De novo transcriptome assembly and RNA-Seq expression analysis in blood from beluga whales of Bristol Bay, AK

Jeanine S Morey[1,2*], Kathy A Burek Huntington[3], Michelle Campbell[4], Tonya M Clauss[5], Caroline E Goertz[6], Roderick C Hobbs[7], Denise Lunardi[8], Amanda J Moors[9], Marion G Neely[1,2], Lori H Schwacke[1†], Frances M Van Dolah[1‡]

[1]Hollings Marine Laboratory, National Centers for Coastal Ocean Science, National Ocean Service, NOAA, 331 Fort Johnson Road, Charleston, SC 29412, USA

[2]Jardon and Howard Technologies Incorporated, 2710 Discovery Drive, Orlando, FL 32826, USA

[3]Alaska Veterinary Pathology Service, 23834 The Clearing Drive, Eagle River, AK 99577, USA

[4]Dolphin Quest, 425 Waikoloa Beach Drive, Waikoloa, HI 96738, USA

[5]Georgia Aquarium, 225 Baker Street, Atlanta, GA 30313, USA

[6]Alaska SeaLife Center, PO Box 1329, Seward, AK 99664, USA

[7]National Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way N.E., Seattle, WA 95115, USA

[8] Department of Life Sciences and Biotechnology, University of Ferrara, via L. Borsari 46, 44121 Ferrara, Italy

[9]Hollings Marine Laboratory, National Institute of Standards and Technology, 331 Fort Johnson Road, Charleston, SC 29412, USA

[*]Corresponding Author: Jeanine.morey@gmail.com

[†]Current Affiliation: National Marine Mammal Foundation, 2240 Shelter Island Drive, Suite 200, San Diego, CA 92106, USA

[‡]Current Affiliation: Department of Biology, School of Science and Mathematics, College of Charleston, 66 George Street, Charleston, SC, 29424, USA

Abstract

Assessing the health of marine mammal sentinel species is crucial to understanding the impacts of environmental perturbations on marine ecosystems and human health. In arctic regions, beluga whales, *Delphinapterus leucas,* are upper level predators that may serve as a sentinel species, potentially forecasting impacts on human health. While gene expression profiling from blood transcriptomes has widely been used to assess health status and environmental exposures in human and veterinary medicine, its use in wildlife has been limited due to the lack of available genomes and baseline data. To this end we constructed the first beluga whale blood transcriptome *de novo* from samples collected during annual health assessments of the healthy Bristol Bay, AK stock during 2012-2014 to establish baseline information on the content and variation of the beluga whale blood transcriptome. The Trinity transcriptome assembly from beluga was comprised of 91,325 transcripts that represented a wide array of cellular functions and processes and was extremely similar in content to the blood transcriptome of another cetacean, the bottlenose dolphin. Expression of hemoglobin transcripts was much lower in beluga (25.6% of TPM, transcripts per million) than has been observed in many other mammals. A T12A amino acid substitution in the HBB sequence of beluga whales, but not bottlenose dolphins, was identified and may play a role in low temperature adaptation. The beluga blood transcriptome was extremely stable between sex and year, with no apparent clustering of samples by principle components analysis and < 4% of genes differentially expressed (EBseq, FDR < 0.05). While the impacts of season, sexual maturity, disease, and geography on the beluga blood transcriptome must be established, the presence of transcripts involved in stress, detoxification, and immune functions indicate that blood gene expression analyses may provide information on health status and exposure. This study provides a wealth of transcriptomic data on beluga whales and provides a sizeable pool of preliminary data for comparison with other studies in beluga whale.

Keywords

Beluga whale, blood transcriptome, common bottlenose dolphin, hemoglobin, RNA-Seq

Introduction

Oceans cover more than 70% of the Earth's surface and play an essential role in climate, habitat, and food chains, thus the impacts of oceanic health are far-reaching. Assessing the health of sentinel species is critical to understanding the impacts of environmental perturbations, whether natural or man-made, on ecosystems, animals, and, ultimately, human health. In particular, the use of marine

mammals as sentinels may serve as an early warning system for human impacts given the similarities in diet and physiology. As marine mammals are constantly exposed to the marine environment, they may exhibit signs of chronic exposure or latent/slow-developing pathologies that are difficult to detect in human populations with less exposure. Alterations at the cellular and molecular level are highly sensitive indicators of environmental change and precede impacts at higher levels. However, as has been recently noted, substantial baseline data and tool development is needed for these species in order to fully understand the impacts of environmental changes before marine mammals can fulfill the role of sentinel species [1-4].

Hematology and serum chemistry are widely used in assessing animal health, including many cetaceans such as beluga, bowhead, fin, minke, and killer whales, common bottlenose dolphins, and harbor porpoise [reviewed in 2, 5]. The central role blood plays in physiological homeostasis, immunity, transport and signaling, combined with its ability to reflect the status of organs throughout the body from molecular messages obtained during circulation, provides a unique measure of a broad range of transcripts by high-throughput gene expression profiling. In human and veterinary medicine blood transcriptomes have been used to identify health status, disease, and exposures to environmental toxicants [6, 7]. However, blood transcriptomic analysis can be confounded by seasonal and geographical variabilities [8-11]. Thus, application in wildlife species has been limited by the lack of baseline data needed to appropriately assess changes in blood transcriptomes, limited genomic data, and sample availability. However, the body of literature continues to grow with numerous reports of transcriptomic responses and genomic data in marine mammals, including several cetaceans [12-20], pinnipeds [21-28], and fissipeds [29]. Recently, the blood transcriptomes of managed bottlenose dolphins housed in Hawaii were assessed and found to be highly stable, with correlations to hematological parameters and few seasonal effects [1].

Beluga whales, *Delphinapterus leucas,* are an upper level predator that may serve as an important sentinel species. The Arctic environment of beluga whales, while relatively isolated, is reported to be a substantial sink for persistent organic pollutants (POPs) and mercury (Hg) [30, 31] and substantial levels have been measured in belugas [32]. Increased transcript levels of classic detoxification genes were positively correlated with POPs concentrations in blubber and liver samples collected as part of beluga subsistence hunts, while alterations in transcripts associated with growth, development, and metabolism were also observed [18]. These results indicate that transcriptomic analyses may be useful in assessing exposure and health in beluga whales. However, development of

robust tools for assessment and monitoring will first require the establishment of baseline data from healthy animals.

In Alaska, there are 5 stocks of beluga whales: Cook Inlet, eastern Chukchi Sea, Bristol Bay, eastern Bering Sea, and Beaufort Sea [33, 34]. The Cook Inlet stock is genetically isolated and geographically isolated from the remaining stocks by the Alaskan Peninsula and was placed on the Endangered Species List in 2008, with a current population of approximately 300 individuals that is declining 1.6% per year [35]. These Cook Inlet animals are thought to be exposed to significant anthropogenic contamination due to their proximity to the highly urbanized Anchorage area. In contrast, the belugas of the Bristol Bay stock number nearly 3000 and are growing at an estimated rate of 4.5% per year while inhabiting a similar ecological niche [35]. Relative to surrounding areas, Bristol Bay is relatively uncontaminated and has been previously proposed as a healthy reference site in many studies including those on beluga hematology [2] and hearing [36]. There is also growing concern that the health of Bristol Bay may decline with increasing industrialization and urbanization introducing environmental changes detrimental to beluga populations [37].

In this study we sought to establish the beluga whale blood transcriptome with samples collected during health assessments over a 3 year period (2012-2014) from the healthy population in Bristol Bay, AK. As no genome is available for beluga whale we assembled the transcriptome *de novo* using Trinity, annotated the assembly, and analyzed gene expression. For direct comparison with another cetacean for which more transcriptomic data is available, a bottlenose dolphin blood transcriptome was also assembled *de novo* by Trinity. While differential expression analysis was limited by the available samples, the results presented here provide crucial baseline data for a critical Arctic species.

Methods

Sample Collection

Beluga Whale

Blood samples were collected as part of annual health assessments of free-ranging *D. leucas* in Bristol Bay, Alaska during August and September of 2012-2014. Field methods were previously described in detail by Norman et al. [2]. Briefly, belugas were captured in a net, held temporarily for examination during which time blood was drawn into PAXgene (Qiagen, Valencia, CA) tubes from the dorsal side of the flukes, then released. All work was conducted under National Marine Fisheries Service

NMML Cetacean permit no. 14245 and in accordance with approval from the Marine Mammal Laboratory of the Alaska Fisheries Science Center (MML/AFSC) IACUC protocols (ID number: AFSC-NWFSC2012-1). Blood tubes were stored at -80°C until RNA extraction. Eight samples from each year were extracted for RNA, comprised of 3 females and 5 males from 2012, 4 females and 4 males from 2013, and 2 females and 6 males from 2014 (Table 1).

Bottlenose Dolphin

Blood samples were collected in PAXgene (Qiagen, Valencia, CA) tubes from the ventral side of the flukes of three managed *T. truncatus* residing at Dolphin Quest, Waikoloa, Hawaii during 2013 as part of a larger study [1]. All research was approved by the Dolphin Quest Scientific Committee and carried out according to standards and guidelines of the AMMPA (Alliance of Marine Mammal Parks and Aquariums). The dolphins sampled for this study included one male, age 5 (83H1, n=1), and two females, ages 12 (8SK7, n=2) and 28 (1FP3, n=2) (Table 2). All animals are trained to participate in monthly veterinary checkups including routine blood draws, which were conducted in the mornings after overnight fasting. Samples were collected from healthy animals; defined as bright, alert, responsive (BAR) animals demonstrating baseline behavior and appetite and blood chemistry within normal ranges. Blood tubes were stored at -80°C until extraction of total RNA.

RNA Extraction

Whole blood RNA was extracted using a PAXgene Blood RNA Kit (Qiagen, Valencia, CA), according to the manufacturer's protocol with on-column DNase digestion to remove contaminating DNA. RNA concentrations were evaluated using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and RNA quality was assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Santa Clara, CA). Only samples with a RIN (RNA Integrity Number) ≥ 7 were sequenced.

Reverse Transcription and qPCR

RNA-seq analysis of a single test sample (DLBB13-03, data not shown) indicated much lower expression levels of hemoglobin transcripts in beluga blood than previously reported in dolphin blood [1]. Therefore, the relative levels of hemoglobin mRNA expression (HBA and HBB) in all beluga blood samples were assessed by quantitative real-time PCR prior to sequencing. Gene specific primers (400 nM) were used for duplicate qPCR reactions on an ABI 7500 using ABI Power SYBR Green master mix

(Applied Biosystems, Foster City, CA) with the following thermocycling parameters: 10 min at 95° C, 40 cycles of 15 s at 94° C, 40 s at 60° C and 30 s at 72° C, followed with the standard dissociation cycle. The primers used were as follows: HBAF-5'-GGCCTCTGCGCCATATT-3', HBAR-5'-CCAGACTCAAAGAGAACTCACC-3', HBBF-5'-TGCATGTGGATCCTGAGAAC-3', HBBR-5'-GGTGAATTCCTTGCCAAAGTG-3'. The specificity of qPCR primers and the size of the amplicon were verified by analysis with an Agilent Bioanalyzer 2100 and further confirmed by the presence of a single peak in the melt curve (dissociation) analysis. One hundred nanograms of total RNA was reverse transcribed with EpiScript-RNase H Reverse Transcriptase (Epicentre, Madision, WI) and oligo(dT) priming and serial dilutions were used to construct a standard curve. The reaction efficiency was determined using the slope of this standard curve of cDNA from total RNA (% efficiency = ($10^{-1/slope}$-1)*100). Using the described conditions, both HBA and HBB produced a single qPCR product with similar and acceptable reactions efficiencies (HBA = 94.5%, HBB = 94.0%) and were used as a preliminary screening tool to assess hemoglobin transcript expression in beluga blood. For all samples, 50 ng of total RNA was reverse transcribed as above for duplicate qPCR reactions normalized to input RNA. A cycle threshold (C_t) was assigned at the beginning of the logarithmic phase of PCR amplification and the difference in the C_t values between DLBB13-03 and other samples was used to determine the relative expression of the gene in each sample (fold change = $2^{\triangle Ct}$).

Sequencing

Total RNA samples were sent to North Carolina State University Genomics Service Laboratory for library preparation using a NEBNext Ultra Directional RNA Library Prep Kit for Illumina and indexed with the NEBNext Mulitplex Oligos for Illumina (New England Biolabs, Ipswich, MA). Prior to sequencing of all samples used in this study, a single sample (DLBB13-03, data not shown) was sequenced at a targeted depth of 195M reads to assess the appropriate depth of sequencing for the beluga blood transcriptome. Sequencing was performed on an Illumina HiSeq 2500 sequencer (Illumina, San Diego, CA), at a targeted depth of 37M (beluga, n=24) or 45M (dolphin, n=5) 100 nt single reads. Bottlenose dolphin samples were sequenced at a greater depth as an initial analysis of preliminary samples indicated a much higher prevalence of hemoglobin transcripts in dolphin compared to beluga blood (data not shown).

***de novo* Transcriptome Assembly and Analysis**

Sequence processing and analysis was carried out in iPlant Collaborative's Discovery Environment using the High-Performance Computing applications [38]. The Illumina BCL output files

were converted to FASTQ-sanger file format and sequence quality trimming was performed using Trimmomatic [39], with a minimum phred quality score >20 over the length of the reads. The trimmed reads were then quality checked using the FASTQC tool. The processed and trimmed reads were used to construct a *de novo* transcriptome using the Trinity assembler v2.0.6 [40] on CyVerse/iPlant Collaborative's Atmosphere cloud computing platform for both beluga whale and bottlenose dolphin. The read files from one female and one male beluga whale from 2013 (DLBB13-07 and DLBB13-02) were concatenated into a single fastq file for assembly using a minimum K-mer coverage of 3, a minimum overlap value of 25 and a minimum contig length of 400 nucleotides (nt). The same assembly parameters were used for the concatenated reads from one female and one male dolphin (8SK7 July and 83H1 June). The assembly completeness for each species was assessed by mapping a set of highly conserved genes using CEGMA [41] and the vertebrate set in transcriptome mode of BUSCO [42]. The transcriptomes were annotated using BLAST+ for blastx searches (E-value ≤ 1e-4) of the human subset of the UniProt-SwissProt database (downloaded 10Jun2016), followed by conserved domain mapping and gene ontology assignment using Blast2GO [43-46][38]. Trimmed and groomed reads from individual animals were mapped to their respective *de novo* Trinity assembly using RSEM v 1.2.18 [47] with Bowtie2 v 2.2.4 [48] as the alignment engine and mapped read counts, as TPM (transcripts per million), were generated in the Atmosphere environment. Differential expression analyses were performed in EBSeq [49] using an FDR of 0.05. Gene enrichment analysis and pathway mapping of the differentially expressed gene sets was analyzed using Fishers Exact test in Blast2GO [43-46] (FDR < 0.05) and pathway mapping with the hypergeometric test for enrichment evaluation in WebGestalt [50, 51] (Benjamini & Hochberg adjusted p-value < 0.05) using a background comprised of all genes expressed in blood with an average TPM ≥ 1 across all samples and TPM > 0 in at least half of the samples. Visualization of pathway mapping was completed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) Mapper Search&Color Pathway tool [52]. Principal component analysis was performed on $\log_2$ transformed TPM values for all genes in the background set. PCA was performed using the prcomp package from the stats library in RStudio (v 1.0.136) and visualized using ggplot2 (v 2.2.1) [53]. The Trinity assemblies, raw reads, summarized TPMs, and differential expression results are available on GEO (beluga GEO accession # GSE98735, dolphin GEO accession # GSE98627).

Sequence Alignment

Amino acid sequences for HBB were obtained from NCBI or translated from sequence data generated here. Multiple sequence alignments were carried out in BioEdit (v 7.2.5) [54] using the

Clustal W accessory (v 1.4) [55]. The ClustalW default parameters within BioEdit were used to construct a neighbor joining tree with 1000 bootstrap replicates for the alignment.

Results and Discussion

de novo assembly and annotation of cetacean blood transcriptomes

To establish baseline molecular data from healthy animals, a *de novo* blood transcriptome for beluga whales of the Bristol Bay, AK stock was assembled and annotated. We have previously described the blood transcriptome of the bottlenose dolphin and observed that approximately 75% of reads from total RNA represented hemoglobin sequences [1]. This preponderance of hemoglobin sequences could be overcome by either globin-reduction of total RNA prior to sequencing or increased depth of sequencing of non-globin reduced total RNA, allowing for sufficient detection of dolphin blood transcripts. As there was no data available on globin transcript expression in beluga blood, we sought to determine this prior to sequencing all samples in the study to ensure sufficient recovery of transcripts, either through globin-reduction of total RNA or increased sequencing depth. Total RNA from a single sample was sequenced and only 15% of reads were mapped to hemoglobin genes (data not shown), significantly less than reported in dolphin. To ensure that this seemingly low expression of hemoglobin transcripts was not sample-specific, all samples in the study were subjected to preliminary analysis by qPCR prior to sequencing. Indeed, most samples expressed HBA and HBB at levels similar to, or lower than, that of DLBB13-03 (Figure 1). Therefore, sequencing of all samples used in this study was carried out on non-globin-reduced total RNA at a depth of 37M reads (beluga) or 45M reads (dolphin) to ensure adequate representation of transcripts expressed in blood of these cetaceans. Currently there is no genome available for the beluga whale and the genome available from Ensembl for the bottlenose dolphin, at the time of this analysis, is of low coverage (2.59X) and not fully annotated. While we have previously described the blood transcriptome of the bottlenose dolphin [1], the globin-reduced samples used in that study are not directly comparable to the non-reduced beluga samples from this study. Therefore, we constructed *de novo* transcriptome assemblies for beluga and dolphin using Trinity with reads from non-globin reduced whole blood samples generated by this study for all analyses.

Beluga whale

The 79,897,657 reads from two beluga whales were assembled by Trinity into 91,325 transcripts assigned to a total of 54,582 genes (Table 3). The majority of beluga genes, 84.3%, are represented by a single isoform. The N50 of the beluga assembly was 2588 nt, with transcripts ranging in length from 424

to 23,817 nt. Overall, 87.5% of reads from samples in this study mapped back to this *de novo* assembly of beluga blood transcripts. The beluga assembly encompasses the breadth of core eukaryotic genes (CEGs) with 94.8% of full length CEGs identified by CEGMA. This increases to 98.8% when partial length CEGS are included. When BUSCO (Benchmarking Universal Single-Copy Orthologs) is applied to the *de novo* assembly, 72% of single-copy orthologs from the vertebrate set were identified in full length, with an additional 5.8% fragmented BUSCOs identified (Table 3). Nearly 33% of BUSCOs were identified in duplicate in the beluga blood transcriptome. However, we have not determined if this truly reflects gene duplication, rather it is more likely indicative of natural sequence variability or ambiguities in the assembly that may be resolved with increased sequencing depth or computational manipulation of assembly parameters. The BUSCO results reported here fall well within the range of 4.4-89% complete BUSCOS identified in vertebrate transcriptomes reported by Simão et al. [42] and is much higher than the 31% identified in greenfinch blood, the only vertebrate blood transcriptome included in their report. Together, the CEGMA and BUSCO data provide evidence that the *de novo* Trinity assembly of beluga whale produced contains full-length, or near full-length, transcripts representative of the full blood transcriptome. Blastx searches of the human subset of the UniProt-SwissProt database were followed by InterProScan searches and Gene Ontology (GO) term mapping in Blast2GO for annotation of the *de novo* transcriptome. Fifty-eight percent of the beluga transcriptome returned blast hits (E-value ≤ 1e-4, Table 3) and 50% was ultimately fully annotated in Blast2GO.

Bottlenose dolphin

The 104,708,450 reads from two dolphins were assembled by Trinity into 56,367 transcripts assigned to a total of 36,596 genes (Table 3), only slightly larger than the 50,000 transcripts reported for a *de novo* Trinity assembly of globin-reduced bottlenose dolphin blood transcriptomes [1], but much smaller than the beluga whale transcriptome assembled here. Overall, the assembly statistics were similar to that of the beluga whale assembly, however the dolphin transcriptome was assembled into somewhat shorter transcripts which is reflected in a lower percentage of identified full lengths BUSCOS (60%, Table 3). Likewise, these dolphin assembly attributes described above are similar to the *de novo* transcriptome assembled from hemoglobin-reduced RNA samples collected in parallel with those utilized here [1]. The assembly presented herein of non-hemoglobin-reduced dolphin blood samples did produce a majority of longer transcripts in comparison to the globin-reduced assembly (N50 = 2293 and N50 = 1331, respectively, [1]) and is reflected in the increased discovery of full-length CEGs here (93.6%

v 87.5%). The increased transcript length may have also contributed to a substantial increase in sequence annotation (54.5% v 38%, Table 3, [1]).

Hemoglobin transcript expression

As expected in a blood transcriptome, the three highest expressed genes in beluga whale mapped to HBB (13.7% of TPM) and HBA2 (6.3 and 5.6% of TPM, 2 separate HBA2 transcripts were produced in the *de novo* assembly). Expression levels of HBA and HBB were approximately equal, as needed for efficient production of functional hemoglobin protein [56]. In addition, there were four other genes annotated as hemoglobins (HBA2, HBB, HBM, and HBQ1), but these only accounted for an additional 0.06% of TPMs in beluga. Overall 25.6% of total TPM in beluga mapped to genes annotated as hemoglobins. This is strikingly less than the 67.3% of total TPM mapped to hemoglobins in dolphin in the current dolphin transcriptome (HBA 32.9%, HBB 34.2%, HBM, HBQ, HBZ) or the 46% to 76% observed in porcine and human blood, respectively [57]. The decreased hemoglobin expression may be impacted by season as lower expression levels of HBA2 and HBB transcripts in warmer months have been observed in humans [9]. While the non-globin reduced dolphin samples described here are primarily from warmer months (April to September), precluding seasonal analysis, an examination of previously published data from globin-reduced dolphin blood [1] identified a trend toward lower HBB expression in warmer months relative to cooler months (Figure 2), although this was not found to be a significant expression change in the previous study. Thus, as all samples from beluga were collected during August and September, during the warmest months and in shallow water estuaries, an examination of seasonal effects on globin transcription in beluga blood is needed to determine if hemoglobin transcript expression is always lower in this species or if it is impacted by season resulting in fewer reads mapping to hemoglobin than reported in other species.

In both the beluga and dolphin transcriptomes, two genes were identified as HBB. When translated, the amino acid sequences are identical within species and vary by only a single amino acid between species (Figure 3). The single amino acid that varies is located at position 12 and is an alanine in beluga and a threonine in dolphin. Three amino acid substitutions in HBB/D have been identified in woolly mammoth that may have played an important role in cold adaptation in the elephantid lineage, T12A, A86S and E101Q [58]. While the HBB/D chimera has not been identified in the Boreoeutheria lineage, there is substantial sequence homology between the wooly mammoth chimera and the beluga and dolphin HBB genes identified here (Figure 3). It is interesting to note that the HBB sequences identified here in beluga whale contain the T12A substitution, whereas the bottlenose dolphin

sequences do not (Figure 3). Likewise, the T12A substitution was not observed in HBB sequences from other cold adapted species including polar bear, orca, minke whale, Weddell seal, and Pacific walrus. HBB sequence data from the narwhal, the closest living relative to belugas, was not available in NCBI databases, thus it is unknown if the T12A substitution is unique to belugas or is conserved among cold-adapted Monodontidae. The T12A substitution in woolly mammoth reduces the oxygenation enthalpy of hemoglobin by enhancing the binding of red-cell ligand 2,3-bisphosphoglycerate which, in turn, promotes the release of oxygen bound to hemoglobin [58]. Thus, the ability of red blood cells to release oxygen near sparsely insulated regions, such as the flippers and flukes, is enhanced. The enhanced oxygen offloading from the T12A substitution may be present in belugas, but not other cold adapted species, to offset the reduction in oxygen release due to the extremely high whole blood viscosity observed in belugas, but not other marine mammals, including orcas [59]. Further research is needed to determine the role of the HBB T12A substitution in cold adaptation of beluga whales. Many other adaptations to enhance oxygen flow in belugas are well known, including high 2,3-disphosphoglycerate [60], myoglobin [61], and s-nitrosothiol [62] content relative to other mammals, but the molecular data made available in this study will offer additional opportunities for research.

Gene expression in cetacean blood

A wide array of GO terms are represented in the *de novo* assemblies, representing the broad range of transcripts present in blood. When the 10 most highly represented GO terms in each category are examined, the rankings are nearly identical between the beluga whale and bottlenose dolphin transcriptomes (Figure 4). A direct comparison of blast hits shows that 12,636 unique top hits are present in both the beluga and dolphin transcriptomes and, in all, 76% (beluga) or 84.6% (dolphin) of the annotated full transcriptomes return identical top hits. However, among the transcripts unique to each species, there is still homology among GO terms, therefore it appears that many of the differences are due to lack of sequence homology, rather than functional homology. For instance, when the transcripts unique to each species are queried, 80% of the top 10 biological process GO terms and 100% of molecular function GO terms are identical between species (data not shown).

In order to exclude unreliable data, reads were mapped to the transcriptome with Bowtie2 and filtered such that only transcripts with an average TPM ≥ 1 across all samples and a TPM > 0 in at least half of the samples were retained and served as the background blood transcriptome for all further analyses. For the beluga whale, this background set contains 84% of genes (45,872) and 90% of transcripts (82,581). Fifty-three percent are fully annotated in Blast2GO and 40,899 are assigned Entrez

Gene IDs by WebGestalt, representing 10,221 unique Entrez Gene IDs for analyses. There was significant overlap in annotation with the bottlenose dolphin background set of 25,877 genes (71%) or 45,474 transcripts (81%) that mapped to 7671 unique Entrez Gene IDs. Relative to the background of protein coding genes from human, the basis of annotation, both the dolphin and beluga blood transcriptome background gene sets were enriched in core pathways for ribosome, signaling, immune function, cell cycle, and metabolism (FDR < 0.01, Table 4). While blood does exhibit expression of a wide array of genes, all protein coding genes will not be expressed. To this end, GO term enrichment and pathway analysis were also completed using the full suite of annotated genes expressed in dolphin and beluga as the background set. Similarly, the transcripts common to both beluga and dolphin reflect basic cellular functions with enrichment of GO terms or KEGG and Wiki pathways associated with ribosome, signaling, binding, transcription, metabolism, and immune function (FDR<0.05, data not shown).

It is possible that gene expression is impacted by the stress of chase and capture in these wild belugas. Increases in blood transcript expression of genes associated with stress response and energy metabolism have been reported in bottlenose dolphins between samples collected immediately after restraint and just prior to release [17]. Similarly, increases in cortisol were measured in the blow and blood between samples collected immediately after restraint and just prior to release from belugas in Bristol Bay, including some analyzed in this study [63]. However, it has been previously proposed that belugas adapt to the stress of handling and capture during the restraint period [64] and observed decreases in white blood cell and neutrophil counts may support this hypothesis [2]. A comparison of transcript expression between wild belugas and managed belugas, trained to present for blood draws, may be needed to fully assess the impact of chase and restraint on these wild animals. All beluga samples utilized in the current study are from animals that experienced chase and restraint of similar duration with blood draws occurring as soon as possible during the handling period. Transcript expression in each species is discussed in more detail below.

Beluga whale transcript expression in blood

As predicted by the significant overlap in annotation between beluga and dolphin transcriptomes, 46% of the top 100 most highly expressed sequences in beluga share top blast hits with the top 100 in dolphin. The most highly expressed transcript, other than HBA2 or HBB, in beluga blood is major histocompatibility complex, class I, B (HLA-B) accounting for approximately 1.4% of TPM. In fact, 5 of the top 100 most highly expressed sequences are HLAs (2.5% of TPMs), and 28 transcripts

annotated as HLAs are in the beluga transcriptome. Twenty-one percent of the top 100 genes are immune/inflammatory related, similar to the 19% in dolphin (Table 5). Twenty-eight of the top 100 expressed sequences encode ribosomal proteins with 5 additional transcripts also associated with ribosome structure or function (Table 5). In all, 120 transcripts, accounting for 4.2% of TPM are annotated as ribosomal proteins, slightly less than the 5.6% of TPM in dolphin. Likewise, fewer transcripts (6%) involved in heme, iron, and/or oxygen binding, transport, and/or synthesis were observed in the 100 most highly expressed beluga transcripts than in dolphin (Table 5), in agreement with the earlier discussion of lower hemoglobin transcript levels in the beluga blood transcriptome. In contrast, 13% of the top 100 transcripts were involved in cytoskeletal structure and the extracellular matrix in the beluga blood transcriptome (4% in dolphin). The increased diversity of annotated sequences in the beluga blood transcriptome also included functions and processes associated with metabolism, catabolism, stress response, and cell proliferation (Table 5).

Genes involved with immune, detoxification, and stress responses are of particular interest given the applicability to assessing health and exposure in these animals. A simple survey of annotated genes reveals 35 genes involved in stress response (heat shock proteins, glutathione-s-transferase, glutathione peroxidase, superoxide dismutase). Likewise, 14 classical detoxification genes (aryl hydrocarbon receptor, cytochrome P450) were identified. A more comprehensive search was completed in Blast2GO where 915 transcripts were annotated with GO terms containing “stress” and 77 contained “detoxification”. In keeping with the large presence of immune related genes among the most highly expressed genes, 1401 transcripts are annotated with GO terms including “immune”. Collectively 2285 unique transcripts are annotated with GO functions or processes involving “stress”, “detoxification” or “immune”. This represents 4.5% of annotated transcripts and indicates that the beluga blood transcriptome contains many messages that are markers of health and exposure in humans and other animals.

Many of the transcripts identified in the beluga blood transcriptome are associated with cytoskeletal and developmental functions and processes. Correspondingly, we observed enrichment, relative to the human set of protein coding genes, for GO terms cell-substrate junction (FDR = $8.05e^{-6}$) and cell-cell adherens junction (FDR = $3.73e^{-6}$), among others. Belugas molt annually during summer, linked to seasonal endocrine cycles, further aided by the warm, shallow, brackish waters of the estuaries from which samples for this study were collected. During this time, cellular proliferation and differentiation of epidermal cells is among the highest observed in cetaceans or terrestrial mammals

[65, 66]. Correspondingly, transcripts mapping to many of the pathways involved in cellular proliferation, differentiation, and cytoskeletal structure were highly represented in the beluga blood transcriptome. For example, the regulation of actin cytoskeleton (hsa04810, 57 of 72 gene products identified), MAPK signaling (hsa04010, 95 of114 identified), and Wnt signaling (hsa04310, 55 of 72 identified) pathways are all very well mapped in the beluga blood transcriptome. An identified product of the regulation of actin cytoskeleton pathway, p21 activated kinase, is a critical component of the JNK and p38 MAP kinase pathway whose products are essential to the Wnt signaling pathway. Through the expression of transcripts involved in these pathways in beluga blood, it is possible to visualize the signaling and basic developmental processes leading to remodeling of the cytoskeleton, cell differentiation, and cell adhesion that are essential to the seasonal molt cycle. The interplay of these pathways, along with others including cellular adhesion molecules and multiple types of cellular junctions similarly identified in the beluga blood transcriptome (data not shown), have also been demonstrated in the skin transcriptome of bottlenose dolphins [67].

An examination of transcripts annotated with hormone related functions in the beluga blood transcriptome revealed few transcripts annotated with hormone activity (GO:0005179, 14 transcripts), hormone-signaling pathway (GO:0009755, 5 transcripts) or hormone metabolic process (GO:0042445, 4 transcripts). A broad search of GO terms containing “hormone” reported 562 transcripts annotated as such and nearly 900 transcripts when more specific terms are added to the search (estrogen, androgen, prolactin, dopamine, and thyroid). Given the reported representation of hormone related transcripts in human blood [68] and the ties between the beluga molt cycle and seasonal endocrine cycles [65, 66], it is somewhat unexpected to observe limited hormone associated-transcripts in beluga blood. It is possible that there is poor sequence homology between hormone related transcripts in human and cetaceans, as evidenced by mapping transcript expression to pathways involved in thyroid hormone synthesis and signaling. The thyroid hormone synthesis pathway is not well represented by identified transcripts in beluga blood (map04918, 19 of 48 gene products identified, Figure 5A). In fact, most of the transcripts mapped are involved in the calcium signaling portion of the pathway, rather than processes more tightly linked to hormone synthesis. However, the thyroid hormone signaling pathway is well represented in the beluga blood transcriptome (map04919, 53 of 78 identified, Figure 5B), indicating the identification of similar transcripts and functions between human and cetaceans. This was also observed in the dolphin blood transcriptome (data not shown) and the presence of hormone signaling pathways in the absence of well mapped synthesis pathways may indicate that the transcripts

encoding the production of hormones may differ between humans and cetaceans or that the transcripts for hormone synthesis are not expressed in blood.

The blood transcriptomes of beluga whale and bottlenose dolphin presented here are strikingly similar despite their adaptation to vastly different climates. As belugas encounter markedly lower water and air temperatures in their arctic habitat, the transcriptome was queried for transcripts relating to cold tolerance (GO:0016048 detection of temperature stimulus, GO:0070417 cellular response to cold, GO:0009409 response to cold, GO:0009631 cold acclimation, GO:0061411 positive regulation of transcription from RNA polymerase II promoter in response to cold, GO: 0001659 temperature homeostasis, GO:1990845 adaptive thermogenesis). Only 73 transcripts were associated with these GO terms, a similar proportion of the whole transcriptome to the 46 transcripts annotated as such in the dolphin blood transcriptome. However, it is also worth noting that the beluga samples were collected during summer months, therefore cold-responsive functions may not be highly utilized and the corresponding transcripts missing from the blood transcriptome. Seven genes annotated with these GO terms were identified in beluga that did not have a corresponding gene identified in the current dolphin blood transcriptome and, in general, gene expression was higher in beluga than in dolphin (Table 6). METRNL, which had relatively high expression levels among this gene set in beluga, is known to be a hormone that promotes thermogenic gene programs with elevated circulating levels [69]. DGIC is also highly expressed in beluga and is known to be a key regulator of brown adipocyte development whose absence is associated with cold intolerance [70]. However, nearly all of the genes identified in this analysis, including Fos and glycerol kinase, have primary roles in metabolism. Exactly what role they may play in cold adaptation or thermoregulation in beluga whales was not explored in the current study, but may be of interest for further investigation in other studies.

Bottlenose dolphin transcript expression in blood

Overall, transcript expression in non-globin reduced blood is highly similar to that reported previously in globin-reduced samples, with 80% of the 100 most highly expressed transcripts returning identical top hits between studies [1], and does not warrant lengthy discussion. There is a significant decrease in transcript abundance for all genes other than HBA and HBB (30% of TMP to 1% of TMP). Forty-six percent of the top 100 most highly expressed genes encode ribosomal proteins, with an additional 7% involved in the structural constituent of the ribosome or translation (Supplemental Table

1). Nineteen percent of the most highly expressed genes are involved in immune or inflammatory functions, 10% are involved in heme, iron, and/or oxygen binding, transport, and/or synthesis, and an additional 4% are involved in cytoskeletal structure (Supplemental Table 1). As observed in the beluga whale blood transcriptome, relatively few transcripts (343) were annotated with hormone related functions of processes, however this is slightly more than previously reported and may be a result of improved sequence annotation. While direct sequence and blast hit comparisons were not made between the entirety of the *de novo* transcriptome constructed here and that previously published using globin-depleted dolphin blood transcriptomes, the similarity in highly expressed genes, GO terms, pathway mapping, and assembly size and statistics indicates the limited impacts of globin-reduction on overall gene expression and repeatability of the Trinity *de novo* assembler. Due to the small sample size employed in this study, differential expression of transcripts was not examined in the blood of bottlenose dolphin. However, differential expression in globin-reduced blood from a broader survey of these dolphin samples has recently been described [1].

Differential transcript expression in beluga blood

While the beluga whale samples from this study do not allow for analysis of seasonal changes as previously described in dolphin [1], variability in transcript expression associated with sex or year were investigated. As shown in Figure 6, samples did not clearly cluster by sex or year by PCA. PC1, which accounts for 52% of variance is loosely associated with year, where samples collected in 2014 generally cluster separately from 2012 and 2013 samples. An additional 7.4% of variance is accounted for by PC2, however there is no apparent clustering of samples by sex or year along this axis.

Minimal differential expression was associated with sex, where only 93 genes from our background set (0.2%) were differentially expressed between males and females (EBseq, FDR < 0.05), an order of magnitude less than observed in the dolphin blood transcriptome [1]. Of the 49 that are annotated (Supplemental Table 2), 16 are located on the X-chromosome in humans and 6 reside on the y-chromosome. The X-linked genes are CAB5 (x2), USP9X, TXLNG (x2), ARMCX6, UBA1, TMEM27 (x2), SEPT6, SYAP1, ZRSR2, PRPS2 (x2), WWC3, and CLCN4. USPY9 (x2), KDM5D (x3), and UTY are the Y-linked genes. Forty-eight genes are expressed only in males (TPM < 1 in all females) with Log_2 fold change of 5.97 – 15.06. There was no overlap among the 26 genes expressed only in male belugas with the previously identified 11 annotated genes expressed only in male dolphins [1]. No genes were expressed only in females (n=9), however, as our background set required TPM > 0 in at least half the samples, genes not expressed in males (n=15) could be overlooked. Expanding our analysis to the full

transcriptome, 7 genes were expressed only in females (log_2 fold change 8.46 to 13.67) and an additional 2 genes were expressed only in males (log_2 fold change 5.44 and 9.92). None of the genes expressed only in females were annotated. A x-linked eukaryotic translation initiation factor 1A (EIF1AX) and an autosomal CCNI protein variant were the additional genes identified in males only. There is no significant enrichment of any GO term or pathway among these genes, likely due to the limited number of annotated genes differentially expressed by sex.

While the available samples did not allow for investigation of seasonal variability in beluga blood transcriptomes, differential expression between years was analyzed. Overall, the beluga blood transcriptome is extremely stable across time. There were only 2 genes significantly different in all 3 years, neither of which met the requirements for inclusion in our background gene set. Nineteen genes exhibited significantly different expression in 2012, relative to 2013-14 (FDR < 0.05). Only 13 are annotated (Supplemental Table 3). Similarly, 25 gene were differentially expressed in 2013 relative to 2012 and 2014 of which 9 are annotated (Supplemental Table 3). Due to the limited number of differentially expressed genes in these sets, there is no significant enrichment of any GO term or pathway. The greatest extent of difference was observed between 2014 and the years 2012-13 in which 1563 genes were differentially expressed. The 699 that were annotated are shown in Supplemental Table 3. The KEGG pathway basal transcription factors was significantly enriched (FDR = 9.33E-4) with lower expression of several general transcription factor IIs and TATA-box binding protein associated factors in 2014. In addition, pathway enrichment analyses indicated perturbations in oxidative phosphorylation that can lead to mitochondrial dysfunction (KEGG pathways oxidative phosphorylation and Alzheimer's, Parkinson's, Huntington's, and non-alcoholic fatty liver diseases; Wikipathways electron transport chain and oxidative phosphorylation; FDR < 0.05). Enrichment was driven by increased expression levels in 2014 of several cytochrome c oxidase subunits, NADH:ubiquinone oxioreductase subunits, ubiquinol-cytochrome c reductases, succinate dehydrogenase complexes, and ATP synthase H+ transporting mitochondrial F_0 and F_1 subunits. [71]Thus, it is possible that belugas sampled in 2014 were experiencing mitochondrial dysfunction and, potentially, resultant increases in oxidative stress and may be of interest when examining other data sets produced during these health assessments.

Conclusions

We constructed and annotated the first beluga whale transcriptome and assessed gene expression in blood over 3 years in animals from the healthy Bristol Bay, AK stock. The blood

519 transcriptome contains a wide array of genes that map to diverse pathways. The presence of many
520 genes involved in stress, detoxification, and immune functions indicate that blood transcriptomic
521 analyses may provide information on health status and exposure. The transcriptome was very stable,
522 with little differential expression between sex or year. This study provides a wealth of transcriptomic
523 data on beluga whales and provides a sizeable pool of baseline data. While the impacts of season,
524 sexual maturity, disease, and geography must still be investigated the data described here serves as a
525 preliminary for comparison with other studies in beluga whales.

526 **Acknowledgements**

527 All beluga work was conducted under NMFS permit no. 14245 and in accordance with approval from the
528 Marine Mammal Laboratory of the Alaska Fisheries Science Center (MML/AFSC) IACUC protocols (ID
529 number: AFSC-NWFSC2012-1). All dolphin samples for this study were collected under the auspices of a
530 USDA public display permit issued to Dolphin Quest. These efforts advance the animal management
531 practices within Dolphin Quest locations. All protocols were approved by the Dolphin Quest Scientific
532 Committee and Animal Wellness Team, which oversees all animal care and use in any scientific activities,
533 as afforded under the USDA public display permit. Sampling of wild belugas was made possible by
534 funding through Georgia Aquarium and MML/AFSC and with the support the Bristol Bay Beluga field
535 team which also included members from the Bristol Bay Native Association, Bristol Bay Marine Mammal
536 Council, Alaska SeaLife Center, Alaska Veterinary Pathology Service, Alaska Department of Fish and
537 Game, Mystic Aquarium, Shedd Aquarium, and National Institute of Standards and Technology. We
538 would like to thank the staff of Dolphin Quest Hawaii for sample collection. This research was funded by
539 NOAA NCCOS programmatic funding.

540

541 **NOAA Disclaimer**

542 The scientific results and conclusions, as well as any opinions expressed herein, are those of the
543 author(s) and do not necessarily reflect the views of NOAA or the Department of Commerce. The
544 mention of any commercial product is not meant as an endorsement by the Agency or Department.

545

546 **Figure 1.** Expression of HBA and HBB transcripts in beluga blood. Relative fold change in expression (to
547 DLBB13-03) was determined by qPCR prior to transcriptome sequencing to ensure adequate depth for

transcript discovery. The line marking 1-fold change indicates hemoglobin expression equal to that of DLBB13-03, the sample for which preliminary RNA-seq data was available.

Figure 2. Expression of HBB transcripts in RNA-seq analyses of globin-reduced dolphin blood indicated lower expression level during warmer months (analysis of previously published data from [1]). Grey bars represent HBB expression plotted on the left axis. Black circles represent temperature, plotted on the right axis.

Figure 3. ClustalW multiple sequence alignment of translated HBB transcripts from beluga whale, bottlenose dolphin, and woolly mammoth. The beluga and dolphin transcripts constructed here are identical with the exception of the T12A substitution, also identified in woolly mammoth. Identical amino acids are shaded. *: highly conserved substitution (Blossum90 Matrix). +: conserved substitution (Blossum65 Matrix).

Figure 4. The 10 most highly expressed GO terms (level 6) among annotated transcripts in bottlenose dolphin or beluga whale. GO terms identified in both species are indicated with an (+) whereas GO terms expressed only in one species are indicated with an (-).

Figure 5. Transcripts identified in beluga blood mapping to gene products in the (A) thyroid hormone synthesis or (B) thyroid hormone signaling KEGG pathways. Gene products shaded pink are present in the beluga whale blood transcriptome. Gene products shaded green are present in the human UniProt/Swiss-Prot database, but not identified in the beluga whale transcriptome.

Figure 6. Principle components analysis (PCA) of beluga whale blood transcriptomes. Samples did not cluster by year or sex along either the PC1 or PC2 axes.

References

1. Morey JS, Neely MG, Lunardi D, Anderson PE, Schwacke LH, Campbell M, Van Dolah FM: **RNA-Seq analysis of seasonal and individual variation in blood transcriptomes of healthy managed bottlenose dolphins**. *BMC Genomics* 2016, **17**(1):720. 10.1186/s12864-016-3020-8.
2. Norman SA, Goertz CEC, Burek KA, Quakenbush LT, Cornick LA, Romano TA, Spoon T, Miller W, Beckett LA, Hobbs RC: **Seasonal Hematology and Serum Chemistry of Wild Beluga WhalesS (*Delphinapeterus leucas*) in Bristol Bay, Alaska, USA**. *Journal of Wildlife Diseases* 2012, **48**(1):21-32. 10.7589/0090-3558-48.1.21.

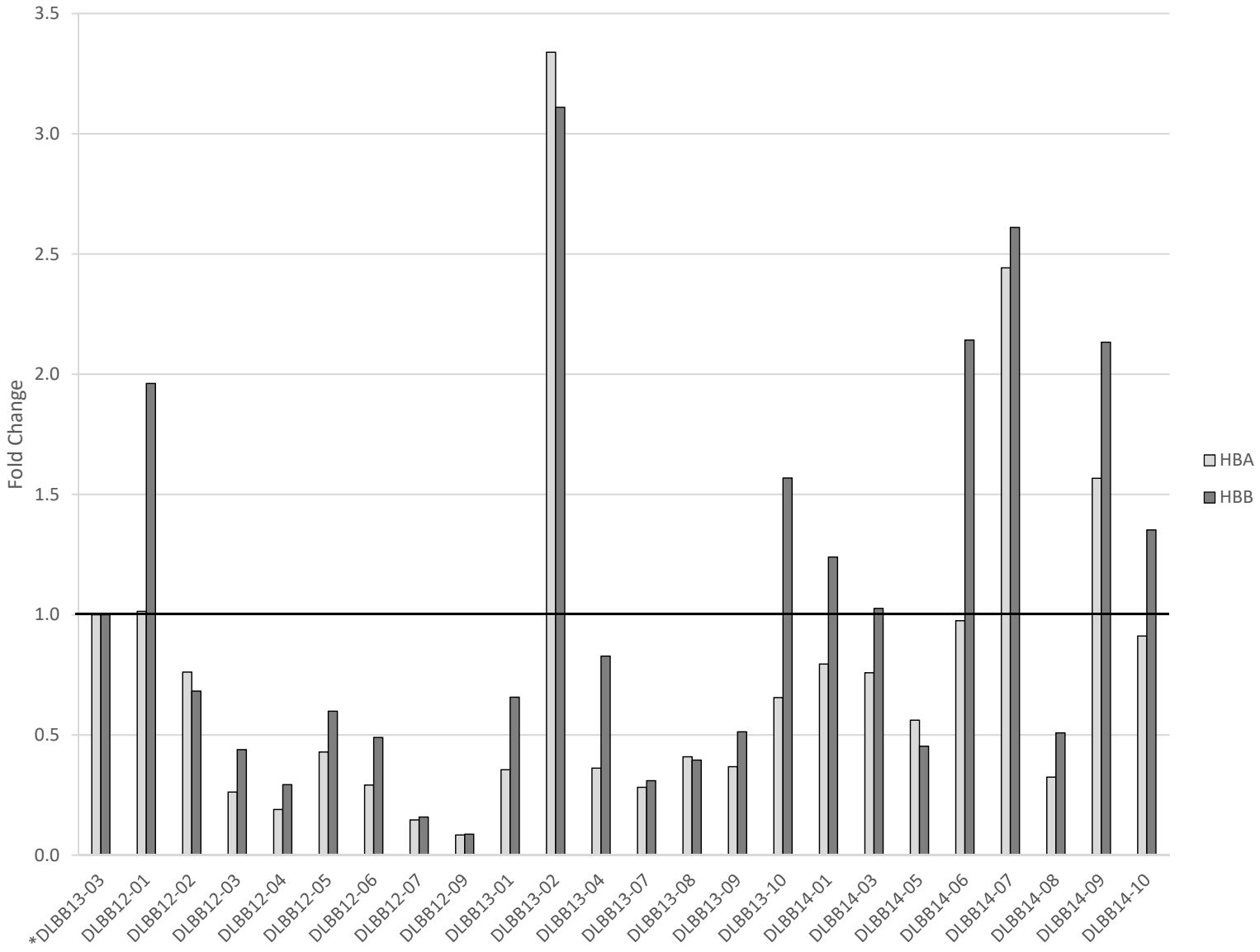
3. Simeone CA, Gulland FMD, Norris T, Rowles TK: **A Systematic Review of Changes in Marine Mammal Health in North America, 1972-2012: The Need for a Novel Integrated Approach**. *PLOS ONE* 2015, **10**(11):e0142105. 10.1371/journal.pone.0142105.
4. National Academies of Sciences EaM: **Approaches to Understanding the Cumulative Effects of Stressors on Marine Mammals**. Washington, DC: The National Academies Press. doi:10.17226/23479.
5. Tsai YL, Chen SY, Lin SC, Li JY: **Effects of Physiological Factors and Seasonal Variations on Hematology and Plasma Biochemistry of Beluga Whales (Delphinapterus leucas) Managed in Pingtung, Taiwan**. *Aquatic Mammals* 2016, **42**(4):494-506. 10.1578/Am.42.4.2016.494.
6. Chaussabel D: **Assessment of immune status using blood transcriptomics and potential implications for global health**. *Seminars in immunology* 2015, **27**(1):58-66. 10.1016/j.smim.2015.03.002.
7. Mohr S, Liew CC: **The peripheral-blood transcriptome: new insights into disease and risk assessment**. *Trends Mol Med* 2007, **13**(10):422-432. 10.1016/j.molmed.2007.08.003.
8. De Boever P, Wens B, Forcheh AC, Reynders H, Nelen V, Kleinjans J, Van Larebeke N, Verbeke G, Valkenborg D, Schoeters G: **Characterization of the peripheral blood transcriptome in a repeated measures design using a panel of healthy individuals**. *Genomics* 2014, **103**(1):31-39. 10.1016/j.ygeno.2013.11.006.
9. De Jong S, Neeleman M, Luykx JJ, ten Berg MJ, Strengman E, Den Breeijen HH, Stijvers LC, Buizer-Voskamp JE, Bakker SC, Kahn RS *et al*: **Seasonal changes in gene expression represent cell-type composition in whole blood**. *Human molecular genetics* 2014, **23**(10):2721-2728. 10.1093/hmg/ddt665.
10. Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, Cooper N, Burren OS, Fulford AJ, Hennig BJ *et al*: **Widespread seasonal gene expression reveals annual differences in human immunity and physiology**. *Nature communications* 2015, **6**:7000. 10.1038/ncomms8000.
11. Karlovich C, Duchateau-Nguyen G, Johnson A, McLoughlin P, Navarro M, Fleurbaey C, Steiner L, Tessier M, Nguyen T, Wilhelm-Seiler M *et al*: **A longitudinal study of gene expression in healthy individuals**. *BMC medical genomics* 2009, **2**:33. 10.1186/1755-8794-2-33.
12. Ball HC, Holmes RK, Londraville RL, Thewissen JGM, Duff RJ: **Leptin in Whales: Validation and Measurement of mRNA Expression by Absolute Quantitative Real-Time PCR**. *PLOS ONE* 2013, **8**(1):e54277. 10.1371/journal.pone.0054277.
13. Buckman AH, Veldhoen N, Ellis G, Ford JK, Helbing CC, Ross PS: **PCB-associated changes in mRNA expression in killer whales (Orcinus orca) from the NE Pacific Ocean**. *Environ Sci Technol* 2011, **45**(23):10194-10202. 10.1021/es201541j.
14. Fossi MC, Urban J, Casini S, Maltese S, Spinsanti G, Panti C, Porcelloni S, Panigada S, Lauriano G, Nino-Torres C *et al*: **A multi-trial diagnostic tool in fin whale (Balaenoptera physalus) skin biopsies of the Pelagos Sanctuary (Mediterranean Sea) and the Gulf of California (Mexico)**. *Mar Environ Res* 2010, **69 Suppl**(1879-0291 (Electronic)):S17-20. 10.1016/j.marenvres.2009.10.006.
15. Gui D, Jia K, Xia J, Yang L, Chen J, Wu Y, Yi M: **De novo Assembly of the Indo-Pacific Humpback Dolphin Leucocyte Transcriptome to Identify Putative Genes Involved in the Aquatic Adaptation and Immune Response**. *PLoS ONE* 2013, **8**(8):e72417. 10.1371/journal.pone.0072417.
16. Mancia A, Ryan JC, Van Dolah FM, Kucklick JR, Rowles TK, Wells RS, Rosel PE, Hohn AA, Schwacke LH: **Machine learning approaches to investigate the impact of PCBs on the transcriptome of the common bottlenose dolphin (Tursiops truncatus)**. *Mar Environ Res* 2014, **100**:57-67. 10.1016/j.marenvres.2014.03.007.

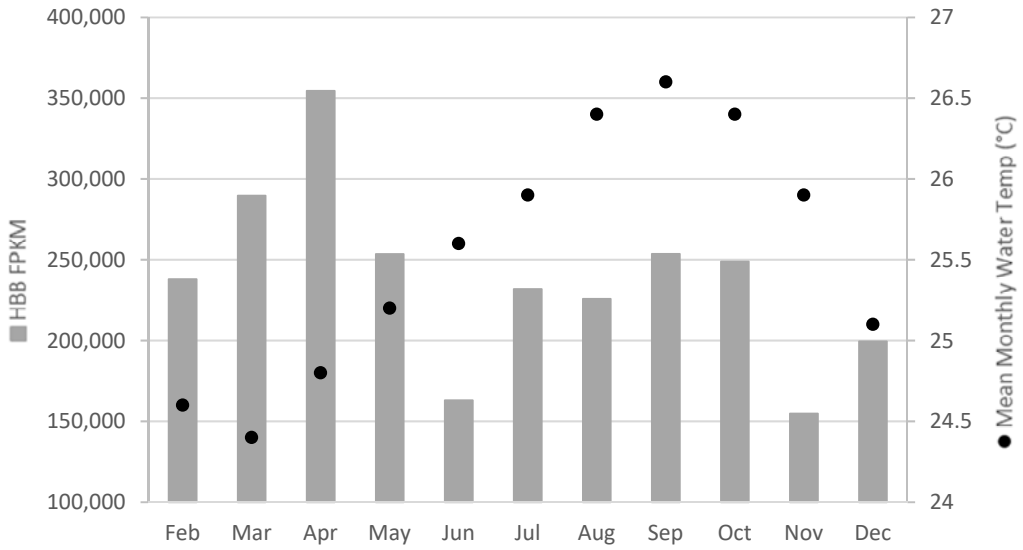
17. Mancia A, Warr GW, Chapman RW: **A transcriptomic analysis of the stress induced by capture-release health assessment studies in wild dolphins (Tursiops truncatus)**. *Mol Ecol* 2008, **17**(11):2581-2589. 10.1111/j.1365-294X.2008.03784.x.
18. Noel M, Loseto LL, Helbing CC, Veldhoen N, Dangerfield NJ, Ross PS: **PCBs are associated with altered gene transcript profiles in arctic Beluga Whales (Delphinapterus leucas)**. *Environ Sci Technol* 2014, **48**(5):2942-2951. 10.1021/es403217r.
19. Sitt T, Bowen L, Blanchard MT, Smith BR, Gershwin LJ, Byrne BA, Stott JL: **Quantitation of leukocyte gene expression in cetaceans**. *Dev Comp Immunol* 2008, **32**(11):1253-1259. 10.1016/j.dci.2008.05.001.
20. Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK *et al*: **Minke whale genome and aquatic adaptation in cetaceans**. *Nat Genet* 2014, **46**(1):88-92. 10.1038/ng.2835.
21. Fabrizius A, Hoff MLM, Engler G, Folkow LP, Burmester T: **When the brain goes diving: transcriptome analysis reveals a reduced aerobic energy metabolism and increased stress proteins in the seal brain**. *BMC Genomics* 2016, **17**(1):583. 10.1186/s12864-016-2892-y.
22. Hoffman JI: **Gene discovery in the Antarctic fur seal (Arctocephalus gazella) skin transcriptome**. *Mol Ecol Resour* 2011, **11**(4):703-710. 10.1111/j.1755-0998.2011.02999.x.
23. Hoffman JI, Thorne MA, Trathan PN, Forcada J: **Transcriptome of the dead: characterisation of immune genes and marker development from necropsy samples in a free-ranging marine mammal**. *BMC Genomics* 2013, **14**(1):52. 10.1186/1471-2164-14-52.
24. Khudyakov JI, Preeyanon L, Champagne CD, Ortiz RM, Crocker DE: **Transcriptome analysis of northern elephant seal (Mirounga angustirostris) muscle tissue provides a novel molecular resource and physiological insights**. *BMC Genomics* 2015, **16**(1):64. 10.1186/s12864-015-1253-6.
25. Mancia A, Ryan JC, Chapman RW, Wu Q, Warr GW, Gulland FM, Van Dolah FM: **Health status, infection and disease in California sea lions (Zalophus californianus) studied using a canine microarray platform and machine-learning approaches**. *Dev Comp Immunol* 2012, **36**(4):629-637. 10.1016/j.dci.2011.10.011.
26. Mos L, Tabuchi M, Dangerfield N, Jeffries SJ, Koop BF, Ross PS: **Contaminant-associated disruption of vitamin A and its receptor (retinoic acid receptor alpha) in free-ranging harbour seals (Phoca vitulina)**. *Aquat Toxicol* 2007, **81**(3):319-328. 10.1016/j.aquatox.2006.12.017.
27. Spitz J, Becquet V, Rosen DA, Trites AW: **A nutrigenomic approach to detect nutritional stress from gene expression in blood samples drawn from Steller sea lions**. *Comparative biochemistry and physiology Part A, Molecular & integrative physiology* 2015, **187**:214-223. 10.1016/j.cbpa.2015.02.006.
28. Tabuchi M, Veldhoen N, Dangerfield N, Jeffries S, Helbing CC, Ross PS: **PCB-related alteration of thyroid hormones and thyroid hormone receptor gene expression in free-ranging harbor seals (Phoca vitulina)**. *Environ Health Perspect* 2006, **114**(7):1024-1031. D - NLM: PMC1513321 EDAT-2006/07/13 09:00 MHDA- 2006/11/03 09:00 CRDT- 2006/07/13 09:00 PST - ppublish.
29. Bowen L, Miles AK, Murray M, Haulena M, Tuttle J, Van Bonn W, Adams L, Bodkin JL, Ballachey B, Estes J *et al*: **Gene transcription in sea otters (Enhydra lutris); development of a diagnostic tool for sea otter and ecosystem health**. *Mol Ecol Resour* 2012, **12**(1):67-74. 10.1111/j.1755-0998.2011.03060.x.
30. Ariya PA, Dastoor AP, Amyot M, Schroeder WH, Barrie L, Anlauf K, Raofie F, Ryzhkov A, Davignon D, Lalonde J *et al*: **The Arctic: a sink for mercury**. *Tellus B* 2004, **56**(5):397-403. 10.1111/j.1600-0889.2004.00118.x.
31. Frank W, Mackay D: **Global Fractionation and Cold Condensation of Low Volatility Organochlorine Compounds in Polar Regions**. *Ambio* 1993, **22**(1):10-18.

32. Hoguet J, Keller JM, Reiner JL, Kucklick JR, Bryan CE, Moors AJ, Pugh RS, Becker PR: **Spatial and temporal trends of persistent organic pollutants and mercury in beluga whales (Delphinapterus leucas) from Alaska**. *Sci Total Environ* 2013, **449**:285-294. 10.1016/j.scitotenv.2013.01.072.
33. Frost KJ, Lowry L: **Distriburtion, abundance and movements of beluga whles, *Delphinapterus leucas*, in coastal waters of western Alaska**. *Canadian Bulletin of Fisheries and Aquatic Sciences* 1990, **224**(Advances in research on the beluga whale, *Delphinapterus leucas*):39-57.
34. O'Corry-Crowe GM, Suydam RS, Rosenberg A, Frost KJ, Dizon AE: **Phylogeography, population structure and dispersal patterns of the beluga whale Delphinapterus leucas in the western Nearctic revealed by mitochondrial DNA**. *Molecular Ecology* 1997, **6**(10):955-970. 10.1046/j.1365-294X.1997.00267.x.
35. Muto M, Helker V, Angliss RP, Allen B, Boveng PL, Breiwick J, Cameron M, Clapham P, Dahle S, Dalhlheim M *et al*: **Alaska marine mammal stock assessments, 2015**. In. Edited by Commerce UDo; 2016: 300. 10.7289/V5/TM-AFSC-323.
36. Castellote M, Mooney TA, Quakenbush L, Hobbs R, Goertz C, Gaglione E: **Baseline hearing abilities and variability in wild beluga whales (Delphinapterus leucas)**. *The Journal of Experimental Biology* 2014, **217**:1682-1691. 10.1242/jeb.093252.
37. Norman S, Hobbs R, Goertz C, A. Burek-Huntington K, Shelden K, A. Smith W, A. Beckett L: **Potential natural and anthropogenic impediments to the conservation and recovery of Cook Inlet beluga whales, Delphinapterus leucas**. *Marine Fisheries Review* 2015, **77**(2):89-105. 10.7755/MFR.77.2.5.
38. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, Matasci N, Wang L, Hanlon M, Lenards A *et al*: **The iPlant Collaborative: Cyberinfrastructure for Plant Biology**. *Frontiers in plant science* 2011, **2**:34. 10.3389/fpls.2011.00034.
39. Bolger AM, Lohse M, Usadel B: **Trimmomatic: A flexible trimmer for Illumina Sequence Data**. *Bioinformatics* 2014. 10.1093/bioinformatics/btu170.
40. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nat Biotechnol* 2011, **29**(7):644-652. 10.1038/nbt.1883.
41. Parra G, Bradnam K, Korf I: **CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes**. *Bioinformatics* 2007, **23**(9):1061-1067. 10.1093/bioinformatics/btm071.
42. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**. *Bioinformatics* 2015, **31**(19):3210-3212. 10.1093/bioinformatics/btv351.
43. Conesa A, Gotz S: **Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics**. *Int J Plant Genomics* 2008, **619832**:12. doi:10.1155/2008/619832.
44. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation and visualization in functional genomics research**. *Bioinformatics* 2005:bti610. 10.1093/bioinformatics/bti610.
45. Götz S, Arnold R, Sebastián-León P, Martín-Rodríguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T, Conesa A: **B2G-FAR, a species-centered GO annotation repository**. *Bioinformatics* 2011, **27**(7):919-924. 10.1093/bioinformatics/btr059.
46. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite**. *Nucl Acids Res* 2008, **36**(10):3420-3435. 10.1093/nar/gkn176.
47. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *BMC Bioinformatics* 2011, **12**:323-323. 10.1186/1471-2105-12-323.

48. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012, **9**(4):357-359. 10.1038/nmeth.1923.
49. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C: **EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments**. *Bioinformatics* 2013, **29**(8):1035-1043. 10.1093/bioinformatics/btt087.
50. Wang J, Duncan D, Shi Z, Zhang B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013**. *Nucleic Acids Res* 2013, **41**(Web Server issue):W77-83. 10.1093/nar/gkt439.
51. Zhang B, Kirov S, Snoddy J: **WebGestalt: an integrated system for exploring gene sets in various biological contexts**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W741-748. 10.1093/nar/gki475.
52. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs**. *Nucl Acids Res* 2017, **45**(D1):D353-D361. 10.1093/nar/gkw1092.
53. Wickham H: **ggplot2: Elegant Graphics for Data Anlaysis**. New York: Springer-Verlag; 2009.
54. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT**. *Nucleic Acids Symposium Series* 1999, **41**:95-98.
55. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucl Acids Res* 1994, **22**(22):4673-4680.
56. Hardison RC: **Evolution of Hemoglobin and Its Genes**. *Cold Spring Harbor Perspectives in Medicine* 2012, **2**(12):a011627. 10.1101/cshperspect.a011627.
57. Choi I, Bao H, Kommadath A, Hosseini A, Sun X, Meng Y, Stothard P, Plastow GS, Tuggle CK, Reecy JM *et al*: **Increasing gene discovery and coverage using RNA-seq of globin RNA reduced porcine blood samples**. *BMC Genomics* 2014, **15**:954. 10.1186/1471-2164-15-954.
58. Campbell KL, Roberts JE, Watson LN, Stetefeld J, Sloan AM, Signore AV, Howatt JW, Tame JR, Rohland N, Shen TJ *et al*: **Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance**. *Nat Genet* 2010, **42**(6):536-540. 10.1038/ng.574.
59. Hedrick MS, Duffield DA: **Haematological and rheological characteristics of blood in seven marine mammal species: physiological implications for diving behaviour**. *Journal of Zoology* 1991, **225**(2):273-283. 10.1111/j.1469-7998.1991.tb03816.x.
60. Dhindsa DS, Metcalfe J, Hoversland AS, Hartman RA: **Comparative studies of the respiratory functions of mammalian blood X. Killer whale (Orcinus orca linnaeus) and beluga whale (Delphinapterus leucas)**. *Respiration Physiology* 1974, **20**(2):93-103. http://dx.doi.org/10.1016/0034-5687(74)90099-1.
61. Noren SR, Suydam R: **Navigating under sea ice promotes rapid maturation of diving physiology and performance in beluga whales**. *J Exp Biol* 2016, **219**(Pt 18):2828-2836. 10.1242/jeb.143644.
62. Fago A, Parraga DG, Petersen EE, Kristensen N, Giouri L, Jensen FB: **A comparison of blood nitric oxide metabolites and hemoglobin functional properties among diving mammals**. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 2017, **205**:35-40. http://dx.doi.org/10.1016/j.cbpa.2016.12.013.
63. Thompson LA, Spoon TR, Goertz CEC, Hobbs RC, Romano TA: **Blow Collection as a Non-Invasive Method for Measuring Cortisol in the Beluga (Delphinapterus leucas)**. *PLOS ONE* 2014, **9**(12):e114062. 10.1371/journal.pone.0114062.
64. J. St. Aubin D, Dierauf L: **Stress and Marine Mammals**. In: *CRC Handbook of Marine Mammal Medicine.* CRC Press; 2001: 253-269. doi:10.1201/9781420041637.ch13.
65. Aubin DJS, Smith TG, Geraci JR: **Seasonal epidermal molt in beluga whales, Delphinapterus leucas**. *Canadian Journal of Zoology* 1990, **68**(2):359-367. 10.1139/z90-051.

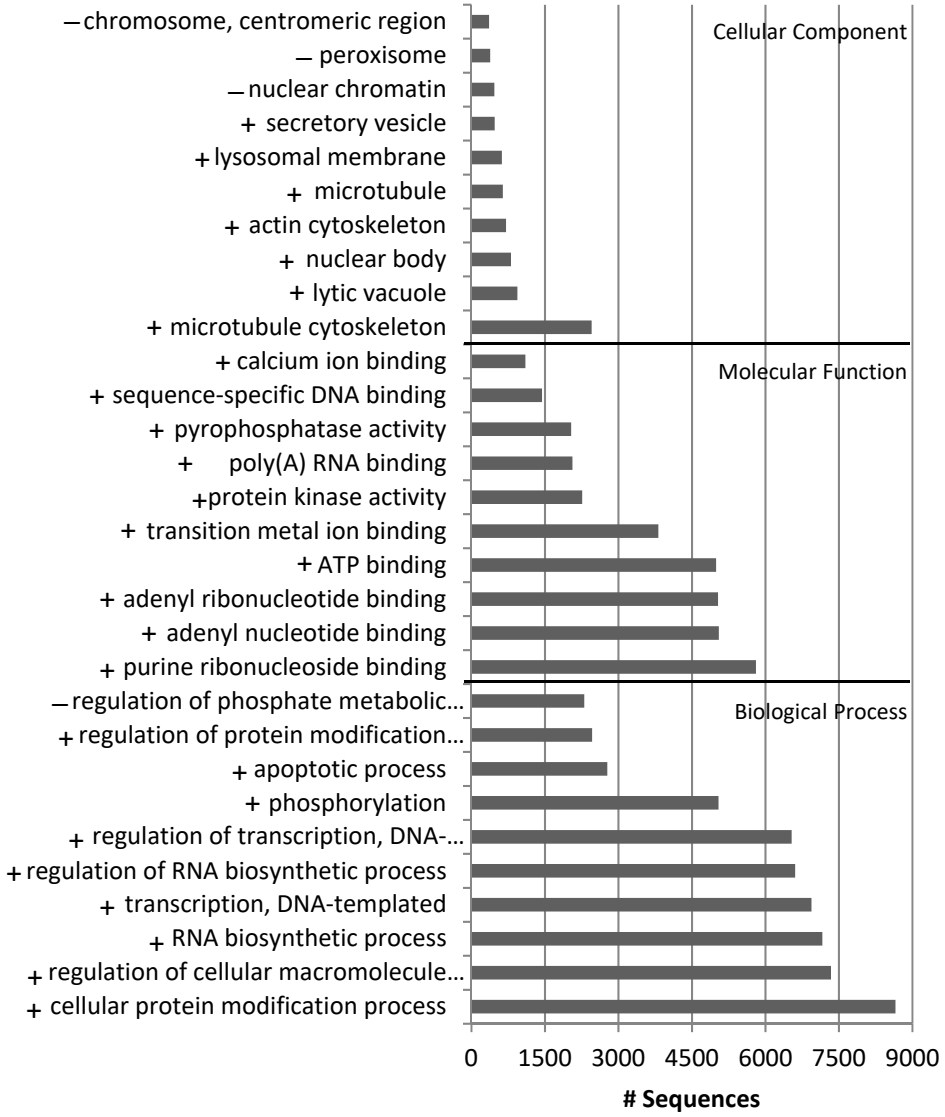
66. Smith TG, Staubin DJ, Hammill MO: **Rubbing Behavior of Belugas, Delphinapterus-Leucas, in a High Arctic Estuary**. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 1992, **70**(12):2405-2409. DOI 10.1139/z92-322.
67. Neely MG, Morey JS, Anderson P, Balmer BC, Ylitalo GM, Zolman ES, Speakman TR, Sinclair C, Bachman MJ, Huncik K *et al*: **Skin Transcriptomes of Common Bottlenose Dolphins (*Tursiops truncatus*) from the Northern Gulf of Mexico and Southeastern U.S. Atlantic Coasts**. *Mar Genomics* submitted.
68. Nicholson AC, Unger ER, Mangalathu R, Ojaniemi H, Vernon SD: **Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells**. *Brain research Molecular brain research* 2004, **129**(1-2):193-197. 10.1016/j.molbrainres.2004.06.036.
69. Rao RR, Long JZ, White JP, Svensson KJ, Lou J, Lokurkar I, Jedrychowski MP, Ruas JL, Wrann CD, Lo JC *et al*: **Meteorin-like is a hormone that regulates immune-adipose interactions to increase beige fat thermogenesis**. *Cell* 2014, **157**(6):1279-1291. 10.1016/j.cell.2014.03.065.
70. Jimenez-Preitner M, Berney X, Uldry M, Vitali A, Cinti S, Ledford Julie G, Thorens B: **Plac8 Is an Inducer of C/EBPB; Required for Brown Fat Differentiation, Thermoregulation, and Control of Body Weight**. *Cell Metab*, **14**(5):658-670. 10.1016/j.cmet.2011.08.008.
71. Shi Q, Gibson GE: **Oxidative stress and transcriptional regulation in Alzheimer disease**. *Alzheimer Dis Assoc Disord* 2007, **21**(4):276-291. 10.1097/WAD.0b013e31815721c3.


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Fold Change
HBA
HBB
*DLBB13-03
DLBB12-01
DLBB12-02
DLBB12-03
DLBB12-04
DLBB12-05
DLBB12-06
DLBB12-07
DLBB12-09
DLBB13-01
DLBB13-02
DLBB13-04
DLBB13-07
DLBB13-08
DLBB13-09
DLBB13-10
DLBB14-01
DLBB14-03
DLBB14-05
DLBB14-06
DLBB14-07
DLBB14-08
DLBB14-09
DLBB14-10

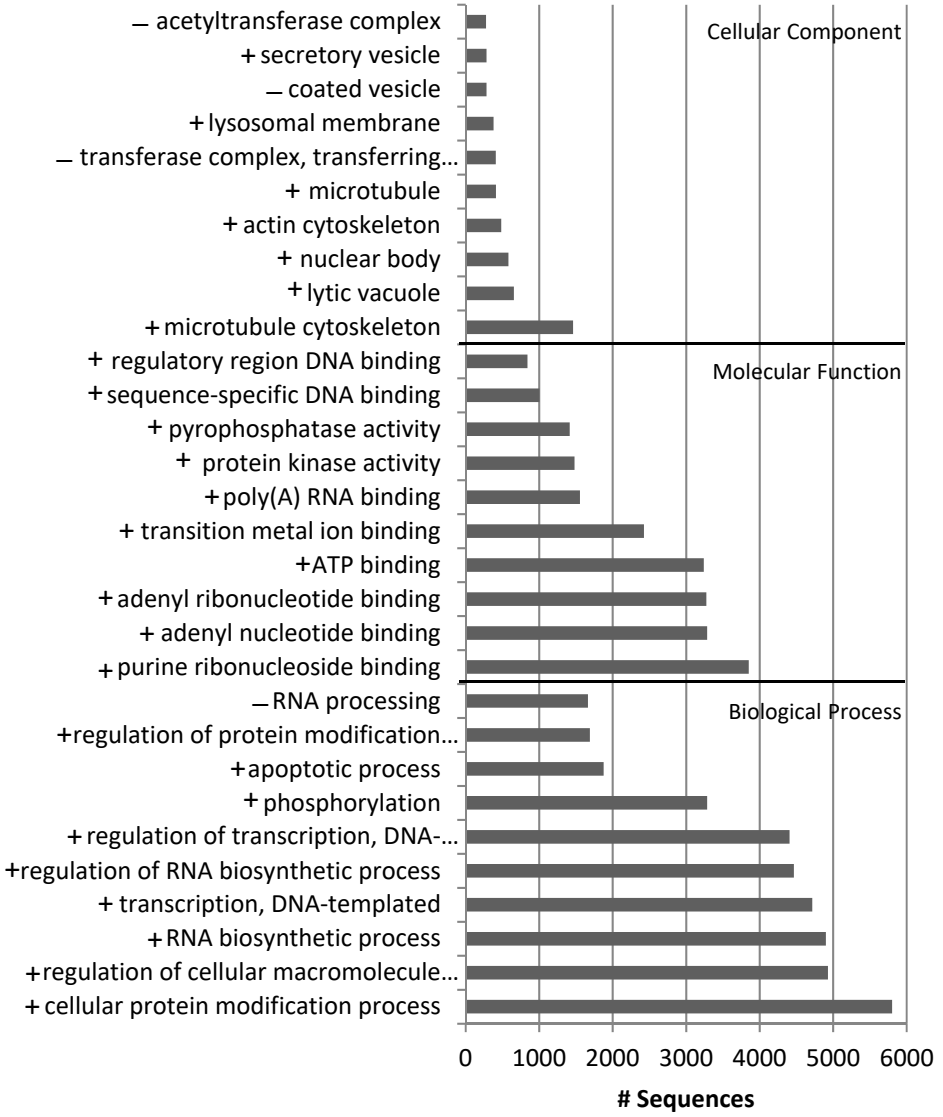

400,000
350,000
300,000
250,000
200,000
150,000
100,000
27
26.5
26
25.5
25
24.5
24
HBB FPKM
Mean Monthly Water Temp (°C)
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

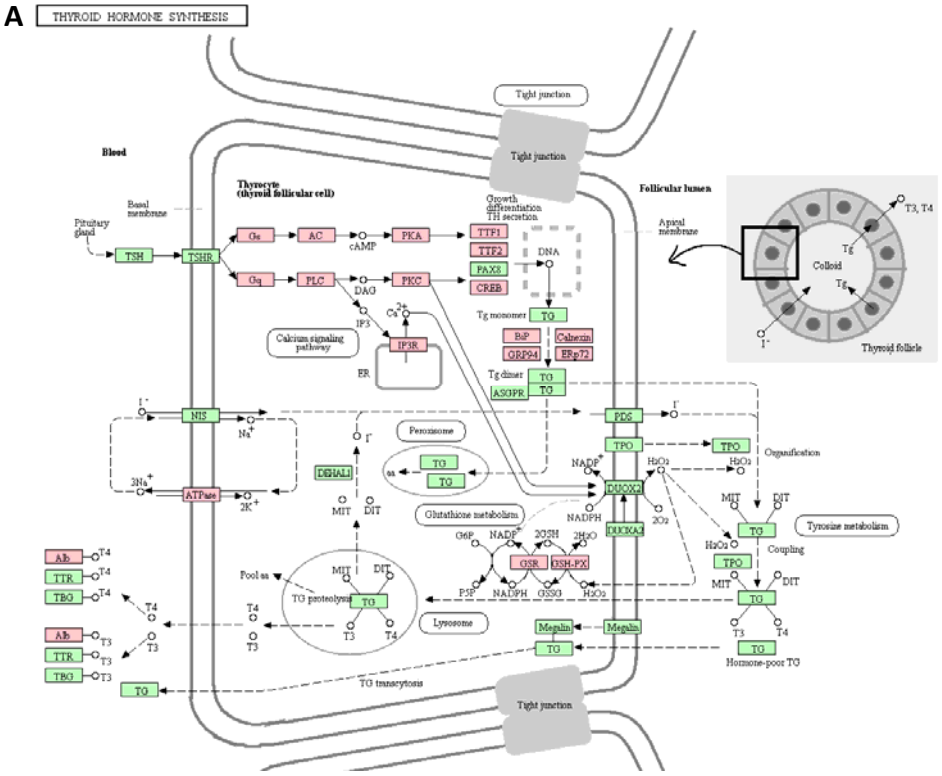

10 20 30 40 50
Bottlenose Dolphin TR15564|c0_g1_i1 HBB MVHLTGEEKSAVTALWGKVNVEEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMKN
Bottlenose Dolphin TR16882|c0_g1_i1 HBB MVHLTGEEKSAVTALWGKVNVEEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMKN
Beluga Whale TR28281|c0_g1_i1 HBB MVHLTGEEKSAVAALWGKVNVEEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMKN
Beluga Whale TR34090|c0_g1_i1 HBB MVHLTGEEKSAVAALWGKVNVEEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMKN
Woolly Mammoth ACV41408.1 HBB/D MVNLTAAEKTQVANLWGKVNVKELGGEALSRLLVVYPWTRRFFEHFGDLSTADAVLHN
70 80 90 100 110
Bottlenose Dolphin TR15564|c0_g1_i1 HBB VKKHGQKVLASFGEGLKHLDDLKGTFAALSELHCDKLHVDPENFRLLGNVLVVVLARH
Bottlenose Dolphin TR16882|c0_g1_i1 HBB VKKHGQKVLASFGEGLKHLDDLKGTFAALSELHCDKLHVDPENFRLLGNVLVVVLARH
Beluga Whale TR28281|c0_g1_i1 HBB VKKHGQKVLASFSEGLKHLDDLKGTFAALSELHCDKLHVDPENFRLLGNVLVVVLARH
Beluga Whale TR34090|c0_g1_i1 HBB VKKHGQKVLASFSEGLKHLDDLKGTFAALSELHCDKLHVDPENFRLLGNVLVVVLARH
Woolly Mammoth ACV41408.1 HBB/D VLAHGEKVLTSFGEGLKHLDNLKGTFSDLSELHCDKLHVDPQNFRLLGNVLVIVLARH
130 140
Bottlenose Dolphin TR15564|c0_g1_i1 HBB KEFTPELQSAYQKVVAGVATALAHKYH
Bottlenose Dolphin TR16882|c0_g1_i1 HBB KEFTPELQSAYQKVVAGVATALAHKYH
Beluga Whale TR28281|c0_g1_i1 HBB KEFTPELQSAYQKVVAGVATALAHKYH
Beluga Whale TR34090|c0_g1_i1 HBB KEFTPELQSAYQKVVAGVATALAHKYH
Woolly Mammoth ACV41408.1 HBB/D KEFTPDVQAAYEKVVAGVANALAHKYH

Beluga Whale



Bottlenose Dolphin




A
THYROID HORMONE SYNTHESIS
Blood
Thyrocyte (thyroid follicular cell)
Follicular lumen
Tight junction
Basal membrane
Apical membrane
Pituitary gland
TSH
TSHR
Gs
AC
cAMP
PKA
Gq
PLC
DAG
PKC
IP3
Ca2+
IP3R
ER
Calcium signaling pathway
TTF1
TTF2
PAX8
CREB
DNA
Tg monomer
TG
BiP
Calnexin
GRP94
ERp72
Tg dimer
ASGPR
Growth differentiation TH secretion
Colloid
Tg
T3, T4
I-
Thyroid follicle
NIS
Na+
3Na+
ATPase
2K+
Peroxisome
DEHAL1
MIT
DIT
Glutathione metabolism
PDS
TPO
H2O2
Organification
DUOX2
DUOXA2
NADP+
NADPH
2O2
Tyrosine metabolism
Coupling
Alb
TTR
TBG
Pool as
TG proteolysis
Lysosome
Megalin
TG transcytosis
Hormone-poor TG
GSR
GSH-PX
G6P
PSP
GSSG
H2O2
B
THYROID HORMONE SIGNALING PATHWAY
T3
T4
S2
ITGA
ITGB
cAMP
Ras
Raf1
MEK
ERK
PKA
MAPK signaling pathway
PLC
IP3
Ca2+
DAG
PKC
Thyroid hormone synthesis
Calcium signaling pathway
STAT
p53
ERα
TRβ1
Enhanced cytokine and growth factor action
p53 signaling pathway
Gene transcription Cell proliferation
Basal / repressed
NCoR1
sin3
HDAC
TR
RXR
DNA
Activation
p/CAF
SRC
CBP
Transcription
TRAPs
TRAP220
NHE1
H+
ATPase
K+
Na+
S1
Src
PI3K
PDK1
PIP3
Akt/PKB
PI3K-Akt signaling pathway
GSK-3β
β-Catenin
Cell cycle Metabolism
MDM2
p53
Cell cycle arrest Apoptosis
TSC2
Rheb
mTOR
Translation
mTOR signaling pathway
AS160
Glucose transport
PFKFB2
Glucose metabolism
BAD
FKHR
Casp9
Survival
Transporter
TRα1
D3
D2
D1/D2
T2
rT3
Regulation of actin cytoskeleton
Actin
F actin
CyclinD1
GATA4
MCIP1
HIF1α
c-Myc
Angiogenesis Glycolysis
HIF-1 signaling pathway
αMHC
SERCA2
PLB
(Cardiomyocyte) Contractility
Cardiac muscle contraction
Wnt
β-Catenin
(Growth plate condrocytes) Terminal differentiation
Wnt signaling pathway
Notch
BMP
Notch signaling pathway
ZAKI-4
GLUT1
PFKP
Glycolysis / Gluconeogenesis
Glucose metabolism

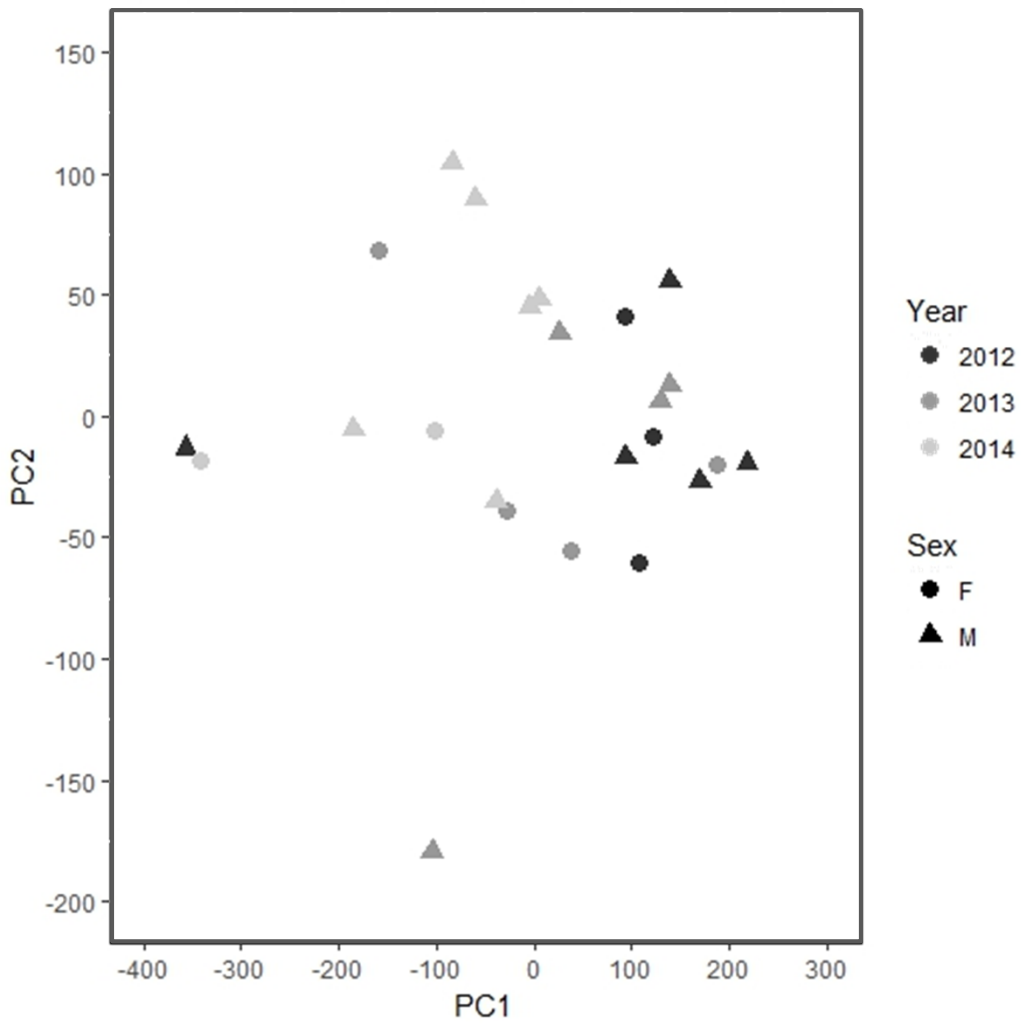

150
100
50
0
-50
-100
-150
-200
PC2
-400
-300
-200
-100
0
100
200
300
PC1
Year
2012
2013
2014
Sex
F
M

Table 1. Beluga whale samples included in this study.

Sample ID	Collection Date	Collection Location	Collection Latitude	Collection Longitude	Sex	GEO Accession
DLBB12-01	01/09/2012	Bristol Bay, Alaska	59.034033	-158.3954	Male	GSM2611035
DLBB12-02	01/09/2012	Bristol Bay, Alaska	59.053067	-158.391983	Female	GSM2611036
DLBB12-03	08/09/2012	Bristol Bay, Alaska	59.053067	-158.391983	Female	GSM2611037
DLBB12-04	08/09/2012	Bristol Bay, Alaska	59.055417	-158.4083	Male	GSM2611038
DLBB12-05	09/09/2012	Bristol Bay, Alaska	58.590467	-158.501517	Female	GSM2611039
DLBB12-06	10/09/2012	Bristol Bay, Alaska	58.950467	-158.501517	Male	GSM2611040
DLBB12-07	12/09/2012	Bristol Bay, Alaska	58.8623	-158.705267	Male	GSM2611041
DLBB12-09	12/09/2012	Bristol Bay, Alaska	58.7598	-158.7742	Male	GSM2611042
DLBB13-01	23/08/2013	Bristol Bay, Alaska	59.0263	-158.4287	Male	GSM2611043
DLBB13-02	24/08/2013	Bristol Bay, Alaska	59.0530	-158.3958	Male	GSM2611044
DLBB13-03	24/08/2013	Bristol Bay, Alaska	59.05109	-158.38335	Female	GSM2611045
DLBB13-04	24/08/2013	Bristol Bay, Alaska	59.0329	-158.36238	Female	GSM2611046
DLBB13-07	28/08/2013	Bristol Bay, Alaska	59.0133	-158.4614	Female	GSM2611047
DLBB13-08	28/08/2013	Bristol Bay, Alaska	59.0530	-158.3981	Male	GSM2611048
DLBB13-09	30/08/2013	Bristol Bay, Alaska	59.0190	-158.4417	Male	GSM2611049
DLBB13-10	30/08/2013	Bristol Bay, Alaska	58.8934	-158.5139	Female	GSM2611050
DLBB14-01	25/08/2014	Bristol Bay, Alaska	59.05273	-158.3881	Male	GSM2611051
DLBB14-03	26/08/2014	Bristol Bay, Alaska	50.052923	-158.393905	Male	GSM2611052
DLBB14-05	28/08/2014	Bristol Bay, Alaska	59.0517	-158.3657	Male	GSM2611053
DLBB14-06	28/08/2014	Bristol Bay, Alaska	58.95997	-158.49783	Female	GSM2611054
DLBB14-07	29/08/2014	Bristol Bay, Alaska	50.052828	-158.383133	Male	GSM2611055
DLBB14-08	31/08/2014	Bristol Bay, Alaska	58.855289	-158.677329	Male	GSM2611056
DLBB14-09	31/08/2014	Bristol Bay, Alaska	58.816445	-158.674507	Female	GSM2611057
DLBB14-10	03/09/2014	Bristol Bay, Alaska	58.816445	-158.674507	Male	GSM2611058

Table 2. Bottlenose dolphin samples included in this study.

Sample ID	Collection Date	Collection Location	Collection Latitude	Collection Longitude	Sex	Age (yr)	GEO Accession
1FP3_062413	24/06/2013	Dophin Quest, Waikoloa, Hawaii	19.9255	-155.889	Female	28	GSM2602137
8SK7_092913	29/09/2013	Dophin Quest, Waikoloa, Hawaii	19.9255	-155.889	Female	12	GSM2602138
8SK7_071413	14/07/2013	Dophin Quest, Waikoloa, Hawaii	19.9255	-155.889	Female	12	GSM2602139
1FP3_041813	18/04/2013	Dophin Quest, Waikoloa, Hawaii	19.9255	-155.889	Female	28	GSM2602140
83H1_061213	12/06/2013	Dophin Quest, Waikoloa, Hawaii	19.9255	-155.889	Male	5	GSM2602141

Table 3. Assembly and annotation statistics.

	Beluga Whale	Bottlenose Dolphin
# reads	79,897,657	104,708,450
%GC	48.78	49.22
# transcripts	91,325	56,357
# genes	54,582	36,596
# single transcript genes	45,987 (84.3%)	30,544 (83.5%)
RMBT	87.50%	90.60%
N50	2,588	2,293
Min. contig length	424	424
Max. contig length	23,817	12,438
Mean contig length	1,707	1,598
Median contig length	1,113	1,121
# blast hits	52,968 (58.0%)	35,361 (62.7%)
# mapped	44,819 (49.1%)	30,012 (53.3%)
# annotated	45,512 (49.9%)	30,731 (54.5%)
CEGMA	94.76% (F)/98.79% (P)	93.55% (F)/ 97.18% (P)
BUSCO	72.2% (F)/5.8% (P)/32.8% (D)	60% (F)/8% (P)/ 25%(D)

RMBT: reads mapping back to transcripts, F: full length, P: partial length, D: duplicated

Table 4. Significantly enriched KEGG pathways in the dolphin and/or beluga blood transcriptomes, relative to the set of human protein coding genes (FDR < 0.01).

			Beluga			Dolphin		
KEGG ID	**KEGG Pathway**	**# human genes**	**# genes**	**# expected genes**	**FDR**	**# genes**	**# expected genes**	**FDR**
hsa04120	Ubiquitin mediated proteolysis	137	122	78.13	6.73E-14	100	61.76	2.13E-09
hsa03010	Ribosome	135	117	76.99	9.02E-12	113	60.86	0.00E+00
hsa03040	Spliceosome	133	115	75.85	1.42E-11	93	59.95	1.36E-07
hsa04142	Lysosome	123	107	70.14	2.89E-11	93	55.45	5.49E-10
hsa04662	B cell receptor signaling pathway	73	69	41.63	2.89E-11	58	32.91	6.13E-08
hsa04660	T cell receptor signaling pathway	105	93	59.88	7.03E-11	79	47.33	1.21E-08
hsa04210	Apoptosis	140	118	79.84	1.33E-10	97	63.11	1.36E-07
hsa04141	Protein processing in endoplasmic reticulum	166	133	94.67	6.35E-09	115	74.83	9.91E-09
hsa05169	Epstein-Barr virus infection	204	159	116.34	6.35E-09	135	91.96	2.58E-08
hsa04110	Cell cycle	124	103	70.72	1.34E-08	75	55.90	1.92E-03
hsa00970	Aminoacyl-tRNA biosynthesis	44	43	25.09	1.59E-08	35	19.83	3.16E-05
hsa04144	Endocytosis	260	194	148.27	3.37E-08	149	117.20	2.72E-04
hsa04722	Neurotrophin signaling pathway	121	100	69.00	3.37E-08	80	54.54	2.42E-05
hsa05221	Acute myeloid leukemia	57	53	32.51	3.37E-08	39	25.69	1.69E-03
hsa03018	RNA degradation	77	68	43.91	4.29E-08	62	34.71	9.91E-09
hsa03013	RNA transport	166	130	94.67	9.26E-08	111	74.83	2.35E-07
hsa05131	Shigellosis	65	58	37.07	2.49E-07	48	29.30	2.42E-05
hsa03420	Nucleotide excision repair	47	44	26.80	4.03E-07	39	21.19	1.96E-06
hsa05211	Renal cell carcinoma	67	59	38.21	5.24E-07	43	30.20	5.41E-03
hsa05161	Hepatitis B	146	114	83.26	8.97E-07	93	65.81	3.85E-05
hsa04380	Osteoclast differentiation	132	104	75.28	1.35E-06	87	59.50	1.32E-05
hsa04666	Fc gamma R-mediated phagocytosis	93	77	53.04	1.35E-06	63	41.92	7.52E-05
hsa04910	Insulin signaling pathway	140	109	79.84	2.01E-06	80	63.11	1.03E-02
hsa04064	NF-kappa B signaling pathway	95	78	54.18	2.08E-06	67	42.82	6.30E-06
hsa04071	Sphingolipid signaling pathway	120	95	68.43	2.82E-06	80	54.09	1.65E-05
hsa04659	Th17 cell differentiation	107	86	61.02	2.91E-06	77	48.23	3.23E-07
hsa05212	Pancreatic cancer	66	57	37.64	2.91E-06	44	29.75	1.72E-03
hsa04668	TNF signaling pathway	110	88	62.73	2.93E-06	70	49.59	3.91E-04
hsa05220	Chronic myeloid leukemia	73	62	41.63	2.96E-06	47	32.91	3.16E-03
hsa05152	Tuberculosis	179	134	102.08	3.95E-06	112	80.69	1.79E-05
hsa04658	Th1 and Th2 cell differentiation	92	75	52.47	4.93E-06	66	41.47	3.06E-06
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	151	115	86.11	5.43E-06	100	68.07	1.99E-06
hsa03030	DNA replication	36	34	20.53	5.57E-06	29	16.23	1.11E-04
hsa05168	Herpes simplex infection	185	137	105.50	7.56E-06	116	83.39	1.19E-05
hsa04621	NOD-like receptor signaling pathway	170	127	96.95	8.29E-06	110	76.63	2.77E-06
hsa05166	HTLV-I infection	258	184	147.13	8.43E-06	146	116.30	6.22E-04
hsa05162	Measles	136	104	77.56	1.13E-05	89	61.31	1.51E-05
hsa05210	Colorectal cancer	62	53	35.36	1.13E-05	43	27.95	5.64E-04
hsa00240	Pyrimidine metabolism	105	83	59.88	1.18E-05	65	47.33	1.91E-03
hsa04068	FoxO signaling pathway	134	102	76.42	1.98E-05	85	60.40	1.03E-04
hsa05164	Influenza A	175	129	99.80	1.99E-05	107	78.89	9.64E-05
hsa05010	Alzheimer's disease	171	126	97.52	2.63E-05	112	77.08	9.83E-07
hsa00520	Amino sugar and nucleotide sugar metabolism	48	42	27.37	3.55E-05	33	21.64	3.57E-03
hsa03008	Ribosome biogenesis in eukaryotes	82	66	46.76	4.01E-05	60	36.96	3.36E-06
hsa05145	Toxoplasmosis	118	90	67.29	5.92E-05	75	53.19	2.69E-04
hsa01100	Metabolic pathways	1266	790	721.98	6.84E-05	640	570.69	8.04E-05
hsa04664	Fc epsilon RI signaling pathway	70	57	39.92	8.39E-05	48	31.55	3.75E-04
hsa04130	SNARE interactions in vesicular transport	34	31	19.39	8.99E-05	28	15.33	7.93E-05
hsa04620	Toll-like receptor signaling pathway	106	81	60.45	1.30E-04	69	47.78	1.79E-04
hsa00280	Valine, leucine and isoleucine degradation	48	41	27.37	1.39E-04	34	21.64	1.53E-03
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	68	55	38.78	1.55E-04	47	30.65	3.34E-04
hsa04370	VEGF signaling pathway	61	50	34.79	1.78E-04	41	27.50	2.00E-03
hsa05142	Chagas disease (American trypanosomiasis)	104	79	59.31	2.21E-04	66	46.88	6.57E-04
hsa03022	Basal transcription factors	45	38	25.66	4.14E-04	36	20.29	1.97E-05
hsa05132	Salmonella infection	86	66	49.04	4.88E-04	58	38.77	1.71E-04
hsa03410	Base excision repair	33	29	18.82	6.52E-04	24	14.88	5.33E-03
hsa04150	mTOR signaling pathway	154	110	87.82	6.52E-04	90	69.42	2.55E-03
hsa04062	Chemokine signaling pathway	187	131	106.64	6.54E-04	104	84.30	9.11E-03
hsa05203	Viral carcinogenesis	205	142	116.91	7.89E-04	116	92.41	2.55E-03
hsa05134	Legionellosis	55	44	31.37	1.21E-03	37	24.79	3.38E-03
hsa05140	Leishmaniasis	73	56	41.63	1.38E-03	48	32.91	1.59E-03
hsa03060	Protein export	23	21	13.12	1.52E-03	20	10.37	2.72E-04
hsa03430	Mismatch repair	23	21	13.12	1.52E-03	20	10.37	2.72E-04
hsa00310	Lysine degradation	59	46	33.65	2.23E-03	38	26.60	8.94E-03
hsa05133	Pertussis	76	57	43.34	2.96E-03	47	34.26	9.58E-03
hsa03020	RNA polymerase	32	27	18.25	3.40E-03	23	14.42	8.50E-03
hsa04010	MAPK signaling pathway	255	169	145.42	4.55E-03	138	114.95	8.51E-03
hsa03015	mRNA surveillance pathway	91	66	51.90	5.14E-03	57	41.02	2.55E-03
hsa05016	Huntington's disease	193	130	110.06	6.03E-03	117	87.00	7.52E-05
hsa04622	RIG-I-like receptor signaling pathway	70	52	39.92	6.41E-03	45	31.55	4.10E-03
hsa00510	N-Glycan biosynthesis	49	38	27.94	6.86E-03	36	22.09	3.34E-04
hsa00020	Citrate cycle (TCA cycle)	30				23	13.52	2.19E-03
hsa04115	p53 signaling pathway	69	56	39.35	1.16E-04			
hsa01524	Platinum drug resistance	75	59	42.77	3.44E-04			
hsa04070	Phosphatidylinositol signaling system	99	75	56.46	3.78E-04			
hsa01522	Endocrine resistance	98	74	55.89	4.88E-04			
hsa00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	25	23	14.26	6.82E-04			
hsa05213	Endometrial cancer	52	42	29.65	1.15E-03			
hsa03460	Fanconi anemia pathway	55	44	31.37	1.21E-03			
hsa04211	Longevity regulating pathway	94	70	53.61	1.31E-03			
hsa01212	Fatty acid metabolism	48	39	27.37	1.38E-03			
hsa04330	Notch signaling pathway	48	39	27.37	1.38E-03			
hsa01521	EGFR tyrosine kinase inhibitor resistance	81	61	46.19	1.72E-03			
hsa05100	Bacterial invasion of epithelial cells	78	59	44.48	1.72E-03			
hsa05214	Glioma	66	51	37.64	1.72E-03			
hsa05215	Prostate cancer	89	66	50.76	2.09E-03			
hsa04012	ErbB signaling pathway	88	65	50.18	2.70E-03			
hsa00562	Inositol phosphate metabolism	71	53	40.49	5.09E-03			
hsa00230	Purine metabolism	175	119	99.80	5.61E-03			
hsa04914	Progesterone-mediated oocyte maturation	98	70	55.89	6.86E-03			
hsa05222	Small cell lung cancer	86	62	49.04	8.34E-03			

Table 5. Top 100 most highly expressed genes in the beluga whale blood transcriptome.

Gene ID	Top Hit Description[1]	Top Hit E-Value	Function/Process	EntrezGene ID	Average TPM[2]
TR28281|c0_g1	tr|D9YZU5|D9YZU5 Beta-globin	1.3E-88	oxygen/heme/iron binding/transport/homeostasis	3043	137115.24
TR7616|c0_g1	tr|D1MGQ2|D1MGQ2 Alpha-2 globin chain	1.5E-85	oxygen/heme/iron binding/transport/homeostasis	3040	62725.18
TR7616|c0_g2	tr|D1MGQ2|D1MGQ2 Alpha-2 globin chain	1.5E-85	oxygen/heme/iron binding/transport/homeostasis	3040	55946.56
TR39022|c0_g2	tr|A2RQE1|A2RQE1 MHC class I antigen (Fragment)	2.1E-122	immune/inflammatory	3106	13920.59
TR38013|c0_g1	tr|Q9NY06|Q9NY06 Integral membrane transporter protein	1.4E-14			13308.01
TR1378|c0_g1					11933.05
TR35614|c0_g5	tr|Q29R63|Q29R63 Uncharacterized protein	9.9E-22			9984.49
TR40068|c0_g1	sp|P63261|ACTG Actin cytoplasmic 2	<1E-180	cytoskeletal	71	5849.46
TR10391|c0_g1					4975.87
TR39022|c0_g1	tr|O19577|O19577 MHC class I antigen (Fragment)	4.3E-44	immune/inflammatory	3106	4607.87
TR35614|c0_g2	tr|Q29R63|Q29R63 Uncharacterized protein	9.9E-22			4315.24
TR39022|c0_g3	tr|F5BX13|F5BX13 MHC class I antigen (Fragment)	5.9E-107	immune/inflammatory	3106	4247.71
TR1743|c0_g2	sp|P61769|B2MG Beta-2-microglobulin	5.7E-47	immune/inflammatory	567	4223.57
TR13427|c0_g2	tr|U6FVB0|U6FVB0 Tyrosine-protein kinase receptor	1.7E-108	immune/inflammatory		3820.77
TR5823|c0_g1	tr|D6W5K2|D6W5K2 Thymosin beta 10 isoform CRA_a (Fragment)	2.3E-24	cytoskeletal	9168	3606.26
TR38008|c0_g1	tr|B2R4C5|B2R4C5 C-type lysozyme	7.3E-79	immune/inflammatory	4069	3141.00
TR23546|c0_g1	tr|Q6NS36|Q6NS36 Ferritin (Fragment)	7.8E-137	oxygen/heme/iron binding/transport/homeostasis	2495	2955.86
TR12254|c0_g1	sp|P06702|S10A9 Protein S100-A9	3.3E-30	immune/inflammatory	6280	2820.38
TR34559|c0_g1	sp|P80511|S10AC Protein S100-A12	1.3E-33	immune/inflammatory	6283	2661.78
TR23255|c6_g1	tr|Q6IPQ0|Q6IPQ0 IGL@ protein	1.4E-55	immune/inflammatory	3535	2648.43
TR23593|c0_g1	sp|P25398|RS12 40S ribosomal protein S12	6.1E-90	ribosomal	6206	2595.32
TR39092|c0_g1	tr|Q0P5Q0|Q0P5Q0 TMSB4X protein (Fragment)	1.2E-11	cytoskeletal; RNA binding	7114	2407.67
TR34432|c0_g1	tr|Q5W0H4|Q5W0H4 Translationally-controlled tumor protein	3.9E-122		7178	2216.47
TR34463|c0_g1	sp|P05109|S10A8 Protein S100-A8	2.6E-42	immune/inflammatory; cytoskeletal	6279	2194.50
TR34457|c0_g1	tr|Q6VIB9|Q6VIB9 Cytochrome c oxidase subunit 3	4.5E-127	respiratory electron transport		2179.91
TR2629|c0_g1					2128.47
TR38003|c0_g1					2080.06
TR14578|c0_g1	tr|A0N0R1|A0N0R1 C-C motif chemokine	8.9E-23	immune/inflammatory	6348	1978.15
TR34599|c0_g1	sp|P02792|FRIL Ferritin light chain	5.7E-108	oxygen/heme/iron binding/transport/homeostasis	2512	1875.46
TR23194|c0_g1	tr|Q6IPS9|Q6IPS9 Elongation factor 1-alpha	<1E-180	ribosomal; translation; cytokine production	1915	1814.17
TR14812|c0_g1	tr|A0AUP3|A0AUP3 LOC339290 protein (Fragment)	2.9E-16			1697.83
TR34505|c0_g1	tr|A0A140TA30|A0A140TA30 HLA class II histocompatibility antigen DRB1-16 beta chain	2E-135	immune/inflammatory	3123	1576.90
TR23252|c0_g1	tr|Q6MZU6|Q6MZU6 Putative uncharacterized protein DKFZp686C15213	2.8E-148	immune/inflammatory		1576.48
TR27300|c0_g1	sp|P62277|RS13 40S ribosomal protein S13	4.2E-106	ribosomal	6207	1535.69
TR1079|c0_g1	tr|V9HVZ4|V9HVZ4 Glyceraldehyde-3-phosphate dehydrogenase	<1E-180	glycolosis; oxioreductase		1434.12
TR25059|c0_g1	sp|P00746|CFAD Complement factor D	1.7E-129	immune/inflammatory	1675	1397.15
TR21884|c0_g1	tr|B0UXD8|B0UXD8 MHC class II antigen	8.5E-127	immune/inflammatory		1377.76
TR29864|c0_g1	tr|A0A024R261|A0A024R261 HCG24487 isoform CRA_c	4.7E-123	ribosomal		1316.53
TR14619|c0_g1	sp|P08708|RS17 40S ribosomal protein S17	6.7E-94	ribosomal	6218	1268.24
TR14822|c0_g1	sp|P06703|S10A6 Protein S100-A6	4.7E-49	cytoskeletal	6277	1256.50
TR38952|c0_g1	tr|Q7L4Q3|Q7L4Q3 Glutathione peroxidase	3.7E-127	oxygen/heme/iron binding/transport/homeostasis	2876	1238.93
TR3535|c0_g1	sp|P62081|RS7 40S ribosomal protein S7	5E-137	ribosomal	6201	1200.04
TR14803|c0_g1	tr|A0A024QZD1|A0A024QZD1 Ribosomal protein L18 isoform CRA_c	3E-101	ribosomal	6141	1198.85
TR5034|c0_g1	sp|P62269|RS18 40S ribosomal protein S18	1.3E-90	ribosomal	6222	1170.32
TR1265|c0_g1	tr|Q5TBB8|Q5TBB8 Lysosomal associated multispanning membrane protein 5 isoform CRA_a	3.3E-147		7805	1164.82
TR36613|c0_g1	tr|W8DBX9|W8DBX9 Cytochrome c oxidase subunit 2	4.5E-76	respiratory electron transport		1130.00
TR15252|c0_g1	sp|P62280|RS11 40S ribosomal protein S11	1.6E-113	ribosomal	6205	1116.56
TR10099|c0_g1	sp|P10124|SRGN Serglycin	2.6E-44	immune/inflammatory	5552	1108.04
TR22989|c0_g1	tr|B2R491|B2R491 40S ribosomal protein S4	<1E-180	ribosomal	6191	1101.88
TR3068|c0_g1	sp|P52209|6PGD 6-phosphogluconate dehydrogenase decarboxylating	<1E-180	oxioreductase	5226	1099.31
TR15583|c0_g1	sp|P15880|RS2 40S ribosomal protein S2	8.1E-175	ribosomal	6187	1064.55
TR26459|c0_g1	sp|P62750|RL23A 60S ribosomal protein L23a	1.8E-85	ribosomal	6147	1052.55
TR3284|c39_g2	tr|Q14754|Q14754 DNA for LINE-1 transposable element ORFI and II	3.8E-112			1045.82
TR12694|c0_g1	tr|G8JLA2|G8JLA2 Myosin light polypeptide 6	1.9E-106	calcium ion binding	4637	1025.98
TR2443|c0_g1	tr|A0A024R611|A0A024R611 Coronin	<1E-180	immune/inflammatory; cytoskeletal	11151	1020.95
TR33177|c0_g1	tr|A0A142L067|A0A142L067 MHC class II antigen (Fragment)	7E-98	immune/inflammatory	3119	1020.58
TR20398|c0_g1	tr|F6U211|F6U211 40S ribosomal protein S10	1.3E-102	ribosomal	6204	1006.43
TR39359|c0_g1	sp|P42766|RL35 60S ribosomal protein L35	6.7E-66	ribosomal	11224	1003.63
TR24545|c0_g1	sp|P23396|RS3 40S ribosomal protein S3	8.6E-176	ribosomal	6188	955.17
TR23409|c0_g1					940.81
TR34328|c0_g1	tr|A0A024R4B7|A0A024R4B7 Thioredoxin interacting protein isoform CRA_a	<1E-180	stress response; cell proliferation	10628	931.61
TR7571|c0_g1	sp|P54368|OAZ1 Ornithine decarboxylase antizyme 1	9.9E-100	protein catabolism	4946	922.38
TR17784|c0_g1	tr|Q3MIH3|Q3MIH3 Ubiquitin A-52 residue ribosomal protein fusion product 1	1.7E-84	ribosomal	7311	917.49
TR5100|c0_g1	sp|P42677|RS27 40S ribosomal protein S27	5.7E-44	ribosomal	6232	916.88
TR35614|c0_g1	tr|Q29R63|Q29R63 Uncharacterized protein	9.9E-22			902.70
TR7025|c0_g1					886.21
TR20048|c0_g1	sp|P27635|RL10 60S ribosomal protein L10	1.5E-157	ribosomal	6134	878.66
TR1550|c0_g1	tr|E2J0J0|E2J0J0 NADH-ubiquinone oxidoreductase chain 4	6.3E-139	mitochondrial electron transport		871.06
TR35092|c0_g1	sp|P14780|MMP9 Matrix metalloproteinase-9	<1E-180	metal binding; extracellular matrix structure	4318	870.54
TR3475|c0_g1	sp|P61247|RS3A 40S ribosomal protein S3a	<1E-180	ribosomal	6189	863.24
TR2712|c0_g1	sp|P62249|RS16 40S ribosomal protein S16	1.6E-101	ribosomal	6217	856.57
TR11472|c0_g1	tr|A0A140VK56|A0A140VK56 Testicular secretory protein Li 56	<1E-180	carbohydrate metabolism		835.06
TR22799|c0_g1	tr|Q8WVX7|Q8WVX7 Ribosomal protein S19 (Fragment)	5.2E-103	ribosomal		832.72
TR1529|c0_g1	tr|Q59GY2|Q59GY2 Ribosomal protein L4 variant (Fragment)	<1E-180	ribosomal		831.58
TR23141|c0_g2	tr|A0A087WUS0|A0A087WUS0 40S ribosomal protein S24	1.1E-64	ribosomal	6229	820.89
TR10287|c0_g2	tr|V9HWN7|V9HWN7 Fructose-bisphosphate aldolase	<1E-180	carbohydrate metabolism		816.27
TR1282|c0_g1	tr|Q53Z07|Q53Z07 NPC-A-16	3.1E-129	ribosomal	6133	802.76
TR15081|c0_g1	tr|A0A024R4Q8|A0A024R4Q8 Ribosomal protein S5 isoform CRA_a	2.1E-150	ribosomal	6193	792.72
TR10303|c0_g1					774.49
TR3251|c9_g29	tr|Q5Y7H0|Q5Y7H0 MHC class II antigen	1.5E-119	immune/inflammatory	3117	769.16
TR36049|c0_g1	tr|C9JIZ6|C9JIZ6 Prosaposin	<1E-180	sphingolipid metabolism	5660	765.82
TR34596|c0_g1	tr|V9HW01|V9HW01 Epididymis secretory protein Li 310	3.5E-77	ribosomal		735.81
TR34466|c0_g1	sp|P40429|RL13A 60S ribosomal protein L13a	1.4E-129	ribosomal	23521	732.57
TR25847|c0_g1	sp|Q14019|COTL1 Coactosin-like protein	8.7E-91	immune/inflammatory; cytoskeletal	23406	729.71
TR2224|c0_g1	tr|B2R4P9|B2R4P9 Histone H3	2.2E-93	DNA binding	3020	717.90
TR33883|c0_g1	tr|A0A024R2P0|A0A024R2P0 40S ribosomal protein SA	<1E-180	ribosomal	3921	699.31
TR23415|c0_g1	tr|V9HWE1|V9HWE1 Epididymis luminal protein 113	<1E-180	cytoskeletal		697.99
TR23331|c0_g1	tr|V9HWI5|V9HWI5 Cofilin 1 (Non-muscle) isoform CRA_b	2.4E-110	cytoskeletal		694.53
TR14854|c0_g1	tr|A0A024R4M0|A0A024R4M0 40S ribosomal protein S9	4.4E-127	ribosomal	6203	693.36
TR27343|c0_g1	sp|P46776|RL27A 60S ribosomal protein L27a	1.4E-82	ribosomal	6157	692.49
TR37954|c0_g1	sp|P62263|RS14 40S ribosomal protein S14	8.3E-97	ribosomal	6208	686.12
TR32031|c0_g1	tr|J3QRS3|J3QRS3 Myosin regulatory light chain 12A	1.3E-111	calcium ion binding; cytoskeletal	10627	667.61
TR32031|c0_g1	sp|O14950|ML12B Myosin regulatory light chain 12B	2.7E-103	calcium ion binding; cytoskeletal	103910	667.61
TR14735|c0_g1	tr|V9HW22|V9HW22 Epididymis luminal protein 33	<1E-180	stress response; cell proliferation		661.92
TR34328|c0_g2	tr|A0A024R4B7|A0A024R4B7 Thioredoxin interacting protein isoform CRA_a	<1E-180	stress response; cell proliferation	10628	661.49
TR15850|c0_g2	sp|P37802|TAGL2 Transgelin-2	2E-140	cytoskeletal	8407	661.23
TR31675|c0_g1	sp|Q02543|RL18A 60S ribosomal protein L18a	1.1E-127	ribosomal	6142	654.19
TR3837|c0_g1	sp|P62917|RL8 60S ribosomal protein L8	3E-178	ribosomal	6132	637.54
TR15263|c0_g2	sp|P07737|PROF1 Profilin-1	1.7E-80	cytoskeletal	5216	629.49
TR4356|c1_g1	tr|A0A087WYJ9|A0A087WYJ9 Ig mu chain C region	<1E-180	immune/inflammatory	3507	625.60

[1]Description of top hit from blastx searches of the human subset of the UniProt-SwissProt database

[2]TPM: transcripts per million

Table 6. Expression of transcripts involved in cold tolerance in beluga whale and bottlenose dolphin.

	Beluga		Dolphin		Ratio of
Description[1]	Sequence	Ave TPM[2]	Sequence	Ave TPM[2]	Expression[3]
Q6FG41 FOS	TR13269|c0_g1	129.84	TR3531|c0_g1	2.26	57.50
GLPK Glycerol kinase	TR4326|c0_g4	30.46	TR9261|c0_g2	1.02	29.86
			TR9261|c0_g1	15.61	1.95
X2BQ60 DGIC Down-regulated in gastrointestinal cancer protein	TR11837|c0_g1	229.94	TR25283|c0_g1	15.86	14.49
			TR25283|c0_g2	150.66	1.53
METRNL Meteorin	TR11280|c0_g2	58.85	TR9140|c0_g1	8.93	6.59
FOXO1 Forkhead box O1	TR35233|c0_g1	12.32	TR10951|c0_g1	1.87	6.59
ACADL Long-chain specific acyl- mitochondrial	TR20138|c0_g1	22.36	TR16032|c0_g1	3.86	5.80
E2AK3 Eukaryotic translation initiation factor 2-alpha kinase 3	TR29274|c0_g2	30.20	TR18766|c0_g1	5.27	5.73
RNF34 E3 ubiquitin- ligase RNF34	TR11302|c0_g1	17.47	TR22038|c0_g1	3.61	4.83
A8KA82 (Hsp40) subfamily member 3	TR29293|c0_g1	22.76	TR25589|c0_g1	4.81	4.74
AAPK15 -AMP-activated kinase catalytic subunit alpha-1	TR39056|c0_g1	24.42	TR25244|c0_g1	5.37	4.55
PCSK1	TR4840|c0_g2	61.97	TR4027|c0_g1	14.26	4.35
	TR4840|c0_g1	4.65			0.33
IMDH1 Inosine-5 -monophosphate dehydrogenase 1	TR24622|c0_g1	4.41	TR15522|c0_g1	1.07	4.13
A0A024RBC7 Calcium-transporting ATPase	TR6827|c0_g2	36.97	TR10324|c4_g1	10.75	3.44
IKBA NF-kappa-B inhibitor alpha	TR15849|c0_g1	74.56	TR5687|c0_g1	21.82	3.42
STAT3 Signal transducer and activator of transcription 3	TR35012|c0_g1	76.65	TR27174|c0_g1	24.02	3.19
A0A024R3X4 Heat shock 60kDa 1 (Chaperonin) isoform CRA_a	TR29009|c0_g1	52.13	TR13447|c0_g1	18.42	2.83
S27A1 Long-chain fatty acid transport 1	TR21434|c0_g1	20.34	TR25223|c0_g1	7.82	2.60
CASP8 Caspase-8	TR18382|c0_g1	34.98	TR15693|c0_g1	13.63	2.57
CIDEA Cell death activator CIDE-A			TR25578|c0_g1	13.74	2.55
THA Thyroid hormone receptor alpha	TR18839|c0_g1	10.10	TR12302|c0_g1	4.99	2.02
GMPR1 GMP reductase 1	TR38108|c0_g2	12.47	TR12541|c0_g1	6.20	2.01
			TR14001|c0_g1	17.81	0.70
A0A024R6B5 Heat shock 70kDa isoform CRA_a	TR6557|c0_g1	2.83	TR1378|c0_g1	2.69	1.05
ACADV Very long-chain specific acyl-coA dehydrogenase mitochondrial	TR17658|c0_g1	40.17			-
LRP11 Low-density lipo receptor-related 11	TR27647|c0_g2	15.77			-
A0A024R9Q9 Peroxisome proliferative activated coactivator isoform CRA_a	TR29637|c0_g2	5.04			-
	TR13727|c0_g1	2.56			-
ACO11 Acyl-coenzyme A thioesterase 11	TR25144|c0_g1	4.88			-
E2AK4 eIF-2-alpha kinase GCN2	TR38676|c0_g1	4.63			-
	TR41271|c0_g2	3.05			-
ZN516 Zinc finger 516	TR10489|c0_g1	1.85			-
A0A024RDA4 C-X-C motif chemokine	TR27317|c0_g1	1.56			-

[1]Description of top hit from blastx searches of the human subset of the UniProt-SwissProt database

[2]TPM: transcripts per million

[3]Beluga TPM/Dolphin TPM