA;nuclear gene for mitochondrial product	Homo sapiens	4176	4176	100%	0	100	2261	NM_000016.6
2.	PREDICTED:Pan paniscus acyl-CoA dehydrogenase medium chain(ACADM),transcript variant x1,mRNA	Pan paniscus	3786	3786	91%	0	99.57	2431	XM_003830572.4
3.	PREDICTED:Gorilla Gorilla acyl-CoA dehydrogenase medium chain(ACADM),transcript variant x1,mRNA	Gorilla Gorilla gorilla	3784	3784	91%	0	99.52	2438	XM_019019925.2
4.	PREDICTED:Pongo abelii acyl-CoA dehydrogenase medium chain(ACADM),transcript variant,x4,mRNA	Pongo abelii	3760	3760	97%	0	97.51	2500	XM_009248888.3
5.	PREDICTED:Pongo pygmaeus acyl-CoA dehydrogenase medium chain(ACADM),transcript variant X2,mRNA	Pongo pygmaeus	3585	3585	91%	0	97.88	2340	XM_054494536.1
6	PREDICTED: Macaca fascicularis acyl-CoA dehydrogenase medium chain(ACADM),transcript variant X1,mRNA	Macaca fascicularis	3278	3278	89%	0	95.88	2027	XM_005542992.3
7.	PREDICTED:Pan troglodytes acyl CoA dehydrogenase medium chain(ACADM),transcript variant X1,mRNA	Pan troglodytes	3254	3786	91%	0	99.77	3316	XM_054664818.1

Table 2 BLAST analysis of CPT1A gene and orthologs

S no	Description	Scientific Name	Max Score	Total Score	Query cover	E value	Per. ident	Acc. Len	Accession

1.	Homo sapiens carnitine palmitoyltransferase 1a(CPT1A),transcript variant X1,mRNA	Homo sapiens	9673	9673	100%	0	100	5238	NM_001876.4
2.	PREDICTED:Pan troglodytes carnitine palmitoyl transferase 1a(CPT1A), transcript variant X1,mRNA	Pan troglodytes	9376	9376	100%	0	98.97	5241	XM_016921421.3
3.	PREDICTED:Pan paniscus carnitine palmitoyltransferase 1a(CPT1A),transcript variant X1,mRNA	Pan paniscus	9367	9367	100%	0	98.93	5298	XM_034933223.2
4.	PREDICTED:Gorilla gorilla carnitine palmitoyltransferase 1a(CPT1A), transcript variant X1,mRNA	Gorilla gorilla	9299	9299	100%	0	98.68	5254	XM_031015971.2
5.	PREDICTED:Pongo pygmaeus carnitine palmitoyltransferase 1a(CPT1A),mRNA	Pongo pygmaeus	8654	8936	100%	0	96.48	5244	XM_054439228.1
6.	PREDICTED:Pongo abelii carnitine palmitoyltransferase 1a(CPT1A),mRNA	Pongo abelii	8600	8879	99%	0	96.4	5226	XM_024255747.2
7.	PREDICTED:Macaca fascicularius carnitine palmitoyltransferase 1a(CPT1A),transcript variant X1,mRNA	Macaca fascicularius	7653	7935	100%	0	92.97	5289	XM_045372278.1

Table 3 BLAST analysis of DGAT1 gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens diacylglycerol o-acyltransferase 1 (DGAT 1) mRNA	Homo sapiens	6746	6746	100%	0	100	3653	NM_012079.6
2.	PREDICTED:Pan troglodytes	Pan troglodyte	4739	6182	97%	0	97.56	4775	XM_016960109.3

	diacylglycerol o-acyltransferase 1 (DGAT 1), transcript variant X2, mRNA								
3.	PREDICTED: Pan Paniscus diacylglycerol o-acyltransferase 1 (DGAT 1), mRNA	Pan Paniscus	3561	3561	53%	0	99.69	5497	XM_003819429.4
4.	PREDICTED: Gorilla gorilla gorilla diacylglycerol o-acyltransferase 1 (DGAT 1), transcript variant X1, mRNA	Gorilla gorilla gorilla	3504	3504	53%	0	99.18	2003	XM_019032174.3
5.	PREDICTED: Pongo pygmaeus diacylglycerol o-acyltransferase 1 (DGAT 1), mRNA	Pongo pygmaeus	3395	3395	53%	0	98.15	1975	XM_054499286.1
6.	PREDICTED: Pongo abelli diacylglycerol o-acyltransferase 1 (DGAT 1), mRNA	Pongo abelli	3384	3384	53%	0	98.05	2012	XM_024251032.2
7.	PREDICTED: Macaca fascicularis diacylglycerol o-acyltransferase 1 (DGAT 1), transcript variant X1, mRNA	Macaca fascicularis	3061	3061	53%	0	95.08	1945	XM_005564342.3

Table 4 BLAST analysis of DGAT1 gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens solute carrier	Homo sapiens	2859	2859	100%	0	100	1548	NM_005984.5

	family 25 member 1 (SLC 25A1), transcript variant 1, mRNA; nuclear gene from mitochondrial product								
2.	PREDICTED: Gorilla gorilla gorilla solute carrier family 25 member 1 (SLC 25A1) transcript variant X2,mRNA	Gorilla gorilla gorilla	2586	2586	100%	0	96.58	2806	XM_004063014.3
3.	PREDICTED: Pan Troglodytes solute carrier family 25 member 1 (SLC 25A1) transcript variant X2,mRNA	Pan troglodytes	2265	2825	100%	0	99.44	1661	XM_001165281.6
4.	PREDICTED: Pan paniscus solute carrier family 25 member 1 (SLC 25A1) transcript variant X2,mRNA	Pan paniscus	2259	2809	100%	0	99.36	1985	XM_003814041.5
5.	PREDICTED: Pongo abellii solute carrier family 25 member 1 (SLC 25A1) transcript variant X2,mRNA	Pongo abellii	2158	2716	99%	0	97.91	1649	XM_024239975.2
6.	PREDICTED: Pongo pygmaeus solute carrier	Pongo pygmaeus	2154	2713	100%	0	97.84	1673	XM_054469293.1

	family 25 member 1 (SLC 25A1) transcript variant X2,mRNA								
7.	PREDICTED: Macaca fascicularis solute carrier family 25 member 1 (SLC 25A1) transcript variant X1,mRNA	Macaca fascicularis	2036	2572	100%	0	96.23	1576	XM_005567997.3

Table 5 BLAST analysis of HADH gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens hydroxyacyl-CoA dehydrogenase(HADH),transcript variant 2,mRNA;nuclear gene for mitochondrial product	Homo sapiens	3330	3330	100%	0	100	1803	NM_005327.7
2.	PREDICTED: Pan paniscus hydroxyacyl CoA dehydrogenase (HADH), mRNA	Pan paniscus	3275	3275	100%	0	99.45	1889	XM_003829995.4
3.	PREDICTED: Pan troglodytes hydroxyacyl CoA dehydrogenase (HADH),transcript variant X1, mRNA	Pan troglodytes	3264	3264	100%	0	99.33	1899	XM_001135650.5
4.	PREDICTED: Gorilla gorilla gorilla hydroxyacyl CoA dehydrogenase (HADH),transcript variant X1, mRNA	Gorilla gorilla gorilla	3214	3214	99%	0	98.89	1916	XM_004040251.4
5.	PREDICTED: Pongo pygmaeus hydroxyacyl CoA dehydrogenase (HADH), mRNA	Pongo pygmaeus	3107	3107	100%	0	97.78	2087	XM_054486662.1

6.	PREDICTED: Pongo abelii hydroxyacyl CoA dehydrogenase (HADH),transcript variant X1, mRNA	Pongo abelii	3096	3096	99%	0	97.78	1913	XM_009240245.3
7.	PREDICTED: Macaca fascicularis hydroxyacyl CoA dehydrogenase (HADH), mRNA	Macaca fascicularis	2824	2824	99%	0	95.06	1838	XM_045392727.1

Table 6 BLAST analysis of PKM gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens Pyruvate kinase M1/2(PKM),transcript variant 1,mRNA	Homo sapiens	4257	4257	100%	0	100	2305	NM_002654.6
2.	PREDICTED:Pan paniscus Pyruvate kinase M1/2(PKM),transcript variant X12,mRNA	Pan paniscus	4218	4218	100%	0	99.7	2387	XM_057299900.1
3.	PREDICTED:Pan troglodytes Pyruvate kinase M1/2(PKM),transcript variant X12,mRNA	Pan troglodytes	4204	4204	99%	0	99.61	2374	XM_001175100.5
4.	PREDICTED:Gorilla gorilla gorilla Pyruvate kinase M1/2(PKM),transcript variant X6,mRNA	Gorilla gorilla gorilla	4159	4159	99%	0	99.26	2402	XM_004056435.4
5.	PREDICTED:Pongo pygmaeus Pyruvate kinase M1/2(PKM),transcript variant X5,mRNA	Pongo pygmaeus	4076	4076	99%	0	98.61	2378	XM_054451495.1
6.	PREDICTED:Macaca fascicularis Pyruvate kinase M1/2(PKM),transcript variant X5,mRNA	Macaca fascicularis	3814	3814	99%	0	96.47	2320	XM_005559965.3
7.	Pongo abelii Pyruvate kinase M1/2(PKM),mRNA	Pongo abelii	3744	3744	99%	0	96.05	2411	NM_001133611.1

Table 7 BLAST analysis of ACLY gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
------	-------------	-----------------	-----------	-------------	-------------	---------	-----------	---------	-----------

1.	Homo sapiens ATP citrate lyase(ACLY),transcript variant 1,mRNA	Homo sapiens	7928	7928	100%	0	100	4293	NM_001096.3
2.	PREDICTED:Pan paniscus ATP citrate lyase(ACLY),transcript variant 3,mRNA	Pan paniscus	7812	7812	100%	0	95.51	4408	XM_034941855.2
3.	PREDICTED:Gorilla gorilla gorilla ATP citrate lyase(ACLY),transcript variant 3,mRNA	Gorilla gorilla gorilla	7810	7810	100%	0	99.51	4447	XM_019027100.3
4.	PREDICTED:Pan troglodytes ATP citrate lyase(ACLY),transcript variant 3,mRNA	Pan troglodytes	7777	7777	100%	0	99.37	4414	XM_024350305.2
5.	PREDICTED:Pongo abelii ATP citrate lyase(ACLY),transcript variant 3,mRNA	Pongo abelii	7590	7590	100%	0	99.6	4372	XM_024234490.2
6.	PREDICTED:Pongo pygmaeus ATP citrate lyase(ACLY),transcript variant X3,mRNA	Pongo pygmaeus	7583	7583	99%	0	98.6	4363	XM_054457539.1
7.	PREDICTED:Macaca fascicularis ATP citrate lyase(ACLY),transcript variant X1,mRNA	Macaca fascicularis	7317	7317	100%	0	97.46	4286	XM_005584204.3

Table 8 BLAST analysis of ACACA gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens acetyl-CoA carboxylase alpha(ACACA),mRNA	Homo sapiens	18491	18491	100%	0	100	10013	NM_198834.3
2.	PREDICTED:Pan troglodytes acetyl-CoA carboxylase alpha(ACACA),mRNA	Pan troglodytes	18114	18114	100%	0	99.33	10008	XM_511428.8
3.	PREDICTED:Pan paniscus acetyl-CoA carboxylase alpha(ACACA),mRNA	Pan paniscus	18061	18061	100%	0	99.24	10009	XM_003817948.4
4.	PREDICTED:Pongo abelii acetyl-CoA	Pongo abelii	17590	17590	100%	0	98.4	9998	XM_0242234623.2

	carboxylase alpha(ACACA),mRNA								
5.	PREDICTED:Pongo pygmaeus acetyl-CoA carboxylase alpha(ACACA),mRNA	Pongo pygmaeus	17579	17579	100%	0	98.38	9998	XM_054456897.1
6.	PREDICTED: Gorilla gorilla gorilla acetyl- CoA carboxylase alpha(ACACA),mRNA	Gorilla gorilla gorilla	17248	17248	95%	0	99.24	9561	XM_031011317.2
7.	PREDICTED: Macaca fascicularis acetyl-CoA carboxylase alpha(ACACA),mRNA	Macaca fascicularis	16903	16903	100%	0	97.15	10022	XM_005583932.3

Table 9 BLAST analysis of MLXIPL gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens MLX interacting protein like(MLXIPL),transcript variant 1,mRNA	Homo sapiens	6006	6006	100%	0	100	3252	NM_032951.3
2.	PREDICTED:Pan paniscus MLX interacting protein like(MLXIPL),transcript variant X1,mRNA	Pan paniscus	5847	5847	99%	0	99.14	3280	XM_003815573.3
3.	PREDICTED:Pan troglodytes MLX interacting protein like(MLXIPL),transcript variant X3,mRNA	Pan troglodytes	5836	5836	99%	0	99.08	3288	XM_003318668.5
4.	PREDICTED:Gorilla gorilla gorilla MLX interacting protein like(MLXIPL),transcript variant X6,mRNA	Gorilla gorilla gorilla	5795	5795	99%	0	98.86	3288	XM_031012705.2
5.	PREDICTED:Pongo abelii MLX interacting protein like(MLXIPL),transcript variant X2,mRNA	Pongo abelii	5651	5651	99%	0	98.09	3243	XM_054560415.1
6.	PREDICTED:Pongo pygmaeus MLX interacting protein like(MLXIPL),transcript variant X5,mRNA	Pongo pygmaeus	5648	5648	99%	0	98.06	3283	XM_054493797.1

7.	PREDICTED:Macaca fascicularis MLX interacting protein like(MLXIPL),transcript variant X9,mRNA	Macaca fascicularis	5387	5387	99%	0	96.62	3272	XM_045389686.1
----	---	---------------------	------	------	-----	---	-------	------	----------------

Table 10 BLAST analysis of SCD gene and orthologs

Sno.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per.Ident	Acc.Len	Accession
1.	Homo sapiens stearoyl-CoA desaturase (SCD), mRNA	Homo sapiens	9686	9686	100%	0	100	5245	NM_005063.5
2.	PREDICTED: Gorilla gorilla gorilla stearoyl-CoA desaturase (SCD), mRNA	Gorilla gorilla gorilla	9380	9380	100%	0	98.93	6018	XM_004049945.3
3.	PREDICTED: Pan paniscus stearoyl-CoA desaturase (SCD), mRNA	Pan paniscus	9164	9164	100%	0	98.25	5886	XM_00384618.4
4.	PREDICTED: Pan troglodytes stearoyl-CoA desaturase (SCD), mRNA	Pan troglodytes	9138	9138	100%	0	98.17	5685	XM_001168563.6
5.	PREDICTED: Pongo pygmaeus stearoyl-CoA desaturase (SCD), mRNA	Pongo pygmaeus	8940	8940	100%	0	97.47	6025	XM_054436912.1
6.	PREDICTED: Macaca fascicularis stearoyl-CoA desaturase	Macaca fascicularis	7589	7589	94%	0	94.34	5291	XM_015456547.2

	(SCD), transcript variant, X1, mRNA								
7.	Pongo abelli stearoyl-CoA desaturase (SCD),mRNA	Pongo abelli	6894	6894	77%	0	97.48	4179	NM_001132259.1

Table 11 BLAST analysis of GHSR gene and orthologs

Sl no.	Description	Scientific Name	Max score	Total score	Query cover	E value	Per ident	Acc len	Accession
1.	Homo sapiens Growth hormone secretagogue receptor(GHSR),transcript variant,1a,mRNA	Homo sapiens	5566	5566	100%	0	100	3014	NM_198407.2
2.	PREDICTED:Gorilla gorilla gorilla Growth hormone secretagogue receptor(GHSR),transcript variant,mRNA	Gorilla gorilla gorilla	5360	5360	100%	0	98.77	3013	XM_004038010.4
3.	PREDICTED: Pan paniscus Growth hormone secretagogue receptor(GHSR),transcript variant,mRNA	Pan paniscus	5319	5319	100%	0	98.54	6648	XM_003831941.3
4.	PREDICTED: Pan troglodytes Growth hormone secretagogue receptor(GHSR),transcript variant,mRNA	Pan troglodytes	5299	5299	100%	0	98.44	5968	XM_054681248.1
5.	PREDICTED: Pongo abelii Growth hormone secretagogue receptor(GHSR),transcript variant,mRNA	Pongo abelii	5221	5221	100%	0	97.94	3188	XM_002814288.4
6.	PREDICTED:Pongo pygmaeus Growth hormone secretagogue receptor(GHSR), transcript variant,mRNA	Pongo pygmaeus	5221	5221	100%	0	97.94	3013	XM_054483135.1
7.	PREDICTED:Macaca fascicularis Growth hormone secretagogue receptor(GHSR), transcript variant,mRNA	Macaca fascicularis	4726	4726	99	0	94.95	5392	XM_005546393.3

Table 12 BLAST analysis of NPY genes and orthologs

Sl no.	Description	Scientific Name	Max Scire	Total Score	Query Cover	E value	Per ident	Acc. Len	Accession
1.	Homo sapiens neuropeptide Y(NPY),mRNA	Homo sapiens	1026	1026	100%	0	100	555	NM_000905.4
2.	PREDICTED: Gorilla gorilla gorilla neuropeptide Y(NPY),mRNA	Gorilla gorilla gorilla	998	998	100%	0	99.1	617	XM_004045187.3
3.	PREDICTED:Pan paniscus neuropeptide Y(NPY),mRNA	Pan paniscus	998	998	100%	0	99.1	631	XM_003807930.4
4.	PREDICTED:Pan troglodytes neuropeptide Y(NPY),mRNA	Pan troglodytes	998	998	100%	0	99.1	603	XM_001159094.5
5.	PREDICTED:Pongo pygmaeus neuropeptide Y(NPY),mRNA	Pongo pygmaeus	970	970	100%	0	98.2	567	XM_054494825.1
6.	PREDICTED:Pongo abelii neuropeptide Y(NPY),mRNA	Pongo pygmaeus	965	965	100%	0	98.02	591	XM_002818138.5
7.	PREDICTED:Macaca fascicularius neuropeptide(NPY),mRNA	Macaca fascicularius	904	904	100%	0	96.04	555	XM_005549975.3

Table 13 BLAST analysis of AGRP gene and orthologs

Sl no.	Description	Scientific Name	Max Scire	Total Score	Query Cover	E value	Per identt	Acc. Len	Accession
1.	Homo sapiens agouti related neuro peptide (AGRP), mRNA	Homo sapiens	920	920	100%	0	100	498	NM_001138.2
2.	PREDICTED: Gorilla gorilla gorilla agouti related neuro peptide (AGRP),mRNA	Gorilla gorilla gorilla	893	893	100%	0	99	764	XM_004057817.2
3.	PREDICTED: Pan troglodytes agouti related neuropeptide (AGRP)mRNA	Pan troglodytes	861	921	100%	0	99.79	1180	XM_054669427.1
4.	PREDICTED: Pan paniscus agouti related neuropeptide (AGRP)mRNA	Pan paniscus	856	916	100%	0	99.57	1178	XM_055100336.2

5.	PREDICTED: Pongo abelii agouti related neuropeptide (AGRP)mRNA	Pongo abelii	854	854	100%	0	97.59	1259	XM_002826540.3
6.	PREDICTED: Pongo pygmaeus agouti related neuropeptide (AGRP)mRNA	Pongo pygmaeus	854	854	100%	0	97.59	498	XM_054453945.1
7.	PREDICTED: Macaca fascicularis agouti related neuropeptide (AGRP)mRNA	Macaca fascicularis	809	809	100%	0	95.98	500	XM_005592267.3

Table 14 BLAST analysis of FOXO1 gene and orthologs

Sl no	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident.	Acc. Lrn	Accession
1.	Homo sapiens forkhead box 01(FOXO1),mRNA	Homo sapiens	10672	10672	100%	0	100	5779	NM_002015.4
2.	PREDICTED:Pan paniscus forkhead box 01(FOXO1),mRNA	Pan paniscus	10430	10430	100%	0	99.26	13305	XM_024925280.4
3.	PREDICTED:Pongo pygmaeus forkhead box 01,mRNA	Pongo pygmaeus	9878	9878	98%	0	98.04	5695	XM_054446024.1
4.	PREDICTED:Pongo abelii forkhead box 01(FOXO1),mRNA	Pongo abelii	9860	9860	98%	0	97.87	5689	XM_002824200.5
5.	PREDICTED:Gorilla gorilla gorilla forkhead box 01(FOXO1),transcript variant X2,mRNA								
6.	PREDICTED:: Macaca fascicularis forkhead box 01(FOXO1),mRNA	Macaca fascicularis	5048	9351	98%	0	95.25	5752	XM_045376895.1

Table 15 BLAST analysis of ADCY3 gene and orthologs

Sl no.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
1.	Homo sapiens adenylate cyclase 3(ADCY3), transcript variant 2,mRNA	Homo sapiens	9141	9142	100%	0	100	4950	NM_0040365.5
2.	PREDICTED:Pan paniscus adenylate cyclase 3(ADCY3), transcript variant X2,mRNA	Pan paniscus	8997	8997	100%	0	99.47	4951	XM_008952400.4
3.	PREDICTED:Pan troglodytes adenylate cyclase 3(ADCY3), transcript variant X2,mRNA	Pan troglodytes	8975	8975	100%	0	99.39	5140	XM_016948151.2
4.	PREDICTED:Gorilla gorilla gorilla adenylate cyclase 3(ADCY3), transcript variant X8,mRNA	Gorilla gorilla gorilla	8865	8865	99%	0	99.13	4926	XM_004028928.4
5.	PREDICTED:Pongo abelii adenylate cyclase 3(ADCY3), transcript variant X4,mRNA	Pongo abelii	7489	8195	94%	0	98.12	4793	XM_024242251.2
6.	PREDICTED:Pongo pygmaeus adenylate cyclase 3(ADCY3),transcript variant X4,mRNA	Pongo pygmaeus	7254	7254	84%	0	98.13	4166	XM_054475220.1
7.	PREDICTED:Macaca fascicularis adenylate cyclase 3(ADCY3), transcript variant X4,mRNA	Macaca fascicularis	6800	7333	91				XM_005576419.3

Table 16 BLAST analysis of APLN gene and orthologs

Sl no.	Description	Scientific name	Max score	Total score	Query cover	E value	Per ident.	Acc. Len	Accession
1.	Homo sapiens apelin(APLN),mRNA	Homo sapiens	5954	5954	100%	0	100	3224	NM_017413.5
2.	PREDICTED:Pan paniscus apelin(APLN),transcript variant X2,mRNA	Pan paniscus	5509	5509	97%	0	98.44	3252	XM_024927383.2

3.	PREDICTED:Pan troglodytes apelin(APLN),mRNA	Pan troglodytes	5509	5509	97%	0	98.44	3252	NM_022927383.2
4.	PREDICTED :Pongo pygmaeus apelin(APLN),mRNA	Pongo pygmaeus	5328	5328	100%	0	96.54	3219	XM_054471173.3
5.	PREDICTED:Pongo abelii apelin(APLN),mRNA	Pongo abelii	5134	5134	95%	0	96.71	3098	XM_002832099.5
6.	PREDICTED:Macaca fascicularius apelin(APLN),mRNA	Macaca fascicularius	4050	4050	82%	0	94.03	2679	XM_005594539.3
7.	PREDICTED:Gorilla gorilla gorilla(APLN),mRNA	Gorilla gorilla gorilla	2726	5034	90%	0	97.56	2919	XM_019018831.2

Table 17 BLASTN analysis of APLNR gene and orthologs

SI	Description	Scientific Name	Max score	Total score	Query cover	E value	Per ident	Acc. Len	Accession
1.	Homo sapiens apelin receptor (APLNR),transcript variant 1, mRNA	Homo sapiens	6759	6759	100%	0	100	3660	NM_005161.6
2.	PREDICTED:Pan troglodytes apelin receptor(APLNR),mRNA	Pan troglodytes	6593	6593	100%	0	99.18	3863	XM_521984.7
3.	PREDICTED: Macaca fascicularius apelin receptor(APLNR),mRNA	Macaca fascicularius	5823	5823	100%	0	95.42	3654	XM_005577873.3
4.	PREDICTED:Pongo pygmaeus apelin receptor (APLNR),mRNA	Pongo pygmaeus	3866	4284	68%	0	97.57	2567	XM_054438351.1
5.	PREDICTED:Pan paniscus apelin receptor(APLNR),mRNA	Pan paniscus	3129	3129	47%	0	99.48	1882	XM_003826439.5
6.	PREDICTED: Gorilla gorilla gorilla apelin receptor(APLN),mRNA	Gorilla gorilla gorilla	2645	3119	47%	0	99.38	1882	XM_055355366.1
7.	Pongo abelii apelin receptor(APLNR),mRNA	Pongo abelii	2483	2907	45%	0	98.44	1690	NM_001131159.2

Table 18 BLAST analysis of MAPK1 gene and orthologs

SI no	Description	Scientific Name	Max score	Total score	Query cover	E value	Per ident	Acc. Len	Accession
1.	Homo sapiens mitogen-activated protein kinase	Homo sapiens	10861	10861	100%	0	100	5881	NM_002745.5

	1(MAPK1),transcript variant 1,mRNA								
2.	PREDICTED:Gorilla gorilla gorilla mitogen-activated protein kinase(MAPK1),transcript variant X1,mRNA	Gorilla gorilla gorilla	9310	9310	88%	0	98.82	5266	XM_004063106.4
3.	PREDICTED:Pongo pygmaeus mitogen-activated protein kinase 1(MAPK1),mRNA	Pongo pygmaeus	9105	9105	88%	0	98.11	5252	XM_054469062.1
4.	PREDICTED:Macaca fascicularis mitogen-activated protein kinase 1(MAPK1),mRNA	Macaca fascicularis	8702	8702	88%	0	96.75	5228	XM_005567863.3
5.	PREDICTED:Pan troglodytes mitogen-activated protein kinase 1(MAPK1),mRNA	Pan troglodytes	5419	5419	50%	0	99.76%	2991	XM_003317123.6
6.	PREDICTED:Pan paniscus mitogen-activated protein kinase 1(MAPK1),mRNA	Pan paniscus	5413	5413	50%	0	99.8	2950	XM_034948791.3
7.	PREDICTED:Pongo abelii mitogen-activated protein kinase 1(MAPK1),mRNA	Pongo abelii	5258	5258	50%	0	98.78	2988	XM_024239685.2

Table 19 BLAST analysis of KLF2 genes and orthologs

Sl. No.	Description	Scientific Name	Max Score	Total Score	Query cover	E value	Per. Ident	Acc Len.	Accession
1.	Homo sapiens KLF transcription factor 2(KLF2)	Homo sapiens	5208	5208	100%	0	100	2820	NM_016270.4
2.	PREDICTED : Pan troglodytes KLF transcription factor 2 (KLF2),mRNA	Pan troglodytes	3064	3064	59%	0	99.7	1886	XM_003316174.5
3.	PREDICTED: Gorilla gorilla gorilla	Gorilla gorilla	3016	3016	59%	0	99.28	1886	XM_031003387.2
4.	PREDICTED :Pongo abelii	Pongo abelii	2898	2898	59%	0	97.97	1688	XM_024237927.2
5.	PREDICTED:Macaca fascicularis Kruppel like factor(KLF2),mRNA	Macaca fascicularis	2791	2791	59%	0	96.9	1697	XM_045379030.1
6.	PREDICTED:Pan paniscus KLF	Pan paniscus	2743	2743	53%	0	99.6	1531	XM_014342534.3

	transcription factor 2(KLF2),mRNA								
7.	PREDICTED:Pongo pygmaeus KLF transcription factor 2(KLF2),mRNA	Pongo pygmaeus	2630	2630	53%	0	98.21	1524	XM_054465869.1

Table 20 BLAST analysis of MC3R genes and orthologs

Sl no.	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
1.	Homo sapiens melanocortin 3 receptor(MC3R)mRNA	Homo sapiens	2002	2002	100%	0	100	1084	NM_019888.3
2.	PREDICTED: Pan troglodytes melanocortin 3 receptor(MC3R) mRNA	Pan troglodytes	1975	1975	100%	0	99.54	1826	XM_525362.5
3.	PREDICTED:Pan paniscus melanocortin 3 receptor(MC3R) mRNA	Pan paniscus	1969	1969	100%	0	99.45	2613	XM_034947207.3
4.	PREDICTED:Gorilla Gorilla melanocortin 3 receptor (MC3R),mRNA	Gorilla Gorilla gorilla	1958	1958	100%	0	99.26	1112	XM_004062397.4
5.	PREDICTED: Pongo abelii melanocortin 3 receptor (MC3R),mRNA	Pongo abelii	1897	1897	100%	0	98.25	1172	XM_024239080.2
6.	PREDICTED: Pongo pygmaeus melanocortin 3 receptor(MC3R),mRNA	Pongo pygmaeus	1892	1892	100%	0	98.15	2693	XM_063659343.1
7.	PREDICTED :Macaca fascicularius melanocortin 3 receptor(MC3R),mRNA	Macaca fascicularius	1637	1637	94%	0	95.6	1846	XM_005569389.3