De Santis C, Bartie K, Olsen RE, Taggart J & Tocher DR (2015) Nutrigenomic profiling of transcriptional processes affected in liver and distal intestine in response to a soybean meal-induced nutritional stress in Atlantic salmon (Salmo salar), *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 15, pp. 1-11.

DOI: 10.1016/j.cbd.2015.04.001

Nutrigenomic profiling of transcriptional processes affected in liver and distal intestine in response to a soybean meal-induced nutritional stress in Atlantic salmon (*Salmo salar*).

C. De Santis[1*], K.L. Bartie[1], R.E. Olsen[2,3], J.B.Taggart[1] and D.R. Tocher[1]

[1] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland

[2] Institute of Marine Research, Matre, 5984 Matredal, Norway

[3] Norwegian University of Science and Technology, Department of Biology, 7491 Trondheim, Norway

Keywords: dietary protein substitution, transcriptome, liver, intestine, Atlantic salmon, *Salmo salar*, soybean meal, nutrigenomic, nutritional stress

Abstract

The aim of the present study was to generate an experimental model to characterize the nutrigenomic profile of a plant-derived nutritional stress. Atlantic salmon (*Salmo salar*) was used as the model species. The nutritional stress was induced by inclusion of dietary defatted soybean meal (SBM), as this ingredient had been previously demonstrated to induce enteropathy in the distal intestine and reduce growth in salmon. Triplicate groups of Atlantic salmon were fed increasing concentrations (0, 100, 200 and 300 g kg^{-1}) of SBM for 12 weeks and reduced growth performance was used as the indicator of nutritional stress. The transcriptome was analysed in two tissues, liver and distal intestine, with the hypothesis being that the liver transcriptome would be characterized by gene expression responses related to overall growth and health performance, whereas intestinal gene expression would be dominated by specific responses to SBM. A set of 133 genes was differentially expressed in liver including 44 genes in common with the intestinal response. The liver specific response included up-regulation of genes involved in protein digestion, energy metabolism and immune functions, whereas genes regulated in other metabolic pathways were generally anabolic and down-regulated. These responses may be more related to general nutritional stress than to SBM *per se*. The transcriptomic profile in the distal intestine was consistent with the enteritis response as described previously. This study provides a comprehensive report on the profiles of liver and distal intestine transcriptomes, highlighting the role of the former tissue in fish undergoing SBM-induced nutritional stress.

1. Introduction

Aquaculture is a major player of the global food security program producing approximately fifty percent of the world seafood supply (The World Bank, 2013). In an effort to provide the growing population with a consistent supply of high-quality and sustainable seafood, alternative sources of protein to the traditionally used fishmeal (FM) are now being sought after, particularly for industries involved in the farming of carnivorous fish species such as Atlantic salmon (*Salmo salar*). Over the last two decades, significant research investment has been made to identify alternative sources of protein that could fulfil the requirements of these fish, provide a competitive growth rate and good flesh properties, and avoid any negative impact on fish welfare. Thus far, plant-based ingredients have proved to be a potentially attractive and sustainable alternative mainly due to a competitive market price and availability. The use of plant protein products such as soybean, pea and other legumes, wheat and corn gluten is either already established in commercial feed manufacturing protocols or is proven to be a viable alternative to FM (Gatlin et al., 2007).

One limitation of plant-based products, however, is the presence of chemical compounds known as antinutritional factors (ANFs), endogenously produced by the plant with a variety of functions including structural (e.g. fibres), storage (e.g. phytic acid) or as defence mechanisms (e.g. saponins, lectins and protease inhibitors; reviewed by Francis et al., 2001, Gatlin et al., 2007, Krogdahl et al., 2010). Processes to concentrate protein in plant products can lower the concentration of ANFs, but ANFs can also co-purify with proteins and be present in protein concentrates utilized as an alternative to FM. Fish are generally not able to metabolize ANFs and when these are present in their diets they can cause a number of effects primarily associated with digestive physiology, health and metabolism, impacting on fish welfare and resulting in reduced productivity (Francis et al., 2001). For example, substances such as fibre and phytic acid can affect the uptake and utilization of micronutrients, while

protease inhibitors, saponins and phytosterols can interfere with the digestion and/or absorption of proteins, lipids and cholesterol, respectively (Gatlin et al., 2007). In addition, some ANFs such as saponins and lectins can have important health consequences primarily affecting the intestinal mucosa, causing inflammation and increasing the permeability to pathogens and other unwanted substances (Krogdhal et al., 2010). It follows that the nutritional profile of plant-based protein raw materials does not always reflect the true value available to the animal for absorption and utilization, but varies depending on the presence and abundance of ANFs.

Understanding how fish respond to and utilize dietary plant ingredients has therefore been a key topic of research in recent years. The study of individual ANFs has been a useful approach to unveil mechanisms behind the biological response of fish to these chemicals and research in this direction will continue to provide crucial information (Buttle et al., 2001, Francis et al., 2002, Krogdahl et al., 2010). However, from a commercial and applied perspective it is critical to maintain a holistic approach particularly as it is increasingly evident that certain effects are only visible when two or more ANFs are present in combination (Kortner et al., 2012, Krogdahl et al., 2010). The high-demand for alternative protein sources continues to be a strong driver for research testing novel raw materials individually or in combination (Azaza et al., 2009, Booth et al. 2001, Glencross et al., 2005, Hartviksen et al., 2014, Kaushik et al., 2004, Kortner et al., 2013, Øverland et al., 2009, Panserat et al., 2009). In addition, novel processing technologies are contributing to the introduction of new products from the same raw materials but with improved nutritional characteristics and reduced levels of ANFs. When testing ingredients with reduced levels of ANFs, a macroscopic physiological response might not be measurable in short-term experimental trials, whereas cumulative adverse effects, gut pathologies or other detrimental physiological conditions might accrue under long-term or more challenging commercial conditions (Krogdahl et al., 2010).

It is therefore increasingly important at experimental levels to have sensitive tools at our disposal that are able to detect subtle physiological changes and help predict long-term detrimental effects of dietary modification. This approach is the foundation of nutrigenomics, a science still in its infancy that studies the influence of food or food constituents on the transcriptome. In nutrigenomics, nutrients are considered signals through which cells interpret information about the environment (diet) and respond, according to necessity, by modifying metabolic pathways through regulation of gene and protein expression towards homeostasis (Muller and Kersten, 2003). Nutrigenomics can provide a means to interpret how dietary ingredients are perceived by fish at the tissue level and also provide a molecular snapshot of the physiological response of specific tissues. Nutrigenomic studies have utilized both candidate gene and high-throughput approaches to investigate the physiological responses of tissues and mechanisms of adaptation to plant-based dietary ingredients in fish or to understand the molecular mechanisms underlying pathologies such as enteritis caused by these ingredients (Gu et al., 2014, Kortner et al., 2012, Kortner et al., 2013, Panserat et al., 2009, Tacchi et al., 2012, Wacyk et al., 2012). However, in order to develop a practical diagnostic use of nutrigenomic approaches, it is essential to determine "dietary signatures" that characterize the physiological response to nutritional stressors.

The overarching aim of the present study was to generate an experimental model to characterize the nutrigenomic profile of a plant-derived nutritional stress. A nutritional stress was induced in Atlantic salmon using dietary defatted soybean meal (SBM), which has been demonstrated to induce enteropathy in the distal intestine (Baeverfjord and Krogdahl, 1996, Urán et al., 2008, Urán et al., 2009). Impaired growth was taken as an indicator of pronounced nutritional stress. Transcriptomic analyses were performed in two tissues, liver and distal intestine. Liver was chosen for being arguably the key metabolically active tissue responding

to diet, while the distal intestine has been studied for the development of specific pathologies associated with plant ingredients such as SBM (Baeverfjord and Krogdahl, 1996, Kortner et al., 2012). The results provided: *a*) an insight into the molecular signatures resulting from a nutritional stress response to which future nutrigenomic studies can refer to; and *b*) a platform for the identification of candidate genes for the molecular phenotyping of several physiological parameters in liver and distal intestine. These data were specifically analysed to test the hypothesis that the liver transcriptome would include gene expression responses that could be more general and related to overall growth/health performance, whereas intestinal gene expression would be dominated by specific responses to SBM.

2. Materials and Methods

2.1. Nutritional trial, diets and sampling

The nutritional trial was conducted at the Institute of Marine Research, Matre, Norway using Atlantic salmon (*Salmo salar*) of the commercial Aquagen strain (Aquagen Ltd, Kyrksæterøra, Norway). The fish were normal smolts (1+) produced in 5 m indoor tanks under natural photoperiod (June 2012). A total of 540 fish were anaesthetized in Finquel, tagged with PIT tags, and equally distributed into 12 white fiberglass tanks each holding 400 L (95 cm x 95 cm x 60 cm, L x W x H). The tanks were closed with lids, and supplied with two fluorescent light tubes (18 W each) and an automatic feeder (ARVO-TEC T Drum 2000, Arvotec, Huutokoski, Finland). After two days, the salinity was gradually increased from freshwater to full salinity (35 ppt) and the temperature adjusted to 12 °C (aerated and temperature-controlled seawater). Water flow was fixed at 20 L min^{-1}. Fish were acclimatized for 4-weeks during which time they were all fed the same feed (S0, Table 1). Two days prior to the start of the trial, the individual fish were weighed (group mean 175 g) and returned to the tanks. Four dietary treatments were investigated with progressively increasing inclusions of solvent-extracted soybean meal (SBM; 0 g kg^{-1}, 100 g kg^{-1}, 200 g kg^{-1}, 300 g kg^{-1})

substituting other protein sources (FM, corn gluten, sunflower cake and horsebeans) referred to as diets S0, S10, S20 and S30, respectively (Table 1). Each experimental feed was fed to triplicate tanks. All feeds were formulated to meet the nutritional requirement of salmon (National Research Council, NRC, 2011). The feeding trial lasted for 87 days and at the end of this period individual weights and lengths of all fish were measured and specific growth rate (SGR) calculated using the following equation [SGR = $\ln WT_{fin}$ - $\ln WT_{in}$ / days]. In addition, the same tip of the liver and a portion of the distal (posterior) intestine were dissected from six individuals per dietary treatment (two per tank replicate). Tissue samples were immediately placed in RNALater (Life Technologies, Paisley, UK) and processed as per manufacturer's instructions before being stored at -20 °C prior to molecular analyses.

2.2. Transcriptomic analysis

Transcriptomic analysis was conducted using an Atlantic salmon custom-designed 4 x 44K oligo microarray (Agilent Technologies, Wokingham, UK; ArrayExpress accession no. A-MEXP-2065) as described in detail previously (Tacchi et al., 2011). The salmon custom array and laboratory procedures utilized have been widely used and extensively validated in previous studies (e.g. Bicskei et al., 2014, Martinez-Rubio et al., 2012, Morais et al., 2012a, Morais et al., 2012b, Tacchi et al., 2011, Tacchi et al., 2012). Briefly, total RNA was extracted from individual samples using TRI Reagent according to the manufacturer's instructions (Sigma-Aldrich, Dorset, UK), including a high salt precipitation as recommended for polysaccharide-rich tissues such as liver (Chomczynski and Mackey, 1995). RNA quantity, integrity and purity were assessed by agarose gel electrophoresis and spectrophotometry (NanoDrop ND-1000, Thermo Scientific, Wilmington, USA). Equal amounts of RNA from six individual fish liver and intestine tissues were analyzed. The resulting RNA samples were amplified using a TargetAmp™ 1-Round Aminoallyl-aRNA Amplification Kit, (Epicentre Technologies Corporation, Madison, Wisconsin, USA)

following recommended procedures and purified through a RNeasy Mini spin column (Qiagen, Manchester, UK). Aminoallyl-amplified RNA (aRNA) samples were individually labelled with Cy3 dye (GE HealthCare Life Sciences, Buckinghamshire, UK), while a pool of all aRNA samples was labelled with Cy5 dye (GE HealthCare Life Sciences) and used as a common reference. Unincorporated dye was removed by purifying the aRNA samples with an Illustra AutoSeq G-50 dye terminator column (GE HealthCare Life Sciences). Successful dye incorporation and sample integrity was assessed for 0.4 µL aliquots of labelled samples by agarose gel electrophoresis followed by fluorescent detection of aRNA products (Typhoon scanner, GE Healthcare Life Sciences). Cy dye concentration and aRNA quantification were measured by Nanodrop mediated spectrophotometry (Thermo Scientific, Wilmington, USA).

Labelled aRNA samples were hybridized to the custom-made array. A dual-label common reference design was adopted, where equal amounts of each individual aRNA sample and the common reference pool were competitively hybridized on an array. The common reference design allowed standardization of inter- and intra-array variability. A total of 24 samples (two tissues x two treatments x six replicates) were processed with the Gene Expression Hybridization Kit (Agilent Technologies), applied to the arrays and immediately incubated using SureHyb hybridization chambers in a DNA Microarray Hybridization Oven (Agilent Technologies) at 65 °C for 17 h. Throughout the experiment samples were randomized, preventing samples from the same treatment being overrepresented in a particular batch in order to avoid unintentional biases. Scanning was performed using a GenePix 4200 AL Scanner (Molecular Devices (UK) Ltd., Wokingham, UK) and the resulting images analyzed with Agilent Feature Extraction Software v.9.5 (Agilent Technologies) to extract the intensity values and identify the features. The foreground intensity was computed as the mean value of pixels, considered a better estimator as being less susceptible to distortion from outlier values (Russell et al., 2009), while background intensities were computed as the median value of

189 pixels. Details of the microarray experiment have been submitted to ArrayExpress and
190 assigned the accession number E-MTAB-3291.

191

192 *2.3. Data pre-processing and differential expression analysis*

193 Transcriptomic data analysis was performed using R v.3.0.1 and Bioconductor v.2.13
194 (Gentleman et al., 2004, R Core Team, 2013). Quality control, data pre-processing and
195 identification of differentially expressed features/genes were conducted using the software
196 package limma (Smyth, 2004). Array quality was assessed by visualizing and comparing
197 boxplots of red (R) and green (G) background and foreground intensities, MA plots [M =
198 $\log_2(G/R)$; A = $1/2\log_2(R*G)$] and spatial heterogeneity to reveal the presence of any
199 technical bias. Following this quality assessment, all arrays were retained for further analyses.
200 Features considered outliers in two or more replicates within at least one treatment were
201 excluded from further analyses. Foreground intensities were background-corrected using the
202 normexp approach (maximum likelihood variant "mle", offset = 50) as previously reported as
203 the most reliable method for two-color microarrays where background estimates are available
204 (Ritchie et al., 2007, Silver et al., 2009). Data were log-transformed and normalized using the
205 function *normalizeWithinArrays* (method = "loess") and *normalizeBetweenArrays* (method =
206 "RQuantile") (Smyth and Speed, 2003). Controls and features expressed just above
207 background (A values < 6) were also removed.

208 Features of the array were annotated using BLAST 2.2.29+ (blastx) against the entire non-
209 redundant protein database as well as using the KEGG Automatic Annotation Server to obtain
210 functional annotations (Altschul et al., 1990, Moriya et al., 2007). A total of 89.6 % of all
211 probes were returned with a BLAST annotation (annotation date May 2014) with e-value
212 < 0.001, while 59 % of probes were returned with a functional annotation (KEGG identifier)
213 using the KAAS server. Differentially expressed features between treatments were estimated
214 by least squares fitting of linear models on a probe-by-probe basis using the entire pre-
215 processed dataset. The function *lmFit* was used to compute differential expression and
216 statistics were extracted using *ebayes* (trend = TRUE), both limma functions (Smyth, 2004).

Features representing the same target gene as implied from KEGG annotation were merged into a unique value obtained by selecting the feature with the highest F-value. A new dataset was therefore generated for further analyses where each gene was only represented by one feature. Merging resulted in a dataset of 6729 annotated features targeting unique genes.

2.4 Data Mining

Hierarchical Clustering. Hierarchical cluster analysis was performed on gene expression normalized values to evaluate overall similarity between samples and analyzed using the R package *pvclust* (Suzuki and Shimodaira, 2011). "Correlation" was used as the distance measure using 1000 bootstrap replication.

Overview of differential expression. Differentially expressed genes were plotted using the R package *ggplot2* (Wickham, 2009). For figures involving functional information, the KEGG database was used as the chosen classification system. Venn diagrams were generated using the function *VennDiagram* from the limma software package and modified using *ggplot2*.

Gene-Set Enrichment Analysis (GSEA). Unique annotated sequences were analyzed using the R function *gage* of the software package GAGE (Generally Applicable Gene-set Enrichment, Luo et al., 2009) to identify mechanistic changes as suggested by coordinated expression changes in gene-sets. For completeness, two types of test were performed: *1 direction (1d)*, testing all genes in a gene-set moving towards the same direction; and *2 directions (2d)*, testing genes in a gene-set that move towards both directions. Gene-sets with a q-value < 0.1 were considered significant, where the q-value represented the p-value adjusted for false discovery rate. KEGG classification was used for these analyses and all figures were produced using the software package *ggplot2*.

3. Results

3.1 Growth performance

There were no significant differences in recorded mortality between treatments. Dietary SBM inclusion of 300 g kg^{-1} (i.e. feed S30) induced statistically significant growth retardation in salmon in terms of both final weight and specific growth rate (SGR) compared to fish fed all other treatments (Table 2). The SGR of salmon fed SBM at an inclusion level of 200 g kg^{-1} was significantly lower than fish fed S0 or S10, but significantly higher than fish fed S30, although final weights were not different to fish fed the diets with lower inclusion of SBM. Based on these data, transcriptome analyses were performed on tissue samples from fish fed diets S0 and S30 to guarantee the largest difference between samples and increase the possibility to detect differential expression.

3.2 Transcriptome analysis: Overview

Hierarchical clustering performed on the normalized gene expression data indicated the existence of a clear structure in the dataset with four main clusters (Fig. 1). A clear separation was evident between the liver and the intestine forming two independent clusters fully supported by bootstrap values (100). Within each tissue cluster, the dietary treatments S0 and S30 also induced clearly identifiable differential gene expression responses. The liver demonstrated a weaker, but significant, bootstrap support (63-80), compared with the response of the intestine that was fully supported by statistical bootstraps (100), indicating a less defined response in the liver compared with that of the intestine.

The analysis of differential expression computed by limma reflected the structure observed by hierarchical clustering, indicating a less pronounced, but still evident response in liver, where 133 genes were differentially expressed ($p < 0.01$, absolute FC > 1.3), compared to a marked response in the intestine where a total of 1918 transcripts were differentially expressed ($p < 0.01$, absolute FC > 1.3). The difference between the two tissues was again apparent in

that the differentially expressed transcripts of the liver showed a generally decreased (fold change) and less significant (*p* value) response compared with that of the intestine (Fig. 2a, b). Due to the minor transcriptional effect observed in the liver associated with primary metabolic role, the use of multiple testing corrections (Benjamini-Hochberg) on this tissue returned only 11 genes below the adjusted cutoff *p* value of 0.05 and proved to be over-conservative. Applying multiple testing correction (adj. $p < 0.05$) in the intestine resulted in 2664 differentially expressed genes. For consistency between the tissues we therefore elected to consider all genes with a $p < 0.01$ and absolute FC > 1.3 to be truly differentially expressed.

A total of 44 genes were significantly affected in both tissues, while the remaining 88 and 1873 genes were represented exclusively as tissue-specific responses in the liver and intestine respectively. Of the 44 genes affected in both liver and intestine, 31 were regulated in the same direction (i.e. either up- or down-regulated in both tissues) possibly indicating a systemic response, while 13 were regulated in opposite directions (Table 3). Expression was always calculated relative to the S0 treatment. Amongst the genes whose transcript abundance increased in both liver and intestine, there was an over-representation of ribosomal subunits (L5, L21, L11, L7, S3A); enzymes involved in ribosome biogenesis such as ribonucleases P/MRP protein subunit RPP25 and casein kinase II subunit alpha; proteins participating in RNA processing such as translation initiation factor 5, cleavage and polyadenylation specificity factor subunit 5 and the spliceosome component U5 snRNP protein and factors engaged in protein processing such as ER degradation enhancer, mannosidase alpha 3 (Table 3). In addition, other important genes that concurrently increased in both tissues included bile-salt stimulated lipase, complement component C7, sulfinoalanine decarboxylase and an enzyme involved in fatty acid β-oxidation, carnitine O-palmitoyltransferase 1. In contrast, down-regulated genes coded for proteins involved in metabolic processes such as lipid metabolism (glycerol kinase, elongation of very long chain fatty acids protein 5, very long

chain-3 hydroxyacyl-dehydratase), sugar metabolism (GDP-L-fucose synthase) and glycan metabolism (alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase).

The following sections (sections 3.3 and 3.4) report on the tissue-specific responses. Gene-set enrichment analysis was utilized as a tool to identify mechanistic changes of groups of genes. However, while the limited number of differentially expressed genes identified in the liver also allowed a more in-depth analysis on individual genes, this approach was not possible with the distal intestine without introducing bias in selecting the gene list. Thus, in the distal intestine only gene-set enrichment analysis was applied representing a more robust approach compared to individual gene analysis. The liver was found to be only marginally affected, whereas a large number of gene-sets were significantly affected ($q < 0.1$) in the intestine, clearly reflecting the trend identified above targeting differentially expressed genes.

3.3 Liver

Compared with the distal intestine, the liver was partially affected by diet S30 (Fig. 2) with only 88 genes uniquely differentially expressed compared to fish fed diet S0, although a set of key pathways were significantly affected (Fig. 3). The *2d* test was able to capture the most significantly perturbed pathways possessing genes both up and down-regulated. The most affected pathway in this category was the complement and coagulation cascades. A closer view on the complement cascade is shown in Fig. 4. Other affected processes identified by the *2d* test included lipid metabolism; in particular lipid digestion and absorption, regulation of steroid biosynthesis and the PPAR signalling pathway, amino acid metabolism; principally glycine/serine/threonine and alanine/aspartate/glutamate metabolism and metabolism of vitamins (retinol) including vitamin digestion and absorption. In contrast, the *1d* GAGE test was able to address expression changes in genes moving in the same direction, either up or down-regulated, and was particularly helpful in identifying the overall trend of expression within a specific pathway. This analysis indicated an up-regulation of the PI3K-Akt signalling

316 pathway and a down regulation of oxidative phosphorylation, proteasome apparatus, protein
317 export and terpenoid backbone biosynthesis.

318 In Table 4, a selection of the genes differentially expressed only in the liver is listed. To
319 minimize bias, we included most of the genes involved in the function of interest (the full list
320 included 40 of the 88 affected genes). These included the up-regulation of several genes
321 coding for pancreatic enzymes such as trypsin, carboxypeptidase B, etc. and bile salt-
322 stimulated lipase, and key genes participating in energy metabolism for instance, 2-
323 oxoglutarate dehydrogenase and succinate dehydrogenase (citric acid cycle). Generally,
324 down-regulated genes were shown to relate to amino acid, lipid and sterol metabolism (Table
325 4). Interestingly, ribosomal genes that were affected only in the liver were all down-regulated,
326 in contrast with those affected in both tissues that were up-regulated. Finally, genes
327 contributing to the innate immune response were also significantly altered in liver, where the
328 complement cascade pathway possessed the highest significance score. Genes of this pathway
329 that were significantly up-regulated included complement components 3 (C3), that was only
330 affected in the liver ($p < 0.001$, Table 4), and C7 that was up-regulated in both tissues in
331 tandem (Table 3). In addition, other genes of the complement cascade such as C5, C6,
332 complement component receptor type 2, the mannan-binding lectin serine protease 1
333 (MASP1) and the CD59 antigen were also perturbed to a lesser extent ($p < 0.05$) and are
334 shown in Fig. 4.

335 *3.4 Distal Intestine*

336 The distal intestine showed a pronounced global response, involving gene-sets regulating a
337 number of processes (Fig. 4). This response at a gene-set level is summarized by the major
338 biological processes affected in the intestine identified in both GAGE *1d* and *2d* analyses and
339 included primarily metabolic (e.g. lipid, amino acid and energy metabolism), organismal

systems (e.g. digestion and absorption, immune and endocrine system) and other cellular processes including pathways involved in the protein synthesis.

For clarity of presentation and due to the number of pathways identified in the intestine, pathways were separated into "highly significantly different" ($q < 0.0001$) and "significantly different" ($0.1 < q < 0.0001$), however this should not minimize the importance of pathways with lower q values. Among the highly significantly regulated gene-sets identified through the *2d* test ($q < 0.0001$) were pathways associated with phagocytosis (cellular organelles lysosome and phagosome), lipid and vitamin digestion and absorption pathways, PPAR signalling pathway of the endocrine system, antigen processing and presentation (immune system) and key metabolic pathways including glycan degradation, amino sugar and nucleotide sugar degradation, retinol metabolism, arginine/proline metabolism, alanine/aspartate/glutamate metabolism and glutathione metabolism (Fig. 5a). Only one immune pathway, antigen processing and presentation, was identified by the *2d* test with a highly significant q value. A further 22 pathways, mainly metabolic (n= 15), were differentially expressed below the $q = 0.0001$ cut-off (Fig. 5a).

Interestingly, only two pathways classified as "genetic information processing" were identified by the *2d* test and included the proteasome complex and aminoacyl-tRNA biosynthesis. Notably, the *1d* test revealed that this category of pathways was significantly up-regulated, with most pathways exhibiting a highly significant q value (< 0.0001) (e.g. ribosome, proteasome, spliceosome, RNA transport, etc., Fig. 5b). Furthermore, other gene-sets with increased expression in fish fed the S30 treatment comprised those regulating functions such as cell growth and death, cell cycle, apoptosis and MAPK signalling (Fig. 5b). An overall increased expression was also observed in a number of immune pathways. The immune response involved up-regulation of the TNF signalling pathway ($p < 0.0001$), NOD-like receptor interaction, NF-kB signalling pathway, cytosolic DNA sensing pathway, Jak-

STAT signalling pathway, cytokine-cytokine receptor interaction and T-cell receptor signalling pathway (Fig. 5b).

.

A number of gene-sets that tested significant in the *2d* test were also significant with the *1d* test, particularly amongst the down-regulated pathways (Fig. 5b). These included gene-sets centred on metabolic pathways and processes such as digestion and absorption of proteins, lipids, vitamins and minerals, indicating the genes underlying these processes were primarily decreased in response to diet S30. Finally, the *1d* test revealed that proteins associated with the lysosome and peroxisome organelles were also reduced (Fig. 5b).

4. Discussion

The present study reports a molecular snapshot reflecting the hepatic and intestinal (distal) physiology of Atlantic salmon exposed to a plant-based nutritional stress. The results herein were produced utilizing an unbiased global analysis approach to identify mechanistic changes of gene expression with the precise aim of avoiding the bias and restrictions often associated with the selection of limited gene lists. The study aimed to provide a platform to screen the most perturbed processes and identify molecular markers useful for phenotyping and discriminating influential processes affected in response to diets formulated with high levels of plant-based ingredients. The complete lists of differentially expressed genes identified are publicly available and can be consulted for further information on Array Express (accession number E-MTAB-3291).

The nutritional trial revealed that dietary SBM reduced growth performance in salmon when included at levels over 200 g kg^{-1}. It was not within the scope of the study to determine in detail the performance of salmon in response to dietary SBM inclusion, since this had already been investigated previously and the effects on growth were in agreement with those observed

389 in the present trial (growth data reviewed in Collins et al., 2013, health effects reviewed in
390 Krogdahl et al., 2010). The aim of this nutritional trial was solely to reproduce the expected
391 effects and provide an experimental model of nutritional stress that would satisfy the
392 objectives of the present gene expression study. The experimental model included two lower
393 inclusion levels (100 g kg^{-1} and 200 g kg^{-1}) that did not show consistent and/or significant
394 growth retardation compared with the control diet (S0). However, the results could provide
395 experimental material to test the potential molecular biomarkers and hypothesis developed in
396 the current study in dietary treatments that have only "subclinical" effects on performance.

397 The detected differential transcriptomic response observed in salmon fed 300 g kg^{-1} SBM
398 compared to FM control was greater in the distal intestine compared to the liver. These results
399 were expected due to the inflammation of the distal intestine and consequent immune
400 response that is generally observed after feeding similar levels of dietary SBM (Baeverfjord
401 and Krogdahl, 1996, Bakke-McKellep et al., 2007, Urán et al., 2008, Urán et al., 2009). Since
402 not directly affected by the contact with the feed itself and hence the ANFs, the liver response
403 was different and possibly accessory to that of the intestine. While in the former we observed
404 processes related to the immune response, cell growth and regeneration and evidence of
405 impairment of metabolic functions, the liver showed activation of mainly metabolic pathways
406 and an immune response that reflected the inflammation observed locally in the distal
407 intestine.

408 The role played by the liver in the physiological response to plant-based nutritional stress is
409 not fully understood. Kortner and colleagues (2013) reported an increase of hepatic
410 cholesterol metabolism suggesting it was a compensatory response to reduced dietary uptake
411 through endogenous cholesterol biosynthesis. However, this was in contrast to results from
412 the present study where all genes of lipid and sterol metabolism were down regulated in liver
413 with the exception of cholest-5-ene-3beta,7alpha-diol 3beta-dehydrogenase (up-regulated),

414 which is involved in the initial stages of bile acid synthesis from cholesterol (Schwarz et al.,
415 2000). The discrepancy observed might indicate that a different response is to be expected
416 depending upon SBM dietary inclusion level and, most importantly, the severity of the
417 immune response. Whichever the case, it is evident that the cholesterol pathway and its
418 underlying genes might represent a reliable marker describing the nutritional status of fish and
419 should be further investigated.

420 A well-known characteristic and important limitation of plant ingredients is the presence of
421 protease inhibitors, proteins involved in the plant's defence mechanisms against herbivores
422 (Hartl et al., 2011). Protease inhibitors are tightly linked to proteins and might consequently
423 be co-extracted with the protein fraction during the production of concentrates. In the gut
424 lumen, protease inhibitors impact on feed digestibility supposedly by interfering with
425 endogenous proteases such as trypsins, chymotrypsins, etc., which in turn can result in over-
426 secretion of these enzymes from pancreatic tissue and consequent loss of endogenous
427 resources (e.g. sulphur-rich amino acids methionine and cysteine), eventually leading to
428 decreased growth (Chikwati et al., 2012, Sarwar-Gilani et al., 2012). In salmon liver, we
429 detected a significantly increased expression of genes coding for pancreatic enzymes such as
430 proteases (trypsin, chymotrypsin, etc.) and lipases. This was not surprising considering that
431 fish liver contains several cell types including pancreatic-like cells (Rust, 2002). These results
432 may resemble the "hyperactive pancreas" syndrome observed in mammals (Sarwar-Gilani et
433 al., 2012) and encourage further investigation of these genes as biomarkers for protein
434 digestibility.

435 At a metabolic level, a number of key hepatic genes involved in amino acid metabolism were
436 down-regulated including those coding for enzymes catalysing the degradation of essential
437 amino acids such as threonine (glycine C-acetyltransferase), tryptophan (indoleamine 2,3-
438 dioxygenase) and methionine (S-adenosylmethionine synthetase) which suggested a

compensatory response to preserve these important nutrients. There was also up-regulation in liver of two genes fundamental to the citric acid cycle (2-oxoglutarate dehydrogenase and succinate dehydrogenase), which suggested an activation of energy metabolism. Interestingly, the up-regulation of the carnitine O-palmitoyltransferase gene, involved in the β-oxidation of fatty acids via the synthesis of acyl-CoA, possibly reflected the undernourished condition of fish fed 300 g kg^{-1} SBM, with the liver approaching a ketotic state. This is consistent with a previous study demonstrating a number of similarities in the hepatic response of Atlantic salmon to restricted feeding and feed containing extracted SBM (Skugor et al., 2011).

Expression data from the liver suggested that this organ also participated in the innate immune response at a systemic level. This was evident from the up-regulation of several genes of the complement cascade, which plays a crucial role in mediating non-specific defence against pathogens in the blood, and the liver is the major site of production of complement proteins (Jain et al., 2014). The up-regulation of the MASP1 genes suggested that activation of the cascade possibly occurred through the lectin pathway, which is stimulated through recognition of carbohydrates or glycoproteins present on invading pathogens or at the lesion. Notably, CD59 antigen abundance decreased, suggesting a stimulation of the formation of the membrane attack complex that is formed as an end result of the complement cascade. Activation of proteins of the complement cascade in the liver was not reported previously in Atlantic salmon fed a lower level (200 g kg^{-1}) of SBM (Skuggor et al., 2011). Since the microarray utilized in the earlier study (NCBI GEO Omnibus GPL6154) contained probes targeting several complement component proteins, it can be hypothesized that the activation of this pathway may occur at SBM inclusions above 200 g kg^{-1} and that below this level the process remains more localized to the intestine.

The gene expression response detected in the distal intestine was consistent with the results of recent studies investigating SBM-induced enteritis (Sahlmann et al., 2013) and that caused by

pea protein concentrate supplemented with soyasaponin (Kortner et al., 2012), and significantly overlapped with the immune response developed during human inflammatory bowel disease (Maloy and Powrie, 2011). The present results confirmed, at an individual gene level (dataset available as supplementary material in Array Express), the majority of the immune-related responses of chronically inflamed distal intestine in salmon reported previously (Kortner et al., 2012, Skugor et al., 2011). These included *a*) the activation of T-cell mediated processes via the up-regulation of the CD86 antigen, CTLA4 IL-18 and IL-22 and increased expression of T-cell receptors, which was also confirmed by the up-regulation of the T-cell receptor signalling pathway identified by gene-set analysis; *b*) activation of TNF- and NF-kB-mediated response and up-regulation of components of the respiratory burst complex through the up-regulation of the TNF signalling pathway, a critical trigger for the release of pro-inflammatory cytokines and also the activation of a number of intracellular pathways eventually leading to apoptosis and cell survival (Chu, 2013); and finally *c*) up-regulation of anti-inflammatory proteins, notably annexin A1, that was found to increase by approximately 10-fold in agreement with previous studies (Kortner et al., 2012). All these genes represent an informative pool of potential molecular markers of the immune response in the salmon intestine. More focused immunological analyses of intestinal transcriptome profiling during the development of enteritis have been reported previously (Kortner et al., 2012, Sahlmann et al., 2013). In addition to these results, we reported a notable intestinal up-regulation of a set of pathways involved in protein synthesis (ribosome, proteasome, tRNA biosynthesis, RNA transport, etc.) and cell proliferation (cell cycle, apoptosis, etc.) that implied active regeneration of the damaged intestinal tissue.

Metabolic pathways and other cellular and physiological processes such as digestion and absorption of nutrients (proteins, lipid, vitamins and minerals) were also significantly affected in the distal intestine. A down-regulation of virtually all metabolic pathways following the SBM-induced nutritional stress was observed, consistent with studies indicating that distal

intestine undergoing severe inflammation was characterized by "tissue malfunction" (Kortner et al., 2012, Sahlmann et al., 2013). Studies from other fish species revealed that impaired metabolism could also develop independently from the immune response (Murray et al., 2010), which suggested that some of the effects on metabolism might be possibly associated with the nutritional properties of SBM or plant products in general rather than tissue malfunction *per se*. To better identify and characterise these processes it will be necessary to analyse dietary treatments inducing a lower (subclinical) level of stress that does not directly affect the functionality of the intestine (e.g. 100 g kg^{-1}). Marked down-regulation of genes involved in cellular processes associated with lysosomes, peroxisomes and, to a lesser extent, phagosomes, whose primary role is the degradation of various cellular metabolites, as well as the digestion of invading pathogens, was possibly also linked to the tissue impairment.

Amongst the processes affected, lipid and in particular sterol metabolism in the distal intestine appeared to be altered by the use of plant proteins. We observed a down-regulation of the genes involved in several pathways of lipid metabolism, digestion and absorption, primary bile acid biosynthesis and secretion. This was in agreement with previous observations in salmon fed 200 g kg^{-1} SBM where processes of lipid and bile uptake were impaired (Kortner et al., 2013). It is unclear whether impaired lipid metabolism in the distal intestine was a consequence of the tissue damage or if it was affected by the presence of certain ANFs such as, for example, soyasaponins, that interfere with the normal absorption of these nutrients and are known to cause hypocholesterolemia (Francis et al., 2001, Francis et al., 2002, Gu et al., 2014, Kortner et al., 2013).

In conclusion, the present study contributes towards the development of a nutrigenomic model describing molecular signatures observed following nutritional stress induced by dietary plant proteins. Our results build on recent nutrigenomic findings in Atlantic salmon fed pea protein concentrate and on those of salmon undergoing restricted/SBM feeding

515 towards the identification of diagnostic markers in the intestine. More importantly, the
516 analyses of the liver transcriptome provided a platform for selection of responses that could
517 be more general and related to the overall performance, and hence have the potential as
518 predictive markers for monitoring health, welfare and performance traits. This study is a
519 comprehensive report on transcriptome profiles of distal intestine and liver, highlighting the
520 role of the latter tissue in fish undergoing SBM-induced nutritional stress.

521 **Acknowledgements**

522 KLB and this study were funded by the European Commission FP7 Combination of
523 Collaborative projects and Coordination and Support Actions (CP-CSA) Project No. 262336,
524 Aquaculture infrastructures for excellence in European fish research (AQUAEXCEL). The
525 authors also acknowledge the support of the MASTS pooling initiative (The Marine Alliance
526 for Science and Technology for Scotland) in the completion of this study. MASTS is funded
527 by the Scottish Funding Council (grant reference HR09011) and contributing Institutions.

528 **References**

529 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment
530 search tool. J. Mol. Biol. 215, 403-410.

531 Azaza, M.S., Wassim, K., Mensi, F., Abdelmouleh, A., Brini, B., Kraïem, M.M., 2009.
532 Evaluation of faba beans (*Vicia faba* L. var. minuta) as a replacement for soybean meal in
533 practical diets of juvenile Nile tilapia *Oreochromis niloticus*. Aquaculture 287, 174-179.

534 Baeverfjord, G., Krogdahl, A., 1996. Development and regression of soybean meal induced
535 enteritis in Atlantic salmon, *Salmo salar* L., distal intestine: a comparison with the intestines
536 of fasted fish. J. Fish Dis. 19, 375-387.

Bakke-McKellep, A.M., Frøystad, M.K., Lilleeng, E., Dapra, F., Refstie, S., Krogdahl, Å, Landsverk, T., 2007. Response to soy: T-cell-like reactivity in the intestine of Atlantic salmon, *Salmo salar* L. J. Fish Dis. 30, 13-25.

Bicskei, B., Bron, J., Glover, K., Taggart, J., 2014. A comparison of gene transcription profiles of domesticated and wild Atlantic salmon (*Salmo salar* L.) at early life stages, reared under controlled conditions. BMC Genomics 15, 884.

Booth, M.A., Allan, G.L., Frances, J., Parkinson, S., 2001. Replacement of fish meal in diets for Australian silver perch, *Bidyanus bidyanus*: IV. Effects of dehulling and protein concentration on digestibility of grain legumes. Aquaculture 196, 67-85.

Buttle, L.G., Burrells, A.C., Good, J.E., Williams, P.D., Southgate, P.J., Burrells, C., 2001. The binding of soybean agglutinin (SBA) to the intestinal epithelium of Atlantic salmon, *Salmo salar* and Rainbow trout, *Oncorhynchus mykiss*, fed high levels of soybean meal. Vet. Immunol. Immunopathol. 80, 237-244.

Chikwati, E.M., Venold, F.F., Penn, M.H., Rohloff, J., Refstie, S., Guttvik, A., Hillestad, M., Krogdahl, Å, 2012. Interaction of soyasaponins with plant ingredients in diets for Atlantic salmon, *Salmo salar* L. Br. J. Nutr. 107, 1570-1590.

Chomczynski, P., Mackey, K., 1995. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques 19, 942-945.

Chu, W., 2013. Tumor necrosis factor. Cancer Lett. 328, 222-225.

Collins, S.A., Øverland, M., Skrede, A., Drew, M.D., 2013. Effect of plant protein sources on growth rate in salmonids: Meta-analysis of dietary inclusion of soybean, pea and canola/rapeseed meals and protein concentrates. Aquaculture 400–401, 85-100.

560 Francis, G., Kerem, Z., Makkar, H.P.S., Becker, K., 2002. The biological action of saponins
561 in animal systems: a review. Br. J. Nutr. 88, 587-605.

562 Francis, G., Makkar, H.P.S., Becker, K., 2001. Antinutritional factors present in plant-derived
563 alternate fish feed ingredients and their effects in fish. Aquaculture 199, 197-227.

564 Gatlin, D.M., Barrows, F.T., Brown, P., Dabrowski, K., Gaylord, T.G., Hardy, R.W., Herman,
565 E., Hu, G., Krogdahl, Å, Nelson, R., Overturf, K., Rust, M., Sealey, W., Skonberg, D., J
566 Souza, E., Stone, D., Wilson, R., Wurtele, E., 2007. Expanding the utilization of sustainable
567 plant products in aquafeeds: a review. Aquacult. Res. 38, 551-579.

568 Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier,
569 L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F.,
570 Li, C., Maechler, M., Rossini, A., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.,
571 Zhang, J., 2004. Bioconductor: open software development for computational biology and
572 bioinformatics. Genome Biol. 5, R80.

573 Glencross, B., Evans, D., Dods, K., McCafferty, P., Hawkins, W., Maas, R., Sipsas, S., 2005.
574 Evaluation of the digestible value of lupin and soybean protein concentrates and isolates when
575 fed to rainbow trout, *Oncorhynchus mykiss*, using either stripping or settlement faecal
576 collection methods. Aquaculture 245, 211-220.

577 Gu, M., Kortner, T.M., Penn, M., Hansen, A.K., Krogdahl, Å, 2014. Effects of dietary plant
578 meal and soya-saponin supplementation on intestinal and hepatic lipid droplet accumulation
579 and lipoprotein and sterol metabolism in Atlantic salmon (*Salmo salar* L.). Br. J. Nutr. 111,
580 432-444.

581 Hartl, M., Giri, A.P., Kaur, H., Baldwin, I.T., 2011. The multiple functions of plant serine
582 protease inhibitors: Defense against herbivores and beyond. Plant Signal. Behav. 6, 1009-
583 1011.

584 Hartviksen, M., Bakke, A., Vecino, J., Ringø, E., Krogdahl, Å, 2014. Evaluation of the effect
585 of commercially available plant and animal protein sources in diets for Atlantic salmon
586 (*Salmo salar* L.): digestive and metabolic investigations. Fish Physiol. Biochem. 40, 1621-37.

587 Jain, U., Otley, A.R., Van Limbergen, J., Stadnyk, A.W., 2014. The complement system in
588 inflammatory bowel disease. Inflamm. Bowel Dis. 20, 1628-37.

589 Kaushik, S.J., Covès, D., Dutto, G., Blanc, D., 2004. Almost total replacement of fish meal by
590 plant protein sources in the diet of a marine teleost, the European seabass, *Dicentrarchus*
591 *labrax*. Aquaculture 230, 391-404.

592 Kortner, T.M., Skugor, S., Penn, M.H., Mydland, L.T., Djordjevic, B., Hillestad, M.,
593 Krasnov, A., Krogdahl, A., 2012. Dietary soyasaponin supplementation to pea protein
594 concentrate reveals nutrigenomic interactions underlying enteropathy in Atlantic salmon
595 (*Salmo salar*). BMC Vet. Res. 8, 101.

596 Kortner, T.M., Gu, J., Krogdahl, Å, Bakke, A.M., 2013. Transcriptional regulation of
597 cholesterol and bile acid metabolism after dietary soyabean meal treatment in Atlantic salmon
598 (*Salmo salar* L.). Br. J. Nutr. 109, 593-604.

599 Krogdahl, Å, Penn, M., Thorsen, J., Refstie, S., Bakke, A.M., 2010. Important antinutrients in
600 plant feedstuffs for aquaculture: an update on recent findings regarding responses in
601 salmonids. Aquacult. Res. 41, 333-344.

602 Luo, W., Brouwer, C., 2013. Pathview: an R/Bioconductor package for pathway-based data
603 integration and visualization. Bioinformatics 29, 1830-1831.

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., Woolf, P.J., 2009. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10, 161.

Maloy, K.J., Powrie, F., 2011. Intestinal homeostasis and its breakdown in inflammatory bowel disease. Nature 474, 298-306.

Martinez-Rubio, L., Morais, S., Evensen, Ø, Wadsworth, S., Ruohonen, K., Vecino, J.L.G., Bell, J.G., Tocher, D.R., 2012. Functional feeds reduce heart inflammation and pathology in Atlantic salmon (*Salmo salar* L.) following experimental challenge with Atlantic Salmon Reovirus (ASRV). PLoS ONE 7, e40266.

Morais, S., Silva, T., Cordeiro, O., Rodrigues, P., Guy, D.R., Bron, J.E., Taggart, J.B., Bell, J.G., Tocher, D.R., 2012a. Effects of genotype and dietary fish oil replacement with vegetable oil on the intestinal transcriptome and proteome of Atlantic salmon (*Salmo salar*). BMC Genomics 13, 448.

Morais, S., Taggart, J., Guy, D., Bell, J., Tocher, D., 2012b. Hepatic transcriptome analysis of inter-family variability in flesh n-3 long-chain polyunsaturated fatty acid content in Atlantic salmon. BMC Genomics 13, 410.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182-5.

Muller, M., Kersten, S., 2003. Nutrigenomics: goals and strategies. Nat. Rev. Genet. 4, 315-322.

Murray, H.M., Lall, S.P., Rajaselvam, R., Boutilier, L.A., Blanchard, B., Flight, R.M., Colombo, S., Mohindra, V., Douglas, S.E., 2010. A nutrigenomic analysis of intestinal response to partial soybean meal replacement in diets for juvenile Atlantic halibut, *Hippoglossus hippoglossus*, L. Aquaculture 298, 282-293.

National Research Council (NRC), 2011. Nutrient Requirements of Fish and Shrimp. National Academy Press, Washington, DC.

Øverland, M., Sørensen, M., Storebakken, T., Penn, M., Krogdahl, Å, Skrede, A., 2009. Pea protein concentrate substituting fish meal or soybean meal in diets for Atlantic salmon (*Salmo salar*) - Effect on growth performance, nutrient digestibility, carcass composition, gut health, and physical feed quality. Aquaculture 288, 305-311.

Panserat, S., Hortopan, G.A., Plagnes-Juan, E., Kolditz, C., Lansard, M., Skiba-Cassy, S., Esquerré, D., Geurden, I., Médale, F., Kaushik, S., Corraze, G., 2009. Differential gene expression after total replacement of dietary fish meal and fish oil by plant products in rainbow trout (*Oncorhynchus mykiss*) liver. Aquaculture 294, 123-131.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (http://www. R-project. org/)

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., Smyth, G.K., 2007. A comparison of background correction methods for two-colour microarrays. Bioinformatics 23, 2700-2707.

Russell, S., Meadows, L.A., Russell, R.R., 2009. Microarray Technology in Practice. Elsevier, San Diego (USA).

Rust, M.B., 2002. Nutritional Physiology. In: Halver, J.E., Hardy, R.W. (Eds.), Fish Nutrition. Academic Press, New York, pp. 367-452.

Sahlmann, C., Sutherland, B.J.G., Kortner, T.M., Koop, B.F., Krogdahl, Å, Bakke, A.M., 2013. Early response of gene expression in the distal intestine of Atlantic salmon (*Salmo salar* L.) during the development of soybean meal induced enteritis. Fish Shellfish Immunol. 34, 599-609.

650 Sarwar Gilani, G., Wu Xiao, C., Cockell, K.A., 2012. Impact of antinutritional factors in food
651 proteins on the digestibility of protein and the bioavailability of amino acids and on protein
652 quality. Br. J. Nutr. 108, S315-S332.

653 Schwarz, M., Wright, A.C., Davis, D.L., Nazer, H., Björkhem, I., Russell, D.W., 2000. The
654 bile acid synthetic gene 3beta-hydroxy-Delta(5)-C(27)-steroid oxidoreductase is mutated in
655 progressive intrahepatic cholestasis. J. Clin. Invest. 106, 1175-1184.

656 Silver, J.D., Ritchie, M.E., Smyth, G.K., 2009. Microarray background correction: maximum
657 likelihood estimation for the normal-exponential convolution. Biostatistics 10, 352-363.

658 Skugor, S., Grisdale-Helland, B., Refstie, S., Afanasyev, S., Vielma, J., Krasnov, A., 2011.
659 Gene expression responses to restricted feeding and extracted soybean meal in Atlantic
660 salmon (*Salmo salar* L.). Aquacult. Nutr. 17, 505-517.

661 Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential
662 expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3, Article3.

663 Smyth, G.K., Speed, T., 2003. Normalization of cDNA microarray data. Methods 31, 265-
664 273.

665 Suzuki, R., Shimodaira, H., 2011. pvclust: Hierarchical Clustering with P-Values via
666 Multiscale Bootstrap Resampling (http://cran.r-project.org/web/packages/pvclust/)

667 Tacchi, L., Bron, J.E., Taggart, J.B., Secombes, C.J., Bickerdike, R., Adler, M.A., Takle, H.,
668 Martin, S.A.M., 2011. Multiple tissue transcriptomic responses to *Piscirickettsia salmonis* in
669 Atlantic salmon (*Salmo salar*). Physiol. Genom. 43, 1241-1254.

Tacchi, L., Secombes, C., Bickerdike, R., Adler, M., Venegas, C., Takle, H., Martin, S., 2012. Transcriptomic and physiological responses to fishmeal substitution with plant proteins in formulated feed in farmed Atlantic salmon (*Salmo salar*). BMC Genomics 13, 363.

The World Bank, 2013. Fish to 2030: Prospects for Fisheries and Aquaculture. (www. worldbank.org)

Urán, P.A., Schrama, J.W., Rombout, J.H.W.M., Taverne-Thiele, J.J., Obach, A., Koppe, W., Verreth, J.A.J., 2009. Time-related changes of the intestinal morphology of Atlantic salmon, *Salmo salar* L., at two different soybean meal inclusion levels. J. Fish Dis. 32, 733-744.

Urán, P.A., Schrama, J.W., Rombout, J.H.W.M., Obach, A., Jensen, L., Koppe, W., Verreth, J.A.J., 2008. Soybean meal-induced enteritis in Atlantic salmon (*Salmo salar* L.) at different temperatures. Aquacult. Nutr. 14, 324-330.

Wacyk, J., Powell, M., Rodnick, K., Overturf, K., Hill, R.A., Hardy, R., 2012. Dietary protein source significantly alters growth performance, plasma variables and hepatic gene expression in rainbow trout (*Oncorhynchus mykiss*) fed amino acid balanced diets. Aquaculture 356–357, 223-234.

Wickham, H., 2009. H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.

Figure Legends

Figure 1. Hierarchical clustering of gene expression profiles in liver and intestine. Two types of *p*-values are provided: AU (Approximately Unbiased) is computed by multiscale bootstrap resampling (1000 rep) and reported in red and BP (Bootstrap Probability) is computed by normal bootstrap resampling and reported in green.

Figure 2. Histograms plotting the results of differential expression analysis in liver and intestine of salmon fed diet S30 compared to fish fed diet S0. Figure 2a plots the number of differentially expressed genes based on *p* value and indicates that a larger number of genes with a $p < 0.01$ (the dashed line depict the *p* value = 0.01) were present in the intestine compared to the liver. Figure 2b only plots genes with $p < 0.01$. FC is fold change.

Figure 3. Bubble graph plotting GAGE results from the *2d* test (a) and *1d* test (b) in the liver. The x-axis plots the fold change (mean statistic) while the y-axis plots the Log_{10} transformed *q*-value. The size of the bubbles is proportional to the number of genes used by the algorithm to test the gene-set. Colours of bubbles refer to the biological process (KEGG classification). Figures are not on the same scale.

Figure 4 Complement cascade generated with the R package *pathview* of affected genes in the liver (Luo and Brouwer, 2013). Red arrows indicate activation, blue arrow inhibition and dotted arrows indirect effect. Highlighted in blue are the names of the differentially expressed genes and the corresponding KEGG identifier.

Figure 5. Bubble graph plotting GAGE results from the *2d* test (a) and *1d* test (b) in the intestine. The x-axis plots the fold change (mean statistic) while the y-axis plots the Log_{10} transformed *q*-value. The size of the bubbles is proportional to the number of genes applied by the algorithm to test the gene-set. Colours of bubbles refer to the biological process type (KEGG classification). The dashed line denotes the cut-off of "highly significant gene-sets" ($q < 0.0001$). Figures are not on the same scale.

713 **Table 1.** Formulation (g / Kg) and analyzed proximate compositions

714 of the experimental diets

	S0	S10	S20	S30
Fishmeal[1]	420.10	368.00	344.00	291.00
Hi-Pro Soymeal[2]	0.00	100.00	200.00	299.50
Corn gluten 60	77.00	77.00	33.00	23.00
Sunflower cake	125.00	104.70	95.80	70.60
Horsebeans	160.00	135.60	105.30	85.30
Fish oil std 18	154.00	158.00	164.00	170.00
Rapeseed oil	40.00	40.00	40.00	40.00
Amino acids	5.70	6.40	7.30	12.10
Vitamins and Minerals	31.90	31.90	31.90	31.90
Proximate composition				
Protein - crude (%)	44.06	44.18	43.83	43.59
Fat - crude (%)	25.97	25.96	25.96	26.02
Ash (%)	7.28	7.47	7.75	7.42
Energy - gross (MJ/kg)	23.65	23.78	23.70	23.75

715

716 [1]NA LT70; [2]Soya cake 48 Hi-Pro (solvent extracted soybean meal cake)

717 All values are represented as g kg^{-1} unless otherwise stated.

718

719

720 **Table 2**. Growth performance and somatic indexes (±SD)

Feed	Initial weight	Final weight	SGR
S0	175±27	424±78[a]	1.03±0.22[a]
S10	170±25	422±76[a]	1.03±0.21[a]
S20	176±28	405±67[a]	0.96±0.16[b]
S30	175±25	378±67[b]	0.87±0.18[c]

721

722 Number of fish in each triplicate tank was 45. Weights are in g. Calculations of SGR (specific
723 growth rate) were based on growth of individual fish. Values within a column not sharing
724 superscript letters are significantly different ($p < 0.05$). Growth data was compared using a
725 one-way ANOVA for statistical significance.

726

Table 3. Forty-four genes differentially expressed ($p < 0.01$) in both liver and intestine of fish fed S30 compared with S0. LogFC is the log2 transformed fold change.

		LIVER		INTESTINE	
KOID	**Description**	**LogFC**	**P.Value**	**LogFC**	**P.Value**
K02377	GDP-L-fucose synthase	-1.01	0.0028	-4.15	< 0.0001
K15985	cAMP-dependent protein kinase inhibitor alpha	-0.68	0.0013	-1.22	< 0.0001
K00736	alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	-0.54	0.0035	-1.17	< 0.0001
K00864	glycerol kinase	-0.53	0.0050	-1.10	< 0.0001
K03940	NADH dehydrogenase (ubiquinone) Fe-S protein 7	-0.53	0.0076	-0.75	0.0004
K11251	histone H2A	-0.53	0.0066	-0.79	0.0002
K10244	elongation of very long chain fatty acids protein 5	-0.52	0.0039	-1.12	< 0.0001
K10703	very-long-chain (3R)-3-hydroxyacyl-dehydratase	-0.50	0.0013	-1.22	< 0.0001
K00522	ferritin heavy chain	-0.40	0.0063	-0.85	< 0.0001
K04644	clathrin light chain A	-0.37	0.0031	-0.51	0.0001
K11428	histone-lysine N-methyltransferase SETD8	-0.30	0.0034	-0.28	0.0057
K06109	Ras-related protein Rab-13	-0.29	0.0074	-0.46	< 0.0001
K02977	small subunit ribosomal protein S27Ae	-0.23	0.0099	-0.24	0.0064
K09850	Ras association domain-containing protein 1	0.26	0.0084	0.27	0.0063
K14525	ribonucleases P/MRP protein subunit RPP25	0.33	0.0023	0.55	< 0.0001
K12859	U5 snRNP protein, DIM1 family	0.34	0.0090	0.40	0.0024
K02932	large subunit ribosomal protein L5e	0.35	0.0091	0.59	< 0.0001
K02889	large subunit ribosomal protein L21e	0.35	0.0049	0.57	< 0.0001
K03262	translation initiation factor 5	0.36	0.0019	0.35	0.0027
K02868	large subunit ribosomal protein L11e	0.38	0.0016	0.51	< 0.0001
K02984	small subunit ribosomal protein S3Ae	0.39	0.0051	0.53	0.0003
K00940	nucleoside-diphosphate kinase	0.40	0.0064	0.60	0.0001
K12298	bile salt-stimulated lipase	0.41	0.0004	0.30	0.0058
K03996	complement component 7	0.41	0.0061	0.52	0.0009
K10086	ER degradation enhancer, mannosidase alpha 3	0.43	0.0076	0.43	0.0072
K03097	casein kinase II subunit alpha	0.47	0.0007	0.58	< 0.0001
K02936	large subunit ribosomal protein L7Ae	0.52	0.0003	0.70	< 0.0001
K08765	carnitine O-palmitoyltransferase 1	0.65	0.0023	0.74	0.0007
K09414	heat shock transcription factor 1	0.77	< 0.0001	0.45	0.0038
K14397	cleavage and polyadenylation specificity factor subunit 5	0.90	0.0002	0.96	< 0.0001
K01594	sulfinoalanine decarboxylase	1.03	0.0060	2.56	< 0.0001
K11262	acetyl-CoA carboxylase / biotin carboxylase	-0.84	0.0015	0.70	0.0068
K00222	delta14-sterol reductase	-0.61	0.0028	0.64	0.0017
K01507	inorganic pyrophosphatase	-0.53	0.0042	1.08	< 0.0001
K01875	seryl-tRNA synthetase	-0.47	0.0043	0.82	< 0.0001
K04440	c-Jun N-terminal kinase	-0.25	0.0062	0.30	0.0015
K10577	ubiquitin-conjugating enzyme E2 I	0.32	0.0057	-0.39	0.0011
K07292	hepatocyte nuclear factor 4-alpha	0.36	0.0035	-0.51	0.0001
K10574	ubiquitin-conjugating enzyme E2 B	0.38	0.0059	-0.48	0.0008
K06624	cyclin-dependent kinase inhibitor 1B	0.55	0.0034	-0.81	< 0.0001
K08550	estrogen receptor alpha	0.63	0.0082	-0.77	0.0015
K12408	cholest-5-ene-3beta,7alpha-diol 3beta-dehydrogenase	0.68	0.0055	-1.40	< 0.0001
K01103	6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase	0.75	0.0004	-2.99	< 0.0001
K04630	guanine nucleotide-binding protein G(i) subunit alpha	1.52	0.0018	-4.41	< 0.0001

Up-regulated genes are highlighted in red, while down-regulated genes in green. Colour intensity is relative to the magnitude of change. All expression values are reported as Log_2 fold change (LogFC) relative to the control treatment S0. KOID = KEGG identifier.

Table 4. Selection of genes differentially expressed only in the liver of fish fed S30 compared with S0. LogFC is the log2 transformed fold change.

	KOID	Gene	LogFC	P Value
Immune System				
Complement Cascade	K03990	complement component 3	1.15	< 0.001
Metabolism				
Amino acid	K00639	glycine C-acetyltransferase	-0.49	< 0.001
Amino acid	K00789	S-adenosylmethionine synthetase	-0.85	0.003
Carbohydrate	K12047	maltase-glucoamylase	0.62	0.010
Energy	K00164	2-oxoglutarate dehydrogenase E1 component	0.61	< 0.001
Energy	K00234	succinate dehydrogenase (ubiquinone) flavoprotein subunit	0.41	0.003
Lipid	K01597	diphosphomevalonate decarboxylase	-0.63	< 0.001
Lipid	K10205	elongation of very long chain fatty acids protein 2	0.54	< 0.001
Lipid	K01830	prostaglandin-H2 D-isomerase	-0.65	0.001
Lipid	K07296	adiponectin	-0.89	0.002
Lipid	K01823	isopentenyl-diphosphate delta-isomerase	-0.63	0.005
Lipid	K00626	acetyl-CoA C-acetyltransferase	-0.49	0.005
Lipid	K00869	mevalonate kinase	-0.48	0.005
Lipid	K01852	lanosterol synthase	-0.41	0.006
Nucleotide	K03010	DNA-directed RNA polymerase II subunit RPB2	-0.43	0.001
Nucleotide	K01489	cytidine deaminase	-0.54	0.004
Vitamin	K01435	biotinidase	-0.86	< 0.001
Vitamin	K18271	retinol-binding protein 4	-0.90	0.009
Cell Proliferation				
Cell Cycle and Apoptosis	K04426	mitogen-activated protein kinase 5	0.44	< 0.001
Cell Cycle and Apoptosis	K10858	DNA mismatch repair protein PMS2	-0.44	< 0.001
Cell Cycle and Apoptosis	K04441	p38 MAP kinase	0.54	0.002
Cell Cycle and Apoptosis	K06245	laminin, beta 4	0.49	0.003
Cell Cycle and Apoptosis	K04436	mitogen-activated protein kinase 8 interacting protein 3	0.48	0.003
Cell Cycle and Apoptosis	K10567	endonuclease VIII-like 1	0.47	0.007
Genetic Information processing				
Proteasome	K05610	ubiquitin carboxyl-terminal hydrolase L5	-1.12	< 0.001
Protein Processing	K13431	signal recognition particle receptor subunit alpha	-0.59	< 0.001
Protein Processing	K07342	protein transport protein SEC61 subunit gamma and related proteins	-0.49	0.004
Ribosomal	K02891	large subunit ribosomal protein L22e	-0.62	< 0.001
Ribosomal	K02911	large subunit ribosomal protein L32	-0.52	< 0.001
Ribosomal	K17418	large subunit ribosomal protein L37, mitochondrial	-0.48	< 0.001
Ribosomal	K02950	small subunit ribosomal protein S12	-0.40	< 0.001
Ribosomal	K17433	large subunit ribosomal protein L52, mitochondrial	-0.52	0.002
Ribosomal	K17430	large subunit ribosomal protein L49, mitochondrial	-0.53	0.003
Ribosomal	K17428	large subunit ribosomal protein L47, mitochondrial	-0.44	0.003
Ribosomal	K17409	small subunit ribosomal protein S30, mitochondrial	-0.45	0.005
Digestive system				
Pancreatic	K01312	trypsin	1.11	< 0.001
Pancreatic	K01310	chymotrypsin	1.23	< 0.001
Pancreatic	K07886	Ras-related protein Rab-27B	0.58	< 0.001
Pancreatic	K01291	carboxypeptidase B	0.92	< 0.001
Pancreatic	K01346	pancreatic elastase II	0.58	0.001

Up-regulated genes are highlighted in red, while down-regulated genes in green. Colour intensity is relative to the magnitude of change. All expression values are reported as Log_2 fold change (LogFC) relative to the control treatment S0. KOID = KEGG identifier.

Fig. 1

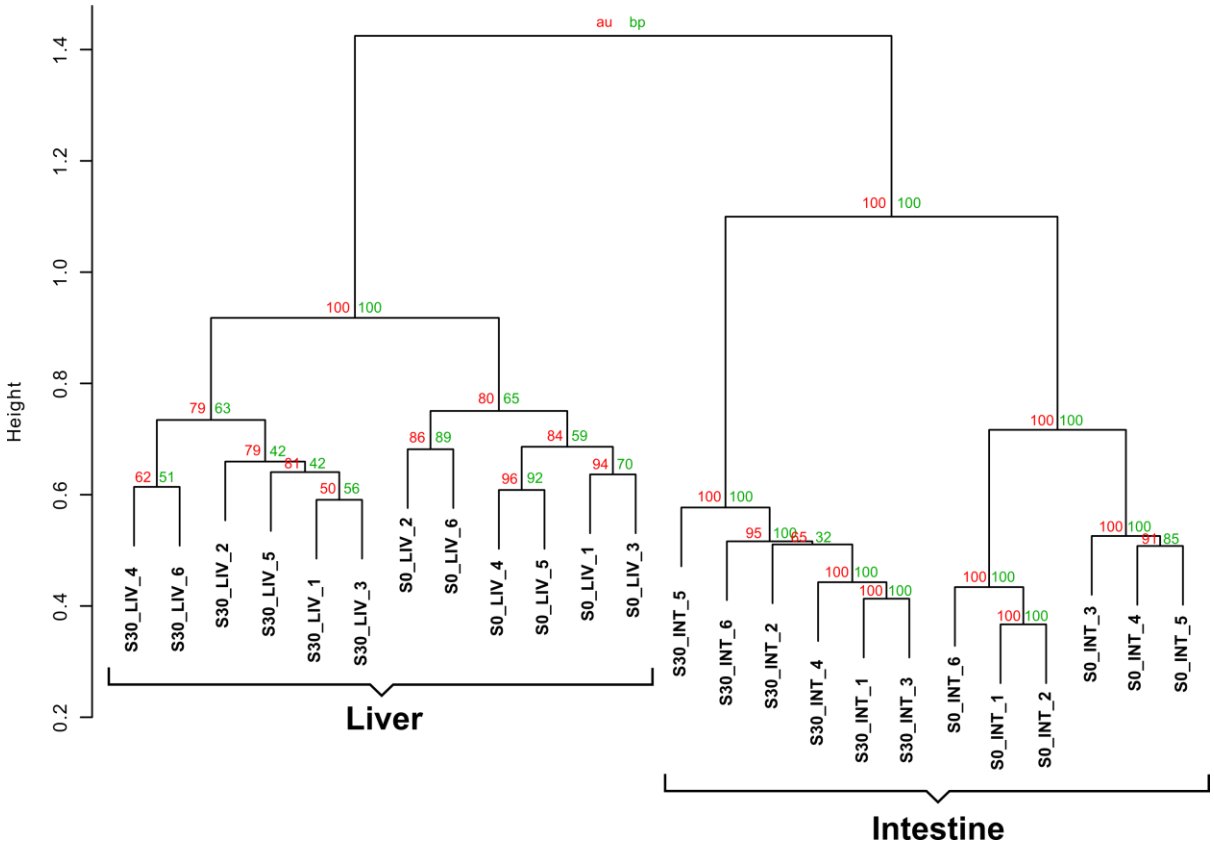

au bp
Height
0.2
0.4
0.6
0.8
1.0
1.2
1.4
S30_LIV_4
S30_LIV_6
S30_LIV_2
S30_LIV_5
S30_LIV_1
S30_LIV_3
S0_LIV_2
S0_LIV_6
S0_LIV_4
S0_LIV_5
S0_LIV_1
S0_LIV_3
Liver
S30_INT_5
S30_INT_6
S30_INT_2
S30_INT_4
S30_INT_1
S30_INT_3
S0_INT_6
S0_INT_1
S0_INT_2
S0_INT_3
S0_INT_4
S0_INT_5
Intestine

Fig. 2

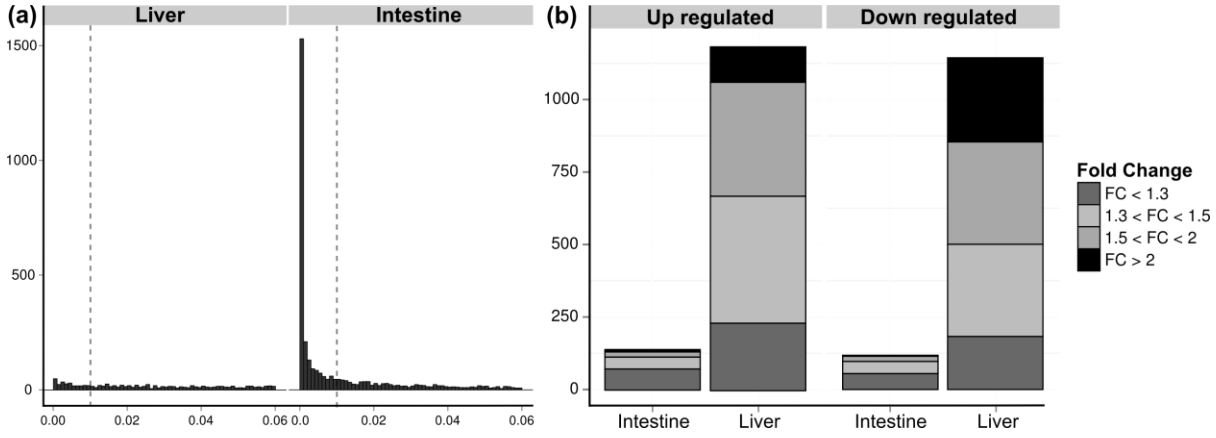

(a)
Liver
Intestine
1500
1000
500
0
0.00
0.02
0.04
0.06
0.0
0.02
0.04
0.06
(b)
Up regulated
Down regulated
1000
750
500
250
0
Intestine
Liver
Intestine
Liver
Fold Change
FC < 1.3
1.3 < FC < 1.5
1.5 < FC < 2
FC > 2

Fig. 3

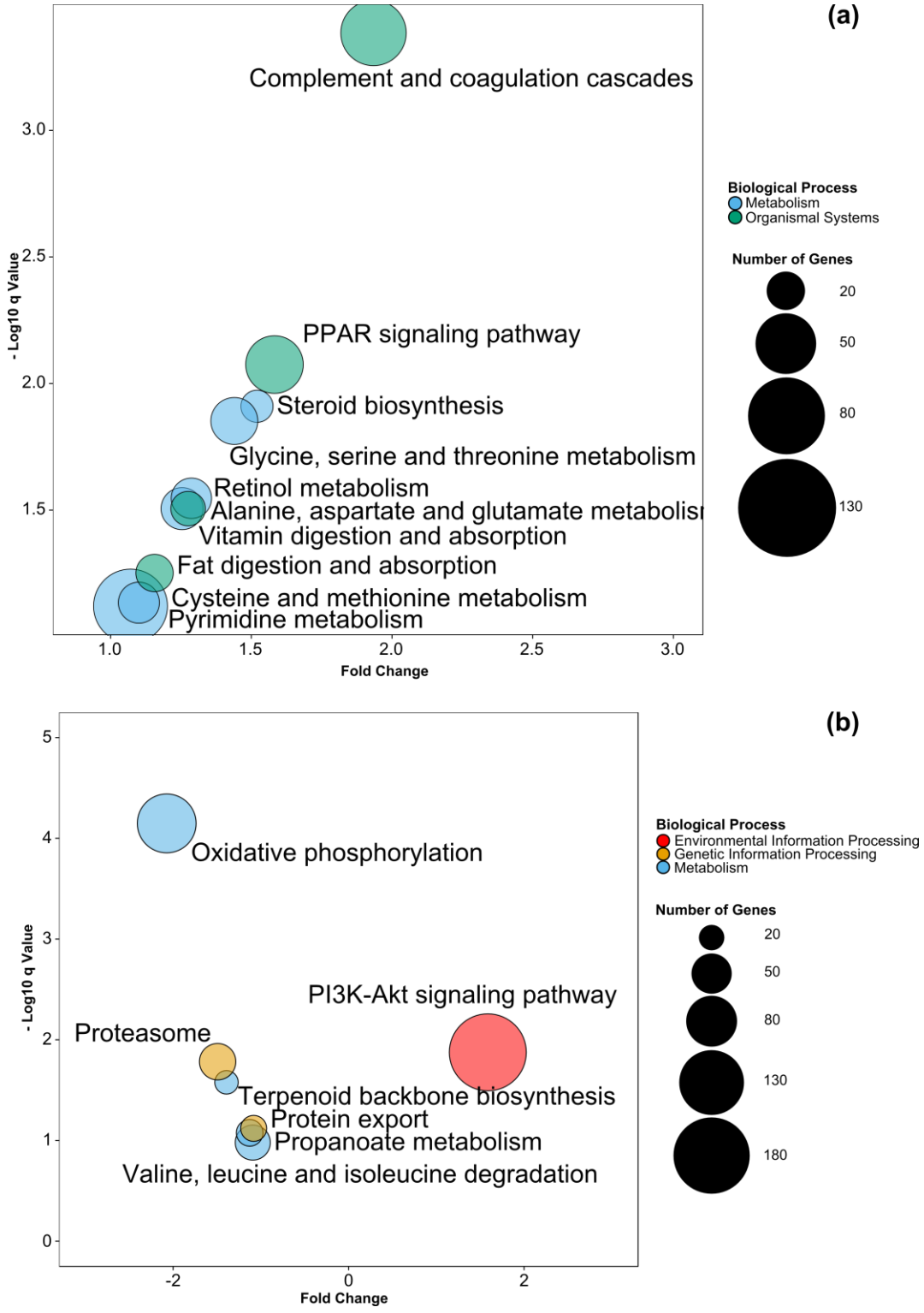

(a)
Complement and coagulation cascades
PPAR signaling pathway
Steroid biosynthesis
Glycine, serine and threonine metabolism
Retinol metabolism
Alanine, aspartate and glutamate metabolism
Vitamin digestion and absorption
Fat digestion and absorption
Cysteine and methionine metabolism
Pyrimidine metabolism
- Log10 q Value
Fold Change
Biological Process
Metabolism
Organismal Systems
Number of Genes
20
50
80
130
(b)
Oxidative phosphorylation
PI3K-Akt signaling pathway
Proteasome
Terpenoid backbone biosynthesis
Protein export
Propanoate metabolism
Valine, leucine and isoleucine degradation
- Log10 q Value
Fold Change
Biological Process
Environmental Information Processing
Genetic Information Processing
Metabolism
Number of Genes
20
50
80
130
180

Fig. 4

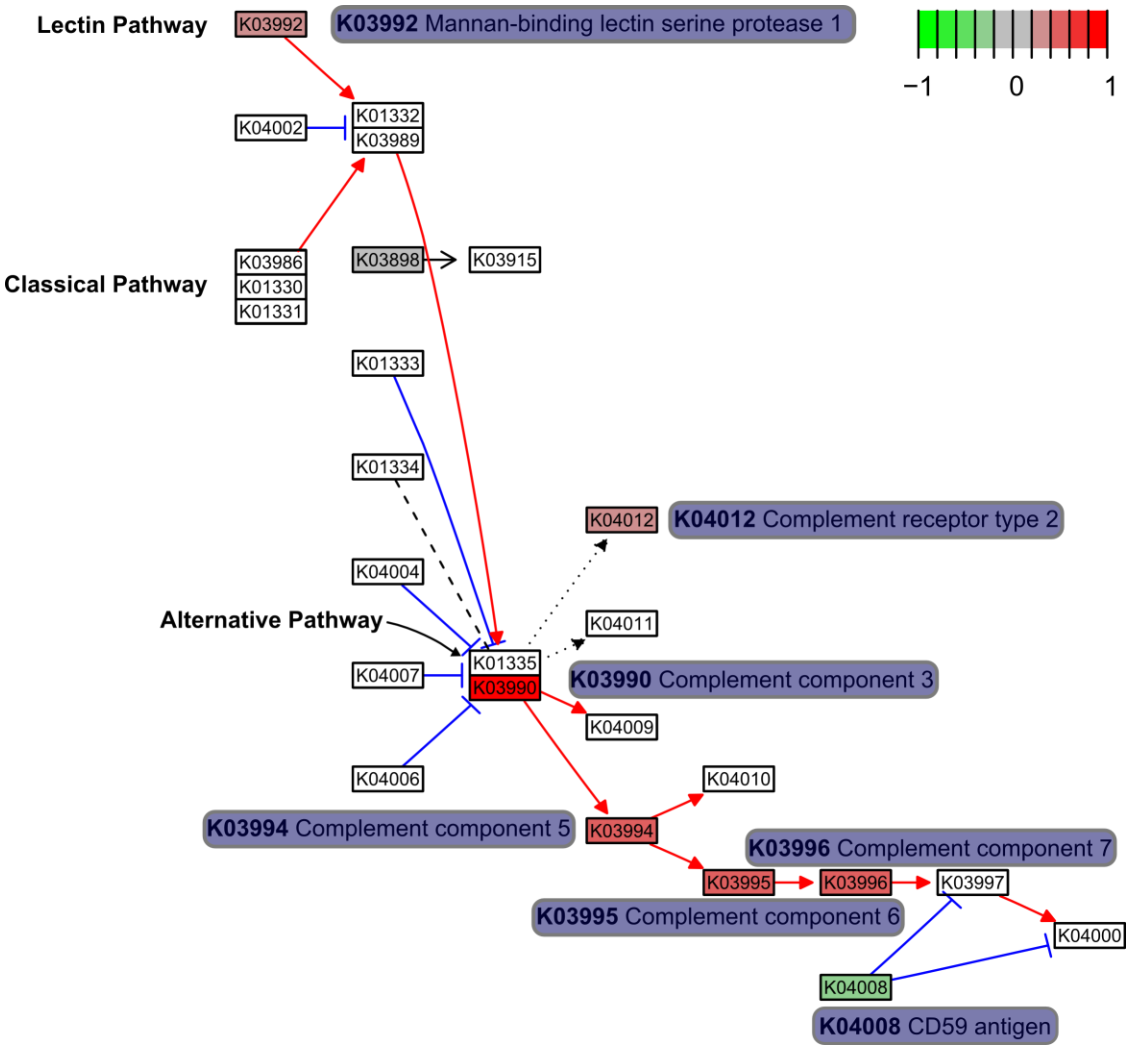
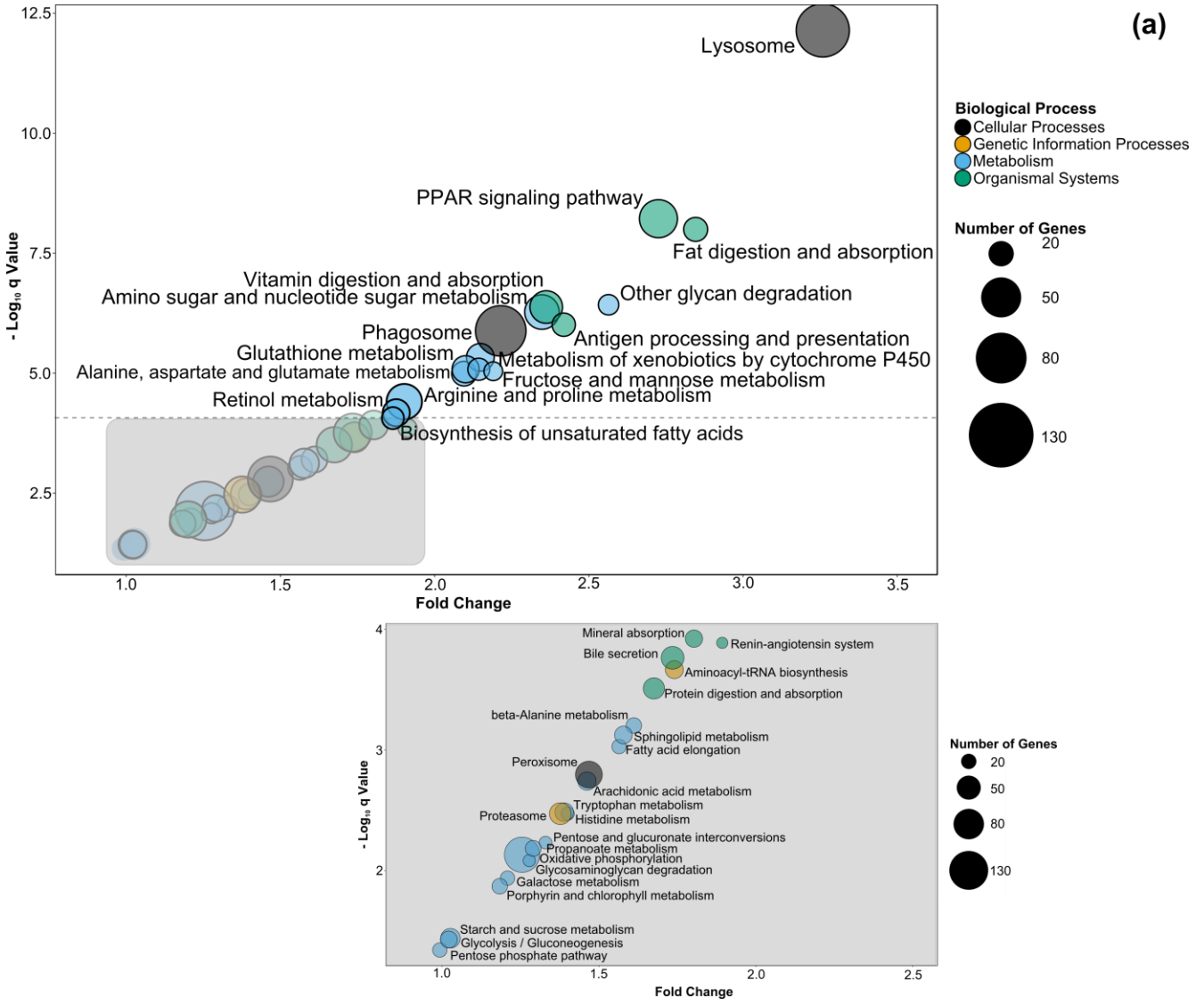
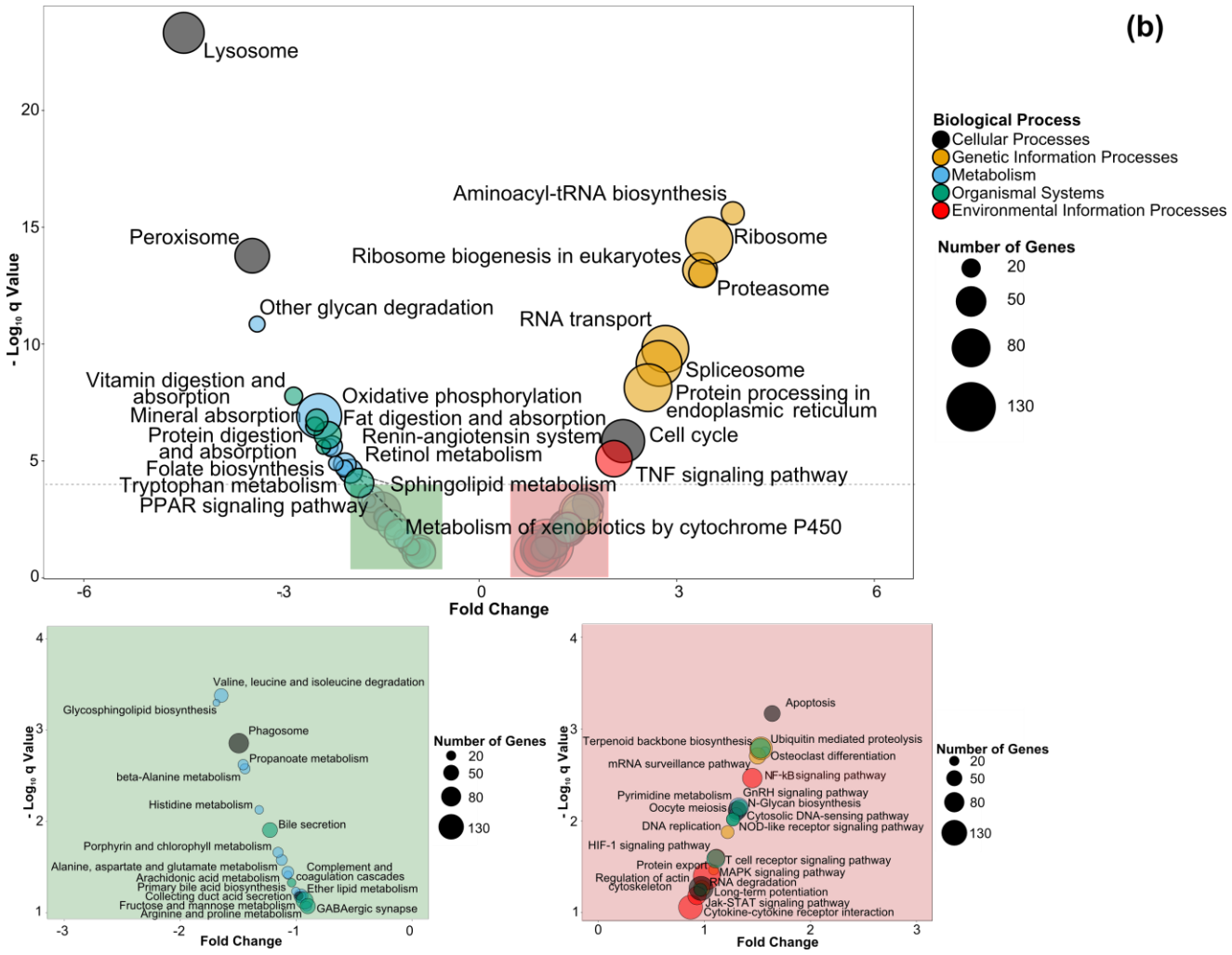

Lectin Pathway
K03992
K03992 Mannan-binding lectin serine protease 1
−1
0
1
K04002
K01332
K03989
Classical Pathway
K03986
K01330
K01331
K03898
K03915
K01333
K01334
K04012
K04012 Complement receptor type 2
K04004
Alternative Pathway
K04011
K01335
K04007
K03990
K03990 Complement component 3
K04009
K04006
K04010
K03994 Complement component 5
K03994
K03996 Complement component 7
K03995
K03996
K03997
K03995 Complement component 6
K04000
K04008
K04008 CD59 antigen

Fig 5


(a)
Lysosome
PPAR signaling pathway
Fat digestion and absorption
Vitamin digestion and absorption
Amino sugar and nucleotide sugar metabolism
Other glycan degradation
Phagosome
Antigen processing and presentation
Glutathione metabolism
Metabolism of xenobiotics by cytochrome P450
Alanine, aspartate and glutamate metabolism
Fructose and mannose metabolism
Retinol metabolism
Arginine and proline metabolism
Biosynthesis of unsaturated fatty acids
- Log10 q Value
Fold Change
Biological Process
Cellular Processes
Genetic Information Processes
Metabolism
Organismal Systems
Number of Genes
20
50
80
130
Mineral absorption
Renin-angiotensin system
Bile secretion
Aminoacyl-tRNA biosynthesis
Protein digestion and absorption
beta-Alanine metabolism
Sphingolipid metabolism
Fatty acid elongation
Peroxisome
Arachidonic acid metabolism
Tryptophan metabolism
Proteasome
Histidine metabolism
Pentose and glucuronate interconversions
Propanoate metabolism
Oxidative phosphorylation
Glycosaminoglycan degradation
Galactose metabolism
Porphyrin and chlorophyll metabolism
Starch and sucrose metabolism
Glycolysis / Gluconeogenesis
Pentose phosphate pathway


(b)
Lysosome
Peroxisome
Other glycan degradation
Aminoacyl-tRNA biosynthesis
Ribosome
Ribosome biogenesis in eukaryotes
Proteasome
RNA transport
Spliceosome
Protein processing in endoplasmic reticulum
Cell cycle
TNF signaling pathway
Vitamin digestion and absorption
Mineral absorption
Protein digestion and absorption
Folate biosynthesis
Tryptophan metabolism
PPAR signaling pathway
Oxidative phosphorylation
Fat digestion and absorption
Renin-angiotensin system
Retinol metabolism
Sphingolipid metabolism
Metabolism of xenobiotics by cytochrome P450
- Log10 q Value
Fold Change
Biological Process
Cellular Processes
Genetic Information Processes
Metabolism
Organismal Systems
Environmental Information Processes
Number of Genes
20
50
80
130
Valine, leucine and isoleucine degradation
Glycosphingolipid biosynthesis
Phagosome
Propanoate metabolism
beta-Alanine metabolism
Histidine metabolism
Bile secretion
Porphyrin and chlorophyll metabolism
Alanine, aspartate and glutamate metabolism
Arachidonic acid metabolism
Primary bile acid biosynthesis
Collecting duct acid secretion
Fructose and mannose metabolism
Arginine and proline metabolism
Complement and coagulation cascades
Ether lipid metabolism
GABAergic synapse
Apoptosis
Terpenoid backbone biosynthesis
mRNA surveillance pathway
Pyrimidine metabolism
Oocyte meiosis
DNA replication
HIF-1 signaling pathway
Protein export
Regulation of actin cytoskeleton
Ubiquitin mediated proteolysis
Osteoclast differentiation
NF-kB signaling pathway
GnRH signaling pathway
N-Glycan biosynthesis
Cytosolic DNA-sensing pathway
NOD-like receptor signaling pathway
T cell receptor signaling pathway
MAPK signaling pathway
RNA degradation
Long-term potentiation
Jak-STAT signaling pathway
Cytokine-cytokine receptor interaction