DIFFERENTIAL GENE EXPRESSION ANALYSIS REVEALS FUNCTIONAL ROLES OF CRYPTIC ORANGE-PIGMENTED ORGAN IN THE COLORLESS HEAD KIDNEY OF ANTARCTIC ICEFISHES

BY

EUN HYE YOON

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biology
with a concentration in Ecology, Ethology, and Evolution
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Master's Committee:

Professor Chi-Hing C. Cheng, Chair and Director of Research
Professor Arthur L. DeVries
Assistant Professor Julian Catchen

ABSTRACT

We discovered an orange-pigmented, discrete body several millimeters in size embedded within the anterior end of the colorless head kidney of the hemoglobinless Antarctic icefishes. No macroscopically visible discrete organ has previously been reported in Antarctic fishes or teleost fishes in general. To assess the tissue type and functions of this cryptic orange body, we performed RNA sequencing and analyzed the orange body transcriptome relative to the surrounding head kidney tissue. We found 5,263 differentially expressed genes in the orange body, of which 4,624 genes were up-regulated and 639 down-regulated relative to the head kidney tissue. Differential gene expression analyses detected the enrichment of two main biological pathways: steroid hormone production and proline biosynthesis. The steroidogenic activities of the orange body indicate it performs functional roles of the interrenal tissue cells – the teleost homolog of the tetrapod adrenal cortex but are normally dispersed among the teleost head kidney tissue. Transcription of many cytochrome P450 enzymes participating in steroidogenesis were upregulated, potentially leading to greater levels of ROS and peroxidation of the lipid-rich steroid producing orange body cells. Proline is known to act as an antioxidant, thus, the detected enrichment in proline synthesis may be related to antioxidation defense mechanism, conferring protection against lipid peroxidation in steroid rich orange body cells.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Chris Cheng-DeVries, not only for her guidance and support throughout my thesis completion, but for also teaching me the importance of paying attention to details during my time in the wet lab. She may never know how much this has helped me in other aspects of life. I thank her for giving me a chance and welcoming me into her lab. I also would like to thank Dr. Art DeVries for encouraging me to learn biology and become a researcher. Little details like leaving the Science magazine every week on my desk made a difference. Next, I would like to thank Dr. Julian Catchen for being part of my committee and for generously providing computer resources that helped me complete my thesis. Special thanks go to Kevin Bilyk, for his insightful comments to my thesis writing and his friendship. Lastly, I would like to thank my husband Guillermo for his unconditional support, I could not have done it without him.

To my little David.

TABLE OF CONTENTS

INTRODUCTION

The Southern Ocean (SO) surrounding Antarctica is a uniquely isolated and frigid marine ecosystem. The fish fauna is highly endemic, dominated by the perciform suborder Notothenioidei, which comprises five Antarctic families: Nototheniidae, Harpagiferidae, Artedidraconidae, Bathydraconidae, and Channicthyidae (Eastman, 2005). Evolution in chronically freezing conditions has led to both adaptive trait gain and maladaptive trait loss in these fishes. The key evolutionary adaptation shared among all five families is the synthesis of plasma antifreeze glycoproteins, which lowers the freezing point of their body fluids below ambient water temperatures making life possible in the icy and freezing SO (DeVries and Cheng, 2005). Paradoxically, the icefish family (Channichthyidae) has suffered the loss of hemoglobin and red blood cells (Cocca et al., 1995; Ruud, 1954) thought to be indispensable for vertebrate life. The loss gives the icefishes the odd appearance of not carrying any visible blood in their blood vessels. Because of this unique characteristic in all channicthyids, they are referred to as the 'white-blooded Antarctic fish'. Along with having lost hemoglobin, six icefish species have additionally lost myoglobin, the intracellular oxygen-storage protein (Sidell et al., 1997).

In the absence of hemoglobin-containing red blood cells in circulation, the tissues and internal organs of the icefishes are characteristically white or colorless, except for the thin black pigmented connective tissue sheets of the peritoneum. It is thus very peculiar to discover the presence of a small but distinctive pinkish-orange to intense orange colored organ (hereon called orange body) superficially embedded in each of the pair of colorless head kidney in icefishes (Figure 1). We have found this orange body in eight of the 16 species of icefishes that we have been able to capture and examine to-date, namely *Champsocephalus gunnari, Champsocephalus esox, Pseudochaenichthys georgianus, Chaenocephalus aceratus, Chionobathyscus dewitti, Cryodraco antarcticus, Chaenodraco wilsoni,* and *Chionodraco rastrospinosus*. Thus, the orange body is most likely a conserved anatomical structure across the icefish family. There was no prior report of its presence until our discovery, thus the anatomical and functional identity of this pair of orange bodies are unknown.

The location of the orange body in the head kidney suggests a number of functional possibilities. The kidney in teleost fish is a composite structure located ventro-lateral along the vertebral column, from the cloaca to the pharyngeal region. It comprises two histologically and functionally distinct components: the non-renal head kidney and the excretory or renal caudal kidney. The head kidney at the anterior consists of bilateral lobes that penetrate underneath the gill arches. It is the main organ for the production of blood cellular components (hematopoiesis) in teleost fishes, both myeloid and lymphoid cells, thus it encompasses the roles associated with the immune system as well as with the bone marrow in tetrapod vertebrates. In addition, teleost head kidneys house the homologous tissues of the tetrapod adrenal gland (Hanke and Kloas, 1995). The tetrapod adrenal gland is organized into two distinct morphological and functional regions: a cortex that produces corticosteroids, and a medulla that produces catecholamines. The adrenal tissues in teleosts however, are not known to form a discrete, macroscopically visible organ, but occur as small patches of cells distributed among the hematopoietic tissue that makes up most of the head kidney. The teleost adrenal homolog consists of interrenal tissues and chromaffin tissues, which are the functional equivalent of the steroidogenic cortex and catecholaminergic medulla respectively of the tetrapod adrenal gland. In most fishes, interrenal steroidogenic and chromaffin catecholaminergic cells usually occur in close proximity or are intermingled with each other. In four perciform fish and three salmonids examined, steroidogenic interrenal cells are found to be localized within the head kidney only, whereas catecholaminergic chromaffin cells can be found either in the head kidney, caudal kidney or dispersed in both organs around vein walls (Grassi Milano et al., 1997). Interrenal cells display a wide range of morphology. They can be scattered and interspersed among the hematopoietic tissue, or arranged as a single group, strand, or as a follicular structure (Abdel-Aziz, 2010). In contrast, the orange body in the icefish is a discrete structure (>5mm) visible to the naked eye, and occupies a specific anatomical location, at or near the anterior tip of the head kidney. Despite these differences, the similar localization of interrenal tissues of various teleosts (Milano et al., 1997) and the orange body of the icefishes to the head kidney raises the possibility that the orange body may also be interrenal tissue that had evolved into a discrete gland in channichthyids.

In this study we test this hypothesis by assessing the functional roles of the cryptic orange body in Antarctic icefishes through gene expression profiling. We performed RNA sequencing (RNA-Seq) of both the orange body and the surrounding head kidney tissue from the icefish *Champsocephalus gunnari* and analyzed for differentially expressed genes in the former relative to the latter. And by comparing the functional annotations of the differentially expressed genes in the orange body to known adrenal functions in vertebrate animals, we assess whether the orange body is the homolog of the adrenal gland and specifically of the steroidogenic interrenal tissue of teleosts or adrenal cortex of higher vertebrates. In the absence of an available genome assembly for *C. gunnari*, we de novo sequenced and assembled a reference transcriptome for RNA-Seq read mapping to quantify differential expression. This reference transcriptome serves to add to genomic resources for other studies of the unique Antarctic icefishes.

MATERIALS AND METHODS

Sample collection

Specimens of Antarctic icefish species *C. gunnari, P. georgianus, C. aceratus, C. dewitti, C. antarcticus, C. wilsoni, and C. rastrospinosus* were collected by trawls from the NSF R/V Laurence Gould from near shore waters along the Western Antarctic Peninsula during the Austral winter of 2014. Fish were transported to the NSF Palmer Station and held in flow through seawater aquarium. The S. American icefish *C. esox* was captured with gill net from the Patagonian fjord near Puerto Natales, Chile, and kept alive in aerated cooler of ambient water. For tissue sampling, fish was sacrificed and dissected on ice. The excised tissues were quickly immersed in glass vials of excess volumes of prechilled -20°C 90% ethanol and then stored in -20°C freezer. The ethanol was refreshed three times within the next 24-36 hours to effectively dehydrate the tissues and inactivate nucleases. All fish handling and sampling followed approved protocol (12123) by the IACUC at the University of Illinois, Urbana-Champaign. Ethanol-preserved samples were shipped to the University of Illinois, Urbana-Champaign on dry ice and stored at -20°C until use.

RNA Extraction and Quality Control

All RNA extractions from tissues of the chosen icefish species *C. gunnari* and quality verifications followed the same procedures for samples used in de novo transcriptome sequencing and RNA-Seq of orange body and head kidney. Tissues were processed using a Bullet Blender homogenizer with Zirconium 0.5 mm beads (Next Advance Inc.) and TRIzol reagent (Thermo Fisher Scientific Inc.) following manufacturer's protocol. The final precipitated RNA pellets were resuspended in 0.5x Tris-EDTA buffer (TE; 5mM Tris-HCl, 0.5 mM EDTA, pH 8.0). RNA concentration was quantified using Epoch™ microplate (Take 3) spectrophotometer (BioTek) and the integrity of each sample was evaluated by running 1µg of each sample on a 1% denaturing formaldehyde agarose gel. After this initial check of total RNA quality, 10 µg of each RNA sample was DNAse treated with 2 U of DNase I (BioLabs) at 37°C for 10 min, then purified using spin columns (Zymo Research) and eluted with 0.5x TE buffer. The

cleaned RNA samples were quantified by Qubit fluorometer 2.0 (Invitrogen) then both quality and integrity were determined using an RNA nano chip on Agilent Bioanalyzer 2100.

Reference Transcriptome Construction and Sequencing

De novo reference transcriptome was constructed from messenger ribonucleic acid (mRNA) populations from a wide variety of tissues and organs isolated from five different *C. gunnari* adult specimens to ensure a comprehensive profile of the set of transcripts that are being expressed (Supplementary Table 1). A total of 20 tissues and organs were used for the construction of the reference transcriptome, namely head kidney, orange body, caudal kidney, heart, liver, spleen, brain, gills, skin, ovary, testis, eye lens, eye cups, esophagus, stomach, anterior intestine, pyloric ceca, mesentery, red muscle, and white muscle. RNA extraction and quality control protocols were consistent as those followed for the experimental RNA-Seq reads. Cleaned RNA samples achieved excellent RNA integrity number (RIN; predominantly ≥9.0) as reported by Agilent 2100 Bioanalyzer analysis. A pooled RNA sample was generated by combining 1 µg of total RNA from each of 12 of the 20 tissue-specific RNA samples, and 0.5 µg from each of the other eight for a total of 15.5 µg of RNA (Supplementary Table 1). The pooled RNA was submitted to the University of Illinois Roy J. Carver Biotechnology Center, where the cDNA library was prepared and paired-end (PE) sequenced (2x250 bp) on a single lane on an Illumina HiSeq 2500 (Illumina Inc.) platform. Raw PE sequencing reads were subjected to read sequence quality control using AFTERQC (v.0.9.6) with default parameter settings.

De novo Reference Transcriptome Assembly Procedure, Quality Assessment, and Functional Annotation

Trinity v.2.8.3 was used to assemble raw reads into putative transcripts with default parameter settings and strand information. From the assembly, transcript abundance was estimated using both the RNA-Seq by Expectation-Maximization (RSEM; v.1.3.0), an abundance estimation method, and bowtie2 (v.2.3.4.3), an aligner method, with the objective of removing low-level expression transcripts that might represent transcripts derived from chimeras or sequencing errors. A threshold of 0.3 TPM was chosen to

retain transcripts in the assembly. This filtration step was carried out with the application of the Trinity built-in script (filter_low_expr_transcripts.pl). Redundancy of the assembly was tackled using CD-HIT-EST (v.4.6.8) with a 95% similarity threshold. To validate the quality of the 'polished' assembled reference transcriptome, we used BUSCO (Benchmarking Universal Single-Copy Orthologs; Simao et al., 2015) v.3.0 to mine for near-universal single-copy orthologs in our de novo assembly against both the actinopterygii and metazoa databases.

Functional annotation was performed using the Trinotate suite (v.3.1.1) including homology identification using BLASTX and hmmer, along with screening options signalP v4, tmhmm v2, and RNAmmer. We used FunctionAnnotator (Chen TW et al., 2017), a web-based annotation tool, to maximize annotation of possible unannotated transcripts.

RNA-Seq of Head Kidney and Orange Body

The orange body was dissected from the head kidney under a dissecting microscope to ensure little or no surrounding hematopoietic tissue was included. Total cellular RNA was isolated from approximately 100 mg of head kidney tissues and up to about 15 mg (maximum mass) of orange body. RNA extraction quality verifications were as described above. The cleaned RNA samples achieved RIN values between 8.7 and 9.7. We constructed a total of eight cDNA libraries (i.e., four replicates for each tissue under investigation) using mRNA selection based-methods with the TruSeq Stranded mRNA Library Prep Kit (Illumina Inc., San Diego, CA, USA) following manufacturer's protocol. The mRNA fraction from total RNA was extracted using poly-T oligo conjugated magnetic beads (Agencourt AMPure XP), and then fragmented into small pieces using divalent cations at 94°C for about 8 min. The RNA fragments were reverse transcribed to first strand cDNA using Super Script II reverse transcriptase and random primers. Strand specificity is achieved by replacing dTTP with dUTP during the second strand synthesis. The products were then amplified with 10 cycles by PCR to create the final cDNA library. Completed libraries were checked for quantity and quality via qPCR and Agilent Bioanalyzer 2100 (using High sensitivity DNA chip) respectively. After performing quality control, the sample libraries were multiplexed and single-end (SE) sequenced (1x150 bp) on a single lane on an Illumina

HiSeq 2500 (Illumina Inc.) platform at the University of Illinois Roy J. Carver Biotechnology Center. Quality control of raw RNA-Seq reads was done using AfterQC (v.0.9.6).

Alignment and RNA-Seq read Quantification

For alignment and RNA-Seq read quantification and estimation, we used salmon v.0.11.3 (Patro et al, 2017), which is an alignment-free transcript quantification method. Salmon implements a novel alignment algorithm based on the quasi-mapping approach. This approach is thus more robust to variations in sequencing reads and transcript composition in the reference transcriptome wherein the alignment takes place (this is in contrast to the base-to-base alignments that traditional alignment methods are relied upon). Salmon principally computes isoform-level abundance estimations and then summarizes isoform-level abundance estimation to generate gene-level counts with the help of tximport (Soneson et al., 2015). We then used the gene-level abundance estimation to perform the differential gene expression (DGE) analysis.

Identifying Differentially Expressed (DE) Genes Between the Orange Body and Head Kidney

We used the R package edgeR v.3.7 (Robinson et al., 2009) to conduct DE analysis of our experimental reads. The experimental design was set to be in paired samples. Internal trimmed mean of M values (TMM) normalization was done to account for RNA-Seq composition bias with the function *calcNormFactors* and then proceeded to identify DE genes using the quasi-likelihood F-test. We applied Benjamini and Hochberg correction for multiple hypothesis testing (at an FDR adjusted p-value <0.05) to consider genes as differentially expressed. The up- and down-regulated DE genes from OB relative to HK were identified (these were interpreted as OB and HK tissue-relevant over-expressed genes, respectively).

Gene Ontology Functional Enrichment analysis

To better understand the biological functions of the set of DE genes that are significantly up- and down-regulated in OB relative to the HK tissue in *C. gunnari*, we used GoSeq (v.3.8) to investigate

overrepresentation of biological functions within the DE set of OB genes obtained from the edgeR analysis. GoSeq was the preferred program for enrichment analysis due to its property of correcting for gene length selection bias that is characteristic of the RNA-Seq data. This is done by GoSeq by applying the probability weighting function (pwf) to all of the transcripts being surveyed as DE or non-DE. This in turn, will form the basis for the null probability distribution for the enrichment analysis. GoSeq employs an extension of the hypergeometric distribution (i.e., Wallenius non-central hypergeometric distribution) for the Gene Ontology category enrichment analysis.

The Gene Ontology enrichment analysis was done separately for the up- and down-regulated DE genes (i.e., the tissue-relevant over-expressed set of genes) to better disclose the relative functional differences between OB and HK tissue.

We used REVIGO (Supek et al., 2011) to summarize and better visualize the list of enriched GO terms obtained from the tissue-relevant over-expressed DE genes. The full set of enriched GO terms along with their respective p-values was summarized using the Revigo default parameter settings. One of the visualization methods offered by REVIGO are scatterplots, which we utilized for further analysis. The scatterplots are constructed using a multidimensional scaling procedure that allows more semantically similar GO terms to be closer in the plot.

RESULTS

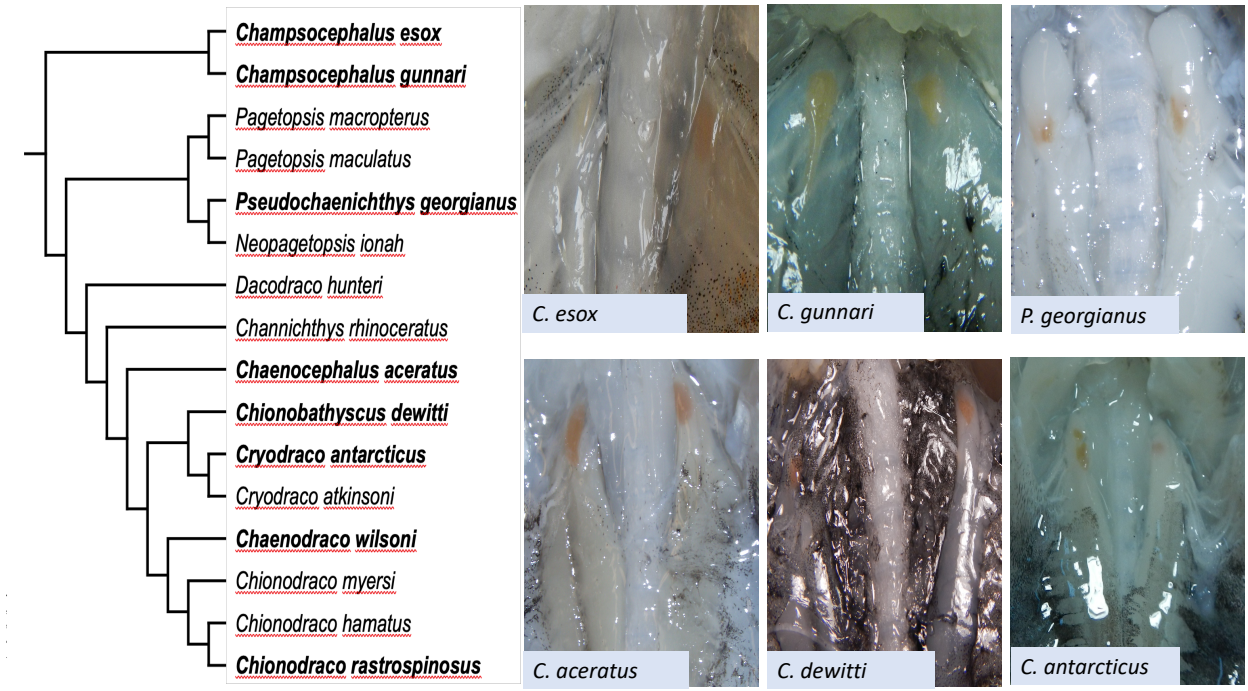


Figure 1. Images of OB (orange body) superficially embedded at or near the anterior tip of each pair of the head kidney lobes in eight icefish species examined. The two icefish species with orange body not shown here are *Chaenodraco wilsoni* and *Chionodraco rastrospinosus*. Icefish phylogeny follows Near et al. (2003).

Orange Body Embedded in the Head Kidney of C. gunnari

We have discovered an orange-colored body superficially embedded at or near the rostral tip of each of the paired anterior lobes of the colorless (white) head kidney of eight Antarctic icefishes, namely Champsocephalus *esox, C. gunnari, Pseudochaenichthys georgianus, Chaenocephalus aceratus, Chionobathyscus dewitti*, *Cryodraco antarcticus*, *Chaenodraco wilsoni*, and *Chionodraco rastrospinosus* (Figure 1). The occurrence of this orange body in half of the 16 species of Channichthyidae, indicates it is likely a shared anatomical trait in the family. This cryptic orange body stands out from its surrounding head kidney tissue by its light orange to intense orange color. Upon dissection, the OB appears to be separated from the surrounding hematopoietic tissue by a thin layer of connective tissue making it a discrete tissue or organ distinct from the surrounding head kidney. The orange body vary in shape and

size. The shape goes from rounded (Figure 1, right OB of *C. antarcticus*), frequently oblong to slightly elongated tear-drop shaped (Figure 1, *C. esox, C. gunnari, C. aceratus*, and *C. dewitti*), to sometimes irregular in outline (Figure 1, *P. georgianus*, and left OB of *C. antarcticus*). The size ranges from about 3mm to 8mm in the long axis. There was no previous record or study of this orange-pigmented body in the Antarctic icefishes or any other notothenioid fishes. Also, as far as we know, there were no published report of a macroscopically visible, discrete organ within the head kidney of teleost fishes in general. Thus, the identity and function of this novel orange body in the Antarctic icefishes are intriguing and entirely unknown. As first approach to understand the functional roles of the OB, we undertook RNA-Seq analyses of the transcriptome of the OB relative to the surrounding HK of the icefish *C. gunnari* to assess what physiological functions were enriched in the OB.

C. gunnari de novo Reference Transcriptome

In the absence of a draft genome for *C. gunnari*, we first de novo sequenced and assembled a reference transcriptome for the species for subsequent use in RNA-Seq analyses of the OB and HK. Messenger RNAs coming from 20 different tissues (Supplementary Table 1) were PE sequenced (2x250 bp) using the 2500 HiSeq Illumina platform. We obtained a total of ~100M raw paired-reads and after using AfterQC (v.0.9.6) ~96M sequencing reads were retained. We discarded low quality reads (PHRED33 score <15), Illumina adaptors, and reads below 35 bases in length, which resulted in a loss of 3% of reads (Supplementary Table 2). De novo assembly with Trinity generated a total of 307,384 putative genes and a total of 469,524 trinity transcripts with an average contig length of 1037 bases. RSEM output consisted in a table with trinity transcripts ids with their respective trinity gene ids, length, effective length, expected count, TPM, FPKM, and IsoPct. RSEM found that 78,985 transcripts (from a total of 469,524) were expressed by at least 1 TPM. We retained 53.58% of transcripts (251,560/469,524) in the de novo assembled transcriptome after removing transcripts with an expression level below the set threshold of 0.3 TPM. Redundancy of the assembly was removed using CD-HIT-EST (v.4.6.8) with a 95% similarity threshold, which reduced the number of transcripts in the assembly to 208,925 from 251,560 (Supplementary Table 3). To additionally assess the quality and completeness of our assembled

reference transcriptome we used BUSCO v.3.0 to mine for near-universal single-copy orthologs in our de novo assembly against both the actinopterygii and metazoa reference databases. BUSCO identified 4,252 (92.8%) and 972 (99.4%) complete genes from a total of 4,584 and 978 highly conserved gene groups corresponding to the actinopterygii and metazoan dataset respectively. BUSCO's results (Supplementary Table 4) suggest that the *C. gunnari* de novo reference transcriptome has a relatively high coverage of genes that are highly conserved in ray-finned fishes and metazoans alike.

From a total of 208,925 transcripts from the *C. gunnari* de novo assembled reference transcriptome, 70,848 transcripts were annotated with the built-in swissprot database from the trinotate suite. We then used the web-based tool FunctionAnnotator to annotate the remaining 138,077 unannotated transcripts. However, only 8,115 transcripts were mapped to Gene Ontology (GO) terms by blast2go within the FunctionAnnotator environment, yielding a total of 78,963 transcripts annotated to their respective GO terms. Studies on the model fish zebrafish provided evidence for the presence of ~26,000 protein-coding genes (Collins et al., 2012), which suggest that the number of annotated transcripts in our reference transcriptome should be a relatively good representation of the universe of genes and its isoforms that are being transcribed in the icefish *C. gunnari*.

Orange Body and Head Kidney Tissue Experimental RNA Sequencing and Transcript Abundance Estimation

An average of ~24M raw sequencing reads were obtained for each of the RNA-Seq libraries. AFTERQC (v.0.9.6) quality check with default parameter settings removed less than 0.5% of the raw reads for each of the eight samples (Supplementary Table 5). This provides a conservative coverage of about 100x assuming the head kidney (HK) expresses a total of about 30,000 transcripts with an average length of 1,500 bp. We further plotted our samples in a multi-dimensional scaling (MDS) graph (Figure 2), which clusters samples based on their differences in gene expression profiles. The graph shows that the distance between each pair of samples is largest by the leading log-fold change dimension 1 (x-axis) while dimension 2 (y-axis) separates samples on an individual basis (confirming the paired nature of the samples). That is, the differences between pair of individuals are larger than those within individuals

indicating that the differences between HK and OB library samples are shown to be statistically significant between replicates. The distance between OB samples and HK samples is ≥4 units of logFC, i.e., a fold change of more or less than 16-fold. In sum, the MDS plot indicates: 1) gene expression differences exist between OB and HK tissues and 2) biological replicates of each tissue cluster together in the graph, which demonstrate consistency between replicates and reproducibility of this RNA-Seq study.

Alignment of experimental raw RNA-Seq SE reads to the de novo reference transcriptome and gene-level abundance estimation were done using Salmon. Salmon identified and quantified a total of 152,826 genes (found either in HK, OB or in both tissues) with more than 23M reads for each of the 8 RNA-Seq experimental samples.

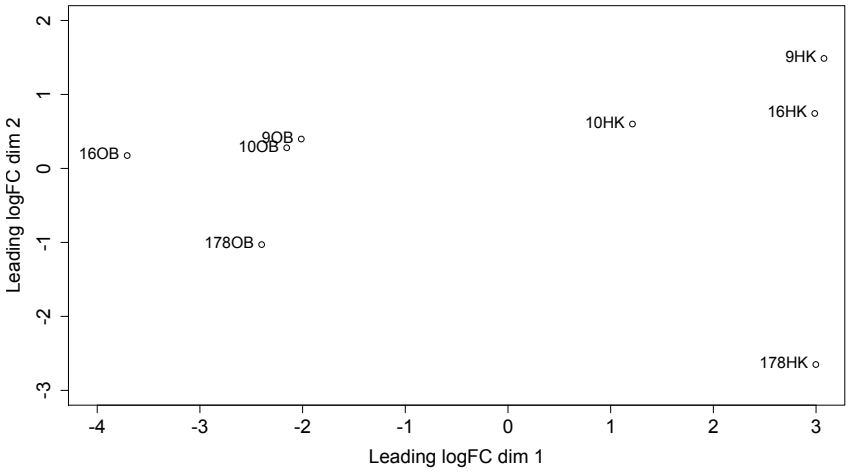


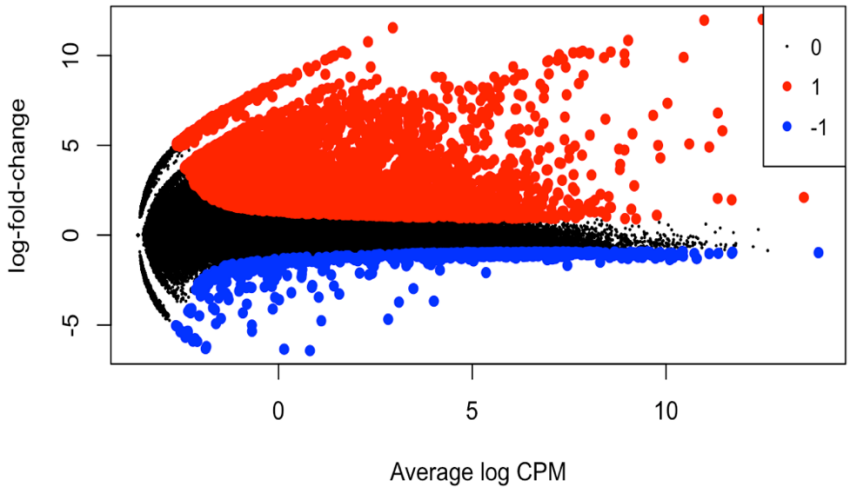
Figure 2. MDS plot of the experimental dataset. Sample labels consist of the fish number followed by an abbreviation of the tissue type (OB, orange body; HK, head kidney). Leading dimension 1 (x-axis) segregates the samples according to tissue while dimension 2 (y-axis) separates the samples on an individual basis.

Identification of DE genes between OB and HK tissue

We performed differential expression analysis with edgeR using its quasi-likelihood F-test method and applied Benjamini and Hochberg correction for multiple hypothesis testing at an FDR p-value adjusted threshold of 0.05. From the 152,826 genes estimated by Salmon, we identified 5,263

differentially expressed genes wherein 4,624 genes (88%) were up-regulated and 639 (12%) down-regulated in OB relative to HK tissue (Figure 3A). These up and down-regulated DE genes can be interpreted as OB and HK tissue-relevant over-expressed genes respectively. That is, if we perform the reciprocal analysis with OB as the reference, we would obtain 639 up- and 4,624 down-regulated genes for the HK tissue. Figure 3B shows a heat map with hierarchical clustering of genes and samples incorporating only the top 30 most DE gene symbols of OB relative to HK tissue, illustrating the differences in expression patterns across all libraries. Gene symbols with their corresponding full gene names, adjusted FDR p-values, and logFCs are reported in table 1. The top 5 genes along with other genes within the top 30 most significantly up-regulated DE genes in OB were steroidogenic acute regulatory protein or StAR (logFC=12.00 and FDR=5.86e-125), steroid 21-hydroxylase or CYP21 (logFC=11.96 and FDR=1.86e-119), cytochrome P450 11B or CYP11B (logFC=10.84 and FDR=3.03e-111), steroid 17-alpha-hydroxylase or CYP17 (logFC=10.20 and FDR=8.40e-107), 3 beta-hydroxysteroid dehydrogenase or 3BHSD (logFC=9.89 and FDR=2.32e-105), cytochrome P450 2U1 or CYP2U1 (logFC=9.37 and FDR=1.39E-86), hydroxysteroid 11-beta dehydrogenase 1-like protein or HSD11B1L (logFC=7.81 and FDR=2.77E-75), and cholesterol side-chain cleavage enzymes or CYPscc (logFC=6.79 and FDR=1.81E-73). All of these genes (except for StAR) encode proteins and enzymes that belong to the cytochrome P450 super family of enzymes and to the hydroxysteroid dehydrogenases that are known to be the main enzymatic machinery of the steroid biosynthetic pathway. Furthermore, StAR plays a role that is considered rate-limiting step in the steroidogenic pathway. StAR expression is tightly associated to steroidogenic tissues as it plays a crucial role delivering cholesterol from the cytoplasm to the inner membrane of the mitochondria (Lin et al., 1995). Remarkably, transcripts from these genes were found to be expressed in OB at astoundingly high levels, at 110- to 4000-fold compared to HK. To further identify the biological functions performed by the up- and down-regulated genes, these were analyzed separately.

(A)



(B)

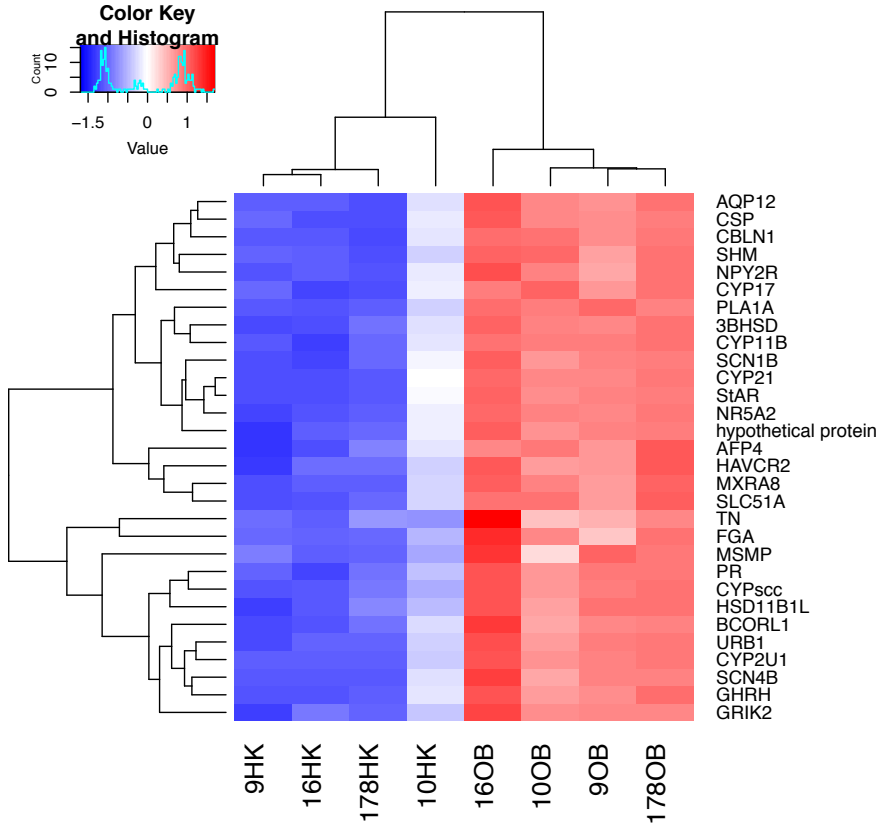


Figure 3. Differential gene expression of OB vs. HK tissue in *C. gunnari*. (A) Mean difference (MD) plot showing gene differential expression between OB relative to HK tissue at an FDR of 5% with log2-fold change plotted against each gene average abundance. Significantly up- and down-regulated DE genes are highlighted in red and blue, respectively. (B) Heat map with hierarchical clustering of genes (y-axis) and samples (x-axis) across all 8 samples using the top 30 most DE genes between OB and HK tissues. Red, blue, and white indicate up-, down-regulated, and unchanged expression level of genes, respectively.

Table 1. Top 30 most significant hits of the OB up-regulated DE genes relative to HK in *C. gunnari* with corresponding FDR adjusted p-values and logFC.

Gene Symbol	Gene full name	Adj. p-value	logFC
StAR	Steroidogenic acute regulatory protein, mitochondrial	1.77E-126	12.00
CYP21	Steroid 21-hydroxylase	5.64E-121	11.96
CYPP11B	Cytochrome P450 11B, mitochondrial	9.18E-113	10.84
CYP17	steroid 17-alpha-hydroxylase/17,20 lyase-like	2.54E-108	10.20
3BHSD	3 beta-hydroxysteroid dehydrogenase	7.02E-107	9.89
AFP4	Type-4 ice-structuring protein LS-12	1.39E-101	9.63
NR5A2	Nuclear receptor subfamily 5 group A member 2	2.61E-97	10.15
MSMP	Prostate-associated microseminoprotein	7.96E-96	10.08
Hypothetical Protein	hypothetical protein VOLCADRAFT	1.75E-94	10.11
GHRH	Somatoliberin	1.72E-93	10.24
SLC51A	Organic solute transporter subunit alpha	1.79E-92	8.89
BCORL1	PREDICTED: BCL-6 corepressor-like protein 1	6.64E-91	9.88
SCN1B	Sodium channel subunit beta-1	8.97E-91	10.17
NPY2R	PREDICTED: neuropeptide Y receptor type 2-like, partial	1.33E-88	9.35
CYP2U1	Cytochrome P450 2U1	1.39E-86	9.37
GRIK2	Glutamate receptor ionotropic, kainate 2	1.19E-84	8.44
CBLN1	Cerebellin-1	4.17E-84	8.93
SCN4B	Sodium channel subunit beta-4	1.40E-83	9.71
URB1	Nucleolar pre-ribosomal-associated protein 1	2.00E-83	8.98
CSP	Circumsporozoite protein	4.51E-81	8.96
PR	Progesterone receptor	5.08E-79	8.14
HAVCR2	PREDICTED: hepatitis A virus cellular receptor 2 homolog	3.36E-78	8.40
FGA	Fibrinogen alpha chain	1.91E-77	7.34
AQP12	Aquaporin-12	3.52E-77	8.62
SHM	Serine hydroxymethyltransferase, mitochondrial	1.82E-76	7.64
MXRA8	Matrix-remodeling-associated protein 8	6.93E-76	8.32
TN	Tenascin	1.37E-75	9.66
HSD11B1L	Hydroxysteroid 11-beta-dehydrogenase 1-like protein	2.77E-75	7.81
PLA1A	Phospholipase A1 member A	6.78E-74	7.75
CYPscc	Cholesterol side-chain cleavage enzyme, mitochondrial	1.81E-73	6.79

Gene Ontology Functional Enrichment Analysis

We performed separate functional classification enrichment analysis using GoSeq to identify enriched GO terms relevant to the up-regulated DE genes as well as the down-regulated DE genes in OB relative to HK tissue. Among the up-regulated 4,624 DE genes we obtained 110 enriched GO terms and among the down-regulated 639 DE genes, we found 136 enriched GO terms after performing a Benjamini

and Hochberg False Discovery Rate (FDR) correction (BH <0.05). The full list of enriched GO terms for the up- and down-regulated DE genes along with the number of genes that fall into each of the GO term categories can be found in Supplementary Figure 1. The 10 most significantly enriched GO terms in OB and HK are listed in table 2. For the orange body, the enriched terms predominantly reflect related functional categories including ornithine-oxo-acid transaminase activity, proline biosynthetic process and L-proline biosynthetic process, membranes, and signaling and channel activities. In contrast, the most significant GO terms in HK were associated with a diverse set of functions that relate to immune system, chemokine activity, cytokine activity, and antigen binding, as expected of this primarily lymphoid tissue in the hemoglobinless icefish (Romano et al., 2000).

Given the large list of enriched GO terms obtained from both OB and HK DE genes, they were further analyzed using REVIGO (Supek et al., 2011) (Figure 4). These enriched GO term categories are subclassified in three categories: biological process (BP), molecular function (MF), and cellular component (CC). There were clear differences in the functions that dominate in the three ontologies between over-expressed DE genes in the OB and HK tissues. For example, enriched BPs in the OB were highly geared towards metabolic processes, whereas in HK tissue BPs were highly enriched for responsive and regulatory processes. As for the MF category, OB tissue showed enrichment for signaling and oxidoreductase, kinase, and monooxygenase activities, whereas in HK tissue these were directed towards receptor activity and binding. In addition, the functional enrichment analysis of the CC ontology in OB cells showed significantly more subcellular activity within the membrane, plasma membrane and mitochondrial matrix, while in HK cells there was significantly more activity occurring in the extracellular space. These marked differences observed between OB and HK indicate that these two tissues have distinctive functional roles. OB was principally involved in pathways that were mainly carried out within membranes and mitochondria, such as proline biosynthesis and pathways with abundant redox reactions, monooxygenase activity, kinase activity, signaling, lyase activity, and cofactor binding. HK in contrast, was associated to have high extracellular activity involving pathways that include that of the immune system and hematopoiesis (concerning only leukocyte cell lineages).

Table 2. Top 10 most significantly enriched GO terms in OB and HK tissues of *C. gunnari.* BP, biological process; CC, cellular component; MF, molecular function.

Category	p-value	Number of DEG in category	Term	Ontology
OB				
GO:0016020	2.00E-12	413	membrane	CC
GO:0004587	1.48E-11	7	ornithine-oxo-acid transaminase activity	MF
GO:0006561	5.40E-11	9	proline biosynthetic process	BP
GO:0005886	4.55E-10	168	plasma membrane	CC
GO:0055129	5.54E-10	8	L-proline biosynthetic process	BP
GO:0038023	2.81E-09	118	signaling receptor activity	MF
GO:0060089	4.39E-09	115	molecular transducer activity	MF
GO:0022838	7.17E-09	71	substrate-specific channel activity	MF
GO:0005216	7.20E-09	71	ion channel activity	MF
GO:0015267	1.08E-08	73	channel activity	MF
HK				
GO:0006955	7.43E-31	40	immune response	BP
GO:0002376	1.36E-27	48	immune system process	BP
GO:0008009	5.96E-19	12	chemokine activity	MF
GO:0006952	8.47E-19	29	defense response	BP
GO:0005576	1.05E-18	39	extracellular region	CC
GO:0050896	2.22E-18	67	response to stimulus	BP
GO:0042379	3.76E-18	12	chemokine receptor binding	MF
GO:0005126	4.33E-17	16	cytokine receptor binding	MF
GO:0005125	2.38E-16	14	cytokine activity	MF
GO:0003823	1.07E-13	10	antigen binding	MF

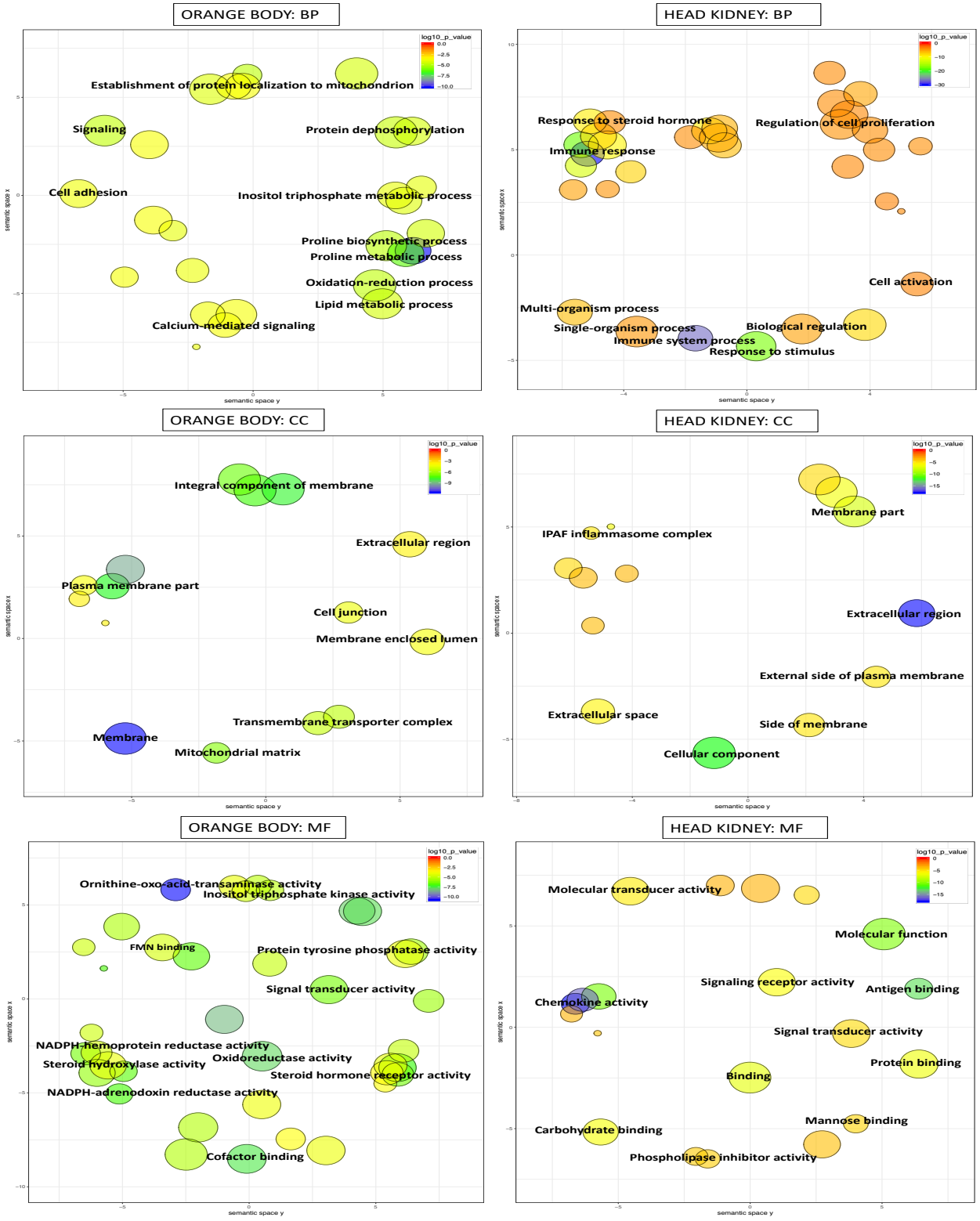


Figure 4. Enriched GO terms summarized by Revigo Scatterplots for each orange body and head kidney tissues with their corresponding BP, CC, MF ontologies. The color of the bubbles makes reference to p-values, where blue and red represent the low- and high-end respectively of the distribution of p-values.

DISCUSSION

This study was motivated by the discovery of an unexpectedly brightly colored body within the colorless head kidney of the hemoglobinless Antarctic icefishes (Figure 1). The localization of the orange body to the head kidney where the teleost adrenal components – interrenal and chromaffin tissue are known to reside, raises the possibility that it may be the adrenal tissues in icefish but occur in discrete glandular form instead of the dispersed cellular distribution among hematopoietic tissue shared by all other teleost species examined. We test this hypothesis by investigating the differential expression of genes of the orange body relative to its surrounding head kidney tissue by RNA-Seq analyses to infer its function and tissue type. We constructed a de novo reference transcriptome of *C. gunnari* species covering a broad selection of tissues including the head kidney and the cryptic orange body for RNA-Seq read mapping to quantify gene expression.

Elucidation of the Functional Roles of OB

The results of the analyses for differential gene expression in the cryptic orange body revealed strong upregulation of genes involved in two major functional pathways – steroid biosynthesis and proline biosynthesis.

Steroid Biosynthesis

The top 5 genes along with three other genes within the top 30 most significantly up-regulated DE genes in OB relative to HK tissue were *StAR, CYP21, CYP11B, CYP17, and 3BHSD, CYP2U1, HSD11B1L*, and *CYPscc* (table 1). The products of these genes are known members of the suite of proteins and enzymes that play key roles in steroid biosynthesis (Castro et al., 2015). Their apparent co-upregulated expression indicates that the orange body is an active steroidogenic tissue.

Steroidogenic acute regulatory protein (StAR) is a transport protein essential in the transfer of cholesterol into the matrix side of the inner membrane of mitochondria where cholesterol is converted to pregnenolone by the cholesterol side chain cleavage enzyme or CYPscc. The cytochromes Cyp11B,

Cyp17, Cyp21, and CYP2U1 are members of the large cytochrome P450 family of heme-containing enzymes with diverse metabolic functions. In steroid biosynthesis, cytochrome P450 enzymes are responsible for the hydroxylation and cleavage of the steroid substrates. They serve as true monooxygenase systems to catalyze the conversion of a substrate (S) to a more polar hydroxylated form (S-OH) using nicotinamide adenine dinucleotide phosphate (NADPH) as the electron donor for the reduction of the molecular oxygen. Interestingly, each type of the cytochrome P450 systems that are found in steroidogenic tissues have their own set of electron transfer proteins that are utilized to transport electron from NADPH to the cytochrome P450 enzymes. Eukaryotic cytochrome P450 enzymes are of two types, one is localized in the mitochondria and the other is said to be in the membrane of the endoplasmic reticulum, which both cellular components were found enriched in OB. Moreover, mitochondrial cytochrome P450 enzymes rely on adrenodoxin proteins (Omura et al., 1966) whereas microsomal cytochrome P450 enzymes utilize coenzymes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) for electron transfer (Vermillion et al., 1981). Interestingly, "NADPH-adrenodoxin reductase activity" and "FMN binding" were enriched in OB, suggesting the expression and thus prevalence of both types of cytochrome P450 systems in OB tissue.

The highly up-regulated *3BHSD* (3 beta-hydroxysteroid dehydrogenase) and *HSD11B1L* (hydroxysteroid 11-beta-dehydrogenase 1-like protein), are non-cytochrome P450 enzymes part of the hydroxysteroid dehydrogenase family of enzymes. 3BHSD is considered a key enzyme as it constitutes an obligate step in the steroidogenic pathway, biosynthesizing all steroid hormones from mineralocorticoids, glucocorticoids, to estrogens and androgens (Simard et al., 2005). Studies on HSD11B1L enzyme are limited in fish; however, in humans, it catalyzes the interconversion of inactive to active glucocorticoids (Tomlinson et al., 2004).

For the rest of the enriched GO terms for OB, some of them showed varying degrees of association with steroid biosynthesis such as "lipid metabolic process", "steroid hormone receptor activity", "steroid hydroxylase activity", "monooxygenase activity", "cofactor binding", "vitamin binding", "steroid 17-alpha monooxygenase activity", "17-alpha-hydroxyprogesterone aldolase activity", "lyase activity", "oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor", and "NADPH-

hemoprotein reductase activity". Additionally, enriched GO terms such as "cyclic nucleotide biosynthetic process", "cyclic nucleotide metabolic process", "protein tyrosine phosphatase activity", "inositol trisphosphate kinase activity", "inositol triphosphate metabolic process", and "calcium-mediated signaling" are described as "second messenger systems" and are said to be important intracellular mediators in the steroid hormone biosynthetic pathway. Gupta and Hanke (1994) demonstrated that corticosterone levels could be increased or decreased by altering intracellular second messenger systems such as cAMP, Phospholipase C (InositolP3), and calcium ions in incubation medium of the adrenocortical cells in axolotl (an amphibian species). Emphasizing the necessary role of calcium ions for the secretion of the steroid hormone by the amphibian steroidogenic adrenocortical cells.

Although we did obtain up-regulated DE genes for OB that are reported to be expressed in the corpuscle of Stannius—another endocrine gland found in teleost kidney responsible for calcium homeostasis—these were of weak signal and therefore, OB enriched functional terms did not show strong evidence supporting that OB cells may be related to the Stannius body tissue. However, we do not discard that OB tissue might be involved in calcium metabolism considering that cortisol in zebrafish has been found to be a hypercalcemic hormone (Lin et al., 2011).

Collectively, the differential analysis of the over-expressed genes of the OB relative to the HK tissue indicates that OB encompasses functions similar to that of the interrenal cells in other teleost fishes. Up-regulated DE genes in OB cells reveal association to the steroid hormone biosynthetic pathway, providing evidence supporting our hypothesis that the cryptic orange body in the icefishes encompasses the steroid-producing interrenal cells of teleost fish.

Up-regulated Gene Expression for Proline Biosynthesis

A strong involvement of the icefish orange body in proline biosynthesis is supported by top 3 most significantly enriched GO terms in the MF and BP categories. They are "ornithine-oxo-acid transaminase activity" (p-value=1.48E-11), "proline biosynthetic process" (p-value=5.40E-11), and "L-proline biosynthetic process" (p-value=5.54E-10) (Table 2). Ornithine-oxo-acid transaminase or OAT is a mitochondrial matrix enzyme that catalyzes the first reaction in the pathway for the eventual production

of the non-essential amino acid L-proline. Thus, these three enriched GO terms are clear signatures of the icefish orange body being an active proline biosynthetic tissue, in addition to steroidogenic. Furthermore, there were other enriched GO terms closely related to proline metabolism. These GO terms were, for example, “proline metabolic process” (p-value=3.97e-7), “glutamine family amino acid biosynthetic process” (p-value=1.2e-6), “cellular amino acid biosynthetic process” (p-value=1.2e-4), and “glutamate receptor activity” (p-value=2.7e-4). Moreover, a key proline biosynthetic enzyme, delta-1-pyrroline-5-carboxylate synthase or P5CS (logFC=2.82 and FDR=1.26E-13) was found among the OB upregulated DE genes along with the mitochondrial enzyme ornithine aminotransferase or OAT (logFC=4.30 and FDR=4.39E-16). P5CS and OAT catalyze the conversion of glutamate to PC5 and transamination or ornithine respectively, alternative first step in the biosynthetic pathway to formation of proline.

Apart from proteogenic usage, proline has been shown to have protective effects on cell survival against various abiotic and biotic stresses including oxidative stress (Krishnan et al., 2008). The stress protection feature that is attributed to intracellular proline levels might operate via different mechanisms, thus designating proline as a multifunctional amino acid capable of altering different aspects of the cellular metabolism. The mechanisms by which proline might aid in stress-relief are reported to occur during protein misfolding and aggregation (Samuel et al., 2000), energy management/utilization (Kowaloff et al., 1977; Kohl et al., 1988; Hayward et al., 1993), stabilizing cellular structure (Rudolph and Crowe, 1985), inhibition of cell programmed death or apoptosis (Chen and Dickman, 2005), and scavenging of free radicals (Mohanty and Matysik, 2001; Kaul et al., 2008). Indeed, as discussed by Krishnan et al. (2008), proline biosynthesis is an adaptive stress response in mammalian cells. The authors reached this conclusion by observing robust up-regulation of P5CS and P5CR transcript levels when exposing human embryonic kidney cells to low concentrations of reactive species hydrogen peroxide (H_2O_2).

Interrenal tissue contains high levels of polyunsaturated fatty acids (PUFA), which are susceptible to lipid peroxidation, along with the high likelihood of generating reactive oxygen species (ROS) from the CYP enzyme monooxygenase activities (Hornsby and Crivello, 1983; Yamazaki et al., 1996; Burczynski et al., 2001), we expected to find antioxidant-related pathways in the DE genes of OB.

However, none of the enzymatic antioxidant defense mechanisms were found to be enriched in OB relative to HK, except for NADPH, which was highly enriched in OB and is considered to be a potent cofactor with indirect antioxidant properties. We did detect several enzymes involved in antioxidant systems such as cytoplasmic thioredoxin reductase 1 (logFC=7.61 and FDR=1.74E-62), microsomal glutathione S-transferase 3 (logFC=3.47 and FDR=6.64E-17), and glutathione peroxidase 1 (logFC=0.91 and FDR=1.59E-3) being up-regulated in OB. However, other antioxidative agent such as extracellular superoxide dismutase (logFC=-4.11 and FDR=7.52E-9) was up-regulated in HK. However, these genes alone were not sufficient to emerge as functionally enriched ontologies in either OB or HK tissue. It might be that both OB and HK tissues express genes that are involved in antioxidant pathways at similar levels, thus we did not capture enrichment for these pathways in OB tissue.

It is therefore of interest that proline biosynthesis was found enriched in OB. Although the mechanisms of how proline might operate under oxidative stress cannot be elucidated through this study, proline has been well documented to have scavenging properties of ROS in plants (Smirnoff and Cumbes, 1989; Matysik et al., 2002). Further, the role of proline metabolism that has been recently established in oxidative stress-ridden cancer cells have shed light on the mechanisms of actions of proline in response to oxidative stress (Phang, 2018). As proline metabolism has been shown to protect against the deleterious effects of ROS, it has also been shown to produce ROS via proline-dependent pathways. This would indicate the intricacies of the proline metabolism, depending on the cellular context, it may serve various purposes by the reprogramming of its metabolic pathways. Moreover, it has been found that proline catabolism might aid in the propagation of tumorigenesis via the up-regulation of the enzyme proline dehydrogenase (Prodh), this finding has led researchers to explore the hypothesis that proline biosynthesis may be a possible future target for cancer therapy (Elia et al., 2017).

Thus, it is possible that the enrichment of proline biosynthesis detected in the OB/interrenal tissue of *C. gunnari* may be related to an antioxidant defense against oxidation of biological macromolecules (such as lipid peroxidation). Due to the presumably high susceptibility of interrenal cells to lipid peroxidation along with enhanced ROS production by cytochrome P450 enzyme activities, it would be

selectively advantageous to impart a cellular defense mechanism against processes of oxidative stress that can lead to significant cellular damage or enzyme inhibition/inactivity.

In summary, this evidence suggests that endowing the OB, the putative interrenal cells of *C. gunnari*, with a cellular antioxidant system (such as the proline metabolism) might be critical for the functionality of this tissue.

The Orange Pigmentation of Orange Body

There are two hypotheses that might explain the distinctive pigmentation of OB. It may be that the heme group of CYP enzymes contributes to the orange color of OB tissue. Or it may be that OB accumulates lipofuscin pigments within its cells. Lipofuscin is an age-related pigment that results from free radical-induced lipid peroxidation in tissues including the kidney of a wide range of animals (Fletcher et al., 1973; Trombly et al., 1975; Csallany and Ayaz, 1976; Kikugawa et al., 1997). Moreover, lipofuscin accumulation has been documented to occur specifically in the interrenal gland of many teleost fishes and the presence of lipofuscin in the interrenal cells confers a distinct yellowish-brown color to this tissue (Oguri, 1986; Oguri and Hibiya, 1957). Hence, it is likely that the lipid peroxidation in the orange body, the putative interrenal tissue, is promoting the buildup of lipofuscin pigments and as a consequence, giving this gland its distinct orange-brown color range.

HK Functional Categorization Analysis

With respect to the *C. gunnari* HK gene expression profile, most of the enriched GO terms obtained from the 639 down-regulated DE list of genes (Supplementary Figure 1; Figure 4) were related to the innate and adaptive immune systems along with multiple cell proliferation processes concerning only leukocyte cell lines. Interestingly, we found no enriched GO terms related to red blood cells differentiation pathways consistent with studies that confirm the absence of erythrocytes (Ruud, 1954) and the non-expression condition of the hemoglobin protein in icefishes (Cocca et al., 1995). It is worth noting that among the enriched GO terms in HK were “response to steroid hormone” (GO:0048545, p-value=1.12E-4), and “response to corticosteroid” (GO:0031960, p-value=1.98E-4), which highlights the

paracrine relationship between HK and interrenal cells and their interconnectedness in maintaining organismal homeostasis. Although there are no transcriptomic studies done on HK teleost fish that explore the differential gene expression profile of HK and interrenal cells, to the best of our knowledge, there are some studies that examine the whole HK fish transcriptomics (Gerdol et al., 2015; Chen et al., 2012; Martinez et al., 2018) and we found that the HK functional enrichment results of *C. gunnari* obtained in this study is in line with the results reached from these studies. As it is reported in the literature that the head kidney in teleost fish encompasses diverse functions from hemopoiesis to lymphocyte and antibody production. This further supports that our RNA-Seq analysis results validate the conclusions reached with respect to the cryptic orange body.

CONCLUSIONS

We employed next generation RNA-Seq to test the hypothesis that the gene expression profile of the light orange body is similar to that of the interrenal cells by examining the differential transcriptome profile of the cryptic OB tissue relative to its surrounding head kidney tissue. The upregulated DE genes of OB tissue revealed the differential expression of key genes related to steroid hormone biosynthesis along with proline metabolism. These findings suggest that the gene expression profile of OB embedded in the head kidney of *C. gunnari* is similar to that of the adrenocortical cells or more specifically the interrenal cells in teleost fish.

Steroidogenesis is considered a critical function as it mediates important biological processes including development, growth, and reproduction as such it is imperative to understand the inner workings of steroidogenic tissues especially in icefish species given the many adaptations as well as the harsh conditions that they sustain during their life cycle. Furthermore, the lack of respiratory pigments in icefish species would allow the OB, the putative interrenal gland, to be highly visible compared to the surrounding HK tissue and therefore it would be straightforward to conduct comparative studies for the more derived and highly cold-specialized Antarctic icefish species. To further validate the results from this study, quantitative real-time qRT-PCR is recommended as this method is considered standard to validate RNA-Seq study results. To confirm that OB is the interrenal gland in icefish, steroidogenic acute regulatory protein (StAR) and cholesterol side chain cleavage enzyme (CYPscc) genes should be the transcripts under investigation since these two genes play roles that are considered rate-limiting steps in steroidogenesis.

On the one hand, interrenal cells house abundant cytochrome P450 enzymes, which are one of the principal enzymes involved in steroidogenesis, that are reported to generate ROS by electron transfer reactions. On the other hand, cholesterol is utilized as substrate for steroid biosynthesis thus interrenal cells contain lipid droplets with high content of PUFAs that are known to be more vulnerable to lipid peroxidation. Given the characteristics of interrenal cells of oxidative stress conditions, it would seem that antioxidant mechanisms within interrenal cells should be well represented to provide protective

mechanisms to counteract the damaging effects of reactive species. Based on the results from proline studies conducted in plants under oxidative stress conditions that demonstrate scavenging properties of ROS by proline, we hypothesize that the enrichment of proline biosynthesis in OB might relate to a form of protection against oxidative stress in this presumably ROS-inducible and lipid-rich tissue. This antioxidant defense mechanism has been extensively studied in plants but little or no study exist for vertebrates such as fish organisms. Furthermore, it would be pertinent to investigate the redox status of this tissue and its ROS concentration relative to the surrounding HK tissue.

Lastly, interrenal cells have not been investigated before in icefish or any other Antarctic notothenioid. To the best of our knowledge, this is the first study to show that the interrenal cells of teleost fishes, historically known to be scattered within the head kidney, might form a visible, discrete compact gland, such as the orange body in the icefishes. Additional histological examination and immunohistochemical verification of key adrenal steroidogenic enzymes in the icefish orange body will add confirmation to its identity as an interrenal tissue. The discovery of the interrenal orange body elicits further intriguing questions. How did the diffused interrenal tissue evolved into a conspicuous compact gland in the Antarctic icefish? Is this a unique condition in the white-blooded notothenioids, or a common condition across the Antarctic notothenioid radiation but has escaped detection in the red-blooded species because it is masked by the deep maroon color of the hemopoietic head kidney tissue. Such questions beg for investigation that may yet further reveal unique phenotypic outcomes in Antarctic notothenioids from evolution in isolation in the freezing South Ocean.

REFERENCES

Abdel-Aziz, E. S., Ali, T. E. S., Abdu, S. B. S., & Fouad, H. F. (2010). Chromaffin cells and interrenal tissue in the head kidney of the grouper, Epinephilus tauvina (Teleostei, Serranidae): a morphological (optical and ultrastructural) study. *Journal of Applied Ichthyology*, *26*(4), 522-527.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Burczynski, J. M., Southard, S. J., Hayes, J. R., Longhurst, P. A., & Colby, H. D. (2001). Changes in mitochondrial and microsomal lipid peroxidation and fatty acid profiles in adrenal glands, testes, and livers from α-tocopherol-deficient rats. *Free Radical Biology and Medicine*, *30*(9), 1029-1035.

Chen, C., & Dickman, M. B. (2005). Proline suppresses apoptosis in the fungal pathogen Colletotrichum trifolii. *Proceedings of the National Academy of Sciences*, *102*(9), 3459-3464.

Chen, T. W., Gan, R. C., Fang, Y. K., Chien, K. Y., Liao, W. C., Chen, C. C., ... & Yeh, Y. M. (2017). FunctionAnnotator, a versatile and efficient web tool for non-model organism annotation. *Scientific reports*, *7*(1), 10430.

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., & Gu, J. (2017). AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC bioinformatics*, *18*(3), 80.

Chen, J., Li, C., Huang, R., Du, F., Liao, L., Zhu, Z., & Wang, Y. (2012). Transcriptome analysis of head kidney in grass carp and discovery of immune-related genes. *BMC veterinary research*, *8*(1), 108.

Cocca, E., Ratnayake-Lecamwasam, M., Parker, S. K., Camardella, L., Ciaramella, M., di Prisco, G., & Detrich, H. W. (1995). Genomic remnants of alpha-globin genes in the hemoglobinless antarctic icefishes. *Proceedings of the National Academy of Sciences*, *92*(6), 1817-1821.

Collins, J. E., White, S., Searle, S. M., & Stemple, D. L. (2012). Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome research*, *22*(10), 2067-78.

Csallany, A. S., & Ayaz, K. L. (1976). Quantitative determination of organic solvent soluble lipofuscin pigments in tissues. *Lipids*, *11*(5), 412-417.

DeVries, A. L., & Cheng, C. H. C. (2005). Antifreeze proteins and organismal freezing avoidance in polar fishes. *Fish physiology*, *22*, 155-201.

Eastman, J.T. Polar Biol (2005) 28: 93. https://doi.org/10.1007/s00300-004-0667-4

Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., … Fendt, S. M. (2017). Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nature communications*, *8*, 15267. doi:10.1038/ncomms15267

Fierro-Castro, C., Santa-Cruz, M. C., Hernández-Sánchez, M., Teles, M., & Tort, L. (2015). Analysis of steroidogenic pathway key transcripts in interrenal cells isolated by laser microdissection (LMD) in stressed rainbow trout. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *190*, 39-46.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, *39*(suppl_2), W29-W37.

Fletcher, B. L., Dillard, C. J., & Tappel, A. L. (1973). Measurement of fluorescent lipid peroxidation products in biological systems and tissues. *Analytical biochemistry*, *52*(1), 1-9.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152.

Gerdol, M., Buonocore, F., Scapigliati, G., & Pallavicini, A. (2015). Analysis and characterization of the head kidney transcriptome from the Antarctic fish Trematomus bernacchii (Teleostea, Notothenioidea): a source for immune relevant genes. *Marine genomics*, *20*, 13-15.

Gupta, O. P., & Hanke, W. (1994). Regulation of interrenal secretion in the axolotl, Ambystoma mexicanum. *Experimental and Clinical Endocrinology & Diabetes*, *102*(04), 299-306.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, *8*(8), 1494.

Hanke, W., & Kloas, W. (1995). Comparative aspects of regulation and function of the adrenal complex in different groups of vertebrates. *Hormone and Metabolic Research*, *27*(09), 389-397.

Hanukoglu, I., & Hanukoglu, Z. (1986). Stoichiometry of mitochondrial cytochromes P-450, adrenodoxin and adrenodoxin reductase in adrenal cortex and corpus luteum: Implications for membrane organization and gene regulation. *European journal of biochemistry*, *157*(1), 27-31.

Hayward, D. C., Delaney, S. J., Campbell, H. D., Ghysen, A., Benzer, S., Kasprzak, A. B., ... & Miklos, G. L. (1993). The sluggish-A gene of Drosophila melanogaster is expressed in the nervous system and encodes proline oxidase, a mitochondrial enzyme involved in glutamate biosynthesis. *Proceedings of the National Academy of Sciences*, *90*(7), 2979-2983.

Kaul, S., Sharma, S. S., & Mehta, I. K. (2008). Free radical scavenging potential of L-proline: evidence from in vitro assays. *Amino acids*, *34*(2), 315-320.

Kikugawa, K., Beppu, M., Sato, A., & Kasai, H. (1997). Separation of multiple yellow fluorescent lipofuscin components in rat kidney and their characterization. *Mechanisms of ageing and development*, *97*(2), 93-107.

Kohl, D. H., Schubert, K. R., Carter, M. B., Hagedorn, C. H., & Shearer, G. (1988). Proline metabolism in N2-fixing root nodules: energy transfer and regulation of purine synthesis. *Proceedings of the National Academy of Sciences*, *85*(7), 2036-2040.

Kowaloff, E. M., Phang, J. M., Granger, A. S., & Downing, S. J. (1977). Regulation of proline oxidase activity by lactate. *Proceedings of the National Academy of Sciences*, *74*(12), 5368-5371.

Krishnan, N., Dickman, M. B., & Becker, D. F. (2008). Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. *Free Radical Biology and Medicine*, *44*(4), 671-681.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, *305*(3), 567-580.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, *35*(9), 3100-3108.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, *12*(1), 323.

Lin, D., Sugawara, T., Strauss, J. F., Clark, B. J., Stocco, D. M., Saenger, P., ... & Miller, W. L. (1995). Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science*, *267*(5205), 1828-1831.

Lin, C. H., Tsai, I. L., Su, C. H., Tseng, D. Y., & Hwang, P. P. (2011). Reverse effect of mammalian hypocalcemic cortisol in fish: cortisol stimulates Ca2+ uptake via glucocorticoid receptor-mediated vitamin D3 metabolism. *PloS one*, *6*(8), e23689.

Martínez, D., Pontigo, J., Morera, F., Yañéz, A., & Vargas-Chacoff, L. (2018). Head Kidney Transcriptome Analysis and Characterization for the Sub-Antarctic Notothenioid Fish Eleginops maclovinus. *Fishes*, *3*(1), 8.

Milano, E. G., Basari, F., & Chimenti, C. (1997). Adrenocortical and adrenomedullary homologs in eight species of adult and developing teleosts: morphology, histology, and immunohistochemistry. *General and comparative endocrinology*, *108*(3), 483-496.

Mohanty, P., & Matysik, J. (2001). Effect of proline on the production of singlet oxygen. *Amino acids*, *21*(2), 195-200.

Morandini, L., Honji, R. M., Ramallo, M. R., Moreira, R. G., & Pandolfi, M. (2014). The interrenal gland in males of the cichlid fish Cichlasoma dimerus: relationship with stress and the establishment of social hierarchies. *General and comparative endocrinology*, *195*, 88-98.

Naseeruddin, S. A., & Hornsby, P. J. (1990). Regulation of 11β-and 17α-hydroxylases in cultured bovine adrenocortical cells: 3′, 5′-cyclic adenosine monophosphate, insulin-like growth factor-I, and activators of protein kinase C. *Endocrinology*, *127*(4), 1673-1681.

Oguri, M. (1986). On the Yellowish Brown Pigment Granules in the Interrenal Cells of Anglerfish and Leopard Shark. *Fisheries*, *52*(5), 801-804.

Oguri, M., & Hibiya, T. (1957). Studies on the adrenal glands of teleosts—I. Some observations from the view point of comparative histology. *Bull. Jap. Soc. Sci. Fish*, *22*, 621-625.

Omura, T., Sanders, E., Estabrook, R. W., Cooper, D. Y., & Rosenthal, O. (1966). Isolation from adrenal cortex of a nonheme iron protein and a flavoprotein functional as a reduced triphosphopyridine nucleotide-cytochrome P-450 reductase. *Archives of Biochemistry and Biophysics*, *117*(3), 660-673.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, *14*(4), 417.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, *8*(10), 785.

Phang J. M. (2018). Proline Metabolism in Cell Regulation and Cancer Biology: Recent Advances and Hypotheses. *Antioxidants & redox signaling*, *30*(4), 635–649. doi:10.1089/ars.2017.7350

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., ... & Heger, A. (2011). The Pfam protein families database. *Nucleic acids research*, *40*(D1), D290-D301.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139-40.

Romano, N., Ceccariglia, S., Abelli, L., Mazzini, M., & Mastrolia, L. (2000). Lymphomyeloid organs of the Antarctic fish Trematomus nicolai and Chionodraco hamatus (Teleostei: Notothenioidea): a comparative histological study. *Polar Biology*, *23*(5), 321-328.

Rudolph, A. S., & Crowe, J. H. (1985). Membrane stabilization during freezing: the role of two natural cryoprotectants, trehalose and proline. *Cryobiology*, *22*(4), 367-377.

Ruud, J. T. (1954). Vertebrates without erythrocytes and blood pigment. *Nature*, *173*(4410), 848.

Samuel, D., Kumar, T. K. S., Ganesh, G., Jayaraman, G., Yang, P. W., Chang, M. M., ... & Yu, C. (2000). Proline inhibits aggregation during protein refolding. *Protein Science*, *9*(2), 344-352.

Sidell, B. D., Vayda, M. E., Small, D. J., Moylan, T. J., Londraville, R. L., Yuan, M. L., ... & Costello, L. (1997). Variable expression of myoglobin among the hemoglobinless Antarctic icefishes. *Proceedings of the National Academy of Sciences*, *94*(7), 3420-3424.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212.

Simard, J., Ricketts, M. L., Gingras, S., Soucy, P., Feltus, F. A., & Melner, M. H. (2005). Molecular biology of the 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase gene family. *Endocrine reviews*, *26*(4), 525-582.

Soneson C, Love MI and Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research* 2016, **4**:1521 (https://doi.org/10.12688/f1000research.7563.2)

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, *6*(7), e21800.

Tomlinson, J. W., Walker, E. A., Bujalska, I. J., Draper, N., Lavery, G. G., Cooper, M. S., ... & Stewart, P. M. (2004). 11β-hydroxysteroid dehydrogenase type 1: a tissue-specific regulator of glucocorticoid response. *Endocrine reviews*, *25*(5), 831-866.

Trombly, R., Tappel, A. L., Coniglio, J. G., Grogan, W. M., & Rhamy, R. K. (1975). Fluorescent products and polyunsaturated fatty acids of human testes. *Lipids*, *10*(10), 591-596.

Vahouny, G. V., Chanderbhan, R., Hodges, V. A., & Treadwell, C. R. (1978). Cholesterol arachidonate as a prostaglandin precursor in adrenocortical cells. *Prostaglandins*, *16*(2), 207-220.

Vermilion, J. L., Ballou, D. P., Massey, V., & Coon, M. J. (1981). Separate roles for FMN and FAD in catalysis by liver microsomal NADPH-cytochrome P-450 reductase. *Journal of Biological Chemistry*, *256*(1), 266-277.

Yamazaki, T. A. K. E. S. H. I., Higuchi, K. A. O. R. I., Kominami, S. H. I. R. O., & Takemori, S. (1996). 15-Lipoxygenase metabolite (s) of arachidonic acid mediates adrenocorticotropin action in bovine adrenal steroidogenesis. *Endocrinology*, *137*(7), 2670-2675.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2012). goseq: Gene Ontology testing for RNA-seq datasets. *R Bioconductor*.

APPENDIX A: SUPPLEMENTARY MATERIAL

Supplementary Table 1. *C. gunnari* RNA pool sequencing information from histologically distinct tissues and corresponding relevant details.

Tissue	Specimen	Qubit RNABR (ng/uL)	RIN	Concentration for pool (ug)
Head Kidney	Pool	195	9.1	1
Orange body	Pool	193	8.9	1
Caudal kidney	Cgun16	176	9.1	1
Heart	Cgun16	130	9.2	1
Liver	Cgun16	163	9.2	1
Spleen	Cgun16	154	9.5	1
Brain	Cgun16	107	8.4	1
Gills	Cgun16	208	9.5	1
Skin	Cgun16	144	9.2	1
Ovary	Cgun9	236	9.8	1
Testis	Cgun10	250	10	1
Eye lens	Cgun16	236	9.4	0.5
Eye cups	Cgun178	226	7.9	0.5
Esophagus & stomach	Cgun16	148	9.3	0.5
Anterior intestine	Cgun16	174	9.8	0.5
Pyloric ceca	Cgun174	208	9.9	0.5
Mesentery	Cgun178	185	9.8	0.5
Red muscle	Cgun16	224	9.0	0.5
White muscle	Cgun16	208	9.5	0.5
Pool	—	206	9.2	—

Supplementary Table 2. Statistics on RNA-Seq raw reads for the construction of *C. gunnari* de novo reference transcriptome.

	Total reads	%GC	Sequence length	Filtered out reads
Raw reads				
Forward	99929740	47	35-250	—
Reverse				
Post-filtration and Quality Control				
Forward	96597757	47	25-240	3331983 (3.33%)
Reverse				

Supplementary Table 3. *C. gunnari* de novo reference transcriptome assembly statistics with Trinity assembler and CD-HIT-EST.

Trinity Statistics:		
Total trinity 'genes'	307384	
Total trinity transcripts	469524	
Percent GC	45.07	
Contig N10	6446	
Contig N20	4819	
Contig N30	3769	
Contig N40	2966	
Contig N50	2254	
Median contig length	475	
Average contig	1037.37	
Total assembled bases	487070261	
Filtration Steps:		
	Number of reads before filtration	Number of reads after filtration
> 0.3 TPM*	469524	251560
95% similarity†	251560	208925

* Transcripts with expression values above 0.3 TPM were retained in the assembly.
† Transcript redundancy was reduced with 95% similarity threshold using CD-HIT-EST (v.4.6.8)

Supplementary Table 4. Quality and completeness assessment of *C. gunnari* de novo reference transcriptome with BUSCO notation results.

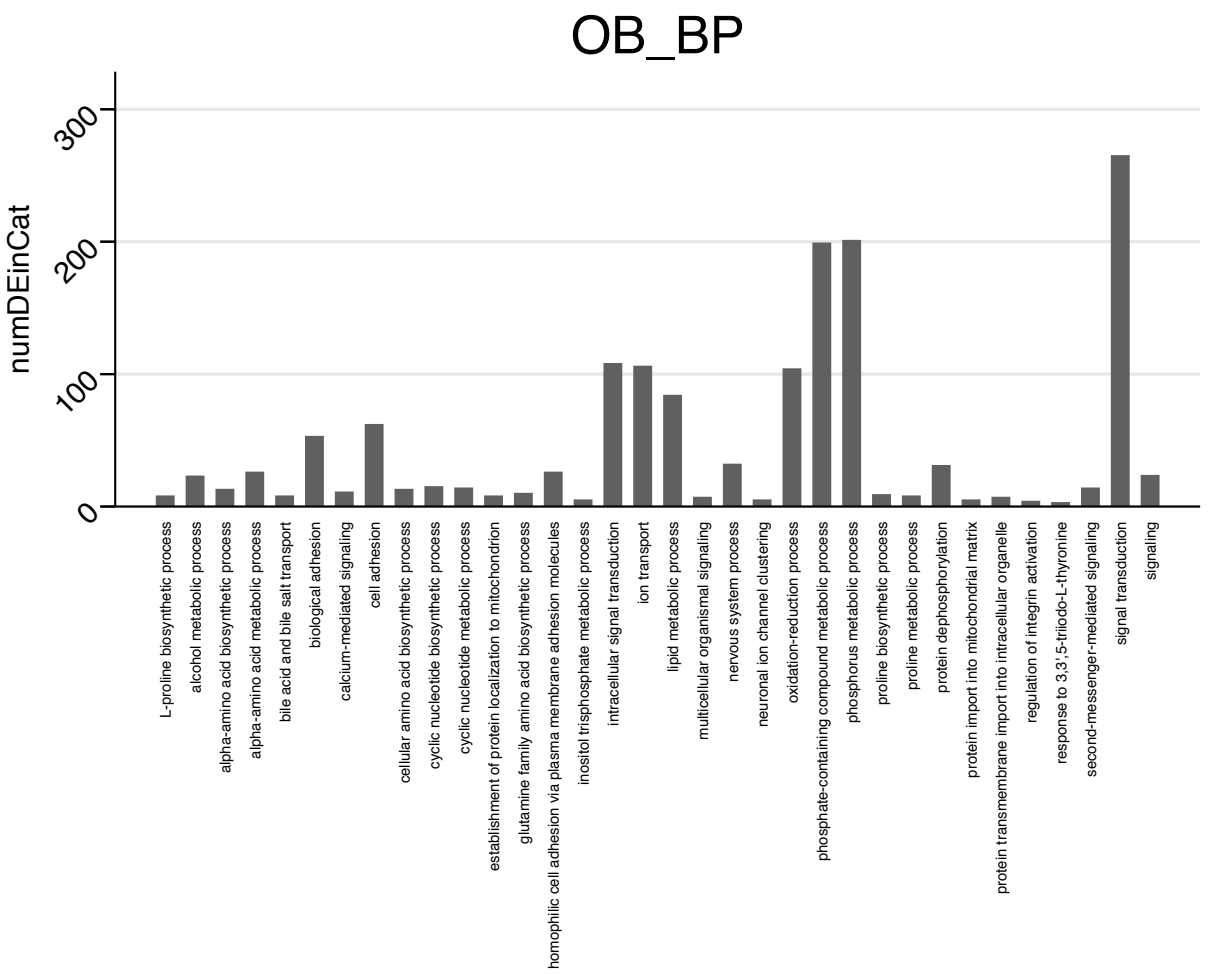
Reference data set	COMPLETE	SINGLE-COPY	DUPLICATED	FRAGMENTED	MISSING
ACTINOPTERYGII*	4,252 (92.8%)	2,824 (61.6%)	1,428 (31.2%)	135 (2.9%)	197 (4.3%)
METAZOA†	972 (99.4%)	573 (58.6%)	399 (40.8%)	3 (0.3%)	3 (0.3%)

* Out of 4,584 highly conserved actinopterygian genes.
† Out of 978 highly conserved metazoan genes.

Supplementary Table 5. Total numbers of experimental RNA-Seq reads before and after performing quality control with AFTERQC.

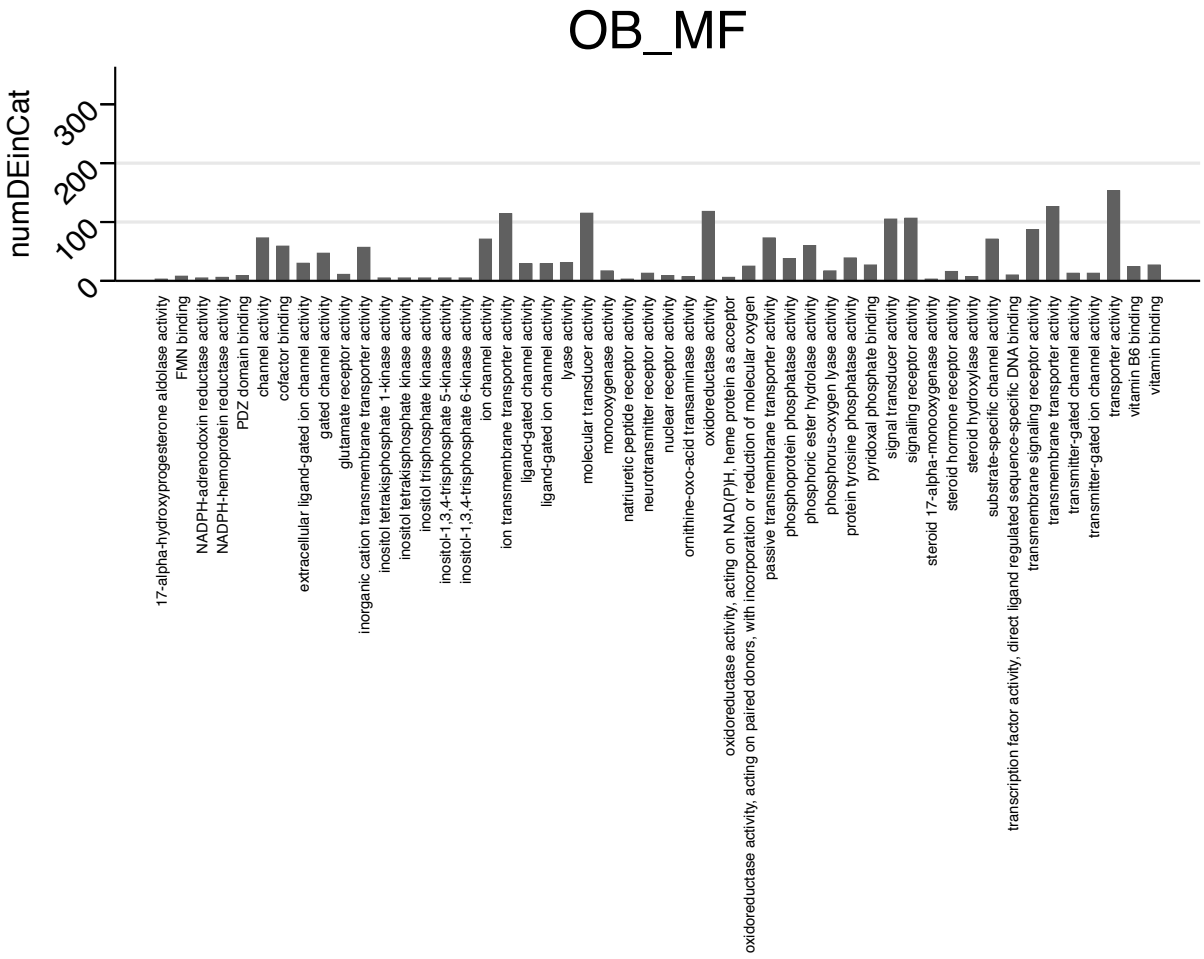
Sample	Total Reads	Filtered Out Reads	% Filtered Out Reads
Head Kidney			
10.HK	23.624 M	76,793	0.310
16.HK	23.449 M	82,355	0.335
178.HK	23.890 M	61,962	0.247
9.HK	25.786 M	65,041	0.241
Orange Body			
10.OB	24.614 M	72,260	0.280
16.OB	25.546 M	68,742	0.257
178.OB	24.466 M	54,551	0.213
9.OB	23.480 M	78,295	0.318

(A)



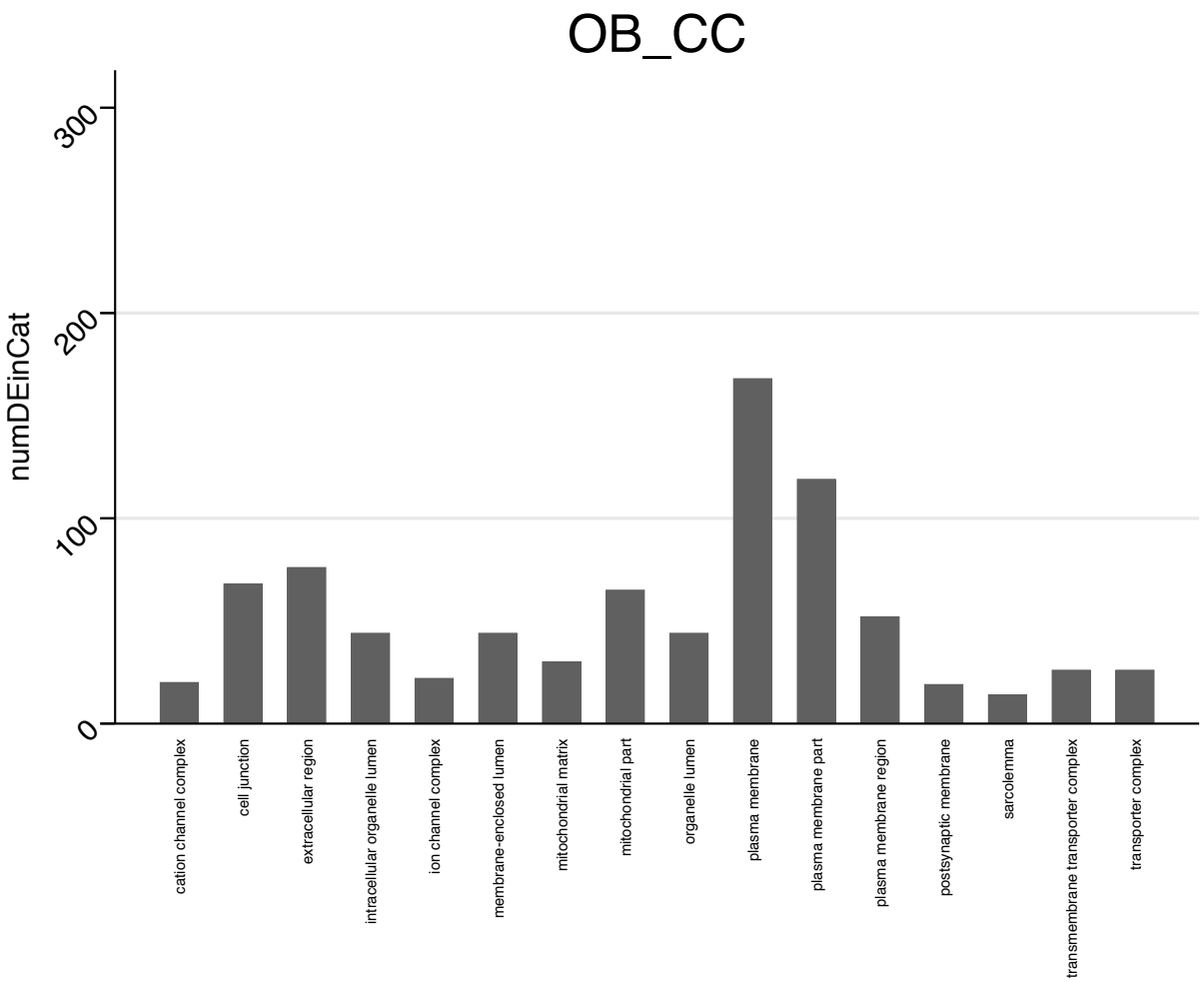
Supplementary Figure 1. Enriched Gene Ontology terms for OB and HK tissue of *C. gunnari.* Full list of OB enriched GO terms for the categories of (A) BP, (B) MF, and (C) CC respectively. Full list of HK enriched GO terms for the categories of (D) BP, (E) MF, and (F) CC respectively. Y-axis corresponds to the number of DE genes that fall into the corresponding GO term category.

(B)



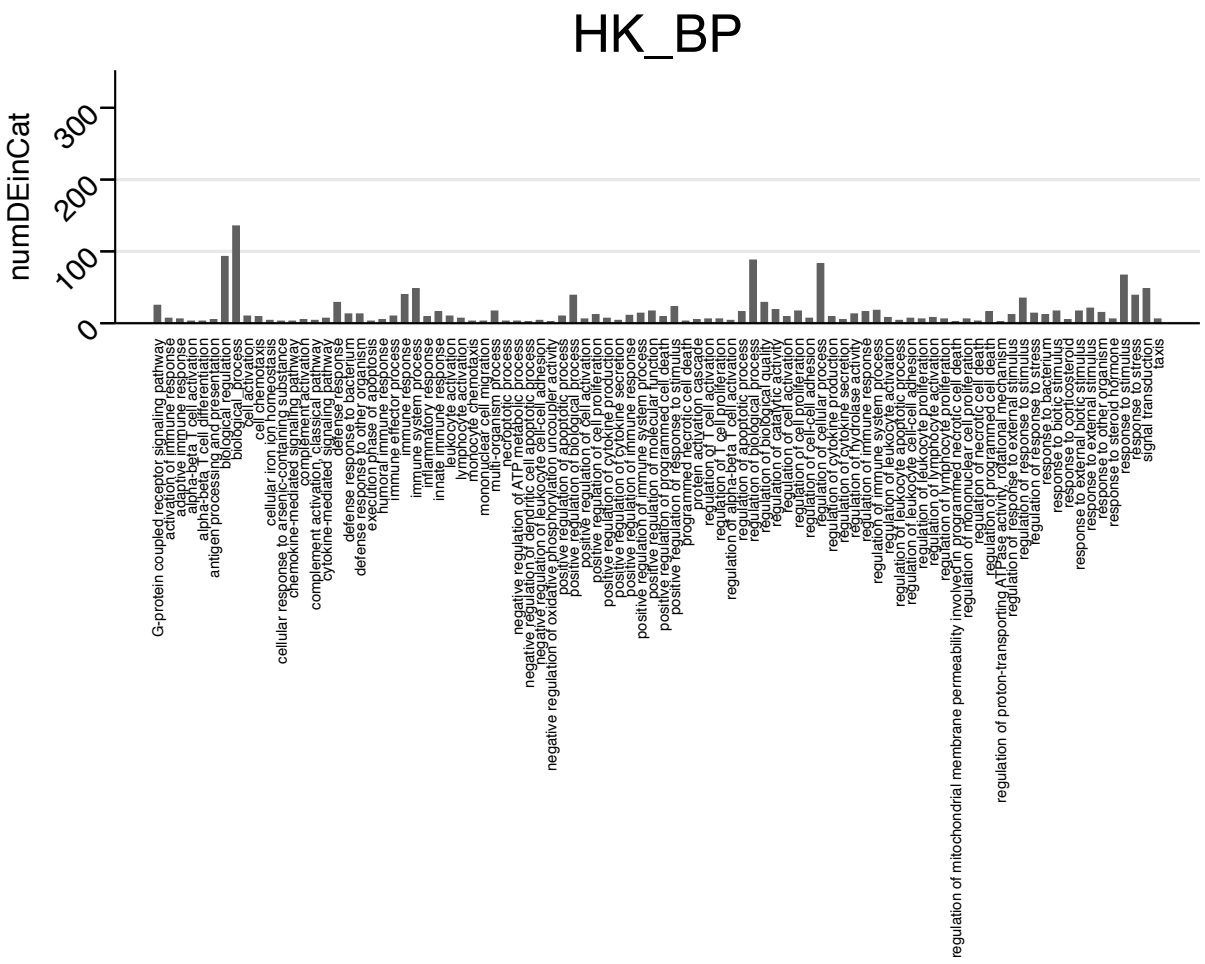
Supplementary Figure 1. (continued)

(C)



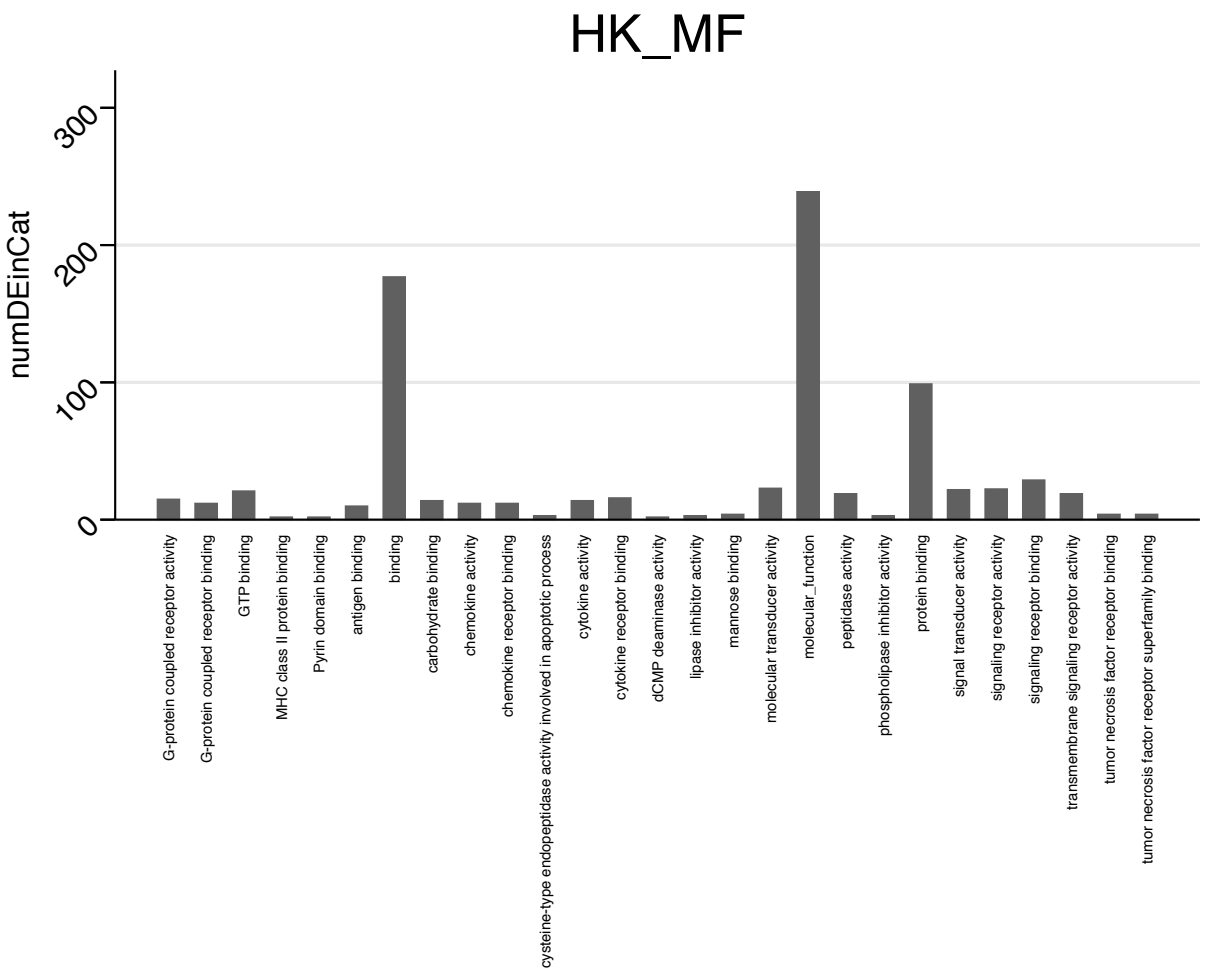
Supplementary Figure 1. (continued)

(D)



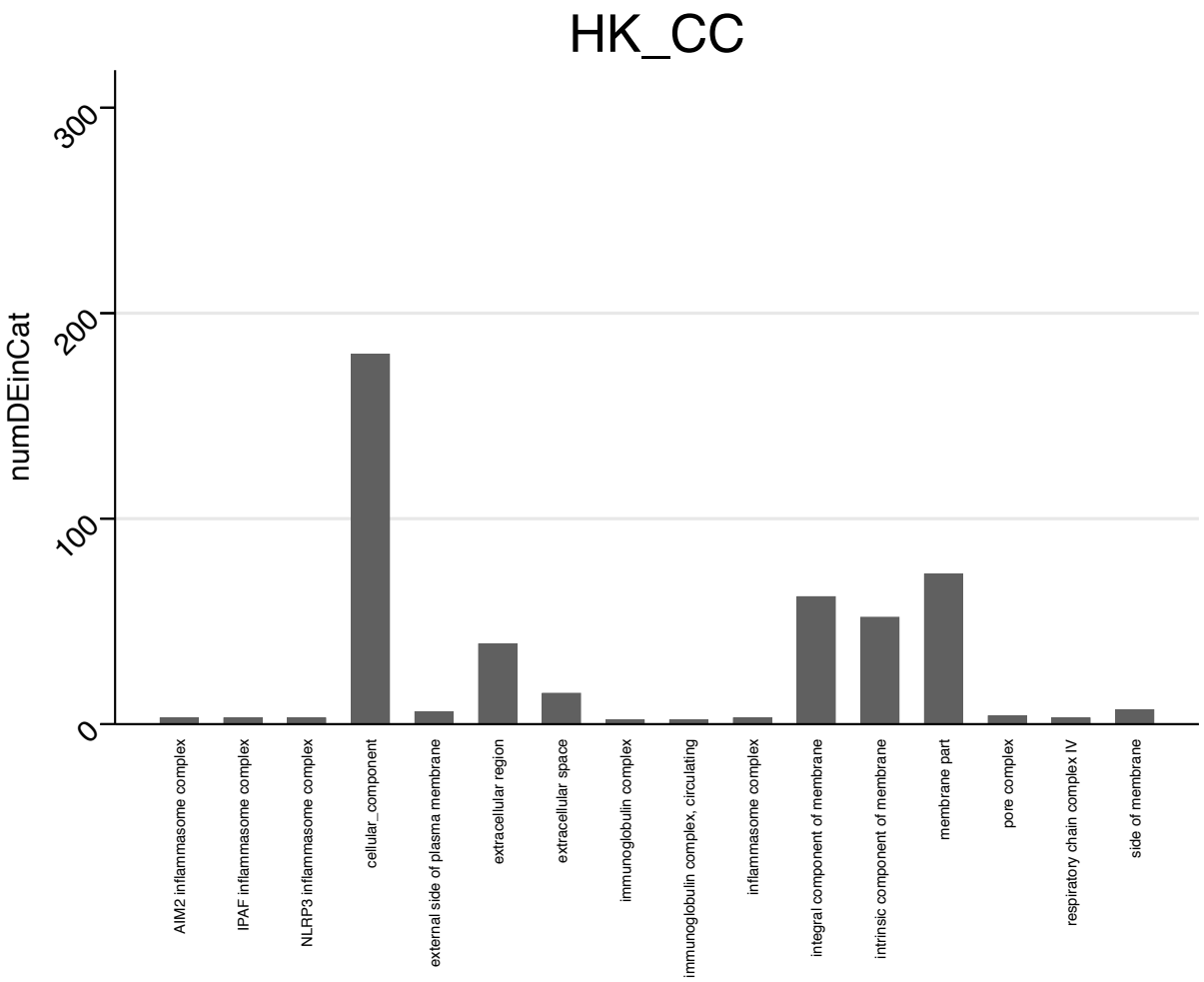
Supplementary Figure 1. (continued)

(E)



Supplementary Figure 1. (continued)

(F)



Supplementary Figure 1. (continued)