mRNA transcriptome of Atlantic cod (*Gadus morhua*) early embryos from farmed and wild broodstocks

Carlos Frederico Ceccon Lanes, Teshome Tilahun Bizuayehu, Jorge Manuel de Oliveira Fernandes, Viswanath Kiron, Igor Babiak*

Faculty of Biosciences and Aquaculture, University of Nordland, N-8049 Bodø, Norway.

* Corresponding author: Faculty of Biosciences and Aquaculture, University of Nordland, N-8049 Bodø, Norway.
Tel.: +47 755 17922; fax: +47 755 17457
E-mail address: igor.babiak@uin.no

Abstract

A number of efforts have been made to elucidate factors regulating egg quality in fish. Recently, we have shown that eggs originating from wild broodstock (WB) of Atlantic cod (*Gadus morhua*) are of superior quality to those derived from farmed broodstock (FB), and this is associated with differences in chemical composition of egg yolk. However, non-yolk cytoplasmic components accumulated during oogenesis, such as maternal transcripts, have not been studied extensively in fish. The aim of the present study was to characterize putative maternal mRNA transcriptome in fertilized eggs of Atlantic cod, and to compare transcript pools between WB and FB in order to investigate the relation between egg developmental potential and putative maternal mRNA deposits. We performed high-throughput 454 pyrosequencing. For each WB and FB groups, five cDNA libraries were individually tagged and sequenced, resulting in 98,687 (WB) and 119,333 (FB) average reads per library. Sequencing reads were *de novo* assembled, annotated, and mapped. Out of 13,726 identified isotigs, 238 were differentially expressed between WB and FB, with 155 isotigs significantly upregulated ($P<0.05$; fold change ≥ 2.0) in WB. The sequence reads were mapped to 11,340 different Atlantic cod transcripts and 158 sequences were differentially expressed between the two groups. Important transcripts involved in fructose metabolism, fatty acid metabolism, glycerophospholipid metabolism and oxidative phosphorylation were differentially represented between the two broodstock groups, showing potential as biomarkers of egg quality in teleosts. Our findings contribute to the hypothesis that maternal mRNAs affect the egg quality and, consequently, early development of fish.

Keywords: Atlantic cod; Egg quality; Maternal transcripts; RNA-seq; Transcriptome.

Introduction

In the last few decades, significant efforts have been made to elucidate the factors affecting egg quality in fish, since such knowledge is needed to overcome the major bottlenecks that restrict the expansion of both marine and freshwater fish farming (Brooks et al. 1997; Bobe and Labbé 2010). In particular, the roles of chemical constituents of yolk on early embryonic development have been intensively investigated in a number of fish species (Furuita et al. 2003; Faulk and Holt 2008; Lanes et al. 2012a). However, less data are available on the role of non-yolk cytoplasmic components such as structural and regulatory proteins, cortical alveoli content and maternal transcripts that are accumulated during oogenesis (Brooks et al. 1997; Bobe and Labbé 2010; Lubzens et al. 2010).

Maternal transcripts are essential for early embryonic development. They support basic cellular functions such as cellular metabolism, nuclear and cellular divisions, intercellular adhesion, as well as the establishment of body axes and specification of early embryonic cells until activation of zygotic transcription (reviewed by Pelegri 2003). Furthermore, maternal transcripts may also function indirectly as a source of nutrition, providing nucleotides and phosphorus during the development (Shen-Orr et al. 2010). Recent studies in rainbow trout (*Oncorhynchus mykiss*) and Atlantic halibut (*Hippoglossus hippoglossus*) have shown that some maternal mRNAs are differentially expressed in eggs of varying quality (Aegerter et al. 2005; Bonnet et al. 2007; Mommens et al. 2010). For instance, high quality eggs were characterized by higher transcript levels of *nucleoplasmin*, *tubulin β*, and *insulin-growth factor 1*, and lower levels of *keratins 8* and *18*, *cathepsin Z* and *prostaglandin synthase 2* in rainbow trout (Aegerter et al. 2005). Bonnet et al. (2007) showed that *prohibitin 2* was negatively associated with development success in rainbow trout. In Atlantic halibut, two maternal genes (*HHC00353* and *HHC01517*) correlated positively with hatching success and one gene (*HHC00255*) correlated negatively with the percentage of normal blastomeres (Mommens et al. 2010).

In recent years, genomic resources for aquaculture species have increased notably with the advent of next-generation sequencing (NGS). Among the different NGS platforms available, the Roche 454 Genome Sequencer has the advantage of

generating longer reads with length between 200-300 base pairs (bp) (Droege and Hill 2008), and it has been used extensively for RNA sequencing (RNA-seq) studies in fish that aimed at: profiling gene expression in larval stages of gilthead seabream (*Sparus aurata*) (Yúfera et al. 2012) and producing a comprehensive transcriptome of fast skeletal muscle (Garcia de la serrana et al. 2012); enhancing the knowledge on transcriptome of Atlantic salmon skin (Micallef al. 2012); generating a reference transcriptome for stress conditions in rainbow trout (Sánchez et al. 2011); providing an ovary transcriptome database representative of all stages of oogenesis and atresia in striped bass (*Morone saxatilis*) (Reading et al. 2012); and characterizing the immune response to bacterial pathogens in the Japanese seabass (*Lateolabrax japonicas*) (Xiang et al. 2010).

In Atlantic cod, factors controlling egg quality are not well understood (Sawanboonchun et al. 2008). Therefore, we undertook research to identify and characterize egg yolk and non-yolk cytoplasmic components, and to investigate their effect on egg quality (Lanes et al. 2012a; Lanes et al. 2012b). We found that Atlantic cod eggs and larvae originating from wild broodstock (WB) were of superior quality than those from farmed broodstock (FB), as also reported previously by Salze et al. (2005). This was found to be associated with differences in chemical composition of egg yolk, including fatty acids, amino acids and minerals (Lanes et al. 2012a). At a molecular level, key genes involved in apoptosis, immunity, oxidative stress, and stress axis were differentially expressed between samples from FB and WB, indicating that quality of eggs and larvae was also associated with some maternal transcripts (Lanes et al. 2012b). In the present study, we have performed RNA-seq using Roche 454 sequencing technology in fertilized eggs obtained from farmed and wild Atlantic cod broodstocks to determine whether and to what extent egg quality is associated with the whole protein-coding transcriptome deposited in eggs during oogenesis, and to have an insight into major transcriptome groups that are important during early development in Atlantic cod. By identifying differentially expressed transcripts between farmed and wild broodstock which are involved in metabolic pathways, we intended to understand their importance during oocyte formation and embryonic development.

Materials and Methods

Broodstock, Sampling and Egg Performance

The experiment was performed at the Research Station of the University of Nordland (UiN), Bodø, Norway. The rearing conditions of WB and FB, as well as their origin and characteristics were described previously by Butts et al. (2011) and Lanes et al. (2012a). WB and FB were allowed to spawn naturally. Five egg batches from each broodstock group were collected in nylon mesh bags using surface collectors at the peak of reproductive season in 2009. Samples of fertilized eggs (2-8 cells stage) were snap-frozen in liquid nitrogen and stored at -80 ºC until RNA extraction.

Fertilization and hatching success were evaluated in four replicates, each comprising of approximately 100 eggs distributed in Petri dishes. Fertilization rate was estimated for 8 h after egg collection under a Stemi SV 11 stereoscopic microscope (Carl Zeiss Vision, Aalen, Germany). Eggs at 2-64 blastomere stage were classified as fertilized. To determine the hatching rate, the Petri dishes were filled with 30 ml of filtered seawater (0.20 µm filters; 33‰ salinity) containing Penicillin-Streptomycin-Neomycin solution (P-S-N; Sigma-Aldrich, St. Louise, Mo, USA; 100 units penicillin, 0.25 mg streptomycin and 0.5 mg neomycin per ml) and kept at 6 ºC in temperature-controlled cell incubator until hatching (approximately 2 weeks after fertilization). Every morning, Petri dishes were checked and dead embryos were removed before taking the counts. Hatching rates are expressed as the percentage of hatched larvae out of the total number of eggs incubated. WB and FB egg batches used in the present study belonged to the major group of samples used in our earlier study (Lanes et al. 2012a).

All procedures adopted in the study were in accordance with the national guidelines detailed in the "Norwegian Regulation on Animal Experimentation" act.

RNA Extraction, Library Construction and Sequencing

RNA extraction from fertilized eggs (2-8 cell stage) was performed using QIAzol (Qiagen, Nydalen, Sweden) as reported by Fernandes et al. (2008). RNA

quantity and quality were checked using Bioanalyzer nanochip (Agilent Technologies, Waldbronn, Germany), and only RNA with a RNA Integrity Number (RIN) higher than 8 was used for downstream applications.

Total RNA samples (~300 μg) were sent to Macrogen Incorporation (Seoul, Korea), for mRNA isolation using the PolyATtract® mRNA Isolation Systems (Promega, Madison, WI, USA) and cDNA synthesis using a cDNA Rapid Library Preparation Kit (Roche Applied Science, Basel, Switzerland). cDNA libraries were individually tagged using multiplex identifier adaptors (MID). The 10 tagged cDNAs were then combined and sequenced on one picotiter plate using the 454 GS-FLX pyrosequencer (Roche Applied Science).

De Novo Assembly and Annotation

Image analysis, signal processing, and data filtering were performed using the Roche GS FLX software (v 2.5.3) with default parameters. Reads were first quality filtered with standard parameters and raw sequencing reads were cleaned from adaptor sequences. Afterwards, sequencing reads were assembled using the 454 GS *De Novo* Assembler (newbler v 2.5.3) software with all parameters set at the default values. Isotigs were used for the annotation and transcriptome analysis.

Sequences were subjected to BLAST (blastx or blastn in the case of unknown proteins) against the NCBI non-redundant protein collection (nr) database and annotated according to the Gene Ontology Consortium (Harris et al. 2004) using Blast2GO software (Götz et al. 2008). Parameters for blastx were: expected value 10^{-3} and HSP Length Cutoff 33. Annotation was done with an E-value Hit Filter of 10^{-6} combined with an Annotation cut-off of 55 and GO weighting of 5. Gene ontology (GO) classes were restricted to those with 50 or more entities (sequence cutoff = 50.0); sequence filter = 50, score alpha = 0.6; and node score filter = 10. To evaluate GO classes for differentially expressed transcripts, parameters were set at the default values. Combined GO graphs for the annotated sequences were created using percentages of 2^{nd} level GO terms.

Mapping Sequencing Reads to the Atlantic Cod Database

Cleaned sequence reads from WB and FB egg batches were mapped separately to the Atlantic cod transcripts (http://codgenome.no/data/ATLCOD1_ANN/; file name: ATLCOD1_ANN_transcripts.fna) using GS Reference Mapper (v 2.5.3) with all parameters set at the default values.

The reads mapped from each run were combined and clustered using TGICL (Pertea et al. 2003). Afterwards, non-redundant sequences were blasted and annotated using Blast2GO software as described above.

Pathway Annotation

In order to gain an overview of pathways networks, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was carried out using the online KEGG Automatic Annotation Server (KAAS; http://www.genome.jp/tools/kaas) for individual sequences. The SBH method, optimized for EST annotation, was used against 16 vertebrates, a fruit fly and a nematode pathway databases.

Comparison of Transcriptome between Farmed and Wild Broodstock Eggs

To compare gene expression in egg batches from FB and WB, and to identify the most frequent transcripts in the two groups, gene expression levels were determined. Expression profiles were characterized by simply counting the number of sequence reads for each putative transcript. Because our libraries were not normalized, read numbers should generally correspond to transcript abundance within each library (Hale et al. 2009). For the identification of differentially expressed transcripts, the analyses were restricted to the transcripts that showed at least five reads in one of the ten libraries. The gene expression level for a given transcript was normalized using the reads per kilobase per million (RPKM) (Mortazavi et al. 2008). Differential expression between the two groups was determined using two-tailed Student's *t*-test. Differentially

expressed genes were identified by fold change (≥2.0) and *P*-value cut-off (<0.05). Hierarchical clustering analyses were performed on transcripts that were differentially expressed between WB and FB samples using the GEPAS 4.0 Suite software (http://gepas.bioinfo.cipf.es/). Data were clustered using the SOTA algorithm (Self Organising Tree Algorithm) and Pearson's correlation coefficient was used as a similarity measurement.

Results

Fertilization and hatching rates were significantly higher in WB group (91±2.4 % and 86±3.7 %) than in FB group (77±4.6 % and 69±4.8 %, respectively).

Sequencing, Assembling and Mapping

Sequencing of non-normalized cDNA from eggs of WB and FB broodstocks (5 batches from each type), yielded average sizes of 36 million bases (Mbp) distributed in 98,687 reads and 44 Mbp distributed in 119,333 reads, respectively. The average read lengths were 361 bp and 367 bp for WB and FB, respectively. The summary of sequencing results is shown in Table 1.

When each library was analyzed individually, at least 46 % of reads were fully assembled in all 10 libraries (Table 1). When all libraries were treated together, 75 % of reads were fully assembled, and 11 % of reads were partially assembled (Table 1). In total, 13,726 isotigs of average length 1,294 bp and minimal length over 100 bp were obtained (Table 1). Of these, 10,313 isotigs (approximately 75 %) showed significant similarity (E value < $1e^{-5}$) to one or more proteins in the GenBank nr database. Only the best hit for each isotig was selected.

Approximately 19 % of reads were fully mapped and 21 % were partially mapped to Atlantic cod transcripts (Table 2). In all libraries, reads mapped to multiple locations (repeats, 2 %) or mapped chimerically (5 %), and too short reads (3 or 4 %) were excluded from further analysis (Table 2). On average, 9,794 and 9,875 contigs

were obtained in WB and FB groups, respectively, of size ≥100 bp (Table 2). In total, the contigs were mapped to 11,340 different Atlantic cod transcripts.

GO and KEGG Analyses

In total, 8,693 isotigs were annotated with specific GO terms (Fig. 1a). In the category biological process, genes involved in cellular (19 %) and metabolic processes (16 %), and biological regulation (11 %), were the most abundant. Regarding the cellular component group, genes involved in cell (41 %), organelle (31 %), macromolecular complex (14 %), and membrane-enclosed lumen (12 %) were the most frequent. In the molecular function set, the most abundant genes were that involved in binding (55 %) and catalytic activity (29 %). KEGG metabolic pathway analysis of the annotated isotigs revealed the involvement of transcripts in amino acid, carbohydrate and energy metabolism (Table 3). Transcripts related to the sub-pathway, signal transduction were the most abundant, while those related to biosynthesis of other secondary metabolites were the least (Table 3). In addition, immune and endocrine related isotigs were predominant among those associated with organismal system (Table 3).

A total of 29,338 non-redundant contig sequences were analyzed with Blast2go and KEGG. Of these, 23,471 contigs were annotated with specific GO terms. In general, the GO terms distribution and KEGG pathways followed the same pattern as those found for isotigs (Fig. 1a; Table 3).

Most Abundant Transcripts in Fertilized Eggs of Atlantic cod

In both Newbler assembly and the GS Reference Mapper approaches, most of the top 20 highly expressed transcripts in WB group were also found in FB group, even though in a different order. Some transcripts including claudin-like protein ZF-A89-like known as claudin D (*cldnd*), cyclin A1 (*ccna1*), cyclin B1 (*ccnb1*), ribonucleotide reductase m2 polypeptide (*rrm2*) and histone H2A type 2-B-like (*h2ab*) were identified using both approaches (Tables 4 and 5).

Transcripts Differentially Expressed in WB and FB Groups

The Newbler assembly approach revealed 427 transcripts differentially expressed between WB and FB groups. However, when the analyses were restricted to transcripts that showed at least five reads in one of samples analyzed, this number was reduced to 238 transcripts, with 155 transcripts highly expressed in WB group and 83 transcripts highly expressed in FB group (Fig. 2a and c). In total, 164 differentially expressed transcripts were annotated with GO terms (Fig. 1b). Regarding KEGG pathways, the transcripts were mainly associated with amino acid and carbohydrate metabolism, translation, transport and catabolism, cell growth and death, signal transduction, and endocrine system (Table 3).

When the reads were mapped against the Atlantic cod transcripts, 158 transcripts were differentially expressed between WB and FB groups. Among these, 118 transcripts were highly expressed in WB, 40 transcripts were highly expressed in FB (Fig. 2b and d), and 140 transcripts were annotated with GO terms (Fig. 1b). KEGG pathways were mostly related to lipid and nucleotide metabolism, in addition to those sub-pathways identified using *De Novo* Assembler as mentioned above (Table 3).

Encoding proteins involved in fructose metabolism, fatty acid metabolism, glycerophospholipid metabolism and oxidative phosphorylation were differentially expressed between the two groups (Fig. 3a-d). For instance, transcripts involved in fructose metabolism including 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1 (*pfkfb1*), and fructose-2,6-bisphosphatase TIGAR B-like or TP53-induced glycolysis and apoptosis regulator B (*tigarb*) were higher (2.1-fold *pfkfb1*, 2.6-fold *tigarb*; $P<0.05$) in WB than in FB group (Fig. 3a). The expression of transcripts related to oxidative phosphorylation such as succinate dehydrogenase (*sdh*), cytochrome c oxidase assembly protein COX11 mitochondrial precursor (*cox11*), and atp synthase-coupling factor mitochondrial precursor (*atp5j*) were higher (5.4-fold *sdh*, 2.0-fold *cox11*, 3.7-fold *atp5j*; $P<0.05$) in samples from WB compared to FB group (Fig. 3b). Regarding transcritps involved in the fatty acid metabolism, long chain fatty acid ligase 4 (*acsl4*) and carnitine o-palmitoyltransferase liver isoform-like (*cpt1*) were more numerous (2.0-

fold *acsl4*, 2.7-fold *cpt1*; $P<0.05$) in samples from WB, whereas propionyl- carboxylase alpha mitochondrial (*pcca*) was 4.0-fold ($P<0.05$) higher in samples from FB (Fig. 3c). Phosphatidylinositol-glycan biosynthesis class f protein (*pigf5*) and group XV phospholipase A2 (*pla2g15*), which are involved in glycerophospholipid metabolism, were more frequent (7.5-fold *pigf5*, 2.7-fold *pla2g15*; $P<0.05$) in samples from WB than FB group (Fig. 3d). In addition, the expression of *cathepsin A* was 2.5-fold higher in samples from FB compared to WB (Fig. 3e).

Discussion

The application of newly available genomic tools has helped to identify thousands of maternal transcripts, which play an important role during early embryonic development of fish (Mathavan et al. 2005; Knoll-Gellida et al. 2006; Vesterlund et al. 2011). For instance, 11,399 different transcripts were described in zebrafish (*Danio rerio*) ovarian follicles (Knoll-Gellida et al. 2006), while 11,187 unique gene transcripts were detected during early zebrafish development from 1-cell stage to 50 % epiboly (Vesterlund et al. 2011). Using the Newbler assembly approach, we identified 13,726 isotigs and annotated 8,693 isotigs. The sequence reads were mapped to 11,340 different Atlantic cod transcripts, and hence we presume that we have investigated a substantial proportion of the total protein-coding gene transcripts present in fertilized eggs. In our study, more than 50 % of reads remained unmapped to Atlantic cod transcripts, although 99 % of sequence reads were mapped to Atlantic cod genome (data not shown). However, it is possible that many transcripts that play important roles during early development are still unknown, as in the case of zebrafish where more than 2,000 putative novel transcribed regions were described recently by Vesterlund et al. (2011). Thus a clear vision of fish transcriptome is far from to be revealed.

One of the methodical steps to be decided prior to RNA-seq studies is the use cDNA normalization. The procedure of cDNA normalization is suitable to characterize as many genes as possible or identify molecular markers (Wall et al. 2009; Yang et al. 2010; Ekblom et al. 2012). Non-normalized libraries are more appropriate to characterize gene expression levels or to investigate differential expression in samples,

since normalization can obscure relative gene expression data (Hale et al. 2009). Recently, Ekblom et al. (2012) found 311 and 328 transcripts that are differentially expressed between two tissues using normalized and non-normalized libraries, respectively. Further, the fold changes of transcripts are more evident in non-normalized library than in the normalized one. Therefore, in the present study, non-normalized libraries were used to investigate the transcripts that are differentially expressed between the samples from FB and WB.

In our study, two parallel approaches were used to identify and characterize the transcripts in fertilized eggs. In general, GS Reference Mapper approach (11,340 unique transcripts) allowed to identify a large number of unique transcripts than Newbler assembly approach (8,693 isotigs annotated), illustrating the importance of a suitable reference database. On the other hand, Newbler assembly approach (238 transcripts) was more appropriate than GS Reference Mapper approach (158 transcripts) to detect the transcripts that are differentially expressed between the eggs of WB and FB groups. The difference in the detection rates may be mainly associated with the difference in the number of reads assembled (86%) and mapped (40%).

Most Abundant Transcripts in Fertilized Eggs of Atlantic cod

Cyclins play an important role in cell cycle control (reviewed by Schafer 1998). In zebrafish, cyclins A2 (*ccna2*), *ccnb1*, B2 (*ccnb2*), and E (*ccne*) were highly abundant in unfertilized eggs and during cleavage stages (Mathavan et al. 2005). Drivenes et al. (2012) found a high expression of *ccnb1* and *ccnb2* during early development of Atlantic cod. In our study too, *ccna1*, *ccnb1*, and *ccnb2* were highly abundant (Table 5). In addition, other important transcripts for cell cycle were abundant in our samples including calmodulin (*cam*), serine threonine-protein kinase plk1-like (*plk1*), cell division cycle protein 20 (*cdc20*), *rrm2*, and ubiquitin (*ubi*). *Cam* acts mainly during G1 and G2/M progression (reviewed by Kahl and Means 2003). *Plk1* plays a central role in G2/M transition phosphorylating *ccnb1* and cell division cycle 25C (*cdc25c*) (Roshak et al. 2000; Yuan et al. 2002). Moreover, *plk1* is crucial in the preservation of genome integrity (Jeong et al. 2010). *Cdc20* is required to activate the anaphase initiation and

exit from mitosis (Kramer et al. 1998), while *rrm2* converts ribonucleoside diphosphates into deoxyribonucleoside diphosphates, and thus provide the precursors for rapid DNA synthesis important for cleavage stages (Standart et al. 1986). Ubiquitin is involved in the degradation of cyclins during the cell cycle, as well as in cytoskeletal regulation, and cell fate determination (reviewed by Bowerman and Kurz 2006).

The zygotic genome activation is a complex process with multiple mechanisms directing the activation and degradation of specific genes over time (Latham and Schultz 2001). It has been shown that zygotic transcription in several species already starts prior to initiation of major zygotic gene expression (reviewed by Andéol 1994; Tadros and Lipshitz 2009). For instance, Mathavan et al. (2005) showed that 125 genes accumulate before the onset of maternal-zygotic transition (MZT) in zebrafish. Vesterlund et al. (2011) also obtained similar results during the early development of zebrafish. Therefore, the high abundance of transcripts such as ataxin-7-like protein 2 (*atxn7l2*), tho complex subunit 4 (*thoc4*), nucleolin-like isoform 3 (*ncl3*), heterogeneous nuclear ribonucleoprotein a1 (*hnrnpa1*), eukaryotic elongation factor 1 alpha (*eef1a*), eukaryotic translation initiation factor 4 gamma 2-like (*eif4g2*), zinc finger protein (*cth1*), and dead (asp-glu-ala-asp) box polypeptide 5 (*ddx5* or *p68*) that are involved in transcriptional and translational control may be needed for the initiation of MZT.

Epigenetic processes, which involve the regulation of gene expression without any actual modification of the DNA sequence, play an important role during embryonic development. DNA methyltransferases, and histones proteins (H2A, H2B, H3 and H4) are important to encode the epigenetic information (reviewed by Meehan 2003). Therefore, the high abundance of transcripts related to *dnmt1*, one of the enzymes involved in methylation of DNA, *h2ab*, and linker histone h1m (*h1m*) in our study might be related to epigenetic processes. *Dnmt1* has also been associated with direct terminal tissue differentiation and knocking down *dnmt1* causes an increase in mortality of zebrafish embryo (Rai et al. 2006).

Chaperonins containing t-complex polypeptide-1 (CCT) are cytosolic molecular chaperone which were implicated in the biogenesis of cytoskeletal proteins by promoting the correct folding of tubulin and actin (reviewed in Frydman 2001). In addition, Dekker et al. (2008) showed that CCT interact with several other processes

including cell cycle, nuclear pore complex, chromatin remodeling, and protein degradation. Therefore, the high abundance of T-complex protein 1 subunit epsilon (*cct5*) found in our study suggest that *cct5* might be involved in several cellular processes during early development of Atlantic cod.

Claudins, the major tight junction transmembrane proteins, are members of the tetraspanin protein superfamily that mediate cellular adhesion and migration (Tsukita and Furuse 2002). In our study, *cldnd* was one of the most abundant transcripts detected in fertilized eggs of Atlantic cod. Similarly in zebrafish, transcripts of *cldnd* and *cldng* were detected among the ten most abundant transcripts from 1-cell stage up to 512-cell stage (Vesterlund et al. 2011). Moreover, a high abundance of *cldnd* has been found in zebrafish and striped bass ovary (Clelland and Kelly 2010; Reading et al. 2012), and in fully-grown follicle of zebrafish (Knoll-Gellida et al. 2006). Recently, Siddiqui et al. (2010) showed that *cldne* is essential for zebrafish epiboly, while *cldnj* is essential for normal ear function and for the formation of the otoliths (Hardison et al. 2005). Taken together, these results indicate that claudin genes might be essential for early development of teleosts. Tumor-associated calcium signal transducer 2 (*tacstd2*), also known as epithelial cell adhesion molecule (*epcam*), was also highly abundant in fertilized eggs of Atlantic cod. Like claudins, *tacstd* participates in cell adhesion, as well as in proliferation, migration and differentiation of cells (reviewed by Trzpis et al. 2008). In zebrafish, *tacstd* (described as zgc:110304) was also among the ten most abundant transcripts at 1- and 16-cell stage (Vesterlund et al. 2011). In zebrafish, transcripts of *tacstd* are required for epithelial morphogenesis and integrity during epiboly and skin development (Slanchev et al. 2009), and in the control of neuromasts migration as well (Villablanca et al. 2006).

In our study, oocyte-specific F-box protein (*fbos*), kinesin-like protein Surhe (*surhe*), and multivesicular body subunit 12a (*mvb12a*) were among the 20 transcripts that were more abundant in fertilized eggs of Atlantic cod. However, the roles of these transcripts during vertebrate development are still unknown. *Fbos* was recently identified through analysis of expressed sequence tags (ESTs) from rainbow trout oocyte cDNA library (Wang et al. 2010). It was shown that *fbos* interact with histone H3.3, adenylate kinase 2, ribosomal protein L36, oocyte protease inhibitor 1, zona pellucida protein 2 and tissue inhibitor of metalloproteinase 2, indicating its essentiality

for early oocyte development. In zebrafish, transcripts of *surhe* that were highly expressed in early embryonic stages were suggested to be involved in dorsalization process as motor molecule (Kim et al. 2008). The multivesicular bodies (MVB) plays an essential role in maintaining proper cell surface protein composition, changing quickly the protein composition of the cell surface during processes such as differentiation and adaptation (Katzmann et al. 2002); particularly, *mvb12* promotes the downregulation of epidermal growth factor (EGF) receptors (Tsunematsu et al. 2010). Considering all these information, *fbos*, *mvb12a*, and *surhe* may be important to control cell proliferation, differentiation and dorsalization in early development of Atlantic cod.

Embryo Survival Success in Relation to Differential Transcriptomes

In rainbow trout and Atlantic halibut, maternal mRNA was differentially expressed in eggs of varying quality (Aegerter et al. 2005; Bonnet et al. 2007; Mommens et al. 2010), and the present study, which performed a massive transcriptome analysis, shows similar findings in Atlantic cod. These results indicate that mRNAs stored in eggs are important to impart full developmental competence of embryos. Therefore, RNA-seq can help in identifying potential biomarkers of egg quality.

The KEGG analysis revealed that a large number of transcripts are involved in carbohydrate metabolism, indicating that carbohydrates are essential for early development of Atlantic cod. Previously, Finn et al. (1995) showed that glycogen was the main metabolic fuel sustaining the development of Atlantic cod up to blastula stage. In gilthead sea bream, whitefish (*Coregonus* spp.), and spotted wolfish (*Anarhichas minor*) also carbohydrate metabolism is essential for early embryonic development and embryos with insufficient carbohydrate metabolism potentially stop their development (Lahnsteiner and Partanello 2003; Lahnsteiner 2005; Desrosiers et al. 2008). In whitefish up to epiboly stage, glycolysis and pentose phosphate pathways are vital, whereas gluconeogenesis is important from epiboly stage to the eye embryo stage (Lahnsteiner 2005). Interestingly, in our study transcripts (*pfkfb1* and *tigarb*) that are involved in the control of glycolysis, pentose phosphate and gluconeogenesis pathways were up-regulated in samples from WB. In mammals, *pfkfb1* is one of four genes

involved in production of fructose 2,6-bisphosphate (Fru-2,6-BP), which is a key regulator of cellular metabolism that activates glycolysis and inhibits gluconeogenesis (reviewed by Bolaños et al. 2010). On the other hand, *tigar* was found to regulate the glycolysis pathway negatively (Bensaad et al. 2006). The expression of *tigar* switches glucose to the pentose phosphate pathway (PPP), along with decreasing reactive oxygen species (ROS) generation and apoptosis by promoting glutathione production. PPP produces more nicotinamide adenine dinucleotide phosphate (NADPH), which is used as reducing agent in anabolic reactions, such as lipid and nucleic acid synthesis (Bensaad et al. 2006). Taken these findings into consideration, *pfkfb1*and *tigar* might be playing a role in the control of carbohydrate metabolism during early development in a marine fish. In addition, the higher expression of *pfkfb1* and *tigarb* in WB embryos compared to FB embryos, suggests an elevated energy production through glycolysis pathway, as well as a high protection against ROS generation and an increase in nucleic acid synthesis, which are essential during cleavage stages.

The roles of fatty acids on egg quality have been extensively studied in fish (reviewed by Sargent et al. 2002). However, information related to maternal mRNAs involved in fatty acids metabolism in fish is still scarce. In the present study, *acls4* and *cpt1*, which play important roles in β-oxidation of fatty acids, were highly expressed in WB embryos, whereas *pcca*, which is involved in synthesis of fatty acids, was highly expressed in FB group. Previously, we showed that the fatty acid profiles were different between WB and FB eggs of Atlantic cod (Lanes et al. 2012a), which could affect the developmental potential. Therefore, in FB embryos which may not be able to meet the fatty acid demands during development, there is an upregulation of *pcca* to support the synthesis of fatty acids.

In Atlantic cod embryos and larvae, the anaerobic metabolism is insignificant (Finn et al. 1995). In lake trout (*Salmo trutta lacustris*), a positive correlation between number of eyed stage embryos and respiration rate, malate dehydrogenase activity, and the ratio of NADH to NAD were found (Lahnsteiner et al. 1999). In cyprinid fishes, respiration rate also correlated positively with the fertilization rate (Lahnsteiner et al. 2001). In the current study, transcripts involved in oxidative phosphorylation in different steps of respiratory chain that drives adenosine triphosphate (ATP) synthesis,

including *sdh*, *cox11*, and *atp5j* were expressed at higher levels in WB than FB eggs, which suggest better potential of WB embryos for aerobic metabolism resulting in higher survival success.

Cell membranes of eukaryote are highly complex structures consisting of hundreds of different lipid molecules and proteins (reviewed by Hermansson et al. 2011). In Atlantic cod, Salze et al. (2005) investigated the different classes of phospholipids in eggs from three different broodstock groups and identified differences in levels of phosphatidylcholine, phosphatidylinositol, and the phosphatic acid/phosphatidylglycerol/cardiolipin ratio among them. In our study, transcripts involved in synthesis (*pigf5*) and degradation (*pla2g15*) of glycerophospholipids were upregulated in WB compared to FB, which may be related to the differences in composition of phospholipids in eggs as observed by Salze et al. (2005).

Cathepsins are important lysosomal enzymes involved in the metabolism of yolk proteins in the oocyte, fertilized egg and larval yolk sac. This process provides free amino acids and is essential for the early development of the embryo (Carnevali et al. 1999; Carnevali et al. 2001; Fabra and Cerdà 2004). In sea bass, high levels of cathepsins B, D, and L were detected in fertilized eggs, whereas no cathepsin A, C, and E activity was found (Carnevali et al. 2001). In rainbow trout, mRNA levels for *cathepsin D* and *L* were high at fertilization, but the expression of *cathepsin B* was very low (Kwon et al. 2001). These studies show that *cathepsin D* and *L* play an important role in yolk mobilization during embryogenesis in teleosts, but the role of *cathepsin B* is not elucidated. We have found high expression of *cathepsins B*, *D* and mainly *L* in early embryos of Atlantic cod. In addition, the expression of *cathepsins A*, *C*, *F*, *S*, and *Z* was profiled in our study. The presence of several cathepsins in fertilized eggs of Atlantic cod suggests that yolk mobilization during embryogenesis is a complex process. In *Fundulus heteroclitus*, Fabra and Cerdà (2004) showed the involvement of several *cathepsins* (*B*, *C*, *F*, *H*, *K*, *L*, *S*, and *Z*) in the processing of yolk proteins during oocytes development. Cathepsins have also been associated with egg quality. In sea bass, cathepsin D activity was significantly higher in sinking eggs, whereas cathepsin L activity was higher in floating eggs (Carnevali et al. 2001). In rainbow trout, *cathepsin*

Z mRNA was more abundant in low quality eggs (Aegerter et al. 2005). In the present study, the expression of *cathepsin A* was higher in FB eggs than in WB group.

Conclusions

In summary, 11,340 mRNA transcripts were detected in fertilized eggs of Atlantic cod which should represent a substantial portion of the total mRNA incorporated during oocyte development. The study revealed that the most abundant transcripts are involved in important cell processes including cell cycle control (*ccna1*, *ccnb1*, *ccnb1*, *cdc20*, *cam*, *plk1*, *rrm2*, and *ubi*), transcriptional and translational control (*atxn7l2*, *thoc4*, *ncl3*, *hnrnpa1*, *eef1a*, *eif4g2*, *cth1*, and *ddx5*), epigenetic control (*dnmt1*, *h2ab*, and *h1m*), and cellular adhesion and migration (*cldnd* and *tacstd2*). Besides, we identified several transcripts that are differentially expressed between eggs from farmed and wild broodstocks. Interestingly, transcripts involved in glycolytic pathway (*pfkfb1*), β-oxidation (*acls4* and *cpt1*) and oxidative phosphorylation (*sdh*, *cox11*, and *atp5j*) were more highly expressed in WB eggs and may be associated with egg quality. It is known that broodstock nutrition plays one of the most important roles in egg quality, affecting directly the chemical constituents of egg yolk and possibly the regulation of nutrient metabolism at a molecular level during oogenesis. Maternal mRNAs involved in fructose metabolism, fatty acid metabolism, glycerophospholid metabolism and oxidative phosphorylation were differentially represented in oocytes originating from the two broodstock groups, showing potential as biomarkers of egg quality in teleosts. Taken together, our findings contribute to the hypothesis that maternal mRNAs affect the egg quality and, consequently, the early development of fish.

Acknowledgments

The study was funded by the Research Council of Norway, project 182653/V10. The authors gratefully thank Bjørnar Eggen for technical assistance during the egg collection and incubation. The authors are also thankful to Morten Krogstad and Marianne Fossum for their help in procuring wild fish for the study. We also thank the

Macrogen Incorporation (Seoul, Korea) for providing the 454 pyrosequencing service. C.F.C. Lanes' PhD scholarship is partially supported by the Research Council of Norway and the University of Nordland.

References

Aegerter S, Jalabert B, Bobe J (2005) Large scale real-time PCR analysis of mRNA abundance in rainbow trout eggs in relationship with egg quality and post-ovulatory ageing. Mol Reprod Dev 72:377-385

Andéol Y (1994) Early transcription in different animal species: implication for transition from maternal to zygotic control in development. Dev Genes Evol 204:3-10

Bensaad K, Tsuruta A, Selak MA, Vidal MNC, Nakano K, Bartrons R, Gottlieb E, Vousden KH (2006) TIGAR, a p53-inducible regulator of glycolysis and apoptosis. Cell 126:107-120

Bobe J, Labbé C (2010) Egg and sperm quality in fish. Gen Comp Endocrinol 165:535-548

Bolaños JP, Almeida A, Moncada S (2010) Glycolysis: a bioenergetic or a survival pathway? Trends Biochem Sci 35:145-149

Bonnet E, Fostier A, Bobe J (2007) Microarray-based analysis of fish egg quality after natural or controlled ovulation. BMC Genomics 8:55

Bowerman B, Kurz T (2006) Degrade to create: developmental requirements for ubiquitin-mediated proteolysis during early *C. elegans* embryogenesis. Development 133:773-784

Brooks S, Tyler CR, Sumpter JP (1997) Egg quality in fish: what makes a good egg? Rev Fish Biol Fish 7:387-416

Butts IAE, Trippel EA, Ciereszko A, Soler C, Słowińska M, Alavi SMH, Litvak MK, Babiak I (2011) Seminal plasma biochemistry and spermatozoa characteristics of Atlantic cod (*Gadus morhua*) of wild and cultivated origin. Comp Biochem Physiol A 159:16-24

Carnevali O, Carletta R, Cambi A, Vita A, Bromage N (1999) Yolk formation and degradation during oocyte maturation in seabream Sparus aurata: involvement of two lysosomal proteinases. Biol Reprod 60:140-146

Carnevali O, Mosconi G, Cambi A, Ridolfi S, Zanuy S, Polzonetti-Magni AM (2001) Changes of lysosomal enzyme activities in sea bass (*Dicentrarchus labrax*) eggs and developing embryos. Aquaculture 202:249-256

Clelland ES, Kelly SP (2010) Tight junction proteins in zebrafish ovarian follicles: stage specific mRNA abundance and response to 17beta-estradiol, human chorionic gonadotropin, and maturation inducing hormone. Gen Comp Endocrinol 168:388-400

Dekker C, Stirling PC, McCormack EA, Filmore H, Paul A, Brost RL, Costanzo M, Boone C, Leroux MR, Willison KR (2008) The interaction network of the chaperonin CCT. EMBO J 27:1827-1839

Desrosiers V, Le François NR, Tveiten H, Andreassen I, Blier PU (2008) Ontogenesis of catabolic and energy metabolism capacities during the embryonic development of spotted wolffish (*Anarhichas minor*). Comp Biochem Physiol B 150:200-206

Drivenes Ø, Taranger GL, Edvardsen RB (2012) Gene expression profiling of Atlantic cod (*Gadus morhua*) embryogenesis using microarray. Mar Biotechnol 14:167-176

Droege M, Hill B (2008) The genome sequencer FLXTM system-longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol 136:3-10

Ekblom R, Slate J, Horsburgh GJ, Birkhead T, Burke T (2012) Comparison between normalised and unnormalised 454-sequencing libraries for small-scale RNA-Seq studies. Comp Funct Genomics 2012:1-8

Fabra M, Cerdà J (2004) Ovarian cysteine proteinases in the teleost *Fundulus heteroclitus*: molecular cloning and gene expression during vitellogenesis and oocyte maturation. Mol Reprod Dev 67:282-294

Faulk CK, Holt GJ (2008) Biochemical composition and quality of captive-spawned cobia *Rachycentron canadum* eggs. Aquaculture 279:70-76

Fernandes JMO, Mommens M, Hagen Ø, Babiak I, Solberg C (2008) Selection of suitable reference genes for real-time PCR studies of Atlantic halibut. Comp Biochem Physiol B 150:23-32

Finn RN, Fyhn HJ, Evjen MS (1995) Physiological energetics of developing embryos and yolk-sac larvae of Atlantic cod (*Gadus morhua*). I. respiration and nitrogen metabolism. Mar Biol 124:355-369

Frydman J (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem 70:603-647

Furuita H, Ohta H, Unuma T, Tanaka H, Kagawa H, Suzuki N, Yamamoto T (2003) Biochemical composition of eggs in relation to egg quality in Japanese eel, *Anguilla japonica*. Fish Physiol Biochem 29:37-46

Garcia de la serrana D, Estévez A, Andree K, Johnston IA (2012) Fast skeletal muscle transcriptome of the gilthead sea bream (*Sparus aurata*) determined by next generation sequencing. BMC Genomics 13:181

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420-3435

Hale MC, McCormick CR, Jackson JR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. BMC Genomics 10:203

Hardison AL, Lichten L, Banerjee-Basu S, Becker TS, Burguess SM (2005) The zebrafish gene claudinj is essential for normal ear function and important for the formation of the otoliths. Mech Dev 122:949-958

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T,

White R (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258-D261

Hermansson M, Hokynar K, Somerharju P (2011) Mechanisms of glycerophospholipid homeostasis in mammalian cells. Prog Lipid Res 50:240-257

Jeong K, Jeong JY, Lee HO, Choi E, Lee H (2010) Inhibition of Plk1 induces mitotic infidelity and embryonic growth defects in developing zebrafish embryos. Dev Biol 345:34-48

Kahl CR, Means AR (2003) Regulation of cell cycle progression by calcium/calmodulin-dependent pathways. Endocr Rev 24:719-736

Katzmann DJ, Odorizzi G, Emr SD (2002) Receptor downregulation and multivesicular-body sorting. Nat Rev Mol Cell Biol 3:893-905

Kim E-J, Ro H, Huh T-L, Lee CJ, Choi J, Rhee M (2008) A novel kinesin-like protein, Surhe is associated with dorsalization in the zebrafish embryos. Animal Cells Syst 12:219-230

Knoll-Gellida A, André M, Gattegno T, Forgue J, Admon A, Babin PJ (2006) Molecular phenotype of zebrafish ovarian follicle by serial analysis of gene expression and proteomic profiling, and comparison with the transcriptomes of other animals. BMC Genomics 7:46

Kramer ER, Gieffers C, Hölzl G, Hengstschläger M, Peters JM (1998) Activation of the human anaphase-promoting complex by proteins of the CDC20/Fizzy family. Curr Biol 8:1207-1210

Kwon JY, Prat F, Randall C, Tyler CR (2001) Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout (*Oncorhynchus mykiss*). Biol Reprod 65:1701-1709

Lahnsteiner F (2005) Carbohydrate metabolism of eggs of the whitefish, *Coregonus* spp. during embryogenesis and its relationship with egg quality. Comp Biochem Physiol B 142:46-55

Lahnsteiner F, Patarnello P (2003) Investigations on the metabolism of viable and nonviable gilthead sea bream (*Sparus aurata*) eggs. Aquaculture 223:159-174

Lahnsteiner F, Weismann T, Patzner RA (1999) Physiological and biochemical parameters for egg quality determination in lake trou, *Salmo trutta lacustris*. Fish Physiol Biochem 20:375-388

Lahnsteiner F, Urbanyi B, Horvath A, Weismann T (2001) Bio-markers for egg quality determination in cyprinid fish. Aquaculture 195:331-352

Lanes CFC, Bizuayehu TT, Bolla S, Martins C, Fernandes JMO, Bianchini A, Kiron V, Babiak I (2012a) Biochemical composition and performance of Atlantic cod (*Gadus morhua* L.) eggs and larvae obtained from farmed and wild broodstocks. Aquaculture 324-325:267-275

Lanes CFC, Fernandes JMO, Kiron V, Babiak I (2012b) Profiling of key apoptotic, stress, and immune-related transcripts during embryonic and postembryonic development of Atlantic cod (*Gadus morhua* L.). Theriogenology 78:1583–1596

Latham KE, Schultz RM (2001) Embryonic genome activation. Front Biosci 6:D748-D759

Lubzens E, Young G, Bobe J, Cerdà J (2010) Oogenesis in teleosts: How fish eggs are formed? Gen Comp Endocrinol 165:367-389

Mathavan S, Lee SGP, Mak A, Miller LD, Murthy KRK, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan Y, Korzh V, Gong Z, Liu ET, Lufkin T (2005) Transcriptome analysis of zebrafish embryogenesis using microarrays. PLoS Genet 1:260-276

Meehan RR (2003) DNA methylation in animal development. Semin Cell Dev Biol 14: 53-65

Micallef G, Bickerdike R, Reiff C, Fernandes JMO, Bowman AS, Martin SAM (2012) Exploring the transcriptome of Atlantic salmon (*Salmo salar*) skin, a major defense organ. Mar Biotechnol 14:559-569

Mommens M, Fernandes JMO, Bizuayehu TT, Bolla SL, Johnston IA, Babiak I (2010) Maternal gene expression in Atlantic halibut (*Hippoglossus hippoglossus* L.) and its relation to egg quality. BMC Res Notes 3:138

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621-628

Pelegri F (2003) Maternal factors in zebrafish development. Dev Dyn 228:535-554

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651-652

Rai K, Nadauld LD, Chidester S, Manos EJ, James SR, Karpf AR, Cairns BR, Jones DA (2006) Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. Mol Cell Biol 26:7077-7085

Reading BJ, Chapman RW, Schaff JE, Scholl EH, Opperman CH, Sullivan CV (2012) An ovary transcriptome for all maturational stages of the striped bass (*Morone saxatilis*), a highly advanced perciform fish. BMC Res Notes 5:111

Roshak AK, Capper EA, Imburgia C, Fornwald J, Scott G, Marshall LA (2000) The human polo-like kinase, PLK, regulates cdc2/cyclin B through phosphorylation and activation of the cdc25C phosphate. Cell Signal 12:405-411

Salze G, Tocher DR, Roy WJ, Robertson DA (2005) Egg quality determinants in cod (*Gadus morhua* L.): egg performance and lipids in eggs from farmed and wild broodstock. Aquac Res 36:1488-1499

Sánchez CC, Weber GM, Gao G, Cleveland BM, Yao J, Rexroad CE (2011) Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. BMC Genomics 12:626

Sargent JR, Tocher DR, Bell JG (2002) In: Halver JH (ed) Fish nutrition, Academic Press, San Diego, California

Sawanboonchun J, Roy WJ, Robertson DA, Bell JG (2008) The impact of dietary supplementation with astaxanthin on egg quality in Atlantic cod broodstock. Aquaculture 283:97-101

Schafer KA (1998) The cell cycle: a review. Vet Pathol 35:461-478

Shen-Orr SS, Pilpel Y, Hunter CP (2010) Composition and regulation of maternal and zygotic transcriptomes reflects species-specific reproductive mode. Genome Biol 11:R58

Siddiqui M, Sheikh H, Tran C, Bruce AEE (2010) The tight junction component claudin E is required for zebrafish epiboly. Dev Dyn 239:715-722

Slanchev K, Carney TJ, Stemmler MP, Koschorz B, Amsterdam A, Schwarz H, Hammerschmidt M (2009) The epithelial cell adhesion molecule EpCAM is required for epithelial morphogenesis and integrity during zebrafish epiboly and skin development. PLoS Genet 5:e1000563

Standart N, Hunt T, Ruderman JV (1986) Differential accumulation of ribonucleotide reductase subunits in clam oocytes: the large subunit is stored as a polypeptide, the small subunit as unstranslated mRNA. J Cell Biol 103:2129-2136

Tadros W, Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts. Development 136:3033-3042

Trzpis M, Bremer E, McLaughlin PM, de Leij LF, Harmsen MC (2008) EpCAM in morphogenesis. Front Biosci 13:5050-5055

Tsukita S, Furuse M (2002) Claudin-based barrier in simple and stratified cellular sheets. Curr Opin Cell Biol 14:531-536

Tsunematsu T, Yamauchi E, Shibata H, Maki M, Ohta T, Konishi H (2010) Distinct functions of human MVB12A and MVB12B in the ESCRIT-I dependent on their posttranslational modifications. Biochem Biophys Res Commun 399:232-237

Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J (2011) The zebrafish transcriptome during early development. BMC Dev Biol 11:30

Villablanca EJ, Renucci A, Sapède D, Lec V, Soubiran F, Sandoval PC, Dambly-Chaudière C, Ghysen A, Allende ML (2006) Control of cell migration in the zebrafish lateral line: implication of the gene "tumour-associated calcium signal transducer," tacstd. Dev Dyn 235:1578-1588

Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho LP, Hu Y, Carlson JE, Ma H, Schuster SC, Soltis DE, Soltis PS, Altman N, dePamphilis CW (2009) Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics 10:347

Wang L, Tripurani S, Wanna W, Rexroad CE, Yao J (2010) Fbos, a novel oocyte-specific protein, interacts with proteins important for oocyte development in rainbow trout (*Oncorhynchus mykiss*). Biol Reprod 83:410

Xiang L, He D, Dong W, Zhang Y, Shao J (2010) Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. BMC Genomics 11:472

Yang P, Li X, Shipp MJ, Shockey JM, Cahoon EB (2010) Mining the bitter melon (Momordica charantia l.) seed transcriptome by 454 analysis of non-normalized and normalized cDNA populations for conjugated fatty acid metabolism-related genes. BMC Plant Biol 10:250

Yuan J, Eckerdt F, Bereiter-Hahn J, Kurunci-Csacsko E, Kaufmann M, Strebhardt K (2002) Cooperative phosphorylation including the activity of polo-like kinase 1 regulates the subcellular localization of cyclin B1. Oncogene 21:8282-8292

Yúfera M, Halm S, Beltran S, Fusté B, Planas JV, Martínez-Rodríguez G (2012) Transcriptomic characterization of the larval stage in gilthead seabream (*Sparus aurata*) by 454 pyrosequencing. Mar Biotechnol 14:423-435

Table 1 Summary of sequencing results and the assembly output for five batches of eggs from wild and five batches of eggs from farmed broodstocks of Atlantic cod

	Wild broodstock					Farmed broodstock					All libraries
	1W	**2W**	**3W**	**4W**	**5W**	**1F**	**2F**	**3F**	**4F**	**5F**	
Reads											
Read count	117,889	66,808	90,491	95,723	122,524	94,205	159,770	105,581	129,303	107,805	1,089,278
Total bases (Mbp)	42.8	23.6	32.8	34.2	44.7	34.8	58.3	38.4	47.6	39.7	395.6
Average read length (bp)	364	353	363	358	365	370	365	364	368	369	364
Assembly											
Assembled	65,774	30,606	49,523	49,598	71,834	52,807	101,272	60,571	77,504	63,406	821,435
(%)	(56)	(46)	(55)	(52)	(59)	(56)	(63)	(57)	(60)	(59)	(75)
Partial	8,976	5,307	6,183	7,830	8,750	5,874	11,761	7,766	8,909	7,244	114,430
(%)	(8)	(8)	(7)	(8)	(7)	(6)	(7)	(7)	(7)	(7)	(11)
Singleton	34,885	25,837	29,049	31,156	34,083	29,302	35,747	29,931	34,500	29,926	56,325
(%)	(30)	(39)	(32)	(33)	(28)	(31)	(22)	(28)	(27)	(28)	(5)
Isotigs											
Number of isotigs	3,180	1,542	2,222	2,532	3,339	2,155	3,885	2,668	3,392	2,506	13,726
Average contig count	1.23	1.27	1.21	1.30	1.21	1.23	1.26	1.34	1.24	1.33	1.48
Largest contig count	10	8	6	11	8	17	11	20	13	20	14
Number with one contig	2,625	1,248	1,859	2,014	2,796	1,796	3,155	2,147	2,802	2,077	10,182
Average isotig size	881	777	836	824	905	828	890	890	900	853	1,294

Table 2 Summary of mapping results for five batches of eggs from wild and five batches of eggs from farmed broodstocks of Atlantic cod

	Wild broodstock						Farmed broodstock					
	1W	**2W**	**3W**	**4W**	**5W**	**Average**	**1F**	**2F**	**3F**	**4F**	**5F**	**Average**
Read count	117,889	66,808	90,491	95,723	122,524	98,687	94,205	159,770	105,581	129,303	107,805	119,333
Fully mapped	23,405	12,069	16,441	18,166	25,920	19,200	15,741	26,532	18,962	23,937	17,930	20,620
(%)	(17)	(17)	(18)	(19)	(17)	(17)	(20)	(18)	(18)	(19)	(21)	(19)
Partially mapped	26,800	14,404	19,106	20,969	28,177	21,891	18,582	31,550	22,279	28,227	21,192	24,366
(%)	(20)	(20)	(21)	(22)	(20)	(20)	(23)	(22)	(21)	(22)	(23)	(22)
Unmapped	55,056	32,989	46,144	45,892	55,271	47,070	50,987	85,891	53,892	64,030	58,588	62,678
(%)	(54)	(54)	(51)	(50)	(54)	(53)	(47)	(49)	(51)	(48)	(45)	(48)
Repeats	2,045	1,127	1,439	1,692	2,270	1,715	1,477	2,571	1,676	2,100	1,738	1,912
(%)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)	(2)
Chimeric	6,398	3,651	4,454	5,459	6,957	5,384	4,307	7,707	5,189	6,772	4,960	5,787
(%)	(5)	(5)	(5)	(5)	(5)	(5)	(5)	(5)	(5)	(6)	(6)	(5)
Too short to map	4,185	2,568	2,907	3,545	3,929	3,358	3,111	5,519	3,583	4,237	3,397	3,934
(%)	(3)	(3)	(3)	(3)	(3)	(3)	(4)	(4)	(3)	(4)	(3)	(4)
Contigs ≥ 100 bp	10,790	8,366	9,039	9,805	10,970	9,794	9,068	11,075	9,529	10,424	9,282	9,875

Table 3 KEGG pathways in the transcriptome of early embryos from Atlantic cod

KEGG Pathways	Sub-Pathways	Number of sequences using *De Novo* Assembler (pooled batches of eggs, n = 10)	Number of sequences using GS Reference Mapper (pooled batches of eggs, n = 10)	Number of differentially expressed transcripts between wild (n=5) and farmed (n=5) groups using *De Novo* Assembler	Number of differentially expressed transcripts between wild (n=5) and farmed (n=5) groups using GS Reference Mapper
Metabolism	**Total**	**1197**	**3460**	**27**	**18**
	Amino acid metabolism	218	625	7	2
	Carbohydrate metabolism	191	676	7	1
	Energy metabolism	179	350	4	0
	Lipid metabolism	167	556	4	6
	Glycan biosynthesis and metabolism	158	450	2	2
	Nucleotide metabolism	119	341	2	4
	Metabolism of cofactors and vitamins	88	242	1	2

	Xenobiotics biodegradation and metabolism	43	117	0	1
	Metabolism of terpenoids and polyketides	27	72	0	0
	Biosynthesis of other secondary metabolites	7	31	0	0
Genetic Information Processing	**Total**	**1249**	**2903**	**18**	**15**
	Folding, sorting, and degradation	466	1111	3	3
	Translation	392	949	10	5
	Transcription	251	393	4	5
	Replication and repair	140	450	1	2
Cellular processes	**Total**	**1243**	**3478**	**19**	**24**
	Transport and catabolism	424	1190	7	17
	Cell communication	329	981	2	2
	Cell growth and death	327	802	8	4
	Cell motility	163	505	2	1
Environmental information processing	**Total**	**689**	**2037**	**9**	**12**
	Signal Transduction	551	1572	6	9

	Signaling molecules and interaction	109	349	3	3
	Membrane transport	29	116	0	0
Organismal system	**Total**	**1417**	**4465**	**16**	**17**
	Immune system	332	917	3	4
	Endocrine system	270	918	7	7
	Nervous system	236	792	3	1
	Circulatory system	139	278	0	1
	Development	136	526	2	1
	Digestive system	129	497	0	0
	Excretory system	107	372	1	2
	Environmental adaptation	41	75	0	1
	Sensory system	27	90	0	0

Table 4 Most abundant 20 transcripts in samples of eggs from wild broodstock of Atlantic cod using *De Novo* Assembler (n=5)

Reference	Rank in farmed group	Length (bp)	Mean of reads after normalization (wild group)	Mean of reads after normalization (farmed group)	Fold change (wild/ farmed)	BLAST2GO Annotation	Accession number	Species	E-value
contig01062	1	591	5,552	6,331	0.88	28S ribosomal RNA	AY141485	*Gadus morhua*	5e-151
isotig00171	2	593	3,347	4,603	0.73	Histone H2A type 2-B-like	XP_003450851	*Oreochromis niloticus*	6e-61
isotig00246	5	932	1,697	1,988	0.85	Claudin-like protein zf-a89	XP_003447029	*Oreochromis niloticus*	8e-81
isotig00170	3	1,278	1,607	2,185	0.74	Late histone	XP_003450851	*Oreochromis niloticus*	6e-57
contig01877	10	508	1,527	1,827	0.84	Multivesicular body subunit 12a	NP_956778	*Danio rerio*	4e-6
contig00071	9	550	1,459	1,898	0.77	ITS1, 5.8S rRNA and ITS2	AJ969261	*Gadus morhua*	0.0
isotig05810	16	1,611	1,424	1,226	1.16	Cyclin A1	AAD49424	*Carassius gibelio*	5e-104
contig00458	14	1,466	1,389	1,400	0.99	Cyclin B1	ACI68598	*Salmo salar*	1e-161
contig03596	13	1,100	1,277	1,432	0.89	16S ribosomal RNA	GU931786	*Gadus morhua*	0.0
isotig00687	22	909	1,146	1,046	1.09	Tho complex subunit 4	XP_003454070	*Oreochromis niloticus*	1e-63
isotig01335	40	114	1,143	923	1.24	No hit	-	-	-
contig00402	17	570	1,118	1,208	0.93	Cytochrome oxidase subunit 1	ACS44839	*Sebastes caurinus*	8e-93
contig00066	24	851	1,104	1,034	1.07	Ribonucleotide reductase m2 polypeptide	NP_001133495	*Salmo salar*	3e-126

isotig00166	26	191	999	1,020	0.98	No hit	-	-	-
isotig00147	43	876	998	837	1.19	Ubiquitin	AEL23099	*Cherax quadricarinatus*	3e-44
isotig03638	19	1,182	977	1,023	0.95	Calmodulin	NP_001734	*Homo sapiens*	4e-92
isotig00532	36	1,112	976	896	1.09	Linker histone h1m	BAH08643	*Pagrus major*	1e-31
isotig01673	30	331	958	925	1.04	No hit	-	-	-
isotig03557	34	1,586	952	906	1.05	Zinc finger protein cth1	NP_571014	*Danio rerio*	2e-91
isotig03672	22	548	938	1,012	0.93	Ferritin, middle subunit	ACI66713	*Salmo salar*	1e-56

Table 5 Most abundant 20 sequences mapped in samples of eggs from wild broodstock of Atlantic cod using GS Reference Mapper (n=5)

Reference	Rank in farmed group	Length (bp)	Mean of reads after normalization (wild group)	Mean of reads after normalization (farmed group)	Fold change (wild/ farmed)	BLAST2GO Annotation	Accession number	Species	E-value
ENSGAUG00000009743	1	390	6,101	5,712	1.07	Ataxin-7-like protein 2	CBN80945	*D. labrax*	2e-06
ENSGAUG00000003512	2	1,063	3,128	2,913	1.07	Ribonucleoside-diphosphate reductase subunit M2	NP_001133495	*S. salar*	0.0
ENSGAUG00000006153	4	1,167	2,354	2,034	1.16	Cyclin A1	AAD49424	*C.gibelio*	2e-110
ENSGAUG00000003491	5	4,591	2,242	2,015	1.11	Dna (cytosine-5)-methyltransferase 1	XP_003442135	*O. niloticus*	0.0
ENSGAUG00000010211	3	1,032	2,153	2,151	1.00	Cyclin-B1	ACI66090	*S. salar*	7e-168
ENSGAUG00000020248	6	633	1,670	1,919	0.87	Claudin-like protein ZF-A89-like	XP_003447029	*O. niloticus*	5e-95
ENSGAUG00000017959	8	698	1,617	1,686	0.96	Serine threonine-protein kinase plk1-like	XP_003457812	*O. niloticus*	7e-130
ENSGAUG00000014993	7	964	1,589	1,718	0.93	Cyclin-B2-like	XP_003452642	*O. niloticus*	4e-166
ENSGAUG00000007360	9	722	1,444	1,359	1.06	Hypothetical protein LOC100700871	XP_003458454	*O. niloticus*	4e-92
ENSGAUG00000001341	10	1,538	1,387	1,347	1.03	Nucleolin-like isoform 3	XP_003449053	*O. niloticus*	2e-159

ENSGAUG0 0000012005	12	1,388	1,141	1,244	0.92	Eukaryotic elongation factor 1 alpha	ABD62881	*G. morhua*	0.0
ENSGAUG0 0000019427	20	1,626	1,107	888	1.25	T-complex protein 1 subunit epsilon	NP_0011673 10	*S. salar*	0.0
ENSGAUG0 0000000441	23	1,506	1,073	816	1.32	Oocyte-specific F-box protein	NP_0011841 37	*O. mykiss*	0.0
ENSGAUG0 0000010469	16	1,491	1,053	1,015	1.04	Cell division cycle protein 20 homolog	XP_0034381 88	*O. niloticus*	0.0
ENSGAUG0 0000017911	13	508	1,028	1,235	0.83	Kinesin-like protein Surhe	CBN81778	*D. labrax*	3e-66
ENSGAUG0 0000006927	14	913	1,017	1,082	0.94	Tumor-associated calcium signal transducer 2 precursor	NP_0011349 32	*S. salar*	4e-126
ENSGAUG0 0000019969	11	420	984	1,334	0.74	Histone H2A type 2-B-like	XP_0034508 51	*O. niloticus*	2e-63
ENSGAUG0 0000017950	32	569	950	732	1.30	Heterogeneous nuclear ribonucleoprotei n a1	ABS70764	*H. burtoni*	1e-108
ENSGAUG0 0000002082	25	2,093	945	799	1.18	Eukaryotic translation initiation factor 4 gamma 2-like	XP_0034492 45	*O. niloticus*	0.0
ENSGAUG0 0000010758	17	1,894	934	965	0.97	Dead (asp-glu-ala-asp) box polypeptide 5	AAH67585	*D. rerio*	0.0

Figure legends

Fig. 1 Functional categorization of transcriptome of fertilized eggs of Atlantic cod using *De Novo* assembler and GS Reference Mapper. Most frequently represented gene ontology (GO) categories are shown within biological process, cellular component and molecular function; (a), distribution of all annotated transcripts and (b), distribution of differentially expressed transcripts between egg batches from wild (n=5) and farmed (n=5) broodstocks (fold change ≥2.0, $P<0.05$)

Fig. 2 Hierarchical clustering analysis of transcripts differentially expressed ($P<0.05$; fold change ≥2.0) between egg batches from wild (n=5) and farmed (n=5) broodstocks of Atlantic cod. a) and b), transcripts with higher expression in wild broodstock eggs using *De novo* assembler and GS Reference Mapper, respectively; c) and d), transcripts with higher expression in farmed broodstock eggs using *De novo* assembler and GS Reference Mapper, respectively. Red and blue represent higher and lower expression levels, respectively. Repeated transcripts in different clusters are indicated in blue and red font, while similar transcripts in the same cluster are underlined. For each transcript, its expression was normalized against the highest expression value

Fig. 3 Expression of transcripts involved in (a) fructose metabolism, (b) oxidative phosphorylation, (c) fatty acid metabolism, and (d) glycerophospholipid metabolism, as well as (e) the different types of cathepsins present in fertilized eggs from wild (n = 5) and farmed broodstock (n = 5) of Atlantic cod. Expression levels were quantified by 454 pyrosequencing and normalized using the RPKM. Data are expressed as mean ± SEM. Asterisks (*) indicate significant differences ($P<0.05$; fold change ≥2) between wild and farmed broodstocks

FIGURE 1

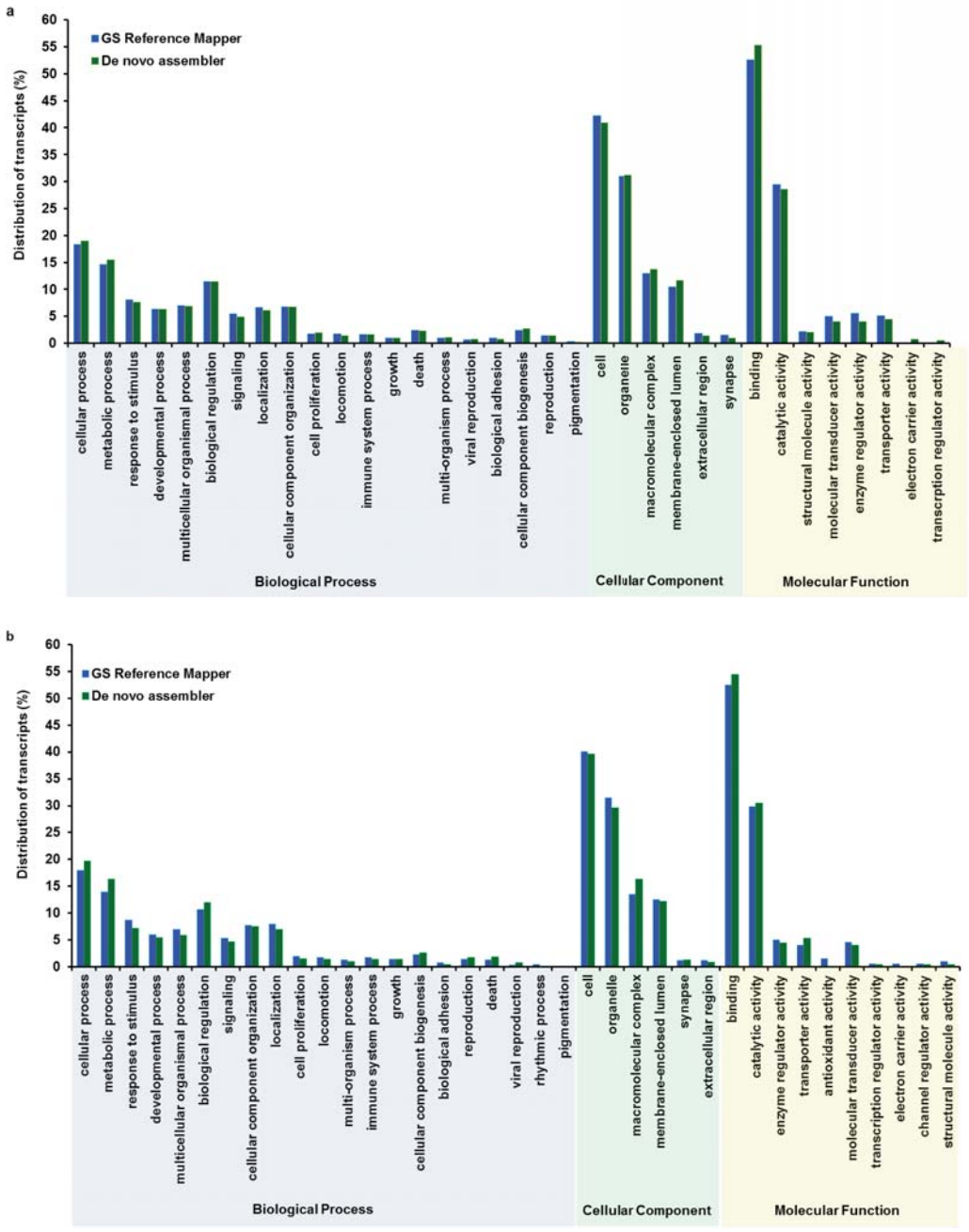

a
GS Reference Mapper
De novo assembler
Distribution of transcripts (%)
cellular process
metabolic process
response to stimulus
developmental process
multicellular organismal process
biological regulation
signaling
localization
cellular component organization
cell proliferation
locomotion
immune system process
growth
death
multi-organism process
viral reproduction
biological adhesion
cellular component biogenesis
reproduction
pigmentation
cell
organelle
macromolecular complex
membrane-enclosed lumen
extracellular region
synapse
binding
catalytic activity
structural molecule activity
molecular transducer activity
enzyme regulator activity
transporter activity
electron carrier activity
transcription regulator activity
Biological Process
Cellular Component
Molecular Function
b
GS Reference Mapper
De novo assembler
Distribution of transcripts (%)
cellular process
metabolic process
response to stimulus
developmental process
multicellular organismal process
biological regulation
signaling
cellular component organization
localization
cell proliferation
locomotion
multi-organism process
immune system process
growth
cellular component biogenesis
biological adhesion
reproduction
death
viral reproduction
rhythmic process
pigmentation
cell
organelle
macromolecular complex
membrane-enclosed lumen
synapse
extracellular region
binding
catalytic activity
enzyme regulator activity
transporter activity
antioxidant activity
molecular transducer activity
transcription regulator activity
electron carrier activity
channel regulator activity
structural molecule activity
Biological Process
Cellular Component
Molecular Function

FIGURE 2

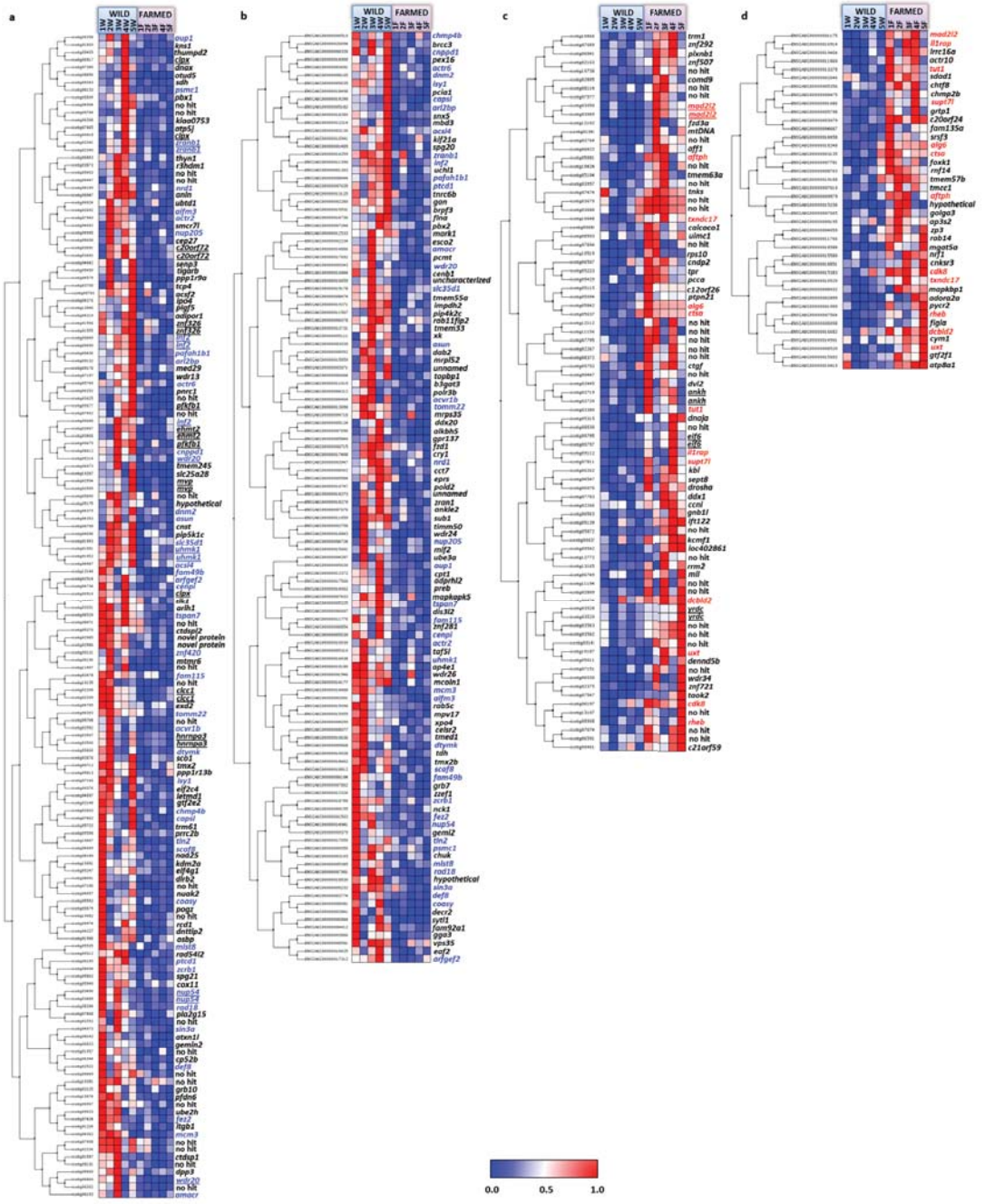
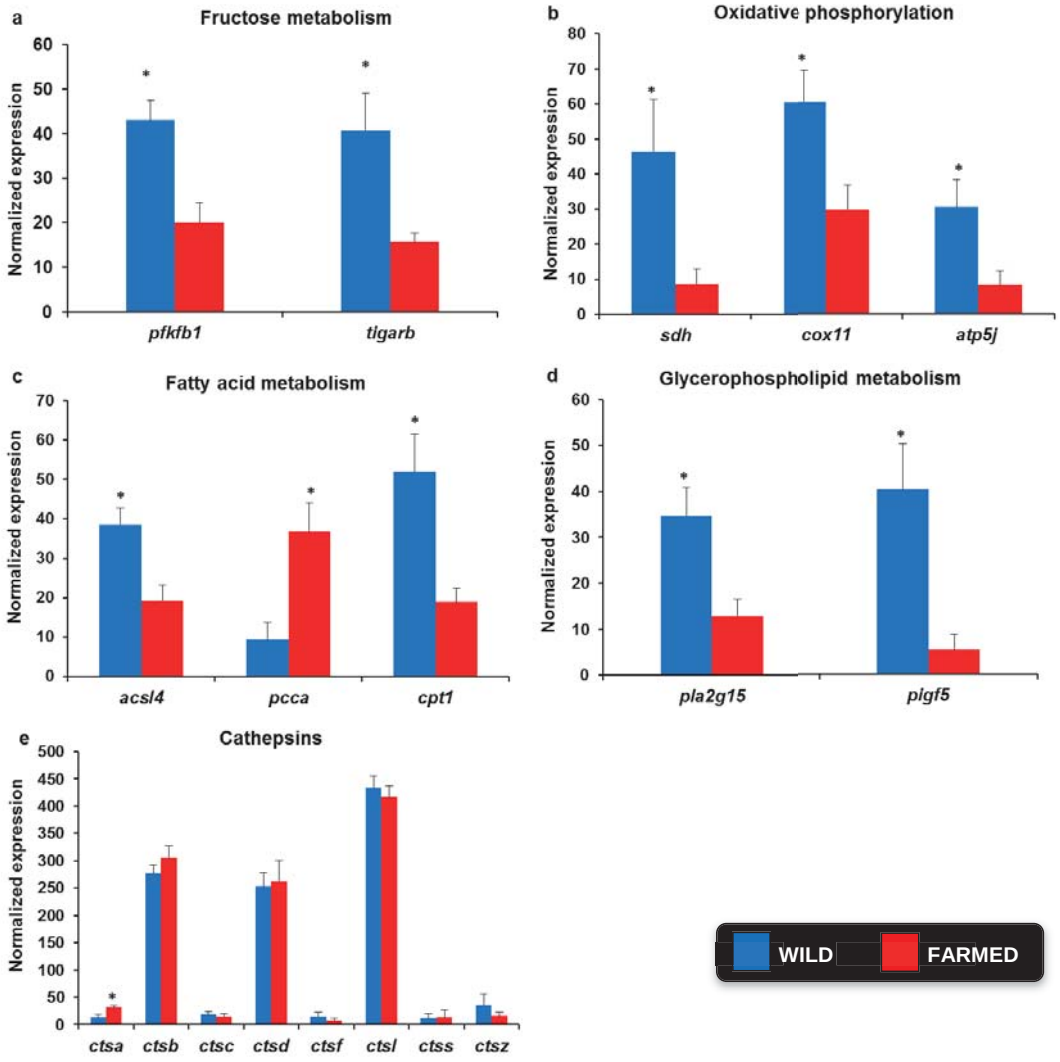


FIGURE 3


a
Fructose metabolism
Normalized expression
pfkfb1
tigarb
b
Oxidative phosphorylation
Normalized expression
sdh
cox11
atp5j
c
Fatty acid metabolism
Normalized expression
acsl4
pcca
cpt1
d
Glycerophospholipid metabolism
Normalized expression
pla2g15
pigf5
e
Cathepsins
Normalized expression
ctsa
ctsb
ctsc
ctsd
ctsf
ctsl
ctss
ctsz
WILD
FARMED