TRANSCRIPTOME ANALYSIS OF THE HYPOTHALAMIC-PITUITARY-ADRENAL AXIS IN THE EXPERIMENTALLY DOMESTICATED FOX

BY

JESSICA PERRY HEKMAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee

Assistant Professor Anna Kukekova, Chair
Professor Jonathan Beever
Associate Professor Allison Bell
Associate Professor Lori Raetzman
Professor Lisa Stubbs

ABSTRACT

Variation in activity of the hormonal stress response, or hypothalamic-pituitary-adrenal (HPA) axis, has been associated with different personality traits and coping styles in humans and animals, while its dysregulation has been implicated in psychological disorders. The molecular basis of HPA axis regulation, however, is not yet well understood. Here, foxes selectively bred for tameness or aggression are used as a model to investigate differences in regulation of the HPA axis. Activity of this axis is markedly reduced in tame compared to aggressive foxes, with reduced levels of HPA axis hormones such as adrenocorticotrophic hormone (ACTH) and cortisol both basally and in response to a stressor. Gene expression differences were analyzed using RNA sequencing in the anterior pituitary and adrenal glands of foxes from the tame and aggressive lines, and variant analysis was performed on RNA reads from hypothalamus, anterior pituitary, and adrenal tissues from the same foxes. Pituitary analysis revealed expression differences in genes related to exocytosis and cellular signaling; adrenals analysis identified differences in similar pathways, in addition to genes related to fatty acid and cholesterol synthesis. Variant analysis also implicated cell signaling and exocytosis, as well as ion transport and DNA damage repair. These findings suggest the importance of regulation of hormone release in the control of ACTH and cortisol levels. They also suggest that metabolism of precursors to cortisol, such as fatty acids and cholesterol, may be of greater importance in HPA axis regulation than synthesis of cortisol itself. Finally, in conjunction with previous genomic findings, they suggest an association between DNA repair mechanisms and selection for tameness. These findings provide possible new lines of investigation into biological underpinnings of the phenotypic differences between the tame and aggressive lines of foxes. More broadly, as the tame foxes are considered experimentally domesticated, the findings from this project may prove applicable to HPA axis regulation differences associated with domestication in other species. Additionally, a deeper understanding of HPA axis regulation and dysregulation may be applicable both to variation in the normal population, particularly as related to behavioral traits such as coping styles, and to a number of psychiatric disorders in humans, as well as to behavioral disorders in other species, such as dogs.

To Jack

ACKNOWLEDGEMENTS

My sincerest thanks to all the members of Kukekova Laboratory for their support, particularly Anna Kukekova, Jennifer Johnson, my fellow graduate students, Halie Rando and Ali Ford, and my research assistant Naomi Won. I am also grateful to my committee members, Jonathan Beever, Allison Bell, Lori Raetzman, and Lisa Stubbs, for their guidance, as well as Sandra Rodriguez-Zas and Carolyn Thomas.

I am grateful for Dr. Lyudmila Trut's fortitude in shepherding the tame fox project for so many years, and for her kindness when I met her. Thank you to everyone at the Institute for Cytology and Genetics, especially my guides when I visited, Anastasia Kharlamova and Darya Shepeleva.

Thank you to Andy Clark and Xu Wang for their hospitality in Ithaca and their help in the early stages of this project; Michael Saul and Waqar Arif for invaluable bioinformatics advice; and Whitney Edwards and Lori Raetzman for, among many other things, the crucial insight about what was going on with the pseudopodia.

I was personally supported by the Jonathan Baldwin Turner Fellowship and by Temple Grandin; without these sources none of this would have been possible. Dr. Grandin particularly provided advice and support that I much appreciated.

Finally, many thanks to Chris, Jack, Jenny, Rosie, and Dashiell, for their (varying amounts of) patience as I worked on this project.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE

Hyper- or hypo-activity of the hypothalamic-pituitary-adrenal (HPA) axis have been associated with psychological disorders such as generalized anxiety disorder (Greaves-Lord et al., 2007; Mantella et al., 2008), major depressive disorder (Burke et al., 2005; Doane et al, 2013; Hardeveld et al., 2014), and post-traumatic stress disorder (Wessa et al., 2006; Meewisse et al., 2007). Moreover, non-pathological variation in HPA activity has been associated with different personality traits (Shoal et al., 2003; Oswald, 2006), with different coping styles in laboratory and wild animals (Koolhaas et al., 2007; Baugh et al., 2013; Clary et al., 2014), and with domesticated compared to wild animals (Künzl et al., 2003; Ericsson et al., 2014). While some genomic variation in HPA axis regulation has been described (Rosmond et al., 2001; Wagner et al., 2006; Keck et al., 2008), a full understanding of HPA axis regulation at the molecular level, and its associated cellular mechanisms, has yet to be reached. Such an understanding could contribute to better description of the mechanisms in animals and humans affecting both pathological HPA axis dysregulation, as well as normal variation in behavior, such as coping styles.

Reaching this understanding is made more difficult by the complexity of the hormonal stress response. The response to a psychological or physical stressor begins in the central nervous system and is transmitted out to organs and tissues via translation into hormones that are released into the systemic circulation. This process begins in the paraventricular nucleus (PVN) of the hypothalamus, initiated by input from higher structures such as the hippocampus, prefrontal cortex, and amygdala (Herman et al, 2003). This region of the hypothalamus releases corticotropin-releasing hormone (CRH), which travels through a portal venous system to the anterior pituitary to trigger the release of adrenocorticotropic releasing hormone (ACTH) from corticotrophic cells. CRH binds with greater affinity to a CRH type 1 receptor, which is mainly expressed in the pituitary, cerebellum, and neocortex. Deficiency of this receptor is associated with reduced anxiety-like behavior in mice (Leonard, 2005) CRH binds with lesser affinity to a type 2 receptor, which is mainly expressed in the hypothalamus, amygdala, dorsal raphe, and

hippocampus, and which appears to have opposing effects on behavior to the type 1 receptor (Leonard, 2005).

ACTH travels through the peripheral circulation where it is taken up by the adrenal glands and triggers the release of cortisol and/or corticosterone (CORT), depending on species, from the adrenal cortex (Bornstein and Chrousos, 1999). CORT release is also modulated by direct sympathetic innervation to the adrenal cortex (Bornstein and Chrousos, 1999). CORT is a steroid hormone that is bound in the bloodstream by corticosteroid binding globulin (CBG); availability of CBG affects the activity of CORT, as protein-bound CORT is not biologically available, so that reduced CBG availability is associated with increased biological activity of CORT (Sapolsky and Meaney, 1986). CORT is bound by the mineralocorticoid receptor (MR) at high affinity and the glucocorticoid receptor (GR) at low affinity (Rupprecht et al., 1993) so that the GR may mediate cellular responses to stress levels of CORT while the MR may mediate cellular responses to basal levels of CORT. When CORT elevation causes increased MR receptor occupancy, binding of CORT to the GR receptor causes it to translocate into the nucleus of the cell and act as a transcription factor. Here it affects gene expression in peripheral tissues to prepare the organism to respond physiologically and behaviorally to a challenge (Sapolsky et al., 2000).

Regulation of the HPA axis relies in large part on negative feedback mechanisms, by which increased levels of downstream hormones result in suppression of upstream hormone release. Glucocorticoid and mineralocorticoid receptors are distributed among multiple brain regions, including hippocampus, amygdala, hypothalamus, and pituitary (Morimoto et al., 1996; Patel et al., 2000; Briassoulis et al., 2011) allowing negative feedback from systemic CORT levels at multiple levels of the HPA axis. This negative feedback acts to decrease the release of CRH and ACTH (Keller-Wood et al., 1984). CORT availability is not the only mechanism of modulating negative feedback, as decreased availability of CBG results in increased free CORT and increased negative feedback to higher levels of the HPA axis (Sapolsky and Meaney, 1986). At the receptor level, GR availability in the hippocampus is implicated in sensitivity of negative feedback to the HPA axis, as increased GR transcription and receptor population are associated with decreased HPA axis reactivity (Sapolsky et al., 1984; Liu et al., 1997). GR availability may be decreased in situations of acute stress (Leonard, 2005).

Additionally, oxytocin (OXT), dehydroepiandrosterone (DHEA), and dehydroepiandrosterone sulfate (DHEA-S) appear to function both as part of the hormonal stress response and as regulators of the HPA axis. OXT is released from the posterior pituitary in response to a stressor, simultaneously with the release of ACTH, which is more classically considered part of the HPA axis stress response (Neumann, 2002); it is also released locally in the anterior pituitary for paracrine effect, i.e., affecting nearby cells, where it mediates ACTH release (Schwartz, 2000; Nakamura et al., 2008). OXT's function as a regulator of the HPA is complex; systemic OXT may suppress the rise of ACTH and CORT levels, but, conversely, OXT released within the PVN appears to stimulate an increase in ACTH and CORT levels (Neumann, 2002). DHEA and DHEA-S are two androgens released from the adrenal cortex simultaneously with CORT as part of the stress response and bind to NMDA and GABA receptors (Kroboth et al., 1999), which are distributed widely throughout the brain (Monyer et al., 1994; Pirker et al., 2000). As with CORT, the association between DHEA and DHEA-S levels and psychological stress is not well understood, but studies comparing DHEA-S and CORT ratios have suggested that the relationship between these hormones is an important component of the stress response (Young et al., 2002; Lennartson et al., 2013). These androgens may in fact serve to buffer the organism from high CORT levels (Morgan et al., 2004).

The complexity of the HPA axis and associated hormonal components of the stress response makes approaching its analysis difficult. Regulatory changes may occur at different levels (hypothalamic, pituitary, or adrenal), and in an array of different mechanisms at each level (receptor density, receptor function, hormone production, hormone release). Therefore, approaching this problem is facilitated by use of a well-understood model with adequate associated genomic resources.

1.2 THE SILVER FOX AS A MODEL

Previous models of non-pathological variation in HPA axis activity have included laboratory rats, both selectively bred (Naumenko et al., 1989; Landgraf, 2003) and not selectively bred (Weaver et al., 2006). However, rodent behavior is significantly less complex than human behavior, so these models cannot allow complete exploration of the effects of HPA axis

regulation on human behavior phenotypes. The relationship of HPA axis activity to personality traits has also been studied in human subjects (Oswald, 2006), but the limitations of human studies have prevented a deep exploration of the causal molecular mechanisms behind these associations. Moreover, because of the complexity of HPA axis regulation, it is likely that different animal models represent alterations at different levels of the axis and in different gene pathways. In fact, no one model may completely represent all relevant variations in non-pathological HPA axis variation. Therefore, there remains a need for models of differential HPA axis activity in a variety of species with complex social behavior.

A novel model for studying HPA axis reactivity as associated with behavioral phenotypes is the silver fox *(Vulpes vulpes,* a color variant of the red fox). Two lines of farmed silver foxes have been selectively bred for several decades, based on a simple behavioral test in which they demonstrate increased or decreased willingness to interact with a human tester, resulting in lines of "tame" and "aggressive" foxes. Breeding of the tame strain began in 1959 and of the aggressive strain in 1970, so that both lines have been under selection for more than 40 generations (Trut at al., 2004). A line of "conventional," or unselected, foxes was maintained as well, representing offspring of the same founding population that was also used for selection of tame and aggressive lines. This selection based on a single inter-specific behavioral response has led to different behavioral traits, such as decreased anxiety in the tame line, which contribute to complex differences in social behavior. Foxes from the tame line actively seek out human contact and display increased tolerance for novelty, whereas foxes from the aggressive line respond to human approach with avoidance and aggression and display anxiety-like behavior. Foxes from the conventional line, while not deliberately selected for behavior, do display fear of and aggression towards humans. High and low HPA axis activity segregates with these distinct behavioral phenotypes; foxes from the tame line display greatly reduced HPA axis activity compared to foxes from the aggressive or conventional farm-bred lines. Specifically, tame foxes display lower basal CORT and ACTH levels, a blunted rise of ACTH and CORT levels in response to a stressor (Trut et al., 2004), and prolonged attenuation of HPA axis response to stressors during early development (Oskina, 1995).

The tame fox model offers a unique opportunity to investigate two populations with dramatically different behavioral phenotypes, selected from the same founding population, in a canid with complex social behavior. Research in rats and mice linking increased anxiety and

decreased tolerance of novelty to increased HPA axis reactivity (Gass et al., 2001; Zhang et al., 2009; Turecki and Meaney, 2014) suggests that differences in personality traits and social behavior may be affected by regulatory differences in the HPA axis. Investigating regulatory mechanisms underlying the differences in HPA axis activity in the tame and aggressive foxes may help illuminate the biological underpinnings of their differences in behavior, and may provide insights into the relationship between HPA axis regulation, coping mechanisms, and psychological disorders.

Of particular interest in veterinary medicine is the potential relevance of this model to investigation of fear and anxiety in the domestic dog, a species in which fear aggression is a common problem and may lead to bite injuries (Reisner, 2003). One of the most marked differences between the dog and its wild ancestor, the wolf, is reduction in neophobia and aggression, and indeed this set of behavioral changes is seen in all domesticated species (Price, 2008; Driscoll et al., 2009; Campler et al., 2009); in fact, a "domestication syndrome" has been proposed, suggesting a unifying underlying mechanism operating in multiple species during the domestication process (Wilkins et al., 2014). Although models of domestication exist in rats (Albert et al., 2008) and guinea pigs (Künzl, et al, 2003), the fox is more closely related to the earliest domesticated animal, the dog, than is the rodent. This opportunity to use the fox model to investigate HPA axis changes that may have occurred during domestication, in addition to the relationship between HPA axis activity and behavioral phenotypes, makes this model a particularly compelling one.

1.3 MOLECULAR RESOURCES AVAILABLE FOR STUDY OF THE FOX MODEL

Multiple experiments have demonstrated that the behavioral differences between the tame and aggressive strains of silver fox have a strong genetic component (Trut et al., 2004; Kukekova et al., 2011). To implicate specific underlying genetic differences, genomic resources for the fox have been developed. A meiotic linkage map for the fox was established and aligned to the canine genome to provide an initial tool for genetic mapping studies (Kukekova et al., 2007). The transcriptome of the prefrontal cortex of one tame and one aggressive fox was sequenced using Roche 454 (Life Sciences, CT) to construct an initial catalog of genes expressed in this fox

brain region (Kukekova et al., 2011a). This study was later expanded using a larger number of samples and Illumina sequencing (Illumina, CA) (Wang et al., under review). Genotyping by sequencing (GBS) of fox reduced genome representation libraries has identified 48,294 fox single nucleotide polymorphisms (SNPs) (Johnson et al., 2015). Finally, the fox genome was sequenced, assembled, and annotated at Beijing Genomics Institute (Kukekova et al., under review). This assembly includes 676,878 scaffolds with a total length of 2,495,544,672 bp and a scaffold N50 of 11,799,617 bp; 21,418 fox protein coding genes were identified using homology information based on canine and human proteins, transcriptome data (including data from this project), and *de novo* gene predictions. The fox genome was used as a reference to identify 8.5 million fox SNPs in the re-sequenced genomes of 30 foxes from tame, aggressive, and conventional farm-bred populations (Kukekova et al., under review). In addition to the genomic resources currently available for the fox, it must be noted that the dog genome sequence is mature and well annotated. The dog is closely related to the fox, having diverged around 9-11 million years ago (Wayne et al., 1993). As the fox genome annotation is still incomplete, the canine genome annotation particularly provides an important resource in fox transcriptome studies.

To identify the genomic targets of selection for behavior, fox three-generation experimental pedigrees were constructed. Tame and aggressive foxes were bred to each other to produce an F1 generation, and F1 offspring were subsequently crossed back to tame or aggressive strains to produce backcross populations, or to each other to produce an F2 population (Kukekova et al, 2011; 2012). The behavior of all foxes in these pedigrees was tested, scored for the presence of 98 behavioral traits, and analyzed using principal component (PC) analysis (Kukekova et al, 2011b, Nelson et al., 2017). Quantitative trait loci (QTL) mapping in F2 pedigrees identified eight unique significant and suggestive loci that explained 2.75-12.42% of the phenotypic variance for the corresponding PCs (Nelson et al., 2017). An additional 14 unique significant or suggestive loci for individual behavioral traits in F2 animals were also mapped (Nelson et al., 2017), two of which were previously identified in fox backcross pedigrees (Kukekova et al, 2011b). QTL mapping is a powerful approach for the identification of genomic regions that harbor genetic changes influencing fox behavior, but it has its own limitations; the identified QTL intervals are usually wide and include hundreds of genes. To pinpoint positional candidate

genes within these regions, other methods must be applied, such as the identification of genomic regions of extreme diversity between fox populations or differential gene expression.

A SNP analysis of tame and aggressive fox populations confirmed that significant genetic diversity has been preserved in these populations despite many years of selective breeding (Johnson et al., 2015; Kukekova et al., under review). Population structure analysis of three populations (tame, aggressive, and conventional farm-bred) clearly differentiated these populations from each other and showed more divergence between the tame and conventional than between the aggressive and conventional populations (Figure 1.1). These findings are consistent with the fact that foxes from the conventional farm-bred population were ancestors to both tame and aggressive strains, but the tame population has been under selection for a decade longer.

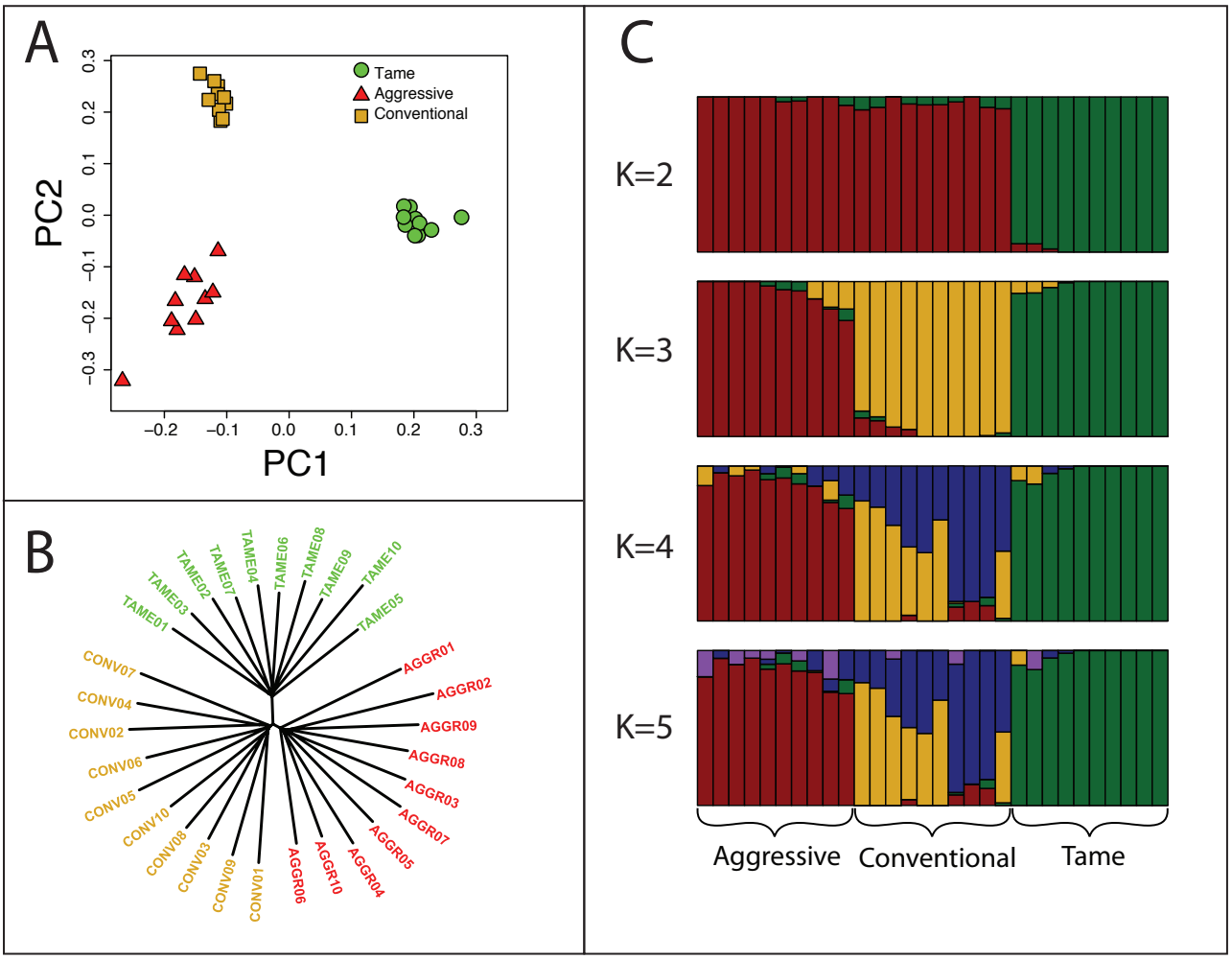

Figure 1.1. Population structure analysis of tame, aggressive, and conventional fox populations. All analyses were performed on SNPs from 30 foxes, 10 from each population. A. Principal component analysis separated fox samples into distinct clouds representing three populations. B. Neighbor-joining tree analysis clustered all individuals from the same line together. C. STRUCTURE analysis clustered tame and aggressive lines separately from the conventional population (Kukekova et al., under review).

Because the tame and aggressive populations were selected solely for specific behavioral phenotypes and efforts were made to minimize inbreeding, these populations are well suited to the identification of the genomic targets of selection for behavior (Trut et al., 2004, 2009).

Recent analysis of the resequenced genomes of 30 foxes from the three populations identified 113 regions with either significantly decreased heterozygosity in one of the three populations or increased divergence between two populations. A strong positional candidate gene for a QTL on fox chromosome 15 was highlighted: *SorCS1*, which encodes the main trafficking protein for AMPA glutamate receptors and neurexins and suggests a role for synaptic plasticity in fox domestication (Kukekova et al., under review).

1.4 THE IMPORTANCE OF GENE EXPRESSION STUDIES IN THE FOX MODEL

Until recently, study of the fox model was limited to genetic mapping studies, as tools for study of the fox transcriptome were lacking. However, recent availability of next generation sequencing opens up new possibilities in transcriptomics for this model. Investigation of gene expression in the tame and aggressive fox will provide important data about the molecular mechanisms involved in differences between the two lines. For example, transcriptome data may enable the annotation of candidate genes identified through genomic approaches with information about differential expression in specific tissues of interest, or inclusion in gene networks with different activities between tissues.

Thus, while previous work in the fox model has implicated several loci and many genes in the differences between the tame and aggressive phenotypes, additional work remains to validate these findings, identify which genes of interest have biological significance in the difference between the lines of foxes, and, critically, to connect genes of interest to mechanisms that may plausibly contribute to those differences. I have chosen the HPA axis, a pathway well described as significantly altered in the tame foxes and additionally strongly associated with human behavioral disorders, as my primary focus of study. Although HPA axis regulation is a complex process affording many different approaches to scientific inquiry, I will focus on gene expression and associated genomic regulatory mechanisms, as well as investigating exonic variants that may contribute to ongoing studies of causative genes.

To this end, I will employ RNA sequencing (RNA-seq), a technology that provides more information about expression of transcript variants than does the older microarray approach, and also allows calling of (mainly exonic) variants. I will perform RNA-seq in anterior pituitary and

adrenal tissue, and call variants in hypothalamus, anterior pituitary, and adrenal tissue, from tame and aggressive foxes. I will use this data to investigate differences in gene expression, gene networks, and SNPs, with the goal of uncovering molecular mechanisms related to the differences in HPA axis regulation between the two lines.

CHAPTER 2. TRANSCRIPTOME ANALYSIS IN DOMESTICATED SPECIES: CHALLENGES AND STRATEGIES[1]

2.1 INTRODUCTION

Over thousands of years, humans have selectively bred domesticated animals for different uses and environments, resulting in a wide diversity of morphology and behavior. In fact, for some traits the variation observed in domesticated animals is much greater than that found in laboratory or wild species. The striking diversity observed among individuals within domesticated species provides advantages for genetic studies of traits with direct relevance to biomedical research as well as traits with economic and cultural value. For example, some traits, such as herding behavior in some varieties of dog or a comfortable riding gait in a gated horse, facilitate specific animal uses. Other traits, such as milk or wool production, represent an increasingly salient avenue of study as the global demand for food and fiber increases. Some adaptations may even be shared between humans and domesticated animals, such as altered metabolism to facilitate living in extreme environments. An improved understanding of the underlying biological mechanisms associated with these traits will help further selection for increased productivity, utility, and health. Indeed, the health of domesticated animal populations is closely tied to that of our own societies and the environments that we share with them (Zinsstag et al., 2011). Domesticated animals are threatened by zoonotic diseases that also threaten us (Kahn, 2006) and have risk factors for hereditary diseases that often closely mimic our own (Khanna et al., 2006; Schoenebeck and Ostrander, 2014; Stefaniuk and Ropka-Molik, 2015). Many of these diseases represent natural models for corresponding human conditions, and clinical studies in domesticated animals receiving advanced veterinary care may facilitate the development of innovative treatment strategies also of use in human medicine (Khanna et al,

1 This chapter was previously published as: Hekman, Jessica P., Jennifer L. Johnson, and Anna V. Kukekova. “Transcriptome analysis in domesticated species: challenges and strategies.” *Bioinformatics and Biology Insights* 9.Suppl 4 (2015): 21.

2006; Paoloni and Khanna, 2008). Progress in understanding human inherited diseases, the genetic architecture of complex phenotypes, and development of treatments against infectious or inherited diseases, can therefore be significantly advanced through studies of similar traits and conditions in domesticated animals.

The history and population structure of domesticated species make them well suited for genetic studies (Lindblad-Toh et al., 2005; Andersson, 2013). Selection of domesticated strains for different morphological and other characteristics has led to the formation of "breeds" which are maintained in complete or partial reproductive isolation from each other. This practice of closed breeding generally results in reduced effective population size (i.e., the number of individuals in a population who contribute genetically to the next generation) within breeds. This reduction is due to three effects: small founding populations, population bottlenecks, and the popular sire effect (Ostrander and Kruglyak, 2000; Gibbs et al., 2009). As a result of closed breeding and intense selection for specific traits, individuals within a breed commonly share long stretches of homozygosity at genomic loci related to the traits under selection (Andersson and Georges, 2004; Lindblad-Toh et al., 2005; Purfield et al., 2012). Although the haplotype associated with a specific trait can be relatively long within a breed, a comparison of haplotypes across breeds characterized by the same phenotype allows a reduction of the critical interval and thereby facilitates identification of the genes under selection (Van Laere et al., 2003; Lindblad-Toh et al., 2005; Setoguchi et al., 2009). For example, small size in dogs is linked to a specific haplotype for *IGF1* that is shared by a majority of toy dog breeds (Chase et al., 2002; Sutter et al., 2007). Closed breeding also leads to accumulation of disease-associated mutations, as has been demonstrated in dogs (Patterson, 2000), cattle (Jolly and Leipold, 1973; Meyers et al., 2010; Venhoranta et al., 2014), and other species (Charlier et al., 2001; Beever et al., 2006; Rosengren et al., 2008; Sironen et al., 2006; Sironen et al., 2011). The identification of disease-associated genes in domesticated animals as compared to human populations is facilitated by domesticated species population structures and access to samples from many individuals in a pedigree (Chase, 1999; Acland et al., 1998; Andersson and Georges, 2004). For example, limited within-breed genetic diversity in dogs and the elevated occurrence of particular cancers in particular breeds facilitates the study of cancer susceptibility loci in that species (Tonomura et al., 2015; Arendt et al., 2015).

With the sequencing of the dog, cow, and pig genomes (Lindblad-Toh et al., 2005; Zimin et al., 2009; Groenen et al., 2012), the genome-wide association study (GWAS) became a common approach for the identification of genomic regions implicated in traits of interest in domesticated animals. This approach tests the probability of association of genetic markers with a trait; in domesticated animals, GWAS have commonly found a relatively small number of loci for such complex phenotypes as height, skull shape, or coat quality (Cadieu et al., 2009; Makvandi-Nejad et al., 2012; Andersson, 2013; Schoenebeck and Ostrander, 2014). The fact that variation in these traits is explained by a small number of loci in these species is likely due to intense selective pressure (Jones et al., 2008; Boyko et al., 2010; Akey et al., 2010). As a result, many GWAS in domesticated species have successfully identified causal genes both for Mendelian traits and for complex traits controlled by loci with large effect size (for review, Schoenebeck and Ostrander, 2014; Andersson, 2013). However, not all traits are amenable to investigation by this technique. Although the longer stretches of homozygosity common to breeds result in relatively large target regions that provide a strong signal, such large loci may contain dozens of genes and therefore may provide poor resolution for identification of causative genes lying within them. Additionally, study design may be complicated by a lack of knowledge about the underlying genotype of a trait shared by multiple breeds, which may be identical by descent or different due to distinct causal mutations. Finally, not all traits in domesticated animals are controlled by a small number of genes of large effect; some complex traits, such as behavior, weight, meat quality, milk production, and some diseases, like hip dysplasia and cancer, are controlled by many genes of small effect (Ashwell et al., 2004; Zhou et al., 2010; Karlsson et al., 2013; Ma et al., 2013). As has been demonstrated by GWAS in human studies, loci of small effect may prove particularly challenging in the elucidation of molecular mechanisms, as they may require large sample sizes to achieve statistical significance (Ng et al., 2008; Lango Allen et al., 2010).

When identified regions of interest are large, when many loci of small effect are implicated, or when the function of the discovered genes is unknown, GWAS alone will not be sufficient to elucidate genetic mechanisms associated with the phenotype under investigation. An alternative approach employs analysis of gene expression differences to pinpoint changes in pathways rather than in specific genes. For example, gene expression studies have proven particularly well suited to investigations of genomic changes in neoplastic cells, illuminating the molecular distinctions

between different types of breast cancer (Reis-Filho and Pusztai, 2011) and contributing, along with copy number variation analysis, to the identification of oncogenes (Huang et al., 2011).

Microarray gene expression studies pioneered the use of genome wide techniques in the hunt for sets of genes or gene networks implicated in complex phenotypes in domesticated species (Saetre et al., 2004; Everts et al., 2005; Klopfleisch et al., 2010). However, microarray technology is limited by its dependence on the use of known probes, requiring a species-specific chip for most accurate results. Cross-species microarray use may result in decreased specificity of the hybridization signal due to species-specific polymorphisms; it should be expected to be most sensitive when performed on closely-related species, ideally less than 10 million years divergent from each other (Renn et al., 2004; Saetre et al., 2004). Moreover, even with a chip designed for the species under study, the dependence of microarray technology on known probes implies that transcripts that do not correspond to known sequences will not be detected, and novel isoforms will not be distinguished from known splice forms.

The advent of next generation sequencing (NGS) revolutionized gene expression studies by obviating the need for pre-existing probes for transcripts. RNA sequencing, or RNA-seq, uses the high throughput reads produced by NGS to represent the entire transcriptome, in other words, all transcripts produced in a tissue sample including previously uncharacterized transcribed sequences and novel isoforms. RNA-seq is used for a variety of applications, most commonly to discover lists of genes that are differentially expressed between experimental groups, such as samples from different tissues (Roy et al., 2013; Tonomura et al., 2015), samples from different treatment groups (Matulova et al., 2012; Duffy et al., 2013), or samples from different populations (Bottomly et al., 2011 Voineagu et al., 2011). To identify gene networks associated with inherited diseases or other genetic traits, individuals can be grouped by disease status (affected versus unaffected) or different haplotypes at the mapped loci (Bottomly et al., 2011; Gautier et al., 2012; Balakrishnan et al., 2014; Rickard et al., 2015; Tonomura et al., 2015). Differential gene expression may complement association studies when used to provide differential expression information about genes in the genomic regions of interest identified by GWAS (Tonomura et al, 2015; Arendt et al., 2015). In addition to gene expression differences, some RNA-seq studies may seek differences in isoform expression (Shalek et al., 2013; Cheng et al., 2014) and allele-specific gene expression (Gregg et al., 2010; Wang et al., 2013). Concurrently with analysis of gene expression, RNA-seq data may be used for calling variants

such as single nucleotide polymorphisms (SNPs) or simple sequence repeats (SSRs) for subsequent use in marker studies (Chepelev et al., 2009; Iorizzo et al., 2011). This use is well suited to non-model species with limited genomic resources, to call variants that are novel for the species or that are enriched in the population under study. Finally, RNA-seq is used to improve genomic annotations through the identification of novel transcripts (Rogers et al., 2014; Du et al., 2015; Hoeppner et al., 2014).

RNA-seq has a particular advantage in non-model species, specifically those less common domesticated species for which species-specific chips for microarray studies are lacking. However, despite the promise of RNA-seq technology, performing differential gene expression experiments with RNA-seq may be challenging in domesticated species with low quality genomes or a lack of high quality reference annotation. In this review, we will discuss strategies for performing differential expression analysis in non-model species, focusing particularly on challenges common to studies in domestic species.

2.2 RNA-SEQ EXPERIMENTAL DESIGN

2.2.1. Number of replicates

One of the first steps in the experimental design of an RNA-seq study is selection of the optimal number of biological replicates. At least a few replicates are necessary in order to characterize biological variation and separate it from technical variation (Auer and Doerge, 2010), and additional replicates provide additional benefits. Specifically, increasing the number of replicates in an RNA-seq study results in increased power available for differential gene expression analysis. When a trade-off must be made between number of replicates and depth of sequencing, replicates may be more instrumental than sequencing depth in increasing power to detect differential expression (Rapaport et al., 2013; Liu et al., 2014). Additionally, studies employing samples from outbred domesticated animals require a larger number of replicates than those employing a group with reduced genetic diversity, such as inbred mouse strains (Williams et al., 2014).

2.2.2. Sample preparation

A variety of commercial RNA extraction kits are available for RNA isolation; though the kits generally extract similar amounts of RNA, they differ in the quality of RNA extracted. Therefore, the choice of kit may affect study results (Sellin Jeffries et al., 2014). After RNA extraction, messenger RNA (mRNA) is isolated using either the polyA capture or rRNA depletion protocol. The polyA capture protocol results in a bias to the 3' end of transcripts, while the rRNA depletion protocol results in more variation in depth of coverage throughout the length of transcripts (Lahens et al., 2014). Decreased depth on the 5' end of transcripts sequenced from libraries built with the polyA capture protocol may result in decreased likelihood of identifying differential exon expression on the 5' end of transcripts, decreased depth of sequencing of long genes compared to short genes, and poor coverage of 5' untranslated regions, particularly important in the use of RNA-seq for improvement of transcriptome annotation. The rRNA depletion protocol depends on known ribosomal RNA sequences, and the probes have not been tested on all species. For example, while the Ribo-Zero Kit (Illumina, San Diego, CA) is predicted to work on all mammalian species due to probe homology, it has only been tested on human, mouse, rat, and dog, as stated in its manual. Its efficiency on avian genomes such as chicken and turkey is also unknown. After mRNA isolation, cDNA libraries are constructed. Libraries may be non-strand-specific, or may support strand-specific RNA reads, which allow transcripts to be identified as sense or antisense. Strand-specific reads have been used in transcriptome assembly (Perkins et al., 2009; Grabherr et al., 2011) and may facilitate differentiation of reads from adjacent or overlapping genes transcribed from opposite strands (Levin et al., 2010). Additionally, Illumina (San Diego, CA) sequencers support either single or paired end read sequencing. Paired end sequencing may be more expensive, but increases the percentage of reads successfully mapped to the genome. Its use is recommended for detection of distinct isoforms; however, its increase in mapping of unique reads may be only marginal, so use of paired end reads is not recommended unless maximizing unique read mapping is critical to the project (Williams et al., 2014).

2.2.3. Sequencing strategy

After library construction, transcripts are typically sequenced on a next generation sequencing platform; currently, Illumina sequencers are the most common. Sequencing considerations include determining appropriate read length and number of lanes (i.e., sequencing depth). Illumina HiSeq 2500 sequencers produce reads of 50-150 bp in length; they employ flow cells with eight lanes, and multiple samples may be run on a single lane. To differentiate reads from different samples after sequencing on the same lane, a unique bar code may be attached to each sample during library preparation. The appropriate number of lanes must be determined by taking into account the necessary depth of sequencing; for example, studies that rely on the detection of rare transcripts or polymorphisms will require greater depth (Tarazona et al., 2011). Artifactual variation per lane may contribute technical variation to a study, but this can be avoided by the use of multiplexing, e.g., ensuring that each lane contains a balanced number of samples from each treatment group (Auer and Doerge, 2010).

Longer reads result in an increased percentage of mapped transcripts and improved handling of splice junctions during alignment (Pepke et al., 2009). Longer reads may therefore prove particularly useful for projects using species without an existing reference genome sequence that require *de novo* transcriptome assembly or identification of alternative transcripts; otherwise, a 50-bp read length should be sufficient (Williams et al., 2014). In the past, 454 pyrosequencing (454 Life Sciences, Branford, CT) has been used to produce RNA-seq reads between 100-500 bp in length (Sugarbaker et al., 2008; Hahn et al., 2009; Maher et al., 2009; Meyer et al., 2009; Schwarz et al., 2009; Zhao et al., 2009; Ferguson et al., 2010; Wang et al., 2010; Kukekova et al., 2011; Ekblom et al., 2012). However, this technology has proven prohibitively expensive and is currently not widely available. Emerging platforms such as PacBio (Pacific Biosciences, Menlo Park, CA) that provide longer read lengths may prove popular in the future, perhaps even providing the ability to sequence entire transcripts in a single read.

2.3 RNA-SEQ BIOINFORMATIC WORKFLOW

A typical bioinformatic workflow using a reference genome and aimed at identification of differentially expressed genes is described below and summarized in Table 2.1. The workflow

begins with raw reads, which are aligned to a reference genome. Gene counts are then quantified from the alignment files and used in differential gene expression analysis.

Table 2.1. RNA-seq bioinformatic workflow for calling differentially expressed genes.

Step	Tools	Challenges
1. Remove low-quality reads, barcodes, and adapters	Fastx-toolkit, FLEXBAR, or Trimmomatic	Follow recommended protocol
2. Remove mitochondrial and ribosomal sequences	Bowtie2	Sequences from the same or related species should be used
3. Align to reference genome	TopHat2	Incomplete or non-existent reference genome
4. Call differentially expressed genes	DESeq2, edgeR, or limma	Incomplete or non-existent reference genome annotation

2.3.1. Read filtering

Post sequencing, several filtering steps are recommended in order to produce a high-quality dataset. Common tools for removal of low quality sequences as well as barcodes and platform-specific adapters added during library construction are fastx-toolkit (Gordon and Hannon, 2010), FLEXBAR (Dodt et al., 2012), and Trimmomatic (Bolger et al., 2014). These tools operate on FASTQ-format files and accept command-line parameters to specify the minimum length or Phred score below which a read should be discarded. Ribosomal and mitochondrial sequences may subsequently be removed, although these types of RNA should have been depleted in large part during library construction (Wilhelm and Landry, 2009). Removal of mitochondrial or ribosomal reads may be accomplished through alignment of all reads to mitochondrial and ribosomal sequences from related species, obtained, for example, from NCBI's RefSeq database (http://www.ncbi.nlm.nih.gov/refseq/). Unaligned reads kept for further processing will then be depleted of mitochondrial and ribosomal sequences. Although the exact percentage of removed reads will depend on details of the chosen pipeline, in our laboratory, following this protocol in

three separate tissues sequenced in different runs resulted in removal of 5.7-13.3% of reads (unpublished data).

2.3.2. Alignment

After filtering, the RNA reads are typically aligned to a reference genome. This mapping process is complicated by the presence of splice junctions in the reads, originating from post processing of mRNA. Two approaches may be used to reduce the number of reads mismapped due to splice junctions. First, a splice junction-aware aligner should always be used for mapping RNA reads, as alignment of this type of reads with a non-splice aware aligner such as Bowtie2 or BWA results in a higher percentage of mapping errors (Fonseca et al., 2014). TopHat2 (Kim et al., 2013) and STAR (Dobin et al, 2012) are two splice-aware aligners that are widely used for RNA read mapping; STAR operates at greater speeds than other aligners but has a correspondingly larger memory footprint (Williams et al., 2014). These aligners have similar mapping performance with a low median error rate (Fonseca et al., 2014). Second, annotation of splice junction locations in the genome should be provided to the aligner when available. Typically, species annotations are archived at Ensembl (ftp://ftp.ensembl.org/), RefSeq (ftp://ftp.ncbi.nlm.nih.gov), and UCSC (http://genome.ucsc.edu/). Selection of an annotation with broader gene coverage will result in an increased percentage of reads mapped to genes. In human, the Ensembl annotation provides the broadest gene coverage and, as a result, corresponds to the highest gene mapping rates (Zhao and Zhang, 2015). Ensembl also provides increased coverage of dog, including intron and UTR annotation that is not available through the other two annotations (personal observation). The appropriate choice of annotation may vary from species to species, but in general, the annotation with the broadest coverage should be selected to maximize mapping rates.

Alignment of reads to the genome is further complicated by reads that map to complementary sequences at multiple locations in the genome. Ambiguous mapping may be due to conserved domains of paralogous genes, pseudogenes, and repeats (Pepke et al., 2009). Such reads are particularly problematic in gene differential expression studies, as some gene count quantification tools discard them (Fonseca et al., 2014). A paired end sequencing approach

results in an increased percentage of uniquely mapped reads, though this improvement may be minimal (Williams et al., 2014).

Visualization of aligned reads offers the opportunity to evaluate the dataset before continuing. Such evaluation can provide opportunities to better understand problems such as coverage bias, intronic or intergenic reads, or overlapping genes. Two such visualizers are GenomeBrowse (Golden Helix, Bozeman, MT) and IGV (Broad Institute, Cambridge, MA). Although exonic reads make up the preponderance of RNA-seq datasets, introns, untranslated regions, and inter-genic regions are often retained, albeit at lower depth. These non-exonic regions may not be artifactual but may be a result of pervasive transcription of the genome (Clark et al., 2011). Reads that align to inter-genic regions may also represent unannotated exons (Pickrell et al., 2010), or long non-coding RNA (lncRNA) transcripts. At least 15,512 lncRNAs have been identified in human (Harrow et al., 2012) and 7,224 in dog (Hoeppner et al., 2014).

2.3.3. Differential gene expression analysis

Gene expression must be quantified in reads before differences in expression can be identified. An assessment of quantification tools shows that while results from different tools are often highly correlated, results from a subset of genes may display differences as great as 10-fold. Identification of the quantification tool with the greatest accuracy is difficult, as accurate counts may not be known for comparison. However, in a comparison of different pipelines composed of a variety of quantification tools and aligners, pipelines including the HTSeq-count quantification tool numbered among those with the best performance (Fonseca et al., 2014).

Outlier samples may influence differential expression results, and should be identified and removed prior to differential gene analysis. George et al., (2015) describe a leave-one-out approach to detection of outliers. Alternatively, some differential expression analysis tools, such as DESeq2 (Love et al., 2014), perform outlier detection and removal automatically. Another differential expression tool, edgeR (Robinson et al., 2009), incorporates outlier detection into its estimate of genewise dispersion when the `robust=TRUE` parameter is specified in the `estimateDisp()` method (Chen et al., 2015).

Differential gene expression analysis tools are confronted with normalization difficulties that are inherent in the analysis of RNA-seq reads, namely, bias due to different depths of sequencing per sample or to gene length (Tarazona et al., 2011). Additionally, these tools must contend with the small replicate numbers that are typical of RNA-seq experiments, often as low as 2-3 replicates (Love et al., 2014). A comparison of eleven different methods for differential expression analysis showed that while some widely used methods have similar accuracy, the sets of differentially expressed genes found by different methods varies significantly (Soneson and Delorenzi, 2013). Therefore, analysis of a single dataset by several methods may provide increased sensitivity, by considering genes identified by any tool as differentially expressed, as well as increased specificity, by considering genes identified only by multiple tools as differentially expressed. The tools DESeq2, edgeR, and limma have been found to have superior specificity and sensitivity (Rapaport et al, 2013) and are widely used. To ensure that the findings from the RNA-seq analysis are not artifactual, it is recommended that real time quantitative PCR (RT-qPCR) be used to evaluate a representative set of differentially expressed genes (Matulova et al., 2012; Duffy et al., 2013; Pérez-Montarelo et al., 2013). The use of RT-qPCR, a well-established method for the evaluation of gene expression, provides technical validation of the RNA-seq procedure and data analysis used for identification of differentially expressed genes.

Although differential expression has conventionally been performed using the gene or the exon as the base unit, it may alternatively be performed at the level of the nucleotide or region of sequential nucleotides. DER Finder (Frazee et al., 2014) analyzes differential expression by nucleotide, and therefore does not require annotation of gene locations. It does require a sequenced genome, which may be a draft assembly. This tool may therefore prove useful in species lacking genome annotation. It may also be used concurrently with the pipeline described above to provide additional information about expression differences at the base, rather than the gene, level.

2.3.4. Analysis of differentially expressed gene lists

Typical differential expression analyses produce lists of hundreds of differentially expressed genes, requiring further analysis to construct a high-level overview of changes between the groups being compared. A commercial package, Ingenuity Pathway Analysis (IPA) (QIAGEN,

Redwood City, CA), provides a graphical user interface to assist in discovery of pathways enriched in differentially expressed genes, generates publication-quality figures, and offers links to peer-reviewed articles about differentially expressed genes and related pathways. For studies with smaller budgets, the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2008) offers a freely available alternative with a web-based interface. DAVID groups differentially expressed genes by related Gene Ontology topics, thereby associating genes with other genes that have similar functions. While DAVID does not provide figure generation, visualization may be accomplished by use of the freely available tool Cytoscape (Shannon et al., 2003; Smoot et al., 2010). This tool, which will run on OS X, Windows, or Linux, provides a graphical interface to allow the user to specify how to visualize a network of genes. Finally, weighted gene co-expression network analysis (WGCNA) may be used to identify clusters of genes with highly correlated co-expression patterns. This tool identifies networks of genes that are perturbed together, thereby suggesting biological pathways affected by the model in question (Langfelder and Horvath, 2008).

2.3.5. SNP detection

In addition to their use in gene expression studies, RNA-seq reads may be used to call SNPs. The Genome Analysis ToolKit (Broad Institute, Cambridge, MA) provides a pipeline for calling SNPs specifically in RNA data. Alternatively, the SAMtools and BCFtools toolkits may be used in conjunction to call SNPs (Li et al., 2009; Li, 2011). Both pipelines require alignment files (BAM) as input rather than raw reads (FASTQ). The identified SNPs will for the most part be in exonic or untranslated regions (UTR); calling of intronic or inter-genic SNPs requires genomic, not transcriptomic, data.

2.3.6. Computing resources

Alignment and SNP calling processes may be computationally expensive and are best performed on a high performance server or cluster rather than a desktop computer. An 8-core cluster with 32 GB of RAM has been recommended as the minimum hardware requirement for a typical alignment process. However, a desktop computer is generally sufficient for calling

differentially expressed genes (Vijay et al., 2012). Additionally, file sizes for reads from an individual sample aligned to a genome may be expected to reach nearly ten gigabytes in size, depending on read depth (unpublished data).

2.4. RNA-SEQ IN DOMESTICATED SPECIES

2.4.1 Challenges of RNA-seq in domesticated species

The RNA-seq analysis pipeline described above has been successfully employed in species with "finished" reference genomes, such as human, mouse, and fruit fly, to identify genes that are differentially expressed between different samples (Duffy et al., 2013, Liao et al., 2014; Rogers et al., 2014). These genomes were constructed using Sanger sequencing (Adams et al., 2000; Lander et al., 2001; Chinwalla et al., 2002), resulting in high quality assemblies with low scaffold numbers, high scaffold N50 lengths, and low total assembly gap lengths. These widely used reference genomes are also extensively annotated, making available a large number of transcripts for gene expression studies, with high percentages of curated transcripts. Curated transcripts have been manually reviewed to remove sequence errors and to ensure association with the correct genomic locus (Pruitt et al., 2007).

Table 2.2. Properties of assemblies of human, mouse, and domesticated species. All data was downloaded from ncbi.nlm.nih.gov on December 2, 2015.

Species	Assembly	Sequencing technology	Scaffolds	Scaffold N50 (bp)	Total gap length (bp)
Human	GRCh38.p5	Sanger	797	59,364,414	161,368,151
Mouse	GRCm38.p4	Sanger	293	52,589,046	79,356,756
Dog	CanFam3.1	Sanger	3,310	45,876,610	18,261,639
Chicken	Gallus_gallus-4.0	Sanger	16,847	12,877,381	14,074,301
Cow	Btau_4.6.1	Sanger	13,387	2,599,288	176,429,395
Pig	Sscrofa10.2	Sanger and NGS	9,906	576,008	289,397,178
Turkey	Turkey 5.0	NGS	233,806	3,801,642	35,294,427
Yak	BosGru 2.0	NGS	41,192	1,407,960	120,154,638
Ferret	MusPutFur1.0	NGS	7,783	9,335,154	132,851,443

Table 2.3. Numbers of curated and uncurated transcripts annotated in the RefSeq database for human, mouse, and domesticated species. All data was downloaded by searching "Species name"[porgn] AND refseq[filter] AND biomol_mrna[PROP] (e.g., "Canis lupus familiaris"[porgn] AND refseq[filter] AND biomol_mrna[PROP]) at http://www.ncbi.nlm.nih.gov/nuccore/ on December 10, 2015.

Species	Species name used in RefSeq search	Total RefSeq transcripts	Curated RefSeq transcripts	Uncurated RefSeq transcripts
Human	Homo sapiens	100,068	39,623	60,445
Mouse	Mus musculus	78,241	29,900	48,341
Dog	Canis lupus familiaris	47,095	1,675	45,420
Chicken	Gallus gallus	32,244	6,197	26,047
Cow	Bos taurus	70,342	13,329	57,013
Pig	Sus scrofa	47,445	4,154	43,291
Turkey	Meleagris gallopavo	26,450	93	26,357
Yak	Bos mutus	28,868	7	28,861
Ferret	Mustela putorius furo	48,113	61	48,052

A small number of domesticated species genomes were also constructed entirely or in part using Sanger sequencing, incuding dog (Lindblad-Toh et al., 2005), chicken (Hillier et al., 2004), cow (Zimin et al., 2009), and pig (Groenen et al., 2012). The assemblies for these species are of variable quality compared to human and mouse. All four have a larger number of scaffolds, and chicken, pig, and cow have a significantly shorter N50 length (Table 2.2). Annotation of these widely used domesticated animal genomes is also less complete than in human or mouse, with fewer total transcripts in the Reference Sequence collection (RefSeq) and a much smaller percentage of manually curated transcripts (Table 2.3). As a result, gene expression pipelines in

these species will have the use of many fewer isoforms and may encounter a higher percentage of transcripts with sequence errors. Moreover, annotation of gene and exon locations may be inadequate to identify important genes even in relatively well-annotated genomes (Hoeppner et al., 2014). For example, the POMC gene is not annotated in *Canis familiaris* 3.1 (verified by search on UCSC Genome Browser, October 20, 2015). This gene encodes pre-pro-opiomelanocortin, which is cleaved to produce β-endorphin and met-enkephalin, which are endogenous opioid peptides; α-melanocyte-stimulating hormone, which is important in feeding behavior and energy homeostasis; and adrenocorticotropic hormone, which is a component of the hypothalamic-pituitary-adrenal axis. POMC's function is therefore well characterized and this gene may be expected to be of interest in a variety of studies. Incomplete annotation is a significant limitation for an RNA-seq pipeline, as existing differential expression tools rely on accurate annotation (Frazee et al., 2014).

Although domesticated animal genomes constructed using Sanger sequencing have some limitations, they are of relatively high quality. However, in the face of the plummeting cost of NGS, the number of unfinished “draft” genomes has increased (Chain et al, 2009; Alkan et al., 2010). Many genomes of less studied domesticated species were constructed entirely with NGS, such as Illumina or 454, as for example turkey (Dalloul et al., 2010), yak (Qiu et al., 2012), and ferret (Peng et al., 2014). NGS uses libraries with smaller insertions than does Sanger sequencing, usually not longer than 10-20 kb; as a trade off, it produces shorter scaffolds. These shorter, sometimes misassembled scaffolds result in fragmented genes and a significant number of missing coding exons (Alkan et al., 2010; Ye et al., 2011). The assemblies for the reference genomes of turkey, yak, and ferret, for example, have many more scaffolds than do the human and mouse assemblies (Table 2.2). Annotation of these newer genomes may also lag behind that of more widely studied species, as is evidenced by the lower number of total transcripts and dramatically lower number of curated transcripts in RefSeq for turkey, yak, and ferret (Table 2.3). Assembly errors in draft genomes such as these have been shown to result in misannotation, particularly by automated annotators; moreover, the completeness of draft genome annotations is difficult to assess (Norgren, 2013).

Overall, domestic animal reference sequences have a wide range of qualities. Some may prove to have assemblies and annotations that are complete enough to support the described pipeline as a sole approach to RNA-seq analysis. Analysis of others using the described pipeline

may prove difficult if the associated reference genome and annotation contain significant missing information. Still others may have no reference genome at all. Therefore, an alternate approach may be required for analysis of RNA-seq reads from less widely studied species.

2.4.2 Alignment of RNA-seq reads to a related reference

One solution to this impasse is the use of a reference genome and annotation from a closely related species. For example, a study of gene expression changes in macrophages of red deer in response to paratuberculosis used the cow genome for alignment of deer RNA-seq reads (Marfell et al., 2013). This solution may be practical for an increasing number of less common domesticated species as more genomes are assembled and annotated. Critically, the reference used must itself be mature and well annotated for this approach to provide real benefit; moreover, the related genome should be not more than 15% divergent for best results (Vijay et al., 2012).

Alignment of reads from one species to the genome of another may prove challenging due to species divergence, even in closely related species. Genomic differences such as SNPs and indels may decrease mapping accuracy when reads are aligned to the genome of a different species, resulting in a decreased depth of coverage due to the loss of reads that cannot be mapped. The default parameters of splice-aware aligners may therefore not be appropriate for use in this situation. For example, without modification of its default parameters, TopHat2 will not accept an alignment of a read to a location if that alignment has more than two mismatches. As a result, three or more differences in a read of 100 or 150 base pairs (2-3% divergent) will result in a rejected alignment. A difference may be a mismatch due to a SNP, or an edit due to either a SNP or a gap (indel). TopHat2's tolerance for mismatches and edits may be increased using the --read-mismatches and --read-edit-dist parameters. In our lab, alignment of *Vulpes vulpes* reads to the *Canis familiaris* 3.1 reference using TopHat2 with default parameters resulted in 69% alignment. Alignment with --read-mismatches=3 and --read-edit-dist=3 increased alignment to 79% (unpublished data). Increasing the allowed mismatches and edit distance should be performed with caution, for fear of false positive alignments (Quinn et al., 2014).

An alignment tool designed to handle divergent genomes, Stampy (Lunter and Goodson, 2010), is tolerant of up to 15% sequence divergence. Stampy assumes 0.001 substitutions per

site, but this default may be modified by the command-line parameter –substititionrate=. Stampy has been successfully used to map RNA-seq reads from white-throated sparrow, song sparrow, and white-crowned sparrow to the zebra finch genome with the substitution rate set to 0.05 (Balakrishnan et al., 2014).

2.4.3. *De novo* assembly of RNA-seq reads

If a high quality reference genome of a closely related species is not available, an alternative solution uses the *de novo* assembly of a reference transcriptome from available RNA-seq reads. This approach was successfully used in our lab for the assembly of the brain transcriptome of silver fox (*Vulpes vulpes*) using 454 reads (Kukekova et al., 2011). *De novo* assembly may be performed in the absence of any genome, or may be guided by a reference genome if one is available. This approach suffers from difficulties in annotating the assembled contigs as well as increased computation requirements (Frazee et al., 2014). Additionally, use of a reference genome from a related species that is 15% or less divergent was found to recover more bases than use of a *de novo* genome from the species under investigation (Vijay et al, 2012). In practice, however, the choice to use a reference approach over a *de novo* assembly approach may depend not just on the divergence of the two genomes, but on the quality of the reference genome assembly and completeness of its annotation. When no reference of a closely related species is available, or the available reference is not sufficiently annotated, the *de novo* approach may be a valid alternative.

Widely used assemblers that will operate independently of a genome include Trinity (Grabherr et al., 2011), Velvet (Zerbino et al., 2008), and Oases (Schulz et al., 2012). Genome-independent assembly typically requires significant time on a high-end server or cluster. Exact requirements vary depending on the number of reads, but may include hundreds of gigabytes of memory and hundreds of hours of runtime (Garber et al, 2011; Wolf, 2013). For example, recommendations for use of Trinity include allocation of about 1 gigabyte for every one million reads assembled, and from 256 gigabytes to 1 terabyte of memory (Haas et al., 2013). Labs that do not have access to their own high-end server may consider purchasing time on a campus cluster or Amazon Web Services (Jackson et al., 2010), or applying for time through XSEDE (Towns et al., 2014).

The assembly process produces putative transcripts, known as contigs. Ideally a single contig is equivalent to a single isoform**,** but in practice a contig may represent an entire isoform (a complete transcript), part of an isoform (an incomplete transcript), or a chimera (a transcript consisting of two transcripts which are biologically independent). Therefore, after assembly, chimeric contigs must be identified and discarded, and remaining contigs must be annotated with the appropriate gene symbol. Some programming, using a scripting language such as Python (The Python Software Foundation, python.org), may be necessary to accomplish this. A set of protocols and scripts exists to aid in analysis of *de novo* assemblies (Brown et al., 2013), or the following protocol may be observed.

Initial analysis of a *de novo* assembly should include masking of repetitive sequences, to avoid false positives during chimera identification. RepeatMasker (Tarailo-Graovac and Chen, 2002) uses a database of known repetitive elements to substitute Ns or Xs for repetitive sequences.

Assembly annotation may be accomplished using BLAST (Altschul, 1990) or BLAT (Kent, 2002). The contigs should be compared to a well-annotated genome that is as closely related as possible. If no closely related genome exists and results using distantly related genomes are insufficient, a protein-protein comparison may be made instead of a nucleotide-nucleotide comparison, thereby eliding synonymous variations. Typically, multiple matches will be found for many genes. This is to be expected, as assemblers may not successfully differentiate between different isoforms, so that multiple contigs may represent different isoforms of the same gene. Additionally, a contig representing an incomplete transcript may match to one set of exons in a gene, while a different contig may match to a different set of exons, together completing the transcript. Chimeric contigs that match multiple genes should be removed from the assembly.

The BLAST or BLAT results may be used to rename individual contigs according to the gene they best match. Tools exist to aid in this process; for example, BioPython (Cock et al., 2009) provides tools to handle and rename sequences as Python objects. The BLAST or BLAT results may then further be used to identify chimeric contigs that match to multiple genes. Some apparently chimeric contigs may actually match two genes from the same gene family that have very similar sequences, and therefore do not need to be removed. To identify these apparently chimeric but actually legitimate contigs, a BLAST or BLAT search comparing the reference

genome to itself may be performed. Genes that match to other genes may be considered pseudo-chimeric, and contigs matching multiple pseudo-chimeric genes need not be removed from the assembly. Additionally, contigs with multiple matches of extremely different lengths (for example, one match ten times longer than the second match) need not be removed from the assembly, as differential gene expression tools will be able to choose the appropriate (longer) match and discard the other (shorter) match.

Non-exonic sequences may be retained in some contigs, comprising both untranslated region, intronic sequence from pre-splicing mRNA, and inter-genic sequence, now known to be pervasively transcribed though at lower levels than intra-genic sequence (Clark et al., 2011). Intron removal is a challenging proposal for *de novo* assemblers (Garber et al., 2011).

The differential expression tools described in the reference genome RNA-seq pipeline require a genome, not a transcriptome, as their reference. This makes them inappropriate for use with a *de novo* assembled transcriptome. However, an alternative differential expression tool, Cuffdiff, was designed to work with the genome-guided assembler Cufflinks, and will perform differential expression analysis against assemblies constructed by alternative assemblers (Trapnell et al., 2012).

2.5 CONCLUSIONS

Species from a marked diversity of taxa have been domesticated, from mammals to fish to birds, including both commonly studied species such as the dog and less commonly studied species such as the yak. Many domesticated species have phenotypes of biomedical or economic value for humans, making them important subjects of research. Compared to wild animals, domesticated animals are well suited for study with RNA-seq, a new technology for evaluating transcriptional activity across the entire genome. First, the striking phenotypic diversity of these species provides opportunities for comparison of traits of interest among individuals. Second, reduced genetic diversity within domesticated breeds results in increased statistical power in many studies. Third, the breeding of domesticated species is under human control, so samples may more easily be collected from individuals with specific phenotypes and at specific time points in their development. Finally, domesticated animals commonly receive veterinary care,

providing an opportunity for sample collection from individuals with well-characterized disease status or subject to advanced treatment. Domesticated animals remain critical for human well-being, and molecular genetic studies provide insights into the mechanisms involved in the regulation of the complex phenotypes for which these animals have been selected. Using this knowledge, we can not only advance human medicine, but also select animals better suited to the changing climate and to human needs.

While many of the most common domesticated species, such as dog, cow, pig, and chicken, have high quality genomes, other species have lower quality, fragmented NGS genomes, and still others are not yet sequenced. While the genomes of all domesticated species may well be sequenced in the coming decades, newly sequenced genomes can be expected to be subject to the limitations of NGS assemblies. RNA-seq is becoming a standard method for annotation of NGS genome assemblies, and its use in improving the annotations of the high quality genome assemblies produced using Sanger sequencing has been demonstrated.

In this review, we have discussed two strategies for the analysis of RNA-seq data in species with lower quality genome assemblies. Use of a mature, well-annotated genome from a closely related species may prove sufficient, especially if the alignment tool used is passed parameters to relax stringent alignments and tolerate an increased rate of sequence divergence. If even a closely related genome annotation is lacking, a *de novo* assembly may be constructed and used as a reference. Using either a reference genome or a *de novo* transcriptome assembly, differentially expressed genes may be called. This list of genes may be further analyzed to ascertain groups of differentially expressed genes with similar functions or networks of genes that are co-expressed. Therefore, even in the absence of the resources available for RNA-seq analysis of model species, RNA-seq analysis is a powerful tool for use in investigation of the genomic underpinnings of phenotypes in domesticated species.

CHAPTER 3. TRANSCRIPTOME ANALYSIS OF TAME AND AGGRESSIVE FOX PITUITARIES SUGGESTS A ROLE FOR PITUITARY CELL NETWORKS AND PSEUDOPODIA IN REGULATION OF ACTH RELEASE

3.1. INTRODUCTION

Domesticated species exhibit greatly reduced fearfulness, increased social tolerance, and increased resistance to stress compared to their ancestral wild species (Price, 2008; Driscoll et al., 2009; Campler et al., 2009). These behavioral changes are closely linked to a suite of physiological and morphological changes, including increased white spotting, curling of the tail, shortening of the cranium, and reduced reactivity of the hormonal stress response, which have together been referred to as the "domestication syndrome" (Hare et al., 2012; Wilkins et al., 2014). The closely correlated behavioral, endocrinological, and morphological changes across multiple species suggest that perturbations at the level of networks of genes, working together for specific biological effects, may be implicated in the domestication process. Therefore, we expect that understanding the biological basis of the endocrinological changes in one domesticated species may shed light on the mechanisms targeted during the domestication process more generally, and potentially also into the biological underpinnings of variation in resilience to stress in animals and humans (Hare et al., 2012).

The characteristic reduction in fearfulness of domesticated animals has been closely linked to reduced reactivity of the hypothalamic-pituitary-adrenal (HPA) axis, the hormonal cascade associated with the stress response in mammals (Trut et al., 2009; Hare et al., 2012; Wilkins et al., 2014). Specifically, reductions in circulating levels of adrenocorticotrophic hormone (ACTH), released by the pituitary, have been demonstrated in domesticated species (Trut, 2004; Gulevich et al., 2004; Kaiser et al., 2015). ACTH stimulates release of glucocorticoids by the adrenals, and reduction in glucocorticoids in domesticated species (possibly related to reduction in ACTH) has also been shown (Albert et al., 2008; Trut et al., 2009; Kaiser et al., 2015). Although the hypothalamus is classically considered the "top" of the HPA axis, the pituitary,

situated to amplify signals from the hypothalamus before release to the systemic circulation, is sometimes known as the “master gland” and may contribute significantly to regulation of the HPA axis.

Although the mechanisms regulating HPA axis attenuation in domesticated animals are not well understood, differences in domesticated and wild HPA axis function have been demonstrated in multiple animal models. Domesticated guinea pigs provide an ideal opportunity to compare domesticated and wild species, as wild cavvies can easily be kept in captivity under similar conditions. Compared to wild cavvies, domesticated guinea pigs display reduced aggression and increased sociopositive behaviors, such as play. In response to psychosocial stressors, their adrenal glucocorticoid responses are both decreased in magnitude and return to baseline after a shorter time (Kaiser et al., 2015). Similar findings have been demonstrated with experimentally domesticated Norway rats. Wild-caught rats were selected for either increased or decreased aggression (tameness) towards humans. After generations of selection, the rats selected for tameness displayed markedly reduced aggressiveness to humans as well as reduced glucocorticoid levels (Naumenko et al., 1989; Albert et al., 2008) and ACTH levels (Shikhevich et al., 2002). Despite the centrality of the pituitary to HPA axis function, changes at that level have not been extensively explored.

In perhaps the best known domestication experiment, farmed foxes have been selectively bred for more than fifty generations. One population of foxes was selected for lack of fear and sociability towards humans, while a second population of foxes was selected for aggression towards humans. A comparison of these tame and aggressive fox populations demonstrates similar changes in the HPA axis as those shown between domesticated guinea pigs and their wild ancestors, as well as between tame and aggressive rats (Trut, 1999; Trut et al., 2004; 2009). Compared to the aggressive line, the tame foxes display reduced basal ACTH levels (Trut, 2004; Gulevich et al., 2004), decreased ACTH response to stress, and shortened time to return to baseline ACTH levels (Oskina and Plyusnina, 2000), as well as markedly reduced glucocorticoid elevation and shortened time to return to baseline glucocorticoid levels after a stressor (Trut et al., 2004). However, the detailed molecular mechanism that led to this domestication phenotype is not known.

Inhibition of the HPA hormonal cascade at the level of the pituitary may occur in at least three ways, any or all of which could have been selected for during the domestication process: reduced synthesis of ACTH, reduced secretion of ACTH, and reduction of the number of cells that synthesize ACTH (Rizotti, 2010). An examination of the details of each of these processes will provide predictions for how they might vary in the wild compared to the domesticated pituitary. ACTH synthesis and secretion are triggered by the binding of corticotrophin releasing hormone (CRH) to CRH receptors in the anterior pituitary (Grammatopoulos, 2012). This process is potentiated by arginine vasopressin (AVP) binding (Denef, 2008). ACTH synthesis begins with the production of a proopiomelanocortin (POMC) precursor, which is regulated by a suite of transcription factors, including NURR77 (NGFI-β), NURR1, and NOR1 (Drouin, 2016). POMC is cleaved into ACTH by the PC1 and PC2 proprotein convertases in a series of reactions that also produce ß-lipotropin, ß-endorphin, and α-melanotropin (Benjannet et al., 1991), so these convertases may also be thought to regulate ACTH production. Therefore, the domesticated pituitary might contain different concentrations of CRH receptors, AVP, POMC, POMC-stimulating transcription factors, or POMC convertases compared to the aggressive pituitary.

Multiple mechanisms regulate ACTH secretion through exocytosis, but the main signal for release of hormone-filled secretory vesicles is a rise in calcium from both intra- and extra-cellular sources, and this rise in calcium is primarily regulated by cAMP (Burgoyne and Morgan, 2003). Some pituitary hormone-releasing cells may also facilitate the release of hormones into the systemic circulation by extending pseudopodia and becoming motile to contact blood vessels (Navratil et al., 2007). Additionally, pituitary endocrine cells are known to form networks, potentially via gap junctions, to coordinate hormone release (Hodson et al., 2012a; Hodson et al., 2012b). Therefore, the domesticated pituitary may demonstrate different levels of motility, cellular adhesion and communication (particularly via gap junctions), or cAMP activity as compared to the wild pituitary.

Finally, tame and aggressive pituitaries may display different population sizes of ACTH-secreting cells (corticotrophs), with a larger population associated with greater ACTH production and/or release. Stem cells in the pituitary differentiate into specialized hormone-producing cells, including corticotrophs, throughout life at a low rate (Florio, 2011). Therefore, key transcription factors involved in corticotroph cell fate determination, such as TBX19/TPIT and PAX7 (Pulichino et al., 2003; Budry et al., 2012), may differ in the wild and domesticated pituitary.

To identify mechanisms involved in the regulation of reduced reactivity of the HPA in domesticated animals, we analyzed gene expression in anterior pituitary of tame and aggressive farmed foxes. Previously, tame fox pituitaries have been shown to have comparable amounts of *POMC* mRNA and POMC protein to aggressive fox pituitaries (Gulevich et al., 2004), despite decreased circulating concentrations of ACTH. Therefore, we specifically tested whether the changes in ACTH release in tame foxes are associated with ACTH secretion (exocytosis) and/or corticotroph differentiation, rather than ACTH synthesis from *POMC* mRNA. Identification of gene networks regulating ACTH release in tame foxes will provide candidate mechanisms for appearance of behavioral characteristics of domestic animals (reduced fearfulness, increased social tolerance, and increased resistance to stress) and facilitate understanding a role of HPA in regulation of mammalian behavior.

3.2. METHODS

3.2.1 Animals and sample preparation

Animals were maintained at the experimental farm of the Institute of Cytology and Genetics (ICG) in Novosibirsk, Russia. All animal procedures at the ICG complied with standards for humane care and use of laboratory animals by foreign institutions. The study was approved by the Institutional Animal Care and Use Committees (IACUC) of the University of Illinois at Urbana-Champaign. Anterior pituitary tissue was dissected from 6 tame and 6 aggressive adult foxes. All foxes were sexually naive 1.5 year old males that were born between March and April of 2009 and raised under standard conditions. The samples were collected in August of 2010. Foxes were euthanized using sodium thiopental; immediately afterwards, the brain case was opened with a saw and the brain was removed, leaving the pituitaries *in situ*. The anterior and posterior pituitaries were immediately dissected out and placed into RNAlater (Life Technologies, Grand Island, NY). The samples were stored at -70°C. Total mRNA was extracted using RNeasy Mini Kit (Qiagen, Valencia, California), according to the manufacturer's protocol.

3.2.2. Sequencing and quality analysis

One microgram of high quality RNA from each sample was used for sequencing. Stranded RNA-seq libraries were prepared using TruSeq SBS Sequencing kit version 3 (Illumina, San Diego, CA). Libraries were barcoded and pooled, and sequenced on two lanes on a HiSeq2500. Reads were single-end, stranded, and 100 nt in length. Sequencing results were processed by CASAVA 1.8 (Illumina, San Diego, CA). The primary sequencing data have been deposited in the GeneBank Sequence Read Archive (SRA) under accession number SRP107226.

Data quality, including base quality per position across reads, GC content, and distribution of sequence length, was initially assessed with FastQC (Andrews, 2014). Reads were processed with flexbar (Dodt et al., 2012) in two passes: the first to trim adapters, remove low quality reads, and remove reads less than 35 bp in length, and the second to remove polyA tails. Subsequently, reads that mapped to fox mitochondrial DNA sequences from NCBI (accession numbers JN711443.1, GQ374180.1, NC_008434.1, and AM181037.1) using Bowtie2 (Langmead and Salzberg, 2012) were discarded. Similarly, any remaining reads that mapped to ribosomal DNA sequences were discarded.

3.2.3. Evaluation and validation of differential expression

Reads were aligned to *Canis familiaris* 3.1 (CanFam3) with Ensembl 1.79 annotation using TopHat 2.0.13 (Kim et al., 2013). The Ensembl annotation used is available at *ftp://ftp.ensembl.org/pub/release-79/gtf/canis_familiaris/Canis_familiaris.CanFam3.1.79.gtf.gz*. The fox and the dog ancestor, the gray wolf, diverged approximately 7-10 million years ago (Wayne, 1993). Therefore, alignment of fox RNA sequences to the dog genome risked a reduction in mapping sensitivity due to a high incidence of mismatches subsequent to divergence between the genomes of the two species. To account for this expected variation between our fox sequences and the CanFam3.1 reference genome, we adjusted the parameters of our aligner, TopHat2, to allow an increased rate of mismatches and increase alignment sensitivity. Specifically, the following TopHat2 parameters were added to the command line, to allow a more robust cross-species alignment with the dog genome: *-N 3 –read-edit-dist=3*. These parameters increased the number of nucleotide mismatches and maximum edit distance allowed per read. Gene expression levels were assessed using HTSeq-count 0.6.1p1 (Anders, 2015).

Differentially expressed genes between samples from tame versus aggressive foxes were evaluated using DESeq2 1.10.1 (Love et al., 2014). Normalized gene counts (FPKMs) were assessed with CuffNorm 2.2.1 (Trapnell et al., 2013) using quartile normalization.

RNA from all 12 samples used for sequencing was also used for quantitative real time PCR (RT-qPCR). Gene assays used were *AKAP7, ANKRD6, BRCA2, CDH4, ITGA1, LAMA2, NELL2, SPOCK1,* and *THBS4*. Primers for qPCR (Table SI-3.1), placed in adjacent exons and spanning exon-intron boundaries where possible, were designed with Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012). cDNA was synthesized from 1ug of RNA with random hexamers using the ThermoScript RT-PCR Kit (Invitrogen, Carlsbad, CA). Real time quantitative PCR was performed on 96 well plates using an Applied Biosystems StepOnePlus Real Time PCR System (Applied Biosystems, Foster City, CA). Wells were loaded with 10 µl solution containing 5 µl SYBR Green reagent (Applied Biosystems), 0.1µg cDNA, and either 2mol forward and 2 mol reverse prmers for the control gene or 0.6 mol forward and 0.6 reverse primers for the test genes.. All samples were analyzed in triplicate. Initial denaturation was performed at 95°C for 20 min, followed by 40 cycles of 95°C for 3 s and 60°C for 3 s, followed by a melting curve analysis at 60°C-95°C. Gene expression was normalized to the expression of *SDHA*, a gene that did not show differential expression between tame and aggressive foxes and that had comparable expression between individuals in the RNA sequencing results. Mean ΔΔ-Ct values for tame foxes were compared to mean ΔΔ-Ct values for aggressive foxes for each gene using the Mann-Whitney test.

3.2.4. Functional analysis

Functional analysis was performed for each tissue with ClueGO version 2.2.5 (Bindea et al., 2009); Gene Ontology (GO) files were downloaded on May 25, 2015. The ClueGO evaluation was performed against a background list of genes with a mean expression level in pooled tame and aggressive pituitary samples reported to be > 0 in that tissue by DESeq2. GO term fusion was used when necessary to reduce large numbers of GO terms to results that could be analyzed. Resulting p-values were adjusted using the Benjamini-Hochberg correction. Enriched GO terms were reviewed for association with stimulus of hormone release, POMC regulation and processing, and corticotroph differentiation.

3.2.5. Weighted gene co-expression network analysis

Modules of genes with highly correlated expression patterns were sought using weighted gene co-expression network analysis. We expect these modules to correspond to networks of genes that interact in shared biological processes (Langfelder, 2008). The full set of genes expressed in the pituitary reads was filtered to remove genes with fewer than 10 FPKMs in each of 3 or more samples. We then constructed unsigned weighted gene co-expression modules using the WGCNA package in R (Langfelder, 2008). The blockwiseModules function was run with the Pearson correlation coefficient and a soft thresholding power of 12. The resulting gene groups, or modules, were named by assigning them arbitrary colors. The first principle component of each module was designated its eigengene. Eigengenes for each module were divided into samples from tame and aggressive foxes; these two groups were compared with a Student's *t* test and their p values were adjusted using the Benjamini-Hochberg correction. Modules selected for investigation were analyzed for gene enrichment using ClueGO as described above. Additionally, highly connected "hub" genes in these modules were identified using Cytoscape (Shannon et al., 2003) to select the genes with the largest number of connections to other genes in the same network, with the expectation that these "hub" genes would represent regulatory genes with control over many genes in that module.

3.2.6. Alternative splicing analysis

Differences in alternative splicing frequencies between tame and aggressive pituitary reads were explored with rMATS (Shen et al., 2014). Because rMATS requires reads of equal lengths, post-filtered reads were further filtered using SAMtools 1.1 (Li et al., 2009) to remove all reads shorter than 100 nucleotides in length. BAM files (previously aligned to CanFam3.1 for differential gene expression calling) and Ensembl annotation, version 1.79, were used with rMATS to investigate reads crossing splicing junctions. Differences in frequencies of reads including or skipping specific exons were used to identify exons that were skipped significantly more frequently in tame or aggressive pituitary reads. P values were adjusted using Benjamini-Hotchberg. The list of differentially skipped exons was filtered to remove exons covered by fewer than 10 reads in either tame or aggressive samples. Differences in skipping were reported

as mean percentage skipped exons, in other words, (mean number of reads skipping the exon)/(mean number of reads skipping the exon + mean number of reads including the exon). Exons identified as significantly more frequently skipped in tame or aggressive samples were further investigated by searching for their amino acid sequence using their nucleotide sequence in blastx (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify isoforms that included or did not include the exon in question. Conserved domains that the exon in question fell within were identified using NCBI's Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd/).

3.3. RESULTS

3.3.1. Gene expression in the anterior pituitary of tame and aggressive foxes

Sequencing of fox pituitaries produced 206 million reads total for six tame samples (194 million reads after filtering) and 200 million reads total for six aggressive samples (191 million reads after filtering) (Table SI-3.2). The filtered fox reads were mapped to the CanFam3 reference with a mean alignment rate of 78.98% (Table SI-3.3). The rate of multiple alignments was low, with a mean of 2.64%. Analysis of filtered pituitary reads identified 18,940 distinct genes expressed in this tissue. A gene ontology analysis was performed on the 0.1% most highly expressed genes (180 genes). Of the most significantly enriched terms (FDR < 10.0e-9; 9 terms), 8 terms were related to RNA processing or splicing (Table SI-3.4). Analysis of gene expression level identified 11 genes with notably higher expression (Figure SI-3.1). One of these genes, *EEF1A1,* was significantly more highly expressed in the tame than the aggressive fox pituitary. The most highly expressed gene by an order of magnitude was *POMC,* which codes for the precursor protein for ACTH and other peptides released by the pituitary.

Table 3.1. Highly expressed genes in fox pituitary, including gene symbol and normalized counts, averaged over both tame and aggressive samples.

Gene Symbol	Normalized Count
POMC (proopiomelanocortin)	1,270,886
SPARCL1 (SPARC like 1)	122,386
EEF1A1 (eukaryotic translation elongation factor 1 alpha 1)	58,368
GNAS (GNAS complex locus)	53,412
DDX5 (DEAD-box helicase 5)	51,106
ENSCAFG00000000101	47,208
SRRM2 (serine/arginine repetitive matrix 2)	46,389
DDX17 (DEAD-box helicase 17)	42,528
PCSK2 (proprotein convertase subtilisin/kkexin type 3)	38,611
PSAP (prosaposin)	32,025
HNRNPH1 (heterogeneous nuclear ribonucleoprotein H1)	29,665

DESeq2 analysis of filtered pituitary reads identified 346 differentially expressed genes (DE genes) (FDR < 0.05). Of these, 191 genes (55%) were up-regulated in aggressive pituitary, and 155 (45%) were up-regulated in tame pituitary. Log_2 fold changes among DE genes ranged from -0.81 to 0.84 (mean -0.043, SD 0.37) (Table SI-3.5). Clustering analysis of DE genes clearly differentiated tame and aggressive samples (Figure 3.1). Additional visualization using a volcano plot showed that the distribution of log_2 fold changes and significance was as expected, and PC analysis grouped tame samples separately from aggressive samples (Figures SI-3.2 and

3.3).The most significantly differentially expressed gene, and the gene with the largest $\log_2$ fold change, was *GRK7,* an important regulator of CRH receptors (Grammatopoulos, 2012). *GRK7* was up-regulated in aggressive pituitaries.

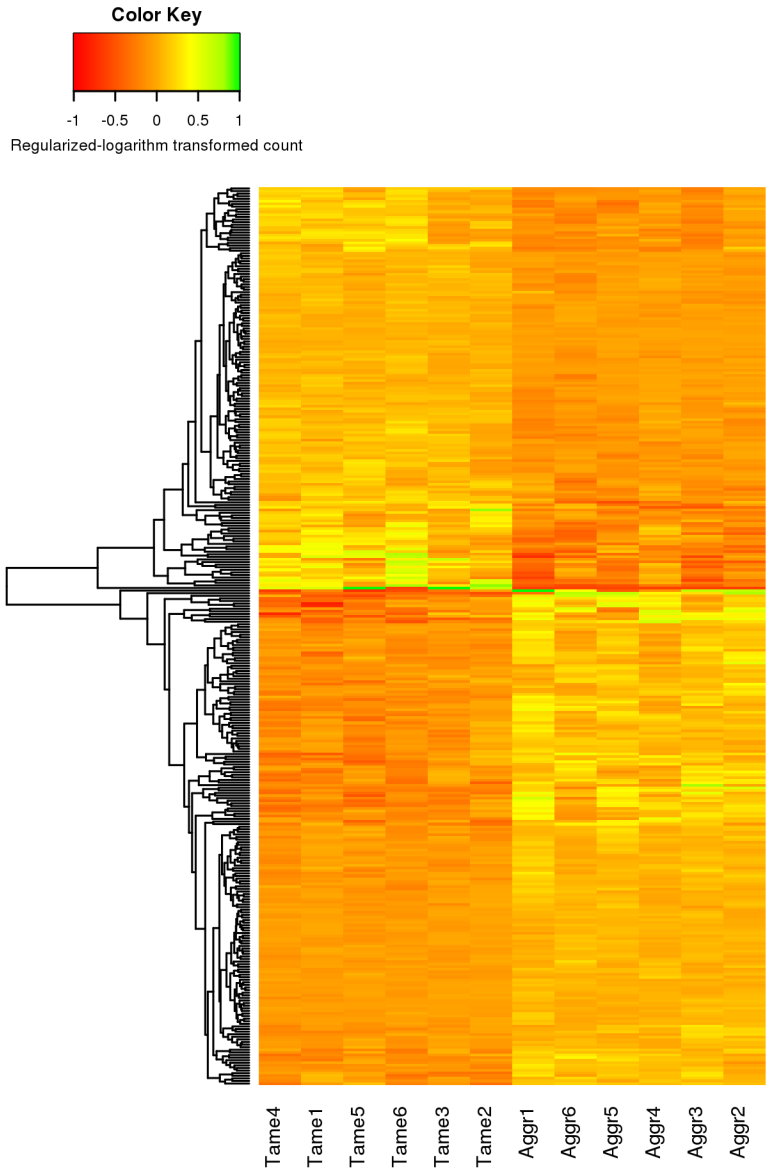

Figure 3.1. Heat map of differentially expressed genes (346) in tame versus aggressive fox pituitary tissue (FDR < 0.05).

Nine genes (AKAP7*, ANKRD6, BRCA2, CDH4, ITGA1, LAMA2, NELL2, PDE7B,* and *SPOCK1)* with a range of FDR values and $\log_2$ fold changes were selected for validation with quantitative PCR. Analysis with RNA-seq data indicated that five of these genes were up-regulated in aggressive foxes and four in tame foxes. Of these nine genes, eight showed differential expression in the same direction (up-regulation or down-regulation) as the RNA-seq results (see Figure SI-3.4 and Table SI-3.6). The smallest p value (p=0.06) was obtained for *SPOCK1,* the gene with the largest $\log_2$ fold change and the second smallest FDR (per RNA-seq results) tested. One gene, *ITGA1,* was shown to be up-regulated in tame foxes according to the RNA-seq results, but down-regulated in tame foxes according to the quantitative PCR results, though this result was not statistically significant. However, this gene had a very small $\log_2$ fold change in the DESeq2 output (0.21). Comparison of RNA-seq and qPCR results demonstrated similar expression levels, with qPCR results nearer to statistical significance in genes that are more significantly differentially expressed in the RNA-seq results.

3.3.2. Functional analysis with ClueGO

ClueGO analysis was performed on the 328 differentially expressed genes that had an associated gene symbol; a background list was used of 18,940 genes for which DESeq2 demonstrated expression in anterior pituitary. The differentially expressed genes were associated with 394 GO terms, of which 74 GO terms were significantly enriched (FDR < 0.05) (Table SI-3.7). The significant terms containing the largest percentage of differentially expressed genes were *cell motility involved in cerebral cortex radial glia guided migration* (75%, 3 out of 4 total genes differentially expressed), *regulation of dendritic cell differentiation* (50%, 3 out of 6 total genes differentially expressed), and *radial glial cell differentiation* (37.5%, 3 out of 8 total genes differentially expressed). Overall, among the significantly enriched terms, the percent of genes in the term which were differentially expressed in the pituitary ranged from 4.28-75.0% (median 9.17%) (Table SI-3.7); for the most significantly enriched terms (FDR < 0.01), the percent of differentially expressed genes ranged from 8.3-50%.

The most significantly enriched terms in the GO analysis at FDR < 0.01 (Table 3.2) cluster into five groups with 11 terms total. Three of these groups are represented by a single GO term

each (*pseudopodium organization, regulation of dendritic cell differentiation,* and *granulocyte chemotaxis*). The most significantly enriched GO term relates to *pseudopodia*, a generic term for different types of membrane projections known to function in cellular motility and endocrine hormone release in gonadotroph cells in the anterior pituitary (Navratil et al., 2006), and possibly other cell types in that organ. This GO term includes the gene *CDC42EP2,* which induces pseudopodia formation (Hirsch et al., 2001); *SRGAP2* and *SRGAP3,* which regulate cell migration through modification of cell protrusion morphology (Guerrier et al., 2009; Endris et al., 2011); and *WASH1,* which promotes pseudopod extension (Zech et al., 2011).

The second of the three single-term GO groups at FDR < 0.01 is *regulation of dendritic cell differentiation,* and includes genes for two galectins related to cellular motility (*LGALS3, LGALS9* [Sano et al., 2000; Bi et al., 2011]) and a gene related to the cross-presentation of antigen (*TMEM176B* [Segovia et al., 2014]]).

The last term of the single-term GO groups, *granulocyte chemotaxis,* contains two genes for receptors for interleukin 17A (IL-17), a pro-inflammatory cytokine, *IL17rA* and *IL17RC* (Gaffen, 2009). This term also contains a galectin, *LGALS3*; an integrin, *ITGA1; VAV3,* a gene associated with cell migration and integrin signaling (Sachdev et al., 2002; Sindrilaru et al., 2009) and implicated in the formation of lamellipodia and membrane ruffles (Movilla and Bustelo, 1999); *PDE4B,* a phosphodiesterase, related to cAMP regulation (Houslay and Adams, 2003); and *THBS4*, related to cellular adhesion, migration, and attachment (Adams, 2001; Mosher and Adams, 2012).

Four GO terms among the most significant GO terms are related to the JAK-STAT pathway (*JAK-STAT cascade, STAT cascade, regulation of JAK-STAT cascade, regulation of STAT cascade)*. The JAK-STAT pathway is associated with the interleukin immune response (Subramaniam et al., 1999; Sansone and Bromberg, 2012), cell migration (Pocha and Montel, 2014), and the cAMP pathway (Chesnokova and Melmed, 2002).

Four other related GO terms from Table 3.2 (SRP-dependent cotranslational protein targeting to membrane, cotranslational protein targeting to membrane, protein targeting, and selenocysteine metabolic process) are related to protein targeting and signal recognition proteins (SRP). Protein targeting is a process by which proteins are translocated to an organelle identified by SRPs during translation.

Table 3.2. GO terms enriched in genes differentially expressed in tame versus aggressive pituitary tissues at FDR < 0.01. **STAT3* is not associated with GO:1904892. ***ZFAND2B* is not associated with GO:0016259. Groups with unique GO terms are highlighted.

Group	GO ID	Term description	Adjusted p-value	Differentially expressed genes annotated with this term	% genes in this term differentially expressed
1	GO:0031268	pseudopodium organization	0.007	*CDC42EP2, SRGAP2, SRGAP3, WASH1*	28.6
2	GO:0007259	JAK-STAT cascade	0.007	*ASPN, CISH, ERBB4, FLRT1, FLRT3, IL6ST, LRRC4B, PTGER4, SOCS2, STAT3*, TNFRFSF1A*	8.3
	GO:0097696	STAT cascade	0.007		8.3
	GO:0046425	regulation of JAK-STAT cascade	0.009		9.2
	GO:1904892	regulation of STAT cascade	0.009		9.2
3	GO:2001198	regulation of dendritic cell differentiation	0.008	*LGALS3, LGALS9, TMEM176b*	50
4	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	0.008	*RPL17, RPL23a, RPL26, RPL4, RPL7a,RPS24, RPS3A, ZFAND2B***	10.7
	GO:0006613	cotranslational protein targeting to membrane	0.008		10.5
	GO:0045047	protein targeting to ER	0.008		10.4
	GO:0016259	selenocysteine	0.009		11.7

Group	GO ID	Term description	Adjusted p-value	Differentially expressed genes annotated with this term	% genes in this term differentially expressed
		metabolic process			
5	GO:0071621	Granulocyte chemotaxis	0.010	*IL17RA, IL17RC, ITGA1, LGALS3, NCKAP1L, PDE4B, TGFB2, THBS4, VAV3*	10.1

Table 3.2 (cont.)

Exploration of additional terms that were significantly enriched in differentially expressed genes (Table SI-3.7) identified terms associated with cell migration (including pseudopodia and cytoskeletal organization), cell adhesion, exocytosis (including formation of exocytotic vesicles), cAMP regulation, and calcium regulation. Specific terms include: *neutrophil migration, negative regulation of neuron migration, granulocyte migration, pseudopodium assembly, cell motility involved in cerebral cortex radial glia guided migration, neutrophil chemotaxis, myeloid leukocyte migration, regulation of neuron migration, leukocyte chemotaxis,* and *glial cell migration*. Several terms are also included relating to cellular adhesion and cellular interactions with the extra-cellular matrix, critical components of cellular migration: *extracellular structure organization, regulation of integrin activation, extracellular matrix organization, integrin activation,* and *positive regulation of cell adhesion mediated by integrin.* Finally, several terms relate to biosynthesis of cAMP, an important signaling molecule in ACTH release by exocytosis (Reisine et al., 1985): *cAMP metabolic process, cAMP biosynthetic process, positive regulation of cAMP biosynthetic process, positive regulation of cAMP metabolic process,* and *regulation of cAMP biosynthetic process.* These terms relating to cAMP biosynthesis include similar gene lists, comprising the genes *ADCY7, ADCY8, AVP, CALCRL, FLNB, NPR3, P2RY11,* and *PTGER4* (Table SI-3.7).

No terms related to CRH-dependent regulation of *POMC* transcription and processing of the POMC protein or differentiation of corticotrophic cells were found in the list of GO terms enriched for differentially expressed genes.

3.3.3. Functional analysis of genes involved in exocytosis, POMC regulation, and cell differentiation

Differentially expressed genes related to exocytosis, POMC regulation, and cell differentiation were sought, to investigate which of these processes were more extensively represented in the differences in gene expression between tame and aggressive fox pituitaries. The gene with the largest $\log_2$ fold change was *GRK7* (up-regulated in aggressive pituitary), a member of the G-protein coupled receptor kinase family, a family which is known to serve as important regulators of phosphorylation of the CRH receptor, activation of which triggers ACTH synthesis and release (Grammatopoulos, 2012). *AVP*, which codes for arginine vasopressin, a neuropeptide with paracrine activity in the anterior pituitary known to potentiate ACTH release (Liu et al., 1994) but not synthesis (Levin et al., 1989), was up regulated in the tame pituitary. Differentially expressed genes associated with membrane proteins of exocytotic vesicles included *VAMP1* (Calakos et al., 1994), and two synaptotagmins (*SYT10* and *SYT13* [Brose et al., 1992]). Therefore, tame and aggressive pituitaries demonstrated potential differences in regulation of ACTH release via CRH receptor regulation, vasopressin signaling, and exocytotic vesicle formation, as well as potential differences in ACTH synthesis via CRH receptor regulation.

Regulation of cAMP was specifically investigated, as cAMP serves as a major cellular signal for exocytosis. Differentially expressed genes associated with cAMP regulation included two members of the adenylyl cyclase family of genes that synthesize cAMP (*ADCY7* and *ADCY8*); two members of the phosphodiesterase family of genes that break down cAMP (*PDE4B* and *PDE7B*); G-protein receptors that activate adenylyl cyclase (*CALCRL /CRLR* [Smith et al., 2002] and *P2RY11* [Qi et al., 2001]); a G-protein receptor which inhibits adenylyl cyclase (*NPR3* [Anand-Srivastava, 2005]); a guanine nucleotide exchange factor that functions in cAMP-dependent exocytosis (*RAPGEF4/EPAC2* [Sugawara et al., 2016]); and a scaffolding protein involved in cAMP compartmentalization, *AKAP7/AKAP15* (Kritzer et al., 2012). Additionally,

the previously discussed differentially expressed gene *AVP* is known to act through the cAMP/PKA pathway in the anterior pituitary (Liu et al., 1994) in its effects on *POMC* regulation. Therefore, differences in cAMP regulation are represented in multiple ways in this data set, from synthesis to metabolism to compartmentalization.

Notably, genes important in CRH-dependent regulation of POMC transcription, such as *NGFIB, NURR1*, and *NOR1* (Bonfiglio et al., 2011), were not differentially expressed, nor were genes associated with processing of POMC into ACTH, such as *PC1* and *PC2* (Benjannet et al., 1991), *CPE* (Cool and Loh, 1998), and *PAM* (Ciccotosto et al., 1999).

Genes associated with pituitary and specifically corticotroph (ACTH-producing) cell differentiation were investigated, to determine the likelihood that tame and aggressive pituitaries have different population sizes of corticotrophic cells. No genes directly associated with corticotroph differentiation were identified as significantly differentially expressed. However, *TBX19/TPIT*, a gene known to be essential in the establishment of the POMC-expressing lineage of corticotrophs and melanotrophs (Pulichino et al., 2003), was significantly up-regulated in aggressive pituitary before adjustment at p=0.047 (FDR = 0.32). *PAX7,* a gene known to suppress corticotroph differentiation in favor of melanotroph differentiation, (Budry et al, 2012), was also significantly up regulated in aggressive pituitary only before adjustment (p=0.03, FDR=0.28).

3.3.4. Weighted gene co-expression network analysis results

Weighted gene co-expression network analysis constructed 66 gene modules with activity in fox anterior pituitary. Two modules, Antique White 4 and Pale Turquoise, had significantly (FDR < 0.10) different eigengenes between tame and aggressive samples (see Table 3.3). Both modules had a negative *t*-statistic, indicating that they are more active in aggressive than tame samples.

Table 3.3. Modules of co-expressed gene networks with significantly different activation in tame versus aggressive fox pituitary tissue.

Module name	*t*-statistic	FDR	Number of genes in module	Percent genes in module differentially expressed
Antique White 4	-4.2	0.09	43	18.6
Pale Turquoise	-4.1	0.09	80	20.0

The Antique White 4 module had 8 differentially expressed genes in total, including *SYT10,* a calcium sensor involved in exocytosis of secretory vesicles (Cao et al., 2011); *RNPEP,* an aminopeptidase (Piesse et al., 2002); and two genes associated with apoptosis, *CPPED1/CSTP1* (Zhuo et al., 2013) and *TRAF4* (Sax and El-Deiry, 2003). (For differentially expressed genes, see Table SI-3.5; for genes in WGCNA modules, see Tables SI-3.8 and SI-3.9.) The Antique White 4 module had no enriched GO terms, possibly due to the small number of genes (43) it contains. Hub (i.e., highly connected) genes in this module included an aminopeptidase, *RNPEP* (Piesse et al., 2002), which was both the most highly connected gene and up-regulated in the aggressive pituitary; *ZIC5*, associated with neural crest cell differentiation (Nakata et al., 2000; Inoue et al., 2004); *SUPT6H/SPT6,* a transcription elongation factor (Diebold et al., 2010); *KLHL2*, involved with cell membrane process formation (Williams et al., 2005); and *ATG5,* involved with autophagic vesicle formation (Walczak and Martens, 2013). Additional genes in this module associated with exocytosis, cell migration, and cell-cell adhesion include *ANXA7* (Caohuy et al., 1996), *LYN* (Malik et al., 2008; Nakata et al., 2006), and *RAP1B* (Mandell et al., 2005); both *LYN* and *RAP1B* are also associated with integrin signaling (Malik et al., 2008; Mandell et al., 2005). Overall, this module demonstrated relationships to processes associated with exocytosis, apoptosis, vesicle formation, cell migration, and cell-cell adhesion.

The Pale Turquoise module had 16 differentially expressed genes. These included a number of genes associated with exocytosis and migration, such as genes associated with the anchoring or biosynthesis of cAMP (*AKAP7* [Kritzer et al., 2012] and *FLNB* [del Valle-Pérez et al., 2010]); *AVP,* a signaling hormone for ACTH release (Denef, 2008); an integrin and a laminin, two proteins known to work together in the formation of pseudopodia (*ITGA6* [Mercurio et al., 2001]

and *LAMA2* [Gu et al., 2001)]); and a cadherin-associated cell-cell adhesion gene, *PTPRT* (Besco et al., 2006). There were two significantly ($p < 0.01$) enriched GO terms for this module, *regulation of cytokinesis* and *intrinsic apoptotic signaling pathway* (Table SI-3.10). Hub genes in this module included genes associated with puberty (*WDR11 [*Hyung-Goo and Layman, 2011]) and carnosine peptidase activity (*CNDP1/CN1* [Teufel et al., 2003]). Hub genes that were also differentially expressed included *LAMA2*, *PTPRT,* and *WLS* (Wnt protein regulation and secretion, Bänziger et al., 2006). Therefore, this module also had strong associations to genes involved in exocytosis and migration, as well as apoptosis.

Notably, the Brown module, which showed increased activation in tame samples only before correction ($p = 0.03$, FDR = 0.20) suggested functionality related to alternative splicing of mRNA. This module was associated with three GO terms significantly enriched for module genes that were related to this function, *mRNA 3'-splice site recognition, regulation of RNA splicing,* and *alternative mRNA splicing, via spliceosome.* The Brown module additionally contained multiple hub genes associated with mRNA alternative splicing and regulation, including *TSEN2/SEN2, TNKS1BP1/TAB182,* and *PRCC* (Trotta et al., 1997; Skalsky et al., 2001; Lau et al., 2009). The mRNA splicing functionality in this module may be associated with cell motility, as significantly enriched GO terms in the Brown module also included the terms *lamellipodium organization, ruffle organization, regulation of fibroblast migration, regulation of focal adhesion assembly, wound healing/spreading of cells,* and *negative regulation of cell-substrate adhesion.* Lamellipodia are pseudopodia formed on the leading edge of migrating cells (Krause and Gautreau, 2014); membrane ruffling allows the cell to form a motile surface to network with other cells (Ridley, 1994). The Brown module had a positive *t*-statistic of 2.7, indicating possible greater activation in tame than aggressive samples.

Overall, gene network analysis with WGCNA demonstrated two modules, Antique White 4 and Pale Turquoise, with significantly increased activity in aggressive samples. These modules were associated with exocytosis, cell migration, and cell adhesion; both modules contained genes suggestive of the importance of integrins in these functionalities. Both modules also indicated the importance of apoptosis, suggesting that increased cellular turnover may have some significance in differing pituitary function between the lines. Finally, analysis of the Brown module suggested the importance of alternative splicing of mRNA in anterior pituitary, which

may or may show greater activity in the tame population. For this reason, an alternative splicing analysis was undertaken.

3.3.5. Alternative splicing analysis

rMATS was used to identify 36 genes with exons skipped at different frequencies in tame compared to aggressive fox pituitary reads at FDR < 0.05 (Table SI-3.11), and 12 genes at FDR < 0.001 (Table 3.4). Exon counts for these 36 genes with significantly differentially skipped exons ranged from 4-46 exons total (median 13, SD 9.3). Of the genes with differentially skipped exons, only one (*MTHFSD,* an RNA-binding protein associated with amyotrophic lateral sclerosis [MacNair et al., 2016]) is itself differentially expressed in pituitary reads. Investigation of GO terms revealed that 9 terms enriched for differentially expressed genes also contained genes with differentially skipped exons. Of these, four terms contained multiple genes with differentially skipped exons: *establishment of protein localization to organelle (ABLIM3, SH2D4A, POT1;* this term is functionally related to the highly significantly enriched group of four terms related to protein targeting); *gliogenesis (ADAM22, NFIB);* and *extracellular matrix organization* and *extracellular structure organization* (both containing *KLKB1* and *MFAP5*). Nine genes with differentially skipped exons were also included in WGCNA modules associated with the state of being tame, including three genes in the Brown module (*DAGLB, MYCBPAP,* and *STAUI*). GO enrichment analysis of the genes with differentially skipped exons revealed no enriched terms.

Table 3.4. Genes containing exons skipped at different frequencies in tame compared to aggressive fox pituitary reads (FDR < 0.001), including conserved domains overlapping the skipped exon.

Gene	Exon location	Percent reads skipping exon (tame)	Percent reads skipping exon (aggr)	FDR	Skipped in isoforms	In conserved domains
MFAP5	chr27:37,034,668-37,034,704	46	0	0	X2, X3	MAGP
SPICE1	chr33:17787918-17788056	13	0	2.7 5E-10	None	SPICE
MECR	chr2:71398391-71398452	0	9.8	1.3 9E-07	None	ETR, Qor
SPATA20	chr9:26,472,931-26,473,143	12	1.9	5.5 7E-06	X4	YyaL
KLKB1	chr16:44493733-44493873	20	0	7.9 8E-00	None	Tryp_SPc
THYN1	chr5:996686-996755	12	2.3	1.8 1E-05	None	Eve
FBRSL1	chr26:542234-542303	13	2.8	1.2 9E-04	None	None
TRIM65	chr9:4716736-4716911	14	2.8	1.4 0E-04	None	None
TMEM198	chr37:26056452-26057028	2.6	0.31	3.2 2E-04	None	DUF4203

Gene	Exon location	Percent reads skipping exon (tame)	Percent reads skipping exon (aggr)	FDR	Skipped in isoforms	In conserved domains
GANC	chr30:9295384-9295464	10	1.8	3.60E-04	None	GH31_GANC_GANAB_alpha, Glyco_hydro_31
SORBS1	chr28:8892641-8892725	67	93	3.76E-04	X7, X8, X12	None
TCFL5	chr24:46680735-46680982	7.1	1.3	8.40E-04	None	None

Table 3.4 (cont.)

Of the genes listed in Table 3.4, *MFAP5/MAGP-2* promotes cell adhesion and spreading via integrin binding (Gibson et al., 1999), and *SORBS1/CAP* is involved in integrin-mediated cell adhesion (Ribon et al., 1998). Exons alternatively spliced at FDR < 0.05 in genes related to cell migration, cell-cell adhesion, and calcium signaling include: *ADAM22,* a disintegrin (Abaco et al., 2006); *CFAP97/HMW*, associated with cilia and flagella (Soulavie et al., 2014); *EFMA5*, integrin-dependent cell-cell adhesion (Davy and Robbins, 2000); *ENKUR*, calcium signaling and flagella (Sutton et al., 2004); *ERG*, adhesion and cell morphology change (McLaughlin et al., 2001); *UNC5D*, cell-cell adhesion/repulsion and motility (Jarjour et al., 2011). Exons alternatively spliced at FDR < 0.05 in genes related to RNA processing include *ELAVL4* (Ince-Dunn et al., 2012) and *FBRSL1* (Baltz et al., 2012). Additional interesting genes just above significance level were *RPH3AL/NOC2* (p=0.10), associated with exocytosis (Matsunaga et al., 2017), and *POLK* (p=0.07), a DNA repair polymerase (Ogi and Lehmann, 2006).

3.4. DISCUSSION

RNA sequencing of pituitary tissue from tame and aggressive foxes identified expression differences in genes related to cAMP regulation, pseudopodia formation, and cell-cell adhesion.

These findings suggest that the pituitaries of experimentally domesticated foxes may differ from the pituitaries of a lineage of foxes selected for increased aggression in regulation of exocytosis. Specifically, cAMP signaling for vesicle release may be modified in the tame foxes to reduce the release of ACTH during the stress response. Additionally, tame fox pituitary cells may differ in their motility or their ability to form networks for cell-cell communication, potentially to coordinate ACTH release, as indicated by differences in the expression of genes related to the organization of pseudopodial membrane extensions and cell adhesion. Notably, significant differences in lineage specification markers were not found between the two populations, suggesting that increased differentiation of pituitary cells involved in the HPA axis did not contribute to the differences in function between the tame and aggressive pituitary. However, non-significant differences were seen, and due to the small relative size of the population of actively differentiating cells in the anterior pituitary (Florio, 2011), significant differences may be difficult to detect in a mixed sample of cell populations such as this one. Additionally, one gene co-expression module suggested differences in alternative splicing functionality between tame and aggressive fox lines, and differentially skipped exons were identified in genes associated with cell migration, cell-cell adhesion, and calcium signaling. Therefore, differential splicing in the tame fox pituitary may affect similar functionality as do gene expression differences.

3.4.1. Alignment and pituitary gene expression

POMC was by far the most highly expressed gene in the pituitary, with an expression level an order of magnitude higher than the next most highly expressed gene, *SPARCL1,* a gene with poorly understood function that appears to code for an extracellular matrix-associated protein (Jakharia et al., 2016). The robust expression of *POMC* was expected in anterior pituitary tissue, as it codes for a prohormone that is spliced into the primary hormonal product of both melanotrophs and corticotrophs (Drouin, 2016). *PCSK2*, a convertase for processing *POMC* (Chrétien and Mbikay, 2016), was also highly expressed. The third and fourth most highly expressed genes, *EEF1A1,* involved in anchoring mRNA in pseudopodial cell projections (Liu et al., 2001), and *GNAS*, an adenylyl cyclase regulator (Weinstein et al., 2004), represented interesting findings in light of the importance of pseudopodium formation and cAMP regulation in differences between tame and aggressive fox pituitary function described by other analyses in

this study. Of particular interest was the significant up-regulation of *EEF1A1* in the tame fox pituitary. Other highly expressed genes were related to transcription and mRNA splicing (*DDX5/P68, SRRM2, DDX17,* and *HNRNPH1* [Jurica et al., 2002; Lin et al., 2005; Fuller-Pace and Ali, 2008; Russo et al., 2010; Wang et al., 2012]) . Along with the highly enriched GO terms associated with RNA processing and splicing for the top 0.1% most highly expressed genes, these findings suggest that the fox pituitary is highly active in RNA production. This enrichment is noteworthy due to the association of the Brown module with mRNA splicing functionality and the differentially spliced exons identified in the two populations of foxes.

Differential expression analysis comparing tame to aggressive samples demonstrated moderate $\log_2$ fold changes, ranging from -0.81 to 0.84. This result is consistent with the outbred genetic background of these lines of foxes (Johnson et al., 2015; Kukekova et al., 2017, in review).

3.4.2. Pseudopodia, cell motility, and networks

Gene ontology analysis identified *pseudopodium organization* as the most significant term associated with differentially expressed genes in pituitary tissue. In addition to the differentially expressed genes included in this GO term, several genes known to be important for pseudopodium formation and functioning are not included, but are differentially expressed between pituitary samples of tame and aggressive foxes. These genes code for laminins (*LAMA2, LAMA5*), a major component of the extra-cellular protein matrix, and integrins (*ITGA1, ITGA6*), transmembrane receptors known to frequently interact with laminins. Together, laminin-10/11 (including the LAMA5 protein as its alpha subunit) and integrin α6β1 (including the ITGA6 protein as its alpha subunit) promote the formation of pseudopodia (Gu et al., 2001), and the integrin family is known to function in epithelial cell migration through pseudopod formation (Mercurio et al., 2001). Two cadherin-family genes, CDH4 (R-cadherin or retinal cadherin) and CDH13 (T-cadherin, H-cadherin, or heart cadherin), are also differentially expressed; cadherins are implicated in cell motility, adhesion, and morphological changes, and these two cadherins particularly have been implicated in vascular remodelling (Dorrell et al., 2002; Ivanov et al., 2004). Weighted gene co-expression network analysis similarly described functionality related to cell migration and adhesion, particularly in the Antique White 4, Pale Turquoise, and Brown

modules, with statistically significant increases in activity in the first two modules associated with aggressive samples. These modules included enriched terms such as *organization of lamellipodia; ruffle organization;* and *integrin-mediated cell signaling,* and contained laminins, integrins, and genes associated with integrin signaling. Significantly alternatively spliced genes in the tame versus aggressive fox pituitary also demonstrated associations to integrins (*MFAP5* [Gibson et al., 1999]) and cell adhesion (*SORBS1* [Ribon et al., 1998]; *EFNA5* [Davy and Robbins, 2000]; *UNC5D* [Jarjour et al., 2011]). Together, these results demonstrate a recurrent theme of cell adhesion and pseudopodia formation throughout analyses of differential gene expression, GO term enrichment, weighted gene co-expression network analysis, and alternative splicing analysis, suggesting that this finding is a robust one.

Regulation of dendritic cell differentiation, another GO term highly significantly enriched for differentially expressed genes in the tame versus aggressive pituitary, also contains genes associated with pseudopodia. Dendritic cells are antigen-presenting white blood cells that pass through the final stages of differentiation in tissues after being activated by exposure to antigen (Ziegler-Heitbrock et al, 2010). Enrichment in genes related to regulation of their differentiation in the pituitary is a surprising finding; examination of the specific genes reveals two galectins related to cellular motility (*LGALS3, LGALS9* [Sano et al., 2000; Bi et al., 2011]) and an intracellular protein associated with cross-presentation of antigen (*TMEM176B* [Segovia et al., 2014]). These findings may result from the migration of dendritic cells through the bloodstream to the pituitary gland and subsequent differentiation there. However, galectins and *TMEM176B* are not solely found in dendritic cells but are represented in numerous other tissues (Wada et al., 1997; Stillman et al., 2006; Cuajungco et al., 2012). Their functionalities are not limited to dendritic cell differentiation; for example, galectin 3 has been associated with vesicle formation (Mehul and Hughes, 1997) and localizes in dendritic cell lamellipodia (a type of pseudopodium) during migration (Hsu et al., 2009); galectin 9 associates with integrins to regulate cell adhesion (Atsushikasamatsu et al., 2005, Nobumuto et al., 2008). Therefore, while the enrichment of a GO term related to dendritic cell differentiation is difficult to explain in relationship to pituitary functionality, examination of specific differentially expressed genes reveals functionality related to vesicle formation (associated with exocytosis) and pseudopodia formation and cell adhesion (associated with cellular networking of hormone-releasing cells).

Another highly enriched GO term, *granulocyte chemotaxis,* again describes a cell type (granulocytic white blood cell) that is not expected to be present in large quantities in the pituitary. This term contains two genes for receptors for interleukin 17A (IL-17), a pro-inflammatory cytokine, *IL17RA* and *IL17RC* (Gaffen, 2009). While IL-17 is most commonly expressed in T cells, it has also been shown to be widely expressed in other tissues, including epithelial tissue (Chesné et al., 2014), which is the tissue type of the anterior pituitary (Conklin, 1968). Interleukins signal through the JAK-STAT pathway (Subramaniam et al., 2002; Sansone et al., 2012), which is also represented by enriched GO terms (described below). IL-17 has different functions in different tissues, so its function in pituitary is difficult to ascertain, but it is known to enhance the production of tight junctions in epithelial tissue (Cua and Tato, 2010). Tight junctions can control cell permeability and are associated with cell signaling and cell-cell adhesion (Steed et al., 2010). This term may therefore relate more to cell migration, adhesion, and hormone release than to granulocyte chemotaxis in this tissue.

These findings are suggestive of differential function of pseudopodia, mediated by differences in integrin, laminin, and galectin expression, in tame and aggressive fox pituitaries, contributing to reduced pituitary ACTH release in response to CRH stimulation in tame foxes. To our knowledge, pseudopodia have not been described in pituitary corticotrophs. However, pituitary lactotrophs and gonadotrophs are known to form pseudopodia. These cellular protrusions may function to both directly contact blood vessels and to increase cell motility to better position endocrine cells near blood vessels for hormone release (Navratil et al., 2007). Pseudopodia may also function in the organization of anterior pituitary cellular networks for regulation of hormone release, allowing a small number of cells to coordinate the timing of hormone release to the bloodstream (Hodson et al., 2012a; Hodson et al., 2012b). These functions, while described in other pituitary endocrine cells, may be similar in ACTH-releasing corticotrophs.

3.4.3. Regulation of exocytosis

cAMP, a major mediator of exocytosis (Seino and Shibasaki, 2005), was associated with a variety of differentially expressed genes in the tame versus aggressive fox pituitary. Of the ten most significant GO terms associated with differentially expressed genes in pituitary tissue, four

are associated with regulation of cAMP. Other differentially expressed genes not included in these GO groups are also associated with cAMP regulation, including two members of the family of phosphodiesterases involved in metabolism of cAMP, *PDE4B* and *PDE7B*. Interestingly, while *PDE4B* is up regulated in the tame pituitary, *PDE7B* is up regulated in the aggressive pituitary, suggesting a difference in function between the two proteins. Two members of the adenylyl cyclase family of proteins involved with synthesis of cAMP, *ADCY7* and *ADCY8*, are both up regulated in the tame pituitary; along with phosphodiesterases, this family of proteins represents the major regulator of cAMP concentration in the cell. Adenylyl cyclase function is additionally regulated by the differentially expressed genes *CALCRL* (a member of the Brown gene co-expression module), *P2RY11* (also a member of the Brown module), and *RAPGEF4* (Smith et al., 2002; Qi et al., 2001; Anand-Srivastava, 2005; Sugawara et al., 2016). A scaffolding protein, *AKAP7* (a member of the Pale Turquoise module), is up regulated in the aggressive pituitary. AKAPs serve as anchoring proteins to spatially restrict cAMP activation (Kritzer et al., 2012), and are also known to bind PDE4s to further spatially regulate cAMP (McCahill et al., 2005; Dodge et al., 2001; Taskén et al., 2001). The cAMP regulation pathway therefore appears to be perturbed in tame pituitaries compared to aggressive pituitaries at several different levels. This supports the hypothesis that release of ACTH through cAMP-mediated exocytosis is an important factor in the different ACTH concentrations in tame and aggressive fox circulations.

The JAK-STAT pathway also appeared to be differentially regulated in tame versus aggressive pituitaries; this pathway has multiple associations with cAMP and cell adhesion. Among the most significant GO terms (Table 3.1), four GO terms are related to the JAK-STAT pathway (*JAK-STAT cascade, STAT cascade, regulation of JAK-STAT cascade, regulation of STAT cascade)*. The JAK-STAT pathway, besides its relationship to IL-17 as previously mentioned, is known to regulate cell migration in epithelial tissues (Pocha and Montel, 2014). In the pituitary, JAK-STAT signaling operates synergistically with the cAMP pathway by mediating gp130 cytokine genes such as *IL-6, LIF, IL-11,* and *CNTF* to regulate *POMC* expression and ACTH release (Chesnokova and Melmed, 2002) even in the absence of CRH (Kariagina et al., 2004). Of the 11 differentially expressed genes in the tame and aggressive fox pituitaries associated with this term, four (*ERBB4, FLRT1, FLRT3,* and *LRRC4B)* are also implicated in cell-cell adhesion and cellular migration (Gambarotta et al., 2004; Jackson et al.,

2016; Yamagishi et al., 2011; Wu et al., 2008). All four are also up regulated in the aggressive pituitary. This pathway appears to link IL-17 signaling, associated with the highly significant enriched GO term *granulocyte chemotaxis,* with cAMP signaling.

Genes associated with secretory vesicles are also implicated in differences between tame and aggressive fox pituitaries. Two members of the synaptotagmin family of membrane trafficking proteins (Südhof, 2002) are differentially expressed: *SYT10,* up regulated in the aggressive pituitary, and *SYT13*, up regulated in the tame pituitary. *SYT10* is a calcium sensor required for exocytosis of secretory vesicles (Brose et al., 1992); this gene is implicated in IGF-1 co-secretion with ACTH (Cao et al., 2011), suggesting IGF-1 functionality in regulation of the stress response at the level of the pituitary (Eppler et al., 2007). *SYT10* is also a hub gene in the Antique White 4 module. *SYT13* is an atypical synaptotagmin, lacking specific amino acid residues which are conserved in other family members, which may limit its calcium binding activity; though its function is not well understood, the SYT13 protein has been associated with constitutive vesicular transport and shown to be up regulated in contextual fear conditioning (Han et al., 2012). Another differentially expressed gene associated with secretory vesicle function is *VAMP1* (synaptobrevin), a membrane protein with binding specificity for vesicular proteins (Calakos et al., 1994); this gene is up regulated in the tame pituitary. Overall, genes associated with exocytosis both at the level of vesicle function and at the level of cAMP display differences in expression between tame and aggressive fox pituitaries, not just individually, but in enriched GO terms and gene co-expression modules.

3.4.4. Regulation of ACTH release

The most significantly differentially expressed gene, as well as the gene with the largest $\log_2$ fold change, was *GRK7*, an important regulator of CRH receptor function (Grammatopoulos, 2012), up regulated in the aggressive pituitary. Activation of the CRH receptor stimulates ACTH synthesis and release. Notably, however, the G protein-coupled receptor kinase (GRK) family is also associated with regulation of cell motility, including pseudopodium formation and cell adhesion (Chodniewicz and Klemke, 2004; Penela et al., 2014). *GRK7* represents a potential candidate gene for future studies of the functional differences between the tame and aggressive fox pituitary.

The *AVP* gene, which codes for the vasopressin protein, is up regulated in the tame fox pituitary. Vasopressin is known to potentiate ACTH release, but not POMC production, in the anterior pituitary (Levin et al., 1989), both in endocrine fashion through release from the posterior pituitary into the systemic circulation, and in paracrine fashion through local synthesis and release (Denef, 2008). However, interspecies analysis of adjacent, recently diverged genes such as *AVP* and *OXT* (Acher, 1980) introduces the potential to conflate paralogy and homology, because sequence divergence has occurred between gene copies both within each species as well as within genes but between species. As a result, some fox *OXT* reads may have inappropriately mapped to dog *AVP*, resulting in apparent up regulation due to artifacts of the alignment algorithm rather than biological shifts. Therefore, the apparent up-regulation of *AVP* should be treated cautiously until more work can more clearly define differences between dog and fox *AVP* sequences.

3.4.5. Protein targeting

Four highly significantly enriched GO terms (SRP-dependent cotranslational protein targeting to membrane, cotranslational protein targeting to membrane, protein targeting, and selenocysteine metabolic process; see Table 3.2) are related to protein targeting and signal recognition proteins (SRP). These terms all describe processes controlling the translocation of proteins during translation to the organelle where they will ultimately reside. Notably, however, all but one of the differentially expressed genes associated with these terms are associated with ribosomal proteins, i.e., with protein elongation rather than protein translocation. The remaining gene, ZFAND2B, is associated with protein folding as well as translocation (Glinka et al., 2014). Interestingly, all of these genes are up regulated in the tame fox. This finding is surprising, as the aggressive pituitary is believed to release more ACTH and therefore might be expected to synthesize a greater number of proteins.

3.5. CONCLUSIONS

Experimentally domesticated foxes provide a novel model for the study of the behavioral phenotype of domesticated animals. An attenuated ACTH response to psychogenic stressors has been shown in the tame foxes, but the cellular mechanisms resulting in the inhibition of this response are not understood. Differences in ACTH concentrations in tame and aggressive fox blood have been demonstrated (Trut et al., 2004), but both *POMC* mRNA and ACTH protein concentrations in tame and aggressive fox pituitaries are similar (Gulevich et al., 2004). These results suggest that differences between the two lines may lie in regulation of ACTH release through exocytosis rather than regulation of ACTH synthesis and processing or in corticotroph cell number. In this study, examination of gene expression through sequencing of mRNA demonstrated differences in formation of pseudopodia and in signaling by cAMP. Pseudopodia formation have been implicated in release of hormones into blood vessels and in the formation of regulatory networks for coordinating hormone release by endocrine cells, while cAMP is a major regulator of exocytosis. Overall, these findings suggest specific networks that may be relevant to domestication-related phenotypes in multiple species. Changes in the same genes demonstrated here may not be expected across different domesticated species. However, genes in networks related to cAMP regulation, cell adhesion, and pseudopod formation may have been targeted as part of the domestication syndrome and may serve as candidates in future studies of domestication or of individual variation in animals and humans.

CHAPTER 4. TRANSCRIPTOME ANALYSIS OF TAME AND AGGRESSIVE FOX ADRENAL GLANDS DESCRIBES DIFFERENTIAL REGULATION OF CHOLESTEROL METABOLISM

4.1 INTRODUCTION

Dangerous or stressful stimuli elicit a highly conserved suite of hormonal responses across mammalian species, concurrent with the behavioral fight or flight response. Both behavioral and hormonal responses show phenotypic variation between individuals (Koolhaas et al., 2010), but the extent to which behavioral variation is affected by hormonal variation versus central control is unknown. The paired endocrinological response includes the immediate release of catecholamines, mediated by the sympathetic nervous system and the adrenal medulla, and the slightly delayed release of glucocorticoids, mediated by the hypothalamic-pituitary-adrenal (HPA) axis. Both catecholamines and glucocorticoids show individual variation correlated with behavioral variation (Koolhaas et al., 2010); however, glucocorticoids are the best candidates for actually affecting the behavioral fight or flight response. As steroid hormones, they can cross the blood-brain barrier, and glucocorticoid receptors are found in several brain regions, including many associated with behavior, such as the cerebral cortex, hippocampus, septal region, and amygdala (Morimoto et al., 1996). In fact, hyper- or hypo-activity of the HPA axis has been associated with psychological disorders such as generalized anxiety disorder (Greaves-Lord et al., 2007; Mantella et al., 2008), major depressive disorder (Burke et al., 2005; Doane et al, 2013; Hardeveld et al., 2014), and post-traumatic stress disorder (Wessa et al., 2006; Meewisse et al., 2007).

The two major hormonal components of the stress response, catecholamines (including epinephrine and norepinephrine) and glucocorticoids (including cortisol and corticosterone), are both released from the adrenal glands. The adrenal medulla is distinct from the adrenal cortex, having developed from ectodermal neural crest cells, whereas the adrenal cortex develops from mesoderm. These two regions of the adrenal gland function separately in the acute stress response; release of steroid hormones from the adrenal cortex is regulated through pituitary secretion of adrenocorticotropic hormone (ACTH), while the adrenal medulla is stimulated by

the sympathetic nervous system (Ehrhart-Bornstein and Bornstein, 2008). The separate responses of the cortex and medulla are also chronologically distinct, with the medullary response delivering catcholamines to the bloodstream in less than a second, and the slower cortical response, functioning as part of the HPA axis, delivering steroidal hormones on a scale of minutes (Vincent and Mitchell, 1992; Kobelt et al., 1993). While the rapidity of the adrenal medullary response makes assessment of catecholamine levels in response to a stressor challenging, the HPA arm of the stress response is commonly assessed through measurement of glucocorticoids synthesized and released by the adrenal cortex (Dickerson and Kemeny, 2004; Burke et al., 2005).

Synthesis and release of androgens and mineralocorticoids from the adrenal cortex are also stimulated by ACTH (Nicholls et al., 1975; Longcope, 1986). Indeed, the adrenal androgens, dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEA-S), are understood to function as an integral part of the hormonal stress response (Boudarene et al., 2002; Morgan et al., 2004). The primary mineralocorticoid product of the adrenals, aldosterone, is not considered part of the classical stress response, but is coming to be appreciated for its potential role in the complex interplay between stress and health (Kubzansky and Adler, 2010).

ACTH binding to the melanocortin receptor (MC2R) throughout the adrenal cortex provides a critical stimulus for hormonal production and release (Abdel-Malek, 2001; Gorrigan et al., 2011). Hormone synthesis in the adrenal cortex is divided across three zones: cholesterol is converted to progestagens and androgens in the innermost layer, the zona reticularis; progestagens are converted to glucocorticoids in the middle layer, the zona fasciculata; and glucocorticoids are converted to mineralocorticoids in the outermost region, the zona glomerulosa (Rosol et al., 2001). The adrenal androgens released in concert with the hormonal stress response, DHEA and DHEA-S, are also stimulated by ACTH, although not solely, as they are additionally regulated by type II 3β-hydroxysteroid dehydrogenase, cytochrome b5, and steroid sulfotransferase (Rainey and Nakamura, 2008). ACTH stimulates the release of the primary adrenal mineralocorticoid, aldosterone, as well, although this hormone is primarily regulated by the rennin–angiotensin–aldosterone system (Brewster and Perazella, 2004).

While the release of catecholamines by the adrenal medulla is primarily regulated not by ACTH binding but by direct sympathetic nervous system innervation, synthesis of

catecholamines is potentiated by adrenal cortical glucocortioids as a paracrine signal. In response to glucocorticoids, chromaffin cells in the adrenal medulla increase expression of *PNMT,* which codes for phenylethanolamine N-methyltransferase (PNMT); PNMT catalyzes the synthesis of epinephrine from norepinephrine (Ehrhart-Bornstein and Bornstein, 2008). In all, ACTH stimulation of the adrenal glands causes increased synthesis of not just glucocorticoids, but catecholamines, androgens (Longcope, 1986), and mineralocorticoids (Nicholls et al., 1975).

Therefore, the symphony of hormones involved in the mammalian stress response all depend directly or indirectly on ACTH coordination of hormone synthesis and release via MC2R. Binding of ACTH to MC2R activates a signaling cascade originating with cAMP, a major intracellular signal for the release of hormones via exocytosis through both a protein kinase A (PKA)-dependent and a PKA-independent mechanism (Seino and Shibasaki, 2005). In addition to its function in exocytosis signaling, cAMP-activated PKA also stimulates three distinct pathways to support adrenal steroid biosynthesis: increased hydrolysis of stored cholesterol; delivery of cholesterol to the inner membrane of the mitochondria for glucocorticoid synthesis; and increased expression of genes involved in long-term regulation of steroidogenesis (Tsai et al., 2011).

In this complex system, an abundance of possibilities exist for regulatory control of hormonal synthesis and/or release, with possible resulting effects on behavior. As different strategies for regulation may have been selected in different species or even different populations, study of a variety of models may be best for a deep understanding of individual variation in regulation of the hormonal stress response. The experimentally domesticated fox, a less widely used model than the laboratory rodent, offers a unique opportunity to investigate HPA axis regulation. As part of an ongoing, multi-decade project, two lines of foxes were selectively bred for more than fifty generations, one for lack of fear and increased sociability to humans ("tame foxes") and another for enhanced aggression to humans ("aggressive foxes") (Trut, 2004). When compared to each other, these two populations demonstrate markedly differing hormonal responses to stressors both in the pituitary, with release of ACTH, and in the adrenals, with release of glucocorticoids, with the tame foxes showing reduced levels of these hormones both basally and during a stress response (Oskina and Plyusnina, 2000; Trut, 2004; Gulevich et al., 2004). Analysis of pituitary gene expression differences between the two lines implicated regulation of hormone release through exocytosis, both through regulation of cAMP, as well as through

formation of pseudopodia and regulation of cellular adhesion, possibly involved in cellular networking to coordinate hormone release (Chapter 3). Changes in the tame fox adrenal gland may mirror these exocytosis-focused regulatory differences seen in tame fox pituitaries, and may also include novel regulatory mechanisms.

To identify molecular differences in tame and aggressive fox adrenal glands, we have analyzed gene expression in adrenals of foxes of both lines. We specifically investigated whether tame foxes adrenals demonstrate differences in genes related to the melanocortin receptor, cAMP regulation, and glucocorticoid synthesis. We also investigated possible similarities between differential regulation in the pituitary and in the adrenals, i.e., genes related to pseudopodia formation.

4.2. METHODS

4.2.1. Animals and sample preparation

Animals were maintained at the experimental farm of the Institute of Cytology and Genetics (ICG) in Novosibirsk, Russia. All animal procedures at the ICG complied with standards for humane care and use of laboratory animals by foreign institutions. The study was approved by the Institutional Animal Care and Use Committees (IACUC) of the University of Illinois at Urbana-Champaign. The adrenal glands were removed from 12 tame and 12 aggressive adult foxes. All foxes were sexually naive 1.5 year old males that were born between March and April of 2009 and raised under standard conditions. The samples were collected in August of 2010. Foxes were euthanized using sodium thiopental, and brain regions and adrenal glands were dissected out immediately afterwards and placed into RNAlater (Life Technologies, Grand Island, NY). The samples were stored at -70°C. The right adrenal gland was sectioned to provide a sample containing both cortical and medullary tissue, from which total mRNA was extracted using the RNeasy Mini Kit (Qiagen, Valencia, California), according to the manufacturer's protocol.

4.2.2. Sequencing and quality analysis

One microgram of high quality RNA from each sample was used for sequencing. Stranded RNA-seq libraries were prepared using TruSeq SBS Sequencing kit version 3 (Illumina, San Diego, CA). Libraries were barcoded and pooled, and sequenced on two lanes on a HiSeq2500. Reads were single-end and 100 nt in length. Sequencing results were processed by CASAVA 1.8 (Illumina, San Diego, CA).

Base quality per position across reads, GC content, and distribution of sequence length were evaluated with FastQC (Andrews, 2014). Reads were processed with flexbar (Dodt et al., 2012) in two passes: the first to trim adapters, remove low quality reads, and remove reads less than 35 bp in length, and the second to remove polyA tails. Subsequently, reads that mapped to fox mitochondrial DNA sequences from NCBI (accession numbers JN711443.1, GQ374180.1, NC_008434.1, and AM181037.1) using Bowtie2 (Langmead and Salzberg, 2012) were discarded. Similarly, any remaining reads that mapped to ribosomal DNA sequences were discarded. Two samples were removed from further analysis: one tame fox sample, which was an outlier in principal components analysis (PCA) using a singular value decomposition matrix, and one aggressive fox sample, which was determined by PLINK2 (*http://www.cog-genomics.org/plink/2.0/*; Chang et al., 2015) to be from a fox that was accidentally sequenced twice. Therefore, downstream analyses were performed using RNA-seq data for 11 tame and 11 aggressive foxes.

4.2.3. Evaluation and validation of differential expression

Evaluation of differential expression was performed as in Chapter 3. Specifically, reads were aligned to *Canis familiaris* 3.1 (CanFam3) with Ensembl 1.79 annotation using TopHat 2.0.13 (Kim et al., 2013). The Ensembl annotation used is available at *ftp://ftp.ensembl.org/pub/release-79/gtf/canis_familiaris/Canis_familiaris.CanFam3.1.79.gtf.gz*. The following TopHat2 parameters were added to the command line, to allow a more robust cross-species alignment with the dog genome: *-N 3 –read-edit-dist=3*. These parameters increased the number of nucleotide mismatches and maximum edit distance allowed per read. Gene expression levels were assessed using HTSeq-count 0.6.1p1 (Anders, 2015). Differentially expressed genes between samples from tame versus aggressive foxes were evaluated using DESeq2 1.10.1 (Love et al., 2014). The

500 genes with the most variance were analyzed using the plotPCA method in DESeq2 to identify principal components associated with tame versus aggressive samples, and to identify outlier samples. Normalized gene counts (FPKMs) were assessed with CuffNorm 2.2.1 (Trapnell et al., 2013) using quartile normalization.

4.2.4. Functional analysis

Gene ontology analysis was performed for the list of most highly expressed genes, differentially expressed genes, genes that were differentially expressed in both the adrenals and the anterior pituitary, genes with the top and bottom scores in principal components differentiating tame from aggressive samples, and genes in individual modules (see section 5) using ClueGO version 2.3.2 (Bindea et al., 2009) in Cytoscape version 3.4.0 (Shannon et al., 2003), with the February 23, 2017 version of the Gene Ontology (GO). All ClueGO evaluations were performed against a background list of 18,855 genes with a mean expression level in pooled tame and aggressive adrenals samples reported to have mean expression > 0 by DESeq2. GO term fusion was used when necessary to reduce large numbers of GO terms to more manageable results. Resulting p-values were adjusted using the Benjamini-Hochberg correction. Enriched GO terms were reviewed for association with stimulus of hormone release, glucocorticoid regulation and processing, and pseudopodia formation.

4.2.5. Weighted gene co-expression network analysis

Modules of genes with highly correlated expression patterns were sought using weighted gene co-expression network analysis. We expect these modules to correspond to networks of genes that interact in shared biological processes (Langfelder, 2008). The full set of genes expressed in the adrenal reads was filtered to remove genes with fewer than 10 FPKMs in each of 3 or more samples. We then constructed unsigned weighted gene co-expression modules using the WGCNA package in R (Langfelder, 2008). The blockwiseModules function was run with the Pearson correlation coefficient and a soft thresholding power of 12. The resulting gene groups, or modules, were named by assigning them arbitrary colors. The first principle component of each module was designated its eigengene. Eigengenes for each module were divided into two groups,

from tame and aggressive foxes; these two groups were compared with a Student's *t* test and their p values were adjusted using the Benjamini-Hochberg correction. Modules selected for investigation were analyzed for gene enrichment using ClueGO as described above. Additionally, highly connected "hub" genes in these modules were identified using Cytoscape (Shannon et al., 2003) to select the genes with the largest number of connections to other genes in the same network, with the expectation that these "hub" genes would represent regulatory genes with control over many genes in that module.

4.2.6. Alternative splicing analysis

rMATS was employed to identify differences in alternative splicing frequencies between tame and aggressive samples (Shen et al., 2014). Because rMATS requires reads of equal lengths, post-filtered reads were further filtered using samtools 1.1 (Li et al., 2009) to remove all reads shorter than 100 nucleotides in length. Previously aligned BAM files and Ensembl annotation, version 1.79, were used with rMATS to investigate reads crossing splicing junctions. Differences in frequencies of reads including or skipping specific exons were used to identify exons that were skipped significantly more frequently in tame or aggressive adrenals reads. Resulting p values were adjusted using Benjamini-Hotchberg. The resulting list of differentially skipped exons was filtered to remove exons covered by fewer than 10 reads in either tame or aggressive samples. Differences in skipping were reported as mean percentage skipped exons, in other words, (mean number of reads skipping the exon)/(mean number of reads skipping the exon + mean number of reads including the exon). Exons identified as significantly more frequently skipped in tame or aggressive samples were further investigated by searching for their amino acid sequence using their nucleotide sequence in blastx (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*) to identify isoforms that included or did not include the exon in question and to identify putative conserved domains overlapping the alternatively spliced sequence.

4.3. RESULTS

4.3.1. Mapping and evaluation of differential expression

Sequencing of fox adrenals produced 366 million reads total for 11 tame samples (348 million reads after filtering) and 347 million reads total for 11 aggressive samples (322 million reads after filtering) (Table SI-3.1). The filtered fox reads were mapped to the CanFam3 reference with a mean alignment rate of 79.23% (Table SI-3.2). The rate of multiple alignments was low, with a mean of 3.58%. Analysis of the filtered adrenals reads with DESeq2 described the expression of 18,855 distinct genes in this tissue.

The most significantly enriched GO terms in the 0.01% most highly expressed genes (188 genes, of which 182 had an associated gene symbol to be used in the GO analysis) were related to translocation of proteins to the endoplasmic reticulum, mRNA metabolic processes, and translational initiation (Table SI-3.3). The 0.001% of the most highly expressed genes in the fox adrenal gland (Figure 4.1) include four genes in the cytochrome P450 superfamily of enzymes; this superfamily is responsible for the majority of steroidogenesis and therefore for biosynthesis of glucocorticoids in the adrenal gland. All four genes (*CYP17A1*, *CYP21A2*, *CYP11A1,* and *CYP11B1*) function in the corticosteroid (glucocorticoid and mineralocorticoid) biosynthetic pathway, as does *HSD3B2* (II 3β-hydroxysteroid dehydrogenase), another highly expressed gene in the adrenals (Payne and Hales, 2004). Also included in this list is *STAR,* which codes for an enzyme that delivers cholesterol to the mitochondrial membrane as part of the rate-limiting step in steroid biosynthesis (Miller and Bose, 2011). Additional highly expressed genes included two genes, *APOE* and *SCARB1,* involved in regulation of lipoproteins (Mahley, 1988; Zanoni et al., 2016), which serve as substrates for the steroid biosynthesis pathway; two genes, *SRRM2* and *DDX5,* involved in alternative splicing of mRNA (Fuller-Pace et al., 2008; Fontrodona et al., 2013); two genes, *CHGA* and *CHGB,* encoding pre-proteins for neurosecretory proteins, with catcholamine regulatory function (Hendy et al., 2006; Zhang et al., 2009); and two translation elongation genes, *EEF1A1* and *EEF2* (Anand et al., 2002; Kanibolotsky et al., 2008).

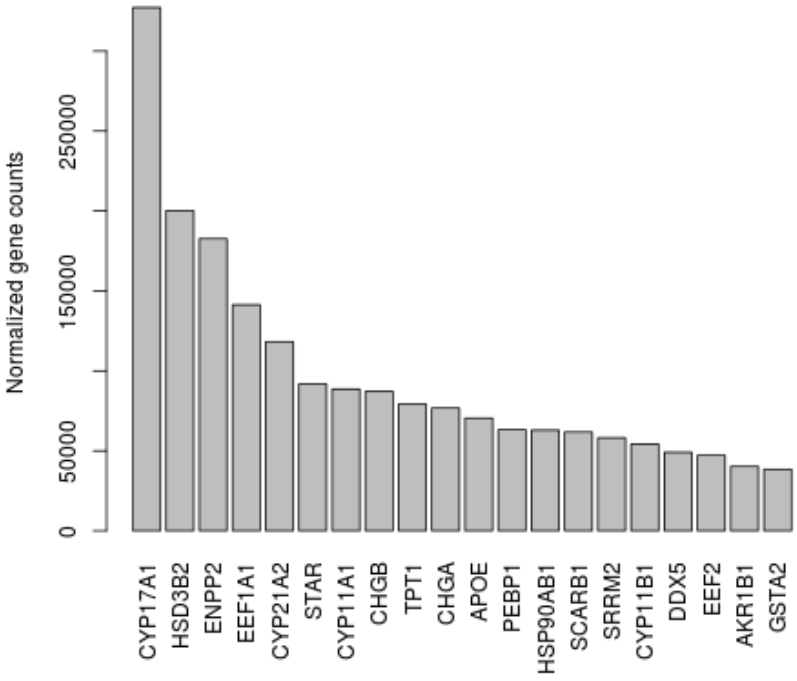

Figure 4.1. Most highly (0.001%) expressed genes in fox adrenals by normalized gene count, assessed by DESeq2.

DESeq2 analysis of filtered adrenal reads identified 775 differentially expressed genes (DE genes) (FDR < 0.05) (Table SI-4.4; Figure SI-4.1). Of these, 387 genes (49.9%) were up-regulated in aggressive adrenals, and 388 (50.1%) were up-regulated in tame adrenals. Log_2 fold changes among DE genes ranged from -0.90 – 0.98 (mean -0.01, SD 0.35). Clustering analysis of DE genes clearly differentiated tame and aggressive samples (Figure 4.2). Principal components analysis (PCA) differentiated tame from aggressive samples along the PC2 axis, but not the PC1 axis (Figure SI-4.2; Table SI-4.5). The 50 highest and 50 lowest scoring genes on PC2 were analyzed for GO term enrichment; 42 GO terms were significantly enriched, including 9 terms relating to lipid or cholesterol biosynthesis (Table SI-6). These 42 enriched terms contained 16 genes that were in the 50 highest scoring genes on PC2, e.g., up-regulated in tame fox adrenals, and 11 genes that were in the 50 lowest scoring genes on PC2, e.g., up-regulated in aggressive fox adrenals, as well as 9 genes that were also differentially expressed between tame and

aggressive adrenals: *ADRA2A, ALDH1B1, FBLN1, GPAT2, NGEF, SCN3A, SEMA5A, SERPINF1,* and *SULT1C4.*

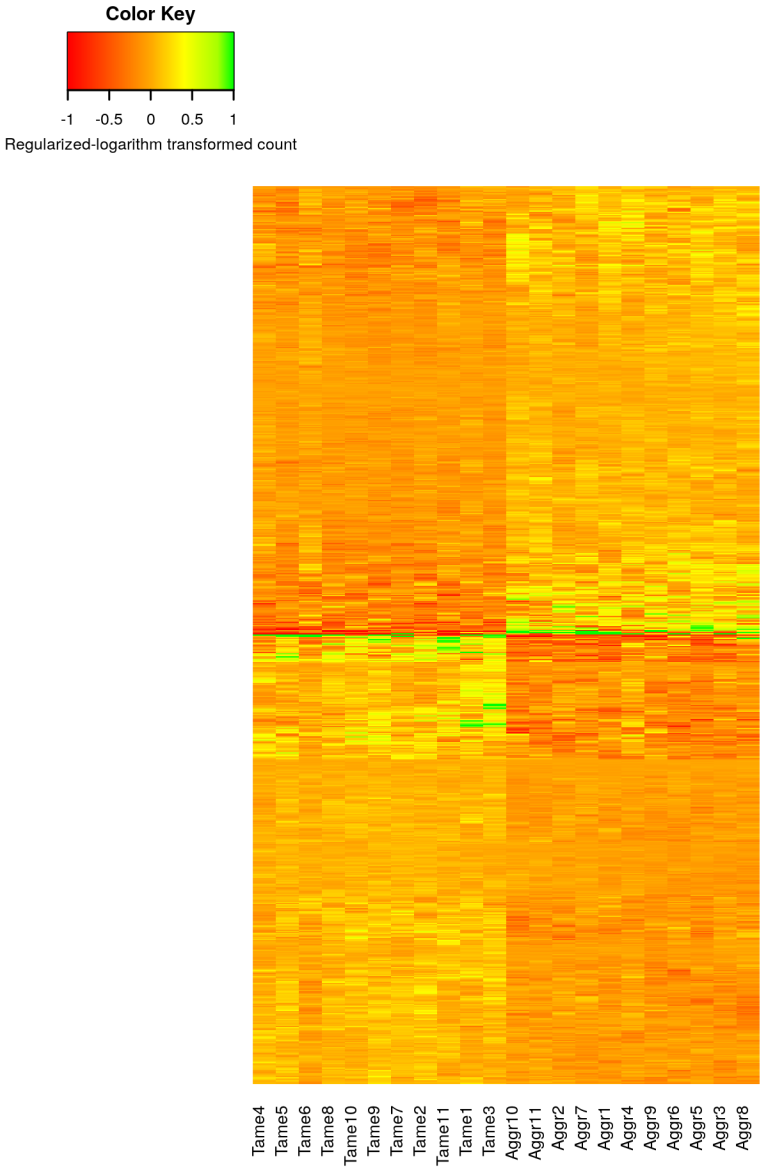

Figure 4.2. Heat map of differentially expressed genes (775) in tame versus aggressive fox adrenal tissue (FDR < 0.05).

The differentially expressed gene gene with the largest positive $\log_2$ fold change (i.e., up-regulated in the tame adrenal), *LOC100686868*, does not have an associated gene symbol; the gene with the second largest positive $\log_2$ fold change was *HS3ST5,* which codes for Heparan Sulfate-Glucosamine 3-Sulfotransferase 5, which catalyzes the transfer of a sulfo group to heparan as part of the biosynthesis of the anticoagulant heparan sulfate and is involved in the insulin secretion pathway (Takahashi et al., 2012). The gene with the largest negative $\log_2$ fold change (i.e., up-regulated in the aggressive adrenal) was *DSC3,* which codes for the desmocollin protein, a component of the intercellular desmosome junction (Nuber et al., 1996). In addition to these three genes, two other outlier genes ($\log_2$ fold change > 0.5, FDR < 0.0000001) were apparent when visualized on a volcano plot (Figure SI-4.1), namely, *THNSL2/SOFAT* (up-regulated in aggressive; a threonine synthase with cytokine activity [Rifas and Weitzmann, 2009]) and *NQO1* (up-regulated in tame; a redox sensor that regulates energy metabolism [Adamovich et al., 2013]).

A comparison of differentially expressed genes in the adrenals with differentially expressed genes in the pituitaries of tame and aggressive foxes (Chapter 3) revealed 59 genes that were differentially expressed in both tissues (Table SI-4.7). Of these, 57 were differentially expressed in the same direction in both tissues; the two exceptions were *NLGN4X* and *NMRK1,* which were both up-regulated in the aggressive adrenal but in the tame anterior pituitary. A ClueGO analysis of the 59 genes revealed two GO terms significantly enriched for genes in the list: *aspartate family amino acid metabolic process* (FDR = 0.0017), including the doubly differentially expressed genes *BHMT2, NIT2,* and *THNSL2;* and *positive regulation of response to DNA damage stimulus* (FDR = 0.0025), including the doubly differentially expressed genes *PAXIP1, PLA2RL,* and *SPIDR*.

4.3.2. Functional analysis

ClueGO analysis was performed on the 627 differentially expressed genes that had an associated gene symbol out of the total of 775. The differentially expressed genes were associated with 1,497 GO terms (Table SI-4.8); percent differentially expressed genes in enriched terms ranged from 4-67% (median 7%). Two GO terms were significantly enriched for

differentially expressed genes (FDR < 0.05): *negative regulation of interleukin-8 production* (containing four genes up-regulated in tame adrenals and two genes up-regulated in aggressive arenals) and *ectodermal cell differentiation* (containing three genes up-regulated in tame adrenals and one gene up-regulated in aggressive adrenals). Terms with more than 30% of their genes differentially expressed were *ectodermal cell differentiation, positive regulation of vesicle fusion, negative regulation of interleukin-8 production, activation of Janus kinase activity, endothelial tube morphogenesis,* and *pseudopodium assembly* (Table 4.1).

Table 4.1. Terms with more than 30% of their genes differentially expressed between tame and aggressive fox adrenals. Gene symbols in **bold** are differentially expressed in both adrenals and anterior pituitary.

GO ID	Term description	Adjusted p-value	Differentially expressed genes annotated with this term	% genes in this term differentially expressed
GO:0010668	ectodermal cell differentiation	0.04	*ETS2, FZD7,* ***ITGA6****, ITGAM*	66.67
GO:0032717	negative regulation of interleukin-8 production	0.03	*ANXA1,* ***BCL3****, IL6R, KLF4, PTPN22, SSC5D*	42.86
GO:0031340	positive regulation of vesicle fusion	0.21	*ANXA1, C2CD5, KIF5B*	42.86
GO:0042976	activation of Janus kinase activity	0.17	*CCL5, IL6R, PRLR, SOCS1*	36.36
GO:0061154	endothelial tube morphogenesis	0.12	*ACVRL1, BMP4, CXCL10, PRKD2, STARD13*	33.33
GO:0031269	pseudopodium assembly	0.19	*CDC42EP1, CDC42EP3,* ***SRGAP2****, WASHC*	30.77

Among all 1,497 GO terms, 16 terms related to cellular adhesion, containing 54 differentially expressed genes; 45 terms related to cellular movement (motility, migration, and chemotaxis), containing 105 differentially expressed genes; 14 terms related to cellular projections (pseudopodia and cilia), containing 81 differentially expressed genes; 38 terms related to exocytosis (including cAMP, vesicles, and protein or hormone secretion), containing 112 genes; and 20 terms related to hormones (including cholesterol, glucocorticoids, and steroids), containing 69 genes. Among terms in the top 0.1% ranked by p-vaues were four terms related to cellular motility (*regulation of cellular component movement, endothelial cell migration, mononuclear cell migration,* and *regulation of endothelial cell migration*); one term related to hormone regulation (*ovulation cycle*); one term related to cellular excitability (*regulation of cardiac muscle cell action potential*); one term related to coagulation (*negative regulation of coagulation*); four terms related to immune signaling (*negative regulation of interleukin-8 production, B cell receptor signaling pathway, negative regulation of cytokine production,* and *positive regulation of immune system process*); two terms related to development (*ectodermal dell differentiation* and *endothelial tube morphogenesis*); and one term related to cell signaling (*activation of Janus kinase activity*).

Genes related to critical steps in steroid biosynthesis were examined for differential expression in tame and aggressive fox adrenals. Differentially expressed genes found included two related to StAR, the rate-limiting enzyme in steroid synthesis (Miller and Bose, 2011), and one of the 0.01% most highly expressed genes in this data set (Figure 4.1): *STARD13* (FDR=0.015) and *STARD8* were up-regulated in the tame adrenal (FDR=0.031). Members of the cytochrome P450 family are involved in the glucocorticoid biosynthesis cascade (Payne and Hales, 2004); the only member of this family differentially expressed in the fox adrenal was *CYP4F8* (up-regulated in aggressive; FDR=0.00057), which codes for a protein involved specifically in eicosanoid rather than glucocorticoid metabolism (Bylund et al., 2000). Also critically involved in the glucocorticoid biosynthesis cascade is *HSD3B1/HSD3B2* (II 3β-hydroxysteroid dehydrogenase) (Payne and Hales, 2004), which was not differentially expressed in tame versus aggressive fox adrenals.

4.3.3. Weighted gene co-expression network analysis results

Weighted gene co-expression network analysis of 8,429 genes with FPKM count ≥ 10 in at least eight samples revealed 19 modules of genes with related activity levels (Table SI-4.9). Three gene modules had significantly different eigengene values (FDR < 0.1) between tame and aggressive samples (Table 4.2; Table SI-4.10). One module, Magenta, had a negative mean eigengene in the tame samples, while two modules, Green and Tan, had positive mean eigengenes in the tame samples. A positive mean eigengene suggests that that genes in that module demonstrate increased activity as a unit in the tame population, while a negative mean eigengene suggests increased activity in the aggressive population.

Table 4.2. Modules of co-expressed gene networks with significantly different activity in tame (versus aggressive) fox adrenal tissue.

Module name	Mean eigengene (tame)	FDR	Number of genes in module	Percent genes in module differentially expressed between tame and aggressive foxes
Magenta	-0.14	0.0056	363	21
Green	0.14	0.0056	686	19
Tan	0.11	0.0758	180	12

The Magenta module, with 77 differentially expressed genes, had 24 significantly enriched GO terms (Table SI-4.11), including *cholesterol biosynthetic process.* This term contained the Magenta module associated genes *APOE* (one of the 0.01% of most highly expressed genes in fox adrenals [Figure 4.1])*, CYP51A1, DHCR24, HMGCR,* and *LBR.* Notably, *HMGCR* codes for HMG-CoA reductase, which is the rate-limiting enzyme for cholesterol synthesis (Panda and Devi, 2004). Other significantly enriched terms associated with this module also included three terms related to membrane organziation (*single-organism membrane organization, membrane organization,* and *single-organism membrane fusion*); five terms related to organelle organization and localization (*single-organism cellular localization; peroxisome organization;*

nucleus organization; nuclear envelope organization; organelle localization); two terms related to mitochondria (*cristae formation; mitochondrial membrane organization)*; three terms related to intracellular transport (*single-organism intracellular transport; transport along microtubule; intracelular protein transmembrane transport)*; and three terms related to protein transport (*protein import; protein complex localization; protein targeting to peroxisome).* The most highly-connected gene in the module, *Snap25,* is associated with vesicle docking and membrane fusion, and has been associated with ADHD and trait neuroticism (Brophy et al., 2002; Mill et al., 2002; Terracciano et al., 2010). Other highly connected genes in this module include *STARD3NL,* which may be involved in cholesterol transport (Wilhelm et al., 2016), and *MIA2,* which is involved in transport of lipoproteins (Santos et al., 2016). Therefore, this module appears to have some relationship to cholesterol processing, in addition to cellular and organelle membrane organization, energy metabolism, and vesicle docking.

The Green module, with 132 differentially expressed genes, had one GO term significantly enriched for genes in that module (Table SI-4.11), *fibroblast migration.* This term contained the Green module associated genes *CORO1C, HYAL2, MTA2, PIP5K1A, PLXNA1, PML, SDC4,* and *THBS1.* This module's most highly-connected gene was *EGLN2,* which suppresses hypoxia-inducible factor 1α (Erez et al., 2003). The module contained 132 differentially expressed genes, including *HYAL2,* a hyaluronidase involved in cellular migration (Ellis et al., 1997). *HYAL2* was also the second most highly connected gene in the module, and a member of the significantly enriched *fibroblast migration* GO term. The other gene in this GO term associated with the Green module was *THBS1,* which is involved in cell adhesion, movement, and angiogenesis (Adams, 1997), and which is up-regulated in the tame adrenal. Additional differentially expressed genes in this module included genes involved in regulation of the cytoskeleton, cellular projections, and pseudopodia (*ANXA1, CDC42EP1, CDC42EP3, FES, DYNC2LI1, ITGA6,* and *ZMYND8/*Spikar [Shaw and Mercurio, 1994; Hirsch et al., 2001; Ernst et al., 2004; Laurent and Smithgall, 2004; Yamazaki et al., 2014; Taylor et. al, 2015]); adhesion and angiogenesis (*ACVRL1* [Lamouille et al., 2002]), migration (*ITGA6* [Golbert et al., 2013]); glycosylation (*GALNT12, GALNT3* [Guo et al., 2002; Kato et al., 2006]); and potassium channels (*KCNIP3, KCNN1*). This module also includes *HS3ST5,* the heparan sulfotransferase that is the gene with the second largest positive (up-regulated in tame) $\log_2$ fold change between

tame and aggressive adrenals, and another differentially expressed sulfotransferase, *SULT14C*. Overall, this module appears to have functions relating to cellular migration and adhesion.

The Tan module, with 12 differentially expressed genes, had 61 GO terms significantly enriched with genes from that module (Table SI-4.11). Thirteen of these terms with the smallest p-values clustered together in five groups with overlapping genes (Table 4.3). These groups related to extracellular matrix organization, collagen metabolism, endodermal development, angiogenesis, and transforming growth factor beta signaling.

Table 4.3. Selected GO terms from those that are significantly enriched for genes in the Tan module. Genes in bold are included in every GO term in the group.

Group	GO IDs	Term descriptions	Adjusted p-value	Genes annotated with this group and associated with the Tan module	% genes in this term differentially expressed
1	GO:0043062	extracellular structure organization	$200e^{-11}$	*COL1A2, COL3A1, COL4A1, COL4A2, COL5A2,* ***CTSK****, CYP1B1,* ***DCN****,* ***ENG****, FBLN1, FBLN5, FMOD,* ***GSN****,* ***LAMA4****,* ***LAMC1****, LUM,* ***MMP14****,* ***MMP2****, MYH11, TEX14*	6.90
	GO:0022617	extracellular matrix disassembly	$260e^{-7}$		9.20

Group	GO IDs	Term descriptions	Adjusted p-value	Genes annotated with this group and associated with the Tan module	% genes in this term differentially expressed
2	GO:0001568	Blood vessel development	$240e^{-11}$	*ACTA2*, ***AQP1***, ***CD34***, *COL1A2*, *COL3A1*, *COL4A1*, *COL4A2*, ***CYP1B1***, *DCN*, ***ECM1***, *EMP2*, ***ENG***, *FBXW8*, *FGFR1*, *LAMA4*, *MCAM*, *MMP14*, *MMP2*, *MYLK*, *NOV*, *NPR1*, *PDGFRB*, ***PRKD1***, ***PTGIS***, ***SOS1***, *TCF21*	4.86
	GO:0048514	Blood vessel morphogenesis	800^{e-10}		4.76
	GO:0001525	Angiogenesis	$820e^{-10}$		5.17
	GO:0045765	Regulation of angiogenesis	$930e^{-8}$		6.12
	GO:0045766	Positive regulation of angiogenesis	$130e^{-6}$		7.27
3	GO:0032963	Collagen metabolic process	$960e^{-9}$	***COL1A2***, ***COL3A1***, *COL4A1*, *COL4A2*, ***COL5A2***, *CTSK*, *CYP1B1*, *ENG*, *FMOD*, *LUM*, *MMP14*, *MMP2*, *PDGFRB*	10.31
	GO:0030574	Collagen catabolic process	$220e^{-8}$		13.33
	GO:0030199	Collagen fibril organization	$970e^{-8}$		18.18

Table 4.3 (cont.)

Group	GO IDs	Term descriptions	Adjusted p-value	Genes annotated with this group and associated with the Tan module	% genes in this term differentially expressed
4	GO:0035987	Endodermal cell differentiation	$290e^{-6}$	***COL4A2, COL5A2, LAMA4,*** *LAMC1*, ***MMP14, MMP2***	12.82
	GO:0007492	Endoderm development	$350e^{-6}$		8.70
5	GO:0071560	Cellular response to transforming growth factor beta stimulus	$290e^{-6}$	***ASPN, CLEC3B, COL1A2, COL3A1, COL4A2, ENG, FERMT2, FMOD, LTBP4, TGFB1I1***	4.98

Table 4.3 (cont.)

The five most highly-connected genes in the Tan module were all collagens (*COL5A2, COL6A2, COL1A2, COL3A1,* and *COL6A1*), and three of these genes were also represented in multiple of the most significantly enriched GO terms in this module (Table 4.3). Other highly-connected genes included *DCN*, involved in collagen fibril assembly (Reed and Iozzo, 2002), significantly up-regulated in the tame adrenal, and also represented in multiple of the most significantly enriched GO terms; *GSN,* an actin-modulating gene involved in ciliogenesis (Yin, 1987; Kim et al., 2010), also represented in multiple of the most significantly enriched GO

terms; *FBLN1,* involved in cell adhesion and motility (Twal et al., 2000), significantly up-regulated in the tame adrenal, and included in *extracellular structure organization,* the most significantly enriched GO term; and *ACTA2,* a member of the actin family, involved in cellular motility (Pollard and Borisy, 2003), significantly up-regulated in the tame adrenal, and represented in the second most significantly enriched GO term, *blood vessel development.* Actin-interacting differentially expressed genes in this module included *FERMT2, MYL6B,* and *TRIOBP* (Adelstein and Eisenberg, 1980; Tu et al., 2003; Bao et al., 2013). Overall, this module's functionality may relate to collagen processing and angiogenesis.

The Black module, while not demonstrating significantly different activity between tame and aggressive samples (FDR=0.139), described functionality directly relevant to HPA axis activity. This module's mean eigengene was -0.09, suggesting increased activity in the aggressive population. Two GO terms in this module were significantly enriched for genes in the Black module: *putrescine metabolic process* and *steroid biosynthetic process* (Table SI-4.11). *Putrescine metabolic process* (FDR = 0.044) contained three genes in the Black module: two involved in polyamine (e.g., putrescine) metabolism (*AZIN1* and *SAT1* [Tang et al., 2009; Li and MacDonald, 2016]) and an amino acid transporter (*SLC38A1/ATA1* [Wang et al., 2000]). *Steroid biosynthetic process* (FDR = 0.049) contained 11 genes in the Black module, including genes involved in cholesterol biosynthesis (*FDFT1, INSIG1*, *CYP39A1*, *OSBPL1A*, *OSBPL6,* and *SREBF1* [Pandit et al., 2000; Yang et al., 2002; Laaksonen et al., 2006; Pikuleva, 2006; Motazacker et al., 2016; Oimet et al., 2016]), genes regulating steroidogenesis in response to long-term ACTH stimulation (*NR0B1/DAX1, PBX1,* and *SF1* [Lichtenauer et al., 2007; Schimmer and White, 2010]), a gene involved in cytoskeletal organization (*SH3D19* [Bai et al., 2011]), and a glucocorticoid-responsive histone deacetylase (*SIRT1* [Amat et al., 2007]). The most highly-connected gene in this module was *YPEL3,* which regulates epithelial cell migration via the Wnt/β-catenin signaling pathway (Zhang et al., 2016). This module contained 42 genes differentially expressed between tame and aggressive adrenal. Overall, this module's functionality appeared to include putrescine metabolism and synthesis of steroid hormones.

4.3.4. Alternative splicing analysis

rMATS was used to identify 31 exons skipped at different frequencies in tame versus aggressive fox adrenals at FDR < 0.05 (Table SI-4.12); these exons were distributed across 28 genes. Exon counts for these genes ranged from 5-46 exons total (median 14.5, SD 9.12). At FDR < 0.001 (Table 4.4), 12 exons were differentially skipped between the populations, distributed across 10 genes.

Genes with differentially spliced exons were compared to differentially expressed genes and genes in WGCNA modules. A single gene with a differentially spliced exons was also differentially expressed, *SYCP2,* up-regulated in the tame adrenal; this gene is associated with chromosome synapsis during meiosis (Yang et al., 2006). The Magenta module contained the differentially spliced genes *IMMT/MINOS2* (associated with the inner mitochondrial membrane [Zerbes et al., 2012]) and *TFF2* (associated with cell migration [Zhang et al., 2010]), and the Tan module contained the differentially spliced gene *FMNL3* (associated with reorganization of the actin cytoskeleton and lamellipodia [Block et al., 2012]).

Table 4.4. Genes containing exons skipped at different frequencies in tame compared to aggressive fox adrenals reads (FDR < 0.001).

Gene	Exon location	Percent reads skipping exon (tame)	Percent reads skipping exon (aggr)	FDR	Skipped in isoforms	In conserved domains
MFAP5	chr27:37034668-37034704	0.39	0	0	2, 3	None
GSTM4	chr6:42206292-42206393	0.06	0	0	2	GST_C_family superfamily
GSTM4	chr6:42206493-42206575	0.08	0	0	2	Thioredoxin-like superfamily
C10orf107	chr4:13796967-13797090	0.09	0.02	3.58E-007	Uncharacterized protein	None

Gene	Exon location	Percent reads skipping exon (tame)	Percent reads skipping exon (aggr)	FDR	Skipped in isoforms	In conserved domains
SPICE1	chr33:17787918-17788056	0.13	0.02	2.29-006	In all reported variants	None
CAPN3	chr30:9381076-9381141	0.02	0.15	7.35-006	In all reported variants	Efh_PEF superfamily
TADA2A	chr9:37231086-37231197	0.11	0.03	1.23E-005	4	None
NAV1	chr7:1168221-1168392	0.04	0.13	1.41-005	4, 5	None
TADA2A	chr9:37232686-37232758	0.12	0.04	0.00011	4	None
HDHD2	chr7:44340420-44340629	0.07	0.02	0.00015	2	NagD superfamily
SYCP2	chr24:44484384-44484555	0.12	0.05	0.00019	6, 8, 9	None
NT5M	chr5:42113879-42113980	0.04	0.15	0.00042	In all reported variants	HAD-like superfamily

Table 4.4 (cont.)

Included among the significantly differently spliced genes (FDR < 0.05) were genes involved in motility and cell morphology (*FMNL3, OFD1, RFX2, SRGAP3, TFF2* [Ferrante et al., 2006; Zhang et al., 2010; Block et al., 2012; Bisgrove et al., 2012; Koschützke et al., 2015]); transcription and translation (*NFIB, NT5B* [Liu et al., 1997; Kviklyte et al., 2017]); gametogenesis (*PIWIL3* [Juliano et al., 2011]); acetylation *(TADA2* [Ogryzko et al., 1998]; and the cell cycle (*MAD2L1, SPICE1, SYCP2*) [Luo et al., 2000; Yang et al., 2006; Archinti et al., 2010].

4.4. DISCUSSION

Because of the known differences in cortisol both basally and in response to stressors between tame and aggressive foxes, we hypothesized that we would see differences in glucocorticoid-specific biosynthesis pathways in tame versus fox adrenal gene expression. Instead, differences were found in pathways relating to the metabolism of cholesterol and fatty acid metabolism, precursors not just to glucocorticoids but to steroid hormones more generally. Adrenal activity between fox lines also appears to differ in release of hormones (exocytosis); cell signaling, possibly to coordinate hormone release; and angiogenesis, potentially related to perfusion levels connecting hormone-producing cells to the systemic circulation. In addition, some medullary-specific changes implicate differences in ion channels and cellular excitability, possibly associated with function in the sympathetic arm of the stress response.

Cholesterol is the precursor for all steroid hormones; fatty acids are metabolized into acetyl Co-A and then into cholesterol in the mitochondria (Griffin and Ojeda, 2004). The GO terms *cholesterol biosynthesis* and *fatty acid metabolism* were both enriched in genes from the Magenta module, which had significantly increased activity in samples from the aggressive line. These genes include *ACADM* (an acyl-CoA dehydrogenase), the enzyme which catalyzes the initial step of fatty acid beta-oxidation (Bross et al., 1990); two genes, *AKT1* and *MORC2*, which regulate ATP-citrate lyase, a key enzyme in cholesterol synthesis (Berwick et al., 2002; Sánchez-Solana et al., 2014); *HMGCR* (HMG-CoA reductase), the rate-limiting enzyme in the mevalonate pathway for cholesterol production (Panda and Devi, 2004); *CYP51A1,* a P450 superfamily enzyme which operates in the cholesterol synthesis pathway (Pikuleva, 2006); and *APOE* (apolipoprotein E), a cholesterol transporter and regulator (Mahley, 1988). Additionally, *NQO1* (NAD(P)H Quinone Dehydrogenase 1), a gene which is up-regulated in tame adrenals, an outlier on the volcano plot of differentially expressed genes, and a member of the Green module, produces a co-factor in the synthesis of lipids, including fatty acids (Liu et al., 2010). These findings, particularly in the critical steps of cholesterol biogenesis mediated by acyl-CoA dehydrogenase and HMC-CoA reductase, strongly associate differences in cholesterol synthesis with the reduction of activity in the tame fox adrenal cortex. Intriguingly, changes in expression of genes related to lipid metabolism, including *HMGCR*, were also found in porcine adrenal glands after a social stressor (Muráni et al., 2011). Just as in our findings, this porcine data did

not demonstrate glucocorticoid-specific changes. It is possible that availability of precursors has greater effect on production of glucocorticoids than does availability of biosynthetic enzymes. Alternatively, it is possible that the cholesterol and fatty acid changes seen between tame and aggressive fox adrenals are related to differences in energy metabolism. The different behavioral traits of the two lines may result in different energy requirements, reflected in their use of high-energy compounds such as fatty acids and cholesterol (Careau et al., 2008).

Rather than cholesterol metabolism differences, we expected to see differences in the expression levels of genes such as *STAR,* the rate-limiting enzyme in cholesterol biosynthesis (Miller and Bose, 2011), and *MC2R,* the ACTH receptor. These genes were specifically examined in previous studies that investigated differences in adrenal activity associated with a stressor or breed characteristics (Muráni et al., 2011; Wang et al., 2014). The adrenal expression of these genes has been demonstrated to change in response to acute stress in some models (Fallahsharoudi et al., 2015) but not others (Muráni et al., 2011). In our study, neither *MC2R* nor *STAR* were significantly differentially expressed, though *MC2R* was up-regulated in the aggressive fox adrenal before Benjamini-Hochberg correction. The Black WGCNA module contained two key regulators of *STAR*, *NR5A1/SF1* and *NR0B1/DAX1* (Stocco, 2001), along with other genes associated with steroid biosynthesis; this module also did not have significantly different activity between the lines, although it did have higher activity in the aggressive fox adrenals before correction. Therefore, it appears that differences in *MC2R* and *STAR* expression probably do not play an important role in the differences between adrenal activity in the tame and aggressive foxes.

The greater demands placed on the adrenal glands of aggressive foxes, in terms of glucocorticoid and catecholamine release, might be expected to result in increased cell turnover in the aggressive adrenal. Stem cells and cellular proliferation have been observed in the adrenal cortex (Mitani et al., 2003; Chang et al., 2013), and therefore increased differentiation of stem cells and/or increased division of terminally differentiated cells represent potential regulatory mechanisms for adrenal activity. One of the two GO terms significantly enriched for differentially expressed genes in tame versus aggressive fox adrenals was *ectodermal cell differentiation.* The adrenal medulla develops from ectoderm via migrating neural crest cells, and differences in neural crest cell migration have been postulated as a key difference between domesticated and wild animals, including the experimentally domesticated foxes (Wilkins et al.,

2014). The GO term *ectodermal cell differentiation* contains four differentially expressed genes. Three are up-regulated in the tame adrenal: *ETS2,* a transcription factor which regulates neural crest cell differentiation (Paratore et al., 2002); *FZD7,* a Frizzled family polarity signaling protein (Seifert and Miodzik, 2007), of particular interest as genes in the Frizzled family were also differentially expressed in the adrenals of domesticated chickens compared to their wild ancestor (Fallasharoudi et al., 2015), as well as in the adrenal medulla of rats subjects to restraint stress (Liu et al., 2008); and *ITGAM,* an integrin, involved in cellular adhesion (Chidlow et al., 2010). The fourth gene in this GO term, *ITGA6,* up-regulated in both aggressive adrenals and aggressive pituitaries, codes for another integrin, involved in stem cell differentiation (Brafman et al., 2013). As expected, these genes do appear to relate to neural crest cells, as well as to cellular adhesion and migration. Another finding with possible relevance to stem cell activity is the pair of GO terms, *endodermal cell differentiation* and *endoderm development,* enriched for genes in the Tan module. This finding is surprising, as parts of the adrenal glands are derived from ectoderm and mesoderm, but not endoderm, and the possibility that this finding is an artifact of the inherent inexactness of GO analyses should be considered. As for division of terminally-differentiated cells, genes related to transcription, translation, and the cell cycle contained significantly differentially spliced exons between tame and aggressive foxes, suggesting that regulatory differences between the lines may affect cell division.

Cellular adhesion also appears to play an important role in differences between tame and aggressive fox adrenals, as six GO terms related to cellular adhesion were enriched (though not significantly) with differentially expressed genes in this tissue. One of these differentially expressed genes, *FBLN1* (fibulin 1, a secreted glycoprotein associated with cell adhesion and migration [Twal et al., 2001]), seems to be of particular importance in the difference between tame and aggressive adrenal function, as it is one of the highest loading genes in PC2, the principal component that separates tame from aggressive adrenals. *FBLN1* is also differentially expressed in tame versus aggressive fox pitutiary, suggesting a possible genomic explanation for its expression increase, and/or similar approaches to modifying cell adhesion across tissues. Specific adhesion mechanisms demonstrating differences between the lines included desmosomes, which function as specialized cell-cell adhesion structures, and heparan sulfate, a proteoglycan modification implicated in cell-cell adhesion, cell-extra cellular matrix adhesion, and focal adhesions (Bernfield et al., 1999). Specifically, the desmosome component *DSC3*

(Nuber et al., 1996) was the most up-regulated gene in the aggressive adrenal, while *ITGA6,* up-regulated in the aggressive adrenal and pituitary, plays a role in the adhesive function of the hemidesmosome (Jones et al., 1991). Heparan sulfate changes were represented by the second most up-regulated gene in the tame adrenal, *HS3ST5,* a heparan sulfotransferase. This gene was also a member of the Green module, along with *SDC4,* a heparan sulfate proteoglycan. The recurring theme of cellular adhesion in the analysis of the tame and aggressive fox adrenal transcriptomes is particularly notable in light of its centrality in the differences between the tame and aggressive pituitary transcriptomes (Chapter 3).

Cellular adhesion directly affects cellular motility, and significantly enriched GO terms related to cellular motility, particularly of pseudopodia formation, were also found in analysis of tame and aggressive fox pituitary gene expression (Chapter 3). Pseudopodia and motile cells may function in some endocrine-producing cells to facilitate the release of hormones into the systemic circulation through contacting blood vessels (Navratil et al., 2007) and/or by forming networks to coordinate hormone release (Hodson et al., 2012a; Hodson et al., 2012b). Pseudopod-like processes have been observed in adrenal medullary and cortical cells (Motta et al., 1979; Kilpatrick et al., 1980) and have been demonstrated to extend with hormonal or electrical stimulation in cultured cells from both adrenal regions (Manivannan and Terakawa, 1994; Shepherd and Holzwarth, 2001 Colonna and Podestá, 2005). Cellular motility was represented in the tame and aggressive fox adrenals through both GO enrichment analysis and differential exon splicing. GO analysis identified 45 terms related to motility, migration, and chemotaxis enriched (but not significantly) for differentially expressed genes, as well as 14 terms relating to cellular projections associated with motility, such as pseudopodia and cilia. Genes with significantly differentially spliced exons between tame and aggressive fox adrenals included two genes associated with cilia, *OFD1* and *RFX2* (Ferrante et al., 2006; Bisgrove et al., 2012) and one associated with cell shape and migration (*FMNL3* [Bai et al., 2011]). While genes relating to pseudopodia are not as widely represented in the adrenals analysis as they were in the pituitary, one particular gene, *CDC42,* may be important in both tissues. *CDC42* functions in the extension and maintenance of filopodia, a type of pseudopodia (Jacquemet et al., 2015); while it is not itself differentially expressed in fox pituitary or adrenals, two of its effector proteins, *CDC42EP1* and *CDC42EP3,* are up-regulated in the tame adrenal, and are members of the adrenal Green module. Another, *CDC42EP2,* is up-regulated in the tame pituitary. In addition, *SRGAP2,* which

inactivates *CDC42* (Guo and Bao, 2010), is up-regulated in the aggressive adrenal. This analysis suggests that cellular motility and formation of pseudopodia may differ between the tame and aggressive adrenals, similarly to the tame and aggressive pituitary.

A critical step in the release of hormones into the bloodstream is exocytosis; in a comparison of tame and aggressive fox anterior pituitary gene expression, expression differences in genes related to exocytosis were identified. Here, differences in expression of genes related to exocytosis are also seen between tame and aggressive fox adrenals. GO analysis of tame and aggressive fox adrenals described 38 terms related to exocytosis enriched for differentially expressed genes. Additionally, the most highly connected gene in the Magenta module, *SNAP25,* is associated with vesicle docking and membrane fusion, both steps in the exocytotic process (Pevsner et al., 1994). This gene is particularly interesting because a risk allele has been identified for attention deficit/hyperactivity disorder (Liu et al., 2017), suggesting its relevance to behavioral traits. Here again the analysis of the fox adrenal gland shows similarity to findings in the fox anterior pituitary.

While differences in activity of the adrenal cortex in tame and aggressive foxes have been described by physiologic assays such as measurement of cortisol concentrations, differences in activity of the adrenal medulla between these two lines are less well understood. Some of our findings are suggestive of medullary-specific differences. The adrenal medulla is considered part of the sympathetic nervous system, similar to a sympathetic ganglion, so changes related to synapses and cellular excitation are likely to be indicative of changes in this region of the adrenals specifically. The most strongly loaded gene on PC2, the principal component differentiating tame from aggressive samples, is *SYNDIG1L,* which is associated with development of excitatory synapses (Kalashnikova et al., 2010). Another highly loading gene on PC2, *SCN3A*, up-regulated in the aggressive adrenal, is a sodium ion channel, an essential component in cellular excitation. A sodium ion channel from the same family, *SCN2A,* is also up-regulated in the aggressive adrenal. These findings suggest a possible difference in excitability in the tame and aggressive adrenal medulla. Another of the strongly loading genes in PC2 is the α2-adrenoceptor (*ADRA2A)*, up-regulated in the aggressive adrenal. α2-adrenoceptors in the adrenal medulla are activated by binding of epinephrine and norepinephrine, and act to inhibit the release of norepinephrine in a negative feedback mechanism (Moura et al., 2006). Their up-regulation in the aggressive adrenal suggests a possible response to likely elevated

catecholamines in these foxes. These findings are similar to those in a gene expression study that implicated cell signaling and ion channel-related genes, among others, in the changes in adrenal activity in the adrenal medulla of rats after repeated immobilization (Liu et al., 2008).

Genes related to angiogenesis also appeared in multiple analyses, including five GO terms relating to angiogenesis or blood vessel formation enriched for genes in the Tan module, and three GO terms relating to collagen (an essential component of arteries), also enriched for genes in the Tan module. One of only two significant GO terms enriched for differentially expressed genes, *negative regulation of interleukin-8 production*, relates to a cytokine which is a known mediator of angiogenesis (Koch et al., 1992) (and of cholesterol [Chen et al., 2014]). One of the most highly loading genes on PC2, *SERPINF1* (up-regulated in the tame adrenal), is a precursor for pigment epithelium-derived factor, a potent inhibitor of angiogenesis (Dawson et al., 1999). Blood flow through the adrenal glad is essential for the step-wise process of steroid biosynthesis, which requires movement of outputs to the adjacent cortical zone, where they become precursors for the next phase. It is also essential for delivery of steroid and catecholamine hormones to the systemic blood stream. Therefore, any functional differences in angiogenesis between tame and aggressive foxes may act as a means of regulating steroidogenesis or hormone release in their adrenal glands.

The GO term *positive regulation of response to DNA damage stimulus* was significantly enriched in genes that are differentially expressed in both the adrenal and pituitary glands of tame and aggressive foxes, suggesting that DNA damage repair may be differentially regulated between the two lines in multiple tissues. Genome analysis of tame and aggressive foxes identified regions with low heterozygosity in each line as putative regions of selection, and the GO term *damaged DNA binding* was enriched for the genes in one such region (Kukekova et al., in review). It is possible that selection for tame behavior may have benefited from selection for DNA instability and therefore increased opportunity for mutations to provide desirable traits. This supposition has not been explored, but is supported by the differential expression of genes related to this mechanism in multiple tissues.

Although the expression differences between tame and aggressive fox adrenals found here were not specific to glucocorticoid regulation, our results echo those from other studies. In a comparison of different breeds of pig, levels of the critical steroidogenesis proteins StAR and

P450scc differed by breed, but did not correlate with serum cortisol concentration across breeds (Wang et al., 2014). An investigation of the adrenal gene expression response to social stress in pigs concluded that most of the differentially expressed genes appeared to be involved in pathways not directly controlled by the response to ACTH (Muráni et al., 2011). A comparison of gene expression differences between domesticated chickens and their ancestor, the red jungle fowl, also demonstrated changes in adrenal genes outside of the steroidogenic pathway (Fallasharoudi et al., 2015). Our findings suggest that instead of steroidogenesis, future studies should look to differences in cell signaling (particularly cell migration and adhesion), exocytosis, and angiogenesis or perfusion of the adrenals for a deeper understanding of regulation of the adrenal glands.

4.5. CONCLUSIONS

Transcriptome analysis of tame and aggressive fox adrenal glands suggests differences in functionality between the two lines related to metabolites early rather than late in steroidogenesis, specifically cholesterol and fatty acids, suggesting that regulation differences between these lines are not glucocorticoid-specific. Findings related to synapse development and sodium ion channels suggest the possibility that excitability of the cells in the adrenal medulla may differ between the lines, though physiologic studies on the foxes have to date focused on the HPA axis rather than the adrenal medullary response. Intriguingly, some adrenal findings suggest similarities to pituitary findings, particularly related to cellular adhesion, motility, and exocytosis, raising the possibility that cellular signaling changes could be consistent across different endocrine tissues in this model. Differences in expression of genes related to DNA damage repair were also similar in both adrenal and pituitary tissues, echoing similar results in whole genome analysis of tame and aggressive foxes.

CHAPTER 5. VARIANT ANALYSIS OF TAME AND AGGRESSIVE FOX RNA READS

5.1. INTRODUCTION

Complex traits are usually influenced by a large number of genes of small effect size, making pinpointing causative variants that affect these phenotypes difficult. However, selection can leave signatures on the genome that allow these loci to be identified. As advantageous variants rise in frequency in a population in response to positive selection, alleles in adjacent regions will "hitchhike" along to also rise in frequency. This linkage disequilibrium (LD) – the association of functionally unrelated alleles due to proximity on a chromosome – breaks down over many generations, but until its decay is complete, reduced heterozygosity provides a tool for finding a variant under selection in a region of a "selective sweep" (Maynard Smith and Haigh, 1974). Additionally, the fixation index (F_{ST}) can be used for identification of regions with significantly different allele frequencies between populations (Nei and Chesser, 1983), which may also pinpoint regions favored by selection.

Discovery of a causal variant within a putative selective sweep can be complicated by extensive LD in the region, as variants of interest may be embedded within long stretches of reduced heterozygosity (Goldstein, 2001). Additionally, random fixation due to genetic drift, particularly likely to occur in small populations, can be difficult to distinguish from true selection events (Orr, 1998). Despite these obstacles, studies using this selective sweep-based approach have been quite successful in establishing genomic regions of interest (Rubin et al., 2010; Vaysse et al., 2011; Axelsson et al., 2013; Wang et al., 2013; Wang et al., 2014; Cagan and Blass, 2016). Once putative regions under selection have been established, genes within them may be considered candidate genes for subsequent studies of the traits under investigation. These studies may sequence the identified regions to further refine them, and may thereby potentially identify a causal gene, causal haplotype, or even the specific causal polymorphism.

One of the first studies of selective sweeps in domesticated animals using next generation sequencing technologies successfully employed this selective sweep approach to implicate

regions of putative selective sweeps using populations of chickens bred for different agricultural purposes (Rubin et al., 2010). Because chickens have been bred for meat and eggs, the phenotypes of interest were growth, particularly of muscle, and laying. Red jungle fowl (the ancestor of the domesticated chicken), layers, and broilers were sequenced. Variant analysis identified regions carrying two selective signatures of interest: first, decreased pooled heterozygosity (Hp) within lines, with the expectation that reduced Hp serves as a proxy for decreased diversity and therefore of selection, and second, increased fixation index (F_{ST}) between lines, with the expectation that differences in allele frequencies between populations indicate increased genetic distance and therefore selection. Domestic chickens are known to be highly diverse, which facilitates the identification of selected variants, and 58 windows of reduced Hp were identified in the domestic chicken populations. Notably, the analysis successfully identified a selective sweep centered around an allele associated with a trait, yellow skin, known to have been under selection during the domestication of chickens, thereby demonstrating the approach's efficacy. This study successfully identified some intriguing candidate genes with functional relevance, such as genes related to muscle growth, or for the thyroid hormone receptor, but their approach using pooled reads did not support the identification of specific causative haplotypes or variants.

In comparison to chickens, dogs are less genetically diverse, resulting in longer regions of LD. Nevertheless, analyses of selective sweeps in the dog genome have successfully identified causal haplotypes (Sutter et al., 2007; Vaysse et al., 2011; Wang et al., 2013; Axelsson et al., 2014; Wang et al., 2014; Cagan and Blass, 2016). An early successful study fine-mapped a region in the dog genome associated with body size to successfully identify a causative haplotype in an individual gene (Sutter et al., 2007). This study used a variety of approaches to overcome the problem of reduced diversity and increased LD in the species. Initial identification of variants was performed within a single breed with extensive phenotypic variance for size, the Portuguese Water Dog, (PWD). The region of interest was subsequently sequenced in a comparison of dogs from small versus giant breeds, and the previously identified variants were employed in these samples to identify a peak in F_{ST} and trough in Hp around the *IGF1* gene, which is known to play a role in growth. Finally, haplotype analysis identified the causative haplotype, which explained 15% of size variation in the PWD. Subsequent studies have also successfully employed the selective sweep approach to identify genes associated with complex

phenotypes in this species (Vaysse et al., 2011; Wang et al., 2013; Wang et al., 2014; Cagan and Blass, 2016).

Despite the difficulties in working with a species with extensive LD, a comparison of Hp and F_{ST} in dog and wolf did identify a causal variant involved in domestication (Axelsson et al., 2013). This comparison identified 14 regions of reduced Hp and 35 of increased F_{ST}. In an attempt to focus on regions of selective sweeps rather than of random fixation, regions were investigated for evidence of functional change, i.e., genes with known functions related to domestication. This approach may have produced some false negatives in regions containing genes with poorly understood functions, but did identify increased copy number variation (CNV) in an amylase gene, involved in starch digestion, in the dog as compared to the wolf. The diet of the domesticated dog includes significantly more starch than does that of the wolf, due to the proximity of the dog to human agriculture as compared to the wolf's specialization in hunting, making an increase in amylase beneficial for dogs. Compared to the causative variant in canine size in the *IGF1* gene, which appears likely to involve a single or small number of nucleotide differences in the identified haplotype, amylase CNV was easier to identify. The study authors noticed increased sequencing depth in that region, a marker not available to the authors of the *IGF1* study who used a SNP array in their experiment. Validation of the amylase variant was also facilitated by the opportunity to perform a gene expression study and establish increased expression in animals with more copies of the gene; not all variants provide researchers with these advantages. This study was able to provide targeted analysis of variation in a particular gene, which allowed support for a long-standing theory of dog domestication, demonstrating the value of sweep-based approaches to the identification of causal variants associated with complex traits.

Discovery of causative variants involved in the tame and aggressive fox phenotypes would provide valuable new resources in the attempt to understand the biological underpinnings of these distinct behavioral suites. Early work using a fox meiotic linkage map established quantitative trait loci (QTL) describing genomic regions associated with differences between the tame and aggressive behavioral phenotypes (Kukekova et al., 2011b; 2012; Nelson et al., 2017), three of which overlapped with previously identified canine domestication regions (vonHoldt et al, 2010; Axelsson et al., 2013). However, these QTL intervals were too broad to be used to identify specific genes of interest (Nelson et al., 2017).

Genotyping-by-sequencing (GBS) and an F_{ST} analysis were used to investigate these QTL regions and to identify new regions of interest (Johnson et al., 2015). This study characterized LD in fox populations, observing an average $r^2 \geq 0.2$ for SNPs located within 500 kb in tame and 100 kb in the aggressive population; these values are comparable to average LD previously reported for dog breeds ($r^2 \geq 0.2$) (Lindblad-Toh et al., 2005). Nine genomic regions of high F_{ST} were identified by comparison of tame and aggressive fox populations, some of which overlapped with previously identified behavioral QTL on fox chromosome 14. The identification of these areas of overlap makes their allele frequency differences less likely to be due to drift and more likely to be due to linkage with causative variants for behavioral differences between the tame and aggressive lines. However, this approach was unable to provide enough resolution to identify causative genes or haplotypes.

More recently, studies seeking to analyze tame, aggressive, and conventional foxes have been able to utilize whole genome sequencing with alignment of fox reads against the newly sequenced fox genome (Kukekova et al., under review). This approach allowed high resolution investigation similar to that in the previously described chicken and dog studies, identifying 113 regions of low pooled heterozygosity (Hp) and/or high F_{ST}. Of these, 29 overlapped with previously identified behavioral QTL, making them likely true targets of selection. One candidate gene, *SorCS1,* was found through investigation of a genomic region that fell within a behavioral QTL and contained a single gene. Identification of the candidate haplotype required additional genotyping, and the haplotype was validated in the F2 population as having an effect on behavior. Identification of additional variants with effects on the tame phenotype are also likely require additional sequencing, but may benefit from the regions defined in this study.

Genome-based approaches excel at the identification of regions of interest, but benefit from synergistic studies using different approaches to help pinpoint causative genes or even candidate causative variants. A recent study employed variant analysis of RNA reads from frontal cortex and basal forebrain samples of tame and aggressive foxes (Wang et al., under review). Samples in this study were sequenced at sufficient depth to support analysis of individual genotypes, and the use of tissues of particular interest to the behavioral phenotype enables focus on genes active in those tissues and therefore more likely to be associated with the relevant phenotype. The use of RNA reads also provides access to gene expression data as an additional means of triangulating on genes of interest. This study again implicated *SorCS1*, and also identified a new

missense variant in the coding region of *GRM3,* a glutamate receptor. Although the relevance of these changes to the tame phenotype is difficult to assess without a comparison of genotype to behavioral assessment, the identification of candidate genes suggests the power of the transcriptome approach.

An examination of reads from HPA-axis related tissues could provide valuable information about variants in genes active in a system that is known to differ between the tame and aggressive foxes, and that is believed to be influential in the development of the distinct behavioral phenotypes. Additionally, a comparison of variants in these tissues to previously identified regions of putative selective sweeps could provide additional information on potential variants of interest in the given region. In this study, exonic variants in RNA reads from hypothalamus, pituitary, and adrenal tissue, the primary tissues associated with the hormonal stress response which is known to differ between tame and aggressive foxes, were annotated with putative impact and F_{ST}. Of particular interest were variants with high impact (such as frameshift or splice acceptor changes), or variants with moderate impact (such as missense variants). High or moderate impact variants in regions with high F_{ST} values in this data as well as regions of putative selective sweeps identified in previous analyses will be selected for potential future analysis to ascertain the likelihood that they are causal variants contributing to the tame phenotype. The use of variants called from RNA data, along with gene expression information and previously described genomic regions of interest, will provide a powerful approach to identification of potentially causative genes and variants implicated in the tame and aggressive fox behavioral phenotypes.

5.2 METHODS

5.2.1. Animals, sample preparation

Foxes (12 tame and 12 aggressive) were maintained and euthanized as described previously. Hypothalamus was removed from all 24 foxes in addition to anterior pituitary and right adrenal glands, as previously described. Samples were stored, prepared, and sequenced as previously described. Reads were filtered for quality, and mitochondrial and ribosomal reads were removed (see methods section in Chapters 3 and 4).

5.2.2. Variant analysis

Variant calling was performed on hypothalamus samples from 12 tame and 12 aggressive foxes, pituitary samples from a subset of 6 tame and 6 aggressive foxes, and adrenals from a subset of 11 tame and 11 aggressive foxes. All reads were aligned against the dog genome (CanFam3.1) as previously described (see methods section in Chapters 3 and 4). SNPs and indels, including both dog-fox and fox-fox variants, were derived using the Genome Analysis Toolkit (GATK) version 3.5 (McKenna et al., 2010). First, reads were split on N cigar elements using SplitNCigarReads (GATK). Reads that mapped to multiple locations in the genome with equal accuracy were removed using SAMtools 1.1 (Li et al., 2009). Indels were realigned using RealignerTargetCreator and IndelRealigner (GATK). Base qualities were recalibrated using BaseRecalibrator (GATK) against a set of 24 million fox SNPs, which included both intraspecies fox variants and fox-dog variants (Kukekova et al., under review). The recalibrated alignments were used to call variants on individual samples using HaplotypeCaller (GATK), and these variants were pooled for joint genotyping using GenotypeGVCFs (GATK). The following hard filters were applied to the resulting variants: FisherStrand (FS) > 60.0, QualByDepth (QD) < 2.0, RMSMappingQuality (MQ) < 18.0, and StrandOddsRatio (SOR) > 10.0. VCFtools (Danecek et al., 2011) was used to remove sites with more than two alleles on the assumption that they were likely to represent sequencing errors. Sites that included fewer than 5 samples containing the dog reference allele (out of 60 total samples) were presumed to represent fox versus dog polymorphisms and not fox versus fox polymorphisms, and were removed using VCFtools. Variants without at least 10 reads in at least 22 samples (hypothalamus), 12 samples (anterior pituitary), or 20 samples (adrenals) were discarded from the analysis. Further analysis was performed in each tissue separately only on these "filtered variants."

SNPeff (Cingolani et al., 2012) was used to annotate the filtered variants for possible functional biological impact in fox orthologs of dog genes. Variants annotated by SNPeff as having "high" or "moderate" impact ("HMI variants") were selected for further analysis. A list of genes associated with these variants was examined for enriched gene ontology terms using ClueGO version 2.3.3 (Bindea et al., 2009) in Cytoscape version 3.5.1 (Shannon et al., 2003), with the February 23, 2017 version of the Gene Ontology (GO). The ClueGO evaluation was performed against a background list of genes associated with all filtered variants. Short

descriptions for each gene were acquired through the biomaRt package in R (Smedley et al., 2015), using the host *www.ensembl.org* and the dataset *cfamiliaris_gene_ensembl.*

Additionally, the fixation index (F_{ST}) was calculated for each filtered variant in R using the estimator formula provided in Karlsson et al., 2007. Variant locations with $F_{ST} > 0.5$ ("high F_{ST} variants") were compared to regions of the genome previously identified as associated with putative selective sweeps; these regions were originally identified in the draft fox genome, but region locations were transformed to CanFam3.1 coordinates by mapping the fox assembly against CanFam3.1 (Kukekova et al., under review). High F_{ST} variants were also analyzed for GO term enrichment using ClueGO as described previously.

5.3. RESULTS

5.3.1. Hypothalamus variants

Hypothalamus sequencing results are described in Table SI-5.1. After filtering for quality and removing putative fox-versus-dog variants, 178,140 variants were identified in the combined hypothalamus, pituitary, and adrenal samples; 146,090 of these variants were polymorphic in the hypothalamus samples. To study variants in hypothalamus reads, analysis proceeded with hypothalamus samples only. After filtering for low coverage and missing data, 19,821 sites remained (Table SI-5.2), associated with 6,512 unique genes. Of these variants, 1,358 were insertion-deletion polymorphisms. Among all the variants (SNPs and small insertions-deletions), 40 were annotated by SNPeff as being high and 1,725 as moderate impact variants (Table SI-5.3), in 1,148 unique genes (Table SI-5.4). ClueGO analysis identified 31 terms significantly (FDR < 0.05) enriched for genes containing high or moderate impact ("HMI") variants (Table SI-5.5). Of these, 25 terms fell into four groups based on overlapping lists of HMI genes (i.e., some HMI genes fall into multiple groups): one group of 19 terms, including 13 of the 14 most significantly enriched, related to transcription regulation; a second group containing 3 terms relating to translation in mitochondria; a third group containing 2 terms relating to ion transport; and a fourth group containing 4 terms associated with phosphorylation and transferase activity. Three significantly enriched terms did not group with any other terms; these were *viral life cycle, non-motile cilium assembly,* and *regulation of signaling.*

F_{ST} analysis identified 647 filtered variants with $F_{ST} > 0.5$ (Table SI-5.6), in 352 unique genes (Table SI-5.7). When variants were aligned to regions of putative selective sweeps as identified in (Kukekova et al., under review), 33 out of 113 regions contained at least one high F_{ST} variant (Table SI-5.8), and 236 of the high F_{ST} variants fell within a region. ClueGO analysis identified one term significantly (FDR < 0.05) enriched for genes containing high F_{ST} variants (Table SI-5.9): *mitochondrion organization* (GO:0007005), FDR = 0.01.

F_{ST} of high impact variants was 0-0.29 (mean 0.07); high impact variants with $F_{ST} > 0.2$ included a splice region variant in *RAB1A,* a GTPase associated with cell migration and adhesion (Wang et al., 2010) and vesicular protein transport (Mukhopadhyay et al., 2011); and a frameshift variant in *RBBP6,* a ubiquitin-protein ligase. F_{ST} of moderate impact variants was 0-0.83 (mean 0.15) (Figures SI-4.1 and SI-4.2). Of the high F_{ST} variants, 60 were also HMI (Table SI-5.10) (all "moderate" impact). HMI variants with $F_{ST} > 0.75$ included missense variants in the genes *PDE4DIP,* which may function as an anchor to sequester components of the cAMP pathway (Verde et al., 2001); *PHTF1,* a putative homeodomain transcription factor (Manuel et al., 2000); *RYR2,* a calcium channel mediating the release of calcium from the sarcoplasmic reticulum (Xiao et al., 2007); and *SORT1,* involved in vesicle-mediated transport (Nielson et al., 2001).

5.3.2. Anterior pituitary variants

Of the 178,140 variants identified in the combined hypothalamus, pituitary, and adrenal samples, 146,031 variants were polymorphic in the anterior pituitary samples. After filtering, 21,926 sites (Table SI-5.11) remained, associated with 7,245 unique genes. Of these variants, 1,555 were insertion-deletion polymorphisms. Among all the variants (SNPs and small insertions-deletions), 48 were annotated by SNPeff as being high and 2,202 as moderate impact variants (Table SI-5.12), in 1,689 unique genes (Table SI-5.13). Of those genes, 36 were differentially expressed between tame and aggressive fox anterior pituitary. ClueGO analysis identified 46 terms significantly (FDR < 0.05) enriched for genes containing HMI variants (Table SI-5.14), which fell into 6 groups based on overlapping genes. These groups were: 12 terms related to ion transport (including 4 of the top 5 most significantly enriched terms); 5 terms related to cellular signal transduction; 2 terms related to organ development; 4 terms related to

translation in mitochondria; 4 terms related to DNA damage repair; and 4 terms related to collagen metabolism. The remaining significantly enriched terms, which did not group with other terms, were *glial cell differentiation, maturation of SSU-rRNA, neurological system process, protein localization to microtubule organizing center, positive regulation of protein phosphorylation, protein retention in ER lumen, single-organism catabolic process,* and *non-motile cilium assembly.*

F_{ST} analysis identified 1,484 filtered variants with $F_{ST} > 0.5$ (Table SI-5.15), in 713 unique genes (Table SI-5.16), and 31 of those genes were differentially expressed between tame and aggressive fox pituitary. When variants were aligned to regions of putative selective sweeps as identified in (Kukekova et al., under review), 46 out of 113 regions contained at least one high F_{ST} variant (Table SI-5.17), and 422 variants fell within a region.

ClueGO analysis identified three GO terms significantly (FDR < 0.05) enriched for genes containing high F_{ST} variants (Table SI-5.18): *negative regulation of signaling, negative regulation of signal transduction,* and *negative regulation of cell communication*. F_{ST} of high impact variants was 0-0.33 (mean 0.07); the high impact variant with the highest F_{ST} was the *RAB1A* splice region variant identified in the hypothalamus analysis. F_{ST} of moderate impact variants was 0-0.92 (mean 0.17) (Figures SI-5.3 and SI-5.4). Of the high F_{ST} variants, 148 were also HMI (Table SI-5.19); HMI variants with $F_{ST} > 0.75$ included *TXNRD1,* a thioredoxin reductase involved in the formation of cell membrane protrusions (Dammeyer et al., 2008); *SLC2A4RG,* a transcription factor associated with a solute carrier; *ZSCAN31,* a zinc finger transcription factor; *TAPBPL,* a component of the antigen processing and presentation pathway (Boyle et al., 2013); *SNX19,* associated with intracellular vesicle trafficking and exocytosis (Harashima et al., 2012); and two genes also identified in the hypothalamus analysis, *PHTF1* and *PDE4DIP*.

Eleven genes were differentially expressed in tame and aggressive fox anterior pituitary, contained at least one variant with $F_{ST} > 0.5$, and also contained at least one HMI variant (Table 5.1). None of these variants were of "high" impact. Some of the variants were both HMI and high F_{ST}, but some were either one or the other.

Table 5.1. Variants associated with genes that are differentially expressed between tame and aggressive fox pituitary; have at least one HMI variant; and also have at least one high F_{ST} variant. For information on regions of putative selective sweeps that variants fall into, see Table SI-5.17.

Gene	Variant location (CanFam3.1)	Region	Ref/alt allele	Tame/aggr alternate allele frequency	F_{ST}	Annotation
A2M	27:36,671,351	46	C/T	0.08/0	0.08	Ala596Val
	27:36,677,013		C/G	0.75/0	0.75	Synonymous
	27:36,683,565		C/A	0.67/0	0.67	
	27:36,689,024		C/G	0.08/0	0.08	Pro1200Ala
	27:36,690,746		A/C	0.08/0	0.08	Arg1264Ser
CYB5R4	12:44,175,243	None	G/A	0.91/0.33	0.53	Asp70Asn
FANCL	10:58,545,880	None	T/A	0/0.58	0.58	Ile342Phe
	10:58,554,231		G/A	1/0.42	0.58	Ala252Val
FLNB	20:32,395,395	None	T/C	0.75/0.08	0.63	3' untranslated region
	20:32,395,924		C/T	0.75/0.08	0.63	Synonymous
	20:32,398,782		G/A	1/0.5	0.5	
	20:32,408,035		T/C	0.5/0.5	0	Met2324Val
MRPL35	17:38,917,454	None	C/T	0.58/0	0.58	Cys60Tyr
OAS1	26:10,385,650	None	G/T	0/0.25	0.25	Ala50Ser
	26:10,394,077		C/T	0.83/0	0.83	Synonymous
	26:10,396,703		A/G	0.17/1	0.83	3' untranslated

Gene	Variant location (CanFam3.1)	Region	Ref/alt allele	Tame/aggr alternate allele frequency	F_{ST}	Annotation
						region
PEX1	14:18,105,234	None	G/C	1/0.5	0.5	Leu579Val
PPP1R26	9:51,199,915	None	G/A	0.75/0.17	0.51	Glu222Lys
	9:51,200,908		G/A			Gly553Ser
	9:51,201,981		C/T			Synonymous
	9:51,202,012		G/A	0.17/0.83	0.62	Ala921Thr
	9:51,202,318		G/A			Gly1023Ser
SNX19	5:4,045,726	None	G/A	0.08/0.92	0.82	Synonymous
	5:4,045,829		G/A			Asp150Asn
	5:4,046,337		C/T	0.42/0.08	0.26	Pro319Leu
	5:4,083,936		T/G	0.08/0.92	0.82	Synonymous
TMEM168	14:52,312,935	None	A/T	0.17/0.75	0.51	3' untranslated region
	14:52,312,952		C/T	0.17/0.75		
	14:52,313,065		G/T	0.83/0.25		
	14:52,313,561		C/G	0.17/0.75		
	14:52,315,145		T/C	0.17/0.75		Asn601Ser
	14:52,337,343		A/G	0.17/0.75		Synonymous
ZDHHC14	1:47,216,111	None	G/A	0/0.5	0.5	Ala419Thr

Table 5.1 (cont.)

5.3.3. Adrenals variants

Of the 178,140 variants identified in the combined hypothalamus, pituitary, and adrenal samples, 138,387 variants were polymorphic in the adrenal samples. Variants without at least 10 reads in at least 20 adrenal samples or without all genotypes called in all samples were discarded, leaving 19,330 sites (Table SI-5.20), associated with 6,491 unique genes. Of these variants, 1231 were insertion-deletion polymorphisms. Among all the variants (SNPs and small insertions-deletions), 41 were annotated by SNPeff as being high impact variants, and 2,026 as moderate impact variants (Table SI-5.21). These high and moderate impact (HMI) variants were distributed among 1,271 unique genes (Table SI-5.22), of which 65 were differentially expressed between tame and aggressive fox anterior pituitary. ClueGO analysis identified 104 terms significantly (FDR < 0.05) enriched for genes containing HMI variants (Table SI-5.23). Nine terms were enriched at FDR < 0.005: 5 terms related to ion transport, 2 terms related to cell signal transduction, *animal organ morphogenesis,* and *regulation of multicellular organismal process*.

F_{ST} analysis identified 767 filtered variants with $F_{ST} > 0.5$ (Table SI-5.24), in 381 unique genes (Table SI-5.25); 49 of those genes were differentially expressed between tame and aggressive fox adrenals. When variants were aligned to regions of putative selective sweeps as identified in (Kukekova et al., under review), 34 out of 113 regions contained at least one high F_{ST} variant (Table SI-5.26), and 255 variants fell within a region.

ClueGO analysis did not identify any terms significantly (FDR < 0.05) enriched for genes containing high F_{ST} variants (Table SI-5.27). F_{ST} of high impact variants was 0-0.37 (mean 0.06); the high impact variant with the highest F_{ST} did not have an associated gene symbol, and the variant with the next highest F_{ST} was the previously identified *RAB1A*. Of the high F_{ST} variants, 76 were also HMI (Table SI-5.28); HMI variants with $F_{ST} > 0.75$ included *CACNA1C*, a calcium voltage-gated channel, and *PHTF1,* which was also identified as HMI with a high F_{ST} in the hypothalamus and pituitary samples.

Thirteen genes were differentially expressed in tame and aggressive fox adrenals, contained at least one variant with FST > 0.5, and also contained at least one HMI variant (Table 5.2). None of these genes contained a “high” impact variant. Some of these variants were both HMI and high F_{ST}, but some were either one or the other.

Table 5.2. Variants associated with genes that are differentially expressed between tame and aggressive fox adrenals; have at least one HMI variant; and also have at least one high F_{ST} variant. For information on regions of putative selective sweeps that variants fall into, see Table SI-5.27.

Gene	Variant location (CanFam3.1)	Region	Ref/alt allele	Tame/aggr alternate allele frequency	F_{ST}	Annotation
ANKZF1	25,779,786	None	T/C	0.5/1	0.5	Synonymous
	25,780,487		A/C	0.05/0.14	0.05	Lys268Thr
	25,780,925		G/A	0.14/0.59	0.36	Arg324His
	25,781,374		G/C	0.5/0	0.5	Gln419His
EXOC6	7,310,874	None	G/A	0.59/0.05	0.52	Ser529Asn
FLNC	14:7,802,919	None	G/A	0.64/0	0.64	Ser1305Leu
	14:7,812,472		G/A	0.27/0.95	0.66	Synonymous
GLB1L	25,784,683	None	G/A	0.5/1	0.5	Synonymous
	25,787,425		A/G			Synonymous
	25,792,288		C/T	0.14/0.59	0.36	Ala43Thr
HNMT	19:40,701,871	None	C/T	0.05/0.59	0.51	Thr221Met
	19:40,703,738		C/T	0.51/0.05	0.59	Downstream
	19:40,703,917		C/T			
	19:40,704,849		A/T			
KCNN1	20:44,957,608	54	T/C	0.86/0.36	0.41	Ile158Val

Gene	Variant location (CanFam3.1)	Region	Ref/alt allele	Tame/aggr alternate allele frequency	F_{ST}	Annotation
	20:44,954,522		G/A	0.68/0	0.68	Synonymous
	20:44,954,699		C/T			
MRPL35	17:38,917,454	52	C/T	0.59/0	0.58	Cys60Tyr
MYOF	28:7,605,345	None	T/C	0.14/0.32	0.09	Asn1974Ser
	28:7,617,379		G/A	0.54/0	0.54	Synonymous
	28:7,638,955		G/T			
	28:7,654,332		G/A	0.60/0.05	0.51	
	28:7,673,275		C/T			
	28:7,688,372		A/T	0.08/0	0.08	Phe321Tyr
PHTF1	17:51,541,448	50	C/T	0.68/0.05	0.61	Synonymous
	17:51,546,388		G/A	0.05/0.86	0.81	Leu204Phe
PTCD3	17:38,980,353	52	G/T	0.27/0.95	0.66	Thr339Asn
SNX19	5:4,045,515	None	T/C	0.32/0	0.32	Leu45Pro
	5:4,045,726		G/A	0.23/0.91	0.64	Synonymous
	5:4,045,829		G/A			Asp150Asn
	5:4,046,215		G/A			Synonymous
	5:4,046,337		C/T	0.32/0	0.32	Pro319Leu
	5:4,083,936		T/G	0.23/0.91	0.64	Synonymous

Table 5.2 (cont.)

Gene	Variant location (CanFam3.1)	Region	Ref/alt allele	Tame/aggr alternate allele frequency	F_{ST}	Annotation
SULT1C4	10:35,503,822	None	A/C	0.86/0.41	0.36	Ile34Ser
	10:35,494,743		G/A	0.18/0.77	0.52	3' untranslated region

Table 5.2 (cont.)

5.3.4. Cross-tissue analysis

Pooling variants from all three tissues (hypothalamus, anterior pitutary, and adrenals), 546 high F_{ST} variants were found in total that fell within regions of putative selective sweeps as identified in (Kukekova et al., under review) (Table SI-5.28). ClueGO analysis was performed on the 311 unique genes associated with this list, using a background list of 8,850 genes, comprising the union of genes associated with filtered variants from all three tissues. No significantly enriched terms were found (Table SI-5.29).

5.4. DISCUSSION

RNA sequencing of hypothalamic, pituitary, and adrenal reads from foxes described 178,140 variants, of which those annotated with high or moderate impact and with high F_{ST} may contribute to the biological differences between tame and aggressive foxes. Because of the RNA origin of the reads, it was possible to contextualize polymorphisms based on previous gene expression studies in anterior pituitary and adrenal tissue, and thereby identify variants more likely to have been under selection in the development of the tame and aggressive phenotypes. Variants were found in genes associated with cell signaling (including cell adhesion, cell morphology, and exocytosis), ion channels, and DNA damage repair, all processes that were implicated in previous gene expression studies. Variants were also found in genes related to organ development and translation in mitochondria, processes that were not implicated in gene expression studies in those tissues. These variants may have been subject to drift that elevated their F_{ST} by chance or may have been under selection during the selection of foxes for behavior, but the specific process that they affect may not operate in the anterior pituitary or adrenal tissues

of adult animals, or they may in fact be associated with different processes, as GO enrichment analysis can return false positive findings.

Many, but not all, of the high F_{ST} variants fell within previously identified regions of putative selective sweeps. High F_{ST} variants discovered outside of regions of putative sweeps may in fact fall within smaller sweep regions that did not rise to the level of significance in a previous study in which the genome was analyzed in 500 Kb windows (Kukekova et al., under review). Therefore, variants in this study that fall within putative selective sweeps are considered higher priority for future investigation.

GO analysis was performed on variants in each tissue with high or moderate impact (as annotated with SNPeff), and on variants with high F_{ST}, to characterize the functionality of groups of genes associated with these variants. Terms relating to cell signaling were significantly enriched in the HMI variant analysis in all three tissues, as well as the high F_{ST} analysis in pituitary tissue; 8 GO terms were enriched for HMI genes across all three tissues studied. These terms showed significant overlap of genes; 12 HMI genes were in every enriched term related to cell signaling. Of these, three genes had variants with $F_{ST} > 0.2$: *GRIK5* ($F_{ST} = 0.32$; Val/Ile missense), a glutamate receptor gene with the alternate allele fixed in the tame population; *ATP2B4* ($F_{ST} = 0.27$, Gln/Leu missense), associated with calcium transport out of the cell (Adamo et al., 1995); and *PIP5K1C* ($F_{ST} = 0.22$, Thr/Asn missense), which codes for a protein that localizes to cellular adherens junctions (Van den Bout and Divecha, 2009). RNA-seq findings support the importance of cell signaling in the differences in activity between tame and aggressive fox anterior pituitary and adrenal glands, particularly cell adhesion. Calcium signaling is an important step in the exocytotic pathway (Burgoyne and Morgan, 2003), which was shown to be differentially regulated in the tame and aggressive fox adrenals, as well as anterior pituitary (Chapters 3 and 4). Those findings identified differentially expressed genes and gene modules implicated in these changes; this set of variants relating to cell signaling may provide candidates in the search for the causal variant perturbing these networks of genes.

Two closely related terms, *animal organ development* and *animal organ morphogenesis,* were significantly enriched with genes associated with anterior pituitary and adrenal variants, and the latter term was enriched in genes associated with high/moderate variants in the hypothalamus at just under significance (FDR = 0.051). These two terms shares 26 HMI genes,

including 7 variants in 6 genes with $F_{ST} > 0.2$. Of these genes, 5 are associated with regulation of the cytoskeleton and the formation of cilia, lamellipodia, or membrane ruffles: *ALMS1* (F_{ST} = 0.55, Ala/Pro missense) (Graser et al., 2007); *BASP1* (F_{ST} = 0.33, Ala/Val missense) (Korshunova et al., 2008); *MKKS/BBS6* (two variants: F_{ST} = 0.26, Gly/Ser missense; F_{ST} = 0.25, Ala/Thr missense) (Seo et al., 2010); *PDGFRB* (F_{ST} = 0.31, Pro/Leu missense) (Farooqi et al., 2011); and *PHACTR4* (F_{ST} = 0.25, Met/Leu missense) (Allen et al., 2004; Zhang et al., 2012). *PHACTR4* is also involved in the migration of enteric neural crest cells (Zhang et al., 2012), an intriguing finding due to the hypothesis that changes in neural crest cell migration might be associated with the proposed domestication syndrome (Wilkins et al., 2014). Notably, two of these variants, in *BASP1* and *PDGFRB,* fall in the putative sweep region 27; another, *ALMS1,* falls into region 50. The categorization of these genes into terms related to organ development suggests developmental differences in HPA axis-associated organs in the tame as compared to the aggressive lines. Indeed, it has been observed that much of the difference between the tame and aggressive phenotypes is apparent at a young age (Plyusnina et al., 1991), at a time when the HPA axis is still developing. Moreover, the identification of variants in genes associated with cytoskeletal morphology and membrane protrusions is notable, as this echoes findings of the importance of pseudopodia and cellular motility in both the anterior pituitary and adrenals gene expression studies (Chapters 3 and 4). These genes may represent possible candidates for genomic variation affecting the expression differences seen in those studies.

Five GO terms related to ion transport were enriched for genes associated with HMI variants in both hypothalamus and adrenals, but not in pituitary; 28 variants in 19 genes were associated with all five of these GO terms. Four of those variants had $F_{ST} > 0.2$: *SLC5A6* (F_{ST} = 0.58, Val/Ile missense), a sodium-dependent multivitamin transporter (Prasad et al., 1998) that may be important in brain development (Subramanian et al., 2017); *GRIK5* (F_{ST} = 0.33, Val/Ile missense), a receptor for glutamate, an excitatory neurotransmitter; *IBTK* (F_{ST} = 0.33, Gln/Leu missense), an inhibitor of calcium mobilization (Liu et al, 2001); and *ITPR2* (F_{ST} = 0.25, Ser/Asn missense), a mediator of intracellular calcium release (Wiel et al., 2014). Polymorphisms in genes related to ion transport are suggestive of changes in excitatory tissue, with possible implications for action potential frequency. Hypothalamus contains neural tissue, and the adrenal medulla is formed from a modified post-synaptic ganglion. As the anterior pituitary contains significantly less excitatory tissue than these two organs, genes associated with ion channel

activity may simply be expressed less frequently in that tissue, making identification of variants in them more difficult. This finding suggests the importance of cellular excitivity in the regulation of HPA axis activity. Additionally, as previously mentioned, calcium regulation may have implications for exocytosis, a critical process in endocrine organs such as the hypothalamus, pituitary, and adrenals.

Terms related to DNA damage repair were enriched for genes associated with HMI variants in the anterior pituitary only. One gene in these terms, *FANCL,* contained two missense variants (Ile/Phe and Ala/Val), both with $F_{ST} = 0.58$, and was significantly up-regulated in the aggressive fox anterior pituitary. This gene codes for a protein functioning as a key step in the Fanconi anemia pathway, which functions in DNA repair (Machida et al., 2006). This finding is intriguing in light of the GO term *positive regulation of response to DNA damage stimulus* being significantly enriched in genes that are differentially expressed in both the adrenal and pituitary glands of tame and aggressive foxes, as reported in Chapter 4. The finding of HMI polymorphisms associated with this term suggests that indeed DNA damage repair may have been unknowingly selected for during the fox domestication process; the gene with two high F_{ST} missense variants, *FANCL,* is a candidate for a causal variant in this process.

The similarities in results from GO analysis of different tissues may suggest that these tissues have similar functions and therefore express similar proteins, a possibility in light of the fact that all three tissues are involved in hormone release. If true, this would suggest that the variants discovered in the genes associated with these GO groups could potentially affect the functionality of proteins specific to endocrine glands or to the HPA axis. Alternatively, the similarities may be due to the expression of genes that are common to many different cell types, and their associated variants may reflect a background of genomic differences between the tame and aggressive fox that may not in fact be associated with the known differences in their HPA axis activity.

High impact variants were fewer in number than moderate impact variants, and none of them were considered high F_{ST}, suggesting that none of them have significantly different allele frequency variation in the two populations. It is therefore unlikely that any of the high impact variants identified in this study were under specific selection in the development of the tame phenotype. However, one splice region variant in *RAB1A* had an F_{ST} of 0.29, one of the highest

F_{ST}s of any high impact variant. (No genes with high impact variants and higher F_{ST} had well described functions.) The *RAB1A* gene is notable because it is associated with cell migration/adhesion and also with vesicular protein transport (Wang et al., 2010; Mukhopadhyay et al., 2011). These processes were implicated in both the anterior pituitary and the adrenal gland gene expression analyses in differences in HPA axis regulation between the tame and aggressive lines. Therefore, a causative variant differentially regulating these processes may exist, and *RAB1A* appears to be a potential candidate.

Genes containing HMI variants with $F_{ST} > 0.5$ (in effect, all of these were "moderate" rather than "high" impact variants) were compared to differentially expressed genes in those tissues. Because differential expression has not been evaluated in hypothalamus, this tissue was exempted from this analysis. Variants associated with genes that had functions relevant to HPA axis regulation were prioritized for investigation. One variant, a Leu/Phe missense in *PHTF1,* was notable for its high F_{ST} (0.81) and its location within a region of putative selective sweep. *PHTF1* is significantly upregulated in the adrenals of tame foxes. This gene has been characterized as a putative homeodomain transcription factor (Manuel et al., 2000), but subsequent work suggested it may actually be an integral membrane protein associated with vesicles in male germ cells as they undergo transformation (Oyhenart et al., 2002). Its function in HPA axis tissues is unknown, but if similar to that in testis, may be related to vesicle regulation, a process implicated in gene expression differences in both pituitary and adrenal analyses.

The other high F_{ST} (0.58), HMI (Cys/Tyr missense) variant associated with a region of putative selective sweeps is in the *MRPL35* gene, which is significantly up-regulated in the aggressive fox anterior pituitary and adrenals. This gene codes for a mitochondrial ribosomal protein, and is associated with the GO terms relating to translation in mitochondria (significantly enriched in hypothalamus and anterior pituitary). Another high F_{ST} (0.66), HMI (Thr/Asn missense) gene, *PTCD3,* codes for a mitochondrial RNA-binding protein (Castello et al., 2012). Differences in mitochondrial activity appear to be important in the differences between the tame and aggressive HPA axis; 7 ribosomal subunit genes are differentially expressed in the fox anterior pituitary, and 4 in the fox adrenal gland, all up-regulated in the tame line in both tissues (Chapters 3 and 4). Differences in activity between the tame and aggressive HPA axis may result in different energy requirements, necessitating different levels of mitochondrial activity in these

tissues. Mitochondrial activity has not been investigated previously in these foxes and may prove a fruitful future avenue of research.

FLNC, or filamin C, contained a Ser/Leu missense variant with $F_{ST} = 0.64$. This gene is differentially expressed (up-regulated in aggressive) in the adrenals. Filamins cross-link to actin and play a role in cytoskeleton regulation (Feng and Walsh, 2004). This finding is particularly interesting in light of the difference in expression of genes associated with regulation of cell morphology in the anterior pituitary and adrenals of tame and aggressive foxes, and may point to a genomic polymorphism influencing those differences.

The gene *SNX19* (sorting nexin 19) also contained a high F_{ST} (FDR = 0.64) and HMI (Asp150Asn) variant, and was differentially expressed in both anterior pituitary and adrenal samples (up-regulated in aggressive foxes). This gene is associated with intracellular vesicle trafficking and exocytosis (Harashima et al., 2012). Similarly, a Gln/His missense variant was found in *EXOC6* ($F_{ST} = 0.5$), a gene associated with exocytic vesicles and cell migration (Liu and Guo, 2012). This gene is significantly up-regulated in the aggressive fox adrenal gland. Differences in gene expression between tame and aggressive anterior pitutiaries and adrenals included other genes associated with vesicle formation and exocytosis; the *SNX19* and *EXOC6* variants may indicate the presence of genomic polymorphisms affecting these processes.

5.5. CONCLUSIONS

Many of the GO term themes identified in this study, such as cellular signaling and DNA damage repair, echo themes from the gene expression analyses in anterior pituitary and adrenal tissue. These similarities suggest that the polymorphisms described here may include variants causal for some of the gene expression differences seen in those studies. Comparing results from different analyses – F_{ST}, variant impact annotation, and differential expression – has provided multiple lines of support for the importance of specific variants in the regulation of the tame and aggressive HPA axis. Overall, identification of variants in hypothalamus, anterior pituitary, and adrenal tissues from tame and aggressive foxes has provided an array of new candidate genes for future investigation.

CHAPTER 6. CONCLUSIONS

6.1 DISCUSSION

As taught in undergraduate endocrinology classes, the HPA axis consists of a cascade of hormones. The steps between the binding of a molecule of hormone to a receptor at one level of that cascade, and the release of a different hormone in response, are left to the imagination of the student. One might expect, therefore, that regulatory differences in the HPA axis of tame and aggressive foxes would involve differences in hormone receptors or in the synthesis of the hormones themselves. The findings of this series of gene expression studies and variant analysis, however, demonstrate that changes in receptors or synthesis of hormones specific to the cascade may not be the only or even most important regulatory mechanisms of the HPA axis.

Instead, these results indicated differences in cell migration and adhesion, possibly related to communication between cells to coordinate hormone release; in cell signals associated with triggering exocytosis; and in vesicular transport, possibly of hormones to the cell membrane for release. In the adrenals, differences between fox lines were additionally associated with processes peripheral to the HPA cascade: metabolism of precursors to all steroid hormones, not just the ones involved in the HPA axis, such as fatty acids and cholesterol; and angiogenesis, perhaps resulting in different levels of perfusion between the lines to carry hormones out to nearby blood vessels during excytosis. Of particular interest is the finding of differences in variants and expression in genes relating to DNA damage repair, a process that has been implicated in genomic studies as well. This intriguing finding has implications beyond the regulation of the HPA axis.

Variant analysis similarly suggested differences between the lines in all three tissues in exocytosis, cell migration, and cell adhesion; the demonstration of comparable results from a different approach provided support for the validity of the findings. In addition, variant analysis identified a set of polymorphisms with potential functional effects and genetic differences between the lines. These variants may provide future candidates in the search for polymorphisms

contributing to the phenotypic differences between tame and aggressive foxes. They will be interpreted in light of ongoing work in Kukekova Lab identifying regions of putative selective sweeps in the tame and aggressive fox genomes, work that has already identified one gene contributing to the tame phenotype.

This fox transcriptome study implicates novel mechanisms in the biological differences between the tame and aggressive fox HPA axis. This unexpected perspective on the importance of cell signaling to HPA axis regulation in foxes will provide new avenues for studies in other models. Specifically, experimental fox domestication has been proposed as a model for domestication of the dog and other species, and the findings from this project may prove applicable to HPA axis regulation differences between other domesticated animals and their wild ancestors. Additionally, HPA axis regulation varies in the normal population, possibly with effects on behavioral traits such as coping styles; and its dysregulation is implicated in a number of psychiatric disorders in humans, as well as potentially in behavioral disorders in other species, such as dogs. Therefore, the new directions suggested by these studies – investigation of cellular signaling and hormone release, rather than hormone receptor function – may provide new avenues of research in a variety of questions and a variety of species.

6.2. FUTURE DIRECTIONS

Future work will be necessary to confirm these transcriptome findings and to understand how the differences identified here manifest in tame and aggressive HPA axis tissues at the cellular level. For example, pseudopodia formation was implicated in differences between tame and aggressive anterior pituitary gene expression, but this finding raises new questions: are pseudopodia formed more or less frequently in one line than the other? In which cell types? If regulation of pseudopodia differs between lines, what is the functional significance of this difference?

Plans for future work include histology studies using immunofluorescent tagging to identify the relative locations of proteins, such as ACTH and growth factors associated with vasculature, in anterior pituitaries of tame and aggressive foxes. This work will provide information about the relative number of anterior pituitary corticotrophs and differences in vascularization in the two

lines. Longer term, functional studies will be necessary to assess the impact of these differences in cell behavior, such as pseudopodial and network formation, migration, and cellular adhesion.

Additionally, this project has identified new candidate variants for investigations into genomic underpinnings of the biological differences in the tame and aggressive phenotypes. The existence of extensive samples from F2 and backcross populations will provide critical resources in the investigation of the potential behavioral effects of these variants.

Finally, exploration of these findings in other models is critical for understanding their importance. The changes identified here in the experimentally domesticated fox may be unique to that species, or may be indicative of a more general domestication syndrome. They may also be relevant to differences between humans with different coping styles or to humans with dysregulated HPA axes subsequent to (or causative of) psychiatric disorders. A logical next step is investigation of the closely related domestic dog, to see if similar changes in cell signaling are reflected in the anterior pituitary and/or adrenals between members of that species with different HPA axis activity. If the suite of changes identified here prove to be duplicated in different species, they may indicate an evolutionary strategy much more widely applicable than only to the domesticated fox.

REFERENCES

Abaco, Giovanna MD, et al. "ADAM22, Expressed in Normal Brain but not in High-Grade Gliomas, Inhibits Cellular Proliferation via the Disintegrin Domain." *Neurosurgery*-Hagerstown 58.1 (2006): 179-186.

Abdel-Malek, Z. A. "Melanocortin receptors: their functions and regulation by physiological agonists and antagonists." Cellular and Molecular Life Sciences 58.3 (2001): 434-441.

Acher, R. "Molecular evolution of biologically active polypeptides." *Proceedings of the Royal Society of London B: Biological Sciences* 210.1178 (1980): 21-43.

Acland, Gregory M., et al. "Linkage analysis and comparative mapping of canine progressive rod–cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans." *Proceedings of the National Academy of Sciences* 95.6 (1998): 3048-3053.

Adamo, Hugo P., et al. "Mutants in the Putative Nucleotide-binding Region of the Plasma Membrane Ca2+-Pump: a reduction in activity due to slow dephosphorylation." *Journal of Biological Chemistry* 270.50 (1995): 30111-30114.

Adamovich, Yaarit, et al. "The protein level of PGC-1α, a key metabolic regulator, is controlled by NADH-NQO1." *Molecular and Cellular Biology* 33.13 (2013): 2603-2613.

Adams, Josephine C. "Thrombospondin-1." *The International Journal of Biochemistry & Cell Biology* 29.6 (1997): 861-865.

Adams, Josephine C. "Thrombospondins: multifunctional regulators of cell interactions." *Annual Review of Cell and Developmental Biology* 17.1 (2001): 25-51.

Adams, Mark D., et al. "The genome sequence of Drosophila melanogaster." *Science* 287.5461 (2000): 2185-2195.

Adelstein, Robert S., and Evan Eisenberg. "Regulation and kinetics of the actin-myosin-ATP interaction." *Annual Review of Biochemistry* 49.1 (1980): 921-956.

Adzhubei, Ivan A., et al. "A method and server for predicting damaging missense mutations." *Nature Methods* 7.4 (2010): 248-249.

Agarwal, Neeraj, et al. “MTBP suppresses cell migration and filopodia formation by inhibiting ACTN4.” *Oncogene* 32.4 (2013): 462-470.

Akey J, Ruhe A, Akey D et al. Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences*. 2010;107(3):1160-1165. doi:10.1073/pnas.0909918107.

Albert, Frank W., et al. “Phenotypic differences in behavior, physiology and neurochemistry between rats selected for tameness and for defensive aggression towards humans.” *Hormones and Behavior* 53.3 (2008): 413-421.

Alkan, Can, Saba Sajjadian, and Evan E. Eichler. "Limitations of next-generation genome sequence assembly." *Nature Methods* 8.1 (2011): 61-65.

Allache, Redouane, et al. “Genetic studies of ANKRD6 as a molecular switch between Wnt signaling pathways in human neural tube defects.” *Birth Defects Research Part A: Clinical and Molecular Teratology* 103.1 (2015): 20-26.

Allen, Patrick B., et al. “Phactrs 1–4: A family of protein phosphatase 1 and actin regulatory proteins.” *Proceedings of the National Academy of Sciences of the United States of America* 101.18 (2004): 7187-7192.

Altschul, Stephen F., et al. "Basic local alignment search tool." *Journal of Molecular Biology* 215.3 (1990): 403-410.

Amat, Ramon, et al. “SIRT1 is involved in glucocorticoid-mediated control of uncoupling protein-3 gene transcription.” *Journal of Biological Chemistry* 282.47 (2007): 34066-34076.

Anand-Srivastava, Madhu B. “Natriuretic peptide receptor-C signaling and regulation.” *Peptides* 26.6 (2005): 1044-1059.

Anand, Nisha, et al. “Protein elongation factor EEF1A2 is a putative oncogene in ovarian cancer.” *Nature Genetics* 31.3 (2002): 301-305.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. “HTSeq—a Python framework to work with high-throughput sequencing data.” *Bioinformatics* 31.2 (2015): 166-169.

Andersson L. “Molecular consequences of animal breeding.” *Current Opinion in Genetics & Development.* 2013;23(3):295-301

Andersson, Leif, and Michel Georges. "Domestic-animal genomics: deciphering the genetics of complex traits." *Nature Reviews Genetics* 5.3 (2004): 202-212.

Andrews, Simon. “FastQC: a quality-control tool for high-throughput sequence data.” http://www. Bioinformaticsbabraham.ac.uk/projects/fastqc (2014).

Antunica-Noguerol, M., et al. “The activity of the glucocorticoid receptor is regulated by SUMO conjugation to FKBP51.” *Cell Death & Differentiation* (2016).

Archinti, Marco, et al. “SPICE–a previously uncharacterized protein required for centriole duplication and mitotic chromosome congression.” *Journal of Cell Science* 123.18 (2010): 3039-3046.

Arendt, Maja L., et al. "Genome-wide association study of golden retrievers identifies germ-line risk factors predisposing to mast cell tumours." *PLoS Genetics* 11.11 (2015): e1005647.

Ashwell, M. S., et al. "Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle." *Journal of Dairy Science* 87.2 (2004): 468-475.

Atsushikasamatsu, Katsuhiro Uzawa, and Hirofumi Koike Dai Nakashima. “Galectin-9 as a regulator of cellular adhesion in human oral squamous cell carcinoma cell lines.” *International Journal of Molecular Medicine 16* (2005): 269-273.

Auer, Paul L., and R. W. Doerge. "Statistical design and analysis of RNA sequencing data." *Genetics* 185.2 (2010): 405-416.

Axelsson, Erik, et al. “The genomic signature of dog domestication reveals adaptation to a starch-rich diet.” *Nature* 495.7441 (2013): 360-364.

Bai, Siau Wei, et al. “Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration.” *BMC Biology* 9.1 (2011): 54.

Balakrishnan C, Mukai M, Gonser R et al. Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. *PeerJ*. 2014;2:e396. doi:10.7717/peerj.396.

Baltz, Alexander G., et al. “The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.” *Molecular Cell* 46.5 (2012): 674-690.

Bänziger, Carla, et al. “Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells.” *Cell* 125.3 (2006): 509-522.

Bao, Jianjun, et al. “R1 motif is the major actin-binding domain of TRIOBP-4.” *Biochemistry* 52.31 (2013): 5256-5264.

Baugh, Alexander T., et al. "Initial reactivity and magnitude of the acute stress response associated with personality in wild great tits (Parus major)." *General and Comparative Endocrinology* 189 (2013): 96-104.

Beever, J. E., et al. "A single‐base change in the tyrosine kinase II domain of ovine FGFR3 causes hereditary chondrodysplasia in sheep." *Animal Genetics* 37.1 (2006): 66-71.

Bell, Scott E., et al. “Differential gene expression during capillary morphogenesis in 3D collagen matrices.” *Journal of Cell Science* 114.15 (2001): 2755-2773.

Benjannet, S., et al. “PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues.” *Proceedings of the National Academy of Sciences* 88.9 (1991): 3564-3568.

Bernfield, Merton, et al. “Functions of cell surface heparan sulfate proteoglycans.” *Annual Review of Biochemistry* 68.1 (1999): 729-777.

Besco, Julie A., et al. “Intracellular substrates of brain-enriched receptor protein tyrosine phosphatase rho (RPTPρ/PTPRT).” *Brain Research* 1116.1 (2006): 50-57.

Bi, Shuguang, et al. “Galectin-9 binding to cell surface protein disulfide isomerase regulates the redox environment to enhance T-cell migration and HIV entry.” *Proceedings of the National Academy of Sciences* 108.26 (2011): 10650-10655.

Bindea, Gabriela, et al. “ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.” *Bioinformatics* 25.8 (2009): 1091-1093.

Bisgrove, Brent W., et al. “RFX2 is essential in the ciliated organ of asymmetry and an RFX2 transgene identifies a population of ciliated cells sufficient for fluid flow.” *Developmental Biology* 363.1 (2012): 166-178.

Block, Jennifer, et al. “FMNL2 drives actin-based protrusion and migration downstream of Cdc42.” *Current Biology* 22.11 (2012): 1005-1012.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* (2014): btu170.

Bonfiglio, Juan José, et al. “The corticotropin-releasing hormone network and the hypothalamic-pituitary-adrenal axis: molecular and cellular mechanisms involved.” *Neuroendocrinology* 94.1 (2011): 12-20.

Bornstein, S. R., and G. P. Chrousos. "Adrenocorticotropin (ACTH)-and non-ACTH-mediated regulation of the adrenal cortex: neural and immune inputs." *The Journal of Clinical Endocrinology & Metabolism* 84.5 (1999): 1729-1736.

Bottomly, Daniel, et al. "Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays." *PloS One* 6.3 (2011): e17820.

Boudarene, M., J. J. Legros, and M. Timsit-Berthier. “Study of the stress response: role of anxiety, cortisol and DHEAs.” *L'Encephale* 28.2 (2001): 139-146.

Bovine HapMap Consortium. "Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds." *Science* 324.5926 (2009): 528-532.

Boyko, Adam R., et al. "A simple genetic architecture underlies morphological variation in dogs." *Plos Biology* 8.8 (2010): e1000451.

Boyle, Louise H., et al. “Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway.” *Proceedings of the National Academy of Sciences* 110.9 (2013): 3465-3470.

Brady, Mark, Nikolina Vlatković, and Mark T. Boyd. “Regulation of p53 and MDM2 activity by MTBP.” *Molecular and Cellular Biology* 25.2 (2005): 545-553.

Brafman, D. A., et al. “Regulation of endodermal differentiation of human embryonic stem cells through integrin-ECM interactions.” *Cell Death & Differentiation* 20.3 (2013): 369-381.

Brewster, Ursula C., and Mark A. Perazella. “The renin-angiotensin-aldosterone system and the kidney: effects on kidney disease.” *The American Journal of Medicine* 116.4 (2004): 263-272.

Briassoulis, George, et al. “The glucocorticoid receptor and its expression in the anterior pituitary and the adrenal cortex: a source of variation in hypothalamic-pituitary-adrenal axis function; Implications for pituitary and andrenal tumors.” *Endocrine Practice* 17.6 (2011): 941-948.

Brophy, K., et al. “Synaptosomal-associated protein 25 (SNAP-25) and attention deficit hyperactivity disorder (ADHD): evidence of linkage and association in the Irish population.” *Molecular Psychiatry* 7.8 (2002): 913.

Brose, Nils, et al. "Synaptotagmin: a calcium sensor on the synaptic vesicle surface." *Science* 256.5059 (1992): 1021-1025.

Bross, Peter, et al. “Characterization of wild-type and an active site mutant of human medium chain acyl-CoA dehydrogenase after expression in Escherichia coli.” *Journal of Biological Chemistry* 265.13 (1990): 7116-7119.

Brown, C. Titus, et al. (2013): khmer-protocols 0.8.4 documentation. khmer-protocols 0.8.4 documentation. 2013. Available at: http://dx.doi.org/10.6084/m9.figshare.878460. Accessed August 26, 2015

Budry, Lionel, et al. “The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling.” *Genes & Development* 26.20 (2012): 2299-2310.

Burgoyne, Robert D., and Alan Morgan. “Secretory granule exocytosis.” *Physiological Reviews* 83.2 (2003): 581-632.

Burke, Heather M., et al. “Depression and cortisol responses to psychological stress: a meta-analysis.” *Psychoneuroendocrinology* 30.9 (2005): 846-856.

Bylund, Johan, et al. “Identification of CYP4F8 in human seminal vesicles as a prominent 19-hydroxylase of prostaglandin endoperoxides.” *Journal of Biological Chemistry* 275.29 (2000): 21844-21849.

Cadieu, Edouard, et al. "Coat variation in the domestic dog is governed by variants in three genes." *Science* 326.5949 (2009): 150-153.

Calakos, Nicole, et al. “Protein-protein interactions contributing to the specificity of intracellular vesicular trafficking.” *Science* 18 (1994): 19.

Campler, Magnus, Markus Jöngren, and Per Jensen. “Fearfulness in red junglefowl and domesticated White Leghorn chickens.” *Behavioural Processes* 81.1 (2009): 39-43.

Cao, Peng, Anton Maximov, and Thomas C. Südhof. “Activity-dependent IGF-1 exocytosis is controlled by the Ca 2+-sensor synaptotagmin-10.” *Cell* 145.2 (2011): 300-311.

Caohuy, Hung, Meera Srivastava, and Harvey B. Pollard. “Membrane fusion protein synexin (annexin VII) as a Ca2+/GTP sensor in exocytotic secretion.” *Proceedings of the National Academy of Sciences* 93.20 (1996): 10797-10802.

Careau, V., et al. “Energy metabolism and animal personality.” *Oikos* 117.5 (2008): 641-653.

Castello, Alfredo, et al. “Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.” *Cell* 149.6 (2012): 1393-1406.

Chain, P. S. G., et al. "Genome project standards in a new era of sequencing." *Science* 326.5950 (2009): 236-237.

Chang, Christopher C., et al. “Second-generation PLINK: rising to the challenge of larger and richer datasets.” *Gigascience* 4.1 (2015): 7.

Chang, Su-Ping, et al. “Cell proliferation, movement and differentiation during maintenance of the adult mouse adrenal cortex.” *PloS One* 8.12 (2013): e81865.

Charlier, Carole, et al. "The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status." *Nature Genetics* 27.4 (2001): 367-369.

Chase, Kevin, et al. "Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton." *Proceedings of the National Academy of Sciences* 99.15 (2002): 9930-9935.

Chen, Yandong, Zheng Wang, and Lijun Zhou. “Interleukin 8 Inhibition Enhanced Cholesterol Efflux in Acetylated Low-Density Lipoprotein–Stimulated THP-1 Macrophages.” *Journal of Investigative Medicine* 62.3 (2014): 615-620.

Chen, Yunshun, Davis McCarthy, Mark Robinson, Gordon K. Smyth. “edgeR: differential expression analysis of digital gene expression data: user's guide.” *Bioconductor.org*. 2015. Available at:

https://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf. Accessed October 12, 2015.

Cheng, Albert W., et al. "Muscleblind-like 1 (Mbnl1) regulates pre-mRNA alternative splicing during terminal erythropoiesis." *Blood* 124.4 (2014): 598-610.

Chepelev, Iouri, et al. "Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq." *Nucleic Acids Research* 37.16 (2009): e106-e106.

Chesné, Julie, et al. "IL-17 in severe asthma. Where do we stand?" *American Journal of Respiratory and Critical Care Medicine* 190.10 (2014): 1094-1101.

Chesnokova, Vera, and Shlomo Melmed. "Minireview: neuro-immuno-endocrine modulation of the hypothalamic-pituitary-adrenal (HPA) axis by gp130 signaling molecules." *Endocrinology* 143.5 (2002): 1571-1574.

Chidlow, John H., et al. "VEGF164 differentially regulates neutrophil and T cell adhesion through ItgaL-and ItgaM-dependent mechanisms." *American Journal of Physiology-Gastrointestinal and Liver Physiology* 299.6 (2010): G1361-G1367.

Chinwalla, Asif T., et al. "Initial sequencing and comparative analysis of the mouse genome." *Nature* 420.6915 (2002): 520-562.

Chodniewicz, David, and Richard L. Klemke. "Guiding cell migration through directed extension and stabilization of pseudopodia." *Experimental Cell Research* 301.1 (2004): 31-37.

Chrétien, Michel, and Majambu Mbikay. "60 Years of POMC: From the prohormone theory to pro-opiomelanocortin and to proprotein convertases (PCSK1 to PCSK9)." *Journal of Molecular Endocrinology* 56.4 (2016): T49-T62.

Ciccotosto, Giuseppe D., et al. "Induction of integral membrane PAM expression in AtT-20 cells alters the storage and trafficking of POMC and PC1." *The Journal of Cell Biology* 144.3 (1999): 459-471.

Cingolani, Pablo, et al. "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3." *Fly* 6.2 (2012): 80-92.

Clark, Michael B., et al. "The reality of pervasive transcription." *PLoS Biology* 9.7 (2011): e1000625.

Clary, Dawson, et al. "Shyness–Boldness, but not Exploration, Predicts Glucocorticoid Stress Response in Richardson's Ground Squirrels (Urocitellus richardsonii)." *Ethology* 120.11 (2014): 1101-1109

Cock, Peter JA, et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics." *Bioinformatics* 25.11 (2009): 1422-1423.

Colonna, Cecilia, and Ernesto J. Podestá. "ACTH-induced caveolin-1 tyrosine phosphorylation is related to podosome assembly in Y1 adrenal cells." *Experimental Cell Research* 304.2 (2005): 432-442.

Conklin, James L. "The development of the human fetal adenohypophysis." *The Anatomical Record* 160.1 (1968): 79-91.

Cool, David R., and Y. Peng Loh. "Carboxypeptidase E is a sorting receptor for prohormones: binding and kinetic studies." *Molecular and Cellular Endocrinology* 139.1 (1998): 7-13.

Creyghton, Menno P., et al. "PR72, a novel regulator of Wnt signaling required for Naked cuticle function." *Genes & Development* 19.3 (2005): 376-386.

Cua, Daniel J., and Cristina M. Tato. "Innate IL-17-producing cells: the sentinels of the immune system." *Nature Reviews Immunology* 10.7 (2010): 479-489.

Cuajungco, Math P., et al. "Abnormal accumulation of human transmembrane (TMEM)-176A and 176B proteins is associated with cancer pathology." *Acta Histochemica* 114.7 (2012): 705-712.

Dalloul, Rami A., et al. "Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis." *PLoS Biology* 8.9 (2010): e1000475.

Dammeyer, Pascal, et al. "Induction of cell membrane protrusions by the N-terminal glutaredoxin domain of a rare splice variant of human thioredoxin reductase 1." *Journal of Biological Chemistry* 283.5 (2008): 2814-2821.

Danecek, Petr, et al. "The variant call format and VCFtools." *Bioinformatics* 27.15 (2011): 2156-2158.

Davy, Alice, and Stephen M. Robbins. “Ephrin‐A5 modulates cell adhesion and morphology in an integrin‐dependent manner.” *The EMBO Journal* 19.20 (2000): 5396-5405.

Dawson, D. W., et al. “Pigment epithelium-derived factor: a potent inhibitor of angiogenesis.” *Science* 285.5425 (1999): 245-248.

De Stefani, Diego, and Rosario Rizzuto. “Molecular control of mitochondrial calcium uptake.” *Biochemical and Biophysical Research Communications* 449.4 (2014): 373-376.

del Valle-Pérez, Beatriz, et al. “Filamin B plays a key role in vascular endothelial growth factor-induced endothelial cell motility through its interaction with Rac-1 and VAV-2.” *Journal of Biological Chemistry* 285.14 (2010): 10748-10760.

Denef, C. “Paracrinicity: the story of 30 years of cellular pituitary crosstalk.” *Journal of Neuroendocrinology* 20.1 (2008): 1-70.

Dickerson, Sally S., and Margaret E. Kemeny. “Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research.” *Psychological Bulletin* 130.3 (2004): 355.

Diebold, Marie-Laure, et al. “Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II.” *Journal of Biological Chemistry* 285.49 (2010): 38389-38398.

Doane, Leah D., et al. "Are flatter diurnal cortisol rhythms associated with major depression and anxiety disorders in late adolescence? The role of life stress and daily negative emotion." *Development and Psychopathology* 25.03 (2013): 629-642.

Dobin, Alexander, et al. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29.1 (2013): 15-21.

Dodge, Kimberly L., et al. “mAKAP assembles a protein kinase A/PDE4 phosphodiesterase cAMP signaling module.” *The EMBO Journal* 20.8 (2001): 1921-1930.

Dodt, Matthias, et al. "FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms." *Biology* 1.3 (2012): 895-905.

Dorrell, Michael I., Edith Aguilar, and Martin Friedlander. “Retinal vascular development is mediated by endothelial filopodia, a preexisting astrocytic template and specific R-cadherin adhesion.” *Investigative Ophthalmology & Visual Science* 43.11 (2002): 3500-3510.

Driscoll, Carlos A., David W. Macdonald, and Stephen J. O'Brien. “From wild animals to domestic pets, an evolutionary view of domestication.” *Proceedings of the National Academy of Sciences* 106.Supplement 1 (2009): 9971-9978.

Drouin, Jacques. “60 years of POMC: Transcriptional and epigenetic regulation of POMC gene expression.” *Journal of Molecular Endocrinology* 56.4 (2016): T99-T112.

Du, Lianming, et al. "First insights into the giant panda (Ailuropoda melanoleuca) blood transcriptome: a resource for novel gene loci and immunogenetics." *Molecular Ecology Resources* 15.4 (2015): 1001-1013.

Duffy, David J., et al. "GSK3 inhibitors regulate MYCN mRNA levels and reduce neuroblastoma cell viability through multiple mechanisms, including p53 and Wnt signaling." *Molecular Cancer Therapeutics* 13.2 (2014): 454-467.

Ehrhart‐Bornstein, Monika, and Stefan R. Bornstein. “Cross‐talk between Adrenal Medulla and Adrenal Cortex in Stress.” *Annals of the New York Academy of Sciences* 1148.1 (2008): 112-117.

Ekblom, Robert, et al. "Comparison between normalised and unnormalised 454-sequencing libraries for small-scale RNA-Seq studies." *Comparative and Functional Genomics* 2012 (2012).

Ellis, I., J. Banyard, and S. L. Schor. “Differential response of fetal and adult fibroblasts to cytokines: cell migration and hyaluronan synthesis.” *Development* 124.8 (1997): 1593-1600.

Endris, Volker, et al. “SrGAP3 interacts with lamellipodin at the cell membrane and regulates Rac-dependent cellular protrusions.” *Journal of Cell Science* 124.23 (2011): 3941-3955.

Eppler, Elisabeth, et al. “Insulin‐like growth factor I (IGF‐I) and its receptor (IGF‐1R) in the rat anterior pituitary.” *European Journal of Neuroscience* 25.1 (2007): 191-200.

Erez, Neta, et al. “Expression of prolyl-hydroxylase-1 (PHD1/EGLN2) suppresses hypoxia inducible factor-1α activation and inhibits tumor growth.” *Cancer Research* 63.24 (2003): 8777-8783.

Ericsson, Maria, et al. "Domestication effects on behavioural and hormonal responses to acute stress in chickens." *Physiology & Behavior* 133 (2014): 161-169.

Ernst, Stefanie, et al. “An annexin 1 N-terminal peptide activates leukocytes by triggering different members of the formyl peptide receptor family.” *The Journal of Immunology* 172.12 (2004): 7669-7676.

Everts, Robin E., et al. "A 7872 cDNA microarray and its use in bovine functional genomics." *Veterinary Immunology and Immunopathology* 105.3 (2005): 235-245.

Fallahsharoudi, Amir, et al. “Domestication effects on stress induced steroid secretion and adrenal gene expression in chickens.” *Scientific Reports* 5 (2015).

Farooqi, Ammad Ahmad, et al. “PDGF: the nuts and bolts of signalling toolbox.” *Tumor Biology* 32.6 (2011): 1057.

Feng, Yuanyi, and Christopher A. Walsh. “The many faces of filamin: a versatile molecular scaffold for cell motility and signalling.” *Nature Cell Biology* 6.11 (2004): 1034-1038.

Ferguson, Laura, et al. "Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus." *Molecular Ecology* 19.s1 (2010): 240-254.

Ferrante, Maria Immacolata, et al. “Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification.” *Nature Genetics* 38.1 (2006): 112-117.

Fischer, Daniel, et al. "Identification of reproduction-related gene polymorphisms using whole transcriptome sequencing in the large white pig population." *G3: Genes| Genomes| Genetics* 5.7 (2015): 1351-1360.

Florio, Tullio. “Adult pituitary stem cells: from pituitary plasticity to adenoma development.” *Neuroendocrinology* 94.4 (2011): 265-277.

Fonseca, Nuno A., John Marioni, and Alvis Brazma. "RNA-seq gene profiling-a systematic empirical comparison." *PloS One* 9.9 (2014): e107026.

Fontrodona, Laura, et al. “RSR-2, the Caenorhabditis elegans ortholog of human spliceosomal component SRm300/SRRM2, regulates development by influencing the transcriptional machinery.” *PLoS Genetics* 9.6 (2013): e1003543.

Frazee, Alyssa C., et al. "Differential expression analysis of RNA-seq data at single-base resolution." *Biostatistics* (2014): kxt053.

Fuller-Pace, Frances V., and Simak Ali. “The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators.” *Biochemical Society Transactions* (2008): 609-612.

Gaffen, Sarah L. “Structure and signalling in the IL-17 receptor family.” *Nature Reviews Immunology* 9.8 (2009): 556-567.

Gambarotta, Giovanna, et al. “ErbB4 expression in neural progenitor cells (ST14A) is necessary to mediate neuregulin-1β1-induced migration.” *Journal of Biological Chemistry* 279.47 (2004): 48808-48816.

Garber, Manuel, et al. "Computational methods for transcriptome annotation and quantification using RNA-seq." *Nature Methods* 8.6 (2011): 469-477.

Gass, Peter, et al. "Mice with targeted mutations of glucocorticoid and mineralocorticoid receptors: models for depression and anxiety?." *Physiology & Behavior* 73.5 (2001): 811-825.

Gautier, Emmanuel L., et al. "Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages." *Nature Immunology* 13.11 (2012): 1118-1128.

George, Nysia I., et al. "An iterative leave-one-out approach to outlier detection in RNA-Seq data." *PloS One* 10.6 (2015): e0125224.

Gibson, Mark A., David I. Leavesley, and Leonie K. Ashman. “Microfibril-associated glycoprotein-2 specifically interacts with a range of bovine and human cell types via αVβ3 integrin.” *Journal of Biological Chemistry* 274.19 (1999): 13060-13065.

Gingras, Sébastien, et al. “Characterization of a family of novel cysteine-serine-rich nuclear proteins (CSRNP).” *PLoS One* 2.8 (2007): e808.

Glinka, Tal, et al. “Signal-peptide-mediated translocation is regulated by a p97–AIRAPL complex.” *Biochemical Journal* 457.2 (2014): 253-261.

Golbert, Daiane Cristina Ferreira, et al. “ITGA6 gene silencing by RNA interference modulates the expression of a large number of cell migration-related genes in human thymic epithelial cells.” *BMC Genomics* 14.6 (2013): S3.

Goldstein, David B. “Islands of linkage disequilibrium.” *Nature Genetics* 29.2 (2001): 109-112.

Gordon A, Hannon G. "Fastx-toolkit." *FASTQ/A short-reads preprocessing tools (unpublished)*. 2010. Available at: http://hannonlab. cshl. edu/fastx_toolkit. Accessed December 12, 2015.

Gorrigan, Rebecca J., et al. “Localisation of the melanocortin-2-receptor and its accessory proteins in the developing and adult adrenal gland.” *Journal of Molecular Endocrinology* 46.3 (2011): 227-232.

Grabherr, Manfred G., et al. "Full-length transcriptome assembly from RNA-Seq data without a reference genome." *Nature Biotechnology* 29.7 (2011): 644-652.

Grammatopoulos, Dimitris K. “Insights into mechanisms of corticotropin-releasing hormone receptor signal transduction.” *British Journal of Pharmacology* 166.1 (2012): 85-97.

Graser, Susanne, et al. “Cep164, a novel centriole appendage protein required for primary cilium formation.” *Journal of Cell Biology* 179.2 (2007): 321-330.

Greaves-Lord, Kirstin, et al. "Higher cortisol awakening response in young adolescents with persistent anxiety problems." *Acta Psychiatrica Scandinavica* 116.2 (2007): 137-144.

Gregg, Christopher, et al. "Sex-specific parent-of-origin allelic expression in the mouse brain." *Science* 329.5992 (2010): 682-685.

Griffin, James E., and Sergio R. Ojeda. *Textbook of Endocrine Physiology,* 5th ed. Oxford University Press, 2004.

Groenen, Martien AM, et al. "Analyses of pig genomes provide insight into porcine demography and evolution." *Nature* 491.7424 (2012): 393-398.

Gu, Jianguo, et al. “Laminin-10/11 and fibronectin differentially regulate integrin-dependent Rho and Rac activation via p130Cas-CrkII-DOCK180 pathway.” *Journal of Biological Chemistry* 276.29 (2001): 27090-27097.

Gu, Jianguo, et al. “Potential roles of N-glycosylation in cell adhesion.” *Glycoconjugate Journal* 29.8-9 (2012): 599-607.

Guerrier, Sabrice, et al. “The F-BAR domain of srGAP2 induces membrane protrusions required for neuronal migration and morphogenesis.” *Cell* 138.5 (2009): 990-1004.

Gulevich, R. G., et al. “Effect of selection for behavior on pituitary–adrenal axis and proopiomelanocortin gene expression in silver foxes (Vulpes vulpes).” *Physiology & Behavior* 82.2 (2004): 513-518.

Guo, Jian-Ming, et al. “Molecular cloning and characterization of a novel member of the UDP‐GalNAc: polypeptide N‐acetylgalactosaminyltransferase family, pp‐GalNAc‐T12 1.” *FEBS Letters* 524.1-3 (2002): 211-218.

Guo, Shaoshi, and Shilai Bao. “srGAP2 arginine methylation regulates cell migration and cell spreading through promoting dimerization.” *Journal of Biological Chemistry* 285.45 (2010): 35133-35141.

Guo, Zhenhuan, et al. “E-cadherin interactome complexity and robustness resolved by quantitative proteomics.” *Science Signaling* 7.354 (2014): rs7-rs7.

Haas, Brian J., et al. "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." *Nature Protocols* 8.8 (2013): 1494-1512.

Hahn, Daniel A., et al. "Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly Sarcophaga crassipalpis." *BMC Genomics* 10.1 (2009): 234.

Hall, Brian K. “The neural crest and neural crest cells: discovery and significance for theories of embryonic organization.” *Journal of Biosciences* 33.5 (2008): 781-793.

Han, Seungrie, et al. “Altered expression of synaptotagmin 13 mRNA in adult mouse brain after contextual fear conditioning.” *Biochemical and Biophysical Research Communications* 425.4 (2012): 880-885.

Hardeveld, Florian, et al. “Increased cortisol awakening response was associated with time to recurrence of major depressive disorder.” *Psychoneuroendocrinology* 50 (2014): 62-71.

Hare, Brian, Victoria Wobber, and Richard Wrangham. “The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression.” *Animal Behaviour* 83.3 (2012): 573-585.

Harrow, Jennifer, et al. "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome Research* 22.9 (2012): 1760-1774.

Hatzirodos, Nicholas, et al. "Transcriptome profiling of granulosa cells of bovine ovarian follicles during growth from small to large antral sizes." *BMC Genomics* 15.1 (2014): 24.

Hendy, Geoffrey N., et al. “Targeted ablation of the chromogranin a (Chga) gene: normal neuroendocrine dense-core secretory granules and increased expression of other granins.” *Molecular Endocrinology* 20.8 (2006): 1935-1947.

Herman, James P., et al. "Central mechanisms of stress integration: hierarchical circuitry controlling hypothalamo–pituitary–adrenocortical responsiveness." *Frontiers in Neuroendocrinology* 24.3 (2003): 151-180.

Hillier, LaDeana W., et al. "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution." *Nature* 432.7018 (2004): 695-716.

Hirsch, Dianne Snow, Dana M. Pirone, and Peter D. Burbelo. “A new family of Cdc42 effector proteins, CEPs, function in fibroblast and epithelial cell shape changes.” *Journal of Biological Chemistry* 276.2 (2001): 875-883.

Hiscott, John, et al. “Triggering the interferon response: the role of IRF-3 transcription factor.” *Journal of Interferon & Cytokine research* 19.1 (1999): 1-13.

Hodson, David J., et al. “Coordination of calcium signals by pituitary endocrine cells in situ.” *Cell Calcium* 51.3 (2012a): 222-230.

Hodson, David J., et al. “Existence of long-lasting experience-dependent plasticity in endocrine cell networks.” *Nature Communications* 3 (2012b): 605.

Hoeppner, Marc P., et al. "An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts." *PloS One* 9.3 (2014): e91172.

Houslay, Miles D., and David R. Adams. “PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization.” *Biochemical Journal* 370.1 (2003): 1-18.

Hsu, Daniel K., et al. "Endogenous galectin-3 is localized in membrane lipid rafts and regulates migration of dendritic cells." *Journal of Investigative Dermatology* 129.3 (2009): 573-583.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." *Nature Protocols* 4.1 (2009): 44-57.

Huang, Norman, Parantu K. Shah, and Cheng Li. "Lessons from a decade of integrating cancer copy number alterations with gene expression profiles." *Briefings in Bioinformatics* 13.3 (2011): 305-316.

Ince-Dunn, Gulayse, et al. “Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability.” *Neuron* 75.6 (2012): 1067-1080.

Inoue, Takashi, et al. “Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives.” *Developmental Biology* 270.1 (2004): 146-162.

Iorizzo, Massimo, et al. "De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity." *BMC Genomics* 12.1 (2011): 389.

Ito, Daisuke, Aric M. Frantz, and Jaime F. Modiano. "Canine lymphoma as a comparative model for human non-Hodgkin lymphoma: recent progress and applications." *Veterinary Immunology and Immunopathology* 159.3 (2014): 192-201.

Ivanov, Danila, et al. “Cell adhesion molecule T-cadherin regulates vascular cell adhesion, phenotype and motility.” *Experimental Cell Research* 293.2 (2004): 207-218.

Jackson, Keith R., et al. "Performance analysis of high performance computing applications on the amazon web services cloud." *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*. IEEE, 2010.

Jackson, Verity A., et al. "Super-complexes of adhesion GPCRs and neural guidance receptors." *Nature Communications* 7 (2016).

Jacquemet, Guillaume, Hellyeh Hamidi, and Johanna Ivaska. "Filopodia in cell adhesion, 3D migration and cancer cell invasion." *Current Opinion in Cell Biology* 36 (2015): 23-31.

Jain, Ruchi, Bappaditya Dey, and Anil K. Tyagi. "Development of the first oligonucleotide microarray for global gene expression profiling in guinea pigs: defining the transcription signature of infectious diseases." *BMC Genomics* 13.1 (2012): 520.

Jakharia, Aniruddha, Biswajyoti Borkakoty, and Sujay Singh. "Expression of SPARC like protein 1 (SPARCL1), extracellular matrix-associated protein is down regulated in gastric adenocarcinoma." *Journal of Gastrointestinal Oncology* 7.2 (2016): 278.

Jarjour, Andrew A., et al. "Autocrine netrin function inhibits glioma cell motility and promotes focal adhesion formation." *PloS One* 6.9 (2011): e25408.

Jeffries, Marlo K. Sellin, et al. "A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples." *BMC Biotechnology* 14.1 (2014): 94.

Johnson, Jennifer L., et al. "Genotyping-by-sequencing (GBS) detects genetic structure and confirms behavioral QTL in tame and aggressive foxes (Vulpes vulpes)." *PloS ONE* 10.6 (2015): e0127013.

Jolly, R. D., and H. W. Leipold. "Inherited diseases of cattle—a perspective." *New Zealand Veterinary Journal* 21.7 (1973): 147-155.

Jones, Paul, et al. "Single-nucleotide-polymorphism-based association mapping of dog stereotypes." *Genetics* 179.2 (2008): 1033-1044.

Juliano, Celina, Jianquan Wang, and Haifan Lin. "Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms." *Annual Review of Genetics* 45 (2011): 447-469.

Jurica, Melissa S., et al. “Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis.” *RNA* 8.4 (2002): 426-439.

Kahn, Laura H. “Confronting zoonoses, linking human and veterinary medicine.” *Emerging Infectious Diseases* 12.4 (2006): 556-561.

Kaiser, Sylvia, Michael B. Hennessy, and Norbert Sachser. “Domestication affects the structure, development and stability of biobehavioural profiles.” *Frontiers in Zoology* 12.1 (2015): 1.

Kalashnikova, Evgenia, et al. “SynDIG1: an activity-regulated, AMPA-receptor-interacting transmembrane protein that regulates excitatory synapse development.” *Neuron* 65.1 (2010): 80-93.

Kanibolotsky, Dmitry S., et al. “Multiple molecular dynamics simulation of the isoforms of human translation elongation factor 1A reveals reversible fluctuations between 'open' and 'closed' conformations and suggests specific for eEF1A1 affinity for Ca 2+-calmodulin.” *BMC Structural Biology* 8.1 (2008): 4.

Kariagina, Anastasia, et al. “Hypothalamic-pituitary cytokine network.” *Endocrinology* 145.1 (2004): 104-112.

Karlsson E, Sigurdsson S, Ivansson E et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biology*. 2013;14(12):R132. doi:10.1186/gb-2013-14-12-r132.

Karlsson, Elinor K., et al. "Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B." *Genome Biology* 14.12 (2013): R132.

Keck, Martin E., et al. "Combined effects of exonic polymorphisms in CRHR1 and AVPR1B genes in a case/control study for panic disorder." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 147.7 (2008): 1196-1204.

Keller-Wood, Maureen E., and Mary F. Dallman. "Corticosteroid Inhibition of ACTH Secretion." *Endocrine Reviews* 5.1 (1984): 1-24.

Kent, W. James. "BLAT—the BLAST-like alignment tool." *Genome Research* 12.4 (2002): 656-664.

Kessler, Efrat, A. Paul Mould, and David JS Hulmes. “Procollagen type I C-proteinase enhancer is a naturally occurring connective tissue glycoprotein.” *Biochemical and Biophysical Research Communications* 173.1 (1990): 81-86.

Khanna, Chand, et al. "The dog as a cancer model." *Nature Biotechnology* 24.9 (2006): 1065-1067.

Kilpatrick, Daniel L., et al. “Stability of bovine adrenal medulla cells in culture.” *Journal of Neurochemistry* 35.3 (1980): 679-692.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013;14(4):R36. doi:10.1186/gb-2013-14-4-r36.

Kim, Daehwan, et al. "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." *Genome Biology* 14.4 (2013): R36.

Kim, Hyejeong Rosemary, et al. “Gli2a protein localization reveals a role for Iguana/DZIP1 in primary ciliogenesis and a dependence of Hedgehog signal transduction on primary cilia in the zebrafish.” *BMC Biology* 8.1 (2010): 65.

Kim, Hyung-Goo, and Lawrence C. Layman. “The role of CHD7 and the newly identified WDR11 gene in patients with idiopathic hypogonadotropic hypogonadism and Kallmann syndrome.” *Molecular and Cellular Endocrinology* 346.1 (2011): 74-83.

Kim, Joon, et al. “Functional genomic screen for modulators of ciliogenesis and cilium length.” *Nature* 464.7291 (2010): 1048-1051.

Klopfleisch, Robert, et al. "Metastatic canine mammary carcinomas can be identified by a gene expression profile that partly overlaps with human breast cancer profiles." *BMC Cancer* 10.1 (2010): 618.

Kobelt, A. J., et al. “Sources of sampling variation in saliva cortisol in dogs.” *Research in Veterinary Science* 75.2 (2003): 157-161.

Koch, Alisa E., et al. “Interleukin-8 as a macrophage-derived mediator of angiogenesis.” *Science* 258 (1992): 1798-1798.

Koolhaas, J. M., et al. “Neuroendocrinology of coping styles: towards understanding the biology of individual variation.” *Frontiers in Neuroendocrinology* 31.3 (2010): 307-321.

Koolhaas, Jaap M., et al. "Individual variation in coping with stress: a multidimensional approach of ultimate and proximate mechanisms." *Brain, Behavior and Evolution* (2007).

Koressaar, Triinu, and Maido Remm. “Enhancements and modifications of primer design program Primer3.” *Bioinformatics* 23.10 (2007): 1289-1291.

Korshunova, Irina, et al. “Characterization of BASP1-mediated neurite outgrowth.” *Journal of Neuroscience Research* 86.10 (2008): 2201-2213.

Koschützke, Leif, et al. “SrGAP3 knockout mice display enlarged lateral ventricles and specific cilia disturbances of ependymal cells in the third ventricle.” *Cell and Tissue research* 361.2 (2015): 645-650.

Köttgen, Michael, et al. “TRPP2 and TRPV4 form a polymodal sensory channel complex.” *The Journal of Cell Biology* 182.3 (2008): 437-447.

Krause, Matthias, and Alexis Gautreau. “Steering cell migration: lamellipodium dynamics and the regulation of directional persistence.” *Nature Reviews Molecular Cell Biology* 15.9 (2014): 577-590.

Kritzer, Michael D., et al. "AKAPs: the architectural underpinnings of local cAMP signaling." *Journal of Molecular and Cellular Cardiology* 52.2 (2012): 351-358.

Kroboth, Patricia D., et al. “DHEA and DHEA‐S: a review.” *The Journal of Clinical Pharmacology* 39.4 (1999): 327-348.

Kubzansky, Laura D., and Gail K. Adler. “Aldosterone: A forgotten mediator of the relationship between psychological stress and heart disease.” *Neuroscience & Biobehavioral Reviews* 34.1 (2010): 80-86.

Kukekova, Anna V., et al. "A meiotic linkage map of the silver fox, aligned and compared to the canine genome." *Genome Research* 17.3 (2007): 387-399.

Kukekova, Anna V., et al. "Genetics of behavior in the silver fox." *Mammalian Genome* 23.1-2 (2012): 164-177.

Kukekova, Anna V., et al. "Mapping loci for fox domestication: deconstruction/reconstruction of a behavioral phenotype." *Behavior Genetics* 41.4 (2011b): 593-606.

Kukekova, Anna V., et al. "Sequence comparison of prefrontal cortical brain transcriptome from a tame and an aggressive silver fox (Vulpes vulpes)." *BMC Genomics* 12.1 (2011): 482.

Kukekova, Anna V., et al. "Sequence comparison of prefrontal cortical brain transcriptome from a tame and an aggressive silver fox (Vulpes vulpes)." *BMC Genomics* 12.1 (2011a): 482.

Kukekova, Anna V., et al. “Measurement of segregating behaviors in experimental silver fox pedigrees.” *Behavior Genetics* 38.2 (2008): 185-194.

Künzl, Christine, and Norbert Sachser. “The behavioral endocrinology of domestication: a comparison between the domestic guinea pig (*Cavia apereaf. porcellus*) and its wild ancestor, the cavy (*Cavia aperea*).” *Hormones and Behavior* 35.1 (1999): 28-37.

Künzl, Christine, et al. "Is a wild mammal kept and reared in captivity still a wild animal?." *Hormones and Behavior* 43.1 (2003): 187-196.

Kviklyte, Samanta, et al. “Effects of genetic deletion of soluble 59-nucleotidases NT5C1A and NT5C2 on AMPK activation and nucleotide levels in contracting mouse skeletal muscles.” *American Journal of Physiology-Endocrinology and Metabolism* (2017): ajpendo-00304.

Laaksonen, Reijo, et al. “Genetic variant of the SREBF-1 gene is significantly related to cholesterol synthesis in man.” *Atherosclerosis* 185.1 (2006): 206-209.

Lahens, Nicholas F., et al. "IVT-seq reveals extreme bias in RNA sequencing." *Genome Biology* 15.6 (2014): R86.

Lamouille, Samy, et al. “Activin receptor–like kinase 1 is implicated in the maturation phase of angiogenesis.” *Blood* 100.13 (2002): 4495-4501.

Lander, Eric S., et al. "Initial sequencing and analysis of the human genome." *Nature* 409.6822 (2001): 860-921.

Landgraf, Rainer. "HAB/LAB rats: an animal model of extremes in trait anxiety and depression." *Clinical Neuroscience Research* 3.4 (2003): 239-244.

Langfelder, Peter, and Steve Horvath. “WGCNA: an R package for weighted correlation network analysis.” *BMC Bioinformatics* 9.1 (2008): 1.

Langmead, Ben, and Steven L. Salzberg. “Fast gapped-read alignment with Bowtie 2.” *Nature Methods* 9.4 (2012): 357-359.

Lango Allen, Hana, et al. "Hundreds of variants clustered in genomic loci and biological pathways affect human height." *Nature* 467.7317 (2010): 832-8.

Lau, Nga-Chi, et al. “Human Ccr4–Not complexes contain variable deadenylase subunits.” *Biochemical Journal* 422.3 (2009): 443-453.

Laurent, Charles E., and Thomas E. Smithgall. “The c-Fes tyrosine kinase cooperates with the breakpoint cluster region protein (Bcr) to induce neurite extension in a Rac-and Cdc42-dependent manner.” *Experimental Cell Research* 299.1 (2004): 188-198.

Lennartsson, Anna-Karin, et al. “DHEA and DHEA-S response to acute psychosocial stress in healthy men and women.” *Biological Psychology* 90.2 (2012): 143-149.

Leonard, Brian E. "The HPA and immune axes in stress: the involvement of the serotonergic system." *European Psychiatry* 20 (2005): S302-S306.

Levin, Joshua Z., et al. "Comprehensive comparative analysis of strand-specific RNA sequencing methods." *Nature Methods* 7.9 (2010): 709-715.

Levin, Nancy, Mariann Blum, and James L. Roberts. “Modulation Of Basal And Corticotropin-Releasing Factor- Stimulated Proopiomelanocortin Gene Expression By Vasopressin In Rat Anterior Pituitary.” *Endocrinology* 125.6 (1989): 2957-2966.

Li, Heng, et al. “The sequence alignment/map format and SAMtools.” *Bioinformatics* 25.16 (2009): 2078-2079.

Li, Heng. "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data." *Bioinformatics* 27.21 (2011): 2987-2993.

Li, Melody MH, and Margaret R. MacDonald. “Polyamines: Small Molecules with a Big Role in Promoting Virus Infection.” *Cell Host & Microbe* 20.2 (2016): 123-124.

Liao, Wei, et al. "Opposing actions of IL-2 and IL-21 on Th9 differentiation correlate with their differential regulation of BCL6 expression." *Proceedings of the National Academy of Sciences* 111.9 (2014): 3508-3513.

Lichtenauer, Urs D., et al. "Pre-B-cell transcription factor 1 and steroidogenic factor 1 synergistically regulate adrenocortical growth and steroidogenesis." *Endocrinology* 148.2 (2007): 693-704.

Lin, Chunru, et al. "ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome." *Molecular and Cellular Biology* 25.17 (2005): 7484-7493.

Lindblad-Toh, Kerstin, et al. "Genome sequence, comparative analysis and haplotype structure of the domestic dog." *Nature* 438.7069 (2005): 803-819.

Liu, Dong, et al. "Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress." *Science* 277.5332 (1997): 1659-1662.

Liu, Gang, et al. "Interactions of elongation factor 1α with F-actin and β-actin mRNA: implications for anchoring mRNA in cell protrusions." *Molecular Biology of the Cell* 13.2 (2002): 579-592.

Liu, Jianglan, and Wei Guo. "The exocyst complex in exocytosis and cell migration." *Protoplasma* 249.3 (2012): 587-597.

Liu, Jun-Ping et al. "A comparative study of the role of adenylate cyclase in the release of adrenocorticotropin from the ovine and rat anterior pituitary." *Molecular and Cellular Endocrinology* 101.1-2 (1994): 173-181.

Liu, Tiangang, Harmit Vora, and Chaitan Khosla. "Quantitative analysis and engineering of fatty acid biosynthesis in E. coli." *Metabolic Engineering* 12.4 (2010): 378-386.

Liu, Weimin, et al. "Direct inhibition of Bruton's tyrosine kinase by IBtk, a Btk-binding protein." *Nature Immunology* 2.10 (2001): 939-946.

Liu, Xiaoping, et al. "Identifying the stress transcriptome in the adrenal medulla following acute and repeated immobilization." *Annals of the New York Academy of Sciences* 1148.1 (2008): 1-28.

Liu, Yichun, Hans-Ulrich Bernard, and Doris Apt. "NFI-B3, a novel transcriptional repressor of the nuclear factor I family, is generated by alternative RNA processing." *Journal of Biological Chemistry* 272.16 (1997): 10739-10745.

Liu, Yun-Sheng, et al. "The association of SNAP25 gene polymorphisms in attention deficit/hyperactivity disorder: a systematic review and meta-analysis." *Molecular Neurobiology* 54.3 (2017): 2189-2200.

Liu, Yuwen, Jie Zhou, and Kevin P. White. "RNA-seq differential expression studies: more sequence or more replication?." *Bioinformatics* 30.3 (2013): 301-304.

Lohse, Marc, et al. "R obi NA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics." *Nucleic Acids Research* 40.W1 (2012): W622-W627.

Longcope, Christopher. "1 Adrenal and gonadal androgen secretion in normal females." *Clinics in Endocrinology and Metabolism* 15.2 (1986): 213-228.

Love, Michael I., Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* 15.12 (2014): 550.

Lu, Shan, et al. "MTBP promotes the invasion and metastasis of hepatocellular carcinoma by enhancing the MDM2-mediated degradation of E-cadherin." *Digestive Diseases And Sciences* 60.12 (2015): 3681-3690.

Lunter, Gerton, and Martin Goodson. "Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads." *Genome Research* 21.6 (2011): 936-939.

Luo, Ruibang, et al. "SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler." *Gigascience* 1.1 (2012): 18.

Luo, Xuelian, et al. "Structure of the Mad2 spindle assembly checkpoint protein and its interaction with Cdc20." *Nature Structural & Molecular Biology* 7.3 (2000): 224-229.

Ma, Junwu, et al. "Genome-wide association study of meat quality traits in a White Duroc× Erhualian F2 intercross and Chinese Sutai pigs." *PloS One* 8.5 (2013): e64047.

Machida, Yuichi J., et al. "UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation." *Molecular Cell* 23.4 (2006): 589-596.

MacNair, Laura, et al. “MTHFSD and DDX58 are novel RNA-binding proteins abnormally regulated in amyotrophic lateral sclerosis.” *Brain* 139.1 (2016): 86-100.

Maher, Christopher A., et al. "Transcriptome sequencing to detect gene fusions in cancer." *Nature* 458.7234 (2009): 97-101.

Mahley, Robert W. “Apolipoprotein E: cholesterol transport protein with expanding role in cell biology.” *Science* 240.4852 (1988): 622.

Makvandi-Nejad, Shokouh, et al. "Four loci explain 83% of size variation in the horse." *PLoS One* 7.7 (2012): e39929.

Malik, Mobeen, et al. "Monocyte migration and LFA-1-mediated attachment to brain microvascular endothelia is regulated by SDF-1α through Lyn kinase.” *The Journal of Immunology* 181.7 (2008): 4632-4637.

Mandell, Kenneth J., et al. “Junctional adhesion molecule 1 regulates epithelial cell morphology through effects on β1 integrins and Rap1 activity.” *Journal of Biological Chemistry* 280.12 (2005): 11665-11674.

Manivannan, Shanthi, and Susumu Terakawa. “Rapid sprouting of filopodia in nerve terminals of chromaffin cells, PC12 cells, and dorsal root neurons induced by electrical stimulation.” *Journal of Neuroscience* 14.10 (1994): 5917-5928.

Mantella, Rose C., et al. "Salivary cortisol is associated with diagnosis and severity of late-life generalized anxiety disorder." *Psychoneuroendocrinology* 33.6 (2008): 773-781.

Marfell, Brooke J., Rory O’Brien, and J. Frank T. Griffin. "Global gene expression profiling of monocyte-derived macrophages from red deer (Cervus elaphus) genotypically resistant or susceptible to Mycobacterium avium subspecies paratuberculosis infection." *Developmental & Comparative Immunology* 40.2 (2013): 210-217.

Matsunaga, Kohichi, et al. “Rab2a and Rab27a cooperatively regulate the transition from granule maturation to exocytosis through the dual effector Noc2.” *Journal of Cell Science* 130.3 (2017): 541-550.

Matulova, Marta, et al. "Characterization of chicken spleen transcriptome after infection with Salmonella enterica serovar Enteritidis." *PLoS One* 7.10 (2012): e48101.

Maynard Smith, John, and John Haigh. “The hitch-hiking effect of a favourable gene.” *Genetical Research* 23.01 (1974): 23-35.

McCahill, Angela, et al. “In resting COS1 cells a dominant negative approach shows that specific, anchored PDE4 cAMP phosphodiesterase isoforms gate the activation, by basal cyclic AMP production, of AKAP-tethered protein kinase A type II located in the centrosomal region.” *Cellular Signalling* 17.9 (2005): 1158-1173.

McKenna, Aaron, et al. “The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.” *Genome Research* 20.9 (2010): 1297-1303.

McLaughlin, Fiona, et al. “Combined genomic and antisense analysis reveals that the transcription factor Erg is implicated in endothelial cell differentiation.” *Blood* 98.12 (2001): 3332-3339.

Meewisse, Marie-Louise, et al. “Cortisol and post-traumatic stress disorder in adults Systematic review and meta-analysis.” *The British Journal of Psychiatry* 191.5 (2007): 387-392.

Mehul, Bruno, and R. Colin Hughes. “Plasma membrane targetting, vesicular budding and release of galectin 3 from the cytoplasm of mammalian cells during secretion.” *Journal of Cell Science* 110.10 (1997): 1169-1178.

Mercurio, Arthur M., Isaac Rabinovitz, and Leslie M. Shaw. “The α6β4 integrin and epithelial cell migration.” *Current Opinion in Cell Biology* 13.5 (2001): 541-545.

Meyer, Eli, et al. "Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx." *BMC Genomics* 10.1 (2009): 219.

Meyers, Stacey N., et al. "A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle." *BMC Genomics* 11.1 (2010): 337.

Mill, Jonathan, et al. “Association study of a SNAP‐25 microsatellite and attention deficit hyperactivity disorder.” *American Journal of Medical Genetics* 114.3 (2002): 269-271.

Mitani, Fumiko, et al. “The undifferentiated cell zone is a stem cell zone in adult rat adrenal cortex.” *Biochimica et Biophysica Acta (BBA)-General Subjects* 1619.3 (2003): 317-324.

Monyer, Hannah, et al. “Developmental and regional expression in the rat brain and functional properties of four NMDA receptors.” *Neuron* 12.3 (1994): 529-540.

Morgan, Charles A., et al. “Relationships Among plasma dehydroepiandrosterone sulfate and cortisol levels, symptoms of dissociation, and objective performance in humans exposed to acute stress.” *Archives of General Psychiatry* 61.8 (2004): 819-825.

Morimoto, Masafumi, et al. “Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat brain: an immunohistochemical and in situ hybridization study.” *Neuroscience Research* 26.3 (1996): 235-269.

Mosher, Deane F., and Josephine C. Adams. “Adhesion-modulating/matricellular ECM protein families: a structural, functional and evolutionary appraisal.” *Matrix Biology* 31.3 (2012): 155-161.

Motazacker, Mahdi M., et al. “A loss-of-function variant in OSBPL1A predisposes to low plasma HDL cholesterol levels and impaired cholesterol efflux capacity.” *Atherosclerosis* 249 (2016): 140-147.

Motta, Pietro, Masaki Muto, and Tsuneo Fujita. “Three dimensional organization of mammalian adrenal cortex.” *Cell and Tissue Research* 196.1 (1979): 23-38.

Moura, E., et al. “α2‐Adrenoceptor subtypes involved in the regulation of catecholamine release from the adrenal medulla of mice.” *British Journal of Pharmacology* 149.8 (2006): 1049-1058.

Movilla, Nieves, and Xosé R. Bustelo. “Biological and regulatory properties of VAV-3, a new member of the VAV family of oncoproteins.” *Molecular and Cellular Biology* 19.11 (1999): 7870-7885.

Mukhopadhyay, Aparna, et al. "Proteomic analysis of endocytic vesicles: Rab1a regulates motility of early endocytic vesicles." *Journal of Cell Science* 124.5 (2011): 765-775.

Muráni, Eduard, et al. "Differential mRNA expression of genes in the porcine adrenal gland associated with psychosocial stress." *Journal of Molecular Endocrinology* 46.3 (2011): 165-174.

Nakamura, Kazuaki, et al. “Effects of vasopressin V1b receptor deficiency on adrenocorticotropin release from anterior pituitary cells in response to oxytocin stimulation.” *Endocrinology* 149.10 (2008): 4883-4891.

Nakata, Katsunori, et al. “A novel member of the Xenopus Zic family, Zic5, mediates neural crest development.” *Mechanisms of Development* 99.1 (2000): 83-91.

Nakata, Yuji, et al. “Integrin inhibition through Lyn-dependent cross talk from CXCR4 chemokine receptors in normal human CD34+ marrow cells.” *Blood* 107.11 (2006): 4234-4239.

Naumenko, E. V., et al. “Behavior, adrenocortical activity, and brain monoamines in Norway rats selected for reduced aggressiveness towards man.” *Pharmacology Biochemistry and Behavior* 33.1 (1989): 85-91.

Navratil, Amy M., et al. “Neuroendocrine plasticity in the anterior pituitary: gonadotropin-releasing hormone-mediated movement in vitro and in vivo.” *Endocrinology* 148.4 (2007): 1736-1744.

Nei, Masatoshi, and Ronald K. Chesser. “Estimation of fixation indices and gene diversities.” *Annals of Human Genetics* 47.3 (1983): 253-259.

Nelson, Ronald M., et al. “Genetics of interactive behavior in silver foxes (*Vulpes vulpes*).” *Behavior Genetics* 47.1 (2017): 88-101.

Neumann, Inga D. "Involvement of the brain oxytocin system in stress coping: interactions with the hypothalamo-pituitary-adrenal axis." *Progress in Brain Research* 139 (2002): 147-162.

Ng, M. Y. M., et al. "Meta-analysis of 32 genome-wide linkage studies of schizophrenia." *Molecular Psychiatry* 14.8 (2009): 774-785.

Nicholls, M. G., E. A. Espiner, and R. A. Donald. “Plasma aldosterone response to low dose ACTH stimulation.” *The Journal of Clinical Endocrinology & Metabolism* 41.1 (1975): 186-188.

Nie, Qinghua, et al. "Deep sequencing-based transcriptome analysis of chicken spleen in response to avian pathogenic Escherichia coli (APEC) infection." *PLoS One* 7.7 (2012): e41645.

Nielsen, Morten S., et al. “The sortilin cytoplasmic tail conveys Golgi–endosome transport and binds the VHS domain of the GGA2 sorting protein.” *The EMBO Journal* 20.9 (2001): 2180-2190.

Nishimura, Junji, et al. “cAMP induces up-regulation of ETA receptor mRNA and increases responsiveness to endothelin-1 of rat aortic smooth muscle cells in primary culture.” *Biochemical and Biophysical Research Communications* 188.2 (1992): 719-726.

Nobumoto, Atsuya, et al. “Galectin-9 suppresses tumor metastasis by blocking adhesion to endothelium and extracellular matrices.” *Glycobiology* 18.9 (2008): 735-744.

Norgren Jr, Robert B. "Improving genome assemblies and annotations for nonhuman primates." *ILAR Journal* 54.2 (2013): 144-153.

Nuber, U. A., et al. “Patterns of desmocollin synthesis in human epithelia: immunolocalization of desmocollins 1 and 3 in special epithelia and in cultured cells.” *European Journal of Cell Biology* 71.1 (1996): 1-13.

Ogi, Tomoo, and Alan R. Lehmann. “The Y-family DNA polymerase κ (pol κ) functions in mammalian nucleotide-excision repair.” *Nature Cell Biology* 8.6 (2006): 640-642.

Ogryzko, Vasily V., et al. “Histone-like TAFs within the PCAF histone acetylase complex.” *Cell* 94.1 (1998): 35-44.

Opii, Wycliffe O., et al. "Proteomic identification of brain proteins in the canine model of human aging following a long-term treatment with antioxidants and a program of behavioral enrichment: relevance to Alzheimer's disease." *Neurobiology of Aging* 29.1 (2008): 51-70.

Orr, H. Allen. “Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data.” *Genetics* 149.4 (1998): 2099-2104.

Oshiumi, Hiroyuki, et al. “TIR-containing adapter molecule (TICAM)-2, a bridging adapter recruiting to Toll-like receptor 4 TICAM-1 that induces interferon-β.” *Journal of Biological Chemistry* 278.50 (2003): 49751-49762.

Oskina, I. N. "Ontogenesis of endocrine function in silver foxes under domestication." *Applied Animal Behaviour Science* 44.2 (1995): 273-274.

Oskina, I. N., and I. Z. Plyusnina. “Pituitary-adrenal system under animals selection for domesticated behavior.” V.K. Shumny, A.L. Markel (Eds.), *Modern Concepts of Evolutionary Genetics*. Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk (2000), pp. 327–333 [In Russian]

Oskina, I. N., I. Z. Plyusnina, and A. Yu Sysoletina. “Effect of selection for behavior on the pituitary-adrenal function of norway rats (*Rattus norvegicus*) in postnatal ontogeny.” *Journal of Evolutionary Biochemistry and Physiology* 36.2 (2000): 161-169.

Ostrander, Elaine A., and Leonid Kruglyak. "Unleashing the canine genome." *Genome Research* 10.9 (2000): 1271-1274.

Oswald, Lynn M., et al. "Relationship between cortisol responses to stress and personality." *Neuropsychopharmacology* 31.7 (2006): 1583-1591.

Ouimet, Mireille, et al. “miRNA targeting of oxysterol-binding protein-like 6 regulates cholesterol trafficking and efflux.” *Arteriosclerosis, Thrombosis, and Vascular Biology* (2016): ATVBAHA-116.

Oyhenart, J., et al. “Phtf1 Is an Integral Membrane Protein Localized in an Endoplasmic Reticulum Domain in Maturing Male Germ Cells 1.” *Biology of Reproduction* 68.3 (2003): 1044-1053.

Panda, T., and V. Amutha Devi. “Regulation and degradation of HMGCo-A reductase.” *Applied Microbiology and Biotechnology* 66.2 (2004): 143-152.

Pandit, Jayvardhan, et al. “Crystal structure of human squalene synthase A key enzyme in cholesterol biosynthesis.” *Journal of Biological Chemistry* 275.39 (2000): 30610-30617.

Paoloni, Melissa, and Chand Khanna. "Translation of new cancer treatments from pet dogs to humans." *Nature Reviews Cancer* 8.2 (2008): 147-156.

Paratore, Christian, et al. “The role of the Ets domain transcription factor Erm in modulating differentiation of neural crest stem cells.” *Developmental Biology* 250.1 (2002): 168-180.

Pastor-Cantizano, Noelia, et al. “p24 family proteins: key players in the regulation of trafficking along the secretory pathway.” *Protoplasma* (2015): 1-19.

Patel, Paresh D., et al. “Glucocorticoid and mineralocorticoid receptor mRNA expression in squirrel monkey brain.” *Journal of Psychiatric Research* 34.6 (2000): 383-392.

Patterson, Donald F. "Companion animal medicine in the age of medical genetics." *Journal of Veterinary Internal Medicine* 14.1 (2000): 1-9.

Paushkin, Sergey V., et al. “Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3′ end formation.” *Cell* 117.3 (2004): 311-321.

Payne, Anita H., and Dale B. Hales. “Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones.” *Endocrine Reviews* 25.6 (2004): 947-970.

Penela, Petronila, Laura Nogués, and Federico Mayor. “Role of G protein-coupled receptor kinases in cell migration.” *Current Opinion in Cell Biology* 27 (2014): 10-17.

Pepke, Shirley, Barbara Wold, and Ali Mortazavi. "Computation for ChIP-seq and RNA-seq studies." *Nature Methods* 6 (2009): S22-S32.

Pérez-Montarelo, Dafne, et al. "Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis." *Physiological genomics* 46.6 (2014): 195-206.

Perkins, Timothy T., et al. "A strand-specific RNA–Seq analysis of the transcriptome of the typhoid bacillus salmonella typhi." *PLoS Genetics* 5.7 (2009): e1000569.

Pevsner, Jonathan, et al. “Specificity and regulation of a synaptic vesicle docking complex.” *Neuron* 13.2 (1994): 353-361.

Pickrell, Joseph K., et al. "Understanding mechanisms underlying human gene expression variation with RNA sequencing." *Nature* 464.7289 (2010): 768-772.

Pielberg, Gerli Rosengren, et al. "A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse." *Nature Genetics* 40.8 (2008): 1004-1009.

Piesse, Christophe, et al. “Human aminopeptidase B (rnpep) on chromosome 1q32. 2: complementary DNA, genomic structure and expression.” *Gene* 292.1 (2002): 129-140.

Pikuleva, Irina A. “Cytochrome P450s and cholesterol homeostasis.” *Pharmacology & Therapeutics* 112.3 (2006): 761-773.

Pirker, S., et al. “GABA A receptors: immunocytochemical distribution of 13 subunits in the adult rat brain.” *Neuroscience* 101.4 (2000): 815-850.

Plyusnina, I. Z., I. N. Oskina, and L. N. Trut. “An analysis of fear and aggression during early development of behaviour in silver foxes (Vulpes vulpes).” *Applied Animal Behaviour Science* 32.2-3 (1991): 253-268.

Pocha, Shirin M., and Denise J. Montell. “Cellular and molecular mechanisms of single and collective cell migrations in Drosophila: themes and variations.” *Annual Review of Genetics* 48 (2014): 295-318.

Pollard, Thomas D., and Gary G. Borisy. “Cellular motility driven by assembly and disassembly of actin filaments.” *Cell* 112.4 (2003): 453-465.

Prasad, Puttur D., et al. “Cloning and functional expression of a cDNA encoding a mammalian sodium-dependent vitamin transporter mediating the uptake of pantothenate, biotin, and lipoate.” *Journal of Biological Chemistry* 273.13 (1998): 7501-7506.

Price, E. O. *Principles and applications of domestic animal behavior: an introductory text.* C.A.B. International, Wallingford, Oxfordshire; Cambridge, MA, 2008.

Pruitt, Kim D., Tatiana Tatusova, and Donna R. Maglott. "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." *Nucleic Acids Research* 35.suppl_1 (2006): D61-D65.

Pulichino, Anne-Marie, et al. “Tpit determines alternate fates during pituitary cell differentiation.” *Genes & Development* 17.6 (2003): 738-747.

Purfield, Deirdre C., et al. "Runs of homozygosity and population history in cattle." *BMC Genetics* 13.1 (2012): 70.

Qi, Ai-Dong, et al. "Differential coupling of the human P2Y11 receptor to phospholipase C and adenylyl cyclase." *British Journal of Pharmacology* 132.1 (2001): 318-326.

Qiu, Qiang, et al. "The yak genome and adaptation to life at high altitude." *Nature Genetics* 44.8 (2012): 946-949.

Quinn, Andrew, Punita Juneja, and Francis M. Jiggins. "Estimates of allele-specific expression in Drosophila with a single genome sequence and RNA-seq data." *Bioinformatics* 30.18 (2014): 2603-2610.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for STATistical Computing, Vienna, Austria. https://www.R-project.org/.

Rainey, William E., and Yasuhiro Nakamura. "Regulation of the adrenal androgen biosynthesis." *The Journal of Steroid Biochemistry and Molecular Biology* 108.3 (2008): 281-286.

Rapaport, Franck, et al. "Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data." *Genome Biology* 14.9 (2013): 3158.

Reed, Charles C., and Renato V. Iozzo. "The role of decorin in collagen fibrillogenesis and skin homeostasis." *Glycoconjugate Journal* 19.4 (2002): 249-255.

Reis-Filho, Jorge S., and Lajos Pusztai. "Gene expression profiling in breast cancer: classification, prognostication, and prediction." *The Lancet* 378.9805 (2011): 1812-1823.

Reisine, Terry et al. "Corticotropin-Releasing Factor-Induced Adrenocorticotropin Hormone Release And Synthesis Is Blocked By Incorporation Of The Inhibitor Of Cyclic AMP-Dependent Protein Kinase Into Anterior Pituitary Tumor Cells By Liposomes." *Proceedings of the National Academy of Sciences* 82.23 (1985): 8261-8265.

Reisner, Ilana R. "Differential diagnosis and management of human-directed aggression in dogs." *Veterinary Clinics of North America: Small Animal Practice* 33.2 (2003): 303-320.

Renn, Susan CP, Nadia Aubin-Horth, and Hans A. Hofmann. "Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray." *BMC Genomics* 5.1 (2004): 42.

Ribon, Vered, et al. "A role for CAP, a novel, multifunctional Src homology 3 domain-containing protein in formation of actin stress fibers and focal adhesions." *Journal of Biological Chemistry* 273.7 (1998): 4073-4080.

Rickard, Amanda M., Lisa M. Petek, and Daniel G. Miller. "Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways." *Human Molecular Genetics* 24.20 (2015): 5901-5914.

Ridley, Anne J. “Membrane ruffling and signal transduction.” *Bioessays* 16.5 (1994): 321-327.

Rifas, Leonard, and M. Neale Weitzmann. “A novel T cell cytokine, secreted osteoclastogenic factor of activated T cells, induces osteoclast formation in a RANKL‐independent manner.” *Arthritis & Rheumatism* 60.11 (2009): 3324-3335.

Rikitake, Yoshiyuki, Kenji Mandai, and Yoshimi Takai. “The role of nectins in different types of cell–cell adhesion.” *Journal of Cell Science* 125.16 (2012): 3713-3722.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics* 26.1 (2010): 139-140.

Rogers, Rebekah L., et al. "Revised annotations, sex-biased expression, and lineage-specific genes in the Drosophila melanogaster group." *G3: Genes| Genomes| Genetics* 4.12 (2014): 2345-2351.

Rosmond, Roland, et al. "A polymorphism in the regulatory region of the corticotropin-releasing hormone gene in relation to cortisol secretion, obesity, and gene-gene interaction." *Metabolism* 50.9 (2001): 1059-1062.

Rosol, Thomas J., et al. “Adrenal gland: structure, function, and mechanisms of toxicity.” *Toxicologic Pathology* 29.1 (2001): 41-48.

Roy, Meenakshi, et al. "Analysis of the canine brain transcriptome with an emphasis on the hypothalamus and cerebral cortex." *Mammalian Genome* 24.11-12 (2013): 484-499.

Rubin, Carl-Johan, et al. “Whole-genome resequencing reveals loci under selection during chicken domestication.” *Nature* 464.7288 (2010): 587-591.

Rupprecht, Rainer, et al. “Pharmacological and functional characterization of human mineralocorticoid and glucocorticoid receptor ligands.” *European Journal of Pharmacology: Molecular Pharmacology* 247.2 (1993): 145-154.

Russo, Annapina, et al. “hnRNP H1 and intronic G runs in the splicing control of the human rpL3 gene.” *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1799.5 (2010): 419-428.

Sachdev, Pallavi, Liyu Zeng, and Lu-Hai Wang. “Distinct role of phosphatidylinositol 3-kinase and Rho family GTPases in VAV3-induced cell transformation, cell motility, and morphological changes.” *Journal of Biological Chemistry* 277.20 (2002): 17638-17648.

Saetre, Peter, et al. "From wild wolf to domestic dog: gene expression changes in the brain." *Molecular Brain Research* 126.2 (2004): 198-206.

Sakamoto, Kei, and Geoffrey D. Holman. “Emerging role for AS160/TBC1D4 and TBC1D1 in the regulation of GLUT4 traffic.” *American Journal of Physiology-Endocrinology and Metabolism* 295.1 (2008): E29-E37.

Sánchez-Solana, Beatriz, Da-Qiang Li, and Rakesh Kumar. “Cytosolic functions of MORC2 in lipogenesis and adipogenesis.” *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1843.2 (2014): 316-326.

Sano, Hideki, et al. “Human galectin-3 is a novel chemoattractant for monocytes and macrophages.” *The Journal of Immunology* 165.4 (2000): 2156-2164.

Sansone, Pasquale, and Jacqueline Bromberg. “Targeting the interleukin-6/Jak/stat pathway in human malignancies.” *Journal of Clinical Oncology* 30.9 (2012): 1005-1014.

Santos, António JM, et al. “TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum.” *Journal of Cell Biology* (2016): jcb-201603072.

Sapolsky, Robert M., and Michael J. Meaney. "Maturation of the adrenocortical stress response: neuroendocrine control mechanisms and the stress hyporesponsive period." *Brain Research Reviews* 11.1 (1986): 65-76.

Sapolsky, Robert M., L. Michael Romero, and Allan U. Munck. "How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions 1." *Endocrine Reviews* 21.1 (2000): 55-89.

Sapolsky, Robert M., Lewis C. Krey, and Bruce S. McEwen. "Glucocorticoid-sensitive hippocampal neurons are involved in terminating the adrenocortical stress response." *Proceedings of the National Academy of Sciences* 81.19 (1984): 6174-6177.

Sax, Joanna K., and Wafik S. El-Deiry. “Identification and characterization of the cytoplasmic protein TRAF4 as a p53-regulated proapoptotic gene.” *Journal of Biological Chemistry* 278.38 (2003): 36435-36444.

Schimmer, Bernard P., and Perrin C. White. “Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease.” *Molecular Endocrinology* 24.7 (2010): 1322-1337.

Schlessinger, Karni, Edward J. McManus, and Alan Hall. “Cdc42 and noncanonical Wnt signal transduction pathways cooperate to promote cell polarity.” *The Journal of Cell Biology* 178.3 (2007): 355-361.

Schoenebeck J, Ostrander E. Insights into morphology and disease from the dog genome project. *Annual Review of Cell and Developmental Biology*. 2014;30(1):535-560. doi:10.1146/annurev-cellbio-100913-012927.

Schulz M, Zerbino D, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012;28(8):1086-1092. doi:10.1093/bioinformatics/bts094.

Schwartz, Jeffrey. “Intercellular Communication in the Anterior Pituitary 1.” *Endocrine Reviews* 21.5 (2000): 488-513.

Schwarz D, Robertson H, Feder J et al. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot Rhagoletis pomonella. *BMC Genomics.* 2009;10(1):633. doi:10.1186/1471-2164-10-633.

Segovia, Mercedes, et al. “Autologous Dendritic Cells Prolong Allograft Survival Through TMEM176b‐Dependent Antigen Cross‐Presentation.” *American Journal of Transplantation* 14.5 (2014): 1021-1031.

Seifert, Jessica RK, and Marek Mlodzik. “Frizzled/PCP signalling: a conserved mechanism regulating cell polarity and directed motility.” *Nature Reviews Genetics* 8.2 (2007): 126-138.

Seino, Susumu, and Tadao Shibasaki. “PKA-dependent and PKA-independent pathways for cAMP-regulated exocytosis.” *Physiological Reviews* 85.4 (2005): 1303-1342.

Seo, Seongjin, et al. “BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly.” *Proceedings of the National Academy of Sciences* 107.4 (2010): 1488-1493.

Setoguchi, Kouji, et al. "Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate." *BMC genetics* 10.1 (2009): 43.

Shalek, Alex K., et al. "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." *Nature* 498.7453 (2013): 236-240.

Shannon, Paul, et al. “Cytoscape: a software environment for integrated models of biomolecular interaction networks.” *Genome Research* 13.11 (2003): 2498-2504.

Shaw, Leslie M., and Arthur M. Mercurio. “Regulation of cellular interactions with laminin by integrin cytoplasmic domains: the A and B structural variants of the alpha 6 beta 1 integrin differentially modulate the adhesive strength, morphology, and migration of macrophages.” *Molecular Biology of the Cell* 5.6 (1994): 679-690.

Shen, Shihao, et al. “rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data.” *Proceedings of the National Academy of Sciences* 111.51 (2014): E5593-E5601.

Shepherd, S. P., and M. A. Holzwarth. “Chromaffin-adrenocortical cell interactions: effects of chromaffin cell activation in adrenal cell cocultures.” *American Journal of Physiology-Cell Physiology* 280.1 (2001): C61-C71.

Shikhevich, S. G., I. N. Oskina, and I. Z. Pliusnina. “Effect of stress and immune stimulus on the pituitary-adrenal axis in gray rats selected for behavior.” *Rossiiskii fiziologicheskii zhurnal imeni IM Sechenova/Rossiiskaia akademiia nauk* 88.6 (2002): 781-789.

Shoal, Gavin D., Peter R. Giancola, and Galina P. Kirillova. "Salivary cortisol, personality, and aggressive behavior in adolescent boys: a 5-year longitudinal study." *Journal of the American Academy of Child & Adolescent Psychiatry* 42.9 (2003): 1101-1107.

Sindrilaru, Anca, et al. “Wound healing defect of VAV3−/− mice due to impaired β2-integrin–dependent macrophage phagocytosis of apoptotic neutrophils.” *Blood* 113.21 (2009): 5266-5276.

Sironen, Anu, et al. "An intronic insertion in KPL2 results in aberrant splicing and causes the immotile short-tail sperm defect in the pig." *Proceedings of the National Academy of Sciences of the United States of America* 103.13 (2006): 5006-5011.

Skalsky, Y. M., et al. “PRCC, the commonest TFE3 fusion partner in papillary renal carcinoma is associated with pre-mRNA splicing factors.” *Oncogene* 19.2 (2001): 178-187.

Smedley, Damian, et al. “The BioMart community portal: an innovative alternative to large, centralized data repositories.” *Nucleic Acids Research* 43.W1 (2015): W589-W598.

Smith, David M., et al. “Adrenomedullin: receptor and signal transduction.” *Biochemical Society Transactions* 30.4 (2002): 432-437.

Smoot, Michael E., et al. "Cytoscape 2.8: new features for data integration and network visualization." *Bioinformatics* 27.3 (2010): 431-432.

Soneson, Charlotte, and Mauro Delorenzi. "A comparison of methods for differential expression analysis of RNA-seq data." *BMC bioinformatics* 14.1 (2013): 91.

Soulavie, Fabien, et al. “hemingway is required for sperm flagella assembly and ciliary motility in Drosophila.” *Molecular Biology of the Cell* 25.8 (2014): 1276-1286.

Steed, Emily, Maria S. Balda, and Karl Matter. “Dynamics and functions of tight junctions.” *Trends in Cell Biology* 20.3 (2010): 142-149.

Stefaniuk, Monika, and Katarzyna Ropka-Molik. "RNA sequencing as a powerful tool in searching for genes influencing health and performance traits of horses." *Journal of applied genetics* 57.2 (2016): 199-206.

Stillman, Brianna N., et al. “Galectin-3 and galectin-1 bind distinct cell surface glycoprotein receptors to induce T cell death.” *The Journal of Immunology* 176.2 (2006): 778-789.

Stocco, Douglas M. “StAR protein and the regulation of steroid hormone biosynthesis.” *Annual Review of Physiology* 63.1 (2001): 193-213.

Subramaniam, S. Vaidyanathan, Roland S. Cooper, and Samuel E. Adunyah. “Evidence for the involvement of JAK/STAT pathway in the signaling mechanism of interleukin-17.” *Biochemical and Biophysical Research Communications* 262.1 (1999): 14-19.

Subramanian, Veedamali S., et al. “Mutations in SLC5A6 associated with brain, immune, bone, and intestinal dysfunction in a young child.” *Human Genetics* 136.2 (2017): 253-261.

Südhof, Thomas C. “Synaptotagmins: why so many?” *Journal of Biological Chemistry* 277.10 (2002): 7629-7632.

Sugarbaker, David J., et al. "Transcriptome sequencing of malignant pleural mesothelioma tumors." *Proceedings of the National Academy of Sciences* 105.9 (2008): 3521-3526.

Sugawara, Kenji, et al. “Structure and functional roles of Epac2 (Rapgef4).” *Gene* 575.2 (2016): 577-583.

Sutter, Nathan B., and Elaine A. Ostrander. "Dog star rising: the canine genetic system." *Nature Reviews Genetics* 5.12 (2004): 900-910.

Sutter, Nathan B., et al. “A single IGF1 allele is a major determinant of small size in dogs.” *Science* 316.5821 (2007): 112-115.

Sutton, Keith A., et al. “Enkurin is a novel calmodulin and TRPC channel binding protein in sperm.” *Developmental Biology* 274.2 (2004): 426-435.

Takahashi, Iwao, Kazuaki Ohashi, and Koji Nata. “Involvement of heparan sulfate 3‐O‐sulfotransferase isoform‐1 in the insulin secretion pathway.” *Journal of Diabetes Investigation* 3.4 (2012): 362-370.

Tang, Hua, et al. “Role of ornithine decarboxylase antizyme inhibitor in vivo.” *Genes to Cells* 14.1 (2009): 79-87.

Tarailo‐Graovac, Maja, and Nansheng Chen. "Using RepeatMasker to identify repetitive elements in genomic sequences." *Current Protocols in Bioinformatics* (2009): 4-10.

Tarazona, Sonia, et al. "Differential expression in RNA-seq: a matter of depth." *Genome research* 21.12 (2011): 2213-2223.

Taskén, Kristin A., et al. “Phosphodiesterase 4D and protein kinase a type II constitute a signaling unit in the centrosomal area.” *Journal of Biological Chemistry* 276.25 (2001): 21999-22002.

Taylor, S. Paige, et al. “Mutations in DYNC2LI1 disrupt cilia function and cause short rib polydactyly syndrome.” *Nature Communications* 6 (2015).

Teitz, Tal, et al. “Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN.” *Nature Medicine* 6.5 (2000): 529-535.

Terracciano, Antonio, et al. “Genome-wide association scan for five major dimensions of personality.” *Molecular Psychiatry* 15.6 (2010): 647-656.

Teufel, Michael, et al. “Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase.” *Journal of Biological Chemistry* 278.8 (2003): 6521-6531.

Tonomura, Noriko, et al. "Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers." *PLoS Genet* 11.2 (2015): e1004922.

Towns, John, et al. "XSEDE: accelerating scientific discovery." *Computing in Science & Engineering* 16.5 (2014): 62-74.

Trapnell, Cole, et al. “Differential analysis of gene regulation at transcript resolution with RNA-seq.” *Nature Biotechnology* 31.1 (2013): 46-53.

Trotta, Christopher R., et al. “The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases.” *Cell* 89.6 (1997): 849-858.

Trut, Lyudmila N., I. Z. Plyusnina, and I. N. Oskina. “An experiment on fox domestication and debatable issues of evolution of the dog.” *Russian Journal of Genetics* 40.6 (2004): 644-655.

Trut, Lyudmila, Irina Oskina, and Anastasiya Kharlamova. “Animal evolution during domestication: the domesticated fox as a model.” *Bioessays* 31.3 (2009): 349-360.

Trut, Lyudmila. “Early Canid Domestication: The Farm-Fox Experiment Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an

interplay between behavioral genetics and development." *American Scientist* 87.2 (1999): 160-169.

Tsai, Li-Chun Lisa, Masami Shimizu-Albergine, and Joseph A. Beavo. "The high-affinity cAMP-specific phosphodiesterase 8B controls steroidogenesis in the mouse adrenal gland." *Molecular Pharmacology* 79.4 (2011): 639-648.

Tu, Yizeng, et al. "Migfilin and Mig-2 link focal adhesions to filamin and the actin cytoskeleton and function in cell shape modulation." *Cell* 113.1 (2003): 37-47.

Turecki, Gustavo, and Michael Meaney. "Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review." *Biological Psychiatry* (2014).

Twal, Waleed O., et al. "Fibulin-1 suppression of fibronectin-regulated cell adhesion and motility." *Journal of Cell Science* 114.24 (2001): 4587-4598.

Untergasser, Andreas, et al. "Primer3 – new capabilities and interfaces." *Nucleic Acids Research* 40.15 (2012): e115-e115.

Van den Bout, Iman, and Nullin Divecha. "PIP5K-driven PtdIns (4, 5) P2 synthesis: regulation and cellular functions." *Journal of Cell Science* 122.21 (2009): 3837-3850.

Van Laere, Anne-Sophie, et al. "A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig." *Nature* 425.6960 (2003): 832-836.

Vaysse, Amaury, et al. "Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping." *PLoS Genetics* 7.10 (2011): e1002316.

Venhoranta, Heli, et al. "In frame exon skipping in UBE3B is associated with developmental disorders and increased mortality in cattle." *BMC genomics* 15.1 (2014): 890.

Verde, Ignacio, et al. "Myomegalin is a novel protein of the Golgi/centrosome that interacts with a cyclic nucleotide phosphodiesterase." *Journal of Biological Chemistry* 276.14 (2001): 11189-11198.

Vijay, Nagarjun, et al. "Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA‐seq experiments." *Molecular ecology* 22.3 (2013): 620-634.

Vincent, I. C., and A. R. Michell. "Comparison of cortisol concentrations in saliva and plasma of dogs." *Research in Veterinary Science* 53.3 (1992): 342-345.

Vinnedge, LM Privette, et al. "The human DEK oncogene stimulates β-catenin signaling, invasion and mammosphere formation in breast cancer." *Oncogene* 30.24 (2011): 2741-2752.

Voineagu, Irina, et al. "Transcriptomic analysis of autistic brain reveals convergent molecular pathology." *Nature* 474.7351 (2011): 380-384.

Vonholdt, Bridgett M., et al. "A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States." *Molecular Ecology* 19.20 (2010): 4412-4427.

Wada, Jun, et al. "Developmental regulation, expression, and apoptotic potential of galectin-9, a beta-galactoside binding lectin." *Journal of Clinical Investigation* 99.10 (1997): 2452.

Wagner, U., et al. "Promoter polymorphisms regulating corticotrophin-releasing hormone transcription in vitro." *Hormone and Metabolic Research* 38.2 (2006): 69-75.

Wahlsten, Douglas. "The hunt for gene effects pertinent to behavioral traits and psychiatric disorders: from mouse to human." *Developmental psychobiology* 54.5 (2012): 475-492.

Walczak, Marta, and Sascha Martens. "Dissecting the role of the Atg12–Atg5-Atg16 complex during autophagosome formation." *Autophagy* 9.3 (2013): 424-425.

Wang, Chenran, et al. "Regulation of integrin β1 recycling to lipid rafts by Rab1a to promote cell migration." *Journal of Biological Chemistry* 285.38 (2010): 29398-29405.

Wang, Daojing, Jing Huang, and Zhi Hu. "RNA helicase DDX5 regulates microRNA expression and contributes to cytoskeletal reorganization in basal breast cancer cells." *Molecular & Cellular Proteomics* 11.2 (2012): M111-011932.

Wang, Guo-dong, et al. "Signature of balancing selection at the MC1R gene in Kunming dog populations." *PloS one* 8.2 (2013): e55469.

Wang, Guo-Dong, et al. “Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau.” *Genome Biology and Evolution* 6.8 (2014): 2122-2128.

Wang, Haiping, et al. “Cloning and functional expression of ATA1, a subtype of amino acid transporter A, from human placenta.” *Biochemical and Biophysical Research Communications* 273.3 (2000): 1175-1179.

Wang, Hao, et al. "Rapid transcriptome and proteome profiling of a non‐model marine invertebrate, Bugula neritina." *Proteomics* 10.16 (2010): 2972-2981.

Wang, Li-na, et al. “Evaluation of adrenocorticotropin regulated glucocorticoid synthesis pathway in adrenal of different breeds of pigs.” *Livestock Science* 169 (2014): 185-191.

Wang, Xu, et al. "Paternally expressed genes predominate in the placenta." *Proceedings of the National Academy of Sciences* 110.26 (2013): 10705-10710.

Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics." *Nature reviews genetics* 10.1 (2009): 57-63.

Wayne, Robert K. "Molecular evolution of the dog family." *Trends in Genetics* 9.6 (1993): 218-224.

Weaver, Ian CG, Michael J. Meaney, and Moshe Szyf. "Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood." *Proceedings of the National Academy of Sciences of the United States of America* 103.9 (2006): 3480-3485.

Weinstein, Lee S., et al. “Minireview: GNAS: normal and abnormal functions.” *Endocrinology* 145.12 (2004): 5459-5464.

Wessa, Michele, et al. “Altered cortisol awakening response in posttraumatic stress disorder.” *Psychoneuroendocrinology* 31.2 (2006): 209-215.

Wiel, Clotilde, et al. “Endoplasmic reticulum calcium release through ITPR2 channels leads to mitochondrial calcium accumulation and senescence.” *Nature Communications* 5 (2014).

Wilhelm, Brian T., and Josette-Renée Landry. "RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing." *Methods* 48.3 (2009): 249-257.

Wilhelm, Léa P., Catherine Tomasetto, and Fabien Alpy. "Touche! STARD3 and STARD3NL tether the ER to endosomes." *Biochemical Society Transactions* 44.2 (2016): 493-498.

Wilkins, Adam S., Richard W. Wrangham, and W. Tecumseh Fitch. "The 'domestication syndrome' in mammals: A unified explanation based on neural crest cell behavior and genetics." *Genetics* 197.3 (2014): 795-808.

Williams, Alexander G., et al. "RNA‐seq Data: Challenges in and Recommendations for Experimental Design and Analysis." *Current Protocols in Human Genetics* (2014): 11-13.

Williams, Sarah K., et al. "Role of Mayven, a kelch‐related protein in oligodendrocyte process formation." *Journal of Neuroscience Research* 81.5 (2005): 622-631.

Wolf, Jochen BW. "Principles of transcriptome analysis and gene expression quantification: an RNA‐seq tutorial." *Molecular ecology resources* 13.4 (2013): 559-572.

Wu, Minghua, et al. "LRRC4 inhibits glioblastoma cell proliferation, migration, and angiogenesis by downregulating pleiotropic cytokine expression and responses." *Journal of Cellular Physiology* 214.1 (2008): 65-74.

Xiao, Jianmin, et al. "Removal of FKBP12. 6 does not alter the conductance and activation of the cardiac ryanodine receptor or the susceptibility to stress-induced ventricular arrhythmias." *Journal of Biological Chemistry* 282.48 (2007): 34828-34838.

Yamagishi, Satoru, et al. "FLRT2 and FLRT3 act as repulsive guidance cues for Unc5‐positive neurons." *The EMBO Journal* 30.14 (2011): 2920-2933.

Yamazaki, Hiroyuki, et al. "Spikar, a novel drebrin‐binding protein, regulates the formation and stabilization of dendritic spines." *Journal of Neurochemistry* 128.4 (2014): 507-522.

Yang, Fang, et al. "Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis." *The Journal of Cell Biology* 173.4 (2006): 497-507.

Yang, Tong, et al. “Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER.” *Cell* 110.4 (2002): 489-500.

Ye, Liang, et al. "A vertebrate case study of the quality of assemblies derived from next-generation sequences." *Genome biology* 12.3 (2011): R31.

Yin, H. L. “Gelsolin: calcium‐and polyphosphoinositide‐regulated actin‐modulating protein.” *Bioessays* 7.4 (1987): 176-179.

Young, Allan H., Peter Gallagher, and Richard J. Porter. “Elevation of the cortisol-dehydroepiandrosterone ratio in drug-free depressed patients.” *American Journal of Psychiatry* 159.7 (2002): 1237-1239.

Zanoni, Paolo, et al. “Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease.” *Science* 351.6278 (2016): 1166-1171.

Zech, Tobias, et al. “The Arp2/3 activator WASH regulates α5β1-integrin-mediated invasive migration.” *Journal of Cell Science* 124.22 (2011): 3753-3759.

Zerbes, Ralf M., et al. “Role of MINOS in mitochondrial membrane architecture: cristae morphology and outer membrane interactions differentially depend on mitofilin domains.” *Journal of Molecular Biology* 422.2 (2012): 183-191.

Zerbino, Daniel R., and Ewan Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." *Genome research* 18.5 (2008): 821-829.

Zhang, Jian, et al. “YPEL3 suppresses epithelial–mesenchymal transition and metastasis of nasopharyngeal carcinoma cells through the Wnt/β-catenin signaling pathway.” *Journal of Experimental & Clinical Cancer Research* 35.1 (2016): 109.

Zhang, Jun, et al. “FNBP1 was involved with the morphological control in the hepatocyte.” *The FASEB Journal* 25.1 Supplement (2011): lb111-lb111.

Zhang, Kuixing, et al. “Autonomic function in hypertension: Role of genetic variation at the catecholamine storage vesicle protein chromogranin B (CHGB).” *Circulation: Cardiovascular Genetics* (2009): CIRCGENETICS-108.

Zhang, Michael Y., et al. "Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy." *Nature Genetics* 47 (2015): 180-185.

Zhang, Ying, Tae-Hee Kim, and Lee Niswander. "Phactr4 regulates directional migration of enteric neural crest through PP1, integrin signaling, and cofilin activity." *Genes & Development* 26.1 (2012): 69-81.

Zhang, Yong, et al. "Bm-TFF2, a toad trefoil factor, promotes cell migration, survival and wound healing." *Biochemical and Biophysical Research Communications* 398.3 (2010): 559-564.

Zhao, Qi, et al. "Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line." *Proceedings of the National Academy of Sciences* 106.6 (2009): 1886-1891.

Zhao, Shanrong, and Baohong Zhang. "A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification." *BMC genomics* 16.1 (2015): 97.

Zhou, Zhengkui, et al. "Differential genetic regulation of canine hip dysplasia and osteoarthritis." *PloS one* 5.10 (2010): e13219.

Zhuo, De-Xiang, et al. "CSTP1, a novel protein phosphatase, blocks cell cycle, promotes cell apoptosis, and suppresses tumor growth of bladder cancer by directly dephosphorylating Akt at Ser473 site." *PloS One* 8.6 (2013): e65679.

Ziegler-Heitbrock, Loems, et al. "Nomenclature of monocytes and dendritic cells in blood." *Blood* 116.16 (2010): e74-e80.

Zimin, Aleksey V., et al. "A whole-genome assembly of the domestic cow, Bos taurus." *Genome biology* 10.4 (2009): R42.

Zinsstag, Jakob, et al. "From "one medicine" to "one health" and systemic approaches to health and well-being." *Preventive veterinary medicine* 101.3 (2011): 148-156.

APPENDIX A: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Supplementary tables and figures for Chapter 3. Descriptions of associated supplementary files are included below for each file separately.

Table SI-3.1. Primer sequences for quantitative PCR

Gene	Sequence of forward (F) and reverse (R) primers	Start location of fox amplicon in CanFam3.1
AKAP7	F: CCTATCCATTCCGATCACCAAC R: AGTCGCTCATCTTGCTGGAG	chr1:69376373
ANKRD6	F: AACACGGCTCTACACGAAGC R: CTCTGGGAGTGGCTGTTCTG	chr12:48811400
BRCA2	F: GCAGAACTGGTAGGCTCTCC R: ATACCAGCAAGCAGGACGAG	chr25:7753591
CDH4	F: GACATCTGGATCCCACTGGC R: TCAGGACACTTCAGCATCCG	chr24:46055878
ITGA1	F: GGCCAATGAGACTGTCCCTG R: CAGCTTAAGTTCTGGCATCGG	chr4:62309206
LAMA2	F: ACCTCTGCTCGCTATATTCGC R:CCGACTGAAATATCCTTGACCG	chr1: 67735809
NELL2	F: CCTTCAGCGCACTCATCAATG R: GGCCGTTTGTAGCTGTCGAG	chr27: 9535623
PDE7B	F: CCCATTCATTGACTTCCGCC R: TCCACGAAGCAACCTCGATG	chr1: 28719580
SDHA	F: CCTGCGTCACAAAGCTCTTT R: TGTCATGTAGTGGATGGCGT	chr34:11957015

Gene	Sequence of forward (F) and reverse (R) primers	Start location of fox amplicon in CanFam3.1
SPOCK1	F: GGGAGTCCTTGCAGATTGGC R: ACTGGTTTGGAGCCCTTCAC	chr11: 24748659
THBS4	F: TGCAAGTTGGTCAGTCTCAGG R: CACTGTGATGGGACGCTTAAAC	chr3: 26970860

Table SI-3.1 (cont.)

Table SI-3.2. Reads from sequencing of tame and aggressive fox anterior pituitary RNA

Sample		Reads (unfiltered)	Reads (filtered)
Tame	1-601	34,577,964	32,479,541
	2-662	31,279,625	29,749,267
	3-663	33,576,358	31,677,842
	4-664	37,936,135	36,086,260
	5-665	34,731,107	31,347,816
	6-666	34,735,178	32,916,697
Aggressive	49-649	32,688,890	32,866,684
	50-710	33,570,511	32,694,392
	51-711	32,901,110	31,043,144
	52-712	33,126,696	31,429,811
	53-713	32,876,971	31,237,749
	54-714	34,760,977	31,328,261
Tame		Mean = 34,168,005 SD = 2,239,555	Mean = 32,376,237 SD = 2,121,953
Aggressive		Mean = 33,625,582	Mean = 31,766,674

	SD = 906,373	SD = 797,449

Table SI-3.2 (cont)

Table SI-3.3. Alignment rate of reads from tame and aggressive fox anterior pituitary RNA

Sample		**Alignment rate (%)**
Tame	1-601	79.0
	2-662	78.9
	3-663	79.1
	4-664	79.5
	5-665	79.2
	6-666	79.3
Aggressive	49-649	78.5
	50-710	78.9
	51-711	78.8
	52-712	79.0
	53-713	79.0
	54-714	78.5

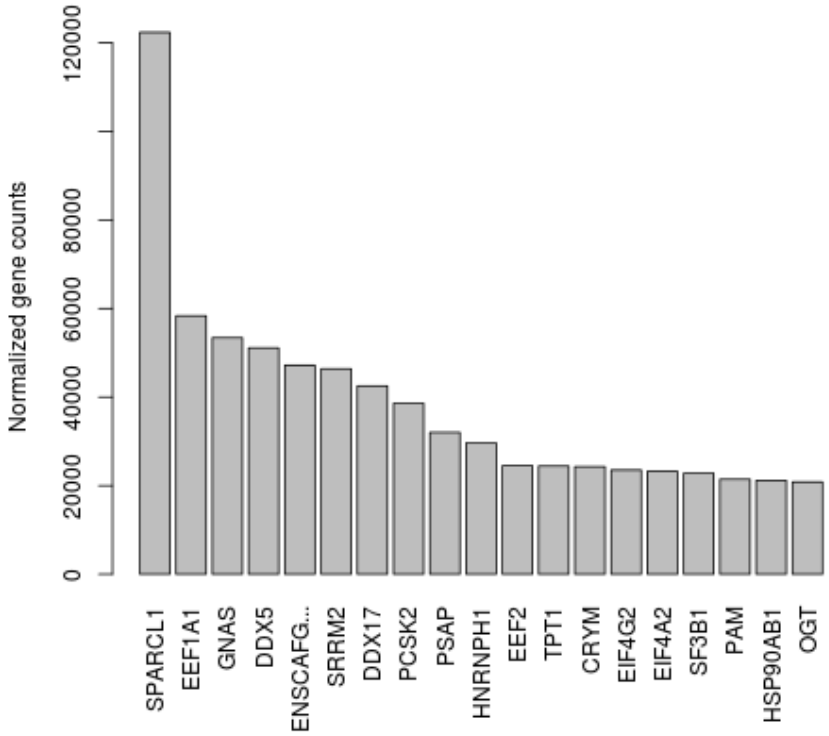

Figure SI-3.1. Most highly (0.01%) expressed genes in fox pituitary by normalized gene count, assessed by DESeq2. *POMC* is excluded, because its order of magnitude larger expression level than the next largest gene results in a scale that does not allow detailed values to be displayed for genes with smaller expression levels.

Table SI-3.4. Significantly enriched GO terms in the 0.1% (180) most highly expressed genes in fox anterior pituitary tissue, including GO identifier, name of enriched GO term, number of DE genes associated with the GO term, percent of genes associated with the GO term which are DE, p-value, corrected p-value (FDR), symbols of the DE genes in the GO term, and symbols of all the genes in the GO term.

File fox-pituitary-topexpr-GO-SI-3.4.xlsx

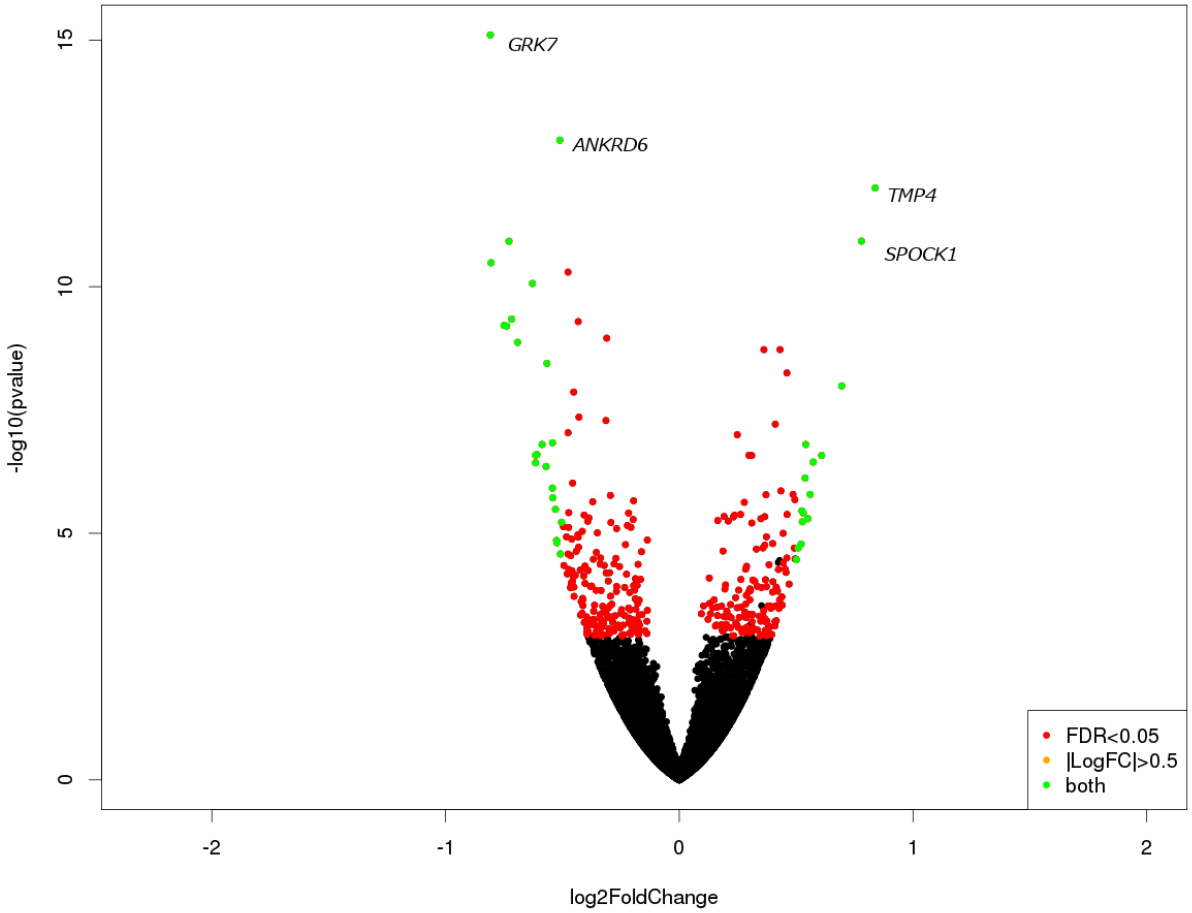

Figure SI-3.2. Volcano plot of gene expression in fox anterior pituitary. Red dots represent genes significantly differentially expressed in tame versus aggressive fox lines at FDR < 0.05. Green dots represent genes that are both significantly differentially expressed and have a log2 fold change > 0.5. Outlier genes are labeled with their gene symbols.

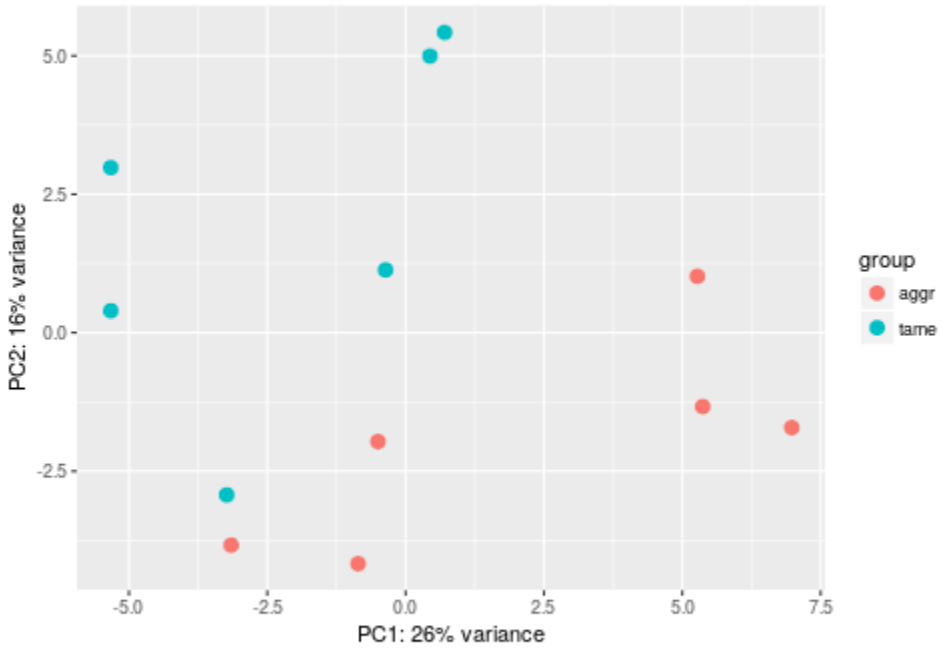

Figure SI-3.3. Principal components analysis of tame versus aggressive fox anterior pituitary samples by gene expression, comparing PC1 vs PC2. Note clustering of tame (blue) versus aggressive (red) samples.

Table SI-3.5. Differentially expressed genes in tame versus aggressive fox anterior pituitary tissue.

File fox-pituitary-DEG-SI-3.5.xlsx

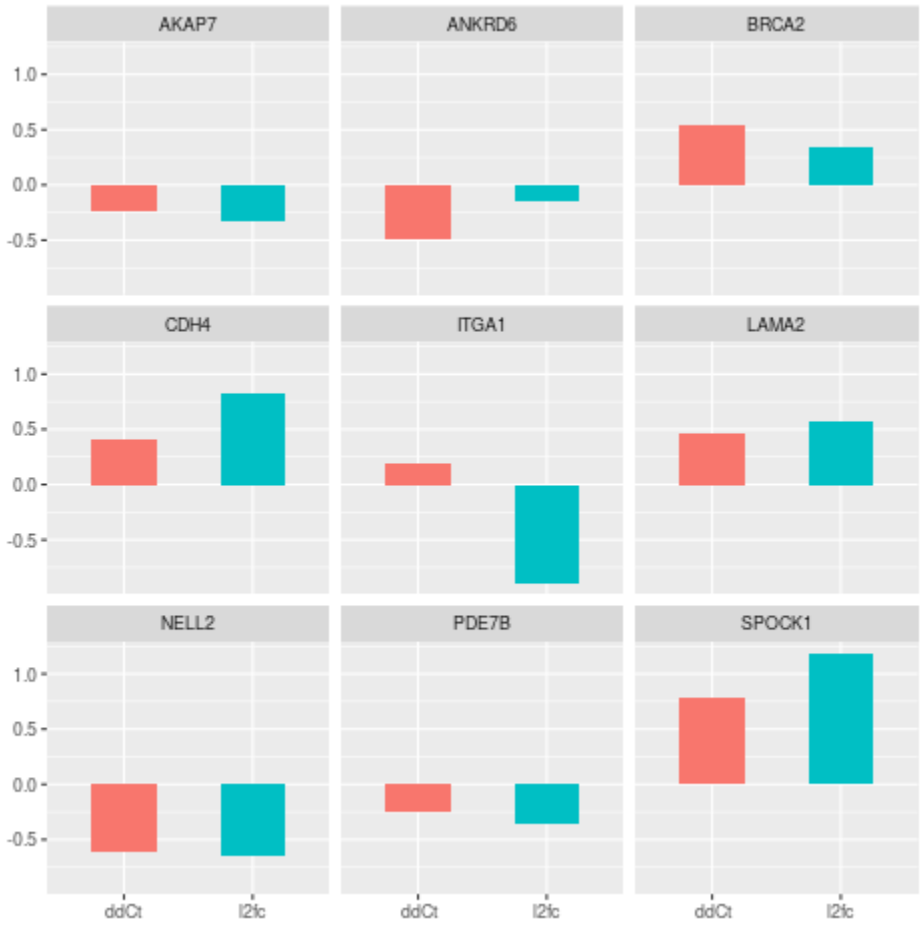

Figure SI-3.4. Comparison of RT-qPCR and RNA sequencing results in tame and aggressive fox anterior pituitary tissue. The y-axis represents ddCt values (qPCR; red bars) and $\log_2$fold change values (RNA-seq; blue bars).

Table SI-3.6. Comparison of RT-qPCR and RNA sequencing results in tame and aggressive fox anterior pituitary tissue. ddCt values for qPCR assays were calculated using the Applied Biosystems StepOnePlus Real Time PCR System software (Applied Biosystems, Foster City, CA) for all tame versus all aggressive samples for each gene. Significant differences in gene ddCt values were calculated by comparing individual ddCt values from tame and aggressive samples for each gene using the Wilcoxon rank-sum test.

Gene	**RNA-seq**		**qPCR**	
	Log_2 fold change	**FDR**	**ddCt**	**Wilcoxon**
AKAP7	-0.23	0.003	-0.32	W=4, p=0.63
ANKRD6	-0.51	7.98e-10	-0.15	W=9, p=0.40
BRCA2	0.54	8.46e-5	0.33	W=0, p=0.10
CDH4	0.40	0.049	0.83	W=1, p=0.11
ITGA1	0.21	0.029	-0.89	W=7, p=0.85
LAMA2	0.46	4.66e-6	0.58	W=1, p=0.11
NELL2	-0.61	0.0001	-0.64	W=11, p=0.11
PDE7B	-0.25	0.024	-0.35	W=12, 0.34
SPOCK1	0.78	3.58e-8	1.19	W=1, 0.06

Table SI-3.7. Significantly enriched GO terms in differentially expressed genes in tame versus aggressive fox anterior pituitary tissue, including GO identifier, name of enriched GO term, number of DE genes associated with the GO term, percent of genes associated with the GO term which are DE, p-value, corrected p-value (FDR), symbols of the DE genes in the GO term, and symbols of all the genes in the GO term.

File fox-pituitary-GO-SI-3.7.xlsx

Table SI-3.8. Antique White 4, Pale Turquoise, and Brown WGCNA modules, including gene symbol and module membership (e.g., the correlation of its gene expression profile with the module eigengene), and whether individual genes are differentially expressed in tame versus aggressive fox anterior pituitary.

File fox-pituitary-WGCNA-SI-3.8.xlsx

Table SI-3.9. All WGCNA modules identified in fox anterior pituitary, including genes in each module and the module membership of each gene (e.g., the correlation of its gene expression profile with the module eigengene).

File fox-pituitary-WGCNA-membership-SI-3.9.xlsx

Table SI-3.10. Enriched GO terms in 13 WGCNA modules in tame and aggressive fox anterior pituitary, including GO identifier, name of enriched GO term, p-value, corrected p-value (FDR), percent of genes in the GO term which are associated with the module, number of genes in the GO term which are associated with the module, and the list of gene symbols in the GO term which are associated with the module. Modules were selected from those with eigengenes that significantly differed between tame and aggressive samples before correction with Benjamini-Hochberg.

File fox-pituitary-WGCNA-GO-SI-3.10.xlsx.

Table SI-3.11. Exons differentially skipped in tame versus aggressive fox anterior pituitary tissue, as identified by rMATS, with Ensembl ID, gene symbol, p value, FDR, percent of reads covering skipped exon in tame samples (mean), and percent of reads covering skipped exon in aggressive samples (mean).

File fox-pituitary-skippedexons-SI-3.11.xlsx

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Supplementary tables and figures for Chapter 4. Descriptions of associated supplementary files are included below for each file separately.

Table SI-4.1. Reads from sequencing of 11 tame and 11 aggressive fox adrenal RNA samples.

Sample		Reads (unfiltered)	Reads (filtered)
Tame	1-61	38,080,258	35,453,022
	2-62	26,619,372	25,116,612
	3-63	33,565,028	31,087,886
	4-64	27,142,330	25,468,776
	5-65	44,616,527	41,830,498
	6-66	27,785,950	30,695,800
	7-67	32,668,307	38,718,553
	8-68	41,337,408	30,423,310
	9-69	32,370,400	28,367,860
	11-71	28,407,990	30,867,683
	12-72	33,082,996	30,337,793
Aggressive	49-109	32,523,228	29,851,993
	50-110	31,633,578	27,808,400
	51-111	29,861,173	26,548,048
	52-112	28,637,677	29,452,195
	53-113	31,537,038	32,666,398
	55-115	32,959,959	35,649,632

Sample		Reads (unfiltered)	Reads (filtered)
	56-116	38,366,741	24,256,345
	57-117	25,922,623	27,518,014
	58-118	29,308,691	27,526,058
	59-119	29,283,118	34,341,866
	60-120	36,992,713	25,997,671
Tame		Mean = 33,243,324 SD = 5,947,638	Mean = 31,669,799 SD = 5,145,833
Aggressive		Mean = 31,547,867 SD = 3,635,886	Mean = 29,237,875 SD = 3,603,062

Table SI-4.1 (cont.)

Table SI-4.2. Alignment rate of reads from tame and aggressive fox adrenal RNA

Sample		Alignment rate (%)
Tame	1-61	79.3
	2-62	79.0
	3-63	79.5
	4-64	79.9
	5-65	79.2
	6-66	79.0
	7-67	79.1
	8-68	79.1
	9-69	79.4
	11-71	79.4

Sample		Alignment rate (%)
	12-72	79.4
Aggressive	49-109	79.0
	50-110	79.2
	51-111	78.9
	52-112	79.0
	53-113	79.1
	55-115	78.9
	56-116	79.6
	57-117	79.1
	58-118	79.0
	59-119	79.9
	60-120	79.7

Table SI-4.2 (cont.)

Table SI-4.3. Significantly enriched GO terms in the 0.01% (188) most highly expressed genes in fox adrenal tissue, including GO identifier, name of enriched GO term, number of DE genes associated with the GO term, percent of genes associated with the GO term which are DE, p-value, corrected p-value (FDR), and symbols of the DE genes in the GO term.

File fox-adrenal-topexpr-GO-4.3.xlsx

Table SI-4. Differentially expressed genes in tame versus aggressive fox adrenal tissue including Ensembl ID, gene symbol, normalized mean expression for tame and aggressive samples combined, $\log_2$ fold change, $\log_2$ fold change standard error, p-value, and FDR.

File fox-adrenals-DEG-SI-4.4.xlsx

Table SI-4.5. Principal components analysis of tame and aggressive gene expression levels in the adrenal gland, limited to analysis of the 500 genes with the most variance, including loadings on PC1 and PC2.

File fox-adrenals-pca-SI-4.5.xlsx

Table SI-4.6. GO terms enriched for genes in the 50 highest and 50 lowest scoring genes on PC2 in principal components analysis of tame and aggressive gene expression levels in the fox adrenal gland, including GO identifier, name of enriched GO term, number of DE genes associated with the GO term, percent of genes associated with the GO term which are DE, p-value, corrected p-value (FDR), and symbols of the DE genes in the GO term.

File fox-adrenals-pc2-GO-SI-4.6.xlsx

Table SI-4.7. Genes that are significantly differentially expressed in both pituitary and adrenal tissue between tame and aggressive foxes, including Ensembl ID, gene symbol, normalized mean expression for adrenals and for pituitary, $\log_2$ fold change and $\log_2$ fold change standard error for adrenals and for pituitary, and p-value and FDR for adrenals and for anterior pituitary.

File fox-adrenals-DEG-pit-SI-4.7.xlsx

Table SI-4.8. GO terms enriched for genes that are differentially expressed in the tame and aggressive fox adrenal, including GO identifier, name of enriched GO term, number of DE genes associated with the GO term, percent of genes associated with the GO term which are DE, p-value, corrected p-value (FDR), and symbols of the DE genes in the GO term.

File fox-adrenals-GO-SI-4.8.xlsx

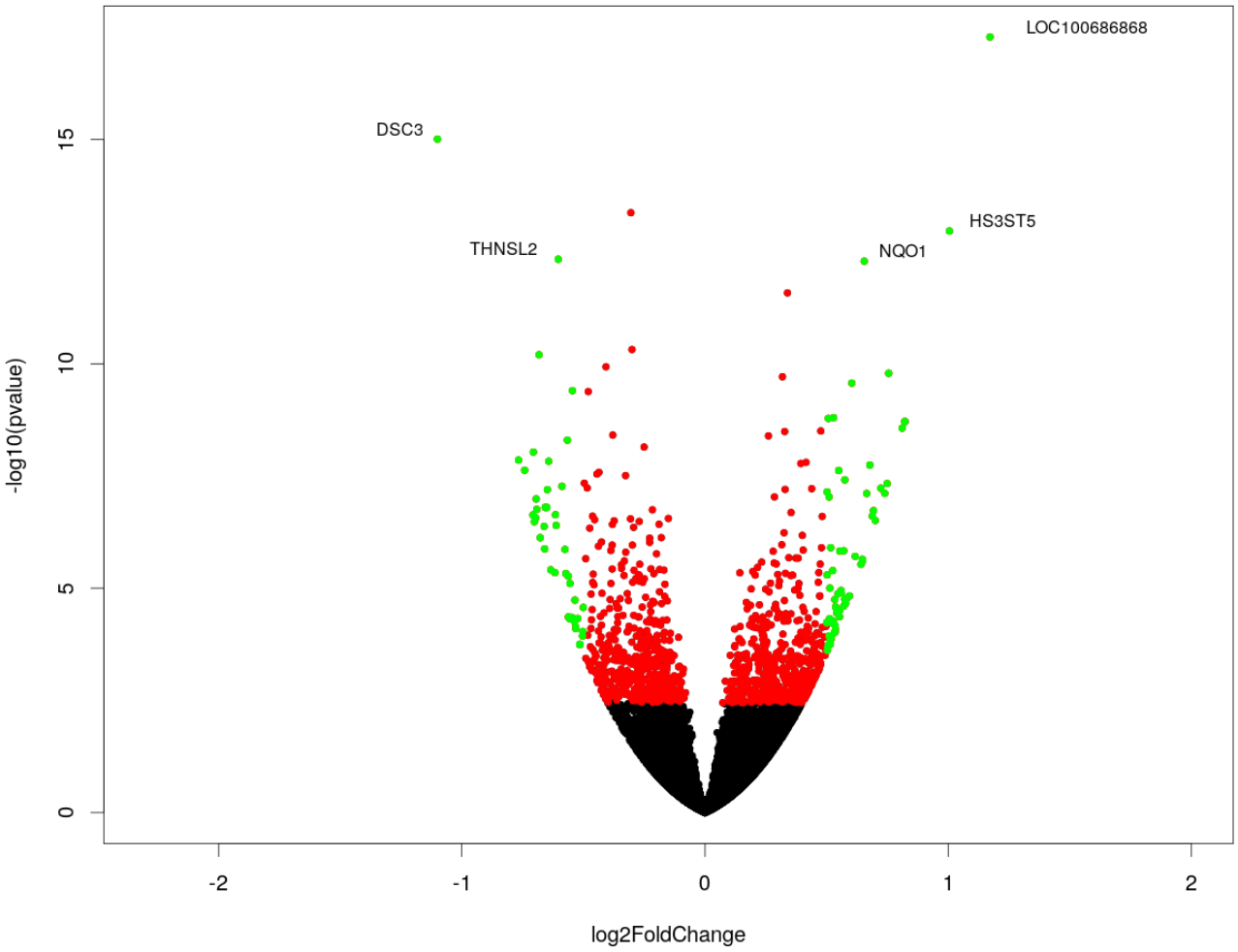

Figure SI-4.1. Volcano plot of gene expression in fox adrenal. Red dots represent genes significantly differentially expressed in tame versus aggressive fox lines at FDR < 0.05. Green dots represent genes that are both significantly differentially expressed and have a $\log_2$ fold change > 0.5. Genes with $\log_2$ fold change > 2 are labeled with their gene symbols.

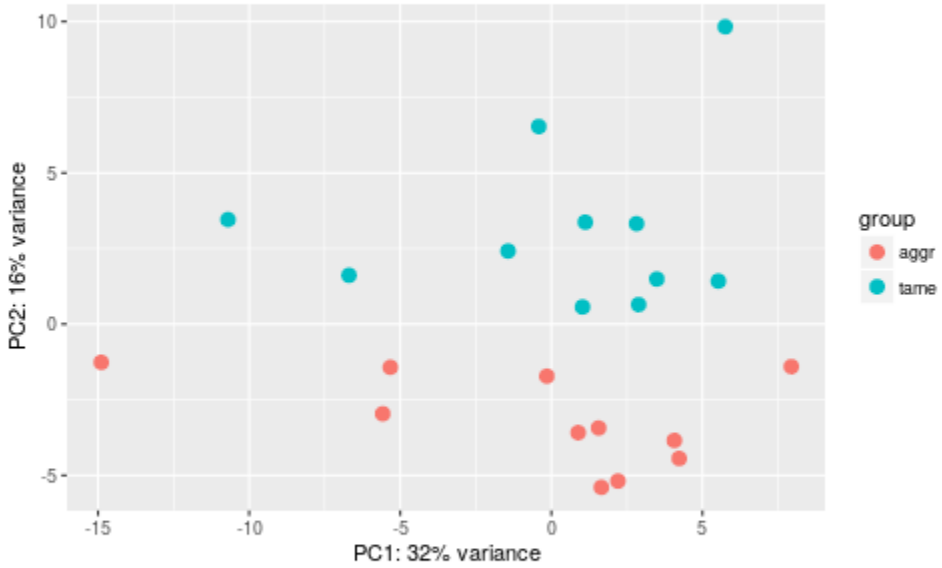

Figure SI-4.2. Principal components analysis of tame versus aggressive fox adrenal samples by gene expression, comparing PC1 vs. PC2. Note clustering of tame (blue) versus aggressive (red) samples.

Table SI-4.9. All WGCNA modules identified in fox adrenals, including genes in each module and the module membership of each gene (e.g., the correlation of its gene expression profile with the module eigengene).

File fox-adrenals-WGCNA-membership-SI-4.9.xlsx

Table SI-4.10. Magenta, Green, Tan, and Black WGCNA modules, including gene symbol, module membership (e.g., the correlation of its gene expression profile with the module

eigengene), and whether individual genes are differentially expressed in tame versus aggressive fox adrenals.

File fox-adrenal-WGCNA-SI-4.9.xlsx

Table SI-4.11. Enriched GO terms in WGCNA modules in tame versus aggressive fox adrenal, including GO identifier, name of enriched GO term, p-value, corrected p-value (FDR), percent of genes in the GO term which are associated with the module, number of genes in the GO term which are associated with the module, and the list of gene symbols in the GO term which are associated with the module.

File fox-adrenals-WGCNA-GO-SI-4.11.xlsx

Table SI-4.12. Exons differentially skipped in tame versus aggressive fox anterior adrenal tissue, as identified by rMATS, with Ensembl ID, gene symbol, p value, FDR, percent of reads covering skipped exon in tame samples (mean), and percent of reads covering skipped exon in aggressive samples (mean).

File fox-adrenals-skippedexons-SI-4.12.xlsx

APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Supplementary tables and figures for Chapter 5. Descriptions of associated supplementary files are included below for each file separately.

Table SI-5.1. Reads from sequencing of 12 tame and 12 aggressive fox hypothalamus RNA samples.

Sample		Reads (unfiltered)	Reads (filtered)
Tame	1-481	25,660,137	22,873,468
	2-124	26,443,307	24,105,777
	3-23	30,665,195	27,329,915
	4-124	22,148,469	20,265,426
	5-125	30,862,922	28,138,900
	6-126	29,082,395	25,825,727
	7-127	26,061,930	23,351,545
	8-128	37,976,178	34,055,594
	9-129	27,698,745	25,250,928
	10-130	23,612,547	20,014,462
	11-131	37,152,640	33,792,092
	12-132	20,148,293	18,085,090
Aggressive	49-169	34,834,952	31,399,500
	50-170	33,604,749	30,789,693
	51-171	35,814,615	32,602,800
	52-172	28,294,005	25,107,221

Sample		Reads (unfiltered)	Reads (filtered)
	53-173	31,324,019	28,425,929
	54-174	28,651,874	26,129,092
	55-535	25,075,833	22,367,895
	56-176	31,059,935	27,235,730
	57-177	42,579,903	38,603,124
	58-178	29,013,282	26,159,574
	59-179	37,280,921	33,827,745
	60-180	31,092,675	28,383,468
Tame		Mean = 28,540,931 SD = 6,233,102	Mean = 25,257,410 SD = 5,024,146
Aggressive		Mean = 32,018,706 SD = 5,271,778	Mean = 29,252,648 SD = 4,431,016

Table SI-5.1 (cont.)

Table SI-5.2. Filtered variants called in hypothalamus samples, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus between tame and aggressive samples; region of putative selective sweeps in which each variant falls, if any, as defined in (Kukekova et al., under review); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File hyp-filt-SI-5.2.xlsx

Table SI-5.3. Variants called in hypothalamus samples with high or moderate impact (as annotated by SNPeff), including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus between tame and aggressive samples; regions of putative selective sweeps in which each variant falls, if any, as defined in (Kukekova et al., under review); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); and amino acid position (amino acid position / total number of amino acids). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File hyp-snpeffhighmod-SI-5.3.xlsx

Table SI-5.4. Genes associated with high and moderate impact variants in hypothalamus samples, including gene name and a short description of the gene.

File hyp-snpeffhighmod-genes-SI-5.4.xlsx

Table SI-5.5. Gene ontology enrichment analysis for genes associated with HMI variants in hypothalamus samples, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with HMI variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with HMI variants.

File hyp-ClueGO-snpeffhighmod-SI-5.5.xls

Table SI-5.6. Variants called in hypothalamus samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File hyp-highfst-SI-5.6.xlsx

Table SI-5.7. Genes associated with variants in hypothalamus samples that have $F_{ST} > 0.5$, including gene name and a short description of the gene.

File hyp-highfst-genes-SI-5.7.xlsx

Table SI-5.8. Windows in regions of putative selective sweeps (as described in Kukekova et al., under review), including the number of hypothalamus variants from this study that fall within each window, number of region that the window falls within, fox and dog chromosomes where the window is found, start and end of the window location on the dog chromosome, type of window (pooled heterozygosity in tame animals, pooled heterozygosity in aggressive animals, or F_{ST} between tame and aggressive animals; some windows are marked "merge" to indicate that they lie between two windows of the indicated type, e.g., between two windows of reduced pooled heterozygosity in tame animals), fox scaffold number with start and end of the window in that scaffold, pooled heterozygosity of tame samples in that window, pooled heterozygosity of aggressive samples in that window, F_{ST} between tame and aggressive samples in that window, and genes in the window with high-F_{ST} variants called in

hypothalamus in this study. (Note that samples used to calculate Hp and F_{ST} in windows are different from the samples used in this study.)

File hyp-regions-SI-5.8.xlsx

Table SI-5.9. Gene ontology enrichment analysis for genes in hypothalamus samples that are associated with variants with $F_{ST} > 0.5$, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with high-F_{ST} variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with high-F_{ST} variants.

File hyp-ClueGO-highfst-SI-5.9.xls

Table SI-5.10. HMI variants called in hypothalamus samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron and total number of exons or introns); amino acid position; and total number of amino acids. Where multiple SNPeff annotations were available for a variant, only the first was reported.

File hyp-highhigh-SI-5.10.xlsx

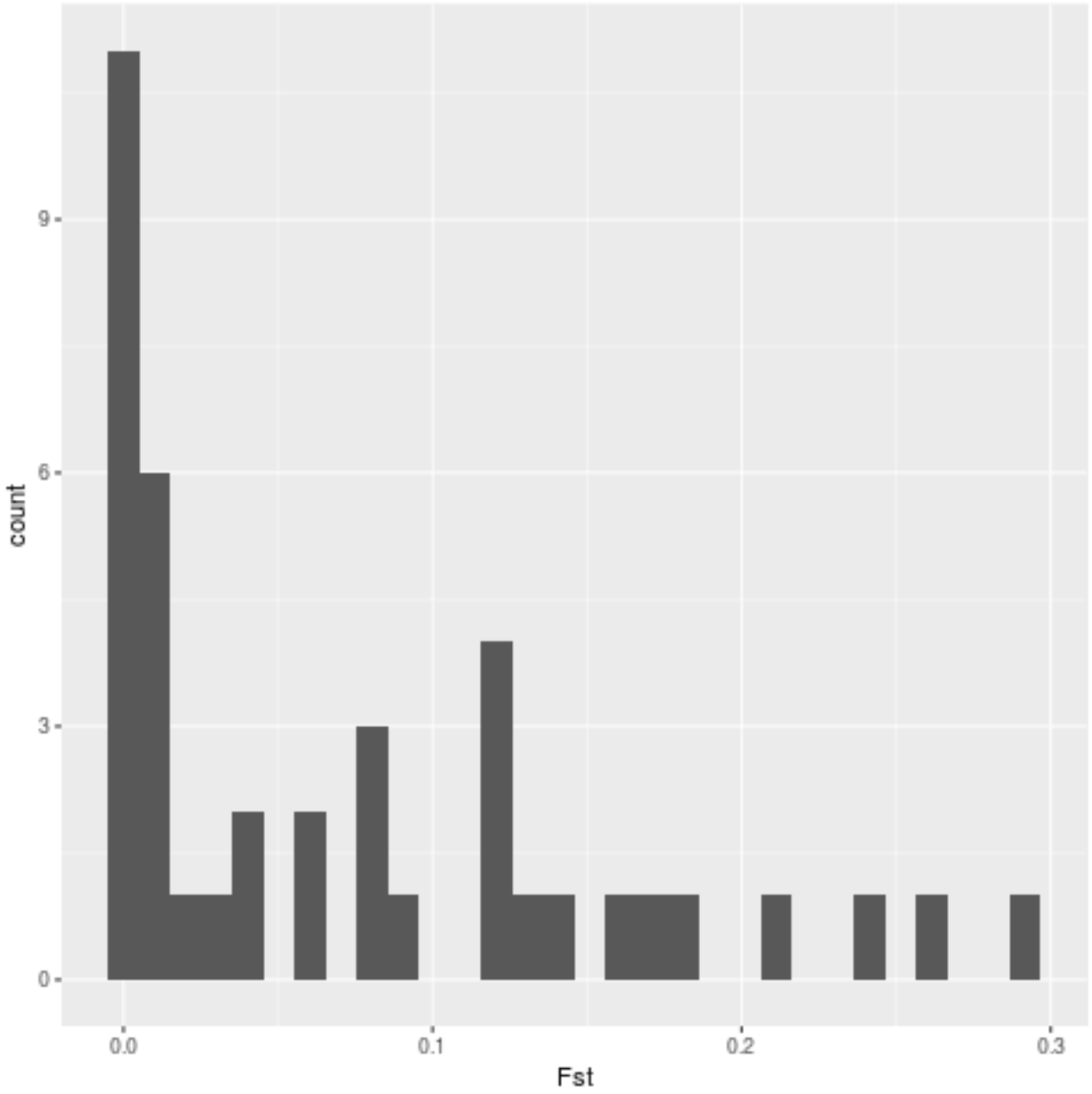

Figure SI-5.1. F_{ST} of high impact variants in hypothalamus samples.

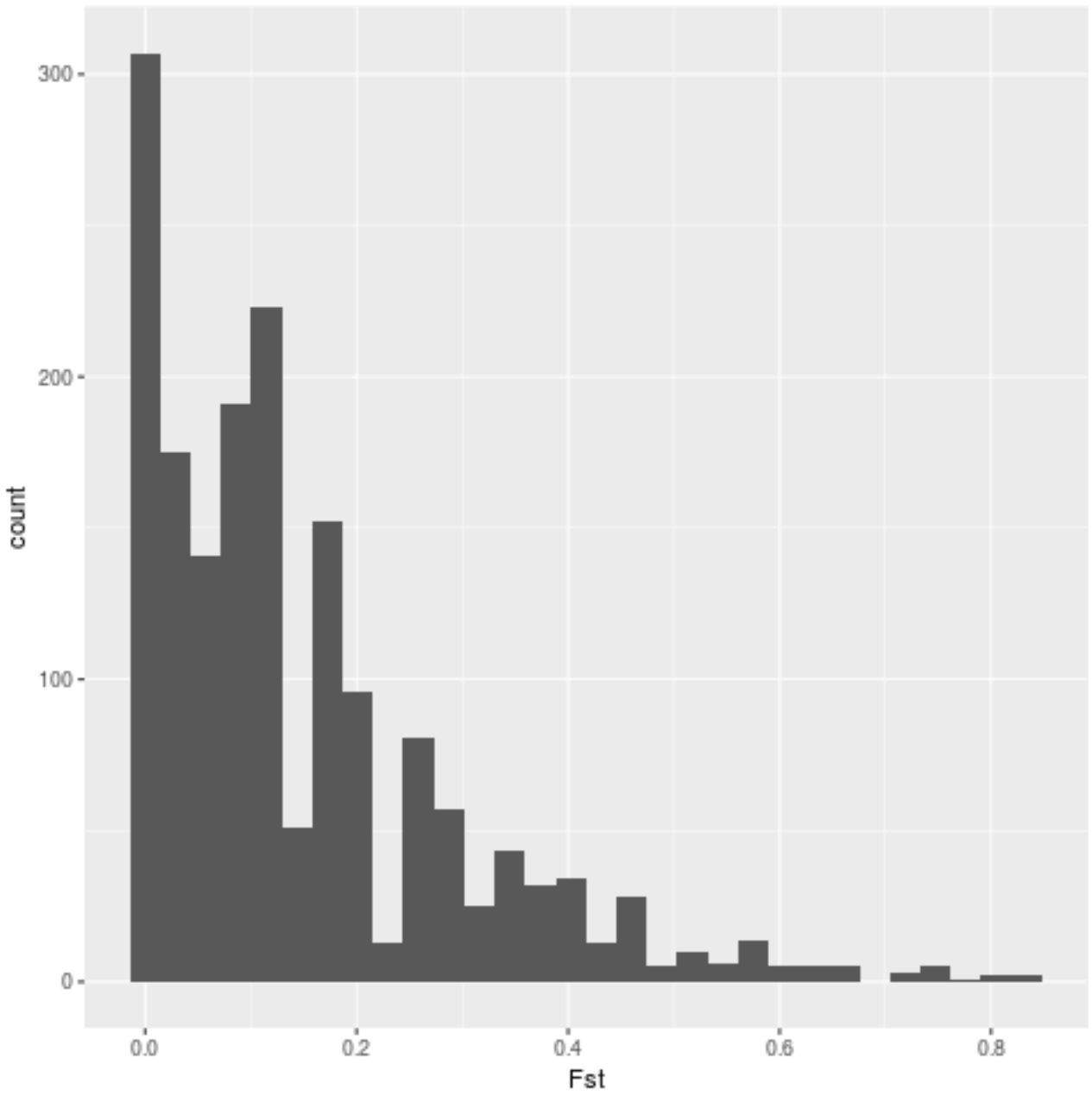

Figure SI-5.2. F_{ST} of moderate impact variants in hypothalamus samples.

Table SI-5.11. Filtered variants called in anterior pituitary samples, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus between tame and aggressive samples; regions of putative selective sweeps in which each variant falls, if any, as defined in (Kukekova et al., under review); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate

allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File pit-filt-SI-5.11.xlsx

Table SI-5.12. Variants called in anterior pituitary samples with high or moderate impact (as annotated by SNPeff), including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus between tame and aggressive samples; regions of putative selective sweeps in which each variant falls, if any, as defined in (Kukekova et al., under review); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); and amino acid position (amino acid position / total number of amino acids). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File pit-snpeffhighmod-SI-5.12.xlsx

Table SI-5.13. Genes associated with high and moderate impact variants in anterior pituitary samples, including gene name, whether or not it is differentially expressed in tame versus aggressive fox anterior pituitary, and a short description of the gene.

File pit-snpeffhighmod-genes-SI-5.13.xlsx

Table SI-5.14. Gene ontology enrichment analysis for genes associated with HMI variants in anterior pituitary samples, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with HMI variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with HMI variants.

File pit-ClueGO-snpeffhighmod-SI-5.14.xls

Table SI-5.15. Variants called in anterior pituitary samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File pit-highfst-SI-5.15.xlsx

Table SI-5.16. Genes associated with variants in anterior pituitary samples that have $F_{ST} > 0.5$, including gene name, whether or not it is differentially expressed in tame versus aggressive fox anterior pituitary, and a short description of the gene.

File pit-highfst-genes-SI-5.16.xlsx

Table SI-5.17. Windows in regions of putative selective sweeps (as described in Kukekova et al., under review), including the number of hypothalamus variants from this study that fall within

each window, number of region that the window falls within, fox and dog chromosomes where the window is found, start and end of the window location on the dog chromosome, type of window (pooled heterozygosity in tame animals, pooled heterozygosity in aggressive animals, or F_{ST} between tame and aggressive animals; some windows are marked "merge" to indicate that they lie between two windows of the indicated type, e.g., between two windows of reduced pooled heterozygosity in tame animals), fox scaffold number with start and end of the window in that scaffold, pooled heterozygosity of tame samples in that window, pooled heterozygosity of aggressive samples in that window, F_{ST} between tame and aggressive samples in that window, genes in the window with high-F_{ST} variants called in anterior pituitary in this study, and differentially expressed genes in anterior pituitary from this study found in this window. (Note that samples used to calculate Hp and F_{ST} in windows are different from the samples used in this study.)

File pit-regions-SI-5.17.xlsx

Table SI-5.18. Gene ontology enrichment analysis for genes in anterior pituitary samples that are associated with variants with $F_{ST} > 0.5$, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with high-F_{ST} variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with high-F_{ST} variants.

File pit-ClueGO-highfst-SI-5.18.xls

Table SI-5.19. HMI variants called in anterior pituitary samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of

exons or introns); and amino acid position (amino acid position / total number of amino acids). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File pit-highhigh-SI-5.19.xlsx

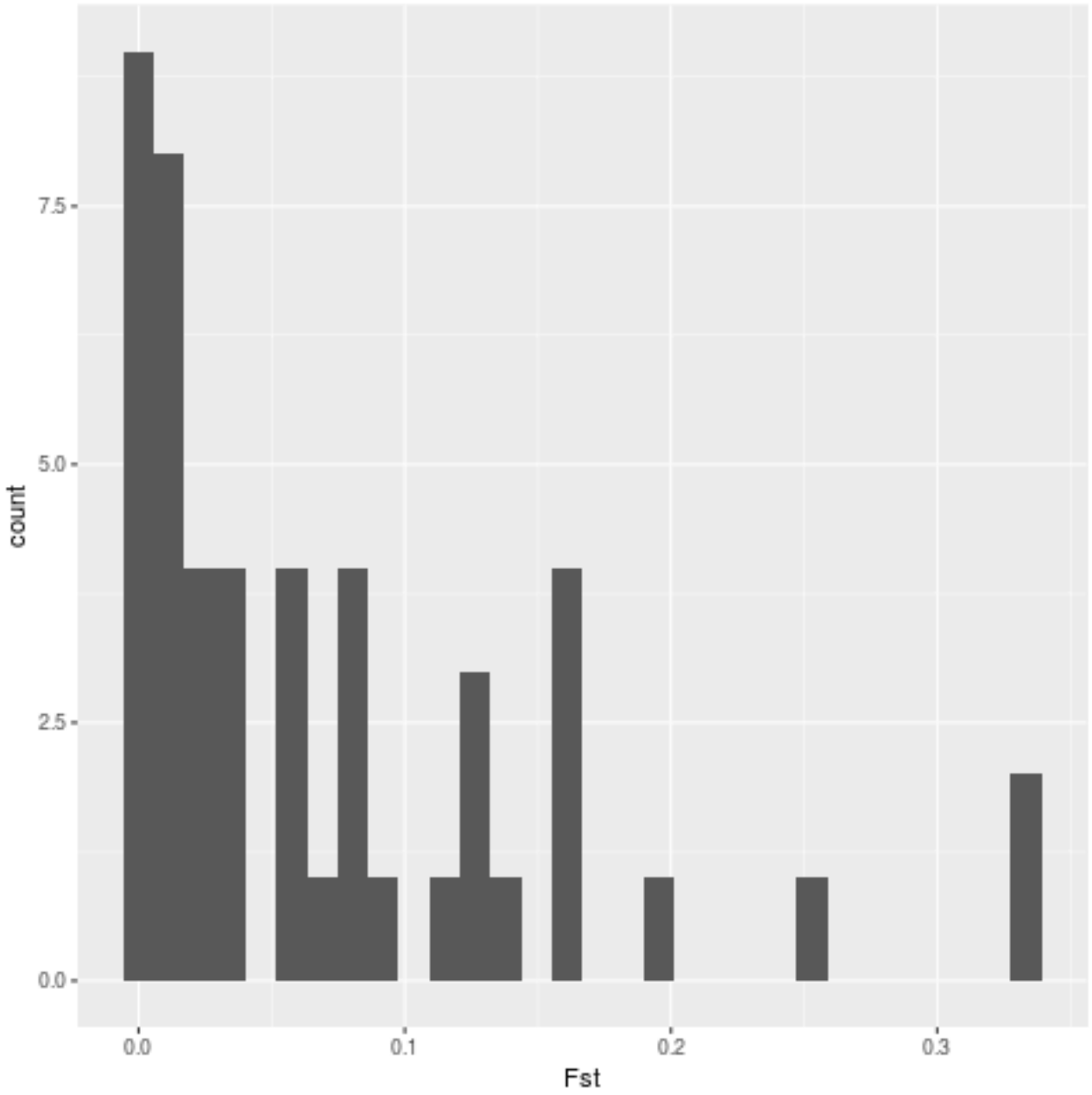

Figure SI-5.3. F_{ST} of high impact variants in anterior pituitary samples.

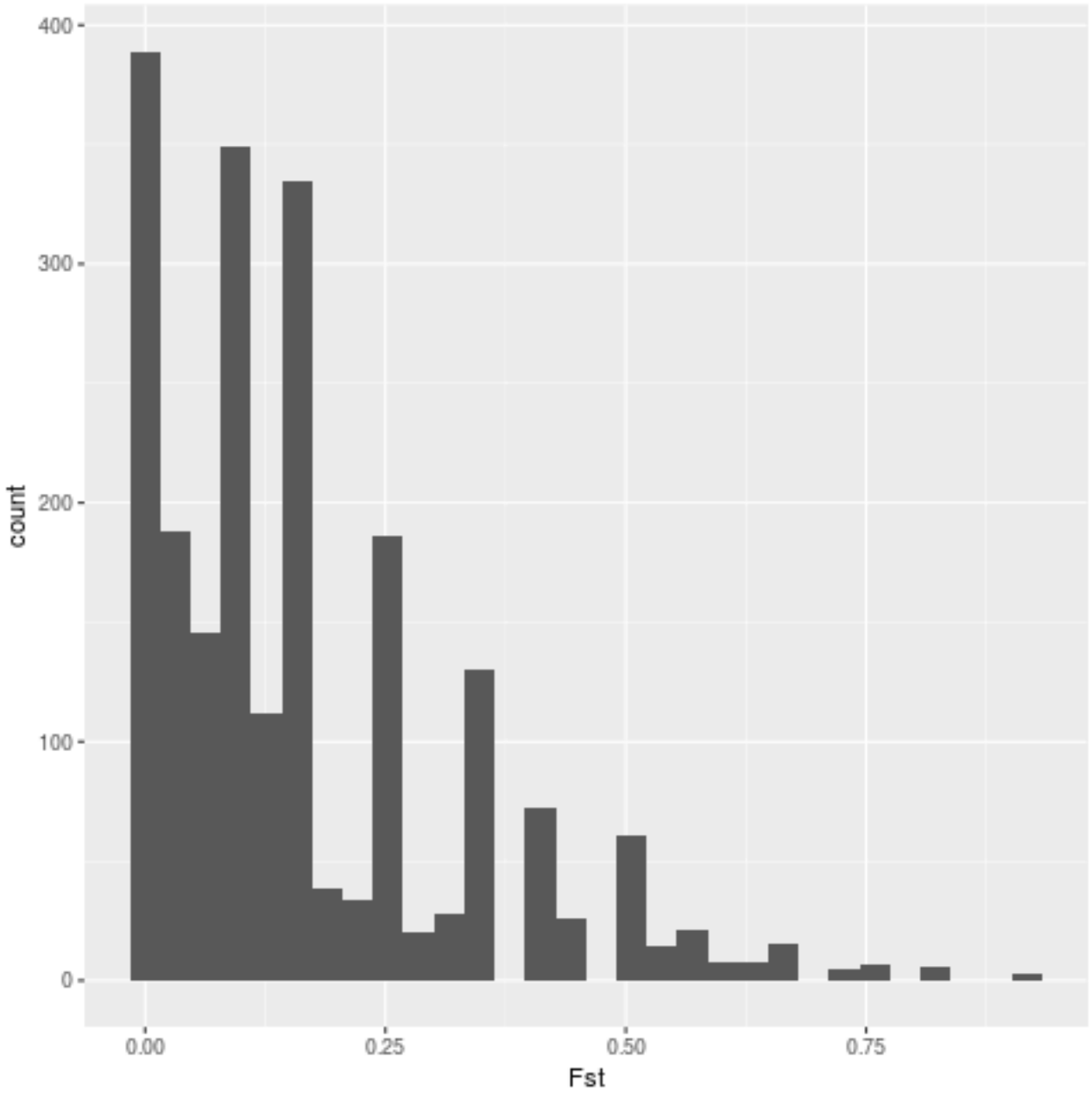

Figure SI-5.4. F_{ST} of moderate impact variants in anterior pituitary samples.

Table SI-5.20. Filtered variants called in adrenal samples, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus

between tame and aggressive samples; regions of putative selective sweeps in which each variant falls, if any, as defined in (Kukekova et al., under review); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File adr-filt-SI-5.20.xlsx

Table SI-5.21. Variants called in adrenal samples with high or moderate impact (as annotated by SNPeff), including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST} calculated at this locus between tame and aggressive samples; regions of putative selective sweeps in which each variant falls, if any,); percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); and amino acid position (amino acid position / total number of amino acids). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File adr-snpeffhighmod-SI-5.21.xlsx

Table SI-5.22. Genes associated with high and moderate impact variants in adrenals samples, including gene name, whether or not it is differentially expressed in tame versus aggressive fox adrenals, and a short description of the gene.

File adr-snpeffhighmod-genes-SI-5.22.xlsx

Table SI-5.23. Gene ontology enrichment analysis for genes associated with HMI variants in adrenal samples, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with HMI variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with HMI variants.

File adr-ClueGO-snpeffhighmod-SI-5.23.xls

Table SI-5.24. Variants called in adrenal samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); amino acid position (amino acid position / total number of amino acids); and distance (in base pairs) to the start of the next feature (e.g., for upstream variants, distance to beginning of coding region). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File adr-highfst-SI-5.24.xlsx

Table SI-5.25. Genes associated with variants in adrenals samples that have $F_{ST} > 0.5$, including gene name, whether or not it is differentially expressed in tame versus aggressive fox adrenals, and a short description of the gene.

File adr-highfst-genes-SI-5.25.xlsx

Table SI-5.26. Windows in regions of putative selective sweeps (as described in Kukekova et al., under review), including the number of hypothalamus variants from this study that fall within

each window, number of region that the window falls within, fox and dog chromosomes where the window is found, start and end of the window location on the dog chromosome, type of window (pooled heterozygosity in tame animals, pooled heterozygosity in aggressive animals, or F_{ST} between tame and aggressive animals; some windows are marked "merge" to indicate that they lie between two windows of the indicated type, e.g., between two windows of reduced pooled heterozygosity in tame animals), fox scaffold number with start and end of the window in that scaffold, pooled heterozygosity of tame samples in that window, pooled heterozygosity of aggressive samples in that window, F_{ST} between tame and aggressive samples in that window, genes in the window with high-F_{ST} variants called in adrenals in this study, and differentially expressed genes in adrenals from this study found in this window. (Note that samples used to calculate Hp and F_{ST} in windows are different from the samples used in this study.)

File adr-regions-SI-5.26.xlsx

Table SI-5.27. Gene ontology enrichment analysis for genes in adrenal samples that are associated with variants with $F_{ST} > 0.5$, sorted by corrected p-value, including GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with high-F_{ST} variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with high-F_{ST} variants.

File adr-ClueGO-highfst-SI-5.27.xls

Table SI-5.28. HMI variants called in adrenal samples with $F_{ST} > 0.5$, including chromosome and location in CanFam 3; reference and alternate alleles; associated gene name; F_{ST}; regions of putative selective sweeps in which each variant falls, if any; percent (frequency) of the alternate allele in the tame samples; percent (frequency) of the alternate allele in the aggressive samples; impact, as annotated by SNPeff; type of variant, as annotated by SNPeff; Ensembl ID of associated gene; Ensembl ID of associated transcript; description of variant; rank of exon or intron (position of exon or intron / total number of exons or introns); and

amino acid position (amino acid position / total number of amino acids). Where multiple SNPeff annotations were available for a variant, only the first was reported.

File adr-highhigh-SI-5.28.xlsx

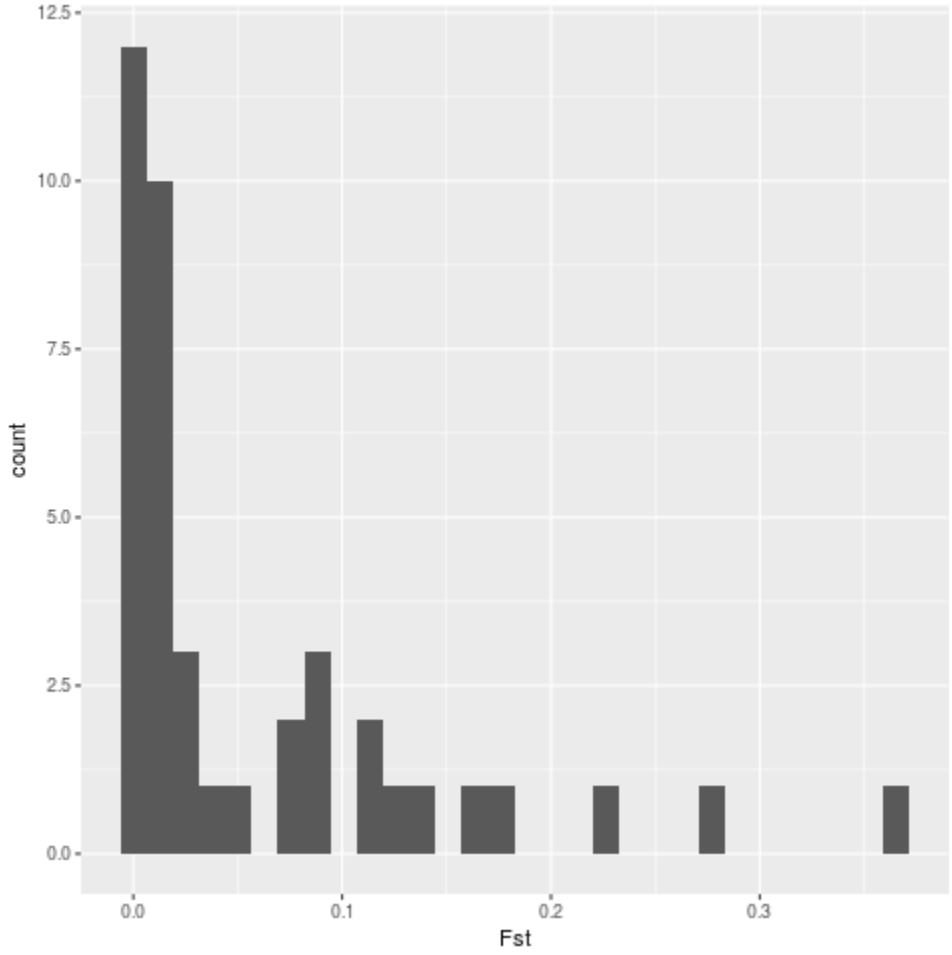

Figure SI-5.5. F_{ST} of high impact variants in adrenal samples.

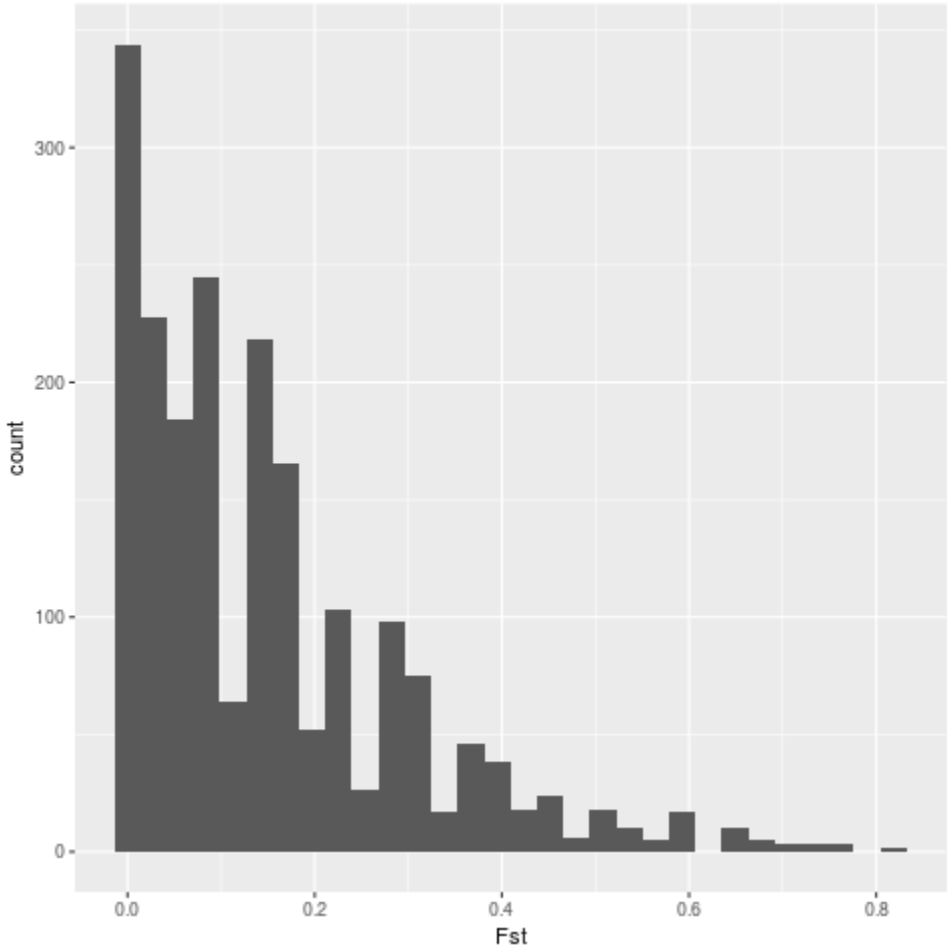

Figure SI-5.6. F_{ST} of moderate impact variants in adrenal samples.

Table SI-5.28. Variants from hypothalamus, anterior pituitary, and adrenals that have FST > 0.5 and are found in at least one putative selective sweep region, as defined by (Kukekova et al., under review), including chromosome, locus, gene name, region(s) in which the variant falls, and the type of variant, as annotated by SNPeff.

File hpa-highfst-regions-SI-5.28.xlsx

Table SI-5.29. Gene ontology enrichment analysis for genes in hypothalamus, anterior pituitary, and adrenals samples that are associated with variants with F_{ST} > 0.5 and are found in at least one putative selective sweep region, as defined in (Kukekova et al., under review). The terms

are sorted by corrected p-value, and columns include GOID, name of GO term, p-value of GO term enrichment (uncorrected), p-value of GO term enrichment (corrected with Benjamini-Hochberg), the percent of genes in this term that are associated with high-F_{ST} variants, the total number of genes in this term, and the symbols of the genes in this term that are associated with high-F_{ST} variants.

File hpa-ClueGO-highfst-regions-SI-5.29.xls