UNIVERSITY OF CALIFORNIA

Los Angeles

Transcriptomic Approaches for Connecting Gene Regulation to Ecologically Important Traits

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biology

by

Rachel A. Johnston

2016

ABSTRACT OF THE DISSERTATION

Transcriptomic Approaches for Connecting

Gene Regulation to Ecologically Important Traits

By

Rachel A. Johnston

Doctor of Philosophy in Biology

University of California, Los Angeles, 2016

Professor Thomas Bates Smith, Co-Chair

Professor Robert Wayne, Co-Chair

Gene regulation is tightly regulated to produce individual phenotypes appropriate to a animal's placement in time and space. How specific genetic and environmental factors shape gene regulation has been extensively studied in laboratory model species. However, far less is known about sources of variance in gene regulation in the wild, in populations where natural selection operates. My dissertation research aimed to identify factors that influence individual variation in gene regulation by evaluating transcriptome-wide gene expression levels within and across populations. In my first chapter, I evaluated gene expression changes in captive Swainson's thrushes (*Catharus ustulatus*) originating from two populations as they transitioned from the non-migratory to migratory condition. Birds exhibited significant seasonal differences in the expression of genes involved in cellular development in the brain, providing support for

seasonal neural plasticity in migratory birds. In my second chapter, I focused on a single population, the North American gray wolves (*Canis lupus*) of Yellowstone National Park, to evaluate the relative impacts of intrinsic (age and sex) and environmental variables (social status and presence of sarcoptic mange) on gene expression variation in blood. I found that, unexpectedly, animal age, but not sex, social status, or sarcoptic mange infestation, impacted gene expression, suggesting that age may have pervasive, evolutionarily conserved effects on gene expression variation in natural populations of mammals. Finally, in my third chapter, I adopted an *in vitro* experimental approach to study how gene regulation is impacted by genotype at the *CBD103* gene, which encodes a protein with antimicrobial properties, exists in two forms (alleles) in North American gray wolves, and is associated with lifetime reproductive success. I established keratinocyte lines from 24 gray wolves and utilized CRISPR/Cas9 to generate one heterozygous and homozygous line. To test whether the fitness effects of *CBD103* genotype are explained by their roles in immune defense, I challenged the cells with synthetic antigens and live canine distemper virus. No effect of *CBD103* genotype was detected, but immune challenge altered expression of thousands of genes. This work demonstrated the ability to establish a population panel of cell lines from a wild mammal, a potentially powerful method for studying gene regulation in natural populations.

The dissertation of Rachel A. Johnston is approved.

Matteo Pellegrini

Barnett Schlinger

Thomas Bates Smith, Committee Co-Chair

Robert Wayne, Committee Co-Chair

University of California, Los Angeles

2016

To my mother for her endless positive encouragement,

and to my father for inspiring my love for science.

TABLE OF CONTENTS

CHAPTER THREE: Population-based cell assays for studying gene function in the North American gray wolf

LIST OF TABLES

LIST OF TABLES PROVIDED AS SUPPLEMENTARY FILES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Bob Wayne, for his continual support and mentorship. None of this work would be possible without his commitment to ensuring that I had the necessary resources and training for this research. I would also like to thank my co-advisor, Tom Smith, and my dissertation committee members, Barney Schlinger and Matteo Pellegrini, for their helpful guidance over the past six years.

Chapter one is a version of Johnston RA, Paxton K, Wayne RK, Smith TB. 2016. Seasonal gene expression in a migratory songbird. In press. Molecular Ecology. Online ahead of print: doi:10.1111/mec.13879, PMID:27747949. I thank John Wiley and Sons for permission to reprint this chapter (license number 4000401508601). The author contributions are: Conceptualization, R.A.J., K.L.P., F.R.M., and T.B.S. Methodology, R.A.J., K.L.P., F.R.M., R.K.W. and T.B.S. Formal analysis, R.A.J. Investigation, R.A.J. Resources, K.L.P. and F.R.M. Writing – Original Draft, R.A.J. Writing – Review and Editing, R.A.J., K.L.P., F.R.M., R.K.W., and T.B.S. Funding Acquisition, F.R.M., R.K.W., and T.B.S. This study would not have been possible without the help of Christine Ames, Emily Cohen, Jill Gautreaux, Kimberly Hollinger, Matthew Johnson, Emijo Laine, Eben Paxton, Abby Powell, C.J. Ralph, Jacqueline Smolinski, Katherine Vowell, Kevin Winker, T.J. Zenzal, and the Migratory Bird Research Group. The authors thank the animal care facility staff at Humboldt State University, University of Alaska Fairbanks, and University of Southern Mississippi for assistance maintaining birds in captivity. I thank Julia Barske and Devaleena Pradhan for guidance on tissue collections. I am thankful to Kristen Ruegg for helpful discussions on Swainson's thrush biology and genotyping and to Barney Schlinger for comments on the manuscript. I am also thankful to Amanda Lea, Noah Snyder-Mackler, and Jenny Tung for providing tremendous guidance on data analysis.

Chapter two is a version of Charrau P*, Johnston R*, Stahler DR, Lea A, Snyder-Mackler N, Smith DW, vonHoldt BM, Cole SW, Tung J, Wayne RK. 2016. Pervasive effects of aging on gene expression in wild wolves. Molecular Biology and Evolution. 33(8):1967-78. doi:10.1093/molbev/msw072, PMID: 27189566. *Co-first author. I thank Oxford University Press for permission to reprint this chapter (license number 4004071169801). The author contributions are: Conceptualization, R.A.J., R.K.W., B.v.H., and P.C.; Resources, D.R.S., D.W.S., and R.K.W.; Funding Acquisition, R.K.W. and B.v.H.; Investigation, R.J. and P.C.; Methodology, A.L., N.S., J.T., S.C., P.C., and R.A.J.; Formal Analysis, R.A.J., P.C., and S.C.; Supervision, R.K.W.; Writing – Original Draft, R.A.J. and P.C.; Writing – Review & Editing, R.A.J., P.C., R.K.W., J.T., A.L., N.S., S.C., and D.R.S. The authors would like to acknowledge Emily Almberg and Erin Stahler for their contribution to data collection, and Devaughn Fraser for her significant help in the blood smear cell counts.

I thank Jim Rheinwald for his incredible mentorship and guidance for experiments in chapter three. This project would have never left the ground without his expertise and invaluable training and feedback. Jim was always there to lend encouragement and support and pushed me to be a better scientist. I am forever grateful for having the opportunity to work with him. I am also thankful to Dan Stahler, Doug Smith, and their teams at Yellowstone National Park, as well as Jennifer Struthers, Jason Husseman, and their teams at the Idaho Department of Fish and Game, for their willingness and hard work to collect the samples for this project. Their patience, perseverance, and collaborative spirits in taking on the herculean task of rapidly getting live samples from wild carnivores of the northwestern wilderness to a sterile laboratory in Los Angeles was truly inspiring. I thank Bill Lowry for taking a leap of faith in this collaboration and providing feedback, lab space, and equipment for all of the cell work. I owe a special thanks to

Jessica Cinkornpumin who was always there, day or night, weekday, weekend, or holiday, to provide help and offer guidance on lab work. I am also grateful to Steven Nguyen, Kashif Iqbal, Nadia Riabkova, and Joey Curtiss for their help with much of the molecular and cell work. I thank Bridgett vonHoldt for her integral role in the conceptualization and acquisition of funding of this project.

I am thankful to all the sources of funding for this dissertation, particularly the Howard Hughes Medical Institute, which provided me the HHMI Gilliam Fellowship and the opportunity to join a Ph.D. program. Other funding sources include the National Science Foundation (grant numbers DEB 1257716 to R.K.W. and B.M.V.; DEB 1245373 to D.R.S. and D.W.S.; IOS 844703 to F.R.M.; and IIA PIRE 1243524 to T.B.S.); the National Institutes of Health (grant numbers P30 AG017265 to S.W.C., S10 RR029668 and S10 RR027303 to work with the Vincent J. Coates GSL at UC Berkeley); the Moore Basic Research Development Account, USM (to F.R.M. and K.L.P.); UCLA's Center for Tropical Research; the Austrian Academy of Sciences (to P.C.); the Max Kade Foundation (to P.C.), the National Park Service (to D.R.S. and D.W.S.); the Yellowstone Park Foundation (to D.R.S. and D.W.S.); the Tapeats Fund (to D.R.S. and D.W.S.); and the Perkin-Prothro Foundation (to D.R.S. and D.W.S.).

Finally, I would like to thank my colleagues and friends that have supported me throughout the years, including labmates in the Wayne lab; Maite Lázaro, who reminded me to relax; Jacqueline Robinson, for being a voice of reason; and Ryan Harrigan, who provided guidance and absolute support in every aspect of my work.

VITA

EDUCATION

2006 - 2010 B.S. Biology, New Mexico State University

FELLOWSHIPS

2015 - 2016 UCLA Dissertation Year Fellowship, $35,500

2010 - 2015 HHMI Gilliam Fellowship for Advanced Study, five year award of annual $30,000 stipend + $3,000 allowance + $13,500 institutional allowance

GRANTS

2014 NSF Research Experience for Undergraduates Supplement, $6,000

2011 - 2013 UCLA Department of Ecology and Evolutionary Biology Grant, $4,600

AWARDS AND TRAINEESHIPS

2014 UCLA Department of Ecology and Evolutionary Biology Service Award

2009 - 2010 Minority Access to Research Careers (MARC) Undergraduate Trainee

2007 - 2010 Howard Hughes Medical Institute (HHMI) Research Trainee

2006 - 2010 NMSU President's Associates Scholar

2009 HHMI Exceptional Research Opportunities Program Trainee

2007 - 2009 Minority Biomedical Research Support-Research Initiative for Scientific Enhancement (MBRS RISE) Undergraduate Trainee

2008 NMSU Chemistry and Biochemistry Award

2006 Outstanding Biology Student Award

2005 National Hispanic Scholar

PUBLICATIONS

2016 **Johnston RA**, Paxton K, Wayne RK, Smith TB. 2016. Seasonal gene expression in a migratory songbird. In press. Molecular Ecology. Online ahead of print: doi:10.1111/mec.13879, PMID:27747949.

2016 Charrau P*, **Johnston RA*,** Stahler DR, Lea A, Snyder-Mackler N, Smith DW, vonHoldt BM, Cole SW, Tung J, Wayne RK. 2016. Pervasive effects of aging on gene expression in wild wolves. Molecular Biology and Evolution. 33(8):1967-78. doi:10.1093/molbev/msw072, PMID: 27189566. ***Co-first author**.

ORAL PRESENTATIONS

2016 RNA-Seq reveals widespread effects of aging on gene expression in wild wolves. EcoEvo Lunch, University of California, Los Angeles.

2014 Identifying Genes Associated with Migration in the Swainson's Thrush. Joint Meeting of the American Ornithologist's Union, Cooper Ornithological Society, Society of Canadian Ornithologists, Estes Park, CO.

2014 New approaches to investigate gene function in natural populations. HHMI Gilliam Fellows Meeting. HHMI Headquarters, Chevy Chase, Maryland.

2014 A systematic approach to investigate allele-specific gene function in the North American gray wolf. Evolution Conference, Raleigh, NC.

2013 On Science and Life. Keynote speaker, HHMI Research Scholars Graduation Banquet. New Mexico State University.

2009 Broad and Complex Role of the Broad-complex Gene in Drosophila Eye Development. New Mexico State University Undergraduate Research and Creative Arts Symposium.

CHAPTER 1:

SEASONAL GENE EXPRESSION IN A MIGRATORY SONGBIRD

MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 5680–5691 doi: 10.1111/mec.13879

Seasonal gene expression in a migratory songbird

RACHEL A. JOHNSTON,* KRISTINA L. PAXTON,†‡ FRANK R. MOORE,† ROBERT K. WAYNE* and THOMAS B. SMITH*§

*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 610 Charles E Young Dr. South, Rm. 4162, Los Angeles, CA 90095, USA, †Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA, ‡Department of Biology, University of Hawaii Hilo, Hilo, HI 96720, USA, §Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA 90095, USA

Abstract

The annual migration of a bird can involve thousands of kilometres of nonstop flight, requiring accurately timed seasonal changes in physiology and behaviour. Understanding the molecular mechanisms controlling this endogenous programme can provide functional and evolutionary insights into the circannual biological clock and the potential of migratory species to adapt to changing environments. Under naturally timed photoperiod conditions, we maintained captive Swainson's thrushes (*Catharus ustulatus*) and performed RNA sequencing (RNA-Seq) of the ventral hypothalamus and optic chiasma to evaluate transcriptome-wide gene expression changes of individuals in migratory condition. We found that 188 genes were differentially expressed in relation to migratory state, 86% of which have not been previously linked to avian migration. Focal hub genes were identified that are candidate variables responsible for the occurrence of migration (e.g. *CRABP1*). Numerous genes involved in cell adhesion, proliferation and motility were differentially expressed (including *RHOJ*, *PAK1* and *TLN1*), suggesting that migration-related changes are regulated by seasonal neural plasticity.

Keywords: *Catharus ustulatus*, circadian rhythm, circannual rhythm, migratory bird, photoperiod, Swainson's thrush

Received 16 June 2016; revision received 17 September 2016; accepted 21 September 2016

Introduction

Circannual rhythms are ubiquitous and include seasonal fluctuations in diverse biological processes such as reproductive status (Nakane & Yoshimura 2014), metabolism (Bairlein 2003; Ebling 2014) and the immune system (Dopico *et al.* 2015). One of the most remarkable examples of circannual rhythms is avian annual migration, the seasonal movement between breeding and nonbreeding areas in response to seasonal change (Dingle & Drake 2007). Migration involves striking seasonal alterations in physiological, morphological and behavioural characteristics (Berthold *et al.* 2003; Dingle 2006), such as feeding behaviour and metabolism to fuel flight muscles (Bairlein 2003), reduction and rebuilding of organs and muscle (McWilliams & Karasov 2005) and traits involved in sleep deprivation in nocturnal migrants (Rattenborg *et al.* 2004; Piersma *et al.* 2005; Fuchs *et al.* 2006). Molecular studies of migration can provide insight into the mechanisms by which such circannual rhythms are coordinated (Gwinner 1996; Dawson *et al.* 2001) and the adaptive potential of migratory species to changing environments. Currently, the molecular underpinnings of migration are poorly understood, and their degree of similarity across avian taxa has not been evaluated (Dingle 2006; van Noordwijk *et al.* 2006; Dingle & Drake 2007; Jones *et al.* 2008; Liedvogel *et al.* 2011; Mueller *et al.* 2011; Lundberg *et al.* 2013; Boss *et al.* 2016; Fudickar *et al.* 2016).

Regulation of seasonal migration potentially involves a variety of dynamic processes in the brain. Many of these processes likely occur in the hypothalamus, which integrates signals of circadian and circannual time (Majumdar *et al.* 2015) and regulates hormone release, energy balance, feeding and reproductive status (Yasuo

Correspondence: Rachel A. Johnston, Fax: +1 310 206 3987; E-mail: racheljo@g.ucla.edu

et al. 2003; Ebling & Barrett 2008; Barrett & Bolborea 2012; Dardente *et al.* 2014; Trivedi *et al.* 2014). An endogenous circannual clock, hypothesized to be linked to two or more of the master circadian pacemakers in the suprachiasmatic nuclei (SCN), pineal gland and retina (Kumar *et al.* 2004; Bartell & Gwinner 2005; Trivedi *et al.* 2016), is thought to ensure that seasonal changes occur at appropriate times of the year (Yasuo *et al.* 2003; Kumar *et al.* 2004; Bartell & Gwinner 2005; Rani *et al.* 2006; Dardente *et al.* 2014; Wood *et al.* 2015). Seasonal cellular remodelling in the hypothalamus, regulated by photoperiod, has been observed in birds (discussed in Yoshimura 2013; Migaud *et al.* 2015) and suggested to play roles in controlling hormone release (Yamamura *et al.* 2004, 2006). Changes in hormone signalling may contribute to alterations in multiple traits, such as impacts of hypothalamic thyroid hormone signalling on seasonal weight gain and increased energy consumption (Woods *et al.* 1998; Ebling & Barrett 2008; Ross *et al.* 2009; Barrett & Bolborea 2012), and, in spring migrants, effects of thyroid hormone signalling and gonadotropin releasing hormone on reproductive status (Dawson *et al.* 2001; Schwartz & Andrews 2013; Yoshimura 2013; Dardente *et al.* 2014; Tavolaro *et al.* 2015). In general, many cellular and physiological signals result in specific alterations of gene expression (Cheung & Kraus 2010). Therefore, assessment of transcriptome-wide gene expression changes in the brain provides the opportunity to evaluate migration-related regulatory processes at molecular, developmental and physiological levels.

Here, we evaluate gene expression changes associated with migration in the Swainson's thrush (*Catharus ustulatus*), a Neotropical, long-distance migratory songbird. We assessed expression of two geographically and genetically distinct subspecies (*C. u. swainsoni* and *C. u. ustulatus*, referred to here as 'inland' and 'coastal', respectively), which diverged approximately 10 000 years ago and exhibit dramatically different migratory pathways separated by a distinct migratory divide (Ruegg & Smith 2002; Ruegg *et al.* 2006a,b, 2014; Ruegg 2008; Delmore *et al.* 2012; Delmore & Irwin 2014). Migratory behaviour of the two subspecies is suggested to be under divergent selection (Ruegg *et al.* 2014; Delmore *et al.* 2015), but knowledge of which genes are involved in migration in this species is unknown and critical for evaluating this inference.

To examine transcriptome-wide expression of genes associated with migratory status, we performed RNA-Seq of Swainson's thrushes during nonmigratory and migratory states in controlled, captive conditions. First, we assessed migration-related gene expression and used network analyses to identify central processes and genes associated with migration, controlling for subspecies and time of day. Second, we tested for migratory expression in pathways hypothesized to be important for migration, including hormone signalling pathways and the circadian system. Finally, we compared migration-related gene expression between subspecies and evaluated overlap of migration-related genes in Swainson's thrushes with genes associated with migration in other bird species. This study provides functional insight into annual migration and reveals novel evidence of seasonal neural plasticity as a mechanism for regulating migration-related changes in physiology and behaviour.

Materials and methods

Bird capture and housing

The capture and handling of Swainson's thrushes was performed under federal Migratory Bird permit MB758364 and state permits SC-11246 (California) and 10–132 (Alaska) and approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Southern Mississippi (USM; protocol 09012601). Hatch-year birds were captured in Humboldt County, California (coastal birds), in June–July 2010 and near Fairbanks, Alaska (inland birds), in July 2010 with mist nets using both playbacks and passive netting. Individuals were housed at the USM animal facilities. Birds were caged individually and fed 70–75 g daily a semisynthetic diet consisting of hard-boiled eggs (25%), insects (freeze-dried crickets and shrimp, waxworms; 20.8%), cottage cheese (17.5%), blueberries (16.7%), crackers (12.5%), egg shells (5%), red meat (2.5%) and a vitamin supplement. Throughout the entire period of animal captivity, photoperiod was adjusted weekly to mimic each population's natural environment (Fig. S1, Supporting information). Using data collected from past research (Mack & Yong 2000), we estimated the timing of departure for fall migration from coastal (August 1) and inland breeding areas (August 15). To mimic the photoperiod that birds would typically experience during migration, changes in photoperiod during autumn migration were estimated using data collected from wild thrushes during migration, based on the estimated weekly location of each population (range: 800–1250 km/week, using an average speed of migration of 265 [±192] km/night; Cochran & Wikelski 2005]. In addition, we supplemented this information with migration passage dates from the literature (Mack & Yong 2000) and from long-term migration-banding data collected along the Gulf of Mexico (F. R. Moore, unpublished). During the year prior to this study, individual body weight, fat level, plasma metabolites levels, plasma hormone levels and caloric content of faeces

were measured weekly for a separate study (K. L. Paxton & F. R. Moore, unpublished).

Migratory activity and sampling time

To classify birds as nonmigratory or migratory, we quantified individual proportion nightly activity for the 2 weeks prior to sample collection using infrared motion detectors which record activity via data loggers (JoAC Electronik, Lund, Sweden) and activity analysis software (NI LabVIEW National Instruments, Austin, TX, USA). Proportion nightly activity was quantified as the proportion of 10-min intervals, between 30 min after lights were turned off and 30 min before lights were turned on, with >20 movements. Birds were considered nonmigratory if they never exhibited more than 0.05 proportion nightly activity during each night of the 2 weeks prior to animal sacrifice and tissue collection. Five birds from each subspecies were sacrificed and tissues collected within 2 days during summer (August 7–8, 2011; $N = 10$; Table S1, Supporting information). Six coastal and five inland birds were sacrificed and tissues collected within 2 days during autumn (October 3–4, 2011; $N = 11$; Table S1, Supporting information). Completion or moderate presence of post-breeding moult by August 1, as well as regressed testes in males, indicated that birds were post-reproductive photorefractory at the time of sampling except one coastal bird sacrificed on October 3 (individual Y56), which had enlarged testes and was fully excluded from the study. Of the remaining birds, birds sacrificed during summer were classified as nonmigratory except for individual R182, which exhibited migratory restlessness during the 2 weeks prior to tissue collection and therefore classified as migratory (Table S1, Supporting information). Birds sacrificed during the autumn were classified as migratory except individual Y57, which showed minimal nightly activity during the 2 weeks prior to tissue collection and therefore classified as nonmigratory (Table S1, Supporting information). The sample sizes of the study treatments were therefore: nonmigratory inland ($N = 4$), nonmigratory coastal ($N = 6$), migratory inland ($N = 6$) and migratory coastal ($N = 4$). All animals were sacrificed during daylight hours, ranging from 2.5 to 8.57 h after lights on (Table S1, Supporting information). As samples were not collected through the full 24-h cycle, we use 'circadian' gene expression in the adjective sense (relating to a 24-h day).

Tissue collection

Birds were deeply anesthetized with isoflurane and decapitated following USGS National Wildlife Health Center guidelines and with approval from the IACUC of USM. The brain was dissected following guidelines described for the zebra finch (Comito *et al.* 2016). Briefly, each brain was removed from the skull and placed ventral side up on a chilled petri dish. The hindbrain was removed, and two parasagittal cuts were made just medial to each optic tectum to remove the optic tecti. To ensure collection of the SCN of the hypothalamus, which is just dorsal to the optic chiasma, we collected the total tissue located medial to the two parasagittal cuts to a depth of approximately 2 mm. Thus, the tissue collected included the ventral hypothalamus and optic chiasma. The tissue was immediately flash frozen on dry ice, with no more than 10 min passing from the time each animal was sacrificed to the time the tissue was frozen on dry ice. Samples were shipped to UCLA in a liquid nitrogen tank and stored at −80 °C until RNA extraction.

RNA extraction and sequencing

Total RNA was extracted from tissues using the Trizol Plus RNA Purification Kit and included DNase treatment with the Ambion RNase-Free DNase Set and column cleanup with the PureLink RNA Mini Kit (Invitrogen, Carlsbad, CA, USA). Complimentary DNA (cDNA) libraries were generated with the TruSeq RNA Sample Preparation Kit v2 which includes poly-A selection (Illumina, San Diego, CA, USA) and sequenced as 100-base pair paired-end reads on the Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. Five barcoded samples were pooled per sequencing lane.

Quality filtering and mapping

A range of 31.2–48.8 million 100-base pair read pairs were generated for each Swainson's thrush sample (Table S1, Supporting information). Reads were trimmed with Trim Galore (available at www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove adaptors (read ends with three or more base pairs matching adapter sequence) and base pairs with Phred score <20 at the ends of reads. Read pairs with either read shorter than 25 base pairs were removed, resulting in 29.8–46 million high-quality read pairs per sample. We used Tophat2 (Kim *et al.* 2013) to map reads to the collared flycatcher genome (*Ficedula albicollis*, version 1.4.74) (Ellegren *et al.* 2012). We used the collared flycatcher genome as a reference because it contains fewer gaps and is better annotated than the draft assembly of the Swainson's thrush genome (Delmore *et al.* 2015). Moreover, because of its availability in Ensembl, the flycatcher genome is accessible for analytical tools that utilize the Ensembl genome database [e.g. gene ontology (GO) analysis with g:profiler]. In order to

allow for base pair differences due to cross-species divergence while maximizing the number of uniquely mapped reads, we used the following parameters for read mapping: read-mismatches 16, read-gap-length 10, read-edit-dist 28, splice-mismatches 1, max-insertion-length 12, max-deletion-length 12, max-multihits 5, b2-D 15, b2-N 1, b2-L 20, b2-i S,1,0.50. After aligning reads, we filtered for only uniquely mapped reads for further analyses. Overall, an average of 58.5% of the total sequenced reads per sample passed all filters and were used for gene expression quantification, ranging from 15 to 28.5 million read fragments per sample.

Gene expression quantification

Gene expression was quantified using HT-Seq (Anders *et al*. 2014) with the 'union' mode. If only one mate of a read pair mapped uniquely within the transcriptome, that read was kept and was counted equally as a read pair for performing gene expression counts, as both a single-end read and a paired-end read represent one RNA molecule. Expression values were normalized using the trimmed mean of *M*-values method in the DESEQ package (Anders & Huber 2010) in R (R Core Team 2013) and adjusted for gene length and GC content using the R package Conditional quantile normalization (CQN; Hansen *et al*. 2012). Genes were included in the analyses if they were annotated as protein coding and if their CQN-normalized and transformed expression values were ≥25 in at least 14 individuals. To identify sample outliers, we used the $Z.k_u$ score in the R package WGCNA (Langfelder & Horvath 2008), which specifies the number of standard deviations a sample's gene expression network connectivity is below the mean of all samples (Horvath 2011). Individual t22 was identified as an outlier and removed from further analyses because its $Z.k_u$ score was greater than the cut-off of 3.

Modelling

We first used principal components analysis of the normalized, log2-transformed expression data to identify variables potentially structuring overall expression data. Variables that significantly correlated with overall normalized expression data included subspecies [correlated with PC1 ($r = 0.7$, $P < 0.001$) and PC2 ($r = 0.596$, $P = 0.007$) of the normalized expression data], time of day [correlated with PC9 ($r = -0.756$, $P < 0.001$)], sex [correlated with PC3 ($r = 0.627$, $P = 0.004$)] and sequencing lanes [correlated with PC2 ($r = -0.6$, $P = 0.007$) and PC11 ($r = 0.602$, $P = 0.006$)]. Sequencing lanes and sex were regressed from the gene expression data (i.e. from the response variables) prior to linear modelling. Migration status, time of day of tissue collection and subspecies were included as covariates in the linear model. Specifically, we performed gene by gene multiple linear regression, with the explanatory variables defined as migration status, time of day, and subspecies, and the response variable defined as expression level of each of the 11 727 genes that passed the sequence coverage criteria. As individuals R182 and Y57 did not show typically timed migratory behaviour, the consistency of results in relation to the migratory classification of individuals R182 and Y57 was evaluated by performing the linear modelling excluding R182 and Y57. *P*-values were corrected for multiple testing using the *q*-value method in R (Storey & Tibshirani 2003). Given the limited sample size, we used a false discovery rate (FDR) threshold of 0.2 after performing 100 random permutations of each explanatory variable of interest (i.e. migration status, time of day and subspecies) to ensure that the empirical null distributions of *P*-values were uniform (Fig. S4, Supporting information). Gene annotations and orthologs for comparison of migration-related genes across species were acquired from Ensembl release 81.

Network analysis

Coexpression of genes significantly associated with migration or time of day (*Q*-value < 0.2) was analysed using the WGCNA package in R. Gene clusters were visualized using the plotNetworkHeatmap function with a network of signed correlations and the topological overlap matrix, which is a measure of gene network interconnectedness based on the number of neighbouring genes shared between two genes (Yip & Horvath 2007; Horvath 2011). Gene connectivity was measured by the K-within metric produced by the intramodularConnectivity function in WGCNA.

Gene ontology analysis

To identify molecular functions and biological processes represented in the gene expression data, we performed GO analyses using g:profiler version r1488_e83_eg30 (Reimand *et al*. 2007). Ensembl identification numbers of genes associated with migration, time of day or subspecies (*Q*-value < 0.2) were input as query lists, and all genes analysed ($N = 11\ 727$; see criteria above) were used as the background gene list. Query lists were divided into significantly upregulated and downregulated genes. The minimum allowed overlap between query genes and genes belonging to a GO term was set to 2. *P*-values were corrected using the Benjamini–Hochberg FDR method (Benjamini & Hochberg 1995).

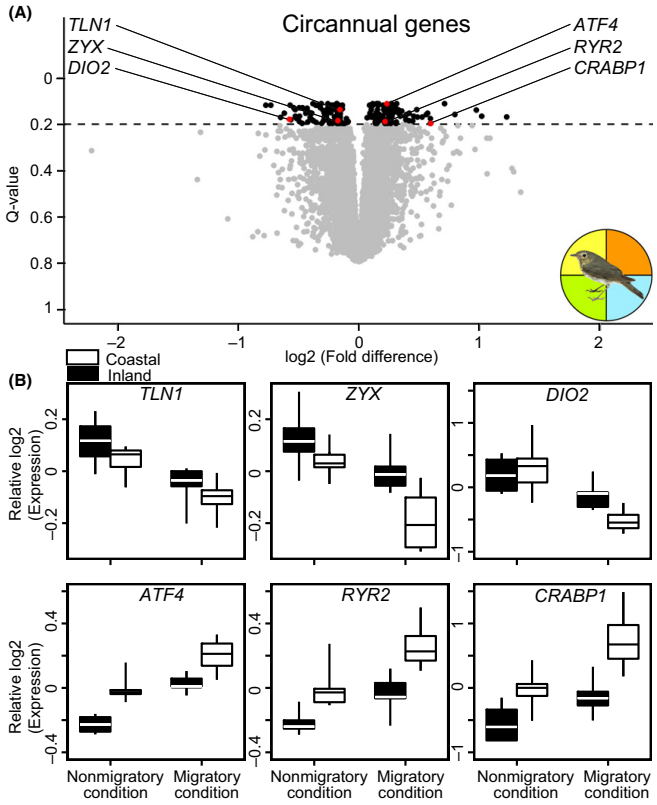

Fig. 1 Gene expression level differences with migratory status. (A) Volcano plot depicting differential gene expression of 11 727 protein-coding genes in Swainson's thrushes in the migratory state relative to the nonmigratory state. Significant genes (Q-value <0.2) are represented by black and red dots. Genes highlighted in red are depicted in (B). See Table S1 (Supporting information) for sample information and Table S2 (Supporting information) for comprehensive gene results. (B) Examples of genes exhibiting expression levels significantly associated with migratory status. Boxes represent the first through third quartiles of gene expression across birds. Horizontal lines represent the mean across birds, and vertical lines extend to the minimum and maximum expression values across birds.

Results

Migratory activity

To assess migratory status of Swainson's thrushes in captivity, we measured individual migratory restlessness, or *zugunruhe*, nightly during summer and fall prior to tissue sampling (Fig. S1, Supporting information). Individuals classified as migratory exhibited significantly more average migratory restlessness activity (mean of migratory group = 0.251, mean of nonmigratory group = 0.003) and greater body weight (mean of migratory group = 45.21 g, mean of nonmigratory group = 34.88 g; Welch's t-test; N = 20; d.f. = 17.376, P = 0.0052) than nonmigratory birds (Table S1, Supporting information). Within birds in the migratory condition, inland Swainson's thrushes exhibited significantly higher average nightly migratory restlessness activity than coastal Swainson's thrushes (mean of inland = 0.34, mean of coastal = 0.11; Welch's t-test; N = 10; d.f. = 7.387, P = 0.0448), concordant with the greater natural migration distance of inland Swainson's thrushes (Delmore *et al.* 2012).

Signature of cellular plasticity with migration

To evaluate transcriptome-wide gene expression changes associated with migration in the SCN, we performed RNA-Seq on the ventral hypothalamus and optic chiasma, located just ventral to the hypothalamus (see Materials and methods). We assessed expression differences with migratory state, controlling for time of day and subspecies, using a Q-value threshold of 0.2 (Storey & Tibshirani 2003). We detected 188 genes differentially expressed between Swainson's thrushes in nonmigratory and migratory states (Fig. 1; Table S2, Supporting information). These results were robust to the exclusion of two individuals that did not exhibit typically timed migratory behaviour (see Materials and methods; 91.5% genes still significant; Table S3, Supporting information). Genes upregulated in birds exhibiting migratory activity were

Table 1 Examples of enriched gene ontology (GO) terms and genes differentially expressed with autumn avian migration

Enriched GO term	GO corrected *P*-value	Gene symbol	Ensembl gene ID	Fold difference	*Q*-value
Cell migration	9.51E-3	*EMX2*	ENSFALG00000002201	1.507	0.176
		GLI3	ENSFALG00000005113	1.220	0.118
		EPHB1	ENSFALG00000013390	1.253	0.137
		FGF13	ENSFALG00000003780	1.228	0.118
Establishment of localization	6.77E-3	*CRABP1*	ENSFALG00000010635	1.511	0.196
		RYR2	ENSFALG00000013270	1.162	0.188
		RASL10B	ENSFALG00000009711	1.240	0.190
		IWS1	ENSFALG00000013574	1.066	0.153
Actin filament-based process	3.27E-3	*TLN1*	ENSFALG00000002197	0.897	0.137
		ZYX	ENSFALG00000004531	0.886	0.185
		RHOJ	ENSFALG00000003356	0.743	0.131
		PAK1	ENSFALG00000007597	0.815	0.141
		MYO1F	ENSFALG00000013814	0.784	0.118
		MYO1C	ENSFALG00000006676	0.837	0.129
		MSN	ENSFALG00000004282	0.830	0.168
		ARPC1B	ENSFALG00000013811	0.736	0.137
		TRIOBP	ENSFALG00000012797	0.826	0.165
Regulation of cell motility	4.31E-3	*PDGFA*	ENSFALG00000013959	0.882	0.137
		LAMA4	ENSFALG00000010965	0.756	0.150
		ADAM10	ENSFALG00000012254	0.866	0.189

Multiple linear regression was used to identify genes associated with avian migration controlling for subspecies and time of day. Genes were annotated based on Ensembl 81 of the *Ficedula albicollis* genome. See Table S4 (Supporting information) for a comprehensive list of GO terms. FD, fold difference.

significantly enriched with genes related to cell migration and cell development and include *EMX2*, *GLI3*, *FGF13* and *EPHB1* (*Q*-values <0.2; Tables 1, S2 and S4, Supporting information). Similarly, genes downregulated during migration were significantly enriched with genes related to cell adhesion and motility of cells, including *RHOJ* and its effector *PAK1*, *myosin IC* (*MYO1C*) *and myosin IF* (*MYO1F*), and adhesion complex molecules *talin-1* (*TLN1*), *zyxin* (*ZYX*) and *moesin* (*MSN*) (*Q*-values <0.2; Tables 1, S2 and S4, Supporting information).

Weighted gene coexpression network analysis (WGCNA) results indicate that within the 188 migration-related genes is a core cluster of 34 highly co-expressed genes (Fig. S2, Supporting information). Gene ontology analysis reveals that these genes are enriched for three groups of GO terms (Table S4, Supporting information), relating to 'anatomical structure morphogenesis' (*P* = 2.23E-2), 'neural precursor cell proliferation', (*P* = 3.13E-3) and 'protein localization to organelle' (*P* = 3.77E-2). This cluster is also enriched for genes involved in action potential and includes potassium channel gene *KCNB1*, potassium channel modulator *DPP6* and the calcium channel component *RYR2*.

Endocrine signalling associated with migration

We examined genes found to be migration-related (*N* = 188) for components of endocrine signalling pathways. *DIO2* was significantly downregulated with migration (*Q*-value = 0.179; fold difference (FD) = 0.671). This gene is known to be photo-responsive and regulates thyroid hormone by converting the prohormone T_4 to the active form of thyroid hormone, T_3 (Kohrle 1999), supporting a potential role for thyroid hormone in regulating seasonal changes in migratory birds. *CRABP1*, which is regulated by T_3 and regulates retinoic acid signalling, was significantly upregulated in migratory birds (*Q*-value = 0.196; FD = 1.511). Further, *CRABP1* had the third highest connectivity of all 188 migration-related genes after *SBK1* and *MTUS2* and was found to be a central hub gene within the core cluster of migration-related genes (Table S2, Supporting information).

Circadian system expression associated with migration

Given the potential role of the circadian system in regulating circannual rhythms (Kumar *et al.* 2004; Bartell & Gwinner 2005; Rani *et al.* 2006; Yoshimura 2013; Dardente *et al.* 2014; Singh *et al.* 2015), we tested whether expression of known avian circadian clock genes were altered in birds in migratory condition. We did not detect migration-related expression differences in the seven genes previously identified in the avian circadian clock [*PER2*, *PER3*, *CRY1*, *CRY2*, *CLOCK*, *ARNTL* (or *BMAL1*) and *ARNTL2* (or *BMAL2*; Kumar *et al.* 2010]. However, additional genes are

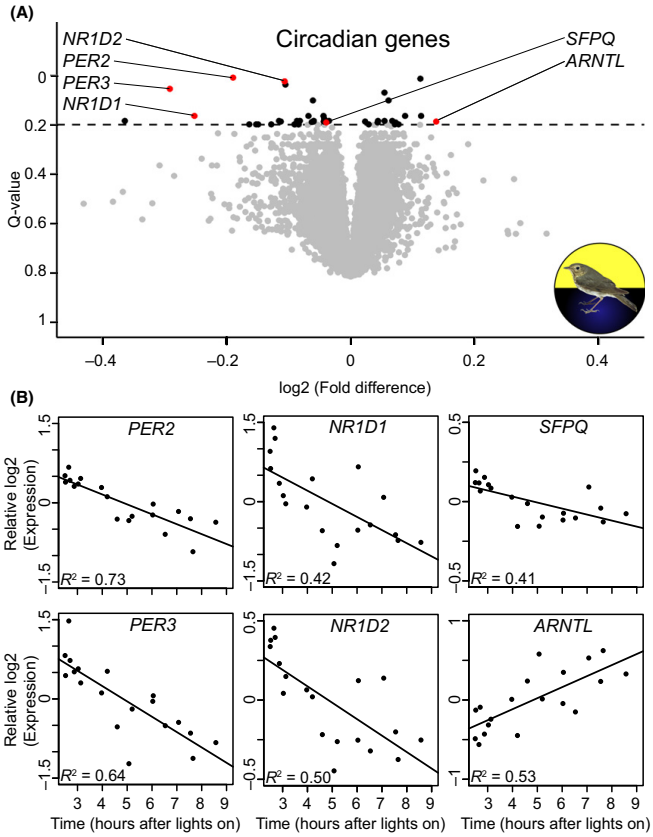

Fig. 2 Gene expression level differences with time of day. (A) Volcano plot depicting changes in gene expression per hour of day (measured only during lights on) in 11 727 protein-coding genes in Swainson's thrushes. Significant genes (Q-value <0.2) are represented by black and red dots. Genes highlighted in red are depicted in (B). See Table S2 (Supporting information) for a comprehensive gene list. (B) Examples of genes exhibiting expression levels significantly associated with time of day. Each dot represents the normalized, log2-transformed expression level of an individual Swainson's thrush.

likely involved in the avian clock, as several avian clock components are proposed based on knowledge of the mammalian clock, but have not been empirically studied (reviewed in Kumar *et al.* 2010). Therefore, we evaluated genes for expression level differences with time of day to identify novel components of the avian circadian clock and its target genes. We observed circadian-related expression in 43 genes, encompassing *PER2*, *PER3*, *ARNTL*, and 40 genes that have not been previously established as components or targets of the avian circadian clock (Q-values <0.2; Fig. 2; Table S2, Supporting information). Novel genes exhibiting circadian-related expression included *SFPQ* (also called *PSF*), which encodes a component of the PER complex in the negative loop of the mammalian circadian clock (Duong *et al.* 2011), and *NR1D1* (also called *Rev-erbα*) and *NR1D2* (also called *Rev-erbβ*), which are components of an accessory circadian feedback loop linking the negative and positive limbs of the mammalian circadian clock (Fig. 2) (Preitner *et al.* 2002; Cho *et al.* 2012). We then used WGCNA to identify genes that may be central drivers of the avian circadian clock. *PER2*, *NR1D2* and *PER3* had the highest intramodular connectivity of genes exhibiting circadian expression and were at the centre of a core cluster of coregulated genes also including *NR1D1* and *SFPQ* (Fig. S3, Supporting information). This is the first evidence that *SFPQ*, *NR1D1* and *NR1D2* are involved in the avian circadian clock. Outside this module, we also detected circadian expression of seven genes encoding heat shock and heat shock-related proteins: *HSF2*, *HSPA5*, *HSPH1*, *DNAJA1*, *DNAJA4*, *AHSA2* and *ENSFALG00000015235* (which belongs to the *HSP20* family) (Q-values <0.2; Table S2, Supporting information). We did not detect migration-related expression differences

in any of the 43 genes exhibiting circadian expression in Swainson's thrushes.

Expression differences between subspecies

Controlling for migration status and time of day of sample collection, 610 genes were differentially expressed between the inland and coastal subspecies (Q-value <0.2; Table S2, Supporting information). Genes differentially expressed between subspecies were highly enriched for migration-related genes, with 47 genes differentially expressed with migration and between subspecies (hypergeometric test; $N = 47$; $P = 4.7E\text{-}20$; Table S2, Supporting information). These 47 genes were over-represented with GO terms relating to regulation of gene expression ($P = 3.51E\text{-}0.5$), cell motility ($P = 7.85E\text{-}3$) and central nervous system neuron axonogenesis ($P = 6.84E\text{-}5$; Table S4, Supporting information). The enrichment of subspecies differences for cellular processes is specific to genes associated with migration, as overall subspecies differences are instead enriched for the opioid receptor signalling pathway ($P = 0.0015$; Table S4, Supporting information). Specifically, the inland subspecies exhibited upregulation of all four known opioid receptors (*OPRD1*, *OPRK1*, *OPRM21* and *OPRL1*), which have central roles in sensory perception (e.g. Mueller *et al*. 2010) and have been implicated in avian behaviours such as birdsong (Kelm *et al*. 2011).

Migration-related genes across species

We compared the 188 migration-related genes to genes with expression associated with migration in other bird species for which homologous annotations are available (Table S2, Supporting information), including willow warblers (*Phylloscopus trochilus*; $N = 1079$ genes) (Lundberg *et al*. 2013; Boss *et al*. 2016), dark-eyed juncos (*Junco hyamalis*; $N = 141$ genes) (Fudickar *et al*. 2016) and white-crowned sparrows (*Zonotrichia leucophrys*; $N = 11$ genes) (Jones *et al*. 2008). Genes differentially expressed in Swainson's thrushes during migration were significantly enriched for candidate migration genes identified in willow warblers (hypergeometric test; $N = 25$ genes; $P = 0.039$) and include the calcium channel component *RYR2* and the transcription factor *ATF4*, which acts as the culmination of multiple pathways to integrate stress signals to impact energy metabolism, amino acid transport, oxidative stress resistance and skeletal myofibre atrophy in mammals (Harding *et al*. 2003; Rutkowski & Kaufman 2003; Yoshizawa *et al*. 2009). One gene, *TMEM132B*, was associated with migration in dark-eyed juncos (Fudickar *et al*. 2016). No migration-related genes were shared with those identified in white-crowned sparrows (Jones *et al*. 2008). We did not detect migration-related expression of *ADCYAP1* or *CLOCK*, whose allele lengths have previously been associated with measures of migratory distance and phenology in blackcaps (*Sylvia atricapilla*) (Mueller *et al*. 2011; Mettler *et al*. 2015), dark-eyed juncos (Peterson *et al*. 2013), Wilson's warblers (*Cardellina pusilla*) (Bazzi *et al*. 2016), nightingales (*Luscinia megarhynchos*) and tree pipits (*Anthus trivialis*) (Saino *et al*. 2015).

Discussion

We identified 188 genes associated with fall migratory restlessness activity. The highly significant enrichment of migration-related genes for cell motility, cell adhesion and cell development provides evidence that migration-related circannual changes involve seasonal neural plasticity. Determining which molecular and cellular processes are necessary for and unique to migration will require functional comparative analyses across migratory and nonmigratory species.

The theoretical 'threshold model of migration' proposes that the tendency to migrate develops when a currently unknown continuous variable (e.g. the concentration of a protein or hormone), termed a 'liability' variable, reaches a certain threshold (Pulido *et al*. 1996; Pulido 2011). Given our focus on the hypothalamus and optic chiasma, where signals of photoperiod are thought to be integrated to elicit downstream responses, the genes we present here are candidate 'liability' variables involved in triggering the occurrence of migratory behaviour. Our utilization of WGCNA, combined with available information on specific functions of genes, highlights particularly strong candidate 'liability' genes. *DIO2*, via T_3 production, may be a driver of neural plasticity, as has been suggested in the nonmigratory Japanese quail (*Coturnix japonica*) for regulating gonadal growth and regression (Yamamura *et al*. 2004, 2006). *CRABP1* may also play a role in regulating the suite of expression changes associated with cellular plasticity, given that: (i) *CRABP1* was a hub gene with high connectivity to other migration-related genes; (ii) *CRABP1* is known to be regulated by T_3, the product of DIO2 activity (reviewed in Balmer & Blomhoff 2002; Mey & McCaffery 2004; Park *et al*. 2005; Maden 2007; Shearer *et al*. 2012); and (iii) thyroid hormone and retinoic acid signalling have been found to drive seasonal neural plasticity in other species (reviewed in Balmer & Blomhoff 2002; Park *et al*. 2005; Yamamura *et al*. 2006; Maden 2007; Horn & Heuer 2010; Shearer *et al*. 2010, 2012).

Elucidating the cellular changes that specifically occur between nonmigratory and migratory states will require additional evidence, such as ultrastructure imaging of

the hypothalamus. The observed differential expression of genes involved in cell motility, adhesion and development may reflect cellular morphological changes (discussed in Migaud *et al.* 2015; Wood *et al.* 2015). One possibility is that glial cells surrounding nerve terminals undergo morphological changes, which has been suggested to regulate hormone secretion in Japanese quail (Yamamura *et al.* 2004, 2006). Another possibility, which could be evaluated with immunohistochemistry, is neurogenesis. Neurogenesis has been reported in the adult hypothalamus of diverse species including zebrafish, sheep and hamsters (reviewed in Sousa-Ferreira *et al.* 2014). Supporting this potential process, we detected expression of markers for stem cells and cell proliferation known to be expressed in the neurogenic niche in the mammalian hypothalamus (Migaud *et al.* 2010, 2015) including *SOX2*, *NES*, *DCX* and *PCNA* (results not shown). Adult hypothalamic neurogenesis is thought to be important in regulating food intake in response to environmental and physiological signals (reviewed in Sousa-Ferreira *et al.* 2014) and could potentially play a role in regulating the striking hyperphagia of birds during migratory fattening.

Among the 188 genes differentially expressed with migration, a significant proportion of genes exhibited expression differences between the two subspecies (N = 47 genes; 25% of migration-related genes). Genetic divergence likely contributes to the observed expression differences between subspecies, given that the subspecies also exhibited expression differences in 558 genes regardless of migration state or time of day (i.e. independent of temporal measures). This possibility supports previous research suggesting divergent selection on migration in the Swainson's thrush (Ruegg *et al.* 2014; Delmore *et al.* 2015). However, the subspecies expression differences could also partly be due to gene expression responsiveness to the difference in photoperiod that each subspecies experienced (see Materials and methods). Research on the molecular evolution of these genes across natural populations, combined with further work on birds in controlled conditions, will be essential in understanding the roles of adaptive and acclimatory responses in generating interindividual variation in migration-related phenotypes.

Expression differences of 188 genes between birds in migratory and nonmigratory states, combined with enrichment of genes associated with cellular plasticity, demonstrate the extensive changes that may seasonally occur in the brain for this remarkable behavioural phenomenon. We suggest the possibility that seasonal neural plasticity has a greater role in this physiological and behavioural transition than previously thought. Network analyses reveal hub genes, potential 'liability' variables, whose expression may be pivotal in regulating the onset of migratory behaviour. Understanding the expression thresholds of such variables for actuating migration, and their sensitivity to environmental conditions, will be critical for predicting the response of migratory animals to future climate and environmental changes.

Acknowledgements

This study would not have been possible without the help of Christine Ames, Emily Cohen, Jill Gautreaux, Kimberly Hollinger, Matthew Johnson, Emijo Laine, Eben Paxton, Abby Powell, C.J. Ralph, Jacqueline Smolinski, Katherine Vowell, Kevin Winker, T.J. Zenzal and the Migratory Bird Research Group. We thank the animal care facility staff at Humboldt State University, University of Alaska Fairbanks and University of Southern Mississippi for assistance maintaining birds in captivity. We are thankful to Kristen Ruegg for helpful discussions on Swainson's thrush biology and genotyping and to Ryan Harrigan and Barney Schlinger for comments on the manuscript. We thank Amanda Lea, Noah Snyder-Mackler and Jenny Tung for guidance on data analysis. This work was supported by the National Science Foundation (grant number IOS 844703 to F.R.M. and grant number DEB 1257716 to R.K.W.), a Howard Hughes Medical Institute Gilliam Fellowship (R.A.J.), UCLA's Center for Tropical Research (T.B.S. and R.A.J.) and the Moore Basic Research Development Account, USM (F.R.M. and K.L.P.).

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.

Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, btu638.

Bairlein F (2003) Nutritional strategies in migratory birds. In: *Avian Migration* (ed. Berthold P, Gwinner E, Sonnenschein E), pp. 321–332. Springer Science & Business Media, Berlin.

Balmer JE, Blomhoff R (2002) Gene expression regulation by retinoic acid. *Journal of Lipid Research*, **43**, 1773–1808.

Barrett P, Bolborea M (2012) Molecular pathways involved in seasonal body weight and reproductive responses governed by melatonin. *Journal of Pineal Research*, **52**, 376–388.

Bartell PA, Gwinner E (2005) A separate circadian oscillator controls nocturnal migratory restlessness in the songbird Sylvia borin. *Journal of Biological Rhythms*, **20**, 538–549.

Bazzi G, Galimberti A, Hays QR *et al.* (2016) Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (Cardellina pusilla). *Ecology and Evolution*, **6**, 3226–3239.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Berthold P, Gwinner E, Sonnenschein E (2003) *Avian Migration*. Springer Science & Business Media, Berlin.

Boss J, Liedvogel M, Lundberg M *et al.* (2016) Gene expression in the brain of a migratory songbird during breeding and migration. *Movement Ecology*, **4**, 4.

Cheung E, Kraus WL (2010) Genomic analyses of hormone signaling and gene regulation. *Annual Review of Physiology*, **72**, 191–218.

Cho H, Zhao X, Hatori M *et al.* (2012) Regulation of circadian behaviour and metabolism by REV-ERB-[agr] and REV-ERB-[bgr]. *Nature*, **485**, 123–127.

Cochran WW, Wikelski M (2005) Individual migratory tactics of New World Catharus thrushes. In: *Birds of Two Worlds: The Ecology and Evolution of Migration* (ed. Greenberg R, Marra PP), pp. 274–289. Johns Hopkins University Press, Baltimore, Maryland.

Comito D, Pradhan DS, Karleen BJ, Schlinger BA (2016) Region-specific rapid regulation of aromatase activity in zebra finch brain. *Journal of Neurochemistry*, **136**, 1177–1185.

Dardente H, Hazlerigg DG, Ebling FJ (2014) Thyroid hormone and seasonal rhythmicity. *Frontiers in Endocrinology*, **5**, 19.

Dawson A, King VM, Bentley GE, Ball GF (2001) Photoperiodic control of seasonality in birds. *Journal of Biological Rhythms*, **16**, 365–380.

Delmore KE, Irwin DE (2014) Hybrid songbirds employ intermediate routes in a migratory divide. *Ecology Letters*, **17**, 1211–1218.

Delmore KE, Fox JW, Irwin DE (2012) Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas, revealed using light-level geolocators. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 4582–4589.

Delmore KE, Hübner S, Kane NC *et al.* (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, **24**, 1873–1888.

Dingle H (2006) Animal migration: is there a common migratory syndrome? *Journal of Ornithology*, **147**, 212–220.

Dingle H, Drake VA (2007) What is migration? *BioScience*, **57**, 113–121.

Dopico XC, Evangelou M, Ferreira RC *et al.* (2015) Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nature Communications*, **6**, 7000.

Duong HA, Robles MS, Knutti D, Weitz CJ (2011) A molecular mechanism for circadian clock negative feedback. *Science*, **332**, 1436–1439.

Ebling FJ (2014) On the value of seasonal mammals for identifying mechanisms underlying the control of food intake and body weight. *Hormones and Behavior*, **66**, 56–65.

Ebling FJ, Barrett P (2008) The regulation of seasonal changes in food intake and body weight. *Journal of Neuroendocrinology*, **20**, 827–833.

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, **30**, 207–210.

Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, **491**, 756–760.

Fuchs T, Haney A, Jechura TJ *et al.* (2006) Daytime naps in night-migrating birds: behavioural adaptation to seasonal sleep deprivation in the Swainson's thrush, Catharus ustulatus. *Animal Behaviour*, **72**, 951–958.

Fudickar AM, Peterson MP, Greives TJ *et al.* (2016) Differential gene expression in seasonal sympatry: mechanisms involved in diverging life histories. *Biology Letters*, **12**, 20160069.

Gwinner E (1996) Circannual clocks in avian reproduction and migration. *Ibis*, **138**, 47–63.

Hansen KD, Irizarry RA, Wu ZJ (2012) Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, **13**, 204–216.

Harding HP, Zhang Y, Zeng H *et al.* (2003) An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Molecular Cell*, **11**, 619–633.

Horn S, Heuer H (2010) Thyroid hormone action during brain development: more questions than answers. *Molecular and Cellular Endocrinology*, **315**, 19–26.

Horvath S (2011) *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer Science & Business Media, New York.

Jones S, Pfister-Genskow M, Cirelli C, Benca RM (2008) Changes in brain gene expression during migration in the white-crowned sparrow. *Brain Research Bulletin*, **76**, 536–544.

Kelm CA, Forbes-Lorman RM, Auger CJ, Riters LV (2011) Mu-opioid receptor densities are depleted in regions implicated in agonistic and sexual behavior in male European starlings (Sturnus vulgaris) defending nest sites and courting females. *Behavioural Brain Research*, **219**, 15–22.

Kim D, Pertea G, Trapnell C *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, **14**, R36.

Kohrle J (1999) Local activation and inactivation of thyroid hormones: the deiodinase family. *Molecular and Cellular Endocrinology*, **151**, 103–119.

Kumar V, Singh BP, Rani S (2004) The bird clock: a complex, multi-oscillatory and highly diversified system. *Biological Rhythm Research*, **35**, 121–144.

Kumar V, Wingfield JC, Dawson A *et al.* (2010) Biological clocks and regulation of seasonal reproduction and migration in birds. *Physiological and Biochemical Zoology*, **83**, 827–835.

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Liedvogel M, Akesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution*, **26**, 561–569.

Lundberg M, Boss J, Canback B *et al.* (2013) Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler Phylloscopus trochilus. *Bmc Genomics*, **14**, 330.

Mack DE, Yong W (2000) Swainson's Thrush (*Catharus ustulatus*). In: *The birds of North America* (ed. Poole A, Gill F), pp. 32. The Birds of North America, Inc., Philadelphia, Pennsylvania.

Maden M (2007) Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience*, **8**, 755–765.

Majumdar G, Rani S, Kumar V (2015) Hypothalamic gene switches control transitions between seasonal life history states in a night-migratory photoperiodic songbird. *Molecular and Cellular Endocrinology*, **399**, 110–121.

McWilliams SR, Karasov WH (2005) Migration takes guts. In: *Birds of two worlds* (ed. Greenberg R, Marra PP), pp. 67–78. Johns Hopkins University Press, Baltimore, Maryland.

Mettler R, Segelbacher G, Schaefer HM (2015) Interactions between a candidate gene for migration (ADCYAP1),

morphology and sex predict spring arrival in blackcap populations. *PLoS One*, **10**, e0144587.
Mey J, McCaffery P (2004) Retinoic acid signaling in the nervous system of adult vertebrates. *Neuroscientist*, **10**, 409–421.
Migaud M, Batailler M, Segura S *et al.* (2010) Emerging new sites for adult neurogenesis in the mammalian brain: a comparative study between the hypothalamus and the classical neurogenic zones. *European Journal of Neuroscience*, **32**, 2042–2052.
Migaud M, Butrille L, Batailler M (2015) Seasonal regulation of structural plasticity and neurogenesis in the adult mammalian brain: focus on the sheep hypothalamus. *Frontiers in Neuroendocrinology*, **37**, 146–157.
Mueller C, Klega A, Buchholz HG *et al.* (2010) Basal opioid receptor binding is associated with differences in sensory perception in healthy human subjects: a [F-18]diprenorphine PET study. *NeuroImage*, **49**, 731–737.
Mueller JC, Pulido F, Kempenaers B (2011) Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 2848–2856.
Nakane Y, Yoshimura T (2014) Universality and diversity in the signal transduction pathway that regulates seasonal reproduction in vertebrates. *Frontiers in Neuroscience*, **8**, 115.
van Noordwijk AJ, Pulido F, Helm B *et al.* (2006) A framework for the study of genetic variation in migratory behaviour. *Journal of Ornithology*, **147**, 221–233.
Park SW, Li G, Lin YP *et al.* (2005) Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. *Molecular Cell*, **19**, 643–653.
Peterson MP, Abolins-Abols M, Atwell JW *et al.* (2013) Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research*, **2**, 115.
Piersma T, Perez-Tris J, Mouritsen H *et al.* (2005) Is there a "migratory syndrome" common to all migrant birds? *Annals of the New York Academy of Sciences*, **1046**, 282–293.
Preitner N, Damiola F, Zakany J *et al.* (2002) The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell*, **110**, 251–260.
Pulido F (2011) Evolutionary genetics of partial migration—the threshold model of migration revis(it)ed. *Oikos*, **120**, 1776–1783.
Pulido F, Berthold P, van Noordwijk AJ (1996) Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. *Proceedings of the National Academy of Sciences, USA*, **93**, 14642–14647.
R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Rani S, Malik S, Trivedi AK *et al.* (2006) A circadian clock regulates migratory restlessness in the blackheaded bunting, Emberiza melanocephala. *Current Science*, **91**, 1093–1096.
Rattenborg NC, Mandt BH, Obermeyer WH *et al.* (2004) Migratory sleeplessness in the white-crowned sparrow (Zonotrichia leucophrys gambelii). *PLoS Biology*, **2**, E212.
Reimand J, Kull M, Peterson H *et al.* (2007) g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, **35**, W193–W200.
Ross AW, Johnson CE, Bell LM *et al.* (2009) Divergent regulation of hypothalamic neuropeptide Y and agouti-related protein by photoperiod in F344 rats with differential food intake and growth. *Journal of Neuroendocrinology*, **21**, 610–619.
Ruegg K (2008) Genetic, morphological, and ecological characterization of a hybrid zone that spans a migratory divide. *Evolution*, **62**, 452–466.
Ruegg KC, Smith TB (2002) Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (Catharus ustulatus). *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 1375–1381.
Ruegg K, Slabbekoorn H, Clegg S, Smith TB (2006a) Divergence in mating signals correlates with ecological variation in the migratory songbird, Swainson's thrush (Catharus ustulatus). *Molecular Ecology*, **15**, 3147–3156.
Ruegg KC, Hijmans RJ, Moritz C (2006b) Climate change and the origin of migratory pathways in the Swainson's thrush, Catharus ustulatus. *Journal of Biogeography*, **33**, 1172–1182.
Ruegg K, Anderson EC, Boone J *et al.* (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, **23**, 4757–4769.
Rutkowski DT, Kaufman RJ (2003) All roads lead to ATF4. *Developmental Cell*, **4**, 442–444.
Saino N, Bazzi G, Gatti E *et al.* (2015) Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology*, **24**, 1758–1773.
Schwartz C, Andrews MT (2013) Circannual transitions in gene expression: lessons from seasonal adaptations. *Current Topics in Developmental Biology*, **105**, 247–273.
Shearer KD, Goodman TH, Ross AW *et al.* (2010) Photoperiodic regulation of retinoic acid signaling in the hypothalamus. *Journal of Neurochemistry*, **112**, 246–257.
Shearer KD, Stoney PN, Morgan PJ, McCaffery PJ (2012) A vitamin for the brain. *Trends in Neurosciences*, **35**, 733–741.
Singh D, Trivedi AK, Rani S *et al.* (2015) Circadian timing in central and peripheral tissues in a migratory songbird: dependence on annual life-history states. *The FASEB Journal*, **29**, 4248–4255.
Sousa-Ferreira L, de Almeida LP, Cavadas C (2014) Role of hypothalamic neurogenesis in feeding regulation. *Trends in Endocrinology and Metabolism*, **25**, 80–88.
Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, **100**, 9440–9445.
Tavolaro FM, Thomson LM, Ross AW *et al.* (2015) Photoperiodic effects on seasonal physiology, reproductive status and hypothalamic gene expression in young male F344 rats. *Journal of Neuroendocrinology*, **27**, 79–87.
Trivedi AK, Kumar J, Rani S, Kumar V (2014) Annual life history–dependent gene expression in the hypothalamus and liver of a migratory songbird insights into the molecular regulation of seasonal metabolism. *Journal of Biological Rhythms*, 0748730414549766.
Trivedi AK, Malik S, Rani S, Kumar V (2016) Pinealectomy abolishes circadian behavior and interferes with circadian clock gene oscillations in brain and liver but not retina in a migratory songbird. *Physiology & Behavior*, **156**, 156–163.
Wood SH, Christian HC, Miedzinska K *et al.* (2015) Binary Switching of Calendar Cells in the Pituitary Defines the Phase of the Circannual Cycle in Mammals. *Current Biology*, **25**, 2651–2662.

Woods SC, Seeley RJ, Porte D, Schwartz MW (1998) Signals that regulate food intake and energy homeostasis. *Science*, **280**, 1378–1383.

Yamamura T, Hirunagi K, Ebihara S, Yoshimura T (2004) Seasonal morphological changes in the neuro-glial interaction between gonadotropin-releasing hormone nerve terminals and glial endfeet in Japanese quail. *Endocrinology*, **145**, 4264–4267.

Yamamura T, Yasuo S, Hirunagi K *et al.* (2006) T3 implantation mimics photoperiodically reduced encasement of nerve terminals by glial processes in the median eminence of Japanese quail. *Cell and Tissue Research*, **324**, 175–179.

Yasuo S, Watanabe M, Okabayashi N *et al.* (2003) Circadian clock genes and photoperiodism: comprehensive analysis of clock gene expression in the mediobasal hypothalamus, the Suprachiasmatic nucleus, and the pineal gland of Japanese quail under various light schedules. *Endocrinology*, **144**, 3742–3748.

Yip AM, Horvath S (2007) Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, **8**, 1.

Yoshimura T (2013) Thyroid hormone and seasonal regulation of reproduction. *Frontiers in Neuroendocrinology*, **34**, 157–166.

Yoshizawa T, Hinoi E, Jung DY *et al.* (2009) The transcription factor ATF4 regulates glucose metabolism in mice through its expression in osteoblasts. *The Journal of Clinical Investigation*, **119**, 2807.

R.A.J., K.L.P., F.R.M. and T.B.S. performed conceptualization. R.A.J., K.L.P., F.R.M., R.K.W. and T.B.S. performed methodology. R.A.J. performed formal analysis and investigation. K.L.P. and F.R.M. provided resources. R.A.J. wrote the original draft of the manuscript. R.A.J., K.L.P., F.R.M., R.K.W. and T.B.S. carried out manuscript reviewing and editing, and F.R.M., R.K.W. and T.B.S. performed funding acquisition.

Data accessibility

The sequence data are available in the NCBI Gene Expression Omnibus repository (Edgar *et al*. 2002) and are accessible through GEO: GSE87549.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Photoperiodic conditions and migratory restlessness activity of Swainson's thrushes during the project.

Fig. S2 Covariation heatmap of gene expression variation across Swainson's thrushes of genes associated with migration.

Fig. S3 Covariation heatmap of gene expression variation across Swainson's thrushes of genes associated with circadian time of day.

Fig. S4 Distribution of *P*-values from the linear analysis vs. the empirically derived null for each variable of interest: (A) migration status, (B) time of day, and (C) subspecies.

Table S1 Biological and technical information of all Swainson's thrushes used in the RNA-Seq analyses.

Table S2 Gene-by-gene results of all 11 727 genes analyzed in the study.

Table S3 Gene-by-gene results excluding samples r182 and y57.

Table S4 Results from all GO analyses performed in g:profiler.

Supplemental Results

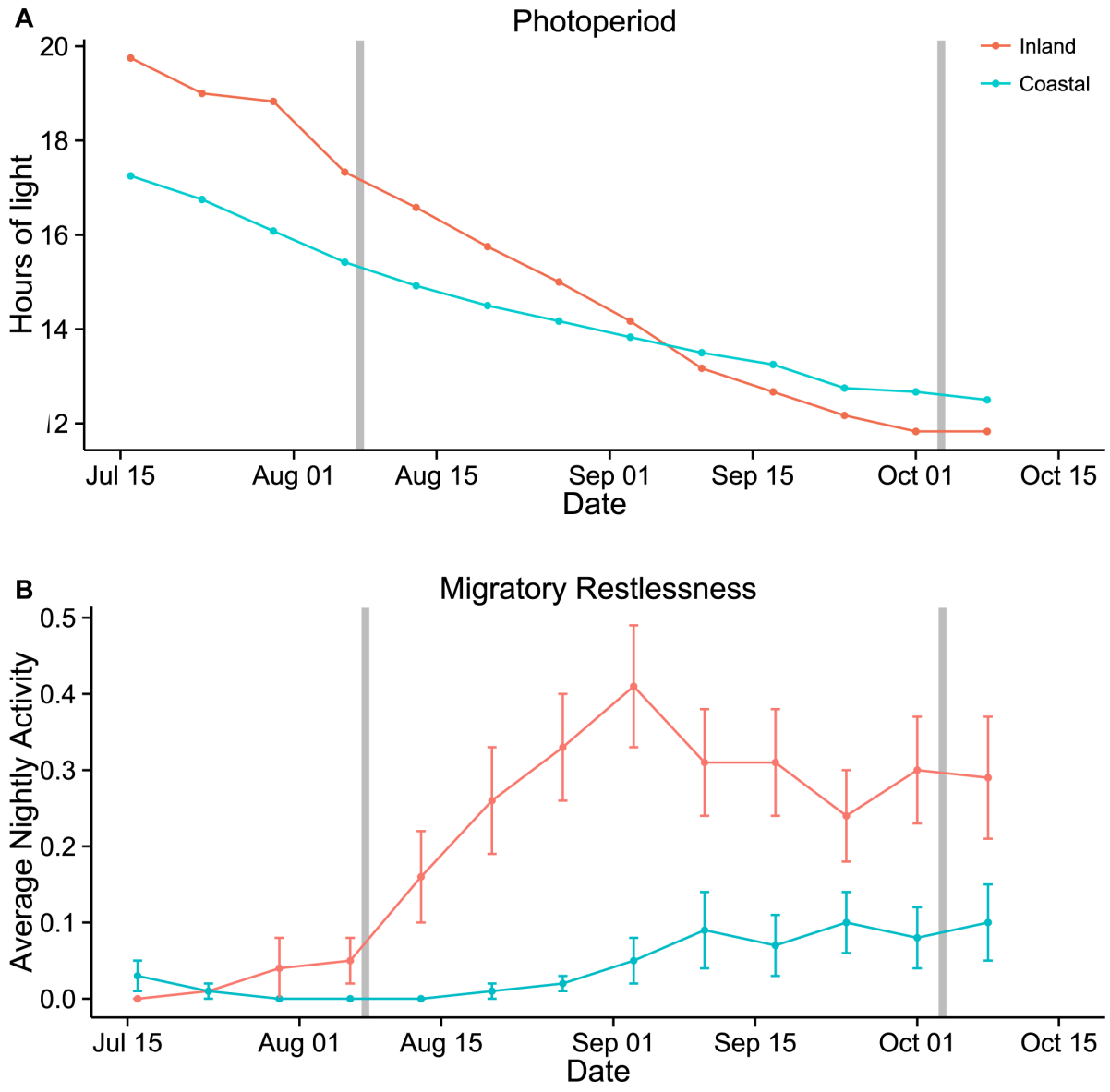

Figure 1-S1. Photoperiodic conditions and migratory restlessness activity of Swainson's thrushes during the project. Gray bars represent days of sample collection for non-migratory and migratory birds. A) Number of hours of daily light (photoperiod) that Swainson's thrushes from each population were exposed to. B) Average nightly activity of inland and Swainson's thrushes per week. Sample sizes prior to summer sacrifice were inland = 19, coastal = 18 and post summer sacrifice were inland = 14, coastal = 13. Error bars represent the standard error of activity across individuals.

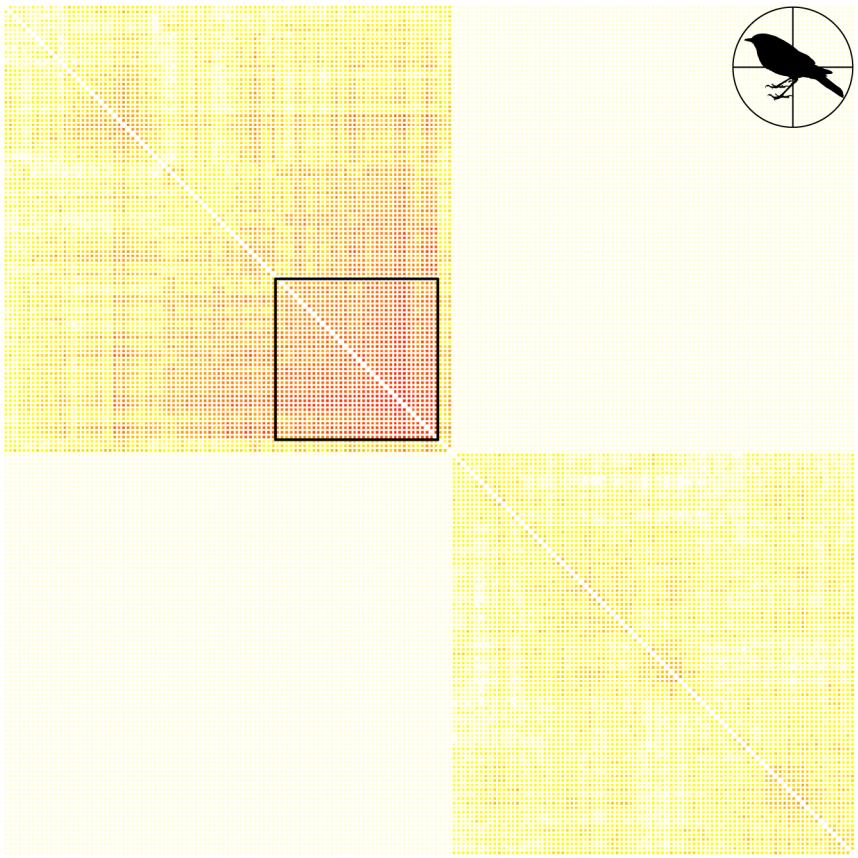

Figure 1-S2. Covariation heatmap of gene expression variation across Swainson's thrushes of genes associated with migration. Each row and column correspond to a gene, and each pixel represents the level of similarity of expression of two genes across Swainson's thrushes using the topological overlap measure in WGCNA (Langfelder and Horvath, 2008). Genes are separated into genes up-regulated (top left) and down-regulated (bottom right) with migration. The black box indicates the module of highly correlated genes. Gene labels and connectivity values are presented in Table S2.

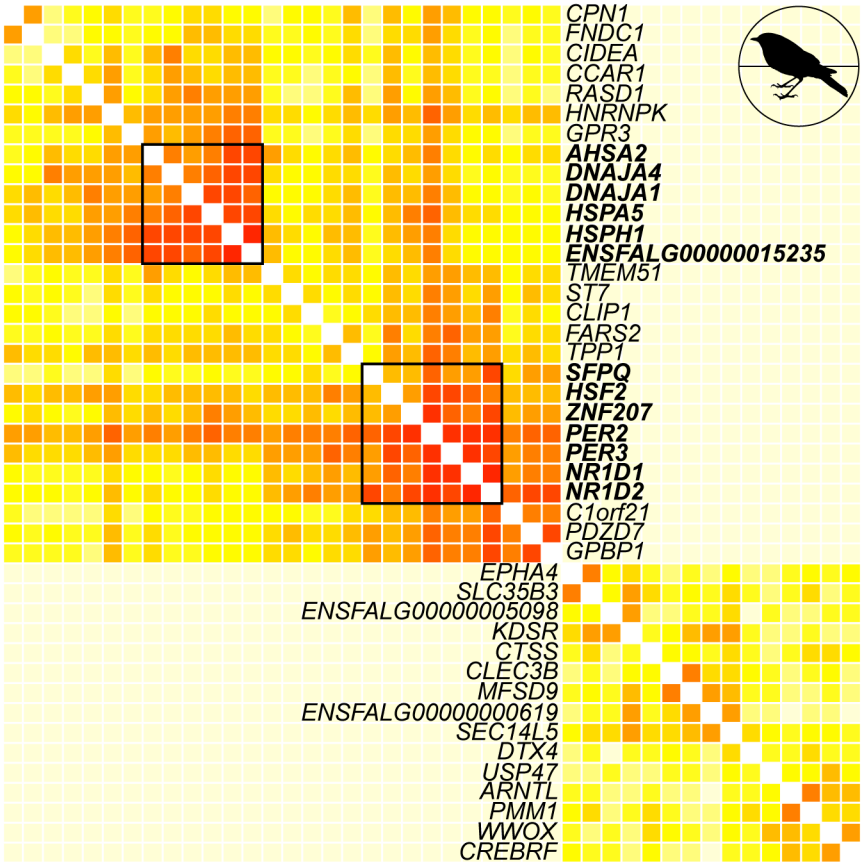

Figure 1-S3. Covariation heatmap of gene expression variation across Swainson's thrushes of genes associated with circadian time of day. Each row and column correspond to a gene, and each pixel represents the level of similarity of expression of two genes across Swainson's thrushes using the topological overlap measure in WGCNA (Langfelder and Horvath, 2008). Genes are separated into genes down-regulated (top left) and up-regulated (bottom right) with time of day. The black boxes indicate the modules of highly correlated genes. Gene connectivity values are presented in Table S2.

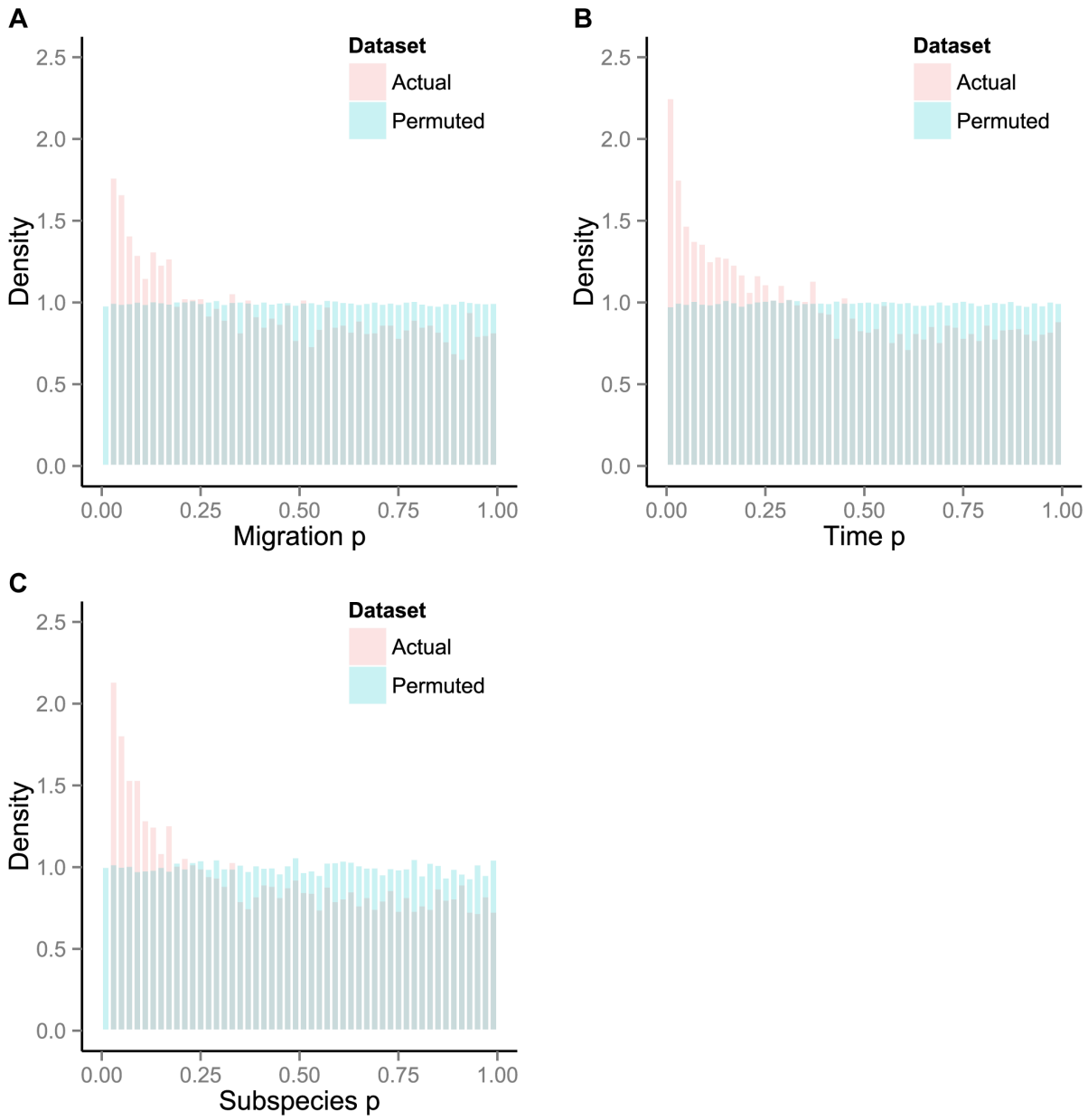

Figure 1-S4. Distribution of p-values from the linear analysis versus the empirically derived null for each variable of interest: A) migration status, B) time of day, and C) subspecies. The empirically derived null distribution for each variable of interest was produced from 100 permutations of the linear model, with the only alteration of each permutation being a random scrambling of the variable of interest. The actual p-value of each gene is presented in Table S2.

Supplemental Results Bibliography

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9, 559.

CHAPTER 2:

PERVASIVE EFFECTS OF AGING ON GENE EXPRESSION IN WILD WOLVES

Pervasive Effects of Aging on Gene Expression in Wild Wolves

Pauline Charruau,[†,‡,1] Rachel A. Johnston,[†,1] Daniel R. Stahler,[2] Amanda Lea,[3] Noah Snyder-Mackler,[4] Douglas W. Smith,[2] Bridgett M. vonHoldt,[5] Steven W. Cole,[6,7] Jenny Tung,[3,4] and Robert K. Wayne[*,1]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles

[2]Yellowstone Center for Resources, National Park Service, Yellowstone National Park

[3]Department of Biology, Duke University

[4]Department of Evolutionary Anthropology, Duke University

[5]Department of Ecology and Evolutionary Biology, Princeton University

[6]Department of Medicine, University of California, Los Angeles

[7]Cousins Center for Psychoneuroimmunology, Semel Institute, University of California, Los Angeles

[†]These authors contributed equally to this work.

[‡]Present address: Department of Ecology and Evolution, University of Lausanne, Switzerland

***Corresponding author:** E-mail: rwayne@biology.ucla.edu.

Associate editor: Meredith Yeager

Abstract

Gene expression levels change as an individual ages and responds to environmental conditions. With the exception of humans, such patterns have principally been studied under controlled conditions, overlooking the array of developmental and environmental influences that organisms encounter under conditions in which natural selection operates. We used high-throughput RNA sequencing (RNA-Seq) of whole blood to assess the relative impacts of social status, age, disease, and sex on gene expression levels in a natural population of gray wolves (*Canis lupus*). Our findings suggest that age is broadly associated with gene expression levels, whereas other examined factors have minimal effects on gene expression patterns. Further, our results reveal evolutionarily conserved signatures of senescence, such as immunosenescence and metabolic aging, between wolves and humans despite major differences in life history and environment. The effects of aging on gene expression levels in wolves exhibit conservation with humans, but the more rapid expression differences observed in aging wolves is evolutionarily appropriate given the species' high level of extrinsic mortality due to intraspecific aggression. Some expression changes that occur with age can facilitate physical age-related changes that may enhance fitness in older wolves. However, the expression of these ancestral patterns of aging in descendant modern dogs living in highly modified domestic environments may be maladaptive and cause disease. This work provides evolutionary insight into aging patterns observed in domestic dogs and demonstrates the applicability of studying natural populations to investigate the mechanisms of aging.

Key words: **RNA-Seq, transcriptome, senescence, immunosenescence, *Canis lupus*, canid.**

Introduction

In contrast to genetic changes occurring across generations, alterations of gene expression allow immediate and acutely sensitive responses to changing conditions. Gene expression can be affected by multiple factors such as age (Göring et al. 2007; Hong et al. 2008; López-Otín et al. 2013; Rahman et al. 2013; Peters et al. 2015), social stressors (Weaver et al. 2006; Cole et al. 2007; McGowan et al. 2009; Cole et al. 2011, 2012; Tung et al. 2012; Murphy et al. 2013; Powell et al. 2013), and infectious disease status (Whitney et al. 2003; Cobb et al. 2005; Ramilo et al. 2007; Blankley et al. 2014; Mejias and Ramilo 2014). Aging is associated with a multitude of gene expression changes in blood including cellular senescence and dysfunctional immune responses (e.g., immunosenescence and inflammaging) (Boren and Gershwin 2004; Kirkwood 2005; Baylis et al. 2013). In contrast, chronic stress due to adverse social environments activates immune pathways and can affect antiviral responses, susceptibility to infectious agents, wound healing abilities, and individual fitness (Sapolsky 2005; Weaver et al. 2006; Miller et al. 2009; Chen et al. 2011; Cole et al. 2011, 2012; Tung et al. 2012; Murphy et al. 2013; Powell et al. 2013).

Investigations of gene expression patterns in natural populations of nonmodel species can reveal which environmental factors are most influential on the regulation of specific molecular pathways. However, with the exception of baboons (Runcie et al. 2013; Tung et al. 2015), the transcriptome-wide effects of intrinsic and environmental variables on gene expression in vertebrates have only been studied in humans and species in captivity. Knowledge regarding the generality of aging effects on gene expression patterns across species is further limited by short lifespan biases of model species and unintended artificial selection on aging rates (Ricklefs 2010). Thus, despite a long history of research on the diverse ecology,

social structures, and life histories of wild animal populations (e.g., Wilson 2000), the impacts of aging and environmental factors on interindividual variation at a functional genomic level in natural animal populations remain poorly understood (Ricklefs 2010).

To address this knowledge gap, we used RNA sequencing (RNA-Seq) to quantify genome-wide expression levels in wild gray wolves (*Canis lupus*) from Yellowstone National Park (YNP), USA. Gray wolves serve as an excellent model for studying the influences of age and social factors on gene expression levels, as wolves live in highly cooperative societies characterized by social hierarchies (Mech 1999; Packard 2003). Further, the divergence time of approximately 80 My between wolves and more extensively studied primate and murine models (dos Reis et al. 2012) affords the opportunity to assess the long-term evolutionary conservation of expression patterns. We take advantage of the exceptional observational database available for the YNP wolf population, in which factors potentially affecting gene expression such as social status, age, and disease presence have been extensively documented since the wolf reintroduction in 1995–1996 (vonHoldt et al. 2008, 2010; Almberg et al. 2009; MacNulty, Smith, Vucetich et al. 2009; Smith et al. 2010; Almberg et al. 2012; Smith et al. 2012; Stahler et al. 2013; Cubaynes et al. 2014).

Wolf packs consist of an adult pair of breeding wolves (denoted alpha), their offspring (denoted subordinates), and additional adult wolves, which may breed but are considered subordinate to alphas (denoted beta) (Mech 1999, 2003; Packard 2003). The breeding pair is considered dominant to other individuals, and social hierarchy is most commonly demonstrated by nonaggressive behavior and submissive postures resulting in unchanged or slightly lower glucocorticoid levels in subordinates (Mech 1999; Packard 2003; Smith et al. 2012; Molnar et al. 2015). This social system is in contrast to hierarchies in which dominance is maintained by displacements and threats toward lower ranking individuals, such as in wild baboons and macaques, which result in higher levels of stress in low-ranking individuals (Sapolsky 2004, 2005). Given the widespread impacts that social stress can have on individual health and the immune system as detected in blood (Sapolsky 2005; Weaver et al. 2006; Miller et al. 2009; Chen et al. 2011; Cole et al. 2011, 2012; Tung et al. 2012; Murphy et al. 2013; Powell et al. 2013), we aimed to compare influences of social rank on gene expression levels in blood of wolves, which could be collected with minimum handling, to those observed in primate social systems.

Active disease infection also impacts gene expression variation (Whitney et al. 2003; Cobb et al. 2005; Ramilo et al. 2007; Blankley et al. 2014; Mejias and Ramilo 2014). Sarcoptic mange is an infectious disease frequently observed in YNP gray wolves, induced by the mite *Sarcoptes scabiei* and analogous to human scabies (Almberg et al. 2012, 2015). *Sarcoptes* mites are known to elicit gene expression changes in humans and trigger allergic reactions causing skin inflammation, itching, and hair loss in wolves (Arlian et al. 2004, 2006; Almberg et al. 2012). Therefore, we hypothesized that presence of mange in YNP wolves results in expression signatures of immunosuppression and inflammatory response (Arlian et al. 2004, 2006; Velavan and Ojurongbe 2011; Gregori et al. 2012).

Finally, aging is also expected to be a factor affecting gene expression in gray wolves. We predict that rapid aging observed at the physical level in wolves (MacNulty, Smith, Vucetich et al. 2009; Stahler et al. 2013; Cubaynes et al. 2014) would be mirrored by gene expression changes with age. Although the molecular impact of age has been well described in humans (Göring et al. 2007; Hong et al. 2008; López-Otín et al. 2013; Rahman et al. 2013; Peters et al. 2015), the relative influence and specific pathways altered by age in a wild population confronted with multiple environmental challenges is unknown. In baboons, the only wild species assessed to date, the impact of age on gene expression was found to be modest, and the molecular pathways altered by age were not evaluated (Runcie et al. 2013; Tung et al. 2015). Therefore, to assess conservation of patterns across the mammalian phylogeny, and because impacts of aging on gene expression of specific pathways are virtually unknown for wild populations, we compared our age results to humans (Göring et al. 2007; Hong et al. 2008).

To investigate natural variation in gene expression levels, we use a linear mixed model to simultaneously assess the effects of social rank (alpha vs. all subordinates), age, disease status (infected with mange vs. not infected), and sex on genome-wide expression in whole blood while controlling for genetic relatedness. As these factors can also influence cell-type abundance in blood (Linton and Dorshkind 2004; Cole et al. 2011, 2012; Baylis et al. 2013; Powell et al. 2013), we perform transcript origin analysis (TOA) (Cole et al. 2011) and transcriptome representation analysis (TRA) (Powell et al. 2013) to infer the cellular origins of differentially expressed transcripts. Among the variables studied, we find that gene expression variation among wild wolves is primarily explained by age, with multiple age-related genes and processes conserved between wolves and humans. In contrast, we observe a surprising absence of expression differences with other variables such as rank, indicating that impacts of rank in primate models cannot be extended to the effects of social hierarchy in other vertebrate species.

Results

Sequence Data

To assess transcriptome variation across gray wolves in their natural environment, we quantified gene expression levels from whole blood samples of 27 wolves using RNA-Seq. A total of 1.26 billion 100 base-pair reads of sufficient quality (Phred score $>$ 20) and length ($>$ 25 bp after trimming) were generated (mean of 46.69 million reads per individual; supplementary table S1, Supplementary Material online). On average, 87.5% of these reads aligned to the dog genome (CanFam3.1.74) and up to 35% of the uniquely mapped reads aligned to annotated protein-coding regions (supplementary table S1, Supplementary Material online). Of the 19,856 protein-coding genes annotated in the dog genome, we obtained 13,558 genes (68.3%) with sufficient read depth across

individuals (≥10 reads in at least 75% of cDNA libraries) to be included in expression analyses. The number of detected genes in whole blood of wolves ($N = 13,558$) was similar to that found in lymphocytes of humans ($N = 12,758$) (Göring et al. 2007) and whole blood of baboons ($N = 10,409$) (Tung et al. 2015).

Association of Gene Expression Levels with Intrinsic and Environmental Factors

We assessed the effects of intrinsic and environmental factors on transcriptome-wide gene expression using a significance cut-off of false discovery rate (FDR) = 0.2 given the limited sample size. Similar to other studies that evaluated effects of sex on blood gene expression in mammals (e.g., Whitney et al. 2003; Tung et al. 2015), we observed minimal expression differences between male and female wolves, with only one gene significantly associated with sex, ENSCAFG00000030886 (fold difference [FD] = 0.08; Q value = 0.1; fig. 1A and supplementary table S2, Supplementary Material online). This protein-coding gene belongs to the thymosin β4 family but has not been formally named. Further, despite the presence of physical symptoms characteristic of mange in seven of the sampled wolves, we did not identify any genes differentially expressed in wolves affected by mange (fig. 1B). Finally, we predicted that social rank would be broadly associated with gene expression levels, comparable to effects of social conditions in humans and nonhuman primates (Cole et al. 2007, 2012; Tung et al. 2012). However, we did not identify any genes in which expression levels were significantly associated with social rank (fig. 1C).

In contrast, controlling for rank, sex, and mange infection status, we detected 625 genes associated with age, comprising 214 genes up-regulated and 411 genes down-regulated with age with diverse functions (fig. 1D and table 1 and supplementary table S2, Supplementary Material online). These age-related gene expression changes may have been induced by intracellular changes in gene transcription or by age-related changes in proportions of blood cell-types (Linton and Dorshkind 2004; Baylis et al. 2013). To evaluate potential changes in blood cell composition with age, we performed TOA (table 2), which infers the cellular origin of transcripts (Cole et al. 2011), in parallel with TRA (table 3), which assesses changes in cell abundance based on genes previously identified to have highly cell-type-specific expression (Powell et al. 2013). In accordance with observations of immunosenescence in humans, genes down-regulated with age were largely associated with B cells ($P < 0.001$; table 2), which appear to partly be due to a lower prevalence of B cells in older wolves (TRA; FD per year = 0.98; $P = 0.001$; table 3). Further, consistent with development of more pro-inflammatory phenotypes in older individuals (reviewed in Boren and Gershwin 2004; Baylis et al. 2013), genes up-regulated with age were significantly associated with innate immunity cells including dendritic cells ($P < 0.001$; table 2) and natural killer cells ($P = 0.004$; table 2). TRA indicated a slightly increased proportion of monocytes in older individuals (FD per year = 1.01; $P < 0.001$; table 3).

To assess the conservation of aging processes across species, we compared the 625 age-related genes in wolf leukocytes to a catalog of age-related genes in human lymphocytes (Göring et al. 2007; Hong et al. 2008). Wolf age-related genes were significantly enriched with human age-related genes, as we found that of the 420 age-related wolf genes represented in the human study, 255 (60.7%) genes were associated with age in humans (hypergeometric test; $P = 0.046$). The majority of these overlapping genes were associated with age in the same direction (60.8% of overlapping genes, $N = 155$ concordant genes, $\chi^2 = 11.863$, df = 1, $P = 0.001$).

A diversity of genes and biological processes we found to be associated with age in wolves (table 1) has previously been associated with aging in humans (Boren and Gershwin 2004; Kirkwood 2005; Baylis et al. 2013). For example, we observed significant age-related down-regulation of *ETS1* (FD per year = 0.900; Q value < 0.001; fig. 1D and 2 and table 1 and supplementary table S2, Supplementary Material online), a transcription factor involved in a variety of functions including regulation of the immune response (Garrett-Sinha 2013) and cellular senescence (Ohtani et al. 2001). Similarly, *CDK4*, which is known to be inhibited by p16, a key protein activated during cellular senescence (Krishnamurthy et al. 2004; Liu et al. 2009; López-Otín et al. 2013), was down-regulated with age (FD = 0.909; Q value = 0.137; fig. 2 and table 1 and supplementary table S2, Supplementary Material online). Additionally, we observed significant age-related expression differences in genes associated with telomere maintenance (e.g., *DKC1*, a component of the telomerase holoenzyme; FD = 0.948; Q value = 0.2; and *NOP56*; FD = 0.958; Q value = 0.198; table 1 and supplementary table S2, Supplementary Material online) (Poncet et al. 2008), DNA repair and genomic stability (e.g., *SIRT6*; FD = 0.896; Q value = 0.2) (Mostoslavsky et al. 2006; Di Mauro and David 2009), and epigenetic regulation (genes down-regulated with age were enriched for the gene ontology (GO) terms "chromatin organization," $P = 0.005$ and "histone modification," $P = 0.011$; table 1 and supplementary table S3, Supplementary Material online).

Paralleling intracellular senescence, aging has been found to dysregulate the propensity of the adaptive immune system to respond to novel infectious disease (a phenomenon known as "immunosenescence") (Baylis et al. 2013; López-Otín et al. 2013). In agreement with previous work, our results suggest reduced activity of B and T cells with age, as interleukin and interleukin-related genes, required for differentiation and proliferation of lymphocytes, were down-regulated in older wolves (*IL7*, *IL21R*, *IL27RA*; Q values < 0.2; fig. 2 and table 1 and supplementary table S2, Supplementary Material online). In addition, we observed age-related down-regulation of multiple MHC class II genes (Dog Leukocyte Antigen genes *DLA-DQA1*, *DLA-DRA*, *DLA-88*, and *DLA-DOB*; Q values < 0.2; fig. 2 and table 1 and supplementary table S2, Supplementary Material online), whose expression is likely limited to antigen-presenting cells (B cells, monocytes) and activated T cells. GO analyses further support immunosenescence, as 30% of terms enriched in the genes down-regulated with age are explicitly related to

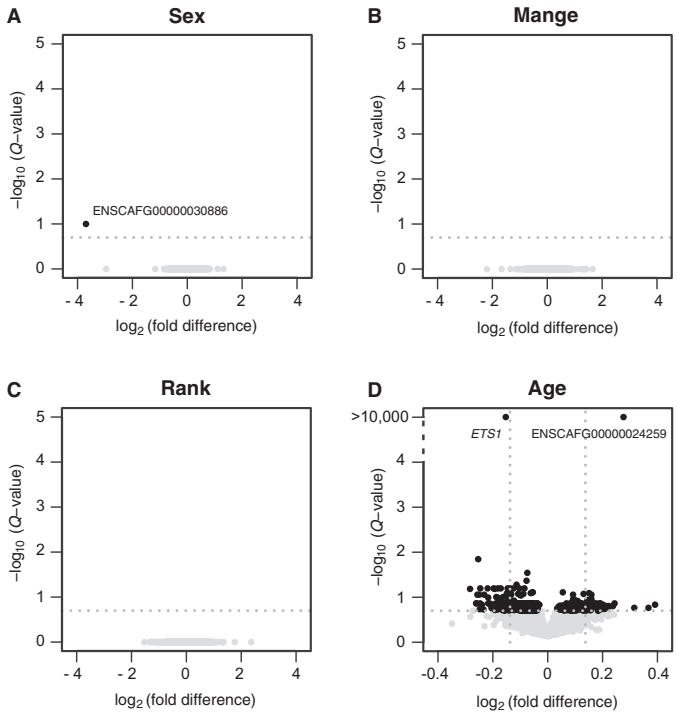

Fig. 1. Difference in gene expression according to sex, mange, social rank, and age in wolf blood. (*A*–*D*) Black dots represent genes for which expression is significantly associated with a variable, whereas gray dots illustrate nonsignificant genes. The significance threshold (Q value = 0.2) is delimited by the horizontal line $y = -\log_{10}(0.2)$. FD is expression of females relative to males (*A*), mange-affected wolves relative to nonaffected individuals (*B*), alpha wolves relative to subordinates (*C*), and difference per year of age (*D*) for 13,558 genes expressed in wolf whole blood. (*D*) Vertical lines depict the thresholds used in the TOA (table 3).

Table 1. Biological Processes and Genes Associated with Age in Gray Wolves.

Enriched GO Term[a]	GO *P* Value	Gene Symbol	Ensembl Gene ID	Annual FD	*Q* Value[b]
Regulation of metabolic process	**2.25E-5**	***ETS1***	**ENSCAFG00000010304**	**0.900**	**<0.001**
		CDK4	**ENSCAFG00000000280**	**0.910**	**0.137**
RNA metabolic process	**1.91E-5**	***DKC1***	**ENSCAFG00000019620**	**0.948**	**0.175**
		NOP56	**ENSCAFG00000006651**	**0.958**	**0.198**
Chromatin modification	**2.97E-3**	***SIRT6***	**ENSCAFG00000019123**	**0.896**	**0.200**
		HIRA	**ENSCAFG00000014619**	**0.974**	**0.137**
Immune response	**6.28E-6**	***DLA-DQA1***	**ENSCAFG00000000812**	**0.945**	**0.137**
		DLA-DRA	**ENSCAFG00000000803**	**0.950**	**0.146**
		DLA-88	**ENSCAFG00000000487**	**0.945**	**0.180**
		DLA-DOB	**ENSCAFG00000000819**	**0.883**	**0.175**
Lymphocyte activation	**1.55E-9**	***IL7***	**ENSCAFG00000008373**	**0.907**	**0.180**
		IL27RA	**ENSCAFG00000016504**	**0.919**	**0.146**
		CD5	**ENSCAFG00000016332**	**0.938**	**0.157**
		CXCR5	**ENSCAFG00000012439**	**0.898**	**0.175**
Lipid metabolic process	**1.17E-2**	***LEP***	**ENSCAFG00000001672**	**1.078**	**0.151**
		LIAS	**ENSCAFG00000015991**	**1.035**	**0.183**
		LIPH	**ENSCAFG00000013246**	**1.113**	**0.184**

[a]A comprehensive gene list and full GO results are presented in supplementary tables S2–S4, Supplementary Material online.
[b]Q values of genes significantly associated with age after controlling for relatedness, social rank, mange presence, and sex using linear mixed effects models.

immune system processes (e.g., lymphocyte activation, antigen receptor-mediated signaling pathway, T- and B-cell activation and proliferation, and interleukin-2 production; supplementary table S3, Supplementary Material online).

In contrast to the abundance of immunity-related GO terms enriched in the genes down-regulated with age in wolves, GO terms enriched in genes up-regulated with age were dominated by biological processes related to lipid

Table 2. Cell Types Inferred to Mediate Age-Related Gene Expression Differences.

Cell Type	*P* Value	
	FD > 1.10[a]	FD < 0.90[b]
$CD14^+$ monocytes	0.035	0.998
$BDCA4^+$ dendritic cells	<0.001*	0.973
$CD56^+$ NK cells	0.004*	0.338
$CD4^+$ T cells	0.762	0.424
$CD8^+$ T cells	0.460	0.768
$CD19^+$ B cells	0.126	<0.001*

NOTE.—NK, natural killer. FD, fold difference per year of age.
[a]TOA results of up-regulated genes ($N = 137$ genes).
[b]TOA results of down-regulated genes ($N = 111$ genes).
*Significant ($P < 0.01$) overrepresentation of genes predominantly expressed by a specific cell type.

Table 3. Leukocyte Prevalence Differences with Age in Wolf Blood.

Cell Types	Genes $Z > 6$[a]	FD	*P* Value[b]
$CD14^+$ monocytes	337	1.01	<0.001*
$BDCA4^+$ dendritic cells	250	1.00	0.912
$CD56^+$ NK cells	349	1.00	0.043
$CD4^+$ T cells	62	1.00	0.606
$CD8^+$ T cells	45	1.01	0.220
$CD19^+$ B cells	87	0.98	0.001*

NOTE.—FD, fold difference in cell prevalence per year of age; NK, natural killer.
[a]Genes diagnostic of a cell type are defined by average expression in a cell type > 6 standard deviations above the mean (Genes $Z > 6$) (Powell et al. 2013).
[b]TRA results of cell-type differences with age.
*Significant ($P < 0.01$) difference in cell-type prevalence.

metabolism (supplementary table S4, Supplementary Material online) and included *leptin* (*LEP*), *lipoic acid synthetase*, and *lipase member H* (*LIPH*; *Q* values < 0.2; fig. 2 and table 1 and supplementary table S2, Supplementary Material online). We also observed an age-related increase in expression of *insulin-like growth factor 1 receptor* (*IGF1R*; FD = 1.038; *Q* value = 0.188; fig. 2 and supplementary table S2, Supplementary Material online), which has been found to regulate lifespan (Holzenberger et al. 2003).

Discussion

We examined factors that we predicted to influence gene expression in gray wolves based on gene expression patterns observed in humans and other primates, which included sex, mange presence, social rank, and age. The negligible effect of sex on gene expression in blood is consistent with similar findings in humans and macaques (Whitney et al. 2003; Tung et al. 2015). However, the lack of detectable differentially expressed genes in mange-infected wolves was unexpected, given the presence of visible symptoms of mange such as hairless patches with skin lesions (Almberg et al. 2015). This result may partly be due to limited statistical power but may also be caused by the immunosuppressive properties that *S. scabiei* evoke in their hosts by impacting IL10 signaling (Arlian et al. 2006). The possibility of altered IL10 signaling in wolves with mange is supported by the presence of *IL10RA* in the top five genes associated with mange in our data set, although this result is not significant at our *Q*-value threshold of 0.2.

The absence of a significant effect of rank on gene expression was also unexpected but likely reflects a less hierarchically based social system in wolves compared to many primates species. In wolves, the term alpha denotes the breeding pair of a pack (which are typically the parents of other pack members) and signifies little to no differences in stress levels, resource access, or received aggression from other wolves (Mech 1999; Cubaynes et al. 2014; Molnar et al. 2015). Additionally, wolf rank and age are positively correlated (Pearson correlation $r = 0.357$; $P = 0.068$) and may have effects on the same biological pathways (Snyder-Mackler et al. 2014), potentially limiting our statistical power to detect rank effects independent of aging.

Overall, we found that age has a much broader impact on gene regulation than sex, mange infection status, or rank in a social top-predator exposed to naturally occurring abiotic and biotic challenges. Wolves show signatures of senescence consistent with aging in model organisms and humans (Baylis et al. 2013; Beirne et al. 2014), which supports the hypothesis that aging effects on cellular processes are evolutionarily conserved through mammal phylogeny (Palacios et al. 2011; Nussey et al. 2013). Theory predicts that higher levels of extrinsic mortality should lead to an increased rate of senescence because a relatively larger amount of energy is expected to be dedicated to reproduction than to somatic maintenance and repair (Kirkwood 1977, 2005). The short time period over which age-related expression differences manifest in wolves (spanning a few years compared to decades in humans) (Göring et al. 2007) is evolutionarily appropriate given the high extrinsic mortality in this species, due mainly to interpack strife (Cubaynes et al. 2014), and rapid physical senescence (MacNulty, Smith, Mech et al. 2009; MacNulty, Smith, Vucetich et al. 2009; Stahler et al. 2013).

Stress related to changes in rank, physical injuries, and disease (Almberg et al. 2015) may augment the expression signature of aging in wolves or result in expression differences unique from that observed in humans. Genes associated with age in wolves were significantly enriched for human age-related genes, yet we detected 165 age-related genes in wolves that were not found to change with age in humans (Göring et al. 2007). Further, 100 genes associated with age in both wolves and humans showed opposite trends of expression levels with age between the two species (Göring et al. 2007). A longitudinal study, and comparison to gene expression in captive wolves, would be valuable in decomposing the specific factors that have long-lasting impacts on gene expression in wolves.

We identified age-related changes in the expression of genes that may be critical in the progression of the aging process, such as up-regulation of *IGF1R*, the primary receptor of the *IGF1* protein product. *IGF1* signaling is thought to be an evolutionarily conserved regulator of the trade-off between growth and survival, as the *IGF1* pathway promotes growth during early development but can increase the rate of aging later in life (Rollo 2002; Salminen and Kaarniranta 2010). The existence of a trade-off in *IGF1* signaling is supported by studies across domestic dog breeds, in which plasma IGF1 protein levels across breeds positively correlate with body size but

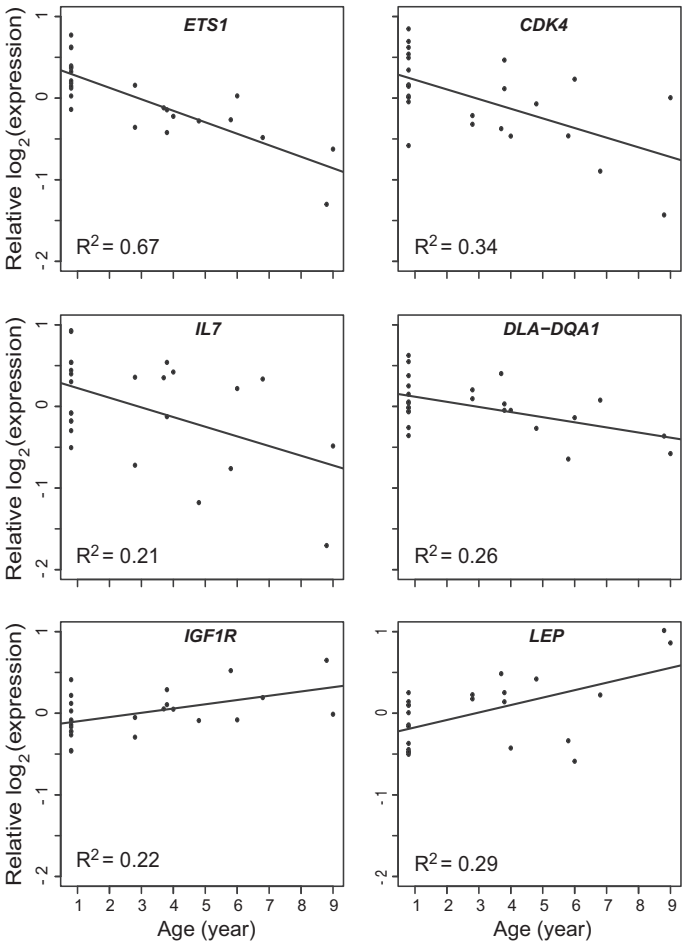

Fig. 2. Expression of a subset of genes significantly associated with age in gray wolves. Relative expression represents the normalized, log 2-transformed expression levels of wolves after controlling for relatedness using linear mixed effects models. A comprehensive gene list is available in supplementary table S2, Supplementary Material online.

negatively correlate with lifespan (Sutter et al. 2007; Hoopes et al. 2012). This mechanism has been suggested to be the reason that small dog breeds can have twice the lifespan of larger breeds (Kraus et al. 2013). The rapid development of yearling wolves (MacNulty, Smith, Mech et al. 2009; MacNulty, Smith, Vucetich et al. 2009) likely requires early *IGF1R* expression. However, the function of up-regulated *IGF1R* we observe in old wolves is unknown and is potentially nonadaptive. *IGF1R* is down-regulated with age in humans (Göring et al. 2007), which may be an important contributor to human longevity. Consequently, up-regulation of *IGF1R* may be a leading driver of the rapid senescence of wolves and other short-lived species.

We also observed increased expression of *LEP* in older wolves, concordant with increased leptin protein levels observed in aging humans, dogs, and model species (Mobbs 1998; Ishioka et al. 2002, 2007; Ma et al. 2002). Leptin is produced by adipocytes and reduces reflex appetite (Sánchez-Rodríguez et al. 2000). Its increase with age, which is often disproportionately more than would follow from an increase in adipose tissue, is thought to indicate a leptin-resistant state in elderly individuals and can contribute to obesity (Ahren et al. 1997; Li et al. 1997; Sánchez-Rodríguez et al. 2000; Ma et al. 2002). Because the weight of wolves is strongly correlated with age (Pearson correlation $r = 0.768$; $P = 2.967 \times 10^{-5}$; supplementary table S1, Supplementary Material online), we do not have the statistical power to test whether LEP expression increases with age more than expected by weight gain alone. However, the overall weight gain and increased expression of *LEP* with age in wolves

supports the hypothesis that wild wolves undergo metabolic aging (MacNulty, Smith, Mech et al. 2009).

Although we find age-related gene expression patterns shared by humans and wolves, their impacts on health and fitness may be vastly different for each species. For example, metabolic aging and obesity are recognized as having negative health impacts in older humans (e.g., Alexander et al. 2003). In contrast, increased weight in wolves enhances reproductive success in females of all ages (Stahler et al. 2013) and dispatching ungulate prey in males and females (MacNulty, Smith, Mech et al. 2009). Further, male wolves exhibit increasing aggressiveness with age, which may facilitate dominance and breeding opportunities within a pack and enhance success during interpack territorial encounters (Cassidy KA, Mech LD, MacNulty DR, Stahler DR, Smith DW, personal communication). Thus, although senescence does occur in wolves, some age-related changes may also allow older wolves to maintain fitness and, consequently, can be maintained by selection (e.g., Schwarz et al. 2016). In dogs, which live in human-dominated environments, such age-related changes may be maladaptive and result in morbidity. Likewise, gene expression patterns that evolved under a hunter-gatherer lifestyle in humans may lead to health problems for individuals with a western lifestyle and diet (Carrera-Bastos et al. 2011; Farooqui 2015). Consequently, for both modern humans and their domestic companions, such legacy effects may result in disease in aging individuals.

Our results establish the ancestral baseline for gene expression in dogs and show that transcriptome-wide gene expression analysis applied to natural populations has the potential to reveal novel functional and evolutionary insights into the mechanisms and drivers of aging. In general, our study suggests that aging is the critical factor affecting gene expression levels in a wild wolf population and provides a new paradigm for investigating the aging process in natural systems.

Materials and Methods

Sampling and Phenotypic Traits

Whole blood was collected from 27 wild gray wolves in YNP, USA. Sampling was conducted during winters 2011, 2012, and 2013 during annual wolf captures according to US national guidelines for handling animals and with all required permits held by the National Park Service (Smith et al. 2012). Individual animal information is provided in supplementary table S1, Supplementary Material online.

Age and Sex

For each of the 27 wolves, sex was determined during handling. Males were coded as 0 and females as 1. Ages were estimated from an assumed birth date of April 15 (mean whelp date in study area) (Stahler et al. 2013) and were continuously coded for the linear models (supplementary table S1, Supplementary Material online and supplementary fig. S2, Supplementary Material online). Age was determined by either capturing pups ($N = 13$), which are easy to distinguish due to body and tooth size, or recapturing individuals that had known birth years. The known ages (years) of the 21 wolves included in the analyses were 0.8 ($N = 13$); 2.8 ($N = 1$); 3.7 ($N = 1$); 3.8 ($N = 3$); 4.8 ($N = 1$); 5.8 ($N = 1$); 6.8 ($N = 1$); and 8.8 ($N = 1$). For the five individuals included in the expression analyses that were not first identified as pups, tooth wear was used to estimate the age of adults (Gipson et al. 2000). The estimated ages (years) of the five wolves used in the analyses were 2.8 ($N = 1$); 4 ($N = 1$); 4.8 ($N = 1$); 6 ($N = 1$); and 9 ($N = 1$). Estimated ages of the two older wolves (6 and 9 years) were corroborated by the fact that they had been monitored by NPS for multiple consecutive years (5 and 6 years of monitoring, respectively).

Social Status

Social rankings of the 27 individuals were determined from behavioral observations and reproductive status and were categorized, specific to the time of sampling, as alpha or nonalpha (supplementary table S1, Supplementary Material online). Individuals that were the only or primary breeders in a pack or behaviorally dominant to all pack members (excluding their mates) were identified as alpha wolves. In well-known and larger packs, individuals that were subordinate to the leading pair but dominant to other pack members were ranked as beta. Individuals showing neither reproductive nor antagonistic behavior were classified in the subordinate category (Mech 1999; Packard 2003; Smith et al. 2012). Because age and social rank are highly correlated in YNP wolves (Pearson correlation $r = 0.614$; $P = 6.55 \times 10^{-5}$), we focused on the distinction between alpha wolves (scored as 1) and nonalpha animals (scored as 0; betas and subordinates), which significantly reduced the magnitude of this correlation ($N = 27$; $P < 0.01$; psych R package; paired.r function to test for a significant decrease in correlation) (Revelle 2015).

Mange Infection Status

The severity of mange infection in each wolf was assessed based on the presence of physical symptoms (hairlessness and scratching lesions) at the time of capture and scored according to the percentage of body surface affected (0: no physical symptoms; 1: $<$ 5% of the body affected by lesions; 2: 6–50%; and 3: $>$ 50%) (Pence et al. 1983; Almberg et al. 2012). As only one individual presented severe symptoms of mange infection, we classified mange as a binary variable (0: no symptom; 1: visible symptoms; supplementary table S1, Supplementary Material online). Continuous coding of mange severity led to qualitatively identical null findings (results not shown).

RNA Extraction, Library Preparation, and Sequencing

Whole blood was preserved in PAXgene Blood RNA tubes (PreAnalytiX, Qiagen, Hombrechtikon, Switzerland) and stored at −80 °C. Total RNA extraction was performed following the manufacturer's instructions of PAXgene Blood RNA kit (PreAnalytiX, Qiagen, Hombrechtikon, Switzerland). RNA quality was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). All samples had a RNA integrity number (RIN) $>$ 7. Total RNA was treated with the GlobinZero kit (Epicentre, Illumina, Madison, WI) and purified with a modified Qiagen RNeasy MinElute (Qiagen Inc., Valencia, CA) cleanup procedure or ethanol precipitation (supplementary table S1, Supplementary Material online). cDNA libraries were synthesized using a strand-specific kit from Epicentre

(ScriptSeq v2 Library Prep kit, Illumina, Madison, WI) with ScriptSeq Index PCR primers (Epicentre, Illumina, Madison, WI). cDNA libraries were quantified with the KAPA SYBR Fast qPCR library quantification kit (Kapa Biosystems Inc., Wilmington, MA) and pooled at five to six samples per lane. Single-end 100 bp sequencing was performed on Illumina's HiSeq2000 platform (supplementary table S1, Supplementary Material online) at the Broad Stem Cell Research Center's biosequencing core, UC Los Angeles, CA (BSCRC) and the Vincent J. Coates Genomics Sequencing Laboratory (GSL), UC Berkeley, CA. The data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO series accession number GSE80440 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80440).

Mapping and Expression Quantification

Quality filtering and adaptor trimming was performed using Trim Galore! v.0.3.1 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). A moderate coverage (25x) genome assembly of the gray wolf is available (Freedman et al. 2014). However, this assembly is based on mapping reads to the domestic dog. Therefore, given the availability of the well-annotated dog genome (*Canis l. familiaris*; Ensembl CanFam3.1.74) (Lindblad-Toh et al. 2005; Hoeppner et al. 2014), as well as the short divergence time between domestic dogs and wolves (< 29 Ka) (Thalmann et al. 2013; Freedman et al. 2014; Skoglund et al. 2015; Fan et al. 2016), we used the domestic dog genome as a reference. Splice-aware mapping to the domestic dog genome (Ensembl CanFam3.1.74, unmasked) was implemented with TopHat2 v.2.0.10 (Engström et al. 2013; Kim et al. 2013). We allowed up to three mismatches, a total of 3-bp length of gaps, and an edit distance no more than 3 bp between the reads and the reference sequence to account for the divergence between the domestic dog and the gray wolf. Only reads that uniquely mapped to the reference were kept in subsequent analyses. Only uniquely mapped reads were kept in subsequent analyses. The expression level of each gene was quantified using the union mode of the python script htseq-count (HTSeq 0.6.1) (Anders et al. 2015). Individual expression profiles were filtered to only include protein-coding genes with at least ten reads in at least 75% of cDNA libraries ($N = 13{,}558$ protein-coding genes). Genome assembly and GTF files of the domestic dog genome (CanFam3.1.74) used as reference in this study can be retrieved at http://dec2013.archive.ensembl.org/Canis_familiaris/Info/Index.

Data Preprocessing

To identify outlier samples, we used the adjacency function in the WGCNA R package (Langfelder and Horvath 2008) to build a Euclidean distance-based sample network from the log 2-transformed read counts. Samples were designated as outliers if their standardized connectivity deviated by more than three standard deviations from the mean (Horvath 2011). After removing two individual outliers from the data (supplementary table S1, Supplementary Material online), we used conditional quantile normalization (cqn package in R) to normalize for sequencing depth (using trimmed mean of *M* values) (Robinson et al. 2010), GC-content, and gene length (Hansen et al. 2012), averaged from the transcripts in the canFam3.1.74 gtf file. We used principal component analysis of the normalized, log 2-transformed expression data to assess general impacts of technical, phenotypic, or environmental variables on gene expression variance. We identified significant effects of technical factors on overall gene expression (supplementary fig. S1*A*, Supplementary Material online) including effects of sequencing core (correlated with PC1 [Pearson correlation $r = 0.59$, $P = 1.892 \times 10^{-3}$] and PC2 [Pearson correlation $r = -0.519$, $P = 7.809 \times 10^{-3}$] of the normalized gene expression data), library preparation procedure (column purifications vs. ethanol precipitation: correlated with PC1 [Pearson correlation $r = 0.791$, $P = 2.557 \times 10^{-6}$]), barcode bleed (Kircher et al. 2012) of RNA-Seq from other samples sequenced in the same lane (Johnston R, unpublished data, March 2015; correlated with PC2 [Pearson correlation $r = -0.566$, $P = 3.178 \times 10^{-3}$] and PC6 [Pearson correlation $r = -0.486$, $P = 1.387 \times 10^{-2}$]), and year of sampling (correlated with PC1 [Pearson correlation $r = 0.623$, $P = 8.86 \times 10^{-4}$]). To control for these effects, we regressed out sequencing core, library preparation procedure, and sequencing lane composition for each gene prior to the main analyses (supplementary fig. S1*B*, Supplementary Material online and supplementary table S1, Supplementary Material online). Year of sampling, which significantly correlated with library preparation procedure (Pearson correlation $r = 0.721$, $P = 4.853 \times 10^{-5}$), was not significantly associated with PC1-12 after regressing out the effect of library preparation (supplementary fig. S1*B*, Supplementary Material online). For the subset of samples in which time of day of collection was available (supplementary table S1, Supplementary Material online), we did not detect a significant effect of time of day on overall gene expression (supplementary fig. S1, Supplementary Material online).

Linear Mixed Effects Models

For each gene, we modeled the fixed effects of age, sex, social rank, and mange infection status on the residuals of the normalized, log 2-transformed gene expression levels of the 25 samples using linear mixed models in gemma v.0.94 (Zhou and Stephens 2012) as done previously (Tung et al. 2012). To fit a random effect controlling for kinship, we calculated pairwise relatedness between individuals using the triadic maximum likelihood approach in Coancestry (Wang 2011), based on genotypes obtained from 24 microsatellite loci (supplementary table S5, Supplementary Material online). Allele frequencies for the 24 markers were estimated using a comprehensive data set of 371 YNP gray wolves (vonHoldt et al. 2008, 2010; vonHoldt BM, unpublished data).

All significance levels for the linear models were adjusted for multiple hypothesis testing using the *Q* value method in R (Storey and Tibshirani 2003), with the null distribution constructed based on 100 random permutations. Because of the difficulty of obtaining samples from animals in the wild for RNA analyses, our sample size limited the statistical power of the linear model analysis. Therefore, we set a relatively liberal FDR threshold of 0.2

to identify candidate genes associated with rank, age, sex, and mange. The FDR threshold of 0.2 corresponded to a minimum detectably significant FD of approximately 2.2% for age and approximately 7.8% for sex (supplementary table S2, Supplementary Material online). Q values of all genes meeting the 0.2 FDR threshold are presented in supplementary table S2, Supplementary Material online.

GO Analysis

To assess which biological functions were significantly enriched in the genes affected by our variables of interest (defined as genes associated at Q value < 0.2), we used G:PROFILER (Reimand et al. 2007) based on the *Canis familiaris* gene annotation (Ensembl 79). Queries were composed of genes up- or down-regulated with each variable of interest and the background set included all genes tested in our analysis ($N = 13{,}558$). We set the minimal overlap between the GO term and the query, as well as the minimal number of genes within a functional category, to 3. All GO analyses were corrected for multiple testing using the Benjamini–Hochberg FDR method (Benjamini and Hochberg 1995).

Differential White Blood Cell Counts

To assess the impacts of age, rank, mange, and sex on the relative abundance of lymphocytes, neutrophils, monocytes, basophils, and eosinophils, blood smears were prepared for eleven of the 27 individuals during the collection of PAXgene-preserved blood (supplementary table S6, Supplementary Material online). Subsequently, the histological slides were fixed and stained according to standard procedures (i.e., fixation using the methanol and staining according to the Wright–Giemsa method). White blood cell (WBC) differentials were determined by quantifying the abundance of lymphocytes, neutrophils, monocytes, basophils, and eosinophils relative to a total number of white cells fixed at $N_{\text{total WBC}} = 100$. WBC counts were calculated as the average of two replicated slides. For each WBC type, counts were correlated in a linear model with each explanatory variable (age, sex, rank, or mange). No significant result was obtained. Therefore, we assessed effects of age on cell population with TRA (Powell et al. 2013), which is more sensitive to small differences in cell prevalence than blood smear cell counts because TRA assesses gene expression of multiple cell markers.

TOA and TRA

TOA was used to assess the extent to which differentially expressed genes were predominately characteristic of one or more subtypes of circulating leukocytes (Cole et al. 2011). Differential expression for TOA was defined by the FD, with differentially expressed genes defined by <0.90 and >1.10 FD per year of age. Thresholds of FD were selected to achieve the largest biological effect size possible while still yielding an input list with at least 50 differentially expressed genes. TRA was used to assess the prevalence of cell-type-specific RNAs (Powell et al. 2013) identified a priori based on highly cell-type-diagnostic transcripts in humans (Cole et al. 2011; Powell et al. 2013).

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to acknowledge Emily Almberg and Erin Stahler for their contribution to data collection. We are thankful to Devaughn Fraser for her significant help in the blood smear cell counts and to Ryan Harrigan for his comments on the manuscript. This work was supported by the National Science Foundation (grant numbers DEB 1257716 to R.K.W. and B.M.V. and DEB 1245373 to D.R.S. and D.W.S.); the National Institutes of Health (grant numbers P30 AG017265 to S.W.C., S10 RR029668 and S10 RR027303 to work with the Vincent J. Coates GSL at UC Berkeley); the Austrian Academy of Sciences (to P.C.); the Max Kade Foundation (to P.C.); the Howard Hughes Medical Institute (to R.A.J.); the National Park Service (to D.R.S. and D.W.S.); the Yellowstone Park Foundation (to D.R.S. and D.W.S.); the Tapeats Fund (to D.R.S. and D.W.S.); the Perkin-Prothro Foundation (to D.R.S. and D.W.S.); and an anonymous donor (to D.R.S. and D.W.S.).

References

Ahren B, Mansson S, Gingerich RL, Havel PJ. 1997. Regulation of plasma leptin in mice: influence of age, high-fat diet, and fasting. *Regul Integr Comp Physiol*. 273:R113–R120.

Alexander CM, Landsman PB, Teutsch SM, Haffner SM. 2003. NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. *Diabetes* 52:1210–1214.

Almberg ES, Cross PC, Dobson AP, Smith DW, Hudson PJ. 2012. Parasite invasion following host reintroduction: a case study of Yellowstone's wolves. *Philos Trans R Soc Lond Ser B Biol Sci*. 367:2840–2851.

Almberg ES, Cross PC, Dobson AP, Smith DW, Metz MC, Stahler DR, Hudson PJ. 2015. Social living mitigates the costs of a chronic illness in a cooperative carnivore. *Ecol Lett*. 18:660–667.

Almberg ES, Mech LD, Smith DW, Sheldon JW, Crabtree RL. 2009. A serological survey of infectious disease in Yellowstone National Park's canid community. *PLoS One* 4:e7042.

Anders S, Pyl PT, Huber W. 2015. HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.

Arlian LG, Morgan MS, Neal JS. 2004. Extracts of scabies mites (Sarcoptidae: *Sarcoptes scabiei*) modulate cytokine expression by human peripheral blood mononuclear cells and dendritic cells. *J Med Entomol*. 41:69–73.

Arlian LG, Morgan MS, Paul CC. 2006. Evidence that scabies mites (Acari: Sarcoptidae) influence production of interleukin-10 and the function of T-regulatory cells (Tr1) in humans. *J Med Entomol*. 43:283–287.

Baylis D, Bartlett D, Patel H, Roberts H. 2013. Understanding how we age: insights into inflammaging. *Longev Healthspan*. 2:1–8.

Beirne C, Delahay R, Hares M, Young A. 2014. Age-related declines and disease-associated variation in immune cell telomere length in a wild mammal. *PLoS One* 9:e108964.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 57:289–300.

Blankley S, Berry MPR, Graham CM, Bloom CI, Lipman M, O'Garra A. 2014. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos Trans R Soc Lond Ser B Biol Sci*. 369:20130427.

Boren E, Gershwin ME. 2004. Inflamm-aging: autoimmunity, and the immune-risk phenotype. *Autoimmun Rev*. 3:401–406.
Carrera-Bastos P, Fontes-Villalba M, O'Keefe JH, Lindeberg S, Cordain L. 2011. The western diet and lifestyle and diseases of civilization. *Res Rep Clin Cardiol*. 2:15–35.
Chen E, Miller GE, Kobor M, Cole SW. 2011. Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. *Mol Psychiatry*. 16:729–737.
Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, et al. 2005. Application of genome-wide expression analysis to human health and disease. *Proc Natl Acad Sci U S A*. 102:4801–4806.
Cole SW, Conti G, Arevalo JMG, Ruggiero AM, Heckman JJ, Suomi SJ. 2012. Transcriptional modulation of the developing immune system by early life social adversity. *Proc Natl Acad Sci U S A*. 109:20578–20583.
Cole SW, Hawkley LC, Arevalo JMG, Cacioppo JT. 2011. Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression in leukocytes. *Proc Natl Acad Sci U S A*. 108:3080–3085.
Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT. 2007. Social regulation of gene expression in human leukocytes. *Genome Biol*. 8:R189.
Cubaynes S, MacNulty DR, Stahler DR, Quimby KA, Smith DW, Coulson T. 2014. Density-dependent intraspecific aggression regulates survival in northern Yellowstone wolves (*Canis lupus*). *J Anim Ecol*. 83:1344–1356.
Di Mauro T, David G. 2009. Chromatin modifications: the driving force of senescence and aging? *Aging* 1:182–190.
dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc Lond Ser B Biol Sci*. 279:3491–3500.
Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 30:207–210.
Engström P, Steijger T, Sipos B, Grant G, Kahles A, Rätsch G, Goldman N, Hubbard T, Harrow J, Guigó R, et al. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods*. 10:1185–1191.
Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, Ramirez O, Pollinger J, Galaverni M, Ortega Del-Vecchyo D, et al. 2016. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res*. 26:163–173.
Farooqui AA. 2015. Effect of long term consumption of high calorie diet and calorie restriction on human health. High calorie diet and the human brain. Cham (Switzerland): Springer International Publishing. p. 1–28.
Freedman AH, Schweizer RM, Gronau I, Han E, Ortega-Del Vecchyo D, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, et al. 2014. Genome sequencing highlights genes under selection and the dynamic early history of dogs. *PLoS Genet*. 10:e1004016.
Garrett-Sinha L. 2013. Review of Ets1 structure, function, and roles in immunity. *Cell Mol Life Sci*. 70:3375–3390.
Gipson PS, Ballard WB, Nowak RM, Mech LD. 2000. Accuracy and precision of estimating age of gray wolves by tooth wear. *J Wildl Manage*. 64:752–758.
Göring HHH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JBM, Abraham LJ, Rainwater DL, Comuzzie AG, et al. 2007. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet*. 39:1208–1216.
Gregori S, Goudy KS, Roncarolo MG. 2012. The cellular and molecular mechanisms of immuno-suppression by human type 1 regulatory T cells. *Front Immunol*. 3:1–12.
Hansen KD, Irizarry RA, Wu Z. 2012. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13:204–216.
Hoeppner MP, Lundquist A, Pirun M, Meadows JRS, Zamani N, Johnson J, Sundström G, Cook A, FitzGerald MG, Swofford R, et al. 2014. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS One* 9:e91172.
Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, Even PC, Cervera P, Le Bouc Y. 2003. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature* 421:182–187.
Hong M-G, Myers AJ, Magnusson PKE, Prince JA. 2008. Transcriptome-wide assessment of human brain and lymphocyte senescence. *PLoS One* 3:e3024.
Hoopes B, Rimbault M, Liebers D, Ostrander E, Sutter N. 2012. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm Genome*. 23:780–790.
Horvath S. 2011. Integrated weighted correlation network analysis of mouse liver gene expression data. Weighted network analysis: applications in genomics and systems biology. New York: Springer Book. p. 421.
Ishioka K, Hosoya K, Kitagawa H, Shibata H, Honjoh T, Kimura K, Saito M. 2007. Plasma leptin concentration in dogs: effects of body condition score, age, gender and breeds. *Res Vet Sci*. 82:11–15.
Ishioka K, Soliman MM, Sagawa M, Nakadomo F, Shibata H, Honjoh T, Hashimoto A, Kitamura H, Kimura K, Saito M. 2002. Experimental and clinical studies on plasma leptin in obese dogs. *J Vet Med Sci*. 64:349–353.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36.
Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 40:e3.
Kirkwood TBL. 1977. Evolution of ageing. *Nature* 270:301–304.
Kirkwood TBL. 2005. Understanding the odd science of aging. *Cell* 120:437–447.
Kraus C, Pavard S, Promislow DEL. 2013. The size–life span trade-off decomposed: why large dogs die young. *Am Nat*. 181:492–505.
Krishnamurthy J, Torrice C, Ramsey MR, Kovalev GI, Al-Regaiey K, Su L, Sharpless NE. 2004. Ink4a/Arf expression is a biomarker of aging. *J Clin Invest*. 114:1299–1307.
Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1): 13.
Li H, Matheny M, Nicolson M, Türner N, Scarpace PJ. 1997. Leptin gene expression increases with age independent of increasing adiposity in rats. *Diabetes* 46:2035–2039.
Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DBKM, Clamp MCJ, Kulbokas EJ 3rd, Zody MCME, Xie XBM, Wayne RK, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–819.
Linton PJ, Dorshkind K. 2004. Age-related changes in lymphocyte development and function. *Nat Immunol*. 5:133–139.
Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Ibrahim JG, Thomas NE, Sharpless NE. 2009. Expression of p16INK4a in peripheral blood T-cells is a biomarker of human aging. *Aging Cell* 8:439–448.
López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The hallmarks of aging. *Cell* 153:1194–1217.
Ma XH, Muzumdar R, Yang XM, Gabriely I, Berger R, Barzilai N. 2002. Aging is associated with resistance to effects of leptin on fat distribution and insulin action. *J Gerontol A Biol Sci Med Sci*. 57:B225–B231.
MacNulty DR, Smith DW, Mech LD, Eberly LE. 2009. Body size and predatory performance in wolves: is bigger better? *J Anim Ecol*. 78:532–539.
MacNulty DR, Smith DW, Vucetich JA, Mech LD, Stahler DR, Packer C. 2009. Predatory senescence in ageing wolves. *Ecol Lett*. 12:1347–1356.
McGowan P, Sasaki A, D'Alessio AC, Dymov S, Labonté B, Szyf M, Turecki G, Meaney MJ. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci*. 12:342–348.
Mech LD. 1999. Alpha status, dominance, and division of labor in wolf packs. *Can J Zool*. 77:1196–1203.
Mech LD. 2003. Wolf social ecology. In: Mech LD, Boitani L, editors. Wolves—behavior, ecology, and conservation. Chicago (IL): The University of Chicago Press. p. 1–34.

Mejias A, Ramilo O. 2014. Transcriptional profiling in infectious diseases: ready for prime time? *J Infect*. 68:S94–S99.
Miller GE, Chen E, Fok AK, Walker HA, Lim A, Nicholls EF, Cole SW, Kobor M. 2009. Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increase proinflammatory signaling. *Proc Natl Acad Sci U S A*. 106:14716–14721.
Mobbs CV. 1998. Leptin and aging. In: Mobbs CV, Hof PR, editors. Functional endocrinology of aging. Basel (Switzerland): Karger. p. 228–240.
Molnar B, Fattebert J, Palme R, Ciucci P, Betschart B, Smith DW, Diehl P-A. 2015. Environmental and intrinsic correlates of stress in free-ranging wolves. *PLoS One* 10:e0137378.
Mostoslavsky R, Chua KF, Lombard DB, Pang WW, Fischer MR, Gellon L, Liu P, Mostoslavsky G, Franco S, Murphy MM, et al. 2006. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* 124:315–329.
Murphy MLM, Slavich GM, Rohleder N, Miller GE. 2013. Targeted rejection triggers differential pro- and anti-inflammatory gene expression in adolescents as a function of social status. *Clin Psychol Sci*. 1:30–40.
Nussey DH, Froy H, Lemaitre J-F, Gaillard J-M, Austad SN. 2013. Senescence in natural populations of animals: widespread evidence and its implications for bio-gerontology. *Ageing Res Rev*. 12:214–225.
Ohtani N, Zebedee Z, Huot TJG, Stinson JA, Sugimoto M, Ohashi Y, Sharrocks AD, Peters G, Hara E. 2001. Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. *Nature* 409:1067–1070.
Packard JM. 2003. Wolf behavior: reproductive, social and intelligent. In: Mech LD, Boitani L, editors. Wolves—behavior, ecology, and conservation. Chicago (IL) and London: The University of Chicago Press. p. 35–65.
Palacios MG, Winkler DW, Klasing KC, Hasselquist D, Vleck CM. 2011. Consequences of immune system aging in nature: a study of immunosenescence costs in free-living Tree Swallows. *Ecology* 92:952–966.
Pence DB, Windberg LA, Pence BC, Sprowls R. 1983. The epizootiology and pathology of sarcoptic mange in coyotes, *Canis latrans*, from South Texas. *J Parasitol*. 69:1100–1115.
Peters MJ, Joehanes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K, et al. 2015. The transcriptional landscape of age in human peripheral blood. *Nat Commun*. 6:8570.
Poncet D, Belleville A, t'Kint de Roodenbeke C, Roborel de Climens A, Ben Simon E, Merle-Beral H, Callet-Bauchu E, Salles G, Sabatier L, Delic J, et al. 2008. Changes in the expression of telomere maintenance genes suggest global telomere dysfunction in B-chronic lymphocytic leukemia. *Blood* 111:2388–2391.
Powell ND, Sloan EK, Bailey MT, Arevalo JMG, Miller GE, Chen E, Kobor MS, Reader BF, Sheridan JF, Cole SW. 2013. Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via beta-adrenergic induction of myelopoiesis. *Proc Natl Acad Sci U S A*. 110:16574–16579.
Rahman AA, Karim AN, Hamid ANA, Harun R, Wan Ngah WZ. 2013. Senescence-related changes in gene expression of peripheral blood mononuclear cells from octo/nonagenarians compared to their offspring. *Oxid Med Cell Longev*. 2013:14.
Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B, Banchereau J, Palucka AK, et al. 2007. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* 109:2066–2077.
Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. g: profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 35:W193–W200.
Revelle W. 2015. Psych: procedures for personality and psychological research. Evanston (IL): Northwestern University.
Ricklefs RE. 2010. Life-history connections to rates of aging in terrestrial vertebrates. *Proc Natl Acad Sci U S A*. 107:10314–10319.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
Rollo CD. 2002. Growth negatively impacts the life span of mammals. *Evol Dev*. 4:55–61.
Runcie DE, Wiedmann RT, Archie EA, Altmann J, Wray GA, Alberts SC, Tung J. 2013. Social environment influences the relationship between genotype and gene expression in wild baboons. *Philos Trans R Soc Lond Ser B Biol Sci*. 368:20120345.
Salminen A, Kaarniranta. K. 2010. Insulin/IGF-1 paradox of aging: regulation via AKT/IKK/NF-κB signaling. *Cell Signal*. 22:573–577.
Sánchez-Rodríguez M, García-Sánchez A Retana-Ugalde R., Mendoza-Núñez VcM 2000. Serum leptin levels and blood pressure in the overweight elderly. *Arch Med Res*. 31:425–428.
Sapolsky RM. 2004. Social status and health in humans and other animals. *Annu Rev Anthropol*. 33:393–418.
Sapolsky RM. 2005. The influence of social hierarchy on primate health. *Science* 308:648–652.
Schwarz F, Springer SA, Altheide TK, Varki NM, Gagneux P, Varki A. 2016. Human-specific derived alleles of CD33 and other genes protect against postreproductive cognitive decline. *Proc Natl Acad Sci U S A*. 113:74–79.
Skoglund P, Ersmark E, Palkopoulou E, Dalén L. 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr Biol*. 25:1515–1519.
Smith DW, Stahler DR, Albers E, McIntyre R, Metz M, Irving J, Raymond R, Anton C, Cassidy-Quimby K, Bowersock N. 2010. Yellowstone wolf project, Annual Report. Yellowstone National Park (WY): National Park Service.
Smith DW, Stahler DR, Stahler E, Metz M, Quimby K, McIntyre R, Ruhl C, Martin H, Kindermann R, Bowersock N, et al. 2012. Yellowstone wolf project, Annual report. Yellowstone National Park (WY): National Park Service.
Snyder-Mackler N, Somel M, Tung J. 2014. Shared signatures of social stress and aging in peripheral blood mononuclear cell gene expression profiles. *Aging Cell* 13:954–957.
Stahler DR, MacNulty DR, Wayne RK, vonHoldt B, Smith DW. 2013. The adaptive value of morphological, behavioural and life-history traits in reproductive female wolves. *J Anim Ecol*. 82:222–234.
Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 100:9440–9445.
Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, et al. 2007. A single *IGF1* allele is a major derterminant of small size in dogs. *Science* 316:112–115.
Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, Germonpré MB, Sablin MV, López-Giráldez F, Domingo-Roura X, et al. 2013. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342:871–874.
Tung J, Barreiro LB, Johnson ZP, Hansen KD, Michopoulos V, Toufexis D, Michelini K, Wilson ME, Gilad Y. 2012. Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc Natl Acad Sci U S A*. 109:6490–6495.
Tung J, Zhou X, Alberts SC, Stephens M, Gilad Y. 2015. The genetic architecture of gene expression levels in wild baboons. *eLife* 4: e04729.
Velavan T, Ojurongbe O. 2011. Regulatory T cells and parasites. *J Biomed Biotechnol*. 2011:8.
vonHoldt BM, Stahler DR, Smith DW, Earl DA, Pollinger JP, Wayne RK. 2008. The genealogy and genetic viablity of reintroduced Yellostone grey wolves. *Mol Ecol*. 17:252–274.
vonHoldt BM, Stahler DR, Bangs EE, Smith DW, Jimenez MD, Mack CM, Niemeyer CC, Pollinger JP, Wayne RK. 2010. A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. *Mol Ecol*. 19:4412–4427.
Wang J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol Ecol Resour*. 11:141–145.
Weaver ICG, Meaney MJ, Szyf M. 2006. Maternal care effects on hippocampal transcriptome and anxiety-mediated behaviors in the

offspring that are reversible in adulthood. *Proc Natl Acad Sci U S A.* 103:3480–3485.
Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. 2003. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A.* 100:1896–1901.
Wilson EO. 2000. Sociobiology: the new synthesis. Cambridge and London: The Belknap Press of Harvard University Press.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 44:821–824.

Supplemental Results

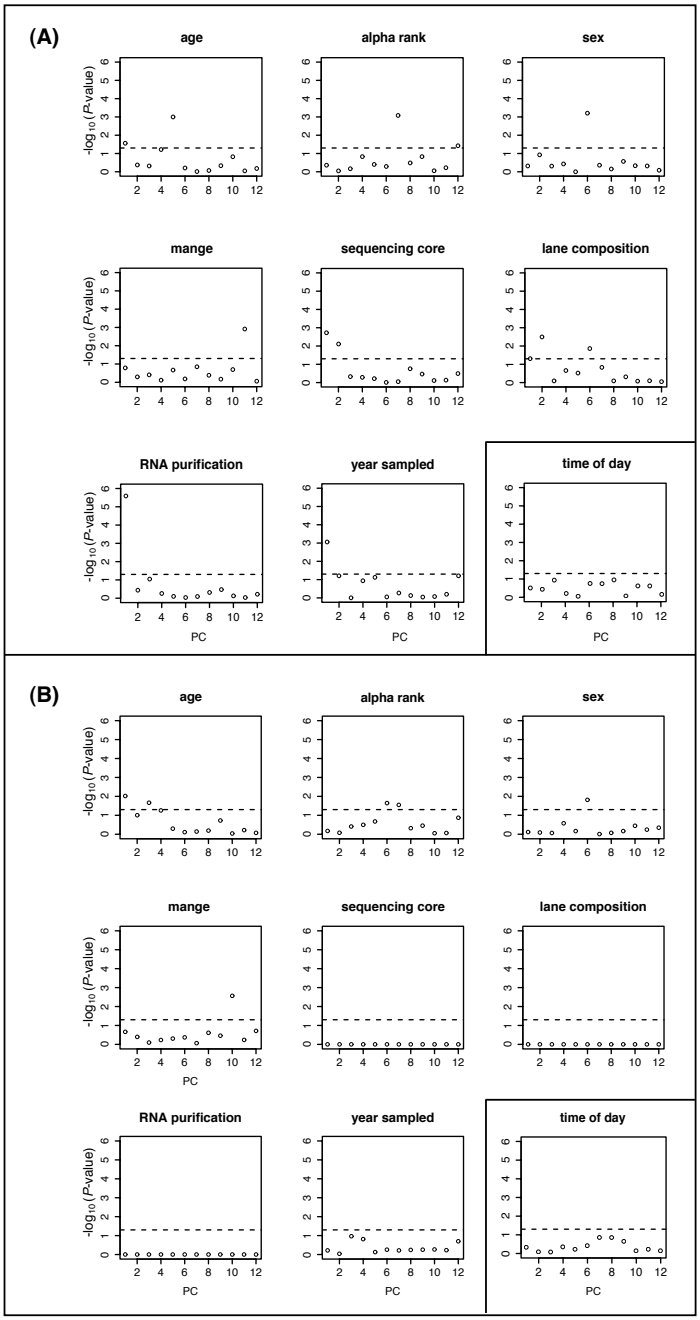

Figure 2-S1. Statistical significance of simple linear regression between each of the first twelve PCs of the expression data and variables that could potentially structure the data. Except for time of day, which represents a subset of 16 samples for which time of day of collection was available, all plots represent results from the dataset of 25 wolves. (A) Significance values before regression of any variables. (B) Significance values after regressing out sequencing core, RNA purification method, and lane composition.

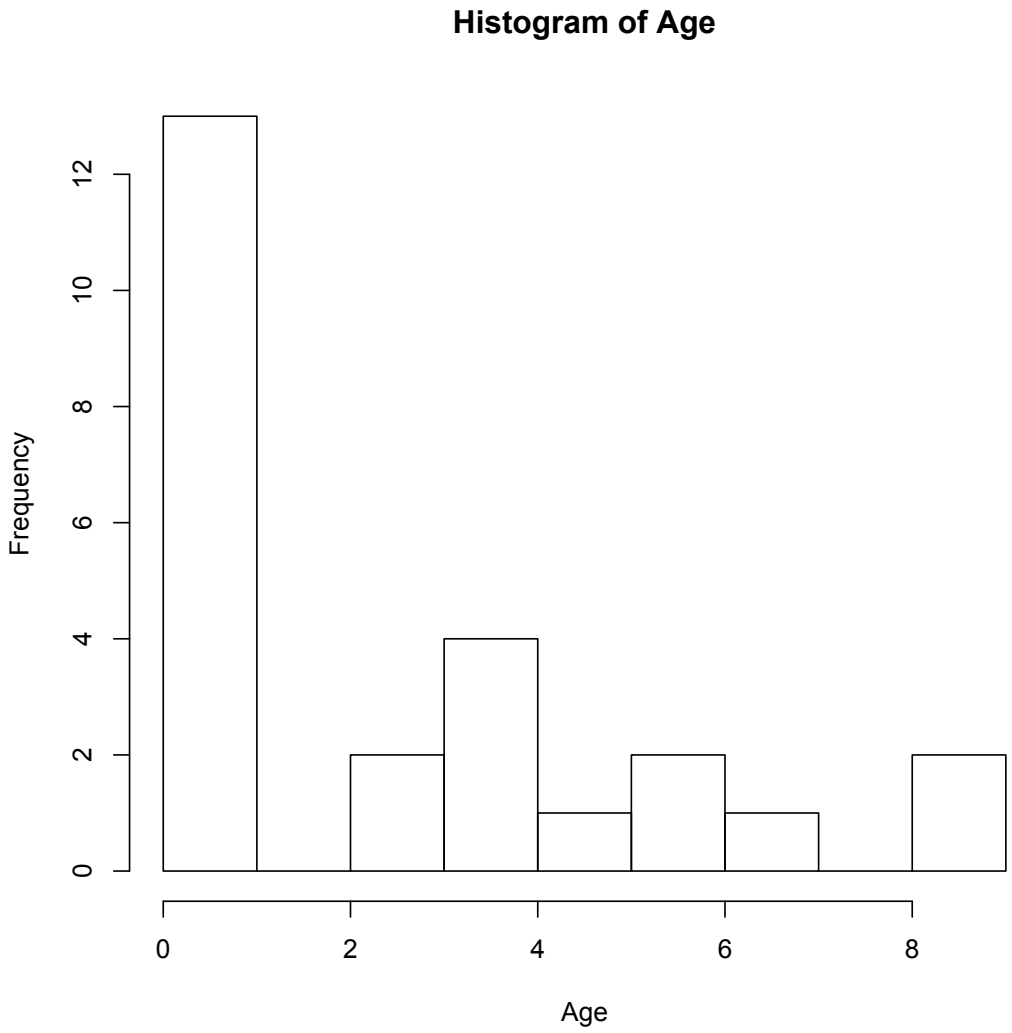

Figure 2-S2. Distribution of ages (years) for the 25 wolves included in the analyses.

Marker name	Chromosome	Repeat type	Reference
PEZ5	cfa12	Tetra-	J. Halverson in (Neff *et al.*, 1999)
PEZ8	cfa17	Tetra-	J. Halverson in (Neff *et al.*, 1999)
PEZ19	cfa20	Tetra-	J. Halverson in (Neff *et al.*, 1999)
FH2001	cfa23	Tetra-	Francisco et al. 1996
FH2004	cfa11	Tetra-	Francisco et al. 1996
FH2010	cfa24	Tetra-	Francisco et al. 1996
FH2054	cfa12	Tetra-	Francisco et al. 1996
FH2088	cfa15	Tetra-	Francisco et al. 1996
FH2137	cfa9	Tetra-	Francisco et al. 1996
FH2324	cfa25	Tetra-	(Mellersh *et al.*, 1997)
FH2611	cfa36	Tetra-	(Guyon *et al.*, 2003)
FH2658	cfa14	Tetra-	(Guyon *et al.*, 2003)
FH2670	cfa16	Tetra-	(Guyon *et al.*, 2003)
FH2766	cfa38	Di-	(Guyon *et al.*, 2003)
FH2785	cfa28	Di-	(Guyon *et al.*, 2003)
FH2790	cfa33	Di-	(Guyon *et al.*, 2003)
FH2869	cfa17	Di-	(Guyon *et al.*, 2003)
FH2914	cfa21	Di-	(Guyon *et al.*, 2003)
FH3047	cfa17	Di-	(Guyon *et al.*, 2003)
FH3398	cfa21	Tetra-	(Guyon *et al.*, 2003)
FH3399	cfa38	Tetra-	(Guyon *et al.*, 2003)
FH3725	cfa14	Tetra-	(Guyon *et al.*, 2003)
FH3853	cfa22	Tetra-	(Guyon *et al.*, 2003)
FH3965	cfa2	Tetra-	(Guyon *et al.*, 2003)

Table 2-S5. Microsatellite loci to control for relatedness. The twenty-four microsatellite loci, originally isolated in domestic dog and previously tested on the North American gray wolf ((vonHoldt *et al.*, 2008; vonHoldt *et al.*, 2010), that were used to control for relatedness among individuals using the triadic Maximum Likelihood (TrioML) approach implemented in COANCESTRY (Wang, 2011).

Animal ID	Eosinophil	Basophil	Neutrophil	Lymphocyte	Monocyte
471F	2	1.5	50.5	45	1
712M	1	1	89	6.5	2.5
sw763M	1	1	81.5	10.5	6
775M	1	0	80	18	1
777M	2	0	63	30	5
778M	1	1	87	8	3
779F	1	1	72	22	4
818F	1	0	87	9	3
819F	4.5	0.5	84.5	7	3.5
869M	4	1.5	71	20	3.5
870F	1.5	0.5	85	11.5	1.5

Table 2-S6. White blood cell counts of eleven Yellowstone National Park wolves.

Supplemental Results Bibliography

Francisco LV, Langston AA, Mellersh CS, Neal CL, Ostrander EA. 1996. A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. Mamm Genome. 7:359-362.

Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B et al. 2003. A 1-Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci USA. 100:5296-5301.

Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA. 1997. A Linkage Map of the Canine Genome. Genomics. 46:326-336.

Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegle JS, Ostrander EA, Rine J. 1999. A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics. 151:803-820.

vonHoldt BM, DR S, DW S, Earl DA, Pollinger JP, Wayne RK. 2008. The genealogy and genetic viablity of reintroduced Yellostone grey wolves. Mol Ecol. 17:252-274.

vonHoldt BM, Stahler DR, Bangs EE, Smith DW, Jimenez MD, Mack CM, Niemeyer CC, Pollinger JP, Wayne RK. 2010. A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. Mol Ecol. 19:4412-4427.

Wang J. 2011. COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol Ecol Resour. 11:141-145.

CHAPTER 3

POPULATION-BASED CELL ASSAYS FOR STUDYING GENE FUNCTION IN THE NORTH AMERICAN GRAY WOLF

Abstract

Cell culture methods are amenable to a diversity of approaches for testing the effects of genetic variation on phenotype. However, these methods have not been widely adopted in evolutionary biology, despite their potential usefulness for understanding functional effects of genetic variants in natural populations. In North American gray wolves, a segregating allele containing a codon deletion in the canine beta defensin 3 (*CBD103*) gene confers dominantly inherited black coat color by altering binding affinity of the beta defensin 3 protein to melanocortin 1 receptor (MC1R). This mutation has been associated with low annual survival in the homozygous state but with high survival in heterozygotes, independent of coat color. Given the known antiviral properties of the beta defensin 3 peptide, we hypothesized that the segregating *CBD103* deletion not only impacts coat color, but also impacts antiviral activity of the peptide or binding affinity to other cellular receptors. To evaluate the allele-specific function of *CBD103* in a tissue where the gene is highly expressed, we optimized keratinocyte culture methods for field conditions to establish a population-level panel of keratinocyte lines from wild gray wolves. We established keratinocyte cell lines from 24 wild wolves, representing 14 wildtype wolves and 10 wolves heterozygous for the *CBD103* mutation. Further, to experimentally test the effects of the mutation, we utilized CRISPR/Cas9 gene editing of a single wildtype cell line to construct cell lines heterozygous and homozygous for the mutation. To evaluate the cellular response to microbial infection, we infected keratinocytes with synthetic antigens as well as live canine distemper virus (CDV), a highly contagious pathogen in carnivores known to have had multiple outbreaks in North American gray wolves. Keratinocytes exhibited gene expression responses to immune challenge, but did not show significant expression differences with *CBD103* genotype, suggesting that the effect of *CBD103* genotype on survival is not mediated by impacts on gene

regulation within keratinocytes. This study describes methods for establishing keratinocytes from wild mammals and highlights the utility of *in vitro* methods for evaluating genetic variants in natural populations.

Introduction

Understanding functional effects of genetic variants in populations is a fundamental goal in evolutionary biology. However, studies that reveal causal connections between genetic variation and phenotype in wild populations are rare (Feder and Mitchell-Olds, 2003). Although high-throughput sequencing has allowed forward genetics methods to be applied to this endeavor (beginning with a phenotype of interest and identifying the genes underlying them) (Stinchcombe and Hoekstra, 2008), forward genetic approaches are limited in that they result only in associations between genotype and phenotype. In contrast, reverse genetic approaches, in which genetic manipulations are utilized, can causally connect a genetic variant to phenotype (Feder and Mitchell-Olds, 2003). However, the resources required to perform genetic manipulations on individuals, as well as logistical and ethical considerations, make this method infeasible for many non-model species. One way to circumvent this challenge is to utilize transgenic model systems to study organismal-level effects of genetic variants (e.g. transgenic mice) (Abzhanov *et al.*, 2008; Bedford and Hoekstra, 2015). However, divergence between the species of interest and model species can constrain this possibility or limit functional interpretation of results (Abzhanov *et al.*, 2008). Another approach is to decompose the candidate organismal-level trait into lower levels (e.g., cellular level), and experimentally test effects of genetic variants. This can be a potentially robust approach, as it permits utilization of the expansive toolkit for *in vitro* methods widely used in model systems, such as genetic manipulations (e.g., CRISPR/Cas9) (Chen *et al.*, 2014) as well as experimental treatments (e.g., immunostimulants). Establishment of cell lines from wild populations may thus be highly amenable to functional genomic studies of ecologically relevant traits in diverse species.

The functional traits that can be studied using cell lines are limited to traits that can be functionally decomposed to the cellular level, and to tissue types available to investigators. Further, because cells have to be kept viable from the moment of collection, not all cell-types will be amenable to collection under field conditions. The immune system is a fundamental determinant of animal survival and fitness (Lochmiller and Deerenberg, 2000), and immune-related traits can be parsed to the cellular level. Further, tissues comprised of immune-related cells, namely blood and skin, are often readily available when working with non-model species. Blood contains the cell types central to the immune system and is commonly collected from animals in the field. However, many blood cell-types have short lifespans and gene expression profiles that are highly sensitive to storage conditions (Tanner *et al.*, 2002), requiring rapid processing of samples that may not be feasible for non-model systems. In contrast, fibroblasts, the predominant cells in connective tissue, are amenable to sampling under field conditions (e.g., can be refrigerated for a week before processing) (Tovar *et al.*, 2008), and methods to establish primary fibroblast lines from the dermal layer of skin biopsies of hundreds of non-model species have long been established (Benirschke, 1984). Notably, in addition to being an important cell in tissue repair, fibroblasts can be used to establish induced pluripotent stem cells (Takahashi *et al.*, 2007; Ben-Nun *et al.*, 2011; Ramaswamy *et al.*, 2015). Keratinocytes can also be cultured from skin, and, as the principal cell type of the epidermis, form a critical barrier between an organism and the environment, imparting a first line of defense against pathogens (Lebre *et al.*, 2007). In addition to forming a physical barrier, keratinocytes produce antimicrobial peptides such as defensins (Wilson *et al.*, 2013) to directly kill microbes, as well as express multiple cytokines and chemokines in response to infection to recruit leukocytes to areas of invasion (Lebre *et al.*, 2007). These functions of keratinocytes may be critical in containing infection of pathogens

known to have devastating impacts on natural animal populations, such as white-nose syndrome in bats (Field *et al.*, 2015) and chytrid fungus in amphibians (Rosenblum *et al.*, 2012). Thus, primary culture of keratinocytes promises to serve as a particularly useful approach for functionally studying immune responses of wild populations to infectious pathogens.

The objectives of the study were to develop methods to establish a population-level panel of primary keratinocyte cell lines of the North American gray wolf (*Canis lupus*)and to use this system to evaluate the allele-specific functions of Canine β-defensin 3 (encoded by *CBD103*, also called K locus; homologous to human *DEFB103*) (Candille *et al.*, 2007), an antimicrobial peptide involved in the immune system (Leonard *et al.*, 2012). North American gray wolves are prone to a diversity of infectious diseases including canine distemper virus (CDV), canine parvovirus, and canine adenovirus-1 (Almberg *et al.*, 2009). Within North American gray wolves, *CBD103* is polymorphic, with an estimated 38% of individuals carrying a three base-pair (bp) in-frame deletion (called the K^B allele) that increases binding affinity of the CBD103 protein to melanocortin 1 receptor (encoded by Mc1r), conferring black coat color (Anderson *et al.*, 2009; Coulson *et al.*, 2011). Notably, homozygotes for the K^B allele (K^{BB}) have dramatically lower mean lifetime reproductive success and annual survival rate than heterozygotes (K^{By}), despite both mutant genotypes conferring black coat color (Coulson *et al.*, 2011). Thus, the difference in annual survival between homozygotes and heterozygotes is likely independent of coat coloration. Individuals with the K^{By} genotype also have higher fitness than wildtype individuals (K^{yy}), explaining the preservation of the derived allele in natural populations (Coulson *et al.*, 2011; Stahler *et al.*, 2013). In humans, HBD3 (the protein product of *DEFB103*, the homolog to *CBD103*) is known to have direct antibacterial and antiviral properties (reviewed in (Wilson *et al.*, 2013); (García *et al.*, 2001; Harder *et al.*, 2001). For example, HBD3 inhibits

herpes simplex virus (HSV) infection by deterring binding and entry of HSV (Hazrati *et al.*, 2006), and inhibits human immunodeficiency virus (Shively and Clarkson) replication via direct binding with virions (Quiñones-Mateu *et al.*, 2003). Therefore, we hypothesized that the protein encoded by *CBD103* also has antiviral roles, and that the antiviral functions of the protein, or binding affinity of the protein to other receptors, is altered by the three bp deletion.

To test this hypothesis, we established keratinocyte cell lines from 24 wild wolves, representing 14 wolves wildtype for *CBD103* and 10 wolves heterozygous for the deletion allele. Additionally, from a single wildtype wolf, we utilized CRISPR/Cas9 gene editing to produce one keratinocyte line heterozygous, and one line homozygous, for the deletion allele. To evaluate the utility of keratinocytes for studying the innate immune system, we used RNA-Seq data to compare expression of immune-related genes in wolf primary keratinocytes, primary fibroblasts, and whole blood. We then used rt-qPCR to assess responsiveness of wolf keratinocytes to different antigens and RNA-Seq to evaluate the response to polyinosinic:polycytidylic acid (poly:IC), a synthetic antigen simulating dsRNA virus, to assess the immune response to simulated viral infection. Finally, to evaluate the response of keratinocytes to viral infection in a more ecologically relevant context, we assessed wolf keratinocyte responses to infection by live canine distemper virus (CDV; belongs to the *Morbillivirus* genus, which also includes measles). Canine distemper is a highly contagious disease in carnivores worldwide, with susceptibility in both marine and terrestrial species and with fatality rates in domestic dogs second only to rabies (Appel and Summers, 1995; Barrett, 1999; Deem *et al.*, 2000), and with known outbreaks in North American gray wolves which have been associated with substantial mortality in wolf-pups (Almberg *et al.*, 2009).

Materials and Methods

Tissue collection

Because the date of availability of animals for sample collection was unpredictable, we pre-prepared medium nutrients in single sample aliquots, called "MEFcubes", which were stored frozen until use (comprised of 1 ml bovine calf serum [BCS; Hyclone, Utah, USA], 0.1 ml L-glutamine [Gibco, Thermo Fisher Scientific, Massachusetts, USA], 20 ul Primocin antibiotic [Invivogen, California, USA], and 252 ul 1M HEPES buffer [Gibco]; MEFcubes were stored frozen up to 1 year). We also pre-prepared and refrigerated aliquots of 9 ml Dulbecco's Modified Eagle's Medium [DMEM; Gibco] in 15 ml conical tubes (up to 1 year refrigeration). 0 - 3 days prior to tissue collection, medium was prepared for each sample by thawing a "MEFcube" and transferring it to a 15 ml tube containing 9 ml DMEM. Each animal was anaesthetized and a 6 mm skin biopsy was collected following US national guidelines by the National Park Service and by the Idaho Department of Fish and Game. Upon collection, each 6 mm skin biopsy was immediately transferred to a 15 ml tube of collection medium and kept in a chilled cooler during the day of fieldwork (Tovar *et al.*, 2008). To reduce space of required reagents in the field, a subset of biopsies were collected into 1 ml aliquots of collection medium (in 1.7 ml eppendorf tubes) during field work and transferred to 15 ml tubes containing 10 ml collection medium at the end of the field day. Samples were transferred to a refrigerator until shipment. Samples were wrapped in paper towels to prevent direct contact with ice and overnight shipped to UCLA with ice.

Keratinocyte and fibroblast culture

Working in a sterile environment (i.e., biosafety class II cabinet), each skin biopsy was transferred to a p100 Petri dish with 50 ul collection medium. The biopsy was then minced into

ca. 0.5 mm pieces using curved iris scissors and evenly dispersed across the dish. A small drop of medium (*ca*. 20 ul) was placed around each skin piece, and 2-3 ml media was added to the rim of the dish. Samples were then incubated at 37° C under 5% CO_2 and atmospheric oxygen levels. Medium was changed every 2-3 days.

To optimize keratinocyte cell growth of initial plating, four biopsies were each split into two p60 Petri dishes. Two biopsies were fed FAD medium (1:1 DMEM:F12 base media [Gibco] + 5% BCS + 0.4 ug/ml hydrocortisone + 10 ng/ml epidermal growth factor (EGF) + 1% Penicillin-Streptomycin [Gibco]), and two biopsies were fed M199/M106 medium (1:1 M199:M106 [Gibco] + 15% BCS + 10 ng/ml EGF + 0.4 ug/ml hydrocortisone + 1% penicillin-streptomycin). After two days incubation, 10 uM ROCK inhibitor Y-27632 (hereafter referred to as RI; Cayman Chemical, Ann Arbor, MI) was added to one p60 of each of the paired samples. We then quantified the number of skin pieces with keratinocyte colonies after 4 days of incubation. All images were taken using an AxioCamMRm camera (Zeiss).

To separate fibroblasts from keratinocytes after initial plating (*ca*. 1 week post-plating), dishes were first washed once with phosphate buffer solution (PBS; Gibco). 1 ml 0.25% trypsin-EDTA (Gibco) was then added to each well and quickly pipetted up and down 3-5 times, which detached fibroblasts but not keratinocytes. The trypsin (containing fibroblasts) was then transferred to a 15 ml tube containing 1 ml DMEM + 10% BCS. The dish was then quickly washed two times with PBS to remove residual fibroblasts and incubated with 1 ml 0.25% trypsin at 37° C for 5-10 minutes to detach keratinocytes for cell passage. After this initial separation, selective growth of each cell type was performed using cell type-specific medium and through selective aspiration.

Fibroblasts were grown on M199/M106 medium (1:1 M199:M106 + 15% BCS + 10 ng/ml EGF + 0.4 ug/ml hydrocortisone + 1% Penicillin-Streptomycin). To optimize medium for keratinocyte growth beyond initial plating, we compared keratinocyte growth rate of two cell lines (wildtype 16361 and heterozygote 16366) on three formulas of medium: KSFM (Keratinocyte serum-free medium [Gibco] + 25 μg/ml bovine pituitary extract [BPE; Gibco] + 0.4 mM $CaCl_2$ + 0.2 ng/ml EGF + 1% Penicillin-Streptomycin), FAD medium, and M199/M106 medium (same medium formulas as described for initial plating). We also tested the effect of including 10 uM RI in the medium and the effect of plating keratinocytes on a feeder layer of 3T3 cells, which is a mouse embryonic fibroblast line (Todaro and Green, 1963) found to improve growth of human epidermal keratinocytes (Rheinwald and Green, 1975). This resulted in 12 growth treatments. To quantify cell growth rate, we plated keratinocytes at 5,000 cells per well of a 6-well plate for each treatment, and quantified the final number of cells after five days of incubation.

Cells were cryopreserved in DMEM/F12 + 10% BCS + 10% dimethyl sulfoxide (DMSO; Sigma) by transferring tubes to a room temperature Mr. Frosty Freezing Container (Thermo Scientific) which was placed in a -80° C freezer overnight. Tubes were then transferred to a liquid nitrogen freezer for storage.

Generation of K^{By} and K^{BB} keratinocyte lines using CRISPR/Cas9

Immortalization of wildtype cells and generation of K^{By} and K^{BB} cells were performed by Applied Stem Cell, Inc. Keratinocytes from a single wildtype (K^{yy}) individual (animal 15071; Table S1) were first immortalized using lentivirus expressing simian virus 40 (SV40) large T antigen. To generate cells carrying the 3 bp mutation in *CBD103*, the following guide RNA (gRNA) sequence was selected based on proximity to the target 3 bp deletion:

CCTGCAGAGGTATTATTGCAGA. To prevent the gRNA/Cas9 complex from recognizing and cutting sequence after the single-stranded donor oligonucleotide (ssODN) was used as repair template, a silent single nucleotide polymorphism (SNP), corresponding to the first bp of the gRNA, was incorporated into the ssODN (C to T). To genotype clones for the 3 bp mutation, the following primers were developed to amplify a 298 bp region containing the 3 bp *CBD103* deletion and two TspE-1 digestion sites: forward 5'-GTGAGGTGTACAATGAGGATTATAACTGAACTCC, reverse 5'-GGAAGAACAGCGGCCTATCTGC. TspE-1 digestion of wildtype PCR product yielded three fragments of lengths 112 bp, 103 bp, and 83 bp, whereas TspE-1 digestion of PCR product containing the 3 bp mutation yielded two fragments of lengths 112 bp and 183 bp. Tentative clones carrying the 3 bp deletion were confirmed with Sanger sequencing.

Cell challenges with synthetic antigens

Two keratinocyte lines were used to evaluate the response of cells to a panel of antigens (from animals 15071 and 890M; Table S1). Cells were plated at $1.5x10^4$ cells per well of a 12-well plate on a 3T3 feeder layer with FAD + 10 uM RI. When cells were at *ca.* 30% confluence, they were switched to a low-serum medium (FAD + 0.5% BCS) without RI and incubated for 24 hours. Cells were then treated with each of the following antigens for 24 hours: 40 ug/ml lipopolysaccharide (LPS; Sigma-Aldrich), 20 ug/ml Gardiquimod (Invivogen), 1 ug/ml poly:IC (Sigma-Aldrich), 400 ng/ml human IFN-gamma (PeproTech), and 40 ug/ml LPS + 400 ng/ml IFN-gamma. Non-challenged control wells were collected at time 0. Wells were quickly washed once with 0.25% trypsin (by pipetting up and down 3 times) and twice with PBS to remove 3T3 cells before collecting keratinocytes into 1 ml Trizol (Invitrogen, California, USA).

To more comprehensively assess the keratinocyte response to virus, we challenged all 24 wolf keratinocyte lines with 1 ug/ml poly:IC. Matched control (non-challenged) and challenged cells were collected into Trizol after 24 hours. Details of the samples used for poly:IC challenge across 24 keratinocyte lines are provided in Table 3-S1.

Cell challenges with live CDV

CDV experiments were performed with CDV 5804PEH-eGFP, a recombinant wild-type CDV that expresses green fluorescent protein (GFP), which was kindly provided by Dr. Veronika von Messling (Von Messling *et al.*, 2004). CDV was propagated in VerodogSLAMtag cells (Von Messling *et al.*, 2004), and $TCID_{50}$/cell was calculated following the Spearman-Kärber method (Finney, 1952) on VerodogSLAMtag cells. Keratinocytes were plated in duplicate on 24-well plates at $8x10^4$ cells/well in FAD + 5% BCS + RI, without feeder cells. After 24 hours incubation, the medium was changed to FAD + 0.5% BCS. After an additional 24 hours, cells were infected at an MOI of 20 $TCID_{50}$/ml or 100 $TCID_{50}$/ml. At various time points (0 – 5 days post-infection), cells were collected in 1 ml Trizol for RNA preservation.

RNA extraction and processing

Total RNA was extracted from keratinocytes using the Trizol Plus RNA Purification Kit with DNAse treatment using the Ambion RNAse-Free DNAse Set and column cleanup with the PureLink RNA Mini Kit (Invitrogen, Carlsbad, CA USA). For quantitative RT-PCR (rt-qPCR), complimentary DNA (cDNA) was synthesized using SuperScript II Reverse Transcriptase (ThermoFisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. The cDNA products were then purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). Rt-qPCR reactions were performed on cDNA representing 20 ng total RNA (based on nanodrop quantification of RNA) using LightCycler 480 SYBR Green Master (Roche,

Switzerland) with 10 ul reaction volumes. The following primer sets (developed by Real Time Primers, Elkins Park, PA, USA) were used: ACTB: forward 5'-ATGCAGAAGGAGATCACTGC, reverse 5'-CTGCGCAAGTTAGGTTTTGT; CXCL10: forward 5'-AGATGATTCCTGCAAGTCCA, reverse 5'-CCCCACTCTTTTTCATTGTG; CCL5: forward 5'-GCAGTCAGGAAGGAGATCAA, reverse 5'-GCAGCGAGAATTTTAATGGA. The following program was used on the LightCycler 480 Instrument for all rt-qPCR reactions: 95°C for 5 min; 45 cycles of 95°C (10 sec); 58°C (45 sec); subsequent melting curve that ramps the temperature from 72°C to 95°C in increments of 0.1°C/sec. rt-qPCR values were standardized to the expression level of ACTB. For RNA-Seq, quality of RNA from all 48 samples were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The average RNA integrity number (RIN) across the 48 samples was 9.83, with the lowest RIN of 7.5 (Table S1). cDNA libraries were generated with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions but with half the volume of all reagents. cDNA libraries were indexed and pooled at 12 samples per lane and sequenced as 100 basepair single-end reads on the Illumina HiSeq 4000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. Technical details of the 48 samples are provided in Table 3-S1.

RNA-Seq gene expression quantification

To compare expression levels of toll receptors and beta defensins across tissues, we quantified transcriptome-wide gene expression of keratinocytes and compared these values to previously published gene expression data of blood and fibroblasts of gray wolves (Charruau *et al.*, 2016). For keratinocyte RNA-Seq data, reads were trimmed with Trim Galore (Krueger, 2015) to remove adapters (read ends with 1 or more basepairs matching adapter sequence) and

basepairs with Phred score < 20 at the ends of reads. We used Tophat2 (Kim *et al.*, 2013) to map reads to the dog genome available from ensembl (*Canis familiaris*, version 3.1, downloaded September 24, 2016). The following parameters were used for mapping reads: read-mismatches 12, read-gap-length 3, read-edit-dist 12, b2-very-sensitive, b2-N 1. Only reads that mapped uniquely were kept for downstream analyses. HT-Seq (Anders *et al.*, 2014) with the "union" mode was used to quantify gene expression levels. After quantifying read counts, only protein-coding genes which exhibited at least 100 read counts in at least 25% of samples were included. Sample outliers were detected using the $Z.k_u$ score (Horvath, 2011) in the WGCNA (Langfelder and Horvath, 2008) R package. Control (non-challenged) sample 14387 was identified as an outlier because its $Z.k_u$ score was greater than the cutoff of 3, and therefore removed from further analyses.

RNA-Seq Analyses

RNA batch group (correlated with principal components [PC] 2, 5, and 6 of the overall gene expression data) and cDNA yield of cDNA library preparations (correlated with PC 2 of the overall expression data) were regressed as variables from the gene expression data. We used the package EMMREML (Akdemir and Godfrey, 2014) to model the resulting residuals of the gene expression data. Individual, as well as pairwise relatedness between individuals, were included as random effects. Relatedness was estimated based on previously described microsatellites (Charruau *et al.*, 2016) using the Lynch and Ritland estimator (Lynch and Ritland, 1999) in GenAlEx (Peakall and Smouse, 2006). *CBD103* genotype and poly:IC treatment were included as fixed effects. We calculated false discovery rates using R scripts kindly provided by Dr. Noah Snyder-Mackler (Snyder-Mackler *et al.*, 2016), which constructs the empirical null distribution

of p-values based on 100 permutations of each variable of interest (here, *CBD103* genotype and poly:IC treatment).

Results

Primary culture of keratinocytes

We first tested different media during initial plating and during subsequent passages to optimize conditions for growth of wolf keratinocytes. Skin biopsies that were plated on FAD medium and switched to FAD + RI two days post-plating showed increased numbers of keratinocyte colonies (Figure 3-S1), more keratinocytes per colony (Figure 3-S2), and healthier keratinocytes (as visualized by colony size and cell morphology; Figure 3-S2) compared to skin biopsies plated on FAD medium alone, M199/M106 medium, or M199/M106 + RI. Upon subsequent passages, keratinocyte health and proliferation rates were greatest when grown in FAD + RI with a feeder layer compared to all other conditions tested (Figures 3-S3 and 3-S4). We therefore used FAD + RI with a feeder layer for culturing keratinocytes from all subsequent samples, resulting in establishment of primary keratinocytes from 24 gray wolves (Table 3-S1). Keratinocyte lifespan (in FAD with a feeder layer) was estimated from a single wildtype wolf (animal 15071), which produced 27.8 cell population doublings over the course of 57 days (Figure 3-S5).

Gene expression of toll receptors and beta defensins

To infer the stimuli that keratinocytes can potentially respond to, we evaluated expression of toll receptors in wolf keratinocytes relative to expression in whole blood and fibroblasts (Figure 3-1A). Keratinocytes exhibited expression of *toll-receptors 1*, *2*, *3*, *5*, *6*, *7*, and *9*, but not *8* or *10* (Figure 3-1A). Expression levels of all toll receptors were lower in keratinocytes than in blood except for *TLR3*, which recognizes viral double-stranded RNA, and *TLR5*, which recognizes bacterial flegellin (reviewed in (Gay and Gangloff, 2007)).

We also evaluated gene expression of beta defensins, which are antimicrobial peptides involved in the innate immune system (Selsted and Ouellette, 2005). All beta defensins evaluated (*CBD1*, *CBD102*, *CBD103*, *CBD105*, *CBD111*, *CBD116*, *CBD118*, *CBD121*, *CBD122*, *CBD123*, *CBD124*, *CBD125*, *CBD127*, *CBD128*, *CBD129*) exhibited low expression levels in blood and non-detectable to low levels of expression in keratinocytes and fibroblasts, except for one gene, *CBD103* (Figure 3-1B), which exhibited high expression levels in keratinocytes.

Response of keratinocytes to immune challenge

To assess the response of wolf keratinocytes to different antigens, we measured keratinocyte expression levels of two chemokines, *CXCL10* and *CCL5*, upon challenge with each of five antigen treatments for 24 hours: poly:IC (TLR3 agonist that mimics viral dsRNA), Gardiquimod (agonist of TLR7, which recognizes viral ssRNA), lipopolysaccharide (LPS; TLR4 agonist), human IFNϒ (immunostimulatory cytokine), and LPS + human IFNϒ. Keratinocytes exhibited strong expression responses of *CXCL10* and *CCL5* to poly:IC, but not to the other treatments tested (Figure 3-2).

Transcriptome-wide gene expression across CBD103 genotypes

Given potential unknown effects of the K^B allele on binding affinity or antiviral activity of the CBD103 protein, we performed transcriptome-wide analyses of wildtype (K^{yy}) and naturally heterozygous (K^{By}) keratinocytes under baseline conditions and in response to simulated viral infection with poly:IC. Transcriptome-wide gene expression analyses revealed extensive expression responses to challenge with poly:IC, with poly:IC treatment significantly correlated with the first PC of overall expression variation across samples ($p < 10^{-10}$; Figure 3-S6). A total of 3,314 genes exhibited significant up-regulation, and 3,608 genes exhibited significant down-regulation, in response to poly:IC (FDR < 0.05; Table 3-S2). We did not detect

significant gene expression differences with *CBD103* genotype in non-challenged or poly:IC challenged keratinocytes.

Response of keratinocytes to CDV

To assess genotype specific responses to a virus existing naturally in wolves, we evaluated the susceptibility and response of keratinocytes to infection with the GFP expressing wild-type 5804PeH strain of live canine distemper virus (CDV) (Von Messling *et al.*, 2004). Both wildtype (K^{yy}) and natural heterozygous (K^{By}) keratinocytes, as well as the CRISPR/Cas9 generated heterozygous and homozygous (K^{BB}) cell lines, exhibited susceptibility to CDV infection, as visually evident by GFP expression in cells 3 – 5 days post-infection (Figures 3-3 and Figure 3-S7). Gene expression analysis of a wildtype cell line (15071) over the course of five days post-infection indicated increased expression of chemokines *CXCL10* and *CCL5* in response to CDV, with *CXCL10* exhibiting peak gene expression levels four days post-infection and *CCL5* exhibiting peak expression two and three days post-infection (Figure 3-4 A,B). We then compared gene expression of *CXCL10* and *CCL5* between three wildtype and three heterozygous keratinocyte lines (Figure 3-4 C,D) at five days post-infection. We did not detect a significant difference in gene expression between genotypes (Welch's t-test; CXCL10: mean K^{yy}: 1.64, mean K^{yB}: 2.61, $N = 6$, $df = 3.99$, $P = 0.562$; *CCL5*: mean K^{yy}: 3.26, mean K^{yB}: 2.71, $N = 6$, $df = 2.97$, $P = 0.618$).

Discussion

We established primary keratinocyte culture from 24 wild wolves, supporting the feasibility to utilize population-level keratinocyte lines for functional studies of non-model species. We detected expression of toll-receptors 1, 2, 3, 5, 6, 7, and 9 in wolf keratinocytes, suggesting that this cell-type is amenable to experimental challenge with diverse antigens. Consistent with this finding, we detected increased expression of *CXCL10* and *CCL5* in response to challenge with poly:IC and with live CDV. Further, keratinocytes heterozygous and homozygous for the *CBD103* deletion were successfully generated using CRISPR/Cas9 technology, exemplifying the potential of *in vitro* reverse genetic techniques to experimentally test gene function in cells of non-model species.

Given that the K^B allele is thought to alter binding affinity of CBD103 to MC1R (Candille *et al.*, 2007), we hypothesized that the K^B allele may alter the binding affinity of CBD103 to other receptors, potentially resulting in gene expression changes in other signaling pathways. However, transcriptome-wide analyses of non-challenged K^{yy} and K^{By} keratinocytes did not reveal significant gene expression differences with *CBD103* genotype, suggesting that the K^B allele does not lead to strong signaling changes within keratinocytes. Likewise, keratinocytes did not exhibit detectable expression differences between *CBD103* genotype when challenged with poly:IC or CDV. These results suggest that *CBD103* genotype does not impact intracellular signaling of keratinocytes during immune challenge. However, there are other mechanisms by which *CBD103* genotype may impact immunity. First, given the direct interactions of beta defensin 3 with bacteria and virus (García *et al.*, 2001; Harder *et al.*, 2001; Wilson *et al.*, 2013), the codon deletion in the K^B allele may reduce the binding affinity of the protein to bacteria or virus, ultimately leading to reduced numbers of viable bacteria or virus

entering cells. This could be tested by measuring bacterial or viral titers of infected K^{yy}, K^{By}, and K^{BB} cell lines through time. Additionally, the antiviral activity of beta defensin 3 may only be evident, or may be more pronounced, in medium with low salt concentrations more similar to the environment of the oral cavity, as has been observed for human beta defensin 3 (Quiñones-Mateu *et al.*, 2003). Such low salt concentrations were not feasible for our experiments (as these conditions would be lethal to keratinocytes). However, this could potentially be addressed using epithelial cells established from the oral cavity or by incubating the virus with beta defensin 3 prior to cell infection (as done by Quiñones-Mateu *et al*., 2003). Finally, the effect of CBD103 genotype on fitness may be mediated by the response to pathogens not evaluated in this study, such as canine parvovirus or canine adenovirus-1 (Almberg *et al.*, 2009).

Primary culture of keratinocytes can potentially be applied to diverse species, as the methods we used to culture primary wolf keratinocytes were originally developed for humans (Rheinwald and Green, 1975; Rheinwald, 1989), and were also applied to establish primary keratinocytes of a zebra finch (*Taeniopygia guttata*; results not shown). This approach could be particularly useful for studying species vulnerable to cutaneous pathogens known to have devastating impacts on wild populations, such as white-nose syndrome in bats (Field *et al.*, 2015). The methods were also robust to variables that may be introduced due to field conditions. For example, following the same protocols, we successfully established keratinocytes from tongue biopsies collected from recently deceased wolves (< 48 hours post-mortem) that were refrigerated in collection medium for up to ten days before they could be plated (results not shown).

My approach could also be adapted for other cell types to study allele-specific effects of genes identified to be under selection. For example, in North American gray wolves,

nonsynonymous single nucleotide polymorphisms (SNPs) hypothesized to impact biological processes including immunity, metabolism, and pigmentation show evidence of local adaptation to distinct environments (Schweizer *et al.*, 2016a; Schweizer *et al.*, 2016b), and could be further studied using cellular approaches. In particular, pigmentation has become a model phenotype for understanding adaptation, and candidate loci potentially impacting pigmentation have been identified in an array of species (Hoekstra, 2006; Hubbard *et al.*, 2010). The cellular approaches we describe here could be adapted for melanocyte culture (Costin *et al.*, 2004) or melanocyte-keratinocyte coculture (Lei *et al.*, 2002) to study effects of genetic variants on pigmentation within the genomic background of the species of interest. Gene editing tools can be particularly valuable for studies in which linkage disequilibrium is present or when a genotype of interest is extremely rare in a wild population (as in the case of K^{BB}, estimated to be at 2% frequency) (Coulson *et al.*, 2011).

Overall, our results suggest that the effects of *CBD103* genotype on fitness in North American gray wolves are not due to changes in gene expression within keratinocytes. Instead, *CBD103* genotype may have consequences on interactions of the secreted protein with microbes outside the cell, which could be further evaluated using the cellular resources developed in this research. More broadly, this study highlights the feasibility to establish population-level panels of cells from non-model, wild animals to experimentally test effects of genetic variants segregating in natural populations.

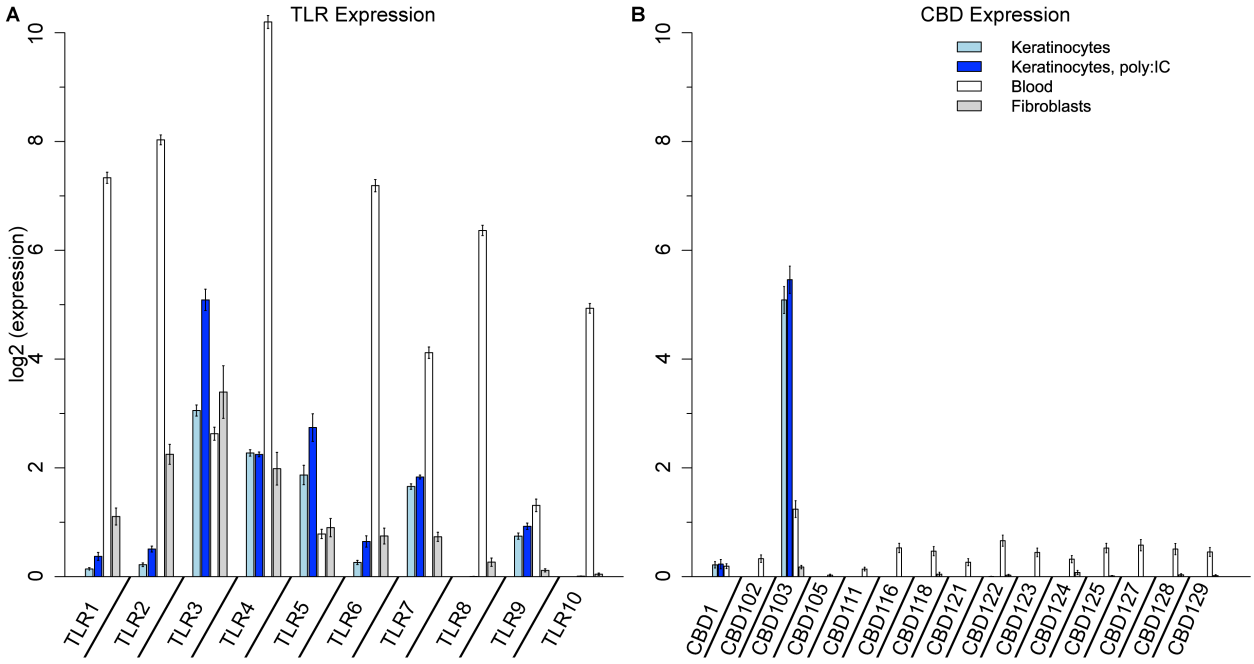

Figure 3-1. Expression of toll receptors and canine beta defensins. Normalized gene expression levels of toll receptors (A) and beta defensins (B) across keratinocytes ($N = 24$ cell lines), keratinocytes challenged with poly:IC ($N = 23$), whole blood ($N = 25$), and fibroblasts ($N = 6$) from wild gray wolves. Error bars represent the mean ± standard error across samples.

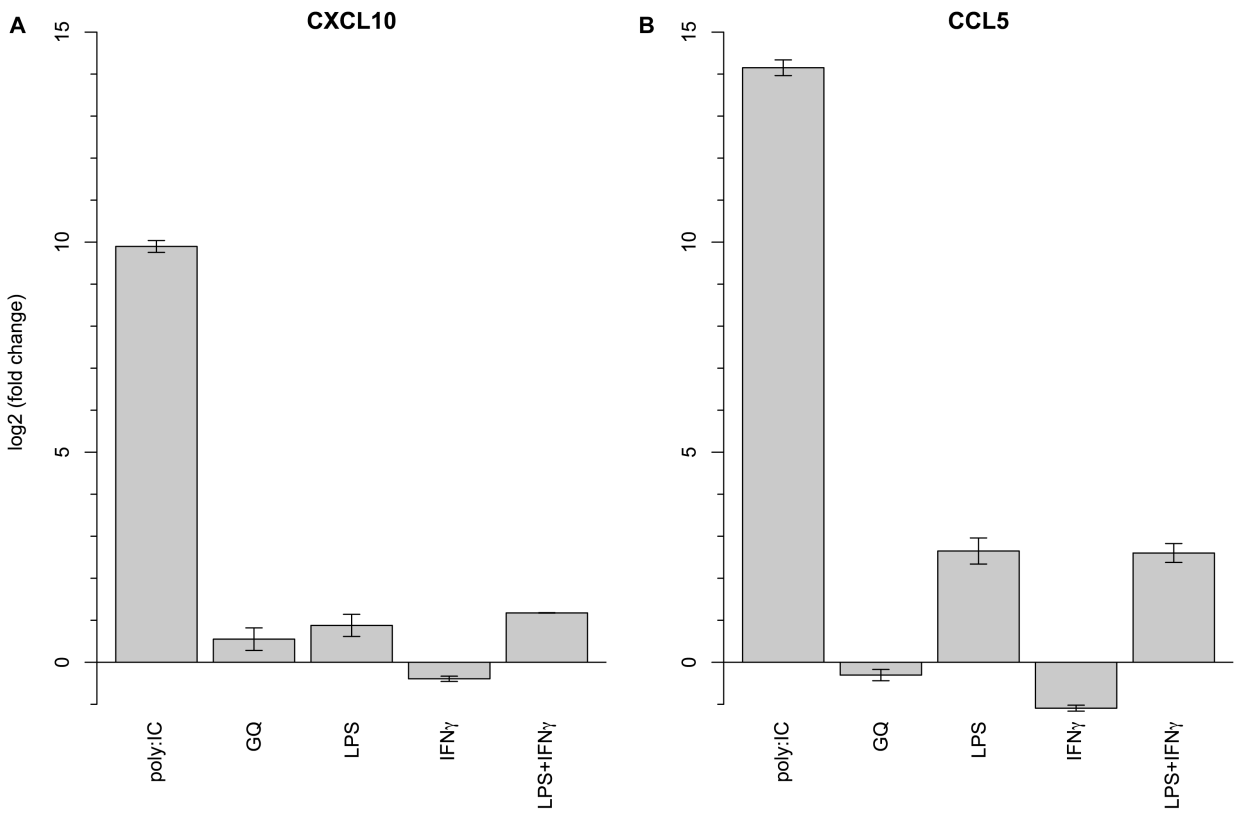

Figure 3-2. Gene expression response of wolf keratinocytes to immune challenge. Expression of *CXCL10* (A) and *CCL5* (B) in response to different antigens 24 hours post-exposure. Values represent the average of two keratinocyte cell lines, a wildtype (15071 K^{yy}) and a heterozygous (890M K^{By}) cell line. Error bars represent the mean ± standard error across the two cell lines.

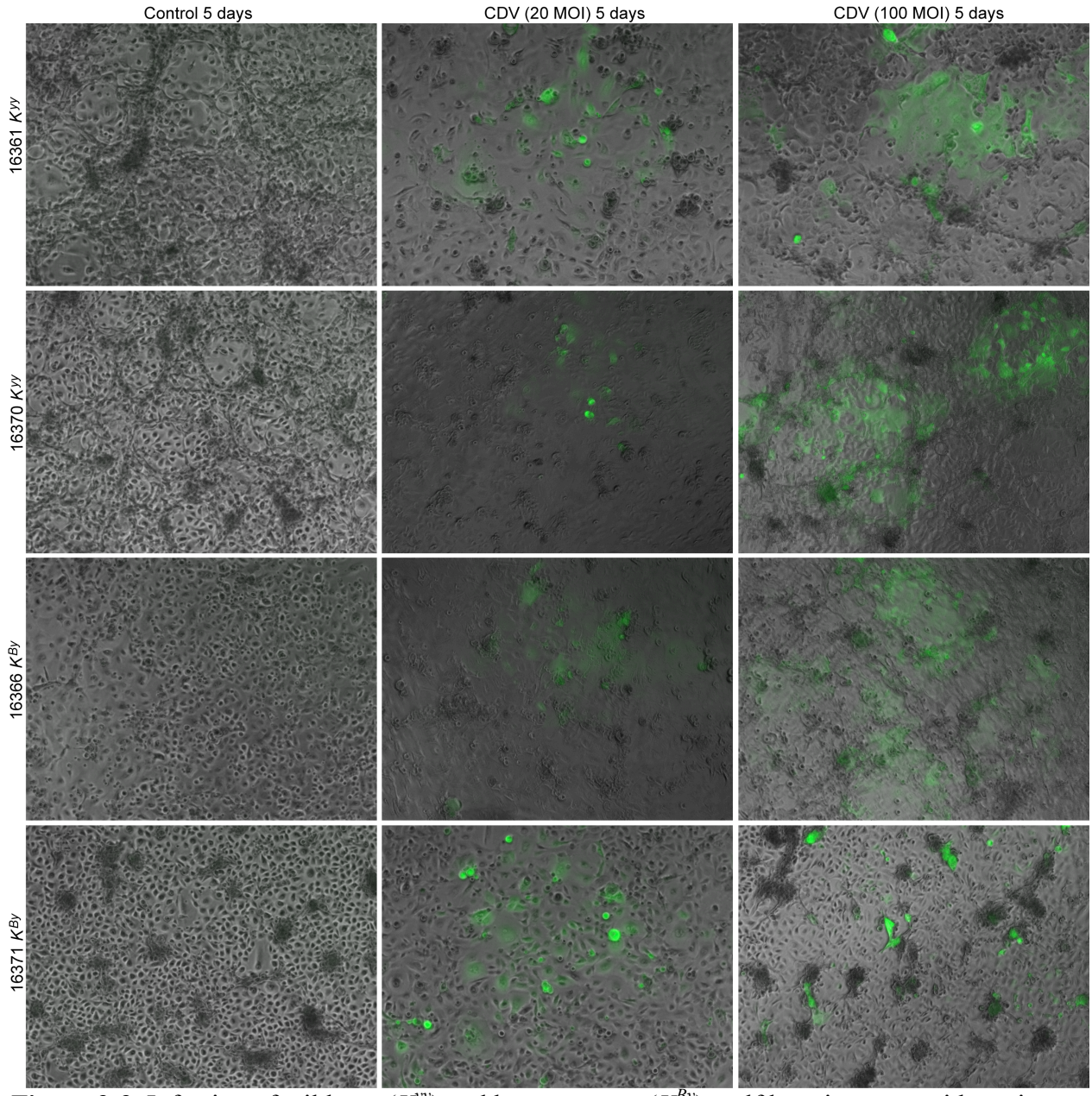

Figure 3-3. Infection of wildtype (K^{yy}) and heterozygous (K^{By}) wolf keratinocytes with canine distemper virus five days post-infection. Keratinocytes were infected at an MOI of 20 or 100 $TCID_{50}$/cell. Fluorescence images were captured to visualize CDV (expressing GFP) and overlaid on phase contrast images.

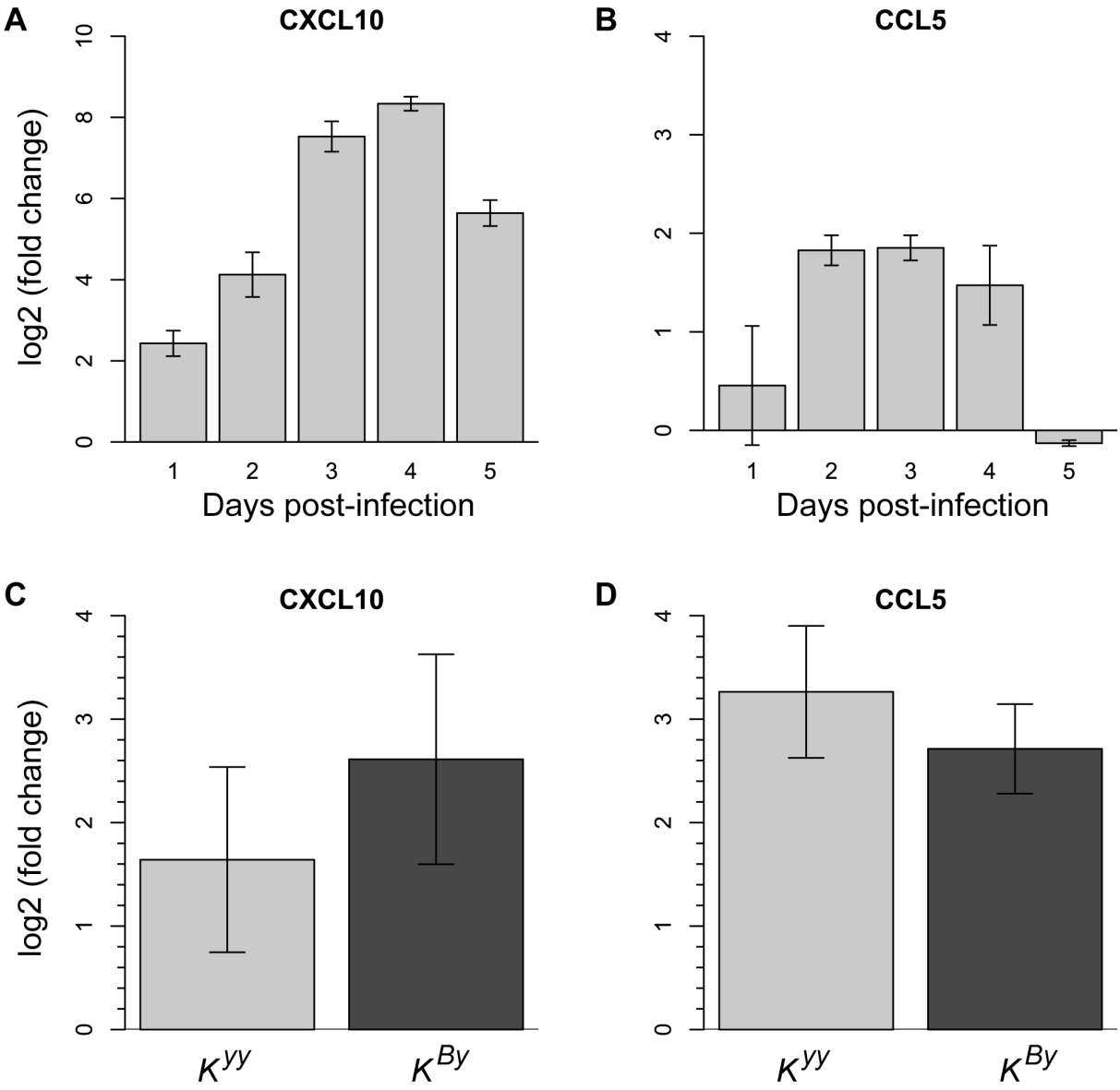

Figure 3-4. Response of wolf keratinocytes to live canine distemper virus. (A, B) Gene expression response of a single wildtype cell line (15071 K^{yy}) to CDV at an MOI of 20 $TCID_{50}$/cell, 1-5 days post-infection relative to day 0. (C, D) Gene expression response of wildtype (K^{yy}, $N = 3$) and heterozygous (K^{By}, $N = 3$) wolf keratinocytes to CDV at an MOI of 100 $TCID_{50}$/cell, 5 days post-infection. Error bars represent the mean ± standard error across duplicates (A, B) and across cell lines (C, D).

Supplemental Results

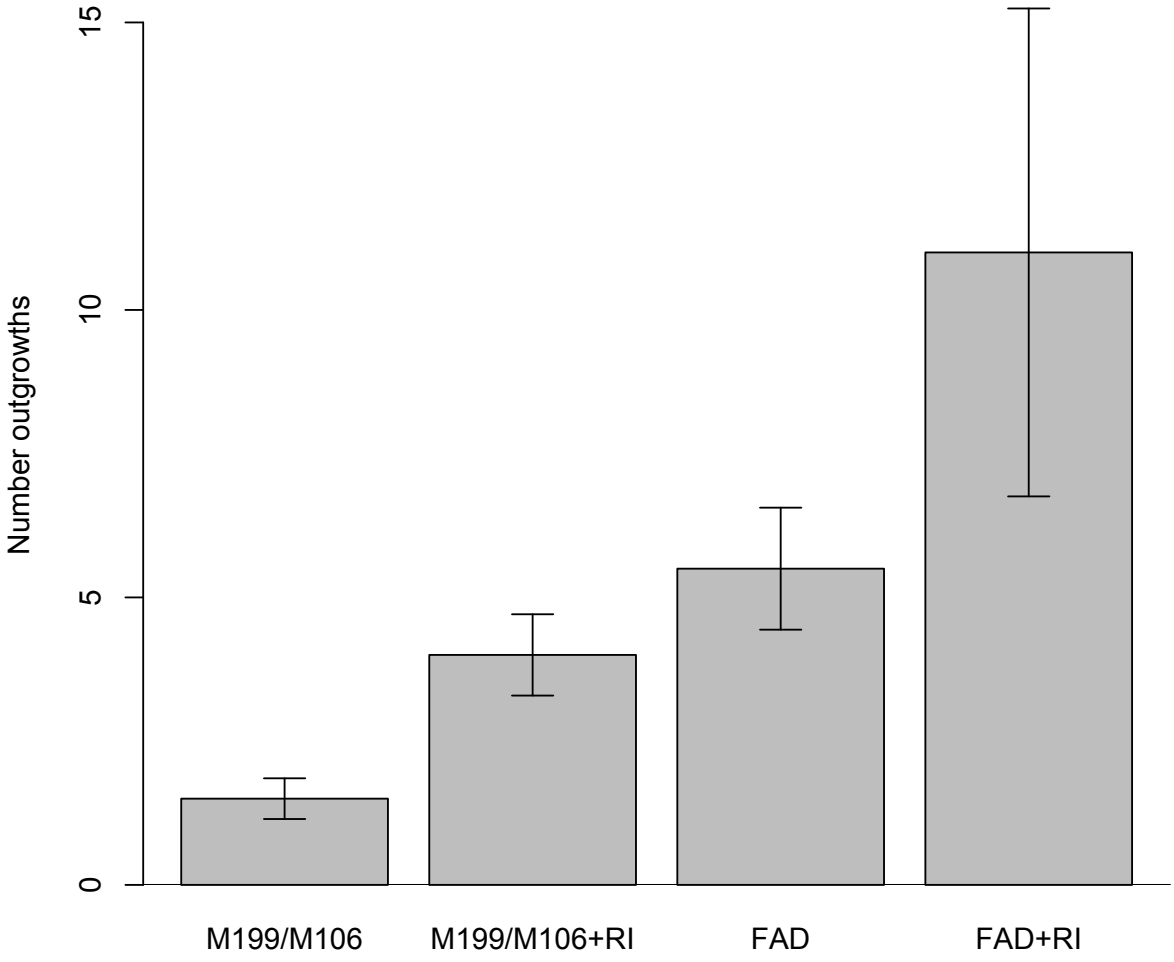

Figure 3-S1. Number of outgrowths of keratinocytes of a sample, six days after initial tissue plating. Skin biopsies plated on FAD + RI showed higher numbers of keratinocyte outgrowths compared to FAD medium alone, M199/M106 + RI, or M199/M106 medium alone. Error bars represent the mean ± standard error across the two cell lines. See Materials and Methods for complete media formulae.

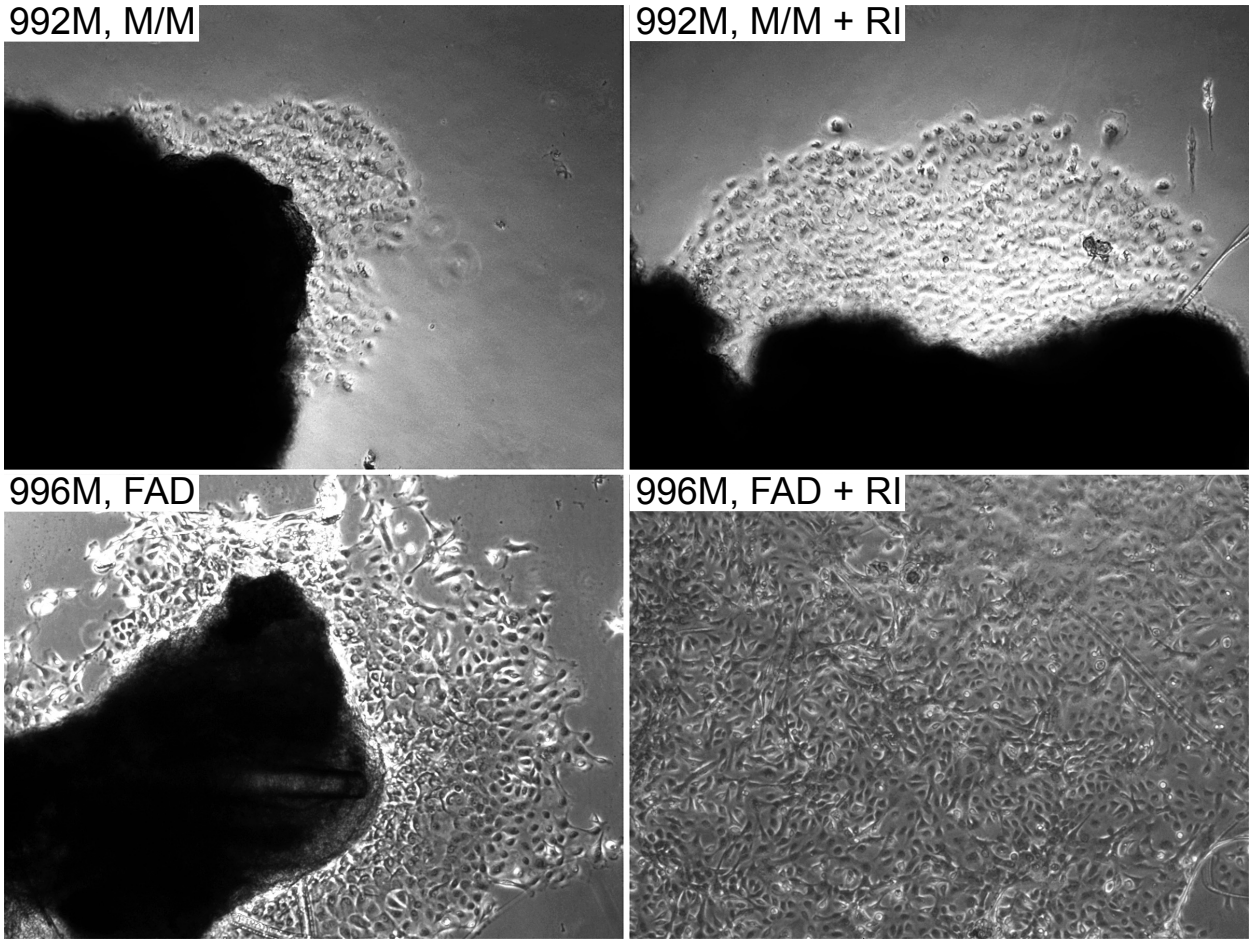

Figure 3-S2. Keratinocytes on different media, six days incubation after initial plating of minced 6 mm biopsies of wolf skin. Keratinocytes in FAD medium exhibited healthier morphology than keratinocytes in M199/M106 (M/M) medium, and keratinocytes in FAD medium with rock inhibitor (RI) exhibited healthy morphology as well as more rapid proliferation than the other treatments. See Materials and Methods for complete media formulae.

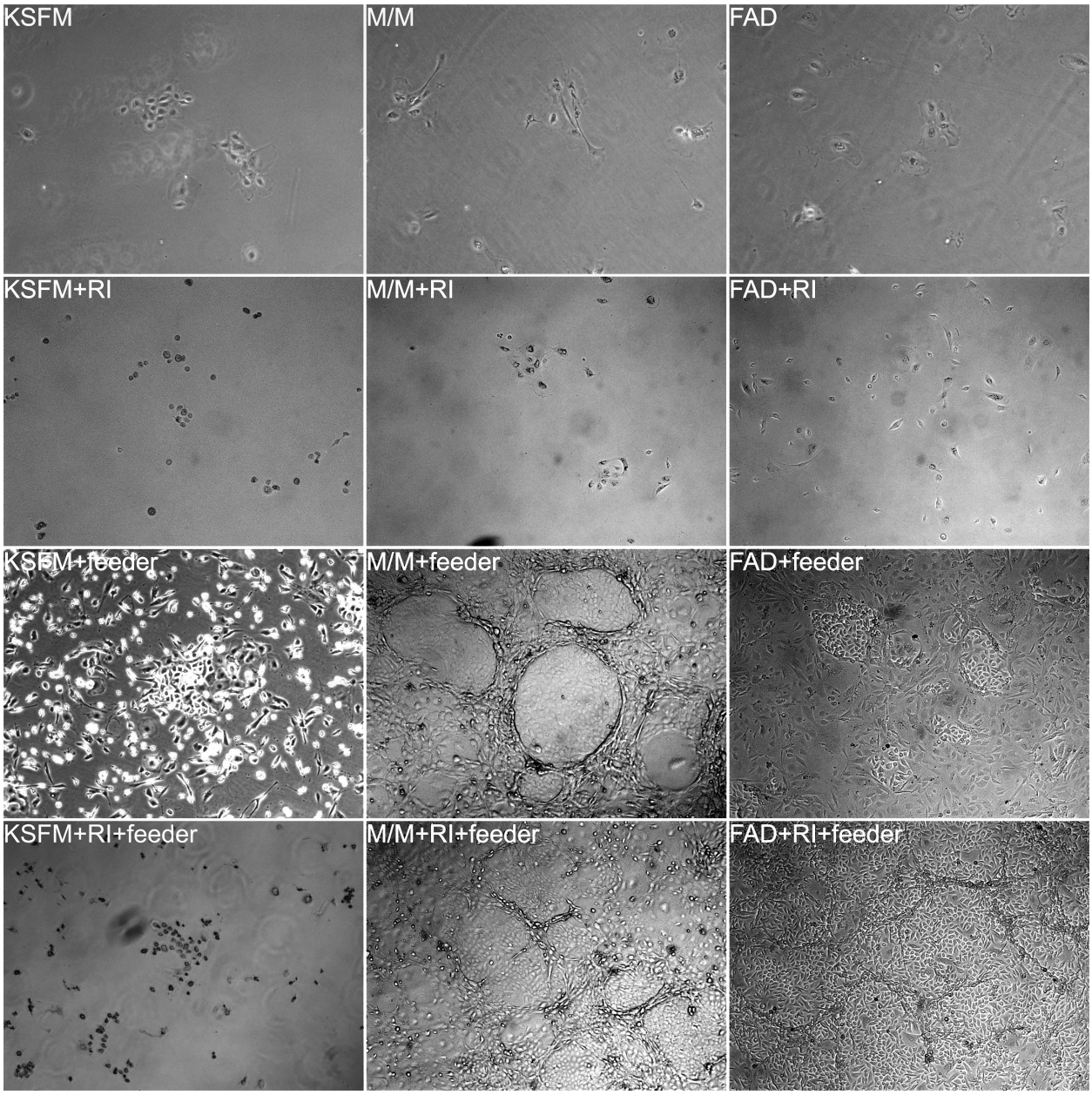

Figure 3-S3. Keratinocytes from a single wolf (16361) grown in different media. Keratinocytes were plated in parallel at $1x10^4$ cells per well in keratinocyte serum-free medium (KSFM), M199:M106 medium (M/M), or DMEM:F12 (FAD) medium with or without feeder cells or rock inhibitor (RI) (see Materials and Methods for media formulas). Images were taken after 5 days of growth. Whereas medium with or without rock inhibitor (RI) did not sustain keratinocyte proliferation, the presence of feeder cells improved keratinocyte health and proliferation in M/M and FAD media (feeders had low survival rate in KSFM), which was further improved with addition of RI (see Figure S4 for corresponding cell counts).

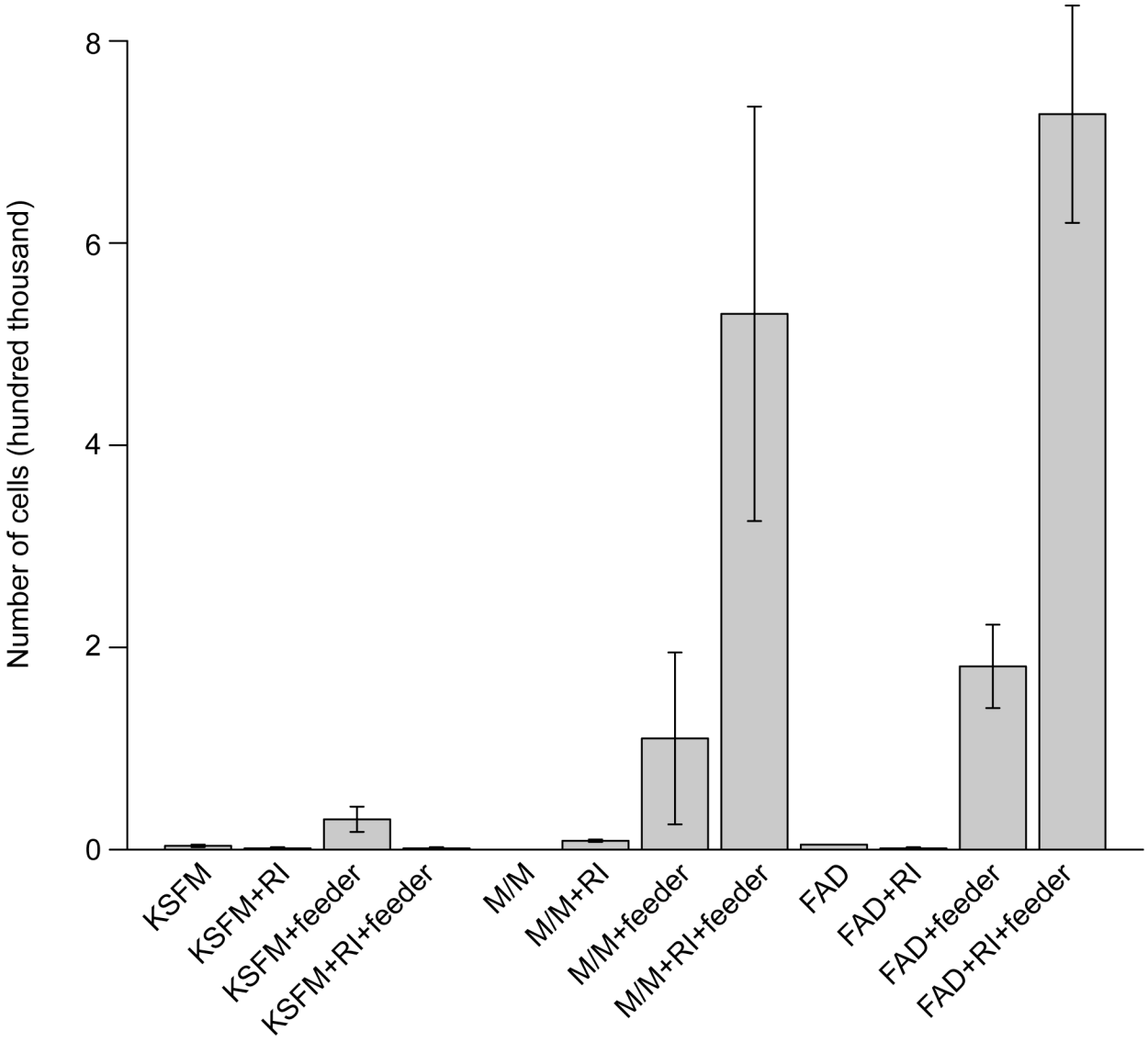

Figure 3-S4. Relative growth of wolf keratinocytes on different media. Keratinocytes were plated at 5,000 cells per well and cultured with different media for five days. Error bars represent standard errors of cells from two individual wolves (animals 16361 and 16366).

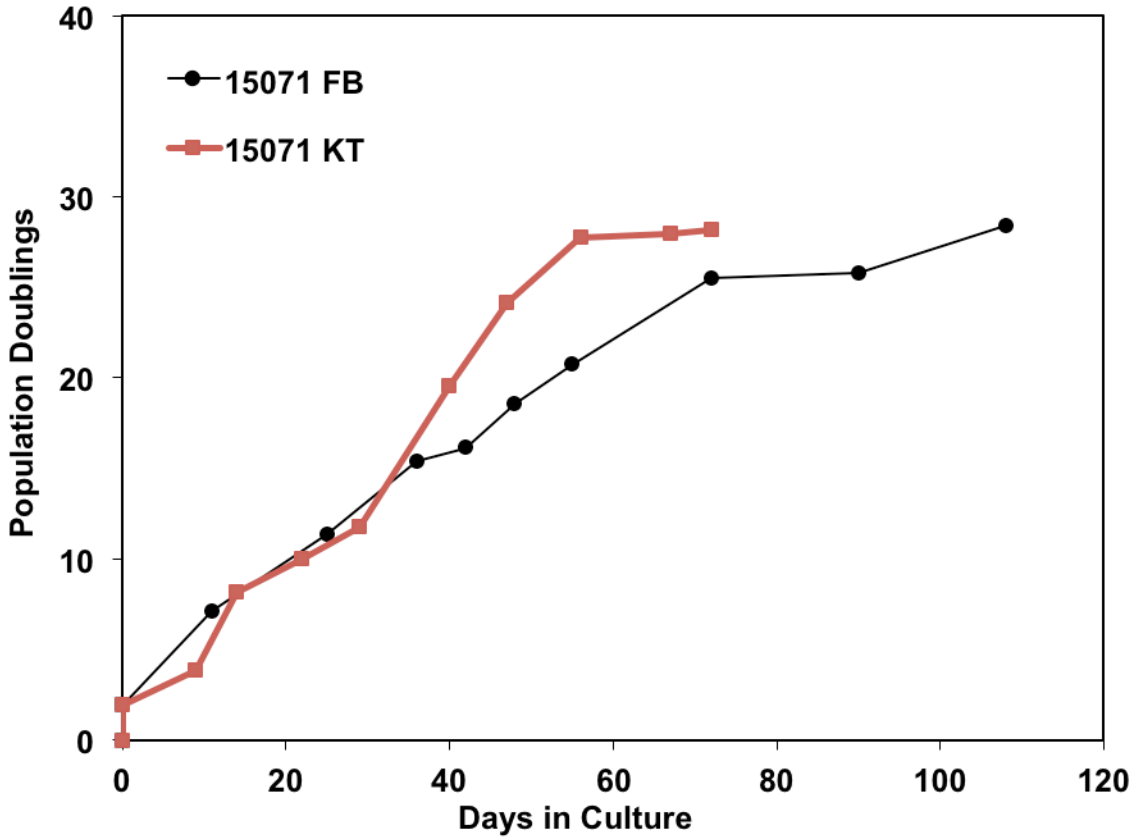

Figure 3-S5. Number of population doublings of keratinocytes (KT) and fibroblasts (FB) established from a single gray wolf (animal 15071). Keratinocytes reached 28 cell doublings after 72 days of culture in FAD medium with feeder cells (without rock inhibitor), while fibroblasts reached 28 cell doublings after 108 days of culture in M199/M106 medium. See Materials and Methods for complete media formulae.

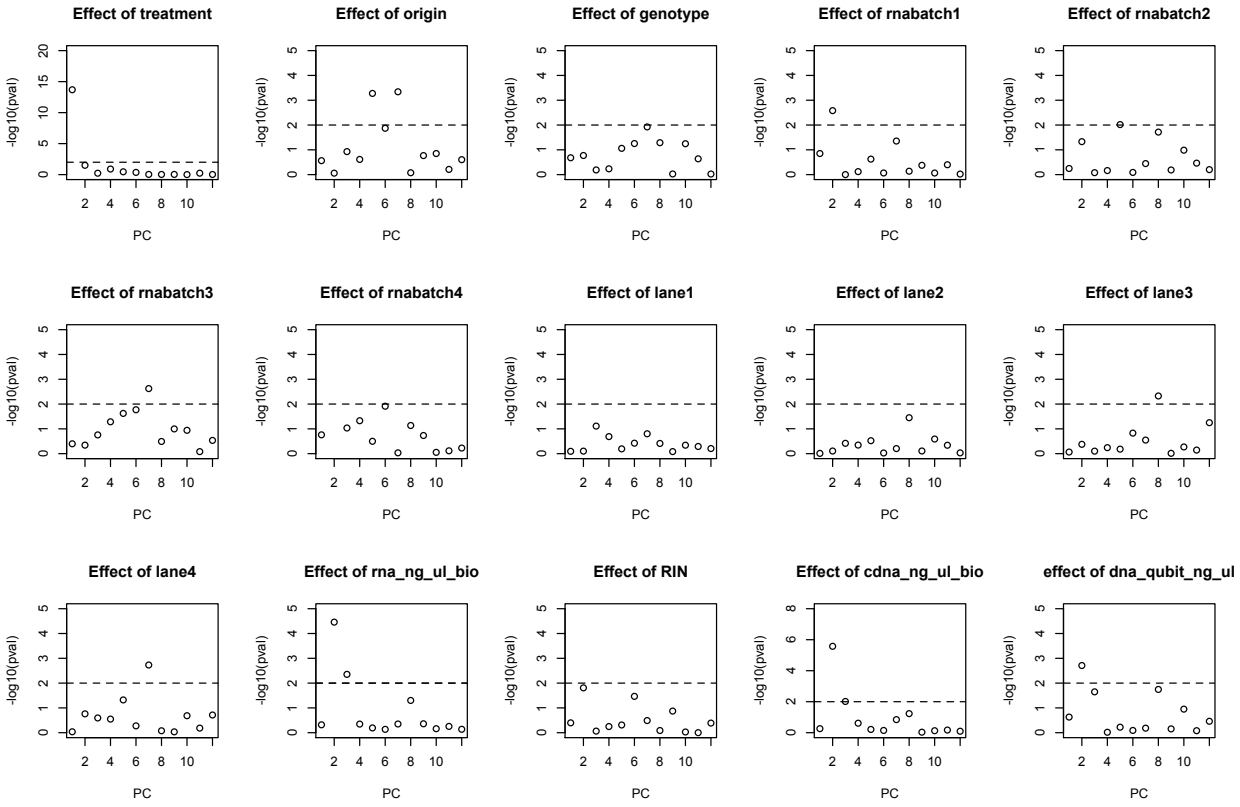

Figure 3-S6. Statistical significance of linear regression between each of the first twelve PCs of the RNA-Seq expression data and variables that could potentially structure the data. Significance values before regression of any variables. Treatment with poly:IC ("Effect of treatment") was significantly correlated ($P < 10^{-10}$) with the first PC. RNA batch and cDNA yield of library preparations as measured by an an Agilent 2100 Bioanalyzer ("Effect of cDNA_ng_ul_bio") were regressed from the data for gene expression analyses.

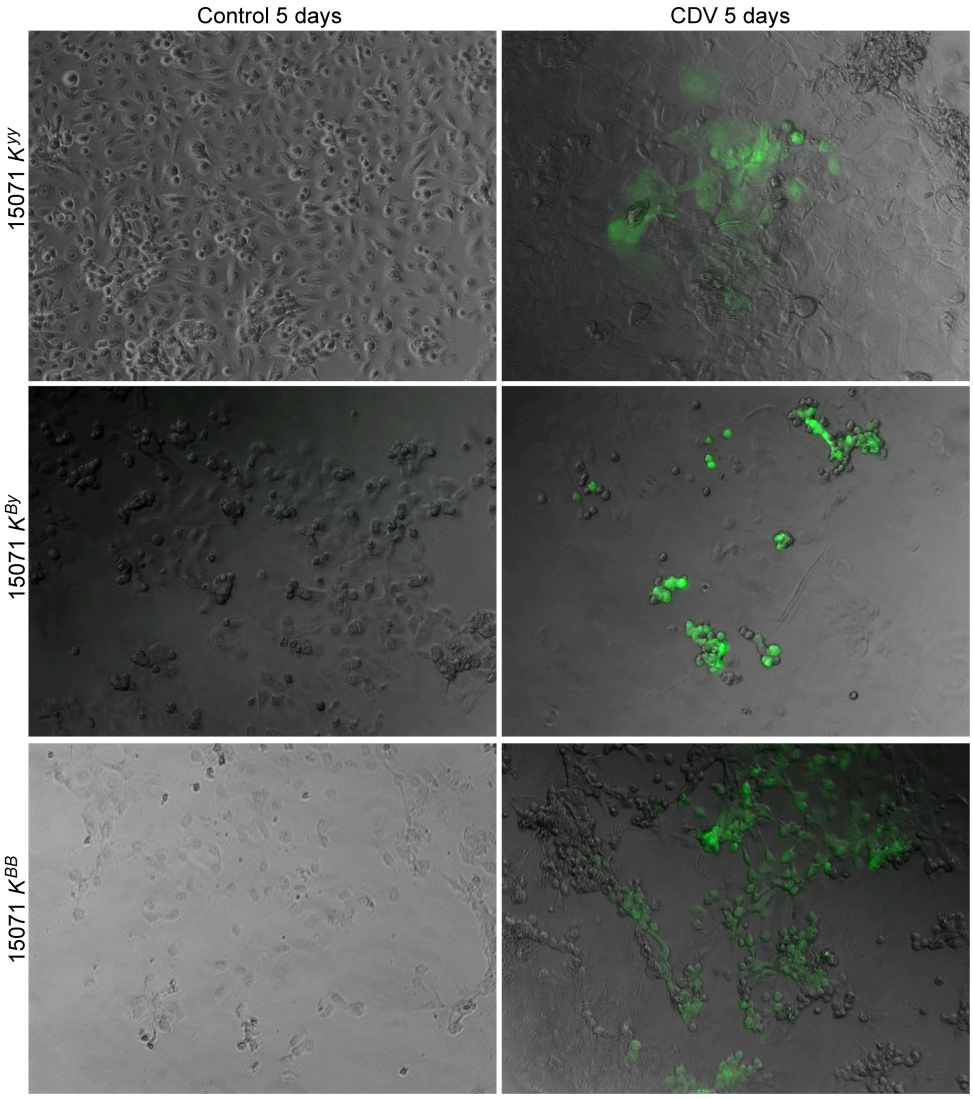

Figure 3-S7. Infection of CRISPR/Cas9 edited cell lines with canine distemper virus (CDV). Immortalized wildtype (K^{yy}) keratinocytes and CRISPR/Cas9 edited heterozygous (K^{By}) and homozygous (K^{BB}) wolf keratinocytes from a single wolf (animal 15071) infected with CDV, five days post-infection. Keratinocytes were infected at an MOI of 100 $TCID_{50}$/cell. Fluorescence images were captured to visualize CDV (expressing GFP) and overlaid on phase contrast images.

Animal ID	Age (yrs)	Sample origin	Treatment (non-challenged or poly:IC)	CBD103 genotype (0= K^{yy}, 1= K^{By})	RNA batch	Sequence lane	RNA 260/280	RIN	cDNA library yield (ng/ul)
1005F	1	Yellowstone	NC	1	3	4	2.06	10	229
1005F	1	Yellowstone	NC	1	3	4	2.07	10	237
1013M	2	Yellowstone	NC	0	4	4	2.02	8.3	219
1013M	2	Yellowstone	PIC	0	4	4	1.96	10	50
1014M	1	Yellowstone	NC	1	1	1	2.07	10	250
1014M	1	Yellowstone	PIC	1	1	1	2.05	10	132
14387	3.5	ID Dpt. F&G	PIC	0	1	4	2.07	10	165
15054	4.5	ID Dpt. F&G	NC	0	1	1	2.04	10	97
15054	4.5	ID Dpt. F&G	PIC	0	1	1	2.07	10	87
15062	1	ID Dpt. F&G	NC	0	2	2	2.1	10	229
15062	1	ID Dpt. F&G	PIC	0	2	2	2.1	10	229
15067	9.5	ID Dpt. F&G	NC	1	2	1	2.1	10	141
15067	9.5	ID Dpt. F&G	PIC	1	2	1	2.02	10	190
15070	5.5	ID Dpt. F&G	NC	0	1	3	2.05	10	366
15070	5.5	ID Dpt. F&G	PIC	0	1	3	2.05	10	199
15071	3.5	ID Dpt. F&G	NC	0	3	1	2.06	10	326
15071	3.5	ID Dpt. F&G	PIC	0	3	1	2.07	10	85
15072	7	ID Dpt. F&G	NC	0	2	1	2.13	10	304
15072	7	ID Dpt. F&G	PIC	0	2	1	2.05	10	494
16230	1	ID Dpt. F&G	NC	0	2	2	2.1	10	289
16230	1	ID Dpt. F&G	PIC	0	2	2	2.11	10	141
16235	9	ID Dpt. F&G	NC	0	4	2	2.02	10	121
16235	9	ID Dpt. F&G	PIC	0	4	2	1.98	10	102
16236	1	ID Dpt. F&G	NC	0	1	3	2.09	10	221
16236	1	ID Dpt. F&G	PIC	0	1	3	2.02	10	157
16240	10	ID Dpt. F&G	NC	0	3	3	2.11	10	174
16240	10	ID Dpt. F&G	PIC	0	3	3	2.11	10	132
16361	3.5	ID Dpt. F&G	NC	0	3	4	2	10	199
16361	3.5	ID Dpt. F&G	PIC	0	3	4	2.08	10	152
16366	5	ID Dpt. F&G	NC	1	3	2	2.04	10	176
16366	5	ID Dpt. F&G	PIC	1	3	2	2.02	10	100
16370	8	ID Dpt. F&G	NC	0	1	4	2.08	10	184
16370	8	ID Dpt. F&G	PIC	0	1	4	2.03	10	72
16371	1	ID Dpt. F&G	NC	1	4	3	2.04	10	71
16371	1	ID Dpt. F&G	PIC	1	4	3	2.02	8.4	58
821F	6.7	Yellowstone	NC	0	2	3	2.12	10	66
821F	6.7	Yellowstone	PIC	0	2	3	1.34	10	62
890M	4.6	Yellowstone	NC	1	4	1	2.03	9.9	36
890M	4.6	Yellowstone	PIC	1	4	1	1.95	9.7	37
910M	5.5	Yellowstone	NC	1	1	2	2.1	10	171
910M	5.5	Yellowstone	PIC	1	1	2	2.02	10	148
992M	6.6	Yellowstone	NC	1	3	3	2.06	10	66
992M	6.6	Yellowstone	PIC	1	3	3	2	10	56
993M	3.5	Yellowstone	NC	1	2	4	2.02	10	27
993M	3.5	Yellowstone	PIC	1	2	4	2.11	10	126
996M	0.6	Yellowstone	NC	1	4	2	1.94	8.4	95
996M	0.6	Yellowstone	PIC	1	4	2	2.02	7.5	86

Table 3-S1. Biological information of the 24 gray wolf keratinocyte lines, as well as corresponding technical data for the 24 cDNA libraries of non-challenged cells and 24 cDNA libraries of cells challenged with poly:IC for RNA-Seq analyses.

References

Abzhanov A, Extavour CG, Groover A, Hodges SA, Hoekstra HE, Kramer EM, Monteiro A (2008) Are we there yet? Tracking the development of new model systems. Trends in Genetics, **24**, 353-360.

Akdemir D, Godfrey OU (2014) EMMREML: Fitting mixed models with known covariance structures. http://CRANR-projectorg/package=EMMREML.

Almberg ES, Mech LD, Smith DW, Sheldon JW, Crabtree RL (2009) A serological survey of infectious disease in Yellowstone National Park's canid community. PloS one, **4**, e7042.

Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics, btu638.

Anderson TM, Candille SI, Musiani M, Greco C, Stahler DR, Smith DW, Padhukasahasram B, Randi E, Leonard JA, Bustamante CD (2009) Molecular and evolutionary history of melanism in North American gray wolves. Science, **323**, 1339-1343.

Appel MJ, Summers BA (1995) Pathogenicity of morbilliviruses for terrestrial carnivores. Veterinary microbiology, **44**, 187-191.

Barrett T (1999) Morbillivirus infections, with special emphasis on morbilliviruses of carnivores. Veterinary microbiology, **69**, 3-13.

Bedford NL, Hoekstra HE (2015) Peromyscus mice as a model for studying natural variation. Elife, **4**, e06813.

Ben-Nun IF, Montague SC, Houck ML, Tran HT, Garitaonandia I, Leonardo TR, Wang Y-C, Charter SJ, Laurent LC, Ryder OA (2011) Induced pluripotent stem cells from highly endangered species. nature methodS, **8**, 829-831.

Benirschke K (1984) The frozen zoo concept. Zoo Biology, **3**, 325-328.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS (2007) A β-defensin mutation causes black coat color in domestic dogs. Science, **318**, 1418-1423.

Charruau P, Johnston RA, Stahler DR, Lea A, Snyder-Mackler N, Smith DW, Cole SW, Tung J, Wayne RK (2016) Pervasive Effects of Aging on Gene Expression in Wild Wolves. Molecular biology and evolution, msw072.

Chen L, Tang L, Xiang H, Jin L, Li Q, Dong Y, Wang W, Zhang G (2014) Advances in genome editing technology and its promising application in evolutionary and ecological studies. Gigascience, **3**, 1.

Costin G-E, Vieira WD, Valencia JC, Rouzaud F, Lamoreux ML, Hearing VJ (2004) Immortalization of mouse melanocytes carrying mutations in various pigmentation genes. Analytical biochemistry, **335**, 171-174.

Coulson T, MacNulty DR, Stahler DR, Wayne RK, Smith DW (2011) Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history. Science, **334**, 1275-1278.

Deem SL, Spelman LH, Yates RA, Montali RJ (2000) Canine distemper in terrestrial carnivores: a review. Journal of Zoo and Wildlife medicine, **31**, 441-451.

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. Nature Reviews Genetics, **4**, 649-655.

Field KA, Johnson JS, Lilley TM, Reeder SM, Rogers EJ, Behr MJ, Reeder DM (2015) The white-nose syndrome transcriptome: activation of anti-fungal host responses in wing tissue of hibernating little brown myotis. Plos Pathog, **11**, e1005168.

Finney DJ: **Statistical method in biological assay**. In.: JSTOR; 1952.

García J-R, Jaumann F, Schulz S, Krause A, Rodríguez-Jiménez J, Forssmann U, Adermann K, Klüver E, Vogelmeier C, Becker D (2001) Identification of a novel, multifunctional β-defensin (human β-defensin 3) with specific antimicrobial activity. Cell and tissue research, **306**, 257-264.

Gay NJ, Gangloff M (2007) Structure and function of Toll receptors and their ligands. Annu Rev Biochem, **76**, 141-165.

Guyon R, Lorentzen TD, Hitte C, Kim L, Cadieu E, Parker HG, Quignon P, Lowe JK, Renier C, Gelfenbeyn B, Vignaux F, DeFrance HB, Gloux S, Mahairas GG, André C, Galibert F, Ostrander EA (2003) A 1-Mb resolution radiation hybrid map of the canine genome. Proc Natl Acad Sci USA, **100**, 5296-5301.

Harder J, Bartels J, Christophers E, Schröder J-M (2001) Isolation and characterization of human β-defensin-3, a novel human inducible peptide antibiotic. Journal of Biological Chemistry, **276**, 5707-5713.

Hazrati E, Galen B, Lu W, Wang W, Ouyang Y, Keller MJ, Lehrer RI, Herold BC (2006) Human α-and β-defensins block multiple steps in herpes simplex virus infection. The Journal of Immunology, **177**, 8658-8666.

Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity, **97**, 222-234.

Horvath S (2011) *Weighted network analysis: applications in genomics and systems biology*. Springer Science & Business Media.

Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ (2010) Vertebrate pigmentation: from underlying genes to adaptive function. Trends in Genetics, **26**, 231-239.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol, **14**, R36.

Trim Galore [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/]

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, **9**, 559.

Lebre MC, van der Aar AM, van Baarsen L, van Capel TM, Schuitemaker JH, Kapsenberg ML, de Jong EC (2007) Human keratinocytes express functional Toll-like receptor 3, 4, 5, and 9. Journal of Investigative Dermatology, **127**, 331-341.

Lei TC, Virador VM, Vieira WD, Hearing VJ (2002) A melanocyte–keratinocyte coculture model to assess regulators of pigmentation in vitro. Analytical biochemistry, **305**, 260-268.

Leonard BC, Marks SL, Outerbridge CA, Affolter VK, Kananurak A, Young A, Moore PF, Bannasch DL, Bevins CL (2012) Activity, expression and genetic variation of canine β-defensin 103: a multifunctional antimicrobial peptide in the skin of domestic dogs. Journal of innate immunity, **4**, 248-259.

Lochmiller RL, Deerenberg C (2000) Trade‐offs in evolutionary immunology: just what is the cost of immunity? Oikos, **88**, 87-98.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics, **152**, 1753-1766.

Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA (1997) A Linkage Map of the Canine Genome. Genomics, **46**, 326-336.

Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM, Aguirre GD, Ziegle JS, Ostrander EA, Rine J (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. Genetics, **151**, 803-820.

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular ecology notes, **6**, 288-295.

Quiñones-Mateu ME, Lederman MM, Feng Z, Chakraborty B, Weber J, Rangel HR, Marotta ML, Mirza M, Jiang B, Kiser P (2003) Human epithelial β-defensins 2 and 3 inhibit HIV-1 replication. Aids, **17**, F39-F48.

Ramaswamy K, Yik WY, Wang X-M, Oliphant EN, Lu W, Shibata D, Ryder OA, Hacia JG (2015) Derivation of induced pluripotent stem cells from orangutan skin fibroblasts. BMC research notes, **8**, 1.

Rheinwald JG (1989) Methods for clonal growth and serial cultivation of normal human epithelial keratinocytes and mesothelial cells. In: Cell growth and division A practical approach, pp. 81-94. Oxford: IRL Press.

Rheinwald JG, Green H (1975) Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. Cell, **6**, 331-343.

Rosenblum EB, Poorten TJ, Settles M, Murdoch GK (2012) Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. Molecular Ecology, **21**, 3110-3120.

Schweizer RM, Robinson J, Harrigan R, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wayne RK (2016a) Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Molecular ecology, **25**, 357-379.

Schweizer RM, Vonholdt BM, Harrigan R, Knowles JC, Musiani M, Coltman D, Novembre J, Wayne RK (2016b) Genetic subdivision and candidate genes under selection in North American grey wolves. Molecular ecology, **25**, 380-402.

Selsted ME, Ouellette AJ (2005) Mammalian defensins in the antimicrobial immune response. Nature immunology, **6**, 551-557.

Shively CA, Clarkson TB (2009) The unique value of primate models in translational research. American Journal of Primatology, **71**, 715-721.

Snyder-Mackler N, Sanz J, Kohn JN, Brinkworth JF, Morrow S, Shaver AO, Grenier J-C, Pique-Regi R, Johnson ZP, Wilson ME (2016) Social status alters immune regulation and response to infection in macaques. Science, **354**, 1041-1045.

Stahler DR, MacNulty DR, Wayne RK, VonHoldt B, Smith DW (2013) The adaptive value of morphological, behavioural and life‐history traits in reproductive female wolves. Journal of Animal Ecology, **82**, 222-234.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity (Edinb), **100**, 158-170.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. cell, **131**, 861-872.

Tanner MA, Berk LS, Felten DL, Blidy AD, Bit SL, Ruff DW (2002) Substantial changes in gene expression level due to the storage temperature and storage duration of human whole blood. Clinical & Laboratory Haematology, **24**, 337-341.

Todaro GJ, Green H (1963) Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. The Journal of cell biology, **17**, 299-313.

Tovar H, Navarrete F, Rodríguez L, Skewes O, Castro FO (2008) Cold storage of biopsies from wild endangered native Chilean species in field conditions and subsequent isolation of primary culture cell lines. In Vitro Cellular & Developmental Biology-Animal, **44**, 309-320.

Von Messling V, Milosevic D, Cattaneo R (2004) Tropism illuminated: lymphocyte-based pathways blazed by lethal morbillivirus through the host immune system. Proceedings of the National Academy of Sciences of the United States of America, **101**, 14216-14221.

vonHoldt BM, DR S, DW S, Earl DA, Pollinger JP, Wayne RK (2008) The genealogy and genetic viablity of reintroduced Yellostone grey wolves. Mol Ecol, **17**, 252-274.

vonHoldt BM, Stahler DR, Bangs EE, Smith DW, Jimenez MD, Mack CM, Niemeyer CC, Pollinger JP, Wayne RK (2010) A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. Mol Ecol, **19**, 4412-4427.

Wang J (2011) COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol Ecol Resour, **11**, 141-145.

Wilson SS, Wiens ME, Smith JG (2013) Antiviral mechanisms of human defensins. Journal of molecular biology, **425**, 4965-4980.