UNIVERSIDADE NOVA DE LISBOA

MScCBBi

MASTER IN
COMPUTATIONAL BIOLOGY & BIOINFORMATICS

JOÃO DUARTE PIEDADE LOPES
BSc in Human Biology

Multilayer molecular networks shaping goby fish physiological responses to marine heatwaves in intertidal habitats

Sep, 2024

iTQB nova
NOVA SCHOOL OF SCIENCE & TECHNOLOGY
NOVA MEDICAL SCHOOL
NOVA IMS Information Management School
DESDE 1902 INSTITUTO DE HIGIENE E MEDICINA TROPICAL
UNIVERSIDADE NOVA DE LISBOA

DEPARTAMENT OF
LIFE SCIENCES

NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

MULTILAYER MOLECULAR NETWORKS SHAPING GOBY FISH PHYSIOLOGICAL RESPONSES TO MARINE HEATWAVES IN INTERTIDAL HABITATS

JOÃO DUARTE PIEDADE LOPES

BSc in Human Biology

Adviser: Pedro Manuel Brôa Costa
Assistant Professor, NOVA School of Science and Technology, NOVA University Lisbon

Co-adviser: Sara Carolina Gusmão Coito Madeira
Researcher, NOVA School of Science and Technology, NOVA University Lisbon

Examination Committee:

Chair: Paula Maria Theriaga Mendes Bernardo Gonçalves
Associate Professor, NOVA School of Science and Technology, NOVA University Lisbon

Rapporteurs: Daniel Vieira Noro e Silva Sobral,
Assistant Researcher at the Institute National Health Doctor Ricardo Jorge

Adviser: Pedro Manuel Brôa Costa,
Assistant Professor, NOVA School of Science and Technology, NOVA University Lisbon

MASTER IN COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

NOVA University Lisbon
Dec, 2024

MULTILAYER MOLECULAR NETWORKS SHAPING GOBY FISH PHYSIOLOGICAL RESPONSES TO MARINE HEATWAVES IN INTERTIDAL HABITATS

Acknowledgments

First, I'm incredibly grateful to Prof. Pedro Costa and Carolina Madeira for all their support and guidance.

To all the Seatox team, especially, Inês, Mariaelena, Ana and Carla, for all the tips and support.

To my colleagues and friends, particularly, David, thank you for always being there to listen and help me troubleshoot my code errors.

To my girlfriend, for all the times you listened to me complain and helped me make the right decision

A huge thank you to my parents and brother, for always believing in me and standing by my side through everything.

Lastly, a special thank you to my uncle Davide and aunt Rute for helping me keep my cool and guiding me through the tough moments. Your knowledge and teaching truly made a difference when I needed it most.

Your support has been invaluable, and I am deeply appreciative of each and every one of you.

Abstract

Climate change has a profound impact on marine life, through ocean warming, water acidification, and the loss of essential habitats for various species. One example of climate change-related phenomena is the surge in marine heatwaves, which are characterized by anomalous increases in water temperature that can occur over long or short periods. These phenomena are already common in estuaries and coastal lagoons, making them one of the greatest threats to the species that inhabit these areas. *Pomatochistus microps,* a common intertidal fish found between Morocco to Norway, as well as in the Mediterranean Sea, was selected as biological model of this study, as there is limited knowledge about the effects of marine heatwaves on fish and their resilience to climate change. This study aims to analyze the impacts of prolonged marine heatwaves compared to sequential marine heatwaves on *Pomatochistus microps*, as well as to explore the benefits of recovery periods between these waves. The research was RNA-seq-based transcriptomics and proteomics performed on *P.microps* muscle. Due to the lack of genomic annotations for this species, various bioinformatics tools, such as Trinity, were used to assemble new transcriptomes. An R pipeline was developed to analyze the obtained data. A greater number of differentially expressed and regulated genes and proteins were identified in fish exposed to two consecutive heatwaves. HSPs, especially HSP70, and Ubiquitin carboxyl-terminal hydrolase isozyme L1, showed negative regulation in fish that had a recovery period. Changes in metabolic pathways, signaling pathways, muscle cell maintenance, and cellular damage repair mechanisms were identified. Overall, animals that went through a recovery period showed more efficient acclimation and likely a comparative reduction in damage caused by thermal stress.

Keywords: Climate change; Marine heatwaves; *Pomatochistus microps*; Transcriptomics; Proteomics

Resumo

A mudança climática tem um impacto profundo na vida marinha, através do aquecimento dos oceanos, acidificação da água e perda de habitats essenciais para várias espécies. Um exemplo de fenómenos relacionados com as alterações climáticas é o aumento de ondas de calor marinhas, que se caracterizam por aumentos anómalos da temperatura da água, podendo ocorrer durante períodos longos ou curtos. Estes fenómenos já são comuns em estuários e lagoas costeiras, tornando-se uma das maiores ameaças para as espécies que habitam estas áreas. *Pomatochistus microps*, um peixe intertidal comum encontrado entre Marrocos e a Noruega, bem como no Mar Mediterrâneo, foi selecionado como modelo biológico deste estudo, dado o conhecimento limitado sobre os efeitos das ondas de calor marinhas nos peixes e a sua resiliência às mudanças climáticas. Este estudo tem como objetivo analisar os impactos de ondas de calor marinhas prolongadas, comparadas a ondas de calor marinhas sequenciais, em *Pomatochistus microps*, bem como explorar os benefícios dos períodos de recuperação entre essas ondas. A investigação baseou-se em transcriptómica por RNA-seq e proteómica realizada em músculo de *P. microps*. Devido à falta de anotações genómicas para esta espécie, várias ferramentas de bioinformática, como o Trinity, foram usadas para montar novos transcriptomas. Foi desenvolvido um pipeline em R para analisar os dados obtidos. Um maior número de genes e proteínas diferencialmente expressos e regulados foi identificado em peixes expostos a duas ondas de calor consecutivas. As proteínas de choque térmico (HSPs), especialmente a HSP70, e a Ubiquitina carboxil-terminal hidrolase isoenzima L1, mostraram uma regulação negativa em peixes que tiveram um período de recuperação. Foram identificadas alterações nas vias metabólicas, vias de sinalização, manutenção das células musculares e nos mecanismos de reparação de danos celulares. No geral, os animais que passaram por um período de recuperação mostraram uma aclimatização mais eficiente e, provavelmente, uma redução comparativa nos danos causados pelo stress térmico.

Palavras-chave: Alterações climáticas; Ondas de calor marinhas; *Pomatochistus microps*; Transcriptómica; Proteómica

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ATP	Adenosine Triphosphate.
BLASTP	Basic Local Alignment Search Tool for Proteins.
bp	Base pair.
cDNA	Complementary DNA.
DGEList	Differentially expressed genes list.
DGEs	Differentially expressed genes.
ELC	Essential light chain.
***e*-value**	Expectation value or Expect value.
FDR	False Discovery Rate.
GHG	Greenhouse Gases.
HSP	Heat Shock Proteins.
HWs	Heat Waves.
IPCC	Intergovernmental Panel on Climate Change.
LC-MS/MS	Liquid Chromatography-Mass Spectrometry/Mass Spectrometry.
log_2FC	log_2 fold change.
MCCEs	Marine Climate Change Experiments.
NCBI	National Center for Biotechnology Information.
NES	Normalized Enrichment Score.
OCLTT	Oxygen and Capacity Limited Thermal Tolerance.
ORFs	Open Reading Frames.
***p*-value**	Probability Value.

R	R Programming Language.
RCP	Representative Concentration Pathways.
RNA	Ribonucleic Acid.
RNA-seq	RNA Sequencing.
RIN	RNA Integrity Number.
SST	Sea surface temperature.
Swiss-Prot	"UniProtKB/Swiss-Prot" (reviewed, manually annotated).
UniProt	Universal Protein Resource database

1

INTRODUCTION

1.1 Global warming and marine heatwaves in the ocean

Analysing and understanding how certain climatic events affect life on Earth is one of the most pressing topics today. Evidence shows that the increase in global warming is related to human activity. One of the main reasons is the emission of greenhouse gases (GHG), which is related to society's consumption patterns and lifestyles (IPCC, 2023; Stott, Stone, & Allen, 2004; Ring et al., 2012). Greenhouse gases emissions began to grow after the industrial revolution. Although emission values remained relatively low, it was possible to observe an increase in the 1950s, with emissions of around 6 billion tons of CO_2. These emissions quadrupled in the 1990s, with an emission of 20 t of CO_2 being recorded. In the last years, GHG emissions skyrocketed, with an average of around 35 t of CO_2 currently being emitted per year (Ritchie, Roser, & Rosado, 2020). These values make emission levels and their consequences worrying in the coming years. There are several risks caused by the increase in global warming, such as the loss of hundreds of species and mortality events on land and in the ocean, desertification, soil degradation, impacts on ecosystems, caused, for example, by acidification of the ocean, sea level rise and regional precipitation decrease (IPCC, 2023). It is also known that 91% of warming in the climate system is due to warming of the oceans, and that land warming, ice loss and atmospheric warming represent only 5%, 3% and 1%, respectively (IPCC, 2023).

An example of climate-change triggered events is the surge in marine heat waves, which are characterized by anomalous, discrete, and prolonged changes of the water temperature rise (Oliver et al., 2021; Hobday et al., 2016). These events can extend over several kilometres and be of long or short duration (Oliver et al., 2021), as they can be caused by several local oceanic and atmospheric processes and climatic phenomena (Holbrook et al., 2019). One of the most recent marine heat waves, which began between February and March 2023 in the Gulf of Mexico and the Caribbean Sea, caused an increase of 1 to 3°C above normal, increasing the risk of disruption to more sensitive marine ecosystems (NOAA, 2023). Between 2015 and 2016, an El Niño, which is an anomalous warming of surface waters in the central-eastern sector of the Pacific Ocean (Yeh et al., 2009), occurred in the North Pacific, primarily affecting California, and causing an increase in water temperature of ~2.3°C (Jacox et al., 2016).

One of the biggest problems we face today is the fact that marine heatwaves have become and will continue to become more frequent (Figure 1.1) and longer-lasting, as researches predicted using climate modelling scenarios such as RCP8.5 and RCP4.5 (Mbokodo et al., 2020). An example is, the 34% increase in the frequency with which they occur and a 17% increase in their duration between 1925 and 2016 (Oliver et al., 2018).

A B

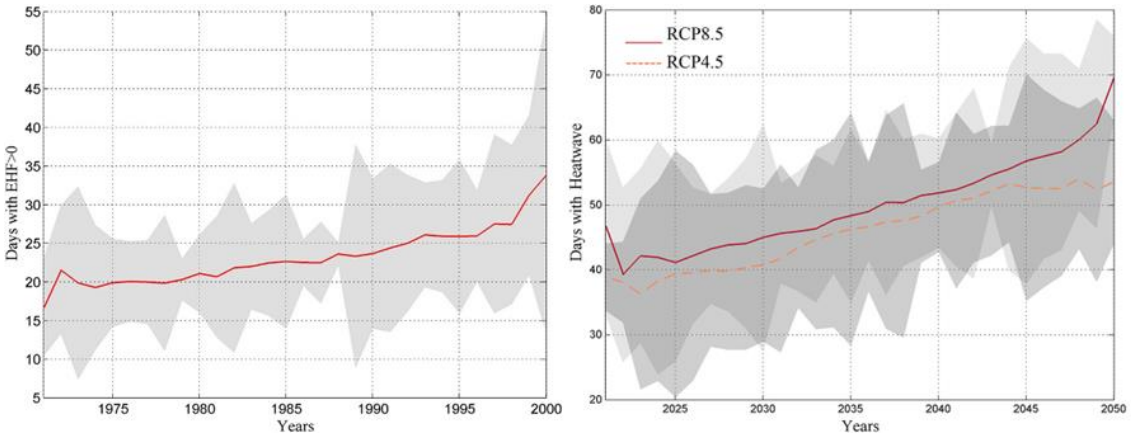

Figure 1.1 Number of heat waves per year. A) Time series of annual days with Excess Heat Factor (EHF) > 0 (a measure of heatwave intensity) on the Iberian Peninsula from 1971 to 2000. B) Predicted time series of annual heatwave days on the Iberian Peninsula from 2021 to 2050. Adapted from Lorenzo et al., 2021.

1.2 A tale of hot climate and seasonal extremes: the Iberian Peninsula

The Iberian Peninsula and Mediterranean Sea will expectedly become one of the most affected regions by the end of the century, where marine heatwaves have become stronger and more intense, in response to the increase of greenhouse gas emissions (Darmaraki et al., 2019). Sea surface temperature (SST) has increased along the Portuguese coast (Figure 1.2) (Biguino et al., 2023; Baptista et al., 2018) during the last decades (1980-2010), more intensely during spring and summer, and non-uniformly along the coast (Baptista et al., 2018). Values such as the maximum annual sea surface temperature were measured and observed during the last decades, and an increase between 0.025ºC and 0.050ºC per year was recorded, coupled with an average increase in SST along the Portuguese coast between 0.010ºC and 0.025ºC. Likewise, the number of days per year in which SST is above average has also increased over recent years (Biguino et al., 2023).

A B

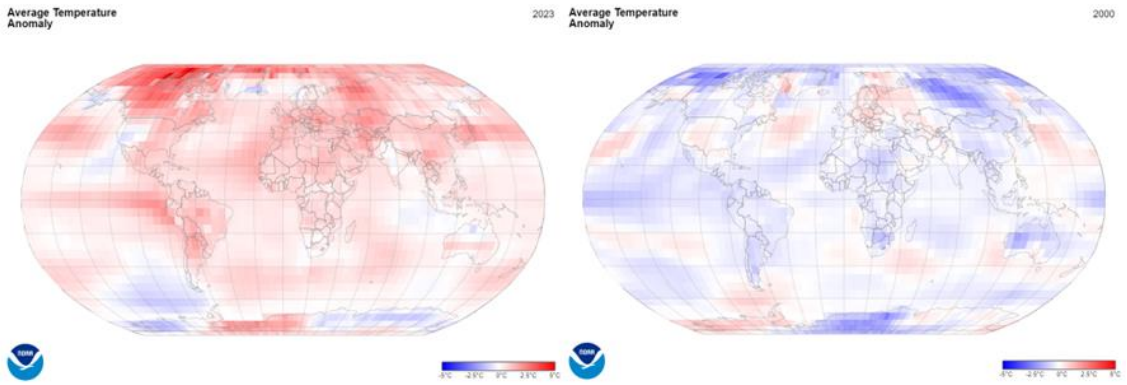

Figure 1.2 Global average temperature anomaly. A) Global annual average temperature anomaly of 2023 B) Global annual average temperature anomaly of 2000. Adapted from NOAA National Centers for Environmental information, Climate at a Glance: Global Mapping, published March 2024, retrieved on April 4, 2024 from https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/mapping

With the observation of the increase in SST and the frequency in which this increase is above average, it is important to study and try to predict the possible scenarios that could occur in the Iberian Peninsula. Regarding the frequency and duration of heat waves, they are expected to increase by an average of 9 HWs every 30 years, with around 30 to 65 days of heat waves expected per year between 2021-2050, with each event projected to be between 6% and 17% longer, compared to those observed during the period from 1971 to 2000. (Parente et al., 2018; Lorenzo et al., 2021). As for SST, it is expected that there will be an increase of 1ºC to 2ºC during the next decades (Baptista et al., 2018). With the increase and duration of marine heatwaves, there will also be an increase in the impacts they cause on different marine species (Oliver et al., 2018), such as the vulnerable seagrass *Posidonia oceanica* which has the morphology of its shoots affected by rising temperatures (Stipcich et al., 2022). In general, the impacts caused by these can lead to problems in the growth, reproduction, metabolism, physiology, and behaviour of fish (Islam et al., 2022), as well as leading to the extinction of habitat-forming species, such as algae and corals, altering the geographical distribution of some species, modifying the structure of communities and ecosystems, and obviously, they can cause economic impacts (Oliver et al., 2018).

1.3 Thermal biology of marine organisms in highly fluctuating environments

Survival of organisms in highly climate variable environments of the Iberian Peninsula, such as shallow ocean areas, including sea-air interface habitats, such as intertidal sandy beaches or rocky reffs, is directly related to the animals' homeostatic control capacity, that is, their ability to maintain a viable state despite recurring changes in their environment (Torday, 2015). To maintain a viable state, animals need a set of processes or mechanisms responsible for maintaining their internal, physical, chemical, and social conditions, called homeostats (Bernhardt et al., 2020). Oxygen and capacity limited thermal tolerance (OCLTT) is a key concept that helps us understand these processes, or how fish respond to climate change (Pörtner, Bock, & Mark, 2017). It is based on the idea that when temperatures approach extreme values, oxygen supply systems, such as cardiovascular circulation, are limited according to the species and life stage, leading to a decrease in fish performance (Pörtner, 2012; Pörtner, Bock, & Mark, 2017). The OCLTT can be used to identify adaptation mechanisms to changes in temperature (Bozinovic & Pörtner, 2015), and in this way, it is a concept used to explain the ecological processes behind climate responses under extreme temperature conditions (Pörtner, 2012). An example of where the OCLTT concept is used is the study of changes in the biogeography of some species. Due to an increase in temperatures, some marine animals' resort to a change in their biogeography to survive these changes (Perry et al., 2005), and their performance decreases if exposed to high temperatures, as is the case of migratory salmon that has a decline in exercise performance, which reflects the existence of a limit in their cardiocirculatory capacity to provide sufficient oxygen (Bozinovic & Pörtner, 2015). In turn, it is also known that fish that live in high latitudes and low latitudes tend to have a smaller tolerance to changes in temperature, while fish that live in medium latitudes tend to have a greater tolerance to changes in temperature, as seasonal temperature changes are typically greater (Pörtner & Peck, 2010). Despite tolerance to these seasonal changes, they also have their impacts on animals, for instance, during winter, low temperatures tend to increase mortality, especially in smaller fish with lower energy reserves and relatively higher metabolic rates (Post & Evans, 1989).

1.4 Experimental biology and multi-omics approaches to mechanistic physiology

Despite the increase in studies related to the impacts of climate change on ecosystems (Bass et al., 2021; Hanson & Walker, 2020), more studies will be necessary to complement knowledge about the mechanisms used by animals to adapt to such changes. Adopting methods that enable analysing thousands of molecules in single runs, i.e. omics, can enable the discovery, quantification and modelling of novel molecular pathways triggered or blocked by extreme weather events. In the past decade, a large increase in the number of papers related to Marine Climate Change Experiments (MCCEs) was observed, compared to the first decade of the millennium, recording 110 papers and 854 papers, respectively (Wernberg, Smale, & Thomsen, 2012; Bass et al., 2021). Although most studies between 2010 and 2019 remain single factor (mostly ocean acidification and temperature), it was possible to observe an increase in multifactorial experiments (mostly combinations between ocean acidification, temperature, and others), with 275 papers published, which corresponds to a greater number than the total number of papers published between 2000-2009 (Bass et al., 2021). Likewise, it is possible to observe an increase in the number of studies on polar and tropical species, however the study of species from temperate regions continues to be more prevalent (Bass et al., 2021). This type of studies becomes very important for understanding the physiological responses and their respective molecular mechanisms of marine animals to climate change.

Many of these studies have implemented the use of multi-omics techniques, such as transcriptomics and proteomics methods. In general, these techniques allow for greater ease in understanding cellular behaviour, which in turn helps in the study of various biological and physiological processes (Graw et al., 2021). Typically, these studies aim to understand metabolism (Esmaeili et al., 2022; Ferrari et al., 2022; Miao et al., 2021), reproduction and growth (Wang & Liu, 2022), mechanisms behind some diseases (Frederick et al., 2022; Fu et al., 2022; Zhan et al., 2022), differences and evolution between species (Campo et al., 2022), among others. With the increase in studies related to MCCEs, the use of these tools has also increased, and in turn, the idea of understanding the consequences caused by an increase in water temperature becomes increasingly clear. One of the main ideas behind this type of study is the study of differentially expressed genes and proteins (Mutz et al., 2013; Dutt & Lee, 2000). Some of these studies prove that genes that encode heat shock proteins, HSPs, (e.g. HSP7C, HSP90a.1, HSP90a.2, HSP90b1 and HSP60) experience an increase in their expression when exposed to high temperatures (Table 1.1) (Clark et al., 2017; Shi et al., 2019; Hassan et al., 2017; Long et al., 2012; Gandar et al., 2017). Likewise, genes related to coding for glucocorticoid receptors and mineralocorticoid receptors, nr3c1 and nr3c2, tend to have an increase in their expression when there is an increase in temperature (Table 1) (Goikoetxea et al., 2021), due to their contribution to the stress response (Sathiyaa & Vijayan, 2003; Faught & Vijayan, 2018). Thus, the use of an integrative analysis of transcriptomics and proteomics helps to understand

which are the main molecular networks that are altered, that is, to acquire greater knowledge about physiological mechanisms (Nie et al., 2007).

Table 1.1 Omics analysis of thermal stress response in fish in different studies.

Species	Tested scenarios	Omics used	Altered genes/proteins	Reference
Dicentrarchus labrax	Short- (4 days) and long-term (4 months) exposure to warm temperature	Transcriptomics	nr3c1; nr3c2	Goikoetxea et al., 2021
Salmo salar	Exposure to 23ºC for 6 or 24h vs. Normal temperature for 13h	Transcriptomics	HSP7C; HSP90A	Shi et al., 2019
Danio rerio	Exposure to heat (34ºC) for 2 and 48h	Transcriptomics	HSP90a.1; HSP90a.2; HSP90b1	Long et al., 2012
Carassius auratus	Exposure to 22ºC and 32ºC for 15 days	Proteomics	HSP60	Gandar et al., 2017

1.5 Gobiids as models in marine research

The fish species used in this study was *Pomatochistus microps* (Krøyer, 1838), which is part of the largest family of marine fish, the gobies. Currently, between 10 and 20 new species are discovered every year, with 212 genera and 1875 species recognised today, *Pomatochistus microps* being one of them. The fact that this family is so large makes it very little known, increasing the number of studies dedicated to this family (Jonna, 2020). Like other gobies in the same genus, *Pomatochistus microps* are usually found in estuaries or coastal lagoons, between Norway and Morocco and in the Mediterranean Sea (Bouchereau et al., 1993), and have a diet based on mysids, shrimps, polychaetes, and foraminifera (Salgado et al., 2004). They are small fish, reaching a maximum of 64 mm, with a short life expectancy, with only a few individuals managing to survive the second winter in their lives (Jones & Miller, 1966). Typically, males are larger than females, and their growth slows down during the winter (Bouchereau & Gueleorget, 1998). Because they have a short life expectancy, reaching a maximum of 22 months of life (Bouchereau & Gueleorget, 1998; Miller, 1975), a short generation time, with reproduction during the spring or summer after the first winter (John & Miller, 1966), being easy to capture and maintain in captivity, these features make them a good model for this study.

1.6 Thesis objectives and hypotheses

The main objective of this thesis is to study the impacts of marine heatwaves on the transcriptomic and proteomic profiles of Gobiidae fish (*Pomatochistus microps*). Our aim is to assemble new transcriptomes, analyze proteomes, and calculate the relative abundance of transcripts and proteins, as well as identify differentially expressed genes and proteins caused by temperature changes, and identify altered molecular networks. We focus on quantifying and comparing the transcriptomes and proteomes of fish subjected to long marine heatwaves and fish subjected to two consecutive marine heatwaves with a recovery period in between. We aim to understand and discover new key mechanisms in the stress response and the importance of recovery time in adaptation. Studying and understanding these changes will also be important for studying the evolution/adaptation of fish to extreme thermal stress.

Little is known about the impacts caused by fluctuating temperatures on fish, where most studies were carried out based on the impacts of either gradual average temperature changes or impacts of maximum and minimum environmental temperatures. Therefore, this study aims to understand the impacts of long marine heatwaves vs. sequential marine heatwaves in *Pomatochistus microp*. The impact caused by temperature changes is expected to be closely related to the amount of stress experienced, including growth problems, alterations in population success and changes in physiological condition (Thomas et al. 1986; Jain, Sharma, & Mathur, 2013; Mugwanya et al. 2022). The amount of stress is largely related to the history of exposure to different temperatures, as is the case with exposure to different seasonal temperatures, making fishes more tolerant to summer and winter temperatures (Bevelhimer & Bennett, 2000). However, it is known that the benefits of acclimatization are exceeded when exposure to high temperatures is prolonged, causing an increase in stress and the fish's tolerance to high temperatures, which may affect the success of the population (Bevelhimer & Bennett, 2000). Regarding the impacts on growth and physiological condition and performance, according to some studies, fish exposed to high temperatures for a long period of time tend to have a reduction in their growth and a worse physiological condition, compared to fish exposed to more constant changes in temperature (Hokanson et al. 1977).

1.7 Relevance of this thesis to science and society

As we all know, we all depend on nature to survive. Food, air, water, energy, and materials from some ecosystems are factors that make life possible, and to do so it is necessary to deal with climate change and maintain a healthy world. The fight against the destruction of ecosystems has gained a lot of relevance in recent years. Currently, 81% of habitats are in poor condition, with the EU's objective being to protect 30% of sea and land (https://environment.ec.europa.eu/topics/nature-and-biodiversity_en), until 2030. To this end, scientific advances must be made with the aim of understanding the impacts caused by climate change, to identify different ways to protect species and ecosystems. The study of proteins and

genes is relevant to understanding the evolution/adaptation of species over time, especially when it comes to their adaptation to climate change (Frankham, 2005). Bioinformatics, or computational biology as it is called by many scientists today, is a crucial tool in the study of these biological molecules, as it unites biology, computational science, and technological information, to reduce time consumption, cost, and laboratory practice, in the context of discovering new biological knowledge, such as understanding the basic principles of biological processes and pathways (Kumar & Chordia, 2017). Here, we use it to study the expression level of messenger RNA molecules and proteins to understand how these molecules are affected by marine heatwaves in *Pomatochistus microps*.

2

MATERIALS AND METHODS

2.1 Animals and temperature assay

The methodology described in this section was performed by collaborators from the MBA (Marine Biotechnology and Aquaculture Research Group), led by Dr. Diana Madeira at University of Aveiro.

Specimens of *P. microps (*n = 400) were collected in Praia da Biarritz, in Ria de Aveiro, Aveiro, Portugal (Figure 2.1) on the 7th of March of 2023, using small hand-operated trawling fishing nets. The fish were transported to CEPAM-ECOMARE (Universidade de Aveiro, Aveiro, Portugal), where they were acclimatized to laboratory conditions for 41 days. Water temperature was kept at 21 ºC with EHEIM thermocontrol® 100Watt heaters, salinity at 33, pH at 7.9 and tanks were equipped with airstones for continuous oxygen supply and maintained at photoperiod cycle of 14h:10h. Fish were fed with mussels every day at 3% of their average weight. Ammonia and nitrite levels were measure twice weekly with commercial colorimetric test kits.

After acclimation, the fish were divided into three groups randomly distributed throughout 15 experimental tanks (n = 25 individuals per tank). Briefly, the three groups consist of the following treatments i) a control group (CTL) where the water temperature was 21ºC for 25 days, simulating summer situations in shallow estuarine water (www.seatemperature.org) (CTL), ii) a group under a marine heat wave treatment (MHW1), with an increase in water temperature of 26ºC for 25 days and iii) a group under a treatment simulating consecutive marine heat waves (MHW2), with an increase in water temperature of 26ºC for 10 days, followed by a recovery time of 5 days with the control water temperature (Recovery, 21ºC), and finally another 10 days with the water temperature at 26ºC. Each treatment group had 5 replicate tanks.

Over the 25 days of the experiment, the fish were sampled twice, on day 10 and on the last day, for molecular analysis (n = 5 fish per treatment, one from each replicate tank; the remaining fish were sampled for purposes outside the scope of this thesis). Prior to sampling, the fish were fasted for 24 hours. To have control over the physical parameters of the fish, they were all measured (total length, mm) and weighed (wet weight, g) using a ruler and a scale.

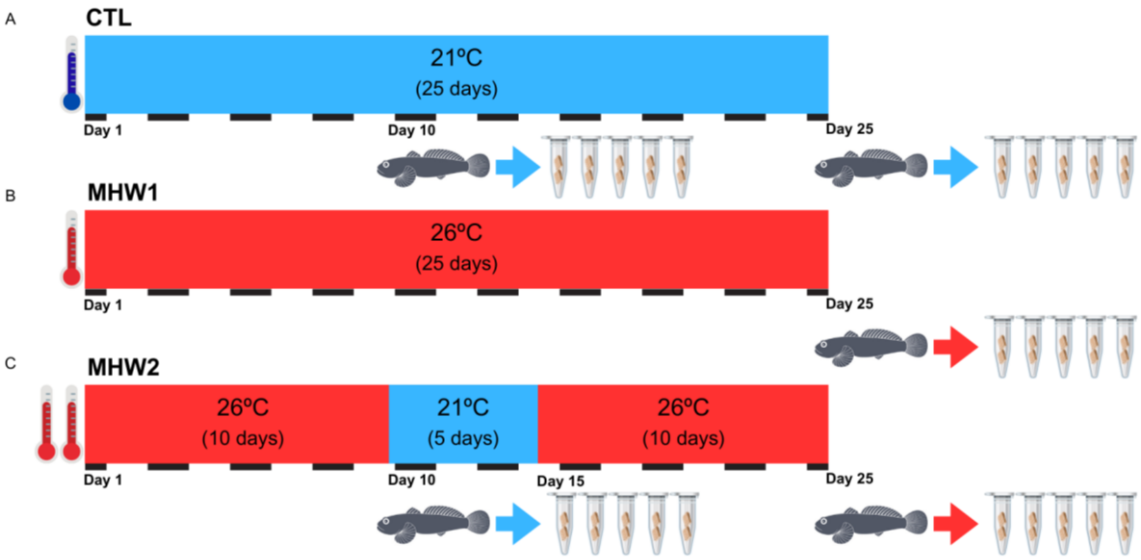

Figure 2.1 Schematic representation of the experimental design illustrating the temperature conditions and sampling points for each group. A) Control group (CTL) maintained at 21°C for 25 days. Samples were collected on day 10 and day 25. B) MHW1 group exposed to 26°C for 25 days. Samples were collected on day 25. C) MHW2 group exposed to 26°C for the first 10 days, followed by 5 days at 21°C, and an additional 10 days at 26°C. Samples were collected on day 10 and day 25.

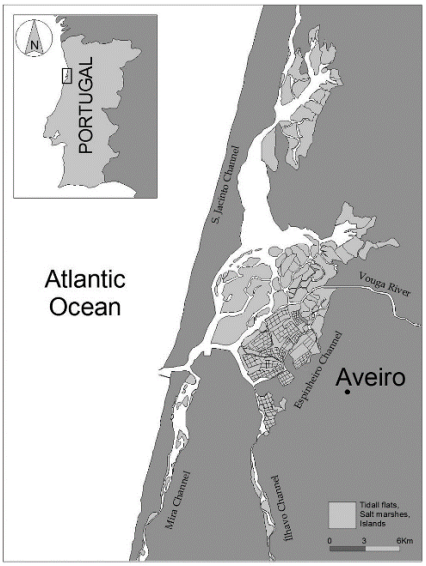

Figure 2.2 Map of the sampling area. *P. microps* were captured in Ria de Aveiro, Aveiro, Portugal. Adapted from Maia et al., 2021.

2.2 RNA extraction

The methodology described from this section onward was performed by the student wiwithin the scope of this thesis, at SeaTox Lab, in NOVA University of Lisbon.

RNA extraction (n = 3 muscle samples per treatment per timepoint n_{total} = 15), was performed on frozen muscle samples pre-homogenised in liquid nitrogen using the RNeasy Mini Kit (Qiagen, Hilden, Germany), following manufacturer's instructions available from https://www.qiagen.com/us/resources/download.aspx?id=f646813a-efbb-4672-9ae3-e665b3045b2b&lang=en). All samples contained between 15-30 mg as suggested by the manufacturer. Small optimizations in the protocol were necessary to increase the amount of RNA, namely the use of 10 µl of proteinase K (20mg/ml), and heating of RNase-free water to 60ºC. Preliminary quantification and quality assessment of the extracted RNAs were performed using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). All RNAs were high quality; RNA concentrations were between 160ng/µl and 400 ng/µl, and RIN higher than 8 (Table A.1 available in the appendix/annex to this thesis).

2.3 Protein extraction, identification and quantification

Frozen muscle samples (15-70mg, n = 3-4 per treatment, n_{total} = 18) were shipped in dry ice top the Clinical Proteomics Mass Spectrometry Facility / Global Proteomics and Proteogenomics Facility at SciLifeLab in Karolinska Institute (Sweden). The samples were dry tissue pulverized using a cryoPREP extraction system, lysed with a 4% SDS lysis buffer, and processed using a modified SP3 protocol for mass spectrometry (Moggridge et al., 2018). Peptides were labeled with TMT18-plex reagents and separated by isoelectric focusing on IPG strips. The extracted peptide fractions were then analyzed using a Thermo Scientific Q Exactive-HF coupled to an online 3000 RSLCnano system.

2.4 RNA sequencing and data-preprocessing

Library samples were prepared using the Stranded mRNA Library Preparation Kit. These cDNA collections were subsequently sequenced on the Illumina Novaseq platform (San Diego, California, U.S.). Throughout the sequencing procedure, pairs of reads measuring 150 base pairs each were produced, totalling approximately 40 to 50 million reads for each sample (Table A.1). To check the quality of the raw data obtained through RNA-Seq, FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used, to obtain some statistics on the data. TrimGalore v0.6.10 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was also used to discard reads with a length shorter than 20 bp and to trim the first 13 bp corresponding to the universal adapter of the Illumina.

2.5 Transcriptome assembly, quality, and quantification

The selected reads were used for de novo transcriptome assembly, to produce a reference transcriptome, using Trinity v2.14.0 (Grabherr et al., 2011; Haas et al., 2013). The quality of the transcriptome assembly was checked using various metrics such as N50 contigs, Ex90N50 and read content statistics. This evaluation used software such as Trinity, Bowtie2 v2.4.2 and Samtools v1.8 (Haas et al., 2013; Danecek et al., 2021; Langmead & Salzberg, 2012). To assess expression levels of individual transcripts, reads from all samples were aligned to the assembled transcriptome and transcript abundance was measured using Kallisto v0.50.1 (Bray et al., 2016).

2.6 Data analysis

All statistical analyses were performed using R (Ihaka & Gentleman, 1996), version 4.4.0. The full script can be found in Appendix. Differentially expressed transcripts and regulated proteins between treatments were identified using the edgeR v4.3.1 and limma v3.61.0 packages (Robinson, McCarthy, & Smyth, 2010; McCarthy, Chen, & Smyth, 2012; Ritchie et al., 2015). Prior to that, the tximport v1.33.0 and seqnir v4.2-36 packages were used to import the data into R (Charif & Lobry, 2007; Soneson, Love, & Robinson, 2015). The normalization factors were calculated using edgeR. A linear model was applied to the normalized RNA-Seq data (in a DGEList) to compare gene expression between treatments. Genes and proteins were considered differentially expressed/regulated if $|\log2FC| > 1.5$ and the FDR-adjusted p-value smaller than 0.05. Visualization was done using gplots v3.1.3.1 (Warnes et al., 2020), RColorBrewer v1.1-3 (Neuwirth, 2014), and ggplot2 v3.5.1 (Wickham, 2016). Pathway analysis was carried out using the multiGSEA v1.15.0 package (Canzler & Hackermüller, 2020).

2.7 Functional annotation

The assembled transcriptomes were examined to identify potential coding regions using TransDecoder v5.5.0 (Haas et al., 2013). Subsequently, these predicted open reading frames (ORFs) underwent functional annotation. This process involved assessing protein domains using the Pfam database (Pfam-A.hmm version 36.0, May 2024) via HMMER v3.4 and conducting homology comparisons against the Swiss-Prot database (version 29-02-2024 11:23, downloaded from https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz) using BLASTP (e-value < 1E-5) from NCBI blast+ v2.12.0+ (Camacho et al., 2009; Eddy, 2009; Mistry et al., 2021). The Proteome Discover version 2.4 including Sequest-Percolator for improved identification was used for protein identification, limited to a false discovery rate of 1%

3

RESULTS

3.1 Differentially expressed gene analysis of transcriptome

The assembled transcriptome from the muscle of *Pomatochistus microps* resulted in a total of 384,337 unique transcripts (Table 3.1). Among these, 56.3%, 216,538, were identified as potential open-reading frames (ORFs). Through homology-matching against SwissProt, 163,451 ORFs were annotated, representing 42.5% of the total transcripts. Additionally, we found a total of 232 (0.06%) differentially expressed transcripts (using a threshold of $|log2FC| > 1.5$ and an FDR-adjusted $p < 0.05$), of which 137 were considered ORFs. Overall, between all the contrasts, 0.036% of the annotated genes were classified as differentially expressed.

Table 3.1 Selection of the transcripts of interest after de novo transcriptome assembly. The table presents the number of sequentially shortlisted transcripts after each stage of analysis.

Analytical stage	Stage objective	Transcripts	Percentages
1	Transcriptome assembly	384337	100%
2	Transcripts with coding regions	216538	56.3%
3	Functional annotation	163451	42.5%
4	Differentially Expressed Transcripts	232	0.06%
5	ORFs with Differential Expression	137	0.036%

The MHW2 treatment (two sequential heatwaves with recovery time between them) showed the highest number of DEGs. When the control group (CTL) was compared with this treatment at day 25, a total of 106 differentially expressed transcripts were observed, of which 56 were classified as overexpressed and 50 as underexpressed (Figure 3.1C). In turn, when the control group was compared with the other treatments—before recovery time (MHW2 at D10) and prolonged heatwaved without recovery time (MHW1 at D25)—only 43 and 51 differentially expressed transcripts were identified, respectively (Figure 3.1E and Figure 3.1B). Large changes

between marine heatwave treatments were also observed. When we compared the fish before and after recovery time (MHW2 at D10 vs. D25), we observed 108 differentially expressed transcripts, 52 being overexpressed and 56 underexpressed. (Figure 3.1D). When the fish subjected to recovery time were compared with those that were not subjected to recovery time (MHW1 vs. MHW2 at D25), 66 differentially expressed transcripts were identified, where 39 were considered overexpressed and 27 under expressed (Figure 3.1A).

When observing only the expression levels of the ORFs, the same pattern was confirmed. The treatment with recovery time (MHW2 at D25) showed the highest number of differentially expressed ORFs. Similarly, when comparing the control group with this treatment (CTL vs. MHW2 at 25), 47 overexpressed ORFs and 36 underexpressed ORFs were identified, totaling 83 differentially expressed ORFs (Figure 3.1C). The group without recovery time (MHW1 at D25) showed 35 differentially expressed ORFs, less than half of those in the group with recovery time (Figure 3.1B). Only 13 differentially expressed ORFs were identified when compared to the control group (CTL vs MHW2 at D10) (Figure 3.1E). A similar number of ORFs with altered expression was identified between the treatments (MHW2 vs MHW1 at D25; MHW2 D10 vs D25). Between the groups with and without recovery time (MHW2 vs MHW1 at D25), 49 ORFs were identified, of which 20 were classified as overexpressed and 29 as underexpressed (Figure 3.1A). Forty-seven differentially expressed ORFs were identified between the treatments before and after recovery time (MHW2 D10 vs D25), where 17 were classified as overexpressed and 30 as underexpressed (Figure 3.1D).

Cluster analysis revealed similarity between treatments (Figure 3.2, upper horizontal dendogram). A total of 3 clusters were created based of the logFC values for each gene (Figure 3.2). The cluster with most of the contrast contained the contrasts against the group with recovery time (Figure 3.2, MHW2, orange cluster). One cluster containing only the contrast that compared the control group with the group without recovery (Figure 3.2, MHW1, grey cluster), and finally, another cluster composed by the contrast between the control group and the group pre-recovery time (Figure 3.2, sampling at D10, red cluster). It was also possible to form 7 clusters of transcripts/ORFs, based on their expression similarity in each contrast (Figure 3.2, left vertical dendogram).

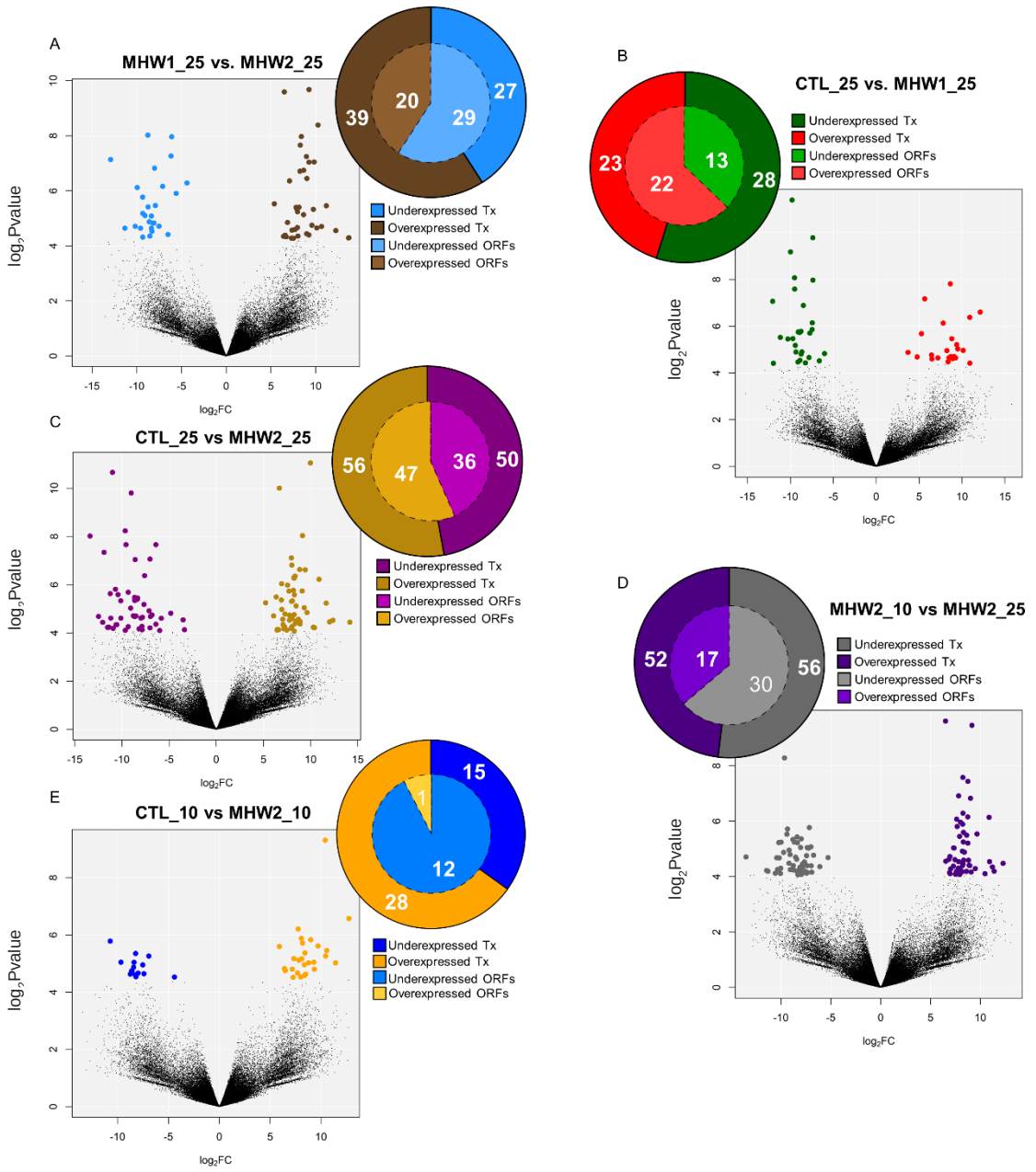

Figure 3.1 Volcano plot and pie charts illustrating differentially expressed transcripts and ORFs. A) between MHW1 group and MHW2 group day 25. B) between control group and MHW1 group day 25. C) between control group and MHW2 group day 25. D) between MHW2 on day 10 and day 25. E) between control group and MHW2 group day 10. The black dots represent transcripts that were not differentially expressed between seasons. "Tx" is the abbreviation for transcripts. Cut-off for differential expression set at | log2FC | > 1.5 and FDR-adjusted p < 0.05.

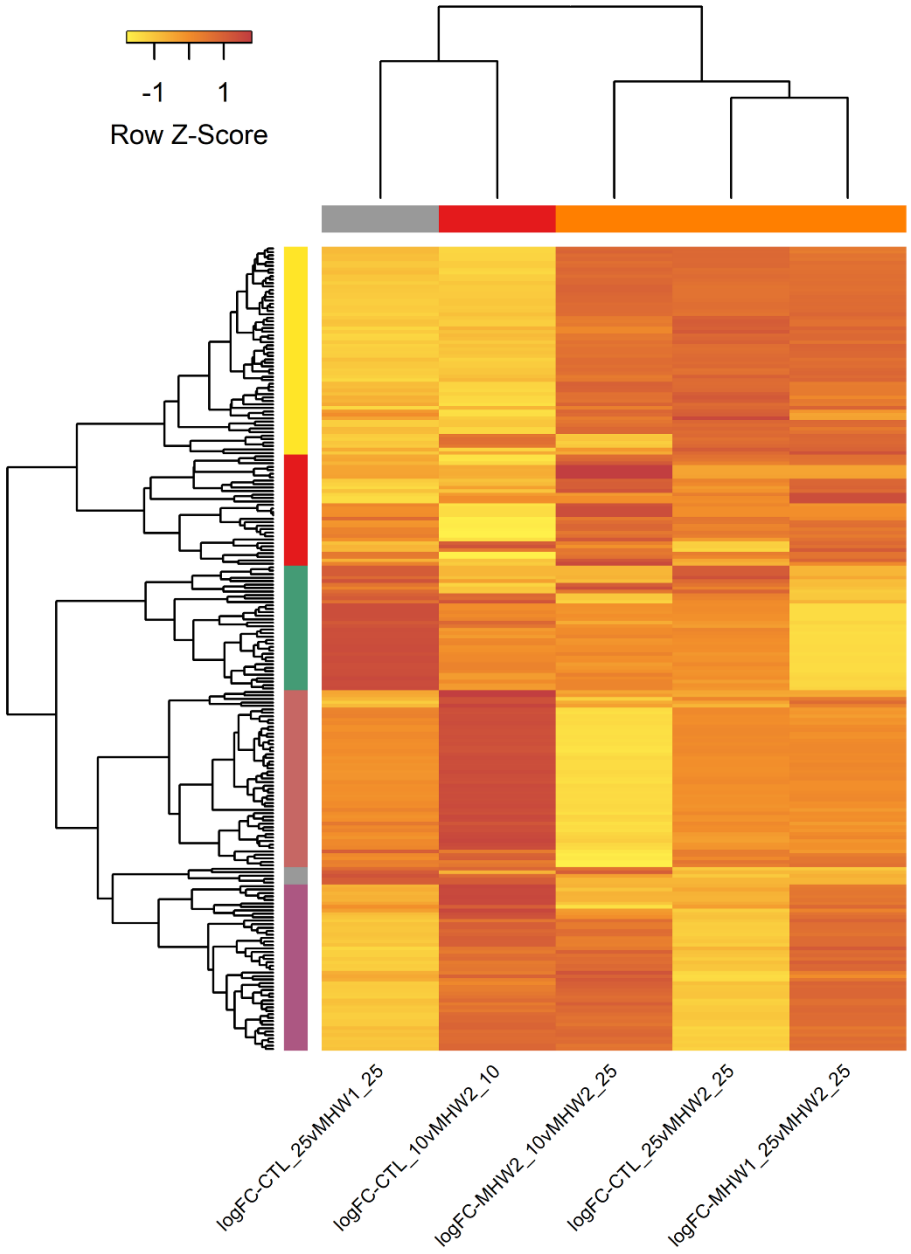

Figure 3.2 Heatmap illustrating relative gene expression of the differentially expressed genes between treatments. Cut-off for differential expression was set at | log2FC | > 1.5 and FDR-adjusted $p < 0.05$. The horizontal dendrogram shows the association between the five contrasts (control group vs. MHW1 group day 25; control group vs. MHW2 group day 25; MHW2 on day 10 vs. day 25; control group vs. MHW2 group day 10; MHW1 group vs. MHW2 group day 25), whereas the vertical dendrogram shows the association between protein-coding genes. Relative high and low expression is indicated by yellow and red, respectively, after normalization per row (gene). Dendrograms were obtained using cluster analysis taking Euclidian distances and complete linkage, as metric and amalgamation rule, respectively.

3.2 Quantitative Proteomics

In total, 7081 proteins were identified from mass spectrometry. Of the putative 7081 proteins, 960 were classified as duplicates based on UniProt accessions. According to normalized abundances, a total of 31 differentially regulated proteins were detected by setting the cut-off | log2FC | > 1.5 and FDR-adjusted $p < 0.05$ (Figure 3.3).

All comparisons between control and treatment groups, revealed proteins with significantly altered regulation (Figure 3.3). The number of down-regulated proteins was always higher than the number of up-regulated proteins. Between the control group and the group without recovery time (MHW1) at day 25, the lowest number of altered proteins was obtained, with only three proteins (Figure 3.3C and Table A.3), while the contrast between the control group and the group with recovery time (MHW2) at day 25 showed the highest number of proteins with altered regulation, with fifteen proteins (Figure 3.3B and Table A.5). Although the comparison between the control group and the group before the recovery time (MHW2) at day 10 involved animals subjected to a shorter period of thermal stress, eight proteins with altered regulation were observed (Figure 3.3A and Table A.2). Nine altered proteins were observed between the groups before and after recovery time (Figure 3.3D and Table A.4, MHW2 at D10 vs. D25). Between the group with recovery time and without recovery time (MHW1 vs. MHW2, respectively) at day 25, no proteins with altered regulation were observed according to the cut-off parameters (Figure 3.3E).

Figure 3.4 illustrates six clusters of proteins based on their regulation (vertical dendogram), and two clusters that show the differences between the contrasts (horizontal dendogram), in which one cluster contains the contrasts between the control groups and the treatments, and another, with the contrasts between treatments. It is possible to notice that the Heat shock 70kDa protein, Heat shock 70kDa protein 4 and Ubiquitin carboxyl-terminal hydrolase isozyme L1, form a cluster based on their regulation. These proteins appeared to be up regulated in all treatments when compared to the control groups. Between treatments, it was observed that these proteins were in lower abundance, or downregulated, in the fish subjected to a recovery time when compared to a continuous prolonged heatwave. There are two more protein clusters, where the regulation between the contrast clusters is similar. One cluster contains thirteen proteins, some of which are Protein-glutamine gamma-glutamyltransferase 4, Leucyl-cytinyl aminopeptidase, Solute carrier Family 2 (facilitated flucose transporter member 2), etc., and another contains five proteins, including Up-regulator of cell proliferation, Interferon-induced very large GTPase 1, Cystospin-B, Apolipoprotein M and Vimentin. In these two protein clusters, a general decrease in proteins was observed across all treatments compared to the controls. An exception was found for the proteins Apolipoprotein M, Cytospin-B, and Vimentin, which were classified as up-regulated in the group with recovery time. It was also observed that all these proteins, present in these two clusters, were in higher abundance in the fish subjected to recovery time compared to

the other treatments. In the cluster of 5 proteins, the proteins Neoverrucotoxin subunit alpha, Probable ubiquitination network signalling protein arcB, Elongation factor 1-alpha and Kelch-like protein 20, which belong to a particular cluster, tend to have their regulation very little altered between all comparisons, except in the MHW2 treatment group before recovery time, where these proteins are seemingly down-regulated, when compared to the control group.

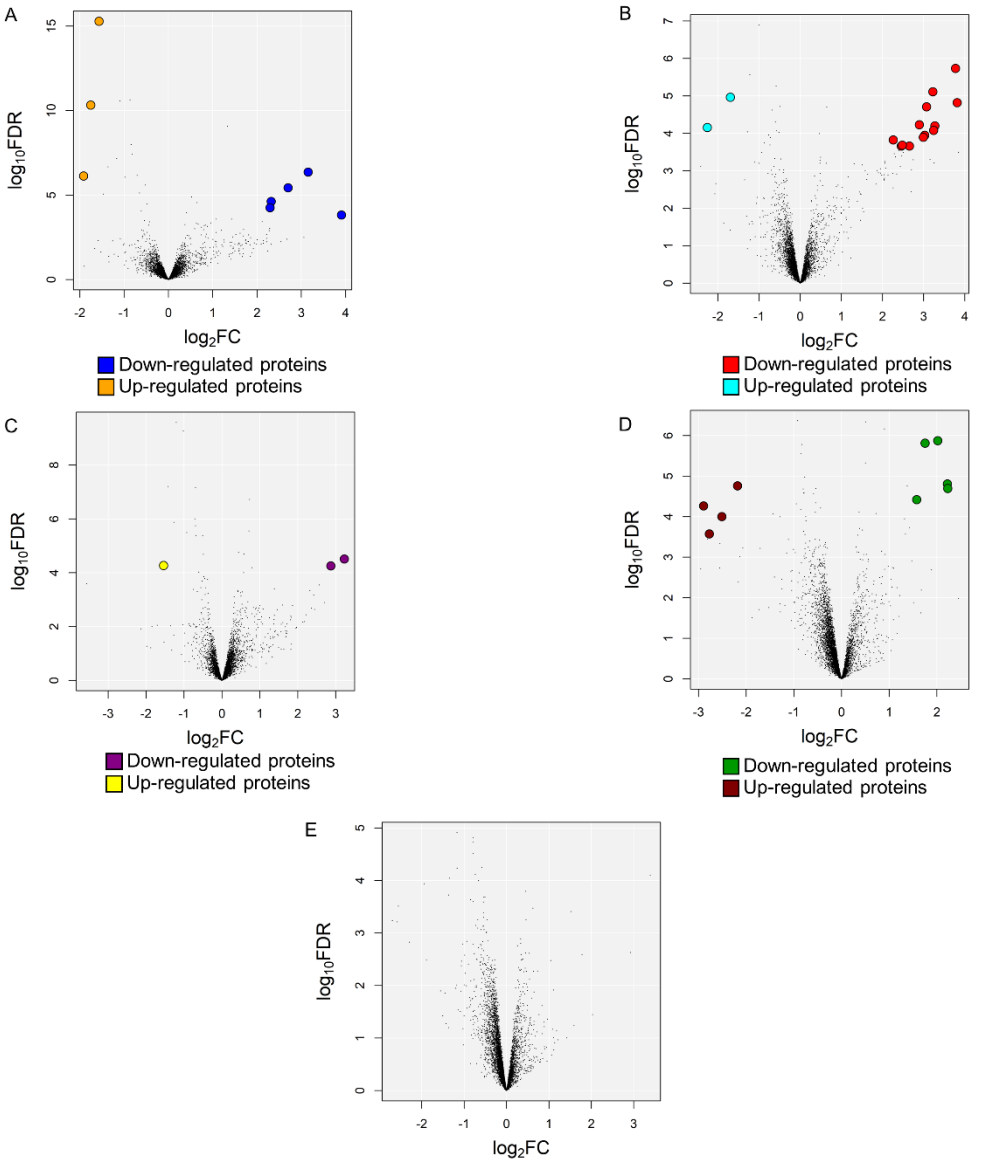

Figure 3.3 Volcano plot illustrating differentially regulated proteins. A) between control group and MHW2 group day 10. B) between control group and MHW2 group day 25. C) between control group and MHW1 group day 25. D) between MHW2 on day 10 and day 25. E) between MHW1 group and MHW2 group day 25. The black dots represent proteins that were not differentially regulated. Cut-off for differential expression set at | log2FC | > 1.5 and FDR-adjusted $p < 0.05$.

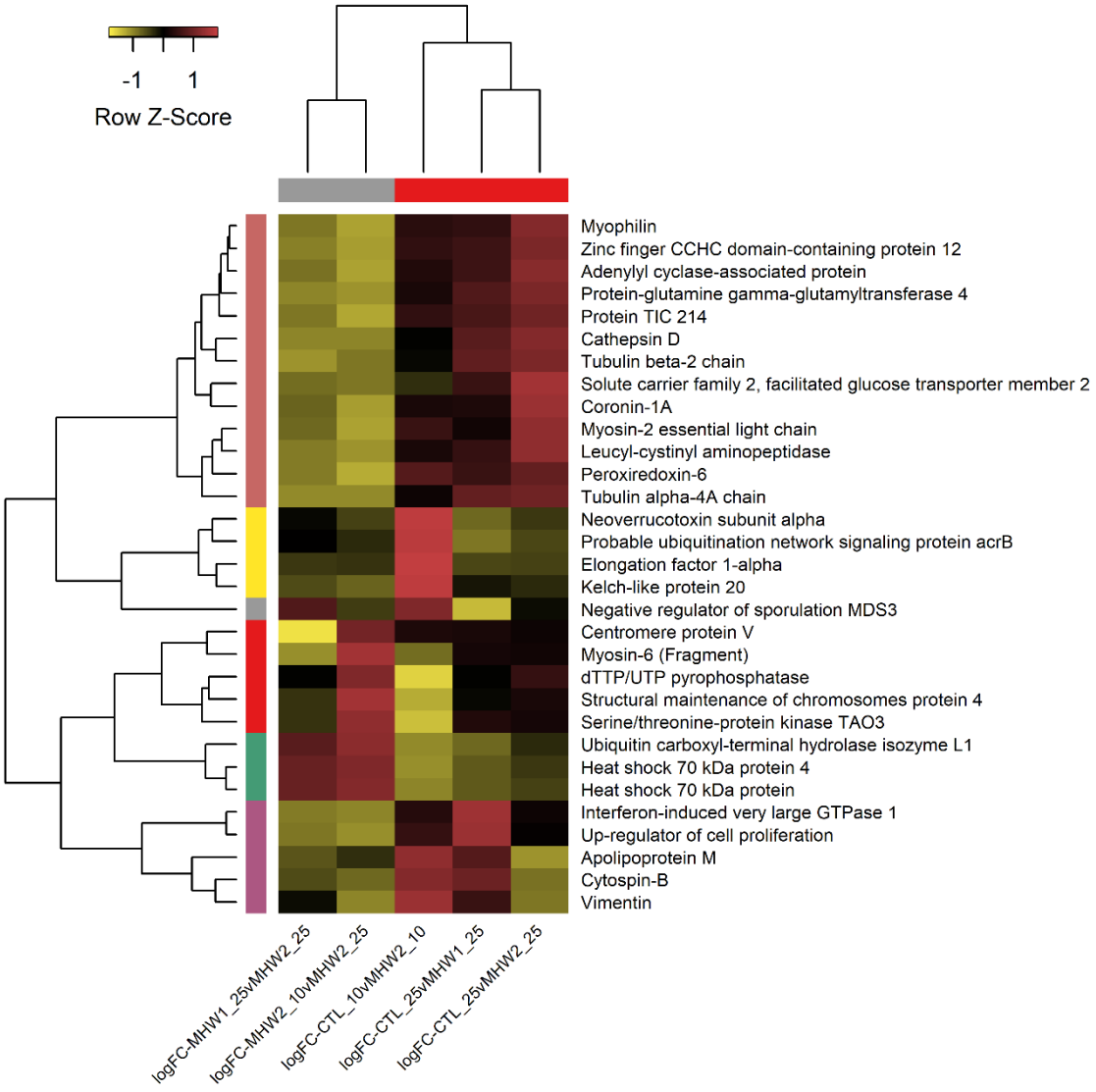

Figure 3.4 Heatmap illustrating relative protein expression of the differentially expressed proteins between treatments. Cut-off for differential expression was set at | log2FC | > 1.5 and FDR-adjusted $p < 0.05$. The horizontal dendrogram shows the association between the five contrasts (MHW1 group vs. MHW2 group day 25; MHW2 on day 10 vs. day 25; control group vs. MHW2 group day 10; control group vs. MHW2 group day 25; control group vs. MHW1 group day 25), whereas the vertical dendrogram shows the association between proteins. Up-regulated proteins are represented in yellow while down-regulated proteins are represented in red. The metric and function of the cluster analysis is Euclidean distances and complete linkage, respectively.

3.3 Altered pathways based on transcriptomic analysis

According to changes in the abundances of different genes in each group, it was possible to observe alterations in various types of pathways. In total, 125 pathways were classified as altered across all contrasts. Focusing on the most important pathways, specifically the top 10 based on adjusted p-values, it could be noted that most of them were related to metabolism, cellular signalling and/or communication, cell cycle, growth and cell death, as well as cellular interactions and some structures (Figure 3.5).

When the animals that, were subjected to a prolonged heatwave with no recovery time (MHW1), were compared with the control group, it was observed that some of the top 10 pathways are related to changes in cellular interactions and some of their structures, such as the decrease in Neuroactive ligand-receptor interaction and ECM-receptor interaction, and the increase in Cell adhesion molecules (Figure 3.5 and Table A.7). Changes were also observed in the metabolism of amino acids, carbohydrates, lipids, and complex compounds, such as the increase in Arginine and N-Glycon biosynthesis, increased metabolism of Ascorbate, aladarate, alanine, aspartate, and glutamate, decreased sulphur metabolism, and decreased Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (Figure 3.5 and Table A.7). Notable changes were observed in the mTOR signaling pathway, where deactivation was noted (Figure 3.5 and Table A.7).

Changes in metabolism were also observed between the control group and the group of fishes sampled before the recovery time (MHW2, day 10). There was an increase in amino acid metabolism, represented by increased metabolism of alanine, aspartate, glutamate, cysteine, and methionine, and a decrease in carbohydrate metabolism, represented by decreased metabolism of ascorbate and aldarate (Figure 3.5 and Table A.6). There was also a general decrease in several pathways related to cellular signalling and communication, such as decreased calcium signalling pathway, C-type lectin receptor signalling pathway, apelin signalling pathway, and adipocytokine signalling pathway (Figure 3.5 and Table A.6). Regarding influential pathways in cell cycle, growth and cell death, there were two opposing processes that increased, namely cellular senescence and cell cycle, and a decrease in apoptosis (Figure 3.5 and Table A.6).

In the group with recovery time (MHW2, day 25), when compared to the control group, several changes in metabolism were also observed. These were characterized by increased metabolism of amino sugars, nucleotide sugars, alanine, aspartate, glutamate, and 2-oxocarboxylic acid, and decreased metabolism of porphyrin (Figure 3.5 and Table A.8). There were also changes in cellular interactions/structures, represented by decreased neuroactive ligand-receptor interaction and adherens junctions (Figure 3.5 and Table A.8). There was also a decrease in mitophagy and autophagy, and in adrenergic signalling in cardiomyocytes (Figure 3.5 and Table A.8).

Additionally, the group with recovery time (MHW2, day 25) showed few changes in metabolism when compared to the group without recovery time (MHW1, day 25), although there was still an increase in amino acid and nucleotide sugar biosynthesis (Figure 3.5 and Table A.9).

It was noted that there was a decrease in some pathways related to cellular signalling, specifically the C-type lectin receptor signalling pathway and adipocytokine signalling pathway (Figure 3.5 and Table A.9). Similarly, there was a potential decrease in the activation of apoptosis-related pathways, which was accompanied by a decrease in efferocytosis (Figure 3.5 and Table A.9). In this contrast, there was a decrease in cell adhesion molecules and in autophagy (Figure 3.5 and Table A.9).

In the group with recovery time (MHW2, day 25), several changes in the metabolism of carbohydrates, sugars and amino acids were also observed, such as an increase in the metabolism of ascorbate, aldarate, amino sugar, nucleotide sugar, alanine, aspartate and glutamate, an increase in arginine biosynthesis, and the decrease in aminoacyl-tRNA biosynthesis, when compared to the same group sampled before the recovery time (MHW2, day 10, Figure 3.5 and Table A.10). When compared with the control group (CTL, day 10), there was a significant decrease in autophagy and mitophagy (Figure 3.5 and Table A.10). There were also several changes in cellular signalling and communication, with increases in the apelin signalling pathway and adipocytokine signalling pathway, and a decrease in adrenergic signalling in cardiomyocytes (Figure 3.5 and Table A.10).

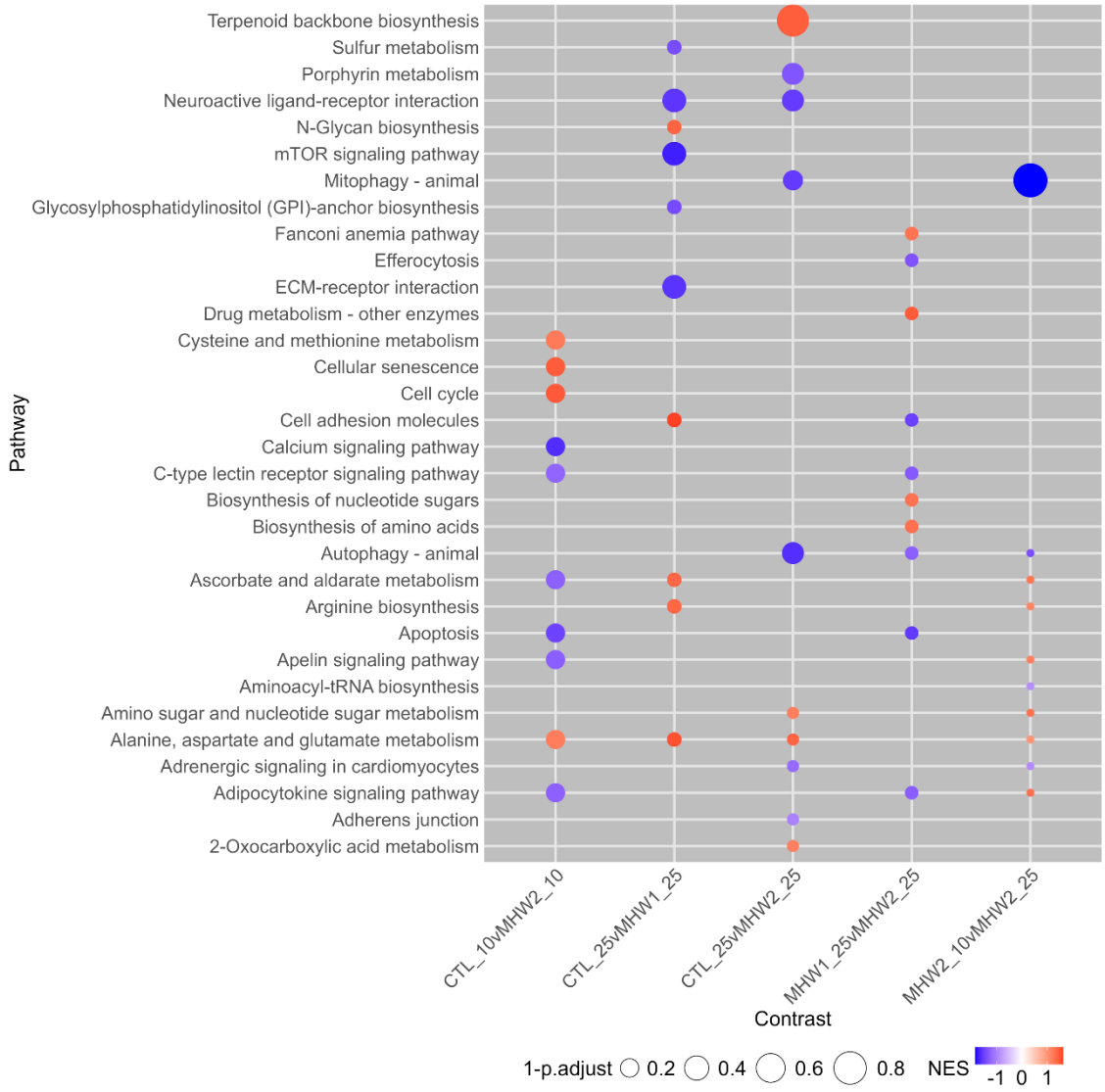

Figure 3.5 Dot plot illustrating top 10 altered pathway for each contrast based on transcriptomic data. Cut-off for significant altered pathway was set at p-adjusted < 0.05. Blue represents depleted pathways (NES < 0), and red represents enriched pathways (NES > 0). Larger dots represent lower p-adjusted and smaller dots represent higher p-adjusted.

3.4 Disrupted pathways through proteomic analysis

A total of 107 pathways were detected, using the abundance of all proteins present in the samples. The top 10 altered pathways for each treatment are represented in Figure 3.6. Most pathways were related to metabolism mechanisms, cellular signaling, immune or inflammatory signaling, biosynthesis (Aminoacyl-tRNA biosynthesis), cardiac function (Cardiac muscle contraction), cell death processes (Apoptosis), adhesion and interactions with the extracellular matrix.

Pathways related to the metabolism of types of organic acids and carbohydrates were classified as altered in animals sampled before recovery time (MHW2, day 10), when compared with the control group at day 10 (Figure 3.6 and Table A.11). It was possible to observe, through the NES value, an enrichment of the 2-Oxocarboxylic acid metabolism and Ascorbate and aldarate metabolism pathways, and a depletion of the Butanoate metabolism and Carbon metabolism (Figure 3.6 and Table A.11). In turn, an enrichment of pathways related to cellular signaling was observed, more specifically related to metabolic and energetic regulation, such as the Adipocytokine signaling pathway and Apelin signaling pathway (Figure 3.6 and Table A.11). It was also observed that there was an increase in aminoacyl-tRNA biosynthesis, cardiac muscle contraction, formation of adherens junctions and also apoptosis (Figure 3.6 and Table A.11).

A large part of the altered pathways in the group without recovery time (MHW1, day 25), when compared to the control, were also related to the metabolism of organic acids and carbohydrates (Figure 3.6 and Table A.12). It was observed that along with the changes of increased Inositol phosphate metabolism and decreased Pyruvate metabolism, there was also a decrease in the Adipocytokine signaling pathway and in MAPK signaling pathway (Figure 3.6 and Table A.12). Most of the altered pathways in this contrast are related to immune or inflammatory signaling (Figure 3.6 and Table A.12). It was notable that there was a depletion of all pathways related to these mechanisms, more specifically, the C-type lectin receptor signaling pathway, the Cytosolic DNA-sensing pathway, the RIG-I-like receptor signaling pathway, and the Toll-like receptor signaling pathway (Figure 3.6 and Table A.12). There were also some changes related to the depletion of calcium signaling and *Salmonella* infection (Figure 3.6 and Table A.12).

When comparing the control group with the group with recovery time (MHW2, day 25), the changes are related to general metabolism, such as lipid metabolism, exemplified by the increase in Arachidonic acid metabolism, decreased Amino sugar and nucleotide sugar metabolism, decreased Alanine, aspartate and glutamate metabolism, and organic acid and carbohydrate metabolism, such as the increase in 2-Oxocarboxylic acid metabolism (Figure 3.6 and Table A.13). Since there were various changes in general metabolism, it was also possible to observe changes in signaling pathways related to metabolic regulation, such as the increase in Adipocytokine signaling pathway and the decrease in Apelin signaling pathway (Figure 3.6 and Table A.13). As well in the fish sampled before recovery time (MHW2, day 10), it was possible to

observe a slight increase in Aminoacyl-tRNA biosynthesis, an increase in Apoptosis, and a very slight increase in Adherens junction (Figure 3.6 and Table A.13). There was also an observed increase in Adrenergic signaling in cardiomyocytes (Figure 3.6 and Table A.13).

The top 10 pathways altered in the group with recovery time (MHW2, day 25), when compared to the group without recovery time (MHW1, day 25), were all enriched, with no depleted pathways detected (Figure 3.6 and Table A.14). Some examples include processes related to immune or inflammatory signaling, such as the Cytosolic DNA-sensing pathway, RIG-I-like receptor signaling pathway, and Toll-like receptor signaling pathway (Figure 3.6 and Table A.14). Similarly, there was an increase in the metabolism of organic acids and carbohydrates, specifically Pyruvate, Propionate, and 2-Oxocarboxylic acid, as well as amino acids like Cysteine and Methionine (Figure 3.6 and Table A.14). There was also an increase in various pathways related to cellular signaling, including the MAPK signaling pathway, Adrenergic signaling in cardiomyocytes, and Adipocytokine signaling pathway (Figure 3.6 and Table A.14).

Between fish sampled before and after the recovery time (MHW2 day 10 vs. day 25), it was observed that most of the altered pathways were related to metabolism (Figure 3.6 and Table A.15). According to the results, there are changes in lipid metabolism, where deactivation of Glycerophospholipid and Glycerolipid metabolism was observed, changes in amino acid metabolism, represented by an increase in the metabolism of Alanine, Aspartate, Glutamate, Cysteine, and Methionine, and changes in organic acid metabolism, such as the increase in 2-Oxocarboxylic acid metabolism (Figure 3.6 and Table A.15). In another aspect, it was also observed that there were changes in the adhesion/interaction mechanism with the extracellular matrix, as there was a decrease in Focal adhesion and ECM-receptor interaction (Figure 3.6 and Table A.15). It was still possible to note an increase in cardiac muscle contraction, as well as changes in cellular and immune signalling, such as the decrease in Adipocytokine signalling pathway and Cytosolic DNA-sensing pathway (Figure 3.6 and Table A.15).

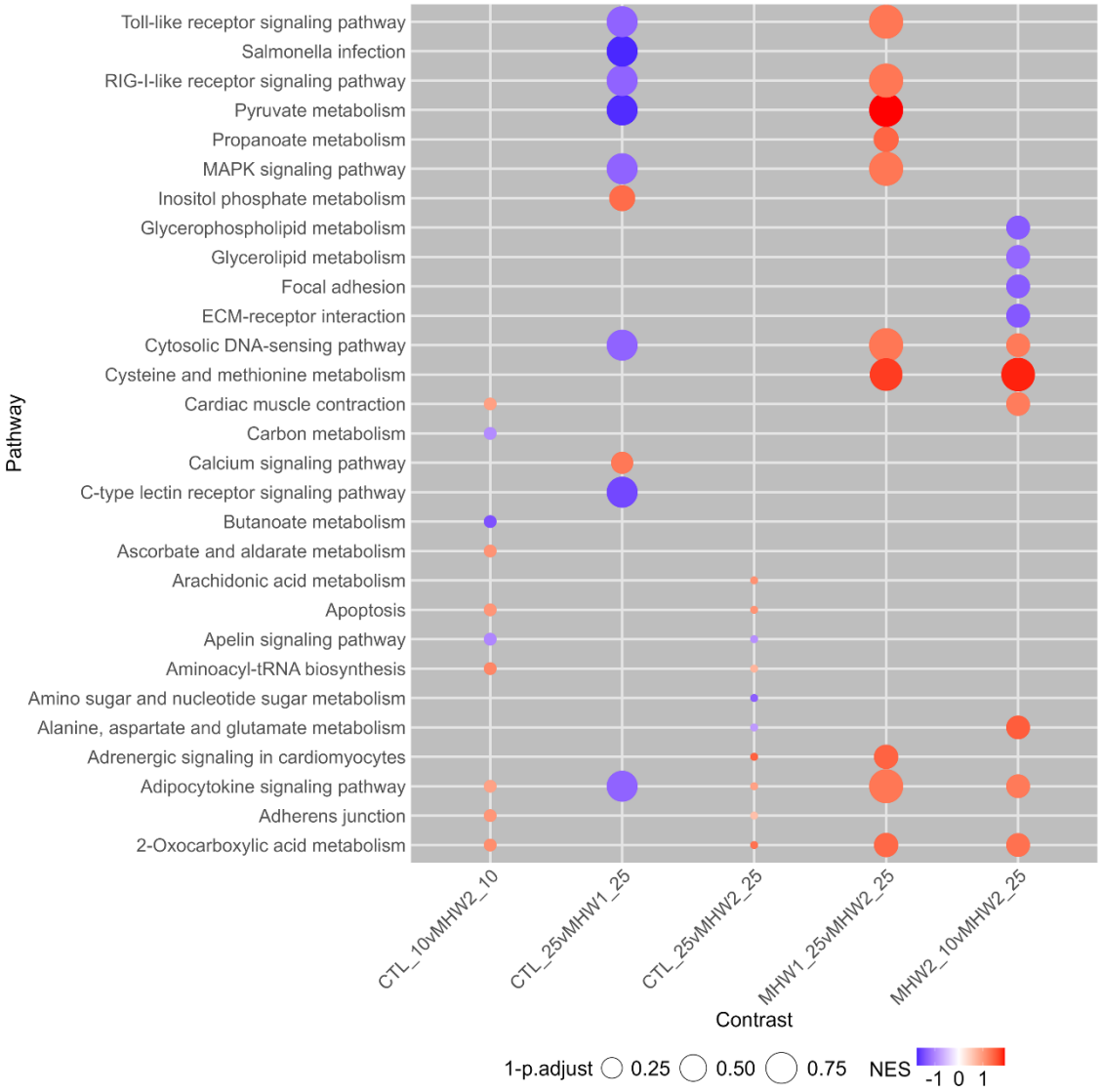

Figure 3.6 Dot plot illustrating top 10 altered pathway for each contrast based on proteomic data. Cut-off for significant altered pathway was set at p-adjusted < 0.05. Blue represents depleted pathways (NES < 0), and red represents enriched pathways (NES > 0). Larger dots represent lower p-adjusted and smaller dots represent higher p-adjusted.

3.5 Cross-validation

A total of 5913 corresponding genes and proteins were identified when crossing transcriptomics with proteomics datasets, corresponding to 96.6% of the total proteins detected. Despite this large number of common genes and proteins, only two genes and proteins were considered differentially expressed/regulated: the *hsp70* gene, which encodes the Heat shock 70 kDa protein (UniProt Accession Code: Q91233), and the *cenpv* gene, which encodes the Centromere protein V (UniProt Accession Code: Q9CXS4).

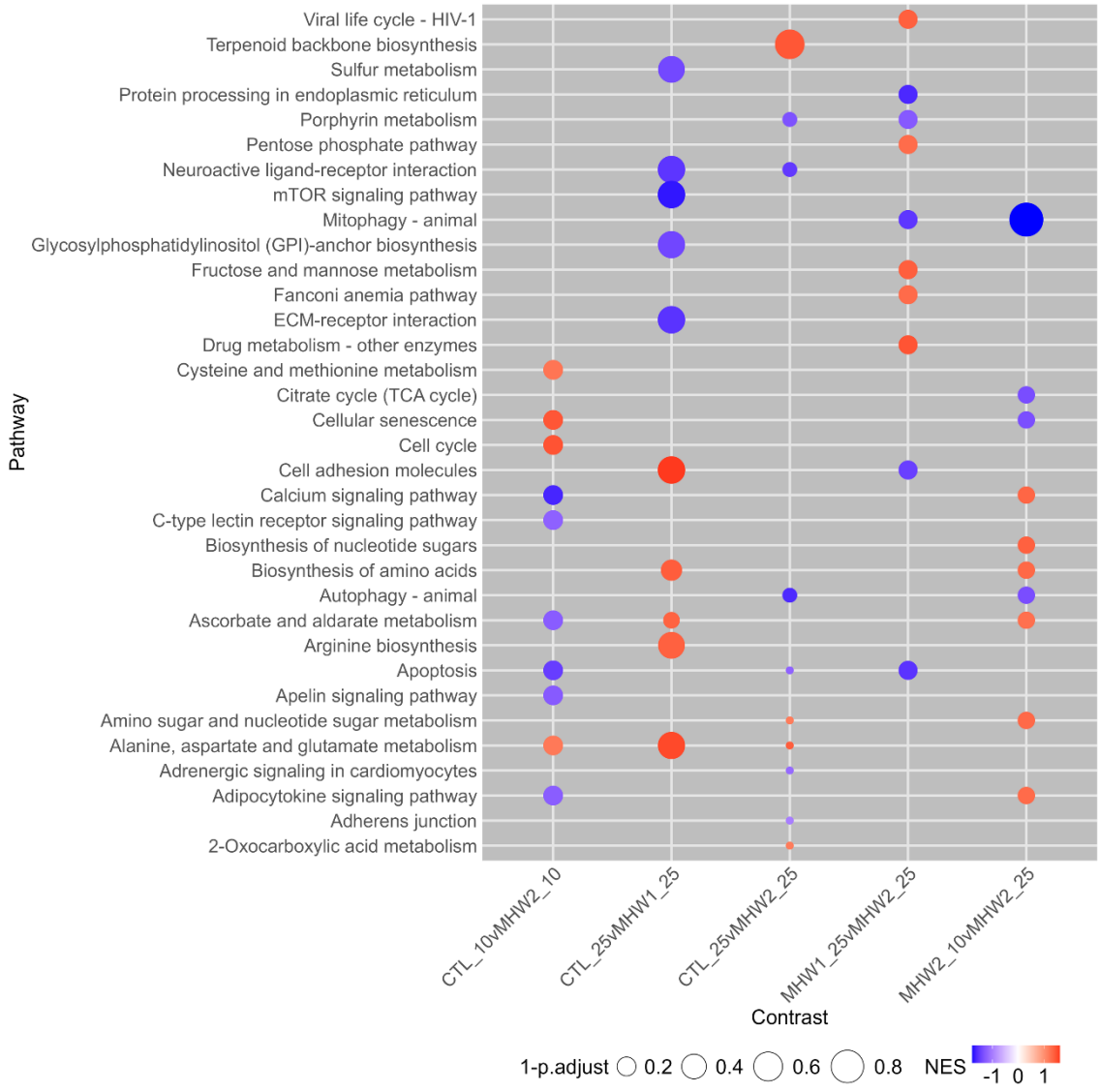

Figure 3.7 Dot plot illustrating top 10 altered pathway for each contrast based on proteomic and transcriptomic data. Cut-off for significant altered pathway was set at p-adjusted < 0.05. Blue represents depleted pathways (NES < 0), and red represents enriched pathways (NES > 0). Larger dots represent lower p-adjusted and smaller dots represent higher p-adjusted.

Nine altered pathways were confirmed by comparing the results of the two omics (Figure 3.5 and Figure 3.6) and using a merged dataset combining the data from both omics (Figure 3.7, Table A.16 and Table A.17). Most of these pathways are related to metabolism and cellular signaling. Both omics confirmed that there is an alteration in 2-oxocarboxylic acid metabolism, alanine, aspartate, and glutamate metabolism, amino sugar and nucleotide sugar metabolism, adherens junctions, and adrenergic signaling in cardiomyocytes in the treatment with recovery time (MHW2, day 25) when compared to the control. In fish sampled before the recovery time (MHW2, day 10), both omics confirmed an alteration in apoptosis, the apelin signaling pathway, the adipocytokine signaling pathway, and ascorbate and aldarate metabolism when compared to the control group. It was also possible to support the idea of several pathways' changes between other treatments. Both omics pointed to alterations in alanine, aspartate, and glutamate metabolism and the adipocytokine signaling pathway between the treatment with recovery time (MHW2) and the other treatment (MHW1).

4

DISCUSSION

Pomatochistus microps is a species that inhabits intertidal zones characterized by periodic changes between low and high tide, and consequently, changes in temperature (among other abiotic stressors). These cyclical changes are followed by several evolutionary adaptations and plastic phenotypic responses (Bang et al., 2018; Leeuwis & Gamperl, 2022), which are relevant to study in order to understand how fish respond to climate change, more precisely to marine heatwaves. Heatwaves have been changing in their duration and intensity, which in turn will cause changes in stress responses, and the molecular networks underpinning them (Oliver et al., 2018). The transcriptomic and proteomic techniques used in this study confirmed that fish respond differently to stress caused by long-duration marine heatwaves and to stress caused by two consecutive marine heatwaves with a recovery period.

The genes and proteins whose expression and regulation were altered in the different treatments were possibly related to the type and level of response to different kinds of thermal stress. Fish subjected to a heatwave, characterized by the same change in temperature for either 25 consecutive days or to only 10 days showed a similar number of differentially expressed genes and differentially regulated proteins, suggesting that fish likely induce physiological adjustments that are kept through time to counteract stress during the exposure time. However, fish subjected to two consecutive temperature changes, with 5 days of recovery between them, showed twice as many altered genes and proteins. We can relate this increase in changes to an increase in the level of response to stress, which could be potentially characterized as a greater acclimation of these fish. As with other studies in fish, the number of DEGs tends to increase during the period of heat stress, so that there is an effective response (Wang et al., 2023). It is also known that these responses tend to decrease with longer exposures to elevated temperatures, which may help to relate this recovery period with an increase in the response to marine heatwaves. The need to maintain a homeostatic balance during long periods of stress is entirely related to energy expenditure levels (Liu et al., 2023), which in turn may explain the discrepancy in DEG expression levels between treatment without recovery time and treatment with recovery time. Therefore, it can be assumed that the recovery time favours the energy levels necessary for the fish to adapt and produce necessary responses to thermal stress.

Some genes and proteins, such as heat shock protein 70 (Hsp70), play a crucial role in the response to different types of stress (Chen, Feder, & Kang, 2018; Beere & Green, 2001; Iwama, Vijayan, Forsyth, & Ackerman, 1999; Zhao & Jones, 2012). Heat shock proteins are known to have chaperonic activities, which ensure the folding, unfolding and refolding of stress-denatured proteins (Archana et al., 2017). In this study, it was observed that there was an increase in the expression of the gene encoding Hsp70, and consequently, an increase in the regulation of the protein, in all fish subjected to elevated temperatures. Several other studies support this result, showing an increase in the abundance of the gene that encodes Hsp70 and in heat shock protein 70 itself in situations of thermal stress (Rout, Kaushik, & Ramachandran, 2016; Gabriel et al., 1996). This result helps confirm that there was a clear response in all treatments. However, it is notable that the changes were not uniform across all treatments. Fish that were not subjected to recovery time between thermal challenge tended to express higher levels of Hsp70, which may support the idea of an eventual stress response. On the other hand, fish subjected to a recovery time showed a lower expression of Hsp70. Despite this decrease, when linking this result with the previously discussed results, we can assume that there was an increase in acclimation when the fish are exposed to two consecutive heatwaves, as the animals still had sufficient energy to respond to thermal changes and a reduced need to express Hsp70 to cope with stress. It is also known that previous thermal history (even a short exposure to 26 ºC for 10 days) can provide a provisional thermal memory, increasing the ability of fish to withstand future similar temperatures. These effects have also been observed in other species, for example in juvenile marbled flounder (*Pseudopleuronectes yokohamae*), where acclimation to an intermediate temperature of 22-24ºC over a period of 24 hours was sufficient to provide a buffering effect on the organisms' tolerance to higher temperatures (26ºC and 28ºC) (Sakurai et al., 2021). We can, therefore, also assume that there was a slight increase in the acclimation of fish subjected to 26ºC for 25 consecutive days, characterized by a decrease in Hsp70 compared to fish subjected to only 10 days at 26ºC, as most studies suggest that fish take between 1 to 3 weeks to fully acclimatize to new environmental conditions (Madeira et al., 2016). Although Hsp70 increases thermotolerance, other studies suggest that the expression of the protein is more influenced by variations in microclimatic differences than by large-scale climate changes, due to various limitations caused by the presence of excessive Hsp70 (Krebs & Bettencourt, 1999; Beckham et al., 2008). Thus, we may infer that the decreased Hsp70 expression in fish that benefited from a recovery period reflects adaptation to thermal challenge.

Since *Pomatochistus microps* has limited genomic annotation available, protein identification was based on homology with sequences from the species *Danio rerio*, due to its large genomic annotation. A protein called Ubiquitin Carboxyl-terminal Hydrolase L1 was identified, which plays a crucial role in protein metabolism, particularly in the ubiquitin-proteasome system (Moss et al., 2002). This protein is known for its predominant expression in neurons (Day & Thompson, 2010), but it is also expressed in skeletal muscle under certain conditions (Gao, Hartnett, & Li, 2017; Vasu et al., 2009). Due to its function, it becomes a protein with potential in the response to heat

stress. Like Heat Shock Protein 70, UCHL1 shows an increase in abundance across all treatments. This result could link the function of UCHL1 to the response to heat stress and suggest its involvement in regulating the stability and degradation of proteins affected by heat. This protein shows lower abundance in animals subjected to a recovery time. This type of regulation appears to be quite like the regulation of Heat Shock Protein 70, leading to the belief that animals evolve similarly regarding the regulation of this protein. The decreased expression of this protein may be related to the reduction in the number of ubiquitin-tagged proteins, suggesting that the occurrence of a recovery time in between exposure to stressful events reduces the amount of damage caused to cell macromolecules, namely denaturation and misfolding of proteins. These changes that are seemingly exclusive to treatment with recovery time, suggest effective acclimation, when compared with the other treatments.

The quality and state of preservation of the muscular system are entirely related to the health and performance of the organism, as most activities (e. g. foraging, social interactions, swimming) heavily rely on normal muscle functioning (Sui, Williams, Holloway-Kew, Hyde, & Pasco, 2020). As this is a study conducted on muscle samples, the identification and quantification of proteins relevant to the quality and maintenance of muscle cells help to detect the type of damage and, consequently, the type of response caused by heat stress. It has previously been shown that heat stress causes various alterations at the level of muscle development, muscle structure, and muscle function (Hao et al., 2016).

In this study, several proteins crucial for maintaining the muscular system were detected. Some of these proteins, such as Tubulin beta 2-chain, Tubulin alpha-4A chain, and Myosin-2 essential light chain, exhibit changes in regulation across all treatments. All experimental groups show a decrease in the regulation of these proteins, when compared to the control. The downregulation of tubulins may be directly related to issues with the structural stability of cells, as these are the main components of microtubules, which are key components of the cytoskeleton responsible for determining cell shape and a variety of cell movements (Cooper, 2000; Tuszynski et al., 2006). The increase in temperature may suggest negative changes in the integrity of muscle cells, as previously demonstrated in other studies (Li et al., 2013; Luo, Zhang, Wu, Sun, & Ma, 2021). Interestingly, from all the treatments, the results indicate higher abundance of these proteins in fish subjected to a recovery time between heatwaves, suggesting that fish in this treatment seemed to be less affected in terms of the structural stability of their muscle cells, which may consequently indicate that the existence of a recovery period significantly alleviates the physical damage caused by temperature. On the other hand, the downregulation of myosin-2 essential light chain (ELC) has a greater impact on motor function. The ELC is an important subunit of myosin (Lowey & Trybus, 2010). Myosin is a protein that converts chemical energy in the form of ATP into mechanical energy, often referred to as a molecular motor, playing a crucial role in muscle contraction (Cooper, 2000). Previous studies have shown that ELC-deficient myosin exhibits a reduction of more than 80% in sliding velocity compared to native myosin

(Vanburen et al., 1994). It can be proposed that heat stress increased the number of ELC-deficient myosin, consequently reducing muscular contractibility. Like the regulation levels of tubulins, this protein was detected in greater abundance in fish subjected to a recovery time, supporting the idea of an improvement in the muscular system in this group compared to the other experimental groups.

The alteration of the regulation of different types of proteins can lead to changes in various pathways. In this study, the abundance of all proteins present in the samples was used to discover which pathways were most affected. Changes were observed in cellular maintenance mechanisms, specifically in apoptosis, autophagy, mitophagy, and efferocytosis. These mechanisms are essential for maintaining homeostasis and cellular health, assisting in the control and repair of cellular damage (Boada-Romero, Martinez, Heckmann, & Green, 2020; Elmore, 2007; Zhang, 2013). Various studies discuss the interconnection of some of these mechanisms in response to cellular damage, with some being caused by heat stress (Sorice, 2022; Thorburn, 2008; Zhang et al., 2012). De-activation of all these mechanisms was observed in fish subjected to a recovery time when compared to other treatments. Changes in apoptosis and efferocytosis between treatments with and without recovery time might suggest that the recovery time is a period during which the fish recover from any cellular damage caused by increased temperature, leading to a greater depletion of these processes after the recovery time. Alternatively, the fact that they are not exposed to high temperatures for a prolonged consecutive period may lead to a reduction in the types of damage caused, which could justify a decrease in the need for response, as represented by the depletion of these biological processes. Other studies discuss that increased exposure to heat stress is related to increased cellular damage and decreased cell viability, and consequently to the induction of cellular maintenance processes (Gu et al., 2014; Li et al., 2014). Proteomic data, when analyzed in isolation, point to an increase in apoptosis, both in the case of fish subjected to a recovery time and in fish sampled before the recovery time, compared to the control group. This result may indicate that there was eventually an increase in apoptosis at an early stage when the fish were initially exposed to 10 days of marine heatwave, with potential repercussions that extended beyond the first exposure and into the recovery time and second exposure. However, transcriptomic data suggest that this increase was followed by a decrease, despite these changes in gene expression not having been detected at proteomic level in the sampled time-window. Nevertheless, these results may suggest that a process of acclimation was still underway, and that changes in gene expression level may eventually translate later into a decrease in the levels of apoptosis-related proteins. However, the only way to confirm this would have been to add an additional sampling point after the 25 days of exposure.

The analysis of both omics together, as well as separately, detected changes in metabolism across all treatments. The most affected compounds are among amino acids, carbohydrates, and organic acids. Some studies suggest that heat stress significantly reduces muscle mass, being directly linked to protein catabolism and the increased use of amino acid substrates for energy

supply (Ma et al., 2021). Certain essential amino acids detected, such as alanine, aspartate, and glutamine, are considered key substrates for gluconeogenesis (Canfield & Bradshaw, 2019), an intermediate process in energy production. Thus, it can be assumed that the rise in temperature is directly related to an increase in the metabolism of some crucial amino acids involved in energy production. This increase in energy may, in turn, be necessary to boost the levels of response to heat stress. The abundances of proteins, on their own, more clearly reveal the changes in metabolism between the treatments. Four types of pathways related to amino acid metabolism, organic acids, propanoate, and pyruvate were identified as enriched in fish subjected to a recovery time, compared to those not subjected to it. Changes in amino acid and organic acid metabolism were also observed between fish sampled before and after recovery time, except for a decrease in the metabolism of some lipids. These changes may provide some clarity regarding the levels of energy available to respond to stress, as fish subjected to a recovery time showed an increase/enrichment in certain types of metabolism necessary for energy production, such as pyruvate and some amino acids (Canfield & Bradshaw, 2019; Zangari, Petrelli, Maillot, & Martinou, 2020).

Some signaling pathways have been identified as altered under heat stress conditions. The adipocytokine signaling pathway and the apelin signaling pathway showed depletion values in the sampled animals before the recovery time, when both transcriptomic data and the integration of both omics were observed. According to *Al-Dawood (2017)*, some adipocytokines, such as Leptin and Adiponectin, either maintain or slightly decrease their concentrations in the early stages when organisms are exposed to heat stress, followed by a significant increase in concentrations. However, when analyzing proteomic data, an enrichment of the adipocytokine signaling pathway was observed in all treatments, except in animals subjected to prolonged heat stress. This result may indicate a lack of clarity in how heat stress regulates adipocytokines; however, it is clear that high temperatures induce changes in these pathways. It is known that adipocytokines regulate lipid and glucose metabolism, energy homeostasis, among others (Hueber, Asquith, McInnes, & Miller, 2010). The increase in this pathway in fish subjected to a recovery time may represent an attempt by the organism to conserve energy to adapt to the stress condition.

5

CONCLUSION

Transcriptomic and proteomic analyses of *Pomatochistus microps* under the effect of different types of marine heatwaves not only allowed for the identification of various stress responses in the fish to each heatwave but also identified several mechanisms affected by temperature changes in general. Fish exposed to a long marine heatwave, and fish exposed to a short marine heatwave – sampled before the recovery time – showed less alteration in the regulation and expression of proteins and genes compared to fish exposed to two consecutive marine heatwaves, with a recovery time. The lower number of altered genes and proteins may be related to a lower energy level, affecting the type of stress response. As this species inhabits estuaries and areas with high temperature fluctuation, its level and rate of acclimatization are higher than in other species. In fact, animals normally subjected to greater temperature fluctuations showed higher levels of altered genes and proteins, as well as changes in crucial pathways involved in heat stress response and acclimatization. The downregulation of Heat Shock Protein 70 and Ubiquitin Carboxyl-terminal Hydrolase L1, identified in fish exposed to two marine heatwaves, may indicate acclimation in these fish, as the function of these proteins was not needed to respond to the thermal challenge. The quality of the muscular system may also indicate greater acclimation in fish that underwent a recovery time. Several proteins related to the maintenance of muscle cells were identified as up regulated in this treatment, pointing to greater cell integrity. Supporting this idea, some cell maintenance mechanisms involved in cellular damage repair were also identified as altered between treatments. The depletion of apoptosis, autophagy, mitophagy, and efferocytosis in this treatment may indicate reduced damage to cellular components caused by heat stress. Since adequate energy levels are necessary to effectively respond to heat-induced stress, various important metabolisms involved in energy production were detected as altered. Fish that had a recovery time clearly showed an increase in these metabolisms, which may be related to the need to produce energy to acclimatize to stress. Overall, the data suggest greater acclimation, or at least, a reduction in heat stress-induced damage in fish that had a recovery time. The development of the pipeline, and the use of various bioinformatics techniques, not only aided in data acquisition and interpretation but were also crucial in reducing laboratory costs and time.

6

REFERENCES

Al-Dawood, A. (2017). Effect of heat stress on adipokines and some blood metabolites in goats from Jordan. Animal Science Journal, 88(2), 356-363. https://doi.org/10.1111/asj.12636

Archana, P. R., Aleena, J., Pragna, P., Vidya, M. K., Niyas, A. P. A., Bagath, M., Krishnan G., Manimaran A., Beena V., Kuriken E. K., Sejian, V., & Bhatta, R. (2017). Role of heat shock proteins in livestock adaptation to heat stress. *J. Dairy Vet. Anim. Res*, *5*(1), 00127. https://doi.org/10.15406/jdvar.2017.05.00127

Bang, C., Dagan, T., Deines, P., Dubilier, N., Duschl, W. J., Fraune, S., Hentschel U., Hirt H., Hülter N., Lachnit T., Picazo D., Pita L., Pogoreutz., Rädecker N., Saad M. M., Schmitz R. A., Schulenburg H., Voolstra C. R., Weiland-Bräuer N., Ziergler, M., & Bosch, T. C. (2018). Metaorganisms in extreme environments: do microbes play a role in organismal adaptation?. *Zoology*, *127*, 1-19. https://doi.org/10.1016/j.zool.2018.02.004

Baptista, V., Silva, P. L., Relvas, P., Teodósio, M. A., & Leitão, F. (2018). Sea surface temperature variability along the Portuguese coast since 1950. International Journal of Climatology, 38(3), 1145-1160. https://doi.org/10.1002/joc.5231

Bass, A., Wernberg, T., Thomsen, M., & Smale, D. (2021). Another decade of marine climate change experiments: Trends, progress and knowledge gaps. Frontiers in Marine Science, 8, 714462. https://doi.org/10.3389/fmars.2021.714462

Beckham, J. T., Wilmink, G. J., Mackanos, M. A., Takahashi, K., Contag, C. H., Takahashi, T., & Jansen, E. D. (2008). Role of HSP70 in cellular thermotolerance. *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery*, *40*(10), 704-715. https://doi.org/10.1002/lsm.20713

Beere, H. M., & Green, D. R. (2001). Stress management–heat shock protein-70 and the regulation of apoptosis. *Trends in cell biology*, *11*(1), 6-10. https://doi.org/10.1016/S0962-8924(00)01874-2

Bernhardt, J. R., O'Connor, M. I., Sunday, J. M., & Gonzalez, A. (2020). Life in fluctuating environments. Philosophical Transactions of the Royal Society B, 375(1814), 20190454. https://doi.org/10.1098/rstb.2019.0454

Bevelhimer, M., & Bennett, W. (2000). Assessing cumulative thermal stress in fish during chronic intermittent exposure to high temperatures. *Environmental Science & Policy*, *3*, 211-216. https://doi.org/10.1016/S1462-9011(00)00056-3

Biguino, B., Antunes, C., Lamas, L., Jenkins, L. J., Dias, J. M., Haigh, I. D., & Brito, A. C. (2023). 40 years of changes in sea surface temperature along the Western Iberian Coast. Science of The Total Environment, 888, 164193. https://doi.org/10.1016/j.scitotenv.2023.164193

Boada-Romero, E., Martinez, J., Heckmann, B. L., & Green, D. R. (2020). The clearance of dead cells by efferocytosis. *Nature Reviews Molecular Cell Biology*, 21(7), 398-414. https://doi.org/10.1038/s41580-020-0232-1

Bozinovic, F., & Pörtner, H. O. (2015). Physiological ecology meets climate change. Ecology and Evolution, 5(5), 1025-1030. ttps://doi.org/10.1002/ece3.1403

Bouchereau, J. L., Quignard, J. P., Joyeux, J. C., & Tomasini, J. A. (1993). Structure du stock des géniteurs de la population de Pomatoschistus microps (Kroyer, 1838)(Gobiidae), dans la lagune de Mauguio, France. *Cybium*, *17*(1), 3-15. https://ictiolab.wordpress.com/wp-content/uploads/2017/07/1993-cybium-bouchereau_quignard_joyeux_tomasini.pdf

Bouchereau, J. L., & Guelorget, O. (1998). Comparison of three Gobiidae (Teleostei) life history strategies over their geographical range. *Oceanologica Acta*, *21*(3), 503-517. https://doi.org/10.1016/S0399-1784(98)80034-0

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature biotechnology, 34(5), 525-527. https://doi.org/10.1038/nbt.3519

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC bioinformatics, 10, 1-9. https://doi.org/10.1186/1471-2105-10-421

Campo, A., Nitzan, T., Slosman, T., Doron-Faigenboim, A., D'Cotta, H., Baroiller, J. F., & Cnaani, A. (2022). Different transcriptomic architecture of the gill epithelia in Nile and Mozambique tilapia after salinity challenge. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 41, 100927. https://doi.org/10.1016/j.cbd.2021.100927

Canfield, C. A., & Bradshaw, P. C. (2019). Amino acids in the regulation of aging and aging-related diseases. *Translational Medicine of Aging*, *3*, 70-89. https://doi.org/10.1016/j.tma.2019.09.001

Canzler, S., & Hackermüller, J. (2020). multiGSEA: A GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinformatics*, 21, 1-13. https://doi.org/10.1186/s12859-020-03910-x

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In Bastolla, U., Porto, M., Roman, H. E., & Vendruscolo, M. (Eds.), Structural Approaches to Sequence Evolution: Molecules, Networks, Populations (pp. 207-232). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35306-5_10

Chen, B., Feder, M. E., & Kang, L. (2018). Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. *Molecular ecology*, *27*(15), 3040-3054. https://doi.org/10.1111/mec.14769

Clark, M. S., Sommer, U., Sihra, J. K., Thorne, M. A., Morley, S. A., King, M., Viant M. R., & Peck, L. S. (2017). Biodiversity in marine invertebrate responses to acute warming revealed by a comparative multi-omics approach. *Global Change Biology*, *23*(1), 318-330. https://doi.org/10.1111/gcb.13357

Cooper, G. M. (2000). *The cell: A molecular approach* (2nd ed.). Sinauer Associates. Microtubules. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK9932/

Cooper, G. M. (2000). *The cell: A molecular approach* (2nd ed.). Sinauer Associates. Actin, myosin, and cell movement. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK9961/

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham A., Keane T., McCarthy S. A., Davies R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10(2), giab008. https://doi.org/10.1093/gigascience/giab008

Darmaraki, S., Somot, S., Sevault, F., Nabat, P., Cabos Narvaez, W. D., Cavicchia, L., Djurdjevic V., Li L., Sannino G., & Sein, D. V. (2019). Future evolution of marine heatwaves in the Mediterranean Sea. *Climate Dynamics*, *53*, 1371-1392. https://doi.org/10.1007/s00382-019-04661-z

Day, I. N., & Thompson, R. J. (2010). UCHL1 (PGP 9.5): neuronal biomarker and ubiquitin system protein. *Progress in neurobiology*, *90*(3), 327-362. https://doi.org/10.1016/j.pneurobio.2009.10.020

Dutt, M. J., & Lee, K. H. (2000). Proteomic analysis. Current opinion in biotechnology, 11(2), 176-179. https://doi.org/10.1016/S0958-1669(00)00078-1

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. In Genome Informatics 2009: Genome Informatics Series Vol. 23 (pp. 205-211). https://doi.org/10.1142/9781848165632_0019

Elmore, S. (2007). Apoptosis: A review of programmed cell death. *Toxicologic Pathology*, 35(4), 495-516. https://doi.org/10.1080/01926230701320337

Esmaeili, N., Carter, C. G., Wilson, R., Walker, S. P., Miller, M. R., Bridle, A. R., & Symonds, J. E. (2022). Protein metabolism in the liver and white muscle is associated with feed efficiency in Chinook salmon (Oncorhynchus tshawytscha) reared in seawater: Evidence from proteomic analysis. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 42, 100994. https://doi.org/10.1016/j.cbd.2022.100994

Faghmous, J. H., & Kumar, V. (2014). A big data guide to understanding climate change: The case for theory-guided data science. *Big data*, *2*(3), 155-163. https://doi.org/10.1089/big.2014.0026

Faught, E., & Vijayan, M. M. (2018). The mineralocorticoid receptor is essential for stress axis regulation in zebrafish larvae. *Scientific Reports*, *8*(1), 18081. https://doi.org/10.1038/s41598-018-36681-w

Frankham, R. (2005). Stress and adaptation in conservation genetics. *Journal of evolutionary biology*, *18*(4), 750-755. https://doi.org/10.1111/j.1420-9101.2005.00885.x

Frederick, A. R., Heras, J., Friedman, C. S., & German, D. P. (2022). Withering syndrome induced gene expression changes and a de-novo transcriptome for the Pinto abalone, Haliotis kamtschatkana. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 41, 100930. https://doi.org/10.1016/j.cbd.2021.100930

I., & Takahashi, L. S. (2022). Carbohydrate tolerance in Amazon tambaqui (Colossoma macropomum) revealed by NMR-metabolomics-Are glucose and fructose different sugars for fruit-eating fish?. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 41, 100928. https://doi.org/10.1016/j.cbd.2021.100928

Fu, C., Li, F., Xu, C., Wang, L., Tian, M., Chang, Y., Xi Y., & Lin, R. (2022). Transcriptome analysis of the potential molecular mechanism of enteritis in intestine and hepatopancreas of Litopenaeus vannamei. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 42, 100964. https://doi.org/10.1016/j.cbd.2022.100964

Gabriel, J. E., Ferro, J. A., Stefani, R. M. P., Ferro, M. I. T., Gomes, S. L., & Macari, M. (1996). Effect of acute heat stress on heat shock protein 70 messenger RNA and on heat shock protein expression in the liver of broilers. *British poultry science*, *37*(2), 443-449. https://doi.org/10.1080/00071669608417875

Gandar, A., Laffaille, P., Marty-Gasset, N., Viala, D., Molette, C., & Jean, S. (2017). Proteome response of fish under multiple stress exposure: effects of pesticide mixtures and temperature increase. Aquatic toxicology, 184, 61-77. https://doi.org/10.1016/j.aquatox.2017.01.004

Gao, H., Hartnett, S., & Li, Y. (2017). Ubiquitin C-Terminal Hydrolase L1 regulates myoblast proliferation and differentiation. *Biochemical and biophysical research communications*, *492*(1), 96-102. https://doi.org/10.1016/j.bbrc.2017.08.027

Goikoetxea, A., Sadoul, B., Blondeau-Bidet, E., Aerts, J., Blanc, M. O., Parrinello, H., Barrachina C., Pratlong M., & Geffroy, B. (2021). Genetic pathways underpinning hormonal stress responses in fish exposed to short-and long-term warm ocean temperatures. *Ecological Indicators*, *120*, 106937. https://doi.org/10.1016/j.ecolind.2020.106937

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., Palma F., Birren B. W., Nusbaum C., Lindbland-Toh K., Friedman N., & Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology, 29(7), 644. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3571712/

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., & Byrum, S. D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular omics*, *17*(2), 170-185. https://doi.org/10.1038%2Fnbt.1883

Gu, Z. T., Wang, H., Li, L., Liu, Y. S., Deng, X. B., Huo, S. F., Yuan F. F., Liu Z. F., Tong H. S., & Su, L. (2014). Heat stress induces apoptosis through transcription-independent p53-mediated mitochondrial pathways in human umbilical vein endothelial cell. *Scientific reports*, *4*(1), 4469. https://doi.org/10.1038/srep04469

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger M. B., Eccles D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N., & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols, 8(8), 1494-1512. https://doi.org/10.1038/nprot.2013.084

Hanson, P. J., & Walker, A. P. (2020). Advancing global change biology through experimental manipulations: Where have we been and where might we go?. Global change biology, 26(1), 287-299. https://doi.org/10.1111/gcb.14894

Hao, Y., Feng, Y., Yang, P., Cui, Y., Liu, J., Yang, C., & Gu, X. (2016). Transcriptome analysis reveals that constant heat stress modifies the metabolism and structure of the porcine longissimus dorsi skeletal muscle. *Molecular Genetics and Genomics*, *291*(6), 2101-2115. https://doi.org/10.1007/s00438-016-1242-8

Hassan, A. M., El Nahas, A. F., Mahmoud, S., Barakat, M. E., & Ammar, A. Y. (2017). Thermal stress of ambient temperature modulate expression of stress and immune-related genes and DNA fragmentation in Nile tilapia (Oreochromis niloticus (Linnaeus, 1758)). *Applied Ecology & Environmental Research*, *15*(3). https://doi.org/10.15666/aeer/1503_13431354

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., Benthuysen J. A., Burrows M. T., Donat M. G., Feng M., Holbrook N. J., Moore P. J., Scannell H. A., Gupta A. S., & Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227-238. https://doi.org/10.1016/j.pocean.2015.12.014

Hokanson, K. E., Kleiner, C. F., & Thorslund, T. W. (1977). Effects of constant temperatures and diel temperature fluctuations on specific growth and mortality rates and yield of juvenile rainbow trout, Salmo gairdneri. Journal of the Fisheries Board of Canada, 34(5), 639-648. https://doi.org/10.1139/f77-100

Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C., Alexander L. V., Burrows M. T., Donat M. T., Hobday A. J., Moore P. J., Perkins-Kirkpatrick S. E., Smale D. A., Straub S. C., & Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. *Nature communications*, *10*(1), 2624. https://doi.org/10.1038/s41467-019-10206-z

Hueber, A. J., Asquith, D. L., McInnes, I. B., & Miller, A. M. (2010). Embracing novel cytokines in RA–complexity grows as does opportunity!. *Best practice & research Clinical rheumatology*, *24*(4), 479-487. https://doi.org/10.1016/j.berh.2010.01.004

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. J Comput Graph Stat, 5(3), 299-314. https://doi.org/10.1080/10618600.1996.10474713

IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf

Islam, M. J., Kunzmann, A., & Slater, M. J. (2022). Responses of aquaculture fish to climate change-induced extreme temperatures: A review. *Journal of the World Aquaculture Society*, *53*(2), 314-366. https://doi.org/10.1111/jwas.12853

Iwama, G. K., Vijayan, M. M., Forsyth, R. B., & Ackerman, P. A. (1999). Heat shock proteins and physiological stress in fish. *American Zoologist*, *39*(6), 901-909. https://doi.org/10.1093/icb/39.6.901

Jacox, M. G., Hazen, E. L., Zaba, K. D., Rudnick, D. L., Edwards, C. A., Moore, A. M., & Bograd, S. J. (2016). Impacts of the 2015–2016 El Niño on the California Current System: Early assessment and comparison to past events. *Geophysical Research Letters*, *43*(13), 7072-7080. https://doi.org/10.1002/2016GL069716

Jain, S., Sharma, G., & Mathur, Y. P. (2013). Effects of temperature variations on fish in lakes. *International Journal of Engineering Research & Technology*, *2*(10), 2516-2523.

https://www.ijert.org/research/effects-of-temperature-variations-on-fish-in-lakes-IJERTV2IS100744.pdf

Jones, D., Miller, P.J. Seasonal migrations of the common Goby, Pomatoschistusmicrops (Kroyer), in Morecambe Bay and elsewhere. *Hydrobiologia* 27, 515–528 (1966). https://doi.org/10.1007/BF00042710

Jonna, R. J. (Author), Animal Diversity Web. (2020, February 5). Gobiidae. Retrieved from https://animaldiversity.org/accounts/Gobiidae/

Krebs, R. A., & Bettencourt, B. R. (1999). Evolution of thermotolerance and variation in the heat shock protein, Hsp70. *American Zoologist*, *39*(6), 910-919. https://doi.org/10.1093/icb/39.6.910

Kumar, A., & Chordia, N. (2017). Role of bioinformatics in biotechnology. *Res Rev Biosci*, *12*(1), 116. https://mccollegeonline.co.in/attendence/classnotes/files/1586267796.pdf

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), 357-359. https://doi.org/10.1038/nmeth.1923

Leeuwis, R. H., & Gamperl, A. K. (2022). Adaptations and plastic phenotypic responses of marine animals to the environmental challenges of the high intertidal zone. In *Oceanography and Marine Biology* (pp. 625-679). CRC Press. https://doi.org/10.1201/9781003288602-13

Li, L., Tan, H., Gu, Z., Liu, Z., Geng, Y., Liu, Y., Tong H., Tang Y., Qiu, J., & Su, L. (2014). Heat stress induces apoptosis through a Ca2+-mediated mitochondrial apoptotic pathway in human umbilical vein endothelial cells. *PloS one*, *9*(12), e111083. https://doi.org/10.1371/journal.pone.0111083

Li, W., Wei, Z., Qiao, Z., Wu, Z., Cheng, L., & Wang, Y. (2013). Proteomics analysis of alfalfa response to heat stress. *PLoS One*, *8*(12), e82725. https://doi.org/10.1371/journal.pone.0082725

Liu, H., Yang, R., Fu, Z., Yu, G., Li, M., Dai, S., Ma, Z., & Zong, H. (2023). Acute thermal stress increased enzyme activity and muscle energy distribution of yellowfin tuna. *Plos one*, *18*(10), e0289606. https://doi.org/10.1371/journal.pone.0289606

Long, Y., Li, L., Li, Q., He, X., & Cui, Z. (2012). Transcriptomic characterization of temperature stress responses in larval zebrafish. PloS one, 7(5), e37209. https://doi.org/10.1371/journal.pone.0037209

Lorenzo, N., Díaz-Poso, A., & Royé, D. (2021). Heatwave intensity on the Iberian Peninsula: Future climate projections. Atmospheric Research, 258, 105655. https://doi.org/10.1016/j.atmosres.2021.105655

Lowey, S., & Trybus, K. M. (2010). Common structural motifs for the regulation of divergent class II myosins. *Journal of Biological Chemistry*, *285*(22), 16403-16407. https://doi.org/10.1074/jbc.R109.025551

Luo, L., Zhang, S., Wu, J., Sun, X., & Ma, A. (2021). Heat stress in macrofungi: effects and response mechanisms. *Applied Microbiology and Biotechnology*, 1-10. https://doi.org/10.1007/s00253-021-11574-7

Ma, B., Zhang, L., Li, J., Xing, T., Jiang, Y., & Gao, F. (2021). Heat stress alters muscle protein and amino acid metabolism and accelerates liver gluconeogenesis for energy supply in broilers. *Poultry science*, *100*(1), 215-223. https://doi.org/10.1016/j.psj.2020.09.090

Madeira, C., Madeira, D., Diniz, M. S., Cabral, H. N., & Vinagre, C. (2016). Thermal acclimation in clownfish: an integrated biomarker response and multi-tissue experimental approach. Ecological Indicators, 71, 280-292. https://doi.org/10.1016/j.ecolind.2016.07.009

Maia, F., Barroso, C. M., & Gaspar, M. B. (2021). Biology of the common cockle Cerastoderma edule (Linnaeus, 1758) in Ria de Aveiro (NW Portugal): Implications for fisheries management. Journal of Sea Research, 171, 102024. https://doi.org/10.1016/j.seares.2021.102024

Mbokodo, I., Bopape, M. J., Chikoore, H., Engelbrecht, F., & Nethengwe, N. (2020). Heatwaves in the future warmer climate of South Africa. Atmosphere, 11(7), 712. https://doi.org/10.3390/atmos11070712

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res, 40(10), 4288-4297. https://doi.org/10.1093/nar/gks042

Miao, L., Zhang, Y., Lin, Y., Liu, B., & Ge, X. (2021). Appropriate leucine supplementation promotes glucose metabolism and enhances energy homeostasis in juvenile crucian carp (Carassius auratus gibelio var. CAS III). Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 40, 100907. https://doi.org/10.1016/j.cbd.2021.100907

Miller, P. J. (1975). Age-structure and life-span in the Common goby, Pomatoschistus microps. Journal of Zoology, 177, 425-448. https://doi.org/10.1111/j.1469-7998.1975.tb02243.x

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto S. C. E., Paladin L., Raj S., Richardson L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic acids research, 49(D1), D412-D419. https://doi.org/10.1093/nar/gkaa913

Moggridge, S., Sorensen, P.H., Morin, G.B., and Hughes, C.S. (2018). Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. Journal of proteome research 17, 1730-1740. http://doi.org/10.1021/acs.jproteome.7b00913

Moss, A., Blackburn-Munro, G., Garry, E. M., Blakemore, J. A., Dickinson, T., Rosie, R., Mitchell, R., & Fleetwood-Walker, S. M. (2002). A role of the ubiquitin–proteasome system in neuropathic pain. *Journal of Neuroscience*, *22*(4), 1363-1372. https://doi.org/10.1523/JNEUROSCI.22-04-01363.2002

Mugwanya, M., Dawood, M. A., Kimera, F., & Sewilam, H. (2022). Anthropogenic temperature fluctuations and their effect on aquaculture: A comprehensive review. *Aquaculture and Fisheries*, *7*(3), 223-243. https://doi.org/10.1016/j.aaf.2021.12.005

Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J. G., & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. Current opinion in biotechnology, 24(1), 22-30. https://doi.org/10.1016/j.copbio.2012.09.004

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. (R package version 1.1-2) https://CRAN.R-project.org/package=RColorBrewer

Nie, L., Wu, G., Culley, D. E., Scholten, J. C., & Zhang, W. (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. Critical reviews in biotechnology, 27(2), 63-75. https://doi.org/10.1080/07388550701334212

NOAA. (2023, July 14). Ongoing Marine Heat Waves in U.S. Waters Explained. National Oceanic and Atmospheric Administration. https://www.noaa.gov/news/ongoing-marine-heat-waves-in-us-waters-explained

Oliver, E. C., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng M., Gupta, A. S., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Staub, S. C., & Wernberg, T. (2018). Longer and more frequent marine heatwaves over the past century. *Nature communications*, *9*(1), 1-12. https://doi.org/10.1038/s41467-018-03732-9

Oliver, E. C., Benthuysen, J. A., Darmaraki, S., Donat, M. G., Hobday, A. J., Holbrook, N. J., Schlegel, R. W., & Sen Gupta, A. (2021). Marine heatwaves. *Annual Review of Marine Science*, *13*, 313-342. https://doi.org/10.1146/annurev-marine-032720-095144

Parente, J., Pereira, M. G., Amraoui, M., & Fischer, E. M. (2018). Heat waves in Portugal: Current regime, changes in future climate and impacts on extreme wildfires. Science of The Total Environment, 631, 534-549. https://doi.org/10.1016/j.scitotenv.2018.03.044

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. science, 308(5730), 1912-1915. https://doi.org/10.1126/science.1111322

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist Reviews*, *30*(1), 19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2643089/

Pörtner, H. O. (2012). Integrating climate-related stressor effects on marine organisms: unifying principles linking molecule to ecosystem-level changes. Marine Ecology Progress Series, 470, 273-290. https://doi.org/10.3354/meps10123

Pörtner, H. O., Bock, C., & Mark, F. C. (2017). Oxygen- and capacity-limited thermal tolerance: bridging ecology and physiology. Journal of Experimental Biology, 220(15), 2685-2696. https://doi.org/10.1242/jeb.134585

Pörtner, H. O., & Peck, M. A. (2010). Climate change effects on fishes and fisheries: towards a cause-and-effect understanding. Journal of Fish Biology, 77(8), 1745-1779. https://doi.org/10.1111/j.1095-8649.2010.02783.x

Post, J. R., & Evans, D. O. (1989). Size-dependent overwinter mortality of young-of-the-year yellow perch (Perca flavescens): laboratory, in situ enclosure, and field experiments. Canadian Journal of Fisheries and Aquatic Sciences, 46(11), 1958-1968. https://doi.org/10.1139/f89-246

Ring, M. J., Lindner, D., Cross, E. F., & Schlesinger, M. E. (2012). Causes of the global warming observed since the 19th century. Atmospheric and Climate sciences, 2(04), 401. http://dx.doi.org/10.4236/acs.2012.24035

Ritchie, H., Roser, M., & Rosado, P. (2020). CO_2 and greenhouse gas emissions. Our world in data. https://ourworldindata.org/co2-and-greenhouse-gas-emissions

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 43(7), e47. https://doi.org/10.1093/nar/gkv007

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. https://doi.org/10.1093/bioinformatics/btp616

Rout, P. K., Kaushik, R., & Ramachandran, N. (2016). Differential expression pattern of heat shock protein 70 gene in tissues and heat stress phenotypes in goats during peak heat stress period. *Cell Stress and chaperones*, *21*(4), 645-651. https://doi.org/10.1007/s12192-016-0689-1

Sakurai, G., Takahashi, S., Yoshida, Y., Yoshida, H., Shoji, J., & Tomiyama, T. (2021). Importance of experienced thermal history: Effect of acclimation temperatures on the high-temperature tolerance and growth performance of juvenile marbled flounder. Journal of Thermal Biology, 97, 102831. https://doi.org/10.1016/j.jtherbio.2020.102831

Salgado, J. P., Cabral, H. N., & Costa, M. J. (2004). Feeding ecology of the gobies Pomatoschistus minutes (Pallas, 1770) and Pomatoschistus microps (Krøyer, 1838) in the upper Tagus estuary, Portugal. *Scientia Marina*, *68*(3), 425-434. https://doi.org/10.3989/scimar.2004.68n3425

Sathiyaa, R., & Vijayan, M. M. (2003). Autoregulation of glucocorticoid receptor by cortisol in rainbow trout hepatocytes. *American Journal of Physiology-Cell Physiology*, *284*(6), C1508-C1515. https://doi.org/10.1152/ajpcell.00448.2002

Shi, K. P., Dong, S. L., Zhou, Y. G., Li, Y., Gao, Q. F., & Sun, D. J. (2019). RNA-seq reveals temporal differences in the transcriptome response to acute heat stress in the Atlantic salmon (Salmo salar). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *30*, 169-178. https://doi.org/10.1016/j.cbd.2018.12.011

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res, 4, 1521. https://doi.org/10.12688/f1000research.7563.2

Sorice, M. (2022). Crosstalk of autophagy and apoptosis. *Cells*, *11*(9), 1479. https://doi.org/10.3390/cells11091479

Stipcich, P., Apostolaki, E. T., Chartosia, N., Efthymiadis, P. T., Jimenez, C. E., La Manna, G., Pansini, A., Principato, E., Resaikos, V., & Ceccherelli, G. (2022). Assessment of Posidonia oceanica traits along a temperature gradient in the Mediterranean Sea shows impacts of marine warming and heat waves. *Frontiers in Marine Science*, *9*, 895354. https://doi.org/10.3389/fmars.2022.895354

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. Nature, 432(7017), 610-614 https://doi.org/10.1038/nature03089

Sui, S. X., Williams, L. J., Holloway-Kew, K. L., Hyde, N. K., & Pasco, J. A. (2020). Skeletal muscle health and cognitive function: a narrative review. *International journal of molecular sciences*, *22*(1), 255. https://doi.org/10.3390/ijms22010255

Thomas, R. E., Gharrett, J. A., Carls, M. G., Rice, S. D., Moles, A., & Korn, S. (1986). Effects of fluctuating temperature on mortality, stress, and energy reserves of juvenile coho salmon. *Transactions of the American Fisheries Society*, *115*(1), 52-59. https://doi.org/10.1577/1548-8659(1986)115<52:EOFTOM>2.0.CO;2

Thorburn, A. (2008). Apoptosis and autophagy: regulatory connections between two supposedly different processes. *Apoptosis*, *13*, 1-9. https://doi.org/10.1007/s10495-007-0154-9

Torday, J. S. (2015). Homeostasis as the mechanism of evolution. Biology, 4(3), 573-590. https://doi.org/10.3390/biology4030573

Tuszynski, J. A., Carpenter, E. J., Huzil, J. T., Malinski, W., Luchko, T., & Luduena, R. F. (2006). The evolution of the structure of tubulin and its potential consequences for the role and function of microtubules in cells and embryos. *International Journal of Developmental Biology*, *50*. https://doi.org/10.1387/ijdb.052063jt

Vanburen, P., Waller, G. S., Harris, D. E., Trybus, K. M., Warshaw, D. M., & Lowey, S. (1994). The essential light chain is required for full force production by skeletal muscle myosin. *Proceedings of the National Academy of Sciences*, *91*(26), 12403-12407. https://doi.org/10.1073/pnas.91.26.12403

Vasu, V. T., Ott, S., Hobson, B., Rashidi, V., Oommen, S., Cross, C. E., & Gohil, K. (2009). Sarcolipin and ubiquitin carboxy-terminal hydrolase 1 mRNAs are over-expressed in skeletal muscles of α-tocopherol deficient mice. *Free radical research*, *43*(2), 106-116. https://doi.org/10.1080/10715760802616676

Wang, Y., Su, C., Liu, Q., Hao, X., Han, S., Doretto, L. B., Rosa, I. F., Yang, Y., Shao, C., & Wang, Q. (2023). Transcriptome analysis revealed the early heat stress response in the brain of Chinese Tongue Sole (Cynoglossus semilaevis). *Animals*, *14*(1), 84. https://doi.org/10.3390/ani14010084

Wang, D., & Liu, B. (2022). Proteomics reveals the changes in energy metabolism associated with reproduction in the clam Meretrix petechialis. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 41, 100954. https://doi.org/10.1016/j.cbd.2021.100954

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. & Gailili, T. (2020). gplots: Various R Programming Tools for Plotting Data. (R package version 3.1.1) https://CRAN.R-project.org/package=gplots

Wernberg, T., Smale, D. A., & Thomsen, M. S. (2012). A decade of climate change experiments on marine organisms: procedures, patterns and problems. Global Change Biology, 18(5), 1491-1498. https://doi.org/10.1111/j.1365-2486.2012.02656.x

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-24277-4

Yeh, S. W., Kug, J. S., Dewitte, B., Kwon, M. H., Kirtman, B. P., & Jin, F. F. (2009). El Niño in a changing climate. *Nature*, *461*(7263), 511-514. https://doi.org/10.1038/nature08316

Zangari, J., Petrelli, F., Maillot, B., & Martinou, J. C. (2020). The multifaceted pyruvate metabolism: role of the mitochondrial pyruvate carrier. *Biomolecules*, *10*(7), 1068. https://doi.org/10.3390/biom10071068

Zhang, J. (2013). Autophagy and mitophagy in cellular damage control. *Redox biology*, *1*(1), 19-23. https://doi.org/10.1016/j.redox.2012.11.008

Zhan, M., Xi, C., Gong, J., Zhu, M., Shui, Y., Xu, Z., Xu, G., & Shen, H. (2022). 16S rRNA gene sequencing analysis reveals an imbalance in the intestinal flora of Eriocheir sinensis with hepatopancreatic necrosis disease. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 42, 100988 https://doi.org/10.1016/j.cbd.2022.100988

Zhang, M., Jiang, M., Bi, Y., Zhu, H., Zhou, Z., & Sha, J. (2012). Autophagy and apoptosis act as partners to induce germ cell death after heat stress in mice. *PloS one*, *7*(7), e41412. https://doi.org/10.1371/journal.pone.0041412

Zhao, L., & Jones, W. A. (2012). Expression of heat shock protein genes in insect stress responses. *Invertebrate Survival Journal*, *9*(1), 93-101. https://www.isj.unimore.it/index.php/ISJ/article/view/264

A

APPENDIX

A.1 Figures and Tables

Table A.1 Sampling

Sample ID	Quality Control Properties (Concentration [ng/uL])	Quality Control Properties (Integrity)	Sequencing Properties (Sequencing reads)
T15CTLPm1 MD10 (M)	188 ng/uL	RIN 9.1	47893540
T1CTLPm1 MD10 (M)	315 ng/uL	RIN 8.9	47622050
T10CTLPm1 MD10 (M)	359 ng/uL	RIN 8.9	47876706
T7MHW2Pm1 MD10 (M)	299 ng/uL	RIN 8.4	47661846
T5MHW2Pm1 MD10 (M)	245 ng/uL	RIN 8.9	50632956
T3MHW2Pm1 MD10 (M)	350 ng/uL	RIN 8.5	48467132
T10CTLPm1 MD25 (M)	393 ng/uL	RIN 8.6	50427090
T1CTLPm1 MD25 (M)	240 ng/uL	RIN 9.2	49069588
T4CTLPm1 MD25 (M)	340 ng/uL	RIN 8.8	49982714
T12MHW1NPm1 MD25 (M)	163 ng/uL	RIN 8.8	43559062
T2MHW2Pm1 MD25(M)	166 ng/uL	RIN 9.1	41636664
T6MHW1Pm1 MD25 (M)	295 ng/uL	RIN 8.6	52625666
T3MHW2Pm1 MD25 (M)	162 ng/uL	RIN 8.6	44383034
T13MHW2Pm MD25 (M)	281 ng/uL	RIN 8.7	46210810
T7MHW2Pm1 MD25 (M)	299 ng/uL	RIN 8.2	51301456

Table A.2 The differentially regulated proteins in CTL_10 vs. MHW2_10

Accession	Description	logFC	Pvalue	padj
Q91233	Heat shock 70 kDa protein	-1.75	4.76E-11	1.91E-10
P34932	Heat shock 70 kDa protein 4	-1.57	5.23E-16	4.18E-15

A0ZSK3	Neoverrucotoxin subunit alpha	3.15	4.42E-07	1.18E-06
Q00981	Ubiquitin carboxyl-terminal hydrolase isozyme L1	-1.92	7.49E-07	1.50E-06
P17506	Elongation factor 1-alpha	2.70	3.71E-06	5.94E-06
Q0CT23	Probable ubiquitination network signaling protein acrB	2.32	2.39E-05	3.19E-05
Q5ADT1	Negative regulator of sporulation MDS3	3.91	1.53E-04	1.53E-04
Q5ZKD9	Kelch-like protein 20	2.30	5.61E-05	6.41E-05

Table A.3 The differentially regulated proteins in CTL_25 vs. MHW1_25

Accession	Description	logFC	Pvalue	padj
Q00981	Ubiquitin carboxyl-terminal hydrolase isozyme L1	-1.53	5.43E-05	5.56E-05
Q94571	Tubulin beta-2 chain	2.88	5.56E-05	5.56E-05
Q05744	Cathepsin D	3.23	3.08E-05	5.56E-05

Table A.4 The differentially regulated proteins in MHW2_10 vs. MHW2_25

Accession	Description	logFC	Pvalue	padj
P04460	Myosin-6 (Fragment)	2.22	1.57E-05	3.64E-05
P48674	Vimentin	-2.51	1.00E-04	1.13E-04
Q7Z2Y8	Interferon-induced very large GTPase 1	-2.77	2.66E-04	2.66E-04
Q8TCY9	Up-regulator of cell proliferation	-2.89	5.46E-05	7.01E-05
Q9CXS4	Centromere protein V	1.58	3.82E-05	5.73E-05
B2UPE6	dTTP/UTP pyrophosphatase	2.02	1.34E-06	6.94E-06
P50532	Structural maintenance of chromosomes protein 4	1.76	1.54E-06	6.94E-06
Q9H2K8	Serine/threonine-protein kinase TAO3	2.23	2.02E-05	3.64E-05
Q5SXY1	Cytospin-B	-2.17	1.74E-05	3.64E-05

Table A.5 The differentially regulated proteins in CTL_25 vs. MHW2_25

Accession	Description	logFC	Pvalue	padj
O35244	Peroxiredoxin-6	2.27	1.50E-04	1.87E-04
Q9XVI9	Myosin-2 essential light chain	2.66	2.19E-04	2.19E-04
Q5R894	Apolipoprotein M	-1.69	1.09E-05	5.46E-05
Q5XIF6	Tubulin alpha-4A chain	2.46	2.19E-04	2.19E-04
Q94571	Tubulin beta-2 chain	3.23	7.77E-06	5.46E-05
Q05744	Cathepsin D	3.78	1.86E-06	2.78E-05
A7M9B2	Protein TIC 214	3.07	1.96E-05	5.87E-05
Q24799	Myophilin	3.03	1.13E-04	1.70E-04
Q5HZA3	Zinc finger CCHC domain-containing protein 12	2.99	1.28E-04	1.74E-04
P49221	Protein-glutamine gamma-glutamyltransferase 4	2.90	5.94E-05	1.31E-04
Q92176	Coronin-1A	3.28	6.36E-05	1.31E-04
P11168	Solute carrier family 2, facilitated glucose transporter member 2	3.82	1.52E-05	5.72E-05
P40122	Adenylyl cyclase-associated protein	3.24	8.39E-05	1.40E-04
Q5SXY1	Cytospin-B	-2.25	6.99E-05	1.31E-04
P97629	Leucyl-cystinyl aminopeptidase	2.48	2.08E-04	2.19E-04

Table A.6 Top 10 altered pathways in CTL_10 vs. MHW2_10 based on transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Adipocytokine signaling pathway	0.23	0.80	0.13	-0.76	-1.25	3	Q9DGD9
Alanine, aspartate and glutamate metabolism	0.27	0.80	0.11	0.59	1.20	7	Q568F6;Q66I24;Q6NSN5
Apelin signaling pathway	0.18	0.80	0.15	-0.67	-1.27	5	A8C984
Apoptosis	0.06	0.80	0.28	-0.78	-1.49	5	Q9DGD9;Q6NW59;Q4G3H4;Q6IQM2
Ascorbate and aldarate metabolism	0.22	0.80	0.14	-0.76	-1.26	3	Q568L5;Q6TLF6
C-type lectin receptor signaling pathway	0.26	0.80	0.12	-0.61	-1.22	6	Q9DGD9
Calcium signaling pathway	0.02	0.80	0.35	-0.73	-1.64	9	A8C984;Q90413;Q9DE49;O57341;A5

							D6R3;Q7SY24;A0JMD4
Cell cycle	0.07	0.80	0.29	0.80	1.48	5	Q1LX51;Q568H3;F1QY75
Cellular senescence	0.07	0.80	0.23	0.70	1.46	8	Q6P3H7;A0AUQ6;A0A2R8RWN9;Q08CM4
Cysteine and methionine metabolism	0.24	0.80	0.12	0.67	1.23	5	Q9PVK4;Q9PVK5;A9JRE2

Table A.7 Top 10 altered pathways in CTL_25 vs. MHW1_25 based on transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
ECM-receptor interaction	0.01	0.65	0.38	-0.90	-1.60	4	Q8JHV6
Neuroactive ligand-receptor interaction	0.02	0.65	0.35	-0.89	-1.58	4	P0C7U5;Q98880;F1QQI2
mTOR signaling pathway	0.01	0.65	0.38	-0.68	-1.72	15	Q7SY24;Q7ZTW4;Q7ZVL2;Q801F7;Q802U2;Q0P3X8;Q92048;Q6PFQ0
Alanine, aspartate and glutamate metabolism	0.05	0.89	0.32	0.76	1.54	7	Q66I24;Q7SY23;Q7SYK7;Q6NSN5;Q568F6
Arginine biosynthesis	0.06	0.89	0.26	0.94	1.38	2	Q66I24;Q7SYK7
Cell adhesion molecules	0.03	0.89	0.35	0.84	1.63	6	A1XQY1;A1XQX0;A3KPA0
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	0.04	0.89	0.32	-0.95	-1.43	2	Q7SXZ1
Sulfur metabolism	0.06	0.89	0.32	-0.94	-1.42	2	Q6PHD9
Ascorbate and aldarate metabolism	0.08	0.90	0.22	0.86	1.39	3	Q6TLF6;Q568L5
N-Glycan biosynthesis	0.08	0.90	0.22	0.81	1.40	4	Q6PBS6;Q7ZV50

Table A.8 Top 10 altered pathways in CTL_25 vs. MHW2_25 based on transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Terpenoid backbone biosynthesis	0.00	0.27	0.43	0.98	1.44	2	Q5U403;Q6AZA0
Autophagy - animal	0.02	0.71	0.35	-0.69	-1.63	10	E7FA21;E7F590;Q6P132;Q5RI56

Neuroactive ligand-receptor interaction	0.02	0.71	0.35	-0.86	-1.54	4	Q98880;P0C7U5
Porphyrin metabolism	0.01	0.71	0.38	-0.99	-1.35	1	Q9PTS2
Mitophagy - animal	0.03	0.77	0.35	-0.57	-1.55	16	Q6NYK8;Q6P132;Q5M7X9;Q5RI56;Q32LU1;A2BGT0;Q9DGD9;Q6NVC5
2-Oxocarboxylic acid metabolism	0.28	0.92	0.11	0.79	1.17	2	Q7SYK7;Q5PRA2
Adherens junction	0.44	0.92	0.10	-0.39	-1.00	14	Q6PCS4;A1A5H8
Adrenergic signaling in cardiomyocytes	0.29	0.92	0.12	-0.53	-1.17	8	O42376;Q6NY64;Q6JWV8;Q1LYG4
Alanine, aspartate and glutamate metabolism	0.09	0.92	0.20	0.71	1.40	7	A4IGD2;Q66I24;Q7SY23;Q6NSN5;Q7SYK7;Q568F6;A8KB34
Amino sugar and nucleotide sugar metabolism	0.27	0.92	0.11	0.69	1.18	4	Q6GMK8;Q6GMI9;Q7ZWD4

Table A.9 Top 10 altered pathways in MHW1_25 vs. MHW2_25 based on transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Adipocytokine signaling pathway	0.16	0.91	0.17	-0.77	-1.28	3	Q9DGD9
Apoptosis	0.06	0.91	0.29	-0.79	-1.56	5	Q9DGD9;Q6IQM2
Autophagy - animal	0.19	0.91	0.17	-0.53	-1.27	10	Q9DGD9;E7FA21;E7F590;Q5RI56
Biosynthesis of amino acids	0.14	0.91	0.16	0.78	1.30	4	Q8JH71;Q66I24
Biosynthesis of nucleotide sugars	0.16	0.91	0.14	0.77	1.28	4	Q6GMK8;Q4V8T0;Q7ZWD4;Q6GMI9
C-type lectin receptor signaling pathway	0.19	0.91	0.16	-0.63	-1.32	6	Q9DGD9
Cell adhesion molecules	0.07	0.91	0.27	-0.72	-1.50	6	A1XQX0;A1XQY1
Drug metabolism - other enzymes	0.05	0.91	0.27	0.76	1.46	7	Q5RGV1;Q6DI51;A5WVX0
Efferocytosis	0.11	0.91	0.20	-0.81	-1.36	3	E7F590;A4IG72;O42376
Fanconi anemia pathway	0.10	0.91	0.20	0.96	1.27	1	Q6NY74

Table A.10 Top 10 altered pathways in MHW2_10 vs. MHW2_25 based on transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Mitophagy - animal	0.00	0.14	0.46	-0.72	-1.83	16	Q5M7X9;Q6NW85;Q6NYK8;Q5RI56;Q32LU1;Q6P132;A2BGT0;Q6DH66
Adipocytokine signaling pathway	0.14	0.95	0.16	0.81	1.29	3	A1A5Z0;Q4G3H4;Q9DGD9
Adrenergic signaling in cardiomyocytes	0.55	0.95	0.07	-0.43	-0.91	8	O42376;Q1LYG4;P13104;Q6JWV8;Q5PQZ7;Q6NY64
Alanine, aspartate and glutamate metabolism	0.49	0.95	0.08	0.50	1.00	7	Q7SY23;A4IGD2;Q66I24;Q6NSN5;Q7SYK7;A8KB34
Amino sugar and nucleotide sugar metabolism	0.17	0.95	0.15	0.77	1.30	4	Q6GMK8;Q6GMI9
(KEGG) Aminoacyl-tRNA biosynthesis	0.64	0.95	0.07	-0.47	-0.88	5	Q6DRC0;F1QAJ4;A2BHB7;Q803A7
Apelin signaling pathway	0.33	0.95	0.10	0.65	1.17	5	A8C984;A0T2N3;Q90Y57;Q6PH57;A1L271
Arginine biosynthesis	0.32	0.95	0.10	0.77	1.13	2	Q66I24;Q7SYK7
Ascorbate and aldarate metabolism	0.18	0.95	0.14	0.79	1.25	3	Q4V8T0;Q6TLF6;Q568L5
Autophagy - animal	0.10	0.95	0.20	-0.62	-1.41	10	E7FA21;Q5RI56;E7F590;Q6P132

Table A.11 Top 10 altered pathways in CTL_10 vs. MHW2_10 based on proteomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
2-Oxocarboxylic acid metabolism	0.39	0.96	0.09	0.72	1.10	2	Q5PRA2;Q7SYK7
Adherens junction	0.46	0.96	0.09	0.55	1.02	5	O93409;B7ZC77
Adipocytokine signaling pathway	0.63	0.96	0.07	0.68	0.91	1	Q4G3H4
Aminoacyl-tRNA biosynthesis	0.31	0.96	0.11	0.77	1.17	2	Q803A7;Q6DRC0
Apelin signaling pathway	0.51	0.96	0.08	-0.76	-0.99	1	Q6PH57

Apoptosis	0.48	0.96	0.08	0.68	1.03	2	Q4G3H4;Q9I9H8
Ascorbate and aldarate metabolism	0.46	0.96	0.08	0.69	1.05	2	Q568L5;Q6TLF6
Butanoate metabolism	0.06	0.96	0.26	-0.89	-1.43	3	Q6AZA0
Carbon metabolism	0.58	0.96	0.07	-0.54	-0.94	4	Q6AZA0
Cardiac muscle contraction	0.59	0.96	0.07	0.70	0.94	1	P13104

Table A.12 Top 10 altered pathways in CTL_25 vs. MHW1_25 based on proteomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Adipocytokine signaling pathway	0.02	0.29	0.35	-0.99	-1.28	1	Q4G3H4
C-type lectin receptor signaling pathway	0.02	0.29	0.35	-0.93	-1.51	3	Q4G3H4
Cytosolic DNA-sensing pathway	0.02	0.29	0.35	-0.99	-1.28	1	Q4G3H4
MAPK signaling pathway	0.02	0.29	0.35	-0.99	-1.28	1	Q4G3H4
Pyruvate metabolism	0.01	0.29	0.38	-0.89	-1.70	6	Q7SX99;Q6AZA0;Q9PVK4;F1QXM5
RIG-I-like receptor signaling pathway	0.02	0.29	0.35	-0.99	-1.28	1	Q4G3H4
Salmonella infection	0.02	0.29	0.38	-0.78	-1.74	10	Q4G3H4;Q66I73;Q6NZW8;O57521
Toll-like receptor signaling pathway	0.02	0.29	0.35	-0.99	-1.28	1	Q4G3H4
Inositol phosphate metabolism	0.05	0.56	0.28	0.90	1.39	3	A5D6R3;Q6IQE1
Calcium signaling pathway	0.07	0.72	0.24	0.97	1.29	1	A5D6R3

Table A.13 Top 10 altered pathways in CTL_25 vs. MHW2_25 based on proteomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
2-Oxocarboxylic acid metabolism	0.11	0.99	0.20	0.90	1.34	2	Q5PRA2;Q7SYK7
Adherens junction	0.88	0.99	0.06	0.32	0.62	5	O93409;Q66I73
Adipocytokine signaling pathway	0.62	0.99	0.07	0.70	0.92	1	Q4G3H4
Adrenergic signaling in cardiomyocytes	0.06	0.99	0.28	0.82	1.47	4	P13104;Q5PQZ7
Alanine, aspartate and glutamate metabolism	0.72	0.99	0.05	-0.45	-0.82	5	Q66I24
Amino sugar and nucleotide sugar metabolism	0.10	0.99	0.29	-0.97	-1.31	1	Q6GMK8
Aminoacyl-tRNA biosynthesis	0.80	0.99	0.06	0.49	0.73	2	Q803A7;Q6DRC0
Apelin signaling pathway	0.63	0.99	0.07	-0.67	-0.91	1	Q6PH57

Apoptosis	0.47	0.99	0.09	0.70	1.04	2	Q9I9H8;Q4G3H4
Arachidonic acid metabolism	0.43	0.99	0.09	0.81	1.06	1	Q7ZUC7

Table A.14 Top 10 altered pathways in MHW1_25 vs. MHW2_25 based on proteomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Adipocytokine signaling pathway	0.01	0.09	0.41	1.00	1.30	1	Q4G3H4
Cytosolic DNA-sensing pathway	0.01	0.09	0.41	1.00	1.30	1	Q4G3H4
MAPK signaling pathway	0.01	0.09	0.41	1.00	1.30	1	Q4G3H4
Pyruvate metabolism	0.00	0.09	0.41	0.85	1.90	6	Q9PVK4;Q7SX99;F1QXM5;Q6P963;Q6AZA0
RIG-I-like receptor signaling pathway	0.01	0.09	0.41	1.00	1.30	1	Q4G3H4
Toll-like receptor signaling pathway	0.01	0.09	0.41	1.00	1.30	1	Q4G3H4
Cysteine and methionine metabolism	0.01	0.18	0.38	0.89	1.72	4	Q9PVK4;A9JRE2;Q7SYK7
Propanoate metabolism	0.04	0.60	0.32	0.91	1.44	2	Q9PVK4;Q6NYL3
2-Oxocarboxylic acid metabolism	0.07	0.63	0.26	0.90	1.42	2	Q5PRA2;Q7SYK7
Adrenergic signaling in cardiomyocytes	0.08	0.63	0.29	0.75	1.46	4	P13104;Q5PQZ7;Q6JWV8

Table A.15 Top 10 altered pathways in MHW2_10 vs. MHW2_25 based on proteomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Cysteine and methionine metabolism	0.00	0.12	0.46	0.95	1.84	4	Q7SYK7;A9JRE2;Q9PVK4;Q32LQ4
2-Oxocarboxylic acid metabolism	0.13	0.66	0.19	0.86	1.35	2	Q7SYK7;Q5PRA2
Adipocytokine signaling pathway	0.11	0.66	0.19	0.95	1.27	1	Q4G3H4
Alanine, aspartate and glutamate metabolism	0.08	0.66	0.29	0.72	1.51	5	Q568F6;Q7SYK7;A8KB34
Cardiac muscle contraction	0.13	0.66	0.18	0.94	1.25	1	P13104
Cytosolic DNA-sensing pathway	0.11	0.66	0.19	0.95	1.27	1	Q4G3H4
ECM-receptor interaction	0.07	0.66	0.21	-0.86	-1.37	3	Q8JGW0

Focal adhesion	0.11	0.66	0.16	-0.71	-1.34	6	Q8JGW0
Glycerolipid metabolism	0.17	0.66	0.14	-0.83	-1.25	2	Q6NYV8
Glycerophospholipid metabolism	0.10	0.66	0.18	-0.84	-1.35	3	Q502J0;Q5XIZ6;Q6NYV8

Table A.16 Top 10 altered pathways in CTL_10 vs. MHW2_10 based on proteomic and transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Adipocytokine signaling pathway	0.23	0.80	0.13	-0.76	-1.25	3	Q9DGD9
Alanine, aspartate and glutamate metabolism	0.27	0.80	0.11	0.59	1.20	7	Q568F6;Q66I24;Q6NSN5
Apelin signaling pathway	0.18	0.80	0.15	-0.67	-1.27	5	A8C984
Apoptosis	0.06	0.80	0.28	-0.78	-1.49	5	Q9DGD9;Q6NW59;Q4G3H4;Q6IQM2
Ascorbate and aldarate metabolism	0.22	0.80	0.14	-0.76	-1.26	3	Q568L5;Q6TLF6
C-type lectin receptor signaling pathway	0.26	0.80	0.12	-0.61	-1.22	6	Q9DGD9
Calcium signaling pathway	0.02	0.80	0.35	-0.73	-1.64	9	A8C984;Q90413;Q9DE49;O57341;A5D6R3;Q7SY24;A0JMD4
Cell cycle	0.07	0.80	0.29	0.80	1.48	5	Q1LX51;Q568H3;F1QY75
Cellular senescence	0.07	0.80	0.23	0.70	1.46	8	Q6P3H7;A0AUQ6;A0A2R8RWN9;Q08CM4
Cysteine and methionine metabolism	0.24	0.80	0.12	0.67	1.23	5	Q9PVK4;Q9PVK5;A9JRE2

Table A.17 Top 10 altered pathways in CTL_25 vs. MHW1_25 based on proteomic and transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Cell adhesion molecules	0.01	0.51	0.38	0.84	1.62	6	A1XQY1;A1XQX0;A3KPA0
ECM-receptor interaction	0.01	0.51	0.38	-0.90	-1.55	4	Q8JHV6
Neuroactive ligand-receptor interaction	0.02	0.51	0.35	-0.89	-1.54	4	P0C7U5;Q98880;F1QQI2
mTOR signaling pathway	0.01	0.51	0.38	-0.68	-1.71	15	Q7SY24;Q7ZTW4;Q7ZVL2;Q801F7;Q802U2;Q0P3

							X8;Q92048;Q6PFQ0
Alanine, aspartate and glutamate metabolism	0.02	0.52	0.35	0.76	1.54	7	Q66I24;Q7SY23;Q7SYK7;Q6NSN5;Q568F6
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	0.02	0.52	0.35	-0.95	-1.42	2	Q7SXZ1
Arginine biosynthesis	0.03	0.54	0.32	0.94	1.38	2	Q66I24;Q7SYK7
Sulfur metabolism	0.03	0.54	0.32	-0.94	-1.41	2	Q6PHD9
Biosynthesis of amino acids	0.05	0.75	0.27	0.80	1.41	4	Q66I24;Q7SYK7;Q08BC6
Ascorbate and aldarate metabolism	0.07	0.87	0.24	0.86	1.37	3	Q6TLF6;Q568L5

Table A.18 Top 10 altered pathways in CTL_25 vs. MHW2_25 based on proteomic and transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Terpenoid backbone biosynthesis	0.00	0.40	0.43	0.98	1.46	2	Q5U403;Q6AZA0
Autophagy - animal	0.03	0.91	0.35	-0.69	-1.61	10	E7FA21;E7F590;Q6P132;Q5RI56
Neuroactive ligand-receptor interaction	0.03	0.91	0.35	-0.86	-1.51	4	Q98880;P0C7U5
Porphyrin metabolism	0.02	0.91	0.35	-0.99	-1.34	1	Q9PTS2
2-Oxocarboxylic acid metabolism	0.25	0.96	0.12	0.79	1.17	2	Q7SYK7;Q5PRA2
Adherens junction	0.47	0.96	0.09	-0.39	-0.99	14	Q6PCS4;A1A5H8
Adrenergic signaling in cardiomyocytes	0.29	0.96	0.12	-0.53	-1.15	8	O42376;Q6NY64;Q6JWV8;Q1LYG4
Alanine, aspartate and glutamate metabolism	0.07	0.96	0.22	0.71	1.41	7	A4IGD2;Q66I24;Q7SY23;Q6NSN5;Q7SYK7;Q568F6;A8KB34
Amino sugar and nucleotide sugar metabolism	0.26	0.96	0.11	0.69	1.18	4	Q6GMK8;Q6GMI9;Q7ZWD4
Apoptosis	0.27	0.96	0.12	-0.63	-1.21	5	Q6IQM2

Table A.19 Top 10 altered pathways in MHW1_25 vs. MHW2_25 based on proteomic and transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Apoptosis	0.04	0.82	0.32	-0.79	-1.55	5	Q9DGD9;Q6IQM2

Cell adhesion molecules	0.06	0.82	0.29	-0.72	-1.47	6	A1XQX0;A1XQY1
Drug metabolism - other enzymes	0.04	0.82	0.29	0.76	1.47	7	Q5RGV1;Q6DI51;A5WVX0
Fanconi anemia pathway	0.06	0.82	0.27	0.96	1.31	1	Q6NY74
Fructose and mannose metabolism	0.05	0.82	0.27	0.88	1.39	3	Q6GMK8;Q8JH71;Q29RA5
Mitophagy - animal	0.06	0.82	0.32	-0.57	-1.53	16	Q9DGD9;Q6NYK8;Q32LU1;Q6NVC5;Q5U3I0;Q5RI56
Pentose phosphate pathway	0.06	0.82	0.25	0.95	1.30	1	Q8JH71
Porphyrin metabolism	0.07	0.82	0.25	-0.97	-1.28	1	Q9PTS2
Protein processing in endoplasmic reticulum	0.01	0.82	0.38	-0.94	-1.62	3	Q9DGD9;Q6NXA9
Viral life cycle - HIV-1	0.05	0.82	0.27	0.88	1.39	3	P79735;Q6NWC6

Table A.20 Top 10 altered pathways in MHW2_10 vs. MHW2_25 based on proteomic and transcriptomic analysis

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Mitophagy - animal	0.00	0.15	0.46	-0.72	-1.78	16	Q5M7X9;Q6NW85;Q6NYK8;Q5RI56;Q32LU1;Q6P132;A2BGT0;Q6DH66
Adipocytokine signaling pathway	0.13	0.86	0.17	0.81	1.31	3	A1A5Z0;Q4G3H4;Q9DGD9
Amino sugar and nucleotide sugar metabolism	0.15	0.86	0.16	0.77	1.32	4	Q6GMK8;Q6GMI9
Ascorbate and aldarate metabolism	0.18	0.86	0.15	0.79	1.27	3	Q4V8T0;Q6TLF6;Q568L5
Autophagy - animal	0.12	0.86	0.18	-0.62	-1.37	10	E7FA21;Q5RI56;E7F590;Q6P132
Biosynthesis of amino acids	0.15	0.86	0.16	0.77	1.33	4	Q8JH71;Q66I24;Q08BC6;Q7SYK7
Biosynthesis of nucleotide sugars	0.10	0.86	0.20	0.80	1.38	4	Q6GMK8;Q4V8T0;Q6GMI9
Calcium signaling pathway	0.15	0.86	0.16	0.63	1.34	9	A8C984;Q9DE49;Q7SY24;O57341;Q2HXK8;Q90413

Cellular senescence	0.13	0.86	0.17	-0.65	-1.36	8	Q08CM4;O42376;A0A2R8RWN9;Q6P3H7
Citrate cycle (TCA cycle)	0.09	0.86	0.21	-0.83	-1.37	3	Q68FN7;A5PL98

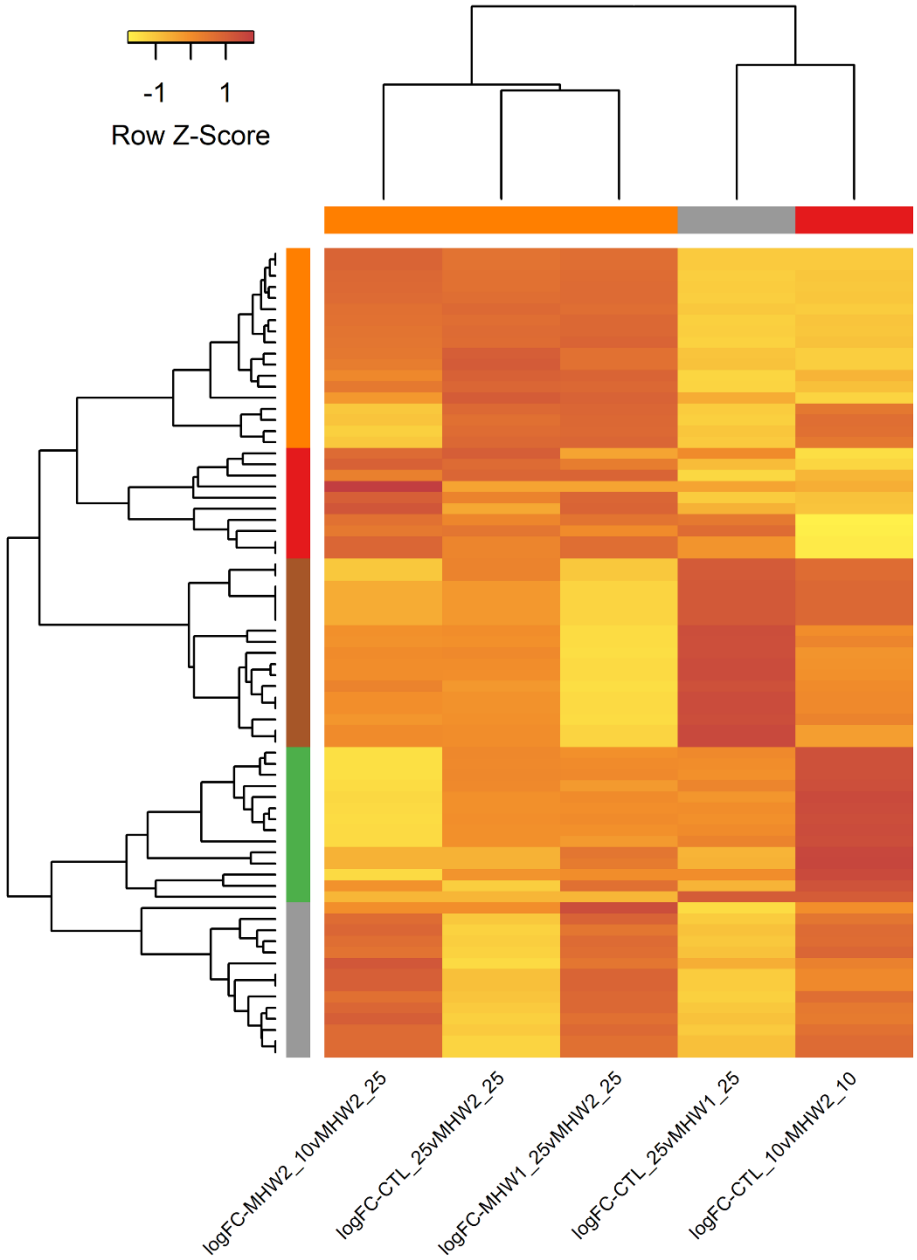

Figure A.1 Heatmap illustrating relative gene expression of the differentially expressed ORFs between treatments. Cut-off for differential expression was set at | log2FC | > 1.5 and FDR-adjusted $p < 0.05$. The horizontal dendrogram shows the association between the five contrasts (control group vs. MHW1 group day 25; control group vs. MHW2 group day 25; MHW2 on day 10 vs. day 25; control group vs. MHW2 group day 10; MHW1 group vs. MHW2 group day 25), whereas the vertical dendrogram shows the association between protein-coding genes. Relative high and low expression is indicated by yellow and red, respectively, after normalization per row (gene). Dendrograms were obtained using cluster analysis taking Euclidian distances and complete linkage, as metric and amalgamation rule, respectively.

A.2 Scripts

A.2.1 Transcriptomics R Script

Please refer to https://github.com/jdplopes/MScThesis_Transcriptomics for the complete R script.

```
##Master's Thesis
##João Lopes
set.seed(1)
setwd("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Transcriptomics")

##Load a R package
library(tximport)
library(limma)
library(edgeR)
library(seqinr)
library(ggplot2)
library(cowplot)
library(gplots)
library(RColorBrewer)
library(gridGraphics)
library(carData)
library(car)
library(rhdf5)
library(data.table)
library(multiGSEA)
library(tidyverse)
library(readxl)
library(tidyr)
library(forcats)
windowsFonts(Calibri = windowsFont("Arial"))

##Create folders
dir.create("Blast_Drerio")
dir.create("Blast_Drerio/Files")
dir.create("Blast_Drerio/Files/Tables")
dir.create("Blast_Drerio/Plots")
dir.create("Blast_Drerio/Plots/Dotplot")

dir.create("Blast_SwissProt")
dir.create("Blast_SwissProt/Files")
dir.create("Blast_SwissProt/Files/Tables")
dir.create("Blast_SwissProt/Plots")
dir.create("Blast_SwissProt/Plots/Heatmap")
dir.create("Blast_SwissProt/Plots/PieCharts")

dir.create("Trinity")
dir.create("Trinity/Files")
dir.create("Trinity/Files/Tables")
dir.create("Trinity/Plots")
dir.create("Trinity/Plots/Volcano plots")
dir.create("Trinity/Plots/Heatmap")
dir.create("Trinity/Plots/PieCharts")

##Directories
##Folders
pathFiles_Trinity <- "Trinity/Files/"
pathTables_Trinity <- "Trinity/Files/Tables/"
pathVolcano_Trinity <- "Trinity/Plots/Volcano plots/"
```

```
pathHeatmap_Trinity <- "Trinity/Plots/Heatmap/"
pathPieCharts_Trinity <- "Trinity/Plots/PieCharts/"

pathFiles_BSwiss <- "Blast_SwissProt/Files/"
pathTables_BSwiss <- "Blast_SwissProt/Files/Tables/"
pathPieCharts_BSwiss <- "Blast_SwissProt/Plots/PieCharts/"
pathHeatmap_BSwiss <- "Blast_SwissProt/Plots/Heatmap/"

pathFiles_BDrerio <- "Blast_Drerio/Files/"
pathTables_BDreio <- "Blast_Drerio/Files/Tables/"
pathDotplot <- "Blast_Drerio/Plots/Dotplot/"

pathBothOmicsPathways <- "C:/Users/jdpl2/OneDrive/Ambiente de Trabalho/Mestrado/2°
Ano/Both Omics/"

#######################################Functions#######################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

relativeExpression<-function(treatment1, treatment2, Design, fit){
  contrast<-paste(treatment1, "-", treatment2, sep = "")
  treatment1vtreatment2<-makeContrasts(contrasts = contrast, levels = Design)
  print(treatment1vtreatment2)
  lrt<-glmLRT(fit, contrast = treatment1vtreatment2)
  print(topTags(lrt))
  return(lrt)
}

volcanoPlot<-function(pathFiles_Trinity, colour, legend, title, contrast, i){
  load(paste(pathFiles_Trinity, "lrt.RData", sep = ""))
  print(colnames(DEG_trinity))
  signif<--log10(DEG_trinity[,paste("FDRp-", contrast, sep = "")])
  plot(DEG_trinity[,paste("logFC-", contrast, sep = "")], signif, pch = ".", xlab =
expression(log[2] * FC), ylab = expression(log[10] * FDR), main = NULL,
       panel.first={
         points(0, 0, pch = 16, cex = 1e6, col = "grey95")
         grid(col = "white", lty = 1)
       }
  )
  points(DEG_trinity[which(DEG_trinity[contrast] == 1), paste("logFC-", contrast, sep =
"")], -log10(DEG_trinity[which(DEG_trinity[contrast] == 1), paste("FDRp-", contrast, sep
= "")]), pch = 20, cex = 1.5, col = colour[1,])
  points(DEG_trinity[which(DEG_trinity[contrast] == -1), paste("logFC-", contrast, sep =
"")], -log10(DEG_trinity[which(DEG_trinity[contrast] == -1), paste("FDRp-", contrast,
sep = "")]), pch = 20, cex = 1.5, col = colour[2,])
}

heatmapGraph<-function(colCutoff, rowCutoff, data, path){
  heatData<-as.matrix(data[,c(2,6,10,14,18)])
  rownames(heatData)<-data$Description
  head(heatData)
  clustFunction <- function(x) hclust(x, method="complete")
  distFunction <- function(x) dist(x,method="euclidean")
  #clusterint (sidebars)
  cCol<-colorRampPalette(brewer.pal(9, "Set1"))
  colFit<-clustFunction(distFunction(t(heatData)))
  cClusters<-cutree(colFit,h=colCutoff)
  cHeight<-length(unique(as.vector(cClusters)))
  colHeight = cCol(cHeight)
  cRow<-colorRampPalette(brewer.pal(9,"Set1"))
  rowFit<-clustFunction(distFunction(heatData))
  rClusters<-cutree(rowFit,h=rowCutoff)
  rHeight<-length(unique(as.vector(rClusters)));
```

```
  rowHeight = cRow(rHeight)
  xDimension<-25
  yDimension<-25
  windows(xDimension, yDimension)
  par(family="Arial")
  #Heatmap
  heatColour<-colorRampPalette(c("#fff04a", "#f28e2b", "#c23e40"))(n = 100)
  heatmap.2(heatData,
            hclust=clustFunction,
            distfun=distFunction,
            ColSideColors=colHeight[cClusters],
            RowSideColors=rowHeight[rClusters],
            density.info="none",
            col=heatColour,
            trace="none",
            scale="row",
            cexCol = 0.8,
            lhei = c(1,5),
            lwid = c(1.5,5),
            offsetRow = 0,
            offsetCol = 0,
            srtCol = 45,
            key.title = NA,
            margins = c(8,13),
            labRow = "")
  recordPlot()
  dev.print(tiff, paste(path, "Heatmap", ".tif", sep =""), height = 15, width = 15,
units = 'cm', res=600)
}

pieChart <- function(data, i,type,contrasts,colour){
  if (type == "transcripts"){
    row_data <- data[i, ]
    long_data <- data.frame(
      Type = c("Overexpressed transcripts", "Underexpressed transcripts"),
      Count = c(row_data$`Overexpressed in treatment 2`, row_data$`Underexpressed in
treatment 2`)
    )
    p <- ggplot(long_data, aes(x = "", y = Count, fill = Type)) +
      geom_bar(width = 1, stat = "identity", color = "black", linewidth = 1,linetype =
"solid") +
      geom_text(aes(label = Count), position = position_stack(vjust = 0.5), color =
"white", size = 12) + # Add numbers inside the pie chart
      coord_polar(theta = "y") +
      theme_void() +
      scale_fill_manual(name = NULL, values = c("Overexpressed transcripts" =
colour[,i][1], "Underexpressed transcripts" = colour[,i][2])) +
      labs(title = NULL) +
      theme(legend.position = "bottom", legend.spacing.y = unit(0.1, "cm"),
            legend.text = element_text(size = 15))
    ggsave(paste(pathPieCharts_Trinity, "PieChart-", row_data[1], ".svg", sep =""), plot
= p, height = 15, width = 15,  units = 'cm', dpi=600)
    ggsave(paste(pathPieCharts_Trinity, "PieChart-", row_data[1], ".tif", sep =""), plot
= p, height = 15, width = 15,  units = 'cm', dpi=600)
  }
  else if (type == "ORFs"){
    row_data <- data[i, ]
    long_data <- data.frame(
      Type = c("Overexpressed ORFs", "Underexpressed ORFs"),
      Count = c(row_data$`Overexpressed in treatment 2`, row_data$`Underexpressed in
treatment 2`)
    )
    p <- ggplot(long_data, aes(x = "", y = Count, fill = Type)) +
```

```
      geom_bar(width = 1, stat = "identity", color = "black", linewidth = 1, linetype =
"dashed") +
      geom_text(aes(label = Count), position = position_stack(vjust = 0.5), color =
"white", size = 12) +
      coord_polar(theta = "y") +
      theme_void() +
      scale_fill_manual(name = NULL, values = c("Overexpressed ORFs" = colour[,i][1],
"Underexpressed ORFs" = colour[,i][2])) +
      labs(title = NULL) +
      theme(legend.position = "bottom", legend.spacing.y = unit(0.1, "cm"),
            legend.text = element_text(size = 15))
    ggsave(paste(pathPieCharts_BSwiss, "PieChart-", row_data[1], ".svg", sep =""), plot
= p, height = 15, width = 15,  units = 'cm', dpi=600)
    ggsave(paste(pathPieCharts_BSwiss, "PieChart-", row_data[1], ".tif", sep =""), plot
= p, height = 15, width = 15,  units = 'cm', dpi=600)
  }
}

dotplot <- function (data,path) {
  o <- ggplot(data, aes(x = contrast, y = pathway)) +
    geom_point(aes(size = 1-padj, color = NES, fill = NES), shape = 21, stroke = 0.5) +
    scale_color_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_size_continuous(range = c(3, 15)) +
    labs(x = "Contrast", y = "Pathway",
         size = "1-p.adjust",
         color = "NES",
         fill = "NES") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 18),
          axis.text.y = element_text(size = 18),
          axis.title.x = element_text(size = 20),
          axis.title.y = element_text(size = 20),
          legend.position = "bottom",
          legend.title = element_text(size = 20),
          legend.text = element_text(size = 20),
          panel.grid.major = element_line(linewidth = 1.2, color = "grey90"),
          panel.grid.minor = element_line(linewidth = 0.8, color = "grey90"),
          panel.background = element_rect(fill = "grey", color = NA))
  ggsave(filename = paste(path, "Dotplot.svg", sep = ""), plot = o, device = "svg",
width = 15, height = 15)
  ggsave(filename = paste(path, "Dotplot.tiff", sep = ""), plot = o, device = "tiff",
width = 15, height = 15)
}

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
######################################Functions#######################################



################################Objects/Descriptions##################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

titleCTL_10vMHW2_10<-expression("CTL day 10 vs. MHW2 day 10")
titleCTL_25vMHW1_25<-expression("CTL day 25 vs. MHW1 day 25")
titleCTL_25vMHW2_25<-expression("CTL day 25 vs. MHW2 day 25")
titleMHW1_25vMHW2_25<-expression("MHW1 day 25 vs. MHW2 day 25")
titleMHW2_10vMHW2_25<-expression("MHW2 day 10 vs. MHW2 day 25")

legend<-data.frame(CTL_10vMHW2_10 = c("Overexpressed in the control group in day 10",
"Overexpressed in the MHW2 group in day 10"),
```

```
                    CTL_25vMHW1_25 = c("Overexpressed in the control group in day 25",
"Overexpressed in the MHW1 group in day 25"),
                    CTL_25vMHW2_25 = c("Overexpressed in the control group in day
25","Overexpressed in MHW2 group in day 25"),
                    MHW1_25vMHW2_25 = c("Overexpressed in the MHW1 group in day
25","Overexpressed in MHW2 group in day 25"),
                    MHW2_10vMHW2_25 = c("Overexpressed in the MHW2 group in day
10","Overexpressed in MHW2 group in day 25"))
contrasts <-
c("CTL_10vMHW2_10","CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25","MHW2_10vMHW2_25"
)

colour<-data.frame(CTL_10vMHW2_10 = c("#FFA500", "#0000FF"),
                    CTL_25vMHW1_25 = c("#FF0000", "#006400"),
                    CTL_25vMHW2_25 = c("#B8860B", "#800080"),
                    MHW1_25vMHW2_25 = c("#654321", "#1E90FF"),
                    MHW2_10vMHW2_25 = c("#4B0082", "#696969"))

sampleName <- c(
  "T13MHW2Pm1 MD25(M)",
  "T7MHW2Pm1 MD25(M)",
  "T3MHW2Pm1 MD10(M)",
  "T6MHW1Pm1 MD25(M)",
  "T1CTLPm1 MD10(M)",
  "T10CTLPm1 MD10(M)",
  "T1CTLPm1 MD25(M)",
  "T4CTLPm1 MD25(M)",
  "T12MHW1Pm1 MD25(M)",
  "T2MHW1Pm1 MD25(M)",
  "T10CTLPm1 MD25(M)",
  "T15CTLPm1 MD10(M)",
  "T3MHW2Pm1 MD25(M)",
  "T7MHW2Pm1 MD10(M)",
  "T5MHW2Pm1 MD10(M)"
)
sampleID <- c("RNA_11_N3066",
              "RNA_13_N3067",
              "RNA_14_N3058",
              "RNA_17_N3064",
              "RNA_18_N3054",
              "RNA_19_N3055",
              "RNA_20_N3060",
              "RNA_21_N3061",
              "RNA_3_N3062",
              "RNA_4_N3063",
              "RNA_5_N3059",
              "RNA_6_N3053",
              "RNA_7_N3065",
              "RNA_8_N3056",
              "RNA_9_N3057")
Day <- c(
  "25",
  "25",
  "10",
  "25",
  "10",
  "10",
  "25",
  "25",
  "25",
  "25",
  "25",
  "10",
```

```
  "25",
  "10",
  "10"
)
Treatment <- c(
  "MHW2",
  "MHW2",
  "MHW2",
  "MHW1",
  "CTL",
  "CTL",
  "CTL",
  "CTL",
  "MHW1",
  "MHW1",
  "CTL",
  "CTL",
  "MHW2",
  "MHW2",
  "MHW2"
)
Levels <- c(
  "MHW2_25",
  "MHW2_25",
  "MHW2_10",
  "MHW1_25",
  "CTL_10",
  "CTL_10",
  "CTL_25",
  "CTL_25",
  "MHW1_25",
  "MHW1_25",
  "CTL_25",
  "CTL_10",
  "MHW2_25",
  "MHW2_10",
  "MHW2_10"
)

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#################################Objects/Descriptions#################################



######################################Files/Data######################################
#||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

#Kallisto folder with the abundance files
Kallistofolder<-"C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Outputs\\Server\\Kallisto_50\\abundance\\"
########################################

#Trinity file
Trinityfile<-"C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Outputs\\Server\\Trinity\\Trinity.Trinity.fasta"
#############

#Blast with swissprot
Blast <- read.csv("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Outputs\\Server\\Blastp\\Blastp_for_analysis\\Pmicrops_Blastp_SwissProt.csv",sep
=";",header = F)
```

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

#Blast with only drerio matches
Blast_drerio <- read.csv("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Outputs\\Server\\Blast_Drerio\\Pmicrops_Blastp_Drerio.csv",sep =";",header = F)
############

#Transdecoder file
TransDecoder <- fread("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2°
Ano\\Outputs\\Server\\TransDecoder\\Trinity.Trinity.fasta.transdecoder.pep",header=FALSE
)
#################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#######################################Files/Data######################################



###################Merge the counts with the respective cDNA sequence################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Import the abundance results and estimate the counts
samples<-dir(Kallistofolder)
file<-c(paste(sep = "", Kallistofolder, samples, "\\", "abundance.h5"))
names(file)=sampleID
print(names(file))
Tsv<-tximport(file, type = "kallisto", txOut = TRUE, countsFromAbundance =
"lengthScaledTPM")
print(head(Tsv$counts))
print(dim(Tsv$counts))
print(colnames(Tsv$counts))
#####################################################

##Read the fasta file to get the ids and sequences
Fa<-read.fasta(Trinityfile, as.string = TRUE)
print(head(Fa))
TxID<-getName(Fa)
TxSEQ<-unlist(getSequence(Fa, as.string = TRUE))
TxSEQ<-as.data.frame(cbind(TxID, TxSEQ))
print(head(TxSEQ))
colnames(TxSEQ)[1]<-"ID"
print(nrow(TxSEQ))
Counts<-Tsv$counts
Counts<-cbind(as.data.frame(row.names(Counts)), Counts)
colnames(Counts)[1]<-"ID"
#################################################

##Create a matrix with  blast ids, sequence and counts
colnames(Blast) <- c("TrinityID","Accession", "%ID", "Aligment_Length", "Mismatches",
"Gap_Openings", "Start_of_Aligment_TrinityID", "End_of_Aligment_TrinityID",
"Start_of_Aligment_Accession", "End_of_Aligment_Accession", "Evalue",
"Bit_Score","Accession","-2","Description")
Blast <- subset(Blast, grepl("^TRINITY", Blast[,1]))
Blast$ID<-unlist(sapply(Blast$TrinityID, function(x) unlist(strsplit(x, ".", fixed =
TRUE))[1]))
blast_trinity_merge<-merge(Blast,Counts, by = "ID")
blast_trinity_merge <- blast_trinity_merge[, -c(2,4:17)]
blast_trinity_merge$Accession <- str_extract(blast_trinity_merge$Accession,
"(?<=\\|)[^|]+(?=\\|)")
Full_blast<-merge(blast_trinity_merge, TxSEQ, by="ID")
write.table(Full_blast, paste(pathBothOmicsPathways, "genes_ID.csv", sep = ""),sep =
```

```
";",col.names = TRUE, row.names = FALSE)
######################################################

##Create a matrix with  trinity ids, sequence and counts
Full_trinity <- merge(TxSEQ,Counts, by = "ID")
colnames(Full_trinity)<-c("ID", "sequence", sampleID)
########################################################

##Create a materix with blast_drerio ids, sequence and counts
colnames(Blast_drerio) <- c("TrinityID","Accession", "%ID", "Aligment_Length",
"Mismatches", "Gap_Openings", "Start_of_Aligment_TrinityID",
"End_of_Aligment_TrinityID", "Start_of_Aligment_Accession", "End_of_Aligment_Accession",
"Evalue", "Bit_Score","Accession","-2","Description")
Blast_drerio <- subset(Blast_drerio, grepl("^TRINITY", Blast_drerio[,1]))
Blast_drerio$ID<-unlist(sapply(Blast_drerio$TrinityID, function(x) unlist(strsplit(x,
".", fixed = TRUE))[1]))
blast_drerio_trinity_merge<-merge(Blast_drerio,Counts, by = "ID")
blast_drerio_trinity_merge <- blast_drerio_trinity_merge[, -c(1,2,4:17)]
sum(duplicated(blast_drerio_trinity_merge$Accession))

##Duplicate with higher mean
w <- c()
for (i in c(1:length(unique(blast_drerio_trinity_merge$Accession)))){
  dups <- which(blast_drerio_trinity_merge$Accession %in%
unique(blast_drerio_trinity_merge$Accession)[i])
  v <- c()
  for (j in dups){
    v <- c(v,mean(as.numeric(blast_drerio_trinity_merge[-c(1,2)][j,])))
  }
  w = c(w,dups[which(v == max(v))[1]])
}
blast_drerio_trinity_merge <- blast_drerio_trinity_merge[w,]
############################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
####################Merge the counts with the respective cDNA sequence################



################################Statistics (with trinity)##############################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create a DGEList object from a table of counts
Data_trinity <-
  DGEList(counts = Full_trinity[, 3:17],
          group = Levels,
          genes = Full_trinity[, 1])
################################################

##Filter out lowly expressed genes
keep_trinity <- filterByExpr(Data_trinity)
Data_trinity <- Data_trinity[keep_trinity, , keep.lib.sizes=FALSE]
##################################

##Calculate the normalization factors
Data_trinity <- calcNormFactors(Data_trinity)
#####################################

##Create the generalized linear models for comparing the expression levels from the
different treatments
Design_trinity <- model.matrix( ~ 0 + Levels, data = Data_trinity$samples)
```

```
colnames(Design_trinity) <- levels(Data_trinity$samples$group)
Data_trinity <- estimateDisp(Data_trinity, Design_trinity)
fit_trinity <- glmFit(Data_trinity, Design_trinity)
################################################################################
################

##Expression levels between the treatments
CTL_10vMHW2_10_trinity<-
relativeExpression(colnames(Design_trinity)[1],colnames(Design_trinity)[4],Design_trinit
y,fit_trinity)
MHW2_10vMHW2_25_trinity<-
relativeExpression(colnames(Design_trinity)[4],colnames(Design_trinity)[5],Design_trinit
y,fit_trinity)
CTL_25vMHW1_25_trinity<-
relativeExpression(colnames(Design_trinity)[2],colnames(Design_trinity)[3],Design_trinit
y,fit_trinity)
CTL_25vMHW2_25_trinity<-
relativeExpression(colnames(Design_trinity)[2],colnames(Design_trinity)[5],Design_trinit
y,fit_trinity)
MHW1_25vMHW2_25_trinity<-
relativeExpression(colnames(Design_trinity)[3],colnames(Design_trinity)[5],Design_trinit
y,fit_trinity)
########################################

##Merge the logFC,logCPM,LR and pvalue with the trinity ids
Results_trinity<-cbind(Data_trinity$genes, CTL_10vMHW2_10_trinity$table,
MHW2_10vMHW2_25_trinity$table, CTL_25vMHW1_25_trinity$table,
CTL_25vMHW2_25_trinity$table, MHW1_25vMHW2_25_trinity$table)
colnames(Results_trinity)<-c(
  "ID",
  "logFC-CTL_10vMHW2_10",
  "logCPM-CTL_10vMHW2_10",
  "LR-CTL_10vMHW2_10",
  "Pvalue-CTL_10vMHW2_10",
  "logFC-MHW2_10vMHW2_25",
  "logCPM-MHW2_10vMHW2_25",
  "LR-MHW2_10vMHW2_25",
  "Pvalue-MHW2_10vMHW2_25",
  "logFC-CTL_25vMHW1_25",
  "logCPM-CTL_25vMHW1_25",
  "LR-CTL_25vMHW1_25",
  "Pvalue-CTL_25vMHW1_25",
  "logFC-CTL_25vMHW2_25",
  "logCPM-CTL_25vMHW2_25",
  "LR-CTL_25vMHW2_25",
  "Pvalue-CTL_25vMHW2_25",
  "logFC-MHW1_25vMHW2_25",
  "logCPM-MHW1_25vMHW2_25",
  "LR-MHW1_25vMHW2_25",
  "Pvalue-MHW1_25vMHW2_25"
)
head(Results_trinity)
write.table(Results_trinity,paste(pathFiles_Trinity,"Results.csv",sep =""),
sep=";",col.names=NA)
#####################################

##Calculating the n° of DEGs in general
expressionTable_trinity<-decideTests(Results_trinity[,grepl("Pvalue",
colnames(Results_trinity))],coefficients = Results_trinity[,grepl("logFC",
colnames(Results_trinity))], adjust.method = "fdr", lfc = 1.5)
expressionTable_trinity<-as.data.frame(expressionTable_trinity)
head(expressionTable_trinity)
colnames(expressionTable_trinity)<-c("CTL_10vMHW2_10", "MHW2_10vMHW2_25",
```

```
"CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25")
DEG_trinity<-cbind(Results_trinity,expressionTable_trinity)
degTable_trinity<-DEG_trinity[which(abs(DEG_trinity$CTL_10vMHW2_10) == 1 |
abs(DEG_trinity$MHW2_10vMHW2_25) == 1 | abs(DEG_trinity$CTL_25vMHW1_25) == 1 |
abs(DEG_trinity$CTL_25vMHW2_25) == 1 | abs(DEG_trinity$MHW1_25vMHW2_25) == 1),]
head(degTable_trinity)
nrow(degTable_trinity)
write.table(degTable_trinity,paste(pathFiles_Trinity,"DEG.csv",sep
=""),sep=";",col.names=NA)
save(CTL_10vMHW2_10_trinity, MHW2_10vMHW2_25_trinity, CTL_25vMHW1_25_trinity,
CTL_25vMHW2_25_trinity, MHW1_25vMHW2_25_trinity, file = "Trinity\\Files\\lrt.RData")
#######################################

##N° of DEGs per treatment
DEG_CTL_10vMHW2_10_trinity<-sum(degTable_trinity$CTL_10vMHW2_10 !=
0);DEG_CTL_10vMHW2_10_trinity
DEG_MHW2_10vMHW2_25_trinity<-sum(degTable_trinity$MHW2_10vMHW2_25 !=
0);DEG_MHW2_10vMHW2_25_trinity
DEG_CTL_25vMHW2_25_trinity<-sum(degTable_trinity$CTL_25vMHW2_25 !=
0);DEG_CTL_25vMHW2_25_trinity
DEG_CTL_25vMHW1_25_trinity<-sum(degTable_trinity$CTL_25vMHW1_25 !=
0);DEG_CTL_25vMHW1_25_trinity
DEG_MHW1_25vMHW2_25_trinity<-sum(degTable_trinity$MHW1_25vMHW2_25 !=
0);DEG_MHW1_25vMHW2_25_trinity
###########################

##N° of overexpressed genes in treatment 2
oedeg_CTL_10vMHW2_10_trinity<-sum(degTable_trinity$CTL_10vMHW2_10 == -
1);oedeg_CTL_10vMHW2_10_trinity
oedeg_MHW2_10vMHW2_25_trinity<-sum(degTable_trinity$MHW2_10vMHW2_25 == -
1);oedeg_MHW2_10vMHW2_25_trinity
oedeg_CTL_25vMHW2_25_trinity<-sum(degTable_trinity$CTL_25vMHW2_25 == -
1);oedeg_CTL_25vMHW2_25_trinity
oedeg_CTL_25vMHW1_25_trinity<-sum(degTable_trinity$CTL_25vMHW1_25 == -
1);oedeg_CTL_25vMHW1_25_trinity
oedeg_MHW1_25vMHW2_25_trinity<-sum(degTable_trinity$MHW1_25vMHW2_25 == -
1);oedeg_MHW1_25vMHW2_25_trinity
############################################

##N° of underexpressed genes in treatment
uedeg_CTL_10vMHW2_10_trinity<-sum(degTable_trinity$CTL_10vMHW2_10 ==
1);uedeg_CTL_10vMHW2_10_trinity
uedeg_MHW2_10vMHW2_25_trinity<-sum(degTable_trinity$MHW2_10vMHW2_25 ==
1);uedeg_MHW2_10vMHW2_25_trinity
uedeg_CTL_25vMHW2_25_trinity<-sum(degTable_trinity$CTL_25vMHW2_25 ==
1);uedeg_CTL_25vMHW2_25_trinity
uedeg_CTL_25vMHW1_25_trinity<-sum(degTable_trinity$CTL_25vMHW1_25 ==
1);uedeg_CTL_25vMHW1_25_trinity
uedeg_MHW1_25vMHW2_25_trinity<-sum(degTable_trinity$MHW1_25vMHW2_25 ==
1);uedeg_MHW1_25vMHW2_25_trinity
############################################

##Table with total differential expressed transcripts, underexpressed transcripts and
overexpressed transcripts per treatment
deg_per_treatment_trinity <- data.frame(Treatments = contrasts,
                                No_of_Genes =
c(DEG_CTL_10vMHW2_10_trinity,DEG_CTL_25vMHW1_25_trinity,DEG_CTL_25vMHW2_25_trinity,DEG_M
HW1_25vMHW2_25_trinity,DEG_MHW2_10vMHW2_25_trinity),
                                No_of_underexpressed_Genes =
c(uedeg_CTL_10vMHW2_10_trinity,uedeg_CTL_25vMHW1_25_trinity,uedeg_CTL_25vMHW2_25_trinity
,uedeg_MHW1_25vMHW2_25_trinity,uedeg_MHW2_10vMHW2_25_trinity),
                                No_of_overexpressed_Genes =
c(oedeg_CTL_10vMHW2_10_trinity,oedeg_CTL_25vMHW1_25_trinity,oedeg_CTL_25vMHW2_25_trinity
```

```
,oedeg_MHW1_25vMHW2_25_trinity,oedeg_MHW2_10vMHW2_25_trinity))
colnames(deg_per_treatment_trinity) <- c("Treatments","transcripts with differential
expression","Overexpressed in treatment 2","Underexpressed in treatment 2")
write.table(deg_per_treatment_trinity,paste(pathTables_Trinity,"DEG_per_treatment_trinit
y.csv",sep =""),sep=";",col.names=NA)
########################################################################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
###############################Statistics (with trinity)###############################



###############################Statistics (with blast)################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create a DGEList object from a table of counts
Data_blast <-
  DGEList(counts = Full_blast[, 3:17],
          group = Levels,
          genes = Full_blast[, 2])
################################################

##Filter out lowly expressed genes
keep_blast <- filterByExpr(Data_blast)
Data_blast <- Data_blast[keep_blast, , keep.lib.sizes=FALSE]
##################################

##Calculate the normalization factors
Data_blast <- calcNormFactors(Data_blast)
#####################################

##Create the generalized linear models for comparing the expression levels from the
different treatments
Design_blast <- model.matrix( ~ 0 + Levels, data = Data_blast$samples)
colnames(Design_blast) <- levels(Data_blast$samples$group)
Data_blast <- estimateDisp(Data_blast, Design_blast)
fit_blast <- glmFit(Data_blast, Design_blast)
########################################################################################
################

##Expression levels between the treatments
CTL_10vMHW2_10_blast<-
relativeExpression(colnames(Design_blast)[1],colnames(Design_blast)[4],Design_blast,fit_
blast)
MHW2_10vMHW2_25_blast<-
relativeExpression(colnames(Design_blast)[4],colnames(Design_blast)[5],Design_blast,fit_
blast)
CTL_25vMHW1_25_blast<-
relativeExpression(colnames(Design_blast)[2],colnames(Design_blast)[3],Design_blast,fit_
blast)
CTL_25vMHW2_25_blast<-
relativeExpression(colnames(Design_blast)[2],colnames(Design_blast)[5],Design_blast,fit_
blast)
MHW1_25vMHW2_25_blast<-
relativeExpression(colnames(Design_blast)[3],colnames(Design_blast)[5],Design_blast,fit_
blast)
##########################################

##Merge the logFC,logCPM,LR and pvalue with the trinity ids
Results_blast<-cbind(Data_blast$genes, CTL_10vMHW2_10_blast$table,
MHW2_10vMHW2_25_blast$table, CTL_25vMHW1_25_blast$table, CTL_25vMHW2_25_blast$table,
```

```
MHW1_25vMHW2_25_blast$table)
colnames(Results_blast)<-c(
  "ID",
  "logFC-CTL_10vMHW2_10",
  "logCPM-CTL_10vMHW2_10",
  "LR-CTL_10vMHW2_10",
  "Pvalue-CTL_10vMHW2_10",
  "logFC-MHW2_10vMHW2_25",
  "logCPM-MHW2_10vMHW2_25",
  "LR-MHW2_10vMHW2_25",
  "Pvalue-MHW2_10vMHW2_25",
  "logFC-CTL_25vMHW1_25",
  "logCPM-CTL_25vMHW1_25",
  "LR-CTL_25vMHW1_25",
  "Pvalue-CTL_25vMHW1_25",
  "logFC-CTL_25vMHW2_25",
  "logCPM-CTL_25vMHW2_25",
  "LR-CTL_25vMHW2_25",
  "Pvalue-CTL_25vMHW2_25",
  "logFC-MHW1_25vMHW2_25",
  "logCPM-MHW1_25vMHW2_25",
  "LR-MHW1_25vMHW2_25",
  "Pvalue-MHW1_25vMHW2_25"
)
head(Results_blast)
write.table(Results_blast,paste(pathFiles_BSwiss,"Results.csv",sep =""),
sep=";",col.names=NA)
#######################################

##Calculating the n° of DEGs in general
expressionTable_blast<-decideTests(Results_blast[,grepl("Pvalue",
colnames(Results_blast))],coefficients = Results_blast[,grepl("logFC",
colnames(Results_blast))], adjust.method = "fdr", lfc = 1.5)
expressionTable_blast<-as.data.frame(expressionTable_blast)
head(expressionTable_blast)
colnames(expressionTable_blast)<-c("CTL_10vMHW2_10", "MHW2_10vMHW2_25",
"CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25")
DEG_blast<-cbind(Results_blast,expressionTable_blast)
degTable_blast<-DEG_blast[which(abs(DEG_blast$CTL_10vMHW2_10) == 1 |
abs(DEG_blast$MHW2_10vMHW2_25) == 1 | abs(DEG_blast$CTL_25vMHW1_25) == 1 |
abs(DEG_blast$CTL_25vMHW2_25) == 1 | abs(DEG_blast$MHW1_25vMHW2_25) == 1),]
head(degTable_blast)
nrow(degTable_blast)
write.table(degTable_blast,paste(pathBothOmicsPathways,"DEG.csv",sep
=""),sep=";",col.names=NA)
save(CTL_10vMHW2_10_blast, MHW2_10vMHW2_25_blast, CTL_25vMHW1_25_blast,
CTL_25vMHW2_25_blast, MHW1_25vMHW2_25_blast, file = "Blast_SwissProt\\Files\\lrt.RData")
#######################################

##N° of DEGs per treatment
DEG_CTL_10vMHW2_10_blast<-sum(degTable_blast$CTL_10vMHW2_10 !=
0);DEG_CTL_10vMHW2_10_blast
DEG_MHW2_10vMHW2_25_blast<-sum(degTable_blast$MHW2_10vMHW2_25 !=
0);DEG_MHW2_10vMHW2_25_blast
DEG_CTL_25vMHW2_25_blast<-sum(degTable_blast$CTL_25vMHW2_25 !=
0);DEG_CTL_25vMHW2_25_blast
DEG_CTL_25vMHW1_25_blast<-sum(degTable_blast$CTL_25vMHW1_25 !=
0);DEG_CTL_25vMHW1_25_blast
DEG_MHW1_25vMHW2_25_blast<-sum(degTable_blast$MHW1_25vMHW2_25 !=
0);DEG_MHW1_25vMHW2_25_blast
#########################

##N° of overexpressed ORFs in treatment 2
```

```
oedeg_CTL_10vMHW2_10_blast<-sum(degTable_blast$CTL_10vMHW2_10 == -
1);oedeg_CTL_10vMHW2_10_blast
oedeg_MHW2_10vMHW2_25_blast<-sum(degTable_blast$MHW2_10vMHW2_25 == -
1);oedeg_MHW2_10vMHW2_25_blast
oedeg_CTL_25vMHW2_25_blast<-sum(degTable_blast$CTL_25vMHW2_25 == -
1);oedeg_CTL_25vMHW2_25_blast
oedeg_CTL_25vMHW1_25_blast<-sum(degTable_blast$CTL_25vMHW1_25 == -
1);oedeg_CTL_25vMHW1_25_blast
oedeg_MHW1_25vMHW2_25_blast<-sum(degTable_blast$MHW1_25vMHW2_25 == -
1);oedeg_MHW1_25vMHW2_25_blast
##############################################

##N° of underexpressed ORFs in treatment 2
uedeg_CTL_10vMHW2_10_blast<-sum(degTable_blast$CTL_10vMHW2_10 ==
1);uedeg_CTL_10vMHW2_10_blast
uedeg_MHW2_10vMHW2_25_blast<-sum(degTable_blast$MHW2_10vMHW2_25 ==
1);uedeg_MHW2_10vMHW2_25_blast
uedeg_CTL_25vMHW2_25_blast<-sum(degTable_blast$CTL_25vMHW2_25 ==
1);uedeg_CTL_25vMHW2_25_blast
uedeg_CTL_25vMHW1_25_blast<-sum(degTable_blast$CTL_25vMHW1_25 ==
1);uedeg_CTL_25vMHW1_25_blast
uedeg_MHW1_25vMHW2_25_blast<-sum(degTable_blast$MHW1_25vMHW2_25 ==
1);uedeg_MHW1_25vMHW2_25_blast
##############################################

##Table with total diferantial expressed ORFs, underexpressed ORFs and overexpressed
ORFs per treatment
deg_per_treatment_blast <- data.frame(Treatments = contrasts,
                                      No_of_Genes =
c(DEG_CTL_10vMHW2_10_blast,DEG_CTL_25vMHW1_25_blast,DEG_CTL_25vMHW2_25_blast,DEG_MHW1_25
vMHW2_25_blast,DEG_MHW2_10vMHW2_25_blast),
                                      No_of_overexpressed_Genes =
c(oedeg_CTL_10vMHW2_10_blast,oedeg_CTL_25vMHW1_25_blast,oedeg_CTL_25vMHW2_25_blast,oedeg
_MHW1_25vMHW2_25_blast,oedeg_MHW2_10vMHW2_25_blast),
                                      No_of_underexpressed_Genes =
c(uedeg_CTL_10vMHW2_10_blast,uedeg_CTL_25vMHW1_25_blast,uedeg_CTL_25vMHW2_25_blast,uedeg
_MHW1_25vMHW2_25_blast,uedeg_MHW2_10vMHW2_25_blast))
colnames(deg_per_treatment_blast) <- c("Treatments","transcripts with differential
expression","Overexpressed in treatment 2","Underexpressed in treatment 2")
write.table(deg_per_treatment_blast,paste(pathTables_BSwiss,"DEG_per_treatment_blast.csv
",sep =""),sep=";",col.names=NA)
##############################################################################################

##Table with n° of transcripts, n° of ORFs, n° of differential expressed transcripts and
n° of ORFs with differential expression
transdecoder_n_transcripts <- grepl("^>", TransDecoder$V1)
transdecoder_filtered <- subset(TransDecoder, transdecoder_n_transcripts)
p_ORFs <- round((nrow(transdecoder_filtered)*100)/nrow(TxSEQ),3)
p_annotated_ORFs <- round((nrow(Blast)*100)/nrow(TxSEQ),3)
p_DETranscripts <- round((nrow(degTable_trinity)*100)/nrow(TxSEQ),3)
p_DEORFs <- round((nrow(degTable_blast)*100)/nrow(TxSEQ),3)
percentages <- c(100,p_ORFs,p_annotated_ORFs,p_DETranscripts,p_DEORFs)
percentages <- paste0(percentages, "%")
transcripts_of_interest <- data.frame(Stage = c(1,2,3,4,5),
                                      Objective = c("Transcriptome
assembly","Transcripts with coding regions","Functional annotation","Differentially-
Expressed Transcripts","ORFs with Differential Expression"),
                                      Specie =
c(nrow(TxSEQ),nrow(transdecoder_filtered),nrow(Blast),nrow(degTable_trinity),nrow(degTab
le_blast)),
                                      Percentagens = percentages)
colnames(transcripts_of_interest) <- c("Analytical stage","Stage
objective","Transcripts","Percentages")
```

```
write.table(transcripts_of_interest,paste(pathTables_BSwiss,"transcripts_of_interest.csv
",sep =""),sep=";",col.names=NA)
########################################################################################
########################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
################################Statistics (with blast)################################



#############################Statistics (with blast_drerio)############################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create a DGEList object from a table of counts
Data_blast_drerio <-
  DGEList(counts = blast_drerio_trinity_merge[, 2:16],
          group = Levels,
          genes = blast_drerio_trinity_merge[, 1])
################################################

##Filter out lowly expressed genes
keep_blast_drerio <- filterByExpr(Data_blast_drerio)
Data_blast_drerio <- Data_blast_drerio[keep_blast_drerio, , keep.lib.sizes=FALSE]
##################################

##Calculate the normalization factors
Data_blast_drerio <- calcNormFactors(Data_blast_drerio)
#####################################

##Create the generalized linear models for comparing the expression levels from the
different treatments
Design_blast_drerio <- model.matrix( ~ 0 + Levels, data = Data_blast_drerio$samples)
colnames(Design_blast_drerio) <- levels(Data_blast_drerio$samples$group)
Data_blast_drerio <- estimateDisp(Data_blast_drerio, Design_blast_drerio)
fit_blast_drerio <- glmFit(Data_blast_drerio, Design_blast_drerio)
########################################################################################
################

##Expression levels between the treatments
CTL_10vMHW2_10_blast_drerio<-
relativeExpression(colnames(Design_blast_drerio)[1],colnames(Design_blast_drerio)[4],Des
ign_blast_drerio,fit_blast_drerio)
MHW2_10vMHW2_25_blast_drerio<-
relativeExpression(colnames(Design_blast_drerio)[4],colnames(Design_blast_drerio)[5],Des
ign_blast_drerio,fit_blast_drerio)
CTL_25vMHW1_25_blast_drerio<-
relativeExpression(colnames(Design_blast_drerio)[2],colnames(Design_blast_drerio)[3],Des
ign_blast_drerio,fit_blast_drerio)
CTL_25vMHW2_25_blast_drerio<-
relativeExpression(colnames(Design_blast_drerio)[2],colnames(Design_blast_drerio)[5],Des
ign_blast_drerio,fit_blast_drerio)
MHW1_25vMHW2_25_blast_drerio<-
relativeExpression(colnames(Design_blast_drerio)[3],colnames(Design_blast_drerio)[5],Des
ign_blast_drerio,fit_blast_drerio)
##########################################

##Merge the logFC,logCPM,LR and pvalue with the trinity ids
Results_blast_drerio<-cbind(Data_blast_drerio$genes, CTL_10vMHW2_10_blast_drerio$table,
MHW2_10vMHW2_25_blast_drerio$table, CTL_25vMHW1_25_blast_drerio$table,
CTL_25vMHW2_25_blast_drerio$table, MHW1_25vMHW2_25_blast_drerio$table)
colnames(Results_blast_drerio)<-c(
```

```
  "ID",
  "logFC-CTL_10vMHW2_10",
  "logCPM-CTL_10vMHW2_10",
  "LR-CTL_10vMHW2_10",
  "Pvalue-CTL_10vMHW2_10",
  "logFC-MHW2_10vMHW2_25",
  "logCPM-MHW2_10vMHW2_25",
  "LR-MHW2_10vMHW2_25",
  "Pvalue-MHW2_10vMHW2_25",
  "logFC-CTL_25vMHW1_25",
  "logCPM-CTL_25vMHW1_25",
  "LR-CTL_25vMHW1_25",
  "Pvalue-CTL_25vMHW1_25",
  "logFC-CTL_25vMHW2_25",
  "logCPM-CTL_25vMHW2_25",
  "LR-CTL_25vMHW2_25",
  "Pvalue-CTL_25vMHW2_25",
  "logFC-MHW1_25vMHW2_25",
  "logCPM-MHW1_25vMHW2_25",
  "LR-MHW1_25vMHW2_25",
  "Pvalue-MHW1_25vMHW2_25"
)
head(Results_blast_drerio)
write.table(Results_blast_drerio,paste(pathFiles_BDrerio,"Results.csv",sep =""),
sep=";",col.names=NA)
#######################################

##Calculating the n° of DEGs in general
expressionTable_blast_drerio<-decideTests(Results_blast_drerio[,grepl("Pvalue",
colnames(Results_blast_drerio))],coefficients = Results_blast_drerio[,grepl("logFC",
colnames(Results_blast_drerio))], adjust.method = "fdr", lfc = 1.5)
expressionTable_blast_drerio<-as.data.frame(expressionTable_blast_drerio)
head(expressionTable_blast_drerio)
colnames(expressionTable_blast_drerio)<-c("CTL_10vMHW2_10", "MHW2_10vMHW2_25",
"CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25")
DEG_blast_drerio<-cbind(Results_blast_drerio,expressionTable_blast_drerio)
degTable_blast_drerio<-DEG_blast_drerio[which(abs(DEG_blast_drerio$CTL_10vMHW2_10) == 1
| abs(DEG_blast_drerio$MHW2_10vMHW2_25) == 1 | abs(DEG_blast_drerio$CTL_25vMHW1_25) == 1
| abs(DEG_blast_drerio$CTL_25vMHW2_25) == 1 | abs(DEG_blast_drerio$MHW1_25vMHW2_25) ==
1),]
head(degTable_blast_drerio)
nrow(degTable_blast_drerio)
write.table(degTable_blast_drerio,paste(pathFiles_BDrerio,"DEG.csv",sep
=""),sep=";",col.names=NA)
save(CTL_10vMHW2_10_blast_drerio, MHW2_10vMHW2_25_blast_drerio,
CTL_25vMHW1_25_blast_drerio, CTL_25vMHW2_25_blast_drerio, MHW1_25vMHW2_25_blast_drerio,
file = "Blast_Drerio\\Files\\lrt.RData")
#######################################

##N° of DEGs per treatment
DEG_drerio_CTL_10vMHW2_10<-sum(degTable_blast_drerio$CTL_10vMHW2_10 !=
0);DEG_drerio_CTL_10vMHW2_10
DEG_drerio_MHW2_10vMHW2_25<-sum(degTable_blast_drerio$MHW2_10vMHW2_25 !=
0);DEG_drerio_MHW2_10vMHW2_25
DEG_drerio_CTL_25vMHW2_25<-sum(degTable_blast_drerio$CTL_25vMHW2_25 !=
0);DEG_drerio_CTL_25vMHW2_25
DEG_drerio_CTL_25vMHW1_25<-sum(degTable_blast_drerio$CTL_25vMHW1_25 !=
0);DEG_drerio_CTL_25vMHW1_25
DEG_drerio_MHW1_25vMHW2_25<-sum(degTable_blast_drerio$MHW1_25vMHW2_25 !=
0);DEG_drerio_MHW1_25vMHW2_25
#########################

##N° of underexpressed genes per treatment
```

```
uedeg_drerio_CTL_10vMHW2_10<-sum(degTable_blast_drerio$CTL_10vMHW2_10 == -
1);uedeg_drerio_CTL_10vMHW2_10
uedeg_drerio_MHW2_10vMHW2_25<-sum(degTable_blast_drerio$MHW2_10vMHW2_25 == -
1);uedeg_drerio_MHW2_10vMHW2_25
uedeg_drerio_CTL_25vMHW2_25<-sum(degTable_blast_drerio$CTL_25vMHW2_25 == -
1);uedeg_drerio_CTL_25vMHW2_25
uedeg_drerio_CTL_25vMHW1_25<-sum(degTable_blast_drerio$CTL_25vMHW1_25 == -
1);uedeg_drerio_CTL_25vMHW1_25
uedeg_drerio_MHW1_25vMHW2_25<-sum(degTable_blast_drerio$MHW1_25vMHW2_25 == -
1);uedeg_drerio_MHW1_25vMHW2_25
#############################################

##N° of overexpressed genes per treatment
oedeg_drerio_CTL_10vMHW2_10<-sum(degTable_blast_drerio$CTL_10vMHW2_10 ==
1);oedeg_drerio_CTL_10vMHW2_10
oedeg_drerio_MHW2_10vMHW2_25<-sum(degTable_blast_drerio$MHW2_10vMHW2_25 ==
1);oedeg_drerio_MHW2_10vMHW2_25
oedeg_drerio_CTL_25vMHW2_25<-sum(degTable_blast_drerio$CTL_25vMHW2_25 ==
1);oedeg_drerio_CTL_25vMHW2_25
oedeg_drerio_CTL_25vMHW1_25<-sum(degTable_blast_drerio$CTL_25vMHW1_25 ==
1);oedeg_drerio_CTL_25vMHW1_25
oedeg_drerio_MHW1_25vMHW2_25<-sum(degTable_blast_drerio$MHW1_25vMHW2_25 ==
1);oedeg_drerio_MHW1_25vMHW2_25
############################################

##Table with total DEG, underexpressed genes and overexpressed genes per treatment
deg_drerio_per_treatment <- data.frame(Treatments = contrasts,
                                No_of_Genes =
c(DEG_drerio_CTL_10vMHW2_10,DEG_drerio_CTL_25vMHW1_25,DEG_drerio_CTL_25vMHW2_25,DEG_drer
io_MHW1_25vMHW2_25,DEG_drerio_MHW2_10vMHW2_25),
                                No_of_underexpressed_Genes =
c(uedeg_drerio_CTL_10vMHW2_10,uedeg_drerio_CTL_25vMHW1_25,uedeg_drerio_CTL_25vMHW2_25,ue
deg_drerio_MHW1_25vMHW2_25,uedeg_drerio_MHW2_10vMHW2_25),
                                No_of_overexpressed_Genes =
c(oedeg_drerio_CTL_10vMHW2_10,oedeg_drerio_CTL_25vMHW1_25,oedeg_drerio_CTL_25vMHW2_25,oe
deg_drerio_MHW1_25vMHW2_25,oedeg_drerio_MHW2_10vMHW2_25))
colnames(deg_drerio_per_treatment) <- c("Treatments","N° of proteins","N° of
underexpressed proteins","N° of overexpressed proteins")
write.table(deg_drerio_per_treatment,paste(pathTables_BSwiss,"DEG_per_treatment.csv",sep
=""),sep=";",col.names=NA)
#####################################################################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
############################Statistics (with blast_drerio)############################



###################################Pathway analysis##################################
#||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

DEG_pathway <- DEG_blast_drerio[-c(22:26)]
DEG_pathway$ID <- str_extract(DEG_pathway$ID, "\\|([^|]+)\\|") %>%
  str_replace_all("\\|", "")
write.table(DEG_pathway, paste(pathBothOmicsPathways, "DEG_pathway.csv", sep = ""),sep =
";",col.names = TRUE, row.names = FALSE)

##Create data frames with Accession, logFC and Pvalue for the pathway analysis
allg_CTL_10vMHW2_10<- data.frame(DEG_pathway$ID,DEG_pathway$`logFC-
CTL_10vMHW2_10`,DEG_pathway$`Pvalue-CTL_10vMHW2_10`)
names(allg_CTL_10vMHW2_10) <- c("Accession","logFC","Pvalue")
```

```
allg_CTL_25vMHW1_25<- data.frame(DEG_pathway$ID,DEG_pathway$`logFC-
CTL_25vMHW1_25`,DEG_pathway$`Pvalue-CTL_25vMHW1_25`)
names(allg_CTL_25vMHW1_25) <- c("Accession","logFC","Pvalue")

allg_CTL_25vMHW2_25<- data.frame(DEG_pathway$ID,DEG_pathway$`logFC-
CTL_25vMHW2_25`,DEG_pathway$`Pvalue-CTL_25vMHW2_25`)
names(allg_CTL_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allg_MHW1_25vMHW2_25<- data.frame(DEG_pathway$ID,DEG_pathway$`logFC-
MHW1_25vMHW2_25`,DEG_pathway$`Pvalue-MHW1_25vMHW2_25`)
names(allg_MHW1_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allg_MHW2_10vMHW2_25<- data.frame(DEG_pathway$ID,DEG_pathway$`logFC-
MHW2_10vMHW2_25`,DEG_pathway$`Pvalue-MHW2_10vMHW2_25`)
names(allg_MHW2_10vMHW2_25) <- c("Accession","logFC","Pvalue")
##########################################################################

##Create a data structure
layers <- c("transcriptome")
odataCTL_10vMHW2_10 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW1_25 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW1_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW2_10vMHW2_25 <- initOmicsDataStructure(layer=layers)
#########################

##Add transcriptome layer
odataCTL_10vMHW2_10$transcriptome <-
rankFeatures(allg_CTL_10vMHW2_10$logFC,allg_CTL_10vMHW2_10$Pvalue)
names(odataCTL_10vMHW2_10$transcriptome) <- allg_CTL_10vMHW2_10$Accession
odataCTL_10vMHW2_10$transcriptome <- sort(odataCTL_10vMHW2_10$transcriptome)
head(odataCTL_10vMHW2_10$transcriptome)

odataCTL_25vMHW1_25$transcriptome <-
rankFeatures(allg_CTL_25vMHW1_25$logFC,allg_CTL_25vMHW1_25$Pvalue)
names(odataCTL_25vMHW1_25$transcriptome) <- allg_CTL_25vMHW1_25$Accession
odataCTL_25vMHW1_25$transcriptome <- sort(odataCTL_25vMHW1_25$transcriptome)
head(odataCTL_25vMHW1_25$transcriptome)

odataCTL_25vMHW2_25$transcriptome <-
rankFeatures(allg_CTL_25vMHW2_25$logFC,allg_CTL_25vMHW2_25$Pvalue)
names(odataCTL_25vMHW2_25$transcriptome) <- allg_CTL_25vMHW2_25$Accession
odataCTL_25vMHW2_25$transcriptome <- sort(odataCTL_25vMHW2_25$transcriptome)
head(odataCTL_25vMHW2_25$transcriptome)

odataMHW1_25vMHW2_25$transcriptome <-
rankFeatures(allg_MHW1_25vMHW2_25$logFC,allg_MHW1_25vMHW2_25$Pvalue)
names(odataMHW1_25vMHW2_25$transcriptome) <- allg_MHW1_25vMHW2_25$Accession
odataMHW1_25vMHW2_25$transcriptome <- sort(odataMHW1_25vMHW2_25$transcriptome)
head(odataMHW1_25vMHW2_25$transcriptome)

odataMHW2_10vMHW2_25$transcriptome <-
rankFeatures(allg_MHW2_10vMHW2_25$logFC,allg_MHW2_10vMHW2_25$Pvalue)
names(odataMHW2_10vMHW2_25$transcriptome) <- allg_MHW2_10vMHW2_25$Accession
odataMHW2_10vMHW2_25$transcriptome <- sort(odataMHW2_10vMHW2_25$transcriptome)
head(odataMHW2_10vMHW2_25$transcriptome)
####################

##Select the databases we want to query and download pathway definitions
databases <- c("kegg")
pathways <- getMultiOmicsFeatures(dbs = databases, layer = layers,
                                  returnTranscriptome = "UNIPROT",
                                  organism = "drerio",
```

```
                                    useLocal =  FALSE)
pathways_short <- lapply(names(pathways), function(name) {
  head(pathways[[name]], 2)
})
names(pathways_short) <- names(pathways)
pathways_short
pathways$transcriptome[8]
#########################################################################

##Run the pathway enrichment
enrichment_scoresCTL_10vMHW2_10 <- multiGSEA(pathways,odataCTL_10vMHW2_10)
Tenrichment_scoresCTL_10vMHW2_10 <-
as.data.frame(enrichment_scoresCTL_10vMHW2_10$transcriptome)
Tenrichment_scoresCTL_10vMHW2_10$leadingEdge <-
sapply(Tenrichment_scoresCTL_10vMHW2_10$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_10vMHW2_10 <-
Tenrichment_scoresCTL_10vMHW2_10[order(Tenrichment_scoresCTL_10vMHW2_10$padj),]
Tenrichment_scoresCTL_10vMHW2_10$contrast <- rep("CTL_10vMHW2_10")
top_10_esCTL_10vMHW2_10 <- head(Tenrichment_scoresCTL_10vMHW2_10,10)
write.table(top_10_esCTL_10vMHW2_10,paste(pathTables_BDreio,"top10_enrichment_scoresCTL_
10vMHW2_10.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW1_25 <- multiGSEA(pathways,odataCTL_25vMHW1_25)
Tenrichment_scoresCTL_25vMHW1_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW1_25$transcriptome)
Tenrichment_scoresCTL_25vMHW1_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW1_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW1_25 <-
Tenrichment_scoresCTL_25vMHW1_25[order(Tenrichment_scoresCTL_25vMHW1_25$padj),]
Tenrichment_scoresCTL_25vMHW1_25$contrast <- rep("CTL_25vMHW1_25")
top_10_esCTL_25vMHW1_25 <- head(Tenrichment_scoresCTL_25vMHW1_25,10)
write.table(top_10_esCTL_25vMHW1_25,paste(pathTables_BDreio,"top10_enrichment_scoresCTL_
25vMHW1_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW2_25 <- multiGSEA(pathways,odataCTL_25vMHW2_25)
Tenrichment_scoresCTL_25vMHW2_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW2_25$transcriptome)
Tenrichment_scoresCTL_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW2_25 <-
Tenrichment_scoresCTL_25vMHW2_25[order(Tenrichment_scoresCTL_25vMHW2_25$padj),]
Tenrichment_scoresCTL_25vMHW2_25$contrast <- rep("CTL_25vMHW2_25")
top_10_esCTL_25vMHW2_25 <- head(Tenrichment_scoresCTL_25vMHW2_25,10)
write.table(top_10_esCTL_25vMHW2_25,paste(pathTables_BDreio,"top10_enrichment_scoresCTL_
25vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW1_25vMHW2_25 <- multiGSEA(pathways,odataMHW1_25vMHW2_25)
Tenrichment_scoresMHW1_25vMHW2_25 <-
as.data.frame(enrichment_scoresMHW1_25vMHW2_25$transcriptome)
Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW1_25vMHW2_25 <-
Tenrichment_scoresMHW1_25vMHW2_25[order(Tenrichment_scoresMHW1_25vMHW2_25$padj),]
Tenrichment_scoresMHW1_25vMHW2_25$contrast <- rep("MHW1_25vMHW2_25")
top_10_esMHW1_25vMHW2_25 <- head(Tenrichment_scoresMHW1_25vMHW2_25,10)
write.table(top_10_esMHW1_25vMHW2_25,paste(pathTables_BDreio,"top10_enrichment_scoresMHW
1_25vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW2_10vMHW2_25 <- multiGSEA(pathways,odataMHW2_10vMHW2_25)
```

```
Tenrichment_scoresMHW2_10vMHW2_25 <-
as.data.frame(enrichment_scoresMHW2_10vMHW2_25$transcriptome)
Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW2_10vMHW2_25 <-
Tenrichment_scoresMHW2_10vMHW2_25[order(Tenrichment_scoresMHW2_10vMHW2_25$padj),]
Tenrichment_scoresMHW2_10vMHW2_25$contrast <- rep("MHW2_10vMHW2_25")
top_10_esMHW2_10vMHW2_25 <- head(Tenrichment_scoresMHW2_10vMHW2_25,10)
write.table(top_10_esMHW2_10vMHW2_25,paste(pathTables_BDreio,"top10_enrichment_scoresMHW
2_10vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)
############################

##Create dataset with all enrichment scores
combined_df <-
rbind(Tenrichment_scoresCTL_10vMHW2_10,Tenrichment_scoresCTL_25vMHW1_25,Tenrichment_scor
esCTL_25vMHW2_25,Tenrichment_scoresMHW1_25vMHW2_25,Tenrichment_scoresMHW2_10vMHW2_25)
combined_df_top10 <-
rbind(top_10_esCTL_10vMHW2_10,top_10_esCTL_25vMHW1_25,top_10_esCTL_25vMHW2_25,top_10_esM
HW1_25vMHW2_25,top_10_esMHW2_10vMHW2_25)
combined_df_top10$pathway <- sub("^\\(KEGG\\) ", "", combined_df_top10$pathway)
###########################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
####################################Pathway analysis###################################


##########################################Plots########################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Volcano Plot
par(mfrow = c(1,1), mar = c(5,5,2,2), family = "Arial")
for(i in 1:length(contrasts)){
  volcanoPlot(pathFiles_Trinity, colour[contrasts[i]], legend[contrasts[i]],
get(paste("title", contrasts[i], sep = "")), contrasts[i], i)
  dev.print(tiff, paste(pathVolcano_Trinity, "VolcanoPlot-", contrasts[i], ".tif", sep
=""), height = 15, width = 15,  units = 'cm', res=600)
}
##############

##Heatmap
heatmapGraph(100, 20,degTable_trinity,pathHeatmap_Trinity)
heatmapGraph(60,24,degTable_blast_drerio,pathHeatmap_BSwiss)
#########

##Pie charts
for(i in 1:length(contrasts)){
  pieChart(deg_per_treatment_blast,i,"ORFs",contrasts,colour)
  pieChart(deg_per_treatment_trinity,i,"transcripts",contrasts,colour)
}
############

#Dotplot
dotplot(combined_df_top10,pathDotplot)
########

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
##########################################Plots########################################
```

A.2.2 Proteomics R Script

Please refer to https://github.com/jdplopes/MScThesis_Proteomics for the complete R script.

```
##Master's Thesis
##João Lopes
set.seed(1)
setwd("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2° Ano\\Proteomics")
##Load a R package
library(forcats)
library(limma)
library(readxl)
library(edgeR)
library(seqinr)
library(tximport)
library(RColorBrewer)
library(gplots)
library(rhdf5)
library(ggplot2)
library(multiGSEA)
library(tidyverse)
library(readxl)
library(tidyr)
windowsFonts(Calibri = windowsFont("Arial"))

########################################Folders########################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create folders
dir.create("Files")
dir.create("Files/Tables")
dir.create("Plots")

dir.create("Plots/Volcano plots")
dir.create("Plots/Heatmap")
dir.create("Plots/Dotplot")


##Directories
pathData <- "Data/"
pathFiles <- "Files/"
pathTables <- "Files/Tables/"

pathVolcano <- "Plots/Volcano plots/"
pathHeatmap <- "Plots/Heatmap/"
pathDotplot <- "Plots/Dotplot/"

pathBothOmicsPathways <- "C:/Users/jdpl2/OneDrive/Ambiente de Trabalho/Mestrado/2°
Ano/Both Omics/"
#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
########################################Folders########################################



#######################################Functions#######################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

relativeExpression<-function(treatment1, treatment2, Design, fit){
```

```
  contrast<-paste(treatment1, "-", treatment2, sep = "")
  treatment1vtreatment2<-makeContrasts(contrasts = contrast, levels = Design)
  print(treatment1vtreatment2)
  lrt<-glmLRT(fit, contrast = treatment1vtreatment2)
  print(topTags(lrt))
  return(lrt)
}

volcanoPlot<-function(pathFiles, colour, legend, title, contrast, i){
  load(paste(pathFiles, "lrt.RData", sep = ""))
  print(colnames(DEP))
  signif<--log10(DEP[,paste("Pvalue-", contrast, sep = "")])
  plot(DEP[,paste("logFC-", contrast, sep = "")], signif, pch = ".", xlab =
expression(log[2] * FC), ylab = expression(log[10] * FDR), main = NULL,
       cex.lab = 1.5,
       panel.first={
         points(0, 0, pch = 16, cex = 1e6, col = "grey95")
         grid(col = "white", lty = 1)
       }
  )
  points(DEP[which(DEP[contrast] == 1), paste("logFC-", contrast, sep = "")], -
log10(DEP[which(DEP[contrast] == 1), paste("Pvalue-", contrast, sep = "")]), pch = 21,
cex = 1.5, col = "black",bg = colour[1,])
  points(DEP[which(DEP[contrast] == -1), paste("logFC-", contrast, sep = "")], -
log10(DEP[which(DEP[contrast] == -1), paste("Pvalue-", contrast, sep = "")]), pch = 21,
cex = 1.5, col = "black",bg = colour[2,])
  #legend("bottomleft", inset = c(-0.06,-0.5), xpd = TRUE, pch = 20, bty = "n",
         #col = c(colour[1,], colour[2,]),
         #legend = c(legend[1,], legend[2,]))
  mtext(side = 3,
        LETTERS[i],
        at =-19.6+(-i+1)*3,
        line = 2,
        cex = 1.5
  )
}

heatmapGraph<-function(heatData,colCutoff, rowCutoff){
  heatData<-as.matrix(depTable[,c(3,7,11,15,19)])
  rownames(heatData)<-depTable$Description
  head(heatData)
  clustFunction <- function(x) hclust(x, method="complete")
  distFunction <- function(x) dist(x,method="euclidean")
  #clusterint (sidebars)
  cCol<-colorRampPalette(brewer.pal(9, "Set1"))
  colFit<-clustFunction(distFunction(t(heatData)))
  cClusters<-cutree(colFit,h=colCutoff)
  cHeight<-length(unique(as.vector(cClusters)))
  colHeight = cCol(cHeight)
  cRow<-colorRampPalette(brewer.pal(9,"Set1"))
  rowFit<-clustFunction(distFunction(heatData))
  rClusters<-cutree(rowFit,h=rowCutoff)
  rHeight<-length(unique(as.vector(rClusters)));
  rowHeight = cRow(rHeight)
  xDimension<-25
  yDimension<-25
  windows(xDimension, yDimension)
  par(family="Arial")
  #Heatmap
  heatColour<-colorRampPalette(c("#fff04a", "black", "#c23e40"))(n = 100)
  p <- heatmap.2(heatData,
            hclust=clustFunction,
            distfun=distFunction,
```

```
            ColSideColors=colHeight[cClusters],
            RowSideColors=rowHeight[rClusters],
            density.info="none",
            col=heatColour,
            trace="none",
            scale="row",
            cexCol = 0.8,
            celRow = 1.5,
            lhei = c(1,5),
            lwid = c(1.5,5),
            offsetRow = 0,
            offsetCol = 0,
            srtCol = 45,
            key.title = NA,
            margins = c(8,21.75),
            labRow = rownames(heatData))
  recordPlot()
  dev.print(tiff, filename = paste(pathHeatmap, "Heatmap", ".tif", sep = ""), height =
15, width = 15, units = "cm", res = 600)
}

dotplot <- function (data,path) {
  o <- ggplot(data, aes(x = contrast, y = pathway)) +
    geom_point(aes(size = 1-padj, color = NES, fill = NES), shape = 21, stroke = 0.5) +
    scale_color_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_size_continuous(range = c(3, 15)) +
    labs(x = "Contrast", y = "Pathway",
         size = "1-p.adjust",
         color = "NES",
         fill = "NES") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 18),
          axis.text.y = element_text(size = 18),
          axis.title.x = element_text(size = 20),
          axis.title.y = element_text(size = 20),
          legend.position = "bottom",
          legend.title = element_text(size = 20),
          legend.text = element_text(size = 20),
          panel.grid.major = element_line(linewidth = 1.2, color = "grey90"),
          panel.grid.minor = element_line(linewidth = 0.8, color = "grey90"),
          panel.background = element_rect(fill = "grey", color = NA))
  ggsave(filename = paste(path, "Prot_Dotplot.svg", sep = ""), plot = o, device = "svg",
width = 15, height = 15)
  ggsave(filename = paste(path, "Prot_Dotplot.tiff", sep = ""), plot = o, device =
"tiff", width = 15, height = 15)
}

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#######################################Functions#######################################



##################################Objects/Descriptions##################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

titleCTL_10vMHW2_10<-expression("CTL day 10 vs. MHW2 day 10")
titleCTL_25vMHW1_25<-expression("CTL day 25 vs. MHW1 day 25")
titleCTL_25vMHW2_25<-expression("CTL day 25 vs. MHW2 day 25")
titleMHW1_25vMHW2_25<-expression("MHW1 day 25 vs. MHW2 day 25")
titleMHW2_10vMHW2_25<-expression("MHW2 day 10 vs. MHW2 day 25")
```

```
legend<-data.frame(CTL_10vMHW2_10 = c("Overexpressed in the control group in day 10",
"Overexpressed in the MHW2 group in day 10"),
                   CTL_25vMHW1_25 = c("Overexpressed in the control group in day 25",
"Overexpressed in the MHW1 group in day 25"),
                   CTL_25vMHW2_25 = c("Overexpressed in the control group in day
25","Overexpressed in MHW2 group in day 25"),
                   MHW1_25vMHW2_25 = c("Overexpressed in the MHW1 group in day
25","Overexpressed in MHW2 group in day 25"),
                   MHW2_10vMHW2_25 = c("Overexpressed in the MHW2 group in day
10","Overexpressed in MHW2 group in day 25"))

contrasts <-
c("CTL_10vMHW2_10","CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25","MHW2_10vMHW2_25"
)

colour<-data.frame(CTL_10vMHW2_10 = c("#0000FF", "#FFA500"),
                   CTL_25vMHW1_25 = c("#800080", "#FFFF00"),
                   CTL_25vMHW2_25 = c("#FF0000", "#00FFFF"),
                   MHW1_25vMHW2_25 = c("#8B4513", "#ADFF2F"),
                   MHW2_10vMHW2_25 = c("#009900", "#800000"))

Levels <- c(
  "CTL_10",
  "MHW2_10",
  "CTL_10",
  "MHW2_10",
  "CTL_10",
  "MHW2_10",
  "CTL_25",
  "MHW2_10",
  "CTL_25",
  "MHW1_25",
  "CTL_25",
  "MHW1_25",
  "MHW2_25",
  "MHW1_25",
  "MHW2_25",
  "MHW1_25",
  "MHW2_25",
  "MHW2_25"
)

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
##################################Objects/Descriptions#################################



#######################################Files/Data######################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##File with all hits
file <- read_xlsx(paste(pathData,"CPMSF_GPPF-CM-47_EACMPC1-
18_pomatochistus_microps_TMT18_HiRIEF_results.xlsx",sep = ""))
####################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#######################################Files/Data######################################
```

```
######Reading the file and creating a matrix with only the normalized abundances#######
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

#Create dataframe with UniProt Accession code, Protein name and normalized abundance of
each protein
abundance_matrix <- file[,c(6,7,26:43)]
##############################################################################################
############

#Remove NAs
abundance_matrix <- na.omit(abundance_matrix)
###########

#Extract the protein name from the description
for (i in 1:nrow(abundance_matrix)) {
  result <- unlist(strsplit(abundance_matrix$Description[i], "OS="))
  previous_part <- result[1]
  abundance_matrix$Description[i] <- previous_part
}
#############################################

#Column names
colnames(abundance_matrix) <- c("Accession",
                                "Protein",
                                "Abundances Normalized T1CTLPm1 MD10(M) - 126",
                                "Abundances Normalized T3MHW2Pm1 MD10(M) - 127_N",
                                "Abundances Normalized T10CTLPm1 MD10(M) - 127_C",
                                "Abundances Normalized T5MHW2Pm1 MD10(M) - 128_N",
                                "Abundances Normalized T15CTLPm1 MD10(M) - 128_C",
                                "Abundances Normalized T7MHW2Pm1 MD10(M) - 129_N",
                                "Abundances Normalized T1CTLPm1 MD25(M) - 129_C",
                                "Abundances Normalized T13MHW2Pm1 MD10(M) - 130_N",
                                "Abundances Normalized T4CTLPm1 MD25(M) - 130_C",
                                "Abundances Normalized T2MHW1Pm1 MD25(M) - 131_N",
                                "Abundances Normalized T10CTLPm1 MD25(M) - 131_C",
                                "Abundances Normalized T6MHW1Pm1 MD25(M) - 132_N",
                                "Abundances Normalized T3MHW2Pm1 MD25(M) - 132_C",
                                "Abundances Normalized T9MHW1Pm1 MD25(M) - 133_N",
                                "Abundances Normalized T5MHW2Pm1 MD25(M) - 133_C",
                                "Abundances Normalized T12MHW1Pm1 MD25(M) - 134_N",
                                "Abundances Normalized T7MHW2Pm1 MD25(M) - 134_C",
                                "Abundances Normalized T13MHW2Pm1 MD25(M) - 135")
#############

#N° of duplicates
sum(duplicated(abundance_matrix$Accession))
#################

#Duplicate with higher mean
w <- c()
for (i in c(1:length(unique(abundance_matrix$Accession)))){
  dups <- which(abundance_matrix$Accession %in% unique(abundance_matrix$Accession)[i])
  v <- c()
  for (j in dups){
    v <- c(v,mean(as.numeric(abundance_matrix[-c(1,2)][j,])))
  }
  w = c(w,dups[which(v == max(v))[1]])
}
abundance_matrix <- abundance_matrix[w,]
####################
```

```
#Save table with UniProt Accession code, Protein Name and Normalized Abundance for each
protein
write.table(abundance_matrix, paste(pathFiles, "Full.csv", sep = ""), sep = ";",
col.names = NA)
########################################################################################
#######

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
######Reading the file and creating a matrix with only the normalized abundances#######



######################################Statistics#####################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create a DGEList object from a table of counts
Data <-
  DGEList(counts = abundance_matrix[3:20],
          group = Levels,
          ids = rownames(abundance_matrix))
################################################

##Create the generalized linear models for comparing the expression levels from the
different treatments
Design <- model.matrix( ~ 0 + Levels, data = Data$samples)
colnames(Design) <- levels(Data$samples$group)
Data <- estimateDisp(Data, Design)
fit <- glmFit(Data, Design)
########################################################################################
################

##Expression levels between the treatments
CTL_10vMHW2_10<-relativeExpression(colnames(Design)[1],colnames(Design)[4],Design,fit)
MHW2_10vMHW2_25<-relativeExpression(colnames(Design)[4],colnames(Design)[5],Design,fit)
CTL_25vMHW1_25<-relativeExpression(colnames(Design)[2],colnames(Design)[3],Design,fit)
CTL_25vMHW2_25<-relativeExpression(colnames(Design)[2],colnames(Design)[5],Design,fit)
MHW1_25vMHW2_25<-relativeExpression(colnames(Design)[3],colnames(Design)[5],Design,fit)
##########################################

##Merge the logFC,logCPM,LR and Pvalue with the proteins unique ids
Results<-cbind(CTL_10vMHW2_10$table, MHW2_10vMHW2_25$table, CTL_25vMHW1_25$table,
CTL_25vMHW2_25$table, MHW1_25vMHW2_25$table)
Results<-cbind(abundance_matrix$Accession,abundance_matrix$Protein,Results)
colnames(Results)<-c(
  "Accession",
  "Description",
  "logFC-CTL_10vMHW2_10",
  "logCPM-CTL_10vMHW2_10",
  "LR-CTL_10vMHW2_10",
  "Pvalue-CTL_10vMHW2_10",
  "logFC-MHW2_10vMHW2_25",
  "logCPM-MHW2_10vMHW2_25",
  "LR-MHW2_10vMHW2_25",
  "Pvalue-MHW2_10vMHW2_25",
  "logFC-CTL_25vMHW1_25",
  "logCPM-CTL_25vMHW1_25",
  "LR-CTL_25vMHW1_25",
  "Pvalue-CTL_25vMHW1_25",
  "logFC-CTL_25vMHW2_25",
  "logCPM-CTL_25vMHW2_25",
  "LR-CTL_25vMHW2_25",
```

```
  "Pvalue-CTL_25vMHW2_25",
  "logFC-MHW1_25vMHW2_25",
  "logCPM-MHW1_25vMHW2_25",
  "LR-MHW1_25vMHW2_25",
  "Pvalue-MHW1_25vMHW2_25"
)
head(Results)
write.table(Results, paste(pathFiles, "Results.csv", sep = ""), sep = ";", col.names =
NA)
###############################################

##Calculating the n° of DEPs in general
expressionTable<-decideTests(Results[,grepl("Pvalue-", colnames(Results))],coefficients
= Results[,grepl("logFC", colnames(Results))], adjust.method = "fdr", lfc = 1.5)
expressionTable<-as.data.frame(expressionTable)
colnames(expressionTable)<-c("CTL_10vMHW2_10", "MHW2_10vMHW2_25",
"CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25")
DEP<-cbind(Results,expressionTable)
depTable<-DEP[which(abs(DEP$CTL_10vMHW2_10) == 1 | abs(DEP$MHW2_10vMHW2_25) == 1 |
abs(DEP$CTL_25vMHW1_25) == 1 | abs(DEP$CTL_25vMHW2_25) == 1 | abs(DEP$MHW1_25vMHW2_25)
== 1),]
nrow(depTable) #N° of DEPs
write.table(depTable, paste(pathBothOmicsPathways, "DEP.csv", sep = ""), sep = ";",
col.names = NA)
save(CTL_10vMHW2_10, MHW2_10vMHW2_25, CTL_25vMHW1_25, CTL_25vMHW2_25, MHW1_25vMHW2_25,
file = "Files/lrt.RData")
#######################################

##N° of DEPs per treatment
DEP_CTL_10vMHW2_10<-sum(depTable$CTL_10vMHW2_10 != 0);DEP_CTL_10vMHW2_10
DEP_MHW2_10vMHW2_25<-sum(depTable$MHW2_10vMHW2_25 != 0);DEP_MHW2_10vMHW2_25
DEP_CTL_25vMHW2_25<-sum(depTable$CTL_25vMHW2_25 != 0);DEP_CTL_25vMHW2_25
DEP_CTL_25vMHW1_25<-sum(depTable$CTL_25vMHW1_25 != 0);DEP_CTL_25vMHW1_25
DEP_MHW1_25vMHW2_25<-sum(depTable$MHW1_25vMHW2_25 != 0);DEP_MHW1_25vMHW2_25
###########################

##N° of underexpressed proteins per treatment
uedep_CTL_10vMHW2_10<-sum(depTable$CTL_10vMHW2_10 == -1);uedep_CTL_10vMHW2_10
uedep_MHW2_10vMHW2_25<-sum(depTable$MHW2_10vMHW2_25 == -1);uedep_MHW2_10vMHW2_25
uedep_CTL_25vMHW2_25<-sum(depTable$CTL_25vMHW2_25 == -1);uedep_CTL_25vMHW2_25
uedep_CTL_25vMHW1_25<-sum(depTable$CTL_25vMHW1_25 == -1);uedep_CTL_25vMHW1_25
uedep_MHW1_25vMHW2_25<-sum(depTable$MHW1_25vMHW2_25 == -1);uedep_MHW1_25vMHW2_25
###########################################

##N° of overexpressed proteins per treatment
oedep_CTL_10vMHW2_10<-sum(depTable$CTL_10vMHW2_10 == 1);oedep_CTL_10vMHW2_10
oedep_MHW2_10vMHW2_25<-sum(depTable$MHW2_10vMHW2_25 == 1);oedep_MHW2_10vMHW2_25
oedep_CTL_25vMHW2_25<-sum(depTable$CTL_25vMHW2_25 == 1);oedep_CTL_25vMHW2_25
oedep_CTL_25vMHW1_25<-sum(depTable$CTL_25vMHW1_25 == 1);oedep_CTL_25vMHW1_25
oedep_MHW1_25vMHW2_25<-sum(depTable$MHW1_25vMHW2_25 == 1);oedep_MHW1_25vMHW2_25
###########################################

##Table with total DEP, underexpressed proteins and overexpressed proteins per treatment
dep_per_treatment <- data.frame(Treatments = contrasts,
                                #No_of_Proteins =
c(DEP_CTL_10vMHW2_10,DEP_CTL_25vMHW1_25,DEP_CTL_25vMHW2_25,DEP_MHW1_25vMHW2_25,DEP_MHW2_
10vMHW2_25),
                                No_of_underexpressed_Proteins =
c(uedep_CTL_10vMHW2_10,uedep_CTL_25vMHW1_25,uedep_CTL_25vMHW2_25,uedep_MHW1_25vMHW2_25,u
edep_MHW2_10vMHW2_25),
                                No_of_overexpressed_Proteins =
c(oedep_CTL_10vMHW2_10,oedep_CTL_25vMHW1_25,oedep_CTL_25vMHW2_25,oedep_MHW1_25vMHW2_25,o
edep_MHW2_10vMHW2_25))
```

```
colnames(dep_per_treatment) <- c("Treatments","N° of underexpressed proteins","N° of
overexpressed proteins")
write.table(dep_per_treatment, paste(pathTables, "dep_per_treatment.csv",sep=""),
sep=";", col.names=TRUE, row.names = FALSE)
########################################################################################

##Create tables with accession, logFC, Pvalue and Description for all proteins
depTable_up <- DEP[which(DEP$CTL_10vMHW2_10 == 1 | DEP$MHW2_10vMHW2_25 == 1 |
DEP$CTL_25vMHW1_25 == 1 | DEP$CTL_25vMHW2_25 == 1 | DEP$MHW1_25vMHW2_25 == 1),]
##############################################################################

##Create tables with accession, logFC, Pvalue and Description for each treatment
depTable_CTL_10vMHW2_10 <- DEP[which(abs(DEP$CTL_10vMHW2_10) == 1),]
depTable_CTL_10vMHW2_10 <- depTable_CTL_10vMHW2_10[, c(1,2,3,6)]
depTable_CTL_10vMHW2_10$padj <- p.adjust(depTable_CTL_10vMHW2_10$`Pvalue-
CTL_10vMHW2_10`, method = "fdr")
write.table(depTable_CTL_10vMHW2_10, paste(pathTables, "dep_CTL_10VMHW2_10.csv", sep =
""), sep = ";", col.names = TRUE, row.names = FALSE)

depTable_MHW2_10vMHW2_25 <- DEP[which(abs(DEP$MHW2_10vMHW2_25) == 1),]
depTable_MHW2_10vMHW2_25 <- depTable_MHW2_10vMHW2_25[, c(1,2,7,10)]
depTable_MHW2_10vMHW2_25$padj <- p.adjust(depTable_MHW2_10vMHW2_25$`Pvalue-
MHW2_10vMHW2_25`, method = "fdr")
write.table(depTable_MHW2_10vMHW2_25, paste(pathTables, "dep_MHW2_10vMHW2_25.csv", sep =
""), sep = ";", col.names = TRUE, row.names = FALSE)

depTable_CTL_25vMHW2_25 <- DEP[which(abs(DEP$CTL_25vMHW2_25) == 1),]
depTable_CTL_25vMHW2_25 <- depTable_CTL_25vMHW2_25[, c(1,2,15,18)]
depTable_CTL_25vMHW2_25$padj <- p.adjust(depTable_CTL_25vMHW2_25$`Pvalue-
CTL_25vMHW2_25`, method = "fdr")
write.table(depTable_CTL_25vMHW2_25, paste(pathTables, "dep_CTL_25vMHW2_25.csv", sep =
""), sep = ";", col.names = TRUE, row.names = FALSE)

depTable_CTL_25vMHW1_25 <- DEP[which(abs(DEP$CTL_25vMHW1_25) == 1),]
depTable_CTL_25vMHW1_25 <- depTable_CTL_25vMHW1_25[, c(1,2,11,14)]
depTable_CTL_25vMHW1_25$padj <- p.adjust(depTable_CTL_25vMHW1_25$`Pvalue-
CTL_25vMHW1_25`, method = "fdr")
write.table(depTable_CTL_25vMHW1_25, paste(pathTables, "dep_CTL_25vMHW1_25.csv", sep =
""), sep = ";", col.names = TRUE, row.names = FALSE)

depTable_MHW1_25vMHW2_25 <- DEP[which(abs(DEP$MHW1_25vMHW2_25) == 1),]
depTable_MHW1_25vMHW2_25 <- depTable_MHW1_25vMHW2_25[, c(1,2,19,22)]
depTable_MHW1_25vMHW2_25$padj <- p.adjust(depTable_MHW1_25vMHW2_25$`Pvalue-
MHW1_25vMHW2_25`, method = "fdr")
write.table(depTable_MHW1_25vMHW2_25, paste(pathTables, "dep_MHW1_25vMHW2_25.csv", sep =
""), sep = ";", col.names = TRUE, row.names = FALSE)
################################################################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#######################################Statistics####################################



####################################Pathway analysis#################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

DEP_pathway <- DEP[c(1,3:22)]
write.table(DEP_pathway, paste(pathBothOmicsPathways, "DEP_pathway.csv", sep = ""),sep =
";",col.names = TRUE, row.names = FALSE)

##Create data frames with Accession, logFC and Pvalue for the pathway analysis (all
```

```
proteins)
allp_CTL_10vMHW2_10<- data.frame(DEP_pathway$Accession,DEP_pathway$`logFC-
CTL_10vMHW2_10`,DEP_pathway$`Pvalue-CTL_10vMHW2_10`)
names(allp_CTL_10vMHW2_10) <- c("Accession","logFC","Pvalue")

allp_CTL_25vMHW1_25<- data.frame(DEP_pathway$Accession,DEP_pathway$`logFC-
CTL_25vMHW1_25`,DEP_pathway$`Pvalue-CTL_25vMHW1_25`)
names(allp_CTL_25vMHW1_25) <- c("Accession","logFC","Pvalue")

allp_CTL_25vMHW2_25<- data.frame(DEP_pathway$Accession,DEP_pathway$`logFC-
CTL_25vMHW2_25`,DEP_pathway$`Pvalue-CTL_25vMHW2_25`)
names(allp_CTL_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allp_MHW1_25vMHW2_25<- data.frame(DEP_pathway$Accession,DEP_pathway$`logFC-
MHW1_25vMHW2_25`,DEP_pathway$`Pvalue-MHW1_25vMHW2_25`)
names(allp_MHW1_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allp_MHW2_10vMHW2_25<- data.frame(DEP_pathway$Accession,DEP_pathway$`logFC-
MHW2_10vMHW2_25`,DEP_pathway$`Pvalue-MHW2_10vMHW2_25`)
names(allp_MHW2_10vMHW2_25) <- c("Accession","logFC","Pvalue")
##########################################################################

##Create a data structure
layers <- c("proteome")
odataCTL_10vMHW2_10 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW1_25 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW1_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW2_10vMHW2_25 <- initOmicsDataStructure(layer=layers)
########################

##Add proteome layer
odataCTL_10vMHW2_10$proteome <-
rankFeatures(allp_CTL_10vMHW2_10$logFC,allp_CTL_10vMHW2_10$Pvalue)
names(odataCTL_10vMHW2_10$proteome) <- allp_CTL_10vMHW2_10$Accession
odataCTL_10vMHW2_10$proteome <- sort(odataCTL_10vMHW2_10$proteome)
head(odataCTL_10vMHW2_10$proteome)

odataCTL_25vMHW1_25$proteome <-
rankFeatures(allp_CTL_25vMHW1_25$logFC,allp_CTL_25vMHW1_25$Pvalue)
names(odataCTL_25vMHW1_25$proteome) <- allp_CTL_25vMHW1_25$Accession
odataCTL_25vMHW1_25$proteome <- sort(odataCTL_25vMHW1_25$proteome)
head(odataCTL_25vMHW1_25$proteome)

odataCTL_25vMHW2_25$proteome <-
rankFeatures(allp_CTL_25vMHW2_25$logFC,allp_CTL_25vMHW2_25$Pvalue)
names(odataCTL_25vMHW2_25$proteome) <- allp_CTL_25vMHW2_25$Accession
odataCTL_25vMHW2_25$proteome <- sort(odataCTL_25vMHW2_25$proteome)
head(odataCTL_25vMHW2_25$proteome)

odataMHW1_25vMHW2_25$proteome <-
rankFeatures(allp_MHW1_25vMHW2_25$logFC,allp_MHW1_25vMHW2_25$Pvalue)
names(odataMHW1_25vMHW2_25$proteome) <- allp_MHW1_25vMHW2_25$Accession
odataMHW1_25vMHW2_25$proteome <- sort(odataMHW1_25vMHW2_25$proteome)
head(odataMHW1_25vMHW2_25$proteome)

odataMHW2_10vMHW2_25$proteome <-
rankFeatures(allp_MHW2_10vMHW2_25$logFC,allp_MHW2_10vMHW2_25$Pvalue)
names(odataMHW2_10vMHW2_25$proteome) <- allp_MHW2_10vMHW2_25$Accession
odataMHW2_10vMHW2_25$proteome <- sort(odataMHW2_10vMHW2_25$proteome)
head(odataMHW2_10vMHW2_25$proteome)
####################
```

```
##Select the databases we want to query and download pathway definitions
databases <- c("kegg")
pathways <- getMultiOmicsFeatures(dbs = databases, layer = layers,
                                  returnProteome = "UNIPROT",
                                  organism = "drerio",
                                  useLocal =  FALSE)
pathways_short <- lapply(names(pathways), function(name) {
  head(pathways[[name]], 2)
})
names(pathways_short) <- names(pathways)
pathways_short
pathways$proteome[8]
#########################################################################

##Run the pathway enrichment
enrichment_scoresCTL_10vMHW2_10 <- multiGSEA(pathways,odataCTL_10vMHW2_10)
Tenrichment_scoresCTL_10vMHW2_10 <-
as.data.frame(enrichment_scoresCTL_10vMHW2_10$proteome)
Tenrichment_scoresCTL_10vMHW2_10$leadingEdge <-
sapply(Tenrichment_scoresCTL_10vMHW2_10$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_10vMHW2_10 <-
Tenrichment_scoresCTL_10vMHW2_10[order(Tenrichment_scoresCTL_10vMHW2_10$padj),]
Tenrichment_scoresCTL_10vMHW2_10$pathway <- sub("^\\(KEGG\\) ", "",
Tenrichment_scoresCTL_10vMHW2_10$pathway)
top_10_Tenrichment_scoresCTL_10vMHW2_10 <- head(Tenrichment_scoresCTL_10vMHW2_10,10)
top_10_Tenrichment_scoresCTL_10vMHW2_10$contrast <- rep("CTL_10vMHW2_10")
write.table(top_10_Tenrichment_scoresCTL_10vMHW2_10,paste(pathTables,"top10_enrichment_s
coresCTL_10vMHW2_10.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW1_25 <- multiGSEA(pathways,odataCTL_25vMHW1_25)
Tenrichment_scoresCTL_25vMHW1_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW1_25$proteome)
Tenrichment_scoresCTL_25vMHW1_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW1_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW1_25 <-
Tenrichment_scoresCTL_25vMHW1_25[order(Tenrichment_scoresCTL_25vMHW1_25$padj),]
Tenrichment_scoresCTL_25vMHW1_25$pathway <- sub("^\\(KEGG\\) ", "",
Tenrichment_scoresCTL_25vMHW1_25$pathway)
top_10_Tenrichment_scoresCTL_25vMHW1_25 <- head(Tenrichment_scoresCTL_25vMHW1_25,10)
top_10_Tenrichment_scoresCTL_25vMHW1_25$contrast <- rep("CTL_25vMHW1_25")
write.table(top_10_Tenrichment_scoresCTL_25vMHW1_25,paste(pathTables,"top10_enrichment_s
coresCTL_25vMHW1_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW2_25 <- multiGSEA(pathways,odataCTL_25vMHW2_25)
Tenrichment_scoresCTL_25vMHW2_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW2_25$proteome)
Tenrichment_scoresCTL_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW2_25 <-
Tenrichment_scoresCTL_25vMHW2_25[order(Tenrichment_scoresCTL_25vMHW2_25$padj),]
Tenrichment_scoresCTL_25vMHW2_25$pathway <- sub("^\\(KEGG\\) ", "",
Tenrichment_scoresCTL_25vMHW2_25$pathway)
top_10_Tenrichment_scoresCTL_25vMHW2_25 <- head(Tenrichment_scoresCTL_25vMHW2_25,10)
top_10_Tenrichment_scoresCTL_25vMHW2_25$contrast <- rep("CTL_25vMHW2_25")
write.table(top_10_Tenrichment_scoresCTL_25vMHW2_25,paste(pathTables,"top10_enrichment_s
coresCTL_25vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW1_25vMHW2_25 <- multiGSEA(pathways,odataMHW1_25vMHW2_25)
Tenrichment_scoresMHW1_25vMHW2_25 <-
as.data.frame(enrichment_scoresMHW1_25vMHW2_25$proteome)
```

```
Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW1_25vMHW2_25 <-
Tenrichment_scoresMHW1_25vMHW2_25[order(Tenrichment_scoresMHW1_25vMHW2_25$padj),]
Tenrichment_scoresMHW1_25vMHW2_25$pathway <- sub("^\\(KEGG\\) ", "",
Tenrichment_scoresMHW1_25vMHW2_25$pathway)
top_10_Tenrichment_scoresMHW1_25vMHW2_25 <- head(Tenrichment_scoresMHW1_25vMHW2_25,10)
top_10_Tenrichment_scoresMHW1_25vMHW2_25$contrast <- rep("MHW1_25vMHW2_25")
write.table(top_10_Tenrichment_scoresMHW1_25vMHW2_25,paste(pathTables,"top10_enrichment_
scoresMHW1_25vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW2_10vMHW2_25 <- multiGSEA(pathways,odataMHW2_10vMHW2_25)
Tenrichment_scoresMHW2_10vMHW2_25 <-
as.data.frame(enrichment_scoresMHW2_10vMHW2_25$proteome)
Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW2_10vMHW2_25 <-
Tenrichment_scoresMHW2_10vMHW2_25[order(Tenrichment_scoresMHW2_10vMHW2_25$padj),]
Tenrichment_scoresMHW2_10vMHW2_25$pathway <- sub("^\\(KEGG\\) ", "",
Tenrichment_scoresMHW2_10vMHW2_25$pathway)
top_10_Tenrichment_scoresMHW2_10vMHW2_25 <- head(Tenrichment_scoresMHW2_10vMHW2_25,10)
top_10_Tenrichment_scoresMHW2_10vMHW2_25$contrast <- rep("MHW2_10vMHW2_25")
write.table(top_10_Tenrichment_scoresMHW2_10vMHW2_25,paste(pathTables,"top10_enrichment_
scoresMHW2_10vMHW2_25.csv",sep=""),sep=";",row.names = FALSE)
############################

##Create dataset with all enrichment scores
combined_df <- rbind(top_10_Tenrichment_scoresCTL_10vMHW2_10,
top_10_Tenrichment_scoresCTL_25vMHW1_25, top_10_Tenrichment_scoresCTL_25vMHW2_25,
top_10_Tenrichment_scoresMHW1_25vMHW2_25, top_10_Tenrichment_scoresMHW2_10vMHW2_25)
###########################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
####################################Pathway analysis###################################



##########################################Plots########################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Volcano Plot
par(mfrow = c(1,1), mar = c(7,7,4,4), family = "Arial")
for(i in 1:length(contrasts)){
  volcanoPlot(pathFiles, colour[contrasts[i]], legend[contrasts[i]], get(paste("title",
contrasts[i], sep = "")), contrasts[i], i)
  dev.print(tiff, filename = paste(pathVolcano, "VolcanoPlot-", contrasts[i], ".tif",
sep = ""), height = 15, width = 15, units = "cm", res = 600)
  dev.copy(svg, filename = paste(pathVolcano, "VolcanoPlot-",contrasts[i], ".svg", sep =
""), height = 15, width = 15)
  dev.off()
}
##############

##Heatmap
colCutoff = 10
rowCutoff = 4
heatmapGraph(heatData,colCutoff, rowCutoff)
#########
```

```
#Dotplot
dotplot(combined_df,pathDotplot)
########

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
###########################################Plots#########################################
```

A.2.3 Integrative Omics R Script

Please refer to https://github.com/jdplopes/MScThesis_MultiOmics for the complete R script.

```
##Master's Thesis
##João Lopes
set.seed(1)
setwd("C:\\Users\\jdpl2\\OneDrive\\Ambiente de Trabalho\\Mestrado\\2° Ano\\Both Omics")

##Load a R package
library(tximport)
library(limma)
library(edgeR)
library(seqinr)
#library(UniprotR)
library(ggplot2)
library(cowplot)
library(gplots)
library(RColorBrewer)
library(gridGraphics)
#library(biomaRt) #Do not use for analyzing DEGs (problems with the getSequence
function)
library(carData)
library(car)
library(rhdf5)
library(data.table)
library(multiGSEA)
library(tidyverse)
library(readxl)
library(tidyr)
library(forcats)
windowsFonts(Calibri = windowsFont("Arial"))

dir.create("Plots")
dir.create("Plots/Dotplot")
dir.create("Files")

pathDotplot <- "Plots/Dotplot/"
pathFiles <- "Files/"

dotplot <- function (data,path) {
  o <- ggplot(data, aes(x = contrast, y = pathway)) +
    geom_point(aes(size = 1-padj, color = NES, fill = NES), shape = 21, stroke = 0.5) +
    scale_color_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    scale_size_continuous(range = c(3, 15)) +
    labs(x = "Contrast", y = "Pathway",
         size = "1-p.adjust",
         color = "NES",
         fill = "NES") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 18),
          axis.text.y = element_text(size = 18),
          axis.title.x = element_text(size = 20),
          axis.title.y = element_text(size = 20),
          legend.position = "bottom",
          legend.title = element_text(size = 20),
          legend.text = element_text(size = 20),
          panel.grid.major = element_line(linewidth = 1.2, color = "grey90"),
          panel.grid.minor = element_line(linewidth = 0.8, color = "grey90"),
          panel.background = element_rect(fill = "grey", color = NA))
```

```
  ggsave(filename = paste(path, "Dotplot.svg", sep = ""), plot = o, device = "svg",
width = 15, height = 15)
  ggsave(filename = paste(path, "Dotplot.tiff", sep = ""), plot = o, device = "tiff",
width = 15, height = 15)
}

#######################################Files/Data#######################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

DEG_pathway <- read.csv("DEG_pathway.csv",sep =";")
DEP_pathway <- read.csv("DEP_pathway.csv",sep = ";")
genes_table <- read.csv("genes_ID.csv",sep =";")
DEG_acc <- read.csv("DEG.csv",sep =";")
DEP_acc <- read.csv("DEP.csv",sep=";")
#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
########################################Files/Data########################################

contrasts <-
c("CTL_10vMHW2_10","CTL_25vMHW1_25","CTL_25vMHW2_25","MHW1_25vMHW2_25","MHW2_10vMHW2_25"
)

proteins <- unique(DEP_pathway$Accession)
genes <- unique(genes_table$Accession)
acc_common <- intersect(proteins,genes);length(acc_common)
de_acc_common <-
intersect(unique(DEG_acc$ID),unique(DEP_acc$Accession));length(de_acc_common)

#####################################Pathway analysis#####################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

##Create data frames with Accession, logFC and Pvalue for the pathway analysis
allg_CTL_10vMHW2_10<-
data.frame(DEG_pathway$ID,DEG_pathway$`logFC.CTL_10vMHW2_10`,DEG_pathway$`Pvalue.CTL_10v
MHW2_10`)
names(allg_CTL_10vMHW2_10) <- c("Accession","logFC","Pvalue")
allp_CTL_10vMHW2_10<-
data.frame(DEP_pathway$Accession,DEP_pathway$`logFC.CTL_10vMHW2_10`,DEP_pathway$`Pvalue.
CTL_10vMHW2_10`)
names(allp_CTL_10vMHW2_10) <- c("Accession","logFC","Pvalue")

allg_CTL_25vMHW1_25<-
data.frame(DEG_pathway$ID,DEG_pathway$`logFC.CTL_25vMHW1_25`,DEG_pathway$`Pvalue.CTL_25v
MHW1_25`)
names(allg_CTL_25vMHW1_25) <- c("Accession","logFC","Pvalue")
allp_CTL_25vMHW1_25<-
data.frame(DEP_pathway$Accession,DEP_pathway$`logFC.CTL_10vMHW2_10`,DEP_pathway$`Pvalue.
CTL_10vMHW2_10`)
names(allp_CTL_25vMHW1_25) <- c("Accession","logFC","Pvalue")

allg_CTL_25vMHW2_25<-
data.frame(DEG_pathway$ID,DEG_pathway$`logFC.CTL_25vMHW2_25`,DEG_pathway$`Pvalue.CTL_25v
MHW2_25`)
names(allg_CTL_25vMHW2_25) <- c("Accession","logFC","Pvalue")
allp_CTL_25vMHW2_25<-
data.frame(DEP_pathway$Accession,DEP_pathway$`logFC.CTL_25vMHW2_25`,DEP_pathway$`Pvalue.
CTL_25vMHW2_25`)
names(allp_CTL_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allg_MHW1_25vMHW2_25<-
data.frame(DEG_pathway$ID,DEG_pathway$`logFC.MHW1_25vMHW2_25`,DEG_pathway$`Pvalue.MHW1_2
```

```
5vMHW2_25`)
names(allg_MHW1_25vMHW2_25) <- c("Accession","logFC","Pvalue")
allp_MHW1_25vMHW2_25<-
data.frame(DEP_pathway$Accession,DEP_pathway$`logFC.MHW1_25vMHW2_25`,DEP_pathway$`Pvalue
.MHW1_25vMHW2_25`)
names(allp_MHW1_25vMHW2_25) <- c("Accession","logFC","Pvalue")

allg_MHW2_10vMHW2_25<-
data.frame(DEG_pathway$ID,DEG_pathway$`logFC.MHW2_10vMHW2_25`,DEG_pathway$`Pvalue.MHW2_1
0vMHW2_25`)
names(allg_MHW2_10vMHW2_25) <- c("Accession","logFC","Pvalue")
allp_MHW2_10vMHW2_25<-
data.frame(DEP_pathway$Accession,DEP_pathway$`logFC.MHW2_10vMHW2_25`,DEP_pathway$`Pvalue
.MHW2_10vMHW2_25`)
names(allp_MHW2_10vMHW2_25) <- c("Accession","logFC","Pvalue")
##########################################################################

##Create a data structure
layers <- c("transcriptome","proteome")
odataCTL_10vMHW2_10 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW1_25 <- initOmicsDataStructure(layer=layers)
odataCTL_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW1_25vMHW2_25 <- initOmicsDataStructure(layer=layers)
odataMHW2_10vMHW2_25 <- initOmicsDataStructure(layer=layers)
#########################

##Add transcriptome layer
odataCTL_10vMHW2_10$transcriptome <-
rankFeatures(allg_CTL_10vMHW2_10$logFC,allg_CTL_10vMHW2_10$Pvalue)
names(odataCTL_10vMHW2_10$transcriptome) <- allg_CTL_10vMHW2_10$Accession
odataCTL_10vMHW2_10$transcriptome <- sort(odataCTL_10vMHW2_10$transcriptome)
head(odataCTL_10vMHW2_10$transcriptome)

odataCTL_25vMHW1_25$transcriptome <-
rankFeatures(allg_CTL_25vMHW1_25$logFC,allg_CTL_25vMHW1_25$Pvalue)
names(odataCTL_25vMHW1_25$transcriptome) <- allg_CTL_25vMHW1_25$Accession
odataCTL_25vMHW1_25$transcriptome <- sort(odataCTL_25vMHW1_25$transcriptome)
head(odataCTL_25vMHW1_25$transcriptome)

odataCTL_25vMHW2_25$transcriptome <-
rankFeatures(allg_CTL_25vMHW2_25$logFC,allg_CTL_25vMHW2_25$Pvalue)
names(odataCTL_25vMHW2_25$transcriptome) <- allg_CTL_25vMHW2_25$Accession
odataCTL_25vMHW2_25$transcriptome <- sort(odataCTL_25vMHW2_25$transcriptome)
head(odataCTL_25vMHW2_25$transcriptome)

odataMHW1_25vMHW2_25$transcriptome <-
rankFeatures(allg_MHW1_25vMHW2_25$logFC,allg_MHW1_25vMHW2_25$Pvalue)
names(odataMHW1_25vMHW2_25$transcriptome) <- allg_MHW1_25vMHW2_25$Accession
odataMHW1_25vMHW2_25$transcriptome <- sort(odataMHW1_25vMHW2_25$transcriptome)
head(odataMHW1_25vMHW2_25$transcriptome)

odataMHW2_10vMHW2_25$transcriptome <-
rankFeatures(allg_MHW2_10vMHW2_25$logFC,allg_MHW2_10vMHW2_25$Pvalue)
names(odataMHW2_10vMHW2_25$transcriptome) <- allg_MHW2_10vMHW2_25$Accession
odataMHW2_10vMHW2_25$transcriptome <- sort(odataMHW2_10vMHW2_25$transcriptome)
head(odataMHW2_10vMHW2_25$transcriptome)
####################

##Add proteome layer
odataCTL_10vMHW2_10$proteome <-
rankFeatures(allp_CTL_10vMHW2_10$logFC,allp_CTL_10vMHW2_10$Pvalue)
names(odataCTL_10vMHW2_10$proteome) <- allp_CTL_10vMHW2_10$Accession
odataCTL_10vMHW2_10$proteome <- sort(odataCTL_10vMHW2_10$proteome)
```

```
head(odataCTL_10vMHW2_10$proteome)

odataCTL_25vMHW1_25$proteome <-
rankFeatures(allp_CTL_25vMHW1_25$logFC,allp_CTL_25vMHW1_25$Pvalue)
names(odataCTL_25vMHW1_25$proteome) <- allp_CTL_25vMHW1_25$Accession
odataCTL_25vMHW1_25$proteome <- sort(odataCTL_25vMHW1_25$proteome)
head(odataCTL_25vMHW1_25$proteome)

odataCTL_25vMHW2_25$proteome <-
rankFeatures(allp_CTL_25vMHW2_25$logFC,allp_CTL_25vMHW2_25$Pvalue)
names(odataCTL_25vMHW2_25$proteome) <- allp_CTL_25vMHW2_25$Accession
odataCTL_25vMHW2_25$proteome <- sort(odataCTL_25vMHW2_25$proteome)
head(odataCTL_25vMHW2_25$proteome)

odataMHW1_25vMHW2_25$proteome <-
rankFeatures(allp_MHW1_25vMHW2_25$logFC,allp_MHW1_25vMHW2_25$Pvalue)
names(odataMHW1_25vMHW2_25$proteome) <- allp_MHW1_25vMHW2_25$Accession
odataMHW1_25vMHW2_25$proteome <- sort(odataMHW1_25vMHW2_25$proteome)
head(odataMHW1_25vMHW2_25$proteome)

odataMHW2_10vMHW2_25$proteome <-
rankFeatures(allp_MHW2_10vMHW2_25$logFC,allp_MHW2_10vMHW2_25$Pvalue)
names(odataMHW2_10vMHW2_25$proteome) <- allp_MHW2_10vMHW2_25$Accession
odataMHW2_10vMHW2_25$proteome <- sort(odataMHW2_10vMHW2_25$proteome)
head(odataMHW2_10vMHW2_25$proteome)
####################

##Select the databases we want to query and download pathway definitions
databases <- c("kegg")
pathways <- getMultiOmicsFeatures(dbs = databases, layer = layers,
                                  returnTranscriptome = "UNIPROT",
                                  returnProteome = "UNIPROT",
                                  organism = "drerio",
                                  useLocal =  FALSE)
pathways_short <- lapply(names(pathways), function(name) {
  head(pathways[[name]], 2)
})
names(pathways_short) <- names(pathways)
pathways_short
pathways$transcriptome[8]
#########################################################################

##Run the pathway enrichment
enrichment_scoresCTL_10vMHW2_10 <- multiGSEA(pathways,odataCTL_10vMHW2_10)
Tenrichment_scoresCTL_10vMHW2_10 <-
as.data.frame(enrichment_scoresCTL_10vMHW2_10$transcriptome)
Tenrichment_scoresCTL_10vMHW2_10$leadingEdge <-
sapply(Tenrichment_scoresCTL_10vMHW2_10$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_10vMHW2_10 <-
Tenrichment_scoresCTL_10vMHW2_10[order(Tenrichment_scoresCTL_10vMHW2_10$padj),]
Tenrichment_scoresCTL_10vMHW2_10$contrast <- rep("CTL_10vMHW2_10")
top_10_esCTL_10vMHW2_10 <- head(Tenrichment_scoresCTL_10vMHW2_10,10)
write.table(top_10_esCTL_10vMHW2_10,paste(pathFiles,"top10_enrichment_scoresCTL_10vMHW2_
10.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW1_25 <- multiGSEA(pathways,odataCTL_25vMHW1_25)
Tenrichment_scoresCTL_25vMHW1_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW1_25$transcriptome)
Tenrichment_scoresCTL_25vMHW1_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW1_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW1_25 <-
```

```
Tenrichment_scoresCTL_25vMHW1_25[order(Tenrichment_scoresCTL_25vMHW1_25$padj),]
Tenrichment_scoresCTL_25vMHW1_25$contrast <- rep("CTL_25vMHW1_25")
top_10_esCTL_25vMHW1_25 <- head(Tenrichment_scoresCTL_25vMHW1_25,10)
write.table(top_10_esCTL_25vMHW1_25,paste(pathFiles,"top10_enrichment_scoresCTL_25vMHW1_
25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresCTL_25vMHW2_25 <- multiGSEA(pathways,odataCTL_25vMHW2_25)
Tenrichment_scoresCTL_25vMHW2_25 <-
as.data.frame(enrichment_scoresCTL_25vMHW2_25$transcriptome)
Tenrichment_scoresCTL_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresCTL_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresCTL_25vMHW2_25 <-
Tenrichment_scoresCTL_25vMHW2_25[order(Tenrichment_scoresCTL_25vMHW2_25$padj),]
Tenrichment_scoresCTL_25vMHW2_25$contrast <- rep("CTL_25vMHW2_25")
top_10_esCTL_25vMHW2_25 <- head(Tenrichment_scoresCTL_25vMHW2_25,10)
write.table(top_10_esCTL_25vMHW2_25,paste(pathFiles,"top10_enrichment_scoresCTL_25vMHW2_
25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW1_25vMHW2_25 <- multiGSEA(pathways,odataMHW1_25vMHW2_25)
Tenrichment_scoresMHW1_25vMHW2_25 <-
as.data.frame(enrichment_scoresMHW1_25vMHW2_25$transcriptome)
Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW1_25vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW1_25vMHW2_25 <-
Tenrichment_scoresMHW1_25vMHW2_25[order(Tenrichment_scoresMHW1_25vMHW2_25$padj),]
Tenrichment_scoresMHW1_25vMHW2_25$contrast <- rep("MHW1_25vMHW2_25")
top_10_esMHW1_25vMHW2_25 <- head(Tenrichment_scoresMHW1_25vMHW2_25,10)
write.table(top_10_esMHW1_25vMHW2_25,paste(pathFiles,"top10_enrichment_scoresMHW1_25vMHW
2_25.csv",sep=""),sep=";",row.names = FALSE)

enrichment_scoresMHW2_10vMHW2_25 <- multiGSEA(pathways,odataMHW2_10vMHW2_25)
Tenrichment_scoresMHW2_10vMHW2_25 <-
as.data.frame(enrichment_scoresMHW2_10vMHW2_25$transcriptome)
Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge <-
sapply(Tenrichment_scoresMHW2_10vMHW2_25$leadingEdge, function(x) paste(x, collapse =
";"))
Tenrichment_scoresMHW2_10vMHW2_25 <-
Tenrichment_scoresMHW2_10vMHW2_25[order(Tenrichment_scoresMHW2_10vMHW2_25$padj),]
Tenrichment_scoresMHW2_10vMHW2_25$contrast <- rep("MHW2_10vMHW2_25")
top_10_esMHW2_10vMHW2_25 <- head(Tenrichment_scoresMHW2_10vMHW2_25,10)
write.table(top_10_esMHW2_10vMHW2_25,paste(pathFiles,"top10_enrichment_scoresMHW2_10vMHW
2_25.csv",sep=""),sep=";",row.names = FALSE)
############################

##Create dataset with all enrichment scores
combined_df <- rbind(top_10_esCTL_10vMHW2_10, top_10_esCTL_25vMHW1_25,
top_10_esCTL_25vMHW2_25, top_10_esMHW1_25vMHW2_25, top_10_esMHW2_10vMHW2_25)
combined_df$pathway <- sub("^\\(KEGG\\) ", "", combined_df$pathway)
###########################################

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
####################################Pathway analysis###################################



##########################################Plots########################################
#|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
#VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV#

#Dotplot
```

```
dotplot(combined_df,pathDotplot)
########

#^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^#
#||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
###########################################Plots########################################
```

A.3 Command line – server

```
#FastQC

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=20

#SBATCH --mem=200G

#SBATCH --partition=hpc

module load fastqc/0.11.9

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_4_N3063_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_4_N3063_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_6_N3053_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_6_N3053_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_7_N3065_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_7_N3065_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_9_N3057_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_9_N3057_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC
```

```
fastqc /users3/seatox/data/Pmicrops/RawData/RNA_11_N3066_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_11_N3066_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_14_N3058_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_14_N3058_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_17_N3064_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_17_N3064_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_19_N3055_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_19_N3055_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_20_N3060_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_20_N3060_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_1.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC

fastqc /users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_2.fq.gz --outdir
/users3/seatox/data/Pmicrops/FastQC
```

```
#TrimGalore

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=20

#SBATCH --mem=200G
```

```
#SBATCH --partition=incd


module load spack

module load python-3.10.8-gcc-11.3.0

module load trimgalore/0.6.10


trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_4_N3063_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_4_N3063_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_6_N3053_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_6_N3053_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_7_N3065_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_7_N3065_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_9_N3057_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_9_N3057_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_11_N3066_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_11_N3066_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_14_N3058_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_14_N3058_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
```

```
/users3/seatox/data/Pmicrops/RawData/RNA_17_N3064_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_17_N3064_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_19_N3055_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_19_N3055_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_20_N3060_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_20_N3060_2.fq.gz

trim_galore --paired --illumina --phred33 --output_dir
/users3/seatox/data/Pmicrops/TrimGalore --length 20 --stringency 1 -e 0.1 --gzip
/users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_1.fq.gz
/users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_2.fq.gz



#Trinity

#!/bin/bash



#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=30

#SBATCH --mem=200G

#SBATCH --partition=hpc



module load jellyfish/2.2.7

module load bowtie2/2.4.2

module load samtools/1.8

module load spack

module load python/3.10.8

module load py-numpy-1.23.4-gcc-11.3.0

module load trinity/2.14.0



Trinity --seqType fq --left
/users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_1.fq.gz,/users3/seatox/data/Pmicrops/Ra
wData/RNA_4_N3063_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_1.fq.gz,/user
s3/seatox/data/Pmicrops/RawData/RNA_6_N3053_1.fq.gz,/users3/seatox/data/Pmicrops/RawData
/RNA_7_N3065_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_1.fq.gz,/users3/se
atox/data/Pmicrops/RawData/RNA_9_N3057_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_
```

```
11_N3066_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_1.fq.gz,/users3/seato
x/data/Pmicrops/RawData/RNA_14_N3058_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_17
_N3064_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_1.fq.gz,/users3/seatox/
data/Pmicrops/RawData/RNA_19_N3055_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_20_N
3060_1.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_1.fq.gz --right
/users3/seatox/data/Pmicrops/RawData/RNA_3_N3062_2.fq.gz,/users3/seatox/data/Pmicrops/Ra
wData/RNA_4_N3063_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_5_N3059_2.fq.gz,/user
s3/seatox/data/Pmicrops/RawData/RNA_6_N3053_2.fq.gz,/users3/seatox/data/Pmicrops/RawData
/RNA_7_N3065_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_8_N3056_2.fq.gz,/users3/se
atox/data/Pmicrops/RawData/RNA_9_N3057_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_
11_N3066_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_13_N3067_2.fq.gz,/users3/seato
x/data/Pmicrops/RawData/RNA_14_N3058_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_17
_N3064_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_18_N3054_2.fq.gz,/users3/seatox/
data/Pmicrops/RawData/RNA_19_N3055_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_20_N
3060_2.fq.gz,/users3/seatox/data/Pmicrops/RawData/RNA_21_N3061_2.fq.gz --max_memory 200G
--CPU 30 --no_salmon --full_cleanup --output
/users3/seatox/data/Pmicrops/Trinity_NoSalmon


#Nx Statistics

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=8

#SBATCH --mem=100G

#SBATCH --partition=hpc


module load jellyfish/2.2.7

module load bowtie2/2.4.2

module load samtools/1.8

module load trinity/2.14.0


TrinityStats.pl /users3/seatox/data/Pmicrops/Trinity.Trinity.fasta


#ExN50 Statistics

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=2

#SBATCH --mem=200G
```

```
#SBATCH --partition=hpc


module load spack

module load kallisto/0.50.1

module load gcc11/R/4.2.2

module load r-edger-3.38.4-gcc-11.3.0

module load r-limma-3.52.4-gcc-11.3.0

module load trinity/2.14.0


abundance_estimates_to_matrix.pl --est_method kallisto --gene_trans_map none --
out_prefix kallisto --name_sample_by_basedir
/users3/seatox/data/Pmicrops/Kallisto/RNA_11_N3066/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_13_N3067/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_14_N3058/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_17_N3064/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_18_N3054/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_19_N3055/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_20_N3060/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_21_N3061/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_3_N3062/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_4_N3063/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_5_N3059/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_6_N3053/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_7_N3065/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_8_N3056/abundance.tsv
/users3/seatox/data/Pmicrops/Kallisto/RNA_9_N3057/abundance.tsv


echo "Kallisto done"


/cvmfs/sw.el8/spack/0.19.1/opt/spack/linux-almalinux8-zen3/gcc-11.3.0/trinity-
2.14.0.FULL-735bglhnbk4skaivsbffy3swclqg2uku/bin/util/misc/contig_ExN50_statistic.pl
kallisto.isoform.TMM.EXPR.matrix
/users3/seatox/data/Pmicrops/Trinity/Trinity.Trinity.fasta > ExN50.stats

/cvmfs/sw.el8/spack/0.19.1/opt/spack/linux-almalinux8-zen3/gcc-11.3.0/trinity-
2.14.0.FULL-735bglhnbk4skaivsbffy3swclqg2uku/bin/util/misc/plot_ExN50_statistic.Rscript
ExN50.stats


#Kallisto - Index

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=10
```

```
#SBATCH --mem=200G

#SBATCH --partition=hpc



module load kallisto/0.50.1

kallisto index --make-unique --index
/users3/seatox/data/Pmicrops/Kallisto/Pmicrops.index
/users3/seatox/data/Pmicrops/Trinity/Trinity.Trinity.fasta



#Kallisto - Quantification

#!/bin/bash



#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=10

#SBATCH --mem=200G

#SBATCH --partition=hpc



module load kallisto/0.50.1

##1. RNA_3_N3062

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_3_N3062

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_3_N3062
/users3/seatox/data/Pmicrops/TrimGalore/RNA_3_N3062_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_3_N3062_2_val_2.fq.gz

##2. RNA_4_N3063

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_4_N3063

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_4_N3063
/users3/seatox/data/Pmicrops/TrimGalore/RNA_4_N3063_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_4_N3063_2_val_2.fq.gz

##3. RNA_5_N3059

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_5_N3059

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_5_N3059
/users3/seatox/data/Pmicrops/TrimGalore/RNA_5_N3059_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_5_N3059_2_val_2.fq.gz

##4. RNA_6_N3053

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_6_N3053
```

```
kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_6_N3053
/users3/seatox/data/Pmicrops/TrimGalore/RNA_6_N3053_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_6_N3053_2_val_2.fq.gz

##5. RNA_7_N3065

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_7_N3065

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_7_N3065
/users3/seatox/data/Pmicrops/TrimGalore/RNA_7_N3065_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_7_N3065_2_val_2.fq.gz

##6. RNA_8_N3056

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_8_N3056

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_8_N3056
/users3/seatox/data/Pmicrops/TrimGalore/RNA_8_N3056_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_8_N3056_2_val_2.fq.gz

##7. RNA_9_N3057

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_9_N3057

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_9_N3057
/users3/seatox/data/Pmicrops/TrimGalore/RNA_9_N3057_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_9_N3057_2_val_2.fq.gz

##8. RNA_11_N3066

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_11_N3066

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_11_N3066
/users3/seatox/data/Pmicrops/TrimGalore/RNA_11_N3066_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_11_N3066_2_val_2.fq.gz

##9. RNA_13_N3067

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_13_N3067

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_13_N3067
/users3/seatox/data/Pmicrops/TrimGalore/RNA_13_N3067_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_13_N3067_2_val_2.fq.gz

##10. RNA_14_N3058

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_14_N3058

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_14_N3058
/users3/seatox/data/Pmicrops/TrimGalore/RNA_14_N3058_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_14_N3058_2_val_2.fq.gz

##11. RNA_17_N3064

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_17_N3064
```

```
kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_17_N3064
/users3/seatox/data/Pmicrops/TrimGalore/RNA_17_N3064_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_17_N3064_2_val_2.fq.gz

##12. RNA_18_N3054

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_18_N3054

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_18_N3054
/users3/seatox/data/Pmicrops/TrimGalore/RNA_18_N3054_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_18_N3054_2_val_2.fq.gz

##13. RNA_19_N3055

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_19_N3055

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_19_N3055
/users3/seatox/data/Pmicrops/TrimGalore/RNA_19_N3055_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_19_N3055_2_val_2.fq.gz

##14. RNA_20_N3060

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_20_N3060

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_20_N3060
/users3/seatox/data/Pmicrops/TrimGalore/RNA_20_N3060_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_20_N3060_2_val_2.fq.gz

##15. RNA_21_N3061

mkdir /users3/seatox/data/Pmicrops/Kallisto_50/RNA_21_N3061

kallisto quant -i /users3/seatox/data/Pmicrops/Kallisto_50/Pmicrops.index -o
/users3/seatox/data/Pmicrops/Kallisto_50/RNA_21_N3061
/users3/seatox/data/Pmicrops/TrimGalore/RNA_21_N3061_1_val_1.fq.gz
/users3/seatox/data/Pmicrops/TrimGalore/RNA_21_N3061_2_val_2.fq.gz



#Read content

#!/bin/bash



#SBATCH --cpus-per-task=2

#SBATCH --mem=200G

#SBATCH --partition=hpc



module load bowtie2/2.4.2

module load samtools/1.8



bowtie2-build /users3/seatox/data/Pmicrops/Trinity/Trinity.Trinity.fasta
Pmicrops_Trinity
```

```
bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_11_N3066_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_11_N3066_2_val_2.fq.gz -S
RNA_11_Pmicrops_Trinity.sam > align_stats_RNA_11.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_13_N3067_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_13_N3067_2_val_2.fq.gz -S
RNA_13_Pmicrops_Trinity.sam > align_stats_RNA_13.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_14_N3058_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_14_N3058_2_val_2.fq.gz -S
RNA_14_Pmicrops_Trinity.sam > align_stats_RNA_14.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_17_N3064_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_17_N3064_2_val_2.fq.gz -S
RNA_17_Pmicrops_Trinity.sam > align_stats_RNA_17.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_18_N3054_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_18_N3054_2_val_2.fq.gz -S
RNA_18_Pmicrops_Trinity.sam > align_stats_RNA_18.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_19_N3055_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_19_N3055_2_val_2.fq.gz -S
RNA_19_Pmicrops_Trinity.sam > align_stats_RNA_19.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_20_N3060_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_20_N3060_2_val_2.fq.gz -S
RNA_20_Pmicrops_Trinity.sam > align_stats_RNA_20.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_21_N3061_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_21_N3061_2_val_2.fq.gz -S
RNA_21_Pmicrops_Trinity.sam > align_stats_RNA_21.txt


bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_3_N3062_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_3_N3062_2_val_2.fq.gz -S
RNA_3_Pmicrops_Trinity.sam > align_stats_RNA_3.txt
```

```
bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_4_N3063_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_4_N3063_2_val_2.fq.gz -S
RNA_4_Pmicrops_Trinity.sam > align_stats_RNA_4.txt



bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_5_N3059_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_5_N3059_2_val_2.fq.gz -S
RNA_5_Pmicrops_Trinity.sam > align_stats_RNA_5.txt



bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_6_N3053_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_6_N3053_2_val_2.fq.gz -S
RNA_6_Pmicrops_Trinity.sam > align_stats_RNA_6.txt



bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_7_N3065_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_7_N3065_2_val_2.fq.gz -S
RNA_7_Pmicrops_Trinity.sam > align_stats_RNA_7.txt



bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_8_N3056_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_8_N3056_2_val_2.fq.gz -S
RNA_8_Pmicrops_Trinity.sam > align_stats_RNA_8.txt



bowtie2 -p 20 -q --no-unal -k 20 -x Pmicrops_Trinity -1
/users3/seatox/data/Pmicrops/TrimGalore/RNA_9_N3057_1_val_1.fq.gz -2
/users3/seatox/data/Pmicrops/TrimGalore/RNA_9_N3057_1_val_1.fq.gz -S
RNA_9_Pmicrops_Trinity.sam > align_stats_RNA_9.txt



echo "Bowtie done"

samtools view -bS RNA_11_Pmicrops_Trinity.sam > RNA_11_Pmicrops_Trinity.bam

samtools view -bS RNA_13_Pmicrops_Trinity.sam > RNA_13_Pmicrops_Trinity.bam

samtools view -bS RNA_14_Pmicrops_Trinity.sam > RNA_14_Pmicrops_Trinity.bam

samtools view -bS RNA_17_Pmicrops_Trinity.sam > RNA_17_Pmicrops_Trinity.bam

samtools view -bS RNA_18_Pmicrops_Trinity.sam > RNA_18_Pmicrops_Trinity.bam

samtools view -bS RNA_19_Pmicrops_Trinity.sam > RNA_19_Pmicrops_Trinity.bam

samtools view -bS RNA_20_Pmicrops_Trinity.sam > RNA_20_Pmicrops_Trinity.bam

samtools view -bS RNA_21_Pmicrops_Trinity.sam > RNA_21_Pmicrops_Trinity.bam

samtools view -bS RNA_3_Pmicrops_Trinity.sam > RNA_3_Pmicrops_Trinity.bam

samtools view -bS RNA_4_Pmicrops_Trinity.sam > RNA_4_Pmicrops_Trinity.bam

samtools view -bS RNA_5_Pmicrops_Trinity.sam > RNA_5_Pmicrops_Trinity.bam
```

```
samtools view -bS RNA_6_Pmicrops_Trinity.sam > RNA_6_Pmicrops_Trinity.bam

samtools view -bS RNA_7_Pmicrops_Trinity.sam > RNA_7_Pmicrops_Trinity.bam

samtools view -bS RNA_8_Pmicrops_Trinity.sam > RNA_8_Pmicrops_Trinity.bam

samtools view -bS RNA_9_Pmicrops_Trinity.sam > RNA_9_Pmicrops_Trinity.bam


#Transdecoder

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --cpus-per-task=20

#SBATCH --mem=200G

#SBATCH --partition=incd


module load transdecoder/5.5.0


TransDecoder.LongOrfs -t /users3/seatox/data/Pmicrops/Trinity.Trinity.fasta

TransDecoder.Predict -t /users3/seatox/data/Pmicrops/Trinity.Trinity.fasta


#Blastp - SwissProt (General)

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=20

#SBATCH --mem=200G

#SBATCH --partition=hpc


module load blast-plus/2.12.0


makeblastdb -in /users3/seatox/data/Pmicrops/Blastp/uniprot_sprot.fasta -out
/users3/seatox/data/Pmicrops/Blastp/uniprot_sprot -dbtype prot


blastp -query
/users3/seatox/data/Pmicrops/TransDecoder/Trinity.Trinity.fasta.transdecoder.pep -db
```

```
/users3/seatox/data/Pmicrops/Blastp/uniprot_sprot -max_target_seqs 1 -max_hsps 1 -outfmt
'10 delim=; qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue
bitscore sacc qcovs stitle' -evalue 1e-5 -num_threads 20 >
/users3/seatox/data/Pmicrops/Blastp/Pmicrops_Blastp_SwissProt.csv


#Blastp - SwissProt (Danio Rerio)

#!/bin/bash


#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=20

#SBATCH --mem=200G

#SBATCH --partition=hpc


module load blast-plus/2.12.0


makeblastdb -in
/users3/seatox/data/Pmicrops/Blast_Drerio/uniprotkb_Danio_rerio_AND_reviewed_true_2024_0
5_28.fasta -out
/users3/seatox/data/Pmicrops/Blast_Drerio/uniprotkb_Danio_rerio_AND_reviewed_true_2024_0
5_28 -dbtype prot


blastp -query
/users3/seatox/data/Pmicrops/Blast_Drerio/Trinity.Trinity.fasta.transdecoder.pep -db
/users3/seatox/data/Pmicrops/Blast_Drerio/uniprotkb_Danio_rerio_AND_reviewed_true_2024_0
5_28 -max_target_seqs 1 -max_hsps 1 -outfmt '10 delim=; qseqid sseqid pident length
mismatch gapopen qstart qend sstart send evalue bitscore sacc qcovs stitle' -evalue 1e-5
-num_threads 20 > /users3/seatox/data/Pmicrops/Blast_Drerio/Pmicrops_Blastp_Drerio.csv


#Hmmer

#!/bin/bash


#SBATCH --cpus-per-task=10

#SBATCH --partition=hpc


module load hmmer/3.4


hmmpress /users3/seatox/data/Pmicrops/Hmmer/Pfam-A.hmm
```

```
hmmscan --cpu 10 --seed 1 --domtblout Pmicrops_pfam.domtblout --tblout
Pmicrops_pfam.tblout /users3/seatox/data/Pmicrops/Hmmer/Pfam-A.hmm
/users3/seatox/data/Pmicrops/TransDecoder/Trinity.Trinity.fasta.transdecoder
```

UNIVERSIDADE NOVA
DE LISBOA

2024

Multilayer molecular networks shaping goby fish physiological responses to marine heatwaves in intertidal habitats

JOÃO DUARTE PIEDADE LOPES

MASTER IN **COMPUTATIONAL BIOLOGY & BIOINFORMATICS**

UNIVERSIDADE NOVA DE LISBOA