International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Identification and Analysis of GeneExpression in Flies Brain: An Insilico Approach

Arpita Parakh[1], Naveen Dhingra[2]

[1]Institute of Biological Science, SAGE University Indore, 452020, M.P., India
Email: *arpita.anchliya7[at]gmail.com*

[2]Institute of Biological Science, SAGE University Indore, 452020, M.P., India
Email: *navlifescience[at]gmail.com*

Abstract: ***Background: Recent advances in unscrambling the epigenetic and genetic keystones of tauopathies have occurred but less development is seen in treatments and effects after treatments. It is a challenge to prognosticate enormously complex area of tauopathy and neurodegeneration by finding out the mystery behind the differential expression of genes and its network during the process of neurodegeneration. Objective: The functional development of neurons is based on regulation of genetic blueprint of brain and is a matter of extensive research to understand the role and differential expression of genes and regulators to identify the novel targets and biomarkers responsible for identification of neurodegenerative disorders occurring due to DNA Damage, aging and tauopathy. Hence highthroughput genome data is exploited to predict the expression and function of genes involved during the process of neurodegeneration in TAU mutated flies brain. Methodology: Here we worked on chain of software to find out the hidden genomic mystery of TAU mutated flies against the control flies by analyzing their next generation sequencing data and identifying important landmarks in genes which express differentially in TAU mutant flies. The study shows differential expression of many genes involved in epigenetic regulation which can be used as prognostic biomarker in neurodegenrative disorders and tauopathies including Alzheimer disease (AD). Results: Our study implicates the role of heterochromatic condensation leads to apoptotic action in tauopathologies. The genes which are involved in notch signaling, autophagy and protein serine/threonine kinase pathway are found to be differentially expressed which gives a clue of their causal role in disrupting cell cycle, DNA damage, heterochromatin condensation, and apoptosis indicating the dreadful future of neurotoxicity in an individual. Conclusion: We focused specifically on pathways, and the genes which are involved in regulation of these pathways, and also found thata large class of highly conserved, small, noncoding RNAs that are responsible for RNA interference and found out their different expression in TAU mediated flies.***

Keywords: Neurodegeneration, tauopathy, biomarker, differential Expression, biological pathways

1. Introduction

Neurodegenerative disorders are the most awful and age dependent with the rare medication although being most studied area[22].Alzheimer's, Parkinson's, Huntington's disease, amyotrophic lateral sclerosis, frontotemporal dementia[1] and the spinocerebellar ataxias and at times loss of memory, inability to move, speak and breath, cognitive impairment are few examples of neurodegeneration [2,3,4].Drosophila models are the most obstinate and economic in understanding and solving the significant glitches in modern medicine, the neurotoxicity and tauopathies [5]. Many techniques to identify novel targets and therapeutics can be deployed on flies'model which provides new insights in disease mechanism [6]. Biological homology allows us to successfully understand aspects of neurodegeneration in humans [7].

The important element i.e. Tau protein aggregation is involved in number of neurodegenerative disorders including Alzheimer's disease. Stress related gene, immune cells, and communication factors also involved in neuronal death as they are directly related to TAU aggregation and ageing too [8]. A study documented that TAU aggregation promotes increased PIWIL1 protein and cell cycle reentry and finally apoptosis in Alzheimer's [9]. Dias-Santagata et al, also documented oxidative stress as causal factor in tau-induced neurodegeneration [10]. HP1 involved in positive regulation of euchromatin and negative regulator of heterochromatin also give a hint of piRNA mediated epigenetic regulation [11]. Recent studies suggest the impact of epigenetic regulation in tauopathies and neurodegenerative disorders including the role of transposons and depletion of a class of small RNA which silence the transposons i.e. piRNA. [12]. Many Alzheimer's related genes has been documented in recent studies [6] has helped to find out the clue for prognostication. It is also studied that expression of noncoding RNA is increased in Alzheimer's [13] which leads to many uncharacterized proteins. Multiple evidences deduce role of repetitive DNA elements in causing neurodegeneration and there are many genes found to have these repeats [14]. Liu et al., also proposed the role of RNA binding protein in neurodegeneration. Steinhilb, M.L.et al suggested that a unique group of serine/threonines has a critical role in controlling TAU toxicity and cooperate to mediate neurodegeneration in vivo [15].Fulga TA et al., showndirect interaction between TAU and actin leads to mediation of tau-induced neurotoxicity in Alzheimer's disease and related disorders [16]. The previous findings suggest the role of various factors and their correlation with heterochromatin progression and repetitive DNA containing genes shows the strong role of specific genes which are not being directly discussed in recent studies may also be involved in cell cycle, ageing and neurodegeneration. Our results may lead us to implicate the role of piRNA mediated silencing genes in neurotoxicity. As they can also act as therapeutic for neurodegenerative disorders like Alzheimer's. It is also documented that piwi-piRNA pathway can set a pace of ageing by reducing DNA damage [17]

2. Materials and Methods

2.1. Data collection and Retrieval

The SRA is NIH's primary archive of high-throughput sequencing data and is part of the International Nucleotide Sequence Database Collaboration (INSDC) that includes at the NCBI Sequence Read Archive (SRA), the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ). Data submitted to any of the three organizations are shared among them. We have downloaded the data from SRA database of National Center for Biotechnology information www.ncbi.nlm.nih.gov.in/sra. The paired end RNAseq of adult fly heads (10 days old) were mined from the database and raw reads were downloaded. 3 samples were control samples having accession number SRR7290065, SRR7290066, SRR7290067, and 3 TAU mutated samples were used for study as data is present in ncbi with the accession no SRR7290068, SRR7290069, SRR7290070. (Table 1)

2.2 Cleaning of contaminated raw reads

The galaxy software is used to create a pipeline for cleaning the raw reads obtained from the samples of SRA datasets. The cutadapt and trimmomatric function of galaxy platform were applied on the samples to remove the adapter contamination and over representative sequences. Cutadaptis used for removing the adaptor sequences and trimmomatic is used to remove contamination from the raw reads. Further the sequences were set for quality testing using FASTQC function. The quality check graphs are provided in (fig.**1**).

2.3 Mapping and Count files generation

The trimmed reads were further used as input for mapping using bowtie 2.0 against the drosophila genome and BAM files are obtained. To further understand the BAM files the results were visualized by multiqc software provided by galaxy platform (fig. **2**). The duplicate regions and unmapped regions were identified and filtered to reduce the time and complexity of machine. The 12 BAM flies generated as paired end reads were used as a input. These BAM files are further used by htseq count software to generate the no. of reads overlapped (expressed) with a particular gene i.e. it gives output in form of count files (fig. **3**). These count files were filtered and analyzed as it contains false positives also i.e. due to deep sequencing there are chances of false expression hence the data is filtered and low counts are removed. Further DEseq2 package was used to identify significant genes which are expressed in mutated fly heads.

2.4 Normalization and Differential Expression of genes

Several R packages exist for the detection of differentially expressed genes from RNA-Seq data. The analysis process includes three main steps, namely normalization, dispersion estimation and test for differential expression. Various normalization techniques applied over count files and differential expression of genes identified. The summary files obtained are mainly used to identify the significant expression of genes. The file contains various attributes like Gene identifiers, Mean normalized counts, averaged over all samples from both conditions, Fold change in log2 (logarithm base 2). The log2 fold changes are based on the primary factor level 1 vs factor level 2, hence the input order of factor levels is important. Here, DESeq2 computes fold changes of 'control' samples against 'TAU mutated' from the first factor 'Treatment', i.e. the values correspond to up- or downregulation of genes in treated samples. Standard error estimate for the log2 fold change estimate, Wald statistic, p-value for the statistical significance of this change, p-value adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls false discovery rate (FDR)

2.5 Z-score calculation and annotation

The genes obtained from normalized table are annotated from the tool "annotatemy IDs" to get the function of the significantly expressed genes. These genes are further used to check out the differentially significant expression. The Z-score gives the number of standard-deviations that a value is away from the mean of all the values in the same group. A Z-score of -2 for the gene X in sample A means that this value is 2 standard-deviations lower than the mean of the values for gene X in all the samples (A, B, C, etc). This analysis gives the significant deviations in gene expression and further the heatmap is used to visualize the data. Total 14 genes are found to be expressed differentially.

3. Results

3.1 *Drosophila* Tauopathy shows altered chromosomal mapping

Fly models provides a platform that can be used to investigate cellular communication and gene regulatory network involved in any disease state as they are found to be the best models to find out the therapeutics behind the particular ailment. To investigate a potential number of genes dysregulation due to tau expression, we began with the fly model of tauopathy i.e. involvement neuron specific mutant form of human TAU which is associated with autosomal dominant taupathy. Multiple evidences suggest the pathways involved in neurodegeneration but there is still a limitation in identifying the regulators and specific genes involved in tau phosphorylation, oxidative stress, DNA damage, constitutive heterochromatin, cell cycle and serine/ threonine associated pathways. Our results will imply the genes involved in these pathways which are differently regulated in Tauopathy and may clear a doubt that this expression and communication of genes are hallmarks or harbingers of neurodegeneration. As, these expressed factors and regulators can be used as biomarkers in future therapeutics.

We mined already publishedRNAseq raw data obtained from next generation sequencing of fly heads from SRA database of ncbi. As the RNA isolation was done at the 10^{th} day of adulthood so that the gene expressed during neuronal death can be identified and defecated easily. We identified out of 17,117 genes annotated in flybase around 4000 genes are expressed and more than 500 genes are related to cell cycle, serine/threonine pathway, piRNA biogenesis,

heterochromatin constitution and ageing being increased in TAU expressing samples of *drosophila*. Raw sequences obtained in fastq format having 3 control and 3 TAU expressed samples are being trimmed to remove the contamination of adapter used during sequencing by cutadapt tool (Galaxy Platform). The multiqc is performed over all the samples to check the quality of sequence and 35% of duplicated regions(fig. 1**).** The obtained data is mapped against the genome of drosophila melanogaster by using bowtie 2.0. The output is summarised shows no. of sequences uniquely mapped against the database (fig. **2**).The repetitions and false positives can be detected by graphic view**.** Control flies sample reads are mapped on chromosome 2L and 3R and TAU sample reads are mapped more with chr 3L, 3R and X**.** Various filters are applied over BAM files to remove the PCR duplicates region, ambiguous mapped regions, non-unique mapped regions which makes the data size small and easily accessible**.** Data statistics can be seen in supplementary (fig.**2**)**.** The mapped genes can be visualised graphically by ucsc genome browser (fig. **3**)**.** The mapped data can be deployed to get the count of RNA sequences of different samples being overlapped with a particular gene in a genome of drosophila and count files of samples are obtained from htseq count platform.

3.2 Genes expressed in Tau mutated flies involved in various pathways

From the count data the numbers of reads overlapped to relative genes are obtained. The major genes which are overlapped i.e. expressed are found to be involved in many signalling pathways. Notch signalling pathway [18, 19] involved in proliferative neurogenesis and development is inhibited by N i.e. notch, numb, adapter protein complex 2 alpha subunit andbre gene involved in neural differentiation and regulation of notch signalling shows increased number of expression in tau mutated flies. The genes like Atg and Epg5 involved in autophagy known to be self-degradation pathway [20]also found have increased expression. Cell cycle also plays a major role in ageing of neurons, their differentiation and neurogenesis. As the genes involved in cell cycle are building blocks of healthy neuron but their improper regulation may lead to various kind of lethal effects in brain. Cdc5 i.e. a cell cycle gene is down regulated in TAU mutated flies as its absence affect chromosome stability and DNA damage and it is upregulated in control flies which helps in maintaining chromosomal aberrations and DNA damage repair. CDC27 and shattered induces metastasis and cancer are found to be upregulated in TAU mutant flies.JAK/stat pathway a communication signalling pathway of various proteins and chain of interactions involved in immunity, cell division, cell death and tumor formation. JAK/stat pathway can also activate mTOR/AKT pathway [21]**.**Slbo and apt is also gene which plays a intermediate role in JAK/STAT pathway. Alg-2 is upregulated,Involved in positive regulation of JNK cascade.Orthologous to human PDCD6 (programmed cell death 6) found to be upregulated in control flies. Diap2 death associated inhibitor of apotosis as it binds with caspase and inhibits apotosis and triggers immune response found to be upregulated in TAU mutated flies. Cytokine induced apoptosis inhibitor 1 is upregulated in control flies, Cytokine induced apoptosis inhibitor 1 found to be upregulated in TAU mutated flies. sesB, Thor, Aldh-III, whd are stress related genes upregulated during TAU mutant flies. As stress related gene expression also leads to epigenetic modifications of genes in flies [22]. Some genes which are involved in G Protein coupled recepetor are also upregulated in TAU mutated flies like 5-HT1B, which promotes the inhibition of serotonin binding. Serine/threonine pathway factors also play a major role in neurodegeneration and apotosis as the genes involved to inhibit and activate the coregulators are the major players in neurodegenerative pathways. Pak, p21-activated kinase (Pak) encodes a serine/threonine effector kinase for the small GTPases, Phospholopase C, PKC delta, RACK, and IP are also upregulated in TAU mutated flies as they interact and communicate with each other in serine/threonine kinase pathway. The product of Pak contributes to growth cone guidance, synaptic development, and epithelial morphogenesis in both the ovary and embryo, tolkin, Exhibitsmetalloendopeptidase activity and Involved in several processes, including axon development; imaginal disc-derived wing vein morphogenesis; and positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway. apontic, apontic (apt) encodes a DNA-binding protein that acts as a transcriptional regulator and can function in translational repression. It plays roles in cell migration, cell fate specification, and morphogenesis via STAT signaling pathway inhibition. Many genes which are involved in RNA interference and immunogenic response are found to be expressed in TAU mutated fly heads like ago1, ago2, ago3, vasa, piwi, ago1, ago2, ago3, pld, dicer, tudor are found to be increased in number in TAU mediated flies than that of control flies. MYC, MAF1, su(tpl), apkc,DIP, HDAC and shaggy are some of the genes shows differential expression in TAU mutant flies. As taken together from the above data it can be deduced that there is strong communication between various regulators of pathway which either express or suppress during tauopathy. Various genes involved in signalling pathways like serine/threonine, cell cycle, autophagy, apoptosis, heterochromatin constitution, RNA interference, JAK/stat and G protein couple receptor interplay with Tauopathy. Genes shown differential regulation are mostly transmembrane protein coding genes and interplay strongly to produce any immune response against taupathy or activation of negative regulation in neurons development(fig. **4**).

3.3. Differential Expression identification in TAU mutated vs. control flies

One of the aims of RNAseq data analysis is the detection of differentially expressed genes. The package DESeq2 provides methods to test for differential expression analysis. In recent years, RNA sequencing (in short RNA-Seq) has become a very widely used technology to analyze the continuously changing cellular transcriptome, i.e. the set of all RNA molecules in one cell or a population of cells. One of the most common aims of RNA-Seq is the profiling of gene expression by identifying genes or molecular pathways that are differentially expressed (DE) between two or more biological conditions.

To be able to identify differential gene expression induced by TAU expression in fly heads all datasets including control flies needs to be analysed. We obtained 6 paired count files i.e. 3 for TAU mutated and 3 for control flies. These files contains many false positives as depending on the depth of the sequence there may be more number of reads mapped in addition to that gene size may also lead to large number of mapping with reads. To remove these false positives we have normalized the data so that the unwanted redundancy can be removed. The normalized data is provided in (table 2).

3.4. Identification of differentially expressed features

DESeq2 [26,27,28] is a great tool for dealing with RNA-seq data and running Differential Gene Expression (DGE) analysis. It takes read count files from different samples, as we used 3 paired count files for TAU mutated samples and 3 control samples and combines them into a big table (with genes in the rows and samples in the columns) and applies normalization for sequencing depth and library composition. Gene length normalization does not need to be accounted for because we are comparing the counts between sample groups for the same gene.Multiple factors with several levels can then be incorporated in the analysis. In our study we have a single factor that contributes to differential gene expression includes TAU mutated flies and control flies. Finally the DEseq2 run generates 3 outputs. A table with the normalized counts for each gene (rows) in the samples (columns)(table 2), a graphical summary of the results, useful to evaluate the quality of the experiment includes principal component analysis plot showing variance in the samples, heaatmap based on sample to sample matrix normalised counts, histograms of p values and MA plot shown in(fig. **5**) and a summary file with the following values for each gene(table 3) which includes Gene identifiers, Mean normalized counts, averaged over all samples from both conditions, Fold change in log2 (logarithm base 2), The log2 fold changes are based on the primary factor level includes two groups Tau mutated vs control, hence the input order of factor levels is important. Here, DESeq2 computes fold changes of 'tau mutated' samples against 'control' from the first factor 'treatment', i.e. the values correspond to up- or down regulation of genes in treated samples, Standard error estimate for the log2 fold change estimate, Wald statistic, p-value for the statistical significance of this change, p-value adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls false discovery rate (FDR). The different type of results obtained from DEseq2 contributes to identify the significant genes expressed in TAU mutated flies. A summary file gives the clear idea how significant a gene is expressed. After normalization we can compare the response of the expression of any gene to the presence of different levels of a factor in a statistically reliable way.

3.5. Extraction and annotation of differentially expressed genes

Now we have extracted the most differentially expressed gene due to treatment with fold change > 1. The DEseq2 result summary file is filtered by using simple expression on column 7 i.e. genes with significant adj. P values and 34 genes are found to be expressed significantly different. The file with the independent filtered results can be used for further downstream analysis as it excludes genes with only few read counts as these genes will not be considered as significantly differentially expressed. Hence we have selected only the genes with absolute fold change abs (log2FC)>1. And we have found 34 differentially expressed genes which are significant. The genes obtained are annotated by using tool Annotate DESeq2/DEXSeq output tables with galaxy.

3.6. Visualization of differentially expressed genes

Heatmaps of expression for these genes in the different samples are generated by extracting the normalized counts for these genes. We proceed in several steps:

Extract and plot the normalized counts for these genes for each sample with a heatmap, using the normalized count file generated by DESeq2. Compute, extract and plot the Z-score of the normalized counts (fig. **6**).

3.7. Z score calculation to compare expression

To compare the gene expression over samples, we could also use the Z-score, which is often represented.The Z-score gives the number of standard-deviations that a value is away from the mean of all the values in the same group, here the same gene. A Z-score of -2 for the gene X in sample A means that this value is 2 standard-deviations lower than the mean of the values for gene X in all the samples (A, B, C, etc).The Z-score zi,j for a gene i in a sample j given the normalized count xi,j is computed as $z_{i,j}=x_{i,j}-\bar{x_i}s_i$ with $\bar{x_i}$the mean and si the standard deviation of the normalized counts for the gene i over all samples.

To compute the Z-score, we break the process into 2 steps:
Subtract each value by the mean of values in the row (i.e. $x_{i,j}-\bar{x_i}$) using the normalized count table.

Divide the previous values by the standard deviation of values of row, using 2 tables (the normalized counts and the table computed in the previous step)

Table Compute tool with the following parameters to first subtract the mean values per row we first customized our option table.sub(table.mean(1), 0) subtracts each value by the mean of the row (computed with table.mean(1)). Than further output is used for calculation of Z score with an expression table2.div(table1.std(1),0). The table1.std(1) expression computes the standard deviations of each row on the 1st table (normalized counts) and table2.div divides the values of 2nd table (previously computed) by these standard deviations. The final Z scores can be visualized by heatmap map generated given in (fig. 6)**.** The Z scores results shows that genes like Ace involved in hydrolysis of acetylcholine neurotransmitter, slik involved in protein serine threonine kinase pathway, expression gives a signal of tumor formation [23, 24], Rab14 involved in phagolysosome and orthologus to human RAB14 family of RAS oncogene. Octopamine β2 receptor involved in G protein coupled receptor and regulates synaptic growth at nueromascular junction [25] are highly differentially expressed in TAU

mutated and control flies. Hence these differential expression are found at early stages of adulthood can act as a prognostic biomarkers for diseased condition caused due to taupathy.

4. Discussion

We found that there are various factors which triggertauopathy at early stages of adulthood in TAU mutants. Our previous results obtained from the deep analysis of RNAseq data of Tau mutated fly heads and control fly heads shown an extraordinary difference in the expression of few genes which are really less explored in various neurodegenerative disorders arised due to TAU expression. Some of the genes code for proteins which are enzymatic in nature[29], some codes for receptor molecules[30], some codes for noncoding portion or small RNA used for RNAi[31,32,33] and some of them code for factors involved in different pathways including heterochromatin condensation. Hence we can use the expression of these genes as a biomarker for diseased condition or neurodegeneration or proliferation in future. Not only differential expression but the overlap regions found in genome of drosophila after mapping with the reads obtained from RNAseq data also shown that Gpcr, cell cycle, protein serine/threonine kinase, RAS/RAF pathway, autophagy, notch signalling are some of the pathways whose disregulation of factors leads to either proliferation of neurons or degeneration of neurons and DAN damage [34, 35].Ace, Rab14, slik, fas, rost, Hrb27C, octbeta2R are some of the genes which are not much studied but are differentially expressed in tauopathy gives a strong clue that further more studies on these gene may give us a strong therapeutic on neurodegenerative. The factors are found to have varied gene expression and involved in different pathways stated above. This above study also proclaims that the huge data generated from Next generation sequencing can be down streamed after chain of operations and the proper information can be mined out to deduce a strong result. There is excess of data available online and is still left unexplored due to lack of understanding and knowledge and hence this workflow can also give an idea to analyse and understand large scale data. The functional development of neuron is based on regulation of genetic blueprint of brain [36, 37, 38]. Next generation sequencing has identified many genomic loci which leads to novel insights into neurological and neurodegenerative disorders [39] and also untangles many mysteries of epigenetic regulation which either guard or attack a gene involved in any network directly or indirectly but tau pathologies is a common link between all neurodegenerative disorder [40] lead us to some remarkable discovery of biomarkers.

5. Acknowledgement

I would like to thank acknowledge Dr. D.N. Begde and S.L. Gargh for their great support in understanding flies genomics.

6. Conflict of interest

The authors have no conflict of interest.

References

[1] Hutton, M. et al. Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 1998; 393(6686):702–705. DOI: 10.1038/31508.
[2] Abeliovich A, Gitler AD. Defects in trafficking bridge Parkinson's disease pathology and genetics. Nature. 2016 Nov 10; 539(7628):207-216. doi:10.1038/nature20414
[3] Canter RG, Penney J, Tsai LH. The road to restoring neural circuits for the treatment of Alzheimer's disease. Nature. 2016;539(7628):187-196. doi:10.1038/nature20412.
[4] Taylor JP, Brown RH Jr, Cleveland DW. Decoding ALS: from genes to mechanism. Nature. 2016;539(7628):197-206. doi: 10.1038/nature20413.
[5] Lee, V.M., Goedert, M. &Trojanowski, J.Q. Neurodegenerative tauopathies. Annu. Rev. Neurosci. 2001; 24: 1121–1159. DOI: 10.1146/annurev.neuro.24.1.1121
[6] Aileen Moloney, David B. Sattelle, David A. Lomas, Damian C. Crowther, Alzheimer's disease: insights from Drosophila melanogaster models,Trends in Biochemical Sciences,2010;35(4),228-235. https://doi.org/10.1016/j.tibs.2009.11.004.
[7] Holtzman, D.M., Mandelkow, E. &Selkoe, D.J. Alzheimer disease in 2020. ColdSpringHarb. Perspect. Med. 2012; 2(11): a011585. doi: 10.1101/cshperspect.a011585.
[8] Wyss-Coray T. Ageing, neurodegeneration and brain rejuvenation. Nature. 2016;539(7628):180-186. doi: 10.1038/nature20411
[9] Heemels MT. Neurodegenerative diseases. Nature. 2016 Nov 10;539(7628):179. doi: 10.1038/539179a.
[10] Dias-Santagata D, Fulga TA, Duttaroy A, Feany MB. Oxidative stress mediates tau-induced neurodegeneration in Drosophila. Version 2. J Clin Invest. 2007; 117(1):236-45. doi: 10.1172/JCI28769.
[11] De Lucia F, Ni JQ, Vaillant C, Sun FL. HP1 modulates the transcription of cell-cycle regulators in Drosophila melanogaster. Nucleic Acids Res. 2005;33(9):2852-8. doi: 10.1093/nar/gki584.
[12] Sun, W., Samimi, H., Gamez, M. et al. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. Nat Neurosci 21, 1038–1048 (2018). https://doi.org/10.1038/s41593-018-0194-1
[13] Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med. 2008 Jul;14(7):723-30. doi: 10.1038/nm1784.
[14] Liu EY, Cali CP, Lee EB. RNA metabolism in neurodegenerative disease. DisModel Mech. 2017;10(5):509-518. doi: 10.1242/dmm.028613.
[15] Steinhilb ML, Dias-Santagata D, Fulga TA, Felch DL, Feany MB. Tau phosphorylation sites work in concert to promote neurotoxicity in vivo. MolBiol Cell. 2007; 18(12):5060-8. doi: 10.1091/mbc.e07-04-0327

[16] Fulga TA, Elson-Schwab I, Khurana V, Steinhilb ML, Spires TL, Hyman BT, Feany MB. Abnormal bundling and accumulation of F-actin mediates tau-induced neuronal degeneration in vivo. Nat Cell Biol. 2007; 9(2):139-48. doi: 10.1038/ncb1528.

[17] Lenart P, Novak J, Bienertova-Vasku J. PIWI-piRNA pathway: Setting the pace of aging by reducing DNA damage. Mech Ageing Dev. 2018; 173:29-38. doi:10.1016/j.mad.2018.03.009.

[18] Yan B. Numb--from flies to humans. Brain Dev. 2010;32(4):293-8. doi: 10.1016/j.braindev.2009.03.008.

[19] Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science. 1999;284(5415):770-6. doi: 10.1126/science.284.5415.770. PMID: 10221902

[20] Filippi-Chiela EC, Villodre ES, Zamin LL, Lenz G. Autophagy interplay with apoptosis and cell cycle regulation in the growth inhibiting effect of resveratrol in glioma cells. *PLoS One*. 2011;6(6):e20849. doi:10.1371/journal.pone.0020849

[21] Aaronson DS, Horvath CM. A road map for those who don't know JAK-STAT. Science. 2002 May 31;296(5573):1653-5. doi: 10.1126/science.1071545.

[22] Sharma V, Kohli S, Brahmachari V. Correlation between desiccation stress response and epigenetic modifications of genes in Drosophila melanogaster: An example of environment-epigenome interaction. BiochimBiophysActa Gene Regul Mech. 2017 Oct;1860(10):1058-1068. doi: 10.1016/j.bbagrm.2017.08.001.

[23] Hipfner DR, Cohen SM. The Drosophila sterile-20 kinase slik controls cell proliferation and apoptosis during imaginal disc development. Plos Biology. 2003;1(2):E35. DOI: 10.1371/journal.pbio.0000035.

[24] Hipfner, D., Cohen, S. Connecting proliferation and apoptosis in development and disease. Nat Rev Mol Cell Biol. 2004; 5, 805–815. https://doi.org/10.1038/nrm1491

[25] VincenzinaReale, Frances Hannan, John M Midgley, Peter D Evans, The expression of a cloned Drosophila octopamine/tyramine receptor in Xenopus oocytes. Brain Research 1997; 769(2):309-320. https://doi.org/10.1016/S0006-8993(97)00723-3.

[26] Boekel J, Chilton JM, Cooke IR, et al. Multi-omic data analysis using Galaxy. Nature Biotechnology. 2015;33(2):137-139. DOI: 10.1038/nbt.3134.

[27] Love MI , Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2015; 15: 550, doi:10.1186/s13059-014-0550-8.

[28] Wagner GP , Kin K & Lynch VJ Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences. 2012;131:281–285, doi:10.1007/s12064-012-0162-3.

[29] Poorkaj, P. et al. Tau is a candidate gene for chromosome 17 frontotemporaldementia. Ann. Neurol. 1998; 43(6): 815–825. 10.1002/ana.410430617

[30] Beisel C, Imhof A, Greene J, Kremmer E, Sauer F. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature. 2002;419(6909):857-62.doi: 10.1038/nature01126.

[31] Dillon N. Heterochromatin structure and function. Biol Cell. 2004; 96(8):631-7. doi: 10.1016/j.biolcel.2004.06.003.

[32] Peng JC, Karpen GH. Heterochromatic genome stability requires regulators of histone H3K9methylation.PLoS Genet. 2009; 5(3):e1000435. doi: 10.1371/journal.pgen.1000435

[33] Lu BY, Ma J, Eissenberg JC. Developmental regulation of heterochromatin-mediated gene silencing in Drosophila. Development. 1998;125(12):2223-34.

[34] Khurana V, Lu Y, Steinhilb ML, Oldham S, Shulman JM, Feany MB. TOR-mediated cell-cycle activation causes neurodegeneration in a Drosophila tauopathy model. Curr Biol. 2006 7;16(3):230-41. 10.1016/j.cub.2005.12.042

[35] Khurana V, Merlo P, DuBoff B, Fulga TA, Sharp KA, Campbell SD, Götz J, Feany MB. A neuroprotective role for the DNA damage checkpoint in tauopathy. Aging Cell. 2012;11(2): 360–2. 10.1111/j.1474-9726.2011.00778

[36] Kruman, I.I. et al. Cell cycle activation linked to neuronal cell death initiated by DNA damage. Neuron. 2004; 41(4): 549–561. 10.1016/s0896-6273(04)00017-0

[37] Blard O, Feuillette S, Bou J, Chaumette B, Frébourg T, Campion D, Lecourtois M. Cytoskeleton proteins are modulators of mutant tau-induced neurodegeneration in Drosophila. Hum Mol Genet. 2007; 16(5):555-6610. 10.1093/hmg/ddm011

[38] Ambegaokar, S.S. & Jackson, G.R. Functional genomic screen and network analysis reveal novel modifiers of tauopathy dissociated from tau phosphorylation. Hum.Mol. Genet. 2011; 20(24): 4947–4977. 10.1093/hmg/ddr432

[39] Smith KT, Workman JL. Chromatin proteins: key responders to stress. PLoS Biol. 2012; 10(7):e1001371. doi: 10.1371/journal.pbio.1001371.

[40] Wittmann CW, Wszolek MF, Shulman JM, Salvaterra PM, Lewis J, Hutton M, Feany MB. Tauopathy in Drosophila: neurodegeneration without neurofibrillary tangles. Science. 2001;293(5530):711-4. doi: 10.1126/science.1062382.

Figures

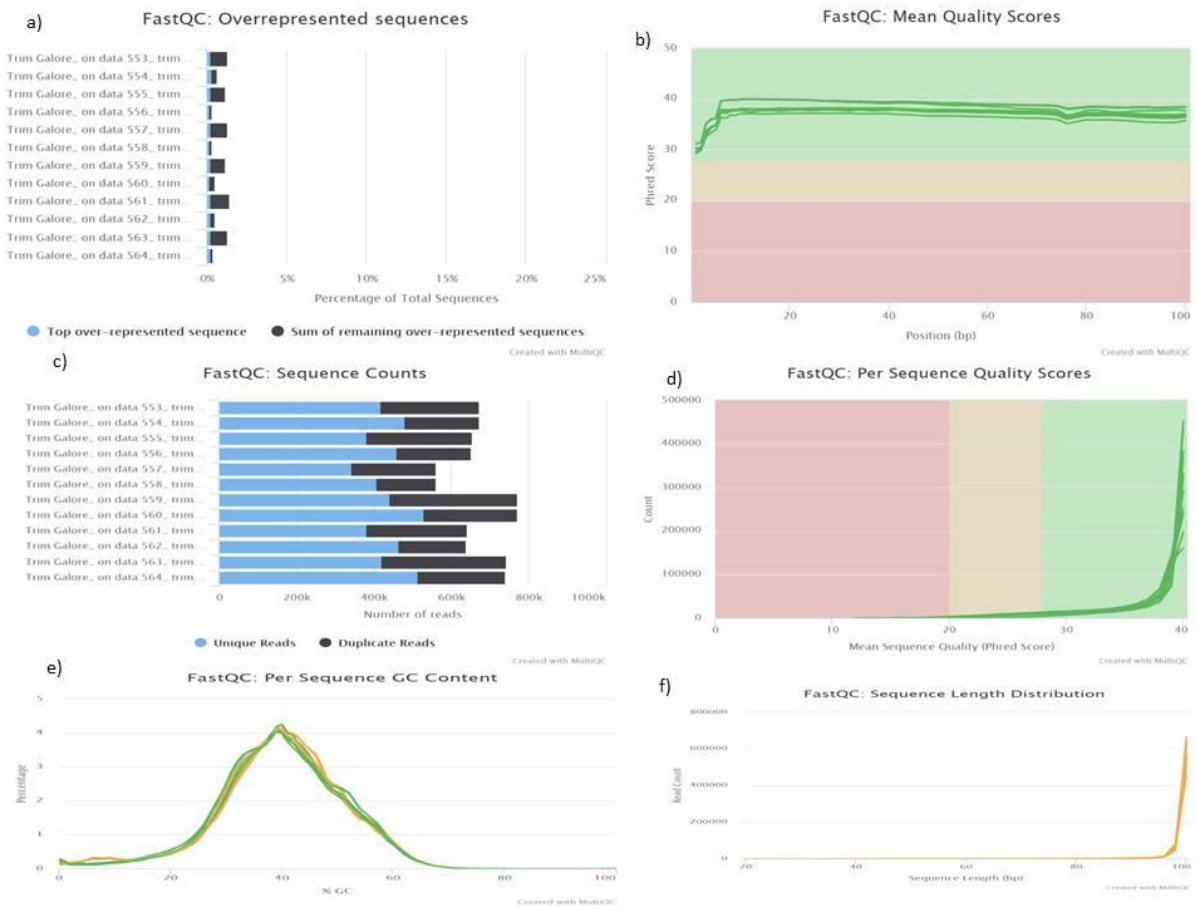

Figure 1: Showing multiqc report obtained from fastqc tool to check overall quality of sequence reads obtained from illumina Next generation RNAseq analysis. a) Showing percentage of over representative sequences. b) showing overall phred score of sequence should be arround 30 shows the best quality of sequencing reads. c) total number of unique reads and duplicate reads in each paired sample. d) graph showing per sequence quality score generated according to phred score in each sample. e) Showing overall GC content of reads in each sample f) total length of the sequence in each sample

Figure 2: Showing mapping summary of genes aligned. a) Feature counts of no. of reads assigned to gene in (blue), no. of reads left unmapped (black), no. of reads unassigned i.e. not featured (green), no. of reads ambigous (orange). b) 60-70% of

reads paired in (blue), 20% unpaired unique in (black), 5% unpaired reads duplicate (green), 20% unmapped reads (orange). C)graph shows mapping statistics i.e. reads mapped with the chromosome location.

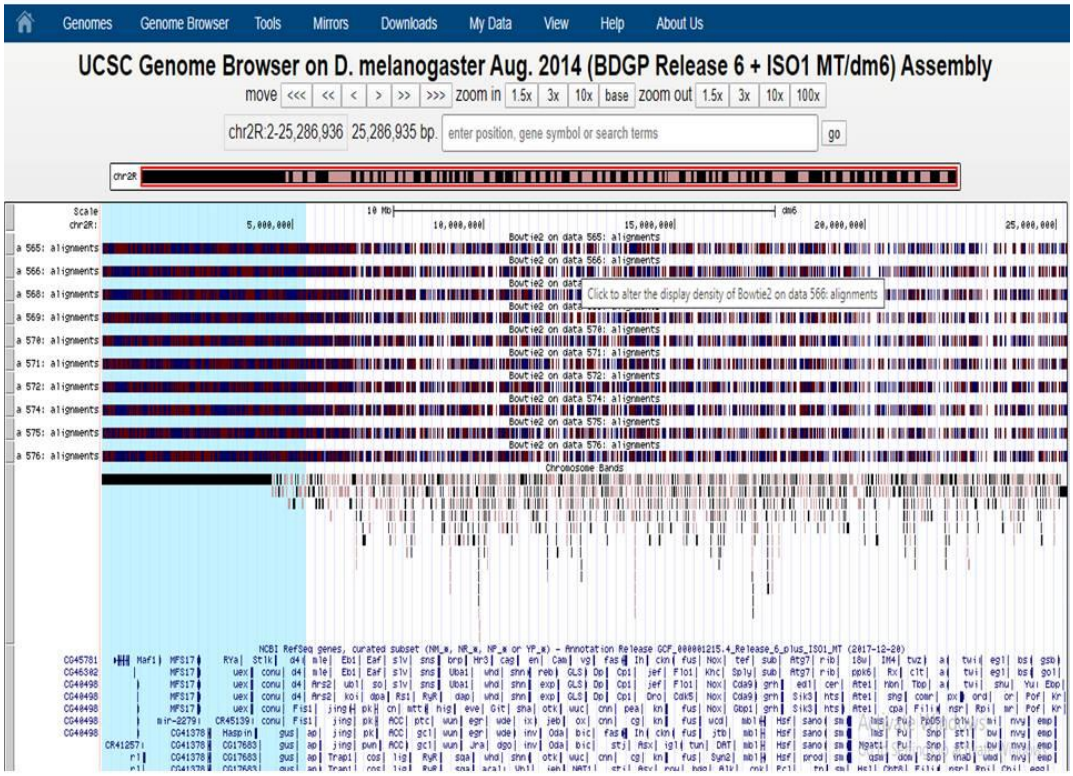

Figure 3: Visual representation of reads mapped with the drosophila genome i.e. BAM files generated from mapping the reads with genome of flies.

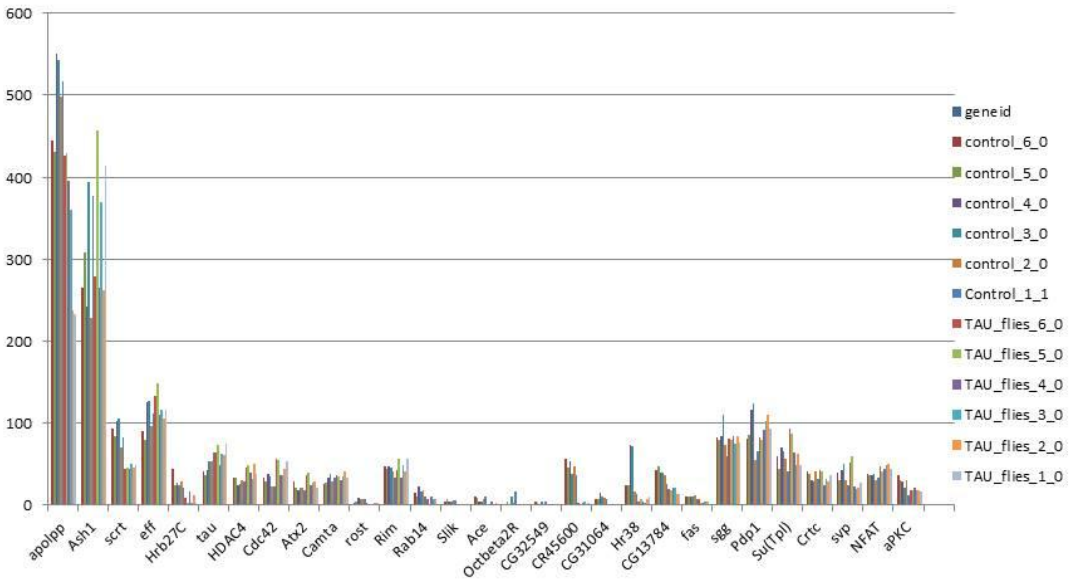

Figure 4: Showing graph of genes expressed in a particular sample. The different color line shows 12 samples. Y-axis shows the number of overlaps and x-axis shows the name of gene to which reads are overlapped.

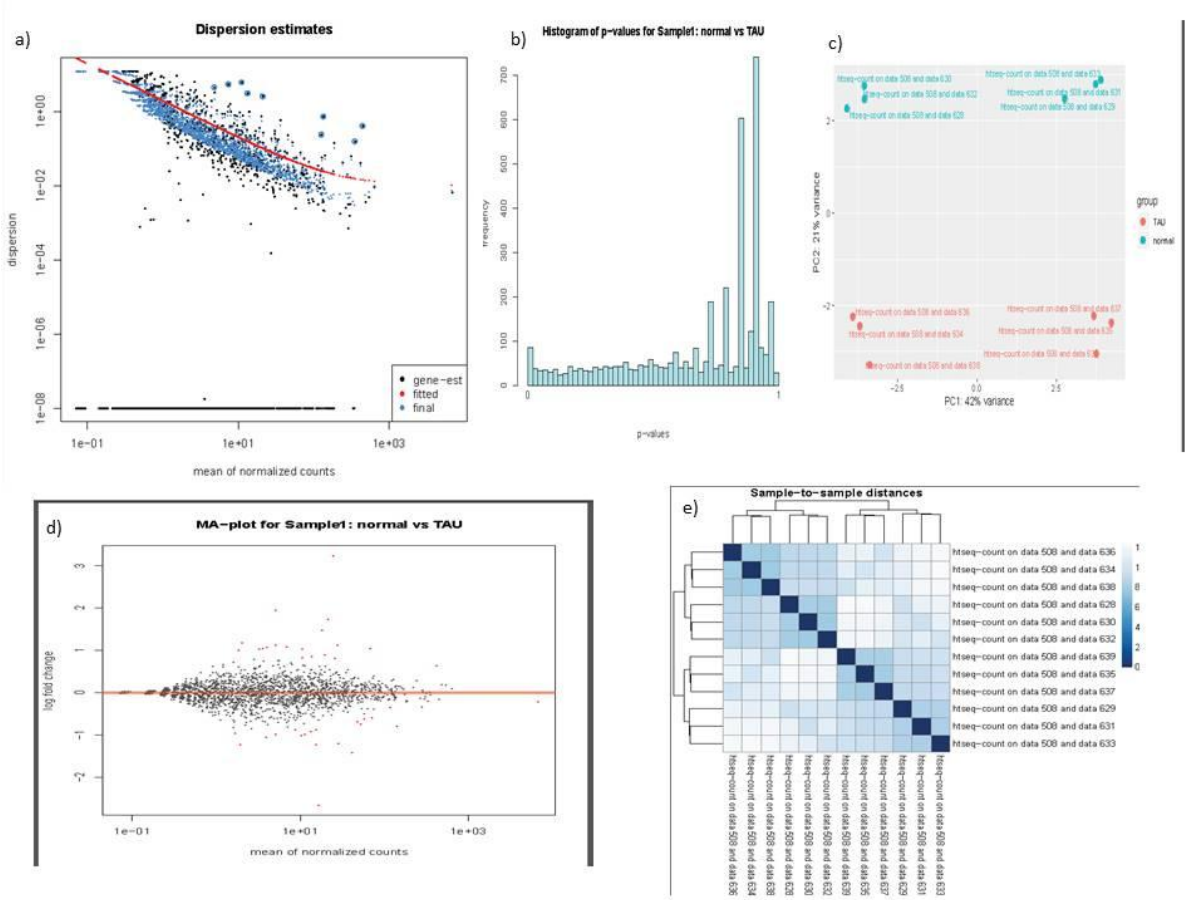

Figure 5: Differential expression of genes generated by DEseq2 package for normal vs. TAU mutated a) Dispersion estimates: gene-wise estimates (black), the fitted values (red), and the final maximum a posterior estimates used in testing (blue). This dispersion plot is typical, with the final estimates shrunk from the gene-wise estimates towards the fitted estimates. Some gene-wise estimates are flagged as outliers and not shrunk towards the fitted value. The amount of shrinkage can be more or less than seen here, depending on the sample size, the number of coefficients, the row mean and the variability of the gene-wise estimates b) next is the histogram generated from p values for comparision between two groups of factor normal vs tau. c) Principal component plot showing the variance between the genes in the different samples showing 21% variance in the expression and 42% variance in the sequencing reads. d) MA-plot generated for sample vs. TAU generated from mean of normalized counts. It displays the global view of the relationship between the expression change of conditions (Tau vs control) (log ratios, M), the average expression strength of the genes (average mean, A), and the ability of the algorithm to detect differential gene expression. The genes that passed the significance threshold (adjusted p-value < 0.1) are colored in red. e) showing the heat map gives sample to sample distance and clustering based on normalized data. The heatmap gives an overview of similarities and dissimilarities between samples: the color represents the distance between the samples. Dark blue means shorter distance, i.e. closer samples given the normalized counts.

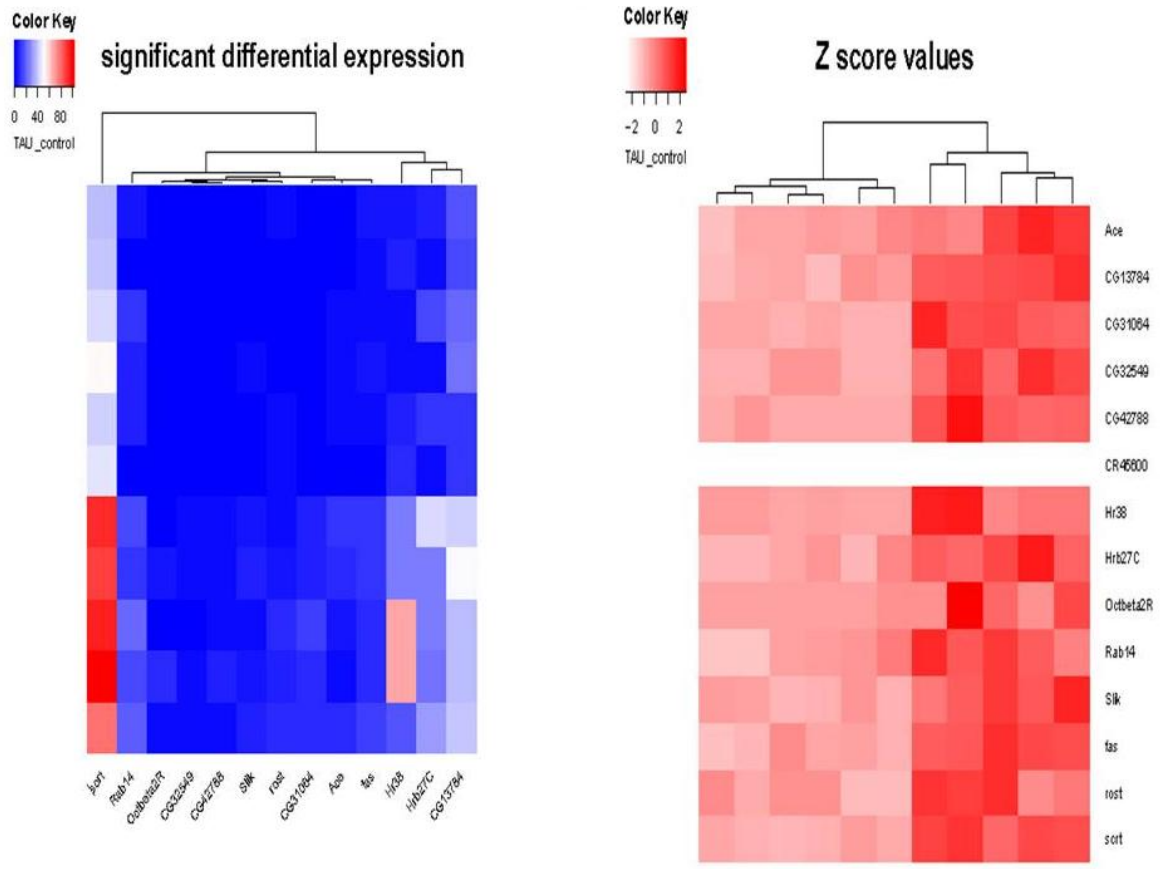

Figure 6: Heat maps generated for significant differential expression observed from the normalized count table and Z scores for most differentially expressed genes in 12 samples.

Table 1: Showing the number and type of sample and accession Id.

S.No.	Accession ID	Strain +/-
1	SRR7290065	wildType +
2	SRR7290066	wildType +
3	SRR7290067	wildType +
4	SRR7290068	TAU mutated +
5	SRR7290069	TAU mutated +
6	SRR7290070	TAU mutated +

Table 2: Showing the gene symbols having significant P value and differentially expressed in TAU mutated flies

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj
CR45600	25.08734	3.229807	0.252627	12.78489	1.99E-37	3.10E-34
CG14464	16.63633	-2.66797	0.266979	-9.99319	1.63E-23	1.27E-20
scrt	67.8131	1.036365	0.129571	7.998435	1.26E-15	6.55E-13
Oda	41.29161	-1.42081	0.215272	-6.60008	4.11E-11	1.60E-08
CG17839	28.81536	-1.2068	0.191099	-6.31508	2.70E-10	7.02E-08
CG31064	5.069166	1.942801	0.306439	6.339936	2.30E-10	7.02E-08
Hr38	21.4842	1.72906	0.293325	5.894687	3.75E-09	8.36E-07
CG13784	28.297	1.116915	0.196811	5.675048	1.39E-08	2.70E-06
Hrb27C	18.10365	1.469021	0.273213	5.376838	7.58E-08	1.31E-05
CG6966	15.61328	-1.2395	0.245329	-5.0524	4.36E-07	6.80E-05
mbl	142.2129	-0.80141	0.18579	-4.31353	1.61E-05	0.00227708
CG40228	47.66094	-0.69857	0.164459	-4.24767	2.16E-05	0.00280624
apolpp	417.7732	0.626761	0.149096	4.203753	2.63E-05	0.00314828
Msp300	19.58673	0.892728	0.214641	4.159161	3.19E-05	0.00331983
Smurf	1.931645	-1.23086	0.295255	-4.16882	3.06E-05	0.00331983
CG42788	2.023276	1.176764	0.296924	3.96318	7.40E-05	0.00678238
grh	72.39727	-0.60577	0.152579	-3.97023	7.18E-05	0.00678238
Dbp80	58.37788	-0.62287	0.159201	-3.91249	9.14E-05	0.00791194
NUCB1	10.48286	-1.02497	0.266902	-3.84026	0.0001229	0.01008477
Idgf4	15.60633	-0.88834	0.235282	-3.77564	0.0001596	0.01184802
fas	7.011308	1.07099	0.283515	3.777541	0.0001584	0.01184802
Ace	5.118111	1.117154	0.30459	3.667735	0.0002447	0.01734093

Rab14	10.71909	1.019553	0.279146	3.652396	0.0002598	0.01761024
Octbeta2R	3.282083	1.028276	0.282449	3.640569	0.000272	0.01767105
mxt	13.56733	-0.92472	0.25535	-3.62136	0.0002931	0.01827481
CG5953	11.49546	-0.9881	0.280399	-3.52389	0.0004253	0.0254993
5-HT1B	3.913425	-1.06117	0.306475	-3.4625	0.0005352	0.03028633
CG7029	58.21553	-0.49988	0.144948	-3.44866	0.0005634	0.03028633
rost	4.093228	1.063321	0.307802	3.454563	0.0005512	0.03028633
CG42750	13.93783	0.84779	0.246679	3.436817	0.0005886	0.03058728
CG32549	1.878915	1.002219	0.295782	3.388369	0.0007031	0.0353589
Slik	3.57435	1.038221	0.307293	3.3786	0.0007286	0.03549451
CR45177	1.700312	-0.96622	0.289753	-3.33463	0.0008541	0.04006239
chrb	26.3333	0.803833	0.241513	3.32832	0.0008737	0.04006239

Table 3: Showing Z score calculated to find the significant gene expressions which are down regulated and upregulated in TAU mutated flies

Ace	-0.78359	-0.76948	-0.17562	-0.6784	-0.51316	-1.2554	1.763837	1.386526	0.034219	-0.18917	1.180227
CG13784	-0.79734	-0.88726	-0.54942	-0.3726	-1.18895	-1.14967	1.050019	1.570051	0.662643	0.752705	0.909805
CG31064	-0.91105	-0.74048	-0.91105	-0.91105	-0.73735	-0.72922	0.630534	0.53156	1.805106	0.960339	1.012643
CG32549	-0.42881	-0.96953	-0.96953	-0.96953	-0.40243	-0.96953	1.54698	1.048993	0.212846	1.474382	0.426148
CG42788	-0.84049	-0.45038	-0.84049	-0.84049	-0.84049	-0.84049	0.481656	0.573522	0.816059	2.155516	0.626056
CR45600											
Hr38	-0.7136	-0.60204	-0.72817	-0.75945	-0.63296	-0.50768	0.116004	0.185612	1.900732	1.916183	-0.17462
Hrb27C	-0.72471	-1.09522	-0.14115	-1.05892	-0.49995	-1.08244	1.920423	0.583263	0.630226	0.411131	1.057348
Octbeta2R	-0.62672	-0.62672	-0.30855	-0.62672	-0.62672	-0.62672	-0.30321	1.103216	-0.32272	2.515041	0.449805
Rab14	-0.7009	-1.33417	0.015317	-0.50086	-0.5227	-1.31675	0.638982	-0.00312	1.625282	0.719286	1.379628
Slik	-1.01937	-0.63834	-0.96749	-0.49955	-1.00183	-0.58946	0.712621	1.752676	0.193415	0.652285	1.405034
fas	-0.20768	-1.0374	-0.96459	-0.65504	-0.80286	-1.22709	1.022684	0.947341	0.626919	0.704581	1.593125
rost	-0.36503	-0.88746	-1.16057	-1.16057	-0.32623	-0.28716	-0.23496	0.15933	1.448801	1.236518	1.577301
scrt	-1.0384	-0.96037	-0.8627	-0.50801	-0.93186	-0.74946	1.076037	0.91691	1.146696	1.417799	0.493353