De novo transcriptome assembly and annotation of the brain of common pipistrelle bat (*Pipistrellus pipistrellus*)

KAMESWARA SHARMA YVR



Supervisor

Professor RICHARD EMES

Thesis submitted to the University of Nottingham
for the degree of

'M. RES'

IN

ADVANCED GENOMICS AND PROTEOMICS

September, 2016

ABSTRACT

Background: Bats are one of the most important, yet least understood, group of animals in the world. Pipistrelles are the most common and most widespread of all British bat species. In contradiction of the metabolic theory of ageing, bats are an anomaly as they have a high metabolic rate but are exceptionally long-lived, they also exhibit unique features such as flight and echolocation. Humans and other mammals share many common features of brain organization, but the differences in theta waves between bats and rats raises questions about how spatial information is represented in all brains. This study aims to sequence and annotate novel genes and transcripts from the bat brain.

Method: In the absence of genome or transcriptome information, transcripts first need to be reconstructed from sequence reads (or read pairs), which is referred to as *de novo* assembly. Further, counting the reads that fall onto a given transcript provides a digital measurement of transcript abundance, which serves as the starting point for biological inference. For each biological replicate brain sample of bat, *de novo* assembly was conducted using the software trinity (http://trinityrnaseq.github.io/). High quality transcripts were determined using transrate (https://github.com/Blahah/ transrate, http://hibberdlab.com/transrate). Intersect of the quality transcripts were identified from multiple replicates and used for transcriptome abundance estimation using HISAT (http://www.ccb.jhu.edu/software/hisat/index) and Stringtie (http://www.ccb. jhu.edu/ software/stringtie/). Annotation of novel transcripts used the software tool dammit (https://github.com/camillescott/dammit). Identification of genes homologous with reference genomes was determined using Inparanoid (http://inparanoid.cgb.ki.se/). Novel genes specific to bat brain tissue were identified by Inparanoid-BLAST. These genes were determined by expression abundance (Transcript per Million, > 5X the median) and hierarchical clustering. Phylogenetic trees were constructed for genes of interest involved in echolocation. Adaptive evolution of genes of interest were determined by comparison of selection strength (ω=dN/dS). The functional characterization of novel genes were confirmed by Gene Ontology and KEGG pathway analysis.

Results: We report the first full brain transcriptome of common pipistrelle and characterized genes associated with echolocation including Prestin (Slc26a5), Gjb2, Foxp2 which provide a remarkable insight into the bat brain transcriptome. We observed high similarity between the transcripts (within replicates) and orthologues genes were found in closely related species (bat, human, mouse and rat). The ratio of nonsynonymous to synonymous substitutions exemplifies that protein sequence of the genes associated with echolocation evolve in response to 'purifying selection'. Phylogenetic analysis revealed that these genes are most closely related with the little brown bat (*Myotis lucifugus*). In addition, this study identifies tissue-specific genes and interactions of classical neural process (Long-term potentiation, Long-term depression and gap junction) in this novel organism.

Conclusion: We comprehensively characterized the complete bat brain transcriptome of the common pipistrelle (*Pipistrellus pipistrellus*). This is the first study of its type and identified novel genes which are prominently involved in regulating the important classical neural processes of brain. This in turn gives profound resource for understanding the evolutionary perspective of these genes and also in brain associated diseases.

Potential Outcome and Future Study: Whole-genome assembly and functional genome annotation with RNA have become a realistic goal, and allow the identification of isoform- or allele-specific expression, coexpression and alternative splicing. Extending network analysis to several brain regions is potentially of high interest given the complexity of behavioural phenotypes. Understanding and identifying the adaptive selection of brain transcriptome genes in non-model organisms will provide insight into vertebrate brain evolution and potentially, identify genes involved in multifactorial human diseases. In order to extend this work we have isolated a high quality of RNA from the brain of zebrafish (whole brain tissue and three different regions of brain tissue separately in replicates). Samples have been submitted for NGS and analysis will follow protocols developed here.

Key words: Bat, Brain, *de novo*, echolocation, LTP, adaptive evolution

PUBLICATIONS AND PRESENTATIONS

Poster Presentation

Sharma, K., Pollard, A., Izquierdo A., Chakrabarti, L and Emes R. 2016. *" De novo transcriptome assembly and annotation of the brain of common pipistrelle bat (Pipistrellus pipistrellus)"* in LIKN-16 organized by University of Nottingham, UK at East Midlands Conference centre, University of Nottingham, Nottingham, UK on 3rd June, 2016. *PP*: 23.

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the test.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Kameswara Sharma Venkata Rama

Date 19th September, 2016

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my supervisor ***Prof. Richard Emes***, for his professional guidance, valuable help, encouragement, understanding and endless patience not only throughout this study but also for other situations.

I would like to thank ***Dr. Lisa Chakrabarti***, Co supervisor for her beneficial suggestions and valuable comments during the wet lab work.

I am immensely thankful to ***Dr. P Hemalatha Reddy***, Principal, Sri Venkateswara College, University of Delhi for her continuous support.
I take this opportunity to sincerely acknowledge ***University of Delhi*** for providing financial assistance under U21 Faculty Training Program which buttressed me to perform my work comfortably.

I thank ***Prof. Chingui Lui*** and ***Prof Dove Stekel,*** Course coordinators & mentors, Students, Administration of University of Nottingham for all the facilities and their timely help. Special thanks to Prof Charlie Hodgman for his inspirational and motivating lectures in the field of Genomics and Biochemical Networking.

I extend heartfelt thanks to my lab mates ***Amelia, Tom, James Freed*** for their help and training me in the lab. I would also like to thank ***Scot Hulmes*** for training me on Agilent Bioanalyzer.

Special thanks to ***Valery Ovens***, International Chaplin, for her kind support and guidance throughout our stay here in the University.

Thanks to my friends in Nottingham ***Eshan, Faraz*** for their friendly chats and guidance and also to ***my colleagues*** from University of Delhi for all the memorable time we have spent together.
A big thanks to my colleague, friend and flat mate, ***Ravindra Varma Polisetty*** for his constructive criticism and friendly time out for a year. Our cricketing spirit in Faculty Cricket League 2016 (UoN) was unforgettable.

Thanks to ***Ping (Mrs. Chingui), Dave*** and ***Di*** for all the social gathering and also keeping me free from the academic stress.

Special thanks to ***Sutton Bonington Cricket Club*** and ***Wollaton Cricket Club*** for making me self-motivated, enthusiaistic and engaging my week ends for four months. A big thanks to all the Lards of my ***cricketing friends*** for keeping my passion alive in the Home of Cricket.

Finally, I want to express my deepest gratitude and love to ***my parents***, ***friends' family*** and ***all my well-wishers*** for their understanding and encouragement that they have showed in every stage of my life, including this study.
I would not be able to finish my study without these people.
Last but not least, thanks to almighty for everything.

Kameswara Sharma Yadavilli

TABLE OF CONTENTS

Titles *Page No*

Abstract i
Publications and Presentations iii
Declaration iv
Acknowledgements v
Table of Contents vi
List of Figures viii
List of Tables xi
1. Introduction 1
1.1 Bats 1
1.1.1 Bats as Bioindicators 2
1.1.2 Current threats to bats 2
1.1.3 Classification of common pipistrelle 3
1.2 Phylogenetic Tree 4
1.3 Brain 6
1.4 Echolocation 8
1.4.1 Frequencies of echolocation call 9
1.4.2 Diversity in call structure 9
1.4.3 Echolocation properties 10
1.4.4 Genes in Echolocation 10
1.5 Transcriptomics 19
1.5.1 Gene expression 20
1.5.2 Transcriptome assembly strategies – NGS 23
1.6 Orthology and Inferring Functions 31
1.6.1 Homologs and orthologs 31
1.6.2 Methods to quantify orthologs 31
1.6.3 Importance of orthologs in genomics and phylogenomics 36
1.6.4 Ortholog inferring functions 37
2. Methods and Materials 39
2.1.1 Sample collection 39
2.1.2 Contamination 39

Titles *Page No*

2.2 RNA Extraction and Quality Assessment 39
2.3 Analysis of RNA-Seq Data 40
2.3.1 Quality control 40
2.4 Transcriptome Characterization (Assembly) 40
2.4.1 Transrate 40
2.4.2 HISAT and StringTie 42
2.4.3 Dammit-BUSCO (Annotation) 42
2.5 Identification of Novel Genes (Inparanoid-BLAST) Specific to Bat Brain Tissue 42
2.6 Highly Expressed Genes – Transcript Abundance (TPM) 44
2.7 Adaptive Evolution 44
2.8 Phylogenetic Tree Construction 44
2.9 Functional Characterization of Novel Genes 44

3. Results 45

3.1 RNA Sequence Data Analysis 45
3.2 *De novo* Assembly 46
3.3 Homologous and Orthologus Search 46
3.3.1 Long-term Potentiation 48
3.3.2 Melatonin Signaling 53
3.3.3 Analysis of genes involved in echolocation 53
3.4 Adaptive Evolution 65
3.5 Mitochondrial Proteome Study 65
3.6 Phylogenetic Study of Genes of Interest (GOI) of UoN Bat Brain Transcripts 78
3.7 Functional Characterization of Novel Genes (Gene Ontology and Pathways) 78

4. Discussion 90

5. Conclusion 97

6. Bibliography 98

7. Appendices 119

LIST OF FIGURES

Figure No.			*Page No*
Figure 1.1	:	Pipistrelle pispitrelle (Common Pipistrelle) in flight.	3
Figure 1.2	:	(a) unrooted tree, (b) rooted tree	5
Figure 1.3	:	An outline of mustached bat brain illustrating reciprocal connections between the amygdala and structures in the auditory forebrain of mammals	8
Figure 1.4	:	Convergent evolution of prestin sequences in echolocating bats and cetaceans	13
Figure 1.5	:	Process of Alternative splicing	21
Figure 1.6	:	Overview of the reference-based transcriptome assembly strategy	28
Figure 1.7	:	Overview of the *de novo* transcriptome assembly strategy	29
Figure 1.8	:	Alternative approaches for combined transcriptome assembly	30
Figure 1.9	:	A hypothetical gene tree and the resulting Inparanoid clusters are shown to illustrate inparalog (and thus co-ortholog) and outparalog assignments	32
Figure 1.10	:	Flow chart of Ortho MCL algorithm for clustering orthologous proteins	35
Figure 2.1	:	Flow chart of a typical RNA experiment with outlines of the study	41
Figure 2.2	:	Pipeline of RNA seq data analysis	43
Figure 3.1	:	Graph illustrating orthology count and type (1:1,1: m,m:1,m:m)	49
Figure 3.2	:	Bar plot depicting the percent count of orthologues to the type of orthologues (1:1,1:m,m;1,m:m) present in UoN bat	50
Figure 3.3	:	Euler plot depicts the sets and quantitative transcript data indicated by the numeric labels in the regions of the corresponding diagrams	51
Figure 3.4	:	Venn diagram showing the overlap of UoN transcripts with Myotis lucifugus, Human and Mouse for known genes	52
Figure 3.5	:	Genes encoding proteins involved in LTP, a classical neural process expressed in UoN transcripts	54
Figure 3.6	:	Highly expressed genes are involved in the process such as Melatonin signaling	55
Figure 3.7	:	Expression of SLC26aS encoded by the Prestin gene associated with echolocation and its known interacting proteins	57

Figure No.			*Page No*
Figure 3.8	:	Heat map showing median- normalized express Expression (TPM) for highly expressed transcripts of Echolocating genes of interest (GOI) across four replicates of UoN bat	61
Figure 3.9	:	Histogram showing the frequency distribution of echolocating genes of interest (UoN bat) - Mouse: Bat	62
Figure 3.10	:	Histogram showing the frequency distribution of echolocating genes of interest (UoN bat) - Mouse: Human	63
Figure 3.11	:	Histogram showing the frequency distribution of echolocating genes of interest (UoN bat) – Mouse: Rat	64
Figure 3.12	:	Cumulative Percentage (distribution) of dN/dS for UoN bat echolocating genes of interest (GOI)	67
Figure 3.13	:	Box and Whisker plots of expression (TPM > 5X Median) of transcript of abundances across four replicates (A,B,C,D, respectively) of UoN bat brain transcript	68
Figure 3.14	:	Heat map showing median- normalized express Expression (TPM) for highly expressed transcripts of mitochondrial proteins of interest associated with the brain functions across four replicates of UoN bat	72
Figure 3.15	:	Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat) – Mouse: Bat	73
Figure 3.16	:	Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat)- Mouse: Human	74
Figure 3.17	:	Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat) - Mouse: Rat	75
Figure 3.18	:	Cumulative Percentage (distribution) of dN/dS for mitochondrial proteins of interest (or genes of interest, GOI) associated with brain function (UoN Bat)	76
Figure 3.19 A:		Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Prestin gene (SLC26a5)	79
Figure 3.19B:		Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Fos protein(gene)	80
Figure 3.19C:		Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Gap Junction Protein (Gjb2) or gene	81

Figure No.		*Page No*
Figure 3.19D:	Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Foxp2 protein	82
Figure 3.19E:	Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Ot of protein	83
Figure 3.19 F:	Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Wnt8a protein	84
Figure 3.20 :	**A,B&C.** Gene Ontology. Visualization of GO terms of protein/genes of interest (GOI), 5X Median identified based on enrichment analysis	85
Figure 3.21 :	Scatter plot of KEGG pathway enrichment statistics and the most enrichment pathway of UoN Bat brain transcript (GOI, 5 X Median)	86
Figure 3.22 :	Visualize GAGE pathway analysis results using pathview (KEGG) – Gjb2 pathway	87
Figure 3.23 :	Visualize GAGE pathway analysis results using pathview (KEGG) – Long-term Potentiation pathway	88
Figure 3.24:	Visualize GAGE pathway analysis results using pathview (KEGG) – Long-term Depression pathway	89

LIST OF TABLES

Table No.			*Page No*
Table 1.1	:	Classification of common pipistrelle	3
Table 1.2	:	A list of splice-aware short read aligners	26
Table 1.3	:	A comparison of the features of existing software for transcriptome assembly	27
Table 3.1	:	Assesment of RNA quality and quantity used in the study	45
Table 3.2	:	Depicting the initial raw reads of NGS	45
Table 3.3	:	Bat Assembly statistics as per the pipeline of the study	47
Table 3.4	:	Count Data orthologues of UoN bat to its type of orthologues in Ensembl Bat, Human and Mouse	47
Table 3.5	:	Percent data orthologues of UoN bat to its type of orthologues in Ensembl Bat, Human and Mouse	48
Table 3.6	:	Genes enriched in Brain which are highly expressed and orthologues in Mouse and Human, Ensembl Bat and UoN Bat	52
Table 3.7	:	Genes of interest in UoN Bat involved in echolocation (Vocalization and hearing)	56
Table 3.8 A	:	Orthologues of Bat brain transcriptomes in Mega bat and micro bat involved in Echolocation along with Ensemble protein ID	58
Table 3.8 B	:	Genes of interest (GOI) involved in echolocation of Bat, Orthologues in Mega bat and micro bat and are unique to UoN Bat	60
Table 3.9	:	Selection strength (ω) of Echolocating genes of interest (UoN bat), orthologues mouse with Ensembl bat, mouse with human and mouse with rat along with dN and dS values, respectively	66
Table 3.10	:	Mitochondrial Bat (UoN) Proteins/Genes involved in neural and Brain functions	69
Table 3.11 A	:	Orthologues of Bat and Mouse Mitochondrial proteins in Mega bat and micro bat involved in brain functions along with Ensemble protein ID	70

Table No.			*Page No*
Table 3.11 B	:	Genes of interest (GOI) of Bat and Mouse mitochondrial proteins involved in brain functions, Orthologues in Mega bat and micro bat and are unique to UoN Bat	71
Table 3.12	:	Selection strength (ω) of UoN Bat and Moue Mitochondrial proteins/Genes of interest, orthologues mouse with Ensembl bat, mouse with human and mouse with rat along with dN and dS values, respectively	77

1. INTRODUCTION

Humans and other mammals share many common features of brain organization, but the differences in theta waves between bats and rats raises questions about how spatial information is represented in all brains. However, with the advent of vocal communication of bats and the fact being bats, the second largest mammalian order, there is a paucity of information particularly at the molecular level and importantly lack of the genomic and transcriptomic information for these species. To address this fundamental gap and the paucity of information this study aims to sequence and annotate novel genes and transcripts from the bat brain (Pipistrelle), which is the pilot analysis.

The first chapter of this thesis represents the background information of Bats and its position in terms of Phylogenetic studies as per the literature evidences. In addition, it comprises of information on bat Brain functions and the variation of brain with phylogeny. Furthermore, literature review explains the transcriptomic approach, methods and emphasis towards Orthologus and Gene ontology studies along with network analysis. This chapter also covers the background information of Adaptive evolution in terms of genes of interest of the study. Second chapter explains the methodology to study the proposed aim of the study, "*de novo* assembly of bat brain transcriptome of common pipistrelle". Third chapter which includes results and findings of the present work with illustrative figures along with tabular data.

In chapter four we exemplified the augmented part of our findings in evidence to the literature. Chapter five comprises of the conclusive part of our work along with the future perspectives and potential outcome of the proposed research work. Bibliography was cited in the chapter six. Supplementary files are listed in Appendices.

1.1 Bats

Bats belong to the order Chiroptera and are diverse with variety of unique features among all mammals. According to the report of Wilson and Reeder (2005), bats constitute an abundant and diverse mammalian lineage comprising approximately 20% of all known mammalian diversity.

Bats diverged from other mammals approximately 50 million years ago (Moratelli and Calisher, 2015) and are divided into two major suborders; the Yinpterochiroptera (megachiroptera) and the Yangochiroptera (microchiroptera). Yinpterochiroptera includes the family Pteropodidae and genera *Rousettes* and *Pteropus* whereas Yangochiroptera includes the family Myotidae including that studied here, and genus *Myotis* (Teeling et al., 2002).

Bats have been recognized as significant natural resources in recent years due to its vital ecological roles as nocturnal insect predators, pollinators, seed dispersers (Gannon et al., 2003). However, there is a predominant decrease in the bat population with many species going extinct (Stebbings, 1988). This is a major concern in different ecosystems as they act as predators to a wide range of insects, other species and also in the pollination (Best, 1997, Racey, 2003).

1.1.1 Bats as Bioindicators

Bats serve as ideal bio indicators to monitor trends in populations, possession by bats of high trophic levels, and the widespread distribution of taxa (Jones et al., 2009). Bats are ecological indicators of ecosystem disturbance (Fenton et al., 1992) , (Medellín et al., 2000), habitat quality (Kalcounis-Rueppell et al., 2007), and have been included as one of the United Kingdom's Biodiversity Indicators (Bat Conservation Trust 2011, http://www.bats.org.uk/). The research outcome of U.S. Fish and Wildlife Service (Fish and Service, 2012) on bat detectors enabled to understand bat ecology and behaviour in turn used to guide vital wildlife management decisions.

1.1.2 Current threats to bats

More than 1200 species of bats have been described (Simmons, 2005) and 172 species are threatened (i.e., vulnerable, endangered, or critically endangered) while five are now extinct (IUCN, 2012). Bat populations are being impacted around the world, from habitat destruction to climate change (Jones et al., 2009), but two current threats facing bats today are wind-energy developments and white-nose syndrome (WNS). Migratory species can collide with the wind turbines and result in death (Kunz et al., 2007) (Rollins et al., 2012). Modification and fragmentation of natural habitats due to the

development of agriculture is also one of the major concerns (Racey, 2003, Kunz et al., 2003) with human demands on the ecosystem one of the prominent threats to bats (Fenton et al., 1992).

1.1.3 Classification of common pipistrelle

Pipistrelles (Figure 1.1) are the most common and most widespread of all British bat species. The common and soprano pipistrelle, are the two commonest pipistrelle species identified in the 1990s. Frequency of their echolocation calls is the easiest way to differentiate between them besides having almost similar characteristics.

Common pipistrelles emit sounds which are above the range of human hearing with the exception of social calls that may be heard by children and adults with good hearing. Bat detector, heterodyne is used to locate the echolocation calls of bats which is in the range of 45 and 70 kHz (Bat conservation Trust 2010, http://www.bats.org.uk/pages/national_bat_monitoring_programme_annual_report_2010.html).

Table 1.1: Classification of common pipistrelle

Kingdom	Animalia
Phylum	Chordata
Class	Mammalia
Order	Chiroptera
Sub order	Microchiroptera
Family	Vespertillionidae
Genus	*Pipistrelles*
Species	*pipistrelles*



Figure 1.1 *Pipistrelle pispitrelle* (Common Pipistrelle) in flight.

Pipistrelles in contradiction of the metabolic theory of ageing, bats are an anomaly as they have a high metabolic rate but are exceptionally long-lived; they also exhibit unique features such as flight and echolocation. Despite unique features and that bats, are the second largest mammalian order (Table 1.1), there is a paucity of information particularly at the molecular level and importantly a lack of genomic and transcriptomic

information for these species. To address this fundamental gap and the paucity of information this study aims to sequence and annotate novel genes and transcripts from the bat brain (Pipistrelle).

1.2 Phylogenetic Tree

Phylogenies form the backbone of evolutionary biology and represent tools that underlie a broad spectrum of evolutionary and ecological studies (Felsenstein, 1985) (Harvey and Pagel, 1991).

Study of phylogeny is noteworthy to address diversified biological questions such as relationships among species or genes, the origin and spread of viral infection and the demographic changes and migration patterns of species.

Phylogenetics is an area of research concerned with finding the genetic relationships between various organisms. Originally, phylogenetics mainly used morphological features such as size, color, fur, or other physical characteristics to determine relationships. In contrast, modern phylogenetics relies on information extracted from genetic material such as DNA, RNA or protein sequences (Shamir, 1984).

“A phylogeny is a tree containing nodes that are connected by branches. Each branch represents the persistence of a genetic lineage through time, and each node represents the birth of a new lineage. If the tree represents the relationship among a group of species, then the nodes represent speciation events” (Yang and Rannala, 2012).

A phylogenetic tree is a binary tree representation of the resulting relationship. It comprises of two main types viz., a) rooted trees -- those that have a single node from which all nodes are derived, and b) unrooted trees -- those that do not originate from one clear node. The tree follows standard graph theory notation were each species is represented as a node or leaf, and the relationship between species is referred to as an edge or branch (Figure 1.2) (Rizzo and Rouchka, 2007). With the advent of new sequencing technologies and the completion of various genome projects, phylogenetics has entered the era of genome-scale data sets.

Phylogenetic Reconstruction of Position of Bats

Bats are an ancient and diverse group that originated approx. 65 million years ago in the late Cretaceous/early Paleocene and underwent a rapid radiation in the Eocene.

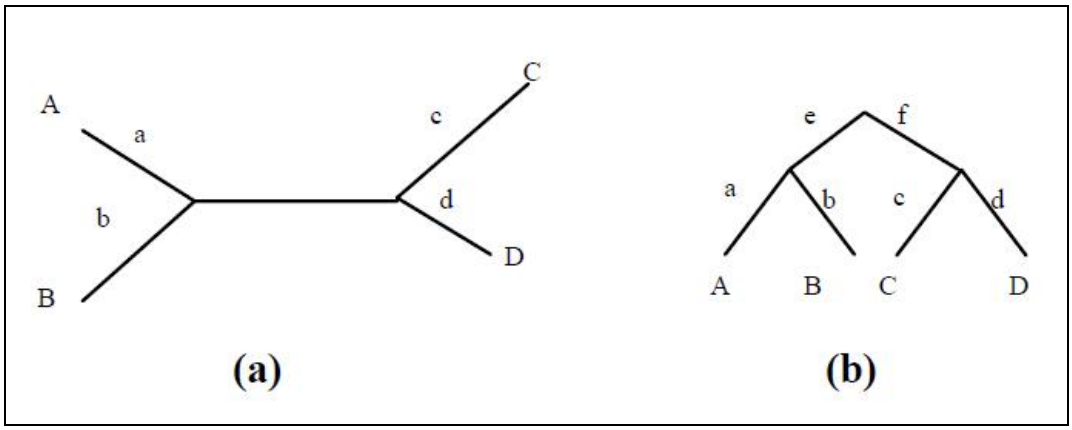


Figure 1.2 (a) unrooted tree, (b) rooted tree (Figure taken from (Rizzo and Rouchka, 2007).

However, the exact evolutionary relationship of bats to their closest mammalian relatives is poorly understood due to their phenomenal morphological features associated with flight, a lack of intermediate forms, and a poor fossil record (Teeling et al., 2005).

There are two most widely accepted methods for the construction of phylogenetic tree viz., distance based and character based. These exhibit enormous options in constructing the trees in two different directions.

Traditional approaches for estimation of phylogeny involved comparisons of phenotypic data from morphology or other organismal characteristics amenable to observation. Methods of phylogenetic reconstruction have been refined during subsequent decades, as more powerful computing machines enabled utilisation of more sophisticated algorithms, and phylogenetic inference has permeated into almost every field of biology. Examination of molecular structure of proteins and nucleic acids was soon implemented into estimation of phylogeny and molecular data became an immense source of information for systematic studies.

During the twentieth century, analysis of DNA sequences has become routine for phylogenetic inference and today systematics routinely employ PCR to amplify the desired fragments of DNA, sequence these on automated sequencers, and analyse many

kilobases of sequence information to reveal relationships that would otherwise remain hidden.

In this study we discuss few issues relating to the analysis of such large data sets and the position of pipistrelle in the phylogenetic tree with reference to the orthologues and paralogues genes or transcripts.

1.3 Brain

Brains are not homogenous structures, different parts of the brain have been demonstrated to serve quite different functions (Brodal, 1981), and the histological structure of a region may vary considerably from that of its neighbours (Baron-Cohen et al., 1996). Likewise, not all parts of brain tissue serve information processing functions. Neurons form complex networks within the brain. These networks are formed from synapses between the axon of one neuron and the dendritic field of another (Zigmond, 1999). While a neuron may only have one cell body, the distribution and complexity of its neuropil (i.e. its dendritic and axonal parts) may vary considerably. Thus an increasing complexity of brain function might result not from an increase in neuron number, but in the quantity and distribution of neuropil per neuron (an increase in white matter relative to grey matter) (Barton and Harvey, 2000).

Interpretation of the function of particular parts of the brain is dependent on an understanding of the connectivity of the nervous system, i.e. how different parts are connected together to interact and exchange information. Although it is commonly assumed that all mammals are basically similar in the patterns of brain connectivity they exhibit, however according to the study of Preuss (2000) this is not universally true. It has been widely recognized that mammal brain size predominantly increases over evolutionary time. Safi et al. (2005), in converse observed reduction in the brain size among bats during the course of evolution in contrast enlargement in others (Yao et al., 2012).

Data for brain structure volumes was analysed using multiple regression to test for correlated volumetric evolution in bats (Mammalia, Chiroptera). Significant partial correlations were found between major brain subdivisions, and between structures within the Auditory, Olfactory and Sensorimotor Systems that were predicted to have

evolved together on the basis of anatomical connectivity and known functional relationships.

Vision is the primary sense for locating food at a distance in Megachiroptera, whereas the visual system is primitive in Microchiroptera (Neuweiler, 2000). The features of nervous system organisation which provide high levels of sensitivity or other measures of visual function reflect similarity between mammals relying on a sense of sight, such as Megachiroptera and Primates. In addition, the relative size of hippocampus is large in Megachiroptera in comparison with Microchiroptera and share many features with Primates (Buhl and Dann, 1991).

Neuweiler (2000) explained the spatial memory of the animal in its search for food by the size of the hippocampus. The qualitative similarities described above might also be features suggestive of the functional demands on spatial memory in both Megachiroptera and Primates.

Among the many tasks that brains perform are the filtering and preparation of sensory information. Different foraging ecologies necessarily entail different sensory requirements. For example, the senses that are most useful for an insectivorous bat (well-developed echolocation) may not be as useful for a nectar-feeding bat. Bats navigate the night sky using echolocation. They emit sound waves that bounce off their surrounding and back to their ears. To support this remarkable skill, the anatomy and brain of some bats have evolved specialized features.

Their vocal cords emit precise ultrasonic pulses at a specific frequency and its harmonics. Each pulse consists of a constant pitch, followed by a punctuating ‘CHIRP’. Some insects, such as certain moths have evolved a jamming response to bat ‘CHIRPS’. By rubbing their genitals, they emit high-frequency signals that disrupt or confuse the bat SONAR.

Special structures in the inner ear (Bat cochlea) are tuned to the second harmonic of the emitted pulse, making it sensitive to echo. In the bat brain, special areas in the cortex have developed to process the ‘BIOSONAR’. The FM-FM area uses the return time of the echo to calculate how far a bat is from things (Figure 1.3). Other areas specialize in detecting and compensating for DOPPLER SHIFTS, or changes in the return pulse

frequency. This area is so sensitive, a bat can process 'Acoustic Features' like the wing beats of insects and even detect textures. The present work emphasis on the brain function of Pipistrelle bat in its response to echolocation and other neural networks. Also the outcome of this study may shed information and be a potential role in understanding 'how the brain varies with phylogeny'.

1.4 Echolocation

Donald Redfield Griffin coined the term "echolocation", he explained how bats use echoes of sounds they produce to locate objects in their path. This discovery about animal behaviour spawned a new field of research that attracted scientists from animal behavior, neurobiology, signal theory, anatomy, and physiology. The evolutionary history of echolocation in bats is not clear, as fossils provide little direct evidence, and mostly ecological approach was explored in the echolocation studies.

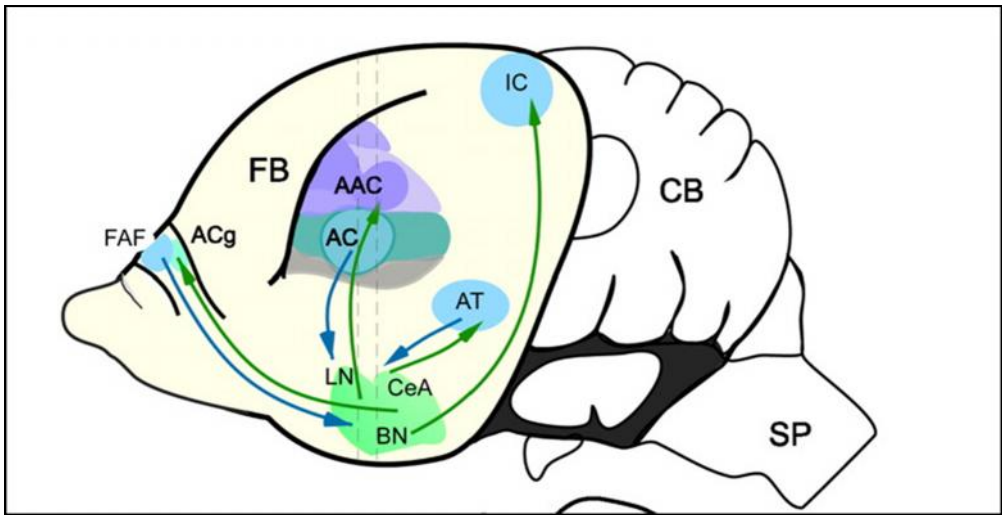


Figure 1.3: An outline of mustached bat brain illustrating reciprocal connections between the amygdala and structures in the auditory forebrain of mammals (Armony et al., 1997; LeDoux et al., 1990). Basolateral amygdala (BLA) also receives projections from the dorsomedial prefrontal cortex (Mc Donald et al., 1996) where a frontal auditory field (FAF) is present in this species. Central nucleus of the amygdala (CeA) projects back to the anterior cingulate (ACg) (Ghashghaei and Barbas, 2002), and basal nuclei (BN) projects to the inferior colliculus (IC) within the brainstem (Marsh et al., 2002). Vertical dashed lines indicate the antero-posterior levels of the sections. AC, anterior cortex; AAC, auditory association cortex; FB, forebrain; CB, cerebellum; LN, lateral nucleus; AT, auditory thalamus; SP, spinal cord. (Source: (Naumann and Kanwal, 2011).

Echolocation call structure of bats is diverse but not clear as how and why they have been evolved, or the impact of these on other aspects of the evolution of bats is unclear.

Bats are unique in having evolved both sophisticated laryngeal echolocation and powered flight. They have more diversified habit and habitat.

Echolocation has evolved in environments where vision is of limited use; the nocturnal niche of bats reduces the efficacy of vision and must reckon on echolocation as their primary sensory method (Nowak, 1994). Echolocation, or biosonar, is the active use of calls and interpretation of their echoes to detect, localises, and classifies objects (Griffin, 1944, Jones, 2005). Predominantly more than 90% of all bat species use echolocation to localize obstacles in their environment by comparing their own high frequency sound pulses with returning echoes (Griffin, 1958).

1.4.1 Frequencies of echolocation call

Bat echolocation calls descend to the ultrasonic frequency range (above the human hearing 20 kHz (Pumphrey, 1950), although some echolocation calls are audible to humans. The lowest frequencies used in echolocation are emitted by the spotted bat (*Euderma maculatum*, Vespertilionidae), which was around 9 kHz (Fullard and Dawson, 1997) which is in contrast with Percival's short-eared trident bat (*Cloeotis percivali*, Hipposideridae), reaches highest known peak energy frequencies about 212 kHz. High frequencies are also directional and can restrict the spread of echolocation calls to their target of interest (Brudzynski and Fletcher, 2010). Most bats emit an intermediate frequency of 20 to 60 kHz (Fenton et al., 1998) as they have to detect small targets in this low range and at the same time have to prevent extremely high frequency because of excess atmospheric attenuation limits (Brudzynski and Fletcher, 2010, Lawrence and Simmons, 1982).

1.4.2 Diversity in call structure

There is a wide diversity in the call structure of echolocation. Some species of Rousettus use tongue clicks as echolocation signals, while all other bats turn out to use vocalizations produced by vocal cords. Hunting flying prey or prey moving on water surface (Kalko, 1995) use high intensity (>110 dB sound pressure level [SPL] @10 cm) echolcation calls. In contrast *Myoits bechsteinii* and *Myotis nattereri* use low intensity echolocation calls to produce feeding buzzes (Von Helversen and Winter, 2003). Jones

and Teeling (2006) have classified bats of suborder, micorchiroptera as constant frequency- frequency model (CF-FM) or frequency modulation (FM) bats based on the echolocation calls. The diversity of echolocation call structure within these broad classes delves into the evolution of diverse spectral and temporal neural processing strategies using behaviourally-relevant sounds (Wenstrup et al., 2012). Extensive studies were performed on the FM sweeps, neural selectivity for the rate and direction of change in frequency (Gittelman et al., 2009, Fuzessery et al., 2011, Washington and Kanwal, 2012).

1.4.3 Echolocation properties

"A single emission of sound is referred to as a call, and a series of calls is a call sequence" (Fenton, 1999). Echolocation calls of bats subsist of three phases: search, approach, and terminal (Griffin et al., 1960). Search phase calls are produced to locate prey, approach phase calls are to identify exact locations of prey, and just before to the capture terminal phase calls are produced. Majority of the bat echolocation calls come under the search phase, and exhibit consistency in structure throughout the call sequence, and may possess species-specific characteristics (Betts, 1998, O'Farrell et al., 1999).

1.4.4 Genes in Echolocation

Understanding the evolutionary history of echolocation was made easy with the advent of the emerging field of genomics, in particular following the sequencing of several bat species' genomes (*Myotis lucifugus* (Vespertilionidae), *Rhinolophus ferrumequinum* (Rhinolophidae), *Carollia perspicillata* (Phyllostomidae), and *Pteropus vampyru*s (Pteropodidae)). Importantly two genes were identified to be having a prominent role in echolocation viz; FOXP2 and PRESTIN.

FoxP2: Transcription factor (FOXP2) is involved in the neural control of orofacial coordination, and also identified as having a prominent role in human speech development and language comprehension (Fisher and Marcus, 2006). This gene has a wide diversity and a higher rate of evolution in comparison with other vertebrates (Li et al., 2007).

The inferior colliculus of mice, a brain region that shows enlarge and morphological specialization in bats shows strong expression of FOXP2 (Ferland et al., 2003, Glezer et al., 2007). "Foxp2 is mostly involved in controlling call parameters during more complex echolocation tasks" (Metzner and Schuller, 2010).

The role of FoxP2 in echolocation behaviour in bats and is exemplified by gene silencing of Foxp2 by lentivirus – mediated RNA interference (Chen et al., 2013). Knockdown experiments by Haesler et al. (2007) show that the mechanism of Foxp2 in the basal ganglia nucleus area X is important for accurate vocal imitation in birds. Foxp2 silencing studies in *H. armiger*, confirms the importance of this gene in echolocation by disrupting the Doppler shift compensation (Metzner and Schuller, 2010). In addition, the Foxp2 gene is highly expressed in the supragenicuIate nucleus and the ACC in the brains of bat species that use laryngeal echolocation (*Rhinolophus ferrumequinum*, *H. armiger*, *Myotis ricketti*), alternatively in species without laryngeal echolocation (*Rousettus leschenaultii* and *Cynopterus sphinx*) the expression was stronger in the olfactory tubercles.

Vernes et al. (2007), (Vernes et al., 2011), reported Foxp2 either represses or activates gene expression at promoter sites involved in the modulation of synaptic plasticity, neurodevelopment, neurotransmission, and axon guidance involving the binding targets in human neuron like cells.

Prestin: Prestin is the second gene of interest; it is a motor protein of cochlear outer hair cells (OHC) (Zheng et al., 2000). Prestin is quicker and amplifies aural sensitivity up to 100 fold compared to other cellular motor proteins. It is also useful for high frequency hearing in mammals (Liberman et al., 2002). All echolocating bats are placed as a monophyletic clade to the exclusion of the old world fruit bats (Pteropodidae) based on the phylogeny of prestin sequence. This depicts the old phylogeny of bats based on morphological characters, but it's prone to be a convergence in the functional gene, prestin (Li et al., 2008).

Genomic comparisons have the potential to lead to further discoveries of genes involved in echolocation.

Knockout studies of mice by Liberman et al. (2002) confirms that Prestin enhance auditory sensitivity by 100 fold as a result of its contraction and mechanical elongation. Kirwan et al. (2013) undertook "phylogenetic and selection analyses of 11 genes implicated in hearing (*Myo15 (Myo15a), Ush1g, Strc, Tecta, Tectb, Otog, Col11a2, Gjb2, Cldn14, Kcnq4* [which was reported as showing parallel evolution by Liu et al. (2012), *Pou3f4*) and found good support for the paraphylogeny of echolocating bats across these loci as well as a high level of evolutionary conservation".

Aside from FOxP2 and Prestin, other genes have been associated with echolocation.

Kcnq4: The convergent amino acid residues in Prestin sequences from echolocating bats were found to be concentrated in the areas of protein involved in voltage sensing (Liu et al., 2011). Kcnq4 gene encodes a protein which acts as a voltage gated potassium channel having a pivotal role in the regulation of electrical signalling (Figure 1.4). One of the most intriguing findings by Kharkovets et al. (2000) is the expression of KCNQ4 in brain and confirmation of expression in the outer hair cells concentrated at the basal part of cochlea by using specific antibodies. Progressive loss of high frequency hearing, DFNA2 (an autosomal dominant form of progressive hearing loss) is caused by the mutations of Kcnq4 gene in humans. Thus, its evolution in bats is of paramount interest (Kharkovets et al., 2006). Trussell (2000) observed the pharmacologic modification of the ion channels mutant in DFNA2 which result in the development of new treatments that are selective for a sensory modality and also for the quality of perception.

Tmc1 (Transmembrane Cochlear Expressed Gene 1): TMC1, a 6-pass integral membrane protein, is a component of mechanotransduction channels in hair cells of inner ear (Pan et al., 2013). The orthologous mouse Tmc1 cDNA was isolated by Kurima et al., 2000 using RT PCR along with 5-prime and 3-prime RACE of mouse inner ear cDNA. It was observed that Tmc1 mRNA is required for normal functioning of cochlear hair cell and is expressed in hair cell of postnatal cochlea and vestibular end organs.

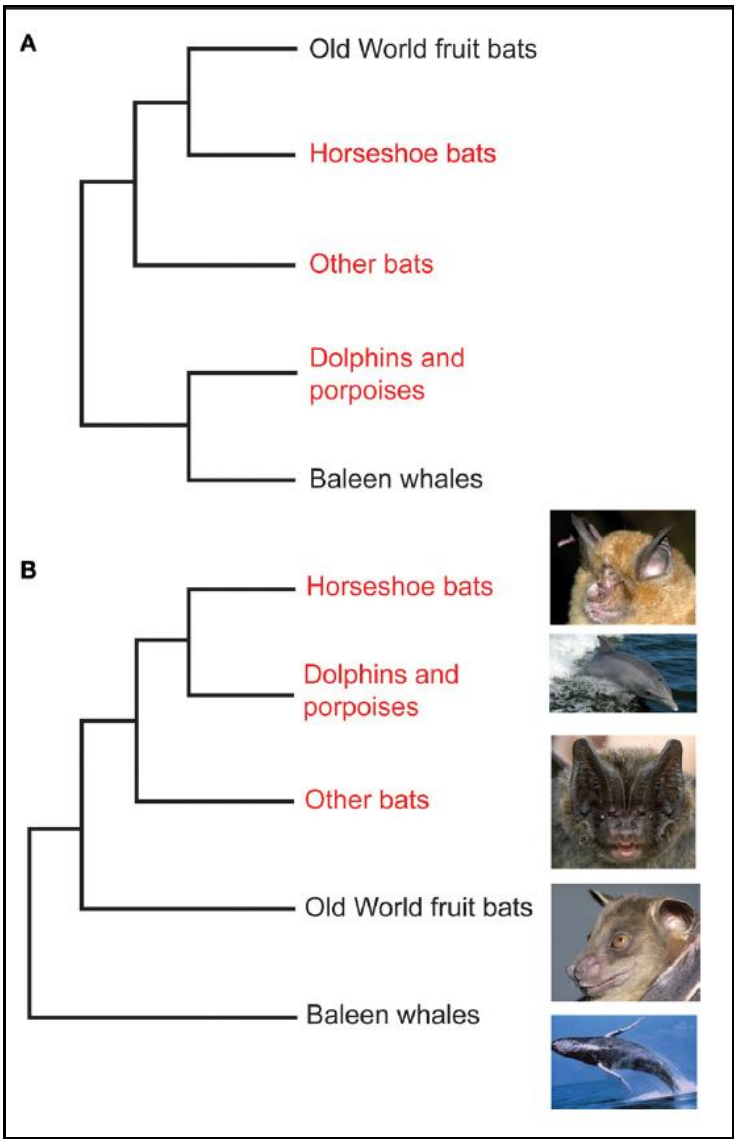


Figure 1.4: Convergent evolution of prestin sequences in echolocating bats and cetaceans. (A) Phylogeny of Old World fruit bats (Pteropodidae), horseshoe bats and their close relatives, other bat lineages studied by Liu et al. (2010) Li et al. (2010), dolphins and porpoises, and baleen whales as determined from large-scale molecular sequence analyses. (B) The arrangement that arises from analysis of the *Prestin* gene. Bat groups highlighted in red use laryngeal echolocation and cetacean groups highlighted in red exhibit biosonar behavior. In (A) these echolocating taxa are paraphyletic—non-echolocating Old World fruit bats are sister to echolocating horseshoe bats, and echolocating dolphins and porpoises are sister to non-echolocating baleen whales. In (B) all echolocating taxa form a monophyletic group, and dolphins and porpoises are the sister group of horseshoe bats. Photographs are species studied by Liu et al. (2010) Li et al. (2010). From top to bottom they are the greater horseshoe bat *Rhinolophus ferrumequinum* (G. Jones), the bottlenose dolphin *Tursiops truncatus* (NASA), the Beijing barbastelle *Barbastella beijingensis* (J. R. Flanders), the greater short-nosed fruit bat *Cynopterus sphinx* (G. Jones), and the humpback whale *Megaptera novaeangliae* (NOAA). (Source: (Jones, 2010).

Mice express 2 Tmc1 spice variants which differ in the presence or absence of exon 2 and in use of translation initiation codon in either exon 1 or exon2. RT-PCR and in situ hybridization studies result in the increased expression of Tmc1 and Tmc2 in cochlear hair cells of mouse within 5 Days of birth and that at postnatal day 8 it was Tmc1. Kurima et al. (2002), also observed the existence of Tmc2 in hair cells of vestibular sensory organs of mouse.

A met4120-to-lys (M412K) missense mutation was identified in Tmc 1 gene of mouse deaf mutant Beethoven (Bth) which therefore proved to be a mouse model for autosomal dominant DFNA 36 (Vreugde et al., 2002).

Pjvk (Dfnb 59 gene): The complete DFNB 59 gene sequence was isolated from human testis cDNA library by Delmaghani et al. (2006) and the gene product was denoted as pejvakin from the Persian word, echo, reflecting the fact that the mutations in the gene caused neuronal deafness (DFNB59; 610220) in four Iranian families. With respect to N-terminal regions, 352-amino acid pejvakin protein is similar to DFNA 5 (608798) with 32% identity over a stretch of 250 amino acids.

Pejvakin protein contains a Nuclear localisation signal (residue 249-258) and a zinc finger motif (residue 305-331) (Delmaghani et al., 2006). As observed by Immunohistoflurorescence microscopy, the cell bodies of neurons of afferent auditory pathway contained pejvakin. Studies conducted by Delmaghani et al. (2006) on the auditory analysis of the Dfnb59 knock in mice confirmed that only afferent neurons were affected with maintenance of outer hair cell function, in contrast sirtaki mice indicated dysfunction in both (Schwander et al., 2007).

Cdh 23 and pcdh 15: CDH 23 and PCDH15 are two cadherins that are linked to inherited forms of deafness which interact to form extracellular filaments (tip links), connect stereocilia and act as mechanoelectrical transduction channel (Kazmierczak et al., 2007). These cadherins (CDH 23 and PCDH 15) were localized to the upper and lower part of tip links, respectively by the Immunohistochemical studies using rodent hair cells. DiPalma et al. (2001) identified the Cdh23 gene as that mutated in waltzer mice. Biochemical experiments of channel Kazmierczak et al. (2007) revealed that CDH 23 homodimers interact in trans with PCDH15 homodimer to form a filament in

structural similarity with tip links. PCDH 15 mutation and ionic effect of tip link integrity result in a recessive form of deafness and disrupted interactions between the two cadherins, CDH 23 and PCDH 15.

A study conducted by Kazmierczak et al. (2007) concludes the mechanisms of sensory impairment associated with mutations in CDH 23 and PCDH15 and also the molecular composition of tip links. The 10.5 kb Cdh23 cDNA encodes a large, single-pass transmembrane protein, otocadherin comprising of 69 exons (Di Palma et al., 2001, Wagatsuma et al., 2007). Zheng et al. (2005) in his studies concluded that, "CDH23 and PCDH15 play an essential long term role in maintaining the normal organization of stereocilia bundle". Bolz et al. (2001) found that CDH23 and harmonin are expressed in the stereocilia of hair cells, and in the retina within the photoreceptor cell layer. CDH 23 deficient mice have splayed stereocilia, this made Siemens et al. (2002) to conclude that, CDH 23 and harmonin are part of transmembrane complex that connects stereocilia into bundles, which in turn disrupts the stereocilia bundles and cause deafness in patients with Usher syndrome type I. Age-related hearing loss (AHL) in common inbred mouse strains is a genetically complex quantitative trait, a synonymous SNP in exon 7 of the Cdh23 gene is the key role in causing the AHL and the deafness modifier mdfw, modifier of deafwaddler (Noben-Trauth et al., 2003).

Otof (Otoferlin): Using RT-PCR, Otof expression was identified in mouse cochlea, vestibule and brain and a 7 kb otoferlin mRNA in the human brain was also identified using Northern blotting (Yasunaga et al., 2000). Several other alternatively spliced transcripts suggested the existence of long isoforms (with 6C2 domains) in humans and mice and short isoforms (3 C2 domains) restricted to humans. Existence of an alternative spliced isoform of OTOF in human cochlea was demonstrated by (Choi et al., 2009). It was observed that exon 48 was exclusively used to encode the C-terminal 60 amino acids of this isoform by human cochlear transcripts which lack exon 47. Since this isoform encodes a unique alternative C terminus affected by some DFNB 9 mutations, it is therefore required for human hearing.

Chrna10 (Cholinergic Receptor Neuronal Nicotinic Alpha Polypeptide 10): Also called as ACETYLCHOLINE RECEPTOR, NEURONAL NICOTINIC, ALPHA-10

SUBUNIT. Neuronal nicotinic acetylcholine receptors (nACHRs) support chemical synaptic transmission throughout the nervous system. Expression of CHRNA10 protein in activated T and B lymphocytes in tonsil and in inner ear hair cells were confirmed by (Lustig et al., 2001) in his Immunohistochemical analysis.

The functional role of CHRNA10 protein in normal olivocochlear activity along with Chrna9 was confirmed by (Vetter et al., 1999).

MYOSIN XVA; MYO15A (Myosin 15): Probst et al. (1998) found MYO15A contains an N-terminal motor domain, which comprises of 2 light-chain binding IQ motifs, and a tail region containing a MyTH4 and a talin like domain. Expression of MYO15A in human fetal and adult brain was confirmed in his Northern blot analysis and it is detected in the human fetal cochlea by RT-PCR analysis.

Wang et al. (1998) isolated a partial MYO15A cDNA from a human chromosome 17-specific human cDNA library. Elongation of stereocilia is regulated by Myosin 15A/Whirlin interaction. Myo15a and whirlin interact with Epidermal growth factor receptor pathway substrate 8 (Eps8) at the tips of stereocilia in mouse inner and outer cochlear and vestibular hair cells. Knockout of this receptor Esp8, result in the shortening of stereocilia and profound deafness (Manor et al., 2011).

Nal et al. (2007) in his study observed 16 novel MYO15A mutations that co-segregated with DFNB3 hearing loss in 20 families from Turkey, India and Pakistan.

Strc (Stereocilin): STRC is a large extracellular structural protein found in the stereocilia of outer hair cells in the inner ear and encodes Stereocilin. It is associated with horizontal top connectors and the tectorial membrane attachment crowns that are important for proper cohesion and positioning of the stereociliary tips (Verpy et al., 2011, Francey et al., 2012).

Otog (Otogelin): An N-glycosylated protein, is present in the acellular membranes spanning around the six sensory epithelial patches of the inner ear; in the cochlea, the tectorial membrane (TM) over the organ of Corti, and in the vestibule (the balance sensory organ), the otoconial membranes over the utricular and saccular maculae, as well as the cupulae over the cristae ampullares of the 3 semi-circular canals. These

membranes are involved in the process of mechanotransduction. Ultrastructural analysis of the TM in the cochlea indicated the involvement of Otogelin in the organization of its fibrillar network. Simmler et al. (2000) also observed the role of Otogelin in the resistance of TM to sound stimulation and it concluded the functional role of Otogelin for human nonsyndromic deafness.

Gjb2 (Gap Junction Protein, beta 2): The *GJB2* gene provides instructions for making a protein called gap junction beta 2, more commonly known as connexin 26. Connexin 26 is a member of the connexin protein family. These proteins form channels called gap junctions that permit the transport of nutrients, charged atoms (ions), and signaling molecules between adjoining cells. The size of the gap junction and the types of particles that move through it are determined by the particular connexin proteins that make up the channel. Gap junctions made with connexin 26 transport potassium ions and certain small molecules. Because of its presence in the inner ear, especially the snail-shaped structure called the cochlea, this research is focussed in this protein's role in hearing. Hearing requires the conversion of sound waves to electrical nerve impulses. This conversion involves many processes, including maintenance of the proper level of potassium ions in the inner ear. Some studies indicate that channels made with connexin 26 help to maintain the correct level of potassium ions. Also literature review suggest that, connexin 26 is required for the maturation of certain cells in the cochlea (Bouchon and Chakraborty, 2014). This protein function and along its network pathway was explored in this study (covered in the results and discussion).

Cldn 14 (Claudin 14): Claudin 14 polymerizes into tight junction fibrils and imparts a key role in selective paracellular permeability. CLDN14 increases transepithelial resistance in inner ear by decreasing the potassium permeability (Lee et al., 2012). CLDN14 expression was detected by Northern blotting in liver and kidney. Its expression was also reported in the sensory epithelium of the organ of Corti (Wilcox et al., 2001).

Pou 3f4 (Brn4, BRN4, otf9): Pou domain, class 3, Transcription Factor 4 (POU3F4), also called as BRAIN-4, BRN4. The POU3F4 gene encodes a transcription factor that restricts the proliferation and lineage potential of neural stem cells (summary by Choi

et al. (2013). Douville et al. (1994) observed the expression of a rat homolog of POU3F4, called RHS2, during embryonic development in the brain, the neural tube, and the otic vesicle at 15.5 and 17.5 days after conception. An X chromosome-linked nonsyndromic mixed deafness, DFN3 (DFNX2; 304400), is reported to be caused by mutations in the BRN4 gene, which encodes a POU transcription factor. However, Brn4-deficient mice created by Minowa et al. (1999) had showed profound deafness.

Studies to date on the molecular basis of sensory biology in bats have focussed on determining patterns of molecular evolution in candidate genes that have known functions in humans and other model organisms. Often these genes have been targeted because of studies that detected phenotypic defects in humans resulting from mutations, as is the case with genes associated with vocalizations (e.g., dysphasia and dyspraxia resulting from mutations in *FOXP2*) and hearing (e.g., non-syndromic deafness resulting from mutations in hearing genes). Recent advances in Transcriptomics and whole genome sequencing allows the genomic comparisons between mammals with different sensory abilities which are to be performed at larger scales and potentially identify novel genomic regions under sensory selection in bats.

Next generation sequencing is making it increasingly possible to identify genetic loci responsible for adaptive evolution in non-model organisms, and the field of adaptation genomics holds great promise (Stapley et al., 2010, Wang and Zhang, 2013).

The importance of alternative splicing in generating proteomic diversity in bats remains largely unknown. Between 40 and 60% of human genes have alternative splice forms, and these comprise one of the major components of functional complexity in the proteomes of humans and other mammals (Modrek and Lee, 2002, Keren et al., 2010).

The importance of splice variants in bats remains largely unknown; however, Li et al. (2008) identified alternative splice forms of the *Prestin* gene in bat brain and cochlea tissue. Such isoforms might be expected to produce a range of functional outcomes from genes associated with audition in bats. Similarly, the importance of other processes contributing toward functional diversity, such as RNA editing (Garrett and Rosenthal, 2012), is not known for bats. Whilst the literature shows clear information

of various genes involved in sensory mechanisms and function, there is a paucity of information of these sensory and brain function genes in bats.

1.5 Transcriptomics

Transcriptomics is a subfield of functional genomics that focuses on gene expression, with specific focus on mRNA (transcripts). The level of gene expression (the amount of mRNA from each gene that is present in a particular cell, tissue, or organism) is a phenotype. Transcriptomics also considered as an "intermediate phenotype" because it is very close to the genotype in the pathway that determines the final phenotype of an organism. The flow of the central DOGMA is

DNA -> mRNA -> protein -> interactions with internal/external environment -> phenotype

The deoxyribonucleic acid (DNA) molecule is the carrier of the genetic information in our cells. Each DNA molecule consists of the four nucleotides Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). It is organized as a double-helix and the two strands of the DNA molecule are complementary to each other. The base-pairing is fixed, where A is complementary to T and G is complementary to C. The formation of the double-helix from two single strand molecules is called hybridization, which is an important step of many of the experimental procedures in transcriptomics. The complete genome is composed of a set of different DNA molecules which are called chromosomes. Eukaryotes, i.e., organisms that have cellular nuclei, store the chromosomes in the nuclei of their cells. For example humans have 23 chromosomes which amount to a total of 3 billion nucleotides and each chromosome appears in two copies in each somatic cell.

An important functional unit on a chromosome is the *gene*. The central dogma in molecular biology is that gene regions are transcribed to ribonucleic acid (RNA), which is then often translated into proteins. In RNA molecules the nucleotide Thymine is replaced by Uracil (U) and is often single stranded compared to DNA. The half-life of RNA molecules is shorter than that of DNA molecules. Because mRNAs correspond to particular genes in the genome, it is often possible to establish a link between a

genotype and an expression phenotype. Large-scale, high-throughput methods have been developed for transcriptomic analysis, and it is now possible to examine the expression level of all genes in the genome in a particular sample, or to detect all of the genes that are expressed differently between two samples.

1.5.1 Gene expression

Exons and introns are the important components in a gene. Exons are the essential parts of the messenger RNA (mRNA) that is the template sequence for the protein, whereas introns mostly have a regulatory rule. The complete gene region is first transcribed into a precursor mRNA (pre-mRNA) molecule that consists of exons and introns (Figure 1.5). The process of transcription is controlled by proteins, called transcription factors, that generally bind in or near the promoter region that resides directly upstream of the gene. Splicing, a process of removal of the intron sequences of the pre-mRNA is initiated by RNA binding proteins called splicing factors. The recognition of intron-exon boundaries is facilitated through the detection of short sequences called splice sites. Polyadenylation is the addition of a poly-A tail, consisting of a series of Adenine nucleotides, to an RNA molecule. The process of polyadenylation is initiated by binding of proteins to a polyadenylation site in an untranslated region of an exon in the pre-mRNA. After splicing and polyadenylation of the pre-mRNA, the final mRNA is obtained. The poly-A tail is important for the transport of the mRNA to the cell cytoplasm and controls further the half-life of the mRNA. The processes of transcription, splicing, and polyadenylation are most likely coupled (Kornblihtt et al., 2004). With the advent of extensive research on genome sequence projects, the other important and fundamental questions arise, ‘the functional role’ of different genes which are involved in their respective cellular process. Abundance of transcript (expression level), regulation of genes and process of interaction between gene and gene products. Change of expression level over growth/development or time (expression profile), Differential gene expression.

Although mRNA is not the ultimate product of a gene, transcription is the first step in gene regulation and information about the transcript levels is needed for understanding

gene regulatory networks. Thus, the new challenge is to identify all genes, their expression patterns and their function.

Transcriptomics or global analysis of gene expression, also called genome-wide expression profiling, is one of the tools that is used to get an understanding of genes and pathways involved in biological processes.

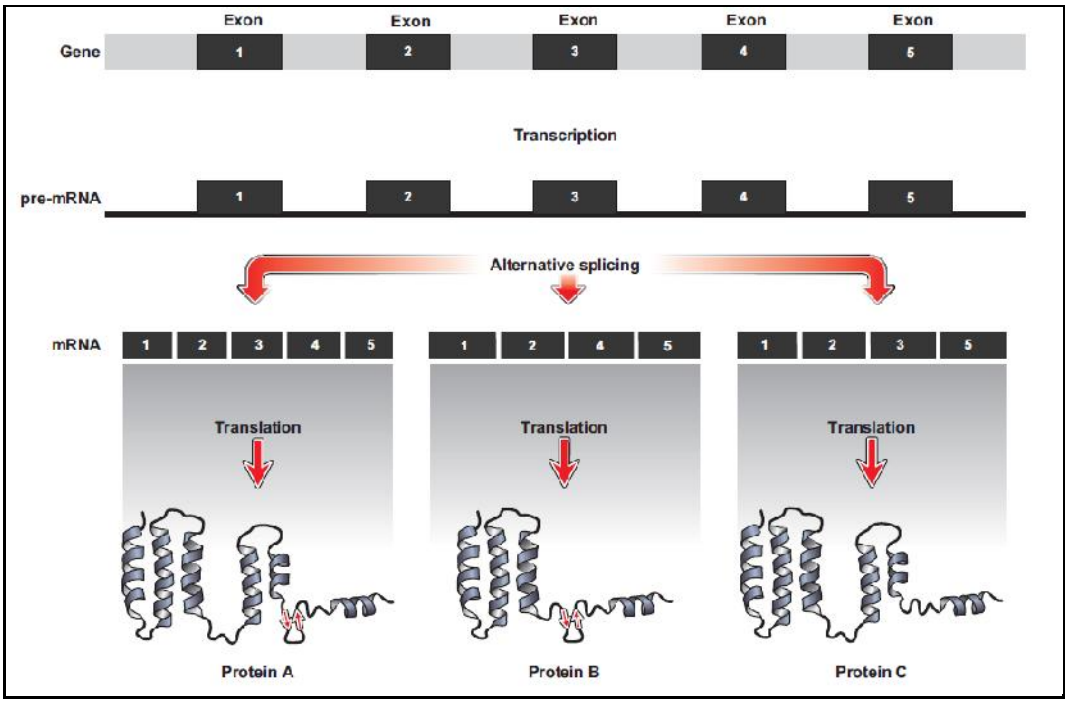


Figure 1.5 All exons of a gene are transcribed into the pre-mRNA. The splicing machinery removes intron sequences and may additionally remove some of the exons through the process of alternative splicing. These mRNAs are then translated into different proteins. Inspired by an illustration in (Source: (Guttmacher and Collins, 2002).

Technologies for genome-wide or high-throughput analysis of gene expression are EST Sequencing, Microarrays, SAGE and RNA – Seq. RNASeq is currently the most ubiquitous and is that used here.

RNA - Seq

The entire set of RNA molecules, including mRNA, rRNA, tRNA, and other noncoding RNA found at cellular, tissue, organ, or whole body levels constitute the transcriptome. Transcriptomes are spatially organized within tissues or even within single cells (nuclear, cytoplasmic synaptic, macromolecular complexes). Transcriptomes can also

be temporally regulated, can change during functional states, or can exhibit circadian dynamics; due to which only a subset of the transcriptome is usually measured at any one time (Montiel et al., 2016). Few years ago, mouse, fruit fly and nematodes were the only genetic model systems available for the study of gene expression. However, with the advancement of parallel sequencing has changed the situation dramatically. Standard practice in molecular ecological studies involving the whole genome or transcriptome analyses has become successful for genetic non model organism (Ellegren et al., 2012, Lamichhaney et al., 2012).

An advantage over other next-generation approaches that reduce the genome to a more manageable size like restriction-site-associated DNA tags (RAD: (Baird et al., 2008)), multiplexed-shotgun genotyping (MSG: (Andolfatto et al., 2011)) or genotyping by-sequencing (GBS: (Elshire et al., 2011)) is that RNA-seq data are directly derived from functional genomic elements, mostly protein-coding genes. Transcriptome sequence thus constitutes a meaningful resource to develop a large number of popular molecular markers such as single-nucleotide polymorphisms and microsatellites (Wolf, 2013).

In situations where full sequencing cannot be afforded, but the application requires the use of many markers (e.g. genome scans), the transcriptome provides a useful functionally relevant subset of the genome. *De novo* assembly of a complete transcriptome without the need for a guiding reference genome is attractive, particularly where the cost and complexity of generating a eukaryote genome is prohibitive (Moreton et al., 2015).

The great advantage of RNA-seq data over other next generation- sequencing applications is that it allows users to investigate differences in gene expression patterns between populations, for example in the context of speciation (Wolf et al., 2010) eco-type-specific adaptation (Lenz et al., 2013) or disease associated mutations and transcriptional changes can be determined by using developmental transcriptome data. Iancu et al. (2012) have demonstrated RNA-Seq data offer better estimation of the gene expression levels and their variability, greatly improving the accuracy of network measurements. Additionally, exon-level data allow the construction and analysis of gene co-splicing networks. It has also been revealed that gene expression levels can be

correlated between distinct tissues (Dobrin et al., 2009). The fields of genomics and transcriptomics maintain their fast development in the improvement of sequencing technologies and associated bioinformatic tools for assembly and annotation. The availability of deep-sequencing techniques that produce hundreds of thousands to millions of reads of variable sizes, at only a fraction of the time, is leading to the accumulation of data in an increasing number of species. Despite these technological advances, sequencing data remain unevenly distributed across phylogenetic groups; for example, 50 of the 72 vertebrate genomes currently available in Ensembl belong to mammals (Welter et al., 2014), most of them originating from the "Mammals Genome Project" (Lindblad-Toh et al., 2011), even though mammals comprise only 8% of the total number of vertebrate species (Lindblad-Toh et al., 2011).

1.5.2 Transcriptome assembly strategies – NGS

Depending on the availability of reference genome, assembly strategies divide into three types viz., A) Reference based strategy, B) *De novo* strategy or C) combined strategy that is the combination of two. A brief overview of these strategies and its associated types of assemblies and software used are given in this section.

A. Reference based strategy or 'ab initio' assembly

This process involves mainly three steps First, RNA sequence reads are aligned to a reference genome using a 'Splice-aware aligner'. It is basically an algorithm which is subdivided in to two types Seed-and-extend aligner and Burrows Wheeler Transform (BWT) aligners. Seed-and-extend - Is an alignment strategy that identifies a seed (k-mer) within the reference genome and then extends these seeds in both the directions to find the best alignment for each read using smith-waterman alignment algorithms.

BWT aligners are optimised to align reads with few errors in them. These are faster in comparison to seed-and-extend methods. The BWT aligner TopHat (Trapnell et al., 2009a), creates a database of all possible combinations of splicing junctions within a locus and then aligns to this database the reads that failed to align to the genome (Table 1.2 and Figure 1.6). Second, overlapping reads from each locus are clustered to build a graph with all possible 'isoforms'. A third, step traverses the graph to resolve the

individual isoforms. Methods used in reference based strategies are Cufflinks (Trapnell et al., 2010), Scripture (Guttman et al., 2010), etc., listed in Table 1.3.

Advantages:

- Transforms larger assembly problems (millions of reads)
- Requires few gigabytes (GB) of RAM to run efficiently on machines and can do parallel computing.
- Sensitive and can assemble transcripts of low abundance.

Disadvantages:

- Depends on the quality of the reference genome taken.
- Probability of generating misassembled or partially assembled transcripts because of many misassemblies and large genome deletions.
- Obviously, not possible without a reference genome.

B. *De novo* assembly

De novo assembly, does not use a reference genome and leverages the redundancy of short-read sequences to find overlaps between reads and assembles them into transcripts. Different types of *de novo* assemblers have been developed (Table 1.3). The Rnnotator (Martin et al., 2010), Multiple-K (Surget-Groba and Montoya-Burgos, 2010) and Trans-Abyss (Robertson et al., 2010) work in a similar way, these assemble the data set multiple times using De Brujin graph-based approach to reconstruct transcripts from a broad range of expression levels. And then post-process the assembly to merge 'contigs' and remove redundancy. Whereas other assemblers e.g. TRINITY and Oases (Grabherr et al., 2011) directly traverse the de Brujin graph to assemble each isoform (Figure 1.7).

Advantages:

- It does not depend on the reference genome, ie for most of the organisms which do not have reference genome or high quality finished genome, *de novo* assembly provide 'initial set of transcripts' which are useful in RNA expression studies.
- It can detect transcripts from unknown exogenous source.

- Does not depend on correct alignment reads to know splice sites (Burset, M. et al., 2000).
- Long introns are of not much concerned in case of *de novo* assembly

C. Combined strategy

Combined strategy is the combination of reference based and *de novo* assembly and should provide a more comprehensive transcriptome. This method gives the advantage of high sensitivity of reference based assemblers while leveraging the ability of *de novo* assemblers to detect novel and trans-spliced transcripts. There are two types Align-then assemble and assemble-then align. Type of the data and the quality of the reference genome, decides the choice of method to be implemented. If a reference genome is available, then the data sets are assembled to a reference genome and followed by 'denovo assembly' (reads that fail to assemble). Otherwise, Transcripts that result from reference based strategy serve as input data to *de novo* assembly if the *de novo* assembler supports both long and short reads (TransAByss and Oases). It also removes unwanted sequences quickly. However, on the other hand, 'Assemble-then align' is important when either quality of reference genome is poor or it is from a closely related species. After the denovo assembly, the contigs are aligned to the reference genome to extend and scaffold contigs (Figure 1.8). The main advantage of this method is errors in the genome assembly do not get propagated into assembled transcripts. In addition, protein sequences can also be used for the alignment step if the sequence similarity at the RNA level is not sufficient. As per the recent findings of Poelchau et al. (2013), Liu et al. (2014) and Huang et al. (2015), the genomic scaffolds and protein sequences serve as a reference to guide transcriptome assembly for non-model organisms.

It results in less fragmented assemblies than *de novo* assembly is a memory intensive, and produces more fragmented assemblies. A new strategy of mapping the *de novo* assembled contigs by Scaffolding using Translation Mapping (STM) was discovered by (Surget-Groba and Montoya-Burgos, 2010).

Table 1.2: A list of splice-aware short read aligners

Aligner	Paired end?	Algorithm type	Finds non-canonical splices?	Independent alignments?	Outputs list of novel splice junctions?	Alignment format	Availability	Reference
Blat	No	Seed and extend	Yes	Yes	NO	PSL	http://users.soe.ucsc.edu/~kent/src/	(Lassmann et al., 2009)
TopHat	Yes	BWT	Yes	NO	Yes	BAM	http://tophat.cbcb.umd.edu/	(Kent, 2002)
GSNAP	Yes	Seed and extend	Yes	Yes	No	SAM	http://research-pub.gene.com/gmap/	(Wang et al., 2010)
SpliceMap	Yes	BWT	No	Yes	Yes	SAM	http://www.stanford.edu/group/wonglab/SpliceMap/	(Trapnell et al., 2009b)
MapSplice	Yes	BWT	Yes	No	Yes	SAM	http://www.netlab.uky.edu/p/bioinfo/MapSplice	(Au et al., 2010)

BAM, binary alignment/map, BWT- Burrows-Wheeler transform; PSL, pat space layout; SAM, sequence alignment/map. (Table taken from (Martin and Wang, 2011)

Table1.3: A comparison of the features of existing software for transcriptome assembly.

Assembler	*De novo?*	Parallelism	Support for paired end reads?	Support for stranded reads?	Support for multiple insert sizes?	Outputs transcript counts?	Software availability	Reference
Cufflinks	No	MP	Yes	Yes	Yes	Yes	http://cufflinks.cbcb.umd.edu//	(Trapnell et al., 2010)
Scripture	No	None	Yes	Yes	Yes	Yes	http://www.broadinstitute.org/software/scripture/	(Guttman et al., 2010)
Rnnotator	Yes	MP	Yes	Yes	Yes	Yes	contact David Gilbert (DEGilbert@lbl.gov)	(Martin et al., 2010)
Trans-ABySS	Yes	MPI	Yes	No	Yes	Yes	http://www.bcgsc.ca/platform/bioinfo/software/trans-abyss	(Robertson et al., 2010)
Oases	Yes	MP	Yes	Yes	Yes	No	http://www.ebi.ac.uk/~zerbino/oases/	-
Trinity	Yes	MP	Yes	Yes	No	Yes	http://trinityrnaseq.sourceforge.net/	(Grabherr et al., 2011)

MP, Multiple processor support (assembler takes advantage of many cores from a single computer); MPI, message-passing interface support (assembler runs in parallel on multiple computers within a cluster). (Table taken from (Martin and Wang, 2011)

It aligns *de novo* assembled contigs to protein sequences from a closely related species. And then re-assembles the contigs according to the alignment results. A reference-based re-assembly, another approach discovered by Poelchau et al. (2013), Liu et al. (2014), which involves two steps, first, it incorporates both protein sequence (like STM) and genomic scaffolds (genomic sequences including 5[l] and 3[l] untranslated regions as well as other non-coding regions.

As per their findings, it improved the *de novo* assembly by increasing contig length and reducing redundancy.

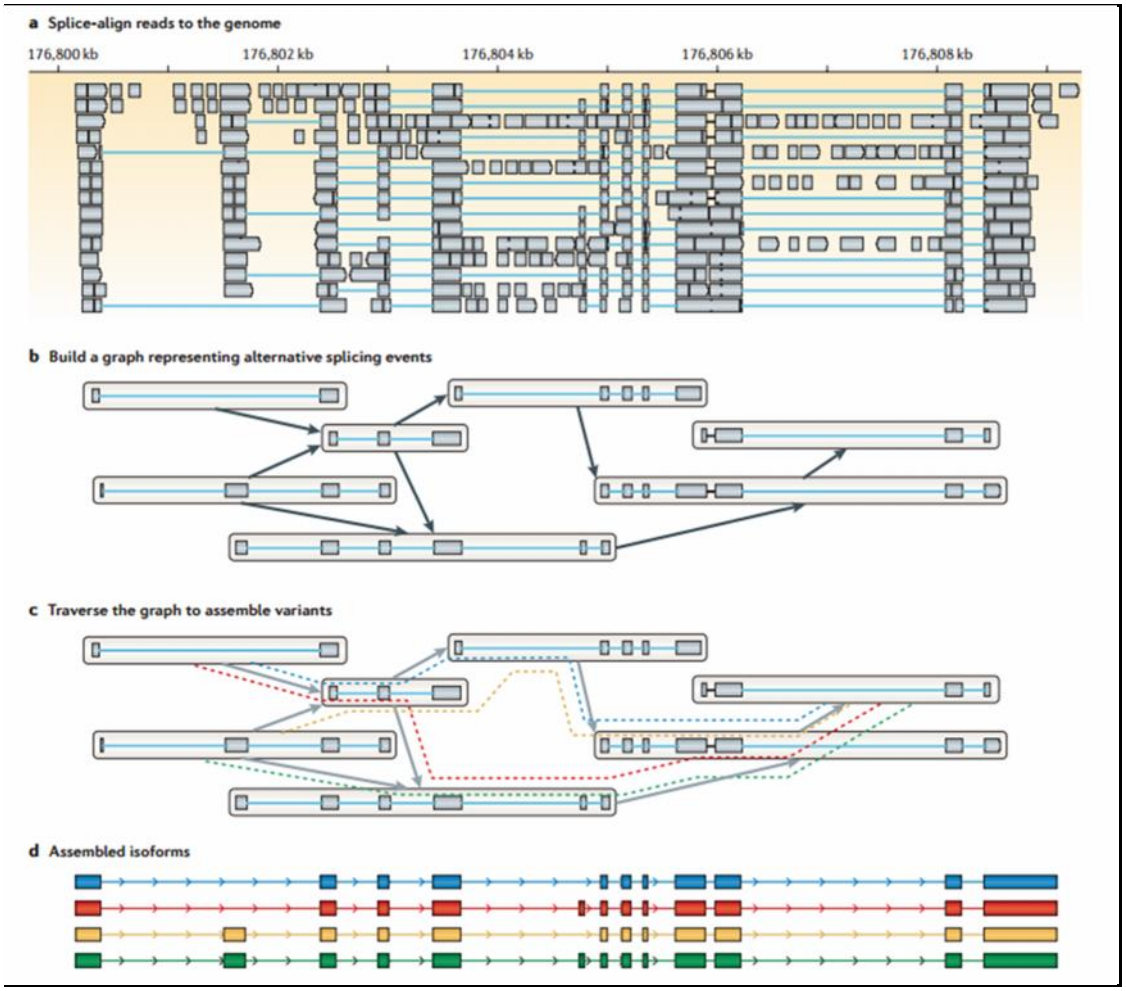


Figure 1.6: Overview of the reference-based transcriptome assembly strategy. The steps of the reference-based transcriptome strategy are shown using an example of a maize gene (GRMZM2G060216). a | Reads (grey) are first splice-aligned to a reference genome. b | A connectivity or splice graph is then constructed to represent all possible isoforms at a locus. c,d | Finally, alternative paths through the graph (blue, red, yellow and green) are followed to join compatible reads together into isoforms. (Taken from (Martin and Wang, 2011).

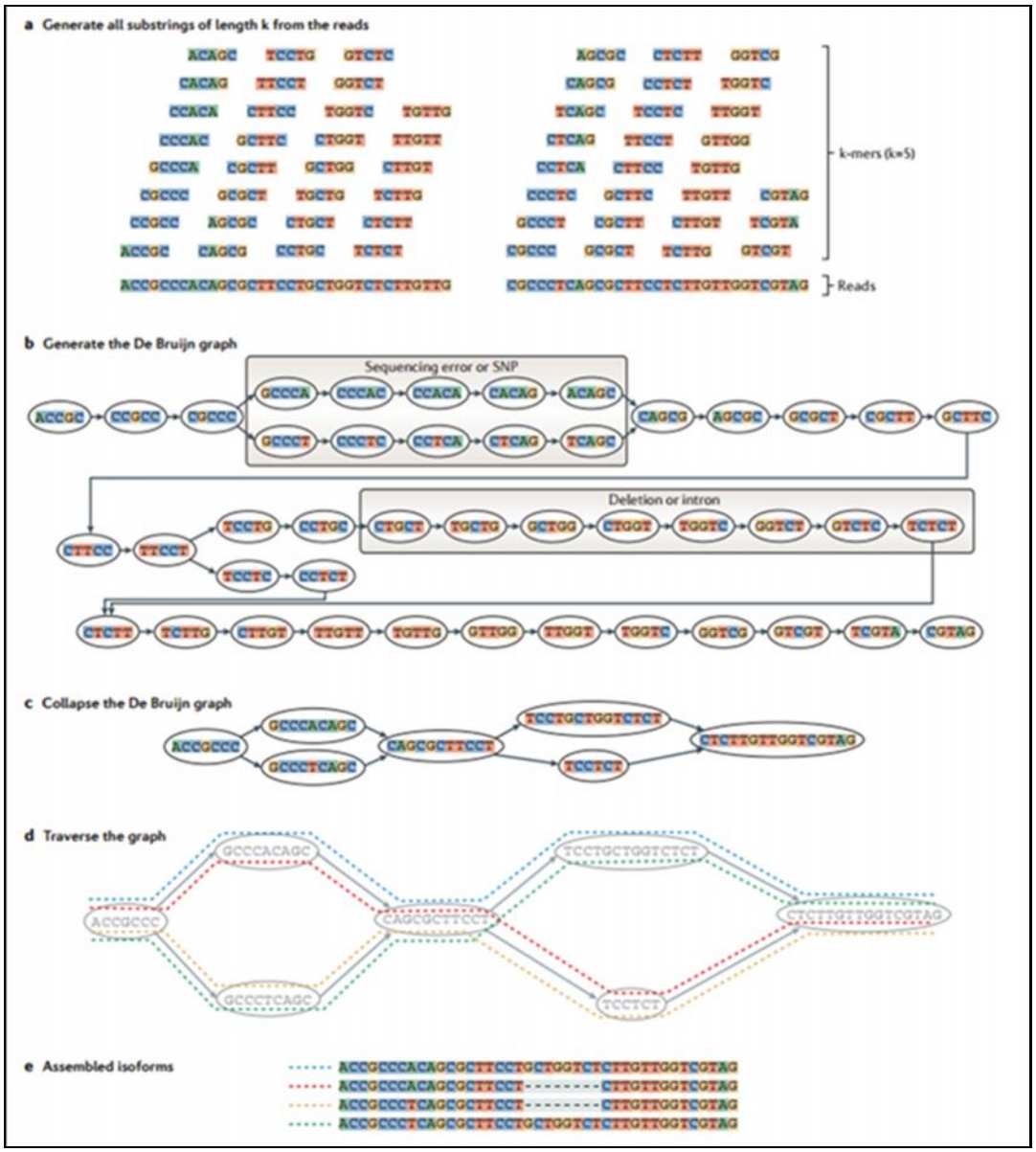


Figure 1.7: Overview of the *de novo* transcriptome assembly strategy. a | All substrings of length k (k-mers) are generated from each read. b | Each unique k-mer is used to represent a node (or vertex) in the De Bruijn graph, and pairs of nodes are connected if shifting a k-mer by one character creates an exact k–1 overlap between the two k-mers. Note that for non-strand-specific RNA sequencing data sets, the reverse complement of each k-mer will also be represented in the graph. Here, a simple example using 5-mers is shown. The example illustrates a SNP or sequencing error (for example, A/T) and an example of an intron or a deletion. Single-nucleotide differences cause 'bubbles' of length k in the De Brujin graph, whereas introns or deletions introduce a shorter path in the graph. c,d | Chains of adjacent nodes in the graph are collapsed into a single node when the first node has an out degree of one and the second node has an in degree of one. Last, as in the reference-based approach, four alternative paths (blue, red, yellow and green) through the graph are chosen. e | The isoforms are then assembled. . (Taken from (Martin and Wang, 2011).

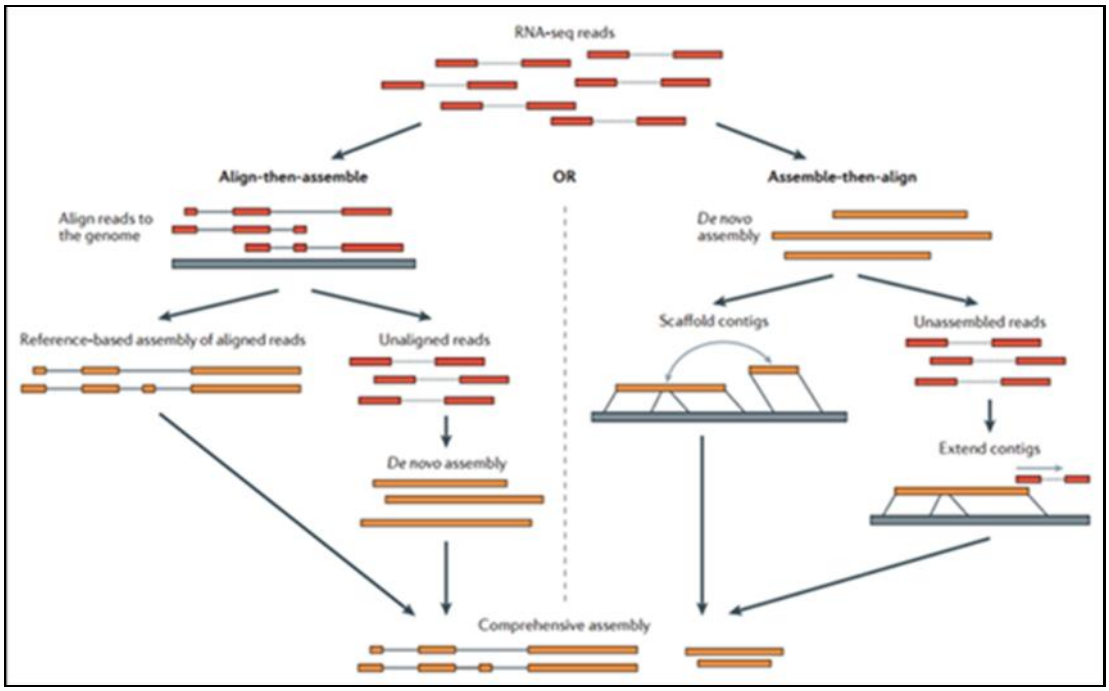


Figure 1.8: Alternative approaches for combined transcriptome assembly. The left choice depicts the align-then-assemble strategy, in which reference-based assembly is followed by *de novo* assembly of reads that failed to align to the genome. The right choice depicts the assemble-then-align strategy, in which the reads are first *de novo* assembled and then scaffolded and extended using a reference genome. RNA sequencing (RNA-seq) reads are shown in red, and assembled transcripts are shown in orange. (Taken from (Martin and Wang, 2011).

Another approach, Transcriptome Post Scaffolding (Trans PS) put forward by Liu et al. (2014), also utilizes protein sequences from a closely related species in order to reassemble the *de novo* contigs according to how they exactly align to the reference protein sequence. In contrast to reference based re-assembly, Trans PS merges contigs with non-overlapping alignments to the same reference protein. However, Trans PS does not utilize genomic sequences from a closely related species. As per the findings of Liu et al. (2014), Trans PS reduces the number of contigs to as few as 10 % of the number of *de novo* assembled contigs, eliminating the contigs that are redundant and increasing the coverage for the reference protein sequences relative to *de novo* assembly.

However, with the advent of different assembly strategies, the optimal strategy of transcriptome assembly will depend upon the downstream analysis and computational resources available. In this study we employed *de novo* assembly with a designed pipeline (see Materials and methods section). Transcriptomic comparison using

unbiased in silico methods are used in this study to derive and identify novel genes which are categorized according to their abundance of transcript, gene functional enrichment and expression profile, networks and pathway interactions, phylogenetic position.

1.6 Orthology and Inferring Functions

1.6.1 Homologs and orthologs

Essential in this study is the comparison of gene sets to determine similar genes between species. Homology in molecular biology refers to a common ancestry. In general, homologous genes are identified through the assessment of the statistical significance of sequence similarity. With the advent of specific species radiation, homologous relations define orthologs – 'equivalent' genes in different species descended from a single ancestral gene (Fitch, 1970, Koonin, 2005, Sonnhammer and Koonin, 2002). Homologs which originate following gene duplications are called paralogs. Paralogy can exist between genes in different species, since gene duplication events occur both before and after speciation. The term 'inparalogs' indicate paralogs that arose through a gene duplication event after speciation, while 'outparalogs' arise following a gene duplication preceding speciation (Remm et al., 2001, O'Brien et al., 2004, Sonnhammer and Koonin, 2002). Clustering inparalogs together enhances the proper identification of both one-to-one and many-to many ortholog cases (Figure 1.9).

1.6.2 Methods to quantify orthologs

There are two major approaches/strategies for orthology delineations viz; Algorithm based and Phylogeny-based methods. In addition to this 'global' or 'local' alignment are used to quantify orthology. In global alignment, the whole sequence is compared with the other entire sequences. On the other hand, local method uses a subset of sequences and aligns it to subset of other sequences. However, global alignment is appropriate for sequences of protein-coding genes and for short stretches of genomic sequences in contrast, local alignment is ideal for large databases and uncovered regions of homology that are related by descent between two otherwise diverse sequences.

A. Algorithm based

Algorithms that cluster all-against-all pair wise sequence comparisons, which includes BLAST (Altschul et al., 1990), COG (Clusters of Orthologous Groups; (Tatusov et al., 2003)), INPARANOID (Remm et al., 2001, O'Brien et al., 2005), RBH (Reciprocal Blast Hits; (Hirsch et al., 2001)) and RSD (Reciprocal Smallest Distance Algorithm; (Wall et al., 2003) to determine gene orthology on a genomic scale.

As per the findings of Koski and Golding (2001), BLAST is not a reliable approach to determine orthology relative to evolutionary history. However, RBH has limitations where multi-gene families with numerous gene duplication events, result in closely related paralogs. As a result, the number of ortholog pairs identified from two genomes using RBH may be an underestimate.

Wall et al. (2003) identified RSD algorithm which uses character-based maximum likelihood estimation of evolutionary distances to identify orthologs between genomes, instead of depending solely on the BLAST scores. This improved the sensitivity of reciprocal genome queries and the number of ortholog gene increased significantly when compared with that identified using RBH (Wall et al., 2003).

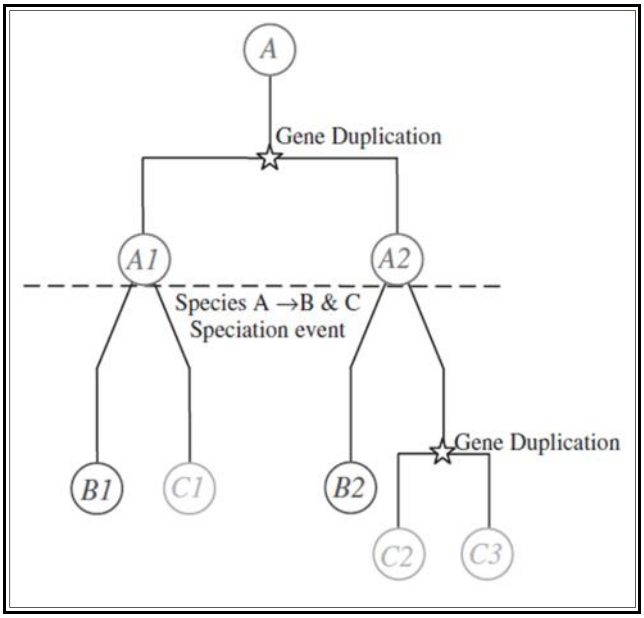


Figure 1.9: A hypothetical gene tree and the resulting Inparanoid clusters are shown to illustrate inparalog (and thus co-ortholog) and outparalog assignments. Protein A in an ancestral species 'A' undergoes a gene duplication. A speciation event occurs which gives rise to the two lineages leading to species 'B' and 'C'. In the C genome the genes C2 and C3 are inparalogs since their

gene duplication occurred after speciation; they are co-orthologous to the B2 gene (one common ancestral protein upon speciation). B1 is an outparalog of the C2 and C3 genes, as are B1 of B2 (duplication and divergence prior to speciation). Taken from O'Brien et al. (2005).

The algorithms discussed above are good in identifying the one-to-one orthology relationships between two genomes.

Two widely and prominent algorithms are built on all-against-all pairwise comparisons, INPARANOID and COG, identify both one-to-many and many-to-many orthologous relationships. COG identifies groups of sequences from at least three species that most likely represent orthologs (Tatusov et al., 2000) and in addition, it also takes advantage of structural data when available (Tatusov et al., 2000). However, in terms of algorithm perspective, INPARANOID can only identify and differentiate between inparalogs (true orthologs by definition) and outparalogs. In addition, it is the only algorithm which can identify orthologs from two genomes simultaneously (Remm et al., 2001).

Ortho DB

OrthoDB is a hierarchical catalogue of orthologs (http://www.orthodb.org) useful in quantifying Orthologus group characteristics that describe evolutionary properties such as gene duplications, or losses and rates of sequence divergence, and provides a basis for the assessment of the confidence of gene function inferences. It basically depends on clustering of best-reciprocal-hits (BRHs) between genes from each species pair, determined from all-against-all Smith–Waterman protein sequence comparisons now using SWIPE (Rognes, 2011). The clustering procedure as described by Li and Godzik (2006) only considers the longest transcript per gene, and only the longest of all gene copies in a single genome with over 97% amino acid identity (using CD-HIT).

Ortho MCL

Recently updated approach by Li et al. (2016), for the construction of orthologous groups from eukaryotic genomes-identifying both orthologs between species and functionally redundant (“recent”) paralogs within species, and grouping these together using the MCL algorithm to cluster graph. The OrthoMCL procedure starts with all-against-all BLASTP (Figure 1.10) comparisons of a set of protein sequences from

genomes of interest. Putative orthologous relationships are identified between pairs of genomes by reciprocal best similarity pairs. For each putative ortholog, probable "recent" paralogs are identified as sequences within the same genome that are (reciprocally) more similar to each other than either is to any sequence from another genome. This method, mainly comprises of a graph representation of sequence relationships which then divided into subgraphs using the MCL algorithm.

INPARANOID

InParanoid is an algorithm designed with the aim to generate ortholog groups that include all inparalogs excluding outparalogs (no outparalogs). It is a graph-based approach that starts with an iterative all-vs-all Basic Local Alignment Search Tool (BLAST) comparison of all protein sequences in two species and then applies a number of clustering rules to build ortholog groups (Remm et al., 2001).

It operates on two species at a time which is an advantage as the orthology assignments are not affected by other species and the orthology query is in pairwise, either one-to-one species queries or one-to-many, both are supported in this programme. The Inparanoid software is one of the few downloadable fully automatic ortholog-inference programs and is ranked best by the online orthology benchmarks at http://orthology.benchmarkservice.org, in terms of both coverage and accuracy.

In a comparison study of phenotype-based gene groups between species, and (Woods et al., 2013) identified significant enrichment of orthologs by Inparanoid. Novel candidate disease genes are predicted by orthologues phenotypes (showing disease phenotypes in human with other phenotypes in chicken, zebrafish, *Escherichia coli* and *Caenorhabditis elegans*), in their study.

Database called FSRD of 1985 fungal stress proteins was built by (Karanyi et al., 2013) using Inparanoid and identified 28 species which includes human and plant pathogens and some other fungi.

Inparanoid BLAST is employed in the present study (usage and commands are explained more in the methodology chapter).

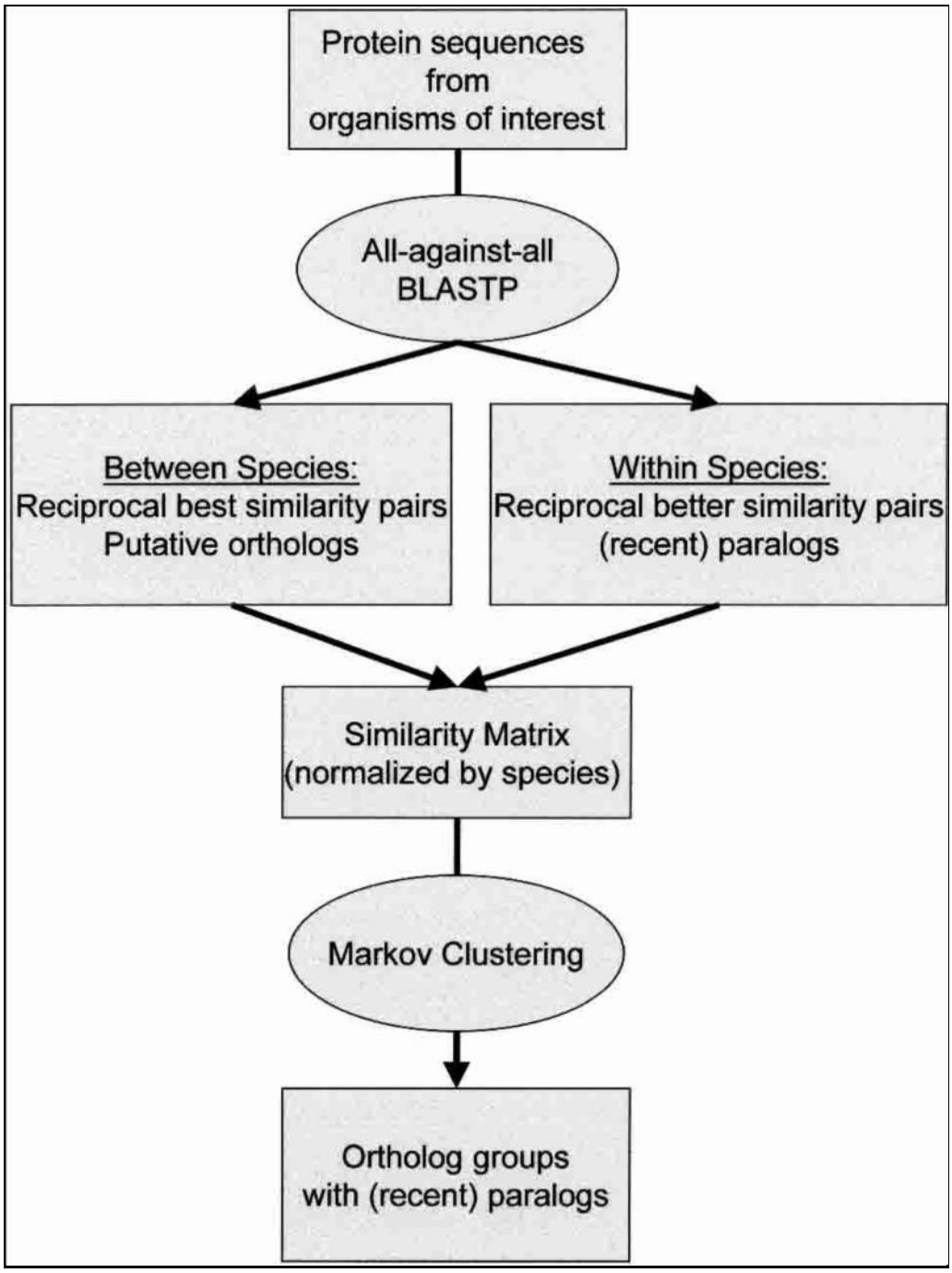


Figure 1.10 Flow chart of Ortho MCL algorithm for clustering orthologous proteins (Taken from (Li et al., 2003).

B. Phylogeny Based

Phylogeny-based approaches have more parameters and therefore yield better accuracy given sufficient data, however, they depend on the quality of multiple sequence alignments (Huerta-Cepas et al., 2011, Vilella et al., 2009, Penel et al., 2009). These

methods first define homologous gene families, build gene trees for each family, and then explicitly or implicitly reconcile them with the species tree often employing assumptions on rates of gene losses and duplications.

Tree building methods are an alternative approach to the identification of Orthologus gene regions (Chiu et al., 2006). Gene family trees are constructed to explore the evolutionary history of gene family members. Complete genomes from multiple species can be included in the analysis. The resulting trees are screened to identify Orthologus groups (clades, clusters, nodes). These methods include MetaPhOrs (Pryszcz et al., 2010) Phylome DB (Huerta-Cepas et al., 2013), Orthology ID (Chiu et al., 2006).

Chiu et al. (2006), developed Ortholog ID, a Web-based tool that fully automates the phylogenetic approach for the identification of orthologs within a character-based parsimony framework. Ortholog ID automatically searches the local database of completely sequenced genomes and clusters gene sequences into putative gene families, performs sequence alignments using multiple sets of alignment parameters and clean or sift alignment-ambiguous regions (Remm et al., 2001). This gives the potential advantage for error that missing 'inparalogs', if partially sequenced genomes were used to construct gene trees (detects the missing inparalogs, true orthologs.

Orthology ID along with INPARANOID can reliably differentiate between in-paralogs and out-paralogs. Orthology ID uses phylogenetic information from outgroup sequences while INPARANOID does so by using pairwise sequence comparisons within and between genomes (as discussed earlier). Another advantage of Ortholog ID over RBH, RSD and INPARANOID is that it is not restricted by reciprocal genome queries, and can therefore determine orthology relationships between more than two complete genomes simultaneously. In addition to it, Ortholog ID identifies diagnostic characters that define all nodes on the phylogenetic trees it generates, including the tree nodes that group Orthologus gene sets by using CAOS algorithm build by (Sarkar et al., 2002).

1.6.3 Importance of orthologs in genomics and phylogenomics

Studies on the organization of Orthologus clusters in a comparative approach exemplifies in understanding the genomic structure and gene function. This in turn acts

as a pilot source for taxonomic classification and phylogenetic studies of organisms, thereby investigating mechanisms underlying the molecular evolution of genes and genomes (Henikoff et al., 1997). The tools for establishing homologies between genes or their products are becoming increasingly important to transfer knowledge from well-studied model organisms to other organisms (Altenhoff and Dessimoz, 2012).

1.6.4 Ortholog inferring functions

Taking orthology as the basis to infer functional annotation between orthologs can accelerate the annotation process and is widely adopted. The Gene Ontology (GO) database (Ashburner et al., 2000) is a major collection possessing consistent descriptions of gene products from different databases. The GO database comprises three structured components (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. Associated with the elaborate label of genes' functionality from the GO database, orthologs offer the ability to accurately convey annotation across organisms. Several ortholog databases are now available online, which consider orthology from the perspective of protein sequence individually, including Homolo Gene (http://www.ncbi.nlm.nih.gov/homologene), Ensembl Compara (Vilella et al., 2009), Inparanoid (Remm et al., 2001, O'Brien et al., 2005), Roundup (DeLuca et al., 2006), OrthoMCL (Li et al., 2003, Chen et al., 2006), and OrthoDB (Kriventseva et al., 2008). There exist ambiguous and incomplete ortholog assignments because of the interference mediated by alternative splicing (AS), the orthologous annotation might be inconsistent across databases without considering alternative splice variants collectively (Ho et al., 2010).

Based on orthology, Gene-Oriented Ortholog Database (GOOD, Ho et al. (2010) additionally comprises functional annotation from the GO database. This information can benefit species which lack of functional. That is, functional annotation for a given species can be predicted using the GO annotation of orthologous genes from other well-annotated species. In addition, it represents graphical view of connections among them. Using graphs, GOOD simultaneously displays all relationships among nodes which are in the paths from the annotated GO term to the root node. That is, graphs of GO terms

can further provide the comprehensive topology for users to reveal the divergence among related GO terms in different GO layers.

GOOD includes functional annotation from the GO database, (ontologies: biological process, cellular component and molecular function) infers to the target species based on orthology.

Functional annotation in the present study was carried out by GO database and interacting proteins and pathway is done by KEGG which gives an illustrative idea of known and interacting proteins. In addition, regulation of gene or protein in its particular pathway can be visualized by up and down regulation (details discussed in methodology section).

2. MATERIALS AND METHODS

Figure 2.1 depicts a generic roadmap for experimental design and analysis using NGS sequencing and computational analysis.

2.1.1 Sample collection

Heads of four bats were taken and shock-frozen with liquid nitrogen was done immediately to prevent fragmentation and eventual loss of RNA due RNase activity and stored in -80º C. Each frozen sample was thawed and the whole brain dissected. Each brain was stored in RNA later® (Ambion by Life technologies) at 4º C till its further process.

2.1.2 Contamination

Special precautions were taken to avoid any form of contamination (RNase-free tubes, RNase zap). To exclude contamination with DNA, treatment with high quality DNase (DNase I) was done in extracting the RNA.

2.2 RNA Extraction and Quality Assessment

Extraction followed that of Rio et al. (2010), brain tissues were removed from RNA later and processed for the extraction of RNA by disruption in Trizol reagent (Ambion). Samples were further passed through a 22 gauge needle using 3 ml syringe. Extraction was with Trizol reagent.

Purity of total RNA was checked by estimating the absorbance ratio at 260/280 and 260/230, and RNA integrity number (RIN) using Agilent Bioanalyzer 2100 System and associated RNA 6000 Nano kit (Agilent) at school of Veterinary Medicine and Science, University of Nottingham, Nottingham.

High quality RNA was processed for high throughput sequencing (2 x 150 bp paired end Illumina) by Deep - Seq, Centre for Genetics and Genomics, The University of Nottingham, Queen's Medical Centre, Nottingham.

Four replicates named LC5, LC6, LC7, LC8 were processed. Raw sequence data (fastq format) was processed by the Advanced Data Analysis Centre (ADAC), University of Nottingham, Nottingham, UK.

2.3 Analysis of RNA-Seq Data

2.3.1 Quality control

Raw data was processed to remove adapters using cutadapt v1.5 phred value < 10 (paired mode) (Martin, 2011).

```
Cutadapt-q10-m 50-b-B fast_1 fastq_2
```

De-duplication of duplicated sequence reads was by PRINSEQ.

```
Prinseq-lite.pl-derep 1 -fastq xxx.R1.fastq-fastq2
   xxx.R2fastq-out_good short_name-out_bad null
```

2.4 Transcriptome Characterization (Assembly)

We performed strand-specific *de novo* transcriptome assembly using Trinity r20140413p1 (Grabherr et al., 2011). The detailed pipeline of study is depicted in the figure 2.2.

2.4.1 Transrate

For each replicate, the generated assembly was filtered with Trans Rate. To generate a nonredundant set of contigs, transcripts with a relative expression value (TPM) < 0.5 were removed and CD-HIT-EST v4.6 (Li and Godzik, 2006) was used to merge

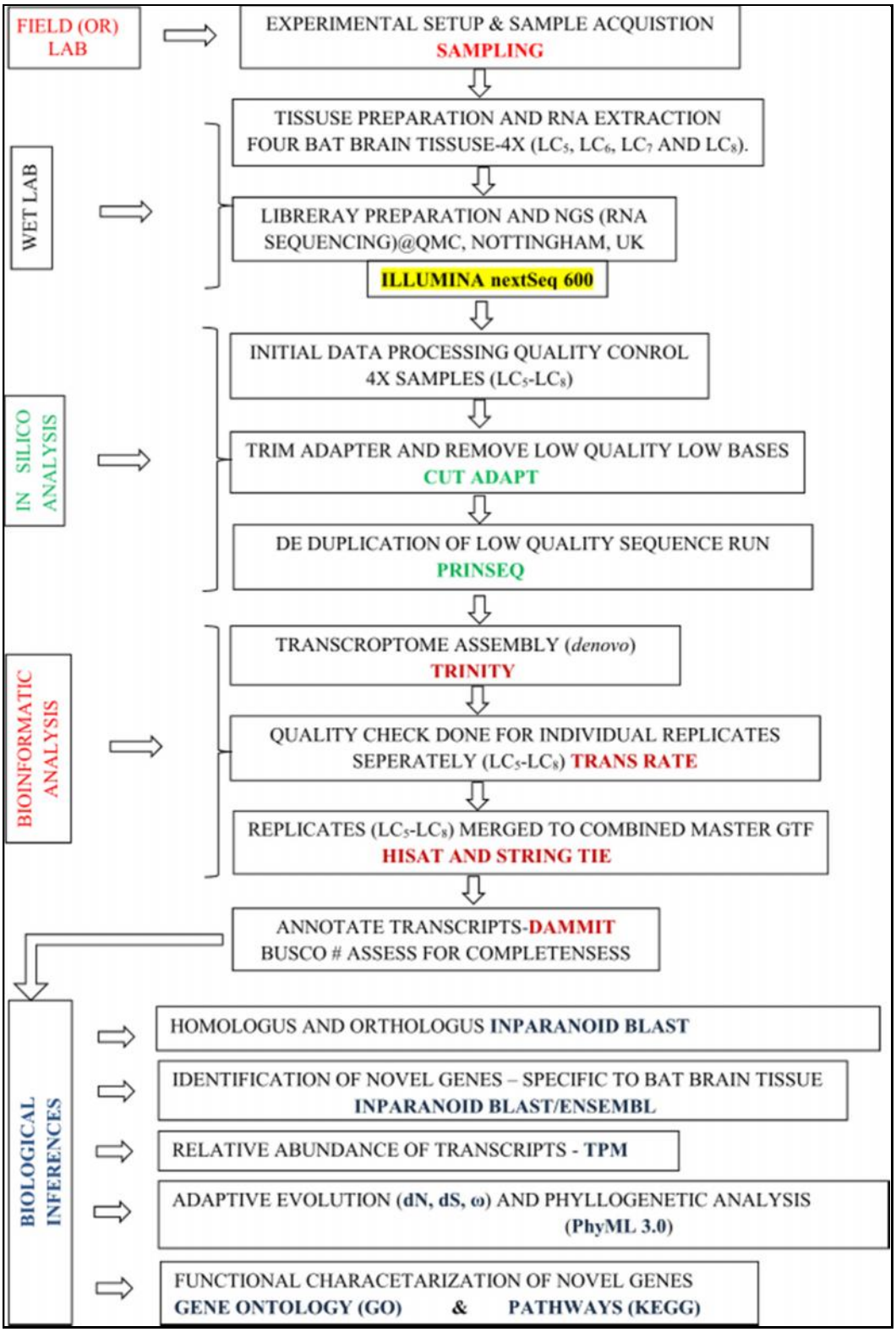


Figure 2.1: Flow chart of a typical RNA experiment with outlines of the study (followed in this thesis).

all sequences with identity threshold of 0.99, which were present in a minimum of 3 of 4 replicates cd-hit-est.cluster.size.filter.pl-f fasta (from cd-hit-est)-cluster file –o name.fa-m2.

2.4.2 HISAT and StringTie

Reads were mapped to the consensus transcriptome HISAT to produce a master GTF file (Kim et al., 2015) . This was processed through StringTie, to quantitate expression (Figure 2.2).

2.4.3 Dammit-BUSCO (Annotation)

The potential transcripts obtained in a GTF file were annotated using dammit v0.3.

Completeness was determined using Benchmarking Universal Single-Copy Orthologs (BUSCO), published by (Simão et al., 2015).

```
dammit annotate --database-dir /path/to/dbs --busco-group
                vertebrata --n_thread.
```

2.5 Identification of Novel Genes (Inparanoid-BLAST) Specific to Bat Brain Tissue

An all-vs-all Basic Local Alignment Search Tool (tBLASTn) with cut off e value < $1e^{-6}$ was used to identify sequences unique to the *de novo* assembly.

Bat specific genes identified were checked for the Brain functions and its associated proteins by IPA pathways (Ohlsson Teague et al., 2009). Genes of interest involved in Echolocation (Vocalization and hearing) were identified by Ensembl IDs (Ensembl.org).

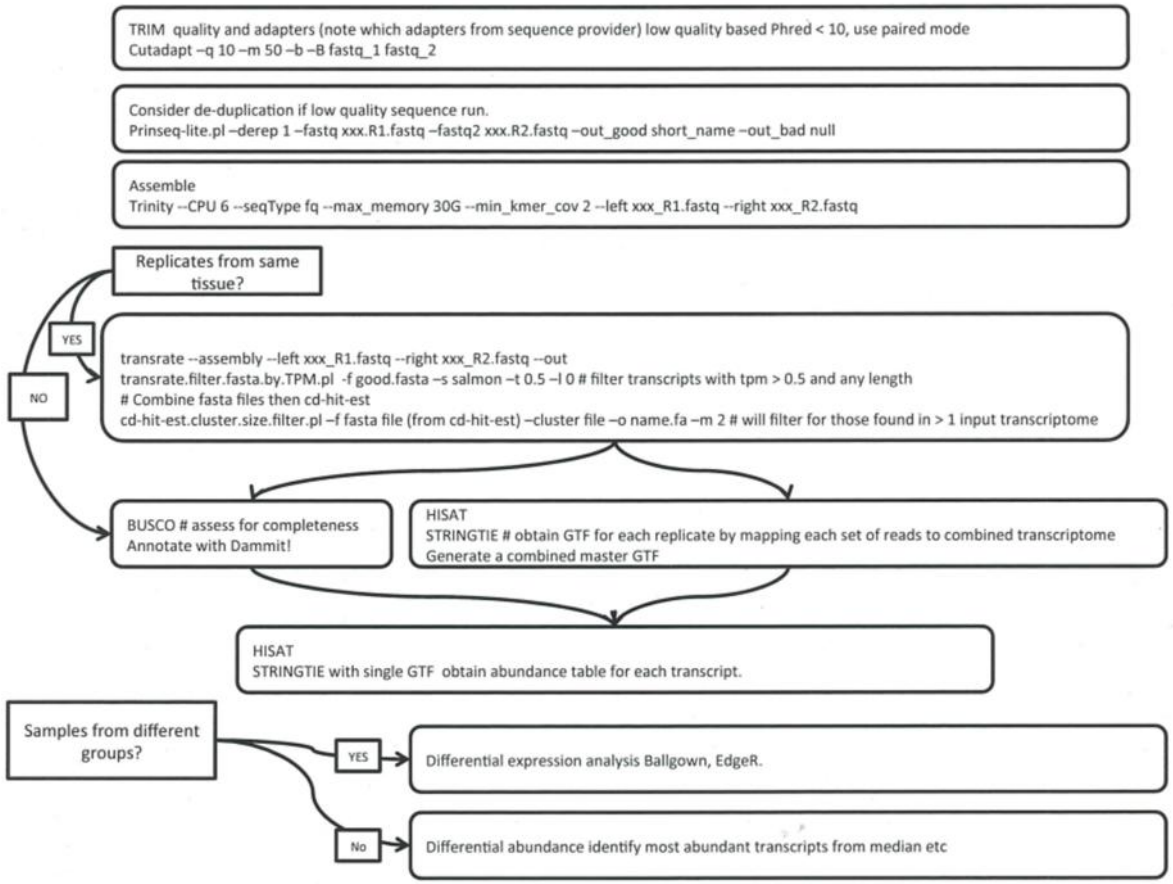


Figure 2.2 Pipeline of *de novo* transcriptome assembly.

2.6 Highly Expressed Genes – Transcript Abundance (TPM)

To identify highly expressed genes the median expression for each replicate was identified and those transcripts above 5 x the median were recorded.

2.7 Adaptive Evolution

The Ratio of synonymous (dS), to nonsynonymous substitution (dN) rates were obtained from Ensembl.

2.8 Phylogenetic Tree Construction

Alignment of the protein sequences was performed using the multiple alignment tool MUSCLE (Edgar, 2004). Phylogenetic trees were generated using PhyML V3.0, (Guindon and Gascuel, 2003). With the following parameters, substitution model:WAG (Whelan and Goldman, 2001) and tree refinement using NNI (Gascuel, 1997). Resulting trees were visualised using FigTree V 3.2.

2.9 Functional Characterization of Novel Genes

Unique bat genes identified by *de novo*, based on highly expression (TPM > 5 X Median) were tested for the significant enrichment of specific Gene Ontology (GO) terms for Biological Processes, Molecular Function (MF) & Cellular Processes (CP) and KEGG pathways using the Bioconductor GO seq package (Young et al., 2010, Alexa and Rahnenfuhrer, 2010). The gage package was used for gene set (enrichment or GSEA) or pathway analysis (Generally Applicable Gene-set Enrichment (GAGE), (Luo et al., 2009).

3. RESULTS

3.1 RNA Sequence Data Analysis

RNA concentration was > 200ng/µl (having 260/280, 260/230 ratios ≥ 1.8) with RIN values as in (Table 3.1). The first step of data analysis was to check the quality of receiving sequencing data by running the FastQC program which generates an output file of raw sequencing data with quality scores across all bases.

Table 3.1: Assessment of RNA quality used in the study

Sample No.	RIN
LC5	6.9
LC6	7.0
LC7	7.1
LC8	6.8

Initial raw reads obtained after deep sequencing were listed in table 3.2. Total number of reads reduced after filtering and thus helps in removing the low quality reads. All three samples showed consistent read quality that decreased steadily over the length of the reads, but in the replicate LC7 it was comparatively less. After trimming and quality check the raw sequence data of four Brain samples (UoN Bat) were assembled by using program 'TRINITY'.

Table 3.2: Data showing raw reads of the study

Method/Step	Description	Number of Raw reads
Initial Raw Reads	LC5	45,781,449
	LC6	44,569,154
	LC7	36,006,885
	LC8	44,936,830

3.2 *De novo* Assembly

The trimmed raw sequencing data was assembled separately for 4 brain replicates (LC5, LC6, LC7, LC8). These samples were assembled by the three independent software modules Inchworm, Chrysalis and Butterfly. Quality & assembly statistics are discussed as per the Table 3.3.

De novo assembly by the TRINITY software produced 266918, 301125, 265488, 352889 transcripts for the four sample of Bat brain LC5, LC6, LC7 and LC8, respectively (Table 3.3). Assembled transcripts after TRINTY were checked for quality of the assembly by 'TransRate' (as per the pipeline of the study Figure 2.2). The transcripts were filtered further, which identified 136758, 161299, 112538, 145015 for the four sample of Bat brain LC5, LC6, LC7 and LC8, respectively (Table 3.3). The quality assessed transcripts were further filtered for TPM > 0.5, and merged the file, and thus a total of 50,856 final transcripts were generated. The filtered transcripts were checked for the cluster of transcripts i.e. the genes in 3 assemblies and was done by using cd hit-est (as mentioned in the methods section) and 45,359 transcripts were retained. Using HISAT and Stringtie a final dataset of 43,800 transcripts named UoNBat was generated.

3.3 Homologous and Orthologus Search

The UoN Bat transcripts were used to identify orthologues in Ensembl bat (*Myotis lucifugus*), Human and Mouse using Inparanoid. Counts and types of orthologs were generated (Tables 3.4 and 3.5) along with histogram plots (Figure 3.1 and 3.2). Count data indicates that, most of the orthologues are in the 1:1 category in all the three comparisons, (6591, 5571 and 5611, for Bat human and mouse, respectively, Table 3.4). A greater number and percentage of transcripts are 1:1 orthologs between UoN Bat and Ensembl bat as would be expected given their recent evolutionary divergence.

Table 3.3: Bat Assembly statistics as per the pipeline of the study.

S No.	Method/ Step	Description	Number of Transcripts
1	TRINITY	LC5.trinity_out/LC5.fixed.fasta	**266918**
		LC6.trinity_out/LC6.fixed.fasta	**301125**
		LC7.trinity_out/LC7.fixed.fasta	**265488**
		LC8.trinity_out/LC8.fixed.fasta	**352889**
2	TRANSRATE	LC5.transrate_out/LC5.fixed/good.LC5.fixed.fasta	**136758**
		LC6.transrate_out/LC6.fixed/good.LC6.fixed.fasta	**161299**
		LC7.transrate_out/LC7.fixed/good.LC7.fixed.fasta	**112538**
		LC8.transrate_out/LC8.fixed/good.LC8.fixed.fasta	**145015**
3	Filter for TPM > 0.5	Combined transcripts found in 3 assemblies (UoN.Ppi.merged.GTF.fa)	**50856**
4	Clusters of transcripts	Genes in 3assemblies.cd.hit.combined.renamed.tpm0.5.minlength0.fa	**45359**
5	HISAT and STRINGTIE	Map reads with HISAT and output Transcripts with measured abundance by StringTie	
		LC5.ST/LC5.abundances	**43800**
		LC6.ST/LC6.abundances	**43800**
		LC7.ST/LC7.abundances	**43800**
		LC8.ST/LC8.abundances	**43800**

Table 3.4: Count Data orthologues of UoN bat to its type of orthologues in Ensembl Bat, Human and Mouse.

Between Orthologues	**Count data and Type of orthologues**			
	1:1	1:m	m:1	m:m
UoN Bat and *Myotis lucifugus*	6591	2634	108	48
UoN Bat and *Homo sapiens*	5571	2048	3413	1735
UoN Bat and Mouse	5611	2095	2279	1216

Table 3.5: Percent data orthologues of UoN bat to its type of orthologues in Ensembl Bat, Human and Mouse.

Between Orthologues	**Percent data and Type of orthologues**			
	1:1	1:m	m:1	m:m
UoN Bat and *Myotis lucifugus*	70.25	28.07	1.15	0.51
UoN Bat and *Homo sapiens*	43.63	16.04	26.73	13.58
UoN Bat and Mouse	50.09	18.70	20.34	10.85

Euler plots were generated between orthologs of Myotis, Human and Mouse vs unique orthologs of UoN Bat (Figure 3.3). These illustrate the number of predicted open reading frames (ORFs) which are unique and shared for each comparison.

The number of ORFs with a clear orthologue between bat species is 9,381 (Figure 3.3 A). 12,767 were identified between UoN bat and Human (Figure 3.3 B) and 11,338 between UoN bat and Mouse (Figure 3.3 C).

Table 3.6 depicts the number of genes which are orthologues in Mouse and human, Ensembl bat and unique to UoNbat which were 1631, 1630 and 190, respectively.

The number of genes were found to be 468 (2.2%), 3261 (15.5%) and 1168 (5.5%) in Ensembl bat, human and mouse, respectively (Figure 3.4). The combined core set of orthologues found in Bat, Mouse and Human is 12,642 (60.1%).

3.3.1 Long-term Potentiation

As an indicator of the completeness of the UoNBat transcriptome we looked for completeness of key pathways associated with brain function. Long-term potentiation (LTP) is a form of activity dependent plasticity which results in a persistent enhancement of synaptic transmission. LTP and Long –term depression (LTD) have another potentially important role in modern neuroscience, and that is the possibility that they may be exploited to treat disorder and disease in the human central nervous system (CNS). (Timothy V.P. Bliss, TVP and Cooke, JSF., 2011).

Figure 3.5 depicts the involvement and expression of wide range of receptors, enzymes and other intracellular signaling molecules in the induction, expression and maintenance of LTP and LTD. Receptors include NMDAR – Voltage dependent subtype Glutamate receptor which allows the permeation of calcium and other cations. Ca^{2+} binds to Ca^{2+}/calmodulin which, in turn activates numerous kinases and phosphatases, including CaMKII, PKC and Calcineurin (PP2B) directly and PKA and PP1 indirectly. C-AMP-dependent signaling cascade involving PKA, mitogen activated protein kinases (MAPK) and the transcription factor cAMP-responsive element binding protein (CREB) are involved in the novel gene expression.

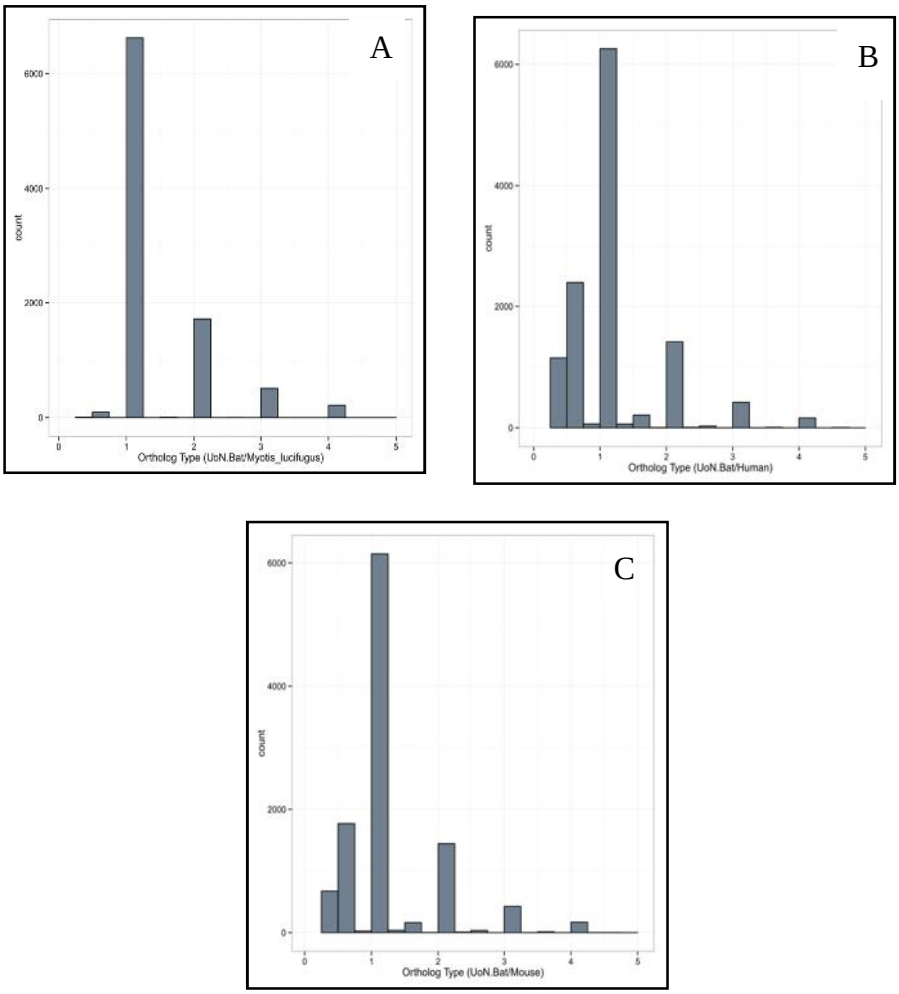


Figure 3.1 A, B, C. Graph illustrating orthology count and type (1:1,1:m,m:1,m:m). Orthologus counts of UoN bat are compared in A) Myotis lucifugus (Ensembl bat), B) Human, and c) Mouse ortholog. The number of orthologues are more in 1:1 orthology type in all three in comparison with 1:m, m:1, m:m, respectively.

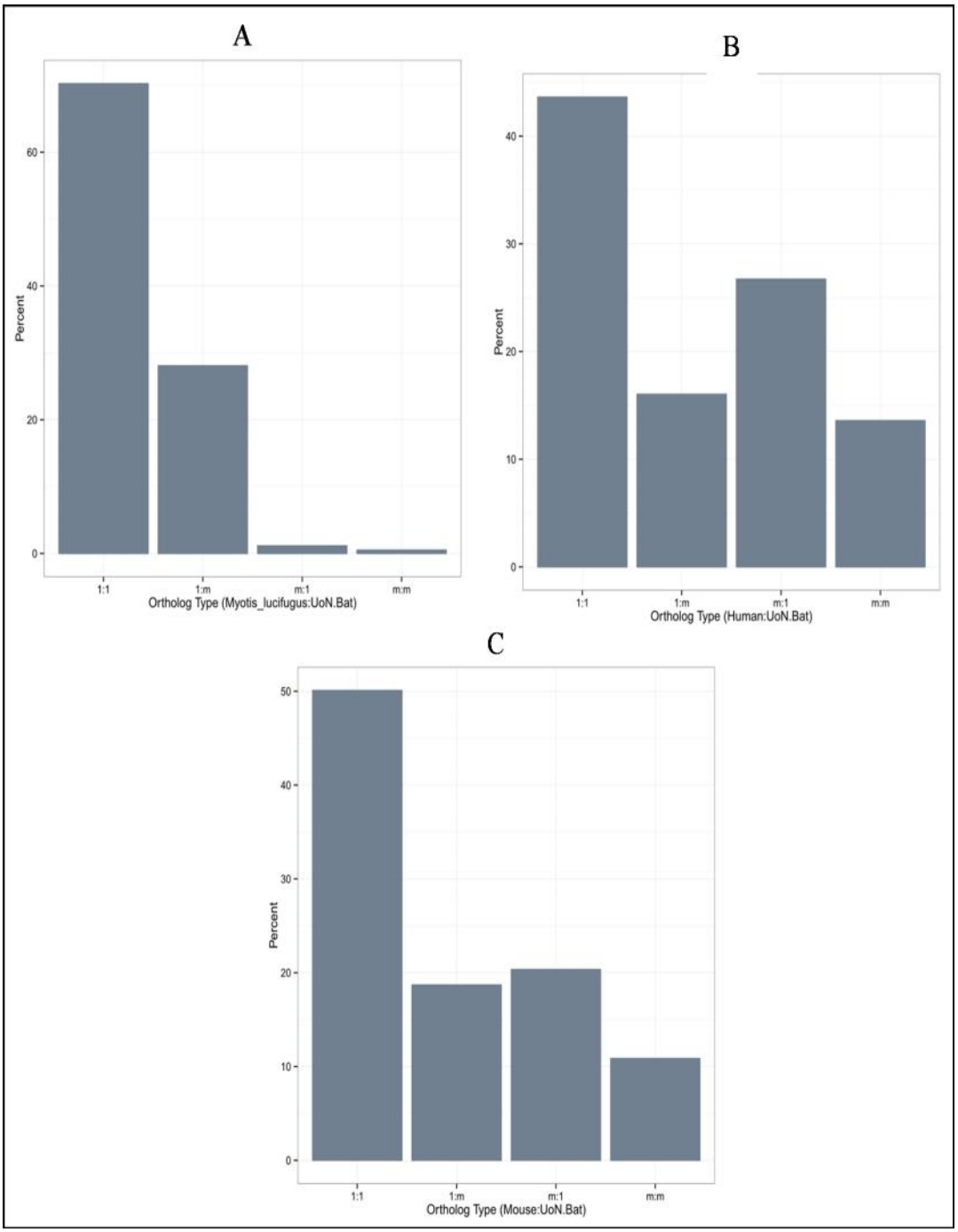


Figure 3.2 A, B, C. Bar plot depicting the percent count of orthologues to the type of orthologues (1:1,1:m,m;1,m:m) present in UoN bat compared with A) M. lucifugus (Ensembl bat), B) Human, C) Mouse. The percent orthologue data is more in 1:1 orthology type in all three in comparison with 1:m, m:1, m:m, respectively.

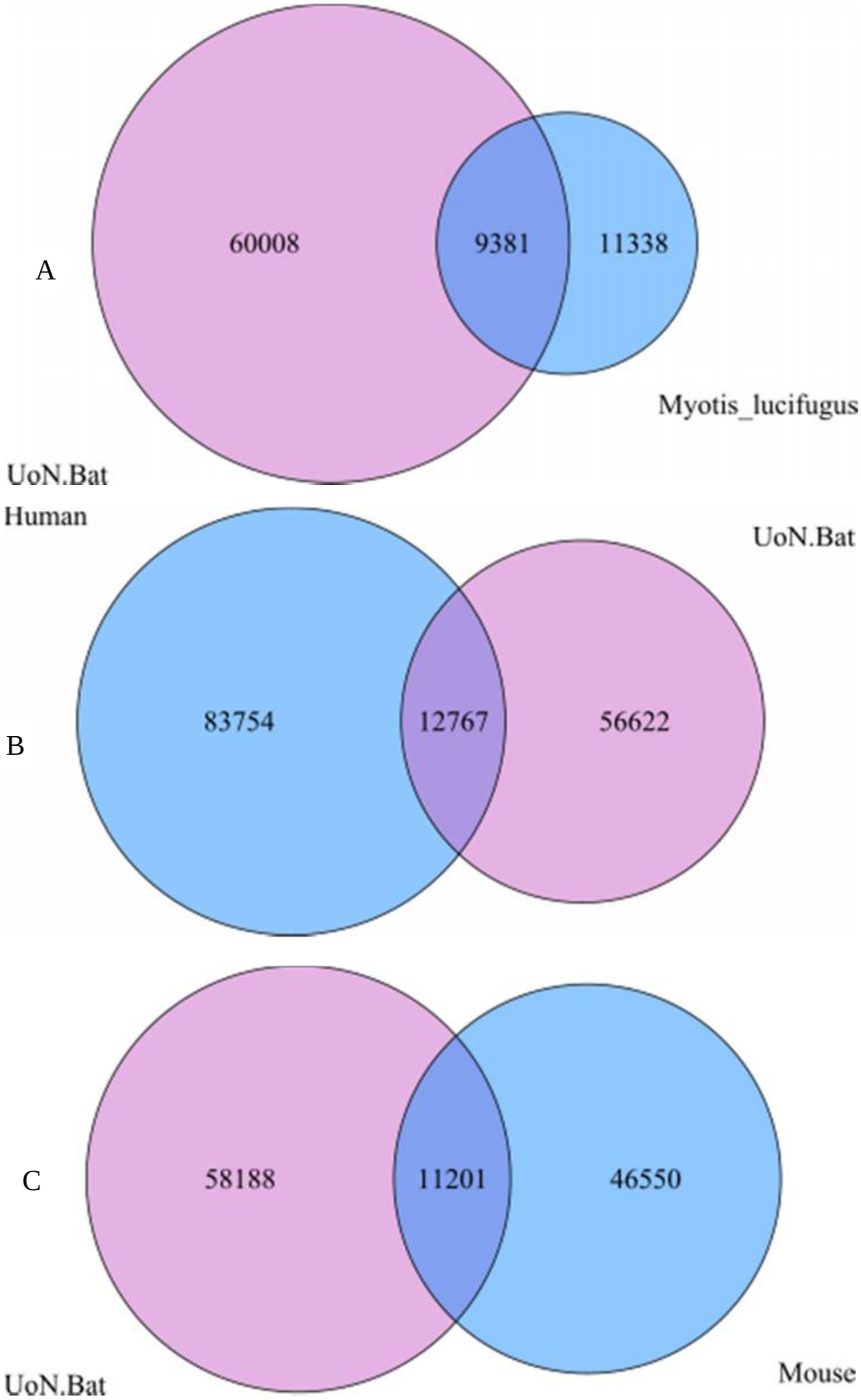


Figure 3.3 A, B, C. Euler plot depicts the sets and quantitative predicted open reading frames (ORF) data indicated by the numeric labels in the regions of the corresponding diagrams, A) Overlap of predicted proteins between UoN bat and known *Myotis lucifugus*, B) UoN bat and human, C) UoN bat and mouse.

Table 3.6: Genes enriched in Brain which are highly expressed and orthologues in Mouse and Human, Ensembl Bat and UoN Bat (Total number of genes after HISAT and StringTie - 43800).

S.No.	Orthologues present in	Number of genes
1	Mouse and Human	1631
2	Ensembl Bat (Myotis)	1630
3	Unique to UoN Bat (Pipistrellus)	190

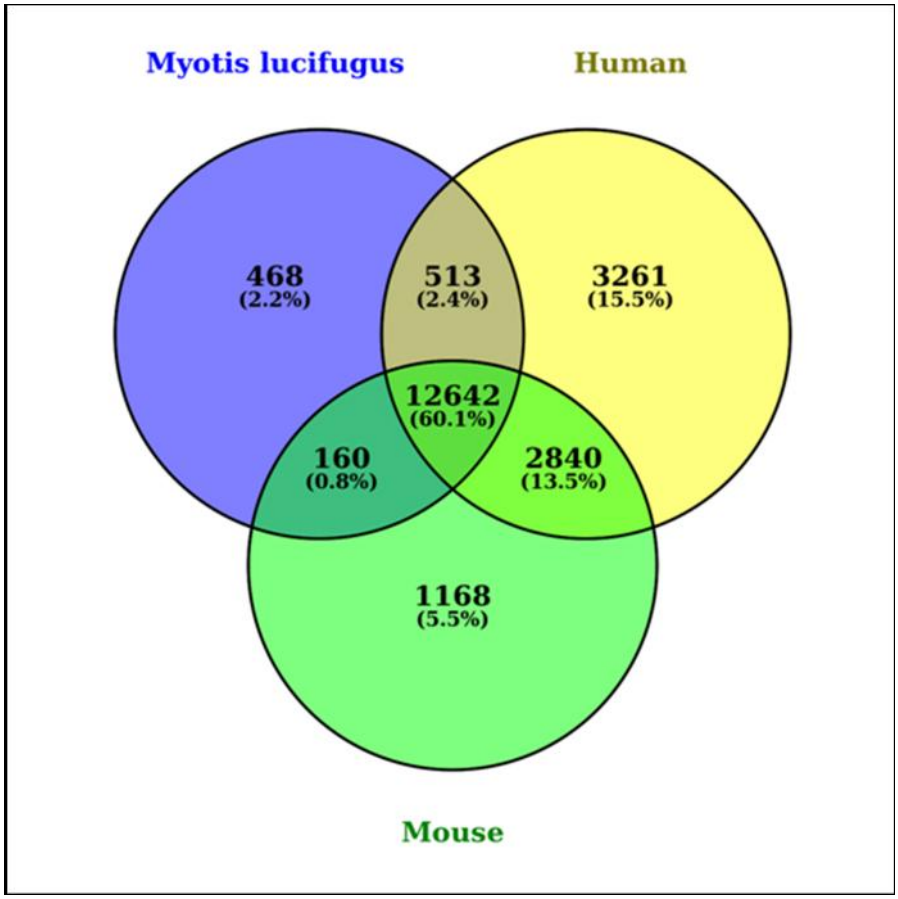


Figure 3.4: Venn diagram showing the overlap of UoN transcripts with *Myotis lucifugus*, Human and Mouse for known genes.

Another class receptor is mGluR, group 1 metabotropic glutamate receptors (mGluR). Glutamate binding to this receptor initiates a signal cascade, involving the breakdown of the membrane lipid PIP2 by phospholipase C (PLC) to the important signaling molecules IP3, which releases Ca^{2+} from Ca^{2+} stores (Figure 3.5) and diacylglycerol (DAG), which leads to the activation of the calcium sensitive kinase PKC. The vast

majority of proteins involved in LTP have an orthologue identified in UoN Bat transcriptome, suggesting that the transcriptome is of relatively good quality.

3.3.2 Melatonin Signaling

Synthesis of Melatonin (N-Acetyl-5 Methoxy-Tryptamine) in the pineal gland functions as a hormonal message encoding for the duration of darkness. Melatonin is a hormone secreted mainly by the pineal gland or epiphysis; it is also produced, but in much smaller quantities, by the retina. Melatonin is produced nocturnally, and is a neurochemical representation of time (Touitou, 2005).

As per the literature and from the findings of Ekmekcioglu (2006), Melatonin activates or inhibits signal transduction cascades through receptors or independent of receptors. The hormone binds with high affinity to membrane receptors (picomole) and to the nuclear receptors (nanomolar), RZR/ROR as well as to Calmodulin. Free radical scavenging mechanism was reported by Carlberg (2000) at higher concentrations Melatonin. Figure 3.6 illustrates the melatonin signaling pathway which was derived by IPA which shows various known genes interaction. The genes colored in the picture viz., MTNR1B, RZR, CAMK II, PKC, MEK, ERK1/2, PKA, PLC, are the highly expressed genes of UoN bat which are involved in the melatonin signaling imparting towards neuronal function.

Melatonin inhibits CalmK II (Ca^{2+}/Calmodulin Dependent Protein Kinase-II) activity, an abundant enzyme in the nervous system. In the brain it phosphorylates a broad spectrum of substrates, thus modulating important neuronal functions and cellular functions may be rhythmically regulated by melatonin modulation of calmodulin dependent protein phosphorylation. Melatonin scavenges oxygen-centered free radicals, especially the highly toxic hydroxyl radical, and neutralizes them by a single electron transfer, which results in detoxified radicals. The hormone may therefore protect macromolecules, particularly DNA, from oxidative damage (Ekmekcioglu, 2006).

3.3.3 Analysis of genes involved in echolocation

From an extensive lieterature search, twenty one genes were identified as associated with echolocation (Table 3.7), orthologs of eight of these were identified in the UoN Bat brain transcriptome; Fos, Foxp2, Gjb2, KCNq4, Otof, SLC26a5, Tmc1 and Wnt8a (Table 3.8B).

The most well-known “echolocation” protein is Prestin encoded by SLC26a5. This gene was encoded by transcript UoN.Ppi.T.43038_UoN.Ppi.T.43038.2|m.5012 and expression was 15.9 in the four replicates. In addition we identified and assessed the expression of Prestin interacting proteins.

Six interacting proteins were identified and five showed measurable expression in our samples (Figure 3.7). Of these, Thyroid hormone receptor alpha (THRA) is the most highly expressed (Figure 3.7). Only the interacting protein CASK (Calcium/ calmodulin-dependent serine protein kinase) is not identified in the UoN Bat data set. The expression of the eight identified echolocation genes are shown in (Figure 3.8). Otof, Gjb2, Tmc1 and Slc26a5 are highly expressed in the UoN bat transcript data in comparison with other genes of interest Foxp2, Kcnq4, Fos, Wnt8a.

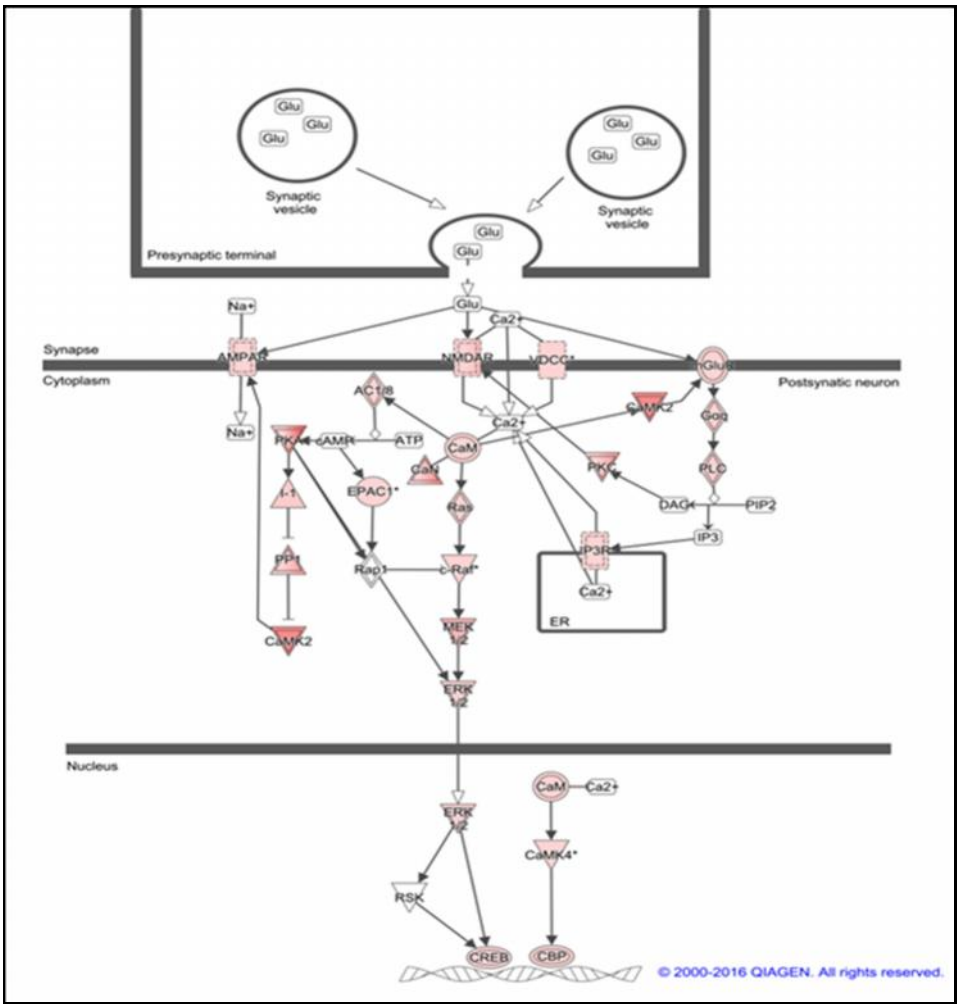


Figure 3.5 Genes encoding proteins involved in LTP, a classical neural process expressed in UoN transcripts. Red shows expression of identified orthologues in UoN Bat transcriptome.

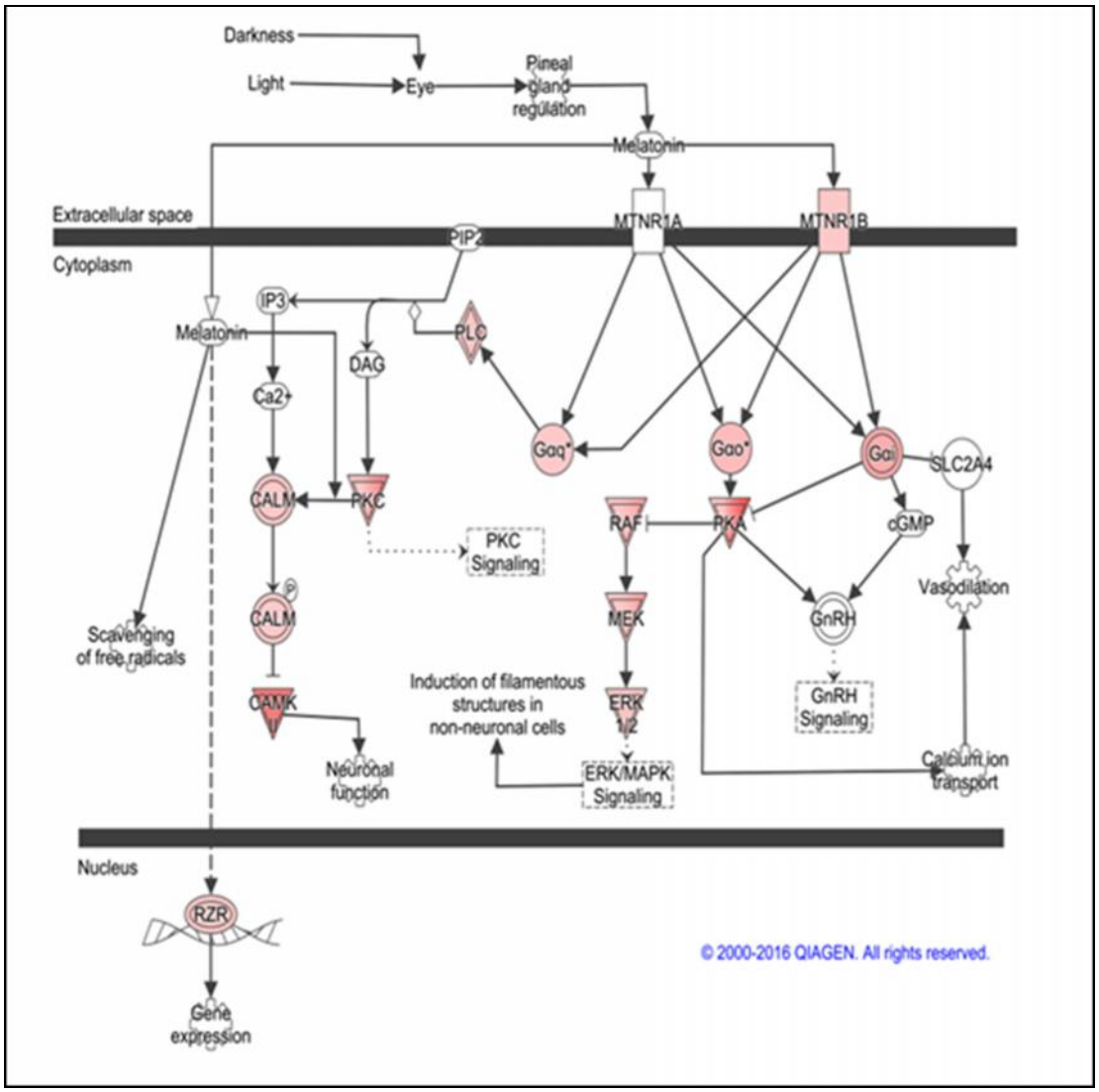


Figure 3.6 Highly expressed genes are involved in the process such as Melatonin signalling. Red shows expression of identified orthologues in UoN Bat transcriptome.

Table 3.7 Genes of interest in UoN Bat involved in echolocation (Vocalization and hearing).

S No.	Name of the Gene	Function	Reference
1.	FoxP2	Transcription factor involved in neural control of Orofacial coordination	(Fisher and Marcus, 2006)
2.	Prestin (Slc 26/Slc 26a5)	Motor protein of cochlear outer hair cells (OHCs)	(Zheng et al., 2002)
3.	Kcnq4	Voltage gated potassium channel Progressive loss of high frequency hearing	(Kharkovets et al., 2000)
4.	Tmc1	Transmembrane protein 1	(Kurima et al., 2002)
5.	Pjvk (Dfnb59)	Encodes 'pejvakin' Non-syndromic hearing loss in mammals Auditory neuropathy in humans Vestibular defects in mice	(Zhang et al., 2015)
6.	Cdh 23	Cadherin	(Kazmierczak et al., 2007)
7.	Pcdh 15	Pro Cadherin	(Kazmierczak et al., 2007)
8.	Otof	Auditory cortex of adult bats that use laryngeal echolocation	(Choi et al., 2009)
9.	Wnt 8a (Wingless related integration site)	Humans encoded by wnt 8a gene Involved in development of early embryos as well as germ cell tumours	(Hanai et al., 2002)
10.	Fos	The Fos promoter - marker of neuronal activity during behaviour Transcription factor AP1 (Activating protein)	(Cohen and Greenberg, 2008)
11.	Chrna 10	Laryngeal echolocation paraphyletic	(Vetter et al., 2007)
12.	Myo 15	Non-syndromic autosomal recessive profound hearing loss disorder (DFNB3)	(Van Camp and Smith, 2002)
13.	Ush 1g	Plays important role in development and maintenance of the auditory and visual systems and functions in the cohesion of hair bundles formed by inner ear sensory cells	(Zou et al., 2011)
14.	Strc (Stereocilin)	Hair bundle of sensory hair cells in the inner ear. Autosomal recessive non syndromic deafness	(Verpy et al., 2000, Francey et al., 2012)
15.	Tect a	Tectorin alpha autosomal dominant non syndromic hearing impairment and recessive form of sensory neural pre-lingual non syndromic deafness	(Schacter et al., 2011)
16.	Tect b Tectorin beta	Tectorial membrane, which covers the auditory hair cells in the cochlea of the inner ear (Low frequency hearing)	(Kammerer et al., 2012)
17.	Oto g (Otogelin)	Role in auditory and vestibular deafness	(Schraders et al., 2012)
18.	Col 11a2	Type – XI collagen/cartilage Pro alpha 2 chain	(Kirwan et al., 2013)
19.	Gjb2	cAMP signaling by intervals with Ras1	(Inoshita et al., 2008)
20.	Cldn 14 (Claudin 14)	Defects in this gene causes Autosomal recessive form of nonsyndromic sensorineural deafness	(Ben-Yosef et al., 2003)
21.	Pou 3f4 (Brn4, BRN4, otf 9)	Pou domain, class 3 transcription factor 4	(Choi et al., 2013)

To compare the evolution of the echolocation genes to the whole genome, pairwise similarity data were obtained. Part A of figures 3.9-3.11 show histograms of the distribution of percent similarity between orthologs of pairs of organisms (mouse:bat, figures 3.9, mouse:human 3.10 and mouse:rat 3.11, respectively). A higher median percent identity (PID) indicates higher similarity. Homologous percent id was closer with the Rat and *Myotis* and was more than 90% for most of the echolocating genes of interest (Table 3.9).

The genes associated with echolocation are generally highly conserved (represented by blue vertical lines in figures 3.9-3.11). Genome wide histogram is shown in (grey) part A of Figures 3.9, 3.10 and 3.11, which indicates the divergence in pair wise comparison of mouse: bat, mouse: human. In contrast, mouse: rat showed conserved distribution both in complete genome wide and echolocating genes (blue vertical lines).

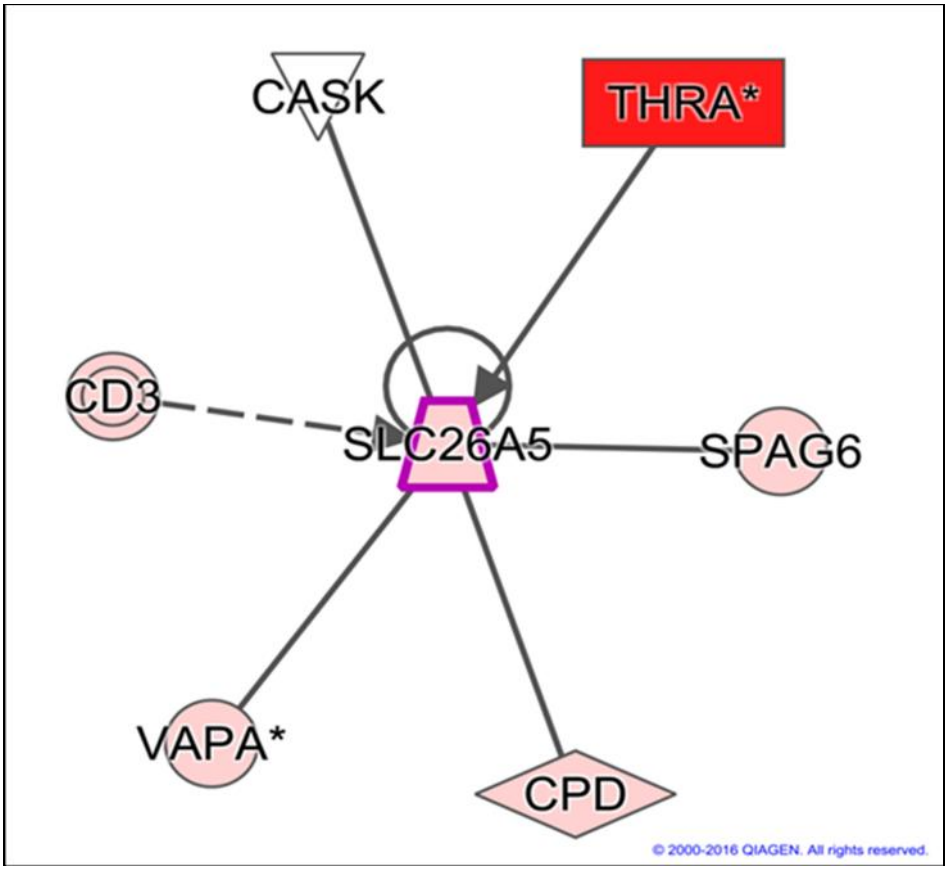


Figure 3.7 Expression of SLC26aS encoded by the Prestin gene associated with echolocation and its known interacting proteins. Red indicates mean expression level of UoN Bat orthologous transcript.

Table 3.8 A: Orthologues of Bat brain transcriptomes in Mega bat and micro bat involved in Echolocation along with Ensembl protein ID

Gene Name	Ensembl Gene ID	Megabat Ensembl Gene ID	Microbat Ensembl Gene ID	Ensembl Protein ID	UoN Transcript	Description
Cdh23	ENSMUSG00000012819	ENSPVAG00000011952	ENSMLUG00000015707	ENSMLUP00000014339	NA	
Chrna10	ENSMUSG00000066279	ENSPVAG00000016436	ENSMLUG00000013586	ENSMLUP00000012363	NA	cholinergic receptor nicotinic alpha 10 subunit [Source:HGNC Symbol;Acc:HGNC:13800]
Cldn14	ENSMUSG00000047109	–	–			
Col11a2	ENSMUSG00000024330	ENSPVAG00000001885	NA	NA	NA	-
Fos	ENSMUSG00000021250	ENSPVAG00000000671	ENSMLUG00000003729	ENSMLUP00000003392	UoN.Ppi.T.43510_UoN.Ppi.T.43510.1|m.7831	Fos proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3796]
Foxp2	ENSMUSG00000029563	ENSPVAG00000005340	ENSMLUG00000015998	ENSMLUP00000014578	UoN.Ppi.T.42052_UoN.Ppi.T.42052.1|m.482	forkhead box P2 [Source:HGNC Symbol;Acc:HGNC:13875]
Gjb2	ENSMUSG00000046352		ENSMLUG00000015730	ENSMLUP00000014330	UoN.Ppi.T.42943_UoN.Ppi.T.42943.1|m.5123	gap junction protein beta 2 [Source:HGNC Symbol;Acc:HGNC:4284]
Kcnq4	ENSMUSG00000028631	ENSPVAG00000010527	ENSMLUG00000007262	ENSMLUP00000006626	UoN.Ppi.T.43296_UoN.Ppi.T.43296.1|m.5694	potassium voltage-gated channel subfamily Q member 4 [Source:HGNC Symbol;Acc:HGNC:6298]
Myo15	ENSMUSG00000042678	ENSPVAG00000005773	ENSMLUG00000001272	ENSMLUP00000001202	NA	myosin XVA [Source:HGNC Symbol;Acc:HGNC:7594]
Otof	ENSMUSG00000062372	ENSPVAG00000001875	ENSMLUG00000010168	ENSMLUP00000009331	UoN.Ppi.T.43189_UoN.Ppi.T.43189.1|m.5874	otoferlin [Source:HGNC Symbol;Acc:HGNC:8515]
Pcdh15	ENSMUSG00000052613	ENSPVAG00000009706	ENSMLUG00000014263	ENSMLUP00000012987	NA	protocadherin related 15 [Source:HGNC Symbol;Acc:HGNC:14674]

Gene Name	Ensembl Gene ID	Megabat Ensembl Gene ID	Microbat Ensembl Gene ID	Ensembl Protein ID	UoN Transcript	Description
Pou3f4	ENSMUSG00000056854	ENSPVAG00000001305	ENSMLUG00000004140	ENSMLUP00000003765	NA	POU class 3 homeobox 4 [Source:HGNC Symbol;Acc:HGNC:9217]
Slc26a5	ENSMUSG00000029015	ENSPVAG00000011364	ENSMLUG00000008900	ENSMLUP00000008129	UoN.Ppi.T.43038_UoN.Ppi.T.43038.2|m.5012	solute carrier family 26 member 5 [Source:HGNC Symbol;Acc:HGNC:9359]
Strc	ENSMUSG00000033498	ENSPVAG00000008857	ENSMLUG00000014042	ENSMLUP00000012784	NA	stereocilin [Source:HGNC Symbol;Acc:HGNC:16035]
Tecta	ENSMUSG00000037705	ENSPVAG00000011726	ENSMLUG00000013553	ENSMLUP00000012347	NA	tectorin alpha [Source:HGNC Symbol;Acc:HGNC:11720]
Tectb	ENSMUSG00000024979	ENSPVAG00000010063	ENSMLUG00000001911	ENSMLUP00000001741	NA	tectorin beta [Source:HGNC Symbol;Acc:HGNC:11721]
Tmc1	ENSMUSG00000024749	ENSPVAG00000003238	ENSMLUG00000007179	ENSMLUP00000006566	UoN.Ppi.T.43040_UoN.Ppi.T.43040.2|m.4239	transmembrane channel like 1 [Source:HGNC Symbol;Acc:HGNC:16513]
Ush1g	ENSMUSG00000045288	ENSPVAG00000016051	ENSMLUG00000017782	ENSMLUP00000016222	NA	USH1 protein network component sans [Source:HGNC Symbol;Acc:HGNC:16356]
Wnt8a	ENSMUSG00000012282	ENSPVAG00000010853	ENSMLUG00000010265	ENSMLUP00000009354	UoN.Ppi.T.42570_UoN.Ppi.T.42570.2|m.2346	Wnt family member 8A [Source:HGNC Symbol;Acc:HGNC:12788]

Table 3.8 B: Genes of interest (GOI) involved in echolocation of Bat, Orthologues in Mega bat and micro bat and are unique to UoN Bat

Gene Name	Ensembl Gene ID	Megabat Ensembl Gene ID	Microbat Ensembl Gene ID	Ensembl Protein ID	UoN Transcript	Description
Fos	ENSMUSG00000021250	ENSPVAG00000000671	ENSMLUG00000003729	ENSMLUP00000003392	UoN.Ppi.T.43510_UoN.Ppi.T.43510.1|m.7831	Fos proto-oncogene, AP-1 transcription factor subunit [Source:HGNC Symbol;Acc:HGNC:3796]
Foxp2	ENSMUSG00000029563	ENSPVAG00000005340	ENSMLUG00000015998	ENSMLUP00000014578	UoN.Ppi.T.42052_UoN.Ppi.T.42052.1|m.482	forkhead box P2 [Source:HGNC Symbol;Acc:HGNC:13875]
Gjb2	ENSMUSG00000046352		ENSMLUG00000015730	ENSMLUP00000014330	UoN.Ppi.T.42943_UoN.Ppi.T.42943.1|m.5123	gap junction protein beta 2 [Source:HGNC Symbol;Acc:HGNC:4284]
Kcnq4	ENSMUSG00000028631	ENSPVAG00000010527	ENSMLUG00000007262	ENSMLUP00000006626	UoN.Ppi.T.43296_UoN.Ppi.T.43296.1|m.5694	potassium voltage-gated channel subfamily Q member 4 [Source:HGNC Symbol;Acc:HGNC:6298]
Otof	ENSMUSG00000062372	ENSPVAG00000001875	ENSMLUG00000010168	ENSMLUP00000009331	UoN.Ppi.T.43189_UoN.Ppi.T.43189.1|m.5874	otoferlin [Source:HGNC Symbol;Acc:HGNC:8515]
Slc26a5	ENSMUSG00000029015	ENSPVAG00000011364	ENSMLUG00000008900	ENSMLUP00000008129	UoN.Ppi.T.43038_UoN.Ppi.T.43038.2|m.5012	solute carrier family 26 member 5 [Source:HGNC Symbol;Acc:HGNC:9359]
Tmc1	ENSMUSG00000024749	ENSPVAG00000003238	ENSMLUG00000007179	ENSMLUP00000006566	UoN.Ppi.T.43040_UoN.Ppi.T.43040.2|m.4239	transmembrane channel like 1 [Source:HGNC Symbol;Acc:HGNC:16513]
Wnt8a	ENSMUSG00000012282	ENSPVAG00000010853	ENSMLUG00000010265	ENSMLUP00000009354	UoN.Ppi.T.42570_UoN.Ppi.T.42570.2|m.2346	Wnt family member 8A [Source:HGNC Symbol;Acc:HGNC:12788]

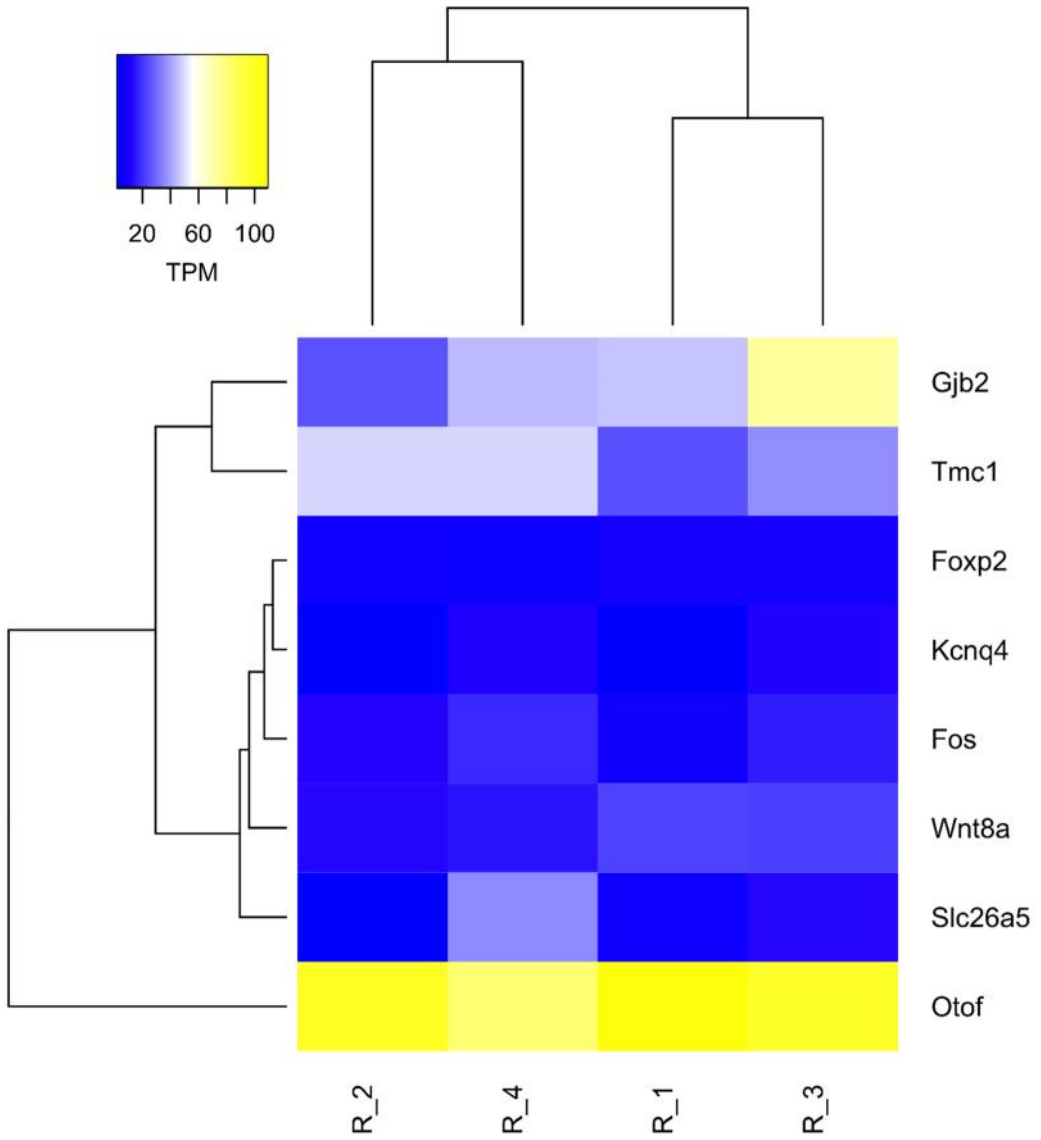


Figure 3.8 Heat map showing median- normalized express Expression (TPM) for highly expressed transcripts of Echolocating genes of interest (GOI) across four replicates of UoN bat. The more the yellow colour the more the more highly expressed that gene is. The heat map was generated using RV2.14.1 heat map package. (Details of the gene name is mentioned in the table 3.6B. R1, R2, R3, R4 represent replicates of UoN bat brain which is refereed as LC5, LC6, LC7, LC8 respectively.

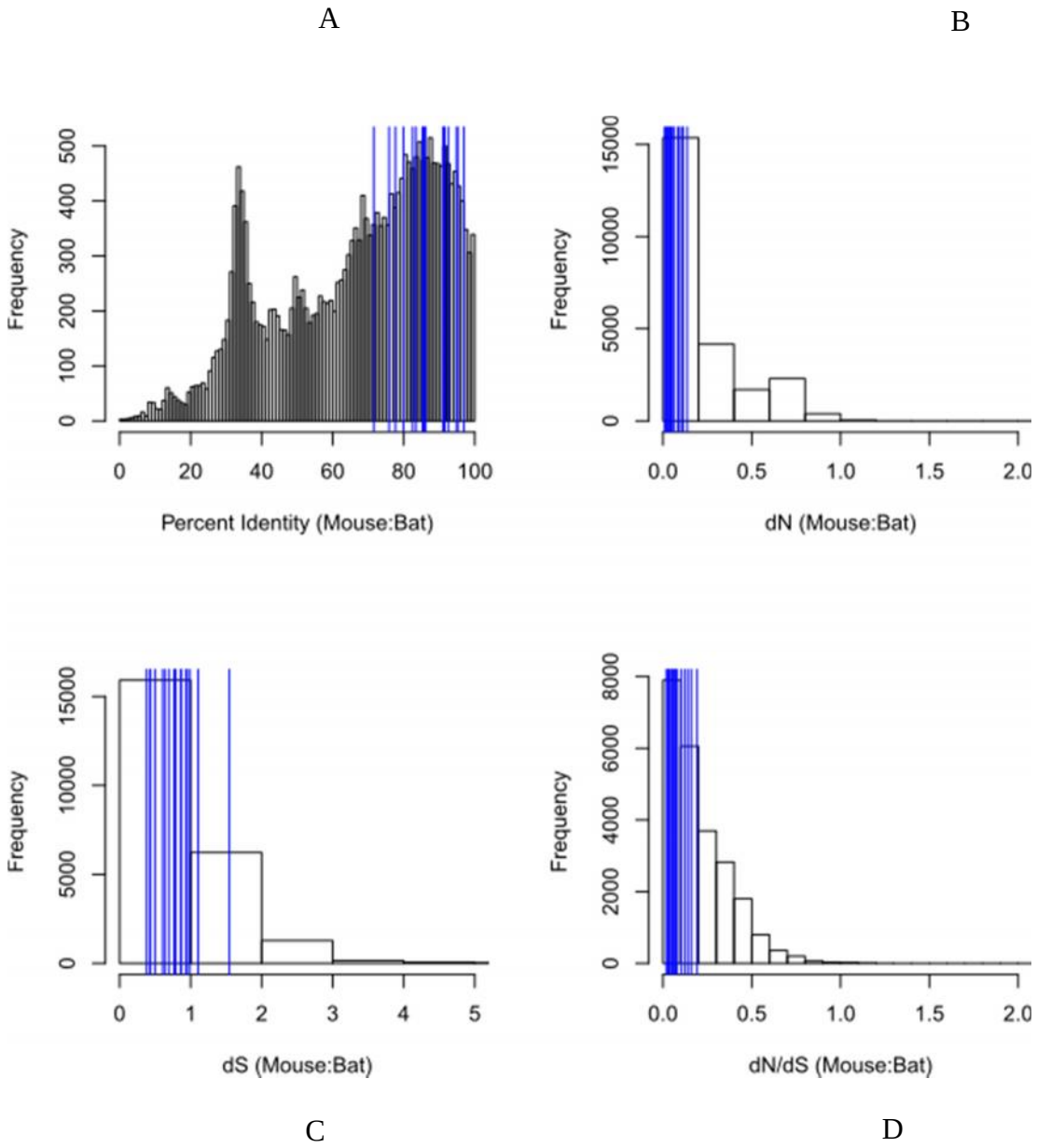


Figure 3.9. A, B, C, D. Histogram showing the frequency distribution of echolocating genes of interest (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Bat). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Bat, C) Synonymous change, dS, mouse: Bat, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Bat, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

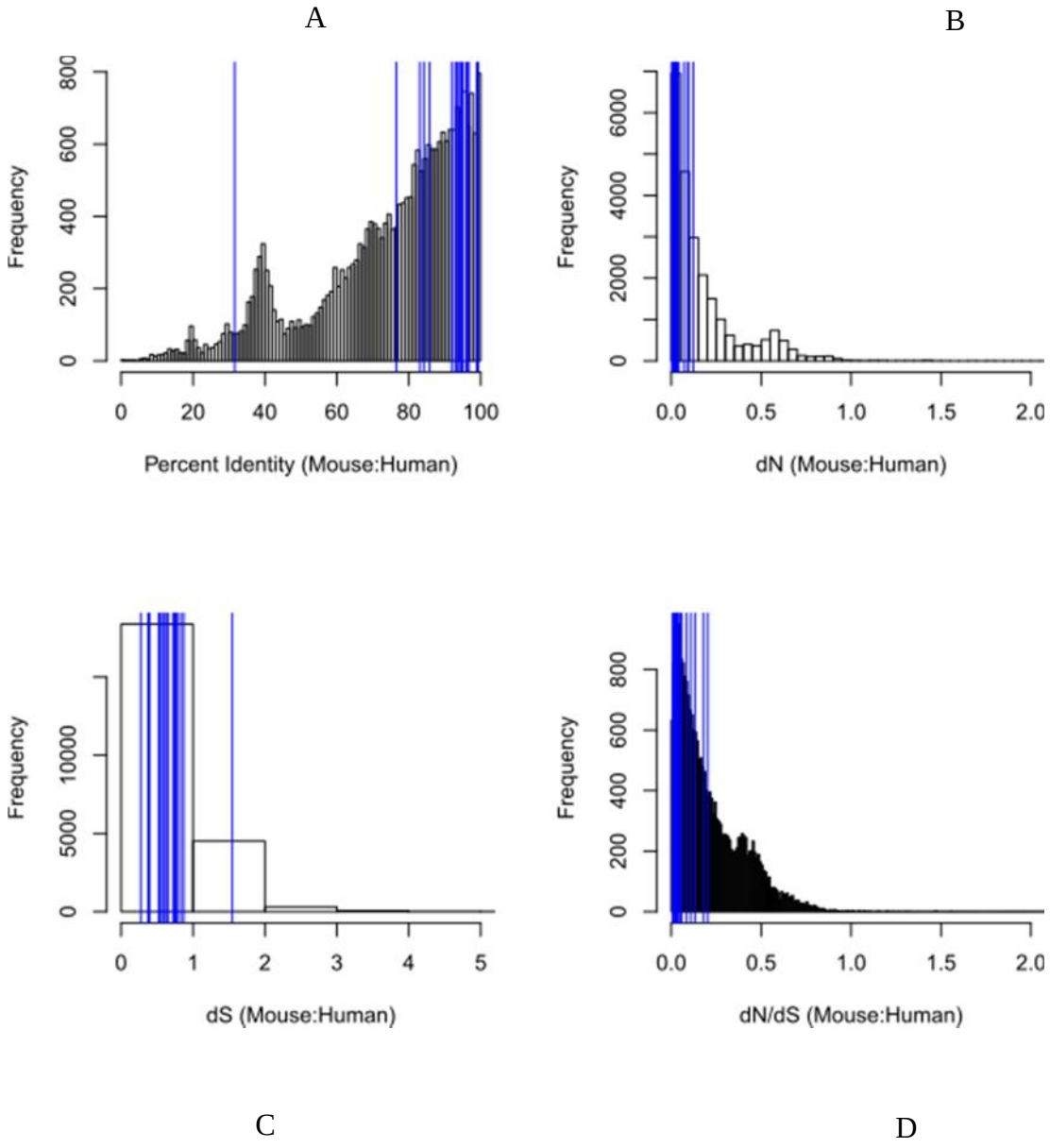
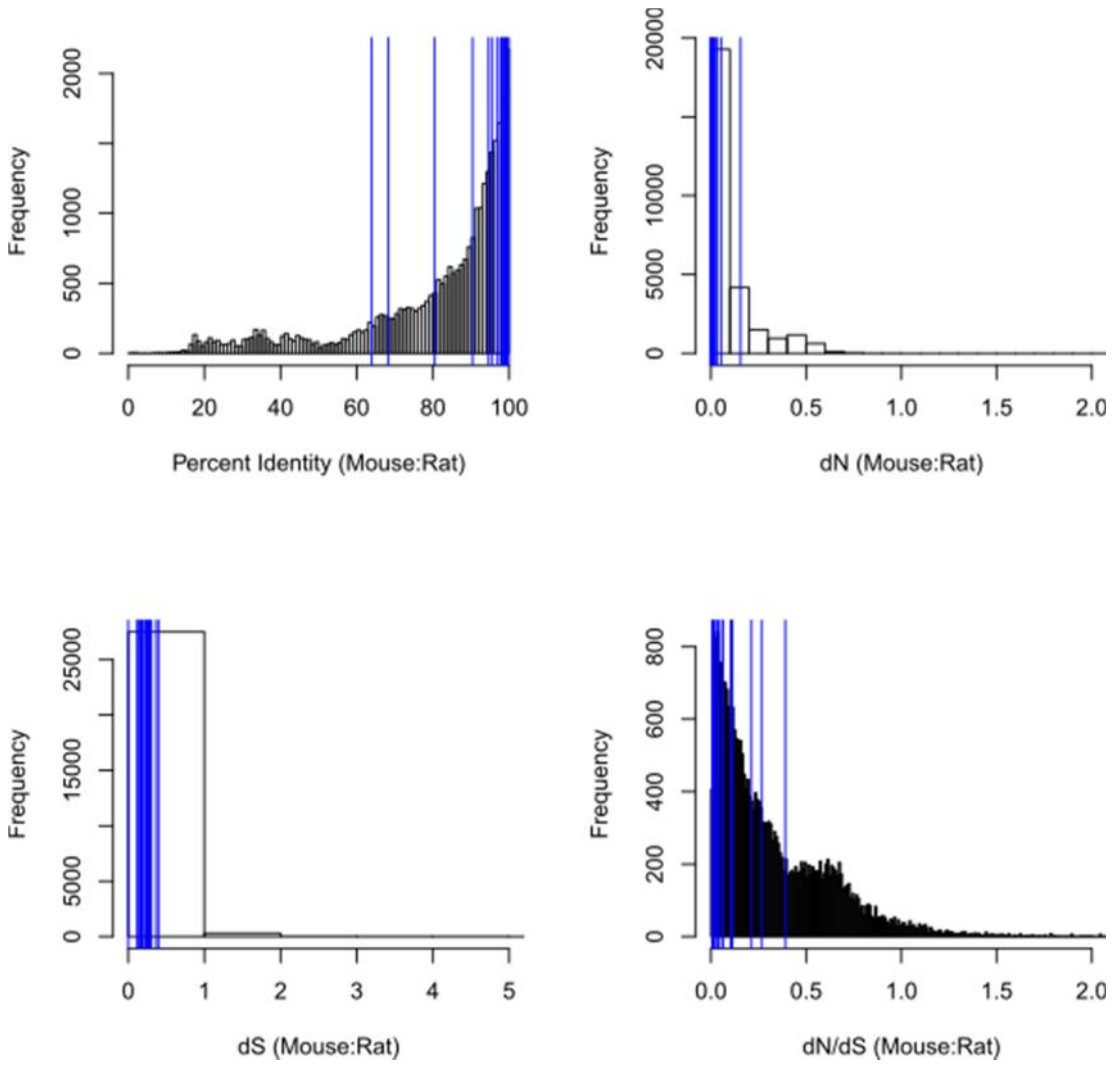


Figure 3.10. A, B, C, D. Histogram showing the frequency distribution of echolocating genes of interest (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Human). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Human, C) Synonymous change, dS, mouse: Human, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Human, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

A B



C D

Figure 3.11: A, B, C, D. Histogram showing the frequency distribution of echolocating genes of interest (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Rat). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Rat, C) Synonymous change, dS, mouse: Rat, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Rat, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

3.4 Adaptive Evolution

In addition to percent protein similarity, the evolutionary selection pressure is shown as non-synonymous substitutions (dN part B of figures), synonymous substitutions (dS part C of figures) and the ration of these dN/dS (part D of figures). The overall distribution of dN/dS ratio of all pairwise comparisosn (as mentioned above) is very similar, with strong purifying selection as the majorty of values are less than 1 and are closer to 0. Synonymous substitutions (d_S), nonsynonymous substitutions (d_N) and selection strength (ω) was determined for echolocating genes and compared to that of the genome as a whole (Table 3.9). Comparison of dN/dS, suggest that, as a class, proteins that are essential in echolocation of UoN bat exhibit relatively low dN/dS and are evolving under strong purifying selection e.g. dN/dS = 0.07, 0.05 for Slc 26 a5 and Gjb2, respectively (Figure 3.12 and Table 3.9).

3.5 Mitochondrial Proteome Study

Given this unique dataset we also investigated expression of genes encoding mitochondrial proteins predicted to be associated with ageing in bat brains. Pollard et al., (2014) identified 12 proteins, ten of these were identified with potential brain and neural functions (Table 3.10).

A total of 6 genes were identified with orthologues in megabat and microbat 'Atp5h', 'Ndufs1', 'Ndufs3', 'Ndufv2', 'Rab14' and 'Tpm1' (Table 3.11 B). The hierarchical clustering in Heat map of mitochondrial proteome study (Figure 3.14) was very consistent among the replicates of UoN bat brain transcriptomes. Tropomyosin 1 (Tpm1) and Ubiquinone oxidoreductase (Ndufs1) was significantly consistent and highly expressed among the replicates. Histograms showing frequency distribution of mitochondrial proteins of interest associated with brain functions were illustrated in Figures 3.15, 3.16, 3.17. Mitochondrial proteins 'Ndufv2', 'Ndufs1', 'Rab14', and 'Tpm1' have high percent identity in Ensembl bat, human and rat (Table 3.12).

Table 3.9: Selection strength (ω) of Echolocating genes of interest (UoN bat), orthologues mouse with Ensembl bat, mouse with human and mouse with rat along with dN and dS values, respectively.

external gene name	M lucifugus				Human				Rat			
	mlucifugus_ homolog_perc_ id	dN	dS	**dN/dS**	hsapiens_ homolog_perc_id	dN	dS	**dN/dS**	rnorvegicus_ homolog_perc _id	dN	dS	**dN/dS**
		mlucifugus_ homolog_ dn	mlucifugus_ homolog_ ds	**Selection strength (ω)**		hsapiens_ homolog_ dn	hsapiens_ homolog_ ds	**Selection strength (ω)**		rnorvegicus_ homolog_ dn	rnorvegicus_ homolog_ ds	**Selection strength (ω)**
Wnt8a	77.68	0.11	0.70	0.16	76.55	0.12	0.60	0.20	90.40	0.05	0.26	0.21
Cdh23	91.59	0.04	0.95	0.04	31.59	0.02	0.78	0.03	80.46	0.03	0.30	0.11
Fos	91.05	0.05	0.61	0.08	93.68	0.03	0.54	0.06	96.84	0.01	0.14	0.11
Col11a2	0.00	0.00	0.00	0.00	94.61	0.03	0.37	0.09	98.30	0.01	0.16	0.06
Tmc1	86.13	0.05	0.78	0.06	96.04	0.02	0.64	0.03	68.30	0.00	0.22	0.01
Tectb	94.83	0.03	0.64	0.04	94.22	0.03	0.60	0.05	98.78	0.01	0.28	0.02
Kcnq4	75.86	0.06	0.86	0.07	96.70	0.01	0.56	0.03	99.14	0.00	0.19	0.02
Slc26a5	92.61	0.04	0.50	0.08	95.03	0.03	0.52	0.05	97.85	0.01	0.29	0.04
Foxp2	85.29	0.01	0.38	0.02	99.30	0.00	0.27	0.01	99.02	0.00	0.18	0.01
Strc	83.47	0.08	0.43	0.19	85.79	0.07	0.39	0.18	94.58	0.03	0.11	0.27
Tecta	82.51	0.11	1.55	0.07	95.87	0.02	0.65	0.03	99.07	0.00	0.25	0.02
Myo15	71.55	0.14	1.11	0.12	83.08	0.09	0.83	0.11	95.56	0.02	0.19	0.11
Ush1g	95.23	0.02	0.99	0.02	96.31	0.02	0.76	0.02	99.57	0.00	0.29	0.01
Gjb2	91.15	0.04	0.77	0.05	92.92	0.03	0.87	0.04	99.12	0.00	0.23	0.02
Cldn14	0.00	0.00	0.00	0.00	93.31	0.03	1.55	0.02	99.16	0.00	0.36	0.01
Pcdh15	79.93	0.11	0.78	0.14	84.20	0.10	0.73	0.13	63.92	0.16	0.40	0.39
Pou3f4	96.95	0.01	0.43	0.03	98.89	0.01	0.39	0.01	100.00	0.00	0.00	0.00
Otof	85.73	0.05	0.93	0.05	94.89	0.02	0.75	0.03	98.10	0.01	0.25	0.03
Chrna10	85.46	0.09	0.87	0.10	91.95	0.04	0.74	0.06	97.76	0.01	0.18	0.06

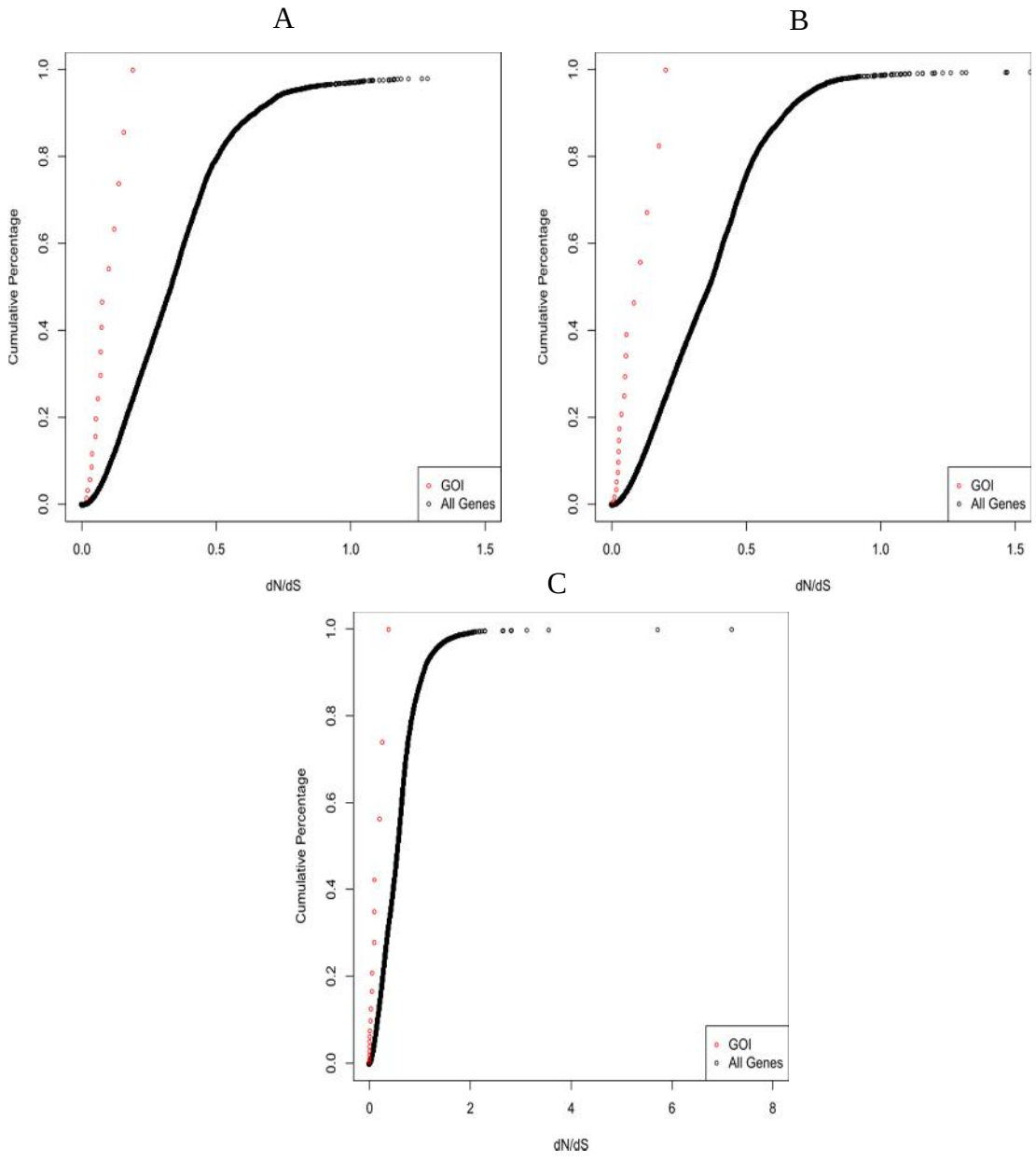


Figure 3.12: A, B, C. Cumulative Percentage (distribution) of dN/dS for UoN bat echolocating genes of interest (GOI), are compared with A) All genes of Ensembl bat, B) All human genes, C) All genes of rat. Significant difference in the distribution of dN/dS was observed between the GOI and All genes of respective species. Numeric details listed in the table no. 3.7

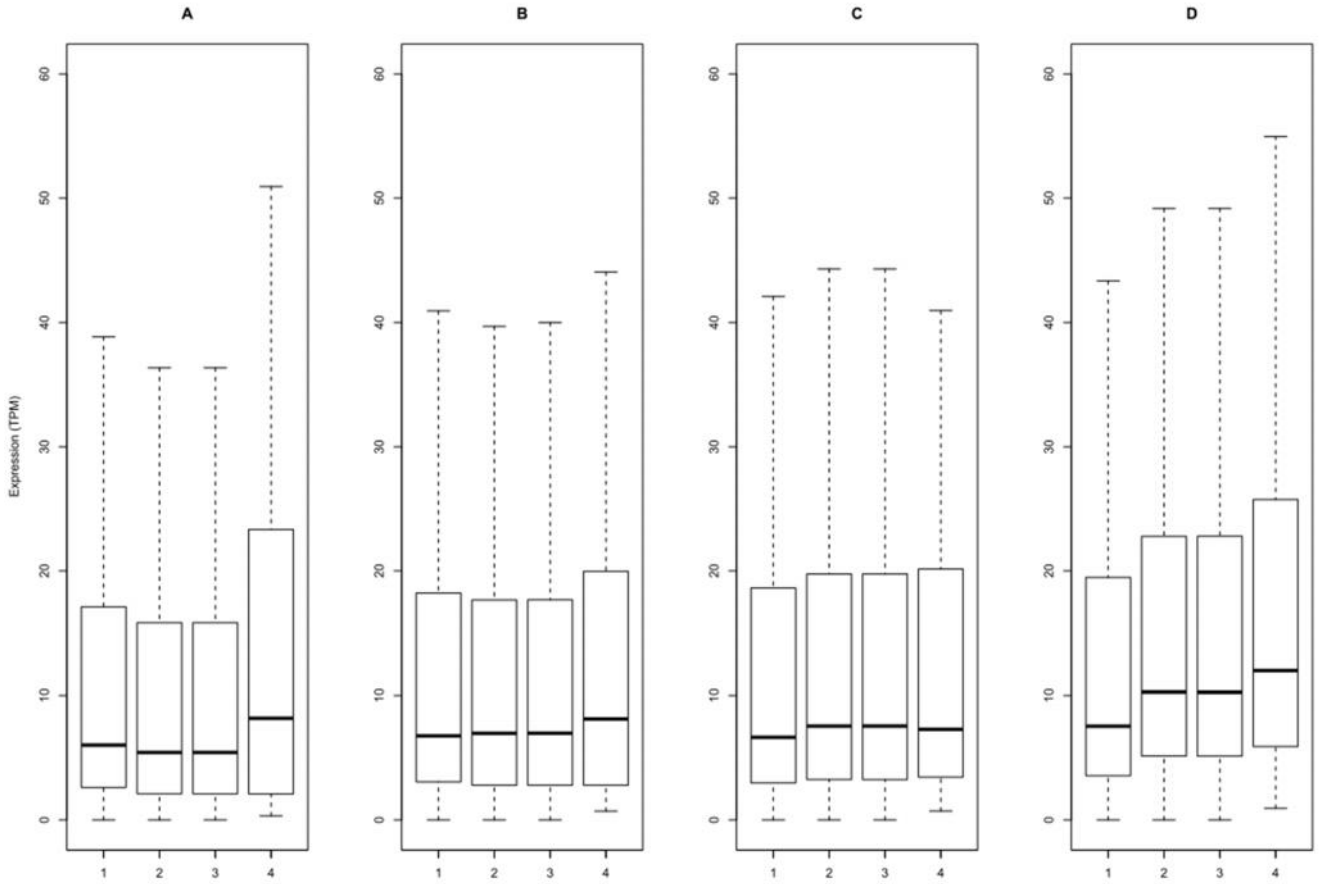


Figure 3.13. Box and Whisker plots of expression (TPM > 5X Median) of transcript of abundances across four replicates (A,B,C,D, respectively) of UoN bat brain transcript for All Transcripts, Orthologous transcripts in Human and Mouse, Orthologus transcripts in Ensemble bat, Unique orthologues in UoN bat (indicated by 1, 2, 3, 4, respectively in the figure).

Table 3.10: Mitochondrial Bat (UoN) Proteins/Genes involved in neural and Brain functions.

S.No.	Gene/Protein	Function	Reference
1.	NDUFS3 NADH dehydrogenase [ubiquinone] iron-sulfur protein 3	NADH:Ubiquinone oxidoreductase core subunit S3 - Disruption of mitochondrial membrane potential and protein binding	(Martinvalet et al., 2008)
2.	NDUFV2 NADH dehydrogenase [ubiquinone] flavoprotein 2	NADH: Ubiquinone oxidoreductase core subunit V2. Involved in the transfer of electrons	(de Coo et al., 1997)
3.	Rab1B ras-related protein Rab-1B	Secretory pathway and Vesicular transport	(Alvarez et al., 2003)
4.	ATP5H ATP synthase subunit d	Catalytic and Protein translocation	(Javed and Wright, 1991)
5.	RAB14	Biosynthetic/recycling pathway between the Golgi and endosomal compartments.	(Junutula et al., 2004)
7.	MBP myelin basic protein	Important role in myelination	(Dayhoff, 1972)
8.	UQCRFS1 Cytochrome b-c1 complex subunit Rieske	Stabilization and electron transfer	(Duncan et al., 1994)
9.	TPM1 Tropomyosin alpha-1 chain	Cardiac function	(Denz et al., 2004)
10.	NDUFS1 NADH dehydrogenase Fe-S protein 1	Involved with several mitochondrial changes associated with apoptosis	(Ricci et al., 2004)

Table 3.11 A: Orthologues of Bat and Mouse Mitochondrial proteins in Mega bat and micro bat involved in brain functions along with Ensemble protein ID.

Gene Name	Ensembl Gene ID	Megabat Ensembl Gene ID	Microbat Ensembl Gene ID	Ensembl Protein ID	UoN Transcript ID	Description
Atp5h	ENSMUSG00000034566		ENSMLUG00000010620	ENSMLUP00000009679	UoN.Ppi.T.43172_UoN.Ppi.T.43172.2|m.4807, UoN.Ppi.T.43172_UoN.Ppi.T.43172.3|m.4813, UoN.Ppi.T.43173_UoN.Ppi.T.43173.1|m.4819, UoN.Ppi.T.43172_UoN.Ppi.T.43172.1|m.4801	ATP synthase, H+ transporting, mitochondrial Fo complex subunit D [Source:HGNC Symbol;Acc:HGNC:845]
Mbp	ENSMUSG00000041607	ENSPVAG00000008208	ENSMLUG00000011157	ENSMLUP00000010163	NA	
Ndufs1	ENSMUSG00000025968	ENSPVAG00000017332	ENSMLUG00000016594	ENSMLUP00000015124	UoN.Ppi.T.42305_UoN.Ppi.T.42305.1|m.1477, UoN.Ppi.T.42304_UoN.Ppi.T.42304.1|m.1468	NADH:ubiquinone oxidoreductase core subunit S1 [Source:HGNC Symbol; Acc:HGNC:7707]
Ndufs3	ENSMUSG00000005510	ENSPVAG00000007344	ENSMLUG00000016649	ENSMLUP00000015171	UoN.Ppi.T.43015_UoN.Ppi.T.43015.3|m.5764, UoN.Ppi.T.43016_UoN.Ppi.T.43016.1|m.5773, UoN.Ppi.T.43015_UoN.Ppi.T.43015.2|m.5755	NADH:ubiquinone oxidoreductase core subunit S3 [Source:HGNC Symbol;Acc: HGNC:7710]
Ndufv2	ENSMUSG00000024099	ENSPVAG00000004820	ENSMLUG00000013503	ENSMLUP00000012287	UoN.Ppi.T.42979_UoN.Ppi.T.42979.1|m.5276	NADH:ubiquinone oxidoreductase core subunit V2 [Source:HGNC Symbol ;Acc:HGNC:7717]
Rab14	ENSMUSG00000026878	ENSPVAG00000007828	ENSMLUG00000009818	ENSMLUP00000008948	UoN.Ppi.T.43458_UoN.Ppi.T.43458.1|m.5863	RAB14, member RAS oncogene family [Source:HGNC Symbol; Acc:HGNC:16524]
Rab1b	ENSMUSG00000024870	ENSPVAG00000016978	ENSMLUG00000023020	ENSMLUP00000017414	NA	RAB1B, member RAS oncogene family [Source:HGNC Symbol; Acc:HGNC:18370]
Tpm1	ENSMUSG00000032366	ENSPVAG00000008366	ENSMLUG00000007003	ENSMLUP00000006404	UoN.Ppi.T.42300_UoN.Ppi.T.42300.1|m.1372, UoN.Ppi.T.43429_UoN.Ppi.T.43429.1|m.7666, UoN.Ppi.T.43123_UoN.Ppi.T.43123.6|m.5468	tropomyosin 1 (alpha) [Source:HGNC Symbol; Acc:HGNC:12010]

Table 3.11 B: Genes of interest (GOI) of Bat and Mouse mitochondrial proteins involved in brain functions, Orthologues in Mega bat and micro bat and are unique to UoN Bat.

Gene Name	Ensembl Gene ID	Megabat Ensembl Gene ID	Microbat Ensembl Gene ID	Ensembl Protein ID	UoN Transcript ID	Description
Atp5h	ENSMUSG00000034566		ENSMLUG00000010620	ENSMLUP00000009679	UoN.Ppi.T.43172_UoN.Ppi.T.43172.2|m.4807, UoN.Ppi.T.43172_UoN.Ppi.T.43172.3|m.4813, UoN.Ppi.T.43173_UoN.Ppi.T.43173.1|m.4819, UoN.Ppi.T.43172_UoN.Ppi.T.43172.1|m.4801	ATP synthase, H+ transporting, mitochondrial Fo complex subunit D [Source:HGNC Symbol;Acc:HGNC:845]
Ndufs1	ENSMUSG00000025968	ENSPVAG00000017332	ENSMLUG00000016594	ENSMLUP00000015124	UoN.Ppi.T.42305_UoN.Ppi.T.42305.1|m.1477, UoN.Ppi.T.42304_UoN.Ppi.T.42304.1|m.1468	NADH:ubiquinone oxidoreductase core subunit S1 [Source:HGNC Symbol;Acc:HGNC:7707]
Ndufs3	ENSMUSG00000005510	ENSPVAG00000007344	ENSMLUG00000016649	ENSMLUP00000015171	UoN.Ppi.T.43015_UoN.Ppi.T.43015.3|m.5764, UoN.Ppi.T.43016_UoN.Ppi.T.43016.1|m.5773, UoN.Ppi.T.43015_UoN.Ppi.T.43015.2|m.5755	NADH:ubiquinone oxidoreductase core subunit S3 [Source:HGNC Symbol;Acc:HGNC:7710]
Ndufv2	ENSMUSG00000024099	ENSPVAG00000004820	ENSMLUG00000013503	ENSMLUP00000012287	UoN.Ppi.T.42979_UoN.Ppi.T.42979.1|m.5276	NADH:ubiquinone oxidoreductase core subunit V2 [Source:HGNC Symbol;Acc:HGNC:7717]
Rab14	ENSMUSG00000026878	ENSPVAG00000007828	ENSMLUG00000009818	ENSMLUP00000008948	UoN.Ppi.T.43458_UoN.Ppi.T.43458.1|m.5863	RAB14, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:16524]
Tpm1	ENSMUSG00000032366	ENSPVAG00000008366	ENSMLUG00000007003	ENSMLUP00000006404	UoN.Ppi.T.42300_UoN.Ppi.T.42300.1|m.1372, UoN.Ppi.T.43429_UoN.Ppi.T.43429.1|m.7666, UoN.Ppi.T.43123_UoN.Ppi.T.43123.6|m.5468	tropomyosin 1 (alpha) [Source:HGNC Symbol;Acc:HGNC:12010]

Mitochondrial proteins of interest associated with brain and neural functions are highly conserved (represented by blue vertical lines in figures 3.15-3.17). Genome wide histogram was shown in (grey) part A of Figures 3.15, 3.16 and 3.17, which indicates the divergence in pair wise comparison of mouse: bat, however, it was partly diverged in mouse: human. In contrast, mouse: rat showed conserved distribution both in complete genome wide and GOI (echolocating genes, blue vertical lines). In addition to percent protein similarity, the evolutionary selection pressure is shown as non-synonymous substitutions (dN part B of figures), synonymous substitutions (dS part C of figures 3.15-3.17) and the ratio of these dN/dS (part D of figures 3.15-3.17).

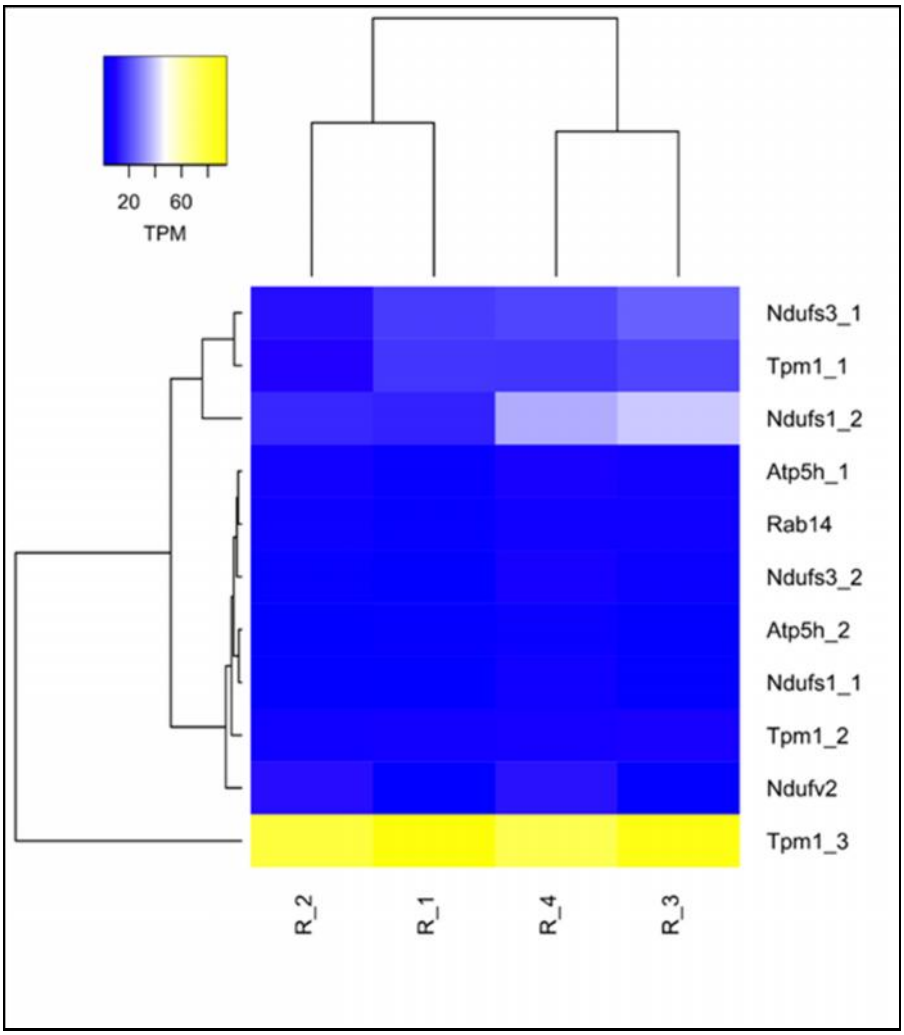


Figure 3.14 Heat map showing median- normalized express Expression (TPM) for highly expressed transcripts of mitochondrial proteins of interest associated with the brain functions across four replicates of UoN bat. The more the yellow colour the more the more highly expressed that gene is. The heat map was generated using RV2.14.1 heat map package. (Details of the gene name is mentioned in the table 3.8B. R1, R2, R3, R4 represent replicates of UoN bat brain which is refereed as LC5, LC6, LC7, LC8 respectively.

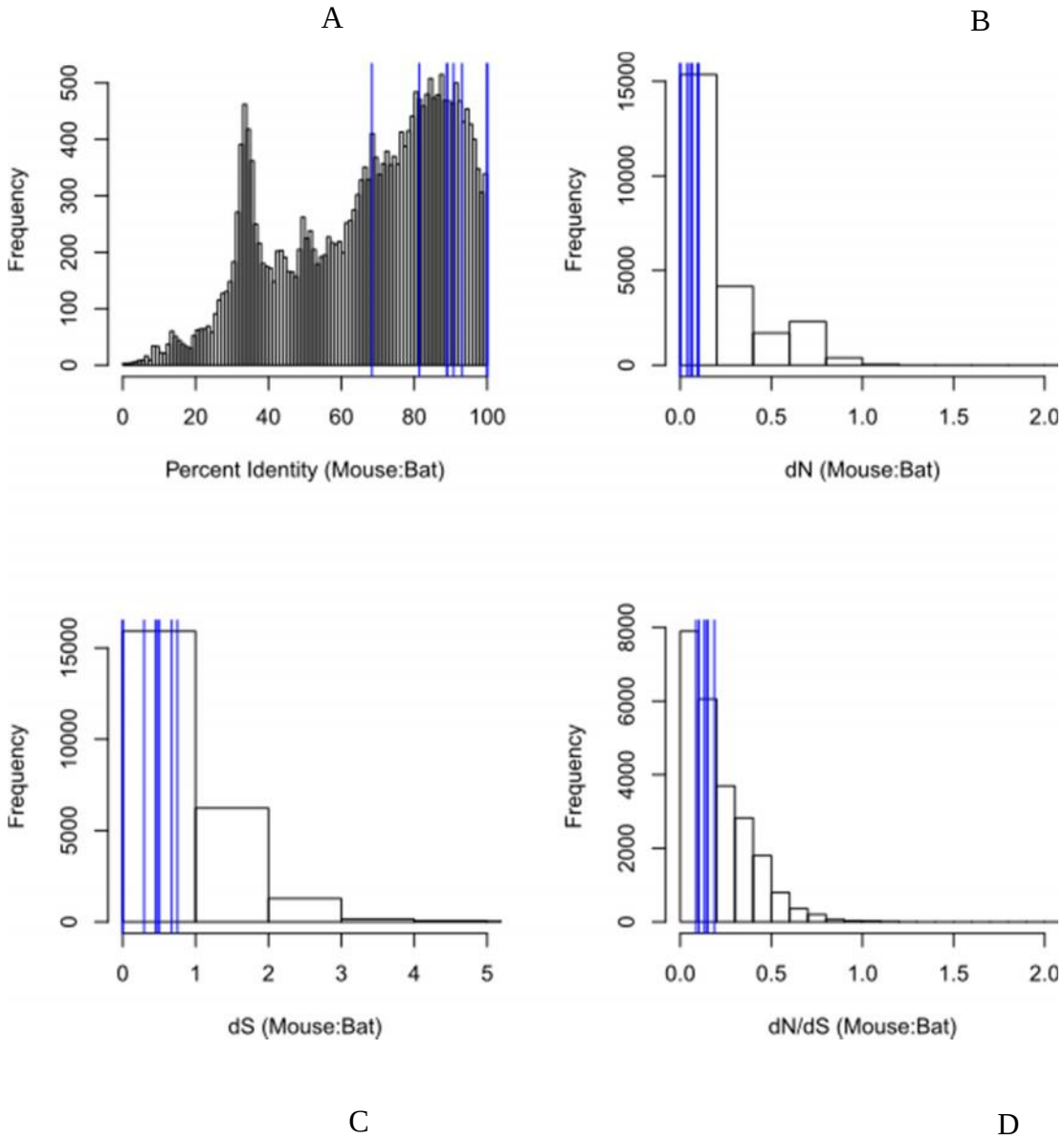


Figure 3.15. A, B, C, D. Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Bat). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Bat, C) Synonymous change, dS, mouse: Bat, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Bat, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

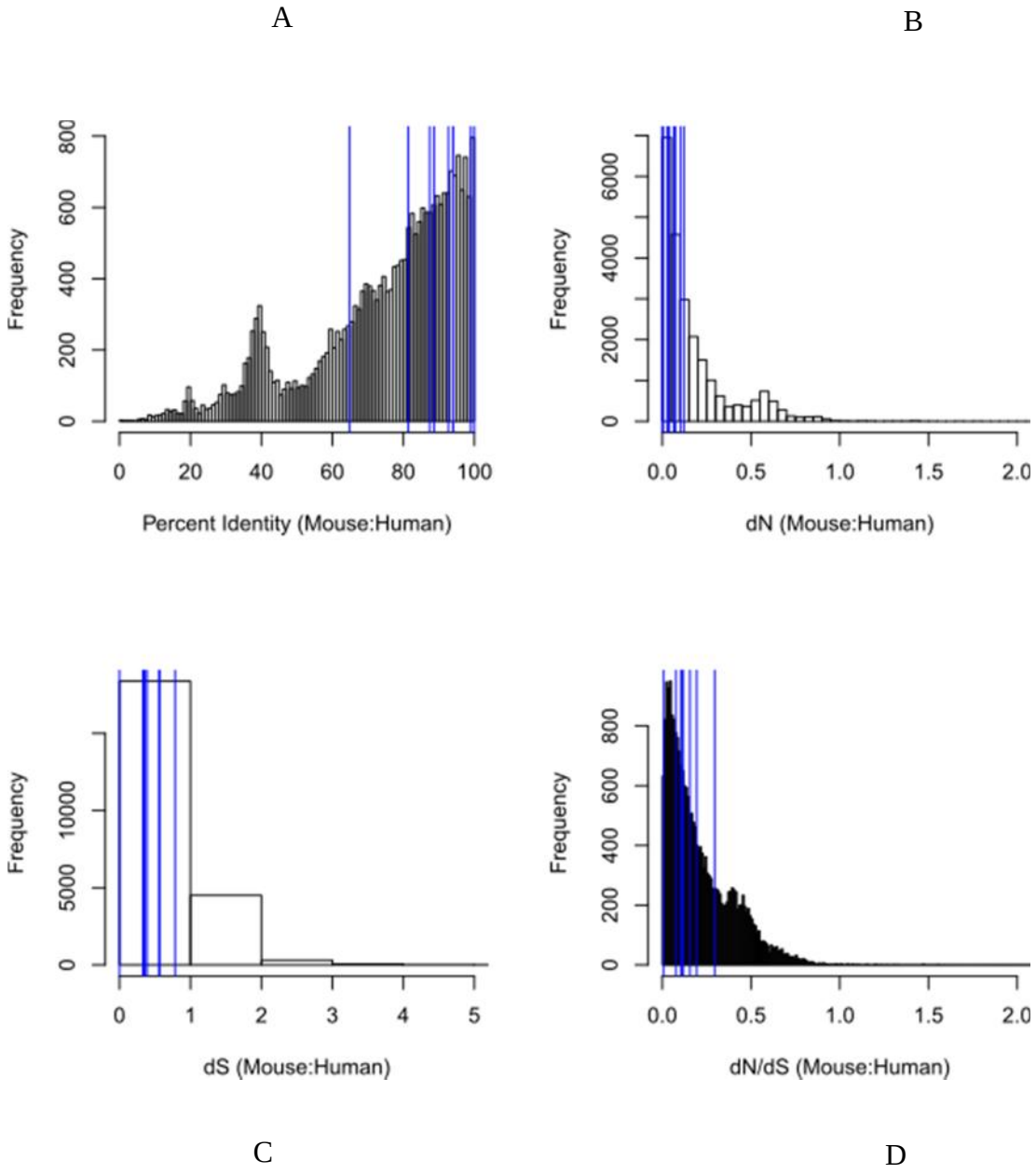


Figure 3.16. A, B, C, D. Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Human). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Human, C) Synonymous change, dS, mouse: Human, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Human, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

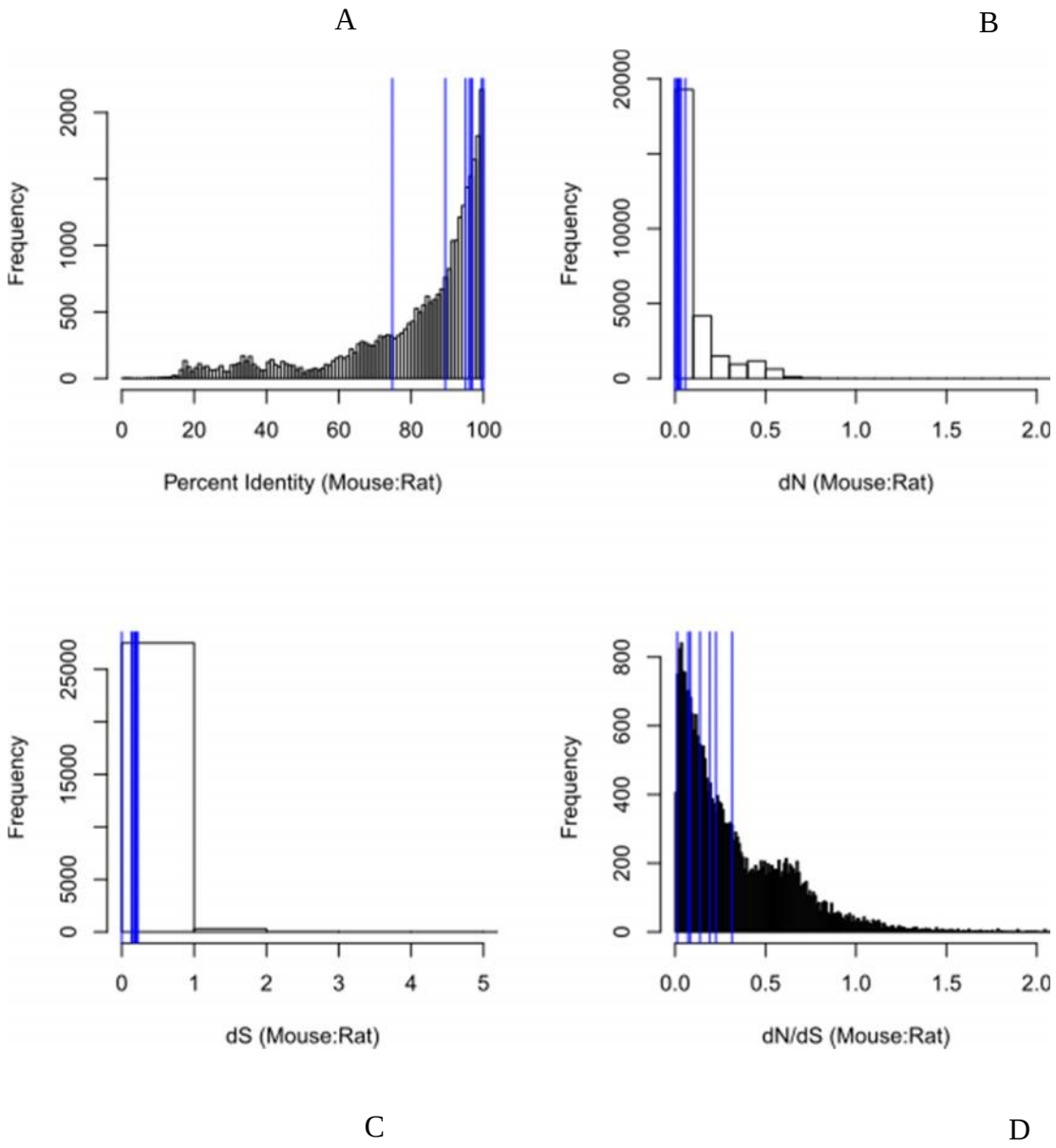


Figure 3.17. A, B, C, D. Histogram showing the frequency distribution of mitochondrial proteins of interest associated with brain functions (UoN bat). A) Percent identity distribution of pairwise comparisons. The graph represents the distribution of protein identity (between Mouse: Rat). The frequency distribution plots used here illustrates the distribution of a percent identity (PID) on X- axis and Y- axis shows the frequency. B) Non-synonymous change dN, mouse: Rat, C) Synonymous change, dS, mouse: Rat, D) The ratio of nonsynonymous substitution rate dN (Ka) and synonymous substitution rate dS (Ks), mouse: Rat, to detect directional selection acting on a protein coding gene (ω=dN/dS gives the purity of selection).

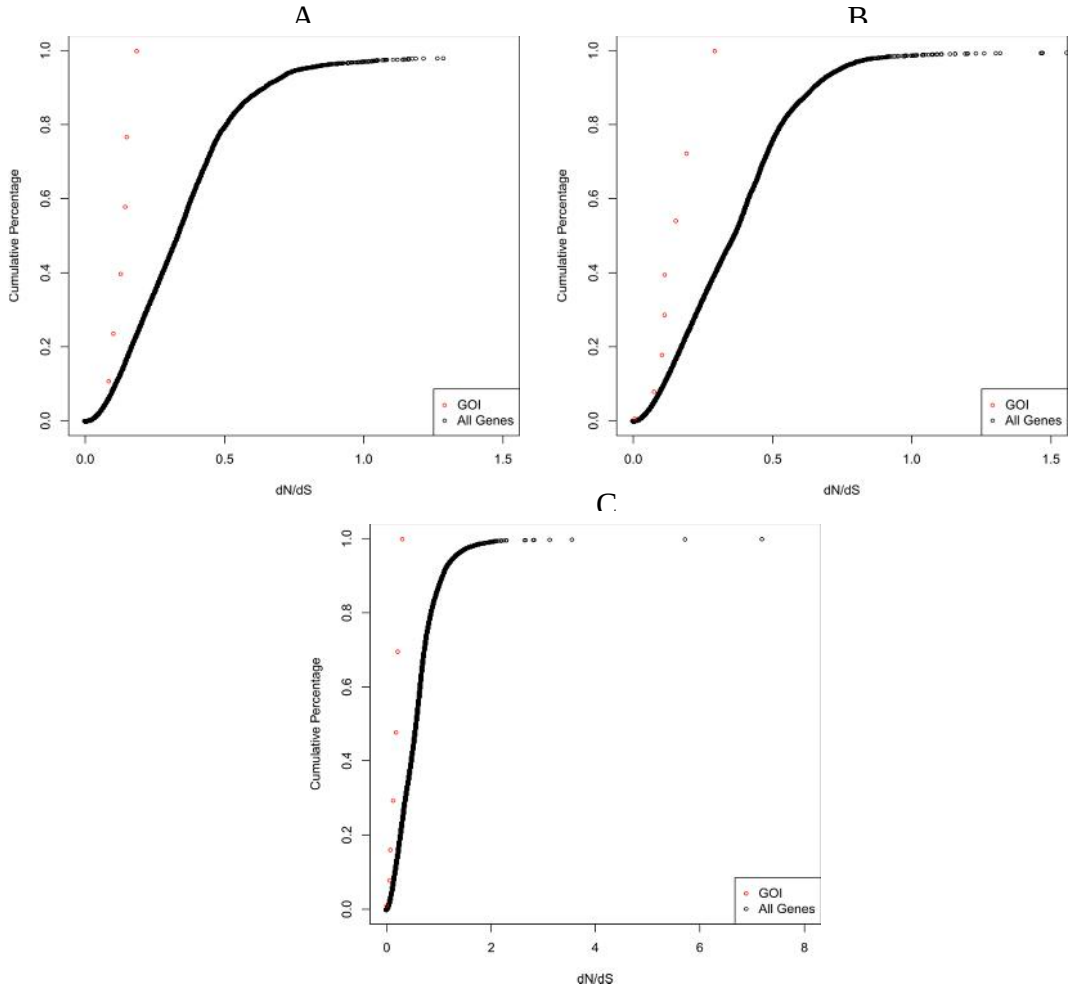


Figure 3.18 A, B, C. Cumulative Percentage (distribution) of dN/dS for mitochondrial proteins of interest (or genes of interest, GOI) associated with brain function (UoN Bat), are compared with A) All genes of Ensemble bat, B) All human genes, C) All genes of rat. Significant difference in the distribution of dN/dS was observed between the GOI and All genes of respective species (Ensembl bat, Human and Rat, respectively). Cumulative percentage (distribution) is the measure of the relative importance of selection and genetic drift in causing amino acid substitutions. Numeric details listed in the table no. 3.10.

The overall distribution of dN/dS ratio of all pairwise comparison (as mentioned above) is very similar, with strong purifying selection as the values are less than 1 and or closer to 0. However, their phylogenetic studies and gene functional analysis need to studied further in understanding its evolutionary and potential functional orientation (GO and pathway interactions) of the proteins (as done in case of Bat brain transcriptomics, present thesis work).

Table 3.12: Selection strength (ω) of UoN Bat and Moue Mitochondrial proteins/Genes of interest, orthologues mouse with Ensembl bat, mouse with human and mouse with rat along with dN and dS values, respectively.

	M lucifugus					**Human**				**Rat**		
		dN	**dS**	**dN/dS**		dN	dS	**dN/dS**		dN	dS	**dN/dS**
external gene name	mlucifugus_homolog_perc_id	mlucifugus_homolog_dn	mlucifugus_homolog_ds	**Selection strength (ω)**	hsapiens homolog perc_id	hsapiens_homolog_dn	hsapiens_homolog_ds	**Selection strength (ω)**	rnorvegicus_homolog_perc_id	rnorvegicus_homolog_dn	rnorvegicus_homolog_ds	**Selection strength (ω)**
Ndufs3	88.97	0.07	0.45	0.15	87.45	0.08	0.39	0.19	95.06	0.03	0.15	0.19
Ndufv2	93.15	0.04	0.29	0.13	92.74	0.04	0.34	0.11	96.77	0.02	0.22	0.07
Rab1b	89.05	0.07	0.75	0.09	99.01	0.00	0.56	0.01	99.50	0.00	0.20	0.01
Ndufs1	90.78	0.05	0.48	0.10	94.09	0.03	0.36	0.08	96.15	0.02	0.14	0.14
Rab14	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
Tpm1	100.00	0.00	0.00	0.00	94.01	0.04	0.33	0.12	0.00	0.00	0.00	0.00
Atp5h	81.37	0.10	0.67	0.15	81.37	0.12	0.78	0.16	89.44	0.06	0.18	0.31
Uqcrfs1	0.00	0.00	0.00	0.00	88.69	0.07	0.56	0.12	96.72	0.02	0.19	0.09
Mbp	68.40	0.09	0.50	0.19	64.80	0.10	0.35	0.29	74.80	0.03	0.14	0.23

3.6 Phylogenetic Study of Genes of Interest (GOI) of UoN Bat Brain Transcripts

The exact evolutionary relationship of bats to their closest mammalian counterparts is not clearly understood. Phylogenetic trees were constructed for the echolocating genes of UoN bat brain transcriptomic data (Table 3.8 B) The phylogenetic tree of echolocating genes of interest viz., Prestin (Slc26a5), Fos, Gjb2, Foxp2, Otof, and Wnt8a was depicted in Figures 3.19 A, 3.19 B, 3.19 C, 3.19 D, 3.19 E, 3.19 F, respectively. Construction of phylogenetic tree of SLc26a5 (Figure 3.19 A) gene depicts that UoN bat Pipistrelle is closer to the rodents and primates. And Figure 3.19 B and 3.19 C illustrate the position of Bats in the super order Laurasiatheria of placental mammals based on the Fos protein and Gjb2 of UoN bat brain transcript.

Expression Foxp2 protein of UoN bat (Figure 3.19 D) in the phylogenetic tree construction is placed in the super order Laurasiatheria, but primates are observed in between the UoN bat (Chiroptera) and *Sorex araneus* L. (Eulipotyphyla). Study of Otof protein (Figure 3.19 E) confirms the position of UoN bat in the super order is placed Laurasiatheria between the order Artiodactyla and Carnivora. Wnt8a Protein (Figure 3.19 F) infers that, it has been placed in between cetacean and carnivore which are the sub orders of Laurasiatheria.

3.7 Functional Characterization of Novel Genes (Gene Ontology and Pathways)

Unique bat genes with high expression (TPM > 5 X Median) were tested for the significant enrichment of specific Gene Ontology (GO) terms for Biological Processes, Molecular Function (MF) & Cellular Processes (CP) and KEGG pathways.

Gene Ontology (GO) enrichment analysis was also performed for the echolocation genes of interest for Biological processes (Figure 3.20 A), Cellular components (Figure 3.20 B) and Molecular functions (Figure 3.20 C). Enrichment of genes in a category is assessed as –log 10 p value from the hypergeometric test. Genes enriched in Biological processes include metabolic processes, cellular processes, localization and transport.

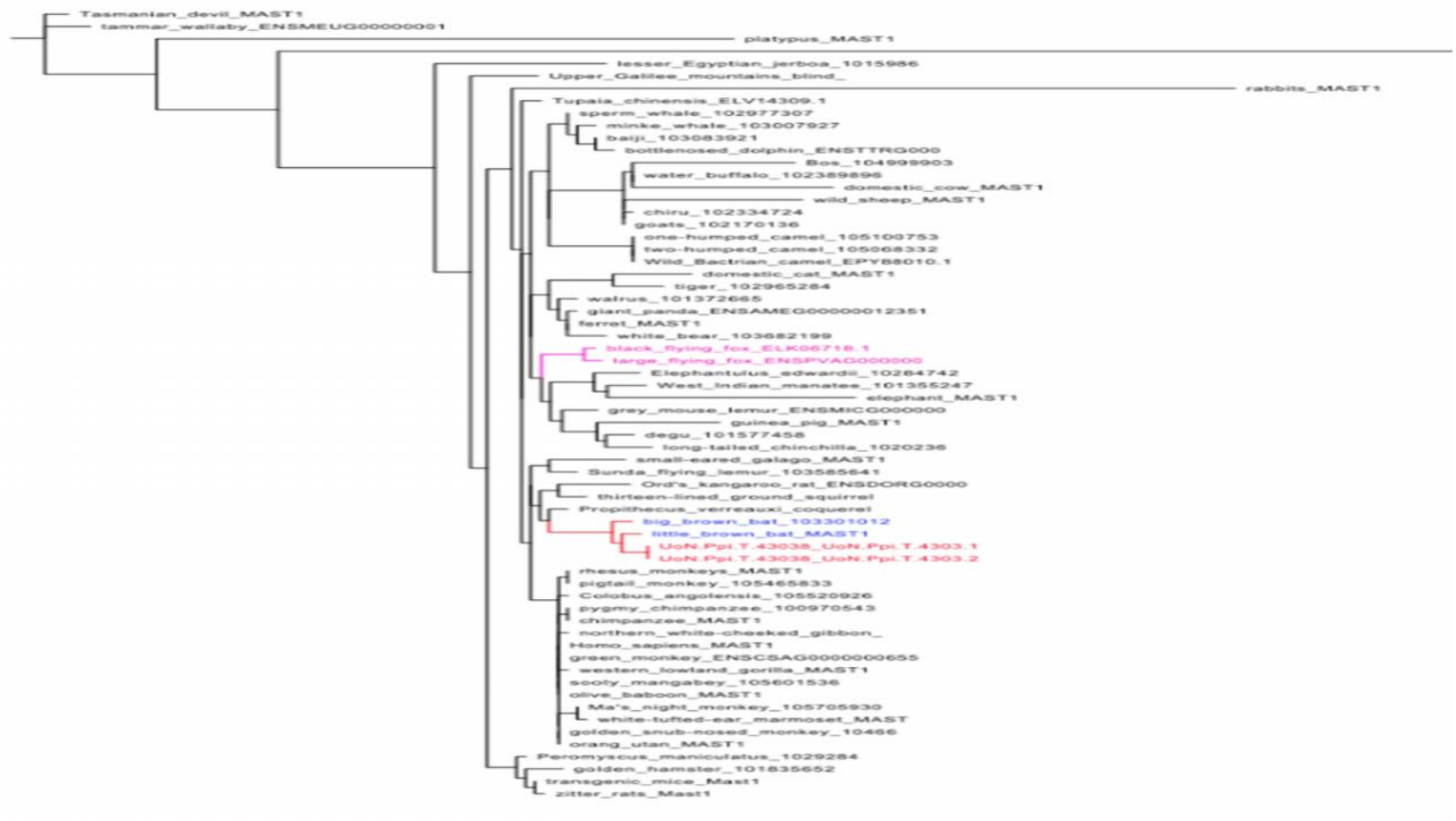


Figure 3.19 A: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of **Prestin** gene (SLC26a5) of UoN bat (transcript ID indicated in orange) which was closely related with brown bat. Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).

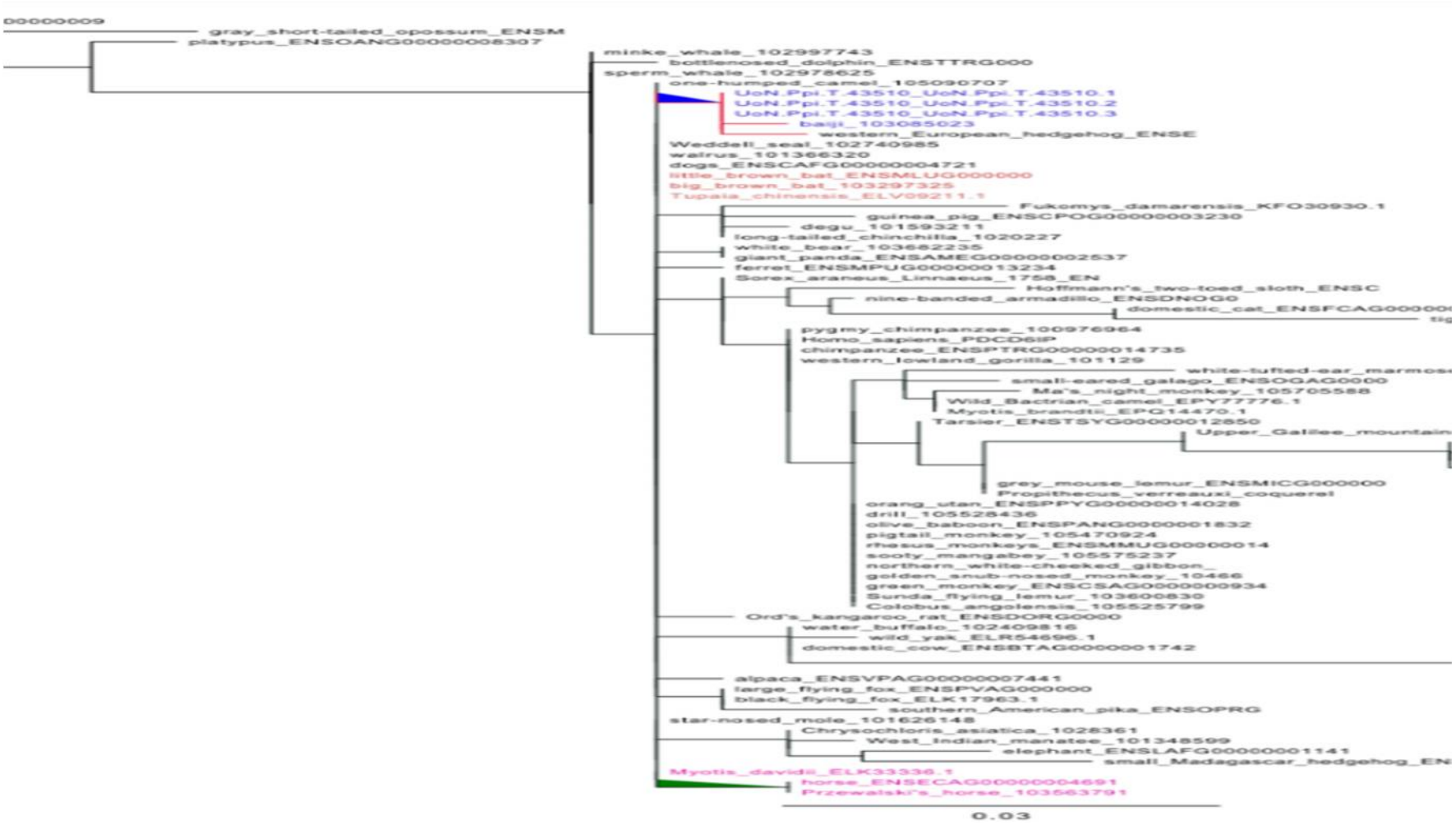


Figure 3.19 B: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of **Fos** protein (gene) of UoN bat (transcript ID indicated in blue) which was closely related with brown bat (indicated in pink). Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).

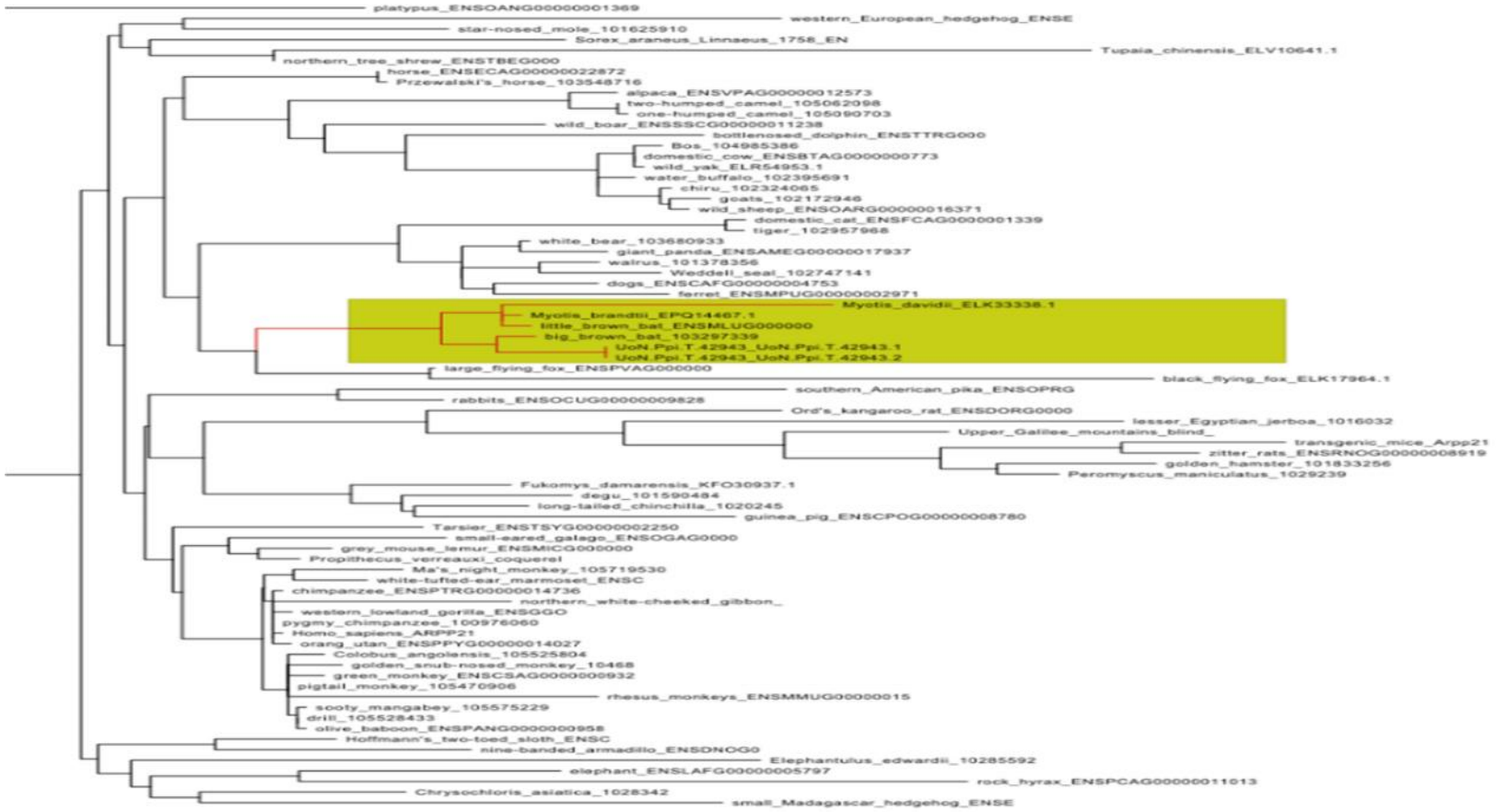


Figure 3.19 C: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Gap Junction protein (**Gjb2)** protein (gene) of UoN bat (transcript ID highlighted in green) which was closely related with brown bat. Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).

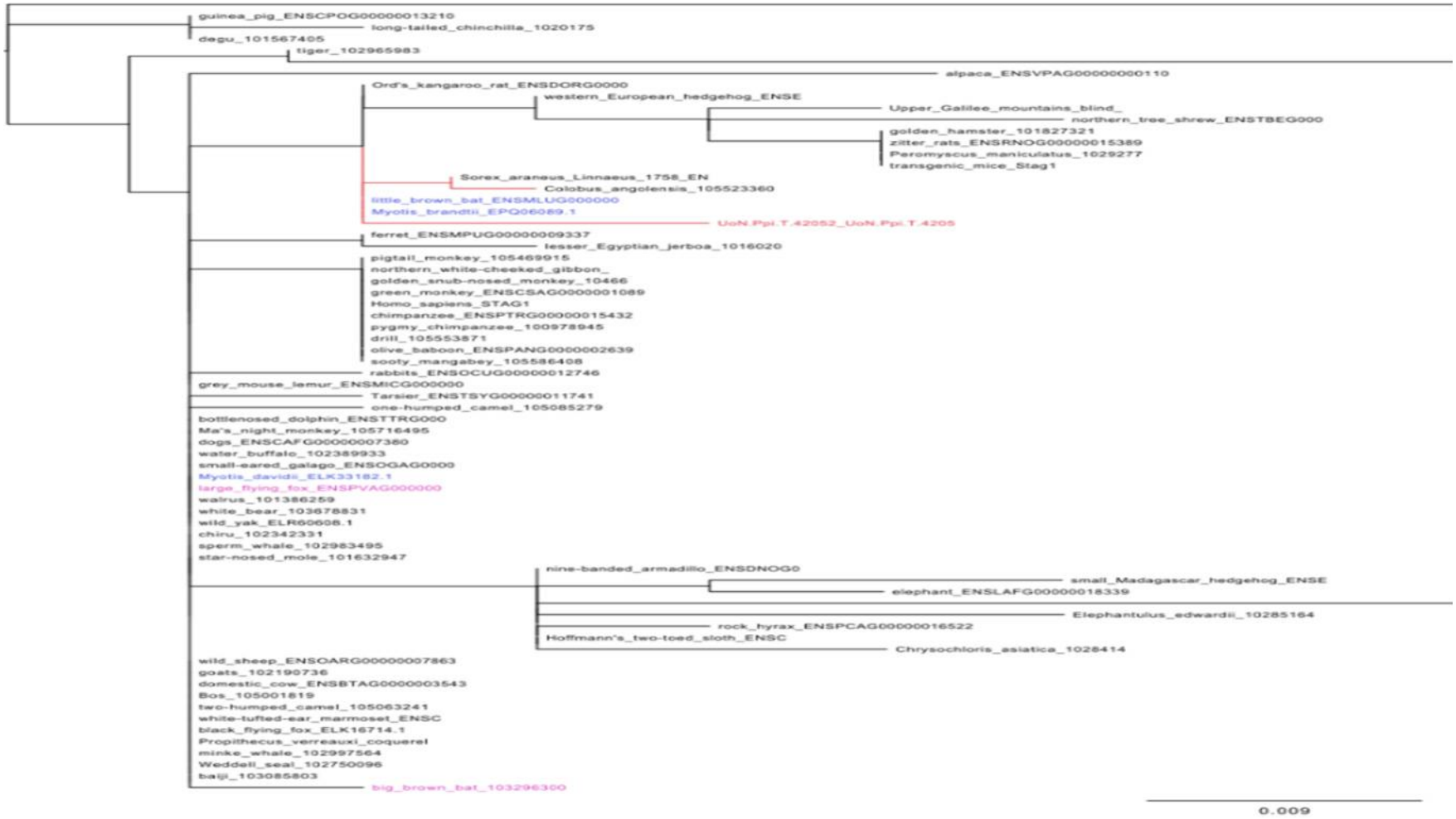


Figure 3.19 D: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of **FoxP2** protein (gene) of UoN bat (transcript ID indicated in red) which was closely related with brown bat. Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).



Figure 3.19 E: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of Otof protein (gene) of UoN bat (transcript ID indicated in red) which was closely related with brown bat (light blue). Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).

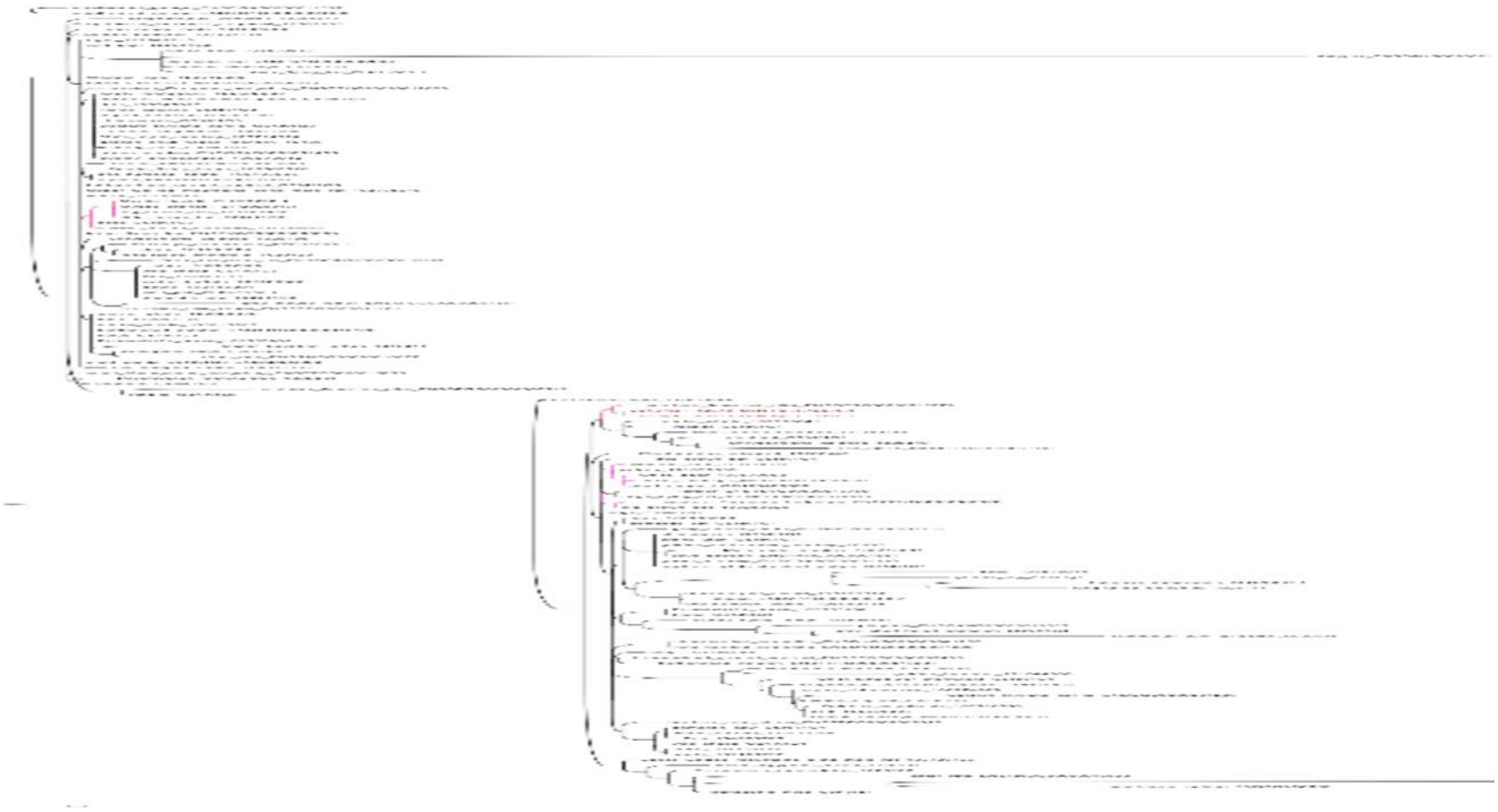


Figure 3.19 F: Bat gene expression phylogeny based on Neighbor-joining tree based pair wise distance matrices for brain transcriptome. Phylogenetic position of wnt 8a protein (gene) of UoN bat (transcript ID indicated in orange) which was closely related with brown bat. Phylogenetic tree showing inferred evolutionary relationships among the sequences of different species. Tree constructed by using PhyML v 3.0 (A fast program for searching for the maximum likelihood trees using nucleotide / protein sequence data) and bootstrapping done with 100 replicates (Guindon and Gascuel, 2003).

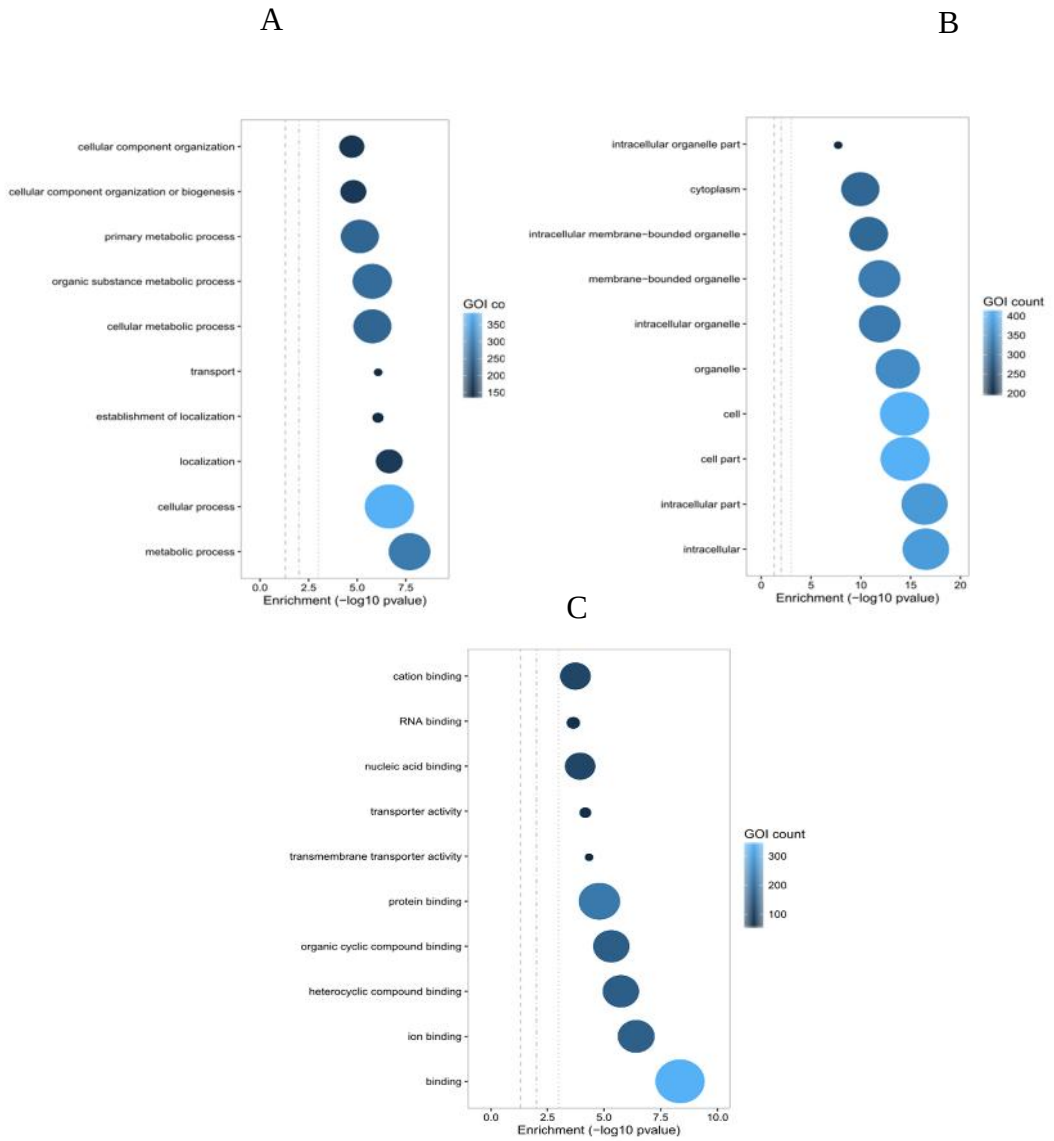


Figure 3.20 A, B, C: Gene Ontology. Visualization of GO terms of protein/genes of interest (GOI), 5X Median identified based on enrichment analysis. Bubble colour denotes the p value in terms of –log 10 p value; bubble size indicates the frequency (count) of the GO terms in the underlying GOA database. A) Visualization of **Biological processes** enriched in the genes, B) Visualization of **Cellular component** terms enriched in the genes, C) Visualization of **Molecular function** terms enriched in the genes.

In cellular components the enrichment of genes were in terms intracellular and membrane bound cell organelle which are involved in various brain functions. However, Molecular functions mostly cover the binding functions with various proteins, ions, cations, and heterocyclic compound little transport activity.

KEGG pathway enrichment (Figure 3.21) was conducted to identify the most enriched pathways for the echolocating genes of interest of UoN bat brain transcripts. For the highly expressed genes (TPM > 5 x median) terms ‘gap junction pathway’ was found as the most significant enriched pathway then two classical neural and brain function pathways, Long-term potentiation and Long-term depression.

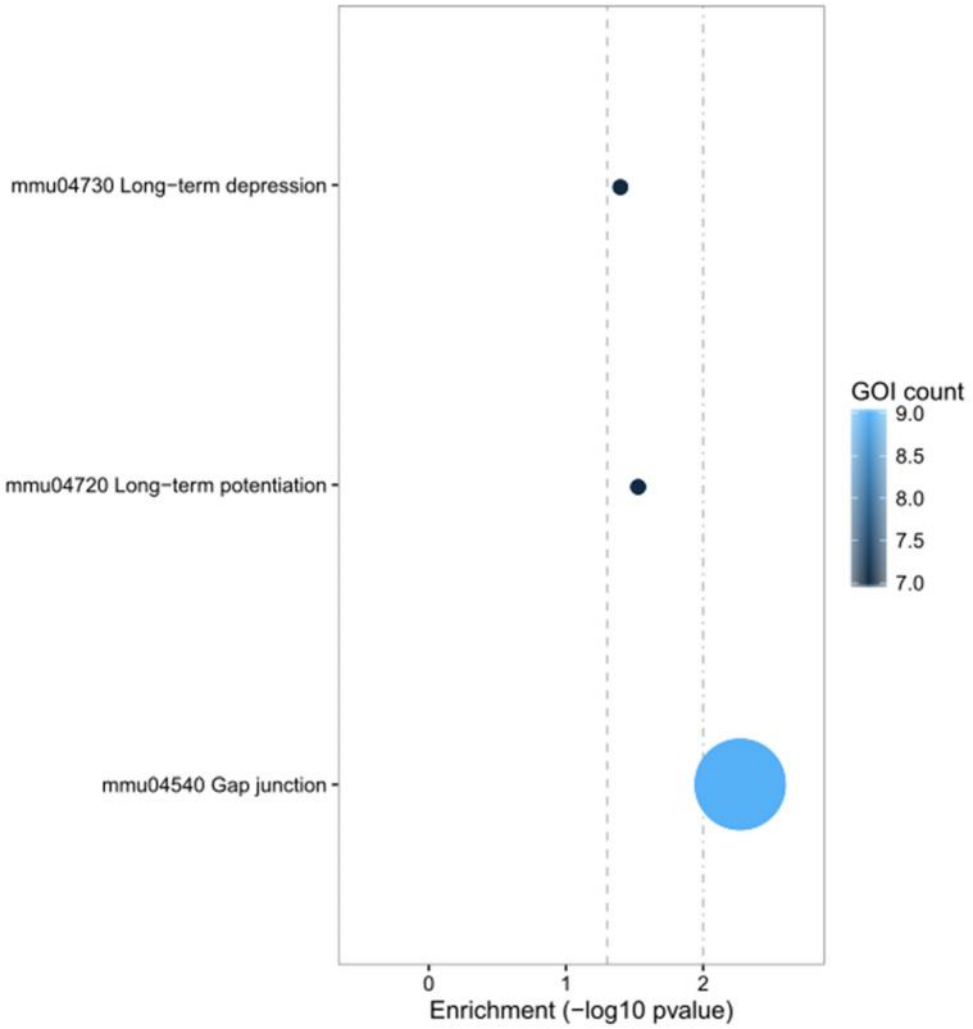


Figure 3.21 Scatter plot of **KEGG pathway enrichment** statistics and the most enrichment pathway of UoN Bat brain transcript (GOI, 5 X Median). Bubble colour denotes the p value in terms of –log 10 p value; bubble size indicates the frequency (count) of the GO terms in the underlying GOA database.

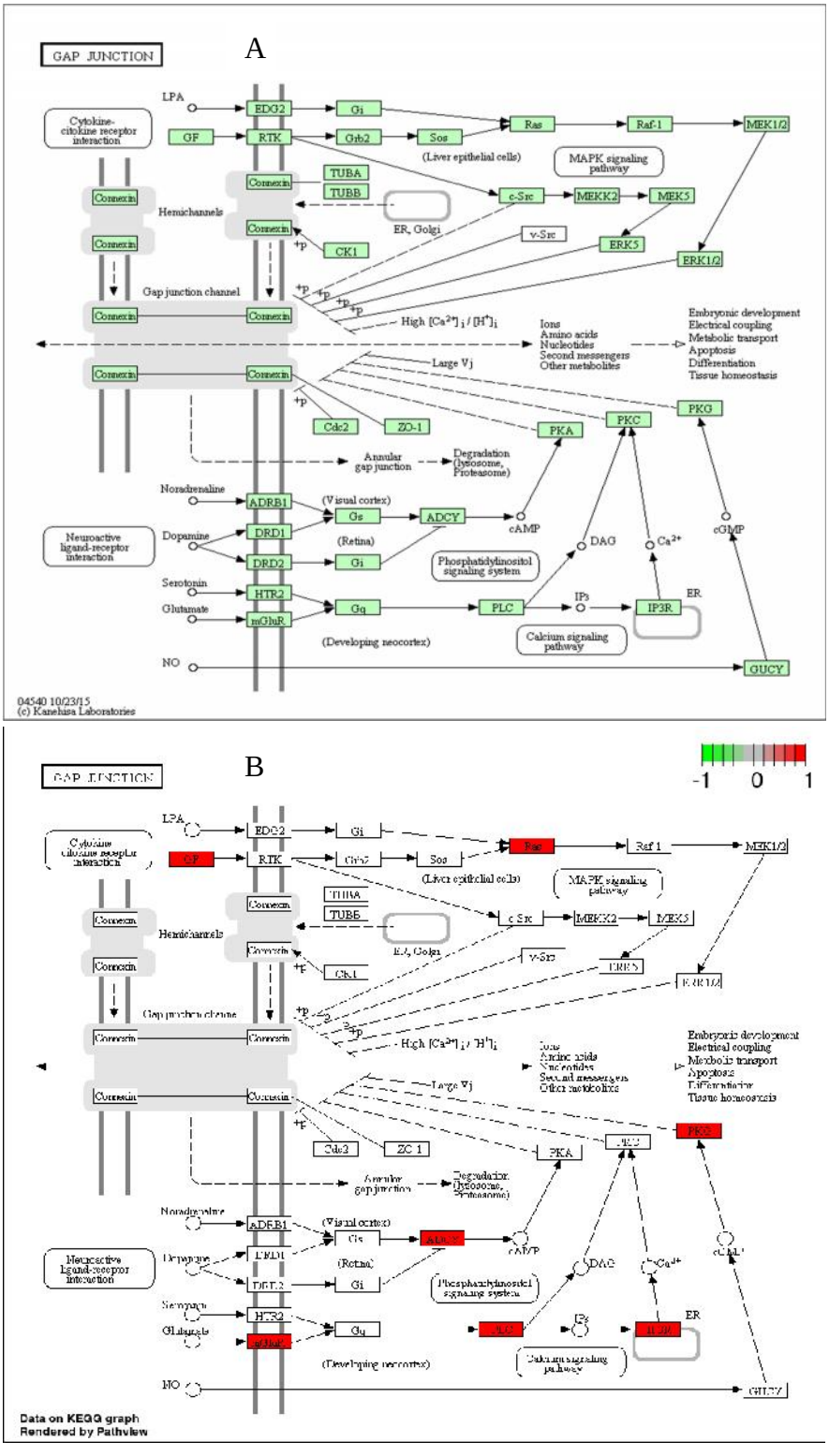


Figure 3.22. Visualize GAGE pathway analysis results using *pathview (KEGG)***.** A) KEGG view of **Gap Junction protein (Gjb),** 5 X Median. Green colour indicates the known interacting proteins associated in its pathway. B) KEGG view of Gap Junction protein (Gjb), 5 X Median. Red colour indicates the regulated proteins associated with Gjb pathway (*up regulated* or highly expressed).

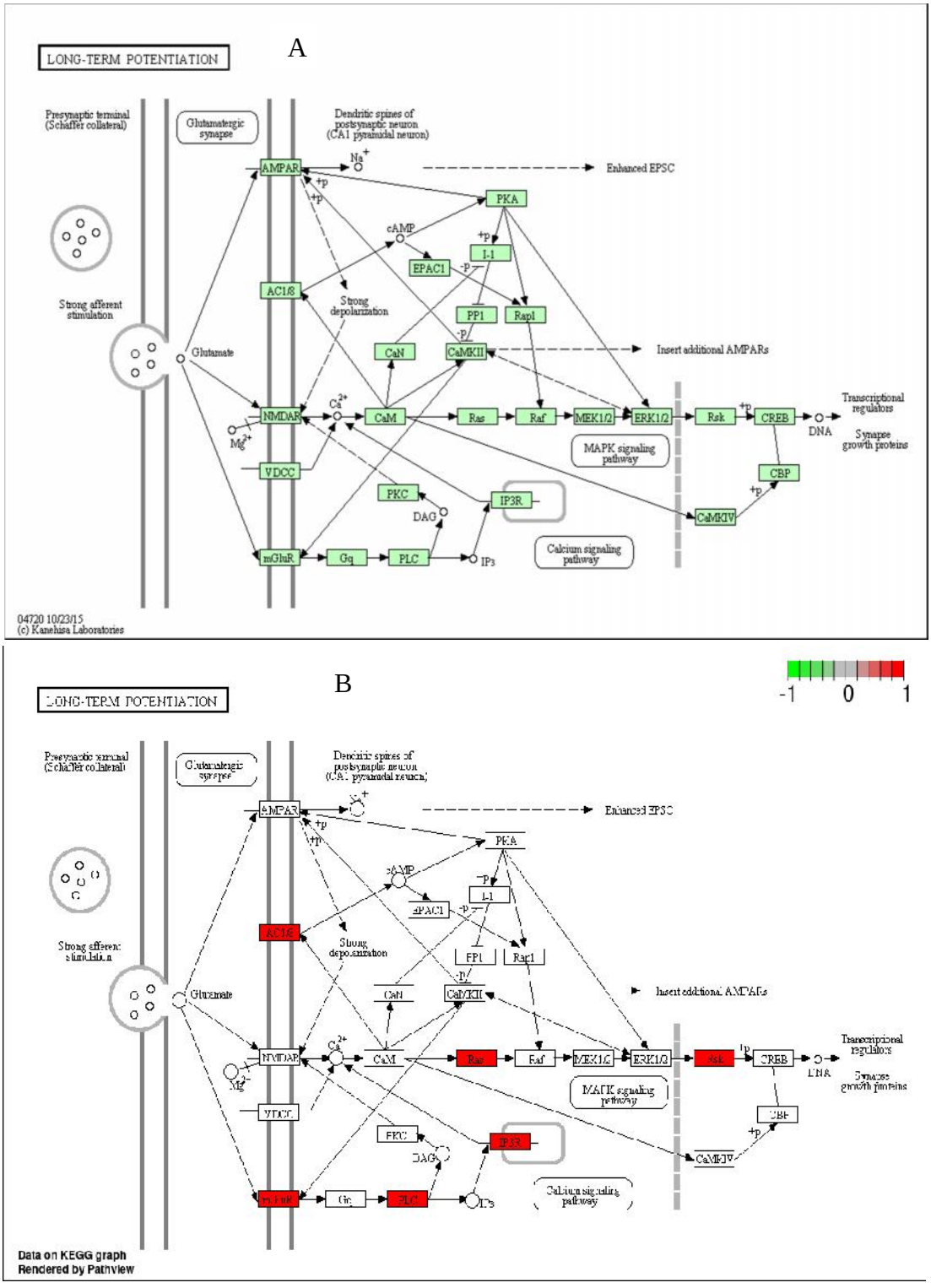


Figure 3.23 Visualize GAGE pathway analysis results using *pathview (KEGG).* A) KEGG view of **Long Term Potentiation (LTP)** gene, 5 X Median. Green colour indicates the known interacting proteins associated in its pathway. B) KEGG view of Long Term Potentiation (LTP) gene, 5 X Median. Red colour indicates the regulated proteins associated with LTP pathway (*up regulated* or highly expressed.

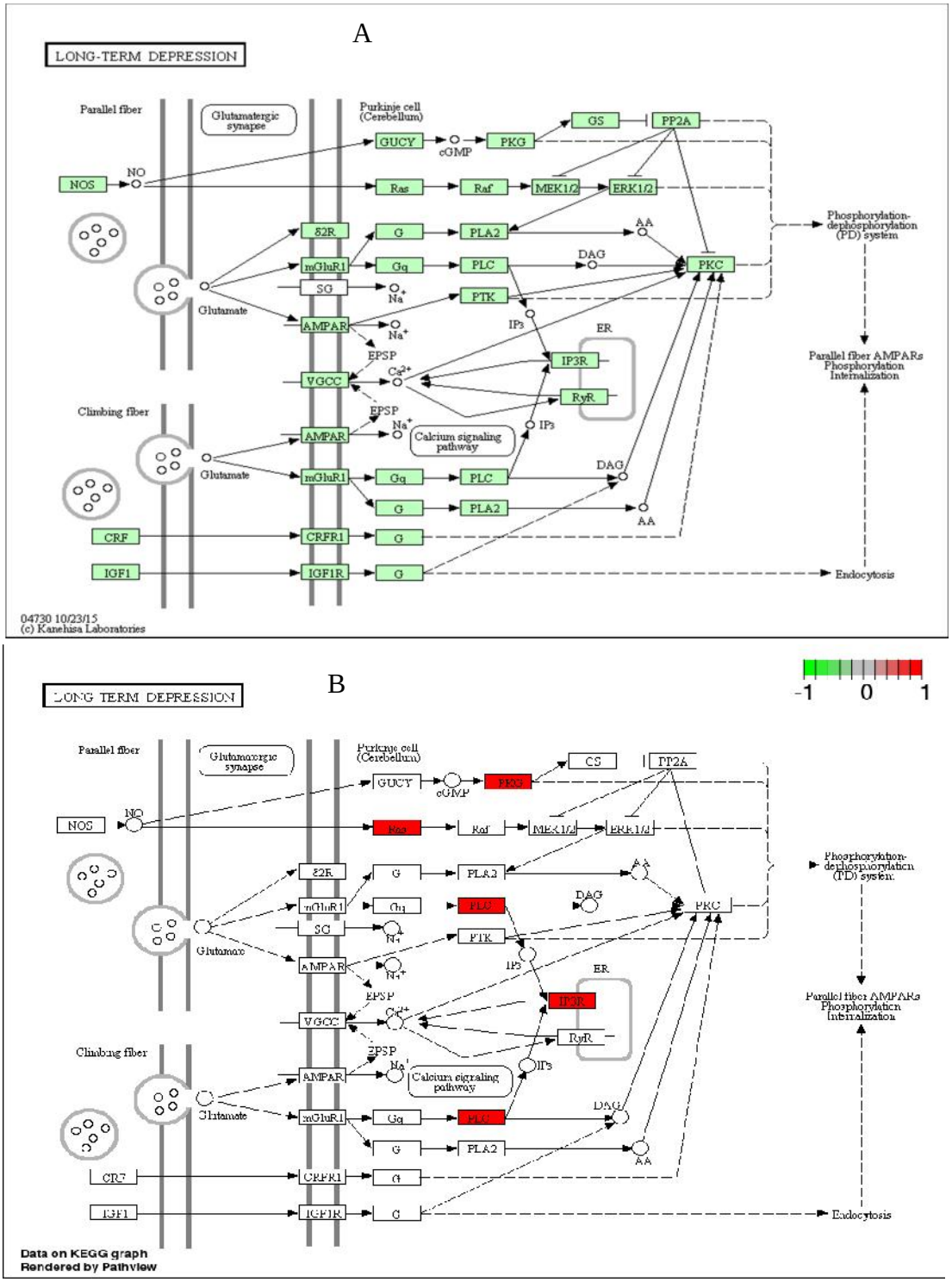


Figure 3.24. Visualize GAGE pathway analysis results using *pathview (KEGG).* A) KEGG view of **Long Term Depression (LTD)** gene, 5 X Median. Green colour indicates the known interacting proteins associated in its pathway (*down regulated*). B) KEGG view of Long Term Depression (LTD) gene, 5 X Median. Red colour indicates the regulated proteins associated with LTD pathway (*up regulated*).

4. DISCUSSION

This study represents the first Brain Transcriptome analysis of *Pipistrellus pisptrellus* (common pipistrelle). Differences of bat and other mammals are their ability to fly, hibernation, relative long life span, and gregarious roosting behaviour which exemplifies as quintessential vectors for the maintenance of viral infections and also as a highly promising model system for the study of echolocation and brain or neural function. And therefore, this study aims to sequence and annotate novel genes and transcripts from the bat brain.

Selection of RNA samples based on high quality and concentration

The selection of RNA samples was based on Agilent Bioanalyzer results. The quality of RNA was considered as good based on the RIN value above 6.5 which is ideal for acceptable for sequencing in this case, may be because of high accumulation of phenolic compounds or contamination. Based on these parameters we selected four samples for sequencing (Table 3.1).

Transcriptome Analysis

Transcriptomes of nonmodel organisms are best studied by *de novo* assembly, which helps in studying unique transcripts and expression levels in multiple organisms in parallel (Ward et al., 2012). In the present thesis, we studied the brain transcriptome of *Pipistrelle pipistrellus* (UoN Bat) using the Trinity platform (Haas et al., 2013) as the sequence data is very limited for this species so far in public databases and in literature.

The processed sequence reads of each biological replicate were then aligned back to the contigs from the assembly using HiSAT where we got good alignment.

Homologous and Orthologus gene prediction was performed using Inparanoid BLAST. Inparanoid is an algorithm designed to generate ortholog groups that include all inparalogs but no outparalogs. Comparative analysis of the organization of orthologous clusters is important for understanding the rules of genome structure and gene/protein function. The most typical orthology query is indeed pairwise, either one-to-one species queries i.e. ‘Find the UoN Bat ortholog to Ensembl bat (Myotis)’, or one-to-many,

eg, 'find all orthologs in any other species (human, mouse) with UoN bat'. Both queries are supported on the Inparanoid website (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). An advantage of pairwise nature of Inparanoid is that, the orthology assignments are not affected by other species. Count data of ortholog and type and percent data of ortholog and type, data (Tables 3.2 and 3.3) was generated between UoN Bat and Ensembl bat (Myotis), UoN Bat and Human and UoN Bat and Mouse orthologs (pairwise) and histogram plots were done by using command script in R (Appendix-1/Script). Euler plots were generated between orthologs of Ensemble bat (Myotis), Human and Mouse vs unique orthologs of UoN Bat pairwise, respectively.

Identification of Overlap of UoN transcripts with human, mouse and Myotis known genes was done by using Venny 2.1, an interactive tool for comparing lists with venn diagrams (http:// bioinfogp.csic.es/tools/venny/). The Euler plots (Figure 3.3) infers the predicted transcripts and all the open reading frames of UoN bat along with the known genes in Ensembl bat, human and mouse, respectively. The overlap of genes between the UoN Bat and Myotis, human and mouse were found to be 9381, 12767, 11201, respectively. These are likely to represent the core genes assembled with high confidence. However, the huge number of predicted genes and ORF were observed in the UoN bat may be comprising of both coding and noncoding genes, which may not filtered properly during the denovo assembly.

One of the simplest but most useful methods of genome wide orthologous comparison is to display the different and overlapping orthologous clusters in a Venn diagram, which allow quick visualization of relationships by revealing intersections (overlaps) and disjunctions (non-overlaps) for large biological data sets obtained from different species, and are often used in the whole genome analysis across species (Cho et al., 2013, Guyon et al., 2012, Argout et al., 2011).

Analysis of Genes of Interest

Twenty one genes were identified from the literature to be involved in echolocation. Only eight genes which were found in the UoN bat transcriptome. (Table 3.8 B and

Figure 3.8). Heat maps of expression (TPM) of echolocating genes between the four replicates of *Pipistrelle* were generated.

Out of 21 echolocating genes identified in this study only 8 were expressed in bat brain specific functions and out of 10 mitochondrial proteins only two were not expressed in the replicates, reason may be, that we have taken only brain tissue for this study. Analysing the other (13) echolocating genes from this study inferred that, these are predominantly expressed in other tissues rather in brain tissue (Appendix-3, Figures 1-12).

In contrast 2 mitochondrial proteins (Mbp and rab1b (Table: 3.9A) are expressed in brain tissue in human (Appendix-3, Figures13 and 14) but not expressed in the bat (pipistrelle) brain transcriptome which may reflect bat specific lack of expression or possible assembly error.

Using percentage similarity as a measure the echolocating genes and mitochondrial proteins of interest were found to be more conserved in all the pairwise comparison (Figures 3.9-3.11 and Figures 3.15-3.17, respectively) whereas the other genes in the genome show a wider range of divergence Higher similarity was seen in Mouse: Rat as is expected because of reduced time in terms of evolution between Mouse and Rat in comparison to other mammalian counterparts.

Adaptive Evolution

According to neutral theory, the major evolutionary changes are due to the random mutations which cause little or no change in the fitness. Alternatively, “adaptive evolution” is where and when evolution has shaped the genome by driving fixation of mutations which increases fitness. Protein coding genes were taken into consideration in this study to explore adaptive evolution.

Selection may alter gene-sequence-evolution either by constraining potential changes of amino acid sequences, synonymous substitutions (purifying or negative selection) or by favouring new and adaptive genetic variants nonsynonymous substitution (adaptive or positive selection). The rate of synonymous substitutions, dS (number of synonymous substitutions per synonymous site), and the rate of nonsynonymous

substitutions dN (number of nonsynonymous substitutions per nonsynonymous site) are determined, where dS is used as a baseline to compare relative fixation of dN by selection.

The ratio of these two rates (dN/dS= ω) gives a measure of selection pressure at the level of the codon. Selection strength, ω has three ranges, ie ω < 1 (Negative or Purifying selection), ω = 1 (Neutral evolution) and ω > 1 (Positive or Adaptive selection).

Synonymous substitutions (dS), nonsynonymous substitutions (dN) and selection strength (ω) was determined for echolocating genes of interest of UoN bat and compared to ensembl Bat, Human &. Cumulative frequency of all genes determined that the echolocating genes showed strong negative selection in Bat, Human and Rat.

While the role of gene expression and protein coding region in adaptation has often been studied separately, there is a need to understand the significance of these changes together in the process of adaption.

Expression (TPM) of all transcripts, and those with Human and Mouse orthologues, only Ensembl Bat orthologues and UoN bat Unique were determined (Figure 3.13). Whilst subtle the UoN bat unique transcripts show an elevated median expression value.

Phylogenetic study

High boot strap and statistical evidence in this study confirmed expectation that, UoN bat (Pipistrelle) of order Chiroptera is placed in the superorder Laurasiatheria (Figures 3.19 A-E). Laurasiatheria is one of the richest and most diverse super orders of placental mammals. Because this group had a rapid evolutionary radiation, the phylogenetic relationships among the six orders of Laurasiatheria remain a subject of intense speculation and several issues related to its phylogeny still unknown. Our results were supported by the recent molecular cladistics research done by Zhou et al. (2011) who confirmed the position of bats in the superorder Laurasiatheria, diverse group containing Carnivora and Artiodactyla and Zhang et al. (2012) placed Chiroptera as a sister taxon to the clade Perissodactyla (which includes horses and other odd-toed ungulates). However, this study does not confirm the exact aim of this section as there

may be some limitation in boot strapping and alignment of tree construction because of the absence of reference genome. However, our attempt in constructing the phylogenetic tree of the echolocating genes of interest of UoN bat sheds light on the phylogenetic relationships of Laurasiatheria. This is a significant case study in the evolutionary biology in understanding the exact position of bats in lieu of other orders in this super order. This insights the implications of biodiversity conservation.

Gene Ontology

Functional genomics is the most widely adopted in applied bioinformatics in which functional interpretation is the key step for the analysis of data. An efficient functional annotation of RNA- Seq is the major requirement for the successful application of these approaches.

Highly enriched pathways were identified. Gap Junction (mmu 04540), Long-term potentiation (mmu 04720) and long-term depression (mmu 04730) are the most enriched pathways of UoN Bat brain.

Gap junction

Cytosolic compartments of adjacent cells are directly mediated or communicated by intracellular channels called Gap junctions. Each gap junction channel is formed by docking of two 'hemichannels', each containing six connexins, contributed by each neighboring cell. These channels permit the direct transfer of small molecules including ions, amino acids, nucleotides, second messengers and other metabolites between adjacent cells. Gap junctional communication is essential for many physiological events, including embryonic development, electrical coupling, metabolic transport, apoptosis, and tissue homeostasis. Communication through Gap Junctions is sensitive to a variety of stimuli, including changes in the level of intracellular Ca^{2+}, pH, transjunctional applied voltage and phosphorylation/ dephosphorylation processes (http://www.genome.jp/kegg-bin/show_pathway?myb 04540).

Figure 3.22-A, represents the possible activation routes of different protein kinases different neuroactive ligand receptor interactions and known interacting proteins

involved in Cx43 found in astrocyte gap junctions and Cx36, found only in neuron (Rash et al., 2001) phosphorylation. Highly expressed or regulated genes of gap junction of UoN bat brain transcriptome in cytokine-cytokine receptor interaction includes GF, Growth factor, Ras, GTPase KRas (GTP coupled receptor) which regulates the MAPK signaling pathway and Long-term potentiation (Figure 3.22 B), a classical neural process involved in the mammals. Regulatory or highly expressed genes in neuroactive ligand-receptor interaction includes mGluR; metabotropic glutamate receptor 5, ADCY, Adenylate cyclase 4, PLC; Phospholipase C, IP3R; inositol 1,4,5-triphosphate receptor type 1, and PKG; Protein Kinase G, which regulates the cAMP signaling pathway, Long-term potentiation, Long-term depression. The interaction between gap junctions and chemical synaptic transmission observed in these developing networks represent a powerful mechanism by which activity across large groups of neurons can be correlated (Roerig and Feller, 2000).

Long term potentiation

As discussed earlier in this chapter Long-term potentiation is a classical neural process involved in brain function. Here we have elucidated the expression and regulatory mechanism of LTP by using GO and KEGG pathway analysis.

Long-term potentiation (LTP), a long-lasting increase in synaptic efficacy, is the molecular basis for learning and memory. Long-term potentiation pathway and its known interacting proteins and receptors are depicted in Figure 3.23 A and its regulatory, highly expressed (in terms of 5 X Median of TPM) genes are illustrated in Figure 3.23 B indicated in red colour. These genes include mGluR; metabotropic glutamate receptor, AC1/8; adenylate cyclase 1, PLC; Phospholipase C, Ras; GTPase KRas (GTP coupled receptor) IP3R; inositol 1,4,5-triphosphate receptor type 1, Rsk; ribosomal protein S6 kinase which are responsible in regulating the neuroactive ligand receptor interaction, calcium signaling pathway, MAPK signaling pathway, long-term potentiation and long-term depression.

Long term Depression

Cerebellar long-term depression (LTD), thought to be a molecular and cellular basis for cerebellar learning, is a process involving a decrease in the synaptic strength between parallel fiber (PF) and Purkinje cells (PCs) induced by the conjunctive activation of PFs and climbing fiber (CF). Multiple signal transduction pathways have been shown to be involved in this process (Figure 3.24 A). Activation of PFs terminating on spines in dendritic branchlets leads to glutamate release and activation of both AMPA and mGluR. Activation of CFs, which make multiple synaptic contacts on proximal dendrites, also via AMPA receptors, opens voltage-gated calcium channels (VGCCs) and causes a generalized influx of calcium. These cellular signals, generated from two different synaptic origins, trigger a cascade of events culminating in a phosphorylation-dependent, long-term reduction in AMPA receptor sensitivity at the PF-PC synapse. This may take place either through receptor internalization and/or through receptor desensitization (http://www.genome.jp/kegg-bin/show_pathway? hsa04730). Expression and regulation of Long-term depression gene (UoN bat brain) was studied in the above mentioned pathway. Highly expressed genes include Ras, PLC, IP3R and PKG (Figure 3.24 B, indicated in red colour) which are in turn responsible for the regulation of vital pathways necessary for brain and cellular functions like Calcium signaling, MAPK signaling and Long-term depression.

This information is useful in identifying the functional approach of the transcripts and the present study indicates the Brain transcriptome and focused with genes enriched in brain functions.

5. CONCLUSION

We comprehensively characterized the complete bat brain transcriptome of the common pipistrelle (*Pipistrelle pipistrellus*). This is the first study of its type and identified novel genes which are prominently involved in regulating the important classical neural processes of brain. This in turn gives profound resource for understanding the evolutionary perspective of these genes and also in brain associated diseases.

Potential Outcome and Future Study: Whole-genome assembly and functional genome annotation with RNA have become a realistic goal, and allow the identification of isoform- or allele-specific expression, coexpression and alternative splicing. Extending network analysis to several brain regions is potentially of high interest given the complexity of behavioural phenotypes. Understanding and identifying the adaptive selection of brain transcriptome genes in non-model organisms will provide insight into vertebrate brain evolution and potentially, identify genes involved in multifactorial human diseases. In order to extend this work we have isolated a high quality of RNA from the brain of zebrafish (whole brain tissue and three different regions of brain tissue separately in replicates). Samples have been submitted for NGS and analysis will follow protocols developed here.

6. BIBLIOGRAPHY

ALEXA, A. & RAHNENFUHRER, J. 2010. topGO: enrichment analysis for gene ontology. *R package version,* 2.

ALTENHOFF, A. M. & DESSIMOZ, C. 2012. Inferring orthology and paralogy. *Evolutionary Genomics: Statistical and Computational Methods, Volume 1*, 259-279.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* 215, 403-410.

ALVAREZ, C., GARCIA-MATA, R., BRANDON, E. & SZTUL, E. 2003. COPI recruitment is modulated by a Rab1b-dependent mechanism. *Molecular biology of the cell,* 14, 2116-2127.

ANDOLFATTO, P., DAVISON, D., EREZYILMAZ, D., HU, T. T., MAST, J., SUNAYAMA-MORITA, T. & STERN, D. L. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome research,* 21, 610-617.

ARGOUT, X., SALSE, J., AURY, J.-M., GUILTINAN, M. J., DROC, G., GOUZY, J., ALLEGRE, M., CHAPARRO, C., LEGAVRE, T. & MAXIMOVA, S. N. 2011. The genome of Theobroma cacao. *Nature genetics,* 43, 101-108.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S. & EPPIG, J. T. 2000. Gene Ontology: tool for the unification of biology. *Nature genetics,* 25, 25-29.

AU, K. F., JIANG, H., LIN, L., XING, Y. & WONG, W. H. 2010. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res,* 38, 4570-8.

BAIRD, N. A., ETTER, P. D., ATWOOD, T. S., CURREY, M. C., SHIVER, A. L., LEWIS, Z. A., SELKER, E. U., CRESKO, W. A. & JOHNSON, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one,* 3, e3376.

BARON-COHEN, S., COX, A., BAIRD, G., SWETTENHAM, J., NIGHTINGALE, N., MORGAN, K., DREW, A. & CHARMAN, T. 1996. Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry,* 168, 158-163.

BARTON, R. A. & HARVEY, P. H. 2000. Mosaic evolution of brain structure in mammals. *Nature*, 405, 1055-8.

BEN-YOSEF, T., BELYANTSEVA, I. A., SAUNDERS, T. L., HUGHES, E. D., KAWAMOTO, K., VAN ITALLIE, C. M., BEYER, L. A., HALSEY, K., GARDNER, D. J. & WILCOX, E. R. 2003. Claudin 14 knockout mice, a model for autosomal recessive deafness DFNB29, are deaf due to cochlear hair cell degeneration. *Human molecular genetics,* 12, 2049-2061.

BEST, T. L. 1997. Altringham, JD 1996. Bats: Biology and Behaviour. Oxford University Press, Inc., New York, 262 pp. ISBN 0-19-854075-2, price (hardcover), $65.00. *Journal of Mammalogy,* 78, 986-987.

BETTS, B. J. 1998. Effects of interindividual variation in echolocation calls on identification of big brown and silver-haired bats. *The Journal of Wildlife Management,* 1003-1010.

BOLZ, H., VON BREDERLOW, B., RAMÍREZ, A., BRYDA, E. C., KUTSCHE, K., NOTHWANG, H. G., SEELIGER, M., CABRERA, M. D. C.-S., VILA, M. C. & MOLINA, O. P. 2001. Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. *Nature genetics,* 27, 108-112.

BOUCHON, A. & CHAKRABORTY, T. 2014. Modified polynucleotides for the treatment of otic diseases and conditions. Google Patents.

BRODAL, A. 1981. *Neurological anatomy in relation to clinical medicine*, Oxford University Press, USA.

BRUDZYNSKI, S. M. & FLETCHER, N. H. 2010. .3-Rat ultrasonic vocalization: short-range communication. *Handbook of Behavioral Neuroscience,* 19, 69-76.

BUHL, E. & DANN, J. 1991. Cytoarchitecture, neuronal composition, and entorhinal afferents of the flying fox hippocampus. *Hippocampus,* 1, 131-152.

CARLBERG, C. 2000. Gene regulation by melatonin. *Annals of the New York Academy of Sciences,* 917, 387-396.

CHEN, F., MACKEY, A. J., STOECKERT, C. J. & ROOS, D. S. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic acids research,* 34, D363-D368.

CHEN, Q., HESTON, J. B., BURKETT, Z. D. & WHITE, S. A. 2013. Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. *Journal of Experimental Biology,* 216, 3682-3692.

CHIU, J. C., LEE, E. K., EGAN, M. G., SARKAR, I. N., CORUZZI, G. M. & DESALLE, R. 2006. OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics,* 22, 699-707.

CHO, Y. S., HU, L., HOU, H., LEE, H., XU, J., KWON, S., OH, S., KIM, H.-M., JHO, S. & KIM, S. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature communications,* 4.

CHOI, B. Y., AHMED, Z. M., RIAZUDDIN, S., BHINDER, M., SHAHZAD, M., HUSNAIN, T., GRIFFITH, A. & FRIEDMAN, T. 2009. Identities and frequencies of mutations of the otoferlin gene (OTOF) causing DFNB9 deafness in Pakistan. *Clinical genetics,* 75, 237-243.

CHOI, B. Y., KIM, D. H., CHUNG, T., CHANG, M., KIM, E. H., KIM, A. R., SEOK, J., CHANG, S. O., BOK, J. & KIM, D. 2013. Destabilization and Mislocalization of POU3F4 by C-Terminal Frameshift Truncation and Extension Mutation. *Human mutation,* 34, 309-316.

COHEN, S. & GREENBERG, M. E. 2008. Communication between the synapse and the nucleus in neuronal development, plasticity, and disease. *Annual review of cell and developmental biology,* 24, 183.

DAYHOFF, M. 1972. 0. 1972. *Atlas of protein sequence and structure,* 5.

DE COO, R. F., BUDDIGER, P., SMEETS, H. J. & VAN OOST, B. A. 1997. Molecular cloning and characterization of the human mitochondrial NADH:oxidoreductase 10-kDa gene (NDUFV3). *Genomics,* 45, 434-7.

DELMAGHANI, S., DEL CASTILLO, F. J., MICHEL, V., LEIBOVICI, M., AGHAIE, A., RON, U., VAN LAER, L., BEN-TAL, N., VAN CAMP, G. & WEIL, D. 2006. Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy. *Nature genetics,* 38, 770-778.

DELUCA, T. F., WU, I.-H., PU, J., MONAGHAN, T., PESHKIN, L., SINGH, S. & WALL, D. P. 2006. Roundup: a multi-genome repository of orthologs and evolutionary distances. *Bioinformatics,* 22, 2044-2046.

DENZ, C. R., NARSHI, A., ZAJDEL, R. W. & DUBE, D. K. 2004. Expression of a novel cardiac-specific tropomyosin isoform in humans. *Biochemical and biophysical research communications,* 320, 1291-1297.

DI PALMA, F., HOLME, R. H., BRYDA, E. C., BELYANTSEVA, I. A., PELLEGRINO, R., KACHAR, B., STEEL, K. P. & NOBEN-TRAUTH, K. 2001. Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. *Nature genetics,* 27, 103-107.

DOBRIN, R., ZHU, J., MOLONY, C., ARGMAN, C., PARRISH, M. L., CARLSON, S., ALLAN, M. F., POMP, D. & SCHADT, E. E. 2009. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome biology,* 10, 1.

DOUVILLE, P., ATANASOSKI, S., TOBLER, A., FONTANA, A. & SCHWAB, M. 1994. The brain-specific POU-box gene Brn4 is a sex-linked transcription factor located on the human and mouse X chromosomes. *Mammalian Genome,* 5, 180-182.

DUNCAN, R., BAZAR, L., MICHELOTTI, G., TOMONAGA, T., KRUTZSCH, H., AVIGAN, M. & LEVENS, D. 1994. A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev,* 8, 465-80.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32, 1792-1797.

EKMEKCIOGLU, C. 2006. Melatonin receptors in humans: biological role and clinical relevance. *Biomedicine & Pharmacotherapy,* 60, 97-108.

ELLEGREN, H., SMEDS, L., BURRI, R., OLASON, P. I., BACKSTRÖM, N., KAWAKAMI, T., KÜNSTNER, A., MÄKINEN, H., NADACHOWSKA-BRZYSKA, K. & QVARNSTRÖM, A. 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature,* 491, 756-760.

ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S. & MITCHELL, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one,* 6, e19379.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *American Naturalist,* 1-15.

FENTON, M., ACHARYA, L., AUDET, D., HICKEY, M., MERRIMAN, C., OBRIST, M., SYME, D. & ADKINS, B. 1992. Phyllostomid bats (Chiroptera: Phyllostomidae) as indicators of habitat disruption in the Neotropics. *Biotropica,* 440-446.

FENTON, M., PORTFORS, C., RAUTENBACH, I. & WATERMAN, J. 1998. Compromises: sound frequencies used in echolocation by aerial-feeding bats. *Canadian Journal of Zoology,* 76, 1174-1182.

FENTON, M. B. 1999. Describing the echolocation calls and behaviour of bats. *Acta Chiropterologica,* 1, 127-136.

FERLAND, R. J., CHERRY, T. J., PREWARE, P. O., MORRISEY, E. E. & WALSH, C. A. 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *Journal of Comparative Neurology,* 460, 266-279.

FISH, U. & SERVICE, W. 2012. *US Fish and Wildlife Service land-based wind energy guidelines*, US Fish and Wildlife Service.

FISHER, S. E. & MARCUS, G. F. 2006. The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics,* 7, 9-20.

FITCH, W. M. 1970. Distinguishing homologous from analogous proteins. *Systematic Biology,* 19, 99-113.

FRANCEY, L. J., CONLIN, L. K., KADESCH, H. E., CLARK, D., BERRODIN, D., SUN, Y., GLESSNER, J., HAKONARSON, H., JALAS, C. & LANDAU, C. 2012. Genome-wide SNP genotyping identifies the Stereocilin (STRC) gene as a major contributor to pediatric bilateral sensorineural hearing impairment. *American Journal of Medical Genetics Part A,* 158, 298-308.

FULLARD, J. & DAWSON, J. 1997. The echolocation calls of the spotted bat Euderma maculatum are relatively inaudible to moths. *Journal of Experimental Biology,* 200, 129-137.

FUZESSERY, Z. M., RAZAK, K. A. & WILLIAMS, A. J. 2011. Multiple mechanisms shape selectivity for FM sweep rate and direction in the pallid bat inferior colliculus and auditory cortex. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol,* 197, 615-23.

GANNON, W. L., SHERWIN, R. E. & HAYMOND, S. 2003. On the importance of articulating assumptions when conducting acoustic studies of habitat use by bats. *Wildlife Society Bulletin,* 31, 45-61.

GARRETT, S. C. & ROSENTHAL, J. J. 2012. A role for A-to-I RNA editing in temperature adaptation. *Physiology,* 27, 362-369.

GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular biology and evolution,* 14, 685-695.

GITTELMAN, J. X., LI, N. & POLLAK, G. D. 2009. Mechanisms underlying directional selectivity for frequency-modulated sweeps in the inferior colliculus revealed by in vivo whole-cell recordings. *J Neurosci,* 29, 13030-41.

GLEZER, I., SIMARD, A. & RIVEST, S. 2007. Neuroprotective role of the innate immune system by microglia. *Neuroscience,* 147, 867-883.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R. & ZENG, Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology,* 29, 644-652.

GRIFFIN, D. R. 1944. Echolocation by blind men, bats and radar. *Science,* 100, 589-590.

GRIFFIN, D. R. 1958. Listening in the dark: the acoustic orientation of bats and men.

GRIFFIN, D. R., WEBSTER, F. A. & MICHAEL, C. R. 1960. The echolocation of flying insects by bats. *Animal behaviour,* 8, 141-154.

GUINDON, S. & GASCUEL, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology,* 52, 696-704.

GUTTMACHER, A. E. & COLLINS, F. S. 2002. Genomic medicine—a primer. *New England Journal of Medicine,* 347, 1512-1520.

GUTTMAN, M., GARBER, M., LEVIN, J. Z., DONAGHEY, J., ROBINSON, J., ADICONIS, X., FAN, L., KOZIOL, M. J., GNIRKE, A. & NUSBAUM, C. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature biotechnology,* 28, 503-510.

GUYON, R., RAKOTOMANGA, M., AZZOUZI, N., COUTANCEAU, J. P., BONILLO, C., D'COTTA, H., PEPEY, E., SOLER, L., RODIER-GOUD, M. & D'HONT, A. 2012. A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC genomics,* 13, 1.

HAAS, B. J., PAPANICOLAOU, A., YASSOUR, M., GRABHERR, M., BLOOD, P. D., BOWDEN, J., COUGER, M. B., ECCLES, D., LI, B. & LIEBER, M. 2013.

De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8, 1494-1512.

HAESLER, S., ROCHEFORT, C., GEORGI, B., LICZNERSKI, P., OSTEN, P. & SCHARFF, C. 2007. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol*, 5, e321.

HANAI, J.-I., GLOY, J., KARUMANCHI, S. A., KALE, S., TANG, J., HU, G., CHAN, B., RAMCHANDRAN, R., JHA, V. & SUKHATME, V. P. 2002. Endostatin is a potential inhibitor of Wnt signaling. *The Journal of cell biology*, 158, 529-539.

HARVEY, P. H. & PAGEL, M. D. 1991. *The comparative method in evolutionary biology*, Oxford university press Oxford.

HENIKOFF, S., GREENE, E. A., PIETROKOVSKI, S., BORK, P., ATTWOOD, T. K. & HOOD, L. 1997. Gene families: the taxonomy of protein paralogs and chimeras. *Science*, 278, 609-614.

HIRSCH, D. J., JINDAL, K. K., WONG, P. & FRASER, A. D. 2001. A simple method to estimate the required dialysis time for cases of alcohol poisoning. *Kidney international*, 60, 2021-2024.

HO, M.-R., CHEN, C.-H. & LIN, W.-C. 2010. Gene-oriented ortholog database: a functional comparison platform for orthologous loci. *Database*, 2010, baq002.

HUANG, X., POELCHAU, M. F. & ARMBRUSTER, P. A. 2015. Global transcriptional dynamics of diapause induction in non-blood-fed and blood-fed Aedes albopictus. *PLoS Negl Trop Dis*, 9, e0003724.

HUERTA-CEPAS, J., CAPELLA-GUTIERREZ, S., PRYSZCZ, L. P., DENISOV, I., KORMES, D., MARCET-HOUBEN, M. & GABALDÓN, T. 2011. PhylomeDB v3. 0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic acids research*, 39, D556-D560.

HUERTA-CEPAS, J., CAPELLA-GUTIERREZ, S., PRYSZCZ, L. P., MARCET-HOUBEN, M. & GABALDÓN, T. 2013. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic acids research*, gkt1177.

IANCU, O. D., KAWANE, S., BOTTOMLY, D., SEARLES, R., HITZEMANN, R. & MCWEENEY, S. 2012. Utilizing RNA-Seq data for *de novo* coexpression network inference. *Bioinformatics,* 28, 1592-1597.

INOSHITA, A., IIZUKA, T., OKAMURA, H.-O., MINEKAWA, A., KOJIMA, K., FURUKAWA, M., KUSUNOKI, T. & IKEDA, K. 2008. Postnatal development of the organ of Corti in dominant-negative Gjb2 transgenic mice. *Neuroscience,* 156, 1039-1047.

IUCN, I. 2012. Red List of Threatened Species: Version 2011.2.

JAVED, M. & WRIGHT, R. 1991. Determination of pentose phosphate and Embden-Meyerhof pathway activities in bovine embryos. *Theriogenology,* 35, 1029-1037.

JONES, G. 2005. Echolocation. *Current Biology,* 15, R484-R488.

JONES, G. 2010. Molecular evolution: gene convergence in echolocating mammals. *Current Biology,* 20, R62-R64.

JONES, G., JACOBS, D. S., KUNZ, T. H., WILLIG, M. R. & RACEY, P. A. 2009. Carpe noctem: the importance of bats as bioindicators. *Endangered Species Research,* 8, 93-115.

JONES, G. & TEELING, E. C. 2006. The evolution of echolocation in bats. *Trends in Ecology & Evolution,* 21, 149-156.

JUNUTULA, J. R., DE MAZIÉRE, A. M., PEDEN, A. A., ERVIN, K. E., ADVANI, R. J., VAN DIJK, S. M., KLUMPERMAN, J. & SCHELLER, R. H. 2004. Rab14 is involved in membrane trafficking between the Golgi complex and endosomes. *Molecular biology of the cell,* 15, 2218-2229.

KALCOUNIS-RUEPPELL, M. C., PAYNE, V., HUFF, S. & BOYKO, A. 2007. Effects of wastewater treatment plant effluent on bat foraging ecology in an urban stream system. *Biological Conservation,* 138, 120-130.

KALKO, E. K. 1995. Insect pursuit, prey capture and echolocation in pipestirelle bats (Microchiroptera). *Animal Behaviour,* 50, 861-880.

KAMMERER, R., RÜTTIGER, L., RIESENBERG, R., SCHÄUBLE, C., KRUPAR, R., KAMP, A., SUNAMI, K., EISENRIED, A., HENNENBERG, M. & GRUNERT, F. 2012. Loss of mammal-specific tectorial membrane component carcinoembryonic antigen cell adhesion molecule 16 (CEACAM16) leads to

hearing impairment at low and high frequencies. *Journal of Biological Chemistry*, 287, 21584-21598.

KARANYI, Z., HOLB, I., HORNOK, L., POCSI, I. & MISKEI, M. 2013. FSRD: fungal stress response database. *Database*, 2013, bat037.

KAZMIERCZAK, P., SAKAGUCHI, H., TOKITA, J., WILSON-KUBALEK, E. M., MILLIGAN, R. A., MÜLLER, U. & KACHAR, B. 2007. Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. *Nature*, 449, 87-91.

KENT, W. J. 2002. BLAT--the BLAST-like alignment tool. *Genome Res*, 12, 656-64.

KEREN, H., LEV-MAOR, G. & AST, G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*, 11, 345-355.

KHARKOVETS, T., DEDEK, K., MAIER, H., SCHWEIZER, M., KHIMICH, D., NOUVIAN, R., VARDANYAN, V., LEUWER, R., MOSER, T. & JENTSCH, T. J. 2006. Mice with altered KCNQ4 K+ channels implicate sensory outer hair cells in human progressive deafness. *The EMBO journal*, 25, 642-652.

KHARKOVETS, T., HARDELIN, J.-P., SAFIEDDINE, S., SCHWEIZER, M., EL-AMRAOUI, A., PETIT, C. & JENTSCH, T. J. 2000. KCNQ4, a K+ channel mutated in a form of dominant deafness, is expressed in the inner ear and the central auditory pathway. *Proceedings of the National Academy of Sciences*, 97, 4333-4338.

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12, 357-360.

KIRWAN, J. D., BEKAERT, M., COMMINS, J. M., DAVIES, K. T., ROSSITER, S. J. & TEELING, E. C. 2013. A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. *Evolutionary applications*, 6, 412-422.

KOONIN, E. V. 2005. Orthologs, paralogs, and evolutionary genomics 1. *Annu. Rev. Genet.*, 39, 309-338.

KORNBLIHTT, A. R., DE LA MATA, M., FEDEDA, J. P., MUNOZ, M. J. & NOGUES, G. 2004. Multiple links between transcription and splicing. *Rna*, 10, 1489-1498.

KOSKI, L. B. & GOLDING, G. B. 2001. The closest BLAST hit is often not the nearest neighbor. *Journal of Molecular Evolution,* 52, 540-542.

KRIVENTSEVA, E. V., RAHMAN, N., ESPINOSA, O. & ZDOBNOV, E. M. 2008. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic acids research,* 36, D271-D275.

KUNZ, T. H., ARNETT, E. B., ERICKSON, W. P., HOAR, A. R., JOHNSON, G. D., LARKIN, R. P., STRICKLAND, M. D., THRESHER, R. W. & TUTTLE, M. D. 2007. Ecological impacts of wind energy development on bats: questions, research needs, and hypotheses. *Frontiers in Ecology and the Environment,* 5, 315-324.

KUNZ, T. H., LUMSDEN, L. F., KUNZ, T. & FENTON, M. 2003. Ecology of cavity and foliage roosting bats. *Bat ecology*, 3-89.

KURIMA, K., PETERS, L. M., YANG, Y., RIAZUDDIN, S., AHMED, Z. M., NAZ, S., ARNAUD, D., DRURY, S., MO, J. & MAKISHIMA, T. 2002. Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. *Nature genetics,* 30, 277-284.

LAMICHHANEY, S., BARRIO, A. M., RAFATI, N., SUNDSTRÖM, G., RUBIN, C.-J., GILBERT, E. R., BERGLUND, J., WETTERBOM, A., LAIKRE, L. & WEBSTER, M. T. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences,* 109, 19345-19350.

LASSMANN, T., HAYASHIZAKI, Y. & DAUB, C. O. 2009. TagDust—a program to eliminate artifacts from next generation sequencing data. *Bioinformatics,* 25, 2839-2840.

LAWRENCE, B. & SIMMONS, J. 1982. Echolocation in bats: the external ear and perception of the vertical positions of targets. *Science,* 218, 481-483.

LEE, K., ANSAR, M., ANDRADE, P. B., KHAN, B., SANTOS-CORTEZ, R. L. P., AHMAD, W. & LEAL, S. M. 2012. Novel CLDN14 mutations in Pakistani families with autosomal recessive non-syndromic hearing loss. *American Journal of Medical Genetics Part A,* 158, 315-321.

LENZ, T. L., EIZAGUIRRE, C., ROTTER, B., KALBE, M. & MILINSKI, M. 2013. Exploring local immunological adaptation of two stickleback ecotypes by

experimental infection and transcriptome-wide digital gene expression analysis. *Molecular ecology,* 22, 774-786.

LI, G., WANG, J., ROSSITER, S. J., JONES, G., COTTON, J. A. & ZHANG, S. 2008. The hearing gene Prestin reunites echolocating bats. *Proceedings of the National Academy of Sciences,* 105, 13959-13964.

LI, G., WANG, J., ROSSITER, S. J., JONES, G. & ZHANG, S. 2007. Accelerated FoxP2 evolution in echolocating bats. *PLoS one,* 2, e900.

LI, L., NAGAI, K. & YIN, F. 2016. Progress in cold roll bonding of metals. *Science and Technology of Advanced Materials*.

LI, L., STOECKERT, C. J. & ROOS, D. S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research,* 13, 2178-2189.

LI, W. & GODZIK, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22, 1658-1659.

LI, Y., LIU, Z., SHI, P. & ZHANG, J. 2010. The hearing gene Prestin unites echolocating bats and whales. *Current Biology,* 20, R55-R56.

LIBERMAN, M. C., GAO, J., HE, D. Z., WU, X., JIA, S. & ZUO, J. 2002. Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature,* 419, 300-304.

LINDBLAD-TOH, K., GARBER, M., ZUK, O., LIN, M. F., PARKER, B. J., WASHIETL, S., KHERADPOUR, P., ERNST, J., JORDAN, G. & MAUCELI, E. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature,* 478, 476-482.

LIU, G.-H., YI, F., SUZUKI, K., QU, J. & BELMONTE, J. C. I. 2012. Induced neural stem cells: a new tool for studying neural development and neurological disorders. *Cell research,* 22, 1087.

LIU, M., ADELMAN, Z. N., MYLES, K. M. & ZHANG, L. 2014. A Transcriptome Post-Scaffolding Method for Assembling High Quality Contigs. *Computational Biology Journal,* 2014.

LIU, Y., COTTON, J. A., SHEN, B., HAN, X., ROSSITER, S. J. & ZHANG, S. 2010. Convergent sequence evolution between echolocating bats and dolphins. *Current Biology,* 20, R53-R54.

LIU, Z., LI, S., WANG, W., XU, D., MURPHY, R. W. & SHI, P. 2011. Parallel evolution of KCNQ4 in echolocating bats. *PLoS One,* 6, e26618.

LUO, W., FRIEDMAN, M. S., SHEDDEN, K., HANKENSON, K. D. & WOOLF, P. J. 2009. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC bioinformatics,* 10, 1.

LUSTIG, L. R., PENG, H., HIEL, H., YAMAMOTO, T. & FUCHS, P. A. 2001. Molecular cloning and mapping of the human nicotinic acetylcholine receptor alpha10 (CHRNA10). *Genomics,* 73, 272-83.

MANOR, U., DISANZA, A., GRATI, M. H., ANDRADE, L., LIN, H., DI FIORE, P. P., SCITA, G. & KACHAR, B. 2011. Regulation of stereocilia length by myosin XVa and whirlin depends on the actin-regulatory protein Eps8. *Current Biology,* 21, 167-172.

MARTIN, J., BRUNO, V. M., FANG, Z., MENG, X., BLOW, M., ZHANG, T., SHERLOCK, G., SNYDER, M. & WANG, Z. 2010. Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC genomics,* 11, 1.

MARTIN, J. A. & WANG, Z. 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics,* 12, 671-682.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal,* 17, pp. 10-12.

MARTINVALET, D., DYKXHOORN, D. M., FERRINI, R. & LIEBERMAN, J. 2008. Granzyme A cleaves a mitochondrial complex I protein to initiate caspase-independent cell death. *Cell,* 133, 681-92.

MEDELLÍN, R. A., EQUIHUA, M. & AMIN, M. A. 2000. Bat diversity and abundance as indicators of disturbance in Neotropical rainforests. *Conservation Biology,* 14, 1666-1675.

METZNER, W. & SCHULLER, G. 2010. .4-Vocal control in echolocating bats. *Handbook of behavioral neuroscience,* 19, 403-415.

MINOWA, O., IKEDA, K., SUGITANI, Y., OSHIMA, T., NAKAI, S., KATORI, Y., SUZUKI, M., FURUKAWA, M., KAWASE, T. & ZHENG, Y. 1999. Altered cochlear fibrocytes in a mouse model of DFN3 nonsyndromic deafness. *Science,* 285, 1408-1411.

MODREK, B. & LEE, C. 2002. A genomic view of alternative splicing. *Nature genetics*, 30, 13-19.

MONTIEL, J. F., VASISTHA, N. A., GARCIA-MORENO, F. & MOLNÁR, Z. 2016. From sauropsids to mammals and back: new approaches to comparative cortical development. *Journal of Comparative Neurology,* 524, 630-645.

MORATELLI, R. & CALISHER, C. H. 2015. Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Memórias do Instituto Oswaldo Cruz,* 110, 1-22.

MORETON, J., IZQUIERDO, A. & EMES, R. D. 2015. Assembly, assessment, and availability of *de novo* generated eukaryotic transcriptomes. *Frontiers in genetics,* 6.

NAL, N., AHMED, Z. M., ERKAL, E., ALPER, Ö. M., LÜLECI, G., DINÇ, O., WARYAH, A. M., AIN, Q., TASNEEM, S. & HUSNAIN, T. 2007. Mutational spectrum of MYO15A: the large N-terminal extension of myosin XVA is required for hearing. *Human mutation,* 28, 1014-1019.

NAUMANN, R. T. & KANWAL, J. S. 2011. Basolateral amygdala responds robustly to social calls: spiking characteristics of single unit activity. *Journal of neurophysiology,* 105, 2389-2404.

NEUWEILER, G. 2000. *The biology of bats,* Oxford University Press on Demand.

NOBEN-TRAUTH, K., ZHENG, Q. Y. & JOHNSON, K. R. 2003. Association of cadherin 23 with polygenic inheritance and genetic modification of sensorineural hearing loss. *Nature genetics,* 35, 21-23.

NOWAK, R. M. 1994. *Walker's bats of the world*, JHU Press.

O'BRIEN, K. P., REMM, M. & SONNHAMMER, E. L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic acids research,* 33, D476-D480.

O'BRIEN, K. P., WESTERLUND, I. & SONNHAMMER, E. L. 2004. OrthoDisease: a database of human disease orthologs. *Human mutation,* 24, 112-119.

O'FARRELL, M. J., MILLER, B. W. & GANNON, W. L. 1999. Qualitative identification of free-flying bats using the Anabat detector. *Journal of Mammalogy,* 80, 11-23.

OHLSSON TEAGUE, E. M. C., VAN DER HOEK, K. H., VAN DER HOEK, M. B., PERRY, N., WAGAARACHCHI, P., ROBERTSON, S. A., PRINT, C. G. & HULL, L. M. 2009. MicroRNA-regulated pathways associated with endometriosis. *Molecular endocrinology,* 23, 265-275.

PAN, B., GÉLÉOC, G. S., ASAI, Y., HORWITZ, G. C., KURIMA, K., ISHIKAWA, K., KAWASHIMA, Y., GRIFFITH, A. J. & HOLT, J. R. 2013. TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear. *Neuron,* 79, 504-515.

PENEL, S., ARIGON, A.-M., DUFAYARD, J.-F., SERTIER, A.-S., DAUBIN, V., DURET, L., GOUY, M. & PERRIÈRE, G. 2009. Databases of homologous gene families for comparative genomics. *BMC bioinformatics,* 10, 1.

POELCHAU, M. F., REYNOLDS, J. A., ELSIK, C. G., DENLINGER, D. L. & ARMBRUSTER, P. A. 2013. Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, Aedes albopictus. *Proceedings of the Royal Society of London B: Biological Sciences,* 280, 20130143.

PREUSS, T. M. 2000. Taking the measure of diversity: comparative alternatives to the model-animal paradigm in cortical neuroscience. *Brain Behav Evol,* 55, 287-99.

PROBST, F. J., FRIDELL, R. A., RAPHAEL, Y., SAUNDERS, T. L., WANG, A., LIANG, Y., MORELL, R. J., TOUCHMAN, J. W., LYONS, R. H. & NOBEN-TRAUTH, K. 1998. Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene. *Science,* 280, 1444-1447.

PRYSZCZ, L. P., HUERTA-CEPAS, J. & GABALDÓN, T. 2010. MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic acids research,* gkq953.

PUMPHREY, R. 1950. Upper limit of frequency for human hearing.

RACEY, P. 2003. Conservation ecology of bats. *Bat ecology*.

RASH, J. E., YASUMURA, T., DUDEK, F. E. & NAGY, J. I. 2001. Cell-specific expression of connexins and evidence of restricted gap junctional coupling between glial cells and between neurons. *The Journal of Neuroscience,* 21, 1983-2000.

REMM, M., STORM, C. E. & SONNHAMMER, E. L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of molecular biology,* 314, 1041-1052.

RICCI, J.-E., MUÑOZ-PINEDO, C., FITZGERALD, P., BAILLY-MAITRE, B., PERKINS, G. A., YADAVA, N., SCHEFFLER, I. E., ELLISMAN, M. H. & GREEN, D. R. 2004. Disruption of mitochondrial function during apoptosis is mediated by caspase cleavage of the p75 subunit of complex I of the electron transport chain. *Cell,* 117, 773-786.

RIO, D. C., ARES, M., HANNON, G. J. & NILSEN, T. W. 2010. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols,* 2010, pdb. prot5439.

RIZZO, J. & ROUCHKA, E. C. 2007. Review of phylogenetic tree construction. *University of Louisville Bioinformatics Laboratory Technical Report Series*, 2-7.

ROBERTSON, G., SCHEIN, J., CHIU, R., CORBETT, R., FIELD, M., JACKMAN, S. D., MUNGALL, K., LEE, S., OKADA, H. M. & QIAN, J. Q. 2010. *De novo* assembly and analysis of RNA-seq data. *Nature methods,* 7, 909-912.

ROERIG, B. & FELLER, M. B. 2000. Neurotransmitters and gap junctions in developing neural circuits. *Brain research reviews,* 32, 86-114.

ROGNES, T. 2011. Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC bioinformatics,* 12, 1.

ROLLINS, K. E., MEYERHOLZ, D. K., JOHNSON, G. D., CAPPARELLA, A. P. & LOEW, S. S. 2012. A forensic investigation into the etiology of bat mortality at a wind farm: barotrauma or traumatic injury? *Veterinary Pathology Online,* 49, 362-371.

SAFI, K., SEID, M. A. & DECHMANN, D. K. 2005. Bigger is not always better: when brains get smaller. *Biology Letters,* 1, 283-286.

SARKAR, I. N., PLANET, P. J., BAEL, T., STANLEY, S., SIDDALL, M., DESALLE, R. & FIGURSKI, D. H. 2002. Characteristic attributes in cancer microarrays. *Journal of biomedical informatics,* 35, 111-122.

SCHACTER, D. L., GUERIN, S. A. & JACQUES, P. L. S. 2011. Memory distortion: An adaptive perspective. *Trends in cognitive sciences,* 15, 467-474.

SCHRADERS, M., RUIZ-PALMERO, L., KALAY, E., OOSTRIK, J., DEL CASTILLO, F. J., SEZGIN, O., BEYNON, A. J., STROM, T. M., PENNINGS, R. J. & SECO, C. Z. 2012. Mutations of the gene encoding otogelin are a cause of autosomal-recessive nonsyndromic moderate hearing impairment. *The American Journal of Human Genetics,* 91, 883-889.

SCHWANDER, M., SCZANIECKA, A., GRILLET, N., BAILEY, J. S., AVENARIUS, M., NAJMABADI, H., STEFFY, B. M., FEDERE, G. C., LAGLER, E. A. & BANAN, R. 2007. A forward genetics screen in mice identifies recessive deafness traits and reveals that pejvakin is essential for outer hair cell function. *The Journal of neuroscience,* 27, 2163-2175.

SHAMIR, R. 1984. Current Research Interests. *University of California, Berkeley,* 1981.

SIEMENS, J., KAZMIERCZAK, P., REYNOLDS, A., STICKER, M., LITTLEWOOD-EVANS, A. & MÜLLER, U. 2002. The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. *Proceedings of the National Academy of Sciences,* 99, 14946-14951.

SIMÃO, F. A., WATERHOUSE, R. M., IOANNIDIS, P., KRIVENTSEVA, E. V. & ZDOBNOV, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* btv351.

SIMMLER, M.-C., COHEN-SALMON, M., EL-AMRAOUI, A., GUILLAUD, L., BENICHOU, J.-C., PETIT, C. & PANTHIER, J.-J. 2000. Targeted disruption of otog results in deafness and severe imbalance. *Nature genetics,* 24, 139-143.

SIMMONS, N. B. 2005. Order chiroptera. *Mammal species of the world: a taxonomic and geographic reference,* 1, 312-529.

SONNHAMMER, E. L. & KOONIN, E. V. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *TRENDS in Genetics,* 18, 619-620.

STAPLEY, J., REGER, J., FEULNER, P. G., SMADJA, C., GALINDO, J., EKBLOM, R., BENNISON, C., BALL, A. D., BECKERMAN, A. P. & SLATE, J. 2010. Adaptation genomics: the next generation. *Trends in ecology & evolution,* 25, 705-712.

STEBBINGS, R. E. 1988. *The conservation of European bats*, Christopher Helm Pub Limited.

SURGET-GROBA, Y. & MONTOYA-BURGOS, J. I. 2010. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome research,* 20, 1432-1440.

TATUSOV, R. L., FEDOROVA, N. D., JACKSON, J. D., JACOBS, A. R., KIRYUTIN, B., KOONIN, E. V., KRYLOV, D. M., MAZUMDER, R., MEKHEDOV, S. L. & NIKOLSKAYA, A. N. 2003. The COG database: an updated version includes eukaryotes. *BMC bioinformatics,* 4, 1.

TATUSOV, R. L., GALPERIN, M. Y., NATALE, D. A. & KOONIN, E. V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research,* 28, 33-36.

TEELING, E. C., MADSEN, O., VAN DEN BUSSCHE, R. A., DE JONG, W. W., STANHOPE, M. J. & SPRINGER, M. S. 2002. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proc Natl Acad Sci U S A,* 99, 1431-6.

TEELING, E. C., SPRINGER, M. S., MADSEN, O., BATES, P., O'BRIEN, S. J. & MURPHY, W. J. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science,* 307, 580-584.

TOUITOU, Y. 2005. [Melatonin: what for?]. *Bull Acad Natl Med,* 189, 879-89; discussion 889-91.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009a. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25, 1105-1111.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009b. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25, 1105-11.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology,* 28, 511-515.

TRUSSELL, L. 2000. Mutant ion channel in cochlear hair cells causes deafness. *Proceedings of the National Academy of Sciences,* 97, 3786-3788.

VAN CAMP, G. & SMITH, R. J. 2002. Hereditary hearing loss homepage.

VERNES, S. C., OLIVER, P. L., SPITERI, E., LOCKSTONE, H. E., PULIYADI, R., TAYLOR, J. M., HO, J., MOMBEREAU, C., BREWER, A. & LOWY, E. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet,* 7, e1002145.

VERNES, S. C., SPITERI, E., NICOD, J., GROSZER, M., TAYLOR, J. M., DAVIES, K. E., GESCHWIND, D. H. & FISHER, S. E. 2007. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *The American Journal of Human Genetics,* 81, 1232-1250.

VERPY, E., LEIBOVICI, M., MICHALSKI, N., GOODYEAR, R. J., HOUDON, C., WEIL, D., RICHARDSON, G. P. & PETIT, C. 2011. Stereocilin connects outer hair cell stereocilia to one another and to the tectorial membrane. *Journal of Comparative Neurology,* 519, 194-210.

VERPY, E., LEIBOVICI, M., ZWAENEPOEL, I., LIU, X.-Z., GAL, A., SALEM, N., MANSOUR, A., BLANCHARD, S., KOBAYASHI, I. & KEATS, B. J. 2000. A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. *Nature genetics,* 26, 51-55.

VETTER, D. E., KATZ, E., MAISON, S. F., TARANDA, J., TURCAN, S., BALLESTERO, J., LIBERMAN, M. C., ELGOYHEN, A. B. & BOULTER, J. 2007. The alpha10 nicotinic acetylcholine receptor subunit is required for normal synaptic function and integrity of the olivocochlear system. *Proc Natl Acad Sci U S A,* 104, 20594-9.

VETTER, D. E., LIBERMAN, M. C., MANN, J., BARHANIN, J., BOULTER, J., BROWN, M. C., SAFFIOTE-KOLMAN, J., HEINEMANN, S. F. & ELGOYHEN, A. B. 1999. Role of α9 nicotinic ACh receptor subunits in the development and function of cochlear efferent innervation. *Neuron,* 23, 93-103.

VILELLA, A. J., SEVERIN, J., URETA-VIDAL, A., HENG, L., DURBIN, R. & BIRNEY, E. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome research,* 19, 327-335.

VON HELVERSEN, O. & WINTER, Y. 2003. Glossophagine bats and their flowers: costs and benefits for plants and pollinators. *Bat ecology,* 346-397.

VREUGDE, S., ERVEN, A., KROS, C. J., MARCOTTI, W., FUCHS, H., KURIMA, K., WILCOX, E. R., FRIEDMAN, T. B., GRIFFITH, A. J. & BALLING, R. 2002. Beethoven, a mouse model for dominant, progressive hearing loss DFNA36. *Nature genetics,* 30, 257-258.

WAGATSUMA, M., KITOH, R., SUZUKI, H., FUKUOKA, H., TAKUMI, Y. & USAMI, S. 2007. Distribution and frequencies of CDH23 mutations in Japanese patients with non-syndromic hearing loss. *Clinical genetics,* 72, 339-344.

WALL, D., FRASER, H. & HIRSH, A. 2003. Detecting putative orthologs. *Bioinformatics,* 19, 1710-1711.

WANG, A., LIANG, Y., FRIDELL, R. A., PROBST, F. J., WILCOX, E. R., TOUCHMAN, J. W., MORTON, C. C., MORELL, R. J., NOBEN-TRAUTH, K. & CAMPER, S. A. 1998. Association of unconventional myosin MYO15 mutations with human nonsyndromic deafness DFNB3. *Science,* 280, 1447-1451.

WANG, K., SINGH, D., ZENG, Z., COLEMAN, S. J., HUANG, Y., SAVICH, G. L., HE, X., MIECZKOWSKI, P., GRIMM, S. A., PEROU, C. M., MACLEOD, J. N., CHIANG, D. Y., PRINS, J. F. & LIU, J. 2010. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res,* 38, e178.

WANG, X. & ZHANG, B. 2013. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics,* btt543.

WARD, J. A., PONNALA, L. & WEBER, C. A. 2012. Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany,* 99, 267-276.

WASHINGTON, S. D. & KANWAL, J. S. 2012. Sex-dependent hemispheric asymmetries for processing frequency-modulated sounds in the primary auditory cortex of the mustached bat. *J Neurophysiol,* 108, 1548-66.

WELTER, D., MACARTHUR, J., MORALES, J., BURDETT, T., HALL, P., JUNKINS, H., KLEMM, A., FLICEK, P., MANOLIO, T. & HINDORFF, L. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research,* 42, D1001-D1006.

WENSTRUP, J. J., NATARAJ, K. & SANCHEZ, J. T. 2012. Mechanisms of spectral and temporal integration in the mustached bat inferior colliculus. *Front Neural Circuits,* 6, 75.

WHELAN, S. & GOLDMAN, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution,* 18, 691-699.

WILCOX, E. R., BURTON, Q. L., NAZ, S., RIAZUDDIN, S., SMITH, T. N., PLOPLIS, B., BELYANTSEVA, I., BEN-YOSEF, T., LIBURD, N. A. & MORELL, R. J. 2001. Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell,* 104, 165-172.

WILSON, D. E. & REEDER, D. M. 2005. *Mammal species of the world: a taxonomic and geographic reference*, JHU Press.

WOLF, J. B. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular ecology resources,* 13, 559-572.

WOLF, J. B., LINDELL, J. & BACKSTRÖM, N. 2010. Speciation genetics: current status and evolving approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365, 1717-1733.

WOODS, J. O., SINGH-BLOM, U. M., LAURENT, J. M., MCGARY, K. L. & MARCOTTE, E. M. 2013. Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. *BMC bioinformatics,* 14, 1.

YANG, Z. & RANNALA, B. 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics,* 13, 303-314.

YAO, L., BROWN, J. P., STAMPANONI, M., MARONE, F., ISLER, K. & MARTIN, R. D. 2012. Evolutionary change in the brain size of bats. *Brain Behav Evol,* 80, 15-25.

YASUNAGA, S. I., GRATI, M. H., CHARDENOUX, S., SMITH, T. N., FRIEDMAN, T. B., LALWANI, A. K., WILCOX, E. R. & PETIT, C. 2000. OTOF encodes multiple long and short isoforms: genetic evidence that the long ones underlie recessive deafness DFNB9. *The American Journal of Human Genetics,* 67, 591-600.

YOUNG, M. D., WAKEFIELD, M. J., SMYTH, G. K. & OSHLACK, A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology,* 11, 1.

ZHANG, G., COWLED, C., SHI, Z., HUANG, Z., BISHOP-LILLY, K. A., FANG, X., WYNNE, J. W., XIONG, Z., BAKER, M. L. & ZHAO, W. 2012. Comparative

analysis of bat genomes provides insight into the evolution of flight and immunity. *Science,* 1230835.

ZHANG, Q.-J., LAN, L., LI, N., QI, Y., ZONG, L., SHI, W., YU, L., WANG, H., YANG, J. & XIE, L.-Y. 2015. Identification of a novel mutation of PJVK in the Chinese non-syndromic hearing loss population with low prevalence of the PJVK mutations. *Acta oto-laryngologica,* 135, 211-216.

ZHENG, J., MADISON, L. D., OLIVER, D., FAKLER, B. & DALLOS, P. 2002. Prestin, the motor protein of outer hair cells. *Audiology and Neurotology,* 7, 9-12.

ZHENG, J. L., SHOU, J., GUILLEMOT, F., KAGEYAMA, R. & GAO, W.-Q. 2000. Hes1 is a negative regulator of inner ear hair cell differentiation. *Development,* 127, 4551-4560.

ZHENG, Q. Y., YAN, D., OUYANG, X. M., DU, L. L., YU, H., CHANG, B., JOHNSON, K. R. & LIU, X. Z. 2005. Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. *Human molecular genetics,* 14, 103-111.

ZHOU, X., XU, S., ZHANG, P. & YANG, G. 2011. Developing a series of conservative anchor markers and their application to phylogenomics of Laurasiatherian mammals. *Molecular ecology resources,* 11, 134-140.

ZIGMOND, M. J. 1999. Otto Loewi and the demonstration of chemical neurotransmission. *Brain Res Bull,* 50, 347-8.

ZOU, J., LUO, L., SHEN, Z., CHIODO, V. A., AMBATI, B. K., HAUSWIRTH, W. W. & YANG, J. 2011. Whirlin replacement restores the formation of the USH2 protein complex in whirlin knockout photoreceptors. *Investigative ophthalmology & visual science,* 52, 2343-2351.

7. APPENDICES

Appendix – I : Inparanoid BLAST script

Appendix – II : Complete BLAST hit table along with e value score

Appendix – III : Bio-GPS (Figures 1-14)

Appendix – IV : Cumulative frequency and selection strength ω, script

Appendix – V : Gene expression script Heatmap (TPM and FPKM)

Appendix – VI : R - Bioconductor commands (Gene Ontology and KEGG)

Appendix – VII : Gene expression Abundances script BOXPLOT and Median (2x, 5x and 10x TPM)

Appendix - I

Inparanoid BLAST - Orthologs

```
# process a parsed inparanoid table. Generate lists and figures.

#####################################################
##### USER TO CHANGE
#####################################################
spp1.name <- "Myotis_lucifugus"
spp2.name <- "UoN.Bat"

home <- "/home/adac_guest/Desktop/16-5 Original Bat data MRes KS/UoN.Bat.vs.Bat"
# set working directory

inparanoid.table <- "/home/adac_guest/Desktop/16-5 Original Bat data MRes
KS/UoN.Bat.vs.Bat/inparanoid.UoN.Bat.vs.Myotis_lucifugus.parsed" # set parsed
inparanoid table

spp1.IDs <- "/home/adac_guest/Desktop/16-5 Original Bat data MRes
KS/UoN.Bat.vs.Bat/Myotis_lucifugus.Myoluc2.0.pep.IDs" # set list of all input IDs for
spp1
spp2.IDs <- "/home/adac_guest/Desktop/16-5 Original Bat data MRes
KS/UoN.Bat.vs.Bat/UoN.Ppi.T.IDs" # set list of all input IDs for spp2

#####################################################
# END
#####################################################


library(ggplot2)
library(dplyr)
library(plyr)
library('VennDiagram')
library('gridExtra')

setwd(home)

# read tables of data
data1 = read.table(inparanoid.table, sep='\t', header=TRUE)

ortho.counts <- as.numeric(table(data1$Type))
ortho.type.factors <- levels(data1$Type)

ortho.counts <- as.data.frame(cbind(ortho.counts,ortho.type.factors))
colnames(ortho.counts) <- c("Counts","Type")

ortho.counts.table <- paste(spp1.name,spp2.name,"ortholog.type.counts.table.txt",
sep=".")
write.table(ortho.counts,file=ortho.counts.table, row.names = FALSE, col.names =
TRUE, sep ="\t", quote=FALSE)
#####################################################
##### Plot ortholog type percentages
#####################################################
```

```
ortho.pc <- as.numeric(100*(table(data1$Type)/length(data1$Type)))

ortho.type.pc <- as.data.frame(cbind(ortho.pc,ortho.type.factors))
colnames(ortho.type.pc) <- c("Percent","Type")
ortho.type.pc$Percent <- as.numeric(paste(ortho.type.pc$Percent))

ortho.type.table <- paste(spp1.name,spp2.name,"ortholog.type.PC.table.txt", sep=".")
write.table(ortho.type.pc,file=ortho.type.table, row.names = FALSE, col.names = TRUE,
sep ="\t", quote=FALSE)

xlabel <- paste("Ortholog Type (",spp1.name,":",spp2.name,")", sep ="")
ortho.type.plot <- ggplot(ortho.type.pc, aes(x=as.factor(Type), y=Percent)) +
  theme_bw()+
  geom_bar(stat = "identity", col="slategrey", fill="slategrey")+
  xlab(xlabel)

ortholog.type.PC.barplot.pdf.name <-
paste(spp1.name,spp2.name,"ortholog.type.PC.barplot.pdf", sep=".")
pdf(ortholog.type.PC.barplot.pdf.name)
print(ortho.type.plot)
dev.off()

#########################################################
##### Plot ratios of orthologs
#########################################################

data1$ratio <- data1$Count_spp2/data1$Count_spp1
data1.short <- data1[which(data1$ratio <= 5), ]
xmin <- min(data1.short$ratio)
xlabel <- paste("Ortholog Type (",spp2.name,"/",spp1.name,")", sep ="")

ortho.ratio.histogram.plot <-
  ggplot(data1.short, aes(ratio)) +
  theme_bw()+
  geom_histogram(lwd=0.5,col="black",fill="slategrey", binwidth=0.25)+
  xlim(xmin,5)+
  xlab(xlabel)

ortholog.ratio.histogram.pdf.name <-
paste(spp1.name,spp2.name,"ortholog.ratio.histogram.pdf", sep=".")
pdf(ortholog.ratio.histogram.pdf.name)
print(ortho.ratio.histogram.plot)
dev.off()

#########################################################
##### Euler Plot of orthologs
#########################################################

spp1.IDs <- unique(scan(spp1.IDs, what=" "))
spp2.IDs <- unique(scan(spp2.IDs, what=" "))

orthologs.spp1 <- unique(unlist(strsplit(as.vector(data1$Homologs_Spp1), ',')))
orthologs.spp2 <- unique(unlist(strsplit(as.vector(data1$Homologs_Spp2), ',')))

orthologs.spp1 <- gsub("\\s", "", orthologs.spp1)
orthologs.spp2 <- gsub("\\s", "", orthologs.spp2)
```

```
spp1.unique.genes <- setdiff(spp1.IDs,orthologs.spp1)
spp2.unique.genes <- setdiff(spp2.IDs,orthologs.spp2)
spp1.unique.genes <- gsub("\\s", "", spp1.unique.genes)
spp2.unique.genes <- gsub("\\s", "", spp2.unique.genes)

# OUTPUT FILES
out.file.1 <- paste(spp1.name,".unique.genes.txt", sep="")
out.file.2 <- paste(spp2.name,".unique.genes.txt", sep="")
out.file.3 <- paste(spp1.name,"vs",spp2.name,spp1.name,"orthologs.txt", sep=".")
out.file.4 <- paste(spp1.name,"vs",spp2.name,spp2.name,"orthologs.txt", sep=".")


write.table(spp1.unique.genes, file=out.file.1 , sep = "," , row.names = FALSE,
col.names = FALSE, quote=FALSE)
write.table(spp2.unique.genes, file=out.file.2 , sep = "," , row.names = FALSE,
col.names = FALSE, quote=FALSE)

write.table(orthologs.spp1, file=out.file.3 , sep = "," , row.names = FALSE, col.names =
FALSE, quote=FALSE)
write.table(orthologs.spp2, file=out.file.4 , sep = "," , row.names = FALSE, col.names =
FALSE, quote=FALSE)

# Euler diagram of orthologs vs unique.

area1 = as.numeric(length(spp1.IDs))
area2 = as.numeric(length(spp2.IDs))

n12 = as.numeric(nrow(data1))

category.names <- c(spp1.name,spp2.name)

venn.plot.1 <- draw.pairwise.venn(area1,area2,n12,
                          category = category.names,
                          euler.d = TRUE,
                          scaled = TRUE,
                          fill = c("dodgerblue", "orchid3"),
                          cat.cex = 1.0,
                          cat.dist = 0.1, # distance of labels from circles.
                          lwd = c(0.8,0.8),
                          margin = 0.2,
                          ind = FALSE
)
euler.pdf.name <- paste(spp1.name,spp2.name,"2.way.Euler.pdf", sep=".")
pdf(euler.pdf.name)
grid.draw(venn.plot.1)
dev.off()
```

Appendix- II

BLAST hit Table

UoN.Ppi.Transcript_I	Query	Query length	Eval	bit score	Query start	Query end	Hit name	Hit description	Hit length	Hit start	Hit end	hsp length	Proportiona	Classification	Reference	GOI
UoN.Ppi.T.42935	Unique.ENSbat_1	670	0	1350	1	669	ref|XP_008157148.1	PREDICTED: forkhead box protein P4 isoform X2 [Eptesicus fuscus] – Big brown bat	669	1	669	669	1	Trancription Factor	Uniprot	
UoN.Ppi.T.42107	Unique.ENSbat_10	333	0	527	65	332	ref|XP_008152626.1	PREDICTED: sulfotransferase family cytosolic 2B member 1 [Eptesicus fuscus] - Big brown bat	338	71	338	268	0.7928994	Enzymatic	Uniprot	
UoN.Ppi.T.43474	Unique.ENSbat_100	760	0	1021	95	759	ref|XP_008150603.1	PREDICTED: NEDD4-binding protein 1 [Eptesicus fuscus]	845	7	691	688	0.9956585			
UoN.Ppi.T.43474	Unique.ENSbat_101	756	0	1021	91	755	ref|XP_008150603.1	PREDICTED: NEDD4-binding protein 1 [Eptesicus fuscus]	845	7	691	688	0.9956585			
UoN.Ppi.T.42474	Unique.ENSbat_102	209	2E-150	428	1	208	ref|XP_006100525.1	PREDICTED: ras-related protein Rab-6A isoform X1 [Myotis lucifugus]ref|XP_008148174.1| PREDICTED: ras-related protein Rab-6A isoform X1 [Eptesicus fuscus]ref|	208	1	208	208	1			
UoN.Ppi.T.42584	Unique.ENSbat_103	304	0	582	1	303	gb|EPQ08730.1|	Putative protein FLJ45252 [Myotis brandtii]	395	93	395	303	0.7670886			
UoN.Ppi.T.42604	Unique.ENSbat_104	441	0	865	14	441	ref|XP_008137806.1	PREDICTED: SH3KBP1-binding protein 1 [Eptesicus fuscus]	704	1	428	428	1			
UoN.Ppi.T.42836	Unique.ENSbat_105	143	2E-84	257	1	142	ref|XP_003227918.2	PREDICTED: 60S ribosomal protein L31 [Anolis carolinensis]	180	34	180	147	0.8166667			Expression
UoN.Ppi.T.42837	Unique.ENSbat_106	140	1E-83	253	15	139	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1			
UoN.Ppi.T.42875	Unique.ENSbat_107	127	4E-84	254	1	126	ref|XP_012596776.1	PREDICTED: histone H2B type 1-like [Microcebus murinus] - The grey mouse-lemur	142	1	126	126	1	Chaperone/ Transcription regulator	Uniprot	
UoN.Ppi.T.43045	Unique.ENSbat_108	994	0	1923	1	923	ref|XP_008138557.1	PREDICTED: DNA-directed RNA polymerase II subunit RPB2 isoform X1 [Eptesicus fuscus]ref|XP_014315655.1| PREDICTED: DNA-directed RNA polymerase II subu	1174	1	923	923	1	Binding protein	Uniprot	Expression
UoN.Ppi.T.43045	Unique.ENSbat_109	994	0	1923	1	923	ref|XP_008138557.1	PREDICTED: DNA-directed RNA polymerase II subunit RPB2 isoform X1 [Eptesicus fuscus]ref|XP_014315655.1| PREDICTED: DNA-directed RNA polymerase II subu	1174	1	923	923	1	Binding protein	Uniprot	
UoN.Ppi.T.43595	Unique.ENSbat_11	129	4E-76	233	11	128	ref|XP_006109357.1	PREDICTED: beta-arrestin-2-like [Myotis lucifugus]	119	2	119	118	0.9915966	Brain development/GPCR	Uniprot	
UoN.Ppi.T.43046	Unique.ENSbat_110	994	0	1923	1	923	ref|XP_008138557.1	PREDICTED: DNA-directed RNA polymerase II subunit RPB2 isoform X1 [Eptesicus fuscus]ref|XP_014315655.1| PREDICTED: DNA-directed RNA polymerase II subu	1174	1	923	923	1	Binding protein	Uniprot	
UoN.Ppi.T.43487	Unique.ENSbat_111	795	0	1406	18	795	ref|XP_008146452.1	PREDICTED: ephrin type-A receptor 8 [Eptesicus fuscus]	1077	84	857	778	0.907818	Nucleotide Binding	Uniprot	
UoN.Ppi.T.42167	Unique.ENSbat_112	163	3E-107	316	1	162	ref|XP_003271942.1	PREDICTED: 40S ribosomal protein S18 [Nomascus leucogenys] - White cheeked gibbon	175	10	175	166	0.9485714			
UoN.Ppi.T.42683	Unique.ENSbat_113	352	0	719	1	351	ref|XP_006097737.1	PREDICTED: THO complex subunit 3 [Myotis lucifugus]	351	1	351	351	1			
UoN.Ppi.T.42693	Unique.ENSbat_114	246	4E-146	424	1	246	gb|ELK04466.1|	Heparan-sulfate 6-O-sulfotransferase 2 [Pteropus alecto] - The black flying fox or black fruit bat	316	11	256	246	0.9609375			
UoN.Ppi.T.42950	Unique.ENSbat_115	389	0	558	1	389	ref|XP_015423726.1	PREDICTED: zinc finger protein 418 [Myotis davidii]gb|ELK25153.1| Zinc finger protein 418 [Myotis davidii]	703	47	442	397	0.89819			
UoN.Ppi.T.42985	Unique.ENSbat_116	201	2E-128	374	1	200	ref|XP_005858576.2	PREDICTED: HSPB1-associated protein 1 [Myotis brandtii]	239	39	239	201	0.8410042			
UoN.Ppi.T.42985	Unique.ENSbat_117	201	2E-128	374	1	200	ref|XP_005858576.2	PREDICTED: HSPB1-associated protein 1 [Myotis brandtii]	239	39	239	201	0.8410042			
UoN.Ppi.T.42992	Unique.ENSbat_118	408	0	826	1	407	ref|XP_014321158.1	PREDICTED: UPF0061 protein ETA_18310-like isoform X5 [Myotis lucifugus]	526	120	526	407	0.7737643			
UoN.Ppi.T.42993	Unique.ENSbat_119	389	0	785	1	388	ref|XP_014321156.1	PREDICTED: UPF0061 protein ETA_18310-like isoform X3 [Myotis lucifugus]	593	206	593	388	0.6543002			
UoN.Ppi.T.42122	Unique.ENSbat_12	463	0	949	1	462	ref|XP_005861918.1	PREDICTED: retinoic acid receptor alpha isoform X1 [Myotis brandtii]ref|XP_006774401.1| PREDICTED: retinoic acid receptor alpha isoform X2 [Myotis davidii]ref|XP_(	462	1	462	462	1	Receptor Family	GenBank	
UoN.Ppi.T.43077	Unique.ENSbat_120	168	4E-108	335	1	168	ref|XP_005078126.1	PREDICTED: leucine-rich repeat-containing protein 4 [Mesocricetus auratus] - The golden hamster or Syrian hamster	651	1	168	168	1			
UoN.Ppi.T.42063	Unique.ENSbat_121	151	7E-85	259	9	150	ref|XP_003227918.2	PREDICTED: 60S ribosomal protein L31 [Anolis carolinensis] - Carolina anole (arboreal lizard)	180	34	180	147	0.8166667			
UoN.Ppi.T.42067	Unique.ENSbat_122	149	1E-84	258	7	148	ref|XP_003227918.2	PREDICTED: 60S ribosomal protein L31 [Anolis carolinensis] - Carolina anole (arboreal lizard)	180	34	180	147	0.8166667			Expression
UoN.Ppi.T.42068	Unique.ENSbat_123	139	7E-84	254	14	138	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1			
UoN.Ppi.T.43176	Unique.ENSbat_124	1400	0	2486	1	1399	ref|XP_008143350.1	PREDICTED: striated muscle preferentially expressed protein kinase isoform X1 [Eptesicus fuscus]	3275	1884	3275	1399	0.4271756	ATP binding/Enzymatic/ Muscle cell (	Uniprot	
UoN.Ppi.T.43268	Unique.ENSbat_125	247	4E-111	345	5	185	ref|XP_008155227.1	PREDICTED: zinc finger protein with KRAB and SCAN domains 7 [Eptesicus fuscus]	624	14	194	181	0.9329897			
UoN.Ppi.T.43343	Unique.ENSbat_126	376	0	704	1	340	ref|XP_005860210.1	PREDICTED: neurexin 3 isoform X20 [Myotis brandtii]ref|XP_006087029.1| PREDICTED: neurexin 3 isoform X20 [Myotis lucifugus]	459	21	360	340	0.9444444			
UoN.Ppi.T.43344	Unique.ENSbat_127	376	0	704	1	340	ref|XP_005860210.1	PREDICTED: neurexin 3 isoform X20 [Myotis brandtii]ref|XP_006087029.1| PREDICTED: neurexin 3 isoform X20 [Myotis lucifugus]	459	21	360	340	0.9444444			
UoN.Ppi.T.43348	Unique.ENSbat_128	171	6E-38	144	15	107	gb|ELK28796.1|	Carcinoembryonic antigen-related cell adhesion molecule 1 [Myotis davidii]	352	258	350	93	0.2657143			
UoN.Ppi.T.43468	Unique.ENSbat_129	396	0	692	1	382	ref|XP_008143128.1	PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1I [Eptesicus fuscus]	2058	1	382	382	1			
UoN.Ppi.T.42468	Unique.ENSbat_13	277	4E-140	418	38	239	ref|XP_008263382.1	PREDICTED: nuclear factor 1 A-type isoform X4 [Oryctolagus cuniculus] - European rabbit	554	30	231	202	0.8744589	Signal transduction/ protein transport	Uniprot	
UoN.Ppi.T.42319	Unique.ENSbat_130	1676	0	3397	1	1675	ref|XP_007957626.1	PREDICTED: U5 small nuclear ribonucleoprotein 200 kDa helicase [Orycteropus afer afer] - Orycteropus means burrowing foot, and the name afer refers to Africa , Aa	2136	1	1675	1675	1			
UoN.Ppi.T.42805	Unique.ENSbat_131	247	1E-180	508	1	246	ref|XP_008139822.1	PREDICTED: phosphomannomutase 2 [Eptesicus fuscus]	246	1	246	246	1			
UoN.Ppi.T.42829	Unique.ENSbat_132	242	5E-145	416	37	241	ref|XP_008153109.1	PREDICTED: mitotic spindle assembly checkpoint protein MAD2A [Eptesicus fuscus]	205	1	205	205	1			
UoN.Ppi.T.42884	Unique.ENSbat_133	587	0	1044	32	586	ref|XP_014310010.1	PREDICTED: vitamin D3 hydroxylase-associated protein-like [Myotis lucifugus]	604	32	604	573	0.9486755			
UoN.Ppi.T.42934	Unique.ENSbat_134	378	0	774	1	377	ref|XP_008139227.1	PREDICTED: histone deacetylase 8 isoform X1 [Eptesicus fuscus]	377	1	377	377	1	Hydrolase/Transcription regulation	Uniprot	
UoN.Ppi.T.42935	Unique.ENSbat_135	378	0	774	1	377	ref|XP_008139227.1	PREDICTED: histone deacetylase 8 isoform X1 [Eptesicus fuscus]	377	1	377	377	1	Hydrolase/Transcription regulation	Uniprot	
UoN.Ppi.T.43047	Unique.ENSbat_136	558	0	1128	1	557	ref|XP_008148933.1	PREDICTED: polypyrimidine tract-binding protein 1 isoform X1 [Eptesicus fuscus]	557	1	557	557	1			
UoN.Ppi.T.43047	Unique.ENSbat_137	583	0	1128	26	582	ref|XP_008148933.1	PREDICTED: polypyrimidine tract-binding protein 1 isoform X1 [Eptesicus fuscus]	557	1	557	557	1			
UoN.Ppi.T.43061	Unique.ENSbat_138	235	5E-172	486	1	234	ref|XP_005871129.1	PREDICTED: proteasome subunit alpha type-2 [Myotis brandtii]ref|XP_006085417.1| PREDICTED: proteasome subunit alpha type-2 [Myotis lucifugus]ref|XP_0081589	234	1	234	234	1			
UoN.Ppi.T.43408	Unique.ENSbat_139	243	2E-164	466	18	242	ref|XP_005860035.1	PREDICTED: pre-mRNA-splicing factor SPF27 [Myotis brandtii]ref|XP_006092551.1| PREDICTED: pre-mRNA-splicing factor SPF27 [Myotis lucifugus]ref|XP_006763(	225	1	225	225	1			
UoN.Ppi.T.42006	Unique.ENSbat_14	107	9E-70	217	1	106	ref|XP_010357954.1	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L36a [Rhinopithecus roxellana] - Golden snub nosed Monkey	143	38	143	106	0.7412587	Translational protein	Uniprot	
UoN.Ppi.T.43416	Unique.ENSbat_140	271	7E-142	424	1	270	ref|XP_008150990.1	PREDICTED: sodium/myo-inositol cotransporter 2 isoform X2 [Eptesicus fuscus]	609	338	609	274	0.4499179			
UoN.Ppi.T.43428	Unique.ENSbat_141	279	2E-164	469	31	278	ref|XP_001036816.1	tropomyosin alpha-3 chain isoform Tpm3.2cy [Homo sapiens]ref|NP_001239140.1| tropomyosin alpha-3 chain [Canis lupus familiaris]ref|NP_001244772.1| tropomyosin	248	1	248	248	1			
UoN.Ppi.T.43429	Unique.ENSbat_142	241	5E-150	431	1	240	ref|NP_001001491.1	tropomyosin alpha-4 chain [Mus musculus]sp|Q6IRU2.3|TPM4_MOUSE RecName: Full=Tropomyosin alpha-4 chain; AltName: Full=Tropomyosin-4gb|AAH70421.1| Tro	248	9	248	240	0.9677419			
UoN.Ppi.T.41959	Unique.ENSbat_143	220	7E-97	309	15	197	gb|EQB78123.1|	eukaryotic translation initiation factor 2D [Camelus ferus] - Bactrian camel	724	62	259	198	0.7644788			
UoN.Ppi.T.41960	Unique.ENSbat_144	324	4E-171	500	57	307	ref|XP_008152401.1	PREDICTED: eukaryotic translation initiation factor 2D isoform X1 [Eptesicus fuscus]	575	1	251	251	1			
UoN.Ppi.T.41972	Unique.ENSbat_145	313	4E-153	442	48	305	ref|XP_014387018.1	PREDICTED: protein FAM205CP, partial [Myotis brandtii]	276	19	276	258	0.9347826			
UoN.Ppi.T.41973	Unique.ENSbat_146	956	0	1683	1	947	ref|XP_008143350.1	PREDICTED: striated muscle preferentially expressed protein kinase isoform X1 [Eptesicus fuscus]	3275	1976	2920	947	0.3243151			
UoN.Ppi.T.41974	Unique.ENSbat_147	956	0	1683	1	947	ref|XP_008143350.1	PREDICTED: striated muscle preferentially expressed protein kinase isoform X1 [Eptesicus fuscus]	3275	1976	2920	947	0.3243151			
UoN.Ppi.T.42629	Unique.ENSbat_148	557	0	936	1	556	ref|XP_014319741.1	PREDICTED: leucine-rich repeat flightless-interacting protein 1 isoform X29 [Myotis lucifugus]	628	1	628	628	1	Transcription Regulation	Uniprot	
UoN.Ppi.T.43238	Unique.ENSbat_149	198	2E-136	394	1	198	ref|XP_012307342.1	PREDICTED: pre-mRNA-splicing factor CWC25 homolog [Aotus nancymaae] - Night monkey	236	1	198	198	1			
UoN.Ppi.T.42007	Unique.ENSbat_15	107	9E-70	217	1	106	ref|XP_010357954.1	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L36a [Rhinopithecus roxellana] - Golden snub nosed Monkey	143	38	143	106	0.7412587	Translational protein	Uniprot	
UoN.Ppi.T.42077	Unique.ENSbat_150	106	1E-71	220	1	105	ref|XP_005883251.1	PREDICTED: dynein light chain 4, axonemal isoform X2 [Myotis brandtii]ref|XP_006107199.1| PREDICTED: dynein light chain 4, axonemal isoform X2 [Myotis lucifugus	105	1	105	105	1			
UoN.Ppi.T.43519	Unique.ENSbat_151	472	0	876	34	471	ref|XP_006757932.1	PREDICTED: clusterin-associated protein 1 isoform X5 [Myotis davidii]	438	1	438	438	1	cell cycle / colon cancer	pfam/interpro	
UoN.Ppi.T.43519	Unique.ENSbat_152	446	0	820	34	445	ref|XP_006757934.1	PREDICTED: clusterin-associated protein 1 isoform X7 [Myotis davidii]	412	1	412	412	1	cell cycle / colon cancer	pfam/interpro	
UoN.Ppi.T.42118	Unique.ENSbat_153	160	3E-100	312	1	160	ref|XP_008158246.1	PREDICTED: leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 2 [Eptesicus fuscus]	606	1	160	160	1			
UoN.Ppi.T.42119	Unique.ENSbat_154	160	3E-100	312	1	160	ref|XP_008158246.1	PREDICTED: leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 2 [Eptesicus fuscus]	606	1	160	160	1			
UoN.Ppi.T.42302	Unique.ENSbat_155	148	9E-104	306	1	147	gb|ELK37008.1|	Protein Wnt-5a [Myotis davidii]	163	17	163	147	0.9018405			Expression
UoN.Ppi.T.42367	Unique.ENSbat_156	515	0	1047	1	515	ref|XP_008138917.1	PREDICTED: LOW QUALITY PROTEIN: sodium channel protein type 8 subunit alpha [Eptesicus fuscus]	1995	1203	1717	515	0.2999418	Transmebrane protein/voltage gated s	Uniprot	
UoN.Ppi.T.42485	Unique.ENSbat_157	230	2E-134	430	1	230	ref|XP_006109819.1	PREDICTED: LOW QUALITY PROTEIN: IgGFc-binding protein [Myotis lucifugus]	3002	2232	2461	230	0.0934579	Methyl transferase/Nucleotide binding	Uniprot	Expression
UoN.Ppi.T.42809	Unique.ENSbat_158	399	0	815	1	399	ref|XP_014443866.1	PREDICTED: LOW QUALITY PROTEIN: ryanodine receptor 3 [Tupaia chinensis] - treeshrew	2407	655	1053	399	0.3789174	Transmembrane protein/intracellular l	Uniprot	
UoN.Ppi.T.42823	Unique.ENSbat_159	137	2E-92	277	1	136	ref|XP_007971598.1	PREDICTED: histone H3.1-like [Chlorocebus sabaeus] - sabaeus or green monkey or the callithrix monkey	181	46	181	136	0.7513812	DNA Binding/chromosomal	Uniprot	
UoN.Ppi.T.42947	Unique.ENSbat_16	338	0	707	1	337	ref|NP_001083.4|	casein kinase I isoform alpha isoform 2 [Homo sapiens]ref|NP_001248164.1| casein kinase I isoform alpha [Macaca mulatta]ref|XP_536470.2| PREDICTED: casein kina	337	1	337	337	1	Enzymatic	Uniprot	
UoN.Ppi.T.42988	Unique.ENSbat_160	246	0	516	1	245	ref|XP_006082367.1	PREDICTED: eukaryotic translation initiation factor 4E type 2 isoform X1 [Myotis lucifugus]ref|XP_006082708.1| PREDICTED: eukaryotic translation initiation factor 4E	245	1	245	245	1	Translation Inititation factor/ Binding	Uniprot	
UoN.Ppi.T.43083	Unique.ENSbat_161	164	1E-115	337	1	164	ref|XP_014115929.1	PREDICTED: cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform X2 [Pseudopodoces humilis] - The ground tit lark-like bird	186	1	164	164	1			
UoN.Ppi.T.43270	Unique.ENSbat_162	861	0	1392	1	860	ref|XP_008155855.1	PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1H isoform X1 [Eptesicus fuscus]	2344	1495	2344	863	0.3681741			
UoN.Ppi.T.43391	Unique.ENSbat_163	208	2E-129	415	1	208	ref|XP_006757669.1	PREDICTED: von Willebrand factor [Myotis davidii]gb|ELK34907.1| von Willebrand factor [Myotis davidii]	2815	584	791	208	0.2629583			
UoN.Ppi.T.43577	Unique.ENSbat_164	324	0	667	1	323	ref|NP_071943.1|	serine/threonine-protein phosphatase PP1-gamma catalytic subunit isoform 1 [Rattus norvegicus]ref|NP_777006.1| serine/threonine-protein phosphatase PP1-gamma c	323	1	323	323	1			
UoN.Ppi.T.42187	Unique.ENSbat_165	573	0	759	107	571	ref|XP_005859300.1	PREDICTED: ribonuclease inhibitor isoform X1 [Myotis brandtii]	465	1	464	465	1.0021552			
UoN.Ppi.T.42726	Unique.ENSbat_166	363	0	588	54	340	ref|XP_008156878.1	PREDICTED: class I histocompatibility antigen, Gogo-C*0203 alpha chain-like [Eptesicus fuscus]	339	1	287	287	1	Immune response/Signal peptide	Uniprot	Expression
UoN.Ppi.T.42835	Unique.ENSbat_167	325	0	624	1	310	ref|XP_008151720.1	PREDICTED: dynein heavy chain 2, axonemal [Eptesicus fuscus]	4424	1142	1451	310	0.2136458	Nucleotide binding	Uniprot	
UoN.Ppi.T.42926	Unique.ENSbat_168	257	5E-156	461	5	257	gb|EPQ16013.1|	Myosin-14 [Myotis brandtii]	634	31	281	253	0.9003559	ATP binding/Motor activity	Uniprot	
UoN.Ppi.T.43493	Unique.ENSbat_169	201	2E-128	374	1	200	ref|XP_005858576.2	PREDICTED: HSPB1-associated protein 1 [Myotis brandtii]	239	39	239	201	0.8410042			
UoN.Ppi.T.42986	Unique.ENSbat_17	155	3E-76	236	1	122	ref|XP_014392234.1	PREDICTED: pyruvate dehydrogenase phosphatase regulatory subunit, mitochondrial [Myotis brandtii]	163	1	122	122	1	Enzymatic/ Calcium binding	Uniprot	
UoN.Ppi.T.43494	Unique.ENSbat_170	201	2E-128	374	1	200	ref|XP_005858576.2	PREDICTED: HSPB1-associated protein 1 [Myotis brandtii]	239	39	239	201	0.8410042			
UoN.Ppi.T.42158	Unique.ENSbat_171	209	1E-124	362	36	208	ref|XP_006758372.1	PREDICTED: 28S ribosomal protein S25, mitochondrial [Myotis davidii]ref|XP_008150588.1| PREDICTED: 28S ribosomal protein S25, mitochondrial [Eptesicus fuscus]	173	1	173	173	1			
UoN.Ppi.T.42158	Unique.ENSbat_172	208	9E-125	362	35	207	ref|XP_006758372.1	PREDICTED: 28S ribosomal protein S25, mitochondrial [Myotis davidii]ref|XP_008150588.1| PREDICTED: 28S ribosomal protein S25, mitochondrial [Eptesicus fuscus]	173	1	173	173	1			
UoN.Ppi.T.42698	Unique.ENSbat_173	374	0	578	1	373	ref|XP_016019607.1	PREDICTED: sorbin and SH3 domain-containing protein 1 isoform X8 [Rousettus aegyptiacus] - The Egyptian fruit bat or Egyptian rousette	1180	808	1180	376	0.3186441			
UoN.Ppi.T.42699	Unique.ENSbat_174	374	0	578	1	373	ref|XP_016019607.1	PREDICTED: sorbin and SH3 domain-containing protein 1 isoform X8 [Rousettus aegyptiacus] - The Egyptian fruit bat or Egyptian rousette	1180	808	1180	376	0.3186441			
UoN.Ppi.T.42699	Unique.ENSbat_175	374	0	578	1	373	ref|XP_016019607.1	PREDICTED: sorbin and SH3 domain-containing protein 1 isoform X8 [Rousettus aegyptiacus] - The Egyptian fruit bat or Egyptian rousette	1180	808	1180	376	0.3186441			Expression
UoN.Ppi.T.43162	Unique.ENSbat_176	292	4E-131	386	47	292	gb|ERE73398.1|	neurexin-1-alpha [Cricetulus griseus] - The Chinese hamster	279	1	230	246	1.0695652			
UoN.Ppi.T.42380	Unique.ENSbat_177	359	0	730	1	359	ref|XP_012320921.1	PREDICTED: neurexin-1-beta isoform X17 [Aotus nancymaae] – Night monkey	438	1	359	359	1	Neuroglin family binding protein	Uniprot	

UoN.Ppi.T.42880	Unique.ENSbat_178	189	9E-19	90.9	69	188	ref|XP_009243147.1	PREDICTED: 60S ribosomal protein L12 [Pongo abelii] - Sumatran Orangutan	228	47	175	134	0.7657143					
UoN.Ppi.T.42133	Unique.ENSbat_179	472	0	822	1	471	ref|XP_008140436.1	PREDICTED: 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-1 isoform X1 [Eptesicus fuscus]	1655	1184	1655	475	0.2870091	Calcium dependent/Signaling	Uniprot			
UoN.Ppi.T.43256	Unique.ENSbat_18	464	0	952	1	461	ref|XP_006094176.1	PREDICTED: histone-arginine methyltransferase CARM1-like [Myotis lucifugus]	461	1	461	461	1	Enzymatic/Methyl transferase	Uniprot			
UoN.Ppi.T.42425	Unique.ENSbat_180	287	5E-166	486	1	258	ref|XP_003215253.1	PREDICTED: general transcription factor II-I repeat domain-containing protein 2-like [Anolis carolinensis]	601	339	596	258	0.4328859					
UoN.Ppi.T.42607	Unique.ENSbat_181	212	2E-146	422	1	210	ref|XP_005865001.1	PREDICTED: E3 ubiquitin-protein ligase TRIM58 [Myotis brandtii]	314	104	313	210	0.6709265					
UoN.Ppi.T.43561	Unique.ENSbat_182	212	2E-146	422	1	210	ref|XP_005865001.1	PREDICTED: E3 ubiquitin-protein ligase TRIM58 [Myotis brandtii]	314	104	313	210	0.6709265					
UoN.Ppi.T.43567	Unique.ENSbat_183	861	0	1297	1	780	ref|XP_008148428.1	PREDICTED: pleckstrin homology-like domain family B member 1 [Eptesicus fuscus]	1337	1	797	797	1					
UoN.Ppi.T.43568	Unique.ENSbat_184	861	0	1297	1	780	ref|XP_008148428.1	PREDICTED: pleckstrin homology-like domain family B member 1 [Eptesicus fuscus]	1337	1	797	797	1					
UoN.Ppi.T.43576	Unique.ENSbat_185	103	4E-66	210	1	102	ref|XP_005861439.1	PREDICTED: ras-related protein Rab-3D [Myotis brandtii]ref|XP_014391212.1| PREDICTED: ras-related protein Rab-3D [Myotis brandtii]gb|EPQ04692.1| Ras-related p	219	118	219	102	0.4657534					
UoN.Ppi.T.43587	Unique.ENSbat_186	786	0	1172	3	786	ref|XP_014318667.1	PREDICTED: microtubule-associated tumor suppressor 1 [Myotis lucifugus]	932	10	807	799	0.9900867					
UoN.Ppi.T.42648	Unique.ENSbat_187	560	0	1143	1	559	ref|XP_004410270.1	PREDICTED: glutamate dehydrogenase 1, mitochondrial [Odobenus rosmarus divergens] - The walrus (large flippered marine mammal)	559	1	559	559	1	Nucleotide Binding	Uniprot			
UoN.Ppi.T.42649	Unique.ENSbat_188	560	0	1143	1	559	ref|XP_004410270.1	PREDICTED: glutamate dehydrogenase 1, mitochondrial [Odobenus rosmarus divergens] - The walrus (large flippered marine mammal)	559	1	559	559	1	Nucleotide Binding	Uniprot			
UoN.Ppi.T.42707	Unique.ENSbat_189	257	5E-158	494	1	257	ref|XP_008136987.1	PREDICTED: sodium channel protein type 9 subunit alpha [Eptesicus fuscus]	1865	1227	1483	257	0.1732974					
UoN.Ppi.T.43311	Unique.ENSbat_19	404	2E-119	378	32	367	ref|XP_015093730.1	PREDICTED: cyclin-dependent kinase 11B isoform X1 [Vicugna pacos] - Alpaca (camel like animal)	808	1	302	336	1.1125828	ATP binding, protein kinase	Uniprot			
UoN.Ppi.T.43079	Unique.ENSbat_190	310	0	587	1	287	ref|XP_008156878.1	PREDICTED: class I histocompatibility antigen, Gogo-C*0203 alpha chain-like [Eptesicus fuscus]	339	1	287	287	1	Immune response/Signal peptide	Uniprot			
UoN.Ppi.T.43087	Unique.ENSbat_191	514	0	1032	1	513	ref|XP_008140415.1	PREDICTED: LOW QUALITY PROTEIN: cytochrome P450 1A4-like [Eptesicus fuscus]	522	1	513	513	1	Transmembrane protein	Uniprot			
UoN.Ppi.T.43088	Unique.ENSbat_192	514	0	1032	1	513	ref|XP_008140415.1	PREDICTED: LOW QUALITY PROTEIN: cytochrome P450 1A4-like [Eptesicus fuscus]	522	1	513	513	1	Transmembrane protein	Uniprot			
UoN.Ppi.T.43098	Unique.ENSbat_193	1548	0	2596	16	1482	ref|XP_008155856.1	PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1H isoform X2 [Eptesicus fuscus]	2338	538	2014	1479	0.7343595					Expression
UoN.Ppi.T.43099	Unique.ENSbat_194	1548	0	2596	16	1482	ref|XP_008155856.1	PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1H isoform X2 [Eptesicus fuscus]	2338	538	2014	1479	0.7343595					
UoN.Ppi.T.43123	Unique.ENSbat_195	326	2E-162	466	81	325	ref|XP_006920362.1	PREDICTED: tropomyosin alpha-1 chain isoform X13 [Pteropus alecto]ref|XP_006079995.1| PREDICTED: tropomyosin alpha-4 chain-like isoform X17 [Bubalus bubalis	245	1	245	245	1	Actin binding/Calcium depepndent re	Uniprot			
UoN.Ppi.T.43497	Unique.ENSbat_196	90	5E-45	162	1	78	ref|XP_008138331.1	PREDICTED: zinc finger protein 382 [Eptesicus fuscus]	551	1	79	79	1					
UoN.Ppi.T.42173	Unique.ENSbat_197	225	1E-156	446	1	224	ref|XP_005866124.1	PREDICTED: claudin-19 [Myotis brandtii]ref|XP_006762650.1| PREDICTED: claudin-19 [Myotis davidii]	224	1	224	224	1					
UoN.Ppi.T.43007	Unique.ENSbat_198	250	0	515	1	249	ref|XP_006757239.1	PREDICTED: general transcription factor IIF subunit 2 [Myotis davidii]ref|XP_008143780.1| PREDICTED: general transcription factor IIF subunit 2 [Eptesicus fuscus]ref	249	1	249	249	1	Trancription Factor	Uniprot			
UoN.Ppi.T.43020	Unique.ENSbat_199	161	9E-61	197	1	160	ref|XP_008158585.1	PREDICTED: keratin-associated protein 13-1-like [Eptesicus fuscus]	156	1	156	166	1.0641026					
UoN.Ppi.T.42936	Unique.ENSbat_2	670	0	1350	1	669	ref|XP_008157148.1	PREDICTED: forkhead box protein P4 isoform X2 [Eptesicus fuscus] - Big brown bat	669	1	669	669	1	Trancription Factor	Uniprot			
UoN.Ppi.T.43491	Unique.ENSbat_20	293	2E-153	441	63	287	emb|CAC37949.1|	RICH1B protein [Homo sapiens]	226	1	225	225	1					
UoN.Ppi.T.43020	Unique.ENSbat_200	161	9E-61	197	1	160	ref|XP_008158585.1	PREDICTED: keratin-associated protein 13-1-like [Eptesicus fuscus]	156	1	156	166	1.0641026					Expression
UoN.Ppi.T.43593	Unique.ENSbat_201	916	0	1759	1	915	ref|XP_008139430.1	PREDICTED: serine/threonine-protein phosphatase 4 regulatory subunit 1-like [Eptesicus fuscus]	916	1	916	916	1					
UoN.Ppi.T.43594	Unique.ENSbat_202	913	0	1716	1	912	gb|EPQ16754.1|	Serine/threonine-protein phosphatase 4 regulatory subunit 1 [Myotis brandtii]	975	41	975	935	0.9589744	Alternative splicing results in multiple	Uniprot			
UoN.Ppi.T.43605	Unique.ENSbat_203	495	0	791	1	495	ref|XP_014392634.1	PREDICTED: FLYWCH-type zinc finger-containing protein 1 isoform X1 [Myotis brandtii]ref|XP_014392636.1| PREDICTED: FLYWCH-type zinc finger-containing prote	699	1	502	507	1.0099602	DNA/Metal binding	Uniprot			
UoN.Ppi.T.42331	Unique.ENSbat_204	342	0	654	1	341	ref|XP_006091821.1	PREDICTED: dysbindin isoform X1 [Myotis lucifugus]	339	1	339	341	1.0058997	Secondary strucutre/coiled coil	Uniprot			
UoN.Ppi.T.42039	Unique.ENSbat_205	166	8E-112	347	1	166	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	166	166	1					
UoN.Ppi.T.43292	Unique.ENSbat_206	250	4E-175	498	1	249	gb|EPQ11808.1|	Heparan-sulfate 6-O-sulfotransferase 2 [Myotis brandtii]	347	99	347	249	0.7175793					
UoN.Ppi.T.43427	Unique.ENSbat_207	749	0	1454	34	748	ref|XP_008140570.1	PREDICTED: methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial [Eptesicus fuscus]	715	1	715	715	1	Enzymatic/Biotin carboxylase activity	Uniprot			
UoN.Ppi.T.43450	Unique.ENSbat_208	514	0	1032	1	513	ref|XP_008140415.1	PREDICTED: LOW QUALITY PROTEIN: cytochrome P450 1A4-like [Eptesicus fuscus]	522	1	513	513	1	Transmembrane protein	Uniprot			
UoN.Ppi.T.43583	Unique.ENSbat_209	721	0	1291	42	720	ref|XP_008157912.1	PREDICTED: zinc finger protein 16 [Eptesicus fuscus]	679	1	679	679	1					
UoN.Ppi.T.42217	Unique.ENSbat_21	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii] – Vesper bat	766	1	241	241	1					
UoN.Ppi.T.41975	Unique.ENSbat_210	156	7E-100	319	1	155	ref|XP_006083130.1	PREDICTED: dual 3',5'-cyclic-AMP and -GMP phosphodiesterase 11A [Myotis lucifugus]	928	774	928	155	0.1670259					
UoN.Ppi.T.42451	Unique.ENSbat_211	680	0	1392	1	679	ref|XP_008139972.1	PREDICTED: stress-70 protein, mitochondrial [Eptesicus fuscus]	679	1	679	679	1					
UoN.Ppi.T.42011	Unique.ENSbat_212	267	7E-180	525	11	267	gb|EPQ07594.1|	Sodium- and chloride-dependent GABA transporter 2 [Myotis brandtii]	728	125	381	257	0.6745407					
UoN.Ppi.T.42694	Unique.ENSbat_213	173	1E-96	303	2	172	ref|XP_008152490.1	PREDICTED: C4b-binding protein alpha chain [Eptesicus fuscus]	607	437	607	171	0.2817133					
UoN.Ppi.T.42497	Unique.ENSbat_214	434	0	624	64	433	ref|XP_008155339.1	PREDICTED: caspase-14-like [Eptesicus fuscus]	362	1	362	370	1.0220994	Endopeptidase/Apoptotic	Uniprot			
UoN.Ppi.T.42497	Unique.ENSbat_215	431	0	624	61	430	ref|XP_008155339.1	PREDICTED: caspase-14-like [Eptesicus fuscus]	362	1	362	370	1.0220994	Endopeptidase/Apoptotic	Uniprot			
UoN.Ppi.T.42520	Unique.ENSbat_216	220	1E-114	360	33	219	ref|XP_014322595.1	PREDICTED: PDZ domain-containing protein 7 isoform X2 [Myotis lucifugus]	916	730	916	187	0.2041485					
UoN.Ppi.T.42520	Unique.ENSbat_217	220	1E-114	360	33	219	ref|XP_014322595.1	PREDICTED: PDZ domain-containing protein 7 isoform X2 [Myotis lucifugus]	916	730	916	187	0.2041485					
UoN.Ppi.T.42524	Unique.ENSbat_218	124	6E-52	174	13	124	ref|XP_014319611.1	PREDICTED: uncharacterized protein LOC102421195 isoform X1 [Myotis lucifugus]	202	7	117	112	0.957265					
UoN.Ppi.T.42524	Unique.ENSbat_219	124	6E-52	174	13	124	ref|XP_014319611.1	PREDICTED: uncharacterized protein LOC102421195 isoform X1 [Myotis lucifugus]	202	7	117	112	0.957265					
UoN.Ppi.T.42218	Unique.ENSbat_22	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					
UoN.Ppi.T.42524	Unique.ENSbat_220	114	1E-49	168	13	109	ref|XP_014319611.1	PREDICTED: uncharacterized protein LOC102421195 isoform X1 [Myotis lucifugus]	202	7	103	97	0.9417476					
UoN.Ppi.T.42536	Unique.ENSbat_221	115	2E-58	188	3	115	gb|ABO15739.1|	leucine rich repeat neuronal 6D [Bos taurus]	138	2	113	113	1					
UoN.Ppi.T.42582	Unique.ENSbat_222	368	0	715	1	367	ref|XP_013818316.1	PREDICTED: nuclease-sensitive element-binding protein 1 [Capra hircus] - The domestic goat	432	66	432	367	0.849537					
UoN.Ppi.T.42651	Unique.ENSbat_223	229	2E-97	298	44	214	ref|XP_004627694.1	PREDICTED: transmembrane protein 178A [Octodon degus] - The degu is a small caviomorph rodent	297	1	172	172	1					
UoN.Ppi.T.42219	Unique.ENSbat_23	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					
UoN.Ppi.T.42220	Unique.ENSbat_24	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					
UoN.Ppi.T.42221	Unique.ENSbat_25	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					
UoN.Ppi.T.42222	Unique.ENSbat_26	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					
UoN.Ppi.T.42223	Unique.ENSbat_27	242	9E-166	489	1	241	ref|XP_015415045.1	PREDICTED: sodium channel protein type 8 subunit alpha-like [Myotis davidii]	766	1	241	241	1					Expression
UoN.Ppi.T.41975	Unique.ENSbat_28	294	2E-166	520	30	283	gb|EPQ06984.1|	Ral GTPase-activating protein subunit alpha-1 [Myotis brandtii] - Vesper bat	1970	453	706	254	0.3597734					
UoN.Ppi.T.42498	Unique.ENSbat_29	169	3E-99	295	25	168	ref|XP_006108582.1	PREDICTED: isocitrate dehydrogenase [NADP] cytoplasmic-like [Myotis lucifugus]	144	1	144	144	1	Enzymatic/Oxidoreductase	Uniprot			
UoN.Ppi.T.43055	Unique.ENSbat_3	810	0	1461	10	809	ref|XP_006094111.1	PREDICTED: transcription elongation factor B polypeptide 3 [Myotis lucifugus]	802	2	802	802	1	Trancription Factor				
UoN.Ppi.T.42565	Unique.ENSbat_30	91	3E-47	169	1	79	ref|XP_008138331.1	PREDICTED: zinc finger protein 382 [Eptesicus fuscus] - Big brown bat	551	1	79	79	1					
UoN.Ppi.T.42617	Unique.ENSbat_31	281	0	592	1	281	ref|XP_010154198.1	PREDICTED: tubulin beta-7 chain [Eurypyga helias]- Sunbittern (neotropical bird) ref|XP_010357708.1| PREDICTED: tubulin beta-7 chain-like [Rhinopithecus roxellana]	298	2	282	281	0.9964539					
UoN.Ppi.T.42874	Unique.ENSbat_32	487	0	775	9	486	ref|XP_006882141.1	PREDICTED: LEM domain-containing protein 2 [Elephantulus edwardii] – cape rock elephant shrew	501	1	501	501	1	NEUROGENESIS	Ensmbl			
UoN.Ppi.T.43391	Unique.ENSbat_33	327	0	591	1	327	ref|XP_008158256.1	PREDICTED: zinc finger and SCAN domain-containing protein 2 [Eptesicus fuscus]	615	1	327	327	1					Expression
UoN.Ppi.T.43474	Unique.ENSbat_34	642	0	962	9	641	ref|XP_008137441.1	PREDICTED: pleckstrin homology domain-containing family G member 3 [Eptesicus fuscus]	1148	520	1148	635	0.5531359					
UoN.Ppi.T.42184	Unique.ENSbat_35	208	5E-98	306	1	207	ref|XP_004604619.1	PREDICTED: wiskott-Aldrich syndrome protein family member 3 isoform X1 [Sorex araneus] – common shrew	501	296	501	207	0.4131737	Actin cytoskeleton regulation	Uniprot			
UoN.Ppi.T.41969	Unique.ENSbat_36	231	1E-83	259	89	230	ref|XP_003227918.2	PREDICTED: 60S ribosomal protein L31 [Anolis carolinensis] Carolina anole (arboreal lizard)	180	34	180	147	0.8166667					
UoN.Ppi.T.42189	Unique.ENSbat_37	682	0	1317	1	681	ref|XP_014405095.1	PREDICTED: protein unc-45 homolog A, partial [Myotis brandtii]	968	288	968	681	0.7035124					
UoN.Ppi.T.42189	Unique.ENSbat_38	682	0	1317	1	681	ref|XP_014405095.1	PREDICTED: protein unc-45 homolog A, partial [Myotis brandtii]	968	288	968	681	0.7035124					
UoN.Ppi.T.42190	Unique.ENSbat_39	682	0	1317	1	681	ref|XP_014405095.1	PREDICTED: protein unc-45 homolog A, partial [Myotis brandtii]	968	288	968	681	0.7035124					
UoN.Ppi.T.43493	Unique.ENSbat_4	290	0	595	1	289	ref|XP_008137234.1	PREDICTED: homeobox protein OTX2 isoform X2 [Eptesicus fuscus] - Big brown bat	289	1	289	289	1	Trancription Factor	Uniprot			
UoN.Ppi.T.42191	Unique.ENSbat_40	768	0	1370	46	767	ref|XP_014405095.1	PREDICTED: protein unc-45 homolog A, partial [Myotis brandtii]	968	247	968	722	0.7458678					
UoN.Ppi.T.41970	Unique.ENSbat_41	126	8E-84	253	1	125	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1					
UoN.Ppi.T.41971	Unique.ENSbat_42	126	8E-84	253	1	125	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1					
UoN.Ppi.T.41972	Unique.ENSbat_43	126	8E-84	253	1	125	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1					
UoN.Ppi.T.41973	Unique.ENSbat_44	126	8E-84	253	1	125	ref|NP_000984.1|	60S ribosomal protein L31 isoform 1 [Homo sapiens]ref|NP_071951.1| 60S ribosomal protein L31 [Rattus norvegicus]ref|NP_444487.1| 60S ribosomal protein L31 [Mus	125	1	125	125	1					
UoN.Ppi.T.41974	Unique.ENSbat_45	231	1E-83	259	89	230	ref|XP_003227918.2	PREDICTED: 60S ribosomal protein L31 [Anolis carolinensis] - Carolina anole (arboreal lizard)	180	34	180	147	0.8166667					
UoN.Ppi.T.42272	Unique.ENSbat_46	1852	0	3818	1	1851	ref|XP_008149892.1	PREDICTED: unconventional myosin-X [Eptesicus fuscus]	2055	205	2055	1851	0.9007299	Motor/ Transport and Actin binding pr	Uniprot			Expression
UoN.Ppi.T.41999	Unique.ENSbat_47	250	2E-134	389	64	249	ref|NP_001012886.1	ADP-ribosylation factor-like protein 8A [Gallus gallus]ref|XP_005426853.1| PREDICTED: ADP-ribosylation factor-like protein 8A [Geospiza fortis]ref|XP_005530237.1| P	186	1	186	186	1					
UoN.Ppi.T.42898	Unique.ENSbat_48	680	0	1392	1	679	ref|XP_008139972.1	PREDICTED: stress-70 protein, mitochondrial [Eptesicus fuscus]	679	1	679	679	1	Chaperonne/Unfolded protein binding	Uniprot			
UoN.Ppi.T.43190	Unique.ENSbat_49	516	0	902	1	487	ref|XP_008150889.1	PREDICTED: snRNA-activating protein complex subunit 4 [Eptesicus fuscus]	1313	1	496	496	1					
UoN.Ppi.T.42431	Unique.ENSbat_5	148	8E-101	310	1	147	ref|XP_006094585.1	PREDICTED: E3 ubiquitin-protein ligase TRIM68 [Myotis lucifugus]ref|XP_014316234.1| PREDICTED: E3 ubiquitin-protein ligase TRIM68 [Myotis lucifugus]ref|XP_014	484	338	484	147	0.303719	Enzymatic and Metal ion binding	Uniprot			
UoN.Ppi.T.43278	Unique.ENSbat_50	939	0	1686	1	939	ref|XP_008139872.1	PREDICTED: CREB-binding protein isoform X2 [Eptesicus fuscus]	2397	1249	2191	943	0.4303971	Transcription activator	Uniprot			
UoN.Ppi.T.42273	Unique.ENSbat_51	638	0	1018	1	605	ref|XP_014302152.1	PREDICTED: pleckstrin homology-like domain family B member 1, partial [Myotis lucifugus]	957	1	629	629	1					
UoN.Ppi.T.42468	Unique.ENSbat_52	143	3E-90	271	1	142	ref|XP_008150697.1	PREDICTED: mitochondrial intermembrane space import and assembly protein 40 [Eptesicus fuscus]	141	1	141	142	1.0070922					
UoN.Ppi.T.42468	Unique.ENSbat_53	143	3E-90	271	1	142	ref|XP_008150697.1	PREDICTED: mitochondrial intermembrane space import and assembly protein 40 [Eptesicus fuscus]	141	1	141	142	1.0070922					
UoN.Ppi.T.43266	Unique.ENSbat_54	203	5E-108	319	39	202	ref|XP_010202881.1	PREDICTED: corticotropin-releasing factor receptor 1 [Colius striatus] – speckled mouse bird	168	5	168	164	0.9761905					
UoN.Ppi.T.43553	Unique.ENSbat_55	363	0	733	1	361	ref|XP_008140277.1	PREDICTED: glutamate--cysteine ligase catalytic subunit isoform X1 [Eptesicus fuscus]	640	1	361	361	1	Enzymatic/Nucleotide binding	Uniprot			
UoN.Ppi.T.43554	Unique.ENSbat_56	363	0	733	1	361	ref|XP_008140277.1	PREDICTED: glutamate--cysteine ligase catalytic subunit isoform X1 [Eptesicus fuscus]	640	1	361	361	1	Enzymatic/Nucleotide binding	Uniprot			
UoN.Ppi.T.42481	Unique.ENSbat_57	1783	0	2373	22	1547	ref|XP_010995206.1	PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1H, partial [Camelus dromedarius] - dromedary (Arabian camel)	2050	533	2050	1542	0.7521951					Expression

UoN.Ppi.T.43379	Unique.ENSbat_58	1157	0	1987	39	1133	ref|XP_008158879.1	PREDICTED: transcriptional regulator ATRX [Eptesicus fuscus]	2446	1	1093	1095	1.0018298					
UoN.Ppi.T.43380	Unique.ENSbat_59	1157	0	1987	39	1133	ref|XP_008158879.1	PREDICTED: transcriptional regulator ATRX [Eptesicus fuscus]	2446	1	1093	1095	1.0018298	transcriptional regulation and chroma	Uniprot			
UoN.Ppi.T.42902	Unique.ENSbat_6	211	1E-65	213	42	184	ref|XP_006764348.1	PREDICTED: transmembrane protein 225-like [Myotis davidii]ref|XP_015419802.1| PREDICTED: transmembrane protein 225-like [Myotis davidii]ref|XP_015419803.1| I	207	65	207	143	0.6908213	Transmembrane protein	Uniprot		Expression	
UoN.Ppi.T.43381	Unique.ENSbat_60	1157	0	1987	39	1133	ref|XP_008158879.1	PREDICTED: transcriptional regulator ATRX [Eptesicus fuscus]	2446	1	1093	1095	1.0018298					
UoN.Ppi.T.43382	Unique.ENSbat_61	1157	0	1987	39	1133	ref|XP_008158879.1	PREDICTED: transcriptional regulator ATRX [Eptesicus fuscus]	2446	1	1093	1095	1.0018298	transcriptional regulation and chroma	Uniprot		Expression	
UoN.Ppi.T.43383	Unique.ENSbat_62	1157	0	1987	39	1133	ref|XP_008158879.1	PREDICTED: transcriptional regulator ATRX [Eptesicus fuscus]	2446	1	1093	1095	1.0018298					
UoN.Ppi.T.43389	Unique.ENSbat_63	2123	0	3344	1	2122	ref|XP_006764329.1	PREDICTED: armadillo repeat-containing X-linked protein 4 isoform X3 [Myotis davidii]	2131	1	2131	2140	1.0042234	Transmembrane protein	Uniprot			
UoN.Ppi.T.43519	Unique.ENSbat_64	243	1E-129	379	1	213	ref|XP_013824041.1	PREDICTED: LOW QUALITY PROTEIN: tropomyosin alpha-4 chain-like [Capra hircus] - Domestic goat	243	9	221	213	0.9638009					
UoN.Ppi.T.41962	Unique.ENSbat_65	290	0	595	1	289	ref|XP_008137234.1	PREDICTED: homeobox protein OTX2 isoform X2 [Eptesicus fuscus]	289	1	289	289	1	Trancription Factor	Uniprot			
UoN.Ppi.T.42131	Unique.ENSbat_66	371	0	743	1	370	ref|XP_005867745.1	PREDICTED: rab GTPase-activating protein 1-like isoform X5 [Myotis brandtii]ref|XP_006088991.1| PREDICTED: rab GTPase-activating protein 1-like isoform X5 [Myo	370	1	370	370	1					
UoN.Ppi.T.43071	Unique.ENSbat_67	112	4E-71	226	1	111	ref|XP_008154271.1	PREDICTED: guanine nucleotide-binding protein subunit alpha-12 [Eptesicus fuscus]	280	170	280	111	0.3964286					
UoN.Ppi.T.42229	Unique.ENSbat_68	345	7E-126	372	140	327	ref|XP_005638175.1	PREDICTED: protein FAM92A1 isoform X1 [Canis lupus familiaris] - Domestic dog	210	7	194	188	0.9690722					
UoN.Ppi.T.42231	Unique.ENSbat_69	345	7E-126	372	140	327	ref|XP_005638175.1	PREDICTED: protein FAM92A1 isoform X1 [Canis lupus familiaris] - Domestic dog	210	7	194	188	0.9690722					
UoN.Ppi.T.42969	Unique.ENSbat_7	263	5E-138	423	12	262	ref|XP_014312944.1	PREDICTED: extracellular sulfatase Sulf-1 isoform X2 [Myotis lucifugus]	902	643	902	260	0.2882483	Calcium ion binding	Uniprot			
UoN.Ppi.T.42337	Unique.ENSbat_70	646	0	1327	1	645	ref|XP_005877221.1	PREDICTED: vascular endothelial growth factor receptor 2 isoform X1 [Myotis brandtii]gb|EPQ14516.1| Vascular endothelial growth factor receptor 2 [Myotis brandtii]	1349	705	1349	645	0.4781319					
UoN.Ppi.T.42606	Unique.ENSbat_71	202	1E-145	416	1	202	ref|XP_006753835.2	PREDICTED: teneurin-1-like [Myotis davidii]	207	1	202	202	1					
UoN.Ppi.T.42968	Unique.ENSbat_72	97	2E-61	194	1	96	gb|EDL08744.1|	mCG145920, isoform CRA_c, partial [Mus musculus]	109	14	109	96	0.8807339					
UoN.Ppi.T.43280	Unique.ENSbat_73	257	2E-156	451	1	256	ref|XP_014397753.1	PREDICTED: retinol dehydrogenase 11 [Myotis brandtii]	330	75	330	256	0.7757576					
UoN.Ppi.T.43425	Unique.ENSbat_74	221	2E-157	447	1	220	ref|XP_006098486.1	PREDICTED: transmembrane emp24 domain-containing protein 10 isoform X3 [Myotis lucifugus]ref|XP_008137040.1| PREDICTED: transmembrane emp24 domain-co	220	1	220	220	1					
UoN.Ppi.T.43488	Unique.ENSbat_75	599	0	1085	1	598	ref|XP_016061495.1	PREDICTED: general transcription factor II-I repeat domain-containing protein 2B-like isoform X1 [Miniopterus natalensis] - The Natal long-fingered bat (Vesper bat) ref|	598	1	598	598	1	Trancription Factor	Uniprot			
UoN.Ppi.T.43592	Unique.ENSbat_76	247	1E-134	414	1	246	ref|XP_014312944.1	PREDICTED: extracellular sulfatase Sulf-1 isoform X2 [Myotis lucifugus]	902	648	902	255	0.2827051					
UoN.Ppi.T.43605	Unique.ENSbat_77	353	9E-178	531	1	329	ref|XP_006108616.1	PREDICTED: LOW QUALITY PROTEIN: lysosomal alpha-glucosidase [Myotis lucifugus]	945	1	328	332	1.0121951				Expression	
UoN.Ppi.T.43606	Unique.ENSbat_78	353	9E-178	531	1	329	ref|XP_006108616.1	PREDICTED: LOW QUALITY PROTEIN: lysosomal alpha-glucosidase [Myotis lucifugus]	945	1	328	332	1.0121951					
UoN.Ppi.T.42152	Unique.ENSbat_79	107	9E-70	217	1	106	ref|XP_010357954.1	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L36a.[Rhinopithecus roxellana] - The golden snub-nosed monkey	143	38	143	106	0.7412587	Translational/Ribonucleo prtoein	Uniprot			
UoN.Ppi.T.42973	Unique.ENSbat_8	452	0	942	1	451	ref|NP_001069723.1	tubulin gamma-1 chain [Bos taurus]ref|XP_001493897.2| PREDICTED: tubulin gamma-1 chain [Equus caballus]ref|XP_004319199.1| PREDICTED: tubulin gamma-1 cl	451	1	451	451	1	Nucleotide binding	Uniprot KB			
UoN.Ppi.T.42153	Unique.ENSbat_80	107	9E-70	217	1	106	ref|XP_010357954.1	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L36a [Rhinopithecus roxellana] - The golden snub-nosed monkey	143	38	143	106	0.7412587					
UoN.Ppi.T.42153	Unique.ENSbat_81	107	9E-70	217	1	106	ref|XP_010357954.1	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L36a [Rhinopithecus roxellana] - The golden snub-nosed monkey	143	38	143	106	0.7412587					
UoN.Ppi.T.42271	Unique.ENSbat_82	317	0	522	58	316	ref|XP_005861969.1	PREDICTED: LIM and SH3 domain protein 1 isoform X1 [Myotis brandtii]gb|EPQ05062.1| LIM and SH3 domain protein 1 [Myotis brandtii]	260	1	260	260	1	Zinc ion binding	Uniprot			
UoN.Ppi.T.42300	Unique.ENSbat_83	277	2E-161	461	30	276	ref|NP_001036818.1	tropomyosin alpha-3 chain isoform Tpm3.4cy [Homo sapiens]ref|XP_001147418.1| PREDICTED: tropomyosin alpha-3 chain isoform X19 [Pan troglodytes] – common c	247	1	247	247	1					
UoN.Ppi.T.42408	Unique.ENSbat_84	452	0	944	1	451	gb|AAX29190.1|	tubulin alpha 3 [synthetic construct]	452	1	451	451	1					
UoN.Ppi.T.42408	Unique.ENSbat_85	335	1E-139	409	1	309	ref|XP_004052011.1	PREDICTED: uncharacterized protein LOC101125632 [Gorilla gorilla gorilla]	268	1	266	309	1.1616541					
UoN.Ppi.T.42410	Unique.ENSbat_86	532	0	872	1	521	gb|AAH58206.1|	Cacna1h protein, partial [Mus musculus] – Mouse	541	1	519	522	1.0057803	Transmembrane/Calcium ion/Voltage	Uniprot			
UoN.Ppi.T.42549	Unique.ENSbat_87	192	1E-126	377	3	191	ref|XP_008158858.1	PREDICTED: UDP-glucuronosyltransferase 2B31-like isoform X3 [Eptesicus fuscus]	445	257	445	189	0.4247191					
UoN.Ppi.T.42694	Unique.ENSbat_88	325	0	576	1	325	ref|XP_008158256.1	PREDICTED: zinc finger and SCAN domain-containing protein 2 [Eptesicus fuscus]	615	1	325	325	1					
UoN.Ppi.T.43045	Unique.ENSbat_89	567	0	953	1	566	ref|XP_008136658.1	PREDICTED: leucine-rich repeat flightless-interacting protein 1 isoform X1 [Eptesicus fuscus]	640	1	640	640	1	coiled coil	Uniprot			
UoN.Ppi.T.42973	Unique.ENSbat_9	452	0	942	1	451	ref|NP_001069723.1	tubulin gamma-1 chain [Bos taurus]ref|XP_001493897.2| PREDICTED: tubulin gamma-1 chain [Equus caballus]ref|XP_004319199.1| PREDICTED: tubulin gamma-1 cl	451	1	451	451	1	Nucleotide binding	Uniprot KB			
UoN.Ppi.T.43046	Unique.ENSbat_90	567	0	953	1	566	ref|XP_008136658.1	PREDICTED: leucine-rich repeat flightless-interacting protein 1 isoform X1 [Eptesicus fuscus]	640	1	640	640	1	coiled coil	Uniprot			
UoN.Ppi.T.43060	Unique.ENSbat_91	150	8E-90	291	1	149	ref|XP_005862691.1	PREDICTED: pyruvate dehydrogenase phosphatase regulatory subunit, mitochondrial isoform X1 [Myotis brandtii]ref|XP_006091286.1| PREDICTED: pyruvate dehydro	878	1	149	149	1					
UoN.Ppi.T.42763	Unique.ENSbat_92	259	2E-159	454	39	258	ref|XP_008142893.1	PREDICTED: BTB/POZ domain-containing protein KCTD17 isoform X3 [Eptesicus fuscus]	220	1	220	220	1					
UoN.Ppi.T.42298	Unique.ENSbat_93	105	1E-68	213	1	104	ref|XP_008143876.1	PREDICTED: small ubiquitin-related modifier 3 [Eptesicus fuscus]	104	1	104	104	1					
UoN.Ppi.T.42427	Unique.ENSbat_94	197	6E-105	331	8	196	ref|XP_008145268.1	PREDICTED: EH domain-binding protein 1-like protein 1 [Eptesicus fuscus]	774	586	774	189	0.244186					
UoN.Ppi.T.43045	Unique.ENSbat_95	166	4E-92	276	24	165	ref|XP_016052281.1	PREDICTED: probable hydrolase PNKD [Miniopterus natalensis] - The Natal long-fingered bat (Vesper bat)	142	1	142	142	1					
UoN.Ppi.T.43177	Unique.ENSbat_96	207	9E-132	383	20	207	gb|AIC59161.1|	PIK3R2, partial [synthetic construct]	253	45	232	188	0.8103448					
UoN.Ppi.T.43361	Unique.ENSbat_97	369	0	749	1	368	ref|XP_008142808.1	PREDICTED: hsc70-interacting protein isoform X1 [Eptesicus fuscus]	368	1	368	368	1	Strucutral domain- 3D /fold	Uniprot			
UoN.Ppi.T.43473	Unique.ENSbat_98	908	0	1266	91	907	ref|XP_008150603.1	PREDICTED: NEDD4-binding protein 1 [Eptesicus fuscus]	845	7	845	843	0.9976331					
UoN.Ppi.T.43474	Unique.ENSbat_99	831	0	1090	99	799	ref|XP_008150603.1	PREDICTED: NEDD4-binding protein 1 [Eptesicus fuscus]	845	7	727	724	0.9958735					

Appendix - 3

BioGPS

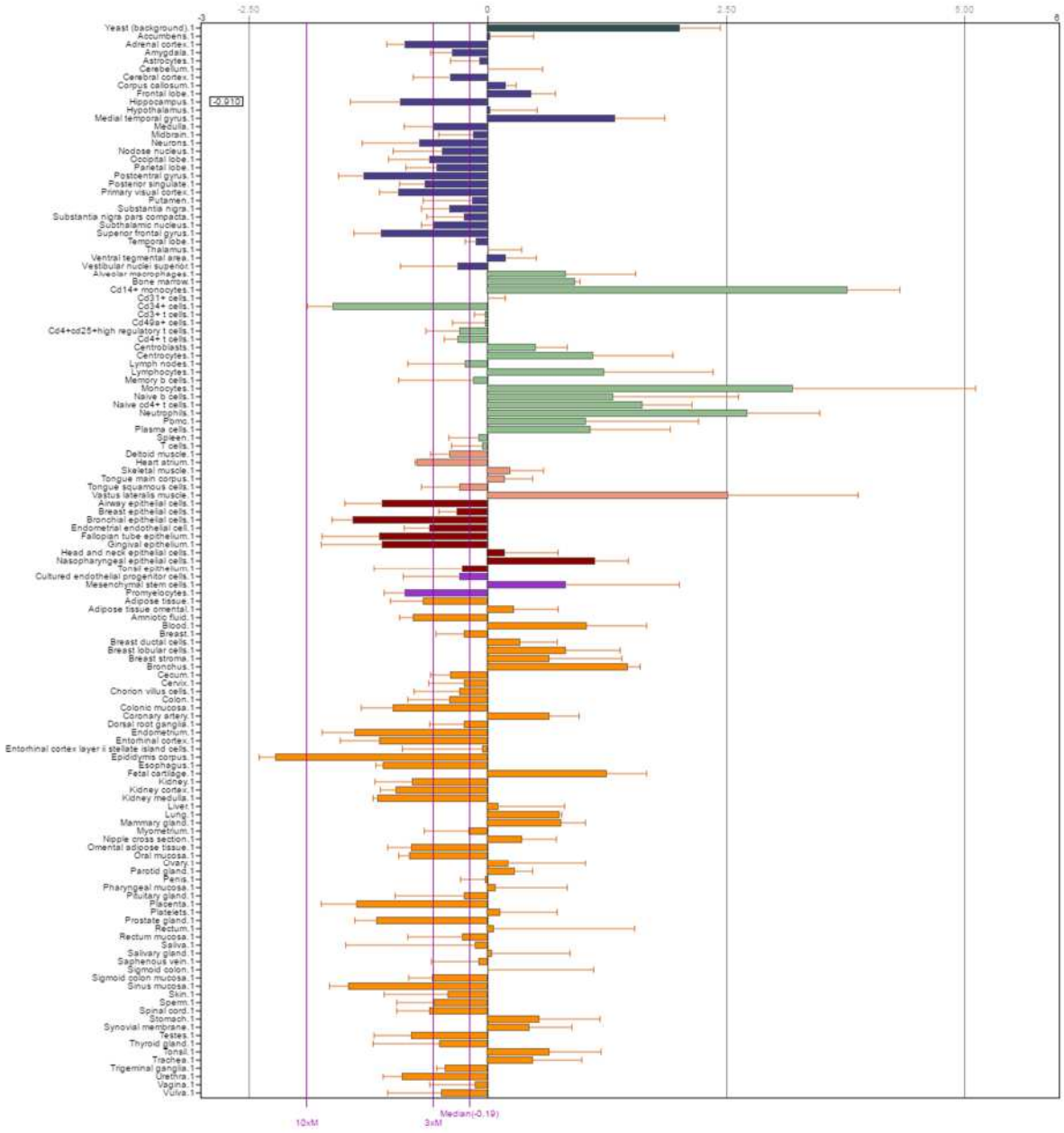


Figure 1: Data set (CDH23), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=22295).

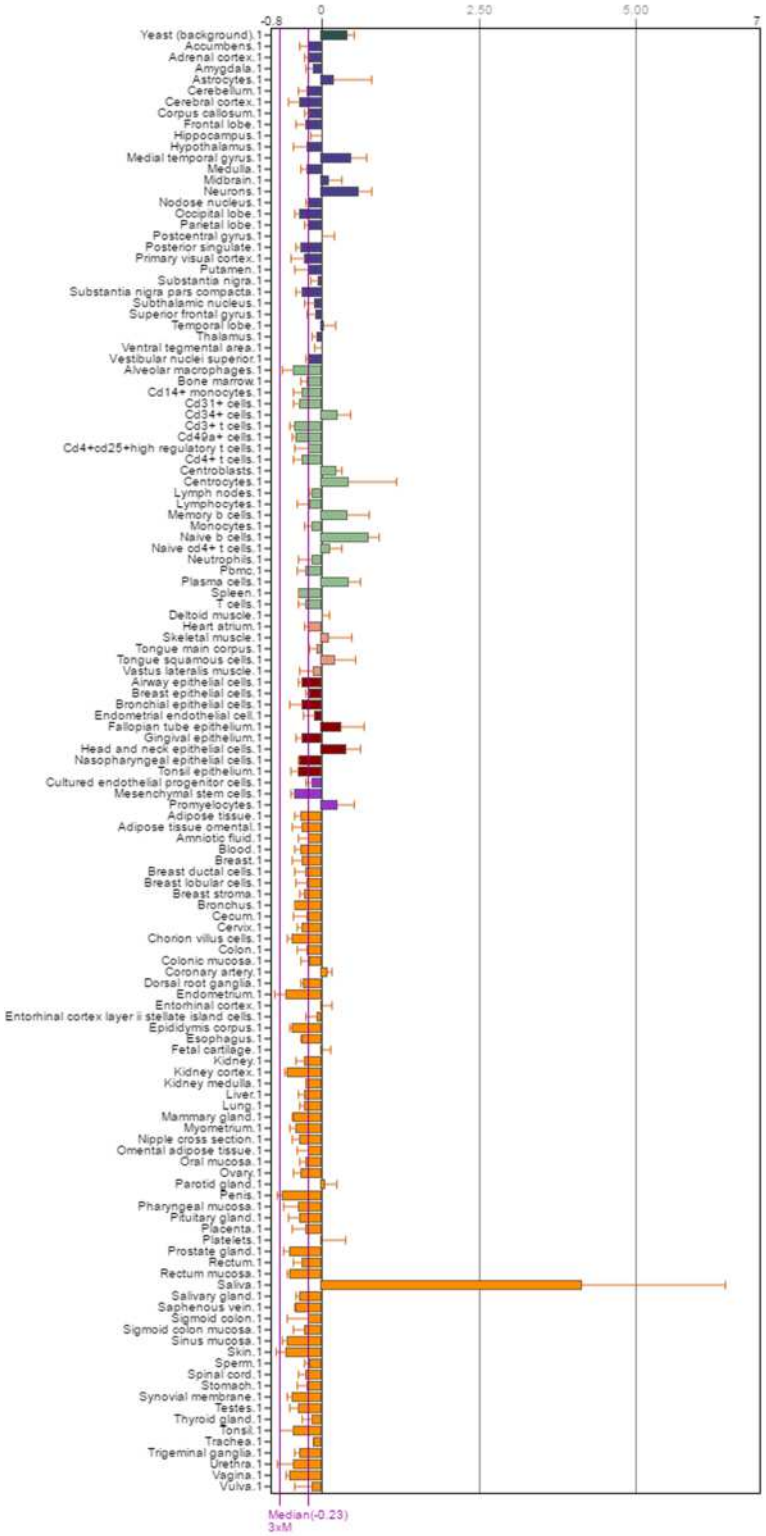


Figure 2: Data set (pcdh15), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=11994).

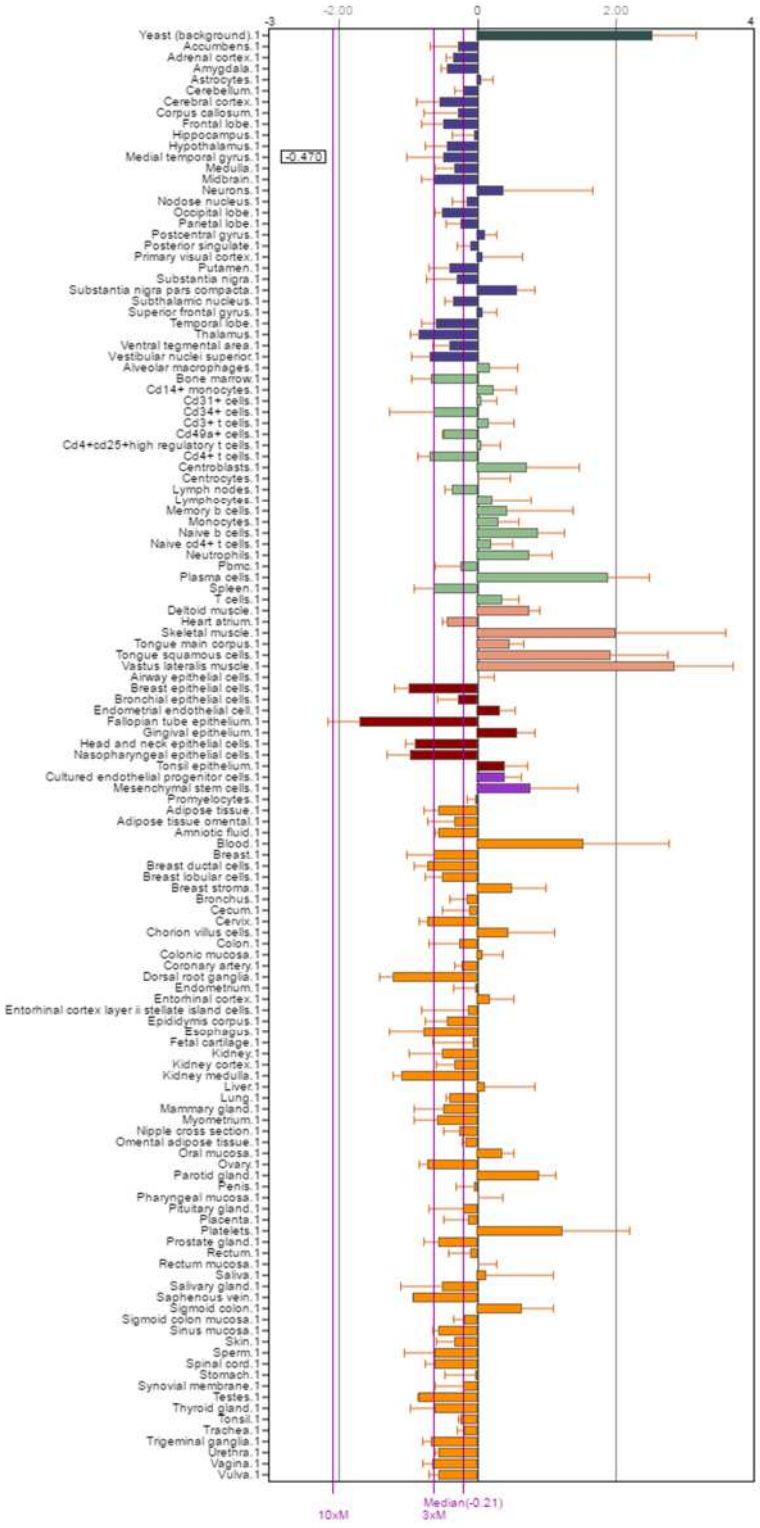


Figure 3: Data set (chrna10), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=504186).

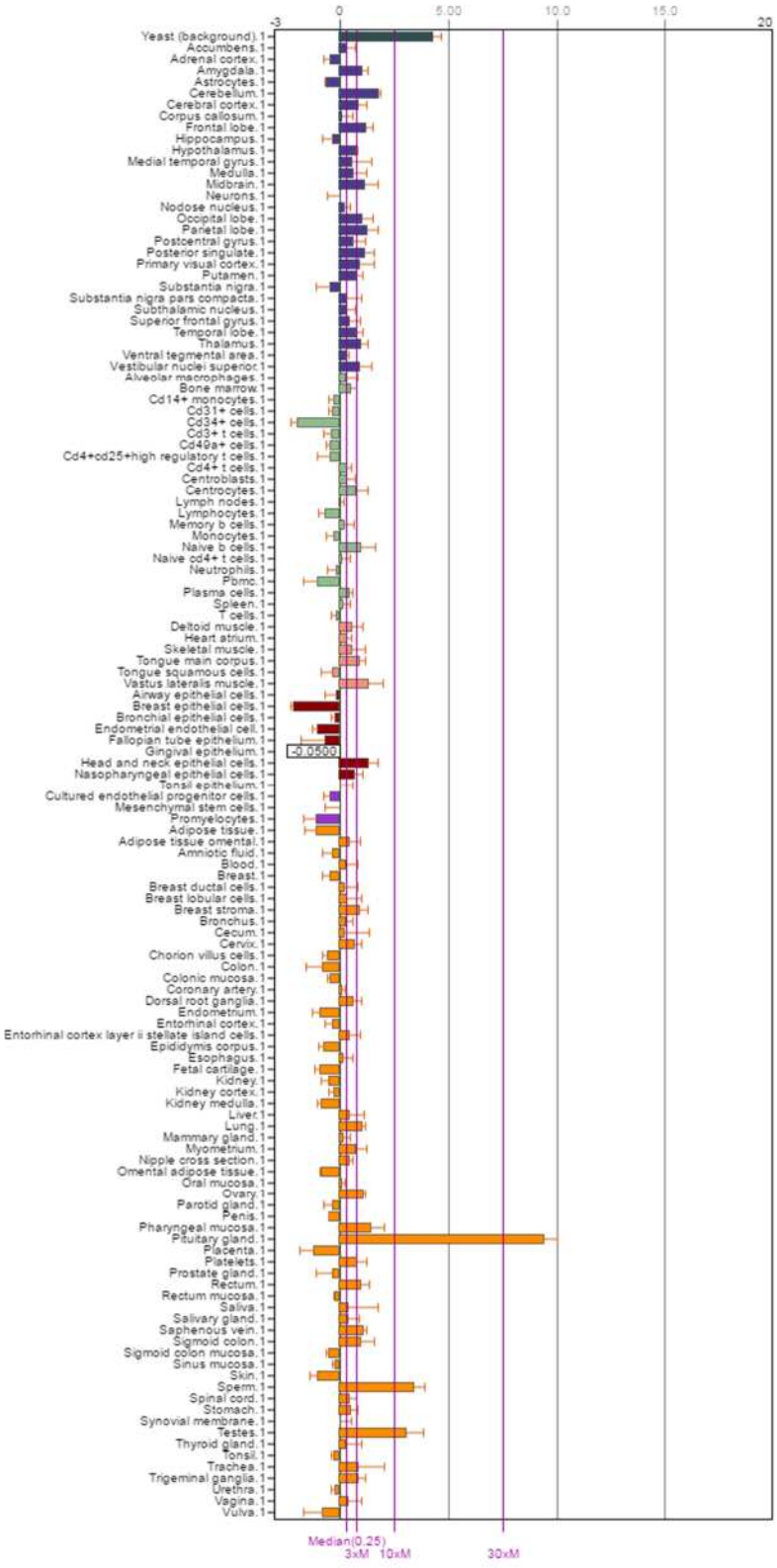


Figure 4: Data set (myo15a), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=17910).

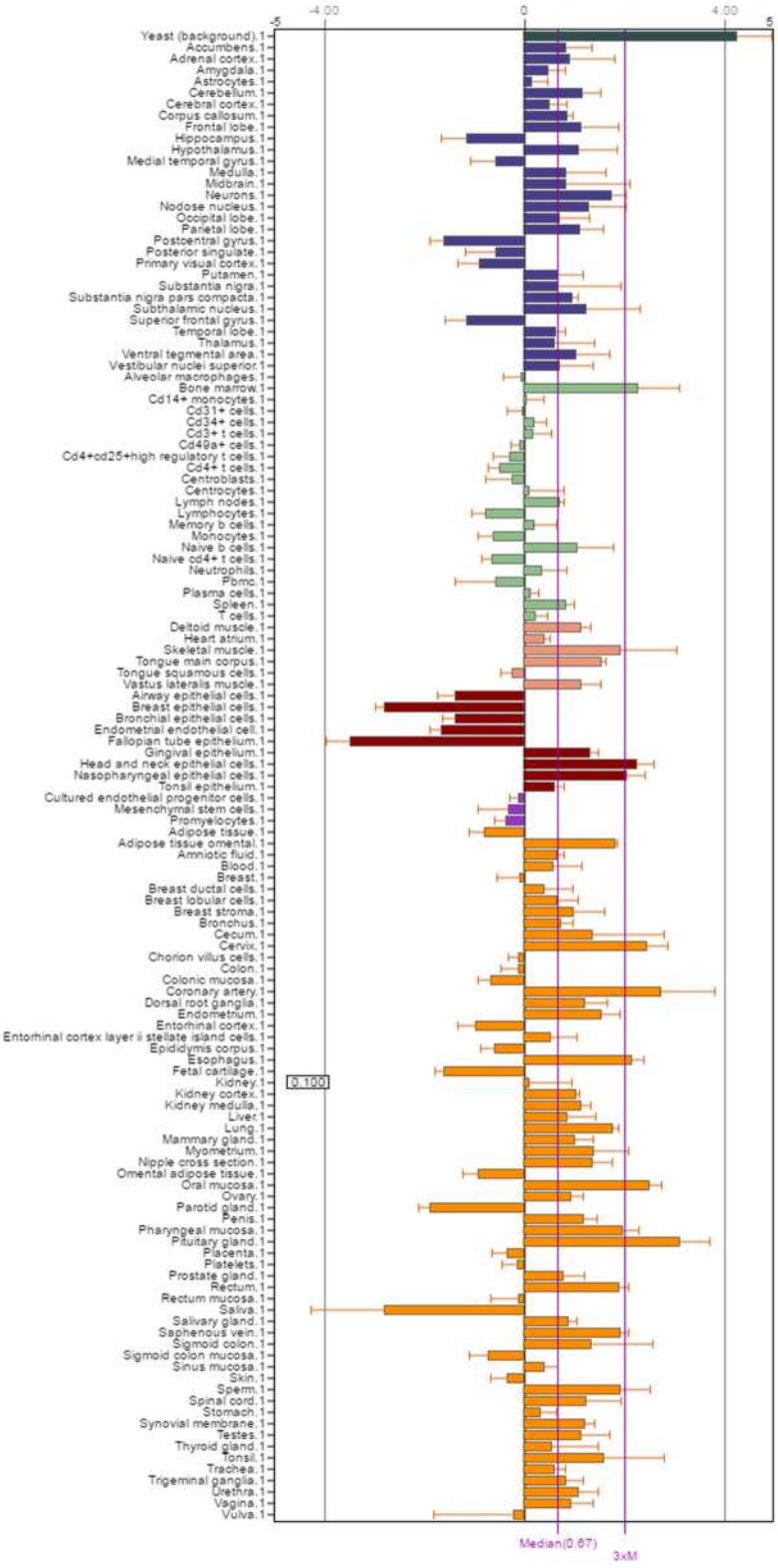


Figure 5: Data set (ush1g), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=16470).

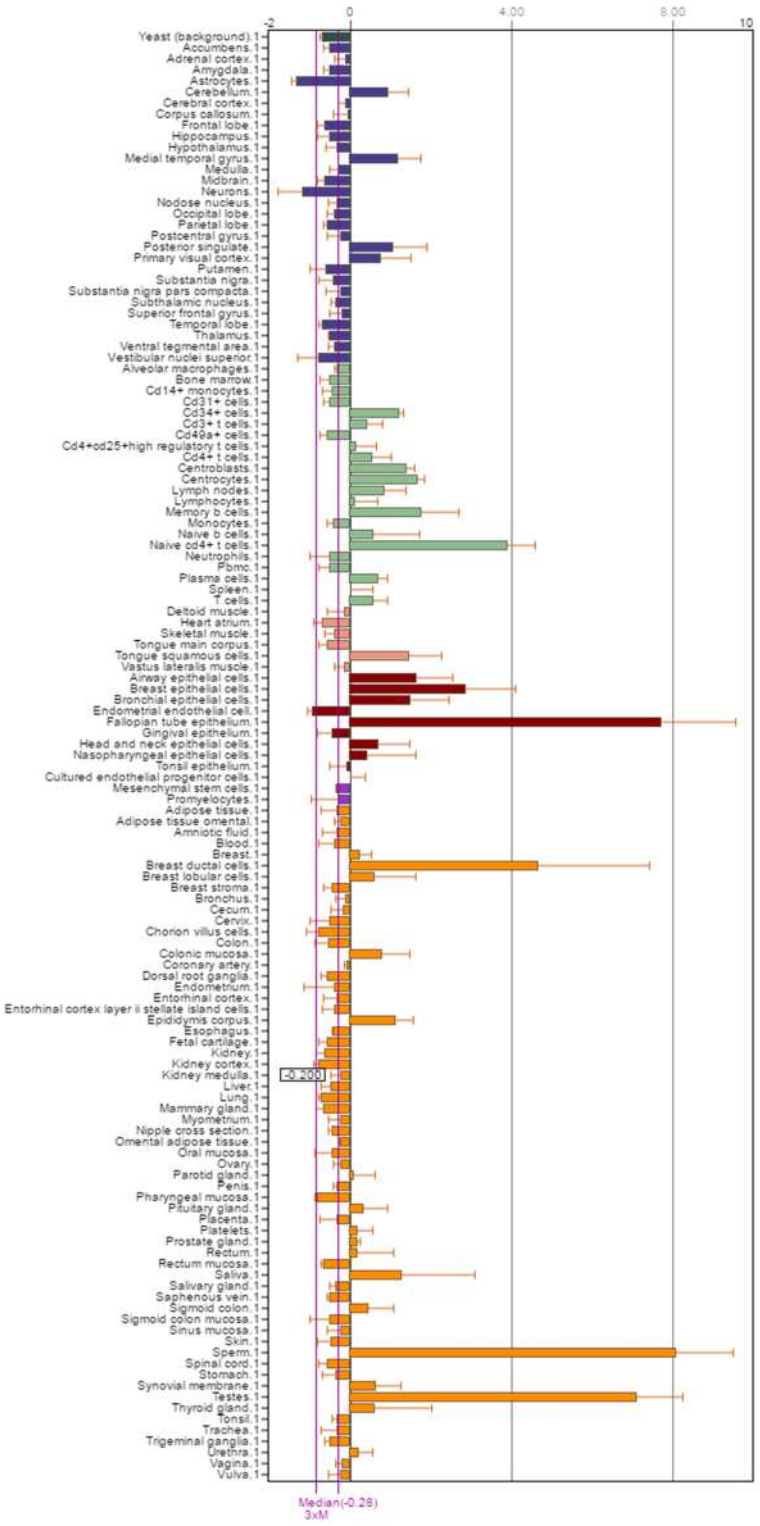


Figure 6: Data set (Strc), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=140476).

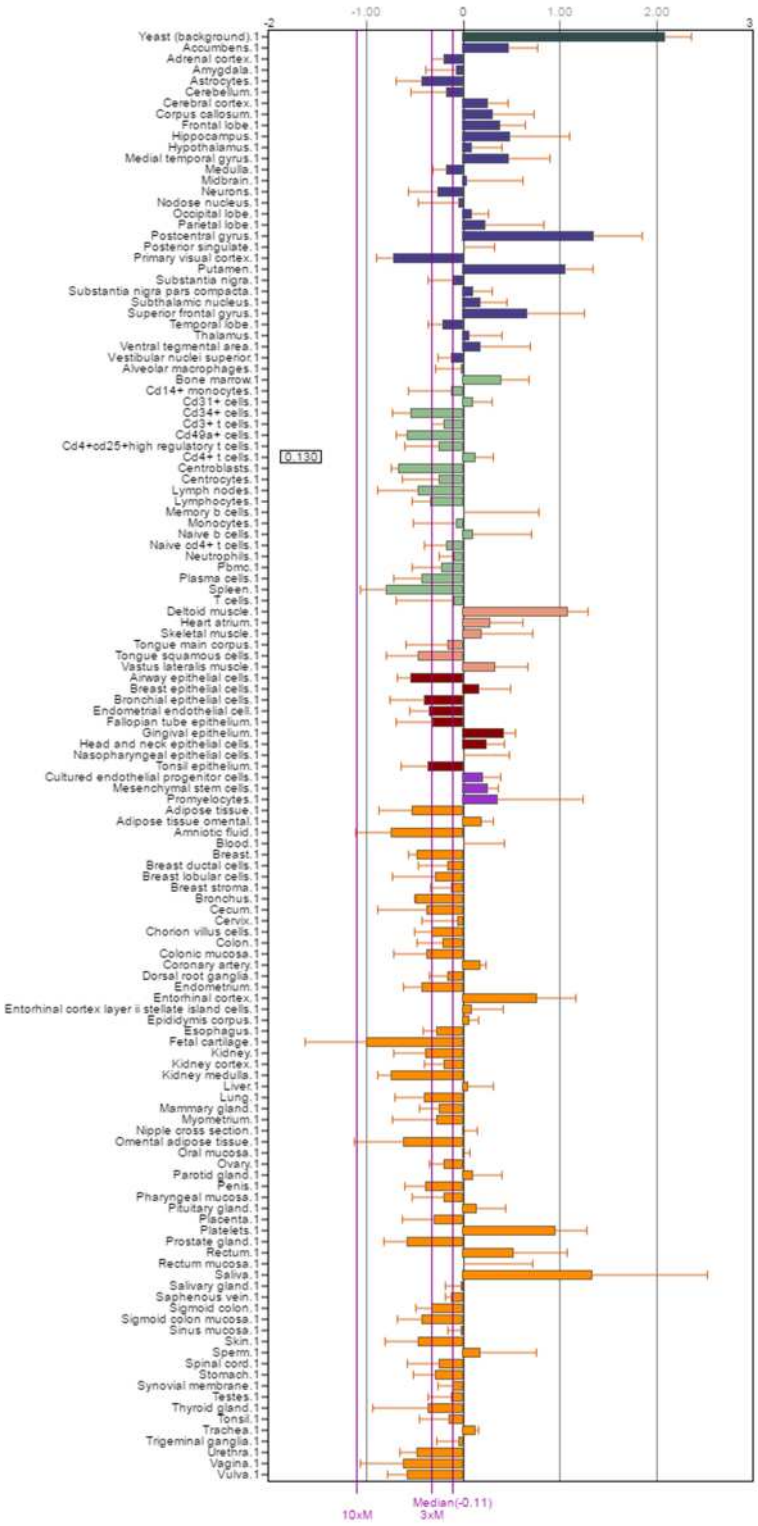


Figure 7: Data set (Tect a), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=21683).

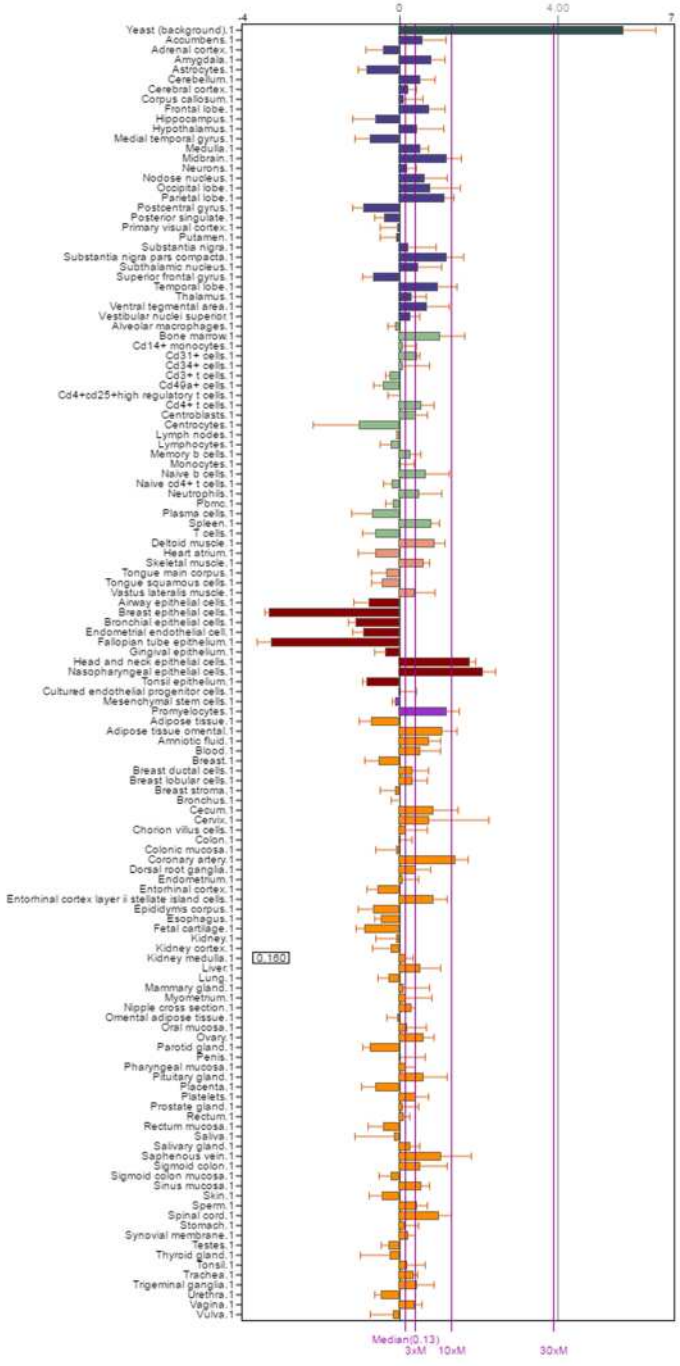


Figure 8: Data set (Tect b), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=21684).

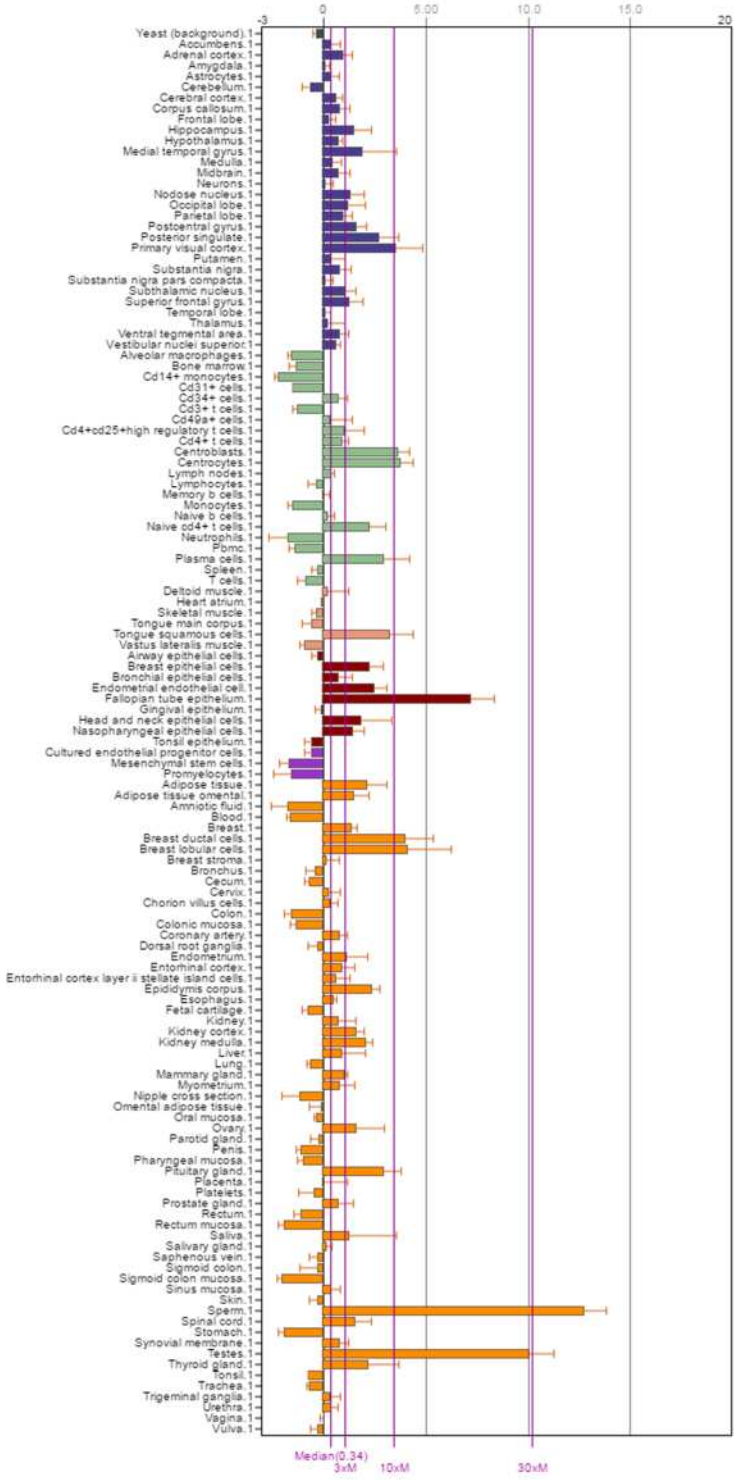


Figure 9: Data set (Dfnb59), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=381375).

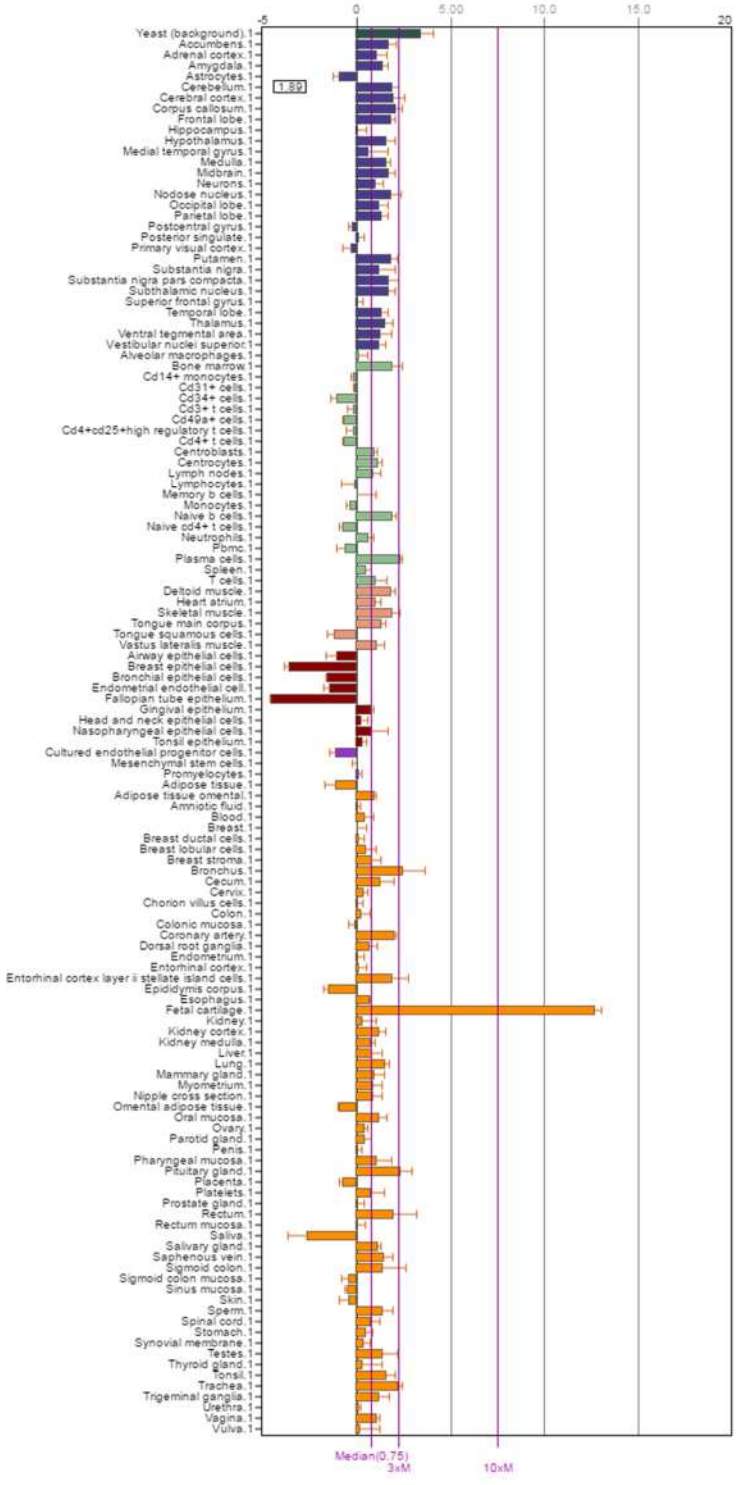


Figure 10: Data set (Col 11a2), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=12815).

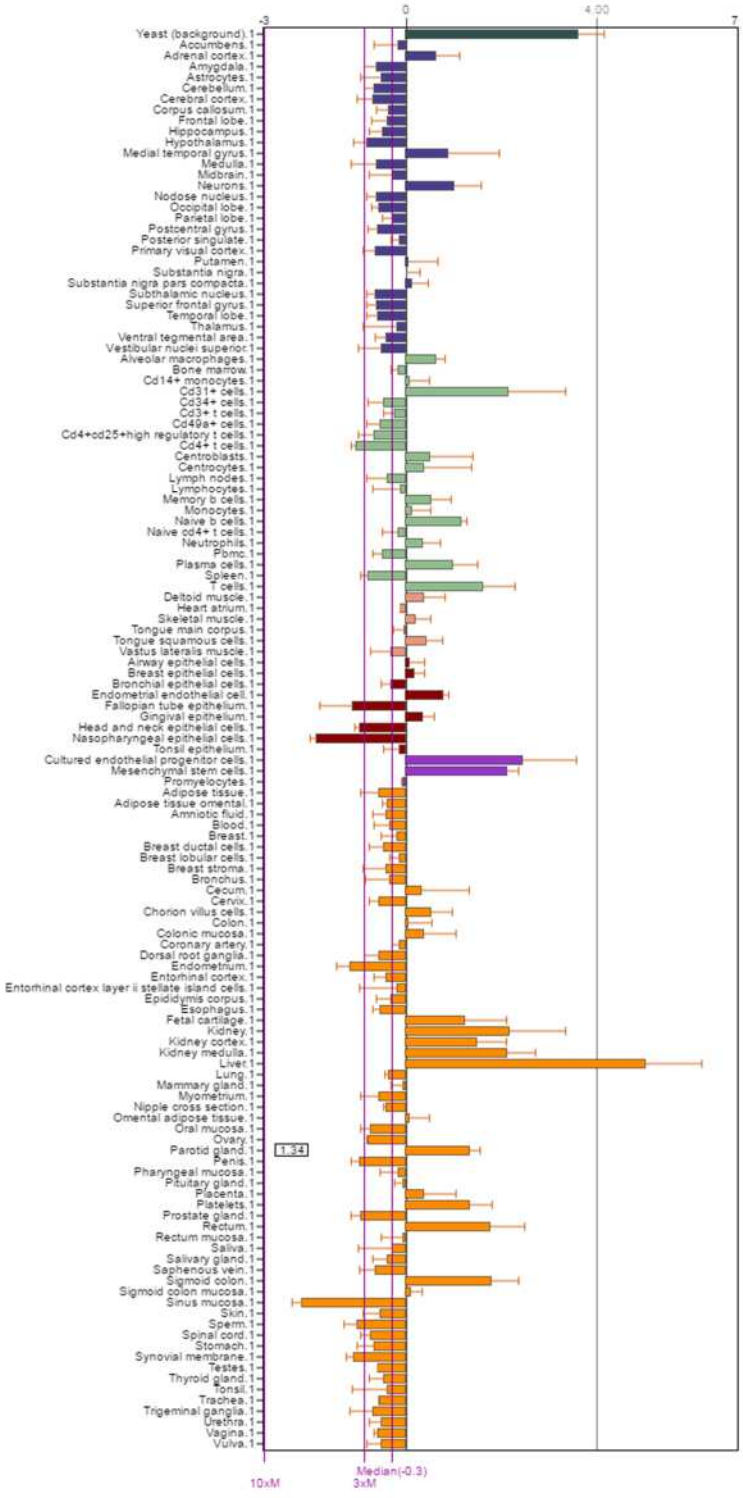


Figure 11: Data set (cldn 14), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=56173).

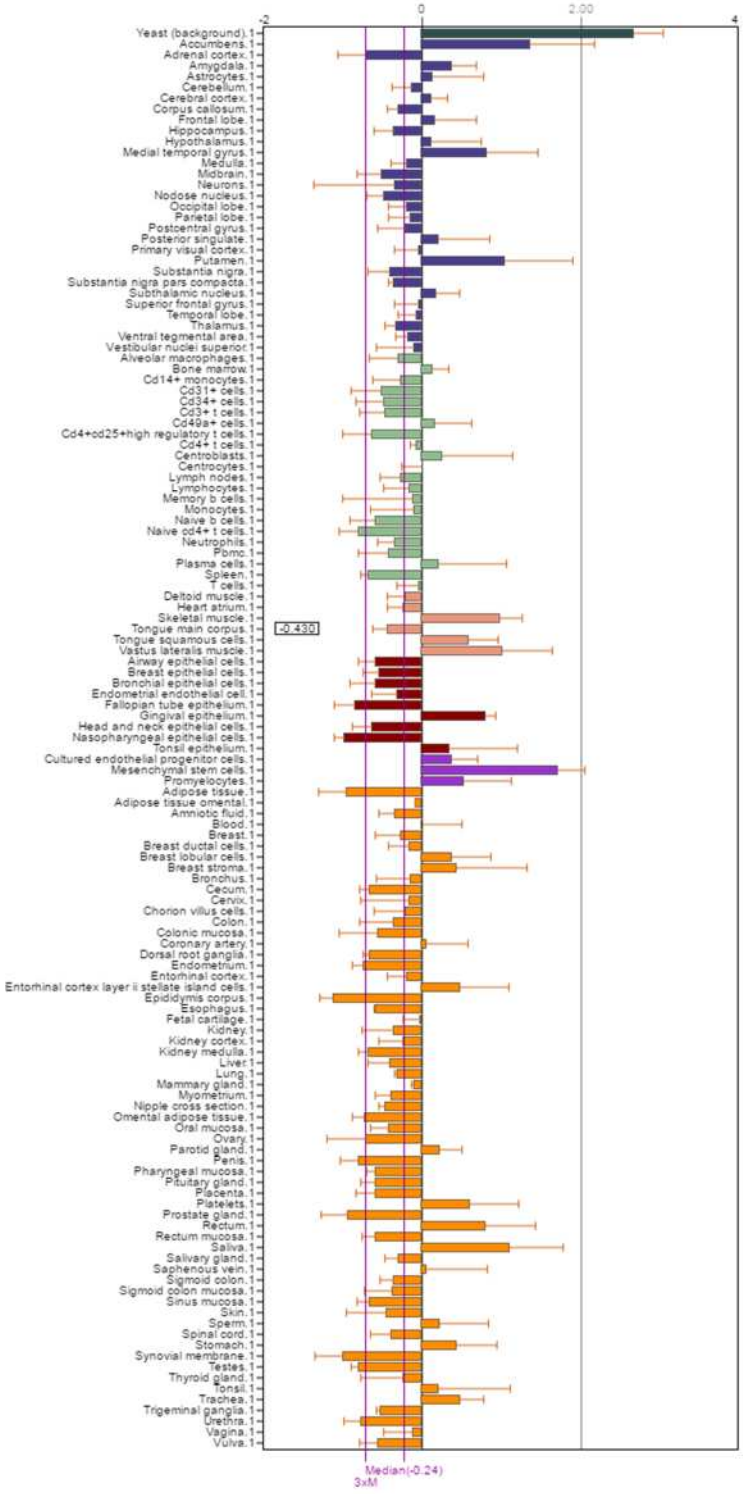


Figure 12: Data set (Pou3f4), displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=18994).

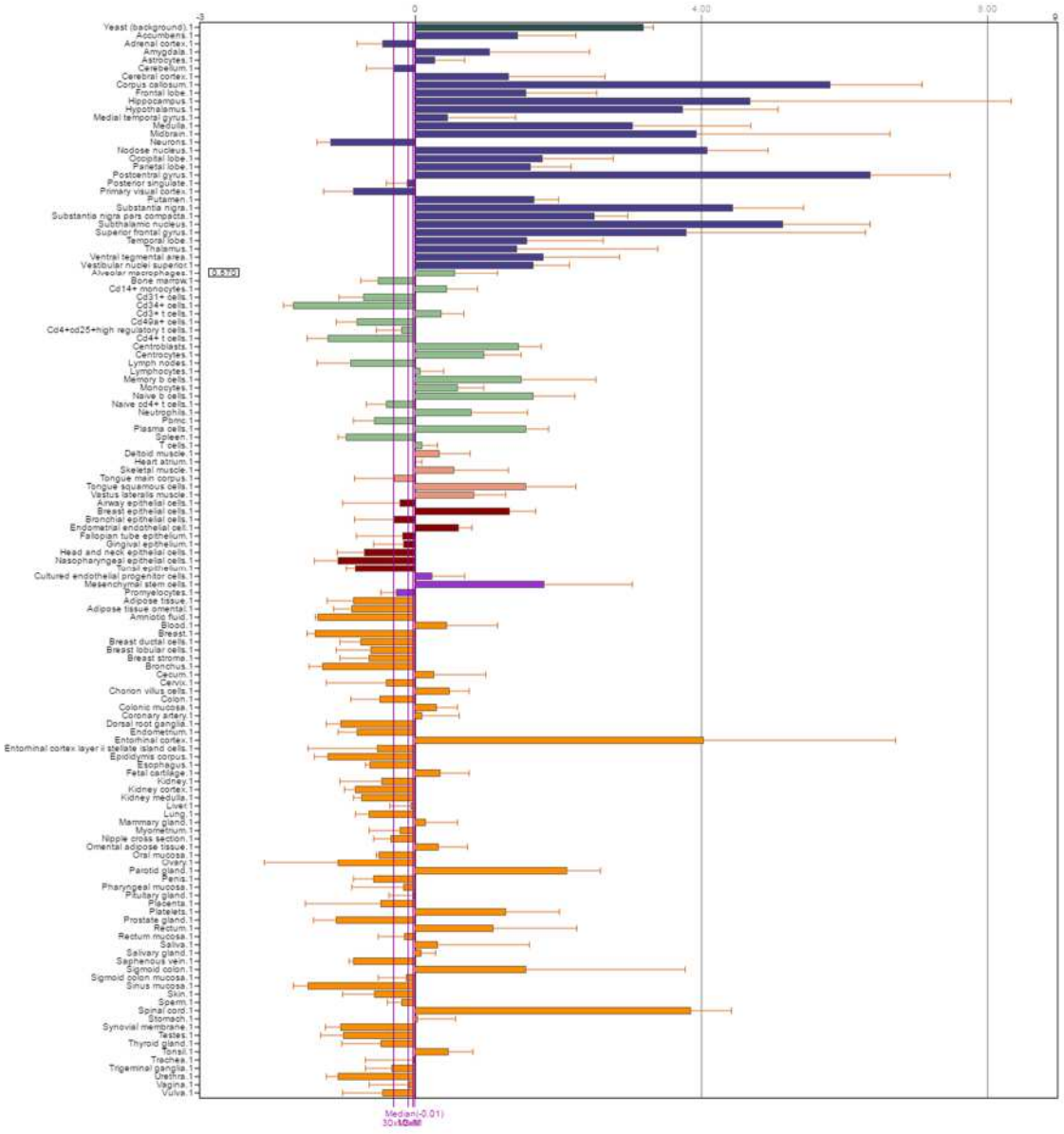


Figure 13: Data set (mbp) mitochondrial protein of interest of this study, displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=4155).

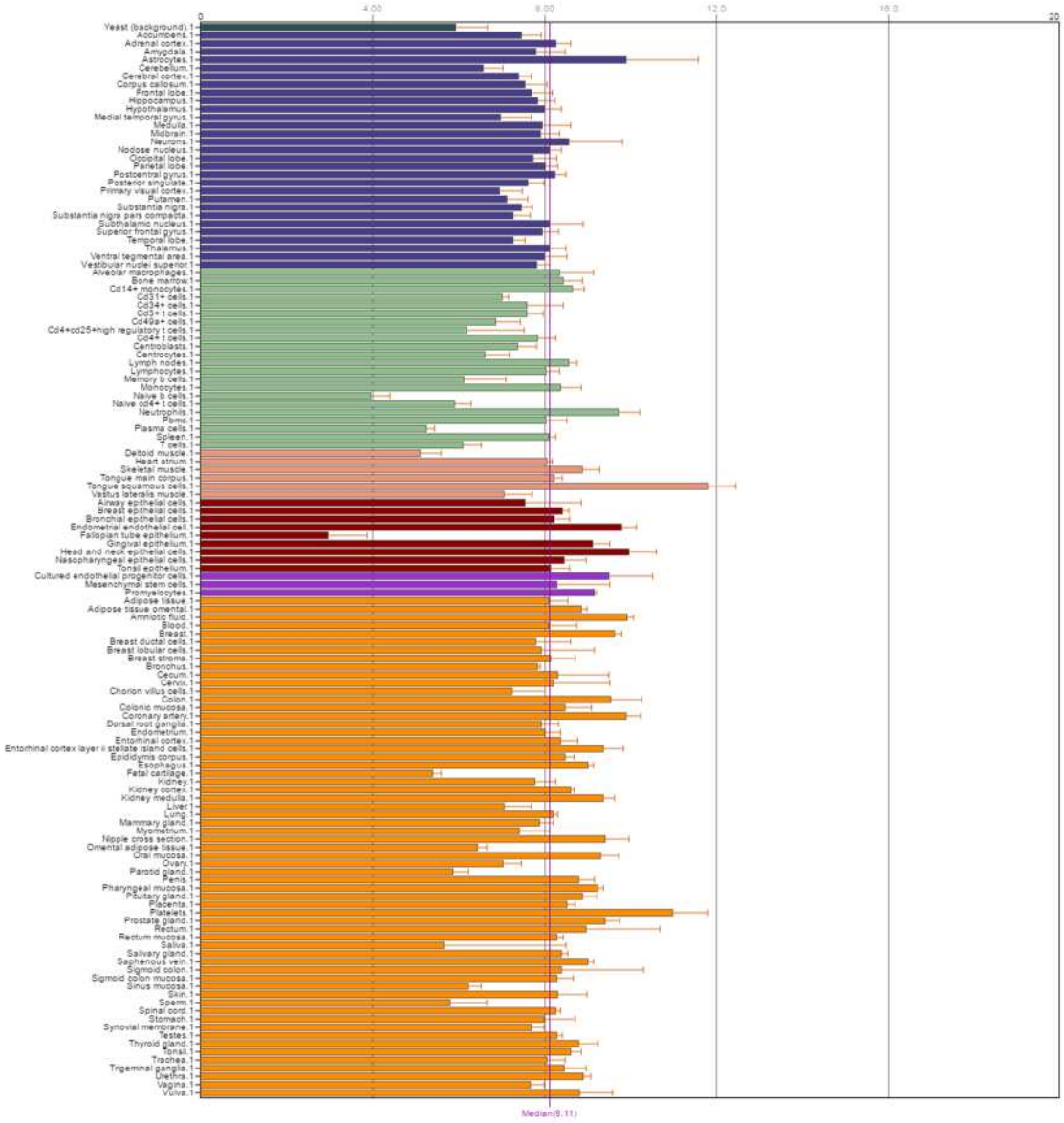


Figure 14: Data set (rab1b) mitochondrial protein of interest of this study, displays a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray. A z-score >5 suggests that the gene is expressed in that tissue. The lines drawn at the median, 3X median and 10X median are defaults of the BioGPS presentation and are not central to this analysis (Ref: http://biogps.org/#goto=genereport&id=81876).

Appendix – IV

Script for Cumulative frequency and Selection Strength

Echolocating genes of Interest – UoN Bat

Bat.R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Echolocation")

microbat.all <- read.table("mlucifugus.All.Homologs.table", header=TRUE, sep
="\t")
microbat.goi <- read.table("mlucifugus.GOI.Homologs.table", header=TRUE, sep
="\t")


microbat.all$dnds <-
microbat.all$mlucifugus_homolog_dn/microbat.all$mlucifugus_homolog_ds
microbat.goi$dnds <-
microbat.goi$mlucifugus_homolog_dn/microbat.goi$mlucifugus_homolog_ds

pdf(file="Similarity.plots.Bat.pdf")
par(mfrow=c(2,2))
hist(microbat.all$mlucifugus_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Bat)",
     breaks= 100,
     main = "")
abline(v = microbat.goi$mlucifugus_homolog_perc_id, col="blue")

hist(microbat.all$mlucifugus_homolog_dn,
     xlab = "dN (Mouse:Bat)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = microbat.goi$mlucifugus_homolog_dn, col="blue")

hist(microbat.all$mlucifugus_homolog_ds,
     xlab = "dS (Mouse:Bat)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = microbat.goi$mlucifugus_homolog_ds, col="blue")

hist(microbat.all$dnds,
     xlab = "dN/dS (Mouse:Bat)",
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = microbat.goi$dnds, col="blue")
dev.off()

summary.all <- summary(microbat.all, na.omit=TRUE)
write.table(summary.all,file="Summary.BAT.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(microbat.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.BAT.GOI.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")



all.dnds <- sort(na.omit(microbat.all$dnds))
goi.dnds <- sort(na.omit(microbat.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)
```

```
pdf(file="dNdS.cumpercent.plots.Bat.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,1.5),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Human.R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Echolocation")

human.all <- read.table("hsapiens.All.Homologs.table", header=TRUE, sep ="\t")
human.goi <- read.table("hsapiens.GOI.Homologs.table", header=TRUE, sep ="\t")


human.all$dnds <- human.all$hsapiens_homolog_dn/human.all$hsapiens_homolog_ds
human.goi$dnds <- human.goi$hsapiens_homolog_dn/human.goi$hsapiens_homolog_ds

pdf(file="Similarity.plots.Human.pdf")
par(mfrow=c(2,2))
hist(human.all$hsapiens_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Human)",
     breaks= 100,
     main = "")
abline(v = human.goi$hsapiens_homolog_perc_id, col="blue")

hist(human.all$hsapiens_homolog_dn,
     xlab = "dN (Mouse:Human)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = human.goi$hsapiens_homolog_dn, col="blue")

hist(human.all$hsapiens_homolog_ds,
     xlab = "dS (Mouse:Human)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = human.goi$hsapiens_homolog_ds, col="blue")

hist(human.all$dnds,
     xlab = "dN/dS (Mouse:Human)",
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = human.goi$dnds, col="blue")
dev.off()

summary.all <- summary(human.all, na.omit=TRUE)
write.table(summary.all,file="Summary.HUMAN.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(human.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.HUMAN.GOI.table",row.names = FALSE,
```

```
col.names = TRUE, quote = FALSE, sep="\t")



all.dnds <- sort(na.omit(human.all$dnds))
goi.dnds <- sort(na.omit(human.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)


pdf(file="dNdS.cumpercent.plots.Human.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,1.5),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Rat.R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Echolocation")

Rat.all <- read.table("rnorvegicus.All.Homologs.table", header=TRUE, sep ="\t")
Rat.goi <- read.table("rnorvegicus.GOI.Homologs.table", header=TRUE, sep ="\t")


Rat.all$dnds <- Rat.all$rnorvegicus_homolog_dn/Rat.all$rnorvegicus_homolog_ds
Rat.goi$dnds <- Rat.goi$rnorvegicus_homolog_dn/Rat.goi$rnorvegicus_homolog_ds

Rat.all[Rat.all == "Inf"] <- NA

pdf(file="Similarity.plots.Rat.pdf")
par(mfrow=c(2,2))
hist(Rat.all$rnorvegicus_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Rat)",
     breaks= 100,
     main = "")
abline(v = Rat.goi$rnorvegicus_homolog_perc_id, col="blue")

hist(Rat.all$rnorvegicus_homolog_dn,
     xlab = "dN (Mouse:Rat)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = Rat.goi$rnorvegicus_homolog_dn, col="blue")

hist(Rat.all$rnorvegicus_homolog_ds,
     xlab = "dS (Mouse:Rat)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = Rat.goi$rnorvegicus_homolog_ds, col="blue")

hist(Rat.all$dnds,
     xlab = "dN/dS (Mouse:Rat)",
```

```
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = Rat.goi$dnds, col="blue")
dev.off()

summary.all <- summary(Rat.all, na.omit=TRUE)
write.table(summary.all,file="Summary.RAT.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(Rat.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.RAT.GOI.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")



all.dnds <- as.vector(sort(na.omit(Rat.all$dnds)))


goi.dnds <- sort(na.omit(Rat.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)


pdf(file="dNdS.cumpercent.plots.Rat.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,8),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Mitochondrial Proteins of Interest – UoN Bat

Mitochondria Bat.R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Mitochondria")

microbat.all <- read.table("mlucifugus.All.Homologs.table", header=TRUE, sep
="\t")
microbat.goi <- read.table("mlucifugus.GOI.Homologs.table", header=TRUE, sep
="\t")


microbat.all$dnds <-
microbat.all$mlucifugus_homolog_dn/microbat.all$mlucifugus_homolog_ds
microbat.goi$dnds <-
microbat.goi$mlucifugus_homolog_dn/microbat.goi$mlucifugus_homolog_ds

pdf(file="Similarity.plots.Bat.pdf")
par(mfrow=c(2,2))
hist(microbat.all$mlucifugus_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Bat)",
     breaks= 100,
     main = "")
abline(v = microbat.goi$mlucifugus_homolog_perc_id, col="blue")

hist(microbat.all$mlucifugus_homolog_dn,
     xlab = "dN (Mouse:Bat)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = microbat.goi$mlucifugus_homolog_dn, col="blue")

hist(microbat.all$mlucifugus_homolog_ds,
     xlab = "dS (Mouse:Bat)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = microbat.goi$mlucifugus_homolog_ds, col="blue")

hist(microbat.all$dnds,
     xlab = "dN/dS (Mouse:Bat)",
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = microbat.goi$dnds, col="blue")
dev.off()

summary.all <- summary(microbat.all, na.omit=TRUE)
write.table(summary.all,file="Summary.BAT.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(microbat.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.BAT.GOI.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")



all.dnds <- sort(na.omit(microbat.all$dnds))
goi.dnds <- sort(na.omit(microbat.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)
```

```
pdf(file="dNdS.cumpercent.plots.Bat.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,1.5),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Mitochondria Human.R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Mitochondria")

human.all <- read.table("hsapiens.All.Homologs.table", header=TRUE, sep ="\t")
human.goi <- read.table("hsapiens.GOI.Homologs.table", header=TRUE, sep ="\t")


human.all$dnds <- human.all$hsapiens_homolog_dn/human.all$hsapiens_homolog_ds
human.goi$dnds <- human.goi$hsapiens_homolog_dn/human.goi$hsapiens_homolog_ds

pdf(file="Similarity.plots.Human.pdf")
par(mfrow=c(2,2))
hist(human.all$hsapiens_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Human)",
     breaks= 100,
     main = "")
abline(v = human.goi$hsapiens_homolog_perc_id, col="blue")

hist(human.all$hsapiens_homolog_dn,
     xlab = "dN (Mouse:Human)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = human.goi$hsapiens_homolog_dn, col="blue")

hist(human.all$hsapiens_homolog_ds,
     xlab = "dS (Mouse:Human)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = human.goi$hsapiens_homolog_ds, col="blue")

hist(human.all$dnds,
     xlab = "dN/dS (Mouse:Human)",
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = human.goi$dnds, col="blue")
dev.off()

summary.all <- summary(human.all, na.omit=TRUE)
write.table(summary.all,file="Summary.HUMAN.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(human.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.HUMAN.GOI.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
```

```
all.dnds <- sort(na.omit(human.all$dnds))
goi.dnds <- sort(na.omit(human.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)


pdf(file="dNdS.cumpercent.plots.Human.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,1.5),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Mitochondria Rat. R

```
setwd("/home/adac_guest/Desktop/GOI Data 26-7-16 /GOI.orthologs/Mitochondria")

Rat.all <- read.table("rnorvegicus.All.Homologs.table", header=TRUE, sep ="\t")
Rat.goi <- read.table("rnorvegicus.GOI.Homologs.table", header=TRUE, sep ="\t")



Rat.all$dnds <- Rat.all$rnorvegicus_homolog_dn/Rat.all$rnorvegicus_homolog_ds
Rat.goi$dnds <- Rat.goi$rnorvegicus_homolog_dn/Rat.goi$rnorvegicus_homolog_ds

Rat.all[Rat.all == "Inf"] <- NA


pdf(file="Similarity.plots.Rat.pdf")
par(mfrow=c(2,2))
hist(Rat.all$rnorvegicus_homolog_perc_id,
     xlab = "Percent Identity (Mouse:Rat)",
     breaks= 100,
     main = "")
abline(v = Rat.goi$rnorvegicus_homolog_perc_id, col="blue")

hist(Rat.all$rnorvegicus_homolog_dn,
     xlab = "dN (Mouse:Rat)",
     breaks= 100,
     main = "",
     xlim=c(0,2))
abline(v = Rat.goi$rnorvegicus_homolog_dn, col="blue")

hist(Rat.all$rnorvegicus_homolog_ds,
     xlab = "dS (Mouse:Rat)",
     breaks= 100,
     main = "",
     xlim=c(0,5))
abline(v = Rat.goi$rnorvegicus_homolog_ds, col="blue")

hist(Rat.all$dnds,
     xlab = "dN/dS (Mouse:Rat)",
```

```
     breaks= 1000,
     main = "",
     xlim=c(0,2))
abline(v = Rat.goi$dnds, col="blue")
dev.off()

summary.all <- summary(Rat.all, na.omit=TRUE)
write.table(summary.all,file="Summary.RAT.All.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")
summary.goi <- summary(Rat.goi, na.omit=TRUE)
write.table(summary.goi,file="Summary.RAT.GOI.table",row.names = FALSE,
col.names = TRUE, quote = FALSE, sep="\t")



all.dnds <- as.vector(sort(na.omit(Rat.all$dnds)))
goi.dnds <- sort(na.omit(Rat.goi$dnds))
goi.dnds.cumsum <- cumsum(goi.dnds)/sum(goi.dnds)


pdf(file="dNdS.cumpercent.plots.Rat.pdf")
par(mfrow=c(1,1))
plot(all.dnds,
     cumsum(all.dnds)/sum(all.dnds),
     xlim=c(0,8),
     ylab = "Cumulative Percentage",
     xlab = "dN/dS"
     )
points(goi.dnds,
       cumsum(goi.dnds)/sum(goi.dnds),
       col="red",
       pch=1
       )
legend("bottomright", c("GOI","All Genes"), col=c("red","black"), pch=c(1,1))
dev.off()
```

Appendix - V

Gene expression script- Heatmap (TPM and FPKM)

Gene Expression.R - Echolocation

```
library(biomaRt)
library(stringr)


# change below to input gene list
goi.list <- "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Echolocation/echolocation.UoN.T"
# change below to working directory where you want output
working.directory = "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Echolocation"


gene.expression.data1 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC5.abund
ances", head=TRUE, sep="\t")
gene.expression.data2 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC6.abund
ances", head=TRUE, sep="\t")
gene.expression.data3 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC7.abund
ances", head=TRUE, sep="\t")
gene.expression.data4 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC8.abund
ances", head=TRUE, sep="\t")

gene.expression.data1 <-
gene.expression.data1[!grepl("STRG",gene.expression.data1$Gene.ID),]
gene.expression.data2 <-
gene.expression.data2[!grepl("STRG",gene.expression.data2$Gene.ID),]
gene.expression.data3 <-
gene.expression.data3[!grepl("STRG",gene.expression.data3$Gene.ID),]
gene.expression.data4 <-
gene.expression.data4[!grepl("STRG",gene.expression.data4$Gene.ID),]

gene.expression.data1$replicate <- "Rep1"
gene.expression.data2$replicate <- "Rep2"
gene.expression.data3$replicate <- "Rep3"
gene.expression.data4$replicate <- "Rep4"

gene.expression.data1 <- gene.expression.data1[c(1,8,9,10)]
gene.expression.data2 <- gene.expression.data2[c(1,8,9,10)]
gene.expression.data3 <- gene.expression.data3[c(1,8,9,10)]
gene.expression.data4 <- gene.expression.data4[c(1,8,9,10)]


##################################################################
######################
```

```
# Get Data
#########################################################
#######################
setwd(working.directory)

myInterestingGenes <- read.table(goi.list,head=FALSE,sep="\t")

gene.expression.data1.goi <- gene.expression.data1[gene.expression.data1$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data1.goi <- merge(gene.expression.data1.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data2.goi <- gene.expression.data2[gene.expression.data2$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data2.goi <- merge(gene.expression.data2.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data3.goi <- gene.expression.data3[gene.expression.data3$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data3.goi <- merge(gene.expression.data3.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data4.goi <- gene.expression.data4[gene.expression.data4$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data4.goi <- merge(gene.expression.data4.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")



TPM <- cbind(gene.expression.data1.goi[c(3,5)],
             gene.expression.data2.goi[c(3,5)],
             gene.expression.data3.goi[c(3,5)],
             gene.expression.data4.goi[c(3,5)])
TPM <- TPM[c(1,2,3,5,7)]
row.names(TPM) <- TPM$V1
TPM <- TPM[c(-2)]
colnames(TPM) <- c("R_1","R_2","R_3","R_4")
write.table(TPM,file="TPM.goi.table", row.names = TRUE, col.names = TRUE, sep="\t",
quote=FALSE)


FPKM <- cbind(gene.expression.data1.goi[c(2,5)],
            gene.expression.data2.goi[c(2,5)],
            gene.expression.data3.goi[c(2,5)],
            gene.expression.data4.goi[c(2,5)])
FPKM <- FPKM[c(1,2,3,5,7)]
row.names(FPKM) <- FPKM$V1
FPKM <- FPKM[c(-2)]
colnames(FPKM) <- c("R_1","R_2","R_3","R_4")
write.table(FPKM,file="FPKM.goi.table", row.names = TRUE, col.names = TRUE,
sep="\t", quote=FALSE)


# plot heatmaps from TPM

minbreak = round(min(TPM)-0.5)
maxbreak = round(max(TPM)+0.5)
```

```
breaks=seq(minbreak, maxbreak, by=0.2)

#create colour panel with length(breaks)-1 colours
mycol <- colorpanel(n=length(breaks)-1,low="blue",mid="white",high="yellow")

# biocLite("gplots")
library("gplots")

tiff(filename = "TPM.heatmap_out.tiff",
    width = 14,
    height = 16,
    units = "cm",
    pointsize = 12,
    compression = c("none"),
    bg = "white", res = 600)

heatmap.2(as.matrix(TPM),
        col=mycol,
        breaks=breaks,
        scale="none",
        key=TRUE,
        key.title = "",
        key.xlab = "TPM",
        symkey=FALSE,
        dendrogram="both",
        density.info="none",
        trace="none",
        cexRow= 1,
        cexCol = 1,
        labRow = rownames(TPM),
        labCol = colnames(TPM),
        margins=c(6,6)
)
dev.off()


# plot heatmaps from FPKM

minbreak = round(min(FPKM)-0.5)
maxbreak = round(max(FPKM)+0.5)
breaks=seq(minbreak, maxbreak, by=0.2)

#create colour panel with length(breaks)-1 colours
mycol <- colorpanel(n=length(breaks)-1,low="blue",mid="white",high="yellow")

# biocLite("gplots")
library("gplots")

tiff(filename = "FPKM.heatmap_out.tiff",
    width = 14,
    height = 16,
    units = "cm",
    pointsize = 12,
    compression = c("none"),
    bg = "white", res = 600)

heatmap.2(as.matrix(FPKM),
```

```
        col=mycol,
        breaks=breaks,
        scale="none",
        key=TRUE,
        key.title = "",
        key.xlab = "FPKM",
        symkey=FALSE,
        dendrogram="both",
        density.info="none",
        trace="none",
        cexRow= 1,
        cexCol = 1,
        labRow = rownames(FPKM),
        labCol = colnames(FPKM),
        margins=c(6,6)
)
dev.off()
```

Gene Expression.R - Mitochondrial Proteins

```
library(biomaRt)
library(stringr)


# change below to input gene list
goi.list <- "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Mitochondria/mito.GOI.T.txt"
# change below to working directory where you want output
working.directory = "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Mitochondria/"


gene.expression.data1 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC5.abund
ances", head=TRUE, sep="\t")
gene.expression.data2 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC6.abund
ances", head=TRUE, sep="\t")
gene.expression.data3 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC7.abund
ances", head=TRUE, sep="\t")
gene.expression.data4 <- read.table("/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/UoN.Bat.Expression.analysis/expression.abundances/LC8.abund
ances", head=TRUE, sep="\t")

gene.expression.data1 <-
gene.expression.data1[!grepl("STRG",gene.expression.data1$Gene.ID),]
gene.expression.data2 <-
gene.expression.data2[!grepl("STRG",gene.expression.data2$Gene.ID),]
gene.expression.data3 <-
gene.expression.data3[!grepl("STRG",gene.expression.data3$Gene.ID),]
gene.expression.data4 <-
gene.expression.data4[!grepl("STRG",gene.expression.data4$Gene.ID),]
```

```
gene.expression.data1$replicate <- "Rep1"
gene.expression.data2$replicate <- "Rep2"
gene.expression.data3$replicate <- "Rep3"
gene.expression.data4$replicate <- "Rep4"

gene.expression.data1 <- gene.expression.data1[c(1,8,9,10)]
gene.expression.data2 <- gene.expression.data2[c(1,8,9,10)]
gene.expression.data3 <- gene.expression.data3[c(1,8,9,10)]
gene.expression.data4 <- gene.expression.data4[c(1,8,9,10)]


#########################################################
#######################
# Get Data
#########################################################
#######################
setwd(working.directory)

myInterestingGenes <- read.table(goi.list,head=FALSE,sep="\t")

gene.expression.data1.goi <- gene.expression.data1[gene.expression.data1$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data1.goi <- merge(gene.expression.data1.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data2.goi <- gene.expression.data2[gene.expression.data2$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data2.goi <- merge(gene.expression.data2.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data3.goi <- gene.expression.data3[gene.expression.data3$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data3.goi <- merge(gene.expression.data3.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")

gene.expression.data4.goi <- gene.expression.data4[gene.expression.data4$Gene.ID
%in% myInterestingGenes$V2, ]
gene.expression.data4.goi <- merge(gene.expression.data4.goi, myInterestingGenes,
by.x="Gene.ID", by.y="V2")



TPM <- cbind(gene.expression.data1.goi[c(3,5)],
             gene.expression.data2.goi[c(3,5)],
             gene.expression.data3.goi[c(3,5)],
             gene.expression.data4.goi[c(3,5)])
TPM <- TPM[c(1,2,3,5,7)]
row.names(TPM) <- TPM$V1
TPM <- TPM[c(-2)]
colnames(TPM) <- c("R_1","R_2","R_3","R_4")
write.table(TPM,file="TPM.goi.table", row.names = TRUE, col.names = TRUE, sep="\t",
quote=FALSE)


FPKM <- cbind(gene.expression.data1.goi[c(2,5)],
           gene.expression.data2.goi[c(2,5)],
```

```
          gene.expression.data3.goi[c(2,5)],
          gene.expression.data4.goi[c(2,5)])
FPKM <- FPKM[c(1,2,3,5,7)]
row.names(FPKM) <- FPKM$V1
FPKM <- FPKM[c(-2)]
colnames(FPKM) <- c("R_1","R_2","R_3","R_4")
write.table(FPKM,file="FPKM.goi.table", row.names = TRUE, col.names = TRUE,
sep="\t", quote=FALSE)


# plot heatmaps from TPM

minbreak = round(min(TPM)-0.5)
maxbreak = round(max(TPM)+0.5)
breaks=seq(minbreak, maxbreak, by=0.2)

#create colour panel with length(breaks)-1 colours
mycol <- colorpanel(n=length(breaks)-1,low="blue",mid="white",high="yellow")

# biocLite("gplots")
library("gplots")

tiff(filename = "TPM.heatmap_out.tiff",
    width = 14,
    height = 16,
    units = "cm",
    pointsize = 12,
    compression = c("none"),
    bg = "white", res = 600)

heatmap.2(as.matrix(TPM),
        col=mycol,
        breaks=breaks,
        scale="none",
        key=TRUE,
        key.title = "",
        key.xlab = "TPM",
        symkey=FALSE,
        dendrogram="both",
        density.info="none",
        trace="none",
        cexRow= 1,
        cexCol = 1,
        labRow = rownames(TPM),
        labCol = colnames(TPM),
        margins=c(6,6)
)
dev.off()


# plot heatmaps from FPKM

minbreak = round(min(FPKM)-0.5)
maxbreak = round(max(FPKM)+0.5)
breaks=seq(minbreak, maxbreak, by=0.2)

#create colour panel with length(breaks)-1 colours
```

```
mycol <- colorpanel(n=length(breaks)-1,low="blue",mid="white",high="yellow")

# biocLite("gplots")
library("gplots")

tiff(filename = "FPKM.heatmap_out.tiff",
    width = 14,
    height = 16,
    units = "cm",
    pointsize = 12,
    compression = c("none"),
    bg = "white", res = 600)

heatmap.2(as.matrix(FPKM),
        col=mycol,
        breaks=breaks,
        scale="none",
        key=TRUE,
        key.title = "",
        key.xlab = "FPKM",
        symkey=FALSE,
        dendrogram="both",
        density.info="none",
        trace="none",
        cexRow= 1,
        cexCol = 1,
        labRow = rownames(FPKM),
        labCol = colnames(FPKM),
        margins=c(6,6)
)
dev.off()
```

Appendix – VI

NIPA script - Bioconductor (Gene Ontology and KEGG pathway)

```
# Need to devleop a robust set of tools for analyis of gene lists.

#########################################################################
########################
# change org.XX.eg.db below to species of interest. e.g XX = Mm for mouse
source("http://www.bioconductor.org/biocLite.R")
biocLite("biomaRt")
biocLite("GOstats")
biocLite("org.Mm.eg.db")
biocLite("ReactomePA")
biocLite("gage")
biocLite("gageData")
biocLite("ggplot2")

library(org.Mm.eg.db)
library(GOstats)
library(biomaRt)
library(pathview)
library(gage)
library(gageData)
library(ReactomePA)
library(ggplot2)

# change below to input gene list
goi.list <- "echo.location.txt"
# change below to working directory where you want output
working.directory = "/home/adac_guest/Desktop/GOI Data 26-7-16
/GOI.orthologs/Echolocation"


# change below to species of interest. e.g hsapiens_gene_ensembl,
mmusculus_gene_ensembl susscrofa_gene_ensembl
ensembl.spp <- "mmusculus_gene_ensembl"
species.ens.code = "Mm"
species.kegg.code = "mmu"
kegg.data.code = "mm"
reactome.spp = "mouse" #one of "human", "rat", "mouse", "celegans", "yeast",
"zebrafish", "fly".

# change "mm" below to match kegg code e.g kegg.sets.hs for human
data(kegg.sets.mm)
data(sigmet.idx.mm)
kegsets.test <- kegg.sets.mm
kegg.sets.mm = kegg.sets.mm[sigmet.idx.mm]

# set variables for hypergeometric cutoff enrichment qval less than this and with greater
or equal to minimum number of genes in pathway or GO term will be drawn
kegg.qval.cutoff = 0.1
min.genes.cutoff = 1
```

```
# change below to determine which test to conduct.
doGO = "yes" # yes or no.       Run GoStats hypergeometric test to find enriched GO
terms in BP, MF and CC category
doReactome = "yes" # yes or no. Run ReactomePA to find enriched pathways in
Reactomedb -- BIT SLOWER
doKEGG = "yes" # yes or no.     Run hypergeometric test to find and plot enriched KEGG
pathways and visualise using PathView


#####################################################################
#########################
# Dont alter below this line
#####################################################################
#########################


#####################################################################
#########################
# Get Data
#####################################################################
#########################
setwd(working.directory)

myInterestingGenes <- scan(goi.list, what="character")

species.db <- paste("org",species.ens.code,"eg.db",sep=".")
ensembl = useEnsembl(biomart="ensembl", dataset=ensembl.spp)

# Get Entrez genes of EnsemblIDs

all.genes <- getBM(attributes=c('ensembl_gene_id', 'entrezgene',
'external_gene_name'), mart = ensembl)
all.genes.entrez <- na.omit(all.genes)

#####################################################################
###
# Match ensembl ids to genes of interest
goi.entrez <-unique(as.character(all.genes.entrez[all.genes.entrez$ensembl_gene_id
%in% myInterestingGenes,2]))
universe <- unique(as.character(all.genes.entrez$entrezgene))


#####################################################################
#########################
# start report and set up variables to catch failing sections.
#####################################################################
#########################
run.report = "NIPA.report.txt"
cat(c("-------------------------------------------------------","The NIPA run has
initiated: Any warnings will appear below.","-----------------------------------------------
---------"), file=run.report, append=FALSE, sep = "\n")

if (length(goi.entrez)==0 )
{
  cat(c("The run has terminated","Conversion of gene/peptide list to entrez failed", "Are
the IDs properly formatted or possibly too few IDs"),
      file=run.report, append=TRUE, sep='\n')
```

```
  stop("Run terminated, see NIPA.report.txt")
}


fail.GO.MF = 0
fail.GO.BP = 0
fail.GO.CC = 0
fail.reactome = 0
fail.KEGG = 0

#########################################################################
#######################
#
# part 1 GO analysis
#
#########################################################################
#######################
#########################################################################
#######################
#########################################################################
###

if (doGO == "yes")
{
  # Biological Process
  params.BP <- new('GOHyperGParams',
                   geneIds=goi.entrez,
                   universeGeneIds=universe,
                   ontology='BP',
                   pvalueCutoff=0.001,
                   conditional=F,
                   testDirection='over',
                   annotation=species.db
  )
  hgOver.BP <- hyperGTest(params.BP)
  result.BP <- summary(hgOver.BP)

  if (nrow(result.BP)==0 )
  {
    fail.GO.BP = 1
    cat(c("GO Biological process search identified no enriched terms","Probably too few
IDs"),
        file=run.report, append=TRUE, sep='\n')
  }
  if (fail.GO.BP !=1)
  {
    top.result.BP <- head(result.BP,10)
    top.result.BP$Term <- as.factor(top.result.BP$Term)
    top.result.BP$Term <- factor(top.result.BP$Term, levels = top.result.BP$Term)

    # draw if Count >= min.genes.cutoff
    top.result.BP <- top.result.BP[top.result.BP$Count >= min.genes.cutoff,]

    if (nrow(top.result.BP) > 0)
    {
      max.y.plot = 1.2*(max(-log10(top.result.BP$Pvalue)))
      sig.BP.plot <-
```

```
      ggplot(data = top.result.BP,
            aes(x = as.factor(Term), y = -log10(top.result.BP$Pvalue),
               colour = Count,
               scale_colour_gradient(low="blue"),
               size = Count))+
      geom_point() +
      scale_color_continuous("GOI count")+
      scale_size_continuous(range = c(5,20), guide=FALSE)+
      geom_hline(yintercept=1.30103,lty=2, color="grey") + # equivalent of p = 0.05
      geom_hline(yintercept=2,lty=4, color="grey") + # equivalent of p = 0.01
      geom_hline(yintercept=3,lty=3, color="grey") + # equivalent of p = 0.001
      coord_flip()+
      geom_point(stat = "identity") +
      theme_bw() +
      theme(axis.text.x = element_text(colour = "black"),
           panel.grid.major = element_blank(),
           panel.grid.minor = element_blank(),
           panel.background = element_rect(fill = "white")) +
      ylim(-0.5,max.y.plot)+
      xlab("") +
      ylab("Enrichment (-log10 pvalue)")

     pdf("GO.BP.Significant.enrichment.plot.pdf")
     print(sig.BP.plot)
     dev.off()
    }

    write.table(result.BP, file="GO.BP.table", row.names = FALSE, col.names=TRUE,sep
= '\t', quote=FALSE)
  }



  # Molecular Function
  params.MF <- new('GOHyperGParams',
               geneIds=goi.entrez,
               universeGeneIds=universe,
               ontology='MF',
               pvalueCutoff=0.001,
               conditional=F,
               testDirection='over',
               annotation=species.db
  )
  hgOver.MF <- hyperGTest(params.MF)
  result.MF <- summary(hgOver.MF)

  if (nrow(result.MF)==0 )
  {
    fail.GO.MF = 1
    cat(c("GO Molecular Function search identified no enriched terms","Probably too few
IDs"),
       file=run.report, append=TRUE, sep='\n')
  }
  if (fail.GO.MF !=1)
  {
    top.result.MF <- head(result.MF,10)
    top.result.MF$Term <- as.factor(top.result.MF$Term)
```

```
  top.result.MF$Term <- factor(top.result.MF$Term, levels = top.result.MF$Term)

  # draw if Count >= min.genes.cutoff
  top.result.MF <- top.result.MF[top.result.MF$Count >= min.genes.cutoff,]

  if (nrow(top.result.MF) > 0)
  {
   max.y.plot = 1.2*(max(-log10(top.result.MF$Pvalue)))
   sig.MF.plot <-
    ggplot(data = top.result.MF,
          aes(x = as.factor(Term), y = -log10(top.result.MF$Pvalue),
             colour = Count,
             scale_colour_gradient(low="blue"),
             size = Count))+
    geom_point() +
    scale_color_continuous("GOI count")+
    scale_size_continuous(range = c(5,20), guide=FALSE)+
    geom_hline(yintercept=1.30103,lty=2, color="grey") + # equivalent of p = 0.05
    geom_hline(yintercept=2,lty=4, color="grey") + # equivalent of p = 0.01
    geom_hline(yintercept=3,lty=3, color="grey") + # equivalent of p = 0.001
    coord_flip()+
    geom_point(stat = "identity") +
    theme_bw() +
    theme(axis.text.x = element_text(colour = "black"),
         panel.grid.major = element_blank(),
         panel.grid.minor = element_blank(),
         panel.background = element_rect(fill = "white")) +
    ylim(-0.5,max.y.plot)+
    xlab("") +
    ylab("Enrichment (-log10 pvalue)")

   pdf("GO.MF.Significant.enrichment.plot.pdf")
   print(sig.MF.plot)
   dev.off()
  }
  write.table(result.MF, file="GO.MF.table", row.names = FALSE, col.names=TRUE,sep
= '\t', quote=FALSE)
 }

 # Cellular Compartment
 params.CC <- new('GOHyperGParams',
             geneIds=goi.entrez,
             universeGeneIds=universe,
             ontology='CC',
             pvalueCutoff=0.001,
             conditional=F,
             testDirection='over',
             annotation=species.db
 )
 hgOver.CC <- hyperGTest(params.CC)



 result.CC <- summary(hgOver.CC)
 if (nrow(result.CC)==0 )
 {
  fail.GO.CC = 1
```

```
   cat(c("GO Cellular location search identified no enriched terms","Probably too few
IDs"),
       file=run.report, append=TRUE, sep='\n')
  }

  if (fail.GO.CC !=1)
  {
   top.result.CC <- head(result.CC,10)
   top.result.CC$Term <- as.factor(top.result.CC$Term)
   top.result.CC$Term <- factor(top.result.CC$Term, levels = top.result.CC$Term)

   # draw if Count >= min.genes.cutoff
   top.result.CC <- top.result.CC[top.result.CC$Count >= min.genes.cutoff,]

   if (nrow(top.result.CC) > 0)
   {
    max.y.plot = 1.2*(max(-log10(top.result.CC$Pvalue)))
    sig.CC.plot <-
     ggplot(data = top.result.CC,
           aes(x = as.factor(Term), y = -log10(top.result.CC$Pvalue),
              colour = Count,
              scale_colour_gradient(low="blue"),
              size = Count))+
     geom_point() +
     scale_color_continuous("GOI count")+
     scale_size_continuous(range = c(5,20), guide=FALSE)+
     geom_hline(yintercept=1.30103,lty=2, color="grey") + # equivalent of p = 0.05
     geom_hline(yintercept=2,lty=4, color="grey") + # equivalent of p = 0.01
     geom_hline(yintercept=3,lty=3, color="grey") + # equivalent of p = 0.001
     coord_flip()+
     geom_point(stat = "identity") +
     theme_bw() +
     theme(axis.text.x = element_text(colour = "black"),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_rect(fill = "white")) +
     ylim(-0.5,max.y.plot)+
     xlab("") +
     ylab("Enrichment (-log10 pvalue)")

    pdf("GO.CC.Significant.enrichment.plot.pdf")
    print(sig.CC.plot)
    dev.off()
   }

   write.table(result.CC, file="GO.CC.table", row.names = FALSE, col.names=TRUE,sep
= '\t', quote=FALSE)
  }
}
##########################################################
########################
##########################################################
########################
#
# part 2 Pathway analysis
# ########################################################
```

```
#########################################################
#######################

if (doReactome == "yes")
{
 reactome.out <- enrichPathway(gene=goi.entrez,
                          #pvalueCutoff=0.05,
                          readable=T,
                          organism = reactome.spp,
                          pAdjustMethod = "BH",
                          qvalueCutoff = 0.01,
                          universe = universe
 )

 reactome.writeout <- (summary(reactome.out))

 if (nrow(reactome.writeout)==0 )
 {
  fail.reactome = 1
  cat(c("Reactome analysis identified no enriched pathways","Probably too few IDs"),
      file=run.report, append=TRUE, sep='\n')
 }

 if (fail.reactome==0)
 {
  write.table(reactome.writeout, file="reactome.pathway.enrichment.table", row.names
= FALSE, col.names = TRUE, quote=FALSE, sep='\t')

  reactome.dot <- dotplot <- dotplot(
   reactome.out,
   showCategory=15,
   font.size = 12
  )
  tiff(filename="reactome.pathway.enrichment.dotplot.tiff",
       width = 320,
       height = 240,
       units = "mm",
       res=800
  )
  print(reactome.dot)
  dev.off()


  tiff(filename="reactome.pathway.enrichment.enrichmap.tiff",
       width = 320,
       height = 240,
       units = "mm",
       res=800,
       type = "Xlib",
       pointsize = 12
  )
  enrichMap(reactome.out,
            layout=igraph::layout.kamada.kawai,
            vertex.label.cex = 0.8
  )
  dev.off()
 }
```

```
}

####
##########################################################
########################
##########################################################
########################
#
# part 3 KEGG analysis
#
##########################################################
########################
##########################################################
########################

if (doKEGG == "yes")
{
  pathview.goi.entrez <- rep.int(1, length(goi.entrez))

  names(pathview.goi.entrez) = goi.entrez

  keggres = gage(pathview.goi.entrez, gsets=kegsets.test, same.dir=TRUE) # determine
kegg membership of all genes.

  keggres.pathways <- as.data.frame(keggres)
  keggres.pathways.out <- keggres.pathways[keggres.pathways$greater.set.size > 0, ]

  if (nrow(keggres.pathways.out) ==0)
  {
    fail.KEGG = 1
    cat(c("KEGG analysis identified no enriched pathways","Probably too few IDs"),
        file=run.report, append=TRUE, sep='\n')
  }

  if (fail.KEGG ==0)
  {
    keggres.pathways.out$KEGGpathways <- rownames(keggres.pathways.out)
    matching.kegg.sets.mm <- kegg.sets.mm[c(keggres.pathways.out$KEGGpathways)]
# named list of matched pathways
    matching.kegg.sets.mm.total.size <- lengths(matching.kegg.sets.mm, use.names =
TRUE) # named list of the number of total number of genes in matched pathway.


    matching.kegg.sets.mm.df <- as.data.frame(unlist(matching.kegg.sets.mm,
use.names = TRUE))
    matching.kegg.sets.mm.df$kegg.id <-  gsub("\\d+$", "",
rownames(matching.kegg.sets.mm.df))
    row.names(matching.kegg.sets.mm.df) <- NULL
    colnames(matching.kegg.sets.mm.df) <- c("entrez.id","kegg.id")

    # make subset of matching.kegg.sets.mm.df with just genes of interest in it
    goi.matching.kegg.sets.mm.df <-
matching.kegg.sets.mm.df[matching.kegg.sets.mm.df$entrez.id %in% goi.entrez, ]


##########################################################
```

```
#########################################################################
   # Stats details

#########################################################################
#########################################################################
   # for each pathway with > 0 goi in it, conduct a hypergeometric test using phyper
   # phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
   # x, q vector of quantiles representing the number of white balls drawn
   # without replacement from an urn which contains both black and white
   # balls.
   # m the number of white balls in the urn.
   # n the number of black balls in the urn.
   # k the number of balls drawn from the urn.
   # if
   # pop size : 5260 # total number of entrez gene in all pathways
   # sample size : 131 # total goi
   # Number of items in the pop that are classified as successes : 1998 # entrez in a
particular pathway
   # Number of items in the sample that are classified as successes : 62 # goi in a
particular pathway
   #
   # phyper(62,1998,5260-1998,131)
   # e.g pathway 100 genes 10 are in goi list of size 400 universe = 20,000
   # phyper(1,100,20000-100,400, lower.tail=FALSE) = 0.597 = probability of finding
this many or greater goi in pathway
   # phyper(80,100,20000-100,400, lower.tail=FALSE) = 4.603708e-122 = probability
of finding this many or greater goi in pathway

#########################################################################
#########################################################################

#########################################################################
#########################################################################

   universe.size = as.numeric(length(universe))
   total.goi.size = as.numeric(length(goi.entrez))


   # do for each pathway in list and generate table of pathways passing cut off after FDR
qvalue calculation
   working.pathways <- unique(matching.kegg.sets.mm.df$kegg.id)

   pathways.hypergeometric.results <- data.frame("Pathway"= character(0),"p.val"=
numeric(0),"FDR q.val"= numeric(0),"ensembl.ids"= character(0), "entrez.ids"=
numeric(0), "external.ids"= character(0))
   pathways.hypergeometric.results.sig <- data.frame("Pathway"= character(0),"p.val"=
numeric(0),"FDR q.val"= numeric(0), "goi.count"= numeric(0))

   for (i in 1:length(working.pathways))
   {
    current.pathway = working.pathways[i]
```

```
    goi.in.pathway <-
as.numeric(nrow(goi.matching.kegg.sets.mm.df[goi.matching.kegg.sets.mm.df$kegg.id
== current.pathway, ]))
    total.genes.in.pathway <-
as.numeric(nrow(matching.kegg.sets.mm.df[matching.kegg.sets.mm.df$kegg.id ==
current.pathway, ]))

    pval <- phyper(goi.in.pathway,total.genes.in.pathway,(universe.size-
total.genes.in.pathway),total.goi.size, lower.tail=FALSE)
    qval <- p.adjust(pval, method = "fdr", n = nrow(keggres.pathways.out))

    current.goi <- goi.matching.kegg.sets.mm.df[goi.matching.kegg.sets.mm.df$kegg.id
== current.pathway, ]
    current.goi <- current.goi[1]
    current.goi.entrez.ids <- as.numeric(as.character(current.goi$entrez.id))

    current.goi.ens <- all.genes.entrez[all.genes.entrez$entrezgene %in%
current.goi.entrez.ids ,]

    current.goi.ens.ids <- unique(current.goi.ens$ensembl_gene_id)
    current.goi.ext.ids <- unique(current.goi.ens$external_gene_name)


    current.goi.ens.ids <- paste(current.goi.ens.ids, collapse=", ")
    current.goi.entrez.ids <- paste(current.goi.entrez.ids, collapse=", ")
    current.goi.ext.ids <- paste(current.goi.ext.ids, collapse=", ")

    current.out <-
as.data.frame(cbind(current.pathway,pval,qval,current.goi.ens.ids,current.goi.entrez.ids,
current.goi.ext.ids))
    current.sig.out <- as.data.frame(cbind(current.pathway,pval,qval,goi.in.pathway))

    pathways.hypergeometric.results <- rbind(pathways.hypergeometric.results,
current.out)

    if (qval < kegg.qval.cutoff & goi.in.pathway >= min.genes.cutoff)
    {
      pid <- substr(current.pathway, start=1, stop=8) # get kegg ids
      pathview(gene.data=pathview.goi.entrez, pathway.id=pid,
species=species.kegg.code)
      pathways.hypergeometric.results.sig <-
rbind(pathways.hypergeometric.results.sig, current.sig.out)
    }

  }

  colnames(pathways.hypergeometric.results) <- c("Pathway","p.val","FDR
q.val","Ensembl.ids","Entrez.ids","External.ids")

  # make FDR q.val numeric and sort
  pathways.hypergeometric.results$`FDR q.val` <-
as.numeric(as.character(pathways.hypergeometric.results$`FDR q.val`))
  pathways.hypergeometric.results <-
pathways.hypergeometric.results[with(pathways.hypergeometric.results,
order(pathways.hypergeometric.results$`FDR q.val`)), ]
```

```
write.table(pathways.hypergeometric.results,"KEGG.enrichment.analysis.results.table",
row.names = FALSE, col.names = TRUE, quote = FALSE, sep ='\t')


#####################################################
   # draw plot of enriched pathways

#####################################################
   colnames(pathways.hypergeometric.results.sig) <- c("Pathway","p.val","FDR
q.val","goi.count")

   if (nrow(pathways.hypergeometric.results.sig)>0)
   {

     pathways.hypergeometric.results.sig$`FDR q.val` <-
as.numeric(as.character(pathways.hypergeometric.results.sig$`FDR q.val`))
     pathways.hypergeometric.results.sig <-
pathways.hypergeometric.results.sig[with(pathways.hypergeometric.results.sig,
order(pathways.hypergeometric.results.sig$`FDR q.val`)), ]

     pathways.hypergeometric.results.sig$goi.count <-
as.numeric(as.character(pathways.hypergeometric.results.sig$goi.count))
     pathways.hypergeometric.results.sig <-
pathways.hypergeometric.results.sig[with(pathways.hypergeometric.results.sig, order(-
pathways.hypergeometric.results.sig$`FDR q.val`)), ]

     max.y.plot = 1.2*(max(-log10(pathways.hypergeometric.results.sig$`FDR q.val`)))

     sig.kegg.plot <-
       ggplot(data = pathways.hypergeometric.results.sig,
            aes(x = as.factor(Pathway), y = -
log10(pathways.hypergeometric.results.sig$`FDR q.val`),
               colour = goi.count,
               scale_colour_gradient(low="blue"),
               size = goi.count))+
       geom_point() +
       scale_color_continuous("GOI count")+
       scale_size_continuous(range = c(5,20), guide=FALSE)+
       geom_hline(yintercept=1.30103,lty=2, color="grey") + # equivalent of p = 0.05
       geom_hline(yintercept=2,lty=4, color="grey") + # equivalent of p = 0.01
       geom_hline(yintercept=3,lty=3, color="grey") + # equivalent of p = 0.001
       coord_flip()+
       geom_point(stat = "identity") +
       theme_bw() +
       theme(axis.text.x = element_text(colour = "black"),
            panel.grid.major = element_blank(),
            panel.grid.minor = element_blank(),
            panel.background = element_rect(fill = "white")) +
       ylim(-0.5,max.y.plot)+
       xlab("") +
       ylab("Enrichment (-log10 pvalue)")

     pdf("KEGG.Significant.enrichment.plot.pdf")
     print(sig.kegg.plot)
     dev.off()
   }}
```

Enseml2Ortholgos – Echolocation

```
# Need to develop a robust set of tools for analysis of gene lists.

#########################################################################
########################
# change org.XX.eg.db below to species of interest. e.g XX = Mm for mouse
# source("http://www.bioconductor.org/biocLite.R")
# biocLite("biomaRt")
# biocLite("stringer")

library(biomaRt)
library(stringr)


# change below to input gene list
goi.list <- "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/echo.location.txt"
# change below to working directory where you want output
working.directory = "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/"


# change below to species of interest. e.g hsapiens_gene_ensembl,
mmusculus_gene_ensembl susscrofa_gene_ensembl
ensembl.spp <- "mmusculus_gene_ensembl"

# change below to species of interest. e.g Hs human, Mm, mouse
species.ens.code = "Mm"



#########################################################################
########################
# Get Data
#########################################################################
########################
setwd(working.directory)

myInterestingGenes <- scan(goi.list, what="character")

species.db <- paste("org",species.ens.code,"eg.db",sep=".")
ensembl = useEnsembl(biomart="ensembl", dataset=ensembl.spp)

all.genes <- getBM(attributes=c('ensembl_gene_id', 'external_gene_name'), mart =
ensembl)


# get possible ensembl ortholog gene names
species.list <- as.data.frame(listDatasets(ensembl))
species.list.names <- species.list$dataset

species.of.interest <- c(
            "rnorvegicus",
            "mlucifugus",
            "pvampyrus",
            "cfamiliaris",
```

```
              "hsapiens",
              "ttruncatus")

for (i in 1:length(species.of.interest))
  {
  species = species.of.interest[i]
  ortholog = paste0(species, "_homolog_ensembl_gene")
  ortholog.pid = paste0(species, "_homolog_perc_id")
  ortholog.dn = paste0(species, "_homolog_dn")
  ortholog.ds = paste0(species, "_homolog_ds")

  species.table = getBM(c("ensembl_gene_id",ortholog, ortholog.pid, ortholog.dn,
ortholog.ds),
        filters="ensembl_gene_id",values = all.genes$ensembl_gene_id,
        mart = ensembl)

  species.table$species <- species
  species.table <- merge(species.table, all.genes, by="ensembl_gene_id", all.x = TRUE)
  species.table.goi <- species.table[species.table$ensembl_gene_id %in%
myInterestingGenes, ]

  table.out1 <- paste0(species,".All.Homologs.table")
  table.out2 <- paste0(species,".GOI.Homologs.table")

  write.table(species.table, file=table.out1, row.names=FALSE, col.names = TRUE,quote
= FALSE, sep="\t")
  write.table(species.table.goi, file=table.out2, row.names=FALSE, col.names =
TRUE,quote = FALSE, sep="\t")
  }
```

Enseml2Ortholgos – Mitochondrial Proteins

```
# Need to devleop a robust set of tools for analyis of gene lists.

##########################################################################
########################
# change org.XX.eg.db below to species of interest. e.g XX = Mm for mouse
# source("http://www.bioconductor.org/biocLite.R")
# biocLite("biomaRt")
# biocLite("stringer")

library(biomaRt)
library(stringr)


# change below to input gene list
goi.list <- "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Mitochondria/mito.GOI.txt"
# change below to working directory where you want output
working.directory = "/Users/svzrde/Documents/Students/MSc
Students/Kamesh.2016/GOI.orthologs/Mitochondria"


# change below to species of interest. e.g hsapiens_gene_ensembl,
mmusculus_gene_ensembl susscrofa_gene_ensembl
ensembl.spp <- "mmusculus_gene_ensembl"

# change below to species of interest. e.g Hs human, Mm, mouse
species.ens.code = "Mm"



##########################################################################
########################
# Get Data
##########################################################################
########################
setwd(working.directory)

myInterestingGenes <- scan(goi.list, what="character")

species.db <- paste("org",species.ens.code,"eg.db",sep=".")
ensembl = useEnsembl(biomart="ensembl", dataset=ensembl.spp)

all.genes <- getBM(attributes=c('ensembl_gene_id', 'external_gene_name'), mart =
ensembl)


# get possible ensembl ortholog gene names
species.list <- as.data.frame(listDatasets(ensembl))
species.list.names <- species.list$dataset

species.of.interest <- c(
            "rnorvegicus",
            "mlucifugus",
            "pvampyrus",
            "cfamiliaris",
```

```
              "hsapiens",
              "ttruncatus")

for (i in 1:length(species.of.interest))
  {
  species = species.of.interest[i]
  ortholog = paste0(species, "_homolog_ensembl_gene")
  ortholog.pid = paste0(species, "_homolog_perc_id")
  ortholog.dn = paste0(species, "_homolog_dn")
  ortholog.ds = paste0(species, "_homolog_ds")

  species.table = getBM(c("ensembl_gene_id",ortholog, ortholog.pid, ortholog.dn,
ortholog.ds),
        filters="ensembl_gene_id",values = all.genes$ensembl_gene_id,
        mart = ensembl)

  species.table$species <- species
  species.table <- merge(species.table, all.genes, by="ensembl_gene_id", all.x = TRUE)
  species.table.goi <- species.table[species.table$ensembl_gene_id %in%
myInterestingGenes, ]

  table.out1 <- paste0(species,".All.Homologs.table")
  table.out2 <- paste0(species,".GOI.Homologs.table")

  write.table(species.table, file=table.out1, row.names=FALSE, col.names = TRUE,quote
= FALSE, sep="\t")
  write.table(species.table.goi, file=table.out2, row.names=FALSE, col.names =
TRUE,quote = FALSE, sep="\t")
  }
```

Appendix - VII

Gene expression Abundances script

BOXPLOT and Median (2x,5x and 10x TPM)

Gene Expression Abundances .R – Echolocation

Expression Abundances.R – BOX PLOT

```
library(ggplot2)
library(dplyr)
library(plyr)
library('VennDiagram')
library('gridExtra')
library('reshape2')

################################################
##### USER TO CHANGE
################################################
spp1.name <- "UoN.Bat"
home <- "/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/" # set working
directory
IDs.of.interest <- "Mouse.Human.UoN.BAT.homologous.UoNBat.IDs"

Unique.bat.IDs <-
"/media/adac_guest/KINGSTON/UoN.Bat.unique.genes/Unique.ENSbat.genes.UoN.Bat.sh
ort.name.IDs"
ENSBat.orthologs.IDs <-
"/media/adac_guest/KINGSTON/UoN.Bat.vs.Bat/Myotis_lucifugus.vs.UoN.Bat.UoN.Bat.or
thologs.SHORT.txt"


setwd(home)

# read tables of data
ids = unique(read.table(IDs.of.interest, sep='\t', header=FALSE, stringsAsFactors =
FALSE))
ids.uniq.UoN.Bat <- unique(read.table(Unique.bat.IDs, sep='\t', header=FALSE,
stringsAsFactors = FALSE))
ids.ENSBat <- unique(read.table(ENSBat.orthologs.IDs, sep='\t', header=FALSE,
stringsAsFactors = FALSE))

abundances.1 <- read.table("expression.abundances/LC5.abundances", sep='\t',
head=TRUE)
abundances.2 <- read.table("expression.abundances/LC6.abundances", sep='\t',
head=TRUE)
abundances.3 <- read.table("expression.abundances/LC7.abundances", sep='\t',
head=TRUE)
abundances.4 <- read.table("expression.abundances/LC8.abundances", sep='\t',
head=TRUE)

# remove STRG named abundances keep only named transcripts
abundances.1 <- abundances.1[-grep("^STRG", abundances.1$Gene.ID), ]
```

```
abundances.2 <- abundances.2[-grep("^STRG", abundances.2$Gene.ID), ]
abundances.3 <- abundances.3[-grep("^STRG", abundances.3$Gene.ID), ]
abundances.4 <- abundances.4[-grep("^STRG", abundances.4$Gene.ID), ]


all.abundances <- cbind(
  abundances.1$TPM,
  abundances.2$TPM,
  abundances.3$TPM,
  abundances.4$TPM
)
row.names(all.abundances) <- abundances.1$Gene.ID

all.abundances.GOI <- all.abundances[row.names(all.abundances) %in% ids$V1, ]
all.abundances.GOI.uniq <- all.abundances[row.names(all.abundances) %in%
ids.uniq.UoN.Bat$V1, ]
all.abundances.ENSBat <- all.abundances[row.names(all.abundances) %in%
ids.ENSBat$V1, ]

par(mfrow=c(1,4))
boxplot(all.abundances[,1], all.abundances.GOI[,1], all.abundances.ENSBat[,1],
all.abundances.GOI.uniq[,1],outline=FALSE, ylim=c(0,60), main="A", ylab="Expression
(TPM)")
axis(side=1,labels=c(1,2,3,4),at=c(1,2,3,4))
boxplot(all.abundances[,2], all.abundances.GOI[,2], all.abundances.ENSBat[,2],
all.abundances.GOI.uniq[,2],outline=FALSE, ylim=c(0,60), main="B")
axis(side=1,labels=c(1,2,3,4),at=c(1,2,3,4))
boxplot(all.abundances[,3], all.abundances.GOI[,3], all.abundances.ENSBat[,3],
all.abundances.GOI.uniq[,3],outline=FALSE, ylim=c(0,60), main="C")
axis(side=1,labels=c(1,2,3,4),at=c(1,2,3,4))
boxplot(all.abundances[,4], all.abundances.GOI[,4], all.abundances.ENSBat[,4],
all.abundances.GOI.uniq[,4],outline=FALSE, ylim=c(0,60), main="D")
axis(side=1,labels=c(1,2,3,4),at=c(1,2,3,4))


par(mfrow=c(1,4))
boxplot(all.abundances,outline=FALSE, ylim=c(0,60), main="All transcripts")
boxplot(all.abundances.GOI,outline=FALSE, ylim=c(0,60), main="Human+Mouse
orthologs")
boxplot(all.abundances.ENSBat,outline=FALSE, ylim=c(0,60), main="Bat orthologs")
boxplot(all.abundances.GOI.uniq,outline=FALSE, ylim=c(0,60), main="UoN Bat Unique
transcripts")


summary(all.abundances)
summary(all.abundances.GOI)
summary(all.abundances.ENSBat)
summary(all.abundances.GOI.uniq)

summary(all.abundances[,1])
summary(all.abundances.GOI[,1])
summary(all.abundances.ENSBat[,1])
summary(all.abundances.GOI.uniq[,1])

summary(all.abundances[,2])
summary(all.abundances.GOI[,2])
summary(all.abundances.ENSBat[,2])
```

```
summary(all.abundances.GOI.uniq[,2])

summary(all.abundances[,3])
summary(all.abundances.GOI[,3])
summary(all.abundances.ENSBat[,3])
summary(all.abundances.GOI.uniq[,3])

summary(all.abundances[,4])
summary(all.abundances.GOI[,4])
summary(all.abundances.ENSBat[,4])
summary(all.abundances.GOI.uniq[,4])



rep1.median <- median(abundances.1$TPM)
rep2.median <- median(abundances.2$TPM)
rep3.median <- median(abundances.3$TPM)
rep4.median <- median(abundances.4$TPM)

all.median.1 <- as.data.frame(abundances.1$TPM/rep1.median)
all.median.2 <- as.data.frame(abundances.2$TPM/rep2.median)
all.median.3 <- as.data.frame(abundances.3$TPM/rep3.median)
all.median.4 <- as.data.frame(abundances.4$TPM/rep4.median)

all.median.1$Gene.ID <- abundances.1$Gene.ID
all.median.2$Gene.ID <- abundances.2$Gene.ID
all.median.3$Gene.ID <- abundances.3$Gene.ID
all.median.4$Gene.ID <- abundances.4$Gene.ID

hist(all.median.1$`abundances.1$TPM/rep1.median`, breaks =1000,
    xlim = c(0,50), main ="")


transcripts.of.interest.1 <- all.median.1[all.median.1$Gene.ID %in% ids$V1, ]
transcripts.of.interest.2 <- all.median.2[all.median.2$Gene.ID %in% ids$V1, ]
transcripts.of.interest.3 <- all.median.3[all.median.3$Gene.ID %in% ids$V1, ]
transcripts.of.interest.4 <- all.median.4[all.median.4$Gene.ID %in% ids$V1, ]

TPMs <- as.data.frame(cbind(as.character(
                    transcripts.of.interest.1$Gene.ID),
                       transcripts.of.interest.1$`abundances.1$TPM/rep1.median`,
                       transcripts.of.interest.2$`abundances.2$TPM/rep2.median`,
                       transcripts.of.interest.3$`abundances.3$TPM/rep3.median`,
                       transcripts.of.interest.4$`abundances.4$TPM/rep4.median`))


TPMs$V2 <- as.numeric(as.character(TPMs$V2))
TPMs$V3 <- as.numeric(as.character(TPMs$V3))
TPMs$V4 <- as.numeric(as.character(TPMs$V4))
TPMs$V5 <- as.numeric(as.character(TPMs$V5))

colnames(TPMs) <- c("Trascript_ID","Rep1","Rep2","Rep3","Rep4")


summary(all.median.1$`abundances.1$TPM/rep1.median`)
summary(transcripts.of.interest.1$`abundances.1$TPM/rep1.median`)
```

```
boxplot(TPMs$Rep1,TPMs$Rep2,TPMs$Rep3,TPMs$Rep4)

ggplot()+
geom_bar(aes(Rep1), colour="blue", data=TPMs, alpha = 0.4) +
geom_density(aes(TPM), colour="red", data=abundances.1, alpha = 1.0) +
coord_cartesian(xlim = c(0, 100), ylim=c(0,0.25))
```

Expression Abundances.R Median (2x,5x and 10x TPM)

Bat Highest Expression

```
# biocLite("ggplot2")
# biocLite("dplyr")
# biocLite("plyr")
# biocLite("VennDiagram")
# biocLite("gridExtra")
# biocLite("reshape2")
# biocLite("tidyr")


library(ggplot2)
library(dplyr)
library(plyr)
library('VennDiagram')
library('gridExtra')
library('reshape2')
library(tidyr)

#####################################################
##### USER TO CHANGE
#####################################################
spp1.name <- "UoN.Bat"
home <- "/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/" # set working
directory

inparanoid.table <-
"/media/adac_guest/KINGSTON/UoN.Bat.vs.Mouse/inparanoid.UoN.Bat.vs.mus_musculu
s.table" # set parsed inparanoid table



setwd(home)

# read tables of data

abundances.1 <-
read.table("/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/expression.abun
dances/LC5.abundances", sep='\t', head=TRUE)
abundances.2 <-
read.table("/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/expression.abun
dances/LC6.abundances", sep='\t', head=TRUE)
```

```
abundances.3 <-
read.table("/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/expression.abun
dances/LC7.abundances", sep='\t', head=TRUE)
abundances.4 <-
read.table("/media/adac_guest/KINGSTON/UoN.Bat.Expression.analysis/expression.abun
dances/LC8.abundances", sep='\t', head=TRUE)

# remove STRG named abundances keep only named transcripts
abundances.1 <- abundances.1[-grep("^STRG", abundances.1$Gene.ID), ]
abundances.2 <- abundances.2[-grep("^STRG", abundances.2$Gene.ID), ]
abundances.3 <- abundances.3[-grep("^STRG", abundances.3$Gene.ID), ]
abundances.4 <- abundances.4[-grep("^STRG", abundances.4$Gene.ID), ]

TPMs <- as.data.frame(cbind(as.character(abundances.1$Gene.ID),
                      abundances.1$TPM,
                      abundances.2$TPM,
                      abundances.3$TPM,
                      abundances.4$TPM))


TPMs$V2 <- as.numeric(as.character(TPMs$V2))
TPMs$V3 <- as.numeric(as.character(TPMs$V3))
TPMs$V4 <- as.numeric(as.character(TPMs$V4))
TPMs$V5 <- as.numeric(as.character(TPMs$V5))

colnames(TPMs) <- c("Trascript_ID","Rep1","Rep2","Rep3","Rep4")

numbers <- TPMs
numbers <- numbers[-1]

TPMs$means <- rowMeans(numbers)
TPMs$stdev <- apply(numbers,1,sd)


ten.times.median <- 10*(median(as.matrix(numbers)))
five.times.median <- 5*(median(as.matrix(numbers)))
two.times.median <- 2*(median(as.matrix(numbers)))




#UoN.Ppi.T.43038.TPM <- TPMs[ which(TPMs$Transcript_ID == "UoN.Ppi.T.43038" ), ]

TPMs.rank.10 <- TPMs[ which(TPMs$means >= ten.times.median & TPMs$stdev < 50), ]
TPMs.rank.5 <- TPMs[ which(TPMs$means >= five.times.median & TPMs$stdev < 50), ]
TPMs.rank.2 <- TPMs[ which(TPMs$means >= two.times.median & TPMs$stdev < 50), ]

write.table(TPMs.rank.10, file="ten.x.median.High.expressed.transcripts.table",
quote=FALSE, col.names = TRUE, row.names = FALSE)
write.table(TPMs.rank.5, file="five.x.median.High.expressed.transcripts.table",
quote=FALSE, col.names = TRUE, row.names = FALSE)
write.table(TPMs.rank.2, file="two.x.median.High.expressed.transcripts.table",
quote=FALSE, col.names = TRUE, row.names = FALSE)


# match highly expressed genes to mouse orthologues.
# read inparanoid table
```

```
inparanoid.table.in <- read.table(inparanoid.table, sep='\t', header=TRUE,
stringsAsFactors = FALSE)

inparanoid.table.in <- as.data.frame(cbind(inparanoid.table.in$Representative_ID,
inparanoid.table.in$Homologs_Spp2))

inparanoid.table.in.long <-  inparanoid.table.in %>%
  mutate(V2=strsplit(as.character(V2), ",")) %>%
  unnest(V2)

inparanoid.table.in.long <- separate(inparanoid.table.in.long, V2, sep = "_",
into=c("A","B"))
inparanoid.table.in.long <- separate(inparanoid.table.in.long, V1, sep = "[.]",
into=c("C","D"))

inparanoid.table.in.long <- inparanoid.table.in.long[c(-2,-4)]
inparanoid.table.in.long <- unique(inparanoid.table.in.long)

inparanoid.table.in.long.TPMs.rank.10 <-
inparanoid.table.in.long[inparanoid.table.in.long$A %in% TPMs.rank.10$Trascript_ID, ]
inparanoid.table.in.long.TPMs.rank.5 <-
inparanoid.table.in.long[inparanoid.table.in.long$A %in% TPMs.rank.5$Trascript_ID, ]
inparanoid.table.in.long.TPMs.rank.2 <-
inparanoid.table.in.long[inparanoid.table.in.long$A %in% TPMs.rank.2$Trascript_ID, ]

mouseP.median10 <- unique(inparanoid.table.in.long.TPMs.rank.10$C)
mouseP.median5 <- unique(inparanoid.table.in.long.TPMs.rank.5$C)
mouseP.median2 <- unique(inparanoid.table.in.long.TPMs.rank.2$C)


write.table(mouseP.median10, file="mouseP.median10.txt", quote=FALSE, col.names =
FALSE, row.names = FALSE)
write.table(mouseP.median5, file="mouseP.median5.txt", quote=FALSE, col.names =
FALSE, row.names = FALSE)
write.table(mouseP.median2, file="mouseP.median2.txt", quote=FALSE, col.names =
FALSE, row.names = FALSE)
```