Norwegian University
of Life Sciences

Master's Thesis 2022 60 ECTS
Faculty of Biosciences

Low effect of β-mannan on gene regulation in Atlantic salmon

Marte Skog
M.Sc. Bioinformatics

ACKNOWLEDGMENTS

This master's thesis was developed as a part of the ongoing ImprovAFish project at the Norwegian University of Life Sciences (NMBU), which is a collaboration between Cargill, NMBU, Oslo Metropolitan University and Uppsala University. The project aims to investigate the effects of β-mannan additives in feed for Atlantic salmon. As well as increase the knowledge about the functional interconnections along the host-microbiome axis for salmon.

I would first and foremost like to thank my main supervisor Simen Sandve (professor, NMBU) for all the guidance during the last year. You have facilitated independent learning and execution of tasks, while being readily available for questions and help. You have also been a really funny and cool guy to work with. I would also like to thank my co supervisor Torgeir Rhoden Hvidsten (professor, NMBU), for help and guidance with the bioinformatics in my thesis. A special thank you to Arturo Vera Ponce de Leon (postdoctoral, NMBU) for helping me figure out my coding problems and patiently answering my questions. I have really appreciated your positivity and kindness.

Lastly, I would like to thank my family and friends for being awesome and cheering me on during the last year.

Marte Skog

SAMMENDRAG

I denne oppgaven undersøkes effekten av prebiotikaen β-mannan på den molekylære funksjonen i ulike vev hos Atlantisk laks. Fibret er tidligere vist å ha en positiv effekt på metabolismen hos andre landdyr og mennesker. Hypotesen er da at tilsetning av β-mannan i fôret kan øke produksjonen av kortkjedede fettsyrer i tarmen hos laksen. Dette kan ha en helsefremmende effekt hos verten siden immunsystemet styrkes.

I et eksperiment utført av forskergrupper ved NMBU testet man tre ulike fôr som inneholder β-mannan. Effekter av disse fôrene på genregulering i gjeller, lever og tarm både i ferskvann og sjøvann ble sammenliknet med en fisk som fikk kontrolldiet. For å finne genreguleringen ble både DESeq2 og EdgeR bruk som analyseverktøy og de ga henholdsvis 54 og 5 differensielt uttrykte gener på tvers av alle vev og dietter. For DESeq2 ble et lite antall gener knyttet til nedregulering av immun respons. Alt i alt viste resultatene at kun et lavt antall gener var differensielt uttrykte mellom de eksperimentelle fôrene og kontrolldietten. Dette indikerer liten eller ingen effekt av inklusjon av β-mannan på molekylære prosesser involvert i metabolisme.

SUMMARY

In this thesis the effect of the prebiotic β-mannan on the molecular function of different tissues in Atlantic salmon was tested. As the fiber has proven metabolism-promoting effects in other terrestrial animals and humans, it is hypothesized that the addition of β-mannans can increase the production of short chain fatty acids in the salmon gut. This could be advantageous to the host as it promotes immune system development, thereby improving overall health.

Three different β-mannan-containing diets were compared to a control diet in the liver, gill, and hindgut tissue for fish in fresh- and saltwater, using differential expression analysis. The analysis was performed with both DESeq2 and EdgeR giving respectively 54 and 5 differentially expressed genes across all diets and tissues. In the DESeq2 results a small number of genes tied to down regulated immune response were found. Overall, the analysis showed a low number of differentially expressed genes for all the β-mannan-containing diets, indicating that the addition of the prebiotic has no or little effect on molecular function linked to metabolism.

TABLE OF CONTENTS

1.INTRODUCTION

In 2018, 7% of protein sources consumed by humans came from aquacultural products, meanwhile the ocean covers around 70% of the world's surface (1). Going forward, a steady increase of the global population will in turn increase the demand for protein. Because of climate change, resources to increase the amount of land-based protein will most likely be scarce. This requires an expanded production of ocean-based protein to fill the demand (1).

Products from aquaculture is Norway's second largest export industry, with Atlantic salmon being the most produced species by far (2). With the rapid increase in demand for salmon, the salmon farming industry is constantly looking for new solutions that will allow expansion in production numbers while keeping product quality high.

Research concerning feed for the salmon farming industry is of great importance as 40-60% of total production cost goes towards the feed (3). Discovering new feed that could contribute to improved health and growth of the fish is imperative for the aquaculture industry to meet the increasing demand in production. The term functional feed is a new generation of food that should cause benefits beyond what a traditional feed would. Additives that give a health promoting effect are an important part of these feeds (3). Prebiotics are additives that have not yet been thoroughly studied in aquatic species (4). β-mannans are a promising group of prebiotics that have shown metabolic-promoting effects for both humans and animals (5). This positive effect raises the question: How will functional feed with added β-mannans impact the physiology and health of farmed Atlantic salmon?

1.1 Atlantic salmon

The Atlantic salmon (*Salmo salar*) is a fish belonging to the taxonomic family Salmonidae, which contains 66 recognized species (6).

1.1.1 The lifecycle of Atlantic salmon

Atlantic salmon is an anadromous fish that lives its first years in rivers, before they migrate to sea in search of better nutrition (7). When they are sexually mature, and it is time for them to spawn they return to freshwater (8). Atlantic salmon will return to their home/birth river with high certainty every migration cycle and only a small percentage of fish will stray to a different river. The fact that salmon return to their birth rivers to spawn creates strong genetic structure (population differentiation) and allows for selection for local river adaptations (7). Some Atlantic salmon never migrate and spend their entire lives in freshwater and this behavior could be a solution to optimize individual fitness (7).

From September to February the Atlantic salmon come to the rivers to spawn (7). During the spawning the females dig little holes in the riverbed and deposit their eggs into them. The males then fertilize the eggs, which will hatch in the spring. The newly hatched fish will stay buried in the riverbed for 3-8 weeks while getting nutrition from their yolk sack (7). When they emerge from the gravel and start their feeding, they are at their fry stage. This is one of the critical stages in a salmon's life, as absence of sufficient feed will quickly lead to starvation. If the fry survives this period, it enters a juvenile(parr) stage that lasts 1-8 years (7), before a smoltification process starts and prepares the salmon for a life in the ocean (7). The migration to sea is another critical stage as the salmon must adapt to a completely new biotic and abiotic environment, including higher water salinity, different food sources, exposure to novel pathogens, and increased risk of predation. The adult salmon spend 1-5 years at sea, before returning home to their river to spawn (7). A male Atlantic salmon can weigh up to 40 kg, while a female can weigh up to 20 kg (9).

1.1.2 Natural diet and habitat

The Atlantic salmon lives in the subarctic regions of the North Atlantic Ocean (7). In the northeast of the Atlantic Ocean salmon are found from the Barents Sea areas of Russia in the north to Portugal in the south. In the northwest Atlantic Ocean, they are found from rivers in Greenland in the north to New England in the United States in the south. Most historical attempts to introduce self-sustained Atlantic Salmon to areas outside of their natural habitat have failed. Many self-sustained populations are extinct or on the brink of extinction due to human impact (7).

In the wild, Atlantic salmon are natural carnivores and opportunistic feeders (10). This means that they can eat a wide variety of different prey and adapt to the food available in the territories they are in. During their time in the sea salmon can increase their weight up to 1000-folds, from 15-50g up to 1-25 kg or more. However, due to huge differences in the structure of fresh and sea-water habitat food chains, salmon diets vary dramatically through their different life stages (10).

In freshwater the available food consists mainly of invertebrates and the salmon often favor those floating on the top (11). At their juvenile stage this is the main food source and they have to eat to stay alive, but when adult salmon return to freshwater to spawn they are believed to stop feeding (11).

In the sea, Atlantic salmon are pelagic feeders reported to eat a wide variety of different organisms (10). They consume 40 different types of fish species from at least 19 families, along with invertebrates from at least 10 taxonomic groups (10). Sand eels, cod, plankton, herring and chapelin are often important components in the diet regardless of location and form the basis for the salmon's sea-water diet (10).

When the post-smolt first enters the sea, they hunt near the shore for the first couple of months (10). Here the salmon has a diet consisting of things like benic, insects, herring, sand eels and plankton. As the salmon gets older and larger, the complexity of their diet expands, and they gradually consume a larger variety of different fish and invertebrates as they age (10).

1.2 The aquaculture industry

Aquaculture is defined as the rearing of aquatic animals or the cultivation of aquatic plants for food (12). Here we find breeding and rearing of fish and shellfish, plus the cultivation of algae (13).

The aquaculture industry has had a rapid increase in production the last 30 years (14). In 2019 the amount of fish and shellfish produced globally was around 170 million tonnes. China is the leading manufacturer of aquacultural products, responsible for ~15% of the global aquaculture production. They are followed by Indonesia, India, and Russia. The amount of aquaculture production in fresh water is larger than in marine water. China is also the top exporting country in the world and here they are followed by Norway, Vietnam, and India. In 2019 aquaculture products worth around USD 160 billion were exported worldwide (14).

1.2.1 The aquaculture industry in Norway

The ocean-based aquaculture industry is Norway's second biggest export industry (2). Almost all Norwegian fish farming happens in sea water and the common farmed species are Atlantic salmon, rainbow trout, cod, halibut, and char. There is also farming of seaweed, scallops, cleaner fish, and fry for food production. The Atlantic salmon is by far the most farmed fish in Norway and the country is the world's biggest exporter of the species (2). In 2020 Norway sold just below 1,4 million tonnes of Atlantic salmon (15) and the export value was 74 billion NOK (2). Almost 100 countries receive Norwegian salmon, and the EU is the most important market, receiving ~70% of exported products (2).

1.2.2 Food production in the aquaculture industry

The amount of wild fish that has been captured for consumption has been steady at around 70 million tonnes for the last 20 years (1). In this period there has been a huge growth in the amount of farmed fish that has been produced (1).

The salmonids account for around 5% of the world's fish supply and Atlantic salmon is the biggest species in the group (1). The amount of farmed Atlantic salmon produced for consumption is much larger than the amount of wild salmon captured. Atlantic salmon is a sought after and versatile ingredient (1). It can be sold fresh or smoked, it can be used for sushi, and it can be processed into ready-made meals. It is a great source of protein and rich in omega-3 fatty acids, as well as vitamins and minerals. Compared to other animal proteins it has a low feed conversion ratio. Which means that the amount of edible meat per 100 kg of feed for the species is high. This makes the salmon resource efficient and more environmentally friendly that their animal counterparts (1).

1.2.3 Salmon Farming

The rearing of Atlantic Salmon began in Norway in the 1960s (11). Trails were set up to see if Atlantic salmon could be grown to marketable size in sea cages and as the trials were successful other countries like Canada, Australia, Chile, Island, and the United Kingdom followed. Most of the farmed salmon production today happens within latitudes 40-70° on the northern hemisphere and 50-60° on the southern hemisphere because salmon prefer colder weather (11).

The process of salmon rearing for food production can be divided into five stages. Fertilization, rearing of fertilized eggs, smolt production, ocean-based growth out and slaughter.

Fertilization of the eggs is the first step in the process (11). The aim of the breeding is to create salmon with favorable traits where robustness (disease resistant) and effectiveness (fast growing) are two traits that are important. The sexually mature fish are stripped of their eggs and milt (seminal fluid), and the eggs are fertilized (11). There are companies, like AquaGen that specialize in manufacturing and producing genetic material and fertilized eggs to be used in salmon farming (16).

The second stage in salmon farming is rearing of the eggs that have been fertilized (11). After fertilization the eggs are very delicate and must not be touched for a period of time. After this stage the dead eggs can be sorted from the live eggs, which are put into an incubator with water. Here they stay for another period, until they hatch. During the first period after hatching, they use their yolk sack for nutrition and stay in their hatchery until they are ready for their first feed. This process takes about two and a half months (11).

Then comes the third phase, smolt production (11). The fry are now transferred to fresh water tanks on land and they stay there for 8-16 months, while they are fry, parr and smolt. The type of freshwater tank used, varies depending on the production. A flow through system uses water that flows through the system in a single pass. If a recirculation system is used up to 99% of the water is recycled and this requires full biofiltering of the water to remove all harmful waste, gasses and toxins (11). A reuse system, filter and reuse around 50% of the water and add additional fresh water. Up till this point the salmon have been maintained in little to no light, but now light intensity increases, and the introduction of 8-to-12-hour dark periods is normal. This mimics the winter season that the wild salmon have, and it lasts for a minimum of 6 weeks, before they get 24-hours of daylight. In the freshwater tanks the salmon are given a diet high in protein and fat (11). There are companies that specialize in producing smolt from fertilized eggs that they buy from manufacturers (17). Examples of companies like this are sisomar and AS sævareid fiskeanlegg.

After the salmon has finished the smoltification, they are moved to new tanks, which are their final grow out site (11). They can either be net pens in the ocean or land based freshwater tanks, but ocean based grow out sites are the most common. The environment of the ocean-based tanks will vary based on its geographical location. Fish stay there for up to two years to reach their desired size. The feed given at this stage varies from farm to farm, but they are high in fats and protein (11). When the salmon has reached a desirable size, they are slaughtered and processed (1). Many companies produce their salmon by buying smolt from companies specializing in smolt production and letting them grow to full size in ocean-based facilities. Some of the biggest contenders in the salmon farming industry like Mowi, Nova Sea and SalMar are completing more of the rearing steps internally. Mowi completes the entire chain from fertilization to production themselves (18).

1.2.4 Ethical challenges

There are many ethical concerns tied to the aquaculture industry and salmon farming (19). Both environmental challenges and challenges tied to the health and welfare of salmon living in captivity (19).

For farmed salmon worldwide, there are many concerning health factors and the use of antibiotics is one of them (20). Overuse of antibiotics in humans and animals can provoke antibiotic resistance, leading to many bacterial infections being harder to treat and therefore a raised mortality (21). Norway has strict laws regarding the use of antibiotics for both humans and animals (21). Salmon louse is another big concern in the salmon farming industry (20). The parasite eats the skin and blood of the fish, and this decreases their health. The high mortality of farmed salmon (15%) is an issue, and most fish die from viral or bacterial infections or parasites like salmon louse. There are several vaccines made for farmed salmon to create increased resilience to many deceased, and more are being developed to help these issues (20).

The impact on the ecosystem and ocean environment around the facilities used for salmon farming is an environmental concern (20). Feed that has not been eaten and waste from the salmon fall to the bottom of the ocean. If the change of water and ocean currents are not strong enough, this buildup can cause chemical and biological changes in the ocean base. The conditions of the ocean base should therefore be controlled regularly (20). Offshore facilities, fully land based salmon farming and closed facilities at sea are some newer technologies explored to lessen the environmental footprint. Fish that has escaped from the ocean tanks is another issue. They go to the same rivers as the wild salmon to spawn and there are concerns that this will affect the genetic material of the wild salmon negatively (20).

The domesticated salmon are subjected to a lot of stressful procedures like handling, transport and chemical treatments that they would not face in the wild (19). On one hand the salmon live in enclosures, safe from predators. They are given food in abundance and are also given vaccines against several diseases. But they live in pens with high density, where diseases spread quickly, and they also lack room for exercise and normal stimuli (19). In Norway there are many laws in place that set standards and demands for the salmon welfare and the environmental footprint that salmon farming leaves (20).

1.3 Aquaculture feed for salmon

Feed is the largest cost in salmon farming, making up 40-60% of the total production cost (3). The feed also has a large impact on the fish health and in turn the quality of the finished product. Developing better feed is therefore of large interest to the industry and the term functional feed is an important part of this (3).

Traditionally the main ingredients in salmon feed have been fishmeal and fish oil (3). The growth in the aquaculture industry over the recent decades has increased the demand for feed and caused a price escalation in these maritime ingredients. Moreover, the global increase in production of aquaculture feed also has a negative impact on the environment and non-sustainable exploitation of marine resources (3). To mitigate this, fishmeal and oil has been exchanged for vegetable substitutes like soy, wheat, corn, beans, and peas. One problem with this solution is that salmon are natural carnivores and have limited capacity to digest starch. Soy is a highly studied substitute ingredient because of its high protein content, but a too high inclusion rate has shown to negatively impact health (3).

Functional feed is feed that induces benefits for the organism it is being fed to beyond what conventional feed would (3). While conventional feed covers all the basic nutritional requirements, functional feed should cause physiological benefits, promote growth and health, and strengthen the immune system beyond what a traditional feed would (3). One way to optimize feed within the functional feed framework in animal production systems is to include additives (22). A wide range of feed additives have been tested and the different additives have many different functions. Additives like acidifiers and exogenous enzymes enhance the digestibility of different feed materials and therefore enhance the animal performance. Prebiotics, probiotics, immune stimulants and phytogenics work by improving intestinal health, upping disease resistance and reducing stress (22).

1.3.1 Prebiotics

Prebiotics are fibers (a type of carbohydrate) that are normally derived from plants (23). Examples of prebiotics are inulin, fructo-oligosaccharides, mannan-oligosaccharides and β-glycans (4). The fibers are indigestible for animals/humans but will enhance beneficial gut bacteria because the bacteria can degrade them. They simulate growth and/or activity from one or more bacteria and can significantly alter composition of the gut microflora. The altered composition will hopefully increase the production of molecules with a positive effect on gut health, like short chained fatty acids. The use of prebiotics in aquatic feed has not been thoroughly investigated, despite the noted health and performance benefits for many terrestrial animals (4).

Addition of prebiotics to a feed could lead to many benefits (4). Prebiotics are shown to have a modulatory effect on the immune system and simulate both systemic and local immunity, boosting health benefits. They are also known to prevent pathogens by increasing humoral and cellular defense mechanisms. As well as simulating feed conversion ratios and increasing live weight gain (4).

1.3.2 β-mannan

β-mannan is a type of prebiotic that is a complex polysaccharide (24). Polysaccharides are an important part of hemicellulose in the cell walls. β-mannans can be sorted into four groups, linear mannan, galactomannan, glucomannan and galactoglucomannan. The β-1,4-linked back bone is shared between the four groups and mainly contains mannose, but there are occasionally substitutions of glucose (24).

One of the bacterias that have been found to be a primary degrader of β-mannan is Firmicute *Roseburia intestinalis*, which is a part of the human gut microbiota (5). The bacteria contain enzymes like *Ri*GH26, *Ri*GH36, *Ri*GH3A and *Ri*GH3B that help break down the fiber into monosaccharides. The monosaccharides enter the glycolytic pathway and are turned into pyruvate, which again is turned into Acetyl-CoA. Acetyl-CoA produces the end product butyrate, which is a shorted chained fatty acid (5). Along with Firmicute *Roseburia intestinali*, β-mannans have also been shown to simulate growth of a number of beneficial gut bacteria (25). These bacteria are bifidobacteria, bacteroides-Prevotella, clostridial cluster IX, and F. prausnitzii groups (25). When the bacteria are stimulated the metabolism production of butyrate and acetate (25). Like butyrate, acetate is a short chained fatty acid and they have both been shown to have a positive effect on energy metabolism (26) (27). As well as having anti-inflammatory attributes that strengthen the immunity of the intestine (26) (27).

β-mannan is found in varying concentrations in many feed ingredients like soybean meal, palm kernel meal and copra meal (28) and is also found in marine ecosystems (24). One example of this is the cuticle of the marine red alga *Porphyra umbilicalis* which contains β-mannan (24). Some terrestrial livestock has shown changes due to the inclusion of β-mannan in their diet (28). A study done previously at the Norwegian University of Life Science has also shown a change in pigs (29). Here no changes in health or growth were shown, but an increased relative butyrate:acetate ratio for the pigs being fed β-mannan (29).

1.4 RNA sequencing

A cell's physiology is defined by the types of gene products (protein and ribonucleic acid (RNA)) cells generate and utilize (30). Hence, by studying differences in gene regulation across contrasting conditions we can infer changes in cellular physiology indirectly. One efficient method to do this is by sequencing all messenger RNA molecules in a sample of cell/tissues, a method referred to as RNA-sequencing (30).

RNA is a component of the cell and has important tasks regarding the production of protein and gene regulation (31). RNA is made from DNA in a process called transcription. Unlike DNA the RNA is single stranded and contains the bases A (Adenine), C (Cytosine), U (Uracil) and G (Guanine). The bases pair together as U-A and C-G. In DNA U is exchanged with T (Thymine) (31).

Since the discovery of the RNA-sequencing (RNA-seq) technology it has since then become an indispensable tool in molecular biology (30).The rise of this technology has given new and better understanding of genomic functions. Since the start RNA-seq technology has evolved and it has become more time efficient and cheaper (30). The capabilities of RNA-seq will most likely continue to improve going forward. This means that the knowledge about both the quantitative and qualitative aspects of the transcriptome will continue to grow. There are several different companies producing technology for RNA-seq. Illumina is the marked leader, and the text going forward in this chapter will be based on their technology (30).

The first step in RNA-seq is isolation of RNA from the cells of the organism that is being researched and the RNA is then used for library prep (32). Constructing a high-quality library is important for the sequencing. The first step of the preparation is isolation of the RNA, followed by fragmentation of the RNA molecules. This is done to break up the RNA that is too long and the fragmentation methods can be physical, enzymatic, and chemical fragmentation. After the fragmentation the RNA strands need to be converted to double stranded DNA, because the DNA is more stable and can be amplified and modified easily (32). The next step is attachment of sequencing adaptors. First an A nucleotide is added to the 3’ end of both strands of DNA. The adaptors have a T nucleotide on the 3’ end and it

recognizes the A in the DNA strands and binds to them. The strands are now amplified by PCA and only the DNA molecules that have adapter sequences on both ends are enriched. This removes the strands where adaptors were unable to adhere. The last step of the preparation is a quality check of the library that has been produced. This is to verify the library concentration and fragment lengths (32).

Following sequencing library construction, the cDNA is now ready for sequencing (33). The sample material is loaded on to a flow cell where the sequencing is performed. On the surface of the flow cell, immobilized sequencing templates are found. The templates are designed to facilitate easy access for enzymes, while also ensuring stability and a low amount of binding for non-specific nucleotides. The principal behind the Illumina sequencing is known as sequencing by synthesis. The technology uses four different fluorescently labeled nucleotides and they can label tens of millions of clusters on a flow cell at once (33). When the sample is loaded on to the flow cell the adapters fitted during the library prep hybridize the single stranded DNA fragments to oligos on flow cell surface. The fragments are then amplified into clonal clusters. This happens through bridge amplification, where the strand bends over when attaching the adapter on the opposite side of the fragment to a different complementary oligo on the surface of the flow cell. Each stand is then amplified with DNA polymerase (33). The process is repeated to form copied clusters of the same strand and the template is ready for sequencing. During the sequencing process fluorescently tagged nucleotides compete for a spot on the chain. The four different bases are differently tagged and the base that is able to bind will be complementary to the base on the sample strand. When the first base is attached, laser excitation makes the tagged base emit a fluorescent signal that is recorded. The emission wavelength and intensity are used to identify the base. This process is repeated until all bases in the strand are identified and it happens to all strands adhered to the flow cell simultaneously. Typically, millions of sequences (referred to also as reads) are produced and used for downstream data analysis (33).

1.5 Bioinformatic software for RNA-seq

Data analysis is an imperative part of RNA-seq (27) and, depending on the aim of the study, can involve identifying patterns of gene expression changes or differential use of exons (differential splicing) (30).

After the RNA sequencing, the information is stored in FASTQ files, and this is the data that is used for further analysis (34). Both aligners, assemblers and QC tools can analyze the FASTQ format. Each file contains the base calls for all the different reads and each file can contain millions of entries. For each read, a single sequence is written to the file and each sequence gets four lines of information. The first line is an identification line with information about the sequence run and the cluster and on the second line the sequenced letters (the read) is found. The third line is a separator line with a plus sign and it can optionally contain further description. The last line contains the quality score for the base call, where each base in the sequencing order gets a numeric quality score of ASCII characters (34).

1.5.1 Mapping

One way of analyzing the RNA-seq data is mapping or alignment (30). Here all the reads in the FASTQ file are mapped to a reference transcriptome or genome. The transcriptome is made from an annotated genome or can be assembled de novo from RNA-seq reads. Each sequence read is associated to one or more genomic coordinates, which are positions on the reference transcriptome. Popular traditional tools used for mapping are TopHat, STAR and HISAT (30).

Recently new methods, known as pseudo mappers have been published (30). They are more computationally effective because they are “alignment-free” (31) i.e.they do not identify the exact base-to-base mapping between the read and the reference. Examples of popular pseudo-mappers are ‘salmon’ and ‘kalisto’ (27). Both use a k-mere based algorithm. Kalisto´s algorithm utilizes a de Bruijn graph of the reference transcriptome to identify likely matches for a read sequence, without alignment (35). A newer feature in salmon called selective alignment also utilizes a de Bruijn graph to make an index (36). This is an extension of the original Salmon method that uses so-called quasi-mapping and a two-stage inference algorithm (37).

After the reads have been mapped to the transcript the next step is quantification (30). Pseudo-mappers perform the quantification within their algorithm, while traditional mappers need an extra tool to quantify the transcript abundance after reads have been assigned a genomic location of origin. Some common quantification tools are RSEM, HTseq and MMseq. In the quantification step, transcript abundance for individual genes are estimated from counting the number of sequence reads that overlap annotated genes. This step creates an expression matrix, where the genes are rows, and

the columns are samples. The values in the matrix can be either the absolute read counts or the estimated relative transcript abundances (TPM - Transcripts Per Million), which are computed using the total number of reads from that sample and the transcript lengths (30).

1.5.2 Differential expression analysis

Differential expression analysis takes count data produced by mapping and tries to find differences in expression levels between different groups (30). The normalization is done as a part of the analysis and accounts for differences in read depth, technical biases and expression patterns. The three main methods of normalization are normalization by distribution, normalization by library size, and normalization by controls (38).

Most RNA-seq experiments are carried out to find transcription differences between sample groups (38). The groups can be divided by tissue, cells, organisms, or conditions. Differential expression analysis is used to decide which transcript features are likely to have changed their expression level between the different groups (30).

Two popular differential expression analysis tools, EdgeR and DESeq2 are based on negative binomial distributions and they provide comparable results and are computationally effective (30).

1.5.3 GO enrichment analysis

Gene ontology (GO) enrichment is a useful analytical approach to help interpret the molecular function and biological context of a gene set (39). This gene set can often come from a differential expression analysis and contain genes that are up or down regulated under different conditions. GO terms are assigned based on some experimental evidence, which are usually predicted based on homology to genes in other species and one gene can be annotated to more than one GO term (39). The term says something about the biological processes that the gene product contributes to, the molecular function of the gene product and the cellular components where the gene products are active. The output from the analysis is often a list of GO terms that are enriched in the group of genes compared to the background (typically all genes expressed in the tissue). Each term gets a p-value that is the probability of getting at least the observed number of genes annotated to the GO term in the geneset by chance. This means that the lower a p-value is, the more significantly enriched the GO term is in that specific gene group (39).

1.6 Aim of the Thesis

Improving growth and health of farmed Atlantic salmon is crucial to continue the rapid increase of the salmon production in the aquaculture industry. Developing feed that helps aid this purpose is therefore an important task. Many additives and ingredients have yet to be tested.

The primary aim of this project is to test if functional feeds with added β-mannan are impacting the Atlantic salmon gut, gill and liver physiology at the molecular level.

All lab work and sequencing were performed by others and FASTQ files produced in this process were utilized for the project. The methods used to achieve the thesis goal included quantification of gene expression levels, data visualization for quality control, differential expression analysis, and GO enrichment analysis.

2. MATERIALS AND METHODS

2.1 Experimental setup and samples used in project

The work on the project was an extension of an ongoing project at NMBU, called ImprovAFish. Samples were collected by researchers working on the project and the feeding trail was carried out by Cargill.

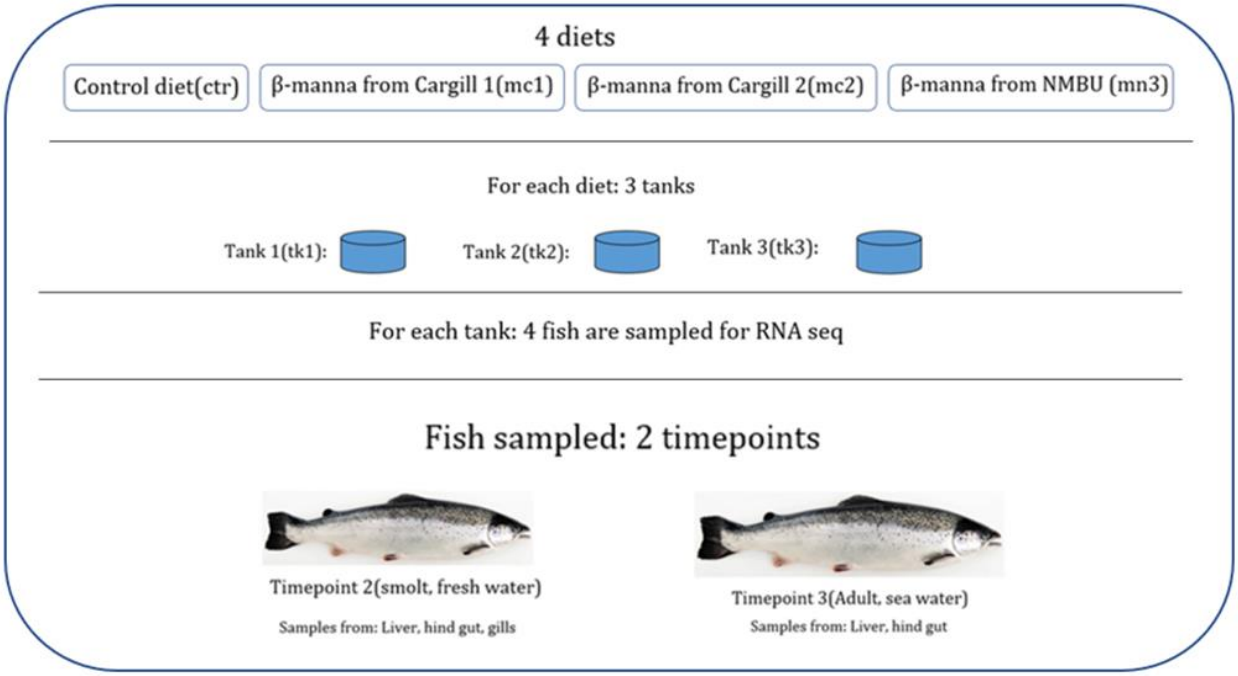

Figure 2.1: Visualization of the setup of samples used in the project. *Fish were sampled and used in the project were fed four different diets (one control and three β-mannan). For each diet, fish lived in three different tanks to exclude differences in tank environment as a source of error. For each tank, four fish were sampled. The project included fish sampled at 2 different timepoints, one at the smolt stage and one at the adult stage. In the smolt stage liver, hind gut and gills were sampled for RNA seq, while in the adult stage hind gut and liver were sampled for RNA seq.*

Figure 2.1 visualizes the sample flow in the project. Fish were fed 4 different diets, where one was a control diet and the other three were diets containing similar levels of β-mannan but slightly different polymerization levels of the mannan. All of the diets were made by Cargill, but NMBU produced one of the three β-mannan pre-biotika. The three experimental feeds all had a 0.2% inclusion of β-mannan. For each diet, fish lived in three different tanks. This was to create replicas and make sure that differences found between the diets were not due to differences in tank conditions. The fish were sampled at different time points. In addition to the time points in Figure 2.1 fish in the ImprovAFish project were also sampled at two additional time points. They were not included in the figure because

data from the time points were not used in this thesis. In time point 2 the salmon sampled were smolt that lived in freshwater. Tissue types sampled for RNA sequencing at this time point were hind gut, liver, and gills. In time point 3 the salmon sampled were adults that lived in seawater. Tissue types sampled for RNA sequencing at this time point were hind gut and liver. In each timepoint four fish from each tank were sampled.

2.2 Preprocessing of samples and lab work

The preprocessing of samples and the lab work for the data used in the project were performed by others.

The first step was to perform an RNA extraction. The samples were preserved in DNA/RNA SHIELD™, obtained by Zymo Research and this was done following the Zymo Research standard procedure. The RNA extractions were executed at the Laboratory of Genomics and Molecular Medicine, Department of Biology, University of Copenhagen, Denmark. The Zymo Research Quick-DNA/RNA (Cat. D2131) was used to isolate the total RNA from liver, gills, and hindgut and this was done according to the manufacturer's instructions. RNA concentration and purity were found using a Nanodrop 8000 (Thermo Scientific, Wilmington, USA) and the RNA integrity was checked by using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The samples with a RIN (RNA integrity number) equal or above two were used. Before the analysis took place, all samples were randomized.

Library preparations were done by Novogene (Beijing, China) and a TruSeq Stranded mRNA kit (Illumina, San Diego, CA, USA) was used, as per manufacturer's protocol. PolyA containing mRNA molecules were isolated from the total RNA, using magnetic oligo dT beads. This was done according to the polyA selection method, and the isolates were then fragmented with the fragmentation buffer. Cleaved RNA fragments were primed with random hexamers into first-strand cDNA using reverse transcriptase and random primers. After this a double-stranded cDNA was synthesized and end-repair, phosphorylation and 'A' base addition, and adapter ligation was performed according to the Illumina's library construction protocol. Libraries were sequenced on the Illumina NovaSeq 6000 platform at Novogene, (Beijing, China), using 300 bp paired-end sequencing.

2.3 QC of sequencing data

As the first step in the bioinformatic analysis of the sequencing data a quality control was performed. FastQC was used to perform quality control of the Illumina sequencing data. The software takes raw reads in fastq format as input (40) and the tool was run on the fastq files for both the forward and reverse reads separately. To summarize the QC analysis results from multiple samples together, MultiQC was then used (41). The results from MultiQC were both used directly and visualized using the ggplot2 package (42).

2.4 Read mapping and quantification

To assign RNA-seq reads to genes, we used the pseudo mapper salmon version 1.5.2 (36). Pseudo-mappers such as Salmon quantifies transcript abundance from RNA-seq reads much faster than the traditional alignment tools and has a high level of accuracy (35). The next sections countais step-by-step description of the different steps to perform read mapping and gene expression quantification.

2.4.1 Creating the transcript and the decoy file

The first step in the mapping process was creating a transcript file, as well as a decoy file. A transcript file was created in a bash script using the program gffread (43). The program could perform many operations on GFF files, like filtering, conversion and verification (43).

To create the transcriptome consisting of the spiced exons for each transcript, gffread was run using the -w option (43). As input the program needed a fasta file for the genome and a GFF file in gff3 or gtf format. The input files were the Salmo salar (Ssal) annotation of V3 genome assembly from rapid.ensembl.org, with accession number: GCA_905237065. To remove unwanted/unneeded instances of “transcript:”, the Linux command sed was used on the data created by gffread and replaced them with nothing.

For this project it was chosen to create an index using a decoy aware method to reduce false mapping (36). Sequencing fragments from unannotated loci could often resemble annotated transcripts. This meant that they could map to the annotated regions and cause false mapping. When salmon was supplied with a decoy sequences the algorithm could take special care when working with these sequences, preventing false mapping. There were two options when creating a decoy aware index. One was using a tool like MashMap to detect decoy sequences. The second method, used in this project was generating a decoy from the entire genome. This would demand more computational power then option one (36). A decoy file was built, using the input fasta file and all the sequence names were extracted for a decoy file that contained all IDs for the sequences of the genome. A concatenated file of the genome and transcriptome file was also made.

2.4.2 Pseudo mapping with the salmon software

To run salmon an index was created. The index would work a bit like the index of a book. A data structure was built based on the reference transcriptome, to make looking for the potential origin of a sequence on said transcriptome more efficient. Salmon built an index from a de Bruijn graph based on k-mers from the reference transcriptome, the same indexing structure used by pufferfish (44).

The salmon indexer was run to build a decoy aware index, using the decoy file and the concatenated transcript and genome fasta file created earlier. This index was used on all samples in the project and only had to be created once. The index data was stored in a folder using the -i option and the -k option was set to 31. This meant that the minimum length of base pairs in a k-mere that could have gotten a valid match between the reference and the samples was 31 (45), as recommended for sequencing reads over 75 base pairs. The keepDuplicates option was also included and it turned off the default setting to discard sequence identical transcripts, but instead retained and quantified them separately (45).

The salmon quantification was run in a mapping-based mode using the - -validateMapping flag which minimizes mapping errors and results in higher sensitivity (36). The quantification step in the project was completed using array scripts. All samples for each tissue type were run using one script, ten samples at a time. The sample IDs were fed to the script through a sample info file, containing a table. The awk linux function was used to extract the first column of the table that contained the sample IDs. Awk was then used to go through each of the sample IDs in the column and run the salmon quantification for the fastq files belonging to the sample. The salmon quant was run with the index created previously as -i. -1 and -2 were the forward and reverse reads for each sample and -o stated the output directory. The library type(-l) was set to ISR, which meant that the samples were a stranded pair ended library where the reads face each other and read 1 came from the reverse strand and read 2 came from the forward strand.

The quantification performed by salmon is called selective alignment and worked by first finding consistent chains of unique maximal exact matches(uni-MEMs) between the index and the sample sequences and chaining them together (36). After this the potential alignments were scored and grouped together. In the first phase the k-mers in the reads were compared to the index in a k-mere lookup. Matches found were extended to the maximum possible length. The uni-MEMs were then projected to their corresponding reference loci and the exact matches were noted. The reference transcript and orientation for each uni-MEM were collected and the transcript/orientation pairs with a too low number of matches were scrapped. All matches on each transcript were then sorted and compressed (36). A chaining algorithm was applied to the exact matches to decide the mapping locations and merge them together. The mappings were now scored using the ksv2 library, and the alignments were put into groups based on their scores. Alignments with scores that fell below a set

threshold were scrapped as well as fragments which were best aligned to a decoy sequence. The valid mappings were summarized and used for quantification (36).

The main output from salmon was the quant.sf files (46). The text files contained a table with a header line of five columns. The five columns were:

- Name: Contained the name of the target transcript
- Length: Contained the length of the transcript in bases
- EffectiveLength: A computed effective length based on all factors that could affect the probability of sampling fragments from this transcript.
- TPM: Stands for transcript per million and is salmon's relative abundance measure.
- NumReads: The number of reads mapped to the transcript during quantification.

Quant.sf files were used for downstream analysis (46).

2.5 Tximport

Quant files for all the samples were imported into R for further data analysis using the R package *tximport* (47). The package estimated counts and transcript lengths from the quantification data and converted transcript-level quantification to gene-level quantifications (47).

First a transcript database was made using the makeTxDbfromGFF() function in tximport. Transcript IDs (TXNAME) and gene IDs (GENEID) were extracted, making a dataframe with the matching IDs.

To create a matrix with all samples from one tissue, a list of paths for the quant.sf files for all samples in the tissue was made. Each path was then named with the corresponding sample IDs. This created a list of named files that were used as input in the tximport() function. The transcript database created earlier was also used as the tx2gene input and the type feature was set to salmon.

The tximport function created a list that consisted of four elements (48). Three of the elements are matrices containing abundances, counts and length of the average transcript for each gene and they had genes as rows and samples as columns. The fourth element was called countsfromAbundance and contained the character arguments used to call tximport (48).

2.6 Differential expression analysis with DESeq2

The goal of the differential expression analysis was to find which genes were differential expressed between each diet containing β-mannan and the control diet in each tissue and time point. DEseq2 (49) and EdgeR (50) were used to evaluate quantitative differences between contrasting groups in the data. Both processes were very similar, so only the procedure using DESeq2 for differential expression analysis is explained below.

To start the differential expression analysis a DESeq dataset was made using the tximport elements made previously and sampleinfo tables. The info tables contained four columns, sample IDs, diet, tank and sampling time point. Sampleinfo tables for hindgut and liver had a fifth column called group, that contained information about both time point and diet. The extra column was used to compare the β-mannan diets with the control diet from their corresponding time point. The DESeq2 package contained a specialized function to make DESeq datasets from tximport data called DESeqDataSetFromTximport(). This function was run to create datasets for each tissue type with the corresponding sample info files and tximport elements. The design of the statistical test was diet~group, where diet was the four different feed treatments and group was the combination of time points and tissues.

Then the DESeq data sets were prepared for differential expression analysis. First non-expressed genes were filtered from each dataset to lighten the computational load. This was done by only keeping rows where the row sum was larger than zero. The data was then normalized and visualized to remove outliers. Variance stabilizing transformation was the chosen normalization method and the function for this in the DESeq package was used. The function created normalized matrices where the variance was constant over the mean values, while also accounting for library size (51). Box plots and PCA plots were now created from the normalized data to visually detect outliers. The PCA plots were created by the DESeq function plotPCA and by default, PC1 and PC2 were plotted. The function made a PCA plot based on only the 500 genes with highest row variance by default (52). Outliers were then removed from the DESeq datasets, and a new round of normalization and visualization was performed to ensure that the correct samples were cut.

The DESeq datasets were now ready to be fed into DESeq2s function that performed differential expression analysis. By default, the function performed the analysis in three steps, estimation of size factor, estimation of dispersion and general linear model (GLM) fitting followed by Wald statistics (49). The goal of the first procedure is to correct both for differences in library size and RNA composition bias. First a normalization was executed by calculating the geometric mean of all genes across the samples in the datasets. All counts in the count matrices were then divided by the mean and the median of these numbers for each sample were the size factor for that sample. In the second step

the data went through a variance adjustment (49). In this process genes were grouped with other genes of similar expression strength, to create a larger number of replicates and a more precise estimated variation. The accuracy of the estimated dispersion was important for the statistical calculations in the downstream analysis. The dispersion estimates were then calculated using a model fit procedure. A maximum likelihood method was used to find gene-wise dispersion estimates and a trend line was fitted to the estimates (49). The final dispersion values were obtained using Empirical Bayes shrinkage to shrink the gene-wise dispersion estimates towards the trend curve values. In the third part a negative binomial general linear model was fitted for each gene (49). The null hypothesis used by DESeq2 was that there was no significant difference in expression between samples belonging to two groups and Wald test was used for hypothesis testing. First a z-statistics was made by dividing the shrunken log fold changes by their standard error. The z-statistics was then compared to a normal distribution and p-values were computed for each gene. If a p-value was small enough the null hypothesis was rejected. The p-values were also independently filtered by a Benjamin and Hochberg procedure to account for multiple testing. A likelihood ratio test was also available as an alternative approach to the function, but Wald test was chosen (49).

The results from the differential analysis were obtained using the result() function in the DESeq2 package. Results for each contrast were found and all β-mannan diets were compared to the control diet in the different tissues and timepoints.

For contrast with a higher number of differentially expressed genes (DEGs) heatmaps were drawn. The heatmaps only contained the DEGs and samples from the specific contrast. These data were extracted from the normalized count matrices and the pheatmap package was used to draw the plots.

2.7 GO enrichment

To test if particular biological processes and molecular functions were overrepresented a GO enrichment analysis was performed. The analysis compared the predicted functional attributes of a list of genes of interest (in our case the differential expressed genes) with the functional attributes of a universe of genes (e.g. all genes expressed in a tissue we study). This enables analyses and interpretations of the biological significance of changes in gene expression between dietary treatments.

The function enrichGO() from the cluster profiler package was used for the GO enrichment analysis. An annotation database was needed as input for the GO enrichment function. The package Annotationhub provided access to genomic files and information from web resources like Ensemble (53). First all information on the hub was retrieved, before the information specific to the salmo salar was extracted. A universe to serve as the background for the analysis also needed to be made. For the

universe, genes from the count table for the specific tissue were used. All genes with row sums over 10 were included in the universe. The gene IDs were extracted and changed to their corresponding NCBI gene IDs. The gene IDs for the DEGs were also extracted and changed to the corresponding NCBI gene IDs. Genes where corresponding gene IDs did not exist were removed.

The NCBI gene IDs for the DEGs, the universe created, and the annotation database were then used as input for the enrichGO() function from cluster profiler (54). The ontology parameter was set to biological processes and the GO terms for each differential expression gene group were then visualized.

3. RESULTS

3.1 Overview of data and quality control

To get an overview of the quality of the RNA-seq data and identify potential outliers we ran FastQC (40) and carried out visualization of sample similarity across tissues, feed type, and sampling times using a PCA approach. In the sections below results from these analyses are presented.

3.1.1. Data quality control using FastQC

The mean number of sequencing reads per sample across all tissues were 49 million. There were some systematic differences in sequencing depth among tissues, with gills having the most reads per sample (mean = 55 million) and hindgut having the least (mean = 40 million). Only 8 samples had fewer than 20 million reads (Figure 3.1) and these were from the hindgut tissue. Furthermore, all samples had high base call accuracy with an average phred score > 35 (Figure 3.2). Taken together, this dataset represents high quality sequencing data for further analyses of gene expression differences.

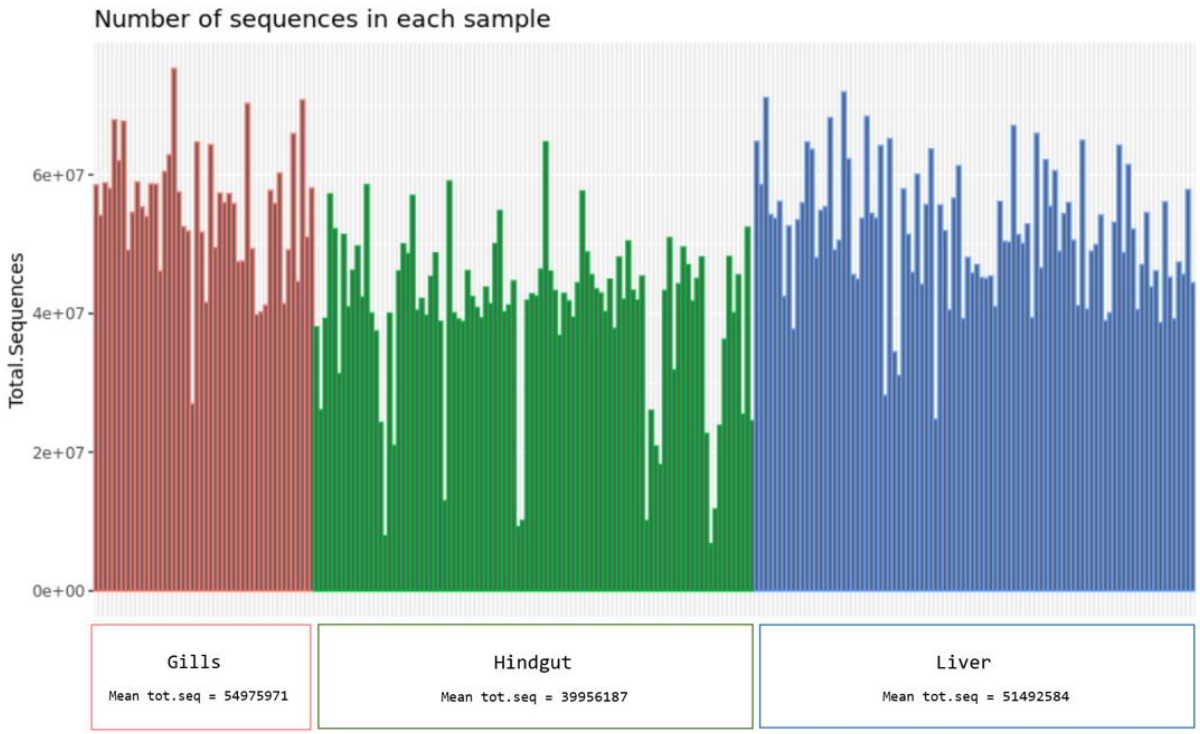

Figure 3.1 Barplot visualizing the amount of sequences in all input samples: *The total amount of sequences in each sample was a bar on the barplot, with total sequences on the y axis. The total sequence amount from the corresponding forward and reverse fastq files were combined by taking the mean, creating one bar for each sample. Each bar was colored by a specific color for the tissue it came from. Gill samples were colored pink, hindgut samples green and liver samples blue. The mean total sequences in each tissue were 549775971 for gills, 39956187 for hindgut and 51492584 for liver.*

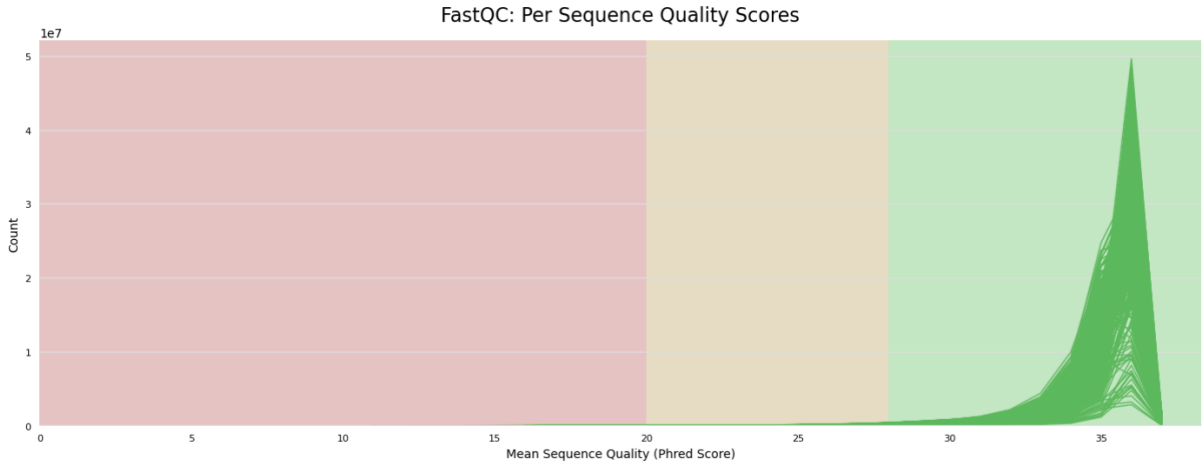

Figure 3.2 Plot of the per sequence quality scores of all input samples: *The graph was created by* ***MultiQC*** *and consisted of one graphline for all fastq files that were run through FastQC. It gave the mean sequence score on the x axis and the number of sequences with that score on the y axis. If MultiQC had colored a graphline green it meant that all sequences in the fastq file were of high quality, and in the plot the lines for all samples were green.*

3.1.2 QC of samples using PCA plots

To get a better overview of data quality and identify potential sample outliers we performed a series of explorative PCA analysis visualizations. First the separation of samples by tissues was assessed (Figure 3.3). In general this showed clear separation of samples by tissue as expected. The hindgut and liver tissues were separated by PC1, while gills were separated from the other two tissues by PC2. However, one gill sample was separated from the rest of the gill cluster by PC2 and lies between the hindgut and liver clusters (Figure 3.3). This extreme outlier was removed from further analyses in the project.

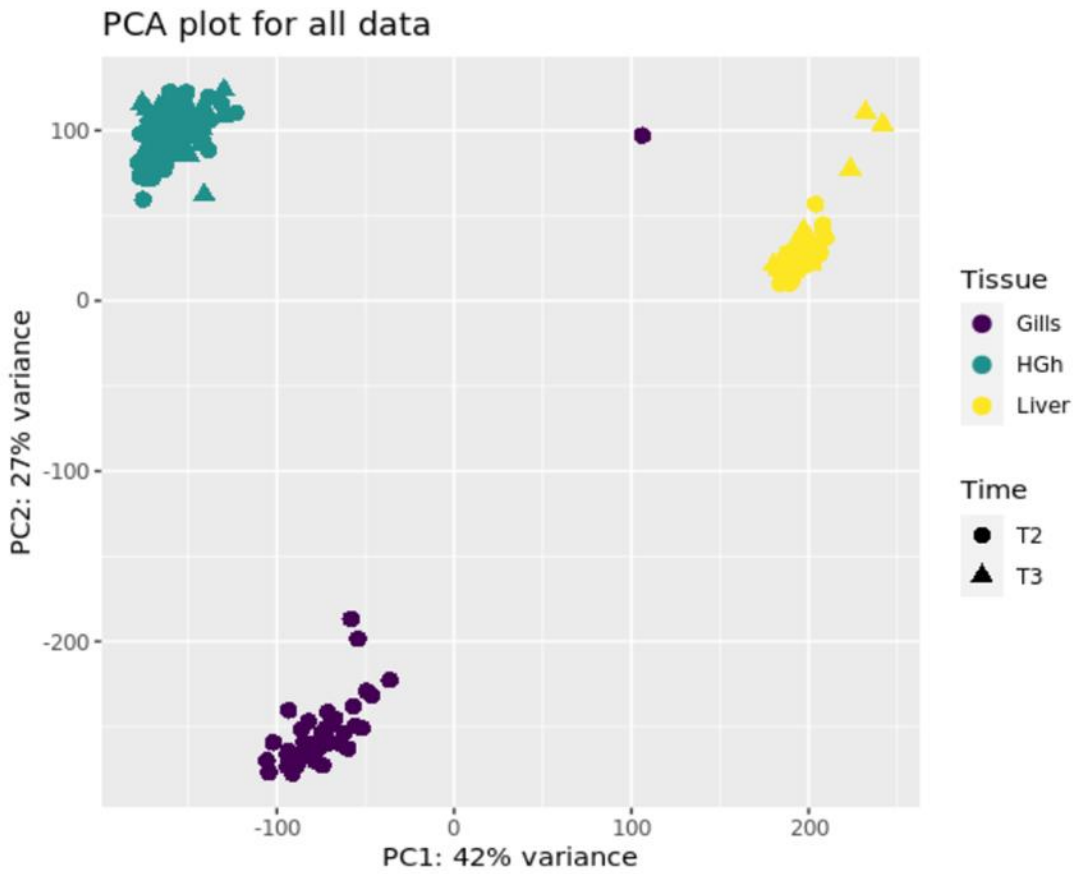

Figure 3.3 PCA plot showing all samples used for the project: *Samples were plotted along the PC1 and PC2 axis. Samples were colored by the tissues. Gill tissue =purple, liver tissue = yellow and hindgut tissue = green. Shapes indicate sampling time point. Sampling time point 2 = circles, sampling time point 3 = triangles.*

The PCA results for each individual tissue were then investigated. For the gill data most samples cluster together, but the plot showed a few outliers (Figure 3.4). Two outliers were removed from the dataset before further analysis was carried out. The lack of separated clusters shows that the PCA detected no strong association between variation in gene expression between samples from different diet groups. The liver data showed separation of the samples into two clusters (Figure 3.5). The clustering was driven by differences between the two sampling time points. The plot also showed individual outliers along PC1 and six of these outliers were removed from the dataset prior to downstream analysis. Furthermore the PCA shows no separation between the different diets in each time point and this indicates that a low fraction of the gene expression variation is associated with different responses to diets . For the hindgut data no sampling time clusters were seen in the visualization of the PCA (Figure 3.6). One outlier was removed from the dataset prior to further analysis, slightly separated from the rest of the cluster by PC1. The lack of separation seen between samples from the different diets could be an indicator of lacking variance in gene expression between samples from the different groups.

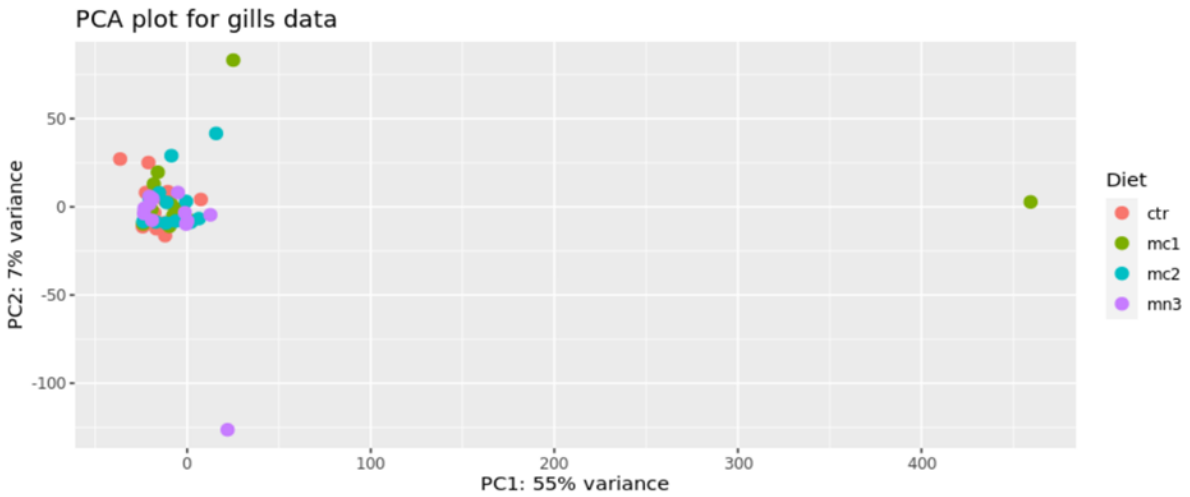

Figure 3.4 PCA plot showing all samples from the gill tissue: *Samples were plotted along the PC1 and PC2 axis. Samples were colored by the diet. Control diet =orange, mc1 = green, mc2 = blue and mn3 = purple.*

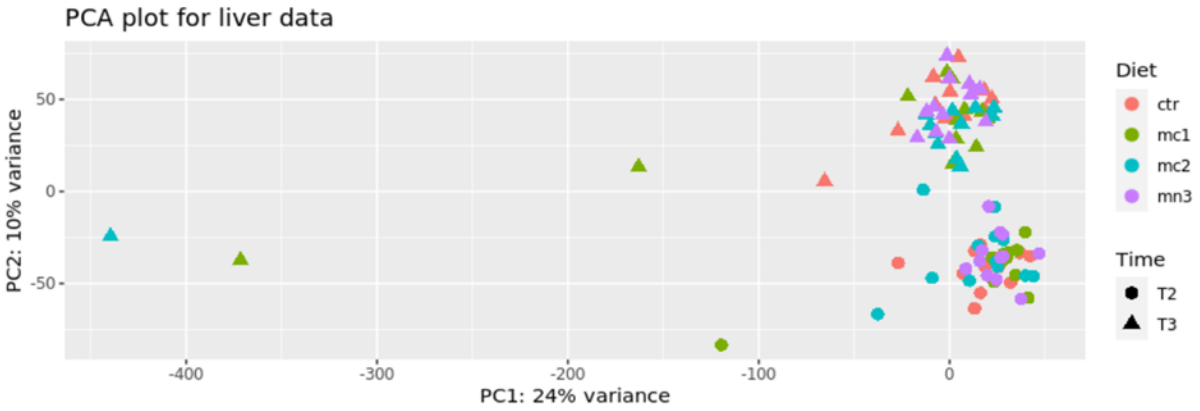

Figure 3.5 PCA plot showing all samples from the liver tissue: *Samples were plotted along the PC1 and PC2 axis. Samples were colored by the diet. Control diet =orange, mc1 = green, mc2 = blue and mn3 = purple. Shapes indicate sampling time point. Sampling time point 2 = circles, sampling time point 3 = triangles.*

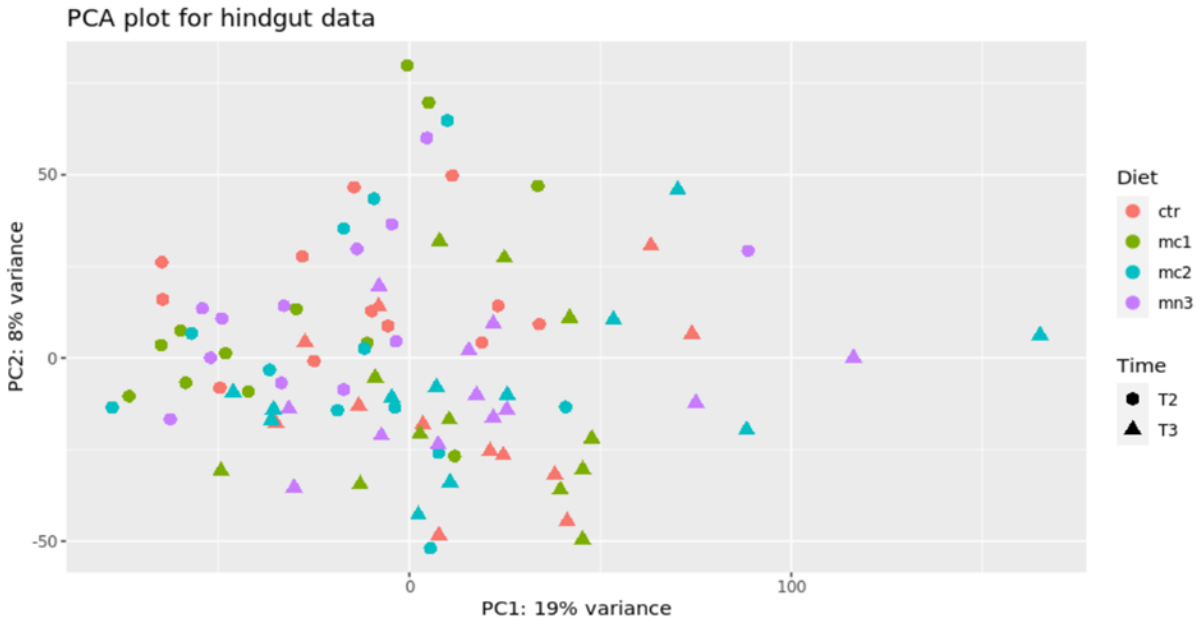

Figure 3.6 PCA plot showing all samples from the hindgut tissue: *Samples were plotted along the PC1 and PC2 axis. Samples were colored by the diet. Control diet =orange, mc1 = green, mc2 = blue and mn3 = purple. Shapes indicate sampling time point. Sampling time point 2 = circles, sampling time point 3 = triangles.*

3.2 Mannan pre-biotics induced little changes in gene expression

3.2.1 DEG analysis using DESeq2

To test for impact of mannan supplement in feed a Wald's test was performed by Deseq2 for differential gene expression between experimental diets and the control feed in the different tissues. In total 54 differentially expressed genes were identified across all contrasts, and 44 were annotated through matching their gene IDs with an annotation table. For the gill tissue, only 1 DEG was found, and this was for the mc1 diet (Table 3.1). In the hindgut tissue there were only identified 2 DEGs in time point 2, and these were found for the mc1 diet (Table 3.1). In time point 3 there were 1, 3 and 2 DEGs found for the mc1, mc2 and mn3 diets, respectively (Table 3.1). For liver, 1 DEG each was identified for the mc2, and mn3 diets in time point 2 (Table 3.1). The biggest effects were seen for liver in time point three and here, 18 and 25 DEGs were found for mc1 and mc2 diets, respectively (Table 3.1). Overall, a low number of DEGs were found, indicating that the addition of β-mannans at industry relevant levels (0.2%) in feed for Atlantic salmon causes little changes in the molecular function of the tissues.

Table 3.1 Table of differential expressed genes from DESeq2: *The table contains information about the DEGs identified with DESeq2. Samples from each β-mannan diet were compared to the samples from the control diet. Hindgut and liver tissues were sampled at two time points. Tissue, time point and diet for the gene is displayed in the first three columns. The fourth and fifth columns contain the adjusted p-value and the log2foldchange. The gene product is displayed in the last column, and they were found from matching the geneIDs, to an annotation table. 10 of the 54 differential expressed genes had no annotation in the table.*

Tissue	Time	Diet	Padj	log2FC	Product / Human readable gene name
Liver	2	mc2	0,029	4,0	protocadherin-11 X-linked-like isoform X1
Liver	2	mn3	0,013	1,5	phosphatidylinositol 3-kinase regulatory subunit gamma-like
Liver	3	mc1	0,035	1,2	nuclear receptor subfamily 1 group D member 1-like
Liver	3	mc1	0,006	-1,2	glutathione S-transferase P
Liver	3	mc1	0,005	-1,0	glutathione S-transferase omega 1
Liver	3	mc1	0,035	-0,8	glutamate--cysteine ligase regulatory subunit-like

Liver	3	mc1	0,048	-1,0	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 8, mitochondrial-like isoform X1
Liver	3	mc1	0,035	-1,2	cytochrome c oxidase subunit 6A1, mitochondrial-like
Liver	3	mc1	0,003	-0,8	ubiquitin carboxyl-terminal hydrolase 4-like
Liver	3	mc1	0,035	-1,1	proteasome subunit alpha type-7-like
Liver	3	mc1	0,035	-1,3	proteasome maturation protein-like
Liver	3	mc1	0,035	-1,0	nucleoside diphosphate kinase 3
Liver	3	mc1	0,026	-1,0	serine/threonine-protein kinase VRK3
Liver	3	mc1	0,006	-1,3	splicing factor 3B subunit 5 isoform X1
Liver	3	mc1	0,048	-1,2	28S ribosomal protein S14, mitochondrial
Liver	3	mc1	0,028	-0,6	actin-related protein 2/3 complex subunit 4-like
Liver	3	mc1	0,006	-3,0	heme oxygenase-like
Liver	3	mc1	0,007	-0,9	akirin 2(2b) isoform X1
Liver	3	mc1	0,003	-1,0	-
Liver	3	mc1	0,03	-1,8	-
Liver	3	mc2	0,042	0,7	plasma protease C1 inhibitor-like
Liver	3	mc2	0,017	0,8	complement C1r-A subcomponent-like isoform X1
Liver	3	mc2	0,017	1,0	protein disulfide-isomerase
Liver	3	mc2	0,047	0,6	Sequestosome-1
Liver	3	mc2	0,017	0,5	transport protein Sec61 subunit alpha

Liver	3	mc2	0,047	-1,0	cysteine dioxygenase type 1-like
Liver	3	mc2	0,047	-0,9	cysteine dioxygenase type 1
Liver	3	mc2	0,047	0,6	glutamate--cysteine ligase catalytic subunit isoform X1
Liver	3	mc2	0,049	-0,7	ADP-ribosylation factor 6-like
Liver	3	mc2	0,047	-0,4	ras-related protein Rab-13
Liver	3	mc2	0,047	-0,8	serine/threonine-protein kinase ICK-like isoform X1
Liver	3	mc2	0,046	-0,9	NADH dehydrogenase 1 alpha subcomplex subunit 4
Liver	3	mc2	0,047	-1,0	cytochrome c oxidase subunit 6B1-like
Liver	3	mc2	0,047	-0,6	Charged multivesicular body protein 3
Liver	3	mc2	0,047	-0,5	Cathepsin B precursor
Liver	3	mc2	0,017	-0,4	insulin-like growth factor-binding protein-like 1
Liver	3	mc2	0,047	-0,7	Ester hydrolase C11orf54 homolog
Liver	3	mc2	0,017	-0,4	peripheral myelin protein 22
Liver	3	mc2	0,047	-0,5	PDZ domain-containing protein 7
Liver	3	mc2	0,017	-0,6	FUN14 domain-containing protein 2
Liver	3	mc2	0,047	-0,6	-
Liver	3	mc2	0,049	-1,0	-
Liver	3	mc2	0,017	-0,8	-
Liver	3	mc2	0,047	-0,9	-

Liver	3	mc2	0,047	-0,5	-
Hindgut	2	mc1	1,30e-03	3,8	uridine phosphorylase 2-like
Hindgut	2	mc1	1,10e-08	5,8	probable lipid phosphate phosphatase PPAPDC3
Hindgut	3	mc1	3,50e-08	-24,3	-
Hindgut	3	mc2	3,60e-08	-17,0	histidine-rich glycoprotein-like
Hindgut	3	mc2	3,10e-03	-5,0	ADP/ATP translocase 2-like
Hindgut	3	mc2	6,80e-07	-23,1	-
Hindgut	3	mn3	2,10e-16	-17,8	complement C3
Hindgut	3	mn3	4,50e-08	-23,9	-
Gills	2	mc1	0,006	-1,8	BTB/POZ domain-containing protein KCTD3-like isoform X1

3.2.2 Heatmap visualization of DEGs

The heatmap showing gene expression levels of the DEGs identified, confirms the overall results from the differential expression analyses, but also highlights DEGs that could be statistical artifacts (Figures 3.7). Most rows in Figure 3.7 show a general pattern of decreased expression in mc1 and mc2 compared to ctrl at time point 3 in liver (Figure 3.7) in line with the DEG results (Table 3.1). However, examples like the two first genes (i.e. rows) reflect little systematic difference in expression levels across samples, and it is possible that these significant DEG results are driven by spurious variation across individuals within each diet group.

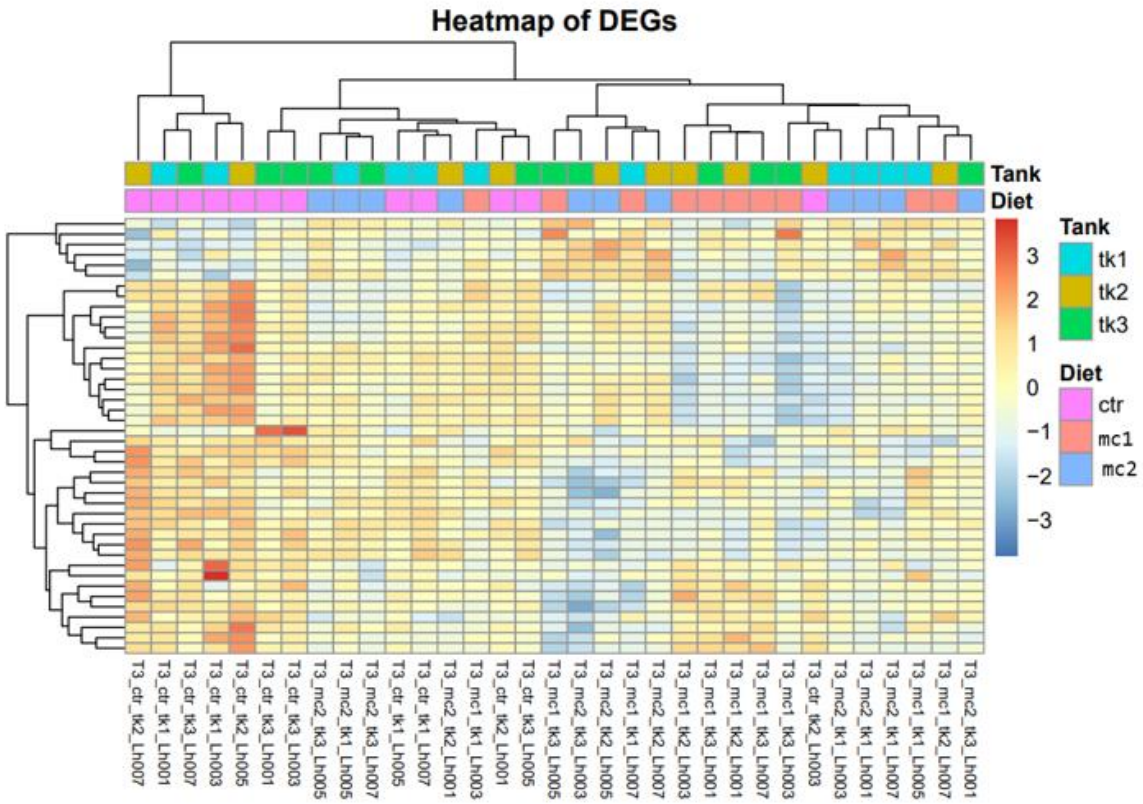

Figure 3.8: *The heatmap is made from the normalized count data for the DEGs from the mc1 and mc2 diet in time point 3 in liver. The rows are the DEGs, and the columns are the samples. Colors represent row-scaled expression levels, where blue indicates lower and red indicates higher expression. At the top of the plot annotational columns for both tank and diet were added. Tank 1= green, tank 2 = yellow, tank 3 = blue. In annotation track 2 diet is indicated by red=control, pink = mc1 and blue = mc2.*

3.2.3 DEG analyses using EdgeR

Because very few DEGs were found using DESeq2 (49), we also wanted to test a different algorithm to see if this would reveal additional DEGs. An additional differential expression analysis was therefore performed using EdgeR (50). However, this analysis resulted in even fewer DEGs (Table 3.3). The gills tissue only had 1 DEG, and this was for the mc1 diet (Table 3.3). In Hindgut 1 DEG was identified, this for the mc1 diet in time point 2 (Table 3.3). The analysis found 4 DEGs for the liver tissue, and they were all for the mc2 diet in time point 2 (Table 3.3). This supports the results found with DESeq2 that the addition of β-mannan in the feed has a low impact on the molecular function of gills, liver and hindgut.

Table 3.3 Table of DEGs from EdgeR: *The table contains information about the DEGs identified with EdgeR. Samples from each β-mannan diet were compared to the samples from the control diet. Hindgut and liver tissues were sampled at two time points. Tissue, time point and diet for the gene is displayed in the first three columns. The fourth and fifth columns contain the adjusted p-value and the log2foldchange. The gene product is displayed in the last column, and they were found from matching the geneIDs to an annotation table.*

Tissue	**Time**	**Diet**	**p-value**	**logFC**	**Product**
Liver	2	mc2	1,40e-08	7,2	protein fosB isoform X4
Liver	2	mc2	2,10e-07	5,4	protein fosB isoform X2
Liver	2	mc2	7,40e-07	-5,4	sperm acrosome membrane-associated protein 4-like
Hindgut	2	mc1	4,80e-07	3,7	uridine phosphorylase 2-like
Gills	2	mc1	6,20e-07	-7,3	neuronal acetylcholine receptor subunit alpha-2-like

3.2.4 GO enrichment

To help biological interpretation of the DEGs, a GO enrichment analysis was performed using results from contrasts that produced >15 DEGs (i.e. mc1 vs ctr and mc2 vs ctr from liver in time point 3).

For the mc1 vs ctr contrast, five GO terms are shown to be significantly enriched (p-adjust < 0.05), meaning that genes involved in these biological processes were down regulated in liver in the mc1 diet at time point 3. These terms included four terms related to different cellular metabolisms and one to oxidative phosphorylation (Figure 3.9). The term with the most genes was peptide metabolic process (4 genes), while amino acid metabolic process, sulfur compound metabolic process and glutathione metabolic process all had three genes.

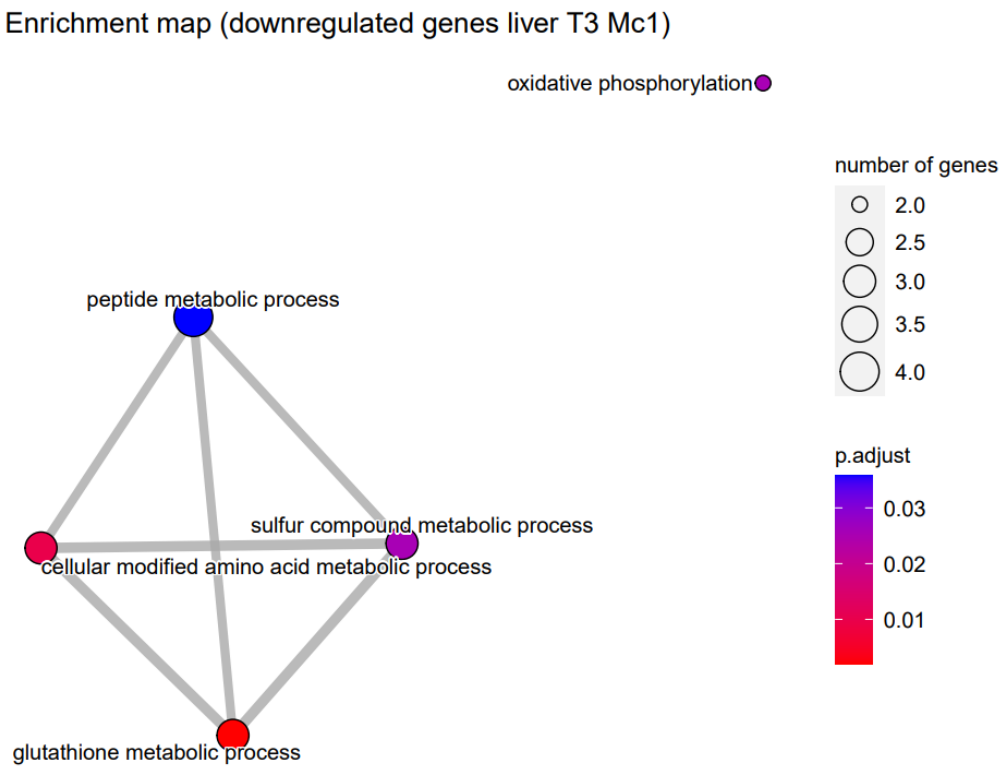

Figure 3.9 Enrichment map of GO terms for the down regulated genes in liver tissue for the mc1/ctr diet contrast in time point 3: *Each GO term is represented by a dot on the map and the number of genes connected to each term decided the size of the circle, ranging from 2 to 4 genes per term. The color of the circles represented the adjusted p-value of the GO term. Red meant a lower p-value while blue a higher. The map shows the 5 GO terms with p value under 0.05 for the 137 GO terms linked to the 17 downregulated genes from the contrast. Four of them clustered together, meaning that the biological processes of the GO terms are related to each other, and one term is single.*

For the mc2/ctr contrast twelve terms are shown to be significantly enriched (p-adjusted<0.05) and they group together in three clusters based on relatedness between the biological processes (Figure 3.9). These processes were all down regulated in the liver for time point 3. The smallest cluster consisted of biological processes concerning cell growth. The largest cluster contains processes concerning organ development and many of the terms are closely linked. In the last cluster biological processes concerning biosynthetic and metabolic processes are found.

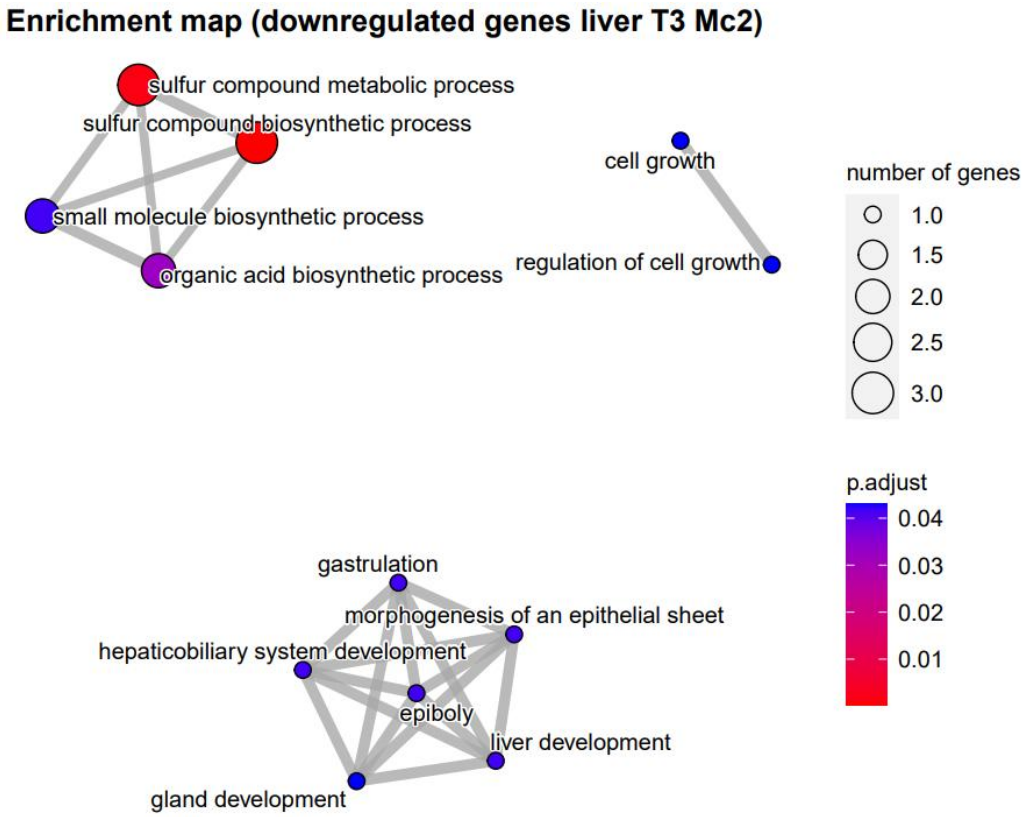

Figure 3.9 Enrichment map of GO terms for the downregulated genes in liver for the mc2/ctr contrast in time point 3: *Each GO term is represented by a dot on the map and the number of genes connected to each term decided the size of the circle, ranging from 1 to 3 genes per term. The color of the circles represented the adjusted p-value of the GO term. Red meant a lower p-value while blue a higher. The map shows the 12 GO terms with p value under 0.05 from the 34 GO terms linked to the 20 down regulated genes from the contrast. The terms created three different clusters where the biological processes of the GO terms are related to each other.*

4. DISCUSSION

4.1 Lack of large effects of industry relevant levels of beta-mannan as prebiotic in salmon feed

Results from the differential expression analysis showed a lacking effect from the β-manna feeds and this can have many explanations. The gut microbiota for salmon is an area where there are large gaps in the information available (55). Perhaps the Atlantic salmon does not have the correct gut microbiomes to utilize and digest the β-mannan or maybe there are not enough of the bacteria that can digest the prebiotic to create an effect. The inclusion level of 0.2% β-manna in the experimental feeds might have been too low to create changes in molecular function. This combined with a short digestion track and thereby quick passthrough time for feed in salmon (56), could be an explanation for the lack of effect. For future work an additional experiment testing higher inclusion levels of β-manna, could confirm if the lack of impact was due to the low inclusion or due to the β-manna not actually having an effect.

Earlier studies considering different prebiotics as additives in salmon feed have also shown meager results (57). The inclusion of 1% fructooligosaccharides in feed has been shown to have no effect on growth, nor affect energy retention or feed conversion ratios for salmon (57). Further no immune modulatory effects were shown when including the probiotic. Mannan oligosaccharide is another prebiotic that is shown to have no effect on growth at a 1% inclusion level (57). The oligosaccharide was shown to have a slight decreasing effect on blood neutrophil oxidative radical production and serum lysozyme activity (57), which can mean a slight decrease in immune stimulation (58). In another study the addition of mannan oligosaccharide and coconut oil to feed showed no effect on growth or ability to resist sea lice infection, but marginal changes to host immune expression occurred (59). Among these were reduction of CD4 expression in blood leukocytes (59).

In mammals, small effects on growth and feed efficiency have been seen when adding β-mannan to their feed, but no immune modulatory effects were detected. When adding galactoglucomannan at a 1-4% inclusion level for nursery pigs an increase of butyrate production in the gut was seen, but no increase in growth or improvements in health were shown (29). In another study piglets were fed diets with 0.4 and 2.8% inclusions of β-mannans and again no increase in growth was seen (28). The study also showed that addition of β-mannans had no effect on the serum concentrations of haptoglobin and C-reactive protein in the piglets. Poultry are often fed diets consisting of ingredients naturally containing β-mannans like soybean meal, copra meal, guar gum and palm kernel meal (60). Studies suggest that copra meal in poultry diets can promote both growth and feed efficiency, but no immune modulatory effects were seen (61). Addition of β-mannanase (mannan degrading enzyme (62)) in

poultry diets along with β-manna containing ingredients have also been researched (63). In combination with soybean meal and guar gum the enzyme has shown to have a positive effect on feed conversion ratios (63).

4.1 Small effect of pre-biotika on inflammatory response related genes

Results from the differential expression analysis with DEseq2 shows a low impact of the three β-manna containing feeds on all tissues. Over 80% of the all DEGs were identified in the liver and most of the feed-associated DEGs were down regulated, with modest effect size (log2FC>-1.5) in the feed with probiotics compared to the control feed.

Galactoglucomannan has shown in vitro prebiotic activity when incubated with human fecal material (25). The simulation of beneficial bacteria resulted in release of the short chain fatty acids, acetate, and butyrate (25). Both of the products have proven anti-inflammatory effects (64) (65) and our analysis results show a small β-mannan effect on pathways linked to inflammatory responses.

The GO enrichment analysis showed down regulation of genes associated with glutathione metabolism related processes in the liver of adult salmon eating the mc1 diet (Figure 3.9). In the same group, down regulation of the gene, glutamate--cysteine ligase regulatory subunit-like tied to glutathione biosynthesis (66) was found (Table 3.1). Glutathione has an important role in many cellular reactions both as a part of the cellular defense system against oxidation processes and in different metabolic processes (67). As an antioxidant glutathione contributes to the neutralization of free radicals and peroxides both directly and indirectly (via enzymatic reaction). In the metabolic processes involving glutathione we find synthesis of leukotrienes and prostaglandins (67). Both of which have important roles as mediators of inflammatory response (68) (69). The metabolic processes also involve formation of glutathione-NO adduct which is an important mediator of downstream signaling for nitric oxide (70). That when produced and released by macrophages kills bacteria and parasites (71).

The two down regulated genes cysteine dioxygenase type 1-like and cysteine dioxygenase type 1 seen in liver for adult salmon being fed the mc2 diet (Table 3.1) are both tied to the biosynthesis of taurine (72) (73). Taurine is involved in a number of vital processes like development of the nervous and renal system, osmoregulation, the formation of bile salts and immune modulation (74). The amino acid has been shown to reduce oxidative stress and have anti-inflammatory properties (75). Down regulation of the gene glutamate--cysteine ligase catalytic subunit isoform X1, means this group also has down regulation of glutathione biosynthesis (76). Up regulation of the plasma protease C1 inhibitor-like leads to increased negative regulation of complement activation (77)**.** The complement

system is an important part of the innate immune system that promotes inflammation, attacks pathogens and helps clear out microbes and damaged cells (78). In hindgut for adult salmon fed the mn3 diet, down regulation of the complement cascade also happens through the complement C3 gene (Table 3.1) (79).

Overall, a small decrease in immune modulatory effects were seen, mostly in the liver and for the mc1 and mc2 diets. The decrease in immune activity could mean a slight decrease in inflammation for the salmon being fed the β-mannan diets vs the salmon being fed the control diet.

5. CONCLUSION

In this thesis it was tested if functional feeds with added β-mannan are impacting the Atlantic salmon gut, gill and liver physiology at the molecular level.

The results from the differential expression analysis from both DESeq2 and EdgeR shows a low level of impact from the three different β-mannan diets in all tissues for both smolt and adult salmon.

REFERENCES

1. Salmon-Industry-Handbook-2021.pdf [Internett]. [sitert 3. mars 2022]. Tilgjengelig på: https://corpsite.azureedge.net/corpsite/wp-content/uploads/2021/05/Salmon-Industry-Handbook-2021.pdf
2. fiskeridepartementet N og. Norsk havbruksnæring [Internett]. Regjeringen.no. regjeringen.no; 2021 [sitert 9. mars 2022]. Tilgjengelig på: https://www.regjeringen.no/no/tema/mat-fiske-og-landbruk/fiskeri-og-havbruk/Norsk-havbruksnaring/id754210/
3. Olmos Soto J, Paniagua-Michel J de J, Lopez L, Ochoa L. Functional Feeds in Aquaculture. I: Kim SK, redaktør. Springer Handbook of Marine Biotechnology [Internett]. Berlin, Heidelberg: Springer; 2015 [sitert 8. mars 2022]. s. 1303–19. (Springer Handbooks). Tilgjengelig på: https://doi.org/10.1007/978-3-642-53971-8_59
4. Akhter N, Wu B, Memon AM, Mohsin M. Probiotics and prebiotics associated with aquaculture: A review. Fish Shellfish Immunol. 1. august 2015;45(2):733–41.
5. La Rosa SL, Leth ML, Michalak L, Hansen ME, Pudlo NA, Glowacki R, mfl. The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. Nat Commun. 22. februar 2019;10(1):905.
6. Genetically Engineered Salmon : Background and Issues [Internett]. [sitert 1. mars 2022]. Tilgjengelig på: https://web.p.ebscohost.com/ehost/ebookviewer/ebook/ZTAwMHh3d19fODM2NjgxX19BTg2?sid=8beeaed6-6316-48e7-8bad-ff927d22a69f@redis&vid=0&format=EB&lpid=lp_i&rid=0
7. Thorstad EB, Whoriskey F, Rikardsen AH, Aarestrup K. Aquatic Nomads: The Life and Migrations of the Atlantic Salmon. I: Atlantic Salmon Ecology [Internett]. John Wiley & Sons, Ltd; 2010 [sitert 1. mars 2022]. s. 1–32. Tilgjengelig på: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444327755.ch1
8. Vøllestad A, Halleraker JH. anadrom. I: Store norske leksikon [Internett]. 2021 [sitert 1. mars 2022]. Tilgjengelig på: http://snl.no/anadrom
9. Vøllestad A. laks. I: Store norske leksikon [Internett]. 2021 [sitert 1. mars 2022]. Tilgjengelig på: http://snl.no/laks
10. Rikardsen AH, Dempson JB. Dietary Life-Support: The Food and Feeding of Atlantic Salmon at Sea. I: Atlantic Salmon Ecology [Internett]. John Wiley & Sons, Ltd; 2010 [sitert 1. mars 2022]. s. 115–43. Tilgjengelig på: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444327755.ch5
11. OECD. Atlantic salmon (Salmo salar) [Internett]. Paris: OECD; 2017 des [sitert 2. mars 2022] s. 107–239. Tilgjengelig på: https://www.oecd-ilibrary.org/environment/safety-assessment-of-transgenic-organisms-in-the-environment-volume-7/atlantic-salmon-salmo-salar_9789264279728-7-en
12. aquaculture [Internett]. [sitert 9. april 2022]. Tilgjengelig på: https://dictionary.cambridge.org/dictionary/english/aquaculture
13. akvakultur. I: Store norske leksikon [Internett]. 2017 [sitert 8. mars 2022]. Tilgjengelig på: http://snl.no/akvakultur
14. FAO. FAO Yearbook. Fishery and Aquaculture Statistics 2019/FAO annuaire. Statistiques des pêches et de l'aquaculture 2019/FAO anuario. Estadísticas de pesca y acuicultura 2019 [Internett]. Rome, Italy: FAO; 2021 [sitert 8. mars 2022]. 110 s. (FAO Yearbook of Fishery and Aquaculture Statistics). Tilgjengelig på: https://www.fao.org/documents/card/en/c/cb7874t
15. Booklets [Internett]. English. [sitert 9. mars 2022]. Tilgjengelig på: https://www.fiskeridir.no/English/Aquaculture/Statistics/Booklets
16. Om AquaGen [Internett]. AquaGen. [sitert 10. april 2022]. Tilgjengelig på: https://aquagen.no/om-aquagen/
17. Produksjon [Internett]. Sevareid.no. [sitert 10. april 2022]. Tilgjengelig på: https://sevareid.no/produksjon/
18. Laks - Bærekraftig oppdrett av laks - Mowi [Internett]. [sitert 10. april 2022]. Tilgjengelig på: https://mowi.com/no/laks/
19. Olesen I, Myhr AI, Rosendal GK. Sustainable Aquaculture: Are We Getting There? Ethical

Perspectives on Salmon Farming. J Agric Environ Ethics. august 2011;24(4):381–408.
20. Misund B. fiskeoppdrett. I: Store norske leksikon [Internett]. 2021 [sitert 2. mars 2022]. Tilgjengelig på: http://snl.no/fiskeoppdrett
21. Tønjum T, Otterholt E. antibiotikaresistens. I: Store medisinske leksikon [Internett]. 2021 [sitert 4. mars 2022]. Tilgjengelig på: http://sml.snl.no/antibiotikaresistens
22. Encarnação P. 5 - Functional feed additives in aquaculture feeds. I: Nates SF, redaktør. Aquafeed Formulation [Internett]. San Diego: Academic Press; 2016 [sitert 8. mars 2022]. s. 217–37. Tilgjengelig på: https://www.sciencedirect.com/science/article/pii/B9780128008737000051
23. Sirevåg R. probiotika. I: Store medisinske leksikon [Internett]. 2021 [sitert 8. mars 2022]. Tilgjengelig på: http://sml.snl.no/probiotika
24. Chen J, Robb CS, Unfried F, Kappelmann L, Markert S, Song T, mfl. Alpha- and beta-mannan utilization by marine Bacteroidetes. Environ Microbiol. 2018;20(11):4127–40.
25. La Rosa SL, Kachrimanidou V, Buffetto F, Pope PB, Pudlo NA, Martens EC, mfl. Wood-Derived Dietary Fibers Promote Beneficial Human Gut Microbiota. mSphere. 27. februar 2019;4(1):e00554-18.
26. Canfora EE, Blaak EE. Acetate: a diet-derived key metabolite in energy metabolism: good or bad in context of obesity and glucose homeostasis? Curr Opin Clin Nutr Metab Care. november 2017;20(6):477–83.
27. Liu H, Wang J, He T, Becker S, Zhang G, Li D, mfl. Butyrate: A Double-Edged Sword for Health? Adv Nutr. 1. januar 2018;9(1):21–9.
28. Huntley NF, Gould SA, Patience JF. Evaluation of the effect of β-mannanase supplementation and mannans on nursery pig growth performance and serum acute-phase protein concentrations. Can J Anim Sci. mars 2020;100(1):111–8.
29. Michalak L, Gaby JC, Lagos L, La Rosa SL, Hvidsten TR, Tétard-Jones C, mfl. Microbiota-directed fibre activates both targeted and secondary metabolic shifts in the distal gut. Nat Commun. 13. november 2020;11(1):5773.
30. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. november 2019;20(11):631–56.
31. Martinsen L, Børresen-Dale AL, Dissen E. RNA. I: Store medisinske leksikon [Internett]. 2021 [sitert 9. mars 2022]. Tilgjengelig på: http://sml.snl.no/RNA
32. Kappelmann-Fenzl M. Library Construction for NGS. I: Kappelmann-Fenzl M, redaktør. Next Generation Sequencing and Data Analysis [Internett]. Cham: Springer International Publishing; 2021 [sitert 9. mars 2022]. s. 39–45. (Learning Materials in Biosciences). Tilgjengelig på: https://doi.org/10.1007/978-3-030-62490-3_3
33. Eisele M, Kappelmann-Fenzl M. NGS Technologies. I: Kappelmann-Fenzl M, redaktør. Next Generation Sequencing and Data Analysis [Internett]. Cham: Springer International Publishing; 2021 [sitert 9. mars 2022]. s. 47–58. (Learning Materials in Biosciences). Tilgjengelig på: https://doi.org/10.1007/978-3-030-62490-3_4
34. FASTQ files explained [Internett]. [sitert 10. mars 2022]. Tilgjengelig på: https://support.illumina.com/bulletins/2016/04/fastq-files-explained.html
35. Reppell M, Novembre J. Using pseudoalignment and base quality to accurately quantify microbial community composition. PLOS Comput Biol. 16. april 2018;14(4):e1006096.
36. Srivastava A, Malik L, Sarkar H, Zakeri M, Almodaresi F, Soneson C, mfl. Alignment and mapping methodology influence transcript abundance estimation. Genome Biol. 7. september 2020;21(1):239.
37. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods. april 2017;14(4):417–9.
38. Bharti R, Grimm DG. Design and Analysis of RNA Sequencing Data. I: Kappelmann-Fenzl M, redaktør. Next Generation Sequencing and Data Analysis [Internett]. Cham: Springer International Publishing; 2021 [sitert 9. mars 2022]. s. 143–75. (Learning Materials in Biosciences). Tilgjengelig på: https://doi.org/10.1007/978-3-030-62490-3_11
39. GO enrichment analysis [Internett]. Gene Ontology Resource. [sitert 10. mars 2022]. Tilgjengelig på: http://geneontology.org/docs/go-enrichment-analysis/

40. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internett]. [sitert 12. april 2022]. Tilgjengelig på: https://www.bioinformatics.babraham.ac.uk/projects/FastQC/
41. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 1. oktober 2016;32(19):3047–8.
42. Create Elegant Data Visualisations Using the Grammar of Graphics [Internett]. [sitert 19. april 2022]. Tilgjengelig på: https://ggplot2.tidyverse.org/
43. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare [Internett]. F1000Research; 2020 [sitert 17. mars 2022]. Tilgjengelig på: https://f1000research.com/articles/9-304
44. Almodaresi F, Sarkar H, Srivastava A, Patro R. A space and time-efficient index for the compacted colored de Bruijn graph. Bioinformatics. 1. juli 2018;34(13):i169–77.
45. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. april 2017;14(4):417–9.
46. Salmon Output File Formats - Salmon 1.8.0 documentation [Internett]. [sitert 21. mars 2022]. Tilgjengelig på: https://salmon.readthedocs.io/en/latest/file_formats.html
47. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [Internett]. F1000Research; 2016 [sitert 22. mars 2022]. Tilgjengelig på: https://f1000research.com/articles/4-1521
48. tximport.pdf [Internett]. [sitert 22. mars 2022]. Tilgjengelig på: https://bioconductor.org/packages/release/bioc/manuals/tximport/man/tximport.pdf
49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [Internett]. Bioinformatics; 2014 feb [sitert 22. mars 2022]. Tilgjengelig på: http://biorxiv.org/lookup/doi/10.1101/002832
50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 1. januar 2010;26(1):139–40.
51. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. oktober 2010;11(10):R106.
52. plotPCA function - RDocumentation [Internett]. [sitert 22. mars 2022]. Tilgjengelig på: https://www.rdocumentation.org/packages/DESeq2/versions/1.12.3/topics/plotPCA
53. AnnotationHub How-To's [Internett]. [sitert 24. mars 2022]. Tilgjengelig på: https://bioconductor.org/packages/devel/bioc/vignettes/AnnotationHub/inst/doc/AnnotationHub-HOWTO.html
54. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, mfl. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 28. august 2021;2(3):100141.
55. Wang J, Jaramillo-Torres A, Li Y, Kortner TM, Gajardo K, Brevik ØJ, mfl. Microbiota in intestinal digesta of Atlantic salmon (Salmo salar), observed from late freshwater stage until one year in seawater, and effects of functional ingredients: a case study from a commercial sized research site in the Arctic region. Anim Microbiome. 28. januar 2021;3(1):14.
56. Aas TS, Sixten HJ, Hillestad M, Sveier H, Ytrestøyl T, Hatlen B, mfl. Measurement of gastrointestinal passage rate in Atlantic salmon (Salmo salar) fed dry or soaked feed. Aquac Rep. 1. november 2017;8:49–57.
57. Song SK, Beck BR, Kim D, Park J, Kim J, Kim HD, mfl. Prebiotics as immunostimulants in aquaculture: A review. Fish Shellfish Immunol. 1. september 2014;40(1):40–8.
58. Siwicki AK, Anderson DP, Rumsey GL. Dietary intake of immunostimulants by rainbow trout affects non-specific immunity and protection against furunculosis. Vet Immunol Immunopathol. mai 1994;41(1–2):125–39.
59. Bledsoe JW, Pietrak MR, Burr GS, Peterson BC, Small BC. Functional feeds marginally alter immune expression and microbiota of Atlantic salmon (Salmo salar) gut, gill, and skin mucosa though evidence of tissue-specific signatures and host–microbe coadaptation remain. Anim Microbiome. 10. mars 2022;4(1):20.
60. Hsiao HY, Anderson DM, Dale NM. Levels of β-Mannan in Soybean Meal. Poult Sci. 1. august 2006;85(8):1430–2.
61. Sundu B, Hatta U, Chaudhry AS. Potential use of beta-mannan from copra meal as a feed additive for broilers. Worlds Poult Sci J. 1. desember 2012;68(4):707–16.

62. Dawood A, Ma K. Applications of Microbial β-Mannanases. Front Bioeng Biotechnol [Internett]. 2020 [sitert 12. mai 2022];8. Tilgjengelig på: https://www.frontiersin.org/article/10.3389/fbioe.2020.598630
63. Caldas JV, Vignale K, Boonsinchai N, Wang J, Putsakum M, England JA, mfl. The effect of β-mannanase on nutrient utilization and blood parameters in chicks fed diets containing soybean meal and guar gum. Poult Sci. 1. august 2018;97(8):2807–17.
64. Balmer ML, Ma EH, Thompson AJ, Epple R, Unterstab G, Lötscher J, mfl. Memory CD8+ T Cells Balance Pro- and Anti-inflammatory Activity by Reprogramming Cellular Acetate Handling at Sites of Infection. Cell Metab. 1. september 2020;32(3):457-467.e5.
65. Siddiqui MT, Cresci GAM. The Immunomodulatory Functions of Butyrate. J Inflamm Res. 18. november 2021;14:6025–41.
66. LOC106584335 - GCS light chain - Salmo salar (Atlantic salmon) - LOC106584335 gene & protein [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/A0A1S3PBH2
67. Wu G, Fang YZ, Yang S, Lupton JR, Turner ND. Glutathione Metabolism and Its Implications for Health. J Nutr. 1. mars 2004;134(3):489–92.
68. Ricciotti E, FitzGerald GA. Prostaglandins and Inflammation. Arterioscler Thromb Vasc Biol. mai 2011;31(5):986–1000.
69. Brain SD, Williams TJ. Leukotrienes and inflammation. Pharmacol Ther. 1. januar 1990;46(1):57–66.
70. Broniowska KA, Diers AR, Hogg N. S-NITROSOGLUTATHIONE. Biochim Biophys Acta. mai 2013;1830(5):3173–81.
71. nitric oxide | Uses, Facts, & Definition | Britannica [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.britannica.com/science/nitric-oxide
72. LOC106580183 - Cysteine dioxygenase - Salmo salar (Atlantic salmon) - LOC106580183 gene & protein [Internett]. [sitert 23. april 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/A0A1S3NMG5
73. LOC106580183 - Cysteine dioxygenase - Salmo salar (Atlantic salmon) - LOC106580183 gene & protein [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/A0A1S3NMG5
74. Chen C, Xia S, He J, Lu G, Xie Z, Han H. Roles of taurine in cognitive function of physiology, pathologies and toxication. Life Sci. 15. august 2019;231:116584.
75. Marcinkiewicz J, Kontny E. Taurine and inflammatory diseases. Amino Acids. 2014;46(1):7–20.
76. GSH1 - Glutamate--cysteine ligase - Salmo salar (Atlantic salmon) - GSH1 gene & protein [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/B5X493
77. LOC106602690 - plasma protease C1 inhibitor-like - Salmo salar (Atlantic salmon) - LOC106602690 gene & protein [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/A0A1S3RGH0
78. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. The complement system and innate immunity. Immunobiol Immune Syst Health Dis 5th Ed [Internett]. 2001 [sitert 12. mai 2022]; Tilgjengelig på: https://www.ncbi.nlm.nih.gov/books/NBK27100/
79. LOC106595495 - complement C3 - Salmo salar (Atlantic salmon) - LOC106595495 gene & protein [Internett]. [sitert 12. mai 2022]. Tilgjengelig på: https://www.uniprot.org/uniprot/A0A1S3QRL5

APPENDIX

Code used to perform the pseudo mapping and differential expression analysis for this masters project can be found in this Gitlab repository:
https://gitlab.com/martesk/master-thesis

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway