Behavioral Ecology

The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2018), XX(XX), 1–11. doi:10.1093/beheco/ary175

Original Article

Gene expression shifts in yellow-bellied marmots prior to natal dispersal

Tiffany C. Armenta,[a,b,*] Steve W. Cole,[c] Daniel H. Geschwind,[d] Daniel T. Blumstein,[a,b,*] and Robert K. Wayne[a,*]

[a]Department of Ecology and Evolutionary Biology, University of California Los Angeles, 612 Charles E Young Dr. East, Los Angeles, CA 90095, USA, [b]Rocky Mountain Biological Laboratory, P.O. Box 519, Crested Butte, CO 81224, USA, [c]Department of Medicine, Division of Hematology-Oncology, University of California Los Angeles, 11-934 Factor Building, Los Angeles, CA 90095, USA, and [d]Department of Psychiatry and Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, 695 Charles E Young Dr. South, Los Angeles, CA 90095, USA

Received 31 May 2018; revised 12 November 2018; editorial decision 15 November 2018; accepted 26 November 2018.

The causes and consequences of vertebrate natal dispersal have been studied extensively, yet little is known about the molecular mechanisms involved. We used RNA-seq to quantify transcriptomic gene expression in blood of wild yellow-bellied marmots (*Marmota flaviventer*) prior to dispersing from or remaining philopatric to their natal colony. We tested 3 predictions. First, we hypothesized dispersers and residents will differentially express genes and gene networks since dispersal is physiologically demanding. Second, we expected differentially expressed genes to be involved in metabolism, circadian processes, and immune function. Finally, in dispersing individuals, we predicted differentially expressed genes would change as a function of sampling date relative to dispersal date. We detected 150 differentially expressed genes, including genes that have critical roles in lipid metabolism and antigen defense. Gene network analysis revealed a module of 126 coexpressed genes associated with dispersal that was enriched for extracellular immune function. Of the dispersal-associated genes, 22 altered expression as a function of days until dispersal, suggesting that dispersal-associated genes do not initiate transcription on the same time scale. Our results provide novel insights into the fundamental molecular changes required for dispersal and suggest evolutionary conservation of functional pathways during this behavioral process.

Key words: **functional genetics, gene expression, immune system, natal dispersal, transcriptomics, vertebrates.**

INTRODUCTION

Natal dispersal, the permanent movement of an individual from their birth site to a new site, is a complex trait influenced by morphology, physiology, and behavior (Clobert et al. 2001). In vertebrates, one sex is typically philopatric, but even for the less dispersing sex there is substantial variation in the likelihood of dispersal (Greenwood 1980; Clutton-Brock and Lukas 2012) and a wide variety of both internal and external factors influence this variation. Internal factors such as mass and body condition (Nilsson and Smith 1985; Meylan et al. 2002; Barbraud et al. 2003), hormones such as testosterone and glucocorticoids (Woodroffe et al. 1993; Silverin 1997; reviewed by Dufty and Belthoff 2001), behavioral traits (Dingemanse et al. 2003; Cote and Clobert 2007; reviewed by Cote et al. 2010), and additive genetic variance (Doligez and Pärt 2008) as well as external forces such as population density (Matthysen 2005), breeding opportunities (Pruett-Jones and Lewis 1990; Alberts and Altmann 1995; Pope 2000), habitat quality (Lin and Batzli 2001; Ekernas and Cords 2007), and social relationships (Bekoff 1977; Gese et al. 1996) may affect an individual's decision to disperse.

There are many important consequences of dispersal. On the population level, it can drive ecological and evolutionary dynamics through gene flow, genetic subdivision, and distributional extent (Chepko-Sade and Halpin 1987; Clobert et al. 2001; Duckworth and Badyaev 2007). For an individual, leaving the natal colony is one of the most profound and potentially costly events to occur in one's lifetime (Smale et al. 1997). Dispersers leave a familiar environment to potentially face a heightened risk of predation (Van Vuren and Armitage 1994; Yoder et al. 2004), physical attacks from unfamiliar conspecifics (Soulsbury et al. 2008), and exposure to novel pathogens in new environments (Srygley et al. 2009; Bonte et al. 2012).

Address correspondence to T.C. Armenta. E-mail: tiffany.armenta@gmail.com.
*These two authors contributed equally to this work.

The many risks that individuals may face during dispersal are not trivial; thus, we expect that dispersers may likely prepare physiologically for them. However, little is known about the physiology underlying dispersal in vertebrates. Insects can display dramatic dispersal polymorphisms due to alternative life-history strategies, including winged dispersers and sedentary wingless morphs of the same species (Zera and Denno 1997). Transcriptomic comparisons of these morphs have revealed underlying coordinated physiological strategies, where dispersers significantly up-regulated genes involved in cellular energy production, metabolism, and muscle building (Brisson et al. 2007; Vellichirammal et al. 2014), yet it remains unknown whether the observed gene expression differences are due to somatic requirements or are associated with the dispersal behavior itself.

Studies examining gene expression during dispersal are limited, but the molecular pathways involved in seasonal migration are well studied. Since dispersal and migration both require endurance for long distance travel, occur at discrete times of year, and incur many of the same risks, the physiological condition necessary for these 2 life-history events may be somewhat analogous (Liedvogel et al. 2011). Genes involved in metabolism and mobilization of cellular energy stores, especially lipids, are commonly differentially expressed (DE) between migratory phenotypes. Metabolic gene expression is significantly higher in migratory versus nonmigratory birds (Boss et al. 2016; Fudickar et al. 2016) and butterflies (Zhu et al. 2009), winged aphids and crickets (Brisson et al. 2007; Vellichirammal et al. 2014), and moths that experimentally fly long vs. short distances (Jones et al. 2015). Lipids are the primary source of energy during migration, so it is not surprising that genes involved in lipid metabolism are altered during long periods without food (Blem 1980; Jenni and Jenni-Eiermann 1998; Jeffs et al. 1999). Genes that dictate circadian processes are also known to be important aspects of migration phenology for invertebrates (Zhu et al. 2009) and vertebrates (Boss et al. 2016; Johnston et al. 2016).

Yellow-bellied marmots (*Marmota flaviventer*) are an excellent system in which to study the molecular pathways involved in vertebrate dispersal because the timing of dispersal is highly predictable (Blumstein et al. 2009; Armitage 2014) and sex-biased differences in dispersal allow a case-controlled approach. In late June to early July of their second summer (as yearlings), nearly all males and approximately 50% of females disperse to new areas outside of their mother's home range (hereafter called dispersers; Van Vuren 1990; Armitage 2014). The remaining 50% of female yearlings never leave their natal site, are recruited into the matriline, and breed in the group for the remainder of their lives (hereafter, residents). A female's affiliative social relationships with other females in her natal colony, particularly matriarchs (her mother, older sisters) explain variation in dispersal (Blumstein et al. 2009; Armitage et al. 2011). As predicted by the social cohesion hypothesis (Bekoff 1977), females that engage in more affiliative behaviors with kin are less likely to disperse (Blumstein et al. 2009).

Sociogenomics is an emerging field that identifies the genes that are subject to social–environmental regulation (Robinson et al. 2005). In vertebrates, social stress can lead to persistent changes in gene expression and often results specifically in an up-regulation of proinflammatory genes and down-regulation of antiviral, innate immune response genes (Cole 2014; Turecki and Meaney 2016). This conserved transcriptional response to the social environment has been observed across many vertebrates, including laboratory rodents (Avitsur et al. 2006; Wu et al. 2014), captive primates (Cole et al. 2012; Tung et al. 2012; Snyder-Mackler et al. 2016), and humans (Cole 2014; Fredrickson et al. 2015). Because dispersing female marmots are significantly less socially integrated than their philopatric counterparts, we expected an analogous up-regulation of antibacterial, proinflammatory genes in less socially integrated dispersers.

To evaluate the molecular pathways associated with dispersal, we conducted RNA sequencing (RNA-seq) of circulating blood cells in female yearling yellow-bellied marmots that either subsequently dispersed or remained residents in their natal colony. We used 2 methods to identify the genes associated with dispersal—one testing for differential expression of individual genes across the genome using linear mixed models, and a second network analysis identifying groups ("modules") of genes that were highly coexpressed across samples (Zhang and Horvath 2005; Oldham et al. 2008) which were then tested for correlation with dispersal. Coexpression of genes suggests they are involved in similar biological processes (Weston et al. 2008; Parikshak et al. 2015), facilitating an understanding of how genes coordinate to achieve a complex phenotype on a more inclusive level than individual genes. Because environmental variation makes gene expression difficult to interpret in natural systems, we collected RNA from all individuals in this population as frequently as possible in order to increase our power of detection and to look at expression profiles across multiple time points. Based on known molecular profiles of dispersing and migratory phenotypes, we tested 3 hypotheses: 1) dispersing and resident marmots differentially express genes prior to the dispersal event; 2) lists of dispersal-related genes are enriched for the biological processes of lipid metabolism, circadian processes, and extracellular pathogen defense; and 3) genes from these lists are likely beneficial in preparing for dispersal, so genes up-regulated in dispersers should increase expression as the date of dispersal approaches.

METHODS

Study system

We studied wild yellow-bellied marmots near the Rocky Mountain Biological Laboratory in Gunnison County, Colorado from 2013 to 2015. Marmots hibernate over-winter and are active from approximately mid-May to mid-September (Blumstein et al. 2004). Each year, we livetrapped this population throughout the active season to collect samples, affix ear tags, and apply unique dorsal fur marks for individual identification from afar (Blumstein 2013). We observed 8 colonies during morning and afternoon peak hours of marmot activity (Armitage 1962) and recorded the date, time, and location of behaviors observed as done previously (Blumstein et al. 2009). In general, each site was under continuous observation during periods of peak marmot activity throughout the active season and most individuals were observed daily, allowing us to identify when yearlings dispersed from the natal colony as done previously (Blumstein et al. 2009).

Behavioral observations

Dispersers leave during their second summer (as yearlings), approximately when young of the year emerge from their natal burrows (Armitage 1991, 2014). Since pup emergence dates vary across years and colony sites, we defined dispersers as those that were observed or trapped earlier in the summer and were not seen later that year up to 10 days after the date of a colony site's first pup emergence (as in Blumstein et al. 2009). We acknowledge that some disappearances may have been due to undetected mortality and

not dispersal. However, there is minimal risk in this assumption as previous work in this population has shown high rates of survival in both dispersing (0.73) and resident female yearlings (0.87; Van Vuren and Armitage 1994). We defined dispersal date as the last date an individual was observed or trapped at its natal colony. Animals known to have died during the year of sampling and those that returned to their natal colony after a prolonged absence were excluded from all analyses.

RNA sampling and sequencing

During each yearling capture, we collected 1 mL whole blood and preserved it in 2.5 mL PAXgene™ Blood RNA solution (PreAnalytiX, Qiagen, Hombrechtikon, Switzerland). This resulted in RNA samples from 43 distinct female marmots (35 individuals sampled once and 8 sampled on 2 or more occasions), yielding a total of 43 samples analyzed (one per individual). To comprehensively evaluate the temporal effects of dispersal on gene expression, we would have had to repeatedly sample RNA of dispersing individuals prior to the dispersal event. However, because most dispersers permanently leave the natal colony within the first 30 days of our trapping season, we rarely trapped dispersing individuals multiple times (5 of 16 dispersers were sampled twice prior to dispersal and only 4 of these samples had RNA Integrity Number [RIN] >4). Thus, our temporal analysis was constrained to accounting for the number of days between sampling and dispersal for single samples per individual.

We sampled blood because it can be obtained through a minimally invasive approach and it is a useful tissue for exploring a variety of physiological functions. Blood is dominated by non-nucleated red blood cells, consequently, white blood cells or leukocytes are the predominant source of RNA transcripts. Leukocytes are immune cells that defend the body against foreign antigens and thus, blood is an appropriate tissue to study the predicted response to pathogens. Furthermore, leukocytes share approximately 80% of mRNA with other tissues (Liew et al. 2006), making blood a powerful surrogate for many tissues including brain (Sullivan et al. 2006; Rudkowska et al. 2011; Kohane and Valtchinov 2012).

Samples were incubated, frozen, and extracted according to PAXgene™ Blood RNA kit manufacturers' instructions. Only samples taken during the dispersal period (up to 10 days after the colony's pup emergence) were included in subsequent steps. We removed globin transcripts using the GLOBINclear™ kit for mouse/rat (Ambion, ThermoFisher Scientific, Waltham, MA) and RNA quality was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). To preserve statistical power, we excluded any samples with RIN < 4 and globally corrected for RNA degradation by regressing any effect of RIN (as suggested by Gallego Romero et al. 2014). cDNA libraries were created from isolated messenger RNA (mRNA) using a TruSeq Library Prep Kit v2 (Illumina, Madison, WI), quantified with the KAPA SYBR® Fast qPCR library quantification kit (Kapa Biosystems Inc., Wilmington, MA), and pooled at 8–10 samples per lane. Single-end 100 bp sequencing was performed on Illumina HiSeq2000 (2013 samples) and HiSeq4000 (2014–2015) platforms at the Vincent J. Coates Genomics Sequencing Laboratory, UC Berkeley, USA.

Read mapping and expression quantification

We trimmed adapters and removed short (<20 base pairs) and low-quality reads (Phred score < 20) using Trim Galore! (Krueger 2015). Resulting reads were mapped to the 13-lined ground squirrel (*Ictidomys tridecemlineatus*) genome (spetri2, GenBank Assembly ID GCA_000236235.1) using TopHat2 v.2.1.0 (Trapnell et al. 2009; Kim et al. 2013). These species are estimated to have diverged approximately 8.6 million years ago (Bininda-Emonds et al. 2007; Soria-Carrasco and Castresana 2012) and exhibit sequence divergence of 13.2% (Thomas and Martin 1993). To maximize uniquely mapped reads, we allowed 8 mismatches, a 10-base pair (bp) gap length, and a 20 bp edit distance between read sequences and the reference genome. We quantified expression of uniquely mapped reads per individual using HT-Seq's "union" mode (Anders et al. 2015) and the squirrel gene annotation file (Ictidomys_tridecemlineatus.spetri2.84.gtf). All statistical analysis hereafter was carried out using R version 3.3.1 (R Core Team 2017). We filtered this dataset to only include protein-coding genes with at least 10 reads in 75% of libraries. We transformed these count data for linear modeling while normalizing according to sequencing depth, gene length, and mean variance across genes using the "voom" function within the LIMMA package (Law et al. 2014; Ritchie et al. 2015). We then built a Euclidean distance-based network of samples using the "adjacency" function in WGCNA (Langfelder and Horvath 2008). Samples were designated as outliers and removed from analyses if their connectivity was more than 3 standard deviations (SD) from the mean (as described by Horvath 2011).

Removal of technical variation

Nonbiological sources of variation, or "batch effects," are ubiquitous in high-throughput genetic studies (Leek et al. 2010). To protect against this bias, we used principal component analysis of the normalized, VOOM transformed expression data to quantify the technically derived variance. Five batch effects significantly influenced the principal components (PCs) of gene expression: sequence platform (HiSeq 2000 vs. 4000; correlated with PC 1 [Pearson correlation $r = -0.74$, $P = 7.435e\text{-}09$]), sequencing lane (samples distributed across 10 lanes; correlated with PC 1 [$r = -0.65$, $P = 1.494e\text{-}06$] and PC 5 [$r = -0.39$, $P = 0.0078$]), RNA extraction date (performed on 6 different days; correlated with PC 1 [$r = -0.53$, $P = 0.0002$] and PC 7 [$r = 0.44$, $P = 0.0027$]), the number of uniquely mapped reads (correlated with PC 1 [$r = -0.51$, $P = 0.0004$]), and RIN (correlated with PC 3 [$r = -0.39$, $P = 0.0079$]). When using a reference genome, a significant number of reads can map to multiple locations due to repetitive sequences or paralogous genes (Li et al. 2010; Conesa et al. 2016). Although the percentage of uniquely mapped reads in our samples (mean of 62 ± 8 SD) was within the range observed in previous studies (e.g., Xiong et al. 2012; Tung et al. 2015; Fraser et al. 2018), we analyzed this potential batch effect to increase our power to detect a true biological signal. To remove technical variation, we iteratively regressed out batch effects in descending order of correlation with PCs, beginning with the sequence platform. That is, we first removed variation associated with the batch effect with the most significant correlation with a PC (sequencing platform). We then re-evaluated the remaining batch effects and again removed variance for the single effect with the highest correlation. We performed this iterative regression until no significant batch effects remained. Cellular composition is also known to influence gene expression in blood samples (Palmer et al. 2006). We controlled for this effect by quantifying cell types in blood smears for all RNA samples (detailed methods described in Lopez et al. 2013) and regressed the effect of the cell type with the strongest signal to a PC (neutrophils;

correlated with PC 4 [$r = -0.42$, $P = 0.0039$]). Samples were balanced across sequencing lanes according to colony, disperser status, and sampling date.

Genome-wide discovery analysis

To identify individual genes that were significantly associated with the dispersal phenotype, we created linear mixed models in EMMREML (Akdemir and Godfrey 2015). Dispersal status (disperser = 1, resident = 0), day of the year, and time of day of blood collection were included as fixed independent effects (to account for seasonal and circadian variation, respectively). To control for heritability of gene expression, we included a random effect of kinship (Wright et al. 2014; Tung et al. 2015). Dependent variables were the residuals of the gene expression counts after filtering, normalizing, and regressing batch effect variance. For each gene model, we extracted the *P*-value associated with dispersal and adjusted for multiple hypothesis testing using a false discovery rate approach (q; Storey and Tibshirani 2003) implemented in QVALUE (Storey et al. 2015). A gene was considered significantly associated with dispersal if $q < 0.1$.

Gene network analysis

We identified coordinated gene expression patterns in dispersers using a weighted gene coexpression network analysis (WGCNA; Zhang and Horvath 2005; Langfelder and Horvath 2008). We performed a signed network analysis on gene expression residuals with a soft thresholding power of 14 (to optimize scale-free topology; Zhang and Horvath 2005), cut height of 0.3, and minimum module size of 50. For each module, we calculated the first PC, or module eigengene (ME) of gene expression, which is the representative expression profile for that module (Oldham et al. 2008). We considered a module associated with dispersal if its ME was significantly correlated with the trait ($P < 0.05$).

Gene ontology analysis

To identify which biological functions were significantly over-represented in genes associated with dispersal, we performed a gene ontology (GO) analysis on 3 resulting gene lists: 1) genes identified as up-regulated in dispersers using mixed models; 2) genes found to be down-regulated in mixed models; and 3) genes in network modules significantly associated with dispersal. Using the ground squirrel genome as a reference, we acquired HGNC (HUGO Gene Nomenclature Committee; Gray et al. 2015) gene symbol information using BiomaRt (Smedley et al. 2015). We identified the enriched biological processes using gProfileR (Reimand et al. 2016). Queries were the 3 dispersal-related gene lists, background lists included all detectably expressed protein-coding genes, and we set the minimum functional category and intersection sizes to 5. We corrected for multiple testing using the "gSCS" method.

Temporal expression patterns in dispersing marmots

One expectation is that if DE genes are beneficial for dispersal, their expression would be related to the timing of the actual dispersal event. Consequently, we tested whether the transcription of genes associated with dispersal varied as a function of the number of days between sampling and dispersal. As previously mentioned, the optimal study design would have analyzed longitudinal samples from dispersers. However, due to limited opportunities to sample dispersers repeatedly, we focused on single RNA samples from each disperser ($n = 16$) and calculated the number of days between sample collection and the dispersal date. Observed days until dispersal ranged from 0 to 54 days; that is, some dispersers were sampled on the day they dispersed whereas others were sampled up to 54 days prior to dispersing. We then fitted linear mixed models in EMMREML with days until dispersal as a fixed predictor variable and kinship as a random predictor variable. Again, dependent variables were gene expression residuals and q was set to 0.1, but here we only fitted models for genes determined to be significantly associated with dispersal in the initial mixed model or network analyses.

RESULTS

RNA-seq samples

We sequenced high-quality mRNA from 43 individual female yearlings with known dispersal status ($n = 16$ dispersers, $n = 27$ residents). Samples from animals known to have died ($n = 2$) or that returned after a prolonged absence ($n = 1$) were excluded from all analyses. Dispersers were sampled between 19 May and 3 July; residents were sampled between 3 June and 10 July across the 3 years. On average, we generated 29 million reads per individual and 62% (18 million) uniquely mapped to the squirrel genome with sufficient quality and length. Of the 22,389 protein-coding genes in the squirrel genome, 11,381 (50.8%) were detectably expressed (≥10 reads in 75% of libraries) and used in subsequent analyses. Clustering analysis revealed no outlier samples.

Genome-wide discovery analysis

Mixed models identified 150 DE dispersal-related genes ($q < 0.1$; Supplementary Table S1). Eighty-one percent (122) had positive ($\log_2$) fold changes, indicating up-regulation, whereas 19% (28) had negative fold changes (down-regulated) in dispersers compared to residents (Figure 1). The mean $\log_2$ fold change of the 150 significant genes was 0.013 (± 0.27 SD). Genes with significantly large fold changes (greater than 3 SD from the mean; −0.81 and 0.81)

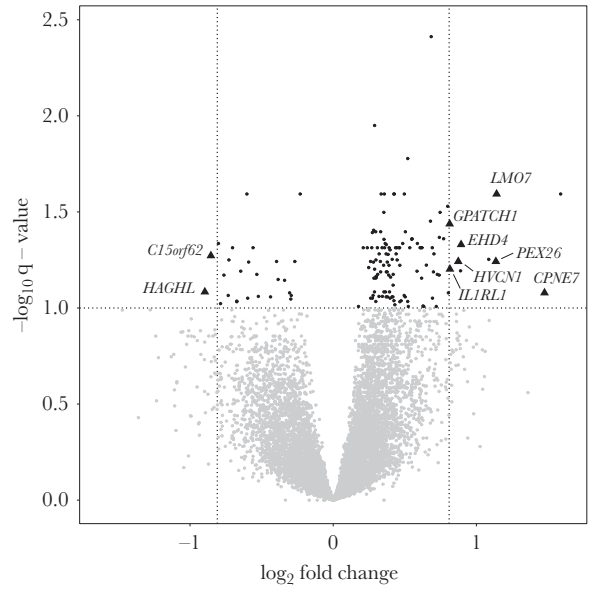

Figure 1
Volcano plot showing fold change in gene expression against q-value for dispersing marmots compared to residents for 11,381 genes. Genes with large fold changes and known homologs are highlighted.

are highlighted in Figure 1. Genes with large fold changes included metabolic genes *CPNE7* and *EHD4* and proinflammatory genes *IL1RL1* and *HVCN1*.

Gene network analysis

WGCNA identified 18 coexpressed networks containing 50 to 2784 genes per module. Of these, one module of 126 genes (designated the "salmon" module by WGCNA; Supplementary Table S2) was significantly and positively correlated with dispersal (correlation = 0.36, $P = 0.02$), indicating dispersers up-regulated this module of genes compared to residents. Many genes in this module are known to regulate inflammation and immune responses to antigen (e.g., cluster of differentiation [CD] and human leukocyte antigen genes [HLA]) and they show a clear trend of coexpression and up-regulation in dispersers compared to residents (Figure 2). This module included several characteristic markers of activated B lymphocytes, including *CD19* (the canonical cell surface marker of B cells), *CD74*, *BLK*, and MHC class II molecules (*HLA-D/DR*). To further highlight the most important genes within this network, we identified the highly-connected nodes, or "hub genes." Intramodular hub genes are often highly associated with a trait of interest (Zhang and Horvath 2005; Geschwind and Konopka 2009) and have important roles within a network (Langfelder et al. 2013). We conservatively defined hub genes as those with a module membership > 0.8 and a correlation with dispersal > 0.3, resulting in 29 hub genes (Supplementary Table S3). Hub genes were largely associated with immune system response (e.g., *MS4A1*, *HLA-DOB*, *CD79B*) and are particularly characteristic of activated B lymphocytes.

GO analysis

GO analysis of the 122 up-regulated genes revealed categorical enrichment of 33 biological processes (Supplementary Table S4). These enriched categories were primarily composed of metabolic functions (e.g., "primary metabolic process," "organic substance metabolic process") and transcription regulatory processes (e.g., "RNA binding," "regulation of gene expression"). There was no categorical enrichment of biological processes in the 28 down-regulated genes.

The 126 salmon module genes were statistically enriched for immunological functions including "B cell receptor signaling pathway," "antigen processing and presentation of peptide antigen," and "MHC class II protein complex binding" (Supplementary Table S1). Genes involved in these pathways included orthologues of several human leukocyte antigens (*HLA-DMA, -DOB, -DPB1, -DRA*) and cluster of differentiation genes characterizing B lymphocytes (*CD19, CD74, CD79A/B*). The overwhelming enrichment of genes associated with antigen defense, and specifically MHC class II processes, confirmed our hypothesis that relative to residents, less socially integrated dispersers activated inflammatory and immunological pathways that protect them from bacterial pathogens.

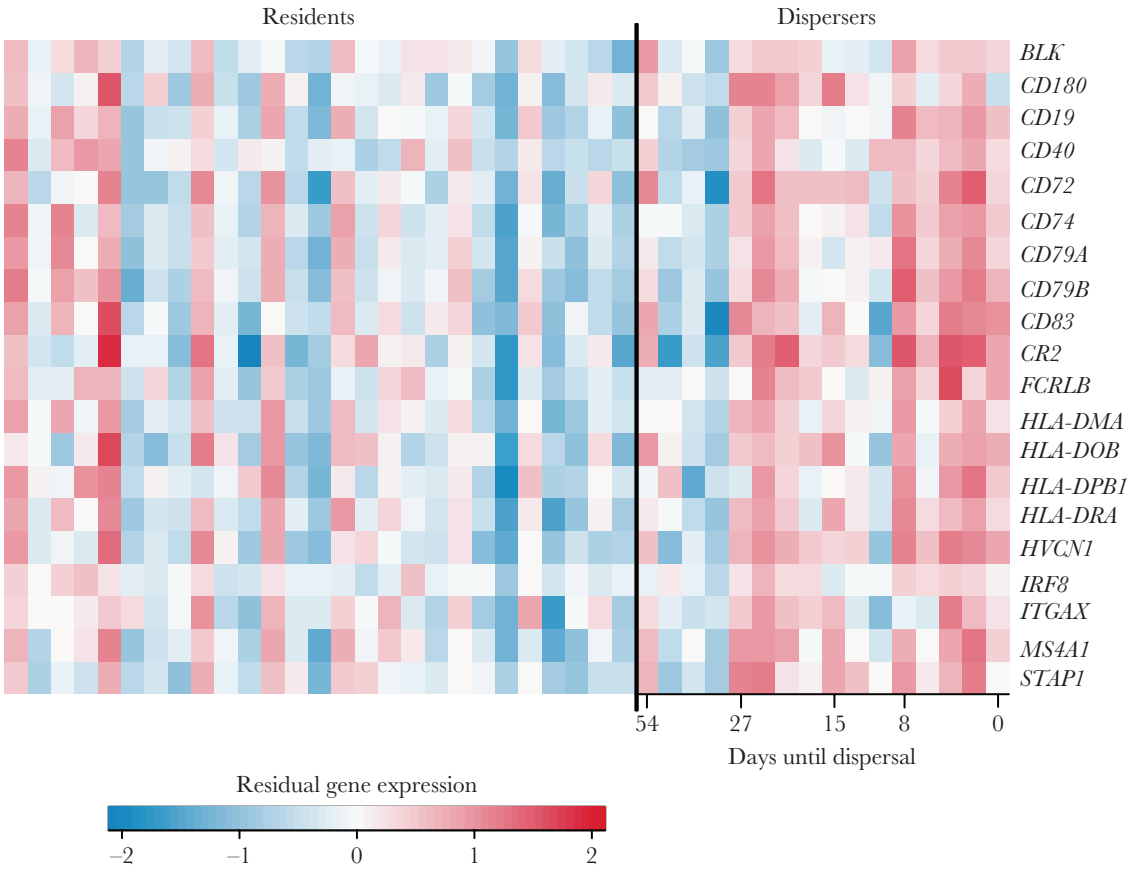

Figure 2
Heat map illustrating gene expression levels of "salmon" module antigen defense genes. Rows represent genes; columns represent individual marmots, grouped by dispersal phenotype.

Temporal expression patterns in dispersing marmots

The number of days between RNA sampling and dispersal date was normally distributed across the 16 dispersers (range = 0–54 days; mean = 19.6; SD = 15.0; Shapiro-Wilk $P = 0.38$). We combined the dispersal-associated genes (150 DE genes and 126 network genes with 16 genes identified by both approaches; Table 1) and tested whether there was a significant fold change ($q < 0.1$) as a function of the number of days until dispersal in these 260 genes. Twenty dispersal-associated genes significantly increased expression as the date of dispersal approached, whereas 2 decreased expression (Figure 3; Supplementary Table S5).

DISCUSSION

Dispersal is a complex behavior that is influenced by multiple factors, including genotype, body condition and social and environmental pressures. We found that prior to dispersal from the natal colony, yellow-bellied marmots activated gene expression in circulating blood cells across numerous pathways and somatic processes. These results echo much of what is known about the genetics of migration and invertebrate dispersal, suggesting that many molecular processes involved in these behaviors are conserved across taxa. Specifically, genome-wide discovery analysis confirmed our a priori hypothesis that genes important for lipid and glucose metabolism would be up-regulated by dispersers. We did not observe evidence of altered circadian or circannual processes prior to dispersal; however, dispersal is not a seasonal or repeated behavior. This suggests that circadian physiological shifts that occur during migration may not apply for a single, undirected dispersal event. As predicted, up-regulated genes were statistically enriched for several metabolic processes and we detected an unexpected enrichment of nucleic acid transcription and regulation. These signals may stem in part from the marked transcriptional induction required for immunological activation (i.e., consistent with the activated B cell signature observed below). This over-representation of transcription processes suggests that protein synthesis shifts are also critical in the preparation to disperse. These data provide a molecular framework for understanding the biology of dispersal.

Our 2 approaches produced complimentary and overlapping results. Specifically, 16 individual genes were found to be significantly related to dispersal in both analyses (Table 1), including several genes that were highly up-regulated in linear models, such as *CPNE7*, *LMO7*, and *HVCN1* (Figure 1) and genes that were hubs of modules related to dispersal *MS4A1*, *STAP1*, and *CNN3*. Consistent with the activated B cell signature seen in mixed model results, genes in the salmon module were categorically enriched for immune system processes, specifically those related to antigen processing (likely reflecting the marked *HLA-D/DR* activation molecule signature). Taken together, the 2 methods illustrate that numerous processes are involved in preparing animals to disperse from the natal site.

Metabolism

The genes that dispersers up-regulated were categorically enriched for metabolic processes, as predicted. Although blood is frequently used to detect metabolic conditions in standard comprehensive blood panels, it is not a primary metabolic tissue (Berg et al. 2002), so this enrichment is correlative. It would be interesting to evaluate metabolic gene expression in liver or kidney of dispersers in systems where these tissues are available, as one would expect an even more pronounced signal in these cells. The gene with the largest fold change in dispersers was *CPNE7* (Figure 1), which is associated with lipid metabolism and food deprivation in chickens (Désert et al. 2008). Interestingly, a related gene, *CPNE4*, was also found to be associated with migration in birds (Jones et al. 2008; Ruegg et al. 2014), though these studies hypothesized this gene was associated with migratory restlessness and not metabolism. *EHD4*, a highly-conserved gene known to induce ATP hydrolysis and carbohydrate derivative binding in many taxa (Pohl et al. 2000; Naslavsky and Caplan 2011), is also strongly up-regulated by dispersers. *GRK5* is implicated in many aspects of physiology, particularly fat uptake and weight gain and a *GRK5* deficiency impairs lipid metabolism (Wang et al. 2012). Similarly, *MAP3K* signaling (including *MAP3K4*)

Table 1

Genes found to be significantly associated with dispersal in both mixed model and gene network analyses

Ensembl ID	Gene symbol	Gene description	Log_2 fold change	*q*-value	WGCNA correlation	*P*-value
ENSSTOG00000019772	*CPNE7*	Copine 7	1.48	0.083	0.35	0.020
ENSSTOG00000013061	*LMO7*	LIM domain 7	1.14	0.026	0.45	0.002
ENSSTOG00000009429	*HVCN1*	Hydrogen voltage gated channel 1	0.81	0.063	0.38	0.012
ENSSTOG00000003129	*DOCK9*	Dedicator of cytokinesis 9	0.80	0.083	0.41	0.007
ENSSTOG00000013018	*STAP1*	Signal transducing adaptor family member 1	0.74	0.043	0.43	0.004
ENSSTOG00000006747	*MS4A1*	Membrane spanning 4-domains A1	0.72	0.098	0.39	0.009
ENSSTOG00000004669	*FCRLB*	Fc receptor like B	0.70	0.065	0.39	0.009
ENSSTOG00000014061	*CNN3*	Calponin 3	0.69	0.089	0.35	0.020
ENSSTOG00000005052	*RALGPS2*	Ral GEF with PH domain and SH3 binding motif 2	0.69	0.049	0.41	0.007
ENSSTOG00000001358	*MCM3*	Minichromosome maintenance complex component 3	0.68	0.004	0.46	0.002
ENSSTOG00000024103	*ZBTB8A*	Zinc finger and BTB domain containing 8A	0.62	0.049	0.41	0.006
ENSSTOG00000016000	*PARP1*	Poly(ADP-ribose) polymerase 1	0.55	0.044	0.40	0.008
ENSSTOG00000007519	*PARN*	Poly(A)-specific ribonuclease	0.52	0.017	0.42	0.005
ENSSTOG00000004790	*NA*	NA	0.50	0.040	0.45	0.002
ENSSTOG00000021692	*NA*	NA	0.37	0.088	0.34	0.024
ENSSTOG00000002553	*GRK5*	G protein-coupled receptor kinase 5	0.37	0.052	0.37	0.015

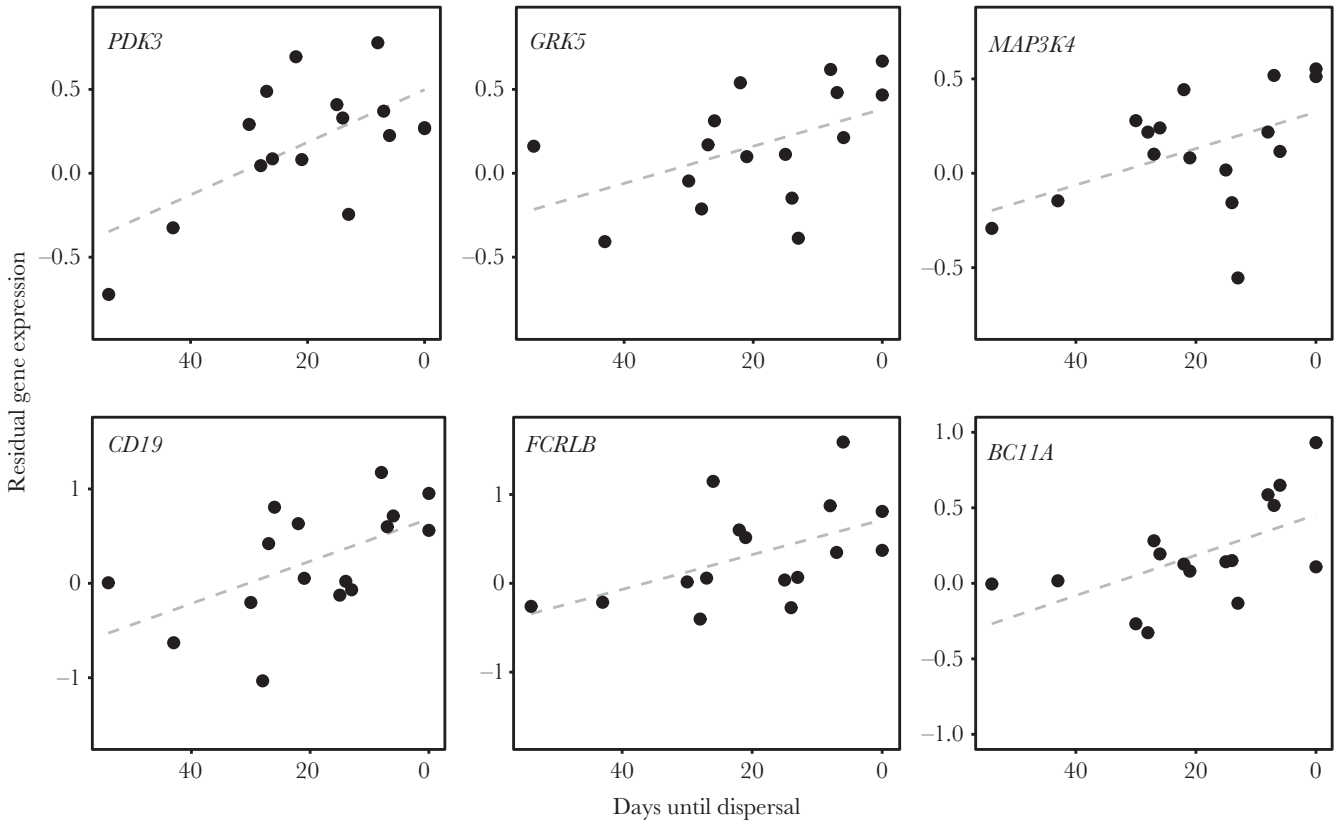

Figure 3
Examples of genes that significantly increase expression relative to the number of days until dispersal for dispersing marmots (n = 16). PDK3, GRK5, and MAP3K4 are associated with metabolism. CD19, FCRLB, and BC11A are involved in antigen defense.

Table 2
GO enrichment of genes in module which WGCNA calls the "salmon" module

Domain	GO term	Genes involved	*P*-value
Biological processes	B cell receptor signaling pathway	*BLK, CD79A, CD79B, STAP1, CD19*	0.04060
	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	*HLA-DOB, HLA-DPB1, CD74, HLA-DMA, HLA-DRA*	0.00075
	antigen processing and presentation of peptide antigen	*HLA-DOB, HLA-DPB1, CD74, HLA-DMA, HLA-DRA*	0.01400
	antigen processing and presentation of peptide antigen via MHC class II	*HLA-DOB, HLA-DPB1, CD74, HLA-DMA, HLA-DRA*	0.00047
Cellular components	plasma membrane protein complex	*ITGAX, HLA-DOB, CD79A, HLA-DPB1, CD79B, CD74, HLA-DMA, HLA-DRA*	0.00582
	MHC protein complex	*HLA-DOB, HLA-DPB1, CD74, HLA-DMA, HLA-DRA*	0.00003
	MHC class II protein complex	*HLA-DOB, HLA-DPB1, CD74, HLA-DMA, HLA-DRA*	0.00001
	receptor complex	*FGFR1, ITGAX, CD79A, CD79B, CD74, CR2, DIABLO, CD40, ERBB3*	0.02440
Molecular functions	antigen binding	*HLA-DOB, HLA-DPB1, MS4A1, CD74, HLA-DMA, HLA-DRA, ATP1B1*	0.00001
	MHC protein complex binding	*HLA-DOB, MS4A1, CD74, HLA-DMA, HLA-DRA, ATP1B1*	0.00001
	MHC class II protein complex binding	*HLA-DOB, MS4A1, CD74, HLA-DMA, HLA-DRA, ATP1B1*	0.00001

is important for growth factors, lipid metabolism, and adipogenesis and has a vital role in energy homeostasis in many mammals (Sale et al. 1995; Zhang et al. 2016). Since dispersers (and animals that migrate) travel long distances in unfamiliar territory, they likely experience unpredictable food sources for extended periods of time. Thus, transcriptional regulation of genes that control lipid metabolism and fat uptake during these life-history phases likely helps conserve the energy required for both long distance travel and food deprivation.

Immune system response

Network analysis revealed coordinated up-regulation of genes that protect the body from extracellular antigens and bacteria, specifically, major histocompatibility complex (MHC) class II genes (Table 2). Moreover, many of the DE genes identified by linear models were specifically characteristic of activated B lymphocytes, including the canonical B cell marker *CD19* and the activation of related molecules *CD74*, *CD79A/B*, *CD40*, *CD83*, and MHC

class II (*HLA-D/DR*). MHC is a cluster of genes found in vertebrates that encodes a complex of receptors expressed on the surface of all somatic cells (Klein 1986). These cell surface receptors bind peptides derived from antigens and present them to immune cells to initiate the antigenic response. Whereas MHC class I genes bind intracellular peptides, MHC II deals with extracellular foreign peptides such as bacteria and presents them to specialized immune cells that distinguish between harmful and neutral antigens (Todd et al. 1988). In conjunction with the up-regulation of multiple B cell activation markers, the observed up-regulation of MHC II molecules suggests that the adaptive immune system may selectively stimulate the antibody-producing B cell subpopulation in advance of the potential microbial challenges associated with dispersal.

Nonetheless, the observed B cell inflammatory/immune activation response may not be entirely due to preparation for dispersal. Rather, it may be stimulated in part by differences in the social environment. Dispersing female marmots engage in fewer affiliative interactions with conspecifics (Armitage et al. 2011) and are less embedded in their social network than residents (Blumstein et al. 2009). Mounting evidence suggests that the social environment alone can regulate major shifts in immunological gene expression profiles (Robinson et al. 2005; Cole 2014), which may explain the observed canonical inflammatory repertoire in less social dispersers compared to more social residents.

We observed up-regulation of antigen processing genes within less social dispersers. However, certain individuals clearly do not fit these patterns (Figure 2). Some dispersers appear to down-regulate these genes, whereas some residents up-regulate them. Thus, it is also possible that this proinflammatory signal results directly from an individual's exposure to extracellular antigens, leading to increased MHC II activity. Future studies examining gene expression and vertebrate dispersal might incorporate additional immune assays to help clarify the immune system's relationship to this phenotype.

Of the genes we found to be associated with the dispersal phenotype, 22 significantly changed expression as the timing of dispersal neared. Importantly, all genes that increased expression during this time were consistently up-regulated in dispersers compared to residents in the initial models and the 2 that decreased expression were down-regulated in dispersers. In addition, many of these temporal genes are related to functions we predicted would be important for dispersal, including metabolic *PDK3*, *GRK5*, and *MAP3K4* and immunological *CD19*, *FCRLB*, and *BC11A* (Figure 3).

We observe that there is a general pattern of increasing immune gene expression as a function of days until dispersal (Figure 2). The individuals sampled within one week of dispersing (columns on the far right) are largely up-regulated (warmer colors), whereas those sampled well before dispersing display cooler expression profiles that are more similar to resident profiles. All DE genes likely aid in successful dispersal; however, these temporal results suggest that physiological preparation does not occur at a constant rate for all genes. Depending on function and regulatory mechanisms, some genes may ramp up expression to synthesize proteins early on whereas others may change expression immediately before dispersal. This temporal trend is consistent with immune expression driven by individual anticipation of the dispersal event. However, social tension may also be increasing toward the dispersal date and in part, drive the final dispersal decision. Thus, stress from an unstable social environment may help prime the immune system to cope with dispersal in this species. Although it is out of the scope of this study, future research may help disentangle the temporal effects of dispersal and the social environment on gene expression.

Finally, we acknowledge these results should be interpreted with caution for 2 reasons. First, gene expression is notoriously difficult to interpret in samples taken from free-living individuals due to environmental heterogeneity. We have taken all available statistical precautions to identify and control for potentially confounding variables (e.g., technical batch effects, cellular composition, seasonal and diurnal fluctuations, age). However, there certainly may be unrecorded factors (such as health variables and exposure to pathogens) that affect gene expression and add noise to the associations estimated here. The statistical impact of that added noise would be to decrease statistical significance of any observed association with dispersal. As such, the present results may underestimate the true range of transcriptomic alterations associated with dispersal, but heterogeneity per se would not produce spurious statistical significance and lead to false positives in the associations reported here. We advise future expression studies in natural populations to record and control for as many potentially confounding variables as possible. Second, these results are based on blood samples. Circulating blood is composed of numerous cell types that have different functional roles and express genes at different rates (Janeway et al. 2005). Of the cell types we measured, we found that the proportion of neutrophils explained significant variation in overall gene expression and so we regressed out this effect. Collecting samples at our isolated field site prevented us from assessing compositions of antigen presenting cells through flow cytometry (e.g., dendritic cells, macrophages, and B cells). Thus, cell populations may have shifted prior to dispersal, leading to the observed increased expression of MHC genes. Nevertheless, whether these results are due to fluctuations in cellular composition or simply gene expression changes in the existing cell repertoire, this is the first time a physiological response of this scale has been documented in dispersers.

Our RNA-seq analysis of yellow-bellied marmots provides novel insights about gene expression preparation for dispersal. To our knowledge, we are the first to document expression differences between dispersing and nondispersing vertebrates. Nonetheless, our results are consistent with previous studies of dispersing insects and species that migrate where genes involved in metabolism are typically up-regulated. We have also demonstrated a substantial immunological activation of the B cell component of the adaptive immune system that occurs prior to dispersal and may be regulated in part by the social environment. These findings suggest that physiological mechanisms are evolutionarily conserved across highly divergent taxa that engage in dispersal and migratory behavior. Dispersal incurs many risks and our findings support the idea that some of the associated costs can be reduced through physiological preparation (Clobert et al. 2009). Indeed, dispersing marmots experience a negligible decrease in survival rates compared to residents (0.73 vs. 0.87; Van Vuren and Armitage 1994). Thus, these gene regulatory profiles may conceivably represent an evolutionarily conserved anticipatory defense mechanism to reduce the risks associated with this major life-history event.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *Behavioral Ecology* online.

FUNDING

This work was supported by the National Science Foundation (grant numbers DGE-1144087, DEB-1119660, DEB-1557130); the National Institutes

of Health (grant numbers NIH S10 OD018174, P30 AG017265); and the University of California Los Angeles.

We thank the many marmoteers who assisted with fieldwork. Special thanks to the UCLA Statistical Consulting Group, Amanda Lea, and Rachel Johnston for statistical advice. We also thank the Blumstein lab, the Wayne lab, and 2 anonymous reviewers for insightful comments.

Ethics statement: Marmots were studied under UCLA research protocol ARC 2001-191-01 and Colorado Division of Wildlife permit 12TR917.

Data accessibility: Analyses reported in this article can be reproduced using the data provided by Armenta et al. (2018). Raw sequencing data, regressed normalized expression counts, and all associated metadata are available in the NCBI Gene Expression Omnibus repository (Edgar et al. 2002) GEO: GSE113744.

Handling editor: Luke Holman

REFERENCES

Akdemir D, Godfrey OU. 2015. EMMREML: fitting mixed models with known covariance structures. http://CRAN.R-project.org/package=EMMREML (accessed 5 December 2018).

Alberts SC, Altmann J. 1995. Balancing costs and opportunities: dispersal in male baboons. Am Nat. 145:279–306.

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

Armenta TC, Cole SW, Geschwind DH, Blumstein DT, Wayne RK. 2018. Data from: gene expression shifts in yellow-bellied marmots prior to natal dispersal. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.6fb8d37

Armitage KB. 1962. Social behaviour of a colony of the yellow-bellied marmot (*Marmota flaviventris*). Anim Behav. 10:319–331.

Armitage KB. 1991. Social and population dynamics of yellow-bellied marmots: results from long-term research. Annu Rev Ecol Evol Syst. 22:379–407.

Armitage KB. 2014. Marmot biology: sociality, individual fitness, and population dynamics. Cambridge: Cambridge University Press.

Armitage KB, Van Vuren DH, Ozgul A, Oli MK. 2011. Proximate causes of natal dispersal in female yellow-bellied marmots, *Marmota flaviventris*. Ecology. 92:218–227.

Avitsur R, Hunzeker J, Sheridan JF. 2006. Role of early stress in the individual differences in host response to viral infection. Brain Behav Immun. 20:339–348.

Barbraud C, Johnson AR, Bertault G. 2003. Phenotypic correlates of post-fledging dispersal in a population of greater flamingos: the importance of body condition. J Anim Ecol. 72:246–257.

Bekoff M. 1977. Mammalian dispersal and the ontogeny of individual behavioral phenotypes. Am Nat. 111:715–732.

Berg J, Tymoczko J, Stryer L. 2002. Each organ has a unique metabolic profile. In: Biochemistry. 5th ed. New York, NY: W. H. Freeman.

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. Nature. 446:507–512.

Blem CR. 1980. The energetics of migration. In: Sidney A, Gauthreaux J, editor. Animal migration, orientation, and navigation. New York: Academic Press. p. 175–224.

Blumstein DT. 2013. Yellow-bellied marmots: insights from an emergent view of sociality. Philos Trans R Soc Lond B Biol Sci. 368:20120349.

Blumstein DT, Im S, Nicodemus A, Zugmeyer C. 2004. Yellow-bellied marmots *(Marmota flaviventris)* hibernate socially. J Mammal. 85:25–29.

Blumstein DT, Wey TW, Tang K. 2009. A test of the social cohesion hypothesis: interactive female marmots remain at home. Proc R Soc Lond B Biol Sci. 276:3007–3012.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, Lehouck V, Matthysen E, Mustin K, Saastamoinen M, et al. 2012. Costs of dispersal. Biol Rev Camb Philos Soc. 87:290–312.

Boss J, Liedvogel M, Lundberg M, Olsson P, Reischke N, Naurin S, Åkesson S, Hasselquist D, Wright A, Grahn M, et al. 2016. Gene expression in the brain of a migratory songbird during breeding and migration. Mov Ecol. 4:4.

Brisson JA, Davis GK, Stern DL. 2007. Common genome-wide patterns of transcript accumulation underlying the wing polyphenism and polymorphism in the pea aphid (*Acyrthosiphon pisum*). Evol Dev. 9:338–346.

Chepko-Sade BD, Halpin ZT. 1987. Mammalian dispersal patterns: the effects of social structure on population genetics. Chicago: University of Chicago Press.

Clobert J, Danchin E, Dhondt AA, Nichols JD, editors. 2001. Dispersal. Oxford: Oxford University Press.

Clobert J, Le Galliard JF, Cote J, Meylan S, Massot M. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecol Lett. 12:197–209.

Clutton-Brock TH, Lukas D. 2012. The evolution of social philopatry and dispersal in female mammals. Mol Ecol. 21:472–492.

Cole SW. 2014. Human social genomics. PLoS Genet. 10:e1004601.

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 USA. 109:20578–20583.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, et al. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17:13. doi:10.1186/s13059-016-0881-8. http://genomebiology.com/2016/17/1/13 (accessed 19 October 2018).

Cote J, Clobert J. 2007. Social personalities influence natal dispersal in a lizard. Proc R Soc Lond B Biol Sci. 274:383–390.

Cote J, Clobert J, Brodin T, Fogarty S, Sih A. 2010. Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. Philos Trans R Soc Lond B Biol Sci. 365:4065–4076.

Désert C, Duclos MJ, Blavy P, Lecerf F, Moreews F, Klopp C, Aubry M, Herault F, Le Roy P, Berri C, et al. 2008. Transcriptome profiling of the feeding-to-fasting transition in chicken liver. BMC Genomics. 9:611.

Dingemanse NJ, Both C, van Noordwijk AJ, Rutten AL, Drent PJ. 2003. Natal dispersal and personalities in great tits (*Parus major*). Proc R Soc Lond B Biol Sci. 270:741–747.

Doligez B, Pärt T. 2008. Estimating fitness consequences of dispersal: a road to 'know-where'? Non-random dispersal and the underestimation of dispersers' fitness. J Anim Ecol. 77:1199–1211.

Duckworth RA, Badyaev AV. 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. Proc Natl Acad Sci USA. 104:15017–15022.

Dufty AM, Belthoff JR. 2001. Proximate mechanisms of natal dispersal: the role of body condition and hormones. In: Clobert J, Danchin E, Dhondt AA, Nichols JD, editors. Dispersal. Oxford: Oxford University Press. p. 223–233.

Edgar R, Domrachev M, Lash AE. 2002. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30:207–210.

Ekernas LS, Cords M. 2007. Social and environmental factors influencing natal dispersal in blue monkeys, *Cercopithecus mitis stuhlmanni*. Anim Behav. 73:1009–1020.

Fraser D, Mouton A, Serieys LEK, Cole S, Carver S, Vandewoude S, Lappin M, Riley SPD, Wayne R. 2018. Genome-wide expression reveals multiple systemic effects associated with detection of anticoagulant poisons in bobcats (*Lynx rufus*). Mol Ecol. 27:1170–1187.

Fredrickson BL, Grewen KM, Algoe SB, Firestine AM, Arevalo JMG, Ma J, Cole SW. 2015. Psychological well-being and the human conserved transcriptional response to adversity. PLoS One. 10:e0121839.

Fudickar AM, Peterson MP, Greives TJ, Atwell JW, Bridge ES, Ketterson ED. 2016. Differential gene expression in seasonal sympatry: mechanisms involved in diverging life histories. Biol Lett. 12:20160069.

Gallego Romero I, Pai AA, Tung J, Gilad Y. 2014. RNA-seq: impact of RNA degradation on transcript quantification. BMC Biol. 12:42.

Geschwind DH, Konopka G. 2009. Neuroscience in the era of functional genomics and systems biology. Nature. 461:908–915.

Gese EM, Ruff RL, Crabtree RL. 1996. Social and nutritional factors influencing the dispersal of resident coyotes. Anim Behav. 52:1025–1043.

Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. 2015. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 43:D1079–D1085.

Greenwood PJ. 1980. Mating systems, philopatry and dispersal in birds and mammals. Anim Behav. 28:1140–1162.

Horvath S, editor. 2011. Weighted network analysis: applications in genomics and systems biology. New York: Springer.

Janeway CA, Travers P, Walport M, Shlomchik MJ. 2005. Immunobiology: the immune system in health and disease. New York: Garland Science Publishing.

Jeffs AG, Willmott ME, Wells RMG. 1999. The use of energy stores in the puerulus of the spiny lobster *Jasus edwardsii* across the continental shelf of New Zealand. Comp Biochem Physiol. 123:351–357.
Jenni L, Jenni-Eiermann S. 1998. Fuel supply and metabolic constraints in migrating birds. J Avian Biol. 29:521–528.
Johnston RA, Paxton KL, Moore FR, Wayne RK, Smith TB. 2016. Seasonal gene expression in a migratory songbird. Mol Ecol. 25: 5680–5691.
Jones CM, Papanicolaou A, Mironidis GK, Vontas J, Yang Y, Lim KS, Oakeshott JG, Bass C, Chapman JW. 2015. Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. Mol Ecol. 24:4901–4911.
Jones S, Pfister-Genskow M, Cirelli C, Benca RM. 2008. Changes in brain gene expression during migration in the white-crowned sparrow. Brain Res Bull. 76:536–544.
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.
Klein J. 1986. Natural history of the major histocompatibility complex. New York: Wiley.
Kohane IS, Valtchinov VI. 2012. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. Bioinformatics. 28:538–545.
Krueger F. 2015. Trim Galore: a wrapper tool around Cutadapt and fastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed 5 December 2018).
Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 9:559.
Langfelder P, Mischel PS, Horvath S. 2013. When is hub gene selection better than standard meta-analysis? PLoS One. 8:e61505.
Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29.
Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 11:733–739.
Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 26:493–500.
Liedvogel M, Akesson S, Bensch S. 2011. The genetics of migration on the move. Trends Ecol Evol. 26:561–569.
Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. 2006. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med. 147:126–132.
Lin Y-TK, Batzli GO. 2001. The influence of habitat quality on dispersal, demography, and population dynamics of voles. Ecol Monogr. 71:245–275.
Lopez J, Wey TW, Blumstein DT. 2013. Patterns of parasite prevalence and individual infection in yellow-bellied marmots. J Zool. 291:296–303.
Matthysen E. 2005. Density-dependent dispersal in birds and mammals. Ecography. 28:403–416.
Meylan S, Belliure J, Clobert J, de Fraipont M. 2002. Stress and body condition as prenatal and postnatal determinants of dispersal in the common lizard (*Lacerta vivipara*). Horm Behav. 42:319–326.
Naslavsky N, Caplan S. 2011. EHD proteins: key conductors of endocytic transport. Trends Cell Biol. 21:122–131.
Nilsson J, Smith HG. 1985. Early fledgling mortality and the timing of juvenile dispersal in the marsh tit *Parus palustris*. Ornis Scandinavica. 16:293–298.
Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. 2008. Functional organization of the transcriptome in human brain. Nat Neurosci. 11:1271–1282.
Palmer C, Diehn M, Alizadeh AA, Brown PO. 2006. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics. 7:115.
Parikshak NN, Gandal MJ, Geschwind DH. 2015. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet. 16:441–458.
Pohl U, Smith JS, Tachibana I, Ueki K, Lee HK, Ramaswamy S, Wu Q, Mohrenweiser HW, Jenkins RB, Louis DN. 2000. EHD2, EHD3, and EHD4 encode novel members of a highly conserved family of EH domain-containing proteins. Genomics. 63:255–262.
Pope TR. 2000. The evolution of male philopatry in neotropical monkeys. In: Kappeler PM, editor. Primate males: causes and consequences of variation in group composition. Cambridge: Cambridge University Press. p. 219.
Pruett-Jones SG, Lewis MJ. 1990. Sex ratio and habitat limitation promote delayed dispersal in superb fairy-wrens. Nature. 348:541–542.
R Core Team. 2017. R: a language and environment for statistical computing, Vienna, Austria. https://www.R-project.org/ (accessed 5 December 2018).
Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. 2016. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 44:W83–W89.
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43:e47.
Robinson GE, Grozinger CM, Whitfield CW. 2005. Sociogenomics: social life in molecular terms. Nat Rev Genet. 6:257–270.
Rudkowska I, Raymond C, Ponton A, Jacques H, Lavigne C, Holub BJ, Marette A, Vohl MC. 2011. Validation of the use of peripheral blood mononuclear cells as surrogate model for skeletal muscle tissue in nutrigenomic studies. OMICS. 15:1–7.
Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol. 23:4757–4769.
Sale EM, Atkinson PG, Sale GJ. 1995. Requirement of MAP kinase for differentiation of fibroblasts to adipocytes, for insulin activation of p90 S6 kinase and for insulin or serum stimulation of DNA synthesis. EMBO J. 14:674–684.
Silverin B. 1997. The stress response and autumn dispersal behaviour in willow tits. Anim Behav. 53:451–459.
Smale L, Nunes S, Holekamp KE. 1997. Sexually dimorphic dispersal in mammals: patterns, causes, and consequences. Adv Study Behav. 26:181–251.
Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, et al. 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 43:W589–W598.
Snyder-Mackler N, Sanz J, Kohn JN, Brinkworth JF, Morrow S, Shaver AO, Grenier JC, Pique-Regi R, Johnson ZP, Wilson ME, et al. 2016. Social status alters immune regulation and response to infection in macaques. Science. 354:1041–1045.
Soria-Carrasco V, Castresana J. 2012. Diversification rates and the latitudinal gradient of diversity in mammals. Proc R Soc Lond B Biol Sci. 279:4148–4155.
Soulsbury CD, Baker PJ, Iossa G, Harris S. 2008. Fitness costs of dispersal in red foxes (*Vulpes vulpes*). Behav Ecol Sociobiol. 62:1289–1298.
Srygley RB, Lorch PD, Simpson SJ, Sword GA. 2009. Immediate protein dietary effects on movement and the generalised immunocompetence of migrating Mormon crickets *Anabrus simplex* (Orthoptera: Tettigoniidae). Ecol Entomol. 34:663–668.
Storey JD, Bass AJ, Dabney A, Robinson D. 2015. QVALUE: Q-value estimation for false discovery rate control. http://github.com/jdstorey/qvalue (accessed 5 December 2018).
Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci USA. 100:9440–9445.
Sullivan PF, Fan C, Perou CM. 2006. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet. 141B:261–268.
Thomas WK, Martin SL. 1993. A recent origin of marmots. Mol Phylogenet Evol. 2:330–336.
Todd JA, Acha-Orbea H, Bell JI, Chao N, Fronek Z, Jacob CO, McDermott M, Sinha AA, Timmerman L, Steinman L. 1988. A molecular basis for MHC class II–associated autoimmunity. Science. 240:1003–1009.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 25:1105–1111.
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 USA. 109:6490–6495.
Tung J, Xiang Z, Alberts SC, Stephens M, Gilad Y. 2015. The genetic architecture of gene expression levels in wild baboons. eLife. 4:e04729.
Turecki G, Meaney MJ. 2016. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. Biol Psychiatry. 79:87–96.

Van Vuren DH. 1990. Dispersal of yellow-bellied marmots [dissertation]. University of Kansas, Lawrence, Kansas. Available from: https://vanvurenlab.weebly.com/uploads/5/3/1/5/53152925/vanvuren_1990_dispersalofyellowbelliedmarmots_dissertation.pdf (accessed 5 December 2018).

Van Vuren D, Armitage KB. 1994. Survival of dispersing and philopatric yellow-bellied marmots: what is the cost of dispersal? Oikos. 69:179.

Vellichirammal NN, Zera AJ, Schilder RJ, Wehrkamp C, Riethoven JJ, Brisson JA. 2014. De novo transcriptome assembly from fat body and flight muscles transcripts to identify morph-specific gene expression profiles in *Gryllus firmus*. PLoS One. 9:e82129.

Wang F, Wang L, Shen M, Ma L. 2012. GRK5 deficiency decreases diet-induced obesity and adipogenesis. Biochem Biophys Res Commun. 421:312–317.

Weston DJ, Gunter LE, Rogers A, Wullschleger SD. 2008. Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. BMC Syst Biol. 2:16.

Woodroffe R, Macdonald DW, da Silva J. 1993. Dispersal and philopatry in the European badger, *Meles meles*. J Zool.. 237:227–239.

Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, Madar V, Jansen R, Chung W, Zhou YH, et al. 2014. Heritability and genomics of gene expression in peripheral blood. Nat Genet. 46: 430–437.

Wu Y, Patchev AV, Daniel G, Almeida OF, Spengler D. 2014. Early-life stress reduces DNA methylation of the Pomc gene in male mice. Endocrinology. 155:1751–1762.

Xiong J, Lu X, Zhou Z, Chang Y, Yuan D, Tian M, Zhou Z, Wang L, Fu C, Orias E, et al. 2012. Transcriptome analysis of the model protozoan, *Tetrahymena thermophila*, using deep RNA sequencing. PLoS One. 7:e30630.

Yoder JM, Marschall EA, Swanson DA. 2004. The cost of dispersal: predation as a function of movement and site familiarity in ruffed grouse. Behav Ecol. 15:469–476.

Zera AJ, Denno RF. 1997. Physiology and ecology of dispersal polymorphism in insects. Annu Rev Entomol. 42:207–230.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 4:Article17.

Zhang Q, Sun X, Xiao X, Zheng J, Li M, Yu M, Ping F, Wang Z, Qi C, Wang T, et al. 2016. Effects of maternal chromium restriction on the long-term programming in MAPK signaling pathway of lipid metabolism in mice. Nutrients. 8:488.

Zhu H, Gegear RJ, Casselman A, Kanginakudru S, Reppert SM. 2009. Defining behavioral and molecular differences between summer and migratory monarch butterflies. BMC Biol. 7:14.