

TISSUE SPECIFIC TRANSCRIPTOME OF ZEBRAFISH IN ACHE MUTANT EMBRYOS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
IN MOLECULAR BIOLOGY AND GENETICS

By
FATMA BETÜL DİNÇASLAN
JUNE, 2019





Abstract

Tissue Specific Transcriptome of Zebrafish in Ache Mutant Embryos

Fatma Betül Dinçaslan
M.Sc. in Molecular Biology and Genetics
Advisor: Özlen KONU KARAKAYALI
June 2019

Differential expression of specific genes in certain tissues provides information about tissue specificity which might define phenotypes of various tissues. It is possible to understand the tissue-specific effects of knockout or knockdown studies performed on zebrafish embryos, using such genes. In this study, publicly available RNA-seq datasets providing data on 15 of the tissues from 5-9 months old zebrafish were used to estimate tissue specificity for zebrafish genes. Three different normalizations (i.e SD, RPKM, TPM) of 15 tissues were performed to compare; and the results were used to understand whether a given zebrafish mutant has significant enrichments for tissue-specific genes, based on different metrics including Tau, TSI, Hg, Spm, Gini, Counts. Application of these pipelines to the publicly available acetylcholinesterase (ache) mutant vs. healthy zebrafish data (GSE74202) revealed that many retina-muscle-, and liver-specific genes were downregulated in ache mutants. The downregulation of retina and liver specific genes (such as *arr3a*, *rpe65a*, *rom1b* and *fabp10a*, respectively) were further confirmed with qPCR on comparative study of ache (+/?) and ache (-/-) 3 days post fertilization (dpf) zebrafish embryos. In addition a pilot experiment testing the effects of constant light and constant dark exposure for 3-5 dpf ache mutant and healthy embryos were performed, suggesting that the expression of retina-specific genes were more prominently affected in 3 dpf mutant embryos regardless of light.

Key Words: *Tissue Specificity, ache mutants, Differential Gene Expression, Tau metric, Zebrafish, Acetylcholinesterase, light, dark, embryos, Tsi, Hg, Spm, Gini, RPKM, Sequencing Depth, TPM*

Özet

Ache Mutasyonlu Zebrabalığı Yavrularında Dokuya Özgü Transkriptomu

Fatma Betül Dinçaslan
Moleküler Biyoloji ve Genetik Yüksek Lisans Programı
Tez Danışmanı: Özlen KONU KARAKAYALI
June 2019

Bazı genler tüm dokularda yakın miktarlarda ifade olunurken bazı genler bir veya birkaç dokuda diğerlerinden fark edilebilir miktarda yüksek olarak ifade edilmekte olup bu genlere “dokuya özgü gen” ismi verilmektedir. Dokuya özgü genler, dokuların fenotipini anlamamıza yardımcı olabilecek genlerdir. Zebrabalığı dokularına ait özgün genleri tespit ederek, susturalan/silinen genlerin işlevlerini anlamamız kolaylaşacaktır. Bu çalışmada, herkese açık olarak sunulan ve 5-9 aylık yetişkin zebrabalığına ait 15 dokunun RNA dizileme örnekleri, zebrabalığına ait dokuya özel genlerin tespiti için kullanıldı. Literatürde mevcut olan SD, RPKM ve TPM normalizasyon metotlarını uygulandıktan sonra yine literatürde varolan dokuya özgü genleri hesaplamak için kullanılan formüller, R kullanılarak uygulanarak sonuçlar birbirleri ile karşılaştırılarak dokuya ait genlerin bulunması hedeflenmiştir. Bir sonraki aşamada, bu veriseti, asetilkolinesteraz (ache) mutant 3 günlük balık (3dpf) embryosu RNAseq veriseti ile karşılaştırılarak, ache geninin rolü hakkında daha fazla bilgi edinilmesi hedeflenmiştir. Analiz sonucunda retina, kas ve karaciğere ait genlerin (örneğin *arr3a*, *rpe65a*, *rom1b* and *fabp10a*) ache geninin yokluğunda daha az ifade edildiği keşfedilmiş ve bazıları 3 ve 5 dpf balık embryolarında qRT-PCR deneyleri ile doğrulanmıştır. Buna ek olarak, ache geninin ışık ve karanlık döngüsünden nasıl etkilendiğinin anlaşılması için değişik ışık-karanlık döngüsünde yetiştirilen balıklarla pilot bir çalışma yapılmış ve bu çalışmada 3 günlük ache mutant balıkların retinaya özel genlerinin bu değişimden ışıktan bağımsız olarak daha fazla etkilendiği bulgulanmıştır.

Anahtar Kelimeler: Dokuya özel genler, ache mutant, Diferansiyel gen ifadesi, Tau metric, Zebrabalığı, Asetilkolinesteraz, ışık, karanlık, embryo, Tsi, Hg, Spm, Gini, RPKM, Dizileme Derinliği, TPM

Acknowledgements

I would like to express my gratitude to Dr. Ozlen KONU for accepting me as a graduate student to her lab and her guidance during Master's Degree. I improved myself a lot thanks to her guidance and patience.

I am very grateful to Dr. Michelle Adams and Dr. Sucularli by being my committee members, and giving valuable feedbacks to my thesis.

I am very happy that I am one of the members of KONU Lab. They have always been supportive and friendly. Especially Seniye, Ayse and Said helped me a lot learn very basics of both wet and dry labs. I deeply thank them. I would like to thank Murat, Tugberk, Damla, Busra, Ilgim, Gizem, Beril, Sena, Melike, Kubra and Rabia for their friendship, and social and academic support.

I would like to thank Afshan Nabi, a previous undergraduate intern in Konu lab for her help in coding.

I would like to thank MBG family, especially the graduate students for their friendship and scientific support, and Tulay Hanim for helping me a lot during zebrafish studies.

I feel very lucky to have my dear friends Tugce, Nazli, Busra, Saliha, Gulsum, Kubra and Irem for being always supportive for me since undergraduate studies.

My greatest gratitude goes to my precious family. They have always put their faith on me in every aspects of life.

Table of Contents

Abstract.....	iv
Özet.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Figures	xi
List of Tables	xvii
Abbreviations.....	xviii
Chapter 1: Introduction	1
1.1 Zebrafish as a model organism	1
1.1.1 Zebrafish embryo	1
1.1.2 ache mutant.....	2
1.2 RNA-Seq Analysis	4
1.2.1 Brief Introduction to RNA-Seq	4
1.3 Tissue Specificity	8
1.3.1 Tissue Specificity in mammals and zebrafish.....	8
1.3.2 Tissue Specificity on AChE Mutant Zebrafish.....	8
1.3.2.1 AChE and visual system.....	8
1.3.2.2 Tissue Specificity of Zebrafish Raised in Different Light Conditions	9
1.3.2.3. ache is highly expressed in tissues such as muscle and heart	9
1.3.2.4 Pineal gland, circadian clock in zebrafish and potential role of ache	9
1.3.2.5 Genes Selected as Specific to a Tissue on AChE Mutant Zebrafish Embryo Dataset....	10
rpe65a	10
arr3a.....	11
rom1b.....	11
desma.....	12
mylz3	12
pck1.....	13
fabp10a	13
1.3.3 Tissue Specificity Metrics Used in This Study.....	13
1.3.4 Existing tools to analyse tissue specificity of gene lists for mammals and/or zebrafish	16
Chapter 2: Aims and Rationale	18
Chapter 3: Materials and Methods.....	20
3.1 Materials	20

3.1.1 General Reagents and Equipment	20
3.1.2 Primers	20
3.1.2.1 Primer Design.....	20
3.1.3 Solutions.....	22
3.1.3.1 Zebrafish Solutions (E3 Medium).....	22
3.1.4 Chemicals	22
3.2 Methods.....	22
3.2.1 RNA-Seq data analysis.....	22
3.2.1.1 RNAseq files used.....	22
3.2.1.2 Seven Bridges Cancer Genomics Cloud (CGC).....	23
3.2.1.2.1 Quality control of fastq files.....	24
3.2.1.2.2 Sequence Alignment	24
3.2.1.2.3 Raw Count Data Retrieval	24
3.2.1.3 Normalization Methodologies	25
3.2.1.4 Determination of Tissue Specificity	26
3.2.1.5 Differential Gene Expression Analysis	27
3.2.1.6 Gene Ontology/ Enrichment Analysis	28
3.2.2 General Methods	28
3.2.2.1 Total RNA Extraction from Zebrafish	28
3.2.2.2 cDNA Synthesis	29
3.2.2.3 Real time RT-qPCR.....	29
3.2.2.4 Q-RT-PCR Expression Analysis.....	31
3.2.3 In vivo zebrafish experiments	31
3.2.3.1 Zebrafish embryo strain	31
3.2.3.2 Breeding Setups	32
3.2.4 Statistical Analysis.....	33
Chapter 4: Results	34
4.1 DEGs of Ache -/- with respect to Ache +/-?	34
4.2 Tissue Specificity Analysis	36
4.2.1 How does normalized data look like?	36
4.2.2 How are the tissues from different datasets clustered?	40
4.3 Application of Tissue Specificity Indices on Ache -/- vs. Ache +/- Dataset.....	44
4.3.1 How does the normalization method affect the distribution of tissue specificity scores for up- and down-regulated genes in ache mutants?	44
4.3.2 How does the p-value or log2fc cut offs used to determine the tissue specificity change the distribution of TS genes of AChE dataset?.....	47

4.3.2.1 Normalizations Compared	48
4.3.2.2 How similar were the genes found as tissue specific across different normalization methods? Does the log transformation have any effect on TS measurement?	61
What about log transformation?	71
4.3.2.3 qPCR Validations on Muscle and Liver and Retina.....	76
4.3.3 Pathway Enrichment using GO BP analysis in R.....	77
Brain- Specific Expression of AChE mutant zebrafish	77
Liver- Specific Expression of AChE mutant zebrafish	78
Retina- Specific Expression of AChE mutant zebrafish	79
Skin- Specific Expression of AChE mutant zebrafish	80
Muscle- Specific Expression of AChE mutant zebrafish	81
Ovary- Specific Expression of AChE mutant zebrafish	82
Testis- Specific Expression of AChE mutant zebrafish	83
Kidney- Specific Expression of AChE mutant zebrafish.....	83
Spleen- Specific Expression of AChE mutant zebrafish.....	84
Heart- Specific Expression of AChE mutant zebrafish.....	85
Intestine- Specific Expression of AChE mutant zebrafish	86
Bones- Specific Expression of AChE mutant zebrafish.....	86
Gill- Specific Expression of AChE mutant zebrafish	87
Embryos- Specific Expression of AChE mutant zebrafish.....	88
UFeggs- Specific Expression of AChE mutant zebrafish	88
4.4 The Relationship between AChE and Retinal Genes.....	89
4.3.1 qPCR results	89
Chapter 5: Discussion.....	92
5.1 Were the tissues from different datasets clustered together?	92
5.2 Did the normalization method chosen have any effect on the distribution of tissue specificity scores?	93
5.3 Should you take log transformation into account while measuring tissue specificity?	95
5.4 Did the metrics chosen have any effect on the definition of tissue specific genes?	95
5.5 Which tissues were affected most in the absence of ache activity in zebrafish embryos?.....	96
5.6 What kind of relationship was present between ache and retinal genes?	99
5.7 The Difference between This Study and Other Available Tools	100
5.8 Conclusions and Future Perspectives	101
Supplementary.....	103
Quality Control of fastq files by FatQC tool	103

How did the Tau values of transformed vs. non transformed and merged vs. original TS data distributed?	105
How similar the distributions of up and down regulated genes when different DEG cut-offs were chosen?	107
Were DEG genes enriched for TS genes?.....	110
Upset Intersection Graphs for log2 transformed vs. non-transformed normalized counts on abs (L2FC)>0.5849 DEGs	111
Upset Intersection Graphs for Different Metrics and Normalizations Combined (log2fc based DEG)	118
Upset Intersection Graphs to compare TS genes (Tau>0.8) in different normalizations for DEG cut-off p-value<0.1	126
Upset Intersection Graphs to compare TS genes (Tau>0.8) in different normalizations for DEG cut-off p-value<0.05	133
Upset Intersection Graphs for Different Metrics and Normalizations Combined (p-value<0.1 based DEG)	134
Upset Intersection Graphs for Different Metrics and Normalizations Combined (p-value<0.05 based DEG).....	135
Upset Intersection Graphs for log2 transformed vs. non-transformed normalized counts on p-value<0.05 DEGs	136
Expression Profiling of the Selected Genes in Expression Atlas	136
The difference between Bgee and TS genes in this thesis.....	137
Structure of Ache	137
References	138

List of Figures

Figure 1.1: The table showing the 1-to-1 ortholog of human ACHE gene in zebrafish with sequence-wise high similarity and the Gene Tree showing the similarity and synteny between the genomes of zebrafish and human genome for ACHE gene	1
Figure 1.2: General pipeline for differential gene expression analysis after RNA-Seq.....	4
Figure 1.3: The general pipeline followed in this study for the tissue specificity analysis of ache zebrafish.....	7
Figure 4.1: The distribution of log2FC and p-values upon analysis of the ache RNAseq dataset. Horizontal dashed lines represent 0.05 cut off of p-value and vertical dashed line shows the cut off value of 0.5849 for log fold change in base 2.	35
Figure 4.2: Boxplots of SD/LS (upper), RPKM (middle), TPM (Downes & Granato) normalized and log transformed counts.	37
Figure 4.3: The distribution of the tissue specificity scores of different metrics for SD, RPKM and TPM normalized counts. Tissue specificity scores were calculated based on the averaged tissue expression values.	39
Figure 4.4: Dendograms of SD, RPKM, TPM normalized and log transformed different tissue counts from different datasets.	41
Figure 4.5: PCA plots of the counts coming from different datasets after normalization and log transformation. (Upper, SD; Middle, RPKM; Bottom, TPM).....	43
Figure 4.6: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were Library Size (LS) normalized.....	45
Figure 4.7: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were RPKM normalized.	46
Figure 4.8: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were TPM normalized.	47
Figure 4.9: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in brain, for SD, RPKM or TPM normalized datasets, respectively.	50
Figure 4.10: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in liver for SD, RPKM and TPM normalized datasets, respectively.	51
Figure 4.11: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in skin for SD, RPKM and TPM normalized datasets, respectively.	51
Figure 4.12: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in retina for SD, RPKM and TPM normalized datasets, respectively.	52
Figure 4.13: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in muscle for SD, RPKM and TPM normalized datasets, respectively.	52
Figure 4.14: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in heart for SD, RPKM and TPM normalized datasets, respectively.	53
Figure 4.15: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in intestine for SD, RPKM and TPM normalized datasets, respectively.	53

Figure 4.16: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in kidney for SD, RPKM and TPM normalized datasets, respectively.	54
Figure 4.17: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in spleen for SD, RPKM and TPM normalized datasets, respectively.	54
Figure 4.18: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in bones for SD, RPKM and TPM normalized datasets, respectively.	55
Figure 4.19: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in gill for SD, RPKM and TPM normalized datasets, respectively.	55
Figure 4.20: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in ovary for SD, RPKM and TPM normalized datasets, respectively.	56
Figure 4.21: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in embryos for SD, RPKM and TPM normalized datasets, respectively.	56
Figure 4.22: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in UFEggs for SD, RPKM and TPM normalized datasets, respectively.	57
Figure 4.23: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in testis for SD, RPKM and TPM normalized datasets, respectively.	57
Figure 4.24: Upset graph showing the intersection between genes specific to Brain across differently normalized datasets.....	62
Figure 4.25: Upset graph showing the intersection between genes specific to Liver across differently normalized datasets.....	63
Figure 4.26: Upset graph showing the intersection between genes specific to skin across differently normalized datasets.....	63
Figure 4.27: Upset graph showing the intersection between genes specific to gill across differently normalized datasets.....	64
Figure 4.28: Upset graph showing the intersection between genes specific to heart across differently normalized datasets.....	64
Figure 4.29: Upset graph showing the intersection between genes specific to intestine across differently normalized datasets.....	65
Figure 4.30: Upset graph showing the intersection between genes specific to kidney across differently normalized datasets.....	66
Figure 4.31: Upset graph showing the intersection between genes specific to muscle across differently normalized datasets.....	66
Figure 4.32: Upset graph showing the intersection between genes specific to spleen across differently normalized datasets.....	67
Figure 4.33: Upset graph showing the intersection between genes specific to Bones across differently normalized datasets.....	68
Figure 4.34: Upset graph showing the intersection between genes specific to embryos across differently normalized datasets.....	68
Figure 4.35: Upset graph showing the intersection between genes specific to UFEggs across differently normalized datasets.....	69

Figure 4.36: Upset graph showing the intersection between genes specific to ovary across differently normalized datasets.....	69
Figure 4.37: Upset graph showing the intersection between genes specific to Retina across differently normalized datasets.....	70
Figure 4.38: Upset graph showing the intersection between genes specific to Testis across differently normalized datasets.....	71
Figure 4.39: Comparison of the distribution of Tau values on differently normalized vs. differently normalized and log2 tranformed values.....	71
Figure 4.40: Overall TS genes intersected in different normalization methods and log transformed (represented with additional “l” letter) vs. non transformed normalized counts. DE genes selected based on abs (L2FC)>0.5849 for ache dataset, and Tau>0.8 to define TS genes.	72
Figure 4.41: Overall TS genes intersected in different normalization methods and log transformed (represented with additional “l” letter) vs. non transformed normalized counts. DE genes selected based on p-value<0.1 for ache dataset, and Tau>0.8 to define TS genes.	73
Figure 4.42: Upset graph showing the intersection between genes specific to tissues across differently normalized datasets and TS metrics used on them. (TS value greater than 0. for all the methods).....	74
Figure 4.43: The qPCR results of the selected Tissue Specific (TS) and Differentially Expressed Genes (DEGs) on ache dataset. desma and mylz3 for muscle, pck1 and fabp10a for liver, arr3a and rpe65a for retina.	76
Figure 4.44: GO Biological Processes affected as liver specific up and down regulated genes in AChE dataset.	78
Figure 4.45: GO Biological Processes affected as retina specific up (up) and down regulated genes in AChE dataset.....	79
Figure 4.46: GO Biological Processes affected as skin specific up and down regulated genes in AChE dataset. None was available for downregulated genes.....	80
Figure 4.47: GO Biological Processes affected as muscle specific up and down regulated genes in AChE dataset.....	81
Figure 4.48: GO Biological Processes affected as ovary specific up and down regulated genes in AChE dataset.	82
Figure 4.49: GO Biological Processes affected as testis specific up and down regulated genes in AChE dataset.	83
Figure 4.50: GO Biological Processes affected as spleen specific up and down regulated genes in AChE dataset.....	84
Figure 4.51: GO Biological Processes affected as heart specific up and down regulated genes in AChE dataset.	85
Figure 4.52: GO Biological Processes affected as intestine specific up regulated genes in AChE dataset. None was available for down regulated genes.....	86
Figure 4.53: GO Biological Processes affected as bone specific up regulated genes in AChE dataset. None was available for down regulated genes.....	86
Figure 4.54: GO Biological Processes affected as gill specific up and down regulated genes in AChE dataset.	87
Figure 4.55: GO Biological Processes affected as embryo specific up and down regulated genes in AChE dataset.....	88
Figure 4.56: absolute log2 fold change greater than 0.5849 DEGs with Tau score greater than 0.8; a) up b) down regulated genes.	89

Figure 4.57: qPCR results of the *ache*, *arr3a*, *rom1b*, *rpe65* genes for WT: *ache* (+/?) and mutant: *ache* (-/-) sibling embryos of the same breedings raised under different lightening conditions. CL: Constant Light, N: Normal cycle; CD: Constant Darkness. dpf denotes days post fertilization..... 90

Supplementary Figure 1: The quality control scores of fastq files retrieved from European Bioinformatics Institute which were trimmed and improved by TrimGalore used in the analysis a) PRJNA263496 datasets b) PRJNA255848 datasets c) PRJNA203029 datasets d) PRJNA297904 datasets e) PRJNA299585 datasets.....	103
Supplementary Figure 2: The quality control scores of fastq files retrieved from European Nucleotide Archive used in the analysis a) PRJNA263496 datasets b) PRJNA255848 datasets c) PRJNA203029 datasets d) PRJNA297904 datasets e) PRJNA299585 datasets	104
Supplementary Figure 3: Tau score distributions of not merged TS data and TS data merged with AChE dataset.....	105
Supplementary Figure 4: The table showing the distributions of Tau(1), Tsi(2), Counts(3), Hg(4), Spm(5) and Gini(6) metrics values across SD(upper), RPKM(middle) and TPM(bottom) normalized datasets in the given order.	105
Supplementary Figure 5: The distributions of up and downregulated genes for 3 methods. Vertical dashed lines were drawn based on median of Tau values.	106
Supplementary Figure 6: Log transformed comparisons for log2fc dependent brain specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	111
Supplementary Figure 7: Log transformed comparisons for log2fc dependent liver specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	111
Supplementary Figure 8: Log transformed comparisons for log2fc dependent skin specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	112
Supplementary Figure 9: Log transformed comparisons for log2fc dependent gill specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	112
Supplementary Figure 10: Log transformed comparisons for log2fc dependent heart specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	113
Supplementary Figure 11: Log transformed comparisons for log2fc dependent intestine specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	113
Supplementary Figure 12: Log transformed comparisons for log2fc dependent kidney specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	114
Supplementary Figure 13: Log transformed comparisons for log2fc dependent muscle specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	114
Supplementary Figure 14: Log transformed comparisons for log2fc dependent spleen specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	115
Supplementary Figure 15: Log transformed comparisons for log2fc dependent bones specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	115
Supplementary Figure 16: Log transformed comparisons for log2fc dependent embryos specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	116
Supplementary Figure 17: Log transformed comparisons for log2fc dependent UFEggs specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts. ...	116
Supplementary Figure 18: Log transformed comparisons for log2fc dependent ovary specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	117
Supplementary Figure 19: Log transformed comparisons for log2fc dependent retina specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	117

Supplementary Figure 20: Log transformed comparisons for log2fc dependent testis specific DEGs of AChE dataset (for TS, $\text{Tau} > 0.8$). Letter l represents log transformation on normalized counts.	118
Supplementary Figure 21: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for embryo specific DEGs.	118
Supplementary Figure 22: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for UFeggs specific DEGs.	119
Supplementary Figure 23: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for ovary specific DEGs.	119
Supplementary Figure 24: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for brain specific DEGs.	120
Supplementary Figure 25: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for liver specific DEGs.	120
Supplementary Figure 26: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for skin specific DEGs.	121
Supplementary Figure 27: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for gill specific DEGs.	121
Supplementary Figure 28: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for heart specific DEGs.	122
Supplementary Figure 29: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for intestine specific DEGs.	122
Supplementary Figure 30: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for kidney specific DEGs.	123
Supplementary Figure 31: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for muscle specific DEGs.	123
Supplementary Figure 32: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for spleen specific DEGs.	124
Supplementary Figure 33: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for bones specific DEGs.	124
Supplementary Figure 34: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for retina specific DEGs.	125
Supplementary Figure 35: Comparisons with metrics and methods on abs (log2FC) > 0.5849 and Tau, Tsi, Gini, Spm, Hg > 0.8 for testis specific DEGs.	125
Supplementary Figure 36: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE bone genes.	126
Supplementary Figure 37: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE brain genes.	126
Supplementary Figure 38: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE embryo genes.	127
Supplementary Figure 39: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE gill genes.	127
Supplementary Figure 40: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE heart genes.	128
Supplementary Figure 41: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE intestine genes.	128
Supplementary Figure 42: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE kidney genes.	129
Supplementary Figure 43: Comparisons with metrics and methods on pvalue < 0.1 and Tau > 0.8 DE liver genes.	129

Supplementary Figure 44: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE muscle genes.....	130
Supplementary Figure 45: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE ovary genes.	130
Supplementary Figure 46: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE retina genes.	131
Supplementary Figure 47: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE skin genes.....	131
Supplementary Figure 48: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE spleen genes.	132
Supplementary Figure 49: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE testis genes.	132
Supplementary Figure 50: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE UFeggs genes.	133
Supplementary Figure 51: Comparisons with metrics and methods on pvalue<0.05 and Tau >0.8 for all TS-DEGs.	133
Supplementary Figure 52: Comparisons with metrics and methods on pvalue<0.1 and Tau, Tsi, Gini, Spm, Hg>0.8. Intersection sizes for all the tissues combined.....	134
Supplementary Figure 53: Comparisons with metrics and methods on pvalue<0.05 and Tau, Tsi, Gini, Spm, Hg>0.8. Intersection size for all TS-DEGs.	135
Supplementary Figure 54: Log transformed comparisons for p-value<0.05 based DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.	136
Supplementary Figure 55: TPM normalized expression of selected genes available for different developmental stages.	136
Supplementary Figure 56: Intersection sizes for Brain specific genes in either of the normalization methods used in this thesis or Bgee tool for zebrafish (Bgee_brain).....	137
Supplementary Figure 57: The ExPASy result for the protein domains, families and functions for zebrafish ache mRNA FASTA sequence.	137

List of Tables

Table 3.1: A list of general reagents and equipment used in this study.....	20
Table 3.2: List of primer pairs used in qPCR experiments.	21
Table 3.3: List of chemicals used for this thesis.	22
Table 3.4: SRA datasets which were analyzed for tissue specificity analyses	23
Table 3.5: The similarities and differences between normalization methodologies used in this study, and the formulas of them.	26
Table 3.6: Different tissue specificity metrics used in this study. N: Number of tissues available in the dataset. x_i represents the expression of a gene in a given tissue.....	27
Table 3.7: The dataset used for DEGs analysis of ache. dpf denotes days post fertilization.....	28
Table 3.8: qPCR reagents and their volumes	30
Table 3.9: qPCR conditions. 'denotes minutes, "denotes seconds.....	31
Table 3.10: A table showing the ideal progeny of ache ^{sb55/+} internal cross.....	32
Table 3.11: A table showing the experimental setup for light-dark experiments.....	33
Table 3.12: The contingency table used in the rest of the analysis.	33
Table 4.1: DEGs profile of ache dataset based on the demonstrated cut-off values of pvalue and log2fold change.....	34
Table 4.2: baseMean (total expression), log2foldchange, pvalue and padj values of selected genes, to be studied by qRT-PCR in ache mutant vs. healthy dataset, were shown.....	35
Table 4.3: The number of up and down regulated genes assigned to the tissues in which they had the highest expression value were shown in the table for different normalization methods applied.....	48
Table 4.4: Kolmogronov Smirnov test applied on Tau distributions for up and down regulated genes in different tissues (DEG cut off were absolute log2foldChange>0.58 difference).....	49
Table 4.5: Fishers Exact Test for up and down regulated genes. DEGs were determined by absolute log2 fold change greater than 0.5849 in AChE dataset.....	59
Table 4.6: Fishers Exact Test for up and down regulated genes. DEGs were determined by differetially expressed genes in AChE having p value smaller than 0.1.....	61
Table 4.7: Different Tissue specificity metric values for different normalization methods with the selected DEGs having Tau>0.8 and abs(log2FC)>0.5849, and ache gene.....	75
 Supplementary Table 1: KS test results to see the distribution difference between up and downregulated genes (DEG cut off log2fold change).....	107
Supplementary Table 2: KS test results to see the distribution difference between up and downregulated genes (DEG cut off pval smaller than 0.1).....	108
Supplementary Table 3: KS test results to see the distribution difference between up and downregulated genes (DEG cut off pval smaller than 0.05).....	109
Supplementary Table 4: Fishers Exact Test results and corresponding Odds Ratios for DEG defined as pval<0.05 and TS defined as Tau>0.8.....	110
Supplementary Table 5: Normalized expression of ache across different tissues. The top 3 tissues where expressed most were highlighted.....	110

Abbreviations

RNA-seq	RNA Sequencing
miRNA-seq	micro RNA Sequencing
DNA-seq	DNA Sequencing
TS	Tissue Specificity
TS-DEGs	Diferentially expressed genes specific to a tissue
TSI	Tissue Specificity Index
UFeggs	Unfertilized Eggs
CGC	Cancer Genomcis Cloud
RPKM	Read Per Kilobase Million
TPM	Transcript Per Kilobase Million
SD/LS	Sequencing Depth/Library Size
TMM	Trimmed Mean Median
FPKM	Fragment Per Kilobase Million
MRN	Median Ratio Normalization
RLE	Relative Log Expression
L2FC	Log fold change in base 2
PCA	Principal Component Analysis
KDE	Kernel Density Estimates
KS-test	Kolmogoronov-Smirnov Test
QC	Quality Control
fastq.gz	Compressed file format of fastq
pval	p value
pAdj	Adjusted p value
BH	Bonjemine-Hochberg
FDR	False Discovery Rate
FP/FN	False Positives/False Negatives
TP/TN	True Positives/True Negatives
dpf	days post fertilization
hpf	hours post fertilization
CD	Constant Dark

CL	Constant Light
N	Regular/normal light-dark cycle
BP	Biological Processes
MF	Molecular Function
CC	Cellular Component
GO	Gene Ontology
mRNA	messenger RNA
cDNA	complementary DNA
PCR	Polymerase Chain Reaction
qPCR	quantitative PCR
RT-qPCR	Real-Time qPCR
DEG/s	Differentially Expressed Gene/s
non-DEG	Genes which are not differentially expressed
ache	acetylcholinesterase
Ach	acetylcholine
ChAT	Choline Acetyltransferase
nAChR	nicotinic Acetylcholine Receptors
eef1a1	Danio rerio eukaryotic translation elongation factor 1
arr3a	arrestin 3a
rom1b	retinal outer segment membrane protein 1b
slc4a5	solute carrier family 4 member 5
rpe65a	retinal pigment epithelium specific 65 kDa
pck1	phosphoenolpyruvate carboxykinase 1
fabp10a	fatty acid binding protein 10a
desma	desmin a
mylz3	myosin, light polypeptide 3, skeletal muscle
GPCR/s	G-Protein Coupled Receptor/s
7TMR/s	7 Transmembrane Receptor/s

Chapter 1: Introduction

1.1 Zebrafish as a model organism

1.1.1 Zebrafish embryo

In vitro cell culture studies provide plethora of information however it is required to test the findings and further investigate the biological question using *in vivo* biological systems that resemble closely the humans. Therefore, it is crucial to study these processes on lower non-human organisms having higher similarity of genomes and functions of the genes with human such as *Danio rerio* (known as zebrafish) which is in fact better explained in **Figure 1.1**.

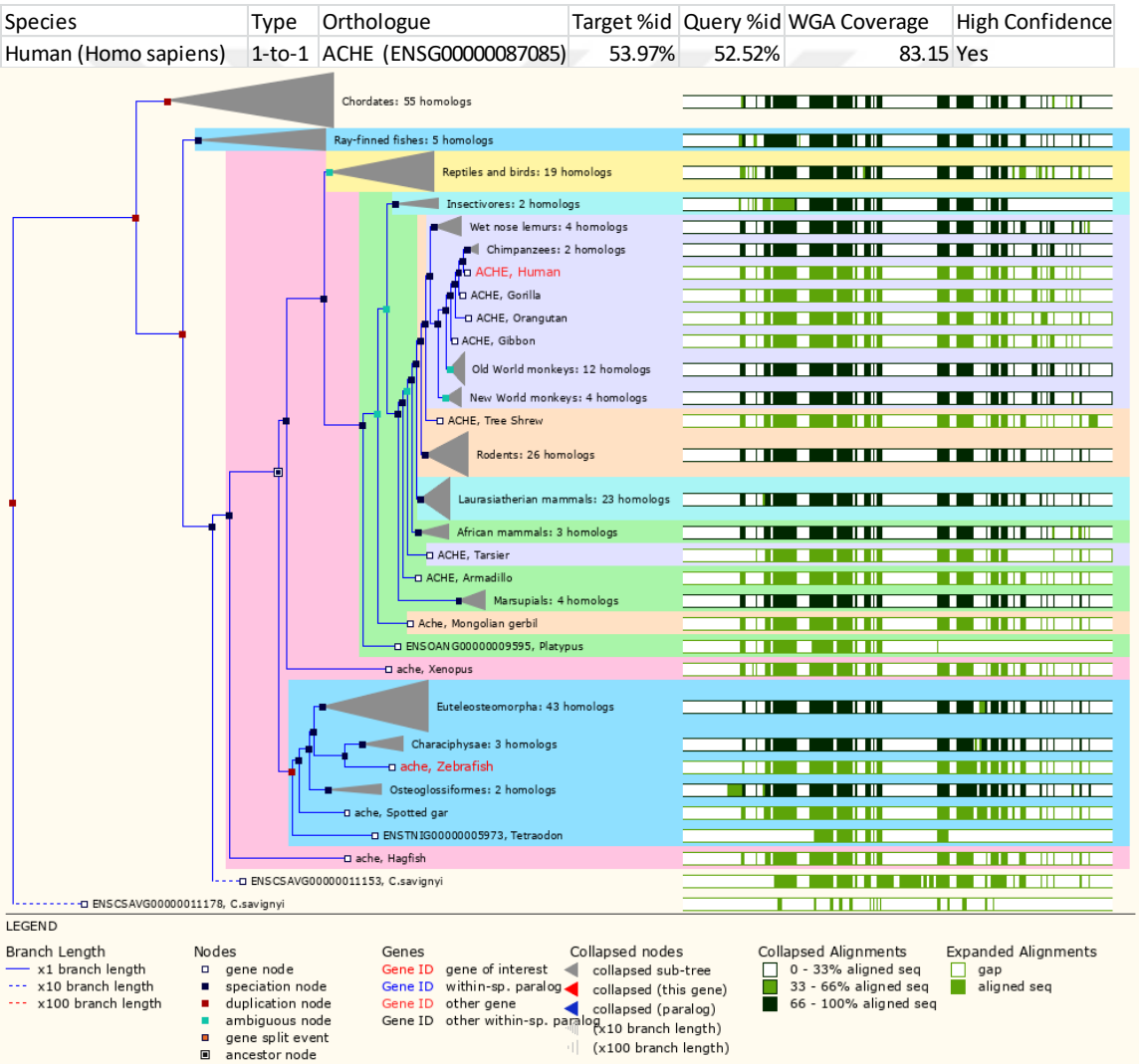


Figure 1.1: The table showing the 1-to-1 ortholog of human ACHE gene in zebrafish with sequence-wise high similarity and the Gene Tree showing the similarity and synteny between the genomes of zebrafish and human genome for ACHE gene

Figure 1 were generated and adapted by using Ensembl genomic database (www.ensembl.org/index.html). (Type: orthology type; Target / Query %id: percentage of sequence matching in human orthologue / zebrafish; WGA Coverage: Whole Genome Analysis Coverage, 100 is the highest score (Herrero et al., 2016)).

Zebrafish as a model organism is in fact very useful in terms of ease of breeding, having diploid chromosomes like human, an external embryonic development after fertilization of eggs, higher egg numbers produced in the breeding setups (e.g., 200-300 for one setup), and transparent embryos available to study under microscope to track the development (Driever, Stemple, Schier, & Solnica-Krezel, 1994; Wixon & O'Kane, 2000). However, it is necessary that genetic similarity and synteny between genomes of zebrafish and human to be comparable of the output received by these studies. Nevertheless, after the sequencing of zebrafish genome completed in 2013 done by Sanger Institute, it is known that of 26,000 protein-coding genes, zebrafish has more than 70% of them with at least one orthologue to human (more than 4, 000 GWAS associated genes) (Barbazuk et al., 2000; Howe et al., 2013). In addition to these, there is a conserved synteny between human and zebrafish having at least two genes located on the same chromosome in the given order (Barbazuk et al., 2000) such as zebrafish *ache* at chromosome 7. Thanks to homology and synteny between these two species, zebrafish has been used for knockout/knockdown/overexpression studies to investigate the role of specific genes in specific conditions or diseases such as role of acetylcholinesterase (*ache*) in Alzheimer's disease, toxicology studies and others (Behra et al., 2002; Behra, Etard, Cousin, & Strahle, 2004; Bradford et al., 2017; Chitramuthu & Bennett, 2013; Colovic, Krstic, Lazarevic-Pasti, Bondzic, & Vasic, 2013; Driever et al., 1996; Driever et al., 1994; Meyers, 2018).

1.1.2 *ache* mutant

Acetylcholine is one of the crucial and firstly discovered neurotransmitters which transmit information between neurons via synapses (DALE, 1914; Loewi & Navratil, 1926). Acetylcholinesterase, also known as *ache*, is a cholinesterase which inactivates acetylcholine by cleaving it to choline and acetate (Behra et al., 2002; Soreq & Seidman, 2001). When the acetylcholinesterase enzyme is not present, it is expected synapses to have accumulated acetylcholine because of no degradation event. Behra and colleagues (Behra et al., 2002) developed an *ache* mutant model

in zebrafish (by changing conserved Ser226 to Asn) to study the function of acetylcholinesterase enzyme and the aberrant production of acetylcholine in synapses. It was shown that this model is a functional one since ache enzymatic function was decreased drastically, and acetylcholine accumulated between neurons which triggered over-activation of muscles. Accordingly, homozygous ache mutant embryos were paralyzed at the third day after fertilization, hence remaining motionless against the tactile stimuli (3dpf- days post fertilization) (Behra et al., 2002).

Although, some higher organisms like mice have more similar Ache protein to humans than zebrafish has, there are more than one type of cholinesterase in such organisms which in turn compensating the function of the other when one of them is absent. For example, mouse has two type of cholinesterases which are butyryl- and acetyl-cholinesterases (Bche and Ache, respectively), and Bche can compensate in the absence of acetylcholinesterase Ache at a certain level in this organism, which makes studying the role of Ache in mouse challenging (Bertrand et al., 2001; W. Xie et al., 2000). One of the great advantages of investigating the function of ache in zebrafish is that ache is the only cholinesterase degrading the acetylcholine in zebrafish and it is one to one orthologous to human ache (Ref: Ensembl-Comperative Genomics-Orthologues analysis as shown in **Figure 1.1**).

On the other hand, interestingly ache expression in zebrafish starts before the 6 somite stage (12 hpf), and catalytically inactive forms (catalytic part of ache degrades the acetylcholine) in early development, which has raised the possibility that ache might have other roles than esterase activity, known as non-classical roles of ache (Appleyard, 1992; Betz, Bourgeois, & Changeux, 1980; Kreutzberg, 1969; Layer, 1990; Soreq & Seidman, 2001). Indeed, previous studies showed that ache has been expressed in non-cholinergic neurons and different tissues such as platelets, erythrocytes without accompanying acetylcholine and ChAT (Correa Mde et al., 2008; Rakonczay, Horvath, Juhasz, & Kalman, 2005). In contrast to ache's action certainly through degradation of acetylcholine, its structure, having catalytic and peripheral sites, raised the possibility of ache functioning as protease or zymogen of proteases with distinct functions such as neurite outgrowth,

morphogenetic stimulation, cellular proliferation, and so on (Soreq & Seidman, 2001).

1.2 RNA-Seq Analysis

1.2.1 Brief Introduction to RNA-Seq

The main aim of RNA-sequencing is to identify the expression of genomic loci in a given tissue/cell(s). It is in fact superior to cDNA microarrays in different ways which are: first, it does not require pre-define selection of probe sets, and so has higher sensitivity; second, several different additional applications such as spliced isoform or protein-RNA interaction domain detection are possible; and third, it allows *de novo* assembly (Nookaew et al., 2012; Z. Wang, Gerstein, & Snyder, 2009). RNA-sequencing technology has been mostly used to detect gene expression changes between different experimental conditions, tissues or other cell types at transcriptomic level.

General workflow of RNA-Seq is shown in the figure below (**Figure 1.2.1**). Upon obtaining RNA from tissues or cells and preparation of library and sequencing the third step is to process the information received via sequencing reads through alignment. It is also important to use proper pre-processing and normalization methodologies to reduce batch effects and variation due to experimental protocols. These are all done to find differentially expressed genes (DEGs) between different experimental conditions, which might shed light on the changes occurring at mRNA level from experimental and control group of interests.



Figure 1.2: General pipeline for differential gene expression analysis after RNA-Seq

RNA isolation, library preparation, strand preference, single vs. paired end selection of the reads, and even the sequencing platform might cause variances among the samples (Corley, MacKenzie, Beverdam, Roddam, & Wilkins, 2017; Leek et al., 2010; Maza, Frasse, Senin, Bouzayen, & Zouine, 2013; S. Zhao et al., 2015).

Therefore, it is important to select datasets that have applied the same approaches and platform used to test dynamicity of tissue specific gene expression profiling to reduce any sort of bias and batch effects, especially combining datasets (Maza et al., 2013).

On the other hand, the statistical design of the data is also important in RNA-Seq experiments as well. For example, it is not possible to estimate variation within group without replicates (Auer & Doerge, 2010; FISHER, 1935), and randomization is crucial to prevent any bias towards any of the datasets or materials used in the study (Auer & Doerge, 2010). To measure the quality of the data and increase the reproducibility of the findings, different quality metrics have been developed by measuring the coverage, GC bias, higher adapter sequence rate and so on for different sequencing methods such as RNA-Seq, miRNA-Seq, DNA-Seq and others (t Hoen et al., 2013). In fact, the number of replicates is important to get reproducible and reliable results. Nicholas Schurch and colleagues (Schurch et al., 2016) stated that at least six replicates for each group are required to get statistically significant results in the case of differential gene expression analysis. This may change depending on the degree of variability within an experiment.

There are different methods available to normalize the datasets and do differential gene expression analysis. Although Reads per Kilobase Million (RPKM) is the most frequently used as a normalization method at the early stages of RNA-Seq technology, some of the researchers later claimed that RPKM has its own bias. This includes conserving the average RPKM measure even among the same genome, so leading to inflated values of statistical significance by having lower p values than expected and inconsistent measure of relative molar concentration of reads, mRNA abundance. Therefore, Transcript per Million (TPM) is introduced as an alternative method as an improved version of RPKM normalization still considering the gene length (Wagner, Kin, & Lynch, 2012). In fact, some other methodologies do not consider gene length while normalizing the dataset for comparative studies, they consider the relative library sizes; and the most frequently used one is by DESeq2. Such methods can be divided into three basic categories: TMM (Trimmed Median M-values), MRN (Median Ration Normalization), and RLE (Relative Log Expression) which are main methodology used in DESeq2. In the paper, Elie Maza and colleagues (Maza et al., 2013) suggest that TMM and RLE behave well while

performing DE analysis, and FPKM (Fragment per kilobase million) should be avoided in differential gene expression analysis because of having poor benchmark profiling. In addition, Nicholas Schurch and his colleagues (Schurch et al., 2016) and Sonesson and Dolerenz (Sonesson & Delorenzi, 2013) and Costa-Silva, J and colleagues (Costa-Silva, Domingues, & Lopes, 2017) have approved that DESeq2 is one of the most convenient methods to measure DE genes across datasets with having lower FDR (False Discovery Rates). When Marie-Agnès Dillies et al. (Dillies et al., 2013) compared DESeq2, RPKM, and TMM, they concluded that RPKM should be avoided in case of DE analysis and DESeq2 and TMM are the most consistent methods especially when you consider most of the genes as non-DE.

Upon obtaining DEGs or a filtered list of genes and their expression values dimension reduction can help differentiate and get reliable results inherent grouping of the samples or the genes in the study. Although there are different methods used for this purpose, PCA is one of the most frequently used and easy to apply/interpret method with application of dendograms after that to know the data better (Meng et al., 2016).

There are also some problems encountered with RNA-Seq technology because tissues are analysed in bulk and represent a variable number of cell types. This can prevent researchers assess the response of individual cells. Luckily the technology is evolving, especially towards a brand new field called single cell RNA sequencing technology that help obtain a more resolute picture of cellular transcriptomes and changes that occur in physiology and pathology. Another problem of RNA-Seq technology comes with the length of the reads. There are short and long read methods, which can be used to study unique (mRNA) and repeated regions (SINEs, LINEs, long non-coding RNAs), respectively (Goodwin, McPherson, & McCombie, 2016). For example, it is a real challenge to get the same accuracy of long reads as in the case of short reads. In fact it is expensive and low throughput compared to short-read sequencing (Goodwin et al., 2016).

Although there is no outperforming workflow for RNA seq (Everaert et al., 2017), one of the most commonly used one is STAR Alignment- HTSeq Raw Count Data retrieval, which explained the paragraphs below (Dundar F and colleagues did before ((Friederike Dünder, 2015)).

Seven Bridges Cancer Genomics Cloud, also known as CGC (Lau et al., 2017), provides a cloud system to upload the fastq files and perform the following analysis on the cloud with the tools available for alignment such as STAR aligner (Alexander Dobin & Gingeras, 2016); quality controls such as FastQC and AlignmentQC; raw count data retrieval by using HTSeq (Anders & Huber, 2010). The overall pipeline was shown in **Figure 1.3** below.

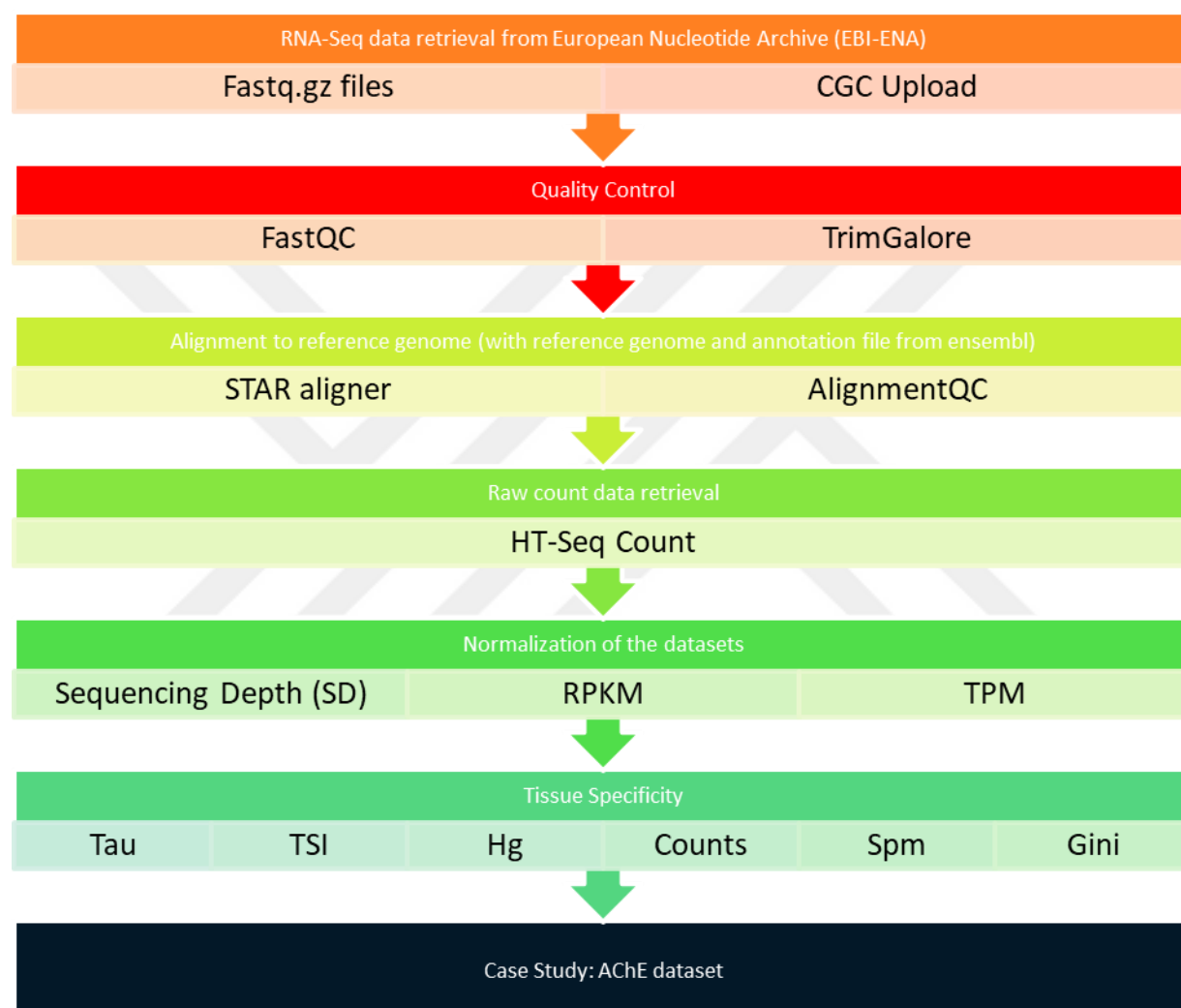


Figure 1.3: The general pipeline followed in this study for the tissue specificity analysis of ache zebrafish

Commonly used Integrated Development Environment (Gosalia, Economides, Dewey, & Balasubramanian) of R, also known as **R Studio** (RStudio Team, 2015), provides excellent environment to do the rest of the analysis which are need to analyse differential gene expression, functional and tissue annotation, and data mining and visualization by the packages available in CRAN or Bioconductor

libraries. These include: DESeq (Anders & Huber, 2010), DESeq2 (Love, Huber, & Anders, 2014) and edgeR (Robinson, McCarthy, & Smyth, 2010) for normalizations and DEGs analysis; ggplot2 (Wickham, 2016), RColorBrewer (Yu, Wang, Han, & He, 2012), extrafont (Yu, Wang et al. 2012), ggExtra (Baker, 2018), clusterProfiler (Yu et al., 2012) for data visualizations; org.Dr.eg.db (Carlson, 2019) and AnnotationDbi, biomaRt (Durinck, Spellman, Birney, & Huber, 2009) for annotations of *Danio rerio* genes (Hervé Pagès, 2019) and reactomePA (Yu & He, 2016) for functional annotations and pathway analysis.

1.3 Tissue Specificity

1.3.1 Tissue Specificity in mammals and zebrafish

As the sequencing technologies are evolving and providing a lot of data on transcriptomic profile of the study of interests and organisms, several new questions have arisen as about the degree of the significance of tissue specific expression profiling in diseases and treatment (Dezso et al., 2008; Xiao, Zhang, Zou, & Ji, 2010). Although the definition of tissue specificity might change depending on the article and metrics used, in general, a tissue specific gene can be defined as a gene relatively more expressed in one tissue with respect to other tissues (Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Yanai et al., 2005). On the other hand, the housekeeping genes are the genes ubiquitously expressed in all the tissues (Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Schug et al., 2005). Previously researchers have studied tissue specificity in human tissues (Dezso et al., 2008) or in human pathologies (e.g., cancer) (Kim et al., 2018)

These studies now recently being extended towards to the tissue specificity in zebrafish as well. For example reserachers (Weirick, John, Dimmeler, & Uchida, 2015) studied the role of non-coding long RNAs (nc-RNAs) in tissue specificity, however, in that study, the age of the fish was not taken into account.

1.3.2 Tissue Specificity on AChE Mutant Zebrafish

1.3.2.1 AChE and visual system

There are different studies trying to depict the relationship between visual system and ache in different species including rabbit (Ahn et al., 1995; Busnyuk, 1983), rat (Clemente et al., 2004; Corvin & Axelrod, 1969; Glow & Rose, 1964; Murakami, Takahashi, & Kawashima, 1984; Sanchez-Chavez, Vidal, & Salceda, 1995; Wood &

Rose, 1979), and hamster (Earnest & Turek, 1983). However, the main drawback in these studies is that there is a secondary enzyme that may replace the role of ache in the absence of it called bche (Behra et al., 2002; Bertrand et al., 2001; Downes & Granato, 2004), hence providing redundancy. This in turn can mask the effects of molecular modifications made on ache gene.

1.3.2.2 Tissue Specificity of Zebrafish Raised in Different Light Conditions

Zebrafish has a well-developed visual system having same class of retinal system with other vertebrates (Branchek, 1984; Laale, 1977) and has been used as an animal model for developmental genetics and neuroscience for a long time (M Barinaga, 1990; M. Barinaga, 1994) including effect of light conditions (J. Bilotta, 2000; Chapman, Tarboush, & Connaughton, 2012; Dekens et al., 2003; Hodel, Neuhauss, & Biehlmaier, 2006; Menger, Koke, & Cahill, 2005; Taylor, Chen, Luo, & Hitchcock, 2012; Villamizar, Vera, Foulkes, & Sanchez-Vazquez, 2014). The retinal structure of zebrafish is similar to human retina in terms of eye structure, eye development and visual processing (Chhetri, Jacobson, & Gueven, 2014).

1.3.2.3. ache is highly expressed in tissues such as muscle and heart

ache mutant model is a well-established model in zebrafish and its role in different functions beyond its classical acetylcholine degradation function and muscle activity has been studied before (Behra et al., 2002; Behra et al., 2004; Fukuto, 1990; Kluver et al., 2011; Paraoanu & Layer, 2008).

1.3.2.4 Pineal gland, circadian clock in zebrafish and potential role of ache

In contrast to suprachiasmatic nucleus in mammals which controls the circadian clock, pineal gland in zebrafish with its endogenous clock acting together with light controls the synthesis of melatonin (Dekens & Whitmore, 2008; Vallone, Lahiri, Dickmeis, & Foulkes, 2005). Pineal gland which controls circadian clock and triggers the rhodopsin production is present at 20-26 hrs after fertilization in zebrafish, while zebrafish is sensitive to visual stimuli in 24 hrs (Asaoka, Mano, Kojima, & Fukada, 2002; Kazimi & Cahill, 1999; P. Li et al., 2005). However studies showed that zebrafish embryos can detect the light at very early stages such as before the gastrula stage (5hpf) (Avanesov & Malicki, 2010; Clark, 1981; Easter & Nicola, 1996;

Tamai, Vardhanabhuti, Foulkes, & Whitmore, 2004; Villamizar et al., 2014; Ziv & Gothilf, 2006).

Previously researchers showed that the expression profile of ache gene has changed in response to a change in regular light-dark cycle (Murakami et al., 1984), and ache is expressed before acetylcholine is produced in the body of zebrafish (Villamizar et al., 2014). In addition, in retinal cone photoreceptors, signal of transduction is generated via activation of light which triggers an enzymatic cascade lowering the outer segment (OS) cGMP and Ca²⁺ in cytoplasm. This controls the cGMP-gated (CNG) ion channels (Korenbrod, 2012; Korenbrot, Mehta, Tserentsoodol, Postlethwait, & Rebrik, 2013), and calcium flux is also affected in the absence of ache at the embryonic stage. On the other hand, recently researchers showed in mice that Ach and light uses same signalling pathways in some part of mouse muscle in eye (Q. Wang et al., 2017), which might also reveal new functions of AChE-Ach relationship. So, there might be other regulators like ache in the retina which might shed light on the new direct/indirect roles of ache including apoptosis (Mena, Ortega, & Estrela, 2009; J. Xie et al., 2011), in Age related macular degeneration (AMD) (L. Cai et al., 2013), as a signalling molecule (Amy & Susan, 2012).

1.3.2.5 Genes Selected as Specific to a Tissue on AChE Mutant Zebrafish Embryo Dataset

In this section I will provide introductory information on the genes whose expression changes tested under different experimental conditions used in this thesis.

Retina Specific Genes

rpe65a

Retinoid isomerohydrolase RPE65a, rpe65a, which is the gene having seven paralogues in zebrafish, is orthologous (1-to-many) to human RPE65, retinal pigment epithelium 65 kDa protein, retinoid isomerohydrolase. Zebrafish rpe65a gene is mainly located in retinal pigment epithelium (Leung, Ma, & Dowling, 2007).

Although rods are functionalized for vision under the dim light, cones are specialized for vision under daylight conditions to enable color vision in high-resolution (Rodieck, 1998). The rods and cones develop in zebrafish together, but the larvae is rod-

dependent at the beginning (Joseph Bilotta, Saszik, & Sutherland, 2001). It was previously shown that although rod-receptor dominated part of the retina is dependent on rpe65a, cone-receptor dominated retina does not depend on rpe65a for visual perception and regeneration in that tissue (Hanovice et al., 2019; Schonthaler et al., 2007).

Human and zebrafish rpe65 genes have high sequence similarity (more than 70%), and found in similar tissues and share the similar molecular functions such as retinol isomerase activity, metal-ion binding and oxireductase activity and are involved in retinal metabolic processes and visual perception (based on Ensembl GO annotation). In addition, human RPE65 gene is also involved in biological processes like circadian rhythm.

arr3a

Arrestin 3a, retinal (X-arrestin), arr3a, has six paralogues in zebrafish, is orthologous (1-to-many) to human ARR3. This gene is located in X chromosome in humans thus is called as X-arrestin, with more than 50% sequence identity to zebrafish. Based on the Ensembl GO annotation, both human and zebrafish arrestins have roles in signal transduction and visual perception.

GPCRs (G-Protein Coupled Receptors) and arrestins are from the same family: 7 Transmembrane Receptors (7TMRs). GPCRs are one of the most important family of 7TMR proteins and are able to sense diverse signals coming from extracellular environment including light, ions, and hormones (Alvarez, 2008; Rompler et al., 2007). arr3 family belongs to cone opsin arrestins in zebrafish and its role is important in photo-response recovery, which is dependent on cone-opsin vision (Renninger, Gesemann, & Neuhauss, 2011). On the other hand, in the regenerating retina of zebrafish, arrestin 3a expression is increased (Eastlake et al., 2017), and in the case of a delay in retinal development, arr3a is decreased (S. Cai et al., 2018) showing the potential role of arr3a in regenerating retina of zebrafish.

rom1b

Retinal outer segment membrane protein 1b, rom1b, having 51 paralogues in zebrafish and is mainly found in the zebrafish eye. It has a 1-to-many orthologous status is is orthologous to human ROM1 (Vihtelic et al., 2005), having more than

35% sequence similarity. ROM1 and rom1b are both found in the integral part of membrane region and have a role in the cell surface receptor signalling pathway and visual perception (based on Ensembl GO annotations).

Even though, ROM1 studied in the context of retinal pigmentosa disease, which is a retinal dystrophy upon degeneration of photoreceptor cells (Q. Wang et al., 2001) in humans in general, the role of rom1b has not been studied widely in zebrafish. Therefore, rom1b was chosen as retina specific gene after the analysis of the datasets.

Muscle Specific Genes

desma

Desmin a, desma, having 57 paralogues in zebrafish, is a 1-to-many ortholog of human desmin, DES, gene, which is more than 65% sequence similarity to zebrafish. Based on Ensembl GO annotations, both human and zebrafish desmin are found in intermediate filaments and a special part of striated muscle fibers, Z disc, and they both are involved in structural muscle activity. DES and desma both regulate the muscle contraction including cardiac muscle.

Previously researchers have studied the role of desmin protein, desma, in zebrafish embryo, and showed that it has a role in myogenesis (Costa, Escaleira, Rodrigues, Manasfi, & Mermelstein, 2002). The knockdown studies showed its importance in muscle organization (M. Li, Andersson-Lendahl, Sejersen, & Arner, 2013) and cardiomyocyte and skeletal muscle function (Ramspacher et al., 2015).

mylz3

Myosin, light polypeptide 3, skeletal muscle, mylz3, gene is part of musculature system in zebrafish, and it has 7 paralogues. It is 1-to-many ortholog of the human myosin light chain 1, MYL1, with around 50% sequence similarity. MYL1 and mylz3 are both involved in calcium ion binding.

Researchers have demonstrated that the functional assembly of acetylcholine receptors are important for muscle contraction in zebrafish (Sepich, Wegner, O'Shea, & Westerfield, 1998; van der Meulen, Schipper, van Leeuwen, &

Kranenbarg, 2005). Myl family genes hence are important for cardiac muscle contraction (Chen et al., 2008).

Liver Specific Genes

pck1

Phosphoenolpyruvate carboxykinase 1 (soluble), pck1 and also known as PEPCK, is a 1-to-1 ortholog of human PCK1 gene with high sequence similarity (more than 70%). Zebrafish and human pck1 and PCK1 are both part of the cytoplasm and mitochondria and have various functions including phosphoenolpyruvate carboxykinase and GTP binding activities. In addition based on Ensembl GO processes, both are involved in biological processes such as gluconeogenesis, insulin response, hemotocyte differentiation, lipid response.

Pck1 regulates glucose metabolism in zebrafish (Elo, Villano, Govorko, & White, 2007), and high levels of glucose trigger the upregulation of pck1 in zebrafish (Z. Wang, Mao, Cui, Tang, & Wang, 2013). In glut12 morphants resulting in heart failure and diabetic phenotype, researchers showed that pck1 is also downregulated (Jimenez-Amilburu, Jong-Raadsen, Bakkers, Spaink, & Marin-Juez, 2015). On the other hand, zebrafish mutants having negative regulators of myogenesis, muscle formation, also causes the upregulation of pck1 (Gao et al., 2016).

fabp10a

Fatty acid binding protein 10a, liver basic, fabp10a, has 24 paralogues in zebrafish and has no direct ortholog to human. Based on Ensembl GO annotations, fabp10a is involved in lipid and bile-acid binding in zebrafish.

It has already been known as liver specific gene (Venkatachalam, Thisse, Thisse, & Wright, 2009) and used to decide on hepatotoxicity of the compounds in zebrafish larvae (Mesens et al., 2015).

1.3.3 Tissue Specificity Metrics Used in This Study

Tau

Tau was first described by Yanai et al. (Yanai et al., 2005); and a value of 0 represents a housekeeping gene and 1 represents a strictly a tissue-specific gene. Because 0 and 1 are more probable than any other intermediate values, there is U-

shaped distribution of the individual values are expected (Yanai et al., 2005). However, many of the values have intermediate tissue specificity scores. The values greater than 0.8 were used to set a gene as tissue-specific in the literature (Kryuchkova-Mostacci & Robinson-Rechavi, 2016).

Tau was used to define specificity in various other studies and databases including Prost! (Processing of Short Transcripts) (Desvignes, Batzel, Sydes, Eames, & Postlethwait, 2019); EnhancerDB (Kang et al., 2019); RATEmiRs (Bushel et al., 2018) for miRNA and mRNA tissue-specific expression measurements; LncBook (Ma et al., 2019) to study long non-coding RNAs Tissue Specificity; RNentropy (Zambelli et al., 2018) to identify genes with higher variation SINCERA (Guo, Wang, Potter, Whitsett, & Xu, 2015) to identify cell type signature in single cells, and G-NEST (Lemay et al., 2012) to delineate co-expressed co-conserved genes .

TSI

The TSI method was described by Julien et. al. (Julien et al., 2012); a value of 0 represents the ubiquitous expression of a gene and 1 represents the tissue-specific expression based on the expression profile of a gene at maximum level across tissues. TSI does not depend on the number of tissues a gene is expressed, and hence it behaves relatively different than Tau.

TSI is used to define tissue specificity in one database, RNentropy (Zambelli et al., 2018) together with Tau to detect variance of the gene expression among multiple RNA-sequencing experiments.

Counts

The specificity determined simply if the gene is expressed in only one tissue. Kryuchkova-Mostacci and colleagues (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) previously showed that it is useful only in the case of good cut offs chosen to decide on tissue specificity since it is binary (absolutely specific or not specific).

Before the RNA-Seq, Expression Sequence Tag (EST) with count of 1 used to define tissue specificity (Duret & Mouchiroud, 2000; Subramanian & Kumar, 2004). Log transformed counts were also used before (Park & Choi, 2010). Counts were used to

define uniformity of housekeeping genes (Lercher, Urrutia, & Hurst, 2002), and to rank tissue specific genes (Ponger, Duret, & Mouchiroud, 2001; Vinogradov, 2003).

Hg

Hg methodology that has been described by Schug et al (Schug et al., 2005) considers the entropy of the distribution of the expression of a gene relative to others: the values range from 0 to show single tissue-specific genes and to 1 for uniformly expressed genes, there is this term called ubiquitous expression of gene which defines the mean expression level of gene any other genes as a background, across tissues. Tissue specificity in Hg scores does not depend on the absolute expression levels but depends on the number of tissues a gene is expressed. To prevent any misunderstanding, Hg values were transformed in a way that they become comparable in the same range with other TS metrics.

Hg metric was used in TraVA, a database built to reveal cold-stress genes in *Arabidopsis thaliana* (Klepikova, Kulakovskiy, Kasianov, Logacheva, & Penin, 2019) DASHR 2.0, to define tissue specificity of small non-coding RNAs, (Kuksa et al., 2019); MAPPIN: as a method to annotate and predict pathogenicity and inheritance mode of nonsynonymous variants (Gosalia et al., 2017); KERIS (P. Li, Tompkins, Xiao, the, & Host Response to Injury Large-Scale Collaborative Research, 2017) to detect the genes differentially expressed in response to inflammation; PepPSy to annotate genes to specific proteins in specific tissues (Sallou et al., 2016); GPSy (Britto et al., 2012) to select specific genes conserved during biological processes via expression differences; QDMR method (Zhang et al., 2011) to detect differentially methylated regions; AtMetExpress (Matsuda et al., 2010) to detect different photochemical processes occurring in different tissues; EBConDB (Mazzarelli et al., 2007) to detect pancreas specific transcripts; REEF (Coppe, Danieli, & Bortoluzzi, 2006) to reveal tissue specific enriched genes; ROKU (Kadota, Ye, Nakai, Terada, & Shimizu, 2006) to detect tissue specific genes by using ranking based Hg values.

Spm

In the Spm method that has been described by Xiao et al (Xiao et al., 2010) the genes were represented at high dimensional spaces. Then, these expression values were projected into direction vector in the multidimensional space and divided by the

number of tissues available (length of the direction vector). If the gene is a major contributor, i.e. having the highest expression in a tissue, then a score of 1 was assigned that gene to consider it as tissue specific, and 0 to mention it as not-tissue specific regardless of absolute expression level of genes.

Spm metric used in CoNekT database (Proost & Mutwil, 2018) to define the expression specificity; and PaGenBase (Pan et al., 2013) to evaluate specific gene patterns to have better understanding of its function.

Gini

Gini method described in the paper of Kryuchkova-Mostacci et al (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) was used to measure the Gini Coefficient; and it was originally used in economics to define inequality (Gini. Corado, 1912) (Ceriani & Verme, 2012). The highest value used to define tissue specific score is dependent on the number of tissues available. For example, if 15 tissues are available for, the score given for tissues specificity is $([15-1]/15=)$ 0.93.

1.3.4 Existing tools to analyse tissue specificity of gene lists for mammals and/or zebrafish

Although there are different tools available for tissue specific expression profiling for human and mice/mouse these have not been commonly available or widely covered for zebrafish until recently.

TissueEnrich (Jain & Tuteja, 2018), which uses fold-change to define tissue specificity, is available for mouse and human as in the case of Human Protein Atlas (HPA) project (Uhlen et al., 2015) to measure specificity of the proteins across tissues. The term the authors used is “tissue enriched” for the genes 5-fold more expressed relative to all the others and “group enriched” for the genes having 5-fold change expression with respect to other tissues in more than one tissue.

TisGeD (Xiao et al., 2010), which uses Spm, as a tissue specificity metric, is available for human, mouse and rat.

TSEA online tool (<http://genetics.wustl.edu/jdlab/tsea/>) is available for human and mouse datasets. TSEA uses their own Specificity Index statistics (pSI) to define genes enriched at specific tissues (Dougherty, Schmidt, Nakajima, & Heintz, 2010), which uses log-fold change-based ranking of the genes.

Expression Atlas includes 6 studies for differential expression of zebrafish from embryonic stage to 120 hours embryo using Ensembl as the main source. However, it does not provide information for adult tissues (Ensembl Blog, 2016) (Collins, White, Searle, & Stemple, 2012; Petryszak et al., 2016). On the other hand, because organs mature at the adult stage, the expression profile of the genes at embryonic stage and adult tissues might differ from each other, in which the term tissue specific expression profiling is used for the genes specific to a tissue at the adult stage in general.

In fact, dynamicity of the expression is also important while deciding on tissue specificity as the transcriptomic alterations in elderly may shift towards degenerative diseases (Aramillo Irizar et al., 2018). Therefore, it is important to decide on the age of zebrafish samples rather than combining all the ages Weirick and colleagues (Weirick et al., 2015) did for **C-It-Loci** database, which does not use specific metrics to define TS but the expression difference for the discovery of tissue-enriched loci for human, mouse and zebrafish combined. However, different adult stages were considered as equal while performing the analysis. Even if 6 months old and 3 years old were both adult, the expression profile might change with the age however the authors did not consider the dynamicity of the expression at different stages of adulthood.

Although Andrea Komljenovic et al. (Komljenovic, Roux, Wollbrett, Robinson-Rechavi, & Bastian, 2016) consider the age of the zebrafish while measuring the differential expressed genes specific to tissues for **Bgee** tool, the outcome generated in the study coming from RNA-Sequencing was restricted to only four tissues at the adults stage which are spleen, heart, intestine and zone of the skin. However, Bgee also takes advantage of data other than RNAseq such as microarray and in situ hybridization. In addition to TSEA, Bgee also uses log-fold change based ranking metrics called pSI (Daugherty et al 2010), Specificity Index Statistics. These tools are mostly directed towards enrichment analysis as in the case of DAVID (Huang da, Sherman, & Lempicki, 2009a) and other similar tools (Huang da, Sherman, & Lempicki, 2009b) rather than defining of the tissue specific expression of the genes.

Chapter 2: Aims and Rationale

The publicly available RNAseq dataset for homozygous ache mutant and heterozygous/homozygous healthy ache siblings in comparison with each other has been used in this thesis while it was also used by two others (Armant, Gourain, Etard, & Strahle, 2016; Etard et al., 2015) and was originally published in GEO database with the accession GSE74202. These researchers previously focused especially on Unc45b and Hsp90aa1.1 genes but not ache found in the same dataset. Now, in this thesis, the main focus has been chosen as ache gene and on finding which tissue specific genes the ache mutation is associated with in the embryonic stage with respect to healthy fish.

Another aim of the present thesis is to identify tissue specific genes in zebrafish to be able to annotate the embryonic ache zebrafish expression profile. This tissue specificity analysis can help us analyze what happens in the absence of ache in zebrafish in a tissue-wise manner and define which tissues are more likely to be affected for future validations. On the other hand, there is no such absolute metric to define tissue specificity (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) but a series of different metrics have been defined and used in the literature. Yet a comparative study of these metrics to reveal zebrafish tissue specificity has not been done before. Therefore, this thesis might provide new insights which tissue-specificity indices behave similar to others and whether these indices agree on each other.

I also performed a preliminary analysis with DEGs between ache mutant and wildtype fish using String and DAVID. These have led to identification of retinal, muscular and liver tissues as suspects. Therefore, one of the aims of this thesis has become to test expression profile of retina specific genes in embryos in day 3 and 4 as well as embryos raised in different light-dark cycles across embryonic development.

Rpe65a, arr3a, rom1b were chosen as a retina specific genes; desma and myl3 were muscle specific; fabp10a and pck1 were liver specific after the analysis of the tissue datasets in zebrafish, and the possible relationship between retinal genes and

ache were further investigated under different light/dark conditions in zebrafish embryos in vivo using RT-qPCRs

Zebrafish has a single copy of ache and no bche therefore, zebrafish might be one of the best candidates to reveal the relationship between retinal genes and ache. However, direct evaluation of zebrafish embryos carrying a mutation in acetylcholinesterase gene reared in different lightening conditions has not been studied before.

Briefly, my aims in this study are:

- To compare the effects of normalization methodologies used to measure differential gene expression on the tissue specificity indices in zebrafish.
- To compare the effect of the metrics used to decide on genes specific to tissues by using 15 different zebrafish tissues while considering the dynamicity of gene expression hence focusing on only 5-9 months old adult fish.
- Apply the findings of tissue specificity and normalization methodology options on embryonic AchE dataset to gain new insights about the role of ache gene at organ development.
- To further investigate this relationship focussing on retinal genes, and confirm *in silico* results on AchE mutant embryos raised in different light/dark (constant light, constant dark, light/dark) conditions using RT-qPCR.
- To validate bioinformatics findings using RT-qPCR studies of selected genes in ache mutants vs. healthy 3 and 4 dpf embryos.

Chapter 3: Materials and Methods

3.1 Materials

3.1.1 General Reagents and Equipment

Here you can find the general reagents and equipment used in this study for the process of RNA isolation, cDNA production and qPCR experiments

Table 3.1: A list of general reagents and equipment used in this study

NAME	COMPANY	CATALOG NUMBER
SEROLOGICAL PIPETTES	CORNING, USA	4487-88-89
QIAzol LYSIS REAGENT	QIAGEN, GERMANY	79306
NUCLEASE FREE WATER	LONZA, SWITZERLAND	BE-51200
LIGHT CYCLER 480 MULTIWELL PLATES-96, WHITE	ROCHE, USA	04729692001
IScript CDNA SYNTHESIS KIT	BIORAD, USA	1708891
LIGHT CYCLER SYBR GREEN I MASTER	ROCHE, USA	04707516001

3.1.2 Primers

3.1.2.1 Primer Design

Primers were designed either using NCBI Primer Blast online tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) or Primer3 online tool (<http://primer3.ut.ee/>). While designing primers three main criteria were considered which were: 1) exon-exon junction spanning, 2) length of product size, in between 100-200 bp, 3) close annealing temperatures for forward and reverse primers. Then these primers were run *in silico* at UCSC Genome Browser – *In silico* PCR online tool (<https://genome.ucsc.edu/cgi-bin/hgPcr>) on zebrafish genome with the most recent assembly (May 2017- GRCz11/danRer11) to check whether these primers align with the zebrafish genome uniquely (one hit), and control whether it is really targeting an exon-exon junction (one expects the target longer than the target size on mRNA in the case of intron occurrence(s) in between two exons).

You can find a list of primers used for this study in the table below.

Table 3.2: List of primer pairs used in qPCR experiments.

Gene Name	Forward Primer (5' – 3')	Reverse Primer (5' – 3')	TM	Target Length
ache	CACAGTTCTGTGCTCTGTGGA	TTCCACTGTCTGCTCAACATC	52/ 54	85
eef1a1	CCCTGGACACAGAGACTTCA	CAGCCTCAAACCTACCAACA	59/ 59	101
arr3a	AGCAAGTGATGTGACAGTAGA G	TGCCCTGTGGAATCTGATATG	58/ 58	87
rpe65a	CATTCCTCTGGATCCCTTCA	CGAAATGGTTGAGTCCCAGT	57/ 57	221
rom1b	GCCCTCGTCAGCTACTACAT	GGTCTCCAGCAATGTCTCCT	59/ 59	201
slc4a5	CTGTTCTTCGGCACTTACTC	CAGCGCCATCCCATAATC	57/ 56	150
fabp10a	ACCGTCACCAACTCCTTCAC	GTTCTGCAGACCAGCTTTCC	59/ 59	
pck1	ATCGCTAAAGAGGAGGGCTG	TGTCGAACTTCATCCATGCG	59/ 57	196
mylz3	CCAGGTTGCCGATATCATGC	TTGGCGTCAACAGTCTTGAG	59/ 57	150
desma	ACGTATCGAGGGTCTTCACG	GCCTCGCTGATATTCTTGGC	59/ 59	195
sagb	AACCTGCCCTGTTCAGTGAC	GCATTATCGGTGGTTATCAGC	59/ 58	184

3.1.3 Solutions

3.1.3.1 Zebrafish Solutions (E3 Medium)

Zebrafish embryo medium also known as E3 medium was used to maintain embryos after fertilization. 60X stock was diluted to 1X by taking 16 ml from the stock and completed to 1000 ml with distilled water. 10 litres of 60 E3 stock includes 172 g NaCl, 7.6 g KCl, 29 g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, and 49 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$. 3 ml of 0.01% Methylene Blue was added to prevent any contamination such as fungi contamination.

3.1.4 Chemicals

Here you can find the chemicals used for this study for the process of RNA isolation.

Table 3.3: List of chemicals used for this thesis.

NAME	COMPANY	CATALOG NUMBER
ETHANOL	SIGMA ALDRICH, USA	32221
CHLOROFORM	SIGMA ALDRICH, USA	24216
ISOPROPANOL	SIGMA ALDRICH, USA	24137

3.2 Methods

3.2.1 RNA-Seq data analysis

3.2.1.1 RNAseq files used

5 replicates of 6 months old zebrafish brain tissue from PRJNA299585, one 5 months old brain sample from PRJNA255848, and a 6 months old zebrafish brain sample from PRJNA263496, in total 3 different brain samples were used. 5 replicates of 6 months old zebrafish skin tissue from PRJNA299585 and 9 replicates of 5 months old skin tissue samples from PRJNA203029 were used while measuring the tissue specificity. One 5 months old liver sample from PRJNA255848, and a 6 months old zebrafish liver sample from PRJNA263496, in total 3 different liver samples were used. For gill, heart, intestine, kidney, muscle and spleen two samples from PRJNA255848 and PRJNA263496 studies were used, replicate number was one for each of it. For, bones, unfertilized eggs, embryos, ovary and testis, one sample from PRJNA255848 was used for each tissue/stage.

Table 3.4: SRA datasets which were analyzed for tissue specificity analyses

Study accession	Platform	Library Layout	Age (months)	Tissues Included	Sex
PRJNA299585	Illumina HiSeq 1000	single	6	Brain, Liver, Skin	Male
PRJNA203029	Illumina Genome Analyzer IIx	single	5	Skin	Male
PRJNA297904	Illumina HiSeq 2000	paired	7-9	Retina	Male
PRJNA255848	Illumina HiSeq 2000	paired	5	Brain, Liver, Gill, Heart, Intestine, Kidney, Muscle, Spleen, Bones, Embryos, Unfertilized eggs, Ovary, Testis	Testis (Male); Embryos (Unknown); Kidney and Heart (Male and Female). Remaining (Female)
PRJNA263496	Illumina HiSeq 2000	paired	6	Brain, Liver, Gill, Heart, Intestine, Muscle, Spleen, Kidney	Both Male and Female

3.2.1.2 Seven Bridges Cancer Genomics Cloud (CGC)

It is a fact that RNA-Sequencing files (fastq) are large files, measured in gigabytes, so it is necessary to use a cloud system to store and process those files. Seven Bridges Cancer Genomics Cloud (<https://www.cancergenomicscloud.org/>) allows easy upload and process (e.g., quality control analysis, alignment, alignment quality

control, retrieval of raw count data). All RNA-Sequencing files were downloaded from SRA (Sequencing Read Archive) from European Nucleotide Archive shown in **Table 3.4**, and uploaded to Seven Bridges Cancer Genomics Cloud for further analysis by using the same pipeline. For the paired end reads, first pair was used to take the counts values.

3.2.1.2.1 Quality control of fastq files

Raw sequencing files were stored in fastq or compressed version of it as fastq.gz. It is important to know the sequencing quality of these files before processing those files and find a conclusion. If the sequencing quality is low (e.g., a lot of duplicates, dominant k-mers/adapters, not equal distribution of the bases), then it is not possible to trust the results coming from the analysis of those files. Therefore, all of the files used in tissue specificity studies were analyzed by FastQC (Fastq file Quality Control tool) developed by Babraham Informatics and available in CGC Cloud system. If necessary, adapter trimming was done by TrimGalore tool developed by Babraham Bioinformatics and available in CGC.

3.2.1.2.2 Sequence Alignment

Raw sequencing files include the sequencing reads with their counts. However it is necessary to align those reads to a specific genome, in this case the most recent zebrafish genome assembly from Ensembl: “Danio_rerio.GRCz11.dna.primary_assembly.fa”, to measure the exact values of unique sequences/reads. STAR Alignment tool (RNA-Seq Alignment – STAR 2.5.4b) (A. Dobin et al., 2013) available on CGC, was used with the default mode except two-pass mode as “basic” with reference genome from Ensembl.

Then, the quality control of the alignments was done with AlignmentQC tool available on CGC.

3.2.1.2.3 Raw Count Data Retrieval

After the alignment to a reference genome, it is possible to retrieve the count of each unique reads/genes. For this purpose, HTSeq Count developed by S. Anders and colleagues (Anders, Pyl, & Huber, 2015), available in CGC, was used with specific arrangements: “Danio_rerio.GRCz11.93.gtf” to annotate the reads to the genome, and trimmed and aligned files sorted by coordinate with union mode and ordering

them in name, and if the sequencing was done as stranded, stranded: “yes” options were chosen.

3.2.1.3 Normalization Methodologies

In this study three different normalization strategies have been tested by evaluating the tissue specificity results. These methods are SD, RPKM, and TPM. Each were explained below mentioning the package was used to calculate:

SD

Library size normalization provided by DESeq2 package [*estimateSizeFactors()* with a type as *iterate*] was used to normalize the different datasets to eliminate any kind of bias might be caused by sequencing depth differences.

RPKM

RPKM, known as also reads per kilobase million, normalizes the datasets first counting all the reads and dividing them into one million, where the term million comes from, (RPM), so according to sequencing depth. Then, these values are divided by the length of each gene in kilobases, where kilobase comes from. Default version of *rpkm()* function with its defaults provided in edgeR package were used. The general formula of rpkm (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008) was provided in the **Table 3.5**.

TPM

TPM, also known as Transcript per Kilobase Million, normalizes the datasets according to length of each gene in kilobases first (Reads Per Kilobase – RPK). Then, these values are counted for each sample, and divided by one million, where per million comes from. The general formula of tpm (Wagner et al., 2012) was provided in the **Table 3.5**. Sample specific library normalization was also done for TPM on the union exon specific gene length normalized counts.

Table 3.5: The similarities and differences between normalization methodologies used in this study, and the formulas of them.

SD	RPKM	TPM
Sequencing Depth	First, sequencing depth in 1 million (RPM)	First, based on gene length
	Second, RPMs for each gene based on gene length	Second, sequencing depth in 1 million
For the detailed formula of estimation, please look for the original papers: (Anders & Huber, 2010; Love et al., 2014)	$RPKM_i = \frac{X_i}{\left[\left(\frac{l_i}{10^3}\right)\left(\frac{N}{10^6}\right)\right]}$ (Mortazavi et al., 2008)	$TPM_i = [X_i / (l_i)] * [1 / \sum_j \frac{X_j}{l_k}] * 10^6$ (Wagner et al., 2012)

i for individual gene; X_i denotes the # of reads mapping to each gene; L_i denotes length of a gene; N denotes total # of mapped reads; $\sum_j X_j / l_k$ count per base for every gene.

3.2.1.4 Determination of Tissue Specificity

There are many available Tissue Specificity metrics/indices available as you can see in the the **Table 3.6**. However, Tau was selected as main metric for this study to determine tissue specificity. If a gene is expressed relatively higher in a specific tissue with respect to other tissues, based on the cut-off value, then these genes were labeled as genes specific to that tissue. The arithmetic average of samples of a tissue was compared with that of the other tissues when deciding which tissue a specified gene was expressed the most.

Table 3.6: Different tissue specificity metrics used in this study. N: Number of tissues available in the dataset. x_i represents the expression of a gene in a given tissue.

Method	Not Specific - Score	Tissue Specific - Score	Formula
Tau	0	1	$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1} ; \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} (x_i)}$
TSI	0	1	$TSI = \frac{\max_{1 \leq i \leq n} (x_i)}{\sum_{i=1}^n x_i}$
Counts	N	1	Counts = # of tissues expressed
H _g	log ₂ N	0	$H_g = -\sum_{i=1}^n p_i * \log_2(p_i) ; p_i = \frac{x_i}{\sum_{i=1}^n x_i}$
Spm	0	1	$SPM = \frac{x_i^2}{\sum_{i=1}^n x_i^2}$
Gini	0	(N-1)/N	$Gini = \frac{n+1}{n} - \frac{2 \sum_{i=1}^n (n+1-i)x_i}{n \sum_{i=1}^n x_i}$

3.2.1.5 Differential Gene Expression Analysis

The ultimate two main aims of this study were A) to find the genes differentially expressed among different tissues, so that B) DEGs from ache mutant vs. wildtype study can be compared with this tissue-specificity dataset. For tissue-specificity

dataset three different normalization methods were used and tissue specific genes were determined via different TS metrics as explained in sections 3.2.2.3 and 3.2.2.4. For ache dataset, DESeq2 package was used to find DEGs. Library size normalization was used, and log2 based differential expression ratios were used to select the DEGs. The ache dataset contained RNAseq data extracted 3 dpf healthy (homozygous or heterozygous) and 3 dpf ache mutant embryos (**Table 3.7**). ache -/- mutant embryos differ from other genotypes because they are paralyzed at 3 dpf due to excess acetylcholine in the absence of ache enzyme activity while heterozygous and homozygous wildtype are indistinguishable in phenotype.

Table 3.7: The dataset used for DEGs analysis of ache. dpf denotes days post fertilization

GEO accession number	Age	Strain(s)
GSE74202	3 dpf	ache (+/?), ache (-/-)

3.2.1.6 Gene Ontology/ Enrichment Analysis

Gene Ontology and enrichment analysis were done on the selected genes by either using bitr and enrichGo function from clusterProfiler (Yu et al., 2012) package with annotations provided by org.Dr.eb.db (Carlson, 2019) package, and visualizations with ggplot2 (Wickham, 2016) to gain insight about the molecular pathways, biological processes changed or specific to which tissue.

3.2.2 General Methods

3.2.2.1 Total RNA Extraction from Zebrafish

Upon performing the tail test, 3 to 5 dpf embryos were washed with 1 ml cold (+4) PBS, then PBS was removed. Next, these samples were snap-frozen in liquid nitrogen, and put into -80 for at least 1h. Embryos were kept on ice while performing the RNA isolation unless stated differently. 600 µl of Trizol (QIAGEN) was added for ~20 embryos, and these were homogenized using 2ml syringes. Then, 130 µl of chloroform (Sigma Aldrich) was added (1/5 of total volume of sample after Trizol addition), and shaken rigorously. The Trizol and chloroform including sample mixture was incubated for more than 3 minutes at room temperature (RT). After incubation at RT, the samples were centrifuged at 13 000 rpm for 20 minutes at +4. The remaining steps were done at +4. The supernatant at the top without any blurry or coloured

liquid was transferred into fresh tube. In this step, it was important to not to touch the cloud or not to take the pinkish coloured liquid at the bottom. Same volume of isopropanol was added onto the supernatant transferred into new tube and mixed gently. Next, isopropanol added mixture was centrifuged at 13 000 rpm for 20 minutes, then the supernatant is removed. The pellet was washed with 1 ml 75% ethanol (EtOH). Ethanol added pellet was centrifuged at 13 000 rpm for 15 minutes, then the supernatant was removed. The remaining pellet was washed with 100% EtOH; and centrifuged at 13 000 rpm for 15-20 minutes. After removing the supernatant, the remaining pellet was left for air dry at least for half an hour under chemical hood with clear air circulation. 20 µl of RNase free water (Lonza) was added onto the dried pellet to dissolve RNA. The concentration of the isolated RNA was measured with NanoDrop (Thermo Fisher Scientific, USA) with ssRNA option in ng/µl.

3.2.2.2 cDNA Synthesis

iScript cDNA synthesis kit was used to make cDNA from RNA templates isolated in the previous step with Trizol (QIAGEN, Germany). A master mix was prepared for all the samples by adding 4 µl reaction mixture and 1 µl of Reverse Transcriptase (RT) enzyme (BIORAD, USA) for each sample (It is important to prepare more, e.g., if you have 12 samples, prepare for 13 sample master mix to make sure in case of any pipetting error). 1000ng of RNA was taken from the isolated RNA sample, and completed to 15 µl by addition of RNase free water (e.g. if ng/µl was 500 for sample1, then take 2 µl from sample1 RNA, and add 13 µl RNase-free water). Next, 5 µl of master mix (4µl +1µl RT enzyme) was added, to make the total volume as 20 µl. For every reaction, one sample was prepared as negative control (without RT enzyme), to make sure there was no DNA contamination in the RNA isolated samples. The mixture was incubated at 25°C for 5 minutes, then 20 minutes at 46°C, and 1 minute at 95°C, and the reaction at PCR machine was finalized. RNA samples were converted to cDNA were stored at -20°C for further use.

3.2.2.3 Real time RT-qPCR

qPCR reactions were done with Roche 2X SYBR Green Mix (ROCHE, USA) in Light Cyclor 480 Multiwell plate 96, white (ROCHE, USA) by using Light Cyclor 480 PCR Machine (ROCHE, USA). 5 µl SYBR Green was added for a total volume of 10 µl reaction, and mixed in 200 µl PCR tubes with 1 µl of desired Forward and 1 µl of

desired Reverse primer and 1 μ l of RNase free water as shown in **Table 3.8**. 8 μ l of master mix was added into 96-well plate for every reaction. 2 μ l of cDNA (1:10 diluted cDNA from previous reaction) was added onto it. All the experiments were performed in duplicates. Negative control including 2 μ l of additional water instead of cDNA was included for every primer pair (10pmol final concentration) specific to a different gene was prepared for every qPCR reaction. After closing the lid, the plate was always centrifuged at 1,500 rpm for one minute to two minutes.

Table 3.8: qPCR reagents and their volumes

Reagent	Volume (μ l)
2X Roche Syber Green Mix	5
Forward Primer (10 μ M)	1
Reverse Primer (10 μ M)	1
cDNA	2
RNase free water	1
Total	10

qPCR reaction protocol used included an initial denaturation step at 95°C for 5 minutes, 40 cycles of denaturation at 95 °C for 10 seconds, annealing at desired temperatures for 20 seconds, and elongation at 72 °C for 20 seconds, acquisition at 95 °C for 5 seconds; 55 °C for 1 second: 95 °C for continuous, and final elongation at 40 °C for 10 seconds for each cycle as shown in **Table 3.9**

Table 3.9: qPCR conditions. ‘denotes minutes, “denotes seconds

Step	Temperature & Denaturation	
Initial Denaturation	95° C, 5'	
Denaturation	95° C, 10 "	} 40 cycles
Annealing	58° C, 20"	
Extension	72° C, 20"	
	95° C, 5 "	
Acquisition	55° C, 1 '	
	95° C, continuous	
Final Extension	40° C, 10"	

3.2.2.4 Q-RT-PCR Expression Analysis

The values representing signal of the expression level above the threshold (Ct) were analyzed via relative quantification of expression. The average of Ct values was taken for each technical duplicate. The average of Ct of control gene (without any treatment) were subtracted from the average of the specified conditions (e.g., treatment, ache mutant), which is called ΔCt . The target gene efficiency ($2^{\Delta Ct}$) was normalized to the reference gene efficiency, which is called $\Delta\Delta Ct$. The $\Delta\Delta Ct$ values were log (base 2) transformed for better representation and visualization, and these are used for the further statistical analysis.

Here you can find the equation to calculate the relative expression rates which is derived from Livak (Livak & Schmittgen, 2001):

$$\text{Log}_2 \left[2^{\Delta Ct_{\text{GOF}}} \left(\frac{\text{Control(ache+/-)} - \text{Treatment(ache+/-)}}{\text{Control(ache+/-)} - \text{Treatment(ache+/-)}} \right) / 2^{\Delta Ct_{\text{HG}}} \left(\frac{\text{Control(ache+/-)} - \text{Treatment(ache+/-)}}{\text{Control(ache+/-)} - \text{Treatment(ache+/-)}} \right) \right]$$

GOF: Gene of interest; **HG:** Housekeeping gene

3.2.3 In vivo zebrafish experiments

3.2.3.1 Zebrafish embryo strain

Homozygous mutant embryos (ache -/-) paralyze at day 3 and however, can be studied until day 5. Heterozygous mutant (ache +/-) are as healthy as the wild-type

strain (*ache* ^{+/+}), and defined in the same frame with *ache* (^{+/+}). *ache* (^{+/?}) was used for them. So, the progenies of *ache*^{*sb55*/+} x *ache*^{*sb55*/+} were used to get *ache*^{*sb55*/*sb55*} (1:4) with *ache*^{*sb55*/+} (2:4) and *ache*^{*+/+*} (1:4) embryos, which also provided internal control for the comparison of the homozygous mutant, heterozygous mutant and wild-type embryos from the same progeny. All the experiments were done under the ethical approval of Bilkent University Animal Ethics Committee (2017/16).

The ideal progeny results of internal cross for *ache* mutant fish was shown in **Table 3.10**.

Table 3.10: A table showing the ideal progeny of *ache*^{*sb55*/+} internal cross

	<i>ache</i> ^{<i>sb55</i>}	<i>ache</i> ^{<i>+</i>}
<i>ache</i> ^{<i>sb55</i>}	<i>ache</i> ^{<i>sb55</i>/<i>sb55</i>}	<i>ache</i> ^{<i>sb55</i>/+}
<i>ache</i> ^{<i>+</i>}	<i>ache</i> ^{<i>sb55</i>/+}	<i>ache</i> ^{<i>+/+</i>}

3.2.3.2 Breeding Setups

It is difficult to distinguish embryos, wt or mut, before 3 days post fertilization (dpf). Although, according to one of the protocols, touching the trunk at 27 hour post fertilization (hpf) might also reveal the phenotype of the embryo (Behra, Cousin et al. 2002), it is important to reduce light exposure here (even the microscope light) (Cousin, Strahle, & Chatonnet, 2005; Vallone et al., 2005; Vallone, Lahiri, Dickmeis, & Foulkes, 2007). Therefore, tail test at 3 dpf was used to separate embryos from each other based on phenotypic difference based on a previous published study (Behra et al., 2002).

Progeny of female and male *ache* (^{+/?}) embryos raised at 28°C under normal LD (14 hrs light and 10 hrs dark) cycle in E3 medium were used to get *ache* (^{+/?}) and *ache* (^{-/-}) genotypes. Fertilized eggs from each spawning was collected and pooled within 2 hours of spawning (before blastula) (Zebrafish can detect light at gastrula stage-5 hpf) (Avanesov & Malicki, 2010; Easter & Nicola, 1996; Tamai et al., 2004; Villamizar et al., 2014; Ziv & Gothilf, 2006).

The hatched embryos were transferred into a new petri dish (hatching--> 48 to 72 h after fertilization). *ache* (^{+/?}) and *ache*(^{-/-}) embryos were raised under control(LD)/ constant darkness (CD) [wrapped in black tape and aquaria will also be covered with

thick black plastic to not permit any light contamination]/ constant light(LL) conditions until 5 dpf.

Table 3.11: A table showing the experimental setup for light-dark experiments

ache (+/?) - LD	ache (+/?) - CD	ache (+/?) – LL
ache (-/-) - LD	ache (-/-) - CD	ache (-/-) – LL

[At least 10 larvae for each group were used and these could be distinguishable at 27hpf and 3dpf after tail test. LD denotes the embryo raised in regular light dark cycle; DD in constant darkness; LL in constant light]

3.2.4 Statistical Analysis

Fishers Exact Tests were performed to see whether differentially expressed genes, genes with log2 fold change greater than 0.5849 or p-value smaller than 0.1/0.05, were enriched for tissue specificity, genes having Tau greater than 0.8, compared to non-differentially expressed genes determined by the specified cut-offs based on log fold change in base two for *in silico* datasets. The analyses were done for up and down regulated genes separately as shown in **Table 3.12**.

Table 3.12: The contingency table used in the rest of the analysis.

	Differentially Expressed [abs(log2FC)>0.5849]	Not Differentially Expressed
Tissue Specific (Tau>0.8)	TS and DEG	TS but not DEG
Not Tissue Specific	Not TS but DEG	Neither TS nor DEG

GraphPad Prism 6.0 was used to perform statistical analysis of the qPCR experiments. One- way ANOVA was used to see the difference between mRNA expression levels of the selected genes for different lightening conditions. Additionally, Sidak's multiple comparison test was used to perform the statistical analysis for comparison of each conditions within the specified conditions.

Chapter 4: Results

4.1 DEGs of Ache -/- with respect to Ache +/-?

The raw counts of ache RNAseq dataset were retrieved from GEO and normalized with DESeq2 library size normalization. As shown in **Table 4.1**, there was a drastic decrease of up and down regulated DE genes when cut-offs to determine differentially expressed genes of ache dataset were used.

Table 4.1: DEGs profile of ache dataset based on the demonstrated cut-off values of pvalue and log2fold change.

Cut Offs	Up	Down
No cut off	14064	14513
pval< 0.1	1280	644
pval < 0.05	902	377
abs(Log2FC)> 0.5849	3102	2377
pval< 0.1 and abs(Log2FC)> 0.5849	597	233
pval < 0.05 and abs(Log2FC)> 0.5849	505	182

[Explanation for L2FC, if the expression in one sample is 10 and another is 15, Log2FC equals to 0.5849 correspond the differential expression between two samples. Namely, 1.5 fold of a sample is 0.5849 in terms of Log2FC. So, this value was chosen as our cut-off for log2FC in the case of consideration of log2FC.]

As shown in **Figure 4.1**, the distribution of up and down regulated genes for L2FC and p-value did not show big difference. However, there were more up regulated genes than down- regulated genes when p-value became 0.1 or 0.05.

AChE Paralyzed vs. Healthy

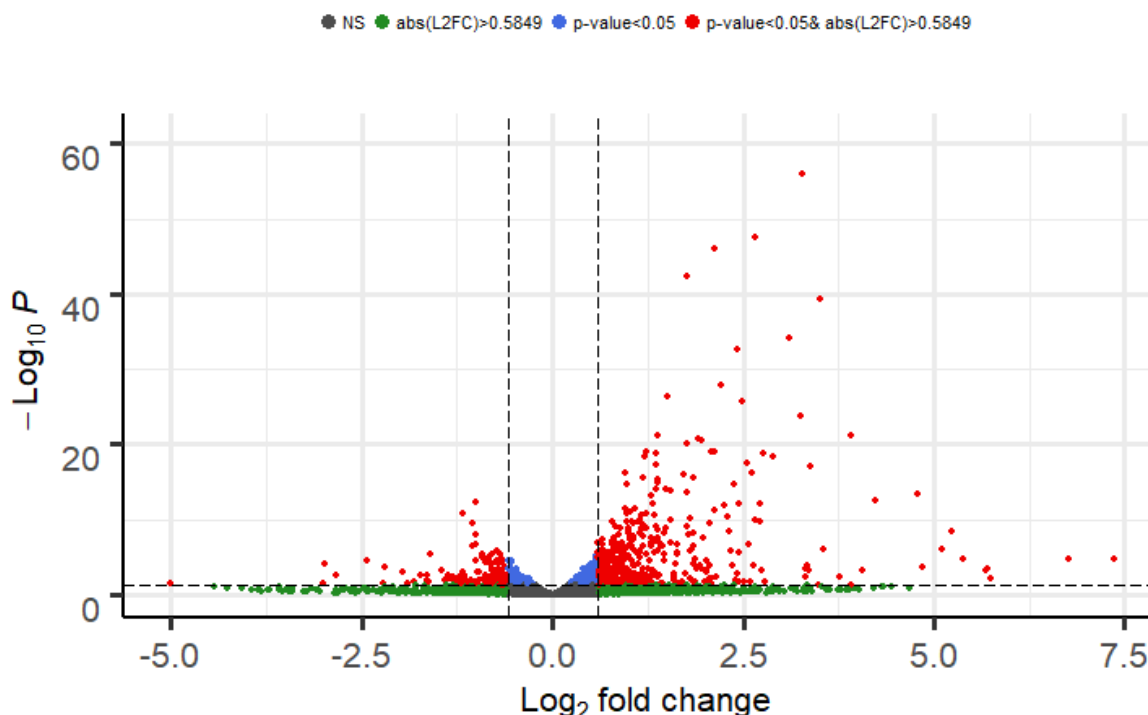


Figure 4.1: The distribution of log2FC and p-values upon analysis of the ache RNAseq dataset. Horizontal dashed lines represent 0.05 cut off of p-value and vertical dashed line shows the cut off value of 0.5849 for log fold change in base 2.

Table 4.2: baseMean (total expression), log2foldchange, pvalue and padj values of selected genes, to be studied by qRT-PCR in ache mutant vs. healthy dataset, were shown.

Ensembl ID	symbol	baseMean	log2FoldChange	pvalue	pAdj
ENSDARG000000031796	<i>ache</i>	6546.84142	0.443020385	0.00427	0.188263
ENSDARG000000056511	<i>arra3</i>	4118.70944	-0.686547993	3.39E-06	0.000461
ENSDARG000000026926	<i>rom1b</i>	181.66922	-0.65608163	0.029522	0.711376
ENSDARG000000007480	<i>rpe65a</i>	532.888406	-0.74727898	4.23E-05	0.004092
ENSDARG000000038439	<i>fabp10a</i>	214.954739	-0.737114302	0.004645	0.199538
ENSDARG000000013522	<i>pck1</i>	3353.69535	1.757234644	4.64E-43	2.86E-39
ENSDARG000000058656	<i>desma</i>	12776.7398	0.965204026	6.87E-17	4.84E-14
ENSDARG000000017441	<i>mylz3</i>	96150.6131	-0.611320884	0.000198	0.015441
ENSDARG000000038378	<i>sagb</i>	1243.08322	-0.689931708	2.09E-05	0.002235

All of the genes had p-value smaller than 0.05 and all of them except *ache* gene had absolute log2FC greater than 0.5849. *arra3a*, *rom1b*, *rpe65a* and *sagb* were selected as retina specific. The *sagb* primers did not work, no result was presented for this gene. *fabp10a* and *pck1* were selected as liver specific; *desma* and *mylz3* were

selected as muscle specific genes to further investigate the tissue specific expression change in ache mutant embryos. The upregulated genes were *pck1* and *desma*; downregulated genes were *arr3a*, *rom1b*, *rpe65a*, *fabp10a*, *mylz3* and *sagb* (Table 4.2).

4.2 Tissue Specificity Analysis

The quality control of RNA-Sequencing was done on the files retrieved from European Nucleotide Archive – European Bioinformatics Institute by FastQC tool available at CGC. The important point about evaluating these files was how good was the report overall. *Red crosses* represent the fields unusual, *orange exclamation signs* represent the fields with slightly abnormal, and *green* represents the results with normal while the sequencing run or before sequencing run due to any kind of handmade bias. These biases can be prevented at some level. Trim Galore tool provides both adapter trimming and improvement in the quality of sequencing. As one can see from the tables provided in **Supplementary Figure 1** and **2**, Trim Galore had no significant effect on the overall quality of the scores. Therefore, the rest of the analysis was done on non-trimmed counts.

4.2.1 How does normalized data look like?

Three normalization methods were used and compared (for tissue specificity dataset), namely, Library Size (or Sequencing Depth), RPKM and TPM. The boxplots of the datasets normalized with either of these methods were shown in **Figure 4.2**.

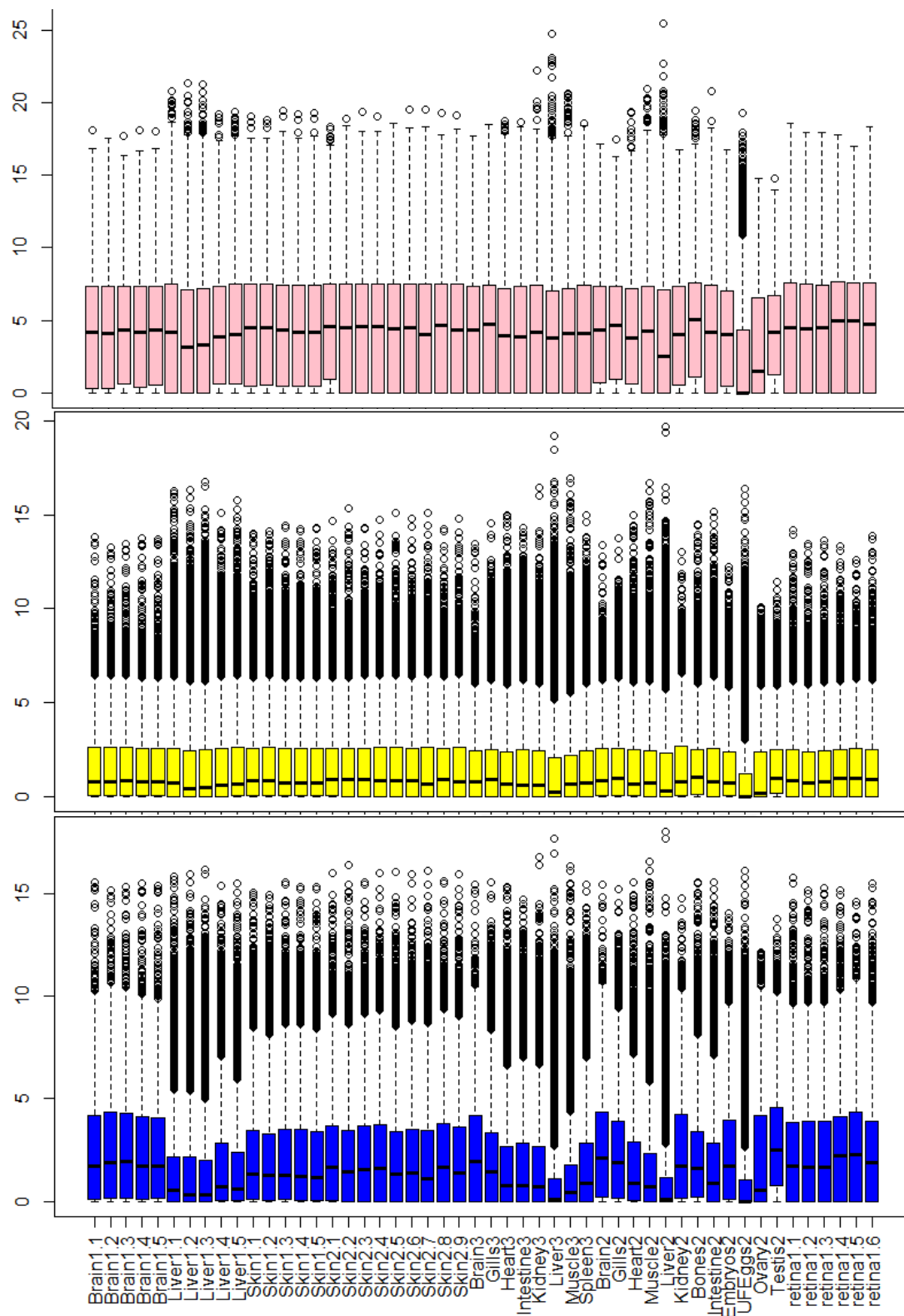


Figure 4.2: Boxplots of SD/LS (upper), RPKM (middle), TPM (Downes & Granato) normalized and log transformed counts.

The distribution of the values after each of the normalization step was standardized across samples. The most equal percentile of distribution in the normalized datasets belonged to SD/LS normalized dataset as shown in the **Figure 4.2** (upper graph). However, the normalization step did not work well on ovary and unfertilized eggs which did not have any replicates. TPM exhibited the most unequal distribution among the three different types of normalization.

The overall distributions of tissue specificity metrics values were almost same for different normalization methods (SD, RPKM and TPM) except the Gini distribution when using TPM (**Figure 4.3**). **Figure 4.3** showed that Tau and Spm metrics values were grouped around 1 and 0 more distinctively than Tsi values. Tau, Hg Spm, Gini distributions were consistent with each other while Tsi, where number of tissues is not used in calculating the metric, was not consistent with any of them except the counts method, which requires only one tissue to be expressed to be considered as tissue specific.

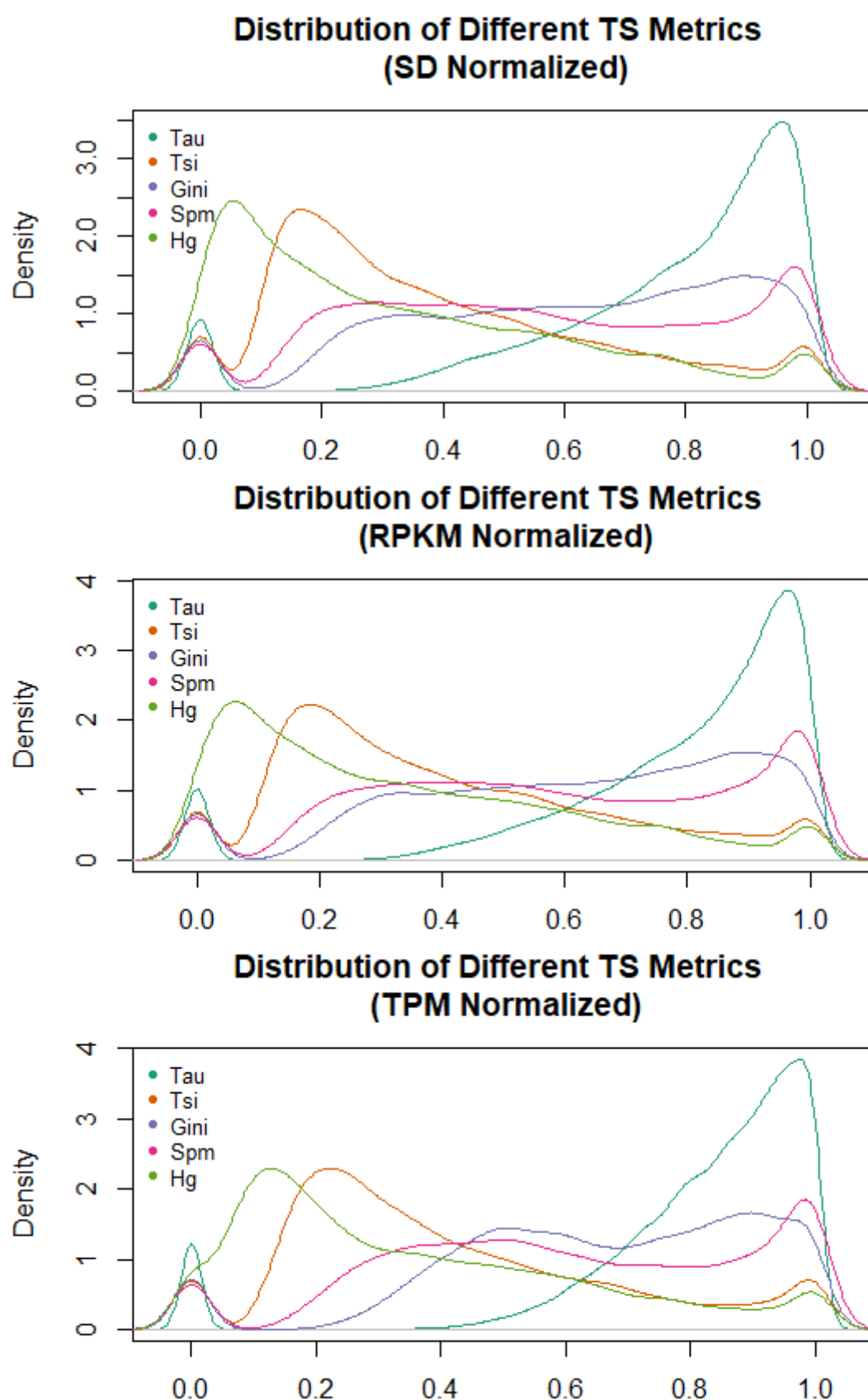


Figure 4.3: The distribution of the tissue specificity scores of different metrics for SD, RPKM and TPM normalized counts. Tissue specificity scores were calculated based on the averaged tissue expression values.

4.2.2 How are the tissues from different datasets clustered?

For each normalization method used, the tissues were clustered based on their expressions to observe the similar tissues/samples using dendograms drawn based on *pearson correlation* and *complete linkage method* as shown in **Figure 4.4**. Same tissues from different datasets were clustered rather distinctly yet but most closely to each other than they were to other tissue clusters.



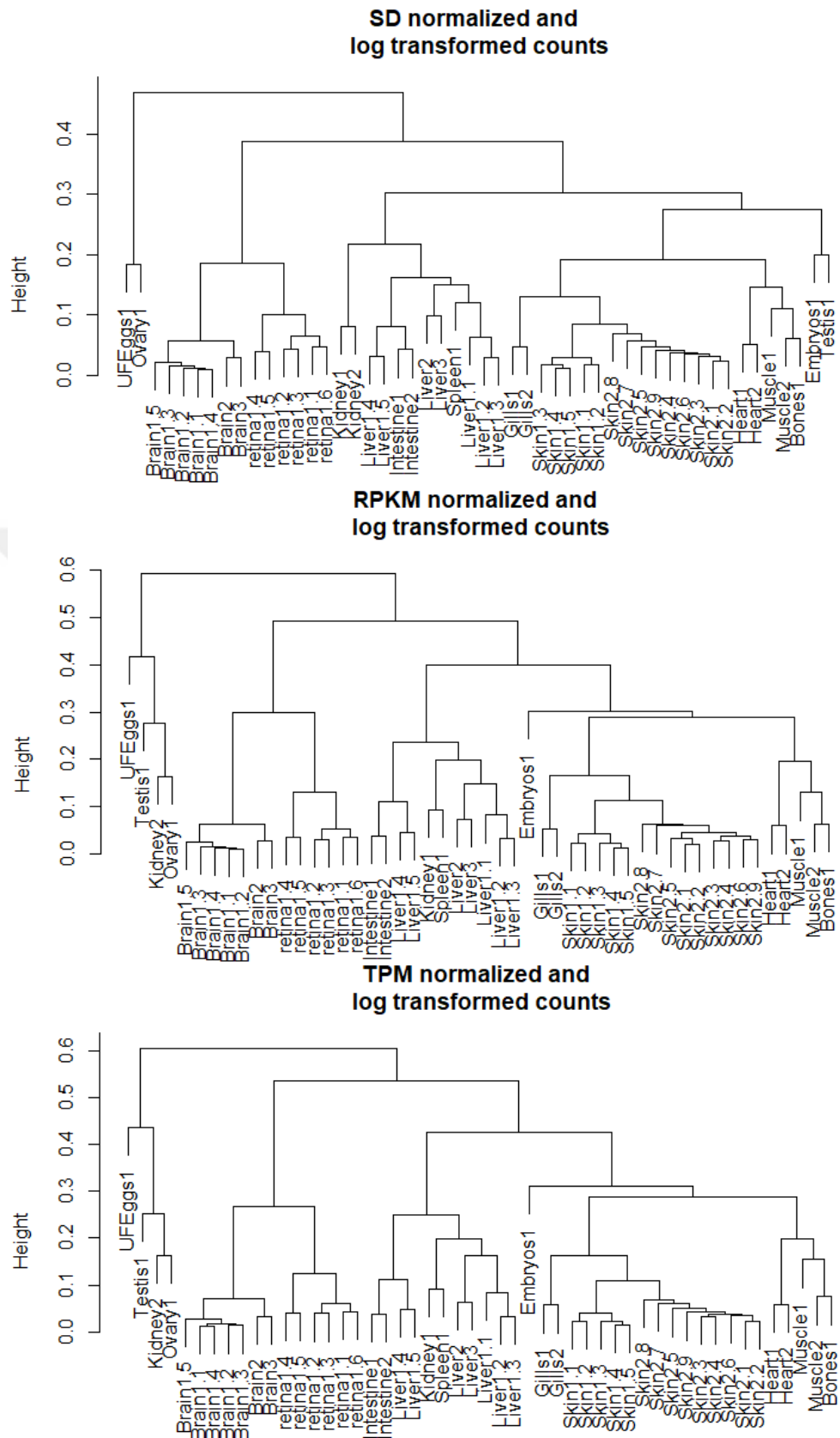


Figure 4.4: Dendrograms of SD, RPKM, TPM normalized and log transformed different tissue counts from different datasets.

Brain and retinal genes clustered together in all of the normalizations (**Figure 4.4**). Skin samples from different experiments were clustered together for all the normalization methods applied. Gill samples were clustered most closely with skin, and also found in the same cluster with muscle and heart and bone tissues that clustered very closely. Liver tissues were clustered together with kidney and spleen, and intestine was in their neighborhood. SD was the only method that correctly clustered the two kidney samples together (**Figure 4.4**). SD was also able to separate egg and ovary samples from the cluster of testis and embryo while the other two methods, RPKM and TPM clustered them all in the same group.



As shown in **Figure 4.5**, Principal Components Analysis (PCA) allowed the visualization of clusters shown in **Figure 4.4**. Different normalization methods indicated similar localizations for similar tissues, e.g., brain and retina closer to each other than were to skin or liver-intestine-kidney clusters. Most variable tissues were ovary, testis and embryo across different normalizations. However, some of the liver samples (Liver2 and Liver3) behaved slightly different for SD and TPM normalized data.

4.3 Application of Tissue Specificity Indices on Ache -/- vs. Ache +/- Dataset

4.3.1 How does the normalization method affect the distribution of tissue specificity scores for up- and down-regulated genes in ache mutants?

As shown in **Figure 4.6**, **4.7** and **4.8**, for each normalization method, respectively, the down-regulated genes were mainly distributed around 1, defined as tissue specific (denoted TS from now on) by Tau, Spm and Gini. When using Hg metrics, there were not many TS genes around 1 as in the case of Tsi, and unlike Tau, Spm or Gini. In addition, TS counts were lower in Tsi and Hg metrics than Tau, Spm or Gini. Because most of the TS genes were determined based on Tau metrics in the literature, a cut off value of 0.8 (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) was used to decide on tissue specificity of a gene among 15 tissues normalized with either of three of the normalization methods. In addition, a value of 0.8 is the starting point of the peak around 1 in both Spm and Gini as well, which also used the value of 1 to define absolute tissue specificity. On the other hand, the peak point starts around 0.4 for Hg, and not as strong as in Tau, Spm or Gini. The distribution of Tsi (b part of each figure) metrics was not compatible with the other metrics tested except Hg.

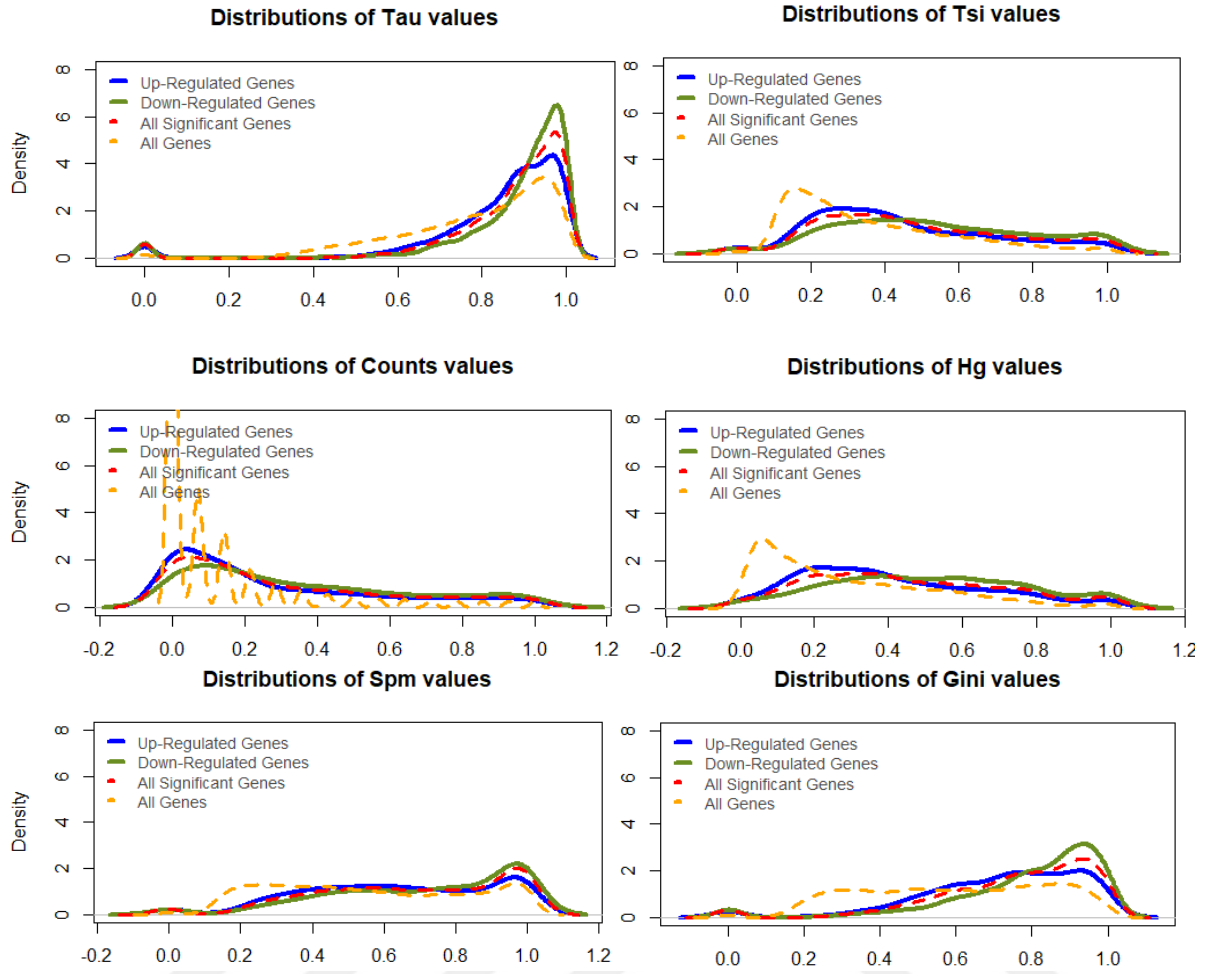


Figure 4.6: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were Library Size (LS) normalized.

Figure 4.6 shows density of up- and down-regulated genes by ache mutation for each of the tissue specificity metric used in the present study when SD normalization was used. Tau metric showed there were more TS-specific downregulated genes than more upregulated genes. Gini and Spm exhibited a very similar distribution to Tau while the distribution profiles of Tsi and Hg were more similar to each other than Tau and Gini and Spm. However, all of these five different metrics showed higher number of down-regulated genes for being tissue-specific. Counts were not defined any kind of tissue specificity at all, which might be due to not defining a lowest expression cutoff for it.

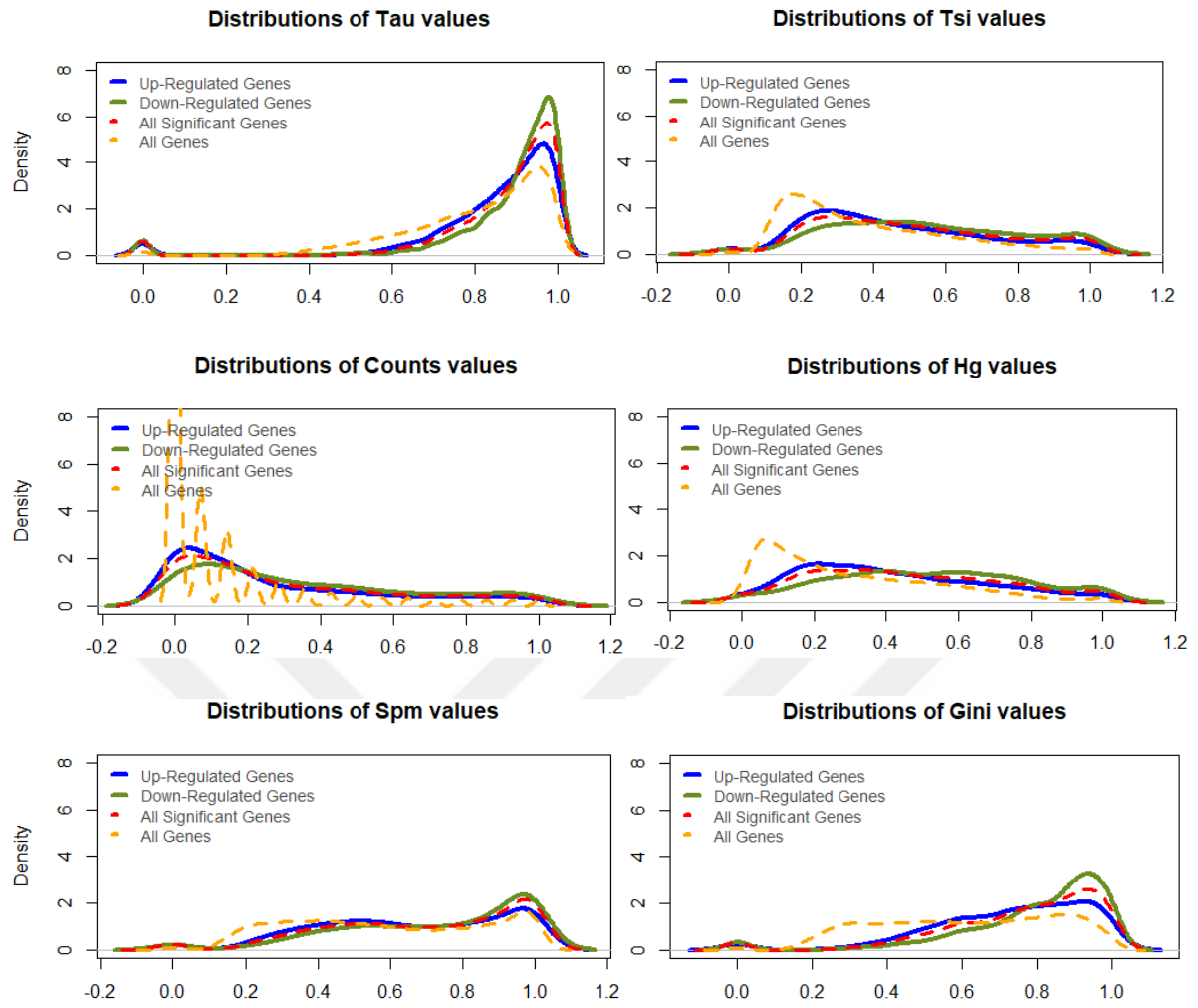


Figure 4.7: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were RPKM normalized.

For **Figure 4.7**, the distribution patterns for all the metrics were highly similar to those shown in **Figure 4.6** suggesting that SD and RPKM normalizations did not significantly change the tissue-specificity patterns.

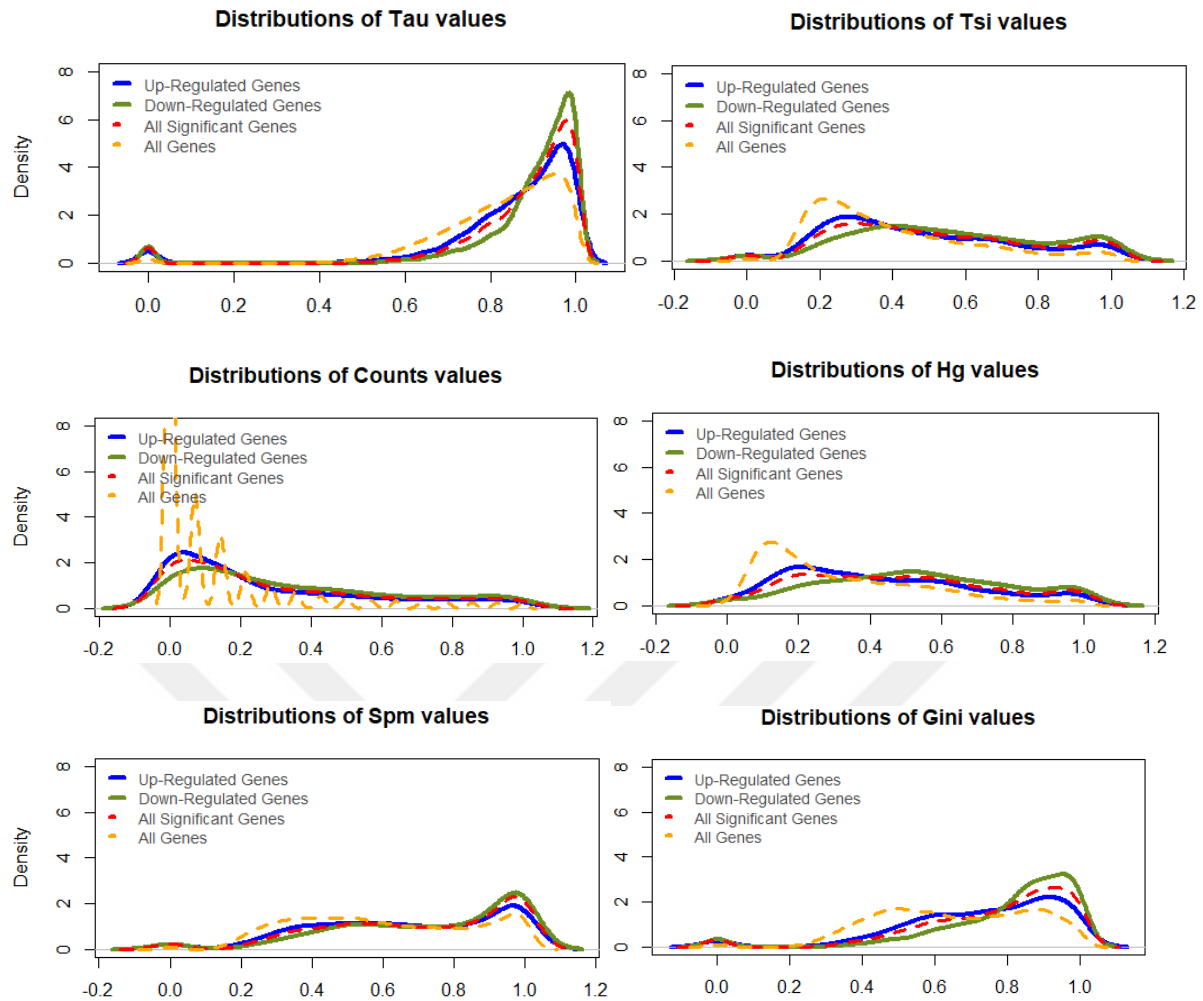


Figure 4.8: Distributions (KDE) of different Tissue Specificity Metrics values on AChE dataset. KDE for Tau, Tsi, Counts, Hg Spm and Gini. Tissues were TPM normalized.

For TPM normalized datasets (**Figure 4.8**) there were similarity to those obtained for SD and RPKM. The second peak of Tau distribution closer to 1 started before the point where the Tau of SD or RPKM normalized datasets were started (**Figure 4.8 – Tau plot**). Tsi, Hg, Spm, Gini and Counts distribution did not show any significant difference for TPM with respect to SD or RPKM.

4.3.2 How does the p-value or log2fc cut offs used to determine the tissue specificity change the distribution of TS genes of AChE dataset?

When a cut-off of $\text{Tau} > 0.8$ applied on AChE dataset DE list, in connection with an absolute log2 fold change greater than 0.5849, the number of genes found as tissue specific for different normalization methods was shown in **Table 4.4**.

4.3.2.1 Normalizations Compared

Before using any TS cut off, it was important to assign the genes to their tissues. To accomplish this, a gene was assigned to a tissue, if it had the highest averaged expression rank among the other tissues. Then, the following tables and figures were generated.

Table 4.3: The number of up and down regulated genes assigned to the tissues in which they had the highest expression value were shown in the table for different normalization methods applied.

Tau>0.8 with abs(l2fc)>0.5849			SD_NT	RPKM	TPM		SD_NT	RPKM	TPM
1	Brain	up	59	58	82	total	150	149	193
		down	91	91	111				
2	Liver	up	114	115	47	total	229	230	103
		down	115	115	56				
3	Skin	up	134	128	127	total	212	196	203
		down	78	68	76				
4	Gill	up	133	123	154	total	209	194	243
		down	76	71	89				
5	Heart	up	85	90	58	total	116	121	79
		down	31	31	21				
6	Intestine	up	137	134	84	total	267	263	204
		down	130	129	120				
7	Kidney	up	42	40	52	total	80	72	86
		down	38	32	34				
8	Muscle	up	118	103	38	total	189	173	64
		down	71	70	26				
9	Spleen	up	97	80	58	total	153	127	97
		down	56	47	39				
10	Bones	up	23	15	21	total	55	41	47
		down	32	26	26				
11	Embryos	up	64	64	90	total	119	121	158
		down	55	57	68				
12	Ufeggs	up	50	69	34	total	75	110	49
		down	25	41	15				
13	Ovary	up	79	93	133	total	150	172	240
		down	71	79	107				
14	Retina	up	47	38	48	total	188	170	185
		down	141	132	137				
15	Testis	up	240	295	431	total	437	528	761
		down	197	233	330				

Most of the tissue specific genes were found on Testis and Intestine. There was significant number of genes downregulated in retina, and up regulated in muscle. The number of genes highest expression found in given tissues did not differ for the

normalization method chosen, especially for RPKM and SD. However, in TPM, more testis specific genes measured compared to SD and RPKM as shown in **Table 4.3**. However at this point significant analyses have not been performed.

To test whether there is a difference between the distributions of up and down regulated genes ranked first in the tissue they were assigned, a Kolmogorov-Smirnov test was applied on log2fc based DE and TS genes (**Table 4.4**). There was a significant difference between the TS value distributions of intestine and retina up and down regulated genes for all the methods of normalizations tested. This pattern for intestine and retina was conserved when other DEG cutoffs (i.e pvalue <0.05 or 0.1) or other metrics were chosen (**Supplementary Table 1-3**). In addition to this, when the strict pvalue cutoffs were applied, there was a significant difference between muscle up and down regulated genes for all the cases. Moreover, heart and liver were two of the tissues that up and down regulated genes showed significant difference in terms of Tau value distributions when pvalue cutoff were applied to define DEGs, and this pattern was consistent for SD and RPKM (**Supplementary Table 2 and 3**).

Table 4.4: Kolmogorov Smirnov test applied on Tau distributions for up and down regulated genes in different tissues (DEG cut off were absolute log2foldChange>0.58 difference).

KS Test		SD		RPKM		TPM	
		D	pval	D	pval	D	pval
Tau	Brain	0.191919	0.108156	0.129556	0.502278	0.233833	0.004568
	Liver	0.235023	0.001186	0.249227	0.000482	0.10873	0.896117
	Skin	0.127569	0.276777	0.14298	0.200934	0.108006	0.503406
	Gill	0.113812	0.395046	0.145274	0.169788	0.16556	0.024351
	Heart	0.135684	0.64267	0.126615	0.682941	0.141107	0.797432
	Intestine	0.371309	7.01E-11	0.355997	1.26E-09	0.34386	1.16E-06
	Kidney	0.262876	0.048327	0.305738	0.021856	0.297673	0.026302
	Muscle	0.217545	0.010277	0.245998	0.004725	0.308547	0.067267
	Spleen	0.132143	0.336357	0.147059	0.278967	0.318129	0.003023
	Bone	0.342342	0.016846	0.431319	0.008768	0.36996	0.016143
	Ovary	0.15973	0.182952	0.073838	0.942409	0.104241	0.441802
	Embryos	0.160418	3.50E-01	0.15757	0.37347	0.179748	0.117034
	UFEggs	0.125683	0.832579	0.093718	0.923943	0.384146	0.048956
	Retina	0.375357	8.79E-06	0.376009	7.11E-05	0.436344	1.09E-07
	Testis	0.128302	0.041402	0.094575	0.162212	0.09284	0.06019

To see the difference better between the distribution and number of up and downregulated genes following graphs, where x axis referred to the value of Tau index and y-axis showed the logFC between ache mutant and healthy embryos, were generated for each tissue (**Figure 4.9-23**). Medians of the Tau distributions for these graphs were always in between 0.8 and 0.85 (**Supplementary Figure 4 and 5**).

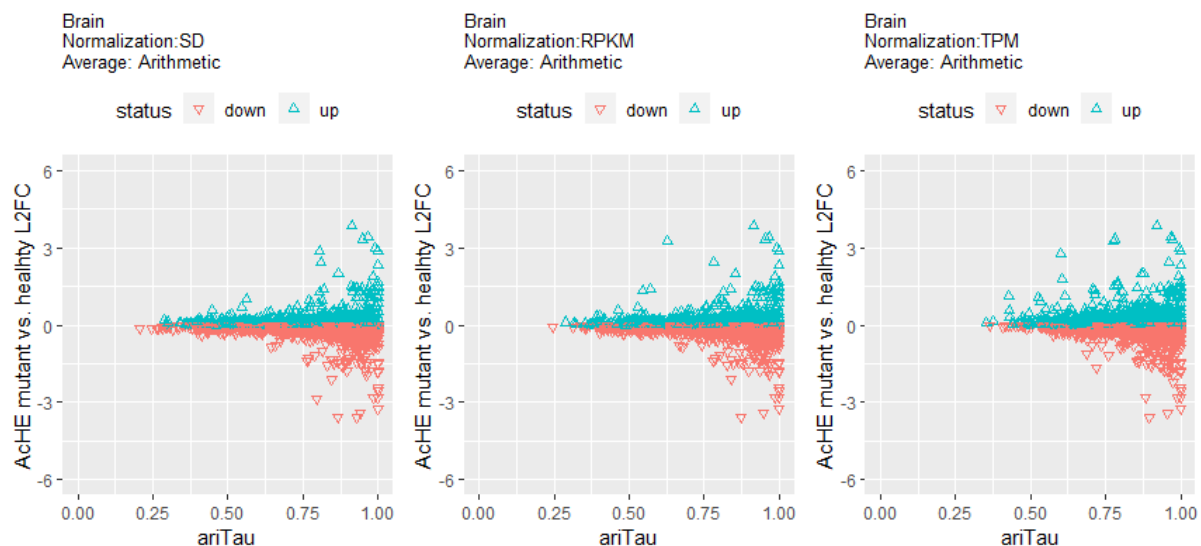


Figure 4.9: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in brain, for SD, RPKM or TPM normalized datasets, respectively.

For TPM, brain genes were accumulated in between 0.5 and 1 compared to SD or RPKM normalized data (**Figure 4.9**). SD and RPKM showed similar ranges for brain tissue.

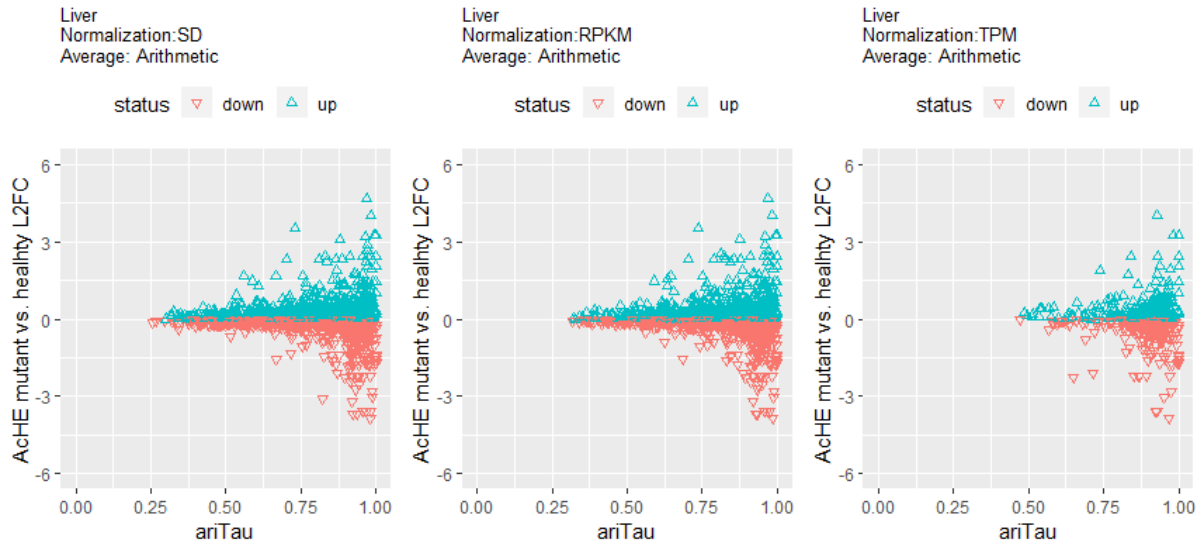


Figure 4.10: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in liver for SD, RPKM and TPM normalized datasets, respectively.

There were more liver specific genes with higher L2FC in RPKM and SD normalized data (**Figure 4.10**). Few numbers of TS genes were found in TPM normalized tissue dataset when compared with AChE dataset.



Figure 4.11: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in skin for SD, RPKM and TPM normalized datasets, respectively.

There a lot of up- and downregulated skin specific genes in AChE mutant embryos and most of them had Tau values greater than 0.9 (**Figure 4.11** and **Table 4.4**).



Figure 4.12: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in retina for SD, RPKM and TPM normalized datasets, respectively.

There were more downregulated retinal genes in AChE mutant zebrafish, and the genes having higher L2FC were generally found around 1 for Tau metric (**Figure 4.12**).

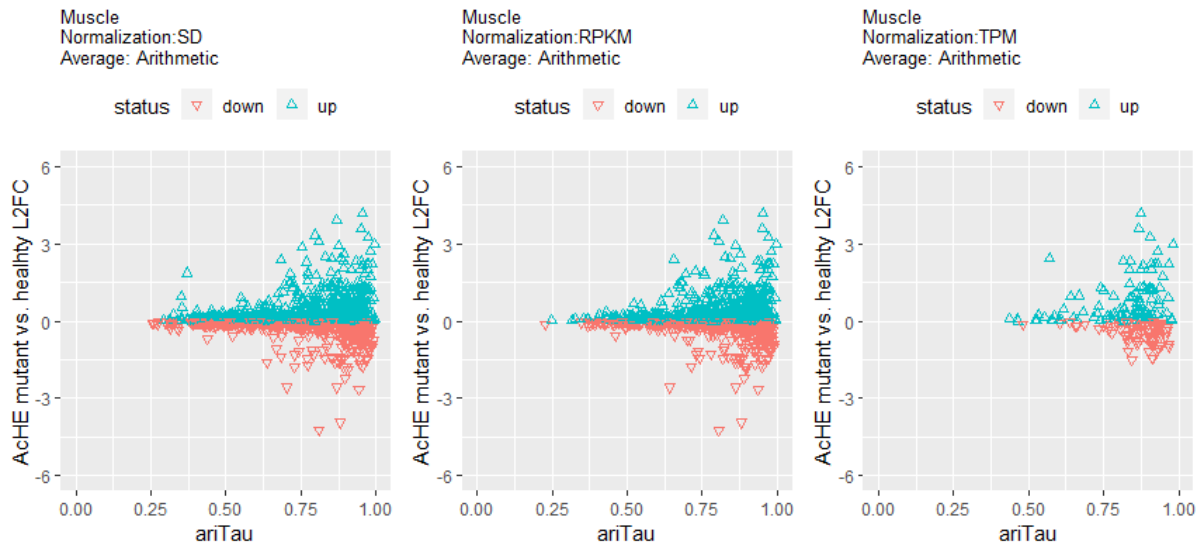


Figure 4.13: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in muscle for SD, RPKM and TPM normalized datasets, respectively.

As shown in **Figure 4.13**, many of the genes defined as muscle specific were upregulated genes in AChE mutant zebrafish embryos. In TPM, the amount of the muscle specific genes was smaller than SD or RPKM.

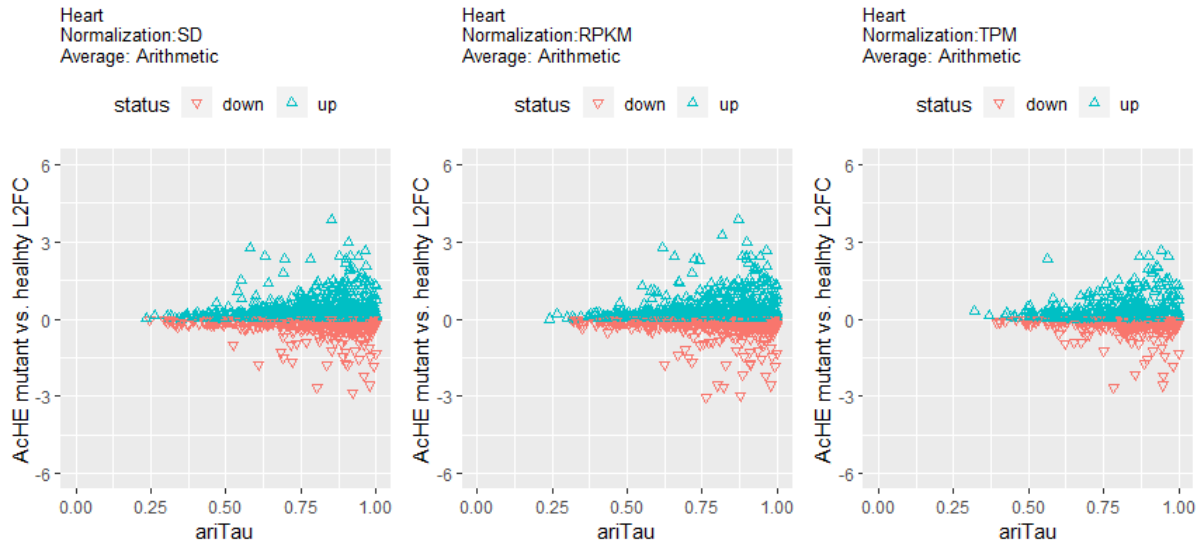


Figure 4.14: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in heart for SD, RPKM and TPM normalized datasets, respectively.

As in the case of heart specific genes, there were more upregulated genes for the heart in AChE mutant embryos (**Figure 4.14**). Tau values of TPM started around 0.4 when the genes assigned to heart.

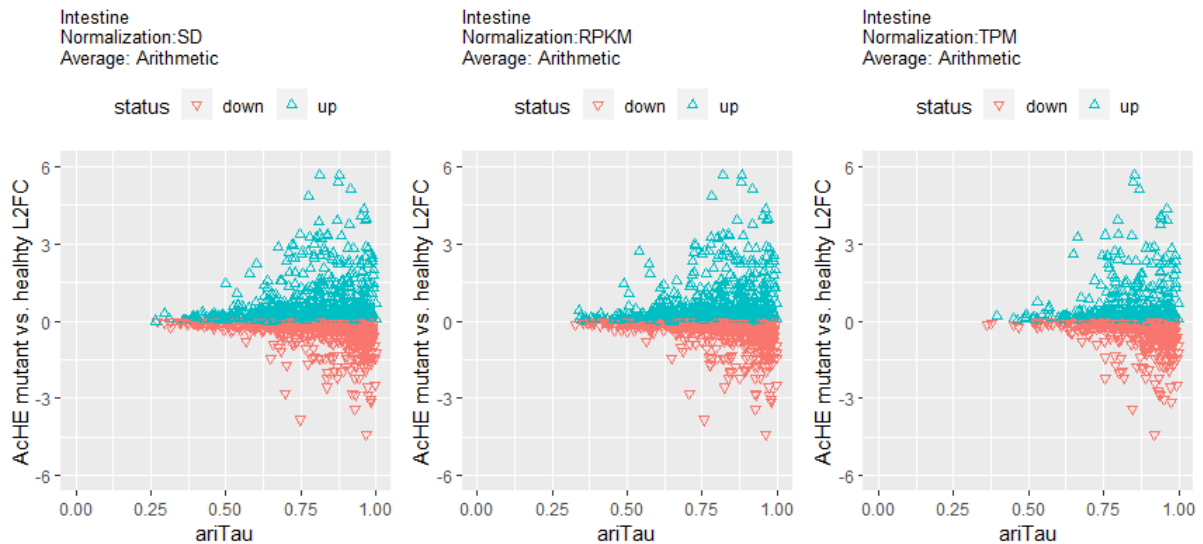


Figure 4.15: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in intestine for SD, RPKM and TPM normalized datasets, respectively.

A high number of genes defined as intestine specific were the upregulated genes with higher L2FC in AChE mutant zebrafish for all the normalization methods when Tau metric used, and these were accumulated around 1 (**Figure 4.15**).

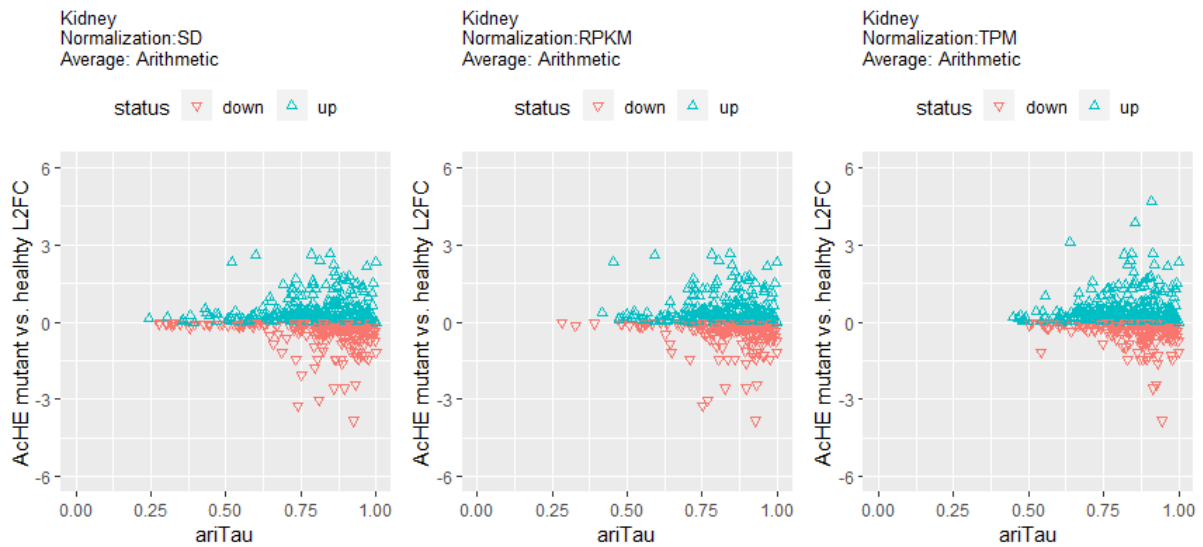


Figure 4.16: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in kidney for SD, RPKM and TPM normalized datasets, respectively.

There were relatively less number of kidney specific genes available for all methods when TS dataset applied on AChE dataset, and these were mostly the genes upregulated in ache mutants (**Figure 4.16 and Table 4.4**).

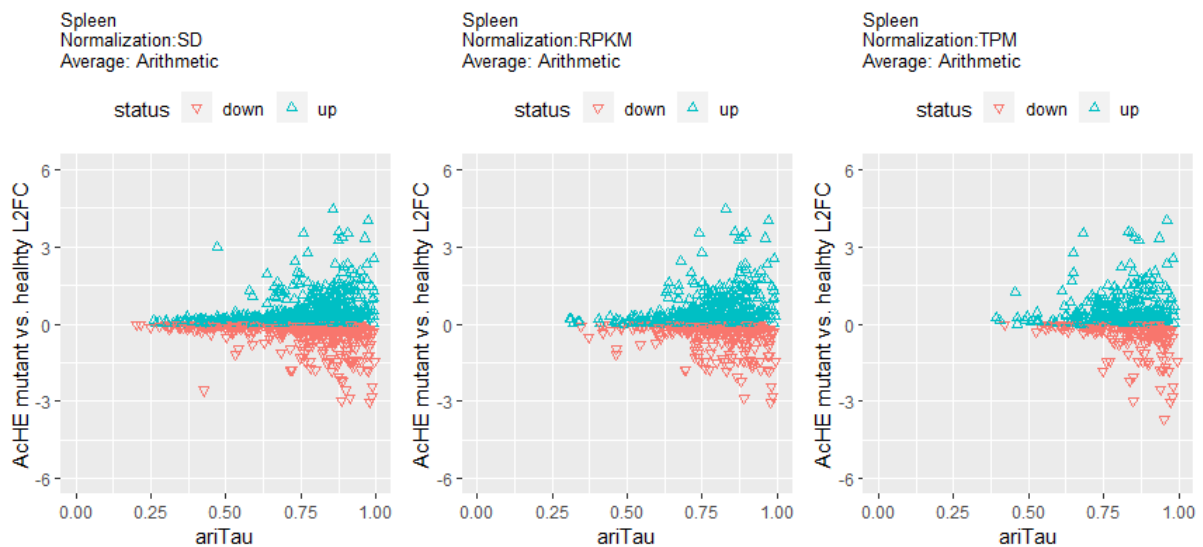


Figure 4.17: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in spleen for SD, RPKM and TPM normalized datasets, respectively.

Although there were more spleen specific genes in AChE dataset than kidney, they had similar distribution pattern for up and down- regulated genes (**Figure 4.17**).



Figure 4.18: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in bones for SD, RPKM and TPM normalized datasets, respectively.

The few number of genes defined as bone specific were both up and down regulated and seemed to be generally located close to 1 in AChE dataset as shown in **Figure 4.18**.



Figure 4.19: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in gill for SD, RPKM and TPM normalized datasets, respectively.

Gill specific genes had same distribution pattern with muscle or heart specific genes in AChE dataset **Figure 4.19**.

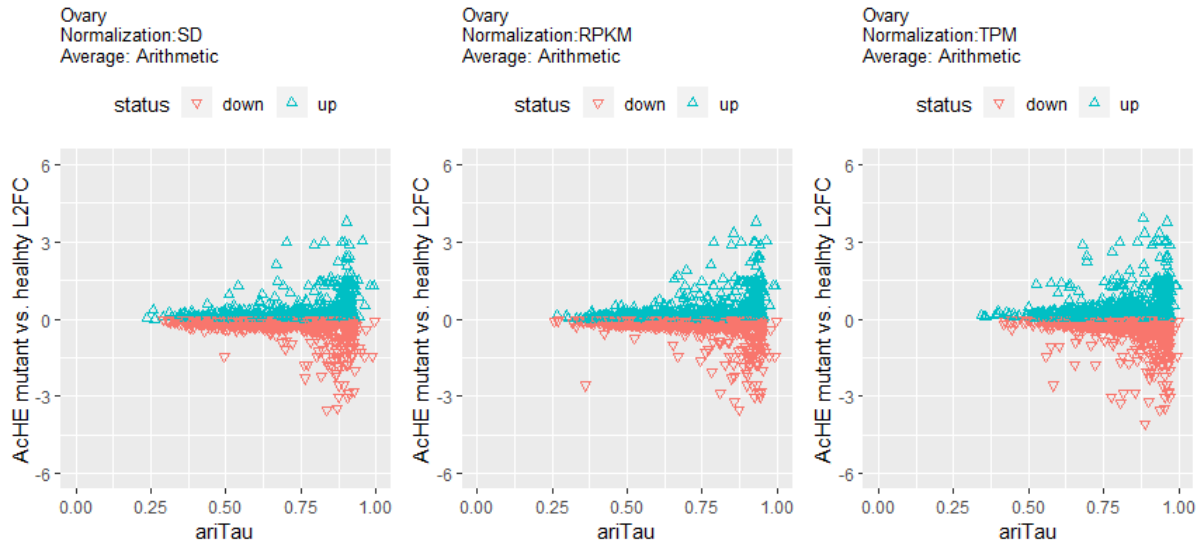


Figure 4.20: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in ovary for SD, RPKM and TPM normalized datasets, respectively.

The ovary specific genes having higher L2FC were around higher TS values for Tau metric as shown in **Figure 4.20**. All three normalizations showed similar patterns for Tau value distributions of the genes assigned to ovary.



Figure 4.21: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in embryos for SD, RPKM and TPM normalized datasets, respectively.

The distributions of up and down regulated embryos specific genes in AChE mutant were almost independent from the method chosen to normalize the datasets (**Figure 4.21**).

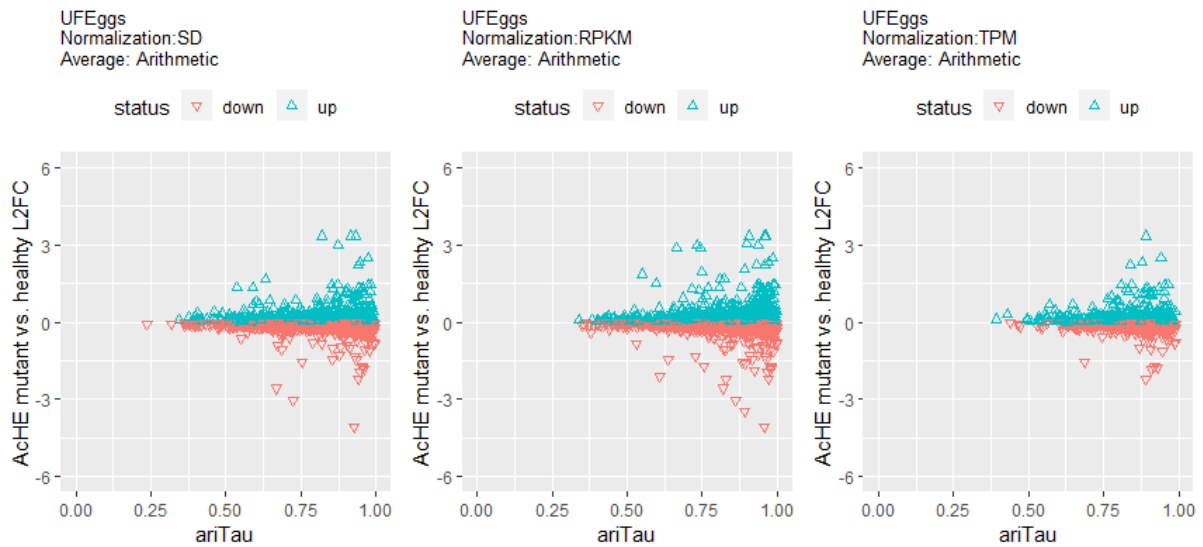


Figure 4.22: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in UFEggs for SD, RPKM and TPM normalized datasets, respectively.

Few unfertilized egg specific genes were found in AChE 3 dpf dataset and they were both up and downregulated and these were distributed between 0.5 and 1 in general. Fewer number of UFEgg genes found when TPM applied (**Figure 4.22**).

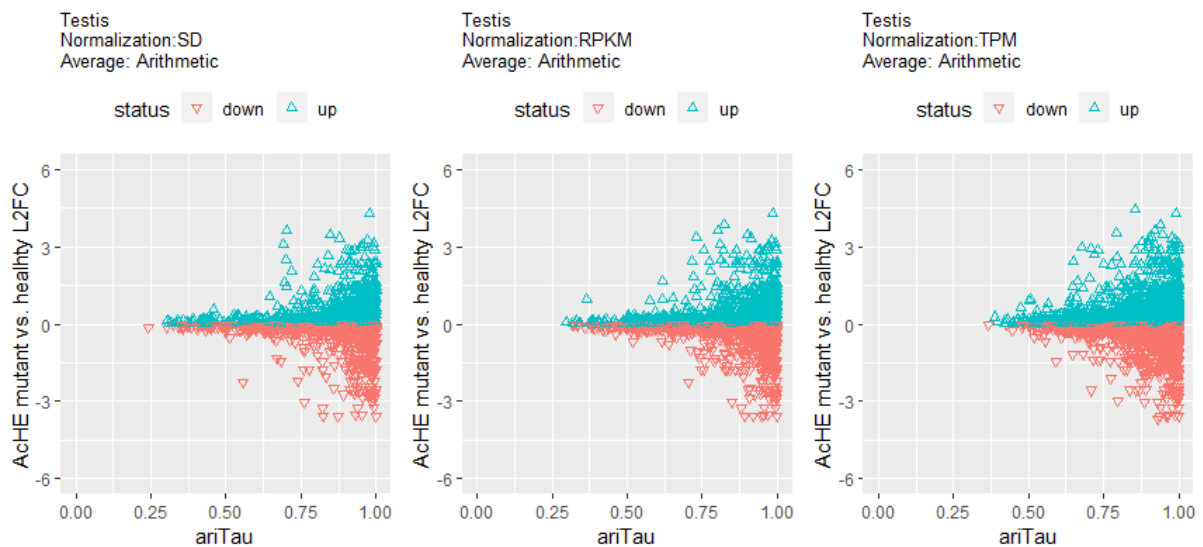


Figure 4.23: The distribution of the L2FC values obtained from Ache mutant vs. healthy comparison against the Tau values of the genes whose highest expression found in testis for SD, RPKM and TPM normalized datasets, respectively.

There were many testis specific genes, either up- or down-regulated in AChE mutant zebrafish. These accumulated around higher TS values and had higher L2FC independent of the normalization method used (**Figure 4.23**).

The analyses were performed with genes that have expression in the zebrafish embryos; therefore genes without counts were removed from AChE dataset. When the AChE dataset (n: 28577) was merged with the tissue dataset (n: 32520) accordingly, 1/3 of the genes were lost (32520-28577). These genes that were lost corresponded to the genes having Tau metrics smaller than 0.2 in general (**Supplementary Figure 3**). Therefore, the graphs shown in **Figure 4.9-23** started with Tau metric values greater than 0.2 and mostly accumulated between 0.8 and 1 (the latter being highest possible TS score for Tau). The genes having higher log2 fold change difference had the higher Tau TS values as well. The distributions of the counts were given in **Table 4.5** and in **Figures 4.9-23** visually.

To see whether DEGs for given threshold were enriched for TS genes, Fisher's Exact Test (FISHER, 1935) was applied in R for each of the tissues for a given status (either up or down). Odds ratios (Szumilas, 2010) were measured based on the contingency tables explained in Statistics part in the Methods section and were performed to reveal any association (at the given DEG cut off) between a higher L2FC or p-value in this case, and tissue specificity.

Table 4.5: Fishers Exact Test for up and down regulated genes. DEGs were determined by absolute log2 fold change greater than 0.5849 in AChE dataset.

TS>0.8 and abs(L2FC)>0.5849		SD		RPKM		TPM	
Tissue	Status	pvalue	OR	pvalue	OR	pvalue	OR
Brain	Up	1.15E-10	8.277934	9.89E-08	4.921774	4.81E-06	2.846666
	Down	3.57E-06	4.369312	3.32E-05	3.82748	7.01E-06	4.766231
Liver	Up	1.80E-15	4.628957	4.96E-13	4.151474	0.009922	3.1991
	Down	1.60E-21	12.71068	1.75E-16	9.691668	0.009832	4.413399
Skin	Up	3.03E-13	3.994749	3.42E-08	2.97385	4.13E-06	2.650065
	Down	2.72E-06	3.367166	0.001065	2.419698	0.030036	1.85241
Gill	Up	8.64E-07	2.486621	8.49E-05	2.117578	3.89E-06	2.009862
	Down	0.000278	2.642855	0.003245	2.306389	0.044172	1.611079
Heart	Up	8.93E-12	4.319536	3.01E-12	4.176277	3.59E-08	4.646938
	Down	0.000713	3.56475	0.00206	2.938529	0.307062	1.647819
Intestine	Up	1.48E-08	2.550167	1.40E-08	2.704063	0.013355	1.844947
	Down	5.92E-17	6.099273	3.15E-14	5.654318	9.13E-08	5.082473
Kidney	Up	0.117464	1.63272	0.182159	1.513878	0.001698	2.505263
	Down	0.013465	2.745448	0.174543	1.861127	0.001124	5.901023
Muscle	Up	1.83E-16	4.727524	5.29E-13	4.463794	0.004914	3.939329
	Down	5.20E-13	6.850223	4.37E-12	8.589372	0.003986	Inf
Spleen	Up	1.54E-13	4.514656	1.77E-07	3.283316	0.158101	1.458583
	Down	6.45E-08	4.141675	0.000402	2.895283	0.000552	4.436843
Bones	Up	1	1.074749	0.215792	0.526194	0.828662	0.906769
	Down	0.186078	2.029403	0.026383	5.180979	0.432417	1.723868
Ovary	Up	1.86E-27	14.4377	4.01E-27	10.48609	8.21E-24	7.116584
	Down	2.88E-38	33.87344	5.20E-30	16.85981	6.27E-15	6.021303
Embryos	Up	5.18E-06	4.145513	2.13E-06	5.136484	1.06E-06	3.437855
	Down	8.71E-15	12.58584	8.68E-13	10.50034	2.26E-11	8.777144
Ufeggs	Up	0.000858	2.938403	0.023907	1.916233	0.074904	2.223852
	Down	0.264224	1.686168	0.057852	2.30298	0.135745	4.643397
Retina	Up	1.97E-05	3.576329	0.018002	2.304186	0.047627	1.86134
	Down	3.98E-10	4.671768	1.69E-06	4.899688	2.03E-06	4.54673
Testis	Up	3.07E-13	4.361341	9.90E-24	6.799826	3.49E-35	5.423936
	Down	2.98E-22	8.808446	1.93E-23	7.510688	6.59E-34	8.100156

Odds ratio equal to 1 refers to no association between being a DEG and being TS at the same time, whereas OR>1 and OR<1 shows positive and negative association between the given conditions, respectively.

As shown in **Table 4.5**, there is an enrichment of both up and down regulated DEGs, determined by a cut-off value based on absolute log2 fold change greater than 0.5849, in a tissue-dependent manner. Among the 15 tissues studied, downregulated DEGs were enriched with TS genes in all tissues except bone and

UF eggs based on SD normalized tissue dataset. Upregulated DEGs were enriched with TS genes in all tissues except kidney and bone. There were a few discrepancies between normalization methods, e.g., upregulated spleen genes being significant in SD and RPKM but not in TPM; upregulated genes in kidney being significant in TPM but not in RPKM and SD. When L2FC cut off was applied, the most enriched tissues were liver, retina and muscle- down regulated, and ovary both up and down regulated genes for zebrafish ache mutant embryos. The OR was not same for the different normalization methods.

When DEG cut off was chosen as p-value smaller than 0.1, most of the tissues lost their enrichment and the enrichment analysis became more specific as shown in **Table 4.6**. Although the downregulated genes in ache mutant embryos were enriched in tissues like retina (consistent for all normalizations), liver down, intestine down and embryos, and the most significant enrichments were found in especially in retina (downregulated), muscle (downregulated), heart, skin and liver. This was not always the case when TPM used. Interestingly, the enrichments specific to tissues were shown for downregulated genes than upregulated genes. On the other hand, this pattern was more pronounced in DEGs with $p\text{-value} < 0.05$ for, intestine, muscle, skin, liver, and especially for retina (**Supplementary Table 3**). In TPM, these enrichments were more apparent than SD and RPKM for especially in retina. In contrast to L2FC based DEG used TS genes, testis were not enriched when p-value cut offs were used (**Table 4.5 and 4.6**).

Table 4.6: Fishers Exact Test for up and down regulated genes. DEGs were determined by differentially expressed genes in AChE having p value smaller than 0.1.

TS>0.8 and pval<0.1		SD		RPKM		TPM	
Tissue	Status	pvalue	OR	pvalue	OR	pvalue	OR
Brain	Up	0.722857	1.184819	0.865427	0.933223	0.61958	0.883658
	Down	0.024879	2.167608	0.035669	2.036182	0.000265	4.755405
Liver	Up	0.020932	1.549735	0.043084	1.480421	0.697403	0.849158
	Down	9.30E-06	4.178513	3.80E-05	4.487846	0.139768	5.040937
Skin	Up	1	0.991187	0.6129	0.891377	0.285437	0.79115
	Down	0.005268	4.008538	0.003841	4.838188	0.091646	2.553173
Gill	Up	0.314592	0.797937	0.201912	0.734827	0.317298	0.825004
	Down	1	0.921533	1	1.249294	0.581076	0.724774
Heart	Up	0.004258	1.837971	0.001339	1.900657	0.001929	2.204816
	Down	8.69E-06	21.44228	4.62E-06	13.10563	0.002536	5.456398
Intestine	Up	0.483595	0.867202	0.669653	0.90103	0.121425	0.619734
	Down	0.000126	3.291001	0.000943	2.828748	0.000786	4.808647
Kidney	Up	0.001949	0.305676	0.005057	0.313558	0.030363	0.463972
	Down	0.666397	2.474403	0.63179	0.617344	0.107679	Inf
Muscle	Up	6.35E-13	3.614729	2.96E-10	3.538998	0.173637	2.127947
	Down	5.44E-19	42.915	4.08E-12	14.93883	0.003604	Inf
Spleen	Up	0.261167	1.400081	0.878592	1.085769	0.612948	1.238671
	Down	0.727796	1.355686	1	1	0.052649	Inf
Bones	Up	0.311089	0.489343	0.67541	0.601324	0.664774	0.802386
	Down	0.230905	2.817854	0.724668	1.631966	0.107368	5.448275
Ovary	Up	0.788004	1.234808	0.552378	0.704576	0.20682	0.760306
	Down	0.804332	1.141668	0.420164	1.372117	0.030958	1.948542
Embryos	Up	6.00E-05	4.018599	0.017331	2.02019	0.047394	1.52368
	Down	0.002002	3.18231	0.001866	3.513045	0.005262	2.790352
Ufeggs	Up	0.483301	1.262473	0.279312	0.770924	0.391916	0.740617
	Down	0.267238	1.937745	0.300243	1.731003	0.047977	Inf
Retina	Up	1	0.988655	1	1.034012	0.06351	0.571165
	Down	1.82E-12	9.224786	9.45E-06	5.510439	3.67E-10	31.04726
Testis	Up	1	1.249701	0.016625	0.42273	7.52E-07	0.372184
	Down	0.001782	0.064927	0.00082	0.131789	0.362874	0.701791

4.3.2.2 How similar were the genes found as tissue specific across different normalization methods? Does the log transformation have any effect on TS measurement?

To answer the questions above, the following analysis were done: “Does it matter to use different normalization method to decide on TS genes?”, and “Does it matter to use log transformed SD, RPKM, TPM normalized counts or not?” Even if the overall

TS genes were not different for SD, RPKM or TPM for Tau or Tsi (**Figure 4.42**), when the graphs were drawn one by one for each tissue, it became much easier to see the difference between TS genes when different methods applied.

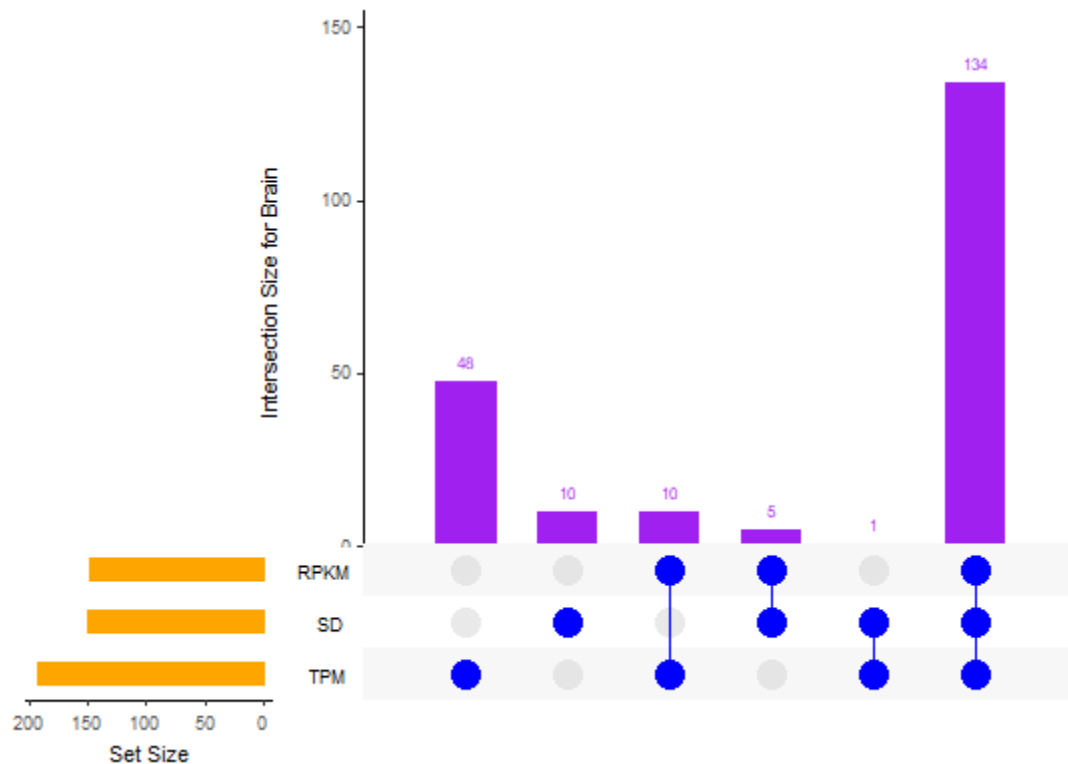


Figure 4.24: Upset graph showing the intersection between genes specific to Brain across differently normalized datasets.

For the brain, more than half of SD and RPKM normalized genes intersected and log transformation had no effect on these genes. On the other hand, there were genes (48) that assigned brain specific in TPM but neither in RPKM or SD, and the genes found only in SD (10) (**Figure 4.24**).

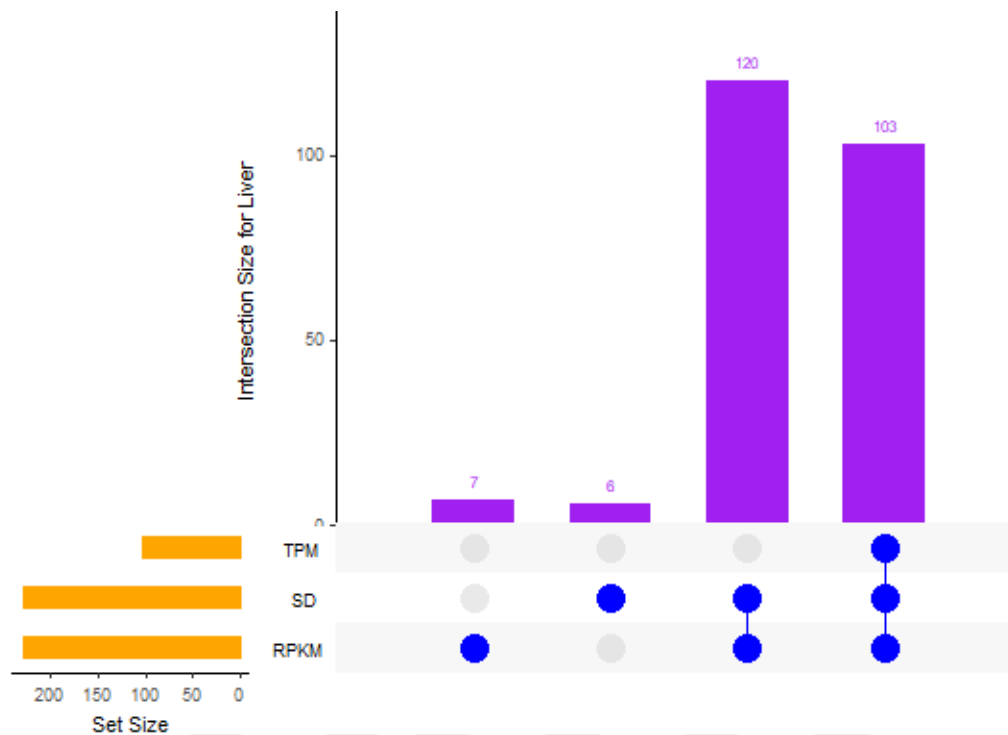


Figure 4.25: Upset graph showing the intersection between genes specific to Liver across differently normalized datasets.

In contrast to brain, liver specific genes were more sensitive to the normalization method applied before any metric calculation for TS, which was shown in **Figure 4.25**. SD and RPKM shared almost half of the overall liver specific genes assigned.

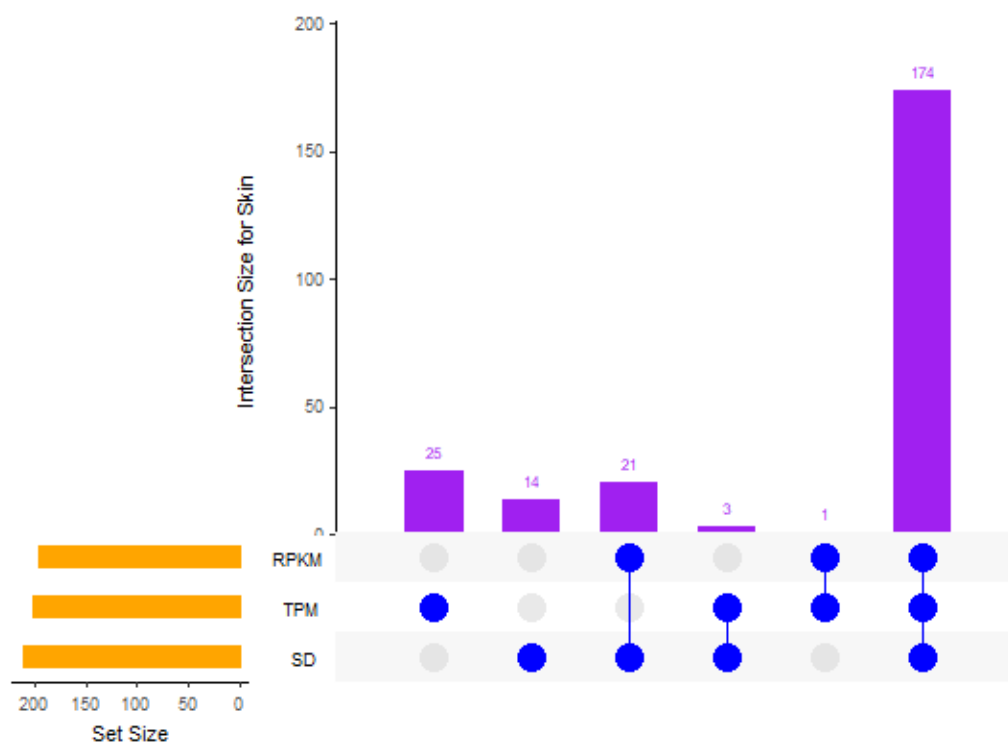


Figure 4.26: Upset graph showing the intersection between genes specific to skin across differently normalized datasets.

Skin resembles brain that more than two third of the genes defined as TS were similar regardless of the metric applied (**Figure 4.26**).

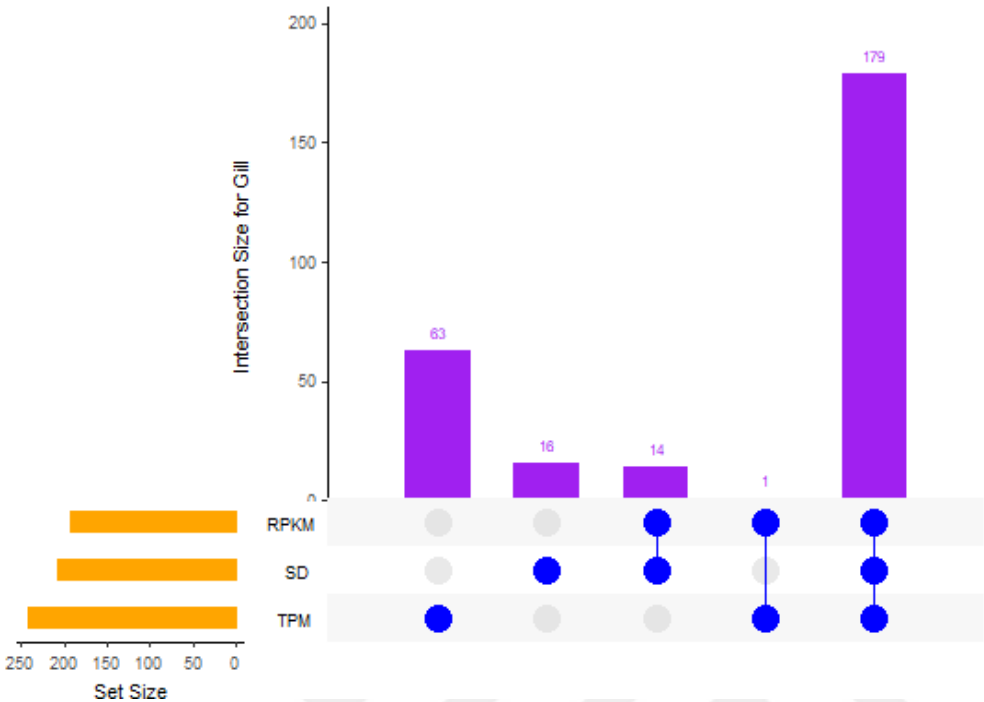


Figure 4.27: Upset graph showing the intersection between genes specific to gill across differently normalized datasets.

As in the case of skin, more than half of the overall gill specific genes available for the normalizations applied were the same among the methods (**Figure 4.27**).

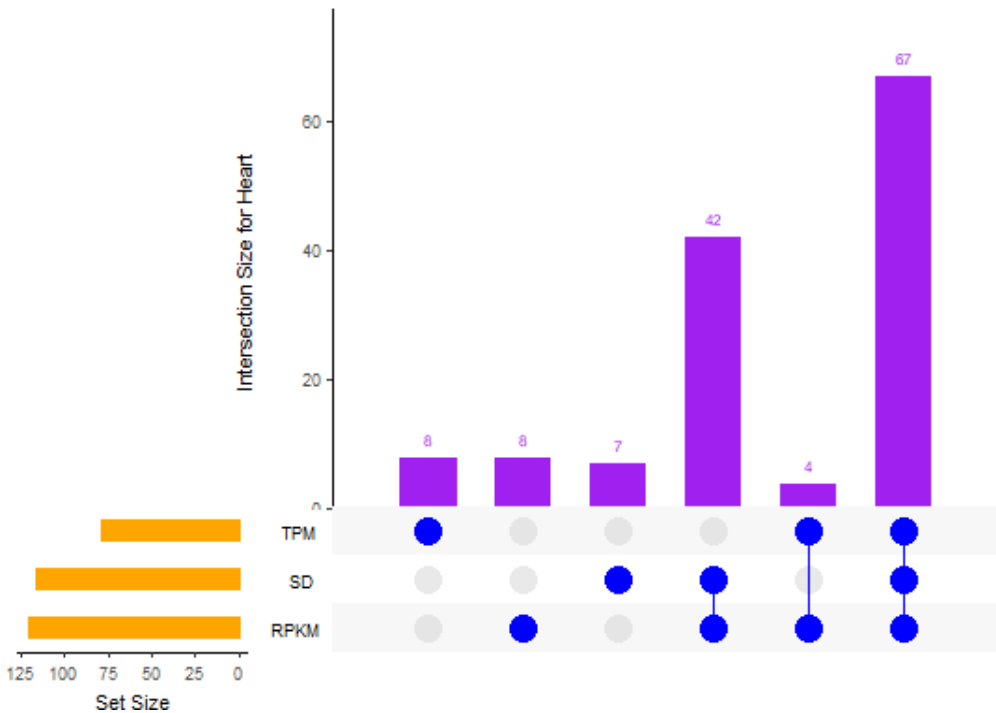


Figure 4.28: Upset graph showing the intersection between genes specific to heart across differently normalized datasets.

Heart tissue was more like liver, and more than half of TS genes overlapped independent on the method used (**Figure 4.28**). However, RPKM and SD together shared almost all the the genes defined as TS.

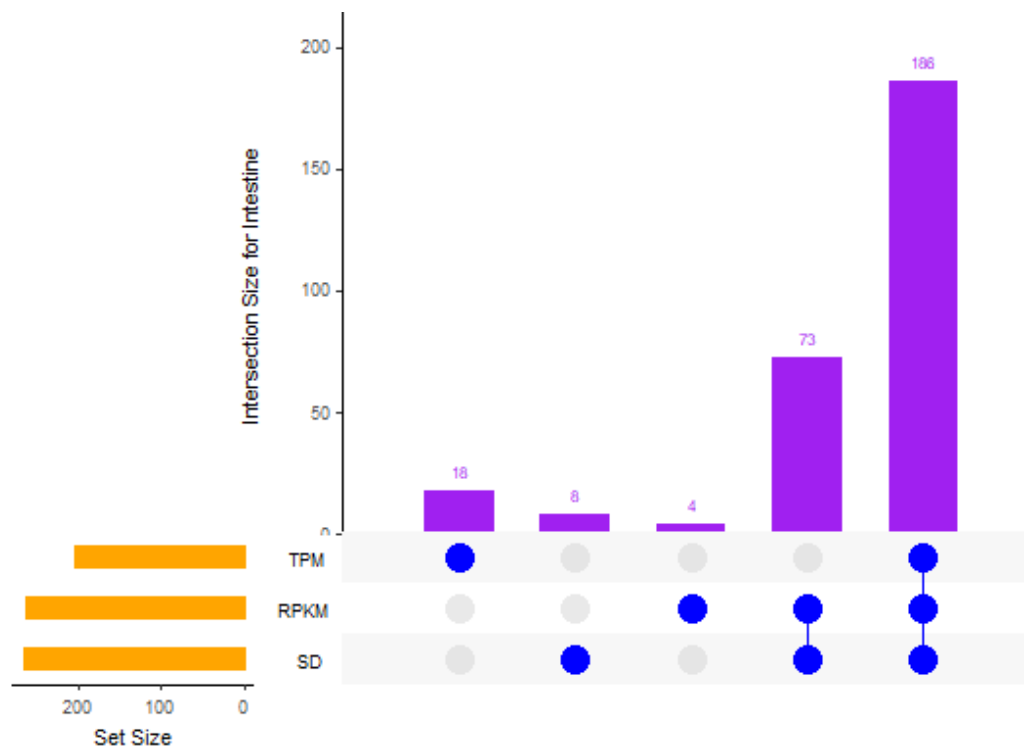


Figure 4.29: Upset graph showing the intersection between genes specific to intestine across differently normalized datasets.

For the intestine specificity, more than $\frac{3}{4}$ of TS genes intersected across all three methods (**Figure 4.29**). RPKM and SD were more similar compared to the TPM TS genes.

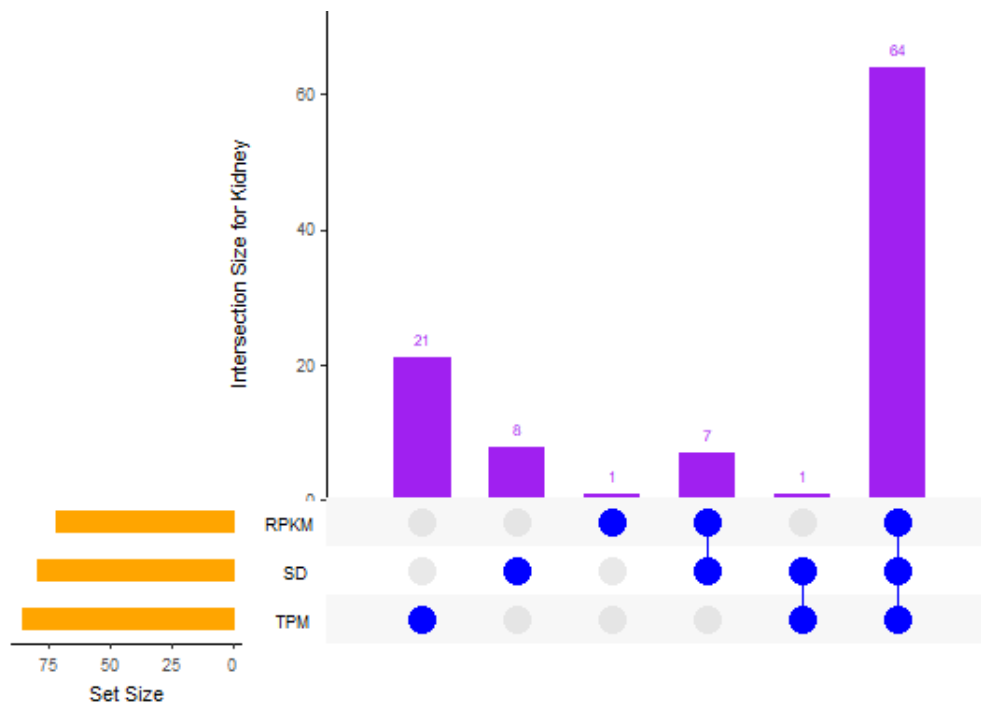


Figure 4.30: Upset graph showing the intersection between genes specific to kidney across differently normalized datasets.

TPM had highest number of TS genes available for Kidney as shown in **Figure 4.30**. More than half of the genes were shared among different normalization methods applied in the case of non-transformed normalized counts.

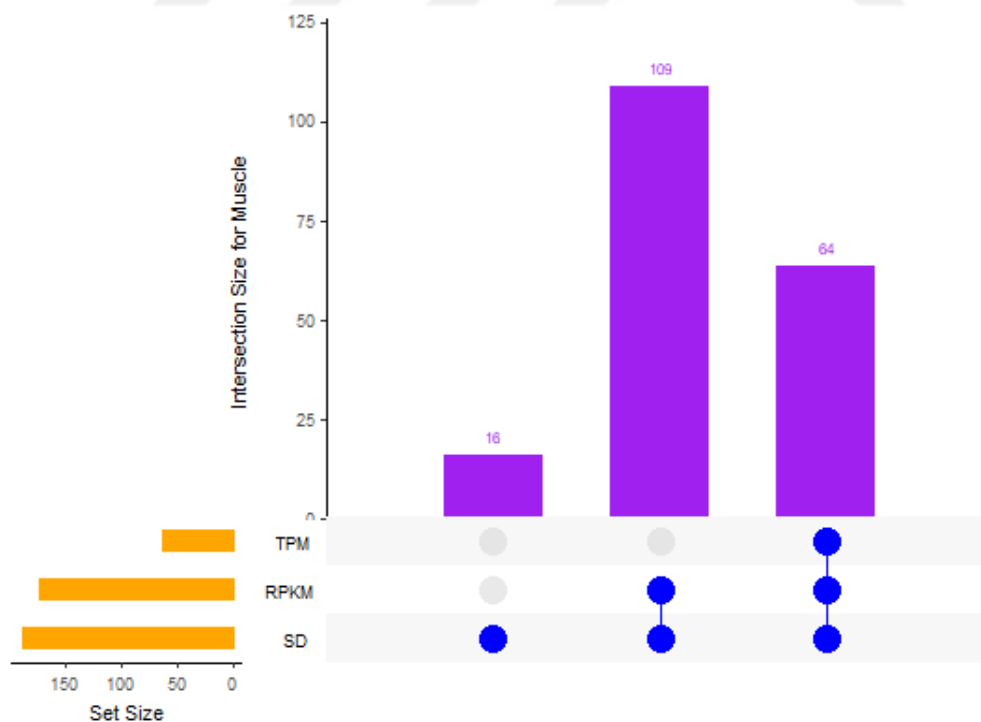


Figure 4.31: Upset graph showing the intersection between genes specific to muscle across differently normalized datasets.

SD and RPKM were more consistent with each other to define TS genes for Muscle (**Figure 4.31**).

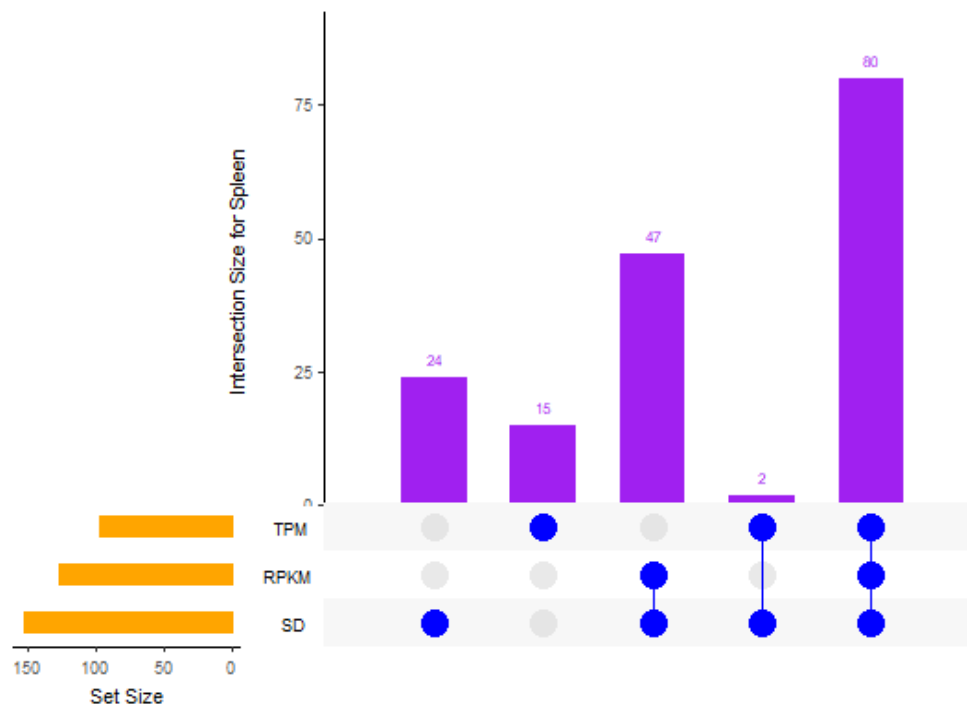


Figure 4.32: Upset graph showing the intersection between genes specific to spleen across differently normalized datasets.

As shown in **Figure 4.32**, at least half of the genes were found as Spleen specific were common TS genes for SD, RPKM and TPM. There were more TS genes identified when SD normalization performed before TS metric calculation. SD and RPKM seemed to be more consistent with each other.

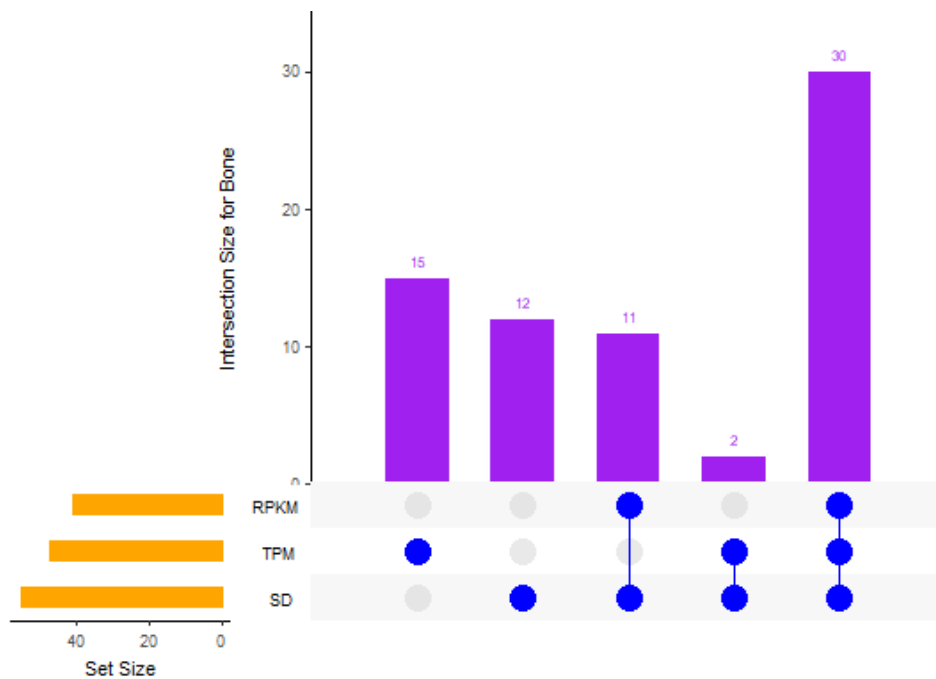


Figure 4.33: Upset graph showing the intersection between genes specific to Bones across differently normalized datasets.

For the bone expressed genes, the numbers of TS genes were smaller relative to other tissues. However, highest intersection size was found for SD, RPKM and TPM together (**Figure 4.33 and Supplementary Figure 15.**). TPM and SD had different bone specific genes.

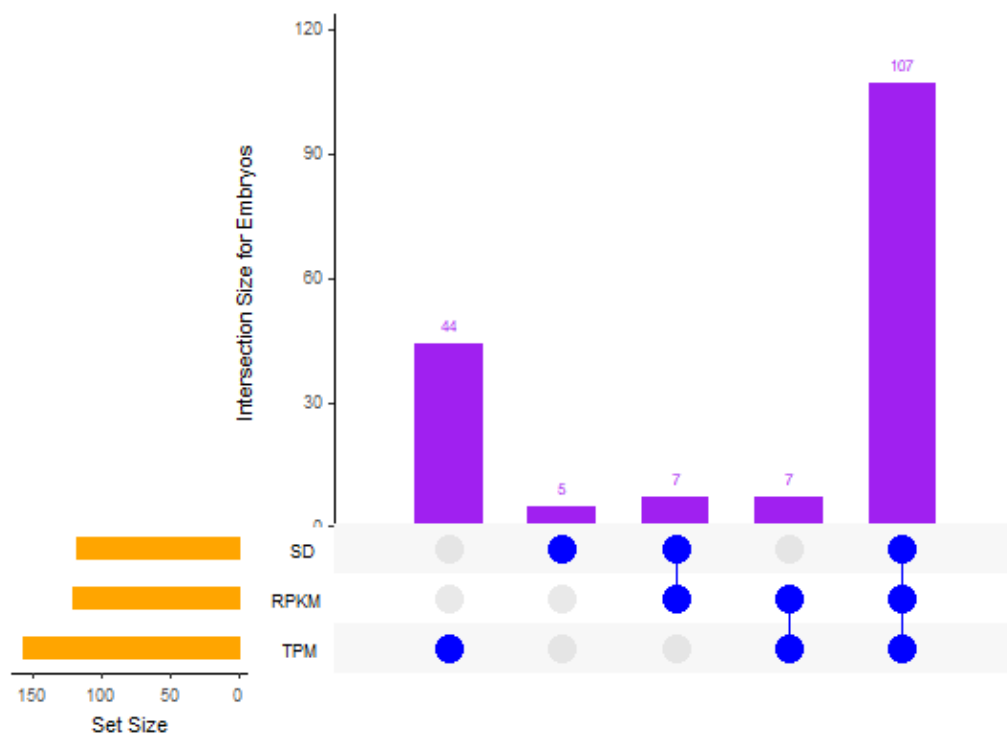


Figure 4.34: Upset graph showing the intersection between genes specific to embryos across differently normalized datasets.

As shown in **Figure 4.34.**, most of the genes defined as embryos specific were defined independent of methods. TPM behaved differently with respect to SD and RPKM.

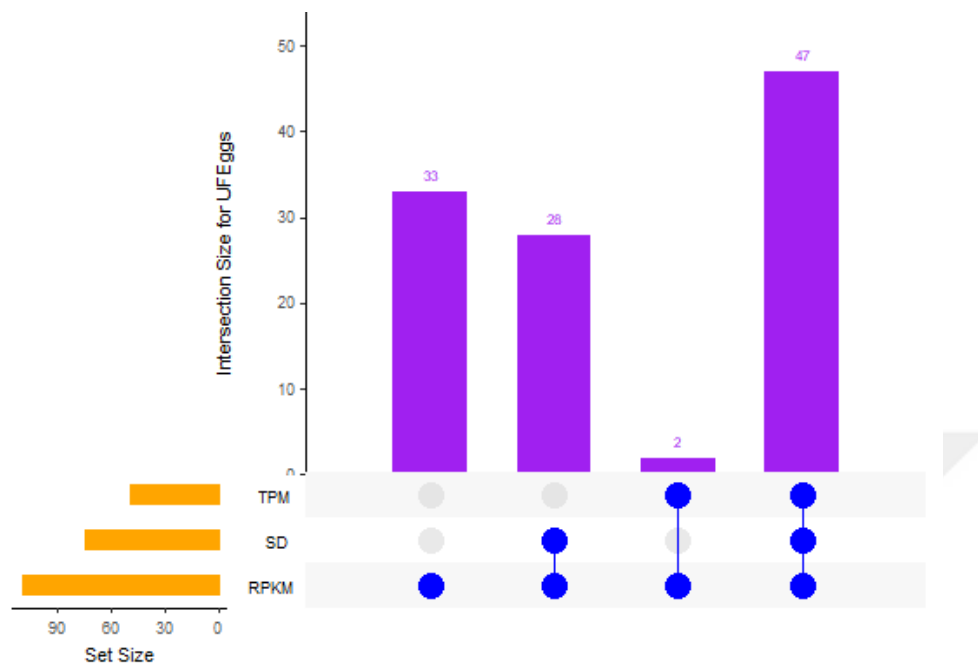


Figure 4.35: Upset graph showing the intersection between genes specific to UFEggs across differently normalized datasets.

High ratio of genes was found as unfertilized egg specific in RPKM but not in two other methods (**Figure 4.35**).

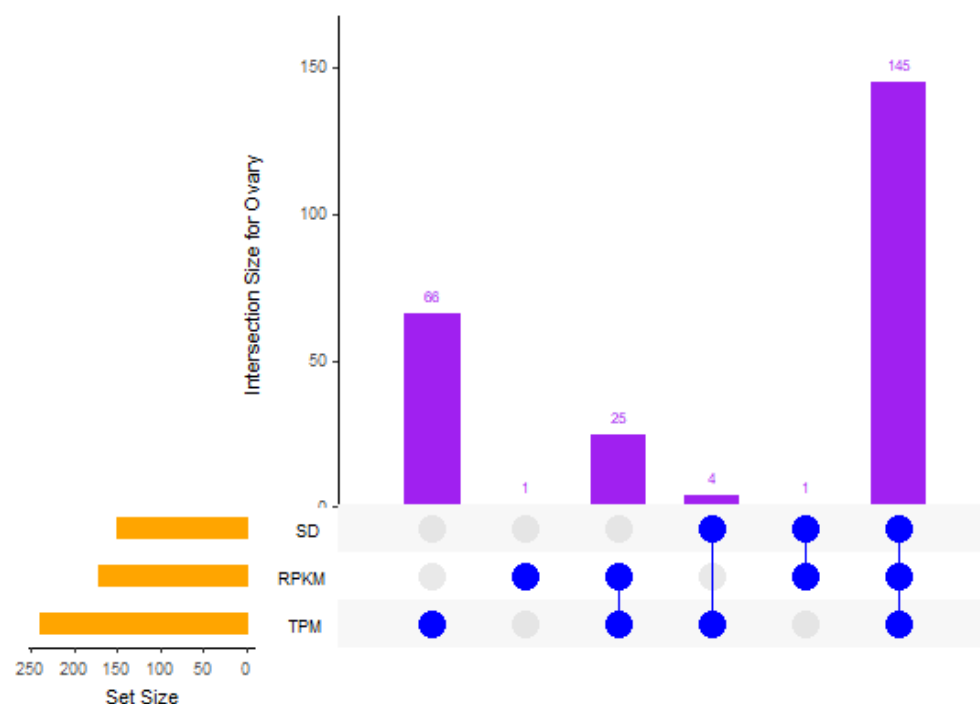


Figure 4.36: Upset graph showing the intersection between genes specific to ovary across differently normalized datasets.

For the ovary, there were more than a hundred of TS genes available for all the methods applied (**Figure 4.36**). High number of genes was defined as ovary specific for SD, RPKM and TPM.

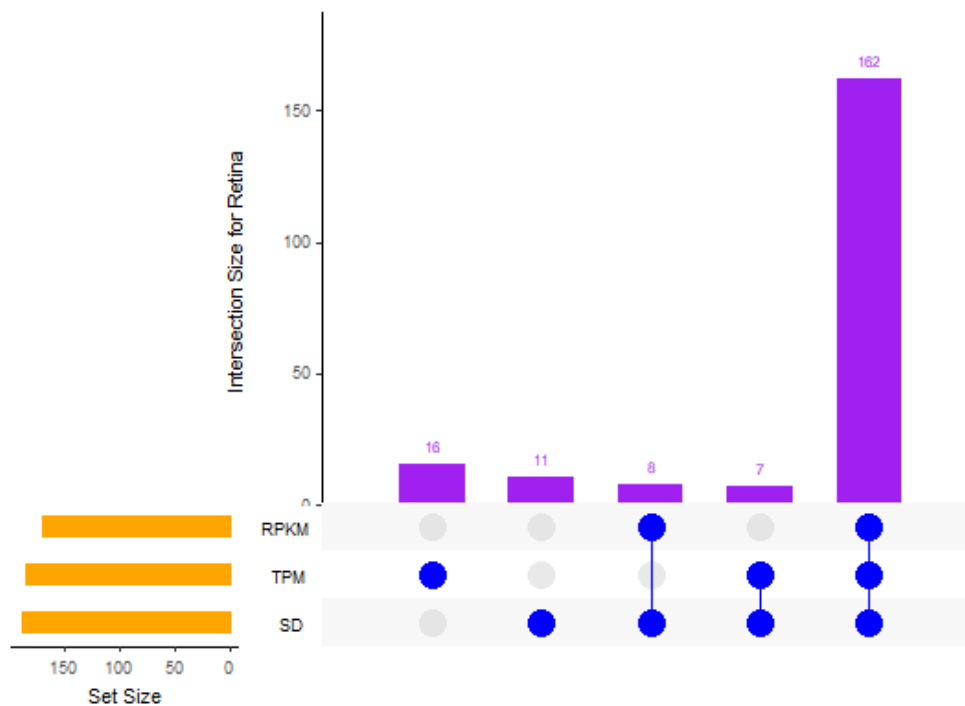


Figure 4.37: Upset graph showing the intersection between genes specific to Retina across differently normalized datasets.

Interestingly, retina specific gene definition was not dependent on the normalization method. Almost all the genes defined in all the methods were the same (**Figure 4.37**).

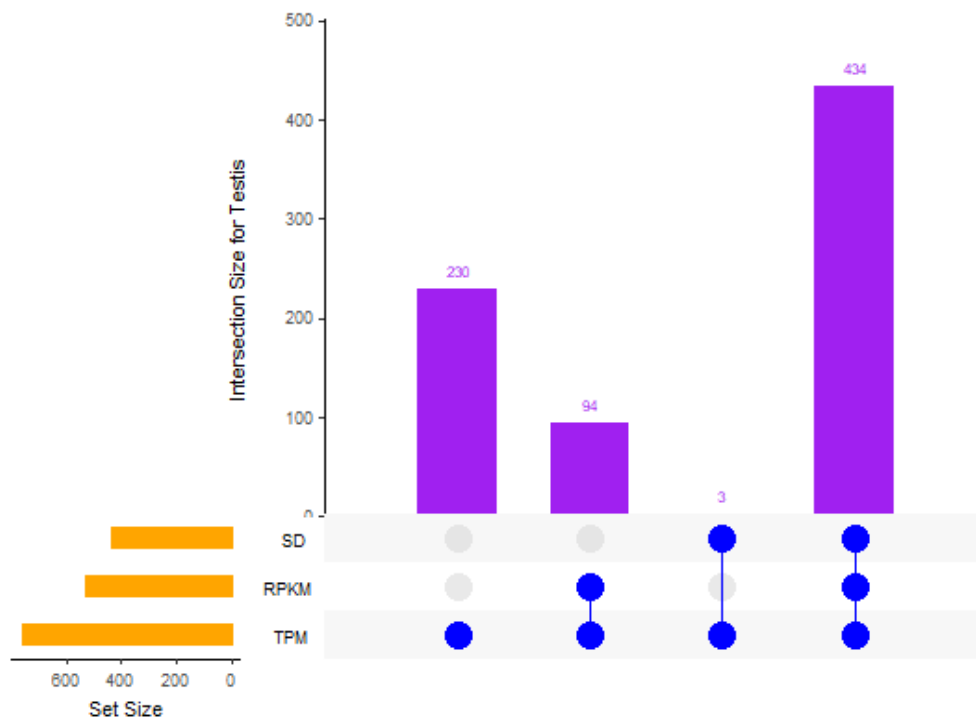


Figure 4.38: Upset graph showing the intersection between genes specific to Testis across differently normalized datasets.

The highest number of DE-TS genes found in testis for the selected lineant DEG cut-off (**Figure 4.38**). Most of the genes found as testis specific intersected with each other regardless of the normalization method chosen. However TPM identified a unique set of TS genes.

What about log transformation?

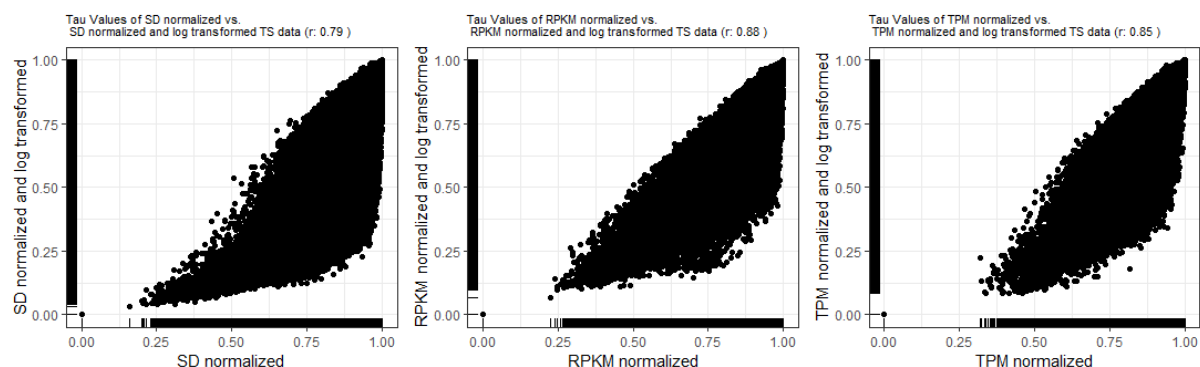


Figure 4.39: Comparison of the distribution of Tau values on differently normalized vs. differently normalized and log2 tranformed values.

Just after the normalization steps, to see whether there was any effect of the log2 transformation on the determination of TS values, Tau value values obtained from unlogged and logged data were compared. When the data were log transformed after normalizations, Tau values corresponding to each gene became more

conservative in comparison with the non-transformed TS dataset (**Figure 4.39**). On the other hand, the correlation between Tau values before and after log2 tranformation were higher in RPKM ($r=0.88$) and TPM ($r=0.85$) than SD ($r=0.79$).

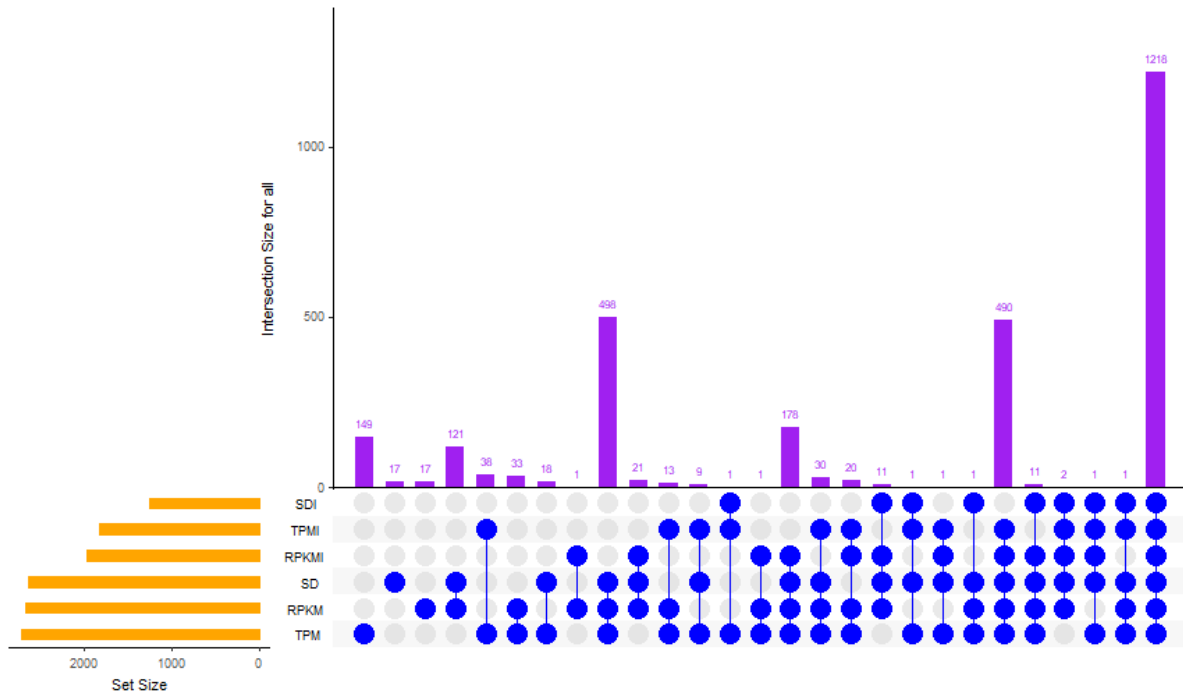


Figure 4.40: Overall TS genes intersected in different normalization methods and log transformed (represented with additional “l” letter) vs. non transformed normalized counts. DE genes selected based on $abs(L2FC) > 0.5849$ for ache dataset, and $Tau > 0.8$ to define TS genes.

As shown in **Figure 4.40**, the numbers of TS genes were decreased when the data were log transformed after the normalization to define TS genes for al the normalization methods.

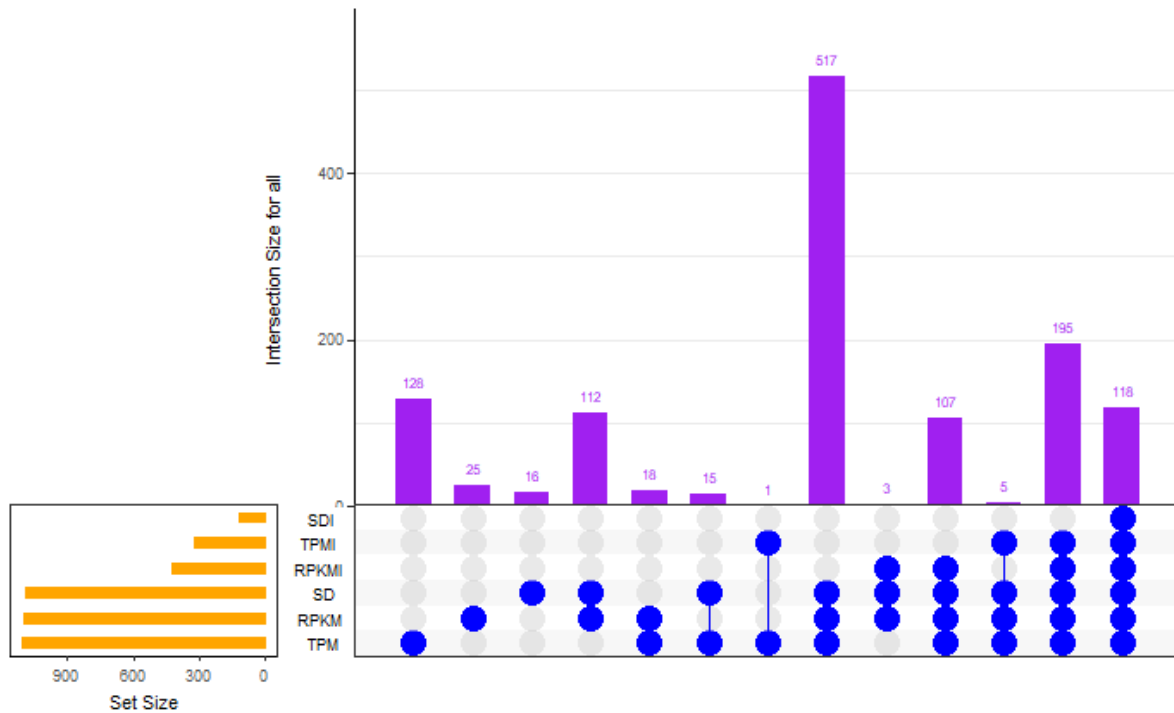


Figure 4.41: Overall TS genes intersected in different normalization methods and log transformed (represented with additional “l” letter) vs. non transformed normalized counts. DE genes selected based on $p\text{-value} < 0.1$ for ache dataset, and $\text{Tau} > 0.8$ to define TS genes.

If the DE cut-off was changed to pvalue cut off smaller than 0.1 or 0.05, as shown in **Figure 4.41** and **Supplementary Figure 54**, the number of TS genes were drastically decreased when the data were log transformed after the normalization to define TS genes for all the normalization methods. Less number of TS-DE genes were found if more strict p-value cut-off (0.05) was chosen with respect to p-value cut-off 0.1 and absolute log2fold change cut-off greater than 0.5849.

If the tissue specificity metrics applied on normalized or normalized-log-transformed dataset, there could be tissue-specific differences arising. To measure how much this affects the result, upset graphs were generated showing the intersected genes between different normalization methods and different applications on them on the transformed vs. non-transformed dataset as shown in **Figure 4.24-38** (**Supplementary Figures 6-20 and 54**). Most of the TS genes intersected for every method included to be retinal, testis, embryo, bone, kidney, intestine and brain specific. However, the application of TS metric on log transformed or non-transformed values had a significant effect on the result. In addition, most of the genes were intersected for RPKM and SD but not in TPM.

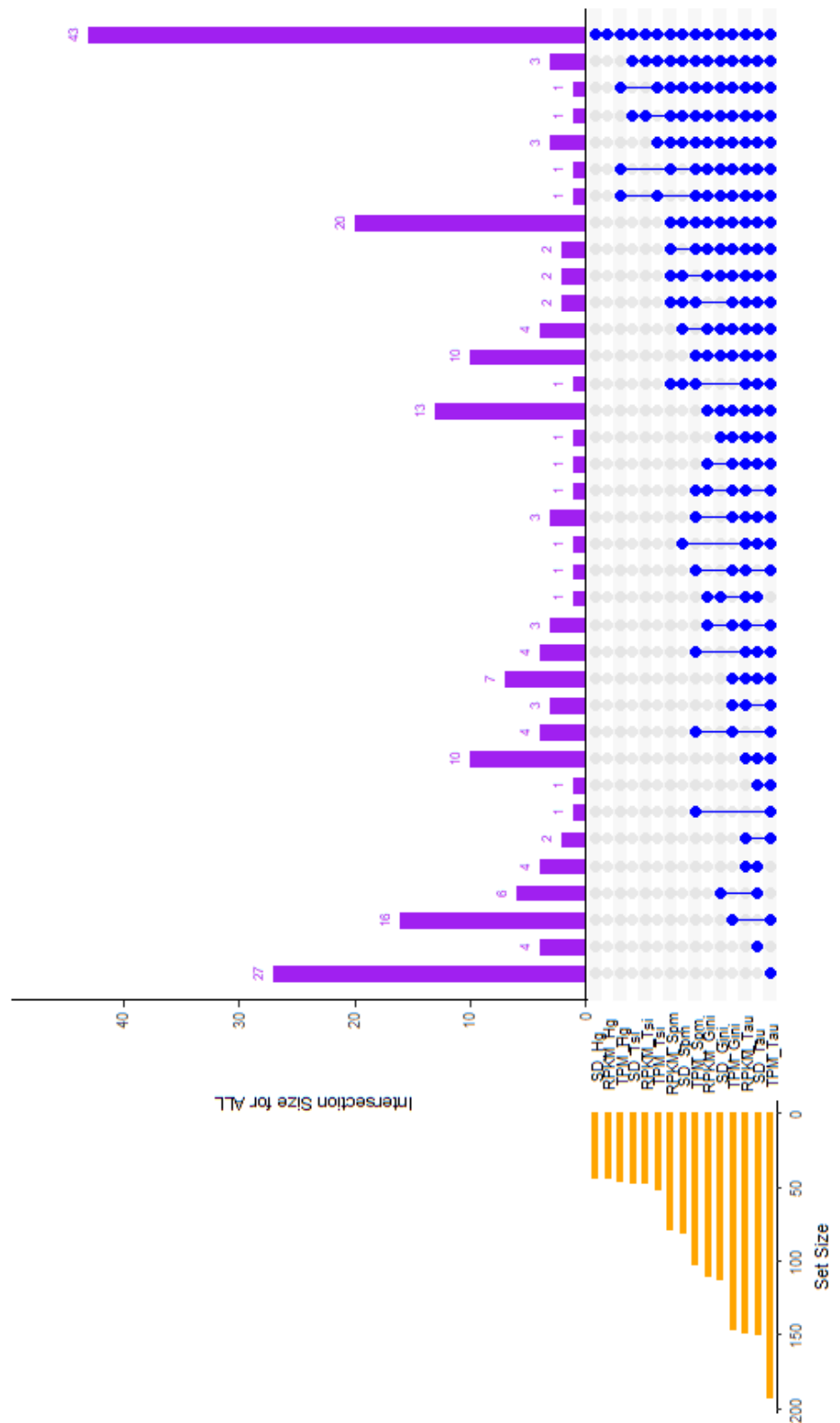


Figure 4.42: Upset graph showing the intersection between genes specific to tissues across differently normalized datasets and TS metrics used on them. (TS value greater than 0. for all the methods)

As shown in **Figure 4.42** most of the genes intersected in SD, RPKM and TPM with Tau as well as Gini metric. There are more TS genes found when Tau metric applied

to define TS. As shown in the **Figure 4.42**, Tau and either of Hg/Tsi were not always consistent with each other to decide TS genes, but Gini were.

In this thesis, several genes specific to retina, muscle, and liver were selected based on their tissue specificities, magnitude of log fold changes, and the tissues where they were expressed very high/highest (**Table 4.7, Supplementary Table 4**).

Table 4.7: Different Tissue specificity metric values for different normalization methods with the selected DEGs having $\text{Tau} > 0.8$ and $\text{abs}(\log_2\text{FC}) > 0.5849$, and ache gene.

		Tau	Tsi	Counts	Hg	Spm	Gini
ache	SD	0.7883	0.2523	0.0000	0.2130	0.4276	0.6147
	RPKM	0.7783	0.2437	0.0000	0.2021	0.4118	0.6022
	TPM	0.8479	0.3195	0.0000	0.2466	0.5697	0.6533
arr3a	SD	0.9999	0.9991	0.2143	0.9969	1.0000	0.9997
	RPKM	0.9999	0.9990	0.2143	0.9964	1.0000	0.9997
	TPM	0.9999	0.9990	0.2143	0.9966	1.0000	0.9997
rom1b	SD	0.9995	0.9926	0.2143	0.9785	1.0000	0.9964
	RPKM	0.9994	0.9917	0.2143	0.9759	1.0000	0.9957
	TPM	0.9997	0.9956	0.2143	0.9860	1.0000	0.9976
rpe65a	SD	0.9995	0.9936	0.5000	0.9827	1.0000	0.9983
	RPKM	0.9994	0.9922	0.5000	0.9794	1.0000	0.9979
	TPM	0.9994	0.9914	0.5000	0.9779	1.0000	0.9978
fabp10a	SD	0.9484	0.5805	0.1429	0.6855	0.6977	0.9306
	RPKM	0.9544	0.6104	0.1429	0.6925	0.7483	0.9350
	TPM	0.9359	0.5270	0.1429	0.6543	0.6117	0.9162
pck1	SD	0.9550	0.6135	0.0000	0.6418	0.8095	0.9240
	RPKM	0.9587	0.6334	0.0000	0.6540	0.8301	0.9284
	TPM	0.9295	0.5032	0.0000	0.5847	0.6438	0.8967
desma	SD	0.8724	0.3589	0.1429	0.4511	0.4678	0.8250
	RPKM	0.8782	0.3697	0.1429	0.4551	0.4923	0.8289
	TPM	0.8518	0.3252	0.1429	0.4164	0.4282	0.8007
myl3	SD	0.9762	0.7498	0.0000	0.7277	0.9294	0.9523
	RPKM	0.9773	0.7590	0.0000	0.7287	0.9384	0.9527
	TPM	0.9370	0.5312	0.0000	0.6050	0.6737	0.9053
sagb	SD	0.9997	0.9963	0.2857	0.9882	1.0000	0.9981
	RPKM	0.9997	0.9956	0.2857	0.9863	1.0000	0.9978
	TPM	0.9997	0.9964	0.2857	0.9887	1.0000	0.9985

Most of the genes we selected as DEGs in AChE data for further investigation had higher TS scores for Tau in any of the normalization method selected and were consistent for most of them. For example, the genes selected in **Table 4.7** had very close and higher TS scores across different metrics (except counts). However, this was not always the case for Tsi and Spm. Furthermore, the selected retinal genes

(*arr3a*, *rpe65a*, *rom1b*, *sagb*) were consistently tissue specific regardless of method and metrics chosen, except counts.

4.3.2.3 qPCR Validations on Muscle and Liver and Retina

To test whether the genes found as TS for AChE dataset were changed in the direction of the *in silico* data, qPCR experiments were done on 3dpf ache embryos.

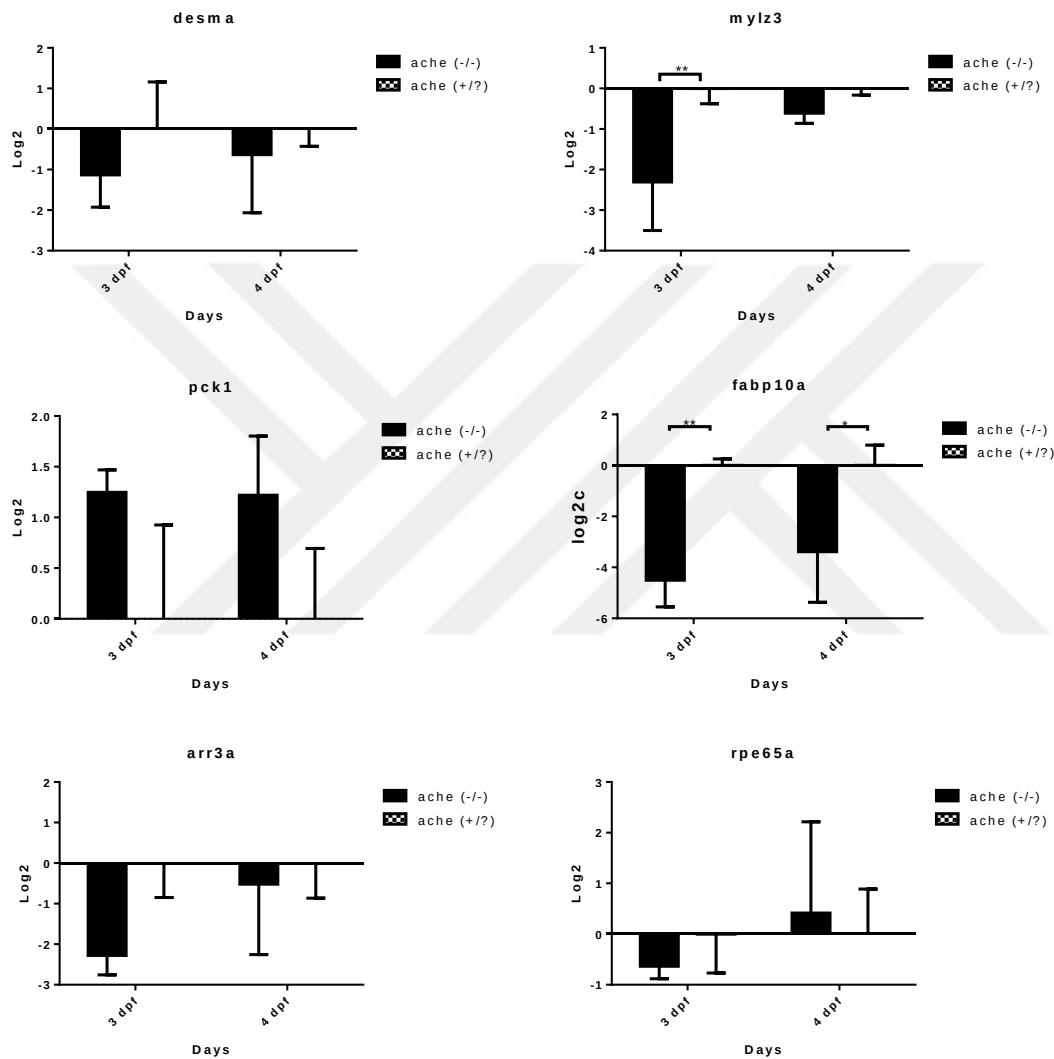


Figure 4.43: The qPCR results of the selected Tissue Specific (TS) and Differentially Expressed Genes (DEGs) on *ache* dataset. *desma* and *mylz3* for muscle, *pck1* and *fabp10a* for liver, *arr3a* and *rpe65a* for retina.

Sidak's multiple comparison test was applied to find the significance of difference between the expression in different strains of zebrafish. *s show how significant the change between the defined conditions. * <0.05 , ** 0.01 , *** <0.001 , **** <0.0001 . The statistical analyses were performed using GraphPad Prism version 6.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com. Error bars

represent SD from the mean, n=3 for each condition except *arr3a* and *rpe65a* (n=2). There were high error associated with the measurements for some of the conditions. The experiments will be repeated for the same set of cDNAs to increase technical replicates and when needed new biological replicates will be added.

In brief, as shown in **Figure 4.43**, muscle specific *mylz3* and liver specific *fabp10* were found to be significantly down regulated in the AChE dataset while *desma*, and *pck1* were up-regulated although not significantly in 3dpf ache homozygous mutant with respect to their healthy (ache +/-) siblings from the same breedings. *Mylz3* and *fabp10* were significantly downregulated and the difference were more pronounced when compared to *in silico* data. In 3dpf embryos, both genes selected as retina specific were downregulated for ache homozygous mutant zebrafish, and downregulation of *arr3a* and *rpe65a* were not significant at 3 dpf and 4 dpf embryos upon post-hoc comparisons.

4.3.3 Pathway Enrichment using GO BP analysis in R

When abs (L2FC)>0.5849 cut-off applied and Tau>0.8 was taken as tissue specificity threshold on RPKM normalized dataset, several pathways were enriched with the genes in the dataset. For Gene Ontology Biological Process enrichment analysis, enrichGO function with p-adjusted value cut off 0.05 was chosen for every up/down regulated tissue enrichments. EnrichGO function requires ENTREZIDs to map the genes to specific processes, and at the annotation part with bitr function, information belong to 1/3 TS gene were lost. The longevity of bars did not show how much gene involved in the process given in the left side of the graph.

Brain- Specific Expression of AChE mutant zebrafish

No specific enrichments were found for brain.

Liver- Specific Expression of AChE mutant zebrafish

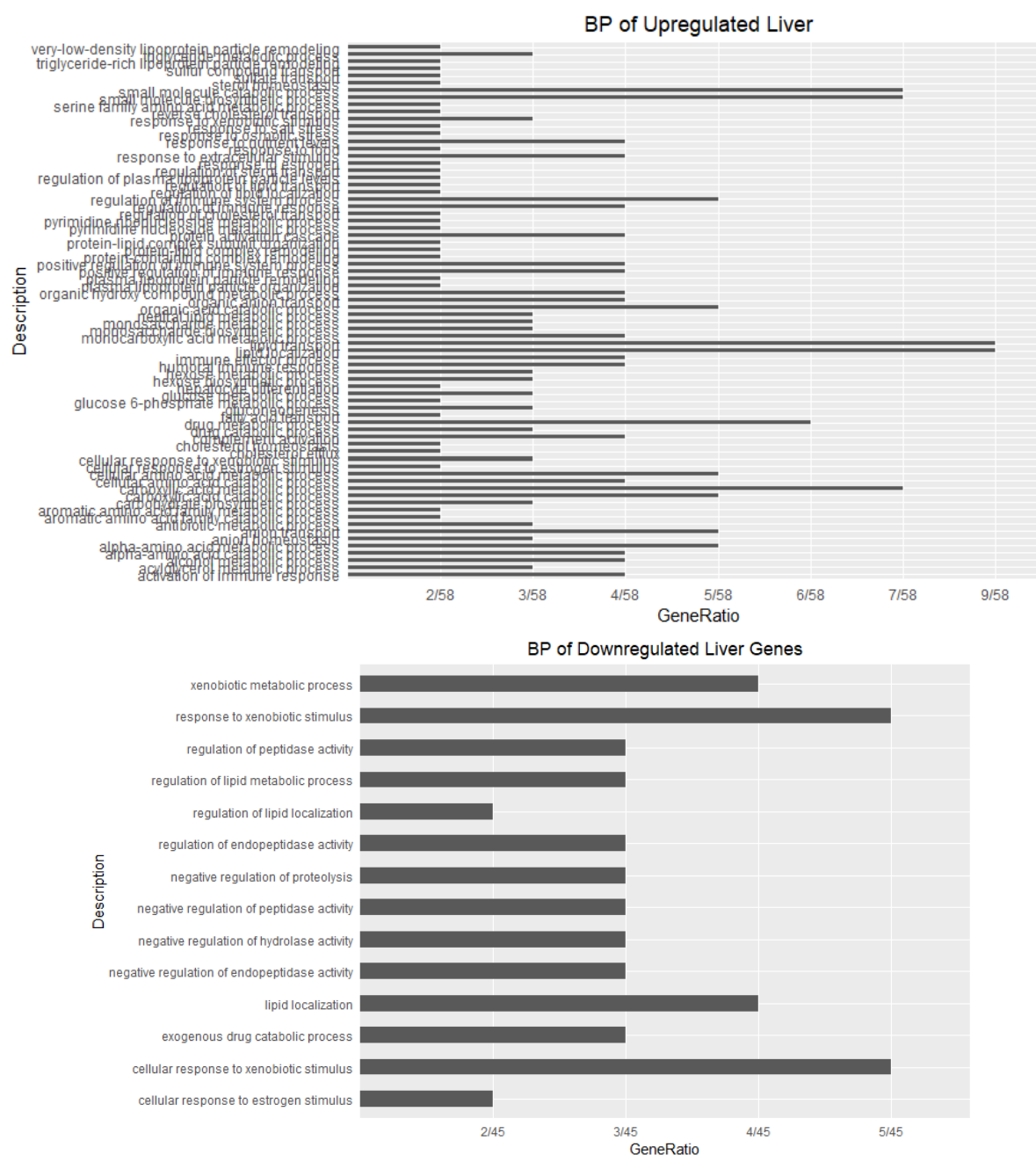


Figure 4.44: GO Biological Processes affected as liver specific up and down regulated genes in AChE dataset.

When ache was absent, the pathways which the up/down-regulated genes were related shown in **Figure 4.44** were shown for liver. 9 of 58 processes of upregulated genes were in lipid transport and lipid localization, and a significant number of genes were found in carboxylic and drug catabolic processes. In addition, immune system regulation related genes were also enriched. The genes downregulated in liver were

involved in processes such as xenobiotic response (5/45) and lipid localization (4/45).

Retina- Specific Expression of AChE mutant zebrafish

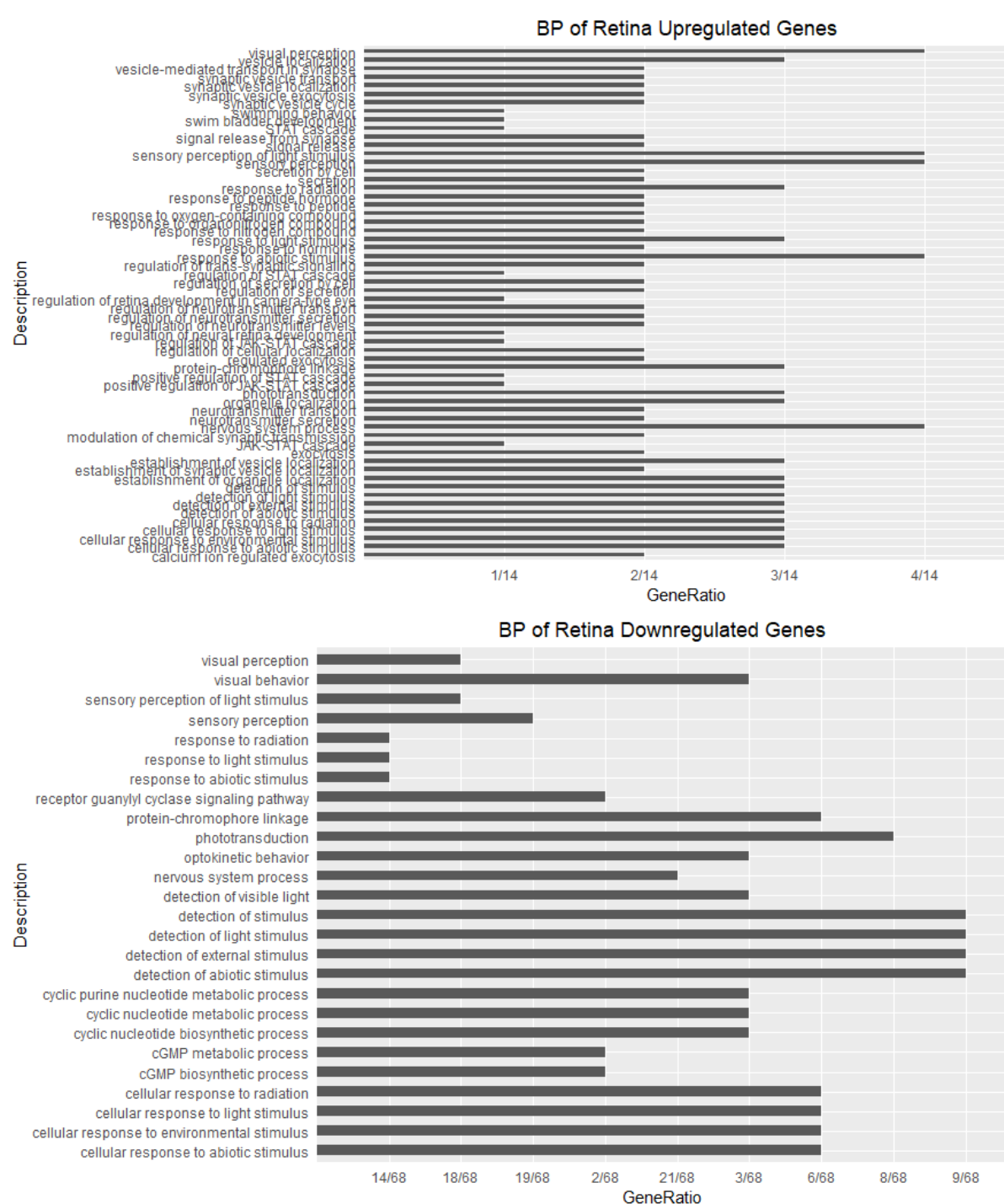


Figure 4.45: GO Biological Processes affected as retina specific up (up) and down regulated genes in AChE dataset.

Although up (14) and down (68)-regulated retinal genes were involved in common biological processes such as visual perception shown in **Figure 4.45** significant amount of the downregulated retinal genes were involved in nervous system

processes (21/68). In fact, the genes having a role in the regulation of neurotransmitter transport were upregulated. On the other hand, there were many downregulated genes in AChE mutant which were involved in sensory perception of light stimulus (14/68), phototransduction (3/68), and response to light stimulus (14/68).

Skin- Specific Expression of AChE mutant zebrafish

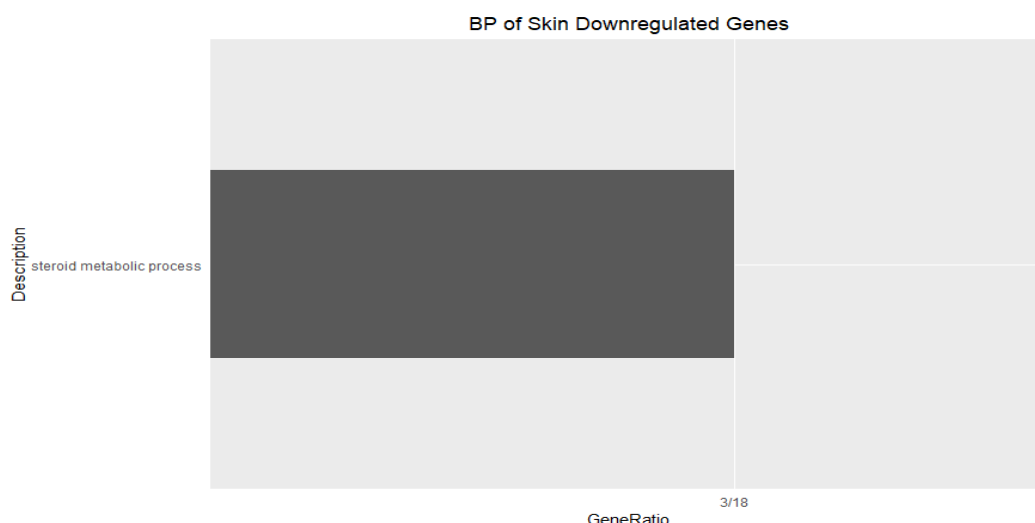


Figure 4.46: GO Biological Processes affected as skin specific up and down regulated genes in AChE dataset. None was available for downregulated genes.

The skin specific downregulated genes were related to steroid metabolic processes as shown in **Figure 4.46**.

Muscle- Specific Expression of AChE mutant zebrafish

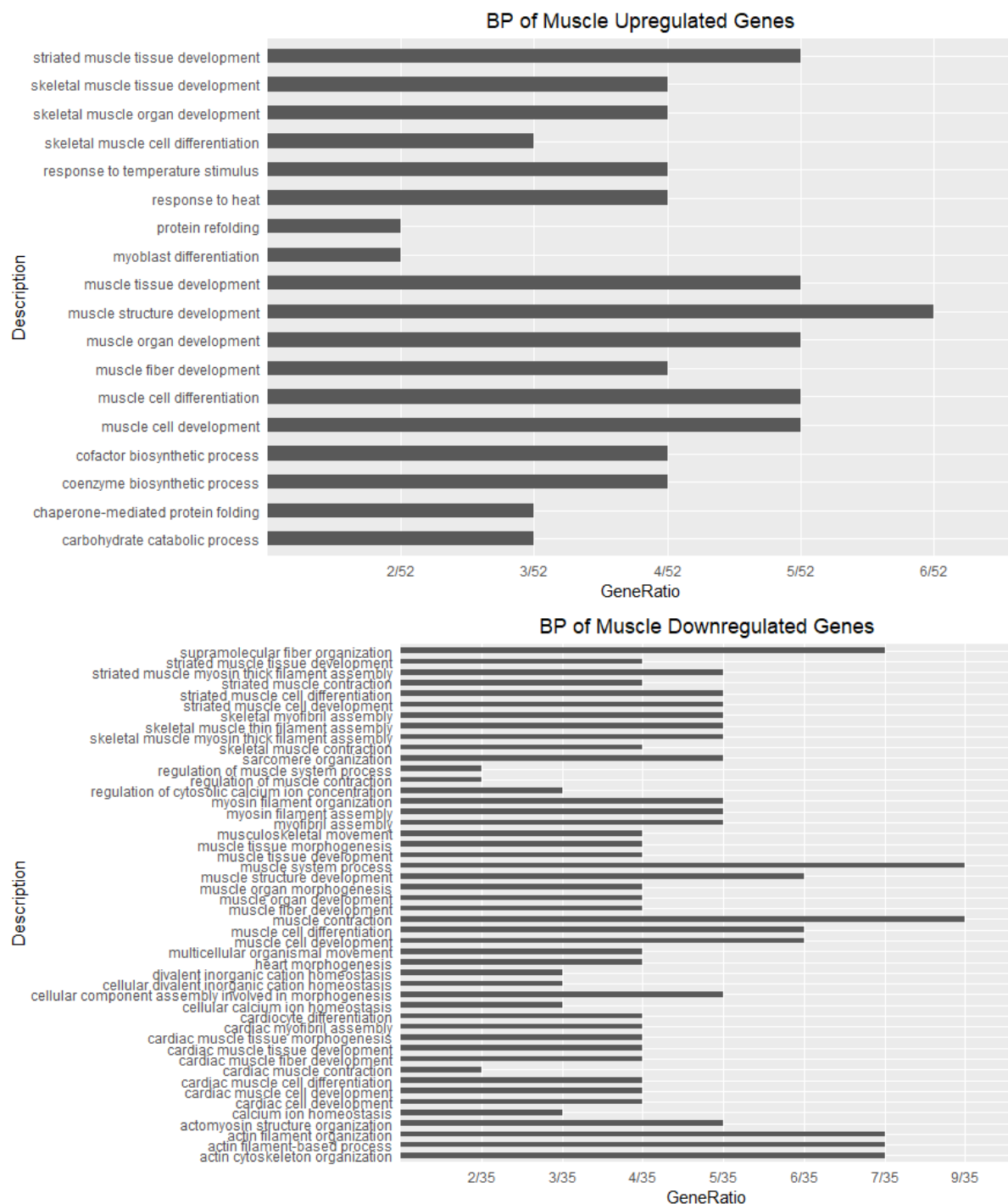


Figure 4.47: GO Biological Processes affected as muscle specific up and down regulated genes in AChE dataset.

As shown in **Figure 4.47**, despite the fact that most of the upregulated (52) muscle genes were involved in muscle development and muscle cell differentiation, the number of the downregulated genes involved in the same processes (35) were higher than upregulated ones; and these downregulated genes were also involved in

cardiomyocyte functioning which might explain the malfunctioning of heart muscles and muscle degradation in general in AChE mutants.

Ovary- Specific Expression of AChE mutant zebrafish

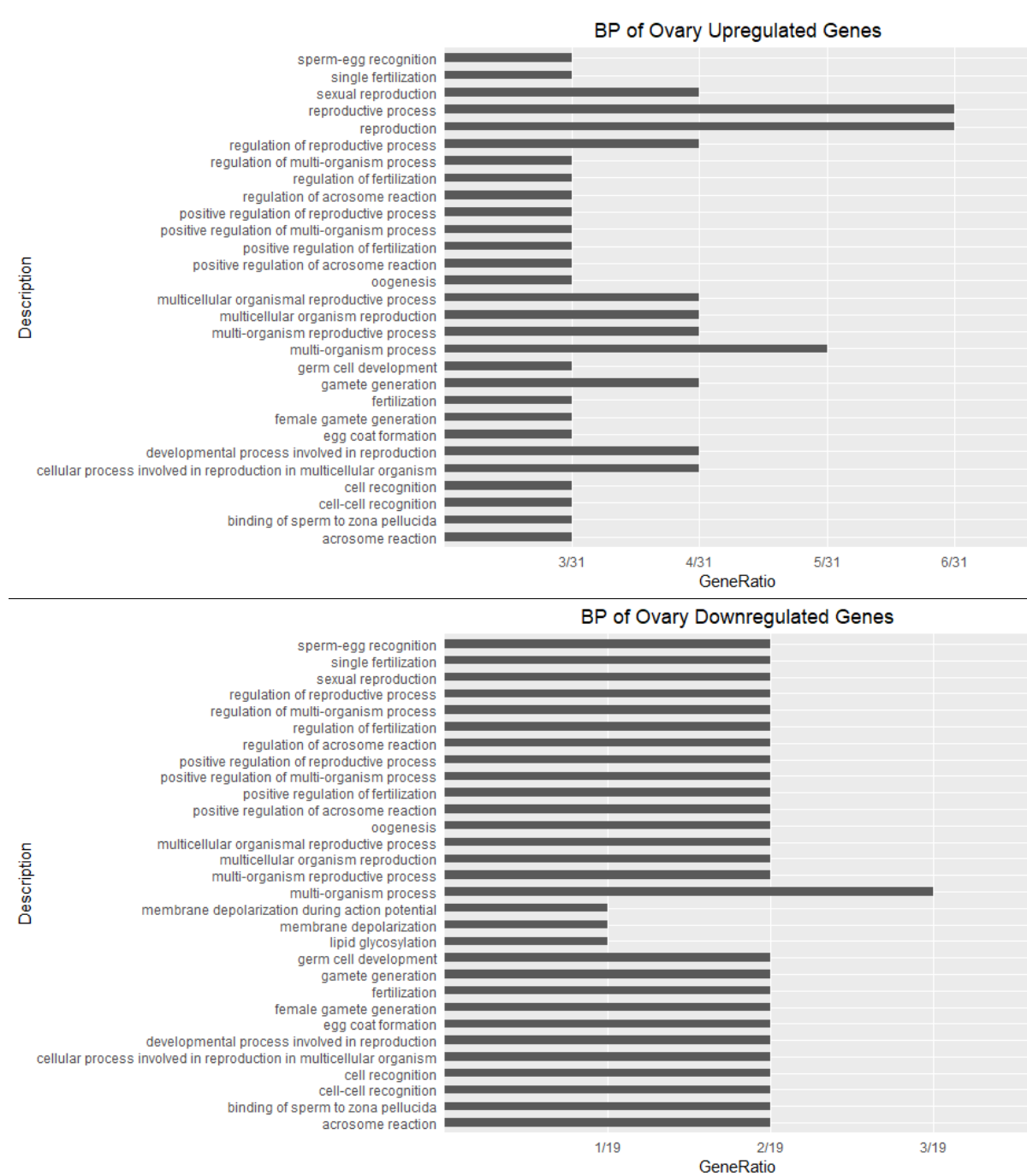


Figure 4.48: GO Biological Processes affected as ovary specific up and down regulated genes in AChE dataset.

Ovary specific up (31) and down (19) regulated genes were involved in the same processes. However, there were more upregulated genes involved for these common processes as shown in **Figure 4.48**. Reproduction related processes were more related to upregulated genes.

Testis- Specific Expression of AChE mutant zebrafish

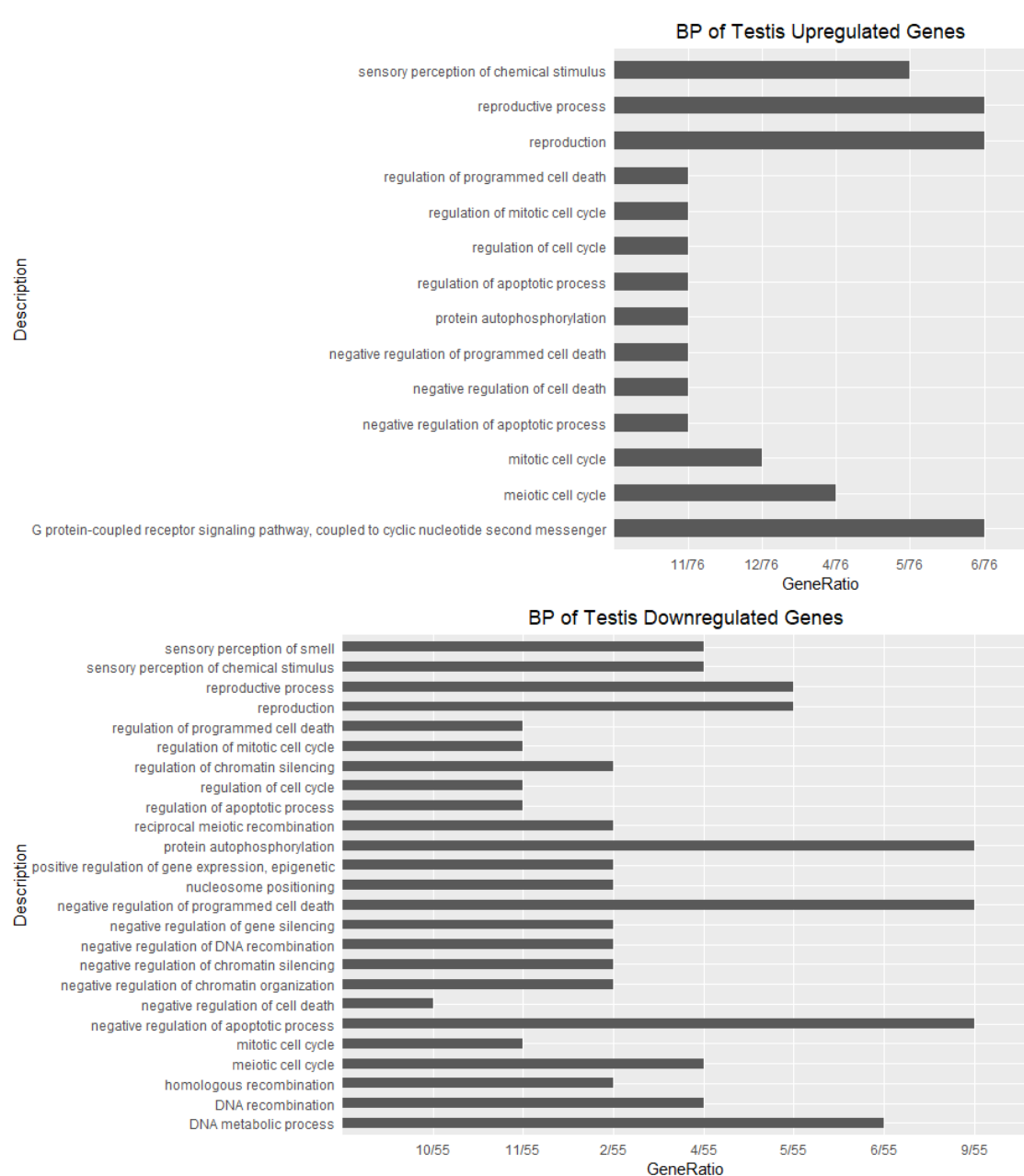


Figure 4.49: GO Biological Processes affected as testis specific up and down regulated genes in AChE dataset.

Among the testicular upregulated genes (75), there were more genes involved in the process of cell cycle regulation than downregulated genes (55) as shown in **Figure 4.49**. However, the processes were mostly shared between up and downregulated genes. On the other hand, DNA recombination and DNA metabolic process related genes were downregulated in AChE mutant embryos.

Kidney- Specific Expression of AChE mutant zebrafish

No enrichments available for kidney.

Spleen- Specific Expression of AChE mutant zebrafish

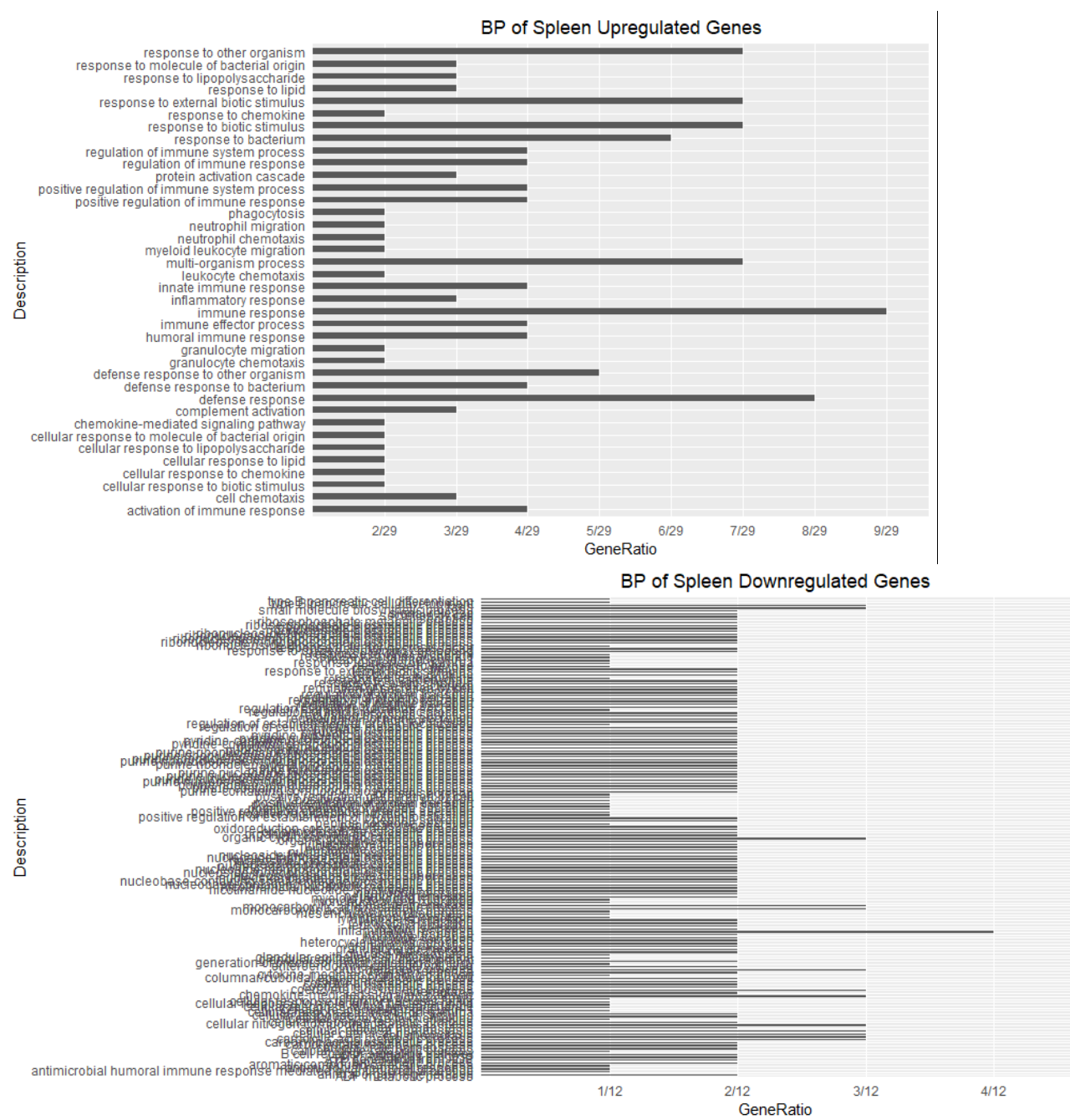


Figure 4.50: GO Biological Processes affected as spleen specific up and down regulated genes in AChE dataset.

Interestingly, there were significant amount of immune system related genes upregulated (29) as spleen specific genes in AChE mutant, and few number of genes enriched for multiple different process for downregulated ache genes, which was shown in **Figure 4.50**.

Heart- Specific Expression of AChE mutant zebrafish

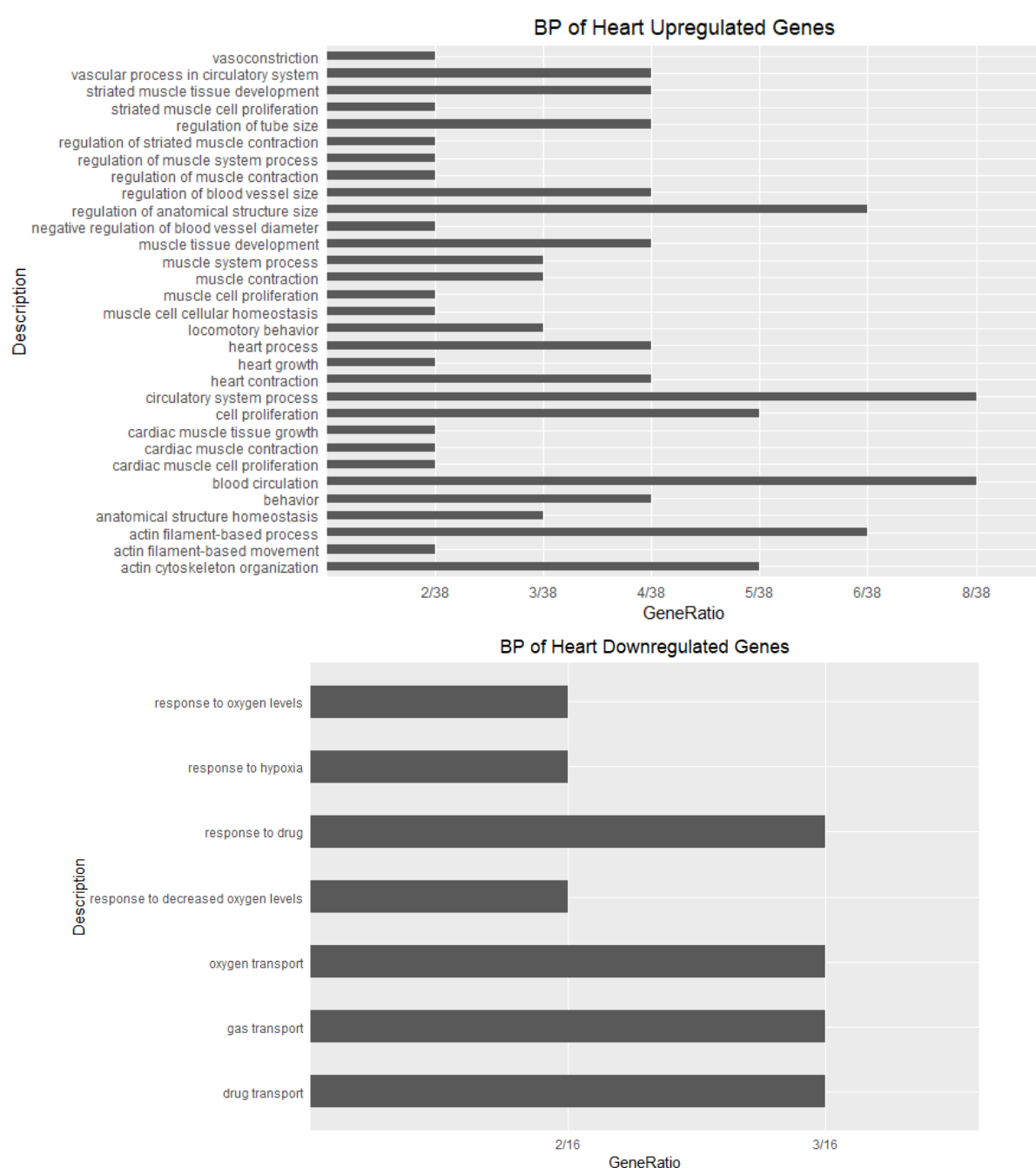


Figure 4.51: GO Biological Processes affected as heart specific up and down regulated genes in AChE dataset.

Heart was one of the main organs affected from the lack of AChE directly. Of 38 processes, most of them were related with muscle and heart growth, proliferation and contraction for upregulated genes as shown in **Figure 4.51**. Because of vascular system also seemed to be involved, there was significant number of genes upregulated regarding the circularatoy system as well.

Intestine- Specific Expression of AChE mutant zebrafish

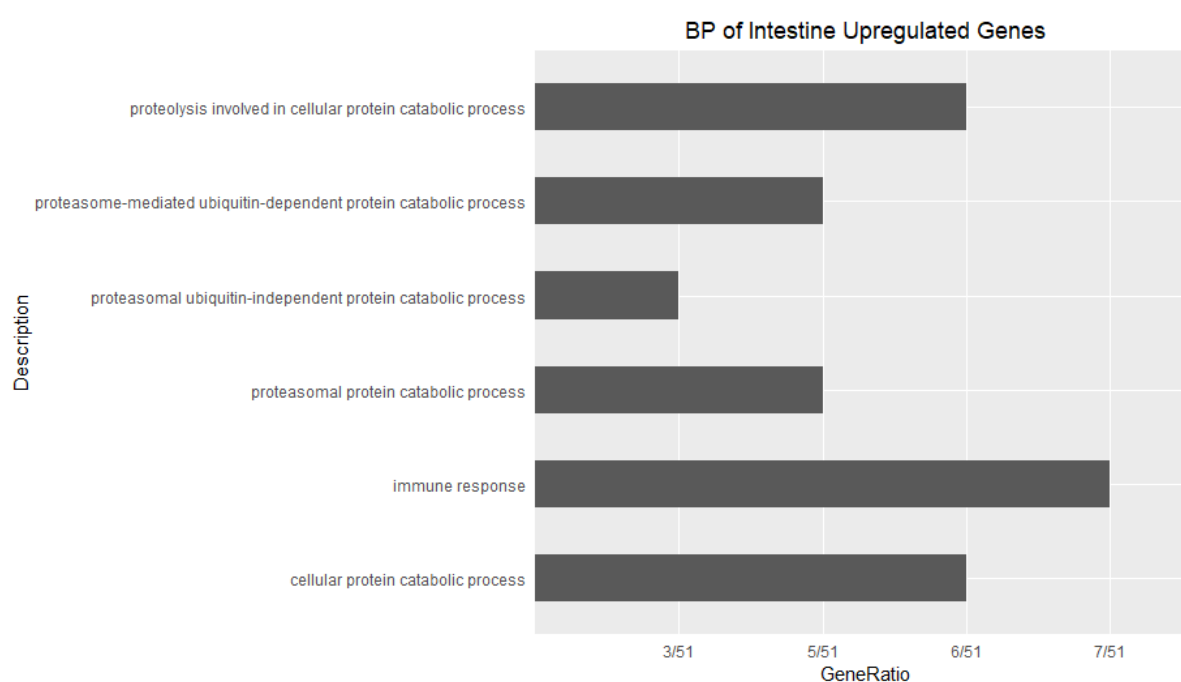


Figure 4.52: GO Biological Processes affected as intestine specific up regulated genes in AChE dataset. None was available for down regulated genes.

As in the case of spleen, many immune system related genes were involved in intestine and were upregulated in AChE mutants which was shown in **Figure 4.52**. In addition, the intestine specific genes involved in proteasome related processes were also upregulated in AChE mutants.

Bones- Specific Expression of AChE mutant zebrafish

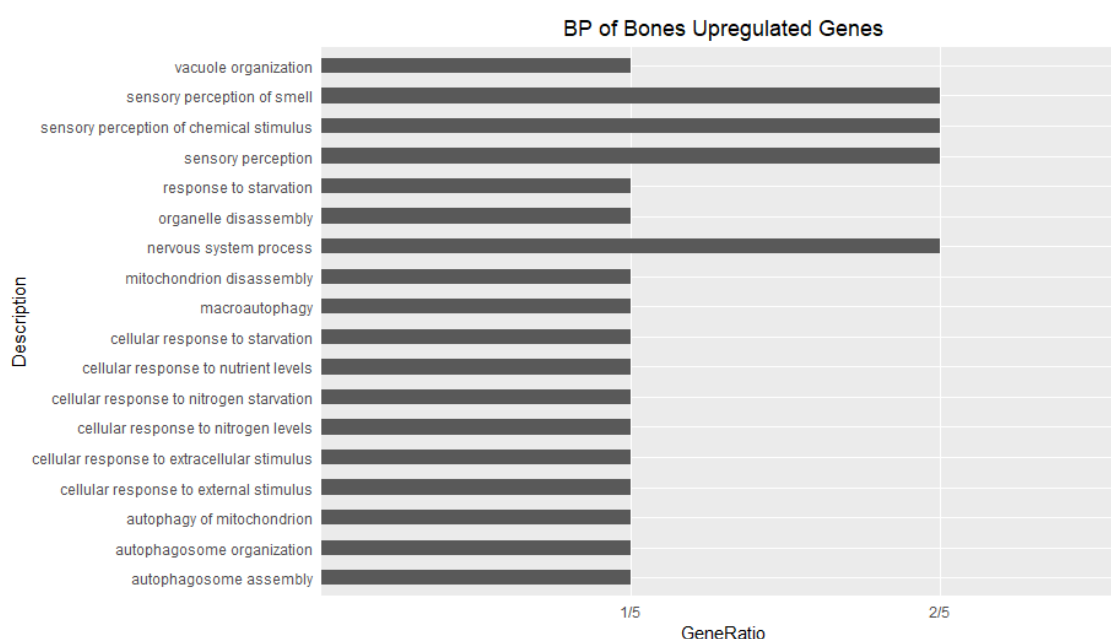


Figure 4.53: GO Biological Processes affected as bone specific up regulated genes in AChE dataset. None was available for down regulated genes.

The upregulated genes specific to bones were involved in processes such as perception of chemical stimuli and nervous system processes, the detailed processes were shown in **Figure 4.53**.

Gill- Specific Expression of AChE mutant zebrafish

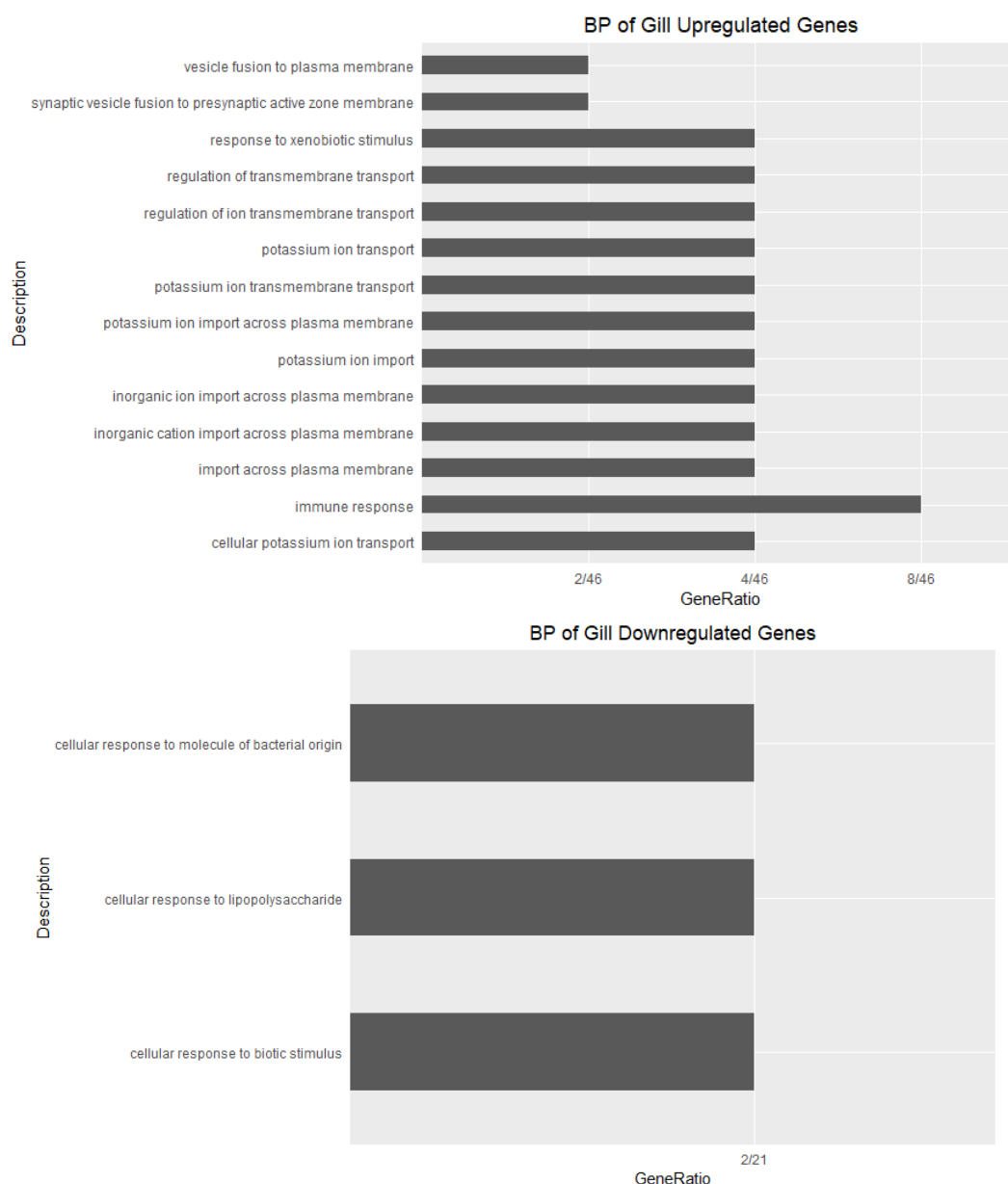


Figure 4.54: GO Biological Processes affected as gill specific up and down regulated genes in AChE dataset.

Most of the upregulated genes specific to gill in AChE dataset were involved in ion import metabolisms, immune response as shown in **Figure 4.54**.

Embryos- Specific Expression of AChE mutant zebrafish

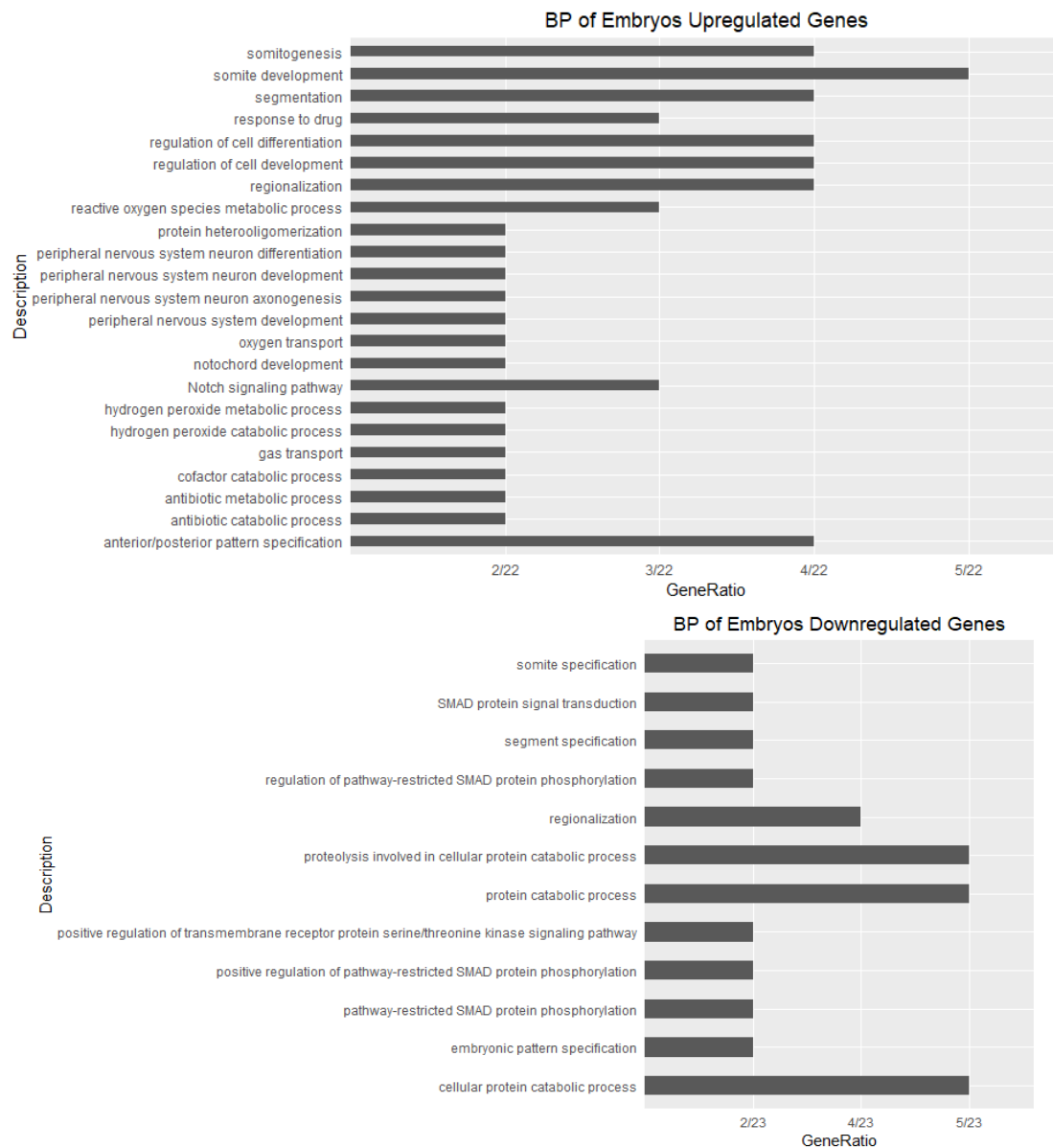


Figure 4.55: GO Biological Processes affected as embryo specific up and down regulated genes in AChE dataset.

Although most of processes were common for up and down regulated genes specific to zebrafish embryos in AChE dataset, SMAD protein related processes were downregulated in AChE embryos as shown in Figure 4.55.

UFeggs- Specific Expression of AChE mutant zebrafish

None of enrichments available for UFeggs.

4.4 The Relationship between AChE and Retinal Genes

Up and Down Regulated Genes of Retinal Genes

As shown in **Figure 4.56**, independent of normalization method chosen for the TS analysis, most of the genes whether up (a) or down (b)-regulated, were same and specific to eye tissue in zebrafish.

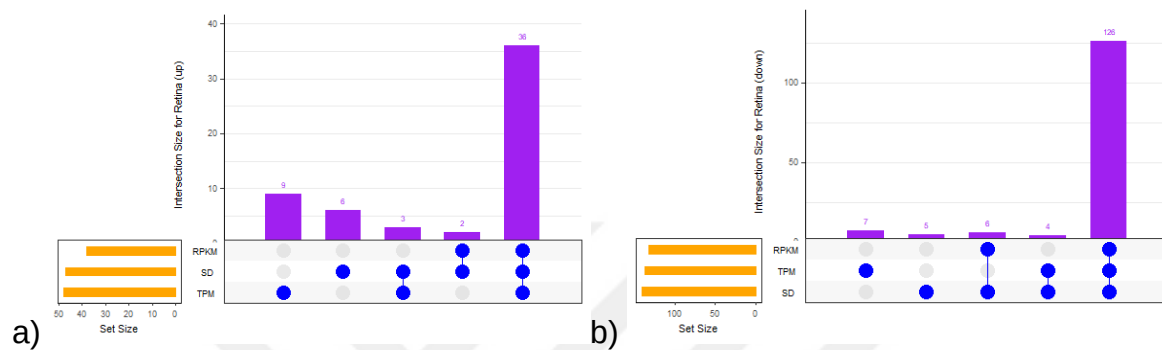


Figure 4.56: absolute log2 fold change greater than 0.5849 DEGs with Tau score greater than 0.8; a) up b) down regulated genes.

4.3.1 qPCR results

The genes which were down regulated and tissue specific in *in silico* dataset of ache mutant 3dpf zebrafish embryos were further confirmed on 3 and 4 dpf ache mutant zebrafish embryos raised at Bilkent University Zebrafish Facility as shown in **Figure 4.43** (e and f).

Because there were significant number of genes that were downregulated as retina specific genes in the ache dataset, and to understand the relationship between ache and retinal genes better, I performed experiments under different dark light regimes using 3, 4 and 5 dpf zebrafish embryos (**Figure 4.57**).

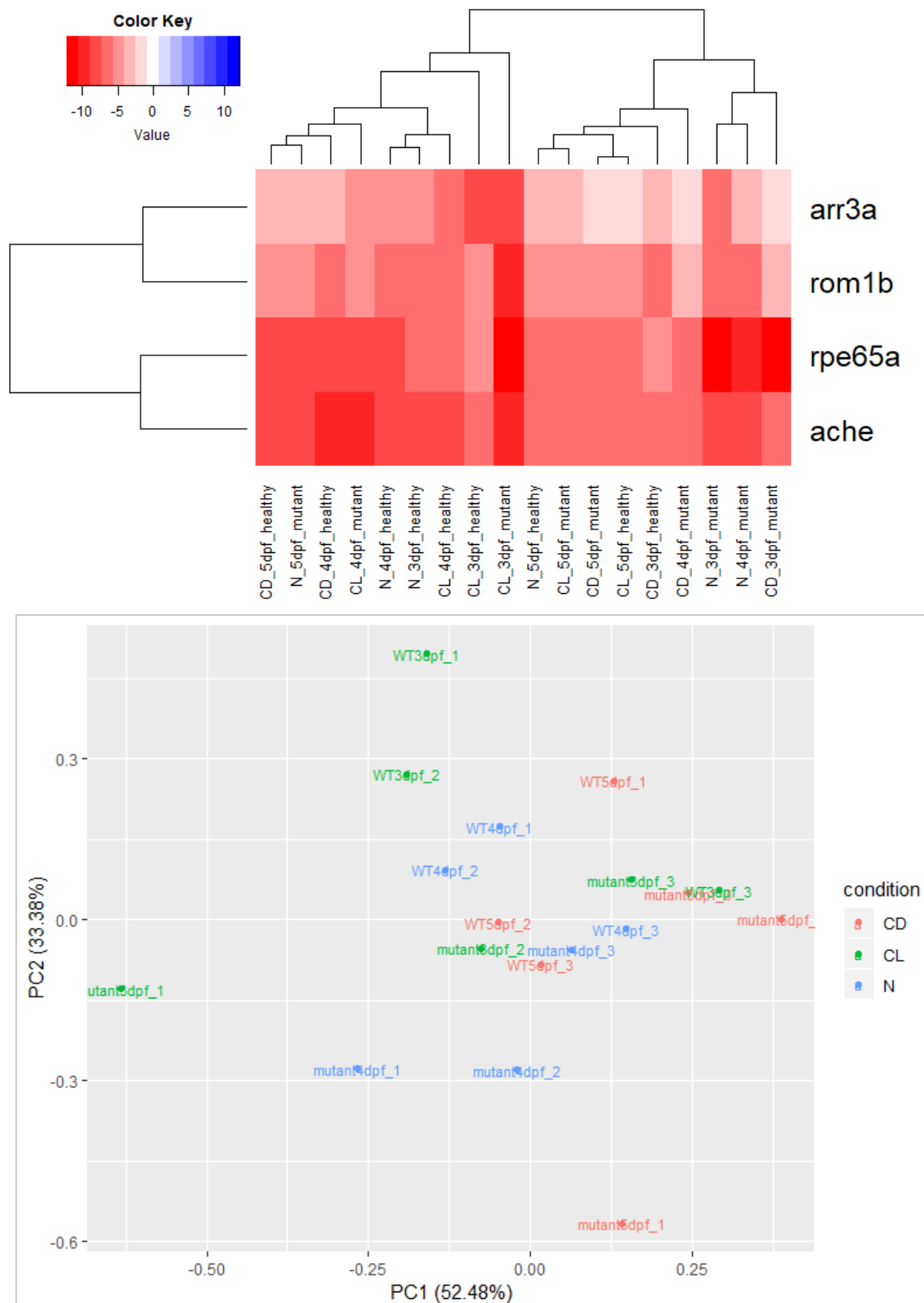


Figure 4.57: qPCR results of the *ache*, *arr3a*, *rom1b*, *rpe65* genes for WT: *ache* (+/?) and mutant: *ache* (-/-) sibling embryos of the same breedings raised under different lightening conditions. CL: Constant Light, N: Normal cycle; CD: Constant Darkness. dpf denotes days post fertilization.

To test how similar the expression pattern of retinal genes with *ache*, $-\Delta Ct$ calculations (reference-target) were done on qPCRs. As shown in **Figure 4.57**, *ache* expression itself was downregulated in *ache* mutant embryos at 3 dpf when raised in constant light, whereas, upregulated in *ache* mutant embryos at 4 dpf *ache* mutant embryos with respect to healthy siblings (heatmap). On the other hand, *arr3a* was upregulated with respect to healthy siblings raised in same conditions, constant light, and in both phenotypes, *arr3a* levels were decreased with respect to embryos raised in normal light-dark cycle (data was not provided). In contrast to constant light conditions, in constant light, *ache* mutant embryos had higher *arr3a* expression with respect to their healthy siblings, and both strain had a higher expression with respect to embryos raised in normal day light conditions. The difference between the expressions was decreased at 4 dpf embryos. *Rom1b* followed the same expression pattern with *arr3a* in 3dpf embryos except upregulation of *arr3a* in healthy siblings raised in constant light with respect to normal day light conditions (heatmap). *Rpe65a* follows the opposite pattern of expression of *rom1b* at 3 dpf embryos, and at 4 dpf its expression was upregulated in *ache* mutant with respect to healthy siblings in both constant light and constant darkness (heatmap). In addition, mutant 3 dpf zebrafish raised in CD, and CL tend to cluster separately (PCA).

Chapter 5: Discussion

In this thesis I aimed to determine tissue specificity (TS) in zebrafish using data from 15 tissues that were publicly available online. RNA-Seq datasets of these tissues were analyzed using same pipeline as shown in **Figure 1.1**. Then, by using different tissue specificity metrics available, which were compared with each other previously (Kryuchkova-Mostacci & Robinson-Rechavi, 2016), the TS index values were determined. After defining TS values, genes with corresponding TS values were merged with ache mutant vs. healthy dataset of 3 dpf zebrafish embryos to understand the role of AChE gene at embryonic stage. TS values gave us the information which tissues were affected most in the absence of AChE activity. As a brief conclusion, retina, along with liver and muscle were among the most effected tissues in the absence of ache activity (**Table 4.4, 4.5 and 4.6**), hence this relationship was further tested for 3 and 4 dpf ache mutant vs healthy siblings using qRT-PCR. In addition, for retinal genes I tested 3 to 5 dpf ache mutant and ache healthy siblings at different light conditions (i.e constant dark and constant light) and the normal light dark cycle (i.e 14hrs light/10hrs night) as a control. On the other hand, 3 different normalization methods were compared while measuring tissue spcecificity on the dataset, which included the RNA-Seq files from 5 different datasets covering 15 tissues. These comparisons showed us that the TS values were both the normalization method and TS metric dependent (**Table 4.5 and Supplementary Figure 6-53**). In addition, log transformation of the normalizations might have also an influence on the definition of TS genes (**Supplementary Figure 6-20**).

5.1 Were the tissues from different datasets clustered together?

There were different confounding factors (i.e sequencing platform, library preparation or library size differences based batch effects) present while combining different datasets of RNA-Seq experiments (Bush, McCulloch, Summers, Hume, & Clark, 2017; Qian Liu, 2016). In fact, I used the datasets originated from 5 different RNA-seq experiments (**Table 3.6**). Therefore, it was important to combine these datasets with minimum variance that might be caused by such factors as previously mentioned (Simoneau & Scott, 2019). To minimize these effects, the experiments using the same platform (Illumina) were chosen. On the other hand, alignment

methods and raw count retrieval methods might cause variation as well. To eliminate such factors, the same pipeline applied on each of the RNA-Seq datasets retrieved from <https://www.ebi.ac.uk/ena> (**Figure 1.1**). After the selected normalization methods (i.e SD, RPKM or TPM), log transformed counts of the tissues were analyzed by dendrograms and PCA plots (clustering with k-mers, 10 in this case) and boxplots. These showed that tissues having more than one sample, even though combined from several experiments, clustered together, and these clusters did not depend on the normalization method (**Figure 4.2, 4.4 and 4.5**). However for example for the two different skin datasets subclustered within the skin tissue cluster. In the future an additional step of normalization, e.g., quantile normalization, might be tested for further batch correction. It was previously shown on miRNA expression based studies that brain and skin or some of the spatially and developmentally closer tissues like liver and kidney, or brain and nerve sheets were very closely clustered together in across human tissue dataset (Ludwig et al., 2016); and it was suggested that it could be possible to predict the cluster IDs based on regional transcriptome (D'Souza et al., 2008; Sudmant, Alexis, & Burge, 2015). Regionally closer tissues in adult zebrafish such as brain and retina, liver and intestine, skin and heart, muscle and bones also tended to be in the same cluster independent of the normalization method (**Figure 4.4**). On the other hand, as shown in **Figure 4.4**, liver, kidney and spleen consistently clustered together. The reason behind such clustering might be the contamination of spleen or kidney tissues with liver because they all are found very close to each other in the zebrafish body. Interestingly, two of the liver samples from the same experimental step did not cluster together due to distinct expression pattern of the genes compared to others. The gender of the fish were partly unknown, this might be because of the differences between gender of the fish taken for experiment or the unknown health conditions of the zebrafish at the time of dissection.

5.2 Did the normalization method chosen have any effect on the distribution of tissue specificity scores?

Normalization was done by adjusting global properties of measurements for different samples to make them appropriately comparable (Simoneau & Scott, 2019). Moreover different normalization methods considered different aspects of the data to minimize any kind of batch effects. Although there was not a best method available to

normalize the data, RPKM and DESeq2 (previously DEseq) were two of the most commonly used normalization strategies (Dillies et al., 2013; Qian Liu, 2016). Later, TPM was introduced to improve RPKM methodology (Wagner et al., 2012). Dündar et al (Friederike Dündar, 2015) recommended to be careful while selecting a normalization methodology for differential gene expression analysis no matter how often they have been suggested in the literature. For example, while measuring differential gene expression between the samples, in the same paper, they have stated to avoid using RPKM, and recommended to use RPKM/FPKM for differential gene/transcript expression analysis between experimental groups. DESeq2 normalization can be used to minimize the batch effects related to sequencing depth/library size differences by using estimate size factors function (Anders & Huber, 2010; Love et al., 2014). The size factor calculated is the geometric mean of all samples for each gene. “iterate” option in this function provides iteration between dispersion estimations with a design of ~ 1 . The size factor vector is found via numerical optimization to the likelihood of the ~ 1 model, which could be useful when comparing more than two samples (i.e. 15 tissues and more than 40 samples). Although, SD normalization does not consider gene length, RPKM (Mortazavi et al., 2008) and TPM (Wagner et al., 2012) do. In this thesis, I first used calcNormFactors to normalize the tissue datasets with RPKM, and took the sequencing depth into account. In this process, while DESeq2 used RLE, in RPKM default option with TMM methodology was chosen to estimate size factors. Then, rpkm function provided in edgeR package was used to normalize the data. The first difference between RPKM and SD were introduced here, i.e., the size factor estimation method. Then, the consideration of gene length in RPKM but not in SD came next. On the other hand, in TPM, the data were normalized considering the gene length, then the sequencing depth in million, and the method took the library sizes into account after normalizing each count to respective gene length. Theoretically, SD considers the difference between sequencing depth while RPKM uses calcNormFactors function of edgeR package to take into account the difference between library sizes. However, the size factor difference was considered after gene length normalization in TPM. These might result in the differences we have observed.

This thesis focused on comparison of methods in finding TS genes. Accordingly, I have found that RPKM and SD were more consistent with each other than they were

with TPM (**Figure 24-38**), which might be explained by the differences in the normalization methodologies such as the consideration of the size factor of libraries (**Table 3.7**). It was also previously (Dillies et al., 2013) stated that, the size factor estimation with TMM methodology prior to RPKM normalization might improve the RPKM normalization via maintaining FDR, and intravariance.

5.3 Should you take log transformation into account while measuring tissue specificity?

Originally, log transformation was used to decrease variance, normalize, visualize the data in gaussian shape, and to deal with skewness. After normalization of the datasets, log transformation in base 2 was applied on these normalized data to test its effects. Then, the specificity metrics measurements were done to assign TS values for each gene. Although almost all of the TS genes in different tissues were also found in TS metrics applied on normalized (in either of SD, RPKM or TPM) then log transformed counts as shown in **Figures 4.24-4.41**, the number of TS genes metrics applied on transformed counts were not as much as log-transformed counts. In fact, this might be due to an inconsistency between log-transformed and not-log-transformed data (Feng et al., 2014). Although Kryuchkova-Mostacci and Robinson-Rechavi (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) performed TS analysis on log transformed data while comparing the different metrics and TS in different datasets, as previously used in RNA-Seq analysis, log transformation might be more suitable for exploratory analysis and clustering such as data visualization than differential gene expression analysis (Friederike Dündar, 2015).

5.4 Did the metrics chosen have any effect on the definition of tissue specific genes?

Six different metrics, namely, Tsi, Tau, Counts, Hg, Spm and Gini, were chosen, and applied on non-transformed counts of the normalized counts. To explore whether the metrics chosen had any effect of the determination of TS genes or the TS genes were consistently found with every metric chosen, the intersection plots (upset graphs) were generated in R. These upset graphs showing the intersection size of the genes found in selected normalization methods were drawn (**Figure 4.44 and Supplementary figures 6-53**). The TS value cutoffs were chosen following the criteria on the reference article followed in this thesis (Kryuchkova-Mostacci & Robinson-Rechavi, 2016). Tau and Gini always intersected. They also intersected

with Spm for all the tissues except brain for the DEG cut-off chosen as absolute log2 fold change greater than 0.5849. When the DEG cut-off was chosen as p-value smaller than 0.1 or 0.05, the number of TS genes decreased for every metric tested. As shown in **Supplementary figures 21-36**, Tsi and Hg metrics were the strictest metric to decide to TS (**Supplementary Figure 4**); they had the lowest amount of TS values if one did not consider the counts metric. In the counts metric, it would be great to put an expression cut-off. When this provided, the count metric can work well. In Hg, the opposite of other metrics was used to define TS (0.8 was used for all of them after transformation of Hg values via 1-Hg to be comparable with the other metrics). Most of the TS genes were found in Testis tissue as previously shown in other papers (Kryuchkova-Mostacci & Robinson-Rechavi, 2016; Xiao et al., 2010). However, when p-value cut-off applied to select testis specific genes, there were less number of TS genes observed (**Table 4.1** and **Supplementary Figures 36-54**).

Although Tau and Tsi used the information provided for each gene each tissue, and took the maximal expression into account, the TS values were not always consistent in between Tsi and Tau (**Figure 4.42, Supplementary Figure 21-35 and 52-54**). In fact, Tau considered the number of tissues available while measuring the TS, whereas Tsi did not, and this made a big difference while finding the TS genes and made Tsi stricter than Tau. On the other hand, Tau found relatively high number of genes as tissue specific. For example, in the Tau metric, if there is a big difference between the expression values of the genes assigned to tissue with respect to other tissues that gene is considered expressed, Tau method still considers that gene as specific to the tissue in which the gene expressed highly differential. It was previously shown that (Kryuchkova-Mostacci & Robinson-Rechavi, 2016) that Tau was the most precise method while measuring TS. Accordingly the choice of metric was Tau for the present study.

5.5 Which tissues were affected most in the absence of ache activity in zebrafish embryos?

The answer was dependent on the selected DEG list and TS cut-off. In the case of L2FC, most affected tissues in ache dataset were liver, retina, muscle, testis and intestine (**Table 4.4**). Most of them were downregulated genes. When p-value cutoffs were chosen to determine DEGs, liver, heart, retina-down, intestine-down, skin-downregulated genes and embryos were significantly enriched. The occurrence of

ache deficiency and muscle defects have been studied before (Behra et al., 2002; Etard et al., 2015). The present thesis showed that the expression profile of ache mutant zebrafish embryos were enriched for downregulated muscle and heart genes (**Figure 4.46 and 4.49**), and some of these were involved in cardiomyocyte functioning which might explain the malfunctioning of heart muscles and muscle degradation in ache mutants. On the other hand, liver related genes were also affected, and a significant number of these genes were also involved in immune system regulation (**Figure 4.44**). Although, the zebrafish inflammatory response has not been defined in drug-induced liver injury (DILI), the enriched pathways of xenobiotic response for downregulated genes and drug catabolism in upregulated genes (**Figure 4.44**) brought up whether ache was involved in these process (Vliegenthart, Tucker, Del Pozo, & Dear, 2014). In addition, for hepatocellular carcinoma, liver related cancer, there was an association between ache as a prognostic marker (Y. Zhao et al., 2011) as revealed before. Moreover, it was important to figure out underlying mechanism in HCC to improve immunotherapeutics (Y. Xie et al., 2018), speculating that the role of ache might be important here to shed light on the underlying mechanism. Based on the analyses of this thesis one can hypothesize that ache has a role in the regulation of immune system at embryonic stage and this would be testable. In fact research showed in mammals the role of ache in the suppression of cytokine release (Pohanka, 2014). This mechanism worked through a “pathway of cholinergic anti-inflammation”. This situation might explain the role of such inhibitors in the regulation of immune system (Gilboa-Geffen, Hartmann, & Soreq, 2012; Hu, Fu, Liu, Zhang, & Dong, 2009).

To study the role of ache in muscle tissue related functions, *desma* for upregulated and *mylz3* as one of the downregulated genes specific to muscle tissue were selected. Previously researchers showed that fast myosin genes like *mylz3* were decreased in the absence of fully functional ache receptors (Sepich et al., 1998; van der Meulen et al., 2005). This may be the case for the acetylcholine itself. The modulation of *mylz3* gene in zebrafish ache mutants, which might also explain the importance of mylz family genes in cardiac muscle contraction (Chen et al., 2008; Sepich et al., 1998; van der Meulen et al., 2005). *desma* gene was shown to be important for muscle organization and cardiomyocyte and muscle function; the association between myogenesis and *desma* gene has been shown previously

(Costa et al., 2002; M. Li et al., 2013; Ramspacher et al., 2015). *In silico* analysis of ache RNAseq dataset revealed that this gene was downregulated (**Table 4.7**). However, qTR-PCR experiments done on 3dpf zebrafish embryo did not lead to a significant downregulation although there seemed to be a trend (**Figure 4.40**). Future studies will involve repetition of the experiment with technical replicates. In the 4 dpf embryo, I did not have the *in silico* data, however, it seemed that the trend observed in 3 dpf was not conserved for 4 dpf, which might indicate presence of adaptation or maybe due to sample size. For future studies increasing sample size might be needed.

Liver was also one of the tissues that had significant enrichments when ache was not present at 3 dpf embryonic stage. *Pck1* as an upregulated gene and *fabp10a* as a downregulated gene were selected for liver qRT-PCR confirmations. Previously researchers found that either high levels of glucose (Z. Wang et al., 2013) or in the presence of negative regulator of myogenesis (e.g. the absence of *desma* gene as stated previously), *pck1* was upregulated; this might be the case for ache mutant zebrafish embryos not having proper functional muscle organization (Gao et al., 2016). It could also be possible the downregulation of *pck1* in later days after fertilization due to glut12 malfunctioning in ache mutants (Jimenez-Amilburu et al., 2015). Although there was a trend for upregulation of *pck1* in ache mutants by qRT-PCR analysis, this was not significant after post-hoc multiple tests and further technical replicates are needed to better conclude because of variability. *fabp10a* was previously shown as a liver specific gene (Venkatachalam et al., 2009), so it was used for positive control to test for a liver specific TS gene. Previous studies showed that abnormal expression of *fabp10a* gene was correlated to hepatotoxicity (Mesens et al., 2015). In the absence of ache, hepatotoxic mechanisms might be triggered and the findings from this study supported the downregulation of *fabp10a* expression in ache mutant embryos high significantly for day 3 and day 4.

According to ExPASy (<https://prosite.expasy.org/c>) (Artimo et al., 2012), protein domain and function analysis for zebrafish *ache* mRNA was shown in **Supplementary Figure 57**. Ache has different domains which might explain its non-classical roles. VWFC domain signature is a repeat site and involved in the

maintaining homeostasis could function as growth-factor binding proteins and it is part of vertebrate membrane protein and might have a role in the process of formation of big protein complexes. EGF-like domain signature 1 is a part of epidermal growth factor (EGF) sequence, and found in extracellular domains, secreted. This domain has different functions in different organism such as adipocyte differentiation inhibition in mouse, and activation of hepatocyte growth factor explained in ExPASy tool. 4Fe-4S ferredoxin-type, is a bacterial type ferredoxin, iron-sulfur binding region signature is iron-sulfur protein controlling the electron transfer in various metabolic processes. Thiolases active site is a part of biosynthetic and degradative pathways. There are two types of this domain, located in mitochondrion and peroxisomes. This domain found proteins related to lipid transport in mammalian systems. Anaphylatoxin domain signature is related to inflammatory stimulus mediation via inducing the contraction of smooth muscle. 2Fe-2S ferredoxin-type iron-sulfur binding region signature is asidic and role in electron transfer, thus involved in redox systems. Tubulin subunits alpha, beta, and gamma signature is major domain of microtubules and constitutes the cytoskeleton. These results from ExPASy tool validated our finding that why that much distinct pathways were affected in the absence of ache in zebrafish and provides insights about its non-classical function.

One of the interesting conclusions was that the relationship between ache and most of the significantly (in any case, either L2FC or p-value cut-offs) downregulated retinal genes (**Table 4.5, 4.6 and Figure 4.37**). This was explained more detailed in the next part.

5.6 What kind of relationship was present between ache and retinal genes?

Previous studies showed that light triggered activation controls the calcium flux which was also affected when ache was absent (Korenbrod, 2012; Korenbrot et al., 2013). In addition, acetylcholine and light used the same signaling pathways in some part of mouse eye (Q. Wang et al., 2017) and when light stimulus present, ache expression was increased (Murakami et al., 1984). Interestingly, ache was expressed before the acetylcholine started to be produced in the zebrafish embryo (Cousin et al., 2005; Villamizar et al., 2014). I suspected that the relationship between ache and retinal genes could be due to one of either the classical or non-

classical roles of *ache*. Therefore, it was hypothesized that if fish was raised without light, the detrimental effect of the absence of *ache* would not be as present as it were raised in constant light. To test this hypothesis, the embryos were raised under different lighting conditions (before the zebrafish can detect the light at gastrula stage (Tamai et al., 2004; Villamizar et al., 2014; Ziv & Gothilf, 2006). These conditions were constant light (CL), constant dark (CD) and regular day-night cycle (N; 14/10 hours light/dark). The embryos raised in CD were never exposed to the daylight, whereas the embryos in CL were never exposed to darkness. However, the PCA and cluster analysis based on the selected genes did not show any proper correlation.

In rod dominated retina, known to be functionalized to work well under dim light conditions, *rpe65a* dependency were observed, especially at embryonic stages (Hanovice et al., 2019; Schonthaler et al., 2007); and it was known that *rpe65a* had role in circadian rhythm. The expression pattern of *rpe65a* was very similar to that of *ache*, and they were clustered together, conditions-wise (**Figure 4.57**). Therefore, this situation raised the possibility whether *ache* was involved in the processes such as regulation of circadian rhythm. On the other hand, *arr3a* was from the same family of GPCRs and it was involved in many processes including the retinal development (S. Cai et al., 2018). The *arr3a* gene at 3 and 4 dpf CL conditions were further down regulated with respect to other conditions, which might indicate that CL could be more detrimental to eye development at embryonic stage. Moreover, *rom1b* and *arr3a* were clustered together in terms of expression pattern, so they might work together in abnormal lighting conditions. However, I had only access to one set of biological replicates and the study should be repeated with 2 or more replicates, to have statistically significant conclusions. In this pilot study, the effects were most prominent on day 3 of embryonic development thus future studies should focus on this day with more biological replicates.

5.7 The Difference between This Study and Other Available Tools

As mentioned before in the introduction, Expression Atlas (<https://www.ebi.ac.uk/gxa/home>) does not provide TS information for adult fish, and it is restricted to 120 hours embryo (**Supplementary Figure 55**). Bgee (<https://bgee.org/>) is a tool available for enrichment analysis for set of genes and provides more detailed information about the expression profile of given set of genes

in different tissues (from highest enrichment to lower). Bgee had more genes assigned for every tissue (an example was given for brain) compared to our study, which showed that its TS selection criteria was more linear (Supplementary Figure 56). Because they are interested in set of genes, a gene might be assigned to different tissues in different sets, so the assignments of TS genes differ in this tool. C-It-Loci database (<http://www.c-it-loci.uni-frankfurt.de/>) was not created for TS expression of specific genes and they did not have as much as tissue I had for this study. ZFIN (<https://zfin.org/>) provides information for the expression profile of the genes by using such *in situ hybridization* data, and it is possible to combine information retrieved from our study and ZFIN to go and check the expression stage and tissue/organ specific expression of given genes (i.e. *arr3a*, *rpe65a* eye specific) in that database.

These different tools provide information about tissue specific gene/gene sets at adult stage to understand the expression pattern in the embryonic stage (except expression atlas, which provides for other organism such as human and mouse but not for zebrafish). On the other hand, tissue specificity is a term that might be defined after tissues organized. TS might not only provide the information about the genes specific the given tissues but also provide deeper understanding the genes only expressed in embryonic stage and adult stage. So, it is important to define TS in adult tissues and apply it on different stages of zebrafish including embryonic stage.

5.8 Conclusions and Future Perspectives

In this study, we showed that it was possible to combine different datasets of different experiments if the same platform was used to obtain tissue specificity indices. Moreover, the normalization method and metrics were important to decide the tissue specificity of genes. SD method and RPKM were more consistent with each other when compared with TPM. Log transformation while measuring the tissue specificity/differential gene expression had also effect on the rest of the analysis significantly reducing the number of genes assigned as tissue-specific. One of the interesting finding about genes having Tau values was that whether the analysis was done on log transformed or not transformed TS dataset, the genes with tau values smaller than 0.2 (Supplementary Figure 3) indicated that either the reference/housekeeping genes expressed in adult tissue was not expressed in

embryos or they were the genes which had no expression. These housekeeping genes could be further investigated by *in vivo* experiments on adult zebrafish.

When TS genes were tested on the L2FC values of a real dataset, DEG selection criteria might also be important in the results obtained. In this study, a linear cut-off, L2FC, with Tau as TS metric were chosen to decide TS DEGs to find all possible candidates. In calculating tissue averages, arithmetic average was used. Geometric average should need to be tested in the future. In addition, there might be a trend that many of the TS genes were the genes downregulated in ache dataset. There is a possibility of TS genes were the genes where highest expression already found in the tissues they assigned, so it might be because of “being on the top” effect. When you are on the top, it is easier to fall down. On the other hand, there were plenty of genes upregulated and assigned as TS. To address this, further statistical tests comparing fold changes of TS and Non TS genes need to be performed. Further studies with other datasets are also required to make clear conclusion about this.

On the other hand, muscle, retina, (**Table 4.5 - 6** and **Supplementary Table 4-5**) and liver were among the most effected tissues in ache mutant zebrafish. As a novelty, the immune system regulation genes, which were more enriched in the intestine, were also significantly changed their expression pattern when ache was absent. However, the relationship between ache and retinal genes were tested in more detail and future studies with increased replicates and other genes are needed. The light and dark regime responses were not clear and this might be due to not having biological replicates for dark-light study or potentially dynamic expression of genes (due to an inherent circadian rhythm) at the embryonic level (Dekens & Whitmore, 2008). Hence, further investigation of these genes would be done with repeated experimentation and embryos raised in different hours post fertilizations. In addition, it is important to validate these results of *in silico* data on ache (+/-) vs. wt adult zebrafish tissues to gain better insight about the function of ache in later stages. To best our knowledge, the immune system and ache in zebrafish has not studied in zebrafish before. Therefore, to reveal the mechanism behind the immune regulation, further studies were required.

Supplementary

Quality Control of fastq files by FatQC tool

a)

FastQC Report	Brain	Heart	Muscle	Liver	Kidney	Bones	Intestine	Ovary	Testis
Accession Code	SRR1524238_1	SRR1524240_1	SRR1524241_1	SRR1524242_1	SRR1524243_1	SRR1524244_1	SRR1524245_1	SRR1524248_1	SRR1524249_1
Basic Statistics	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	x	/	/	/	x	/	/	x
Per Tile Sequence Quality	!	x	!	!	!	x	/	!	x
Per Sequence Quality Score	/	x	/	/	/	x	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	/	!	!	!	!	!	!	!
Per Base N Content	/	x	/	/	/	/	/	/	!
Sequence LengthDistribution	/	/	/	/	/	/	/	/	/
Sequence Duplication Levels	!	!	x	x	!	x	x	x	!
Overrepresented Sequences	/	/	!	!	!	/	/	/	/
Adaptor Content	/	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x

b)

FastQC Report	Brain	Gill	Heart	Intestine	Kidney	Liver	Muscle	Spleen
Accession Code	SRR1609735_1	SRR1609738_1	SRR1609741_1	SRR1609744_1	SRR1609747_1	SRR1609750_1	SRR1609753_1	SRR1609756_1
Basic Statistics	/	/	/	/	/	/	/	/
Per Base Sequence Quality	x	/	/	/	/	/	/	x
Per Tile Sequence Quality	x	!	x	!	x	!	!	x
Per Sequence Quality Score	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	x	x	!	!	!	x	!
Per Base N Content	/	/	/	/	/	/	/	/
Sequence LengthDistribution	/	/	/	/	/	/	/	/
Sequence Duplication Levels	/	!	x	x	x	x	x	x
Overrepresented Sequences	/	/	/	/	/	/	/	!
Adaptor Content	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x

c)

FastQC Report	Skin	Skin	Skin	Skin	Skin	Skin	Skin	Skin	Skin
Accession Code	SRR850589	SRR850590	SRR850591	SRR850592	SRR850593	SRR850594	SRR850595	SRR850596	SRR850597
Basic Statistics	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	x	/	/	/	/	/	/	/
Per Tile Sequence Quality	/	x	/	/	x	!	!	!	!
Per Sequence Quality Score	/	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	/	!	!	!	!	/	!	!
Per Base N Content	/	/	/	/	/	/	/	/	/
Sequence LengthDistribution	/	/	/	/	/	/	/	/	/
Sequence Duplication Levels	x	!	x	x	x	x	x	x	x
Overrepresented Sequences	/	/	/	!	/	/	/	/	/
Adaptor Content	/	!	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x

d)

FastQC Report	Retina	Retina	Retina	Retina	Retina	Retina
Accession Code	SRR2567887_1	SRR2567888_1	SRR2567889_1	SRR2567890_1	SRR2567891_1	SRR2567892_1
Basic Statistics	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/
Per Tile Sequence Quality	x	x	x	x	x	x
Per Sequence Quality Score	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x
Per Sequence GC Content	!	/	!	!	!	!
Per Base N Content	/	/	/	/	/	/
Sequence LengthDistribution	/	/	/	/	/	/
Sequence Duplication Levels	/	!	/	/	/	/
Overrepresented Sequences	!	x	!	!	!	!
Adaptor Content	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x

e)

FastQC Report	Brain	Brain	Brain	Brain	Brain	Liver	Liver	Liver	Liver	Liver	Skin	Skin	Skin	Skin	Skin
Accession Code	SRR2750987	SRR2750988	SRR2750989	SRR2750990	SRR2750991	SRR2751010	SRR2751011	SRR2751012	SRR2751013	SRR2751014	SRR2751035	SRR2751036	SRR2751037	SRR2751038	SRR2751039
Basic Statistics	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Tile Sequence Quality	/	/	/	/	/	/	/	/	/	/	/	/	/	!	/
Per Sequence Quality Score	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	/	/	/	/	/	!	/	/	/	!	!	!	!	!	!
Per Base N Content	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Sequence LengthDistribution	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Sequence Duplication Levels	!	!	!	!	!	x	x	x	x	x	x	x	x	x	x
Overrepresented Sequences	!	/	/	/	/	/	/	!	/	/	!	/	!	/	/
Adaptor Content	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Supplementary Figure 1: The quality control scores of fastq files retrieved from European Bioinformatics Institute which were trimmed and improved by TrimGalore used in the analysis a) PRJNA263496 datasets b) PRJNA255848 datasets c) PRJNA203029 datasets d) PRJNA297904 datasets e) PRJNA299585 datasets

FastQC Report	Brain	Gill	Heart	Muscle	Liver	Kidney	Bones	Intestine	Ovary	Testis
Accession Code	SRR1524238_1	SRR1524239_1	SRR1524240_1	SRR1524241_1	SRR1524242_1	SRR1524243_1	SRR1524244_1	SRR1524245_1	SRR1524248_1	SRR1524249_1
Basic Statistics	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	/	x	x	/	/	/	/	/	/
Per Tile Sequence Quality	/	/	x	/	!	!	x	/	!	x
Per Sequence Quality Score	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	!	/	!	!	!	!	!	!	!
Per Base N Content	/	/	x	/	/	/	/	/	/	/
Sequence LengthDistribution	!	!	!	!	!	!	!	!	!	!
Sequence Duplication Levels	/	x	!	!	x	/	x	x	x	!
Overrepresented Sequences	/	/	/	/	/	/	/	/	/	/
Adaptor Content	/	/	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x	x

a)

FastQC Report	Brain	Gill	Heart	Intestine	Kidney	Liver	Muscle	Spleen
Accession Code	SRR1609735_1	SRR1609738_1	SRR1609741_1	SRR1609744_1	SRR1609747_1	SRR1609750_1	SRR1609753_1	SRR1609756_1
Basic Statistics	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/	/	/
Per Tile Sequence Quality	x	!	x	x	!	!	!	x
Per Sequence Quality Score	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	x	x	!	!	!	!	!
Per Base N Content	/	/	/	!	/	/	/	/
Sequence LengthDistribution	!	!	!	!	!	!	!	!
Sequence Duplication Levels	/	!	x	x	x	x	x	x
Overrepresented Sequences	/	/	/	/	/	/	/	/
Adaptor Content	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x

b)

FastQC Report	Skin	Skin	Skin	Skin	Skin	Skin	Skin	Skin	Skin
Accession Code	SRR850589	SRR850590	SRR850591	SRR850592	SRR850593	SRR850594	SRR850595	SRR850596	SRR850597
Basic Statistics	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/	/	/	/
Per Tile Sequence Quality	/	x	/	/	x	!	!	!	x
Per Sequence Quality Score	/	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	!	/	!	!	!	!	!	!	!
Per Base N Content	/	/	/	/	/	/	/	/	/
Sequence LengthDistribution	!	!	!	!	!	!	!	!	!
Sequence Duplication Levels	!	!	x	x	x	x	x	x	x
Overrepresented Sequences	/	/	/	/	/	/	/	/	/
Adaptor Content	/	!	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x

c)

FastQC Report	Retina	Retina	Retina	Retina	Retina	Retina
Accession Code	SRR2567887_1	SRR2567888_1	SRR2567889_1	SRR2567890_1	SRR2567891_1	SRR2567892_1
Basic Statistics	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/
Per Tile Sequence Quality	/	/	/	/	/	/
Per Sequence Quality Score	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x
Per Sequence GC Content	!	/	!	!	!	!
Per Base N Content	/	/	/	/	/	/
Sequence LengthDistribution	!	!	!	!	!	!
Sequence Duplication Levels	/	/	/	/	/	/
Overrepresented Sequences	/	/	/	/	/	/
Adaptor Content	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x

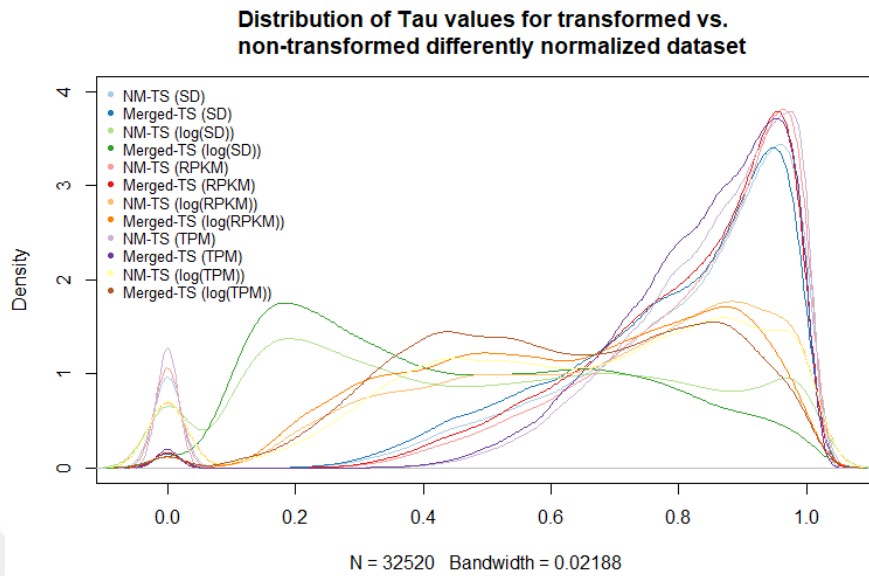
d)

FastQC Report	Brain	Brain	Brain	Brain	Brain	Liver	Liver	Liver	Liver	Liver	Skin	Skin	Skin	Skin	Skin
Accession Code	SRR2750987	SRR2750988	SRR2750989	SRR2750990	SRR2750991	SRR2751010	SRR2751011	SRR2751012	SRR2751013	SRR2751014	SRR2751035	SRR2751036	SRR2751037	SRR2751038	SRR2751039
Basic Statistics	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Quality	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Tile Sequence Quality	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Sequence Quality Score	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Per Base Sequence Content	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Per Sequence GC Content	/	/	/	/	/	!	/	/	!	!	!	!	!	!	!
Per Base N Content	/	/	/	/	/	!	/	/	/	/	/	/	/	/	/
Sequence LengthDistribution	/	/	/	/	/	!	!	!	!	!	!	!	!	!	!
Sequence Duplication Levels	!	!	!	!	!	x	x	x	x	x	x	x	x	x	x
Overrepresented Sequences	!	!	!	!	!	/	/	/	/	/	/	/	/	/	/
Adaptor Content	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Kmer Content	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

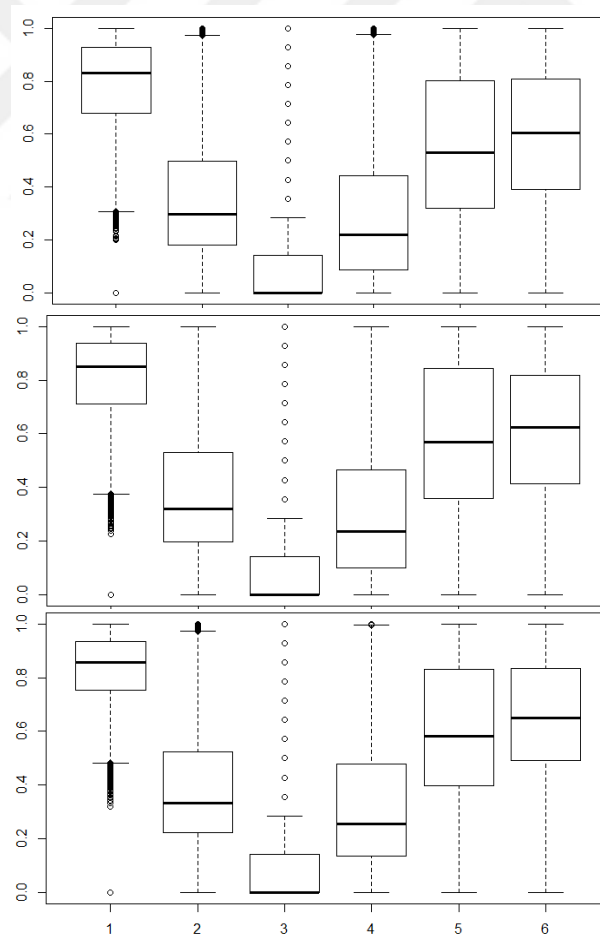
e)

Supplementary Figure 2: The quality control scores of fastq files retrieved from European Nucleotide Archive used in the analysis a) PRJNA263496 datasets b) PRJNA255848 datasets c) PRJNA203029 datasets d) PRJNA297904 datasets e) PRJNA299585 datasets

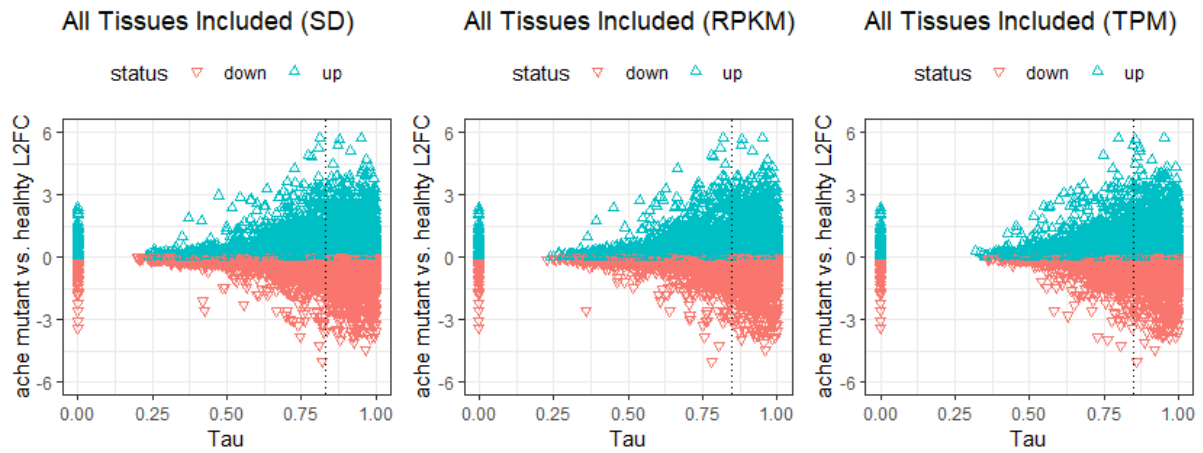
How did the Tau values of transformed vs. non transformed and merged vs. original TS data distributed?



Supplementary Figure 3: Tau score distributions of not merged TS data and TS data merged with AChE dataset.



Supplementary Figure 4: The table showing the distributions of Tau(1), Tsi(2), Counts(3), Hg(4), Spm(5) and Gini(6) metrics values across SD(upper), RPKM(middle) and TPM(bottom) normalized datasets in the given order.



Supplementary Figure 5: The distributions of up and downregulated genes for 3 methods. Vertical dashed lines were drawn based on median of Tau values.

How similar the distributions of up and down regulated genes when different DEG cut-offs were chosen?

Supplementary Table 1: KS test results to see the distribution difference between up and downregulated genes (DEG cut off log2fold change).

Tsi	Brain	0.191919	0.108156	0.129556	0.502278	0.233833	0.004568
	Liver	0.235023	0.001186	0.249227	0.000482	0.10873	0.896117
	Skin	0.127569	0.276777	0.14298	0.200934	0.108006	0.503406
	Gill	0.113812	0.395046	0.145274	0.169788	0.16556	0.024351
	Heart	0.135684	0.64267	0.126615	0.682941	0.141107	0.797432
	Intestine	0.371309	7.01E-11	0.355997	1.26E-09	0.34386	1.16E-06
	Kidney	0.262876	0.048327	0.305738	0.021856	0.297673	0.026302
	Muscle	0.217545	0.010277	0.245998	0.004725	0.308547	0.067267
	Spleen	0.132143	0.336357	0.147059	0.278967	0.318129	0.003023
	Bone	0.342342	0.016846	0.431319	0.008768	0.36996	0.016143
	Ovary	0.15973	0.182952	0.073838	0.942409	0.104241	0.441802
	Embryos	0.160418	0.350425	0.15757	0.37347	0.179748	0.117034
	UFEggs	0.125683	0.832579	0.093718	0.923943	0.384146	0.048956
	Retina	0.375357	8.79E-06	0.376009	7.11E-05	0.436344	1.09E-07
	Testis	0.128302	0.041402	0.094575	0.162212	0.09284	0.06019
Hg	Brain	0.131313	0.501951	0.133509	0.463301	0.29721	0.000108
	Liver	0.248848	0.000482	0.264582	0.000167	0.186121	0.288197
	Skin	0.174687	0.049267	0.152455	0.146805	0.105815	0.530115
	Gill	0.097238	0.597796	0.11269	0.448315	0.184738	0.008266
	Heart	0.215385	0.125832	0.232189	0.063063	0.260436	0.115964
	Intestine	0.415715	1.57E-13	0.415681	5.72E-13	0.389744	1.94E-08
	Kidney	0.307358	0.012323	0.347131	0.005922	0.315315	0.015502
	Muscle	0.245658	0.00241	0.242461	0.005615	0.393162	0.008292
	Spleen	0.232143	0.00828	0.233193	0.015167	0.446784	4.95E-06
	Bone	0.335412	0.020739	0.431319	0.008768	0.399194	0.008193
	Ovary	0.146875	0.263867	0.08096	0.889716	0.124023	0.23925
	Embryos	0.20619	0.113551	0.18684	0.190567	0.24478	0.010355
	UFEggs	0.308495	0.025751	0.178318	0.253835	0.460366	0.009595
	Retina	0.371338	1.14E-05	0.350756	0.000269	0.437001	1.04E-07
	Testis	0.109434	0.119072	0.100984	0.114128	0.094194	0.054302
Spm	Brain	0.171717	0.193377	0.13307	0.467556	0.178681	0.05738
	Liver	0.138029	0.154081	0.163831	0.054677	0.204989	0.19126
	Skin	0.119799	0.348507	0.153316	0.142529	0.098007	0.628861
	Gill	0.122768	0.304718	0.115334	0.418765	0.131269	0.12513
	Heart	0.137821	0.623017	0.111296	0.822005	0.163793	0.626453
	Intestine	0.261605	1.29E-05	0.256616	3.32E-05	0.183401	0.033615
	Kidney	0.237458	0.095874	0.254918	0.086577	0.209835	0.23208
	Muscle	0.202532	0.020745	0.234177	0.008332	0.222222	0.327954
	Spleen	0.096429	0.731069	0.086134	0.87755	0.231579	0.062712
	Bone	0.288288	0.065947	0.373626	0.032817	0.27621	0.13053
	Ovary	0.087056	0.853227	0.077586	0.91693	0.069183	0.896087
	Embryos	0.163034	0.331274	0.134023	0.581053	0.178742	0.1208
	UFEggs	0.087432	0.989194	0.108409	0.821963	0.272866	0.294076
	Retina	0.369963	1.25E-05	0.395243	2.43E-05	0.406678	9.82E-07
	Testis	0.116981	0.079627	0.09987	0.121524	0.079273	0.155426
Gini	Brain	0.131313	0.501951	0.135266	0.446484	0.29721	0.000108
	Liver	0.246324	0.00057	0.270879	0.000107	0.180365	0.323794
	Skin	0.175188	0.04823	0.155326	0.132943	0.098692	0.620054
	Gill	0.085374	0.754191	0.100463	0.597224	0.171964	0.017201
	Heart	0.223932	0.100606	0.209487	0.119925	0.286751	0.063366
	Intestine	0.423585	4.94E-14	0.407376	1.79E-12	0.391903	1.58E-08
	Kidney	0.288294	0.02272	0.315164	0.016472	0.313814	0.016235
	Muscle	0.251987	0.001697	0.240413	0.006198	0.437607	0.002209
	Spleen	0.226786	0.010636	0.220588	0.025252	0.436257	9.15E-06
	Bone	0.335412	0.020739	0.431319	0.008768	0.40121	0.007041
	Ovary	0.138603	0.326595	0.086207	0.840305	0.120501	0.269223
	Embryos	0.180471	0.221857	0.185299	0.198044	0.254214	0.006848
	UFEggs	0.327372	0.014948	0.249747	0.03738	0.484756	0.005341
	Retina	0.339609	8.24E-05	0.341213	0.000434	0.421784	3.27E-07
	Testis	0.128302	0.041402	0.107806	0.076514	0.10562	0.021469

Supplementary Table 2: KS test results to see the distribution difference between up and downregulated genes (DEG cut off pval smaller than 0.1).

KS Test		SD		RPKM		TPM	
		D	pval	D	pval	D	pval
Tau	Brain	0.524955	3.43E-06	0.486111	1.44E-05	0.540451	1.88E-10
	Liver	0.361111	2.52E-05	0.402174	2.68E-06	0.256993	0.188784
	Skin	0.356593	0.003955	0.366324	0.003664	0.311494	0.01473
	Gill	0.202128	0.870609	0.28254	0.576876	0.243357	0.338522
	Heart	0.506672	7.40E-05	0.445574	0.000296	0.294974	0.044549
	Intestine	0.531005	7.28E-11	0.49747	3.09E-09	0.528622	1.09E-07
	Kidney	0.555556	0.060152	0.533333	0.134725	0.813953	0.000134
	Muscle	0.468117	4.24E-09	0.423343	1.99E-07	0.380174	0.018244
	Spleen	0.258621	0.64282	0.377551	0.561943	0.5	0.045323
	Bone	0.493056	0.084271	0.428571	0.257172	0.458111	0.013073
	Ovary	0.225543	0.621258	0.275862	0.222135	0.3525	0.00037
	Embryos	0.263482	0.085129	0.155303	0.582184	0.23878	0.051006
	UFEggs	0.41904	0.004177	0.332048	0.007966	0.564935	0.00104
	Retina	0.574603	1.05E-10	0.551657	1.08E-06	0.608849	7.06E-13
	Testis	0.717105	0.002793	0.377622	0.100231	0.207951	0.162366
Tsi	Brain	0.524955	3.43E-06	0.486111	1.44E-05	0.540451	1.88E-10
	Liver	0.361111	2.52E-05	0.402174	2.68E-06	0.256993	0.188784
	Skin	0.356593	0.003955	0.366324	0.003664	0.311494	0.01473
	Gill	0.202128	0.870609	0.28254	0.576876	0.243357	0.338522
	Heart	0.506672	7.40E-05	0.445574	0.000296	0.294974	0.044549
	Intestine	0.531005	7.28E-11	0.49747	3.09E-09	0.528622	1.09E-07
	Kidney	0.555556	0.060152	0.533333	0.134725	0.813953	0.000134
	Muscle	0.468117	4.24E-09	0.423343	1.99E-07	0.380174	0.018244
	Spleen	0.258621	0.64282	0.377551	0.561943	0.5	0.045323
	Bone	0.493056	0.084271	0.428571	0.257172	0.458111	0.013073
	Ovary	0.225543	0.621258	0.275862	0.222135	0.3525	0.00037
	Embryos	0.263482	0.085129	0.155303	0.582184	0.23878	0.051006
	UFEggs	0.41904	0.004177	0.332048	0.007966	0.564935	0.00104
	Retina	0.574603	1.05E-10	0.551657	1.08E-06	0.608849	7.06E-13
	Testis	0.717105	0.002793	0.377622	0.100231	0.207951	0.162366
Hg	Brain	0.43672	0.000218	0.430556	0.000201	0.514768	1.91E-09
	Liver	0.392593	3.12E-06	0.437888	2.06E-07	0.309441	0.067648
	Skin	0.349451	0.005118	0.411453	0.000708	0.394828	0.000746
	Gill	0.276596	0.54012	0.349206	0.323146	0.352914	0.050256
	Heart	0.462469	0.000389	0.452459	0.000221	0.414021	0.001129
	Intestine	0.587883	2.29E-13	0.568944	4.13E-12	0.637992	3.12E-11
	Kidney	0.5	0.117728	0.466667	0.252393	0.790698	0.000248
	Muscle	0.468492	4.10E-09	0.414108	3.98E-07	0.511983	0.000399
	Spleen	0.431034	0.107143	0.459184	0.316377	0.738095	0.000376
	Bone	0.604167	0.015884	0.494505	0.136558	0.497326	0.003671
	Ovary	0.307065	0.26878	0.206897	0.572177	0.390417	5.15E-05
	Embryos	0.191577	0.359758	0.195707	0.303233	0.295276	0.007504
	UFEggs	0.394303	0.008163	0.2777	0.041508	0.590909	0.000466
	Retina	0.580952	5.97E-11	0.582846	1.75E-07	0.677225	1.33E-15
	Testis	0.75	0.001068	0.496503	0.012842	0.13685	0.623606
Spm	Brain	0.464349	7.00E-05	0.5	7.05E-06	0.437351	6.70E-07
	Liver	0.264815	0.004728	0.32764	0.000261	0.181818	0.576114
	Skin	0.354396	0.004304	0.354497	0.005489	0.326437	0.009114
	Gill	0.148936	0.987644	0.226984	0.820073	0.233566	0.384362
	Heart	0.543786	1.33E-05	0.467869	0.000115	0.300265	0.039287
	Intestine	0.409551	1.56E-06	0.368438	3.45E-05	0.342181	0.001995
	Kidney	0.5	0.117728	0.533333	0.134725	0.813953	0.000134
	Muscle	0.471305	3.22E-09	0.416186	3.58E-07	0.285403	0.134126
	Spleen	0.198276	0.898188	0.357143	0.629058	0.357143	0.286704
	Bone	0.493056	0.084271	0.428571	0.257172	0.426025	0.025912
	Ovary	0.288043	0.333856	0.310345	0.12286	0.325	0.001356
	Embryos	0.284027	0.051604	0.17803	0.413325	0.165157	0.332923
	UFEggs	0.375562	0.014098	0.30014	0.021786	0.49026	0.007511
	Retina	0.526984	5.51E-09	0.54191	1.79E-06	0.546506	2.54E-10
	Testis	0.664474	0.007308	0.347319	0.157156	0.164628	0.394005
Gini	Brain	0.407308	0.000726	0.430556	0.000201	0.523941	8.74E-10
	Liver	0.392593	3.12E-06	0.445135	1.19E-07	0.309441	0.067648
	Skin	0.364835	0.002988	0.430128	0.000328	0.385057	0.0011
	Gill	0.276596	0.54012	0.338095	0.360256	0.3669	0.037391
	Heart	0.453711	0.000541	0.444262	0.000309	0.416667	0.001032
	Intestine	0.596579	8.80E-14	0.582226	1.07E-12	0.651103	9.39E-12
	Kidney	0.5	0.117728	0.466667	0.252393	0.790698	0.000248
	Muscle	0.442048	3.68E-08	0.402332	9.66E-07	0.511983	0.000399
	Spleen	0.431034	0.107143	0.459184	0.316377	0.761905	0.000202
	Bone	0.604167	0.015884	0.571429	0.054076	0.467023	0.007935
	Ovary	0.307065	0.26878	0.172414	0.791228	0.412083	1.52E-05
	Embryos	0.191577	0.359758	0.209596	0.232741	0.318898	0.002966
	UFEggs	0.29985	0.079757	0.298387	0.023493	0.545455	0.00171
	Retina	0.565079	2.40E-10	0.55653	8.07E-07	0.666164	2.33E-15
	Testis	0.75	0.001068	0.496503	0.012842	0.152396	0.497975

Supplementary Table 3: KS test results to see the distribution difference between up and downregulated genes (DEG cut off pval smaller than 0.05).

KS Test		SD		RPKM		TPM	
		D	pval	D	pval	D	pval
Tau	Brain	0.493734	0.010072	0.430909	0.016918	0.44709	0.000758
	Liver	0.360089	0.001105	0.41844	0.000125	0.315789	0.135961
	Skin	0.397251	0.023757	0.41573	0.012273	0.313253	0.097921
	Gill	0.528571	0.465571	0.523077	0.477612	0.660377	0.015037
	Heart	0.596639	0.000158	0.518617	0.000608	0.47549	0.004802
	Intestine	0.689931	7.29E-13	0.649781	4.47E-11	0.625138	1.78E-08
	Kidney	0.55	0.207039	0.604167	0.235294	0.769231	0.002304
	Muscle	0.515615	3.08E-09	0.459091	6.65E-07	0.351852	0.065331
	Spleen	0.339394	0.586311	0.425287	0.575403	0.4	0.249538
	Bone	0.480519	0.207642	0.583333	0.139194	0.576923	0.002456
	Ovary	0.5	0.165958	0.354167	0.299727	0.275434	0.099578
	Embryos	0.202703	0.541625	0.134188	0.878184	0.371951	0.004311
	UFEggs	0.413534	0.196795	0.30625	0.323079	0.666667	0.117241
	Retina	0.631491	1.29E-08	0.622807	7.25E-06	0.674561	8.59E-11
Tsi	Testis	0.618182	0.105769	0.254386	0.853744	0.329837	0.146439
	Brain	0.493734	0.010072	0.430909	0.016918	0.44709	0.000758
	Liver	0.360089	0.001105	0.41844	0.000125	0.315789	0.135961
	Skin	0.397251	0.023757	0.41573	0.012273	0.313253	0.097921
	Gill	0.528571	0.465571	0.523077	0.477612	0.660377	0.015037
	Heart	0.596639	0.000158	0.518617	0.000608	0.47549	0.004802
	Intestine	0.689931	7.29E-13	0.649781	4.47E-11	0.625138	1.78E-08
	Kidney	0.55	0.207039	0.604167	0.235294	0.769231	0.002304
	Muscle	0.515615	3.08E-09	0.459091	6.65E-07	0.351852	0.065331
	Spleen	0.339394	0.586311	0.425287	0.575403	0.4	0.249538
	Bone	0.480519	0.207642	0.583333	0.139194	0.576923	0.002456
	Ovary	0.5	0.165958	0.354167	0.299727	0.275434	0.099578
	Embryos	0.202703	0.541625	0.134188	0.878184	0.371951	0.004311
	UFEggs	0.413534	0.196795	0.30625	0.323079	0.666667	0.117241
Hg	Retina	0.631491	1.29E-08	0.622807	7.25E-06	0.674561	8.59E-11
	Testis	0.618182	0.105769	0.254386	0.853744	0.329837	0.146439
	Brain	0.43609	0.031202	0.423636	0.020534	0.473545	0.000265
	Liver	0.44398	2.07E-05	0.501773	1.55E-06	0.365497	0.056246
	Skin	0.435052	0.009206	0.426966	0.009254	0.399008	0.015531
	Gill	0.585714	0.36385	0.538462	0.448666	0.688679	0.009388
	Heart	0.573109	0.000345	0.506649	0.000894	0.467647	0.005856
	Intestine	0.731693	1.33E-14	0.702314	4.08E-13	0.718715	2.96E-11
	Kidney	0.6	0.13834	0.479167	0.504644	0.730769	0.004668
	Muscle	0.471243	9.67E-08	0.434091	3.34E-06	0.513889	0.001325
	Spleen	0.478788	0.203318	0.597701	0.190323	0.633333	0.010755
	Bone	0.480519	0.207642	0.541667	0.192474	0.637363	0.000424
	Ovary	0.7	0.012021	0.25	0.721644	0.368486	0.00968
	Embryos	0.154791	0.834224	0.176923	0.605938	0.422474	0.000677
Spm	UFEggs	0.398496	0.231795	0.265625	0.494033	0.62963	0.16601
	Retina	0.638139	7.32E-09	0.630117	4.74E-06	0.739496	3.69E-13
	Testis	0.6	0.125	0.403509	0.361355	0.282051	0.293469
	Brain	0.303258	0.251605	0.385455	0.043455	0.410053	0.002634
	Liver	0.215719	0.130819	0.306147	0.0114	0.184211	0.715843
	Skin	0.40756	0.018811	0.375702	0.031327	0.265769	0.225388
	Gill	0.471429	0.647887	0.446154	0.717322	0.481132	0.155169
	Heart	0.609244	0.000103	0.527926	0.000446	0.40098	0.027642
	Intestine	0.51659	3.99E-07	0.458099	1.49E-05	0.410853	0.001047
	Kidney	0.45	0.44777	0.604167	0.235294	0.807692	0.001066
	Muscle	0.517934	2.56E-09	0.482576	1.31E-07	0.199074	0.61235
	Spleen	0.339394	0.586311	0.333333	0.836694	0.3	0.576666
	Bone	0.480519	0.207642	0.708333	0.042624	0.5	0.012969
	Ovary	0.375	0.485306	0.4375	0.113179	0.238213	0.207996
Gini	Embryos	0.256757	0.268477	0.176923	0.605938	0.333624	0.014336
	UFEggs	0.413534	0.196795	0.29375	0.367171	0.740741	0.059113
	Retina	0.589946	1.52E-07	0.562865	7.95E-05	0.603514	1.31E-08
	Testis	0.527273	0.226648	0.263158	0.82218	0.363636	0.080186
	Brain	0.43609	0.031202	0.429091	0.017933	0.509259	6.44E-05
	Liver	0.44398	2.07E-05	0.512411	8.26E-07	0.365497	0.056246
	Skin	0.424742	0.01189	0.420646	0.010857	0.400425	0.015165
	Gill	0.614286	0.317684	0.569231	0.393487	0.650943	0.017448
	Heart	0.608403	0.000106	0.517287	0.000635	0.490196	0.00326
	Intestine	0.731693	1.33E-14	0.713258	1.65E-13	0.719823	1.63E-11
	Kidney	0.6	0.13834	0.479167	0.504644	0.730769	0.004668
	Muscle	0.471243	9.67E-08	0.434091	3.34E-06	0.513889	0.001325
	Spleen	0.478788	0.203318	0.597701	0.190323	0.666667	0.005934
	Bone	0.480519	0.207642	0.666667	0.060606	0.598901	0.001245
	Ovary	0.7	0.012021	0.229167	0.801688	0.378412	0.007282
	Embryos	0.154791	0.834224	0.193162	0.49541	0.434669	0.000418
	UFEggs	0.319549	0.481017	0.296875	0.357864	0.62963	0.16601
	Retina	0.650187	3.27E-09	0.590643	2.60E-05	0.723071	1.56E-12
	Testis	0.6	0.125	0.403509	0.361355	0.343823	0.119702

Were DEG genes enriched for TS genes?

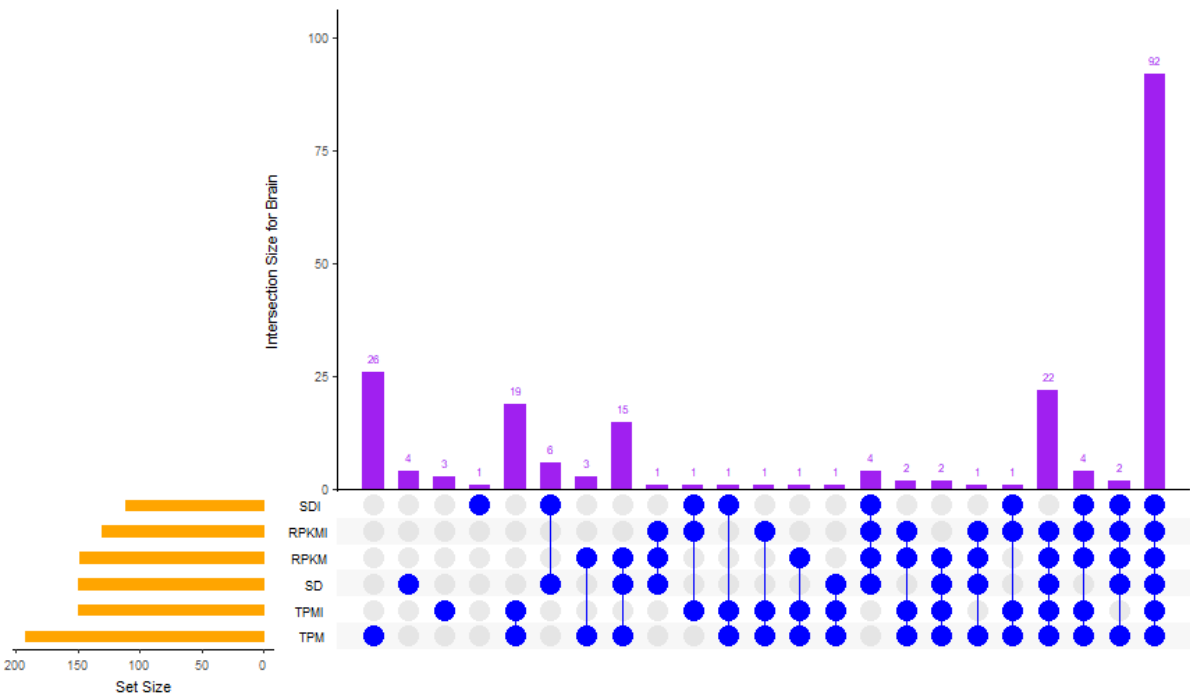
Supplementary Table 4: Fishers Exact Test results and corresponding Odds Ratios for DEG defined as $pval < 0.05$ and TS defined as $Tau > 0.8$

TS>0.8 and pval<0.05		SD		RPKM		TPM	
Tissue	Status	pvalue	OR	pvalue	OR	pvalue	OR
Brain	Up	0.121181	2.207625	0.415899	1.506413	0.676157	0.888076
	Down	0.434153	1.961303	0.321109	2.049793	0.146898	3.224497
Liver	Up	0.000982	2.10242	0.002257	2.012795	0.537394	0.771354
	Down	4.31E-05	5.548678	0.000125	6.887169	0.081823	Inf
Skin	Up	0.830262	0.947923	0.422766	0.82458	0.713123	0.910994
	Down	0.007246	9.226005	0.062375	4.141717	0.174823	2.898631
Gill	Up	0.373669	0.792957	0.350162	0.762386	0.67979	0.899542
	Down	0.540248	Inf	1	Inf	0.177597	Inf
Heart	Up	2.07E-05	2.769818	1.08E-06	3.071856	4.74E-05	3.172599
	Down	0.00011	Inf	0.000176	16.53556	0.001324	Inf
Intestine	Up	1	0.989552	0.815329	1.065881	0.127124	0.594947
	Down	2.17E-06	9.3299	1.24E-05	8.123385	0.001734	6.78006
Kidney	Up	0.001672	0.180761	0.007638	0.219562	0.154365	0.551102
	Down	1	1.467032	1	0.830479	0.182651	Inf
Muscle	Up	4.13E-13	4.331905	4.83E-10	4.075289	0.051583	3.252729
	Down	2.24E-14	31.56665	1.39E-08	10.70946	0.008045	Inf
Spleen	Up	0.283285	1.468538	1	1.048523	0.117811	2.048005
	Down	0.654714	2.037141	1	2.010072	0.097118	Inf
Bones	Up	0.430642	0.462913	0.378688	0.445071	1	1.123937
	Down	0.725322	1.702054	0.191613	Inf	0.034644	Inf
Ovary	Up	0.721529	1.231101	0.597339	0.618398	0.697901	0.887403
	Down	0.708167	1.386609	0.763991	0.740687	0.324647	1.560349
Embryos	Up	7.89E-05	6.972592	0.070091	2.076046	0.030021	1.78406
	Down	0.000261	6.205263	0.004154	3.700212	0.000285	6.771135
Ufeggs	Up	0.864523	0.925106	0.032832	0.565429	0.281036	0.615521
	Down	1	1.330352	1	1.319524	1	Inf
Retina	Up	0.849336	0.877859	0.806168	1.224555	0.266672	0.664481
	Down	2.28E-10	17.88978	3.35E-05	8.990623	1.09E-07	Inf
Testis	Up	1	1.499358	0.124305	0.491502	3.84E-05	0.351008
	Down	0.037339	0.114915	0.07837	0.223223	1	1.011378

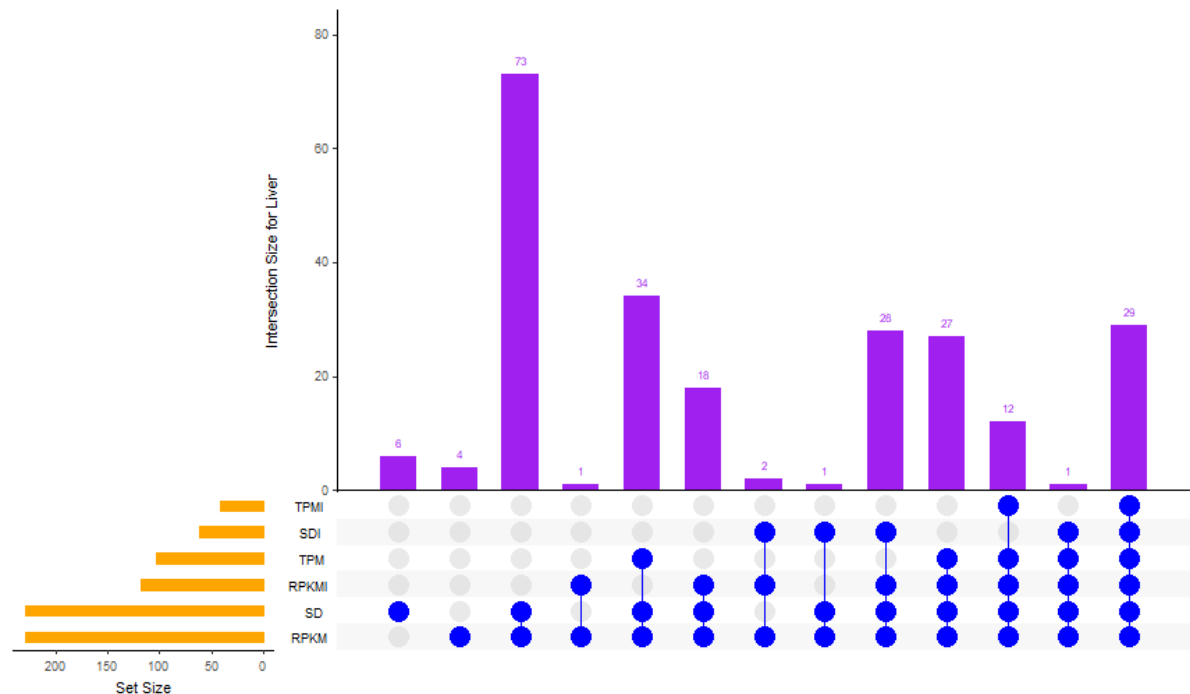
Supplementary Table 13: Normalized expression of ache across different tissues. The top 3 tissues where expressed most were highlighted.

ache	Brain	Liver	Skin	Gill	Heart	Intestine	Kidney	Muscle	Spleen	Bone	Embryos	UFEggs	Ovary	Testis	Retina
SD	1952.3	370.1	471.5	262.5	730.1	637.3	44.4	2810.5	159.9	1372.8	432.0	1.2	3.5	79.2	1814.2
RPKM	31.1	5.7	7.0	3.9	12.1	9.6	0.6	39.8	2.2	18.2	7.3	0.0	0.1	1.9	23.9
TPM	114.1	5.1	15.0	9.5	16.9	12.5	1.1	29.6	3.0	37.8	25.4	0.0	0.3	9.2	77.5

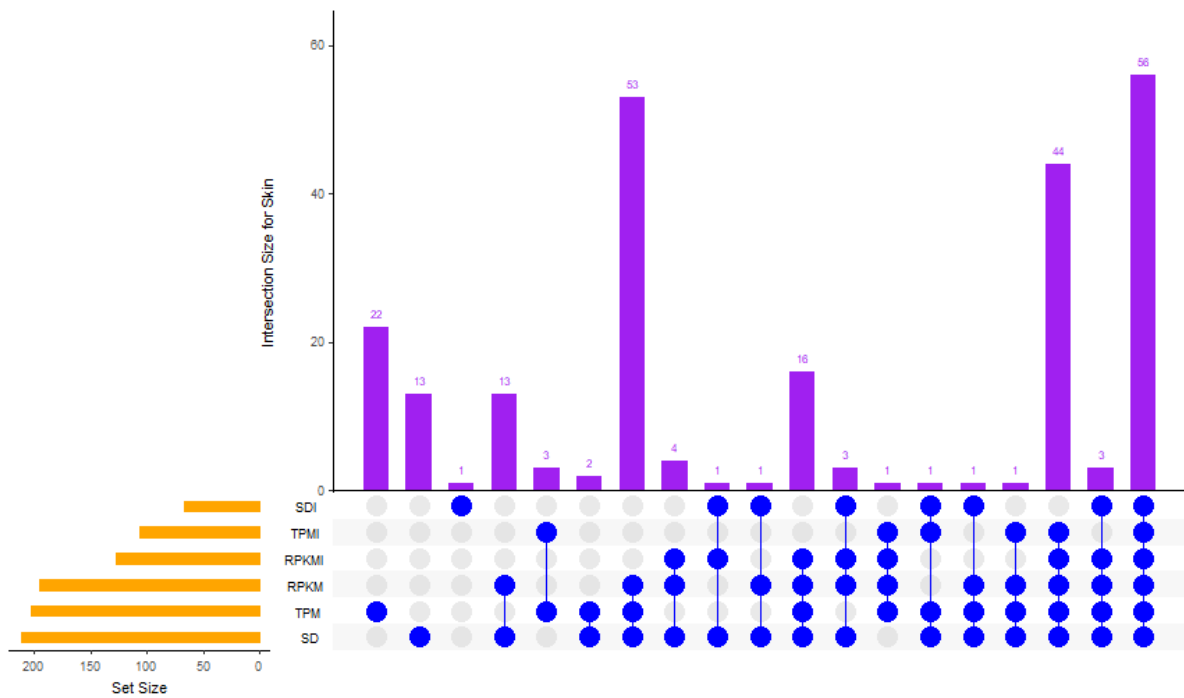
Upset Intersection Graphs for log2 transformed vs. non-transformed normalized counts on abs (L2FC)>0.5849 DEGs



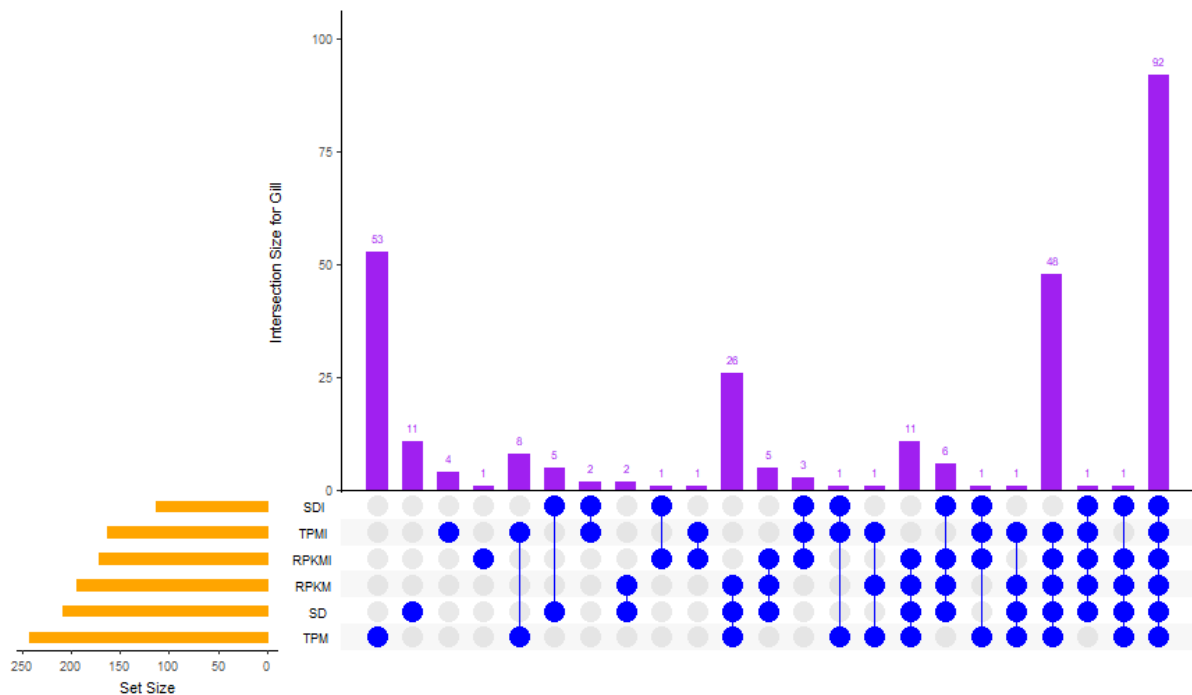
Supplementary Figure 6: Log transformed comparisons for log2fc dependent brain specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



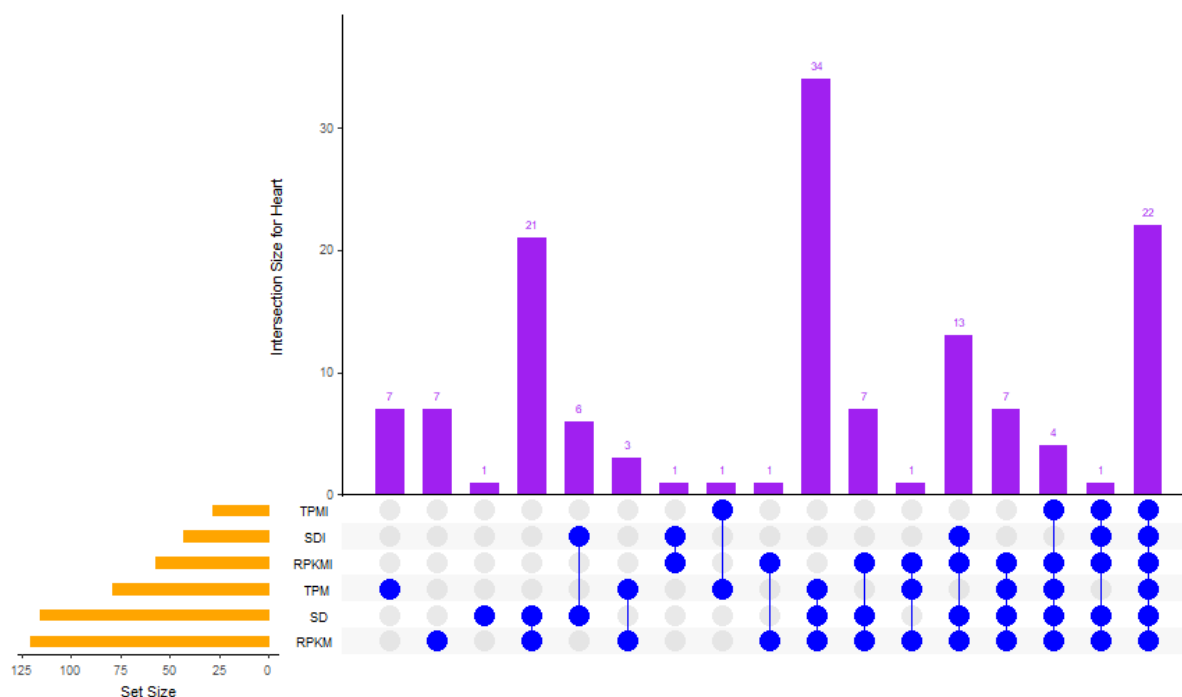
Supplementary Figure 7: Log transformed comparisons for log2fc dependent liver specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



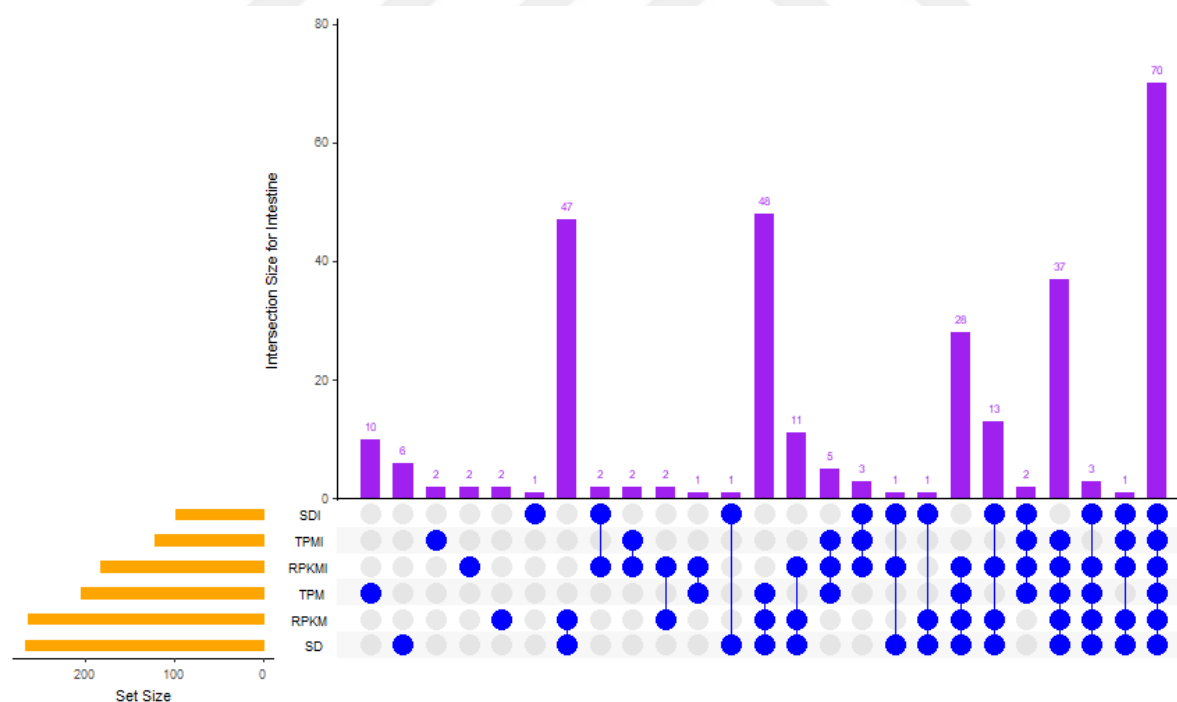
Supplementary Figure 8: Log transformed comparisons for log2fc dependent skin specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



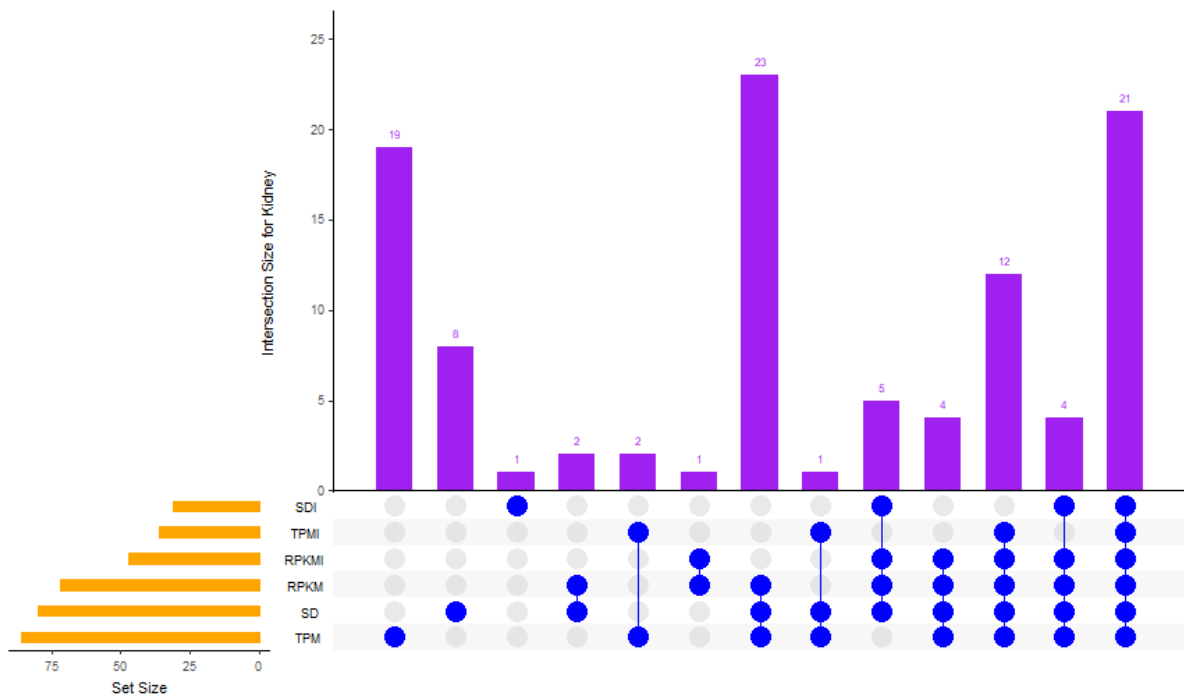
Supplementary Figure 9: Log transformed comparisons for log2fc dependent gill specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



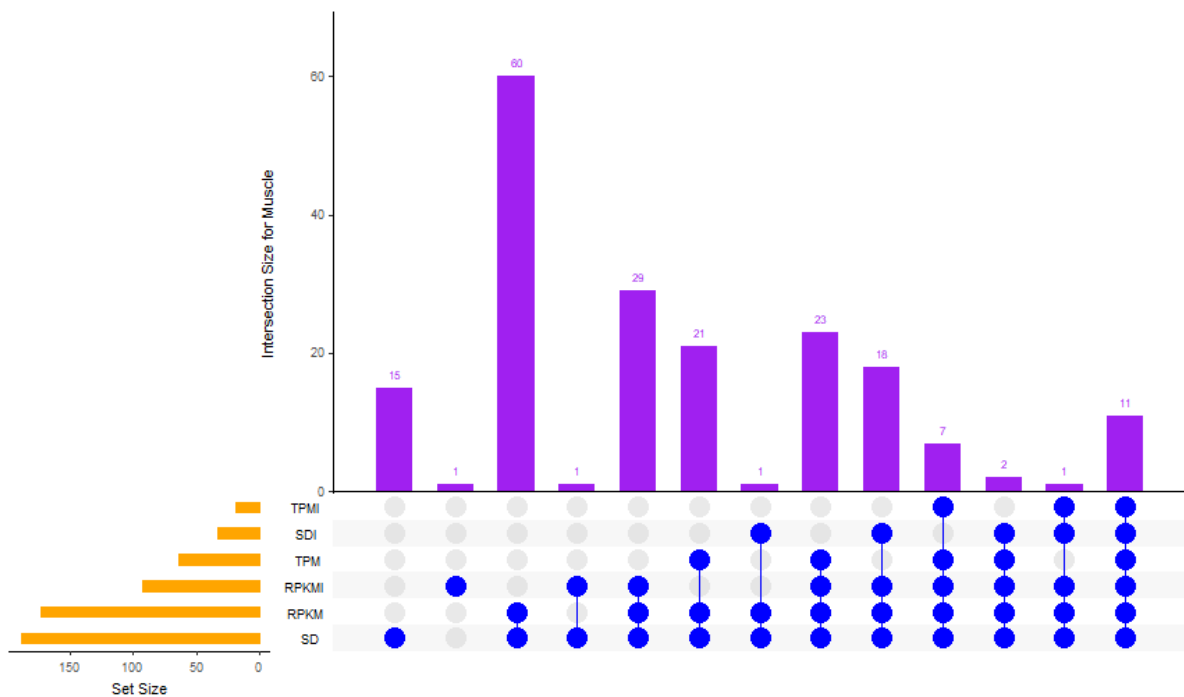
Supplementary Figure 10: Log transformed comparisons for log₂fc dependent heart specific DEGs of AChE dataset (for TS, Tau>0.8). Letter L represents log transformation on normalized counts.



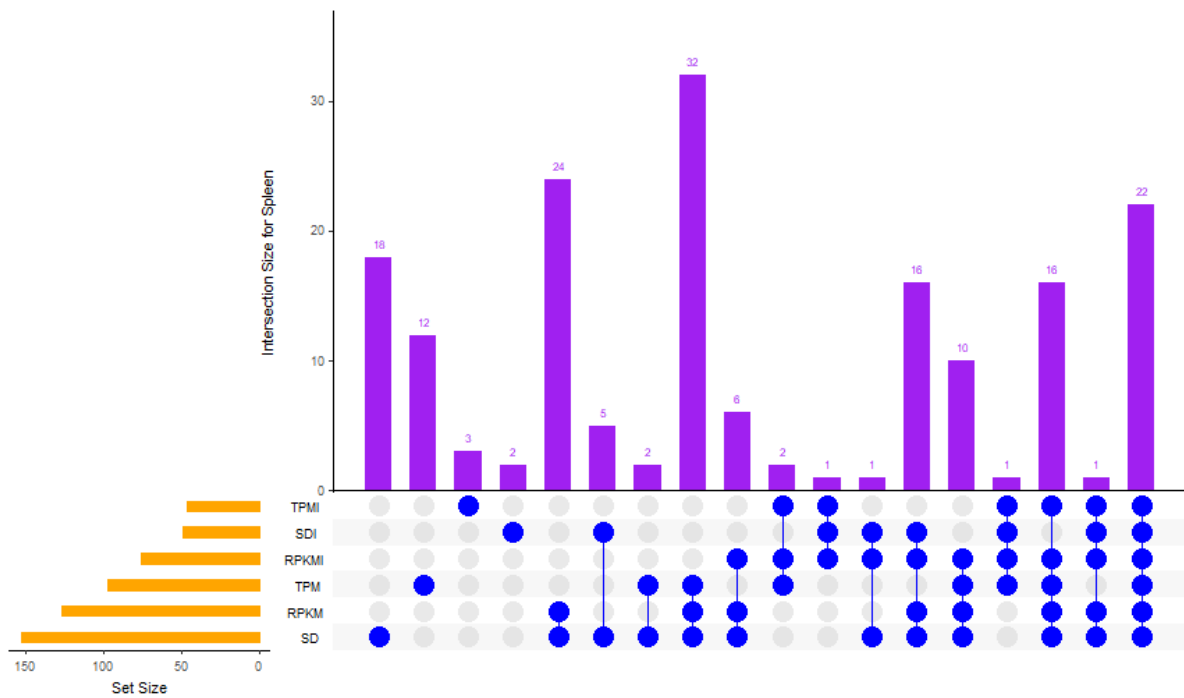
Supplementary Figure 11: Log transformed comparisons for log₂fc dependent intestine specific DEGs of AChE dataset (for TS, Tau>0.8). Letter L represents log transformation on normalized counts.



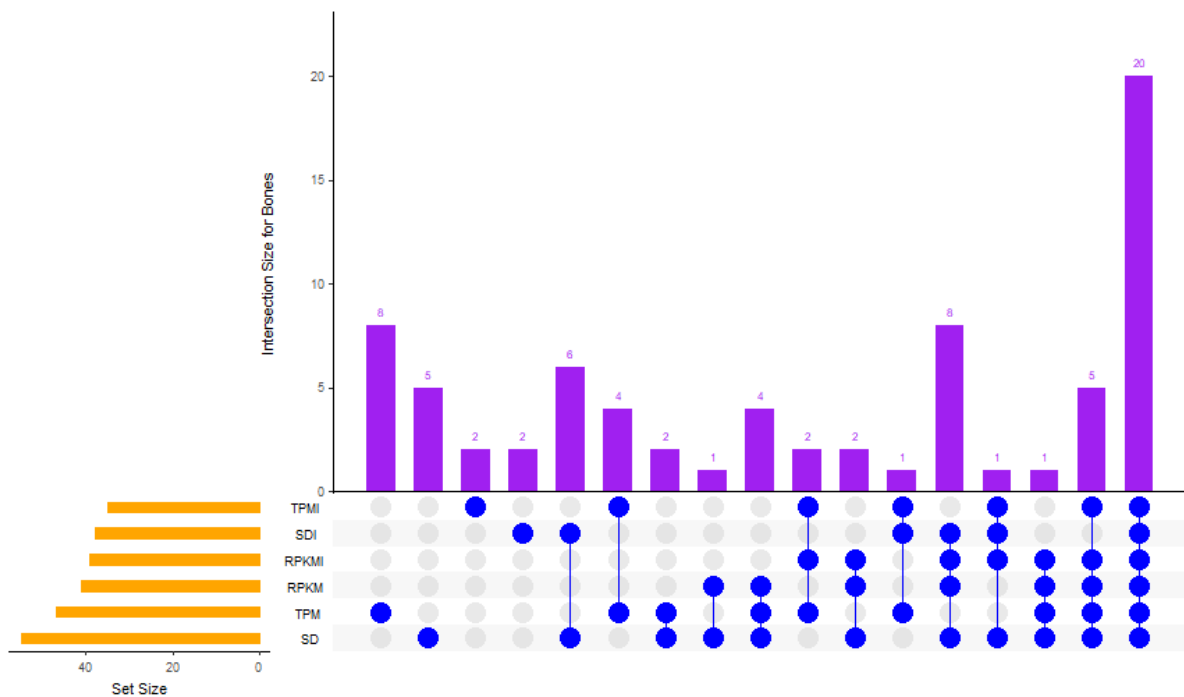
Supplementary Figure 12: Log transformed comparisons for log2fc dependent kidney specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



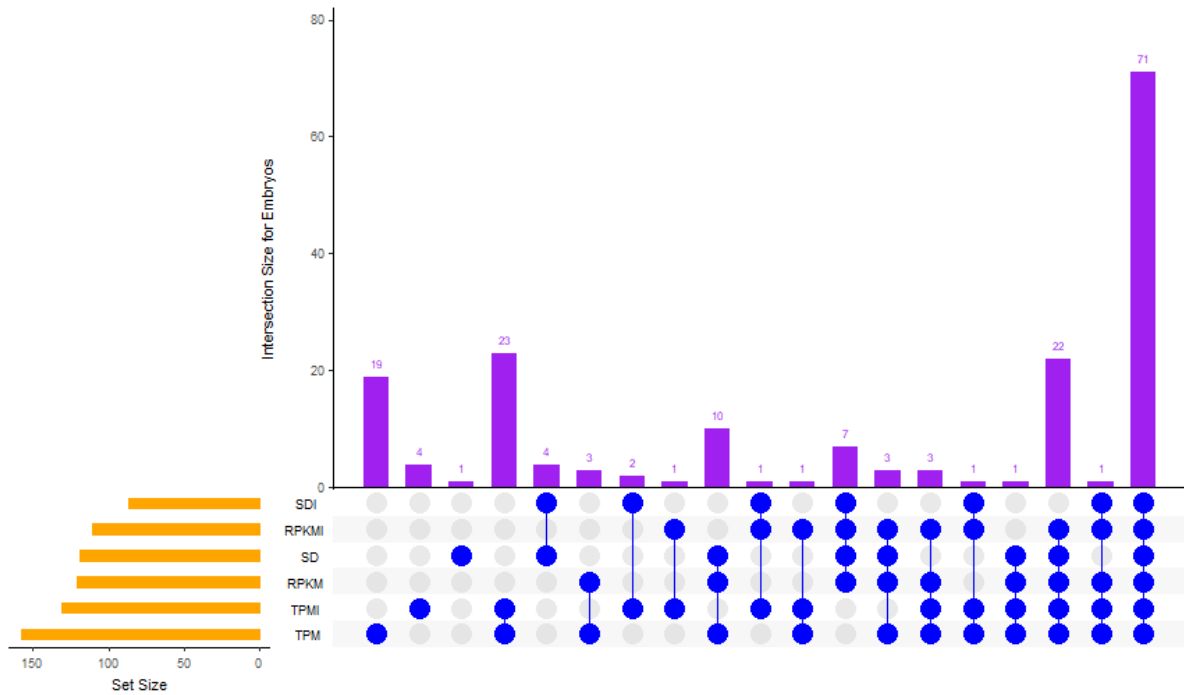
Supplementary Figure 13: Log transformed comparisons for log2fc dependent muscle specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



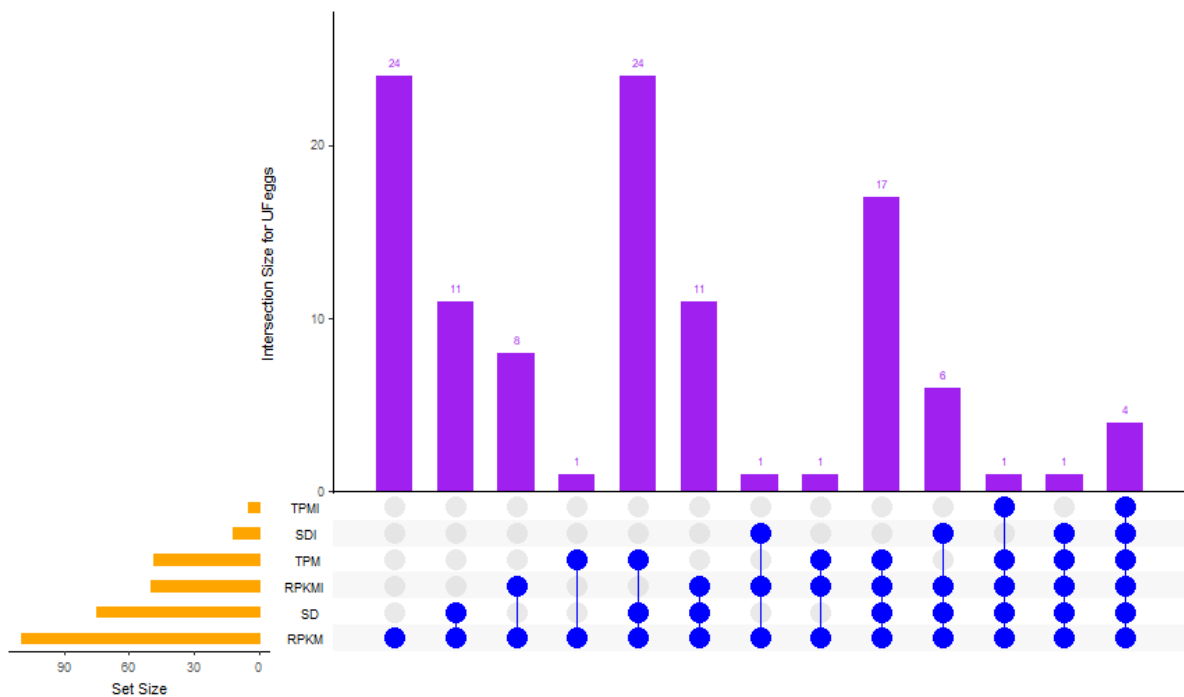
Supplementary Figure 14: Log transformed comparisons for log₂fc dependent spleen specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



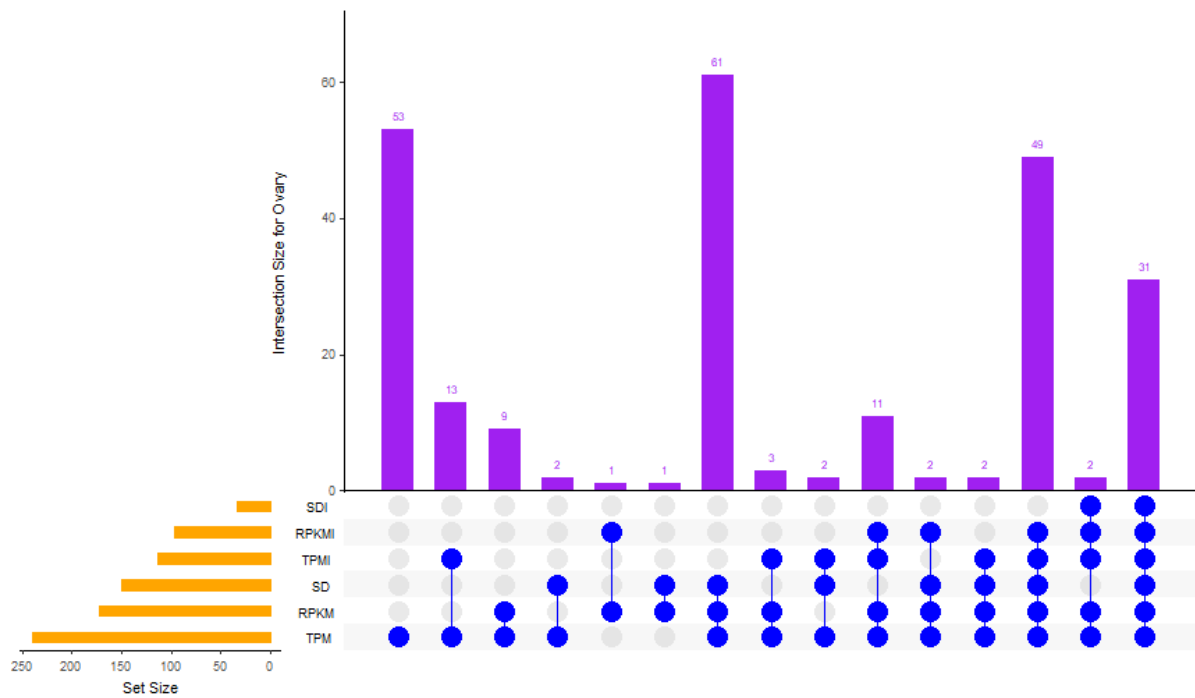
Supplementary Figure 15: Log transformed comparisons for log₂fc dependent bones specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



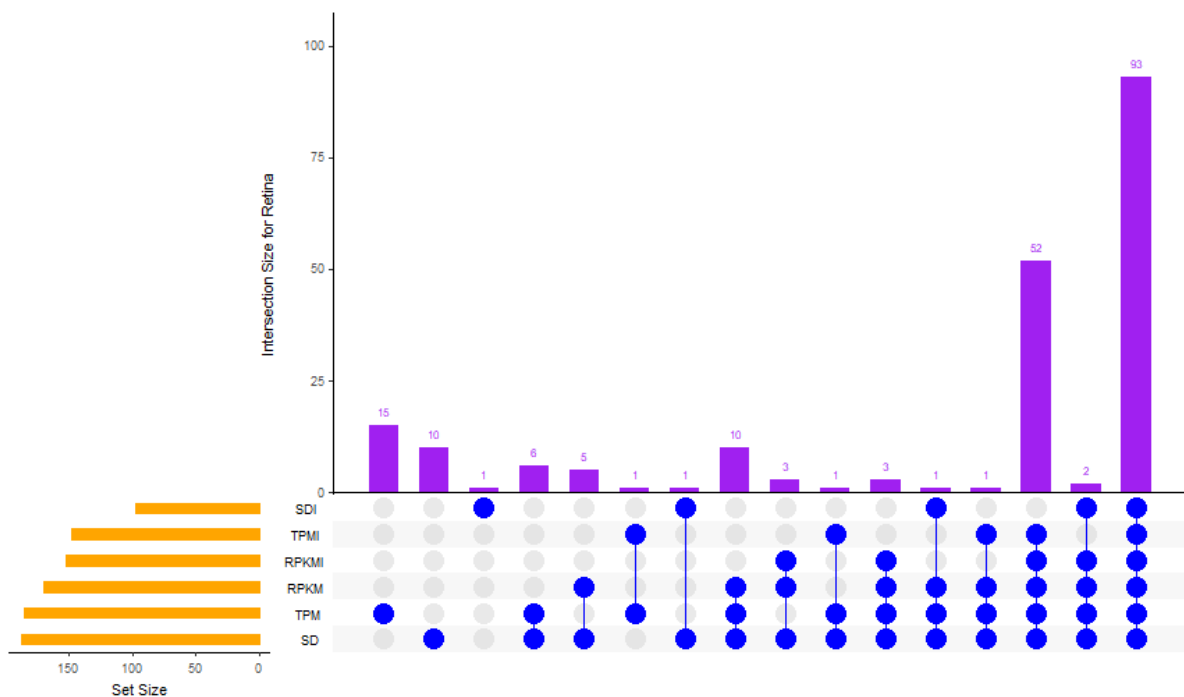
Supplementary Figure 16: Log transformed comparisons for log₂fc dependent embryos specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



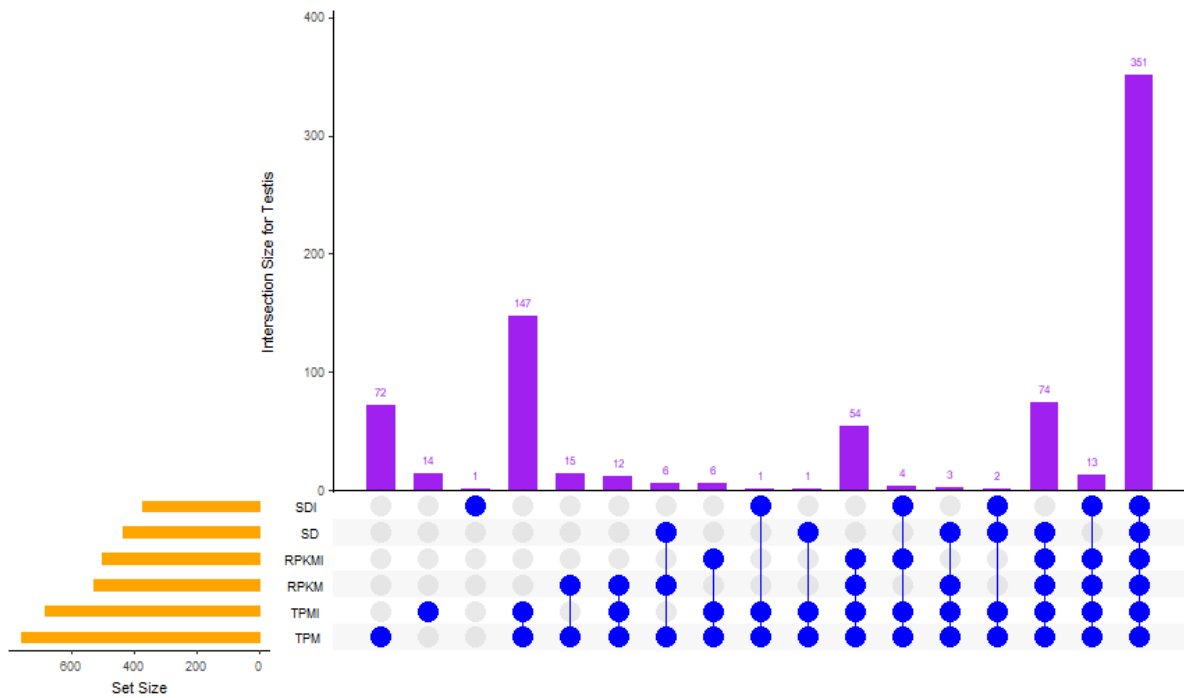
Supplementary Figure 17: Log transformed comparisons for log₂fc dependent UFeggs specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.



Supplementary Figure 18: Log transformed comparisons for log2fc dependent ovary specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.

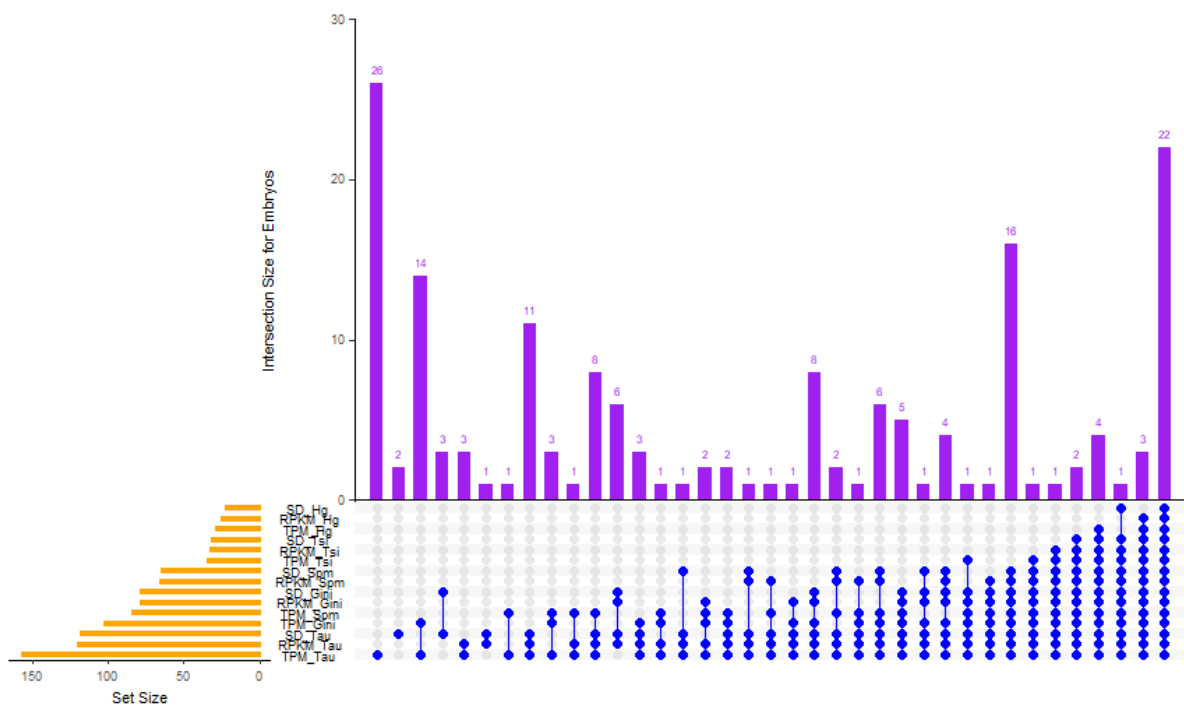


Supplementary Figure 19: Log transformed comparisons for log2fc dependent retina specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.

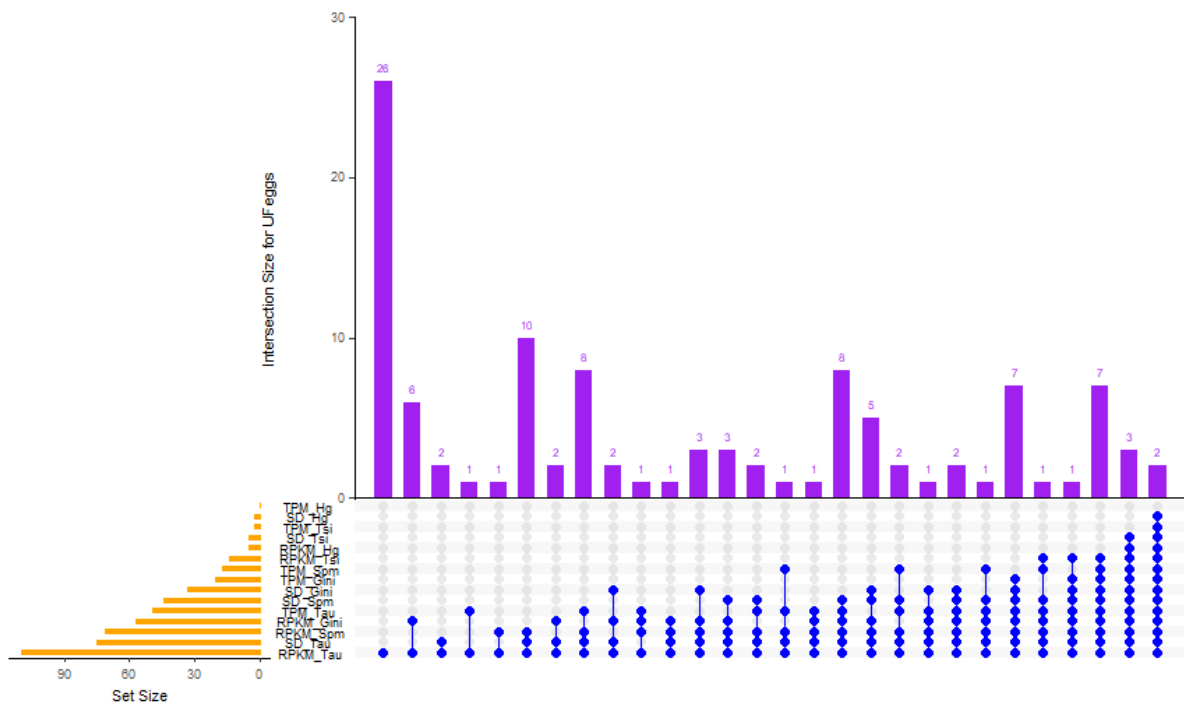


Supplementary Figure 20: Log transformed comparisons for log2fc dependent testis specific DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.

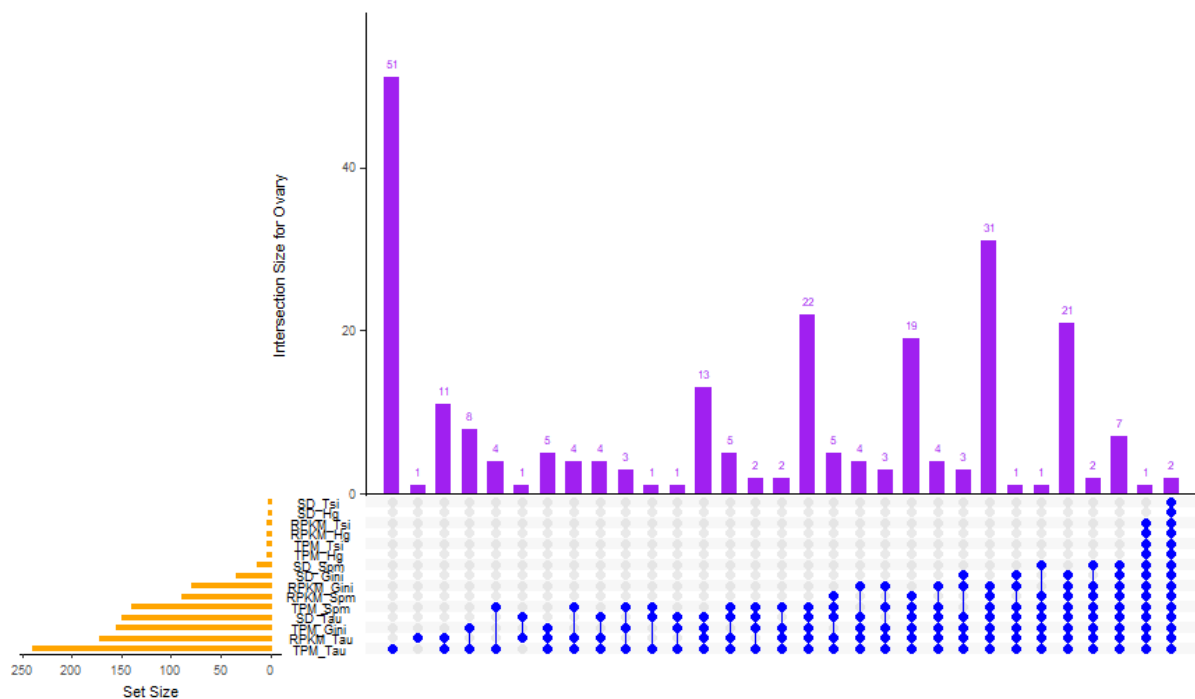
Upset Intersection Graphs for Different Metrics and Normalizations Combined (log2fc based DEG)



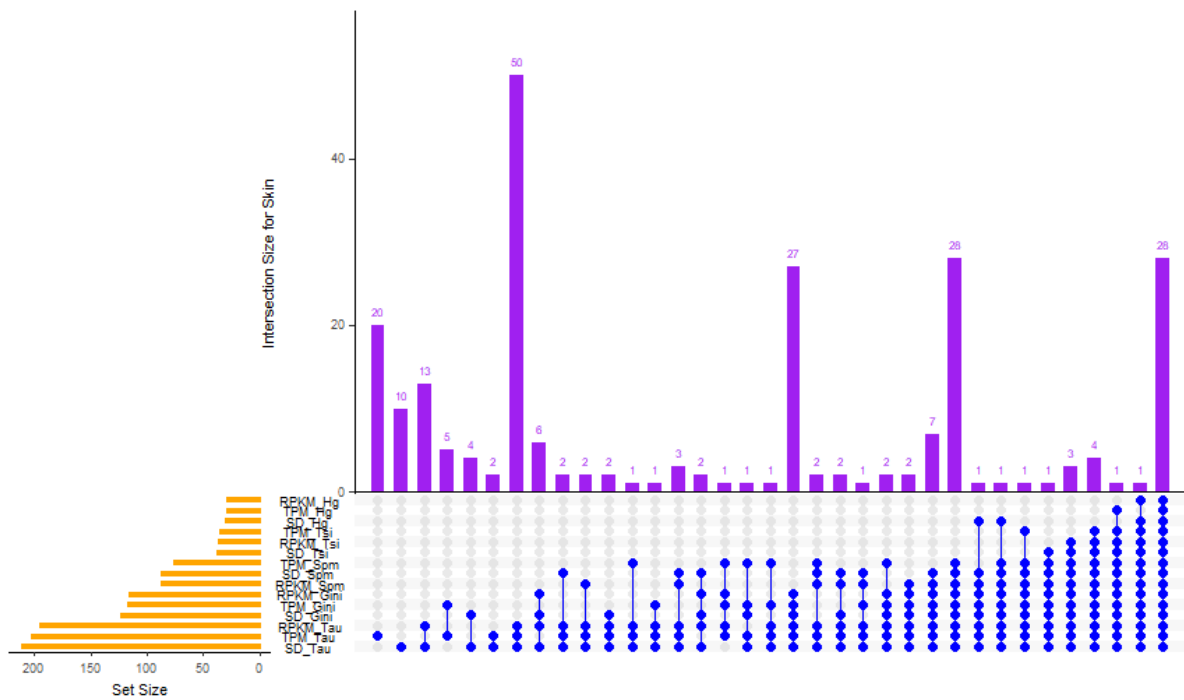
Supplementary Figure 21: Comparisons with metrics and methods on abs (log2FC)>0.5849 and Tau, Tsi, Gini, Spm, Hg>0.8 for embryo specific DEGs.



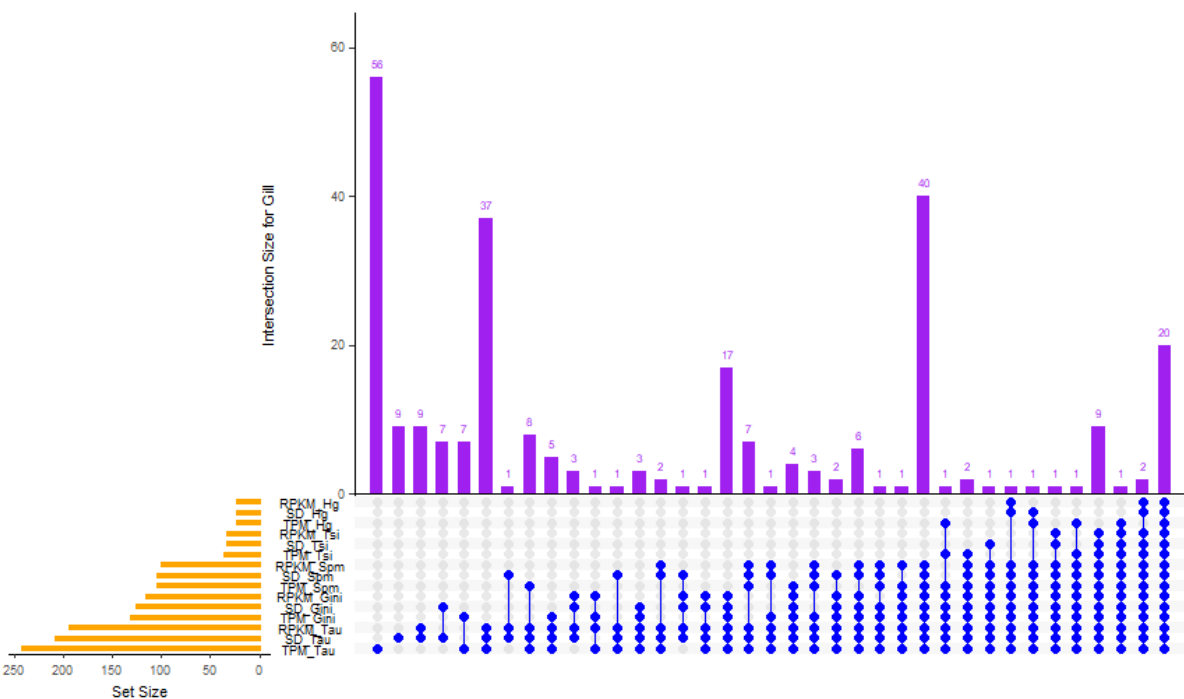
Supplementary Figure 22: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for Uterine Follicle (UF) specific DEGs.



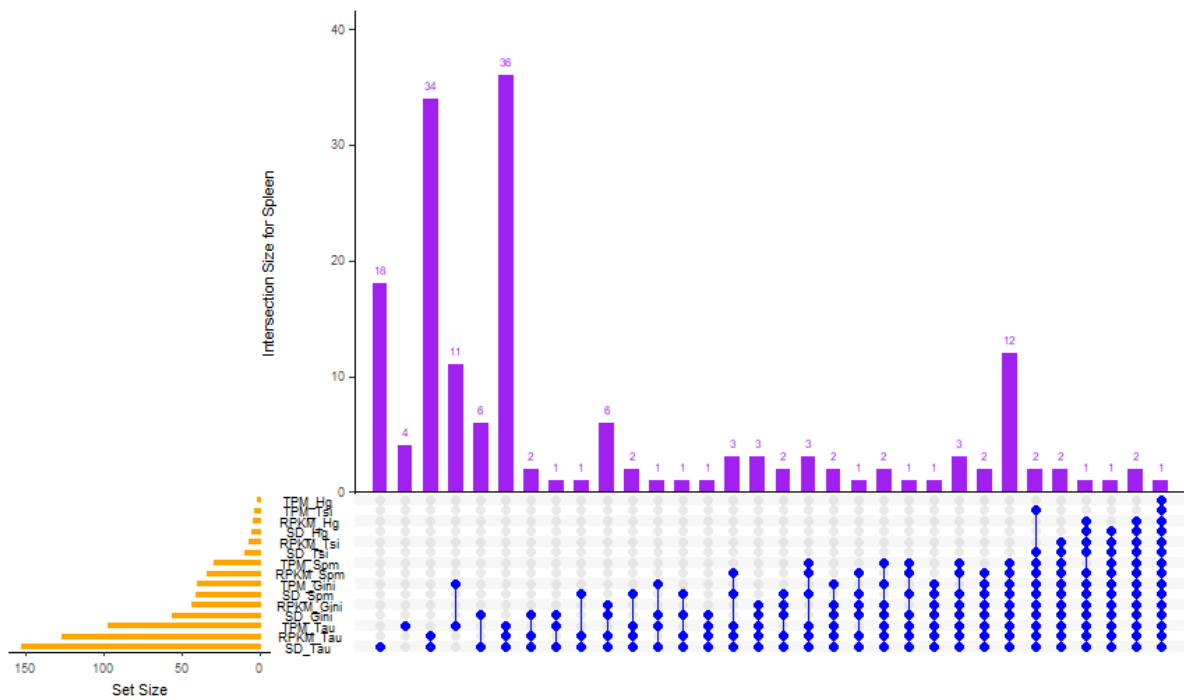
Supplementary Figure 23: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for ovary specific DEGs.



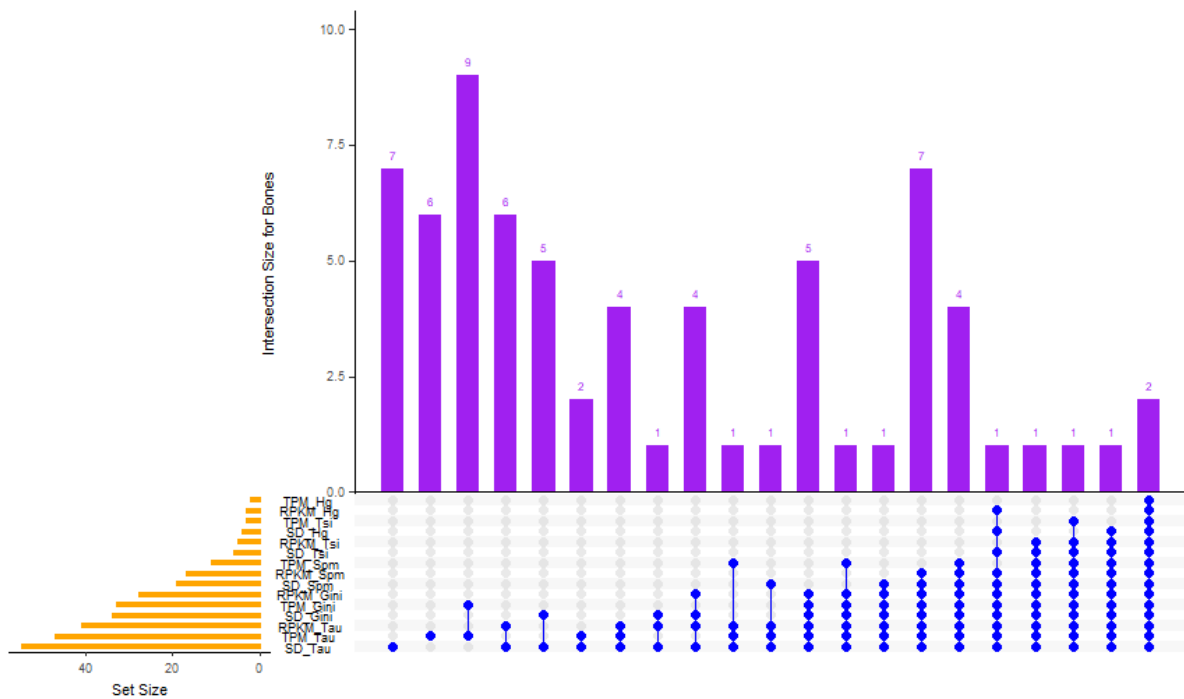
Supplementary Figure 26: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for skin specific DEGs.



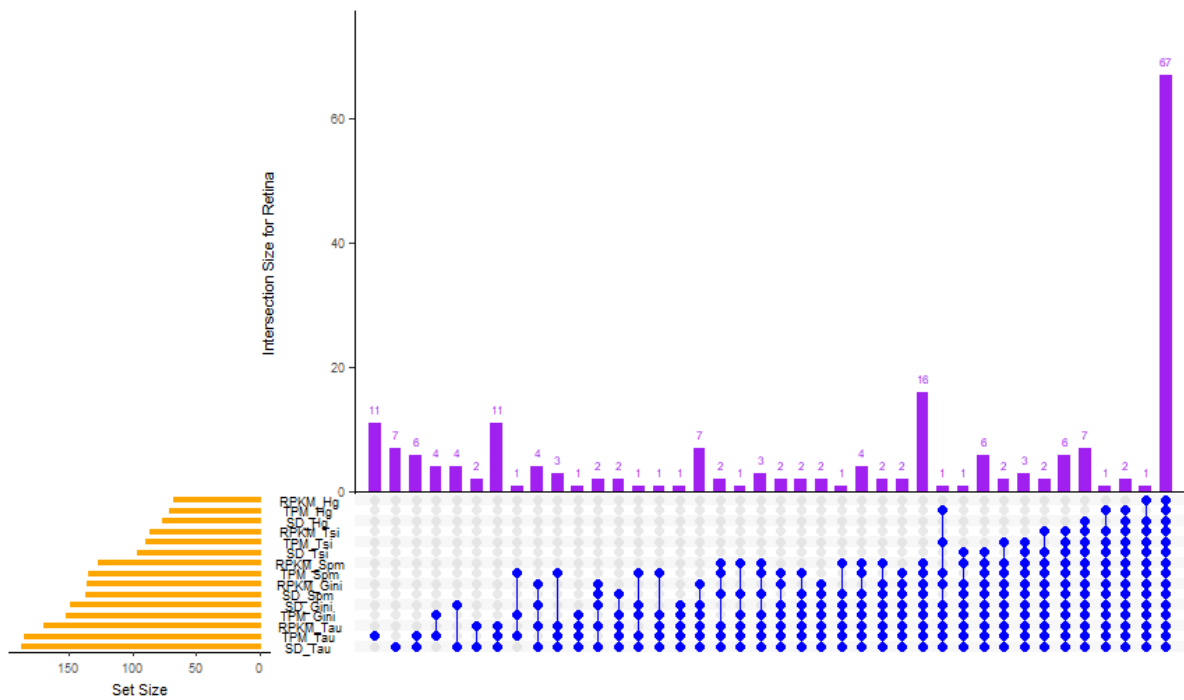
Supplementary Figure 27: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for gill specific DEGs.



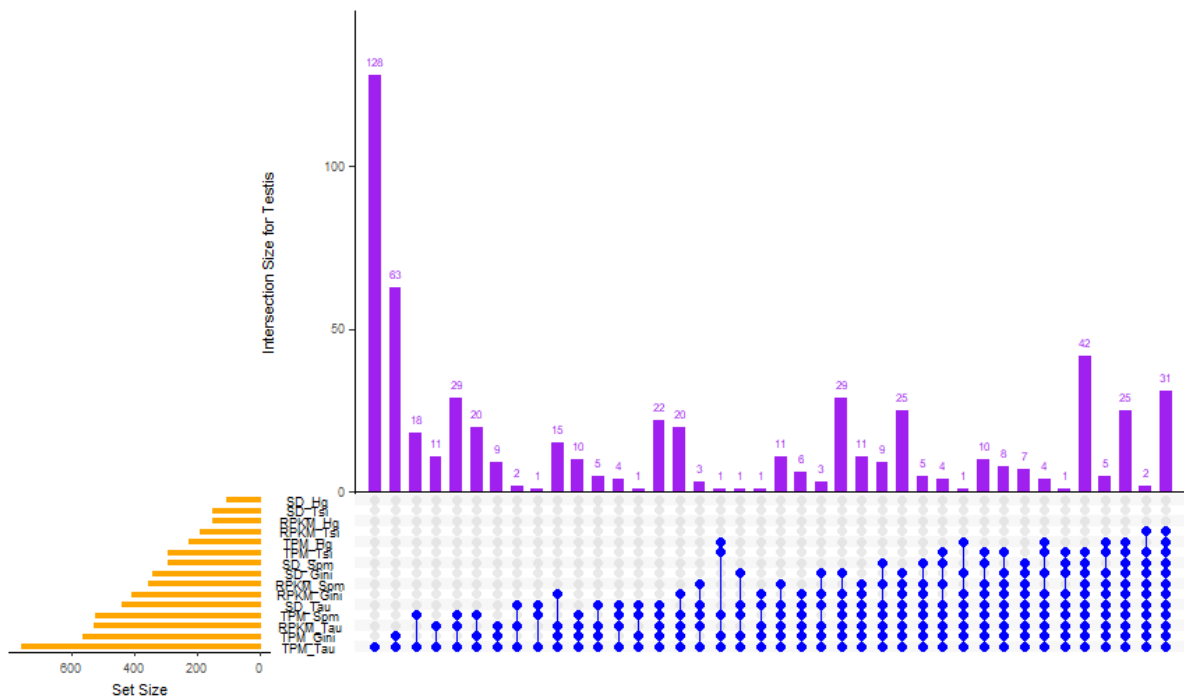
Supplementary Figure 32: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for spleen specific DEGs.



Supplementary Figure 33: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and Tau, Tsi, Gini, Spm, Hg > 0.8 for bones specific DEGs.

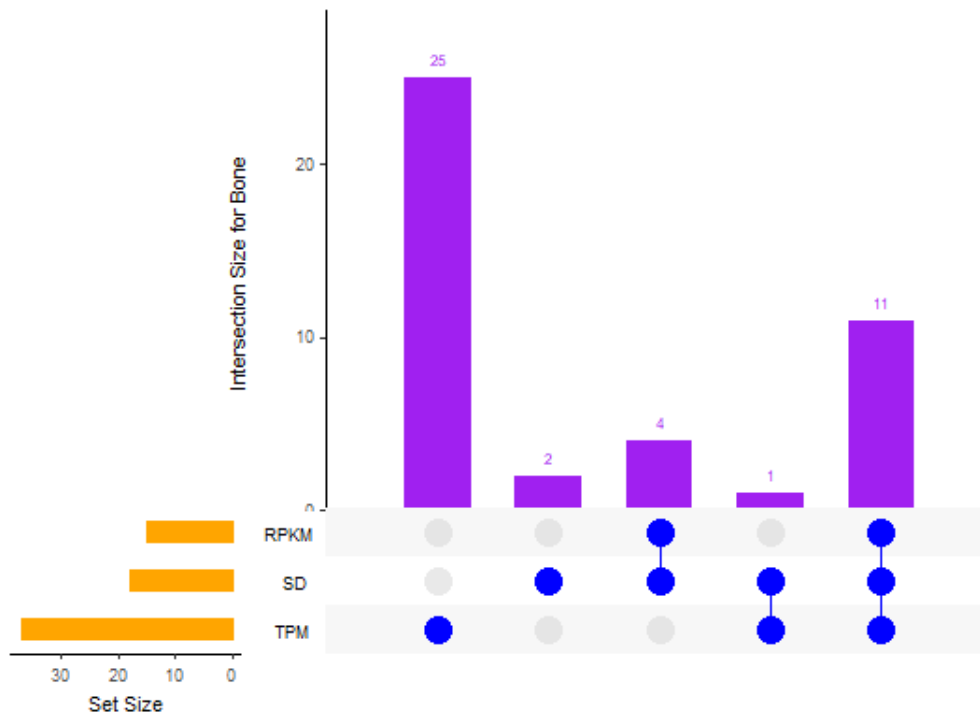


Supplementary Figure 34: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and $\text{Tau}, \text{Tsi}, \text{Gini}, \text{Spm}, \text{Hg} > 0.8$ for retina specific DEGs.

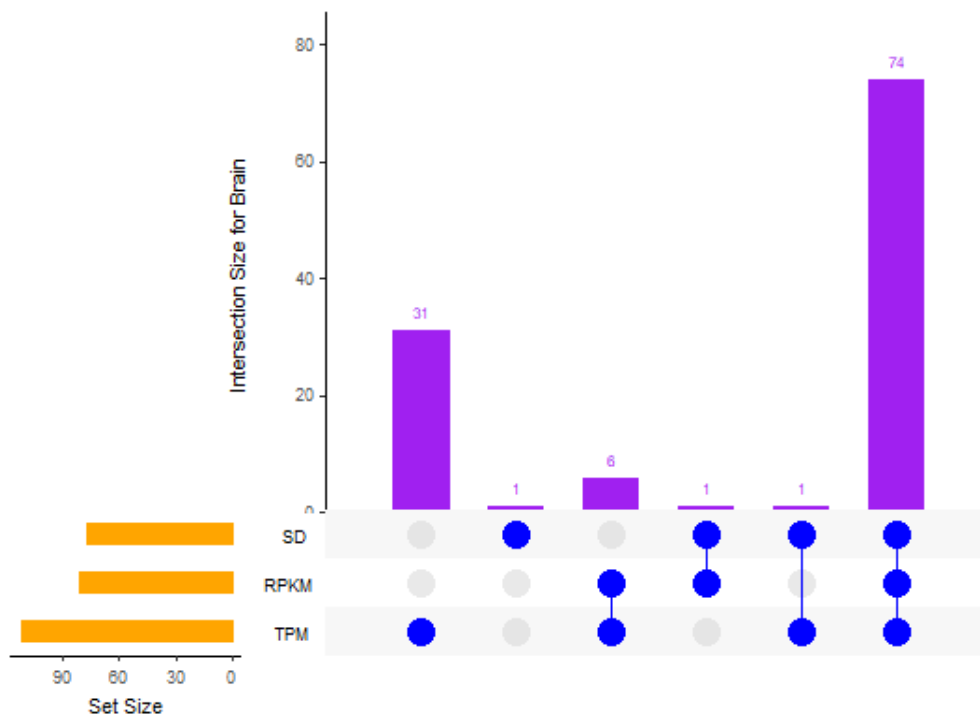


Supplementary Figure 35: Comparisons with metrics and methods on $\text{abs}(\log_2\text{FC}) > 0.5849$ and $\text{Tau}, \text{Tsi}, \text{Gini}, \text{Spm}, \text{Hg} > 0.8$ for testis specific DEGs.

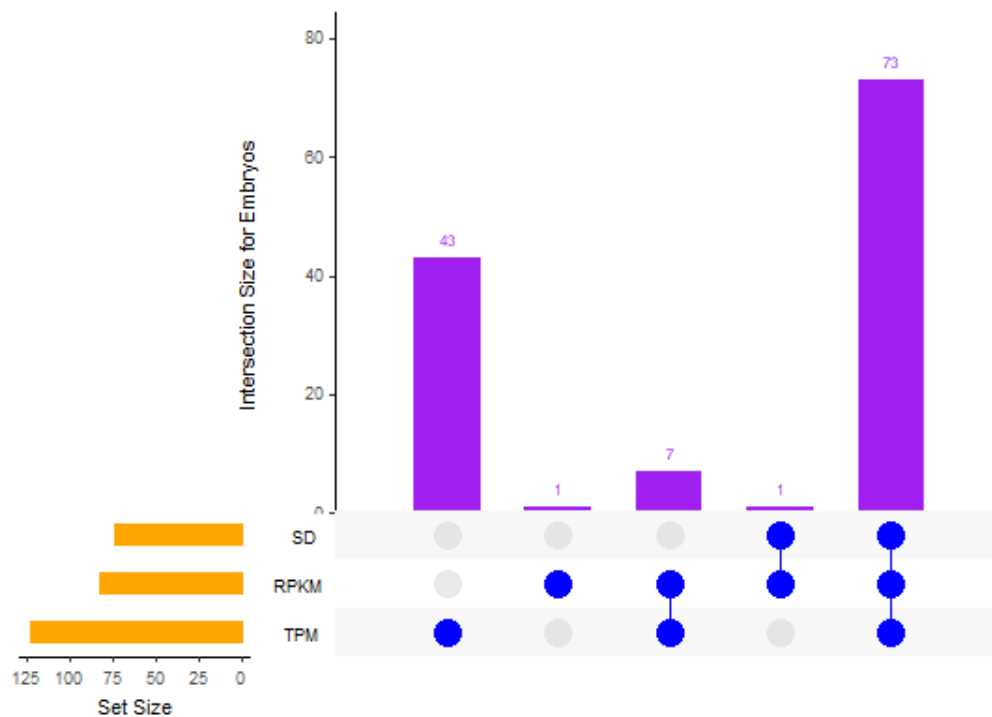
Upset Intersection Graphs to compare TS genes (Tau>0.8) in different normalizations for DEG cut-off p-value<0.1



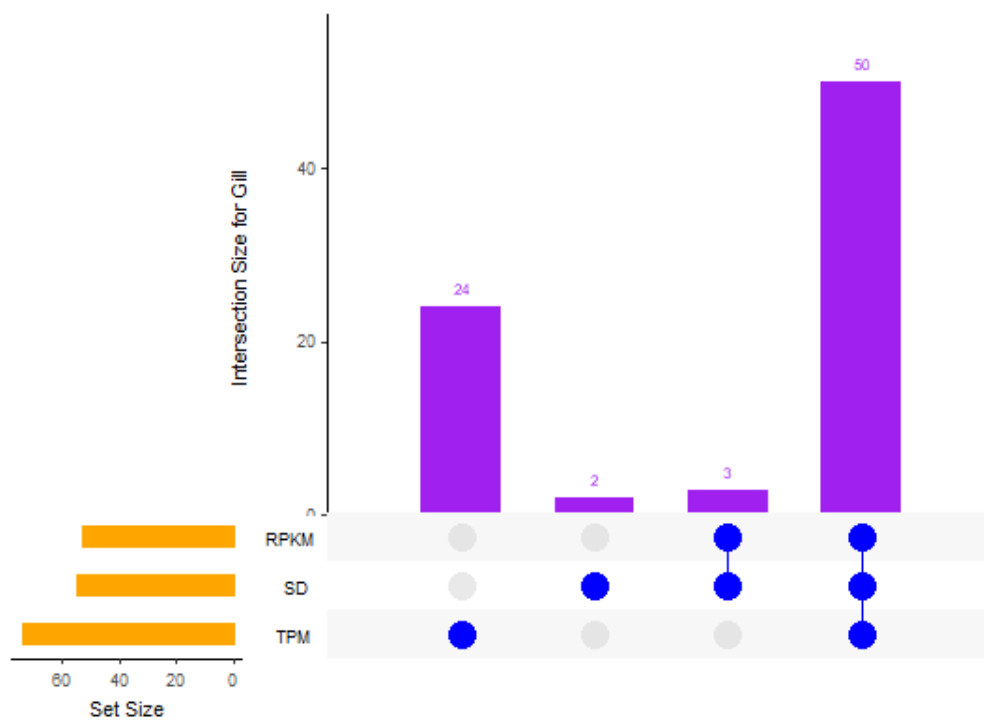
Supplementary Figure 36: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE bone genes.



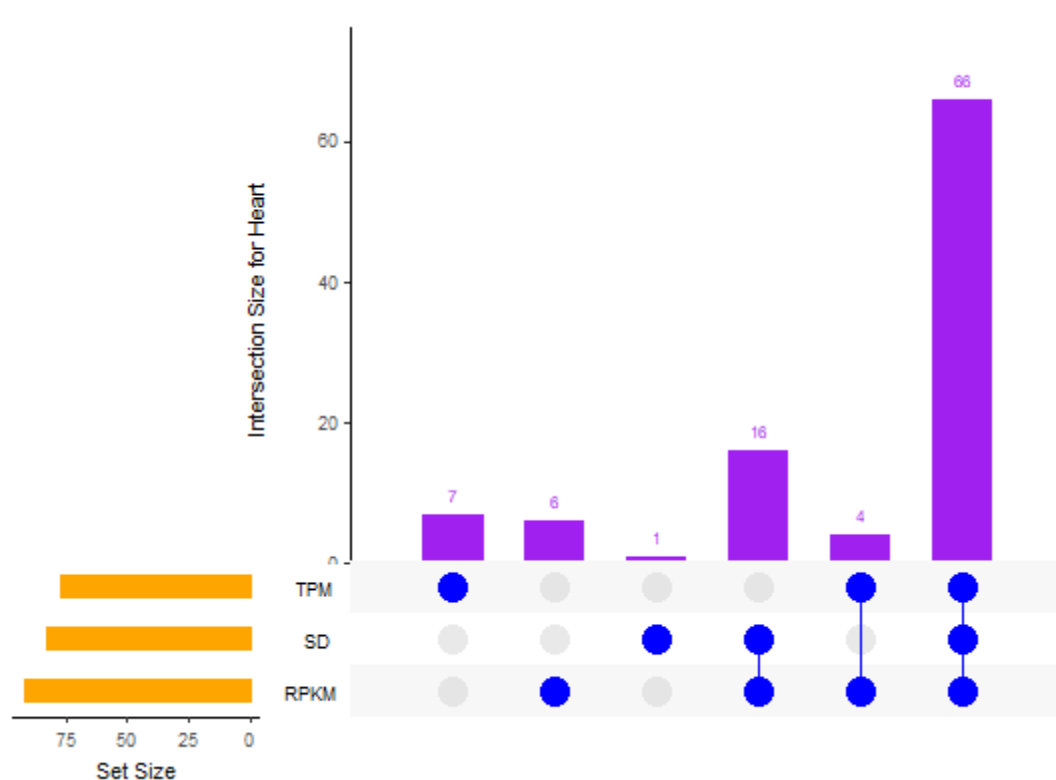
Supplementary Figure 37: Comparisons with metrics and methods on pvalue<0.1 and Tau >0.8 DE brain genes.



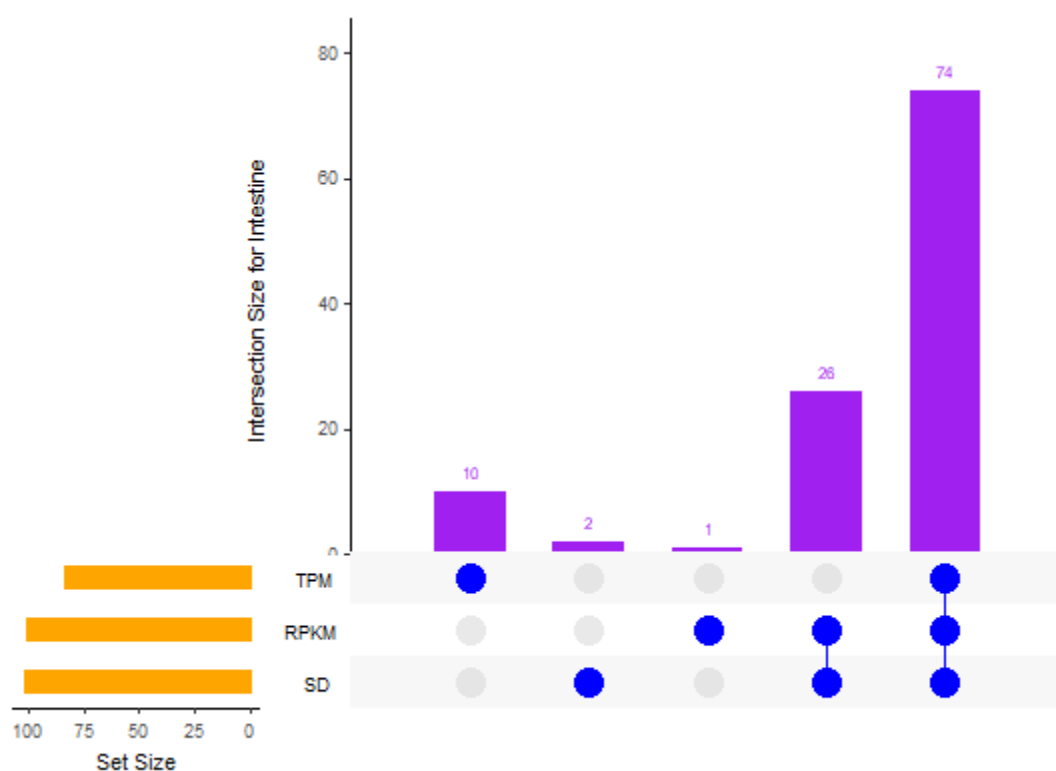
Supplementary Figure 38: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE embryo genes.



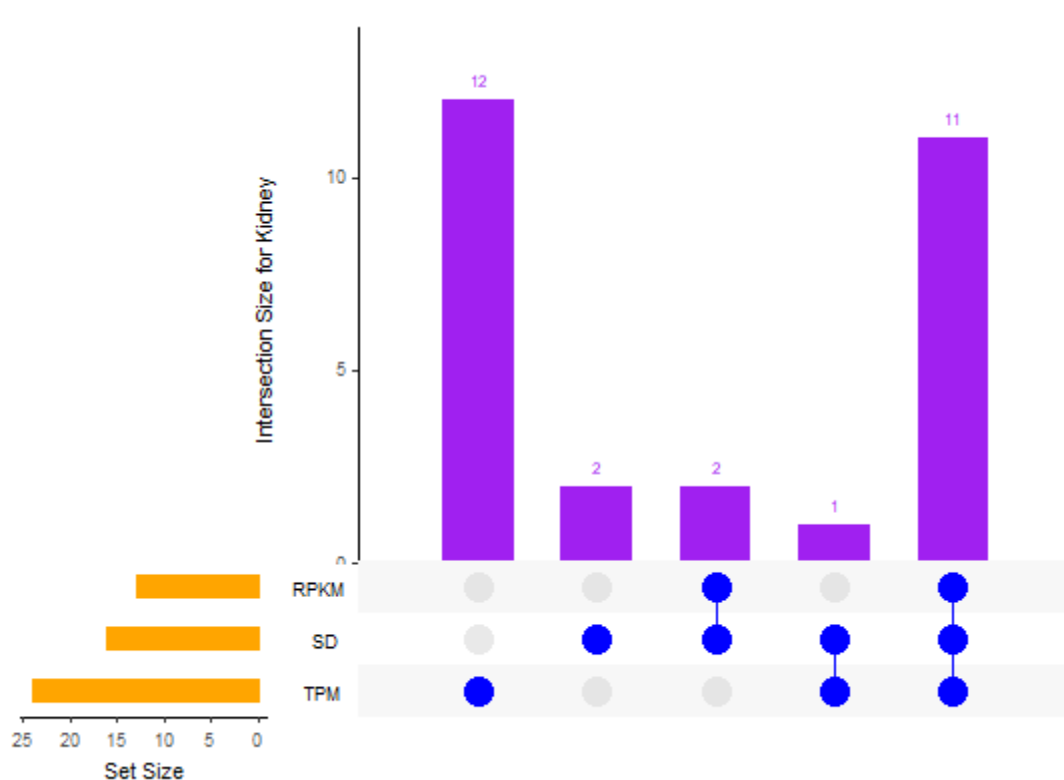
Supplementary Figure 39: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE gill genes.



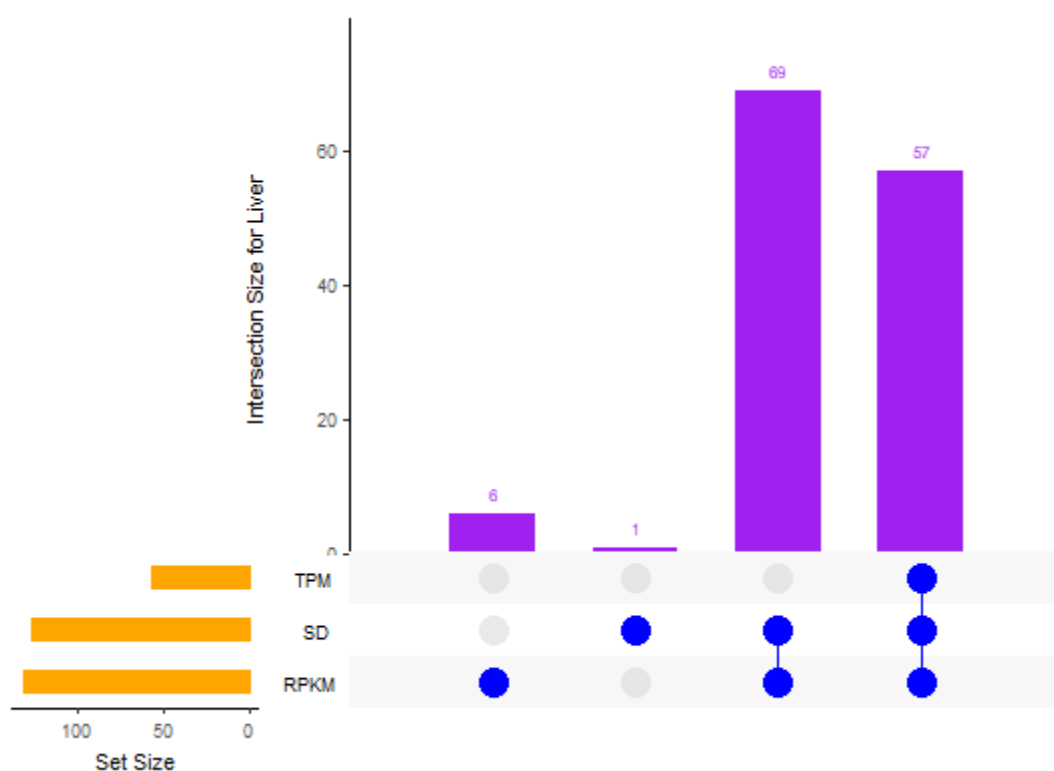
Supplementary Figure 40: Comparisons with metrics and methods on $pvalue < 0.1$ and $\text{Tau} > 0.8$ DE heart genes.



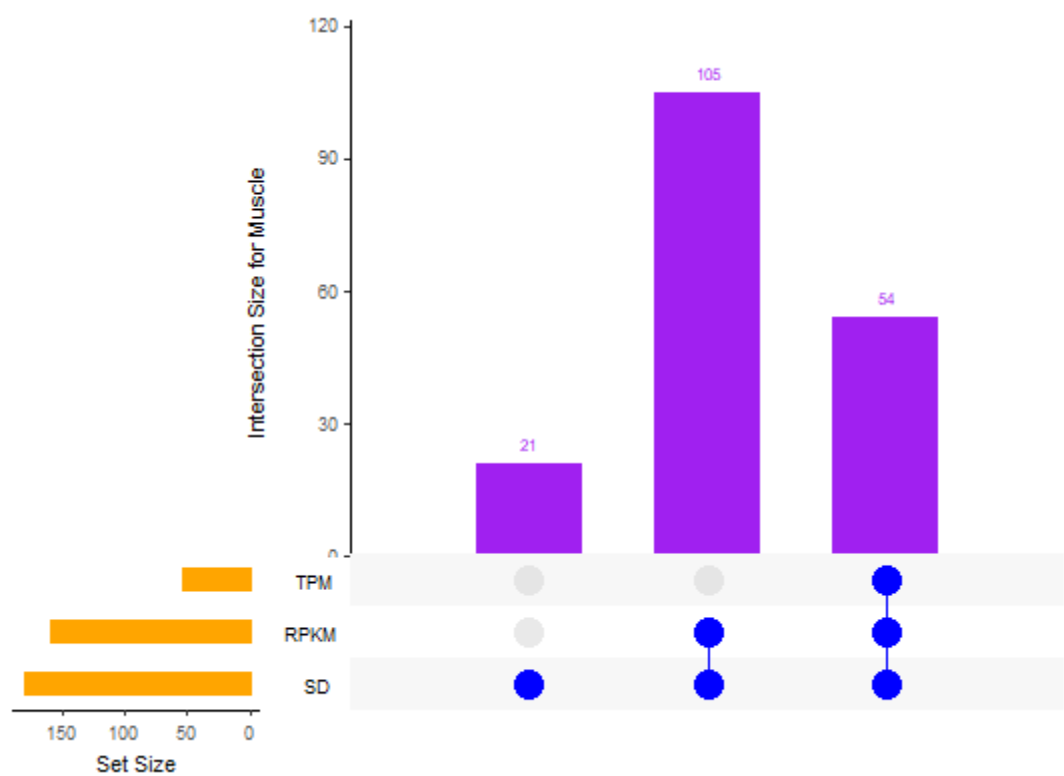
Supplementary Figure 41: Comparisons with metrics and methods on $pvalue < 0.1$ and $\text{Tau} > 0.8$ DE intestine genes.



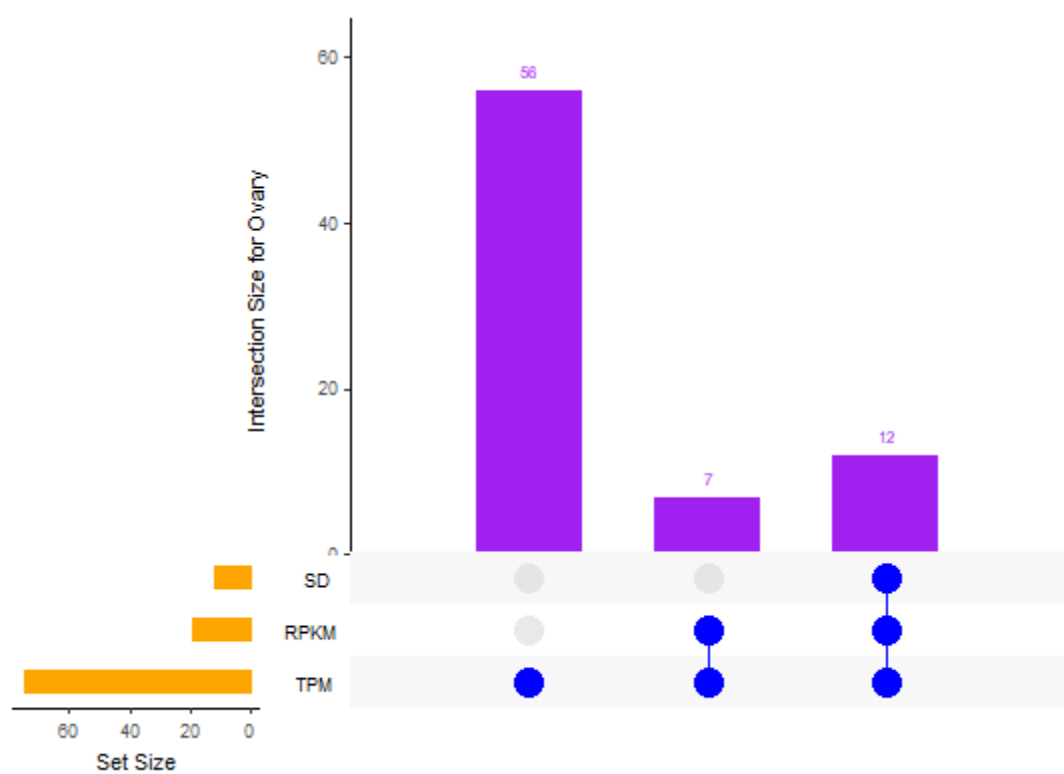
Supplementary Figure 42: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE kidney genes.



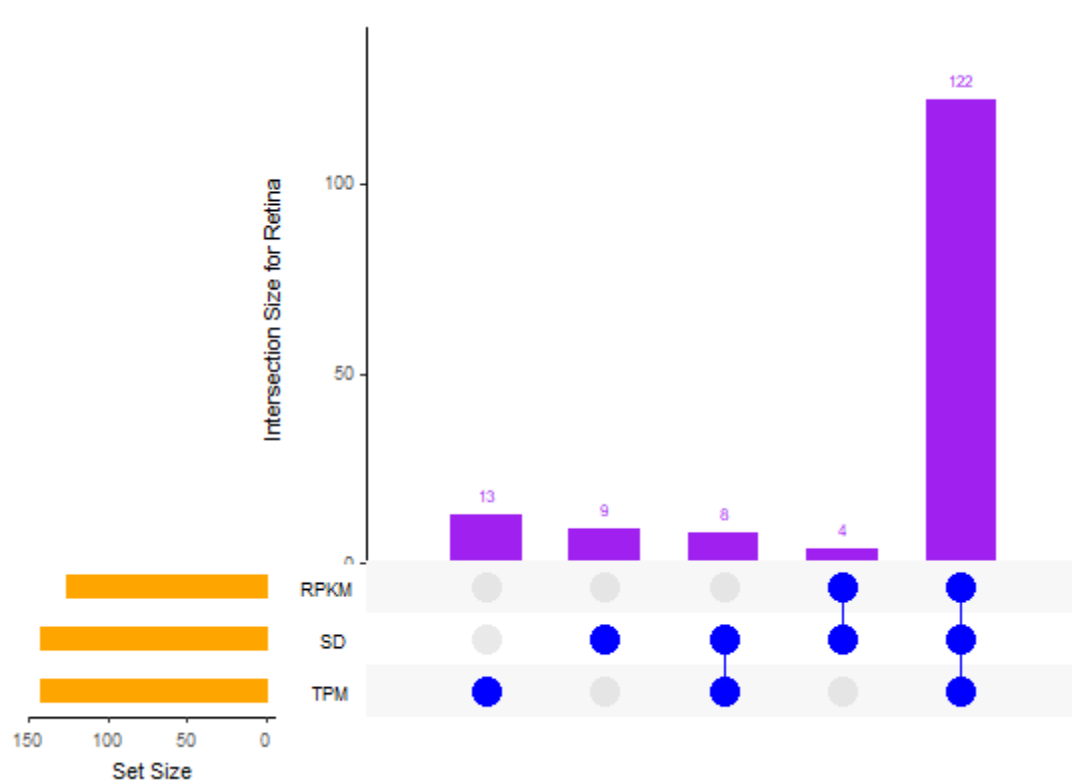
Supplementary Figure 43: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE liver genes.



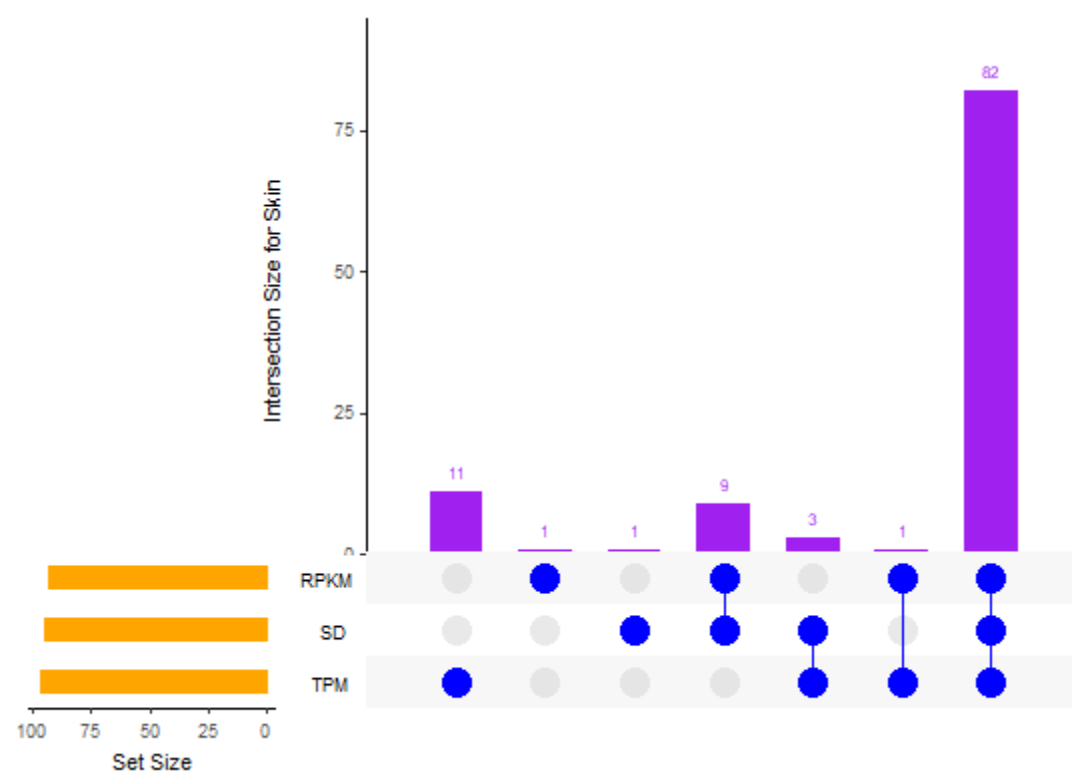
Supplementary Figure 44: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE muscle genes.



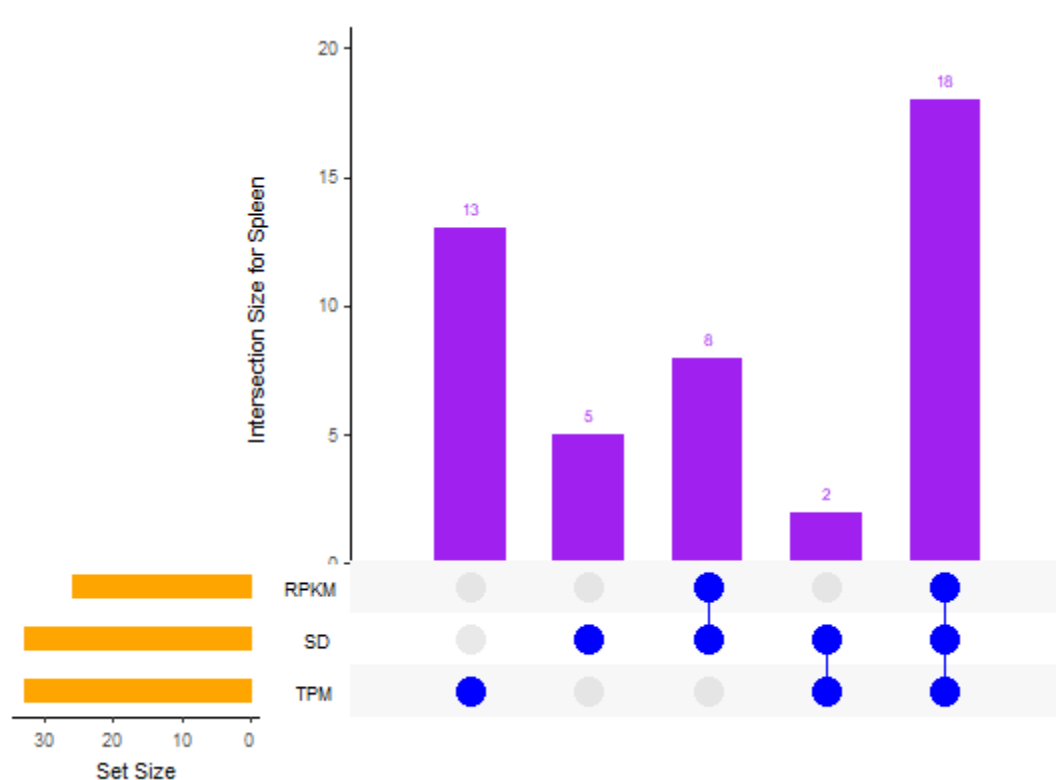
Supplementary Figure 45: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE ovary genes.



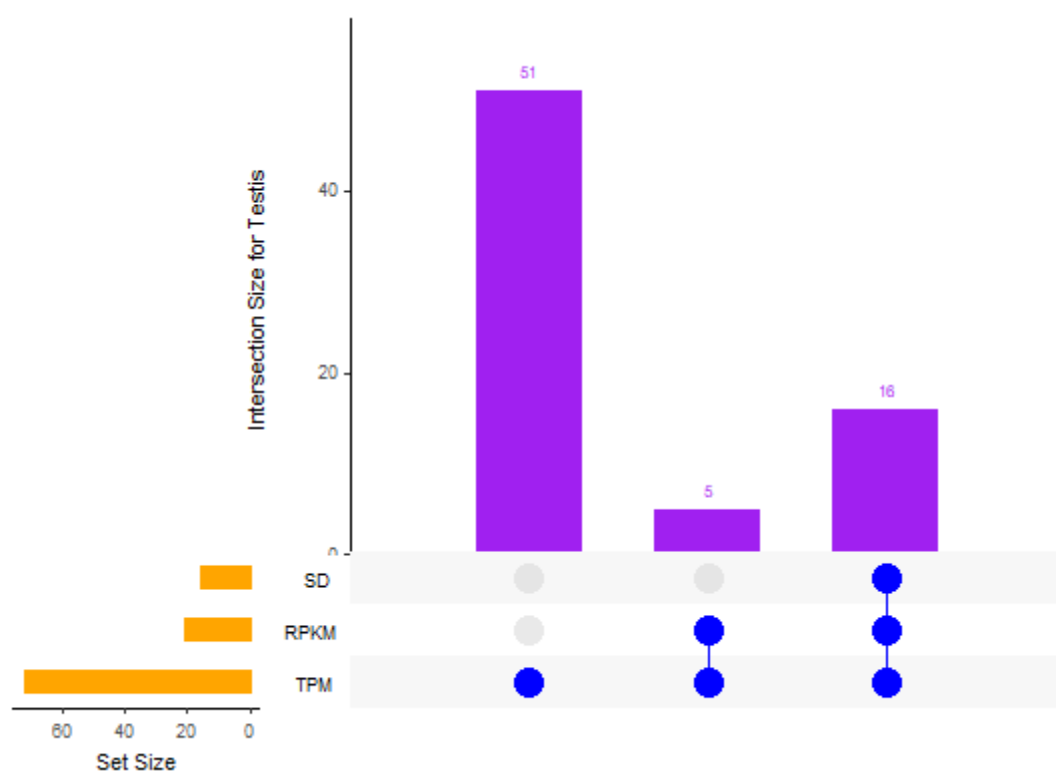
Supplementary Figure 46: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE retina genes.



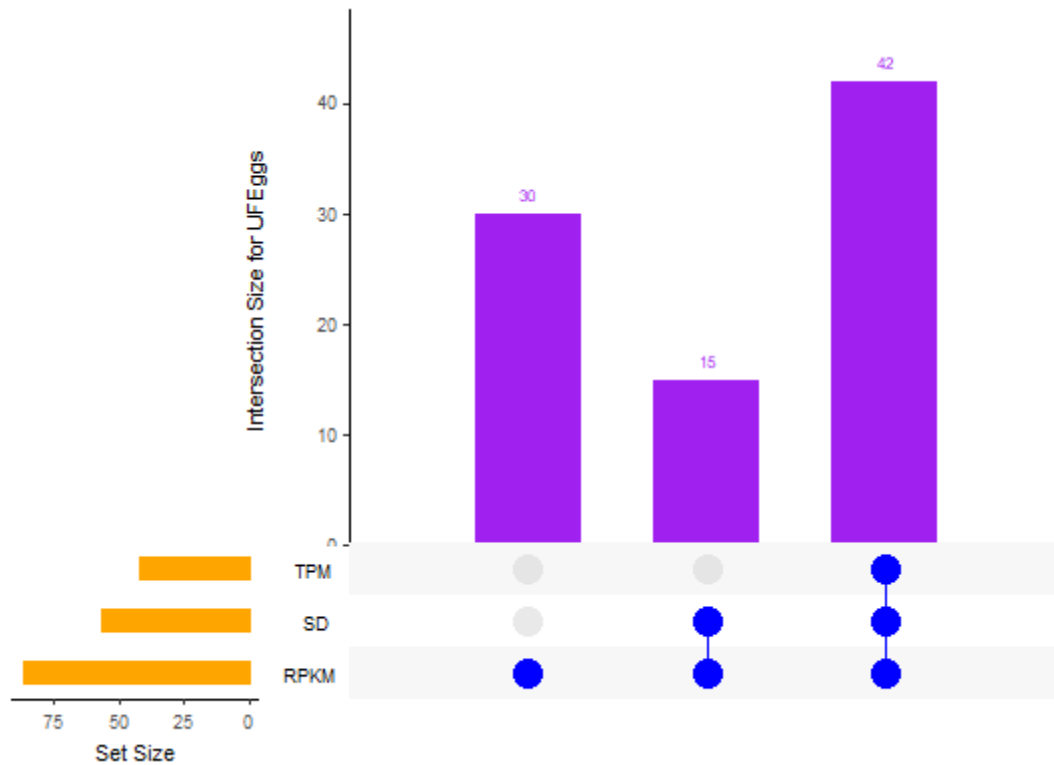
Supplementary Figure 47: Comparisons with metrics and methods on $p\text{value} < 0.1$ and $\text{Tau} > 0.8$ DE skin genes.



Supplementary Figure 48: Comparisons with metrics and methods on $pvalue < 0.1$ and $\text{Tau} > 0.8$ DE spleen genes.

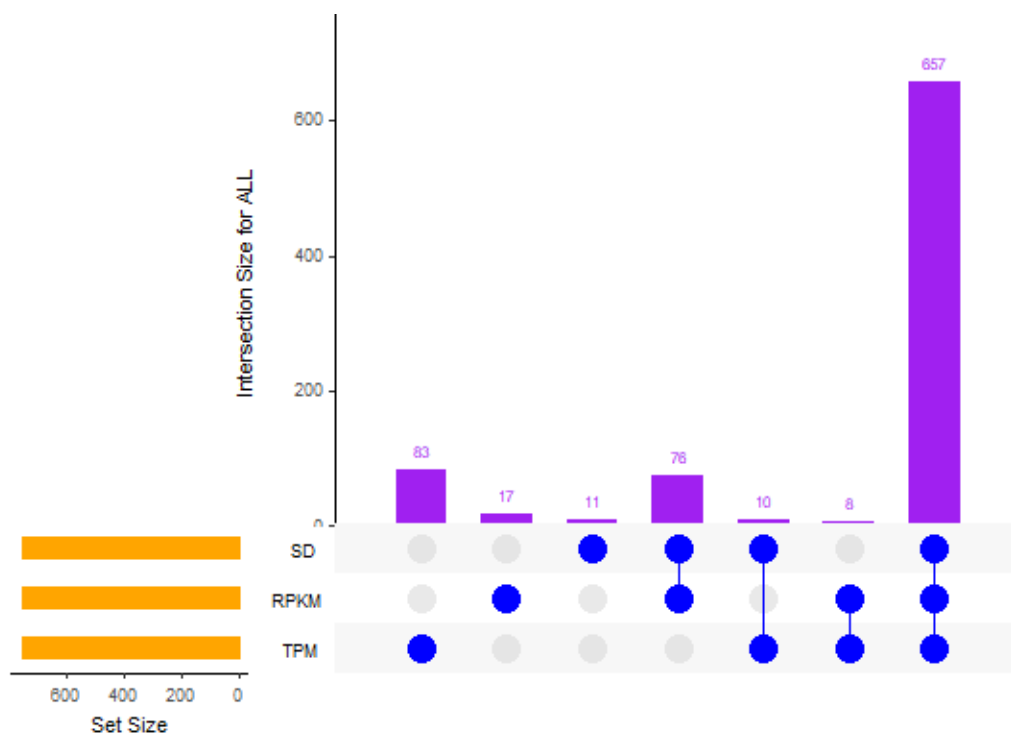


Supplementary Figure 49: Comparisons with metrics and methods on $pvalue < 0.1$ and $\text{Tau} > 0.8$ DE testis genes.



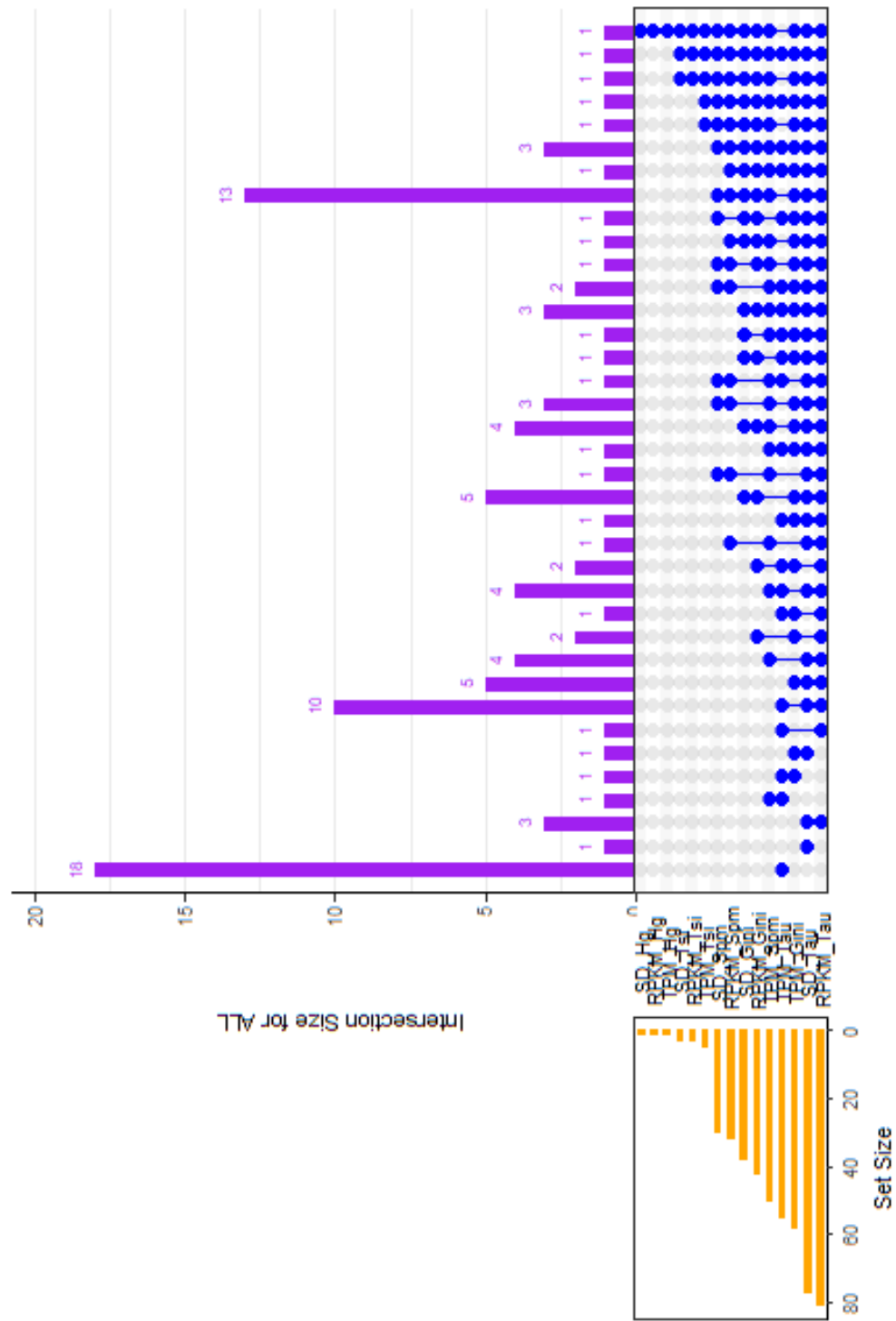
Supplementary Figure 50: Comparisons with metrics and methods on $p\text{-value} < 0.1$ and $\text{Tau} > 0.8$ DE UFEggs genes.

Upset Intersection Graphs to compare TS genes ($\text{Tau} > 0.8$) in different normalizations for DEG cut-off $p\text{-value} < 0.05$



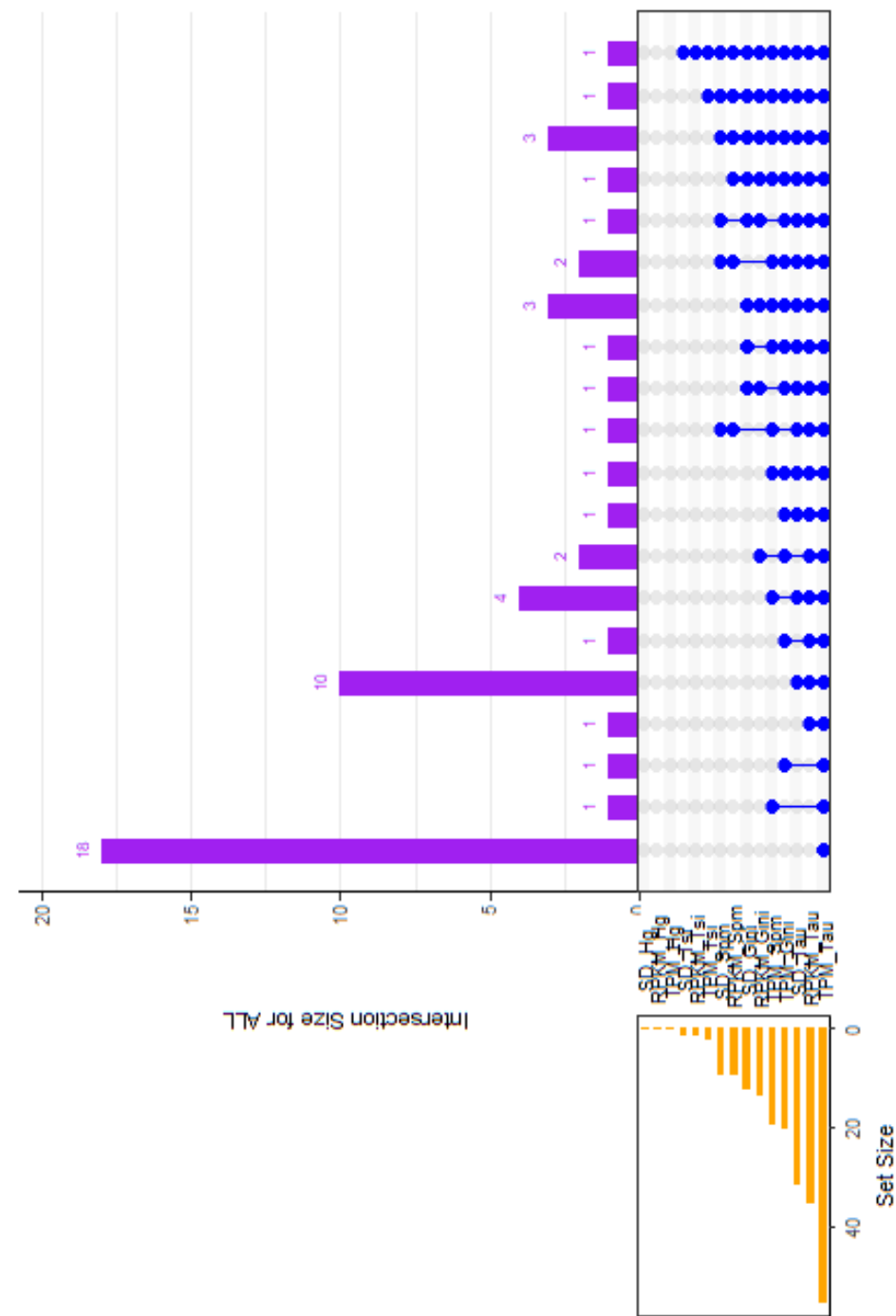
Supplementary Figure 51: Comparisons with metrics and methods on $p\text{-value} < 0.05$ and $\text{Tau} > 0.8$ for all TS-DEGs.

Upset Intersection Graphs for Different Metrics and Normalizations Combined (p-value<0.1 based DEG)



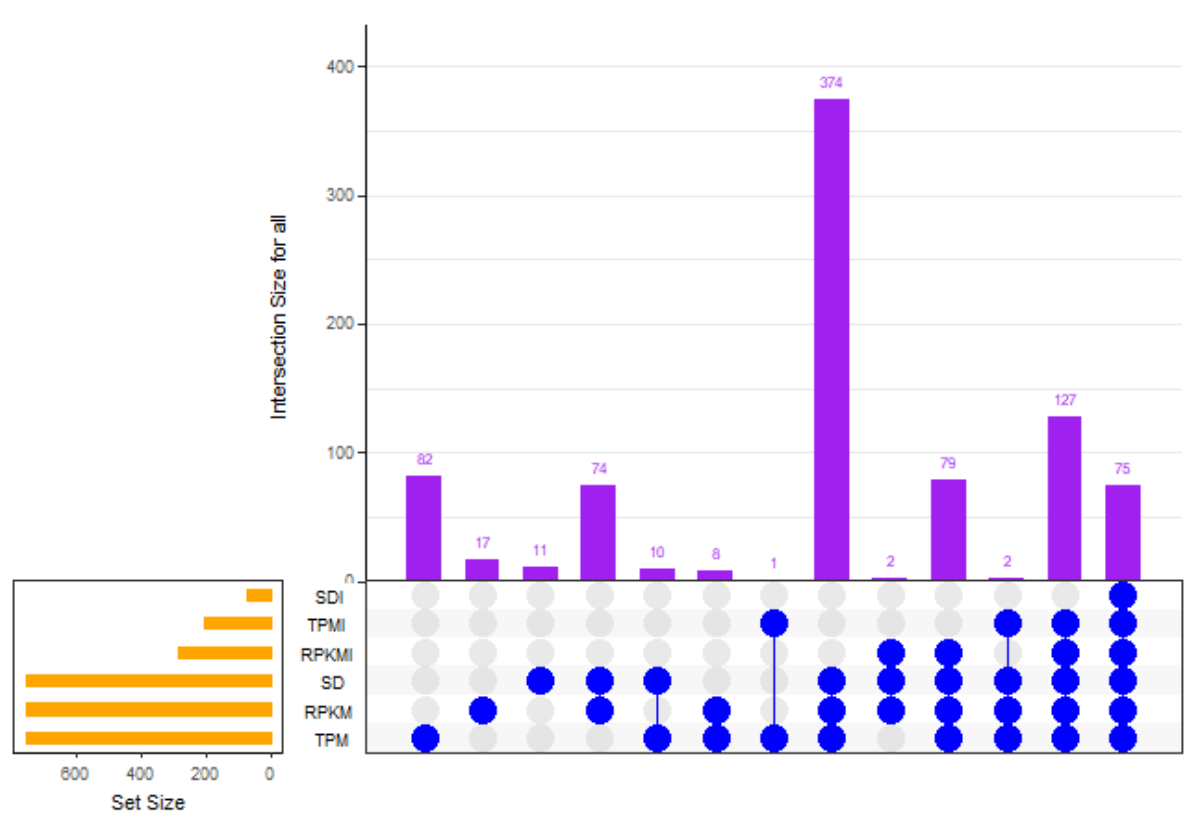
Supplementary Figure 52: Comparisons with metrics and methods on pvalue ≤ 0.1 and Tau, Tsi, Gini, Spm, Hg ≥ 0.8 . Intersection sizes for all the tissues combined.

Upset Intersection Graphs for Different Metrics and Normalizations Combined (p-value<0.05 based DEG)



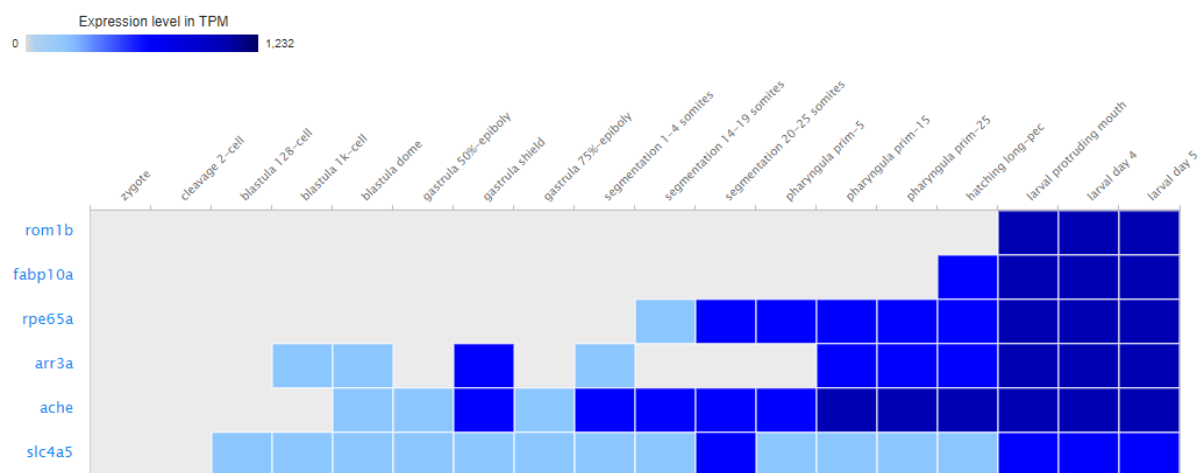
Supplementary Figure 53: Comparisons with metrics and methods on pvalue<0.05 and Tau, Tsi, Gini, Spm, Hg>0.8. Intersection size for all TS-DEGs.

Upset Intersection Graphs for log2 transformed vs. non-transformed normalized counts on p-value<0.05 DEGs



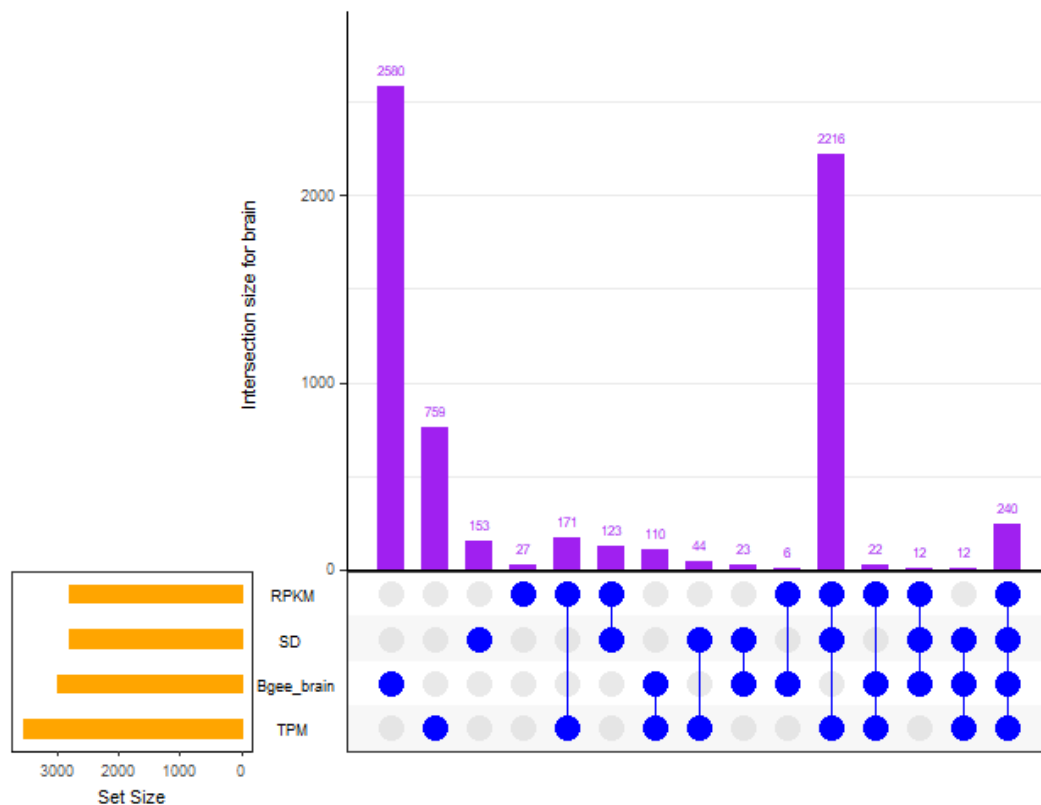
Supplementary Figure 54: Log transformed comparisons for p-value<0.05 based DEGs of AChE dataset (for TS, Tau>0.8). Letter I represents log transformation on normalized counts.

Expression Profiling of the Selected Genes in Expression Atlas



Supplementary Figure 55: TPM normalized expression of selected genes available for different developmental stages.

The difference between Bgee and TS genes in this thesis



Supplementary Figure 56: Intersection sizes for Brain specific genes in either of the normalization methods used in this thesis or Bgee tool for zebrafish (Bgee brain).

Structure of Ache

P801208 VWFC_1 VWFC domain signature :
46 - 84: [confidence level: <1] CcscCagGccccc.....CcgcCagCgacGcgGccggccccgc.....CGCa..C
906 - 965: [confidence level: <1] CcA.CcAgctccccc....CcAcGcttctgttt.....CcAcgcttgca.....CCGcac
C
4734 - 4778: [confidence level: <1] CgcCccCgagcagat.....CcAcCagaaq.....CcCcCagcagc.....CGCa..
P800022 EGF_1 EGF-like domain signature :
538 - 547: [confidence level: <1] CcCagCgcCgaAC
1872 - 1883: [confidence level: <1] CcCgcCagCgpcC
2117 - 2128: [confidence level: <1] CcAcTccCagaDcT
4811 - 4822: [confidence level: <1] CcCtcCtCgDcC
4738 - 4747: [confidence level: <1] CcCcccccGGAC
P800198 4FE4F_FER_1 4Fe-4S ferredoxin-type iron-sulfur binding region signature :
894 - 1006: [confidence level: <1] CcCrcCAcAAcG
3382 - 3403: [confidence level: <1] CcAcCrcCAcTAcG
P800099 THIOALBE_3 Thioredoxin active site :
1284 - 1377: [confidence level: <1] AAAGCAAAAGGAAG
Predicted feature:
ACT_SITE 1369 Proton acceptor [condition: none]
P801771 ANAPHYLATOXIN_1 Anaphylatoxin domain signature :
1424 - 1480: [confidence level: <1] CCAGcAGCGAGcAGcAGcGcGcGcGcGAGcAGcAGc..CC
1818 - 1866: [confidence level: <1] CcCgAGcAGcAGcAGcAGcGCGG.....AAACCGAAC
P800197 2FE2F_FER_1 2Fe-2S ferredoxin-type iron-sulfur binding region signature :
2039 - 2047: [confidence level: <1] CCcCAcCAcC
2262 - 2280: [confidence level: <1] CcAGAcAGcG
3802 - 3810: [confidence level: <1] CAAGCCTTC
4297 - 4306: [confidence level: <1] CAACATCAcC
4742 - 4760: [confidence level: <1] CrGdCAAG
P800227 TUBULIN Tubulin subunits alpha, beta and gamma signature :
3832 - 3838: [confidence level: <1] aaGaaGaG

Supplementary Figure 57: The ExPASy result for the protein domains, families and functions for zebrafish ache mRNA FASTA sequence.

References

- Ahn, D. S., Flores-Aguilar, M., Kirsch, L., Munguia, D., Gangan, P., & Freeman, W. R. (1995). Evaluation of retinal toxicity of acetylcholine in rabbit eyes. *Retina*, 15(4), 327-331.
- Alvarez, C. E. (2008). On the origins of arrestin and rhodopsin. *BMC Evol Biol*, 8, 222. doi: 10.1186/1471-2148-8-222
- Amy, C. H., & Susan, A. G. (2012). From Protein to Peptides: a Spectrum of Non-Hydrolytic Functions of Acetylcholinesterase. *Protein & Peptide Letters*, 19(2), 165-172. doi: <http://dx.doi.org/10.2174/092986612799080149>
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol*, 11(10), R106. doi: 10.1186/gb-2010-11-10-r106
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166-169. doi: 10.1093/bioinformatics/btu638
- Appleyard, M. E. (1992). Secreted acetylcholinesterase: non-classical aspects of a classical enzyme. *Trends Neurosci*, 15(12), 485-490.
- Aramillo Irizar, P., Schauble, S., Esser, D., Groth, M., Frahm, C., Priebe, S., . . . Kaleta, C. (2018). Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun*, 9(1), 327. doi: 10.1038/s41467-017-02395-2
- Armant, O., Gourain, V., Etard, C., & Strahle, U. (2016). Whole transcriptome data analysis of zebrafish mutants affecting muscle development. *Data Brief*, 8, 61-68. doi: 10.1016/j.dib.2016.05.007
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., . . . Stockinger, H. (2012). ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res*, 40(Web Server issue), W597-603. doi: 10.1093/nar/gks400
- Asaoka, Y., Mano, H., Kojima, D., & Fukada, Y. (2002). Pineal expression-promoting element (PIPE), a cis-acting element, directs pineal-specific gene expression in zebrafish. *Proc Natl Acad Sci U S A*, 99(24), 15456-15461. doi: 10.1073/pnas.232444199
- Auer, P. L., & Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics*, 185(2), 405-416. doi: 10.1534/genetics.110.114983
- Avanesov, A., & Malicki, J. (2010). Analysis of the retina in the zebrafish model. *Methods Cell Biol*, 100, 153-204. doi: 10.1016/B978-0-12-384892-5.00006-2
- Baker, D. A. a. C. (2018). ggExtra: Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements.
- Barbazuk, W. B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., . . . Johnson, S. L. (2000). The syntenic relationship of the zebrafish and human genomes. *Genome Res*, 10(9), 1351-1358.
- Barinaga, M. (1990). Zebrafish: swimming into the development mainstream. *Science*, 250(4977), 34-35. doi: 10.1126/science.2218513
- Barinaga, M. (1994). Looking to development's future. *Science*, 266(5185), 561-564.
- Behra, M., Cousin, X., Bertrand, C., Vonesch, J. L., Biellmann, D., Chatonnet, A., & Strahle, U. (2002). Acetylcholinesterase is required for neuronal and muscular development in the zebrafish embryo. *Nat Neurosci*, 5(2), 111-118. doi: 10.1038/nn788
- Behra, M., Etard, C., Cousin, X., & Strahle, U. (2004). The use of zebrafish mutants to identify secondary target effects of acetylcholine esterase inhibitors. *Toxicol Sci*, 77(2), 325-333. doi: 10.1093/toxsci/kfh020

- Bertrand, C., Chatonnet, A., Takke, C., Yan, Y. L., Postlethwait, J., Toutant, J. P., & Cousin, X. (2001). Zebrafish acetylcholinesterase is encoded by a single gene localized on linkage group 7. Gene structure and polymorphism; molecular forms and expression pattern during development. *J Biol Chem*, 276(1), 464-474. doi: 10.1074/jbc.M006308200
- Betz, H., Bourgeois, J. P., & Changeux, J. P. (1980). Evolution of cholinergic proteins in developing slow and fast skeletal muscles in chick embryo. *J Physiol*, 302, 197-218. doi: 10.1113/jphysiol.1980.sp013238
- Bilotta, J. (2000). Effects of abnormal lighting on the development of zebrafish visual behavior. *Behav Brain Res*, 116(1), 81-87.
- Bilotta, J., Saszik, S., & Sutherland, S. E. (2001). Rod contributions to the electroretinogram of the dark-adapted developing zebrafish. *Developmental Dynamics*, 222(4), 564-570. doi: 10.1002/dvdy.1188
- Bradford, Y. M., Toro, S., Ramachandran, S., Ruzicka, L., Howe, D. G., Eagle, A., . . . Westerfield, M. (2017). Zebrafish Models of Human Disease: Gaining Insight into Human Disease at ZFIN. *ILAR J*, 58(1), 4-16. doi: 10.1093/ilar/ilw040
- Branchek, T. (1984). The development of photoreceptors in the zebrafish, *Brachydanio rerio*. II. Function. *Journal of Comparative Neurology*, 224(1), 116-122. doi: 10.1002/cne.902240110
- Britto, R., Sallou, O., Collin, O., Michaux, G., Primig, M., & Chalmel, F. (2012). GPSy: a cross-species gene prioritization system for conserved biological processes--application in male gamete development. *Nucleic Acids Res*, 40(Web Server issue), W458-465. doi: 10.1093/nar/gks380
- Bush, S. J., McCulloch, M. E. B., Summers, K. M., Hume, D. A., & Clark, E. L. (2017). Integration of quantitated expression estimates from polyA-selected and rRNA-depleted RNA-seq libraries. *BMC Bioinformatics*, 18(1), 301. doi: 10.1186/s12859-017-1714-9
- Bushel, P. R., Caiment, F., Wu, H., O'Lone, R., Day, F., Calley, J., . . . Li, J. (2018). RATEmiRs: the rat atlas of tissue-specific and enriched miRNAs database. *BMC Genomics*, 19(1), 825. doi: 10.1186/s12864-018-5220-x
- Busnyuk, M. M. (1983). Effect of early light deprivation on acetylcholinesterase activity in the visual cortex. *Neurosci Behav Physiol*, 13(2), 122-125.
- Cai, L., Liao, H. F., Zhang, X. J., Shao, Y., Xu, M., & Yi, J. L. (2013). Acetylcholinesterase function in apoptotic retina pigment epithelial cells induced by H₂O₂. *Int J Ophthalmol*, 6(6), 772-777. doi: 10.3980/j.issn.2222-3959.2013.06.06
- Cai, S., Chen, Y., Shang, Y., Cui, J., Li, Z., & Li, Y. (2018). Knockout of zebrafish interleukin 7 receptor (IL7R) by the CRISPR/Cas9 system delays retinal neurodevelopment. *Cell Death Dis*, 9(3), 273. doi: 10.1038/s41419-018-0337-z
- Carlson, M. (2019). org.Dr.eg.db: Genome wide annotation for Zebrafish.
- Ceriani, L., & Verme, P. (2012). The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3), 421-443. doi: 10.1007/s10888-011-9188-x
- Chapman, G. B., Tarboush, R., & Connaughton, V. P. (2012). The effects of rearing light level and duration differences on the optic nerve, brain, and associated structures in developing zebrafish larvae: a light and transmission electron microscope study. *Anat Rec (Hoboken)*, 295(3), 515-531. doi: 10.1002/ar.22403

- Chen, Z., Huang, W., Dahme, T., Rottbauer, W., Ackerman, M. J., & Xu, X. (2008). Depletion of zebrafish essential and regulatory myosin light chains reduces cardiac function through distinct mechanisms. *Cardiovasc Res*, 79(1), 97-108. doi: 10.1093/cvr/cvn073
- Chhetri, J., Jacobson, G., & Gueven, N. (2014). Zebrafish--on the move towards ophthalmological research. *Eye (Lond)*, 28(4), 367-380. doi: 10.1038/eye.2014.19
- Chitramuthu, B. P., & Bennett, H. P. (2013). High resolution whole mount in situ hybridization within zebrafish embryos to study gene expression and function. *J Vis Exp*(80), e50644. doi: 10.3791/50644
- Clark, D. T. (1981). Visual Responses in Developing Zebrafish (*Brachydanio rerio*).
- Clemente, D., Porteros, A., Weruaga, E., Alonso, J. R., Arenzana, F. J., Aijon, J., & Arevalo, R. (2004). Cholinergic elements in the zebrafish central nervous system: Histochemical and immunohistochemical analysis. *J Comp Neurol*, 474(1), 75-107. doi: 10.1002/cne.20111
- Collins, J. E., White, S., Searle, S. M., & Stemple, D. L. (2012). Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res*, 22(10), 2067-2078. doi: 10.1101/gr.137901.112
- Colovic, M. B., Krstic, D. Z., Lazarevic-Pasti, T. D., Bondzic, A. M., & Vasic, V. M. (2013). Acetylcholinesterase inhibitors: pharmacology and toxicology. *Curr Neuroparmacol*, 11(3), 315-335. doi: 10.2174/1570159X11311030006
- Coppe, A., Danieli, G. A., & Bortoluzzi, S. (2006). REEF: searching REgionally Enriched Features in genomes. *BMC Bioinformatics*, 7, 453. doi: 10.1186/1471-2105-7-453
- Corley, S. M., MacKenzie, K. L., Beverdam, A., Roddam, L. F., & Wilkins, M. R. (2017). Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, 18(1), 399. doi: 10.1186/s12864-017-3797-0
- Correa Mde, C., Maldonado, P., da Rosa, C. S., Lunkes, G., Lunkes, D. S., Kaizer, R. R., . . . Schetinger, M. R. (2008). Oxidative stress and erythrocyte acetylcholinesterase (AChE) in hypertensive and ischemic patients of both acute and chronic stages. *Biomed Pharmacother*, 62(5), 317-324. doi: 10.1016/j.biopha.2007.10.002
- Corvin, A., & Axelrod, S. (1969). Light deprivation from birth and acetylcholinesterase activity in cortex of mature rats. *Psychonomic Science*, 14(5), 251-252. doi: 10.3758/bf03332822
- Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One*, 12(12), e0190152. doi: 10.1371/journal.pone.0190152
- Costa, M. L., Escalera, R. C., Rodrigues, V. B., Manasfi, M., & Mermelstein, C. S. (2002). Some distinctive features of zebrafish myogenesis based on unexpected distributions of the muscle cytoskeletal proteins actin, myosin, desmin, α -actinin, troponin and titin. *Mechanisms of Development*, 116(1), 95-104. doi: https://doi.org/10.1016/S0925-4773(02)00149-1
- Cousin, X., Strahle, U., & Chatonnet, A. (2005). Are there non-catalytic functions of acetylcholinesterases? Lessons from mutant animal models. *Bioessays*, 27(2), 189-200. doi: 10.1002/bies.20153
- D'Souza, C. A., Chopra, V., Varhol, R., Xie, Y. Y., Bohacec, S., Zhao, Y., . . . Jones, S. J. (2008). Identification of a set of genes showing regionally enriched expression in the mouse brain. *BMC Neurosci*, 9, 66. doi: 10.1186/1471-2202-9-66

- DALE, H. H. (1914). THE ACTION OF CERTAIN ESTERS AND ETHERS OF CHOLINE, AND THEIR RELATION TO MUSCARINE. *Journal of Pharmacology and Experimental Therapeutics*, 6(2), 147-190.
- Dekens, M. P., Santoriello, C., Vallone, D., Grassi, G., Whitmore, D., & Foulkes, N. S. (2003). Light regulates the cell cycle in zebrafish. *Curr Biol*, 13(23), 2051-2057.
- Dekens, M. P., & Whitmore, D. (2008). Autonomous onset of the circadian clock in the zebrafish embryo. *EMBO J*, 27(20), 2757-2765. doi: 10.1038/emboj.2008.183
- Desvignes, T., Batzel, P., Sydes, J., Eames, B. F., & Postlethwait, J. H. (2019). miRNA analysis with Prost! reveals evolutionary conservation of organ-enriched expression and post-transcriptional modifications in three-spined stickleback and zebrafish. *Sci Rep*, 9(1), 3913. doi: 10.1038/s41598-019-40361-8
- Dezso, Z., Nikolsky, Y., Sviridov, E., Shi, W., Serebriyskaya, T., Dosymbekov, D., . . . Nikolskaya, T. (2008). A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol*, 6, 49. doi: 10.1186/1741-7007-6-49
- Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., . . . French StatOmique, C. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*, 14(6), 671-683. doi: 10.1093/bib/bbs046
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. doi: 10.1093/bioinformatics/bts635
- Dobin, A., & Gingeras, T. R. (2016). Optimizing RNA-Seq Mapping with STAR. In O. Carugo & F. Eisenhaber (Eds.), *Data Mining Techniques for the Life Sciences* (pp. 245-262). New York, NY: Springer New York.
- Dougherty, J. D., Schmidt, E. F., Nakajima, M., & Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res*, 38(13), 4218-4230. doi: 10.1093/nar/gkq130
- Downes, G. B., & Granato, M. (2004). Acetylcholinesterase function is dispensable for sensory neurite growth but is critical for neuromuscular synapse stability. *Developmental Biology*, 270(1), 232-245. doi: <https://doi.org/10.1016/j.ydbio.2004.02.027>
- Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., . . . Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development*, 123, 37-46.
- Driever, W., Stemple, D., Schier, A., & Solnica-Krezel, L. (1994). Zebrafish: genetic tools for studying vertebrate development. *Trends Genet*, 10(5), 152-159.
- Duret, L., & Mouchiroud, D. (2000). Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol*, 17(1), 68-74. doi: 10.1093/oxfordjournals.molbev.a026239
- Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4, 1184. doi: 10.1038/nprot.2009.97
- <https://www.nature.com/articles/nprot.2009.97#supplementary-information>
- Earnest, D., & Turek, F. (1983). Role for acetylcholine in mediating effects of light on reproduction. *Science*, 219(4580), 77-79. doi: 10.1126/science.6849121

- Easter, J. S. S., & Nicola, G. N. (1996). The Development of Vision in the Zebrafish (*Danio rerio*). *Developmental Biology*, 180(2), 646-663. doi: <https://doi.org/10.1006/dbio.1996.0335>
- Eastlake, K., Heywood, W. E., Tracey-White, D., Aquino, E., Bliss, E., Vasta, G. R., . . . Limb, G. A. (2017). Comparison of proteomic profiles in the zebrafish retina during experimental degeneration and regeneration. *Sci Rep*, 7, 44601. doi: 10.1038/srep44601
- Elo, B., Villano, C. M., Govorko, D., & White, L. A. (2007). Larval zebrafish as a model for glucose metabolism: expression of phosphoenolpyruvate carboxykinase as a marker for exposure to anti-diabetic compounds. *J Mol Endocrinol*, 38(4), 433-440. doi: 10.1677/JME-06-0037
- Etard, C., Armant, O., Roostalu, U., Gourain, V., Ferg, M., & Strahle, U. (2015). Loss of function of myosin chaperones triggers Hsf1-mediated transcriptional response in skeletal muscle cells. *Genome Biol*, 16, 267. doi: 10.1186/s13059-015-0825-8
- Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., Dinger, M. E., Hellemans, J., & Mestdagh, P. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci Rep*, 7(1), 1559. doi: 10.1038/s41598-017-01617-3
- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*, 26(2), 105-109. doi: 10.3969/j.issn.1002-0829.2014.02.009
- FISHER, K. A. (1935). Statistical Tests. *Nature*. doi: 10.1038/136474b0
- Friederike Dündar, L. S., Paul Zumbo. (2015). Introduction to differential gene expression analysis using RNA-seq.
- Fukuto, T. R. (1990). Mechanism of action of organophosphorus and carbamate insecticides. *Environ Health Perspect*, 87, 245-254. doi: 10.1289/ehp.9087245
- Gao, Y., Dai, Z., Shi, C., Zhai, G., Jin, X., He, J., . . . Yin, Z. (2016). Depletion of Myostatin b Promotes Somatic Growth and Lipid Metabolism in Zebrafish. *Front Endocrinol (Lausanne)*, 7, 88. doi: 10.3389/fendo.2016.00088
- Gilboa-Geffen, A., Hartmann, G., & Soreq, H. (2012). Stressing hematopoiesis and immunity: an acetylcholinesterase window into nervous and immune system interactions. *Front Mol Neurosci*, 5, 30. doi: 10.3389/fnmol.2012.00030
- Glow, P. H., & Rose, S. A. M. (1964). Effects of Light and Dark on the Acetylcholinesterase Activity of the Retina. *Nature*, 202(4930), 422-423. doi: 10.1038/202422b0
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, 17(6), 333-351. doi: 10.1038/nrg.2016.49
- Gosalia, N., Economides, A. N., Dewey, F. E., & Balasubramanian, S. (2017). MAPPIN: a method for annotating, predicting pathogenicity and mode of inheritance for nonsynonymous variants. *Nucleic Acids Res*, 45(18), 10393-10402. doi: 10.1093/nar/gkx730
- Guo, M., Wang, H., Potter, S. S., Whitsett, J. A., & Xu, Y. (2015). SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput Biol*, 11(11), e1004575. doi: 10.1371/journal.pcbi.1004575
- Hanovice, N. J., Leach, L. L., Slater, K., Gabriel, A. E., Romanovicz, D., Shao, E., . . . Gross, J. M. (2019). Regeneration of the zebrafish retinal pigment epithelium after widespread genetic ablation. *PLoS Genet*, 15(1), e1007939. doi: 10.1371/journal.pgen.1007939

- Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., . . . Flicek, P. (2016). Ensembl comparative genomics resources. *Database (Oxford)*, 2016. doi: 10.1093/database/bav096
- Hervé Pagès, M. C., Seth Falcon, Nianhua Li (2019). AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor.
- Hodel, C., Neuhauss, S. C., & Biehlmaier, O. (2006). Time course and development of light adaptation processes in the outer zebrafish retina. *Anat Rec A Discov Mol Cell Evol Biol*, 288(6), 653-662. doi: 10.1002/ar.a.20329
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., . . . Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498-503. doi: 10.1038/nature12111
- Hu, T., Fu, Q., Liu, X., Zhang, H., & Dong, M. (2009). Increased acetylcholinesterase and capase-3 expression in the brain and peripheral immune system of focal cerebral ischemic rats. *J Neuroimmunol*, 211(1-2), 84-91. doi: 10.1016/j.jneuroim.2009.04.002
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1), 1-13. doi: 10.1093/nar/gkn923
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57. doi: 10.1038/nprot.2008.211
- Jain, A., & Tuteja, G. (2018). TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics*. doi: 10.1093/bioinformatics/bty890
- Jimenez-Amilburu, V., Jong-Raadsen, S., Bakkers, J., Spaink, H. P., & Marin-Juez, R. (2015). GLUT12 deficiency during early development results in heart failure and a diabetic phenotype in zebrafish. *J Endocrinol*, 224(1), 1-15. doi: 10.1530/JOE-14-0539
- Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schutz, F., . . . Kaessmann, H. (2012). Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol*, 10(5), e1001328. doi: 10.1371/journal.pbio.1001328
- Kadota, K., Ye, J., Nakai, Y., Terada, T., & Shimizu, K. (2006). ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinformatics*, 7, 294. doi: 10.1186/1471-2105-7-294
- Kang, R., Zhang, Y., Huang, Q., Meng, J., Ding, R., Chang, Y., . . . Guo, Z. (2019). EnhancerDB: a resource of transcriptional regulation in the context of enhancers. *Database (Oxford)*, 2019. doi: 10.1093/database/bay141
- Kazimi, N., & Cahill, G. M. (1999). Development of a circadian melatonin rhythm in embryonic zebrafish. *Brain Res Dev Brain Res*, 117(1), 47-52.
- Kim, P., Park, A., Han, G., Sun, H., Jia, P., & Zhao, Z. (2018). TissGDB: tissue-specific gene database in cancer. *Nucleic Acids Res*, 46(D1), D1031-D1038. doi: 10.1093/nar/gkx850
- Klepikova, A. V., Kulakovskiy, I. V., Kasianov, A. S., Logacheva, M. D., & Penin, A. A. (2019). An update to database TraVA: organ-specific cold stress response in *Arabidopsis thaliana*. *BMC Plant Biol*, 19(Suppl 1), 49. doi: 10.1186/s12870-019-1636-y
- Kluver, N., Yang, L., Busch, W., Scheffler, K., Renner, P., Strahle, U., & Scholz, S. (2011). Transcriptional response of zebrafish embryos exposed to neurotoxic compounds reveals a muscle activity dependent hspb11 expression. *PLoS One*, 6(12), e29063. doi: 10.1371/journal.pone.0029063

- Komljenovic, A., Roux, J., Wollbrett, J., Robinson-Rechavi, M., & Bastian, F. B. (2016). BgeeDB, an R package for retrieval of curated expression datasets and for gene list expression localization enrichment tests. *F1000Res*, 5, 2748. doi: 10.12688/f1000research.9973.2
- Korenbrodt, J. I. (2012). Speed, sensitivity, and stability of the light response in rod and cone photoreceptors: facts and models. *Prog Retin Eye Res*, 31(5), 442-466. doi: 10.1016/j.preteyeres.2012.05.002
- Korenbrodt, J. I., Mehta, M., Tserentsoodol, N., Postlethwait, J. H., & Rebrink, T. I. (2013). EML1 (CNG-modulin) controls light sensitivity in darkness and under continuous illumination in zebrafish retinal cone photoreceptors. *J Neurosci*, 33(45), 17763-17776. doi: 10.1523/JNEUROSCI.2659-13.2013
- Kreutzberg, G. W. (1969). Neuronal dynamics and axonal flow. IV. Blockage of intra-axonal enzyme transport by colchicine. *Proc Natl Acad Sci U S A*, 62(3), 722-728. doi: 10.1073/pnas.62.3.722
- Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2016). Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. *PLoS Comput Biol*, 12(12), e1005274. doi: 10.1371/journal.pcbi.1005274
- Kuksa, P. P., Amlie-Wolf, A., Katanic, Z., Valladares, O., Wang, L. S., & Leung, Y. Y. (2019). DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. *Bioinformatics*, 35(6), 1033-1039. doi: 10.1093/bioinformatics/bty709
- Laale, H. W. (1977). The biology and use of zebrafish, *Brachydanio rerio* in fisheries research. *Journal of Fish Biology*, 10(2), 121-173. doi: 10.1111/j.1095-8649.1977.tb04049.x
- Lau, J. W., Lehnert, E., Sethi, A., Malhotra, R., Kaushik, G., Onder, Z., . . . Seven Bridges, C. G. C. T. (2017). The Cancer Genomics Cloud: Collaborative, Reproducible, and Democratized-A New Paradigm in Large-Scale Computational Research. *Cancer Res*, 77(21), e3-e6. doi: 10.1158/0008-5472.CAN-17-0387
- Layer, P. G. (1990). Cholinesterases preceding major tracts in vertebrate neurogenesis. *Bioessays*, 12(9), 415-420. doi: 10.1002/bies.950120904
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., . . . Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*, 11(10), 733-739. doi: 10.1038/nrg2825
- Lemay, D. G., Martin, W. F., Hinrichs, A. S., Rijnkels, M., German, J. B., Korf, I., & Pollard, K. S. (2012). G-NEST: a gene neighborhood scoring tool to identify co-conserved, co-expressed genes. *BMC Bioinformatics*, 13, 253. doi: 10.1186/1471-2105-13-253
- Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genetics*, 31(2), 180-183. doi: 10.1038/ng887
- Leung, Y. F., Ma, P., & Dowling, J. E. (2007). Gene expression profiling of zebrafish embryonic retinal pigment epithelium in vivo. *Invest Ophthalmol Vis Sci*, 48(2), 881-890. doi: 10.1167/iovs.06-0723
- Li, M., Andersson-Lendahl, M., Sejersen, T., & Arner, A. (2013). Knockdown of desmin in zebrafish larvae affects interfibrillar spacing and mechanical properties of skeletal muscle. *J Gen Physiol*, 141(3), 335-345. doi: 10.1085/jgp.201210915
- Li, P., Temple, S., Gao, Y., Haimberger, T. J., Hawryshyn, C. W., & Li, L. (2005). Circadian rhythms of behavioral cone sensitivity and long wavelength opsin mRNA expression: a correlation study in zebrafish. *J Exp Biol*, 208(Pt 3), 497-504. doi: 10.1242/jeb.01424

- Li, P., Tompkins, R. G., Xiao, W., the, I., & Host Response to Injury Large-Scale Collaborative Research, P. (2017). KERIS: kaleidoscope of gene responses to inflammation between species. *Nucleic Acids Res*, 45(D1), D908-D914. doi: 10.1093/nar/gkw974
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4), 402-408. doi: 10.1006/meth.2001.1262
- Loewi, O., & Navratil, E. (1926). Über humorale Übertragbarkeit der Herznervenwirkung. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 214(1), 678-688. doi: 10.1007/bf01741946
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550. doi: 10.1186/s13059-014-0550-8
- Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., . . . Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Res*, 44(8), 3865-3877. doi: 10.1093/nar/gkw116
- Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., . . . Zhang, Z. (2019). LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res*, 47(D1), D128-D134. doi: 10.1093/nar/gky960
- Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., . . . Saito, K. (2010). AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant Physiol*, 152(2), 566-578. doi: 10.1104/pp.109.148031
- Maza, E., Frasse, P., Senin, P., Bouzayen, M., & Zouine, M. (2013). Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments: A matter of relative size of studied transcriptomes. *Commun Integr Biol*, 6(6), e25849. doi: 10.4161/cib.25849
- Mazzarelli, J. M., Brestelli, J., Gorski, R. K., Liu, J., Manduchi, E., Pinney, D. F., . . . Stoeckert, C. J., Jr. (2007). EPConDB: a web resource for gene expression related to pancreatic development, beta-cell function and diabetes. *Nucleic Acids Res*, 35(Database issue), D751-755. doi: 10.1093/nar/gkl748
- Mena, S., Ortega, A., & Estrela, J. M. (2009). Oxidative stress in environmental-induced carcinogenesis. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 674(1), 36-44. doi: https://doi.org/10.1016/j.mrgentox.2008.09.017
- Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., & Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief Bioinform*, 17(4), 628-641. doi: 10.1093/bib/bbv108
- Menger, G. J., Koke, J. R., & Cahill, G. M. (2005). Diurnal and circadian retinomotor movements in zebrafish. *Vis Neurosci*, 22(2), 203-209. doi: 10.1017/S0952523805222083
- Mesens, N., Crawford, A. D., Menke, A., Hung, P. D., Van Goethem, F., Nuyts, R., . . . Esguerra, C. V. (2015). Are zebrafish larvae suitable for assessing the hepatotoxicity potential of drug candidates? *J Appl Toxicol*, 35(9), 1017-1029. doi: 10.1002/jat.3091
- Meyers, J. R. (2018). Zebrafish: Development of a Vertebrate Model Organism. *Current Protocols Essential Laboratory Techniques*, 16(1), e19. doi: 10.1002/cpet.19
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5(7), 621-628. doi: 10.1038/nmeth.1226

- Murakami, N., Takahashi, K., & Kawashima, K. (1984). Effect of light on the acetylcholine concentrations of the suprachiasmatic nucleus in the rat. *Brain Res*, 311(2), 358-360.
- Nookaew, I., Papini, M., Pornputtpong, N., Scalcinati, G., Fagerberg, L., Uhlen, M., & Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 40(20), 10084-10097. doi: 10.1093/nar/gks804
- Pan, J. B., Hu, S. C., Shi, D., Cai, M. C., Li, Y. B., Zou, Q., & Ji, Z. L. (2013). PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. *PLoS One*, 8(12), e80747. doi: 10.1371/journal.pone.0080747
- Paraoanu, L. E., & Layer, P. G. (2008). Acetylcholinesterase in cell adhesion, neurite growth and network formation. *The FEBS Journal*, 275(4), 618-624. doi: 10.1111/j.1742-4658.2007.06237.x
- Park, S. G., & Choi, S. S. (2010). Expression breadth and expression abundance behave differently in correlations with evolutionary rates. *BMC Evol Biol*, 10, 241. doi: 10.1186/1471-2148-10-241
- Petryszak, R., Keays, M., Tang, Y. A., Fonseca, N. A., Barrera, E., Burdett, T., . . . Brazma, A. (2016). Expression Atlas update--an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res*, 44(D1), D746-752. doi: 10.1093/nar/gkv1045
- Pohanka, M. (2014). Inhibitors of acetylcholinesterase and butyrylcholinesterase meet immunity. *Int J Mol Sci*, 15(6), 9809-9825. doi: 10.3390/ijms15069809
- Ponger, L., Duret, L., & Mouchiroud, D. (2001). Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Res*, 11(11), 1854-1860. doi: 10.1101/gr.174501
- Proost, S., & Mutwil, M. (2018). CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic Acids Res*, 46(W1), W133-W140. doi: 10.1093/nar/gky336
- Qian Liu, M. M. (2016). Evaluation of Methods in Removing Batch Effects on RNA-seq Data. *Infect Dis Transl Med*, 2(1), 3-9. doi: 10.11979/itdm.201601002
- Rakonczay, Z., Horvath, Z., Juhasz, A., & Kalman, J. (2005). Peripheral cholinergic disturbances in Alzheimer's disease. *Chem Biol Interact*, 157-158, 233-238. doi: 10.1016/j.cbi.2005.10.034
- Ramspacher, C., Steed, E., Boselli, F., Ferreira, R., Faggianelli, N., Roth, S., . . . Vermot, J. (2015). Developmental Alterations in Heart Biomechanics and Skeletal Muscle Function in Desmin Mutants Suggest an Early Pathological Root for Desminopathies. *Cell Rep*, 11(10), 1564-1576. doi: 10.1016/j.celrep.2015.05.010
- Renninger, S. L., Gesemann, M., & Neuhauss, S. C. (2011). Cone arrestin confers cone vision of high temporal resolution in zebrafish larvae. *Eur J Neurosci*, 33(4), 658-667. doi: 10.1111/j.1460-9568.2010.07574.x
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. doi: 10.1093/bioinformatics/btp616
- Rompler, H., Staubert, C., Thor, D., Schulz, A., Hofreiter, M., & Schoneberg, T. (2007). G protein-coupled time travel: evolutionary aspects of GPCR research. *Mol Interv*, 7(1), 17-25. doi: 10.1124/mi.7.1.5

- Sallou, O., Duek, P. D., Darde, T. A., Collin, O., Lane, L., & Chalmel, F. (2016). PepPSy: a web server to prioritize gene products in experimental and biocuration workflows. *Database (Oxford)*, 2016. doi: 10.1093/database/baw070
- Sanchez-Chavez, G., Vidal, C. J., & Salceda, R. (1995). Acetyl- and butyrylcholinesterase activities in the rat retina and retinal pigment epithelium. *J Neurosci Res*, 41(5), 655-662. doi: 10.1002/jnr.490410512
- Schonthaler, H. B., Lampert, J. M., Isken, A., Rinner, O., Mader, A., Gesemann, M., . . . von Lintig, J. (2007). Evidence for RPE65-independent vision in the cone-dominated zebrafish retina. *Eur J Neurosci*, 26(7), 1940-1949. doi: 10.1111/j.1460-9568.2007.05801.x
- Schug, J., Schuller, W. P., Kappen, C., Salbaum, J. M., Bucan, M., & Stoeckert, C. J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol*, 6(4), R33. doi: 10.1186/gb-2005-6-4-r33
- Schurch, N. J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., . . . Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6), 839-851. doi: 10.1261/rna.053959.115
- Sepich, D. S., Wegner, J., O'Shea, S., & Westerfield, M. (1998). An altered intron inhibits synthesis of the acetylcholine receptor alpha-subunit in the paralyzed zebrafish mutant *nic1*. *Genetics*, 148(1), 361-372.
- Simoneau, J., & Scott, M. S. (2019). In silico analysis of RNA-seq requires a more complete description of methodology. *Nat Rev Mol Cell Biol*. doi: 10.1038/s41580-019-0137-z
- Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14, 91. doi: 10.1186/1471-2105-14-91
- Soreq, H., & Seidman, S. (2001). Acetylcholinesterase--new roles for an old actor. *Nat Rev Neurosci*, 2(4), 294-302. doi: 10.1038/35067589
- Subramanian, S., & Kumar, S. (2004). Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics*, 168(1), 373-381. doi: 10.1534/genetics.104.028944
- Sudmant, P. H., Alexis, M. S., & Burge, C. B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biol*, 16, 287. doi: 10.1186/s13059-015-0853-4
- Szumilas, M. (2010). Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*, 19(3), 227-229.
- t Hoen, P. A., Friedlander, M. R., Almløf, J., Sammeth, M., Pulyakhina, I., Anvar, S. Y., . . . Lappalainen, T. (2013). Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat Biotechnol*, 31(11), 1015-1022. doi: 10.1038/nbt.2702
- Tamai, T. K., Vardhanabhuti, V., Foulkes, N. S., & Whitmore, D. (2004). Early embryonic light detection improves survival. *Current Biology*, 14(3), R104-R105. doi: <https://doi.org/10.1016/j.cub.2004.01.014>
- Taylor, S., Chen, J., Luo, J., & Hitchcock, P. (2012). Light-Induced Photoreceptor Degeneration in the Retina of the Zebrafish. In S.-Z. Wang (Ed.), *Retinal Development: Methods and Protocols* (pp. 247-254). Totowa, NJ: Humana Press.
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220), 1260419. doi: 10.1126/science.1260419

- Vallone, D., Lahiri, K., Dickmeis, T., & Foulkes, N. S. (2005). Zebrafish cell clocks feel the heat and see the light! *Zebrafish*, 2(3), 171-187. doi: 10.1089/zeb.2005.2.171
- Vallone, D., Lahiri, K., Dickmeis, T., & Foulkes, N. S. (2007). Start the clock! Circadian rhythms and development. *Dev Dyn*, 236(1), 142-155. doi: 10.1002/dvdy.20998
- van der Meulen, T., Schipper, H., van Leeuwen, J. L., & Kranenbarg, S. (2005). Effects of decreased muscle activity on developing axial musculature in nicb107 mutant zebrafish (*Danio rerio*). *J Exp Biol*, 208(Pt 19), 3675-3687. doi: 10.1242/jeb.01826
- Venkatachalam, A. B., Thisse, C., Thisse, B., & Wright, J. M. (2009). Differential tissue-specific distribution of transcripts for the duplicated fatty acid-binding protein 10 (fabp10) genes in embryos, larvae and adult zebrafish (*Danio rerio*). *The FEBS Journal*, 276(22), 6787-6797. doi: 10.1111/j.1742-4658.2009.07393.x
- Vihtelic, T. S., Fadool, J. M., Gao, J., Thornton, K. A., Hyde, D. R., & Wistow, G. (2005). Expressed sequence tag analysis of zebrafish eye tissues for NEIBank. *Mol Vis*, 11, 1083-1100.
- Villamizar, N., Vera, L. M., Foulkes, N. S., & Sanchez-Vazquez, F. J. (2014). Effect of lighting conditions on zebrafish growth and development. *Zebrafish*, 11(2), 173-181. doi: 10.1089/zeb.2013.0926
- Vinogradov, A. E. (2003). Isochores and tissue-specificity. *Nucleic Acids Res*, 31(17), 5212-5220. doi: 10.1093/nar/gkg699
- Vliegenthart, A. D., Tucker, C. S., Del Pozo, J., & Dear, J. W. (2014). Zebrafish as model organisms for studying drug-induced liver injury. *Br J Clin Pharmacol*, 78(6), 1217-1227. doi: 10.1111/bcp.12408
- Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*, 131(4), 281-285. doi: 10.1007/s12064-012-0162-3
- Wang, Q., Chen, Q., Zhao, K., Wang, L., Wang, L., & Traboulsi, E. I. (2001). Update on the molecular genetics of retinitis pigmentosa. *Ophthalmic Genet*, 22(3), 133-154.
- Wang, Q., Yue, W. W. S., Jiang, Z., Xue, T., Kang, S. H., Bergles, D. E., . . . Yau, K. W. (2017). Synergistic Signaling by Light and Acetylcholine in Mouse Iris Sphincter Muscle. *Curr Biol*, 27(12), 1791-1800 e1795. doi: 10.1016/j.cub.2017.05.022
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1), 57-63. doi: 10.1038/nrg2484
- Wang, Z., Mao, Y., Cui, T., Tang, D., & Wang, X. L. (2013). Impact of a combined high cholesterol diet and high glucose environment on vasculature. *PLoS One*, 8(12), e81485. doi: 10.1371/journal.pone.0081485
- Weirick, T., John, D., Dimmeler, S., & Uchida, S. (2015). C-It-Loci: a knowledge database for tissue-enriched loci. *Bioinformatics*, 31(21), 3537-3543. doi: 10.1093/bioinformatics/btv410
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.
- Wixon, J., & O'Kane, C. (2000). Featured organism: *Drosophila melanogaster*. *Yeast*, 17(2), 146-153. doi: 10.1002/1097-0061(20000630)17:2<146::AID-YEA24>3.0.CO;2-A
- Wood, N., & Rose, S. P. R. (1979). Changes in acetylcholinesterase with light exposure, time of day, and motor activity in the rat. *Behavioral and Neural Biology*, 25(1), 79-89. doi: https://doi.org/10.1016/S0163-1047(79)90790-8
- Xiao, S. J., Zhang, C., Zou, Q., & Ji, Z. L. (2010). TiSGeD: a database for tissue-specific genes. *Bioinformatics*, 26(9), 1273-1275. doi: 10.1093/bioinformatics/btq109

- Xie, J., Jiang, H., Wan, Y. H., Du, A. Y., Guo, K. J., Liu, T., . . . Zhang, X. J. (2011). Induction of a 55 kDa acetylcholinesterase protein during apoptosis and its negative regulation by the Akt pathway. *J Mol Cell Biol*, 3(4), 250-259. doi: 10.1093/jmcb/mjq047
- Xie, W., Stribley, J. A., Chatonnet, A., Wilder, P. J., Rizzino, A., McComb, R. D., . . . Lockridge, O. (2000). Postnatal developmental delay and supersensitivity to organophosphate in gene-targeted mice lacking acetylcholinesterase. *J Pharmacol Exp Ther*, 293(3), 896-902.
- Xie, Y., Xiang, Y., Sheng, J., Zhang, D., Yao, X., Yang, Y., & Zhang, X. (2018). Immunotherapy for Hepatocellular Carcinoma: Current Advances and Future Expectations. *J Immunol Res*, 2018, 8740976. doi: 10.1155/2018/8740976
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., . . . Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5), 650-659. doi: 10.1093/bioinformatics/bti042
- Yu, G., & He, Q. Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst*, 12(2), 477-479. doi: 10.1039/c5mb00663e
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, 16(5), 284-287. doi: 10.1089/omi.2011.0118
- Zambelli, F., Mastropasqua, F., Picardi, E., D'Erchia, A. M., Pesole, G., & Pavesi, G. (2018). RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res*, 46(8), e46. doi: 10.1093/nar/gky055
- Zhang, Y., Liu, H., Lv, J., Xiao, X., Zhu, J., Liu, X., . . . Cui, Y. (2011). QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res*, 39(9), e58. doi: 10.1093/nar/gkr053
- Zhao, S., Zhang, Y., Gordon, W., Quan, J., Xi, H., Du, S., . . . Zhang, B. (2015). Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics*, 16, 675. doi: 10.1186/s12864-015-1876-7
- Zhao, Y., Wang, X., Wang, T., Hu, X., Hui, X., Yan, M., . . . He, X. (2011). Acetylcholinesterase, a key prognostic predictor for hepatocellular carcinoma, suppresses cell growth and induces chemosensitization. *Hepatology*, 53(2), 493-503. doi: 10.1002/hep.24079
- Ziv, L., & Gothilf, Y. (2006). Circadian time-keeping during early stages of development. *Proc Natl Acad Sci U S A*, 103(11), 4146-4151. doi: 10.1073/pnas.0600571103