

2012

Health status, infection and disease in California sea lions (*Zalophus californianus*) studied using a canine microarray platform and machine-learning approaches

Annalaura Mancia
Medical University of South Carolina

James C. Ryan
Center for Coastal Environmental Health and Biomolecular Research

Robert W. Chapman
Hollings Marine Laboratory

Qingzhong Wu
Hollings Marine Laboratory

Gregory W. Warr
Medical University of South Carolina

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/usdeptcommercepub>

 Part of the [Environmental Sciences Commons](#)

Mancia, Annalaura; Ryan, James C.; Chapman, Robert W.; Wu, Qingzhong; Warr, Gregory W.; Gulland, Frances M.D.; and Van Dolah, Frances M., "Health status, infection and disease in California sea lions (*Zalophus californianus*) studied using a canine microarray platform and machine-learning approaches" (2012). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 348.

<http://digitalcommons.unl.edu/usdeptcommercepub/348>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Annalaura Mancia, James C. Ryan, Robert W. Chapman, Qingzhong Wu, Gregory W. Warr, Frances M.D. Gulland, and Frances M. Van Dolah



Health status, infection and disease in California sea lions (*Zalophus californianus*) studied using a canine microarray platform and machine-learning approaches

Annalaura Mancía^{a,*}, James C. Ryan^{b,c}, Robert W. Chapman^d, Qingzhong Wu^c, Gregory W. Warr^{a,1}, Frances M.D. Gulland^e, Frances M. Van Dolah^{b,c}

^a Medical University of South Carolina, Marine Biomedicine and Environmental Sciences Center, Hollings Marine Laboratory, 331 Ft Johnson Rd., Charleston, SC 29412, USA

^b NOAA/Marine Biotoxins Program, Center for Coastal Environmental Health and Biomolecular Research, 219 Ft Johnson Rd., Charleston, SC 29412, USA

^c NOAA Center of Excellence in Oceans and Human Health, Hollings Marine Laboratory, 331 Ft Johnson Rd., Charleston, SC 29412, USA

^d The South Carolina Department of Natural Resources, Hollings Marine Laboratory, 331 Ft Johnson Rd., Charleston, SC 29412, USA

^e The Marine Mammal Center, 2000 Bunker Road, Sausalito, CA 94965, USA

ARTICLE INFO

Article history:

Received 19 May 2011

Revised 21 October 2011

Accepted 22 October 2011

Available online 31 October 2011

Keywords:

California sea lion

Cross-hybridization

Microarray

Domoic acid toxicity

Leptospirosis

Blood transcriptomics

ABSTRACT

Conservation biologists face many challenges in assessing health, immune status and infectious diseases in protected species. These challenges include unpredictable sample populations, diverse genetic and environmental backgrounds of the animals, as well as the practical, legal and ethical issues involved in experimentation. The use of whole genome scale transcriptomics with animal samples obtained in a minimally invasive manner is an approach that shows promise for health assessment. In this study we assessed the utility of a microarray to identify changes in gene expression predictive of health status by interrogating blood samples from California sea lions (*Zalophus californianus*) in rehabilitation. A custom microarray was developed from the commercially available dog microarray (*Canis familiaris*) by selecting probes that demonstrated reliable cross-hybridization with RNA in sea lion blood. This custom microarray was used for the analysis of RNA from 73 sea lion blood samples, from animals with a broad spectrum of health changes. Both traditional classifying techniques and newer artificial neural network approaches correctly classified sea lions with respect to health status, primarily distinguishing between leptospirosis infection and domoic acid exposure. Real time PCR validation for a small set of genes, followed by sequencing, showed good correlation with array results and high identity (96–98%) between the dog and sea lion sequences. This approach to health status classification shows promise for disease identification in a clinical setting, and assessment of health status of wildlife.

Published by Elsevier Ltd.

1. Introduction

The use of global scale (or “-omic”) data sets is widely accepted as a route to the better understanding of health and disease in both human medicine and in agricultural and veterinary practice. However, the utility of such approaches in conservation biology is limited by several factors. First, little is known of the basic biology (including the genome and proteome) of most species of interest. Even in the few cases for which genomes have been sequenced and transcriptomes described (such as the bottlenose dolphin *Tursiops truncatus* and the American alligator *Alligator mississippiensis*) these data sets are usually not complemented by adequate

correlative information on, for example, physiology and disease status. Second, for ethical and legal reasons it is often difficult (and in some cases logistically impossible) to conduct carefully designed experiments on protected, threatened or endangered species. Third, wild populations are often heterogeneous in their genetic structure, habitats and life style. Despite these limitations, studies utilizing large data sets (especially transcriptomic information) have been attempted, with some success, in the context of both the ecology and the health status of wild populations (Mancia et al., 2010, 2008).

One animal of particular interest, due to its ability to act as a sentinel species for coastal habitats, is the California sea lion (*Zalophus californianus*), a protected marine mammal inhabiting the western coast of North America. While the males range from the US/Mexico border to the Alaska panhandle, female sea lions reside primarily in central California sites around their breeding grounds (Melin, 2002; Riedman, 1991). In 2009 alone, over 600 stranded sea lions were admitted to The Marine Mammal Center, in Sausalito, CA for rehabilitation. Some of the strandings were due to

* Corresponding author. Present address: Department of Biology and Evolution, University of Ferrara, via L. Borsari, 46, Ferrara, FE, Italy. Tel.: +39 0532 455704; fax: +39 0532 455715.

E-mail address: annalaura.mancia@unife.it (A. Mancía).

¹ Present address: Division of Molecular and Cellular Biosciences, National Science Foundation, Arlington, VA 22230, USA.

malnutrition, cancer, pneumonia, entanglement in debris or fishing gear and gunshots, but the two most common causes of stranding are reported to be domoic acid toxicosis (DAT) and leptospirosis infection (Lepto) (Greig et al., 2005). DAT is a condition caused by ingestion of domoic acid, a potent neurotoxin produced by diatom species of the genus *Pseudo-nitzschia*, which is known to cause neuronal necrosis especially in the hippocampus (Ryan et al., 2011; Silvagni et al., 2005; Van Dolah, 2000) and also have direct effects on lymphocytes in vitro (Levin et al., 2010). The other most common cause of stranding and mortality in adult California sea lions is the infectious pathogen *Leptospira*, one of the most widespread bacterial zoonoses in the world, best characterized from terrestrial species (Bharti et al., 2003). *Leptospira* infection, or leptospirosis, in sea lions is characterized by bacteria colonizing the liver and kidneys of the host, causing acute renal failure and death.

Although extensive studies to better understand causes and epidemiology of sea lion strandings have been conducted, a test to identify a specific disease at an early stage, thus allowing the appropriate treatment and therapy of the animal, has yet to be developed. Additionally, rapid identification of zoonotic disease and toxicosis from sea food diets in this sentinel species may provide important warnings for health threats to the millions of people that inhabit and recreate in coastal areas.

Because the two most common causes of strandings are likely to elicit a response from the immune system, we studied the gene expression in blood from stranded animals with well characterized clinical diagnoses. Here we report a study testing the ability of transcriptomic signatures, captured using a custom-designed dog microarray, to accurately classify stranded sea lions in terms of infection, disease, intoxication or other causes of compromised health status.

2. Material and methods

2.1. Animals, tissue samples and RNA

Blood samples were collected from the caudal gluteal vein into PaxGene tubes (Qiagen, Valencia, CA) from 100 stranded California sea lions in rehabilitation at The Marine Mammal Center (TMMC), in Sausalito, California. In addition to blood, bone marrow, brain, liver and kidney were collected from 3 sea lions (not used for blood sampling) euthanized due to poor prognosis at TMMC and shipped in RNAlater (Applied Biosystems/Ambion, Austin, TX) to the Hollings Marine Laboratory. All samples were stored at -80°C prior to shipping. Disease status was determined at the TMMC using clinical signs, serum biochemistry, hematology, diagnostic imaging and post mortem examinations, including histology as described in Greig et al. (2005).

Of the 100 blood samples collected 73 animals were chosen for analysis due to quality and quantity of RNA or age of animal (pups were not used in this study). Of these 73, four animals were considered clinically healthy since their capture was due to nuisance rather than illness (Healthy, group 1), 33 were diagnosed with domoic acid toxicosis (DAT, group 2), 24 had leptospirosis infection (Lepto, group 3) and 12 were considered an outgroup (Outgroup, group 4), e.g. presenting with other non-specific diseases or trauma (cancer, gunshots, pneumonia, etc.). Of the 73 animals, 45 were males, 28 were females (age, disease and gender distribution is shown in Table 1).

RNA from bone marrow, brain, liver and kidney was extracted using silica-membrane RNeasy spin columns (Qiagen, Valencia, CA), DNase-digested on-column, and eluted with 40 μl of nuclease-free water. RNA from blood was extracted using the PaxGene Blood RNA Kit according to manufacturer (Qiagen, Valencia, CA). RNA was quantified using a NanoDrop ND-1000 (Wilmington,

Table 1

Distribution of gender and health classification for animals used in this study.

		Healthy	DAT	Lepto	Outgroup	Total/group
Females	Adults	–	14	1	3	18
	Subadults	–	5	1	3	9
	Juveniles	–	–	–	1	1
	Total F	0	19	2	7	28
Males	Adults	4	5	–	1	10
	Subadults	–	6	12	3	21
	Juveniles	–	3	10	1	14
	Total M	4	14	22	5	45
Total F + M		4	33	24	12	73

DAT, domoic acid toxicosis; Lepto, Leptospirosis infection. Age classes were defined as: Juvenile female (2–3 years), Subadult female (3–5 years), and Adult female (5+ years); Juvenile male (2–4 years), Subadult male (4–8 years), Adult male (8+ years).

DE), qualified on an Agilent 2100 Bioanalyzer (Foster City, CA) and stored at -80°C until needed for gene expression profiling.

2.2. Microarray labeling and hybridization

All RNA labeling and microarray hybridizations were performed according to the manufacturer's instructions in both the Two-Color Microarray-Based Gene Expression Analysis and One-Color Microarray-Based Gene Expression Analysis manuals (Agilent Technologies, Santa Clara, CA). To select sea lion cross-hybridizing probes from the commercial dog microarray, two-color dye techniques with Cy3 and Cy5 labeled CTP were used with the Agilent Quick Amp labeling kit (Agilent Technologies, Santa Clara, CA). For this purpose we used RNA from different tissues for hybridization, as described in the following section. For gene expression profiling, blood RNA from 73 sea lions was hybridized on a custom sea lion $8 \times 15 \text{ K}$ array (see below) after amplification and labeling with only cyanine 3-CTP dye using the Agilent Quick Amp labeling kit. For both instances, 500 ng of RNA was amplified and fluorescently labeled with the appropriate dyes. This amplification product was measured for quantity and dye incorporation using the Nanodrop 1000. One and a half micrograms of amplified, fluorescently labeled RNA was hybridized to the microarray at 65°C . After 17 h, one and/or two-color hybridizations arrays were washed consecutively in solutions of 6X SSPE with 0.005% N-lauroylsarcosine and 0.06X SSPE with 0.005% N-lauroylsarcosine for 1 min each at room temperature. This was followed by a final 15 s wash in Agilent Stabilization and Drying solution.

2.3. Canine and custom sea lion microarray

The Agilent (Santa Clara, CA) 44,000 feature (44 K) canine array was used in pilot hybridizations to determine the cross-reactivity with the sea lion transcriptome. To maximize sea lion specific content on the microarray, RNA from blood, bone marrow, brain, liver and kidney collected from sea lions was hybridized to the array and probes with good sea lion cross-reactivity based on signal intensity and signal to noise ratios, as well as reproducibility, were determined. By removing probes of poor hybridization efficiency, the content of the $4 \times 44 \text{ K}$ feature canine array was reduced to a custom $8 \times 15 \text{ K}$ microarray (8 individual arrays each containing 15,000 probes) and manufactured by Agilent (Santa Clara, CA). The DNA sequences of the probes represented on the custom, sea lion-fitted 15 K array were re-blasted in Blast2go (Conesa and Gotz, 2008) with the last update relative to July 2010.

2.4. Microarray data analysis

Microarrays were imaged on an Agilent microarray scanner, extracted with Agilent Feature Extraction software version A8.5.3,

and data analyzed with Rosetta Resolver 7.0 gene expression analysis system (Rosetta Informatics, Seattle, WA). The microarray build and hybridization data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series Accession Number GSE29497 (www.ncbi.nlm.nih.gov/geo/).

One color gene expression arrays were normalized by removing control and flagged data, applying a trimming function to remove the top and bottom 5% of intensity data, then scaling to the mean intensity. The arrays from each treatment group (Healthy, DAT, Lepto, and Outgroup) were then combined and feature intensities underwent a weighted averaging for each probe using the error model for Agilent data in the Resolver software environment (Weng et al., 2006). These data were then used to build ratios between the 4 groups. These ratio data were initially filtered using a $p < 0.001$ cutoff for significant differential expression between groups. In particular, the DAT vs. Lepto analysis described resulted in the selection of 348 genes at a $p < 0.001$ (Table S1). These genes were then used to drive an SVM (support vector machines) classifier algorithm that selected 110 genes that allowed for complete stratification of animals (Table S2). This gene set was then used for unsupervised clustering using *K*-means with both Euclidean and Pearson metrics, and principal component analysis using *z*-score values.

2.5. Artificial neural network analysis

Raw intensity microarray data for the 73 animal samples hybridized were variance stabilized normalization (VSN) transformed (Huber et al., 2002) and 500 genes showing the greatest statistical significance (lowest *p*-value) in pair wise comparisons of each group were selected using *t*-tests (Table S3). These genes were then used in artificial neural network (ANN) modeling following established protocols (Chapman et al., 2009; Mancia et al., 2010). The comparisons were balanced by randomly selecting roughly 70% of the individuals from the smaller of the groups compared and randomly selecting an equal number of individuals from the larger group. These individuals comprised the training

Table 2
ANN accuracy of discrimination (R^2 values) of diseased status in training set and cross validation sets.

	R^2	Lepto	Outgroup	Healthy
DAT	T	1	1	1
	CV	0.981	0.365	0.535
Lepto	T		1	1
	CV		0.684	1
Outgroup	T			1
	CV			0.979

T, training set (70% of the sample set); CV, Cross Validation set (remaining 30% of the sample set).

Table 3
ANN misclassification probabilities for the cross validation set from ROC curves.

		Lepto	Outgroup	Healthy
DAT	AUC	0.986	0.853	1
	sd	0.005	0.018	0
Lepto	AUC		0.999	1
	sd		0.002	0
Outgroup	AUC			1
	sd			0

ROC, receiver operating characteristic; AUC, area under the curve; sd, standard deviation.

group for the ANNs. The remaining individuals were a cross validation set used to test the robustness of the ANNs training. The entire selection and training process was repeated 20 times for each pair wise comparison. Following training, the R^2 from both training and cross validation data selected were computed (Table 2). In addition, as the comparisons were binary, receiver operating characteristic (ROC) curves were employed to determine the misclassification probabilities. This analysis consisted solely of individuals in the cross validation sets, i.e. not data used to train the ANNs. The area under the curve (AUC) along with standard deviations (sd) is presented in Table 3.

2.6. Quantitative real time PCR (qPCR)

Results from the microarray analysis were tested by measuring mRNA expression of 9 selected genes for 30 different animals by qPCR (animals were selected based on quantity of RNA remaining after microarrays and age class): elongation factor RNA polymerase II, which was used as a control (ELL; Acc. No. XM_001115395), tumor necrosis factor, alpha-induced protein 6 (TNFAIP6; Acc. No. XM_533354), ring finger protein 122 (RNF122; Acc. No. XM_845404), calcium binding atopy-related autoantigen 1 (CBAR-A1; Acc. No. XM_001503799), neuromedin U receptor 1 (NMUR; Acc. No. XM_001495180), neuronal calcium sensor 1 (NCS1; Acc. No. EZ492132), scavenger receptor class B, member 2 (SCARB2; Acc. No. XM_001096458), nesprin-1 (SYNE-1; Acc. No. XM_541154), and phosphoglucomutase 5 (PGM5; Acc. No. XM_533534).

Relative levels of mRNA were determined using qPCR on an ABI7500 (Applied Biosystems, Carlsbad, CA). Given the lack of genomic resources for the California sea lion, we designed gene-specific primers over conserved regions across closely related species using sequences available at NCBI (Table S4). Species used, when possible, were the dog (*Canis familiaris*), panda (*Ailuropoda melanoleuca*), horse (*Equus caballus*), cow (*Bos taurus*), pig (*Sus scrofa*), ferret (*Mustela putorius*), human (*Homo sapiens*), orangutan (*Pongo abeeli*) and monkey (*Macaca mulatta*). Each gene primer set was optimized for efficiency and specificity by running standard curves on pooled cDNA samples from 4 healthy animals resulting in a correlation coefficient $R^2 > 96.5\%$ and efficiency = 92–113% (Amp = 1.92–2.13) (Table S4). Quantitative RT-PCR efficiencies were calculated using the equation: $m = -(1/\log E)$, where m is the slope of the line and E is the efficiency (Dhar et al., 2009). Five-hundred nanogram of total RNA was reverse transcribed to cDNA for each of the 30 samples ($n = 15$ domoic acid toxicosis, 7 leptospirosis, 8, outgroup) using an RNA-to-cDNA Kit (Applied Biosystems, Carlsbad, CA) according to manufacturer's instructions. Optimized qPCR parameters for each gene were determined using diluted cDNA reverse transcribed from 500 ng of total RNA from 4 healthy animals (at 1:5) and 12.5 μ l SyBr Green Master Mix (Applied Biosystems, Carlsbad, CA) to a final volume of 25 μ l were run as follows: 1 cycle of 95 °C for 10 min, 40 cycles of 95 °C for 15 s, 52 °C (or 56 °C) for 40 s, and 72 °C for 40 s. At the end of each RT-PCR reaction, PCR products were subjected to a melt curve analysis. PCR amplicons were also run by capillary gel electrophoresis (Agilent 2100 bioanalyzer) to confirm the presence of a single amplicon.

Negative controls (minus reverse transcriptase) were run for all animals for each gene to confirm the absence of contaminating DNA. Each biological replicate was run in triplicate and the averaged C_t values were normalized to the corresponding measured mRNA C_t value of the housekeeping gene ELL (Zhu et al., 2008), which did not show significant variation across samples in microarray hybridizations. Comparative C_t method of analysis ($2^{-\Delta\Delta C_t}$) was used to determine changes in gene expression between control and treated samples.

2.7. Sequencing

PCR products of four genes were purified, cloned and sequenced (TNFAIP6, RNF122, CBARA1, SCARB2). Fifty microliter PCR reactions from above were purified using the Qiagen QIAquick PCR purification kit (Qiagen, Valencia, CA). These templates were then cloned with a TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Seventy-five microliters from each transformation was spread on LB plates containing 50 µg/ml ampicillin, incubated overnight at 37 °C, and colonies subjected to blue white screening. White colonies were picked and grown overnight in LB broth and purified with a QIAprep spin miniprep kit (Qiagen). Plasmids were subjected to PCR with sequence specific consensus primers to ensure insertion of appropriate sequence. Two positive plasmids from each gene were sequenced with M13 primer in both forward and reverse orientation by SeqWright Inc. (Houston, TX).

3. Results

3.1. A microarray for sea lion studies

Little genomic or transcriptomic information is available for sea lions (984 entries, including the complete mitochondrial sequence, were in the NCBI Nucleotide database for *Zalophus californianus* as of 15 March, 2011), and as an alternative to undertaking a sea lion cDNA sequencing project, a cross species hybridization approach was examined. The California sea lion (family Otariidae) and the domestic dog (*Canis familiaris*, family Canidae) are both members of the suborder Caniformia (www.itis.gov) and thus the potential for using a modified dog microarray to detect sea lion gene expression was explored. RNA from the blood of a single animal was amplified and replicate samples were labeled with either Cy3 or Cy5 dyes, then hybridized to the canine microarray in a self against self format (Fig. 1). Those genes with intensities greater than 10^2 counts for both Cy3 and Cy5 channels were identified and are highlighted in red. Roughly one-third of the 42,000 non-control probes on the canine array hybridized to message expressed in sea lion

blood with high affinity and robust signal to noise ratios. To identify additional sea lion cross-reactive features on the microarray, RNA from bone marrow, brain, liver and kidney were also hybridized to the canine array, and probes with good sea lion cross-reactivity based on signal intensity and signal to noise ratios were determined (data not shown). In addition to the signal intensity cut-offs, we searched the canine array annotation for terms used to describe genes with immune functions (including interleukins and their receptors, cytokines and receptors, chemokines and receptors, immunoglobulins, Toll-like receptors, TCR genes, MHC I & II genes, interferons, etc.). A combination of those canine probes displaying good cross reactivity (signal intensity >100 counts) along with canine immune function genes were used, through the Agilent eArray interface, to design a custom 8 × 15 K dog-for-sea lion microarray. Technical replicate measurements showed high reproducibility (correlation of 98–99%) when using the custom canine array to interrogate sea lion gene expression (Fig. S1).

The new custom dog-for-sea lion 8 × 15 K microarray (platform and datasets can be found at <http://www.ncbi.nlm.nih.gov/geo/> with GEO Series Accession Number GSE29497) was employed for further experiments to investigate the correlation of sea lion transcriptomic profiles with their health status, in terms of infection, immune function, intoxication and other diseases.

3.2. Differential gene expression and classification analysis

Several methods to identify gene expression patterns in California sea lions were assessed with the goal of developing robust diagnostic classifiers that would aid in rapid diagnosis and treatment of stranded animals. First, the animals were placed into one of four groups; (1) DAT, (2) Lepto, (3) Healthy and (4) Outgroup. Then, the data were subjected to a combination of traditional analyses including analysis of variance (ANOVA), 2 dimensional clustering and SVM classifier algorithms.

Results of several attempts to separate all animals into their respective groups based on clinical diagnoses were not always successful. This was due to the Outgroup samples, which had no reproducible gene expression patterns for the $n = 12$ used here,

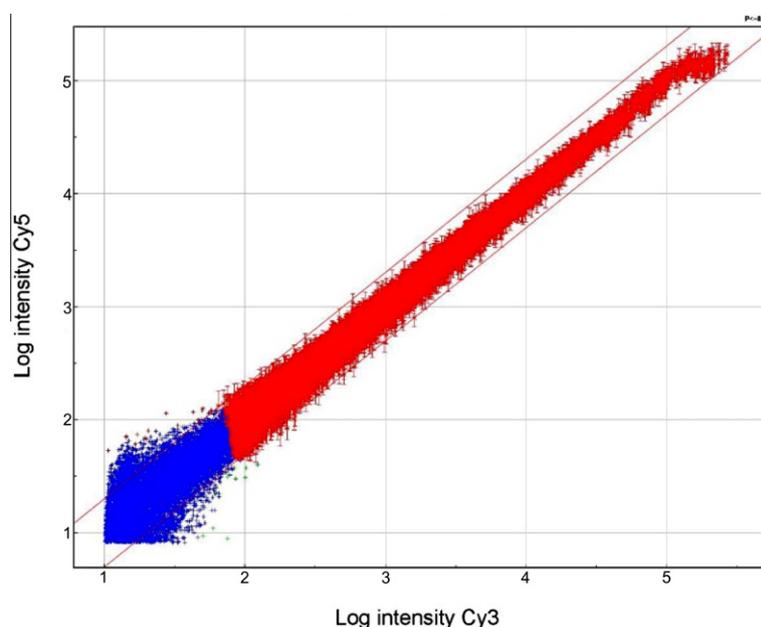


Fig. 1. Cross reactivity between dog and sea lion genes. RNA (500 ng) from the blood sample of a stranded sea lion was amplified, and technical replicates were labeled with Cy3 or Cy5 dyes. The samples were hybridized to the commercial Agilent 4 × 44 K canine microarray in a self against self format. The log signal intensity for each gene was plotted with the Cy5 intensity value on the y-axis, and the intensity of Cy3 binding on the x-axis. Those genes with intensities greater than 10^2 for both Cy3 and Cy5 were identified and are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

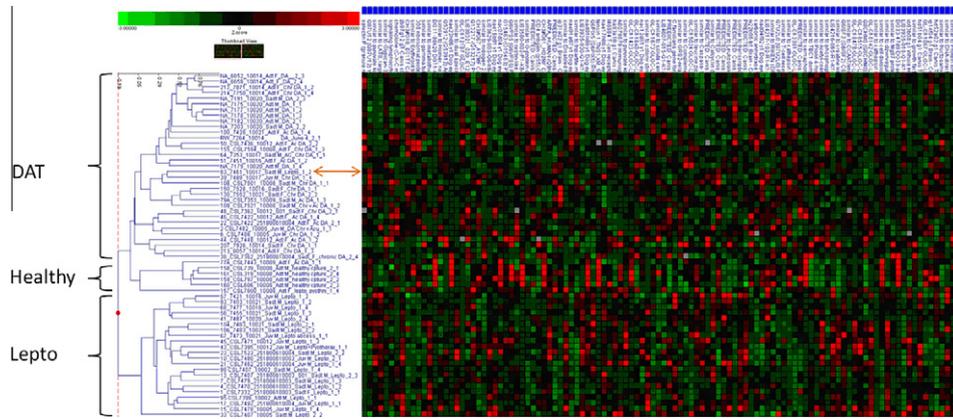


Fig. 2. Agglomerative cluster of all animals representing the Healthy, DAT and Lepto groups. The cluster was generated using the 110 genes selected by SVM in the disease classifications. Samples are clustered by similarity of gene expression. The animals are shown on the left, the 110 genes on the top (Table S2). The DAT animals are grouped on the top left, the Healthy in the center left and the Lepto at the bottom left (green parenthesis). The orange arrow indicates one Lepto animal misclassified to the DAT group. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and resulted in these animals quite frequently populating the other groups in classification attempts (shown in the principal component analysis below and in Fig. 3). To determine if there was any value in attempting to identify a gene expression fingerprint in the remaining animals, we eliminated the Outgroup and focused on the DAT and Lepto groups. As a first test, we compared the DAT and Lepto groups by building a differential expression ratio between the two, which resulted in 348 genes of interest ($p < 0.001$, Table S1).

As expected, some of the genes differentially expressed between DAT and Lepto reflected an activation of the immune response. For example, the following were all significantly up-regulated in Lepto- sea lions: macrophage erythroblast attachment-like, which is known to play a role in the development of the mature macrophages, active in both innate and adaptive immunity (Hanspal et al., 1998); the macrophage scavenger receptor 1, a recognition receptor implicated in innate immunity, apoptotic cell clearance, atherosclerosis and Alzheimer’s disease (Emi et al., 1993; Neyen et al., 2009); the Toll interacting protein (Tollip), a component of the signaling pathway of IL-1 and Toll-like receptors, and inhibits cell activation by microbial products (Zhang and Ghosh, 2002); the tumor necrosis factor receptor superfamily, member 25, that mediates activation of NF-kappa-B, induces apoptosis and may play a role in regulating lymphocyte homeostasis (Jiang et al., 1999); the signal transducer and activator of transcription STAT3, which is an acute-phase response transcription factor

that binds to the interleukin-6 (IL-6)-responsive elements identified in the promoters of various acute-phase protein genes (Collum et al., 2000). In contrast, the T-lymphocyte antigen 86 (downregulated in Leptospirosis infected sea lions) is a receptor that plays a critical role in the early events of T-cell activation and costimulation of naive T-cells (Engel et al., 1994).

The 348 genes were then used to drive an SVM classifier algorithm that selected 110 genes for complete stratification of animals (Table S2). This gene set was then employed in un-supervised clustering methods with success. Fig. 2 shows an agglomerative cluster with the animals representing the Healthy, DAT and Lepto groups. The animals were clustered into 3 distinct groups as indicated by parentheses with only one animal clustered in the wrong group, a sea lion diagnosed with leptospirosis was misplaced by the algorithm in the DAT group (orange arrow, Fig. 2). The same 110 gene set was used in a principal component analysis (PCA), which shows the inconsistency of the Outgroup animals signature (Fig. 3A) and the characteristics of the animals populating the other 3 groups (Fig. 3B). Admittedly, the gene set used here was derived from comparison of DAT vs. Lepto groups, but this plot is indicative of the Outgroup, with highly variable and no clear reproducible gene expression pattern exhibited in most analyses. After the removal of the Outgroup, the remaining animals show clear distinction of their different classes (Fig. 3B).

The entire set of 348 genes selected from the DAT/Lepto ratio as differentially regulated was run through the web tool DAVID

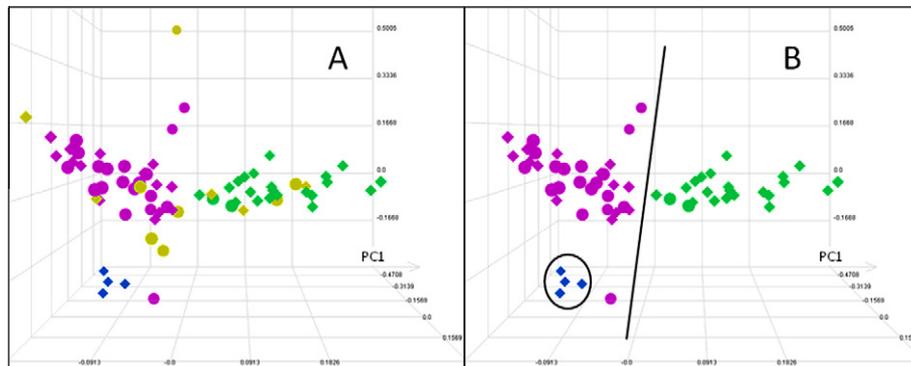


Fig. 3. Principal component analysis (PCA) plots of all animals. PCA generated from the SVM selected list of 110 genes using all animals (A) and with Outgroup animals removed (B). Circles are females; diamonds are males. Blue – animals belonging to group 1 (healthy); purple – animals belonging to group 2 (DAT); green – animals belonging to group 3 (Lepto); yellow – animals belonging to group 4 (Outgroup). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Database for Annotation, Visualization of Integrated Discovery) software (Huang da et al., 2009a,b) and placed (where possible) in one or several KEGG pathways, thus mapping individual gene expression to higher-level pathway functions. A limiting factor was the level of gene functions and annotations for the canine genome that was used as a reference. Of the 348 genes, 110 had no match or matched a DNA sequence present on the NCBI database, but with no known function. The remaining 238 were mapped to KEGG pathways ($p < 0.05$). The pathways with greatest enrichment were (1) Ubiquitin mediated proteolysis (6 genes, $p = 0.045$), (2) SNARE interaction in vesicular transport (4 genes, $p = 0.018$), (3) Alzheimer's disease (AD) (9 genes, $p = 0.021$) and (4) Oxidative phosphorylation (7 genes, $p = 0.013$).

These cross-species hybridization microarray results were challenged using the more conventional quantitative real time PCR (qPCR) to confirm differential gene expression and a less conservative and independent non-linear method such as the artificial neural networks (ANNs), to confirm the classifier property analysis.

3.3. Validation of gene expression monitoring of the cross-species microarray: qPCR

For confirmation of microarray data we decided to use only adult animals in our PCR validation, as the juveniles tended to show the highest variability. Eight test genes plus one control were chosen from the DAT vs. Lepto comparisons to confirm the microarray results. The 8 test genes have a variety of functions from intercellular signaling to basic metabolism. Some were chosen because they showed large changes in expression between groups, some because they were annotated and their specialized functions in inflammation and/or neuro- and muscular-disorders could have a potential link with the health disturbances studied here, and one because it could prove to be easily tested in a clinical setting. Due to the limitation of RNA quantity for the initial step of reverse transcription, a total of 30 adult animals were used for qPCR, consisting of 15 DAT, 7 Lepto and 8 Outgroup animals.

These 8 genes and their known functions are (1) tumor necrosis factor alpha-induced protein 6 (TNFAIP6), involved in extracellular matrix stability and cell migration (Kahmann et al., 2000); (2) ring finger protein 122 (RNF122), involved in protein-protein and protein-DNA interactions (Wang et al., 2006); (3) calcium binding atopy-related autoantigen 1 (CBARA1), key regulator of mitochondrial calcium uptake and also induces T-helper 1-mediated autoreactivity, which is accompanied by the release of interferon gamma (Aichberger et al., 2005); (4) neuromedin U receptor 1 (NMUR), leads to intracellular signal transduction via calcium mobilization, phosphoinositide signaling, and the inhibition of cAMP production through the binding of neuromedin U (or NmU, a neuropeptide found in the brain involved in the contraction of smooth muscle, regulation of blood pressure, pain perception, appetite, bone growth, and hormone release) (Howard et al., 2000); (5) scavenger receptor class B, member 2 (SCARB2), lysosome membrane protein 2 that participates in membrane transportation and the reorganization of endosomal/lysosomal compartment (deficiency of the similar protein impairs cell membrane transport processes and cause pelvic junction obstruction, deafness, and peripheral neuropathy; mutations of SCARB2 cause progressive myoclonic epilepsy (Berkovic et al., 2008)); (6) nesprin-1 (SYNE1), regulates G protein-coupled receptor phosphorylation in a calcium-dependent manner and can substitute for calmodulin. Nesprin-1 is an outer nuclear membrane protein that is thought to link the nucleus to the actin cytoskeleton: mutations in Nesprin 1 lead to a form of autosomal recessive cerebellar ataxia and may also be involved in the pathogenesis of Emery-Dreifuss muscular dystrophy (Gros-Louis et al., 2007; Zhang et al., 2010, 2007); (7) neuronal calcium sensor-1 (NCS-1), is a calcium sensor with high-affinity,

low-capacity, calcium-binding activity with a known role in learning and memory. The expression of NCS-1 increases in bipolar disorder and some forms of schizophrenia (Zhang et al., 2010); (8) phosphoglucosyltransferase 5 (PGM5), involved in interconversion of glucose-1-phosphate and glucose-6-phosphate (Edwards et al., 1995). Although this gene presented with a very low change in expression between DAT and Lepto (30%), its enzymatic activity make this protein an easily tested biomarker in blood.

Analysis of PCR amplicons by capillary gel electrophoresis, for each gene in two randomly selected animals, showed that 5 of the 8 reactions produced strong single bands while 3 reactions showed multiple bands (Fig. S2). Although these 3 reactions produced multiple bands, we proceeded with the assays anyway, since further optimization was difficult without the sea lion specific sequence data. The data were analyzed by averaging the C_t values of all animals for each one of the 3 groups. Of the 8 genes tested 5 confirmed the microarray data and their level of expression is shown in Fig. 4. The 3 remaining genes that did not exhibit changes between groups by PCR were the same 3 assays that produced multiple bands (Fig. S2, lanes 9–10, 11–12, 13–14). Not surprisingly, these genes also showed high variability by PCR across animals, above all in the Lepto group (the average C_t and sd in the Lepto animals for NCS-1 was 27.85 ± 2.31 , for NMUR was 26.76 ± 1.61 and for PGM5 was 27.01 ± 1.83). It is necessary to confirm that the PCR reaction produces a single band of the expected size.

The PCR comparisons of the Outgroup animals to both DAT and Lepto groups reproduce what was observed with both the linear and non-linear methods of analysis of the microarray records. The results were confounded by the high variability of C_t values across animals in the Outgroup, most likely reflecting the variety of the stranding causes, diseases and diagnosis described. Specifically, the 8 animals of the Outgroup used in qPCR were identified with carcinoma, neoplasia, blindness, corneal ulcer, pneumonia, pyothorax, trauma and malnutrition.

The difference in expression observed by qPCR between the DAT and Lepto animals for 4 of the 5 confirmatory genes (but not CBARA1) was statistically significant with $p < 0.05$. In particular TNFAIP6 showed the biggest change between the two groups (and lowest p value $p < 0.00058$), being up-regulated in the DAT group ($\Delta C_t = 2.53$) and down-regulated in the Lepto group ($\Delta C_t = -0.75$) relative to the control gene expression.

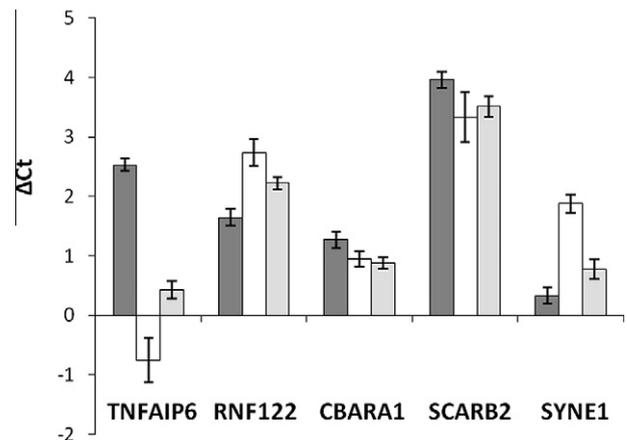


Fig. 4. Real time PCR data for 5 selected genes. RNA from 30 different animals was individually reverse transcribed, each gene was amplified in triplicate from the cDNA. Data were normalized to ELL, elongation factor. ΔC_t values of animals belonging to the same group were pooled. Dark grey bars, DAT ($n = 15$); White bars, Lepto ($n = 7$); light grey bars, Outgroup ($n = 8$).

3.4. Sequencing

PCR products were purified, cloned and then sent for sequencing by a commercial facility (SeqWright Inc., Houston, TX). Sequences were vector trimmed using VecScreen and aligned with canine sequences using BioEdit. Identity between California sea lion and canine sequences was quite high; TNAIP shared 139 of 141 base identity (98.6%), RNF122 shared 132 of 135 bases (97.8%), CBARA1 shared 112 of 115 (97.4%), and SCARB2 shared 115 of 119 (96.6)% identity including the consensus primers used for PCR (Table S5).

3.5. Validation of the classificatory power of the cross-species microarray: the artificial neural networks

After exploring the capacity of traditional classifiers to distinguish disease state by blood gene expression, in a combination of supervised and unsupervised methods, we employed a non-linear method (the artificial neural network, ANN) to test classification of diseased status. Both linear and non-linear methods were able to generate highly precise models of the relationship between gene expression and disease status as evidenced by the perfect correlations of diseased and healthy animals (Table 2). The value of ROC curves generated by ANN provide confirmation in that there were no false positive or negative diagnoses in comparison of health to diseased individuals, and the discrimination of specific disease conditions were better than 85%. The perfect discrimination between healthy and diseased animals (Table 3, ROC value for DAT vs. Healthy-Lepto vs. Healthy- and Outgroup vs. Healthy- was always equal to 1 or 100% chance of correct classification) was to be expected given the origin and the characteristics of the healthy animals (Table 3). The healthy sea lions used in the study were in fact, only 4, all males and they were sampled from Oregon, north of the stranding locations of most other animals, which was near the rehabilitation center in Sausalito CA. The discrimination of the Lepto animals was surprising. The transcriptomic profile expressed a unique signature and a probability of correct classification very close to 100% (98.8% in Lepto vs. DAT and 99.9% in Lepto vs. Outgroup). The weakest value in probability of discriminating animals from different groups was 85.5% in the DAT vs. Outgroup comparison, which was still an excellent ROC value.

The ANNs were initially run using 500 genes with the highest sensitivities, but the same classificatory power was obtained even when only the top 50 genes (first 50 lines in the comparisons for each one of the 500 gene set/comparisons) were used in the run (Table S3). Table 4 shows the number of genes that were found to be up and down regulated in the different comparisons. Genes found unique to one comparison were the majority

Table 4
Summary of UP and DOWN regulated genes selected by ANN in pairwise comparisons.

	Genes	DAT vs. Lepto	DAT vs. Outgroup	Lepto vs. Outgroup
UP regulated	Unique	157	144	119
	Shared	25	35	12
	TOT	182	179	131
DOWN regulated	Unique	247	236	288
	Shared	71	85	81
	TOT	318	321	369
UP and DOWN – ANN selected	TOT	500	500	500

Total number of the genes selected by ANN was 500 for each comparisons; 'unique' were the genes that were selected in only one comparison; '<shared>' were those genes that were selected in at least 2 of the 3 comparisons.

(157 + 247 = 404, 144 + 236 = 380 and 119 + 288 = 407 for DAT vs. Lepto, DAT vs. Outgroup, Lepto vs. Outgroup, respectively) and only one was shared by all comparisons, sequence coding for the canine collagen alpha 1 (NCBI Acc. No. XM_858055). Overall, of the 500 genes found to be significant in the ANNs runs, the down-regulated genes were always predominant in the discrimination of diseased states (182up- 318down, 179up- 321down, 131up- 369down for DAT vs. Lepto, DAT vs. Outgroup, Lepto vs. Outgroup, respectively). The pair wise comparisons with the healthy animals were removed from all further analysis and discussion of the data, given the strong difference observed in the preliminary runs and their known difference in geographical history.

4. Discussion

Any study using wild populations can be complicated by factors related to seasonal variations in environment, diet, reproductive cycle and migration. Known confounding factors in this study were season of stranding, nutritional status of the animal, reproductive status and coinfection with macroparasites. Additionally, 24 animals were diagnosed with leptospirosis, and of these animals, 22 were male. Domoic acid exposure on the other hand was much more evenly distributed between sexes with 14 males and 19 females sampled. The classification of DAT and Lepto was successfully achieved with a similar outcome through both linear and non-linear methods of analysis. While using SVM selected genes for unsupervised agglomerative clustering, DAT and Lepto animals were correctly classified in their respective groups, with only one Lepto animal misclassified in the DAT group (Figs. 2 and 3B), the ANN selected and used 500 genes which allowed the discrimination between DAT and Lepto animals with a 99% rate of success (Table 3). Fig. 3A and Table 3 also show that the classification of the Outgroup via both PCA and ANN, although still successful, was less striking. The Outgroup was consistently difficult to properly classify, because of the range of health complications such as cancer, blindness, and entanglement. Also, this category does not preclude previous exposure to contaminants or toxins such as domoic acid. For example, an exposure to DA could have disoriented an animal to the point where it became entangled in a net, leading to the need for rehabilitation at The Marine Mammal Center. Since the toxin or symptoms could have disappeared by the time of examination, an animal could be classified in an inappropriate health category for our analysis.

Some of the gene expression results are indicative of illness. In fact, in Lepto animals, consistent with the indication of a bacterial infection, we find up-regulated genes such as the macrophage erythroblast attachment-like (development of the mature macrophages and active in both innate and adaptive immunity), the macrophage scavenger receptor 1 (involved in innate immunity) and the Toll interacting protein Tollip (component of the signaling pathway of IL-1 and Toll-like receptors, inhibits cell activation by microbial products). The first case of leptospirosis in sea lions was reported in 1970 from a mass stranding and mortality on the northern California and Oregon coasts (Greig et al., 2005). Since then, outbreaks of acute leptospirosis infection have been reported periodically; in particular they seem to follow a seasonal pattern with peaks in August–October, affecting mostly male sea lions that travel extensively north while females remain near the rookeries nursing pups (Greig et al., 2005; Melin, 2002).

Sea lions in California are frequently exposed to domoic acid during and after harmful algal blooms through ingestion of contaminated prey items such as sardines and anchovies (Scholin et al., 2000) which can result in both acute and chronic effects (Gulland et al., 2002). Domoic acid binds to ionotropic glutamate receptors in the CNS causing sustained neuronal excitation, or

excitotoxicity, which in turn can lead to dysfunction and death of the cells (Olney et al., 1979). A single exposure can lead to damage, necrosis and atrophy of the hippocampus, which has been previously described through histopathology and MRI (Goldstein et al., 2008). Moreover, severe cases of DAT often present with abnormal behaviors, and even seizures and epilepsy are amongst the clinical conditions observed (Goldstein et al., 2008). It was noteworthy that the KEGG pathway for Alzheimer's disease (AD) was enriched, which has a neuropathology profile including features shared by the California sea lions intoxicated with domoic acid, in particular in the recently discussed link between hippocampal lesions and epilepsy (Goldstein et al., 2008; Noebels, 2011). Seizures are part of the natural history of many pedigrees with autosomal dominant early onset AD, including those with mutations in presenilin-1, presenilin-2, or the amyloid precursor protein (APP), or with duplications of wild-type APP. This can result in amyloid beta (A β) deposits and buildup in the brain, which are the fundamental cause of the disease (Hardy and Allsop, 1991; Hsiao et al., 1996). Our microarray results showed that presenilin-2 and beta-site APP-cleaving enzyme 2, are two of the genes showing differential expression between DAT and Lepto animals. Although AD targets the CNS, several studies have linked immune proteins found in blood to the illness (Reddy et al., 2011; Weeraratna et al., 2007). The KEGG pathway shown in Fig. S3 does not imply the existence of AD in California sea lions but does point out the similarity of the condition induced by the neurotoxin. Even the activation of the oxidative phosphorylation pathway of the energy metabolism group was not surprising as it has been shown before that a stressful or diseased condition could activate the energy machinery (Chapman et al., 2009). Moreover, in AD pathogenesis, increasing evidence implicates mitochondrial dysfunction resulting from molecular defects in oxidative phosphorylation (Manczak et al., 2004).

One of the genes selected by ANN for its sensitivity and weight in the classificatory outcome, was TNFAIP6, which showed a 3.5 fold difference of expression between the DAT (up) and Lepto (down) in qPCR. The expression of TNFAIP6 (or TSG-6) is regulated by a number of signaling molecules, principally tumor necrosis factor α (TNF- α) and interleukin-1 (IL-1) and its activation seems to be involved in intercellular and cell-matrix interactions and migration during inflammation and tumorigenesis (Johnatty et al., 2011; Maina et al., 2009). Through its hyaluronan-binding domains TNFAIP forms a stable complex with the inter-alpha-inhibitor protein (α 1), and enhances the serine protease inhibitory activity of α 1, important in the protease network associated with inflammation (Lee et al., 1993, 1992). This gene showed the greatest difference between the DAT and Lepto groups by PCR. Thus its use as a biomarker and its involvement in DAT warrants further investigation.

5. Conclusions

As the immune system surveys the entire organism, its ability to report changes to the health of an animal, from a wide range of pathologies, is unparalleled. Several methods were assessed for their ability to identify gene expression patterns from leukocytes of California sea lions, with the goal of developing robust diagnostic classifiers that would allow for rapid diagnosis and treatment of stranded animals. The results achieved with this work indicate that cross species microarray technologies, using a selected gene set for microarray design and analysis, can be successfully used to classify some disease states in California sea lions, and could be of use to wildlife managers and veterinarians in handling species with little genomic data. Of course the limitation of cross species applications is the depth of genome characterization and

identified gene functions in the model organism. That being said, until we have whole genome scale sequence data for the hundreds of protected species whose care is in our charge, the combination of cross-species microarray analysis coupled with machine learning approaches (coordinated with a high level of data exchange between veterinarian examinations and bioinformaticists) may be of use in the identification of environmental threats and diseased conditions.

Disclaimers

This publication does not constitute an endorsement of any commercial product or intend to be an opinion beyond scientific or other results obtained by the National Oceanic and Atmospheric Administration (NOAA). No reference shall be made to NOAA, or this publication furnished by NOAA, to any advertising or sales promotion which would indicate or imply that NOAA recommends or endorses any proprietary product mentioned herein, or which has as its purpose an interest to cause the advertised product to be used or purchased because of this publication.

This material is based in part on work supported by the National Science Foundation. Any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements

This work was funded by the NOAA Marine Biotoxins Program, the NOAA Oceans and Human Health Initiative, and a grant from the Oiled Wildlife Care Network.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.dci.2011.10.011.

References

- Aichberger, K.J., Mittermann, I., Reiningger, R., Seiberler, S., Swoboda, I., Spitzauer, S., Kopp, T., Stingl, G., Sperr, W.R., Valent, P., Repa, A., Bohle, B., Kraft, D., Valenta, R., 2005. Hom s 4, an IgE-reactive autoantigen belonging to a new subfamily of calcium-binding proteins, can induce Th cell type 1-mediated autoreactivity. *J. Immunol.* 175, 1286–1294.
- Berkovic, S.F., Dibbens, L.M., Oshlack, A., Silver, J.D., Katerelos, M., Vears, D.F., Lullmann-Rauch, R., Blanz, J., Zhang, K.W., Stankovich, J., Kalnins, R.M., Dowling, J.P., Andermann, E., Andermann, F., Faldini, E., D'Hooge, R., Vadlamudi, L., Macdonell, R.A., Hodgson, B.L., Bayly, M.A., Savige, J., Mulley, J.C., Smyth, G.K., Power, D.A., Saftig, P., Bahlo, M., 2008. Array-based gene discovery with three unrelated subjects shows SCARB2/LIMP-2 deficiency causes myoclonus epilepsy and glomerulosclerosis. *Am. J. Hum. Genet.* 82, 673–684.
- Bharti, A.R., Nally, J.E., Ricaldi, J.N., Matthias, M.A., Diaz, M.M., Lovett, M.A., Levett, P.N., Gilman, R.H., Willig, M.R., Gotuzzo, E., Vinetz, J.M., 2003. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* 3, 757–771.
- Chapman, R.W., Mancía, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., Sanger, D., Holland, A.F., Warr, G.W., Didonato, G., 2009. A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica*, in the South Atlantic bight. *Mol. Ecol.* 18, 2415–2425.
- Collum, R.G., Brutsaert, S., Lee, G., Schindler, C., 2000. A Stat3-interacting protein (StIP1) regulates cytokine signal transduction. *Proc. Natl. Acad. Sci. USA* 97, 10120–10125.
- Conesa, A., Gotz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 619832.
- Dhar, A.K., Bowers, R.M., Licon, K.S., Veazey, G., Read, B., 2009. Validation of reference genes for quantitative measurement of immune gene expression in shrimp. *Mol. Immunol.* 46, 1688–1695.
- Edwards, Y.H., Putt, W., Fox, M., Ives, J.H., 1995. A novel human phosphoglucosyltransferase (PGM5) maps to the centromeric region of chromosome 9. *Genomics* 30, 350–353.
- Emi, M., Asaoka, H., Matsumoto, A., Itakura, H., Kurihara, Y., Wada, Y., Kanamori, H., Yazaki, Y., Takahashi, E., Lepert, M., et al., 1993. Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene. *J. Biol. Chem.* 268, 2120–2125.

- Engel, P., Gribben, J.G., Freeman, G.J., Zhou, L.J., Nozawa, Y., Abe, M., Nadler, L.M., Wakasa, H., Tedder, T.F., 1994. The B7-2 (B70) costimulatory molecule expressed by monocytes and activated B lymphocytes is the CD86 differentiation antigen. *Blood* 84, 1402–1407.
- Goldstein, T., Mazet, J.A., Zabka, T.S., Langlois, G., Colegrove, K.M., Silver, M., Bargu, S., Van Dolah, F., Leighfield, T., Conrad, P.A., Barakos, J., Williams, D.C., Dennison, S., Haulena, M., Gulland, F.M., 2008. Novel symptomatology and changing epidemiology of domoic acid toxicosis in California sea lions (*Zalophus californianus*): an increasing risk to marine mammal health. *Proc. Biol. Sci.* 275, 267–276.
- Greig, D.J., Gulland, F.M.D., Kreuder, C., 2005. A decade of live California sea lion (*Zalophus californianus*) strandings along the central California coast: causes and trends, 1991–2000. *Aquat. Mammals* 31, 40–51.
- Gros-Louis, F., Dupre, N., Dion, P., Fox, M.A., Laurent, S., Verreault, S., Sanes, J.R., Bouchard, J.P., Rouleau, G.A., 2007. Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia. *Nat. Genet.* 39, 80–85.
- Gulland, F.M., Haulena, M., Fauquier, D., Langlois, G., Lander, M.E., Zabka, T., Duerr, R., 2002. Domoic acid toxicity in Californian sea lions (*Zalophus californianus*): clinical signs, treatment and survival. *Vet. Rec.* 150, 475–480.
- Hanspal, M., Smockova, Y., Uong, Q., 1998. Molecular identification and functional characterization of a novel protein that mediates the attachment of erythroblasts to macrophages. *Blood* 92, 2940–2950.
- Hardy, J., Allsop, D., 1991. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol. Sci.* 12, 383–388.
- Howard, A.D., Wang, R., Pong, S.S., Mellin, T.N., Strack, A., Guan, X.M., Zeng, Z., Williams Jr., D.L., Feighner, S.D., Nunes, C.N., Murphy, B., Stair, J.N., Yu, H., Jiang, Q., Clements, M.K., Tan, C.P., McKee, K.K., Hreniuk, D.L., McDonald, T.P., Lynch, K.R., Evans, J.F., Austin, C.P., Caskey, C.T., Van der Ploeg, L.H., Liu, Q., 2000. Identification of receptors for neuromedin U and its role in feeding. *Nature* 406, 70–74.
- Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., Yang, F., Cole, G., 1996. Correlative memory deficits, Aβ elevation, and amyloid plaques in transgenic mice. *Science* 274, 99–102.
- Huang da, W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Huang da, W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocol.* 4, 44–57.
- Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., Vingron, M., 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18 (Suppl. 1), S96–S104.
- Jiang, Y., Woronicz, J.D., Liu, W., Goeddel, D.V., 1999. Prevention of constitutive TNF receptor 1 signaling by silencer of death domains. *Science* 283, 543–546.
- Johnatty, S.E., Beesley, J., Chen, X., Macgregor, S., Duffy, D.L., Spurdle, A.B., deFazio, A., Gava, N., Webb, P.M., Rossing, M.A., Doherty, J.A., Goodman, M.T., Lurie, G., Thompson, P.J., Wilkens, L.R., Ness, R.B., Moysich, K.B., Chang-Claude, J., Wang-Gohrke, S., Cramer, D.W., Terry, K.L., Hankinson, S.E., Tworoger, S.S., Garcia-Closas, M., Yang, H., Lissowska, J., Chanock, S.J., Pharoah, P.D., Song, H., Whitmore, A.S., Pearce, C.L., Stram, D.O., Wu, A.H., Pike, M.C., Gayther, S.A., Ramus, S.J., Menon, U., Gentry-Maharaj, A., Anton-Culver, H., Ziogas, A., Hogdall, E., Kjaer, S.K., Hogdall, C., Berchuck, A., Schildkraut, J.M., Iversen, E.S., Moorman, P.G., Phelan, C.M., Sellers, T.A., Cunningham, J.M., Vierkant, R.A., Rider, D.N., Goode, E.L., Haviv, I., Chenevix-Trench, G., 2011. Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility “hot-spot”. *PLoS Genet.* 6, e1001016.
- Kahmann, J.D., O'Brien, R., Werner, J.M., Heinegard, D., Ladbury, J.E., Campbell, I.D., Day, A.J., 2000. Localization and characterization of the hyaluronan-binding site on the link module from human TSG-6. *Structure* 8, 763–774.
- Lee, T.H., Klampfer, L., Shows, T.B., Vilcek, J., 1993. Transcriptional regulation of TSG6, a tumor necrosis factor- and interleukin-1-inducible primary response gene coding for a secreted hyaluronan-binding protein. *J. Biol. Chem.* 268, 6154–6160.
- Lee, T.H., Wisniewski, H.G., Vilcek, J., 1992. A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. *J. Cell Biol.* 116, 545–557.
- Levin, M., Joshi, D., Draghi 2nd, A., Gulland, F.M., Jessup, D., De Guise, S., 2010. Immunomodulatory effects upon in vitro exposure of California sea lion and southern sea otter peripheral blood leukocytes to domoic acid. *J. Wildl. Dis.* 46, 541–550.
- Maina, V., Cotena, A., Doni, A., Nebuloni, M., Pasqualini, F., Milner, C.M., Day, A.J., Mantovani, A., Garlanda, C., 2009. Coregulation in human leukocytes of the long pentraxin PTX3 and TSG-6. *J. Leukoc. Biol.* 86, 123–132.
- Mancía, A., Warr, G., Almeida, J.S., Veloso, A., Wells, R., Chapman, R.W., 2010. Transcriptome profiles: diagnostic signature of dolphin populations. *Estuaries Coasts* 33, 919–929.
- Mancía, A., Warr, G.W., Chapman, R.W., 2008. A transcriptomic analysis of the stress induced by capture-release health assessment studies in wild dolphins (*Tursiops truncatus*). *Mol. Ecol.* 17, 2581–2589.
- Manczak, M., Park, B.S., Jung, Y., Reddy, P.H., 2004. Differential expression of oxidative phosphorylation genes in patients with Alzheimer's disease: implications for early mitochondrial dysfunction and oxidative damage. *Neuromolecular Med.* 5, 147–162.
- Melin, S.R., 2002. The foraging ecology and reproduction of the California sea lion (*Zalophus californianus*). University of Minnesota.
- Neyen, C., Pluddemann, A., Roversi, P., Thomas, B., Cai, L., van der Westhuyzen, D.R., Sim, R.B., Gordon, S., 2009. Macrophage scavenger receptor A mediates adhesion to apolipoproteins A-I and E. *Biochemistry* 48, 11858–11871.
- Noebels, J., 2011. A perfect storm: converging paths of epilepsy and Alzheimer's dementia intersect in the hippocampal formation. *Epilepsia* 52 (Suppl. 1), 39–46.
- Olney, J.W., Fuller, T., de Gubareff, T., 1979. Acute dendrotoxic changes in the hippocampus of kainate treated rats. *Brain Res.* 176, 91–100.
- Reddy, M.M., Wilson, R., Wilson, J., Connell, S., Gocke, A., Hynan, L., German, D., Kodadek, T., 2011. Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. *Cell* 144, 132–142.
- Riedman, M., 1991. The Pinnipeds: Seals, Sea Lions, and Walrus University of California Press.
- Ryan, J.C., Cross, C.A., Van Dolah, F.M., 2011. Effects of COX inhibitors on neurodegeneration and survival in mice exposed to the marine neurotoxin domoic acid. *Neurosci. Lett.* 487, 83–87.
- Scholin, C.A., Gulland, F., Doucette, G.J., Benson, S., Busman, M., Chavez, F.P., Cordaro, J., DeLong, R., De Vogelaere, A., Harvey, J., Haulena, M., Lefebvre, K., Lipscomb, T., Loscutoff, S., Lowenstine, L.J., Marin 3rd, R., Miller, P.E., McLellan, W.A., Moeller, P.D., Powell, C.L., Rowles, T., Silvagni, P., Silver, M., Spraker, T., Trainer, V., Van Dolah, F.M., 2000. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature* 403, 80–84.
- Silvagni, P.A., Lowenstine, L.J., Spraker, T., Lipscomb, T.P., Gulland, F.M., 2005. Pathology of domoic acid toxicity in California sea lions (*Zalophus californianus*). *Vet. Pathol.* 42, 184–191.
- Van Dolah, F.M., 2000. Marine algal toxins: origins, health effects, and their increased occurrence. *Environ. Health Perspect.* 108 (Suppl. 1), 133–141.
- Wang, L., Shi, T.P., Yu, C.F., Zeng, L.E., Wang, J., 2006. Cloning, expression and subcellular localization of a novel human gene-RNF122. *Beijing Da Xue Xue Bao* 38, 239–243.
- Weeraratna, A.T., Kalebua, A., Deleon, I., Bertak, D., Maher, G., Wade, M.S., Lustig, A., Becker, K.G., Wood 3rd, W., Walker, D.G., Beach, T.G., Taub, D.D., 2007. Alterations in immunological and neurological gene expression patterns in Alzheimer's disease tissues. *Exp. Cell Res.* 313, 450–461.
- Weng, L., Dai, H., Zhan, Y., He, Y., Stepaniants, S.B., Bassett, D.E., 2006. Rosetta error model for gene expression analysis. *Bioinformatics* 22, 1111–1121.
- Zhang, G., Ghosh, S., 2002. Negative regulation of toll-like receptor-mediated signaling by Tollip. *J. Biol. Chem.* 277, 7059–7065.
- Zhang, J., Felder, A., Liu, Y., Guo, L.T., Lange, S., Dalton, N.D., Gu, Y., Peterson, K.L., Mizisin, A.P., Shelton, G.D., Lieber, R.L., Chen, J., 2010. Nesprin 1 is critical for nuclear positioning and anchorage. *Hum. Mol. Genet.* 19, 329–341.
- Zhang, Q., Bethmann, C., Worth, N.F., Davies, J.D., Wasner, C., Feuer, A., Ragnauth, C.D., Yi, Q., Mellad, J.A., Warren, D.T., Wheeler, M.A., Ellis, J.A., Skepper, J.N., Vorgerd, M., Schlotter-Weigel, B., Weissberg, P.L., Roberts, R.G., Wehnert, M., Shanahan, C.M., 2007. Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Hum. Mol. Genet.* 16, 2816–2833.
- Zhu, J., He, F., Song, S., Wang, J., Yu, J., 2008. How many human genes can be defined as housekeeping with current expression data? *BMC Genomics* 9, 172.